July 2023

Proceeding of
The 12th International Seminar on Linear Algebra and its Applications

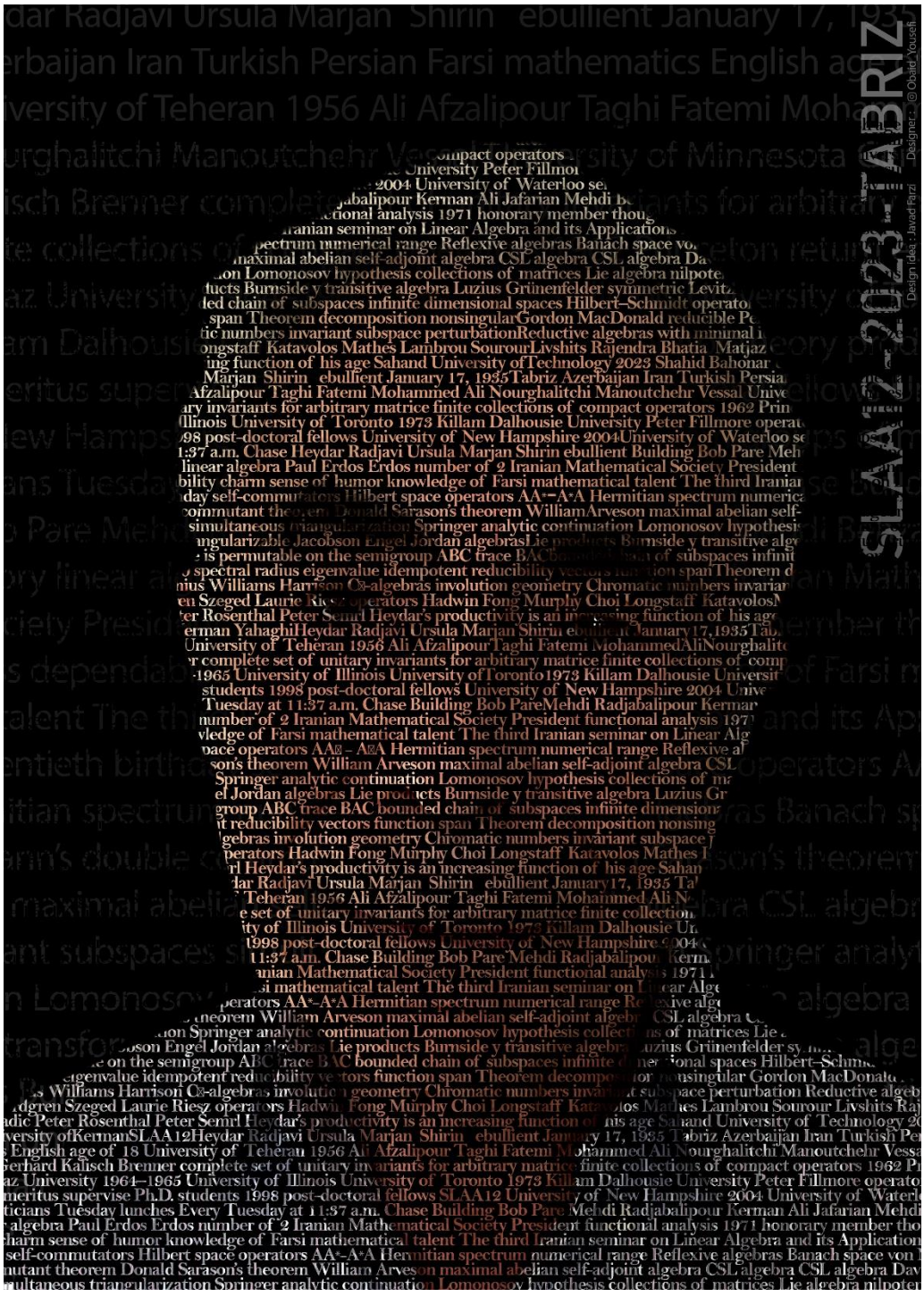
Javad Farzi
Yousef Zamani
Ildar Sadeghi
Sahand University of Technology

Proceeding of
The 12th International Seminar on Linear Algebra and its Applications

Javad Farzi
Yousef Zamani
Ildar Sadeghi
Sahand University of Technology

Seminar Sponsors

In honor of **Heydar Radjavi**



Typography of Heydar Radjavi

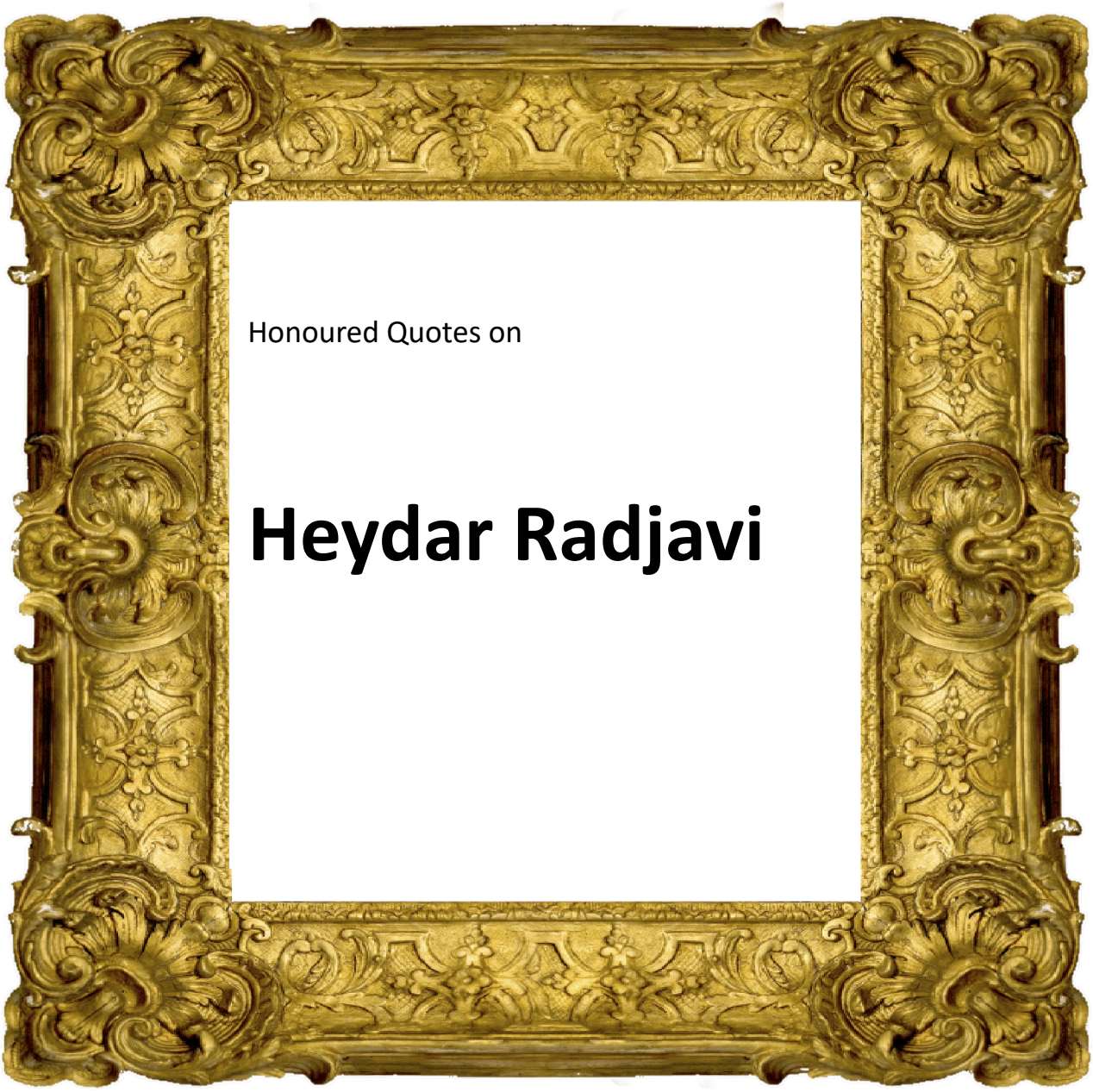
Table of Contents

Preface

January 16, 1996, is the birthday of the “Seminar on Linear Algebra and its Applications”, a biannual Iranian Seminar on Linear Algebra. The 12th SLAA took place from 18-19 July 2023 at Sahand University of Technology in Tabriz, Iran. There were about 80 contributions in this seminar, including keynote talks, a workshop, and oral and poster presentations. This event was dedicated to appreciating Professor Heydar Radjavi, for his outstanding research in linear algebra. Heydar was born in Tabriz and continued to be one of the pioneering Iranian people who got a PhD in mathematics. Although Heydar spent most of his career outside of Iran, he had a role in establishing the Iranian mathematical society. He has also trained many students and has kept in touch with Iranian researchers to help develop mathematical research in Iran. In December 2004, the 3rd SLAA was held in honor of the 70th birthday of Professor Heydar Radjavi in Kerman, Iran. Now, near his 90th birthday, participants celebrated him virtually on ZOOM in the closing ceremony of the seminar.

Thanks are due to scientific committee members who helped in assuring timely and professionally peer-reviewed submissions. Thanks are also due to the Sahand University of Technology team and many others behind the scenes. Finally, special thanks go to the artists who brought the ideas into the two beautiful portraits of Heydar, one of them is an oil painting and the other is a typography portrait of him, in which the words are mostly borrowed from a paper from a Special Issue of Linear Algebra and its Applications in honor of Heydar's 70th birthday[1].

[1] R. Bhatia, M. Omladic, P. Rosenthal and P. Semrl, A survey of Heydar Radjavi, Linear Algebra and its Applications 383 (2004) 1–15.



Honoured Quotes on

Heydar Radjavi

PETER ROSENTHAL

EMERITUS PROFESSOR OF MATHEMATICS,

UNIVERSITY OF TORONTO, CANADA.

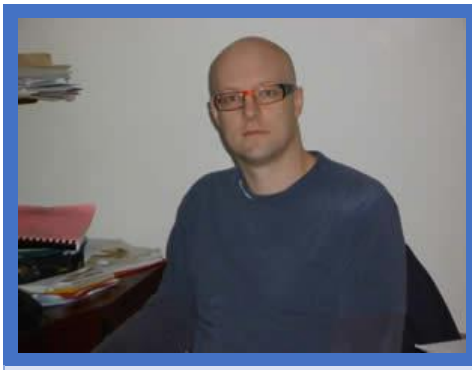
The third Iranian Seminar on Linear Algebra and its Applications, held in Kerman in December 2004, was in honour of the seventieth birthday of Heydar Radjavi. It is fitting that the twelfth such Seminar is also in Heydar's honour. Issue 383 (2004) of the journal Linear Algebra and its Applications was in Heydar's honour. I was co-author (along with Rajendra Bhatia, Matjaz Omladic, and Peter Semrl) of the Preface to that issue, titled A Survey of Heydar Radjavi. The Preface contains a brief description of his life and an overview of his mathematics up until 2004. It begins as follows.

★★★★★★★

Heydar Radjavi is seventy years old? Impossible; he's too vigorous! He cant be seventy; he's too productive! Seventy? That can't be true; he's too good-looking! It is true; vigorous, productive and good-looking as he is, Heydar Radjavi is seventy years old as of January 17, 2005. Most of us slow down, at least a bit, as we enter our sixties. Not Heydar. As his list of publications establishes (see the end of this article for complete references to his research papers to date, followed by a list of his books), Heydar's productivity is an increasing function of his age. As his many collaborators know, it is a great pleasure to work with Heydar. He is a very talented and knowledgeable mathematician. He loves thinking and talking about mathematics and working with others. His enthusiasm never seems to wane, even when numerous attacks on a problem fail, and even on those occasions when an unfillable gap is found in what the collaborating group had thought was a really nice discovery. He is helpful and pleasant to everyone, and is not at all competitive. Heydar is almost always in great spirits: the joy he finds in mathematics is part of his overall joy in life. He is one of the few people to whom the word ebullient is truly applicable. His lectures are wonderful: they are invariably very interesting and clear, and peppered with Heydar's special humor. He is, overall, the nicest kind of human being. It is a great pleasure to visit Heydar and his wife Ursula. In particular, they are great cooks: A nontrivial corollary of working with Heydar is the opportunity to sample the delicious Iranian meals he and Ursula prepare. This survey consists of a brief look at Heydar's background followed by discussions of a few of the highlights of his mathematical work.

★★★★★★★

It is hard to believe but, over the time period from Heydar's seventieth birthday to today, Heydar's productivity has continued to be an increasing function of his age. Moreover, he is still ebullient.



MITJA MASTNAK
PROFESSOR OF MATHEMATICS
SAINT MARY'S UNIVERSITY, CANADA.

It is important to acknowledge and celebrate the exceptional contributions of individuals who made significant advancements in their fields and also enriched the lives of everyone around them; Heydar Radjavi is undoubtedly one of them.

Heydar's influence on my professional journey and my life in general cannot be overstated. I first briefly met him during his visit to Matjaz Omladic, my honours thesis supervisor, at University of Ljubljana in the 1990's. Even though I was a lowly undergrad, Heydar had time to have a nice chat we me. We talked mostly about mathematics, but also discussed some worldly things. For example, Heydar's underappreciation for the amount of egg yolk in omelets in his hotel.

When I started my graduate studies at Dalhousie, I had the privilege of taking the last course he ever taught there: a reading course on simultaneous triangularization. Heydar also ran an informal linear algebra seminar that allowed me to join the fellowship of Heydar's coauthors even before I graduated.

Then there were Tuesday lunches and, when Heydar had visitors, lavish dinners at the Radjavi household. Heydar taught me how to properly cook rice. I was able to return that favour about two decades later when my daughters taught him how to make tacos out of the weeds in our back yard.

During a crucial time when I was seeking employment, Heydar extended a helping hand and offered me a position as a postdoctoral fellow at the University of Waterloo. I will always be grateful for his support and the trust he placed in me. Working alongside Heydar Radjavi has been a truly transformative experience. Even in projects where he wasn't a direct coauthor, his influence has been immense.

Recent pandemic also has some good side-effects: I now get to Zoom with Heydar every week. Mathematicians have one of the best jobs in the world, but those few of us (actually not quite so few) that collaborate with Heydar are fortunate indeed. I was thinking hard about my favourite joint project with Heydar and came to the obvious conclusion: it is our next one. Well, probably the one after that.

In Halifax, July 14, 2023
Mitja Mastnak

KEN DAVIDSON
EMERITUS PROFESSOR OF MATHEMATICS
UNIVERSITY OF WATERLOO, CANADA.

Dear Heydar,

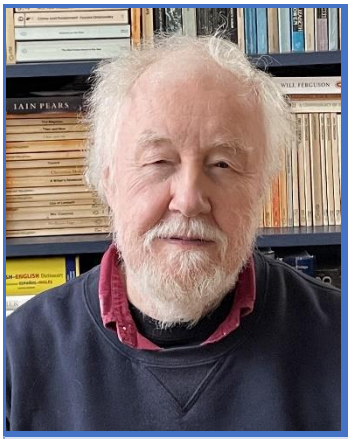
It has been a pleasure to know you, discuss math with you, and to call you my friend. In particular during the past 20 years, during your ``retirement'', you have been anything but retired.

Your continued interest in mathematical questions is a source of inspiration to me as a newly ``retired'' person.

Also we have had many interesting conversations over a Friday get togethers, even in the past three years thanks to Zoom!

I wish you well on the occasion of this conference.

Regards, Ken



BOB PARE
EMERITUS PROFESSOR OF MATHEMATICS
DALHOUSIE UNIVERSITY, CANADA.

Heydar Radjavi

There once was a prof from Tabriz
Who said solving problems' a breeze
Just make a big matrix
Then use everyday tricks
You can prove whatever you please

Cheers,

Bob

CHELLURI SASTRI
EMERITUS PROFESSOR OF MATHEMATICS
DALHOUSIE UNIVERSITY, CANADA.

Heydar and I have been friends since 1980, and I must say that that friendship has sustained me through all these years. What we have in common is a love of reading and walking; he is also a writer who has written two memoirs; although I haven't published anything, I have written a bit.

Heydar and I have only one joint paper, in which he helped answer a question of mine in the area of unobserved probability. What impressed me most was his showing that an example I had come up with, with some help from my late son Raju, encapsulated everything that the general case entailed. Remarkable!

Regards and best wishes,

Sastri

YUANHANG ZHANG
ASSISTANT PROFESSOR OF MATHEMATICS
JILIN UNIVERSITY, CHINA.

I got to know Professor Heydar Radjavi in 2005 when I prepared my undergraduate thesis. My supervisor gave me a Chinese translation of his famous book “Invariant Subspaces”, co-authored with Professor Peter Rosenthal. This book introduced me to the fascinating field of operator theory.

In 2019, I had the opportunity to visit Professor Laurent Marcoux and work with him and Professor Heydar Radjavi. I often attended the long-standing Tuesday Lunches, where I learned math and other interesting stories from Heydar. I have many precious memories of that time!

After returning to China, we continued to have regular weekly Zoom meetings. Sometimes, Laurent or I would find results in the literature that may be useful for our project, and Heydar would say: “That’s a nice result! Who proved it?” We would reply: “Haha, you did!” Yes, Heydar has so many influential and substantial results, and (I bet that) he will keep doing math even more productively. I am honored and grateful for the collaboration with Heydar and Laurent, which has influenced my career.

Thank you!

The 12th International Seminar on Linear Algebra and its Applications

Sahand University of Technology, 18-19 July 2023

دوازدهمین سمینار بین‌المللی جبرخطی و کاربردهای آن،

دانشگاه صنعتی سهند، ۲۷ و ۲۸ تیرماه ۱۴۰۲

برنامه دوازدهمین سمینار بین‌المللی جبرخطی و کاربردهای آن

سه شنبه قبل از ظهر ۲۷ تیرماه ۱۴۰۲

7:30 – 8:30	پذیرش و ثبت نام	
8:30 – 9:00	مراسم افتتاحیه (آمفی تئاتر شهید آوینی)	
9:00 – 9:45	سخنرانی کلیدی	دکتر مهدی رجبعلی پور، دکتر بامدادرضا یاحقی: در باب استادم حیدر رجوی (آمفی تئاتر شهید آوینی)
9:45 – 10:20		مراسم تجلیل (آمفی تئاتر شهید آوینی)
10:20 – 10:50		پذیرایی
10:50 – 11:40	سخنرانی کلیدی	آمفی تئاتر شهید آوینی(دکتر مهدی رجبعلی پور) کلاس A (دکتر علی محمد نظری)
		Dr. Abbas Salemi Parizi Spectra and pseuduo-spectra of matrices
		Dr. Fatemeh Panjeh Ali Beik A two-dimensional minimum residual technique for accelerating two-step iterative solvers

محل برگزاری: رئیس جلسه:	کلاس A دکتر فرشید عبداللهی	کلاس B دکتر قدرت عبادی	کلاس C دکتر عباس سالمی
11:40 – 12:00	**احمد صفا پور** : قاب ها، نیم-قاب ها و نمایش های القا شده	**فرنگیس کیان فر:** شکست روشهای RRGMRES و DGMRES	**بامداد رضا یاحقی:** توسیع هایی از قضیه اساسی جبر
12:00 – 12:20	**مهدی رشیدی کوچی:** پایداری g- قابهای تلفیقی در هم تنیده	**مریم حاجی صادقی اصفهانی:** کاربرد روش منظم سازی تیخونوف تکراری آرنولدی برای بازسازی شکل یک جسم ناهمسانگرد در پراکندگی معکوس امواج الکترومغناطیسی	**علی محمد نظری:** بعضی از خواص یک ماتریس همراه خاص و توان‌های آن
12:20 – 12:40	**اصغر رحیمی:** آشکارسازی کوانتومی با استفاده از قابهای همجوش	**مریم شمس سولاری :** برخی خواص ماتریسهای هسنبرگ توپلیتز بلوکی	**محسن کیان:** نمایش ماتریسی نگاشت های چندخطی
12:40 – 14:00		ناهار و نماز	

سه شنبه بعدازظهر ۲۷ تیرماه ۱۴۰۲

14:00 – 14:50	سخنرانی کلیدی	آمفی تئاتر شهید آوینی(رئیس: دکتر محمدصال مصلحیان) **Dr. Ali Zamani** Numerical Radius: New Extensions and Inequalities	کلاس A(رئیس جلسه: دکتر فاطمه پنجه علی بیگ) **Dr. Mansoor Rezghi** Matrix and tensor modeling in Artificial intelligence and data science

محل برگزاری: رئیس جلسه:	کلاس A دکتر علی تقوی	کلاس B دکتر منصور رزقی	کلاس C دکتر بامداد یاحقی
14:50 – 15:10	**لیلا عابدینی:** نگاشت های غیر خطی حافظ ضرب های چند گانه	**سریاس وکیلی :** یک پیش شرط ساز شکافت- انتقال سریع جدید برای مسائل نقطه زینی	**علی محمد نظری :** تاریخچهٔ ماتریس و دترمینان قبل از قرن بیستم
15:10 – 15:30	**روجا حسین زاده:** نگاشت های به طور کامل نگهدارنده		**سید صادق غلامی :** کلاسهای تقارن دکارتی متناظر با گروه دو وجهی
15:30 – 15:50	**علی آرمند نژاد:** نگهدارنده های خطی G- ماتریس ها روی M_2	**بنت الهدی زالی:** روش تکراری برای حل معادلات تنسوری مقید	**یوسف زمانی :** کلاسهای تقارن دکارتی تعمیم یافته
15:50-16:20	پذیرایی و ارائه پوستر		

رئیس جلسه ارائه پوستر: دکتر حنیف میرزایی (۱- مریم شمس سولاری: برخی حقایق در مورد ماتریس های شبه بلوکی توپلیتز، ۲- ماجد حمادی : کران‌هایی برای نرم توابع ماتریسی، ۳- مهدی کشتکار : بهینه سازی هندسی تک جمله ای با استفاده از سیستم نامعادلات رابط – فازی، ۴- فاطمه شایان فرد: بررسی حل پذیری دستگاه معادلات حاصل از روش هم محلی برای حل معادله انتگرال ولترای نوع سوم)

محل برگزاری: رئیس جلسه:	کلاس A دکتر محسن کیان	کلاس B دکترحنیف میرزایی	کلاس C دکتر فرشید عبداللهی
16:20 – 16:40	**جواد فرخی استاد:** ویژگی فوگلد-پاتنم و وارون مور- پنروز یک عملگر	**فرناز خیر خواه :** روش شبه طیفی لژاندر برای مسأله کنترل بهینه حاصل از یک معادله نفوذ کسری زمان	**دکتر فرشید عبداالهی** کارگاه آموزشی جبر خطی تجربی
16:40 – 17:00	**علی تقوی:** *-مشتق های سه تایی جردن روی *- جبرهای اول	**رعنا اکبری:** کاربرد ماتریس های معین مثبت در روش های عددی برای ODE ها	
17:00 – 17:20	**اسماعیل نیکوفر:** معرفی برخی آنتروپی های عملگری	**فاطمه شایان فرد:** یک روش تقریبی برای محاسبه ماتریس انتقال حالت	

17:20-18:10	سخنرانی کلیدی	آمفی تئاتر شهید آوینی (رئیس جلسه: دکتر عباس سالمی) **Dr. Tin-Yau Tam** Ballantine's type theorem for complex symplectic group	کلاس A (رئیس جلسه: دکتر علی محمد نظری) **Dr. Kazem Ghanbari** Positive Classes of Matrices

بازدید از دانشکده معماری دانشگاه هنر اسلامی تبریز و صرف شام سمینار

The 12th International Seminar on Linear Algebra and its Applications

Sahand University of Technology, 18-19 July 2023

دوازدهمین سمینار بین‌المللی جبرخطی و کاربردهای آن،

دانشگاه صنعتی سهند، ۲۷ و ۲۸ تیرماه ۱۴۰۲

چهار شنبه قبل از ظهر ۲۸ تیرماه ۱۴۰۲

8:30 – 9:20	سخنرانی کلیدی	آمفی تئاتر شهید آوینی (رئیس جلسه: دکتر فرشید عبداللهی)	کلاس A (رئیس جلسه: دکتر معصومه آذغانی)
		Dr. Rajab Ali Kamyabi Gol Fourier Like Systems, Frame of Translates and their Oblique Duals on LCA-groups	**Dr. Arash Amini** Stationary Graph Signals

		کلاس A دکتر اصغر رحیمی	کلاس B دکتر پیام مختاری	کلاس C دکتر رحیم رحمتی اصغر
	محل برگزاری: رئیس جلسه:			
9:20 – 9:40	سخنرانی تخصصی	**فهیمه سلطان زاده :** برخی اتحادها و نامساویها برای دو- قاب ها در فضاهای دو ضرب داخلی	**محمد جواد خسرو جردی :** جایابی ساختار ویژه: ابزاری موثر از جبرخطی جهت تشخیص عیب	**حمید رضا بامداد :** تخمین اندیس استرادا
9:40 – 10:00		**وحید ابراهیمی :** برخی خواص قابهای فازی	**محمد شریفی:** کاربرد سری فوریه در بدست آوردن چند جمله‌ای پایداری روش‌های چند مقداری برای معادلات دیفرانسیل معمولی	
10:00 – 10:20		**محمد رضا فروتن:** هسته های معین مثبت و هسته بازتولید فضاهای هیلبرت	**لیلا طاهری کلتپه:** تحلیل پایداری دسته ای از روش های هم محلی چند گامی برای معادلات دیفرانسیل معمولی	**حسین مهدوی پور :** پیرامون کدگذاری اندیسی انعطاف پذیر منبع
10:20 – 10:50	پذیرایی و ارائه پوستر			
		رئیس جلسه ارائه پوستر: دکتر مجتبی حاجی پور (۱- سید ابوالفضل شاهزاده فاضلی: حل معادلات ماتریسی خطی به روش بهینه سازی تکامل تفاضلی، ۲- وحید ابراهیمی: برخی خواص فضاهای ضرب داخلی فازی، ۳- حسن کریمی: نگاشت های حافظ جمع موازی عملگرها)		

10:50 – 11:40	سخنرانی کلیدی	آمفی تئاتر شهید آوینی (رئیس جلسه: دکتر محمد صال مصلحیان)	کلاس A (رئیس جلسه: دکتر فاطمه پنجه علی بیگ)
		Dr. Qing-Wen Wang New Developments of Matrix and Tensor Equations	**Dr. Daniel Kressner** Randomization for solving difficult linear algebra problems

		کلاس A دکتر علی زمانی	کلاس B دکتر مجتبی حاجی پور	کلاس C دکتر سجاد مرادی
	محل برگزاری: رئیس جلسه:			
11:40 – 12:00	سخنرانی تخصصی	**سید ابوالفضل شاهزاده فاضلی:** یافتن جواب بهینه دستگاه معادلات غیر خطی با روش بهینه سازی کپک مخاطی	**هما افراز:** تقریبی برای معادله انتشار کسری زمان - مکان سازگار	**سامان بابائی کفاکی:** مقیاس بندی بهینه فرمول های بهنگام سازی شبه نیوتن کم حافظه
12:00 – 12:20		**هژیر هومئی:** توزیع حاصلضرب ماتریسهای تصادفی با بکار بردن توزیع دیریکله	**سید رسول کافی:** کنترل عدد شرطی در پیاده‌سازی روش طیفی گالرکین برای حل معادلات انتگرال آبل تعمیم یافته	**حسین سلمه ای:** تکنیک اسکالرسازی هیبرید برای حل مسائل چند هدفه درجه دوم با قیود درجه دوم
12:20 – 12:40		**سولماز سیف اللهی:** برآوردگرهای بهبود یافته نوع ریج در مدل‌های خطی چندگانه چندمتغیره		**مهدی کشتکار:** بهینه سازی هندسی با استفاده از سیستم نامعادلات رابط- فازی
12:40 – 14:00	ناهار و نماز			

چهار شنبه بعدازظهر ۲۸ تیرماه ۱۴۰۲

		کلاس A دکتر رجبعلی کامیابی گل	کلاس B دکتر قدرت عبادی	کلاس C
	محل برگزاری: رئیس جلسه:			
14:00 - 14:20	سخنرانی تخصصی	**علی مرصعی:** آنتروپی برای توابع h- محدب	**محمد شهریاری :** یک روش عددی برای حل عملگر دیراک یک بعدی کسری	**دکتر فرشید عبداالهی** کارگاه آموزشی جبر خطی تجربی
14:20 - 14:40		**اسماعیل نیکوفر:** برخی خواص توابع ضربی - جمعی با کاربردهایش	**مژگان اقبال :** دسته جدیدی از روش های چند گامی مشتق دوم برای معادلات دیفرانسیل معمولی سخت	
14:40 – 15:00		**جواد فرخی استاد :** رفتار یک عملگر مدولی بر حسب اجزایش در C*- مدول هیلبرت		
15:00 – 15:20		**سیده سمیه جعفری:** وجود عملگرهای مثبت متناظر با گروههای فشرده موضعی	**جواد فرضی:** یک روش چند شبکه ای FAS برای قوانین بقای هذلولوی	
15:20 – 15:50	پذیرایی و ارائه پوستر			
		رئیس جلسه ارائه پوستر: دکتر سجاد مرادی(۱- محمد شهریاری: الگوریتمی کارآمد برای حل عملگرهای دیفرانسیل اشتورم - لیوویل کسری با تاخیر ثابت، ۲- رحیم رحمتی اصغر : محاسبه پایه های گروبنر مجموعه توسیع یافته از چند جمله ای ها، ۳- هژیر هومئی و منیژه جلیلوند : باز نویسی توزیع ناداراجا و کاتز، ۴- آیدا موسوی پایداری روش‌های دو-گامی رانگ- کوتا ، ۵- محمدمهدی منصوریان: تبدیل λ میانگین از عملگرها)		

		کلاس A دکتر پیام مختاری	کلاس B دکتر علی آرمند نژاد	کلاس C دکتر اصغر رحیمی
	محل برگزاری: رئیس جلسه:			
15:50 – 16:10	سخنرانی تخصصی	**مجتبی حاجی پور:** آنالیز همگرایی و پایداری یک روش عددی با مرتبه دقت بالا برای مسآله نفوذ کسری	**حنیف میرزایی :** ساختن ماتریسهای کاملا مثبت	**رحیم رحمتی اصغر:** کلاس هایی از مجتمع های سادکی مِنگِرَین
16:10 - 16:30		**عبداله الهی:** حل معادلات دیفرانسیل کسری با استفاده از موجک‌های بی اسپلاین خطی	**محمد رضا فروتن :** نظریه قاب ها و هسته بازتولید فضاهای هیلبرت	**بیاض دارابی :** مطالعه برخی نامساویهای انتگرالی برای شبه انتگرالها
16:30	آمفی تئاتر شهید آوینی **مراسم اختتامیه**			

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Ballantine's type theorem for complex symplectic group[1]

Tin-Yau Tam*

University of Nevada, Reno, Nevada, USA.

Abstract

In the late 1960 Ballantine showed that every matrix with positive determinant is a product of five positive definite matrices.

We consider the complex symplectic group $\mathrm{Sp}(2n,\mathbb{C})$:

$$\mathrm{Sp}(2n,\mathbb{C}) = \left\{A \in \mathrm{GL}(2n,\mathbb{C}) : A^\top J_n A = J_n\right\},$$

where

$$J_n = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

The symplectic group is a classical group defined as the set of linear transformations of a $2n$-dimensional vector space over $\mathbb{C}$, which preserve the non-degenerate skew-symmetric bilinear form that is defined by J_n. We show that every symplectic matrix is a product of five positive definite symplectic matrices. We also show that five is the best in the sense that there are symplectic matrices which are not product of less.

This is a joint work with Daryl Q. Granario, De La Salle University, Philippines.

Keywords: Ballantine's theorem, Radjavi's theorem, complex symplectic group, symplectic positive definite matrices

References

[1] C.S. Ballantine, Products of positive definite matrices. I, *Pacific J. Math.,* **23** (1967) 427–433.

[2] C.S. Ballantine, Products of positive definite matrices. II, *Pacific J. Math.,* **24** (1968) 7–17.

[3] C.S. Ballantine, Products of positive definite matrices. III, *J. Algebra*, **10** (1968) 174–182.

[4] C.S. Ballantine, Products of positive definite matrices. IV, *Linear Algebra Appl.* **3** (1970) 79–114.

[5] J. Cui, C.-K. Li, and N.-S. Sze, Products of positive semi-definite matrices, *Linear Algebra Appl.,* **528** (2017) 17–24.

[1] This is a joint work with Daryl Q. Granario, De La Salle University, Philippines.

*Speaker. Email address: ttam@unr.edu

[6] D.Q. Granario and T.-Y. Tam, Products of positive definite symplectic matrices, *Linear Algebra Appl.,* **626** (2021) 188–203.

[7] H. Radjavi, Products of Hermitian matrices and symmetries, *Proc. Amer. Math Soc.,* **21** (1969) 369–372; Errata, *Proc. Amer. Math. Soc.,* **26** (1970) 701.

[8] P.Y. Wu, Products of positive semidefinite matrices, *Linear Algebra Appl.,* **111** (1988) 53–61.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The new developments of matrix and tensor equations

Qing-Wen Wang*

Department of Mathematics, Shanghai University, P. R. China

Abstract

In this talk, I give a brief introduction to some new developments in systems of Sylvester-type matrix equations and tensor equations.

Keywords: Quternion algebra, matrix equation, rank, Moore-Penrose inverse

References

[1] X.Y. Chen, Q.W. Wang, The η-(anti-)Hermitian solution to a constrained Sylvester-type generalized commutative quaternion matrix equation, *Banach J. Math. Anal.*, (2023) DOI: 10.1007/s43037-023-00262-5.

[2] X.L. Xu, Q.W. Wang, The consistency and the general common solution to some quaternion matrix equations, *Ann. Funct. Anal.*, (2023) DOI: 10.1007/s43034-023-00276-y.

[3] B.Y. Ren, Q.W. Wang, X.Y. Chen, The η-anti-Hermitian solution to a system of constrained matrix equations over the generalized Segre quaternion algebra, *Symmetry*, 15 (2023) DOI: 10.3390/sym15030592.

[4] T. Li, Q.W. Wang, Structure preserving quaternion full orthogonalization method with applications, *Numer. Linear Algebra Appl.*, (2023) DOI: 10.1002/nla.2495.

[5] L.M. Xie, Q.W. Wang, A system of matrix equations over the commutative quaternion ring, *Filomat*, 37 (1) (2023) 97-106 DOI: 10.2298/FIL2301097X.

[6] L.S. Liu, Q.W. Wang, The reducible solution to a system of matrix equations over the Hamilton quaternion algebra, *Filomat*, 37 (9) (2023) DOI: 10.2298/FIL2309731L.

[7] W.W. Li, H. Xin, Q.W. Wang, The canonical forms of permutation matrices, *Symmetry*, 15 (2) (2023) DOI: 10.3390/sym15020332.

[8] L.S. Liu, Q.W. Wang, J.F. Chen, Y.Z. Xie, An exact solution to a quaternion matrix equation with an application, *Symmetry*, 14 (2) (2022) DOI: 10.3390/sym14020375.

*Speaker. Email address: wqwshu9@126.com

[9] S. Dong, Q.W. Wang, L. Hou, Determinantal inequalities for block Hadamard product and Khatri-Rao product of positive definite matrices, *Mathematics*, 7 (6) (2022) DOI: 10.3390/math2022536.

[10] A. Farouk, Q.W. Wang, Construction of new Hadamard matrices using known Hadamard matrices, *Filomat*, 36 (6) (2022) 2025-2042 DOI: 10.2298/FIL2206025F.

[11] M.Y. Xie, Q.W. Wang, Z.H. He, Mahmoud S. Mehany, A System of Sylvester-type Quaternion Matrix Equations with Ten Variables, *Acta Mathematica Sinica-English Series*, 38 (8) (2022) 1399-1420 DOI: 10.1007/s10114-022-9040-1.

[12] M.Y. Xie, Q.W. Wang, Y. Zhang, The Minimum-Norm Least Squares Solutions to Quaternion Tensor Systems, *Symmetry*, 14 (7) (2022) DOI: 10.3390/sym14071460.

[13] Y.F. Xu, Q.W. Wang, L.S. Liu, Mahmoud S. Mehany, A constrained system of matrix equations, *Computational and Applied Mathematics*, 41 (4) (2022) DOI: 10.1007/s40314-022-01873-8.

[14] T. Li, Q.W. Wang, X.F. Zhang, A Modified Conjugate Residual Method and Nearest Kronecker Product Preconditioner for the Generalized Coupled Sylvester Tensor Equations, *Mathematics*, 10 (10) (2022) DOI: 10.3390/math10101730.

[15] L.S. Liu, Q.W. Wang, Mahmoud S. Mehany, A Sylvester-Type Matrix Equation over the Hamilton Quaternions with an Application, *Mathematics*, 10 (10) (2022) DOI: 10.3390/math10101758.

[16] R.N. Wang, Q.W. Wang, L.S. Liu, Solving a System of Sylvester-like Quaternion Matrix Equations, *Symmetry*, 14 (5) (2022) DOI: 10.3390/sym14051056.

[17] Mahmoud S. Mehany, Q.W. Wang, Three Symmetrical Systems of Coupled Sylvester-like Quaternion Matrix Equations, *Symmetry*, 14 (3) (2022) DOI: 10.3390/sym14030550.

[18] X.F. Zhang, Q.W. Wang, On RGI Algorithms for Solving Sylvester Tensor Equations, *Taiwanese J. Math.*, 26 (3) (2022) 501-519 DOI: 10.11650/tjm/220103.

[19] W.J. Yuan, Q.W. Wang, The Common Solution of Twelve Matrix Equations over the Quaternions, *Filomat*, 36 (3) (2022) 887-903 DOI: 10.2298/FIL2203887Y.

[20] T. Li, Q.W. Wang, X.F. Zhang, Gradient based iterative methods for solving symmetric tensor equations, *Numer. Linear Algebra Appl.*, 29 (2) (2022) DOI: 10.1002/nla.2414.

[21] D.M. Zhou, X.X. Ye, Q.W. Wang, J.W. Ding, W.Y. Hu, Explicit Solutions of the Yang-Baxter-Like Matrix Equation for a Singular Diagonalizable Matrix With Three Distinct Eigenvalues, *Filomat*, 35 (12) (2021) 3971-3982 DOI: 10.2298/FIL2112971Z.

[22] Q.W. Wang, X.J. Xu, X.F. Duan, Least squares solution of the quaternion Sylvester tensor equation, *Linear Multilinear Algebra*, 69 (1) (2021) 104–130.

[23] H.H. Zhu, L.Y. Wu, Q.W. Wang, Suitable elements, *-clean elements and Sylvester equations in rings with involution, *Commun. Algebra*, (2021) DOI: 10.1080/00927872.2021.1985129.

[24] D. Cvetkovic-Ilic, Q.W. Wang, Y.M. Wei, Different completions of A plus CX, *J. Spectr. Theor.*, 11 (3) (2021) 905–914 DOI: 10.4171/JST/356.

[25] H.H. Zhu, Q.W. Wang, Weighted Moore-Penrose inverses and weighted core inverses in rings with Involution, *Chinese Ann. Math. Ser. B*, 42 (4) (2021) 613-624.

[26] X.F. Zhang, Q.W. Wang, Developing iterative algorithms to solve Sylvester tensor equations, *Appl. Math. Comput.*, 409 (2021) 126403.

[27] L.Q. Qi, Z.Y. Luo, Q.W. Wang, Quaternion matrix optimization: motivation and analysis, *J. Optim. Theory Appl.*, (2021) DOI: 10.1007/s10957-021-01906-y.

[28] X.J. Xu, Q.W. Wang, On the solutions of a class of tensor equations, *Linear Multilinear Algebra*, (2021) DOI: 10.1080/03081087.2021.1948492.

[29] D. S. Cvetković-Ilić, J. N. Radenković, Q.W. Wang, Algebraic conditions for the solvability to some systems of matrix equations, *Linear Multilinear Algebra*, 69 (9) (2021) 1579-1609 DOI: 10.1080/03081087.2019.1633993.

[30] X.F. Duan, S.Q. Duan, J. Li, J.f. Li, Q.W. Wang, An efficient algorithm for solving the nonnegative tensor least squares problem, *Numer. Linear Algebra Appl.*, (2021) DOI: 10.1002/nla.2385.

[31] L. Li, Q.W. Wang, S.Q. Shen, M. Li, Quantum coherence measures based on Fisher information with applications, *Phys. Rev. A*, 103 (1) (2021) DOI: 10.1103/PhysRevA.103.012401.

[32] T. Li, Q.W. Wang, X.F. Zhang, Hermitian and skew-Hermitian splitting methods for solving a tensor equation, *Int. J. Comput. Math.*, 98 (6) (2021) 1274–1290.

[33] S. Dong, Q.W. Wang, More generalizations of Hartfiel's inequality and the Brunn-Minkowski inequality, *Bull. Iranian Math. Soc.*, 47 (1) (2021) 21–29.

[34] X.F. Duan, J. Li, S.Q. Duan, Q.W. Wang, Numerical method for the generalized nonnegative tensor factorization problem, *Numer. Algorithms*, 87 (2) (2021) 499–510.

[35] Q.W. Wang, X. Wang, A system of coupled two-sided Sylvester-type tensor equations over the quaternion algebra, *Taiwanese J. Math.*, 24 (6) (2020) 1399–1416.

[36] Q.W. Wang, X. Wang, Y. Zhang, A constraint system of coupled two-sided Sylvester-like quaternion tensor equations, *Comput. Appl. Math.*, 39 (4) (2020) DOI: 10.1007/s40314-020-01370-w.

[37] M. Xie, Q.W. Wang, Reducible solution to a quaternion tensor equation, *Front. Math. China*, 15 (5) (2020) 1047–1070.

[38] Q.W. Wang, R.Y. Lv, Y. Zhang, The least-squares solution with the least norm to a system of tensor equations over the quaternion algebra, *Linear Multilinear Algebra*, (2020) DOI: 10.1080/03081087.2020.1779172.

[39] X.F. Zhang, Q.W. Wang, T. Li, The accelerated overrelaxation splitting method for solving symmetric tensor equations, *Comput. Appl. Math.*, 39 (3) (2020) DOI: 10.1007/s40314-020-01182-y.

[40] T. Li, Q.W. Wang, X.F. Duan, Numerical algorithms for solving discrete Lyapunov tensor equation, *J. Comput. Appl. Math.*, 370 (2020) DOI: 10.1016/j.cam.2019.112676.

[41] Q.W. Wang, X.X. Wang, Arnoldi method for large quaternion right eigenvalue problem, *J. Sci. Comput.*, 82 (3) (2020) DOI: 10.1007/s10915-020-01158-4.

[42] D. Cvetkovic-Ilic, Q.W. Wang, Q. Xu, Douglas' plus Sebestyen's lemmas = a tool for solving an operator equation problem, *J. Math. Anal. Appl.*, 482 (2) (2020) DOI: 10.1016/j.jmaa.2019.123599.

[43] C. Song, Q.W. Wang, Modified CGLS iterative algorithm for solving the generalized Sylvester-conjugate matrix equation, *Filomat*, 34 (4) (2020) 1329–1346.

[44] A. Farouk, Q.W. Wang, An infinite family of Hadamard matrices constructed from Paley type matrices, *Filomat*, 34 (3) (2020) 815–834.

[45] Z. Chen, L. Cao, Q.W. Wang, The extreme points of certain polytopes of doubly substochastic matrices, *Linear Multilinear Algebra*, 68 (10) (2020) 1956-1971.

[46] H.H. Zhu, Q.W. Wang, Weighted pseudo core inverses in rings, *Linear Multilinear Algebra*, 68 (12) (2020) 2434-2447.

[47] L. Li, Q.W. Wang, S.Q. Shen, M. Li, Coherence measures based on coherence eigenvalue and their applications, *Quantum Inf. Process.*, 18 (11) (2019) DOI: 10.1007/s11128-019-2461-9.

[48] X. Liu, Q.W. Wang, Y. Zhang, Consistency of quaternion matrix equations AX*-XB=C and X-AX*B=C*, *Electron. J. Linear Algebra*, 35 (2019) 394–407.

[49] L. Li, S.Q. Shen, M. Li, Q.W. Wang, Connection of coherence measure and unitary evolutions, *Quantum Inf. Process.*, 18 (6) (2019) DOI: 10.1007/s11128-019-2304-8.

[50] X.J. Xu, Q.W. Wang, Extending BiCG and BiCR methods to solve the Stein tensor equation, *Comput. Math. Appl.*, 77 (12) (2019) 3117-3127.

[51] L. Li, Y. N. Chen, M. Li, Q.W. Wang, L.Q. Qi, Computing the maximal violation of Bell inequalities for multipartite qubit via partially symmetric tensor, *Int.J. Theor. Phys.*, 58 (4) (2019) 1161-1171.

[52] Q.W. Wang, Z.H. He, Y. Zhang, Constrained two-sided coupled Sylvester-type quaternion matrix equations, *Automatica*, 101 (2019) 207–213.

[53] Z.H. He, Q.W. Wang, Y. Zhang, A simultaneous decomposition for seven matrices with application, *J. Comput. Appl. Math.*, 349 (2019) 93–113.

[54] Q.W. Wang, X.J. Xu, Iterative algorithms for solving some tensor equations, *Linear Multilinear Algebra*, 67 (7) (2019) 1325–1349.

[55] C.M. Li, X.F. Duan, L.Z. Lu, Q.W. Wang, S.Q. Shen, Iterative algorithm for solving a class of convex feasibility problem, *J. Comput. Appl. Math.*, 352 (2019) 352-367.

[56] Q.W. Wang, X.X. Yang, S.F. Yuan, The least square solution with the least norm to a system of quaternion matrix equations, *Iran J. Sci. Technol. Trans. Sci.*, 42 (2018) 1317–1325.

[57] Z.H. He, Q.W. Wang, Y. Zhang, The complete equivalence canonical form of four matrices over an arbitrary division ring, *Linear Multilinear Algebra*, 66 (1) (2018) 74–95.

[58] Z.H. He, Q.W. Wang, Y. Zhang, A system of quaternary coupled Sylvester-type real quaternion matrix equations, *Automatica*, 87 (2018) 25–31.

[59] F.O. Farid, X.R. Nie, Q.W. Wang, On the solutions of two systems of quaternion matrix equations, *Linear Multilinear Algebra*, 66 (12) (2018) 2355–2388.

[60] S.W. Huang, Q.W. Wang, Y. Zhang, Equivalence on some Rotfel'd type theorems, *Linear Multilinear Algebra*, 66 (8) (2018) 1626–1632.

[61] G.J. Song, Q.W. Wang, S.W. Yu, Cramer's rule for a system of quaternion matrix equations with application, *Appl. Math. Comput.*, 336 (2018) 490–499.

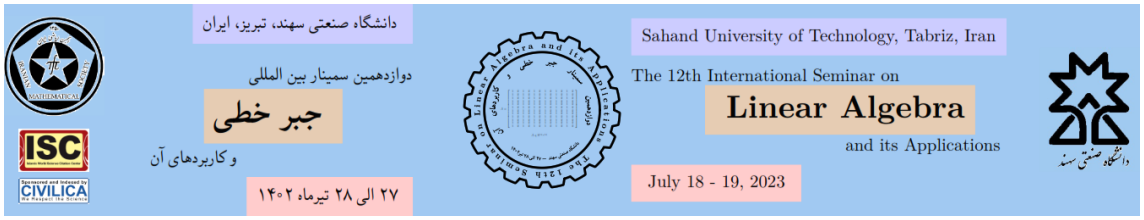
[62] S.W. Huang, C.K. Li, Y.T. Poon, Q.W. Wang, Inequalities on generalized matrix functions, *Linear Multilinear Algebra*, 65 (10) (2017) 1947–1961.

[63] J.T. Liu, Y.T. Poon, Q.W. Wang, A generalized Hölder type eigenvalue inequality, *Linear Multilinear Algebra*, 65 (10) (2017) 2145–2151.

[64] A. Rehman, Q.W. Wang, I. Ali, M. Akram, M. O. Ahmad, A constraint system of generalized Sylvester quaternion matrix equations, *Advances in Appl.* Clifford Algebras, 27(4) (2017) 3183–3196.

[65] J.T. Liu, Q.W. Wang, F.F. Sun, On Hayajneh and Kittaneh's Conjecture on unitary invariant norm, *J. Math. Inequal.*, 11 (4) (2017) 1019–1022.

[66] S.F. Yuan, Q.W. Wang, Y.B. Yu, Y. Tian, On Hermitian solutions of the split quaternion matrix equation AXB+CXD=E, *Advances in Appl.* Clifford Algebras, 27 (4) (2017) 3235–3252.



Randomization for solving difficult linear algebra problems[1]

Daniel Kressner *

EPFL, Switzerland

Abstract

Randomization is becoming an increasingly popular tool in numerical linear algebra, sometimes leading to surprisingly simple algorithms that frequently outperform existing deterministic algorithms. The poster child of these developments, the randomized singular value decomposition is nowadays one of the state-the-of-art approaches to perform low-rank approximation for large-scale matrices. In this talk, we will discuss numerous further examples for the potential of randomization to facilitate the solution of notoriously difficult linear algebra problems. This includes a simple numerical algorithm for jointly diagonalizing a family of nearly commuting matrices, a topic to which Heydar Radjavi has made seminal contributions. We will also discuss the solution of several other challenging flavors of eigenvalue problems as well as the low-rank approximation of matrix functions and matrix-valued functions. A common theme of all these developments is that randomization turns identities that only hold generically into robust numerical algorithms that come with reliability guarantees.

References

[1] Stefan Güttel, Daniel Kressner, Bart Vandereycken, Randomized sketching of nonlinear eigenvalue problems, Preprint, https://arxiv.org/abs/2211.12175.

[2] Haoze He and Daniel Kressner, Randomized joint diagonalization of symmetric matrices, Preprint, https://arxiv.org/abs/2212.07248.

[3] Daniel Kressner and Ivana Sain Glibic, Singular quadratic eigenvalue problems: Linearization and weak condition numbers, Preprint, https://arxiv.org/abs/2204.07424.

[4] Daniel Kressner and Hei Yin Lam, Randomized low-rank approximation of parameter-dependent matrices, Preprint, https://arxiv.org/abs/2302.12761.

[5] David Persson, Alice Cortinovis, Daniel Kressner, Improved variants of the Hutch++ algorithm for trace estimation, Preprint, https://arxiv.org/abs/2109.10659.

[6] David Persson and Daniel Kressner, Randomized low-rank approximation of monotone matrix functions, Preprint, https://arxiv.org/abs/2209.11023.

[1] This talk is based on joint work with Alice Cortinovis, Stefan Güttel, Haoze, Hysan Lam, David Persson, Bor Plestenjak, Ivana Sain Glibic, and Bart Vandereycken, see [1–6].

* Speaker. Email address: daniel.kressner@epfl.ch

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Positive Classes of Matrices

Kazem Ghanbari[1,2,*]

[1]Department of Mathematics, Sahand University of Technology, Tabriz, Iran

[2]School of Mathematics and Statistics, Carleton University, Ottawa, Canada

Abstract

In this lecture we present different types of positivity concept in matrix analysis. Any kind of positive matrix has own typical applications. Entrywise positivity, definite positivity, complete positivity and total positivity are the main types of positivity in matrix analysis. The concept of a positive definite matrix (PD) is well-known for most people having the elementary course in linear algebra, but the other types of positivity are not quiet well-know as PD matrices, thus we present other types of positivities in matrix theory. For more complete information on PD matrices see [1]. In linear algebra any real matrix with nonnegative entries is called *Nonnegative Matrix* (NM).

A matrix which is both nonnegative and positive semi-definite is called *doubly nonnegative matrix* (DNM). The Perron–Frobenius theorem, proved by Oskar Perron (1907) and Georg Frobenius (1912), is the most important result stating that a real square matrix with positive entries has a unique largest real eigenvalue and that the corresponding eigenvector can be chosen to have strictly positive entries. This theorem has signifant applications [2, 4–6]. If a symmetric matrix A can be factorized of the form $A = BB^T$ where B is a non-negative matrix, then A is called a *Completely Matrix* (CP). Completely positive matrices have arisen in some situations in economic modelling and appear to have some applications in statistics, and they are also appear in quadratic optimisation, for more details see [3]. Any real matrix with nonnegative minors are called *Totally Non-Negative* (TN) matrix. If all minors are strictly positive then A is called *Totally Positive* (TP). This topic appears in the spectral properties of kernels of ordinary differential equations whose Green's function is totally positive (studied by M. G. Krein and some colleagues in the mid-1930s) [7–9]. In this presentation we give a detailed picture of all kinds of positivity mentioned above.

Keywords: Positivity, Total Positivity, Complete Positivity

Mathematics Subject Classification [2010]: 15A18, 15A23

References

[1] R. Bhatia, *Positive definite Matrices*, Princeton University Press 2007.

[2] Abraham Berman, Robert J. Plemmons, Nonnegative Matrices in the Mathematical Sciences, 1994, SIAM.

*Speaker. Email address: kazemghanbari@math.carleton.ca

[3] Abraham Berman, Completely Positive Matrices, World Scientific Publishing (2003)

[4] A. Graham, Nonnegative Matrices and Applicable Topics in Linear Algebra, John Wiley and Sons, New York, 1987.

[5] R. A. Horn and C.R. Johnson, Matrix Analysis, Cambridge University Press, 1990

[6] Seneta, E. Nonnegative matrices and Markov chains. 2nd rev. ed., 1981, XVI, 288 p., Softcover Springer Series in Statistics. (Originally published by Allen and Unwin Ltd., London, 1973)

[7] George M. Total Positivity, Interpolation and Approximation by Polynomials, Springer (2003)

[8] Graham Gladwell, Inverse problems in vibration, Kluwer Academic Publishers, (2004)

[9] Shaun Fallat and Charles R. Johnson, Totally Nonnegative Matrices, Princeton University Press (2011)

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Spectra and pseuduospectra of matrices[1]

Abbas Salemi*

Department of Applied Mathematics, Shahid Bahonar University of Kerman, Iran.

Abstract

Given $A \in M_n(\mathbb{C})$ and $\varepsilon > 0$ be given. The ε-pseudospectrum of A is defined to be the set

$$\Lambda_\varepsilon(A) := \{z \in \mathbb{C} : \|(zI - A)^{-1}\| \geq \varepsilon^{-1}\}.$$

The concept of pseudospectra has its roots in the study of the behavior of non-normal matrices and their spectra. One of the early works on pseudospectra was done in 1967 by Jim Varah in his Stanford PhD thesis. The idea was further developed by other researchers, among them Lloyd N. Trefethen and Mark Embree who published a book in 2005 entitled *Spectra and Pseudospectra*.

In this lecture we review recent results in spectra and ε-pseudospectra of matrices. Also, we study the shapes and behavior of the connected components of ε-pseudospectra for special kinds of matrices. Moreover, growth rate of ε-pseudospectra is considered.

Keywords: spectrum, pseudospectra, connected component, convergence rate
Mathematics Subject Classification [2010]: 15A18, 15A60, 65F15

References

[1] J.V. Burke, A.S. Lewis, and M.L. Overton, *Spectral conditioning and pseudospectral growth;* Numer. Math., 2007.

[2] A. Greenbaum, R.C. Li, and M.L. Overton *First-Order Perturbation Theory for Eigenvalues and Eigenvectors;* SIAM Review, 2020.

[3] A. Greenbaum, L.N. Trefethen, *Do the Pseudospectra of a matrix determine its behaivior?* Cornell University; 1993.

[4] L.N. Trefethen, M. Embree, *Spectra and Pseudospectra;* Princeton University Press, Princeton, 2005.

[1] This is a joint work with Anne Greenbaum and Faranges Kyanfar

*Speaker. Email address: salemi@uk.ac.ir

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

A two-dimensional minimum residual technique for accelerating two-step iterative solvers[1]

Fatemeh Panjeh Ali Beik[1,*], Michele Benzi[2] and Mehdi Najafi-Kalyani[1]

[1]Department of Mathematics, Vali-e-Asr University of Rafsanjan,
P.O. Box 518, Rafsanjan, Iran

[2]Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126, Pisa, Italy

Abstract

In this talk, we present a technique to speed up the convergence of a class of two-step iterative methods for solving linear systems of equations. To implement the acceleration technique, the residual norm associated with computed approximations for each sub-iterate is minimized over a certain two-dimensional subspace. Convergence properties of the resulting method will be discussed in detail. It will be further shown that the approach can be developed to solve (regularized) normal equations arising from the discretization of ill-posed problems. Numerical experiments will be disclosed to illustrate the performance of exact and inexact variants of the method for some test problems.

Keywords: Iterative methods, minimum residual technique, convergence, normal equations, ill-posed problems

Mathematics Subject Classification [2010]: 65F10

References

[1] F. P. A. Beik, M. Benzi, M. Najafi-Kalyani, A two-dimensional minimum residual technique for accelerating two-step iterative solvers with applications to discrete ill-posed problems, Preprint, https://arxiv.org/abs/2303.12473.

[1]The presented results in this talk are the summary of authors' recently submitted manuscript which is accessible in Arxiv, see [1].

*Speaker. Email address: f.beik@vru.ac.ir

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Fourier Like Systems, Frame of Translates and their Oblique Duals on LCA-groups

R. A. Kamyabi Gol*

Department of Mathematics, Ferdowsi University of Mashhad , Mashhad, Iran

Abstract

The theory of frames of translates has an essential role in many areas of mathematics and its applications such as wavelet theory and reconstruction of signals from sample values [1 − 4, 6, 11, 12, 13]. A lattice system of translates is a sequence in $L^2(\mathbb{R})$ that has the form $\mathcal{T}(g) = \{g(. - ak)\}_{k\in\mathbb{Z}}$ where $g \in L^2(\mathbb{R})$ and $a > 0$ are fixed. In the setting of $L^2(\mathbb{R})$, it is known that frames of translates can be characterized in terms of a 1-periodic function ([3, 6]). More precisely, for $g \in L^2(\mathbb{R})$, if we define $\Phi_g(\omega) = \sum_{k\in\mathbb{Z}} |\widehat{\phi}(\omega+k)|^2$, then Φ_g is a 1-periodic function which characterizes frames of translates as follows.

(a) $\mathcal{T}(g)$ is a frame sequence if and only if there exist $0 < A \leq B < \infty$ such that $A \leq \Phi_g \leq B$, a.e. on the zero set of Φ_g.

(b) $\mathcal{T}(g)$ is a Riesz basis for the closure span of $\mathcal{T}(g)$ if and only if there exist $0 < A \leq B < \infty$ such that $A \leq \Phi_g \leq B$, a.e.

(c) $\mathcal{T}(g)$ is an orthonormal basis for the closure span of $\mathcal{T}(g)$ if and only if $\Phi_g = 1$ a.e.

Our goal in this presentation is a generalization of frames of translates in the setting of locally compact abelian groups. Let G be a locally compact abelian (LCA) group and Γ be a uniform lattice in G (i.e. a discrete subgroup of G which is co-compact), with the annihilator Γ^* in $\widehat{G}$ (the dual group of G)[5, 7, 8, 10, 14 − 16]. For $g \in L^2(G)$, a system of translates generated by g via Γ, is defined as

$$\mathcal{T}(g) = \{g(. + \gamma)\}_{\gamma\in\Gamma}$$

We define a Γ^*-periodic function Φ_g on $\widehat{\Gamma}$ and investigate a characterization of translates of $g \in L^2(G)$ to have some properties. We achieve our goal by using an isometry from $L^2(G)$ into $L^2(\widehat{\Gamma})$, in such a way that the system of translates in $L^2(G)$ is transferred to a nice Fourier-like system in $L^2(\widehat{\Gamma})$. To do so, we consider a fix $\varphi \in L^2(\widehat{\Gamma})$ and define the Fourier-like system generated by φ as $\mathcal{E}(\varphi) = \{X_\gamma\varphi\}_{\gamma\in\Gamma}$, where X_γ is the corresponding character γ on $\widehat{\Gamma}$. We deduce the structure of the canonical dual frame of a frame sequence $\mathcal{T}(g)$. Using the fact that the frame operator of a frame of translates commutes with the translation operator, it is shown that the canonical dual frame of $\mathcal{T}(g)$ has the same form $\mathcal{T}(h)$ for some $h \in \overline{\text{span}}(\mathcal{T}(g))$. Some properties of Φ_g which are useful in the study of the translates sequence generated by g are

*Speaker. Email address: kamyabi@um.ac.ir

investigated. In particular, it is shown that if Φ_g is continuous, then $\mathcal{T}(g)$ can not be a redundant frame.

Keywords: locally compact abelian group, Fourier-like system, Fourier-like frame, frame of translates, oblique dual.

Mathematics Subject Classification [2010]: Primary 47A15 ; Secondary 42B99, 22B99.

References

[1] C. Cabrelli, C. Mosquera and V. Patemostro, Linear combinations of frame generators in systems of translates, *J. Math. Anal. Appl.*, 413 (2014), 776–788.

[2] P. Casazza, O. Christensen and N.J. Kalton, Frames of translates, *Collect. Math.*, (2001), 35–54.

[3] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhauser, 2015.

[4] C. Demeter, Linear independence of time frequency translates for special configurations, *Math. Res. Lett.*, 17 (2010), 761–779.

[5] G.B. Folland, *A Course in Abstract Harmonic Analysis*, CRS Press, 1995.

[6] C. Heil, *A Basis Theory Primer*, Birkhauser, 2011.

[7] E. Hewitt and K.A. Ross, *Abstract Harmonic Analysis*, vol. 1, Springer-Verlag, 1963.

[8] R.A. Kamyabi Gol and R. Raisi Tousi, Bracket products on locally compact abelian groups, *J. Sci. Islam. Repub. Iran*, 19 (2008), No. 2, 153–157.

[9] H.O. Kim and J.K. Lim, New characterizations of Riesz bases, *Appl. Comput. Harmon. Anal.*, 4 (1997), 222–229.

[10] G. Kutyniok, Time frequency analysis on locally compact groups, Ph.D thesis, Padeborn University, 2000.

[11] M. Nielsen and H. Sikie, Schauder bases of integer translates, *Appl. Comput. Harmon. Anal.*, 23 (2007), 259–262.

[12] A. Olevskii and A. Ulanavskii, Almost linear translates. Do nice generators exist?, *J. Foutier Anal. Appl.*, 10 (2004), 93–104.

[13] T.E. Olson and R.A. Zalik, *Nonexistence of a Riesz basis of translates, in: Approximation Theory*, Lecture Notes in Pure and Applied Math. Vol. 138, Dekker, New York, 1992, 401–408.

[14] R. Raisi Tousi, Shift invariant spaces, MRA and bracket products on LCA groups, Ph.D. thesis, Ferdowsi University of Mashhad, 2008.

[15] A. Safapour and R.A. Kamyabi Gol, A necessary condition for Weil-Heisenberg frames, *Bull. Iranian Math. Soc.*, 2 (2004), 67–79.

[16] N. Seyedi and R.A. Kamyabi Gol , On the frames of translates on locally compact abelian groups, *Bull. Iranian Math. Soc.*, (2021), 1–22.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Stationary Graph Signals

Arash Amini* and Mohammad-Bagher Iraji

Electrical Engineering Department, Sharif University of Technology, Tehran, Iran

Abstract

While conventional discrete signals are represented over grids, we currently deal with a number of signal types for which no well-defined grid is applicable; data related to social networks is among the examples. An alternative way for representing such signals is to assume a graph, where each node plays the role of a grid point. In other words, each node contains a part of the whole signal and based on the connections in the graph, these parts could be thought of as related to each other. In contrast to the conventional 1D discrete signals where each signal tap is adjacent to its predecessor and successor taps, in graph signals, the adjacency of a signal tap is not necessarily limited to 2 other taps. Obviously, there are more degrees of freedom in graph signals, which makes them a more versatile modeling platform. The downside is that the processing techniques which are well-studied for decades for conventional discrete signals shall be revisited and redefined. As we will see in this talk, some of the equivalent processes and definitions in the graph signal domain are quite non-trivial.

Keywords: Discrete signals, Fourier transform, frequency domain, and graphs.
Mathematics Subject Classification [2010]: 15A06, 15A18, 15A20.

1 Introduction

For a graph signal, we first need a graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{v_1, v_2 \dots, v_n\}$ denotes the set of vertices and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ stands for the set of edges. In general, the edges could be directed, which implies that for some $v_i, v_j \in \mathcal{V}$ we might have $(v_i, v_j) \in \mathcal{E}$ but $(v_j, v_i) \notin \mathcal{E}$. Besides, the graph is equipped with a weight matrix $\mathbf{W}_{n\times n}$, which is very similar to the adjacency matrix except that the non-zero elements are not necessarily 1:

$$i, j \in \{1, 2 \dots, n\}: \quad (\mathbf{W})_{i,j} = \begin{cases} w_{i,j} \geq 0, & (v_i, v_j) \in \mathcal{E}, \\ 0, & (v_i, v_j) \notin \mathcal{E}. \end{cases}$$

The weight matrix actually encodes the connectivity of the graph. Edges with larger weights are assumed to be more strongly connected.

Now that we introduced the graph, we need to define the graph signal $\mathbf{x}_{n\times 1}$, which takes a scalar value x_i at each vertex v_i. Although we express the signal graph using a vector, the order of elements in this vector depend on the labeling of the graph vertices. As this labeling does not affect the structure and properties of the graph, we need to

*Speaker. Email address: aamini@sharif.edu

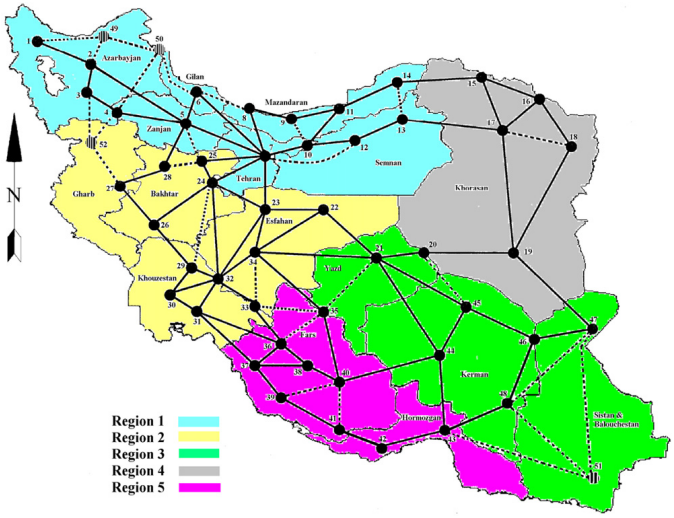


Figure 1: Iran's power distribution grid for 400KV lines [1, 2].

devise signal processing tools and techniques that commute with reordering of the graph signal.

To better illustrate the concept of graph signals, let us consider Iran's power distribution grid for 400KV lines in Figure 1. Each node in this graph is either connected to a power generator, or a cluster of power consumers (after reducing the voltage in a number of steps). The connection between the nodes correspond to the existence of physical power lines in between. Here, the edge weights could be defined as the inverse of the effective total resistance between the nodes. In simple words, nodes that are physically far apart are expected to have small edge weights; similarly, nodes that are not connected can be interpreted as being connected by an infinite-length power line.

A simple example of a graph signal here is the voltage, i.e., at vertex v_i the value x_i of the graph signal is given as the measured voltage of the substation at a given time instance. Ideally, this value should be 400KV equally at all substations; however, due to line losses, we observe voltage drops. Based on the definition of the graph and the electrical properties, the graph signal values at neighboring vertices that are connected with strong edges shall be almost equal. This suggests that the graph signal is almost smooth on the graph; in other words, when we move along the edges, signal values do not change drastically. This behavior resembles the lowpass nature of conventional graph signals.

In this talk, upon simple graph operators such as the *shift*, we build graph Fourier transform (GFT) and interpret classical concepts such as the frequency domain, lowpass property and etc based on GFT. Next, we consider stochastic signals over graphs and investigate the notion of stationarity. The basics of graph signal processing reviewed in this talk are taken from [3] and [4].

Acknowledgment

I greatly thank the technical committee of the 12th seminar on linear algebra and its applications (SLAA2023) for inviting me.

References

[1] P. Maghouli, S.H. Hosseini, M.O. Buygi and M. Shahidehpour, A scenario-based multi-objective model for multi-stage transmission expansion planning, *IEEE Transactions on Power Systems*, 36 (2011), No. 1, 470–-478.

[2] M. Hasani-Marzooni and S.H. Hosseini, ADynamic analysis of various investment incentives and regional capacity assignment in Iranian electricity market, *Energy Policy*, 56 (2013), 271-284.

[3] L. Stanković and E. Sejdić, *Vertex-frequency analysis of graph signals*, Springer, 2019.

[4] A. Ortega, *Introduction to graph signal processing*, Cambridge University Press, 2022.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Matrix and tensor modeling in Artificial intelligence and data science

Mansoor Rezghi*

Department of computer science, Tarbiat Modares University, Iran

Abstract

Physics and engineering have been the primary sources of problems in matrix computations for several decades. However, in recent years, significant progress in artificial intelligence and data analysis has given rise to challenging problems that require efficient matrix techniques. Additionally, these fields contain vast data with multi-dimensional structures, for which tensors serve as the appropriate structure. In this lecture, we intend to discuss the main approaches and concepts in the field of utilizing matrix and tensor modeling in artificial intelligence and data science.

Keywords: Matrix computation, Tensor decomposition, Artificial Intelligence, Data science

Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36

References

[1] M. Rezghi, A novel fast tensor-based preconditioner for image restoration, **IEEE Transactions on Image Processing**, 9 (2017), 4499- 4508.

[2] M. Rezghi, L. Elden, Diagonalization of tensors with circulant structure, **Linear Algebra and its Applications**, 435 (2011), 422-447.

[3] M.Rezghi, S. M. Hosseini, L. Elden, Best Kronecker product approximation of the blurring operator in three dimensional image restoration problems, **SIAM Journal on Matrix Analysis and Applications**, 35 (2014), 1086-1104.

[4] S Ahmadi, M Rezghi, Generalized low-rank approximation of matrices based on multiple transformation pairs, **Pattern Recognition**, 108(2020), 107545, 1–16.

[5] T. Saeedi, M. Rezghi, A Novel Enriched Version of Truncated Nuclear Norm Regularization for Matrix Completion of Inexact Observed Data, **IEEE Transactions ON Knowledge and Data Engineering**, 34 (2020) , 519–530.

[6] A. Noroozi, M. Rezghi, A Tensor-Based Framework for rs-fMRI Classification and Functional Connectivity Construction, **Frontiers in Neuroinformatics** 14(2020), 1–13.

*Speaker. Email address: Rezghi@modares.ac.ir

[7] P. Parvaside, M. Rezghi, A novel dictionary learning method based on total least squares approach with application in high dimensional biological data, **Advances in Data Analysis and Classification** , 15 (2021), 575–597.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Numerical radius: New Extensions and Inequalities

Ali Zamani*

Department of Mathematics, Farhangian University, Tehran, Iran

Abstract

We firstly define a seminorm on the space of bounded linear operators on a Hilbert space, which generalizes the numerical radius norm. We investigate basic properties of this seminorm and prove inequalities involving it. Further, for a positive element a in a unital C^*-algebra $\mathfrak{A}$ we define a semi-norm on $\mathfrak{A}$, which generalizes the a-operator semi-norm and the a-numerical radius.

Keywords: Numerical range, numerical radius, inequality.
Mathematics Subject Classification [2010]: 47A12, 47A30, 46L05.

1 Introduction and preliminaries

Let $\mathfrak{A}$ be a C^*-algebra with unit denoted by $\mathbf{1}$ and let $a \in \mathfrak{A}$ be a positive element. Let $\mathbb{B}(\mathcal{H})$ be the C^*-algebra of all bounded linear operators on a complex Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$. By a state on $\mathfrak{A}$ we mean a positive linear functional f on $\mathfrak{A}$ such that $\|f\| = 1$ and let $\mathcal{S}(\mathfrak{A})$ denote the set of states on $\mathfrak{A}$. For an element $x \in \mathfrak{A}$, let $V(x)$ denote the (algebraic) numerical range of $x \in \mathfrak{A}$, that is, the set $V(x) = \{f(x) : f \in \mathcal{S}(\mathfrak{A})\}$. This set generalizes the classical numerical range in the sense that the numerical range $V(T)$ of a Hilbert space operator T (considered as an element of a C^*-algebra $\mathbb{B}(\mathcal{H})$) coincides with the closure of its classical numerical range $W(T) = \{\langle T\xi, \xi\rangle : \xi \in \mathcal{H}, \|\xi\| = 1\}$. The numerical radius of $x \in \mathfrak{A}$ is defined as $v(x) = \sup\{|\lambda| : \lambda \in V(x)\}$.

Recently, Bourhim and Mabrouk in [3] introduced and studied a-numerical range and a-numerical radius of elements in C^*-algebras. Also, the authors in [1] continued the work on the a-numerical range and the a-numerical radius. In particular, some ideas from the recent papers are extended.

Set $\mathcal{S}_a(\mathfrak{A}) = \left\{\frac{f}{f(a)} : f \in \mathcal{S}(\mathfrak{A}),\ f(a) \neq 0\right\}$ and for an element $x \in \mathfrak{A}$, let $\|x\|_a = \sup\left\{\sqrt{f(x^*ax)} : f \in \mathcal{S}_a(\mathfrak{A})\right\}$. It is worth observing that $\|\cdot\|_{\mathbf{1}} = \|\cdot\|$ and $\|x\|_a = 0$ if and only if $ax = 0$. Further the set $\mathcal{S}_a(\mathfrak{A})$ is a non empty, convex and closed subset of the topological dual space of $\mathfrak{A}$, but it is compact if and only if a is invertible in $\mathfrak{A}$; see [3, proposition 2.3]. In particular if a is not invertible and due to the lack of compactness of $\mathcal{S}_a(\mathfrak{A})$, it may happen that $\|x\|_a = \infty$ for some $x \in \mathfrak{A}$; see [3, Example 3.2]. In the sequel we will denote $\mathfrak{A}^a = \{x \in \mathfrak{A} : \|x\|_a < \infty\}$. The set $\mathfrak{A}^a$ is a subalgebra of $\mathfrak{A}$ not necessarily closed. Also by [3, Proposition 3.3], $\|\cdot\|_a$ is a semi-norm on $\mathfrak{A}^a$ and satisfies

*Speaker. Email address: zamani.ali85@yahoo.com

$\|xy\|_a \leq \|x\|_a \|y\|_a$ for all $x, y \in \mathfrak{A}^a$. Denote by $\mathfrak{A}_a$ the set of all elements in $\mathfrak{A}$ that admit a-adjoints. Recall that for an element $x \in \mathfrak{A}$, an element $x^{\sharp_a} \in \mathfrak{A}$ is said to be an a-adjoint of x if $ax^{\sharp_a} = x^*a$. Basic properties of $\mathfrak{A}_a$ were investigated in [3]. In particular, $\mathfrak{A}_a$ is a subalgebra of $\mathfrak{A}^a$ which is neither closed nor dense in $\mathfrak{A}$. Further if $x \in \mathfrak{A}_a$ and $x^{\sharp_a}$ is an a-adjoint of it, then by

$$\|x\|_a^2 = \left\|xx^{\sharp_a}\right\|_a = \left\|x^{\sharp_a}x\right\|_a = \left\|x^{\sharp_a}\right\|_a^2. \tag{1}$$

An element $x \in \mathfrak{A}$ is said to be a-self-adjoint if ax is self-adjoint, i.e., $ax = x^*a$. We say that x is a-positive if ax is positive. Every element x in $\mathfrak{A}_a$ can be written as $x = y+iz$ where y and z are a-self-adjoint but, in general, this decomposition is not unique. In fact if $x^{\sharp_a}$ is an a-adjoint of x, then $x = \Re(x)+i\Im(x)$, where $\Re(x) = \frac{x+x^{\sharp_a}}{2}$ and $\Im(x) = \frac{x-x^{\sharp_a}}{2i}$ are a-real and a-imaginary parts of x, respectively. The a-numerical range (respectively, the a-numerical radius) of an element $x \in \mathfrak{A}$ are defined by $V_a(x) = \{f(ax) : f \in \mathcal{S}_a(\mathfrak{A})\}$ (respectively, $v_a(x) = \sup\{|\lambda| : \lambda \in V_a(x)\}$). In contrast of the classical algebraic numerical range, the a-numerical range $V_a(x)$ of $x \in \mathfrak{A}$ may be unbounded. Note that these concepts were introduced in [3] as generalizations of the A-numerical range (respectively, the A-numerical radius) for Hilbert space operator T given by $W_A(T) = \{\langle AT\xi, \xi\rangle : \xi \in \mathcal{H}, \|\xi\|_A = 1\}$ (respectively, $w_A(T) = \sup\{|\lambda| : \lambda \in W_A(T)\}$), where A is a positive operator on $\mathcal{H}$ and $\|\xi\|_A = \sqrt{\langle A\xi, \xi\rangle}$ for all $\xi \in \mathcal{H}$. In particular, when A is the identity operator on $\mathcal{H}$, then A-numerical range and A-numerical radius of T coincide with the classical numerical range and numerical radius, respectively, i.e., $W_A(T) = W(T)$ and $w_A(T) = w(T)$.

An important and useful identity for the a-numerical radius (see [3, Theorem 4.11]) is as follows:

$$v_a(x) = \sup_{\theta\in\mathbb{R}} \left\|\Re(e^{i\theta}x)\right\|_a.$$

By [3, Proposition 3.3 and Corollary 4.10], observe that $v_a(\cdot)$ defines a semi-norm on $\mathfrak{A}_a$, which is equivalent to the a-operator semi-norm $\|\cdot\|_a$. Namely, for $x \in \mathfrak{A}_a$, it holds that

$$\frac{1}{2}\|x\|_a \leq v_a(x) \leq \|x\|_a. \tag{2}$$

The first inequality becomes equality if $ax \neq 0$ and $ax^2 = 0$ and the second inequality becomes equality if x is a-self-adjoint (see, [3, Corollary 4.6]).

2 A generalization of the numerical radius for Hilbert space operators

The notion of orthogonality in an arbitrary normed linear space may be introduced in various ways. Among them, the one which is frequently studied in literature is the *Birkhoff–James orthogonality* [2, 4]: if x, y are elements of a normed linear space E equipped with the norm $N(\cdot)$, then x is orthogonal to y in the Birkhoff–James sense, in short $x \perp_B^N y$, if

$$N(x + \lambda y) \geq N(x), \quad \forall \lambda \in \mathbb{C}.$$

Moreover, $\|\cdot\|_N^*\colon E^* \longrightarrow [0, +\infty)$ stands for the dual norm, i.e. $\|\cdot\|_N^*$ is a norm in $E^* = \big(E, N(\cdot)\big)^*$. For fixed $x \in E$ let $J_N(x)$ denote the set of its supporting functionals:

$$J_N(x) := \Big\{f \in E^* : \|f\|_N^* = 1,\ f(x) = N(x)\Big\}.$$

The Hahn-Banach theorem implies that $J_N(x) \neq \emptyset$. Recall that a unit vector point $u \in E$ is called a *vertex* of the closed unit ball in E if $J_N(u)$ is total over E.

Now, let $N(\cdot)$ be an arbitrary norm on $\mathbb{B}(\mathcal{H})$. According to the beginning of this section, for fixed $T \in \mathbb{B}(\mathcal{H})$ we have

$$J_N(T) = \Big\{ f \in \mathbb{B}(\mathcal{H})^* : \|f\|_N^* = 1,\ f(T) = N(T) \Big\}.$$

Since Birkhoff -James orthogonality has the property of right existence, we obtain $\big\{\xi \in \mathbb{C} : I \perp_B^N (T - \xi I)\big\} \neq \emptyset$. Let $I \perp_B^N (T - \xi I)$ for some $\xi \in \mathbb{C} \setminus \{0\}$. Hence $N\big(I + \frac{1}{\xi}(T - \xi I)\big) \geq N(I)$ and so $|\xi| \leq \frac{N(T)}{N(I)}$. Thus the set $\big\{\xi \in \mathbb{C} : I \perp_B^N (T - \xi I)\big\}$ is also bounded in $\mathbb{C}$. This motivates the following definition (see [9]).

Definition 2.1. Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$. The function $w_N \colon \mathbb{B}(\mathcal{H}) \to [0, +\infty)$ is defined as

$$w_N(T) := \sup \Big\{ |\xi| : \xi \in \mathbb{C},\ I \perp_B^N (T - \xi I) \Big\}$$

for every $T \in \mathbb{B}(\mathcal{H})$.

Remark 2.2. Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$ and let $T \in \mathbb{B}(\mathcal{H})$. For every $\xi \in \mathbb{C}$, we have

$$\begin{aligned} I \perp_B^N (T - \xi I) &\iff N\big(I + \lambda(T - \xi I)\big) \geq N(I) \quad \forall \lambda \in \mathbb{C} \\ &\iff N\big(I + \frac{1}{\xi - \lambda}(T - \xi I)\big) \geq N(I) \quad \forall \lambda \in \mathbb{C} \setminus \{\xi\} \\ &\iff N\big((\xi - \lambda)I + T - \xi I\big) \geq |\xi - \lambda| N(I) \quad \forall \lambda \in \mathbb{C} \\ &\iff N\big(T - \lambda I\big) \geq |\xi - \lambda| N(I) \quad \forall \lambda \in \mathbb{C}. \end{aligned}$$

Thus

$$I \perp_B^N (T - \xi I) \iff N\big(T - \lambda I\big) \geq |\xi - \lambda| N(I) \quad \forall \lambda \in \mathbb{C}. \tag{3}$$

Remark 2.3. For any $T \in \mathbb{B}(\mathcal{H})$, it is well-known (see [7]) that

$$\overline{W(T)} = \bigcap_{\lambda \in \mathbb{C}} \big\{ \xi : \|T - \lambda I\| \geq |\xi - \lambda| \big\}.$$

Therefore, by (3), we have

$$w(T) = \sup \Big\{ |\xi| : \xi \in \mathbb{C},\ I \perp_B^{\|\cdot\|} (T - \xi I) \Big\}.$$

In view of the previous relation, it is now obvious that $w_N(\cdot)$ generalizes the classical numerical radius $w(\cdot)$.

Proposition 2.4. *Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$ and let $T \in \mathbb{B}(\mathcal{H})$. Then the following properties hold:*

(i) $w_N(I) = 1$.

(ii) $w_N(T) \leq \frac{N(T)}{N(I)}$.

(iii) If $N(\cdot)$ is self-adjoint, then so is $w_N(\cdot)$.

(iv) If $N(\cdot)$ is weakly unitarily invariant, then so is $w_N(\cdot)$.

Theorem 2.5. *Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$. Then $w_N(\cdot)$ is a seminorm on $\mathbb{B}(\mathcal{H})$.*

Remark 2.6. Let $N(\cdot)$ be an arbitrary norm on $\mathbb{B}(\mathcal{H})$. By Theorem 2.5, $w_N(\cdot)$ is a seminorm on $\mathbb{B}(\mathcal{H})$. Therefore, for $T \in \mathbb{B}(\mathcal{H})$, if $T = 0$, then $w_N(T) = 0$. The converse is however not true, in general (see Theorem 2.7).

From now on we assume that the considered norm $N\colon \mathbb{B}(\mathcal{H}) \longrightarrow [0, +\infty)$ satisfies $N(I) = 1$. There is no loss in generality in assuming this. In particular, the classical norms on $\mathbb{B}(\mathcal{H})$ satisfy such equality for the identity operator I. Therefore, we think that such assumption is interesting for investigations.

Now, we are going to prove a condition for checking when $w_N(\cdot)$ is a norm on $\mathbb{B}(\mathcal{H})$.

Theorem 2.7. *Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$ with $N(I) = 1$. The following conditions are equivalent:*

(i) $w_N(\cdot)$ is a norm on $\mathbb{B}(\mathcal{H})$.

(ii) The operator I is a vertex of the closed unit ball in $\big(\mathbb{B}(\mathcal{H}), N(\cdot)\big)$.

The following result says that the spaces $\big(\mathbb{B}(\mathcal{H}), N(\cdot)\big)$ and $\big(\mathbb{B}(\mathcal{H}), w_N(\cdot)\big)$ are similar (in some sense) in the point I.

Theorem 2.8. *Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$ with $N(I) = 1$. If $w_N(\cdot)$ is a norm on $\mathbb{B}(\mathcal{H})$, then*

$$J_N(I) = J_{w_N}(I). \tag{4}$$

In this case, the operator I is a vertex of the closed unit ball in $\big(\mathbb{B}(\mathcal{H}), w_N(\cdot)\big)$.

Now we may consider the function $w_{w_N}\colon \mathbb{B}(\mathcal{H}) \longrightarrow [0, +\infty)$. Suppose that $w_N(\cdot)$ is a norm on $\mathbb{B}(\mathcal{H})$. It follows from Proposition 2.4(i)-(ii) that $w_{w_N}(\cdot) \leq w_N(\cdot)$. Moreover, Theorem 2.5 yields the subadditivity of $w_{w_N}(\cdot)$. It is amazing that these remarks can be strengthen as follows.

Theorem 2.9. *Let $N(\cdot)$ be a norm on $\mathbb{B}(\mathcal{H})$ with $N(I) = 1$. If $w_N(\cdot)$ is a norm on $\mathbb{B}(\mathcal{H})$, then $w_{w_N}(\cdot)$ is also a norm on $\mathbb{B}(\mathcal{H})$. Moreover, $w_{w_N}(\cdot) = w_N(\cdot)$.*

Our next result reads as follows.

Theorem 2.10. *Let $N(\cdot)$ is a weakly unitarily invariant norm on $\mathbb{B}(\mathcal{H})$ and let T and S be self-adjoint operator in the norm-unit ball of $\mathbb{B}(\mathcal{H})$. Then*

$$w_N(TS \pm ST) \leq \sup_{U \in \mathcal{U}} \big\{ w_N(TU \pm U^*T), w_N(SU \pm U^*S) \big\},$$

where $\mathcal{U}$ is the unitary group of all unitary operators in $\mathbb{B}(\mathcal{H})$.

As a consequence of Theorem 2.10, we have the following result.

Corollary 2.11. *Let $N(\cdot)$ is a weakly unitarily invariant norm on $\mathbb{B}(\mathcal{H})$ and let T be an operator in the norm-unit ball of $\mathbb{B}(\mathcal{H})$. Then*

$$w_N(TT^* - T^*T) \leq 2 \sup_{U \in \mathcal{U}} \Big\{ w_N\big(\Re(T)U \pm U^*\Re(T)\big), w_N\big(\Im(T)U \pm U^*\Im(T)\big) \Big\},$$

where $\mathcal{U}$ is the unitary group of all unitary operators in $\mathbb{B}(\mathcal{H})$.

3 An extension of the a-numerical radius on C^*-algebras

First, let us define notions weighted a-real and a-imaginary parts of elements in $\mathfrak{A}_a$. Let s and t be two nonnegative reals such that $s+t>0$. Define the weighted a-real and a-imaginary parts of $x\in\mathfrak{A}_a$ by $\Re_{(s,t)}(x)=sx+tx^{\sharp_a}$ and $\Im_{(s,t)}(x)=s(-ix)+t(-ix)^{\sharp_a}$, respectively. When $s=t=\frac{1}{2}$, we clearly have $\Re_{(\frac{1}{2},\frac{1}{2})}(x)=\Re(x)$ and $\Im_{(\frac{1}{2},\frac{1}{2})}(x)=\Im(x)$. Also define the function $v_{(a,(s,t))}(\cdot)\colon \mathfrak{A}_a\longrightarrow[0,+\infty)$ by

$$v_{(a,(s,t))}(x)=\sup_{\theta\in\mathbb{R}}\left\|\Re_{(s,t)}(e^{i\theta}x)\right\|_a. \tag{5}$$

Remark 3.1. For $x\in\mathfrak{A}_a$, it is easy to see that $v_{(a,(s,t))}(x)=\sup_{\theta\in\mathbb{R}}\left\|\Im_{(s,t)}(e^{i\theta}x)\right\|_a$.

Remark 3.2. Obviously, $v_{(a,(1,0))}(x)=v_{(a,(0,1))}(x)=\|x\|_a$, and $v_{(a,(\frac{1}{2},\frac{1}{2}))}(x)=v_a(x)$. Hence $v_{(a,(s,t))}(\cdot)$ generalizes the a-operator semi-norm $\|\cdot\|_a$ and the a-numerical radius $v_a(\cdot)$, which have been introduced in [3].

Remark 3.3. Let $\mathfrak{A}=\mathbb{B}(\mathcal{H})$ and let $0\leq\nu\leq 1$. We have

$$v_{(I,(\nu,1-\nu))}(T)=\sup_{\theta\in\mathbb{R}}\left\|\nu e^{i\theta}T+(1-\nu)(e^{i\theta}T)^*\right\|:=w_\nu(T).$$

Thus $v_{(a,(s,t))}(\cdot)$ also generalizes the weighted numerical radius $w_\nu(\cdot)$, which has been recently introduced in [6] (see also [8]).

Our first result reads as follows.

Theorem 3.4. *Let $x\in\mathfrak{A}_a$. The following statements hold.*

(i) $v_{(a,(s,t))}(x)=\sup_{\alpha,\beta\in\mathbb{R},\alpha^2+\beta^2=1}\left\|\alpha\,\Re_{(s,t)}(x)+\beta\,\Im_{(s,t)}(x)\right\|_a$.

(ii) $v_{(a,(s,t))}(x)=\frac{1}{2}\sup_{\theta,\varphi\in\mathbb{R}}\left\|\Re_{(s,t)}\left((e^{i\theta}-ie^{i\varphi})x\right)\right\|_a$.

The next result establishes that $v_{(a,(s,t))}(\cdot)$ and $\|\cdot\|_a$ are two equivalent semi-norm on $\mathfrak{A}_a$.

Theorem 3.5. *$v_{(a,(s,t))}(\cdot)$ is a semi-norm on $\mathfrak{A}_a$ and for every $x\in\mathfrak{A}_a$ the following inequalities hold:*

$$\max\{s,t\}\|x\|_a\leq v_{(a,(s,t))}(x)\leq(s+t)\|x\|_a. \tag{6}$$

Remark 3.6. For $x\in\mathfrak{A}_a$, by (1), we have

$$\begin{aligned}v_{(a,(s,t))}(x^{\sharp_a})&=\sup_{\theta\in\mathbb{R}}\left\|se^{i\theta}x^{\sharp_a}+te^{-i\theta}(x^{\sharp_a})^{\sharp_a}\right\|_a\\&=\sup_{\theta\in\mathbb{R}}\left\|\left(se^{-i\theta}x+te^{i\theta}x^{\sharp_a}\right)^{\sharp_a}\right\|_a\\&=\sup_{\theta\in\mathbb{R}}\left\|se^{-i\theta}x+te^{i\theta}x^{\sharp_a}\right\|_a=v_{(a,(s,t))}(x),\end{aligned}$$

and hence $v_{(a,(s,t))}(x^{\sharp_a})=v_{(a,(s,t))}(x)$.

In the following result, we give a condition equivalent to $v_{(a,(s,t))}(x) = \max\{s,t\}\|x\|_a$.

Theorem 3.7. *Let $x \in \mathfrak{A}_a$. The following are equivalent:*

(i) $\left\|\Re_{(s,t)}(e^{i\theta}x)\right\|_a = \max\{s,t\}\|x\|_a$ *for all* $\theta \in \mathbb{R}$.

(ii) $v_{(a,(s,t))}(x) = \max\{s,t\}\|x\|_a$.

In the following theorem, a refinement of the inequality (6) is given.

Theorem 3.8. *Let $x \in \mathfrak{A}_a$. Then*

$$v_{(a,(s,t))}(x) \leq \sqrt{(s^2+t^2)\|x\|_a^2 + 2st\, v_a\left(x^2\right)} \leq (s+t)\|x\|_a.$$

Corollary 3.9. *If $x \in \mathfrak{A}_a$ is such that $v_{(a,(s,t))}(x) = (s+t)\|x\|_a$, then $\|x^2\|_a = \|x\|_a^2$.*

Our final result extends and refines an inequality for the numerical radius of Hilbert space operators obtained by Kittaneh in [5].

Theorem 3.10. *Let $x \in \mathfrak{A}_a$. Then*

$$st\left\|xx^{\sharp_a} + x^{\sharp_a}x\right\|_a + \frac{1}{2}\sup_{\theta\in\mathbb{R}}\left|\left\|\Re_{(s,t)}(e^{i\theta}x)\right\|_a^2 - \left\|\Im_{(s,t)}(e^{i\theta}x)\right\|_a^2\right| \leq v_{(a,(s,t))}^2(x).$$

References

[1] A. Alahmari, M. Mabrouk and A. Zamani, *Further results on the a-numerical range in C^*-algebras*, Banach J. Math. Anal. **16**, 25 (2022).

[2] G. Birkhoff, *Orthogonality in linear metric spaces*, Duke Math. J. **1** (1935), 169 -172.

[3] A. Bourhim and M. Mabrouk, *a-numerical range on C^*-algebras*, Positivity **25** (2021), 1489-1510.

[4] R. C. James, *Orthogonality in normed linear spaces*, Duke Math. J. **12** (1945), 291–302.

[5] F. Kittaneh, *Numerical radius inequalities for Hilbert space operators*, Stud. Math. **168** (2005), 73-80.

[6] A. Sheikhhosseini, M. Khosravi and M. Sababheh, *The weighted numerical radius*, Ann. Funct. Anal. **13**, (3) (2022).

[7] J. G. Stampfli and J. P. Williams, *Growth conditions and the numerical range in a Banach algebra*, Tôhoku Math. J. **20** (1968), no. 2, 417–424.

[8] A. Zamani, *The weighted Hilbert–Schmidt numerical radius*, Linear Algebra Appl. **675** (2023), 225–243.

[9] A. Zamani and P. Wójcik, *Another generalization of the numerical radius for Hilbert space operators*, Linear Algebra Appl. **609** (2021), 114–128.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Perturbation of Woven g-fusion Frames

Mehdi Rashidi-Kouchi[1,*] and Maryam Mohammadrezaee[2]

[1] Department of Mathematics, Ghaderabad Center, Safashahr Beranch, Islamic Azad University, Ghaderabad, Iran.

[2] Department of Mathematics, Kerman Branch, Islamic Azad University, Kerman, Iran

Abstract

In this paper, we show that those of g-fusion frames that are small perturbations of each other, constitute woven g-fusion frame. We start with Paley-Wiener perturbation of weaving g-fusion frames and continue two results of perturbations in the sequel.

Keywords: Frame, Weaving g-fusion frame, Perturbation.

Mathematics Subject Classification [2010]: 42C15, 42C30,42C40.

1 Introduction

Frames for a Hilbert space were first introduced by Duffin and Schaeffer [10] in 1952. Daubechies, Grossmann and Meyer [8] reintroduced frames, in 1986 and considered from then. Frame theory has applications in signal processing, image processing, data compression and sampling theory.

Orthonormal bases are special case of frames in Hilbert space. Any element in a Hilbert space can be present as an infinite linear combination, not necessary unique, of the frame elements. For more information, readers can refer to [7, 11].

Some new types and generalizations of frame were introduced by researchers such as fusion frames, g-frames, woven frames, ... Frame of subspaces or fusion frames are a generalization of frames which were introduced by Cassaza and Kutyniok [4] in 2003 and were investigated in [1, 5, 12, 13]. Generalized frames or in abbreviation g-frames were introduced by Sun [18] in 2006. Most recently, g-fusion frames in Hilbert space were introduced by Sadri et.al. [17].

In other side, weaving frames were introduced by Bemrose et.al. [2] and [6] in 2016. Weaving frames are powerful tools for pre-processing signals and distributed data processing. Many researchers studied and generalized weaving frames. Some of these generalizations are weaving g-frames, weaving fusion frames [14], Weaving K-frames [9] and controlled weaving frames [16].

In this paper, we review the concept of weaving g-fusion frame from [15]. This frame includes weaving g-frames and weaving fusion frames. Then, we study perturbations of woven g-fusion frames.

*Speaker. Email address: rashidimehdi20@gmail.com

2 Basic Definitions and Preliminaries

As a preliminary of frames, at the first, we mention fusion frames. Also we review g-frames, g-fusion frames and woven frames . Throughout this paper, $\mathbb{I}$ is the indexing set where it can be a finite or countably infinite set, and $[m]$ is the set consisting of the of natural numbers$\{1, 2, ..., m\}$. Also, $\mathcal{H}$ and $\mathcal{K}$ are separable Hilbert spaces and$B(\mathcal{H}, \mathcal{K})$ is the collection of all the bounded linear operators of $\mathcal{H}$ into $\mathcal{K}$. If $\mathcal{H} = \mathcal{K}$, then $B(\mathcal{H}, \mathcal{H})$ will be denoted by$B(\mathcal{H})$ and P is the orthogonal projection.

In 2003, a new type of generalization of frames were introduced by Cassaza and Kutyniok to the science world that today we know them as fusion frames. In this section, we briefly recall some basic notations, definitions and some important properties of fusion frames that are useful for our study. For more detailed information one can see [1, 4, 5, 12, 13].

Definition 2.1. Let $\{v_i\}_{i\in\mathbb{I}}$ be a family of real weights such that $v_i > 0$ for all $i \in \mathbb{I}$. A family of closed subspaces $\{W_i\}_{i\in\mathbb{I}}$ of a Hilbert space $\mathcal{H}$ is called a fusion frame (or frame of subspaces) for $\mathcal{H}$ with respect to weights $\{v_i\}_{i\in\mathbb{I}}$, if there exist constants $C, D > 0$ such that

$$C \|f\|^2 \leq \sum_{i\in\mathbb{I}} v_i^2 \|P_{W_i}(f)\|^2 \leq D \|f\|^2, \quad \forall f \in \mathcal{H}, \tag{1}$$

where P_{W_i} is the orthogonal projection of $\mathcal{H}$ to W_i. The constants C and D are called the lower and upper fusion frame bounds, respectively. If the right inequality in (1) holds, the family of subspace $\{W_i\}_{i\in\mathbb{I}}$ is called a Bessel sequence of subspaces with respect to $\{v_i\}_{i\in\mathbb{I}}$ with Bessel bound D. Also it is called a tight fusion frame with respect to $\{v_i\}_{i\in\mathbb{I}}$, if $C = D$ and is called parseval fusion frame, if $C = D = 1$. We say $\{W_i\}_{i\in\mathbb{I}}$ an orthogonal fusion basis for $\mathcal{H}$, if $\mathcal{H} = \bigoplus_{i\in\mathbb{I}} W_i$.

Sun [18] introduced g-frames which are generalized frames and include ordinary frames and many recent generalizations of frames.

Definition 2.2. Let $\{\mathcal{H}_i\}_{i\in\mathbb{I}}$ be a family of Hilbert spaces. We call $\Lambda = \{\Lambda_i \in B(\mathcal{H}, \mathcal{H}_i),\ i \in \mathbb{I}\}$ a g-frame for $\mathcal{H}$ with respect to $\{\mathcal{H}_i\}_{i\in\mathbb{I}}$, or simply, a g-frame for H, if there exist two positive constants C, D such that

$$C \|f\|^2 \leq \sum_{i\in\mathbb{I}} \|\Lambda_i f\|^2 \leq D \|f\|^2, \quad \forall f \in \mathcal{H}. \tag{2}$$

The positive numbers C and D are called the lower and upper g-frame bounds, respectively. We call Λ a tight g-frame, if $C = D$ and we call it a parseval g-frame, if $C = D = 1$. If only the second inequality holds, we call it g-Bessel sequence. If Λ is a g-frame, then the g-frame operator S_Λ is defined by

$$S_\Lambda f = \sum_{i\in\mathbb{I}} \Lambda_i^* \Lambda_i f, \quad f \in \mathcal{H},$$

which is a bounded, positive and invertible operator such that

$$AI \leq S_\Lambda \leq BI,$$

and for each $f \in \mathcal{H}$, we have

$$f = S_\Lambda S_\Lambda^{-1} f$$

$$
\begin{aligned}
&= S_\Lambda^{-1} S_\Lambda f \\
&= \sum_{i\in\mathbb{I}} S_\Lambda^{-1} \Lambda_i^* \Lambda_i f \\
&= \sum_{i\in\mathbb{I}} \Lambda_i^* \Lambda_i S_\Lambda^{-1} f.
\end{aligned}
$$

The canonical dual g-frame for Λ is defined by $\{\Lambda_i S_\Lambda^{-1}\}_{i\in\mathbb{I}}$ with bounds $\frac{1}{B}, \frac{1}{C}$. In other words, $\{\Lambda_i S_\Lambda^{-1}\}_{i\in\mathbb{I}}$ and $\{\Lambda_i\}_{i\in\mathbb{I}}$ are dual g-frames with respect to each other.

It is easy to show that by letting $\mathcal{H}_i = W_i$, $\Lambda_i = P_{W_i}$ and $v_i = 1$, a fusion frame is a g-frame.

Generalized fusion frames (g-fusion frames) in Hilbert space were introduced by Sadri et.al. [17].

Let

$$
\left(\sum_{i\in\mathbb{I}} \oplus \mathcal{H}_i\right)_{\ell^2} = \left\{ \{f_i\}_{i\in\mathbb{I}} | f_i \in \mathcal{H}_i \text{ and } \sum_{i\in\mathbb{I}} \|f_i\|^2 < \infty \right\},
$$

with the inner product defined by

$$
\langle \{f_i\}_{i\in\mathbb{I}}, \{g_i\}_{i\in\mathbb{I}} \rangle = \sum_{i\in\mathbb{I}} \langle f_i, g_i \rangle ,
$$

is a Hilbert space.

Definition 2.3. Let $W = \{W_i\}_{i\in\mathbb{I}}$ be a family of closed subspaces of $\mathcal{H}$, $\{v_i\}_{i\in\mathbb{I}}$ be a family of weights, i.e. $v_i > 0$ and $\Lambda_i \in B(\mathcal{H}, \mathcal{H}_i)$ for all $i \in \mathbb{I}$. We say $\Lambda := (\Lambda_i, W_i, v_i)$ is a generalized fusion frame (or g-fusion frame) for $\mathcal{H}$, if there exist $0 < A \leq B < \infty$ such that for each $f \in \mathcal{H}$

$$
A \|f\|^2 \leq \sum_{i\in\mathbb{I}} v_i^2 \|\Lambda_i P_{W_i} f\|^2 \leq B \|f\|^2 . \tag{3}
$$

We call Λ a parseval g-fusion frame, if $A = B = 1$. When the right hand of (3) holds, Λ is called a g-fusion Bessel sequence for $\mathcal{H}$ with bound B. If $\mathcal{H}_i = \mathcal{H}$ for all $i \in \mathbb{I}$ and $\Lambda_i = I_\mathcal{H}$, then we get the fusion frame (W_i, v_i) for $\mathcal{H}$. Throughout this paper, Λ will be a triple (Λ_i, W_i, v_i) with $i \in \mathbb{I}$ unless otherwise stated.

Woven frames in Hilbert spaces, were introduced in 2015 by Bemrose et.al. [2, 3, 6], after that, Vashisht, Deepshikha, and others. have done more research [9, 19–21]. They have studied a variety of different types of generalized weaving frames, such as g-frame, K-frame, and continuous frame. In the following, we mention the definition of woven frames.

Definition 2.4. Let $F = \{f_{ij}\}_{i\in\mathbb{I}}$ for $j \in [m]$ (where $[m]$ is the set $\{1, 2, \ldots, m\}$) be a family of frames for separable Hilbert space $\mathcal{H}$. If there exist universal constants A' and B' such that for every partition $\{\sigma_j\}_{j\in[m]}$, the family $F_j = \{f_{ij}\}_{i\in\sigma_j}$ is a frame for $\mathcal{H}$ with bounds A' and B', then F is said Woven frames and for every $j \in [m]$, the frames F_j are called Weaving frame.

Woven g-fusion frames extend and improve the notions of g-fusion frames and weaving frames. Woven g-fusion frames Introduced in [15].

Definition 2.5. A family of g-fusion frames $\{(\Lambda_{ij}, W_{ij}, v_{ij})\}_{i\in\mathbb{I}}$ for $j \in [m]$, is said woven g-fusion frames if there exist universal constants A and B, such that for every partition $\{\sigma_j\}_{j\in[m]}$ of $\mathbb{I}$, the family $\{(\Lambda_{ij}, W_{ij}, v_{ij})\}_{i\in\sigma_j, j\in[m]}$ is a g-fusion frame for $\mathcal{H}$ with lower and upper frame bounds A and B. Each family $\{(\Lambda_{ij}, W_{ij}, v_{ij})\}_{i\in\sigma_j, j\in[m]}$ is called a Weaving g-fusion frame.

3 Main results

In this section, shown that those of g-fusion frames that are small perturbations of each other, constitute woven g-fusion frame. We start by Paley-Wiener perturbation of weaving g-fusion frames and continue two results of perturbations in the sequel.

Theorem 3.1. *Let $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \mu_i)\}_{i\in\mathbb{I}}$ be g-fusion frames for $\mathcal{H}$ with respect to $\{\mathcal{H}_i\}_{i\in\mathbb{I}}$ with g-fusion frame bounds (A_Λ, B_Λ) and (A_Γ, B_Γ), respectively. If there exist constants $0 < \lambda_1, \lambda_2, \mu < 1$ such that:*

$$\frac{2}{A_\Lambda}\left(\sqrt{B_\Lambda}+\sqrt{B_\Gamma}\right)\left(\lambda_1\sqrt{B_\Lambda}+\lambda_2\sqrt{B_\Gamma}+\mu\right)\leq 1$$

and

$$\|T_\Lambda(f)-T_\Gamma(f)\|\leq\lambda_1\|T_\Lambda(f)\|+\lambda_2\|T_\Gamma(f)\|+\mu, \tag{4}$$

where T_Λ, T_Γ are the analysis operators for these g-fusion frames, then $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \mu_i)\}_{i\in\mathbb{I}}$ are woven g-fusion frames.

Theorem 3.2. *Let $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \mu_i)\}_{i\in\mathbb{I}}$ be g-fusion frames for $\mathcal{H}$ with respect to $\{\mathcal{H}_i\}_{i\in\mathbb{I}}$ and g-fusion frame bounds (A_Λ, B_Λ) and (A_Γ, B_Γ), respectively. If there exist constants $0 < \lambda, \mu, \gamma < 1$, such that $\lambda B_\Lambda + \mu B_\Gamma + \gamma\sqrt{B_\Lambda} < A_\Lambda$. We have*

$$S_\Lambda^\sigma < \lambda S_\Lambda^\sigma + \mu S_\Gamma^\sigma + \gamma U_\Lambda^\sigma,$$

where S_Λ, U_Λ are g-fusion frame operators of $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$. Then $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \mu_i)\}_{i\in\mathbb{I}}$ are woven g-fusion frame with universal woven bounds

$$\left(A_\Lambda-\lambda B_\Lambda-\mu B_\Gamma-\gamma\sqrt{B_\Lambda}\right),\qquad\left(B_\Gamma+\lambda B_\Lambda+\mu B_\Gamma+\gamma\sqrt{B_\Lambda}\right).$$

Theorem 3.3. *Let $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \nu_i)\}_{i\in\mathbb{I}}$ be g-fusion frames of $\mathcal{H}$ with respect to $\{\mathcal{H}_i\}_{i\in\mathbb{I}}$ and g-fusion frame bounds (A_Λ, B_Λ) and (A_Γ, B_Γ), respectively. Also, if there exists a constant $K > 0$, such that for every $\sigma \subseteq \mathbb{I}$:*

$$\sum_{i\in\sigma}\nu_i^2\|\Lambda_i P_{W_i}f-\Gamma P_{\nu_i}f\|\leq K\min\left\{\sum_{i\in\sigma}\nu_i^2\|\Lambda_i P_{W_i}f\|,\sum_{i\in\sigma}\nu_i^2\|\Gamma_i P_{\nu_i}f\|\right\},$$

then $\{(\Lambda_i, W_i, \nu_i)\}_{i\in\mathbb{I}}$ and $\{(\Gamma_i, V_i, \nu_i)\}_{i\in\mathbb{I}}$ are woven g-fusion frame.

4 Conclusion

In this paper, we show that those of g-fusion frames that are small perturbations of each other, constitute woven g-fusion frame. We start with Paley-Wiener perturbation of weaving g-fusion frames and continue two results of perturbations in the sequel.

References

[1] M.S. Asgari and A. Khosravi, *Frames and bases of subspaces in Hilbert spaces*, J. Math. Anal. Appl., 308 (2005),pp. 541-553.

[2] T. Bemrose, P.G. Casazza, K. Gröchenig, M.C. Lammers and R.G. Lynch, *Weaving Frames*, J. Oper. Matrices., 10 (2016), pp. 1093-1116.

[3] P.G. Casazza, D. Freeman and R.G. Lynch, *Weaving Schauder frames*, J. Approx. Theory., 211 (2016), pp. 42-60.

[4] P.G. Casazza and G. Kutyniok, *Frames of Subspaces*, Contemp Math. Amer. Math. Soc., 345, (2004), pp. 87-113.

[5] P.G. Casazza, G. Kutyniok and Sh. Li, *Fusion frames and distributed processing*, Appl. Comput. Harmon. Anal., 25 (2008), pp. 114-132.

[6] P.G. Casazza and R.G. Lynch, *Weaving properties of Hilbert space frames*, J. Proc. SampTA., (2015), pp. 110-114.

[7] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhauser, Boston, (2016).

[8] I. Daubechies, A. Grossmann and Y. Meyer, *Painless nonorthogonal expansions*, J. Math. Phys., 27 (1986), pp. 1271-1283.

[9] Deepshikha and L.K. Vashisht, *Weaving K-frames in Hilbert spaces*, Results Math., 73 (2018), pp. 1-20.

[10] R.J. Duffin and A.C. Schaeffer, *A class of nonharmonic Fourier series*, Trans. Amer. Math. Soc., 72, (1952), pp. 341-366.

[11] D. Han and D. Larson, *Frames, bases and group representations*, Mem. Amer. Math. Soc., 147 (697) (2000).

[12] A. Khosravi and M.S. Asgari, *Frames of subspaces and approximation of the inverse frame operator*, Houston. J. Math., 33 (2007), pp. 907-920.

[13] A. Khosravi and K. Musazadeh, *Fusion frames and g-frames*, J. Math. Anal. Appl., 342 (2008), pp. 1068-1083.

[14] A. Khosravi and J. Sohrabi Banyarani, *Weaving g-frames and weaving fusion frames*, Bull. Malays. Math. Sci. Soc., 42 (2019), pp. 3111-3129.

[15] M. Mohammadrezaee, M. Rashidi-Kouchi, A. Nazari and A. Oloomi *Woven g-Fusion Frames in Hilbert Spaces* Sahand Commun. Math. Anal., 18 (2021), pp. 133-151.

[16] R. Rezapour, A. Rahimi, E. Osgooei and H. Dehghan, *Controlled weaving frames in Hilbert spaces*, Inf. Dim. Anal. Quan. Prob. Rel. Top., 22 (2019), pp. 1-18.

[17] Sadri, V. Rahimlou, Gh. Ahmadi R. and Zarghami Farfar, R.:*Generalized Fusion Frames in Hilbert Spaces*, Inf. Dim. Anal. Quan. Prob. Rel. Top., (to appear).

[18] W. Sun, *g-frames and g-Riesz bases*, J. Math. Anal. Appl., 322 (2006), pp. 437-452.

[19] L.K. Vashisht and Deepshikha, *On continuous weaving frames*, Adv. Pure Appl. Math., 8 (2017), pp. 15-31.

[20] L.K. Vashisht and Deepshikha, *Weaving properties of generalized continuous frames gen erated by an iterated function system*, J. Geom. Phys., 110 (2016), pp. 282-295.

[21] L.K. Vashisht and Deepshikha, S. Garg and G. Verma , *On weaving fusion frames for Hilbert spaces*, In: Proceedings of SampTA., (2017), pp. 381-385.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Quantum Detection Problem via Fusion Frames

Asghar Rahimi[1,*] and Abdollah Ghasemzadeh[2]

[1,2]Department of Mathematics, University of Maragheh, Maragheh, Iran

Abstract

The quantum detection problem was considered recently for discrete frames in both finite and infinite dimensional Hilbert spaces and also for continuous frames. The quantum detection problem ask to characterize the POVM (positive operator-valued measures) that are informationally complete. It can be split as follows: the injective or state separability problem and the rang analysis or state estimation problem. Improving and extending this notion, in this note we will study the quantum detection problem for fusion frames.

Keywords: Quantum detection, Fusion frame, Positive operator-valued measure, Injectivity problem, State estimation

Mathematics Subject Classification [2010]: 42C15

1 Introduction

Frames in Hilbert spaces at the first were introduced by Duffin and Schaeffer to deal with nonharmonic Fourier series in 1952 [7] and widely studied from 1986 since the great work by Daubechies, Grossmann and Meyer constructed [6]. For a more complete treatment of frame theory we recommend the excellent book of Christensen [5].

In the early 20 th century, the concept of frame nicely generalized to new notion under the name of "frame of subspaces", which are now known as "fusion frames". Fusion frames is a generalization of frames which were introduced by Cassaza and Kutyniok [4] in 2003. The significance of fusion frame is the construction of global frames from local frames in Hilbert space, so the characteristic fusion frame is special suiting for application such as distributed sensing, parallel processing, and packet encoding and so on.

Quantum detection has applications in optical communications, including the detection of coherent light signals such as radio, radar, and laser signals. Quantum detection theory is a reformulation, in quantum mechanical terms, of statistical decision theory.

The problem of quantum detection is to uniquely determine a state from quantum measurements described as positive operator-valued measures (POVMs) acting on a state. This problem was recently settled in Botelho-Andrade et al. [1, 2] mainly for finite or infinite but discrete frames and Han et al. [9] and Hong and Li [10] for continuous frames by constructing some kinds of frame POVMs.

*Speaker. rahimi@maragheh.ac.ir

The quantum detection problem ask to characterize the POVM that are informationally complete. The quantum detection problem can be split as follows: the injective or state separability problem and the rang analysis or state estimation problem.

A quantum injective frame is a frame that can be used to distinguish density operators (states) from their frame measurement and the frame quantum detection problem asks to characterize all of such frames. Recently, D. Han et. al. studied quantum injectivity of multi-window Gabor frames in finite dimensions, [8]. The purpose of this paper is to investigate the quantum detection for fusion frames and gives some equivalent conditions which help us to classify injective fusion frames.

Definition 1.1. A countable family of elements $\{f_i\}_{i\in I}$ in H is a frame for H, if there exist constant $A, B \geq 0$, such that

$$A\|f\|^2 \leq \sum_{i\in I} |\langle f, f_i\rangle|^2 \leq B\|f\|^2 \quad \forall f \in H. \tag{1}$$

The constant A and B are called lower and upper frame bounds. If only the upper inequality in (1) satisfies, it called a Bessel sequence. A frame is tight if $A = B$ and if $A = B = 1$ it is called Parseval frame. We call a frame $\{f_i\}_{i\in I}$ uniform (or equal norm), if we have $\|f_i\| = \|f_j\|$ for all $i, j \in I$. A frame is exact if it ceases to be a frame whenever any single element is deleted. A sequence is called a frame sequence, if it is a frame for its closed linear span. Moreover, we say that a frame $\{f_i\}_{i\in I}$ is a Riesz frame, if every subfamily of the sequence $\{f_i\}_i$ is a frame sequence with uniform frame bound A and B.

Definition 1.2. Let I be some index set, and let $\{v_i\}_{i\in I}$ be family of weights, i.e. $v_i > 0$ for all $i \in I$, the family of closed subspace $\{W_i\}_{i\in I}$ of the Hilbert space H is a fusion frame with respect to $\{v_i\}_{i\in I}$ for H, if there exist constants $0 < C < D < \infty$ such that

$$C\|f\|^2 \leq \sum_i v_i^2 \|\pi_{W_i}(f)\|^2 \leq D\|f\|^2 \qquad \forall f \in H, \tag{2}$$

where π_{W_i} is the orthogonal projection onto W_i. We call C and D the bounds for the fusion frame. The family $\{W_i\}_{i\in I}$ is called a tight fusion frame with respect to $\{v_i\}_{i\in I}$ if in (2) the constants C and D can be chosen so that $C = D$. A Parseval fusion frame with respect to $\{v_i\}_{i\in I}$ provided that $C = D = 1$. Moreover, we call a fusion frames with respect to $\{v_i\}_{i\in I}$, v-uniform if $v := v_i = v_j$ for all $i, j \in I$. If we only have the upper bound, we call $\{W_i\}_{i\in I}$ a fusion Bessel sequence with respect to $\{v_i\}_{i\in I}$ with Bessel bound D.

1.1 Quantum detection and positive operator-valued measure (POVM)

Positive operator-valued measure (POVM) also called generalized observable a mathematical object, consisting of a family of operators in Hilbert spaces that accurse in quantum the arterial formulas for probability distribution of the random outcome a quantum mechanical experiment. The concept of POVM contains as a special case that of observable represented by self- adjoint operator and quantum theory to predict the probability of observing outcomes from a sequence of measurement of the system in an unknown state. This process is called quantum state tomography (QST) [3] and the quantum statistic are described by a positive operator-valued measure (POVM) [1, 9, 10]. In fact quantum measurement extremes the transmitted information from received quantum signals and therefore performs an important role in quantum communications. Actually, quantum state tomography asks to recover a state (sometimes known density operator) from the

probability of observing outcomes from a collection of measurements of the system on this state, and retrieving data from quantum systems is carried out according to quantum measurement theory [4, 11]. In this process positive operator-valued measure plays the central role.

The simplest quantum measurement is the projection-valued measure (PVM), also called standard measurement on von Neumenn measurement.

Let X denote a set of outcomes (e. g. this called be finite or infinite subset of $\mathbb{Z}^d$ of $\mathbb{R}^d$) and β denote a sigma algebra of subset of X.

Definition 1.3. A positive operator-valued measure (POVM) is a function $\Pi : \beta \to \mathrm{Sym}^+(H)$ satisfying

1. $\Pi(\emptyset) = 0$ (the zero operator);
2. for every at most countable disjoint family $\{V_i\} \subset \beta$ and $x, y \in H$ we have $\langle \Pi(\cup_i V_i)x, y\rangle = \sum_i \langle \Pi(V_i)x, y\rangle$;
3. $\Pi(X) = I$ (the identity operator).

Definition 1.4. A quantum system is defined as a von Neumann algebra $\mathcal{A}$ of operators on H. The set of state on H is

$$\mathbb{S}(H) = \{T \in S_1, T = T^* \geq 0, \mathrm{tr}(T) = 1\}.$$

It is known that [1] the set of quantum state are in one to one correspondence with the linear functionals on $\mathcal{A}$ of the form:

$$\rho : \mathcal{A} \to \mathbb{C} \quad \text{for} \quad S \in \mathbb{S}(H), \rho(T) = \mathrm{tr}(ST), \qquad \forall T \in \mathcal{A}.$$

Let $L(\beta, \mathbb{R})$ denote the set of real-valued bounded functions defined on β.

Given a POVM Π associated to a von Neumann algebra $\mathcal{A}$, the quantum detection problem asks two questions:

1. **Injectivity Problem**. Is the following map injective

$$M : \mathbb{S}(\mathcal{A}) \to L(\beta, \mathbb{R}), \quad M(\rho)(U) = \rho(\Pi(U))?$$

2. **Range Analysis or State Estimation**. Assume M is injective. Then given a map $p \in L(\beta, \mathbb{R})$ determine of p is the rang of M. Hence is of the form $p = M(\rho)$, for some unique $\rho \in \mathbb{S}(\mathcal{A})$. If not find a quantum state ρ that best approximation p in some sense (e. g. robustness to noise).

2 Injectivity Problem by Frames and Fusion Frames

In this section we define the notion of injectivity for fusion frames and we study some of its properties.

2.1 Injectivity Problem by Frame and equivalent frames

Definition 2.1. A family of vectors $X = \{x_k\}_{k \in I}$ in a Hilbert space H is said to be injective, if given a Hilbert Schmidt, self-adjoint T satisfying $\langle Tx_k, x_k\rangle = 0$ for all $k \in I$, then $T = 0$.

For finite dimensional Hilbert space H^n, an injective family of vectors is a frame. Since if $\{x_k\}_{k=1}^m$ in a Hilbert space H^n is injective and P is the orthogonal projection onto $(span\{x_k\}_{k=1}^m)^\perp$, then $\langle Px_k, x_k\rangle = 0$ for all $k = 1, 2, ..., m$ and $P \neq 0$ therefor $(span\{x_k\}_{k=1}^m)^\perp = \{0\}$ which means $\{x_k\}_{k=1}^m$ is a frame sequence and the words frame sequences and frames on finite dimensional Hilbert spaces are coincide.

But inverse of this is not the case in general.

Proposition 2.2. *[2] Let $\{x_k\}_{k\in I}$ be a frame for H which gives injectivity. If F is a bounded invertible operator on H, then $\{Fx_k\}_{k\in I}$ also gives injectivity .*

Definition 2.3. Let $\{W_i, v_i\}_{i\in I}$ be a fusion frame for H, and $\pi_{W_i} : H \to W_i$ is the orthogonal projection onto W_i, if $\operatorname{tr}(\pi_{W_i} T) = 0$ for all $i \in I$, $T = T^*$ and $T \in S_j$, $j = 1, 2$ imply that $T = 0$, then $\{W_i, v_i\}_{i\in I}$ is called S_j- injective. (S_p Schatten p-classes operators).

The following theorem gives some equivalent conditions for injectivity of fusion frames.

Theorem 2.4. *Let $\{W_i, v_i\}_{i\in I}$ be a fusion frame for H^n, then following are equivalent:*

1. *if $T, S \in S_j (j = 1, 2)$ are positive, self adjoint and $\operatorname{tr}(\pi_{W_i} T) = \operatorname{tr}(\pi_{W_i} S), \forall i \in I$, then $T = S$.*
2. *if $T, S \in S_j$ are self -adjoint and $\operatorname{tr}(\pi_{W_i} T) = \operatorname{tr}(\pi_{W_i} S), \forall i \in I$, then $T = S$.*
3. *$\{W_i, v_i\}_{i\in I}$ is injective.*

Now we will give another classification of injectivity for the quantum detection problem. The advantage is, if a fusion frame gives injectivity in the quantum detection problem, then it must satisfy these complex requirements.

Theorem 2.5. *Let $\{W_i, v_i\}_{i=1}^m$ be a fusion frame for H^n (real or complex). Then*

1 $\{W_i, v_i\}_{i=1}^m$ gives injectivity.

2 For any ONB $\{e_i\}_{i=1}^n$ for H^n, we have

$$U = \operatorname{span}\{(\langle \pi_{W_k} e_1, e_1\rangle, \langle \pi_{W_k} e_2, e_2\rangle, \ldots, \langle \pi_{W_k} e_n, e_n\rangle)\}_{k=1}^m = \mathbb{R}^n.$$

The states or the positive operators of trace one has essential role in the studying of quantum theory and in the next theorem we give a classification of injectivity for states. It is known that Hilbert-Schmidt and trace classes are not coincide in infinite dimensional Hilbert spaces. If we consider operators which are trace class, then we have the following classification for the infinite dimensions: first, we need a definition.

Definition 2.6. We define a subspace of the real space ℓ_1 as follows:

$$\mathbb{D} := \left\{ (\lambda_1, \lambda_2, \ldots) \in \ell_1 : \sum_{j=1}^{\infty} \lambda_j = 0 \right\}.$$

Theorem 2.7. *Let $\{W_k, v_k\}_{k\in I}$ be a fusion frame for an infinite dimension real or complex Hilbert space H. Then the following are equivalent:*

1 . T is trace class operator and

$$\operatorname{tr}(\pi_{W_k} T) = 0 \quad \forall k \in I \Rightarrow T = 0;$$

2 . for $\lambda \in \mathbb{D}$ and given ONB $\{e_j\}_{j=1}^{\infty}$, if

$$\sum_j \lambda_j \|\pi_{W_k} e_j\| = 0 \quad \forall k \in I \Rightarrow \lambda = 0.$$

References

[1] S. Botelho-Andrade, P. G. Casazza, D. Cheng and T.T. Tran, The solution to the frame quantum detection problem, *J. Fourier Anal. Appl.*, 25 (2019), 2268-2323.

[2] S. Botelho-Andrade, P. G. Casazza, D. Cheng, J. Haas and T.T. Tran, The Quantum Detection Problem: A Survey, *Springer Proc.Math.Stat.* 255 (2017), 337-352.

[3] P. Busch, P. Lahti, J. P. Pellonpa and K. Ylinen, *Quantum Measurement*, Springer, Berlin, 2016.

[4] P. G. Casazza and G. Kutyniok, Frames of Subspaces, *Contemp Math. Amer. Math. Soc.*, 345 (2004), 87-113.

[5] O. Christensen, *An Introduction to Frames and Riesz bases*, Birkhauser, Boston, 2016.

[6] I. Daubechies, A. Grossman and Y. Meyer, Painless nonorthogonal expansions, *J. Math. Phys.*, 27, (1986), No.5, 1271-1283.

[7] R. J. Duffin and A. C. Schaeffer, A class of nonharmonic Fourier series, *Trans. Am. Math. Soc.* 72, (1952), 341-366.

[8] D. Han, Q. Hu, L. Rui and H. Wang, Quantum injectivity of multi-window Gabor frames in finite dimensions, *Ann. Funct. Anal.* (2022) 13:59 , https://doi.org/10.1007/s43034-022-00208-2

[9] D. Han, Q. Hu and R. Liu, Injective continuous frames and quantum detections. *Banach J. Math. Anal.*, 15, (2021), 1-24.

[10] G. Hong and P. Li, On the Continuous Frame Quantum Detection Problem. *Results in Math.*, 78(2), 64.(2023).

[11] M. Paris and J. Rehacek, *Quantum State Estimation*, (Lecture Notes in Physics, 649), Springer, Berlin, 2004.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

BREAKDOWNS OF RRGMRES and DGMRES

Faranges Kyanfar*

Department of Applied Mathematics, Shahid Bahonar University , Kerman, Iran

Abstract

The GMRES method is one of the most common iterative methods to solve linear systems of equations with an $n \times n$ large nonsingular matrix . When the matrix is singular, the GMRES method may break down before determining an acceptable approximate solution. The RRGMRES and DGMRES are modified GMRES restricting the Krylov subspaces within the range of A and the range of A^m to make solutions more stable, where m is the index of A. The aim of this paper is to characterize breakdowns and least square solutions of the RRGMRES and DGMRES algorithms for solving a singular linear system equations.

Keywords: singular linear systems, breakdown, RRGMRES, DGMRES
Mathematics Subject Classification [2010]: 65F10, 65F20,15A03

1 Introduction

The GMRES method [2] is a Krylov method for solving the nonsingular linear system

$$Ax = b, \quad A \in \mathbb{R}^{n\times n}, \quad x, b \in \mathbb{R}^n.$$

Application of the Arnoldi process to the matrix A with initial vector b yields at step k

$$AV_k = V_k H_k - h_{j+1,j} v_{j+1} e_k^T = AV_k = V_k H_k - w_k e_k^T = V_{k+1} \bar{H}_k$$

The GMRES method seeks an approximate solution $x_k \in \mathcal{K}_k(A, r_0)$ such that

$$\|r_k\| = \|b - Ax_k\| = \min_{z\in\mathcal{K}_k} \|b - Az\| = \min_{y\in\mathbb{R}_k} \|e_1 - H_k y\|, \quad x_k := V_k y_k, \ y_k \in \mathbb{R}^k.$$

If the matrix $A \in \mathbb{R}^{n\times n}$ is nonsingular, then the Hessenberg matrix H_k is nonsingular and hence y_k is determined uniquely. If A is singular, then the Hessenberg matrix H_k at step k may be singular and the vector $x_k := V_k y_k$ is not approximated the solution of the linear system of equations.

The Arnoldi process breaks down when $h_{k+1,k} = 0$ equivalent to $w_k = 0$ for some k. The GMRES method breaks down at step k if the Arnoldi process breakdown at step k.

We remind that the smallest nonnegative integer m for which $\text{rank}(A^m) = \text{rank}(A^{m+1})$ is called the *index* of A and is denoted by $ind(A)$. Also, a matrix is said to be rank-deficient

*Email address: kyanfar@uk.ac.ir

if it does not have full rank and also is said to be ill-conditioned if the smallest singular value of A is zero or the ratio of the largest singular value and the smallest singular value is very large.

Rank-Deficient Linear Least-Squares Problems occurs when the matrix is ill-conditioned or rank-deficient.

Lemma 1.1. *Two kinds of breakdowns at step $k+1$ are as follows:*

1. *A degeneracy of the Krylov subspace occurs when*

$$dim(\mathcal{K}_k) = dim(A\mathcal{K}_k) = k \quad \& \quad dim(\mathcal{K}_{k+1}) < k+1.$$

2. *A rank-deficient of the least-squares problem occurs when*

$$dim(\mathcal{K}_k) = k \quad \& \quad dim(A\mathcal{K}_k) < k.$$

The second case is called a hard breakdown.

Proposition 1.2. *[4] Let matrix $A \in \mathbb{R}^{n\times n}$ with $rank(A) = N < n$ and $b \in \mathcal{R}(A)$. The GMRES method breaks down at step N. Then one of the following hold.*

- *The GMRES method determines a solution of the linear system at breakdown, if $dim(A\mathcal{K}_N) = N$.*
- *The linear system has a solution in $\mathcal{K}_N + \mathcal{R}(A^T)$, if $dim(A\mathcal{K}_N) < N$.*

The RRGMRES algorithm [1] computes a sequence of approximate solutions of $Ax = b$. The k^{th} approximation $x_k \in x_0 + \mathcal{K}_k^*$, satisfies:

$$\|b - Ax_k\| = \min_{z\in\mathcal{K}_k^*} \|b - Az\|, \quad x_k \in \mathcal{K}_k^*(A, r_0),$$

$$\mathcal{K}_k^*(A, r_0) := \mathcal{K}_k(A,\ Ar_0) = \text{Span}\{Ar_0,\ A^2r_0, \ldots,\ A^kr_0\}.$$

Note that the Krylov subspace is restricted to the range of A.

Let A be a matrix of index m. The DGMRES algorithm [3] computes a sequence of approximate solutions of $Ax = b$. The k^{th} approximation $x_k \in x_0 + \mathcal{K}_k^D$, satisfies:

$$\|b - Ax_k\| = \min_{z\in\mathcal{K}_k^D} \|b - Az\|, \quad x_k \in \mathcal{K}_k^D(A, r_0),$$

$$\mathcal{K}_k^D(A, r_0) := \mathcal{K}_k(A,\ A^m r_0) = \text{Span}\{A^m r_0,\ A^{m+1}r_0, \ldots,\ A^{m+k-1}r_0\}.$$

Note that the Krylov subspace is restricted to the range of A^m.

2 Main results

In this section, we study the breakdowns and least square solutions of GMRES, RRGMRES and DGMRES methods for singular systems [1, 4, 5].

Proposition 2.1. *The RRGMRES method breaks down at step $k+1$, if k is the smallest positive integer such that one of the following holds.*

1. $dim(\mathcal{K}_{k+1}^*(A, r_0)) = dim(A(\mathcal{K}_k^*(A, r_0)) = k$, *(degeneracy of the Krylov subspace).*
2. $dim(A(\mathcal{K}_k^*(A, r_0))) < dim(\mathcal{K}_k^*(A, r_0)) = k$, *(hard breakdown of RRGMRES).*

Proposition 2.2. *Let A be a matrix of index m. Then the DGMRES method breaks down at step $k+1$, if k is the smallest positive integer such that one of the following holds.*

1. $dim(\mathcal{K}^D_{k+1}(A, r_0)) = dim(A(\mathcal{K}^D_k(A, r_0)) = k$, *(degeneracy of the Krylov subspace).*
2. $dim(A(\mathcal{K}^D_k(A, r_0))) < dim(\mathcal{K}^D_k(A, r_0)) = k$, *(hard breakdown of DGMRES).*

We know that for nonsingular case, the breakdown of GMRES at step $k+1$ indicates that the solution lives in $\mathcal{K}_k(A, r_0)$. Unlike to the nonsingular case, anything may happen when A is singular, the GMRES, RRGMRES and DGMRES may or may not determine the pseudo-inverse solution.

The following example shows that the GMRES breaks down (degeneracy of the Krylov subspace) at the step two after the least-square solution has been found in the first step.

Example 2.3. Suppose that $A = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ and $x_0 = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$. Then $r_0 = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) = Ar_0$, and hence $\mathcal{K}_2 = \mathcal{K}_1 = span\{\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)\}$. Therefore, the GMRES determines $x_1 = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$ as a least-squares solution. The pseudo-inverse solution $x^\dagger = \left(\begin{smallmatrix} .5 \\ .5 \end{smallmatrix}\right)$ and hence x_1 is not pseudo-inverse solution. Note that

$$dim(A\mathcal{K}_2) = dim(K_2) < 2.$$

The following example shows that the GMRES breaks down (hard breakdown) at the step two after the pseudo-inverse solution has been found in the first step.

Example 2.4. Suppose that $A = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ and $b = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right)$, then $r_0 = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right)$ and $Ar_0 = \left(\begin{smallmatrix} 0 \\ 4 \end{smallmatrix}\right)$, and GMRES determines the pseudo inverse solution $x^\dagger = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$. Note that

$$dim(A(\mathcal{K}_2)) < dim(\mathcal{K}_2) = 2.$$

Theorem 2.5. *Let A be an $n \times n$ matrix with index m and $b \in \mathcal{R}(A^m)$. Then the system of linear equations $Ax = b$ has a solution in $\mathcal{K}^D_k(A, b)$ and the hard breakdown of DGMRES method can not occur.*

Theorem 2.6. *Let A be an $n \times n$ matrix with index one and $b \in \mathcal{R}(A)$. Then the system of linear equations $Ax = b$ has a solution in $\mathcal{K}^*_k(A, b)$ and the hard breakdown of RRGMRES method can not occur for the system of linear equations $Ax = b$.*

Remind that an $n \times n$ matrix A is called EP (equal projector), if $\mathcal{R}(A) = \mathcal{R}(A^T)$ or $AA^\dagger = A^\dagger A$ [5]. In the following theorem, we consider a matrix A such that A^m is an ED matrix, where m is the index of A.

Theorem 2.7. *Let A be an $n \times n$ matrix with index$(A) = m$. Assume $\mathcal{R}(A^m) = \mathcal{R}((A^T)^m)$. Then the DGMRES method produces the pseudo inverse solution.*

The following example shows that the condition $\mathcal{R}(A^m) = \mathcal{R}((A^T)^m)$ is necessary in the above theorem.

Example 2.8. Suppose that $A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ and $x_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$. Then $A^2 r_0 = \begin{pmatrix} 4 \\ 8 \\ 4 \\ 0 \\ 0 \end{pmatrix}$ and $A^3 r_0 = \begin{pmatrix} 12 \\ 24 \\ 12 \\ 0 \\ 0 \end{pmatrix}$. Therefore,

$$\mathcal{K}^D_1(A, r_0) = \mathcal{K}^D_2(A, r_0).$$

Corollary 2.9. *Let A be a singular matrix. Then RRGMRES method at some step, either*

1. *The RRGMRES method determines a solution without breakdown and then breaks down at the next step through degeneracy of the Krylov subspace or*

2. *The RRGMRES method breaks down through rank deficiency of the least-squares problem without determining a solution.*

Proposition 2.10. *Let k be the smallest positive integer such that*

$$dim(A(\mathcal{K}_k^*(A, r_0)) = dim(\mathcal{K}_{k+1}^*(A, r_0)) = k.$$

Then there is a matrix $H_k \in \mathbb{R}^{k\times k}$ such that

$$AK_k^* = K_k^* H_k,$$

where H_k is nonsingular and the matrix $K_k^* := [Ar_0,\ A^2r_0, \ldots,\ A^k r_0]$.

The following example shows that RRGMRES may breaks down before getting any solution, even when the system has a solution, or it may determine a least-squares solution.

Example 2.11. Suppose that $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ and $x_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Then $Ar_0 = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$, $A^2r_0 = \begin{pmatrix} 4 \\ 8 \\ 4 \end{pmatrix}$ and $A^3r_0 = \begin{pmatrix} 12 \\ 24 \\ 12 \end{pmatrix}$. Therefore,

$$\mathcal{K}_1^*(A, r_0) \neq \mathcal{K}_2^*(A, r_0) = \mathcal{K}_3^*(A, r_0).$$

RRGMRES method breaks down at step $m = 3$, but RRGMRES method cannot produce pseudo inverse solution of the linear system $Ax = b$.

Theorem 2.12. *Let A be a singular matrix of rank r. Assume that the RRGMRES method breaks down at step $k + 1$.*

- *If $dim(A\mathcal{K}_k^*) = k$, then the method yields a least-squares solution at breakdown.*
- *If $dim(A\mathcal{K}_k^*) < k$, then the least-square solution belongs to $\mathcal{K}_k^* + \mathcal{R}(A^T)$.*

References

[1] D. Calvetti, B. Lewis and L. Reichel, *GMRES-type methods for inconsistent systems*, Linear Algebra and its Applications, **316** (2000) 157–169.

[2] Y. Saad and M.H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM Journal on scientific and statistical computing, **7-3** (1986) 856-869.

[3] A. Sidi, *DGMRES: A GMRES-type algorithm for Drazin-inverse solution of singular non-symmetric linear systems*, Linear Algebra and its Applications, **335** (2001) 189-204.

[4] L. Reichel and Q. Ye, *Breakdown-free GMRES for singular systems*, SIAM Journal on Matrix Analysis and Applications, **26-4** (2005)1001–1021.

[5] K. Sugihara, K. Hayami and , L. Zeyu , *GMRES using pseudoinverse for range symmetric singular systems*, Journal of Computational and Applied Mathematics, **422** (2023) 114865.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some properties of the block Toeplitz-Hessenberg matrices

M. Shams Solary*

Department of Mathematics, Payame Noor University, Tehran, IRAN

Abstract

In this paper, we have a $t \times t$ matrix polynomial that is defined by

$$L(\lambda) = \lambda^r I - \sum_{j=1}^{r} \lambda^{r-j} \mathcal{C}_j, \quad \mathcal{C}_j \in \mathbb{C}^{t\times t}, \quad j = 1, \ldots, r$$

of degree r that $\sigma(L) = \{\lambda \in \mathbb{C}^1 : \ det\ L(\lambda) = 0\}$.
Here, we try to show a process for finding determinant of the block Toeplitz-Hessenberg matrices from matrix polynomials by the block companion matrices.
Also, we generalize a new version of Trudi's formula for the block Toeplitz-Hessenberg matrices. These tools are used to find eigenvalues, singularities, and stability and instability of systems.

Keywords: Matrix polynomial, Toeplitz-Hessenberg matrices, Determinant, Block
Mathematics Subject Classification [2010]: 65F30, 15A15

1 Introduction

We know that the structure of matrix polynomials instead of ordinary polynomials (the scalar case) arise in a wide variety of applications such as physics and applied mathematics, appearing naturally in the description of systems with multiple discrete variables such as quantum spin, the linear dynamical systems, statistical problems, multigrid techniques, and etc. This paper is highly motivated by [2, 5, 6]. Here, is solved the study of matrix polynomials of arbitrary degree.

2 Main results

Let

$$\boldsymbol{C}_L = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & I \\ \mathcal{C}_r & \cdots & \mathcal{C}_3 & \mathcal{C}_2 & \mathcal{C}_1 \end{bmatrix} \tag{1}$$

*Speaker. Email address: shamssolary@pnu.ac.ir or shamssolary@gmail.com

and

$$L(\lambda) = \lambda^r I - \sum_{j=1}^{r} \lambda^{r-j} \mathcal{C}_j$$

with or without commuting coefficients

$$(\mathcal{C}_i\mathcal{C}_j = \mathcal{C}_j\mathcal{C}_i, \quad or \quad \mathcal{C}_i\mathcal{C}_j \neq \mathcal{C}_j\mathcal{C}_i \quad for \;\; \mathcal{C}_i \in \mathbb{C}^{t\times t}, \;\; i,j = 1,\ldots,r).$$

Throughout, I is the identity matrix and 0 is the zero matrix of any size to satisfy the conformability requirement of a particular operation.
Now, we try to explain a method for finding determinants of block Toeplitz- Hessenberg matrices from matrix polynomials by the block companion matrices. We know that a way for finding the roots of the polynomial is by the companion matrix method. Here, our aim is to study some properties of the block companion matrices by block-Toeplitz ones.

Theorem 2.1. *Let $\boldsymbol{S}$ be the $r-block \times r-block$ lower triangular matrix*

$$\boldsymbol{S} = \begin{bmatrix} I & 0 & 0 & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & \cdots & 0 \\ S_{21} & 0 & I & \ddots & \ddots & \vdots \\ S_{31} & S_{22} & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ S_{(r-1)1} & \cdots & \cdots & S_{2(r-2)} & 0 & I \end{bmatrix}, \tag{2}$$

then $\boldsymbol{S}^{-1}$ is as an $r-block \times r-block$ block lower triangular matrix of the form

$$\boldsymbol{S}^{-1} = \begin{bmatrix} I & 0 & 0 & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & \cdots & 0 \\ T_{21} & 0 & I & \ddots & \ddots & \vdots \\ T_{31} & T_{22} & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ T_{(r-1)1} & \cdots & \cdots & T_{2(r-2)} & 0 & I \end{bmatrix}, \tag{3}$$

with all blocks of size $t \times t$.
Where

$$T_{2i} = -S_{2i}, \quad i = 1,2,\ldots(r-2),$$
$$T_{3i} = -S_{3i}, \quad i = 1,2,\ldots(r-3),$$
$$T_{mk} = -S_{mk} - \sum_{i=2}^{m-2} T_{(m-i)(k+i)} S_{ik}, \quad m = 4,\ldots,(r-1), \quad k = 1,\ldots,(r-4), \tag{4}$$

such that $\boldsymbol{S}\boldsymbol{S}^{-1} = \boldsymbol{S}^{-1}\boldsymbol{S} = \boldsymbol{I}$. $\square$

Theorem 2.2. *Let $\boldsymbol{C}$ be the $r \times r$ for $r \geq 5$, block companion matrix as follows*

$$\boldsymbol{C} = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & I \\ \mathcal{C}_r & \cdots & \mathcal{C}_3 & \mathcal{C}_2 & 0 \end{bmatrix}. \tag{5}$$

Then $\boldsymbol{C}$ *is similar to an* $r-$*block* $\times\, r-$*block Toeplitz matrix of the form*

$$\boldsymbol{T}=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ A_2 & 0 & I & \ddots & \ddots \\ A_3 & A_2 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & I \\ A_r & \cdots & A_3 & A_2 & 0 \end{bmatrix}, \tag{6}$$

with all blocks of size $t \times t$.

By a standard triple for matrix polynomial $L(\lambda)$ in [1, 4], Laurent expansion, and residue theorem in the complex plane for the generating matrix polynomial $L(\lambda)$ in the interior, we found determinant of block-Toeplitz band Matrices.
Theorem 2.2 above was proved in a more general context in [6] but here, was done with different process. The matrix $\boldsymbol{T}$ is obtained from Theorem 2.2 and the transpose of the matrix $\boldsymbol{T}$ in [2].

Example 2.3. Let $n = 6$, $r = 3$, $t = 2$. Then

$$\boldsymbol{C}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ c_{32} & 0 & c_{22} & 0 & 0 & 0 \end{bmatrix},\quad \mathcal{C}_2=\begin{bmatrix} 0 & 1 \\ c_{22} & 0 \end{bmatrix},\quad \mathcal{C}_3=\begin{bmatrix} 0 & 1 \\ c_{32} & 0 \end{bmatrix},$$

$$\boldsymbol{S}=\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ S_{21} & 0 & I \end{bmatrix},\quad \boldsymbol{S}^{-1}=\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -S_{21} & 0 & I \end{bmatrix},\quad \text{with}\quad S_{21}=\frac{1}{2}\mathcal{C}_2.$$

Then

$$\boldsymbol{T}=\boldsymbol{S}^{-1}\,\boldsymbol{C}\boldsymbol{S}=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 1 & 0 \\ \frac{1}{2}c_{22} & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{1}{2} & 0 & 0 \\ c_{32} & 0 & \frac{1}{2}c_{22} & 0 & 0 & 0 \end{bmatrix}$$

is a 3×3 block Toeplitz matrix with blocks of size 2×2.

We know that if the blocks $\mathcal{C}_k$ of the companion matrix $\boldsymbol{C}$ in

$$\boldsymbol{C}=\begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ 0 & 0 & I & \ddots & \ddots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & I \\ \mathcal{C}_r & \cdots & \mathcal{C}_3 & \mathcal{C}_2 & 0 \end{bmatrix} \tag{7}$$

are nonnegative, then the corresponding Toeplitz matrix T in (6) will be also nonnegative, otherwise there is a contradiction.

Example 2.4. For $r = 6$, after some calculations we have:

$$\mathbf{S} = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ \frac{1}{5}\mathcal{C}_2 & 0 & I & 0 & 0 & 0 \\ \frac{1}{4}\mathcal{C}_3 & \frac{2}{5}\mathcal{C}_2 & 0 & I & 0 & 0 \\ \frac{4}{25}\mathcal{C}_2^2 + \frac{1}{3}\mathcal{C}_4 & \frac{2}{4}\mathcal{C}_3 & \frac{3}{5}\mathcal{C}_2 & 0 & I & 0 \\ \frac{1}{2}\mathcal{C}_5 + \frac{9}{40}\mathcal{C}_2\mathcal{C}_3 + \frac{7}{40}\mathcal{C}_3\mathcal{C}_2 & \frac{9}{25}\mathcal{C}_2^2 + \frac{2}{3}\mathcal{C}_4 & \frac{3}{4}\mathcal{C}_3 & \frac{4}{5}\mathcal{C}_2 & 0 & I \end{bmatrix},$$

and the first column of the Toeplitz matrix $T = \mathbf{S}^{-1}\boldsymbol{C}\mathbf{S}$ is

$$T_1 = \begin{bmatrix} 0 \\ \frac{1}{5}\mathcal{C}_2 \\ \frac{1}{4}\mathcal{C}_3 \\ \frac{2}{25}\mathcal{C}_2^2 + \frac{1}{3}\mathcal{C}_4 \\ \frac{1}{2}\mathcal{C}_5 + \frac{3}{40}\mathcal{C}_2\mathcal{C}_3 + \frac{3}{40}\mathcal{C}_3\mathcal{C}_2 \\ \mathcal{C}_6 + \frac{3}{125}\mathcal{C}_2^3 + \frac{1}{15}\mathcal{C}_2\mathcal{C}_4 + \frac{1}{15}\mathcal{C}_4\mathcal{C}_2 + \frac{1}{16}\mathcal{C}_3^2 \end{bmatrix} = \begin{bmatrix} 0 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \end{bmatrix},$$

$$T_2 = \begin{bmatrix} I \\ 0 \\ \frac{1}{5}\mathcal{C}_2 \\ \frac{1}{4}\mathcal{C}_3 \\ \frac{2}{25}\mathcal{C}_2^2 + \frac{1}{3}\mathcal{C}_4 \\ \frac{1}{2}\mathcal{C}_5 + \frac{3}{40}\mathcal{C}_2\mathcal{C}_3 + \frac{3}{40}\mathcal{C}_3\mathcal{C}_2 \end{bmatrix} = \begin{bmatrix} I \\ 0 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \end{bmatrix},$$

$$T_3 = \begin{bmatrix} 0 \\ I \\ 0 \\ \frac{1}{5}\mathcal{C}_2 \\ \frac{1}{4}\mathcal{C}_3 \\ \frac{2}{25}\mathcal{C}_2^2 + \frac{1}{3}\mathcal{C}_4 \end{bmatrix} = \begin{bmatrix} 0 \\ I \\ 0 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix},$$

$$T_4 = \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \\ \frac{1}{5}\mathcal{C}_2 \\ \frac{1}{4}\mathcal{C}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ I \\ 0 \\ A_2 \\ A_3 \end{bmatrix},$$

$$T_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \\ 0 \\ \frac{1}{5}\mathcal{C}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ I \\ 0 \\ A_2 \end{bmatrix},$$

$$T_6 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ I \\ 0 \end{bmatrix}.$$

Here, we generalize a new version of Trudi's formula in [3] for block Toeplitz-Hessenberg matrix:

Theorem 2.5. *Let* $F(x)=\sum\limits_{n\geq 0}A_nx^n$ *and* $G(x)=\sum\limits_{n\geq 0}B_nx^n$ *be two matrix polynomials such that* $F(x)G(x)=I$. *Then*

$$det\begin{bmatrix} A_1 & A_0 & & \cdots & 0\\ A_2 & A_1 & A_0 & \ddots & 0\\ A_3 & A_2 & A_1 & \ddots & 0\\ \vdots & \ddots & \ddots & A_1 & A_0\\ A_r & A_{r-1} & \cdots & A_2 & A_1 \end{bmatrix}=(-1)^r det\ B_r\ \left(det(A_0)\right)^{r+1}.$$

For proof we have

$$F(x)G(x)=\sum_{n\geq 0}\left(\sum_{k=0}^{n}A_kB_{n-k}\right)x^n=I\rightarrow\sum_{k=0}^{n}A_kB_{n-k}=\Delta_{0,n}, \tag{8}$$

that

$$\Delta_{0,n}=\begin{cases} I, & if\ n=0,\\ \mathbf{0}, & if\ n\neq 0. \end{cases} \tag{9}$$

For more details see [2].

Example 2.6. If

$$\boldsymbol{T}_5=\begin{bmatrix} A_1 & I & 0 & 0 & 0\\ A_2 & A_1 & I & 0 & 0\\ A_3 & A_2 & A_1 & I & 0\\ A_4 & A_3 & A_2 & A_1 & I\\ A_5 & A_4 & A_3 & A_2 & A_1 \end{bmatrix},$$

and the matrices $A_i,\ i=1,2,3,4,5$ commute, because the number 5 has seven partitions:

$$1+1+1+1+1,\ 1+4,\ 1+1+3,\ 1+1+1+2,\ 2+2+1,\ 3+2\ and\ 5,$$

then

$$det\ \boldsymbol{T}_5=det\left(A_1^5-2A_1A_4+3A_1^2A_3-4A_1^3A_2+3A_1A_2^2-2A_2A_3+A_5\right).$$

3 Conclusion

In this paper, we show that Block companion matrices are similar to block-Toeplitz ones, Also, we generalize Trudi's formula in scalar case to block case, and we get Determinant of a block Toeplitz-Hessenberg matrix that each result can be used independently in other research articles and activities. These processes are applied to find eigenvalues, singularities, and stability and instability of systems.

References

[1] Z. Gong, M. Aldeen, L. Elsner, A note on a generalized Cramer's rule, *Linear Algebra Appl.*, 340 (2002), No. 13, 253–254.

[2] M. Shams Solary, From matrix polynomial to determinant of block Toeplitz-Hessenberg matrix, *Numerical Algorithms*, Accepted (2023).

[3] M. Merca , A note on the determinant of a Toeplitz-Hessenberg matrix, *Special Matrices*, (2013), 10–16.

[4] M. Tismenetsky, Determinant of block Toeplitz band matrices. *Linear Algebra Appl.*, 85 (1987), 165–184.

[5] L. N. Trefethen, *Approximation Theory and Approximation Practice*, Society for Industrial and Applied Mathematics, 2018.

[6] H. K. Wimmer, Similarity of block companion and block Toeplitz matrices, *Linear Algebra Appl.*, 343–344 (2002), 381–387.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Extensions of the fundamental theorem of algebra

Bamdad R. Yahaghi*

Department of Mathematics, Faculty of Sciences, Golestan University, Gorgan 19395-5746, Iran

Abstract

In this talk motivated by the celebrated fundamental theorem of algebra and its standard proof utilizing Liouville's Theorem, we prove the fundamental theorem of algebra type results for both commutative and noncommutative polynomials in several settings, e.g., the setting of associative locally convex complex algebras and that of such real algebras whose centers contain certain copies of complex numbers. An application of one of the main results of the paper is the existence of eigenvalues for matrices with entries from arbitrary finite-dimensional complex algebras. A conjecture extending the fundamental theorem of algebra to noncommutative polynomials with coefficients from locally convex associative real algebras containing a copy of the complex numbers is proposed.

Keywords: (Noncommutative) Polynomials, Real/Complex algebras, Commutative/ Alternative/ Associative/Noncommutative/Nonassociative algebras, (Left/Right) eigenvalues, Singular elements, Copies of complex numbers.

Mathematics Subject Classification [2010]: 46H70, 13J30, 15A18, 15A60, 16W99, 17D05.

1 Introduction

The celebrated fundamental theorem of algebra needs no introduction. It is safe to say that we all were exposed to the theorem, perhaps the real version of it, in high school. From then on, with the fundamental theorem at our disposal, we did various interesting problems and theorems such as *any polynomial with real or complex coefficients that is nonnegative on the real line is the sum of squares of two real polynomials* and Lucas' theorem which asserts *the roots of the derivative of a given nonconstant complex polynomial lie in the convex hull of those of the given polynomial.* All that being said, we refer the reader to [3, Chapter 4] and [5] for a detailed introduction as well as a comprehensive account of the fundamental theorem and its history. For a fundamental theorem of algebra for for polynomial equations over real composition algebras, see [10].

In this note, we consider arbitrary real or complex unital algebras of the following kinds: finite-dimensional algebras, normed algebras, and locally convex algebras, all real algebras containing a copy of complex numbers. First, we prove a Fundamental Theorem of Algebra type result for such commutative and associative algebras and for finite-dimensional commutative and alternative algebras. We also prove the existence of right eigenvalues for

*bamdad5@hotmail.com, bbaammddaadd55@gmail.com

matrices with entries in such finite-dimensional associative real algebras. Next, we present a Fundamental Theorem of Algebra type result for both real and complex locally convex associative algebras with an identity element. We use this result to prove the existence of eigenvalues for matrices with entries from arbitrary finite-dimensional complex unital algebras as well as certain real unital algebras that contain a copy of complex numbers.

Let us begin by setting the stage. A vector space $\mathbb{A}$ over reals (resp. complex numbers) together with a multiplication coming from a bilinear form on $\mathbb{A}$ is said to be a real (resp. complex) algebra. The algebra $\mathbb{A}$ is said to be unital or to have an identity if its multiplication has an identity element, denoted by 1. The identity element of the addition operation of the algebra is denoted by 0. The algebra $\mathbb{A}$ is called associative (resp. commutative) if its multiplication is associative (resp. commutative). Throughout, by an algebra we mean an arbitrary real or complex algebra not necessarily associative or commutative. A nonzero element $a \in \mathbb{A}$ is said to be invertible if there exists a unique element of the algebra, denoted by a^{-1}, satisfying the relations $aa^{-1} = a^{-1}a = 1$. The symbol $\mathbb{A}^{-1}$ is used to denote the set of all invertible elements of the algebra $\mathbb{A}$. A nonzero element $a \in \mathbb{A}$ is said to be a nonzero-divisor if $ab = 0$ or $ba = 0$ with $b \in \mathbb{A}$ implies that $b = 0$.

An algebra $\mathbb{A}$ is said to be alternative if for every $a, b \in \mathbb{A}$, the real subalgebra generated by the elements a and b is associative. By a theorem of E. Artin, [9, Theorem 3.1], an algebra $\mathbb{A}$ is alternative if and only if $a(ab) = (aa)b$ and $a(bb) = (ab)b$ for all $a, b \in \mathbb{A}$. Note that if an algebra $\mathbb{A}$ is associative, then the uniqueness of the inverse element is a redundant hypothesis in the definition of the invertible elements of $\mathbb{A}$. Also the uniqueness in the definition of invertible elements of an algebra $\mathbb{A}$ is a redundant hypothesis whenever the algebra $\mathbb{A}$ is alternative and finite-dimensional.

Let $\mathbb{A}$ be a real or complex algebra with identity. The algebra $\mathbb{A}$ together with a Hausdorff topology is said to be a topological algebra if the operations addition, multiplication, and the scalar product of the algebra $\mathbb{A}$ are all continuous and that the inversion, defined on $\mathbb{A}^{-1}$, the set of the invertible elements of $\mathbb{A}$, is continuous on $\mathbb{A}^{-1}$. An algebra norm $||.||$ of a unital algebra $\mathbb{A}$ is said to be unital if $||1|| = 1$, where the first 1 denotes the identity element of the algebra $\mathbb{A}$.

Let R be a commutative ring with identity. As is usual, the symbol $R[x]$ stands for the ring of all polynomials in the indeterminate x with coefficients in the ring R. Let $f = f_0 + f_1x + \cdots + f_nx^n \in R[x]$ be of degree $n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, i.e., $f_n \neq 0$, where R is a commutative ring with identity; the coefficient $f_n \in R \setminus \{0\}$ is called the leading coefficient of the polynomial f. We say that an element $r \in R$ is a singular element for the polynomial f if $f(r) := f_0 + f_1r + \cdots + f_nr^n \notin R^{-1}$. Let R be a ring with identity, not necessarily commutative nor associative. An expression of the form $f_0xf_1xf_2 \cdots xf_n$ with $n \in \mathbb{N}_0$ and $f_i \in R \setminus \{0\}$ $(0 \leq i \leq n)$ is said to be a noncommutative monomial of degree n with coefficients f_i in the indeterminate x. Note that if the ring is nonassociative, proper parentheses must be inserted in the expression $f_0xf_1xf_2 \cdots xf_n$ to make it sensible. By a noncommutative polynomial in the indeterminate x and with coefficients in the ring R, we mean a finite sum of noncommutative monomials with coefficients in R, each of which is called a monomial summand of the noncommutative polynomial. The sum of all noncommutative monomials of the greatest degree in a noncommutative polynomial is called the leading polynomial part of the noncommutative polynomial. For instance, $f_0 + f_1(xf_1') + (f_1x)f_1' + f_1''x + xf_1'''$, where the coefficients come from R, is an example of a noncommutative polynomial of degree at most 1 with coefficients in R whose leading noncommutative polynomial part is $f_1(xf_1') + (f_1x)f_1' + f_1''x + xf_1'''$. The notion of singular elements of noncommutative polynomials with coefficients in a unital ring R can be defined

in a similar fashion.

Let R be a ring and $A \in M_n(R)$, the set of all $n \times n$ matrices with entries from R. An element $\lambda \in R$ is said to be a left (resp. right) eigenvalue of the matrix A if there is a nonzero $n \times 1$ column matrix $X \in R^n := M_{n\times 1}(R)$ such that $AX = \lambda X$ (resp. $AX = X\lambda$). An element $\lambda \in R$ is said to be an eigenvalue of the matrix A if there is a nonzero $n \times 1$ column matrix $X \in R^n := M_{n\times 1}(R)$ such that $AX = \lambda X = X\lambda$.

The following theorem, taken from [6], see [6, Propositions 1.1.7, 1.1.111] and [1, Theorem 4], is quoted here for reader's convenience. For a thorough treatment of theory of (complete) normed algebras, we refer the reader to the classical references [7] and [2] and to the more recent excellent reference [6].

Theorem 1.1. *Every finite-dimensional real or complex algebra can be normed. Moreover, if a real or complex algebra is finite-dimensional and unital, it can be equipped with a unital norm.*

2 Main results

We start off with a useful lemma.

Lemma 2.1. (i) *Let $(\mathbb{A}, ||.||)$ be a normed associative algebra over reals. If the algebra $\mathbb{A}$ contains a copy of complex numbers, say, $\mathbb{C}_I := \{a+bI : a, b \in \mathbb{R}\}$ with $I \in \mathbb{A}$ and $I^2 = -1$, then there exists an algebra norm $||.||' : \mathbb{A} \longrightarrow \mathbb{R}$ on $\mathbb{A}$ such that $||zaw||' = |z|||a||'|w|$ for all $z, w \in \mathbb{C}_I$ and $a \in \mathbb{A}$.*

(ii) *Let $\mathbb{A}$ be a finite-dimensional associative algebra over reals. If the algebra $\mathbb{A}$ contains a copy of the complex numbers, say, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$ with $I \in A$ and $I^2 = -1$, then there exists an algebra norm $||.|| : \mathbb{A} \longrightarrow \mathbb{R}$ on $\mathbb{A}$ such that $||zaw|| = |z|||a|||w|$ for all $z, w \in \mathbb{C}_I$ and $a \in \mathbb{A}$.*

The following theorem can be thought of as an extension of the Fundamental Theorem of Algebra to normed (resp. finite-dimensional, locally convex) commutative and associative real algebras containing a copy of the complex numbers.

Theorem 2.2. (i) *Let $\mathbb{A}$ be a commutative and associative real algebra that is either normed or finite-dimensional and contains a copy of the complex numbers, say, $\mathbb{C}_I := \{a+bI : a, b \in \mathbb{R}\}$, where $I \in \mathbb{A}$ is such that $I^2 = -1$. Then every nonconstant polynomial with coefficients in $\mathbb{A}$ and with an invertible leading coefficient has a singular element in $\mathbb{C}_I$.*

(ii) *Let $\mathbb{A}$ be a commutative and associative locally convex real algebra containing a copy of the complex numbers, say, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$, where $I \in \mathbb{A}$ is such that $I^2 = -1$. Then every nonconstant polynomial with coefficients in $\mathbb{A}$ and with an invertible leading coefficient has a singular element in $\mathbb{C}_I$.*

In fact when the real algebra is finite-dimensional, we can say more.

Theorem 2.3. *Let $\mathbb{A}$ be a commutative and alternative real finite-dimensional algebra containing a copy of the complex numbers, say, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$, where $I \in \mathbb{A}$ is such that $I^2 = -1$. Then every nonconstant polynomial with coefficients in $\mathbb{A}$ and with an invertible leading coefficient has a singular element in $\mathbb{C}_I$.*

Motivated by the preceding theorems and the main result of [4], here are extensions of the fundamental theorem of algebra for noncommutative polynomials in the settings of real and complex locally convex associative algebras.

Theorem 2.4. (i) *Let $\mathbb{A}$ be a unital locally convex associative real algebra containing a copy of the complex numbers, namely, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$ with $I \in \mathbb{A}$ and $I^2 = -1$, such that $IA = AI$ for all $A \in \mathbb{A}$. Then every nonconstant and noncommutative polynomial with coefficients in $\mathbb{A}$ whose leading noncommutative polynomial part evaluated at some $A_0 \in \mathbb{A}$ is invertible has a singular element in $\mathbb{C}_I A_0$.*

(ii) *Let $\mathbb{A}$ be a locally convex unital complex associative algebra. Then every nonconstant and noncommutative polynomial with coefficients in $\mathbb{A}$ whose leading noncommutative polynomial part evaluated at some $A_0 \in \mathbb{A}$ is invertible has a singular element in $\mathbb{C}A_0$.*

Remarks. 1. Just as in Theorem 2.3, when the algebras are finite-dimensional, the associativity hypothesis on the algebras can be replaced by a weaker hypothesis, namely the algebras being alternative as opposed to being associative (and locally convex).

2. Motivated by the theorem we suggest the following conjecture. *Let $\mathbb{A}$ be a unital locally convex associative real algebra containing a copy of the complex numbers, namely, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$ with $I \in \mathbb{A}$ and $I^2 = -1$, and $A_0 \in \mathbb{A}$. Then every nonconstant and noncommutative polynomial with coefficients in $\mathbb{A}$ whose leading noncommutative polynomial part evaluated at every $zA_0 \in \mathbb{A}$ (resp. $A_0 z \in \mathbb{A}$) with $z = a + bI \in \mathbb{C}_I$ and $a^2 + b^2 = 1$ is invertible has a singular element in $\mathbb{C}_I A_0$ (resp. $A_0 \mathbb{C}_I$).*

With the preceding theorem at our disposal, the following corollary is immediate.

Theorem 2.5. *Let $n \in \mathbb{N}$ and $\mathbb{A}$ be an arbitrary finite-dimensional real or complex unital algebra. If the algebra $\mathbb{A}$ happens to be a real algebra, assume further that $\mathbb{A}$ contains a copy of the complex numbers, namely, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$ with $I \in \mathbb{A}$ and $I^2 = -1$, such that $IA = AI$ for all $A \in \mathbb{A}$ and that $A(BC) = (AB)C$ for all $A, B, C \in \mathbb{A}$ with $I \in \{A, B, C\}$. Then every element of $M_n(\mathbb{A})$ has eigenvalues in $\mathbb{C}_I$ or in $\mathbb{C}$ depending on whether $\mathbb{A}$ is a real or a complex algebra.*

With a method similar to that used in the proof of Theorem 2.2, we can prove a result on the existence of right eigenvalues for matrices with entries in finite-dimensional associative real algebras containing a copy of the complex numbers.

Theorem 2.6. *Let $n \in \mathbb{N}$ and $\mathbb{A}$ be an associative finite-dimensional real algebra containing a copy of the complex numbers, namely, $\mathbb{C}_I := \{a + bI : a, b \in \mathbb{R}\}$ with $I \in \mathbb{A}$ and $I^2 = -1$. Then every $A \in M_n(\mathbb{A})$ has right eigenvalues in $\mathbb{C}_I$. In particular, every quaternion matrix has right eigenvalues in any copy of the complex numbers within the quaternions.*

References

[1] A.A. Albert, Absolute valued real algebras. Ann. of Math. 48 (1947), 495-501.

[2] F.F. Bonsall and J. Duncan, *Complete Normed Algebras*, Ergeb. Math. Grenzgeb. 80, Springer, Berlin, 1973.

[3] H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel, and R. Remmert, *Numbers*, Graduate Texts in Mathematics, Readings in Mathematics, Springer-Verlag, New York, 1991.

[4] S. Eilenberg and I. Niven, The "Fundamental Theorem of Algebra" for quaternions, Bulletin of the American Mathematical Society. 50 (4): 246-248 (April 1944).

[5] B. Fine and G. Rosenberger, *The Fundamental Theorem of Algebra*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997.

[6] M.C. García and Á.R. Palacios, *Non-Associative Normed Algebras, Vol. I: The Vidav-Palmer and Gelfand-Naimark Theorems*, Cambridge University Press, Cambridge, 2014.

[7] C.E. Rickart, *General Theory of Banach Algebras*, The University Series in Higher Mathematics, D. van Nostrand Co., Inc., Princeton, NJ, 1960

[8] W. Rudin, *Functional Analysis*, 2nd edition, McGraw-Hill, Inc., New York, 1991.

[9] R.D. Schafer. *An Introduction to Nonassociative Algebras*, Academic Press, New York, 1966.

[10] Dariusz M. Wilczyński, On the fundamental theorem of algebra for polynomial equations over real composition algebras, Journal of Pure and Applied Algebra 218 (2014) 1195-1205.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some properties of a special companion matrices and their powers

A.M. Nazari* and S. Asghari

Department of Mathematics, Arak University , Arak, Iran

Abstract

In this paper, we will discuss the properties of an interesting companion matrix that is widely used in k-circulant matrices. Finding eigenvalues, singular values, qr factorization, and many other interesting properties for this matrix and its integer power are considered.

Keywords: Companion matrix, k-ciculant matrix, qr factorization, Singular value decomposision.

Mathematics Subject Classification [2010]: 15A18, 15A60, 93B10

1 Introduction

Some of the properties of the circulant-like matrix are given in [1]. This matrix has the following form

$$Q = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{n\times n}, \tag{1}$$

for $k \in \mathbb{C} - \{0\}$. This matrix is a sparse, invertible and upper companion matrix. The k-circulant matrix is defined in [2] as follows:

$$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ ka_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ ka_{n-2} & ka_{n-1} & a_0 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ ka_1 & ka_2 & ka_3 & \cdots & a_0 \end{bmatrix}.$$

*Speaker. Email address: a-nazari@araku.ac.ir

When $k = 1$, A becomes a circulant matrix. The matrix Q is a k-circulant and since the sum and product of the two k-circulant matrices also are k-circulant then $A + Q$ and AQ are k-circulant. Although the multiplication of two circulant matrices has commutative property, unfortunately this property does not hold for two k-circulant matrices. If A is a k-circulant matrix, then we have

$$AQ = QA,$$

and by simple induction for two arbitrary positive or negative integers numbers m_1 and m_2 we will have the following relationship

$$A^{m_1}Q^{m_2} = Q^{m_2}A^{m_1}.$$

It is also shown in [2] that the matrix A can be produced as follows using the matrix Q and its powers. i.e.

$$A = \sum_{i=0}^{n-1} a_i Q^i.$$

2 The powers of Q and its properties

In this section, we provide the interesting properties of matrix Q. It is eeasy to see that

$$Q^2 = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & k & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad Q^3 = \begin{bmatrix} 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & k & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & k & 0 & \cdots & 0 & 0 \end{bmatrix},$$

and then $Q^n = \text{diag}(k, k, \cdots, k)$. Also we have

$$Q^{n+1} = \begin{bmatrix} 0 & k & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & k & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & k & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & k \\ k^2 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad Q^{n+2} = \begin{bmatrix} 0 & 0 & k & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & k \\ k^2 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & k^2 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix},$$

and hence $Q^{2n} = \text{diag}(k^2, k^2, \cdots, k^2)$, therefore for all integer and positive m we have $Q^{mn} = \text{diag}(k^m, k^m, \cdots, k^m)$. It is well know that in the qr factorization of a companion matrix, the q-matrix will have the same form as the companion matrix. The qr

factorization of $Q = qr$ will be calculated as follows:

$$q = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad r = \begin{bmatrix} k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{2}$$

Also the LU factorization of this matrix is very similar to its qr factorization, and only instead of the entry $(n, 1)$ of the q matrix, we have to put 1 to obtain the U matrix, and instead of the entry $(1, 1)$ of the r matrix, we have to put k to get the L matrix be achieved.

The inverse of this matrix(Q) can be easily calculated and it will be a lower companion matrix as shown below:

$$Q^{-1} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

and its exponents can be easily calculated just like the exponents of the matrix Q, for instance we have

$$Q^{-2} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{k} \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix}, \qquad \cdots, \qquad Q^{-n} = \operatorname{diag}\left(\frac{1}{k}, \cdots, \frac{1}{k}\right). \tag{3}$$

Since Q is a upper companion matrix, its characteristic polynomial coefficients are in the first column, so its characteristic polynomial is:

$$P(\lambda) = \lambda^n - k,$$

Therefore, all the n-th roots of k are the eigenvalues of the matrix Q. Also the singular

values of $Q, Q^2, Q^3, \cdots, Q^n, Q^{n+1}, Q^{n+2}, \cdots$ respectively are

$$\begin{aligned}
\sigma_1 &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}} & 1 & 1 & \cdots & 1 \end{array}\right], \\
\sigma_2 &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & 1 & \cdots & 1 \end{array}\right], \\
\sigma_3 &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & \cdots & 1 \end{array}\right], \\
&\vdots \\
\sigma_n &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & \cdots & \sqrt{k\,\overline{k}} \end{array}\right], \\
\sigma_{n+1} &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}^2} & \sqrt{k\,\overline{k}} & \sqrt{k\,\overline{k}} & \cdots & \sqrt{k\,\overline{k}} \end{array}\right], \\
\sigma_{n+2} &= \left[\begin{array}{ccccc} \sqrt{k\,\overline{k}^2} & \sqrt{k\,\overline{k}^2} & \sqrt{k\,\overline{k}} & \cdots & \sqrt{k\,\overline{k}} \end{array}\right], \\
&\vdots
\end{aligned}$$

respectively.

Another important companion matrix similar matrix Q in (1) introduce as follows:

$$Q(k,m) = \begin{bmatrix} 0 & m & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{n\times n}, \tag{4}$$

then the charactristic polynomial of this matrix equal $P(Q(k,m),\lambda) = \lambda^n - km$, so its eigenvalues very easy are calculated.

$$Q^2(k,m) = \begin{bmatrix} 0 & 0 & m & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ & & & & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & km & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \qquad Q^3(k,m) = \begin{bmatrix} 0 & 0 & 0 & m & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ & & & & & & \\ k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & km & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & km & 0 & \cdots & 0 & 0 \end{bmatrix},$$

and therefor $Q^n = \mathrm{diag}(km, km, \cdots, km)$. Similar (3) we can obtain the inverse of matrix $Q(k,m)$ and its powers as follows:

$$Q^{-1}(k,m) = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} \\ \frac{1}{m} & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix},$$

$$Q^{-2}(k,m)=\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{k} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \frac{1}{km} \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix},\cdots,Q^{-n}(k,m)=\operatorname{diag}\left(\frac{1}{km},\cdots,\frac{1}{km}\right). \quad (5)$$

The following matrix introduce in [1]:

$$A(k,m)=\begin{bmatrix} a_0 & a_1m & a_2m & \cdots & a_{n-1}m \\ ka_{n-1} & a_0m & a_1 & \cdots & a_{n-2} \\ ka_{n-2} & ka_{n-1}m & a_0m & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ ka_1 & ka_2m & ka_3m & \cdots & a_0 \end{bmatrix}.$$

Theorem 2.1. *For two arbitrary positive or negative integers numbers m_1 and m_2 we will have the following relationship*

$$A(k,m)^{m_1}Q(k,m)^{m_2}=Q(k,m)^{m_2}A(k,m)^{m_1}.$$

Proof. In [1] this is proved that $A(k,m)Q(k,m)=Q(k,m)A(k,m)$. Now by induction on m_1 we have

$$A(k,m)^{m_1}Q(k,m)=A(k,m)A(k,m)^{m_1-1}Q(k,m)=A(k,m)Q(k,m)A(k,m)^{m_1-1}=Q(k,m)A(k,m)A(k,m)^{m_1-1}=Q(k,m)A(k,m)^{m_1}.$$

And again we take the induction on m_2 and obtaine $A(k,m)Q(k,m)^{m_2}=Q(k,m)^{m_2}A(k,m)$ and then the Theorem will be proved similarly. □

Similar (2) we can find the qr-factorization of matrix $Q(k,m)$ as follows:

$$q=\begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix},\qquad r=\begin{bmatrix} k & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & m & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}. \quad (6)$$

3 Squar root

The squar root of a matrix Q is a matrix P such that $P^2=Q$. If Q is diagonalizable, we have $P=VEV^{-1}$, where V is a matrix, rows of which are eigenvectors of Q and also E is diagonal matrix, and the entries on its diagonal are square root of eigenvalues of Q [3]. This means $P^2=VEV^{-1}VEV^{-1}=VEV^{-1}=Q$. Since $Q(k,m)$ also is diagonalizable with similar method we can calculate its square root.

Example 3.1. Find the squar foot of the follwing matrix :

$$Q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 4 & 0 & 0 & 0 \end{bmatrix}.$$

Solution. The eigenvalues of the Q matrix are as follows:

$$\begin{bmatrix} \sqrt[4]{2} & -\sqrt[4]{2} & i\sqrt[4]{2} & -i\sqrt[4]{2} \end{bmatrix}.$$

By finding the corresponding eigenvectors of the above set, the eigenvector matrix has the following form:

$$V = \begin{bmatrix} i/4\sqrt{2} & -i/4\sqrt{2} & 1/4\sqrt{2} & -1/4\sqrt{2} \\ -1/2 & -1/2 & 1/2 & 1/2 \\ -i/2\sqrt{2} & i/2\sqrt{2} & 1/2\sqrt{2} & -1/2\sqrt{2} \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Then we have

$$\sqrt{E} = \begin{bmatrix} \sqrt{i\sqrt{2}} & 0 & 0 & 0 \\ 0 & \sqrt{-i\sqrt{2}} & 0 & 0 \\ 0 & 0 & \sqrt[4]{2} & 0 \\ 0 & 0 & 0 & \sqrt{-\sqrt{2}} \end{bmatrix}.$$

Therefore $P = V\sqrt{E}V^{-1} =$

$$\begin{bmatrix} 1/4\,2^{3/4} + 1/4\,\sqrt[4]{2} + i/4\,\sqrt[4]{2} & 1/4\,\sqrt[4]{2} + 1/8\,2^{3/4} - i/8 2^{3/4} & -1/8\,2^{3/4} + 1/8\,\sqrt[4]{2} + i/8\,\sqrt[4]{2} & -1/8\,\sqrt[4]{2} + 1/16\,2^{3/4} - i/16 2^{3/4} \\ -1/2\,\sqrt[4]{2} + 1/4\,2^{3/4} - i/4 2^{3/4} & 1/4\,2^{3/4} + 1/4\,\sqrt[4]{2} + i/4\,\sqrt[4]{2} & 1/4\,\sqrt[4]{2} + 1/8\,2^{3/4} - i/8 2^{3/4} & -1/8\,2^{3/4} + 1/8\,\sqrt[4]{2} + i/8\,\sqrt[4]{2} \\ -1/2\,2^{3/4} + 1/2\,\sqrt[4]{2} + i/2\,\sqrt[4]{2} & -1/2\,\sqrt[4]{2} + 1/4\,2^{3/4} - i/4 2^{3/4} & 1/4\,2^{3/4} + 1/4\,\sqrt[4]{2} + i/4\,\sqrt[4]{2} & 1/4\,\sqrt[4]{2} + 1/8\,2^{3/4} - i/8 2^{3/4} \\ \sqrt[4]{2} + 1/2\,2^{3/4} - i/2 2^{3/4} & -1/2\,2^{3/4} + 1/2\,\sqrt[4]{2} + i/2\,\sqrt[4]{2} & -1/2\,\sqrt[4]{2} + 1/4\,2^{3/4} - i/4 2^{3/4} & 1/4\,2^{3/4} + 1/4\,\sqrt[4]{2} + i/4\,\sqrt[4]{2} \end{bmatrix}.$$

It is easy te see that $P^2 = Q$, then P is squar root of Q.

References

[1] E. Andrade, D. Carrasco-Olivera, C. Manzaneda, On circulant like matrices properties involving Horadam, Fibonacci, Jacobsthal and Pell numbers, Linear Algebra and its Applications, 617 (2021) 100-120.

[2] R. E. Cline, R. J. Plemmons, G. Worm, Generalized inverses of certain Toeplitz matrices, Linear Algebra and its Applications, 8 (1974) 25-33.

[3] A. Nazari, H. Fereydooni, and M. Bayat, A manual approach for calculating the root of square matrix of dimension ≤3. Math Sci 7, 44 (2013).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Matrix representation of multilinear mappings

Mohsen Kian*

Department of Mathematics, University of Bojnord, Bojnord, Iran

Abstract

Let $\mathbb{M}_n$ be the algebra of all n by n complex matrices. The well-known Jamiolkowski isomorphism gives a bijection between the space of linear maps (from $\mathbb{M}_n$ to $\mathbb{M}_k$) and the matrix algebra $\mathbb{M}_n \otimes \mathbb{M}_k$. We investigate this isomorphism in the case of multilinear mappings between matrix algebras.

Keywords: Multilinear mapping, matrix representation, positive mapping)
Mathematics Subject Classification [2010]: 15A69, 47C15

1 Introduction

Throughout the paper, assume that $\mathbb{M}_n$ is the algebra of all n by n matrices with complex entries. A Hermitian matrix A is called positive semi-definite (positive definite) if all of its eigenvalues are non-negative (positive). The set of all positive semi-definite (positive definite) matrices in $\mathbb{M}_n$ is denoted by $\mathbb{P}_n$ ($\mathbb{P}_n^+$). It is known that the dual space of $\mathbb{M}_n$ is identified with itself [1] via the duality

$$\langle A, B\rangle = \mathrm{tr}(A^*B), \qquad A, B \in \mathbb{M}_n.$$

Let $\mathcal{L}(k, m)$ be the space of all linear mappings from $\mathbb{M}_k$ to $\mathbb{M}_m$. It is known that $\mathcal{L}(k, m)$ is identified by the matrix algebra $\mathbb{M}_k(\mathbb{M}_m)$ of block matrices.

Every $\Phi \in \mathcal{L}(k, m)$ has a matrix representation (see [2]) defined by

$$A_\Phi = \sum_{i,j} E_{ij}\Phi(E_{ij})$$

in which $\{E_{ij}\}$ is the canonical set of matrix units for $\mathbb{M}_k$. In fact, there is a one to one correspondence between $\mathcal{L}(k, m)$ and $\mathbb{M}_k(\mathbb{M}_m)$. Every block matrix $A \in \mathbb{M}_k(\mathbb{M}_m)$ produces a linear map from $\mathbb{M}_k$ into $\mathbb{M}_m$. This correspondence is known as the Jamiolkowski isomorphism.

A mapping $\Phi \in \mathcal{L}(k, m)$ is said to be positive if $\Phi(\mathbb{P}_k) \subseteq \mathbb{P}_m$. It is known that positive linear mappings are automatically continuous. See [4] in the case of positive non-linear mappings. The mapping Φ induces a linear mapping $\Phi_n : \mathbb{M}_n(\mathbb{M}_k) \to \mathbb{M}_n(\mathbb{M}_m)$ for every $n \in \mathbb{N}$, by $\Phi_n([A_{ij}] = [\Phi(A_{ij})]$. If Φ_n is positive for every $n \in \mathbb{N}$, then Φ is called completely positive.

A famous result regarding the matrix representation of linear mappings states that A mapping $\Phi \in \mathcal{L}(k, m)$ is completely positive if and only if its matrix representation is positive definite.

*Speaker. Email address: kian@ub.ac.ir

2 Main results

Assume that $\Phi : \mathbb{M}_{k_1} \times \mathbb{M}_{k_2} \times \cdots \times \mathbb{M}_{k_p} \to \mathbb{M}_m$ is a multilinear map. We say that Φ is positive, if

$$A_i \in \mathbb{P}_{k_i} \ \ (i = 1, \ldots, p) \qquad \Longrightarrow \qquad \Phi(A_1, \ldots, A_p) \in \mathbb{P}_m.$$

Typical example of positive multilinear mappings are tensor product of matrices, $(A_1, \ldots, A_p) \mapsto A_1 \otimes \cdots \otimes A_p$.

For more examples and basic facts concerning positive multilinear mappings, see [5].

Here, we are interested in matrix representations for multilinear mappings. For a multilinear map

$$\Phi : \mathbb{M}_{k_1} \times \mathbb{M}_{k_2} \times \cdots \times \mathbb{M}_{k_p} \to \mathbb{M}_m$$

we define

$$A_\Phi = \sum_{i_1, j_1 = 1}^{k_1} \cdots \sum_{i_p, j_p = 1}^{k_p} (E^1_{i_1 j_1} \otimes \cdots \otimes E^p_{i_p j_p}) \otimes \varphi(E^1_{i_1 j_1}, \ldots, E^p_{i_p j_p})$$

in which $\{E^\ell_{i_\ell j_\ell}\}$ is the standard set of matrix units in $\mathbb{M}_{k_\ell}$.

The matrix representation of Φ depends on the choice of the set of matrix units. Changing the basis affects on it as follows.

Lemma 2.1. *Let $\mathbf{A}_\Phi$ be matrix representation of a multilinear map*

$$\Phi : \mathbb{M}_{k_1} \times \mathbb{M}_{k_2} \times \cdots \times \mathbb{M}_{k_p} \to \mathbb{M}_m$$

with respect to the system of matrix units $\{E^\ell_{i_\ell j_\ell}\} \subseteq \mathbb{M}_{k_\ell}$ $(\ell = 1, \ldots, p)$. If U_ℓ is the transition matrix of $\{E^\ell_{i_\ell j_\ell}\}$ to a new basis $\{F^\ell_{i_\ell j_\ell}\}$, then

$$\tilde{\mathbf{A}_\varphi} = (\pi_1^* \otimes \cdots \otimes \pi_p^*) \otimes \mathrm{id}_m(\mathbf{A}_{\varphi \circ (\pi_1, \ldots, \pi_p)})$$

is the matrix representation of φ with respect to the basis $\{F_{i_\ell j_\ell}\}$, where each π_ℓ is the unitary congruence via U_ℓ.

It is known that the building terms of completely positive linear maps between matrix algebras are of the form $A \mapsto V^*AV$ for some linear mapping V. In the next result, we give the matrix representation of such mappings in the multilinear setting.

Theorem 2.2. *Let $V : \mathbb{C}^m \to \mathbb{C}^{nk}$ be a linear operator and let $\Phi : \mathbb{M}_n \times \mathbb{M}_k \to \mathbb{M}_m$ be defined by $\Phi(A, B) = V^*(A \otimes B)V$. Let $\{e_i\}_{1 \le i \le n}$, $\{f_i\}_{1 \le i \le k}$ and $\{g_i\}_{1 \le i \le m}$ are orthonormal basis for $\mathbb{C}^n$, $\mathbb{C}^k$ and $\mathbb{C}^m$, respectively and let $\{E_{ij}\}_{1 \le i,j \le n}$, $\{F_{ij}\}_{1 \le i,j \le k}$ and $\{G_{ij}\}_{1 \le i,j \le m}$ are corresponding system of matrix units in $\mathbb{M}_n$, $\mathbb{M}_k$ and $\mathbb{M}_m$. Then*

$$\mathbf{A}_\Phi = \sum_{i,j,t,s,\ell,u} \overline{v}_{it\ell} v_{jsu} (E_{ij} \otimes F_{ts} \otimes G_{\ell u})$$

in which v_{ijr} is the coordinate of Vg_r in the basis of $\mathbb{C}^{nk}$ made by $\{e_i\}_{1 \le i \le n}$ and $\{f_j\}_{1 \le j \le k}$.

Example 2.3. Let $\Phi : \mathbb{M}_2 \times \mathbb{M}_2 \to \mathbb{M}_2$ be defined by $\Phi(A, B) = A \circ B$, the Hadamard product. Then

$$\mathbf{A}_\Phi = \sum_{i,j=1}^{2} \sum_{r,s=1}^{2} (E_{ij} \otimes E_{rs}) \otimes (E_{ij} \circ E_{rs})$$

$$= \sum_{i,j=1}^{2} (E_{ij} \otimes E_{ij}) \otimes E_{ij} = \sum_{i,j=1}^{2} E_{ij}^{\otimes 3}$$
$$= \begin{bmatrix} E_{11} & O & O & E_{12} \\ O & O & O & O \\ O & O & O & O \\ E_{21} & O & O & E_{22} \end{bmatrix},$$

in which O is the 2×2 zero matrix.

References

[1] R. Bhatia, *Matrix analysis*, Springer-Verlage, New York, 1997.

[2] E. Christensen and A. Sinclair, Representations of completely bounded multilinear operators, *J. Funct. Anal.*, 72 (1987), 151–181.

[3] A. Dadkhah, M. Kian, and M.S. Moslehian, Decomposition of tracial positive maps and applications in quantum information, arXiv:2202.12798.

[4] A. Dadkhah, M. S. Moslehian, and M. Kian, Continuity of positive non-linear maps between C^*-algebras, *Studia Math.*, 263 (2022), 241–265.

[5] M. Dehghani, M. Kian, and Y. Seo, Developed matrix inequalities via positive multilinear mappings, *Linear Algebra Appl.*, 484 (2015), 63–85.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Nonlinear maps preserving the mixed product

Leila Abedini* and Ali Taghavi

Department of Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, P. O. Box 47416-1468, Babolsar, Iran

Abstract

Let $\mathcal{A}$ and $\mathcal{B}$ be two von Neumann algebras. For $A, B \in \mathcal{A}$, define by $A \bullet B = A^*B + B^*A$ and $A \circ B = A^*B - B^*A$ the new products of A and B. Suppose that a bijective map $\Phi : \mathcal{A} \longrightarrow \mathcal{B}$ satisfies $\Phi(A \bullet B \circ C) = \Phi(A) \bullet \Phi(B) \circ \Phi(C)$ for all $A, B, C \in \mathcal{A}$. In this paper, it is proved that if $\mathcal{A}$ and $\mathcal{B}$ be two von Neumann algebras with no central abelian projections, then the map $\Phi(I)\Phi$ is sum of a $*$-isomorphism and a conjugate linear $*$-isomorphism, where $\Phi(I)$ is a self-adjoint central element in $\mathcal{B}$ with $\Phi(I)^2 = I$.

Keywords: Jordan $*$- product, isomorphism, von Neumann algebras.
Mathematics Subject Classification [2010]: 46J10, 47B48

1 Introduction

Let $\mathcal{A}$ be a $*$-algebra over the complex field $\mathbb{C}$. For $A, B \in \mathcal{A}$, define, the Jordan $*$-product of A and B by $A \blacklozenge B = AB + BA^*$, the Jordan product of A and B by $A \lozenge B = AB + BA$, the skew Lie product of A and B by $[A, B]_* = AB - BA^*$, $A \bullet B = A^*B + B^*A$ and $A \circ B = A^*B - B^*A$, which are different kinds of new products. These products have recently attracted many *author's* attention(for example, see ([2], [3]). Liu and Ji [9] proved that a bijective map Φ on factor von Neumann algebras preserves, $A \bullet B = A^*B + B^*A$ if and only if Φ be a $*$-isomorphism. Recently C. Li, F. Zhao, Q. Chen [10] discussed some bijective maps preserving the new product $A \bullet B = A^*B + B^*A$ between von Neumann algebras with no central abelian projections. In fact, it is shown that a bijective map $\Phi : \mathcal{A} \longrightarrow \mathcal{B}$ satisfies $\Phi(A^*B + B^*A) = \Phi(A)^*\Phi(B) + \Phi(B)^*\Phi(A)$ for all $A, B \in \mathcal{A}$. Then Φ is a sum of a linear a $*$-isomorphism and a conjugate linear a $*$-isomorphism. Recently, nonlinear maps preserving the products of the mixture of (skew) Lie product and Jordan $*$- product have received a fair amount of attention (see [4], [5], [7], [8], [12], [13]). For example, C. Li et al. studied the nonlinear maps preserving the skew Lie triple product $[[A, B]_*, C]_*$ (see [5], [8]) and the jordan triple $*$-product $A \bullet B \bullet C$ (see [7], [13]) on von Neumann algebras. Z. Yang and J.Zhang in ([12]) studied the nonlinear maps preserving the mixed skew Lie triple product $[[A, B]_*, C]$ and $[[A, B], C]_*$ on factor von Neumann algebras. Changjing Li, Yuanyuan Zhao, Fangfang Zhao studied the nonlinear maps preserving the mixed product $[A \bullet B, C]_*$ on von Neumann algebras (see [6]). Dongfang Zhang, Changjing Li, Yuanyuan

*Speaker. Email address: lelaabediny@gmail.com

Zhao, studied the nonlinear maps preserving mixed Jordan triple products $A\blacklozenge B\diamondsuit C$ on von Neumann algebras.

In the present paper, we will establish the structure of the nonlinear maps preserving the mixed product $(A \bullet B \circ C)$ on von Neumann algebras.

2 Main results

Lemma 2.1. Let $\mathcal{A}$ be a von Neumann algebra with no central abelian projections. Then there exists a projection $P \in \mathcal{A}$ such that $\underline{P} = 0$ and $\overline{P} = I$.

Lemma 2.2. Let $\mathcal{A}$ be a von Neumann algebra on a Hilbert space H. Let A be an operator in $\mathcal{A}$ and $P \in \mathcal{A}$ is a projection with $\overline{P} = I$. If $ABP = 0$ for all $B \in \mathcal{A}$, then $A = 0$.

Theorem 2.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be two von Neumann algebras with no central abelian projections. Suppose that a bijective map $\Phi : \mathcal{A} \longrightarrow \mathcal{B}$ satisfies $\Phi(A\bullet B\circ C) = \Phi(A)\bullet\Phi(B)\circ\Phi(C)$ for all $A, B, C \in \mathcal{A}$. Then the map $\Phi(I)\Phi$ is sum of a linear $*$-isomorphism and a conjugate linear $*$-isomorphism, where $\Phi(I)$ is a self-adjoint central element in $\mathcal{B}$ with $\Phi(I)^2 = I$.*

Proof. We organize the proof in a series of steps.

Step 1. $\Phi(0) = 0$.

Step 2. $\Phi(\frac{I}{2}) = \Phi(\frac{I}{2})^* \in Z(B)$.

Step 3. i) Φ preserves the self-adjoint and skew self-adjoint elements in both direction.
ii) $\Phi(\frac{I}{2})^2 = \frac{I}{4}$.

Step 4. $\Phi(\frac{iI}{2})^* = -\Phi(\frac{iI}{2}) \in Z(B)$.

Step 5. $\Phi(\frac{iI}{2})^2 = -\frac{I}{4}$.

Step 6. For every $A \in \mathcal{A}_s$, $\Phi(iA) = 4\Phi(\frac{I}{2})\Phi(\frac{iI}{2})\Phi(A)$.

Step 7. For every $A,\ B \in \mathcal{A}_s$, we have

$$\Phi(A_{11} + \boldsymbol{B}_{12}) = \Phi(A_{11}) + \Phi(\boldsymbol{B}_{12}).$$

and

$$\Phi(\boldsymbol{B}_{12} + A_{22}) = \Phi(\boldsymbol{B}_{12}) + \Phi(A_{22}).$$

Step 8. For every $A,\ B,\ C \in \mathcal{A}_s$, we have

$$\Phi(A_{11} + \boldsymbol{B}_{12} + C_{22}) = \Phi(A_{11}) + \Phi(\boldsymbol{B}_{12}) + \Phi(C_{22}).$$

Step 9. For every $A,\ B \in \mathcal{A},\ 1 \leq j \neq k \leq 2$, we have

$$\Phi(\boldsymbol{A}_{jk} + \boldsymbol{B}_{jk}) = \Phi(\boldsymbol{A}_{jk}) + \Phi(\boldsymbol{B}_{jk}).$$

Step 10. For every $A, B \in \mathcal{A}_s$, $\Phi(A_{jj} + B_{jj}) = \Phi(A_{jj}) + \Phi(B_{jj})$.

Step 11. Φ is additive on $\mathcal{A}_s$.

Step 12. Φ is additive on $\mathcal{A}_{sk}$.

Step 13. Φ is additive on $\mathcal{A}$.

Step 14. For every $A \in \mathcal{A}_s$, $\Psi(iA) = \Psi(iI)\Psi(A)$.

Step 16. $\Psi(A^*B + B^*A) = \Psi(A)^*\Psi(B) + \Psi(B)^*\Psi(A)$, for all $A, B \in \mathcal{A}$.

Now, by the Theorem 2.2 of [11], we have the map $\Psi = \Phi(I)\Phi$ is sum of a linear $*$-isomorphism and a conjugate linear $*$-isomorphism.

□

References

[1] Abedini, L. Taghavi, A. (2022) *Nonlinear maps preserving the mixed product $A \bullet B \circ C$ on von Neumann algebras*, Rocky Mountain Journal of Mathematics. (Accepted).

[2] Z. Bai, S. Du,*Maps preserving product $XY - YX^*$ on von Neuman algebras*, J. Math. Anal. Appl. 386(2012)103-109.

[3] J. Cui, C. K. Li, *Maps preserving product $XY - YX^*$ on factor von Neuman algebras*, Linear Algebra Appl. 431(2009) 833- 842.

[4] D. Huo, B. Zheng and H. Liu,*Nonlinear maps preserving Jordan triple η-$*$-Products*, J. Math. Anal. Appl. 430(2015)830-844.

[5] C. Li, Q. Chen, T. Wang, *Nonlinear maps preserving Jordan triple $*$- product on factors*, chin. Ann. Math. Ser. B 39(2018)633-642.

[6] C. Li, Yu. Z and Fang. Z, *Nonlinear maps preserving the mixed product $[A \bullet B, C]_*$ on von Neumann algebras*, Filomat, 35:8(2021)2775-2781.

[7] C. Li, F. Lu, *Nonlinear mappings preserving the Jordan triple 1-$*$-Product, on von Neuman algebras*, Complex Anal. Oper. Theory. 11(2017)109-117.

[8] C. Li, F. Lu, *Nonlinear mappings preserving the Jordan triple $*$-Products, on von Neuman algebras*, Ann. Funct. Anal. 7(2016)496-507.

[9] L. Liu, G. X. Ji, *Maps preserving product $X^*Y + YX^*$ on factor von Neuman algebra*, Linear and Multilinear Algebra. 59 (2011), 951-955.

[10] C. Li, F. Zhao, Q. Chen, *Nonlinear Maps preserving product $X^*Y + YX^*$ on von Neuman Algebras*, Bull. Iran. Math. Soc. 44 (33)(2018) 729738.

[11] A. Taghavi, S. gholampoor, *Maps preserving product $A^*B + B^*A$ on C^*-algebras*, Bulletin of the Iranian Mathematical society, 41(2015), No. 7, 85-98.

[12] Z. Yang, J. Zhang, *Nonlinear maps preserving the mixed skew Lie triple product on factor von Neumann algebras*, Ann. Funct. Anal. 10(2019)325-336.

[13] F. Zhao, C. Li, *Nonlinear maps preserving the Jordan triple $*$-Products between factors*, Indag. Math. 29(2018)619-627.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On completely preserving maps

Roja Hosseinzadeh*

Department of Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, P. O. Box 47416-1468, Babolsar, Iran

Abstract

Let $\mathcal{A}$ and $\mathcal{B}$ be two standard operator algebra on Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively. In this paper, we determine the forms of the surjective maps from $\mathcal{A}$ onto $\mathcal{B}$ such that completely preserve zero triple Jordan product in both directions.

Keywords: Preserver problem, Standard operator algebra, Triple Jordan product
Mathematics Subject Classification [2010]: 46J10, 47B48

1 Introduction

Mappings that preserve a certain property is a subject that has attracted the attention of many mathematicians and they seek to obtain other properties of these maps as well as their forms. In the field of preserving problems, the issue of maps that completely preserve a specific property has recently been taken into consideration. For example, you can see papers [1-4].

Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $\mathcal{B}(\mathcal{X})$ denote the Banach algebra of all bounded linear operators on $\mathcal{X}$. Let $\mathcal{S} \subseteq B(X)$ and $\mathcal{T} \subseteq B(Y)$ be linear subspaces and $\phi : \mathcal{S} \longrightarrow \mathcal{T}$ be a map. Define for each $n \in \mathbb{N}$, a map $\phi_n : \mathcal{S} \otimes \mathcal{M}_n(\mathbb{F}) \longrightarrow \mathcal{T} \otimes \mathcal{M}_n(\mathbb{F})$ by

$$\phi_n((s_{ij})_{n\times n}) = (\phi(s_{ij}))_{n\times n} \qquad (\forall s_{ij} \in \mathcal{S}).$$

Let (p) be a property. We say that ϕ preserves $n-(p)$, whenever ϕ_n preserves (p) and ϕ completely preserves (p), whenever ϕ_n preserves (p) for each n.

Recall that a standard operator algebra on $\mathcal{X}$ is a norm closed subalgebra of $\mathcal{B}(\mathcal{X})$ which contains the identity and all finite rank operators. Let $\mathcal{A}$ and $\mathcal{B}$ be standard operator algebras on Banach spaces X and Y, respectively. Recently in [3] completely idempotent and completely square-zero preserving maps and in [4] completely commutativity and completely Jordan zero product preserving maps are discussed. Let $\phi : \mathcal{A} \longrightarrow \mathcal{B}$ be a map. If for every $A_{ij} \in \mathcal{A}$, $1 \leq i, j \leq n$ we have

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}^2 = 0 \Leftrightarrow \begin{pmatrix} \times\phi(A_{11}) & \cdots & \phi(A_{1n}) \\ \vdots & & \\ \phi(A_{n1}) & \cdots & \phi(A_{nn}) \end{pmatrix}^2 = 0$$

*Speaker. Email address: ro.hosseinzadeh@umz.ac.ir

for each n, then we say that ϕ completely preserves square-zero operators in both directions. If for every $A_{ij} \in \mathcal{A}$, $1 \leq i, j \leq 2$ we have

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^2 = 0 \Leftrightarrow \begin{pmatrix} \phi(A_{11}) & \phi(A_{12}) \\ \phi(A_{21}) & \phi(A_{22}) \end{pmatrix}^2 = 0,$$

then we say that ϕ preserves 2-square-zero operators in both directions.

Triple Jordan product of two operators $A, B \in \mathcal{A}$ is defined as ABA. In this paper, surjective maps from $\mathcal{A}$ onto $\mathcal{B}$ such that completely preserve zero triple Jordan product, are determined.

Let $\phi : \mathcal{A} \longrightarrow \mathcal{B}$ be a map. If for every $A_{ij}, B_{ij} \in \mathcal{A}$, $1 \leq i, j \leq n$ we have

$$\begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \times \begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \\ B_{n1} & \cdots & B_{nn} \end{pmatrix} \times \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} = 0$$

$$\Leftrightarrow \begin{pmatrix} \times\phi(A_{11}) & \cdots & \phi(A_{1n}) \\ \vdots & & \\ \phi(A_{n1}) & \cdots & \phi(A_{nn}) \end{pmatrix} \times \begin{pmatrix} \phi(B_{11}) & \cdots & \phi(B_{1n}) \\ \vdots & & \\ \phi(B_{n1}) & \cdots & \phi(B_{nn}) \end{pmatrix} \times$$

$$\begin{pmatrix} \phi(A_{11}) & \cdots & \phi(A_{1n}) \\ \vdots & & \\ \phi(A_{n1}) & \cdots & \phi(A_{nn}) \end{pmatrix} = 0$$

for each n, then we say that ϕ completely preserves zero triple Jordan product of operators in both directions.

See the following result from [3]. We use from these theorems in the proof of our main results.

Theorem 1.1. *[3] Let $\mathcal{X}$, $\mathcal{Y}$ be infinite dimensional Banach spaces and $\mathcal{A}$ and $\mathcal{B}$ be standard operator algebras on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\phi : \mathcal{A} \to \mathcal{B}$ be a surjective map. Then the following statements are equivalent:*

(1) *ϕ is completely square-zero preserving in both directions.*

(2) *ϕ is 2-square-zero preserving operators in both directions.*

(3) *There exist a bounded invertible linear or (in the complex case) conjugate- linear operator $A : \mathcal{X} \to \mathcal{Y}$ and a scalar c such that*

$$\phi(T) = cATA^{-1},$$

for all $T \in \mathcal{A}$.

Proposition 1.2. *[3] Let $\phi : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_n(\mathbb{F})$ $(n \geq 3)$ be a surjective map, where $\mathbb{F}$ is the real or complex field. Then the following statements are equivalent:*

(1) *ϕ is completely square-zero preserving in both directions.*

(2) *ϕ is 2-square-zero preserving in both directions.*

(3) *There exist an invertible matrix $A \in \mathcal{M}_n$, a scalar c and an automorphism $\tau : \mathbb{F} \to \mathbb{F}$ such that*

$$\phi(T) = cAT_\tau A^{-1},$$

for all $T \in \mathcal{M}_n(\mathbb{F})$. Here $T_\tau = (\tau(t_{ij}))$ for $T = (t_{ij})$.

2 Main results

Main results of this paper are as following.

Theorem 2.1. *Let $\mathcal{X}$, $\mathcal{Y}$ be infinite dimensional Banach spaces and $\mathcal{A}$ and $\mathcal{B}$ be standard operator algebras on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\phi : \mathcal{A} \to \mathcal{B}$ be a bijective map. Then the following statements are equivalent:*

(1) *ϕ is completely preserving zero triple Jordan product in both directions.*

(2) *ϕ is 2-zero triple Jordan product preserving in both directions.*

(3) *There exist a bounded invertible linear or (in the complex case) conjugate- linear operator $A : \mathcal{X} \to \mathcal{Y}$ and a scalar λ such that*

$$\phi(T) = \lambda ATA^{-1},$$

for all $T \in \mathcal{A}$.

Theorem 2.2. *Let $\phi : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_n(\mathbb{F})$ $(n \geq 3)$ be a bijective map, where $\mathbb{F}$ is the real or complex field. Then the following statements are equivalent:*

(1) *ϕ is completely preserving zero triple Jordan product in both directions.*

(2) *ϕ is 2-zero triple Jordan product preserving in both directions.*

(3) *There exist an invertible matrix $A \in \mathcal{M}_n$, a scalar λ and an automorphism $\tau : \mathbb{F} \to \mathbb{F}$ such that*

$$\phi(T) = \lambda AT_\tau A^{-1},$$

for all $T \in \mathcal{M}_n(\mathbb{F})$. Here $T_\tau = (\tau(t_{ij}))$ for $T = (t_{ij})$.

Acknowledgements: This research was financed by a research grant from the University of Mazandaran.

References

[1] R. Hosseinzadeh, *Maps completely preserving the quadratic operators*, Sahand Communications in Mathematical Analysis, 20 (2023), 123-132.

[2] R. Hosseinzadeh, I. Sharifi, A. Taghavi, *Maps completely preserving fixed points and maps completely preserving kernel of operators*, Analysis Mathematica, 44 (2018), 451-459.

[3] J. Hou and L. Huang, *Maps completely preserving idempotents and maps completely preserving square-zero operators*, Israel Journal of Mathematics, 176 (2010), 363-380.

[4] L. Huang, Y. Liu, *Maps completely preserving commutativity and maps completely preserving jordan-zero products*, Israel J. Math. Anal. 462 (2014), 233-249.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Linear preservers of G-matrices on $\mathbb{M}_2$

Ali Armandnejad[1,2,*] and Setareh Golshan[2]

[1]Department of Pure Mathematics, Shahid Bahonar University of Kerman, Kerman, Iran

[2]Department of Mathematics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

Abstract

Let $\mathbf{M}_n$ be the set of all $n \times n$ real matrices. A nonsingular matrix $A \in \mathbf{M}_n$ is called a G-matrix if there exist nonsingular diagonal matrices D_1 and D_2 such that $A^{-T} = D_1 A D_2$, where A^{-T} denotes the transpose of the inverse of A. Let $\mathbb{G}_n$ be the set of all $n \times n$ G-matrices. A linear operator $T : \mathbf{M}_n \to \mathbf{M}_n$ is called a linear preserver of G-matrices if $T(\mathbb{G}_n) \subseteq \mathbb{G}_n$. The purpose of this paper is to find the structure of the linear operator preserving G-matrices on $\mathbf{M}_2$.

Keywords: G-matrices, linear preserver, J-orthogonal matrices
Mathematics Subject Classification [2010]: 15A30, 15B10

1 Introduction

Let $\mathbf{M}_n$ be the set of all $n \times n$ real matrices. A nonsingular matrix $A \in \mathbf{M}_n$ is called a G-matrix if there exist nonsingular diagonal matrices D_1 and D_2 such that $A^{-T} = D_1 A D_2$, where A^{-T} denotes the transpose of the inverse of A. For a survey of the basic properties of G-matrices and connections to other classes of matrices the reader can see [1], [2] and [6], and the references therein. For fixed nonsingular diagonal matrices D_1 and D_2, let the class of $n \times n$ G-matrices

$$\mathbb{G}(D_1, D_2) = \{A \in \mathbf{M}_n : A^{-T} = D_1 A D_2\}.$$

We call such a class of matrices a G-class of matrices. $\mathbb{G}_n$ is the set of all $n \times n$ G-matrix matrices, that is,

$$\mathbb{G}_n = \bigcup_{D_1, D_2} \mathbb{G}(D_1, D_2),$$

for nonsingular diagonal matrices D_1, D_2. Some preliminary results of G-matrices are as follows:

Theorem 1.1. *If A is an $n \times n$ G-matrix and D is an $n \times n$ nonsingular diagonal matrix, then both AD and DA are G-matrices.*

Theorem 1.2. *If A is an $n \times n$ G-matrix and P is an $n \times n$ permutation matrix, then both AP and PA are G-matrices.*

*Speaker. Email address:armandnejad@uk.ac.ir, armandnejad@vru.ac.ir

Theorem 1.3. *A 2×2 matrix is G-matrix if and only if it is nonsingular and has four or two nonzero entries.*

A matrix $J \in \mathbf{M}_n$ is said to be a signature matrix if J is diagonal and its diagonal entries are ± 1. If J is a signature matrix, a nonsingular matrix $A \in \mathbf{M}_n$ is said to be a J-orthogonal matrix if $A^\top JA = J$. Some properties of J-orthogonal matrices were investigated in [4]. For a fixed signature matrix J, $\Gamma_n(J) = \{A \in \mathbf{M}_n : A^\top JA = J\}$. In fact,

$$\Gamma_n(J) = \mathbb{G}(J, J).$$

There are some interesting relations between J-orthogonal and G-class of matrices, see [5]. Also note that when J is I or $-I$, $\Gamma_n(J) = \mathcal{O}_n$, is the set of all $n \times n$ orthogonal matrices.

The inertia matrix of a Hermitian matrix A is the diagonal matrix

$$\text{diag}(1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0),$$

where the number of $1's$, $-1's$, $0's$ is the number of positive, negative, zero eigenvalues, respectively of A. The following theorem shows the relationship between G-matrices and J-orthogonal matrices.

Theorem 1.4. *[6, Theorem 2.2] Let D_1 and D_2 be nonsingular diagonal matrices with the same inertia matrix J. Then there exist permutation matrices P and Q such that*

$$\mathbb{G}(D_1, D_2) = \{|D_1|^{-1/2} P^T AQ |D_2|^{-1/2} : A \in \Gamma_n(J)\}.$$

This characterization shows that $\mathbb{G}(D_1, D_2)$ is in fact nonempty.

A matrix $A \in \mathbf{M}_n$ is called a generalized permutation matrix if $A = PD$, for some permutation matrix P and some nonsingular diagonal matrix D. The set of n-by-n generalized permutation matrices is a subgroup of $GL(n, C)$. Let $\mathbb{GP}_n$ be the set of all $n \times n$ generilized permutation matrices. In fact, $\mathbb{GP}_n$ is the set of $n \times n$ matrices with exactly one nonzero entry in each row and in each column. A linear operator $T : \mathbf{M}_n \to \mathbf{M}_n$ defined by $T(X) = A^T XA$ or $T(X) = A^T X^T A$ for some $A \in \mathbf{M}_n$ is called a standard linear operator on $\mathbf{M}_n$. It is said that a linear operator $T : \mathbf{M}_n \to \mathbf{M}_n$ preserves a set G if $T(G) \subseteq G$. In this paper, we show that if A is a generalized permutation matrix, a standard linear operator $T : \mathbf{M}_n \to \mathbf{M}_n$ defined by $T(X) = A^T XA$ preserves the set of G-matrices. We guess that the opposite is also true but we prove it just for $n = 2$.

2 Main results

In this section we show that if A is a generalized permutation matrix then $A^T \mathbb{G}_n A = \mathbb{G}_n$. For $n = 2$, we show that $A^T \mathbb{G}_2 A = \mathbb{G}_2$ if and only if A is a generalized permutation matrix.

Lemma 2.1. *Let $A \in \mathbf{M}_n$. If $A^T \mathbb{G}_n A \subseteq \mathbb{G}_n$, then A is nonsingular.*

Proposition 2.2. *Let $A \in \mathbf{M}_n$. If A is a generalized permutation matrix then $A^T \mathbb{G}_n A = \mathbb{G}_n$.*

Proof. Since A is a generalized permutation matrix, $A = PD$ for some permutation matrix P and nonsingular diagonal matrix D. By using Theorem 1.1 and Theorem 1.2 it is clear that $A^T \mathbb{G}_n A = \mathbb{G}_n$. □

Theorem 2.3. *The linear operator $T : \boldsymbol{M}_2 \to \boldsymbol{M}_2$ defined by $T(X) = P^T XP$ is a linear preservers of G-matrices if and only if P is a generalized permutation matrix.*

Proof. The proof of the necessity follows from Proposition 2.2. We now prove the sufficiency. Let $P = \begin{pmatrix} d_1 & d_3 \\ d_2 & d_4 \end{pmatrix}$. For every $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{G}_2$, assume that $P^T XP \in \mathbb{G}_2$.

We have

$$P^T XP = \begin{pmatrix} ad_1^2 + bd_2d_1 + cd_1d_2 + dd_2^2 & ad_1d_3 + bd_1d_4 + cd_2d_3 + dd_2d_4 \\ ad_1d_3 + bd_2d_3 + cd_1d_4 + dd_2d_4 & ad_3^2 + bd_3d_4 + cd_4d_3 + dd_4^2 \end{pmatrix}.$$

Since $X \in \mathbb{G}_2$, by Theorem 1.3, we have three cases for X:

Case (i) : $a, d \neq 0$, and $b, c = 0$, so that $P^T XP = \begin{pmatrix} ad_1^2 + dd_2^2 & ad_1d_3 + dd_2d_4 \\ ad_1d_3 + dd_2d_4 & ad_3^2 + dd_4^2 \end{pmatrix}$.

$P^T XP \in \mathbb{G}_2$, by Theorem 1.3, we have the following relations:

1) $ad_1d_3 + dd_2d_4 = 0, ad_1^2 + dd_2^2 \neq 0$ and $ad_3^2 + dd_4^2 \neq 0$. Or

2) $ad_1^2 + dd_2^2 = ad_3^2 + dd_4^2 = 0$ and $ad_1d_3 + dd_2d_4 \neq 0$. Or

3) all entries of $P^T XP \neq 0$ and det $(P^T XP) \neq 0$.

Case (ii) : $b, c \neq 0$, and $a, d = 0$, so that $P^T XP = \begin{pmatrix} bd_2d_1 + cd_1d_2 & bd_1d_4 + cd_2d_3 \\ bd_2d_3 + cd_1d_4 & bd_3d_4 + cd_4d_3 \end{pmatrix}$.

$P^T XP \in \mathbb{G}_2$, of Theorem 1.3, we have the following relations:

$1^{'}$) $bd_1d_4 + cd_2d_3 = bd_2d_3 + cd_1d_4 = 0$, $bd_2d_1 + cd_1d_2 \neq 0$ and $bd_3d_4 + cd_4d_3 \neq 0$. Or

$2^{'}$) $bd_2d_1 + cd_1d_2 = bd_3d_4 + cd_4d_3 = 0$ and $bd_1d_4 + cd_2d_3 \neq 0, bd_2d_3 + cd_1d_4 \neq 0$. Or

$3^{'}$) all entries of $P^T XP \neq 0$ and det $(P^T XP) \neq 0$.

Case (iii) : All entries of $X \neq 0$ and det $X \neq 0$. So that

$$P^T XP = \begin{pmatrix} ad_1^2 + bd_2d_1 + cd_1d_2 + dd_2^2 & ad_1d_3 + bd_1d_4 + cd_2d_3 + dd_2d_4 \\ ad_1d_3 + bd_2d_3 + cd_1d_4 + dd_2d_4 & ad_3^2 + bd_3d_4 + cd_4d_3 + dd_4^2 \end{pmatrix}.$$

$P^T XP \in \mathbb{G}_2$, by Theorem 1.3, we have the following relations:

$1^{''}$) $ad_1d_3 + bd_1d_4 + cd_2d_3 + dd_2d_4 = ad_1d_3 + bd_2d_3 + cd_1d_4 + dd_2d_4 = 0$ and $ad_1^2 + bd_2d_1 + cd_1d_2 + dd_2^2 \neq 0, ad_3^2 + bd_3d_4 + cd_4d_3 + dd_4^2 \neq 0$. Or

$2^{''}$) $ad_1^2 + bd_2d_1 + cd_1d_2 + dd_2^2 = ad_3^2 + bd_3d_4 + cd_4d_3 + dd_4^2 = 0$ and $ad_1d_3 + bd_1d_4 + cd_2d_3 + dd_2d_4 \neq 0, ad_1d_3 + bd_2d_3 + cd_1d_4 + dd_2d_4 \neq 0$. Or

$3^{''}$) all entries $P^TXP \neq 0$ and det $(P^TXP) \neq 0$.

From the above relations it follows that on all rows and columns of P there exists exactly one nonzero entry and hence P is a generalized permutation matrix.

□

References

[1] M. Fiedler, F.J. Hall, *G-matrices,* Linear Algebra Appl. 436 (2012), 731-741.

[2] M. Fiedler, T. L. Markham, *More on G-matrices,* Linear Algebra Appl. 438 (2013), 231–241.

[3] S. Golshan, A. Armandnejad and F. J. Hall, Two $n \times n$ G-classes of matrices having finite intersection, Spacial Matrices, **11** (2023) 1–4.

[4] N. J. Higham, *J-orthogonal matrices: properties and generation,* SIAM Review 45 (3) (2003) 504–519.

[5] S. M. Motlaghian, A. Armandnejad and F. J. Hall, *Topological properties of J-orthogonal matrices,* Linear and Multilinear Algebra, 66 no. 12, (2018) 2524–2533.

[6] S. M. Motlaghian, A. Armandnejad and F. J. Hall, *A note on some classes of G-matrices*, Operators and Matrices, 16 (2022) 251–263.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

A new fast shift-splitting preconditioner for saddle point problems

G. Ebadi and S. Vakili*

Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran

Abstract

In this paper, a new fast shift-splitting (NFSS) method and its induced preconditioner is proposed for solving nonsymmetric saddle point problems. The convergence analysis of the NFSS iteration method is discussed. Finally, the efficiency of methods is illustrated by giving one example.

Keywords: Saddle-point, Shift-splitting, Preconditioner, Convergence
Mathematics Subject Classification [2010]: 65F10, 65F08

1 Introduction

We consider a nonsymmetric saddle point problem as

$$\mathfrak{A}u = \begin{pmatrix} A & B \\ -B^T & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} f \\ -g \end{pmatrix} \equiv b, \tag{1}$$

where $A \in \mathbb{R}^{m\times m}$ is nonsymmetric positive definite, $B \in \mathbb{R}^{m\times n}$ has full column rank, $f \in \mathbb{R}^m$ and $g \in \mathbb{R}^n$, with $m \geq n$. Here, B^T is the transpose of B.

When the matrices of coefficient matrix $\mathfrak{A}$, i.e., A and B are large and sparse, iterative methods are better suited for solving saddle point problems compared to direct methods [1]. If B in (1) has full column rank, then the coefficient matrix $\mathfrak{A}$ is nonsingular. In this case, the problem is called a nonsingular saddle point problem. When B has a rank deficiency, equation (1) is called the singular saddle point problem and the coefficient matrix $\mathfrak{A}$ is singular. In recent years, various authors have proposed a number of useful iterative methods to solve (1). Cao et al. [2] presented the SS preconditioner as

$$\mathcal{P}_{SS} = \frac{1}{2}\begin{pmatrix} \alpha I + A & B \\ -B^T & \alpha I \end{pmatrix},$$

where $\alpha \geq 0$ and I is the unit matrix with suitable dimension. Quan et al. [3] by using $A = H + S$ introduced FSS method for solving the saddle-point problem with nonsymmetric positive definite (1,1) part of the form

$$\mathcal{P}_{FSS} = \begin{pmatrix} \alpha I + H & B \\ -B^T & \alpha I \end{pmatrix}.$$

*Speaker. Email address: s.vakili@tabrizu.ac.ir

Jian et al. [4] substituted parameter α in the final block of $\mathcal{P}_{FSS}$ by parameter β and constructed a new preconditioner from the saddle point matrix $\mathfrak{A}$. Salkuyeh et al. [5] presented a modification of the generalized shift-splitting method for singular saddle point problems. Vakili et al. [6] have recently utilized constant l and symmetric positive definite matrices P, Q within the matrix $\mathfrak{A}$ to introduce a parameterized extended shift-splitting preconditioner $PESS$ as

$$\mathfrak{A} = \mathcal{P}_{PESS} - \mathcal{Q}_{PGSS} = \begin{pmatrix} \alpha P + lA & lB \\ -lB^T & \beta Q \end{pmatrix} - \begin{pmatrix} lA + \alpha P - A & lB - B \\ -(l-1)B^T & \beta Q \end{pmatrix},$$

such that $\hat{\alpha} \geq 0$, $\hat{\beta} > 0$.

These studies led us to introduce the new fast shift-splitting (NFSS) preconditioner in order to improve the convergence rate of (1) problems. In the current study the convergence of the proposed iteration method, and the spectral properties of $NFSS$ preconditioned matrix are investigated. We carry out a numerical example in order to show the efficiency of $NFSS$ method and the GMRES method with the $NFSS$ preconditioner for solving (1). This paper is structured as follows: Section 2 will introduce the new fast shift-splitting preconditioner and its implementation. Section 3 presents the convergence properties of the $NFSS$ iteration method. Section 4 presents the spectral analysis of the $NFSS$ preconditioned matrix. The numerical results are provided in Section 5.

2 The new fast shift-splitting preconditioner

In this Section, using idea of [3, 4], a new splitting of matrix $\mathfrak{A}$ is presented as

$$\begin{aligned} \mathfrak{A} &= \mathcal{P}_{NFSS} - \mathcal{Q}_{NFSS} \\ &= \begin{pmatrix} \alpha I + 2A & B \\ -B^T & \beta I \end{pmatrix} - \begin{pmatrix} \alpha I + A & 0 \\ 0 & \beta I \end{pmatrix}, \end{aligned} \tag{2}$$

where $\alpha \geq 0$, $\beta > 0$. Therefore, using (2), we present a new method as follows:
The $NFSS$ iteration method: Let $\alpha \geq 0$ and $\beta > 0$. Assume $(x^{(0)^T}, y^{(0)^T})^T$ be an initial guess for $k = 0, 1, 2, ...$, until $(x^{(0)^T}, y^{(k)^T})^T$ converges, compute

$$\mathcal{P}_{NFSS} \begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \mathcal{Q}_{NFSS} \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} + \begin{pmatrix} f \\ -g \end{pmatrix}, \tag{3}$$

The iteration scheme (3) can be rewritten as follows

$$\begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} = \Gamma(\alpha, \beta) \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} + \begin{pmatrix} \alpha I + 2A & B \\ -B^T & \beta I \end{pmatrix}^{-1} \begin{pmatrix} f \\ -g \end{pmatrix}, \tag{4}$$

where

$$\Gamma(\alpha, \beta) = \begin{pmatrix} \alpha I + 2A & B \\ -B^T & \beta I \end{pmatrix}^{-1} \begin{pmatrix} \alpha I + A & 0 \\ 0 & \beta I \end{pmatrix}$$

is the iteration matrix of the NFSS method, and

$$\mathcal{P}_{NFSS} = \begin{pmatrix} \alpha I + 2A & B \\ -B^T & \beta I \end{pmatrix},$$

is called the $NFSS$ preconditioner for $\mathfrak{A}$. At each step of (4), we need to solve a linear system in the following form.

$$\begin{pmatrix} \alpha I + 2A & B \\ -B^T & \beta I \end{pmatrix} z = r.$$

3 The convergence of the $NFSS$ iteration method

To demonstrate the convergent properties of the $NFSS$ iteration method, we provide some necessary lemmas.

Lemma 3.1. *Both roots of the complex quadratic equation $x^2 - \phi x + \psi = 0$ are less than one in modulus if and only if $|\phi - \bar{\phi}\psi| + |\psi|^2 < 1$, where $\bar{\phi}$ denotes the conjugate complex of ϕ.*

Lemma 3.2. *Assume $A \in \mathbb{R}^{m\times m}$ is a positive definite matrix, $B \in \mathbb{R}^{m\times n}$ has full column rank, $\alpha \geq 0$ and $\beta > 0$. If λ is an eigenvalue of the $\Gamma(\alpha, \beta)$, then $\lambda \neq \pm 1$.*

Lemma 3.3. *Assume λ be an eigenvalue of $\Gamma(\alpha, \beta)$ and $(u^*, v^*)^* \in \mathbb{C}^{m\times n}$, be the corresponding eigenvector and all the conditions in Lemma 3.2 are satisfied, then $u \neq 0$. Moreover, if $v = 0$, then $|\lambda| < 1$.*

Theorem 3.4. *Assume that the conditions in Lemma 3.2 are satisfied. Let $(\lambda, (u^*, v^*)^*)$ be an eigenpair of $\Gamma(\alpha, \beta)$ of the $NFSS$ iteration method. Then the $NFSS$ iteration method converges to the exact solution of the saddle point problem (1).*

4 The spectral analysis of the $NFSS$ preconditioned matrix

The rate of convergence is closely related to the distribution of eigenvalues and eigenvectors of the $NFSS$ preconditioned matrix $\mathcal{P}_{NFSS}^{-1}\mathfrak{A}$. Therefore, we investigate the spectral features of the preconditioned matrix $\mathcal{P}_{NFSS}^{-1}\mathfrak{A}$.

Theorem 4.1. *Let the preconditioner of the $NFSS$ method be defined as shown in (4). Let λ be an eigenvalue of $\mathcal{P}_{NFSS}^{-1}\mathfrak{A}$ and $(u^*, v^*)^*$ be the corresponding eigenvector. If B has full column rank and $B^T u = 0$, then*

$$\frac{\alpha\lambda_{min}(H) + 2\lambda_{min}(H)^2}{(\alpha + 2\rho(H))^2 + 4\rho(S)^2} \leq Re(\lambda) \leq \frac{\alpha\rho(H) + 2\rho(H)^2 + 2\rho(S)^2}{(\alpha + 2\lambda_{min}(H))^2},$$
$$|Im(\lambda)| \leq \frac{\alpha\rho(S)}{(\alpha + 2\lambda_{min}(H))^2}, \tag{5}$$

5 Numerical experiments

We provide an example to explain the feasibility and effectiveness of the $NFSS$ method for solving (1).

Example 5.1. The problem structured as (1) was considered with the following coefficient sub-matrices

$$A = \begin{pmatrix} I \otimes T + T \otimes I & 0 \\ 0 & I \otimes T + T \otimes I \end{pmatrix} \in \mathbb{R}^{2p^2\times 2p^2},$$
$$B = \begin{pmatrix} I \otimes F \\ F \otimes I \end{pmatrix} \in \mathbb{R}^{2p^2\times p^2},$$

$T = \frac{\mu}{h^2}.tridiag(-1, 2, -1) + \frac{1}{2h}.tridiag(-1, 0, 1) \in \mathbb{R}^{p\times p}$, $F = \frac{1}{h}.tridiag(-1, 1, 0) \in \mathbb{R}^{p\times p}$,

where $h = \frac{1}{p+1}$ and $\otimes$ denotes the Kronecker product.

Table 1: Numerical results for the example with $\mu = 0.1$.

Method	p	16	32	64
GSS	α	20	40	90
	β	9.993	9.992	9.991
	IT.	52	93	161
	CPU	0.11	0.85	3.56
	RES	$7.67e-07$	$9.70e-07$	$9.40e-08$
FSS	α	2.7	2	1
	IT.	37	42	40
	CPU	0.04	0.21	0.97
	RES	$29.47e-07$	$8.09e-07$	$8.96e-08$
GFSS	α	2.99	3.85	3.3
	β	0.099	0.1007	0.1007
	IT.	34	29	26
	CPU	0.037	0.137	0.64
	RES	$8.52e-07$	$7.86e-07$	$9.16e-07$
NFSS	α	0.1	1	0.2
	β	0.1	0.1	0.1
	IT.	20	20	20
	CPU	0.01	0.072	0.39
	RES	$5.62e-07$	$6.85e-07$	$6.42e-07$

Table 1 shows the efficiency of the $NFSS$ method through the selection of small values for α and β. Table 1 presents the results of numerical experiments for different iteration methods, where the optimal parameters have been determined through experimental minimization of iterations for $\mu = 0.1$ on various grids. Compared to the other two methods, the $NFSS$ iteration method for solving Example 5.1 requires less processing time. We

Table 2: Numerical of results for the three preconditioned GMRES methods with $\mu = 0.2$.

Method	p	16	32
I	IT.	115	240
	CPU	0.1326	3.4868
	RES	$9.50e-07$	$9.34e-07$
$\mathcal{P}_{GSS}$	α	0.3	0.6
	β	1.8	0.7
	IT.	9	8
	CPU	0.039	0.148
	RES	$7.68e-07$	$4.14e-07$
$\mathcal{P}_{FSS}$	α	0.03	0.09
	IT.	8	9
	CPU	0.11	0.162
	RES	$7.69e-07$	$2.78e-07$
$\mathcal{P}_{GFSS}$	α	0.02	0.03
	β	0.01	0.01
	IT.	8	8
	CPU	0.035	0.149
	RES	$4.65e-07$	$3.41e-07$
$\mathcal{P}_{NFSS}$	α	0.01	0.02
	β	0.02	0.01
	IT.	7	7
	CPU	0.023	0.102
	RES	$4.27e-08$	$3.82e-08$

present numerical experiments of the GSS, FSS, $GFSS$, and $NFSS$ preconditioned GMRES methods on different uniform grids with $\mu = 0.2$ in Tables 2. Note that I in Table 2 indicates the GMRES method without preconditioning. The GMRES method with $\mathcal{P}_{NFSS}$ preconditioning has been shown to be both feasible and efficient in Table 2.

References

[1] M. Benzi, G.H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numer., 14 (2015) 1-137.

[2] Y. Cao, J. Du, and Q. Niu, *Shift-splitting preconditioners for saddle point problems*, J. Comput. Appl. Math., 272 (2014) 239-250.

[3] Dou, Q.Y, Yin, J.F, and Liao, Z.Y, *A fast shift-splitting iteration method for nonsymmetric saddle point problems*, East Asian J. on Applied Math., 7 (2017) 172-191.

[4] Zheng, J.H, Chen, X.P, and Zhao, J, *Generalized fast shift-splitting preconditioner for nonsymmetric saddle-point problems*, J. Comput. Appl. Math., 92 (2019) 91-114.

[5] Salkuyeh, D.K., Rahimian, M., *A modification of the generalized shift-splitting method for singular saddle point problems*, Comput. Math. with Appl., 74 (2017) 2940-2949.

[6] Vakili, S, Ebadi, G, and Vuik, C, *A parameterized extended shift-splitting preconditioner for nonsymmetric saddle point problems*, Numer. Linear Algebra Appl., 2022;e2478. https://doi.org/10.1002/nla.2478.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

An iterative method for solving the constrained tensor equation using the Einstein product

B. Zali* and S. Karimi

Department of Mathematics, Faculty of Intelligent Systems Engineering and Data Science, Persian Gulf University, Bushehr, Iran

Abstract

In this paper, we will propose an iterative method for solving the tensor equation $\mathcal{A} *_N \mathcal{X} = \mathcal{B}$ with the constraint $\mathcal{X}^T = \mathcal{X}$, where $*_N$ is the symbol of Einstein product. The proposed iterative method is based on the generalized least squares method.

Keywords: Tensor equation, Global least squares, Constrained equation, Einstein product

Mathematics Subject Classification [2010]: 15A10, 15A69, 15A72

1 Introduction

Tensor equations have many applications in image processing and deep learning. In some practical problems, such as control problems and physics and tensor equations have constraints.

A tensor is often thought of as a generalized matrix. Some basic definitions related to a tensor that is used throughout this paper are introduced in the following. Throughout this paper, let $\mathcal{H} = \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$, and $\mathcal{N} = \mathbb{R}^{J_1 \times \cdots \times J_N \times J_1 \times \cdots \times J_N}$.

Definition 1.1. For a positive integer N, an order N tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ consists of $\prod_{i=1}^{N} I_i$ elements in the real field $\mathbb{R}$.

Suppose $I_1, \cdots I_N \in \mathbb{N}$, $\mathcal{A} = (a_{i_1 \cdots i_N})$ is a multidimensional array with M ($M = \prod_{i=1}^{N} I_i$) entries.

Definition 1.2. [1] Let $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$, $\mathcal{B} \in \mathbb{R}^{J_1 \times \cdots \times J_N \times K_1 \times \cdots \times K_M}$, the Einstein product of the tensors $\mathcal{A}$ and $\mathcal{B}$ is the tensor of size $\mathbb{R}^{I_1 \times \cdots \times I_N \times K_1 \times \cdots \times K_M}$ whose elements are defined by

$$(\mathcal{A} *_N \mathcal{B})_{i_1 \cdots i_N k_1 \cdots k_M} = \sum_{j_1, \cdots, j_N} a_{i_1 \cdots i_N j_1, \cdots j_N} b_{j_1 \cdots j_N k_1 \cdots k_N}, \quad 1 \leq j_i \leq J_i, \quad i = 1, \cdots N.$$

Definition 1.3. [1] For a tensor $\mathcal{A} = (a_{i_1 \cdots i_N j_1 \cdots j_M}) \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$, let $\mathcal{B} = (b_{i_1 \cdots i_M j_1 \cdots j_N}) \in \mathbb{R}^{J_1 \times \cdots \times J_M \times I_1 \times \cdots \times I_N}$ be the transpose of $\mathcal{A}$, where $b_{i_1 \cdots i_M j_1 \cdots j_N} = a_{j_1 \cdots j_N i_1 \cdots i_M}$. The tensor $\mathcal{B}$ is denoted by $\mathcal{A}^T$.

*Speaker:hoda5964@yahoo.com

Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$, we define

$$< \mathcal{X}, \mathcal{Y} >= tr(\mathcal{X}^T *_N \mathcal{Y}), \tag{1}$$

where the trace of an even-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ is given by

$$tr(\mathcal{X}) = \sum_{i_1}^{I_1} \cdots \sum_{i_N}^{I_N} x_{i_1 \ldots i_N i_1 \ldots i_N}.$$

It is easy to show that equation (1) is an inner product on the subspace $\mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$. Then, for a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$, the tensor norm induced by this inner product is Frobenius norm $\|\mathcal{A}\|^2 =< \mathcal{X}, \mathcal{X} >^2 = \sum_{i_1 \ldots, i_N, j_1 \ldots, j_N} \mid a_{i_1 \ldots i_N j_1 \ldots j_N} \mid^2$.

Definition 1.4. [1] The tensor $\mathcal{D} = (d_{i_1 \cdots i_N j_1 \cdots j_N}) \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_N}$ is called a diagonal tensor if $d_{i_1 \cdots i_N j_1 \cdots j_N} = 0$ in the case that the indices $i_1 \cdots i_N$ are different from $j_1 \cdots j_N$. If all diagonal entries $d_{i_1 \cdots i_N i_1 \cdots i_N} = 1$ then, $\mathcal{D}$ is called the unit tensor and denoted by $\mathcal{I}$.

Proposition 1.5. *[1] Let $\mathcal{A} \in \mathcal{H}$ and $\mathcal{C} \in \mathcal{N}$, then*

$$(\mathcal{A} *_N \mathcal{C})^T = \mathcal{C}^T *_N \mathcal{A}^T.$$

2 Main results

In this section, we propose a global least squares tensor (GLS-T) method for solving the following constrained tensor equation:

$$\begin{cases} \mathcal{A} *_N \mathcal{X} = \mathcal{B}, \\ \mathcal{X} = \mathcal{X}^T, \end{cases} \tag{2}$$

where $\mathcal{A}, \mathcal{B} \in \mathcal{H}$ are known and $\mathcal{X} \in \mathcal{N}$ is an unknown tensor to be determined. Now, we introduce a equivalent system to constrained tensor equation (2), the following remark can be easily shown.

Remark 2.1. Any solution of equation (2) is a solution of the following pair of equations:

$$\begin{cases} \mathcal{A} *_N \mathcal{X} = \mathcal{B}, \\ \mathcal{A} *_N \mathcal{X}^T = \mathcal{B}. \end{cases} \tag{3}$$

The pair equation (3) can be rewritten as the following tensor operator equation form:

$$\hat{\mathcal{A}} : \mathcal{N} \longrightarrow \mathcal{H} \times \mathcal{H},$$

$$\hat{\mathcal{A}}(\mathcal{X}) = (\mathcal{A} *_N \mathcal{X}, \mathcal{A} *_N \mathcal{X}^T) = \hat{\mathcal{B}}, \tag{4}$$

where $\hat{\mathcal{B}} = (\mathcal{B}, \mathcal{B})$.

Definition 2.2. [4] Let $\hat{\mathcal{A}}$ be the linear operator (4). Then the linear operator

$$\hat{\mathcal{A}}^* : \mathcal{H} \times \mathcal{H} \longrightarrow \mathcal{N},$$

that satisfies

$$< \hat{\mathcal{A}}(\mathcal{X}), \mathcal{Y} >=< \mathcal{X}, \hat{\mathcal{A}}^*(\mathcal{Y}) >,$$

for all $\mathcal{X} \in \mathcal{N}$ and $\mathcal{Y} \in \mathcal{H} \times \mathcal{H}$, is called the adjoint of $\hat{\mathcal{A}}$.

As a consequence of the tensor operator (4), and the above definition, we have the following remark which is easy to prove.

Remark 2.3. Let $\hat{\mathcal{A}}$ be the linear operator (4), then the adjoint operator $\hat{\mathcal{A}}^* : \mathcal{H} \times \mathcal{H} \longrightarrow \mathcal{N}$ is

$$\hat{\mathcal{A}}^*(Y, Z) = \mathcal{A}^T *_N Y + Z^T *_N \mathcal{A}.$$

Now, we present the GLS-T algorithm which is a generalization of the GL-LSQR for solving the constrained matrix equation [5]. Similar to the bidiagonal process for the matrix equation, we give the bidiagonalization process for the tensor operator equation (4).

Algorithm 2.4. Bidiagonalization Process (starting tensor $\hat{\mathcal{B}}$)

1. $\beta_1 \mathcal{U}_1 = \hat{\mathcal{B}}, \qquad \alpha_1 \mathcal{V}_1 = \hat{\mathcal{A}}^*(\mathcal{U}_1),$
2. For $i = 1, 2, \ldots n$
3. $\quad \beta_{i+1} \mathcal{U}_{i+1} = \hat{\mathcal{A}} *_N \mathcal{V}_i - \alpha_i \mathcal{U}_i,$
4. $\quad \alpha_{i+1} \mathcal{V}_{i+1} = \hat{\mathcal{A}}^*(\mathcal{U}_{i+1}) - \beta_{i+1} \mathcal{V}_i,$
5. End.

Where $\mathcal{U}_i \in \mathcal{H} \times \mathcal{H}$, $\mathcal{V}_i \in \mathcal{N}$, and the scalers $\alpha_i \geq 0$ and $\beta_i \geq 0$ are chosen such that $\|\mathcal{U}_i\| = 1$ and $\|\mathcal{V}_i\| = 1$. Now the approximate solution (4) is obtained by the GLS-T algorithm expressed as follows :

Algorithm 2.5. GLS-T algorithm

1. Set $\mathcal{X} = 0 \in \mathcal{N}$,
2. $\beta_1 = \|\hat{\mathcal{B}}\|, \quad \mathcal{U}_1 = \frac{\hat{\mathcal{B}}}{\beta_1}, \quad \alpha_1 = \|\hat{\mathcal{A}}^*(\mathcal{U}_1)\|, \quad \mathcal{V}_1 = \frac{\hat{\mathcal{A}}^*(\mathcal{U}_1)}{\alpha_1},$
3. Set $\mathcal{W}_1 = \mathcal{V}_1, \quad \hat{\Phi}_1 = \beta_1, \quad \hat{\rho}_1 = \alpha_1,$
4. For $i = 1, 2, \ldots$ until convergence, Do:
5. $\quad \hat{\mathcal{W}}_i = \hat{\mathcal{A}}(\mathcal{V}_i) - \alpha_i \mathcal{U}_i,$
6. $\quad \beta_{i+1} = \|\hat{\mathcal{W}}_i\|,$
7. $\quad \mathcal{U}_{i+1} = \frac{\hat{\mathcal{W}}_i}{\beta_{i+1}},$
8. $\quad \hat{\mathcal{S}}_i = \hat{\mathcal{A}}^*(\mathcal{U}_{i+1}) - \beta_{i+1} \mathcal{V}_i,$
9. $\quad \alpha_{i+1} = \|\hat{\mathcal{S}}_i\|,$
10. $\quad \mathcal{V}_{i+1} = \frac{\hat{\mathcal{S}}_i}{\alpha_{i+1}},$
11. $\quad \rho_i = \sqrt{(\hat{\rho}_i^2 + \beta_{i+1}^2)},$
12. $\quad c_i = \frac{\hat{\rho}_i}{\rho_i},$
13. $\quad s_i = \frac{\beta_{i+1}}{\rho_i},$

14. $\theta_{i+1} = s_i\alpha_{i+1}$,

15. $\hat{\rho}_{i+1} = c_i\alpha_{i+1}$,

16. $\Phi_i = c_i\hat{\Phi}_i$,

17. $\hat{\Phi}_{i+1} = -s_i\hat{\Phi}_i$,

17. $\mathcal{X}_i = \mathcal{X}_{i-1} + \frac{\Phi_i}{\rho_i}\mathcal{W}_i$,

18. $\mathcal{W}_{i+1} = \mathcal{V}_{i+1} - \frac{\theta_i}{\rho_i}\mathcal{W}_i$,

19. If $\mid \hat{\Phi}_{i+1} \mid$ is smal enough then stop,

20 EndDo .

The approximate solution of the constrained equation (2) can be obtained by $\hat{\mathcal{X}} = \frac{\mathcal{X}+\mathcal{X}^T}{2}$, where $\mathcal{X}$ is the approximate solution of the associated unconstrained reduced equation (4) obtained by GLS-T.

Remark 2.6. [2] The stopping criterion can be chosen as $\|\mathcal{R}_k\| = \mid \hat{\Phi}_{i+1} \mid$, where $\mathcal{R}_k$ is the kth residual.

3 Numerical results

In this section, we report two numerical examples to show the performance of the proposed algorithm on the tensor equation (2). All tests were run on the Intel(R), Core(TM) $i7-8565U$, CPU 2.00 GHz, and 16.00 GB RAM. The programming language was MATLAB $R2021b$, using the code from the MATLAB tensor toolbox developed by Bader and Kolda [3].

Example 3.1. In this example, we solve the tensor equation (2) with $\mathcal{A} = s\mathcal{I} - \mathcal{C} \in \mathbb{R}^{n\times n\times n\times n}$ and $\mathcal{C}$ being a nonnegative tensor with

$$c_{i_1i_2i_3i_4} = |sin(i_1 + i_2 + i_3 + i_4)|.$$

By taking $s = n^3$, it follows from [6] that $\mathcal{A}$ is a nonsingular $\mathcal{M}-$tensor. We chose the right-hand side tensor as $\mathcal{B} = \mathcal{A} *_2 tenones([n, n, n, n])$, where $tenones([n, n, n, n])$ is a MATLAB style which is the 4-order n-dimensional tensor with all entries equal to 1.

Example 3.2. Consider tensor equation (2), creates a random sparse tensor of the specified x with approximately nnz nonzero entries $\mathcal{A}$ in MATLAB style as follows:

$$\mathcal{A} = sptenrand(x, nnz).$$

In this example, we choose $x = [a\ b\ c\ a\ b\ c]$ a 6-dimensional vector, and $nnz = a \times b \times c$. Also, by taking $\mathcal{X}^* = tenones([a, b, c, a, b, c]) - \mathcal{I}$ as the exact solution of $\mathcal{A} *_3 \mathcal{X} = \mathcal{B}$, we generate the right-hand tensor $\mathcal{B}$.

We implemented the GLS-T algorithm for Examples (3.1) and (3.2) to investigate the numerical solution of the tensor equation (2). In both examples, we take the initial tensor to be zero tensors, and the stopping criterion is that the kth-iteration residual satisfies

$$\mid \hat{\Phi}_{k+1} \mid = \|\mathcal{R}_k\| < 10^{-8}.$$

The numerical results are shown in Table 1. Also, in Table 1, notations $Iter$, CPU, and RES are the number of iteration steps, the elapsed CPU time in seconds, and the residual norm, respectively.

Table 1: Numerical results of Examples 3.1, 3.2.

Example	order	Iter	CPU	RES
3.1($n = 10$)	4	7	0.0827	8.29×10^{-10}
3.1($n = 20$)	4	6	0.2177	5.59×10^{-9}
3.2($x = [2\ 4\ 6\ 2\ 4\ 6]$)	6	550	3.6933	6.59×10^{-9}
3.2($x = [3\ 5\ 7\ 3\ 5\ 7]$)	6	999	8.9310	6.48×10^{-9}

4 Conclusion

In this paper, we proposed an iterative method for obtaining the approximate solution of the tensor equation $\mathcal{A} *_N \mathcal{X} = \mathcal{B}$ with the constraint $\mathcal{X} = \mathcal{X}^T$. For this purpose, first we reduced the constrained tensor equation to an equivalent unconstrained coupled tensor equation. Then, we applied the GLS-T algorithm for obtaining the approximate solution $\mathcal{X}$ of the new coupled tensor equation. Finally, the approximate solution of the original equation gets as $\hat{\mathcal{X}} = \frac{\mathcal{X}+\mathcal{X}^T}{2}$. Numerical experiments showed the efficiency of the new method.

References

[1] Baohua Huanga, Changfeng Ma, Global least squares methods based on tensor form to solve a class of generalized Sylvester tensor equations, *Appl. Math. Comput.*, 369 (2020).

[2] C. C. Painge, M. A. Saunders, LSQR an Algorithm for spares linear equations and sparse least squares, *ACM Trans. Math Softw.*, 8 (1982) 43-71.

[3] T. G. Kolda, B. W. Bader, Tensor Decompositions and Application, *SIAM. Rew.*, (2009), No. 3, 455-500.

[4] S. Karimi, M. Dehghan, A general iterative approach for solving the general constrained linear matrix equations system,*Tran. Inst. Meas. control.*, (2015) 1-14.

[5] F. Toutounian, S. Karimi, Global least squares method (GL-LSQR) for solving general linear systems with several right-hand sides, *Appl. Math. Comput.*, 178 (2006) 452-460.

[6] Xie ZJ, Jin XQ, Wei YM, Tensor methods for solving symmetric $\mathcal{M}$-tensor systems, *J. Sci. Comput.*, 74 (2017), 412-25.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Cartesian symmetry classes associated with dihedral group[1]

S.S. Gholami* and Y. Zamani

Department of Mathematics, Sahand University of Technology, Tabriz, Iran

Abstract

This paper provides a necessary and sufficient condition for the existence of an orthogonal basis consisting of standard symmetrized vectors for Cartesian symmetry classes associated with the dihedral group. In addition, the dimensions of these classes are also computed.

Keywords: Irreducible characters, dihedral groups, Cartesian symmetry classes, generalized trace functions.

Mathematics Subject Classification [2010]: Primary: 20C15; Secondary: 15A69

1 Introduction

In this section, we give a review of Cartesian symmetry classes. The reader can be find a detailed introduction in [5, 6].

Let V be a complex inner product space of dimension n.9 Let G be a subgroup of S_m and $\mathbb{E} = \{e_1, \cdots, e_n\}$ is an orthonormal basis of V. Let $\times^m V$ be the Cartesian product of m-copies of V. We have an induced inner product of $\times^m V$, which is defined by

$$< u^{\times}, v^{\times} >= \sum_{i=1}^{m} < u_i, v_i >,$$

where

$$u^{\times} = (u_1, \cdots, u_m),\ v^{\times} = (v_1, \cdots, v_m).$$

For every $1 \leq i \leq n,\ 1 \leq j \leq m$, we define

$$e_{ij} = (\delta_{1j}e_i, \delta_{2j}e_i, \cdots, \delta_{mj}e_i) \in \times^m V.$$

Then the set

$$\mathbb{E}^{\times} = \{e_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq m\}$$

is an orthonormal basis of $\times^m V$.

[1]The presented results in this talk are the summary of the authors' recently submitted manuscript which is accessible in Arxiv (see [1]).

*Speaker. Email address: rgolamie@yahoo.com

Let G be a subgroup of S_m. For any $\sigma \in G$, the linear operator

$$Q_\sigma : \times^m V \longrightarrow \times^m V$$

defined by

$$Q_\sigma(v_1, \cdots, v_m) = (v_{\sigma^{-1}(1)}, \cdots, v_{\sigma^{-1}(m)})$$

is called Cartesian permutation operator with respect to σ. It is easy to see that $Q_{\sigma\tau} = Q_\sigma Q_\tau$, for all $\sigma,\ \tau \in G$. Moreover, Q_σ is invertible. Therefore $Q : \sigma \to Q_\sigma$ defines a faithful unitary representation of G over $\times^m V$.

Let χ be a complex irreducible character of G. We define the *Cartesian symmetrizer* $C(G, \chi)$ as follows:

$$C(G, \chi) = \frac{\chi(1)}{|G|} \sum_{\sigma \in G} \chi(\sigma) Q_\sigma.$$

It is proved [6] that $C(G, \chi)$ is an orthogonal projection on $\times^m V$. The image of $\times^m V$ under the map $C(G, \chi)$ is called *the Cartesian symmetry class associated with G and χ* and is denoted by $V^\chi(G)$. It is proved that $\times^m V$ is the orthogonal direct sum of the Cartesian symmetry classes $V^\chi(G)$ as χ ranges over $Irr(G)$.

Clearly $V^\chi(G)$ is spanned by *the standard symmetrized vectors*

$$e^\chi_{ij} = C(G, \chi)(e_{ij}).$$

Let $\mathcal{D}$ be a set of representatives of orbits of the set $\{1, 2, \cdots, m\}$. Now suppose

$$\mathcal{O} = \{j \mid 1 \leq j \leq m,\ [\chi, 1_{G_j}] \neq 0\},$$

where $\bar{\mathcal{D}} = \mathcal{D} \cap \mathcal{O}$ and $[\ ,\ ]$ is the inner product of characters (see [3]). It is easy to see that the set

$$\{e^\chi_{ij} \mid 1 \leq i \leq n, j \in \bar{D}\}$$

is an orthogonal set of non-zero vectors in $V^\chi(G)$.

For any $1 \leq i \leq n$ and $j \in \bar{\mathcal{D}}$, define the cyclic subspace

$$V^\chi_{ij} = \langle e^\chi_{i\sigma(j)} \mid \sigma \in G \rangle.$$

It is proved that

$$V^\chi(G) = \sum_{i,j}^{\perp} V^\chi_{ij},$$

the orthogonal direct sum of the cyclic subspaces V^χ_{ij} $(1 \leq i \leq n,\ j \in \bar{\mathcal{D}})$.
Also

$$\dim V^\chi(G) = \dim(V)\chi(1) \sum_{j \in \bar{\mathcal{D}}} [\chi, 1_{G_j}].$$

If χ is a linear character of G and $j \in \bar{\mathcal{D}}$, then it is easy to see that $e^\chi_{i\sigma(j)} = \chi(\sigma^{-1}) e^\chi_{ij}$, so $\dim V^\chi_{ij} = 1$ and the set

$$\{e^\chi_{ij} \mid 1 \leq i \leq n, j \in \bar{\mathcal{D}}\}$$

is an orthogonal basis of $V^\chi(G)$. Suppose χ is a non-linear irreducible character of G. We now construct a basis of $V^\chi(G)$. For each $j \in \bar{\mathcal{D}}$, we choose the set $\{j_1, \cdots, j_{s_j}\}$ from the orbit of j such that $\{e^\chi_{ij_1}, \cdots, e^\chi_{ij_{s_j}}\}$ is a basis of the cyclic subspace V^χ_{ij}. Execute this procedure for each $k \in \bar{\mathcal{D}}$. If $\bar{\mathcal{D}} = \{j, k, l, \cdots\}$ $(j < k < l < \cdots)$, take

$$\hat{\mathcal{D}} = \{j_1, \cdots, j_{s_j}; k_1, \cdots, k_{s_k}, \cdots\}$$

to be ordered as indicated. Then

$$\mathbb{E}^\chi = \{e^\chi_{ij} \mid 1 \leq i \leq n,\ j \in \hat{\mathcal{D}}\}$$

is a basis of $V^\chi(G)$. Note that it may not be orthogonal basis. So

$$\dim V^\chi(G) = (\dim V)|\hat{\mathcal{D}}|.$$

If the subspace W of $\times^m V$ has a basis consisting of orthogonal standard symmetrized vectors, we will say that W has an *orthogonal O-basis*.

2 The Dihedral Group

The subgroup D_{2m} of S_m $(m \geq 3)$ generated by the elements

$$r = (1\ 2\ \cdots\ m) \quad \text{and} \quad s = \begin{pmatrix} 1 & 2 & 3 & \cdots & m-1 & m \\ 1 & m & m-1 & \cdots & 3 & 2 \end{pmatrix}$$

is the *dihedral group of degree m*. The generators r and s satisfy (see [2, P. 50])

$$r^m = 1 = s^2 \text{ and } s^{-1}rs = r^{-1}.$$

If m is even, i.e., $m = 2k$ $(k \geq 2)$, then D_{2m} has $k+3$ conjugacy classes. If m is odd, i.e., $m = 2k+1$ $(k \geq 1)$, then D_{2m} has $k+2$ conjugacy classes.

For each integer h with $0 < h < m/2$, D_{2m} has an irreducible character χ_h of degree 2 given by

$$\psi_h(r^k) = 2\cos(\frac{2kh\pi}{m}), \quad \psi_h(sr^k) = 0, \quad 0 \leq k < m.$$

The other characters of D_{2m} are of degree 1, namely χ_j. The character table of D_{2m} is shown in Table 1 (see [4, P. 182]).

Table 1: Character table of D_{2m}

m is odd	r^k	sr^k	m is even	r^k	sr^k
χ_1	1	1	χ_1	1	1
χ_2	1	-1	χ_2	1	-1
ψ_h	$2\cos(\frac{2kh\pi}{m})$	0	χ_3	$(-1)^k$	$(-1)^k$
-	-	-	χ_4	$(-1)^k$	$(-1)^{k+1}$
-	-	-	ψ_h	$2\cos(\frac{2kh\pi}{m})$	0

3 Main Results

In this section, we obtain the dimensions of Cartesian symmetry classes associated with the irreducible characters of the dihedral group D_{2m}. Also we give a necessery and sufficient condition for the existence an orthogonal O -basis for Cartesian symmetry classes $V^{\psi_h}(G)$ $(0 < h < \frac{m}{2})$.

Theorem 3.1. *Let $G = D_{2m}$ $(m \geq 3)$. Assume $n = \dim V \geq 2$. If m is even, then*

(a) $\dim V^{\chi_1}(G) = \dim V^{\chi_3}(G) = n$,

(b) $\dim V^{\chi_2}(G) = \dim V^{\chi_4}(G) = 0$,

(c) $\dim V^{\psi_h}(G) = 2n$ $(0 < h < \frac{m}{2})$.

Theorem 3.2. *Let $G = D_{2m}$ $(m \geq 3)$. Assume $n = \dim V \geq 2$. If m is odd, then*

(a) $\dim V^{\chi_1}(G) = n$,

(b) $\dim V^{\chi_2}(G) = 0$,

(c) $\dim V^{\psi_h}(G) = 2n$ $(0 < h < \frac{m}{2})$.

Theorem 3.3. *Let $G = D_{2m}$ $(m \geq 3)$ and $\psi = \psi_h$ $(0 < h < \frac{m}{2})$. Then $V^{\psi}(G)$ has orthogonal O-basis if and only if $m \equiv 0 \ (mod\ 4h_2)$, where $h = h_2 h_{2'}$ with h_2 a power of 2 and $h_{2'}$ odd.*

Corollary 3.4. *Let $G = D_{2m}$ and assume $\dim V \geq 2$. Then $\times_1^m V$ has an orthogonal O-basis if and only if m is a power of 2.*

References

[1] S.S. Gholami and Y. Zamani, *Cartesian symmetry classes*, Preprint, https://arxiv.org/abs/2304.13990.

[2] T.W. Hungerford, *Algebra*, New York: Holt, Rinehart and Wilson, 1974.

[3] M. Isaacs, *Character Theory of Finite Groups*, Academic Press, 1976.

[4] G. James and M. Liebeck, *Representations and Characters of Groups*, Cambridge University press, 1993.

[5] T.G. Lei, *Notes on Cartesian symmetry classes and generalized trace functions*, Linear Algebra Appl., 292 (1999), 281-288.

[6] Y. Zamani and M. Shahryari, *On the dimensions of Cartesian symmetry classes*, Asian-Eur. J. Math., 5(3) (2012), 1250046 (7 pages).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Generalized Cartesian Symmetry Classes[1]

Yousef Zamani* and Seyyed Sadegh Gholami

Department of Mathematics, Sahand University of Technology, Tabriz, Iran

Abstract

Assuming V is a finite-dimensional inner product space, G is a subgroup of the full symmetric group S_m, and $\mathfrak{X}$ is an irreducible unitary representation of G. In this paper, we introduce the generalized Cartesian symmetry class over V associated with G and $\mathfrak{X}$. We proceed to investigate some important properties of this vector space.

Keywords: Irreducible unitary representation, Cartesian symmetry class, generalized trace function, orthogonal basis, o.b.-representation

Mathematics Subject Classification [2010]: 15A69, 20C30

1 Introduction

Let S_m denote the full symmetric group of degree m, and let G be a subgroup of S_m. Let U be a unitary space. The set of all linear operators on U is denoted by $End\ (U)$. Assume that $\mathfrak{X}$ is an irreducible unitary representation of G over U. The generalized trace function $Tr_{\mathfrak{X}} : \mathbb{C}_{m\times m} \longrightarrow \text{End } (U)$ is defined by

$$Tr_{\mathfrak{X}}(A) = \sum_{\sigma\in G} \mathfrak{X}(\sigma) \sum_{i=1}^{m} a_{i\sigma(i)}$$

for $A = (a_{ij}) \in \mathbb{C}_{m\times m}$.
It is proved that $Tr_{\mathfrak{X}}(A^*) = Tr_{\mathfrak{X}}(A)^*$. In particular, if A is Hermitian, then $Tr_{\mathfrak{X}}(A)$ is Hermitian (see [5]).

Let V be a unitary space of dimension n and denote by $\times^m V$ be the Cartesian product of m-copies of V. Then $U \otimes V^{\times m}$ is a unitary space with an induced inner product given by

$$\langle u \otimes x^{\times}, v \otimes y^{\times}\rangle = \langle u, v\rangle \sum_{i=1}^{m} \langle x_i, y_i\rangle,$$

where $u, v \in U$ and $x^{\times} = (x_1, \cdots, x_m)$, $y^{\times} = (y_1, \cdots, y_m) \in \times^m V$.

[1]The presented results in this talk are the summary of the authors' recently submitted manuscript which is accessible in Arxiv (see [1])

*Speaker. Email address: zamani@sut.ac.ir

The *generalized Cartesian symmetrizer associated with* G *and* $\mathfrak{X}$ is defined by

$$C_{\mathfrak{X}} = \frac{1}{|G|}\sum_{\sigma\in G}\mathfrak{X}(\sigma)\otimes Q(\sigma),$$

where

$$Q(\sigma)(v_1,\cdots,v_m) = (v_{\sigma^{-1}(1)},\cdots,v_{\sigma^{-1}(m)})$$

is Cartesian permutation operator with respect to $\sigma \in G$.

Theorem 1.1. *The linear operator* $C_{\mathfrak{X}}$ *is an orthogonal projection on* $U\otimes\times^m V$.

Definition 1.2. The image of $U\otimes\times^m V$ under $C_{\mathfrak{X}}$ is denoted by $V^{\mathfrak{X}}(G)$ and we call it the *generalized Cartesian symmetry class* over V associated with G and $\mathfrak{X}$.

If dim $U = 1$, then $V^{\mathfrak{X}}(G)$ reduces to $V^{\chi}(G)$, which is the Cartesian symmetry class associated with G and the irreducible character χ of G corresponding to the representation $\mathfrak{X}$ (see [2,4,6]). The elements of $V^{\mathfrak{X}}(G)$ of the form $C_{\mathfrak{X}}(u\otimes x^{\times})$ are called the generalized Cartesian symmetrized vectors.

2 Main results

The following theorem states the inner product two generalized symmetrized vectors in terms of the generalized trace function.

Theorem 2.1. *For all* $u,v\in U$ *and* $x^{\times},y^{\times}\in\times^m V$ *we have*

$$\left\langle C_{\mathfrak{X}}(u\otimes x^{\times}), v\otimes y^{\times}\right\rangle = \frac{1}{|G|}\langle Tr_{\mathfrak{X}}(A)u,v\rangle,$$

where $A=[a_{ij}]\in\mathbb{C}_{m\times m}$ *and* $a_{ij}=\langle x_i,y_j\rangle$.

In this paper, we will refer to the following lemma frequently.

Lemma 2.2. *Let* $\sigma\in G$, $u\in U$ *and* $x^{\times}\in\times^m V$. *Then*

$$C_{\mathfrak{X}}(u\otimes x^{\times}_{\sigma}) = C_{\mathfrak{X}}(\mathfrak{X}(\sigma)u\otimes x^{\times}).$$

Suppose $\mathbb{F}=\{u_1,\cdots,u_r\}$ and $\mathbb{E}=\{e_1,\cdots,e_n\}$ are orthonormal bases for unitary spaces U and V, respectively. For $1\le i\le n$ and $1\le j\le m$, let

$$e_{ij} = (\delta_{1j}e_i,\delta_{2j}e_i,\cdots,\delta_{mj}e_i)\in\times^m V.$$

Then the set

$$\mathbb{B} = \{u_k\otimes e_{ij}\mid 1\le k\le r, 1\le i\le n, 1\le j\le m\}$$

is an orthonormal basis of $U\otimes\times^m V$. Therefore,

$$V^{\mathfrak{X}}(G) = \langle C_{\mathfrak{X}}(u_k\otimes e_{ij})\mid 1\le k\le r, 1\le i\le n, 1\le j\le m\rangle.$$

The elements

$$C_{\mathfrak{X}}(u_k\otimes e_{ij}),\ 1\le k\le r,\ 1\le i\le n,\ 1\le j\le m$$

of $V^{\mathfrak{X}}(G)$ are called *the generalized Cartesian standard symmetrized vectors.*

Definition 2.3. For any $1 \leqslant j, s \leqslant m$, we define the linear map $T_{sj} : U \longrightarrow U$ by

$$T_{sj} = \frac{1}{|G_{sj}|} \sum_{\sigma \in G_{sj}} \mathfrak{X}(\sigma),$$

where

$$G_{sj} = \{\sigma \in G \mid \sigma(j) = s\}.$$

If G_{sj} is empty, then we define $T_{sj} = 0$. If $s = j$, then $G_{jj} = G_j$, the stabilizer of j in G and so $T_{jj} = T_j$, the linear map corresponding to j.

It is proved that T_j is an orthogonal projection on U. Also

$$\text{rank } T_j = \frac{1}{|G_j|} \sum_{\sigma \in G_j} \chi(\sigma),$$

where χ is the irreducible character of G corresponding to the representation $\mathfrak{X}$. So $T_j \neq 0$ if and only if $\sum_{\sigma \in G_j} \chi(\sigma) \neq 0$.

Theorem 2.4. *For any $1 \leqslant j, s \leqslant m, 1 \leqslant i, r \leqslant n, 1 \leqslant k, l \leqslant r$, we have*

$$\langle C_{\mathfrak{X}}(u_k \otimes e_{ij}), C_{\mathfrak{X}}(u_l \otimes e_{rs})\rangle = \begin{cases} 0 & s \nsim j \\ \delta_{ir} \dfrac{|G_{sj}|}{|G|} \langle T_{sj} u_k, u_l\rangle & s \sim j \end{cases}$$

In particular,

$$\| C_{\mathfrak{X}}(u_k \otimes e_{ij}) \|^2 = \frac{1}{[G : G_j]} \| T_j u_k \|^2 .$$

From the above Theorem, we deduce that $C_{\mathfrak{X}}(u_k \otimes e_{ij}) = 0$ if and only if $T_j u_k = 0$. For any $1 \leq k \leq r$, let

$$\Omega_k = \{1 \leq j \leq m \mid T_j u_k \neq 0\}.$$

Put $\Omega = \bigcup_{k=1}^{r} \Omega_k$. Then

$$\Omega = \{1 \leq j \leq m \mid [\chi, 1_{G_j}] \neq 0\},$$

where $[\ ,\ ]$ is the inner product of characters (see [3]).

Let $\bar{\mathcal{D}} = \mathcal{D} \cap \Omega$. For each $1 \leq j \leq m$ and $1 \leq i \leq n$, the subspace

$$V_{ij}^{\mathfrak{X}}(G) = \langle C_{\mathfrak{X}}(u_k \otimes e_{ij}) \mid 1 \leq k \leq r\rangle$$

is called the *generalized cyclic subspace*. If $\dim U = 1$, then $V_{ij}^{\mathfrak{X}}(G)$ reduces to $V_{ij}^{\chi}(G)$, the cyclic subspace associated with G and the irreducible character χ of G (see [2, 4, 6]).

Since $\langle \mathfrak{X}(\sigma)u_1 : \sigma \in G\rangle$ is a non-zero submodule of the irreducible $C[G]$-module U, we have $\langle \mathfrak{X}(\sigma)u_1 : \sigma \in G\rangle = U$. Therefore we can see that for every $1 \leq j \leq m$ and $1 \leq i \leq n$,

$$V_{ij}^{\mathfrak{X}}(G) = \langle C_{\mathfrak{X}}(u_1 \otimes e_{i\sigma(j)}) \mid \sigma \in G\rangle.$$

Now by using Theorem 2.4, we obtain

$$V^{\mathfrak{X}}(G) = \bigoplus_{i=1}^{n} \bigoplus_{j \in \bar{\mathcal{D}}} V_{ij}^{\mathfrak{X}}(G) \text{ (orthogonal)}.$$

The following theorem provides a formula for computing the dimension of the generalized cyclic subspace.

Theorem 2.5. *Let $\mathfrak{X}$ be an irreducible unitary representation of G over a unitary space U. Suppose $\mathfrak{X}$ affords the irreducible character χ of G. If $j \in \bar{\mathcal{D}}$ then*

$$\dim\ V_{ij}^{\mathfrak{X}}(G) = [\chi, 1_{G_j}].$$

Now we construct a basis for the generalized Cartesian symmetry class $V^{\mathfrak{X}}(G)$. Since $V^{\mathfrak{X}}(G) = \bigoplus_{i=1}^{n} \bigoplus_{j\in\overline{D}} V_{ij}^{\mathfrak{X}}(G)$, in order to find a basis for $V^{\mathfrak{X}}(G)$, it suffices to find a basis for the generalized cyclic subspace $V_{ij}^{\mathfrak{X}}(G)$ for every $1 \le i \le n$ and $j \in \bar{\mathcal{D}}$. Let $j \in \overline{\mathcal{D}}$ and $\dim V_{ij}^{\mathfrak{X}}(G) = s_j$. Since

$$V_{ij}^{\mathfrak{X}}(G) = \langle C_{\mathfrak{X}}(u_1 \otimes e_{i\sigma(j)}) \mid \sigma \in G\rangle,$$

so we can choose the ordered subset $\{j_1, \cdots, j_{s_j}\}$ from the orbit of j, such that the set

$$\{C_{\mathfrak{X}}(u_1 \otimes e_{ij_1}), \cdots, C_{\mathfrak{X}}(u_1 \otimes e_{ij_{s_j}})\}$$

is a basis for the generalized cyclic subspace $V_{ij}^{\mathfrak{X}}(G)$. Execute this procedure for each $k \in \bar{\mathcal{D}}$. If $\bar{\mathcal{D}} = \{j, k, l, \cdots\}$ $(j < k < l < \cdots)$, take

$$\hat{\mathcal{D}} = \{j_1, \cdots, j_{s_j}; k_1, \cdots, k_{s_k}; \cdots\}$$

to be ordered as indicated. Then

$$\{C_{\mathfrak{X}}(u_1 \otimes e_{ij}) \mid 1 \le i \le n,\ j \in \hat{\mathcal{D}}\}$$

is a basis of $V^{\mathfrak{X}}(G)$. Hence

$$\dim V^{\mathfrak{X}}(G) = (\dim V)|\hat{\mathcal{D}}| = n\sum_{j\in\bar{\mathcal{D}}} s_j = n\sum_{j\in\bar{\mathcal{D}}} [\chi, 1_{G_j}].$$

If $\mathfrak{X}$ is a linear representation of G, then $\dim V_{ij}^{\mathfrak{X}}(G) = 1$ and the set

$$\{C_{\mathfrak{X}}(u_1 \otimes e_{ij}) \mid 1 \le i \le n,\ j \in \hat{\mathcal{D}}\}$$

is an orthogonal basis of $V^{\mathfrak{X}}(G)$ (such representations of G are called *o.b.-representations*).

3 Open problem

Problem 3.1. Characterize the subgroups of S_m whose irreducible representations are all o.b.-representations.

Problem 3.2. Let G be a subgroup of S_m and $\mathfrak{X}$ be an irreducible unitary representation of G. Determine the conditions on $\mathfrak{X}$ such that $V^{\mathfrak{X}}(G)$ has an orthogonal basis consisting the generalized Cartesian standard symmetrized vectors.

4 Conclusion

In this paper, we introduce the generalized Cartesian symmetry class over V that is associated with G and $\mathfrak{X}$. We provide a formula for dimension of the generalized cyclic subspace $V_{ij}^{\mathfrak{X}}(G)$ and present a basis for the generalized symmetry class $V^{\mathfrak{X}}(G)$. Additionally, we identify some open problems for further research in this area.

References

[1] S.S. Gholami and Y. Zamani, *Generalized Cartesian symmetry classes*, Preprint, https://arxiv.org/abs/2304.11917.

[2] S.S. Gholami and Y. Zamani, *Cartesian symmetry classes associated with certain groups*, Preprint, https://arxiv.org/abs/2304.13990.

[3] M. Isaacs, *Character Theory of Finite Groups*, Academic Press, 1976.

[4] T.G. Lei, Notes on Cartesian symmetry classes and generalized trace functions, *Linear Algebra Appl.*, 292 (1999) 281-288.

[5] T.G. Lei, Generalized Schur functions and generalized decomposable symmetric tensors, *Linear Algebra Appl.*, 263 (1997), 311-332.

[6] Y. Zamani and M. Shahryari, On the dimensions of Cartesian symmetry classes, *Asian-Eur. J. Math.*, 5 (2012), No. 3, Article ID 1250046 (7 pages).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On Fuglede-Putnam property of Moore-Penrose inverse

Javad Farokhi Ostad*

Department of Basic Sciences, Birjand University, Birjand, Iran

Abstract

In this paper, we intend to investigate the relationship between some types of bounded linear operators on the Hilbert space H and their Moore-Penrose inverse deals with terms of the Fugled-Putnam property. It has been shown that if two bi-dagger operators apply to the Fugled-Putnam property, then their Moore-Penrose inverse also applies to this property.

Keywords: Fugled-Putnam property, Moore-Penrose inverse, bi-dagger operator

Mathematics Subject Classification [2010]: 47A05, 47A06

1 Introduction

Let H and K be Hilbert spaces and let $\mathcal{B}(H,K)$ denote the algebra of all bounded linear operators from H to K. Several useable definitions are only briefly mentioned. The operator $T \in \mathcal{B}(H)$ is called sel-adjoint, normal, isometry, unitary and projection when $T = T^*$, $TT^* = T^*T$, $T^*T = I$, $TT^* = T^*T = I$ and $T^2 = T = T^*$ respectively.

The Fuglede-Putnam theorem states that; if T and S are bounded normal operators on a Hilbert space H such that for some nonzero operator $X \in \mathcal{B}(H)$, $TX = XS$, then $T^*X = XS^*$.

On the other hand, the Moore-Penrose inverse is a generalization of the inverse of an operator on a Hilbert space. Given a bounded linear operator T on a Hilbert space H, its Moore-Penrose inverse $T^\dagger$ is defined as the unique operator satisfying four properties:

1. $T\,T^\dagger T = T$,
2. $T^\dagger\,T\,T^\dagger = T^\dagger$,

*Speaker. Email address: J.farrokhi@birjandut.ac.ir

3. $(T\,T^{\dagger})^{*} = T\,T^{\dagger}$,

4. $(T^{\dagger}\,T)^{*} = T^{\dagger}T$.

The Moore-Penrose inverse has many applications in linear algebra, functional analysis, and signal processing.

The Fuglede-Putnam theorem and the Moore-Penrose inverse are two important concepts in operator theory and functional analysis. Although there are some connections between them, more research is needed to establish a new and novel relation between these topics.

For example, it has been shown that if T and S are self-adjoint operators on a Hilbert space H such that $TS = ST$ and $T^{\dagger}S$ is invertible, then T and S are simultaneously diagonalizable by a unitary operator. This result has important implications for the spectral theory of self-adjoint operators and the theory of quantum mechanics.

Moreover, if T and S are bounded linear operators on a Hilbert space H such that TS is invertible, then $(TS)^{\dagger} = S^{\dagger}T^{\dagger}$, which is said the reverse order law satisfied. This result can be used to derive some properties of the Fuglede-Putnam theorem and related topics.

2 The Main Results

The operator $T \in \mathcal{B}(H)$ is called EP (stands for Equal Projections), hypo-EP, star dagger and bi-dagger when ran(T) and ran(T*) have the same closure, $T^{\dagger}T - TT^{\dagger}$ is a positive operator, $T^{*}T^{\dagger} = T^{\dagger}T^{*}$ and $(T^{2})^{\dagger} = (T^{\dagger})^{2}$, respectively. Also, the operator $T \in \mathcal{B}(H)$ is said to be compact if it can be written in the form $T = \Sigma_{n=1}^{\infty}\lambda_{n}\,\langle f_{n}, .\rangle\, g_{n}$, where $\{f_{1}, f_{2}, \dots\}$ and $\{g_{1}, g_{2}, \dots\}$ are orthonormal sets (not necessarily complete), and $\lambda_{1}, \lambda_{2}, \dots$ is a sequence of positive numbers with limit zero, called the singular values of the operator.

Theorem 2.1. *Let T and S in $\mathcal{B}(H)$ be compact operator which the reverse order law hold for them. If T and S satisfy the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

Remark 2.2. It is not true in general that if T and S are two compact operators satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.

One such counterexample is given by considering the operators T and S on the Hilbert space $l^{2}(\mathbb{N})$ defined as follows:

$T(x_{1}, x_{2}, x_{3}, ...) = (0, x_{1}, \frac{x_{2}}{2}, \frac{x_{3}}{3}, ...)$ and $S(x_{1}, x_{2}, x_{3}, ...) = (\frac{x_{1}}{2}, \frac{x_{2}}{3}, \frac{x_{3}}{4}, ...)$

It can be shown that T and S are compact operators satisfying the Fuglede-Putnam property $TS = ST$. However, their Moore-Penrose inverses are given by:

$T^{\dagger}(x_{1}, x_{2}, x_{3}, ...) = (0, x_{1}, 2x_{2}, 3x_{3}, ...)$ and $S^{\dagger}(x_{1}, x_{2}, x_{3}, ...) = (\frac{x_{1}}{2}, 3\frac{x_{2}}{4}, 4\frac{x_{3}}{5}, ...)$

It can be verified that $T^{\dagger}$ and $S^{\dagger}$ do not satisfy the Fuglede-Putnam property.

Theorem 2.3. *Let $T, S \in \mathcal{B}(H)$ have closed ranges. If T and S are star-dagger and satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

Remark 2.4. If T or S is not star-dagger, the above result may not be true. Let

$$T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \; S = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \text{ and } X = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that, in this case $T^\dagger = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $S^\dagger = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and so $T^\dagger X \neq XS^\dagger$.

The following Theorem has important applications in the theory of spectral flow and index theory.

Theorem 2.5. *Let T and S in $\mathcal{B}(H)$ be two operators which their ranges are closed and also satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

This result is a consequence of the fact that the Moore-Penrose inverse of a bounded operator is also bounded.

In particular, if T and S are two self-adjoint bounded operators satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses are also self-adjoint and satisfy the same property.

Theorem 2.6. *Let T and S in $\mathcal{B}(H)$ be two positive operators satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses may not necessarily satisfy the same property.*

This can be seen by considering the example of $T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and $S = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$. Then, both T and S are positive operators and satisfy the Fuglede-Putnam property. However, their Moore-Penrose inverses are $T^\dagger = T$ and $S^\dagger = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$, which do not satisfy the Fuglede-Putnam property.

Theorem 2.7. *Let T and S in $\mathcal{B}(H)$ be two operators with closed ranges. Let T and S be two star-dag operators, also satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

Theorem 2.8. *Let T and S in $\mathcal{B}(H)$ be two normal operators satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

This can be seen by using the spectral theorem for normal operators. Let $T = U|T|U^*$ and $S = V|S|V^*$ be the polar decompositions of T and S, where U and V are unitary operators and $|T|$ and $|S|$ are positive operators.

Theorem 2.9. *Let T and S in $\mathcal{B}(H)$ have closed ranges satisfy the Fuglede-Putnam property if and only if $|T|$ and $|S|$ satisfy the same property.*

Furthermore, it can be shown that the Moore-Penrose inverse of a normal operator is given by $T^\dagger = U|T|^\dagger U^*$, where $|T|^\dagger$ is the Moore-Penrose inverse of $|T|$. Since $|T|$ satisfies the Fuglede-Putnam property if and only if T does, it follows that $T^\dagger$ also satisfies the same property.

An operator T is bi-normal if and only if $TT^* = T^*T$,

Theorem 2.10. *Let T and S in $\mathcal{B}(H)$ be two bi-normal and bi-dagger operators. If T and S satisfying the Fuglede-Putnam property, then their Moore-Penrose inverses also satisfy the same property.*

There are two examples of a bi-normal operator that satisfies the Fuglede-Putnam property but whose Moore-Penrose inverse does not is the following:

Example 2.11. Let T be the operator on l^2 given by $T(x_1, x_2, x_3, ...) = (0, x_1, \frac{x_2}{2}, \frac{x_3}{3}, ...)$. It can be shown that T is bi-normal and satisfies the Fuglede-Putnam property. However, its Moore-Penrose inverse does not exist.

Another example is the following:

Example 2.12. Let S be the operator on $L^2([0,1])$ given by $S(f)(x) = \int_0^1 f(t)\sin(\pi(x-t))dt$. It can be shown that S is bi-normal and satisfies the Fuglede-Putnam property. However, its Moore-Penrose inverse does not satisfy the Fuglede-Putnam property.

Acknowledgement

I would like to express my gratitude to the organizers of the 12th Seminar on Linear Algebra and its Applications, Sahand University of Technology.

References

[1] J. Farokhi-ostad and M. Mohammadzadeh karizaki, *The reverse order law for EP modular operators*, J. Math. Computer Sci., 16 (2016), 412–418.

[2] A. R. Janfada and J. Farokhi-ostad, *Convergence of Bounded Sequences in Hilbert C^*-Modules*, Lobachevskii Journal of Mathematics., 43(9), (2022), 2493–2500.

[3] T. Furuta, Invitation to Linear Operators From Matrices to Bounded Linear Operators on a Hilbert Space, CRC Press, 2001.

[4] Jalaeian M, Mohammadzadeh Karizaki M, Hassani M. *Conditions that the product of operators is an EP operator in Hilbert C^*-module*, Linear Multilinear Algebra., 68(10), (2020), 1990–2004.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Jordan triple $*$-derivations on prime $*$-algebras

Ali Taghavi* and Yousef Ahmadi

Department of Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, P. O. Box 47416-1468, Babolsar, Iran.

Abstract

Let $\mathcal{A}$ be a unital prime $*$-algebra containing a non-trivial projection P_1. In this paper, it is shown that a map $\Phi : \mathcal{A} \to \mathcal{A}$ is a multiplicative $*$-Jordan triple derivation if and only if Φ is an additive $*$-derivation.

Keywords: Jordan triple derivation, Prime $*$-algebra, Additive map
Mathematics Subject Classification [2010]: 46J10, 47B48

1 Introduction

Let $\mathcal{R}$ be a $*$-ring. For $A, B \in \mathcal{R}$, the $*$-Jordan product and bi-skew Jordan product are defined as $A \diamond B = AB + BA^*$ and $A * B = AB^* + BA^*$, respectively. These products have recently attracted the attention of many researchers ([1–6]). For more examples of maps preserving triple product, one can refer to [1–6]. We say that a (non necessarily linear) mapping Φ with the property of $\Phi(A \diamond B) = \Phi(A) \diamond B + A \diamond \Phi(B)$ is a Jordan $*$-derivation. It should be noted that $\diamond$ and $*$ are not necessarily associative. For clarifying this, we set $A \diamond B \diamond C := (A \diamond B) \diamond C$ and $A * B * C := (A * B) * C$. We should mention here whenever we say that Φ is a derivation, it means that the identity $\Phi(AB) = \Phi(A)B + A\Phi(B)$ holds for all $A, B \in \mathcal{A}$. Set $\mathcal{A}_s = \{A \in \mathcal{A} : A^* = A\}$ and $\mathcal{A}_{sk} = \{A \in \mathcal{A} : A^* = -A\}$.
In [3], Taghavi et al., showed the following result.

Theorem 1.1. *Let Φ preserves triple $*$-Jordan derivation on prime $*$-algebra $\mathcal{A}$, i.e.,*

$$\Phi(A \diamond B \diamond C) = \Phi(A) \diamond B \diamond C + A \diamond \Phi(B) \diamond C) + A \diamond B \diamond \Phi(C), \tag{1}$$

for all $A, B, C \in \mathcal{A}$, where $A \diamond B = AB + BA^*$, then Φ is additive. Moreover, if $\Phi(\alpha I)$ is self-adjoint for $\alpha \in \{1, i\}$, then Φ is a $*$-derivation.

In this paper, we replace $*$-Jordan product with bi-skew Jordan products in Theorem 1.1. It is proved that if $\Phi : \mathcal{A} \to \mathcal{A}$ is a multiplicative bi-skew Jordan triple derivations, i.e.,

$$\Phi(A * B * C) = \Phi(A) * B * C + A * \Phi(B) * C + A * B * \Phi(C), \tag{2}$$

for all $A, B, C \in \mathcal{A}$, where $A * B = AB^* + BA^*$, then Φ preserves self-adjoint elements. Therefore, the two relations (1) and (2) will be equivalent for every $A \in \mathcal{A}$ and $B, C \in \mathcal{A}_s$.

*Speaker. Email address: taghavi@umz.ac.ir

This result makes possible to unify the conclusions in a single statement which implies the above two results. In fact, it is shown that if map Φ on a unital prime $*$-algebra $\mathcal{A}$ containing a non-trivial projections satisfies the conditions in (1) or (2) for every $A, B, C \in \mathcal{A}$, then Φ is an additive $*$-derivation.

We recall that $\mathcal{A}$ is prime if for $A, B \in \mathcal{A}$ the condition $A\mathcal{A}B = \{0\}$, implies $A = 0$ or $B = 0$.

2 Main results

Main results of this paper are as following.

Theorem 2.1. *Let $\mathcal{A}$ be a unital prime $*$-algebra containing a non-trivial projection P_1. Then a (non-necessarily linear) mapping $\Phi : \mathcal{A} \to \mathcal{A}$ satisfies (2) for every $A, B, C \in \mathcal{A}$ if and only if Φ is an additive $*$-derivation.*

Clearly, the sufficient implication is obvious.

Let P_1 be a nontrivial projection in $\mathcal{A}$ and $P_2 = I - P_1$. Denote $\mathcal{A}_{ij} = P_i\mathcal{A}P_j,\ i, j = 1, 2$, then $\mathcal{A} = \sum_{i,j=1}^{2} \mathcal{A}_{ij}$. Denote $\mathbf{A}_{12} = P_1AP_2 + P_2AP_1$ whenever $A \in \mathcal{A}_{sa}$. Hence, for every $A \in \mathcal{A}_{sa}$ we may write $A = A_{11} + \mathbf{A}_{12} + A_{22}$. In all what follows, when we write A_{ij} it indicates that $A_{ij} \in \mathcal{A}_{ij}$.To prove the additivity of Φ on $\mathcal{A}$ we shall use the above partition of $\mathcal{A}$ and we shall establish some claims proving that Φ is additive on each $\mathcal{A}_{ij},\ i, j = 1, 2$. We prove this theorem in several steps.

Step 1. $\Phi(0) = 0$.

Step 2. Φ preserves self-adjoint and skew self-adjoint elements.

Step 3. For every $A \in \mathcal{A}_{sa}$, $\Phi(iA) = i\Phi(A)$.

Step 4. Φ is additive on $\mathcal{A}_{sa}$ and $\mathcal{A}_{sk}$.

Step 5. Φ is additive on $\mathcal{A}$.

In the rest of this paper we show that Φ is a $*$-derivation.

Step 6. Φ preserves the involution.

Step 7. Φ is a derivation.

Corollary 2.2. *Let $\mathcal{A}$ be a unital prime $*$-algebra containing a non-trivial projection P_1. Then a (non-necessarily linear) mapping $\Phi : \mathcal{A} \to \mathcal{A}$ satisfies (1) for every $A, B, C \in \mathcal{A}$ if and only if Φ is an additive $*$-derivation.*

References

[1] C. Li, F. Lu, X. Fang, *Nonlinear $\xi-$Jordan $*$-derivations on von Neumann algebras*, Linear and Multilinear Algebra. 62 (2014) 466-473.

[2] C. Li, F. Lu, X. Fang, Nonlinear mappings preserving product $XY + YX^*$ on factor von Neumann algebras, Linear Algebra Appl. 438 (2013) 2339-2345.

[3] A. Taghavi, H. Rohi, V. Darvish, *Non-linear $*$-Jordan derivations on von Neumann algebras*, Linear Multilinear Algebra 64 (2016) 426–439.

[4] W. Yu, J. Zhang, *Nonlinear $*$-Lie derivations on factor von Neumann algebras*, Linear Algebra Appl. 437 (2012) 1979-1991.

[5] F. Zhao, C. Li, *Nonlinear $*$-Jordan triple derivations on von Neumann algebras*, Math. Slovaca 68 (2018) 163-170.

[6] F. Zhao, C. Li, *Nonlinear maps preserving the Jordan triple $*$-product between factors*, Indag. Math. 99 (2018) 619-627.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

An invitation to some operator entropies

Ismail Nikoufar*

Department of Mathematics, Payame Noor University, Tehran, Iran

Abstract

In this paper, we find upper and lower bounds of some operator entropies. We also refine and improve the lower and upper bound of these operator entropies. As a consequence of our result, we improve the bounds of the relative operator entropy announced by Fujii and Kamei.

Keywords: perspective function, relative operator entropy, Tsallis relative operator entropy, generalized relative operator, operator geometric mean

Mathematics Subject Classification [2010]: 47A63, 81P45, 15A39

1 Introduction

There is a classical perspective function associated to the function f which was defined on a convex set $\mathcal{C} \subseteq \mathbb{R}^n$. The classical perspective function is a function of two variable on the subset

$$K := \{(t, s) : s > 0, \quad \frac{t}{s} \in \mathcal{C}\} \subseteq \mathbb{R}^{n+1}.$$

This function was defined by $P_f(t, s) := f(\frac{t}{s})s$. Marechal defined the generalized perspective function by $P_{f\Delta g}(x, y) := f(\frac{x}{g(y)})g(y)$ on $\mathbb{R}^{n+m}$ for functions $f : \mathbb{R}^n \to (-\infty, \infty)$ and $g : \mathbb{R}^m \to (0, \infty)$ [2]. This generalization of perspectivity of functions has a natural operator version.

Effros [4] considered an operator version of perspectivity for commuting operators and proved in this way that the perspective of an operator convex function is operator convex as a function of two variables. Let f and h be real valued continuous functions on the closed interval $\mathbb{I}$. By recalling that if for every continuous function f, $f(A)$ commutes with every operator commuting with A (including A itself) and by restricting to positive commuting operators, Effros defined the generalized perspective function by

$$P_{f\Delta h}(A, B) := f(\frac{A}{h(B)})h(B).$$

We defined in [2] a fully noncommutative generalized perspective of two variable (associated to f and h) by choosing an appropriate ordering and by setting

$$P_{f\Delta h}(A, B) := h(B)^{1/2} f(h(B)^{-1/2} A h(B)^{-1/2}) h(B)^{1/2},$$

*Speaker. Email address: nikoufar@pnu.ac.ir

where A is a self–adjoint operator and B is a strictly positive operator on a Hilbert space $\mathcal{H}$ with spectra in the closed interval $\mathbb{I}$ containing 0. The perspective of the function f is denoted by P_f and is defined by $P_f(A,B) := B^{1/2} f(B^{-1/2}AB^{-1/2})B^{1/2}$. In this approach all references to commutativity can be removed and this contribution can surely be affected quantum information theory and quantum statistical mechanics. We then proved the necessary and sufficient conditions for jointly convexity of a fully noncommutative perspective and generalized perspective function.

Generalized entropies are used as alternate measures of an informational content. In particular, they may be used to study properties of the standard entropy in more general setting.

The notion of relative operator entropy was considered on strictly positive operators in noncommutative information theory [1] as follows:

$$S(A|B) := A^{\frac{1}{2}}(\log A^{-\frac{1}{2}}BA^{-\frac{1}{2}})A^{\frac{1}{2}}.$$

This is an extension of the operator entropy defined by Nakamura and Umegaki and the relative operator entropy introduced by Umegaki [5]. More generally, the generalized relative operator entropy for strictly positive operators A, B and $q \in \mathbb{R}$ defined by Furuta by setting

$$S_q(A|B) = A^{1/2}(A^{-1/2}BA^{-1/2})^q(\log A^{-1/2}BA^{-1/2})A^{1/2}.$$

In particular, when $q = 0$, we have $S_0(A|B) = S(A|B)$. Using the notion of generalized relative operator entropy, Furuta obtained the parametric extension of operator Shannon inequality and its reverse one. We also notify that the relative operator entropy $S(A|B)$ is perspective of $\log t$ in the sense that $S(A|B) = P_{\log t}(B|A)$ and the generalized relative operator entropy $S_q(A|B)$ is perspective of $t^q \log t$ in the sense that $S_q(A|B) = P_{t^q \log t}(B|A)$. For strictly positive operators A, B and $0 < \lambda \leq 1$,

$$T_\lambda(A|B) := \frac{A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^\lambda A^{\frac{1}{2}} - A}{\lambda}$$

is called Tsallis relative operator entropy between A and B. This notion can be rewritten as

$$T_\lambda(A|B) = A^{\frac{1}{2}} \ln_\lambda(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})A^{\frac{1}{2}},$$

where $\ln_\lambda X \equiv \frac{X^\lambda - 1}{\lambda}$ for the positive operator X. Yanagi et. al [6] proved some properties of Tsallis relative operator entropy and the generalized Shannon inequalities. Some other operator inequalities related to Tsallis relative operator entropy were also proved by Furuichi. Various generalizations of the Shannon inequalities have played an important role in classical information theory. Surprisingly, it has been discovered that many of these inequalities have operator generalizations, in which one replaces random variables by Hilbert space operators. The latter are the variables of quantum thermodynamics and quantum information theory.

We introduced in [3] the notions of relative operator (α,β)-entropy (two parameter relative operator entropy) and Tsallis relative operator (α,β)-entropy as following:

$$S_{\alpha,\beta}(A|B) = A^{\frac{\beta}{2}}(A^{-\frac{\beta}{2}}BA^{-\frac{\beta}{2}})^\alpha(\log A^{-\frac{\beta}{2}}BA^{-\frac{\beta}{2}})A^{\frac{\beta}{2}}$$

for strictly positive operators A, B and real numbers α, β and

$$T_{\alpha,\beta}(A|B) := A^{\frac{\beta}{2}} \ln_\alpha(A^{-\frac{\beta}{2}}BA^{-\frac{\beta}{2}})A^{\frac{\beta}{2}}$$

for strictly positive operators A, B and real numbers $\alpha \neq 0, \beta$. We applied a perspective approach to prove the convexity or concavity of these notions, under certain conditions concerning α and β. In particular, we have $S_{q,1}(A|B) = S_q(A|B)$, $S_{0,1}(A|B) = S(A|B)$, $T_{\lambda,1}(A|B) = T_\lambda(A|B)$, and

$$\lim_{\alpha\to 0} T_{\alpha,\beta}(A|B) = S_{0,\beta}(A|B).$$

2 Main results

In this section, we find upper and lower bounds of relative operator (α,β)-entropy and Tsallis relative operator (α,β)-entropy according to operator (α,β)-geometric mean.

We announced the notion of operator (α,β)-geometric mean for real numbers α, β as a generalization of the notion of operator α-geometric mean of two strictly positive operators A, B as following

$$A\#_{(\alpha,\beta)}B := A^{\frac{\beta}{2}}(A^{-\frac{\beta}{2}}BA^{-\frac{\beta}{2}})^{\alpha}A^{\frac{\beta}{2}}.$$

Note that every operator α-geometric mean is an operator $(\alpha,1)$-geometric mean. By using concavity and convexity of operator (α,β)-geometric mean of strictly positive operators A, B, we gave the simplest proof of the well-known Lieb concavity theorem and Ando convexity theorem. We would remak that $A\#_{(-1,\beta)}B = A^{\beta}B^{-1}A^{\beta}$, $A\#_{(0,\beta)}B = A^{\beta}$, and $A\#_{(1,\beta)}B = B$.

Theorem 2.1. *Let r, q, k, and h be real valued functions on the closed interval $\mathbb{I}$ such that $h > 0$. If $r(t) \leq q(t) \leq k(t)$ for $t \in \mathbb{I}$, then*

$$P_{r\Delta h}(B, A) \leq P_{q\Delta h}(B, A) \leq P_{k\Delta h}(B, A)$$

for strictly positive operator A and self adjoint operator B.

Corollary 2.2. *For any strictly positive operators A and B, $0 < \lambda \leq 1$ and $\beta > 0$, we have*

$$T_{-\lambda,\beta}(A, B) \leq S_{0,\beta}(A, B) \leq T_{\lambda,\beta}(A, B). \tag{1}$$

Remark 2.3. We note that inequalities (1) recover the inequalities obtained by Furuichi, if we put $\beta = 1$. Then, we have

$$T_{-\lambda}(A, B) \leq S(A, B) \leq T_\lambda(A, B)$$

for strictly positive operator A and self adjoint operator B.

Corollary 2.4. *For any strictly positive operators A and B, $0 < \lambda \leq 1$ and $\beta > 0$,*

$$A\#_{(0,\beta)}B - A\#_{(-1,\beta)}B \leq T_{\lambda,\beta}(A, B) \leq A\#_{(1,\beta)}B - A\#_{(0,\beta)}B. \tag{2}$$

Moreover, $T_{\lambda,\beta}(A, B) = 0$ if and only if $A^{\beta} = B$.

Corollary 2.5. *For any strictly positive operators A and B, $0 < \lambda \leq 1$, $\beta > 0$ and any positive real number s, we have the following inequalities:*

$$A\#_{(\lambda,\beta)}B - \frac{1}{s}A\#_{(\lambda-1,\beta)}B + (\ln_\lambda \frac{1}{s})A\#_{(0,\beta)}B \leq T_{\lambda,\beta}(A, B) \tag{3}$$

$$T_{\lambda,\beta}(A, B) \leq \frac{1}{s}A\#_{(1,\beta)}B - (\ln_\lambda \frac{1}{s})A\#_{(\lambda,\beta)}B - A\#_{(0,\beta)}B. \tag{4}$$

Remark 2.6. We notice that inequalities (3) and (4) generalize the inequalities obtained by Furuta:

$$(1-\log s)A\#_{(0,\beta)}B-\frac{1}{s}A\#_{(-1,\beta)}B\leq S_{0,\beta}(A,B)\leq\frac{1}{s}A\#_{(1,\beta)}B+(\log s-1)A\#_{(0,\beta)}B \quad (5)$$

as $\lambda\to 0$. Moreover, if we put $s=1$, then we have

$$A\#_{(0,\beta)}B-A\#_{(-1,\beta)}B\leq S_{0,\beta}(A,B)\leq A\#_{(1,\beta)}B-A\#_{(0,\beta)}B$$

for strictly positive operator A and self adjoint operator B, which generalize the inequalities obtained by Fujii and Furuichi, cf. (2).

Corollary 2.7. *For any strictly positive operators A and B, $0<\alpha\leq 1$ and $\beta>0$,*

$$A\#_{(0,\beta)}B-A\#_{(-1,\beta)}B\leq S_{\alpha,\beta}(A,B)\leq A\#_{(\alpha+1,\beta)}B-A\#_{(\alpha,\beta)}B. \quad (6)$$

Corollary 2.8. *For any strictly positive operators A and B, $0<\alpha\leq 1$, $\beta>0$ and any positive real number s, we have the following inequalities:*

$$A\#_{(\alpha,\beta)}B-\frac{1}{s}A\#_{(\alpha-1,\beta)}B+(\ln_\alpha\frac{1}{s})A\#_{(0,\beta)}B\leq S_{\alpha,\beta}(A,B) \quad (7)$$

$$S_{\alpha,\beta}(A,B)\leq\frac{1}{s}A\#_{(\alpha+1,\beta)}B-(\ln_\alpha\frac{1}{s})A\#_{(2\alpha,\beta)}B-A\#_{(\alpha,\beta)}B. \quad (8)$$

Remark 2.9. In the inequalities (7) and (8), if we put $s=1$, then we have

$$A\#_{(\alpha,\beta)}B-A\#_{(\alpha-1,\beta)}B\leq S_{\alpha,\beta}(A,B)\leq A\#_{(\alpha+1,\beta)}B-A\#_{(\alpha,\beta)}B$$

and also as $\alpha\to 0$ we get (5).

3 Improvements of the results

It is well known that the following Hermite-Hadamard integral inequality for convex functions plays a central role in the theory of convex functions and includes a basic property of convex functions. As can be seen in a large number of research articles and books devoted to this field, the Hermite-Hadamard inequality is the first fundamental result for convex functions with a natural geometrical interpretation and has attracted much interest in elementary mathematics with many applications.

Lemma 3.1. *Let f be a real-valued function which is convex on the interval $[a,b]$. Then,*

$$f(\frac{a+b}{2})\leq\frac{1}{b-a}\int_a^b f(t)dt\leq\frac{f(a)+f(b)}{2}.$$

Theorem 3.2. *For any positive invertible operators A and B with $A^\beta\leq B$, and $\alpha\geq 0,\beta>0$, we have*

$$2\big(1-2A^\beta(A^\beta+B)^{-1}\big)A\#_{(\alpha,\beta)}B\leq S_{\alpha,\beta}(A|B)\leq\frac{1}{2}(A\#_{(\alpha+1,\beta)}B-A\#_{(\alpha-1,\beta)}B). \quad (9)$$

For $B\leq A^\beta$ the reverse inequalities in (9) hold.

Corollary 3.3. *Let A, B be two positive invertible operators and $\alpha\geq 0,\beta>0$.*

(i) If $A^\beta \leq B$, then

$$\begin{aligned} A\#_{(\alpha,\beta)}B - A\#_{(\alpha-1,\beta)}B &\leq 2\big(1 - 2A^\beta(A^\beta + B)^{-1}\big)A\#_{(\alpha,\beta)}B \\ &\leq S_{\alpha,\beta}(A|B) \\ &\leq \frac{1}{2}(A\#_{(\alpha+1,\beta)}B - A\#_{(\alpha-1,\beta)}B) \\ &\leq A\#_{(\alpha+1,\beta)}B - A\#_{(\alpha,\beta)}B. \end{aligned} \tag{10}$$

(ii) If $B \leq A^\beta$, then

$$\begin{aligned} A\#_{(\alpha,\beta)}B - A\#_{(\alpha-1,\beta)}B &\leq \frac{1}{2}(A\#_{(\alpha+1,\beta)}B - A\#_{(\alpha-1,\beta)}B) \\ &\leq S_{\alpha,\beta}(A|B) \\ &\leq 2\big(1 - 2A^\beta(A^\beta + B)^{-1}\big)A\#_{(\alpha,\beta)}B \\ &\leq A\#_{(\alpha+1,\beta)}B - A\#_{(\alpha,\beta)}B. \end{aligned} \tag{11}$$

In particular, when we put $\beta = 1$ in (10), we derive a refined and sharp bounds for the generalized relative operator entropy with $A \leq B$:

$$\begin{aligned} A\#_\alpha B - A\#_{\alpha-1}B &\leq 2\big(1 - 2A(A + B)^{-1}\big)A\#_\alpha B \\ &\leq S_\alpha(A|B) \\ &\leq \frac{1}{2}(A\#_{\alpha+1}B - A\#_{\alpha-1}B) \\ &\leq A\#_{\alpha+1}B - A\#_\alpha B. \end{aligned} \tag{12}$$

Moreover, when we put $\alpha = 0$ in (12) we observe that the lower and upper bound of the relative operator entropy with $A \leq B$ is sharper than the bounds discovered by Fujii and Kamei:

$$A - AB^{-1}A \leq 2(A - 2A(A + B)^{-1}A) \leq S(A|B) \leq \frac{1}{2}(B - AB^{-1}A) \leq B - A. \tag{13}$$

In light of Corollary 3.3 (ii) and putting $\alpha = 0, \beta = 1$ in (11), we realize the other sharp bounds of the relative operator entropy with $B \leq A$:

$$A - AB^{-1}A \leq \frac{1}{2}(B - AB^{-1}A) \leq S(A|B) \leq 2(A - 2A(A + B)^{-1}A) \leq B - A. \tag{14}$$

4 Conclusion

We introduced some operator entropies and gave the upper and lower bounds of them. We also refined and improved the lower and upper bound of these operator entropies. As a new consequence of our result, we improved the bounds of the relative operator entropy announced by Fujii and Kamei.

References

[1] J. I. Fujii and E. Kamei, Relative operator entropy in noncommutative information theory, *Math. Japonica*, 34 (1989), 341–348.

[2] I. Nikoufar, On operator inequalities of some relative operator entropies, *Adv. Math.*, 259 (2014), 376–383.

[3] I. Nikoufar, Convexity of parameter extensions of some relative operator entropies with a perspective approach, *Glasgow Math. J.*, 62(3) (2020), 737–744.

[4] I. Nikoufar, Improved operator inequalities of some relative operative entropies, *Positivity*, 24(1) (2020), 241–251.

[5] H. Umegaki, Conditional expectation in operator algebra IV (entropy and information), *Kodai Math. Sem. Rep.*, 14 (1962), 59–85.

[6] K. Yanagi, K. Kuriyama, and S. Furuichi, Generalized Shannon inequalities based on Tsallis relative operator entropy, *Linear Alg. Appl.*, 394 (2005), 109–118.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The Legendre pseudospectral method for a time-fractional optimal control problem

Farnaz Kheirkhah* and Mojtaba Hajipour

Department of Mathematics, Sahand University of Technology, Tabriz, Iran

Abstract

In this article a numerical scheme based on the Legendrepseudo spectral and finite diffference methods is presented to solve an optimal control problem (OCP) governed by a fractional diffusion equation. Using the pseudospectral derivative matrices, this OCP reduces to a nonlinear optimization problem (NOP). Moreover, it is shown that the Karush-Kuhn-Tucker conditions of the derived NOP are exactly equivalent to the discretized form of its Ponteryagin optimal conditions. To demonstrate the efficiency of the proposed method, some numerical results are provided.

Keywords: Optimal control problem, Time-fractional diffusion equation, Legendre pseudospectral method.

Mathematics Subject Classification [2010]: 49J15, 49J20.

1 Introduction

The main purpose of the present paper is to develop an accurate numerical scheme to solve an optimal control problem governed by a time-fractional diffusion equation in the following form

$$\min_{q\in\mathcal{A}} \; J(u,q) = \frac{1}{2}\left\|u-u_d\right\|^2_{L^2(I_T)} + \frac{\gamma}{2}\left\|q\right\|^2_{L^2(I_T)} \tag{1}$$

subject to

$$\begin{aligned} {}_0^C D_t^\alpha u(x,t) - u_{xx}(x,t) &= f(x,t)+q(x,t), \qquad (x,t)\in I_T, \\ u(x,t) &= 0, \qquad (x,t)\in \Gamma_T, \\ u(x,0) &= g(x), \qquad x\in I, \end{aligned} \tag{2}$$

Where $I=(a,b)$, $I_T = I\times(0,T)$, $\Gamma_T = \{a,b\}\times(0,T)$, q is the control variable, u is the state variable and u_d is desird state. also the term ${}_0^C D_t^\alpha u(x,t)$ denotes the left Caputo fractional derivative of order $\alpha(0<\alpha<1)$ of the state u defined by:

$${}_0^C D_t^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)}\int_0^t \frac{f'(s)}{(t-s)^\alpha} ds, \tag{3}$$

*Speaker. Email address: f_kheirkhah98@sut.ac.ir

During this decade, the fractional (or non-integer) calculus has attracted increasing attention in various fields of science and engineering. The fractional derivative simultaneously possesses memory which makes it a powerful tool in modeling complex dynamical systems related to non-locality and memory effect [1]. In the past decade, compared with optimal control problems governed by differential equations with integer derivatives [2], only limited research was committed to the theoretical analysis or numerical techniques of optimal control problem governed by time-fractional diffusion equations [3]. In [4], control constrained optimal control problem governed by the time-fractional diffusion equation with the Caputo-type time-fractional derivative was studied. In the present paper, we approximate optimal control problems (1) and (2) by the Legendre pseudo-spectral method (PSM). For the spatial discretization, Lagrange interpolating basis polynomials associated with Legendre–Gauss–Lobatto (LGL) points are used to approximate the state, and the second derivative of the state is discretized by using the differentiation matrices. Furthermore, the objective functional is approximated by using the LGL numerical quadrature rule. Then, the optimal control problem is converted to matrix form. We consider a finite difference method for the time discretization. A fully discrete first-order optimality condition is obtained. A projected gradient algorithm is designed based on the fully discrete optimality condition. Numerical examples are presented to verify the effectiveness of the presented method.

2 The Legendre pseudo-spectral method

In this section the semi-discrete formulation of optimal control problem (1) and (2) is obtained by using the Legendre PSM for spatial discretization. Furthermore, we derive the semi-discrete first order optimality condition by using Pontryagin's minimum principle. First, we approximate the state u by the Lagrange polynomials as

$$u(x,t) \approx \tilde{u}(x,t) = \sum_{i=0}^{n} u_i(t)\varphi_i(x), \qquad \forall (x,t) \in I_T, \tag{4}$$

where $u_i(t) = \tilde{u}(x_i,t), u_1(t), \cdots, u_{n-1}(t)$ are unknown functions. Moreover, $\phi_i(x)$ are the Lagrange interpolating basis polynomials of degree n,

$$\varphi_i(x) = \prod_{k=0,k\neq i}^{n} \frac{x-x_k}{x_i-x_k}, \qquad i = 0,1,\ldots,n, \tag{5}$$

where the interpolating points $x_i (i = 0,1,\cdots,n)$ are LGL points associated with interval $[a,b]$. Then we obtain

$$u(x_0,t) = \tilde{u}(a,t) = 0, \qquad u(x_n,t) = \tilde{u}(b,t) = 0, \qquad \forall t \in (0,T).$$

If we set $\mathbf{u}(t) = [u_1(t),\ldots,u_{n-1}(t)]^T$ and $\boldsymbol{\phi}(x) = [\varphi_1(x),\ldots,\varphi_{n-1}(x)]^T$. Then

$$\tilde{u}(x,t) = \phi^T(x)\mathbf{u}(t), \quad \forall (x,t) \in I_T. \tag{6}$$

Furthermore we define $\mathbf{W} = \mathrm{diag}[\omega_1,\omega_2,...,\omega_{n-1}]$, where ω_i are the LGL weights. which are positive and defined by

$$\omega_i = \frac{2}{n(n+1)(P_n(\xi_i))^2}, \quad i = 0,1,\cdots,n,$$

Consequently the semi-discrete scheme of optimal control problem (1)-(2) can be characterized as

$$\begin{aligned}\min_{q\in\mathcal{A}_h}\tilde{J}(\mathbf{q},\mathbf{u}) &= \frac{b-a}{4}\int_0^T \{\mathbf{u}(t)-\mathbf{u_d}(t)]^T\mathbf{W}[\mathbf{u}(t)-\mathbf{u_d}(t)] \\ &+ \gamma\mathbf{q}^T(t)\mathbf{W}\mathbf{q}(t)\}dt, \end{aligned} \tag{7}$$

subject to

$$\begin{cases} {}_0^C D_t^\alpha \mathbf{u}(t) = \mathbf{D}\mathbf{u}(t)+\mathbf{f}(t)+\mathbf{q}(t),\ t\in(0,T), \\ \mathbf{q}(t)\in\mathcal{A}_h=\big\{\mathbf{q}(t): q_a\le q(x_i,t)\le q_b,\ 1\le i\le n-1,\ t\in(0,T)\big\}, \\ \mathbf{u}(0)=\mathbf{g}, \end{cases}$$

where $\mathbf{q}(t)=[q(x_1,t),\dots,q(x_{n-1},t)]^T$, $\mathbf{f}(t)=[f(x_1,t),\dots,f(x_{n-1},t)]^T$.

Hence, by using Pontryagin's minimum principle, the semi-discrete optimality conditions can be summarized as

$$\begin{cases} {}_0^C D_t^\alpha \mathbf{u}(t) = \mathbf{D}\mathbf{u}(t)+\mathbf{f}(t)+\mathbf{q}(t), & t\in(0,T), \\ {}_t^C D_T^\alpha \lambda(t) = \frac{b-a}{2}\mathbf{W}[\mathbf{u}(t)-\mathbf{u_d}(t)]+\mathbf{D}^T\lambda(t), & t\in(0,T), \\ \mathbf{u}(0)=\mathbf{g}, \quad \lambda(T)=0. \end{cases}$$

where the components of optimal control $q(t)$ can be obtained as:

$$q_i(t)=\begin{cases} q_a, & if \quad -\dfrac{2\lambda_i(t)}{\gamma(b-a)\omega_i}<q_a, \\ -\dfrac{2\lambda_i(t)}{\gamma(b-a)\omega_i}, & if\ \ q_a\le -\dfrac{2\lambda_i(t)}{\gamma(b-a)\omega_i}\le q_b, \\ q_b, & if\ \ q_b\le -\dfrac{2\lambda_i(t)}{\gamma(b-a)\omega_i}. \end{cases} \tag{8}$$

2.1 A fully discrete scheme

In this section, a fully discrete scheme is formulated to derive the fully discrete optimality conditions. To approximate the left Caputo fractional derivative (3), we define the following discrete left fractional differential operator

$${}_0L_t^\alpha u_i(t_{m+1}) = {}_0L_t^\alpha {U_i}^{m+1} = \frac{1}{\Gamma(2-\alpha)}\sum_{j=0}^{m} B_j \frac{u_i(t_{m+1-j})-u_i(t_{m-j})}{\tau^\alpha}.$$

where $B_j=(j+1)^{1-\alpha}-j^{1-\alpha}$. If $u_i(t)$ be a twice continuously differentiable function, then the above approximation is of order $2-\alpha$. Consequently, we obtain the fully discrete scheme of optimal control problem as:

$$\begin{aligned}\min_{Q^{m+1}\in\mathcal{A}_{h,\tau}} \frac{\tau(b-a)}{4}\sum_{m=0}^{M-1}\big\{&[\mathbf{U}^{m+1}-{\mathbf{u_d}}^{m+1}]^T\mathbf{W}[\mathbf{U}^{m+1}-{\mathbf{u_d}}^{m+1}] \\ &+\gamma(\mathbf{Q}^{m+1})^T\mathbf{W}\mathbf{Q}^{m+1}\big\}, \end{aligned} \tag{9}$$

subject to

$$\begin{cases} {}_0L_t^\alpha \mathbf{U}^{m+1} = \mathbf{D}\mathbf{U}^{m+1}+\mathbf{f}^{m+1}+\mathbf{Q}^{m+1}, \\ \mathbf{Q}^{m+1}\in\mathcal{A}_{h,\tau}=\big\{\mathbf{Q}^{m+1}: q_a\le q(x_i,t_{m+1})\le q_b,\ 1\le i\le n-1\big\}, \\ \mathbf{U}^0=\mathbf{g}. \end{cases} \tag{10}$$

Table 1: The L^2-errors for various values of α and $\tau$with $n = 50$.

α	τ	1/41	1/82	order $(\tau^{2-\alpha})$
1/8	*error*	$5.63e-05$	$1.64e-05$	$\approx 1.79(1.80)$
1/2	*error*	$8.74e-04$	$3.12e-04$	$\approx 1.47(1.5)$
7/8	*error*	$8.74e-04$	$3.12e-04$	$\approx 1.47(1.5)$

Here, the control constraint is to be understood as $q_a \leq q(x_i, t_{m+1}) \leq q_b$ for all collocation points at each time step. Finally using the Karush-Kuhn-Tucker conditions, the fully discrete optimality conditions for (9)-(10) are expressed as

$$\begin{cases} {}_0L_t^{\alpha}\mathbf{U}^{m+1} = \mathbf{D}\mathbf{U}^{m+1} + \mathbf{f}^{m+1} + \mathbf{Q}^{m+1}, \\ {}_tL_T^{\alpha}\Lambda^m = \frac{b-a}{2}\mathbf{W}\left[\mathbf{U}^{m+1} - {\mathbf{u}_d}^{m+1}\right] + \mathbf{D}^T\Lambda^m, \\ \left[\frac{\gamma(b-a)}{4}(\mathbf{Q}^{m+1})^T\mathbf{W} + (\Lambda^m)^T\right](\mathbf{v} - \mathbf{Q}^{\mathbf{m+1}}) \geq \mathbf{0}, \quad \forall \mathbf{v} \in \mathbf{A}_{\mathbf{h},\tau}, \\ \mathbf{U}^0 = \mathbf{g}, \qquad \Lambda^m = 0, \qquad m = 0, 1, \cdots, M-1. \end{cases} \tag{11}$$

By applying a projected gradiant algorithm for (11), the discrete state and adjoint equations can be solved.

3 Numerical example

Example 3.1. In problems (1) and (2) with $I = (0, 1)$, $\gamma = \dfrac{1}{\pi^2}$. The exact solutions are given by

$$\begin{cases} u = t^4 \sin(\pi x), \\ \lambda = (T-t)^4 \sin(\pi x), \\ q = \max\left(-1, \min(-\dfrac{1}{\gamma}(\lambda), -0.3)\right). \end{cases}$$

where u is the state variable and q is the control variable. also the right-hand term f and the desired state u_d can be calculated by the exact solutions and governing equations. By applying the proposed scheme with $n = 50$, $\alpha = 1/10$ and $T = 1$, the exact solutions u, q and numerical solutions U, Q with $t = 1/4$ are plotted in Figure 1. We tabulate in Table 1 the L^2-errors of the state variable with $T = 1$, and the order of the convergence with respect to time variable for the proposed scheme, where $error = \left\|u - U^{M+1}\right\|_{L^2(I)}$. FromTable 1, we observe that the convergence rate of the state variable with respect to the time variable agree with the theoretical findings for various values of α.

4 Conclusion

In this paper, a numerical method based on the Legendre-Gauss Lobato pseudo-spectral methods and finite differences is presented to solve an optimal control problem derived from the fractional diffusion equation. For this purpose, we first formulate the first and second order derivative matrices of pseudo-spectral methods and then the differential equation with partial derivatives in the problem conditions with respect to space and time, respectively, using the Legendre-Gauss Lobato pseudo-spectral method and finite difference

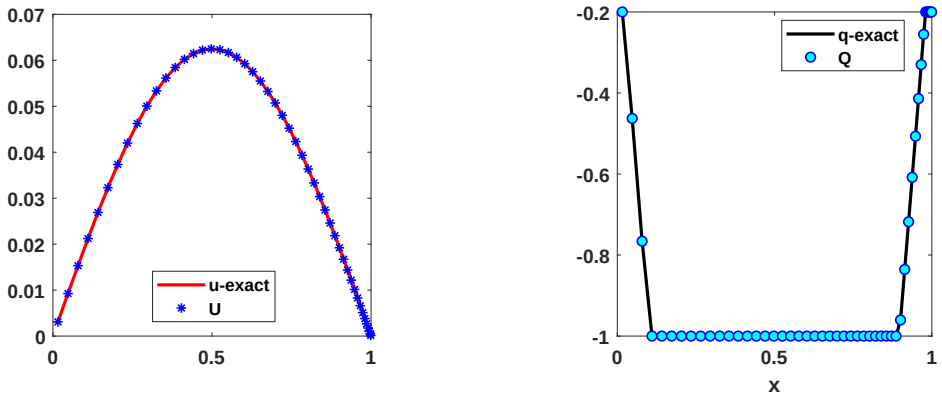


Figure 1: The exact(u, q) and numerical(U, Q) solutions with $t = 1/4$.

formulas. We discretize and turn the desired optimal control problem into a nonlinear optimization problem. Also, we present the necessary conditions for Pontryagin's first-order optimality for the desired optimal control problem and show that the Krusch-Kan Tucker condition corresponding to the nonlinear optimization problem obtained after applying the numerical method is equivalent to the discretized form of Pontryagin's conditions, . Finally, some numerical results are presented to show the effectiveness of the method.

References

[1] F. Kheirkhah, M. Hajipour and D. Baleanu, The performance of a numerical scheme on the variable-order time-fractional advection-reaction-subdiffusion equations, *Applied Numerical Mathematics*, 178 (2022), 25–40.

[2] D E. Kirk, *Optimal Control Theory: an Introduction*, Springer-Verlage, New York, 1970.

[3] A. Jajarmi, M. Hajipour, E. Mohammadzadeh and D. Baleanu, A new approach for the nonlinear fractional optimal control problems with external persistent disturbances, *Journal of the Franklin Institute*, 355 (2018), 3938–3967.

[4] S. Li and Z.J. Zhou, Legendre pseudo-spectral method for optimal control problem governed by a time fractional diffusion equation, *International journal of Computer Mathematics*, 95 (2018), 1308–1325.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Applications of positive definite matrices in the numerical methods for ODEs

Rana Akbari*, Gholamreza Hojjati and Ali Abdi

Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran

Abstract

In designing general linear methods for the numerical soluition of ODEs, a concept known as algebraic stability is defined based on a non-linear test problem. By this concet, the behavior of the methods applying to the non-linear problems is analyzed. In this paper, we discuss the role of positive definite matrices in defining the algebraic stability and its properties.

Keywords: General linear methods, algebraic stability, symmetric, non-negative definite matrix, positive definite matrix,Second derivative methods.

Mathematics Subject Classification [2010]: 65L05,65L20

1 Introduction

To consider the applicability of a numerical method for solving stiff ordinary differential equations, we need to check the stability properties of the method. A general linear method (GLM) for the numerical solution of the non-autonomous initial value problem (IVP)

$$\begin{aligned} &y'(x) = f(t, y(t)), \qquad f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m, \\ &y(0) = 0, \end{aligned} \tag{1}$$

has both multistage and multistep structure. This methods are based on r input and output values, and s stage values. Defining $y^{[n-1]} = [y_i^{[n-1]}]_{i=1}^r$ and $y^{[n]} = [y_i^{[n]}]_{i=1}^r$ as the quantities imported at the beginning of step number n and the exported at the end of this step, and $Y^{[n]} = [Y_i^{[n]}]_{i=1}^s$ as approximations of stage order q to the vector $y(x_{n-1} + ch) = [y(x_{n-1} + c_i h)]_{i=1}^s$, a GLMs for solving (1) is defined by [3]

$$\begin{cases} Y^{[n]} = h(A \otimes I_m) f(Y^{[n]}) + (U \otimes I_m) y^{[n-1]}, \\ y^{[n]} = h(B \otimes I_m) f(Y^{[n]}) + (V \otimes I_m) y^{[n-1]}. \end{cases} \tag{2}$$

Also, an SGLM for numerical solution of (1) is defined as

$$\begin{cases} Y^{[n]} = h(A \otimes I_m) F(Y^{[n]}) + h^2(\overline{A} \otimes I_m) G(Y^{[n]}) + (U \otimes I_m) y^{[n-1]}, \\ y^{[n]} = h(B \otimes I_m) F(Y^{[n]}) + h^2(\overline{B} \otimes I_m) G(Y^{[n]}) + (V \otimes I_m) y^{[n-1]}, \end{cases} \tag{3}$$

*Speaker. Email address: r.akbari@tabrizu.ac.ir

where $n = 1, 2, \ldots, N$ with $Nh = \overline{x} - x_0$.
$F(Y^{[n]}) := [f(Y_i^{[n]})]_{i=1}^s$, and $G(Y^{[n]}) := [g(Y_i^{[n]})]_{i=1}^s$ denote the vectors of first and second derivative stage values with $g(\cdot) = f'(\cdot)f(\cdot)$. The matrices $A, \overline{A} \in \mathbb{R}^{s\times s}$, $U \in \mathbb{R}^{s\times r}$, $B, \overline{B} \in \mathbb{R}^{r\times s}$ and $V \in \mathbb{R}^{r\times r}$ are called as the coefficients matrices of the method which are usually written as a partitioned $(s + r) \times (2s + r)$ matrix

$$\left[\begin{array}{c|c|c} A & \overline{A} & U \\ \hline B & \overline{B} & V \end{array}\right]. \tag{4}$$

The syudies of nonlinear stability property has been discussed for multistep methods as G-stability [6] and for Runge–Kutta methods as B-stability [4]. In this paper, by recaling the concept of algebraic stability of GLMs, we determine the conditions under which the method remains algebraically stable by using the positive definite conditions of a matrix.Also, examples of GLMs are presented and their algebraic stability is evaluated. Finally, the concept of algebraic stability for second-derivative GLMs (SGLMs) is presented, and using the positive definite condition of a matrix as for GLMs, we determine the conditions under which this methods remain algebraically stable.

2 Algebraic stability of GLMs

To analyze the linear stability properties of a GLM, it is applied to the linear test problem of Dahlquist, $y' = \xi y, \xi \in \mathbb{C}$, to get $y^{[n]} = M(z)y^{[n-1]}$, where the matrix $M(z)$ is defined by

$$M(z) = V + zB(I - zA)^{-1}U.$$

The $M(z)$ is called the stability matrix of GLM. For methods used to solve stiff problems, the stepsizes required for stability may be very small. This means that stability rather than accuracy limits the step size. To ensure that there is no bound on the step size, A-stability condition is desired. A GLM is A-stable if for all $z \in \mathbb{C}^-$, $I - zA$ is non-singular and $M(z)$ is a stable matrix.
A-stability describes the behavior of the method when applied to a linear, autonomous differential equation. This concept is noteworthy, but it says little about the action of the method applied to problems that are either non-autonomous or non-linear or both. Lately, another stability definitions have been proposed for overcoming this defect. In the case of linear multistep methods, the stability of non-linear problems has been studied through the idea of G-stability [6] while in the case of Runge–Kutta methods, the similar type of model has been used under the name B-stability [4]. The concepts of AN-, BN-, B-, and algebraic stability were investigated by Butcher [5, 7, 8]. The AN-stability is a generalization of A-stability and is relevant to the scalar, linear, and non-autonomous test equation

$$\begin{cases} y' = \xi(t)y, & t \geq 0, \\ y(0) = y_0, \end{cases}$$

$Re(\xi(t)) \leq 0$, where $\xi(t)$ has an arbitrarily complex-valued. Consider (1) for $t \geq 0$, where $(f(t, y_1) - f(t, y_2))^T(y_1 - y_2) \leq 0$, for all $t \geq 0$ and $y_1, y_2 \in \mathbb{R}^m$. Applying GLM to this test equation, we obtain

$$y^{[n+1]} = \mathbf{S}(\xi)y^{[n]}, \qquad n = 0, 1, \ldots.$$

where $\xi = \mathrm{diag}(\xi_1, \xi_2, \ldots, \xi_s) = \mathrm{diag}(h\xi(t_n + c_1 h), \ldots, h\xi(t_n + c_s h))$, and

$$\mathbf{S}(\xi) = \mathbf{V} + \mathbf{B}\xi(\mathbf{I} - \mathbf{A}\xi)^{-1}\mathbf{U}.$$

To define AN-, G- and algebaric stability, consider $G = [g_{ij}]_{i,j=1}^{r}$ be a real, symmetric and positive definite matrix, and for a vector $y \in \mathbb{R}^{mr}$, let the inner product norm

$$\|y\|_G^2 = \sum_{i=1}^{r}\sum_{j=1}^{r} g_{ij} y_i^T y_j.$$

Definition 2.1. GLM is AN-stable if there exists a real, symmetric, and positive definite matrix G for all $\xi = \mathrm{diag}(\xi_1, \ldots, \xi_s)$, so that

$$\|\mathbf{S}(\xi)y\|_G \leq \|y\|_G,$$

and $\xi_i = \xi_j$ while $c_i = c_j$ and for $i = 1, 2, \ldots, s$, we have $Re(\xi_i) \leq 0$.

Definition 2.2. GLM is said to be G-stable if for two numerical solutions, $\{y^{[n]}\}_{n=0}^{N}$ and $\{\widetilde{y}^{[n]}\}_{n=0}^{N}$, there is a real, symmetric, and positive definite matrix $G \in \mathbb{R}^{r\times r}$ that

$$\|y^{[n+1]} - \widetilde{y}^{[n+1]}\|_G \leq \|y^{[n]} - \widetilde{y}^{[n]}\|_G,$$

and $\|.\|_G$ is the norm defined for all $h > 0$. For given $G \in \mathbb{R}^{r\times r}$ and $D \in \mathbb{R}^{s\times s}$, define the matrix M by the formula

$$M := \left[\begin{array}{c|c} DA + A^T D - B^T G B & DU - B^T G V \\ \hline U^T D - V^T G B & G - V^T G V \end{array}\right]. \tag{5}$$

Definition 2.3. GLM is said to be algebraically stable if there exist a real, symmetric, and positive definite matrix G and a real, diagonal, and positive definite matrix D such that the matrix M is nonnegative definite [3].

There are relationships between the types of sustainability defined above. Algebraic stability is equivalent to AN-stability and both of them imply A-stability.
We derive equivalent conditions for the algebraic stability of GLMs by using Albert's theorem [2] which is a theory about block symmetric nonnegative definite matrices.In the following, we use the notation $X \geq 0$ to denote that the matrix X is symmetric non-negative definite matrix.

Theorem 2.4. *[2] The matrix*

$$\left[\begin{array}{c|c} M_{11} & M_{12} \\ \hline M_{21} & M_{22} \end{array}\right], \tag{6}$$

is non-negative definite if and only if there exists a matrix Z such that

$$\begin{aligned} &M_{22} \geq 0, \\ &M_{22}Z = M_{21}, \\ &M_{11} - M_{12}Z \geq 0. \end{aligned}$$

Applying these results to the matrix M defined by (5), we can conclude that GLM is algebraically stable if and only if there exist matrices Z so that

$$\begin{aligned} &G - V^T G V \geq 0, \\ &\big(G - V^T G V\big)Z = U^T D - V^T G B, \\ &DA + A^T D - B^T G B - \big(DU - B^T G V\big)Z \geq 0. \end{aligned} \tag{7}$$

3 Examples

Finally, we consider some examples of GLMs and examine the characteristic of algebraic stability in these methods.

Example 3.1. We study the property of algebraic stability on the implicit trapezoidal method

$$Y_1 = y_n + \frac{1}{2}hf(t_{n-1} + \frac{h}{2}, Y_1),$$

$$y_{n+1} = y_n + hf(t_{n-1} + \frac{h}{2}, Y_1).$$

Coefficients partitioned matrix for this method is given by

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] = \left[\begin{array}{c|c} \frac{1}{2} & 1 \\ \hline 1 & 1 \end{array}\right].$$

We find

$$M = \left[\begin{array}{c|c} \frac{D}{2} + \frac{D}{2} - G & D - G \\ \hline D - G & 0 \end{array}\right]. \tag{8}$$

According to the three conditions mentioned above for the non-negative definiteness of the matrix M, it can be concluded that $G = D$ must be satisfied for the algebraic stability of the method. Therefore, the implicit trapezoidal method is algebraically stable.

Example 3.2. As the another example, for the implicit Euler method

$$Y_1 = y_n + hf(t_{n-1} + h, Y_1),$$
$$y_{n+1} = y_n + hf(t_{n-1} + h, Y_1),$$

where can be represented by a partitioned matrix

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] = \left[\begin{array}{c|c} 1 & 1 \\ \hline 1 & 1 \end{array}\right],$$

this yields

$$M = \left[\begin{array}{c|c} 2D - G & D - G \\ \hline D - G & 0 \end{array}\right].$$

It can be shown that for $G = D$, the matrix M is a non-negative definite matrix. Consequently, the implicit Euler method is algebraically stable.

Example 3.3. Consider the one-stage method

$$\begin{array}{c|c} \lambda & \lambda \\ \hline & 1 \end{array}.$$

Coefficients partitioned matrix for this method is given by

$$\left[\begin{array}{c|c} \lambda & 1 \\ \hline 1 & 1 \end{array}\right]$$

By calculating the Mmatrix, we have

$$M = \left[\begin{array}{c|c} 2\lambda D - G & D - G \\ \hline D - G & 0 \end{array}\right].$$

which is a nonnegative definite matrix with $D = G$ and $\lambda \geq \frac{1}{2}$. Therefore, the method with this assumptions is algebraically stable.

4 Algebraic stability for second derivative GLMs

As for GLMs, the algebraic stability and irreducibility concepts for second derivative GLMs (SGLMs) have been discussed in [1] and some examples of the methods equipped with these properties up to order four have been constructed.
In the SGLMs, For given $G \in \mathbb{R}^{r\times r}$ and $D \in \mathbb{R}^{s\times s}$,define the matrix M by the formula

$$M := \begin{bmatrix} G - V^T GV & U^T D - V^T GB & -V^T G\overline{B} \\ DU - B^T GV & DA + A^T D - B^T GB & D\overline{A} - B^T G\overline{B} \\ -\overline{B}^T GV & \overline{A}^T D - \overline{B}^T GB & -\overline{B}^T G\overline{B} \end{bmatrix}. \tag{9}$$

We now introduce the definition of algebraic stability of SGLMs.

Definition 4.1. SGLM is said to be algebraically stable if there exist a real, symmetric, and positive definite matrix G and a real, diagonal, and positive definite matrix D such that the matrix M of (9) is nonnegative definite [1].

Based on the Albert's theorem (2.4), we derive equivalent conditions for the the algebraic stability of SGLMs. The result of this theorem can be generalized to 3×3 block matrices as follows. Indeed, the matrix

$$\left[\begin{array}{c|cc} M_{11} & M_{12} & M_{13} \\ \hline M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{array}\right],$$

is non-negative definite if and only if there exist matrices Z_1, Z_2 and Z_3 such that

$$\begin{aligned} &M_{33} \geq 0, \\ &M_{33}Z_1 = M_{32}, \\ &M_{22} - M_{23}Z_1 \geq 0, \\ &M_{22}Z_2 + M_{23}Z_3 = M_{21}, \\ &M_{32}Z_2 + M_{33}Z_3 = M_{31}, \\ &M_{11} - [M_{12}Z_2 + M_{13}Z_3] \geq 0. \end{aligned}$$

Applying these results for the matrix M defined by (9), we find that an SGLM is algebraically stable if and only if matrices Z_1, Z_2 and Z_3 exist such that

$$\begin{aligned} &-\overline{B}^T G\overline{B} \geq 0, \\ &-\overline{B}^T G\overline{B}Z_1 = -\overline{A}^T D - \overline{B}^T GB, \\ &DA + A^T D - B^T GB - \big(D\overline{A} - B^T G\overline{B}\big)Z_1 \geq 0, \\ &\big(DA + A^T D - B^T GB\big)Z_2 + \big(D\overline{A} - B^T G\overline{B}\big)Z_3 = DU - B^T GV, \\ &\big(-\overline{A}^T D - \overline{B}^T GB\big)Z_2 + \big(-\overline{B}^T G\overline{B}\big)Z_3 = -\overline{B}^T GV, \\ &G - V^T GV - \big((U^T D - V^T GB)Z_2 - V^T G\overline{B}Z_3\big) \geq 0. \end{aligned}$$

References

[1] A. Akbari, G. Hojjati, A. Abdi, Algebraic stability and irreducibility of second derivative methods, *Appl. Numer. Math.*, to appear.

[2] A. Albert, Conditions for positive and non-negative definiteness in terms of pseudoinverses, *SIAM J. Appl. Math.* 17 (1969), 434–440.

[3] K. Burrage, J.C. Butcher, Non-linear stability for a general class of differential equation methods, *BIT* , 20 (1980), 185–203.

[4] J.C. Butcher, A stability property of implicit Runge–Kutta methods, *BIT*, 15 (1975), 358–361.

[5] J.C. Butcher, Thirty years of G-stability, *BIT* 46 (2006), 479–489.

[6] G. Dahlquist, On stability and error analysis for stiff non-linear problems, *1, Dept. of Information Processing, Royal Institute of Technology, Stockholm.*, Report NA 75.08.

[7] E. Hairer and G. Wanner, Solving Ordinary Differential Equations 11: Stiff and Differential-Algebraic Problems, *2nd rev. ed., Springer-Verlag, New York*, 1996.

[8] L.L. Hewitt, A.T. Hill, Algebraically stable general linear methods and the G-matrix, *BIT* 49 (2009), 93–111.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On some identities and inequalities for 2-Frames in 2-Inner product Spaces

Fahimeh Sultanzadeh*

Department of Mathematics, Mashhad Branch, Islamic Azad University, Mashhad, Iran

Abstract

2-Frames share many useful properties with frames. In this paper we introduce some identities and inequalities for 2-Frames in 2-Inner product Spaces and give several results in this field. These identities and inequalities discussed by Gavruta and Balan et al. in Hilbert space.

Keywords: 2-Inner product Space, 2-Frame, 2-Frame operator
Mathematics Subject Classification [2010]: 46C50, 42C15

1 Introduction

Frames are a key tool in ingineering, applied mathematics, and computer sciences. The concept of linear 2-normed spaces has been investigated by S. Gahlor [5] in 1965 and has been developed extensively by many mathematicians.

Let $\mathcal{X}$ be a linear space of dimension greater than 1 over the field $\mathcal{K}$, where $\mathcal{K}$ is the real or complex numbers field. Suppose that $(\cdot,\cdot|\cdot)$ is a $\mathcal{K}$-valued function defined on $\mathcal{X}\times\mathcal{X}\times\mathcal{X}$ satisfying the foliowing conditions:

(I1) $(x,x|z)\geq 0$ and $(x,x|z)=0$ if and only if x and z are linerly dependent,

(I2) $(x,x|z)=(z,z|x)$,

(I3) $(y,x|z)=\overline{(x,y|z)}$,

(I4) $(\alpha x,y|z)=\alpha(x,y|z)$ for any scalar $\alpha\in\mathcal{K}$,

(I5) $(x+x^{'},y|z)=(x,y|z)+(x^{'},y|z)$,

where x,$x^{'}$,y,$z\in\mathcal{X}$.
$(\cdot,\cdot|\cdot)$ is called a 2-inner Product on $\mathcal{X}$ and $(\mathcal{X},(\cdot,\cdot|\cdot))$ is called a 2-inner Product space (or 2-pre Hilbert spece). Some properties of 2-inner Product $(\cdot,\cdot|\cdot)$ can be obtained as follows ([4]):

- $(0,y|z)=(x,0|z)=(x,y|0)=0$,

*Speaker. Email: fsultanzadeh@gmail.com

- $(x, \alpha y|z) = \overline{\alpha}(x, y|z)$,
- $(x, y|\alpha z) = |\alpha|^2(x, y|z)$ for all $x, y, z \in \mathcal{X}$ and $\alpha \in \mathcal{K}$.

Using the above properties, we can prove the Cauchy-Schwarz inequality

$$|(x, y|z)|^2 \leq (x, x|z)(y, y|z). \tag{1}$$

Example 1.1. If $(\mathcal{X}, \langle \cdot, \cdot \rangle)$ is an inner Product space, then the standard 2-inner Product $(\cdot, \cdot|\cdot)$ is defined on $\mathcal{X}$ by

$$(x, y|z) = \begin{vmatrix} \langle x, y \rangle & \langle x, z \rangle \\ \langle z, y \rangle & \langle z, z \rangle \end{vmatrix} = \langle x, y \rangle \langle z, z \rangle - \langle x, z \rangle \langle z, y \rangle, \tag{2}$$

for all $x, y, z \in \mathcal{X}$.

In any given 2-inner Product space $(\mathcal{X}, (\cdot, \cdot|\cdot))$, we can define a function $\|\cdot, \cdot\|$ on $\mathcal{X} \times \mathcal{X}$ by

$$\|x, z\| = (x, x|z)^{\frac{1}{2}}, \tag{3}$$

for all $x, z \in \mathcal{X}$.

It is easy to see that, this function satisfies the following conditions:

(N1) $\|x, z\| \geq 0$ and $\|x, z\| = 0$ if and only if x and z are linerly dependent,

(N2) $\|x, z\| = \|z, x\|$,

(N3) $\|\alpha x, z\| = |\alpha| \|x, z\|$ for any scalar $\alpha \in \mathbb{K}$,

(N4) $\|x_1 + x_2, z\| \leq \|x_1, z\| + \|x_2 z\|$.

Any function $\|\cdot, \cdot\|$ defined on $\mathcal{X} \times \mathcal{X}$ and satisfying the above conditions is called a 2-norm on $\mathcal{X}$ and $(\mathcal{X}, \|\cdot, \cdot\|)$ is called a linear 2-normed space. Whenever a 2-inner product space $(\mathcal{X}, (\cdot, \cdot|\cdot))$ is given, we consider it as a linear 2-normed space $(\mathcal{X}, (\cdot, \cdot|\cdot))$ with 2-norm defined by (3).

Let $\mathcal{X}$ be a 2-inner product space , $h \in \mathcal{X}$ and $x, y \in \mathcal{X} \setminus L_h$, where L_h is the subspace of $\mathcal{X}$ generated by h, then

$$x \perp^h y \Leftrightarrow (x, y|h) = 0.$$

Let $\mathcal{X}$ be a 2-inner product space. A sequence $\{a_n\}_{n=1}^{\infty}$ is said to be convergent if there exists an element $a \in \mathcal{X}$ such that $\lim_{n \to \infty} \|a_n - a, x\| = 0$, for all $x \in \mathcal{X}$. Similary, we can define a Cauchy sequence in $\mathcal{X}$. A 2-linear product space $\mathcal{X}$ is called a 2-Hilbert space if it is compelete. That is, every Cauchy sequence in $\mathcal{X}$ is convergent in this space. Using Theorem 1.7 [6], the following example is obtained:

Example 1.2. The space $l^2(I)$ with 2-linear product defined by

$$(\{a_i\}_{i \in I}, \{b_i\}_{i \in I} | \{c_i\}_{i \in I}) = \langle \{a_i\}_{i \in I}, \{b_i\}_{i \in I} \rangle \langle \{c_i\}_{i \in I}, \{c_i\}_{i \in I} \rangle - \langle \{a_i\}_{i \in I}, \{c_i\}_{i \in I} \rangle \langle \{c_i\}_{i \in I}, \{b_i\}_{i \in I} \rangle. \tag{4}$$

is a 2-Hilbert space.

Suppose that $\mathcal{X}$ and $\mathcal{Y}$ are 2-Hilbert spaces and $T : \mathcal{X} \to \mathcal{Y}$. The adjoint of T is $T^* : \mathcal{Y} \to \mathcal{X}$ such that

$$(Tx, y|h') = (x, T^* y|h) \quad \text{for all } x, h \in \mathcal{X}, y, h' \in \mathcal{Y}.$$

The operator $T : \mathcal{X} \to \mathcal{Y}$ is bounded if there is a positive number K such that

$$\| Tx, h^{'} \| \leq K \| x, h \| \quad \text{for all } x, h \in \mathcal{X}, h^{'} \in \mathcal{Y}.$$

Recently Arefijamaal and Sadeghi [1] defined frames in 2-inner product spaces. R. Balan, P.G. Gasazza, D. Edidin, and G. Kutyniok in [2] established several identities and inequalities for frames in Hibert spaces. In the present paper, 2-Parseval frame identity in 2-inner product space has discussed and an extension is given. In particular, we derive intriguing equivalent condititions for both sides of the identity to be equal to zer.

2 Main results

Let $(\mathcal{X}, (\cdot, \cdot|\cdot))$ be a 2-Hilbert space and $h \in \mathcal{X}$. A sequence $\{f_i\}_{i\in I}$ of elements in $\mathcal{X}$ is called a 2-frame (associated to h) if there exist $A, B > 0$ such that

$$A\|f, h\|^2 \leq \sum_{i\in I} |\langle f, f_i|h\rangle|^2 \leq B\|f, h\|^2 \quad \text{for every } f \in \mathcal{X}. \tag{5}$$

A sequence satisfying the upper 2-frame condition is called a 2-Bessel sequence. In (5) we may assume that every f_i is linearly independent to h, by (1) and the property (I1).

Definition 2.1. Let $(\mathcal{X}, (\cdot, \cdot|\cdot))$ be a 2-Hilbert space and $h \in \mathcal{X}$. A sequence $\{f_i\}_{i\in I}$ is said to be a 2-tight frame if $A\|f, h\|^2 = \sum_{i\in I} |(f, f_i|h)|^2$, for all $f \in \mathcal{X}$. Moreover, $\{f_i\}_{i\in I}$ is said to be a 2-Parseval frame if it is a 2-tight frame for the 2-Hilbert space $\mathcal{X}$ and in addition $A = 1$.

In the standard 2-inner Product spaces every frame is a 2-frame, but the converse is not true. In fact, every 2-frame is a frame for a closed subspase of $\mathcal{H}$ with codimension 1. Every 2-frame associated to h is a frame for $L_h^{\perp}$ [1]. Assume that $(\mathcal{X}, (\cdot, \cdot|\cdot))$ is a 2-Hilbert space, the algebraic complement of L_h in $\mathcal{X}$ is denoted by M_h, i.e $L_h \oplus M_h = \mathcal{X}$. One may see that

$$\langle f, g\rangle_h = (f, g|h), \quad f, g \in \mathcal{X}.$$

defines a semi-inner product on $\mathcal{X}$ [1] This semi-inner product induces the following inner product on the quotient space $\dfrac{\mathcal{X}}{L_h}$ as

$$\langle f + L_h, g + L_h\rangle_h = \langle f, g\rangle_h, \quad f, g \in \mathcal{X}.$$

So M_h with respect to $\|f\|_h = \sqrt{\langle f, f\rangle_h}$, where$f \in M_h$, is a normed space. Now if $\{f_i\}_{i\in I} \subseteq \mathcal{X}$ is a 2-frame associated to h with bounds A and B, then we can rewrite (5) as

$$A\|f\|_h^2 \leq \sum_{i\in I} |\langle f, f_i\rangle_h|^2 \leq B\|f\|_h^2, \quad \text{for every } f \in M_h.$$

The completion of the inner product space M_h is denoted by $\mathcal{X}_h$. The sequence $\{f_i\}_{i\in I}$ is also a frame for $\mathcal{X}_h$ with the same bounds [3].

Proposition 2.2. *[1] Let $(\mathcal{X}, (\cdot, \cdot|\cdot))$ be a 2-Hilbert space. Then $\{f_i\}_{i\in I} \subseteq \mathcal{X}$ is a 2-frame associated to h with bounds A and B if and only if it is a frame for the Hilbert space $\mathcal{X}_h$ with bounds A and B.*

Let $\{f_i\}_{i\in I}$ be a 2-Bessel sequence in $\mathcal{X}$. Then the 2-frame operator $T_h : l^2 \to \mathcal{X}_h$ defined by $T_h f = \sum_{i\in I}(f, e_i|h)f_i$ is well-defined and bounded.
Also the adjoint of $T_h^* : \mathcal{X}_h \to l^2$ defined by $T_h^* f = \{(f, f_i|h)\}_{i\in I}$ is well-defined and bounded.
Let $\{f_i\}_{i\in I}$ be a 2-frame associated to h with bounds A and B in a 2-Hilbert space $\mathcal{X}$. The operator $S_h : \mathcal{X}_h \to \mathcal{X}_h$ defined by $S_h f = \sum_{i\in I}(f, f_i|h)f_i$ is called the 2-frame operator for $\{f_i\}_{i\in I}$. The operator S_h is invertible, selfadjoint and positive. Furtheremore, each $f \in \mathcal{X}_h$ has an expansion of the following

$$f = S_h S_h^{-1} f = \sum_{i\in I}(S_h^{-1} f, f_i|h)f_i.$$

The family $\{\tilde{f}_i\}_{i\in I}$, where $\tilde{f}_i = S_h^{-1} f_i, i \in I$, is a 2-frame for $\mathcal{X}$, called the canonical dual 2-frame of the $\{f_i\}_{i\in I}$, see [1].

Lemma 2.3. *Let $\mathcal{X}$ be a 2-Hilbert space. If S, T are operators on $\mathcal{X}$ such that $S+T = Id$, then $S - T = S^2 - T^2$.*

Theorem 2.4. *Let $\{f_i\}_{i\in I}$ be a 2-frame associated to h in a 2-Hilbert space $\mathcal{X}$ with canonical dual 2-frame $\{\tilde{f}_i\}_{i\in I}$. Then for all $J \subset I$, we have*

$$\sum_{i\in J} |(f, f_i|h)|^2 - \sum_{i\in I} |(S_{h_J} f, \tilde{f}_i|h)|^2 = \sum_{i\in J^c} |(f, f_i|h)|^2 - \sum_{i\in I} |(S_{h_{J^c}} f, \tilde{f}_i|h)|^2, \quad \text{for } f \in \mathcal{X}_h.$$

Proof. Since $S_h = S_{h_J} + S_{h_{J^c}}$, it follows that $I = S_h^{-1} S_{h_J} + S_h^{-1} S_{h_{J^c}}$. Now we apply Lemma 2.3 and we have

$$S_h^{-1} S_{h_J} - S_h^{-1} S_{h_J} S_h^{-1} S_{h_J} = S_h^{-1} S_{h_{J^c}} - S_h^{-1} S_{h_{J^c}} S_h^{-1} S_{h_{J^c}}.$$

For every $f, g \in \mathcal{X}_h$,

$$\langle S_h^{-1} S_{h_J} f, g\rangle_h - \langle S_h^{-1} S_{h_J} S_h^{-1} S_{h_J} f, g\rangle_h = \langle S_h^{-1} S_{h_{J^c}} f, g\rangle_h - \langle S_h^{-1} S_{h_{J^c}} S_h^{-1} S_{h_{J^c}} f, g\rangle_h,$$

or

$$\langle S_{h_J} f, S_h^{-1} g\rangle_h - \langle S_h^{-1} S_{h_J} f, S_{h_J} S_h^{-1} g\rangle_h = \langle S_{h_{J^c}} f, S_h^{-1} g\rangle_h - \langle S_h^{-1} S_{h_{J^c}} f, S_{h_{J^c}} S_h^{-1} g\rangle_h.$$

By replacing $g = S_h f$, we obtain

$$\langle S_{h_J} f, f\rangle_h - \langle S_h^{-1} S_{h_J} f, S_{h_J} f\rangle_h = \langle S_{h_{J^c}} f, f\rangle_h - \langle S_h^{-1} S_{h_{J^c}} f, S_{h_{J^c}} f\rangle_h,$$

and so

$$\sum_{i\in J} |\langle f, f_i\rangle_h|^2 - \sum_{i\in I} |\langle S_{h_J} f, \tilde{f}_i\rangle_h|^2 = \sum_{i\in J^c} |\langle f, f_i\rangle_h|^2 - \sum_{i\in I} |\langle S_{h_{J^c}} f, \tilde{f}_i\rangle_h|^2.$$

Therefore

$$\sum_{i\in J} |(f, f_i|h)|^2 - \sum_{i\in I} |(S_{h_J} f, \tilde{f}_i|h)|^2 = \sum_{i\in J^c} |(f, f_i|h)|^2 - \sum_{i\in I} |(S_{h_{J^c}} f, \tilde{f}_i|h)|^2.$$

□

Theorem 2.5. *(2-Parsval Frame Identity) Let $\{f_i\}_{i\in I}$ be a 2-parsval frame associated to h in a 2-Hilbert space $\mathcal{X}$. Then for all $J \subset I$, we have*

$$\sum_{i\in J}|(f,f_i|h)|^2 - \|\sum_{i\in J}(f,f_i|h)f_i,h\|^2 = \sum_{i\in J^c}|(f,f_i|h)|^2 - \|\sum_{i\in J^c}(f,f_i|h)f_i,h\|^2, \quad \text{for } f\in\mathcal{X}_h. \tag{6}$$

Proof. Since $\{f_i\}_{i\in I}$ is a 2-parseval frame, its 2-frame operator equals the identity operator, hence $\tilde{f}_i = f_i$ for all $i \in I$. By applying Theorem 2.4 and the fact that $\{f_i\}_{i\in I}$ is a 2-parseval frame, it follows that

$$\begin{aligned}
\sum_{i\in J}|(f,f_i|h)|^2 - \|\sum_{i\in J}(f,f_i|h)f_i,h\|^2 &= \sum_{i\in J}|(f,f_i|h)|^2 - \|S_Jf,h\|^2\\
&= \sum_{i\in J}|(f,f_i|h)|^2 - \sum_{i\in I}|(S_Jf,f_i|h)|^2\\
&= \sum_{i\in J}|(f,f_i|h)|^2 - \sum_{i\in I}|(S_Jf,\tilde{f}_i|h)|^2\\
&= \sum_{i\in J^c}|(f,f_i|h)|^2 - \sum_{i\in I}|(S_{J^c}f,\tilde{f}_i|h)|^2\\
&= \sum_{i\in J^c}|(f,f_i|h)|^2 - \|S_{J^c}f,h\|^2\\
&= \sum_{i\in J^c}|(f,f_i|h)|^2 - \|\sum_{i\in J^c}(f,f_i|h)f_i,h\|^2.
\end{aligned}$$

□

Corollary 2.6. *Let $\{f_i\}_{i\in I}$ be a 2-Parseval frame associated to h in a 2-Hilbert space $\mathcal{X}$, for every $J \subset I$, and every $f \in \mathcal{X}_h$, if S_Jf and h are linearly dependent, then*

$$\|\sum_{i\in J^c}(f,f_i|h)f_i,h\|^2 = \sum_{i\in J^c}|(f,f_i|h)|^2 - \sum_{i\in J}|(f,f_i|h)|^2,$$

Analogosely, if $S_{J^c}f$ and h are linearly dependent, then

$$\|\sum_{i\in J}(f,f_i|h)f_i,h\|^2 = \sum_{i\in J^c}|(f,f_i|h)|^2 - \sum_{i\in J}|(f,f_i|h)|^2.$$

Since each $2A$-tight frames can be turned into a 2-Parseval frame by a change of scale, we have the following corollary.

Corollary 2.7. *Let $\{f_i\}_{i\in I}$ be a $2A$-tight frame associated to h in a 2-Hilbert space $\mathcal{X}$. Then for all $J \subset I$, we have*

$$A\sum_{i\in J}|(f,f_i|h)|^2 - \|\sum_{i\in J}(f,f_i|h)f_i,h\|^2 = A\sum_{i\in J^c}|(f,f_i|h)|^2 - \|\sum_{i\in J^c}(f,f_i|h)f_i,h\|^2, \quad \text{for } f\in\mathcal{X}_h.$$

A version of the 2-Parsval Frame Identity for overlapping divisions is derived in the following result.

Proposition 2.8. *Let $\{f_i\}_{i\in I}$ be a 2-parsval frame associated to h in a 2-Hilbert space $\mathcal{X}$. Then for all $J \subset I$ and $E \subset J^c$, we have*

$$\begin{aligned}
&\|\sum_{i\in J\cup E}(f,f_i|h)f_i,h\|^2 - \|\sum_{i\in J^c\setminus E}(f,f_i|h)f_i,h\|^2\\
&= \|\sum_{i\in J}(f,f_i|h)f_i,h\|^2 - \|\sum_{i\in J^c}(f,f_i|h)f_i,h\|^2 + 2\sum_{i\in E}|(f,f_i|h)|^2, \quad \text{for } f\in\mathcal{X}_h.
\end{aligned}$$

Proof. We apply Theorem 2.5 twice and the result is obtained. □

Theorem 2.9. *Let $\{f_i\}_{i\in I}$ be a 2-parsval frame associated to h in a 2-Hilbert space $\mathcal{X}$. Then for all $J \subset I$, we have*

$$\sum_{i\in J} |(f, f_i|h)|^2 + \| \sum_{i\in J^c} (f, f_i|h) f_i, h\|^2 \geq \frac{3}{4}\|f, h\|^2, \quad \text{for } f \in \mathcal{X}_h.$$

Proof. Since

$$\|f, h\|^2 = \|Sf, h\|^2 = \|S_J f + S_{J^c} f, h\|^2 = \|S_J f + S_{J^c} f\|_h^2,$$

and

$$\|S_J f + S_{J^c} f\|_h^2 \leq \|S_J f\|_h^2 + \|S_{J^c} f\|_h^2 + 2\|S_J f\|_h \|S_{J^c} f\|_h \leq 2(\|S_J f\|_h^2 + \|S_{J^c} f\|_h^2),$$

hence

$$\frac{1}{2}\|f\|_h^2 \leq \|S_J f\|_h^2 + \|S_{J^c} f\|_h^2,$$

or

$$\frac{1}{2}\langle Idf, f\rangle_h \leq \langle S_J f, S_J f\rangle_h + \langle S_{J^c} f, S_{J^c} f\rangle_h.$$

This implies that

$$\frac{1}{2} Id \leq S_J^2 + S_{J^c}^2,$$

so

$$\frac{3}{2} Id \leq S_J + S_{J^c}^2 + S_{J^c} + S_J^2 = 2(S_J + S_{J^c}^2).$$

Now we apply Lemma 2.3, it follows that

$$\frac{3}{4} Id \leq S_J + S_{J^c}^2.$$

That is

$$(S_J f, f|h) + (S_{J^c} f, S_{J^c} f|h) \geq \frac{3}{4}\|f, h\|^2.$$

Therefore, we have

$$\sum_{i\in J} |(f, f_i|h)|^2 + \| \sum_{i\in J^c} (f, f_i|h) f_i, h\|^2 \geq \frac{3}{4}\|f, h\|^2.$$

□

Theorem 2.10. *Let $\{f_i\}_{i\in I}$ be a 2-parsval frame associated to h in a 2-Hilbert space $\mathcal{X}$. Then for each $J \subset I$ and $f \in \mathcal{X}_h$, the following conditions are equivalent:*

(i) $\sum_{i\in J} |(f, f_i|h)|^2 = \| \sum_{i\in J} (f, f_i|h) f_i, h\|^2$,

(ii) $\sum_{i\in J^c} |(f, f_i|h)|^2 = \| \sum_{i\in J^c} (f, f_i|h) f_i, h\|^2$,

(iii) $S_{J^c} S_J f \perp^h f$.

References

[1] A.A. Arefijamaal, Gh. Sadeghi, *Frames in* 2*-inner product spaces*, Iran. J. Math. Sci. Inform., **8**(2), (2013), 123–130.

[2] R. Balan, P.G. Casazza, D. Edidin, G. Kutyniok, *A new identity for Parseval frames* , Proc. Amer. Math. Soc.**135** (2007), 1007–1015.

[3] O. Christensen, Frames and bases, *An Introductory course.* Birkhauser Boston, (2008).

[4] R.W. Freese, S.S. Dragomir, Y.J. Cho and S.S. Kim, *Some componions of Gruss inequality in* 2*-inner product space and applications for determinantal integral inequalities*, Commum. Korean Math. Soc., 20(3), (2005), 487–503.

[5] S. Gahler, *Lineare* 2*-normierte Raume*, Math. Nachr., **28**, (1965), 1–43.

[6] Z. Lewandowska, *Bounded* 2*-linear operators on* 2*-normed sets*, Gals. Mat. Ser. III,**39**(59), (2004), 303-314.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some properties of fuzzy frames

V. Ebrahimi* and B. Daraby

Department of Mathematics, University of Maragheh, Maragheh, Iran

Abstract

Some results of fuzzy frames on fuzzy Hilbert spaces at the point of view of Bag and Samanta are proved.and we showed that all results in classical Hilbert spaces are immediate consequences of the corresponding results for Felbin-fuzzy Hilbert spaces.

Keywords: Fuzzy norm, Fuzzy inner product space, Fuzzy Hilbert space, Fuzzy frame
Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36 (At least one and at most three codes)

1 Introduction

The idea of fuzzy norms on a linear space first introduced by Katsaras [7] in 1984. Recently, B. Daraby and et al. [4] studied some properties of fuzzy Hilbert spaces and they showed that all results in classical Hilbert spaces are immediate consequences of the corresponding results for Felbin-fuzzy Hilbert spaces. Also by an example, they showed that the spectrum of the category of Felbin- fuzzy Hilbert spaces is broader than the category of classical Hilbert spaces [5].

One of the important concepts in the study of vector spaces is basis, which allows every vector to be uniquely represented as a linear combination of the basis elements. The main feature of a basis $\{x_k\}$ in a Hilbert space H is that every $x \in H$ can be represented as a linear combinations of the elements x_k in the form:

$$x = \sum_{k=1}^{\infty} c_k(x) x_k. \tag{1}$$

The coefficients $c_k(x)$ are unique. However, the linear independence property for a basis -which implies the uniqueness of coefficients- is restrictive in applications; sometimes it is impossible to find vectors which both fulfill the basis requirements and also satisfy external conditions demanded by applied problems. For such purposes, a more flexible types of spanning sets is needed. Frames provide these alternatives.

Many physical systems are inherently nonlinear functions and must be described by non-linear models. But some systems have of uncertain structured and it is not possible to provide an accurate mathematical model. Therefore, to these systems, we need to use a new concept namely fuzzy frames theory and fuzzy waveletes. Fuzzy frame and fuzzy wavelet inspired from frame theory, wavelet theory and fuzzy concepts. For achieving approximation functions, control and identification of nonlinear systems are presented.

*Speaker. Email address: m.ebrahimi@stu.maragheh.ac.ir

2 Some preliminaries

In this section, some definitions and preliminary results are given which will be used in this paper.

Definition 2.1. Let U be a linear space over the field $\mathbb{C}$ of complex numbers. Let $\mu : U \times U \times \mathbb{C} \longrightarrow I = [0,1]$ be a mapping such that the following hold:

(FIP1) for $s,t \in \mathbb{C}, \mu(x+y,z,|t|+|s|) \geq \min\{\mu(x,z,|t|), \mu(y,z,|s|)\}$;

(FIP2) for $s,t \in \mathbb{C}, \mu(x,y,|st|) \geq \min\left\{\mu\left(x,x,|s|^2\right), \mu\left(y,y,|t|^2\right)\right\}$;

(FIP3) for $t \in \mathbb{C}, \mu(x,y,t) = \mu\left(y,x,\overline{t}\right)$;

(FIP4) $\mu(\alpha x,y,t) = \mu(x,y,\frac{t}{|\alpha|}), \alpha(\neq 0) \in \mathbb{C}, t \in \mathbb{C}$;

(FIP5) $\mu(x,x,t) = 0, \forall t \in \mathbb{C}\backslash\mathbb{R}^+$;

(FIP6) $(\mu(x,x,t) = 1, \forall t > 0)$ iff $x = \underline{0}$;

(FIP7) $\mu(x,x,.) : \mathbb{R} \longrightarrow I$ is a monotonic non-decreasing function on $\mathbb{R}$ and $\lim_{t\to\infty} \mu(\alpha x,x,t) = 1$.

We call μ fuzzy inner product function on U and (U,μ) fuzzy inner product space (FIP space).

Theorem 2.2. *Let U be a linear space over $\mathbb{C}$. Let μ be a FIP on U. Then*

$$N(x,t) = \begin{cases} \mu(x,x,t^2) & \text{if } t \in \mathbb{R}, t > 0, \\ 0 & \text{if } t \leq 0. \end{cases}$$

is a fuzzy norm on U. Now if μ satisfies the following conditions:

(FIP8) $\left(\mu\left(x,x,t^2\right) > 0, \forall t > 0\right) \Rightarrow x = \underline{0}$ *and*

(FIP9) for all $x,y \in U$ and $p,q \in \mathbb{R}$,

$$\mu\left(x+y,x+y,2q^2\right) \bigwedge \mu\left(x-y,x-y,2p^2\right) \geq \mu\left(x,x,p^2\right) \bigwedge \mu\left(y,y,q^2\right),$$

then $\|x\|_\alpha = \bigwedge\{t > 0 : N(x,t) \geq \alpha\}, \alpha \in (0,1)$ is an ordinary norm satisfying parallelogram law. By using polarization identity, we can get ordinary inner product, called the $\langle .,. \rangle_\alpha$-inner product, as follows:

$$\langle x,y\rangle_\alpha = \frac{1}{4}\left(\|x+y\|_\alpha^2 - \|x-y\|_\alpha^2\right) + \frac{1}{4}i\left(\|x+iy\|_\alpha^2 - \|x-iy\|_\alpha^2\right), \forall \alpha \in (0,1).$$

Definition 2.3. Let (U,μ) be a FIP space satisfying $(FIP8)$. The linear space U is said to be level complete if for any $\alpha \in (0,1)$, every Cauchy sequence converges w.r.t. $\|.\|_\alpha$ (the α-norm generated by the fuzzy norm N which is induced by fuzzy inner product μ).

3 Main results

In this section, after a short introduction to history of frame, we define fuzzy frame and prove some new results.
Frames were introduced already in 1952 by Duffin and Schaeffer in their fundamental paper they used frames as a tool in the study of nonharmonic Fourier series, i.e., sequences of the type $\{e^{i\lambda_n}x\}_{n\in Z}$, where $\{\lambda_n\}_{n\in Z}$ is a family of real or complex numbers. Apparently, the importance of the concept was not realized by the mathematical community; at least it took almost 30 years before the next treatment appeared in print. Frames were presented in the abstract setting, and again used in the context of nonharmonic Fourier series. Then, in 1985, as the wavelet area began, Daubechies, Grossmann and Meyer observed that frames can be used to find series expansions of functions in $L^2(\mathbb{R})$ which are very similar to expansions using orthogonal bases.

Recall that for a Hilbert space H and a countable index set I, a family of vectors $\{x_i\}_{i\in I}\subseteq H$ is called a discrete frame for H, if there exist constants $0<A\leq B<+\infty$ such that

$$A\|x\|^2\leq\sum_{i\in I}|\langle x,x_i\rangle|^2\leq B\|x\|^2,\quad x\in H,$$

the constants A and B are called frame bounds. The frame $\{x_i\}_{i\in I}$ is called tight if $A=B$ and Parseval if $A=B=1$. For a very good and useful reference, we refer to the comprehensive book by Christensen [2]. The concept of frame improved and generalized to Banach spaces, Ferchet spaces and a lot of papers published in both pure and applied mathematics concerning frames. In this manuscript, we will try to present the fuzzy frame version of frame theorems and related concepts.

In a fuzzy Hilbert space (U,μ) satisfying $(FIP8)$ and $(FIP9)$ when $\alpha\in(0,1)$ and $\{e_k\}_{k=1}^{\infty}$ is an α-fuzzy orthonormal sequence in U, we say that $\{e_k\}_{k=1}^{\infty}$ is a basis for U if the following two conditions are satisfied:

(i) $U=span\,\{e_k\}_{k=1}^{\infty}$

(ii) $\{e_k\}_{k=1}^{\infty}$ is linearly independent.

So, every $x\in U$ has a unique representation in terms of the elements in the basis, i.e., there exist unique coefficients $\{\beta_k\}_{k=1}^{\infty}$ such that $x=\sum_{k=1}^{\infty}\beta_k e_k$. (By Theorem 2.2), if (U,μ) is a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$ and $x\in U$, then $\{e_k\}_{k=1}^{\infty}$ is fuzzy orthonormal sequence in U. Then, since $U=span\,\{e_k\}_{k=1}^{\infty}$, we can write $x=\sum_{k=1}^{\infty}\beta_k e_k$ and $\beta_k=\langle x,e_k\rangle_\alpha$.

Definition 3.1. Let (U,μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$. A countable family of elements $\{x_k\}_{k=1}^{\infty}$ in U is a fuzzy frame for U if there exist constants $A,B>0$ such that for all $x\in U$ and $\alpha\in(0,1)$:

$$A\|x\|_\alpha^2\leq\sum_{k=1}^{\infty}|\langle x,x_k\rangle_\alpha|^2\leq B\|x\|_\alpha^2. \tag{2}$$

The numbers A and B are called fuzzy frame bounds. Fuzzy frame bounds are not unique. The optimal lower frame bound is supremum over all lower frame bounds, and the optimal upper frame bound is the infimum over all upper frame bounds. Note that the optimal fuzzy frame bounds are actually fuzzy frame bounds. If $\|x_k\|_\alpha=1$, the fuzzy

frame is normalized. A fuzzy frame $\{x_k\}_{k=1}^{\infty}$ is tight if $A = B$ and in case $A = B = 1$, we call Parseval fuzzy frame. In case the upper inequality in (2) satisfy, $\{x_k\}_{k=1}^{\infty}$ is called fuzzy Bessel sequence. It follows from the definition that if $\{x_k\}_{k=1}^{\infty}$ is a fuzzy frame for (U, μ), then $\overline{span}\,\{x_k\}_{k=1}^{\infty} = U$.

Theorem 3.2. *Let (U, μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$ and $\alpha \in (0, 1)$ and $\{e_k\}_{k=1}^{\infty}$ be an α-fuzzy orthonormal sequence of U. Then for every $x \in U$,*

$$\sum_{k=1}^{\infty} |\langle x, x_k\rangle_\alpha|^2 \leq B\|x\|_\alpha^2.$$

Proof. Since α-fuzzy orthonormal sequence is orthonormal sequence in $(U, \langle ., .\rangle_\alpha)$, so by Bessel's inequality in crisp inner product we have

$$\sum_{k=1}^{\infty} |\langle x, x_k\rangle_\alpha|^2 \leq B\|x\|_\alpha^2 \quad \forall x \in U.$$

□

Example 3.3. Let $(U, \langle\ .\ \rangle)$ be a classic Hilbert space and let $\{x_n\}_{n=1}^{\infty}$ be a frame for U with frame bound A and B. Then $\{x_n\}_{n=1}^{\infty}$ is a fuzzy frame in fuzzy Hilbert space $(U, \langle\ .\ \rangle)$ satisfying (FIP8) and (FIP9) and $\alpha \in (0, 1)$.
Since $\{x_n\}_{n=1}^{\infty}$ is frame for U then there exist constants A, $B > 0$ such that

$$A\|x\|^2 \leq \sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2 \leq B\|x\|^2, \quad \forall x \in U$$

for all $x \in U$ and for any $\alpha \in (0, 1)$ we have

$$\frac{\alpha}{1-\alpha}A\|x\|^2 \leq \sum_{n=1}^{\infty} \frac{\alpha}{1-\alpha}\langle x, x_n\rangle|^2 \leq \frac{\alpha}{1-\alpha}B\|x\|^2.$$

It follows that

$$A\|x\|_\alpha^2 \leq \textstyle\sum_{n=1}^{\infty} |\langle x, x_n\rangle_\alpha|^2 \leq B\|x\|_\alpha^2, \quad \forall x \in U, \forall \alpha \in (0, 1).$$

Consider now a vector space U equipped with a fuzzy frame $\{x_k\}_{k=1}^{\infty}$, and define a linear mapping

$$T : l^2(\mathbb{N}) \longrightarrow U, \quad T\{\beta_k\}_{k=1}^{\infty} = \textstyle\sum_{k=1}^{\infty} \beta_k x_k$$

T is usually called the pre-fuzzy frame operator or the **fuzzy synthesis operator**. The adjoint operator is given by

$$T^* : U \longrightarrow l^2(\mathbb{N}), \quad T^*x = \{\langle x, x_k\rangle_\alpha\}_{k=1}^{\infty},$$

and it called the **fuzzy analysis operator**. Composing T with its adjoint T^*, we obtain the fuzzy frame operator,

$$S : U \longrightarrow U, \quad Sx = TT^*x = \sum_{k=1}^{\infty} \langle x, x_k\rangle_\alpha x_k.$$

Note that in terms of the fuzzy frame operator, we have

$$\langle Sx, x\rangle_\alpha = \sum_{k=1}^{\infty} |\langle x, x_k\rangle_\alpha|^2, \quad \forall x \in U.$$

Analogs to Theorem 3.2.3 of [2], the following Theorem shows that for given fuzzy Bessel its pre-frame operator is bounded and versa.

Theorem 3.4. *Let $\{x_k\}_{k=1}^{\infty}$ be a sequence in the fuzzy Hilbert space (U, μ) satisfying $(FIP8)$ and $(FIP9)$ and $B > 0$ be given. Then $\{x_k\}_{k=1}^{\infty}$ is a fuzzy Bessel sequence with fuzzy Bessel bound B if and only if the operator*

$$T : \{\beta_k\}_{k=1}^{\infty} \longrightarrow \sum_{k=1}^{\infty} \beta_k x_k$$

defines a fuzzy bounded linear operator from $l^2(\mathbb{N})$ into U and $\|T\|_\alpha \leq \sqrt{B}$ for all $\alpha \in (0, 1)$.

References

[1] B. T. Bilalov, S. M. Farahani, F. A. Guliyeva, The Intuitionistic Fuzzy Normed Space of Cofficients, Abstr. Appl. Anal., 2012(1): 1-11, 2018.

[2] O. Christensen, An introduction to frames and Riesz bases, 2003, Birkhauser.

[3] B. Daraby, Z. Solimani, A. Rahimi, Some Properties of Fuzzy Hilbert Spaces, Complex Anal. Oper. Theory., 11(1): 119-138, 2017.

[4] B. Daraby, Z. Solimani, A. Rahimi, A Note on Fuzzy Hilbert Spaces, J. Intell. Fuzzy Syst., 31(1): 313-319, 2016

[5] C. Felbin, Finite Dimensional Fuzzy Normed Linear Spaces, Fuzzy Sets Syst., 48(2): 239-248, 1992.

[6] A. Hasankhani, A. Nazari and M. Saheli, Some Properties if Fuzzy Hilbert Spaces and Norm of Operators, Iran. J. Fuzzy Syst., 7(3): 129- 157, 2010.

[7] A. K. Katsaras, Fuzzy Topological Vector Spaces II, Fuzzy Sets and Systems, 12(2): 143-154, 1984.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Positive definite kernels and reproducing kernel Hilbert spaces

Mohammadreza Foroutan[1,*] and Farzad Farzanfar[2]

[1]Department of Mathematics, Payame Noor University, P.O.Box 19395-3697, Tehran, Iran

[2]Department of Computer Engineering and Information Technology, Payame Noor University, P.O. Box 19395-3697 Tehran, Iran

Abstract

In this paper, we give aspects of positive definite functions with their respective reproducing kernel Hilbert spaces and applications. The following topics helps to understand reproducing kernel Hilbert spaces and their boundary spaces. We provide a characterization of kernel functions and reproducing kernel, derive their properties, and discuss methods for designing them. Given a kernel and a training set, we can form the matrix known as the kernel, or Gram matrix: the matrix containing the evaluation of the kernel function on all pairs of data points.

Keywords: Positive definite matrices, Gram matrix, Positive definite kernel, Reproducing kernel Hilbert space, Reproducing property.

Mathematics Subject Classification [2010]: 46E22, 47Bxx, 46Cxx

1 Introduction

In Linear algebra, a branch of mathematics, a positive-definite kernel is a generalization of a positive-definite function or a positive-definite matrix. It was first introduced by James Mercer in the early 20th century, in the context of solving integral operator equations. Since then, positive-definite functions and their various analogues and generalizations have arisen in diverse parts of mathematics. They occur naturally in Fourier analysis, probability theory, operator theory, complex function-theory, moment problems, integral equations, boundary-value problems for partial differential equations, machine learning, embedding problem, information theory, and other areas [4–7].
A positive definite function is synonymous with what we call a reproducing kernel, and the existence of are reproducing kernel implies the existence of a reproducing kernel Hilbert space. Reproducing kernel Hilbert spaces are Hilbert spaces H of functions with the property that the values $f(x)$ for $f \in H$ are reproduced from the inner product in H.

*Speaker. Email address: mr_forootan@pnu.ac.ir, foroutan_mohammadreza@yahoo.com

2 Positive definite matrices

All the eigenvalues of any symmetric matrix are real; this section is about the case in which the eigenvalues are positive. These matrices, which arise whenever optimization problems are encountered, have countless applications throughout science and engineering. They also arise in statistics and in geometry [7].

Definition 2.1. Let $M = (a_{ij})$ be an $n \times n$ complex matrix. We say that M is a positive semi-definite matrix if and only if for every $\{c_1, c_2, \cdots, c_n\} \subset \mathbb{C}$, we have that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}a_{ij} \geq 0.$$

When M is a positive semi-definite matrix, we write $M \geq 0$.

It is well known that this is the case if and only if M is hermitian (i.e. $a_{jk} = \overline{a_{kj}}$ for $j, k = 1, \cdots, n$) and the eigenvalues of M are all a non-negative.
We say M is a positive, $M > 0$, if $\sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}a_{ij} = 0$ implies that $c_i\overline{c_j} = 0$ for all $i, j = 1, 2, \cdots, n$.

Theorem 2.2. *If M is a positive semi-definite, $\langle Mx, x\rangle \geq 0$ for $x \in \mathbb{C}^n$ and for each eigenvalue λ_i of M, $\lambda_i \geq 0$ for $i = 1, 2, \cdots, n$.*

Proof. A proof is described in [2]. □

Corollary 2.3. *If M is a positive definite, $\langle Mx, x\rangle > 0$ for $x \in \mathbb{C}^n$ and for each eigenvalue λ_i of M, $\lambda_i > 0$ for $i = 1, 2, \cdots, n$.*

Theorem 2.4. *If M is a positive definite, then M is invertible.*

Proof. A proof is described in [2]. □

Theorem 2.5. *If M is a positive semi-definite, $M = M^*$. Moreover, there exists a matrix $N \in M_n(\mathbb{C})$ such that $M = N^*N$.*

Proof. A proof is described in [2]. □

3 Positive definite kernels

In this section, we consider their positive definiteness. A necessary and sufficient condition for the dot product kernel K to be positive definite is given. Given a kernel and a training set, we can form the matrix known as the kernel, or Gram matrix: the matrix containing the evaluation of the kernel function on all pairs of data points [1].

Definition 3.1. Let X be a non-empty set. Then a function $k : X \times X \longrightarrow \mathbb{K}$ is called a kernel on X if there exists a $\mathbb{K}$-Hilbert space H and a map $\Phi : X \longrightarrow H$ such that for all $x, x^{'} \in X$ we have

$$k(x, x^{'}) = \langle \Phi(x^{'}), \Phi(x)\rangle. \tag{1}$$

We call Φ a feature map and H a feature space of k.

Note that in the real case condition (1) can be replaced by the more natural equation $k(x,x^{'}) = \langle \Phi(x), \Phi(x^{'})\rangle$. In the complex case, however, $\langle .,. \rangle$ is anti-symmetric and hence (1) is equivalent to $k(x,x^{'}) = \overline{\langle \Phi(x), \Phi(x^{'})\rangle}$.
Given a kernel, neither the feature map nor the feature space are uniquely determined. Let us illustrate this with a simple. To this end, let $X := \mathbb{R}$ and $k(x,x^{'}) := xx^{'}$ for all $x, x^{'} \in \mathbb{R}$. Then k is a kernel since obviously the identity map $id_{\mathbb{R}}$ on $\mathbb{R}$ is a feature map with feather space $H := \mathbb{R}$. However, the map $\Phi : X \longrightarrow \mathbb{R}^2$ defined by $\Phi(x) := (\frac{x}{\sqrt{2}}, \frac{x}{\sqrt{2}})$ for all $x \in X$ is also a feature map of k since we have

$$\langle \Phi(x^{'}), \Phi(x)\rangle = \frac{x^{'}}{\sqrt{2}}\frac{x}{\sqrt{2}} + \frac{x^{'}}{\sqrt{2}}\frac{x}{\sqrt{2}} = x^{'}x = k(x,x^{'}),$$

for all $x, x^{'} \in X$. Moreover, note that a similar construction can be made for arbitrary kernels, and consequently every kernel has many different feature spaces.

Lemma 3.2. *Let X be a non-empty set and $f_n : X \longrightarrow \mathbb{K}$, $n \in \mathbb{N}$, be functions such that $(f_n(x)) \in \ell_2$ for all $x \in X$. Then*

$$k(x,x^{'}) := \sum_{n=1}^{\infty} f_n(x)\overline{f_n(x^{'})}, \quad x, x^{'} \in X, \tag{2}$$

defines a kernel on X.

Proof. Using Hölderś inequality for the sequences ℓ_1 and ℓ_2, we obtain

$$\sum_{n=1}^{\infty} |f_n(x)f_n(x^{'})| \leq \|f_n(x)\|_{\ell_2}\|f_n(x^{'})\|_{\ell_2},$$

and hence the series in (2) converges absolutely for all $x, x^{'} \in X$. Now, we write $H := \ell_2$ and define $\Phi : X \longrightarrow H$ by $\Phi(x) := (\overline{f_n(x)})$, $x \in X$. Then (2) immediately gives the assertions. □

Suppose $k : X \times X \longrightarrow \mathbb{R}$ is symmetric, i.e., $k(x,y) = k(y,x)$, $x, y \in X$. For $x_1, x_2, \cdots, x_n \in X$ $(n \geq 1)$, we say that the matrix

$$\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & \cdots & k(x_n,x_n) \end{pmatrix} \in \mathbb{R}^{n\times n}$$

is the Gram matrix w.r.t. k of order n. We say that k is a positive definite kernel if the Gram matrix of order n is positive semi-definite matrix for any $n \geq 1$ and $x_1, x_2, \cdots, x_n \in X$.

Example 3.3. We use the kernel $k : X \times X \longrightarrow \mathbb{R}$ such that

$$k_\lambda := D\Big(\frac{|x-y|}{\lambda}\Big),$$

where

$$D(t) = \begin{cases} \frac{3}{4}(1-t^2), & |t| \leq 1, \\ 0, & \text{otherwise}, \end{cases}$$

for $\lambda > 0$. The kernel k_λ does not satisfy positive definiteness. In fact, when $\lambda = 2,\ n = 3$ and $x_1 = -1,\ x_2 = 0,\ x_3 = 1$, the matrix consisting of $k_\lambda(x_i,\ y_i)$ can be written as

$$\begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & k(x_1,x_3) \\ k(x_2,x_1) & k(x_2,x_2) & k(x_2,x_3) \\ k(x_3,x_1) & k(x_3,x_2) & k(x_3,x_3) \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{9}{16} & 0 \\ \frac{9}{16} & \frac{3}{4} & \frac{9}{16} \\ 0 & \frac{9}{16} & \frac{3}{4} \end{pmatrix}$$

and the determinant is computed as $\frac{3^3}{2^6} - \frac{3^5}{2^{10}} - \frac{3^5}{2^{10}} = -\frac{3^3}{2^9}$. In general, the determinant of a matrix is the product of its eigenvalues, and we find that at least one of the eigenvalue is negative.

Example 3.4. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a positive semi-definite 2×2 matrix. Then

$$0 \leq \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = a + b + c + d,$$

from definition 2.1, we conclude that $Imb = -Imc$. On the other hand,

$$0 \leq \begin{pmatrix} 1 & i \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = a - ib + ic + d,$$

From definition 2.1, we conclude that $Reb = Rec$, i.e., $b = \overline{c}$. It follows immediately that any positive definite kernel is hermitian. Furthermore, for $z, w \in \mathbb{C}$ we have

$$\begin{aligned} \begin{pmatrix} w & z \end{pmatrix} \begin{pmatrix} a & \overline{c} \\ c & d \end{pmatrix} \begin{pmatrix} \overline{w} \\ \overline{z} \end{pmatrix} =& a|w|^2 + 2Re(cz\overline{w}) + d|z|^2 \\ =& a|w + \frac{c}{a}z|^2 + \frac{|z|^2}{a}(ad - |c|^2), \quad (\text{for } a \neq 0). \end{aligned}$$

The matrix is therefore positive definite if and only if $a \geq 0,\ d \geq 0$ and

$$det \begin{pmatrix} a & \overline{c} \\ c & d \end{pmatrix} = ad - |c|^2 \geq 0.$$

Hence for any positive definite kernel φ, since φ is symmetric, then with supposes $a := \varphi(x,x)$, $c := \varphi(x,y)$ and $d := \varphi(y,y)$ with using above equation, we have

$$|\varphi(x,y)|^2 \leq \varphi(x,x).\varphi(y,y).$$

4 Reproducing kernel Hilbert spaces

Positive-definite kernels provide a framework that encompasses some basic Hilbert space constructions. In the following we present a tight relationship between positive-definite kernels and two mathematical objects, namely reproducing Hilbert spaces and feature maps.

Definition 4.1 ([3]). Let H be a Hilbert space whose elements are functions $f : X \longrightarrow \mathbb{R}$. A function $k : X \times X \longrightarrow \mathbb{R}$ is said to be a reproducing kernel of a Hilbert space H with an inner product $\langle .,. \rangle_H$ if it satisfies the following two conditions.

1. For each $x \in X$, we have
$$k(x,.) \in H. \tag{3}$$

2. Reproducing property: for each $f \in H$ and $x \in X$,

$$f(x) = \langle f, k(x, .)\rangle_H. \tag{4}$$

When H has a reproducing kernel, we say that H is a reproducing kernel Hilbert space. The reproducing property (4) is called a kernel trick.

Example 4.2. Let $\{e_1, e_2, \cdots, e_n\}$ be an orthonormal basis of a finite-dimensional Hilbert space H. If we define

$$k(x, y) = \sum_{i=1}^{n} e_i(x)e_i(y), \tag{5}$$

for $x, y \in X$, then we have $k(x, .) \in H$ and

$$\langle e_j(.), k(x, .)\rangle_H = \sum_{i=1}^{n} \langle e_j, e_i\rangle_H e_i(x) = e_j(x),$$

for each $1 \leq j \leq n$. Thus, for any $f(.) = \sum_{i=1}^{n} f_i e_i(.) \in H$, $f_i \in \mathbb{R}$, we have $\langle f, k(x, .)\rangle_H = f(x)$ (reproducing property). Therefore, H is a reproducing kernel Hilbert space, and (5) is a reproducing kernel.

Example 4.3. Let E be a finite set $\{x_1, x_2, \cdots, x_n\}$, and let $k : E \times E \longrightarrow \mathbb{R}$ be a positive definite kernel. Then, the linear space

$$H := \{\sum_{i=1}^{n} \alpha_i k(x_i, .) \mid \alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}\},$$

is a reproducing kernel Hilbert space. We define the inner product by

$$\langle f(.), g(.)\rangle_H = a^T K b,$$

for $f(.), g(.) \in H$, where $f(.) = \sum_{j=1}^{n} a_j k(x_j, .) \in H$, $a = [a_1, a_2, \cdots, a_n]^T \in \mathbb{R}^n$ and $g(.) = \sum_{j=1}^{n} b_j k(x_j, .) \in H$, $b = [b_1, b_2, \cdots, b_n]^T \in \mathbb{R}^n$ via the Gram matrix

$$K := \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{pmatrix}$$

Then, for each x_i, $i = 1, 2, \cdots, n$, we have

$$\langle f(.), k(x_i, .)\rangle_H = [a_1, a_2, \cdots, a_n] K e_i = \sum_{j=1}^{n} a_j k(x_j, x_i) = f(x_i),$$

(reproducing property), where e_i is an n-dimensional column vector in which we set component i and the other components to 1 and 0, respectively.

Lemma 4.4. *The reproducing kernel k of the reproducing kernel Hilbert space H is unique, symmetric $k(x, y) = k(y, x)$, and positive semi-definite.*

Proof. If k_1 and k_2 are reproducing kernel Hilbert space of H, then by the reproducing property, we have that

$$f(x) = \langle f, k_1(x, .)\rangle_H = \langle f, k_2(x, .)\rangle_H.$$

In other words, $\langle f, k_1(x,.) - k_2(x,.)\rangle_H = 0$ holds for all $f \in H$, $x \in X$ for which $k_1 = k_2$. Additionally, the symmetry of a reproducing kernel follows from that of its inner product:

$$k(x,y) = \langle k(x,.), k(y,.)\rangle_H = \langle k(y,.), k(x,.)\rangle_H = k(y,x).$$

The positive semi-definiteness of the reproducing kernel can be shown as follows.

$$\sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_j k(x_i,x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_j \langle k(x_i,.), k(x_j,.)\rangle_H = \langle \sum_{i=1}^{n} z_i k(x_i,.), \sum_{j=1}^{n} z_j k(x_j,.)\rangle_H.$$

□

References

[1] N. Aronszajn, *Reproducing and pseudo-reproducing kernels and their application to the partial differential equations of physics. Studies in partial differential equations*, Technical report 5, preliminary note. Harvard University, Graduate School of Engineering., 1948.

[2] R. Bhatia, *Positive definite matrices. In Positive Definite Matrices*, Princeton university press, 2009.

[3] M. R. Foroutan, A. Ebadian and R. Asadi, Reproducing kernel method in Hilbert spaces for solving the linear and nonlinear four-point boundary value problems, *Int. J. Comput. Math.,*95(10) (2018), 2128–2142.

[4] M. R. Foroutan, A. S. Gholizadeh, Sh. Najafzadeh, R. H. Haghi, Laguerre reproducing kernel method in Hilbert spaces for unsteady stagnation point flow over a stretching/shrinking sheet, *Appl. Math. J. Chinese Univ.,* 36(3) (2021), 354–-369.

[5] M. R. Foroutan, M. S. Hashemi, L. Gholizadeh, A. Akgül, F. Jarad,A new application of the Legendre reproducing kernel method , *AIMS Mathematics,* 7(6) (2022), 10651–-10670.

[6] P. E. T. Jorgensen, F. Tian, Discrete reproducing kernel Hilbert spaces: sampling and distribution of Dirac-masses, *J. Mach. Learn. Res.,* 16 (2015), 3079-–314.

[7] P. E. T. Jorgensen, F. Tian, Positive definite kernels and boundary spaces, *Adv. Oper. Theory,* 1 (2016), 123-–133.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Application of Fourier series in deriving stability polynomial of multivalue methods for ODEs

M. Sharifi[1,*], A. Abdi[1,2]

[1]Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran

[2]Research Department of Computational Algorithms and Mathematical Models, University of Tabriz, Tabriz, Iran

Abstract

We construct second derivative diagonally implicit multistage integration methods (SDIMSIMs) as a subclass of second derivative general linear methods (SGLMs) with Runge–Kutta stability property (RKS). The conditions arise from RKS which are a system of polynomial equations can not be produced by symbolic manipulation packages for the methods of order $p \geq 5$. In this paper, we describe an approach to construct SDIMSIMs with RKS property by using some variant of the Fourier series method. We construct explicit and implicit SDIMSIMs of order 5 and 6 which respectively are appropriate for both non–stiff and stiff differential systems and show their efficiency by applying to some well–known problems.

Keywords: Second derivative methods, Order conditions, A– and L–stability, Fourier series.

Mathematics Subject Classification [2010]: 65L05

1 Introduction

Second derivative diagonally implicit multistage integration methods (SDIMSIMs) as a subclass of SGLMs for the numerical solution

$$\begin{cases} y'(x) = f\big(y(x)\big), \quad x \in [x_0, \bar{x}], \\ y(x_0) = y_0, \end{cases} \tag{1}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ and $y : \mathbb{R} \to \mathbb{R}^m$, have been introduced in [1]. SDIMSIMs for solving (1) on the uniform grid $x_n = x_0 + nh$, $n = 1, 2, \ldots, N$, $Nh = \bar{x} - x_0$, take the form

$$\begin{aligned} Y_i^{[n]} &= h \sum_{j=1}^{s} a_{ij} f\big(Y_j^{[n]}\big) + h^2 \sum_{j=1}^{s} \overline{a}_{ij} g\big(Y_j^{[n]}\big) + \sum_{j=1}^{r} u_{ij} y_j^{[n-1]}, \ i = 1, 2, \ldots, s, \\ y_i^{[n]} &= h \sum_{j=1}^{s} b_{ij} f\big(Y_j^{[n]}\big) + h^2 \sum_{j=1}^{s} \overline{b}_{ij} g\big(Y_j^{[n]}\big) + \sum_{j=1}^{r} v_{ij} y_j^{[n-1]}, \ i = 1, 2, \ldots, r, \end{aligned} \tag{2}$$

*Speaker. Email address: moh.sharifi@tabrizu.ac.ir

where s is the number of internal stages, r is the number of external stages, and h is the stepsize. Here, the vector of stage values $Y^{[n]} := \left[Y_i^{[n]}\right]_{i=1}^{s}$ is an approximation of stage order q to the vector $\left[y(x_{n-1}+c_ih)\right]_{i=1}^{s}$, and the function g stands for the second derivative of the solution, $g(\cdot) = f'(\cdot)f(\cdot)$. Moreover, the input and output vectors $y^{[n-1]} = \left[y_i^{[n-1]}\right]_{i=1}^{r}$ and $y^{[n]} = \left[y_i^{[n]}\right]_{i=1}^{r}$ are approximation of order p to the linear combinations of scaled derivatives of the solution at the points x_{n-1} and x_n.

According to a standard linear stability analysis, the stability matrix of these methods takes the form $M(z) = V + (zB + z^2\overline{B})(I_s - zA - z^2\overline{A})^{-1}U$ in which z is a complex number and I_s stands for the identity matrix of dimension s. Then the stability function is defined as the characteristic polynomial of the stability matrix $M(z)$, given by

$$p(w, z) = \det(wI - M(z)),$$

where $w \in \mathbb{C}$. The method is said to possess Runge–Kutta stability (RKS) property, if its stability function has the form

$$p(w, z) = w^{r-1}(w - R(z)).$$

Depending on the nature of the differential system to be solved and the computer architecture that is used to implement these methods, the authors in [1], by considering the matrices A and $\overline{A}$ in lower triangular form, have divided SGLMs into four types. Types 1 and 2 are those with arbitrary a_{ij}, $\overline{a}_{ij}$ where $\lambda = \mu = 0$ and $\lambda > 0$, $\mu < 0$, respectively. Such methods are appropriate respectively for non–stiff and stiff differential systems in a sequential computing environment. Requiring $a_{ij} = \overline{a}_{ij} = 0$, cases $\lambda = \mu = 0$ and $\lambda > 0$, $\mu < 0$ lead respectively to types 3 and 4 methods which can be useful respectively for non-stiff and stiff systems in a parallel computing environment.

2 The structure of order conditions

The structure of the order conditions for SGLMs in their general form has been investigated in [3]. The fundamental concept is to use input vector of the form

$$y_i^{[n-1]} = \sum_{k=0}^{p} h^k \alpha_{ik} y^{(k)}(x_{n-1}) + \mathcal{O}(h^{p+1}), \qquad i = 1, 2 \ldots, r, \tag{3}$$

for some real parameters α_{ik}, $i = 1, 2, \ldots, r$, $k = 0, 1, \ldots, p$, where $y_i^{[n]}$ denotes approximation number i at integration point number n. We then request that the stage values $Y_i^{[n]}$ within the current step with stepsize h be approximations of order q to the solution at the points $x_{n-1} + c_ih$, that is,

$$Y_i^{[n-1]} = \sum_{k=0}^{p} \frac{c_i^k}{k!} h^k y^{(k)}(x_{n-1}) + \mathcal{O}(h^{q+1}), \qquad i = 1, 2 \ldots, s, \tag{4}$$

and the output values computed at the end of current step satisfy

$$y_i^{[n]} = \sum_{k=0}^{p} h^k \alpha_{ik} y^{(k)}(x_n) + \mathcal{O}(h^{p+1}), \qquad i = 1, 2 \ldots, r, \tag{5}$$

for the same numbers α_{ik}. Let us denote $\alpha_k := \begin{bmatrix}\alpha_{1k} & \alpha_{2k} & \cdots & \alpha_{rk}\end{bmatrix}^T$ for $k = 0, 1, \ldots, p$. Pre-consistency and consistency vectors are denoted by α_0 and α_1, respectively. Also let us denote $Z := \begin{bmatrix}1 & z & \cdots & z^p\end{bmatrix}^T \in \mathbb{C}^{p+1}$ and collect the vectors α_k in the matrix W as

$$W = \begin{bmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_p\end{bmatrix}.$$

Theorem 2.1. *[3] Assume that $y^{[n-1]}$ satisfies (3). Then the SGLM (2) of order p and stage order $q = p$ satisfies (4) and (5) if and only if*

$$\exp(cz) = zA\exp(cz) + z^2\overline{A}\exp(cz) + UWZ + \mathcal{O}(z^{p+1}), \tag{6}$$

$$\exp(z)WZ = zB\exp(cz) + z^2\overline{B}\exp(cz) + VWZ + \mathcal{O}(z^{p+1}). \tag{7}$$

Here, the exp *function is applied component–wise to a vector.*

3 High order SDIMSIMs with RKS property

SDIMSIMs of orders up to four in various types with RKS property have been derived in [2] by solving the generated nonlinear equations related to RKS conditions by symbolic manipulation packages such as MATHEMATICA or MAPLE. Symbolic manipulation tools could no longer produce the corresponding systems of nonlinear equations in a reasonable form for higher orders ($p \geq 5$); therefore another approach to construct such methods is required. In this paper, we describe the construction of high order SDIMSIMs of type 1 and 2 using the Fourier series approach which has been already used in the context of diagonally implicit multistage integration methods (DIMSIMs) in [4].

For type 1 and 2 the stability function $p(w, z)$ has the form

$$p(w, z) = w^s - p_{s-1}(z)w^{s-1} + \cdots + (-1)^{s-1}p_1(z)w + (-1)^s p_0(z),$$

and

$$p(w, z) = (1 - \lambda z - \mu z^2)^s w^s - p_{s-1}(z)w^{s-1} + \cdots + (-1)^{s-1}p_1(z)w + (-1)^s p_0(z),$$

respectively, where $p_i(z)$, $i = 0, 1, \ldots, s-1$, are polynomial of degree less than or equal to $2s-1$ with respect to z. Then, the method (2) has RKS property if we impose the conditions

$$p_i(z) \equiv 0, \qquad i = 0, 1, \ldots, s-2,$$

or equivalently,

$$p_{i,k} = 0, \qquad i = 0, 1, \ldots, s-2, \qquad k = s-1-i, s-i, \ldots, s+i.$$

Here, for producing $p_{i,k}$, we use Fourier series approach. Let

$$w_\zeta = \exp\left(-\frac{2\pi\zeta\,\mathrm{i}}{N_1}\right), \qquad \zeta = 0, 1, \ldots, N_1 - 1,$$

and

$$z_\eta = \exp\left(-\frac{2\pi\eta\,\mathrm{i}}{N_2}\right), \qquad \eta = 0, 1, \ldots, N_2 - 1,$$

with i as the imaginary unit, are complex numbers uniformly distributed on the unit circle, and N_1 and N_2 are sufficiently large integers. Multiplying $p(w_\zeta, z)$ by w_ζ^{-j} and summing on ζ we obtain

$$p_j(z) = (-1)^{s-j}\frac{1}{N_1}\sum_{\zeta=0}^{N_1-1} w_\zeta^{-j}\, p\left(w_\zeta, z\right), \qquad j = 0, 1, \ldots, s-1. \tag{8}$$

Similarly, multiplying

$$p_j\left(z_\eta\right) = p_{j,0} + p_{j,1}z_\eta + \cdots + p_{j,s}z_\eta^s,$$

by z_η^{-k} and summing on η, we get

$$p_{j,k} = \frac{1}{N_2} \sum_{\eta=0}^{N_2-1} z_\eta^{-k} p_j(z_\eta), \qquad k = s-1-j, s-j, \ldots, s+j.$$

Then, substituting (8) into the last relation, we obtain

$$p_{j,k} = (-1)^{s-j} \frac{1}{N_1 N_2} \sum_{\zeta=0}^{N_1-1} \sum_{\eta=0}^{N_2-1} w_\zeta^{-j} z_\eta^{-k} p(w_\zeta, z_\eta), \tag{9}$$

for $j = 0, 1, \ldots, s-1$, $k = s-1-j, s-j, \ldots, s+j$. We solve numerically systems of nonlinear equations (9) corresponding to the RKS conditions using subroutine `fsolve.m` utilizing the algorithm 'levenberg–marquardt' in MATLAB.

We will be mainly interested in explicit and implicit SDIMSIMs of order 5 and 6 with $p = q = r = s$, which the implicit proposed methods are A–, $A(\alpha)$–, or L–stable.

4 Numerical results

The proposed types 1 and 2 SDIMSIMs are implemented on some non–stiff, mildly stiff and stiff problems. To compare, the results of DIMSIMs in [4] of the same orders are also reported.

Computational experiments for the proposed methods are done by applying these methods on the following problems:

P1. The famous nonlinear **Van Der Pol** system [5]

P2. The **BRUSSELATOR** problem [5]

Table 1: Numerical results of type 1 SDIMSIM of orders $p = q = r = s = 5$ and 6 for the problem P2 with $\epsilon = 10^{-1}$.

h	Type 1 SDIMSIM of order 5		Type 1 SDIMSIM of order 6	
	$\|e_h(\overline{x})\|$	p	$\|e_h(\overline{x})\|$	p
$\frac{1}{8}$	1.27e-6		3.48e-4	
$\frac{1}{16}$	3.04e-8	5.39	1.57e-7	11.11
$\frac{1}{32}$	7.96e-10	5.26	3.51e-9	5.49
$\frac{1}{64}$	1.44e-11	5.79	4.67e-11	6.23
$\frac{1}{128}$	1.50e-13	6.58	5.34e-13	6.45

For problem P1, we have presented the numerical results in Tabel 1 and Figures 1–2 for various value of ϵ, respectively. Alos, in Figure 3 numerical results have reported for problem P2.

5 Conclusion

In this paper, we described the construction of SDIMSIMs of order $p = 5$ and $p = 6$, with $p = q = r = s$, using a variant of Fourier series approach [4]. The numerical experiments

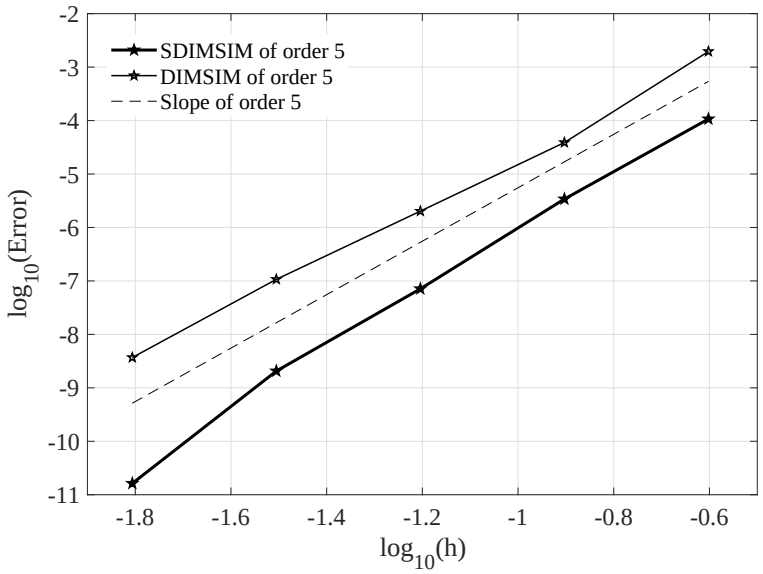


Figure 1: Numerical results of type 2 methods of orders $p = 5$ for the problem P2 with $\epsilon = 10^{-6}$.

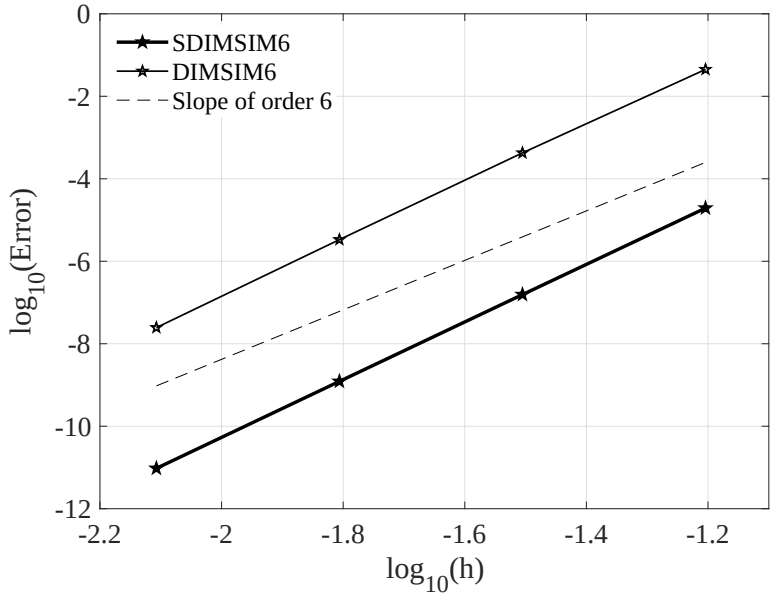


Figure 2: Numerical results of type 2 methods of orders $p = 6$ for the problem P2 with $\epsilon = 10^{-6}$.

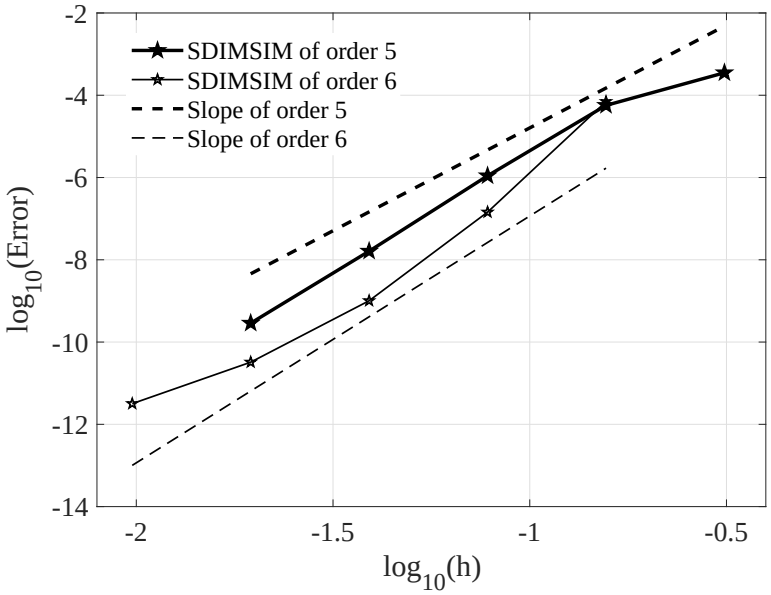


Figure 3: Numerical results of type 2 SDIMSIMs of orders $p = 5$ and $p = 6$ for the problem P3.

confirmed that the proposed methods are efficient in solving non–stiff and stiff problems and also verified their order of convergence.

References

[1] A. Abdi, G. Hojjati, An extension of general linear methods, *Numer. Algorithms*, 57 (2011), 149–167.

[2] A. Abdi, M. Braś, G. Hojjati, On the construction of second derivative diagonally implicit multistage integration methods for ODEs, *Appl. Numer. Math*, 76 (2014), 1–18.

[3] A. Abdi, G. Hojjati, Maximal order for second derivative general linear methods with Runge–Kutta stability, *Appl. Numer. Math*, 61 (2011), 1046–1058.

[4] J.C. Butcher, Z. Jackiewicz, Construction of high order diagonally implicit multistage integration methods for ordinary differential equations, *Appl. Numer. Math*, 27 (1998), 1–12.

[5] E. Hairer, G. Wanner, Solving ordinary differential equations II: Stiff and Differential–Algebraic Problems, *Springer Berlin Heidelberg*, 2010.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On the stability analysis of a class of multistep collocation methods for ODEs

L. Taheri koltape*, G. Hojjati, S. Fazeli

Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran

Abstract

In this paper, we study a class of of multistep collocation methods for the numerical solution of initial value problems in ordinary differential equations. By analysing the linear stability of the introduced methods, we aim to construct algorithms with substantial stability properties of high convergence order.

Keywords: Initial value problem, Collocation methods, Linear Stability, Stiff problems.

Mathematics Subject Classification [2010]: 65L05.

1 Introduction

We consider the initial value problem (IVP) in ordinary differential equations (ODEs)

$$\begin{aligned} y'(x) &= f(y(t)), \quad t \in [t_0, T], \\ y(t_0) &= y_0, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ is assumed to satisfy the conditions to guarantee the existence of a unique solution and d is the dimensionality of the system. Collocation is a widely applied and powerful technique in the construction of numerical methods for ODEs. The systematic study of collocation methods for IVPs has its origin in the late 60's [1, 2]. Collocation methods provide an approximation over the entire integration interval to the solution of the equation and they have high order of convergence, strong stability properties and flexibility. The idea of multistep collocation was first introduced by Lie and Nørsett in [5]. Important special cases of multistep collocation methods are the one-leg methods of Dahlquist [3] and the BDF-methods. The multistep collocation method uses information from multiple previous time steps to approximate the solution at the current time step that this allows for greater accuracy and stability in the numerical solution. Due to its concurrency capabilities, this method can be also used to solve large-scale problems. For instance, in economic and social simulations, this method is used to solve dynamic system equations and analyze their behavior. The stability of the method can be affected by the stiffness of the ODE problem being solved. Stiff problems require special treatment to ensure stability, such as using implicit methods or adaptive time-stepping. Overall, the stability of the multistep collocation method can be ensured through careful selection of the method parameters and appropriate treatment of stiff problems.

*Speaker. l.taheri@tabrizu.ac.ir

2 Collocation based methods for ODEs

We are going to briefly review the structure of collocation based methods for ODEs. Let us suppose that the integration interval $[t_0, T]$ is discretized in an uniform grid $\{t_h : t_0 < t_1 < ... < t_N = T\}$. Here, $t_n + c_i h, i = 1, 2, \ldots, m$, where $c_1, c_2, \ldots, c_m$ are given collocation points.

2.1 One-step Collocation methods

In classical one-step collocation methods [4], the collocation function is given by an algebraic polynomial $P(t), t \in [t_n, t_{n+1}]$ satisfying

$$\begin{aligned} &P(t_n) = y_n, \\ &P'(t_n + c_i h) = f(P(t_n + c_i h)), \quad i = 1, 2, \ldots, m, \end{aligned}$$

i.e. interpolating the numerical solution in t_n and exactly satisfying the given system in collocation points. The solution at t_{n+1} can then be computed from the function evaluation $y_{n+1} = P(t_{n+1})$.

2.2 Multistep Collocation methods

Multistep collocation methods are constructed by adding interpolation conditions in the previous k step points, so that the collocation polynomial is defined by

$$\begin{cases} P(t_{n-l}) = y_{n-l}, \quad l = 0, 1, \ldots, k-1, \\ P'(t_n + c_j h) = f(P_n(t_n + c_j h)), \quad j = 1, 2, \ldots, m. \end{cases}$$

The continuous approximant is given by

$$P(t_n + sh) = \sum_{l=0}^{k-1} \varphi_l(s) y_{n-l} + h \sum_{j=1}^{m} \chi_j(s) f(P(t_n + c_j h)), \tag{2}$$

with $s \in [0, 1]$ and the numerical solution is then $y_{n+1} = P(t_{n+1})$. It is proved that these methods have uniform order $m + k - 1$ on the whole interval of integration.

3 Stability analysis of a class of multistep collocation methods

A class of multistep collocation method is introduced by the formula

$$P(t_n + sh) = \sum_{l=0}^{k-1} \varphi_l(s) y_{n-l} + h \sum_{j=1}^{m} \chi_j(s) f(P(t_n + c_j h)) + h^2 \sum_{j=1}^{m} \overline{\chi}_j(s) g(P(t_n + c_j h)), \tag{3}$$

with $s \in [0, 1]$ and $g(.) = f'(.)f(.)$. The collocation polynomial (3) is an algebraic polynomial which is derived in order to satisfy some interpolation and collocation conditions, i.e.

$$\begin{cases} P(t_{n-l}) = y_{n-l}, \quad l = 0, 1, \ldots, k-1, \\ P'(t_n + c_j h) = f(P(t_n + c_j h)), \quad P''(t_n + c_j h) = g(P(t_n + c_j h)), \quad j = 1, 2, \ldots, m. \end{cases}$$

To analyze the stability properties of the methods (3), we apply the method to the Dahlquist test problem $y' = \lambda y,\ \ \lambda \in \mathbb{C}$. Defining

$$P(t_n + ch) = \begin{bmatrix} P(t_n + c_1 h) \\ P(t_n + c_2 h) \\ \vdots \\ P(t_n + c_m h) \end{bmatrix}, \qquad \overline{y} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{n-(k-1)} \end{bmatrix},$$

$$w^T = \begin{bmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_m(1) \end{bmatrix}, \qquad u^T = \begin{bmatrix} \overline{\chi}_1(1) & \overline{\chi}_2(1) & \dots & \overline{\chi}_m(1) \end{bmatrix},$$

$$A = \begin{bmatrix} \varphi_0(c_1) & \varphi_1(c_1) & \dots & \varphi_{k-1}(c_1) \\ \varphi_0(c_2) & \varphi_1(c_2) & \dots & \varphi_{k-1}(c_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_0(c_m) & \varphi_1(c_m) & \dots & \varphi_{k-1}(c_m) \end{bmatrix},$$

$$B = \begin{bmatrix} \chi_1(c_1) & \chi_2(c_1) & \dots & \chi_m(c_1) \\ \chi_1(c_2) & \chi_2(c_2) & \dots & \chi_m(c_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(c_m) & \chi_2(c_m) & \dots & \chi_m(c_m) \end{bmatrix},$$

$$C = \begin{bmatrix} \overline{\chi}_1(c_1) & \overline{\chi}_2(c_1) & \dots & \overline{\chi}_m(c_1) \\ \overline{\chi}_1(c_2) & \overline{\chi}_2(c_2) & \dots & \overline{\chi}_m(c_2) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\chi}_1(c_m) & \overline{\chi}_2(c_m) & \dots & \overline{\chi}_m(c_m) \end{bmatrix},$$

we have

$$\begin{cases} P(t_n + ch) = A\overline{y} + zBP(t_n + ch) + z^2 CP(t_n + ch), \\ y_{n+1} = \sum_{l=0}^{k-1} \varphi_l(s) y_{n-l} + zw^T P(t_n + ch) + z^2 u^T P(t_n + ch), \end{cases}$$

$n = 1, 2, \dots, N-1$. Hence,

$$P(t_n + ch) = (I - zB - z^2 C)^{-1} A\overline{y},$$

where $z = \lambda h$. Substituting this relation into the equation for y_{n+1} leads to

$$y_{n+1} = \sum_{l=0}^{k-1} \varphi_l(s) y_{n-l} + (zw^T + z^2 u^T)(I - zB - z^2 C)^{-1} A\overline{y}.$$

This equation can be written in the following form

$$\det(I - zB - z^2 C) y_{n+1} = \det(I - zB - z^2 C) \sum_{l=0}^{k-1} \varphi_l(s) y_{n-l} + (zw^T + z^2 u^T) adj(I - zB - z^2 C) A\overline{y},$$

which can be rewritten in the form

$$q_0(z) y_{n+1} + q_1(z) y_n + \ldots + q_k(z) y_{n-(k-1)} = 0.$$

So, the stability properties of the methods are governed by the characteristic polynomial

$$P(w,z) = w^k - \sum_{i=1}^{k} q_i(z)(w)^{i-1},$$

where $w \in \mathbb{C}$. The stability behaviour of the method corresponds to the roots of the *stability function* $P(w,z)$. In the construction of the introduced methods, after satisfying the order conditions, the main free parameters are chosen in such away that the method has large absolute stability region.

References

[1] H. Brunner, *Collocation Methods for Volterra Integral and Related Functional Equations*, Cambridge University Press, 2004.

[2] J.C. Butcher, *Numerical Methods for Ordinary Differential Equations*, Wiley, New York, 2016.

[3] G. Dahlquist, On one-leg multistep methods, *SIAM J. Numer. Anal.*, 20 (1983), No. 6, 1130–1138.

[4] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II – Stiff and Differential–Algebraic Problems*, Springer–Verlag, Berlin, 2002.

[5] I. Lie, S.P. Nørsett, Superconvergence for Multistep Collocation, *Math. Comp.*, 52 (1989), No. 185, 65–79.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Estimating the Estrada Index

Hamid Reza Bamdad*

[1]Department of Mathematics, Payame Noor University (PNU), Tehran, Iran

Abstract

Let G be an n-vertex graph. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of its adjacency matrix, the Estrada index of G defined as $\mathrm{EE}(G) = \sum_{i=1}^{n} e^{\lambda_i}$. In this paper by use of the linier algebraic tools we obtain a lower bound for $\mathrm{EE}(G)$ and show that it is best possible.

Keywords: Estrada index, Adjacency matrix, Complete graph, Spectrom of graphs
Mathematics Subject Classification [2010]: 05C50

1 Introduction

Throughout this paper we consider simple graphs, that is finite and undirected graphs without loops and multiple edges. If G is a graph with vertex set $\{1, \ldots, n\}$, the *adjacency matrix* of G is an $n \times n$ matrix $A = [a_{ij}]$, where $a_{ij} = 1$ if there is an edge between the vertices i and j, and $a_{ij} = 0$ otherwise. Since A is a real symmetric matrix, its eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are real numbers. These are referred to as the eigenvalues of G. In what follows we assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The multiset of eigenvalues (represent a set of eigenvalues, where the same eigenvalues can appear multiple times) of A is called the *spectrum* of G. For details of the theory of graph spectra see [1, 2]. We denote the complete graph on n vertices by K_n, the complete multipartite graph whose parts are of orders $n_1, \ldots, n_k$ by $K_{n_1,\ldots,n_k}$. We occasionally denote the regular complete k-partite graph $K_{t,\ldots,t}$ by $K_{k\times t}$. The *energy* of G is defined as [5, 6]

$$\mathrm{E} = \mathrm{E}(G) = \sum_{i=1}^{n} |\lambda_i| \ .$$

The *Estrada index* of G, recently put forward by Ernesto Estrada [3], is defined as

$$\mathrm{EE} = \mathrm{EE}(G) = \sum_{i=1}^{n} e^{\lambda_i} \ ,$$

wher e denotes the Napier's constant. The Estrada index has a wide range of applications in various fields. It is a powerful and versatile tool for analyzing the structure and function of complex systems in various fields, ranging from physics and chemistry to biology, social sciences, and engineering. Its applications are continually expanding as researchers develop new techniques for social network analysis, physics, chemistry, biology and computer sciences as follows:

*Speaker. Email address: hr_bamdad@pnu.ac.ir

1. Social network analysis: The Estrada index has been used to study social networks, such as online social networks and communication networks. It has been shown that the Estrada index can provide insights into the influence and centrality of individuals in the network.

2. Quantum systems: The Estrada index has been used to study the entanglement properties of quantum systems. Specifically, it has been shown that the Estrada index of the density matrix of a system can be used to quantify the amount of entanglement in the system.

3. Chimistry: The Estrada index in chemistry is in the study of molecular structure and properties. The Estrada index has been used to quantify the complexity and aromaticity of molecules, which are important properties that affect the reactivity and stability of molecules.

4. Biological networks: The Estrada index has been used to study the structure and function of biological networks, such as protein-protein interaction networks and gene regulatory networks.

5. Machine learning: The Estrada index has been used as a feature in machine learning algorithms for network analysis, such as link prediction and community detection.

2 Main results

We begin with the following:

Lemma 2.1. *Let G be a connected graph with adjacency spectrum $\{[r]^1, [0]^t, [s]^{k-1}\}$, such that $r > 0 > s$, $t \geq 0$, and $k \geq 3$. Then G is a regular complete multipartite graph.*

Proof. As G has only one positive eigenvalue, by ([1], Theorem 6.7), it is a complete multipartite graph $K_{n_1,\dots,n_k}$, say, where $n_1 + \cdots + n_k = n$. It is known that the characteristic polynomial of the adjacency matrix of $K_{n_1,\dots,n_k}$ is of the form (see [1, p. 74])

$$\phi(\lambda) = \lambda^{n-k} \left(\prod_{i=1}^{k} (\lambda + n_i) - \sum_{i=1}^{k} \frac{n_i \prod_{j=1}^{k} (\lambda + n_j)}{\lambda + n_i} \right) .$$

If for some $i \neq j$, $n_i = n_j$, then from the above formula it is seen that $\phi(-n_i) = 0$. It follows that, if $n_{i_1} = n_{i_2} \neq n_{i_3} = n_{i_4}$ for four different indices i_1, i_2, i_3, i_4, then $-n_{i_1}$ and $-n_{i_2}$ are eigenvalues of G. This is impossible since G has only one negative eigenvalue. Therefore, we see that at least $k-1$ of n_i's must be mutually equal. Suppose that, without loss of generality, $n_1 = \cdots = n_{k-1}$. This means that $-n_1$ is a root of $\phi(\lambda)$ with multiplicity $\geq k-2$. If $n_1 = n_k$, then G is regular, and we are done. So assume that $n_k \neq n_1$. The only remaining negative root of $\phi(\lambda)$ is the root of the auxiliary polynomial $\psi(\lambda)$,

$$\psi(\lambda) := \frac{\phi(\lambda)}{(\lambda + n_1)^{k-2}} = \lambda^{n-k} \left[(\lambda + n_1)(\lambda + n_k) - n_1(k-1)(\lambda + n_k) - n_k(\lambda + n_1) \right] .$$

We see that $\psi(-n_1)\,\psi(-n_k) < 0$. This implies that G has an eigenvalue other than $-n_1$ which is a contradiction. Therefore, G must be regular. □

Theorem 2.2. *Let p, η, and q be, respectively, the number of positive, zero, and negative adjacency eigenvalues of G. Then*

$$\mathrm{EE}(G) \geq \eta + p\, e^{\mathrm{E}(G)/(2p)} + q\, e^{-\mathrm{E}(G)/(2q)} \ . \tag{1}$$

Equality holds if and only if G is either

(i) *a union of complete bipartite graphs $K_{a_1,b_1} \cup \cdots \cup K_{a_p,b_p}$ with (possibly) some isolated vertices, such that $a_1\, b_1 = a_2 b_2 = \cdots = a_p\, b_p$, or*

(ii) *a union of copies of $K_{k\times t}$, for some fixed positive integers k, t, with (possibly) some isolated vertices.*

Proof. Let $\lambda_1, \ldots, \lambda_p$ be the positive, and $\lambda_{n-q+1}, \ldots, \lambda_n$ be the negative eigenvalues of G. As the sum of eigenvalues of a graph is zero, one has $\mathrm{E}(G) = 2\sum_{i=1}^{p} \lambda_i = -2\sum_{i=n-q+1}^{n} \lambda_i$. By the arithmetic–geometric mean inequality, we have $\sum_{i=1}^{p} e^{\lambda_i} \geq p\, e^{(\lambda_1+\cdots+\lambda_p)/p} = p\, e^{\mathrm{E}(G)/(2p)}$. Similarly, $\sum_{i=n-q+1}^{n} e^{\lambda_i} \geq q\, e^{-\mathrm{E}(G)/(2q)}$. For the zero eigenvalues, we also have $\sum_{i=p+1}^{n-q} e^{\lambda_i} = \eta$. Now, (1) is obtained by combining these inequalities.

Equality is attained in (1) if and only if $\lambda_1 = \cdots = \lambda_p = r$ and $\lambda_{n-q+1} = \cdots = \lambda_n = s$, for some r, s. By the Perron–Frobenius theorem (see [4], Theorem 8.8.1), G has exactly p non-trivial components. Moreover, each non-trivial component of G has exactly one positive eigenvalue, and thus they must be complete multipartite graphs.

If $p = q$, then $r = -s$, and so each non-trivial component of G is a complete multipartite graph $K_{a,b}$, where $r = \sqrt{ab}$. Therefore, G must be a graph of the form described in part (i) of Theorem 2.2.

If $p \neq q$, then, by the Perron–Frobenius theorem, $r > -s$. Therefore each non-trivial component of G has a spectrum of the form $\{[r]^1, [0]^t, [s]^k\}$, such that $t \geq 0$ and $k \geq 2$. Hence, by Lemma 2.1, each non-trivial component of G is a regular complete multipartite graph. As regular complete multipartite graphs are uniquely determined by their spectrum (see [1, p. 163]), all non-trivial components of G are the same. This implies that G is a graph of the form described in part (ii) of Theorem 2.2.

By this the proof of Theorem 2.2 is complete. □

The following corollary follows immediately from the above theorem. We recall that for bipartite graphs, one has $p = q$.

Corollary 2.3. *If G is a bipartite graph, then* $\mathrm{EE}(G) \geq \eta + r\cosh\left(\frac{\mathrm{E}(G)}{r}\right)$, *where r is rank of the adjacency matrix of G. Equality holds if and only if G is a union of complete bipartite graphs $K_{a_1,b_1} \cup \cdots \cup K_{a_p,b_p}$ with (possibly) some isolated vertices, such that $a_1\, b_1 = a_2 b_2 = \cdots = a_p\, b_p$.*

3 Conclusion

In conclusion, linear algebraic approaches have proven to be effective in estimating the Estrada index bounds. By utilizing tools such as matrix eigenvalues and the spectrom of graphs, researchers have developed various techniques for computing the Estrada index and its bounds. These methods have applications in diverse fields such as network analysis, computational biology, and physics, where the Estrada index provides valuable insights into the properties of complex systems.

References

[1] D. M. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs – Theory and Application*, Barth, Heidelberg, 1995.

[2] D. Cvetković, P. Rowlinson, S. K. Simić, *An Introduction to the Theory of Graph Spectra*, Cambridge Univ. Press, Cambridge, 2009.

[3] E. Estrada, Characterization of 3D molecular structure, *Chem. Phys. Lett.* **319** (2000) 713–718.4

[4] C. Godsil, G. Royle, *Algebraic Graph Theory*, Springer–Verlag, New York, 2001.

[5] I. Gutman, The energy of a graph: Old and new results, in: A. Betten, A. Kohnert, R. Laue, A. Wassermann (Eds.), *Algebraic Combinatorics and Applications*, Springer–Verlag, Berlin, 2001, pp. 196–211.

[6] I. Gutman, X. Li, J. Zhang, Graph energy, in: M. Dehmer, F. Emmert–Streib (Eds.), *Analysis of Complex Networks. From Biology to Linguistics*, Wiley–VCH, Weinheim, 2009, pp. 145–174.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On pliable source index coding

Javad B. Ebrahimi[1,2] and Hossein Mahdavipour[1],*

[1]Department of Mathematics, Sharif University of Technology, Tehran, Iran
[2]Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Abstract

Index coding problem (IC), introduced in [1], is a canonical problem in the field of communication theory. It is connected to many problems in the theory of communication and combinatorics. A relaxed version of IC, called pliable index coding(PIC) is introduced in [2]. In this work, we introduce the source code counterpart of PIC which we call it "pliable source index coding problem (PSCI)" problem. We use linear algebraic tools and techniques to show that for the case of linear PIC and linear PSCI problems, there exists a strong linear algebraic duality.

Keywords: Pliable index coding, pliable source index coding, network coding
Mathematics Subject Classification [2010]: 94A29, 68P30

1 Introduction

The main practical motivation for IC is coming from satellite communication. However, it founds strong connections to many other problems in computer science, combinatorics, the theory of communication, and more. Besides numerous applications in data transmission and data storage, it has interesting theoretical aspects. For example its relation to network coding has been proved in [3].

A variant of this problem, known as pliable index coding(PIC) has been introduced by Brahma and Fragouli in [2]. It is a relaxed version of the original problem but later on, PIC receives a lot of attention among researchers. In this work, we introduce the source index coding, a new version of this problem. We prove that for the case of linear algebraic pliable code, they are equivalent meaning in the sense that for any linear pliable index code, there exists a corresponding pliable source index code, which is its orthogonal complement in the scene of linear algebra, and vice versa.

More precisely, we show that if we look at a PIC as vectors in a W, a vector space over a finite field, we can find a linear pliable source index coding(LPSIC) from the complement of original vectors and vice versa.

An informal description of the index coding problem is as follows. Imagine a satellite has m different messages and there are n different persons listening to the satellite, each of them needs a specific message while knowing some of the other messages as side information beforehand. We model this with a bipartite graph G called side information graph with

*Speaker. Email address: hosseinmp76@hotmail.com

two parts. Each vertex in part B represents a message and each vertex in part C represents a person. We use a directed edge from c_i to b_j if c_i knows b_j as side information. We show the side information of $c_i(N(S_i)$ in graph) with S_i. The satellite uses an encoding function En to transmit some information (which all clients will receive). Then, each client uses its specific decoding function $\overrightarrow{f}_i$ to find its desired message from its side information and received data.

For the PIC problem, the clients are pliable meaning they do not necessarily seek a specific message. Instead, they want to know a message not in their side information. That means, for a given side information graph G, the PIC solution includes an index vector I of length n of indices from the set $[m]$ such that its i-th coordinate is the index of the message where the i-th person will learn. Therefor $I = (I_1, \ldots, I_n) : I_i \in [m], I_i \notin S_i$. Imagine that a letter x_j from $\mathbb{F}$ is assigned to each data node b_j. The encoding function En takes $X = (x_1, \ldots, x_m)$ as an input and outputs string $En(X) \in \mathbb{F}^k$. Now, each client node c_j can recover the value x_{I_j}.

We can imagine this setting in a real-life situation: n persons visit a news website to know something they don't know previously. That's why we suppose $deg(c_i) \leq m$ so for every client there is at least a data that is unknown.

We now try to formally define the PSIC problem. Let G be a bipartite graph with parts $C = \{c_1, c_2, \ldots, c_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$. Each c_i is called a client node and each b_j is called a data (message, bit) node. Suppose that the degree of each user node is smaller than m. Let $\mathbb{F}$ be a finite field. A pliable index code with side information (PIC) for the underlying graph G over the alphabet set $\mathbb{F}$ is a tuple $(\Sigma, E, I, \overrightarrow{f})$ such that:

1. $En : \mathbb{F}^m \to \Sigma$ (encoding function)
2. $I = (I_1, I_2, \ldots, I_n)$ such that each I_j is an element of the set $[m]$. Also, for each j, $I_j \notin N(c_j)$. (index of guessing data for each client)
3. $\overrightarrow{f} = (\overrightarrow{f}_1, \ldots, \overrightarrow{f}_n)$ such that $\overrightarrow{f}_i : \mathbb{F}^{|S_i|} \times \Sigma \to \mathbb{F}$. (decoding vector)
4. For any choice of $X = (x_1, x_2, \ldots, x_m) \in \mathbb{F}^m$ (message tuple) and for each $j \in [n]$, we have: $\overrightarrow{f}_j\big(X[S_j], En(X)\big) = X[I_j]$.

The range of encoding function in PIC is not necessarily $\mathbb{F}$. and it can be any other set so it isn't necessarily linear too. But in this paper, we focus on linear encoding functions on finite fields. i.e. Σ is $\mathbb{F}^k$ for some k(k is the code length). therefor we show PIC with $(En, I, \overrightarrow{f})$

One of the applications of IC is in designing distributed systems. Different variants appear in many problems. For example, [4] uses PIC for a better data shuffling scheme.

2 Preliminaries

2.1 Linear algebra

For a matrix M the $span(M)$ is the row span of the matrix M.
Let V be a subspace of $\mathbb{F}^l$. The co-sets of V are the shifts of V by a vector $b \in \mathbb{F}^l$. The co-sets of a subspace are also called Affine subspaces. In mathematical notation, if W is a co-set V then $\forall a, b \in W : a - b \in V$. For a matrix M, we define $ker(M)$ as the kernel(null space) of M. For a vector space W, define $W^\perp$ as the set of all vectors $v \in \mathbb{F}^n$ such that for all $w \in W$ we have $v^T.w = 0$. $W^\perp$ is also called the orthogonal complement of W with respect to the canonical inner product. It is well-known that for every finite dimensional

vector space W we have: $dim(W) + dim(W^{\perp}) = n$.
For every one-to-one linear function h, the inverse function h^{-1} is a linear function too.

2.2 Graph Theory:

A graph G is a pair (V, E) of a finite set V, called the vertex set and a set E of 2-tuple elements of V called the edge set. For a vertex $v \in V$, the set of all other vertices u such that $\{u, v\} \in E$ are called the neighbors of v. We denote the set of the neighbors of v by $N(v)$. A graph G is called bipartite with parts A and B when V is the disjoint union of A and B and each edge has exactly one vertex from each part. For a given bipartite graph G with parts B, C and vertices $B = \{B_1, \ldots, b_m\}, C = \{c_1, \ldots, c_n\}$ we call B the "data" part and C the "client" part. We assign a variable x_i to each b_i vertex. Each x_i is supposed to take a value from a finite field $\mathbb{F}$. For a vector $\mathcal{V} = (v_1, \ldots, v_m)$ and a set $S = \{s_1, \ldots, s_k\} \subseteq [m]$, we define $\mathcal{V}[S]$ as a k-coordinates vector that consists of k coordinates of V with indices from S, that is $\mathcal{V}[S] = (v_{s_1}, \ldots, v_{s_k})$. Also we define $\mathcal{V}[i] := \mathcal{V}[\{i\}]$

3 Main results

In this section, we first formally define the problem of pliable source index coding. Then we state the main theorem of this paper regarding the connection between the linear pliable index code problem (PIC) and linear pliable source index code (PSID). Note that all functions are linear in this paper unless stated otherwise.

Before we start to define the PSID, we overview the PIC problem both in arbitrary and linear form. Roughly speaking, the pliable index code problem says that if a side information graph is given, (i.e. each client vertex knows certain values of X_j's) then the sender transmits further information (i.e. an element from some finite set Σ) such that each client can retrieve some X_j from the information he already had (the side information) and the new piece of information he received from the transmitter. The goal is to minimize the size of Σ such that this task can be accomplished.

As an example, when each client c_i knows all the values of X_j's except X_i, then the transmitter can transmit $X_1 + X_2 + \ldots + X_n$. Now, each client can subtract its side information from the transmitted message to recover X_i which he does not have previously.

An alternative way to look at this solution is that if $\Sigma = \{z_1, z_2, \ldots, z_k\}$, then we can partition all the vectors in $\mathbb{F}^n$ into k parts; namely, those whose encoding is z_i are grouped in one part. Now, when the transmitter picks some vector $X \in \mathbb{F}^n$, it tells all the clients that the chosen x belongs to which part. It is then the task of the clients to look at their side information and to the group which X belongs, and then learn a new entry of X that they did not know before.

In the pliable source index code problem, there is no transmitter to tell the clients which part X is chosen from. Instead, there exists a set of vectors of $\mathbb{F}^m$ called code-book (also called a table) such that X belongs to it. Then, given the side information and the code-book, each client must learn a coordinate of X beyond the coordinates he already knew. For instance, in the side information graph in the above example, suppose that the code-book consists of all the vectors $(X_1, X_2, \ldots, X_n)$ such that $X_1 + X_2, \ldots + X_n = 0$. Since every c_i knows all the X_j's except X_i, he can retrieve the remaining coordinate X_i.

Notice that in the PIC problem, the goal is to minimize the size of the transmitted messages while in the PSIC problem, the goal is to maximize the size of a code-book(feasible table). We now formally define a feasible table.

Definition 3.1 (**Feasible table**)**.** For a given alphabet $\mathbb{F}$ and the side information graph G, a feasible table T is a set with elements from $\mathbb{F}^m$ such that for every client c_i if its side information is the same in some different elements there is at least an index i not in his side information that has the same value in all those elements. i.e.:

$$\forall i : \exists j \notin S_i : \forall X, X' : X[S_i] = X'[S_i] \Rightarrow X[j] = X'[j]$$

We use the term table because we show elements of this set in a table. In this manner, each column will show one of the variables X_i and each row is a possible message tuple.

The pliable source index coding problem, PSIC, is to find the largest possible feasible table for a given G.

Remark 3.2. We can always partition $\mathbb{F}^n$ into feasible tables, namely taking every $X \in \mathbb{F}^n$ as a feasible table with one row.

Lemma 3.3. *Let G be a graph. If $\mathbb{F}^m$ is partitioned into disjoint feasible tables, then the encoding function which takes the vector $X \in \mathbb{F}^m$ and encodes it into the index of the feasible table which X belongs to, is a valid solution for the PIC problem for the graph G.*

Proof. Consider client c_i. Client c_i knows $X[S_i]$. By the definition of the feasible table, there exists an index j outside S_i which c_i can learn from the table whose index has been transmitted. □

Corollary 3.4. *If we partition all possible message tuples (i.e. $\mathbb{F}^t$) in l different feasible tables then we can just send $\lceil \log(l) \rceil$ bits of data to users to find the index of the selected table and find their desired message.*

Note 3.5. In case of linear PIC, the encoding function En can be described by a matrix. $\overrightarrow{f}$ are linear functions. Also, if the server sends k messages $Y_i = a_{i,1}X_1 + a_{i,2}X_2 \ldots + a_{i,n}X_n$ then

$$En = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \cdots & a_{k,n} \end{pmatrix}$$

Definition 3.6 (**Linear feasible table**)**.** Let $\mathbb{F}$ be a finite field. A feasible table T over $\mathbb{F}$ is called "linear" if the rows of T form a vector space over $\mathbb{F}$.

The linear pliable source index coding problem, LPSIC, is to find the largest possible linear feasible table for a given graph G.

Remark 3.7. There is always at least one linear feasible table, $\{\overrightarrow{0}\}$. In the linear case, we can partition $\mathbb{F}^n$ with a linear feasible table and its co-sets, simply by considering the linear feasible table and all its shifts. This way, we cover all $\mathbb{F}^n$.

Example 3.8. Consider G_1 and G_2 in Figure 1. For G_1 we can partition message tuples in 4 different sets(tables). In this example, they are all feasible tables. For G_2 we have a linear feasible table and its only co-set which together cover $\mathbb{F}_2^3$:

Lemma 3.9. *In a feasible table, we have: $\forall i : \exists j \in [m], j \notin S_i : \forall X_1, \ldots, X_k \in T : X_1[S_i] = X_2[c_i] = \ldots = X_k[S_i] \Rightarrow X_1[j] = X_2[j] = \ldots = X_k[j]$*

Proof. The lemma follows directly from the definition of feasible table. □

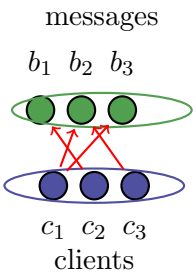


Figure 1: G_1

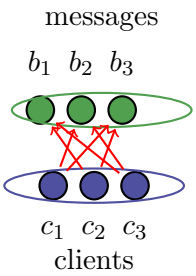


Figure 2: G_2

0000	001	010	011
101	110	111	100

000	001
011	010
110	100
101	111

Table 1: four feasible table for G_1 and a linear feasible table and its only co-set for G_2. Both of these partition 2^3

Lemma 3.10. *Let G be a bipartite graph on the parts C, B. Let $(En, I, \overrightarrow{f})$ be a linear pliable index code for G, then there exists a LPSIC $(W, J, \overrightarrow{g})$ such that $W = ker(En)$*

Proof. First, we overview the idea of the proof. For every linear pliable index coding problem, consider the linear combinations that the server broadcasts to the users. The coefficients will form the kernel of a linear feasible table.

The LPSIC $(W, J, \overrightarrow{g})$ is defined as follows: $W = ker(En)$, $\overrightarrow{g}_i = \overrightarrow{f}_i(\overrightarrow{0}, X[S_i])$, $J = I$ Consider the $span(W)$. If the message tuple V is from this vector space then, the transmitted message would be all zero since $En \times V = \overrightarrow{0}$. Knowing V is from $span(W)$ means the transmitted encoding is $\overrightarrow{0}$ and since the i-th client can guess $X[I_i]$ in the PIC then it can use the $\overrightarrow{f}_i$ with $\overrightarrow{0}$ as the transmitted code as its decoding function. So $span(W)$ forms a linear feasible table and therefore $span(W)$ is a linear feasible table and $(W, J, \overrightarrow{g})$ is a LPSIC solution. □

Lemma 3.11. *Let $(W, J, \overrightarrow{g})$ be a LPSIC for G. Then there exists a LPIC $(En, I, \overrightarrow{f})$ such that $span(En) = W^{\perp}$*

Proof. Let $r = dim(W)$ and $\{V_1, V_2, \ldots, V_{n-r}\}$ be a basis for $n-r$ dimensional space $W^{\perp}$. Define the encoding matrix En as an $(n-r) \times n$ matrix whose rows are $V_1, V_2, \ldots, V_{n-r}$. Clearly $span(En) = W^{\perp}$. Also, let $J = I$ and $\overrightarrow{f}_i(\overrightarrow{R}, En(X[S_i]) = f(En(X[S_i] - R[S_i])) + R[I_i]$ where $\exists! V \in W : V + R = X$. Now, we should prove if i-th user can retrieve the $X[I_i]$ in every message tuples. In another words there aren't different message tuples with different $X[I_i]$ that with same side information for i-th client and same transmitted code(i.e. $En(X)$). Suppose otherwise. Then there exists X_1, X_2, i such that $En(X_1) = En(X_2)$ and $X_1[S_i] = X_2[S_i]$ but $X_1[I_i] \neq X_2[I_i]$. Let $W_1, \ldots, W_t$ be the co-sets of W Then there exists unique vectors that $X_1 = V_1 + V'_1, X_1 = V_2 + V'_2$, where $V_1, V_2 \in W$ and $V'_1 \in W_p, V'_2 \in W_q$. By the definition of the feasible table, we have $En(V_1) = En(V_2) = 0$. We should prove that client i can always recognize the I_i message. $En(X_1) = En(X_2)$ means both X_1, X_2 come from same co-set, i.e. $V'_1 = V'_2$. We claim every for two different

co-sets W_1, W_2 we have: $\forall U_1, U_2 : U_1 \in W_1, U_2 \in W_2 : En(U_1) \neq En(U_2)$. This means the clients can distinguish the message tuple co-set from all other co-sets. Let W_j be the co-set that $X_1, X_2 \in W_j$ then: $R = U - U' : U, U' \in W_j$. Now, $X_1 - R, X_2 - R$ are both in W meaning the client can differentiate between them. But this is a contradiction because after adding $R[I_i]$ to the $X[I_i]$ message, the i-th user can differentiate between X_1, X_2.
It only remains to prove the claim. Suppose otherwise. Since En is linear, we have: $En(U_1) = En(U_2) \Rightarrow En(U_1 - U_2) = 0 \Rightarrow U_1 - U_2 \in W \Rightarrow \exists i : U_1, U_2 \in W_i$. But we assumed $W_1 \neq W_2$. □

Theorem 3.12. *For any side information graph G, LPSIC $(W, J, \overrightarrow{g})$ is linear algebraic dual of LPIC $(En, I, \overrightarrow{f})$ in the sense that:*

$$(W, J, \overrightarrow{g}) = \begin{cases} W = ker(En) \\ I = J \\ \overrightarrow{g}_i(En(X[S_i])) = \overrightarrow{f}_i(\overrightarrow{0}, En(X[S_i]) \end{cases}$$

$$(En, I, \overrightarrow{f}) = \begin{cases} span(En) = W^{\perp} \\ J = I \\ \overrightarrow{f}_i(\overrightarrow{R}, En(X[S_i]) = g(En(X[S_i] - R[S_i])) + R[I_i] : \exists! V \in W : V + R = X \end{cases}$$

and $dim(W) = n - dim(En)$ i.e. if $En : \mathbb{F}^n \to \mathbb{F}^k$ and code-book W has l message tuples then $\log(l) + k = n$

Proof. This is a direct consequence of the last two lemmas. We just need to show the equality of dimensions. This comes from the fact that: $dim(W) + dim(W^{\perp}) = n$. □

4 Conclusion

Inspired by the work of [5] on the dual index code problem, we defined the pliable source index code problem and showed that in the linear case, there exists a strong duality between the minimum size transmission for PIC on the graph G and the maximum feasible table for PSIC for the same graph. This, in particular, shows that the minimum transmission plus the dimension of maximum linear PSIC is equal to n.

References

[1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.

[2] S. Brahma and C. Fragouli, "Pliable index coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 11, pp. 6192–6203, 2015.

[3] M. Effros, S. El Rouayheb, and M. Langberg, "An equivalence between network coding and index coding," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2478–2487, 2015.

[4] L. Song, C. Fragouli, and T. Zhao, "A pliable index coding approach to data shuffling," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1333–1353, 2020.

[5] A. Mazumdar, "On a duality between recoverable distributed storage and index coding," in *2014 IEEE International Symposium on Information Theory*, pp. 1977–1981, 2014.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The Distribution of Product Random Stochastic Matrices: By Dirichlet Distribution

Hazhir Homei* and Parisa Rafiee

Department of Mathematics, Statistics, and Computer Science, University of Tabriz, Tabriz, Iran

Abstract

Following the study of the concept of averaging dynamics, the well-known concept of multiplication of random matrices has been frequently investigated in the form of specific assumptions. In this paper, we argument the distributional properties of product of random stochastic matrices by using the Dirichlet distribution. In the following, we will answer the raised questions posted at the conclusion of two separate articles related to real lifetime and solving some differential equations and generalize some of the results obtained.

Keywords: Averaging Dynamics, Real Lifetime, Random Stochastic Matrices, Deferential Equation

Mathematics Subject Classification [2010]: MSC 65Cxx, MSC 62E15

1 Introduction

The product of random stochastic matrices is one of the concepts that has attracted the attention of many mathematicians [6]. The behavior of this concept in science and engineering has been investigated based on various assumptions. The study and application of this concept are closely related to the study of averaging dynamics.
Let us note that some of the examples of applications, similar to some of the examples and applications, presented by Touri and Nedic [6].
In the present paper, considering the independent random matrices each of which has independent rows and are identically distributed with Dirichlet distribution, we have investigated some distributional and statistical properties of the product of random matrices. For this purpose, we have addressed the mixture random variables and followed their traces in the applied fields.

Random Convex Combination: A stochastic linear combination

$$\hat{C}_1.Z_1 + \hat{C}_2.Z_2 + \ldots + \hat{C}_m.Z_m \tag{1}$$

of random variables $Z_1, \ldots, Z_m$ where $\hat{C}_i, 1 \leq i \leq m$, are random variables such that

(i) $\hat{C}_i \geq 0,\ 1 \leq i \leq m,$ and
(ii) $\sum_{i=1}^{m} \hat{C}_i = 1, a.s.,$

*Speaker. Email address: homei@tabrizu.ac.ir

is called a random convex combination of the random variables $Z_1, \ldots, Z_m$ (for more details see Homei and Nadaraja [5]).
Of course, another form of real lifetime is provided by Homei [1], which is not far from the statistic defined. Let $Z_i, i = 1, \ldots, n$, be the lifetime measured in a lab and $0 \leq C_i \leq 1$ be the random effect of the environment on it, so $C_i Z_i \leq Z_i$ and thus $\sum_{i=1}^{n} C_i Z_i$ is the average lifetime in the environment see Homei [3], Homei and Nadaraja [5]. In parentheses, if Y_i is the real lifetime in the ith area, $C_i = \frac{Y_i}{\sum_{i=1}^{n} Y_i}$ is the random effect ratio in the ith area. Therefore, it is clear that a good choice for distribution $\mathbf{C} = \langle C_1, \ldots, C_n \rangle$ can be Dirichlet distribution. It is important that the product of random stochastic matrices connect us directly to stochastic linear combination.

2 Product Moments Of Random Stochastic Matrices

The concept of the product of random stochastic matrices are motivated us to discuss on the distributional properties of random convex combination. These properties include product moments, mean and variance of components. Throughout the paper, $\langle W_1, \cdots, W_r \rangle$ is called random coefficient vector of environmental effect in $r-$position. In the introduction recall that the rows are independent and have Dirichlet distribution in random stochastic matrices.

Theorem 2.1. *Suppose that the independent random vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_r$ have identical distributions with mean μ and variance $\boldsymbol{S}$ and that the random vector $W = \langle W_1, \cdots, W_r \rangle$ is independent from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ such that $\sum_{j=1}^{r} W_j = 1$, a.s. Then the mean and variance of $\boldsymbol{Z} = \sum_{j=1}^{r} W_j \boldsymbol{X}_j$ are*

$$E(\boldsymbol{Z}) = \mu \quad and \quad Var(\boldsymbol{Z}) = \sum_{j=1}^{r} EW_j^2 \boldsymbol{S},$$

where $\boldsymbol{S}$ is variance-covariance matrix.

The following theorem results in product moments of random convex combination when we consider the random vectors by Dirichlet distribution.

Theorem 2.2. *Suppose that the independent random vectors $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_r$ have respectively, Dirichlet$(n_{11}, \cdots, n_{1k}), \cdots,$ Dirichlet$(n_{r1}, \cdots, n_{rk})$ distributions and that the random vector $W = \langle W_1, \cdots, W_r \rangle$ is independent from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ and has Dirichlet$(\alpha_1, \cdots, \alpha_r)$ distribution. Then the product moments in $(s_1, \ldots, s_k)$ of $\boldsymbol{Z} = \sum_{j=1}^{r} W_j \boldsymbol{X}_j$ are*

$$E(L_1^{s_1} L_2^{s_2} \ldots L_k^{s_k}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \sum_{h_1} \cdots \sum_{h_k} \left(\prod_{j=1}^{k} \binom{s_j}{h_{1j} \ldots h_{rj}} \times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)} \right.$$

$$\left. \frac{\Gamma(n_{i.})}{\Gamma(n_{i.} + h_{i.})} \prod_{i=1}^{r} \prod_{j=1}^{k} \frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})} \right),$$

where L_j's are components of vector Z, $\sum_{i=1}^{r} h_{i.} = h$ and $\sum_{i=1}^{r} \alpha_i = \alpha$.

Proof. We find the general moments $(s_1, s_2, \cdots, s_k)$ of Z as follow

$$E(L_1^{s_1}L_2^{s_2}\dots L_k^{s_k}) = E\Big(\prod_{j=1}^{k}(\sum_{i=1}^{r} W_i\mathbf{X}_{ij})^{s_j}\Big)$$

$$= E\Big(\prod_{j=1}^{k}(\sum_{h_j}\binom{s_j}{h_{1j}, h_{2j}, \cdots, h_{nj}} \times \prod_{i=1}^{r}(W_iX_{ij})^{h_{ij}})\Big),$$

where $\sum h_{ij}$ denotes summation over all nonnegative integers $h_j = (h_{1j}, h_{2j}, \cdots, h_{rj})$ subject to

$$\sum_{i=1}^{r} h_{ij} = s_j, \quad (j = 1, 2, \cdots, k).$$

Equation above can be rearranged as

$$= E\Big(\sum_{h_1}\sum_{h_2}\cdots\sum_{h_k}(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \cdots, h_{rj}}\prod_{j=1}^{k}\prod_{i=1}^{r}(W_i\mathbf{X}_{ij})^{h_{ij}})\Big)$$

$$= E\Big(\sum_{h_1}\sum_{h_2}\cdots\sum_{h_k}(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \cdots, h_{rj}}\prod_{i=1}^{r}W_i^{h_{i.}}\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}})\Big),$$

where $h_{i.} = \sum_{j=1}^{k} h_{ij}$ and we have,

$$= \sum_{h_1}\cdots\sum_{h_k}(\prod_{j=1}^{k}\binom{s_j}{h_{1j}, h_{2j}, \cdots, h_{rj}}E(\prod_{i=1}^{r}W_i^{h_{i.}})E(\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}})), \tag{2}$$

now we find two expectations in equation (2):

$$E(\prod_{i=1}^{r}W_i^{h_{i.}}) = \frac{\Gamma(\sum_{i=1}^{r}\alpha_i)}{\Gamma(\sum_{i=1}^{r}(\alpha_i + h_{i.}))} \times \prod_{i=1}^{r}\frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}.$$

By using the Dirichlet distribution, we have

$$E(\prod_{i=1}^{r}W_i^{h_{i.}}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \times \prod_{i=1}^{r}\frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}. \tag{3}$$

Also, we have

$$\begin{aligned} E(\prod_{j=1}^{k}\prod_{i=1}^{r}X_{ij}^{h_{ij}}) &= \prod_{i=1}^{r}E(\prod_{j=1}^{k}X_{ij}^{h_{ij}}) \\ &= \prod_{i=1}^{r}\Big(\frac{\Gamma(\sum_{j=1}^{k}n_{ij})}{\Gamma(\sum_{j=1}^{k}(n_{ij} + h_{ij})} \times \prod_{j=1}^{k}\frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})}\Big), \end{aligned}$$

now we have $\sum_{j=1}^{k} n_{ij} = n_{i.}$ and $\sum_{j=1}^{k} h_{ij} = h_{i.}$

$$= \prod_{i=1}^{r}\Big(\frac{\Gamma(n_{i.})}{\Gamma(n_{i.} + h_{i.})} \times \prod_{j=1}^{k}\frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})}\Big), \tag{4}$$

and by using the Dirichlet distribution, we have

$$E(\prod_{j=1}^{k} X_{ij}^{h_{ij}}) = \frac{\Gamma(\sum_{j=1}^{k} \alpha_j^{(i)})}{\Gamma(\sum_{j=1}^{k} \alpha_j^{(i)} + h_{i.})} \prod_{j=1}^{k} \frac{\Gamma(\alpha_j^{(i)} + h_{ij})}{\Gamma(\alpha_j^{(i)})}.$$

So, by using 3 and 4 in 2

$$= \sum_{h_1} \cdots \sum_{h_k} \Big(\prod_{j=1}^{k} \binom{s_j}{h_{1j}, h_{2j}, \cdots, h_{rj}} \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)}$$

$$\times \prod_{i=1}^{r} \frac{\Gamma(n_{i.})}{\Gamma(n_{i.} + h_{i.})} \prod_{j=1}^{k} \frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{i)})} \Big)$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha + h)} \sum_{h_1} \cdots \sum_{h_k} \prod_{j=1}^{k} \binom{s_j}{h_{1j} \dots h_{rj}}$$

$$\times \prod_{i=1}^{r} \frac{\Gamma(\alpha_i + h_{i.})}{\Gamma(\alpha_i)} \frac{\Gamma(n_{i.})}{\Gamma(n_{i.} + h_{i})} \prod_{i=1}^{r} \prod_{j=1}^{k} \frac{\Gamma(n_{ij} + h_{ij})}{\Gamma(n_{ij})}.$$

Therefore the proof of the product moments on $(s_1, \dots, s_k)$ is complete.

□

2.1 Some Characterizations

In this section, some characterizations of random stochastic linear combinations (real lifetime, random convex combination) in Dirichlet random vectors are introduced.

Theorem 2.3. *Suppose that* $\mathbf{U}$ *and* V *are independent (absolutely continuous) nonnegative random variable, respectively, such that* $\mathbf{U}$ *has bounded support and* $\mathbf{Z} = \mathbf{U}V$. *Then for arbitrary positive* $\alpha_i; i = 1, \dots, k$ *and any two of the following three conditions imply the third.*
i) $\mathbf{Z} \sim < Gamma_1(\alpha_1, \frac{1}{\mu}), \dots, Gamma_k(\alpha_k, \frac{1}{\mu}) >$ *where* $Gamma(\alpha_i, \frac{1}{\mu})$ *are independent;*
ii) $\mathbf{U} \sim Dirichlet(\alpha_1, \dots, \alpha_k)$;
iii) $V \sim Gamma(\alpha^+, \frac{1}{\mu}), \quad \alpha^+ = \sum \alpha_i$.

Theorem 2.4. *Suppose that the independent random variables* $\boldsymbol{X}_1, \dots, \boldsymbol{X}_r$ *have* $Dirichlet(\frac{1}{2} + \alpha_1, \dots, \frac{1}{2} + \alpha_1), \dots, Dirichlet(\frac{1}{2} + \alpha_r, \dots, \frac{1}{2} + \alpha_r)$ *distributions and that the random vector* $\boldsymbol{W} = \langle W_1, \dots, W_r \rangle$ *has* $Dirichlet(\alpha_1, \dots, \alpha_r)$ *distribution. Then* $\boldsymbol{Z} = \sum_{j=1}^{r} W_j \boldsymbol{X}_j$ *has* $Dirichlet(\frac{1}{2} + \sum_{i=1}^{r} \alpha_i, \dots, \frac{1}{2} + \sum_{i=1}^{r} \alpha_i)$ *distribution.*

Proof Let $Y_j \ (j = 1, \cdots, n)$ be independent random variables independent from $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ that have the distribution $Gamma(\alpha_j, \frac{1}{\mu})$, respectively. It can be seen, by some classic ways (e.g. $E(e^{t'T}) = [\Psi(t)]^{\sum_j \alpha_j}$ from Homei([1]), Table 2), that the distribution of $\mathbf{T} = \sum_j \mathbf{T}_j = \sum_j Y_j \mathbf{X}_j$ is the same distribution of $\mathbf{T}_j$ with the parameter $(\sum_j \alpha_j, \cdots, \sum_j \alpha_j)$. We can also write $\mathbf{T} \overset{\mathrm{d}}{=} Y\mathbf{X}$ in which Y has the gamma distribution with the parameter $(\sum_{i=1}^{k} \alpha_j, \frac{1}{\mu})$ and $\mathbf{T}$ has the $Dirichlet(\sum \alpha_i, \cdots, \sum \alpha_i)$ distribution, and Y and $\mathbf{X}$ are independent from each other. By using theorems the proof is completed.

2.2 A Generalization

In this subsection we want to review some of the works by others and generalize to multivariate case, see Homei and Nadarajah [5].

Theorem 2.5. *Suppose* $\boldsymbol{X}$ *and* Y *have Drichlet*$(\alpha_1, \ldots, \alpha_r)$ *and Gamma*(α, β) *distribution and the distribution of* $\boldsymbol{Z} = \boldsymbol{X}Y$ *can be expressed as:*

$$f(z_1, \ldots, z_r) = \frac{\beta^{-\alpha}\Gamma(\sum_{i=1}^{r}\alpha_i)}{\Gamma(\alpha)\prod_{i=1}^{r}\Gamma(\alpha_i)}(\sum_{i=1}^{r} z_i)^{\alpha-\sum_{i=1}^{r}\alpha_i} e^{-\frac{\sum_{i=1}^{r} z_i}{\beta}} \prod_{i=1}^{r} z_i^{\alpha_i-1}. \tag{5}$$

3 Application

3.1 In Statistics

Some results from the last few decades show that some researchers are interested in obtaining the distribution of $\mathbf{Z}$, the random convex combination, and in more detail it can be seen that they are still interested in finding the distribution of $\mathbf{X}_i$ while assuming the distribution of $\mathbf{Z}$ is known, for example see Theorem 2.3 in [4], Theorem 1 in [3] and also references therein.
By considering the main theorem in Homei [1] and equation 6, one can obtain the distribution of $\mathbf{Z}$ or $\mathbf{X}_i$'s by placing c-characteristic and solving a differential equation, of course if can solve it. We can easily obtain Theorem 2.3 in [4] and Theorem 1 in [3] by equation 6, and also with this equation we can obtain the main result of Homei [1], so this equation can play a fundamental role in obtaining the distribution of $\mathbf{Z}$ provided that an according differential equation can be solved.

3.2 In Mathematics

The main theorem in Homei [1] may be useful in solving some differential equations and some interesting mathematical facts which may be very difficult to solve. In this subsection, we change our view of this theorem.
First, we assume equation 6 then we recognize the solutions of the equation from the main theorem of Homei [1], so we were able to use the results of the main theorem to obtain the answer to equation 6. See some applications in a very special case in Homei [1].

The multivariate c-characteristic function has appeared to be an appropriate tool for investigating how the random convex combination is related to the distributions of $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Definition 3.1. If $u = (U_1, \cdots, U_k)$ is a random vector, its multivariate c-characteristic function is defined as

$$g(t_1, \cdots, t_k; u, c) = E\left\{(1 - it_1u_1 - \cdots - it_ku_k)^{-c}\right\},$$

where c is a positive real number.

Homei([1], Lemma 2.2) show that there is a one-to-one correspondence between random vector and its multivariate c-characteristic function. The following is the main theorem of this subsection.

Theorem 3.2. *For independent and continuous random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, and F denote the distribution of Z, the following equality holds:*

$$g(t_1, \ldots, t_k; Z, c) = \prod_{j=1}^{n} \prod_{r=1}^{k} g(t_1, \ldots, t_k; X_{jr}, c_j), \tag{6}$$

where $g(\cdot)$ is c-characteristic function.

4 Conclusion

For product random stochastic matrices, we studied the distribution of the random convex combination Z when the $\mathbf{X}_i$'s have Dirichlet distributions. It is interesting to note that with certain conditions the distribution of $\mathbf{Z}$ is the well-known Dirichlet distribution.

References

[1] H. Homei, The stochastic linear combination of Dirichlet distributions, *Communications in Statistics: Theory and Methods*, 50 (2021), No. 10, 2354-2359.

[2] H. Homei, Characterizations of arcsin and related distributions based on a new generalized unimodality, *Communications in Statistics: Theory and Methods* 46 (2017a), 1024-1030.

[3] H. Homei, A Novel Extension of Randomly Weighted Averages, *Statistical Papers*, 56 (2015), 933-946.

[4] H. Homei, Randomly Weighted Averages with Beta Random Proportions, *Statistics and Probability Letters*, 82 (2012), 1515-1520.

[5] H. Homei, and S. Nadarajah, On Products and Mixed Sums of Gamma and Beta Random Variables Motivated by Availability, *Methodology and Computing in Applied Probability*, 20 (2018), 799-810.

[6] B. Touri and A. Nedi c, Distributed consensus over network with noisy links, in Proceedings of the 12th International Conference on Information Fusion, 2009, pp. 146- 154.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Improved Ridge-Type Estimators in Multivariate Multiple Linear Model

Solmaz Seifollahi[1]*, Hossein Bevrani[1] and Kaniav Kamary[2]

[1]Department of Statistics, University of Tabriz, Tabriz, Iran

[2]CentraleSuplec, Laboratoire MICS, Paris, France

Abstract

This paper focus on estimating of the multivariate multiple linear models (MMLMs) when multicollinearity presents in the design matrix. Among the estimators proposed to tackle the problem of multicollinearity, the ridge estimator is popular. We propose the ridge-type estimators of the coefficient matrix when the multicollinearity exists along with the prior information about the coefficients. A simulation study is utilized to compare the relative efficiency of proposed estimators and finally, we analyze a real dataset by the proposed estimators to show the usefulness of the proposed estimators.

Keywords: Ordinary ridge estimator, restricted ridge-type estimator, James-Stein ridge-type estimator, preliminary test ridge-type estimator.

Mathematics Subject Classification [2010]: 62F10, 62J02.

1 Introduction

Multivariate multiple linear models (MMLMs) are considered as a generalized version of the multiple linear models in which several response variables are predicted from a set of independent variables or covariates. The MMLMs have recently found a wide range of applications in a variety of areas such as machine learning theory, psychology and education and other fields.

The use and interpretation of MMLMs depend on the quality of the coefficient matrix estimation. If the model parameters are poorly calibrated, the model output will not be relevant to make predictions or data assimilation. A popular estimation method is the least square estimation, which is not always efficient especially when there is multicollinearity in the design matrix. Some authors tried to improve the least square estimation of the coefficient matrix. Recently, the uncertain prior information on some of the parameters in a statistical model has been used in statistical inference. The uncertain prior information incorporated in the model through some restrictions on parameters that lead a submodel. The candidate submodel can be obtained by using variable selection techniques such as AIC or BIC. When the candidate submodel is true or the restrictions hold, analysis of such submodel leads to efficient statistical inferences than would be obtained via the

*Speaker. Email address: s.seifollahi@tabrizu.ac.ir

fullmodel. However, when considered restrictions are suspected, one may combine the restricted estimators and unrestricted estimators based on shrinkage strategies, such as the preliminary test method, James-Stein, and positive James-Stein methods, to obtain new estimators with better performance.

In MMLMs, [4] considered the general linear restriction on the coefficient matrix and introduced the restricted estimator. Later shrinkage strategies are investigated to improve the least square estimator when the subspace is true. When the design matrix suffers from ill-condition, these estimators will be poor. In this case, the absolute value of the least square estimates will be significant and unstable. To overcome these problems, some alternative techniques have been suggested such as the partial least squares estimator, principle components estimator, the Liu and Liu-type estimator and the ordinary ridge regression estimator by [3]. But among them, the ordinary ridge regression estimator is the most popular method. In the case of MMLMs, a remedy was introduced by [2] in the form of a multivariate ridge estimator. In this paper, we consider some linear restrictions on the coefficient matrix and use shrinkage strategies to improve the ordinary ridge estimator.

The paper is organized as follows: We start with a description of the model in which we are interested. In section 3, we review the shrinkage least square estimators. In section 4, we deal with multicollinearity issues and use some estimation strategies such as the James-Stein estimator, positive James-Stein estimator, and preliminary test estimator to improve the ordinary multivariate ridge estimator. A simulation study has been conducted in section 5 to compare the proposed estimators in terms of the estimated relative efficiency. Analysis of real data examples is provided in section 6. Finally, some conclusions are given in section 7.

2 Model definition

Consider a MMLM with q response and p independent variables. Suppose that all variables are measured for n subjects. Let $\boldsymbol{Y} = [Y_1, Y_2, \cdots, Y_n]'$ be the response matrix and $\boldsymbol{X} = [X_1, X_2, \cdots, X_n]'$ be the design matrix where $Y_i = (y_{i1}, y_{i2}, \cdots, y_{iq})'$ and $X_i = (x_{i1}, x_{i2}, \cdots, x_{ip})'$ for $i = 1, 2, \cdots, n$. So the MMLM can be expressed in matrix notation as follows

$$\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{E}_{n\times q} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]'$ and $\boldsymbol{B}_{p\times q} = [\beta_1, \beta_2, \cdots, \beta_q]$ such that $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{iq})'$ and $\beta_j = (\beta_{1j}, \beta_{2j}, \cdots, \beta_{pj})'$. The usual assumptions on the errors are $\mathbb{E}(\varepsilon_j) = \boldsymbol{0}$ and $cov(\varepsilon_i, \varepsilon_k) = \boldsymbol{\Sigma}$ for $i \neq k$ where $\boldsymbol{\Sigma}$ is a $q \times q$ positive-defined matrix. If $\boldsymbol{X}$ is a full rank matrix, the usual estimator of $\boldsymbol{B}$ can be the ordinary least square estimator which is obtained by minimizing $tr\{(\boldsymbol{Y} - \boldsymbol{XB})'(\boldsymbol{Y} - \boldsymbol{XB})\}$ with respect to $\boldsymbol{B}$. The estimator is then given by $\hat{\boldsymbol{B}}^{UE} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$ which is called unrestricted estimator.

3 Shrinkage Least Square Estimators

Consider the following general linear restriction on the coefficient matrix:

$$H_0 : \boldsymbol{FBG} = \boldsymbol{d}, \tag{2}$$

where $\boldsymbol{F}$ and $\boldsymbol{G}$ are respectively $m \times p$ and $q \times r$ known full rank matrixes and $\boldsymbol{d}$ is a $m \times r$ known matrix. This subspace can be written in a vector form as $H_0 : (\boldsymbol{G'} \otimes \boldsymbol{F})vec(\boldsymbol{B}) =$

$vec(\boldsymbol{d})$. The symbol $\otimes$ stands for Kronecker product. By considering subspace defined in (2), the restricted least square estimator will be

$$\hat{\boldsymbol{B}}^{RE} = \hat{\boldsymbol{B}}^{UE} - \boldsymbol{P}(\boldsymbol{F}\hat{\boldsymbol{B}}^{UE}\boldsymbol{G} - \boldsymbol{d})\boldsymbol{Q}, \tag{3}$$

where $\boldsymbol{P} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{F}'(\boldsymbol{F}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{F}')^{-1}$, $\boldsymbol{Q} = (\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'$. The test statistic for the null hypothesis is obtained as

$$\Lambda_n = \left[vec(\boldsymbol{F}\hat{\boldsymbol{B}}^{UE}\boldsymbol{G} - \boldsymbol{d})\right]'\left(\boldsymbol{G}'\boldsymbol{\Sigma}\boldsymbol{G} \otimes \boldsymbol{F}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{F}'\right)^{-1}\left[vec(\boldsymbol{F}\hat{\boldsymbol{B}}^{UE}\boldsymbol{G} - \boldsymbol{d})\right]. \tag{4}$$

Under H_0, this statistic follows a chi-square distribution with mr degrees of freedom.

3.1 James-Stein Type Estimator

As noted, shrinkage foundation-based estimate procedures are often used to address uncertainty about the linear restrictions in H_0. Shrinkage estimating methods optimally blend restricted and unrestricted estimators to dominate unrestricted estimator. The James-Stein estimator is one of them. If we denote Λ_n as a test statistic forH_0, the James-Stein Type estimator is defined as follows:

$$\hat{\boldsymbol{B}}^{JS} = \hat{\boldsymbol{B}}^{RE} + \{1 - c\Lambda_n^{-1}\}(\hat{\boldsymbol{B}}^{UE} - \hat{\boldsymbol{B}}^{RE}); \quad mr > 2 \tag{5}$$

where c is allowed to vary over $[0, 2(mr-2))$, and often is taken as $c = mr - 2$. This estimator shrinks the unrestricted estimator toward the restricted estimator. But when $0 \leq \Lambda_n < c$, consequently $1 - c\Lambda_n^{-1} < 0$, the James-Stein estimator will suffer with the over-shrinkage problem. To avoid this issue, a truncated form of $\hat{\boldsymbol{B}}^{JS}$, called positive James-Stein estimator, is suggested. Formally, the positive James-Stein estimator, $\hat{\boldsymbol{B}}^{PJS}$ is defined as follows

$$\hat{\boldsymbol{B}}^{PJS} = \hat{\boldsymbol{B}}^{RE} + \{1 - c\Lambda_n^{-1}\}^{+}(\hat{\boldsymbol{B}}^{UE} - \hat{\boldsymbol{B}}^{RE}); \quad mr > 2 \tag{6}$$

where $z^{+} = max(0, z)$.

3.2 Preliminary Test Estimator

In the case where the linear restriction is uncertain, it may be reasonable to construct preliminary test estimator. The preliminary test estimator is defined as

$$\hat{\boldsymbol{B}}^{PTE} = \hat{\boldsymbol{B}}^{UE} - (\hat{\boldsymbol{B}}^{UE} - \hat{\boldsymbol{B}}^{RE})I_{(\Lambda_n \leq u)}, \tag{7}$$

where u is the upper $100\alpha\%$ point of the chi-square distribution with mr degrees of freedom.

4 Shrinkage Ridge Estimators

As mentioned before, the least square estimation of the model parameters may be poor when the multicollinearity presents in the model. To overcome these problems, the ordinary ridge regression estimator proposed by [3] is the most popular. In the case of MMLMs, a remedy was introduced by [2] in the form of a multivariate ridge estimator. The unrestricted ridge estimator (URE) will be obtained by minimizing $tr\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B})'(\boldsymbol{Y} -$

$\boldsymbol{XB}) + \lambda \boldsymbol{B}'\boldsymbol{B}$ with respect to $\boldsymbol{B}$ leading to $\hat{\boldsymbol{B}}^{UR} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I}_p)^{-1}\boldsymbol{X}'\boldsymbol{Y}$, where $\lambda > 0$ is the ridge parameter. $\hat{\boldsymbol{B}}^{UR}$ is a function of $\hat{\boldsymbol{B}}^{UE}$ such that $\hat{\boldsymbol{B}}^{UR} = \boldsymbol{\mathcal{R}}(\lambda)\hat{\boldsymbol{B}}^{UE}$, where $\boldsymbol{\mathcal{R}}(\lambda) = (\boldsymbol{I}_p + \lambda(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}$.
Following [1], we propose ridge-type estimators of $\boldsymbol{B}$ as:

- Restricted ridge-type estimator (RRE): $\hat{\boldsymbol{B}}^{RR} = \boldsymbol{\mathcal{R}}(\lambda)\hat{\boldsymbol{B}}^{RE}$
- James-Stein ridge-type estimator (JSRE): $\hat{\boldsymbol{B}}^{JSR} = \boldsymbol{\mathcal{R}}(\lambda)\hat{\boldsymbol{B}}^{JS}$
- Positive James-Stein ridge-type estimator (PJSRE): $\hat{\boldsymbol{B}}^{PJSR} = \boldsymbol{\mathcal{R}}(\lambda)\hat{\boldsymbol{B}}^{PJS}$
- Preliminary test ridge-type estimator (PTRE): $\hat{\boldsymbol{B}}^{PTR} = \boldsymbol{\mathcal{R}}(\lambda)\hat{\boldsymbol{B}}^{PT}$

5 Simulation Study

In this section, we compare the performance of the ridge-type estimators defined in section 3 by a simulation study for different sample sizes and by supposing different degrees of multicollinearity. We consider two response variables and five independent variables. The independent variables are generated using the following equation:

$$x_{lj} = (1-\rho^2)^{1/2} z_{lj} + \rho z_{lp}; \qquad l : 1, 2, \cdots, n \qquad j : 1, 2, \cdots, p \tag{8}$$

where ρ represents the correlation between two independent variables and z_{lj}'s are independent standard normal pseudo-random numbers. Four different sets of correlation corresponding to 0.5, 0.7, and 0.9, three different values of n, $n = \{50, 100\}$, and $\alpha = 0.05$ are considered. The coefficient matrix and covariance matrix of the error term are selected as

$$\boldsymbol{B} = \begin{bmatrix} -1 & 1.25 & 0.5 & 0.75 & 1 \\ 0.5 & 1.2 & 1.7 & 0.3 & -0.5 \end{bmatrix}' \qquad and \qquad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}$$

For given values of n and ρ, a set of independent variables is generated, then n observations are determined by $Y_l = (y_{l1}, y_{l2}) = X_l\boldsymbol{B} + e_l;\ l : 1, 2, \cdots, n$ where e_l's are independent and generated from $\mathcal{N}_2(\boldsymbol{0}, \boldsymbol{\Sigma})$. In order to determine the subspace produced under the hypothesis H_0 defined in (2), we use the following matrixes:

$$\boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0.5 & 0 & 1 \\ 0 & 1.2 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 1 \end{bmatrix} \qquad and \qquad \boldsymbol{G} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We determine λ by $\hat{\lambda}_{HK} = \hat{\sigma}^2/\hat{\beta}^2_{max}$ where $\hat{\beta}_{max}$ is the maximum element of $\hat{\boldsymbol{B}}^{UE}$. We repeat each simulation procedure 2000 times by generating new random numbers for each repetition. In order to compare the proposed estimators, $\hat{\boldsymbol{B}}^*$, we use the relative efficiency of estimator $\hat{\boldsymbol{B}}^*$ obtained by

$$R.E(\hat{\boldsymbol{B}}^*) = \frac{MSE(\hat{\boldsymbol{B}}^{UR})}{MSE(\hat{\boldsymbol{B}}^*)}$$

where

$$MSE(\hat{\boldsymbol{B}}^*) = \frac{1}{2000}\sum_{m=1}^{2000}\sum_{i=1}^{q}\sum_{\omega=1}^{p}\left(\hat{\beta}^*_{\omega i(m)} - \beta_{\omega i}\right)^2$$

Table 1: Relative efficiency of the suggested estimators for different values of n and ρ.

	$n = 50$				$n = 100$			
ρ	RRE	JSRE	PJSRE	PTRE	RRE	JSRE	PJSRE	PTRE
0.5	2.5248	1.6797	1.8657	2.1308	2.1802	1.6308	1.8543	2.1204
0.7	2.5243	1.6791	1.8650	2.1311	2.1794	1.6312	1.8542	2.1203
0.9	2.5216	1.6762	1.8616	2.1319	2.1733	1.6320	1.8530	2.1172

and $\hat{\beta}^*_{\omega i(m)}$ is the estimation of $\beta_{\omega i}$ in the mth repetition. When null the hypothesis is true, the simulation results illustrated in Table 1 show that: When ρ increases, the relative efficiency of suggested estimators with respect to unrestricted ridge estimator reduces. Nevertheless, these estimators are still more efficient than unrestricted ridge estimator. When n increases, the relative efficiency of suggested estimators with respect to unrestricted ridge estimator increases. in all cases, we have:

$$R.E(\hat{\boldsymbol{B}}^{JSR}) \leq R.E(\hat{\boldsymbol{B}}^{PJSR}) \leq R.E(\hat{\boldsymbol{B}}^{PTR}) \leq R.E(\hat{\boldsymbol{B}}^{RR}).$$

6 Real Data Analysis

The real dataset is the tobacco data set that has been initially studied by [5]. Dataset involves a sample of size 25 of tobacco leaf for organic and inorganic chemical constituents. There are three response variables and six independent variables. The response variables are the rate of cigarette burn in inches per 1000 seconds y_1, the percent sugar in the leaf y_2, and the percent nicotine y_3. The independent variables are the percentages of total nitrogen x_1, of chlorine x_2, of potassium x_3, of phosphorus x_4, of calcium x_5, and of magnesium x_6. In order to determine a prior information on a possible best model for the response variable, we first do an analysis on each response variable $y_i, i = 1, 2, 3$ based on stepwise selection and AIC criterion. The results show that the best model for y_1 contains x_2, x_3, x_5, and x_6 for y_2, it contains x_1, x_2, x_4, and x_6 and about y_3, the best model contains x_1, x_2, and x_6. Since x_3, x_4, and x_5 appear only in one model, we impose the following restriction on the multivariate multiple linear model which means x_3, x_4, and x_5 are deleted from the model. Thus, our restrictions will be:

$$\begin{aligned}
\beta_{31} = \beta_{32} = \beta_{33} = 0 \\
\beta_{41} = \beta_{42} = \beta_{43} = 0 \\
\beta_{51} = \beta_{52} = \beta_{53} = 0
\end{aligned}$$

Based on the above restrictions, we can select the following matrixes

$$F = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad and \quad G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \quad and \quad d = \mathbf{0}_{3\times 3}.$$

We conducted the bootstrap method to compute the relative efficiency of the suggested estimators. The bootstrap sample size is chosen $m = 15$ with 2000 times replication. We determined λ by $\hat{\lambda}_{HK}$ and considered $\alpha = 0.05$. The results are presented in Table 2 which shows that the proposed estimators are more efficient than unrestricted ridge estimator, and the positive James-Stein ridge-type estimator is the best in terms of efficiency.

Table 2: Relative efficiency of estimators for the tobacco data set.

Estimators	RRE	JSRE	PJSRE	PTRE
R. E	2.8232	1.3903	2.8383	2.6678

7 Conclusion

In this article, we consider the multicollinearity issue in the MMLMs. We use some shrinkage strategies to improve the ordinary ridge estimator. We introduce the restricted ridge estimator, James-Stein ridge-type estimator, positive James-Stein ridge-type estimator, and preliminary ridge-type estimator. By numerical example, we simulate the relative efficiency of proposed estimators such that when the subspace holds on parameters, all proposed estimators perform better than unrestricted ridge estimator. The performance of suggested estimators is considered in real data set and the relative efficiency of them is estimated. The results show that the proposed estimators are efficient than unrestricted ridge estimator.

References

[1] M. Arashi, S. M. M. Tabatabaey, and M. Hassanzadeh Bashtian, Shrinkage Ridge Estimators in Linear Regression, *Commun. Stat. Simul*, 43(4) (2014), 871–904.

[2] E. Füle, On ecological regression and ridge estimation, *Commun. Stat. Simul*, 24(2) (1995), 385–398.

[3] A. E. Hoerl and R. W. Kennard, Ridge regression: biased estimation for non-orthogonal problems, *Technometrics*, 12 (1970), 69–82.

[4] K. Kim and N. Timm, *Univariate and Multivariate General Linear Models: Theory and Applications with SAS*, Second Edition, Chapman & Hall, 2007.

[5] R. S. Sparks, W. Zucchini, and D. Coutsourides, On Variable Selection In Multivariate Regression, *Commu. Stat. Theor. Meth.*, 14(7) (1985), 1569–1587.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

An approximation for the conformable time-space fractional diffusion equation

Haniye Hajinezhad and Homa Afraz*

Department of Mathematics, Payame Noor University, Tehran, Iran

Abstract

The present paper suggests a finite difference method to approximate the one-dimensional time-space fractional diffusion equation, with the initial and boundary conditions, that employs the conformable fractional derivative. The stability and convergence of this approach are proved, and numerical experiments confirm its second-order accuracy. One of the advantages of this method is that it can be readily extended to the two or three-dimensional time-space fractional diffusion equation that uses the conformable fractional derivative.

Keywords: Time-space fractional diffusion equation, Conformable derivative, Crank-Nicolson method, Convergence, Stability

Mathematics Subject Classification [2010]: 65N25;65M06; 65M12; 26A33; 35R11

1 Introduction

The scientific community has displayed a notable interest in the time-space fractional diffusion equation (TSFDE), which replaces integer-order derivatives in the conventional diffusion equation with fractional derivatives. This equation has widespread applications in studying anomalous diffusive processes in various fields.

Arshad et al. [1] introduced a finite difference technique for solving the one-dimensional TSFD with second-order accuracy in time and space. They utilized Caputo and Riesz derivatives to represent the time and space fractional derivatives, respectively. The centered difference method was used to estimate the Riesz fractional derivative in space. The fractional ordinary differential equations resulting from this approach were then transformed into Volterra integral equations. Finally, the trapezoidal rule was utilized to estimate these integral equations.

In the realm of fractional derivatives, the Riemann-Liouville and Caputo definitions are the most commonly used. Although these definitions adhere to linearity properties, they do not satisfy other properties of the ordinary derivative, such as the quotient rule or product rule. To address this, Khalil et al. [4] introduced a well-behaved definition of the fractional derivative known as the conformable fractional derivative. This definition adheres to the product rule, quotient rule, chain rule, and composition rule, among other properties. Consequently, the conformable fractional derivative facilitates a more effective

*Speaker. Email address: homaafraz@pnu.ac.ir

analysis of fractional partial differential equations.
Bayrak et al. [2] investigated the solution of a one-dimensional TSFDE with both initial and boundary conditions. Their study incorporated the conformable derivatives and utilized the residual power series method to develop a solution for one-dimensional TSFDEs. The researchers employed a specific transformation to convert the TSFDEs into either space or time fractional diffusion equations. Subsequently, they obtained solutions in the form of fractional power series through the semi-analytical residual power series method. Bayrak et al. [2] used two different approaches: one that utilized the initial condition and another that employed the boundary conditions. These approaches yielded significantly different results. However, in this work, an approximation is presented that considers both the initial and boundary conditions of the one-dimensional TSFDE simultaneously, leading to a more accurate solution.
This paper presents a finite difference scheme based on the Crank-Nicolson method for the one-dimensional TSFDE with the initial and boundary conditions using the conformable derivative. The rigorous proofs of the stability and convergence of this approach are provided. Furthermore, Some computational experiments using a test problem that features an analytical solution with the conformable derivative are provided. The numerical tests verify the accuracy and efficiency of the proposed method. This method is easier to use compared to the method presented by Bayrak et al. [2]. It is worth noting that the development of the proposed scheme, as well as its stability and convergence analysis for two-dimensional and three-dimensional TSFDEs, are similar to those of the one-dimensional case.
In the next Section, the conformable fractional diffusion equation is defined. In Section 3, the discretization of TSFDE with conformable derivative is explained. Section 4 proves the Stability and convergence of the proposed scheme. Section 5 verifies the accuracy and efficiency of the proposed scheme. Finally, the conclusion is given.

2 Conformable fractional diffusion equation

In 2014, Khalil et al. [4] defined the β-order conformable fractional derivative of g, as follows:

$$D^{\alpha}g(x) = \lim_{\epsilon \to 0} \frac{g^{(\lceil \beta \rceil - 1)}(x + \epsilon x^{\lceil \beta \rceil - \beta}) - g^{(\lceil \beta \rceil - 1)}(x)}{\epsilon}, \qquad n < \beta \leqslant n+1,$$

where $\lceil \beta \rceil$ is the smallest integer greater than or equal to β. As a consequence of this definition, it is straightforward to show [4]:

$$D^{\beta}g(x) = t^{n+1-\beta} g^{(n+1)}(x), \qquad n < \beta \leqslant n+1, \tag{1}$$

where that g is $(n+1)$-differentiable at $x > 0$. This definition follows various fundamental rules of calculus such as the product rule, quotient rule, chain rule, composition rule, and other related properties. As a result, the conformable fractional derivative provides a more efficient approach to analyzing fractional partial differential equations.
In this paper, the one-dimensional time-space fractional diffusion equation (TSFDE)

$$\frac{\partial^{\alpha} u(x,t)}{\partial t^{\alpha}} = D(x) \frac{\partial^{\beta} u(x,t)}{\partial x^{\beta}} + f(x,t), \qquad 0 < x < L, 0 < t \leqslant T, 0 < \alpha \leqslant 1, 1 < \beta \leqslant 2, \tag{2}$$

with initial and boundary conditions

$$u(x,0) = \Theta(x), \qquad 0 \leqslant x \leqslant L, \tag{3}$$

$$u(0,t) = \phi(t), \qquad 0 \leqslant t \leqslant T, \tag{4}$$
$$u(L,t) = \psi(t), \qquad 0 \leqslant t \leqslant T, \tag{5}$$

is considered, where using relation (1), the conformable fractional derivatives are as follows:

$$\frac{\partial^\alpha u(x,t)}{\partial t^\alpha} = t^{1-\alpha}\frac{\partial u(x,t)}{\partial t}, \qquad 0 < \alpha \leqslant 1, \tag{6}$$
$$\frac{\partial^\beta u(x,t)}{\partial x^\beta} = x^{2-\beta}\frac{\partial^2 u(x,t)}{\partial x^2}, \qquad 1 < \beta \leqslant 2. \tag{7}$$

Therefore, equation (2) using relations (6)-(7) is as follows:

$$t^{1-\alpha}\frac{\partial u(x,t)}{\partial t} = D(x)x^{2-\beta}\frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t), \qquad 0 < x < L, 0 < t \leqslant T, 0 < \alpha \leqslant 1, 1 < \beta \leqslant 2. \tag{8}$$

In this work, a discretization for the one-dimensional TSFDE (8) with conditions (3)-(5) using the Crank-Nicolson method is presented. Then the stability and convergence of the proposed scheme are demonstrated. The primary advantage of the proposed scheme is its flexibility in extending to two or three-dimensional TSFDEs with ease. Specifically, the two-dimensional TSFDE is defined by the equation:

$$\frac{\partial^\alpha u(x,y,t)}{\partial t^\alpha} = D_1(x,y)\frac{\partial^{\beta_1} u(x,y,t)}{\partial x^{\beta_1}} + D_2(x,y)\frac{\partial^{\beta_2} u(x,y,t)}{\partial y^{\beta_2}} + f(x,y,t),$$
$$(x,y) \in (0,L_1)\times(0,L_2), 0 < t \leqslant T, 0 < \alpha \leqslant 1, 1 < \beta_1, \beta_2 \leqslant 2,$$

Similarly, the three-dimensional TSFDE is defined by the equation:

$$\frac{\partial^\alpha u(x,y,z,t)}{\partial t^\alpha} = D_1(x,y,z)\frac{\partial^{\beta_1} u(x,y,z,t)}{\partial x^{\beta_1}} + D_2(x,y,z)\frac{\partial^{\beta_2} u(x,y,z,t)}{\partial y^{\beta_2}}$$
$$+ D_3(x,y,z)\frac{\partial^{\beta_3} u(x,y,z,t)}{\partial z^{\beta_3}} + f(x,y,z,t),$$
$$(x,y,z) \in (0,L_1)\times(0,L_2)\times(0,L_3), 0 < t \leqslant T, 0 < \alpha \leqslant 1, 1 < \beta_1, \beta_2, \beta_3 \leqslant 2,$$

where, in both of these equations the conformable fractional derivatives and initial and boundary conditions similar to the one-dimensional TSFDE are used.

3 Crank-Nicolson method

The discretization of equation (8) with conditions (3)-(5) using the Crank-Nicolson method is presented in this section.
Consider the grid size in space and in time to be Δx and Δt, respectively. Then, $x_j = j\Delta x$ $(j = 0, 1, ..., J)$, $t^n = n\Delta t$ $(n = 0, 1, ..., N)$, and $t^{n+\frac{1}{2}} = (n+\frac{1}{2})\Delta t$ $(n = 0, 1, ..., N-1)$, where $J\Delta x = J$ and $N\Delta t = T$. Also, u_j^n is the value of $u(x_j, t^n)$ for $j = 0, 1, ..., J$, and $n = 0, 1, ..., N$.
Assume the following discretization using the Crank-Nicolson method:

$$(t^{1-\alpha}\frac{\partial u(x,t)}{\partial x})\Big|_j^{n+\frac{1}{2}} = (t^{n+\frac{1}{2}})^{1-\alpha}\{\frac{u_j^{n+1} - u_j^n}{\Delta t} + \mathrm{O}(\Delta t)^2\}, \qquad 0 \leqslant n \leqslant N-1,\ 1 \leqslant j \leqslant J-1, \tag{9}$$

$$(x^{2-\beta}\frac{\partial^2 u(x,t)}{\partial x^2})|_j^{n+\frac{1}{2}} = {x_j}^{2-\beta}\frac{1}{2}\{\frac{u_{j+1}^{n+1}-2u_j^{n+1}+u_{j-1}^{n+1}}{(\Delta x)^2}+\frac{u_{j+1}^{n}-2u_j^{n}+u_{j-1}^{n}}{(\Delta x)^2} + \mathrm{O}(\Delta x)^2+\mathrm{O}(\Delta t)^2\}, \quad 0\leqslant n\leqslant N-1,\ 1\leqslant j\leqslant J-1. \tag{10}$$

Therefore, by disregarding the truncation errors, the discretization of equation (8) with conditions (3)-(5) using relations (9)-(10) are as follows:

$$u_1^{n+1}-\gamma^n\gamma_1(-2u_1^{n+1}+u_2^{n+1}) = u_1^n+\gamma^n\gamma_1(-2u_1^n+u_2^n) + \gamma^n\gamma_1(\phi(t^n)+\phi(t^{n+1}))+\gamma^n f_1^{n+\frac{1}{2}}, \quad 0\leqslant n\leqslant N-1, \tag{11}$$

$$u_j^{n+1}-\gamma^n\gamma_j(u_{j+1}^{n+1}-2u_j^{n+1}+u_{j-1}^{n+1}) = u_j^n+\gamma^n\gamma_j(u_{j+1}^{n}-2u_j^{n}+u_{j-1}^{n}) + \gamma^n f_j^{n+\frac{1}{2}}, \quad 2\leqslant j\leqslant J-2, \quad 0\leqslant n\leqslant N-1, \tag{12}$$

$$u_{J-1}^{n+1}-\gamma^n\gamma_{J-1}(u_{J-2}^{n+1}-2u_{J-1}^{n+1}) = u_{J-1}^n+\gamma^n\gamma_{J-1}(u_{J-2}^{n}-2u_{J-1}^{n}) + \gamma^n\gamma_{J-1}(\psi(t^n)+\psi(t^{n+1}))+\gamma^n f_{J-1}^{n+\frac{1}{2}}, \quad 0\leqslant n\leqslant N-1, \tag{13}$$

where $\gamma_j = D(x_j)\frac{(j\Delta x)^{2-\beta}}{2(\Delta x)^2}$, $\gamma^n = \frac{\Delta t}{((n+\frac{1}{2})\Delta t)^{1-\alpha}}$, and $f_j^{n+\frac{1}{2}} = f(x_j, t^{n+\frac{1}{2}})$, for $1\leqslant j\leqslant J-1$ and $0\leqslant n\leqslant N-1$. Now the following theorem can be easily proved.

Theorem 3.1. *The discretization equation (8) with conditions (3)-(5) using relation (9)-(10) is consistent with accuracy* $\mathrm{O}(\Delta x)^2+\mathrm{O}(\Delta t)^2$.

4 Stability and convergence

This section is dedicated to prove the stability and convergence of obtained scheme (11)-(13).
The matrix form of equations (11)-(13) is as follows:

$$(I-\gamma^n A)U^{n+1} = (I+\gamma^n A)U^n+\gamma^n F^{n+\frac{1}{2}}, \quad 0\leqslant n\leqslant N-1, \tag{14}$$

where
$U^n = [u_1^n, u_1^n, ..., u_{J-1}^n]^T$ for $0\leqslant n\leqslant N$,
$F^{n+\frac{1}{2}} = [f_1^{n+\frac{1}{2}}+\gamma^n(\phi(t^n)+\phi(t^{n+1})), f_2^{n+\frac{1}{2}}, f_3^{n+\frac{1}{2}}, ..., f_{J-2}^{n+\frac{1}{2}}, f_{J-1}^{n+\frac{1}{2}}+\gamma_{J-1}(\psi(t^n)+\psi(t^{n+1}))]^T$ for $0\leqslant n\leqslant N-1$,
I is a $(J-1)\times(J-1)$ identity matrix,
and A is a $(J-1)\times(J-1)$ matrix as follow:

$$A = \begin{bmatrix} -2\gamma_1 & \gamma_1 & 0 & 0 & \cdots & 0 & 0 \\ \gamma_2 & -2\gamma_2 & \gamma_2 & 0 & \cdots & 0 & 0 \\ 0 & \gamma_3 & -2\gamma_3 & \gamma_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -2\gamma_{n-2} & \gamma_{n-2} \\ 0 & 0 & 0 & 0 & \cdots & \gamma_{n-1} & -2\gamma_{n-1} \end{bmatrix}.$$

Theorem 4.1. *The discretization of equation (8) with conditions (3)-(5), defined by (14), is unconditionary stable.*

Proof. By using the Gereshgorin theorem ([3] p. 294) , for Matrix A, we have

$$\begin{cases} |\lambda_1 + 2\gamma_1| \leqslant \gamma_1, \\ |\lambda_j + 2\gamma_j| \leqslant 2\gamma_j, \quad 2 \leqslant j \leqslant J-2, \\ |\lambda_{J-1} + 2\gamma_{J-1}| \leqslant \gamma_{J-1}, \end{cases}$$

where λ_j $(1 \leqslant j \leqslant J-1)$ is the eigenvalue of matrix A. Additionally, matrix A is invertible. Therefore the real-part of $\lambda_j (j = 1, 2, ..., J-1)$ is not positive. So $|\frac{1+\lambda_j}{1-\lambda j}| < 1$, for $1 \leqslant j \leqslant J-1$.
As we know, λ_j is the eigenvalue of the matrix A if and only if $\frac{1+\lambda_j}{1-\lambda j}$ is an eigenvalue of the matrix $(I-A)^{-1}(I+A)$. These result in the equations system (14) is unconditionally stable. □

Therefore the convergence of the proposed scheme (14) is proved according to the Lax equivalence theorem [5] and by using theorems 3.1 and 4.1.

5 Evaluation

In order to confirm the accuracy of the suggested method, the maximum absolute error as the verification parameter is defined. This is done by assuming certain values for Δx and Δt and using the following relation:

$$L_\infty(\Delta x, \Delta t) = max_{1 \leqslant j \leqslant J-1, 1 \leqslant n \leqslant N} |\widehat{u}_j^n - u_j^n|,$$

where $\widehat{u}_j^n$ and u_j^n are the approximated and exact solutions, respectively, of the equation (2) under the conditions (3)–(5) at the location x_j and the time t^n. To test the proposed method, the following example is used where the exact solution is known.

Example Consider the equation (8) where

$$f(x,t) = 2t^{2-\alpha}x^2(-1+x)^2 - 12(x^2 - x + \frac{1}{2})t^2x^{2-\beta}$$

and conditions (3)–(5) are defined by the exact solution $u(x,t) = t^2x^2(1-x)^2$.
The effectiveness of the proposed method is evaluated by testing it with different values of α and β, as well as by varying Δx and Δt. Tables 1 demonstrates that the maximum absolute errors are small. So, the proposed method is accurate enough. The ratio of errors as the refinement of the grids, using the following equation is defined.

$$Error\ rate = \frac{L_\infty((\Delta x)_1, (\Delta t)_1)}{L_\infty((\Delta x)_2, (\Delta t)_2)},$$

where $(\Delta x)_2 = (\Delta t)_2 < (\Delta x)_1 = (\Delta t)_1$. Tables 1 indicates that the proposed method has second-order accuracy. This means that the ratio of $L_\infty((\Delta x)_1, (\Delta t)_1)$ to $L_\infty((\Delta x)_2, (\Delta t)_2)$ is approximately equal to the square of the ratio of $(\Delta x)_1$ or $(\Delta t)_1$ to $(\Delta x)_2$ or $(\Delta t)_2$.

Table 1: The maximum absolute error and error rate with different Δx, Δt, α, and β.

		$\Delta x, \Delta t = \frac{1}{10}$	$\Delta x, \Delta t = \frac{1}{100}$	$\Delta x, \Delta t = \frac{1}{1000}$
$\alpha = 0.5$	L_∞	$1.7084\text{e}-03$	$1.7185\text{e}-05$	$1.7186\text{e}-07$
$\beta = 1.5$	Error rate	-	$99.4122 \approx \left(\frac{100}{10}\right)^2$	$99.9941 \approx \left(\frac{1000}{100}\right)^2$
$\alpha = 0.3$	L_∞	$1.8312\text{e}-03$	$1.8397\text{e}-05$	$1.8399\text{e}-07$
$\beta = 1.8$	Error rate	-	$99.5379 \approx \left(\frac{100}{10}\right)^2$	$99.9891 \approx \left(\frac{1000}{100}\right)^2$
$\alpha = 0.8$	L_∞	$1.8422\text{e}-03$	$1.8516\text{e}-05$	$1.8517\text{e}-07$
$\beta = 1.9$	Error rate	-	$99.4923 \approx \left(\frac{100}{10}\right)^2$	$99.9945 \approx \left(\frac{1000}{100}\right)^2$
$\alpha = 0.2$	L_∞	$1.5489\text{e}-03$	$1.5629\text{e}-05$	$1.5630\text{e}-07$
$\beta = 1.1$	Error rate	-	$99.1042 \approx \left(\frac{50}{20}\right)^2$	$99.9936 \approx \left(\frac{150}{50}\right)^2$

6 Conclusion

This paper introduces a finite difference scheme by the Crank-Nicolson method to approximate the solution of the one-dimensional time-space fractional diffusion equation (TSFDE) with the initial and boundary conditions using the conformable derivatives. The conformable derivatives are used to simplify the analysis of fractional derivatives. Moreover, the stability of the proposed scheme has been demonstrated through the Gershgorin theorem, and convergence has been demonstrated through the Lax equivalence theorem. The numerical experiments showed that the proposed scheme is accurate enough. Furthermore, numerical tests demonstrated second-order accuracy. These results can be easily extended to two- or three-dimensional time-space fractional diffusion equations.

References

[1] Arshad, S., et al. "Trapezoidal scheme for time–space fractional diffusion equation with Riesz derivative." *Journal of Computational Physics*, 350 (2017), 1-15.

[2] Bayrak, M. A., et al. "A novel approach for the solution of fractional diffusion problems with the conformable derivative." *Numerical Methods for Partial Differential Equations*, (2021), 1-18.

[3] Datta, B. N., *Numerical linear algebra and applications*, 2nd edition, Siam, 2010.

[4] Khalil, R., et al. "A new definition of fractional derivative." *Journal of Computational and Applied Mathematics*, 264 (2014), 65-70.

[5] Lax, P. D., and Richtmyer, R. D., Survey of the stability of linear finite difference equations. *Communications on Pure and Applied Mathematics*, 9 (1956). No. 2, 267-293.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Control of condition number in Spectral Galerkin implementation for solving generalized Abel integral equation

Sayyed Rasoul Kafi[1]*, Payam Mokhtary[2] and Esmail Hesameddini[1]

[1]Department of Mathematics, Shiraz University of Technology, Shiraz, Iran

[2]Department of Mathematics, Sahand University of Technology, Tabriz, Iran

Abstract

Numerical discretization of functional equation using spectral methods usually leads to a linear system of algebraic equations with a high conditional and full coefficient matrix. This drawback destroys the accuracy specially for large values of approximation degree. In this paper, we propose a new spectral Galerkin implementation for solving a class of generalized Abel integral equations, with the purpose of controlling condition number and recovering the familiar spectral accuracy.

Keywords: Generalized Abel integral equation, spectral Galerkin, condition number.
Mathematics Subject Classification [2010]: 65F35, 65G30, 65K05

1 Introduction

Consider the following generalized Abel integral equation

$$y(x) = g(x) + \int_0^x K(x,t)\big(x^\eta - t^\eta\big)^{\alpha-1} y(t)dt \qquad \eta > 1, \quad x \in S = [0,T], \tag{1}$$

which y is unknown, $\alpha \in (0,1), \eta(1-\alpha) = 1$. Also, the functions g and K are real valued and continuous on S and $D = \big\{(x,t)|t \in [0,T], t \in [0,x)\big\}$, respectively. From [1], we can deduce in (1) smooth data conclude smooth solution, so applying spectral methods to acquire the suitable approximate solution is reasonable. However, applying the classical version of spectral methods [2] usually turns into solving a full and high conditioning algebraic systems. To fix this difficulty, we intend to design a new strategy that not only avoids producing full and complex systems but also obtains the approximate solution using some recursive relations.

This paper is organized as follows: In the next section, we explain the new implementation process. In Section 3, we examine the proposed strategy on a test problem. The last section devotes for our conclusions.

*Speaker. Email address: r.kaafi@sutech.ac.ir

2 Implementation approach

In this part, a new approach is described for implementing the spectral Galerkin method for solving (1). We consider the approximate solution of (1) by

$$y_N(x) = \sum_{p=0}^{\infty} y_p J_p^{\alpha,\beta}(x) = \underline{y}\underline{J}^{\alpha,\beta}, \tag{2}$$

and assume that

$$\begin{aligned} g(x) &\simeq \sum_{p=0}^{\infty} \tilde{g}_p J_p^{\alpha,\beta}(x) = \sum_{p=0}^{\infty} g_p x^p = \underline{g}\underline{X}, \\ K(x,t) &\simeq \sum_{p=0}^{N}\sum_{q=0}^{N} \tilde{k}_{pq} J_p^{\alpha,\beta}(x) J_p^{\alpha,\beta}(t) = \sum_{p=0}^{N}\sum_{q=0}^{N} k_{pq} x^p t^q \end{aligned} \tag{3}$$

where we have

$$\begin{aligned} &\underline{y} = [y_0, y_1, ..., y_N, 0, ...], \quad \underline{J}^{\alpha,\beta} = [J_0^{\alpha,\beta}(x), J_1^{\alpha,\beta}(x), ..., J_N^{\alpha,\beta}(x), ...] = J^{\alpha,\beta}\underline{X}, \\ &\underline{g} = [g_0, g_1, ..., g_N, 0, ...], \quad \underline{X} = [1, x, ..., x^N, ...], \end{aligned}$$

such that $J_i^{\alpha,\beta}(x)$ is the i-th Jacobi polynomials with parameters $\alpha, \beta > -1$ and $J^{\alpha,\beta}$ is an infinite coefficient matrix.

Substituting the relations (2) and (3) into (1), we obtain

$$\underline{y}J^{\alpha,\beta}\underline{X} = \underline{g}\underline{X} + \underline{y}J^{\alpha,\beta}E\underline{X}, \tag{4}$$

or equivalently

$$\underline{y}J^{\alpha,\beta}(Id - E)\underline{X} = \underline{g}\underline{X}, \tag{5}$$

where Id is the identity matrix and the only non-zero entries of E are given by

$$E_{\zeta,\zeta+\tau} = \sum_{\theta=0}^{\tau} k_{\theta,\tau-\theta}\mathcal{B}_{\tau-\theta,\zeta}^{\alpha,\beta}, \quad \forall \zeta, \tau \in \mathbb{N} \cup \{0\}.$$

such that $\mathcal{B}_{q,\theta}^{\alpha,\beta} = \eta^{-1}B\left(\frac{q+\theta+1}{\eta}, \alpha\right)$ and $B(.,.)$ is the well-known beta function. Projecting (5) into the finite dimensional space, $< \{J_0^{\alpha,\beta}(x), J_1^{\alpha,\beta}(x), ..., J_N^{\alpha,\beta}(x)\} >$, concludes the following $(N+1)\times(N+1)$ algebraic system

$$\underline{y}J^{\alpha,\beta}(Id - E)\left(J^{\alpha,\beta}\right)^{-1} = \underline{g}\left(J^{\alpha,\beta}\right)^{-1}, \tag{6}$$

in view of using the relation $\underline{X} = \left(J^{\alpha,\beta}\right)^{-1}\underline{J}^{\alpha,\beta}$. Actually, the linear algebraic system (6) is full and complex with high conditioning property which causes low accuracy results especially for large values of N. To fix this drawback, we design a well-conditioned strategy as follows: Assume $\overline{y} = \underline{y}J^{\alpha,\beta} = [\tilde{y}_0, \tilde{y}_1, ..., \tilde{y}_N, 0, ...]$, so multiplying both sides of (6) by $J^{\alpha,\beta}$, we can obtain

$$\overline{y}(Id - E) = \underline{g}, \tag{7}$$

Table 1: Comparisons between condition numbers for different N.

N	the new approach	the classical approach
10	1.9520	8.268×10^{2}
20	1.9999	8.418×10^{5}
30	2.0017	5.382×10^{8}
40	2.0017	3.872×10^{11}
50	2.0017	7.683×10^{14}
60	2.0017	7.603×10^{17}

Table 2: Comparisons between numerical errors for different N.

N	the new approach	the classical approach
10	1.47789×10^{-5}	1.47789×10^{-5}
20	8.78504×10^{-15}	8.78504×10^{-15}
30	1.63698×10^{-16}	1.63698×10^{-16}
40	1.63698×10^{-16}	4.00722×10^{-15}
50	1.63698×10^{-16}	3.16758×10^{-11}
60	1.63698×10^{-16}	7.14705×10^{-6}

which has a sparse and lower triangular coefficient matrix that can be solved recursively by

$$\tilde{y}_0 = \frac{g_0}{J^{\alpha,\beta}_{0,0}}, \ \tilde{y}_i = \frac{g_i - \sum_{\kappa=0}^{i-1} \tilde{y}_\kappa J^{\alpha,\beta}_{i,\kappa}}{J^{\alpha,\beta}_{i,i}} \quad \forall i \in \{0,1,2,...,N\},$$

and eventually, we can obtain the unknown vector $\underline{y}$ by solving the triangular system $\overline{y} = \underline{y} J^{\alpha,\beta}$.

3 Numerical results

In this section, we focus on effect of controlling condition number in recovering familiar spectral accuracy.

Example 3.1. Consider (1) with $K(x,t) = \frac{1}{2}\sin(2x^2 t)$, $\eta = 4$ and $\alpha = \frac{3}{4}$. $g(x)$ is chosen so that the exact solution be $y(x) = \sin(x)$.

We implement both classical and new approaches of spectral Galerkin method by solving both systems (6) and (7), respectively. The obtained results are reported in Tables 1 and 2 by making a comparison between two approach monitoring numerical errors and condition numbers. In Figure 1, different solution based on two different condition number is showed, too. Indeed, the results confirm superiority of our proposed scheme in controlling condition number and regularly decaying of the numerical errors.

4 Conclusion

In this work, we designed a new spectral Galerkin implementation approach to catch the approximate solution of a class of generalized Abel integral equations by means of solving a sparse and low conditional linear system. We approved that our strategy has a significant superiority over the classical one.

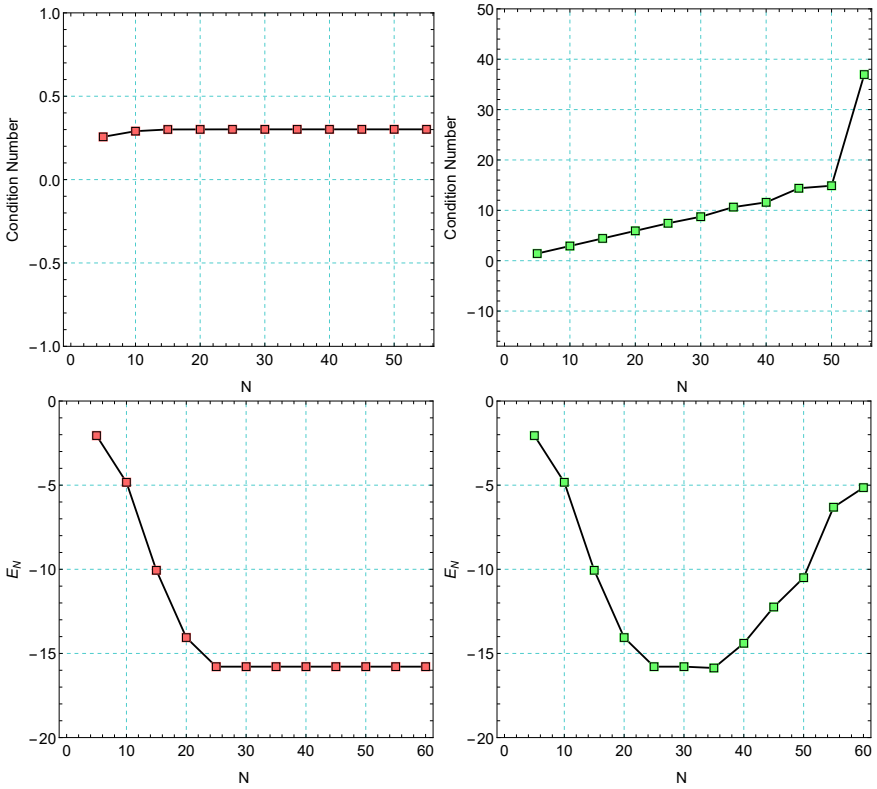


Figure 1: Comparing of numerical error graphs and also condition number graphs. In each figure, one is attained from the classical method and the former is attained from the new method

References

[1] H. Brunner, *Volterra Integral Equations: An Introduction to Theory and Applications*, Cambridge University Press, Jan 2017.

[2] J. Shen, T. Tang and L.L. Wang, *Spectral Methods: Algorithms, Analysis and Applications*, Springer Science & Business Media, Aug 2011.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Optimal scaling of the memoryless quasi–Newton updating formulas

Saman Babaie–Kafaki*

Faculty of Engineering, Free University of Bozen–Bolzano, Bolzano, Italy

Abstract

Matrix approximations generated by the quasi–Newton (QN) updates may be generally vulnerable to ill-conditioning. Thus, the QN algorithms for unconstrained optimization may fail to suggest a proper trajectory to the solution. Here, by matrix analyses, it is discussed that how the classic scaling schemes of the QN algorithms can be modified to make further improvement in the computational stability of the methods. The argument mainly centers on a well-know open problem.

Keywords: Nonlinear programming, quasi–Newton update, scaling, condition number, eigenvalue

Mathematics Subject Classification [2010]: 65K05, 90C53, 15A18

1 Introduction

Because of emerging high dimensional data sets in the real world models, majority of the scientific researches is devoted to devise effective memoryless versions of the classic algorithms. Among such efforts in continuous optimization, there are the memoryless (limited memory) QN algorithms.

As known, QN algorithms are a class of line search methods in which the search directions are iteratively determined by

$$\mathbf{d}_0 = -\mathbf{g}_0, \ \mathbf{d}_{k+1} = -\mathbf{H}_{k+1}\mathbf{g}_{k+1}, \quad \text{for all } k \geq 0,$$

where $\mathbf{g}_k$ is the gradient vector at the iterate $\mathbf{x}_k$ and $\mathbf{H}_k$ is an $n \times n$ (symmetric) positive definite approximation of $\nabla^2\mathbf{f}(\mathbf{x}_{k+1})^{-1}$ in the unconstrained optimization model $\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{f}(\mathbf{x})$ [6]. A holistic framework of the QN updates has been attributed to Broyden [3], i.e.

$$\mathbf{H}_{k+1}^{\phi} = \mathbf{H}_k + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{\mathbf{H}_k\mathbf{y}_k\mathbf{y}_k^T\mathbf{H}_k}{\mathbf{y}_k^T\mathbf{H}_k\mathbf{y}_k} + \phi\mathbf{v}_k\mathbf{v}_k^T, \tag{1}$$

in which $\phi \in \mathbb{R}$ is called the Broyden parameter, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k = \alpha_k\mathbf{d}_k$ with the step size α_k resulted from a line search along the direction $\mathbf{d}_k$, and

$$\mathbf{v}_k = \sqrt{\mathbf{y}_k^T\mathbf{H}_k\mathbf{y}_k}\left(\frac{\mathbf{s}_k}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{\mathbf{H}_k\mathbf{y}_k}{\mathbf{y}_k^T\mathbf{H}_k\mathbf{y}_k}\right).$$

*Speaker. Email address: saman.babaiekafaki@unibz.it

It is worthwhile to mention that $\phi = 0$ and $\phi = 1$ in (1) respectively yield the well-known DFP (Davidon–Fletcher–Powell) and BFGS (Broyden–Fletcher–Goldfarb–Shanno) updating formulas [6]. Also, when the line search fulfills the Wolfe conditions [6], $\mathbf{H}_{k+1}^{\phi}$ is well-defined because $\mathbf{s}_k^T\mathbf{y}_k > 0$. Moreover, if $\phi \geq 0$, then $\mathbf{H}_{k+1}^{\phi}$ is positive definite and so, the corresponding search direction is descent.

As seen, the classic form of the Broyden updating formula is dense in the sense that it needs to save the $n \times n$ matrix $\mathbf{H}_k$ to determine $\mathbf{H}_{k+1}^{\phi}$ as the new approximation of the inverse Hessian. Also, it has been known as a matter of fact in the literature that when the eigenvalues of the Hessian at the solution are large but close to each other, then $\mathbf{H}_{k+1}^{\phi}$ can be an ill-conditioned matrix [6]. Thus, in real computations $\kappa(\mathbf{H}_{k+1}^{\phi})$ may be larger that $\kappa(\nabla^2\mathbf{f}(\mathbf{x}^*)^{-1})$, where $\kappa(.)$ and $\mathbf{x}^*$ respectively signify the (spectral) condition number and the optimal solution.

To control the growth of $\kappa(\mathbf{H}_{k+1}^{\phi})$ in contrast to $\kappa(\nabla^2\mathbf{f}(\mathbf{x}^*)^{-1})$, scaled QN updates have been developed by making the eigenvalues of $\mathbf{H}_{k+1}^{\phi}$ well-distributed [3]. The measure has been traditionally taken by the setting $\mathbf{H}_k \leftarrow \theta_k\mathbf{H}_k$ in (1), where the nonnegative parameter θ_k is called the scaling factor. On the other side, to adopt the algorithms for large scale cases, the memoryless versions of $\mathbf{H}_{k+1}^{\phi}$ have been put forward, initially by the simple setting $\mathbf{H}_k \leftarrow \mathbf{I}$. Thus, in aggregate the scaled memoryless QN algorithms have been developed, principally by using the following modified version of (1) with a special setting of the parameter ϕ:

$$\tilde{\mathbf{H}}_{k+1}^{\phi} = \theta_k\mathbf{I} + \frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k} - \theta_k\frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{y}_k} + \phi\tilde{\mathbf{v}}_k\tilde{\mathbf{v}}_k^T,\ \tilde{\mathbf{v}}_k = \sqrt{\theta_k}\|\mathbf{y}_k\|\left(\frac{\mathbf{s}_k}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{\mathbf{y}_k}{\mathbf{y}_k^T\mathbf{y}_k}\right), \tag{2}$$

where $\|.\|$ stands for the Euclidean norm. Hence, the search directions can be calculated by a few vector inner products, not needing significant memory intake. Generally, the (scaled) memoryless QN algorithms are in a close connection with the (scaled) conjugate gradient algorithms [5], another class of efficient tools for unconstrained optimization.

As known, reasonable choices for the scaling factor θ_k have been originally suggested as a result of an analytical spectrum in the framework of the matrix analyses [3]; that is,

$$\theta_k^{\mathrm{OS}} = \frac{\mathbf{s}_k^T\mathbf{y}_k}{\mathbf{y}_k^T\mathbf{y}_k},\ \text{and}\ \theta_k^{\mathrm{OL}} = \frac{\mathbf{s}_k^T\mathbf{s}_k}{\mathbf{s}_k^T\mathbf{y}_k},$$

where OS and OL are respectively abbreviations of Oren–Spedicato and Oren–Luenberger, who devised the above effective formulas for θ_k. Bearing the analysis conducted by Barzilai and Borwein in mind [6], as an inspiring fact it is notable that θ_k^{OS} and θ_k^{OL} can be regarded as scalar approximations of the inverse Hessian. Recently, Babaie–Kafaki [3] established that θ_k^{OS} can be seen as an optimal choice for the scaled memoryless BFGS formula, while θ_k^{OL} can be sort of optimal for the scaled memoryless DFP formula. Here, the main focus is to analyze other optimal choices for the scaling factor θ_k in (2), known as a classic open problem of the literature [1].

2 On optimal choices for the scaling factor

The main source of optimality of the scaling factor in this study is to control the condition number of $\tilde{\mathbf{H}}_{k+1}^{\phi}$ which straightly affects the error bounds of the QN algorithms. The analysis is here carried out on the scaled memoryless BFGS updating formula, known as an effective update of the Broyden family, given by

$$\hat{\mathbf{H}}_{k+1} = \theta_k\mathbf{I} - \theta_k\frac{\mathbf{s}_k\mathbf{y}_k^T + \mathbf{y}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k} + \left(1 + \theta_k\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}\right)\frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k}.$$

Similar discussions can be presented for the other members of the Broyden family of QN updating formulas. In what follows, we assume that the Wolfe line search conditions hold [6], and so, we have $\mathbf{s}_k^T\mathbf{y}_k > 0$.

As the core of our analytical efforts, principally we need to determine the eigenvalues of $\hat{\mathbf{H}}_{k+1}$. By this issue in the forefront, since $\mathbf{s}_k$ and $\mathbf{y}_k$ are nonzero vectors, there exists mutually orthonormal vectors $\mathbf{p}_k^1, \mathbf{p}_k^2, ..., \mathbf{p}_k^{n-2}$ in $\mathbb{R}^n$ such that $\mathbf{s}_k^T\mathbf{p}_k^i = \mathbf{y}_k^T\mathbf{p}_k^i = 0$. As a result,

$$\hat{\mathbf{H}}_{k+1}\mathbf{p}_k^i = \theta_k\mathbf{p}_k^i,\ \ i = 1, 2, \ldots, n-2.$$

Thus, θ_k is the eigenvalue of $\hat{\mathbf{H}}_{k+1}$ with the multiplicity $n-2$. Also, after some algebraic calculations, the two remaining eigenvalues of $\hat{\mathbf{H}}_{k+1}$ can be given by

$$\mu_k^{\pm} = \frac{1}{2}\left(1 + \theta_k\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}\right)\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} \pm \frac{1}{2}\sqrt{\left(1 + \theta_k\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}\right)^2\frac{\|\mathbf{s}_k\|^4}{(\mathbf{s}_k^T\mathbf{y}_k)^2} - 4\theta_k\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}},$$

for which $0 < \mu_k^- \leq \theta_k \leq \mu_k^+$ [3]. Now, since $\kappa(\hat{\mathbf{H}}_{k+1}) = \dfrac{\mu_k^+}{\mu_k^-}$, we feel a meaningful need to make the two border eigenvalues μ_k^- and μ_k^+ close to each other as much as possible. This yields

$$\theta_k^{\mathrm{E}} = \frac{\mathbf{s}_k^T\mathbf{y}_k}{\|\mathbf{y}_k\|^2}\left(\frac{2(\mathbf{s}_k^T\mathbf{y}_k)^2}{\|\mathbf{s}_k\|^2\|\mathbf{y}_k\|^2} - 1\right),$$

which requires further modification to ensure positivity [4].

Now, we move from the Euclidean norm toward the nonsmooth ℓ_∞ norm. That is, based on the norm consistency relations [6], we plan to analyze

$$\kappa_\infty(\hat{\mathbf{H}}_{k+1}) = \|\hat{\mathbf{H}}_{k+1}\|_\infty\|\hat{\mathbf{H}}_{k+1}^{-1}\|_\infty.$$

To proceed, firstly note that

$$\hat{\mathbf{H}}_{k+1}^{-1} = \frac{1}{\theta_k}\mathbf{I} - \frac{1}{\theta_k}\frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{s}_k} + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{s}_k^T\mathbf{y}_k}.$$

Also, it can be observed that

$$\|\hat{\mathbf{H}}_{k+1}\|_\infty \leq \theta_k + \theta_k\frac{\|\mathbf{s}_k\|_\infty\|\mathbf{y}_k\|_1}{\mathbf{s}_k^T\mathbf{y}_k} + \theta_k\frac{\|\mathbf{y}_k\|_\infty\|\mathbf{s}_k\|_1}{\mathbf{s}_k^T\mathbf{y}_k} + \left(1 + \theta_k\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}\right)\frac{\|\mathbf{s}_k\|_\infty\|\mathbf{s}_k\|_1}{\mathbf{s}_k^T\mathbf{y}_k},$$

and

$$\|\hat{\mathbf{H}}_{k+1}^{-1}\|_\infty \leq \frac{1}{\theta_k} + \frac{1}{\theta_k}\frac{\|\mathbf{s}_k\|_\infty\|\mathbf{s}_k\|_1}{\|\mathbf{s}_k\|^2} + \frac{\|\mathbf{y}_k\|_\infty\|\mathbf{y}_k\|_1}{\mathbf{s}_k^T\mathbf{y}_k}.$$

Thus, an upper bound for $\kappa_\infty(\hat{\mathbf{H}}_{k+1})$ is at hand which by its minimization, we can obtain

$$\theta_k^{\ell_\infty} = \sqrt{\frac{\|\mathbf{s}_k\|(\mathbf{s}_k^T\mathbf{y}_k)^2}{2\|\mathbf{y}_k\|^3(\mathbf{s}_k^T\mathbf{y}_k) + \|\mathbf{s}_k\|\|\mathbf{y}_k\|^4}},$$

as a result of performing some relaxations when $n \to \infty$ [2]. Moreover, in an extension scheme, we can suggest the following hybrid formula for the scaling factor:

$$\theta_k^{\mathrm{H}} = \sqrt{\frac{\|\mathbf{s}_k\|(\mathbf{s}_k^T\mathbf{y}_k)^2}{2\rho\|\mathbf{y}_k\|^3(\mathbf{s}_k^T\mathbf{y}_k) + \|\mathbf{s}_k\|\|\mathbf{y}_k\|^4}},$$

with the real constant ρ, where for $\rho = 0$ and $\rho = 1$ respectively reduces to θ_k^{OS} and $\theta_k^{\ell_\infty}$. Since the matrix $\hat{\mathbf{H}}_{k+1}$ is symmetric, similar results can be obtained by analyzing its ℓ_1-norm condition number.

In the final part of our study, considering the relationship between the formulas θ_k^{OS} and θ_k^{OL} which originates from the dual relationship between the BFGS and DFP updating formulas, we propose a new choice for the scaling factor based upon the framework of θ_k^{H} as follows:

$$\theta_k^{\mathrm{N}} = \sqrt{\frac{2\xi\|\mathbf{s}_k\|^3(\mathbf{s}_k^T\mathbf{y}_k) + \|\mathbf{y}_k\|\|\mathbf{s}_k\|^4}{\|\mathbf{y}_k\|(\mathbf{s}_k^T\mathbf{y}_k)^2}},$$

where ξ is a real constant. It can be seen that if $\xi = 0$, then θ_k^{N} reduces to θ_k^{OL}.

3 Conclusions

Analyzing the condition number of the well-known scaled memoryless BFGS updating formula has been targeted here. In this context, firstly the distribution of the eigenvalues of the updating matrix has been improved and then, an upper bound of the ℓ_∞-norm condition number of the matrix has been approximately minimized. It has been also briefly shown that how adaptive, sort of optimal choices for the scaling factor can be obtained as a result of the analyses. Finally, a new hybrid choice for the scaling factor has been suggested, based on the structure of a recent hybrid formula of the scaling factor as well as the dual relationship between the BFGS and DFP updating formulas.

In the computational standpoint, preliminary implementations showed that the given formulas for the scaling factor are capable to yield promising outputs. Especially, the two last (hybrid) one-parameter formulas require more numerical investigations to determine the proper values of their inner parameters. Recently, as an extension of the scalar scaling, the diagonal scaling of the QN updates has been put forward. So, generally the sparse (matrix) scaling for the QN updates can be an important subject of research, being helpful for the high dimensional data analysis as well. It should be mainly noted that the simplicity of the scaling formula can be regarded as the greatest need for the large scale problems.

References

[1] N. Andrei, Open problems in conjugate gradient algorithms for unconstrained optimization, *B. Malays. Math. Sci. So.*, 34 (2011), No. 2, 319–330.

[2] S. Babaie–Kafaki, A hybrid scaling parameter for the scaled memoryless BFGS method based on the ℓ_∞ matrix norm, *Int. J. Comput. Math.*, 96 (2019), No. 8, 1595–1602.

[3] S. Babaie–Kafaki, On optimality of the parameters of self-scaling memoryless quasi–Newton updating formulae, *J. Optim. Theory Appl.*, 167 (2015), No. 1, 91–101.

[4] S. Babaie–Kafaki, A modified scaling parameter for the memoryless BFGS updating formula, *Numer. Algorithms*, 72 (2016), No. 2, 425–433.

[5] S. Babaie–Kafaki, A survey on the Dai–Liao family of nonlinear conjugate gradient methods, *RAIRO–Oper. Res.*, 57 (2023), No. 1, 43–58.

[6] W. Sun and Y.X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer–Verlage, New York, 2006.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Hybrid scalarization technique for solving multiobjective quadratically constrained quadratic programming

Hossein Salmei*

Department of Mathematics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran

Abstract

In this paper, the Hybrid scalarization technique is exploited for solving multiobjective quadratically constrained quadratic programming problems with (non)convex quadratic function. To this end, a linear programming relaxation is derived that computes a lower bound on the optimal objective value of the scalarization problem. Basically, the proposed algorithm aims to find efficient solutions to the problem by solving the linear relaxation sequentially on the subsets of the feasible region.

Keywords: Multiobjective programming, quadratic programming, Linear relaxation, Convex and concave envelopes.

Mathematics Subject Classification [2010]: 15B48, 90C20, 90C26.

1 Introduction

Consider the following multiobjective quadratically constrained quadratic programming (MQCQP) problem of the form

$$\begin{array}{lll} \min f(x) = (f_1(x), ..., f_p(x)) & & \\ s.t. & f_k(x) \geqslant 0, & k = p+1, ..., m, \\ & x \in [a, b], & \end{array} \tag{1}$$

where $a, b \in \mathbb{R}^n_{\leqq}$ and $x \in [a, b]$ means that $a_i \leqslant x_i \leqslant b_i$ for all $i = 1, ..., n$. Each $f_k : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function in the form of

$$f_k(x) = x^t H^k x + c_k^t x + d_k, \qquad k = 1, ..., p, \tag{2}$$

where $H^1, ..., H^p$ are real and symmetric $n \times n$ matrixes, $c_1, ..., c_p \in \mathbb{R}^n$ and $d_1, ..., d_p \in \mathbb{R}$.

If problem (1) involves a single objective and we use the term SQCQP instead of MQCQP.

Quadratic programming with quadratic constraints is an important and well known technique for formulating and dealing with various mathematical programming problems (see, for example [3, 6]).

Throughout the paper, $\mathbb{R}^n$ denotes the n dimensional Euclidean space and the feasible set of problem (1) is specified by $X = \{x \in \mathbb{R}^n | f_k(x) \geqslant 0, k = p+1, ..., m, \ x \in [a, b]\}$. If $x, y \in \mathbb{R}^n$ then $x \leqq y (x < y)$ if and only if $x_i \leqslant y_i (x_i < y_i), \forall i = 1, ..., n$. In addition, $x \leq y$ means that $x \leqq y$ and $x \neq y$. We will denote by $\mathbb{R}^n_{\leqq}$ the set $\{x \in \mathbb{R}^n | x \geqq 0\}$.

*Speaker. Email address: salmei@vru.ac.ir

Definition 1.1. ([2]) Consider an MQCQP problem. The feasible solution $\hat{x} \in X$ is called efficient (weak efficient) if there is no another $x \in X$ such that $f(x) \leq f(\hat{x})(f(x) < f(\hat{x}))$. If $\hat{x} \in X$ is efficient (weak efficient) then $\hat{y} = f(\hat{x})$ is called a nondominated (weak nondominated) point. The set of all efficient solutions and nondominated points are called the efficient set and efficient frontier, respectively

The sets of weakly efficient solutions and efficient solutions are denoted by X_{wE} and X_E, respectively.

Definition 1.2. ([6]) A symmetric $n \times n$ matrix H is called

- Positive definite if and only if $x^t Hx > 0$ for all $x \in \mathbb{R}^n$ and $x \neq 0$.
- Positive semidefinite if and only if $x^t Hx \geqslant 0$ for all $x \in \mathbb{R}^n$.

Proposition 1.3. *([6]) Let C be a convex subset of $\mathbb{R}^n$ and let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be twice continuously differentiable over $\mathbb{R}^n$.*

- *If $\bigtriangledown^2 f(x)$ (Hessian of f) is positive semidefinite for all $x \in C$, then f is convex over C.*
- *If $\bigtriangledown^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex over C.*

Corollary 1.4. *Consider the quadratic function $f(x) = x^t Hx + c^t x + d$, where H is a symmetric $n \times n$ matrix, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$. Then, f is convex if the Hessian matrix H is positive semidefinite. Moreover, f is strictly convex if H is positive definite.*

Definition 1.5. ([6]) A function $h : \mathbb{R}^n_{\leqq} \longrightarrow \mathbb{R}$ is called an increasing function if $h(x) \leqslant h(y)$ for $x \leqq y$. It is a *d.m* (difference of monotonic) function if $h(x) = h^+(x) - h^-(x)$, where h^+ and h^- are increasing functions.

Remark 1.6. Each quadratic function can be represented as a difference of two quadratic functions with nonnegative coefficients. So, every quadratic function is a *d.m* function.

Definition 1.7. ([5]) Let X be a convex and compact subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$. The convex envelop of the function f over X is denoted by $Vex_X f$ and for all $x \in X$ is defined as

$$Vex_X f(x) = \sup\{g(x) : \text{g is convex on X}, \ g(y) \leqslant f(y), \quad \forall y \in X\}$$

Definition 1.8. ([5]) Let X be a convex and compact subset of $\mathbb{R}^n$ and $f : X \to \mathbb{R}$. The concave envelop of the function f over X is denoted by $Cav_X f$ and for all $x \in X$ is defined as

$$Cav_X f(x) = \inf\{g(x) : \text{g is convex on X}, \ f(y) \leqslant g(y), \quad \forall y \in X\}$$

Theorem 1.9. *([1]) The convex envelop and concave envelop of the two dimensional bilinear function $f(x, y) = xy$ on the hyperrectangle $R = \{(x, y) \in \mathbb{R}^2 : \ell \leqslant x \leqslant u, m \leqslant y \leqslant M\}$ are respectively*

$$Vex_R(xy) = \max\{\ell y + mx - \ell m, uy + Mx - uM\},$$

$$Cav_R(xy) = \min\{\ell y + Mx - \ell M, uy + mx - um\}.$$

Definition 1.10. ([2]) Let $\hat{x}$ be an arbitrary feasible point of the MQCQP (1). The Hybride method is defined as follows

$$\begin{aligned} \min \quad & \sum_{k=1}^{p} \lambda_k f_k(x) \\ & f_k(x) \leqslant f_k(\hat{x}), \qquad k = 1, ..., p, \\ & x \in X, \end{aligned} \tag{3}$$

where $\lambda \in \mathbb{R}^p_{\geq}$.

Theorem 1.11. *([2]) Let $\lambda \in \mathbb{R}^p_{\geq}$. A feasible point $\hat{x} \in X$ is an efficient solution of MQCQP (1) if and only if $\hat{x}$ is an optimal solution of problem (3).*

2 Main results

Assume that $\lambda \in \mathbb{R}^p_{\geq}$ and $\hat{x}$ be a feasible solution of the MQCQP (1). According to Theorem 1.11, $\hat{x}$ is an efficient solution of MQCQP (1) if and only if $\hat{x}$ is an optimal solution of the scalarization problem

$$\begin{aligned} \min \quad & \sum_{k=1}^{p} \lambda_k f_k(x) \\ s.t. \quad & f_k(x) \leqslant f_k(\hat{x}), \quad k = 1, ..., p, \\ & f_k(x) \geqslant 0, \qquad k = p+1, ..., m, \\ & x \in [a, b], \end{aligned} \tag{4}$$

An approach to find approximate solutions of the SQCQP (4) is to solve a linear relaxation of this problem. Here, we use a linear relaxation of problem (4), which is based on the convex and concave envelops of the bilinear terms in the quadratic functions $f_k(x)$. We denote this linear relaxation by $LP(a, b, \hat{x})$.

$$\begin{aligned} \min \quad & \sum_{k=1}^{p} \lambda_k \Big(\sum_{j=1}^{p} t_j^k + c_k^t x + d_k \Big) \\ s.t. \quad & t_j^k \geqslant a_j H_j^k x + m_j^k x_j - a_j m_j^k, \quad j = 1, ..., n, \quad k = 1, ..., p, \\ & t_j^k \geqslant b_j H_j^k x + M_j^k x_j - b_j M_j^k, \quad j = 1, ..., n, \quad k = 1, ..., p, \\ & t_j^k \leqslant a_j H_j^k x + M_j^k x_j - a_j M_j^k, \quad j = 1, ..., n, \quad k = p+1, ..., m, \\ & t_j^k \leqslant b_j H_j^k x + m_j^k x_j - b_j m_j^k, \quad j = 1, ..., n, \quad k = p+1, ..., m, \\ & \textstyle\sum_{j=1}^{n} t_j^k + c_k^t x + d_k \leqslant f_k(\hat{x}), \quad k = 1, ..., p, \\ & \textstyle\sum_{j=1}^{n} t_j^k + c_k^t x + d_k \geqslant 0, \qquad k = p+1, ..., m, \\ & x \in [a, b], \end{aligned} \tag{5}$$

where t_j^k is the corresponding variable to the convex (concave) envelop of the bilinear function $x_j y_j^k$ such that $y_j^k = H_j^k x$ and H_j^k is the j-th row of the matrix H^k. Also m_j^k and M_j^k are the minimum and maximum of the linear function $H_j^k x$ on the interval $[a, b]$, respectively. Therefore,

$$m_j^k = \min\{H_j^k x : x \in [a, b]\} = \sum_{q=1}^{n} \min\{H_{jq}^k a_q, H_{jq}^k b_q\},$$

$$M_j^k = \max\{H_j^k x : x \in [a,b]\} = \sum_{q=1}^{n} \max\{H_{jq}^k a_q, H_{jq}^k b_q\},$$

wherein, H_{jq}^k is element (j,q) of the matrix H^k ([1]).
The next theorem shows that an optimal objective value of the linear programming problem (5) is a lower bound to an optimal objective value of the quadratic programming problem (4).

Theorem 2.1. *Assume $(x^\star, t^{\star 1}_1, ..., t^{\star m}_n)$ be the optimal solution of the linear problem (5) and $\bar{x}$ be the optimal solution of the quadratic problem (4). Then*

$$\sum_{k=1}^{p} \lambda_k \Big(\sum_{j=1}^{p} t^{\star k}_j + c_k^t x^\star + d_k \Big) \leqslant \sum_{i=1}^{k} \lambda_k f_k(\bar{x})$$

2.1 Proposed Algorithm

In the following, we propose an algorithm to solve problem (1) when $[a,b] \subseteq \mathbb{R}^n_{\leqq}$. At first, we divide the cell $[a,b]$ into smaller subcells. Then, for each subcells, we solve the linear problem (5) to find a set of approximate (weakly) efficient solutions of the quadratic problem (1). By repeating this procedure and removing the non efficient solutions of this set at each iteration of the algorithm, we will have a better approximation of the efficient solutions set of problem (1).

Algorithm 5.

- **Input.** $f = (f_1, ..., f_p), \lambda = (\lambda_1, ..., \lambda_p) \in \mathbb{R}^p_{\leqq}, a, b \in \mathbb{R}^n_{\leqq}$, positive integer m and positive real number Δ.

1. $t := 1, [a^t, b^t] := [a,b], \mathcal{X}_E^{t-1} := \emptyset, \mathcal{A} := \emptyset$.

2. Divide cell $[a^t, b^t]$ into $(tm)^2$ subcell $[l^t, u^t]$ such that

$$l^t = (a_1^t + (i_1 - 1)s_1^t,\ a_2^t + (i_2 - 1)s_2^t, ..., a_n^t + (i_n - 1)s_n^t),$$

$$u^t = (a_1^t + i_1 s_1^t,\ a_2^t + i_2 s_2^t, ..., a_n^t + i_n s_n^t), \qquad i_1, ..., i_n = 1, ..., tm,$$

$$s_r^t := \frac{b_r^t - a_r^t}{tm}, \qquad r = 1, 2, ..., n.$$

3. For each subcell $[l^t, u^t]$ solve the linear problem (5), where $[a,b] = [a^t, b^t]$ and $\hat{x}$ is an arbitrary point in $[l^t, u^t]$. Set $\mathcal{A} := \mathcal{A} \cup \{\bar{x}\}$, where $(\bar{x}, \bar{t}^1_1, ..., \bar{t}^m_n)$ is the optimal solution of (5).

4. Construct the set $\mathcal{X}_E^t$ which is obtained by removing the non efficient points of problem (1) from $\mathcal{X}_E^{t-1} \cup \mathcal{A}$. Set, $\mathcal{X}_E := \mathcal{X}_E^t$.

5. If $\frac{\|b^t - a^t\|}{tm} > \Delta$
   Set $t := t+1, a^t := (a_1^t, a_2^t, ..., a_n^t)$ and $b^t := (b_1^t, b_2^t, ..., b_n^t)$ where $a_i^t = \min_{\bar{x} \in \mathcal{X}_E^t} \bar{x}_i$ and $b_i^t = \max_{\bar{x} \in \mathcal{X}_E^t} \bar{x}_i$, for $i = 1, 2, ..., n$ then goto Step 1, else stop.

6. end if

- **Output.** The sets $\mathcal{X}_E$ and $\mathcal{Y}_E := f(\mathcal{X}_E)$ as a discrete approximations of efficient set and efficient frointer set of problem (1), respectively.

Theorem 2.2. *For each $\Delta > 0$, Algorithm 2.1 terminates after a finite number of iterations.*

In the following an example from [3] is considered with the proposed algorithm.

Example 2.3. Consider the following biobjective quadratic programming problem

$$\begin{array}{ll} \min & (f_1(x), f_2(x)) \\ s.t. & -2x_1 - x_2 + 3 \leqslant 0, \\ & -x_1 - 2x_2 + 3 \leqslant 0, \\ & -2x_1 + 3x_2 - 3 \leqslant 0, \\ & x \in [(0.5, 0.5), (3, 3)], \end{array}$$

where $f_1(x) = 0.5(5x_1^2 + x_2^2)$ and $f_2(x) = 0.5(x_1^2 + 5x_2^2)$. By [3], the efficient set is two line segments between the points of $\{(\frac{3}{4}, \frac{3}{2}), (1, 1)\}$ and $\{(1, 1), (\frac{5}{3}, \frac{2}{3})\}$. Figures 1 and 2 show the output of Algorithm 5 with $a = (0.5, 0.5), b = (3, 3), m = 5, \Delta = 0.05$, and $(\lambda_1, \lambda_2) = (0.4, 0.6)$ in feasible and objective spaces, respectively.

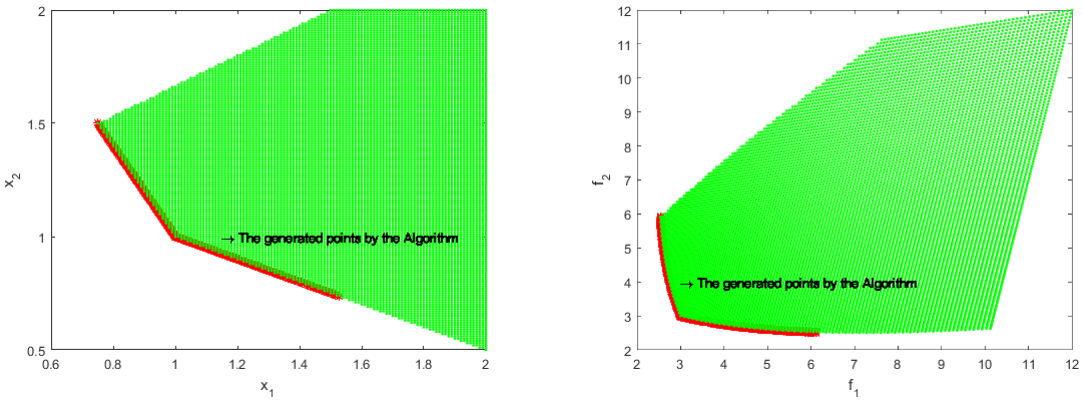


Figure 1: The sets $\mathcal{X}_E$ and $\mathcal{Y}_E$ for example 2.3.

3 Conclusion

In this paper, the hybrid scalarization technique is used to solve multiobjective quadratically constrained quadratic programming (MQCQP) problem. While most of existing method for solving MQCQPs consider convex objective functions and linear constraints, the hybrid method able to solve this problems with (non)convex quadratic functions and

constraints. In fact, the proposed scalarization converts MQCQP to an SQCQP. A linear relaxation of SQCQP is extracted which it's optimal objective value is a lower bound of the optimal objective value of SQCQP on a given box. Basically, The study of the solutions of the MQCQP is based on solving LPRs successively over smaller hyperrectangles of $[a, b] \subseteq \mathbb{R}^n_{\leqq}$. The proposed algorithm implement this procedure.

References

[1] F. A. Al-Khayyal, C. Larsen and T. Van Voorhis, A relaxation method for nonconvex quadratically constrained quadratic programs,*Journal of Global Optimization*, 6 (1995), 215–230.

[2] M. Ehrgott, *Multicriteria Optimization*, Springer, Berlin, 2005.

[3] C. J. Goh and X. Q. Yang, Analytic efficient solution set for multi-criteria quadratic programs, *European Journal of Operational Research*, 92 (1996), 166–181.

[4] N. Rastegar and E. khorram, A combined scalarizing method for multiobjective programming problems, *European Journal of Operational Research*, 236 (2014), 229–237.

[5] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[6] H. Tuy, *Convex Analysis and Global Optimization*, Springer, Berlin, 2016.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Geometric optimization via system of fuzzy relation inequalities

Mahdi Keshtkar[1,*] and Elyas Shivanian[2]

[1]Department of Mathematics, Buein Zahra Technical University, Buein Zahra, Qazvin, Iran

[2]Department of Mathematics, Faculty of Science, Imam Khomeini International University, Qazvin 34194-288, Iran

Abstract

In this paper, an optimization model with geometric objective function is presented. Regarding this matter, we present geometric programming model with a monomial objective function subject to the fuzzy relation inequalities constraints with max-product composition. Simplification operations have been given to accelerate the resolution of the problem by removing the components having no effect on the solution process. Also, an algorithm is presented to abbreviate the problem resolution.

Keywords: Geometric programming, Fuzzy relation inequalities, Max-product composition.

Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36

1 Introduction

The fundamental result for fuzzy relation equations with max-product composition goes back to Pedrycz [5]. The problem of optimization subject to FRE and FRI is one of the most interesting and on-going research topic among the problems related to FRE and FRI theory [1]. They extended the study of an inverse solution of a system of fuzzy relation equations with max-product composition. They provided theoretical results for determining the complete sets of solutions as well as the conditions for the existence of resolutions. Their results showed that such complete sets of solutions can be characterized by one maximum solution and a number of minimal solutions. A problem of optimization was studied by Loetamonfong and Fang with max-product composition [4] which was improved by Guu and Wu by shrinking the search region [3]. Also, Guo and Xia presented an algorithm to accelerate the resolution of this problem [2]. In view of the importance of geometric programming and the fuzzy relation equation in theory and applications, Yang and Cao have proposed a fuzzy relation geometric programming, discussed optimal solutions with two kinds of objective functions based on fuzzy max product operator [6].

*Speaker. Email address: keshtkarmahdi@gmail.com

In this paper, we generalize the geometric programming of the FRE with the max-product operator [6] by considering the fuzzy relation inequalities instead of the equations in the constraints. This problem can be formulated as follows:

$$\begin{array}{rcll} \min Z & = & \max\limits_{j=1,2,3,...,n} c_j.x_j^{\alpha_j} & \\ \text{s.t.} & & A \bullet x \geq d^1 & \\ & & B \bullet x \leq d^2 & \\ & & x \in [0,1]^n & \end{array} \tag{1}$$

Where $c_j, \alpha_j \in R$,$c_j \geq 0$ and $A = (a_{ij})_{m\times n}$,$a_{ij} \in [0,1]$, $B = (b_{ij})_{l\times n}$, $b_{ij} \in [0,1]$, are fuzzy matrices, $d^1 = (d_i^1)_{m\times 1} \in [0,1]^m$,$d^2 = (d_i^2)_{l\times 1} \in [0,1]^l$ are fuzzy vectors, $c = (c_j)_{n\times 1} \in R^n$ is the vector of cost coefficients, and $x = (x_j)_{n\times 1} \in [0,1]^n$ is an unknown vector, and "$\bullet$" denotes the fuzzy max-product operator as defined below. Problem (1) can be rewritten as the following problem in detail:

$$\begin{array}{rcll} \min Z & = & \max\limits_{j\in J} c_j.x_j^{\alpha_j} & \\ \text{s.t.} & & a_i \bullet x \geq d_i^1, & i \in I^1 = \{1,2,...,m\} \\ & & b_i \bullet x \leq d_i^2, & i \in I^2 = \{1,2,...,l\} \\ & & 0 \leq x_j \leq 1, & j \in J = \{1,2,...,n\} \end{array} \tag{2}$$

where a_i and b_i are the ith row of the matrices A and$B$, respectively, and the constraints are expressed by the max-product operator definition as:

$$\begin{array}{ll} a_i \bullet x = \max\limits_{j\in J}\{a_{ij} \cdot x_j\} \geq d_i^1 & \forall i \in I^1 \\ b_i \bullet x = \max\limits_{j\in J}\{b_{ij} \cdot x_j\} \leq d_i^2 & \forall i \in I^2 \end{array} \tag{3}$$

2 The characteristics of the set of feasible solution

Lemma 2.1. *(a) $S(A, d^1) \neq \phi$ if and only if for each $i \in I^1$ there exists some $j_i \in J$ such that $a_{ij_i} \geq d_i^1$.*
(b) If $S(A, d^1) \neq \phi$ then $\overline{1} = [1, 1, ..., 1]_{1\times n}^t$ is the greatest element in set $S(A, d^1)$.

Theorem 2.2. *If $S(A, B, d^1, d^2) \neq \phi$, then for each $i \in I^1$ there exist $j \in J$ such that $a_{ij} \geq d_i^1$.*

Definition 2.3. Set $\overline{x} = (\overline{x}_j)_{n\times 1}$ where

$$\overline{x}_j = \begin{cases} 1 & \forall i : b_{ij} \leq d_i^2 \\ \min\limits_{i=1,...l}\left\{\dfrac{d_i^2}{b_{ij}} : \ b_{ij} > d_i^2\right\} & \text{otherwise} \end{cases}$$

Definition 2.4. Let $J_i = \{j \in J : a_{ij} \geq d_i^1\}$, $\forall i \in I^1$. For each $j \in J_i$, we define $i_{x(j)} = (i_{x(j)_k})_{n\times 1}$ such that

$$i_{x(j)_k} = \begin{cases} \frac{d_i^1}{a_{ij}} & k = j \\ 0 & k \neq j \end{cases}$$

Definition 2.5. Let $e = (e(1), e(2), ...e(m)) \in J_1 \times J_2 \times ... \times J_m$ such that $e(i) = j \in J_i$.We define$x(e) = (x(e)_j)_{n\times 1}$, in which $x(e)_j = \max\limits_{i\in I_j^e}\{i_{x(e(i))j}\} = \max\limits_{i\in I_j^e}\left\{\frac{d_i^1}{a_{ij}}\right\}$ if $I_j^e \neq \phi$ and $x(e)_j = 0$ if$I_j^e = \phi$, where$I_j^e = \{i \in I^1 : e(i) = j\}$.

Corollary 2.6. *(a) If $d_i^1 = 0$ for some $i \in I^1$, then we can remove the ith row of matrix A with no effect on the calculation of the vectors $x(e)$ for each $e \in J_I = J_1 \times J_2 \times ... \times J_m$. (b) If $j \notin J_i$, $\forall i \in I^1$, then we can remove the jth column of the matrix A before calculating the vectors $x(e)$, $\forall e \in J_I$ and set $x(e)_j = 0$ for each $e \in J_I$*

Theorem 2.7. *If $S(A, B, d^1, d^2) \neq \phi$, then $S(A, B, d^1, d^2) = \bigcup\limits_{X_0(e)} [x(e), \overline{x}]$.*

3 Simplification operations and the resolution algorithm

In order to solve problem (1), we first convert it into the two sub-problems below:

$$\begin{array}{lll} \min Z & = & \max\limits_{j\in R^+} c_j.x_j^{\alpha_j} \\ \text{s.t.} & & A \bullet x \geq d^1 \quad (4a) \\ & & B \bullet x \leq d^2 \\ & & x \in [0,1]^n \end{array} \qquad \begin{array}{lll} \min Z & = & \max\limits_{j\in R^-} c_j.x_j^{\alpha_j} \\ \text{s.t.} & & A \bullet x \geq d^1 \quad (4b) \\ & & B \bullet x \leq d^2 \\ & & x \in [0,1]^n \end{array}$$

where $R^+ = \{j \,|\alpha_j \geq 0, j \in J\}$ and $R^- = \{j \,|\alpha_j < 0, j \in J\}$.

Theorem 3.1. *Assume that $x(e_0)$ be an optimal solution of problem (4a) (it is possible that don't be unique) then, the optimal solution of problem (1) is x^* that defined as follow:*

$$x_j^* = \begin{cases} \overline{x}_j & j \in R^- \\ x(e_0)_j & j \in R^+ \end{cases}$$

Theorem 3.2. *The set of feasible solutions for problem (1), namely$S(A, B, d^1, d^2)$, is nonempty if and only if for each $i \in I^1$ set $\overline{J}_i = \left\{j \in J_i : \frac{d_i^1}{a_{ij}} \leq \overline{x}_j\right\}$ is nonempty.*

Theorem 3.3. *If$S(A, B, d^1, d^2) \neq \phi$,then$S(A, B, d^1, d^2) = \bigcup\limits_{\overline{X}(e)} [x(e), \overline{x}]$where$\overline{X}(e) = \{x(e) :$ $e \in \overline{J}_I = \overline{J}_1 \times \overline{J}_2 \times ... \times \overline{J}_m\}$.*

Definition 3.4. We define $J_i^* = \{\quad j : \quad j \in R^- \quad \text{and} \quad j \in \overline{J}_i \ \}$ for $i \in I^1$.

Theorem 3.5. *Suppose $x(e_0)$ is the optimal solution in (4a) and $J_{i'}^* \neq \phi$ for some $i' \in I^1$, then there exist $x(e')$ such that $e'(i') \in J_{i'}^*$, and also $x(e')$ is the optimal solution in (4a).*

Corollary 3.6. *If $J_i^* \neq \phi$ for some $i \in I^1$ then, we can remove the ith row of matrix A without any effect on finding the optimal solution of problem (4a).*

Definition 3.7. Let $j_1, j_2 \in J$, $\alpha_{j_1} > 0$ and $\alpha_{j_2} > 0$. We say j_2 dominates j_1 if and only if
(a) $j_1 \in \overline{J}_i$ implies $j_2 \in \overline{J}_i$, $\forall i \in I^1$.
(b) For each $i \in I^1$such that$j_1 \in \overline{J}_i$we have $c_{j_1}.(\frac{d_i^1}{a_{ij_1}})^{\alpha_{j_1}} \geq c_{j_2}.(\frac{d_i^1}{a_{ij_2}})^{\alpha_{j_2}}$.

Theorem 3.8. *Suppose $x(e_0)$ is the optimal solution in (4a) and j_2 dominates j_1 for$j_1, j_2 \in R^+$, then there exists $x(e')$ such that $I_{j_1}^{e'} = \phi$, and also $x(e')$ is the optimal solution in (4a). (Notification: $\alpha_{j_1} > 0$ and $\alpha_{j_2} > 0$)*

Corollary 3.9. *If j_2 dominates j_1 for some $j_1, j_2 \in R^+$, then we can remove the j_1th column of the matrix A without any effect on finding the optimal solution $x(e_0)$ in (4a).*

4 Algorithm for finding an optimal solution

Definition 4.1. Consider problem (1). We call $\overline{A} = \overline{(a_{ij})}_{m\times n}$ and $\overline{B} = (\overline{b}_{ij})_{l\times n}$the characteristic matrices of matrix$A$ and matrix B, respectively, where $\overline{a}_{ij} = \frac{d_i^1}{a_{ij}}$ for each $i \in I^1$and$j \in J$, also $\overline{b}_{ij} = \frac{d_i^2}{b_{ij}}$ for each $i \in I^2$ and $j \in J$. (set $\frac{0}{0} = 1$ and $\frac{k}{0} = \infty$)

Algorithm 4.2. Given problem (2),

1. Find matrices $\overline{A}$ and $\overline{B}$.
2. If there exists $i \in I^1$ such that $\overline{a}_{ij} > 1$,$\forall j \in J$, then stop. Problem 2 is infeasible.
3. Calculate $\overline{x}$ from $\overline{B}$.
4. If there exists $i \in I^1$ such that $d_i^1 = 0$, then remove the i'th row of matrix $\overline{A}$.
5. If $\overline{a}_{ij} > \overline{x}_j$, then set $\overline{a}_{ij} = 0$,$\forall i \in I^1$ and $\forall j \in J$.
6. If there exists $i \in I^1$ such that $\overline{a}_{ij} = 0$,$\forall j \in J$, then stop. Problem (2) is infeasible.
7. If there exists$j' \in J$ such that $\overline{a}_{ij'} = 0$,$\forall i \in I^1$, then remove the j'th column of the matrix $\overline{A}$ and set $x(e_0)_{j'} = 0$.
8. For each $i \in I^1$, if $J_i^* \neq \phi$ then remove the ith row of the matrix $\overline{A}$.
9. Remove each column $j \in J$ from $\overline{A}$ such that $j \in R^-$ and set $x(e_0)_j = 0$.
10. If j_2 dominates j_1, $(j_1, j_2 \in R^+)$ then remove column j_1 from $\overline{A}$, $\forall j_1, j_2 \in J$ and set $x(e_0)_{j_1} = 0$.
11. Let $J_i^{new} = \{j \in \overline{J}_i : \overline{a}_{ij} \neq 0\}$ and $J_I^{new} = J_1^{new} \times J_2^{new} \times ... \times J_m^{new}$. Find the vectors $x(e)$, $\forall e \in J_I^{new}$.
12. Find x^*.

5 Conclusion

In this paper, we studied the geometric programming with fuzzy relational inequalities constraints defined by the max-product operator. Since the difficulty of this problem is finding the minimal solutions optimizing the same problem with the objective function $\max_{j\in R^+}\{c_j.{x_j}^{\alpha_j}\}$, we presented an algorithm together with some simplification operations to accelerate the problem resolution.

References

[1] A. Ghodousian, A. Babalhavaeji, An efficient genetic algorithm for solving nonlinear optimization problems defined with fuzzy relational equations and max-Lukasiewicz composition, Applied Soft Computing 69 (2018) 475–492.

[2] F. F. Guo, Z. Q. Xia, An algorithm for solving optimization Problems with one linear objective Function and Finitely Many Constraints of Fuzzy Relation Inequalities, Fuzzy optimization and Decision making 5, 33-47, 2006

[3] S.M. Guu, and Y. K. Wu, Minimizing a Linear Objective Function with Fuzzy Relation Equation Constraints, Fuzzy Optimization and Decision Making 12, (2002) 1568–4539.

[4] J. Loetamonphong, and S.-C. Fang, Optimization of Fuzzy Relation Equations with Max-product Composition Fuzzy Sets and Systems 118, (2001) 509–517

[5] W. Pedrycz, On Generalized fuzzy relational equations and their applications, Journal of Mathematical Analysis and Applications 107 (1985), 520-536.

[6] J. H. Yang, & B. Y. Cao, Geometric programming with max-product fuzzy relation equation constraints. Proceedings of the 24th North American Fuzzy Information Processing Society, Ann Arbor, Michigan, June 22–25, 650–653 (2005b).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Entropy for h-convex functions

Ali Morassaei* and Maryam Samadi

Department of Mathematics, Faculty of Sciences, University of Zanjan, University Blvd., Zanjan 45371-38791, Iran

Abstract

In this paper, we present numerical form for generalization entropy involve with h-convex functions and propound some results for them. So, we state operator form for this concept.

Keywords: Operator h-convex function, Entropy, Entropy inequality, Relative entropy

Mathematics Subject Classification [2010]: 47A63

1 Introduction and Preliminaries

In [10], Varošanec defined the concept of h-convex functions as follows:
Let $h : J \subseteq \mathbb{R} \to \mathbb{R}$ be a non-negative function, $h \not\equiv 0$. We say that $f : I \to \mathbb{R}$ is an h-convex function, or that f belongs to the class $SX(h, I)$, if f is non-negative and for all $x, y \in I$, $t \in (0, 1)$ we have

$$f(tx + (1 - t)y) \leq h(t)f(x) + h(1 - t)f(y)\,. \tag{1}$$

If inequality (1) is reversed, then f is said to be h-concave, that is $f \in SV(h, I)$.

If $h(t) = t$, then $SX(h, I)$ is exactly equal to the set all non-negative convex functions are definded on the interval I and the set of all non-negative concave functions is equal to $SV(h, I)$.

If the domain of a function h is closed under multiplication, then h is called a *super-multiplicative function* if

$$h(xy) \geq h(x)h(y)\,, \tag{2}$$

for all $x, y \in J$ [10]. If inequality (2) is reversed, then h is called a *sub-multiplicative function*. If the equality holds in (2), then h is called a *multiplicative function*.

Example 1.1. [10] Consider the function $h : [0, +\infty) \to \mathbb{R}$ by $h(x) = (c + x)^{p-1}$, where p is a non-negative real number. If $c = 0$, then the function h is multiplicative. If $c \geq 1$, then for $p \in (0, 1)$ the function h is super-multiplicative and for $p > 1$ the function h is sub-multiplicative.

*Speaker. Email address: morassaei@znu.ac.ir

Definition 1.2. Let $h : J \subseteq \mathbb{R} \to \mathbb{R}$ be a non-negative function, $h \not\equiv 0$. We say that $f : I \to \mathbb{R}$ is an *operator h-convex function*, if f is non-negative continuous and for all $A, B \in \mathbb{B}(\mathcal{H})$ with $\sigma(A), \sigma(B) \subseteq I$ and $t \in (0,1)$,

$$f\left(tA + (1-t)B\right) \leq h(t)f(A) + h(1-t)f(B)\,. \tag{3}$$

If inequality (3) is reversed, then f is said to be *operator h-concave*.

If $t = \frac{1}{2}$ in (3), then f is called *h-mid-convex function*.

Example 1.3. Assume that h is a function on $[0,\infty)$ such that $h(t) \geq t$ and $f : I \to \mathbb{R}$ given by $f(t) = t^2$. Then f is operator h-mid-convex function.

Corollary 3.7 in [2] state that if Φ is a normalized positive linear map and f is an operator h-convex function on an interval I, then

$$f(\Phi(A)) \leq 2h\left(\frac{1}{2}\right)\Phi(f(A))\,, \tag{4}$$

for every self-adjoint operator A with $\sigma(A) \subseteq I$. This inequality is said to be *Davis-Choi-Jensen's inequality*.

A relative operator entropy of strictly positive operators A, B was introduced in the noncommutative information theory by Fujii and Kamei [3] and is defined by

$$S(A|B) = A^{1/2}\log(A^{-1/2}BA^{-1/2})A^{1/2}.$$

For positive operators A, B, one may set $S(A|B) := s - \lim_{\epsilon\to+0} S(A+\epsilon I|B)$ if it exists. The relative entropy satisfies $S(U^*AU, U^*BU) = S(A,B)$ for all unitaries. $S(A,B)$ is tangent vector of the geodesic $\gamma(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^tA^{1/2}$ joining A to B at $t = 0$.

Let $a_1, \cdots, a_n > 0$ be such that $\sum_{j=1}^n a_j = 1$. The entropy of $a_1, \cdots, a_n$ is defined by

$$H(a_1, \cdots, a_n) = -\sum_{j=1}^n a_j \log a_j\,. \tag{5}$$

Entropy inequality states that

$$H(a_1, \cdots, a_n) \leq \log n\,, \tag{6}$$

or equivalently $\frac{1}{n} \leq \prod_{j=1}^n a_j^{a_j}$, see [1,6]. Rooin and Morassaei have studied some numerical refinements of the entropy and information inequalities [8]. The Shannon inequality as an extension of the entropy inequality asserts that if $(a_1, \cdots, a_n), (b_1, \cdots, b_n)$ are two probability vectors, then $0 \geq \sum_{j=1}^n a_j \log(\frac{b_j}{a_j})$, see [9].

Furuta [4] obtained a parametric extensions of Shannon's inequality and its reverse one in $\mathbb{B}(\mathcal{H})$. In particular, it is shown that $0 \geq \sum_{i=1}^n S(A_j \mid B_j)$ for n-tuples of positive operators with the unit operator as sum.

Moslehian, Mirzapour and Morassaei in [7] present an extension of the entropy inequality for Hilbert space operators. More precisely, let $p \in [0,1]$ and let $\mathbf{A} = (A_1, \cdots, A_n)$ and $\mathbf{B} = (B_1, \cdots, B_n)$ be two sequences of strictly positive contractions on a Hilbert space $\mathcal{H}$ such that $\sum_{j=1}^n A_j = \sum_{j=1}^n B_j = I$. If f is operator concave, then

$$f\left[\sum_{j=1}^n (A_j \natural_{p+1} B_j) + t_0\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right)\right] - f(t_0)\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right)$$

$$\geq S_p^f(\mathbf{A}|\mathbf{B}) \tag{7}$$

for all $p \in [0,1]$ and for any fixed real number $t_0 > 0$, and

$$-f\left[\sum_{j=1}^{n}(A_j \natural_{p-1} B_j) + t_0\left(I - \sum_{j=1}^{n} A_j \sharp_p B_j\right)\right] + f(t_0)\left(I - \sum_{j=1}^{n} A_j \sharp_p B_j\right)$$
$$\leq S_p^f(\mathbf{A}|\mathbf{B}) \tag{8}$$

for all $p \in [2,3]$ and for any fixed real number $t_0 > 0$, where $S_q^f(A|B)$ is defined by

$$S_q^f(A|B) = \sum_{j=1}^{n} A_j^{\frac{1}{2}}\left(A_j^{-\frac{1}{2}} B_j A_j^{-\frac{1}{2}}\right)^q f\left(A_j^{-\frac{1}{2}} B_j A_j^{-\frac{1}{2}}\right) A_j^{\frac{1}{2}},$$

in which q is a real number and f is an operator monotone function.

2 Main Result

If $\mathbf{a} = (a_1, a_2, \cdots, a_n)$ and $\mathbf{b} = (b_1, b_2, \cdots, b_n)$; a_j and b_j $(1 \leq j \leq n)$ are positive real number such that $\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j = 1$ and $p \geq 1$, the *relative entropy* is

$$S_p^f(\mathbf{a}|\mathbf{b}) := \sum_{j=1}^{n} a_j \left(\frac{b_j}{a_j}\right)^p f\left(\frac{b_j}{a_j}\right). \tag{9}$$

Now, if h is non-negative function and f is h-convex function, then we define

$${}_hS_p^f(\mathbf{a}|\mathbf{b}) := \sum_{j=1}^{n} h\left({a_j}^{\frac{1}{2}}\right)^2 h\left(\frac{b_j}{a_j}\right)^p f\left(\frac{b_j}{a_j}\right). \tag{10}$$

In additional, if h is multiplicative function, then

$${}_hS_p^f(\mathbf{a}|\mathbf{b}) := \sum_{j=1}^{n} h\left(a_j^{1-p} b_j^p\right) f\left(\frac{b_j}{a_j}\right). \tag{11}$$

Lemma 2.1. *Let $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$, $\boldsymbol{b} = (b_1, b_2, \cdots, b_n)$ and $p \geq 1$, then*

$$\sum_{j=1}^{n}(a_j \sharp_p b_j) \leq \sum_{j=1}^{n} a_j \sharp_p \sum_{j=1}^{n} b_j.$$

Theorem 2.2. *If $\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j = 1$ and f be h-conve function then*

$$f\left(\sum_{j=1}^{n}(a_j \sharp_{p+1} b_j) + t_0\Big(1 - \sum_{j=1}^{n} a_j \sharp_p b_j\Big)\right)$$
$$\geq \sum_{j=1}^{n} h({a_j}^{1-p} {b_j}^p) f({a_j}^{-1} b_j) + h\Big(1 - \sum_{j=1}^{n} a_j \sharp_p b_j\Big) f(t_0) \tag{12}$$
$$\geq \ {}_hS_p{}^f(a|b) + h\Big(1 - \sum_{j=1}^{n} a_j \sharp_p b_j\Big) f(t_0).$$

Now, we state the generalization of (7) and (8) as follows:

Theorem 2.3. *Let $p \in [0,1]$ and let $\mathbf{A} = (A_1, \cdots, A_n)$ and $\mathbf{B} = (B_1, \cdots, B_n)$ be two sequences of strictly positive contractions on a Hilbert space $\mathcal{H}$ such that $\sum_{j=1}^n A_j = \sum_{j=1}^n B_j = I$. If h is multiplicative function such that spectrum of A_j and B_j contained in domain of h and f is operator h-concave, then*

$$f\left[\sum_{j=1}^n (A_j \natural_{p+1} B_j) + t_0\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right)\right] - f(t_0)h\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right) \geq {}_hS_p^f(\mathbf{A}|\mathbf{B}) \tag{13}$$

for all $p \in [0,1]$ and for any fixed real number $t_0 > 0$, and

$$-f\left[\sum_{j=1}^n (A_j \natural_{p-1} B_j) + t_0\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right)\right] + f(t_0)h\left(I - \sum_{j=1}^n A_j \sharp_p B_j\right) \leq {}_hS_p^f(\mathbf{A}|\mathbf{B}) \tag{14}$$

for all $p \in [2,3]$ and for any fixed real number $t_0 > 0$, where ${}_hS_q^f(A|B)$ is defined by

$${}_hS_q^f(A|B) = \sum_{j=1}^n h\left(A_j^{\frac{1}{2}}\right) h\left(A_j^{-\frac{1}{2}} B_j A_j^{-\frac{1}{2}}\right)^q f\left(A_j^{-\frac{1}{2}} B_j A_j^{-\frac{1}{2}}\right) h\left(A_j^{\frac{1}{2}}\right),$$

in which q is a real number.

References

[1] R. Bhatia, *Positive definite matrices*, Princeton University Press, 2007.

[2] T.H. Dinh, K.T.B. Vo, Some inequalities for operator (p,h)-convex functions, *Linear Multilinear A.* **66** (2018), 580–592.

[3] J.I. Fujii and E. Kamei, Relative operator entropy in noncommutative information theory, *Math. Japonica* **34** (1989), 341–348.

[4] T. Furuta, Parametric extensions of Shannon inequality and its reverse one in Hilbert space operators, *Linear Algebra Appl.* **381** (2004), 219–235.

[5] T. Furuta, J. Mićić Hot, J. Pečarić, Y. Seo, *Mond-Pečarić method in operator inequalities*, Element, Zagreb, 2005.

[6] G.A. Jones and J.M. Jones, *Information and coding theory*, Springer-Verlag, 2000.

[7] M.S. Moslehian, F. Mirzapour and A. Morassaei, Operator entropy inequalities, *Coll. Math.* **130** (2013), 159–168.

[8] J. Rooin and A. Morassaei, Some refinements of relative information inequality, *Creative Math. & Inf.* **16** (2007), 95–98.

[9] C.E. Shannon, *The mathematical theory of communication*, University of Illinois Press, Urbana, 1949.

[10] S. Varošanec, On h-convexity, *J. Math. Anal. Appl.* **326** (2007), 303–311.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some properties of multiplicative-additive functions with applications

Ismail Nikoufar*

Department of Mathematics, Payame Noor University, Tehran, Iran

Abstract

In this paper, we find some properties of operator monotone and multiplicative-additive functions and its converse with applications for some relative operator entropies.

Keywords: Loewner-Heinz inequality, perspective function, relative operator entropy, generalized relative operator

Mathematics Subject Classification [2010]: 47A63, 46L05, 46L60

1 Introduction

The Loewner-Heinz inequality is a fundamental tool for treating operator inequalities. The Loewner-Heinz inequality means that the power function t^α is operator monotone on $[0,\infty)$ for $0<\alpha<1$. Uchiyama [6] showed that $A\le B$ if and only if $(A+\lambda)^\alpha\le(B+\lambda)^\alpha$ for every $\lambda>0$. He proved a converse of Loewner-Heinz inequality and applied it to the operator mean and spectral order. The converse of the Loewner-Heinz inequality in the view point of the perspective and generalized perspective of operator monotone and multiplicative functions was investigated in [4], where perspective inequalities were given equivalent to the Loewner-Heinz inequality.

Let A and B be two bounded self-adjoint operators on a Hilbert space $\mathcal{H}$. The partial order $A\le B$ means that $\langle Ax,x\rangle\le\langle Bx,x\rangle$ for every $x\in\mathcal{H}$. A real continuous function $f(t)$ defined on a real interval is said to be operator monotone provided that $A\le B$ implies that $f(A)\le f(B)$ for any two self-adjoint operators A and B whose spectra are in the domain of f. A non-negative operator monotone function is considered as a variation of an operator mean by the theory of operator means introduced by Kubo and Ando. However, this theory does not include the logarithm and the entropy function which are operator monotone and often used in information theory.

2 The main results

We denote the set of invertible positive operators by $\mathcal{B}(\mathcal{H})_{++}$. Throughout this paper A,B,C,D stand for invertible positive operators and for a real number λ we write $A+\lambda$,

*Speaker. Email address: nikoufar@pnu.ac.ir

for short, instead of $A + \lambda I$. Recall that the perspective of the one variable continuous function f was defined in [1] by setting

$$\mathcal{P}_f(A, B) = B^{1/2} f(B^{-1/2} A B^{-1/2}) B^{1/2}$$

for $A, B \in \mathcal{B}(\mathcal{H})_{++}$ with the spectrum of the operator $B^{-1/2} A B^{-1/2}$ in the domain of f.

Throughout this paper, let $f : (0, \infty) \to \mathbb{R}$ and $h : (0, \infty) \to (0, \infty)$ be two continuous functions. The generalized perspective of two variables (associated with f and h) was defined by

$$\mathcal{P}_{f\Delta h}(A, B) = h(B)^{1/2} f(h(B)^{-1/2} A h(B)^{-1/2}) h(B)^{1/2},$$

where $A, B \in \mathcal{B}(\mathcal{H})_{++}$ [1]. For any continuous function $f : (0, \infty) \longrightarrow \mathbb{R}$ the transpose $\tilde{f}$ of f is defined by

$$\tilde{f}(x) = x f(x^{-1}), \quad x > 0.$$

It is well-known that the transpose $\tilde{f}$ of an operator monotone function f on $(0, \infty)$ is operator monotone again. The function f is called multiplicative if $f(ts) = f(t)f(s)$ for every $t, s \in \mathbb{R}$. We say that f is multiplicative-additive if $f(ts) = f(t) + f(s)$. For example, the functions t^α and $\log t$ are multiplicative and multiplicative-additive, respectively.

Theorem 2.1. *[6] Let $f(t)$ be a non-constant operator monotone function in a neighbourhood of $t = a$. Then, $A \leq B$ if and only if there exists a sequence $\{t_n\}$ such that $t_n \to 0$ and*

$$f(a + t_n A) \leq f(a + t_n B).$$

Lemma 2.2. *[4] Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be a continuous function and $\tilde{f}$ the transpose of f. Then,*

$$\mathcal{P}_{\tilde{f}}(A, B) = \mathcal{P}_f(B, A)$$

for every $A, B \in \mathcal{B}(\mathcal{H})_{++}$.

Theorem 2.3. *[4] If $A \leq B$, $C \leq D$ and f is non-negative and operator monotone, then*

$$\mathcal{P}_f(A, C) \leq \mathcal{P}_f(B, D).$$

Theorem 2.4. *Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be a non-constant operator monotone and multiplicative-additive function and $A, B \geq 0$. Then the following statements are equivalent:*

(i) $A \leq B$

(ii) $A + \lambda \leq B + \lambda$ *for every* $\lambda \geq 0$,

(iii) $f(A + \lambda) \leq f(B + \lambda)$ *for every* $\lambda \geq 0$,

(iv) $f(A + \lambda_n) \leq f(B + \lambda_n)$ *for a sequence* $\{\lambda_n\}$ *such that* $\lambda_n > 0$ *and* $\lambda_n \to \infty$ *as* $n \to \infty$,

(v) $f(t_n A + 1) \leq f(t_n B + 1)$ *for a sequence* $\{t_n\}$ *such that* $t_n > 0$ *and* $t_n \to 0$ *as* $n \to \infty$.

Lemma 2.5. *Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be a multiplicative-additive function. Then,*

(i) $\tilde{f}(ts) = t\tilde{f}(s) + s\tilde{f}(t)$ *for every* $t, s > 0$,

(ii) $f(1) = \tilde{f}(1) = 0$.

Lemma 2.6. *Let $f : (0, \infty) \longrightarrow \mathbb{R}$ be multiplicative-additive and $A, B \in \mathcal{B}(\mathcal{H})_{++}$. Then,*

(i) $\mathcal{P}_f(A+\lambda, B+\mu) = \tilde{f}(\mu)(\frac{B}{\mu}+1) + \mu\mathcal{P}_f(A+\lambda, \frac{B}{\mu}+1)$ *for every* $\lambda, \mu > 0$,

(ii) $\mathcal{P}_f(A+\lambda, B+\mu) = f(\lambda)(B+\mu) + \mathcal{P}_f(\frac{A}{\lambda}+1, B+\mu)$ *for every* $\lambda, \mu > 0$.

Theorem 2.7. *Let* $f : (0,\infty) \longrightarrow \mathbb{R}$ *be non-negative, operator monotone and multiplicative-additive. Then the following statements are equivalent for* $A, B, C, D \in \mathcal{B}(\mathcal{H})_{++}$.

(i) $A \leq B$ *and* $C \leq D$,

(ii) $\mathcal{P}_f(\lambda^{-1}, \mu^{-1})C + \mu^{-1}\mathcal{P}_f(A+\lambda, C+\mu) \leq \mathcal{P}_f(\lambda^{-1}, \mu^{-1})D + \mu^{-1}\mathcal{P}_f(B+\lambda, D+\mu)$ *for every* $\lambda, \mu > 0$,

(iii) $\mathcal{P}_f(sA+1, tC+1) \leq \mathcal{P}_f(sB+1, tD+1)$ *for every* $s, t > 0$.

Corollary 2.8. *Let* $f : (0,\infty) \longrightarrow \mathbb{R}$ *be non-negative, operator monotone and multiplicative-additive. Then the following statements are equivalent for* $A, B, C, D \in \mathcal{B}(\mathcal{H})_{++}$.

(i) $A \leq B$ *and* $C \leq D$,

(ii) $\mathcal{P}_f(A+\lambda, C+\lambda) \leq \mathcal{P}_f(B+\lambda, D+\lambda)$ *for every* $\lambda > 0$,

(iii) $\mathcal{P}_f(sA+1, sC+1) \leq \mathcal{P}_f(sB+1, sD+1)$ *for every* $s > 0$.

Lemma 2.9. *Let* f *be multiplicative-additive and let* h *be multiplicative and* $A, B \in \mathcal{B}(\mathcal{H})_{++}$. *Then*

$$\mathcal{P}_{f\Delta h}(A+\lambda, B+\mu) = \tilde{f}(h(\mu))h(\frac{B}{\mu}+1) + h(\mu)\mathcal{P}_{f\Delta h}(A+\lambda, \frac{B}{\mu}+1)$$

for every $\lambda, \mu > 0$.

Lemma 2.10. *Let* f *be multiplicative-additive and* $A, B \in \mathcal{B}(\mathcal{H})_{++}$. *Then*

$$\mathcal{P}_{f\Delta h}(A+\lambda, B+\mu) = f(\lambda)h(B+\mu) + \mathcal{P}_{f\Delta h}(\frac{A}{\lambda}+1, B+\mu)$$

for every $\lambda, \mu > 0$.

Theorem 2.11. *[4] Let* f *be a non-negative, operator monotone function and* h *an operator monotone function. If* $A \leq B$, $C \leq D$, *then*

$$\mathcal{P}_{f\Delta h}(A, C) \leq \mathcal{P}_{f\Delta h}(B, D).$$

Theorem 2.12. *Let* f *be a non-negative, operator monotone, and multiplicative-additive function. If* h *is operator monotone and multiplicative, then the following statements are equivalent for* $A, B, C, D \in \mathcal{B}(\mathcal{H})_{++}$.

(i) $A \leq B$ *and* $C \leq D$,

(ii) $\mathcal{P}_{f\Delta h}(\lambda^{-1}, \mu^{-1})h(C+\mu) + h(\mu^{-1})\mathcal{P}_{f\Delta h}(A+\lambda, C+\mu) \leq \mathcal{P}_{f\Delta h}(\lambda^{-1}, \mu^{-1})h(D+\mu) + h(\mu^{-1})\mathcal{P}_{f\Delta h}(B+\lambda, D+\mu)$ *for every* $\lambda, \mu > 0$,

(iii) $\mathcal{P}_{f\Delta h}(sA+1, tC+1) \leq \mathcal{P}_{f\Delta h}(sB+1, tD+1)$ *for every* $s, t > 0$.

Corollary 2.13. *Let* f *be a non-negative, operator monotone, and multiplicative-additive function. If* h *is operator monotone and multiplicative, then the following statements are equivalent for* $A, B, C, D \in \mathcal{B}(\mathcal{H})_{++}$.

(i) $A \leq B$ *and* $C \leq D$,

(ii) $\mathcal{P}_{f\Delta h}(A+h(\mu), C+\mu) \leq \mathcal{P}_{f\Delta h}(B+h(\mu), D+\mu)$ *for every* $\mu > 0$,

(iii) $\mathcal{P}_{f\Delta h}(h(t)A+1, tC+1) \leq \mathcal{P}_{f\Delta h}(h(t)B+1, tD+1)$ *for every* $t > 0$.

3 Applications

Fujii and Kamei [2] introduced the relative operator entropy of invertible positive operators A and B by

$$S(A,B) = A^{\frac{1}{2}} \log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) A^{\frac{1}{2}}.$$

Indeed, the relative operator entropy is the perspective of the function $f(t) = \log t$, i.e.,

$$S(A,B) = \mathcal{P}_f(B,A).$$

We now apply Theorem 2.7 to the relative operator entropy.

Corollary 3.1. *For $A,B,C,D \in \mathcal{B}(\mathcal{H})_{++}$ with $A \le C$ and $B \le D$ the following are equivalent.*

(i) $A \le B$ *and* $C \le D$,

(ii) $S(A+\lambda, C+\mu) \le S(B+\lambda, D+\mu)$ *for every* $\lambda,\mu > 0$ *with* $\lambda \le \mu$,

(iii) $S(rA+1, tC+1) \le S(rB+1, tD+1)$ *for every* $r,t > 0$ *with* $r \le t$.

The generalized relative operator entropy for positive invertible operators A, B and $q \in \mathbb{R}$ was defined by Furuta as follows:

$$S_q(A|B) = A^{1/2}(A^{-1/2} B A^{-1/2})^q (\log A^{-1/2} B A^{-1/2}) A^{1/2}.$$

Using the notion of the generalized relative operator entropy, Furuta obtained the parametric extension of the operator Shannon inequality and its reverse one. Note that for $q = 0$ this reduces to $S_0(A|B) = S(A|B)$.

We introduced the notion of the relative operator (α,β)-entropy (two parameter relative operator entropy) as follows:

$$S_{\alpha,\beta}(A,B) := A^{\frac{\beta}{2}} (A^{-\frac{\beta}{2}} B A^{-\frac{\beta}{2}})^{\alpha} (\log A^{-\frac{\beta}{2}} B A^{-\frac{\beta}{2}}) A^{\frac{\beta}{2}}$$

for positive invertible operators A, B and real numbers α, β. In particular, we have $S_{\alpha,1}(A|B) = S_\alpha(A|B)$ and $S_{0,1}(A|B) = S(A|B)$. The bounds of the generalized relative operator entropy in a general form were determined in [3, 5].

It would be remark that the relative operator (α,β)-entropy is the generalized perspective of the functions $f(t) = t^\alpha \log t$ and $h(t) = t^\beta$, $0 < \beta < 1$, i.e.,

$$S_{\alpha,\beta}(A,B) = \mathcal{P}_{f\Delta h}(B,A).$$

We apply Theorem 2.12 to the relative operator $(0,\beta)$-entropy.

Corollary 3.2. *For $A,B,C,D \in \mathcal{B}(\mathcal{H})_{++}$ with $I \le A \le C$ and $I \le B \le D$ and $0 < \beta < 1$ the following are equivalent.*

(i) $A \le B$ *and* $C \le D$,

(ii) $S_{0,\beta}(A+\lambda, C+\mu) \le S_{0,\beta}(B+\lambda, D+\mu)$ *for every* $\lambda,\mu > 0$ *with* $1 \le \lambda \le \mu$,

(iii) $S_{0,\beta}(rA+1, tC+1) \le S_{0,\beta}(rB+1, tD+1)$ *for every* $r,t > 0$ *with* $r \le t \le 1$.

4 Conclusion

In this paper, we obtained some properties of operator monotone and multiplicative-additive functions and its converse with applications. We also identified the equivalence relations among inequalities for some relative operator entropies in a view point of the perspective.

References

[1] A. Ebadian, I. Nikoufar, and M. Eshagi Gordji, Perspectives of matrix convex functions, *Proc. Natl. Acad. Sci.*, 108 (2011), 7313-7314.

[2] J. I. Fujii and E. Kamei, Relative operator entropy in noncommutative information theory, *Math. Japonica*, 34 (1989), 341-348.

[3] I. Nikoufar, Bounds of some relative operator entropies in a general form, *Filomat*, 32 (2018), 5105-5114.

[4] I. Nikoufar and M. Shamohammadi, The converse of the Loewner-Heinz inequality via perspective, *Linear Multilinear Alg.*, 66 (2018), 243-249.

[5] I. Nikoufar and M. Fazlolahi, Equivalence relations among inequalities for some relative operator entropies, *Positivity*, 24 (2020), 1503-1518.

[6] M. Uchiyama, A converse of the Loewner-Heinz inequality, geometric mean and spectral order, *Proc. Edinburgh Math. Soc.*, 57 (2014), 565-571.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The behavior of an operator in terms of its components on Hilbert C^*-modules

Javad Farokhi Ostad*

Department of Basic Sciences, Birjand University, Birjand, Iran

Abstract

In this paper, the behavior of an operator on Hilbert C^*-module via the matrix decomposition, have been studied. This relationship has been verified in terms of boundedness, closed ranges, homogeneity, normality and some other characteristics. Also, we have shown that if T is EP, then its components are also EP. Finally, we have provided conditions under which T is a hypo-EP operator.

Keywords: Hilbert C^*-module, Moore-Penrose inverse, EP operator
Mathematics Subject Classification [2010]: 47A05, 47A06

1 Pereliminaries

A Hilbert C^*-module is a generalization of a Hilbert space, where the algebra of complex numbers is replaced by a possibly more general C^*-algebra $\mathcal{A}$. In fact, a Hilbert C^*-module has an inner product which takes values not in $\mathbb{C}$, but in $\mathcal{A}$ (i. e. its inner product as a generalization of a complex-valued inner product has been considered).

Let $\mathcal{A}$ be a C^*-algebra (not necessarily unital). A right pre-Hilbert module over $\mathcal{A}$ is a complex linear space $\mathcal{X}$, which is an algebraic right $\mathcal{A}$-module and $\lambda(xa) = (\lambda x)a = x(\lambda a)$ equipped with an $\mathcal{A}$-valued inner product $\langle .,. \rangle : \mathcal{X} \times \mathcal{X} \to \mathcal{A}$ satisfied the following conditions;

1. $\langle x, x\rangle \geq 0$, and $\langle x, x\rangle = 0$ if and only if $x = 0$,

2. $\langle x, y + \lambda z\rangle = \langle x, y\rangle + \lambda\langle x, z\rangle$,

3. $\langle x, ya\rangle = \langle x, y\rangle a$,

4. $\langle y, x\rangle = \langle x, y\rangle^*$.
for each $x, y, z \in \mathcal{X}$, $\lambda \in \mathbb{C}$, $a \in \mathcal{A}$.

*Speaker. Email address: J.farrokhi@birjandut.ac.ir

Left pre-Hilbert C^*-modules are defined similarly.

Hilbert C^*-modules form a category in between Banach spaces and Hilbert spaces. The basic idea was to consider module over C^*-algebra instead of linear space and to allow the inner product to take values in a more general C^*-algebra than $\mathbb{C}$. For example, every inner product space is a left Hilbert $\mathbb{C}$-module.
Every C^*-algebra $\mathcal{A}$ is a Hilbert $\mathcal{A}$-module with respect to the inner product $\langle x, y\rangle = x^*y$.

A pre-Hilbert C^*-module $\mathcal{X}$ is called a Hilbert C^*-module if $\mathcal{X}$ equipped with the $\|x\| = \|\langle x, x\rangle\|^{\frac{1}{2}}$, for any $x \in \mathcal{X}$. For comprehensive accounts we refer to the lecture note of Lance [3].

Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert $\mathcal{A}$-modules, $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ is called the space of adjointable maps $\mathcal{L}(\mathcal{X}, \mathcal{Y}) = \{T : \mathcal{X} \to \mathcal{Y} | \ \exists\ T^* : \mathcal{Y} \to \mathcal{X}; \langle Tx, y\rangle = \langle x, T^*y\}$, and we put $\mathcal{L}(\mathcal{X}) = \mathcal{L}(\mathcal{X}, \mathcal{X})$. We use the notations $\ker(\cdot)$ and $\mathrm{ran}(\cdot)$ for the kernel and the range of operators, respectively.

Let $T \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$. The Moore-Penrose inverse $T^\dagger$ of T (if it exists) is an element of $\mathcal{L}(\mathcal{Y}, \mathcal{X})$ which satisfies:

1. $T\,T^\dagger T = T$,
2. $T^\dagger\,T\,T^\dagger = T^\dagger$,
3. $(T\,T^\dagger)^* = T\,T^\dagger$,
4. $(T^\dagger\,T)^* = T^\dagger T$.

- The operator $T \in \mathcal{L}(\mathcal{X})$ is normal, if $T^*T - TT^* = 0$.
- The operator $T \in \mathcal{L}(\mathcal{X})$ with closed range is called EP operator (As already mentioned in the Introduction, EP stands for Equal Projections), if ran(T) = ran(T*).

  Equvalently, the operator $T \in \mathcal{L}(\mathcal{X})$ is EP if and only if $T^\dagger T - TT^\dagger = 0$.
- The operator T is hypo-EP provided that T has closed range and $T^\dagger T - TT^\dagger \geq 0$.
- Let $\mathcal{M}$ and $\mathcal{N}$ are closed orthogonally complemented submodules of $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\mathcal{X} = \mathcal{M} \oplus \mathcal{M}^\perp$, $\mathcal{Y} = \mathcal{N} \oplus \mathcal{N}^\perp$, then T can be written as the following 2×2 matrix

$$T = \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} \tag{1}$$

  where, $T_1 \in \mathcal{L}(\mathcal{M}, \mathcal{N})$, $T_2 \in \mathcal{L}(\mathcal{M}^\perp, \mathcal{N})$, $T_3 \in \mathcal{L}(\mathcal{M}, \mathcal{N}^\perp)$ and $T_4 \in \mathcal{L}(\mathcal{M}^\perp, \mathcal{N}^\perp)$. Note that $P_\mathcal{M}$ denotes the projection corresponding to $\mathcal{M}$.

  In fact $T_1 = P_\mathcal{N} T P_\mathcal{M}$, $T_2 = P_\mathcal{N} T(1 - P_\mathcal{M})$ $T_3 = (1 - P_\mathcal{N}) T P_\mathcal{M}$, $T_4 = (1 - P_\mathcal{N}) T (1 - P_\mathcal{M})$.

2 Main results

Regarding an operator in $T \in \mathcal{L}(\mathcal{X})$, with the above matrix representation 1, if all its components T_1, T_2, T_3, and T_4 are bounded, then the operator T itself is also bounded. In the next Theorem, we see that the reverse of this issu is not necessarily true.

The operator $T \in \mathcal{L}(\mathcal{X})$ is called sel-adjoint, normal, EP (stands for Equal Projections) and hypo-EP when $T = T^*$, $TT^* = T^*T$, ran(T) and ran(T*) have the same closure and $T^\dagger T - TT^\dagger$ is a positive operator, respectively.

Theorem 2.1. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$,*

1. *For $T = \begin{bmatrix} T_1 & 0 \\ 0 & T_4 \end{bmatrix}$, if T_1 or T_4 is not bounded, then T is not bounded.*

2. *For $T = \begin{bmatrix} 0 & T_2 \\ T_3 & 0 \end{bmatrix}$, if T_2 or T_3 is not bounded, then T is not bounded.*

Remark 2.2. Note that sometimes all components of an operator may be unbounded, while the operator itself is bounded. For example, suppose that T_1, T_2, T_3, and T_4 be unbounded with $\mathrm{ran(T_1)} \cap \mathrm{ran(T_3)} = \mathrm{ran(T_2)} \cap \mathrm{ran(T_4)} = \{0\}$. Thus, $\mathrm{ran(T)} = \{0\} \oplus \{0\}$ and so $T = \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix}$ is bounded on $\mathrm{ran(T)}$.

Theorem 2.3. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$.*

1. *For $T = \begin{bmatrix} T_1 & 0 \\ 0 & T_4 \end{bmatrix}$, if the range of T_1 or T_4 is not closed, then T has not the closed range.*

2. *For $T = \begin{bmatrix} 0 & T_2 \\ T_3 & 0 \end{bmatrix}$, if the range of T_2 or T_3 s not closed, then T has not the closed range.*

Remark 2.4. It is even possible that the ranges of the components are not all closed, while the range of the operator itself is closed. For example, suppose that the range of T_1 is not closed, but $\overline{\mathrm{ran(T_1)}} = \mathcal{X}$. Take, $T = \begin{bmatrix} T_1 & T_1 \\ T_1 & T_1 \end{bmatrix}$. Let $x \in \mathcal{X}$, there are $\{x_n\} \subseteq \mathcal{X}$ and $\{-x_n\} \subseteq \mathcal{X}$such that $x_n \longrightarrow x$ and $-x_n \longrightarrow -x$ respectivly.

Now, $T \begin{pmatrix} x_n \\ -x_n \end{pmatrix} = \begin{bmatrix} T_1 & T_1 \\ T_1 & T_1 \end{bmatrix} \begin{pmatrix} x_n \\ -x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$,
while $(x, -x) \in \mathrm{ran(T_1)}$.

Theorem 2.5. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$. Also, let $T_2, T_3 \in \mathcal{L}(\mathcal{X})$ be such that $T_2T_3 = T_3T_2 = 0$, then for $T = \begin{bmatrix} 0 & T_2 \\ T_3 & 0 \end{bmatrix}$, it must* $\mathrm{ran(T)} = \mathrm{ran(T^2)}$.

Note that, in this case $T^2 = \begin{bmatrix} T_2T_3 & 0 \\ 0 & T_3T_2 \end{bmatrix}$, and so

$$\begin{aligned} \mathrm{ran(T^2)} &= \mathrm{ran(T_2T_3)} \oplus \mathrm{ran(T_3T_2)} \\ &= \mathrm{ran(T_2)} \oplus \mathrm{ran(T_3)} \\ &= \mathrm{ran(T)} \end{aligned}$$

Remark 2.6. Even, this issue can be properly generalized up to power n. In fact, there is nilpotent operator T_2 with closed range, and $T = \begin{bmatrix} 0 & T_2 \\ T_2 & 0 \end{bmatrix}$, which also,

$$\mathrm{ran(T)} = \mathrm{ran(T^2)} = \cdots = \mathrm{ran(T^n)} \neq \mathcal{X}.$$

Theorem 2.7. *Let the range of T_1 has not closed, and $T_1^2 = 0$, put $T = \begin{bmatrix} T_1 & I \\ I & -T_1 \end{bmatrix}$, then T is not closed range operator, but $T^2 = I$.*

Theorem 2.8. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$. For normal operator T, we have $T^4 = T^5$, if and only if T is projection ($T^* = T = T^2$).*

Remark 2.9. Even, this issue can be properly generalized up to power n. In fact, For normal operator T, we have $T^{2n} = T^{2n+1}$; $\quad n \geq 3$, if and only if T is projection.

Theorem 2.10. *Let $\mathcal{X}$ be Hilbert C^*-module. Also, let $T_2^2 = I$, then $T = \begin{bmatrix} 0 & T_2 \\ I & 0 \end{bmatrix}$, and $S = \begin{bmatrix} 0 & I \\ T_2 & 0 \end{bmatrix}$ then $TS = ST$.*

Remark 2.11. Note that, not only $T \neq S$, but also,

$$\begin{aligned} TS &= \begin{bmatrix} 0 & T_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} 0 & I \\ T_2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} T_2^2 & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & T_2^2 \end{bmatrix} \\ &= ST. \end{aligned}$$

Theorem 2.12. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$. If T is EP, then so is all components. In addition, in this case the $|T| := (T^*T)^{\frac{1}{2}}$ is EP too.*

In the next theorem, via the matrix decomposition of operator $T \in \mathcal{L}(\mathcal{X})$ we show that under certain conditions T becomes hypo-EP.

Theorem 2.13. *Let $\mathcal{X}$ be Hilbert C^*-module and $T \in \mathcal{L}(\mathcal{X})$. If $(T^\dagger T - I)TT^\dagger = 0$, then T is hypo-EP operator.*

Acknowledgement

I would like to express my gratitude to the organizers of the 12th Seminar on Linear Algebra and its Applications, Sahand University of Technology.

References

[1] J. Farokhi-ostad and M. Mohammadzadeh karizaki, *New identities for Moore-Penrose inverses of some operator products and their reverse-order laws*, Acta Mathematica Sinica, English Series, (2023), To apear.

[2] A. R. Janfada and J. Farokhi-ostad, *Two equal range operators on Hilbert C^*-Modules*, Sahand Communications in Math. Anal. 18(2), 85-96 (2021).

[3] E. C. Lance, *Hilbert C^*-Modules*, LMS Lecture Note Series 210 (1995).

[4] Jalaeian M, Mohammadzadeh Karizaki M, Hassani M. *Conditions that the product of operators is an EP operator in Hilbert C^*-module*, Linear Multilinear Algebra., 68(10), (2020), 1990–2004.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Existence of positive operators associated with locally compact groups

Seyedeh Somayeh Jafari*

Department of Mathematics, Payame Noor University, Tehran, Iran

Abstract

Given a unitary representation (π, H_π) of a locally compact group G, we study the notions on (topological) invariant subspaces to non-empty subsets of $B(H_\pi)$. We also characterize the existence of a positive operator equipped with special properties.

Keywords: Amenibility, Bounded linear operator, Positivity, Unitary representation
Mathematics Subject Classification [2010]: 46H25, 22D10

1 Introduction

Throughout this paper, G is a locally compact group. As usual, $L^1(G)$ denotes the group algebra of G equipped with convolution product and $\|\cdot\|_1$ as defined in [3]. The notation l_x is the left translation operator by $x \in G$; i.e., $l_x f(y) = f(xy)$ for all complex-valued function f on G. Let $L^\infty(G)$ is usual Lebesgue space as defined in [3] equipped with the essential supremum $\|\cdot\|_\infty$. Then $L^\infty(G)$ can be identified by the first dual space of $L^1(G)$ under the pairing

$$\langle f, \phi\rangle = \int_G f(x)\phi(x)\,dx \qquad (f \in L^\infty(G),\ \phi \in L^1(G)).$$

We can also consider $L^\infty(G)$ as a right Banach $L^1(G)$-module by the following action.

$$f \cdot \phi = \int_G l_x f\ \phi(x)\,dx \qquad (f \in L^\infty(G),\ \phi \in L^1(G)).$$

Let also, $LUC(G)$ denote the C^*-algebra of left uniformly continuous functions; i.e., $f \in LUC(G)$ when the map $x \mapsto l_x f$ from G into $L^\infty(G)$ is norm continuous. A representation of G will always mean a continuous unitary representation of G as defined in [3]; i.e., a pair (π, H_π) where π is a homomorphism of G into the group unitary operators on the Hilbert space H_π that is continuous with respect to the strong operator topology on $B(H_\pi)$, consisting of all bounded linear operators on H_π. Note that $B(H_\pi)$ is a right G-module under the following action

$$T \cdot_\pi x = \pi(x^{-1})T\pi(x) \quad (T \in B(H_\pi),\ x \in G).$$

*Speaker. Email address: ss.jafari@pnu.ac.ir

In general, $B(H_\pi)$ is not Banach G-module in terms of Johnson's notion, [5]. In fact, for $T \in B(H_\pi)$, the map $x \mapsto T \cdot_\pi x$, $G \longrightarrow B(H_\pi)$ is not norm continuous, necessarily. We say that T is *uniformly G-continuous operator* if the mapping $x \mapsto T \cdot_\pi x$ are norm continuous. Suppose that the notation $UCB(\pi)$ refers to the collection of such operators. Then $UCB(\pi)$ is a C^*-subalgebra of $B(H_\pi)$, and also it is a right Banach G-module.

Moreover, $B(H_\pi)$ is a right Banach $L^1(G)$-module as follows.

$$T \cdot_\pi \phi = \int_G T \cdot_\pi x\, \phi(x)\, dx \qquad (T \in B(H_\pi),\ \phi \in L^1(G)).$$

Also, Cohen's factorization theorem implies that

$$B(H_\pi) \cdot_\pi L^1(G) = UCB(\pi) \cdot_\pi L^1(G) = UCB(\pi).$$

See [1] for more details and the survey article.

In recent years, some authors have studied the relations of G-module and $L^1(G)$-module maps, in the sense of the map commute with the translations and convolutions; see for example [4] and [6]. The mission of this note is to study this notions and applications on (topological) invariant subspaces to non-empty subsets of $B(H_\pi)$. We also characterize the existence of a positive operator equipped with special properties.

2 Main results

We commence with the known following definitions.

Definition 2.1. Let (π, H_π) be a unitary representation of a locally compact group G and $\mathcal{X}$ is a non-empty subset of $B(H_\pi)$ or $UCB(\pi)$. Then

(a) $\mathcal{X}$ is called invariant if $\mathcal{X} \cdot_\pi G \subseteq \mathcal{X}$, where

$$\mathcal{X} \cdot_\pi G = \{T \cdot_\pi x \,|\, T \in \mathcal{X},\ x \in G\}.$$

(b) $\mathcal{X}$ is called topologically invariant if $\mathcal{X} \cdot_\pi L^1(G) \subseteq \mathcal{X}$, where

$$\mathcal{X} \cdot_\pi L^1(G) = \{T \cdot_\pi \phi \,|\, T \in \mathcal{X},\ \phi \in L^1(G)\}.$$

Using the Dirac approximate identity of $L^1(G)$, the reader observes that topological invariant norm closed sets of $UCB(\pi)$ are invariant. Also, one can easily check that these concepts coincide on any closed subset of $B(H_\pi)$ with respect to weak operator topology.

For each M in the dual of $\mathcal{X} \cdot_\pi L^1(G)$, we define the bounded linear operator $\gamma_M : \mathcal{X} \longrightarrow L^\infty(G)$ given by

$$\langle \gamma_M(T),\, \phi \rangle = \langle M,\, T \cdot_\pi \phi \rangle \quad (T \in \mathcal{X},\ \phi \in L^1(G)).$$

Lemma 2.2. *Let (π, H_π) be a unitary representation of a locally compact group G. Let also, $\mathcal{X}$ be a topologically invariant closed subspace of $B(H_\pi)$. Then*

(a) $\mathcal{X} \cdot_\pi L^1(G)$ is a topologically invariant closed subspace of $UCB(\pi)$.

(b) $\|\gamma_M\| = \|M\|$ for each M in the dual of $\mathcal{X} \cdot_\pi L^1(G)$.

Recently, the author has investigated and studied the following notion; see [4].

Definition 2.3. Let (π, H_π) be a unitary representation of a locally compact group G, and let $\gamma : B(H_\pi) \longrightarrow L^\infty(G)$ be a bounded linear operator.

(a) γ is said *to commute with the action as $L^1(G)$-module* if

$$\gamma(T \cdot_\pi \phi) = \gamma(T) \cdot \phi \qquad (T \in B(H_\pi),\, \phi \in L^1(G)). \tag{1}$$

(b) γ is said *to commute with the action as G-module* if

$$\gamma(T \cdot_\pi x) = l_x \gamma(T) \qquad (T \in B(H_\pi),\, x \in G), \tag{2}$$

By the assumptions of the previous lemma, the following theorem reveals that only operators such form γ_M from $\mathcal{X}$ into $L^\infty(G)$ can commute with action as $L^1(G)$-module.

Theorem 2.4. *Let (π, H_π) be a unitary representation of a locally compact group G and $\mathcal{X}$ be a topologically invariant norm-closed subspace of $B(H_\pi)$. Let also, $\gamma : \mathcal{X} \longrightarrow L^\infty(G)$ be a linear bounded operator . Then the following statements are equivalent.*

(a) γ commutes with the action as $L^1(G)$-module,

(b) $\gamma = \gamma_M$ for some M in the dual of $\mathcal{X} \cdot_\pi L^1(G)$.

Also, if $\mathcal{X} \subseteq UCB(\pi)$ or $\mathcal{X}$ is closed with respect to weak operator topology of $B(H_\pi)$, then the above statements imply the following part.

(c) γ commutes with the action as G-module.

Now, we state one of the pivotal results of Chan's work that was proven in Proposition 2.3 of [2]. Before stating, note that for all $M \in B(H_\pi)^*$ and $T \in B(H_\pi)$, we can define the complex-valued function MT on G by

$$MT(x) = \langle M, T \cdot_\pi x \rangle \quad (x \in G).$$

Obviously, MT is bounded by $\|M\|\|T\|$. Note that $T \in UCB(\pi)$ if and only if $MT \in LUC(G)$ for all $M \in B(H)^*$.

Proposition 2.5. *Let (π, H_π) be a unitary representation of a locally compact group G. Then $UCB(\pi)^*$ is a left Banach $LUC(G)^*$-module, via the bounded bilinear mapping $LUC(G)^* \times UCB(\pi)^* \longrightarrow UCB(\pi)^*$ defined by $(m, M) \mapsto m \cdot M$, where $\langle m \cdot M, T\rangle = \langle m, MT\rangle$.*

Let $\mathcal{X}$ be a topologically invariant closed subspace of $UCB(\pi)$ and let $\Gamma(\mathcal{X}, G)$ be the space of all bounded linear operators from $\mathcal{X}$ into $L^\infty(G)$ that commuting with the action as $L^1(G)$-module. Note that automatically the range of such operators lies in $LUC(G)$.

Proposition 2.6. *Let (π, H_π) be a unitary representation of a locally compact group G and $\mathcal{X}$ be a topological invariant subspace of $UCB(\pi)$. Then $\Gamma(\mathcal{X}, G)$ is a left Banach $LUC(G)^*$-module by the following action.*

$$(m \cdot \gamma)(T) = m \cdot (\gamma(T)),$$

where $m \in LUC(G)^*$, $\gamma \in \Gamma(\mathcal{X}, G)$ and $T \in \mathcal{X}$.

The following result is one of the important aims of this memoir.

Theorem 2.7. *Let (π, H_π) be a unitary representation of a locally compact group G and let $\mathcal{X}$ be a topological invariant closed subspace of $UCB(\pi)$. Then there exists*

1. *an isometric isomorphism as left Banach $LUC(G)^*$-modules between the dual of $\mathcal{X} \cdot_\pi L^1(G)$ and $\Gamma(\mathcal{X}, G)$.*
2. *a homeomorphism as topological spaces between the dual of $\mathcal{X} \cdot L^1(G)$ and $\Gamma(\mathcal{X}, G)$.*

We end the section with a consequence of Theorem 2.7.

Corollary 2.8. *Let (π, H_π) be a unitary representation of a locally compact group G. Then $UCB(\pi)^*$ and $\Gamma\big(UCB(\pi), G\big)$ are isometrically isomorphic as $LUC(G)^*$-modules.*

3 The applications on amenability

One of the motivations of this note is the existence of a vast body of results on equivalents of locally compact groups equipped with the amenable property. Our conclusion in this regard realizes here.

It's known that (topological) invariant means can develop to any (topological) admissible subspace of $L^\infty(G)$. Let here (π, H_π) be a unitary representation of a locally compact group G. An invariant (resp. topological invariant) subspace $\mathcal{X}$ of $B(H_\pi)$ is called admissible (resp. topological admissible) subspace of $B(H_\pi)$ if it is a norm closed and conjugate closed, containing the identity operator on H_π; such as $UCB(\pi)$.

Definition 3.1. Let (π, H_π) be a unitary representation of a locally compact group G. Then

(a) a bounded linear functional M on an admissible (resp. topological admissible) subspace $\mathcal{X}$ of $B(H_\pi)$ is called a state on $\mathcal{X}$ if it satisfies $\|M\| = M(I) = 1$.

(b) a state M on an admissible subspace $\mathcal{X}$ of $B(H_\pi)$ is called an invariant mean on $\mathcal{X}$ if $M(T \cdot_\pi x) = M(T)$ for all $T \in \mathcal{X}$ and $x \in G$.

(c) a state M on a topological admissible subspace $\mathcal{X}$ of $B(H_\pi)$ is called a topological invariant mean on $\mathcal{X}$ if $M(T \cdot_\pi \phi) = M(T)$ for all $T \in \mathcal{X}$ and $\phi \in L^1(G)_1^+$, where

$$L^1(G)_1^+ = \{\phi \in L^1(G) \,|\, \phi \geq 0, \int_G \phi = 1\}.$$

We recall that the concept of amenability for unitary representations as defined in [1]; that is, a unitary representation (π, H_π) of a locally compact group G is called amenable if there exists an invariant mean on $B(H_\pi)$.

Theorem 3.2. *Let (π, H_π) be a unitary representation of a locally compact group G. Let also, $\mathcal{X}$ be a topological admissible subspace of $UCB(\pi)$. Then the following statements are equivalent.*

(a) $\mathcal{X}$ has a topological invariant mean,

(b) There exists a weakly compact positive operator of norm 1 in $\Gamma(\mathcal{X}, G)$.

Remark 3.3. For any unitary representation (π, H_π), let $\mathcal{X}$ is a topological admissible subspace of $B(H_\pi)$. Then automatically $\mathcal{X} \cdot_\pi L^1(G)$ is a topological admissible subspace of $UCB(\pi)$. So, we can replace the role of $\mathcal{X}$ in the Theorem 3.2 by $\mathcal{X} \cdot_\pi L^1(G)$ that $\mathcal{X}$ regarded as a topological admissible subspace of $B(H_\pi)$.

References

[1] M. E. B. Bekka, Amenable unitary representations of locally compact groups, *Invent. Math.* ,100 (1990), 383–401.

[2] P. K. Chan, Topological centers of module actions induced by unitary representations, *J. Funct. Anal.*, 259 (2010), 2193–2214.

[3] G. B. Folland, *A course in abstract harmonic analysis*, CRC Press, Boca Raton, 1995.

[4] S. S. Jafari, *Operators commuting with certain module actions*, Int. J. Nonlinear Anal. Appl., In press, (2022).

[5] E. B. Johnson, *Cohomology in Banach algebras*, (Mem. Amer. Math. Soc. **127**, 1972).

[6] A. T. Lau, Operators which commute with convolutions on subspace of $L^\infty(G)$, *Colloquium Math.*, 39 (1978), 357–359.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

A numerical method for solving fractional one-dimensional Dirac operator

Mohammad Shahriari[1,*], Behzad Nemati Saray[2], and Bahareh Mohammadalipour[1]

[1]Department of Mathematics, Faculty of Science, University of Maragheh, Maragheh, Iran.

[2]Department of Mathematics, Institute for Advanced Studies in Basic Sciences, (IASBS), Zanjan, 45137-66731, Iran.

Abstract

In this manuscript, the pseudospectral method for solving the one–dimensional Caputo fractional Dirac operator are presented. The base of this method is transforming the problem to a weakly singular Volterra integro-differential equation. For this purpose first, the matrices obtained from the representation of the fractional integration operator based on Chebyshev cardinal functions. To obtain approximation of the eigenvalues of the problem, the roots of the characteristics matrix function are find. Finally, Some numerical examples are presented to illustrate the ability and accuracy of the method.

Keywords: Dirac operator, fractional differential equation, pseudospectral method.
Mathematics Subject Classification [2010]: 34B24, 34B27

1 Introduction

Let us consider the system of fractional Dirac operator

$$\ell_\alpha[y(t)] := B\,{}^C\mathcal{D}_0^\alpha y(t) + \Omega(t)y(t) = \lambda y(t), \quad t \in (0,1) \tag{1}$$

with the boundary conditions

$$U(y) := r_{11}y_1(0) + r_{12}y_2(0) = 0, \tag{2}$$
$$V(y) := r_{21}y_1(1) + r_{22}y_2(1) = 0, \tag{3}$$

where $\frac{1}{2} < \alpha \leq 1$,
λ is the spectral parameter,

$$B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \ \Omega(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{bmatrix}, \text{ and } \ y(t) = (y_1(t), y_2(t))^T.$$

Throughout this paper, $p_{ij}(t)$ in $C(0,1)$, and r_{ij} are real, for $i,j = 1,2$. Here ${}^C\mathcal{D}_a^\alpha$ denotes the Caputo fractional differential operator of order α.

*Speaker. Email address: shahriari@maragheh.ac.ir

In the case $\alpha = 1$, if $p_{12}(t) = p_{21}(t) = 0$, $p_{22}(t) = V(t) + m$, and $p_{11}(t) = V(t) - m$, such that, the function $V(t)$ is the potential function and m is the particle mass, then (1) is called a *one–dimensional stationary Dirac system* in relativistic quantum theory.

Applying an orthogonal and smooth transformation of a two dimensional space, we have

$$\Omega(t) = \begin{bmatrix} p(t) & 0 \\ 0 & q(t) \end{bmatrix}, \text{ or } \Omega(t) = \begin{bmatrix} p(t) & q(t) \\ q(t) & -p(t) \end{bmatrix}.$$

These are called the *canonical forms of Dirac operator* [5]. Define the function

$$\Delta(\lambda) := V(\varphi(\lambda)),$$

such that $\varphi(t, \lambda) = (\varphi_1(t, \lambda), \varphi_2(t, \lambda))^T$ is the solution of Eq. (1) satisfying the condition (2). $\Delta(\lambda)$ is called the characteristic function of problem (1)–(3) and does not depend on x.

2 Preliminaries

In this section, we recall some resualts in fractional integral and dirivative and cardinal Chebechev polynomials.

2.1 Definition and theorems of the fractional calculation

We denote the notation ${}^C\mathcal{D}_0^\alpha$ for any $\alpha \in \mathbb{R}^+$ to the left sided Caputo fractional derivative defined by

$${}^C\mathcal{D}_0^\alpha y(t) = \frac{1}{\Gamma(m-\alpha)} \int_0^t (t-x)^{m-\alpha-1} y^{(m)}(x) dx, \quad t > 0, \tag{4}$$

where $m = \lceil \alpha \rceil$ ($\lceil\ \rceil$ is ceiling function) and Γ represents the Euler's gamma function. We give some important information of the fractional calculus theory that will be anxiously used in this paper. At the first, the Riemann–Liouville fractional integral operator of order α are introduced.

Definition 2.1. For the function $y \in L_1[0, T]$, the left sided Riemann–Liouville fractional integral of order α is defined by

$$J^\alpha y(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-x)^{\alpha-1} y(x) dx \tag{5}$$

where $x \leq T$, and $\alpha \in \mathbb{R}^+$.

Lemma 2.2. *For $\alpha \in \mathbb{R}^+$, $m = [\alpha]$, and $f \in L_1[0, T]$, the following relations are satisfied:*

1. ${}^C\mathcal{D}_0^\alpha J^\alpha f(t) = f(t)$,
2. $J^\alpha\, {}^C\mathcal{D}_0^\alpha f(t) = f(t) - \sum_{k=0}^{m-1} f^{(k)}(0+)\frac{t^k}{k!}$,
3. ${}^C\mathcal{D}_0^\alpha t^r = \begin{cases} \frac{\Gamma(r+1)}{\Gamma(r+1-\alpha)} t^{r-\alpha}, & \text{for } [\alpha] < r, \\ 0, & \text{for } [\alpha] \geq r. \end{cases}$

Lemma 2.3 (Lemma 2.1(a), [3])**.** *The integral operators J^α, for $1 \leq p \leq \infty$, are bounded in $L^p([a, b])$ as follows.*

$$\|J^\alpha(g)\|_p \leq K\|g\|_p, \quad K := \frac{(b-a)^\alpha}{\Gamma(\alpha+1)}. \tag{6}$$

2.2 Chebyshev cardinal functions

We consider the Chebyshev polynomials, $T_{m+1}(x) = \cos((m+1)\cos^{-1}x)$ on $[-1,1]$. Suppose that $\mathcal{X}$ be the set of all roots of T_{m+1}, i.e.

$$\mathcal{X} := \{x_j |\ T_{m+1}(x_j) = 0,\ j \in \mathcal{M}\}, \quad \mathcal{M} := \{1, 2, \ldots, m+1\}.$$

It is easy to check that the roots of T_{m+1} are

$$x_j := \cos\left(\frac{(2j-1)\pi}{2m+2}\right), \quad \forall j \in \mathcal{M}. \tag{7}$$

Using variable change $x = 2t-1$, we get the shifted Chebyshev polynomials for the interval $[0,1]$ of the following form

$$T^*_{m+1}(t) := T_{m+1}(2t-1), \quad t \in [0,1]. \tag{8}$$

The roots of $T^*_{m+1}(t)$ are $t_i = \frac{x_i+1}{2}$. Define the Chebyshev cardinal functions of the following form

$$C_j(t) = \frac{T^*_{m+1}(t)}{T^*_{m+1,t}(t_j)(t-t_j)}, \quad j \in \mathcal{M}, \tag{9}$$

where $T^*_{m+1,t}(t_j) = \frac{d}{dt}T^*_{m+1}(t)|_{t=t_j}$.

Using (9), we get

$$C_j(t_i) = \delta_{ji} = \begin{cases} 1, & j = i; \\ 0, & j \neq i. \end{cases} \tag{10}$$

Suppose that $g(t)$ is a given function of $L^2[0,1]$, applying (10), one can approximate this function as follows

$$g(t) \approx \sum_{i=1}^{m+1} g(t_i)C_i(t). \tag{11}$$

The cardinal function $C_i(t)$ can be rewrithen as follows

$$C_i(t) = \eta_i \prod_{k=1, k\neq i}^{m+1} (t - t_k), \tag{12}$$

where $\eta_i = \frac{2^{2m+1}}{T^*_{m+1,t}(t_i)}$.

2.3 Representation of J^α in the Chebyshev cardinal functions

Using the definition 2.1 and from (12), the Chebyshev cardinal functions $\{C_i\}$ are constructed. So, a matrix $\mathbf{I}^\alpha$ of dimension $(m+1)(m+1)$ is obtained of the following form

$$J^\alpha \mathcal{C}(t) = \mathbf{I}^\alpha \mathcal{C}(t). \tag{13}$$

Our purpose is to determine the entries of the matrix $\mathbf{I}^\alpha$. It may be shown that the (i,j)-th element of this matrix is given by

$$\mathbf{I}^\alpha_{i,j} = J^\alpha C_i(t_j). \tag{14}$$

So, by a simple calculation, we get

$$\mathbf{I}^\alpha_{i,j} = \eta \sum_{k=0}^{m} r_{i,k} \frac{\Gamma(m-k+1)}{\Gamma(m-k+\alpha+1)} t_j^{m-k+\alpha}. \tag{15}$$

The matrix representation of $\mathbf{I}^\alpha$ is invertible, because the operator J^α is an invertible [1].

3 Formulas and method

Actually, in this section we formulated a set of the fractional differential equations (1) with the boundary conditions (2). Unfortunately, calculating the exact solution, eigenvalues and eigenfunctions of Eqs. (1) and (2) is too difficult. Therefore, the numerical methods must be proposed to solve such problems. If $y(t) \in C^n[0,1]$, from [3] we get

$$J^{\alpha}\,{}^{C}\mathcal{D}_0^{\alpha}(y)(t) = y(t) - \sum_{i=0}^{n-1} \frac{y^{(i)}(0)}{i!} t^i, \tag{16}$$

where $n = -[-\alpha]$. The Eq. (16) helps us to reduce attention fractional Dirac operator (FDP) (1) to the following Volterra integro-differential equation

$$B(y(t) - y(0)) + J^{\alpha}\left(\Omega(t)y(t) - \lambda y(t)\right) = 0. \tag{17}$$

Here we either have the values $y(0)$ from (2) or we will consider them unknown.

3.1 Pseudospectral method

For $m \in \mathbb{N}_0$, the space P^m is defined all polynomials of degree Less than or equal to m. According to the Pseudospectral method, we first introduce the projection operator $\mathcal{P}$ that maps any continuous function onto P^m. Thus the solution of equation (17) can be approximated by the operator P^m, i.e.,

$$y(t) \approx \mathcal{P}(y)(t) = Y^T \otimes \mathcal{C}(t) := y_m(t), \tag{18}$$

where $Y = [Y_1, Y_2]^T$ is a vector of dimension $2m+2$. Also, the vectors Y_i, for $i = 1, 2$, can be obtained by

$$y_i(t) \approx \mathcal{P}(y_i)(t) = Y_i^T \mathcal{C}(t), \quad i = 1, 2.$$

Replacing (18) into (17), we have

$$B(y_m(t) - \bar{y}(t)) + \mathcal{J}^{\alpha}\left(\Omega y_m + \lambda y_m\right)(t) = 0, \tag{19}$$

where $\bar{y} = [\bar{y}_1, \bar{y}_2]$ and $\bar{y}_i(t) := \sum_{j=0}^{n-1} \frac{y_i^{(j)}(0)}{j!} t^j$ for $i = 1, 2$.

So we use the 2 boundary conditions that we have not used yet, to get a system with $2(m+2)$ equations and $2(m+2)$ unknowns as follows.

$$\Lambda(\lambda)\bar{Y} = 0, \tag{20}$$

where $\bar{Y}$ consists of unknowns and $\Lambda(\lambda)$ is a matrix function of λ. Since equation (1) have nonzero eigenvectors then $\Lambda(\lambda)$, when λ be an eigenvalue of (1), should be singular. Equivalently, we have

$$\det(\Lambda(\lambda)) = 0. \tag{21}$$

Note that $\det(\Lambda(\lambda))$ is a polynomial of degree $2m+2$ and all roots of the polynomial is an approximation of eigenvalues. To find the eigenvalues, we apply the Maple's fsolve command.
In order to find the eigenvector $\bar{Y}$ corresponding to eigenvalues, Noting that $\bar{Y} \in ker\{\Lambda(\lambda)\}$ where

$$\ker\{\Lambda(\lambda)\} = \{\bar{Y} \in Q^{m+1+L} : \Lambda(\lambda)\bar{Y} = 0\},$$

where $Q := \mathbb{R}$ or $Q := \mathbb{C}$. It is easy to see that the matrix $\Lambda(\lambda)$ has a non-zero kernel (because $\bar{Y}$ should be non-zero). In action, the function $\det(\Lambda(\lambda))$, for each choice λ, is not totally equal to zero and has a value close to zero. Thus we select the eigenvector corresponding to the eigenvalues of $\Lambda(\lambda)$. Thus, we obtain the eigenfunction $y_m(t)$ corresponding the particular λ, via

$$y_m(t) = \frac{\sum_{i=1}^{m+1} \bar{Y}_i C_i(t)}{\| \sum_{i=1}^{m+1} \bar{Y}_i C_i(t) \|_2}.$$

Suppose that y_m be the solution of problem (1)-(2), approximated by the proposed method introduced by the previous subsection. Thus, we get

$$By_m(t) - B\bar{y}_m(t) + \mathcal{J}^{\alpha}\left(\Omega y_m + \lambda y_m\right)(t) \approx 0, \tag{22}$$

where $\bar{y}_m = \mathcal{P}(\bar{y})$. Subtracting (22) from (17), and using $e_m := y - y_m$, we have

$$E_m(t) := Be_m(t) - B(\bar{y}(t) - \bar{y}_m(t)) + \mathcal{J}^{\alpha}\left(\Omega e_m + \lambda e_m\right)(t). \tag{23}$$

If λ is an eigenvalue of (1)-(2), the corresponding eigenfunction is not equal to zero, i.e. $y_m(t) \neq 0$.

Theorem 3.1. *Assume that $y(t) \in H^n([0,1])$ for $n > 1$ (sufficiently smooth function) is the exact solution of (1) and the approximate solution $y_m(t)$ is given by the proposed method. Thus the L^2–error $\|E_m(t)\|_2$ decays exponentially in m, i.e.,*

$$\lim_{m\to\infty} \|E_m(t)\|_2 \to 0. \tag{24}$$

4 Numerical results

In this section, we consider some examples to show the performance and efficiency of these algorithms.

Table 1: The numerical values of the first 8 eigenvalues in example 4.1 for $n = 10$, 15 and $\alpha = 1$, 0.95, 0.9, 0.85

k	λ_k^{10} exact,$\alpha = 1$	Λ_k^{10} $\alpha = 1$	Λ_k^{10} $\alpha = 0.95$	Λ_k^{10} $\alpha = 0.9$	Λ_k^{10} $\alpha = 0.85$
-4	-10.99557429	-10.98881693	-9.78449452	-8.77944334	-7.97259590
-3	-7.85398163	-7.85354167	-7.10896762	-6.47846114	-5.94454161
-2	-4.71238898	-4.71239232	-4.38186989	-4.10909726	-7.97259590
-1	-1.57079632	-1.57079632	-1.52765229	-1.49397669	-1.47078426
1	1.57079632	1.57079632	1.52765229	1.49397669	1.47078426
2	4.71238898	4.71239232	4.38186989	4.10909726	3.89581424
3	7.85398163	7.85354167	7.10896762	6.47846114	5.94454161
4	10.99557429	10.98881693	9.78449452	8.77944334	7.97259590
k	λ_k^{15} exact, $\alpha = 1$	λ_k^{15} $\alpha = 1$	λ_k^{15} $\alpha = 0.95$	λ_k^{15} $\alpha = 0.9$	λ_k^{15} $\alpha = 0.85$
-4	-10.99557429	-10.99556936	-9.78927604	-8.78559342	-7.97989364
-3	-7.85398163	-7.85398167	-7.10715868	-6.47385186	-5.93563436
-2	-4.71238898	-4.71238898	-4.38067552	-4.10674607	-3.89253867
-1	-1.57079632	-1.57079633	-1.52717253	-1.49283728	-1.46872383
1	1.57079632	1.57079633	1.52717253	1.49283728	1.46872383
2	4.71238898	4.71238899	4.38067552	4.10674607	3.89253867
3	7.85398163	7.85398167	7.10715868	6.47385186	5.93563436
4	10.99557429	10.99556936	9.78927604	8.78559342	7.97989364

Example 4.1. Consider the FDP (1) with $\Omega(t) = 0$ and the boundary conditions

$$y_2(0) = 0, \qquad y_1(1) = 0.$$

Example 4.2. Consider the FDP (1) with

$$\Omega(t) = \begin{bmatrix} \cos(t) & t^2 \\ t^2 & -\cos(t) \end{bmatrix}$$

and the boundary conditions

$$y_1(0) = 0, \qquad y_1(1) - y_2(1) = 0.$$

Table 2: The numerical values of the first 8 eigenvalues in example 4.2 for $n = 10$, 15 and $\alpha = 1$, 0.95, 0.9, 0.85

k	Λ_k^{10} $\alpha = 1$	Λ_k^{10} $\alpha = 0.95$	Λ_k^{10} $\alpha = 0.9$	Λ_k^{10} $\alpha = 0.85$
-4	-12.55002225	-10.94099079	-9.52893212	-8.12847207
-3	-9.36861847	-8.29506386	-7.42432642	-6.83926111
-2	-6.23088128	-5.58819887	-5.02840738	-4.51356233
-1	-3.07586188	-2.85206955	-2.68112975	-2.57468620
1	0.33576536	0.35673261	0.37562432	0.39243154
2	3.33228505	3.24418979	3.17773999	3.13661805
3	6.37363800	5.93850368	5.56584722	5.24396650
4	9.49771151	8.64984174	7.95040680	7.41762521
k	Λ_k^{15} $\alpha = 1$	Λ_k^{15} $\alpha = 0.95$	Λ_k^{15} $\alpha = 0.9$	Λ_k^{15} $\alpha = 0.85$
-4	-12.51492477	-10.92554518	-9.56061315	-8.21343673
-3	-9.37370510	-8.29670835	-7.42260392	-6.82785392
-2	-6.23101265	-5.58637166	-5.02444405	-4.50713905
-1	-3.0758608	-2.85095149	-2.67878402	-2.57099104
1	0.33576535	0.35695644	0.37616221	0.39340152
2	3.33228514	3.24356372	3.17631473	3.13411340
3	6.37363662	5.93721096	5.56337757	5.24071463
4	9.49889185	8.64860192	7.94586130	7.40667784

References

[1] R. A. Beezer, *A First Course in Linear Algebra*, Congruent Press; 3rd edition, 2012.

[2] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral methods fundamentals in single domains*, Springer-Verlag, Berlin, 2006.

[3] A. Kilbas, H. M. Srivastava and J. J. Trujillo, *Theory and applications of fractional differential equations*, 24. Elsevier B. V., Amsterdam, 2006.

[4] F. Mainardi and R. Gorenflo, *Time-fractional derivatives in relaxation processes: a tutorial survey*, arXiv preprint arXiv:0801.4914, (2008).

[5] I. Sargsjan et. al., *Sturm–Liouville and Dirac Operators*, Springer Science and Business Media, 2012.

[6] M. Shahriari, and J. Manafian, An efficient algorithm for solving the fractional Dirac differential operator. *Adv. Math. Models Appl.,* 5(3), (2020), 289-297.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

A new class of second derivative multistep methods for stiff ODEs

M. Egbaljoo*, G. Hojjati

Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran

Abstract

In this paper, we study a new class of linear methods for the numerical solution of initial value problems in ordinary differential equations. These methods have multivalue structure in which the second derivative of the solution together with a free parameter have been included. Deriving the stability matrix of the methods and checking its eigenvalues, we aim to fine the free parameter so that construct methods with better stability properties of high convergence order.

Keywords: Initial value problem, Second derivative methods, Stability, Stiff systems.
Mathematics Subject Classification [2010]: 65L05.

1 Introduction

In recent years, a number of studies have explored the development of more advanced and efficient methods for the solution of stiff problems

$$\begin{aligned} y'(x) &= f(x, y(x)), \qquad x \in [t_0, X], \\ y(x_0) &= y_0, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ is continuous and m is the dimensionality of the system. In the class of linear multistep methods (LMMs) many methods have been introduced that solve stiff initial value problems with good accuracy and a reasonably wide region of absolute stability. Most of these improvments have been achieved through the use of backward differentitaion formulae (BDF), which have several interesting properties. In order to improve the stability of BDF, Fredebeul [4] introduced the A-BDF method to provide better stability properties. A free parameter was added to the EBDF algorithm by Hojjati et al. [6] to increase the stability region of the A-BDF and EBDF algorithms. Using of the second derivative of the solution, more advanced methods have been introduced by Enright [3] and Cash [1].

In this paper, we are going to introduce a new class of algorithms based on second derivative BDF (SDBDF) methods. Deriving the stability matrix of the methods and its eigenvalues, we find the free parameter to construct methods with a wider region of absolute stability.

*Speaker. m.eghbaljoo@tabrizu.ac.ir

2 A-SDBDF method

In this section, we introduce a new class of methods which we call A-SDBDF methods. In this methods, we insert a free parameter t into the algorithm of the SDBDF method in order to improve the stability properties. The methods are defined as

$$A-SDBDF := \overline{SDBDF} - t \cdot \widehat{SDBDF},$$

in which

$$\overline{SDBDF} := \sum_{j=0}^{k} \overline{\alpha}_j y_{n+j} = h\overline{\beta}_k f_{n+k} + h^2 \overline{\gamma}_k g_{n+k},$$

$$\widehat{SDBDF} := \sum_{j=0}^{k} \widehat{\alpha}_j y_{n+j} = h\widehat{\beta}_{k-1} f_{n+k-1} + h^2 \widehat{\gamma}_{k-1} g_{n+k-1},$$

are respectively the implicit and explicit SDBDF methods.

We are going to study A-SDBDFs as SGLMs form. An SGLM used for the numerical solution of (1) is given by

$$\begin{aligned} Y^{[n]} &= h(\mathbf{A} \otimes \mathbf{I}_m) f(Y^{[n]}) + h^2(\overline{\mathbf{A}} \otimes \mathbf{I}_m) g(Y^{[n]}) + (\mathbf{U} \otimes \mathbf{I}_m) y^{[n-1]}, \\ y^{[n]} &= h(\mathbf{B} \otimes \mathbf{I}_m) f(Y^{[n]}) + h^2(\overline{\mathbf{B}} \otimes \mathbf{I}_m) g(Y^{[n]}) + (\mathbf{V} \otimes \mathbf{I}_m) y^{[n-1]}, \end{aligned} \tag{2}$$

where $n = 1, 2, \cdots, N$, $Nh = x - x_0$, h is the stepsize and $\otimes$ is the Kronecker product of two matrices. Here $\mathbf{A}, \overline{\mathbf{A}} \in \mathbb{R}^{s\times s}$, $\mathbf{U} \in \mathbb{R}^{s\times r}$, $\overline{\mathbf{B}} \in \mathbb{R}^{r\times s}$, $\mathbf{V} \in \mathbb{R}^{r\times r}$.

Now for the algorithm based on formulas (2), written in the SGLM form (2) with $s = 1$, $r = k + 2$, the vectors of internal approximations $Y^{[n]}$, $f(Y^{[n]})$, $g(Y^{[n]})$, and the vector of external approximations $y^{[n]}$ are defined by

$$Y^{[n]} = \left[y_{n+k}\right], \quad f(Y^{[n]}) = \left[f_{n+k}\right], \quad g(Y^{[n]}) = \left[g_{n+k}\right], \quad y^{[n]} = \begin{bmatrix} y_{n+k} \\ y_{n+k-1} \\ \vdots \\ y_{n+1} \end{bmatrix},$$

and the coefficient matrices $\mathbf{A}$, $\overline{\mathbf{A}}$, $\mathbf{U}$, $\mathbf{B}$, $\overline{\mathbf{B}}$, $\mathbf{V}$ are given by

$$\mathbf{A} = \left[\frac{\overline{\beta}_k}{1-t}\right], \quad \overline{\mathbf{A}} = \left[\frac{\overline{\gamma}_k}{1-t}\right],$$

$$\mathbf{U} = \begin{bmatrix} \dfrac{-(\overline{\alpha}_{k-1} - t\widehat{\alpha}_{k-1})}{1-t} & \dfrac{-(\overline{\alpha}_{k-2} - t\widehat{\alpha}_{k-2})}{1-t} & \cdots & \dfrac{-(\overline{\alpha}_0 - t\widehat{\alpha}_0)}{1-t} & \dfrac{-\widehat{\beta}_{k-1}}{1-t} & \dfrac{\widehat{\gamma}_{k-1}}{1-t} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \dfrac{\overline{\beta}_k}{1-t} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \overline{\mathbf{B}} = \begin{bmatrix} \dfrac{\overline{\gamma}_k}{1-t} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\mathbf{V}=\begin{bmatrix} \frac{-(\overline{\alpha}_{k-1}-t\widehat{\alpha}_{k-1})}{1-t} & \frac{-(\overline{\alpha}_{k-2}-t\widehat{\alpha}_{k-2})}{1-t} & \cdots & \frac{-(\overline{\alpha}_{0}-t\widehat{\alpha}_{0})}{1-t} & \frac{-\widehat{\beta}_{k-1}}{1-t} & \frac{\widehat{\gamma}_{k-1}}{1-t} \\ 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix},$$

where $\mathbf{A}$, $\overline{\mathbf{A}} \in \mathbb{R}^{1\times 1}$, $\mathbf{U} \in \mathbb{R}^{1\times k+2}$, $\overline{\mathbf{B}} \in \mathbb{R}^{k+2\times 1}$, $\mathbf{V} \in \mathbb{R}^{k+2\times k+2}$.

To study the stability properties of the constructed methods, we first recall that the stability polynomial of SGLM (2) is defined as

$$p(w,z) = det(wI - M(z)),$$

where

$$M(z) = V + (zB + z^2\overline{\mathbf{B}})(I - zA - z^2\overline{\mathbf{A}})^{-1}U,$$

is the stability matrix. The stability polynomial for A-SDBDF can be found in the form

$$p(w,z) = \frac{1}{(-\overline{\alpha}_k + t\widehat{\alpha}_k + z\overline{\beta}_k + z^2\overline{\gamma}_k)} \sum_{j=0}^{k} C_j(z)w^j. \tag{3}$$

A-SDBDF methods are A-stable up to order $p = 4$ ($k = 3$) and $A(\alpha)$-stable up to order $p = 11$ ($k = 10$).

Table 1: The angles of $A(\alpha)$-stability of BDF, A-BDF, A-EBDF and A-SDBDF methods for $k = 1, 2, \dots, 10$.

	A-BDF		A-EBDF		SDBDF		A-SDBDF	
k	p	α	p	α	p	α	p	α
1	1	90°	2	90°	2	90°	2	90°
2	2	90°	3	90°	3	90°	3	90°
3	3	89.99°	4	90 °	4	90°	4	90°
4	4	85.94°	5	89.14°	5	89.36°	5	89.83°
5	5	73.20°	6	83.94°	6	86.35°	6	86.43°
6	6	51.62°	7	75.25°	7	80.82°	7	82.35°
7	7	17.47°	8	60.79°	8	72.53°	8	72.90°
8	8	——	9	37.87°	9	60.71°	9	65.30°
9	9	——	10	——	10	43.39°	10	45.32°
10	10	——	11	——	11	12.34°	11	18.65°

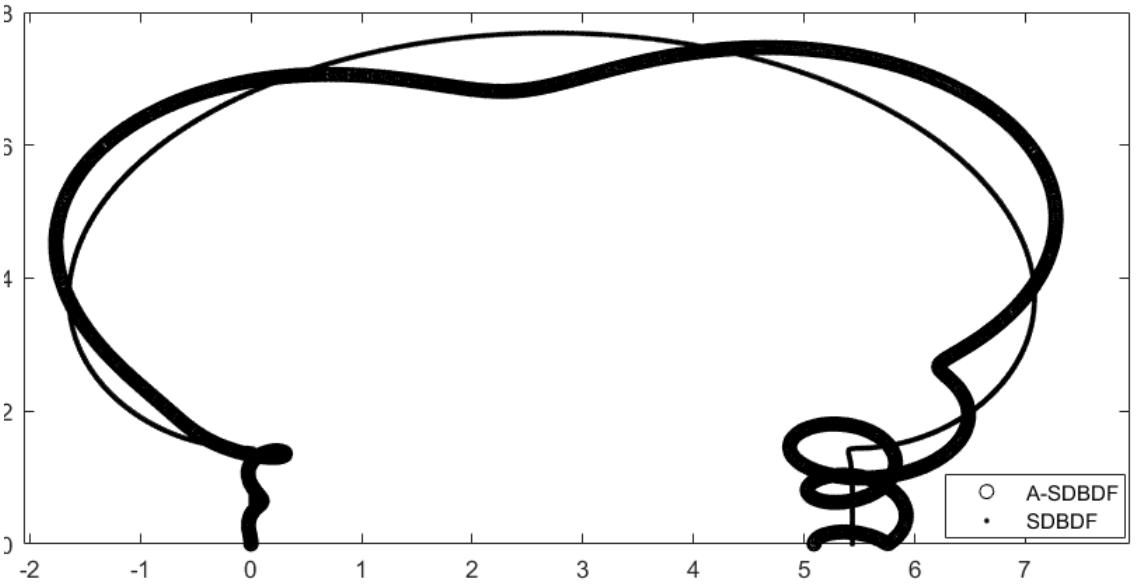


Figure 1: Regions of absolute stability of SDBDF (thin line) and A-SDBDF (medium line) for $k = 8$.

3 Conclusion

The stability properties of the numerical methods for stiff IVPs play an important role in the success of the methods. In this paper, we introduced the new class of A-SDBDF methods which have a free parameter and analyzed their linear stability properties. Representing the methods in SGLMs form, we derived the stability matrix of the methods and found the optimal values for the free parameter to get $A(\alpha)$-stable methods with maximum values of α. The introduced methods, comparison with SDBDF methods, have the same order but with more extensive region of absolute stability.

References

[1] J.R. Cash, *Second derivative extended backward differentiation formula for the numerical integration of stiff systems*, SIAM J. Numer. Anal. 18 (1981) 21–36.

[2] G. Dahlquist, *A special stability problem for linear multistep methods*, BIT 3 (1963) 27–43.

[3] W.H. Enright, *Second derivative multistep methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. 11 (1974) 321–331.

[4] C. Fredebeul, *A-BDF: a generalization of the backward differentiation formulae*, SIAM J. Numer. Anal. 35 (1998) 1917–1938.

[5] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential Algebraic Problems*, Springer, Berlin, 2016.

[6] G. Hojjati, M. Rahimi, S.M. Hosseini, *A-EBDF: an adaptive method for numerical solution of stiff systems of ODEs*, Math. Comp. Simul. 66 (2004) 33–41.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

A FAS multigrid scheme for hyperbolic conservation laws

Javad Farzi*

Department of Mathematics, Sahand University of Technology, Tabriz, Iran

Abstract

In this paper we present a multigrid scheme for time stepping of hyperbolic conservation laws. It is well-known that the solution of hyperbolic conservation laws may encounter with discontinuities or shocks. Therefore, the major concern is to capture the true entropy satisfying unique solution. We show that the provided method fulfills the monotonicity preserving and total variation diminishing (TVD) properties.

Keywords: Multigrid, Hyperbolic problems, Convergence acceleration
Mathematics Subject Classification [2010]: 65M55, 65D05, 65F10

1 Introduction

We consider the nonlinear hyperbolic conservation laws

$$u_t + f(u)_x = 0, \tag{1}$$

with the initial data $u(x,0) = u_0(x)$ and appropriate boundary conditions, where f is the flux function. This formulation, models any problems in science and engineering such as traffic model, Chemical convection, electromagnetic and so on. The solution of the hyperbolic problems is characterized by their eigenvalues and eigenvectors. Therefore, the flux function play a major role in identifying the different behaviors in the solution. We suppose that the flux is a convex function, which deals with a situation that a shock or rarefaction wave may appear in the solution. The first order upwind method has a non-oscillatory behavior, however, by increasing the order of the accuracy, we observe a serious nonphysical oscillations in the solution. To avoid such unpleasant quality we need to enforce a monotonicity and TVD properties. In other words, the oscillations would be disappeared with adding some sort of diffusion in the solution.

Consider the finite volume method for (1) on a computational domain Ω_x:

$$\frac{du_j}{dt}h_j + (f_{j+1/2} - f_{j-1/2}) = 0, \tag{2}$$

where, u_j may refer to a point value or cell average of the solution function and $f_{j\pm1/2}$ is the numerical flux function..

Multigrid techniques are a powerful tools for elliptic partial differential equations and an extensive study has been done to establish the convergence theory, coarsing, interpolation, smoothing. In next section we introduce the multigrid scheme for hyperbolic conservation laws.

*Speaker. Email address: farzi@sut.ac.ir

2 Full approximation scheme (FAS)

Multigrid methods has shown to be an efficient technique for solving elliptic partial differential equations. The smoothing nature of elliptic equations is well simulated by reducing the high and low frequncies errors. However, the fundamental wave propagation nature of hyperbolic equations is also should be reflected in multigrid methods for hyperbolic problems. In spite the elliptic equations, the discretization matrices of hyperbolic equations are in general non-symmetric. Although the main goal of the multigid schemes is to rapidly spell out the low frequency disturbancesof the boundary, we observe that the numerical oscillations cause a delay in propagation. In [1], Wan and Jameson demonstrated such a situation with an example by advection equation, where initial condition is a square wave. We follow the methodology of [1, 2] to present a monotonicity preserving scheme with a conservative scheme.

We consider the following general wave propagation problem with initial and inflow boundary conditions

$$\begin{aligned} u_t + f(u)_x &= h(x), \quad 0 < x < 1,\ t > 0, \\ u(0,t) &= g, \quad t > 0, \\ u(x,0) &= u^0(x), \quad 0 < x < 1, \end{aligned}$$

where f is the flux function, h and g are known functions independent of time. To run a multigrid platform we write (3) in a semi-discrete form by summation by part (SBP) upwind form. For simplicity we consider the linear wave propagation problem by the linear flux $f(u) = u$. The solution of this problem is right-going wave that enables us to incorporate the upwind methodology. We describe the backbone of the multigrid scheme for (3).

2.1 Evolution

The main part of a numerical scheme is how the data evolve in next steps.

Definition 2.1. Let the diagonal matrix P defines a discrete norm.The difference operators $D_+ = P^{-1}(Q_+ + \frac{B}{2})$ and $D_- = P^{-1}(Q_- + \frac{B}{2})$, with $B = \boldsymbol{e}_N \boldsymbol{e}_N^T - \boldsymbol{e}_0 \boldsymbol{e}_0^T$, N being the number of partitions, are said to be pth order diagonal-norm upwind SBP operators for the first derivative if

(i) $D_\pm$ is pth and $[p/2]$th order accurate in the interior and at the left/right boundary,

(ii) $Q_+ + Q_-^T = 0$,

(iii) $Q_\pm + Q_\pm^T$ are positive/negative semi-definite.

With and equidistant grid $\Omega_1 = \{x_j = j\Delta x,\ j = 0, 1, 2, \ldots, N\}$ on $[0, 1]$, where $N\Delta x = 1$, and discretization of u_x with upwind SBP operators we obtain

$$\begin{aligned} \boldsymbol{U}_t + D_+\boldsymbol{U} &= \boldsymbol{h} - P^{-1}(U_0 - g)\boldsymbol{e}_0, \quad t > 0, \\ \boldsymbol{U}(0) &= \boldsymbol{U}^0, \end{aligned}$$

where the we have used SAT method for imposing the physical BC [3]. In (3), $U = [U_0, U_1, \ldots, U_N]^T$ is a vector indicating the approximate solution: $Uj(t) \approx u(x_j, t)$ and $\boldsymbol{h} = [h(x_0), h(x_1), \ldots, h(x_N)]^T$.

Applying Euler Forward (EF) we get

$$\boldsymbol{U}^{n+1} = S_1\boldsymbol{U}^n + (I_1 - S_1)L_1^{-1}\boldsymbol{F},$$

where, $\boldsymbol{F} = \boldsymbol{h} + g(\Delta x)^{-1}\boldsymbol{e}_0$, $\boldsymbol{U}^n$ denotes the approximation at time $t^n = n\Delta t$, and we have

$$L_1 = D_+ + P^{-1}\boldsymbol{e}_0\boldsymbol{e}_0^T = \frac{1}{\Delta x}\begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix},$$

and

$$S_1 = I_1 - \Delta t L_1 = \begin{bmatrix} 1-\lambda & & & \\ \lambda & 1-\lambda & & \\ & \ddots & \ddots & \\ & & \lambda & 1-\lambda \end{bmatrix}.$$

For fourth order Runge-Kutta schemes it is suffices to replace S_1 as follow

$$S_1 = I_1 - \Delta t L_1 + \frac{1}{2}(\Delta t L_1)^2 - \frac{1}{6}(\Delta t L_1)^3 + \frac{1}{24}(\Delta t L_1)^4.$$

Note that the from second up to ninth order upwind SBP operators are availabe in [4].

2.2 Fine and Coarse grid

We use the subscript 1 for finest grid and define the similar grid levels. Accordingly, we name matrices in first grid by $L_1, S_1, ...$, and similarly for next level grids. We outline the tools for describing a two level algorithm:

- A coarse grid $\Omega_2 = \{x_j^{(2)} = j\Delta x^{(2)}, j = 0, 2, \ldots, N\}$. Note that $\Delta x^{(2)} = 2\Delta x^{(1)}$;
- Restriction operator $\mathcal{R}_u$ such that $(\mathcal{R}_u\boldsymbol{v}^{(1)})_j = v_j^{(2)}$, $j = 0, 2, \ldots, N$,
- Residual restriction operator I_r transfers the data from Ω_1 to Ω_2 witch uses a first order interpolation.
- The prolongation operator $I_p = I_p^I + I_p^E$, is define on the fine grid that split the nodes for coarse (acting by I_p^I) and other grid points (acting by I_p^E.

For more details we refer to [5].

3 Numerical experiments

In this section we present an example of two-grid solution of linear wave propagation problem with $h(x) = 0$, $g(x) = 0$ and initial data $u^0(x) = 1$. The solution is illustrated in Figure 1.

4 Conclusion

In this paper we have presented the performance of a multigrid scheme for wave propagation scheme. The results are TVD and we don't observe non-physical oscillations.

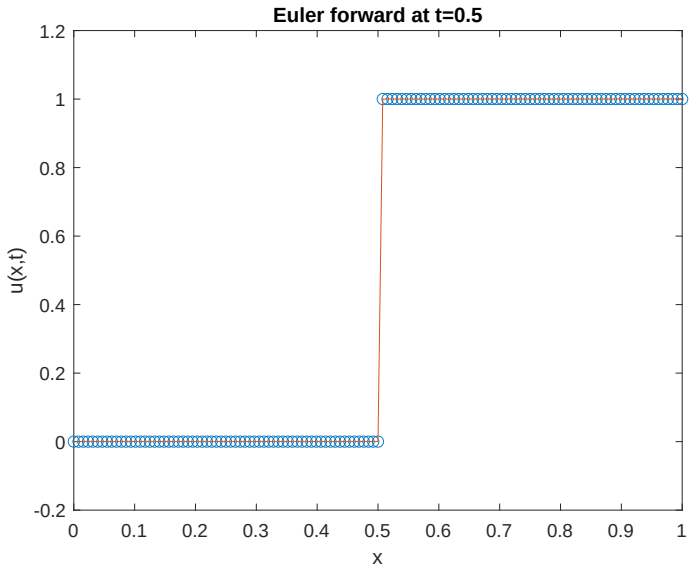


Figure 1: Two-grid solution of linear wave propagation with Euler Forward time marching.

References

[1] J. W. L. Wan, A. Jameson, Monotonicity preserving multigrid time stepping schemes for conservation laws, *Computing and Visualization in Science* , 11 (2008), 41–58.

[2] P. Lötstedt, B. Gustafsson, Fourier analysis of multigrid methods for general systems of PDEs, *Math. Comp.*, 60, (1993) 473–493.

[3] M. Svärd, J. Nordström, Review of summation-by-parts schemes for initial–boundary-value problems, *J. Comput. Phys.* 268, (2014) 17–38.

[4] K. Mattsson, Diagonal-norm upwind SBP operators, *J. Comput. Phys.*, (2017), 335, 283–310.

[5] A. A. Ruggiu, J. Nordström1, Multigrid Schemes for High Order Discretizations of Hyperbolic Problems, *Journal of Scientific Computing*, 82, Article number: 62 (2020).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Construction of completely positive matrices

Kazem Ghanbari[1,2] and Hanif Mirzaei[1,*]

[1]Department of Mathematics, Sahand University of Technology, Tabriz, Iran

[2]School of Mathematics and Statistics, Carleton University, Ottawa, Canada

Abstract

If a symmetric matrix A can be factorized of the form $A = BB^T$ where B is an entry wise nonnegative matrix, then A is called a Completely Matrix (CP). Completely positive matrices have arisen in some situations in economic modelling and appear to have some applications in statistics, and they are also appear in quadratic optimization. If we pick a random matrix, most probably it is not CP. In this paper we give an algorithm to construct a CP matrix from two given nonnegative spectrum.

Keywords: Complete Positivity, Jacobi Matrix
Mathematics Subject Classification [2010]: 15A48, 15A23

1 Introduction

The concept of positivity in matrix theory is one of the well-studied topics. Positive definite matrices are quite well-known for most of mathematicians.

If a symmetric matrix A can be factorized of the form $A = BB^T$ where B is a nonnegative matrix, then A is called a *Completely positive* (CP) matrix. Completely positive matrices have arisen in some situations in economic modelling and appear to have some applications in statistics, and they are also appear in quadratic optimisation, for more details see [2]. There is no golden algorithm to determine if a given matrix is CP. Therefore it is quite interesting to see how can we construct a CP matrix from given data. The set of eigenvalues of a matrix A is called the spectrum of A. Any procedure leading to construction of a matrix from spectral data is called *inverse eigenvalue problem*. Inverse eigenvalue problem is well-studied for some classes of special structured matrices such as tridiagonal and pentadiagonal matrices, for example see [3]

2 Preliminary Materials

In this section we give some definitions and preliminary materials.

Definition 2.1. If A is nonnegative and positive definite, then A is called *doubly nonnegative* (DNN). The set of CP matrices of size n is denoted by CP_n and the set of all (DNN) matrices of size n is denoted by DNN_n.

*Speaker. Email address: h_mirzaei@sut.ac.ir

By definition it is clear that $CP_n \subset DNN_n$ but the reverse is not true in general. We use the following notation for nonnegative orthant of $\mathbb{R}^n$

$$\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, for \quad i = 1, 2, \cdots, n\}.$$

Definition 2.2. If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n \in \mathbb{R}^n$ then the matrix A defined by

$$a_{ij} = \mathbf{v}_i^T \mathbf{v}_j = Gram(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$$

is called *Gram matrix*.

The following result is well know for positive semi-definite matrices:

Theorem 2.3. *[1, 2] Let A be an $n \times n$ symmetric matrix such that $rank(A) = k$. Then the following statements are equivalent*

(1) A is positive semi-definite.

(2) Any eigenvalue of A is nonnegative.

(3) There exists a lower triangular $n \times n$ matrix L such that $A = LL^T$.

(4) There exists some $n \times k$ matrix B such that $A = BB^T$.

(5) There exists a $k-$ dimensional vector space V and vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n \in V$ such that $A = Gram(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$.

(6) There exist k vectors $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_k \in \mathbb{R}^n$ such that $A = \sum_{i=1}^k b_i b_i^T$.

Definition 2.4. The comparison matrix of an $n \times n$ matrix A denoted by $M(A)$ is defined as follows

$$M(A)_{ij} = \begin{cases} |a_{ij}|, & \text{if } i = j \\ -|a_{ij}|, & \text{if } i \neq j. \end{cases}$$

3 Main results

In this section we present some results for CP matrices. Moreover, we give an algorithm for constructing a CP matrix from two given sets of interlacing positive real numbers.

3.1 Completely Positive Matrices

In contrary to Positive Definite (PD) matrices and entrywise nonnegative matrices, it is not easy to determine if a given matrix is CP or not. In this section we give some characteristics of CP matrices. Based on the definition of CP matrix A is a symmetric matrix with decomposition of the form

$$A = BB^T = \begin{bmatrix} b_1 & b_2 & \cdots & b_k \end{bmatrix} \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_k^T \end{bmatrix} = \sum_{i=1}^k b_i b_i^T, \tag{1}$$

where $b_i \geq 0$. This representation of A is known as *rank 1 representation*. If we partition B with rows $\tilde{b_1}, \tilde{b_2}, \cdots, \tilde{b_n}$, then we have the following representation for A

$$A = BB^T = \begin{bmatrix} \tilde{b_1}^T \\ \tilde{b_2}^T \\ \vdots \\ \tilde{b_k}^T \end{bmatrix} \begin{bmatrix} \tilde{b_1} & \tilde{b_2} & \cdots & \tilde{b_k} \end{bmatrix} = \begin{pmatrix} < \tilde{b_1}, \tilde{b_1} > & \cdots & < \tilde{b_1}, \tilde{b_n} > \\ \vdots & \ddots & \vdots \\ < \tilde{b_n}, \tilde{b_1} > & \cdots & < \tilde{b_n}, \tilde{b_n} > \end{pmatrix} \tag{2}$$

The last matrix is called Gram matrix. Therefore we have the following theorem for CP matrices.

Theorem 3.1. *The following conditions are equivalent for any $n \times n$ CP matrix A:*

(1) $A = BB^T$ for some $B \in \mathbb{R}_+^{n \times k}$.

(2) $A = \sum_{i=1}^{k} b_i b_i^T$, where $b_i \geq 0$ for $i = 1, 2, \cdots, k$.

(3) $A = Gram(\tilde{b_1}, \cdots, \tilde{b_n})$, with $\tilde{b_i} \in \mathbb{R}_+^k$ for $i = 1, 2, \cdots, n$.

Theorem 3.2. *If $f(x)$ is a polynomial with nonnegative coefficients and A is CP, then $f(A)$ is also CP.*

Theorem 3.3. *If A and C are CP then*

(1) $A + C$ is CP and any powers of A is CP.

(2) The Kronecker product $A \otimes C$ is CP.

(3) For any permutation matrix P, the product $P^T AP$ is CP.

As we mentioned above $CP_n \subset DNN_n$ but the reverse is not true in general. But if the comparison matrix $M(A)$ is PD then the inverse statement is also true, i.e.,

Theorem 3.4. *For any $n \times n$ DNN matrix A if the comparison matrix $M(A)$ is positive semi-definite, then A is CP.*

Definition 3.5. An $n \times n$ matrix $A = [a_{ij}]$ is called a tridiagonal matrix if $a_{ij} = 0$ for $|i - j| > 1$.

Tridiagonal matrices are perhaps one of the most studied classes of matrices. There are two reason for this claim. First, tridiagonal matrices represent the discrete form of some important differential equations such as Sturm-Liouville problems. Second, many of algorithms in linear algebra require significantly less computational work when they are applied to tridiagonal matrices. Using some transformations such as Householder transformation a given matrix can bed reduced to tridiagonal matrix without changing the spectrum. Thus tridiagonal matrix can be used in computing eigenvalues of matrices. The determinant of a given tridiagonal matrix $A = [a_{ij}]$ can be evaluated easily by the following recursive formula:

$$\det A = a_{11} \det A\left[\{2, 3, \cdots n\}\right] - a_{12}a_{21} \det A\left[\{3, 4, \cdots n\}\right],$$

which is immediate result of the expansion of determinant across row (or column) 1.

Definition 3.6 (Jocobi Matrix)**.** A *Jacobi matrix* is a tridiagonal matrix of the form

$$A = \begin{pmatrix} a_1 & c_1 & 0 & . & . & . \\ c_1 & a_2 & c_2 & 0 & . & . \\ 0 & c_2 & a_3 & c_3 & 0 & . \\ . & . & . & . & . & . \\ . & . & . & . & . & c_{N-1} \\ . & . & . & . & c_{N-1} & a_N \end{pmatrix} \tag{3}$$

where $c_i > 0$, for $i = 1, 2, ..., N$.

Remark 3.7. Now it is important to note that given a random matrix most probably will not be a CP matrix. The question is how to construct a CP matrix. There exists a systematic procedure to construct a unique positive definite tridiagonal symmetric matrix from two given sets of positive real interlacing sequences $0 < \lambda_1 < \mu_1 < \lambda_2 < \cdots < \mu_{n-1} < \lambda_n$.

Theorem 3.8 (Inverse Eigenvalue Problem)**.** *Let $\{\lambda_i\}_1^N$ and $\{\mu_j\}_1^{N-1}$ be two sets of real numbers which satisfy the following interlacing property*

$$\lambda_1 < \mu_1 < \lambda_2 < \mu_2 < ... < \lambda_{N-1} < \mu_{N-1} < \lambda_N.$$

Then, there exists a Jacobi matrix J such that $\sigma(J) = \{\lambda_i\}_1^N$ and $\sigma(J_{N-1}) = \{\mu_j\}_1^{N-1}$. If the eigenvalues are positive then J is a nonsingular DNN matrix. Moreover, the comparison matrix $M(J)$ is also positive definite. As a result J will be completely positive matrix.

4 Numerical results

Example 4.1. Suppose that the interlacing eigenvalues

$$\{\lambda_i\} = \{0.2538, 1.7923, 3.0000, 4.2077, 5.7462\}, \quad \{\mu_i\} = \{1.2547, 2.8227, 4.1773, 5.7453\},$$

are given. According to Theorem 3.8, using Lanczos algorithm we find the unique tridiagonal matrix J as follows

$$J = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix},$$

where $\{\lambda_i\}$ are the eigenvalues of J and $\{\mu_i\}$ are the eigenvalues of J_1, where J_1 is the submatrix of J by deleting the last row and last column of J. Clearly the matrix J is DNN and the comparison matrix

$$M(J) = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 3 & -1 & 0 \\ 0 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & -1 & 5 \end{bmatrix}$$

is PD. Therefor J is CP.

5 Conclusion

In this paper we introduced the concept of completely positive matrix. There is no golden criterion to determine if a given matrix is CP. We introduced an algorithm for constructing a CP matrix from given spectral data.

References

[1] R. Bhatia, *Positive definite Matrices*, Princeton University Press 2007.

[2] Abraham Berman, Completely Positive Matrices, World Scientific Publishing (2003)

[3] K. Ghanbari, A survey on inverse and generalized inverse eigenvalue problems for Jacobi matrices, Applied Mathematics and Computation, Vol. 195 (2008) 355-363

[4] Graham Gladwell, Inverse problems in vibration, Kluwer Academic Publishers, (2004)

[5] Shaun Fallat and Charles R. Johnson, Totally Nonnegative Matrices, Princeton University Press (2011)

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Frame theory and reproducing kernel Hilbert spaces

Mohammadreza Foroutan[1,*] and Farzad Farzanfar[2]

[1]Department of Mathematics, Payame Noor University, P.O.Box 19395-3697, Tehran, Iran

[2]Department of Computer Engineering and Information Technology, Payame Noor University, P.O. Box 19395-3697 Tehran, Iran

Abstract

In this paper, we discuss various aspects of orthonormal, discrete frame expansions, together with some admissibility conditions with the reproducing kernel Hilbert spaces. We show that all Parseval frames in a Hilbert space are connected by transformations that are unitary between reproducing kernel Hilbert spaces. The purpose of this paper is to present a method for constructing a reproducing kernel Hilbert space and its associated kernel by means of frame theory.

Keywords: Parseval frame, Frame coefficients, Frame operator, Reproducing kernel Hilbert space, Reproducing property.

Mathematics Subject Classification [2010]: 46E22, 42C15, 46C05

1 Introduction

Reproducing kernel Hilbert spaces have developed into an important tool in many areas, especially statistics and machine learning, and they play a valuable role in complex analysis, probability, group representation theory, and the theory of integral operators [3,4,6,7].

Frames have important applications to machine learning. Every frame on a finite dimensional Hilbert space can be viewed as a matrix whose columns are the frame elements. Parseval frames on a finite-dimensional Hilbert space are characterized by the singular value decompositions of their matrix representations; all singular values of any such Parseval frame are 1.

Since the reproducing property of kernel functions is closely related to the reconstruction property of frames, so frames can be used to construct a variety of kernel functions that are useful for kernel method based machine learning algorithms such as support vector machines [7]. In this paper, we study a relationships between frames and reproducing kernel Hilbert spaces.

*Speaker. Email address: mr_forootan@pnu.ac.ir, foroutan_mohammadreza@yahoo.com

2 Reproducing kernel Hilbert spaces

In this section, we introduce the reader to the formal definition of the reproducing kernel Hilbert space and present a few of their most basic properties.

Definition 2.1 ([3]). Let X be a set. We will call a subset $H \subseteq \mathrm{F}(X, \mathbb{F})$ a reproducing kernel Hilbert space on X if

1. H is a vector subspace of $\mathrm{F}(X, \mathbb{F})$;
2. H is endowed with an inner product, $\langle ., . \rangle$, with respect to which H is a Hilbert space;
3. for every $x \in X$, the linear evaluation functional, $\delta_x :\longrightarrow \mathbb{F}$, defined by $\delta_x(f) = f(x)$, is bounded.

If H is a reproducing kernel Hilbert space on X, then an application of the Riesz representation theorem shows that the linear evaluation functional is given by the inner product with a unique vector in H. Therefore, for each $x \in X$, there exists a unique vector, $k_x \in H$, such that for every $f \in H$, $f(x) = \delta_x(f) = \langle f, k_x \rangle$. The function k_x is called the reproducing kernel for the point x and the function $K : X \times X \longrightarrow \mathbb{F}$ defined by $K(x, y) = k_y(x) = \langle k_y, k_x \rangle$. Also,

$$\|\delta_y\|^2 = \|k_y\|^2 = \langle k_y, k_y \rangle = K(y, y).$$

Example 2.2. Let $\{e_1, e_2, \cdots, e_n\}$ be an orthonormal basis of a finite-dimensional Hilbert space H. If we define

$$K(x, y) = \sum_{i=1}^{n} e_i(x) e_i(y), \tag{1}$$

for $x \in X$, then we have $k_x \in H$ and

$$\langle e_j, k_x \rangle_H = \sum_{i=1}^{n} \langle e_j, e_i \rangle_H e_i(x) = e_j(x),$$

for each $1 \leq j \leq n$. Thus, for any $f(.) = \sum_{i=1}^{n} f_i e_i(.) \in H$, $f_i \in \mathbb{R}$, we have $\langle f, k_x \rangle_H = f(x)$ (reproducing property). Therefore, H is a reproducing kernel Hilbert space, and (1) is a reproducing kernel.

In separable Hilbert spaces, countable orthonormal systems are used to expand any element as an infinite sum. In separable reproducing kernel Hilbert space the reproducing kernel can be expressed through orthonormal systems as stated in the following theorem.

Theorem 2.3. *Let H be a separable Hilbert space with reproducing kernel K. For any complete orthonormal system $(\phi_i)_{i \in \mathbb{N}}$ in H, we have*

$$\forall t \in X : \quad K(., t) = \sum_{i=1}^{\infty} \overline{\phi_i(t)} \phi_i(.) \quad (\textit{convergence in } H). \tag{2}$$

Conversely, if (2) holds for an orthonormal system $(\phi_i)_{i \in \mathbb{N}}$ then this system is complete and H is separable. Moreover, (2) implies that

$$\forall s \in X, \ \forall t \in X : \quad K(s, t) = \sum_{i=1}^{\infty} \overline{\phi_i(t)} \phi_i(s) \quad (\textit{convergence in } \mathbb{C}).$$

Proof. A proof is described in Theorem 14 of [1]. □

Theorem 2.4. *Let H be a reproducing kernel Hilbert space on X with reproducing kernel k, let $H_0 \subseteq H$ be a closed subspace and let $P_0 : H \longrightarrow H_0$ be the orthogonal projection onto H_0. Then H_0 is a reproducing kernel Hilbert space on X with reproducing kernel*

$$K_0(x, y) = \langle P_0(k_y), k_x \rangle.$$

Proof. Since evaluation of a point in X defines a bounded linear functional on H, it remains bounded when restricted to the subspace H_0. Thus, H_0 is a reproducing kernel Hilbert space on X. Let $f \in H_0$, we have

$$f(x) = \langle f, k_x \rangle = \langle P_0(f), k_x \rangle = \langle f, P_0(k_x) \rangle.$$

Hence, $P_0(k_x)$ is the kernel function for H_0 and we have

$$K_0(x, y) = \langle P_0(k_y), P_0(k_x) \rangle = \langle P_0(k_y), k_x \rangle.$$

□

3 Frames in reproducing kernel Hilbert spaces

We denote the closure in H of the linear span of $\{k(x_n, .)\}$ by H_∞, and require $\{k(x_n, .)\}$ to be a frame in H_∞, i.e., that there be constants A, B such that

$$A\|f\|^2 \leq \sum_{n=1}^{\infty} \Big| \int_{-\infty}^{\infty} k(x, x_n) f(x) dx \Big|^2 \leq B\|f\|^2, \tag{3}$$

for all $f \in H_\infty$ (see [6]). This may be expressed as

$$A\|f\|^2 \leq \sum_{n=1}^{\infty} |\langle f, \phi_n \rangle_{H_\infty}|^2 \leq B\|f\|^2, \tag{4}$$

for $f(x) = k(x_i, x)$ and $\phi_n := k(x_n, .)$ which gives us a necessary condition that (3) hold. The frame is said to be tight if A and B are equal and a frame with $A = B = 1$, is called Parseval frame. We again form the matrix $K = [k(x_i, x_j)]$ each of whose rows belongs to ℓ^2 by (4). Again K is formally self-adjoint and maps ℓ^2 into a space of sequences dominated by $k(x_n, x_n)$.
We suppose further that H_∞ contains series with coefficients in ℓ^2, i.e.,

$$\{a_n\} \in \ell^2 \quad \textit{implies} \quad \| \sum_n a_n k(x_n, .) \| < \infty. \tag{5}$$

This ensures that the matrix K maps ℓ^2 into ℓ^2.

Theorem 3.1 ([6])**.** *Let K be the operator on ℓ^2 given by the infinite matrix $[k(x_i, x_j)]$ and let $k(x_n, x)$ satisfy (5). Then K is a self-adjoint from ℓ^2 to ℓ^2 with a positive spectrum contained in the interval $[A, B]$, where A and B are as in (3).*

Proof. By (5) the series $\sum_m^p a_n k(x_n, x)$ converges to an element f of H_∞ for $\{a_n\} \in \ell^2$. Then

$$\overline{f(x_n)} = \int_{-\infty}^{\infty} \overline{k(x_n, u)} f(u) du = \sum_i a_n k(x_i, x_n).$$

By (3), $\overline{f(x_n)} \in \ell^2$ and hence the range of K is contained in ℓ^2. By (3) again if $f(x_n) = 0$ for all n, then $\|f\| = 0$, which implies that K is one-to-one. If $\{\alpha_n\}$ is an eigenvector of K with eigenvalue λ and $\varphi(x) = \sum \alpha_n k(x_n, x)$ with $\sum |\alpha_n|^2 = 1$, then (3) becomes

$$A\|\varphi\|^2 = A\sum_n \alpha_n \overline{\sum_m \alpha_m k(x_n, x_m)} = A\sum_n \alpha_n \lambda \overline{\alpha_n} = A\lambda$$
$$\leq \sum_n \Big|\sum_m \alpha_m k(x_n, x_m)\Big|^2 = \sum \lambda^2 |\alpha_n|^2 = \lambda^2 \leq B\lambda.$$

The same result holds if λ is an approximate eigenvalue. □

If the set ϕ_n satisfies the frame condition then the frame operator U can be defined as

$$\begin{aligned} U :& H_\infty \longrightarrow \ell^2 \\ & f \longrightarrow \{\langle f, \phi_n\rangle_{H_\infty}\}. \end{aligned} \tag{6}$$

The reconstruction of f from its frame coefficients needs the definition of a dual frame. For this purpose, one introduces the adjoint operator U^* of U which exists and is unique because it lies on a Hilbert space:

$$\begin{aligned} U^* :& \ell^2 \longrightarrow H_\infty \\ & \{c_n\} \longrightarrow \sum c_n \phi_n. \end{aligned} \tag{7}$$

A key role in frame theory is played by the so called frame operator S which is defined by

$$S : H_\infty \longrightarrow H_\infty, \quad Sf := U^*Uf = \sum_n \langle f, \phi_i\rangle \phi_i.$$

It is known that S is always bounded, invertible, self-adjoint, and positive.

Theorem 3.2. *Let $\{\phi_n\}$ be a frame of H_∞ with frame bounds A and B. Let us define the dual frame $\{\overline{\phi_n}\}$ as $\overline{\phi_n} = S^{-1}\phi_n$. For all $f \in H_\infty$, we have*

$$\frac{1}{B}\|f\|^2 \leq \sum_{n=1}^{\infty} |\langle f, \overline{\phi_n}\rangle_{H_\infty}|^2 \leq \frac{1}{A}\|f\|^2, \tag{8}$$

and

$$f = \sum_{n=1}^{\infty} \langle f, \overline{\phi_n}\rangle_{H_\infty} \phi_n = \sum_{n=1}^{\infty} \langle f, \phi_n\rangle_{H_\infty} \overline{\phi_n}. \tag{9}$$

If the frame is tight then $\overline{\phi_n} = \frac{1}{A}\phi_n$.

Proof. A proof is described in [2]. □

4 Parseval Frames and reproducing kernel Hilbert spaces

In this section, we show the connection between reproducing kernels and Parseval frames. The following result shows one of the most common way that Parseval frame arises.

Lemma 4.1. *Let H be a Hilbert space, let $H_0 \subseteq H$ be a closed subspace and let $P_0 : H \longrightarrow H_0$ be the orthogonal projection onto H_0. If $\{u_n\}_{n=1}^{\infty}$ is an orthonormal basis for H then $\{P_0(u_n)\}_{n=1}^{\infty}$ is a Parseval frame for H_0.*

Proof. A proof is described in [2]. □

The numbers $\langle f, \overline{\phi_n}\rangle_{H_\infty}$ in equation (9) are called frame coefficients. Note that the operator U^* occurring in equation (7) is not required to be injective. An element $f \in H_\infty$ might therefore, have different expansions, i.e. $c, d \in \ell^2,\ c \neq d$, but $f = \sum_i c_i\phi_i = \sum_i d_i\phi_i$. The next theorem show that every f has one canonical expansion and this canonical expansion is given by the frame coefficients $\langle f, \overline{\phi_n}\rangle_{H_\infty}$.

Theorem 4.2. *Let $\{\phi_n\}$ be a frame for H_∞ and let f be an arbitrary element of H_∞ with an arbitrary representation $f = \sum_{n=1}^{\infty} c_i\phi_i$, then*

$$\sum_{n=1}^{\infty} c_i^2 = \sum_{n=1}^{\infty} \langle f, S^{-1}\phi_i\rangle^2 + \sum_{n=1}^{\infty} (c_i - \langle f, S^{-1}\phi_i\rangle)^2. \tag{10}$$

Proof. A proof is described in [2]. □

The theorem above leads to a characterization of the frame coefficients with referring to the frame operator S.

Theorem 4.3. *Let H be a Hilbert space of functions with reproducing kernel K. Let $\{\phi_n\}$ be a set of functions belonging to H. Then, the reproducing kernel $K(s,t)$ has the representation*

$$K(s,t) = \sum_n \overline{\phi_n}(s)\phi_n(t), \tag{11}$$

if and only if $\{\phi_n\}$ is a Parseval frame of H.

Proof. We assume that $K(s,t) = \sum_n \overline{\phi_n}(s)\phi_n(t)$ is the reproducing kernel of H. We obtain

$$f(x) = \langle f, K(x,.)\rangle = \sum_n \langle f, \phi_i\rangle\phi_i(x),$$

and therefore, $\|f\|^2 = \sum_n \langle f, \phi_i\rangle^2$ for every $f \in H$. By definition the set $\{\phi_n\}$ is a Parseval frame in H.
Conversely, we assume that $\{\phi_n\}$ is a Parseval frame in H. Due to the reproducing property of the kernel K for each $t \in X$ we have

$$\sum \phi_n^2(t) = \sum |\langle K(t,.), \phi_n(.)\rangle_{H_\infty}|^2 \leq B\|K(t,.)\|^2 < \infty.$$

and therefore the kernel $K(s,t) = \sum_n \overline{\phi_n}(s)\phi_n(t)$ is well defined. Since the operator U^* is an isometry between H and ℓ^2, so $\langle f, g\rangle_H = \langle U^*(f), U^*(g)\rangle_{\ell^2}$ for all $f, g \in H$. This means

$$\langle f, g\rangle_H = \sum_n \langle f, \phi_i\rangle_H \langle g, \phi_i\rangle_H,$$

for every $f, g \in H$. Particularly, for $g := K(t,.),\ \ t \in X$ we obtain

$$f(t) = \langle f, K(t,.)\rangle = \sum_{n=1}^{\infty} \langle f, \overline{\phi_n}\rangle_H \phi_n(t) = \langle f(.), \sum_{n=1}^{\infty} \overline{\phi_n}(.)\phi_n(t)\rangle_H.$$

Due to the uniqueness of the reproducing kernel we have $K(s,t) = \sum_{n=1}^{\infty} \overline{\phi_n}(s)\phi_n(t)$. □

Theorem 4.4. *Let H be a reproducing kernel Hilbert space and $K(x,y) = \sum_{i=1}^{\infty} \overline{\phi_i}(x)\phi_i(y)$ be a Parseval frame expansion of the reproducing kernel K. We consider a function f of the form $f = \sum_{j=1}^{n} \alpha_j K(x_j, .) \in H$, where $\{x_1, x_2, \cdots, x_n\}$ is a set of points belonging to X and α_i are some real numbers. Then, the frame coefficients of f are given by*

$$\langle f, S^{-1}\phi_i \rangle_H = \sum_{j=1}^{n} \alpha_j \phi_i(x_j).$$

Proof. We define the coefficients $c_i = \sum_{j=1}^{n} \alpha_j \phi_i(x_j)$. Then f has the expansion

$$f = \sum_{j=1}^{n} \alpha_j K(x_j, .) = \sum_{j=1}^{n} \alpha_j \Big(\sum_{i=1}^{\infty} \overline{\phi_i}(x_j)\phi_i(.) \Big) = \sum_{i=1}^{\infty} c_i \phi_i.$$

By $\mathcal{N}_U \subset \ell^2$ we denote the nullspace of U and by $\mathcal{N}_U^{\perp}$ we denote its orthogonal complement where the operator U defined in equation (6). Let d be an arbitrary element of $\mathcal{N}_U$, i.e. $\sum_{i=1}^{\infty} d_i \phi_i(x) = 0$ for all $x \in X$. we obtain

$$\langle c, d \rangle_{\ell^2} = \sum_{i=1}^{\infty} c_i d_i = \sum_{i=1}^{\infty} d_i \Big(\sum_{j=1}^{n} \alpha_j \phi_i(x_j) \Big) = \sum_{j=1}^{n} \alpha_j \Big(\sum_{i=1}^{\infty} d_i \phi_i(x_j) \Big) = 0,$$

and therefore, $c \in \mathcal{N}_U^{\perp}$. Let now $d \in \ell^2$ define arbitrary coefficients such that $f = \sum_{i=1}^{\infty} d_i \phi_i$. Since $c - d \in \mathcal{N}_U$ and $c \in \mathcal{N}_U^{\perp}$ Pythagoras theorem yields

$$\sum_{i=1}^{\infty} d_i^2 = \sum_{i=1}^{\infty} (c_i - d_i)^2 + \sum_{i=1}^{\infty} c_i^2,$$

and therefore, $\sum_{i=1}^{\infty} c_i^2 \leq \sum_{i=1}^{\infty} d_i^2$, for all $d \in \ell^2$ with $f = \sum_{i=1}^{\infty} d_i \phi_i$. Theorem 4.2 states that $\langle f, S^{-1}\phi_i \rangle_H = c_i$. □

References

[1] A. Berlinet and T. A. Christine, *Reproducing kernel Hilbert spaces in probability and statistics*, Springer Science and Business Media, 2011.

[2] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, 2016.

[3] M. R. Foroutan, A. Ebadian and H. R. Fazli, Generalized Jacobi reproducing kernel method in Hilbert spaces for solving the Black-Scholes option pricing problem arising in financial modelling, *Mathematical Modelling and Analysis,* 23(4) (2018), 538–553.

[4] M. R. Foroutan, A. S. Gholizadeh, Sh. Najafzadeh, R. H. Haghi, Laguerre reproducing kernel method in Hilbert spaces for unsteady stagnation point flow over a stretching/shrinking sheet, *Appl. Math. J. Chinese Univ.,* 36(3) (2021), 354–-369.

[5] M. Levy, *Introduction to reproducing kernel Hilbert spaces and the continuous wavelet transform*, The Open University. 2007.

[6] M. Z. Nashed, G. G. Walter, General sampling theorems for functions in reproducing kernel Hilbert spaces, *Mathematics of Control, Signals and Systems,* 4(4) (1991), P. 363.

[7] A. Rakotomamonjy, S. Canu, A. Smola, Frames, Reproducing Kernels, Regularization and Learning, *Journal of Machine Learning Research,* 6(9) 2005.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some classes of Mengerian simplicial complexes

Rahim Rahmati-Asghar*

Department of Mathematics, Faculty of Basic Sciences, University of Maragheh, P. O. Box 55181-83111, Maragheh, Iran.

Abstract

In this paper, we introduce two classes of hypergraphs whose flag complexes are Mengerian.

Keywords: Mengerian simplicial complex, Mengerian Matrix, hypergraph.
Mathematics Subject Classification [2010]: 13B25,13F20,05E40

1 Introduction

A *simplicial complex* Δ on the vertex set $V(\Delta) = \{v_1, \ldots, v_n\}$ is a colection of subsets of $V(\Delta)$ that satisfy the following conditions:

1. $\{v_i\} \in \Delta$, for all $v_i \in V(\Delta)$;
2. If $F \in \Delta$ and G is a subset of F, then $G \in \Delta$.

For convenience we will denote the vertex set of Δ by $[n] = \{1, 2, \ldots, n\}$.

The elements of the simplicial complex are called *faces*. The *dimension* of a face, F, is denoted by $\dim(F)$ and $\dim(F) = |F| - 1$. Denote by $d = \max\{|F| : F \in \Delta\}$. Then the dimension of the simplicial complex Δ is $d - 1$. Throughout this paper Δ is a simplicial complex of dimension $d - 1$ on $[n]$.

A *facet* of Δ is a maximal face (with respect to the inclusion). If all the facets have the same dimension, we say that the simplicial complex is *pure*, otherwise, it is called *nonpure*.

Notation. In this paper we use the term "simplicial complex" to mean a simplicial complex which is not necessarily pure, unless otherwise indicated. Denote by $\mathcal{F}(\Delta)$ the set of all the facets of Δ. It is clear that $\mathcal{F}(\Delta)$ determines Δ. When $\mathcal{F}(\Delta) = \{F_1, \ldots, F_r\}$, we write $\Delta = \langle F_1, \ldots, F_r\rangle$. More generally, if we have a set $\{G_1, \ldots, G_s\}$ of faces of Δ, we denote by $\langle G_1, \ldots, G_s\rangle$ the subcomplex of Δ consisting of those faces of Δ which are contained in some G_i.

A *nonface* of Δ is a subset F of $[n]$ with $F \notin \Delta$. Let $\mathcal{N}(\Delta)$ denote the set of minimal nonfaces of Δ. Recall that a simplicial complex is called *flag*, if all minimal nonfaces consist of two elements [2].

Let $\mathcal{V} = \{x_1, \ldots, x_n\}$ be a finite set and $\mathcal{E} = \{e_1, \ldots, e_m\}$ a finite collection of distinct subsets of $\mathcal{V}$. The pair $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is called a *hypergraph* if $e_i \neq \emptyset$ for each i. The elements

*Speaker. Email address: rahmatiasghar.r@gmail.com

of $\mathcal{V}$ and $\mathcal{E}$ are called the *vertices* and *edges*, respectively, of $\mathcal{H}$. We may write $\mathcal{V}(\mathcal{H})$ and $\mathcal{E}(\mathcal{H})$ for the vertices and edges of $\mathcal{H}$, respectively. The hypergraph $\mathcal{H}$ is *simple* if: (1) $|e| \geq 2$ for all $e \in \mathcal{E}$ and (2) whenever $e_i, e_j \in \mathcal{E}$ and $e_i \subseteq e_j$, then $i = j$. In this paper we consider only simple hypergraphs, and hence, by hypergraph we will always mean a simple hypergraph. In the literature, a simple hypergraph is also called a clutter. A simple hypergraph is as simple graph if every $e \in \mathcal{E}$ has cardinality 2. If $V \subseteq \mathcal{V}$, the *induced subhypergraph* on V, $\mathcal{H}_V$, is a hypergraph with $\mathcal{V}(\mathcal{H}_V) = V$ and with $\mathcal{E}(\mathcal{H}_V)$ consisting of the edges of $\mathcal{E}(\mathcal{H})$ that lie entirely in V. The hypergraph $\mathcal{H}$ is called *d-uniform* if $|e| = d$ for each $e \in \mathcal{E}(\mathcal{H})$. The vertex $x \in \mathcal{V}(\mathcal{H})$ is called isolated if it belongs to no edge of $\mathcal{H}$. A hypergraph with vertex set $[n] := \{x_1, \ldots, x_n\}$ is *complete* if its edge set is the set of all subsets of $[n]$ and it is denoted by $\mathcal{K}_n$. We will also denote by $\mathcal{K}_n^d$ the complete d-uniform hypergraph. A *clique* is a complete induced subhypergraph of a hypergraph.

The *complementary hypergraph* $\mathcal{H}^c$, of a d-uniform hypergraph $\mathcal{H}$, is defined a hypergraph on the same set of vertices as $\mathcal{H}$, and edge set

$$\mathcal{E}(\mathcal{H}^c) = \{e \subseteq \mathcal{V}(\mathcal{H}) : |e| = d, e \notin \mathcal{E}(\mathcal{H})\}.$$

We denote by $\bar{\mathcal{H}}$ the hypergraph on the vertex set $\mathcal{V}(\mathcal{H})$ and the edge set $\mathcal{E}(\bar{\mathcal{H}}) = \{e^c \subseteq \mathcal{V}(\mathcal{H}) : e \in \mathcal{E}(\mathcal{H})\}$, where $e^c = \mathcal{V}(\mathcal{H}) \backslash e$.

The *flag* complex of the hypergraph $\mathcal{H}$ denoted by $\Delta(\mathcal{H})$ is defined as follows:

$$\Delta(\mathcal{H}) = \{F \subset [n] : \ every\ d-element\ subset\ of\ F\ is\ in\ \mathcal{E}(\mathcal{H})\}.$$

Definition 1.1. [1] A *chordal* hypergraph is a d-uniform hypergraph, obtained inductively as follows:

- $\mathcal{K}_n^d$ is a chordal hypergraph, $n, d \in \mathcal{N}$.
- If $\mathcal{G}$ is chordal, then so is $\mathcal{C} = \mathcal{G} \cup_{\mathcal{K}_j^d} \mathcal{K}_i^d$ for $0 \leq j < i$. ($\mathcal{K}_i^d$ is attached to $\mathcal{G}$ in a common (under identification) $\mathcal{K}_j^d$.)

Definition 1.2. [3] A d-partite d-uniform hypergraph $\mathcal{F}$ on the vertex set $V^{(1)} \dot{\cup} \ldots \dot{\cup} V^{(d)}$ and with the edge set $\mathcal{E}(\mathcal{F}) = \{\{x_{i_1}, \ldots, x_{i_d}\} : x_{i_j} \in V^{(j)} \text{ for all } j\}$ is a *Ferrers hypergraph* if for $\{x_{i_1}, \ldots, x_{i_d}\} \in \mathcal{E}(\mathcal{F})$ and $\{x_{i'_1}, \ldots, x_{i'_d}\}$ with $i'_j \leq i_j$ for all j, one also has $\{x_{i'_1}, \ldots, x_{i'_d}\} \in \mathcal{E}(\mathcal{F})$.

In [4] we studied some algebraic properties of chordal and Ferrers hypergraphs. Actually, we described some combinatorial and homological structures of edge ideals of these hypergraphs.

2 Main results

Let $F(\Delta) = \{F_1, \ldots, F_m\}$, and let M be the incidence matrix of Δ, that is, the $m \times n$-matrix $M = (e_{ij})$ with $e_{ij} = 1$ if $j \in F_i$ and $e_{ij} = 0$ if $j \notin F_i$.

Recall from [2] that a simplicial complex Δ on $[n]$ with incidence matrix M is called a *Mengerian* simplicial complex if for all $\mathbf{a} \in \mathbb{Z}_+^n$,

$$\min\{\langle \mathrm{c}, \mathbf{a}\rangle : \mathrm{c} \in \mathbb{Z}_+^n, M \cdot \mathrm{c} \geq \mathbf{1}\} = \max\{\langle \mathbf{b}, \mathbf{1}\rangle : \mathbf{b} \in \mathbb{Z}_+^m, M^t \cdot \mathbf{b} = \mathrm{c}\}.$$

Theorem 2.1. *Let Δ be a flag complex of a connected chordal hypergraph. Then Δ is Mengerian.*

Theorem 2.2. *Let Δ be a flag complex of a the complement of a connected Ferrers hypergraph. Then Δ is Mengerian.*

References

[1] E. Emtander, *A class of hypergraphs that generalizes chordal graphs*, Math. Scand. **106** (2010), 50-66.

[2] J. Herzog, T. Hibi, *Monomial ideals*, Graduate Texts in Mathematics *260*, Springer-Verlag. (2011).

[3] U, Nagel, V. Reiner, *Betti numbers of monomial ideals and shifted skew shapes*, the electronic journal of combinatorics **16** (2009), no. 2, Special volume in honor of Anders Bjorner, Research Paper 3, 59 pp.

[4] R. Rahmati-Asghar, S. Yassemi, *On the weakly polymatroidal property of the edge ideals of hypergraphs*, Communications in Algebra, *42*: 1011-1021, 2014.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Study on Some Integral Inequalities for Pseudo-Integrals

Bayaz Daraby*

Department of Mathematics, University of Maragheh, P. O. Box 55181-83111, Maragheh, Iran.

Abstract

In this paper, we express and prove Stolars ky, Feng Qi type inequalities for two classes of pseudo-integrals. One of them concerning the pseudo-integrals based on a function reduces on the g-integral where pseudo-operations are defined by a monotone and continuous function g. The other one concerns the pseudo-integrals based on a semiring $([a,b],\max,\odot)$, where $\odot$ is generated. The integral inequalities are appling in multivariate approximation theory and probability theory and etc.

Keywords: Feng Qi type inequality, Stolarsky type inequality , Sugeno integral, pseudo-integral, Fuzzy integral inequality

Mathematics Subject Classification [2010]: 03E72, 28E10, 26E50

1 Introduction

Pseudo-analysis is a generalization of the classical analysis, where instead of the field of real numbers, a semiring is taken on a real interval $[a,b]\subset[-\infty,\infty]$ endowed with pseudo-addition $\bigoplus$ and with pseudo-multiplication $\odot$ (see [4]). Pseudo-analysis would be an interesting topic to generalize an inequality from the framework of the classical analysis to that of some integrals which contain the classical analysis as special cases [1, 9].

The integral inequalities are good mathematical tools both in theory and application.

In this paper, firstly we generalize the stolarsky type inequality and some version of Feng Qi type inequalities for pseudo-integrals of monotone functions.

2 Preliminaries

In this section, it is going to be reviewed some well-known results of pseudo-operations, pseudo-analyses and pseudo-integrals in details, we refer to [3, 4]

Let $[a,b]$ be a closed (in some cases can be considered semiclosed) subinterval of $[-\infty,\infty]$. The full order on $[a,b]$ will be denoted by $\preceq$.

Definition 2.1. The operation $\oplus$ (pseudo-addition) is a function $\oplus:[a,b]\times[a,b]\to[a,b]$ which is commutative, nondecreasing (with respect to $\preceq$), associative and with a zero (natural) element denoted by $\mathbf{0}$, i.e., for each $x\in[a,b]$, $\mathbf{0}\oplus x=x$ holds (usually $\mathbf{0}$ is either a or b).

Let $[a,b]_+=\{x|x\in[a,b],\mathbf{0}\preceq x\}$.

*Speaker. Email address: bdaraby@maragheh.ac.ir

Definition 2.2. The operation $\odot$ (pseudo-multiplication) is a function $\odot : [a,b] \times [a,b] \to [a,b]$ which is commutative, positively non-decreasing, i.e., $x \preceq y$ implies $x \odot z \preceq y \odot z$ for all $z \in [a,b]_+$, associative and for which there exists a unit element $\mathbf{1} \in [a,b]$, i.e., for each $x \in [a,b], \mathbf{1} \odot x = x$.

We also assume $\mathbf{0} \odot x = \mathbf{0}$ that $\odot$ is a distributive pseudo-multiplication with respect to $\oplus$, i.e., $x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z)$. The structure $([a,b], \oplus, \odot)$ is a semiring (see [3,8]). We will consider the semiring $([a,b], \oplus, \odot)$ for two important (with completely different behavior) cases.

The first case is when pseudo-operations are generated by a monotone and continuous function $g : [a,b] \to [0,\infty]$, i.e., pseudo-operations are given with:

$x \oplus y = g^{-1}(g(x) + g(y))$ and $x \odot y = g^{-1}(g(x)g(y))$.

Then, the pseudo-integral for a function $f : [c,d] \to [a,b]$ reduces on the g-integral [6,7],

$$\int_{[c,d]}^{\oplus} f(x)dx = g^{-1}\left(\int_c^d g(f(x))dx\right). \tag{1}$$

More on this structure as well as on corresponding measures and integrals can be found in [5,6]. The second class is when $x \oplus y = max(x,y)$ and $x \odot y = g^{-1}(g(x)g(y))$, the pseudo-integral for a function $f : \mathbb{R} \to [a,b]$ is given by

$$\int_{\mathbb{R}}^{\oplus} f \odot dm = \sup\left(f(x) \odot \psi(x)\right), \tag{2}$$

where function ψ defines sup-measure m. Any sup-measure generated as essential supremum of a continuous density can be obtained as a limit of pseudo-additive measures with respect to generated pseudo-addition [4]. For any continuous function $f : [0,\infty] \to [0,\infty]$ the integral $\int^{\oplus} f \odot dm$ can be obtained as a limit of g-integrals, [4].

We denote by μ the usual Lebesgue measure on $\mathbb{R}$. We have

$$m(A) = esssup(x|x \in A) = \sup\{a|\mu(\{x|x \in A, x > a\}) > 0\}.$$

Theorem 2.3 ([4]). *Let m be a sup-measure on $([0,\infty], \mathbb{B}([0,\infty]))$, where $\mathbb{B}([0,\infty])$ is the Borel σ-algebra on $[0,\infty]$, $m(A) = esssup_\mu(\psi(x)|x \in A)$, and $\psi : [0,\infty] \to [0,\infty]$ is a continuous density. Then, for any pseudo-addition $\oplus$ with a generator g there exists a family $\{m_\lambda\}$ of $\oplus_\lambda$–measure on $([0,\infty), \mathbb{B})$, where $\oplus_\lambda$ is generated by g^λ (the function g of the power λ), $\lambda \in (0,\infty)$, such that $lim_{\lambda\to\infty} m_\lambda = m$.*

Theorem 2.4 ([4]). *Let $([0,\infty], sup, \odot)$ be a semiring, when $\odot$ is generated with g, i.e.,we have $x \odot y = g^{-1}(g(x)g(y))$ for every $x, y \in (0,\infty)$. Let m be the same as in Theorem 2.8. Then, there exists a family $\{m_\lambda\}$ of $\oplus_\lambda$-measures, where $\oplus_\lambda$ is generated by $g^\lambda, \lambda \in (0,\infty)$ such that for every continuous function $f : [0,\infty] \to [0,\infty]$,*

$$\begin{aligned} \int^{sup} f \odot dm &= \lim_{\lambda\to\infty} \int^{\oplus_\lambda} f \odot dm_\lambda \\ &= \lim_{\lambda\to\infty} \left(g^\lambda\right)^{-1} \left(\int g^\lambda(f(x))dx\right). \end{aligned} \tag{3}$$

Now easily we can obtain indicate the properties listed in the following proposition.

Proposition 2.5 ([2]). *Let $(X, \mathcal{F}, \mu, \mathbb{R}_+^-, \oplus, \odot)$ is a pseudo-space and $f, g \in \mathcal{F}$, then:*

(1) If $f = 0$ on A a.e., then $\int_A^{\oplus} f d\mu = 0$.

(2) If $\mu(A)=0$, *then* $\int_A^{\oplus} f d\mu = 0$.

(3) $\int_A^{\oplus} a d\mu \geq a \odot \mu(A)$.

(4) If $f \leq g$ *on* A, *then* $\int_A^{\oplus} f d\mu \leq \int_A^{\oplus} g d\mu$.

(5) If $A \subset B$, *then* $\int_A^{\oplus} f d\mu \leq \int_B^{\oplus} f d\mu$.

3 Stolarsky Type Inequality for Pseudo-Integrals

Theorem 3.1 (Pseudo Stolarsky type inequality: decreasing case)**.** *Let* $a, b > 0$, $f : [0,1] \to [0,1]$ *be a continuous and strictly decreasing function and* μ *be the Lebesgue measure on* $\mathbb{R}$. *If the pseudo-operations are defined by a continuous and increasing function* $g : [0,1] \to [0,1]$, *then the inequality*

$$\int_{[0,1]}^{\oplus} f\left(x^{\frac{1}{a+b}}\right) dx \geq \left(\int_{[0,1]}^{\oplus} f\left(x^{\frac{1}{a}}\right) dx\right) \odot \left(\int_{[0,1]}^{\oplus} f\left(x^{\frac{1}{b}}\right) dx\right) \tag{4}$$

holds.

Example 3.2. 1. Let $g(x) = x^{\alpha}$ for some $\alpha \in [1, \infty)$, so $x \oplus y = \sqrt[\alpha]{x^{\alpha} + y^{\alpha}}$ and $x \odot y = xy$. Then (5) reduces on the following inequality

$$\sqrt[\alpha]{\int_0^1 f\left(x^{\frac{1}{a+b}}\right)^{\alpha} dx} \geq \left(\sqrt[\alpha]{\int_0^1 f\left(x^{\frac{1}{a}}\right)^{\alpha} dx}\right)\left(\sqrt[\alpha]{\int_0^1 f\left(x^{\frac{1}{b}}\right)^{\alpha} dx}\right).$$

2. Let $g(x) = e^x$. The corresponding pseudo-operations are $x \oplus y = \ln(e^x + e^y)$ and $x \odot y = x + y$. Then (4) reduces on the following inequality

$$\ln \int_0^1 e^{f\left(x^{\frac{1}{a+b}}\right)} dx \geq \ln \int_0^1 e^{f\left(x^{\frac{1}{a}}\right)} dx + \ln \int_0^1 e^{f\left(x^{\frac{1}{b}}\right)} dx.$$

Theorem 3.3. *Let* $a, b > 0$, *if* $f : [0,1] \to [0,1]$ *is a continuous and strictly increasing function and let* m *be the same as in the Theorem 2.3. If* $\odot$ *is represented by a decreasing multiplicative generator* g, *then the inequality*

$$\int_{[0,1]}^{\sup} f\left(x^{\frac{1}{a+b}}\right) dm \geq \left(\int_{[0,1]}^{\sup} f\left(x^{\frac{1}{a}}\right) dm\right)\left(\int_{[0,1]}^{\sup} f\left(x^{\frac{1}{b}}\right) dm\right) \tag{5}$$

holds.

4 Feng Qi Type Inequalities for Pseudo-Integrals

Theorem 4.1. *For a given measurable space* (X, A), *let* $f : [0,1] \to [0,1]$ *be a real valued function such that* $(S)\int_0^1 f d\mu = p$. *If* f *is a continuous and strictly decreasing function, such that* $f\left(p^{n+1}\right) \geq p^{\left(\frac{n+1}{n+2}\right)}$ *and let a generator* $g : [0,1] \to [0,\infty)$ *of pseudo-addition* $\oplus$ *and psudo-multipication* $\odot$ *be decreasing function. Then the inequality:*

$$\int_{[0,1]}^{\oplus} f_{\odot}^{n+2} \odot dm \geq \left(\int_{[0,1]}^{\oplus} f_{\odot} \odot dm\right)_{\odot}^{n+1}$$

holds for all $n \geq 0$ *and* $\sigma - \oplus$*-measure* m.

Proof. We apply the classical Feng Qi inequality and we obtain:

$$\int_0^1 (g \circ f)^{n+2} d(g \circ m) \geq \left(\int_0^1 (g \circ f) d(g \circ m) \right)^{n+1}.$$

since function g is decreasing function, then g^{-1} is also decreasing function and we obtain:

$$g^{-1}\left(\int_0^1 (g \circ f)^{n+2} d(g \circ m) \right) \geq g^{-1}\left(\int_0^1 (g \circ f) d(g \circ m) \right)^{n+1}$$

for left side of inequality we have:

$$\begin{aligned} g^{-1}\left(\int_0^1 (g \circ f)^{n+2} d(g \circ m) \right) &= g^{-1}\left(\int_0^1 g\left(g^{-1}(g \circ f)^{n+2} \right) d(g \circ m) \right) \\ &= \int_{[0,1]}^{\oplus} f_{\odot}^{n+2} \odot dm. \end{aligned}$$

For right side of the inequality we have:

$$\begin{aligned} g^{-1}\left(\int_0^1 (g \circ f) d(g \circ m) \right)^{n+1} &= g^{-1}\left(\int_0^1 g\left(g^{-1}(g \circ f) \right) d(g \circ m) \right)^{n+1} \\ &= \left(\int_{[0,1]}^{\oplus} f_{\odot} \odot dm \right)^{n+1}. \end{aligned}$$

Hence we have:

$$\int_{[0,1]}^{\oplus} f_{\odot}^{n+2} \odot dm \geq \left(\int_{[0,1]}^{\oplus} f_{\odot} \odot dm \right)_{\odot}^{n+1}$$

In this generalation we have term $f\left(p^{n+1}\right) \geq p^{\left(\frac{n+1}{n+2}\right)}$, because this theorem is ture in fuzzy case. □

Theorem 4.2. *For a given measurable space (X, A), let $f : [0,1] \to [0,1]$ be a ral valued function such that $(S)\int_0^1 f d\mu = p$. If f is a continuous and strictly increasing function, such that $f\left(1 - p^{n+1}\right) \geq p^{\left(\frac{n+1}{n+2}\right)}$ and let a generator $g : [0,1] \to [0,\infty)$ of pseudo-addition $\oplus$ and psudo-multipication $\odot$ be an increasing function. Then the inequality:*

$$\int_{[0,1]}^{\oplus} f_{\odot}^{n+2} \odot dm \geq \left(\int_{[0,1]}^{\oplus} f_{\odot} \odot dm \right)_{\odot}^{n+1}$$

holds for all $n \geq 0$ and $\sigma - \oplus$-measure m.

Proof. The proof is similar whit Theorem 4.1. □

Example 4.3. Let $g(x) = \ln(x)$, then

$$x \oplus y = xy, \qquad x \odot y = e^{\ln x . \ln y}.$$

by Theorem 4.1, the following inequality

$$\ln \int_0^1 e^{(\ln f(x))^{n+2}} \geq \left(\ln \int_0^1 e^{(\ln f(x))} \right)^{n+1}$$

holds.

In the sequel, we generalize the Feng Qi inequality by the semiring $([a,b],\max,\odot)$, where $\odot$ is generated .

Theorem 4.4. *Let $f:[0,1]\to[0,1]$ be a real valued, continuous and strictly increasing function such that $(S)\int_0^1 f d\mu = p$. If $\odot$ is represented by a increasing generator g and m is complet* sup*-measure same as in Theorem 2.4, then whit condition $f\left(1-p^{n+1}\right)\geq p^{\left(\frac{n+1}{n+2}\right)}$*

$$\int_{[0,1]}^{\sup} f_{\odot}^{n+2}\odot dm \geq \left(\int_{[0,1]}^{\sup} f\odot dm\right)_{\odot}^{n+1}$$

holds for all $n\geq 0$ and $\sigma-\oplus$-measure m.

5 Conclusion

We have proved the Stolarsky and Feng Qi type inequalities for pseudo-integrals when we consider the semiring $([0,1],\oplus,\odot)$. We concentrate for two classes of pseudo-integrals: The first class includes the pseudo-integral based on a function reduces on the g-integral, where $\oplus$ and $\odot$ are defined by a monotone and continuous function g. The second class includes the pseudo-integral based on the semiring $([0,1],max,\odot)$ is given by *sup*-measure where $x\odot y$ is generated by $g^{-1}\left(g(x)g(y)\right)$.

References

[1] H. Agahi, R. Mesiar and Y. Ouyang, *Chebyshev type inequalities for pseudo-integrals*, Nonlinear Analysis, 72 (2010), pp. 2737-2743.

[2] B. Daraby, *Generalizations of the Well-Known Chebyshev type inequalities for pseudo-integrals*, Gen. Math. Notes, 38 (1) (2017), pp. 32-45.

[3] W. Kuich, *Semirings, Automata, Languages*, Springer-Verlag, Berlin, 1986.

[4] R. Mesiar and E. Pap, *Idempotent integral as limit of g-integrals*, Fuzzy Sets Syst., 102 (1999), pp. 385-392.

[5] E. Pap, *An integral generated by decomposable measure*, Univ. Novom Sadu Zb. Rad. Prirod. -Mat. Fak. Ser. Mat., 20 (1) (1990), pp. 135-144.

[6] E. Pap, *g-calculus*, Univ. Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat., 23 (1) (1993), pp. 145-156.

[7] E. Pap, *N. Ralević, Pseudo-Laplace transform*, Nonlinear Analysis, 33 (1998), pp. 553-560.

[8] E. Pap, *Pseudo-additive measures and their applications*, in: E. Pap (Ed.), Handbook of Measure Theory, Elsevier, Amsterdam, 2002, pp. 1403-1465.

[9] E. Pap and M. Štrboja, *Generalization of the Jensen inequality for pseudo-integral*, Inf. Sci., 180 (2010), pp. 543-548.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some facts about Quasi-Block Toeplitz matrices

M. Shams Solary*

Department of Mathematics, Payame Noor University, Tehran, IRAN

Abstract

In this paper, we try to find some analytical and approximate solutions for quasi-block Toeplitz (QBT) matrices with some MATLAB commands. We say that, these matrices are semi-infinite block matrices of the kind $\mathbf{F} = T(F) + E$ where $T(F) = (F_{j-k})_{j,k\in\mathbb{Z}}$, that F_k are $m \times m$ matrices such that $\sum_{i\in\mathbb{Z}} |F_i|$ has bounded entries, and $E = (e_{i,j})_{i,j\in\mathbb{Z}^+}$ is a compact correction.
Here, we have the norms $\| \mathbf{F} \|_w = \sum_{i\in\mathbb{Z}} \| F_i \|$ and $\| E \|_2$ are finite.

Keywords: Quasi-Block, Toeplitz matrix, Banach algebra
Mathematics Subject Classification [2010]: 65F30, 60B20

1 Introduction

In [1, 2], Some properties and descriptions are provided for the r-Toeplitz matrices and finitely representable QT matrices. The interest of the study of these matrices appears to be very important not only from a theoretical point of view in linear algebra or numerical analysis, e.g., but also in applications such as ranging from imaging to Markov chains, queuing models, sound propagation problem, finance to the solution of PDEs, Yule-Walker equations, etc [3, 5].
So, this motivated us to study and to collect more information about this matter in this way.

2 Main results

The semi-infinite quasi-Toeplitz (QT) matrices which comes from a Banach algebra and a suitable norm, are implementation of an approximate matrix arithmetic. They are shown by matrices of the kind $A = T(a) + E$ where, in general, $a(z) = \sum_{i\in\mathbb{Z}} a_i z^i$ is a Laurent series such that $\| a \|_w := \sum_{i\in\mathbb{Z}} |a_i|$ is finite, E is a compact correction. The norm which makes the class QT a Banach algebra is defined by

$$\| A \|_{\mathcal{QT}} := \alpha \| a \|_w + \| E \|_2, \quad \alpha = (1 + \sqrt{5})/2.$$

Here, we try to analyze the representation of quasi-Block Toeplitz (QBT) matrices with the finite floating point representation of a finite number of parameters. In fact, we show

*Speaker. Email address: shamssolary@pnu.ac.ir or shamssolary@gmail.com

the results in [1] for QBT matrices, that still makes the set QBT of a Banach algebra

$$\| \mathbf{F} \|_{\mathcal{QBT}} = \alpha \| \mathbf{F} \|_w + \| E \|_2, \quad \alpha = (1+\sqrt{5})/2.$$

$\mathbf{F}$ is a matrix-valued function in the Wiener class specify that $F(z) = \sum_{i\in\mathbb{Z}} z^i F_i$, and $T(F) = (F_{j-k})_{j,k\in\mathbb{Z}}$, $\{F_k\}_{k\in\mathbb{Z}}$ is the sequence of Fourier coefficients of $\mathbf{F}$, then

$$\| \mathbf{F} \|_w = \sum_{i=-\infty}^{\infty} \| F_i \| < \infty,$$

and E is a compact correction, see [1].
The starting point of a class of QBT matrices will be to indicate with a function in the Wiener class $\mathcal{W}$ formed by $F(w) = \sum\limits_{k=-\infty}^{\infty} e^{kwi} F_k$ where $w \in \mathbb{R}$, $F_k \in \mathbb{C}^{N\times N}$ with $k \in \mathbb{Z}$, and i denotes the imaginary unit.
The matrix-valued function $F : \mathbb{R} \to \mathbb{C}^{N\times N}$ is continuous and $2\pi-$periodic, and $\sum\limits_{k=-\infty}^{k=\infty} |[F_k]_{r,s}| < \infty$, $1 \le r \le N$, $1 \le s \le N$, that $\{F_k\}_{k=-\infty}^{\infty}$ are the sequence of Fourier coefficients of F:

$$F_k = \frac{1}{2\pi} \int_0^{2\pi} F(w) e^{-ikw} dw. \tag{1}$$

Gohberg and Krein [4] say that the operator induced by $T(F)$ is invertible if and only if $F(w)$ has a canonical right Wiener-Hopf factorization.
Here,

$$T(F) = \begin{pmatrix} F_0 & F_{-1} & F_{-2} & \dots & \dots \\ F_1 & F_0 & F_{-1} & \ddots & \vdots \\ F_2 & F_1 & F_0 & F_{-1} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{2}$$

and $\tilde{\mathbf{F}}(w) := \mathbf{F}(1/w)$ that ($w = e^{i\theta} \in \mathbb{T}$), $\mathbb{T}$ stands for the complex unit circle $\{z \in \mathbb{C} : |z| = 1\}$.
Also, the formula of Theorem 2.1 of [1] remains true in the matrix case, namely:

$$T(FG) = T(F)T(G) + H(F)H(\tilde{G}),$$

here, $\mathbf{F}$, $\mathbf{G}$ are the matrix functions and $H(F)$, $H(\tilde{G})$ are the Hankel operators.

Definition 2.1. The semi-infinite block matrix $\mathbf{F}$ is quasi-block Toeplitz matrix (QBT) if it can be written in the form

$$\mathbf{F} = T(F) + E$$

where $F(w) = \sum\limits_{k=-\infty}^{\infty} e^{kwi} F_k$ is in the Wiener class, and $E = (e_{i,j})_{i,j\in\mathbb{Z}^+}$ is compact.

If E has finite support, i.e., if only a finite number of entries is nonzero. In this case the nonzero entries of E will stay in a sufficiently large leading principal submatrix (in the top left corner). We can get $\mathbf{F} = \{F_i\}_{i\in\mathbb{Z}}$, $\| \mathbf{F} \|_w = \sum\limits_{i=-\infty}^{\infty} \| F_i \| < \infty$, while E is a matrix having only a finite number of nonzero entries containing the information concerning the boundary conditions.
Also, we can extend the norm $\| \cdot \|_{\mathcal{QT}}$ and form the Banach algebra for the class of QBT

matrices that shows by $\| \, . \, \|_{\mathcal{QBT}}$.
Now, we get

$$\| \mathbf{F} \|_{\mathcal{QBT}} = \alpha \| \mathbf{F} \|_w + \| E \|_2, \quad \alpha = (1+\sqrt{5})/2,$$

and

$$\| \mathbf{FG} \|_{\mathcal{QBT}} \leq \| \mathbf{F} \|_{\mathcal{QBT}} \| \mathbf{G} \|_{\mathcal{QBT}} \, .$$

This norm tries to solve the difficult to compute numerically by the l^2 norm and to complete the linear space of QBT matrices, similar $\| \, . \, \|_{\mathcal{QT}}$ of QT matrices.

Theorem 2.2. *Let $\mathbf{F} = T(F) + E \in \mathcal{QBT}$ and $\epsilon > 0$. Then, we can have the negative integers n_-, n_+, n_r, n_c such that the matrix $\hat{\mathbf{F}} = T(\hat{F}) + \hat{E}$ and*

$$\hat{E}_{i,j} = \begin{cases} E_{i,j} & if \quad 1 \leq i \leq n_r \\ 0 & otherwise. \end{cases} \tag{3}$$

Also, we can show $\| \mathbf{F} - \hat{\mathbf{F}} \|_{\mathcal{QBT}} \leq \| \mathbf{F} \|_{\mathcal{QBT}} .\epsilon$.

Remark 2.3. If matrix E is block form we can have $\mathbf{F} = T(F) + E \in \mathcal{QBT}$, ordinary. Otherwise, we can set these partitions for continuing to prove.
These commands do in MATLAB by

»mat2cell(.)

»cell2mat(.)

Here

$$fl(a) = a + \varepsilon, \quad |\varepsilon| \leq |a|.\epsilon,$$

that a is the real number in floating point format $fl(a)$ and ϵ is the so-called unit roundoff. Also,

$$\mathcal{QBT}(\mathbf{F}) = T(\hat{F}) + \hat{E} = \mathbf{F} + \varepsilon, \quad \| \varepsilon \| \leq \| \mathbf{F} \|_{\mathcal{QBT}} .\epsilon, \tag{4}$$

where ε is some prescribed tolerance set a priori, $\mathcal{QBT}(\mathbf{F})$ is a finite support of QBT-matrix $\mathbf{F} = T(F) + E$ by the sum of a banded block Toeplitz matrix $T(\hat{F})$ and a semi-infinite matrix $\hat{E}$. $\mathcal{QBT}(\mathbf{F})$ is a class of finitely representable quasi-block Toeplitz matrices, the lengths of the representations are not constant and can vary in order to guarantee a uniform bound to the relative error in norm.
The cqbt-commands collect some commands for operating with finitely representable quasi block Toeplitz matrices. The following MATLAB commands help us to build a block Toeplitz matrix.

» F=cqbt(neg, pos, E);

For the floating point operations where $\odot$ is any basilar arithmetic operation (sum, subtraction, multiplication, and division), we get

$$fl(a \odot b) = a \odot b + \varepsilon, \quad |\varepsilon| \leq (a \odot b).\epsilon.$$

Also, for the set of finitely representable QT matrices, we have

$$\mathcal{QT}(A \odot B) + \varepsilon, \quad \| \varepsilon \| \leq \epsilon \| A \odot B \|_{QT},$$

for any pair of finitely representable $A, B \in \mathcal{QT}$ and $\odot \in \{+, -, *, /, \backslash\}$.
Some properties of sequences of block Toeplitz matrices generated by continuous matrix-valued functions is proven by asymptotically equivalent sequences of Matrices [3, 6].
The factorization $E_{\mathbf{H}} = U_{\mathbf{H}} V_{\mathbf{H}}$ is given by

$$U_{\mathbf{H}} = [U_{\mathbf{F}}, U_{\mathbf{G}}], \quad V_{\mathbf{H}} = [V_{\mathbf{F}}, V_{\mathbf{G}}]. \tag{5}$$

$\hat{U}_{\mathbf{H}}$ and $\hat{V}_{\mathbf{H}}$ show a lower number of columns $U_{\mathbf{H}}$ and $V_{\mathbf{H}}$ such that $\| E_{\mathbf{H}} - \hat{U}_{\mathbf{H}} \hat{V}_{\mathbf{H}}^T \|_2$ is sufficiently small. Then, we get

$$\mathcal{QBT}(\mathbf{F} + \mathbf{G}) = \mathbf{F} + \mathbf{G} + \varepsilon, \quad \| \varepsilon \|_{QT} \leq \epsilon \| \mathbf{F} + \mathbf{G} \|_{QBT} .$$

Here, ε is shown the local error of the addition. The original QBT matrices $\mathbf{F}$ and $\mathbf{G}$ are represented by

$$\hat{\mathbf{F}} = \mathbf{F} + \varepsilon_{\mathbf{F}}, \quad \hat{\mathbf{G}} = \mathbf{G} + \varepsilon_{\mathbf{G}}, \tag{6}$$

and

$$\mathcal{QBT}(\hat{\mathbf{F}} + \hat{\mathbf{G}}) - (\mathbf{F} + \mathbf{G}) = \varepsilon_{\mathbf{F}} + \varepsilon_{\mathbf{G}} + \varepsilon,$$

where $\varepsilon_{\mathbf{F}} + \varepsilon_{\mathbf{G}}$ is the inherent error, ε is the local error, and the sum of the local error and the inherent error is the global error.Also, as we know from Theorem 2.1 of [1], multiplication remains true in the matrix case:

$$\mathbf{H} = \mathbf{FG} = T(FG) - H(F)H(\tilde{G}) = T(H) + E_{\mathbf{H}}, \tag{7}$$

where $H(w) = F(w)G(w)$, $\mathbf{F}$, $\mathbf{G}$ are the matrix functions and $H(F)$, $H(\tilde{G})$ are the Hankel operators.
Then, we have

$$\mathbf{H} = \mathbf{FG} = (T(F) + E_F)(T(G) + E_G) = T(F)T(G) + E_F T(G) + E_G T(F) + E_F E_G.$$

From asymptotic result about the product of block Toeplitz matrices,

$$\mathbf{H} = T(FG) + E_H$$

that inherent error is

$$E_H = E_F T(G) + E_G T(F) + E_F E_G.$$

The local error is defined by

$$\mathcal{QBT}(\mathbf{H}) - \mathcal{QBT}(\mathbf{FG}) = \varepsilon,$$

that $\| \varepsilon \|_{QT} \leq \epsilon \| \mathbf{FG} \|_{QBT}$.
The global error is

$$\mathcal{QBT}(\hat{\mathbf{F}}\hat{\mathbf{G}}) - \mathbf{FG} = \varepsilon_{\mathbf{F}}\mathbf{G} + \varepsilon_{\mathbf{G}}\mathbf{F} + \varepsilon_{\mathbf{F}}\varepsilon_{\mathbf{G}} + \varepsilon.$$

A famous theorem by Gohberg and Krein says that the operator induced by $T(F)$ is invertible if and only if $F(w)$ has a canonical right Wiener-Hopf factorization [6].
In the following way, we try to show a new finitely representable $\mathcal{QBT}$ matrix by the cqbt function:

» F=cqbt(neg, pos, E);

or

» F=cqbt(neg, pos, U, V);

Here, if we have $\mathbf{F} = T(F) + E$ that the vectors pos and neg contain the coefficients of the symbol $F(w)$ with non positive and non negative indices, respectively. E is a finite section of the correction representing its nonzero part and it is in the upper left corner. It is possible that the correction representing its nonzero part in the lower right corner. If the corrections overlap, then we switch to a single correction format, as in the semi-infinite case. This is done by storing it as an upper left correction and setting the lower right to the empty matrix. Also, in this way, we lost the sparsity and it to be convenient, the rank of the correction needs to stay small compared to the size of the matrix.
When, we represent finite quasi- block Toeplitz matrices by storing two additional matrices that represent the lower right correction in factorized form.
See the following steps:

1- Compute $\| \mathbf{F} \|_{QBT}$.

2- Obtain a truncated version $\hat{F}(w)$ of the symbol $F(w)$ by discarding the sequence of Fourier coefficients of F, such that $\| \mathbf{F} - \hat{\mathbf{F}} \|_w \leq \| \mathbf{F} \|_{QBT} \cdot \frac{\epsilon}{2\alpha}$.

3- Compute a compressed version $\hat{E}$ (a truncation error bounded by $\| \mathbf{F} \|_{QBT} \cdot \frac{\epsilon}{2\alpha}$) of the correction using the SVD and dropping negligible rows and columns.

The truncation of a QBT matrix $\mathbf{F} = T(F) + E$ is described by the above steps.
It is performed by some details of the operator $\mathcal{QBT}$ on a finitely generated QBT matrix. The QBT norm enables to recognize unbalanced representations and to completely drop the negligible part.
We set $\hat{\epsilon} = \frac{\epsilon}{4} \| \mathbf{F} \|_{QBT}$,
if $min(\| \mathbf{F}_{n-} \|, \| \mathbf{F}_{n+} \|) < \hat{\epsilon}$
then by the following commands
$F(w) = F(w) - F_{n-} w^{n-}$, $\hat{\epsilon} = \hat{\epsilon} - \| \mathbf{F}_{n-} \|$,
or
$F(w) = F(w) - F_{n+} w^{n+}$, $\hat{\epsilon} = \hat{\epsilon} - \| \mathbf{F}_{n+} \|$,
the sequence of Fourier coefficients of F and by
$\hat{\epsilon} = \hat{\epsilon} - \| E \|$, the correction's support,
will be truncated to provide a specified threshold.
The analysis of a random walk on the semi-infinite strip $\{0, ..., m\} \times \mathbb{N}$ is considered.
The random walk to be a Markov chain, and that movements are possible only to adjacent states; that is, from (i, j), one can reach only (i', j') with $|i - i'|, |j - j'| \leq 1$, with probabilities of moving up/down and left/right not depending on the current state. Then, the transition matrix $\mathbf{F}$ is an infinite quasi-Toeplitz-block-quasi-Toeplitz matrix of the form

$$\mathbf{F} = \begin{pmatrix} \hat{F}_0 & F_{-1} & & \\ F_1 & F_0 & F_{-1} & \\ & \ddots & \ddots & \ddots \end{pmatrix}.$$

If $min(\| F_{-2} \|, \| F_2 \|) < \hat{\epsilon}$ then we only have the transition probabilities are chosen in a

way that gives the following matrices:

$$F_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \\ & \ddots & \ddots & \ddots \end{pmatrix}, \quad F_0 = \begin{pmatrix} 0 & \frac{1}{5} & & \\ \frac{1}{10} & 0 & \frac{1}{5} & \\ & \ddots & \ddots & \ddots \end{pmatrix}, \quad F_{-1} = \begin{pmatrix} 1 & \frac{1}{3} & & \\ \frac{1}{2} & 1 & \frac{1}{3} & \\ & \ddots & \ddots & \ddots \end{pmatrix},$$

properly rescaled in order to make $F_{-1}+F_0+F_1$ a row-stochastic matrix. The matrices F_i are non negative tridiagonal Toeplitz matrices with corrections to the elements in position $(1,1)$ and (m,m), and satisfy $(F_{-1}+F_0+F_1)e = e$, where e is the vector of all ones.

3 Conclusions and Suggested Future Work

In this paper, we have introduced a suitable norm for approximating any QBT matrix by means of a finitely representable matrix within a given relative error bound. Then, we have analyzed the class of quasi- block Toeplitz matrices by this norm. We expand some computational aspects of a block matrix arithmetic by Matlab toolbox.
We suggest for future work that Hankel compression to store a low-rank approximation of $H(F)$ and $H(G)$ for computing the multiplication of two block Toeplitz matrices $T(F)$ and $T(G)$ in Equation (7) is possible by some strategies such as random sampling techniques and reblocking.

References

[1] D.A., Bini, S., Massei, L., Robol, Quasi-Toeplitz matrix arithmetic: a MATLAB toolbox, *Numer. Algorithm.*, 81 (2019), 741–769.

[2] D.A., Bini, S., Massei, B., Meini, On functions of quasi Toeplitz matrices, *Sb. Math.*, 208 (2017), No. 11, 56–74.

[3] D.A., Bini, S., Massei, On the exponential of semi-infinite quasi-Toeplitz matrices, *Numerische Mathematik*, 141 (2019), 319–351.

[4] A., Böttcher, S.M., Grudsky, Spectral properties of banded Toeplitz matrices. SIAM, PA, (2005).

[5] A., Böttcher, M., Halwass, A Newton method for canonical Wiener-Hopf and spectral factorization of matrix polynomials, *Electron. J. Linear Algebra*, 26 (2013), 873–897.

[6] M. Shams Solary, Quasi-Block Toeplitz matrix in MATLAB, *TWMS J. App. Eng. Math.*, Accepted (2022).

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Bounds for Norms of Matrix Functions

Majed Hamadi* and Nezam Mahdavi-Amiri

Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran

Abstract

Here we investigate the bounds for norm of matrix functions, considering the Crouzeix's conjecture. We provide bounds for the norm of a matrix function using the numerical range and Faber polynomials. In particular, we obtain bounds for norm of the sine and cosine and also the hyperbolic sine and cosine functions of matrices as exponential functions.

Keywords: Crouzeix's conjecture, Numerical range, Matrix functions
Mathematics Subject Classification [2010]: 47A30, 47A25, 65F60

1 Introduction

It is well-known that analytic function f of a square matrix A can be represented as a contour integral,

$$f(A) = \frac{1}{2\pi i}\int_{\Gamma} f(z)(zI - A)^{-1}\mathrm{d}z,$$

where f is analytic on and inside a closed contour Γ that encloses $\Lambda(A)$. For both theoretical and practical purposes it is useful to be able to bound the norm of $f(A)$. Recently, it has been of interest to find sets on the complex plane that can be associated with a square matrix A to provide more information about the norms of functions of A. Consider $\Omega \subset \mathbb{C}$ to be a smooth, bounded, convex domain. Delyon and Delyon [2] showed the existence of a best constant C_Ω such that, for all rational functions f, there holds

$$\|f(A)\| \leq C_\Omega \sup_{z\in\Omega} |f(z)|, \tag{1}$$

whenever A is a bounded linear operator in a complex Hilbert space $(H, \langle\cdot,\cdot\rangle, \|\cdot\|)$ with numerical range $\mathcal{W}(A) := \{\langle Ax, x\rangle \ : \ x \in H, \ \|x\| = 1\}$, where closure of a $\mathcal{W}(A)$ lies in Ω. Let us recall that the numerical radius $w(A)$ of a linear operator A in the Hilbert space H is the number $w(A) = \sup_{z\in\mathcal{W}(A)} |z|$. Then, the interesting result of [4] is given by

$$w\big(f(A)\big) \leq \sqrt{2} \sup_{z\in\mathcal{W}(A)} |f(z)|. \tag{2}$$

*Speaker. Email address: majed.hamadi77@sharif.edu

Using $\|A\| \leq 2w(A)$ for any linear operator A, we have $\|f(A)\| \leq 2\sqrt{2}\sup_{z\in\mathcal{W}(A)}|f(z)|$. Crouzeix and Palencia [4] showed that $C_\Omega \leq 1+\sqrt{2}$ holds in (1). The conjecture is $\sup_\Omega C_\Omega = 2$. This is a well-known conjecture, known as Crouzeix's conjecture. Other bounds on $\|f(A)\|_2$ have also been introduced in the literature.
Here, we use numerical range and the Faber polynomial to derive some bounds for $\|f(A)\|_2$. We make use of the matrix 2-norm compatible with vector 2-norm, given by the largest singular value.

2 Numerical Range and Bounds for Norm of Matrix Function

In the following, $\|f\|_S$ denotes the supremum of $|f(z)|$ over set S, and A is an $n\times n$ complex matrix.

Theorem 2.1. *[3] Let $A\in\mathbb{C}^{n\times n}$. Suppose there is $t\in[0,2\pi)$ such that $e^{it}\mathcal{W}(A)$ lies in a rectangle R centered at $z_0\in\mathbb{C}$ with vertices $z_0\pm\alpha\pm i\beta$ and $z_0\pm\alpha\mp i\beta$, where $\alpha,\beta>0$, so that $z_1=z_0+\alpha+i\beta$ has the largest magnitude. Then, we have*

$$\|A\|_2 \leq \begin{cases} |z_1|, & R\subseteq \text{conv}\{z_1,\bar{z}_1,-\bar{z}_1\},\\ \alpha+\beta, & \text{otherwise.}\end{cases}$$

Note that the bound in each case is attainable.

Let us denote $M(A,R):=\max\{|z_1|,\alpha+\beta\}$. Since $w(A)\leq\|A\|_2$, we obtain $w(A)\leq M(A,R)$. There is a constant $1\leq C(A,R)$ such that $M(A,R)\leq C(A,R)w(A)$. We now find a new bound for $\|f(A)\|_2$ as given below.

Theorem 2.2. *Let $A\in\mathbb{C}^{n\times n}$. Then, we have*

$$\|f(A)\|_2 \leq \mathcal{C}\sqrt{2}\|f\|_{\mathcal{W}(A)},$$

where $\mathcal{C}:=\sup_{B\in\mathbb{C}^{n\times n}}\inf_R C(B,R)$.

Proof. Since $f(A)\in\mathbb{C}^{n\times n}$, $\|f(A)\|_2\leq\mathcal{C}w(f(A))$, and by using inequality (2), the proof is complete. □

Remark 2.3. If $\mathcal{C}\leq\sqrt{2}$, then $\|f(A)\|_2\leq 2\|f\|_{\mathcal{W}(A)}$.

Remark 2.4. In the particular case of the unit disk $\Omega=\mathbb{D}$, if $f(0)=0$, then we have $w(f(A))\leq\|f\|_{\mathcal{W}(A)}$, and therefore,

$$\|f(A)\|_2 \leq 2w(f(A)) \leq 2\|f\|_{\mathcal{W}(A)}.$$

The result given in Remark 2.4 was obtained by Crouzeix and Palencia [4].

3 Bounds on Norms of Special Functions of Matrices

For every complex number z, we have $|e^z|=e^{\text{Re}(z)}$. Here, $\text{Re}(A)$ will denote $\text{Re}(A)=\dfrac{A+A^*}{2}$. Additionally, we will denote the real part of the numerical range of A as $\text{Re}\big(\mathcal{W}(A)\big)$. The next result has been proven in [6].

Theorem 3.1. *[6] Let $A \in \mathbb{C}^{n\times n}$. Then,*

$$\|e^A\| \leq \|e^{\mathrm{Re}(A)}\|,$$

for every unitarily invariant norm.

We have obtained the following result, which establishes a bound for $\|e^A\|_2$, using the numerical range of matrix A.

Corollary 3.2. *Let $A \in \mathbb{C}^{n\times n}$. Then,*

$$\|e^A\|_2 \leq \|e^z\|_{\mathcal{W}(A)}. \tag{3}$$

Proof. It is easy to see $\mathcal{W}\big(\mathrm{Re}(A)\big) = \mathrm{Re}\big(\mathcal{W}(A)\big)$. Then, we have

$$\begin{aligned}\|e^{\mathrm{Re}(A)}\|_2 &= \sup_{z\in\mathcal{W}\left(\mathrm{Re}(A)\right)} |e^z|,\\ &= \sup_{z\in\mathcal{W}(A)} |e^{\mathrm{Re}(z)}|,\\ &= \sup_{z\in\mathcal{W}(A)} |e^z| = \|e^z\|_{\mathcal{W}(A)}.\end{aligned}$$

Now, by Theorem 3.1, inequality (3) is established. □

In the following, for a sufficiently differentiable function f, we use $f_+(\cdot)$ to denote that for every $k \geq 0$, $f_+^{(k)}(0) \geq 0$. There are many functions of this type, such as $\exp(\cdot)$, $\sinh(\cdot)$, $\cosh(\cdot)$, etc. Next, we establish a bound for $\|f_+\big(e^A\big)\|_2$ using numerical range.

Theorem 3.3. *Let $A \in \mathbb{C}^{n\times n}$. Then,*

$$\|f_+\big(e^A\big)\|_2 \leq \|f_+(e^z)\|_{\mathrm{Re}\left(\mathcal{W}(A)\right)}.$$

Proof. Knowing $\|f_+\big(e^A\big)\|_2 \leq f_+\big(\|e^A\|\big)$ and applying Theorem 3.1, we have $\|f_+\big(e^A\big)\|_2 \leq \|f_+(e^z)\|_{\mathrm{Re}\left(\mathcal{W}(A)\right)}$. □

For functions such as $\sin(\cdot)$ and $\cos(\cdot)$ as well as hyperbolic functions $\sinh(\cdot)$ and $\cosh(\cdot)$ we can use Theorem 3.1 to find the bounds of the matrix norm of these functions. Here, we find bounds for $\|\cos(A)\|_2$ and $\|\cosh(A)\|_2$. Since $\cos(z) = \dfrac{e^{iz}+e^{-iz}}{2}$, using inequality (3) we get

$$\|\cos(A)\|_2 = \left\|\frac{e^{iA}+e^{-iA}}{2}\right\|_2 \leq \frac{1}{2}\Big(\|e^{iz}\|_{\mathcal{W}(A)} + \|e^{-iz}\|_{\mathcal{W}(A)}\Big),$$

and also for $\cosh(z) = \dfrac{e^z+e^{-z}}{2}$, we have

$$\|\cosh(A)\|_2 = \left\|\frac{e^{A}+e^{-A}}{2}\right\|_2 \leq \frac{1}{2}\Big(\|e^{z}\|_{\mathcal{W}(A)} + \|e^{-z}\|_{\mathcal{W}(A)}\Big).$$

Note that we can likewise derive bounds for $\sin(\cdot)$ and $\sinh(\cdot)$ using an exponential function and numerical range.

4 Faber Polynomial and Bounds for Norm of Matrix Function

In this section, we find a bound for $\|f(A)\|_2$ using Faber polynomials. For that, let us define continuum sets and Faber polynomials. We call $\mathcal{K}$ continuum if $\mathcal{K}$ is compact, connected, and not reduced to a point. If $\mathcal{K}$ has a connected complement, then the Riemann Mapping theorem ensures the existence of a function ϕ that maps the exterior of $\mathcal{K}$ conformally onto the set $\{z \in \mathbb{C} : |z| > 1\}$ so that $\phi(\infty) = \infty$, and $\lim_{z\to\infty} \dfrac{\phi(z)}{z} = d > 0$. With ϕ having the Laurent expansion,

$$\phi(z) = dz + \sum_{j=0}^{\infty} \frac{a_j}{z^j},$$

for every $k \geq 0$, we have

$$\big(\phi(z)\big)^k = d^k z^k + a_{k-1}^{(k)} z^{k-1} + \cdots + a_0^{(k)} + \sum_{j=1}^{\infty} \frac{a_{-j}^{(k)}}{z^j}.$$

The polynomial parts, $F_k(z) = d^k z^k + a_{k-1}^{(k)} z^{k-1} + \cdots + a_0^{(k)}$, are known as the Faber polynomials produced by the continuum $\mathcal{K}$. The following theorem of [5] shows that any analytic function can be expanded in a Faber series.

Theorem 4.1. *[5] Every function $f(z)$ analytic on a continuum $\mathcal{K}$ can be expanded in a Faber series converging uniformly on the whole $\mathcal{K}$, that is,*

$$f(z) = f_0 + \sum_{k=1}^{\infty} f_k F_k(z), \qquad z \in \mathcal{K},$$

where the coefficients, for $k \geq 0$, are given by

$$f_k = \frac{1}{2\pi i} \int_{|z|=\tau} \frac{f(\phi^{-1}(z))}{z^{k+1}} \, \mathrm{d}z.$$

We can select $\tau > 1$ in such a way that f is analytic on the complement of the set $\{\phi^{-1}(z); |z| > \tau\}$.

The following theorem gives an important result due to Beckermann.

Theorem 4.2 (Beckermann's Theorem [1])**.** *Let $\mathcal{K} \subseteq \mathbb{C}$, be convex and compact. If $A \in \mathbb{C}^{n\times n}$ is such that $\mathcal{W}(A) \subseteq \mathcal{K}$, then for all $k \in \mathbb{N}$ the Faber polynomials generated by $\mathcal{K}$ satisfy*

$$\|F_k(A)\|_2 \leq 2.$$

Now, we are ready to present a new bound for $\|f(A)\|_2$ using the Faber polynomials.

Theorem 4.3. *Let $A \in \mathbb{C}^{n\times n}$ be such that $\mathcal{W}(A) \subseteq \mathcal{K}$, with $\mathcal{K}$ compact and convex and function f be analytic on $\mathcal{K}$. With $\tau > 1$ in Theorem 4.1, we have*

$$\big\|f(A)\big\|_2 \leq |f_0| + \frac{2}{\tau - 1} \max_{|z|=\tau} |f(\phi^{-1}(z))|.$$

Proof. Function f being analytic on $\mathcal{K}$, using Theorem 4.1 we easily obtain

$$\big\|f(A)\big\|_2 = \Big\|f_0 I + \sum_{k=1}^{\infty} f_k F_k(A)\Big\|_2 \leq |f_0| + \sum_{k=1}^{\infty} |f_k| \|F_k(A)\|_2.$$

For $k \geq 0$, we have $|f_k| \leq \big(\frac{1}{\tau}\big)^k \max_{|z|=\tau} |f(\phi^{-1}(z))|$. Using Theorem 4.2, we have $\|F_k(A)\|_2 \leq 2$, and therefore

$$\begin{aligned}\big\|f(A)\big\|_2 &\leq |f_0| + 2 \max_{|z|=\tau} \big|f(\phi^{-1}(z))\big| \sum_{k=1}^{\infty} \big(\frac{1}{\tau}\big)^k, \\ &= |f_0| + \frac{2}{\tau - 1} \max_{|z|=\tau} \big|f(\phi^{-1}(z))\big|,\end{aligned}$$

completing the proof. □

A result of Theorem 4.3 is that for $\tau \geq 3$, we have

$$\big\|f(A)\big\|_2 \leq 2 \max_{|z|=\tau} |f(\phi^{-1}(z))|. \tag{4}$$

This result is not exactly Crouzeix's conjecture! Here is an example.

Example 4.4. Let $A \in \mathbb{C}^{n\times n}$ and numerical rang of A be a unit circle, that is, $\mathcal{W}(A) = \{z : |z| = 1\}$. Then, $\phi(z) = z$ and if we use (4), we obtain

$$\|f(A)\|_2 \leq 2 \sup_{|z|=3} |f(z)|.$$

However, this bound is bigger than $2\|f\|_{\mathcal{W}(\mathcal{A})}$ for some functions f.

5 Conclusion

We have proposed some new bounds for the spectral norm of matrix functions using a subset of complex plane associated with a given square matrix A, exploiting results related to the numerical range and maximum absolute value of the function over the numerical range. We have also made use of Faber polynomials to obtain alternative bounds.

References

[1] B. Beckermann, Image numérique, GMRES et polynomes de Faber, C. R. Acad. Sci. Paris Ser. I, 340 (2005), 855-860.

[2] B. Delyon and F. Delyon, Generalization of Von Neumann's spectral sets and integral representation of operators, Bull. Soc. Math. France, 127 (1999), 25-41.

[3] L. Hogben, Handbook of Linear Algebra, CRC Press, Ames, 2013, page 25-10.

[4] M. Crouzeix and C. Palencia, The numerical range is a $(1+\sqrt{2})$-spectral set, SIAM J. Matrix Anal. and Applications, 38 (2017), 649-655.

[5] P. K. Suetin, Series of Faber polynomials, Gordon and Breach Science Publishers, Amsterdam, 1998. Translated from the 1984 Russian original by E. V. Pankratiev.

[6] R. Bhatia, Matrix Analysis, Springer-Verlage, New York, 1997.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Monomial geometric optimization through fuzzy relation inequalities

Mahdi Keshtkar[1,*] and Elyas Shivanian[2]

[1]Department of Mathematics, Buein Zahra Technical University, Buein Zahra, Qazvin, Iran

[2]Department of Mathematics, Faculty of Science, Imam Khomeini International University, Qazvin 34194-288, Iran

Abstract

In this paper, an optimization model with geometric objective function is presented. Monomials are basic structural units of geometric programming and are widely used. Regarding this matter, we present geometric programming model with a monomial objective function subject to the fuzzy relation inequalities constraints with max-product composition. Simplification operations have been given to accelerate the resolution of the problem by removing the components having no effect on the solution process. Also, an algorithm is presented to abbreviate and illustrate the steps of the problem resolution.

Keywords: Monomial geometric programming, Fuzzy relation inequalities, Max-product composition.

Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36

1 Introduction

Fuzzy relation equations (FRE), fuzzy relation inequalities (FRI) and their connected problems have been investigated by many researchers in both theoretical and applied areas. Sanchez [5] started a development of the theory and applications of FRE treated as a formalized model for non-precise concepts. Generally, FRE and FRI have a number of properties that make them suitable for formulizing the uncertain information upon which many applied concepts are usually based. Fang and Li solved the linear optimization problem with respect to the FRE constraints by considering the max-min composition [1]. The max-min composition is commonly used when a system requires conservative solutions in the sense that the goodness of one value cannot compensate the badness of another value [3]. The fundamental result for fuzzy relation equations with max-product composition goes back to Pedrycz [4]. Recently, many interesting generalizations of the linear programming subject to a system of fuzzy relations have been introduced and developed based on composite operations used in FRE, fuzzy relations used in the definition of the constraints, some developments on the objective function of the problems and other ideas [2]. They extended the study of an inverse solution of a system of fuzzy relation equations

*Speaker. Email address: keshtkarmahdi@gmail.com

with max-product composition. They provided theoretical results for determining the complete sets of solutions as well as the conditions for the existence of resolutions. Their results showed that such complete sets of solutions can be characterized by one maximum solution and a number of minimal solutions. In view of the importance of geometric programming and the fuzzy relation equation in theory and applications, Yang and Cao have proposed a fuzzy relation geometric programming, discussed optimal solutions with two kinds of objective functions based on fuzzy max product operator [6].
In this paper, we consider the monomial geometric programming of the FRI with the max-product operator. This problem can be formulated as follows:

$$\begin{aligned} \min \quad & c\prod_{j=1}^{n} {x_j}^{\alpha_j} \\ \text{s.t.} \quad & A \bullet x \geq d^1 \\ & B \bullet x \leq d^2 \\ & x \in [0,1]^n \end{aligned} \tag{1}$$

Where $c, \alpha_j \in R$,$c > 0$ and $A = (a_{ij})_{m\times n}$,$a_{ij} \in [0,1]$, $B = (b_{ij})_{l\times n}$, $b_{ij} \in [0,1]$, are fuzzy matrices, $d^1 = (d_i^1)_{m\times 1} \in [0,1]^m$, $d^2 = (d_i^2)_{l\times 1} \in [0,1]^l$ are fuzzy vectors, $c = (c_j)_{n\times 1} \in R^n$ is the vector of cost coefficients, and $x = (x_j)_{n\times 1} \in [0,1]^n$ is an unknown vector, and "$\bullet$" denotes the fuzzy max-product operator as defined below. Problem (1) can be rewritten as the following problem in detail:

$$\begin{aligned} \min \quad & c\prod_{j=1}^{n} {x_j}^{\alpha_j} \\ \text{s.t.} \quad & a_i \bullet x \geq d_i^1, \quad i \in I^1 = \{1,2,...,m\} \\ & b_i \bullet x \leq d_i^2, \quad i \in I^2 = \{1,2,...,l\} \\ & 0 \leq x_j \leq 1, \quad j \in J = \{1,2,...,n\} \end{aligned} \tag{2}$$

where a_i and b_i are the ith row of the matrices A and$B$, respectively, and the constraints are expressed by the max-product operator definition as:

$$\begin{aligned} a_i \bullet x = \max_{j\in J}\{a_{ij} \cdot x_j\} \geq d_i^1 \quad & \forall i \in I^1 \\ b_i \bullet x = \max_{j\in J}\{b_{ij} \cdot x_j\} \leq d_i^2 \quad & \forall i \in I^2 \end{aligned} \tag{3}$$

2 The characteristics of the set of feasible solution

Theorem 2.1. *If $S(A,B,d^1,d^2) \neq \phi$, then for each $i \in I^1$there exist $j \in J$ such that $a_{ij} \geq d_i^1$.*

Definition 2.2. Set $\overline{x} = (\overline{x}_j)_{n\times 1}$ where

$$\overline{x}_j = \begin{cases} 1 & \forall i : b_{ij} \leq d_i^2 \\ \min\limits_{i=1,...l}\left\{\dfrac{d_i^2}{b_{ij}} : \ b_{ij} > d_i^2\right\} & \text{otherwise} \end{cases}$$

Therefore,$x \in [\overline{0}, \overline{x}]$.

Definition 2.3. Let $J_i = \{j \in J : a_{ij} \geq d_i^1\}$,$\forall i \in I^1$. For each $j \in J_i$, we define $i_{x(j)} = (i_{x(j)_k})_{n\times 1}$ such that

$$i_{x(j)_k} = \begin{cases} \frac{d_i^1}{a_{ij}} & k = j \\ 0 & k \neq j \end{cases}$$

Definition 2.4. Let $e = (e(1), e(2), ...e(m)) \in J_1 \times J_2 \times ... \times J_m$ such that $e(i) = j \in J_i$. We define $x(e) = (x(e)_j)_{n\times 1}$, in which $x(e)_j = \max\limits_{i\in I_j^e}\{i_{x(e(i))j}\} = \max\limits_{i\in I_j^e}\left\{\frac{d_i^1}{a_{ij}}\right\}$ if $I_j^e \neq \phi$ and $x(e)_j = 0$ if$I_j^e = \phi$, where $I_j^e = \{i \in I^1 : e(i) = j\}$.

Theorem 2.5. *(a) If $d_i^1 = 0$ for some $i \in I^1$, then we can remove the ith row of matrix A with no effect on the calculation of the vectors $x(e)$ for each $e \in J_I = J_1 \times J_2 \times ... \times J_m$. (b) If $j \notin J_i$, $\forall i \in I^1$, then we can remove the jth column of the matrix A before calculating the vectors $x(e)$, $\forall e \in J_I$ and set $x(e)_j = 0$ for each $e \in J_I$*

3 Simplification operations and the resolution algorithm

In order to solve problem (1), we first convert it into the two sub-problems below:

$$\begin{array}{lll} \min & \prod\limits_{j\in R^+} {x_j}^{\alpha_j} & \\ \text{s.t.} & A \bullet x = b \quad (4a) & \\ & x \in [0,1]^n & \end{array} \qquad \begin{array}{lll} \min & \prod\limits_{j\in R^-} {x_j}^{\alpha_j} & \\ \text{s.t.} & A \bullet x = b \quad (4b) & \\ & x \in [0,1]^n & \end{array}$$

where $R^+ = \{j\,|\alpha_j \geq 0, j \in J\}$ and $R^- = \{j\,|\alpha_j < 0, j \in J\}$.

Theorem 3.1. *Assume that $x(e_0)$ be an optimal solution of problem (4a) (it is possible that don't be unique) then, the optimal solution of problem (1) is x^* that defined as follow:*

$$x_j^* = \begin{cases} \overline{x}_j & j \in R^- \\ x(e_0)_j & j \in R^+ \end{cases}$$

Theorem 3.2. *The set of feasible solutions for problem (1), namely $S(A, B, d^1, d^2)$, is nonempty if and only if for each $i \in I^1$ set $\overline{J}_i = \left\{j \in J_i : \frac{d_i^1}{a_{ij}} \leq \overline{x}_j\right\}$ is nonempty.*

Theorem 3.3. *If $S(A, B, d^1, d^2) \neq \phi$, then*
$S(A, B, d^1, d^2) = \bigcup\limits_{\overline{X}(e)} [x(e), \overline{x}]$ *where* $\overline{X}(e) = \{x(e) : e \in \overline{J}_I = \overline{J}_1 \times \overline{J}_2 \times ... \times \overline{J}_m\}$.

Definition 3.4. Let $j_1, j_2 \in J$, $\alpha_{j_1} > 0$ and $\alpha_{j_2} > 0$. We say j_2 dominates j_1 if and only if
(a) $j_1 \in \overline{J}_i$ implies $j_2 \in \overline{J}_i$, $\forall i \in I^1$.
(b) For each $i \in I^1$such that $j_1 \in \overline{J}_i$we have $(\frac{d_i^1}{a_{ij_1}})^{\alpha_{j_1}} \geq (\frac{d_i^1}{a_{ij_2}})^{\alpha_{j_2}}$.

Theorem 3.5. *Suppose that j_2 dominates j_1 for $j_1, j_2 \in J$ then, the minimum value of objective function is zero. (Notification: $\alpha_{j_1} > 0$ and $\alpha_{j_2} > 0$)*

4 Algorithm for finding an optimal solution

Definition 4.1. Consider problem (1). We call $\overline{A} = \overline{(a_{ij})}_{m\times n}$ and $\overline{B} = (\overline{b}_{ij})_{l\times n}$the characteristic matrices of matrix A and matrix B, respectively, where $\overline{a}_{ij} = \frac{d_i^1}{a_{ij}}$ for each $i \in I^1$and$j \in J$, also $\overline{b}_{ij} = \frac{d_i^2}{b_{ij}}$ for each $i \in I^2$ and $j \in J$. (set $\frac{0}{0} = 1$ and $\frac{k}{0} = \infty$)

Algorithm 4.2. Given problem (2),

1. Find matrices $\overline{A}$ and $\overline{B}$.
2. If there exists $i \in I^1$ such that $\overline{a}_{ij} > 1$,$\forall j \in J$, then stop. Problem 2 is infeasible.
3. Calculate $\overline{x}$ from $\overline{B}$.
4. If there exists $i \in I^1$ such that $d_i^1 = 0$, then remove the i'th row of matrix $\overline{A}$.
5. If$\overline{a}_{ij} > \overline{x}_j$, then set $\overline{a}_{ij} = 0$,$\forall i \in I^1$ and $\forall j \in J$.
6. If there exists $i \in I^1$ such that $\overline{a}_{ij} = 0$,$\forall j \in J$, then stop. Problem (2) is infeasible.
7. If there exists $j' \in J$ such that $\overline{a}_{ij'} = 0$,$\forall i \in I^1$, then remove the j'th column of the matrix $\overline{A}$ and set $x(e_0)_{j'} = 0$. If $j' \in R^+$ then $\forall e \in J_I$, $x(e)$ is the optimal solution of (4a) and minimum value of objective function is zero. Therefore, stop.
8. If j_2 dominates j_1, $(j_1, j_2 \in R^+)$ then remove column j_1 from $\overline{A}$, $\forall j_1, j_2 \in J$ and set $x(e_0)_{j_1} = 0$ then$\forall e \in J_I$, $x(e)$ is the optimal solution of (4a) and minimum value of the objective function is zero. Therefore, stop.
9. Let $J_i^{new} = \{j \in \overline{J}_i : \overline{a}_{ij} \neq 0\}andJ_I^{new} = J_1^{new} \times J_2^{new} \times ... \times J_m^{new}$. Find the vectors $x(e)$, $\forall e \in J_I^{new}$.
10. Find x^*.

5 Conclusion

In this paper, we studied the monomial geometric programming with fuzzy relational inequalities constraints defined by the max-product operator. Since the difficulty of this problem is finding the minimal solutions optimizing the same problem with the objective function $\prod_{j\in R^+} {x_j}^{\alpha_j}$, we presented an algorithm together with some simplification operations to accelerate the problem resolution.

References

[1] S. C. Fang, G. Li, Solving fuzzy relations equations with a linear objective function, Fuzzy Sets and systems. 103(1999) 107–113.

[2] A. Ghodousian, On The Frank FREs and Its Application in Optimization Problems, Journal of Computer Science Applications and Information Technology 3(2) (2018) 1-14.

[3] J. Loetamonphong, and S.-C. Fang, Optimization of Fuzzy Relation Equations with Max-product Composition Fuzzy Sets and Systems 118, (2001) 509–517

[4] W. Pedrycz, On Generalized fuzzy relational equations and their applications, Journal of Mathematical Analysis and Applications 107 (1985), 520-536.

[5] E. Sanchez, Solution in composite fuzzy relation equations: Application to medical diagnosis in Brouwerian logic, In Fuzzy Automata and Decision Processes, (Edited by M. M. Gupta , G. N. Saridis and B R Games), pp.221-234, North-Holland, New York , (1977).

[6] Yang, J. H., & Cao, B. Y. (2005a). Geometric programming with fuzzy relation equation constraints. 2005 IEEE International Fuzzy Systems Conference Proceedings, Reno, Nevada, May 22–25, 557–560.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Maps preserving the parallel sum of operators

Hasan Karimi*

Department of Mathematics, Tehran University, Tehran, Iran

Abstract

The general form of bijective transformations of the set of all positive linear operators on a Hilbert space which preserve means is described.

Keywords: Positive operator, harmonic mean, parallel sum, arithmetic mean
Mathematics Subject Classification [2010]: 47A05, 47A56, 47B65

1 Introduction

Let $\mathcal{H}$ be a complex Hilbert space with inner product $\langle\cdot,\cdot\rangle$. Denote by $\mathbf{B}(\mathcal{H})$ the algebra of all bounded linear operators on $\mathcal{H}$. As usual, an operator $A \in \mathbf{B}(\mathcal{H})$ is called positive if $\langle Ax, x\rangle \geq 0$ holds for every $x \in \mathcal{H}$ and in that case we write $A \geq 0$. The set of all positive operators on $\mathcal{H}$ is denoted by $\mathbf{B}(\mathcal{H})^+$. For self-adjoint operators $A, B \in \mathbf{B}(\mathcal{H})$ we write $B \geq A$ if and only if $B - A \geq 0$.

For arbitrary positive operators $A, B \in \mathbf{B}(\mathcal{H})^+$, their harmonic mean $A!B$ is defined by

$$A!B = \max\left\{X \geq 0 : \begin{bmatrix} 2A & 0 \\ 0 & 2B \end{bmatrix} \geq \begin{bmatrix} X & X \\ X & X \end{bmatrix}\right\}.$$

This concept was introduced by Ando in [2]. We list some important properties of the harmonic mean (see [4]).

(1) $A!B = B!A$.

(2) For any $\lambda \in \mathbb{R}^+$ we have $(\lambda A)!(\lambda B) = \lambda(A!B)$.

(3) If $C \geq A$ and $D \geq B$, then $C!D \geq A!B$.

(4) We have $S(A!B)S^* = (SAS^*)!(SBS^*)$ for every invertible bounded linear or conjugate-linear operator S on $\mathcal{H}$.

(5) Suppose $A_1 \geq A_2 \geq \cdots \geq 0, B_1 \geq B_2 \geq \cdots \geq 0$ and $A_n \to A, B_n \to B$ strongly. Then we have that $A_n!B_n \to A!B$ strongly.

(6) $A!A = A, I!A = 2A(I + A)^{-1}$ and $0!A = 0$.

*Speaker. Email address:Hassankarimi7799@gmail.com

(7) $A!B = 2A(A+B)^{-1}B$ if A or B is invertible.

(8) $A!B = 2(A^{-1}+B^{-1})^{-1}$ if A and B are both invertible.

The harmonic mean is well-known to have important applications in operator theory but recently it has found serious applications in other areas, for example, in quantum information theory as well (see [5]).

There is a concept closely related to the harmonic mean called parallel sum. For arbitrary positive operators $A, B \in \mathbf{B}(\mathcal{H})^+$, their parallel sum $A : B$ is expressed as

$$A : B = \frac{1}{2}(A!B).$$

This notion originally defined by Anderson and Duffin [1] in a different way has many important applications in operator theory and in electrical network theory, too. See [3] and the references therein.

Finally, for arbitrary positive operators $A, B \in \mathbf{B}(\mathcal{H})^+$, their arithmetic mean $A\nabla B$ is defined by

$$A\nabla B = \frac{1}{2}(A+B).$$

For the most classical results concerning this operation we refer the reader to [6].

In the next section, we described the structure of all bijective maps on $\mathbf{B}(\mathcal{H})^+$ which preserve means.

2 Main results

The transfer property shows that for an arbitrary invertible bounded linear or conjugate-linear operator S, the transformation $A \longrightarrow SAS^*$ is a bijective map of $\mathbf{B}(\mathcal{H})^+$ respecting the operation of the harmonic mean. The content of our first result is that the converse is also true.

Theorem 2.1. *Let $\Phi : \mathbf{B}(\mathcal{H})^+ \longrightarrow \mathbf{B}(\mathcal{H})^+$ be a bijective map satisfying*

$$\Phi(A!B) = \Phi(A)!\Phi(B) \qquad (A, B \in \mathbf{B}(\mathcal{H})^+).$$

Then there is an invertible bounded linear or conjugate-linear operator S on $\mathcal{H}$ such that Φ is of the form

$$\Phi(X) = SXS^* \qquad (X \in \mathbf{B}(\mathcal{H})^+).$$

As a result of the above theorem we shall obtain the following description of bijective maps preserving the parallel sum.

Theorem 2.2. *Let $\Phi : \mathbf{B}(\mathcal{H})^+ \longrightarrow \mathbf{B}(\mathcal{H})^+$ be a bijective map satisfying*

$$\Phi(A : B) = \Phi(A) : \Phi(B) \qquad (A, B \in \mathbf{B}(\mathcal{H})^+).$$

Then Φ respects the operation of the harmonic mean. Consequently, there exists an invertible bounded linear or conjugate-linear operator S on $\mathcal{H}$ such that Φ is of the form

$$\Phi(X) = SXS^* \qquad (X \in \mathbf{B}(\mathcal{H})^+).$$

Finally, to make our investigation more complete we conclude with the following rather result concerning the structure of bijective maps of $\mathbf{B}(\mathcal{H})^+$ preserving the arithmetic mean.

Theorem 2.3. *Let* $\Phi : \mathbf{B}(\mathcal{H})^+ \longrightarrow \mathbf{B}(\mathcal{H})^+$ *be a bijective map satisfying*

$$\Phi(A\nabla B) = \Phi(A)\nabla\Phi(B) \qquad (A, B \in \mathbf{B}(\mathcal{H})^+).$$

Then there exists an invertible bounded linear or conjugate-linear operator S *on* $\mathcal{H}$ *such that* Φ *is of the form*

$$\Phi(X) = SXS^* \qquad (X \in \mathbf{B}(\mathcal{H})^+).$$

References

[1] W. N. Anderson and R. J. Duffin, *Series and parallel addition of matrices*, J. Math. Anal. Appl. **26** (1969), 576–594.

[2] T. Ando, *Topics on Operator Inequalities*, Mimeographed Lecture Notes, Hokkaido University, Sapporo, 1978.

[3] C. Fu, M. S. Moslehian and Q. Xu and A. Zamani, *Generalized parallel sum of adjointable operators on Hilbert* C^**-modules*, Linear Multilinear Algebra **70** (2022), no. 12, 2278–2296.

[4] F. Kubo and T. Ando, *Means of positive linear operators*, Math. Ann. **246** (1980), 205–224.

[5] L. Molnár, *Maps preserving the harmonic mean or the parallel sum of positive operators*, Linear Algebra Appl. **430** (2009), 3058–3065.

[6] F. Zhang, *Matrix Theory. Basic Results and Techniques (2nd edition)*, Springer-Verlag, New York, 2011.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

An efficient algorithm for solving fractional Sturm-Liouville differential operators with a constant delay

Mohammad Shahriari*

Department of Mathematics, Faculty of Science, University of Maragheh, Box 55136-553, Maragheh, Iran.

Abstract

In this manuscript, we present a simple and efficient computational algorithm for solving eigenvalue problems of fractional second-order differential operators with a constant delay inside the interval. By transforming the governing fractional differential equations with a constant delay into a linear system of algebraic equations, we can obtain the corresponding polynomial characteristic equations for kinds of boundary conditions based on the polynomial expansion and integral technique. Then, the eigenvalues can be calculated by finding the roots of the corresponding characteristic polynomial. The numerical results demonstrate reliability and efficiency of the proposed algorithm.

Keywords: Fractional Sturm–Liouville problems, Eigenvalues, Eigenfunctions, Delay differential equations, Polynomial expansion.

Mathematics Subject Classification [2010]: 34B24, 34B27

1 Introduction

The Sturm–Liouville Problems (SLPs) play a significant role in many areas of science, engineering, and mathematics [1] and [6]. Since many thousands of papers, many review articles, monographs and books have been written on this and related subjects, and yet this field remains an active field of research [6].

In this paper, we present a simple and efficient computational algorithm for solving fractional eigenvalue problems with a constant delay. Differential equations with delay arise in various problems of mathematics as well as in applications (see the monographs [4] and the references therein). To this end, we introduce the notion fractional Sturm–Liouville differential operators with a constant delay.

*Speaker. Email address: shahriari@maragheh.ac.ir

2 Fractional Sturm–Liouville differential operators with a constant delay

We consider the following special class of the fractional Sturm–Liouville problems with a constant delay:

$$\sum_{s=1}^{2} q_s(x) D^{\alpha_s} y(x) + q_0(x) y(x) + q(x) y(x-d) = \lambda w(x) y(x), \qquad x \in (a,b) \tag{1}$$

with the boundary conditions

$$\sum_{l=0}^{1} p_{1l} y^{(l)}(a) = 0, \qquad \sum_{l=0}^{1} p_{2l} y^{(l)}(b) = 0, \tag{2}$$

where $s-1 < \alpha_s \leq s$, $d \in (a,b)$, $q(x) = 0$ for $x < d$, and $q_2(x) > 0$, $w(x) > 0$ on (a,b); the functions $q_s(x)$ $(s = 0,\ 1,\ 2)$, $q(x)$, $w(x)$ are real. The notation D^{α} for any $\alpha \in \mathbb{R}^+$ denotes the left sided Caputo fractional derivative defined by

$$D^{\alpha} y(x) = \frac{1}{\Gamma(m-\alpha)} \int_0^x (x-t)^{m-\alpha-1} y^{(m)}(t) dt, \ x > 0. \tag{3}$$

where $m = [\alpha]$. We present some essential information about fractional calculus theory that will be intensively used in this paper.

Definition 2.1. The left sided Riemann–Liouville fractional integral operator of order α is defined by

$$J^{\alpha} y(x) = \frac{1}{\Gamma(\alpha)} \int_0^x (x-t)^{\alpha-1} y(t) dt \tag{4}$$

where $y \in L_1[0,T]$ and $\alpha \in \mathbb{R}^+$.

Some useful properties of the operator J^{α} are summarized in the following lemma [5]

Lemma 2.2. *Let $\alpha,\ \beta,\ x > 0$, and $\gamma > -1$. Then*

1. $J^{\alpha} J^{\beta} = J^{\alpha+\beta} = J^{\beta} J^{\alpha}$,
2. $J^{\alpha} x^{\gamma} = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+\alpha+1)} x^{\gamma+\alpha}$.

We note that the left side of Caputo fractional derivative (3) is originally defined via the left sided Riemann–Liouville fractional integral (4), (see [5]). So we have

$$D^{\alpha} y(x) = J^{m-\alpha} y^{(m)}(x) \quad x > 0.$$

Lemma 2.3. *For $\alpha \in \mathbb{R}^+$, $m = [\alpha]$ and $y \in L_1[0,T]$ we have*

1. $D^{\alpha} J^{\alpha} y(x) = y(x)$,
2. $J^{\alpha} D^{\alpha} y(x) = y(x) - \sum_{k=0}^{m-1} y^{(k)}(0+) \frac{x^k}{k!}$,
3. $D^{\alpha} x^r = \begin{cases} \frac{\Gamma(r+1)}{\Gamma(r+1-\alpha)} x^{r-\alpha}, & \text{for } [\alpha] < r; \\ 0, & \text{for } [\alpha] \geq r. \end{cases}$

For Case 3 we use the notation $\Gamma_{\alpha,r} = \frac{\Gamma(r+1)}{\Gamma(r+1-\alpha)}$. Unfortunately, it is difficult to obtain exact eigenvalues of Eqs. (1)–(2). Therefore, the numerical methods must be proposed to solve such an eigenvalue problem. In this part, avoiding solving Eq. (1) directly, by using the similar methods of [3], we introduce a simple method to determine the eigenvalues of the Sturm–Liouville equation with constant delay. For this purpose, we expand the solution $y(x)$ of Eq. (1) in the following polynomial form:

$$y(x) = \sum_{i=0}^{N} c_i x^i + R_N(x)$$

where c_i are unknown coefficients, R_N is the rest, and N is a certain positive integer which is chosen large enough such that the rest has a negligible error. We further assume that our solution can be approximated by

$$y(x) \approx \sum_{i=0}^{N} c_i x^i, \tag{5}$$

and the delay function with a constant d, $y(x-d)$, can be approximated by

$$y(x-d) \approx \sum_{i=0}^{N} c_i (x-d)^i = \sum_{i=0}^{N} c_i \sum_{k=0}^{i} \frac{i!}{k!(i-k)!} x^k d^{i-k}. \tag{6}$$

Ultimately, the main idea is to obtain a homogeneous system of equation in the unknowns c_i, $i = 0, 1, \ldots, N$, the roots of whose characteristic equation constitute the eigenvalues of the problem. First, from the boundary conditions (2) and using (5) we get

$$\sum_{i=0}^{N} (f_{m,i} - \lambda k_{m,i}) c_i = 0, \qquad m = 0, 1,$$

where

$$f_{0,i} = \sum_{m=0}^{N} \sum_{n=0}^{1} D_m^n P_{1n} a^{m-n}, \quad k_{0,i} = 0, \qquad i = 1, 2, \ldots N \tag{7}$$

and

$$f_{1,i} = \sum_{m=0}^{N} \sum_{n=0}^{1} D_m^n P_{2n} b^{m-n}, \quad k_{1,i} = 0, \qquad i = 1, 2, \ldots N \tag{8}$$

with

$$D_0^n = 1, \qquad D_m^n = \prod_{i=m-n+1}^{m} i.$$

Next, substituting (5) and (6) in (1), we get

$$\sum_{i=0}^{N} c_i \left(\sum_{s=0}^{2} \Gamma_{\alpha_s,i} q_s(x) x^{i-\alpha_s} + q(x)(x-d)^i \right) - \lambda \sum_{i=0}^{N} c_i w(x) x^i = 0. \tag{9}$$

where $\Gamma_{\alpha_0,i} = 1$ and $\alpha_0 = 0$. Notice that we have $N+1$ unknown coefficients c_i,$i = 0, 1, \ldots, N$ and only two equations, of the boundary conditions (2). we need an additional $N-1$ equations. To obtain these equations, we multiply Eq. (9) in x^l for $l = 0, 1, \ldots, N-2$ and integrate with respect to x from a to b. So, we get the following coefficients

$$f_{ji} = \sum_{s=0}^{2} \left(\int_a^b \Gamma_{\alpha_s,i} q_s(x) x^{i+j-\alpha_n-2} \right) dx + \int_a^b q(x)(x-d)^{i+j-2} dx,$$

$$k_{ji} = \int_a^b w(x)x^{i+j-2}dx. \tag{10}$$

From (7)–(8) and (10) we obtain the following linear system of $N+1$ equations.

$$\sum_{i=0}^{N}(f_{ji} - \lambda k_{ji})c_i = 0, \qquad j = 0, 1, \ldots, N. \tag{11}$$

For simplicity, the system (11) can be written in matrix form

$$(\mathbf{F} - \lambda\mathbf{K})\mathbf{c} = 0$$

where $\mathbf{F}$ and $\mathbf{K}$ are square $(N+1)(N+1)$ matrices with $F_{mi} = f_{mi}$ and $K_{mi} = k_{mi}$, and $\mathbf{c} = (c_0, c_1, \ldots, c_N)^t$. To obtain a non-trivial solution of the system of equations, the determinant of the coefficient matrix of the system must be vanish; then we get a characteristic function in eigenvalues λ:

$$\det(\mathbf{F} - \lambda\mathbf{K}) = 0 \tag{12}$$

such that $\det(\mathbf{F} - \lambda\mathbf{K})$ is a polynomial of degree $N-1$ in λ. The eigenvalues of the original problem would be those that satisfy (12). In our simulations, we solve (12) using the Matlab built-in function *solve()*. Using the Eq. (12) with a simple modified in **algorithms 2.4, 2.5**, we obtain the approximation of eigenvalues in Eq. (1) with the boundary conditions (2).

Algorithm 2.4. Find eigenvalues of problem (1)–(2).

1. Define the symbol x and λ.
2. Define $a,\ b,\ p_{kj},\ q_j(x),\ q(x),\ w(x)$.
3. Get the numbers $N,\ d$, the degree of power series and delay constant.
4. Construct the matrices $\mathbf{F}$ and $\mathbf{K}$ using (10)
5. Define the symbolic matrix $\mathbf{A} = \mathbf{F} - \lambda\mathbf{K}$
6. Calculate $P = \det(\mathbf{A})$
7. Solve for the eigenvalues of $\mathbf{A}$ by calling the Matlab function $solve(P)$.

Algorithm 2.5. Find the eigenfunction for a particular eigenvalues λ of problem (1)–(2).

1. Substitute the value of the particular λ in the matrix $\mathbf{A}$ by using the **Algorithm 2.4**.
2. Use the function $null(\mathbf{A})$ to fined a basis for the null space of $\mathbf{A}$.
3. Pike the vector from $null(\mathbf{A})$ and construct the eigenfunction, $y(x)$ by (5).
4. Normalize the eigenfunction such that $y(0) = 1$.

3 Numerical results

In this section, we consider three examples to demonstrate the performance and efficiency of the present algorithm.

Example 3.1. Consider the following fractional Sturm-Liouville differential operator with a constant delay

$$-D^{\alpha_2}y(x) + q(x)y(x-d) = \lambda y(x), \qquad 1 < \alpha_2 \leq 2, \quad 0 < d \leq 1 \tag{13}$$

with the following Dirichlet boundary conditions

$$y(0) = 0, \quad y(1) = 0. \tag{14}$$

Remark 3.2. Let us consider the function $S(x,\lambda)$ be the solution of (13) in $\alpha_2 = 2$ with the initial conditions $S(0,\lambda) = 0,\ S'(0,\lambda) = 1$. Using the method of [2], the solution of $S(x,\lambda)$ satisfy in the following integral equation

$$S(x,\lambda) = \frac{\sin\rho x}{\rho} + \int_0^x \frac{\sin\rho(x-t)}{\rho} q(t)S(t-d,\lambda)dt \tag{15}$$

where $\lambda = \rho^2$. Let $M \in \mathbb{N}$ be such that $dM < 1 \leq d(M+1)$, applying the method of [2], and using successive approximation of the solution, we get

$$S(x,\lambda) = S_0(x,\lambda) + S_1(x,\lambda) + \cdots + S_M(x,\lambda),$$

where

$$S_0(x,\lambda) = \frac{\sin\rho x}{\rho}, \qquad x > 0$$

$$S_k(x,\lambda) = \begin{cases} 0, & x \leq kd; \\ \int_{kd}^x \frac{\sin\rho(x-t)}{\rho} q(t)S_{k-1}(t-d,\lambda)dt, & x \geq kd; \end{cases}$$

with $k = 0, 1, \ldots, M$. For the value of $d = \frac{1}{2}$, the function $S(x,\lambda)$ and the characteristic function $\Delta(\lambda)$ of the problem (13)–(14) are the following form

$$S(x,\lambda) = \frac{\sin\rho x}{\rho} + \frac{1}{\rho^2}\int_{0.5}^x \sin\rho(x-t)q(t)\sin\rho(t-0.5)dt,$$

and

$$\Delta(\lambda) = S(1,\lambda) = \frac{\sin\rho}{\rho} + \frac{1}{\rho^2}\int_{0.5}^1 \sin\rho(1-t)q(t)\sin\rho(t-0.5)dt.$$

Using the Maple we get the exact eigenvalues.

Example 3.3. Consider the following second order fractional eigenvalue problem

$$-D^{\alpha_2}y(x) + y'(x) + q(x)y(x-d) = \lambda y(x), \qquad 1 < \alpha_2 \leq 2, \quad 0 < d \leq 1$$

subject to

$$y'(0) = 0, \qquad y(1) = 0$$

where $\alpha \in (1,2]$ and $d \in [0,1]$.

Table 1: The numerical and exact values of the first 5 eigenvalues for $d = 0.5$ in example 3.1

k	$q(x) = x$, λ_k $\alpha_2 = 2$	Λ_k^{14} $\alpha_2 = 1.999$	Λ_k^{14}, $\alpha_2 = 1.99$	Λ_k^{14}, $\alpha_2 = 1.9$	Λ_k^{14}, $\alpha_2 = 1.8$
1	10.111188381	10.107029283	10.0707261115	9.8301667944	9.9078432482
2	39.100688361	39.035103202	38.450376634	33.074177518	27.704259111
3	88.747593705	88.564931786	86.944041972	72.907275127	61.929691970
4	158.28800860	157.90834521	154.54090890	125.05003657	98.183372017
5	246.78813615	246.12160642	240.29223183	191.28914444	156.26481771
time		8.731452s	8.709697s	8.655266s	8.737939s
k	$q(x) = 10x^2$, $\Lambda_k^{18}\mathbf{d} = 1$	$\alpha_2 = 1.8$ Λ_k^{18} $d = 0.8$	Λ_k^{18}, $d = 0.6$	Λ_k^{18}, $d = 0.4$	Λ_k^{18}, $d = 0.2$
1	9.4568542544	10.003484537	12.790356410	16.279898757	14.426826949
2	28.476913809	26.435171515	21.850852185	21.812526078	30.297250524
3	62.200127487	71.388162429	66.652152471	57.03149016	63.317894533
4	97.064166077	84.059665889	102.48943632	97.930998240	91.804669312
5	155.43383353		140.91736445	167.0503499	150.00113927
time	11.512792	14.115463s	13.556020s	13.796776s	14.236312s

Table 2: The numerical values of the first 5 eigenvalues for $\alpha = 1.95$ $\alpha = 1.9$ $\alpha = 1.85$ in example 3.1 for different value of $d = 1, 0.9, 0.8, 0.7$.

k	$q(x) = x^2$, λ_k [3], $d = 1$	$\alpha = 1.95$ Λ_k^{14} $d = 1$	Λ_k^{14}, $d = 0.9$	Λ_k^{14}, $d = 0.8$	Λ_k^{14}, $d = 0.7$
1	3.628756	3.62875668861	3.65493703149	3.71198751715	3.77665574530
2	22.217377	22.2173745753	22.1351247869	21.9864490632	21.8864632581
3	57.581008	57.5810213745	57.7238921759	57.8916941330	57.8140827622
4	109.571053	109.57104279	109.370712555	109.303265831	109.690458471
5	177.718889	177.718878641	177.966612015	177.807850090	177.267167768
k	$q(x) = x^2$, λ_k [3], $d = 1$	$\alpha = 1.9$ Λ_k^{14} $d = 1$	Λ_k^{14}, $d = 0.9$	Λ_k^{14}, $d = 0.8$	Λ_k^{14}, $d = 0.7$
1	3.648054	3.645754164698	3.67999698010	3.74564063136	3.81708640051
2	21.215116	21.21841069258	21.1051244377	21.9864490632	20.8157158096
3	52.999304	52.99314000255	53.2143457196	53.4315378641	53.2948012012
4	98.902945	98.91110938039	98.5629052855	98.4980572451	99.1079951264
5	157.892919	157.9369978455	158.331262676	158.024364006	157.128495247
k	$q(x) = x^2$, λ_k [3], $d = 1$	$\alpha = 1.85$ Λ_k^{20} $d = 1$	Λ_k^{20}, $d = 0.9$	Λ_k^{20}, $d = 0.8$	Λ_k^{20}, $d = 0.7$
1	3.681348	3.6813486081	3.7205356816	3.796393212	3.8754551858
2	20.429199	20.429194203	20.279517899	20.0548166738	19.940970973
3	49.087410	49.08742698	49.4179200346	49.7047387885	49.465438672
4	89.939834	89.939833559	89.351407076	89.3068005612	90.308996876
5	141.072730	141.07144165	142.00458721	141.259425561	139.75998119

References

[1] W. O. Amrein, A. M. Hinz, and D. B. Pearson, *Sturm–Liouville theory: past and present*. Springer Science and Business Media, 2005.

[2] G. Freiling and V. A. Yurko, Inverse problems for Sturm–Liouville differential operators with a constant delay. *Applied Mathematics Letters*, 25(11), (2012), 1999-2004.

[3] M. A. Hajji, Q. M. Al-Mdallal, and F. M. Allan, An efficient algorithm for solving higher–order fractional Sturm–Liouville eigenvalue problems. *Journal of Computational Physics*, 272, (2014), 550-558.

[4] F. Hartung and M. Pituk, *Recent advances in delay differential and difference equations*. Springer International Publishing, 2014.

[5] I. Podlubny, *An introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. Math. Sci. Eng, 198, 340, 1999.

[6] A. Zettl, *Sturm–liouville theory* (No. 121). American Mathematical Soc., 2005.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Computing Gröbner bases of an expanded set of polynomials

Rahim Rahmati-Asghar*

Department of Mathematics, Faculty of Basic Sciences, University of Maragheh, P. O. Box 55181-83111, Maragheh, Iran.

Abstract

In this paper, we prepare some structural results on Gröbner bases of expanded toric ideals in polynomial rings. In this way, we can use the algorithm presented in [3] to computing Gröbner bases of polynomial ideals by using Macaulay matrices.

Keywords: Macaulay matrix; polynomials idels; Gröbner basis.
Mathematics Subject Classification [2010]: 00A69,11D45,05A17

1 Introduction

Let K be a field and $X = \{x_1, \ldots, x_n\}$ a set of n indeterminates. $[X]$ denotes the commutative monoid of power-products in X, i. e. the set of all terms of the form $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ for $\alpha_i \in \mathbb{N}$ $(i = 1, \ldots, n)$ endowed with the usual multiplication of such terms, and $K[X]$ denotes the polynomial ring in X over K, i. e. all K-linear combinations of power-products in $[X]$ with the usual addition and multiplication. A polynomial of the form ct, for $c \in K\backslash\{0\}$ and $t \in [X]$, is called a *monomial*.

Let F be a finite list of polynomials and $T \subset [X]$ finite; let m be the length of F and $l = |T|$. The *Macaulay matrix* $\mathrm{Mac}(F, T)$ of F with respect to T is the matrix $A \in K^{m\times l}$ such that the (i, j)-th entry of $\mathrm{Mac}(F, T)$ is the coefficient of the j-th largest power-product in T in the i-th polynomial in F (see [3]).

let I be a monomial ideal with the set of minimal generators $G(I) = \{\mathbf{x}^{\mathbf{a}_1}, \ldots, \mathbf{x}^{\mathbf{a}_r}\}$ where $\mathbf{x}^{\mathbf{a}_i} = x_1^{\mathbf{a}_i(1)} \ldots x_n^{\mathbf{a}_i(n)}$ for $\mathbf{a}_i = (\mathbf{a}_i(1), \ldots, \mathbf{a}_i(n)) \in \mathbb{Z}_+^n = \{\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{Z}^n : u_i \geq 0\}$. For the n-tuple $\alpha = (k_1, \ldots, k_n) \in \mathbb{N}^n$, Bayati and Herzog [1] defined the expansion of I with respect to α in the following form:

Let $S^\alpha = K[x_{11}, \ldots, x_{1k_1}, \ldots, x_{n1}, \ldots, x_{nk_n}]$ be a polynomial ring over K and set $P_j = (x_{j1}, \ldots, x_{jk_j})$ a prime monomial ideal in S^α for all $1 \leq j \leq n$. The expansion of I with respect to α, denoted by I^α, is the monomial ideal

$$I^\alpha = \sum_{i=1}^{r} P_1^{\mathbf{a}_i(1)} \ldots P_n^{\mathbf{a}_i(n)} \subset S^\alpha$$

where $\mathbf{a}_i(j)$ is the j-th component of the vector $\mathbf{a}_i$.

*Speaker. Email address: rahmatiasghar.r@gmail.com

Let $\mathcal{A} = \{u_1, \ldots, u_m\}$ be a set of monomials belonging to $S = K[x_1, \ldots, x_n]$ and suppose that the affine semigroup ring $K[\mathcal{A}] = K[u_1, \ldots, u_m]$ is a homogeneous K-algebra. Let $S_{\mathcal{A}} := K[y_{u_1}, \ldots, y_{u_m}]$ be the polynomial ring in n variables over K with each $\deg(y_{u_i}) = 1$ and let $I_{\mathcal{A}}$ denote the kernel of the surjective homomorphism $\varphi_{\mathcal{A}} : S_{\mathcal{A}} \to K[\mathcal{A}]$ defined by $\varphi_{\mathcal{A}}(y_{u_i}) = u_i$ for all $1 \leq i \leq m$. $I_{\mathcal{A}}$ and $K[\mathcal{A}]$ are, respectively, called *toric ideal* and *toric ring* of $\mathcal{A}$. For more details on toric rings refer to [2].

We recall the concept of *combinatorial pure subring* of a toric ring, introduced in [4], which we will use in the rest of the paper. Let $T \subseteq [n] := \{x_1, \ldots, x_n\}$. If T is a nonempty subset of $[n]$, then we set $\mathcal{A}_T := \mathcal{A} \cap K[\{x_i : x_i \in T\}]$. A subring of $K[\mathcal{A}]$ of the form $K[\mathcal{A}_T]$ with $\emptyset \neq T \subseteq [n]$ is called a combinatorial pure subring of $K[\mathcal{A}]$. For $\mathcal{A}_T = \{u_{i_1}, \ldots, u_{i_r}\}$, we set $S_{\mathcal{A}_T} = \{y_{u_{i_1}}, \ldots, y_{u_{i_r}}\}$. Therefore $I_{\mathcal{A}_T} = I_{\mathcal{A}} \cap S_{\mathcal{A}_T}$.

2 Computing Gröbner basis

Proposition 2.1. *[5] Let $\alpha \in \mathbb{N}^n$ and let $\mathcal{A} = \{u_1, \ldots, u_m\}$ be a set of monomials belonging to $S = K[x_1, \ldots, x_n]$. If $\mathcal{G}$ is the reduced Gröbner basis of $I_{\mathcal{A}^\alpha}$ with respect to a term order $<$ on $S_{\mathcal{A}^\alpha}$, then $\mathcal{G} \cap K[\mathcal{A}]$ is the reduced Gröbner basis of $I_{\mathcal{A}}$ with respect to a term order induced by $<$ on $S_{\mathcal{A}}$.*

Proposition 2.2. *Let $\mathcal{A}$ be a finite set of monomials belonging to $S = K[x_1, \ldots, x_n]$ and let $\alpha = (k_1, \ldots, k_n) \in \mathbb{N}^n$. For $\beta = \alpha + \epsilon_i$ there exists a K-algebra isomorphism*

$$\varphi : K[(\mathcal{A}^\alpha)^\gamma] \to K[\mathcal{A}^\beta]$$

where $\gamma = \mathbf{1} + \epsilon_{ik_i} \in \mathbb{N}^{|\alpha|}$. Here $\mathbf{1}$ is the vector in $\mathbb{N}^{|\alpha|}$ with all components 1.

Let $\mathcal{A} = \{u_1, \ldots, u_m\}$ be a set of monomials belonging to $S = K[x_1, \ldots, x_n]$. Define the term order "$<^\sharp_{\text{lex}}$" on the variables of $\{y_{u_1}, \ldots, y_{u_m}\}$ in the following form:

$$y_u <^\sharp_{\text{lex}} y_v \Leftrightarrow u <_{\text{lex}} v \text{ and } y_u = y_v \Leftrightarrow u = v.$$

Also, consider the ordering $<_{\text{Lex}}$ induced by

$$x_{11} > \ldots > x_{1k_1} > \ldots > x_{n1} > \ldots > x_{nk_n}$$

on the monomials of $\mathcal{A}^\alpha$ for $\alpha = (k_1, \ldots, k_n)$. Again, let "$<^\sharp_{\text{Lex}}$" be a term order on the variables of $\{y_{u'} : u' \in \mathcal{A}^\alpha\}$ in the following form:

$$y_{u'} <^\sharp_{\text{Lex}} y_{v'} \Leftrightarrow u' <_{\text{Lex}} v' \text{ and } y_{u'} = y_{v'} \Leftrightarrow u' = v'.$$

Theorem 2.3. *[5] Let $\mathcal{A}$ be a set of monomials belonging to $S = K[x_1, \ldots, x_n]$ and let $\alpha = \mathbf{1} + \epsilon_i \in \mathbb{N}^n$. If $\mathcal{G}_{\mathcal{A}}$ is a Gröbner basis of $I_{\mathcal{A}}$ with respect to a term order induced by $<^\sharp_{\text{lex}}$ on $S_{\mathcal{A}}$, then the Gröbner basis of $I_{\mathcal{A}^\alpha}$, $\mathcal{G}_{\mathcal{A}^\alpha}$, with respect to the term order induced by $<^\sharp_{\text{Lex}}$ on $S_{\mathcal{A}^\alpha}$ is the union of the set*

$$\mathcal{G}_0 := \{y_{u'}y_{v'} - y_{(u'/x_{i1})x_{i2}}y_{(v'/x_{i2})x_{i1}} : \ u', v' \in \mathcal{A}^\alpha \text{ and } x_{i1}|u', x_{i2}|v'\}$$

and the set, call $\mathcal{G}_1$, containing all binomials $\prod_l^r y_{u'_l} - \prod_l^s y_{v'_l}$ with the property that $\prod_l^r y_{\pi(u'_l)} - \prod_l^s y_{\pi(v'_l)} \in \mathcal{G}_{\mathcal{A}}$ and $\prod_l^r u'_l = \prod_l^s v'_l$ for $u'_l, v'_l \in \mathcal{A}^\alpha$.

Corollary 2.4. *For a given $\alpha \in \mathbb{N}^n$ and a set $\mathcal{A}$ of monomials in S, the Gröbner basis of $I_{\mathcal{A}}$ is consists of quadratic binomials if and only if the Gröbner basis of $I_{\mathcal{A}^\alpha}$ has a such construction.*

Theorem 2.5. *Let $F = \{f_1, \ldots, f_m\} \subset K[X]$ be an arbitrary set of polynomials and set $d := \max_{i=1}^m(\deg(f_i))$. Then, for every admissible ordering $\preceq$, there exists a Gröbner basis G of F such that for every $g \in G$ there exist $q_1, \ldots, q_m \in K[X]$ with $g = \sum_{i=1}^m q_i f_i$ and $\deg(q_i f i) \geq Dube_{n+1,d}$ for all $1 \leq i \leq m$.*

References

[1] S. Bayati, J. Herzog, *Expansions of monomial ideals and multigraded modules*, To appear in Rocky Mountain J. Math.

[2] J. Herzog, T. Hibi, *Monomial ideals*, Graduate Texts in Mathematics **260**, Springer-Verlag. 2011.

[3] A. Maletzky, *Gröbner Bases and Macaulay Matrices in Isabelle/HOL*, `https://www3.risc.jku.at/publications/download/risc_5929/Paper.pdf`.

[4] H. Ohsugi, J. Herzog, T. Hibi, *Combinatorial pure subrings*, Osaka J. Math. **37** (2000), 745-757.

[5] R. Rahmati-Asghar, S. Yassemi, *The Behaviors of Expansion Functor on Monomial Ideals and Toric Rings*, Communications in Algebra, **44**: 3874-3889, 2016.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The distribution of Nadarajah and Kotz revisited

Hazhir Homei* and Manizheh Jalilvand

Department of Mathematics, University of Tabriz, Tabriz, Iran

Abstract

A new model for real lifetimes is proposed, here That can be used in topics such as vehicle speed, asphalt, etc. The distributional properties of this model are discussed, also the Nadarajah and Kotz distributions are generalized by using it.

Keywords: Real Lifetime, Product, Deferential Equation

Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36 (At least one and at most three codes)

1 Introduction

In statistics, the generalized distributions play a very important role in the lifetime. In this paper, we use the generalized distribution of the Nadarajah and Kotz (2005) if it is possible, otherwise, we suggest the approximation of Nadarajah (2006a,2006b) by using MLE. We will answer the questions posed at the conclusion of two separate articles about the lifetime in Homei and Nadarajah (2018) and Hadad et al. (2021) and generalize some of the results obtained by solving some differential equations.

2 The distribution of a real lifetime and some properties

The product of random variables has found many interesting applications theoretically. There are various examples of random variables in the literature that their products are analyzed theoretically and practically (see section 2 in Adamska et al. 2022) of which are being reviewed in this section and generalized to the multivariate cases; see Nadarajah and Kotz (2005).

Theorem 2.1. *Let the random vector* $\boldsymbol{X}$*, effective coefficient, and the random variable* Y*, lifetime in the laboratory, be with* $CE(\alpha_1,\ldots,\alpha_r)$ *and* $L(\alpha,\beta)$ *distributions. Then the distribution of the real lifetime,* $\boldsymbol{Z}=\boldsymbol{X}Y$*, can be expressed by:*

$$f(z_1,\ldots,z_r)=\frac{\beta^{-\alpha}\Gamma(\sum_{i=1}^r\alpha_i)}{\Gamma(\alpha)\prod_{i=1}^r\Gamma(\alpha_i)}(\sum_{i=1}^r z_i)^{\alpha-\sum_{i=1}^r\alpha_i}exp\{-\frac{\sum_{i=1}^r z_i}{\beta}\}\prod_{i=1}^r z_i^{\alpha_i-1}. \qquad (1)$$

Where $z_i>0,\ i=0,\cdots,r.$

*Speaker. Email address: homei@tabrizu.ac.ir

Proof. The distribution of $\mathbf{Z}$ can be found by using classical methods of transformations. Thus we will have

$$J = (\frac{1}{\sum_{i=1}^{r} z_i})^{r-1} \quad , \quad y = \sum_{i=1}^{r} z_i \quad , \quad x_i = \frac{z_i}{\sum_i z_i} \tag{2}$$

We also know that the probability density function $\mathbf{Z} = (z_1, \ldots, z_r)$ is as follows:

$$f_{\mathbf{Z}}(z_1, \ldots, z_r) = f_Y(\sum_{i=1}^{r} z_i) f_X(\frac{z_i}{\sum_1 z_i}, \ldots, \frac{z_r}{\sum_i z_i})|J| \tag{3}$$

$$= \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(\sum_{i=1}^{r} z_i)^{\alpha-1} e^{-\frac{\sum_{i=1}^{r} z_i}{\beta}} \frac{\Gamma(\sum_{i=1}^{r}\alpha)}{\prod_{i=1}^{r}\Gamma(\alpha_i)} \prod_{i=1}^{r}(\frac{z_i}{\sum_{i=1}^{r} z_i})^{\alpha_i-1}(\frac{1}{\sum_{i=1}^{r} z_i})^{r-1}, \tag{4}$$

the proof is completed. □

The Nadarajah and Kotz (2005) distribution will be obtained exactly for $r = 2$ and thus the distribution of Z_1 should be calculated (marginal distribution). Theorem 2.2 shows the distribution of the product of the random variables Gamma and Dirichlet random vector. Moreover, if $\alpha = \sum \alpha_i$ then the distribution of $\mathbf{Z}$ in (2) shows that $Z_1, \ldots, Z_k$ are independent with Gamma distribution.

Theorem 2.2. *If* $\mathbf{X} \sim Ga * D(\alpha, \beta, \alpha_1, \cdots, \alpha_n)$ *then we have*

$$E(Z_i) = \frac{\alpha_i}{\sum_{i=1}^{n}\alpha_i}\alpha\beta$$

and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Where Σ *denotes the covariance matrix of* Z_j*'s and*

$$\sigma_j = \frac{\beta^{\alpha}\alpha(\alpha+1)\alpha_j(\alpha_j+1)}{(\sum_{i=1}^{n}\alpha_i)(\sum_{i=1}^{n}\alpha_i+1)} - (\frac{\alpha_j\alpha\beta}{\sum_{i=1}^{n}\alpha_i})^2$$

$$\sigma_{jk} = \frac{\alpha\alpha_j\alpha_k}{(\alpha+\beta)^2(\sum_{i=1}^{n}\alpha_i)(\sum_{i=1}^{n}\alpha_i+1)}(\frac{\beta}{\alpha+\beta+1} - \frac{\alpha}{\sum_{i=1}^{n}\alpha_i}).$$

3 Lifetime in the independent r-position

By the most important properties of the distribution quoted in Nadarajah et al (2022), the following theorems are stated.

Theorem 3.1. *Let* $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_r$ *be the independent random vectors, effective coefficient, with* $CE(n_{11}, \cdots, n_{1k}), \cdots, CE(n_{r1}, \cdots, n_{rk})$ *distributions respectively and that the random variable* Y_j, $j = 1, 2, \cdots, r$, *lifetime in the laboratory, is independent from* $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r$ *with* $L(\alpha_j, 1)$, $j = 1, 2, \cdots, r$, *distribution. Then the product moments in* $(s_1, \ldots, s_k)$ *of the real lifetime in the independent r-position,* $\boldsymbol{T} = \sum_{j=1}^{r} Y_j \boldsymbol{X}_j$, *are*

$$E(T_1^{s_1}T_2^{s_2}\ldots T_k^{s_k}) = \frac{\Gamma(\alpha+s)}{\Gamma(\alpha+h)}\sum_{h_1}\cdots\sum_{h_k}\left(\prod_{j=1}^{k}\binom{s_j}{h_{1j}\ldots h_{rj}}\right) \times \prod_{i=1}^{r}\frac{\Gamma(\alpha_i+h_i)}{\Gamma(\alpha_i)}$$

$$\frac{\Gamma(n_i)}{\Gamma(n_i+h_i)}\prod_{i=1}^{r}\prod_{j=1}^{k}\frac{\Gamma(n_{ij}+h_{ij})}{\Gamma(n_{ij})}\Big),$$

where T_j's are the components of vector T, $\sum_{i=1}^{r} h_i = h$, $\sum_{i=1}^{r}\alpha_i = \alpha$ and $\sum_{i=1}^{r} s_i = s$.

Proof.

$$\begin{aligned}
E(T_1^{s_1}T_2^{s_2}\dots T_k^{s_k}) &= E\Big(\prod_{j=1}^{k}(\sum_{i=1}^{r} Y_j\mathbf{X}_{ij})^{s_j}\Big)\\
&= E\Big(\prod_{j=1}^{k}(\sum_{j=1}^{r} Y_j\frac{\sum_{i=1}^{r} Y_j}{\sum_{j=1}^{r} Y_j}\mathbf{X}_{ij})^{s_j}\Big)\\
&= E\Big(\prod_{j=1}^{k}(\sum_{j=1}^{r} Y_j\sum_{i=1}^{r}\frac{Y_j}{\sum_{j=1}^{r} Y_j}\mathbf{X}_{ij})^{s_j}\Big)\\
&= E\Big((\sum_{j=1}^{r} Y_j)^{\sum_{j=1}^{r} s_j}\Big)\Big(\prod_{j=1}^{k}(\sum_{i=1}^{r}\frac{Y_j}{\sum_{j=1}^{r} Y_j}\mathbf{X}_{ij})^{s_j}\Big)\\
&= E\Big((\sum_{j=1}^{r} Y_j)^{s}\Big)E\Big(\prod_{j=1}^{k}(\sum_{i=1}^{r}\frac{Y_j}{\sum_{j=1}^{r} Y_j}\mathbf{X}_{ij})^{s_j}\Big)
\end{aligned}$$

The previously mentioned mathematical expectations can be easily calculated separately.

$$E(L_1^{s_1}L_2^{s_2}\dots L_k^{s_k}) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+h)}\sum_{h_1}\dots\sum_{h_k}\Big(\prod_{j=1}^{k}\binom{s_j}{h_{1j}\dots h_{rj}}\times\prod_{i=1}^{r}\frac{\Gamma(\alpha_i+h_i)}{\Gamma(\alpha_i)}$$

where L_j's are components of vector $Z(=\sum_{i=1}^{n}\frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i)$

The result is obtained by placing the moments in the above expressions. □

The most important properties of any distribution are the first moment and variance.

Theorem 3.2. *Suppose $\boldsymbol{X}_1,\dots,\boldsymbol{X}_r$ are some independent effective coefficients with $CE(n_{11},\dots,n_{1k}),\dots,CE(n_{r1},\dots,n_{rk})$ distributions and $Y_1,\dots,Y_r$ are some independent lifetime in the laboratory with $L(n,\frac{1}{\theta})$ distributions and also independent from $\boldsymbol{X}_i$'s. Then the distribution of the real lifetime $\bar{T} = \frac{\sum_{i=1}^{r}\boldsymbol{X}_iY_i}{r}$ equals to*

$$f(z_1,\dots,z_r) = \frac{r^{rn-r+k}\theta^{rn}\Gamma(\sum_{i=1}^{r}\alpha_i)}{\Gamma(rn)\prod_{i=1}^{r}\Gamma(\alpha_i)}(\sum_{i=1}^{r} t_i)^{rn-\sum_{i=1}^{r}\alpha_i}e^{-\theta\sum_{i=1}^{r} rt_i}\prod_{i=1}^{r} t_i^{\alpha_i-1}. \qquad (5)$$

Where $t_i > 0, \sum_{j=1}^{k} n_{ij} = n, i = 1,\dots,r$ *and* $\sum_{j=1}^{r} n_{ji} = \alpha_i,\ j = 1,\dots,r.$

Proof. Here $\sum_{i=1}^{n} Y_i$ has a gamma distribution and the main theorem of Hadad et al. (2021) concludes that $\sum_{i=1}^{n} Y_i\mathbf{X}_i$ has a Dirichlet distribution. Of course, it is easy to prove that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n}\frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i$ are independent. Therefore, we can use the theorem **??** and obtain the desired distribution with a simple variable change.

$$\mathbf{T} = \sum_{i=1}^{n} Y_i\mathbf{X}_i = \sum_{i=1}^{n} Y_i.\sum_{i=1}^{n}\frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i.$$

□

The following theorem is another characterization on the real lifetime in Homei et al.(2022). It has been tried not to use Stieljes transformation for proof .

Table 1: Checking robustness of the approximation for r=2

(n_1, m_1)	(n_2, m_2)	(α_1, α_2)	p	q	$P-Value$
(1,1)	(1,1)	(1,1)	1.78	1.71	0.50
(1,1)	(1,1)	(0.2,0.3)	1.24	1.22	0.65
(2,4)	(3,5)	(0.2,0.3)	3.13	5.61	0.58
(2,4)	(3,5)	(1.75,2)	4.3	7.49	0.13
(3,7)	(5,9)	(0.2,0.3)	4.95	9.67	0.23
(3,7)	(5,9)	(1.75,2)	6.92	13.99	0.97
(4,10)	(7,13)	(0.2,0.3)	6.03	12.5	0.26
(4,10)	(7,13)	(1.75,2)	8.81	18.53	0.91
(5,13)	(9,17)	(0.2,0.3)	7.91	16.87	0.25
(5,13)	(9,17)	(1.75,2)	11.26	24.62	0.70

4 Approximation of Nadarajah by MLE

It is not easy to calculate the distribution of T and it requires a lengthy calculation to find the distribution of T. Inspired by the proof of the previous theorems, we suggest approximating the distribution of $\sum_{i=1}^{n} \frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i$ first and then obtaining an approximation of the distribution of T. In this section, we propose an approximation for the distribution of $\sum_{i=1}^{n} \frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i$. By Theorem 2 and the fact that the support of Z is Dirichlet distribution, we are interested in approximating its distribution by the Dirichlet family of distributions. The main idea of this approximation is taken from Nadarajah's works; see e.g. Nadarajah et al.(2013), Nadarajah (2006a) and Nadarajah (2006b). The idea of approximating distributions involving complicated formulas by the beta distribution are very well established in the statistics literature. The next procedure is based on the simulation and the Kolmogorov-Smirnov test and thus, the following steps are applied:
step 1 : Generate random numbers of $\sum_{i=1}^{n} \frac{Y_i}{\sum_{i=1}^{n} Y_i}\mathbf{X}_i$ by simulating the $\mathbf{X}_i's$ and $Y's$.
step 2 : Obtain MLE for unknown parameters.
step 3 : We do the forth step in Homei and Nadarajah (2018).

4.1 A comparison with the work of others

The distribution of Z can be approximated by the new method similar to Homei and Nadarajah (2018) and then compared with the method in Homei and Nadarajah (2018) showing to be better very close to it. Moreover, the result can be improved by increasing n. It is worth noting that the previous method is applicable to a univariate problem, but the presented method is applicable to the vectors. To evaluate the approximation of the distribution of Z, we used the p-values given in table 1, which show to be robust for most of the chosen parameters. We illustrate another example from Homei and Nadarajah (2018) graphically. Thus, let W, X_1, X_2 be independent with uniform $[0, 1]$ distribution. The density function of Z is $f_Z(z) = -2(log(1-z)^{(1-z)}z^z)I_{(0,1)}(z)$. By using the new method, $B(1.78, 1.71)$ will be a good approximation for Z.Of course, another approximation that was obtained for the distribution is $B(1.7, 1.7)$. Figure 1 shows the exact and approximate density of Z.

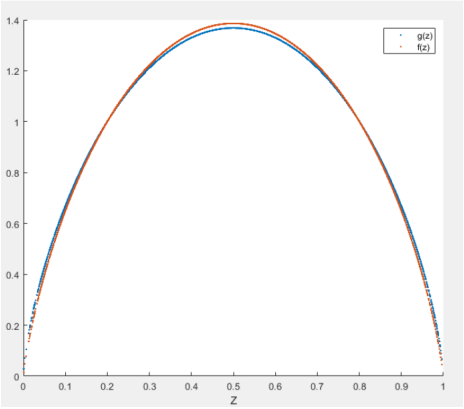


Figure 1: Exact and approximated distribution of Z when n=8000

5 Conclusions

By considering the environmental conditions, a model is introduced for analyzing a real lifetime in this paper and some of its distributional properties are discussed. Also, an approximation of the distribution is proposed if it is very complicated. In order to compare better the model with the work of Nadarajah, the beta distribution has been considered, but, using the quality of the approximation for Dirichlet distribution directly for evaluation is under investigation.

References

[1] Hadad, H., Homei, H., Behzadi, M., and Farnoosh, R. 2021. Solving some stochastic differential equation using Dirichlet distributions, *Computational Methods for Differential Equations* 9(2):393-398.

[2] Homei,H. 2021. The stochastic linear combination of dirichlet distributions,*Communications in Statistics: Theory and Methods,* 50(10):2354-2359.

[3] Homei,H., and Nadarajah, S. 2018. On products and mixed sums of Gamma and Beta random variables motivated by availability, *Methodology and Computing in Applied Probability* 20(2): 799 810.

[4] Homei, H., Nadarajah, S., and Taherkhani, A. 2022 Randomly weighted averages on multivariate dirichlet distributions with generalized parameters, REVSTAT Statistical Journal, Submited

[5] Nadarajah, S., and Kotz, S. 2005. On the Product and Ratio of Gamma and Beta random variables, *Allgemeines Statistisches Archiv* 89(4):435 49.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

On the stability of two-step Runge–Kutta methods

A. Mousavi[1,*], A. Abdi[1,2]

[1]Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran

[2]Research Department of Computational Algorithms and Mathematical Models, University of Tabriz, Tabriz, Iran

Abstract

Construction of efficient explicit two-step Runge–Kutta (TSRK) methods for ordinary differential equations is discussed. By obtaining the stability matrix of these methods, we present an overview of stability properties of such methods with the order p and stage order $q = p$. Some special conditions are then applied to obtain efficient methods with a large region of absolute stability.

Keywords: Two-step Runge–Kutta methods; Stability matrix; Stability region; Order conditions

Mathematics Subject Classification [2010]: 65L05

1 Introduction

Many efficient numerical methods have been proposed for the numerical solution of initial value problems (IVPs) of systems of ordinary differential equations (ODEs) in the form

$$\begin{cases} y'(x) = f(y(x)), & x \in [x_0, X], \\ y(x_0) = y_0, \end{cases} \tag{1}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ and m is the dimensionality of the system. In this paper, we are concerned with the class of explicit two-step Runge–Kutta (TSRK) methods, on the nonuniform grid $x_0 < x_1 < x_2 < \ldots < x_N, x_N \geq X$, defined by

$$\begin{cases} Y_i^{[n]} = u_i \bar{y}_{n-1} + (1-u_i) y_n + h_n \sum_{j=1}^{s} (a_{ij} f(\bar{Y}_j^{[n-1]}) + b_{ij} f(Y_j^{[n]})), \\ y_{n+1} = \eta \bar{y}_{n-1} + (1-\eta) y_n + h_n \sum_{j=1}^{s} (v_j f(\bar{Y}_j^{[n-1]}) + w_j f(Y_j^{[n]}), \end{cases} \tag{2}$$

for $n = 1, 2, \ldots . N-1$. In these formulae, $\eta \in \mathbb{R}$, $h_n = x_{n+1} - x_n$ is the stepsize and $y_n \approx y(x_n)$, $\bar{y}_{n-1} \approx y(x_n - h_n)$, $y_{n+1} \approx y(x_{n+1})$, $\bar{Y}_i^{[n-1]} \approx y(x_n + (c_i - 1)h_n)$, $Y_i^{[n]} \approx y(x_n + c_i h_n)$ are approximations to the exact solution where $c = [c_1 \quad c_2 \quad \cdots \quad c_s]^T$ is a given abscissa vector and

$$u = [u_1 \quad u_2 \quad \ldots \quad u_s]^T \in \mathbb{R}^s, \qquad A = [a_{ij}] \in \mathbb{R}^{s \times s}, \qquad B = [b_{ij}] \in \mathbb{R}^{s \times s},$$
$$v = [v_1 \quad v_2 \quad \ldots \quad v_s]^T \in \mathbb{R}^s, \quad w = [w_1 \quad w_2 \quad \ldots \quad w_s]^T \in \mathbb{R}^s,$$

*Speaker. Email address: a.moosavi1401@ms.tabrizu.ac.ir

are the coefficients of the method. These methods were introduced by Jackiewicz and Tracogna [1]. We can represent this class of the methods in the vector form

$$\begin{cases} Y^{[n]} = (u \otimes I_m)\bar{y}_{n-1} + ((e-u) \otimes I_m)y_n \\ \qquad +h_n((A \otimes I_m)f(\bar{Y}^{[n-1]}) + (B \otimes I_m)f(Y^{[n]})), \\ y_{n+1} = \eta\bar{y}_{n-1} + (1-\eta)y_n \\ \qquad +h_n((v^T \otimes I_m)f(\bar{Y}^{[n-1]}) + (w^T \otimes I_m)f(Y^{[n]})), \end{cases} \tag{3}$$

for $n = 1, 2, \ldots, N-1$, where $e = [1 \quad 1 \quad \ldots \quad 1]^T \in \mathbb{R}^s$, I_m is the identity matrix of dimension m, and

$$Y^{[n]} = \begin{bmatrix} Y_1^{[n]} \\ Y_2^{[n]} \\ \vdots \\ Y_s^{[n]} \end{bmatrix}, \quad f(Y^{[n]}) = \begin{bmatrix} f(Y_1^{[n]}) \\ f(Y_2^{[n]}) \\ \vdots \\ f(Y_s^{[n]}) \end{bmatrix}.$$

The class of the methods (3) can also be represented as general linear methods (GLMs) in the form

$$\left[\begin{array}{c} Y^{[n]} \\ \hline y_{n+1} \\ y_n \\ hf(\bar{Y}^{[n-1]}) \end{array}\right] = \left[\begin{array}{c|ccc} A & e-u & u & B \\ \hline v^T & 1-\eta & \eta & w^T \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} hf(\bar{Y}^{[n-1]}) \\ \hline y_n \\ \bar{y}_{n-1} \\ hf(Y^{[n]}) \end{array}\right],$$

This representation was first proposed by Butcher [2].

2 Order conditions for TSRK methods

The method (2) or (3) has order p and stage order $q = p$ when

$$\begin{aligned} y_{n+1} &= y(x+h) + O(h^{p+1}), \\ Y^{[n]} &= y(x+ch) + O(h^{p+1}), \end{aligned}$$

in which $y(x+ch)$ stands for

$$y(x+ch) := \begin{bmatrix} y(x+c_1h) \\ y(x+c_2h) \\ \vdots \\ y(x+c_sh) \end{bmatrix}.$$

We have the following theorem about the stage order and order conditions.

Theorem 2.1. *[3] The method (2) or (3) has the order p and stage order $q = p$ if and only if*

$$\begin{cases} A(c-e)^{j-1} + Bc^{j-1} = \frac{c^j - (-1)^j u}{j}, & j = 1, 2, \ldots, p, \\ v^T(c-e)^{j-1} + w^T c^{j-1} = \frac{1-(-1)^j\eta}{j}, & j = 1, 2, \ldots, p. \end{cases} \tag{4}$$

Moreover, the error constant C_p and the vector of error constants $\xi = [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_s]^T$ are given by

$$C_p = \frac{1}{(p+1)!} - \frac{v^T(c-e)^p + w^T c^p}{p!(1-(-1)^{p+1}\eta)}, \tag{5}$$

and

$$\xi = \frac{c^{p+1}}{(p+1)!} - (-1)^{p+1}\left(\frac{1}{(p+1)!} - E\right)u - \frac{A(c-e)^p}{p!} - \frac{Bc^p}{p!}. \tag{6}$$

Introducing the notations

$$C := \left[e \quad \tfrac{c}{1!} \quad \tfrac{c^2}{2!} \quad \cdots \quad \tfrac{c^p}{p!}\right], \quad \tilde{C} := \left[e \quad c-e \quad \tfrac{(c-e)^2}{2!} \quad \cdots \quad \tfrac{(c-e)^p}{p!}\right],$$
$$\hat{C} := \left[0 \quad \tfrac{c}{1!} \quad \tfrac{c^2}{2!} \quad \cdots \quad \tfrac{c^p}{p!}\right], \quad K := [\,0 \quad e_1 \quad e_2 \quad \cdots \quad e_p\,],$$
$$E_p := \left[0 \quad \tfrac{1}{1!} \quad \tfrac{1}{2!} \quad \cdots \quad \tfrac{1^p}{p!}\right], \quad \bar{E}_p := \left[0 \quad \tfrac{-1}{1!} \quad \tfrac{1}{2!} \quad \cdots \quad \tfrac{(-1)^p}{p!}\right],$$

conditions (4) are equivalent to

$$\begin{cases} A\tilde{C}K + BCK = \hat{C} - u\bar{E}_p, \\ v^T\tilde{C}K + w^TCK = E_p - \bar{E}_p\eta. \end{cases}$$

3 Linear Stability analysis

To obtain the stability matrix of TSRK methods in the fixed stepsize environment, we first apply the methods (3) to the standard test problems of Dahlquist

$$y' = \xi y,$$

where $\xi \in \mathbb{C}$ (possibly complex). This leads the stage values $Y^{[n]}$ and the output vector y_{n+1} to be in the form

$$\begin{cases} Y^{[n]} = (u \otimes I_m)\bar{y}_{n-1} + ((e-u) \otimes I_m)y_n \\ \qquad\quad +z((A \otimes I_m)\bar{Y}^{[n-1]} + (B \otimes I_m)Y^{[n]}), \\ y_{n+1} = \eta\bar{y}_{n-1} + (1-\eta)y_n \\ \qquad\quad +z((v^T \otimes I_m)\bar{Y}^{[n-1]} + (w^T \otimes I_m)Y^{[n]}), \end{cases} \tag{7}$$

where $z = \xi h$. Assume that the matrix $I - zB$ is nonsingular, then the *stability matrix* of TSRK methods takes the form

$$M(z) = \begin{bmatrix} 1-\eta+zw^TD(z)(e-u) & \eta+zw^TD(z)u & zw^TD(z)A+v^T \\ 1 & 0 & 0 \\ zD(z)(e-u) & zD(z)u & zD(z)A \end{bmatrix}, \tag{8}$$

where $D(z) = (I - zB)^{-1}$. The characteristic polynomial of $M(z)$, known as the *stability function* of the methods, is defined by

$$p(w, z) = \det(wI - M(z)),$$

where $w \in \mathbb{C}$. The *absolute stability region* S of the method (2) or (3) is the set of all $z \in \mathbb{C}$ such that

$$S = \{z \in \mathbb{C} : |w_i(z)| < 1, i = 1, 2, \ldots, s\}.$$

4 An example of TSRK method of order 3

Here, we construct a method of order $p = 3$ and stage order $q = p = 3$. In the construction of such methods, assuming $\eta = 0$ and $c = [0, 1/2, 1]$, and by solving the stage order and order conditions (4) and the relation (5), we obtain some free parameters; these free parameters are used in such a way to maximize the area of the region of absolute stability of the method for which **fminsearch** command from MATLAB is used. The coefficients of the constructed TSRK method with the error constant $C_3 = 1/12$ are as

$$B = \begin{bmatrix} 0 & 0 & 0 \\ 0.97846258 & 0 & 0 \\ 1.84458236 & 0.43911817 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.01511862 & 0.06047448 & 0.01511862 \\ 0.20333554 & -0.68665785 & -0.02512704 \\ 0.80855800 & -1.69194079 & 0.08573930 \end{bmatrix},$$

$$u = \begin{bmatrix} 0.09071173 \\ -0.02998677 \\ 0.48605705 \end{bmatrix}, \quad w = \begin{bmatrix} 0.60652150 \\ -1.75941934 \\ 0.21195023 \end{bmatrix}, \quad v = \begin{bmatrix} 1.59384545 \\ 0.24058066 \\ 0.10652150 \end{bmatrix}.$$

The stability region of this method is plotted in Fig.1. Also, to compare, we have plotted the stability region of explicit Runge–Kutta method (RK) of order 3 with the coefficients

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ 1 & -1 & 2 & \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

.

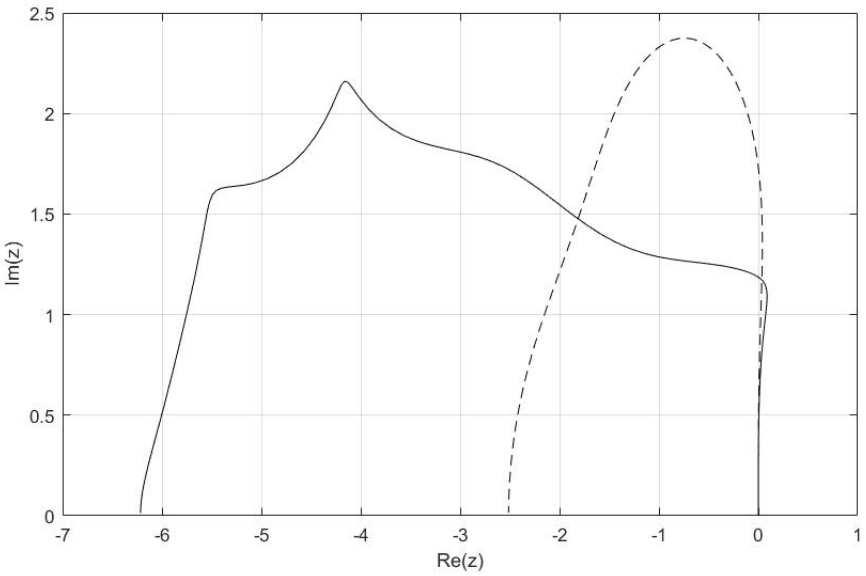


Figure 1: Stbility regions of TSRK method of order three with $C_3 = 1/12$ (solid line) and RK method of order three (dashed line).

5 Conclusion

We illustrated the construction of explicit TSRK methods of order p and stage order $q = p$. Using the free parameters in such methods, we derived a method of order three with a large region of absolute stability region that can solve IVPs with larger stepsize than explicit RK method of order $p = 3$.

References

[1] Z. Jackiewicz and S. Tracogna, A general class of two-step Runge-Kutta methods for ordinary differential equations, *SIAM J. Numer. Anal.* 32 (1995), 1390–1427.

[2] J.C. Butcher, On the convergence of numerical solutions to ordinary differential equations, *Math. Comp*, 20 (1966), 1–10.

[3] D. Conte, R. D'Ambrosio, and Z. Jackiewicz, Two-step Runge–Kutta methods with quadratic stability functions, *J. Sci. Comput.* 44 (2010), 191–218.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

The λ-mean transform of operators

Mohammad Mahdi Mansourian* and Ali Zamani

Department of Mathematics, Farhangian University, Tehran, Iran

Abstract

For $\lambda \in [0,1]$, we introduce the λ-mean transform $M_\lambda(T)$ of a Hilbert space operator T as an extension of some operator transforms based on the Duggal transform T^D by $M_\lambda(T) := \lambda T + (1-\lambda)T^D$, and present estimates for the operator norm and the numerical radius of $M_\lambda(T)$ in terms of the original operator T.

Keywords: Duggal transform, mean transform, numerical radius, inequality
Mathematics Subject Classification [2010]: 47A05, 47B49, 47A12

1 Introduction

Let $\mathbb{B}(\mathcal{H})$ denote the C^*-algebra of all bounded linear operators on a complex Hilbert space $\big(\mathcal{H}, \langle\cdot,\cdot\rangle\big)$. For every $T \in \mathbb{B}(\mathcal{H})$ its spectral radius is denoted by $r(T)$, and its numerical radius by $w(T)$. It is well known that for all $T \in \mathbb{B}(\mathcal{H})$,

$$\max\big\{r(T), \frac{1}{2}\|T\|\big\} \leq w(T) \leq \|T\|.$$

For an operator $T \in \mathbb{B}(\mathcal{H})$, there exists a unique polar decomposition $T = U|T|$ (called the canonical polar decomposition), where $|T| = (T^*T)^{1/2}$ and U is the appropriate partial isometry satisfying $\ker(U) = \ker(T)$. Let us introduce some transforms of Hilbert space operators. Let $T = U|T|$ be the canonical polar decomposition of $T \in \mathbb{B}(\mathcal{H})$. The Aluthge transform $\widetilde{T}$ of T is defined by $\widetilde{T} := |T|^{1/2}U|T|^{1/2}$. The Duggal transform T^D of T is defined by $T^D := |T|U$. The mean transform $\widehat{T}$ of T is defined by $\widehat{T} := \frac{1}{2}(T + T^D)$. This transform was first introduced in [4] and has received much attention in recent years. A kind of operator transform is the generalized mean transform $\widehat{T}(t)$ of T, introduced recently in [1], by

$$\widehat{T}(t) := \frac{1}{2}(|T|^t U|T|^{1-t} + |T|^{1-t}U|T|^t),$$

for $t \in [0,1/2]$. Clearly, $\widehat{T}(0) = \widehat{T}$ and $\widehat{T}(1/2) = \widetilde{T}$.

For more information about the transforms and their properties, interested readers are referred to [1–5].

Now, we introduce (see [5]) a new transform of the given operator $T \in \mathbb{B}(\mathcal{H})$ based on the Duggal transform T^D.

*Mohammad Mahdi Mansourian. Email address:mmahdimans@gmail.com

Definition 1.1. Let $T = U|T|$ be the canonical polar decomposition of $T \in \mathbb{B}(\mathcal{H})$. For $\lambda \in [0,1]$, the λ-mean transform $M_\lambda(T)$ of T is defined by

$$M_\lambda(T) := \lambda T + (1-\lambda)T^D,$$

where $T^D = |T|U$ denotes the Duggal transform of T. In particular, $M_0(T) = T^D$ and $M_{1/2}(T) = \widehat{T}$ is the mean transform of T.

For $t \in (0, 1/2)$, it is so hard to find the generalized mean transform $\widehat{T}(t)$ of the given operator $T \in \mathbb{B}(\mathcal{H})$ because it involves $|T|^t$, and it is quite difficult to find $|T|^t$ in general. By contrast, for $\lambda \in [0,1]$, the λ-mean transform $M_\lambda(T)$ of T involves T^D, so it is easy to get $M_\lambda(T)$ if we know the canonical polar decomposition of T. Hence the λ-mean transform $M_\lambda(T)$ is more convenient than the generalized mean transform $\widehat{T}(t)$ in the practical use.

In the next section we present estimates for the operator norm and the numerical radius of $M_\lambda(T)$ in terms of the original operator T.

2 Main results

Our first result reads as follows.

Theorem 2.1. *[5, Theorem 3.2] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in [0,1]$. Then*

$$2\sqrt{\lambda - \lambda^2}\,\|\widetilde{T}\| \le \big\|M_\lambda(T)\big\| \le \lambda\|T\| + (1-\lambda)\|T^D\|. \tag{1}$$

In particular,

$$2\sqrt{\lambda - \lambda^2}\,r(T) \le \big\|M_\lambda(T)\big\| \le \|T\|.$$

In the following theorem we give a necessary and sufficient condition for the equality $\big\|M_\lambda(T)\big\| = \|T\|$.

Theorem 2.2. *[5, Theorem 3.9] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in (0,1)$. Then the following statements are equivalent:*

(i) $\big\|M_\lambda(T)\big\| = \|T\|$.

(ii) There exists a sequence of unit vectors $\{x_n\}$ in $\mathcal{H}$ such that

$$\lim_{n\to+\infty} \langle Tx_n, T^D x_n\rangle = \|T\|^2.$$

In the following theorem we present an improvement of the second inequality in (1).

Theorem 2.3. *[5, Theorem 3.5] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in [0,1]$. Then*

$$\big\|M_\lambda(T)\big\| \le \frac{\big\|\lambda|T| + (1-\lambda)|T^D|\big\| + \big\|\lambda|T^*| + (1-\lambda)|(T^D)^*|\big\|}{2} \le \lambda\|T\| + (1-\lambda)\|T^D\|.$$

In particular,

$$\|\widehat{T}\| \le \frac{\big\|\,|T| + |T^D|\big\| + \big\|\,|T^*| + |(T^D)^*|\big\|}{4} \le \|T\|. \tag{2}$$

Example 2.4. [5, Example 3.6] Let $T = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$. Simple computations show that

$$\|\widehat{T}\| \simeq 1.1180 < \frac{\big\| |T| + |T^D| \big\| + \big\| |T^*| + |(T^D)^*| \big\|}{4} \simeq 1.1218 < \frac{\|T\| + \|T^D\|}{2} \simeq 1.2071 < \|T\| \simeq 1.4142.$$

Therefore, the inequality (2) is a nontrivial improvement.

Now, we state several the numerical radius inequalities for the λ-mean transform of Hilbert space operators.

Theorem 2.5. *[5, Theorem 4.1] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in [0,1]$. Then*

$$2\sqrt{\lambda - \lambda^2}\,\omega(\widetilde{T}) \leq w\big(M_\lambda(T)\big) \leq \lambda w(T) + (1-\lambda) w(T^D). \tag{3}$$

In particular,

$$2\sqrt{\lambda - \lambda^2}\, r(T) \leq w\big(M_\lambda(T)\big) \leq w(T).$$

In the following result we present an improvement of the second inequality in (3).

Theorem 2.6. *[5, Theorem 4.6] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in [0,1]$. Then*

$$w\big(M_\lambda(T)\big) \leq 2\int_0^1 w\Big(\lambda t T + (1-\lambda)(1-t)T^D\Big)dt \leq \frac{1}{2}\Big(\lambda w(T) + (1-\lambda) w(T^D) + w\big(M_\lambda(T)\big)\Big).$$

In particular,

$$w(\widehat{T}) \leq \int_0^1 w\big(M_t(T)\big)dt \leq \frac{1}{4}\Big(w(T) + w(T^D) + 2w(\widehat{T})\Big) \leq w(T).$$

Example 2.7. [5, Remark 4.7] Let $T = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$. It is easy to see that $w(T) = \frac{1+\sqrt{2}}{2}, w(T^D) = 1, w(\widehat{T}) = \frac{2+\sqrt{5}}{4}$, and

$$\int_0^1 w\big(M_t(T)\big)dt = \int_0^1 \frac{1+\sqrt{1+t^2}}{2}dt \simeq 1.0739.$$

Thus

$$\begin{aligned} w(\widehat{T}) \simeq 1.0590 &< \int_0^1 w\big(M_t(T)\big)dt \simeq 1.0739 \\ &< \frac{1}{4}\Big(w(T) + w(T^D) + 2w(\widehat{T})\Big) \simeq 1.0812 \\ &< \frac{1}{2}\big(w(T) + w(T^D)\big) \simeq 1.1035 < w(T) \simeq 1.2071. \end{aligned}$$

Therefore, the inequalities in Theorem 2.6 are nontrivial improvements.

Finally, we present another improvement of the second inequality in (3).

Theorem 2.8. *[5, Theorem 4.8] Let $T \in \mathbb{B}(\mathcal{H})$ and let $\lambda \in [0,1]$. Then*

$$\begin{aligned} w\big(M_\lambda(T)\big) &\leq \frac{1}{2}\big(\lambda w(T) + (1-\lambda) w(T^D)\big) \\ &\quad + \frac{1}{2}\sqrt{\big(\lambda w(T) - (1-\lambda) w(T^D)\big)^2 + 4(\lambda - \lambda^2) \sup_{\theta \in \mathbb{R}} \big\| \mathfrak{Re}(e^{i\theta}T)\mathfrak{Re}(e^{i\theta}T^D) \big\|} \\ &\leq \lambda w(T) + (1-\lambda) w(T^D). \end{aligned}$$

In particular,

$$w(\widehat{T}) \leq \frac{w(T) + w(T^D) + \sqrt{\big(w(T) - w(T^D)\big)^2 + 4 \sup_{\theta \in \mathbb{R}} \big\| \mathfrak{Re}(e^{i\theta}T)\mathfrak{Re}(e^{i\theta}T^D) \big\|}}{4} \leq w(T).$$

Example 2.9. [5, Remark 4.9] The inequality obtained by us in Theorem 2.8 is a non-trivial improvement. Consider $T = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$. It is easy to check that

$$\sup_{\theta\in\mathbb{R}} \left\|\mathfrak{Re}(e^{i\theta}T)\mathfrak{Re}(e^{i\theta}T^D)\right\| = \frac{\sqrt{5}}{2}.$$

Therefore,

$$\begin{aligned} w(\widehat{T}) \simeq 1.0590 \leq \frac{w(T) + w(T^D) + \sqrt{\left(w(T) - w(T^D)\right)^2 + 4\sup_{\theta\in\mathbb{R}} \left\|\mathfrak{Re}(e^{i\theta}T)\mathfrak{Re}(e^{i\theta}T^D)\right\|}}{4} &\simeq 1.0829 \\ < \frac{1}{2}\left(w(T) + w(T^D)\right) \simeq 1.1035 < w(T) &\simeq 1.2071. \end{aligned}$$

References

[1] C. Benhida, M. Chō, E. Ko and J. E. Lee, *On the generalized mean transforms of complex symmetric operators*, Banach J. Math. Anal. **14** (2020), 842–855.

[2] F. Chabbabi, R. Curto and M. Mbekhta, *The mean transform and mean limit of an operator*, Proc. Amer. Math. Soc. **147** (2019), no. 3, 1119–1133.

[3] C. Foias, I. Jung, E. Ko and C. Pearcy, *Complete contractivity of maps associated with the Aluthge and Duggal transformations*, Pacific J. Math. **209** (2003), 249–259.

[4] S. Lee, W. Lee and J. Yoon, *The mean transform of bounded linear operators*, J. Math. Anal. Appl. **410** (2014), 70–81.

[5] A. Zamani, *On an extension of operator transforms*, J. Math. Anal. Appl. **493** (2021) 124546.

دانشگاه صنعتی سهند، تبریز، ایران
دوازدهمین سمینار بین المللی
جبر خطی
و کاربردهای آن
۲۷ الی ۲۸ تیرماه ۱۴۰۲
Sahand University of Technology, Tabriz, Iran
The 12th International Seminar on
Linear Algebra
and its Applications
July 18 - 19, 2023

Some properties of fuzzy inner product spaces

V. Ebrahimi* and B. Daraby

Department of Mathematics, University of Maragheh, Iran

Abstract

In this paper, we present some properties of the space $(B(U,V), \|.\|_\alpha)$ and some consequences of fuzzy linear spaces analogous to the ordinary normed spaces. .

Keywords: Fuzzy norm, Fuzzy inner product space, Fuzzy Hilbert space, Fuzzy frame
Mathematics Subject Classification [2010]: 15A03, 15A23, 15B36 (At least one and at most three codes)

1 Introduction

The idea of fuzzy norms on a linear space first introduced by Katsaras [6] in 1984. Later on, many authors Felbin [7], Cheng, Mordeson [2], Bag and Samanta [1] etc. gave different definitions of fuzzy normed linear spaces. R. Biswas and A. M. El-Abye and H. M. El-Hamouly tried to give a meaningful definition of fuzzy inner product space and associated fuzzy norm function with those definition are restricted to the real linear space only.. Recently, B. Daraby and et al. [4] studied some properties of fuzzy Hilbert spaces and they showed that all results in classical Hilbert spaces are immediate consequences of the corresponding results for Felbin-fuzzy Hilbert spaces. Also by an example, they showed that the spectrum of the category of Felbin- fuzzy Hilbert spaces is broader than the category of classical Hilbert spaces [7].

2 Some preliminaries

In this section, some definitions and preliminary results are given which will be used in this paper.

Definition 2.1. [1] Let U be a linear space over the field $\mathbb{C}$ of complex numbers. Let $\mu : U \times U \times \mathbb{C} \longrightarrow I = [0,1]$ be a mapping such that the following hold:

(FIP1) for $s,t \in \mathbb{C}, \mu(x+y,z,|t|+|s|) \geq \min\{\mu(x,z,|t|), \mu(y,z,|s|)\}$;

(FIP2) for $s,t \in \mathbb{C}, \mu(x,y,|st|) \geq \min\{\mu(x,x,|s|^2), \mu(y,y,|t|^2)\}$;

(FIP3) for $t \in \mathbb{C}, \mu(x,y,t) = \mu(y,x,\bar{t})$;

*Speaker. Email address:m.ebrahimi@stu.maragheh.ac.ir

(FIP4) $\mu(\alpha x, y, t) = \mu(x, y, \frac{t}{|\alpha|}), \alpha (\neq 0) \in \mathbb{C}, t \in \mathbb{C}$;

(FIP5) $\mu(x, x, t) = 0, \forall t \in \mathbb{C} \backslash \mathbb{R}^+$;

(FIP6) $(\mu(x, x, t) = 1, \forall t > 0)$ iff $x = \underline{0}$;

(FIP7) $\mu(x, x, .) : \mathbb{R} \longrightarrow I$ is a monotonic non-decreasing function on $\mathbb{R}$ and $\lim_{t\to\infty} \mu(\alpha x, x, t) = 1$.

We call μ fuzzy inner product function on U and (U, μ) fuzzy inner product space (FIP space).

Theorem 2.2. *[1] Let U be a linear space over $\mathbb{C}$. Let μ be a FIP on U. Then*

$$N(x,t) = \begin{cases} \mu(x, x, t^2) & \text{if } t \in \mathbb{R}, t > 0, \\ 0 & \text{if } t \leq 0. \end{cases}$$

is a fuzzy norm on U. Now if μ satisfies the following conditions:

(FIP8) $\left(\mu\left(x, x, t^2\right) > 0, \forall t > 0\right) \Rightarrow x = \underline{0}$ *and*

(FIP9) for all $x, y \in U$ and $p, q \in \mathbb{R}$,

$$\mu\left(x + y, x + y, 2q^2\right) \bigwedge \mu\left(x - y, x - y, 2p^2\right) \geq \mu\left(x, x, p^2\right) \bigwedge \mu\left(y, y, q^2\right),$$

then $\|x\|_\alpha = \bigwedge \{t > 0 : N(x, t) \geq \alpha\}, \alpha \in (0, 1)$ is an ordinary norm satisfying parallelogram law. By using polarization identity, we can get ordinary inner product, called the $\langle ., . \rangle_\alpha$-inner product, as follows:

$$\langle x, y \rangle_\alpha = \frac{1}{4}\left(\|x + y\|_\alpha^2 - \|x - y\|_\alpha^2\right) + \frac{1}{4} i \left(\|x + iy\|_\alpha^2 - \|x - iy\|_\alpha^2\right), \forall \alpha \in (0, 1).$$

Theorem 2.3. *[1] Let $T : (U, N_1) \longrightarrow (V, N_2)$ be a linear operator where (U, N_1) and (V, N_2) are fuzzy normed linear spaces satisfying (N_6). Then T is strongly fuzzy bounded if and only if it is uniformly bounded with respect to α-norms of N_1 to N_2.*

3 Main results

Definition 3.1. Let (U, μ) and (V, μ) be two fuzzy Hilbert spaces satisfying $(FIP8)$ and $(FIP9)$ where μ is the same fuzzy inner product. Let T be a strongly fuzzy bounded linear operator from U to V. If there exists an operator T^* from V to U such that for all $\alpha \in (0, 1)$

$$\langle Tx, y \rangle_\alpha = \langle x, T^* y \rangle_\alpha, \quad \forall x \in U, y \in V,$$

then the operator T^* is called fuzzy adjoint of T.

In the following example , we give that the fuzzy inner product , results the classic inner product.

Example 3.2. Let $(U, \langle ., . \rangle)$ be a real inner product space. Define a function $\mu : U \times U \times \mathbb{C} \to [0, 1]$ by

$$\mu(x, y, t) = \begin{cases} \dfrac{|t|}{|t| + \|x\|\|y\|} & \text{if } t > \|x\|\|y\|, \\ 0 & \text{if } t \leq \|x\|\|y\|, \\ 0 & \text{if } t \in \mathbb{C} \setminus \mathbb{R}^+. \end{cases}$$

It is clear that (FIP8), (FIP9) holds. Using polarization identity, the α-inner product follows from classic inner product.

$$\|x - y\|_\alpha^2 + \|x + y\|_\alpha^2 = 2(\|x\|_\alpha^2 + \|y\|_\alpha^2).$$

It follows that

$$\begin{aligned} \langle x, y \rangle_\alpha &= \frac{1}{4}(\|x + y\|_\alpha^2 - \|x - y\|_\alpha^2) + \frac{i}{4}(\|x + iy\|_\alpha^2 - \|x - iy\|_\alpha^2) \\ &= \frac{\alpha}{4(1-\alpha)}(\|x + y\|^2 - \|x - y\|^2) + \frac{\alpha i}{4(1-\alpha)}(\|x + iy\|^2 - \|x - iy\|^2) \\ &= \frac{\alpha}{1-\alpha}\langle x, y \rangle. \end{aligned}$$

The example show that the fuzzy inner product , results the classic inner priduct.

Definition 3.3. [1] Let (U, μ) be a FIP space satisfying $(FIP8)$ and $(FIP9)$. Now if $x, y \in U$ be such that $\langle x, y \rangle_\alpha = 0$, for all $\alpha \in (0, 1)$, then we say that x, y are fuzzy orthogonal to each other and is denoted by $x \perp y$.
Thus $x \perp y$ if and only if $x \perp_\alpha y$, for all $\alpha \in (0, 1)$. The set of all fuzzy orthogonal elements to each other is called fuzzy orthogonal set.

Theorem 3.4. *[3] Let (U, μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$, $\alpha \in (0, 1)$ and $\{e_k\}_{k=1}^\infty$ be an α-fuzzy orthonormal sequence in U. If the series $\sum_{k=1}^\infty \gamma_k e_k$ converges w.r.t. N induced by μ, then*

$$\gamma_k = \langle x, e_k \rangle_\alpha = \langle x, e_k \rangle_\beta, \quad \forall \alpha, \beta \in (0, 1),$$

where $\langle ., . \rangle$ denotes the α-inner product induced by μ, x denotes the sum of $\sum_{k=1}^\infty \gamma_k e_k$. Hence

$$x = \sum_{k=1}^\infty \langle x, e_k \rangle_\alpha e_k = \sum_{k=1}^\infty \langle x, e_k \rangle_\beta e_k, \quad \forall \alpha, \beta \in (0, 1).$$

Theorem 3.5. *[1] Let (U, μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$ and $\{e_k\}_{k=1}^\infty$ is fuzzy orthonormal sequence in U. Then the following statements are equivalent:*

(i) $\{e_k\}_{k=1}^\infty$ is complete fuzzy orthonormal;

(ii) if $x \perp e_i$ for $i = 1, 2, \ldots$ then $x = \underline{0}$;

(iii) For every $x \in U, x = \sum_{k=1}^\infty \langle x, e_i \rangle_\alpha e_i$ for all $\alpha \in (0, 1)$ and hence

$$\langle x, e_k \rangle_\alpha = \langle x, e_k \rangle_\beta, \quad \forall \alpha, \beta \in (0, 1);$$

i.e. x is independent on α.

(iv) For every $x \in U$, $\|x\|_\alpha^2 = \sum_{k=1}^\infty |\langle x, e_i \rangle_\alpha|$ for all $\alpha \in (0, 1)$ and hence

$$\|x\|_\alpha^2 = \|x\|_\beta^2, \quad \forall \alpha, \beta \in (0, 1).$$

Proposition 3.6. *Let (U,μ) be a FIP space satisfying $(FIP8)$ and $(FIP9)$. A fuzzy inner product space (U,μ) with its corresponding norm N satisfies the Schwartz inequality*

$$|\langle x,y\rangle_\alpha| \leq \|x\|_\alpha \|y\|_\alpha \quad \forall \alpha \in (0,1].$$

Proof. At the first, we show that for all $\alpha \in (0,1)$, $\langle x,x\rangle_\alpha = \|x\|_\alpha^2$.
According to the definition of α-fuzzy inner product by supposing $x=y$ we have:

$$\begin{aligned}\langle x,x\rangle_\alpha &= \frac{1}{4}(\|x+x\|_\alpha^2 - \|x-x\|_\alpha^2) + \frac{i}{4}(\|x+ix\|_\alpha^2 - \|x-ix\|_\alpha^2)\\ &= \frac{1}{4}(4\|x\|_\alpha^2 - 0) + \frac{i}{4}x(\|1+i\|_\alpha^2 - \|1-i\|_\alpha^2)\\ &= \|x\|_\alpha^2.\end{aligned}$$

Therefore $\langle x,x\rangle_\alpha = \|x\|_\alpha^2$. Let $x,y \in U$ be arbitrary, in the special case where $y=0$, the corollary is trivially true. Now assume that $y \neq 0$. Considering $\lambda \in \mathbb{C}$ and by $\lambda = \frac{\langle x,y\rangle_\alpha}{\|y\|_\alpha^2}$ for all $\alpha \in (0,1)$, we have:

$$\begin{aligned}0 &\leq \|x-\lambda y\|_\alpha^2\\ &= \langle x,x\rangle_\alpha - \langle \lambda y,x\rangle_\alpha - \langle x,\lambda y\rangle_\alpha + \langle \lambda y,\lambda y\rangle_\alpha\\ &= \langle x,x\rangle_\alpha - \lambda\langle y,x\rangle_\alpha - \overline{\lambda}\langle x,y\rangle_\alpha + \lambda\overline{\lambda}\langle y,y\rangle_\alpha\\ &= \|x\|_\alpha^2 - \lambda\overline{\langle x,y\rangle_\alpha} - \overline{\lambda}\langle x,y\rangle_\alpha + \lambda\overline{\lambda}\|y\|_\alpha^2\\ &= \|x\|_\alpha^2 - \frac{|\langle x,y\rangle_\alpha|^2}{\|y\|_\alpha^2} - \frac{|\langle x,y\rangle_\alpha|^2}{\|y\|_\alpha^2} + \frac{|\langle x,y\rangle_\alpha|^2}{\|y\|_\alpha^2}\\ &= \|x\|_\alpha^2 - \frac{|\langle x,y\rangle_\alpha|^2}{\|y\|_\alpha^2}.\end{aligned}$$

Therefore

$$0 \leq \|x\|_\alpha^2 - \frac{|\langle x,y\rangle_\alpha|^2}{\|y\|_\alpha^2},$$

It follows that $|\langle x,y\rangle_\alpha| \leq \|x\|_\alpha\|y\|_\alpha$. □

Theorem 3.7. *If $T : (U,N_1) \longrightarrow (V,N_2)$ is a strongly fuzzy bounded operator, where (U,N_1) and (V,N_2) are fuzzy normed linear spaces that N_1 and N_2 induce from fuzzy inner products on U and V respectively, then there exists $T^* : (V,N_2) \longrightarrow (U,N_1)$ such that for all $x \in U, y \in V$ and for all $\alpha \in (0,1)$*

$$\langle x,T(y)\rangle_\alpha = \langle T^*(x),y\rangle_\alpha . \tag{1}$$

Proof. For existing of T^*, we have to show that for every $x \in U$, there is a vector $z \in U$, depending linearly on x, such that

$$\langle z,y\rangle_\alpha = \langle T^*(x),y\rangle_\alpha \quad \forall \alpha \in (0,1).$$

By Theorem 2.3, T is uniformly bounded and there exists $M>0$ such that

$$\|T(x)\|_\alpha^2 \leq M\|x\|_\alpha^1 \quad \forall \alpha \in (0,1).$$

Suppose that $\alpha \in (0,1)$, for fixed x, consider the mapping φ_x, defined by

$$\varphi_x(y) = \langle x,T(y)\rangle_\alpha .$$

The mapping φ_x is a fuzzy bounded linear functional on U crossponding with α i.e. $\varphi_x \in U_\alpha^*$ and $\|\varphi_x\|_\alpha \leq M\|x\|_\alpha$. By the Riesz Representation Theorem, there is a unique $z \in U$ such that $\varphi_x(y) = \langle z, y\rangle_\alpha$.Thus, the Equality (1) holds. So, we set $T^*(x) = z$. The linearity of T^* follows from its uniqueness by Riesz Representation Theorem and from the linearity of the inner product. Since we have

$$\begin{aligned}
\|T^*(x)\|_\alpha = \|z\| &= \bigvee_{\|y\|_\alpha=1} |\langle y, z\rangle_\alpha| \\
&= \bigvee_{\|y\|_\alpha=1} |\langle T(y), x\rangle_\alpha| \\
&\leq \bigvee_{\|y\|_\alpha=1} \|T(y)\|_\alpha \|x\|_\alpha \\
&\leq \bigvee_{\|y\|_\alpha=1} \|T\|_\alpha \|y\|_\alpha \|x\|_\alpha = \|T\|_\alpha \|x\|_\alpha,
\end{aligned}$$

It follows that T^* is bounded and $\|T^*\|_\alpha \leq \|T\|_\alpha$ for any $\alpha \in (0,1)$. Finally, we show that T^* is unique. Suppose that $S \in B(U,V)$ and $\langle T(x), y\rangle_\alpha = \langle S(x), y\rangle_\alpha$ for all $x \in U$, $y \in V$ and $\alpha \in (0,1)$. For each fixed y and for every x, we have $\langle x, S(y) - T^*(y)\rangle_\alpha = 0$. It follows that $S(y) - T^*(y) = 0$. Hence $S = T^*$. □

Proposition 3.8. *If $T \in B(U,V)$, then for all $\alpha \in (0,1)$, $\|T^*\|_\alpha = \|T\|_\alpha$.*

Proof. In the Theorem 3.7, we already showed that

$$\|T^*\|_\alpha \leq \|T\|_\alpha \quad \forall \alpha \in (0,1). \tag{2}$$

For $x \in U$, we have

$$\begin{aligned}
\|T(x)\|_\alpha^2 &= \langle T(x), T(x)\rangle_\alpha \\
&= \langle T^*T(x), x\rangle_\alpha \\
&\leq \|T^*T(x)\|_\alpha \|x\|_\alpha \\
&\leq \|T^*\|_\alpha \|T(x)\|_\alpha \|x\|_\alpha.
\end{aligned}$$

Hence $\|T(x)\|_\alpha \leq \|T^*\|_\alpha \, \|x\|_\alpha$. It follows that

$$\|T\|_\alpha \leq \|T^*\|_\alpha \quad \forall \alpha \in (0,1). \tag{3}$$

From the inequalites (2) and (3) we have

$$\|T\|_\alpha = \|T^*\|_\alpha \quad \forall \alpha \in (0,1).$$

□

Theorem 3.9. *Let (U,μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$ and $\alpha \in (0,1)$. Let T be a fuzzy operator on (U,μ). Then T^* is also a fuzzy linear operator on (U,μ) and following properties hold:*

i) $(T^*)^* = T$;

ii) $(T_1 + T_2)^* = T_1^* + T_2^*$;

iii) $(\lambda T)^* = \overline{\lambda} T^*, \quad \forall \lambda \in \mathbb{C}$;

iv) $(ST)^* = T^* S^*$.

Proof. Suppose that $y_1, y_2 \in U$ and $\lambda, \beta \in \mathbb{C}$. For each $x \in U$, we have

$$\begin{aligned} \langle x, T^*(\lambda y_1 + \beta y_2)\rangle_\alpha &= \langle Tx, \lambda y_1 + \beta y_2\rangle_\alpha \\ &= \overline{\lambda}\langle Tx, y_1\rangle_\alpha + \overline{\beta}\langle Tx, y_2\rangle_\alpha \\ &= \langle x, \lambda T^* y_1 + \beta T^* y_2\rangle_\alpha. \end{aligned}$$

It follows that $T^*(\lambda y_1 + \beta y_2) = \lambda T^* y_1 + \beta T^* y_2$, that is, T^* is linear.
For each $x, y \in U$,

$$\langle y, (T^*)^* x\rangle_\alpha = \langle T^* y, x\rangle_\alpha = \overline{\langle x, T^* y\rangle_\alpha} = \overline{\langle Tx, y\rangle_\alpha} = \langle y, Tx\rangle_\alpha.$$

Hence $(T^*)^* = T$, so we have (i).
For proving (ii), obviously we have

$$\begin{aligned} \langle x, (T_1 + T_2)^* y\rangle_\alpha &= \langle (T_1 + T_2)x, y\rangle_\alpha \\ &= \langle T_1 x, y\rangle_\alpha + \langle T_2 x, y\rangle_\alpha \\ &= \langle x, T_1^* y\rangle_\alpha + \langle x, T_2^* y\rangle_\alpha \\ &= \langle x, (T_1^* + T_2^*) y\rangle_\alpha. \end{aligned}$$

(iii) For each $\alpha \in (0, 1]$ and $\lambda \in \mathbb{C}$, we have

$\langle \lambda Tx, y\rangle_\alpha = \lambda\langle Tx, y\rangle_\alpha = \lambda\langle x, T^* y\rangle_\alpha = \langle x, \overline{\lambda} T^* y\rangle_\alpha$, so we get (iii).

For each $x, y \in U$,

$$\langle STx, y\rangle_\alpha = \langle Tx, S^* y\rangle_\alpha = \langle x, T^* S^* y\rangle_\alpha.$$

Therefore $(ST)^* = T^* S^*$. □

Corollary 3.10. *Let (U, μ) be a fuzzy Hilbert space satisfying $(FIP8)$ and $(FIP9)$ and $\alpha \in (0, 1)$. Let T be a fuzzy operator on (U, μ). Then*

$$\|T^* T\|_\alpha = \|TT^*\|_\alpha = \|T\|_\alpha^2.$$

Proof. Using by Theorem (3.9), proof is straightforward. For all $x \in U$,

$$\|T^* Tx\|_\alpha \leq \|T^*\|_\alpha \|Tx\|_\alpha \leq \|T\|_\alpha^2 \|x\|_\alpha$$

and therefore $\|T^* T\|_\alpha \leq \|T\|_\alpha^2$.
Also, we can write

$$\begin{aligned} \|T^* T\|_\alpha &= \bigvee_{\beta \leq \alpha} \|T^* Tx\|'_\beta \\ &= \bigvee_{\beta \leq \alpha} \left(\bigvee_{x \in U, x \neq 0} \frac{\|T^* Tx\|_\beta^2}{\|x\|_\beta^1} \right) \\ &\geqslant \bigvee_{\beta \leq \alpha} \left(\bigwedge_{x \in U, x \neq 0} \frac{\|T^* Tx\|_\beta^2}{\|x\|_\beta^1} \right) \\ &= \bigvee_{\beta \leq \alpha} \left(\bigwedge_{x \in U, x \neq 0} \frac{\|T\|_\beta^2 \|Tx\|_\beta^2}{\|x\|_\beta^1} \right) \end{aligned}$$

$$\begin{aligned} &= \bigvee_{\beta\leq\alpha} \frac{\|T^2x\|_\beta^2}{\|x\|_\beta^1} \\ &= \|T\|_\alpha^2. \end{aligned}$$

Hence the equality is obtained. □

References

[1] T. Bag, S. K. Samanta, Fuzzy Bounded Linear Operators, Fuzzy Sets Syst., 151(3): 513-547, 2005.

[2] S. C. Cheng, J. N. Mordeson, Fuzzy Linear Operators and Fuzzy Normed Linear Spaces, Bull. Cal. Math. Soc., 86(5): 429-436, 1994.

[3] C O. Christensen, An introduction to frames and Riesz bases, 2003, Birkhauser.

[4] B. Daraby, Z. Solimani, A. Rahimi, Some Properties of Fuzzy Hilbert Spaces, Complex Anal. Oper. Theory., 11(1): 119-138, 2017.

[5] B. Daraby, Z. Solimani, A. Rahimi, A Note on Fuzzy Hilbert Spaces, J. Intell. Fuzzy Syst., 31(1): 313-319, 2016.

[6] A. K. Katsaras, Fuzzy Topological Vector Spaces II, Fuzzy Sets and Systems, 12(2): 143-154, 1984.

[7] C. Felbin, Finite Dimensional Fuzzy Normed Linear Spaces, Fuzzy Sets Syst., 48(2): 239-248, 1992.

Author index and list of Participants

*Speaker
p(Persian)