



روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

تألیف:

دکتر داود خجسته سالکویه

استاد دانشکده علوم ریاضی دانشگاه گیلان

چاپ اول

مرکز نشر دانشگاه گیلان

روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات
خطی بزرگ و تُنک

Iterative Methods and Preconditioning for Solving Large and Sparse System of Linear Equations

By:

Davod Khojasteh Salkuyeh, Ph.D.

University of Guilan Press

تألیف: دکتر داود خجسته سالکویه

حل تقریبی بسیاری از مدل‌های ریاضی و مهندسی منجر به حل دستگاه معادلات خطی بزرگ و تُنک می‌شود که محاسبهٔ جواب آنها توسط روش‌های مستقیم ناممکن یا با خطای زیادی همراه است. از این‌رو، برای حل این گونه دستگاه‌ها از روش‌های تکراری و برای بهبود همگرایی آنها از راه‌کار پیش‌شرط‌سازی استفاده می‌شود. این کتاب به مرور روش‌های تکراری رایج و شیوه‌های پیش‌شرط‌سازی متداول می‌پردازد. دانشجویان تحصیلات تکمیلی رشته‌های مختلف ریاضی و مهندسی می‌توانند از این کتاب به‌عنوان منبع درسی استفاده نمایند.



9 786001 533389

ISBN: 978-600-153-338-9



روش‌های تکراری و پیش شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

تألیف:

دکتر داود خجسته سالکویه

استاد دانشکده علوم ریاضی دانشگاه گیلان

مرکز نشر دانشگاه گیلان

۱۴۰۴



دانشگاه گیلان
1353-1974

شابک: ۹-۳۳۸-۱۵۳-۶۰۰-۹۷۸

سرشناسه	: خجسته سالکویه، داود، ۱۳۴۹ -
عنوان و نام پدیدآور	: روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ/تالیف داود خجسته‌سالکویه؛ ویراستار علمی و ادبی حسین امینی‌خواه.
مشخصات نشر	: رشت: دانشگاه گیلان، انتشارات، ۱۴۰۴.
مشخصات ظاهری	: ۴۳۲ص.
شابک	: 978-600-153-338-9
وضعیت فهرست نویسی	: فیپا
یادداشت	: واژه‌نامه.
یادداشت	: کتابنامه: ص. ۴۰۳ - ۴۱۲.
یادداشت	: نمایه.
موضوع	: معادله‌های خطی -- راهنمای آموزشی (عالی). linear equations -- Study and teaching (Higher)※ معادله‌های خطی -- مسائل، تمرین‌ها و غیره (عالی) linear equations -- Problems, excercises, etc. (Higher)※
شناسه افزوده	: امینی‌خواه، حسین، ۱۳۵۷ -، ویراستار
شناسه افزوده	: دانشگاه گیلان. انتشارات
رده بندی کنگره	: QA۲۴۲
رده بندی دیویی	: ۵۱۵/۲۵۲
شماره کتابشناسی ملی	: ۱۰۱۲۴۴۷۷

مرکز نشر دانشگاه گیلان

عنوان کتاب	: روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ
مؤلف	: دکتر داود خجسته سالکویه
ویراستار علمی و ادبی	: دکتر حسین امینی خواه
نوبت چاپ	: اول، ۱۴۰۴
ناشر	: مرکز نشر دانشگاه گیلان
شمارگان	: ۱۰۰۰ جلد

* هر گونه چاپ و تکثیر صرفاً در اختیار مرکز نشر دانشگاه گیلان است.

تقدیم به

همسر

و

دخترهای نازنینم

فاطمه و فائزه

فهرست اختصارات و روش‌ها

Lanczos Algorithm	الگوریتم لانکزوس
Direct Lanczos (D-Lanczos) Algorithm	الگوریتم لانکزوس مستقیم
Alternating Direction Implicit (ADI)	تبدیل مستقیم ضمنی
Incomplete LU (ILU) factorization	تجزیه ناقص LU
Incomplete LU is Skyline (ILUS)	تجزیه ناقص LU در قالب اسکی‌لاین
Incomplete Cholesky (IC) factorization	تجزیه ناقص چولسکی
Incomplete Cholesky with Threshold (ICT) Parameter	تجزیه ناقص چولسکی با پارامتر ازپیش‌تعیین‌شده
Approximate Inverse (AINV)	تقریب معکوس
Factored Approximate Inverse (FAPINV)	تقریب معکوس تجزیه‌شده
Forward Factored Approximate Inverse (FFAPINV)	تقریب معکوس تجزیه‌شده پیش‌رو
Backward Factored Approximate Inverse (BFAPINV)	تقریب معکوس تجزیه‌شده پس‌رو
Singular Value Decomposition (SVD)	تجزیه مقدار تکین
Uzawa Iteration Method	روش تکراری اوزاوا
Richardson Iteration Method	روش تکراری ریچاردسون
Jacobi Iteration Method	روش تکراری ژاکوبی
Alternating Directions Implicit (ADI) method	روش جهت‌های متناوب ضمنی
Chebyshev Acceleration Method	روش شتاب چبیشف
Augmented Lagrangian Method	روش لاگرانژی افزوده
Full Orthogonalization Method (FOM)	روش متعامدسازی کامل
Steepest Descent (SD)	سریع‌ترین کاهش
Residual Norm Steepest Descent (RNSD)	سریع‌ترین کاهش نرم مانده

Positive Definite and skew-Hermitian Splitting (PSS)	شکافت معین مثبت و هرمیتی کج
Deteriorated Positive-Definite and skew-Hermitian Splitting (DPSS)	شکافت معین مثبت و هرمیتی کج تباهیده
Normal and skew-Hermitian Splitting (NSS)	شکافت نرمال و هرمیتی کج
Positive Semidefinite and Positive Semidefinite Splitting (PPS)	شکافت نیمه معین مثبت و نیمه معین مثبت
Hemitian and skew-Hermitian Splitting (HSS)	شکافت هرمیتی و هرمیتی کج
Inexact Hemitian and skew-Hermitian Splitting (IHSS)	شکافت هرمیتی و هرمیتی کج ناقص
Modified Hemitian and skew-Hermitian Splitting (MHSS)	شکافت هرمیتی و هرمیتی کج اصلاح شده
Preconditioned Modified Hemitian and skew-Hermitian Splitting (PMHSS)	شکافت هرمیتی و هرمیتی کج اصلاح شده پیش شرط سازی شده
Accelerated Overrelaxation (AOR)	فوق تخفیف شتاب داده شده
Successive Overrelaxation (SOR)	فوق تخفیف متوالی
Symmetric Successive Overrelaxation (SSOR)	فوق تخفیف متوالی متقارن
Gauss-Seidel (GS)	گوس-سایدل
Symmetric Gauss-Seidel (SGS)	گوس-سایدل متقارن
Bi-Conjugate Gradient (BiCG)	گرادیان دومزدوج
Bi-Conjugate Gradient Stabilized (BiCGSTAB)	گرادیان دومزدوج پایدار شده
Conjugate Gradient (CG)	گرادیان مزدوج
Conjugate Gradient Squared (CGS)	گرادیان مزدوج مربع شده
Two-Sided Lanczos	لانکروزس دوطرفه
Quasi-Minimal Residual (QMR)	مانده شبه مینیمال
Minimal Residual (MR)	مانده مینیمال
Generalized Minimal Residual (GMRES)	مانده مینیمال تعمیم یافته
Flexible Generalized Minimal Residual (FGMRES)	مانده مینیمال تعمیم یافته انعطاف پذیر
Symmetric Approximate Minimum Degree (SYMAMD)	مینیمم درجه تقریبی متقارن

فهرست مطالب

ط	پیشگفتار
۱	۱ پیش‌زمینهٔ جبر خطی عددی
۱	۱-۱ مفاهیم، تعاریف و قضایای مورد نیاز
۱۲	۲-۱ مسئلهٔ مقدار ویژه
۱۶	۳-۱ ماتریس‌های معین مثبت
۲۶	۴-۱ تجزیهٔ QR
۲۸	۵-۱ تجزیهٔ مقدار تکین
۳۲	۶-۱ مسئلهٔ کمترین توان‌های دوم و دستگاه معادلات نرمال
۳۸	۷-۱ روش حذفی گاوس و تجزیهٔ LU
۴۳	۸-۱ یک مدل برای آزمایش‌های عددی
۴۶	۹-۱ ماتریس‌های تُنک و به‌کارگیری آنها
۴۷	۱-۹-۱ قالب مختصاتی
۴۹	۲-۹-۱ قالب سطری تُنک فشرده
۵۱	۳-۹-۱ قالب ستونی تُنک فشرده
۵۱	تمرینات
۵۷	۲ روش‌های تکراری ایستا
۵۷	۱-۲ مقدمه
۵۹	۲-۲ روش R و خواص آن

۶۴	۳-۲	روش‌های تکراری ایستا در حالت کلی
۶۸	۱-۳-۲	نرخ همگرایی
۷۲	۲-۳-۲	انتخاب ماتریس M
۷۲	۳-۳-۲	روش تکراری ریچاردسون
۷۳	۴-۳-۲	روش تکراری AOR
۷۵	۵-۳-۲	روش تکراری SSOR
۷۸	۴-۲	همگرایی روش تکراری ریچاردسون
۸۰	۵-۲	همگرایی روش‌ها برای ماتریس‌های معین مثبت هرمیتی
۸۵	۶-۲	محاسبه پارامتر بهینه در روش SOR
۹۶	۷-۲	همگرایی AOR برای ماتریس‌های غالب قطری
۱۰۲	۸-۲	ماتریس‌های نامنفی و همگرایی روش‌های تکراری ایستا
۱۱۶	۹-۲	یک قضیه مهم و کاربردهای آن
۱۲۳	۱۰-۲	روش شتاب چبیشف
۱۳۱		تمرینات

۱۳۷	۳	نیمه‌همگرایی روش‌های تکراری ایستا
۱۳۷	۱-۳	مقدمه
۱۳۸	۲-۳	شاخص یک ماتریس
۱۴۲	۳-۳	زیرفضاهای پایا و خواص آن
۱۴۴	۴-۳	تجزیه هسته-پوچ توان
۱۴۹	۵-۳	معکوس درازین
۱۵۲	۶-۳	نیمه‌همگرایی
۱۶۱		تمرینات

۱۶۵	۴	عملگرهای تصویر
۱۶۵	۱-۴	معرفی عملگر تصویر و خواص آن
۱۷۰	۲-۴	نمایش ماتریسی یک عملگر تصویر
۱۷۲	۳-۴	عملگر تصویر متعامد

۱۷۶	۴-۴ خواص عملگرهای تصویر متعامد
۱۷۷	تمرینات

۵ روش‌های تصویری برای حل دستگاه معادلات خطی

۱۷۹	۱-۵ روش‌های تصویری در حالت کلی
۱۸۲	۲-۵ نمایش ماتریسی
۱۸۴	۳-۵ آنالیز روش‌های تصویری
۱۸۹	۴-۵ روش‌های تصویری یک‌بُعدی
۱۹۰	۱-۴-۵ روش سریع‌ترین کاهش
۱۹۵	۲-۴-۵ ماندهٔ مینیمال
۱۹۷	۳-۴-۵ سریع‌ترین کاهش نرم مانده
۲۰۰	تمرینات

۶ روش‌های مبتنی بر زیرفضای کرایلف (قسمت ۱)

۲۰۳	۱-۶ معرفی زیرفضای کرایلف
۲۰۵	۲-۶ خواص زیرفضای کرایلف و محاسبهٔ پایه برای آن
۲۱۴	۳-۶ الگوریتم متعامدسازی کامل
۲۱۶	۱-۳-۶ الگوریتم FOM با شروع مجدد
۲۱۷	۲-۳-۶ الگوریتم FOM کامل
۲۱۷	۴-۶ الگوریتم ماندهٔ مینیمال تعمیم‌یافته
۲۲۰	۵-۶ به‌کارگیری الگوریتم‌های FOM و GMRES
۲۲۰	۱-۵-۶ به‌کارگیری الگوریتم GMRES
۲۲۴	۲-۵-۶ به‌کارگیری الگوریتم FOM
۲۲۴	۶-۶ همگرایی الگوریتم $GMRES(m)$ با شروع مجدد
۲۲۷	۷-۶ الگوریتم GMRES در متلب
۲۲۹	۸-۶ الگوریتم لانکزوس و گرادیان مزدوج
۲۳۰	۱-۸-۶ الگوریتم لانکزوس
۲۳۵	۲-۸-۶ الگوریتم گرادیان مزدوج

۲۴۵	۹-۶	گرایان مزدوج برای حل دستگاه معادلات نرمال
۲۴۵	۱-۹-۶	الگوریتم CGNR
۲۴۸	۲-۹-۶	الگوریتم CGNE
۲۵۰		تمرینات

۷ روش‌های مبتنی بر زیرفضای کرایلف (قسمت ۲)

۲۵۵	۱-۷	الگوریتم لانکزوس برای ماتریس‌های نامتقارن
۲۶۰	۲-۷	الگوریتم لانکزوس دوطرفه برای حل دستگاه معادلات خطی
۲۶۲	۳-۷	الگوریتم گرایان دومزدوج (BiCG)
۲۶۸	۴-۷	الگوریتم مانده شبه‌مینیمال (QMR)
۲۷۵	۵-۷	الگوریتم گرایان مزدوج مربع‌شده (CGS)
۲۷۹	۶-۷	الگوریتم BiCG پایدارشده (BiCGSTAB)
۲۸۶		تمرینات

۸ پیش‌شرط‌سازی روش‌های تکراری

۲۸۸	۱-۸	روش‌های تکراری پیش‌شرط‌سازی شده
۲۸۹	۲-۸	الگوریتم گرایان مزدوج پیش‌شرط‌سازی شده
۲۹۲	۳-۸	الگوریتم $GMRES(m)$ پیش‌شرط‌سازی شده
۲۹۲	۱-۳-۸	پیش‌شرط‌سازی از چپ
۲۹۳	۲-۳-۸	پیش‌شرط‌سازی از راست
۲۹۵	۳-۳-۸	پیش‌شرط‌سازی دوطرفه
۲۹۵	۴-۳-۸	مقایسه پیش‌شرط‌ساز چپ و راست
۲۹۷	۵-۳-۸	الگوریتم $GMRES$ انعطاف‌پذیر (FGMRES)
۳۰۰	۴-۸	پیش‌شرط‌سازی دستگاه معادلات نرمال
۳۰۳	۵-۸	روش‌های تکراری پیش‌شرط‌سازی شده در متلب
۳۰۳	۱-۵-۸	پیش‌شرط‌سازی از چپ
۳۰۵	۲-۵-۸	پیش‌شرط‌سازی راست
۳۰۶		تمرینات

۳۰۷	۹ شیوه‌های ساخت پیش شرط‌ساز
۳۰۷	۱-۹ پیش شرط‌سازهای تولیدشده توسط روش‌های تکراری ایستا
۳۰۸	۱-۱-۹ پیش شرط‌ساز AOR
۳۰۹	۲-۱-۹ پیش شرط‌ساز SSOR
۳۱۱	۲-۹ پیش شرط‌سازهای مبتنی بر تجزیه ناقص LU
۳۱۱	۱-۲-۹ محاسبه تجزیه LU با کمک الگوریتم حذفی گاوس
۳۱۸	۲-۲-۹ تجزیه چولسکی ناقص
۳۲۱	۳-۲-۹ بهبود کارایی تجزیه چولسکی با کمک بازمرتب‌سازی
۳۲۲	۴-۲-۹ الگوریتم ILUS
۳۲۶	۳-۹ پیش شرط‌سازهای مبتنی بر معکوس تقریبی تُنک
۳۲۷	۱-۳-۹ تقریب مستقیم معکوس ماتریس
۳۳۰	۲-۳-۹ معکوس تقریبی تُنک تجزیه‌شده
۳۳۲	۳-۳-۹ الگوریتم FFAPINV
۳۳۶	۴-۳-۹ محاسبه تجزیه ILU بر پایه الگوریتم FFAPINV
۳۴۴	تمرینات
۳۴۷	۱۰ روش تکراری شکافت هرمیتی و هرمیتی‌کج
۳۴۷	۱-۱۰ روش تکراری شکافت هرمیتی و هرمیتی‌کج
۳۵۶	۲-۱۰ روش شکافت هرمیتی و هرمیتی‌کج نادقیق
۳۶۶	۳-۱۰ پیش شرط‌ساز HSS
۳۶۷	۴-۱۰ برای مطالعه بیشتر
۳۶۸	تمرینات
۳۷۱	۱۱ مسئله نقطه زینی
۳۷۱	۱-۱۱ برخی از کاربردهای مسئله نقطه زینی
۳۷۱	۱-۱-۱۱ مسائل بهینه‌سازی نقطه زینی
۳۷۳	۲-۱-۱۱ مسئله کمترین توان‌های دوم
۳۷۴	۳-۱-۱۱ حل مسئله ناویر-استوکس به روش عناصر متناهی

۳۷۵	۲-۱۱ وجود جواب برای مسئله نقطه زینی
۳۷۸	۳-۱۱ روش تکراری اوزاوا و پیش شرط ساز اوزاوا
۳۷۹	۴-۱۱ پیش شرط سازهای قطری بلوکی و بالامثلشی بلوکی
۳۸۵	۵-۱۱ پیش شرط ساز HSS
۳۹۵	۶-۱۱ روش لاگرانژی افزوده
۳۹۹	تمرینات

۴۰۲ منابع و مأخذ

۴۱۳ نمایه

۴۲۱ واژه نامه فارسی به انگلیسی

۴۲۷ واژه نامه انگلیسی به فارسی

پیشگفتار

بسیاری از مدل‌های مختلف در علوم و مهندسی توسط معادلات دیفرانسیل فرمول‌بندی می‌شوند؛ برای مثال مسئله‌های انتقال حرارت، حرکت موج، حرکت سیال و حل عددی این‌گونه مسائل به کمک روش‌های تفاضلات متناهی یا اجزای محدود منجر به حل دستگاه معادلات خطی می‌شوند که اغلب ابعاد آنها بزرگ و تُنک هستند. همچنین، مسائل مختلفی در علوم داده وجود دارند که برای حل آنها لازم است دستگاه‌هایی تنک با ابعاد بزرگ حل شوند؛ برای مثال مسئله رتبه‌صفحه یکی از این مسائل است.

حل این‌گونه دستگاه‌ها با استفاده از روش‌های مستقیم، مثل روش حذفی گاوس، با هزینه‌های زیادی همراه است و حتی گاهی ناممکن می‌باشد. دو دلیل عمده این است که اولاً انباشتگی خطا در روش‌های مستقیم زیاد است، ثانیاً حافظه زیادی از رایانه را اشغال می‌کنند. از این رو، از روش‌های تکراری برای حل آنها استفاده می‌شود.

به طور کلی، دو دسته روش تکراری برای حل دستگاه معادلات خطی وجود دارد: روش‌های تکراری ایستا و غیرایستا. در این کتاب تعداد زیادی از روش‌های تکراری را مطالعه می‌کنیم. نخست، یک روش تکراری ایستا را در حالت کلی تعریف کرده و شرایط همگرایی آن را مطالعه می‌کنیم. سپس چند حالت خاص آن را معرفی می‌کنیم. از روش‌های تکراری ایستا، روش‌های تکراری فوق‌تخفیف شتاب داده شده (AOR) (در حالت خاص، روش‌های تکراری ژاکوبی-گاوس-سایدل، فوق‌تخفیف متوالی (SOR))، شکافت هرمیتی و هرمیتی-کج (HSS)، ریچاردسون و ... را مطالعه می‌کنیم. همچنین، نیمه‌همگرایی روش‌های تکراری ایستا را نیز بررسی می‌کنیم. سپس به معرفی عملگرهای تصویر و خواص آنها می‌پردازیم. عملگرهای تصویر نقش اساسی در آنالیز عددی و اساس روش‌های تصویری برای حل دستگاه معادلات خطی است. اغلب روش‌های تکراری غیرایستا برای حل دستگاه معادلات خطی براساس عملگرهای تصویر است که فضای‌های مورد استفاده در آنها زیرفضای کرایلف می‌باشد. از روش‌های تکراری غیرایستا،

روش‌های تکراری مانده مینیمال (MR)، سریع‌ترین کاهش (SD)، گرادیان مزدوج (CG)، شتاب چیبیشف، مانده مینیمال تعمیم‌یافته (GMRES) گرادیان دومزدوج (BiCG) گرادیان دومزدوج پایدارشده (BiCGSTAB)، و ... را بررسی می‌کنیم.

یکی از مشکلات اساسی روش‌های تکراری کُندبودن سرعت همگرایی آنهاست و برای بهبود سرعت همگرایی از راه‌کار پیش‌شرط‌سازی استفاده می‌شود. یک پیش‌شرط‌ساز ماتریسی است که یک دستگاه را به دستگاه معادل دیگری تبدیل می‌کند که خوش‌حالت‌تر است. در این کتاب، مفاهیم اساسی پیش‌شرط‌سازی و شیوه پیش‌شرط‌سازی روش‌های تکراری را بیان کرده و در ادامه، چندین شیوه محاسبه پیش‌شرط‌سازها را ارائه می‌کنیم. اغلب روش‌های تکراری ارائه‌شده و شیوه‌های پیش‌شرط‌سازی آنها، با مثال‌های عددی مورد آزمایش قرار می‌گیرند.

در فصل پایانی، مسئله نقطه زینی که یکی از مسائل مهم در بهینه‌سازی و مکانیک سیالات می‌باشد را مطالعه کرده و روش‌های پیش‌شرط‌سازی این‌گونه مسائل را بررسی می‌کنیم.

سعی شده است که مطالب کتاب به‌صورت ساده و روان و برای هر مورد مثال‌هایی ارائه شود. همچنین، برای درک بهتر برخی از مطالب از شکل استفاده شده است. در بسیاری از موارد چگونگی به‌کارگیری و اجرای روش‌های تکراری و پیش‌شرط‌سازی آنها در محیط متلب شرح داده شده است. در پایان هر فصل تعدادی تمرین مطابق با مباحث همان فصل گنجانده شده است که حل آنها می‌تواند به دانشجو در درک مفاهیم همان فصل کمک نماید.

برای درک بهتر مطالب هر فصل، به دانشجویان گرامی توصیه می‌شود که برای الگوریتم‌های ارائه‌شده در هر فصل برنامه‌ای در محیط متلب یا هر زبان دیگر نوشته و با مثال‌های مختلف مورد آزمایش قرار دهند. پیشنهاد می‌کنم که برنامه‌ها در محیط متلب نوشته شوند، زیرا بسیار از روش‌های تکراری مورد مطالعه به‌صورت توابعی در کتابخانه متلب وجود دارند.

این کتاب را می‌توان به‌عنوان یک مرجع کامل برای چندین درس در دوره‌های تحصیلات تکمیلی رشته‌های ریاضی و علوم کامپیوتر و همچنین دوره‌های تحصیلات تکمیلی رشته‌های مهندسی مورد استفاده قرار داد. بدیهی است تذکرات خوانندگان گرامی در مورد لغزش‌های احتمالی موجب امتنان و از دریافت نظرات و پیشنهادات آنها سپاسگزار خواهم بود.

در پایان از واحد پژوهشی دانشگاه گیلان که مدیریت مراحل مختلف داوری، ویرایش و انتشار کتاب را برعهده گرفتند بسیار سپاسگزارم. از داوران محترم که کتاب را با دقت خاصی بررسی نموده و اینجانب را از نکات ارزشمند ویرایشی بهره‌مند کردند قدردانی می‌نمایم. از جناب آقای دکتر حسین امینی‌خواه که زحمت

ویرایش علمی و ادبی کتاب را کشیدند، بسیار ممنونم. این کتاب به صورت جزوه درسی در سال‌های گذشته به دانشجویان تحصیلات تکمیلی تدریس و روی مباحث مختلف آن بحث و تبادل نظر شد که به بهبود کتاب کمک شایانی نمود. بدین وسیله از همه آنها تشکر می‌کنم. همچنین، از خانواده‌ام به خاطر حمایت‌هایشان در طی مراحل آماده‌سازی کتاب سپاسگزاری می‌کنم.

دکتر داود خجسته سالکویه

دانشکده علوم ریاضی دانشگاه گیلان

اردیبهشت ۱۴۰۴

پیش‌زمینه جبر خطی عددی

در این فصل برخی از مفاهیم، تعاریف و قضایای مورد نیاز در فصل‌های بعدی را مرور می‌نماییم. اغلب این مطالب را می‌توان در کتاب‌های مختلف از جمله [۳، ۳۲، ۵۰، ۷۷] یافت.

۱-۱ مفاهیم، تعاریف و قضایای مورد نیاز

در سراسر این کتاب مجموعه اعداد حقیقی و مختلط را به ترتیب با $\mathbb{R}$ و $\mathbb{C}$ نشان می‌دهیم. همچنین، مجموعه بردارهای n تایی با مؤلفه‌های حقیقی را با $\mathbb{R}^n$ و مجموعه ماتریس‌های $m \times n$ با درایه‌های حقیقی را با $\mathbb{R}^{n \times n}$ نشان می‌دهیم. به همین ترتیب، مجموعه $\mathbb{C}^n$ و $\mathbb{C}^{m \times n}$ تعریف می‌شود. برای هر ماتریس A نمادهای $A(:, j)$ یا A_{*j} برای نمایش ستون j ام و از $A(i, :)$ یا A_{i*} برای نمایش سطر i ام ماتریس A به کار می‌رود. از نماد A^T و A^H به ترتیب برای نمایش ترانهاد و ترانهاد هرمیتی ماتریس A استفاده می‌کنیم.

ماتریس $A \in \mathbb{R}^{n \times n}$ را متقارن می‌گویند هرگاه $A^T = A$ و ماتریس $A \in \mathbb{C}^{n \times n}$ را هرمیتی می‌گویند هرگاه $A^H = A$. ماتریس $A \in \mathbb{R}^{n \times n}$ را متقارن کج می‌گویند هرگاه $A^T = -A$ و ماتریس $A \in \mathbb{C}^{n \times n}$ را هرمیتی کج می‌نامند هرگاه $A^H = -A$. ماتریس $A \in \mathbb{R}^{n \times n}$ یک ماتریس متعامد نامیده می‌شود هرگاه $A^T A = A A^T = I$ و در صورتی که ماتریس A مختلط باشد و $A^H A = A A^H = I$ ، می‌گوییم ماتریس A یکانی است. ماتریس مربعی حقیقی یا مختلط A نرمال نامیده می‌شود هرگاه $A^H A = A A^H$. ماتریس $P \in \mathbb{R}^{n \times n}$ را یک ماتریس جایگشت می‌گویند هرگاه

۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

در هر سطر و ستون آن یک و فقط یک ۱ وجود داشته باشد. می‌توان دید که هر ماتریس جایگشت یک ماتریس متعامد است.

ماتریس $A \in \mathbb{C}^{n \times n}$ را **قطری** می‌گویند، هرگاه به ازای $i \neq j$ ، $a_{ij} = 0$. یک ماتریس قطری معمولاً با $\text{diag}_n(a_{11}, a_{22}, \dots, a_{nn})$ نشان داده می‌شود. ماتریس A را **سه‌قطری** می‌گویند، هرگاه به ازای هر i ، $|i - j| > 1$ ، $a_{ij} = 0$. ماتریس سه‌قطری

$$T = \begin{pmatrix} b & c & & & \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ & & & a & b \end{pmatrix} \in \mathbb{C}^{n \times n},$$

را با $T = \text{tridiag}_n(a, b, c)$ نمایش می‌دهیم.

ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ را **بالامثلثی** می‌گوییم، هرگاه به ازای هر $i > j$ ، $a_{ij} = 0$ و A را **پائین‌مثلثی** می‌گوییم هرگاه به ازای هر $i < j$ ، $a_{ij} = 0$. ماتریس $A \in \mathbb{C}^{n \times n}$ را **شبه‌بالامثلثی** می‌گوییم، هرگاه به صورت

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ & A_{22} & \cdots & A_{2p} \\ & & \ddots & \vdots \\ & & & A_{pp} \end{pmatrix},$$

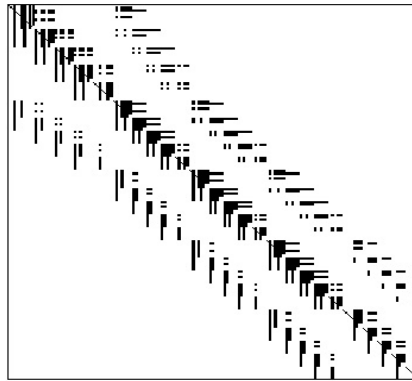
باشد که در آن $i = 1, 2, \dots, p$ ، A_{ii} ماتریس‌های مربعی هستند.

همچنین، ماتریس A را **بالاهسنبرگی** می‌گویند، هرگاه به ازای هر $i - j > 1$ ، $a_{ij} = 0$. در واقع یک ماتریس بالاهسنبرگی به شکل

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ & \ddots & \ddots & \vdots & \vdots \\ & & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ & & & a_{n,n-1} & a_{nn} \end{pmatrix},$$

است.

ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ را یک ماتریس نواری می‌گویند هرگاه به‌ازای $j < i - k_1$ یا $j > i + k_2$ داشته باشیم $a_{ij} = 0$ ، که در آن k_1 و k_2 دو عدد صحیح نامنفی هستند. اعداد k_1 و k_2 را به‌ترتیب پهنای پائینی و بالانی نواری می‌گویند. همچنین، مقدار $k = \max\{k_1, k_2\}$ را پهنای نواری ماتریس A می‌گویند. روشن است که هر ماتریس قطری یک ماتریس نواری با $k_1 = k_2 = 0$ است و پهنای نواری برابر است با $k = 0$. به‌علاوه، هر ماتریس سه‌قطری یک ماتریس نواری با $k_1 = k_2 = 1$ است و پهنای نواری برابر است با $k = 1$. شکل ۱-۱ نمونه‌ای دیگر از یک ماتریس نواری را نشان می‌دهد.



شکل ۱-۱ یک نمونه از ماتریس نواری

تعریف ۱.۱ یک فضای برداری شامل دو مجموعه V و $\mathbb{F}$ و دو عمل جمع برداری و ضرب اسکالر است. اعضای V را بردار و اعضای $\mathbb{F}$ را اسکالر می‌گویند. در این کتاب $\mathbb{F}$ ، میدان اعداد حقیقی $\mathbb{R}$ یا میدان اعداد مختلط $\mathbb{C}$ است. اگر عمل جمع برداری را با “+” و ضرب اسکالر را با “.” نشان دهیم، آنگاه می‌گوییم V یک فضای برداری روی میدان $\mathbb{F}$ است، هرگاه

$$(۱) \text{ به‌ازای هر } u, v \in V, u + v \in V.$$

$$(۲) \text{ به‌ازای هر } \alpha \in \mathbb{F} \text{ و } v \in V, \alpha \cdot v \in V.$$

$$(۳) \text{ به‌ازای هر } u, v \in V, u + v = v + u.$$

$$(۴) \text{ به‌ازای هر } u, v, w \in V, (u + v) + w = u + (v + w).$$

$$(۵) \text{ عضوی مثل } 0 \in V \text{ وجود دارد به‌طوری‌که به‌ازای هر } v \in V, v + 0 = v.$$

$$(۶) \text{ به‌ازای هر } v \in V \text{ عضوی مثل } -v \in V \text{ وجود دارد به‌طوری‌که } v + (-v) = 0.$$

(۷) برای هر $\alpha, \beta \in \mathbb{F}$ و $v \in V$ $(\alpha\beta).v = \alpha.(\beta.v)$.

(۸) برای هر $\alpha, \beta \in \mathbb{F}$ و $v \in V$ $(\alpha + \beta).v = \alpha.v + \beta.v$.

(۹) برای هر $\alpha \in \mathbb{F}$ و $v, w \in V$ $\alpha.(v + w) = \alpha.v + \alpha.w$.

(۱۰) برای هر $v \in V$ $1.v = v$ (در اینجا $1 \in \mathbb{F}$).

یک فضای برداری معمولاً به صورت $(V, \mathbb{F})$ نشان داده می‌شود و در صورتی که جای اشتباه وجود نداشته باشد گاهی با V نیز نشان داده می‌شود.

فرض کنید V یک فضای برداری روی میدان $\mathbb{F}$ باشد و $W = \{v_1, v_2, \dots, v_n\} \subseteq V$ بردار $w \in V$ یک ترکیب خطی از مجموعه بردارهای W نامیده می‌شود هرگاه اسکالرهای $\alpha_i \in \mathbb{F}$ ، $i = 1, 2, \dots, n$ وجود داشته باشند به طوری که

$$w = \sum_{i=1}^n \alpha_i v_i.$$

می‌گوییم $W = \{v_1, v_2, \dots, v_n\}$ یک مجموعه مستقل خطی نامیده می‌شوند، اگر به ازای $\alpha_i \in \mathbb{F}$ ، $i = 1, 2, \dots, m$ داشته باشیم

$$\sum_{i=1}^n \alpha_i v_i = 0,$$

آنگاه $\alpha_i = 0$ ، $i = 1, 2, \dots, m$. در غیر این صورت، v_i ها را وابسته خطی می‌گویند. فضای تولیدشده توسط مجموعه بردارهای $\mathcal{B} = \{v_1, v_2, \dots, v_m\}$ را با $\text{span}(\mathcal{B})$ نشان داده و به صورت

$$\text{span}(\mathcal{B}) = \left\{ \sum_{i=1}^m \alpha_i v_i : \alpha_i \in \mathbb{F}, i = 1, 2, \dots, m \right\},$$

تعریف می‌شود. مجموعه بردارهای $\mathcal{B}$ یک پایه برای فضای برداری V نامیده می‌شود هرگاه یک مجموعه مستقل خطی باشد و $V = \text{span}(\mathcal{B})$. در این صورت، عدد m را بُعد فضای V می‌نامند و با $\dim(V)$ نشان می‌دهند.

تعریف ۲.۱ فرض کنید R و S دو زیرفضای برداری از فضای برداری V باشند. جمع این دو زیرفضا را با $R + S$ نشان داده و به صورت

$$R + S = \{r + s : r \in R, s \in S\},$$

تعریف می‌شود. می‌گوییم T جمع مستقیم زیرفضاهای R و S است هرگاه $R \cap S = \{0\}$ و $R + S = T$. در این صورت، می‌نویسیم $T = R \oplus S$.

قضیه ۱.۱ فرض کنید R و S دو زیرفضای برداری از فضای برداری V باشند. در این صورت،

$$\dim(R + S) = \dim(R) + \dim(S) - \dim(R \cap S).$$

به علاوه، اگر $T = R \oplus S$ ، آنگاه $\dim(R \oplus S) = \dim(R) + \dim(S)$.

مثال ۱.۱ فرض کنید $T = \mathbb{C}^3$ و

$$R = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} : x_1, x_2 \in \mathbb{C} \right\}, \quad S = \left\{ \begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix} : x_2, x_3 \in \mathbb{C} \right\}.$$

داریم

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ \frac{1}{2}x_2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{2}x_2 \\ x_3 \end{pmatrix}.$$

بنابراین، هر عضو مجموعه T را می‌توان به صورت مجموع یک عضو از R و یک عضو از S نوشت. از طرفی، به ازای هر مقدار ناصفر $\alpha \in \mathbb{C}$ ، داریم

$$\begin{pmatrix} 0 \\ \alpha \\ 0 \end{pmatrix} \in R \cap S.$$

از این رو، داریم $T = R + S$ ، اما T به صورت جمع مستقیم R و S نیست.

مثال ۲.۱ فرض کنید $T = \mathbb{C}^{n \times n}$ و

$$R = \{A \in T : A^H = A\}, \quad S = \{A \in T : A^H = -A\}.$$

در این صورت، هر ماتریس $A \in T$ را می‌توان به صورت

$$A = \frac{1}{2}(A + A^H) + \frac{1}{2}(A - A^H),$$

نوشت. ماتریس $(A + A^H)/2$ هرمیتی و ماتریس $(A - A^H)/2$ هرمیتی‌کج هستند. بنابراین، به‌ترتیب متعلق به R و S هستند. ازاین‌رو، داریم $T = R + S$. ازطرفی، اگر $A \in R \cap S$ ، آنگاه خواهیم داشت $A^H = A = -A$ ، که نتیجه می‌دهد $A = 0$. بنابراین، داریم $T = R \oplus S$.

تعریف ۳.۱ فرض کنید $A \in \mathbb{C}^{m \times n}$. دراین‌صورت، برد A و فضای پوچ A را به‌ترتیب با $\mathcal{R}(A)$ و $\mathcal{N}(A)$ نشان می‌دهیم و به‌صورت زیر تعریف می‌شوند

$$\mathcal{R}(A) = \{Ax : x \in \mathbb{C}^n\}, \quad \mathcal{N}(A) = \{x \in \mathbb{C}^n : Ax = 0\}.$$

قضیه ۲.۱ [۷۷] فرض کنید M و N دو زیرفضای برداری از فضای برداری V باشند به‌طوری‌که $M \subseteq N$ ، دراین‌صورت،

$$\dim(M) \leq \dim(N) \quad (\text{الف})$$

(ب) اگر $\dim(M) = \dim(N)$ ، آنگاه $M = N$.

تعریف ۴.۱ فرض کنید $A \in \mathbb{C}^{m \times n}$. تعداد ستون‌های مستقل خطی ماتریس A را رتبه A می‌نامند و با $\text{rank}(A)$ نشان داده می‌شود.

قضیه ۳.۱ [۷۷] به ازای هر $A \in \mathbb{C}^{m \times n}$ ، داریم

$$\text{rank}(A) = \text{rank}(A^H), \quad \text{rank}(A) = \dim(\mathcal{R}(A)).$$

تعریف ۵.۱ فرض کنید V یک فضای برداری روی میدان $\mathbb{F}$ باشد. یک نرم روی V ، تابعی است مثل $\|\cdot\|$ از V به $\mathbb{R}$ ، به‌طوری‌که در شرایط زیر صدق کند

(الف) به‌ازای هر $x \in V$ ، $\|x\| \geq 0$ ، به‌علاوه، $\|x\| = 0$ اگر و تنها اگر $x = 0$ ؛

(ب) به‌ازای هر $\alpha \in \mathbb{F}$ و $x \in V$ ، $\|\alpha x\| = |\alpha| \|x\|$ ؛

(ج) به‌ازای هر $x, y \in V$ ، $\|x + y\| \leq \|x\| + \|y\|$ (نامساوی مثلث).

اگر $\|\cdot\|$ یک نرم روی V باشد، آنگاه به‌ازای هر $x, y \in V$ ، داریم

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

از اینجا به بعد، نرم‌های تعریف‌شده روی $\mathbb{C}^n$ و $\mathbb{C}^{m \times n}$ را به‌ترتیب نرم برداری و نرم ماتریسی می‌نامیم.

برای هر $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{C}^n$ نرم‌های برداری متداول عبارتند از

$$\|x\|_1 = \sum_{i=1}^n |x_i|, \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|, \quad \|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} = (x^H x)^{\frac{1}{2}},$$

که به ترتیب نرم-۱، نرم بینهایت، و نرم-۲ (نرم اقلیدسی) می‌باشند. همچنین، سه نرم ماتریسی مهم روی $\mathbb{C}^{m \times n}$ عبارتند از

$$\begin{aligned} \|A\|_1 &= \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|, \\ \|A\|_\infty &= \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|, \\ \|A\|_F &= \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}, \end{aligned}$$

که به ترتیب به نرم-۱، نرم بینهایت، و نرم فروبنیوس معروفند.

تعریف ۶.۱ (نرم طبیعی) فرض کنید $\|\cdot\|$ یک نرم برداری روی $\mathbb{C}^n$ باشد و $A \in \mathbb{C}^{m \times n}$. در این صورت، **نرم طبیعی** متناظر به این نرم برداری به صورت

$$\|A\| = \max_{\|x\|=1} \|Ax\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}, \quad (1-1)$$

تعریف می‌شود.

می‌توان دید که نرم‌های طبیعی تولیدشده توسط نرم-۱ برداری و نرم بینهایت برداری به ترتیب نرم‌های ماتریسی طبیعی نرم-۱ و نرم بینهایت هستند. همچنین، اگر $\|\cdot\|_v$ یک نرم برداری و $\|\cdot\|_M$ نرم طبیعی متناظر باشد، آنگاه به‌ازای هر $A, B \in \mathbb{C}^{n \times n}$ و $x \in \mathbb{C}^n$ داریم

$$\|AB\|_M \leq \|A\|_M \|B\|_M \quad (1)$$

$$\|Ax\|_v \leq \|A\|_M \|x\|_v \quad (2)$$

خاصیت اول را **خاصیت ضربی** نرم و خاصیت دوم را **خاصیت سازگاری** نرم ماتریسی نسبت به نرم برداری مولد آن می‌گویند.

رد (یا اثر) ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ به صورت $\text{trace}(A) = \sum_{k=1}^n a_{kk}$ تعریف می‌شود.

به سادگی می‌توان دید

۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

۱. به‌ازای هر $\alpha \in \mathbb{C}$ ، $\text{trace}(\alpha A) = \alpha \text{trace}(A)$ ؛

۲. $\text{trace}(A^H) = \overline{\text{trace}(A)}$ ؛

۳. $\|A\|_F = \sqrt{\text{trace}(A^H A)}$ ؛

۴. اگر $A, B \in \mathbb{C}^{n \times n}$ ، آنگاه $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$ ؛

۵. اگر $A \in \mathbb{C}^{m \times n}$ و $B \in \mathbb{C}^{n \times m}$ ، آنگاه خواهیم داشت $\text{trace}(AB) = \text{trace}(BA)$ ، حتی اگر $AB \neq BA$.

فرض کنید $U \in \mathbb{C}^{n \times n}$ یک ماتریس یکانی باشد. در این صورت، به‌ازای هر $x \in \mathbb{C}^n$ ، داریم

$$\|Ux\|_2 = \sqrt{(Ux)^H Ux} = \sqrt{x^H U^H Ux} = \sqrt{x^H x} = \|x\|_2.$$

همچنین، اگر $A \in \mathbb{C}^{n \times n}$ ، آنگاه به‌طور مشابه می‌توان دید که

$$\|UA\|_F = \|AU\|_F = \|A\|_F, \quad \|UA\|_2 = \|AU\|_2 = \|A\|_2.$$

تعریف ۷.۱ فرض کنید V یک فضای برداری روی میدان $\mathbb{F}$ باشد. یک ضرب داخلی روی V ، تابعی است که به هر زوج از بردارهای x و y در V یک اسکالر مثل $\langle x, y \rangle$ در $\mathbb{F}$ را نسبت می‌دهد به‌طوری‌که در شرایط زیر صدق کند

(الف) $\langle x, x \rangle$ حقیقی باشد و $\langle x, x \rangle \geq 0$. به‌علاوه، $\langle x, x \rangle = 0$ اگر و تنها اگر $x = 0$ ؛

(ب) برای هر $\alpha \in \mathbb{F}$ ، $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ ؛

(ج) برای هر $x, y, z \in V$ ، $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ ؛

(د) $\langle x, y \rangle = \overline{\langle y, x \rangle}$. در اینجا علامت “ $-$ ” به‌معنای مزدوج مختلط است.

هر فضای برداری که یک ضرب داخلی در آن تعریف شده باشد را یک فضای ضرب داخلی می‌نامند.

با توجه به تعریف ضرب داخلی، به‌ازای هر $\alpha \in \mathbb{F}$ و $x, y \in V$ ، داریم

$$\langle x, \alpha y \rangle = \overline{\langle \alpha y, x \rangle} = \overline{\alpha \langle y, x \rangle} = \overline{\alpha} \overline{\langle y, x \rangle} = \overline{\alpha} \langle x, y \rangle.$$

همچنین، به‌ازای هر $x, y, z \in V$ ، داریم

$$\begin{aligned} \langle x, y + z \rangle &= \overline{\langle y + z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} \\ &= \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle. \end{aligned}$$

قضیه ۴.۱ (نامساوی کوشی-شوارتز) [۷۷] فرض کنید V یک فضای ضرب داخلی روی میدان $\mathbb{F}$ و $\langle \cdot, \cdot \rangle$ یک ضرب داخلی دلخواه در V باشد. در این صورت، به ازای هر $x \in V$ ، داریم

$$|\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle}.$$

این نامساوی، به نامساوی کوشی-شوارتز معروف است.

قضیه ۵.۱ [۷۷] فرض کنید V یک فضای ضرب داخلی روی میدان $\mathbb{F}$ و $\langle \cdot, \cdot \rangle$ یک ضرب داخلی دلخواه در V باشد. در این صورت، اگر به ازای هر $x \in V$ تعریف کنیم

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad (۲-۱)$$

آنگاه $\|\cdot\|$ یک نرم روی V تعریف می‌کند.

با توجه به تعریف (۲-۱) نامساوی کوشی-شوارتز را می‌توان به صورت زیر نوشت

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

مثال ۳.۱ فرض کنید $V = \mathbb{C}^{n \times n}$. به ازای هر $x, y \in V$ ، تعریف می‌کنیم

$$\langle x, y \rangle = \sum_{i=1}^n x_i \bar{y}_i = y^H x, \quad (۳-۱)$$

که در آن x_i ها و y_i ها به ترتیب درایه‌های x و y هستند. به سادگی می‌توان دید که $\langle \cdot, \cdot \rangle$ تعریف شده در (۳-۱) در شرایط ضرب داخلی صدق می‌کند. این ضرب داخلی را ضرب داخلی استاندارد می‌گویند. نرم تولیدشده توسط ضرب داخلی استاندارد، همان نرم-۲ است، زیرا

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x^H x} = \|x\|_2.$$

به ازای هر ماتریس $A \in \mathbb{C}^{n \times n}$ و $x, y \in \mathbb{C}^n$ ، داریم

$$\langle x, Ay \rangle = (Ay)^H x = y^H A^H x = \langle A^H x, y \rangle.$$

در حالت خاص، اگر ماتریس A هرمیتی باشد، آنگاه $\langle x, Ay \rangle = \langle Ax, y \rangle$. عکس این مطلب نیز درست است. در واقع، اگر به ازای هر $x, y \in \mathbb{C}^n$ داشته باشیم $\langle x, Ay \rangle = \langle Ax, y \rangle$ ، آنگاه A هرمیتی است (تمرین ۴.۱ را ببینید).

۱۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و مُتَنگ

مثال ۴.۱ فرض کنید $V = \mathbb{C}^{m \times n}$. به‌ازای هر $X = (x_{ij}), Y = (y_{ij}) \in V$ ، تعریف می‌کنیم

$$\langle X, Y \rangle = \text{trace}(Y^H X) = \sum_{i=1}^n \sum_{j=1}^m \overline{y_{ji}} x_{ji}. \quad (۴-۱)$$

در این صورت، $\langle \cdot, \cdot \rangle$ یک ضرب داخلی روی V است. نرم تولیدشده توسط ضرب داخلی استاندارد، همان نرم فروبنیوس است، زیرا

$$\begin{aligned} \|X\| &= \sqrt{\langle X, X \rangle} = \sqrt{\text{trace}(X^H X)} \\ &= \left(\sum_{i=1}^n \sum_{j=1}^m \overline{x_{ji}} x_{ji} \right)^{\frac{1}{2}} \\ &= \left(\sum_{i=1}^n \sum_{j=1}^m |x_{ji}|^2 \right)^{\frac{1}{2}} \\ &= \|X\|_F. \end{aligned}$$

تعریف ۸.۱ مجموعه بردارهای $V_m = \{v_1, v_2, \dots, v_m\}$ از یک فضای ضرب داخلی با ضرب داخلی $\langle \cdot, \cdot \rangle$ را یک مجموعه یکامتعامد می‌گویند، هرگاه

$$\langle v_i, v_j \rangle = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

در تعریف ۸.۱، به‌ازای هر بردار $v_i \in V_m$ داریم

$$\|v_i\|^2 = \sqrt{\langle v_i, v_i \rangle} = 1.$$

بنابراین، $\langle v_i, v_1 \rangle = 1$ اگر و تنها اگر $\|v_i\| = 1$.

مثال ۵.۱ در $\mathbb{R}^3$ مجموعه بردارهای

$$B = \left\{ v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, v_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, v_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} \right\}, \quad (۵-۱)$$

نسبت به ضرب داخلی استاندارد، یک مجموعهٔ یکامتعامد است؛ زیرا

$$\|v_1\|_2 = \|v_2\|_2 = \|v_3\|_2 = 1, \quad \langle v_1, v_2 \rangle = \langle v_1, v_3 \rangle = \langle v_2, v_3 \rangle = 0.$$

فرض کنید $B = \{v_1, v_2, \dots, v_n\}$ یک پایهٔ یکامتعامد برای فضای ضرب داخلی V با ضرب داخلی $\langle \cdot, \cdot \rangle$ باشد و $x \in V$ یک بردار دلخواه باشد. در این صورت، اسکالرهایی α_i ، $i = 1, 2, \dots, n$ وجود دارند به طوری که

$$x = \sum_{j=1}^n \alpha_j v_j.$$

به ازای هر $i = 1, 2, \dots, n$ داریم

$$\langle x, v_i \rangle = \left\langle \sum_{j=1}^n \alpha_j v_j, v_i \right\rangle = \sum_{j=1}^n \alpha_j \langle v_j, v_i \rangle = \alpha_i \|v_i\|^2 = \alpha_i.$$

بنابراین،

$$x = \sum_{j=1}^n \langle x, v_j \rangle v_j.$$

این بسط را بسط فوریهٔ x نامند.

مثال ۶.۱ مجموعهٔ B داده شده در رابطهٔ (۵-۱) یک پایهٔ یکامتعامد برای $\mathbb{R}^3$ است. فرض کنید $x = (-1, 2, 1)^T$ در این صورت، خواهیم داشت

$$\alpha_1 = \langle x, v_1 \rangle = -\frac{3}{\sqrt{2}}, \quad \alpha_2 = \langle x, v_2 \rangle = \frac{2}{\sqrt{3}}, \quad \alpha_3 = \langle x, v_3 \rangle = \frac{1}{\sqrt{6}}.$$

بنابراین، بسط فوریهٔ x بر حسب B به صورت زیر خواهد بود

$$x = -\frac{3}{\sqrt{2}}v_1 + \frac{2}{\sqrt{3}}v_2 + \frac{1}{\sqrt{6}}v_3.$$

تعریف ۹.۱ فرض کنید $\langle \cdot, \cdot \rangle$ یک ضرب داخلی روی فضای ضرب داخلی V باشد. اگر W یک زیرفضای V باشد، آنگاه زیرفضای مکمل متعامد W به صورت

$$W^\perp = \{v \in V : \langle v, w \rangle = 0, \forall w \in W\}.$$

تعریف می شود.

در سراسر این کتاب تمام ضرب‌های داخلی استفاده شده، استاندارد هستند، مگر اینکه ذکر گردد.

۲-۱ مسئله مقدار ویژه

این بخش را با تعریف مسئله مقدار ویژه شروع می‌کنیم.

تعریف ۱۰.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$. (λ, x) را یک زوج ویژه A می‌نامیم هرگاه $Ax = \lambda x$ و $x \neq 0$. λ را مقدار ویژه و x را بردار ویژه A می‌نامند.

اگر (λ, x) یک زوج ویژه A باشد، آنگاه $Ax = \lambda x$ ، یا به عبارت دیگر

$$(\lambda I - A)x = 0.$$

این دستگاه فقط وقتی جواب ناصفر x دارد که ماتریس $\lambda I - A$ منفرد باشد. بنابراین، باید داشته باشیم

$$p(\lambda) = \det(\lambda I - A) \neq 0.$$

می‌توان دید که p یک چندجمله‌ای از درجه n به صورت

$$p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \dots + a_1\lambda + a_0,$$

است که به چندجمله‌ای مشخصه معروف است. مجموعه مقادیر ویژه A را طیف A می‌گویند و با $\sigma(A)$ نشان می‌دهند. می‌توان دید که اگر λ_i ، $i = 1, 2, \dots, n$ مقادیر ویژه ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ باشند، آنگاه

$$\det(A) = \prod_{i=1}^n \lambda_i, \quad \text{trace}(A) = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i, \quad a_0 = (-1)^n \det(A).$$

بنابراین، به عنوان یک نتیجه، ماتریس A نامنفرد است اگر و تنها اگر همه مقادیر ویژه آن ناصفر باشند. در نرم افزار متلب، تابع eig زوج‌های ویژه یک ماتریس را به دست می‌دهد. در واقع، دستور $[X, D] = \text{eig}(A)$ ماتریس‌های $X = (x_1, x_2, \dots, x_n)$ و $D = (d_{ij})$ را برمی‌گرداند که در آن $i = 1, 2, \dots, n$ ، (d_{ii}, x_i) زوج‌های ویژه A هستند.

مثال ۷.۱ فرض کنید

$$A = \begin{pmatrix} 5 & 4 & 2 \\ 0 & 2 & -1 \\ -1 & -1 & 3 \end{pmatrix}.$$

در این صورت، دستور $[X, D] = \text{eig}(A)$ ماتریس‌های زیر را به دست می‌دهد

$$X = \begin{pmatrix} 0.7559 & 0.5774 & -0.7559 \\ 0.2336 & -0.5774 & 0.6116 \\ -0.6116 & 0.5774 & -0.2336 \end{pmatrix}, \quad D = \begin{pmatrix} 4.6180 & 0 & 0 \\ 0 & 3.0000 & 0 \\ 0 & 0 & 2.3820 \end{pmatrix}.$$

بنابراین $\lambda_1 = 4.6180$ ، $\lambda_2 = 3.0000$ و $\lambda_3 = 2.3820$ مقادیر ویژه A و ستون‌های اول، دوم و سوم ماتریس X ، به ترتیب بردارهای ویژه متناظر هستند.

تعریف ۱۱.۱ ماتریس $A \in \mathbb{C}^{n \times n}$ را با ماتریس $B \in \mathbb{C}^{n \times n}$ متشابه می‌گوییم و می‌نویسیم $A \sim B$ ، هرگاه یک ماتریس نامنفرد مثل P وجود داشته باشد به طوری که

$$P^{-1}AP = B.$$

در صورتی که P یکانی باشد، می‌گوییم ماتریس A به طور یکانی با ماتریس B متشابه است. همچنین، در صورتی که B یک ماتریس قطری باشد، می‌گوییم ماتریس A قطری‌شدنی است و به علاوه، اگر P یک ماتریس یکانی باشد، می‌گوییم ماتریس A به طور یکانی قطری‌شدنی است.

فرض کنید $A \sim B$ و $P^{-1}AP = B$. همچنین، فرض کنید (λ, x) یک زوج ویژه A باشد. در این صورت، داریم $Ax = \lambda x$. لذا،

$$P^{-1}APP^{-1}x = \lambda P^{-1}x.$$

از این رو، با فرض $y = P^{-1}x$ ، داریم

$$By = \lambda y, \quad y \neq 0.$$

لذا (λ, y) یک زوج ویژه ماتریس B است.

قضیه ۶.۱ مقادیر ویژه یک ماتریس هرمیتی یا متقارن حقیقی‌اند. همچنین مقادیر ویژه یک ماتریس هرمیتی کج یا متقارن کج به صورت βi هستند که در آن $\beta \in \mathbb{R}$ و $i = \sqrt{-1}$ واحد موهومی است. **برهان.** فرض کنید $A \in \mathbb{C}^{n \times n}$ یک ماتریس هرمیتی کج و (λ, x) یک زوج ویژه آن باشد. داریم

$$Ax = \lambda x.$$

طرفین این رابطه را در x ضرب داخلی (استاندارد) می‌کنیم:

$$\langle x, Ax \rangle = \langle x, \lambda x \rangle = \bar{\lambda} \langle x, x \rangle.$$

بنابراین،

$$\begin{aligned} \bar{\lambda} \langle x, x \rangle &= \langle x, Ax \rangle = \langle A^H x, x \rangle \\ &= \langle -Ax, x \rangle = \langle -\lambda x, x \rangle \\ &= -\lambda \langle x, x \rangle. \end{aligned}$$

از اینکه x یک بردار ویژه است، داریم $\langle x, x \rangle \neq 0$. بنابراین، $\bar{\lambda} = -\lambda$. این رابطه نشان می‌دهد که قسمت حقیقی λ برابر با صفر است. در نتیجه، مقادیر ویژه A به صورت $i\beta$ هستند که $\beta \in \mathbb{R}$. به طور مشابه می‌توان قضیه را برای ماتریس‌های هرمیتی نیز ثابت کرد. $\square$

شعاع طیفی ماتریس A را با $\rho(A)$ نشان می‌دهند و به صورت

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|,$$

تعریف می‌شود.

با توجه به نکات اشاره‌شده، نرم-۲ ماتریسی را می‌توان از رابطه

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\rho(A^H A)},$$

به دست آورد.

تعریف ۱۲.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $p(x) = a_0 + a_1 x + \dots + a_m x^m$ در این صورت، ماتریس $p(A)$ به صورت زیر تعریف می‌شود

$$p(A) = a_0 I + a_1 A + \dots + a_m A^m = \sum_{k=0}^m a_k A^k,$$

که در آن I ماتریس همانی $n \times n$ است و $A^0 = I$.

قضیه ۷.۱ فرض کنید (λ, x) یک زوج ویژه ماتریس $A \in \mathbb{C}^{n \times n}$ باشد و p و q دو چندجمله‌ای دلخواه باشند، به طوری که $p(A)$ نامنفرد است. در این صورت، زوج‌های ویژه $q(A)p(A)^{-1}$ به صورت $(q(\lambda)p(\lambda)^{-1}, x)$ هستند.

برهان. از رابطه $Ax = \lambda x$ داریم

$$A^k x = A^{k-1} Ax = \lambda A^{k-1} x = \dots = \lambda^k x, \quad k \geq 0.$$

بنابراین، اگر فرض کنیم $p(x) = a_m x^m + a_{m-1} x^{m-1} + \dots + a_1 x + a_0$ ، آنگاه خواهیم داشت

$$p(A)x = \sum_{k=0}^m a_k A^k x = \sum_{k=0}^m a_k \lambda^k x = p(\lambda)x.$$

ازاین‌رو، چون ماتریس $p(A)$ نامنفرد است، داریم $p(\lambda) \neq 0$. درنتیجه،

$$p(A)^{-1}x = \frac{1}{p(\lambda)}x.$$

سرانجام، داریم

$$q(A)p(A)^{-1}x = \frac{1}{p(\lambda)}q(A)x = \frac{1}{p(\lambda)}q(\lambda)x,$$

□ که اثبات قضیه را کامل می‌کند.

قضیه ۸.۱ (کیلی-همیلتون) [۹۳] فرض کنید p چندجمله‌ای مشخصه ماتریس $A \in \mathbb{C}^{n \times n}$ باشد. دراین‌صورت، $p(A) = 0$.

تعریف ۱۳.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$. چندجمله‌ای مینیمال A یک چندجمله‌ای تکیین مثل q با کوچک‌ترین درجه ممکن است به‌طوری‌که $q(A) = 0$.

می‌توان دید که چندجمله‌ای مینیمال یک ماتریس، چندجمله‌ای مشخصه را عادی می‌کند.

مثال ۸.۱ چندجمله‌ای مشخصه ماتریس

$$A = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix},$$

برابر است با $p(x) = (\lambda - 2)^2(\lambda - 3)$. بنابراین، طبق قضیه کیلی-همیلتون، داریم

$$p(A) = (A - 2I)^2(A - 3I) = 0.$$

همچنین، چندجمله‌ای مینیمال ماتریس A برابر است با $q(\lambda) = (\lambda - 2)(\lambda - 3)$. ازاین‌رو،

$$q(A) = (A - 2I)(A - 3I) = 0.$$

۳-۱ ماتریس‌های معین مثبت

ماتریس‌های معین مثبت نقش مهمی در مسائل کاربردی و آنالیز ماتریسی دارند. در این بخش، به معرفی این‌گونه از ماتریس‌ها و خواص آنها می‌پردازیم.

تعریف ۱۴.۱ ماتریس $A \in \mathbb{R}^{n \times n}$ را **معین مثبت** متقارن می‌گوییم، هرگاه

$$(الف) \quad A^T = A \text{ (متقارن باشد)};$$

$$(ب) \quad \text{به‌ازای هر بردار ناصفر } x \in \mathbb{R}^n, \quad x^T A x > 0.$$

ماتریس A را **نیمه‌معین مثبت** متقارن می‌گویند هرگاه (الف) برقرار باشد و برای هر $x \in \mathbb{R}^n$ ،
 $x^T A x \geq 0$.

تعریف ۱۵.۱ ماتریس $A \in \mathbb{C}^{n \times n}$ را **معین مثبت** هرمیتی می‌گوییم، هرگاه

$$(الف) \quad A^H = A \text{ (هرمیتی باشد)};$$

$$(ب) \quad \text{به‌ازای هر بردار ناصفر } x \in \mathbb{C}^n, \quad x^H A x > 0.$$

ماتریس A را **نیمه‌معین مثبت** هرمیتی می‌گویند هرگاه (الف) برقرار باشد و برای هر $x \in \mathbb{C}^n$ ،
 $x^H A x \geq 0$.

در اینجا لازم به ذکر است که اگر به‌ازای هر بردار ناصفر $x \in \mathbb{C}^n$ ، داشته باشیم $x^H A x > 0$ ، آنگاه نتیجه می‌شود که ماتریس A هرمیتی است (تمرین ۲۰.۱). بنابراین، شرط (الف) در تعریف ۱۵.۱ را می‌توان نادیده گرفت، هرچند برای تأکید معمولاً این شرط آورده می‌شود.

مثال ۹.۱ ماتریس

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

معین مثبت متقارن است. زیرا، نخست اینکه ماتریس A متقارن است. همچنین، فرض کنید x یک بردار ناصفر در $\mathbb{R}^3$ باشد. اگر $x = (x_1, x_2, x_3)^T$ ، آنگاه با کمی محاسبه خواهیم داشت

$$x^T Ax = x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 \geq 0.$$

ازطرفی، $x^T Ax \neq 0$ ، زیرا درغیراین صورت، داریم

$$x_1^2 = (x_1 - x_2)^2 = (x_2 - x_3)^2 = x_3^2 = 0,$$

که نتیجه می‌دهد $x_1 = x_2 = x_3 = 0$. بنابراین، $x = 0$ ، که با فرض $x \neq 0$ در تناقض است. $\square$

مثال ۱۰.۱ ماتریس

$$A = \begin{pmatrix} 4 & 3 + 4i \\ 3 - 4i & 25 \end{pmatrix}.$$

معین مثبت هرمیتی است. زیرا اولاً ماتریس A هرمیتی است. ثانیاً فرض کنید x یک بردار ناصفر در $\mathbb{C}^2$ باشد. با فرض $x = (x_1, x_2)^T$ داریم

$$\begin{aligned} x^H Ax &= \begin{pmatrix} \overline{x_1} & \overline{x_2} \end{pmatrix} \begin{pmatrix} 4 & 3 + 4i \\ 3 - 4i & 25 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= 4|x_1|^2 + 2\Re((3 + 4i)\overline{x_1}x_2) + 25|x_2|^2 \\ &\geq 4|x_1|^2 - 2|3 + 4i||\overline{x_1}||x_2| + 25|x_2|^2 \\ &= 4|x_1|^2 - 10|x_1||x_2| + 25|x_2|^2 \\ &= 3|x_1|^2 + (|x_1| - 5|x_2|)^2. \end{aligned}$$

لازم به ذکر است که در نوشتن تساوی سوم در روابط بالا از

$$\Re(z) \geq -|z|,$$

استفاده نموده‌ایم. رابطه بالا نشان می‌دهد که $x^H Ax \geq 0$. اگر $x^H Ax = 0$ ، آنگاه

$$3|x_1|^2 = (|x_1| - 5|x_2|)^2 = 0,$$

که نتیجه می‌دهد $x_1 = x_2 = 0$. ازاین‌رو، داریم $x = 0$ ، که یک تناقض است. بنابراین، به‌ازای هر بردار ناصفر x ، داریم $x^H Ax > 0$.

۱۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

همان‌گونه که در دو مثال پیش دیدیم، استفاده از تعریف برای نشان دادن معین مثبت بودن یک ماتریس کاری دشوار است، به‌ویژه وقتی که ابعاد ماتریس بزرگ باشند. در ادامه، روش‌های کاراتری برای این کار معرفی می‌کنیم.

تعریف ۱۶.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $1 \leq i_1 < i_2 < \dots < i_k \leq n$. در این صورت، ماتریس

$$A_k = \begin{pmatrix} a_{i_1 i_1} & a_{i_1 i_2} & \dots & a_{i_1 i_k} \\ a_{i_2 i_1} & a_{i_2 i_2} & \dots & a_{i_2 i_k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_k i_1} & a_{i_k i_2} & \dots & a_{i_k i_k} \end{pmatrix},$$

یک زیرماتریس اصلی A نامیده می‌شود. در صورتی که $i_j = j$ ، $j = 1, 2, \dots, k$ ، A_k یک زیرماتریس اصلی پیشرو ماتریس A نامیده می‌شود.

قضیه ۹.۱ [۹۳] فرض کنید ماتریس A معین مثبت هرمیتی باشد. در این صورت،

(الف) A معکوس‌پذیر است و A^{-1} نیز معین مثبت هرمیتی می‌باشد.

(ب) دترمینان ماتریس A مثبت است.

قضیه ۱۰.۱ [۹۳] فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ هرمیتی باشد. ماتریس A معین مثبت هرمیتی است اگر و تنها اگر دترمینان همهٔ زیرماتریس‌های اصلی پیشروی آن مثبت باشند.

قضیه ۱۱.۱ برای هر ماتریس هرمیتی $A \in \mathbb{C}^{n \times n}$ ، داریم

$$\min_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \lambda_{\min}, \quad \max_{x \neq 0} \frac{\langle Ax, x \rangle}{\langle x, x \rangle} = \lambda_{\max},$$

که در آن $\lambda_{\min}$ و $\lambda_{\max}$ به ترتیب بزرگ‌ترین و کوچک‌ترین مقدار ویژهٔ ماتریس A هستند.

قضیه ۱۲.۱ [۹۳] فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ هرمیتی باشد. ماتریس A معین مثبت هرمیتی است اگر و تنها اگر مقادیر ویژهٔ آن مثبت باشند.

قضیه ۱۳.۱ [۹۳] ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت هرمیتی است اگر و تنها اگر یک ماتریس پائین‌مثلثی یکتا با درایه‌های قطری مثبت مثل L وجود دارد به‌طوری‌که $A = LL^H$.

تجزیه ماتریس A به صورت $A = LL^H$ ، که در آن L یک ماتریس پائین‌مثلثی با درایه‌های قطری مثبت است، به تجزیه چولسکی

معروف است. برای ماتریس معین مثبت هرمیتی A ، دستور $L = \text{chol}(A, 'lower')$ در متلب، عامل L در تجزیه چولسکی A را محاسبه می‌کند.

مثال ۱۱.۱ ماتریس

$$A = \begin{pmatrix} 4 & -2 & 6 \\ -2 & 2 & 1 \\ 6 & 1 & 5 \end{pmatrix},$$

معین مثبت متقارن است (چرا؟). دستور $L = \text{chol}(A, 'lower')$ عامل L در تجزیه چولسکی را به صورت زیر محاسبه می‌کند

$$L = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & 4 & 5 \end{pmatrix}.$$

همه نتایج را که تاکنون برای ماتریس‌های معین مثبت هرمیتی نوشته شده‌اند، برای هر ماتریس معین مثبت متقارن نیز برقرار هستند. کافیست در نتایج ارائه شده به جای “هرمیتی” کلمه “متقارن” جایگزین شود.

مثال ۱۲.۱ ماتریس متقارن

$$A = \begin{pmatrix} 4 & -2 & 4 \\ -2 & 5 & 4 \\ 4 & 4 & 14 \end{pmatrix},$$

را در نظر بگیرید. زیرماتریس‌های اصلی پیشروی ماتریس A عبارتند از

$$A_1 = (4), \quad A_2 = \begin{pmatrix} 4 & -2 \\ -2 & 5 \end{pmatrix}, \quad A_3 = A.$$

داریم $\det(A_1) = 4 > 0$ ، $\det(A_2) = 16 > 0$ و $\det(A_3) = 16 > 0$. لذا، با استناد به قضیه ۱۰.۱، ماتریس A معین مثبت متقارن است.

۲۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

طیف ماتریس A برابر با $\sigma(A) = \{0.15, 6.53, 16.32\}$ است. حال با توجه به اینکه مقادیر ویژه A همگی مثبت هستند، با استناد به قضیه ۱۲.۱ نتیجه می‌گیریم که این ماتریس معین مثبت متقارن است. همچنین، اگر قرار دهیم

$$L = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 2 & 3 & 1 \end{pmatrix},$$

آنگاه خواهیم داشت $A = LL^T$. از این رو، بنا به قضیه ۱۳.۱ ماتریس A معین مثبت متقارن است.

قضیه ۱۴.۱ اگر ماتریس A معین مثبت هرمیتی باشد، آنگاه

$$\langle x, y \rangle_A = \langle Ax, y \rangle = y^H Ax,$$

یک ضرب داخلی روی $\mathbb{C}^n$ تعریف می‌کند و نرم تولیدشده توسط ضرب داخلی $\langle x, y \rangle_A$ به صورت $\|x\|_A = \sqrt{x^H Ax}$ است.

تعریف ۱۷.۱ ماتریس $A \in \mathbb{C}^{n \times n}$ را معین مثبت می‌گوییم هرگاه به ازای هر $x \in \mathbb{C}^n$ $x \neq 0$ قسمت حقیقی $x^H Ax$ مثبت باشد، یعنی $\Re(x^H Ax) > 0$.

قضیه ۱۵.۱ ماتریس A معین مثبت است اگر و تنها اگر $A + A^H$ معین مثبت هرمیتی باشد. برهان. به ازای هر $x \in \mathbb{C}^{n \times n}$ داریم

$$\begin{aligned} x^H(A + A^H)x &= x^H Ax + x^H A^H x = x^H Ax + (x^H Ax)^H \\ &= x^H Ax + \overline{x^H Ax} = 2\Re(x^H Ax). \end{aligned}$$

لذا،

$$x^H(A + A^H)x > 0 \iff \Re(x^H Ax) > 0,$$

□

که اثبات قضیه را کامل می‌کند.

مثال ۱۳.۱ ماتریس

$$A = \begin{pmatrix} 4 & 2+2i & 2i \\ 2 & 6 & 2+1i \\ i & 1-i & 2 \end{pmatrix},$$

هرمیتی نیست اما معین مثبت است؛ زیرا ماتریس

$$A + A^H = \begin{pmatrix} 8 & 4+2i & i \\ 4-2i & 12 & 3+2i \\ -i & 3-2i & 4 \end{pmatrix},$$

معین مثبت هرمیتی است.

قضیه ۱۶.۱ (شکل شور) [۹۳] فرض کنید $A \in \mathbb{C}^{n \times n}$. در این صورت، یک ماتریس یکانی مثل U وجود دارد به طوری که $U^H A U = T$ ، که در آن T یک ماتریس بالامثلثی است. درایه‌های روی قطر T مقادیر ویژه A هستند.

فرض کنید $A \in \mathbb{R}^{n \times n}$. اگر ماتریس A چند مقدار ویژه مختلط داشته باشد، آنگاه برخی از درایه‌های قطری T مختلط خواهند بود. به علاوه، ماتریس U نیز ممکن است مختلط باشد. اما در حالت کلی یک ماتریس متعامد مثل U و یک ماتریس شبه بالامثلثی (حقیقی) مثل T می‌توان یافت به طوری که

$$U^T A U = T = \begin{pmatrix} T_{11} & * & \cdots & * \\ & T_{22} & & \vdots \\ & & \ddots & * \\ & & & T_{rr} \end{pmatrix},$$

است، که در آن $i = 1, 2, \dots, r$ ، S_{ii} یا 1×1 یا 2×2 هستند. در اینجا S را شکل شور حقیقی ماتریس A می‌گویند [۴۳].

در متلب دستور `[U,T]=schur(A,flag)` تجزیه شور یک ماتریس را به دست می‌دهد. اگر `flag='real'`، آنگاه این دستور شکل شور حقیقی را برمی‌گرداند و اگر `flag='complex'`، آنگاه شکل شور مختلط (U و T ممکن است مختلط باشد، اما T بالامثلثی است) را محاسبه می‌کند.

مثال ۱۴.۱ اگر فرض کنیم

$$A = \begin{pmatrix} 2 & -1 & 2 & 4 \\ -1 & 1 & 4 & 2 \\ 3 & 2 & 3 & 2 \\ 2 & -1 & 2 & 1 \end{pmatrix},$$

آنگاه دستور $[U, T] = \text{schur}(A, 'real')$ ماتریس‌های

$$U = \begin{pmatrix} -0.4145 & 0.6675 & 0.2000 & -0.5853 \\ -0.4974 & -0.6121 & 0.5983 & -0.1414 \\ -0.7047 & -0.1271 & -0.6878 & 0.1190 \\ -0.2902 & 0.4045 & 0.3591 & 0.7895 \end{pmatrix},$$

$$T = \begin{pmatrix} 7.0000 & -2.3435 & 0.9046 & -2.1512 \\ 0 & 3.5043 & 0.3602 & 0.0626 \\ 0 & 0 & -1.7522 & 2.0384 \\ 0 & 0 & -0.5938 & -1.7522 \end{pmatrix},$$

را به دست می‌دهد.

همچنین، اگر قرار دهیم

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -2 & 3 & 4 \\ 2 & 1 & 1 \end{pmatrix},$$

آنگاه دستور $[U, T] = \text{schur}(A, 'complex')$ ماتریس‌های زیر را محاسبه می‌کند

$$U = \begin{pmatrix} -0.1102 + 0.1811i & -0.0205 + 0.9733i & 0.0856 + 0.0000i \\ -0.0100 - 0.7676i & 0.0869 + 0.0882i & 0.6288 + 0.0000i \\ 0.0203 + 0.6044i & -0.0685 - 0.1796i & 0.7728 + 0.0000i \end{pmatrix},$$

$$T = \begin{pmatrix} 0.3174 + 0.3583i & 3.1230 + 0.0000i & -0.1800 + 2.4693i \\ 0.0000 + 0.0000i & 0.3174 - 0.3583i & 0.2796 - 1.5892i \\ 0.0000 + 0.0000i & 0.0000 + 0.0000i & 4.3652 + 0.0000i \end{pmatrix}.$$

قضیه ۱۷.۱ [۹۳] ماتریس $A \in \mathbb{C}^{n \times n}$ به‌طور یکانی قطری‌شدنی است اگر و تنها اگر نرمال باشد.

نتیجه ۱۰.۱ ماتریس‌های هرمیتی و هرمیتی‌کج نرمال هستند. لذا، این‌گونه ماتریس‌ها به‌طور یکانی قطری‌شدنی هستند، به این معنا که ماتریس T در قضیه ۱۶.۱ قطری می‌باشد.

قضیه ۱۸.۱ اگر $\|\cdot\|$ یک نرم طبیعی روی $\mathbb{C}^{n \times n}$ باشد، آنگاه $\rho(A) \leq \|A\|$. به‌علاوه، اگر ماتریس A نرمال باشد، آنگاه $\rho(A) = \|A\|_2$.

برهان. فرض کنید (λ, x) یک زوج ویژه A باشد. در این صورت، داریم

$$|\lambda||x| = \|\lambda x\| = \|Ax\| \leq \|A\|\|x\|.$$

این رابطه نشان می‌دهد که به‌ازای هر مقدار ویژه λ ، $\lambda \leq \|A\|$ که قسمت اول قضیه را نتیجه می‌دهد. اکنون، فرض کنید ماتریس A نرمال باشد. در این صورت، با توجه به قضیه ۱۷.۱ یک ماتریس یکانی مثل U وجود دارد به‌طوری‌که $A = UDU^H$ ، که در آن D یک ماتریس قطری است که درایه‌های قطری آن مقادیر ویژه A هستند. بنابراین، داریم

$$\|A\|_2 = \|UDU^H\|_2 = \|D\|_2 = \rho(D) = \rho(A),$$

□

و به این ترتیب اثبات قضیه کامل می‌شود.

نتیجه ۲۰.۱ ماتریس‌های هرمیتی، هرمیتی‌کج و یکانی همگی نرمال هستند. بنابراین، همه آنها به‌طور یکانی قطری‌شدنی هستند. از این رو، اگر A یکی از این ماتریس‌ها باشد، آنگاه $\rho(A) = \|A\|_2$.

قضیه ۱۹.۱ اگر ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت هرمیتی باشد، آنگاه یک ماتریس معین مثبت هرمیتی مثل B وجود دارد به‌طوری‌که $A = B^2$.

برهان. طبق نتیجه ۱۰.۱ یک ماتریس یکانی مثل P وجود دارد به‌طوری‌که

$$P^{-1}AP = P^HAP = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \ddots \\ & & & \lambda_n \end{pmatrix} = D,$$

که در آن $\lambda_i > 0, i = 1, 2, \dots, n$. قرار می‌دهیم

$$D^{\frac{1}{2}} = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \ddots \\ & & & \sqrt{\lambda_n} \end{pmatrix},$$

و $B = PD^{\frac{1}{2}}P^{-1}$. در این صورت، به‌سادگی می‌توان دید که ماتریس B معین مثبت هرمیتی است. از طرفی، داریم

$$B^2 = PD^{\frac{1}{2}}P^{-1}PD^{\frac{1}{2}}P^{-1} = PD^{\frac{1}{2}}D^{\frac{1}{2}}P^{-1} = PDP^{-1} = A.$$

□

به این ترتیب اثبات قضیه کامل می‌شود.

در قضیه ۱۹.۱ ماتریس B را معمولاً با $A^{\frac{1}{2}}$ نشان داده و آن را ریشه دوم ماتریس A می‌گویند. در نرم‌افزار متلب، تابع `sqrtm` ریشه دوم یک ماتریس را محاسبه می‌کند.

مثال ۱۵.۱ اگر فرض کنیم

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix},$$

آنگاه دستور `B=sqrtm(A)` ماتریس

$$B = \begin{pmatrix} ۱.۷۰۱۳ & ۰.۳۲۴۹ \\ ۰.۳۲۴۹ & ۱.۳۷۶۴ \end{pmatrix},$$

را به دست می‌دهد و داریم $B^2 = A$.

قضیه ۲۰.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ یک ماتریس نیمه‌معین مثبت متقارن باشد و $x \in \mathbb{C}^n$. اگر $Ax = 0$ ، آنگاه $x^H Ax = 0$.

برهان. از اینکه ماتریس A هرمیتی است از نتیجه ۱.۱ درمی‌یابیم که یک ماتریس یکانی مثل $U \in \mathbb{C}^{n \times n}$ وجود دارد به‌طوری‌که $A = UDU^H$ و $D = \text{diag}_n(\lambda_1, \lambda_2, \dots, \lambda_n)$. بنابراین،

$$0 = x^H Ax = x^H UDU^H x = y^H Dy = \sum_{i=1}^n \lambda_i |y_i|^2,$$

که در آن، $y = U^H x$. در نتیجه، $y = U^H x$ ، در نتیجه، $\lambda_i |y_i|^2 = 0$ ، $i = 1, 2, \dots, n$. از این رو، نتیجه می‌شود که به‌ازای هر i ، $y_i = 0$ یا $\lambda_i = 0$ ، که معادل است با $Dy = 0$. سرانجام، داریم

$$Ax = UDU^H x = y^H Dy = 0,$$

اثبات قضیه را کامل می‌کند. $\square$

قضیه ۲۱.۱ (شکل کانونی جردن) [۹۳، ۷۷] برای هر ماتریس $A \in \mathbb{C}^{n \times n}$ با مقادیر ویژه متمایز $\lambda_1, \lambda_2, \dots, \lambda_s$ ، یک ماتریس نامنفرد مثل $P \in \mathbb{C}^{n \times n}$ وجود دارد به‌طوری‌که

$$P^{-1}AP = J = \begin{pmatrix} J(\lambda_1) & 0 & \dots & 0 \\ 0 & J(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J(\lambda_s) \end{pmatrix},$$

و

$$J(\lambda_j) = \begin{pmatrix} J_1(\lambda_j) & 0 & \dots & 0 \\ 0 & J_2(\lambda_j) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{t_j}(\lambda_j) \end{pmatrix}, \quad J_r(\lambda_j) = \begin{pmatrix} \lambda_j & 1 & & \\ & \lambda_j & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{pmatrix},$$

ماتریس J ، شکل جردن ماتریس A و $J_r(\lambda_j)$ بلوک‌های جردن نامیده می‌شوند.

در نرم‌افزار متلب، تابع jordan شکل جردن یک ماتریس را به‌دست می‌دهد. درحقیقت، دستور

$[P, J] = \text{jordan}(A)$ ماتریس‌های مربعی P و J را به‌دست می‌دهد به‌طوری‌که $P^{-1}AP = J$ و

شکل کانونی جردن ماتریس A است.

مثال ۱۶.۱ فرض کنید

$$A = \begin{pmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{pmatrix}.$$

در این صورت، دستور $[P, J] = \text{jordan}(A)$ در متلب ماتریس‌های

$$P = \begin{pmatrix} 1 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & J_3 \end{pmatrix},$$

را به دست می‌دهد، که در آن

$$J_1 = \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix}, \quad J_2 = (1), \quad J_3 = (2).$$

قضیه ۲۲.۱ (دیسک‌های گرشگورین) [۹۳، ۷۷] فرض کنید $A \in \mathbb{C}^{n \times n}$.

(الف) مقادیر ویژه ماتریس A در اجتماع n دیسک گرشگورین زیر قرار دارند

$$C_r : |z - a_{ii}| \leq r_i, \quad r_i = \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

(ب) مقادیر ویژه ماتریس A در اجتماع دیسک‌های زیر قرار دارند

$$C_c : |z - a_{jj}| \leq c_j, \quad c_j = \sum_{i=1, i \neq j}^n |a_{ij}|, \quad j = 1, 2, \dots, n.$$

(ج) مقادیر ویژه A در $C_r \cap C_c$ قرار دارند.

۴-۱ تجزیه QR

فرض کنید $A \in \mathbb{C}^{m \times n}$ ، به طوری که $m \geq n$ و $\text{rank}(A) = n$. به سادگی می‌توان نتایج به دست آمده را برای ماتریس‌های دلخواه نیز تعمیم داد. تجزیه QR ماتریس A یک تجزیه به صورت $A = QR$ است

که در آن $Q \in \mathbb{C}^{m \times m}$ یک ماتریس یکانی ($Q^H Q = Q Q^H = I$) و ماتریس $R \in \mathbb{C}^{m \times n}$ یک ماتریس بالامثلثی یا شبه‌بالامثلثی است. درحقیقت، در حالتی که ماتریس A مربعی است ($m = n$) ماتریس شبه‌بالامثلثی R بالامثلثی خواهد بود، و اگر $m > n$ ، آنگاه ماتریس شبه‌بالامثلثی R به صورت زیر خواهد بود

$$R = \begin{pmatrix} \tilde{R} \\ 0 \end{pmatrix}, \quad (۶-۱)$$

که در آن $\tilde{R} \in \mathbb{C}^{n \times n}$ یک ماتریس بالامثلثی است و 0 ماتریس صفر با بعد $n \times (m - n)$ است. تجزیه QR یک ماتریس کاربردهای مختلفی دارد، از جمله می‌توان به حل دستگاه معادلات خطی، حل مسئله کمترین توان‌های دوم و محاسبه زوج‌های ویژه یک ماتریس اشاره کرد. روش‌های مختلفی برای محاسبه تجزیه QR یک ماتریس وجود دارد: با استفاده از الگوریتم گرام-اشمیت، با کمک ماتریس‌های هوسهولدر، و با کمک ماتریس‌های گایوز. برای جزئیات بیشتر درباره شیوه‌های محاسبه تجزیه QR به [۱] مراجعه کنید.

تابع qr در متلب تجزیه QR یک ماتریس را محاسبه می‌کند. درحقیقت، با کمک دستور $[Q, R] = qr(A)$ می‌توان عامل‌های Q و R در تجزیه QR ماتریس A را به دست آورد.

مثال ۱۷.۱ با استفاده از دستور $[Q, R] = qr(A)$ در متلب، عامل‌های تجزیه QR ماتریس

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ -1 & 1 \end{pmatrix},$$

به صورت زیر محاسبه می‌شوند

$$Q = \begin{pmatrix} -0.5774 & 0.6172 & -0.5345 \\ -0.5774 & 0.1543 & 0.8018 \\ 0.5774 & 0.7715 & 0.2673 \end{pmatrix}, \quad R = \begin{pmatrix} -1.7321 & -1.1547 \\ 0 & 2.1602 \\ 0 & 0 \end{pmatrix}.$$

۵-۱ تجزیه مقدار تکین

فرض کنید $A \in \mathbb{C}^{m \times n}$. قرار می‌دهیم $B = A^H A$. در این صورت، ماتریس B نیمه‌معین مثبت هرمیتی است، زیرا به ازای هر $x \in \mathbb{C}^n$ داریم

$$x^H B x = x^H A^H A x = (Ax)^H A x = \|Ax\|_2^2 \geq 0.$$

بنابراین، مقادیر ویژه آن نامنفی‌اند، چون اگر (λ, x) یک زوج ویژه B باشد، آنگاه داریم $Bx = \lambda x$ ،

$$\lambda x^H x = x^H B x = x^H A^H A x = (Ax)^H (Ax) = \|Ax\|_2^2.$$

از این رو،

$$\lambda = \frac{\|Ax\|_2^2}{\|x\|_2^2} \geq 0.$$

فرض کنید $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ مقادیر ویژه B باشند. قرار می‌دهیم

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, 2, \dots, n.$$

در این صورت، σ_i ها را مقادیر تکین A می‌گویند.

قضیه ۲۳.۱ (تجزیه مقدار تکین) [۹۳، ۲۲] فرض کنید $A \in \mathbb{C}^{m \times n}$. در این صورت، ماتریس‌های یکانی $U \in \mathbb{C}^{m \times m}$ و $V \in \mathbb{C}^{n \times n}$ و ماتریس

$$\Sigma = \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}, \quad D_r = \text{diag}_r(\sigma_1, \dots, \sigma_r),$$

وجود دارند به طوری که $A = U \Sigma V^H$ ، که در آن $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. σ_i ها مقادیر تکین ماتریس A ، ستون‌های $U = (u_1, \dots, u_m)$ را بردارهای تکین چپ A و ستون‌های ماتریس $V = (v_1, \dots, v_n)$ را بردارهای تکین راست A می‌نامند. به علاوه، اگر σ_r کوچک‌ترین مقدار تکین ناصفر A باشد، آنگاه $\text{rank}(A) = r$.

نتیجه ۳.۱ [۹۳، ۲۲] فرض کنید $A = U \Sigma V^H$ تجزیه مقدار تکین ماتریس A و U ، V و Σ ماتریس‌هایی باشند که در قضیه ۲۳.۱ تعریف شده‌اند. در این صورت،

(الف) مجموعه بردارهای $\{u_1, u_2, \dots, u_r\}$ یک پایه یکامتعامل برای بُرد ماتریس A است.

(ب) مجموعه بردارهای $\{v_{r+1}, v_{r+2}, \dots, v_n\}$ یک پایه یکامتعامل برای فضای پوچ A است.

در نرم‌افزار متلب، تابع svd تجزیه مقدار تکین یک ماتریس را به‌دست می‌دهد. درحقیقت، دستوری مثل $[U, S, V] = \text{svd}(A)$ ماتریس‌های متعامد U و V و همچنین ماتریس شبه‌قطری $S = \Sigma$ را برمی‌گرداند. نکته قابل ذکر این است که مقادیر تکین ماتریس، به‌ترتیب از بزرگ‌تر به کوچک‌تر روی قطر اصلی Σ ظاهر می‌شوند.

مثال ۱۸.۱ فرض کنید

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

دراین صورت، نرم‌افزار متلب، ماتریس‌های زیر را به‌دست می‌دهد

$$U = \begin{pmatrix} -0.8165 & 0.0000 & -0.5774 \\ -0.4082 & -0.7071 & 0.5774 \\ -0.4082 & 0.7071 & 0.5774 \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} 1.7321 & 0 \\ 0 & 1.0000 \\ 0 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} -0.7071 & -0.7071 \\ -0.7071 & 0.7071 \end{pmatrix}.$$

تعریف ۱۸.۱ فرض کنید $A = U\Sigma V^H$ تجزیه مقدار تکین ماتریس A باشد و $\text{rank}(A) = r$. دراین صورت، معکوس تعمیم‌یافته مور-پنرز ماتریس A را با A^+ نشان می‌دهند و به‌صورت

$$A^+ = V\Sigma^+U^H, \quad (7-1)$$

تعریف می‌شود، که در آن

$$\Sigma^+ = \begin{pmatrix} D_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n \times m}.$$

نتیجه ۴.۱ اگر ماتریس $A \in \mathbb{C}^{n \times n}$ نامنفرد باشد، آنگاه $A^+ = A^{-1}$.

۳۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

در نرم‌افزار متلب تابع pinv معکوس تعمیم‌یافته مور-پنرز را به‌دست می‌دهد.

مثال ۱۹.۱ فرض کنید

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$

در این صورت، دستور $P = \text{pinv}(A)$ در متلب، معکوس تعمیم‌یافته مور-پنرز ماتریس A را به‌صورت زیر به‌دست می‌دهد

$$P = \begin{pmatrix} -0.1667 & 0.8333 & 0.3333 \\ 0.5000 & -0.5000 & -0.5000 \end{pmatrix}.$$

تعریف ۱۹.۱ فرض کنید $A \in \mathbb{C}^{m \times n}$. در این صورت، عدد حالت ماتریس A را با $\text{cond}(A)$ نشان می‌دهند به‌صورت

$$\text{cond}(A) = \|A\| \|A^+\|,$$

تعریف می‌شود، که در آن $\|\cdot\|$ یک نرم ماتریسی است. در صورتی‌که عدد حالت یک ماتریس بزرگ باشد آن ماتریس را بدحالت، و در غیر این صورت خوش‌حالت می‌گویند. اینکه عدد حالت چقدر بزرگ باشد تا یک ماتریس بدحالت در نظر گرفته شود به دقت کامپیوتری که محاسبات با آن انجام می‌شود مربوط می‌شود. برای جزئیات بیشتر دربارهٔ این موضوع، بخش ۳.۵ از مرجع [۱] را ببینید.

فرض کنید $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ مقادیر تکین ناصفر ماتریس $A \in \mathbb{C}^{m \times n}$ باشند. در این صورت،

داریم

$$\|A\|_2 = \sqrt{\rho(A^H A)} = \sigma_1,$$

و

$$\begin{aligned} \|A^+\|_2 &= \sqrt{\rho((A^+)^H A^+)} \\ &= \sqrt{\rho(U \Sigma^+{}^H \Sigma^+ U^H)} \\ &= \sqrt{\rho(\Sigma^+{}^H \Sigma^+)} \end{aligned}$$

$$= \sqrt{\rho \left(\text{diag}_n \left(\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0 \right) \right)}$$

$$= \frac{1}{\sigma_r}.$$

بنابراین،

$$\text{cond}_2(A) = \|A\|_2 \|A^+\|_2 = \frac{\sigma_1}{\sigma_r} \geq 1, \quad (8-1)$$

که در آن σ_1 و σ_r به ترتیب بزرگ‌ترین و کوچک‌ترین مقدار تکین ناصفر ماتریس A هستند.

یکی از خواص مهم عدد حالت این است که به ازای هر $\alpha \neq 0$ ، داریم

$$\text{cond}(\alpha A) = \text{cond}(A).$$

حل مسائل بدحالت یکی از چالش‌های مهم در آنالیز عددی است. در مسائلی که بدحالت هستند، یک تغییر ناچیز در داده‌های مسئله منجر به یک تغییر اساسی در جواب مسئله می‌شود. این موضوع از این جهت حائز اهمیت است که برای حل یک مسئله توسط رایانه، به محض ذخیره‌سازی آن، همه داده‌ها به صورت تقریبی ذخیره می‌شوند (به دلیل محدودیت در ذخیره‌سازی ارقام). از این رو، زمانی که مسئله بدحالت است، این تقریب‌ها ممکن است روی جواب‌های واقعی تأثیر زیادی بگذارند و یک جواب نامعقول به دست آید. مثال زیر یک نمونه از مسائل بدحالت را نشان می‌دهد.

مثال ۲۰.۱ [۱۰] دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} 1000 & 999 \\ 999 & 998 \end{pmatrix}, \quad b = \begin{pmatrix} 1999 \\ 1997 \end{pmatrix}.$$

جواب یکتای این دستگاه بردار $x^* = (1, 1)^T$ است. داریم

$$A^{-1} = \begin{pmatrix} -998 & 999 \\ 999 & -1000 \end{pmatrix}.$$

همچنین، داریم

$$\text{cond}_1(A) = \|A\|_1 \|A^{-1}\|_1 = 1999^2 \approx 3.996 \times 10^6,$$

و

$$\text{cond}_{\infty}(A) = \|A\|_{\infty} \|A^{-1}\|_{\infty} = 1999^2 \approx 3/996 \times 10^6,$$

که نشان می‌دهد این مسئله بدحالت است. برای مثال، اگر تغییری به صورت

$$\Delta b = \begin{pmatrix} -0.01 \\ 0.01 \end{pmatrix},$$

در بردار سمت راست ایجاد کرده و دستگاه $Ax = b + \Delta b$ را حل کنیم، جواب واقعی دستگاه به صورت $\hat{x} = (2097, 1899)^T$ خواهد بود. می‌بینیم که یک تغییر کوچک در سمت راست دستگاه $Ax = b$ باعث شده است که تغییرات زیادی در جواب دستگاه ایجاد شود. بنابراین، دستگاه $Ax = b$ یک نمونه از دستگاه‌های بدحالت است.

۱-۶ مسئله کمترین توان‌های دوم و دستگاه معادلات نرمال

فرض کنید $A \in \mathbb{R}^{m \times n}$ به طوری که $\text{rank}(A) = r$. همه نتایج این بخش را می‌توان برای ماتریس‌های مختلط نیز نوشت. کافیت به جای ترانواده، ترانوادهٔ هرمیتی را جایگزین کنید. اگر $A = U\Sigma V^T$ تجزیه مقدار تکیین ماتریس A باشد، آنگاه می‌توان این تجزیه را به صورت

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T, \quad (9-1)$$

نوشت که در آن u_i و v_i به ترتیب ستون i ام ماتریس‌های U و V هستند. این نمایش از SVD را نمایش اقتصادی SVD ماتریس A می‌گویند. به سادگی می‌توان دید که

$$A^+ = \sum_{i=1}^r \frac{1}{\sigma_i} v_i u_i^T.$$

لم ۱.۱ اگر $A \in \mathbb{R}^{m \times n}$ و $b \in \mathbb{R}^n$ ، آنگاه $A^T b \in \mathcal{R}(A^T A)$.

برهان. فرض کنید $A = U\Sigma V^T$ تجزیه مقدار تکیین A باشد و $y \in \mathcal{R}(A^T A)$ در این صورت، یک $x \in \mathbb{R}^n$ وجود دارد به طوری که $y = A^T A x$ در نتیجه،

$$y = V\Sigma^T U^T U\Sigma V^T x = V\Sigma^T \Sigma V^T x = \sum_{k=1}^r \sigma_k^2 (v_k^T x) v_k.$$

این رابطه نشان می‌دهد که

$$\mathcal{R}(A^T A) = \text{span}\{v_1, v_2, \dots, v_r\}.$$

ازطرفی، $\{v_1, v_2, \dots, v_r\}$ یک مجموعهٔ یکامتعامد است. بنابراین، $\{v_1, v_2, \dots, v_r\}$ یک پایه برای $\mathcal{R}(A^T A)$ است. ازاین‌رو،

$$A^T b = V \Sigma^T U^T b = \sum_{k=1}^r \sigma_k (u_k^T b) v_k \in \mathcal{R}(A^T A),$$

□

که اثبات لم را کامل می‌کند.

قضیهٔ ۲۴.۱ فرض کنید $A \in \mathbb{R}^{m \times n}$. دراین‌صورت،

$$\|Ax^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|Ax - b\|_2, \quad (۱۰-۱)$$

اگر و تنها اگر x^* یک جواب دستگاه معادلات نرمال

$$A^T A x = A^T b, \quad (۱۱-۱)$$

باشد.

برهان. نخست توجه کنید که دستگاه معادلات نرمال (۱۱-۱) همیشه دارای جواب است، زیرا با استفاده

از لم ۱۰.۱ داریم $A^T b \in \mathcal{R}(A^T A)$.

برای اثبات قضیه، فرض کنید x^* جواب دستگاه معادلات نرمال باشد. قرار می‌دهیم

$$f(x) = \|Ax - b\|_2^2.$$

دراین‌صورت، به‌ازای هر $\delta \in \mathbb{R}^n$ داریم

$$\begin{aligned} f(x^* + \delta) &= \|A(x^* + \delta) - b\|_2^2 \\ &= (Ax^* - b + A\delta)^T (Ax^* - b + A\delta) \\ &= \|Ax^* - b\|_2^2 + 2(A\delta)^T (Ax^* - b) + \|A\delta\|_2^2. \end{aligned}$$

درنتیجه، از اینکه $A^T A x^* = A^T b$ داریم

$$f(x^* + \delta) - f(x^*) = 2\delta^T (A^T A x^* - A^T b) + \|A\delta\|_2^2 = \|A\delta\|_2^2 \geq 0.$$

این رابطه نشان می‌دهد که x^* تابع f را کمینه می‌کند.

اکنون، فرض کنید x^* یک جواب مسئله (۱۰-۱) باشد. در این صورت، داریم

$$f(x^* + \delta) - f(x^*) = 2\delta^T A^T r + \|A\delta\|^2 \geq 0, \quad \forall \delta \in \mathbb{R}^n.$$

که در آن $r = Ax^* - b$. به‌برهان خلف فرض کنید $A^T r \neq 0$. قرار می‌دهیم $\delta = -\lambda A^T r$ ، که در آن $\lambda > 0$. در نتیجه،

$$f(x^* + \delta) - f(x^*) = -2\lambda \|A^T r\|^2 + \lambda^2 \|AA^T r\|^2 \geq 0,$$

که معادل است با

$$0 \leq 2\|A^T r\|^2 \leq \lambda \|AA^T r\|^2.$$

داریم

$$\lim_{\lambda \rightarrow 0^+} \lambda \|AA^T r\|^2 = 0 \implies 2\|A^T r\|^2 = 0 \implies A^T r = 0,$$

□

که اثبات قضیه را کامل می‌کند.

هنگامی که ماتریس A یک ماتریس مستطیلی با $m > n$ باشد، برای حل دستگاه فرامعین $Ax = b$ از حل مسئله کمترین توان‌های دوم (۱۰-۱)، یا دستگاه معادلات نرمال (۱۱-۱) استفاده می‌شود.

نتیجه ۵.۱ فرض کنید $A \in \mathbb{R}^{m \times n}$ ، $m \geq n$ و $\text{rank}(A) = n$. در این صورت، جواب مسئله (۱۰-۱) یکتا و به‌صورت $x^* = (A^T A)^{-1} A^T b$ است.

برهان. روشن است که ماتریس $A^T A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن است. بنابراین، با توجه به قضیه ۲۴.۱ دستگاه معادلات نرمال یک جواب یکتا دارد که به‌صورت

$$x^* = (A^T A)^{-1} A^T b,$$

□

است.

اگر شرایط نتیجه ۵.۱ برقرار باشد، آنگاه برای محاسبه x^* کفایت دستگاه $A^T Ax = A^T b$ حل شود. برای حل این دستگاه، می‌توان از تجزیه QR ماتریس A استفاده کرد. فرض کنید $A = QR$ تجزیه

QR ماتریس A باشد. در این صورت، این دستگاه به صورت

$$Rx = Q^T b,$$

نوشته می‌شود که به سادگی با جایگذاری پیشرو قابل حل است (بخش ۷-۱ را ببینید). اکنون دستگاه $Ax = b$ را در نظر بگیرید که در آن $A \in \mathbb{R}^{m \times n}$ ، $m < n$ و $\text{rank}(A) = m$. این دستگاه را دستگاه فرومعی می‌گویند. روشن است که این دستگاه بی‌شمار جواب دارد. با فرض

$$x = A^T u,$$

که در آن $u \in \mathbb{R}^n$ ، می‌توان دستگاه $Ax = b$ را به صورت

$$AA^T u = b, \quad (12-1)$$

نوشت. روشن است که ماتریس AA^T یک ماتریس معین مثبت متقارن است و بنابراین، این دستگاه یک جواب یکتا دارد. فرض کنید x^* یک جواب دلخواه برای دستگاه $Ax = b$ باشد. در این صورت، $\tilde{u}$ جواب یکتای دستگاه (۱۲-۱) است اگر و تنها اگر

$$\|A^T \tilde{u} - x^*\|_2 = \min_{u \in \mathbb{R}^m} \|A^T u - x^*\|_2.$$

در حقیقت A^T یک ماتریس با رتبه ستونی کامل است و بنابراین، با استناد به قضیه ۲۴.۱ رابطه بالا برقرار است اگر و تنها اگر

$$(A^T)^T A^T \tilde{u} = (A^T)^T x^* \implies AA^T \tilde{u} = Ax^* = b,$$

که همان دستگاه (۱۲-۱) به ازای $\tilde{u}$ است. درحقیقت، $\tilde{x} = A^T \tilde{u}$ برداری است که در $\mathcal{R}(A^T)$ قرار دارد و کمترین فاصله (تحت نرم-۲) تا بردار x^* را دارد. توجه کنید که اگر A نامنفرد باشد، آنگاه $\tilde{u} = x^*$.

قضیه ۲۵.۱ جواب عمومی مسئله (۱۰-۱) به صورت

$$x(v) = A^+ b + v, \quad v \in \mathcal{N}(A).$$

است. به علاوه، به ازای هر $v \neq 0$ ، داریم

$$\|x(0)\|_2 < \|x(v)\|_2.$$

۳۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

برهان. با توجه به قضیه ۲۴.۱ بردار x یک جواب مسئله کمترین توان‌های دوم است اگر و تنها اگر $A^T Ax = A^T b$. با جایگذاری تجزیه مقدار تکین ماتریس A در دستگاه معادلات نرمال، داریم

$$\Sigma^T \Sigma (V^T x) = \Sigma^T (U^T b).$$

از این رابطه نتیجه می‌شود

$$\begin{cases} (V^T x)_i = \frac{1}{\sigma_i} (U^T b)_i, & i = 1, \dots, r, \\ (V^T x)_i = t_i, & i = r+1, \dots, n, \end{cases} \quad (13-1)$$

که در آن $t_i \in \mathbb{R}$ ها مقادیری دلخواه هستند. بنابراین، داریم

$$V^T x = \begin{pmatrix} \frac{1}{\sigma_1} (U^T b)_1 \\ \vdots \\ \frac{1}{\sigma_r} (U^T b)_r \\ \circ \\ \vdots \\ \circ \end{pmatrix} + \begin{pmatrix} \circ \\ \vdots \\ \circ \\ t_{r+1} \\ \vdots \\ t_n \end{pmatrix} = \Sigma^+ U^T b + \begin{pmatrix} \circ \\ \vdots \\ \circ \\ t_{r+1} \\ \vdots \\ t_n \end{pmatrix}.$$

با ضرب طرفین این رابطه در V و اینکه $VV^T = I$ ، خواهیم داشت

$$\begin{aligned} x &= V \Sigma^+ U^T b + \sum_{i=r+1}^n t_i v_i \\ &= A^+ b + \sum_{i=r+1}^n t_i v_i \\ &= A^+ b + z =: x(z), \end{aligned}$$

که در آن

$$z = \sum_{i=r+1}^n t_i v_i.$$

با توجه به نتیجه ۳.۱، روشن است که $z \in \mathcal{N}(A)$. بنابراین، قسمت نخست قضیه ثابت می‌شود. برای اثبات قسمت دوم قضیه، ابتدا توجه کنید که (چرا؟)

$$A^+b \in \text{span}\{v_1, v_2, \dots, v_r\}.$$

بنابراین، $A^+b \perp z$ و در نتیجه، داریم

$$\begin{aligned} \|x(z)\|_2^2 &= \|A^+b\|_2^2 + \left\| \sum_{i=r+1}^n t_i v_i \right\|_2^2 \\ &= \|A^+b\|_2^2 + \sum_{i=r+1}^n |t_i|^2 \\ &> \|A^+b\|_2^2. \end{aligned}$$

□

که نتیجه لازم را به دست می‌دهد.

مثال ۲۱.۱ دستگاه معادلات خطی زیر را در نظر بگیرید

$$Ax = \begin{pmatrix} 1 & 2 \\ -1 & -2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = b.$$

به سادگی می‌توان دید که این دستگاه ناسازگار است (جواب ندارد). بنابراین، سعی می‌کنیم مسئله کمترین توان‌های دوم متناظر را حل کنیم. دستگاه معادلات نرمال متناظر، یعنی $A^T Ax = A^T b$ ، به صورت زیر است

$$\begin{pmatrix} 3 & 6 \\ 6 & 12 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \quad (14-1)$$

با توجه به اینکه ماتریس ضرایب دستگاه منفرد است، این دستگاه بی‌شمار جواب دارد. یکی از جواب‌های دستگاه به صورت

$$x^* = A^+b = \begin{pmatrix} \frac{1}{15} \\ \frac{2}{15} \end{pmatrix}.$$

این جواب از بین همه جواب‌های دستگاه (۱۴-۱)، کمترین نرم-۲ را دارد. برای مثال، بردار

$$\tilde{x} = \left(-\frac{2}{3} \quad \frac{1}{2} \right)^T,$$

جواب دیگری برای دستگاه (۱۴-۱) است، اما داریم

$$\|x^*\|_2 = 0.1491 < \|\tilde{x}\|_2 = 0.8333.$$

۷-۱ روش حذفی گاوس و تجزیه LU

دستگاه

$$Ax = b, \quad (15-1)$$

را در نظر بگیرید که در آن $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ و $x, b \in \mathbb{R}^n$ ، که در اینجا x بردار مجهول است. در روش حذفی گاوس برای حل دستگاه ۱۵-۱، این دستگاه به دستگاه معادل (دستگاه بالامثلثی) $Ux = c$ تبدیل می‌شود که در آن U یک ماتریس بالامثلثی است. این دستگاه به‌سادگی قابل حل است، زیرا اگر دستگاه بالامثلثی را به‌صورت

$$\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix},$$

بنویسیم، آنگاه با شرط اینکه برای هر i ، $u_{ii} \neq 0$ ، داریم

$$x_n = \frac{c_n}{u_{nn}},$$

$$x_k = \frac{1}{u_{kk}} \left(c_k - \sum_{j=k+1}^n u_{kj} x_j \right), \quad k = n-1, n-2, \dots, 1.$$

این روند را جایگذاری پسر می‌گویند. به همین ترتیب اگر ماتریس ضرایب دستگاه یک ماتریس پایین‌مثلثی باشد، یعنی اگر دستگاه به صورت

$$\begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix},$$

باشد، آنگاه با شرط اینکه برای هر i ، $l_{ii} \neq 0$ ، داریم

$$x_1 = \frac{c_1}{l_{11}},$$

$$x_k = \frac{1}{l_{kk}} \left(c_k - \sum_{j=1}^{k-1} l_{kj} x_j \right), \quad k = 2, 3, \dots, n.$$

این روند را جایگذاری پیشرو می‌گویند. دستگاه (۱-۱۵) را به صورت

$$E_1 : a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1,$$

$$E_2 : a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2,$$

$\vdots$

$$E_n : a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n,$$

می‌نویسیم، که در آن E_i نشان دهنده معادله i ام است. فرایند روش حذفی گاوس در $n - 1$ مرحله انجام می‌شود و ایده اصلی روش این است که با سه عمل

(۱) تعویض دو معادله؛

(۲) افزودن مضربی از طرفین یک معادله به طرفین معادله دیگر؛

(۳) ضرب طرفین یک معادله در یک عدد ناصفر؛

که اعمال سطری مقدماتی نامیده می‌شوند، دستگاه را به دستگاه معادل دیگری تبدیل می‌کنیم (دستگاهی که جواب آن با جواب دستگاه اولیه یکسان است). برای این کار، ماتریس افزوده

$$(A, b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array} \right),$$

را تشکیل می‌دهیم. انجام اعمال سطری مقدماتی روی دستگاه $Ax = b$ ، معادل با انجام این اعمال روی سطرهای ماتریس افزوده است. $n - 1$ مرحله روش حذفی گاوس به صورت زیر انجام می‌گیرد.

گام ۱: متغیر x_1 در همه معادلات به جز معادله اول حذف می‌شود. برای این کار، فرض کنید $a_{11} \neq 0$. قرار می‌دهیم

$$l_{i1} = \frac{a_{i1}}{a_{11}}, \quad i = 2, 3, \dots, n.$$

سپس به ازای $i = 2, 3, \dots, n$ ، $-l_{i1}$ برابر سطر اول ماتریس افزوده (A, b) را به سطر i ام این ماتریس اضافه می‌کنیم. در این صورت، ماتریس افزوده به ماتریس $(A^{(1)}, b^{(1)})$ تبدیل می‌شود که در آن

$$\begin{aligned} a_{ij}^{(1)} &= a_{ij} - l_{i1}a_{1j}, \quad i = 2, 3, \dots, n, \quad j = 2, 3, \dots, n, \\ b_i^{(1)} &= b_i - l_{i1}b_1, \quad i = 2, 3, \dots, n. \end{aligned}$$

در اینجا توجه کنید که $a_{i1}^{(1)} = 0$ ، $i = 2, 3, \dots, n$ و سایر درایه‌های ماتریس افزوده $(A^{(1)}, b^{(1)})$ با درایه‌های متناظر در ماتریس (A, b) برابرند.

گام ۲: متغیر x_2 را از معادلات سوم تا n ام حذف می‌کنیم. فرض می‌کنیم $a_{22}^{(1)} \neq 0$ و قرار می‌دهیم

$$l_{i2} = \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}, \quad i = 3, 4, \dots, n.$$

سپس به ازای $i = 3, 4, \dots, n$ ، $-l_{i2}$ برابر سطر دوم ماتریس افزوده $(A^{(1)}, b^{(1)})$ را به سطر i ام این ماتریس اضافه می‌کنیم. در این صورت، ماتریس $(A^{(2)}, b^{(2)})$ به دست می‌آید که در آن

$$\begin{aligned} a_{ij}^{(2)} &= a_{ij}^{(1)} - l_{i2}a_{2j}^{(1)}, \quad i = 3, 4, \dots, n, \quad j = 3, 4, \dots, n, \\ b_i^{(2)} &= b_i^{(1)} - l_{i2}b_2^{(1)}, \quad i = 3, 4, \dots, n, \\ a_{i2}^{(2)} &= 0, \quad i = 3, 4, \dots, n. \end{aligned}$$

توجه کنید که سایر درایه‌های ماتریس $(A^{(2)}, b^{(2)})$ با درایه‌های متناظر در ماتریس $(A^{(1)}, b^{(1)})$ برابرند.

گام k : متغیر x_k را از معادلات $(k+1)$ ام تا n ام حذف می‌کنیم. برای این کار، فرض می‌کنیم $a_{kk}^{(k-1)} \neq 0$ و قرار می‌دهیم

$$l_{ik} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}, \quad i = k+1, k+2, \dots, n.$$

سپس، به ازای $i = k+1, k+2, \dots, n$ ، $-l_{ik}$ برابر سطر k ام ماتریس افزوده $(A^{(k-1)}, b^{(k-1)})$ را به سطر i ام این ماتریس اضافه می‌کنیم. در این صورت، ماتریس $(A^{(k)}, b^{(k)})$ به دست می‌آید که در آن

$$\begin{aligned} a_{ij}^{(k)} &= a_{ij}^{(k-1)} - l_{ik}a_{kj}^{(k-1)}, \quad i = k+1, k+2, \dots, n, \quad j = k+1, k+2, \dots, n, \\ b_i^{(k)} &= b_i^{(k-1)} - l_{ik}b_k^{(k-1)}, \quad i = k+1, k+2, \dots, n, \\ a_{ik}^{(k)} &= 0, \quad i = k+1, k+2, \dots, n. \end{aligned}$$

در اینجا درایه‌های دیگر ماتریس $(A^{(k)}, b^{(k)})$ با درایه‌های متناظر در ماتریس $(A^{(k-1)}, b^{(k-1)})$ برابر می‌باشند.

به وضوح در پایان مرحله $(n-1)$ ام ماتریس $A^{(n-1)}$ یک ماتریس بالامثلثی خواهد بود و دستگاه متناظر به سادگی حل می‌شود. روند فوق به روش حذفی گاوس بدون محورگیری (یا به طور خلاصه، روش حذفی گاوس) معروف است.

به درایه‌های $a_{11}^{(1)}, \dots, a_{nn}^{(n-1)}$ که در $n-1$ مرحله از روش حذفی گاوس ظاهر می‌شوند، **درایه‌های محوری** می‌گویند. در صورتی که یکی از درایه‌های محوری برابر با صفر باشد، آنگاه روش حذفی گاوس شکست خورده و الگوریتم متوقف می‌شود. همان‌گونه که ملاحظه شد، اگر تمام درایه‌های محوری ناصفر باشند، آنگاه الگوریتم با موفقیت به پایان می‌رسد. الگوریتم روش حذفی گاوس را می‌توان به صورت الگوریتم ۱.۱ جمع‌بندی کرد. در اینجا لازم به ذکر است که تمام $a_{ij}^{(k)}$ ها و $b_i^{(k)}$ ها به ترتیب در A و b ذخیره می‌شوند. در واقع، در پایان اجرای الگوریتم، قسمت بالامثلثی $A^{(n-1)}$ در قسمت بالامثلثی A و بردار $b^{(n-1)}$ در بردار b ذخیره شده‌اند.

در حالت کلی روش حذفی گاوس ناپایدار است. اما می‌توان با عمل محورگیری، پایداری این الگوریتم را بهبود بخشید. در نسخه حذفی گاوس با محورگیری جزئی و در هر مرحله از روش، از بین درایه محوری و درایه‌های زیر آن در ستون مربوطه، بزرگترین درایه از حیث قدرمطلق را می‌یابیم و سطر مربوطه را با سطر

۴۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

درایه‌محوری تعویض می‌کنیم. در صورتی که بزرگ‌ترین درایه از حیث قدرمطلق صفر باشد، ماتریس A منفرد است (چرا؟).

الگوریتم ۱۰.۱ حذفی گاوس

1. Input $A = (a_{ij})$ and $b = (b_1, \dots, b_n)^T$
2. For $k = 1, \dots, n - 1$, Do
3. For $i = k + 1, \dots, n$ Do
4. If $a_{kk} \neq 0$ then $l_{ik} := \frac{a_{ik}}{a_{kk}}$, else stop
5. $b_i := b_i - l_{ik}b_k$
6. For $j = k + 1, \dots, n$ Do
7. $a_{ij} := a_{ij} - l_{ik}a_{kj}$
8. EndDo
9. EndDo
10. EndDo

فرض کنید $A^{(n-1)} = (a_{ij}^{(n-1)})$. بعد از اجرای الگوریتم ۱۰.۱، از روی ماتریس به دست آمده، ماتریس‌های $L = (\ell_{ij})$ و $U = (u_{ij})$ را با درایه‌های

$$\ell_{ij} = \begin{cases} l_{ij}, & i > j, \\ 1, & i = j, \\ 0, & i < j, \end{cases} \quad u_{ij} = \begin{cases} a_{ij}^{(n-1)}, & i \leq j, \\ 0, & i > j, \end{cases}$$

تعریف می‌کنیم. می‌توان دید که

$$A = LU. \quad (۱۶-۱)$$

با توجه به بحث‌های ارائه شده ماتریس L یک ماتریس پائین‌مثلثی با درایه‌های قطری ۱ و U یک ماتریس بالامثلثی با درایه‌های قطری ناصفر است. تجزیه (۱۶-۱) به تجزیه LU معروف است. اکنون، اگر قرار دهیم $D = \text{diag}_n(u_{11}, u_{22}, \dots, u_{nn})$ و ماتریس $\tilde{U} = (\tilde{u}_{ij})$ را با درایه‌های

$$\tilde{u}_{ij} = \begin{cases} u_{ij}/u_{ii}, & i \leq j, \\ 0, & i > j, \end{cases}$$

تعریف کنیم، آنگاه خواهیم داشت

$$A = LD\tilde{U}.$$

توجه کنید در اینجا $\tilde{U}$ یک ماتریس بالامثلثی با درایه‌های قطری ۱ است. این تجزیه به تجزیه LDU معروف است. در صورتی که ماتریس A متقارن باشد، داریم $\tilde{U} = L^T$ و در این صورت، تجزیه A به صورت

$$A = LDL^T,$$

خواهد بود. این تجزیه، به تجزیه LDL^T معروف است.

۸-۱ یک مدل برای آزمایش‌های عددی

معادله دیفرانسیل

$$-\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \sigma \frac{\partial u}{\partial x} + \tau \frac{\partial u}{\partial y} = f(x, y), \quad (x, y) \in \Omega = (0, 1) \times (0, 1), \quad (17-1)$$

با شرط مرزی $u = 0$ روی مرز Ω را در نظر بگیرید. وقتی که

$$\sigma = \tau = 0,$$

این معادله به معادله پواسون معروف است. بازه $[0, 1]$ را در هر دو جهت x و y به $N + 1$ قسمت مساوی تقسیم کرده و قرار می‌دهیم

$$h = \frac{1}{N+1}, \quad x_i = ih, \quad y_j = jh, \quad i = 0, 1, \dots, N+1,$$

و $u_{ij} = u(x_i, y_j)$ در این صورت، داریم

$$\frac{\partial^2 u}{\partial x^2}(x_i, y_j) = \frac{u_{i-1,j} - 2u_{ij} + u_{i+1,j}}{h^2} + \mathcal{O}(h^2), \quad (18-1)$$

$$\frac{\partial^2 u}{\partial y^2}(x_i, y_j) = \frac{u_{i,j-1} - 2u_{ij} + u_{i,j+1}}{h^2} + \mathcal{O}(h^2), \quad (19-1)$$

$$\frac{\partial u}{\partial x}(x_i, y_j) = \frac{u_{i+1,j} - u_{i-1,j}}{2h} + \mathcal{O}(h^2), \quad (20-1)$$

$$\frac{\partial u}{\partial y}(x_i, y_j) = \frac{u_{i,j+1} - u_{i,j-1}}{2h} + \mathcal{O}(h^2), \quad (21-1)$$

که در آن، $\mathcal{O}(\cdot)$ نماد « $\mathcal{O}$ ی بزرگ» است [۱]. با نوشتن معادله (۱۷-۱) در هر نقطه‌ی درونی (x_i, y_j) از ناحیه Ω و جایگذاری این چهار رابطه در آن و صرف‌نظر کردن از $\mathcal{O}(h^2)$ ، به ازای $i, j = 1, 2, \dots, N$

داریم

$$-\hat{\beta}u_{i-1,j} - \tilde{\beta}u_{i+1,j} + \mathfrak{F}u_{ij} - \hat{\gamma}u_{i,j-1} - \tilde{\gamma}u_{i,j+1} = h^2 f_{ij}, \quad (22-1)$$

که در آن

$$\beta = \frac{\sigma h}{\mathfrak{F}}, \quad \hat{\beta} = 1 + \beta, \quad \tilde{\beta} = 1 - \beta, \quad \gamma = \frac{\tau h}{\mathfrak{F}}, \quad \hat{\gamma} = 1 + \gamma, \quad \tilde{\gamma} = 1 - \gamma,$$

و $f_{ij} = f(x_i, y_j)$ اگر بردارهای x و b را به صورت

$$x = [u_{11}, u_{21}, \dots, u_{N1}, u_{12}, \dots, u_{N2}, \dots, u_{1N}, \dots, u_{NN}]^T,$$

$$b = h^2 [f_{11}, \dots, f_{N1}, \dots, f_{1N}, \dots, f_{NN}]^T,$$

تعریف کنیم، آنگاه (22-1) را می‌توان به شکل ماتریسی

$$Ax = b,$$

نوشت که در آن A یک ماتریس از مرتبه $n = N^2$ به صورت

$$A = \begin{pmatrix} J & L & & \\ K & J & L & \\ & \ddots & \ddots & \ddots \\ & & K & J & L \\ & & & K & J \end{pmatrix}, \quad (23-1)$$

است، به طوری که

$$J = \begin{pmatrix} \mathfrak{F} & -\tilde{\beta} & & \\ -\hat{\beta} & \mathfrak{F} & -\tilde{\beta} & \\ & \ddots & \ddots & \ddots \\ & & -\hat{\beta} & \mathfrak{F} & -\tilde{\beta} \\ & & & -\hat{\beta} & \mathfrak{F} \end{pmatrix} \in \mathbb{R}^{N \times N},$$

کرونکر، ماتریس A را می‌توان به صورت

$$A = I_N \otimes T_{\backslash} + T_{\uparrow} \otimes I_N,$$

نوشت، که در آن

$$T_{\backslash} = \text{tridiag}_N(-\hat{\beta}, \mathbb{I}, -\tilde{\beta}), \quad T_{\uparrow} = \text{tridiag}_N(-\hat{\gamma}, \circ, -\tilde{\gamma}).$$

لازم به ذکر است که برای هر دو ماتریس $A = (a_{ij})$ و B ضرب کرونکر A در B به صورت $A \otimes B = (a_{ij}B)$ تعریف می‌شود. برای مثال اگر

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ 3 & 1 \end{pmatrix},$$

آنگاه

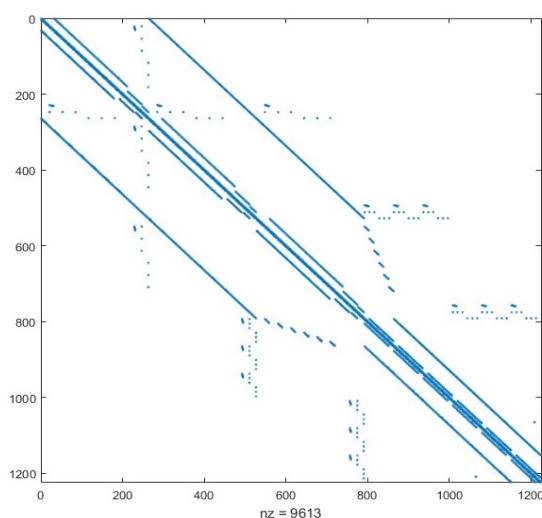
$$A \otimes B = \begin{pmatrix} 1B & 2B & 1B \\ 2B & 3B & 2B \end{pmatrix} = \begin{pmatrix} 2 & 3 & 2 & 3 & 2 & 2 \\ 3 & 1 & 2 & 6 & 3 & 2 \\ 2 & 6 & 2 & 9 & 6 & 2 \\ 6 & 2 & 2 & 3 & 6 & 2 \end{pmatrix}.$$

تابع KronConvDiff که به زبان متلب نوشته شده است با دریافت γ, β و N ماتریس A را تولید می‌کند (صفحه ۱۹۹ از [۵] را ببینید).

```
function A=KronConvDiff(sigma,tau,N);
h=1/(N+1); beta=(sigma*h)/2; gamma=(tau*h)/2;
ee=ones(N,1);
a=4; b=-1-gamma; c=-1-beta; d=-1+beta; e=-1+gamma;
t1=spdiags([c*ee,a*ee,d*ee],-1:1,N,N);
t2=spdiags([b*ee,zeros(N,1),e*ee],-1:1,N,N);
A=kron(speye(N),t1)+kron(t2,speye(N));
```

ماتریس (۱-۲۳) را در نظر بگیرید. بسیاری از درایه‌های این ماتریس صفرند. این‌گونه ماتریس‌ها را ماتریس‌های تُنک می‌گویند. یک تعریف کلی (نادقیق) برای یک ماتریس تُنک این است که کمتر از پنج

درصد از درایه‌های ماتریس ناصفر باشند. ماتریسی که تُنک نباشد، ماتریس چگال نامیده می‌شود. شکل ۲-۱ یک ماتریس تُنک 1224×1224 را نمایش می‌دهد که تعداد درایه‌های ناصفر آن برابر با ۹۶۱۳ است (درایه‌های ناصفر به صورت نقاطی در ماتریس نشان داده شده‌اند). تنها ۰/۶۴ درصد از درایه‌های ماتریس ناصفرند. این ماتریس یک نمونه از یک ماتریس تُنک است.



شکل ۲-۱ یک نمونه از ماتریس تُنک

در بخش بعد، در خصوص این‌گونه ماتریس‌ها، شیوه‌های ذخیره‌سازی و به‌کارگیری آنها مطالبی را ارائه می‌کنیم.

۹-۱ ماتریس‌های تُنک و به‌کارگیری آنها

با توجه به اینکه تعداد کمی از درایه‌های یک ماتریس تُنک، ناصفر هستند، معمولاً تنها درایه‌های ناصفر ماتریس ذخیره‌سازی می‌شوند. در این صورت، باید دسترسی به یک درایه خاص، سطر و ستون‌های ماتریس به سادگی امکان‌پذیر باشد. همچنین، اعمالی مثل ضرب ماتریس در بردار به سادگی امکان‌پذیر باشد؛ زیرا مهم‌ترین عمل در بسیاری از روش‌های تکراری، ضرب ماتریس در بردار است. در این بخش، سه قالب ذخیره‌سازی یک ماتریس تُنک را شرح می‌دهیم.

۱-۹-۱ قالب مختصاتی

در قالب مختصاتی (قالب COO)، قالب مختصاتی درایه‌های ناصفر ماتریس و آدرس هر درایه در ماتریس، یا به عبارت دیگر اندیس سطر و ستون هر درایه ناصفر، ذخیره می‌شوند. برای مثال، یک ماتریس سه‌قطری با بعد 1000×1000 ، دارای یک میلیون درایه است، اما تنها ۲۹۹۸ درایه آن ناصفرند. در این صورت، اگر بخواهیم تنها درایه‌های ناصفر و آدرس درایه‌ها را ذخیره نمایم، به تعداد $8994 = 3 \times 2998$ عدد ذخیره می‌شوند. ملاحظه می‌شود که این راهکار ذخیره‌سازی بسیار مقرون به صرفه است. قالب مختصاتی ساده‌ترین و رایج‌ترین شیوه ذخیره‌سازی یک ماتریس در حافظه کامپیوتر است.

فرض کنید تعداد درایه‌های ناصفر ماتریس $A \in \mathbb{R}^{n \times n}$ برابر با nz باشد. در قالب مختصاتی، سه بردار ia ، ja و a با طول nz برای ذخیره‌سازی A استفاده می‌شود. در بردار ia ، اندیس سطری درایه‌ها، در بردار ja ، اندیس ستونی درایه‌ها و در بردار a ، درایه‌های ناصفر ماتریس ذخیره می‌شوند. درواقع، اگر $a(k)$ محل ذخیره‌سازی a_{ij} باشد، آنگاه $ia(k) = i$ و $ja(k) = j$. برای روشن شدن مطلب، مثال بعدی را ارائه می‌کنیم.

مثال ۲۲.۱ ماتریس

$$A = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ 2 & 0 & -2 & 0 & 3 \\ 0 & -3 & 0 & 0 & 0 \\ 0 & 4 & 0 & -4 & 0 \\ 5 & 0 & -5 & 0 & 6 \end{pmatrix},$$

را در نظر بگیرید. ذخیره‌سازی این ماتریس در قالب مختصاتی به صورت جدول ۱-۱ است. در این مثال، درایه‌های ماتریس به صورت دلخواه و بدون هیچ ترتیبی چیده شده‌اند، ولی معمولاً به صورت سطر به سطر یا ستون به ستون ذخیره می‌شوند.

در قالب COO برای ذخیره‌سازی یک ماتریس $n \times n$ ، $3nz$ عدد ذخیره می‌شود. توجه کنید که تعداد درایه‌های ماتریس n^2 است که برای یک ماتریس $n \times n$ بزرگ، در مقایسه با $3nz$ ، عدد بسیار بزرگی است.

جدول ۱-۱ نمایش ماتریس مثال ۲۲.۱ در قالب مختصاتی

k	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
ia	۱	۲	۲	۱	۵	۳	۴	۵	۲	۴	۵
ja	۴	۵	۱	۱	۵	۲	۴	۳	۳	۲	۱
a	-۱	۳	۲	۱	۶	-۳	-۴	-۵	-۲	۴	۵

در بسیاری از روش‌های تکراری برای حل دستگاه معادلات خطی نیاز به چندین ضرب ماتریس در بردار می‌باشد. در این صورت، وقتی که ماتریس ضرایب دستگاه در قالب COO ذخیره شده است، لازم است الگوریتمی برای ضرب ماتریس در یک بردار طراحی شود. در واقع، اگر ماتریس A در قالب COO ذخیره شده باشد، برای بردار $x \in \mathbb{R}^n$ ، می‌خواهیم $y = Ax$ را محاسبه کنیم. ابتدا توجه کنید که

$$y_i = \sum_{j=1}^n a_{ij}x_j, \quad i = 1, 2, \dots, n. \quad (24-1)$$

همان‌گونه که ملاحظه می‌شود مقدار a_{ij} تنها در مؤلفه y_i از بردار y ظاهر می‌شود و تنها یک جمله از مجموع $(24-1)$ شامل a_{ij} ، و به صورت $a_{ij}x_j$ است. بنابراین، برای محاسبه y ، نخست قرار می‌دهیم $y = 0$. سپس کافی است بردار a یک بار پیمایش شود. در مرحله k ام پیمایش، داریم

$$i = ia(k), \quad j = ja(k), \quad a_{ij} = a(k).$$

بنابراین، قرار می‌دهیم $y_i := y_i + a_{ij}x_j$ یا

$$y(ia(k)) = y(ia(k)) + a(k) * x(ja(k)).$$

بعد از اتمام پیمایش، بردار y برابر با Ax خواهد بود. توجه کنید که درایه‌های صفر ماتریس A در محاسبه y_i ها تأثیری ندارند و جملات متناظر برابر با صفرند. با جمع‌بندی مطالب بالا، الگوریتم ۲.۱ برای محاسبه بردار $y = Ax$ وقتی که ماتریس A در قالب COO ذخیره شده است، ارائه می‌شود.

الگوریتم ۲.۱ ضرب ماتریس در بردار در قالب مختصاتی

1. Input a, ia, ja, nz and x
2. $y := 0$
3. For $k = 1, 2, \dots, nz$, Do
4. $y(ia(k)) := y(ia(k)) + a(k) x(ja(k))$
5. EndDo

در زبان برنامه‌نویسی متلب این امکان وجود دارد که یک ماتریس چگال را به قالب COO یا برعکس تبدیل کرد. اگر A یک ماتریس دلخواه باشد، آنگاه دستور $B = \text{sparse}(A)$ ماتریس A را به قالب COO تبدیل کرده و در B ذخیره می‌کند. برای مثال، اگر

$$A = \begin{pmatrix} 0 & 2 & 0 \\ -2 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

آنگاه دستور $B = \text{sparse}(A)$ ، قالب مختصاتی A را به صورت

```
>> B
```

```
B=
```

```
(2,1)      -2
```

```
(1,2)       2
```

```
(3,3)       3
```

نمایش می‌دهد. به علاوه، دستور $[ia, ja, a] = \text{find}(A)$ بردارهای ia ، ja و a را برمی‌گرداند. همچنین، اگر برای ماتریس A بردارهای ia ، ja و a معلوم باشند، دستور $A = \text{sparse}(ia, ja, a)$ قالب مختصاتی A را به دست خواهد داد. سرانجام، اگر ماتریس A در قالب COO در متلب ذخیره شده باشد، آنگاه دستور $y = A * x$ حاصلضرب ماتریس A در x را با استفاده از الگوریتم ۲.۱ محاسبه کرده و حاصل را در بردار y ذخیره می‌نماید.

۲-۹-۱ قالب سطری تُنک فشرده

قالب سطری تُنک فشرده (CSR) شامل سه آرایه به صورت زیر است

- آرایه حقیقی a شامل درایه‌های ناصفر ماتریس A است، که به صورت سطر به سطر، از سطر اول تا n ام ذخیره شده‌اند، طول این آرایه برابر nnz می‌باشد.
- آرایه صحیح ja شامل اندیس ستون درایه‌های نظیر در آرایه a است. به عبارت دیگر، $ja(k)$ محتوی اندیس ستون درایه ذخیره شده در $a(k)$ است. طول این آرایه نیز برابر nnz است.
- آرایه صحیح ia شامل اشاره‌گرهایی به مکان وقوع نخستین درایه ناصفر هر سطر ماتریس، در آرایه a است. طول این آرایه برابر با $n + 1$ است و $ia(n + 1) = nnz + 1$. بنابراین، اگر

۵۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

جدول ۲-۱ نمایش ماتریس مثال ۲۲.۱ در قالب CSR.

k	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱
a	۱	-۱	۲	-۲	۳	-۳	۴	-۴	۵	-۵	۶
ja	۱	۴	۱	۳	۵	۲	۲	۴	۱	۳	۵
ia	۱	۳	۶	۷	۹	۱۲					

دارند. $ia(i) = r$ و $ia(i+1) = s$ ، آنگاه $a(k)$ ، $k = r, \dots, s-1$ در سطر i ام ماتریس A قرار

در قالب CSR تعداد $2nnz + n + 1$ عدد حقیقی یا صحیح ذخیره می‌شوند.

مثال ۲۳.۱ با توجه توضیحات بالا، ذخیره‌سازی ماتریس مثال ۲۲.۱ در قالب CSR به صورت جدول ۲-۱ خواهد بود.

برخلاف قالب مختصاتی، در اینجا جست‌وجوی یک درایه به سادگی انجام می‌گیرد. برای مثال، اگر بخواهیم درایه a_{ij} را بیابیم، کافی است این درایه را در

$$a(k), \quad k = ia(i), \dots, ia(i+1) - 1,$$

جست‌وجو کنیم. درواقع، یک بار $a(k)$ ، $k = ia(i), \dots, ia(i+1) - 1$ را می‌پیماییم. در این صورت، اگر به ازای k ای $ja(k) = j$ ، آنگاه $a_{ij} = a(k)$ ، و در غیر این صورت، $a_{ij} = 0$.

اینک، به نحوه ضرب یک ماتریس در بردار، زمانی که ماتریس A در قالب CSR ذخیره شده است، می‌پردازیم. فرض کنید $x \in \mathbb{R}^n$ و $y = Ax$. با توجه به رابطه (۲۴-۱) فقط درایه‌های سطر i ام ماتریس A در y_i سهم دارند. درایه‌های ناصفر سطر i ام به صورت زیر هستند

$$a(k), \quad k = ia(i), \dots, k = ia(i+1) - 1.$$

حال با توجه به اینکه اندیس ستونی درایه‌ی $a(k)$ برابر است با $ja(k)$ ، داریم

$$y_i = \sum_{k=ia(i)}^{ia(i+1)-1} a(k)x(ja(k)).$$

با توجه به این رابطه، الگوریتم ضرب یک ماتریس در بردار، وقتی که ماتریس در قالب CSR ذخیره شده است، را می‌توان به صورت زیر جمع‌بندی کرد.

الگوریتم ۳.۱ ضرب ماتریس در بردار در قالب CSR

1. Input a, ia, ja, nnz and x
2. $y := 0$
3. For $i = 1, 2, \dots, n$, Do
4. For $k = ia(i), \dots, ia(i+1) - 1$, Do
5. $y(i) := y(i) + a(k)x(ja(k))$
6. EndDo.
7. EndDo.

۳-۹-۱ قالب ستونی تُنک فشرده

قالب دیگر ذخیره‌سازی یک ماتریس، قالب ستونی تُنک فشرده (CSC قالب) است که کاملاً شبیه به قالب CSR است، با این تفاوت که در اینجا ستون‌های ماتریس A در قالب CSR ذخیره می‌شوند. به عبارت دیگر، A^T در قالب سطرِی تُنک فشرده ذخیره می‌شود. می‌توان الگوریتمی شبیه به الگوریتم ۳.۱ برای ضرب یک ماتریس تُنک در بردار، برای حالتی که ماتریس A در قالب CSC ذخیره شده باشد، طراحی کرد (تمرین ۲۲.۱ را ببینید).

تمرینات

۱.۱ فرض کنید $A \in \mathbb{C}^{m \times n}$. نشان دهید که به ازای هر $x \in \mathbb{C}^n$ ، داریم

$$\|Ax\|_2 \leq \|A\|_F \|x\|_2.$$

و از آنجا نتیجه بگیرید که $\|A\|_2 \leq \|A\|_F$.

۲.۱ فرض کنید $S \in \mathbb{C}^{m \times n}$ یک ماتریس هرمیتی‌کج باشد و $\alpha > 0$. نشان دهید که هر دو ماتریس $\alpha I + S$ و $\alpha I - S$ نامنفرد هستند، که در آن I ماتریس همانی $n \times n$ است.

۳.۱ فرض کنید $A, B \in \mathbb{C}^{n \times n}$ به طوری که $AB = BA$. همچنین، فرض کنید p و q دو چندجمله‌ای دلخواه باشند. نشان دهید

$$p(A)q(B) = q(B)p(A).$$

به علاوه، نشان دهید اگر ماتریس $q(B)$ نامنفرد باشد، آنگاه

$$p(A)q(B)^{-1} = q(B)^{-1}p(A).$$

۵۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنک

۴.۱ نشان دهید اگر به‌ازای هر $x, y \in \mathbb{C}^n$ داشته باشیم $\langle x, Ay \rangle = \langle Ax, y \rangle$ ، آنگاه A هرمیتی است.

۵.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $B \in \mathbb{C}^{n \times m}$. نشان دهید که مقادیر ویژهٔ ناصفر دو ماتریس A و B یکسان هستند. سپس، نتیجه بگیرید که $\rho(AB) = \rho(BA)$.

۶.۱ نشان دهید اگر $A \in \mathbb{C}^{n \times n}$ یک ماتریس دلخواه و ماتریس $U \in \mathbb{C}^{n \times n}$ یک ماتریس یکانی باشد، آنگاه

$$\|UA\|_F = \|AU\|_F = \|A\|_F, \quad \|UA\|_2 = \|AU\|_2 = \|A\|_2.$$

۷.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$. همچنین، فرض کنید

$$H = \frac{1}{2}(A + A^H), \quad S = \frac{1}{2}(A - A^H).$$

نشان دهید A نرمال است اگر و تنها اگر $HS = SH$.

۸.۱ فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ هرمیتی و مقادیر ویژهٔ آن مثبت باشند. نشان دهید که A معین مثبت هرمیتی است.

۹.۱ فرض کنید

$$A = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix},$$

که در آن $i = 1, 2, \dots, n, x_i, y_i \in \mathbb{R}$. با محاسبهٔ $\det(AA^T)$ نشان دهید

$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2.$$

۱۰.۱ فرض کنید که $A = LU$ تجزیهٔ LU ماتریس A باشد. نشان دهید

$$U_{11} = A_{11}, \quad U_{ii} = \frac{\det(A_i)}{\det(A_{i-1})}, \quad i = 2, 3, \dots, n$$

که در آن A_i و A_{i-1} به‌ترتیب زیرماتریس‌های اصلی پیشرو مرتبهٔ i ام و $(i-1)$ ام ماتریس A می‌باشند. با استفاده از این رابطه، نتیجه بگیرید که اگر ماتریس A معین مثبت هرمیتی باشد، آنگاه $U_{ii} > 0$. با استفاده از این نتیجه، قضیهٔ ۱۳.۱ را ثابت کنید.

۱۱.۱ فرض کنید ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ غالب قطری اکید باشد، یعنی

$$|a_{ii}| > \sum_{j=1}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

با استفاده از قضیه دیسک‌های گرشگورین نشان دهید A نامنفرد است. به علاوه، اگر درایه‌های روی قطر A مثبت و ماتریس A هرمیتی باشد، آنگاه معین مثبت هرمیتی است.

۱۲.۱ نشان دهید مقادیر ویژه یک ماتریس هرمیتی کج به صورت $i\beta$ است که در آن $\beta \in \mathbb{R}$ و $i = \sqrt{-1}$ واحد موهومی است.

۱۳.۱ فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ هرمیتی باشد به طوری که تمام درایه‌های قطری آن برابر با ۱ می‌باشند. به علاوه، فرض کنید

$$\sum_{j=1}^n |a_{ij}| \leq 2, \quad i = 1, 2, \dots, n.$$

(الف) نشان دهید A نیمه‌معین مثبت هرمیتی است.

(ب) نشان دهید اگر λ یک مقدار ویژه A باشد، آنگاه $0 \leq \lambda \leq 2$.

(ج) نشان دهید $1 \leq \det(A) \leq 2^n$.

(راهنمایی: صفحه ۸۲ از [۱۰۰] را ببینید)

۱۴.۱ فرض کنید λ_i ها مقادیر ویژه $A \in \mathbb{C}^{n \times n}$ و σ_i ها مقادیر تکین آن باشند. نشان دهید

$$\sum_{i=1}^n |\lambda_i|^2 \leq \text{trace}(A^H A) = \text{trace}(A A^H) = \|A\|_F^2 = \sum_{i=1}^n \sigma_i^2. \quad (25-1)$$

نشان دهید در رابطه (۲۵-۱) تساوی اتفاق می‌افتد اگر و تنها اگر ماتریس A نرمال باشد. فرض کنید

$$A = \begin{pmatrix} 0 & \sqrt{x_1} & & & \\ & 0 & \sqrt{x_2} & & \\ & & 0 & \ddots & \\ & & & 0 & \sqrt{x_{n-1}} \\ \sqrt{x_n} & & & & 0 \end{pmatrix},$$

۵۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

با استفاده از رابطه (۱-۲۵)، نامساوی هندسی-حسابی

$$\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^n x_i,$$

را ثابت کنید.

۱۵.۱ فرض کنید ماتریس $W \in \mathbb{C}^{n \times n}$ معین مثبت متقارن و ماتریس $T \in \mathbb{C}^{n \times n}$ نیمه‌معین مثبت متقارن باشند. نشان دهید که مقادیر ویژه ماتریس $W^{-1}T$ نامنفی هستند.

۱۶.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $v \in \mathbb{C}^n$ ، $v \neq 0$. به‌علاوه، فرض کنید به‌ازای یک عدد طبیعی k داریم $A^k v = 0$ و $A^{k-1} v \neq 0$. نشان دهید مجموعه $B = \{v, Av, \dots, A^{k-1}v\}$ یک مجموعه مستقل خطی است.

۱۷.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ معین مثبت هرمیتی باشد. به‌ازای هر $x \in \mathbb{C}^n$ نشان دهید که

$$\sqrt{\lambda_1} \|x\|_2 \leq \|x\|_A \leq \sqrt{\lambda_n} \|x\|_2,$$

که در آن λ_1 و λ_n به‌ترتیب کوچک‌ترین و بزرگ‌ترین مقادیر ویژه A هستند.

۱۸.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ معین مثبت باشد (نه لزوماً هرمیتی). نشان دهید که قسمت حقیقی مقادیر ویژه A مثبتند.

۱۹.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$ معین مثبت باشد (نه لزوماً هرمیتی) و $\alpha > 0$. نشان دهید

$$\|(\alpha I + A)^{-1}(\alpha I - A)\|_2 < 1.$$

۲۰.۱ فرض کنید $A \in \mathbb{C}^{n \times n}$.

(الف) ثابت کنید اگر به‌ازای هر $x \in \mathbb{C}^n$ ، $x^H A x = 0$ ، آنگاه به‌ازای هر $x, y \in \mathbb{C}^n$ ، $x^H A y = 0$ و در نتیجه $A = 0$.

(ب) نشان دهید اگر به‌ازای هر $x \in \mathbb{C}^n$ ، $x^H A x \in \mathbb{R}$ ، آنگاه ماتریس A هرمیتی خواهد بود.

راهنمایی: برای اثبات قسمت (الف)، از اینکه به‌ازای هر $\alpha, \beta \in \mathbb{C}$ و $x, y \in \mathbb{C}^n$ داریم

$$(\alpha x + \beta y)^H A (\alpha x + \beta y) = 0,$$

استفاده کنید.

۲۱.۱ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ نامنفرد باشد و $\text{cond}_2(A) = 1$. نشان دهید که ماتریس A را می‌توان به صورت $A = \alpha Q$ نوشت که در آن α یک عدد مثبت و Q یک ماتریس متعامد است.

۲۲.۱ یک الگوریتم شبیه به الگوریتم ۳.۱ برای ضرب یک ماتریس $n \times n$ در بردار، برای حالتی که ماتریس A در قالب CSC ذخیره شده باشد، بنویسید.

روش‌های تکراری ایستا

در این فصل اصول پایه‌ای روش‌های تکراری و سپس روش تکراری AOR که توسط حاجیدیموس در [۴۵] ارائه شده است را معرفی نموده و بعضی از خواص آن را بررسی می‌کنیم. همچنین ماتریس‌های نامنفی را معرفی، خواص آنها را مطالعه نموده و سپس ارتباط این‌گونه ماتریس‌ها با همگرایی روش تکراری AOR را مطالعه می‌کنیم. در پایان برخی از روش‌های تکراری AOR پیش‌شرط‌سازی شده را معرفی می‌کنیم.

۱-۲ مقدمه

دو مشکل اساسی در روش‌های مستقیم اشغال حافظه رایانه و انباشتگی خطا می‌باشد. با دو مثال این دو موضوع را روشن می‌کنیم.

مثال ۱.۲ دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن

$$A = \text{tridiag}_n(-1, 2, -1) + E,$$

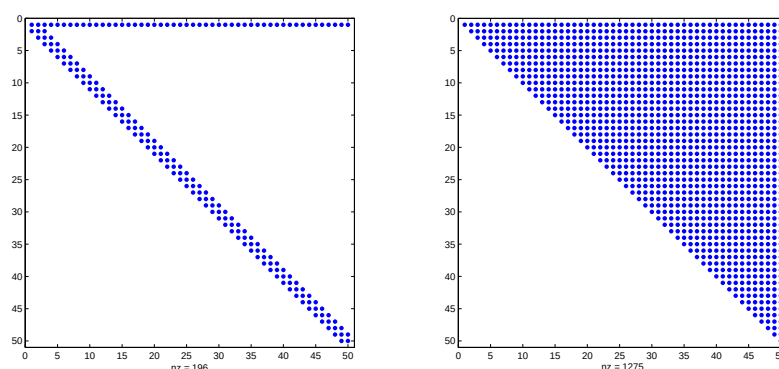
به‌طوری‌که درایه‌های ماتریس $E = (e_{ij})$ به‌صورت

$$e_{ij} = \begin{cases} -1, & i = 1, j = 3, \dots, n, \\ 0, & \text{درغیراین صورت} \end{cases}$$

تعریف می‌شود و b یک بردار دلخواه است. قرار می‌دهیم $n = 50$. اگر روش حذفی گاوس را برای حل دستگاه به‌کار ببریم ماتریس ضرایب دستگاه نهایی یک ماتریس بالامثلثی U خواهد بود. در شکل ۱-۲

۵۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

هر دو ماتریس A و U را نمایش داده‌ایم. همان‌گونه که ملاحظه می‌شود تعداد درایه‌های ناصفر ماتریس A برابر با $nz = ۱۹۶$ است، اما تعداد درایه‌های U برابر با ۱۲۷۵ است. می‌بینیم که برای حل مسئله با استفاده از روش حذفی گاوس استفاده از حافظه کامپیوتر به میزان ۵.۶ برابر افزایش می‌یابد. این نسبت به‌ازای $n = ۸۰۰$ برابر با ۱۰۰/۲۵ است!



شکل ۱-۲ نمایش ماتریس A و U برای مثال ۱.۲.

مثال ۲.۲ دستگاه معادلات خطی $H_n x = b$ را در نظر بگیرید که در آن $H_n = (h_{ij})$ ماتریس هیلبرت $n \times n$ است. یادآوری می‌کنیم که درایه‌های ماتریس هیلبرت به‌صورت

$$h_{ij} = \frac{1}{i+j-1}, \quad i, j = 1, 2, \dots, n,$$

تعریف می‌شوند. ماتریس هیلبرت معین مثبت متقارن است، اما بدحالت می‌باشد. برای مثال، به‌ازای $n = ۱۳$ داریم

$$\text{cond}_2(H_{13}) = \|H_{13}\|_2 \|H_{13}^{-1}\|_2 = ۱.۷۶ \times ۱۰^{۱۸} \gg ۱.$$

همچنین، قرار می‌دهیم $n = ۱۳$ و

$$b_i = \sum_{j=1}^{13} h_{ij}, \quad i = 1, \dots, 13.$$

در این صورت، جواب واقعی دستگاه $H_1 x = b$ ، بردار $x^* = (1, 1, \dots, 1)^T$ است. اگر این دستگاه را با روش حذفی گاوس حل کنیم جواب به دست آمده به صورت

$$\tilde{x} = (1/0000, 1/0000, 1/0006, 0/9898, 1/0920, 0/5010, 2/7409, \\ -3/0361, 7/2839, -5/4935, 5/2709, -0/6182, 1/2688)^T,$$

خواهد بود. خطای نسبی در جواب به دست آمده برابر با

$$\frac{\|x - \tilde{x}\|_2}{\|x\|_2} = 3/07,$$

می‌باشد. بنابراین، با استفاده از روش حذفی گاوس جوابی با $3/07\%$ خطا به دست آمده است. با توجه به این دو مثال می‌بینیم که استفاده از یک روش تکراری برای دستگاه‌هایی که بزرگ هستند ضروری به نظر می‌رسد. به طور کلی روش‌های تکراری دو دسته‌اند: روش‌های تکراری ایستا و غیرایستا. از روش‌های تکراری ایستا می‌توان به روش‌های ژاکوبی، گاوس-سایدل، SOR ، AOR ، $SSOR$ ، HSS ، و از روش‌های تکراری غیرایستا می‌توان به روش‌های تکراری مانده‌مینیمال، روش سریع‌ترین کاهش، CG ، $GMRES$ ، CGS ، $BICG$ ، $BICGSTAB$ ، و $IHSS$ اشاره کرد [۸۷، ۱۱]. در این فصل به معرفی روش‌های تکراری ایستا در حالت کلی و در حالت خاص روش تکراری AOR می‌پردازیم.

۲-۲ روش-R و خواص آن

دستگاه $Ax = b$ را در نظر بگیرید که در آن $A \in C^{n \times n}$ نامنفرد است. این دستگاه را به صورت

$$x = (I - A)x + b,$$

می‌نویسیم که در آن I ماتریس همانی است. سپس یک روند تکراری به صورت

$$x^{(k+1)} = (I - A)x^{(k)} + b, \quad (1-2)$$

می‌سازیم به طوری که $x^{(0)}$ یک حدس اولیه برای جواب واقعی دستگاه، یعنی $x^* = A^{-1}b$ است. این روش را روش-R می‌نامیم. دلیل این نامگذاری را در ادامه بیان می‌کنیم. قرار می‌دهیم $G_R = I - A$. در این صورت، G_R را ماتریس تکرار روش-R می‌گوییم. در ادامه، می‌خواهیم ببینیم که تحت چه شرایطی روش-R به جواب دستگاه همگراست. برای این کار، ابتدا لم زیر را بیان می‌کنیم.

۶۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

لم ۱۰۲ به‌ازای هر ماتریس $A \in \mathbb{C}^{n \times n}$ و عدد $\epsilon > 0$ ، یک نرم طبیعی مثل $\|\cdot\|$ روی $\mathbb{C}^{n \times n}$ وجود دارد به‌طوری‌که $\|A\| \leq \rho(A) + \epsilon$.

برهان. با استفاده از قضیه ۱۰۱، یک ماتریس نامنفرد مثل P وجود دارد به‌طوری‌که $P^{-1}AP = J$ ، که در آن ماتریس J فرم جردن ماتریس A است. فرض کنید

$$D_\epsilon = \text{diag}_n(1, \epsilon, \epsilon^2, \dots, \epsilon^{n-1}).$$

دراین صورت،

$$(PD_\epsilon)^{-1}A(PD_\epsilon) = D_\epsilon^{-1}JD_\epsilon = \begin{pmatrix} J_\epsilon(\lambda_1) & 0 & \dots & 0 \\ 0 & J_\epsilon(\lambda_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_\epsilon(\lambda_s) \end{pmatrix},$$

و

$$J_\epsilon(\lambda_j) = \begin{pmatrix} J_{1,\epsilon}(\lambda_j) & 0 & \dots & 0 \\ 0 & J_{2,\epsilon}(\lambda_j) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_{t_j,\epsilon}(\lambda_j) \end{pmatrix},$$

که در آن

$$J_r(\lambda_j) = \begin{pmatrix} \lambda_j & \epsilon & & \\ & \lambda_j & \ddots & \\ & & \ddots & \epsilon \\ & & & \lambda_j \end{pmatrix}.$$

تعریف می‌کنیم

$$\|x\|_\epsilon = \|(PD_\epsilon)^{-1}x\|_\infty. \quad (2-2)$$

می‌توان دید که $\|\cdot\|_\epsilon$ یک نرم روی $\mathbb{C}^n$ است. نرم طبیعی متناظر با این نرم برداری به‌صورت

$$\|A\|_\epsilon = \max_{x \neq 0} \frac{\|Ax\|_\epsilon}{\|x\|_\epsilon},$$

تعریف می‌شود. داریم

$$\begin{aligned}
 \|A\|_\epsilon &= \max_{x \neq 0} \frac{\|Ax\|_\epsilon}{\|x\|_\epsilon} \\
 &= \max_{x \neq 0} \frac{\|(PD_\epsilon)^{-1}Ax\|_\infty}{\|(PD_\epsilon)^{-1}x\|_\infty} \\
 &= \max_{x \neq 0} \frac{\|(PD_\epsilon)^{-1}A(PD_\epsilon)y\|_\infty}{\|y\|_\infty} \quad (y = (PD_\epsilon)^{-1}x \neq 0) \\
 &= \|(PD_\epsilon)^{-1}A(PD_\epsilon)\|_\infty \\
 &= \max_{\lambda_j \in \sigma(A)} |\lambda_j| + \epsilon \\
 &= \rho(A) + \epsilon.
 \end{aligned}$$

□ بدین ترتیب اثبات قضیه کامل می‌شود.

لم ۲.۲ فرض کنید $A \in \mathbb{C}^{n \times n}$. در این صورت،

(الف) $\lim_{k \rightarrow \infty} A^k = 0$ اگر و تنها اگر $\rho(A) < 1$.

(ب) سری $\sum_{k=0}^{\infty} A^k$ همگراست اگر و تنها اگر $\rho(A) < 1$. به علاوه، در صورت همگرایی سری، به $(I - A)^{-1}$ همگرا می‌شود.

برهان. (الف) اگر $\rho(A) < 1$ ، آنگاه عدد $\epsilon > 0$ را طوری انتخاب می‌کنیم که $\rho(A) + \epsilon < 1$. با استفاده از لم ۱.۲، یک نرم ماتریسی طبیعی مثل $\|\cdot\|$ وجود دارد به طوری که

$$\|A\| \leq \rho(A) + \epsilon < 1.$$

از این رو، بنا به خاصیت ضربی نرم، وقتی که $k \rightarrow \infty$ ، داریم

$$\|A^k\| \leq \|A\|^k \leq (\rho(A) + \epsilon)^k \rightarrow 0.$$

بنابراین، $\lim_{k \rightarrow \infty} A^k = 0$. برعکس، فرض کنید (λ, x) یک زوج ویژه ماتریس A باشد. در این صورت، وقتی که $k \rightarrow \infty$ ،

$$\lambda^k x = A^k x \rightarrow 0.$$

این رابطه نشان می‌دهد که $\lim_{k \rightarrow \infty} \lambda^k = 0$ و این بدین معنا است که $|\lambda| < 1$. از اینکه λ یک مقدار ویژه دلخواه است، نتیجه می‌گیریم $\rho(A) < 1$.

۶۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

(ب) فرض کنید $\rho(A) < 1$. مقادیر ویژه ماتریس $I - A$ به صورت $1 - \lambda_i$ ، $i = 1, 2, \dots, n$ می‌باشند که در آن یک مقدار ویژه A است. از اینکه $\rho(A) < 1$ داریم $1 - \lambda_i \neq 0$ ، $i = 1, 2, \dots, n$. بنابراین، ماتریس $I - A$ نامنفرد است. اکنون، داریم

$$(I - A) \sum_{i=0}^k A^i = I - A^{k+1}.$$

در نتیجه،

$$\sum_{i=0}^k A^i = (I - A)^{-1} (I - A^{k+1}).$$

لذا، از اینکه $\rho(A) < 1$ داریم $\lim_{k \rightarrow \infty} A^{k+1} = 0$ و در نتیجه،

$$\begin{aligned} \sum_{i=0}^{\infty} A^i &= \lim_{k \rightarrow \infty} \sum_{i=0}^k A^i \\ &= \lim_{k \rightarrow \infty} (I - A)^{-1} (I - A^{k+1}) \\ &= (I - A)^{-1}. \end{aligned}$$

برعکس، فرض کنید سری $\sum_{k=0}^{\infty} A^k$ همگرا و (λ, x) یک زوج ویژه A باشد. در این صورت، داریم

$$\left(\sum_{k=0}^m A^k \right) x = \sum_{k=0}^m A^k x = \sum_{k=0}^m \lambda^k x = \left(\sum_{k=0}^m \lambda^k \right) x.$$

این رابطه نشان می‌دهد، وقتی که $m \rightarrow \infty$ ، سمت چپ این رابطه همگراست. لذا، سری $\sum_{k=0}^{\infty} \lambda^k$ همگراست که نشان می‌دهد $|\lambda| < 1$. در نتیجه، $\rho(A) < 1$. $\square$

قضیه ۱.۲ روش R -به‌ازای هر حدس اولیه $x^{(0)}$ به جواب دستگاه $Ax = b$ همگراست اگر و تنها اگر $\rho(G_R) < 1$.

برهان. فرض کنید روش R -به‌ازای هر حدس اولیه $x^{(0)}$ به جواب دستگاه $Ax = b$ همگرا باشد. با فرض $x^* = A^{-1}b$ داریم

$$x^* = G_R x^* + b.$$

با تفریق طرفین این رابطه و رابطه (۱-۲)، داریم

$$x^{(k+1)} - x^* = G_R(x^{(k)} - x^*),$$

و ازاین‌رو، با کمک استقرا، خواهیم داشت

$$x^{(k+1)} - x^* = G_R^{k+1}(x^{(0)} - x^*).$$

فرض کنید $\rho(G_R) = |\lambda_r|$ و x_r بردار ویژه متناظر به λ_r باشد. با توجه به اینکه حدس اولیه $x^{(0)}$ دلخواه است فرض می‌کنیم $x^{(0)} = x_r + x^*$. دراین‌صورت، داریم

$$\begin{aligned} \circ &= \lim_{k \rightarrow \infty} x^{(k+1)} - x^* \\ &= \lim_{k \rightarrow \infty} G_R^{k+1}(x_r + x^* - x^*) \\ &= \lim_{k \rightarrow \infty} \lambda_r^{k+1} x_r. \end{aligned}$$

حال با توجه به اینکه $\circ \neq x_r$ ، باید داشته باشیم $|\lambda_r| < 1$ ، که اثبات را کامل می‌کند. برعکس، فرض کنید $\rho(G_R) < 1$. به‌سادگی می‌توان دید که

$$x^{(k+1)} = G_R^{k+1}x^{(0)} + \left(\sum_{i=0}^k G_R^i \right) b. \quad (3-2)$$

فرض کنید $\rho(G_R) < 1$. با توجه به لم ۲.۲، داریم

$$\lim_{k \rightarrow \infty} G_R^{k+1} = \circ, \quad \lim_{k \rightarrow \infty} \sum_{i=0}^k G_R^i = (I - G_R)^{-1}.$$

بنابراین، با توجه به (۳-۲)، داریم

$$\lim_{k \rightarrow \infty} x^{(k+1)} = (I - G_R)^{-1}b = A^{-1}b,$$

و بدین ترتیب اثبات قضیه کامل می‌گردد. $\square$

نتیجه ۱.۲ اگر $\| \cdot \|$ یک نرم طبیعی باشد و $\|G_R\| < 1$ ، آنگاه روش-R به‌ازای هر حدس اولیه $x^{(0)}$ به جواب دستگاه $Ax = b$ همگرا است. برهان. با توجه به قضیه ۱۸.۱، داریم

$$\rho(G_R) \leq \|G_R\| < 1.$$

بنابراین، با استفاده از قضیه ۱.۲ همگرایی نتیجه می‌شود. $\square$

همان‌گونه که دیدیم (قضیه ۱.۲) روش R-همگراست اگر و تنها اگر $\rho(I - A) < 1$. روشن است که به‌ازای هر $\lambda \in \sigma(A)$ ، $1 - \lambda$ یک مقدار ویژه $I - A$ است. بنابراین، روش R-همگراست اگر و تنها اگر به‌ازای هر $\lambda \in \sigma(A)$ داشته باشیم $|1 - \lambda| < 1$. با فرض $\lambda = \Re(\lambda) + i\Im(\lambda)$ ، رابطه اخیر معادل با

$$(1 - \Re(\lambda))^2 + \Im(\lambda)^2 < 1,$$

است که نشان می‌دهد روش R-همگراست اگر و تنها اگر مقادیر ویژه A در دایره‌ای به شعاع ۱ و با مرکز $(1, 0)$ قرار داشته باشند.

مثال ۳.۲ فرض کنید درایه‌های ماتریس $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ به‌صورت

$$a_{ij} = \begin{cases} \frac{1}{2} + \frac{3i}{2n}, & i = j, \\ \frac{1}{1 + 4(i-j)}, & i \neq j, \end{cases}$$

تعریف شده باشد و $b = Ae$ که در آن $e = (1, 1, \dots, 1)^T \in \mathbb{R}^n$. فرض کنید $n = 200$. مقادیر ویژه A و $G = I - A$ در شکل ۲-۲ نمایش داده شده‌اند. همان‌گونه که ملاحظه می‌شود مقادیر ویژه A در دایره‌ای به مرکز $(1, 0)$ و به شعاع ۱ قرار دارند. ازاین‌رو، همگرایی روش تضمین می‌شود. اگر روش R-را برای حل دستگاه به‌کار ببریم، خواهیم دید که با حدس اولیه $x^{(0)} = 0$ و شرط توقف

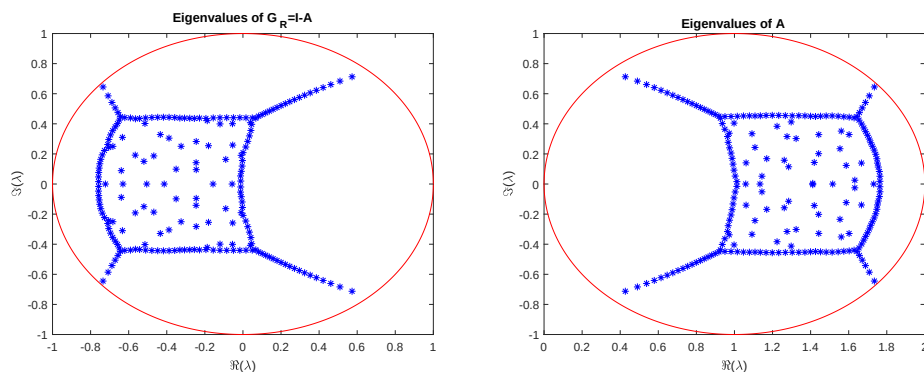
$$\frac{\|b - Ax^{(k)}\|_2}{\|b\|_2} < 10^{-7},$$

در ۸۹۲ تکرار همگرا شده است.

۳-۲ روش‌های تکراری ایستا در حالت کلی

در این بخش شکل کلی یک روش تکراری ایستا را معرفی کرده و نشان می‌دهیم که یک روش تکراری ایستا در حقیقت تعمیمی از روش R-است. ایده ارائه روش‌های تکراری ایستا براساس روش R-از [۵۶] گرفته شده است. ابتدا تعریف یک شکافت را معرفی می‌کنیم.

تعریف ۱.۲ فرض کنید $A \in \mathbb{C}^{n \times n}$. یک تجزیه مثل $A = M - N$ یک شکافت برای A نامیده می‌شود هرگاه M نامنفرد باشد.



شکل ۲-۲ نمایش مقادیر ویژه A و $G = I - A$ برای مثال ۳.۲

برای معرفی یک روش تکراری ایستا در حالت کلی، فرض کنید $A = M - N$ یک شکافت از ماتریس A باشد. نخست دستگاه $Ax = b$ را به صورت

$$(M - N)x = b,$$

و سپس به صورت

$$x = M^{-1}Nx + M^{-1}b,$$

می‌نویسیم. اکنون، شکل کلی یک روند تکراری ایستا به صورت

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b, \quad k = 0, 1, \dots, \quad (۴-۲)$$

نوشته می‌شود. با فرض $f^{(k)} = Nx^{(k)} + b$ رابطه (۴-۲) به صورت

$$Mx^{(k+1)} = f^{(k)}, \quad (۵-۲)$$

نوشته می‌شود. این رابطه نشان می‌دهد که برای محاسبه $x^{(k+1)}$ باید دستگاه (۵-۲) حل شود. از این رو، ماتریس M طوری انتخاب می‌شود که دستگاه (۵-۲) به راحتی حل شود. معمولاً ماتریس M یک ماتریس بالامثلثی، پائین‌مثلثی، یا یک حاصلضرب از این دو ماتریس انتخاب می‌شود. با جایگذاری ماتریس $N = M - A$ در رابطه (۵-۲) خواهیم داشت

$$x^{(k+1)} = (I - M^{-1}A)x^{(k)} + M^{-1}b. \quad (۶-۲)$$

۶۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

با مقایسه رابطه (۶-۲) و روش-R می‌بینیم که روش تکراری ایستای (۵-۲) چیزی نیست، جز همان روش-R برای حل دستگاه

$$M^{-1}Ax = M^{-1}b. \quad (۷-۲)$$

دستگاه (۷-۲) را دستگاه پیش‌شرط‌سازی شده $Ax = b$ و ماتریس M را یک پیش‌شرط‌ساز می‌گویند.

رابطه (۶-۲) را می‌توان به صورت

$$x^{(k+1)} = Gx^{(k)} + f, \quad k = 0, 1, \dots, \quad (۸-۲)$$

نوشت که در آن

$$G = M^{-1}N = I - M^{-1}A,$$

ماتریس تکرار روش است و $f = M^{-1}b$. با توجه به قضیه ۱۰.۲، می‌بینیم که روند تکراری (۸-۲) به‌ازای هر حدس اولیه دلخواه همگراست اگر و تنها اگر $\rho(G) < 1$.

مثال ۴.۲ دستگاه معادلات خطی

$$Ax = \begin{pmatrix} 2 & 3 & -1 \\ 1 & 2 & 2 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 10 \\ 12 \end{pmatrix} = b,$$

را در نظر بگیرید. با فرض

$$M = \begin{pmatrix} 2 & 3 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -2 \\ -1 & -2 & 0 \end{pmatrix},$$

داریم $A = M - N$. از اینکه ماتریس M نامنفرد است نتیجه می‌گیریم که این تجزیه یک شکافت از ماتریس A است. از طرفی،

$$G = M^{-1}N = \frac{1}{4} \begin{pmatrix} 0 & 0 & 32 \\ 0 & 0 & -20 \\ -1 & -2 & 0 \end{pmatrix}, \quad f = M^{-1}b = \begin{pmatrix} -6 \\ 8 \\ 3 \end{pmatrix},$$

و

$$\sigma(G) = \left\{ 0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right\}.$$

بنابراین،

$$\rho(G) = \frac{\sqrt{2}}{2} \approx 0.7071 < 1,$$

که نشان می‌دهد روند تکراری

$$x^{(k+1)} = Gx^{(k)} + f, \quad (9-2)$$

به‌ازای هر حدس اولیه دلخواه به جواب دستگاه $Ax = b$ همگراست. با جایگذاری ماتریس G و بردار f در رابطه (۹-۲) خواهیم داشت

$$\begin{pmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 0 & 0 & 32 \\ 0 & 0 & -20 \\ -1 & -2 & 0 \end{pmatrix} \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{pmatrix} + \begin{pmatrix} -6 \\ 8 \\ 3 \end{pmatrix},$$

یا

$$\begin{cases} x_1^{(k+1)} = 8x_3^{(k)} - 6, \\ x_2^{(k+1)} = -5x_3^{(k)} + 8, \\ x_3^{(k+1)} = \frac{1}{4}(-x_1^{(k)} - 2x_2^{(k)}) + 3. \end{cases}$$

حدس اولیه را بردار

$$x^{(0)} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

انتخاب می‌کنیم. در این صورت، به‌ازای $k = 0$ ، داریم

$$\begin{cases} x_1^{(1)} = 8x_3^{(0)} - 6 = -6, \\ x_2^{(1)} = -5x_3^{(0)} + 8 = 8, \\ x_3^{(1)} = \frac{1}{4}(-x_1^{(0)} - 2x_2^{(0)}) + 3 = 3. \end{cases}$$

همچنین، به‌ازای $k = ۱$ داریم

$$\begin{cases} x_1^{(2)} = ۸x_3^{(1)} - ۶ = ۸ \times ۳ - ۶ = ۱۸, \\ x_2^{(2)} = -۵x_3^{(1)} + ۸ = -۵ \times ۳ + ۸ = -۷, \\ x_3^{(2)} = \frac{1}{4}(-x_1^{(1)} - ۲x_2^{(1)}) + ۳ = \frac{1}{4}(۶ - ۲ \times ۸) + ۳ = ۰.۵. \end{cases}$$

با ادامه این روند در تکرار سی‌ام خواهیم داشت (با دقت چهار رقم اعشار)

$$x^{(30)} = \begin{pmatrix} ۲.۰۰۱۰ \\ ۲.۹۹۹۴ \\ ۱.۰۰۰۰ \end{pmatrix}.$$

لازم به ذکر است که جواب واقعی دستگاه بردار $x^* = (۲, ۳, ۱)^T$ است. ملاحظه می‌کنیم که روند تکراری داده‌شده جواب مناسبی به‌دست داده است.

۲-۳-۱ نرخ همگرایی

در این زیربخش، می‌خواهیم ببینیم زمانی که یک روش تکراری ایستا مثل (۲-۸) همگراست، چگونه می‌توان سرعت همگرایی آن را برآورد کرد. تعیین یک معیار مناسب برای سرعت همگرایی می‌تواند ما را در مقایسه روش‌ها کمک کند. فرض کنید دنباله (۲-۸) به‌ازای هر حدس اولیه دلخواه به جواب دستگاه $Ax = b$ ، یعنی $x^* = A^{-1}b$ ، همگرا باشد. در این صورت، روشن است که

$$x^* = Gx^* + f.$$

با تفریق طرفین این رابطه و رابطه (۲-۸)، خواهیم داشت

$$x^{(k+1)} - x^* = G(x^{(k)} - x^*).$$

از این رو، داریم $e^{(k+1)} = Ge^{(k)}$ ، که در آن $e^{(k)} = x^{(k)} - x^*$. به‌استقرا می‌توان دید که

$$e^{(k)} = G^k e^{(0)}, \quad k = ۱, ۲, \dots$$

در نتیجه،

$$\|e^{(k)}\| \leq \|G^k\| \|e^{(0)}\|, \quad (۲-۱۰)$$

که در آن، نرم ماتریسی مورد استفاده یک نرم طبیعی است که توسط نرم برداری استفاده شده تولید شده است. این رابطه نشان می‌دهد که سرعت همگرایی دنباله $x^{(k)}$ به جواب دستگاه $Ax = b$ ، با سرعت همگرایی دنباله $\|G^k\|$ به صفر ارتباط مستقیم دارد. به عبارت دیگر، سرعت همگرایی $\|G^k\|$ به صفر، سرعت همگرایی دنباله $\{x^k\}$ را توصیف می‌کند. از این رو، برای به دست آوردن یک معیار مناسب برای برآورد سرعت همگرایی دنباله (۲-۸)، روی سرعت همگرایی $\|G^k\|$ به ماتریس صفر تمرکز می‌کنیم. نخست فرض می‌کنیم ماتریس G نرمال باشد. در این صورت، بنا به قضیه ۱۷.۱ یک ماتریس یکانی مثل U وجود دارد به گونه‌ای که $G = UDU^H$ ، که در آن D یک ماتریس قطری است که درایه‌های قطری آن مقادیر ویژه G هستند. از این رو، داریم

$$\|G^k\|_2 = \|UD^kU^H\|_2 = \|D^k\|_2 = \|D\|_2^k = \rho(G)^k.$$

در نتیجه،

$$\lim_{k \rightarrow \infty} \|G^k\|_2 = \lim_{k \rightarrow \infty} \rho(G)^k = 0.$$

توجه کنید که دنباله $\|G^k\|_2$ یک دنباله اکیداً نزولی است که به صفر همگراست. بنابراین، در حالتی که ماتریس G نرمال باشد، همه چیز مطلوب خواهد بود و شعاع طیفی ماتریس G به تنهایی همگرایی را توصیف می‌کند. اما این موضوع برای یک ماتریس دلخواه درست نیست. به مثال زیر توجه کنید.

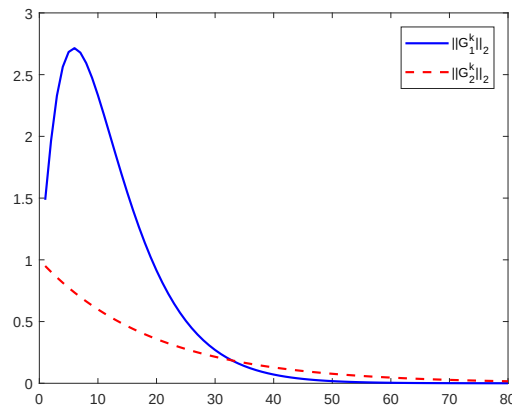
مثال ۵.۲ فرض کنید ماتریس‌های تکرار دو روش تکراری ایستا به صورت

$$G_1 = \begin{pmatrix} 0.85 & 1 \\ 0 & 0.85 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 0.85 & 0 \\ 0 & 0.95 \end{pmatrix},$$

باشند. داریم

$$\rho(G_1) = 0.85 < \rho(G_2) = 0.95 < 1.$$

بنابراین، همگرایی هر دو روش تضمین می‌شود. در شکل ۲-۳ مقادیر $\|G_1^k\|_2$ و $\|G_2^k\|_2$ به ازای $k = 1, 2, \dots, 80$ نمایش داده شده‌اند. این شکل نشان می‌دهد با اینکه شعاع طیفی ماتریس G_1 از شعاع طیفی ماتریس G_2 کوچکتر است، می‌بینیم که به ازای $1 \leq k \leq 32$ ، مقادیر $\|G_1^k\|_2$ از $\|G_2^k\|_2$ بیشتر است. به علاوه، همگرایی $\|G_1^k\|_2$ به بردار صفر یکنوا نیست. اما باید توجه کرد که به ازای مقادیر بزرگ k نه تنها $\|G_1^k\|_2$ کمتر $\|G_2^k\|_2$ است، بلکه هر دو به صفر نزدیک هستند. این مثال نشان می‌دهد که



شکل ۳-۲ نمودار مقادیر $\|G_1^k\|_2$ و $\|G_2^k\|_2$ بر حسب k برای مثال ۵.۲.

طیف ماتریس به‌تنهایی سرعت همگرایی را توصیف نمی‌کند، هرچند این موضوع برای زمانی که k به سمت بینهایت میل می‌کند درست است.

اکنون، برای به‌دست آوردن یک معیار مناسب به رابطه (۱۰-۲) برمی‌گردیم. این رابطه را می‌توان به‌صورت

$$\eta_k := \left(\frac{\|e^{(k)}\|}{\|e^{(0)}\|} \right)^{\frac{1}{k}} \leq \|G^k\|^{\frac{1}{k}}, \quad (11-2)$$

نوشت. روشن است که عدد η_k میانگین ضریب کاهش نرم خطا در k تکرار است. از رابطه (۱۱-۲)، داریم

$$\eta_k \leq \log_{10} \|G^k\|^{\frac{1}{k}} = \log_{10} \|G^k\|^{\frac{1}{k}} = 10^{-R_k}, \quad (12-2)$$

که در آن

$$R_k = -\frac{1}{k} \log_{10} \|G^k\|. \quad (13-2)$$

رابطه (۱۲-۲) نشان می‌دهد که به‌طور متوسط در هر تکرار نرم بردار خطا به‌نسبت R_k کاهش می‌یابد. عدد R_k را میانگین نرخ همگرایی در k تکرار می‌گویند. روشن است که محاسبه R_k بسیار پرهزینه است. قضیه زیر این مشکل را حل می‌کند.

قضیه ۲.۲ برای هر نرم طبیعی $\|\cdot\|$ ، داریم

$$\lim_{k \rightarrow \infty} \|G^k\|^{\frac{1}{k}} = \rho(G).$$

برهان. نخست، توجه کنید که

$$\rho(G)^k = \rho(G^k) \leq \|G^k\| \implies \rho(G) \leq \|G^k\|^{\frac{1}{k}}.$$

اکنون، فرض کنید که $\epsilon > 0$ داده شده باشد. در این صورت، داریم

$$\rho\left(\frac{1}{\rho(G) + \epsilon} G\right) = \frac{\rho(G)}{\rho(G) + \epsilon} < 1.$$

بنابراین، با استفاده از قسمت نخست از لم ۲.۲، داریم

$$\lim_{k \rightarrow \infty} \left(\frac{G}{\rho(G) + \epsilon}\right)^k = 0,$$

یا

$$\lim_{k \rightarrow \infty} \frac{\|G^k\|}{(\rho(G) + \epsilon)^k} = 0.$$

در نتیجه، یک عدد طبیعی مثل N_ϵ وجود دارد به طوری که به ازای هر $k \geq N_\epsilon$ ، داریم

$$\frac{\|G^k\|}{(\rho(G) + \epsilon)^k} < 1,$$

یا

$$\|G^k\|^{\frac{1}{k}} < \rho(G) + \epsilon.$$

با جمع‌بندی این مطالب، داریم

$$\forall k \geq N_\epsilon : \rho(G) \leq \|G^k\|^{\frac{1}{k}} < \rho(G) + \epsilon.$$

□ اکنون، با توجه به اینکه ϵ دلخواه است، داریم $\rho(G) = \lim_{k \rightarrow \infty} \|G^k\|^{\frac{1}{k}}$.

اکنون، با استفاده از این قضیه، با جایگزین کردن $\rho(G)$ به جای $\|G^k\|^{\frac{1}{k}}$ در R_k ، نرخ همگرایی

مجانبی به دست می‌آید. در حقیقت، نرخ همگرایی مجانبی برابر است با

$$R_\infty = \lim_{k \rightarrow \infty} R_k = - \lim_{k \rightarrow \infty} \log_{10} \|G^k\|^{\frac{1}{k}} = - \log_{10} \rho(G).$$

در مثال ۵.۲، داریم

$$R_{\infty}^1 = -\log_{10} \rho(G_1) = -\log_{10} 0.85 = 0.07,$$

$$R_{\infty}^2 = -\log_{10} \rho(G_2) = -\log_{10} 0.95 = 0.02.$$

بنابراین به‌طور مجانبی نرخ همگرایی دنباله متناظر به G_1 بیشتر از دنباله متناظر به G_2 است.

۲-۳-۲ انتخاب ماتریس M

روند تکراری (۲-۶) همگراست اگر و تنها اگر مقادیر ویژه $M^{-1}A$ در دایره‌ای به مرکز $(1, 0)$ و به شعاع ۱ قرار داشته باشند، یا به عبارت دیگر مقادیر ویژه در درون این دایره، حول نقطه $(1, 0)$ خوشه شده باشند. علاوه بر این، در حالتی که ماتریس $M^{-1}A$ نرمال باشد، هرچه مقادیر ویژه ماتریس $M^{-1}A$ حول نقطه $(1, 0)$ بهتر خوشه شده باشند، آنگاه همگرایی سریع‌تر خواهد بود. این نتیجه برای یک ماتریس دلخواه به صورت مجانبی درست است. از این رو، ماتریس پیش‌شرط‌ساز M باید طوری انتخاب شود که مقادیر ویژه $M^{-1}A$ حول نقطه $(1, 0)$ بهتر خوشه شوند. در بهترین حالت اگر $M = A$ اختیار شود، آنگاه خواهیم داشت $M^{-1}A = I$ و تمام مقادیر ویژه $M^{-1}A$ برابر با ۱ خواهند بود و

$$M^{-1}N = I - M^{-1}A = 0.$$

این رابطه نشان می‌دهد که همه مقادیر ویژه ماتریس تکرار برابر با صفر خواهند بود و روش تکراری، سریع‌ترین همگرایی ممکن را خواهد داشت. اما در عمل برای ماتریس‌های بزرگ محاسبه A^{-1} مقرون به صرفه نیست. بنابراین، ماتریس M یک تقریب از A در نظر گرفته می‌شود. همان‌گونه که پیش‌تر ذکر شد ماتریس M همچنین باید طوری انتخاب شود که یک دستگاه با ماتریس ضرایب M به سادگی حل شود. در عمل معمولاً ماتریس M یک ماتریس بالامثلثی، پائین‌مثلثی، یا یک حاصلضرب از این دو ماتریس انتخاب می‌شود. با در نظر گرفتن تقریب‌های مختلف روش‌های تکراری مختلف از جمله روش‌های تکراری ریچاردسون، ژاکوبی، گاوس-سایدل، SOR و AOR به دست می‌آیند. در ادامه، این روش‌ها را مرور می‌کنیم.

۲-۳-۳ روش تکراری ریچاردسون

فرض کنید $0 < \alpha$. داریم

$$A = \frac{1}{\alpha}I - \left(\frac{1}{\alpha}I - A\right) = M - N,$$

که در آن

$$M = \frac{1}{\alpha}I, \quad N = \frac{1}{\alpha}I - A.$$

ازاین‌رو، روش-R برای حل دستگاه $M^{-1}Ax = M^{-1}b$ به‌صورت

$$x^{(k+1)} = (I - \alpha M^{-1}A)x^{(k)} + M^{-1}b = (I - \alpha A)x^{(k)} + \alpha b,$$

خواهد بود. این روش به روش تکراری ریچاردسون معروف است. به‌ازای $\alpha = 1$ روش ریچاردسون معادل با روش-R است و دلیل نامگذاری روش-R به این نام، همین است.

۲-۳-۴ روش تکراری AOR

همان‌گونه که پیش‌تر اشاره شد، روش تکراری فوق تخفیف شتاب داده‌شده (AOR) توسط حاجیدیموس در [۴۵] ارائه شده است. ماتریس ضرایب دستگاه $Ax = b$ را به‌صورت

$$A = D - E - F,$$

تجزیه می‌کنیم که در آن D یک ماتریس قطری، E پائین‌مثلثی اکید و F بالامثلثی اکید است. فرض کنید درایه‌های قطری ماتریس A ناصفر باشند. به‌ازای $w \in \mathbb{R}, w \neq 0$ ، شکافت

$$A = \frac{1}{\omega}(D - \gamma E) - \frac{1}{\omega}[(1 - \omega)D + (\omega - \gamma)E + \omega F] = M_{\gamma, \omega} - N_{\gamma, \omega},$$

را در نظر می‌گیریم که در آن

$$M_{\gamma, \omega} = \frac{1}{\omega}(D - \gamma E), \quad N_{\gamma, \omega} = \frac{1}{\omega}[(1 - \omega)D + (\omega - \gamma)E + \omega F],$$

و $\gamma \in \mathbb{R}$ یک پارامتر دلخواه است. ازاین‌رو، روش-R برای حل دستگاه

$$M_{\gamma, \omega}^{-1}Ax = M_{\gamma, \omega}^{-1}b,$$

به‌صورت

$$x^{(k+1)} = M_{\gamma, \omega}^{-1}N_{\gamma, \omega}x^{(k)} + M_{\gamma, \omega}^{-1}b = (I - M_{\gamma, \omega}^{-1}A)x^{(k)} + M_{\gamma, \omega}^{-1}b,$$

خواهد بود. این روش به روش تکراری AOR معروف است. با فرض

$$\mathcal{L}_{\gamma, \omega} = M_{\gamma, \omega}^{-1}N_{\gamma, \omega}, \quad f_{\gamma, \omega} = M_{\gamma, \omega}^{-1}b,$$

می‌توان روش تکراری AOR را به صورت

$$x^{(k+1)} = \mathcal{L}_{\gamma, \omega} x^{(k)} + f_{\gamma, \omega}, \quad (14-2)$$

جمع‌بندی نمود. توجه کنید که اگر قرار دهیم $U = D^{-1}F$ و $L = D^{-1}E$ ، آنگاه

$$\mathcal{L}_{\gamma, \omega} = (I - \gamma L)^{-1}[(1 - \omega)I + (\omega - \gamma)L + \omega U].$$

در روش AOR، ω را پارامتر فوق‌تخفیف و γ را پارامتر شتاب می‌گویند. به‌ازای مقادیر خاصی از پارامترهای ω و γ ، روش AOR به روش‌های مشهور زیر منجر می‌شود
۱. اگر $\gamma = 0$ و $\omega = 1$ ، آنگاه روش AOR هم‌ارز با روش ژاکوبی است، که به صورت

$$x^{(k+1)} = D^{-1}(E + F)x^{(k)} + D^{-1}b,$$

نوشته می‌شود. درواقع روش ژاکوبی همان روش-R برای حل دستگاه معادلات خطی $D^{-1}Ax = D^{-1}b$ است. این بدین معنا است که روش-R برای حل دستگاه $Ax = b$ با پیش‌شرط‌ساز $M = D$ به کار رفته است. در اینجا توجه کنید که $N = M - A = E + F$.

۲. اگر $\gamma = 1$ و $\omega = 1$ ، آنگاه روش AOR هم‌ارز با روش گaus-سایدل است، که به صورت

$$x^{(k+1)} = (D - E)^{-1}Fx^{(k)} + (D - E)^{-1}b,$$

نوشته می‌شود. همانند روش ژاکوبی روشن است که روش گaus-سایدل همان روش-R برای حل دستگاه $Ax = b$ با پیش‌شرط‌ساز $M = D - E$ می‌باشد. لازم به ذکر است که $N = M - A = F$.
۳. اگر $\gamma = \omega$ ، آنگاه روش AOR هم‌ارز با روش فوق‌تخفیف متوالی (SOR) است، که به صورت

$$x^{(k+1)} = (D - \omega E)^{-1}[(1 - \omega)D + \omega F]x^{(k)} + (D - \omega E)^{-1}\omega b,$$

خلاصه می‌شود. روشن است که روش SOR چیزی نیست جز روش-R برای حل دستگاه $Ax = b$ با پیش‌شرط‌ساز

$$M_{\omega} = \frac{1}{\omega}(D - \gamma E).$$

همچنین، اگر قرار دهیم

$$N_{\omega} = M_{\omega} - A = \frac{1}{\omega}((1 - \omega)D + \omega F),$$

آنگاه $A = M_\omega - N_\omega$ شکافت مربوطه خواهد بود.

در ادامه، برخی از خواص این روش‌ها را مطالعه می‌کنیم. برای جزئیات بیشتر به [۳، ۹۳] مراجعه نمایید.

مثال ۶.۲ مثال ۳.۲ را در نظر گرفته و دستگاه حاصل را با روش R -با پیش‌شرط‌سازهای

$$(الف) \quad M = \frac{3}{4}I \quad (\text{روش ریچاردسون با پارامتر } \alpha = \frac{2}{3}),$$

$$(ب) \quad M = D \quad (\text{روش ژاکوبی}),$$

$$(ج) \quad M = D - E \quad (\text{روش گاوس-سایدل}),$$

$$(د) \quad M = \frac{1}{0.8}(D - 0.8E) \quad (\text{روش SOR با پارامتر } \omega = 0.8),$$

$$(ه) \quad M = \frac{1}{0.8}(D - 0.6E) \quad (\text{روش AOR با پارامترهای } \omega = 0.8 \text{ و } \gamma = 0.6),$$

حل می‌کنیم. سایر شرایط همانند مثال ۳.۲ هستند.

در شکل ۴-۲ توزیع مقادیر ویژه $M^{-1}A$ به‌ازای پیش‌شرط‌سازهای مختلف نمایش داده شده است. روش‌های ریچاردسون، گاوس-سایدل، SOR و AOR به ترتیب ۱۱۰۰، ۲۶، ۲۷ و ۲۶ تکرار به جواب تقریبی مورد نظر همگرا شده‌اند، درحالی‌که روش ژاکوبی واگرا شده است. شکل ۴-۲ به‌خوبی این موضوع را تأیید می‌کند.

۵-۳-۲ روش تکراری SSOR

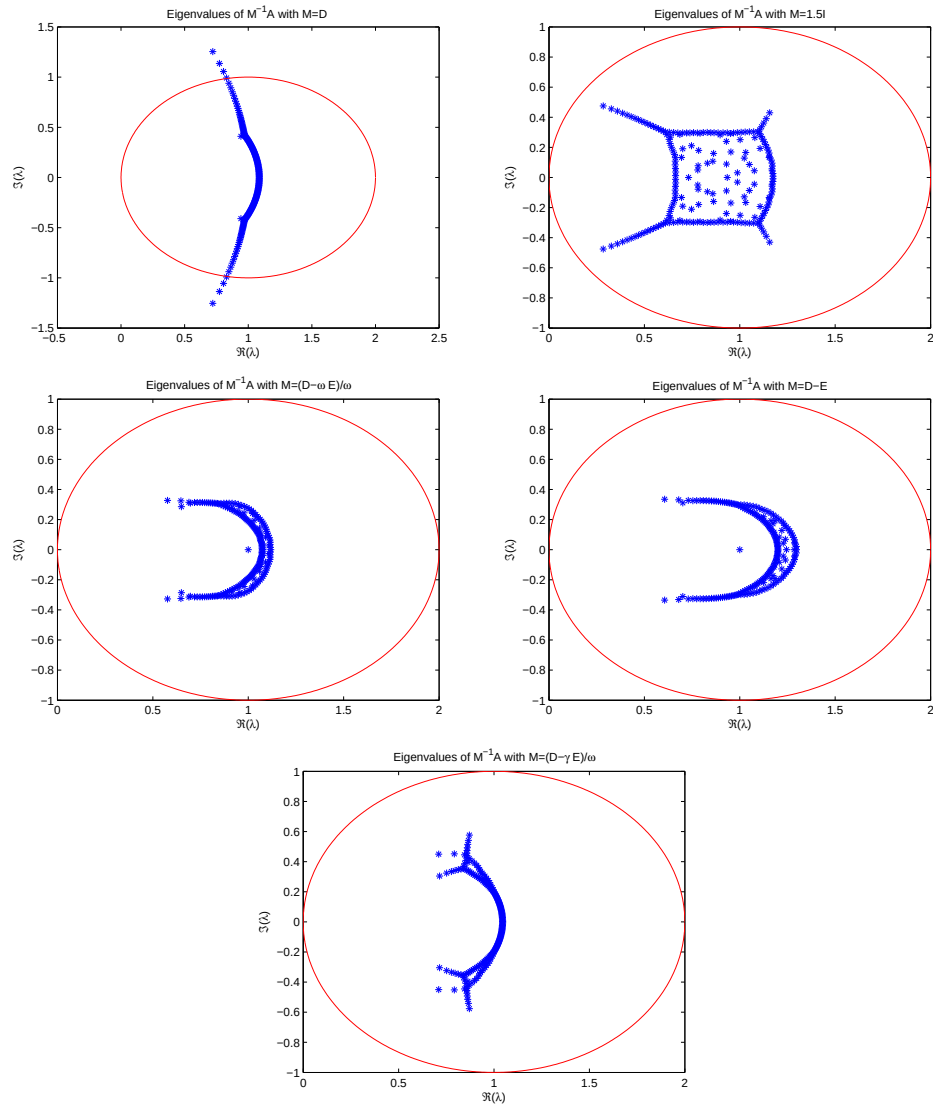
روش تکراری SOR متقارن یا به‌طور خلاصه روش تکراری SSOR یک روش تکراری دومرحله‌ای است، به این معنا که هر تکرار آن از دو گام تشکیل شده است. دو شکافت

$$\begin{cases} A = \frac{1}{\omega}(D - \omega E) - \frac{1}{\omega}((1 - \omega)D + \omega F), \\ A = \frac{1}{\omega}(D - \omega F) - \frac{1}{\omega}((1 - \omega)D + \omega E), \end{cases}$$

از ماتریس A را در نظر گرفته و همانند روش تکراری SOR، روش تکراری دوگامی

$$\begin{cases} (D - \omega E)x^{(k+\frac{1}{2})} = ((1 - \omega)D + \omega F)x^{(k)} + \omega b, \\ (D - \omega F)x^{(k+1)} = ((1 - \omega)D + \omega E)x^{(k+\frac{1}{2})} + \omega b, \end{cases} \quad (۱۵-۲)$$

را می‌سازیم. روند تکراری (۱۵-۲) به روش تکراری SSOR معروف است. در گام نخست روش، بردار $x^{(k+\frac{1}{2})}$ با حل یک دستگاه با ماتریس ضرایب پائین‌مثلثی $(D - \omega E)$ با استفاده از جایگذاری پیشرو



شکل ۴-۲ توزیع مقادیر ویژه $M^{-1}A$ برای مثال ۶.۲.

به دست می‌آید. سپس، با جایگذاری بردار $x^{(k+\frac{1}{p})}$ در گام دوم روش و با حل یک دستگاه با ماتریس ضرایب بالامثلثی $(D - \omega F)$ با استفاده از جایگذاری پسرو، بردار $x^{(k+1)}$ محاسبه می‌شود. با محاسبه $x^{(k+\frac{1}{p})}$ از گام نخست و جایگذاری در گام دوم یک روند یک‌مرحله‌ای به صورت

$$x^{(k+1)} = \mathcal{G}_\omega x^{(k)} + g_\omega,$$

به دست می‌آید که در آن

$$\mathcal{G}_\omega = (D - \omega F)^{-1}((1 - \omega)D + \omega E)(D - \omega E)^{-1}((1 - \omega)D + \omega F),$$

ماتریس تکرار روش تکراری SSOR است و

$$g_\omega = \omega(D - \omega F)^{-1}(I + ((1 - \omega)D + \omega E))(D - \omega E)^{-1}b.$$

در ادامه، شکافت متناظر با این روش را به دست می‌آوریم. داریم

$$\begin{aligned}\mathcal{G}_\omega &= (D - \omega F)^{-1}((1 - \omega)D + \omega E)(D - \omega E)^{-1}((1 - \omega)D + \omega F) \\ &= (I - \omega U)^{-1}((1 - \omega)I + \omega L)(I - \omega L)^{-1}((1 - \omega)I + \omega U),\end{aligned}$$

که در آن $L = D^{-1}E$ و $U = D^{-1}F$. با توجه به تمرین ۳.۱ می‌بینیم که

$$((1 - \omega)I + \omega L)(I - \omega L)^{-1} = (I - \omega L)^{-1}((1 - \omega)I + \omega L).$$

در نتیجه،

$$\begin{aligned}\mathcal{G}_\omega &= (I - \omega U)^{-1}(I - \omega L)^{-1}((1 - \omega)I + \omega L)((1 - \omega)I + \omega U), \\ &= ((I - \omega L)(I - \omega U))^{-1}((1 - \omega)I + \omega L)((1 - \omega)I + \omega U) \\ &= \left((D - \omega E)D^{-1}(I - \omega F)\right)^{-1}((1 - \omega)D + \omega E)D^{-1}((1 - \omega)D + \omega F).\end{aligned}$$

با فرض

$$\hat{M} = (D - \omega E)D^{-1}(D - \omega F),$$

$$\hat{N} = ((1 - \omega)D + \omega E)D^{-1}((1 - \omega)D + \omega F),$$

داریم $\mathcal{G}_\omega = \hat{M}^{-1}\hat{N}$. با کمی محاسبه، می‌توان دید که

$$\hat{M} - \hat{N} = \omega(2 - \omega)A.$$

لذا،

$$A = \frac{1}{\omega(2 - \omega)}\hat{M} - \frac{1}{\omega(2 - \omega)}\hat{N}.$$

درنتیجه، با فرض

$$M = \frac{1}{\omega(2-\omega)} \hat{M} = \frac{1}{\omega(2-\omega)} (D - \omega E) D^{-1} (D - \omega F),$$

$$N = \frac{1}{\omega(2-\omega)} \hat{N} = \frac{1}{\omega(2-\omega)} ((1-\omega)D + \omega E) D^{-1} ((1-\omega)D + \omega F),$$

خواهیم داشت $A = M - N$ و $G_\omega = M^{-1}N$.

در حالت خاص، اگر $\omega = 1$ ، آنگاه روش تکراری SSOR به روش تکراری گاوس-سایدل متقارن (SGS) تبدیل می‌شود.

۴-۲ همگرایی روش تکراری ریچاردسون

فرض کنید

$$G_R(\alpha) = I - \alpha A,$$

ماتریس روش تکراری ریچاردسون باشد. به سادگی می‌توان دید که هر مقدار ویژه $G_R(\alpha)$ به صورت $\mu = 1 - \alpha\lambda$ است که در آن λ یک مقدار ویژه A می‌باشد. فرض می‌کنیم مقادیر ویژه A حقیقی باشند. درنتیجه، خواهیم داشت

$$1 - \alpha\lambda_{\max} \leq 1 - \alpha\lambda \leq 1 - \alpha\lambda_{\min},$$

که در آن $\lambda_{\max}$ و $\lambda_{\min}$ به ترتیب کوچک‌ترین و بزرگ‌ترین مقادیر ویژه A هستند. بنابراین، روش ریچاردسون همگرا است اگر و تنها اگر

$$1 - \alpha\lambda_{\min} < 1, \quad -1 < 1 - \alpha\lambda_{\max}.$$

نامساوی نخست با $\lambda_{\min} > 0$ معادل است. درنتیجه، نامساوی دوم با

$$\alpha < \frac{2}{\lambda_{\max}},$$

معادل خواهد بود. بنابراین، برای $\alpha > 0$ ، روش ریچاردسون همگرا است اگر و تنها اگر

$$\lambda_{\min} > 0, \quad \alpha < \frac{2}{\lambda_{\max}}.$$

به‌عنوان یک نتیجه، می‌توان گفت که اگر ماتریس A معین مثبت متقارن باشد و

$$0 < \alpha < \frac{2}{\lambda_{\max}},$$

آنگاه روش ریچاردسون همگرا خواهد بود.

در ادامه مقدار بهینه α را محاسبه می‌کنیم. فرض کنید کوچک‌ترین مقدار ویژه A مثبت باشد و

$$0 < \alpha < \frac{2}{\lambda_{\max}}.$$

به‌سادگی می‌توان دید که شعاع طیفی ماتریس $G_R(\alpha)$ در رابطه زیر صدق می‌کند

$$\rho(G_R(\alpha)) = \max\{|1 - \alpha\lambda_{\max}|, |1 - \alpha\lambda_{\min}|\}.$$

فرض کنید

$$\rho(G_R(\alpha_{\text{opt}})) = \min_{0 < \alpha < \frac{2}{\lambda_{\max}}} \rho(G_R(\alpha)).$$

برای محاسبه α_{opt} از نمودار مربوط به $y = |1 - \alpha\lambda_{\max}|$ و $y = |1 - \alpha\lambda_{\min}|$ کمک می‌گیریم. تابع اول از نقطه $(0, 1/\lambda_{\max})$ و تابع دوم از $(0, 1/\lambda_{\min})$ و هر دو تابع از $(0, 1)$ می‌گذرند. شکل ۵-۲ را ملاحظه کنید. در این شکل، نمودار مربوط به $\rho(G_R(\alpha))$ پررنگ شده است. به‌وضوح α_{opt} در رابطه زیر صدق می‌کند

$$1 - \alpha_{\text{opt}}\lambda_{\min} = -(1 - \alpha_{\text{opt}}\lambda_{\max}),$$

که با محاسبه α_{opt} خواهیم داشت

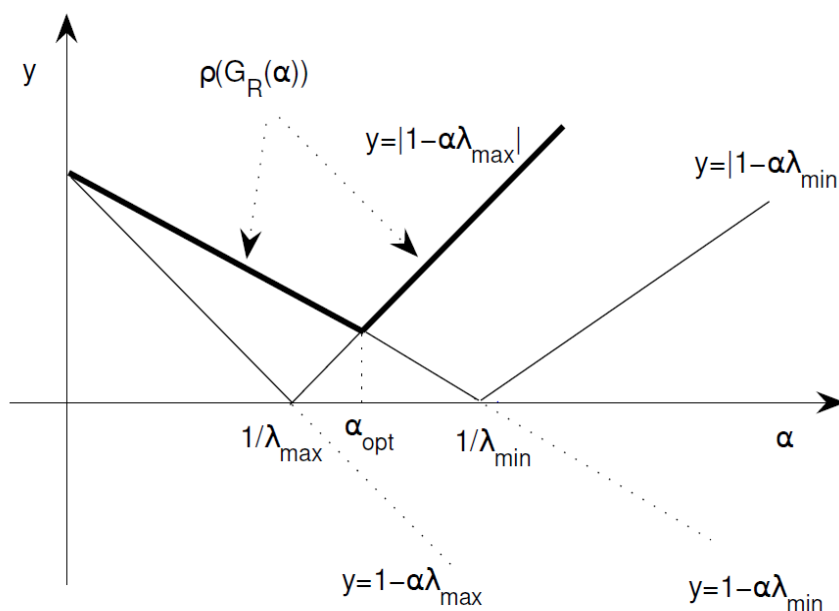
$$\alpha_{\text{opt}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}.$$

در این صورت، داریم

$$\begin{aligned} \rho(G_R(\alpha_{\text{opt}})) &= 1 - \frac{2}{\lambda_{\min} + \lambda_{\max}} \lambda_{\min} \\ &= \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\lambda_{\max}/\lambda_{\min} - 1}{\lambda_{\max}/\lambda_{\min} + 1}. \end{aligned}$$

این رابطه نشان می‌دهد اگر $\lambda_{\max}/\lambda_{\min}$ خیلی بزرگ باشد، آنگاه $\rho(G_R(\alpha_{\text{opt}})) \approx 1$. در نتیجه، سرعت همگرایی روش ریچاردسون بسیار کند خواهد بود. در صورتی که ماتریس A متقارن باشد، خواهیم

۸۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ



شکل ۵-۲ نمودار $\rho(G_R(\alpha_{opt}))$.

داشت

$$\text{cond}_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\lambda_{\max}}{\lambda_{\min}},$$

و در نتیجه، خواهیم داشت

$$\rho(G_R(\alpha_{opt})) = \frac{\text{cond}_2(A) - 1}{\text{cond}_2(A) + 1}.$$

بنابراین، اگر ماتریس A بد حالت باشد، یعنی اگر $\text{cond}_2(A)$ خیلی بزرگ باشد، آنگاه سرعت همگرایی بسیار کند خواهد بود.

۵-۲ همگرایی روش‌ها برای ماتریس‌های معین مثبت هرمیتی

نخست همگرایی روش تکراری SOR را برای ماتریس‌های معین مثبت هرمیتی بررسی می‌کنیم. برای این کار، قضیه بعد را معرفی می‌کنیم.

قضیه ۳.۲ فرض کنید A یک ماتریس معین مثبت هرمیتی و $A = M - N$ یک شکافت از آن باشد. در این صورت، ماتریس $M^H + N$ هرمیتی است. به علاوه، اگر $M^H + N$ معین مثبت باشد، آنگاه

$$\rho(M^{-1}N) < 1.$$

برهان. ابتدا توجه کنید که $N = M - A$. بنابراین،

$$M^H + N = M^H + M - A.$$

با توجه به اینکه $M^H + M$ و A هرمیتی هستند، نتیجه می‌گیریم که $M^H + N$ نیز هرمیتی است. برای اثبات قسمت دوم قضیه، نرم برداری

$$\|x\|_A = \sqrt{\langle Ax, x \rangle} = \sqrt{x^H Ax}, \quad \forall x \in \mathbb{C}^n,$$

را در نظر بگیرید (قضیه ۱۴.۱ را ببیند). فرض کنید $\|\cdot\|$ نرم طبیعی تولیدشده توسط $\|\cdot\|_A$ باشد. نشان می‌دهیم $\|M^{-1}N\| < 1$. در این صورت، با توجه به نتیجه ۱.۲، اثبات قضیه کامل می‌شود. با توجه به تعریف نرم طبیعی، برداری مثل v با فرض $\|v\|_A = 1$ وجود دارد به طوری که

$$\|M^{-1}N\|^\vee = \max_{\|x\|_A=1} \|M^{-1}Nx\|_A^\vee = \|M^{-1}Nv\|_A^\vee.$$

با فرض $w = M^{-1}Av$ و اینکه $N = M - A$ ، داریم

$$M^{-1}Nv = M^{-1}(M - A)v = v - M^{-1}Av = v - w.$$

لذا، با توجه به اینکه $\|v\|_A^\vee = 1$ ، داریم $\langle Av, v \rangle = \|v\|_A^\vee = 1$

$$\begin{aligned} \|M^{-1}Nv\|_A^\vee &= \|v - w\|_A^\vee \\ &= \langle A(v - w), v - w \rangle \\ &= \langle Av, v \rangle - \langle Aw, v \rangle - \langle Av, w \rangle + \langle Aw, w \rangle \\ &= 1 - \langle w, Av \rangle - \langle Av, w \rangle + \langle Aw, w \rangle \\ &= 1 - \langle w, MM^{-1}Av \rangle - \langle MM^{-1}Av, w \rangle + \langle Aw, w \rangle \\ &= 1 - \langle w, Mw \rangle - \langle Mw, w \rangle + \langle Aw, w \rangle \\ &= 1 - \langle M^H w, w \rangle - \langle (M - A)w, w \rangle \end{aligned}$$

$$\begin{aligned} &= 1 - \langle M^H w, w \rangle - \langle Nw, w \rangle \\ &= 1 - \langle (M^H + N)w, w \rangle. \end{aligned}$$

به‌وضوح $w \neq 0$. بنابراین، با توجه به اینکه $M^H + N$ معین مثبت هرمیتی است، داریم $\langle (M^H + N)w, w \rangle > 0$. در نتیجه، $\|M^{-1}Nv\|_A < 1$ و این اثبات قضیه را کامل می‌کند. $\square$

قضیه ۴.۲ (کاهان). اگر $\omega \notin (0, 2)$ ، آنگاه روش SOR واگرا است.

برهان. همان‌گونه که دیدیم ماتریس تکرار روش SOR به‌صورت

$$G_{\text{SOR}}(\omega) = (D - \omega E)^{-1}((1 - \omega)D + \omega F).$$

است. به‌سادگی می‌توان دید

$$\det((D - \omega E)^{-1}) = \prod_{i=1}^n \frac{1}{a_{ii}} = \frac{1}{\prod_{i=1}^n a_{ii}},$$

و

$$\det((1 - \omega)D + \omega F) = \prod_{i=1}^n (1 - \omega)a_{ii} = (1 - \omega)^n \prod_{i=1}^n a_{ii}.$$

اکنون، اگر λ_i ، $i = 1, 2, \dots, n$ ، مقادیر ویژه $G_{\text{SOR}}(\omega)$ باشند، آنگاه

$$\prod_{i=1}^n \lambda_i = \det(G_{\text{SOR}}(\omega)) = (1 - \omega)^n.$$

بنابراین،

$$|1 - \omega|^n = \prod_{i=1}^n |\lambda_i| \leq \rho(G_{\text{SOR}}(\omega))^n,$$

که معادل با

$$|1 - \omega| \leq \rho(G_{\text{SOR}}(\omega)).$$

است. اینک، اگر $|1 - \omega| \geq 1$ ، آنگاه روش SOR همگرا نخواهد شد و این رابطه با اینکه $\omega \geq 2$ یا

$\square$

$\omega \leq 0$ معادل است.

این قضیه نشان می‌دهد اگر روش SOR همگرا باشد، آنگاه $\omega \in (0, 2)$.

قضیه ۵.۲ (آستروفسکی و ریچ). اگر ماتریس A معین مثبت هرمیتی باشد، آنگاه به ازای هر $0 < \omega < 2$ روش SOR همگرا است.

برهان. همان‌گونه که دیدیم، در روش تکراری SOR داریم

$$M = \frac{1}{\omega}(D - \omega E), \quad N = \frac{1}{\omega}((1 - \omega)D + \omega F).$$

با توجه به اینکه A متقارن است، داریم $F = E^H$. بنابراین، ماتریس

$$M^H + N = \frac{1}{\omega}(D - \omega E^H + (1 - \omega)D + \omega E^H) = \frac{1}{\omega}(2 - \omega)D,$$

به ازای هر $0 < \omega < 2$ معین مثبت هرمیتی است. اکنون، با استفاده از قضیه ۳.۲ نتیجه لازم به دست می‌آید. $\square$

اثبات دیگری از این قضیه را می‌توان در [۹۳] یافت.

نتیجه ۲.۲ اگر A معین مثبت هرمیتی باشد، آنگاه روش گاوس-سایدل همگرا خواهد بود.

برهان. با توجه به قضیه ۵.۲ و اینکه روش SOR به ازای $\omega = 1$ ، روش گاوس-سایدل را به دست می‌دهد، نتیجه لازم به دست می‌آید. $\square$

قضیه ۶.۲ اگر ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت هرمیتی باشد و $\omega \in (0, 2)$ ، آنگاه روش تکراری SSOR به ازای هر حدس اولیه دلخواه همگرا است.

برهان. اگر A معین مثبت هرمیتی باشد، آنگاه درایه‌های قطری آن مثبت هستند و $F = E^H$. قرار می‌دهیم $\hat{D} = D^{\frac{1}{2}}$. در نتیجه، ماتریس تکرار روش به صورت

$$\begin{aligned} \mathcal{G}_\omega &= M^{-1}N \\ &= \left((D - \omega E)D^{-1}(D - \omega E^H) \right)^{-1} ((1 - \omega)D + \omega E)D^{-1}((1 - \omega)D + \omega E^H) \\ &= D^{-\frac{1}{2}} \left((I - \omega \hat{E})(I - \omega \hat{E}^H) \right)^{-1} ((1 - \omega)I + \omega \hat{E})((1 - \omega)I + \omega \hat{E}^H)D^{\frac{1}{2}} \\ &= D^{-\frac{1}{2}} \hat{\mathcal{G}}_\omega D^{\frac{1}{2}}, \end{aligned}$$

خواهد بود، که در آن $\hat{E} = D^{-\frac{1}{2}}ED^{-\frac{1}{2}}$. بنابراین، ماتریس‌های $\mathcal{G}_\omega$ و $\hat{\mathcal{G}}_\omega$ متشابه هستند و مقادیر ویژه یکسانی دارند. لذا، برای بررسی همگرایی روش تکراری SSOR روی مقادیر ویژه $\hat{\mathcal{G}}_\omega$ تمرکز می‌کنیم.

فرض کنید (λ, x) یک زوج ویژه ماتریس $\hat{G}_\omega$ باشد. در این صورت، داریم $\hat{G}_\omega x = \lambda x$ که با

$$((1 - \omega)I + \omega \hat{E})(1 - \omega)I + \omega \hat{E}^H x = \lambda(I - \omega \hat{E})(I - \omega \hat{E}^H)x,$$

معادل است. با ضرب طرفین این رابطه در بردار x^H ، خواهیم داشت

$$\lambda = \frac{\|(1 - \omega)x + \omega \hat{E}^H x\|_2^2}{\|x - \omega \hat{E}^H x\|_2^2}.$$

اگر صورت و مخرج این کسر را به ترتیب با t و s نشان دهیم، آنگاه

$$\begin{aligned} t &= (1 - \omega)^2 \langle x, x \rangle + \omega(1 - \omega) \langle (\hat{E} + \hat{E}^H)x, x \rangle + \omega^2 \langle \hat{E}^H x, \hat{E}^H x \rangle \\ &= \langle x - \omega \hat{E}^H x, x - \omega \hat{E}^H x \rangle - \omega(2 - \omega) \left(\langle x, x \rangle - \langle (\hat{E} + \hat{E}^H)x, x \rangle \right) \\ &= \|x - \omega \hat{E}^H x\|_2^2 - \omega(2 - \omega) \left(\langle x, x \rangle - \langle (I - \hat{A})x, x \rangle \right) \\ &= s - \omega(2 - \omega) \langle \hat{A}x, x \rangle, \end{aligned}$$

که در آن $\hat{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. روشن است که ماتریس $\hat{A}$ معین مثبت هرمیتی است و در نتیجه، $\langle \hat{A}x, x \rangle > 0$. لذا از اینکه $(0, 2) \in \omega$ ، داریم $\omega(2 - \omega) > 0$ و

$$0 \leq \lambda = \frac{t}{s} = \frac{s - \omega(2 - \omega) \langle \hat{A}x, x \rangle}{s} = 1 - \omega(2 - \omega) \frac{\langle \hat{A}x, x \rangle}{s} < 1,$$

□

که اثبات قضیه را کامل می‌کند.

نتیجه ۳.۲ اگر ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت هرمیتی باشد، آنگاه روش تکراری SGS به‌ازای هر حدس اولیه دلخواه به جواب دستگاه $Ax = b$ همگرا است.

نتیجه ۴.۲ فرض کنید ماتریس A معین مثبت هرمیتی باشد و $\omega \in (0, 2)$. در این صورت، مقادیر ویژه ماتریس $M^{-1}A$ در بازه $(0, 1)$ قرار دارند، که در آن

$$M = \frac{1}{\omega(2 - \omega)} (D - \omega E) D^{-1} (D - \omega F).$$

برهان. فرض کنید $N = M - A$. در این صورت، داریم

$$M^{-1}A = M^{-1}(M - N) = I - G_\omega.$$

با توجه به قضیه ۶.۲، مقادیر ویژه ماتریس $\mathcal{G}_\omega$ در بازه $[0, 1)$ قرار دارند، و لذا، مقادیر ویژه $M^{-1}A$ در بازه $[0, 1)$ قرار دارند. $\square$

۶-۲ محاسبه پارامتر بهینه در روش SOR

مسئله محاسبه پارامتر بهینه در روش SOR بسیار پیچیده است و تاکنون قضیه‌ای در این خصوص برای ماتریس‌های دلخواه ارائه نشده است. اما در حالت‌های خاص قضایایی وجود دارد که در این بخش بعضی از آنها را مرور می‌کنیم.

تعریف ۲.۲ گوییم ماتریس A دارای “خاصیت A” است، اگر یک ماتریس جایگشت مثل P وجود داشته باشد به طوری که

$$PAP^T = \begin{pmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{pmatrix},$$

که در آن D_1 و D_2 ماتریس‌های قطری هستند.

مثال ۷.۲ ماتریس

$$A = \begin{pmatrix} -4 & 1 & 0 & 0 & 1 & 0 \\ 1 & -4 & 1 & 0 & 0 & 0 \\ 0 & 1 & -4 & 1 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 0 \\ 1 & 0 & 0 & 0 & -4 & 1 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{pmatrix},$$

دارای خاصیت A است. زیرا اگر قرار دهیم

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

و

$$D_1 = D_2 = \begin{pmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -4 \end{pmatrix}, \quad A_{12} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$A_{21} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

آنگاه

$$PAP^T = \begin{pmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{pmatrix}.$$

قضیه ۷.۲ فرض کنید ماتریس $A = (a_{ij})$ دارای خاصیت A است و $a_{ii} \neq 0$ ، $i = 1, 2, \dots, n$. در این صورت، یک ماتریس جایگشت مثل P وجود دارد به طوری که

$$\bar{A} = PAP^T = D(I - L - U),$$

که در آن L ، U و D به ترتیب ماتریس‌های پائین‌مثلثی اکید، بالامثلثی اکید و قطری هستند و به علاوه، مقادیر ویژه

$$J(\alpha) = \alpha L + \alpha^{-1}U,$$

مستقل از α هستند که در آن $\alpha \in \mathbb{C}$ و $\alpha \neq 0$.

برهان. از اینکه ماتریس A دارای خاصیت A است یک ماتریس جایگشت مثل P وجود دارد به طوری که

$$\begin{aligned} PAP^T &= \begin{pmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{pmatrix} \\ &= \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ -A_{21} & 0 \end{pmatrix} - \begin{pmatrix} 0 & -A_{12} \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \left[I - \begin{pmatrix} 0 & 0 \\ -D_2^{-1}A_{21} & 0 \end{pmatrix} - \begin{pmatrix} 0 & -D_1^{-1}A_{12} \\ 0 & 0 \end{pmatrix} \right]. \end{aligned}$$

کافی است ماتریس‌های D ، L و U را به صورت زیر معرفی کنیم

$$D = \begin{pmatrix} D_1 & \circ \\ \circ & D_2 \end{pmatrix}, \quad L = \begin{pmatrix} \circ & \circ \\ -D_2^{-1}A_{21} & \circ \end{pmatrix}, \quad U = \begin{pmatrix} \circ & -D_1^{-1}A_{12} \\ \circ & \circ \end{pmatrix}.$$

برای تکمیل اثبات قضیه، داریم

$$\begin{aligned} J(\alpha) &= \alpha L + \alpha^{-1}U \\ &= \begin{pmatrix} \circ & -\alpha^{-1}D_1^{-1}A_{12} \\ -\alpha D_2^{-1}A_{21} & \circ \end{pmatrix} \\ &= S_\alpha J(1) S_\alpha^{-1}, \end{aligned}$$

که در آن

$$S_\alpha = \begin{pmatrix} I_1 & \circ \\ \circ & \alpha I_2 \end{pmatrix},$$

و I_1 و I_2 ماتریس‌های همانی هستند. این رابطه نشان می‌دهد که $J(\alpha)$ و $J(1)$ متشابه هستند و بنابراین، مقادیر ویژه این دو ماتریس یکسان هستند و این اثبات قضیه را کامل می‌کند. $\square$ خاصیت دیگری که در اینجا بررسی می‌شود، ماتریس «به‌طور سازگار مرتب‌شده» است.

تعریف ۳.۲ تجزیه $A = D - E - F$ را در نظر بگیرید که در آن D نامنفرد است. در این صورت،

$$A = D(I - L - U),$$

که در آن $U = D^{-1}F$ و $L = D^{-1}E$. گوییم ماتریس A به‌طور سازگار مرتب‌شده است، هرگاه به‌ازای هر $\alpha \neq 0$ مقادیر ویژه $J(\alpha) = \alpha L + \alpha^{-1}U$ مستقل از α باشند.

قضیه ۷.۲ نشان می‌دهد که اگر ماتریس A خاصیت A داشته باشد، آنگاه می‌توان با استفاده از جابه‌جایی هم‌زمان سطرها و ستون‌های A ، آن را به ماتریسی تبدیل کرد که به‌طور سازگار مرتب‌شده باشد. ماتریس‌هایی که به‌طور سازگار مرتب‌شده هستند، لزوماً به‌صورت

$$A = \begin{pmatrix} D_1 & A_{12} \\ A_{21} & D_2 \end{pmatrix},$$

که در آن D_1 و D_2 قطری هستند، نمی‌باشند. برای مثال، ماتریس‌های سه‌قطری بلوکی به‌شکل

$$A = \begin{pmatrix} D_1 & A_{12} & & \\ A_{21} & D_2 & \ddots & \\ & \ddots & \ddots & A_{m-1,m} \\ & & A_{m,m-1} & D_m \end{pmatrix},$$

را در نظر بگیرید که در آن D_i ها ماتریس‌های قطری هستند. فرض کنید D_i ها نامنفرد باشند. داریم

$$J(\alpha) = - \begin{pmatrix} \circ & \alpha^{-1} D_1^{-1} A_{12} & & \\ \alpha D_1^{-1} A_{21} & \circ & \alpha^{-1} D_2^{-1} A_{23} & \\ & \ddots & \ddots & \ddots \\ \alpha D_{m-1}^{-1} A_{m-1,m-2} & & \circ & \alpha^{-1} D_{m-1}^{-1} A_{m-1,m} \\ & & \alpha D_m^{-1} A_{m,m-1} & \circ \end{pmatrix}.$$

در این صورت، $J(\alpha) = S_\alpha J(1) S_\alpha^{-1}$ که در آن

$$S_\alpha = \begin{pmatrix} I_1 & & & \\ & \alpha I_2 & & \\ & & \ddots & \\ & & & \alpha^{m-1} I_{m-1} \end{pmatrix},$$

و I_i ها ماتریس‌های همانی هستند. بنابراین، ماتریس A به‌طور سازگار مرتب‌شده است.

هر ماتریس به‌طور سازگار مرتب‌شده لزوماً دارای خاصیت A نیست. برای مثال، کافی است قرار دهید

$$A = \begin{pmatrix} 1 & \circ & \circ \\ 1 & 1 & \circ \\ 1 & 1 & 1 \end{pmatrix}.$$

به‌سادگی می‌توان دید که برای این ماتریس، داریم

$$J(\alpha) = \alpha L + \alpha^{-1} U = \begin{pmatrix} \circ & \circ & \circ \\ -\alpha & \circ & \circ \\ -\alpha & -\alpha & \circ \end{pmatrix}.$$

مقادیر ویژه این ماتریس همگی صفر و مستقل از α هستند. بنابراین، ماتریس A به‌طور سازگار مرتب‌شده است. می‌توان دید که این ماتریس دارای خاصیت A نیست.

قضیه ۸.۲ (یانگ و وارگا). فرض کنید A به‌طور سازگار مرتب‌شده باشد و $\omega \neq 0$.
(الف) اگر μ یک مقدار ویژه ماتریس تکرار روش ژاکوبی، G_J ، باشد، آنگاه $-\mu$ نیز مقدار ویژه G_J است.
(ب) اگر μ یک مقدار ویژه G_J باشد و

$$(\lambda + \omega - 1)^2 = \lambda \omega^2 \mu^2, \quad (16-2)$$

آنگاه λ یک مقدار ویژه $G_{SOR}(\omega)$ است.

(ج) اگر $\omega \neq 0$ یک مقدار ویژه $G_{SOR}(\omega)$ باشد و $(16-2)$ برقرار باشد، آنگاه μ یک مقدار ویژه G_J است.

برهان. (الف) همان‌گونه که دیدیم، اگر $A = D - E - F$ ، آنگاه $G_J = L + U$ ، که در آن $L = D^{-1}E$ و $U = D^{-1}F$. قرار می‌دهیم $J(\alpha) = \alpha L + \alpha^{-1}U$ ، $\alpha \in \mathbb{C}$ ، $\alpha \neq 0$. با توجه به اینکه A به‌طور سازگار مرتب‌شده است، مقادیر ویژه

$$J(-1) = -L - U = -(L + U) = -J(1) = -G_J,$$

و $J(1) = L + U = G_J$ یکسان هستند. بنابراین، اگر μ یک مقدار ویژه G_J باشد، آنگاه μ مقدار ویژه $-G_J$ نیز هست. پس

$$\exists x \neq 0 : -G_J x = \mu x.$$

بنابراین، داریم $G_J x = -\mu x$. یعنی، $-\mu$ نیز مقدار ویژه G_J است.
(ب) ابتدا توجه کنید که

$$\begin{aligned} G_{SOR}(\omega) &= (D - \omega E)^{-1}((1 - \omega)D + \omega F) \\ &= (I - \omega D^{-1}E)^{-1}D^{-1}((1 - \omega)D + \omega F) \\ &= (I - \omega L)^{-1}((1 - \omega)I + \omega D^{-1}F) \\ &= (I - \omega L)^{-1}((1 - \omega)I + \omega U). \end{aligned}$$

فرض کنید μ یک مقدار ویژه G_J و رابطه $(16-2)$ برقرار باشد. داریم

$$\lambda + \omega - 1 = \pm \sqrt{\lambda} \omega \mu.$$

با توجه به قسمت (الف) و بدون از دست دادن کلیت، فرض می‌کنیم

$$\lambda + \omega - 1 = \sqrt{\lambda} \omega \mu.$$

با استفاده از این رابطه و اینکه $\det(I - \omega L)^{-1} = 1$ داریم

$$\begin{aligned} \det(\lambda I - G_{\text{SOR}}(\omega)) &= \det(\lambda I - (I - \omega L)^{-1}((1 - \omega)I + \omega U)) \\ &= \det(I - \omega L)^{-1} \det(\lambda(I - \omega L) - (1 - \omega)I - \omega U) \\ &= \det((\lambda + \omega - 1)I - \omega(\lambda L + U)) \\ &= \det(\omega \mu \sqrt{\lambda} I - \omega(\lambda L + U)). \end{aligned}$$

اکنون، اگر $\lambda = 0$ ، آنگاه

$$\det(0 \times I - G_{\text{SOR}}(\omega)) = \det(\omega \mu \sqrt{0} I - \omega(0 \times L + U)) = \det(-\omega U) = 0.$$

بنابراین، $\lambda = 0$ یک مقدار ویژه $G_{\text{SOR}}(\omega)$ است. در صورتی که $\lambda \neq 0$ ، داریم

$$\begin{aligned} \det(\lambda I - G_{\text{SOR}}(\omega)) &= \det(\omega \mu \sqrt{\lambda} I - \omega(\lambda L + U)) \\ &= \det(\omega \mu \sqrt{\lambda} I - \omega \sqrt{\lambda} (\sqrt{\lambda} L + \frac{1}{\sqrt{\lambda}} U)) \\ &= (\omega \sqrt{\lambda})^n \det(\mu I - (\sqrt{\lambda} L + \frac{1}{\sqrt{\lambda}} U)) \\ &= (\omega \sqrt{\lambda})^n \det(\mu I - J(\sqrt{\lambda})). \end{aligned}$$

با توجه به اینکه مقادیر ویژه $J(\sqrt{\lambda})$ با $J(1) = G_J$ یکسان هستند و μ یک مقدار ویژه G_J است

داریم $\det(\mu I - J(\sqrt{\lambda})) = 0$. بنابراین،

$$\det(\lambda I - G_{\text{SOR}}(\omega)) = 0,$$

و این رابطه نشان می‌دهد که λ یک مقدار ویژه $G_{\text{SOR}}(\omega)$ است.

(ج) دیدیم که

$$\det(\lambda I - G_{\text{SOR}}(\omega)) = (\omega\sqrt{\lambda})^n \det(\mu I - J(\sqrt{\lambda})).$$

فرض کنید λ مقدار ویژه $G_{\text{SOR}}(\omega)$ باشد. در این صورت، سمت چپ این رابطه برابر با صفر است و در نتیجه،

$$\det(\mu I - J(\sqrt{\lambda})) = 0.$$

یعنی μ یک مقدار ویژه $J(\sqrt{\lambda})$ است. با توجه به اینکه A به طور سازگار مرتب شده است، نتیجه می‌گیریم μ یک مقدار ویژه $G_J = J(1)$ نیز می‌باشد. $\square$

نتیجه ۵.۲ فرض کنید ماتریس A به طور سازگار مرتب شده باشد. در این صورت

$$\rho(G_{\text{GS}}) = \rho(G_J)^2,$$

که در آن G_{GS} و G_J به ترتیب ماتریس‌های تکرار روش‌های ژاکوبی و گاوس-سایدل هستند. برهان. به ازای $\omega = 1$ روش تکراری SOR به روش گاوس-سایدل تبدیل می‌شود. بنابراین، با استفاده از رابطه (۱۶-۲) به ازای $\omega = 1$ ، داریم $\lambda^2 = \lambda\mu^2$ ، که با استفاده از آن نتیجه لازم به دست می‌آید. $\square$

قضیه ۹.۲ (یانگ و وارگا). فرض کنید ماتریس A به طور سازگار مرتب شده، مقادیر ویژه ماتریس تکرار ژاکوبی حقیقی باشند و $\omega \neq 0$. اگر $\rho(G_J) < 1$ و

$$\rho(G_{\text{SOR}}(\omega_{\text{opt}})) = \min_{0 < \omega < 2} \rho(G_{\text{SOR}}(\omega)),$$

آنگاه

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - \rho(G_J)^2}},$$

و

$$\rho(G_{\text{SOR}}(\omega_{\text{opt}})) = \omega_{\text{opt}} - 1 = \left(\frac{\rho(G_J)}{1 + \sqrt{1 - \rho(G_J)^2}} \right)^2.$$

برهان. فرض کنید μ یک مقدار ویژه G_J باشد. در این صورت، $1 < \rho(G_J) \leq \mu$. با محاسبه λ از رابطه (۱۶-۲)، داریم

$$\lambda_1, \lambda_2 = \frac{1}{4}\mu^2\omega^2 - (\omega - 1) \pm \mu\omega\sqrt{\frac{1}{4}\mu^2\omega^2 - (\omega - 1)}. \quad (17-2)$$

همچنین، با استفاده از رابطه (۲-۱۶)، داریم

$$\frac{\lambda + \omega - 1}{\omega} = \pm \sqrt{\lambda \mu}.$$

قرار می‌دهیم

$$f_{\omega}(\lambda) = \frac{\lambda + \omega - 1}{\omega}, \quad g(\lambda) = \pm \sqrt{\lambda \mu}.$$

به‌ازای یک مقدار دلخواه از ω ، نمودار این دو تابع در شکل ۲-۶ رسم شده‌اند. نقاط برخورد توابع f_{ω} و g را A و B می‌نامیم. این نقاط متناظر به مقادیر λ_1 و λ_2 می‌باشند به‌طوری‌که $\lambda_1 \geq \lambda_2$. شیب خط $f_{\omega}(\lambda)$ برابر با $1/\omega$ است و از نقطه $(1, 1)$ نیز می‌گذرد. با کاهش ω ، شیب این خط افزایش می‌یابد و در نتیجه، λ_1 و $\rho(G_{\text{SOR}}(\omega))$ افزایش می‌یابند. بنابراین، حتی‌الامکان بایستی ω را در فاصله صفر تا ۲ افزایش دهیم. با افزایش ω ، شیب خط کاهش می‌یابد تا جایی که f_{ω} و g بر هم مماس می‌شوند. در این حالت، خواهیم داشت $\lambda_1 = \lambda_2$ و

$$\frac{1}{4}\mu^2\omega^2 - (\omega - 1) = 0.$$

با حل این معادله، داریم

$$\omega_{\pm}(\mu) = \frac{2}{1 \pm \sqrt{1 - \mu^2}}.$$

با توجه به اینکه $1 < \rho(G_J) \leq \mu$ ، می‌بینیم که

$$\omega_{-}(\mu) = \frac{2}{1 - \sqrt{1 - \mu^2}} \notin (0, 2).$$

بنابراین، $\omega_{+}(\mu)$ را در نظر می‌گیریم که

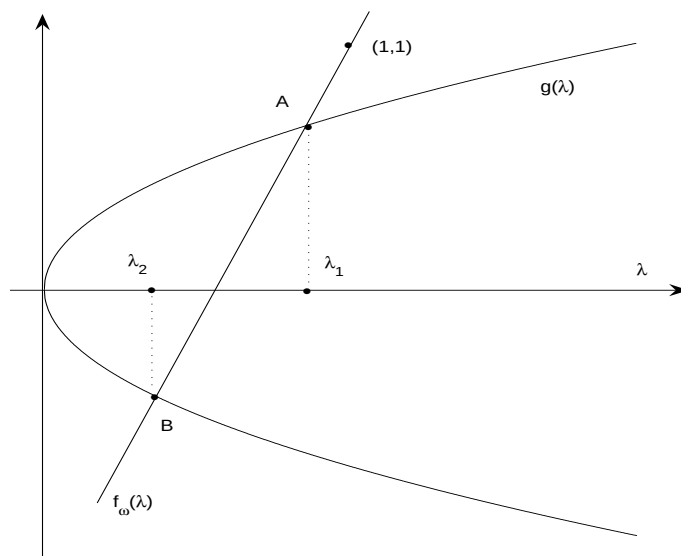
$$\omega_{+}(\mu) = \frac{2}{1 + \sqrt{1 - \mu^2}} \in [1, 2).$$

در این حالت، داریم

$$\frac{1}{4}\mu^2\omega_{+}(\mu)^2 - (\omega_{+}(\mu) - 1) = 0,$$

یا

$$\mu^2\omega_{+}(\mu)^2 = 4(\omega_{+}(\mu) - 1).$$



شکل ۶-۲ نمایش مقادیر ویژه ماتریس تکرار روش SOR.

درنتیجه، با توجه به این رابطه و رابطه (۱۷-۲)، خواهیم داشت

$$\begin{aligned}\lambda_1 &= \lambda_2 = \frac{1}{\omega} \mu^2 \omega_+(\mu)^2 - (\omega_+(\mu) - 1) \\ &= 2(\omega_+(\mu) - 1) - (\omega_+(\mu) - 1) \\ &= \omega_+(\mu) - 1.\end{aligned}$$

قرار می‌دهیم

$$\lambda_+(\mu) = \omega_+(\mu) - 1. \quad (18-2)$$

بنابراین،

$$|\lambda_1| = |\lambda_2| = \omega_+(\mu) - 1.$$

اکنون، فرض می‌کنیم که $\omega > \omega_+(\mu) \geq 1$. در این صورت، خواهیم داشت

$$\frac{1}{\omega} < \frac{1}{\omega_+(\mu)},$$

و f_ω و g یکدیگر را قطع نمی‌کنند. در نتیجه، λ_1 و λ_2 ریشه‌های مختلط هستند، یعنی

$$\lambda_1, \lambda_2 = \frac{1}{4}\mu^2\omega^2 - (\omega - 1) \pm i\mu\omega\sqrt{(\omega - 1) - \frac{1}{4}\mu^2\omega^2},$$

که در آن i واحد موهومی است. در این حالت، داریم $|\lambda_1| = |\lambda_2| = \omega - 1$. بنابراین، با توجه به (۱۸-۲) و اینکه $\omega > \omega_+(\mu) - 1$ داریم $\omega - 1 > \omega_+(\mu) - 1$. این رابطه نشان می‌دهد که اگر μ یک مقدار ویژه G_J باشد، آنگاه بهترین انتخاب برای پارامتر w مقدار $\omega_+(\mu)$ است. فرض کنید $\sigma(G_J)$ مجموعه مقادیر ویژه G_J باشد. در این صورت،

$$\begin{aligned}\rho(G_{\text{SOR}}(\omega_{\text{opt}})) &= \max_{\mu \in \sigma(G_J)} |\lambda_+(\mu)| \\ &= \max_{\mu \in \sigma(G_J)} \omega_+(\mu) - 1 \\ &= \omega_+(\rho(G_J)) - 1.\end{aligned}$$

□ در نتیجه، $\omega_{\text{opt}} = \omega_+(\rho(G_J))$ که اثبات قضیه را کامل می‌کند.

یک نتیجه این قضیه این است که اگر ماتریس A به‌طور سازگار مرتب‌شده باشد، آنگاه در حالت کلی داریم

$$\rho(G_{\text{SOR}}(\omega)) = \begin{cases} 1 - \omega + \frac{1}{4}\omega^2\mu^2 + \omega\mu\sqrt{1 - \omega + \frac{1}{4}\omega^2\mu^2}, & 0 < \omega < \omega_{\text{opt}}, \\ \omega - 1, & \omega_{\text{opt}} \leq \omega < 2, \end{cases}$$

که در آن $\mu = \rho(G_J)$.

مثال ۸.۲ دستگاه معادلات خطی

$$\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

را در نظر بگیرید. فرض می‌کنیم $n = 20$. به‌سادگی می‌توان دید که جواب واقعی این دستگاه، برابر با بردار $x = (1, 1, \dots, 1)^T$ است. به‌وضوح ماتریس ضرایب دستگاه به‌طور سازگار مرتب‌شده است. با محاسبه

جدول ۱-۲ نتایج عددی برای مثال ۸.۲

۱/۸	۱/۸	$\omega_{\text{opt}} = ۱/۷۴۰۶$	۱/۷	۱/۶	ω
۲۳۰	۱۰۵	۸۶	۱۳۱	۲۱۰	iters

ماتریس تکرار روش ژاکوبی، داریم

$$G_J = \begin{pmatrix} 0 & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & \frac{1}{2} & 0 \end{pmatrix},$$

است. همان‌گونه که در پایان فصل ۴ دیدیم مقادیر ویژه این ماتریس عبارتند از

$$\lambda_j = 0 + 2 \times \frac{1}{2} \sqrt{\frac{1}{2} \cos \frac{j\pi}{20+1}} = \cos \frac{j\pi}{21}, \quad j = 1, 2, \dots, 20.$$

بنابراین،

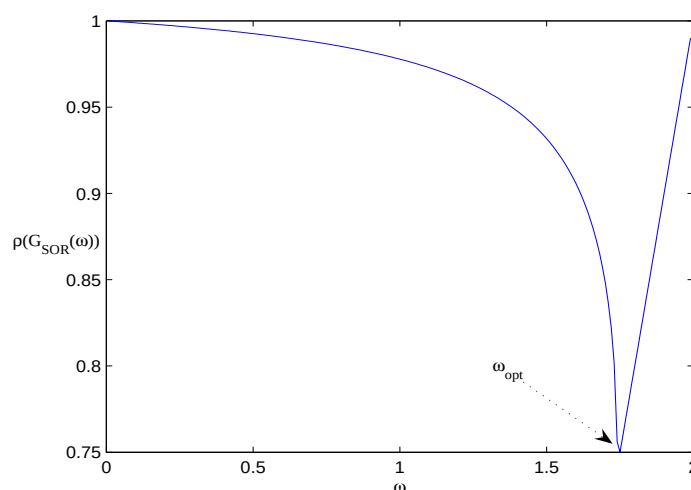
$$\rho(G_J) = \cos \frac{\pi}{21} \approx 0.9888.$$

درنتیجه، با توجه قضیه ۹.۲، داریم

$$\omega_{\text{opt}} = \frac{2}{1 + \sqrt{1 - 0.9888^2}} \approx 1/7406.$$

برای نشان دادن کارایی مقدار ω_{opt} به‌دست‌آمده، دستگاه معادلات خطی بالا را با استفاده از روش SOR به‌ازای ω_{opt} و چند مقدار دیگر حل کرده و نتایج عددی را با یکدیگر مقایسه می‌کنیم. برای این کار، حدس اولیه را $x^{(0)} = 0$ ، حداکثر تکرارها را $\text{itmax} = 1000$ و شرط توقف را $\|r_k\|_2 < 10^{-10} \|r_0\|_2$ می‌گیریم که در آن r_k مانده تکرار k ام است. نتایج عددی در جدول ۱-۲ داده شده‌اند. در این جدول iters نشان‌دهنده تعداد تکرارها است.

همان‌گونه که این جدول نشان می‌دهد، روش SOR به‌ازای ω_{opt} سریع‌تر از سایر مقادیر همگرا شده است. در شکل ۷-۲ تابع $\rho(G_{\text{SOR}}(\omega))$ را برای $0 < \omega < 2$ رسم کرده‌ایم. این نمودار رفتار $\rho(G_{\text{SOR}}(\omega))$ را نشان می‌دهد. همچنین می‌بینیم که مقدار بهینه ω برابر $1/7406$ است.



شکل ۷-۲ نمودار مثال ۸.۲.

۷-۲ همگرایی AOR برای ماتریس‌های غالب قطری

در این بخش به بررسی خواص همگرایی روش تکراری AOR برای ماتریس‌های غالب قطری می‌پردازیم. نخست تعریف زیر را ارائه می‌کنیم.

تعریف ۴.۲ ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ را تحویل‌پذیر می‌گویند اگر یک مجموعهٔ ناتهی مثل S از $N = \{1, 2, \dots, n\}$ وجود داشته باشد به‌طوری‌که به‌ازای هر $i \in S$ و $j \in N - S$ ، $a_{ij} = 0$. درغیراین‌صورت، ماتریس A را تحویل‌ناپذیر می‌گویند.

مثال ۹.۲ ماتریس

$$A = \begin{pmatrix} a_{11} & \dots & a_{1r} & \dots & a_{1n} \\ \vdots & & & & \vdots \\ 0 & \dots & 0 & a_{rr} & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nr} & \dots & a_{nn} \end{pmatrix},$$

تحویل‌پذیر است، زیرا با فرض $S = \{r\}$ شرایط تعریف ۴.۲ برقرار خواهد بود.

قضیه ۱۰.۲ ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ تحویل‌پذیر است اگر و تنها اگر یک ماتریس مثل P وجود داشته باشد به طوری که

$$PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ \circ & A_{22} \end{pmatrix}, \quad (19-2)$$

که در آن A_{11} و A_{22} ماتریس‌های مربعی هستند. برهان. فرض کنید ماتریس A تحویل‌پذیر باشد و

$$P = (e_{i_1}, \dots, e_{i_r}, e_{i_{r+1}}, \dots, e_{i_n}),$$

که در آن e_{i_k} ستون i_k ام ماتریس همانی است، $i_k \neq i_l$ و $k \neq l$ و $S = \{e_{i_{r+1}}, \dots, e_{i_n}\}$ در این صورت، داریم

$$(PAP^T)_{st} = a_{kl} = \circ, \quad s = i_k \in S, t = i_l \in N-S, s = r+1, \dots, n, t = 1, \dots, r.$$

این رابطه نشان می‌دهد که ماتریس PAP^T به صورت (۱۹-۲) است. برای اثبات عکس قضیه، می‌بینیم که اگر ماتریس $B = PAP^T$ به صورت (۱۹-۲) باشد، آنگاه ماتریس $A = P^T B P$ تحویل‌پذیر است. $\square$

مثال ۱۰.۲ ماتریس مثال ۹.۲ را در نظر بگیرید. با فرض

$$P = (e_1, \dots, e_{r-1}, e_n, e_{r+1}, \dots, e_{n-1}, e_r)^T,$$

می‌توان دید که ماتریس PAP^T به صورت (۱۹-۲) است.

تعریف ۵.۲ گراف جهت‌دار وابسته به ماتریس $A \in \mathbb{C}^{n \times n}$ شامل n رأس $P_1, P_2, \dots, P_n$ است به طوری که یک یال جهت‌دار از P_i به P_j مثل $(\overrightarrow{P_i P_j})$ وجود دارد اگر و تنها اگر $a_{ij} \neq \circ$.

تعریف ۶.۲ یک گراف جهت‌دار را قویاً همبند می‌گویند هرگاه برای هر زوج از رأس‌های متمایز P_i و P_j یک مسیر جهت‌دار مانند $\overrightarrow{P_{i_0} P_{i_1}}, \overrightarrow{P_{i_1} P_{i_2}}, \dots, \overrightarrow{P_{i_{r-1}} P_{i_r}}$ در گراف وجود داشته باشد به طوری که $i_0 = i$ و $i_r = j$.

قضیه ۱۱.۲ یک ماتریس تحویل‌ناپذیر است اگر و تنها اگر گراف جهت‌دار آن قویاً همبند باشد.

برهان. فرض کنید A تحویل‌ناپذیر و i_0 یک اندیس دلخواه باشد. در این صورت، دست‌کم یک اندیس j وجود دارد به‌طوری‌که $a_{i_0,j} \neq 0$ ، زیرا در غیر این صورت، A تحویل‌پذیر است (مثال ۹.۲ را ببینید). فرض کنید T مجموعهٔ همهٔ اندیس‌هایی مانند j باشد به‌طوری‌که یک مسیر جهت‌دار از P_{i_0} به P_j وجود دارد و $S = N - T$ ، که در آن $N = \{1, 2, \dots, n\}$. لذا، S و T مجزا هستند و T ناتهی است. به برهان خلف، فرض کنید S ناتهی باشد، $k \in S$ و $j \in T$. در این صورت، خواهیم داشت $a_{jk} = 0$ ، زیرا اگر $a_{jk} \neq 0$ ، آنگاه یک مسیر جهت‌دار از P_{i_0} به P_k از طریق P_j ، مثل

$$\overrightarrow{P_{i_0}P_{i_1}}, \overrightarrow{P_{i_1}P_{i_2}}, \dots, \overrightarrow{P_{i_{r-1}}P_j}, \overrightarrow{P_jP_k},$$

وجود دارد، یعنی $k \in T$. لذا، برای هر $j \in T$ و $k \in S$ ، $a_{jk} = 0$ و از اینکه $S \cap T = \emptyset$ ، بنا به تعریف ۴.۲ ماتریس A تحویل‌پذیر است. این تناقض ایجاب می‌کند که S باید تهی باشد. بنابراین، گراف ماتریس باید قویاً همبند باشد.

برعکس، فرض کنید که گراف قویاً همبند باشد. اگر A تحویل‌پذیر باشد، آنگاه مجموعه‌های ناتهی و مجزای S و T وجود دارند به‌طوری‌که $a_{jk} = 0$ ، به‌ازای هر $k \in T$ و به‌ازای $j \in S$. از اینکه گراف قویاً همبند است، مسیر جهت‌داری مانند

$$\overrightarrow{P_{i_0}P_{i_1}}, \overrightarrow{P_{i_1}P_{i_2}}, \dots, \overrightarrow{P_{i_{r-1}}P_{i_r}},$$

وجود دارد که در آن $i_r = k \in T$ ، $i_0 = j \in S$. در دنبالهٔ $i_0, \dots, i_{r-1}$ یک اندیس نهایی i_s وجود دارد به‌طوری‌که $i_{s-1} \in S$ و $i_s \in T$ (شاید $i_s = k$). اما $a_{i_{s-1}, i_s} \neq 0$ ، که این متناقض با تعریف S و T است. بنابراین A تحویل‌ناپذیر است. $\square$

مثال ۱۱.۲ ماتریس سه‌قطری

$$A = \begin{pmatrix} & 2 & 1 & & \\ 1 & & 2 & 1 & \\ & \ddots & & \ddots & \ddots \\ & & 1 & 2 & 1 \\ & & & 1 & 2 \end{pmatrix} \in \mathbb{R}^{n \times n},$$

دارای گراف جهت‌دار نمایش داده‌شده در شکل ۸-۲ است که قویاً همبند است. بنابراین، ماتریس A تحویل‌ناپذیر است.



شکل ۸-۲ گراف جهت‌دار متناظر با ماتریس مثال ۱۱.۲.

تعریف ۷.۲ ماتریس $A = (a_{ij}) \in \mathbb{C}^{n \times n}$ را

(الف) غالب قطری اکید می‌گویند، هرگاه

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

(ب) غالب قطری ضعیف می‌گویند، هرگاه

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

(ج) غالب قطری تحویل‌ناپذیر می‌گویند، هرگاه تحویل‌ناپذیر بوده و

$$|a_{ii}| \geq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n,$$

و دست‌کم برای یک j ، نامساوی به‌صورت اکید باشد.

قضیه ۱۲.۲ فرض کنیم A یک ماتریس تحویل‌ناپذیر باشد. اگر یک مقدار ویژه λ از A روی مرز اجتماع

n دیسک گرشگورین قرار داشته باشد، آنگاه λ روی مرز همه دیسک‌های گرشگورین قرار دارد.

برهان. فرض کنید x یک بردار ویژه ماتریس A متناظر با مقدار ویژه λ باشد. بدون از دست دادن کلیت

مسئله، می‌توان فرض کرد $\|x\|_\infty = |x_m| = 1$. با استفاده از معادله m ام از دستگاه $Ax = \lambda x$ داریم

$$|\lambda - a_{mm}| \leq \sum_{\substack{j=1 \\ j \neq m}}^n |a_{mj}| |x_j| \leq \sum_{\substack{j=1 \\ j \neq m}}^n |a_{mj}| = \rho_m.$$

با توجه به اینکه λ روی مرز n دیسک گرشگورین A قرار دارد، نمی‌تواند درون دیسک m ام باشد.

بنابراین، داریم

$$|\lambda - a_{mm}| = \sum_{\substack{j=1 \\ j \neq m}}^n |a_{mj}| |x_j| = \sum_{\substack{j=1 \\ j \neq m}}^n |a_{mj}| = \rho_m. \quad (20-2)$$

حال فرض کنید $1 \leq j \leq n$ عددی دلخواه باشد. از اینکه A تحویل‌ناپذیر است، بنا به قضیه ۱۱.۲ گراف جهت‌دار آن قویاً همبند بوده و لذا، یک مسیر از گره m به گره j وجود دارد. فرض کنید این مسیر به صورت

$$\overrightarrow{P_m P_{m_1}}, \overrightarrow{P_{m_1} P_{m_2}}, \dots, \overrightarrow{P_{m_{k-1}} P_{m_k}}, \quad m_k = j,$$

باشد، بنابراین $a_{m, m_1} \neq 0$. با توجه به رابطه (۲۰-۲) برای هر x_j ای که ناصفر باشد داریم $|x_j| = 1$ ، پس $|x_{m_1}| = 1$. با تکرار این روند برای m_1 به جای m خواهیم داشت

$$|\lambda - a_{m_1 m_1}| = \sum_{\substack{j=1 \\ j \neq m_1}}^n |a_{m_1 j}| |x_j| = \sum_{\substack{j=1 \\ j \neq m_1}}^n |a_{m_1 j}| = \rho_{m_1}.$$

و در هر حال، داریم $|\lambda - a_{m_i m_i}| = \rho_{m_i}$ ، که برای هر $i = 1, 2, \dots, k$ برقرار است و نشان می‌دهد که λ متعلق به مرز j امین دیسک، با j دلخواه است. $\square$

نتیجه ۶.۲ اگر ماتریس A غالب قطری اکید یا غالب قطری تحویل‌ناپذیر باشد، آنگاه نامنفرد است. برهان. اگر ماتریس A غالب قطری اکید باشد، آنگاه اجتماع دیسک‌های گرشگورین A ، شامل مبدأ نیست. بنابراین، $\lambda = 0$ نمی‌تواند مقدار ویژه آن باشد، لذا، A نامنفرد است. حال فرض کنید A غالب قطری تحویل‌ناپذیر باشد. اگر A منفرد باشد، آنگاه مقدار ویژه صفر روی مرز اجتماع دیسک‌های گرشگورین قرار دارد و بنا به قضیه ۱۲.۲ این مقدار ویژه باید روی مرز همه دیسک‌ها باشد، یعنی

$$|a_{ii}| = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n,$$

که متناقض با فرض غالب قطری تحویل‌ناپذیر بودن است. $\square$

قضیه ۱۳.۲ فرض کنید $0 < w \leq 1$ و $0 \leq \gamma \leq 1$.

(الف) اگر ماتریس A غالب قطری اکید باشد، آنگاه روش AOR همگراست.

(ب) اگر A یک ماتریس غالب قطری تحویل‌ناپذیر باشد، آنگاه روش AOR همگراست.

برهان. فرض کنید λ یک مقدار ویژه $\mathcal{L}_{\gamma,\omega}$ باشد و $|\lambda| \geq 1$. برای این مقدار ویژه، داریم

$$\det(\mathcal{L}_{\gamma,\omega} - \lambda I) = 0, \quad (21-2)$$

و با جایگذاری ماتریس تکرار روش AOR در این رابطه، خواهیم داشت

$$\begin{aligned} \det(\mathcal{L}_{\gamma,\omega} - \lambda I) &= \det \left((I - \gamma L)^{-1} [(1 - \omega)I + (\omega - \gamma)L + \omega U] - \lambda I \right) \\ &= \det(I - \gamma L)^{-1} \det((1 - \omega - \lambda)I + (\omega - \gamma + \lambda\gamma)L + \omega U) \\ &= (1 - \omega - \lambda)^n \det \left(I - \frac{\gamma(\lambda - 1) + \omega}{\lambda - 1 + \omega} L - \frac{\omega}{\lambda - 1 + \omega} U \right). \end{aligned}$$

لازم به ذکر است که $1 - \omega - \lambda \neq 0$. زیرا، درغیراین صورت، داریم

$$|\lambda| = |1 - \omega| < 1,$$

که یک تناقض است. بنابراین، $\det(Q) = 0$ ، که در آن

$$Q = I - \frac{\gamma(\lambda - 1) + \omega}{\lambda - 1 + \omega} L - \frac{\omega}{\lambda - 1 + \omega} U. \quad (22-2)$$

نشان می‌دهیم که ضرایب U و L در رابطه (22-2) از حیث قدرمطلق کمتر یا مساوی یک و ناصفر هستند. روشن است که ضریب U ناصفر است. اگر ضریب L صفر باشد، آنگاه Q یک ماتریس بالامتثلی با درایه‌های قطری ۱ خواهد بود و درنتیجه، $\det(Q) = 1$ ، که یک تناقض است. بنابراین، کافی است نشان دهیم

$$|\lambda - 1 + \omega| \geq |\gamma(\lambda - 1) + \omega|, \quad |\lambda - 1 + \omega| \geq |\omega|. \quad (23-2)$$

با توجه به اینکه $|\lambda| \geq 1$ ، می‌توان فرض کرد $\lambda^{-1} = qe^{i\theta}$ ، که در آن θ و q اعداد حقیقی هستند به‌طوری‌که $0 < q \leq 1$. دراین صورت، نامساوی اول در (23-2) معادل با

$$|\omega - 1 + q^{-1} \cos \theta - iq^{-1} \sin \theta| \geq |\omega - \gamma + \gamma q^{-1} \cos \theta - i\gamma q^{-1} \sin \theta|,$$

است که این خود با

$$(1 - \gamma^2) + (1 - \gamma^2)q^2 - (1 - \gamma^2)2q \cos \theta + (1 - \gamma)2q\omega \cos \theta - (1 - \gamma)2q^2\omega \geq 0,$$

معادل است. با تقسیم رابطه اخیر بر $(1 - \gamma)$ داریم (توجه کنید که به‌ازای $\gamma = 1$ همه چیز روشن است)

$$(1 + \gamma) + (1 + \gamma)q^2 - [(1 + \gamma) - \omega]2q \cos \theta - 2q^2\omega \geq 0. \quad (24-2)$$

از اینکه عبارت درون کروشه نامنفی است، رابطه (24-2) برای هر مقدار حقیقی θ برقرار است اگر و تنها اگر برای $\cos \theta = 1$ برقرار باشد، یعنی

$$(1 - q)[(1 + \gamma)(1 - q) + 2q\omega] \geq 0,$$

که همواره برقرار است. به‌طور مشابه، برای نامساوی دوم در (23-2)، داریم

$$(1 - q)[(1 - q) + 2q\omega] \geq 0,$$

که همواره درست می‌باشد. حال دو حالت (الف) و (ب) را در نظر می‌گیریم. در حالت (الف) اگر ماتریس A غالب قطری اکید باشد، آنگاه با توجه به اینکه ضرائب U و L در (22-2) از لحاظ قدرمطلق کمتر از یک هستند، ماتریس Q نیز غالب قطری اکید است (توجه کنید که ماتریس $D^{-1}A = I - L - U$ غالب قطری اکید است). از این رو، ماتریس Q نیز غالب قطری اکید خواهد بود و طبق نتیجه ۶.۲ نامنفرد است که یک تناقض است. در حالت (ب) با توجه به اینکه ضرائب U و L در (22-2) از لحاظ قدرمطلق کمتر یا مساوی یک و ناصفر می‌باشند، ماتریس Q نیز غالب قطری تحویل‌ناپذیر است و طبق نتیجه ۶.۲، نامنفرد است که متناقض با (21-2) است. لذا، در هر دو حالت (الف) و (ب)، داریم $|\lambda| < 1$ و بنابراین، $\rho(\mathcal{L}_{\gamma, \omega}) < 1$. $\square$

نتیجه ۷.۲ فرض کنید ماتریس A غالب قطری اکید یا غالب قطری تحویل‌ناپذیر باشد. در این صورت،
(الف) روش‌های تکراری ژاکوبی و گاوس-سایدل همگرا هستند.
(ب) اگر $1 > \omega \geq 0$ ، آنگاه روش SOR همگراست.

۸-۲ ماتریس‌های نامنفی و همگرایی روش‌های تکراری ایستا

ماتریس‌های نامنفی نقش مهمی در مسائل مختلف علوم و مهندسی دارند. برای مثال، این‌گونه ماتریس‌ها در مباحث مربوط به زنجیرهای مارکوف، نظریه بازی‌ها به‌وفور ظاهر می‌شوند [۲۱]. در این بخش ضمن معرفی این‌گونه ماتریس‌ها، خواص همگرایی روش AOR را برای این رده از ماتریس‌ها بررسی می‌کنیم.

تعریف ۸.۲ فرض کنید $A, B \in \mathbb{R}^{n \times n}$. ماتریس $A = (a_{ij})$ را نامنفی (مثبت) می‌گویند و با $A \geq 0$ نشان می‌دهند هرگاه برای هر i و j داشته باشیم $a_{ij} \geq 0$ ($a_{ij} > 0$). می‌گویند A کوچکتر یا مساوی B (A کوچکتر از B) است و با $A \leq B$ ($A < B$) نشان می‌دهند هرگاه $B - A \geq 0$ ($B - A > 0$). به همین ترتیب این تعریف را می‌توان به‌طور مشابه برای بردارها نیز بیان کرد.

قضیه ۱۴.۲ فرض کنید $A, B, C \in \mathbb{R}^{n \times n}$

(الف) اگر $A \leq B$ و $B \leq A$ ، آنگاه $A = B$. به‌علاوه اگر $A \leq B$ و $B \leq C$ ، آنگاه $A \leq C$ ؛

(ب) اگر $A, B \in \mathbb{R}^{n \times n}$ نامنفی (مثبت) باشند، آنگاه AB و $A + B$ نامنفی (مثبت) هستند؛

(ج) اگر A نامنفی (مثبت) باشد، آنگاه A^k نیز نامنفی (مثبت) است؛

(د) اگر $A \leq B$ باشد، آنگاه $A^T \leq B^T$ ؛

(ه) اگر $0 \leq A \leq B$ ، آنگاه $\|A\|_1 \leq \|B\|_1$ و $\|A\|_\infty \leq \|B\|_\infty$.

□

برهان. اثبات این قضیه ساده است و به خواننده واگذار می‌شود.

فرض کنید $A, B \in \mathbb{R}^{n \times n}$ و $x, y \in \mathbb{R}^n$. در این صورت، به سادگی می‌توان دید که

$$A > 0, x \geq 0, x \neq 0 \implies Ax > 0, \quad (2-25)$$

$$A \geq 0, x \geq y \geq 0 \implies Ax \geq Ay, \quad (2-26)$$

$$A \geq 0, x > 0, Ax = 0 \implies A = 0, \quad (2-27)$$

$$A \geq 0, \det(A) \neq 0, x > y > 0 \implies Ax > Ay. \quad (2-28)$$

یکی از قضایای مهم در خصوص ماتریس‌های مثبت، قضیه پرون-فروبنیوس است. قبل از بیان این قضیه، چند نکته در خصوص ماتریس‌های مثبت را بیان می‌کنیم. به‌وضوح هر ماتریس مثبت تحویل‌ناپذیر است. اگر $A > 0$ ، آنگاه $\rho(A) > 0$. زیرا اگر $\rho(A) = 0$ ، آنگاه همه مقادیر ویژه A برابر با صفر خواهند بود و با توجه به شکل جردن ماتریس A (قضیه ۲۱.۱)، به‌ازای یک k ، داریم $A^k = 0$. اما با توجه به اینکه A مثبت است، بنا به قضیه ۱۴.۲ داریم، $A^k > 0$. بنابراین، $\rho(A)$ نمی‌تواند صفر باشد.

فرض کنید $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$. در این صورت، بردار $|x|$ به‌صورت

$$|x| = (|x_1|, \dots, |x_n|)^T,$$

تعریف می‌شود. به‌طور مشابه، برای ماتریس $A \in \mathbb{R}^{n \times n}$ ، ماتریس $|A|$ تعریف می‌شود.

۱۰۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تَنگ

قضیه ۱۵.۲ (پرون-فروبنیوس برای ماتریس‌های مثبت) اگر $A > 0$ ، آنگاه $\rho(A)$ یک مقدار ویژه A است و به علاوه یک بردار مثبت مثل u وجود دارد به طوری که $Au = \rho(A)u$. برهان. با توجه به اینکه $A/\rho(A) > 0$ و

$$\rho\left(\frac{1}{\rho(A)}A\right) = 1,$$

بدون از دست دادن کلیت، می‌توان فرض کرد $\rho(A) = 1$. فرض کنید (λ, x) یک زوج ویژه A باشد به طوری که $|\lambda| = 1$. در این صورت،

$$|x| = |\lambda||x| = |\lambda x| = |Ax| \leq |A||x| = A|x| \implies |x| \leq A|x|. \quad (29-2)$$

اکنون، نشان می‌دهیم که در اینجا تساوی برقرار است. فرض کنید $z = A|x|$ و $y = z - |x|$. با توجه به رابطه (۲۵-۲)، $z > 0$ و با توجه به (۲۹-۲) داریم $y \geq 0$. به برهان خلف فرض کنید $y \neq 0$ ، یعنی به ازای یک i ، داریم $y_i \neq 0$. در نتیجه، با توجه به رابطه (۲۵-۲) داریم $Ay > 0$. بنابراین، یک $\epsilon > 0$ وجود دارد به طوری که $Ay > \epsilon z$ ، که این رابطه نیز با رابطه

$$\frac{A}{1+\epsilon}z > z,$$

معادل است. با فرض $B = A/(1+\epsilon)$ ، این رابطه را می‌توان به صورت $Bz > z$ نوشت. به استقرا نشان می‌دهیم که

$$B^k z > z, \quad k = 1, 2, \dots$$

به ازای $k = 1$ ، به وضوح حکم برقرار است. فرض کنید حکم برای $k = m$ برقرار باشد، یعنی

$$B^m z > z.$$

در این صورت، از اینکه $B, z > 0$ ، با استفاده از (۲۸-۲)، داریم

$$B(B^m z) > Bz \implies B^{m+1} z > Bz > z.$$

اینک، داریم

$$\rho(B) = \rho\left(\frac{A}{1+\epsilon}\right) = \frac{1}{1+\epsilon} < 1 \implies \lim_{k \rightarrow \infty} B^k = 0.$$

بنابراین،

$$^{\circ} = \lim_{k \rightarrow \infty} B^k z \geq z.$$

این رابطه با $^{\circ} > z$ در تناقض است. ازاین‌رو، فرض خُلف باطل است و داریم

$$^{\circ} = y = A|x| - |x| \implies A|x| = |x|.$$

این رابطه نشان می‌دهد که $u = |x|$ یک بردار ویژه A متناظر به مقدار ویژه $\rho(A) = 1$ است. با توجه به اینکه $^{\circ} > u = |x| = A|x| = z$ اثبات قضیه کامل می‌شود. $\square$

در قضیه ۱۵.۲ می‌توان فرض کرد $A \geq ^{\circ}$. در این صورت، $\rho(A)$ یک مقدار ویژه A است و یک بردار نامنفی مثل u وجود دارد به‌طوری‌که $u \neq ^{\circ}$ و $Au = \rho(A)u$ [۷۷]. توجه کنید که در اینجا $\rho(A)$ می‌تواند برابر با صفر باشد.

در صورتی‌که ماتریس A نامنفی و تحویل‌ناپذیر باشد، $\rho(A)$ یک مقدار ویژه A است و به‌علاوه بردار مثبتی مثل u وجود دارد به‌طوری‌که $Au = \rho(A)u$ [۷۷]. توجه کنید که در اینجا نیز همانند قضیه ۱۵.۲، $\rho(A) > ^{\circ}$.

قضیه ۱۶.۲ فرض کنید ماتریس‌های A و B نامنفی باشند و $A \leq B$.

(الف) برای هر ماتریس نامنفی C ، داریم $AC \leq BC$ و $CA \leq CB$ ؛

(ب) برای هر $m \geq ^{\circ}$ ، $A^m \leq B^m$.

برهان. اثبات قسمت (الف) ساده است و به خواننده واگذار می‌شود. به استقرا قسمت (ب) را ثابت می‌کنیم. به‌وضوح نامساوی برای $m = ^{\circ}$ برقرار است. اکنون، فرض می‌کنیم نامساوی برای $m = k$ برقرار باشد، یعنی $A^k \leq B^k$. با توجه به قسمت (الف)، اگر A را از سمت چپ در طرفین این نامساوی ضرب کنیم، خواهیم داشت

$$A^{k+1} \leq AB^k.$$

از طرفی بنا به قضیه ۱۴.۲، ماتریس B^k نامنفی است. با ضرب B^k از راست در طرفین $A \leq B$ ، داریم

$$AB^k \leq B^{k+1}.$$

بنابراین، $A^{k+1} \leq B^{k+1}$ و این اثبات را کامل می‌کند. $\square$

قضیه ۱۷.۲ اگر $A, B \in \mathbb{R}^{n \times n}$ ، به‌طوری‌که $^{\circ} \leq A \leq B$ ، آنگاه $\rho(A) \leq \rho(B)$.

برهان. با توجه به قضیه ۱۶.۲، داریم

$$A^k \leq B^k.$$

با انتخاب نرم-۱ به‌عنوان نرم ماتریسی و قسمت (ه) از قضیه ۱۴.۲، داریم

$$\|A^k\|_1 \leq \|B^k\|_1.$$

بنابراین، با توجه به قضیه ۲.۲ نتیجه می‌گیریم که

$$\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|_1^{\frac{1}{k}} \leq \lim_{k \rightarrow \infty} \|B^k\|_1^{\frac{1}{k}} = \rho(B),$$

□

و بدین ترتیب اثبات قضیه کامل می‌شود.

تعریف ۹.۲ ماتریس A را یک M -ماتریس می‌گویند، هرگاه یک عدد حقیقی مثبت مثل s و یک ماتریس نامنفی مثل B وجود داشته باشند به‌طوری‌که

$$A = sI - B, \quad \rho(B) < s,$$

که در آن I ماتریس همانی است.

مثال ۱۲.۲ ماتریس

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -2 & 8 & -2 \\ 0 & -1 & 1 \end{pmatrix},$$

یک M -ماتریس است. زیرا اگر قرار دهیم $s = 8$ و

$$B = \begin{pmatrix} 6 & 1 & 1 \\ 2 & 0 & 2 \\ 0 & 1 & 7 \end{pmatrix},$$

آنگاه داریم

$$A = 8I - B, \quad \rho(B) = 7.5311 < 8.$$

بنابراین، طبق تعریف، A یک M -ماتریس است.

تعریف ۱۰.۲ ماتریس $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ را یک Z-ماتریس می‌گویند هرگاه به ازای هر $i \neq j$ ، $a_{ij} \leq 0$. ماتریس A را یک L-ماتریس می‌گویند هرگاه یک Z-ماتریس باشد و به ازای هر $i = 1, \dots, n$ ، $a_{ii} > 0$.

قضیه ۱۸.۲ ماتریس $A = (a_{ij})$ یک M-ماتریس است اگر و تنها اگر

(الف) A یک L-ماتریس باشد؛

(ب) $A^{-1} \geq 0$ موجود و $A^{-1} \geq 0$.

برهان. فرض کنید A یک M-ماتریس باشد. در این صورت، به ازای یک عدد مثبت s داریم

$$A = sI - B, \quad B \geq 0, \quad \rho(B) < s. \quad (30-2)$$

هر مقدار ویژه λ از A به صورت $\lambda = s - \mu$ است که در آن μ یک مقدار ویژه B است. از اینکه $\rho(B) < s$ ، نتیجه می‌گیریم که $|\mu| < s$. بنابراین، $\lambda \neq 0$ و در نتیجه، A نامنفرد است. قرار می‌دهیم $T = B/s \geq 0$ داریم

$$\rho(T) = \frac{1}{s} \rho(B) < 1.$$

بنابراین با توجه به قضیه ۲.۲، داریم

$$A^{-1} = [s(I - T)]^{-1} = \frac{1}{s} \sum_{k=0}^{\infty} T^k \geq 0.$$

با توجه به (۳۰-۲)، به وضوح برای هر i و j با شرط $i \neq j$ ، داریم $a_{ij} \leq 0$. از طرفی، فرض کنید که $A^{-1} = C = (c_{ij})$. در این صورت، $AC = I$ و در نتیجه، به ازای هر $i = 1, \dots, n$ ، داریم

$$(AC)_{ii} = 1 \implies \sum_{j=1}^n a_{ij} c_{ji} = 1 \implies a_{ii} c_{ii} = 1 - \sum_{j=1, j \neq i}^n a_{ij} c_{ji}.$$

سمت راست این عبارت مثبت است و $c_{ii} \geq 0$. بنابراین، $a_{ii} > 0$. در نتیجه، A یک L-ماتریس است.

برعکس، فرض کنید شرایط (الف) و (ب) برقرار باشند. قرار می‌دهیم $s = \max_i a_{ii}$ و

$B = sI - A$. در این صورت، داریم

$$A = sI - B, \quad B \geq 0, \quad s > 0.$$

ادعا می‌کنیم که $s > \rho(B)$. فرض کنید (λ, x) یک زوج ویژه B باشد. در نتیجه،

$$\begin{aligned} \lambda x = Bx &\implies |\lambda||x| = |Bx| \leq |B||x| \\ &\implies (sI - B)|x| \leq (s - |\lambda|)|x| \\ &\implies A|x| \leq (s - |\lambda|)|x| \\ &\implies 0 \leq |x| \leq (s - |\lambda|)A^{-1}|x|. \end{aligned} \quad (31-2)$$

از اینکه $0 \leq A^{-1}|x|$ ، نتیجه می‌گیریم که $s - |\lambda| \geq 0$. اما $s - |\lambda| \neq 0$ ، زیرا در غیر این صورت، بنا به رابطه (31-2)، داریم $x = 0$ که غیرممکن است؛ زیرا x یک بردار ویژه B است. بنابراین، $0 < s - |\lambda|$ و این رابطه نشان می‌دهد که $\rho(B) < s$. $\square$

مثال ۱۳.۲ دیدیم که ماتریس A در مثال ۱۲.۲ یک M -ماتریس است. ملاحظه می‌شود که به‌ازای هر $i = 1, \dots, n$ ، داریم $a_{ii} > 0$ ، و برای هر i و j با شرط $i \neq j$ ، $a_{ij} \leq 0$. به‌علاوه

$$A^{-1} = \begin{pmatrix} 0.75 & 0.25 & 1.25 \\ 0.25 & 0.25 & 0.75 \\ 0.25 & 0.25 & 1.75 \end{pmatrix} \geq 0.$$

قضیه ۱۹.۲ فرض کنید $A = (a_{ij})$ یک ماتریس $n \times n$ و L -ماتریس باشد. اگر ماتریس A غالب قطری اکید باشد، آنگاه یک M -ماتریس است.

برهان. فرض کنید $D = \text{diag}_n(a_{11}, \dots, a_{nn})$. به‌وضوح ماتریس $D^{-1}A$ را می‌توان به‌صورت $D^{-1}A = I - B$ نوشت که در آن $B \geq 0$. اگر قرار دهیم $B = (b_{ij})$ ، آنگاه از اینکه A غالب قطری اکید است، داریم

$$\sum_{j=1}^n |b_{ij}| < 1, \quad i = 1, 2, \dots, n.$$

از این رابطه نتیجه می‌گیریم

$$\rho(B) \leq \|B\|_{\infty} < 1.$$

بنابراین، با فرض $s = 1$ ، داریم

$$D^{-1}A = sI - B, \quad \rho(B) < s, \quad B \geq 0.$$

این رابطه نشان می‌دهد که $C = D^{-1}A$ یک M -ماتریس است. در نتیجه،

$$A = DC \implies A^{-1} = C^{-1}D^{-1} \geqslant 0,$$

□ که اثبات قضیه را کامل می‌کند.

مثال ۱۴.۲ با توجه به قضیه ۱۹.۲، ماتریس

$$A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & 0 & 2 \end{pmatrix},$$

یک M -ماتریس است، زیرا اولاً یک L -ماتریس است و ثانیاً غالب قطری اکید می‌باشد.

قضیه ۲۰.۲ فرض کنید A یک L -ماتریس باشد. در این صورت، A یک M -ماتریس است اگر و تنها اگر بردار مثبتی مثل x وجود داشته باشد به طوری که $Ax > 0$.

برهان. فرض کنید A یک M -ماتریس باشد و $e = (1, 1, \dots, 1)^T$. قرار می‌دهیم $x = A^{-1}e$. به وضوح $x \geqslant 0$. فرض کنید به ازای i ای $x_i = 0$. در این صورت،

$$\sum_{j=1}^n (A^{-1})_{ij} = 0,$$

که نتیجه می‌گیریم سطر i ام ماتریس A^{-1} برابر با صفر است. این نتیجه با نامنفرد بودن ماتریس A در تناقض است. بنابراین، $x > 0$ و در نتیجه، برای بردار x داریم $Ax = e > 0$.

برای اثبات عکس قضیه، فرض کنید $x = (x_1, x_2, \dots, x_n)^T > 0$ و $Ax > 0$. تعریف می‌کنیم $D = \text{diag}_n(x_1, x_2, \dots, x_n)$ در این صورت، داریم

$$ADe = Ax > 0,$$

و بنابراین،

$$\sum_{j=1}^n a_{ij}x_j > 0 \implies a_{ii}x_i > - \sum_{j=1, j \neq i}^n a_{ij}x_j \implies |a_{ii}x_i| > \sum_{j=1, j \neq i}^n |a_{ij}x_j|.$$

لذا، ماتریس AD غالب قطری اکید است. به وضوح AD یک L -ماتریس است. بنابراین، با استفاده از

۱۹.۲، ماتریس AD یک M -ماتریس است. به سادگی می‌توان دید که A نیز یک M -ماتریس است. □

مثال ۱۵.۲ L-ماتریس

$$A = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 3 & -1 \\ -1 & -1 & 2 \end{pmatrix},$$

را در نظر بگیرید. با فرض $x = (2, 1, 2)^T$ ، داریم $Ax = (1, 1, 1)^T > 0$. بنابراین، بنا به قضیه ۲۰.۲ ماتریس A یک M-ماتریس است.

قضیه ۲۱.۲ فرض کنید $B \geq 0$. در این صورت، $\rho(B) < 1$ اگر و تنها اگر $I - B$ نامنفرد باشد و $(I - B)^{-1} \geq 0$.

برهان. فرض کنید $\rho(B) < 1$. در این صورت، با فرض $s = 1 > 0$ ، داریم

$$I - B = sI - B, \quad \rho(B) < s, \quad B \geq 0.$$

بنابراین بنا به تعریف $I - B$ یک M-ماتریس است و در نتیجه، $I - B$ نامنفرد است و به علاوه $(I - B)^{-1} \geq 0$.

برای اثبات عکس قضیه ابتدا توجه کنید که بنا به قضیه پرون-فروبنیوس (قضیه ۱۵.۲)، $\rho(B)$ یک مقدار ویژه B است و یک بردار ویژه نامنفی مثل x وابسته به $\rho(B)$ برای ماتریس B وجود دارد، یعنی

$$Bx = \rho(B)x, \quad x \geq 0, \quad x \neq 0.$$

بنابراین، از اینکه $C = I - B$ نامنفرد است، داریم

$$C^{-1}x = \frac{1}{1 - \rho(B)}x.$$

حال با توجه به اینکه x و C^{-1} نامنفی هستند، نتیجه می‌گیریم که

$$1 - \rho(B) > 0,$$

و این اثبات قضیه را کامل می‌کند. □

قضیه ۲۲.۲ فرض کنید $A = (a_{ij})$ یک L-ماتریس باشد و $D = \text{diag}_n(a_{11}, \dots, a_{nn})$. در این صورت، A یک M-ماتریس است اگر و تنها اگر $\rho(B) < 1$ ، که در آن $B = I - D^{-1}A$.

برهان. فرض کنید A یک M -ماتریس باشد. به‌وضوح $B \geq 0$ و $I - B = D^{-1}A$. با توجه به اینکه ماتریس‌های A و D نامنفرد هستند، ماتریس $I - B$ نیز نامنفرد است و

$$(I - B)^{-1} = A^{-1}D \geq 0.$$

بنابراین، با توجه قضیه ۲۱.۲، نتیجه می‌گیریم $\rho(B) < 1$. برعکس فرض کنید A یک L -ماتریس باشد و $\rho(B) < 1$. داریم $B = I - D^{-1}A \geq 0$. با فرض $s = 1$ ، داریم

$$D^{-1}A = I - B = sI - B, \quad B \geq 0, \quad \rho(B) < s.$$

بنابراین $D^{-1}A$ یک M -ماتریس است. به‌سادگی می‌توان دید که A نیز یک M -ماتریس است. $\square$

قضیه ۲۳.۲ فرض کنید $A = (a_{ij})$ و $B = (b_{ij})$ دو ماتریس $n \times n$ باشند و
(الف) $A \leq B$ ؛

(ب) $i \neq j, b_{ij} \leq 0$.

اگر A یک M -ماتریس باشد، آنگاه B نیز یک M -ماتریس است.

برهان. فرض کنید A یک M -ماتریس باشد و

$$D_A = \text{diag}_n(a_{11}, \dots, a_{nn}), \quad D_B = \text{diag}_n(b_{11}, \dots, b_{nn}).$$

درایه‌های قطری D_B مثبتند، زیرا درایه‌های قطری A مثبت هستند و $D_B \geq D_A$. از اینکه $A \leq B$ ، داریم $0 \leq D_A - A \leq D_B - B$. با ضرب طرفین این رابطه از سمت چپ در D_A^{-1} ، خواهیم داشت

$$I - D_A^{-1}A \geq D_A^{-1}(D_B - B) \geq D_B^{-1}(D_B - B) = I - D_B^{-1}B \geq 0.$$

حال با استفاده از قضیه ۱۷.۲، می‌بینیم که

$$\rho(I - D_B^{-1}B) \leq \rho(I - D_A^{-1}A).$$

از طرفی، با استفاده از قضیه ۲۲.۲، داریم $\rho(I - D_A^{-1}A) < 1$. بنابراین،

$$\rho(I - D_B^{-1}B) < 1.$$

حال یک بار دیگر با استفاده از قضیه ۲۲.۲ نتیجه می‌گیریم که ماتریس B نیز یک M -ماتریس است. $\square$

۱۱۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

تعریف ۱۱.۲ (شکافت منظم) فرض کنید $A, M, N \in \mathbb{R}^{n \times n}$. در این صورت، $A = M - N$ را یک شکافت منظم برای A می‌گویند، هرگاه M نامنفرد باشد، $N \geq 0$ و $M^{-1} \geq 0$.

قضیه ۲۴.۲ فرض کنید $A = M - N$ یک شکافت منظم برای ماتریس A باشد. در این صورت، $\rho(M^{-1}N) < 1$ اگر و تنها اگر A نامنفرد باشد و $A^{-1} \geq 0$.

برهان. ابتدا توجه کنید که $G = M^{-1}N \geq 0$. فرض کنید $\rho(G) < 1$. در این صورت، بنا به قضیه ۲۱.۲ ماتریس $I - G$ نامنفرد است و $(I - G)^{-1} \geq 0$. بنابراین، از اینکه $A = M(I - G)$ نتیجه می‌شود که A نامنفرد است و

$$A^{-1} = (I - G)^{-1}M^{-1} \geq 0.$$

برای اثبات عکس قضیه، فرض می‌کنیم A نامنفرد باشد و $A^{-1} \geq 0$. از اینکه $A = M(I - G)$ نتیجه می‌شود که $I - G$ نامنفرد است. از طرفی،

$$A^{-1}N = (M(I - G))^{-1}N = (I - G)^{-1}M^{-1}N = (I - G)^{-1}G. \quad (32-2)$$

از اینکه ماتریس G نامنفی است، بنا به قضیه پرون-فروبنیوس $\rho(G)$ یک مقدار ویژه G است و برداری مثل x وجود دارد به طوری که

$$Gx = \rho(G)x, \quad x \geq 0, \quad x \neq 0.$$

از اینکه ماتریس $I - G$ نامنفرد است نتیجه می‌گیریم که $\rho(G) \neq 1$. بنابراین، با استفاده از رابطه (۳۲-۲)، می‌توان نوشت

$$A^{-1}Nx = \frac{\rho(G)}{1 - \rho(G)}x.$$

چون x و $A^{-1}N$ نامنفی هستند، داریم

$$\frac{\rho(G)}{1 - \rho(G)} \geq 0.$$

این رابطه وقتی می‌تواند برقرار باشد که

$$0 \leq \rho(G) < 1,$$

□

که اثبات قضیه را کامل می‌کند.

فرض کنید $A = M - N$ یک شکافت منظم برای A باشد و $A^{-1} \geq 0$. در این صورت، با استفاده از قضیه ۲۴.۲ روند تکراری ایستای

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b.$$

برای حل دستگاه $Ax = b$ به ازای هر حدس اولیه دلخواه همگرا خواهد بود.

مثال ۱۶.۲ ماتریس

$$A = \begin{pmatrix} 2 & -1 & -1 & -1 \\ -2 & 8 & -2 & 0 \\ -1 & -1 & 4 & -1 \\ -1 & -2 & -1 & 3 \end{pmatrix},$$

یک M -ماتریس است. قرار می‌دهیم

$$M = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -2 & 8 & -2 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 3 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{pmatrix}.$$

در این صورت، $A = M - N$ و $N \geq 0$. از طرفی، داریم $A \leq M$. بنابراین، با توجه به قضیه ۲۳.۲، M یک M -ماتریس است و $M^{-1} \geq 0$. در نتیجه، $A = M - N$ یک شکافت منظم برای A است و با توجه به قضیه ۲۴.۲، داریم $\rho(M^{-1}N) < 1$. با محاسبه شعاع طیفی ماتریس $M^{-1}N$ می‌بینیم که $\rho(M^{-1}N) = 0.9043 < 1$.

قضیه ۲۵.۲ اگر A یک M -ماتریس باشد، آنگاه روش‌های تکراری ژاکوبی و گaus-سایدل با هر حدس اولیه دلخواه همگرا خواهند بود.

برهان. همان‌گونه که دیدیم، در روش ژاکوبی داریم $M = D$ و $N = E + F$. روشن است که $N \geq 0$ و $D^{-1} \geq 0$. بنابراین، $A = M - N$ یک شکافت منظم برای A است و بنا به قضیه ۲۴.۲، داریم $\rho(G_J) = \rho(M^{-1}N) < 1$ ، که در آن G_J ماتریس تکرار روش ژاکوبی است. در نتیجه، روش ژاکوبی با هر حدس اولیه دلخواه همگرا خواهد بود. به طور مشابه در روش گaus-سایدل داریم $M = D - E$ و $N = F$. به وضوح $N \geq 0$. از طرفی، $A \leq D - E = M$ ، و درایه‌های غیرقطری M نامثبت‌اند.

۱۱۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

در نتیجه، بنا به قضیه ۲۳.۲ ماتریس M نیز یک M -ماتریس است. بنابراین، M نامنفرد است و $M^{-1} \geq 0$. لذا، $A = M - N$ یک شکافت منظم برای A است و بنا به قضیه ۲۴.۲، داریم $\rho(G_{GS}) = \rho(M^{-1}N) < 1$ ، که در آن G_{GS} ماتریس تکرار روش گاوس-سایدل است. در نتیجه، روش گاوس-سایدل با هر حدس اولیه دلخواه همگرا خواهد بود. $\square$

مثال ۱۷.۲ ماتریس مثال ۱۶.۲ را در نظر بگیرید. در این صورت، با محاسبه شعاع طیفی ماتریس‌های تکرار روش‌های ژاکوبی و گاوس-سایدل، داریم

$$\rho(G_J) = 0.9462 < 1, \quad \rho(G_{GS}) = 0.8937 < 1.$$

همان‌گونه که ملاحظه می‌شود قضیه ۲۵.۲ در اینجا صادق است.

قضیه ۲۶.۲ اگر A یک M -ماتریس باشد و $0 < \omega \leq 1$ ، آنگاه روش تکراری SOR به ازای هر حدس اولیه دلخواه همگرا است.

برهان. همان‌گونه که دیدیم در روش SOR، داریم

$$M = D - \omega E, \quad N = (1 - \omega)D + \omega F.$$

به وضوح داریم $N \geq 0$. از طرفی، $A \leq D - \omega E = M$ و درایه‌های غیرقطری M نامثبت هستند. بنابراین، بنا به قضیه ۲۳.۲، ماتریس M یک M -ماتریس است و $M^{-1} \geq 0$. در نتیجه، $A = M - N$ یک شکافت منظم برای ماتریس A است. اکنون با استفاده از قضیه ۲۴.۲، داریم

$$\rho(B_{SOR}(\omega)) = \rho(M^{-1}N) < 1.$$

$\square$ بنابراین، روش SOR با هر حدس اولیه دلخواه همگرا خواهد بود.

مثال ۱۸.۲ ماتریس مثال ۱۶.۲ را در نظر بگیرید. در این صورت، با محاسبه شعاع طیفی ماتریس تکرار روش‌های SOR به ازای $\omega = 0.25, 0.5, 0.75$ ، داریم

$$\rho(B_{SOR}(0.25)) = 0.9847,$$

$$\rho(B_{SOR}(0.5)) = 0.9645,$$

$$\rho(B_{SOR}(0.75)) = 0.9363.$$

همان‌گونه که ملاحظه می‌شود قضیه ۲۶.۲ در اینجا صادق است.

قضیه ۲۷.۲ فرض کنید A یک L -ماتریس باشد. در این صورت، به ازای هر $\omega \neq 0$ و γ ای که در رابطه $0 \leq \gamma \leq \omega \leq 1$ صدق کنند، روش AOR همگراست اگر و تنها اگر روش ژاکوبی همگرا باشد. برهان. روشن است که اگر روش AOR همگرا باشد، آنگاه روش ژاکوبی نیز همگراست. برای اثبات عکس قضیه، به برهان خلف فرض کنید $\mu = \rho(\mathcal{L}_{\gamma, \omega}) \geq 1$. با توجه به اینکه، داریم $0 \leq \gamma \leq \omega \leq 1$ و $(I - \gamma L)^{-1}$ ، نتیجه می‌گیریم

$$\mathcal{L}_{\gamma, \omega} = (I - \gamma L)^{-1}[(1 - \omega)I + (\omega - \gamma)L + \omega U] \geq 0.$$

بنا به قضیه پرون-فروبنیوس، μ یک مقدار ویژه $\mathcal{L}_{\gamma, \omega}$ است. اگر $v \neq 0$ بردار ویژه متناظر با این مقدار ویژه باشد، آنگاه $\mu v = \mathcal{L}_{\gamma, \omega} v$ و با توجه به تعریف $\mathcal{L}_{\gamma, \omega}$ خواهیم داشت

$$\left(\frac{\omega - \gamma + \gamma\mu}{\omega} L + U \right) v = \frac{\mu - 1 + \omega}{\omega} v.$$

این رابطه ایجاب می‌کند که $(\mu - 1 + \omega)/\omega$ یک مقدار ویژه $(\omega - \gamma + \gamma\mu)/\omega L + U$ است. بنابراین،

$$\frac{\mu - 1 + \omega}{\omega} \leq \rho \left(\frac{\omega - \gamma + \gamma\mu}{\omega} L + U \right). \quad (2-33)$$

با توجه به اینکه $(\omega - \gamma + \gamma\mu)/\omega \geq 1$ ، داریم

$$0 \leq \frac{\omega - \gamma + \gamma\mu}{\omega} L + U \leq \frac{\omega - \gamma + \gamma\mu}{\omega} (L + U) = \frac{\omega - \gamma + \gamma\mu}{\omega} J,$$

که در آن J ماتریس تکرار ژاکوبی است. با استفاده از قضیه ۱۷.۲ و رابطه اخیر، داریم

$$\rho \left(\frac{\omega - \gamma + \gamma\mu}{\omega} L + U \right) \leq \frac{\omega - \gamma + \gamma\mu}{\omega} \rho(J).$$

حال با استفاده از این رابطه و رابطه (۲-۳۳)، نتیجه می‌گیریم که

$$\frac{\mu - 1 + \omega}{\omega} \leq \frac{\omega - \gamma + \gamma\mu}{\omega} \rho(J),$$

یا

$$\rho(J) \geq \frac{\mu - 1 + \omega}{\omega - \gamma + \gamma\mu} = \frac{\omega + (\mu - 1)}{\omega + \gamma(\mu - 1)} \geq 1,$$

□

که یک تناقض است.

۹-۲ یک قضیه مهم و کاربردهای آن

در این بخش ابتدا یک قضیه اساسی در مبحث ماتریس‌های نامنفی را مطرح کرده، سپس کاربردهایی از این قضیه را بیان می‌کنیم.

قضیه ۲۸.۲ فرض کنید $A \geq 0$.

(الف) اگر به‌ازای یک بردار $x \geq 0$ و $x \neq 0$ و $\alpha > 0$ ، $Ax \geq \alpha x$ ، آنگاه $\rho(A) \geq \alpha$.

(ب) اگر به‌ازای یک $x > 0$ و $\beta > 0$ ، $Ax \leq \beta x$ ، آنگاه $\rho(A) \leq \beta$.

(ج) فرض کنید A تحویل‌ناپذیر باشد و برای یک بردار $x \geq 0$ ،

$$0 \neq \alpha x \leq Ax \leq \beta x,$$

آنگاه $\alpha \leq \rho(A) \leq \beta$ و $x > 0$.

□

برهان. به [۲۱، ۳] رجوع کنید.

با استفاده از این قضیه می‌توان روش‌های تکراری مختلف برای حل یک دستگاه را از لحاظ نرخ همگرایی مجانبی با یکدیگر مقایسه کرد یا نرخ همگرایی یک روش تکراری برای صورت‌های معادل یک دستگاه را مقایسه کرد. برای مثال قضیه بعدی در این خصوص ارائه می‌شود.

قضیه ۲۹.۲ فرض کنید A یک L-ماتریس باشد و J و G به‌ترتیب ماتریس‌های تکرار روش‌های ژاکوبی و گaus-سایدل باشند. در این صورت،

(الف) اگر ماتریس A تحویل‌ناپذیر باشد و $\rho(J) < 1$ ، آنگاه $\rho(G) \leq \rho(J) < 1$.

(ب) اگر $\rho(J) \geq 1$ ، آنگاه $\rho(G) \geq \rho(J) \geq 1$.

برهان. با توجه به اینکه A یک L-ماتریس است داریم

$$J = D^{-1}(E + F) \geq 0,$$

و از اینکه A تحویل‌ناپذیر است به‌روشنی J نیز تحویل‌ناپذیر است. بنابراین، $\lambda = \rho(J) > 0$ یک مقدار

ویژه J ، متناظر با بردار ویژه $x > 0$ است. لذا، $Jx = \lambda x$ که معادل با

$$(E + F)x = \lambda Dx,$$

است. اکنون، با استفاده از این رابطه، نتیجه می‌گیریم

$$\begin{aligned} Gx - \lambda x &= (D - E)^{-1}Fx - \lambda x \\ &= (D - E)^{-1}(Fx - \lambda(D - E)x) \\ &= (\lambda - 1)(D - E)^{-1}Ex. \end{aligned}$$

از اینکه $1 < \rho(D^{-1}E) = 0$ داریم

$$(D - E)^{-1} = (I - D^{-1}E)^{-1}D^{-1} \geq 0.$$

لذا، اگر $1 < \rho(J)$ ، آنگاه $0 \leq (\lambda - 1)(D - E)^{-1}Ex \leq Gx \leq \lambda x$ و بنابراین، بنا به قضیه ۲۸.۲ خواهیم داشت

$$\rho(G) \leq \lambda = \rho(J).$$

به‌طور مشابه در حالتی که ماتریس A تحویل‌پذیر است داریم $x \geq 0$. بنابراین، اگر $1 < \rho(J)$ ، آنگاه $0 \leq (\lambda - 1)(D - E)^{-1}Ex \leq Gx \leq \lambda x$ و با استفاده از قضیه ۲۸.۲، داریم $\rho(G) \geq \lambda = \rho(J)$. $\square$

این قضیه نشان می‌دهد که اگر A یک L -ماتریس و تحویل‌ناپذیر و روش ژاکوبی همگرا باشد آنگاه نرخ همگرایی مجانبی روش گاوس-سایدل دست‌کم به خوبی روش ژاکوبی خواهد بود. به‌علاوه اگر A یک L -ماتریس باشد و روش ژاکوبی واگرا باشد، آنگاه روش گاوس-سایدل دست‌کم به بدی روش ژاکوبی خواهد بود.

در ادامه، یکی دیگر از کاربردهای قضیه ۲۸.۲ را بیان می‌کنیم. دستگاه

$$Ax = b, \quad (2-34)$$

را در نظر بگیرید که در آن $A \in \mathbb{R}^{n \times n}$ و $b \in \mathbb{R}^n$. به‌علاوه فرض کنید که درایه‌های قطری A ناصفر باشند. بنابراین، بدون از دست دادن کلیت، می‌توان فرض کرد که درایه‌های قطری A برابر با یک می‌باشند (چرا؟). در این صورت، ماتریس A را به‌صورت $A = I - E - F$ ، تجزیه می‌کنیم که در آن E و F به‌ترتیب ماتریس‌های پائین‌مثلثی اکید و بالامثلثی اکید هستند. همان‌گونه که دیدیم ماتریس تکرار روش AOR به‌صورت

$$\mathcal{L}_{\gamma, \omega} = (I - \gamma E)^{-1}[(1 - \omega)I + (\omega - \gamma)E + \omega F],$$

است. قرار می‌دهیم $P = I + S$ که در آن S ماتریسی است که تمام درایه‌های آن صفرند به جز درایه $(n, 1)$ که برابر با $-a_{n1}$ است. درواقع، ماتریس P به صورت

$$P = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ -a_{n1} & & & & 1 \end{pmatrix}, \quad (2-35)$$

می‌باشد. طرفین دستگاه (2-34) را در P ضرب کرده و قرار می‌دهیم $\tilde{A} = PA$ و $\tilde{b} = Pb$. در اینجا P را یک پیش‌شرط‌ساز و دستگاه

$$\tilde{A}x = \tilde{b}, \quad (2-36)$$

را دستگاه پیش‌شرط‌سازی شده می‌نامند. در خصوص مفهوم پیش‌شرط‌سازی در فصل ۸ به‌طور جامع بحث خواهد شد. ماتریس $\tilde{A}$ را به صورت $\tilde{A} = \tilde{D} - \tilde{E} - \tilde{F}$ تجزیه می‌کنیم که در آن

$$\tilde{D} = (\tilde{d}_{ij}), \quad \tilde{d}_{ij} = \begin{cases} 1, & i = j, i \neq n, \\ 1 - a_{n1}a_{1n}, & i = j = n, \\ 0, & i \neq j, \end{cases}$$

$$\tilde{E} = (\tilde{e}_{ij}), \quad \tilde{e}_{ij} = \begin{cases} 0, & i \leq j, \\ -(a_{nj} - a_{n1}a_{1j}), & i = n \text{ و } 1 \leq j \leq n-1, \\ -a_{ij}, & \text{در غیر این صورت} \end{cases}$$

$$\tilde{F} = F.$$

توجه کنید که $SE = 0$ و

$$\tilde{A} = (I + S)(I - E - F) = (I - D') - (E + E' - S) - F,$$

که در آن $SF = D' + E'$ از این رو، خواهیم داشت $\tilde{A} = \tilde{D} - \tilde{E} - \tilde{F}$ که در آن

$$\tilde{D} = I - D', \quad \tilde{E} = E + E' - S \quad \tilde{F} = F.$$

اگر ماتریس $\tilde{D}$ نامنفرد باشد، آنگاه ماتریس تکرار روش AOR برای دستگاه پیش‌شرط‌سازی شده (۳۶-۲) به صورت

$$\tilde{\mathcal{L}}_{\gamma,\omega} = (\tilde{D} - \gamma\tilde{E})^{-1}[(1-\omega)\tilde{D} + (\omega-\gamma)\tilde{E} + \omega\tilde{F}],$$

خواهد بود که در آن $\tilde{D}$ ، $\tilde{E}$ و $\tilde{F}$ به ترتیب ماتریس‌های قطری، پائین‌مثلثی اکید و بالامثلثی اکید هستند. در ادامه، می‌خواهیم ببینیم که تحت چه شرایطی

$$\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) \leq \rho(\mathcal{L}_{\gamma,\omega}).$$

این رابطه نشان می‌دهد که نرخ همگرایی مجانبی روش تکراری AOR برای حل دستگاه پیش‌شرط‌سازی شده بزرگ‌تر از نرخ همگرایی مجانبی روش تکراری AOR برای حل دستگاه است.

لم ۳.۲ فرض کنید A و $\tilde{A}$ به ترتیب ماتریس ضرایب دستگاه‌های (۳۴-۲) و (۳۶-۲) باشند. اگر $0 \leq \gamma \leq \omega \leq 1$ ، $\omega \neq 0$ و $\gamma \neq 1$ و ماتریس A یک L -ماتریس تحویل‌ناپذیر باشد به طوری که $a_{1n}a_{n1} < 1$ ، آنگاه ماتریس‌های $\mathcal{L}_{\gamma,\omega}$ و $\tilde{\mathcal{L}}_{\gamma,\omega}$ نامنفی و تحویل‌ناپذیرند. برهان. ماتریس A یک L -ماتریس است. بنابراین، $E \geq 0$. حال با توجه به $0 < \rho(\gamma E) < 1$ داریم

$$(I - \gamma E)^{-1} = \sum_{k=0}^{n-1} (\gamma E)^k \geq 0.$$

از طرفی $0 \leq (1-\omega)I + (\omega-\gamma)E + \omega F$. در نتیجه، $\mathcal{L}_{\gamma,\omega} \geq 0$. به علاوه، داریم

$$\begin{aligned} \mathcal{L}_{\gamma,\omega} &= (I - \gamma E)^{-1}[(1-\omega)I + (\omega-\gamma)E + \omega F] \\ &= \sum_{k=0}^{n-1} (\gamma E)^k [(1-\omega)I + (\omega-\gamma)E + \omega F] \\ &= (1-\omega)I + \omega(1-\gamma)E + \omega F + T, \end{aligned}$$

که در آن $T \geq 0$. با توجه به اینکه A تحویل‌ناپذیر است، این رابطه نشان می‌دهد که ماتریس $\mathcal{L}_{\gamma,\omega}$ نیز تحویل‌ناپذیر است. با توجه به شرط $a_{1n}a_{n1} < 1$ و تعریف ماتریس $\tilde{D}$ می‌بینیم که درایه‌های قطری $\tilde{A}$ مثبت هستند. به روشنی $\tilde{E} \geq 0$. بنابراین، $\tilde{L} = \tilde{D}^{-1}\tilde{E} \geq 0$ و $\tilde{U} = \tilde{D}^{-1}\tilde{F} \geq 0$. لذا،

$$\begin{aligned} \tilde{\mathcal{L}}_{\gamma,\omega} &= (\tilde{D} - \gamma\tilde{E})^{-1}[(1-\omega)\tilde{D} + (\omega-\gamma)\tilde{E} + \omega\tilde{F}] \\ &= (I - \gamma\tilde{L})^{-1}[(1-\omega)I + (\omega-\gamma)\tilde{L} + \omega\tilde{U}]. \end{aligned}$$

بنابراین، همانند قسمت قبل می‌توان دید که $\tilde{\mathcal{L}}_{\gamma,\omega}$ نامنفی و تحویل‌ناپذیر است. $\square$

قضیه ۳۰.۲ فرض کنید تمام شرایط لم ۳.۲ برقرار باشند. در این صورت،

(الف) اگر $\rho(\mathcal{L}_{\gamma,\omega}) < 1$ ، آنگاه $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) \leq \rho(\mathcal{L}_{\gamma,\omega})$.

(ب) اگر $\rho(\mathcal{L}_{\gamma,\omega}) = 1$ ، آنگاه $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) = \rho(\mathcal{L}_{\gamma,\omega}) = 1$.

(ج) اگر $\rho(\mathcal{L}_{\gamma,\omega}) > 1$ ، آنگاه $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) \geq \rho(\mathcal{L}_{\gamma,\omega})$.

برهان. با توجه به لم ۳.۲ ماتریس $\mathcal{L}_{\gamma,\omega}$ نامنفی و تحویل‌ناپذیر است. از این رو، با استفاده از قضیه پرون-فروبنیوس یک بردار $x > 0$ وجود دارد به طوری که $\mathcal{L}_{\gamma,\omega}x = \lambda x$ ، که در آن $\lambda = \rho(\mathcal{L}_{\gamma,\omega})$. لذا،

$$[(1 - \omega)I + (\omega - \gamma)E + \omega F]x = \lambda(I - \gamma E)x. \quad (37-2)$$

با ضرب طرفین این رابطه در S ، داریم

$$\omega SFx = (\lambda + \omega - 1)Sx. \quad (38-2)$$

اکنون، داریم

$$\begin{aligned} \tilde{\mathcal{L}}_{\gamma,\omega}x - \lambda x &= (\tilde{D} - \gamma\tilde{E})^{-1}[(1 - \omega)\tilde{D}x + (\omega - \gamma)\tilde{E}x + \omega\tilde{F}x - \lambda(\tilde{D} - \gamma\tilde{E})x] \\ &= (\tilde{D} - \gamma\tilde{E})^{-1}[(1 - \omega)(I - D')x + (\omega - \gamma)(E + E' - S)x \\ &\quad + \omega Fx - \lambda((I - D') - \gamma(E + E' - S))x]. \end{aligned}$$

از رابطه (۳۷-۲)، داریم

$$\begin{aligned} \tilde{\mathcal{L}}_{\gamma,\omega}x - \lambda x &= (\tilde{D} - \gamma\tilde{E})^{-1}[(\omega - 1)D'x + (\omega - \gamma)(E' - S)x \\ &\quad + \lambda(D' + \gamma(E' - S))x] \\ &= (\tilde{D} - \gamma\tilde{E})^{-1}[(\lambda + \omega - 1)D'x + (\omega - \gamma + \lambda\gamma)(E' - S)x]. \end{aligned}$$

اکنون، با استفاده از اتحاد

$$\omega - \gamma + \lambda\gamma = (\omega + \lambda - 1) + (\gamma - 1)(\lambda - 1),$$

و (۳۸-۲) نتیجه می‌گیریم

$$\tilde{\mathcal{L}}_{\gamma,\omega}x - \lambda x = (\tilde{D} - \gamma\tilde{E})^{-1}[(\lambda + \omega - 1)(D' + E')x - (\lambda + \omega - 1)Sx]$$

$$\begin{aligned}
 & +(\lambda - 1)(\gamma - 1)(E' - S)x] \\
 & = (\tilde{D} - \gamma\tilde{E})^{-1}[(\lambda + \omega - 1)(D' + E')x - \omega SFx \\
 & \quad +(\lambda - 1)(\gamma - 1)(E' - S)x] \\
 & = (\tilde{D} - \gamma\tilde{E})^{-1}[(\lambda + \omega - 1)(D' + E')x - \omega(D' + E')x \\
 & \quad +(\lambda - 1)(\gamma - 1)(E' - S)x] \\
 & = (\lambda - 1)(\tilde{D} - \gamma\tilde{E})^{-1}[D' + \gamma E' + (1 - \gamma)S]x.
 \end{aligned}$$

داریم $D' \geq 0$ ، $E' \geq 0$ ، $S \geq 0$ و

$$(\tilde{D} - \gamma\tilde{E})^{-1} = (I - \gamma\tilde{D}^{-1}\tilde{E})^{-1}\tilde{D}^{-1} = \sum_{k=0}^{\infty}(\gamma\tilde{D}^{-1}\tilde{E})^k\tilde{D}^{-1} \geq 0.$$

در نتیجه، اگر $\lambda < 1$ ، آنگاه $\tilde{\mathcal{L}}_{\gamma,\omega}x - \lambda x \leq 0$ و بنابراین، با استفاده از قضیه ۲۸.۲، داریم $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) \leq \lambda = \rho(\mathcal{L}_{\gamma,\omega})$. اگر $\lambda = 1$ ، آنگاه $\tilde{\mathcal{L}}_{\gamma,\omega}x = \lambda x$ و از این رو، خواهیم داشت $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) = \lambda = \rho(\mathcal{L}_{\gamma,\omega})$. سرانجام اگر $\lambda > 1$ ، آنگاه $\tilde{\mathcal{L}}_{\gamma,\omega}x \geq \lambda x$ و بنابراین، با استفاده از قضیه ۲۸.۲، داریم $\rho(\tilde{\mathcal{L}}_{\gamma,\omega}) \geq \lambda = \rho(\mathcal{L}_{\gamma,\omega})$. □

تاکنون پیش‌شرط‌سازهای مختلفی شبیه به (۲-۳۵) و شکل کلی‌تری ارائه شده‌اند. کلی‌ترین نوع از این نوع پیش‌شرط‌سازها به صورت

$$P_{\alpha} = (p_{ij}), \quad (2-39)$$

می‌باشد که در آن

$$p_{ij} = \begin{cases} -\alpha_{ij}a_{ij}, & i \neq j, \\ 1, & i = j, \end{cases}$$

و $0 \leq \alpha_{ij} \leq 1$. در حالت خاصی از این نوع پیش‌شرط‌سازها، وُ و همکارانش در [۹۷] فرض کرده‌اند

$$\alpha_{ij} = \begin{cases} -\alpha_{ij}a_{ij}, & j = i + 1, \\ 0, & \text{درغیراین صورت} \end{cases}$$

و خواص پیش‌شرط‌ساز را بررسی نموده‌اند. حالت‌های خاص دیگری از P_{α} در [۴۸] مطالعه شده است. در [۹۵]، برخی از خواص P_{α} در حالت کلی مطالعه شده است. در [۵۷]، نشان داده شد که تحت شرایط

ناسختی، بهترین پیش‌شرط‌ساز از نوع P_α به‌ازای $\alpha_{ij} = 1, i, j = 1, 2, \dots, n$ بدست می‌آید. این قضیه را در اینجا ارائه می‌نماییم.

قضیه ۳۱.۲ فرض کنید $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ یک Z-ماتریس نامنفرد باشد و $k = 1, 2, P_{\alpha^{(k)}}$ مشابه (۳۹-۲) تعریف شوند. به‌علاوه فرض کنید $0 < \gamma \leq \omega \leq 1$ و به‌ازای هر $i \neq j$ داشته باشیم $0 < \alpha_{ij}^{(k)} \leq 1, k = 1, 2$. همچنین، فرض کنید $\mathcal{L}_{\gamma, \omega}(\alpha^{(1)})$ و $\mathcal{L}_{\gamma, \omega}(\alpha^{(2)})$ به‌ترتیب ماتریس تکرار روش AOR برای حل $Ax = b$ ، $P_{\alpha^{(1)}}Ax = P_{\alpha^{(1)}}b$ و $P_{\alpha^{(2)}}Ax = P_{\alpha^{(2)}}b$ باشند. در این صورت، اگر $\rho(\mathcal{L}_{\gamma, \omega}) < 1$ و به‌ازای هر i و j داشته باشیم $\alpha_{ij}^{(1)} \leq \alpha_{ij}^{(2)}$ ، آنگاه

$$\tilde{\mathcal{L}}_{\gamma, \omega}(\alpha^{(2)}) \leq \tilde{\mathcal{L}}_{\gamma, \omega}(\alpha^{(1)}).$$

نتیجه ۸.۲ قضیه ۳۱.۲ نتیجه می‌دهد که از بین پیش‌شرط‌سازهای P_α ، پیش‌شرط‌سازی که به‌ازای $\alpha_{ij} = 1$ بدست می‌آید، یعنی $P = I + L + U$ ، بهترین است.

مثال ۱۹.۲ معادله دیفرانسیل

$$-(u_{xx} + u_{yy}) + u_x + 2u_y = f(x, y), \quad (x, y) \in \Omega = (0, 1) \times (0, 1), \quad (40-2)$$

با شرایط مرزی دیریکله را در نظر گرفته و همانند روش ارائه‌شده در بخش ۱-۸، این معادله را گسسته‌سازی می‌کنیم. تعداد تقسیمات در هر کدام از محورهای x و y را $N + 1$ می‌گیریم. همچنین، قرار می‌دهیم $h = 1/(N + 1)$. در این صورت، ماتریس ضرایب دستگاه حاصل به‌صورت $A = I \otimes P + Q \otimes I$ خواهد بود که در آن

$$P = \text{tridiag}_N(-\frac{2+h}{h}, 1, -\frac{2-h}{h}), \quad Q = \text{tridiag}_N(-\frac{1+h}{4}, 0, -\frac{1-h}{4}).$$

توجه کنید که A یک ماتریس $n \times n$ است به‌طوری‌که $n = N^2$. شش پیش‌شرط‌ساز به‌صورت

$$P_0 = I,$$

$$P_1 = I + \omega L,$$

$$P_2 = I + \omega L + \omega U,$$

$$P_3 = I + L(\alpha) + U(\alpha),$$

$$P_4 = I + \text{tridiag}(\omega, \omega, -a_{i, i+1}),$$

جدول ۲-۲ نتایج عددی برای مثال ۱۹.۲.

n	(γ, ω)	P_0	P_1	P_2	P_3	P_4	P_5
۲۵	$(0.7, 0.8)$	۰.۸۳۱۷	۰.۷۹۶۴	۰.۷۴۰۴	۰.۷۴۸۶	۰.۷۶۵۷	۰.۶۳۲۳
۲۵	$(0.8, 1)$	۰.۷۷۳۹	۰.۷۳۰۵	۰.۶۵۴۰	۰.۶۴۴۴	۰.۶۷۹۸	۰.۵۱۳۸
۱۰۰	$(0.7, 0.8)$	۰.۹۴۷۴	۰.۹۳۵۰	۰.۹۱۲۵	۰.۹۱۱۶	۰.۹۲۳۰	۰.۸۶۷۷
۱۰۰	$(0.8, 1)$	۰.۹۲۸۹	۰.۹۱۳۵	۰.۸۸۲۱	۰.۸۸۱۵	۰.۸۹۳۳	۰.۸۲۲۱

$$P_5 = I + L + U,$$

در نظر می‌گیریم که در پیش‌شرط‌ساز P_3 مقادیر α_{ij} ها اعدادی تصادفی هستند که به‌طور یکنواخت در بازه $(0, 1)$ توزیع شده‌اند. یادآوری می‌کنیم وقتی که $P_0 = I$ ، به این معنا است که دستگاه، پیش‌شرط‌سازی نشده است. در جدول ۲-۲ شعاع طیفی روش تکراری AOR برای حل دستگاه

$$P_i Ax = P_i b, \quad i = 0, 1, \dots, 5$$

به‌ازای مقادیر مختلف γ و ω ارائه شده است. همان‌گونه که ملاحظه می‌شود از بین پیش‌شرط‌سازهایی که مورد آزمایش، بهترین پیش‌شرط‌ساز P_5 است.

۱۰-۲ روش شتاب چبیشف

الگوریتم شتاب چبیشف توسط گلوب برای بهبود همگرایی روش ریچاردسون ارائه شد [۴۰]. همان‌گونه که در زیربخش ۳-۳-۲ دیدیم روش تکراری ریچاردسون را می‌توان به‌صورت

$$x_k = (I - \alpha A)x_{k-1} + \alpha b = x_{k-1} + \alpha r_{k-1}, \quad (41-2)$$

نوشت، که در آن $r_{k-1} = b - Ax_{k-1}$. لازم به ذکر است که برای سادگی از اندیس پائین برای شماره تکرار استفاده می‌کنیم، زیرا در این روش با مؤلفه‌های بردارها سروکار نداریم. با محاسبه بردار مانده x_k ، خواهیم داشت

$$r_k = b - Ax_k = (I - \alpha A)r_{k-1}.$$

با فرض $e_k = x^* - x_k$ ، داریم $r_k = Ae_k$ ، که در آن x^* جواب دستگاه $Ax = b$ است. از این‌رو، با ضرب طرفین رابطه (۴۲-۲) در A^{-1} ، داریم

$$e_k = A^{-1}r_k = A^{-1}(I - \alpha A)r_{k-1}$$

۱۲۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

$$= (I - \alpha A)A^{-1}r_{k-1} = (I - \alpha A)e_{k-1}. \quad (۴۲-۲)$$

ایده اصلی روش شتاب چبیشف این است که به جای استفاده از یک پارامتر ثابت در همه تکرارها، در هر تکرار از یک پارامتر مخصوص به همان تکرار استفاده شود. در این صورت، روابط (۴۱-۲) و (۴۲-۲) به ترتیب به صورت

$$x_k = (I - \alpha_k A)x_{k-1} + \alpha_k b = x_{k-1} + \alpha_k r_{k-1}, \quad (۴۳-۲)$$

و

$$e_k = (I - \alpha_k A)e_{k-1},$$

خواهند بود. از این رو، به استقرا خواهیم داشت

$$e_k = \prod_{j=1}^k (I - \alpha_j A)e_0.$$

بنابراین، با فرض

$$p_k(t) = \prod_{j=1}^k (1 - \alpha_j t),$$

داریم

$$e_k = p_k(A)e_0, \quad p_k(0) = 1. \quad (۴۴-۲)$$

روشن است که ریشه‌های چندجمله‌ای p_k به صورت $t_j = 1/\alpha_j$, $j = 1, 2, \dots, k$ هستند. در نتیجه، اگر بتوانیم t_j ها را مقادیر ویژه A اختیار کنیم، یا به عبارت دیگر قرار دهیم

$$\alpha_j = \frac{1}{\lambda_j}, \quad j = 1, 2, \dots, n, \quad (۴۵-۲)$$

که در آن λ_j ها مقادیر ویژه A هستند، آنگاه براساس قضیه کیلی-همیلتون ۸.۱، خواهیم داشت $p_n(A) = 0$. از این رو، داریم $e_n = 0$ و روش در حداکثر n تکرار همگرا خواهد شد. اما معمولاً مقادیر ویژه ماتریس A در دست نیست و محاسبه آنها نیز کار دشواری است. بنابراین، این انتخاب، عملی نیست. برای حل این مشکل، خود را به رده‌ای خاص از ماتریس‌ها محدود می‌کنیم.

فرض کنید ماتریس A معین مثبت متقارن و مقادیر ویژه آن به صورت

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n,$$

باشند. ازاینکه ماتریس A متقارن است، با استناد به قضیه ۱۷.۱ یک ماتریس متعامد مثل U وجود دارد به گونه‌ای که

$$A = UDU^T, \quad D = \text{diag}_n(\lambda_1, \lambda_2, \dots, \lambda_n).$$

با توجه به اینکه ماتریس U متعامد است، داریم

$$\|U\|_2 = 1, \quad \|U^T e_0\|_2 = \|e_0\|_2.$$

ازاین‌رو، از رابطه (۵۳-۲)، داریم

$$\begin{aligned} \|e_k\|_2 &= \|p_k(A)e_0\|_2 \\ &= \|p_k(UDU^T)e_0\|_2 \\ &= \|Up_k(D)U^T e_0\|_2 \\ &\leq \|U\|_2 \|p_k(D)\|_2 \|U^T e_0\|_2 \\ &= \max_{1 \leq i \leq n} |p_k(\lambda_i)| \|e_0\|_2. \end{aligned} \quad (۴۶-۲)$$

با توجه به اینکه مقادیر ویژه A نامعلوم هستند، مقدار $\max_{1 \leq i \leq n} |p_k(\lambda_i)|$ همچنان نامعلوم خواهد بود. برای رفع این مشکل، فرض کنید یک کران از مقادیر ویژه A در دست باشد؛ به عبارت دیگر، فرض کنید

$$\alpha \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq \beta.$$

دراین صورت، با توجه به رابطه (۴۶-۲)، داریم

$$\|e_k\|_2 \leq \max_{1 \leq i \leq n} |p_k(\lambda_i)| \|e_0\|_2 \leq \max_{t \in [\alpha, \beta]} |p_k(t)| \|e_0\|_2.$$

فرض کنید $\mathbb{P}_k$ مجموعه چندجمله‌ای‌های از درجه حداکثر k باشد. اکنون، از بین چندجمله‌ای‌های $p_k \in \mathbb{P}_k$ با شرط $p_k(0) = 1$ ، چندجمله‌ای که مقدار $\max_{t \in [\alpha, \beta]} |p_k(t)|$ را کمینه می‌کند را می‌یابیم. کلید حل این مسئله، چندجمله‌ای‌های چبیشف است. نخست این چندجمله‌ای‌ها را معرفی کرده و چند خاصیت مهم آنها را مطالعه می‌کنیم.

چندجمله‌ای جیبش نوع اول از درجه k به صورت

$$T_k(t) = \cos(k \cos^{-1} t), \quad -1 \leq t \leq 1, \quad (47-2)$$

تعریف می‌شود. به سادگی می‌توان دید که ریشه‌های T_k عبارتند از

$$t_j = \cos \frac{(j+1)\pi}{2n}, \quad j = 0, 1, \dots, k-1. \quad (48-2)$$

با فرض $\theta = \cos^{-1} t$ داریم

$$\cos((k+1)\theta) + \cos((k-1)\theta) = 2 \cos \theta \cos k\theta,$$

که رابطه بازگشتی

$$T_{k+1}(t) = 2tT_k(t) - T_{k-1}(t), \quad (49-2)$$

برای چندجمله‌ای‌های جیبش را به دست می‌دهد. توجه کنید که به ازای $k=1$ و $k=0$ به ترتیب چندجمله‌ای‌های $T_1(t) = t$ و $T_0(t) = 1$ به دست می‌آیند.

اگرچه چندجمله‌ای‌های جیبش توسط رابطه (47-2) برای $t \in [-1, 1]$ تعریف می‌شود، می‌بینیم که چندجمله‌ای‌هایی که توسط رابطه بازگشتی (49-2) تولید می‌شوند برای هر $t \in \mathbb{R}$ تعریف می‌شوند. از این رو، می‌توانیم چندجمله‌ای‌های جیبش را برای هر $t \in \mathbb{R}$ تعمیم دهیم. به ازای هر $t \in [-1, 1]$ داریم

$$\begin{aligned} T_k(t) &= \cos k\theta = \frac{1}{2} (e^{ik\theta} + e^{-ik\theta}) = \frac{1}{2} \left((e^{i\theta})^k + (e^{i\theta})^{-k} \right) \\ &= \frac{1}{2} \left((\cos \theta + i \sin \theta)^k + (\cos \theta + i \sin \theta)^{-k} \right) \\ &= \frac{1}{2} \left(\left(\cos \theta + i \sqrt{1 - \cos^2 \theta} \right)^k + \left(\cos \theta + i \sqrt{1 - \cos^2 \theta} \right)^{-k} \right) \\ &= \frac{1}{2} \left(\left(t + i \sqrt{1 - t^2} \right)^k + \left(t + i \sqrt{1 - t^2} \right)^{-k} \right). \end{aligned}$$

بنابراین، کافیت به ازای هر $|t| > 1$ ، چندجمله‌ای T_k را به صورت

$$T_k(t) = \frac{1}{2} \left(\left(t + \sqrt{t^2 - 1} \right)^k + \left(t + \sqrt{t^2 - 1} \right)^{-k} \right), \quad (50-2)$$

تعریف کنیم.

قضیه ۳۲.۲ فرض کنید $[\alpha, \beta]$ یک بازهٔ ناتهی از $\mathbb{R}$ باشد به‌طوری‌که عدد حقیقی γ در خارج این بازه قرار دارد. در این صورت،

$$\min_{p \in \mathbb{P}_m, p(\gamma)=1} \max_{t \in [\alpha, \beta]} |p(t)| = \max_{t \in [\alpha, \beta]} \left| \frac{T_m(1 + 2 \frac{t-\beta}{\beta-\alpha})}{T_m(1 + 2 \frac{\gamma-\beta}{\beta-\alpha})} \right| = \frac{1}{\left| T_m(2 \frac{\gamma-\mu}{\beta-\alpha}) \right|},$$

که در آن $\mu = (\alpha + \beta)/2$.

□

برهان. مراجع [۲۶، ۳۹] را ببینید.

با توجه به این قضیه و انتخاب $\gamma = \circ$ ، اگر قرار دهیم

$$q_k(t) = \frac{T_k \left(1 + 2 \frac{\alpha-t}{\beta-\alpha} \right)}{T_k \left(1 + 2 \frac{\alpha}{\beta-\alpha} \right)}, \quad (51-2)$$

آنگاه خواهیم داشت

$$q_k = \arg \min_{p_k \in \mathbb{P}_k, p_k(\circ)=1} \max_{\alpha \leq t \leq \beta} |p_k(t)|. \quad (52-2)$$

درحقیقت، چندجمله‌ای q_k مقدار $\max_{t \in [\alpha, \beta]} |p_k(t)|$ را از بین همهٔ چندجمله‌ای‌های $p_k \in \mathbb{P}_k$ با شرط $p_k(\circ) = 1$ کمینه می‌کند. از این رو، قرار می‌دهیم

$$e_k = q_k(A)e_{\circ}. \quad (53-2)$$

لازم به ذکر است که ریشه‌های چندجمله‌ای q_k و $T_k \left(1 + 2 \frac{\alpha-t}{\beta-\alpha} \right)$ یکسان هستند. با توجه به رابطه (۴۸-۲)، این ریشه‌ها عبارتند از

$$t_j = \frac{1}{2} \left(\beta + \alpha - (\beta - \alpha) \cos \frac{(2j+1)\pi}{2n} \right), \quad j = \circ, 1, \dots, k-1.$$

از این رو، اگر بخواهیم m تکرار از (۴۳-۲) را انجام دهیم، با توجه به (۴۵-۲) کافی است قرار دهیم

$$\alpha_j = \frac{1}{t_{j-1}}, \quad j = 1, 2, \dots, m.$$

در [۶۸] نشان داده شده است که این فرایند متأسفانه ناپایدار است. از طرفی، تعیین پارامتر m از پیش غیرممکن است. برای رفع این دو مشکل از رابطه بازگشتی (۴۹-۲) برای چندجمله‌ای چبیشف استفاده می‌شود. در ادامه، روشی که با این ایده به دست می‌آید را معرفی می‌کنیم.

با فرض

$$\theta = \frac{\beta + \alpha}{2}, \quad \delta = \frac{\beta - \alpha}{2},$$

تابع q_k در رابطه (۵۱-۲) را می‌توان به صورت زیر نوشت

$$q_k(t) = \frac{1}{\sigma_k} T_k \left(\frac{\theta - t}{\delta} \right), \quad \sigma_k = T_k \left(\frac{\theta}{\delta} \right). \quad (54-2)$$

با توجه به رابطه بازگشتی (۴۹-۲) برای چندجمله‌ای چبیشف، داریم

$$\begin{aligned} T_0 \left(\frac{\theta}{\delta} \right) &= 1, \quad T_1 \left(\frac{\theta}{\delta} \right) = \frac{\theta}{\delta}, \\ T_{k+1} \left(\frac{\theta}{\delta} \right) &= 2 \frac{\theta}{\delta} T_k \left(\frac{\theta}{\delta} \right) - T_{k-1} \left(\frac{\theta}{\delta} \right), \quad k = 1, 2, \dots \end{aligned}$$

بنابراین، با توجه به تعریف σ_k ، داریم

$$\sigma_0 = 1, \quad \sigma_1 = \frac{\theta}{\delta}, \quad (55-2)$$

$$\sigma_{k+1} = 2 \frac{\theta}{\delta} \sigma_k - \sigma_{k-1}, \quad k = 1, 2, \dots \quad (56-2)$$

از طرفی، با استفاده از رابطه بازگشتی (۴۹-۲) برای چندجمله‌ای چبیشف، داریم

$$\begin{aligned} q_{k+1}(t) &= \frac{1}{\sigma_{k+1}} T_{k+1} \left(\frac{\theta - t}{\delta} \right) \\ &= \frac{1}{\sigma_{k+1}} \left[2 \frac{\theta - t}{\delta} \sigma_k q_k(t) - \sigma_{k-1} q_{k-1}(t) \right] \\ &= \frac{\sigma_k}{\sigma_{k+1}} \left[2 \frac{\theta - t}{\delta} q_k(t) - \frac{\sigma_{k-1}}{\sigma_k} q_{k-1}(t) \right], \quad k = 1, 2, \dots, \quad (57-2) \end{aligned}$$

به طوری که

$$\begin{aligned} q_1(t) &= \frac{1}{\sigma_1} T_1 \left(\frac{\theta - t}{\delta} \right) = \frac{\delta}{\theta} \times \frac{\theta - t}{\delta} = 1 - \frac{t}{\theta}, \\ q_0(t) &= 1. \end{aligned}$$

فرض کنید

$$\rho_k = \frac{\sigma_k}{\sigma_{k+1}}, \quad k = 1, 2, \dots$$

در این صورت، از (۵۶-۲)، داریم

$$\frac{\sigma_{k+1}}{\sigma_k} = \frac{\theta}{\delta} - \frac{\sigma_{k-1}}{\sigma_k},$$

که با استفاده از (۵۵-۲) نتیجه می‌دهد

$$\frac{1}{\rho_k} = \frac{\theta}{\delta} - \frac{\rho_{k-1}}{\rho_k},$$

یا

$$\rho_k = \frac{1}{\frac{\theta}{\delta} - \rho_{k-1}}. \quad (58-2)$$

از این رو، با استفاده از رابطه (۵۷-۲)، چند جمله‌ای q_{k+1} را می‌توان به صورت زیر نوشت

$$q_{k+1}(t) = \rho_k \left[\frac{\theta}{\delta} \left(\sigma_1 - \frac{t}{\delta} \right) q_k(t) - \rho_{k-1} q_{k-1}(t) \right], \quad k = 1, 2, \dots \quad (59-2)$$

توجه کنید که اگر قرار دهیم $q_{-1} = 0$ و $\rho_{-1} = 0$ ، آنگاه خواهیم داشت $1/(\frac{\theta}{\delta} - \rho_{-1}) = \rho_0$ و رابطه‌های بازگشتی (۵۸-۲) و (۵۹-۲) را می‌توانیم از $k = 0$ آغاز کنیم.

اکنون، فرض کنید $d_k = x_{k+1} - x_k$. در این صورت، با توجه به رابطه (۵۳-۲)، داریم

$$\begin{aligned} d_k &= x_{k+1} - x_k = x_{k+1} - x^* - (x_k - x^*) \\ &= -(e_{k+1} - e_k) = -(q_{k+1}(A) - q_k(A))e_0 \\ &= -(q_{k+1}(A) - q_k(A))A^{-1}r_0 \\ &= -A^{-1}(q_{k+1}(A) - q_k(A))r_0 \\ &= -g_k(A)r_0, \end{aligned}$$

که در آن $g_k(t) = (q_{k+1}(t) - q_k(t))/t$. اینک، با استفاده از اتحاد $1 = (\frac{\theta}{\delta} - \rho_{k-1})\rho_k$ و رابطه (۵۹-۲)، داریم

$$\begin{aligned} g_k(t) &= \frac{1}{t} [q_{k+1}(t) - q_k(t)] \\ &= \frac{1}{t} \left(\rho_k \left[\frac{\theta}{\delta} \left(\sigma_1 - \frac{t}{\delta} \right) q_k(t) - \rho_{k-1} q_{k-1}(t) \right] - (\frac{\theta}{\delta} - \rho_{k-1}) \rho_k q_k(t) \right) \\ &= \rho_k \left[\rho_{k-1} \frac{q_k(t) - q_{k-1}(t)}{t} - \frac{\theta}{\delta} q_k(t) \right] \end{aligned}$$

۱۳۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

$$= \rho_k \left[\rho_{k-1} g_{k-1}(t) - \frac{1}{\delta} q_k(t) \right].$$

ازاین‌رو،

$$g_k(A)r_0 = \rho_k \left[\rho_{k-1} g_{k-1}(A)r_0 - \frac{1}{\delta} q_k(A)r_0 \right].$$

درنتیجه، داریم

$$-d_k = \rho_k \left[-\rho_{k-1} d_{k-1} - \frac{1}{\delta} r_k \right],$$

که معادل است با

$$d_k = \rho_k \rho_{k-1} d_{k-1} + \frac{1}{\delta} \rho_k r_k.$$

اکنون با استفاده از نتایج به‌دست‌آمده الگوریتم شتاب چبیشف را به‌صورت زیر جمع‌بندی می‌کنیم.

الگوریتم ۱.۲ شتاب چبیشف

1. Input a lower bound α and an upper bound β for the eigenvalues of A
2. Set $\theta = (\beta + \alpha)/2$ and $\delta = (\beta - \alpha)/2$
3. Choose an initial guess x_0
4. $r_0 := b - Ax_0$, $\sigma_1 := \theta/\delta$, $\rho_0 := 1/\sigma_1$, and $d_0 := r_0/\theta$
5. For $k = 0, 1, \dots$, until convergence Do
6. $x_{k+1} := x_k + d_k$
7. $r_{k+1} := r_k - Ad_k$
8. $\rho_{k+1} := 1/(2\sigma_1 - \rho_k)$
9. $d_{k+1} := \rho_{k+1}\rho_k d_k + \frac{2\rho_{k+1}}{\delta} r_{k+1}$
10. EndDo

چند نکته درخصوص روش شتاب چبیشف قابل ذکر است.

۱. در این بخش الگوریتم شتاب چبیشف را برای ماتریس‌های معین مثبت متقارن به‌دست آورده‌ایم، اما باید توجه داشت که می‌توان این الگوریتم را برای هر ماتریس نامنفردی که مقادیر ویژه آن حقیقی و در بازه $[\alpha, \beta]$ قرار داشته باشند به‌کار برد. هرچه این بازه کوچک‌تر باشد سرعت همگرایی روش بیشتر خواهد بود.
۲. در این بخش، استخراج روش شتاب چبیشف براساس روش تکراری ریچاردسون بود. اما، باید توجه کرد که هر روش تکراری ایستا چیزی نیست جز روش تکراری ریچاردسون برای دستگاه پیش‌شرط‌سازی‌شده $M^{-1}Ax = M^{-1}b$. درحقیقت، اگر $A = M - N$ یک شکافت از ماتریس A باشد، آنگاه روند

تکراری ایستای متناظر را می‌توان به صورت زیر نوشت

$$x_k = (I - M^{-1}A)x_{k-1} + M^{-1}b. \quad (۶۰-۲)$$

با فرض $H = M^{-1}A$ و $f = M^{-1}b$ روند $(۶۰-۲)$ را می‌توان به صورت زیر نوشت

$$x_k = (I - H)x_{k-1} + f, \quad (۶۱-۲)$$

که همان روش تکراری ریچاردسون برای حل دستگاه $Hx = f$ است. از این رو، کافیت در الگوریتم چبیشف به جای ماتریس A و بردار b ، به ترتیب ماتریس H و بردار f را جایگزین کنیم. این الگوریتم را می‌توان به صورت زیر جمع‌بندی کرد.

الگوریتم ۲.۲ شتاب چبیشف برای حل دستگاه $M^{-1}Ax = M^{-1}b$

1. Input a lower bound α and an upper bound β for the eigenvalues of $M^{-1}A$
2. Set $\theta = (\beta + \alpha)/2$ and $\delta = (\beta - \alpha)/2$
3. Choose an initial guess x_0
4. $r_0 := b - Ax_0$, $z_0 = M^{-1}r_0$, $\sigma_1 := \theta/\delta$, $\rho_0 := 1/\sigma_1$, and $d_0 := z_0/\theta$
5. For $k = 0, 1, \dots$, until convergence Do
6. $x_{k+1} := x_k + d_k$
7. $r_{k+1} := r_k - Ad_k$
8. $z_{k+1} = M^{-1}r_{k+1}$
9. $\rho_{k+1} := 1/(2\sigma_1 - \rho_k)$
10. $d_{k+1} := \rho_{k+1}\rho_k d_k + \frac{2\rho_{k+1}}{\delta} z_{k+1}$
11. EndDo

تمرینات

۱.۲ روش تکراری

$$x^{(k+1)} = x^{(k)} + \alpha P^{-1}r^{(k)}, \quad (۶۲-۲)$$

برای حل دستگاه $Ax = b$ را در نظر بگیرید که در آن $r^{(k)} = b - Ax^{(k)}$ یک ماتریس نامنفرد است و $\alpha > 0$. این روند تکراری را با روش تکراری ریچاردسون مقایسه نمایید. فرض کنید

$$P^{-1} = D^{-1} \left(I + CD^{-1} + (CD^{-1})^2 \right),$$

۱۳۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

که در آن D بخش قطری ماتریس A و $C = D - A$. فرض کنید مقادیر ویژه $D^{-1}C$ حقیقی باشند و $\rho(D^{-1}C) < 1$. حدود α را طوری بیابید که روش تکراری (۶۲-۲) همگرا باشد.

۲.۲ فرض کنید که $A \in \mathbb{C}^{n \times n}$ و

$$A = M_i - N_i, \quad i = 1, 2,$$

دو شکافت برای ماتریس A و $x_0 \in \mathbb{C}^n$ یک حدس اولیه برای جواب واقعی دستگاه $Ax = b$ باشد. همچنین، فرض کنید دنباله $\{x^k\}$ توسط روند تکراری دومرحله‌ای

$$\begin{cases} M_1 x^{(k+\frac{1}{2})} = N_1 x^k + p, \\ M_2 x^{(k+1)} = N_2 x^{(k+\frac{1}{2})} + q, \end{cases} \quad k = 0, 1, 2, \dots,$$

تولید می‌شود. این روش را به‌صورت یک روند تکراری یک مرحله‌ای نوشته و نشان دهید که به‌ازای هر حدس اولیه دلخواه همگراست اگر و تنها اگر

$$\rho(M_2^{-1} N_2 M_1^{-1} N_1) < 1.$$

۳.۲ معادله دیفرانسیل

$$\begin{cases} -u'' + cu = f, & x \in (0, 1), \\ u(0) = u(1) = 0, \end{cases}$$

را در نظر بگیرید که در آن $c > 0$. بازه $[0, 1]$ را به $n+1$ قسمت مساوی تقسیم کرده و قرار می‌دهیم $h = 1/(n+1)$. سپس، با استفاده تقریب تفاضل مرکزی برای مشتق دوم این معادله را گسسته‌سازی می‌کنیم. نشان دهید دستگاهی به‌صورت $Ax = b$ به‌دست می‌آید که در آن

$$A = \text{tridiag}_n(-1, 2 + \gamma, -1),$$

به‌طوری‌که $\gamma = c/(n+1)^2$. نشان دهید که مقادیر ویژه ماتریس A به‌صورت

$$\lambda_k = \gamma + 4 \sin^2 \left(\frac{k\pi}{2(n+1)} \right), \quad k = 1, \dots, n,$$

می‌باشند. یک روند تکراری برای حل دستگاه $Ax = b$ به‌صورت

$$x^{(k+1)} = G_\omega x^{(k)} + \frac{w}{2 + \gamma} b,$$

تعریف می‌کنیم که در آن

$$G_{\omega} = I - \frac{\omega}{2 + \gamma} A.$$

حدودی برای پارامتر ω بیابید تا روش همگرا شود. مقدار بهینه پارامتر ω را بیابید.

۴.۲ فرض کنید $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ و $a_{ii} \neq 0$ ، $i = 1, 2, \dots, n$. ماتریس A را به صورت

$$A = T - E - F,$$

تجزیه می‌کنیم که در آن

$$T = \begin{pmatrix} a_{1,1} & & & & \\ a_{2,1} & a_{2,2} & & & \\ & a_{3,2} & a_{3,3} & & \\ & & \ddots & \ddots & \\ & & & a_{n,n-1} & a_{n,n} \end{pmatrix},$$

$$F = \begin{pmatrix} 0 & -a_{1,2} & -a_{1,3} & \cdots & -a_{1,n} \\ & 0 & -a_{2,3} & & \vdots \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & 0 & -a_{n-1,n} \\ & & & & & 0 \end{pmatrix},$$

و $E = T - F - A$. اکنون، یک روند تکراری به صورت زیر برای حل دستگاه معادلات خطی $Ax = b$ می‌سازیم

$$x^{(m+1)} = T^{-1}(E + F)x^{(m)} + T^{-1}b, \quad m = 0, 1, 2, \dots \quad (۶۳-۲)$$

نشان دهید اگر A غالب قطری اکید یا یک M -ماتریس باشد، آنگاه روند (۶۳-۲) با هر حدس اولیه دلخواه $x^{(0)}$ به جواب دستگاه $Ax = b$ همگراست.

راهنمایی: رجوع کنید به [۵۵].

۱۳۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

۵.۲ فرض کنید $A = C - R$ یک شکافت از A باشد (در اینجا C نامنفرد است). یک شکافت کلی‌تر به صورت

$$A = C(\alpha) - R(\alpha),$$

در نظر می‌گیریم که در آن

$$C(\alpha) = (1 + \alpha)C, \quad R(\alpha) = C(\alpha) - A = R + \alpha C.$$

فرض کنید $1 > \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n < 1$ مقادیر ویژه $C^{-1}R$ باشند. نشان دهید که روند تکراری

$$C(\alpha)x^{(k+1)} = R(\alpha)x^{(k)} + b, \quad k = 0, 1, \dots,$$

به ازای هر $\alpha > -(1 + \lambda_1)/2$ همگراست و

$$\min_{\alpha} \rho(C(\alpha)^{-1}R(\alpha)) = \rho(C(\alpha^*)^{-1}R(\alpha^*)), \quad \alpha^* = -\frac{\lambda_1 + \lambda_n}{2}.$$

۶.۲ فرض کنید $A \in \mathbb{R}^{m \times n}$ به طوری که $m \geq n$ و $\text{rank}(A) = n$. با فرض $b \in \mathbb{R}^m$ همگرایی روش ریچاردسون برای حل دستگاه $A^T A x = A^T b$ را بررسی کنید. پارامتر بهینه را بدست آورید.

۷.۲ فرض کنید $A, B \in \mathbb{R}^{n \times n}$ و $\|I - AB\| = c < 1$.

(الف) نشان دهید هر دو ماتریس A و B نامنفرد هستند.

(ب) دنباله $\{B_k\}$ را به صورت

$$B_k = B_{k-1} + B_{k-1}(I - AB_{k-1}) = B_{k-1}(2I - B_{k-1}A), \quad k = 1, 2, \dots,$$

تعریف می‌کنیم. نشان دهید

$$\lim_{k \rightarrow \infty} B_k = A^{-1}, \quad \|A^{-1} - B_k\| \leq \|B_0\| \frac{c^{2^k}}{1 - c}.$$

(ج) فرض کنید ماتریس A قطری‌شدنی باشد، یعنی یک ماتریس نامنفرد مثل T وجود دارد به طوری که $A = T \text{diag}_n(\lambda_1, \lambda_2, \dots, \lambda_n) T^{-1}$. به علاوه، فرض کنید به ازای $i = 1, 2, \dots, n$ ، $\Re(\lambda_i) \neq 0$.

نشان دهید که روند تکراری

$$x_0 = A, \quad X_{s+1} = \frac{1}{2}(X_s + X_s^{-1}), \quad s = 0, 1, \dots,$$

به تابع علامت

$$\operatorname{sgn}(A) = T \operatorname{diag}_n(\operatorname{sgn}(\lambda_1), \operatorname{sgn}(\lambda_2), \dots, \operatorname{sgn}(\lambda_n)) T^{-1},$$

همگرا می‌شود که در آن

$$\operatorname{sgn}(\lambda_i) = \begin{cases} +1, & \Re(\lambda_i) > 0, \\ -1, & \Re(\lambda_i) < 0. \end{cases}$$

همچنین نشان دهید $X^2 = I$.

۸.۲ فرض کنید A یک ماتریس معین مثبت متقارن باشد و به علاوه، مقادیر ویژه آن به صورت $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ باشند. برای به دست آوردن جواب تقریبی دستگاه $Ax = b$ ، روش تکراری

$$x^{(k+1)} = (I - \alpha_k A)x^{(k)} + \alpha_k b,$$

را تعریف می‌کنیم که در آن $\alpha_k > 0$.

(الف) نشان دهید $e^{(k)} = q_k(A)e^{(0)}$ که در آن $e^{(k)} = x^{(k)} - x$ و $q_k(t) = \prod_{i=0}^{k-1} (1 - \alpha_i t)$.

(ب) نشان دهید

$$\|q_k(A)\|_2 \leq \max_{\lambda_1 \leq t \leq \lambda_n} |q_k(t)|.$$

(ج) در حالتی که $0 < \alpha_k = \alpha, k = 0, 1, 2, \dots$ باید در چه شرایطی صدق کند تا روش همگرا گردد و در این صورت مقدار بهینه α را به دست آورید.

۹.۲ فرض کنید مجموع درایه‌های هر سطر (یا ستون) ماتریس $A > 0$ برابر با r باشد. نشان دهید $\rho(A) = r$.

۱۰.۲ فرض کنید $A = (a_{ij}) > 0$ یک ماتریس $n \times n$ باشد. نشان دهید

$$\min_i \sum_{j=1}^n a_{ij} \leq \rho(A) \leq \max_i \sum_{j=1}^n a_{ij}.$$

۱۱.۲ دستگاه $Az = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} I_n & M \\ -M & I_n \end{pmatrix}, \quad z = \begin{pmatrix} x \\ y \end{pmatrix}, \quad b = \begin{pmatrix} p \\ q \end{pmatrix},$$

۱۳۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

ماتریس $M \in \mathbb{R}^{n \times n}$ معین مثبت متقارن و I_n ماتریس همانی $n \times n$ است. قرار دهید $A = I_n - E - F$ به‌طوری‌که

$$E = \begin{pmatrix} \circ & \circ \\ M & \circ \end{pmatrix}, \quad F = \begin{pmatrix} \circ & -M \\ \circ & \circ \end{pmatrix},$$

و I_n ماتریس همانی $(2n) \times (2n)$ است. با استفاده از این شکافت، روند تکراری SOR را نوشته و مشخص کنید که به‌ازای چه پارامتری روش SOR همگراست. پارامتر بهینه را محاسبه کنید.

۱۲.۲ فرض کنید ماتریس‌های A و M معین مثبت متقارن باشند و $N = M - A$. نشان دهید ماتریس $M^{-1}A$ قطری‌شدنی است. به‌علاوه، اگر روند تکراری

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b,$$

برای حل دستگاه $Ax = b$ همگرا باشد آنگاه مقادیر ویژه $M^{-1}A$ در بازه $(\circ, 1)$ قرار دارند.

۱۳.۲ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ متقارن و $A = M - N$ یک شکافت از A باشد که در آن ماتریس M معین مثبت متقارن و N متقارن است. نشان دهید اگر $\rho(N) < \lambda_{\min}(M)$ ، آنگاه روند تکراری

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b,$$

با هر حدس اولیه دلخواه به جواب دستگاه $Ax = b$ همگرا می‌شود.

۱۴.۲ فرض کنید $A \in \mathbb{R}^{m \times n}$ به‌طوری‌که $m < n$ و $\text{rank}(A) = n$. به‌ازای $\mu > \circ$ روند تکراری

$$\begin{cases} u^{(k+1)} = \frac{1}{\mu} \left(a + (\mu - 1)u^{(k)} + Av^{(k)} \right), \\ v^{(k+1)} = \frac{1}{\mu} \left(b + (\mu - 1)v^{(k)} - A^T u^{(k)} \right), \end{cases} \quad (64-2)$$

را برای حل دستگاه

$$\begin{cases} u - Av = a, \\ A^T u + v = b, \end{cases} \quad (65-2)$$

به‌کار می‌بریم. به‌ازای چه مقادیری از μ روند تکراری (۶۴-۲) با هر حدس اولیه $u^{(\circ)}$ و $v^{(\circ)}$ به جواب دستگاه (۶۵-۲) همگرا است؟ مقدار بهینه این پارامتر را بیابید.

نیمه همگرایی روش‌های تکراری ایستا

در این فصل، همگرایی روش‌های تکراری ایستا برای حل دستگاه معادلات خطی که ماتریس ضرایب آنها منفرد باشند را مطالعه می‌کنیم.

۱-۳ مقدمه

دستگاه معادلات خطی سازگار (دارای دست‌کم یک جواب)

$$Ax = b, \quad (1-3)$$

را در نظر بگیرید، که در آن $A \in \mathbb{C}^{n \times n}$ یک ماتریس منفرد است ($\det(A) \neq 0$), $b \in \mathbb{C}^n$ و $x \in \mathbb{C}^n$. بردار مجهول است. فرض کنید $A = M - N$ یک شکافت از A و

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b, \quad k = 0, 1, 2, \dots,$$

روند تکراری ایستای متناظر به این شکافت برای حل دستگاه (۱-۳) باشد که در آن $x^{(0)}$ یک حدس اولیه دلخواه برای جواب دستگاه است. با فرض $T = M^{-1}N$ و $f = M^{-1}b$ ، روند تکراری اخیر به صورت

$$x^{(k+1)} = Tx^{(k)} + f, \quad k = 0, 1, 2, \dots, \quad (2-3)$$

در می‌آید. در حالتی که ماتریس A نامنفرد باشد، دستگاه (۱-۳) جواب یکتای

$$x^* = A^{-1}b,$$

۱۳۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

را دارد و در فصل ۲ دیدیم که شرط لازم و کافی برای همگرایی روش تکراری (۲-۳) به x^* ، به‌ازای هر حدس اولیه دلخواه، این است که $\rho(T) < 1$.

اکنون، فرض کنید که ماتریس A منفرد باشد. با توجه به اینکه دستگاه (۱-۳) سازگار است، به‌شمار جواب دارد. در این حالت، ماتریس تکرار روند تکراری (۲-۳) را برای حل دستگاه (۱-۳) به‌صورت

$$T = M^{-1}N = M^{-1}(M - A) = I - M^{-1}A,$$

می‌نویسیم، که در آن I ماتریس همانی $n \times n$ است. بنابراین،

$$M^{-1}A = I - T.$$

با توجه به اینکه A منفرد است، ماتریس $I - T$ نیز منفرد خواهد بود. روشن است که مقادیر ویژه $I - T$ به‌صورت $1 - \lambda$ هستند، که در آن λ یک مقدار ویژه T است. با توجه به اینکه $I - T$ منفرد است، به‌ازای یک λ ، $1 - \lambda = 0$. در نتیجه، داریم

$$\rho(T) \geq 1.$$

بنابراین، نمی‌توان نتیجه گرفت که دنباله تولیدشده توسط (۲-۳) به‌ازای هر حدس اولیه دلخواه همگراست. اما در ادامه، خواهیم دید که اگر ماتریس A منفرد باشد، آنگاه تحت شرایطی دنباله تولیدشده توسط (۲-۳) به‌ازای هر حدس اولیه دلخواه، به جوابی از دستگاه همگرا خواهد شد که به این حدس اولیه وابسته است. این شرایط را شرایط نیمه‌همگرایی می‌گویند و دنباله (۲-۳) را نیمه‌همگرا می‌نامند. برای ارائه شرایط نیمه‌همگرایی، مقدماتی نیاز است که در ادامه معرفی می‌کنیم.

۲-۳ شاخص یک ماتریس

این بخش را با تعریف شاخص یک ماتریس شروع می‌کنیم.

تعریف ۱.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$. **شاخص** ماتریس A را با $\text{index}(A)$ نشان می‌دهند و بنا به تعریف کوچک‌ترین عدد صحیح نامنفی k است، به‌طوری‌که

$$\text{rank}(A^k) = \text{rank}(A^{k+1}).$$

اگر λ یک مقدار ویژه A باشد، آنگاه شاخص λ ، شاخص ماتریس $\lambda I - A$ تعریف می‌شود.

مثال ۱.۳ فرض کنید

$$A = \begin{pmatrix} ۲ & ۰ & ۰ \\ ۰ & ۱ & ۱ \\ ۰ & -۱ & -۱ \end{pmatrix}.$$

واضح است که $\text{rank}(A) = ۲$ از طرفی،

$$A^۲ = \begin{pmatrix} ۴ & ۰ & ۰ \\ ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ \end{pmatrix}, \quad A^۳ = \begin{pmatrix} ۸ & ۰ & ۰ \\ ۰ & ۰ & ۰ \\ ۰ & ۰ & ۰ \end{pmatrix}.$$

بنابراین،

$$\text{rank}(A^۲) = \text{rank}(A^۳) = ۱.$$

لذا، $\text{index}(A) = ۲$.

تعریف ۲.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $\sigma(A) = \{\lambda_1, \lambda_2, \dots, \lambda_s\}$

(الف) **تکرار جبری** مقدار ویژه λ_i ، تعداد تکرارهای λ_i به عنوان ریشه ای از چندجمله ای مشخصه ماتریس A است و با $\text{alg mult}(\lambda_i)$ نشان داده می شود. به عبارت دیگر، $\text{alg mult}(\lambda_i) = a_i$ اگر و تنها اگر

$$p(\lambda) = (\lambda - \lambda_1)^{a_1} (\lambda - \lambda_2)^{a_2} \dots (\lambda - \lambda_s)^{a_s}.$$

(ب) اگر $\text{alg mult}(\lambda_i) = ۱$ ، آنگاه λ_i یک مقدار ویژه ساده نامیده می شود.

(ب) **تکرار هندسی** مقدار ویژه λ_i با $\text{geo mult}(\lambda_i)$ نشان داده می شود و بنا به تعریف تعداد بردارهای ویژه مستقل خطی نظیر λ_i است. به عبارت دیگر،

$$\text{geo mult}(\lambda_i) = \dim \mathcal{N}(A - \lambda_i I),$$

(د) اگر $\text{alg mult}(\lambda_i) = \text{geo mult}(\lambda_i)$ ، آنگاه λ_i یک مقدار ویژه نیمه ساده نامیده می شود.

مثال ۲.۳ چندجمله ای مشخصه ماتریس

$$A = \begin{pmatrix} ۰ & ۱ \\ ۰ & ۰ \end{pmatrix},$$

۱۴۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

برابر است با λ^2 ، بنابراین، $p(\lambda) = \lambda^2$ ریشه مضاعف p است. لذا، $\text{alg mult}_A(\circ) = 2$. به سادگی می‌توان دید که

$$\mathcal{N}(A - \circ I) = \text{span} \left\{ \begin{pmatrix} 1 \\ \circ \end{pmatrix} \right\}.$$

بنابراین،

$$\dim \mathcal{N}(A - \circ I) = \dim \mathcal{N}(A) = 1.$$

پس $\text{geo mult}(\circ) = 1$. به عبارت دیگر، گرچه مقدار ویژه $\lambda = \circ$ دو بار تکرار شده است، اما فقط یک بردار ویژه مستقل خطی متناظر با این مقدار ویژه وجود دارد. قضیه زیر در آنالیز نیمه همگرایی کمک زیادی به ما می‌کند. اثبات این قضیه ساده است؛ هرچند برای اثبات آن خواننده می‌تواند به [۷۷] مراجعه کند.

قضیه ۱.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$ و شکل کانونی جردن ماتریس A به صورت

$$P^{-1}AP = J = \begin{pmatrix} J(\lambda_1) & \circ & \cdots & \circ \\ \circ & J(\lambda_2) & \cdots & \circ \\ \vdots & \vdots & \ddots & \vdots \\ \circ & \circ & \cdots & J(\lambda_s) \end{pmatrix},$$

باشد، که در آن

$$J(\lambda_j) = \begin{pmatrix} J_1(\lambda_j) & \circ & \cdots & \circ \\ \circ & J_2(\lambda_j) & \cdots & \circ \\ \vdots & \vdots & \ddots & \vdots \\ \circ & \circ & \cdots & J_{t_j}(\lambda_j) \end{pmatrix}, \quad J_r(\lambda_j) = \begin{pmatrix} \lambda_j & 1 & & \\ & \lambda_j & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_j \end{pmatrix}.$$

در این صورت، داریم

(الف) اگر بزرگ‌ترین بلوک جردن در $J(\lambda_j)$ ، $k_j \times k_j$ باشد، آنگاه $\text{index}(\lambda_j) = k_j$.

(ب) تعداد بلوک‌های جردن وابسته به مقدار ویژه λ_j ، برابر با تکرار هندسی مقدار ویژه λ_j است.

(ج) شاخص مقدار ویژه λ_j برابر با ۱ است اگر و تنها اگر $\text{alg mult}_A(\lambda_j) = \text{geo mult}_A(\lambda_j)$.

به عبارت دیگر، $\text{index}(\lambda_j) = 1$ اگر و تنها اگر λ_j یک مقدار ویژه نیمه ساده باشد.

(د) چندجمله‌ای مینیمال A یک چندجمله‌ای به صورت

$$p(\lambda) = (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_s)^{m_s},$$

است که در آن m_k برابر است با بُعد بزرگ‌ترین بلوک جردن $J_i(\lambda_k)$ در $J(\lambda_k)$.

مثال ۳.۳ اگر شکل کانونی جردن ماتریس A به صورت

$$J = \begin{pmatrix} 4 & 1 & & & & & \\ & 4 & 1 & & & & \\ & & 4 & & & & \\ & & & 4 & 1 & & \\ & & & & 4 & & \\ & & & & & 3 & 1 \\ & & & & & & 3 \\ & & & & & & & 2 \\ & & & & & & & & 2 \end{pmatrix},$$

باشد، آنگاه $\sigma(A) = \{2, 3, 4\}$. چندجمله‌ای مشخصه ماتریس A به صورت

$$p(\lambda) = (\lambda - 4)^5 (\lambda - 3)^2 (\lambda - 2)^2,$$

است. بنا به تعریف، داریم $\text{alg mult}(4) = 5$. با توجه به اینکه دو بلوک جردن متناظر به مقدار ویژه ۴ وجود دارد، تکرار هندسی این مقدار ویژه برابر با ۲ است، یعنی $\text{geo mult}(4) = 2$. بُعد بزرگ‌ترین بلوک جردن متناظر با مقدار ویژه ۴ برابر با ۳ است. بنابراین، $\text{index}(4) = 3$. به همین ترتیب، برای دو مقدار ویژه دیگر، داریم

$$\text{alg mult}(3) = 2, \quad \text{geo mult}(3) = 1, \quad \text{index}(3) = 2,$$

$$\text{alg mult}(2) = 2, \quad \text{geo mult}(2) = 2, \quad \text{index}(2) = 1.$$

همچنین، با توجه به اینکه

$$\text{alg mult}(2) = \text{geo mult}(2) = 2,$$

۱۴۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

مقدار ویژه ۲ نیمه‌ساده است. سرانجام، چندجمله‌ای مینیمال ماتریس A (و البته ماتریس J) به صورت

$$p(\lambda) = (\lambda - 2)(\lambda - 3)^2(\lambda - 4)^3,$$

است.

۳-۳ زیرفضاهای پایا و خواص آن

این بخش را با تعریف زیر آغاز می‌کنیم.

تعریف ۳.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$. زیرفضای M از $\mathbb{C}^n$ را A -پایا می‌گویند هرگاه $AM \subseteq M$ ، یعنی

$$AM = \{Ax : x \in M\} \subseteq M.$$

مثال ۴.۳ فرض کنید

$$A = \begin{pmatrix} 4 & 4 & 4 \\ -2 & -2 & -5 \\ 1 & 2 & 5 \end{pmatrix}, \quad x_1 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.$$

اگر $M = \text{span}\{x_1, x_2\}$ ، آنگاه داریم $AM \subseteq M$ ، زیرا، داریم

$$Ax_1 = 2x_1 \in M,$$

$$Ax_2 = x_1 + 2x_2 \in M.$$

اکنون، اگر $x \in M$ ، آنگاه دو مقدار $\alpha_1, \alpha_2 \in \mathbb{C}$ وجود دارند به طوری که $x = \alpha_1 x_1 + \alpha_2 x_2$ ، در نتیجه،

$$\begin{aligned} Ax &= A(\alpha_1 x_1 + \alpha_2 x_2) \\ &= \alpha_1 Ax_1 + \alpha_2 Ax_2 \\ &= (2\alpha_1 + \alpha_2)x_1 + (2\alpha_2)x_2 \in M. \end{aligned}$$

این رابطه نشان می‌دهد $AM \subseteq M$. از این رو، زیرفضای M ، A -پایا است.

فرض کنید زیرفضای M از $\mathbb{C}^n$ ، A -پایا باشد. به علاوه، فرض کنید $B_M = \{x_1, x_2, \dots, x_r\}$ یک پایه برای M باشد. این پایه را به صورت $B = \{x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_n\}$ به عنوان یک پایه برای $\mathbb{C}^n$ توسعه می‌دهیم. قرار می‌دهیم $S = [x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_n]$ داریم

$$AS = [Ax_1, Ax_2, \dots, Ax_r, Ax_{r+1}, \dots, Ax_n]. \quad (3-3)$$

اما Ax_j ها را می‌توان به صورت

$$Ax_j = \sum_{i=1}^r \alpha_{ij} x_i, \quad j = 1, 2, \dots, r, \quad (4-3)$$

$$Ax_j = \sum_{i=1}^r \alpha_{ij} x_i + \sum_{i=r+1}^n \alpha_{ij} x_i, \quad j = r+1, \dots, n, \quad (5-3)$$

نمایش داد که α_{ij} ها در $\mathbb{C}$ هستند. با جایگذاری (4-3) و (5-3) در (3-3)، خواهیم داشت

$$AS = S \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1r} & \alpha_{1,r+1} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{r1} & \cdots & \alpha_{rr} & \alpha_{r,r+1} & \cdots & \alpha_{rn} \\ \circ & \cdots & \circ & \alpha_{r+1,r+1} & \cdots & \alpha_{r+1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \circ & \cdots & \circ & \alpha_{n,r+1} & \cdots & \alpha_{n,n} \end{pmatrix} \\ = S \begin{pmatrix} A_1 & A_2 \\ \circ & A_2 \end{pmatrix}.$$

اکنون، فرض کنید $\mathbb{C}^n$ برابر با جمع مستقیم دو زیرفضای M و N باشد، یعنی

$$\mathbb{C}^n = M \oplus N,$$

به طوری که M و N ، A -پایا می‌باشند. به علاوه، فرض کنید مجموعه‌های $B_M = \{x_1, x_2, \dots, x_r\}$ و $B_N = \{x_{r+1}, x_{r+2}, \dots, x_n\}$ به ترتیب پایه‌ای برای M و N باشند. در این صورت، $B_M \cup B_N$ یک پایه برای $\mathbb{C}^n$ است (تمرین ۱.۳ را ببیند) و روابط (4-3) و (5-3) به صورت

$$Ax_j = \sum_{i=1}^r \alpha_{ij} x_i, \quad j = 1, 2, \dots, r,$$

$$Ax_j = \sum_{i=r+1}^n \alpha_{ij} x_i, \quad j = r+1, \dots, n,$$

خواهند بود. در نتیجه، داریم

$$AS = S \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1r} & \circ & \cdots & \circ \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{r1} & \cdots & \alpha_{rr} & \circ & \cdots & \circ \\ \circ & \cdots & \circ & \alpha_{r+1,r+1} & \cdots & \alpha_{r+1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \circ & \cdots & \circ & \alpha_{n,r+1} & \cdots & \alpha_{n,n} \end{pmatrix} \\ = S \begin{pmatrix} A_1 & \circ \\ \circ & A_2 \end{pmatrix}.$$

به‌طور معادل، خواهیم داشت

$$S^{-1}AS = \begin{pmatrix} A_1 & \circ \\ \circ & A_2 \end{pmatrix}.$$

با جمع‌بندی این مطالب قضیه زیر را ارائه می‌کنیم.

قضیه ۲.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $\mathbb{C}^n = M \oplus N$ که در آن M و N دو زیرفضای A -پایا از $\mathbb{C}^n$ می‌باشند. در این صورت، یک ماتریس نامنفرد مثل S وجود دارد به‌طوری‌که

$$S^{-1}AS = \begin{pmatrix} A_1 & \circ \\ \circ & A_2 \end{pmatrix}.$$

۳-۴ تجزیه هسته-پوچ توان

همان‌گونه که می‌دانیم برای هر ماتریس $A \in \mathbb{C}^{n \times n}$ ، داریم ([۷۷] را ببینید)

$$\dim(\mathcal{R}(A)) + \dim(\mathcal{N}(A)) = n.$$

با توجه به این رابطه ممکن است تصور شود که $\mathbb{C}^n = \mathcal{R}(A) \oplus \mathcal{N}(A)$ ، اما این رابطه همیشه برقرار

نیست. برای مثال، اگر

$$A = \begin{pmatrix} \circ & 1 \\ \circ & \circ \end{pmatrix},$$

آنگاه

$$\begin{pmatrix} 1 \\ \circ \end{pmatrix} \in \mathcal{R}(A) \cap \mathcal{N}(A).$$

از این رو، $\mathbb{C}^n$ را نمی‌توان به صورت جمع مستقیم $\mathcal{R}(A)$ و $\mathcal{N}(A)$ نوشت.

در هر صورت توجه کنید که اگر ماتریس A نامنفرد باشد، آنگاه خواهیم داشت $\mathbb{C}^n = \mathcal{R}(A) \oplus \mathcal{N}(A)$. در ادامه، خواهیم دید که حتی اگر ماتریس A منفرد باشد، به ازای دستکم یک عدد صحیح نامنفی k خواهیم داشت $\mathbb{C}^n = \mathcal{R}(A^k) \oplus \mathcal{N}(A^k)$.

قضیه ۳.۳ برای هر ماتریس $A \in \mathbb{C}^{n \times n}$ یک عدد صحیح نامنفی k وجود دارد به طوری که

$$\mathbb{C}^n = \mathcal{R}(A^k) \oplus \mathcal{N}(A^k). \quad (۳-۶)$$

به علاوه، کوچک‌ترین k که در رابطه (۳-۶) صدق می‌کند برابر با شاخص A است. برهان. به سادگی می‌توان دید که

$$\mathcal{N}(A^0) \subseteq \mathcal{N}(A^1) \subseteq \cdots \subseteq \mathcal{N}(A^k) \subseteq \mathcal{N}(A^{k+1}) \subseteq \cdots, \quad (۳-۷)$$

$$\mathcal{R}(A^0) \supseteq \mathcal{R}(A^1) \supseteq \cdots \supseteq \mathcal{R}(A^k) \supseteq \mathcal{R}(A^{k+1}) \supseteq \cdots. \quad (۳-۸)$$

ابتدا نشان می‌دهیم که در یک نقطه از هریک از زنجیرهای (۳-۷) و (۳-۸) تساوی وجود دارد. نخست زنجیر مربوط به فضای پوچ را در نظر می‌گیریم و به برهان خلف فرض می‌کنیم تمام زیرمجموعه‌ها اکید باشند. در این صورت، خواهیم داشت

$$\dim(\mathcal{N}(A^0)) < \dim(\mathcal{N}(A^1)) < \dim(\mathcal{N}(A^2)) < \cdots < \dim(\mathcal{N}(A^k)) < \cdots.$$

این رابطه نشان می‌دهد که $n < \dim(\mathcal{N}(A^{n+1}))$ ، که غیرممکن است. چنین نتیجه‌ای برای فضای برد نیز برقرار است. نشان می‌دهیم که اگر $\mathcal{N}(A^k) = \mathcal{N}(A^{k+1})$ ، آنگاه

$$\mathcal{N}(A^0) \subset \mathcal{N}(A^1) \subset \cdots \subset \mathcal{N}(A^k) = \mathcal{N}(A^{k+1}) = \mathcal{N}(A^{k+2}) = \cdots,$$

$$\mathcal{R}(A^0) \supset \mathcal{R}(A^1) \supset \cdots \supset \mathcal{R}(A^k) = \mathcal{R}(A^{k+1}) = \mathcal{R}(A^{k+2}) = \cdots.$$

فرض کنید k کوچک‌ترین عدد صحیحی باشد که $\mathcal{R}(A^k) = \mathcal{R}(A^{k+1})$. در این صورت، به ازای هر $i \geq 1$ داریم

$$\mathcal{R}(A^{i+k}) = \mathcal{R}(A^i A^k) = A^i \mathcal{R}(A^k) = A^i \mathcal{R}(A^{k+1}) = \mathcal{R}(A^{i+k+1}).$$

با توجه به اینکه

$$\dim(\mathcal{N}(A^p)) = n - \dim(\mathcal{R}(A^p)),$$

می‌بینیم که تساوی در زنجیر مربوط به فضای پوچ، دقیقاً در مکانی ظاهر می‌شود که در حلقهٔ مربوط به فضای برد ظاهر شده است.

اکنون، نشان می‌دهیم اگر k کوچک‌ترین عدد صحیحی باشد که نخستین تساوی به ازای آن در هر دو زنجیر رخ دهد، آنگاه خواهیم داشت $\mathcal{N}(A^k) \cap \mathcal{R}(A^k) = \{0\}$. فرض کنید $x \in \mathcal{N}(A^k) \cap \mathcal{R}(A^k)$. از اینکه $x \in \mathcal{R}(A^k)$ ، می‌بینیم که به ازای یک $y \in \mathbb{C}^n$ داریم $x = A^k y$ و از اینکه $x \in \mathcal{N}(A^k)$ داریم $A^k x = 0$. در نتیجه، $A^{2k} y = A^k x = 0$ و از این رو، $y \in \mathcal{N}(A^{2k}) = \mathcal{N}(A^k)$. بنابراین، $x = A^k y = 0$. سرانجام، داریم

$$\begin{aligned} \dim(\mathcal{R}(A^k) + \mathcal{N}(A^k)) &= \dim(\mathcal{R}(A^k)) + \dim(\mathcal{N}(A^k)) \\ &\quad - \dim(\mathcal{R}(A^k) \cap \mathcal{N}(A^k)) \\ &= \dim(\mathcal{R}(A^k)) + \dim(\mathcal{N}(A^k)) = n. \end{aligned}$$

از اینکه $\mathcal{R}(A^k) + \mathcal{N}(A^k) \subseteq \mathbb{C}^n$ و $\dim(\mathcal{R}(A^k) + \mathcal{N}(A^k)) = n$ بنابر قضیهٔ ۲.۱، داریم

$$\mathcal{R}(A^k) + \mathcal{N}(A^k) = \mathbb{C}^n,$$

□

و بدین ترتیب اثبات قضیه کامل می‌شود.

نتیجهٔ ۱.۳ در صورتی که شاخص ماتریس A معلوم نباشد، می‌توان از

$$\mathbb{C}^n = \mathcal{R}(A^n) \oplus \mathcal{N}(A^n),$$

استفاده کرد.

نتیجه ۲.۳ فرض کنید شاخص ماتریس $A \in \mathbb{C}^{n \times n}$ برابر با k باشد. در این صورت،

$$A(\mathcal{N}(A^k)) \subseteq \mathcal{N}(A^k), \quad A(\mathcal{R}(A^k)) \subseteq \mathcal{R}(A^k).$$

تعریف ۴.۳ ماتریس $A \in \mathbb{C}^{n \times n}$ را پوچ توان می‌گویند هرگاه به ازای یک عدد صحیح نامنفی k داشته باشیم $A^k = 0$. کوچک‌ترین عدد k که به ازای آن $A^k = 0$ ، را شاخص پوچ توانی A می‌گویند.

قضیه ۴.۳ (تجزیه هسته-پوچ توان) فرض کنید $A \in \mathbb{C}^{n \times n}$ و $\text{index}(A) = k$. به علاوه، فرض کنید $\text{rank}(A^k) = r$. در این صورت، یک ماتریس نامنفرد مثل S وجود دارد به طوری که

$$S^{-1}AS = \begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix}, \quad (9-3)$$

که در آن C یک ماتریس $r \times r$ نامنفرد است و N یک ماتریس پوچ توان با شاخص پوچ توانی k است. برهان. با استناد به قضیه ۳.۳، داریم

$$\mathbb{C}^n = \mathcal{R}(A^k) \oplus \mathcal{N}(A^k).$$

از طرفی،

$$\dim(\mathcal{R}(A^k)) = \text{rank}(A^k) = r.$$

فرض کنید $x_1, \dots, x_r$ یک پایه برای $\mathcal{R}(A^k)$ و $x_{r+1}, \dots, x_n$ یک پایه برای $\mathcal{N}(A^k)$ باشند. قرار می‌دهیم $S = (x_1, x_2, \dots, x_r, x_{r+1}, \dots, x_n) \in \mathbb{C}^{n \times n}$. با استفاده از نتیجه ۲.۳، هر دو زیرفضای $\mathcal{R}(A^k)$ و $\mathcal{N}(A^k)$ -پایا هستند. بنابراین، با توجه به قضیه ۲.۳، ماتریس $S^{-1}AS$ به صورت

$$S^{-1}AS = \begin{pmatrix} C & 0 \\ 0 & N \end{pmatrix}, \quad (10-3)$$

خواهد بود. اگر طرفین رابطه (۱۰-۳) را به توان k برسانیم، خواهیم داشت

$$(S^{-1}AS)^k = S^{-1}A^kS = \begin{pmatrix} C^k & 0 \\ 0 & N^k \end{pmatrix}. \quad (11-3)$$

از طرفی، با قرار دادن $S_1 = (x_1, x_2, \dots, x_r)$ و $S_2 = (x_{r+1}, \dots, x_n)$ داریم $S = (S_1, S_2)$.

با نمایش ماتریس S^{-1} به صورت

$$S^{-1} = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

خواهیم داشت

$$\begin{aligned} (S^{-1}AS)^k &= S^{-1}A^kS = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} A^k(S_1 \ S_2) \\ &= \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} (A^kS_1 \ 0) = \begin{pmatrix} T_1 A^k S_1 & 0 \\ T_2 A^k S_1 & 0 \end{pmatrix}. \end{aligned}$$

با مقایسه روابط (۱۱-۳) و (۱۲-۳) می‌بینیم که $N^k = 0$ و این رابطه نشان می‌دهد که ماتریس N پوچ‌توان است. ماتریس C^k یک ماتریس $r \times r$ است و

$$\text{rank}(A^k) = r = \text{rank}(S^{-1}A^kS) = \text{rank} \begin{pmatrix} C^k & 0 \\ 0 & 0 \end{pmatrix} = \text{rank}(C^k).$$

بنابراین، C^k یک ماتریس با رتبه کامل است. لذا، C نامنفرد است. سرانجام نشان می‌دهیم که شاخص پوچ‌توانی N برابر با k است. به برهان خلف، اگر شاخص پوچ‌توانی N کوچک‌تر از k باشد، آنگاه $N^{k-1} = 0$ و در نتیجه، داریم

$$\begin{aligned} \text{rank}(A^{k-1}) &= \text{rank}(S^{-1}A^{k-1}S) = \text{rank} \begin{pmatrix} C^{k-1} & 0 \\ 0 & N^{k-1} \end{pmatrix} \\ &= \text{rank} \begin{pmatrix} C^{k-1} & 0 \\ 0 & 0 \end{pmatrix} = \text{rank}(C^{k-1}) = r = \text{rank}(A^k). \end{aligned}$$

این رابطه نشان می‌دهد که شاخص ماتریس A برابر با $k-1$ است که یک تناقض است. $\square$

مثال ۵.۳ فرض کنید

$$A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

داریم

$$A^{\mathfrak{r}} = \begin{pmatrix} \circ & \circ & \circ \\ ۱\circ & ۷ & ۲ \\ ۱\circ & ۷ & ۲ \end{pmatrix}, \quad A^{\mathfrak{r}} = \begin{pmatrix} \circ & \circ & \circ \\ ۳\circ & ۲۱ & ۶ \\ ۳\circ & ۲۱ & ۶ \end{pmatrix}.$$

لذا،

$$\text{rank}(A^{\mathfrak{r}}) = \text{rank}(A^{\mathfrak{r}}) = ۱ = r.$$

به راحتی می توان دید که

$$\mathcal{R}(A^{\mathfrak{r}}) = \text{span} \left\{ \begin{pmatrix} \circ \\ ۱ \\ ۱ \end{pmatrix} \right\}, \quad \mathcal{N}(A^{\mathfrak{r}}) = \text{span} \left\{ \begin{pmatrix} ۱ \\ \circ \\ -۵ \end{pmatrix}, \begin{pmatrix} \circ \\ ۲ \\ -۷ \end{pmatrix} \right\}.$$

قرار می دهیم

$$S = \begin{pmatrix} \circ & ۱ & \circ \\ ۱ & \circ & ۲ \\ ۱ & -۵ & -۷ \end{pmatrix}.$$

در این صورت،

$$S^{-۱}AS = \begin{pmatrix} ۳ & \circ & \circ \\ \circ & ۶ & ۹ \\ \circ & -۴ & -۶ \end{pmatrix} = \begin{pmatrix} C & \circ \\ \circ & N \end{pmatrix},$$

که در آن

$$C = (۳), \quad N = \begin{pmatrix} ۶ & ۹ \\ -۴ & -۶ \end{pmatrix}.$$

در اینجا ماتریس C یک ماتریس با رتبه $r = ۱$ (رتبه کامل) است و ماتریس N پوچ توان با شاخص پوچ توانی ۲ است.

۵-۳ معکوس درازین

در این بخش معکوس درازین یک ماتریس که نقش مهمی در آنالیز نیمه همگرایی دارد، معرفی می شود.

۱۵۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

تعریف ۵.۳ فرض کنید شاخص ماتریس $A \in \mathbb{C}^{n \times n}$ برابر با k باشد. معکوس درازین ماتریس A را با A^D نشان داده و ماتریسی است که در سه شرط

$$(a) \quad A^D A A^D = A^D, \quad (b) \quad A A^D = A^D A, \quad (c) \quad A^{k+1} A^D = A^k,$$

صدق کند.

نتیجه ۳.۳ با توجه به شرط (b) می‌بینیم که شرط (c) معادل است با $A^k A^D A = A^k$ یا $A^D A^{k+1} = A^k$.

قضیه ۵.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $\text{index}(A) = k$. یک و فقط یک معکوس درازین وجود دارد که در شرایط (a)، (b) و (c) صدق می‌کند. **برهان.** فرض کنید

$$A = S \begin{pmatrix} C & \circ \\ \circ & N \end{pmatrix} S^{-1},$$

تجزیه هسته-پوچ توان ماتریس A باشد. تعریف می‌کنیم

$$A^D = S \begin{pmatrix} C^{-1} & \circ \\ \circ & \circ \end{pmatrix} S^{-1}.$$

به سادگی می‌توان نشان داد که A^D در هر سه شرط (a)، (b) و (c) صدق می‌کند. برای اثبات یکتایی جواب، فرض کنید X و Y دو معکوس درازین ماتریس A باشند. در این صورت، با توجه به نتیجه ۳.۳ خواهیم داشت

$$A^k = A^k X A = A^k Y A.$$

با ضرب طرفین این رابطه از راست در X می‌بینیم که

$$A^k X = A^k Y A X. \quad (۱۲-۳)$$

بنابراین، $A^{k-1} A X = A^{k-1} A Y A X$ یا $A^{k-1} A Y A X = A^{k-1} X A$. با ضرب طرفین این رابطه از سمت راست در X خواهیم داشت

$$A^{k-1} X = A^{k-1} A Y A X^2,$$

یا

$$\begin{aligned} A^{k-1}X &= A^{k-1}AYAX^2 \\ &= A^{k-1}AYXAX \\ &= A^{k-1}AYX \\ &= A^{k-1}YAX. \end{aligned}$$

این رابطه نشان می‌دهد که توانسته‌ایم یک واحد از توان A در رابطه (۱۲-۳) کم کنیم. با ادامه این روند خواهیم داشت $X = YAX$. به‌طور مشابه، می‌توان نشان داد که $Y = XAY$. اکنون، با استفاده از (c) داریم $A^k(AY - I) = 0$. بنابراین، $XA^k(AY - I) = 0$. از این‌رو،

$$A^k(XAY - X) = 0, \quad (13-3)$$

و با استفاده از (a)، داریم $A^k(XAY - XAX) = 0$ یا $A^kXA(Y - X) = 0$. با ضرب طرفین این رابطه از سمت چپ در X ، داریم

$$\begin{aligned} 0 &= XA^kXA(Y - X) \\ &= XA^{k-1}AXA(Y - X) \\ &= A^{k-1}XAXA(Y - X) \\ &= A^{k-1}XA(Y - X) \\ &= A^{k-1}(XAY - XAX) \\ &= A^{k-1}(XAY - X). \end{aligned}$$

این رابطه نشان می‌دهد که یک واحد از توان A در رابطه (۱۳-۳) کاسته‌ایم. با ادامه این روند، خواهیم داشت $XAY = X$. این رابطه به همراه $Y = XAY$ ، نتیجه می‌دهد $X = Y$. $\square$

مثال ۶.۳ معکوس درازین ماتریس مثال ۵.۳ به‌صورت زیر است

$$A^D = S^{-1} \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} S.$$

نتیجه ۴.۳ اگر ماتریس A نامنفرد باشد، آنگاه $A^D = A^{-1}$.

۳-۶ نیمه‌همگرایی

فرض کنید ماتریس A منفرد و x^* یک جواب دلخواه برای دستگاه (۳-۱) باشد. روشن است که

$$x^* = Tx^* + f, \quad (۳-۱۴)$$

با تفریق طرفین این رابطه و رابطه (۳-۲)، داریم

$$e^{(k)} = Te^{(k-1)}, \quad k = 1, 2, \dots,$$

که در آن $e^{(j)} = x^{(j)} - x^*$ به استقرا می‌توان دید که

$$e^{(k)} = T^k e^{(0)}. \quad (۳-۱۵)$$

همان‌گونه که در بخش ۳-۱ اشاره شد اگر A نامنفرد باشد، شرط لازم و کافی برای همگرایی دنباله (۳-۲) به جواب یکتای $x^* = A^{-1}b$ این است که $\lim_{k \rightarrow \infty} T^k = 0$ ، و این خود معادل است با اینکه $\rho(T) < 1$. در اینجا توجه کنید که اگر $\rho(T) < 1$ ، آنگاه خواهیم داشت

$$\lim_{k \rightarrow \infty} e^{(k)} = 0,$$

و حدس اولیه $x^{(0)}$ تأثیری در حد دنباله ندارد.

در ادامه، فرض می‌کنیم که ماتریس A منفرد باشد. همان‌گونه که دیدیم $\rho(T) \geq 1$. بنابراین، در این حالت، $\lim_{k \rightarrow \infty} T^k$ ناصفر است یا موجود نیست. فرض کنید که این حد موجود و برابر با ماتریس C باشد، یعنی

$$\lim_{k \rightarrow \infty} T^k = C \neq 0.$$

در این صورت، از (۳-۱۵) داریم

$$\lim_{k \rightarrow \infty} x^{(k)} = x^* + C(x^{(0)} - x^*).$$

این رابطه نشان می‌دهد که در این حالت حد دنباله به حدس اولیه وابسته است. در ادامه، وجود حد برای دنباله $\{T^k\}$ را بررسی می‌کنیم. برای روشن شدن مطلب، نخست مثال زیر را ارائه می‌کنیم.

مثال ۷.۳ فرض کنید

$$T = \begin{pmatrix} \frac{1}{2} & 5 \\ 0 & 1 \end{pmatrix}.$$

با کمک استقرا می‌توان دید که

$$T^k = \begin{pmatrix} \frac{1}{2^k} & 5(2 - \frac{1}{2^{k-1}}) \\ 0 & 1 \end{pmatrix}.$$

بنابراین،

$$\lim_{k \rightarrow \infty} T^k = \begin{pmatrix} 0 & 10 \\ 0 & 1 \end{pmatrix}.$$

این مثال نشان می‌دهد با وجود اینکه $\rho(T) = 1$ ، حد دنباله $\{T^k\}$ موجود و ناصفر است.

تعریف ۶.۳ دنباله $\{T^k\}$ را نیمه همگرا می‌گویند هرگاه $\lim_{k \rightarrow \infty} T^k$ موجود (نه لزوماً برابر با صفر) باشد.

اکنون، آماده‌ایم که شرایط همگرایی را مطالعه کنیم. فرض کنید J شکل کانونی جردن ماتریس T باشد، یعنی یک ماتریس نامنفرد مثل P وجود داشته باشد به‌طوری‌که

$$P^{-1}TP = J.$$

به‌سادگی می‌توان دید که

$$T^k = PJ^kP^{-1} = P \begin{pmatrix} \ddots & & \\ & J_*^k & \\ & & \ddots \end{pmatrix} P^{-1},$$

که در آن

$$J_* = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} = \lambda I + S \in \mathbb{C}^{m \times m}, \quad \lambda \in \sigma(T),$$

به‌طوری‌که

$$S = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}.$$

به‌سادگی می‌توان دید که به‌ازای هر $j \geq m$ ، $S^j = 0$. در نتیجه، به‌ازای هر $k \geq m$ خواهیم داشت

$$\begin{aligned} J_*^k &= (\lambda I + S)^k \\ &= \sum_{j=0}^k \binom{k}{j} \lambda^{k-j} S^j = \sum_{j=0}^{m-1} \binom{k}{j} \lambda^{k-j} S^j \\ &= \begin{pmatrix} \lambda^k & \binom{k}{1} \lambda^{k-1} & \binom{k}{2} \lambda^{k-2} & \dots & \binom{k}{m-1} \lambda^{k-m+1} \\ & \lambda^k & \binom{k}{1} \lambda^{k-1} & \ddots & \vdots \\ & & \ddots & \ddots & \binom{k}{2} \lambda^{k-2} \\ & & & \lambda^k & \binom{k}{1} \lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix}. \end{aligned}$$

با توجه به ساختار ماتریس J_*^k ، واضح است که اگر $|\lambda| > 1$ ، آنگاه $\lim_{k \rightarrow \infty} J_*^k$ موجود نیست. بنابراین، برای وجود این حد، باید $|\lambda| \leq 1$. در حالتی که $|\lambda| < 1$ ، دنباله همگراست و چیزی برای بررسی باقی نمی‌ماند. بنابراین، تنها حالتی که باید بررسی شود این است که $|\lambda| = 1$. برای این کار، دو حالت را در نظر می‌گیریم:

حالت اول: فرض کنید $|\lambda| = 1$ و $\lambda \neq 1$. در این صورت، λ را می‌توان به‌صورت

$$\lambda = e^{i\theta} = \cos \theta + i \sin \theta, \quad 0 < \theta < 2\pi,$$

نوشت که در آن i واحد موهومی است. بنابراین،

$$\lambda^k = (\cos \theta + i \sin \theta)^k = \cos k\theta + i \sin k\theta.$$

در نتیجه، حد درایه های قطری J_*^k و در نتیجه، T^k موجود نیستند.

حالت دوم: فرض کنید $|\lambda| = 1$ و $\lambda \neq 1$. در این حالت، داریم

$$J_*^k = \begin{pmatrix} 1 & \binom{k}{1} & \binom{k}{2} & \cdots & \binom{k}{m-1} \\ & 1 & \binom{k}{1} & \ddots & \vdots \\ & & \ddots & \ddots & \binom{k}{2} \\ & & & 1 & \binom{k}{1} \\ & & & & 1 \end{pmatrix}.$$

اگر $m \neq 1$ ، داریم

$$\lim_{k \rightarrow \infty} \binom{k}{j} = \infty, \quad 1 \leq j \leq m-1.$$

بنابراین $\lim_{k \rightarrow \infty} J_*^k$ و در نتیجه، $\lim_{k \rightarrow \infty} T^k$ موجود است اگر و تنها اگر $m = 1$ (J_* یک ماتریس 1×1 است)، که این خود معادل است با اینکه $\lambda = 1$ یک مقدار ویژه نیمه ساده باشد. اکنون، با توجه به قسمت (ج) از قضیه ۱.۳ نتایج به دست آمده را به صورت قضیه ۶.۳ جمع بندی می کنیم.

قضیه ۶.۳ دنباله (۲-۳) نیمه همگرا است اگر و تنها اگر

$$(\text{الف}) \quad \rho(T) = 1;$$

(ب) شاخص مقدار ویژه $\lambda \in \sigma(T)$ برابر با ۱ باشد. به عبارت دیگر

$$\text{rank}(I - T) = \text{rank}(I - T)^2;$$

(ج) اگر $\lambda \in \sigma(T)$ و $|\lambda| = 1$ ، آنگاه $\lambda = 1$.

فرض کنید دنباله (۲-۳) نیمه‌همگرا باشد. در این صورت، تعریف می‌کنیم

$$v(T) = \max\{|\lambda| : \lambda \in \sigma(T), \lambda \neq 1\} < 1.$$

$v(T)$ را شعاع شبه‌طیفی ماتریس T می‌نامند. شعاع شبه‌طیفی همانند شعاع طیفی در بحث همگرایی عمل می‌کند. به عبارت دیگر، هرچه $v(T)$ کوچکتر باشد سرعت نیمه‌همگرایی بیشتر خواهد بود. دیدیم که چه وقت ماتریس T نیمه‌همگرا است. در ادامه می‌بینیم که اگر T نیمه‌همگرا باشد، آنگاه با هر حدس اولیه $x^{(0)}$ دنباله (۲-۳) به جوابی از دستگاه همگرا می‌شود که این جواب وابسته به $x^{(0)}$ است. بنابراین فرض می‌کنیم T نیمه‌همگرا باشد و $p = \text{alg mult}_A(1)$. در این صورت، p بلوک جردن به شکل $J_* = (1)$ در J وجود دارد، یعنی شکل کانونی جردن ماتریس T به صورت است

$$J = P^{-1}TP = \begin{pmatrix} I & \circ \\ \circ & K \end{pmatrix}, \quad (۱۶-۳)$$

که در آن $1 < \rho(K)$ و I یک ماتریس همانی $p \times p$ است.

قضیه ۷.۳ روند تکراری ایستای

$$x^{(k+1)} = Tx^{(k)} + f, \quad (۱۷-۳)$$

برای حل دستگاه $Ax = b$ را در نظر بگیرید. فرض کنید $\text{index}(I - T) \leq 1$. اگر به‌ازای حدس اولیه $x^{(0)}$ دنباله $\{x^{(k)}\}$ همگرا باشد، آنگاه حد دنباله به صورت

$$x^* = (I - T)^D M^{-1}b + (I - E)x^{(0)}, \quad (۱۸-۳)$$

خواهد بود که در آن $E = (I - T)(I - T)^D$ و شاخص D نشانگر معکوس درازین است. برهان. اگر $\text{index}(I - T) = 0$ ، آنگاه

$$n = \text{rank}(I - T)^0 = \text{rank}(I - T)^1,$$

که نشان می‌دهد ماتریس $I - T$ نامنفرد است. از طرفی، چون

$$\begin{aligned} I - T &= I - M^{-1}N \\ &= M^{-1}(M - N) \end{aligned}$$

$$= M^{-1}A,$$

نتیجه می گیریم که ماتریس A نامنفرد است و نیمه همگرایی به همگرایی تبدیل می شود. در این صورت، داریم $(I - T)^D = (I - T)^{-1}$ و $E = 0$. بنابراین، رابطه (۱۸-۳) به صورت زیر خواهد بود

$$x^* = (I - T)^{-1} M^{-1} b = (M^{-1} A)^{-1} M^{-1} b = A^{-1} b.$$

به عبارت دیگر، روند تکراری (۱۷-۳) به تنها جواب دستگاه $Ax = b$ همگرا می شود. اکنون، فرض می کنیم $\text{index}(I - T) = 1$. در این صورت، شاخص مقدار ویژه ۱ برای ماتریس T برابر با ۱ خواهد بود و از این رو، یک ماتریس نامنفرد مثل P وجود دارد به طوری که

$$J = P^{-1}TP = \begin{pmatrix} I & 0 \\ 0 & K \end{pmatrix}, \quad (۱۹-۳)$$

که در آن $\rho(K) < 1$ و I یک ماتریس همانی $p \times p$ است (در اینجا، تکرار جبری مقدار ویژه ۱ برابر با p است و $p \geq 1$). فرض کنید $x^{(*)}$ حد دنباله $x^{(k)}$ باشد. در این صورت،

$$x^* = Tx^* + M^{-1}b. \quad (۲۰-۳)$$

با تفريق طرفین این رابطه و رابطه (۱۷-۳)، خواهیم داشت

$$x^{(k)} - x^* = T(x^{(k-1)} - x^*) = \dots = T^k(x^{(0)} - x^*).$$

فرض کنید

$$y = P^{-1}(x^{(0)} - x^*) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

که در آن $y_1 \in \mathbb{R}^p$ و $y_2 \in \mathbb{R}^{n-p}$. از اینکه $\lim_{k \rightarrow \infty} (x^{(k)} - x^*) = 0$ ، می بینیم که

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} T^k(x^{(0)} - x^*) \\ &= \lim_{k \rightarrow \infty} P \begin{pmatrix} I & 0 \\ 0 & K^k \end{pmatrix} P^{-1}(x^{(0)} - x^*) \\ &= P \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \end{aligned}$$

$$= P \begin{pmatrix} y_1 \\ \circ \end{pmatrix}.$$

این رابطه نشان می‌دهد که $y_1 = \circ$ ازطرفی، با توجه به تعریف معکوس درازین، داریم

$$\begin{aligned} I - E &= I - (I - T)(I - T)^D \\ &= I - P \begin{pmatrix} \circ & \circ \\ \circ & I - K \end{pmatrix} P^{-1} P \begin{pmatrix} \circ & \circ \\ \circ & (I - K)^{-1} \end{pmatrix} P^{-1} \\ &= P \begin{pmatrix} I & \circ \\ \circ & \circ \end{pmatrix} P^{-1}. \end{aligned}$$

بنابراین،

$$\begin{aligned} (I - E)(x^{(\circ)} - x^*) &= P \begin{pmatrix} I & \circ \\ \circ & \circ \end{pmatrix} y \\ &= P \begin{pmatrix} I & \circ \\ \circ & \circ \end{pmatrix} \begin{pmatrix} \circ \\ y_2 \end{pmatrix} \\ &= \circ. \end{aligned}$$

اکنون، با محاسبه x^* خواهیم داشت

$$x^* = Ex^* + (I - E)x^{(\circ)}. \quad (21-3)$$

ازطرفی، با استفاده از (20-3)، داریم $(I - T)x^* = M^{-1}b$ و درنتیجه،

$$(I - T)^D(I - T)x^* = (I - T)^D M^{-1}b,$$

که معادل است با

$$Ex^* = (I - T)^D M^{-1}b. \quad (22-3)$$

توجه کنید که

$$E = (I - T)^D(I - T) = (I - T)(I - T)^D.$$

با جایگذاری Ex^* از رابطه (22-3) در رابطه (21-3)، نتیجه لازم به‌دست می‌آید. $\square$

مثال ۸.۳ دستگاه $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 12 \end{pmatrix}.$$

ماتریس A منفرد و دستگاه $Ax = b$ سازگار است؛ زیرا بردار $x^* = (1, 1)^T$ یکی از جواب های دستگاه است. با فرض

$$M = \begin{pmatrix} 1 & 0 \\ 2 & 8 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & -2 \\ -2 & 0 \end{pmatrix},$$

داریم $A = M - N$. همچنین،

$$T = M^{-1}N = \begin{pmatrix} 0 & -2 \\ -\frac{1}{4} & \frac{1}{4} \end{pmatrix}, \quad f = M^{-1}b = \begin{pmatrix} 3 \\ 3 \\ \frac{3}{4} \end{pmatrix}.$$

مقادیر ویژه ماتریس T عبارتند از $\lambda_1 = -\frac{1}{4}$ و $\lambda_2 = 1$. بنابراین، شرایط (الف) و (ج) از نیمه همگرایی برقرارند. از طرفی، داریم

$$I - T = \begin{pmatrix} 1 & 2 \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix},$$

$$(I - T)^2 = \begin{pmatrix} \frac{3}{4} & \frac{3}{4} \\ \frac{3}{8} & \frac{3}{4} \end{pmatrix}.$$

به وضوح، داریم

$$\text{rank}(I - T) = \text{rank}(I - T)^2 = 1.$$

بنابراین، $\text{index}(1) = 1$. در نتیجه، روند تکراری (۳-۱۷) به ازای هر حدس اولیه به یک جواب دستگاه $Ax = b$ همگرا می شود. برای مثال، اگر حدس اولیه برای جواب دستگاه را بردار صفر و شرط توقف برای روند تکراری (۳-۱۷) را به صورت $\|b - Ax^{(k)}\| < 10^{-7}$ در نظر بگیریم، در ۲۷ تکرار جواب تقریبی زیر به دست می آید

$$x^{(27)} = \begin{pmatrix} 2.000000001490116 \\ -0.500000000372529 \end{pmatrix}.$$

۱۶۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

توجه کنید که $x^* = (2, 0/5)^T$ یکی از جواب‌های دستگاه است. در صورتی که حدس اولیه را بردار $(-1, -1)^T$ در نظر بگیریم، در ۲۸ تکرار جواب تقریبی

$$x^{28} = \begin{pmatrix} 2/99999998509884 \\ -0/000000000372529 \end{pmatrix},$$

به دست می‌آید. توجه کنید که این جواب، یک تقریب از جواب واقعی $x^* = (3, 0)^T$ است.

مثال ۹.۳ دستگاه $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} -1 & 1 & -2 & 1 \\ -1 & 4 & 1 & -1 \\ 4 & -4 & 3 & 1 \\ 2 & 1 & 2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 3 \\ 4 \\ 6 \end{pmatrix}.$$

ماتریس A منفرد و دستگاه $Ax = b$ سازگار است؛ زیرا بردار $x^* = (1, 1, 1, 1)^T$ یکی از جواب‌های دستگاه است. با فرض

$$M = \begin{pmatrix} -1 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 \\ 4 & -4 & 3 & 0 \\ 2 & 1 & 2 & 1 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

داریم $A = M - N$ و M نامنفرد است. بنابراین، $A = M - N$ یک شکافت از A است. داریم

$$T = M^{-1}N = \begin{pmatrix} 0 & 1 & -2 & 1 \\ 0 & \frac{1}{4} & -\frac{3}{4} & \frac{1}{2} \\ 0 & -1 & \frac{5}{4} & -1 \\ 0 & -\frac{1}{4} & \frac{17}{12} & -\frac{1}{2} \end{pmatrix}, \quad f = M^{-1}b = \begin{pmatrix} 1 \\ 1 \\ 4 \\ \frac{3}{4} \end{pmatrix}.$$

طیف ماتریس T عبارت است از

$$\sigma(T) = \{0, -\frac{1}{4}, 1, \frac{2}{3}\}.$$

بنابراین، شرایط (الف) و (ج) از قضیه ۶.۳ برقرار هستند. با کمی محاسبه می توان دید

$$\text{rank}(I - T) = \text{rank}(I - T)^2 = 3.$$

بنابراین، $\text{index}(1) = 1$ و در نتیجه، دنباله $(17-3)$ به ازای هر حدس اولیه دلخواه به یک جواب دستگاه همگرا است. برای مثال، اگر حدس اولیه را بردار صفر و شرط توقف را $\|b - Ax^{(k)}\|_2 < 10^{-7}$ بگیریم آنگاه روند تکراری در ۴۳ تکرار به جواب

$$x^{(43)} = \begin{pmatrix} -2.99999983922179 \\ 0.00000002923240 \\ 3.99999989281452 \\ 3.9999998358213 \end{pmatrix},$$

همگرا می شود. جواب به دست آمده یک تقریب از جواب واقعی $x^* = (-3, 0, 4, 4)^T$ است.

تمرینات

۱.۳ فرض کنید M و N دو زیرفضای $\mathbb{C}^n$ باشند به طوری که $C^n = M \oplus N$. ثابت کنید اگر B_M و B_N به ترتیب یک پایه برای M و N باشند، آنگاه $B_M \cup B_N$ یک پایه برای $\mathbb{C}^n$ است.

۲.۳ فرض کنید $P \in \mathbb{R}^{4 \times 4}$ نامنفرد باشد و

$$A = P \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} P^{-1}.$$

(الف) شاخص ماتریس A را محاسبه کنید.

(ب) معکوس درازین ماتریس A را محاسبه کنید.

۳.۳ فرض کنید

$$A = \begin{pmatrix} 2 & 1 \\ 8 & 4 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 12 \end{pmatrix}.$$

۱۶۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنگ

نشان دهید روش تکراری ژاکوبی برای حل دستگاه $Ax = b$ نیمه‌همگرا نیست، اما روش گاوس-سایدل نیمه‌همگراست.

۴.۳ فرض کنید ماتریس P نامنفرد باشد. نشان دهید $A = PBP^{-1}$ اگر و تنها اگر $A^D = PB^DP^{-1}$.

۵.۳ فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ منفرد باشد. نشان دهید $(A^D)^D = A$ اگر و تنها اگر $\text{index}(A) = 1$.

۶.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$ و

$$A = TJT^{-1} = T \begin{pmatrix} J_1 & \circ \\ \circ & J_\circ \end{pmatrix} T^{-1},$$

شکل جردن ماتریس A باشد که در آن $J_\circ$ و J_1 به‌ترتیب متناظر با بلوک‌های متناظر با مقادیر صفر و ناصفر A هستند. نشان دهید

$$A^D = T \begin{pmatrix} J_1^{-1} & \circ \\ \circ & \circ \end{pmatrix} T^{-1}.$$

۷.۳ فرض کنید $A \in \mathbb{R}^{n \times n}$ و $\text{index}(A) = k$. نشان دهید اگر دستگاه $Ax = b$ در $\mathcal{R}(A^k)$ یک جواب داشته باشد، آنگاه این جواب، یکتاست و به‌صورت $x = A^D b$ خواهد بود.

۸.۳ فرض کنید $A \in \mathbb{C}^{n \times n}$. دراین‌صورت،

$$A^D = \lim_{\alpha \rightarrow 0} \left(A^{k+1} + \alpha^k I \right)^{-1} A^k,$$

که در آن I ماتریس همانی $n \times n$ است.

۹.۳ تمرین ۶.۲ را وقتی‌که ماتریس A با رتبهٔ ناقص است، یعنی $\text{rank}(A) < n$ ، حل کنید.

۱۰.۳ فرض کنید $A \in \mathbb{R}^{n \times n} \neq \circ$ یک ماتریس نیمه‌معین مثبت متقارن باشد و $\omega > 0$. روش تکراری

$$x_{k+1} = (I - \omega A)x_k + \omega b,$$

برای حل دستگاه سازگار $Ax = b$ را در نظر بگیرید. نخست، نشان دهید که به ازای هر ω یکی از مقادیر ویژه ماتریس $I - \omega A$ برابر با یک است. سپس، نشان دهید که به ازای هر

$$0 < \omega < \frac{2}{\lambda_{\max}},$$

این روش تکراری با هر حدس اولیه به یک جواب دستگاه همگرا است، که در آن $\lambda_{\max}$ بزرگترین مقدار ویژه ماتریس A است.

۱۱.۳ قسمت های (الف) و (ج) از قضیه ۱۰.۳ را ثابت کنید.

۱۲.۳ فرض کنید

$$A = \begin{pmatrix} 1 & -2 & 2 \\ 2 & -4 & 1 \\ 0 & 0 & 5 \end{pmatrix}, \quad M = \begin{pmatrix} 1 & 0 & 0 \\ 2 & -4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

و $N = M - A$. نیمه همگرایی روند تکراری

$$x^{(k+1)} = M^{-1} N x^{(k)} + M^{-1} b,$$

را برای حل دستگاه $Ax = b$ بررسی کنید که در آن $b = (1, 0, 0)^T$.

۱۳.۳ فرض کنید

$$A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & -1 \\ -1 & 0 & 1 \end{pmatrix} = M - N,$$

که در آن

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

نشان دهید که روند تکراری

$$x^{(k+1)} = M^{-1} N x^{(k)} + M^{-1} b,$$

۱۶۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و مُتَنَک

برای حل دستگاه $Ax = b$ نیمه‌همگرا نیست، که در آن $b = (0, 1, 0)^T$. نشان دهید به‌ازای
روند تکراری $0 < \omega < 1$

$$x^{(k+1)} = ((1 - \omega)I + \omega T) x^{(k)} + \omega M^{-1}b,$$

برای حل دستگاه $Ax = b$ نیمه‌همگرا است که در آن $T = M^{-1}N$.

عملگرهای تصویر

در این فصل عملگرهای تصویر را معرفی کرده و خواص آنها را بررسی می‌کنیم. این‌گونه عملگرها کاربرد زیادی در نظریه تقریب دارند، مثل حل عددی معادلات دیفرانسیل با استفاده از روش‌های عناصر متناهی و حل دستگاه معادلات خطی با استفاده از روش‌های تصویری. اگرچه بنا به نیاز ما عملگرهای تصویر روی $\mathbb{C}^n$ را مطالعه می‌کنیم، این‌گونه عملگرها روی فضاهای مختلفی تعریف می‌شوند.

۱-۴ معرفی عملگر تصویر و خواص آن

تعریف ۱.۴ عملگر خطی P از $\mathbb{C}^n$ به $\mathbb{C}^n$ یک عملگر تصویر نامیده می‌شود هرگاه خودتوان باشد، یعنی $P^2 = P$.

مثال ۱.۴ فرض کنید $x = (x_1, x_2, x_3)^T \in \mathbb{C}^3$ و $Px = (x_1, x_2, 0)^T$. در این صورت، P یک عملگر تصویر روی $\mathbb{C}^3$ است. توجه کنید که P هر نقطه در $\mathbb{C}^3$ را روی صفحه xoy به‌طور عمودی تصویر می‌کند (مفهوم تصویر عمودی و غیرعمودی در حالت کلی را در ادامه بیان خواهیم کرد).

قضیه ۱.۴ فرض کنید P یک عملگر تصویر باشد. در این صورت،

$$x \in \mathcal{R}(P) \iff Px = x.$$

به‌علاوه، $I - P$ نیز یک عملگر تصویر است و $\mathcal{N}(P) = \mathcal{R}(I - P)$.

برهان. برای اثبات قسمت نخست قضیه، داریم

$$x \in \mathcal{R}(P) \implies \exists y \in \mathbb{C}^n : x = Py \implies Px = P^2 y = Py = x \in \mathcal{R}(P).$$

برعکس، اگر $Px = x$ ، آنگاه روشن است که $x \in \mathcal{R}(P)$. برای اثبات قسمت دوم قضیه، ابتدا توجه کنید که $I - P$ خطی است. فرض کنید $x \in \mathbb{C}^n$ در نتیجه،

$$\begin{aligned} (I - P)^2 x &= (I - P)(x - Px) = (I - P)x - Px + P^2 x \\ &= (I - P)x - Px + Px = (I - P)x, \end{aligned}$$

که نشان می‌دهد $(I - P)^2 = (I - P)$. برای اثبات قسمت آخر، داریم

$$x \in \mathcal{N}(P) \implies Px = 0 \implies x - Px = x \implies (I - P)x = x \in \mathcal{R}(I - P).$$

اکنون، فرض کنید $x \in \mathcal{R}(I - P)$. در این صورت، به‌ازای یک $y \in \mathbb{C}^n$ داریم $x = (I - P)y$. در نتیجه،

$$Px = P(I - P)y = Py - P^2 y = Py - Py = 0 \implies x \in \mathcal{N}(P),$$

و بدین ترتیب اثبات قضیه کامل می‌شود. □ □

قضیه ۲.۴. اگر P یک عملگر تصویر باشد، آنگاه $\mathbb{C}^n = \mathcal{N}(P) \oplus \mathcal{R}(P)$.

برهان. ابتدا نشان می‌دهیم که $\mathbb{C}^n = \mathcal{R}(P) \oplus \mathcal{R}(I - P)$. به‌ازای هر $x \in \mathbb{C}^n$ ،

$$x = Px + (I - P)x.$$

داریم $Px \in \mathcal{R}(P)$ و $(I - P)x \in \mathcal{R}(I - P)$. فرض کنید $x \in \mathcal{R}(P) \cap \mathcal{R}(I - P)$. در این صورت، به‌ازای یک y و z ، داریم $x = Py$ و $x = (I - P)z$. بنابراین، $Py = (I - P)z$. در نتیجه،

$$P^2 y = P(I - P)z = Pz - P^2 z = Pz - Pz = 0 \implies Py = 0.$$

این رابطه نشان می‌دهد که $x = Py = 0$. بنابراین، $\mathbb{C}^n = \mathcal{R}(P) \oplus \mathcal{R}(I - P)$. اینک، با توجه به قضیه ۱.۴ نتیجه لازم به‌دست می‌آید. □

اکنون، فرض کنید M و S دو زیرفضای برداری $\mathbb{C}^n$ باشند، به طوری که $\mathbb{C}^n = M \oplus S$. عملگر P را روی $\mathbb{C}^n$ به صورت زیر تعریف می‌کنیم. فرض کنید $x \in \mathbb{C}^n$ و

$$x = x_1 + x_2, \quad x_1 \in M, x_2 \in S.$$

در این صورت، تعریف می‌کنیم $Px = x_1$. اولاً، P خطی است (چرا؟). ثانیاً، طبق تعریف داریم $Px = x_1$. روشن است که $Px_1 = x_1$. در نتیجه، $Px_1 = x_1 = Px$. این رابطه نشان می‌دهد که $P^2 = P$. یعنی P یک عملگر تصویر است. توجه کنید که

$$\forall x \in \mathbb{C}^n : x = x_1 + x_2 = Px + x_2 \implies x_2 = x - Px = (I - P)x.$$

این رابطه نشان می‌دهد که $M = \mathcal{R}(P)$ و $S = \mathcal{N}(I - P)$. با تعریفی که برای P ارائه نموده‌ایم، داریم

$$\begin{cases} Px \in M, \\ (I - P)x \in S, \end{cases} \quad (1-4)$$

و گوییم P یک عملگر تصویر روی M و به موازات S است.

مثال ۲.۴ فرض کنید

$$M = \left\{ (x, y)^T \in \mathbb{R}^2 : y = \frac{1}{2}x \right\}, \quad S = \left\{ (x, y)^T \in \mathbb{R}^2 : y = 2x \right\}.$$

داریم

$$\begin{pmatrix} x \\ y \end{pmatrix} = \overbrace{\begin{pmatrix} \frac{1}{2}(2x - y) \\ \frac{1}{2}(2x - y) \end{pmatrix}}^{\in M} + \overbrace{\begin{pmatrix} \frac{1}{2}(2y - x) \\ \frac{1}{2}(2y - x) \end{pmatrix}}^{\in S}.$$

از طرفی، $M \cap S = \{0\}$. بنابراین، $\mathbb{R}^2 = M \oplus S$. اکنون، تعریف می‌کنیم

$$\begin{cases} P : \mathbb{R}^2 \longrightarrow M, \\ P \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(2x - y) \\ \frac{1}{2}(2x - y) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \end{cases}$$

در این صورت، P یک عملگر تصویر است.

اگر $\dim(M) = m$ ، آنگاه $\dim(S) = n - m$. قرار می‌دهیم $L = S^\perp$. در این صورت، $S = L^\perp$. بنابراین، رابطه (۱-۴) معادل با

$$\begin{cases} Px \in M, \\ (I - P)x \in L^\perp, \end{cases}$$

خواهد بود که خود معادل با

$$\begin{cases} Px \in M, \\ (I - P)x \perp L, \end{cases}$$

است. اگر قرار دهیم $u = Px$ ، آنگاه عملگر P بردار x را به بردار u می‌نگارد به طوری که

$$\begin{cases} u \in M, \\ x - u \perp L. \end{cases} \quad (۲-۴)$$

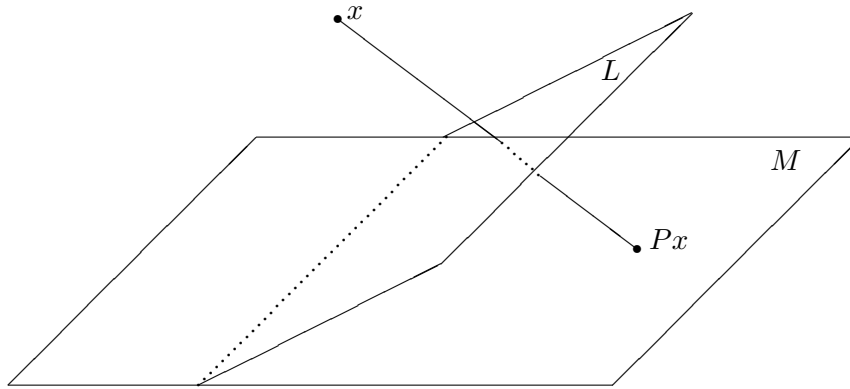
با توجه به این رابطه، می‌گوییم P یک عملگر تصویر روی M و عمود بر L است (شکل ۱-۴ را ببینید). از اینکه $u \in M$ ، برای محاسبه u ، m ضریب نامعلوم داریم. توجه کنید که اگر بخواهیم u را به صورت یک ترکیب خطی از بردارهای پایه M بنویسیم، تعداد ضرایب نامعلوم m خواهد بود. از طرفی، تعداد بردارهای پایه L نیز برابر با m است. بنابراین $x - u \perp L$ ایجاب می‌کند که $x - u$ بر m بردار پایه‌ای L عمود باشد که m معادله را بدست می‌دهد و با استفاده از آن می‌توان ضرایب نامعلوم را مشخص کرد. اما ببینیم که آیا برای هر دو زیرفضای M و L با بعد m می‌توان یک عملگر تصویر روی M و عمود بر L نوشت؟ لم بعدی این موضوع را روشن می‌کند.

لم ۱.۴ برای هر دو زیرفضای M و L با بعد m ، گزاره‌های زیر معادلند.

(الف) هیچ عضو ناصفری از M بر L عمود نیست، به عبارت دیگر $M \cap L^\perp = \{0\}$.

(ب) به ازای هر $x \in \mathbb{C}^n$ ، یک بردار یکتا مثل u وجود دارد به طوری که در شرایط $u \in M$ و $x - u \perp L$ صدق کند.

برهان. روشن است که گزاره (ب) معادل با $M \cap L^\perp = \{0\}$ است.



شکل ۱-۴ نمایش یک عملگر تصویر روی M و عمود بر L

اکنون، فرض کنید (الف) برقرار باشد. از اینکه $\dim(L^\perp) = n - m$ داریم

$$\begin{aligned}\dim(M + L^\perp) &= \dim(M) + \dim(L^\perp) - \dim(M \cap L^\perp) \\ &= m + (n - m) + 0 = n.\end{aligned}$$

حال از اینکه $M + L^\perp = \mathbb{C}^n$ یک زیرفضای $\mathbb{C}^n$ است، با استفاده از قضیه ۲.۱ نتیجه می‌گیریم که $\mathbb{C}^n = M \oplus L^\perp$ و این یعنی به‌ازای هر $x \in \mathbb{C}^n$ ، یک بردار یکتا مثل $u \in M$ و یک بردار یکتا مثل $v \in L^\perp$ وجود دارند به‌طوری‌که $x = u + v$. بنابراین، $v = x - u \in L^\perp$ که معادل است با $x - u \perp L$. $\square$

نتیجه ۱.۴ اگر $\dim(M) = \dim(L) = m$ به‌طوری‌که $M \cap L^\perp = \{0\}$ ، آنگاه یک عملگر تصویر مثل P روی M و عمود بر L می‌توان تعریف کرد به‌طوری‌که $\mathcal{R}(P) = M$ و $\mathcal{N}(P) = L^\perp$.

مثال ۳.۴ عملگر تصویر

$$P : \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$Px = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} x,$$

را در نظر بگیرید. به‌ازای هر $x = (x_1, x_2)^T \in \mathbb{R}^2$ داریم

۱۷۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنک

$$P \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ \circ \end{pmatrix}, \quad (I - P) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_2 \end{pmatrix}.$$

بنابراین،

$$\mathcal{R}(P) = \left\{ \begin{pmatrix} x_1 + x_2 \\ \circ \end{pmatrix} : x_1, x_2 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} x_1 \\ \circ \end{pmatrix} : x_1 \in \mathbb{R} \right\},$$

و

$$\mathcal{N}(P) = \mathcal{R}(I - P) = \left\{ \begin{pmatrix} -x_2 \\ x_2 \end{pmatrix} : x_2 \in \mathbb{R} \right\}.$$

همچنین، داریم

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \overbrace{\begin{pmatrix} -x_2 \\ x_2 \end{pmatrix}}^{\in \mathcal{N}(P)} + \overbrace{\begin{pmatrix} x_1 + x_2 \\ \circ \end{pmatrix}}^{\in \mathcal{R}(P)}.$$

شکل ۲-۴ به خوبی عملگر P و خواص آن را نمایش می‌دهد.

۲-۴ نمایش ماتریسی یک عملگر تصویر

فرض کنید $V = (v_1, v_2, \dots, v_m)$ یک ماتریس پایه برای زیرفضای $M = \mathcal{R}(P)$ و $W = (w_1, w_2, \dots, w_m)$ یک ماتریس پایه برای L باشند. این دو پایه را دومی‌زوج گویند هرگاه

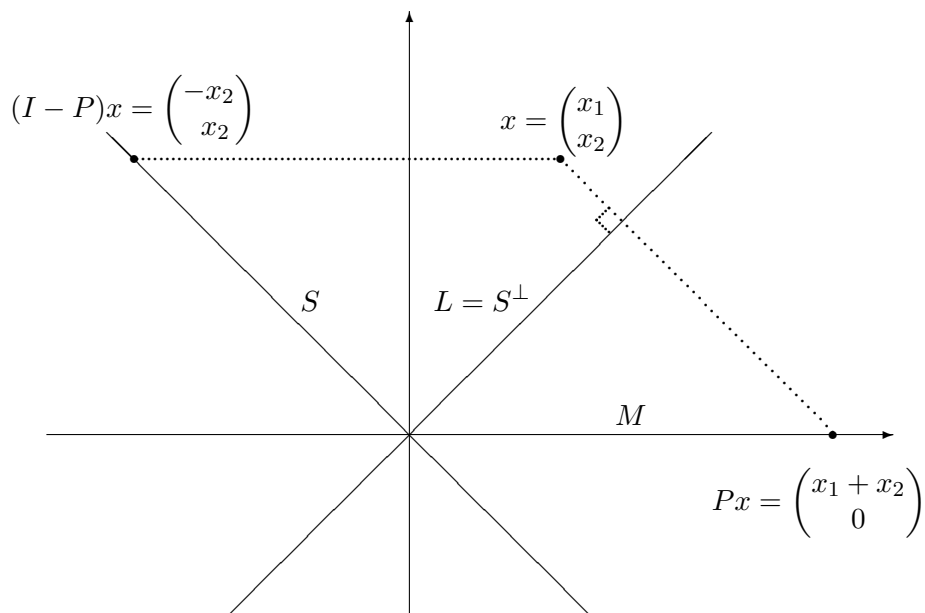
$$\langle v_i, w_j \rangle = \delta_{ij} = \begin{cases} 1, & i = j, \\ \circ, & i \neq j. \end{cases}$$

به شکل ماتریسی این رابطه معادل با $W^H V = I$ است. از اینکه $Px \in M$ ، برداری مثل $y \in \mathbb{C}^m$ وجود دارد به‌طوری‌که $Px = Vy$. از طرفی، $x - Px \perp L$ معادل با

$$\langle (x - Vy), w_j \rangle = \circ, \quad j = 1, 2, \dots, m,$$

است. به شکل ماتریسی این رابطه معادل با

$$W^H (x - Vy) = \circ,$$



شکل ۲-۴ نمایش عملگر تصویر P برای مثال ۳.۴

است که این خود معادل با $W^H V y = W^H x$ می‌باشد. اینک، اگر دو پایه V و W دومیزوج باشند، آنگاه خواهیم داشت $y = W^H x$. از این رو، خواهیم داشت

$$Px = VW^H x,$$

و بنابراین،

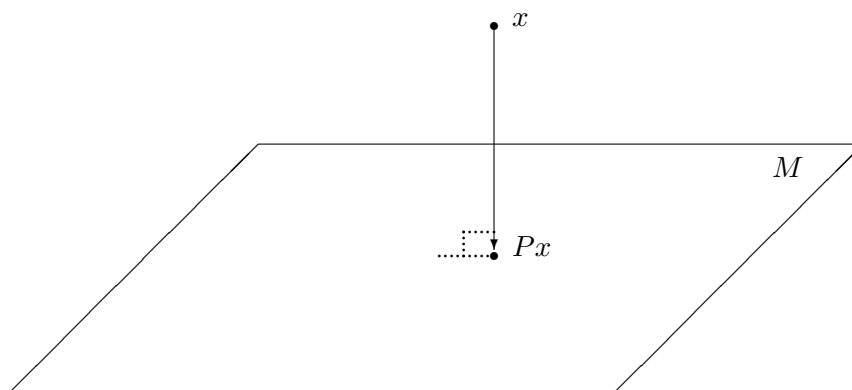
$$P = VW^H.$$

حال اگر پایه‌های V و W دومیزوج نباشند و هیچ عضوی از M بر L عمود نباشد، به عبارت دیگر $M \cap L^\perp = \{0\}$ ، آنگاه $W^H V$ نامنفرد خواهد بود (چرا؟) و خواهیم داشت

$$y = (W^H V)^{-1} W^H x,$$

و در نتیجه،

$$Px = V(W^H V)^{-1} W^H x,$$



شکل ۳-۴ نمایش یک عملگر متعامد

که نشان می‌دهد

$$P = V(W^H V)^{-1} W^H.$$

۳-۴ عملگر تصویر متعامد

اگر برای یک عملگر تصویر، دو زیرفضای M و L برابر باشند، یا به عبارت دیگر

$$\mathcal{N}(P) = \mathcal{R}(P)^\perp,$$

آنگاه عملگر تصویر را متعامد گویند. در صورتی که عملگر تصویر متعامد نباشد، مورب می‌گویند. بنابراین، اگر P یک عملگر تصویر متعامد روی M باشد، آنگاه برای هر $x \in \mathbb{C}^n$ داریم

$$Px \in M, \quad (I - P)x \perp M,$$

یا به طور معادل

$$Px \in M, \quad \langle (I - P)x, y \rangle = 0 \quad \forall y \in M.$$

عملگر معرفی شده در مثال ۱۰.۴، یک عملگر تصویر متعامد است. شکل ۳-۴ یک عملگر تصویر متعامد را نمایش می‌دهد.

تعریف ۲.۴ فرض کنید P یک عملگر خطی از $\mathbb{C}^n$ به $\mathbb{C}^n$ باشد. عملگر الحاقی P را با P^H نشان می‌دهیم و عملگری است که در رابطه

$$\langle P^H x, y \rangle = \langle x, Py \rangle, \quad \forall x, y \in \mathbb{C}^n, \quad (۳-۴)$$

صدق کند. عملگر P را هرمیتی می‌گویند هرگاه $P^H = P$. ابتدا توجه کنید که اگر P یک عملگر تصویر باشد، آنگاه P^H نیز یک عملگر تصویر است، زیرا به‌ازای هر x و y داریم

$$\langle (P^H)^\dagger x, y \rangle = \langle P^H x, Py \rangle = \langle x, P^\dagger y \rangle = \langle x, Py \rangle = \langle P^H x, y \rangle.$$

لم ۲.۴ برای هر عملگر خطی P ، داریم

$$\mathcal{N}(P^H) = \mathcal{R}(P)^\perp, \quad (۴-۴)$$

$$\mathcal{N}(P) = \mathcal{R}(P^H)^\perp. \quad (۵-۴)$$

برهان. رابطه (۴-۴) را ثابت می‌کنیم. فرض کنید $x \in \mathcal{N}(P^H)$ در این صورت، داریم $P^H x = 0$. حال با توجه به رابطه (۳-۴)، به‌ازای هر y داریم $\langle x, Py \rangle = 0$. این رابطه نشان می‌دهد که $x \in \mathcal{R}(P)^\perp$.

اکنون، فرض کنید $x \in \mathcal{R}(P)^\perp$ در این صورت، به‌ازای هر y داریم $\langle x, Py \rangle = 0$. از این رو، با استفاده از رابطه (۳-۴)، داریم $\langle P^H x, y \rangle = 0$. با توجه به اینکه y برداری دلخواه است، نتیجه می‌گیریم که $P^H x = 0$ و این یعنی $x \in \mathcal{N}(P^H)$.

اثبات رابطه (۵-۴) شبیه به اثبات رابطه (۴-۴) است و به‌عنوان تمرین به خواننده واگذار می‌کنیم. $\square$

قضیه ۳.۴ فرض کنید P یک عملگر تصویر روی $\mathbb{C}^n$ باشد. P متعامد است اگر و تنها اگر هرمیتی باشد. **برهان.** ابتدا فرض کنید P یک عملگر تصویر متعامد باشد. با توجه به تعریف یک عملگر تصویر متعامد، داریم $\mathcal{N}(P) = \mathcal{R}(P)^\perp$. از این رو، با استفاده از رابطه‌های ۴-۴ و ۵-۴ به‌ترتیب داریم $\mathcal{N}(P) = \mathcal{N}(P^H)$ و $\mathcal{R}(P) = \mathcal{R}(P^H)$. یک عضو دلخواه $x \in \mathbb{C}^n$ را در نظر بگیرید. در این صورت، وجود دارد $x_1 \in \mathcal{R}(P)$ و $x_2 \in \mathcal{N}(P)$ به‌طوری‌که $x = x_1 + x_2$. با توجه به اینکه $x_1 \in \mathcal{R}(P) = \mathcal{R}(P^H)$ ، نتیجه می‌گیریم که $x_1 = Px_1 = P^H x_1$. همچنین از اینکه

$x_2 \in \mathcal{N}(P) = \mathcal{N}(P^H)$ می‌بینیم که $Px_2 = P^Hx_2 = 0$. اکنون، داریم

$$\begin{aligned} Px &= P(x_1 + x_2) = Px_1 = Px_1 + 0 = P^Hx_1 + P^Hx_2 \\ &= P^H(x_1 + x_2) = P^Hx, \end{aligned}$$

که نشان می‌دهد $P = P^H$.

برعکس، فرض کنید $P = P^H$. در این صورت، با استفاده از رابطه (۴-۴) خواهیم داشت $\mathcal{N}(P) = \mathcal{R}(P)^\perp$ و این یعنی عملگر P متعامد است. $\square$

لم ۳.۴ هر عملگر تصویر متعامد، یکتاست.

برهان. فرض کنید P و Q دو عملگر تصویر متعامد روی M باشند. کفایت نشان دهیم به ازای هر $y \in \mathbb{C}^n$ ، $(P - Q)y = 0$. برای این کار، داریم

$$\begin{aligned} \|(P - Q)y\|^2 &= y^H(P - Q)^H(P - Q)y \\ &= y^H(P - Q)^2y \\ &= y^H(P - PQ - QP + Q)y \\ &= y^H(I - P)Qy + y^H(I - Q)Py. \end{aligned}$$

از اینکه Q یک عملگر تصویر روی M است، داریم $Qy \in M$. بنابراین، بنا به لم (۱.۴)، خواهیم داشت $PQy = Qy$ ، لذا،

$$(I - P)Qy = Qy - PQy = Qy - Qy = 0.$$

به همین ترتیب داریم، $(I - Q)Py = 0$. از این رو، $\|(P - Q)y\|^2 = 0$ ، که نتیجه لازم بدست می‌آید. $\square$

برای هر ماتریس یکانی $V \in \mathbb{C}^{n \times m}$ که ستون‌های آن یک پایه یکامتعامد برای زیرفضای $M = \mathcal{R}(P)$ تشکیل دهند، عملگر تصویر متعامد P متناظر را می‌توان به صورت

$$P = VV^H,$$

نمایش داد. توجه کنید که $P^H = P$ و $P^2 = P$. همچنین برای محاسبه Px ، داریم

$$VV^Hx \in M, \quad \text{and} \quad (I - VV^H)x \in M^\perp.$$

توجه دیگر این است که نمایش عملگر P به صورت VV^H یکتا نیست، زیرا ماتریس پایه V برای M یکتا نیست. در هر صورت، اگر $P = V_1 V_1^H$ و $P = V_2 V_2^H$ دو نمایش مختلف برای P باشند، آنگاه $V_1 V_1^H = V_2 V_2^H$.

مثال ۴.۴ یک عملگر تصویر متعامد روی زیرفضای

$$S = \left\{ (x, y, z)^T \in \mathbb{R}^3 : 2x + y - z = 0 \right\},$$

را بیابید.

حل. سه نقطه (دلخواه) $P = (-2, 1, -3)^T$ ، $Q = (1, -1, 1)^T$ ، و $R = (2, -1, 3)^T$ در S (روی صفحه $2x + y - z = 0$) را در نظر می‌گیریم. دو بردار v_1 و v_2 را به صورت

$$v_1 := \overrightarrow{QP} = \overrightarrow{OP} - \overrightarrow{OQ} = \begin{pmatrix} -3 \\ 2 \\ -4 \end{pmatrix},$$

$$v_2 := \overrightarrow{RP} = \overrightarrow{OP} - \overrightarrow{OR} = \begin{pmatrix} -1 \\ 0 \\ -2 \end{pmatrix},$$

محاسبه می‌کنیم. پرواضح است که این دو بردار روی صفحه $2x + y - z = 0$ قرار دارند. با توجه به اینکه دو بردار v_1 و v_2 مستقل خطی هستند، نتیجه می‌شود که مجموعه $\{v_1, v_2\}$ یک پایه برای S است. توجه کنید که $\dim(S) = 2$. ازاین‌رو، ماتریس پایه برای S به صورت زیر خواهد بود

$$V = (v_1 \ v_2) = \begin{pmatrix} -3 & -1 \\ 2 & 0 \\ -4 & -2 \end{pmatrix}.$$

لذا، عملگر تصویر متعامد روی S به صورت زیر محاسبه می‌شود

$$P = V(V^T V)^{-1} V^T = \frac{1}{6} \begin{pmatrix} 2 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 5 \end{pmatrix}.$$

برای مثال، تصویر متعامد نقطه $x = (3, 10, 1)^T$ روی S به صورت زیر خواهد بود

$$Px = \frac{1}{7} \begin{pmatrix} -4 \\ 15 \\ 7 \end{pmatrix}.$$

۴-۴ خواص عملگرهای تصویر متعامد

فرض کنید P یک عملگر تصویر متعامد روی $\mathbb{C}^n$ باشد. در این صورت، برای هر $x \in \mathbb{C}^n$ داریم $x = Px + (I - P)x$ به طوری که $Px \in \mathcal{R}(P)$ و $(I - P)x \in \mathcal{R}(P)^\perp$. بنابراین، Px و $(I - P)x$ بر هم عمودند. از این رو، داریم

$$\|x\|_2^2 = \|Px\|_2^2 + \|(I - P)x\|_2^2.$$

به عنوان یک نتیجه، به ازای هر $x \in \mathbb{C}^n$ داریم

$$\|Px\|_2 \leq \|x\|_2.$$

این رابطه نشان می‌دهد که

$$\|P\|_2 = \sup_{x \neq 0} \frac{\|Px\|_2}{\|x\|_2} \leq 1.$$

این نامساوی به ازای هر $x \in \mathcal{R}(P)$ ، به صورت تساوی در می‌آید. بنابراین، $\|P\|_2 = 1$.

قضیه ۴.۴ فرض کنید P یک عملگر تصویر متعامد روی زیرفضای M باشد. در این صورت، به ازای هر $x \in \mathbb{C}^n$ داریم

$$\min_{y \in M} \|x - y\|_2 = \|x - Px\|_2.$$

برهان. داریم $x - Px \perp M$ و به ازای هر $y \in M$ داریم $Px - y \in M$. بنابراین

$$\begin{aligned} \|x - y\|_2^2 &= \|x - Px + Px - y\|_2^2 \\ &= \|x - Px\|_2^2 + \|Px - y\|_2^2 \\ &\geq \|x - Px\|_2^2. \end{aligned}$$

این نامساوی به ازای $y = Px$ به تساوی تبدیل می‌شود. بنابراین، اثبات این قضیه کامل می‌شود. $\square$

نتیجه ۲.۴ فرض کنید M یک زیرفضای $\mathbb{C}^n$ و بردار $x \in \mathbb{C}^n$ داده شده باشد. در این صورت،

$$\min_{y \in M} \|x - y\|_2 = \|x - y^*\|_2,$$

اگر و تنها اگر شرایط

$$\begin{cases} y^* \in M, \\ x - y^* \perp M, \end{cases}$$

برقرار باشند.

تمرینات

۱.۴ در مثال ۲.۴، به طور مستقیم نشان دهید که P یک عملگر تصویر است.

۲.۴ زیرفضای برداری M از $\mathbb{R}^4$ را در نظر بگیرید که توسط بردارهای

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ -1 \end{pmatrix},$$

تولید می‌شود، یعنی $M = \text{span}\{v_1, v_2\}$.

(الف) ماتریسی که عملگر تصویر متعامد P روی M را نمایش می‌دهد، بیابید.

(ب) فضای پوچ عملگر P را بیابید.

(ج) برد عملگر P را بیابید.

(د) برداری مثل $x \in M$ را بیابید به طوری که کمترین فاصله را (تحت نرم-۲) تا بردار $c = (1, 1, 1, 1)^T$ داشته باشد.

۳.۴ مثال ۳.۴ را برای عملگر تصویر

$$P: \mathbb{R}^2 \rightarrow \mathbb{R}^2$$

$$Px = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x,$$

تکرار کنید.

۴.۴ رابطه (۴-۵) را ثابت کنید.

۵.۴ فرض کنید P_1 و P_2 دو عملگر تصویر از $\mathbb{C}^n$ به $\mathbb{C}^n$ باشند. نشان دهید

$$\mathcal{R}(P_1) = \mathcal{R}(P_2) \iff P_1 P_2 = P_2 P_1 = P_2.$$

۶.۴ فرض کنید $P_1, P_2 : \mathbb{C}^n \rightarrow \mathbb{C}^n$ دو عملگر تصویر باشند به‌طوری‌که $P_1 P_2 = P_2 P_1$. نشان دهید $P = P_1 P_2$ یک عملگر تصویر است، $\mathcal{N}(P_1 P_2) = \mathcal{N}(P_1) + \mathcal{N}(P_2)$ و $\mathcal{R}(P_1 P_2) = \mathcal{R}(P_1) \cap \mathcal{R}(P_2)$.

۷.۴ نشان دهید که هر ماتریس $A \in \mathbb{R}^{n \times n}$ را می‌توان به‌صورت $A = H + S$ نوشت که در آن H یک ماتریس متقارن و S یک ماتریس متقارن کج است. قرار می‌دهیم

$$\mathcal{M} = \{A \in \mathbb{R}^{n \times n} : A^T = A\}, \quad \mathcal{S} = \{A \in \mathbb{R}^{n \times n} : A^T = -A\}.$$

نشان دهید $\mathbb{R}^{n \times n} = \mathcal{M} \oplus \mathcal{S}$. یک عملگر تصویر مثل P از $\mathbb{R}^{n \times n}$ روی $\mathcal{M}$ بنویسید؟ فضای بوج چیست؟ فرض کنید $A = H + S$ به‌طوری‌که $H \in \mathcal{M}$ و $S \in \mathcal{S}$. نشان دهید $\langle H, S \rangle_F = 0$ ، که در آن $\langle H, S \rangle_F = \text{trace}(H^T S)$.

۸.۴ فرض کنید $A \in \mathbb{R}^{m \times n}$. نشان دهید $P_{\mathcal{N}(A)} = I - A^+ A$ و $P_{\mathcal{R}(A)} = A A^+$ به‌ترتیب عملگرهای تصویر روی $\mathcal{N}(A)$ و $\mathcal{R}(A)$ می‌باشند.

۹.۴ نشان دهید هر عملگر تصویر P تنها دو مقدار ویژه ۰ و ۱ دارد. بردارهای ویژه P را مشخص کنید.

۱۰.۴ یک عملگر تصویر متعامد روی هریک از زیرفضاهای

$$S_1 = \left\{ (x, y, z)^T \in \mathbb{R}^3 : x + 2y + z = 0 \right\},$$

$$S_2 = \left\{ (x, y, z)^T \in \mathbb{R}^3 : x + y - z = 0, x - y + 2z = 0 \right\},$$

را بیابید.

روش‌های تصویری برای حل دستگاه معادلات خطی

دستگاه

$$Ax = b, \quad (۱-۵)$$

را در نظر بگیرید که در آن $A \in \mathbb{R}^{n \times n}$ و $x, b \in \mathbb{R}^n$. هدف از یک روش تصویری برای حل دستگاه (۱-۵) به دست آوردن یک جواب تقریبی در یک زیرفضای m -بُعدی مثل $\mathcal{K}$ از $\mathbb{R}^n$ برای دستگاه (۱-۵) می‌باشد. زیرفضای $\mathcal{K}$ را زیرفضای جستجو می‌نامند. بنابراین، به‌طور کلی m شرط لازم است تا به چنین جواب تقریبی برسیم. یک شیوه بیان این شرایط، اعمال m شرط متعامد مستقل است؛ یعنی بردار مانده $b - Ax$ بر m بردار مستقل خطی عمود باشد. برای این کار، یک زیرفضای دیگر با بُعد m تعریف می‌شود که آن را زیرفضای شرایط نامیده و با $\mathcal{L}$ نشان می‌دهیم. این چهارچوب ساده، که همان شرایط پتروف-گالرکین می‌باشند، برای روش‌های مختلف، یکسان است. به‌طور کلی، روش‌های تصویری دو دسته‌اند: روش‌های تصویری متعامد و روش‌های تصویری مورب (مایل). در یک روش تصویری متعامد، داریم $\mathcal{K} = \mathcal{L}$ ، ولی در یک روش تصویری مورب، زیرفضاهای $\mathcal{K}$ و $\mathcal{L}$ با هم متفاوتند و حتی ممکن است هیچ ارتباطی با هم نداشته باشند. این تمایز، انواع متنوعی از الگوریتم‌ها را نتیجه می‌دهد.

۱-۵ روش‌های تصویری در حالت کلی

فرض کنید $\mathcal{L}$ و $\mathcal{K}$ دو زیرفضای m -بُعدی از $\mathbb{C}^n$ باشند. یک روش تصویری روی زیرفضای $\mathcal{K}$ و عمود بر $\mathcal{L}$ ، روندی است که یک جواب تقریبی مثل $\tilde{x}$ را برای (۱-۵) با اعمال این شرایط که اولاً $\tilde{x}$ متعلق به $\mathcal{K}$

۱۸۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

باشد و ثانیاً بردار مانده جدید بر $\mathcal{L}$ عمود باشد، پیدا می‌کند؛ یعنی

$$(2-5) \quad \tilde{x} \in \mathcal{K} \text{ را به قسمی پیدا کنید که } b - A\tilde{x} \perp \mathcal{L}.$$

شرایط (۲-۵) به شرایط پتروف-گالرکین معروف است. فرض کنید x_* یک حدس اولیه برای جواب دستگاه (۱-۵) باشد. در این صورت، داریم

$$Ax - Ax_* = b - Ax_* \implies A(x - x_*) = b - Ax_* \implies A\delta = r_*,$$

که در آن $r_* = b - Ax_*$ و $\delta = x - x_*$. بنابراین، به جای حل دستگاه (۱-۵) می‌توانیم دستگاه $A\delta = r_*$ را حل کنیم. از این رو، روش تصویری (۲-۵) برای حل این دستگاه به صورت زیر خواهد بود

$$\tilde{\delta} = \tilde{x} - x_* \in \mathcal{K} \text{ را به قسمی بیابید که } r_* - A\tilde{\delta} \perp \mathcal{L},$$

که این خود معادل است با

$$\tilde{x} \in x_* + \mathcal{K} \text{ را به قسمی بیابید که } r_* - A\tilde{\delta} \perp \mathcal{L}.$$

این رابطه نشان می‌دهد که جواب تقریبی را می‌توان با استفاده از دو رابطه

$$(3-5) \quad \tilde{x} = x_* + \tilde{\delta}, \quad \tilde{\delta} \in \mathcal{K},$$

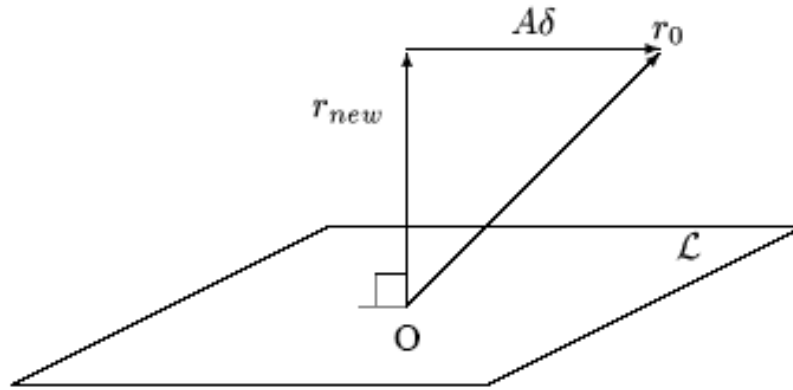
$$(4-5) \quad \langle r_* - A\tilde{\delta}, w \rangle = 0, \quad \forall w \in \mathcal{L},$$

تعیین کرد. شرط (۴-۵) که بر مانده جدید $r_{new} = r_* - A\tilde{\delta}$ تحمیل شده است در شکل ۱-۵ دیده می‌شود.

روابط (۳-۵) و (۴-۵) شکل کلی گام اصلی یک روش تصویری را مشخص می‌کند. بسیاری از روش‌های تصویری استاندارد یک دنباله از چنین گام‌هایی را به کار می‌برند. در یک گام تصویری جدید، از یک جفت جدید از زیرفضاهای $\mathcal{K}$ و $\mathcal{L}$ و یک حدس اولیه x_* که همان تقریب به دست آمده در گام تصویری قبلی است، استفاده می‌شود.

در روش‌های تصویری متعامد زیرفضاهای $\mathcal{K}$ و $\mathcal{L}$ یکی هستند. در حالت $\mathcal{L} = \mathcal{K}$ ، شرایط پتروف-گالرکین را شرایط گالرکین می‌نامند.

مثال ۱.۵ فرض کنید درایه‌های قطری ماتریس $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ ناصفر باشند و دو زیرفضای $\mathcal{K}$ و $\mathcal{L}$ به صورت $\mathcal{L} = \mathcal{K} = \text{span}\{e_i\}$ انتخاب شوند، که در آن $1 \leq i \leq n$. بنابراین، برای زیرفضاهای



شکل ۱-۵ نمایش بردار مانده جدید در یک روش تصویری

انتخاب شده، روش تصویری برداری مثل $\tilde{x} = x_0 + \tilde{\delta}$ را به دست می‌دهد به طوری که $\tilde{\delta} \in \mathcal{K}$ و

$$r_0 - A\tilde{\delta} \perp \mathcal{L} = \text{span}\{e_i\}.$$

از اینکه $\tilde{\delta} \in \mathcal{K}$ ، به ازای یک اسکالر $\alpha \in \mathbb{R}$ ، خواهیم داشت $\tilde{\delta} = \alpha e_i$. از این رو، $\tilde{x} = x_0 + \alpha e_i$ و

$$\begin{aligned} r_0 - A\tilde{\delta} = r_0 - \alpha A e_i \perp e_i &\implies e_i^T (r_0 - \alpha A e_i) = 0 \\ &\implies \alpha = \frac{e_i^T r_0}{e_i^T A e_i} = \frac{e_i^T r_0}{a_{ii}}. \end{aligned}$$

بنابراین،

$$\tilde{x} = x_0 + \frac{e_i^T r_0}{a_{ii}} e_i.$$

این رابطه نشان می‌دهد که درایه i ام بردار $\tilde{x}$ به صورت

$$\begin{aligned} \tilde{x}_i &= (x_0)_i + \frac{e_i^T r_0}{a_{ii}} \\ &= (x_0)_i + \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^n a_{ij} (x_0)_j \right) \\ &= \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j \neq i}^n a_{ij} (x_0)_j \right), \end{aligned}$$

۱۸۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

بهبود می‌یابد و سایر درایه‌ها بدون تغییر باقی می‌مانند. واضح است اگر به‌ازای $n, \dots, 1, i$ و به‌ترتیب از 1 تا $i = n$ روش تصویری روی $\mathcal{L} = \mathcal{K} = \text{span}\{e_i\}$ انجام گیرد، معادل با یک تکرار روش گaus-سایدل خواهد بود.

۲-۵ نمایش ماتریسی

فرض کنید $V = (v_1, v_2, \dots, v_m)$ یک ماتریس $n \times m$ باشد به‌طوری که ستون‌های آن یک پایه برای $\mathcal{K}$ و همچنین $W = (w_1, w_2, \dots, w_m)$ نیز یک ماتریس $n \times m$ باشد که ستون‌های آن پایه‌ای برای $\mathcal{L}$ است. اکنون، اگر جواب تقریبی به‌صورت $\tilde{x} = x_* + \tilde{\delta}$ و $\tilde{\delta} \in \mathcal{K}$ نوشته شود، از اینکه $\tilde{\delta} \in \mathcal{K}$ برداری مانند y وجود دارد به‌طوری که $\tilde{\delta} = Vy$ ، لذا، داریم

$$\tilde{x} = x_* + Vy.$$

اینک، با توجه به شرایط گالرکین و با فرض اینکه $\det(W^H AV) \neq 0$ ، داریم

$$\begin{aligned} \forall w \in W : r_* - A\tilde{\delta} \perp w &\implies W^H(r_* - A\tilde{\delta}) = 0 \\ &\implies W^H r_* = W^H AVy \\ &\implies y = (W^H AV)^{-1} W^H r_*, \end{aligned}$$

ازاین‌رو،

$$\tilde{x} = x_* + V(W^H AV)^{-1} W^H r_*. \quad (5-5)$$

با جمع‌بندی این مطالب نمونه اصلی یک روش تصویری را می‌توان به‌صورت الگوریتم ۱.۵ نوشت.

الگوریتم ۱.۵ نمونه اصلی روش تصویری

1. Until convergence, Do:
2. Select a pair of subspaces $\mathcal{K}$ and $\mathcal{L}$
3. Choose bases V and W for $\mathcal{K}$ and $\mathcal{L}$, respectively
4. $r := b - Ax$
5. $y := (W^H AV)^{-1} W^H r$
6. $x := x + Vy$
7. EndDo

همان‌گونه که اشاره شد، این الگوریتم زمانی می‌تواند مورد استفاده قرار گیرد که ماتریس $W^H AV$ نامنفرد باشد. توجه کنید که حتی اگر A نامنفرد باشد، لزومی ندارد که $W^H AV$ نامنفرد باشد. به مثال بعدی توجه کنید.

مثال ۲.۵ ماتریس

$$A = \begin{pmatrix} 0 & I \\ I & I \end{pmatrix},$$

را در نظر بگیرید که در آن I ماتریس همانی $m \times m$ و 0 ماتریس صفر $m \times m$ است. همچنین، فرض کنید $V = W = (e_1, e_2, \dots, e_m)$. لذا، داریم

$$V = W = \begin{pmatrix} I \\ 0 \end{pmatrix},$$

که در آن I ماتریس همانی $m \times m$. در نتیجه،

$$W^H AV = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} 0 & I \\ I & I \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} 0 \\ I \end{pmatrix} = 0.$$

بنابراین، ملاحظه می‌شود که $W^H AV$ منفرد است درحالی‌که ماتریس A نامنفرد است. قضیه بعدی نشان می‌دهد که در دو حالت خاص ماتریس $W^H AV$ نامنفرد خواهد بود.

قضیه ۱.۵ اگر A ، $\mathcal{L}$ و $\mathcal{K}$ در یکی از دو شرط

(الف) A معین مثبت هرمیتی باشد و $\mathcal{L} = \mathcal{K}$ ؛

(ب) A نامنفرد باشد و $\mathcal{L} = A\mathcal{K}$ ؛

صدق کنند، آنگاه ماتریس $B = W^H AV$ به‌ازای هر پایه V برای $\mathcal{K}$ و هر پایه W برای $\mathcal{L}$ نامنفرد خواهد بود.

برهان. حالت (الف) را در نظر بگیرید. فرض کنید V پایه‌ای برای $\mathcal{K}$ و W پایه‌ای برای $\mathcal{L}$ باشد. درحقیقت، چون $\mathcal{L}$ و $\mathcal{K}$ برابرند، می‌توانیم بنویسیم $W = VG$ که در آن G یک ماتریس نامنفرد $m \times m$ است (چرا؟). بنابراین، داریم

$$B = W^H AV = G^H V^H AV.$$

اکنون، چون ماتریس A معین مثبت هرمیتی است، ماتریس $V^H AV$ نیز معین مثبت هرمیتی خواهد بود. زیرا اگر $u \neq 0$ یک بردار دلخواه باشد، آنگاه

$$u^H V^H AV u = (Vu)^H A (Vu) > 0.$$

توجه کنید که $Vu \neq 0$ ؛ زیرا اگر $Vu = 0$ ، آنگاه $V^H Vu = 0$ و این یعنی $u = 0$ ، که یک تناقض است.

برای اثبات حالت (ب)، فرض کنید V یک ماتریس پایه برای $\mathcal{K}$ و W یک ماتریس پایه برای $\mathcal{L}$ باشد. از اینکه $\mathcal{L} = AK$ ، یک ماتریس نامنفرد $m \times m$ مثل G وجود دارد به‌طوری‌که $W = AVG$. لذا، داریم

$$B = W^H AV = G^H (AV)^H AV.$$

چون A نامنفرد است، نتیجه می‌شود که ماتریس AV با رتبه کامل است. لذا، ماتریس $(AV)^H AV$ و در نتیجه، ماتریس B نامنفرد است. $\square$

اکنون، حالت خاصی را در نظر بگیرید که در آن ماتریس A هرمیتی باشد. در این حالت چون زیرفضاها برابرند، می‌توان ماتریس پایه را برای هر دو فضا یکسان در نظر گرفت. اگر این ماتریس پایه را V بگیریم، آنگاه ماتریس $B = V^H AV$ هرمیتی خواهد بود. به‌علاوه، اگر A معین مثبت هرمیتی باشد، آنگاه $V^H AV$ نیز معین مثبت هرمیتی خواهد بود.

۳-۵ آنالیز روش‌های تصویری

نخست، دو قضیه اساسی ارائه می‌شود که ارتباط بین یک روش تصویری با یک نوع بهینگی را بیان می‌کند.

قضیه ۲.۵ فرض کنید ماتریس A معین مثبت هرمیتی باشد و $\mathcal{L} = \mathcal{K}$. در این صورت، بردار $\tilde{x}$ از یک روش تصویری (متعامد) روی $\mathcal{K}$ با حدس اولیه x_0 به‌دست می‌آید اگر و تنها اگر $\tilde{x}$ ، A -نرم خطا را روی $x_0 + \mathcal{K}$ کمینه کند؛ یعنی اگر و تنها اگر

$$\|x^* - \tilde{x}\|_A = \min_{x \in x_0 + \mathcal{K}} \|x^* - x\|_A, \quad (۶-۵)$$

که در آن $\|z\|_A = (Az, z)^{1/2}$ و x_* جواب واقعی دستگاه (۱-۵) است.

برهان. فرض کنید $x = x_0 + \delta$ و $\tilde{x} = x_0 + \tilde{\delta}$ که در آن $\delta, \tilde{\delta} \in \mathcal{K}$. در این صورت، رابطه (۵-۶) معادل با

$$\|(x^* - x_0) - \tilde{\delta}\|_A = \min_{\delta \in \mathcal{K}} \|(x^* - x_0) - \delta\|_A,$$

است. اکنون، با توجه به نتیجه (۲۰۴)، این رابطه برقرار است اگر و تنها اگر

$$\begin{cases} \tilde{\delta} \in \mathcal{K}, \\ (x^* - x_0) - \tilde{\delta} \perp_A \delta, \quad \forall \delta \in \mathcal{K}, \end{cases}$$

که در آن $r \perp_A s$ اگر و تنها اگر $(r, s)_A = 0$. رابطه اخیر معادل است با

$$\begin{cases} x_0 + \tilde{\delta} \in x_0 + \mathcal{K}, \\ \langle A(x^* - (x_0 + \tilde{\delta})), \delta \rangle = 0, \quad \forall \delta \in \mathcal{K}, \end{cases}$$

و این خود نیز معادل است با

$$\begin{cases} \tilde{x} \in x_0 + \mathcal{K}, \\ \langle b - A\tilde{x}, \delta \rangle = 0, \quad \forall \delta \in \mathcal{K}, \end{cases}$$

□

که اثبات قضیه را کامل می‌کند.

قضیه ۳.۵ فرض کنید A یک ماتریس مربعی نامنفرد باشد و $\mathcal{L} = AK$. آنگاه بردار $\tilde{x}$ از یک روش تصویری بر روی $\mathcal{K}$ و عمود بر $\mathcal{L}$ ، با حدس اولیه x_0 به دست می‌آید اگر و تنها اگر $\tilde{x}$ نرم-۲ بردار مانده $b - Ax$ را روی $x_0 + \mathcal{K}$ کمینه کند؛ یعنی اگر و تنها اگر

$$\|b - A\tilde{x}\|_2 = \min_{x \in x_0 + \mathcal{K}} \|b - Ax\|_2. \quad (۷-۵)$$

برهان. فرض کنید $x = x_0 + \delta$ و $\tilde{x} = x_0 + \tilde{\delta}$ که در آن $\delta, \tilde{\delta} \in \mathcal{K}$. در این صورت، رابطه (۷-۵) معادل است با

$$\|b - Ax_0 - A\tilde{\delta}\|_2 = \min_{\delta \in \mathcal{K}} \|b - Ax_0 - A\delta\|_2,$$

که این خود معادل با

$$\|r_0 - A\tilde{\delta}\|_2 = \min_{\delta \in \mathcal{K}} \|r_0 - A\delta\|_2,$$

است. همچنین، رابطهٔ اخیر را می‌توان به صورت

$$\|r_\circ - \tilde{w}\|_2 = \min_{w \in \mathcal{L} = AK} \|r_\circ - w\|_2,$$

نوشت. اما بنا به (۲.۴)، رابطهٔ اخیر معادل است با

$$\begin{cases} \tilde{w} \in \mathcal{L} = AK, \\ \langle r_\circ - \tilde{w}, w \rangle = 0, \quad \forall w \in \mathcal{L} = AK, \end{cases}$$

که این خود با

$$\begin{cases} A\tilde{\delta} \in AK, \\ \langle r_\circ - A\tilde{\delta}, w \rangle = 0, \quad \forall w \in \mathcal{L}, \end{cases}$$

معادل است. اما این رابطه معادل است با

$$\begin{cases} \tilde{\delta} \in \mathcal{K}, \\ \langle b - A(x_\circ + \tilde{\delta}), w \rangle = 0, \quad \forall w \in \mathcal{L}, \end{cases}$$

یا

$$\begin{cases} \tilde{x} \in x_\circ + \mathcal{K}, \\ \langle b - A\tilde{x}, w \rangle = 0, \quad \forall w \in \mathcal{L} = AK, \end{cases}$$

□ که دقیقاً همان شرایط پتروف-گالرکین برای تقریب $\tilde{x}$ می‌باشد.

قضیهٔ ۴.۵ فرض کنید $\tilde{x}$ جواب تقریبی به دست آمده توسط یک روش تصویری روی $\mathcal{K}$ و عمود بر $\mathcal{L} = AK$ باشد. اگر $\tilde{r} = b - A\tilde{x}$ بردار ماندهٔ متناظر باشد، آنگاه

$$\tilde{r} = (I - P)r_\circ,$$

که در آن P عملگر تصویر متعامد روی زیرفضای AK است.

برهان. فرض کنید $\tilde{x} = x_\circ + \tilde{\delta}$ ، که در آن $\tilde{\delta} \in \mathcal{K}$ و بردار ماندهٔ متناظر به $x_\circ$ به صورت $r_\circ = b - Ax_\circ$ باشد. از اینکه $\tilde{x}$ یک جواب تقریبی به دست آمده توسط روش تصویری است، باید داشته باشیم

$$b - A\tilde{x} \perp \mathcal{L}.$$

ازطرفی،

$$\tilde{r} = b - A\tilde{x} = b - A(x_0 + \tilde{\delta}) = b - Ax_0 - A\tilde{\delta} = r_0 - A\tilde{\delta}.$$

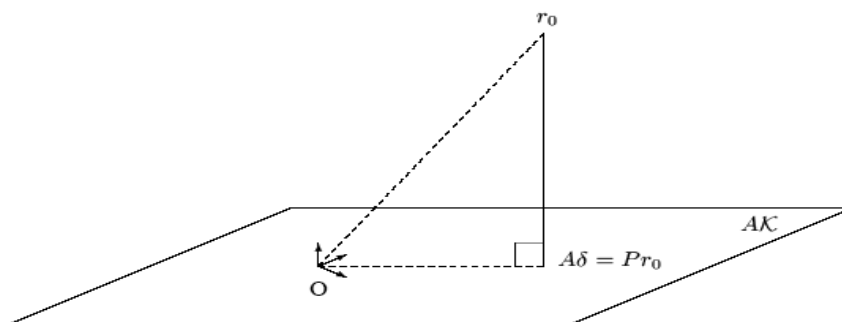
درنتیجه،

$$r_0 - A\tilde{\delta} \perp \mathcal{L} = AK.$$

از اینکه $r_0 - A\tilde{\delta}$ عمود بر AK است نتیجه می‌شود که $A\tilde{\delta}$ تصویر عمودی r_0 روی زیرفضای AK است. اکنون، اگر P عملگر تصویر متعامد باشد، می‌توان نوشت $A\tilde{\delta} = Pr_0$ (به شکل ۲-۵ توجه کنید). لذا، داریم

$$\tilde{r} = r_0 - A\tilde{\delta} = r_0 - Pr_0 = (I - P)r_0,$$

□ که اثبات قضیه را کامل می‌کند.



شکل ۲-۵ نمایش عملگر تصویر متعامد در قضیه ۴.۵

نتیجه ۱.۵ اگر P یک عملگر تصویر باشد، آنگاه $I - P$ نیز یک عملگر تصویر است و از اینکه P متعامد است نتیجه می‌شود $\|I - P\|_2 = 1$. بنابراین،

$$\|\tilde{r}\|_2 = \|(I - P)r_0\|_2 \leq \|I - P\|_2 \|r_0\|_2 = \|r_0\|_2.$$

یعنی

$$\|\tilde{r}\|_2 \leq \|r_0\|_2,$$

که نشان می‌دهد که در هر مرحله، نرم-۲ بردار مانده کمتر یا مساوی نرم-۲ بردار مانده مرحله قبل است.

قضیه ۵.۵ فرض کنید A یک ماتریس معین مثبت هرمیتی و $\tilde{x}$ جواب تقریبی دستگاه (۵-۱) باشد که توسط روش تصویری متعامد روی $\mathcal{K}$ ($\mathcal{L} = \mathcal{K}$) به دست آمده و $\tilde{d} = x_* - \tilde{x}$ در این صورت،

$$\tilde{d} = (I - P_A)d_*,$$

که در آن P_A عملگر تصویر A -عمود روی $\mathcal{K}$ می‌باشد.

برهان. فرض کنید ماتریس A معین مثبت هرمیتی باشد، $\mathcal{L} = \mathcal{K}$ ، x_* جواب واقعی دستگاه (۵-۱) باشد و

$$\tilde{d} = x_* - \tilde{x}, \quad d_* = x_* - x_0.$$

با توجه به روش تصویری، $\tilde{\delta}$ طوری تعیین می‌شود که $\tilde{\delta} = b - A\tilde{x} \perp \mathcal{L} = \mathcal{K}$ ؛ یعنی

$$\langle r_* - A\tilde{\delta}, w \rangle = 0, \quad \forall w \in \mathcal{K}.$$

اما داریم

$$\begin{aligned} r_* - A\tilde{\delta} &= b - Ax_* - A\tilde{\delta} = Ax_* - Ax_* - A\tilde{\delta} \\ &= A(x_* - x_0) - A\tilde{\delta} = Ad_* - A\tilde{\delta} \\ &= A(d_* - \tilde{\delta}). \end{aligned}$$

بنابراین،

$$\langle A(d_* - \tilde{\delta}), w \rangle = 0, \quad \forall w \in \mathcal{K}.$$

اکنون، چون ماتریس A معین مثبت هرمیتی است، می‌توان نوشت

$$\langle d_* - \tilde{\delta}, w \rangle_A = 0, \quad \forall w \in \mathcal{K}.$$

و این یعنی بردار $\tilde{\delta}$ تصویر A -عمود بردار d_* روی زیرفضای $\mathcal{K}$ است. اگر عملگر تصویر A -عمود را با P_A نمایش دهیم، داریم

$$\tilde{\delta} = P_A d_*. \quad (۵-۸)$$

ازطرفی، داریم

$$\tilde{d} = x_* - \tilde{x} = x_* - (x_* + \tilde{\delta}) = (x_* - x_*) - \tilde{\delta} = d_* - \tilde{\delta}.$$

بنابراین، با توجه به رابطه (۵-۸)، داریم

$$\tilde{d} = d_* - \tilde{\delta} = d_* - P_A d_* = (I - P_A) d_*.$$

یعنی $\tilde{d} = (I - P_A) d_*$ که بدین ترتیب اثبات قضیه کامل می‌گردد. $\square$

نتیجه ۲.۵ اگر عملگر تصویر A -عمود را با P_A نشان دهیم، آنگاه خواهیم داشت

$$\|\tilde{d}\|_A = \|(I - P_A) d_*\|_A \leq \|I - P_A\|_A \cdot \|d_*\|_A = \|d_*\|_A.$$

درنتیجه، $\|\tilde{d}\|_A \leq \|d_*\|_A$ ؛ یعنی در هر مرحله A -نرم خطا، کوچک‌تر یا مساوی A -نرم خطای مرحله قبل است.

۴-۵ روش‌های تصویری یک‌بعدی

فرض کنید x_k جواب تقریبی به دست آمده در تکرار k ام باشد. به‌طور کلی، یک روش تصویری یک‌بعدی با استفاده $\mathcal{K}_k = \text{span}\{v_k\}$ و $\mathcal{L}_k = \text{span}\{w_k\}$ به دست می‌آیند، که در آن w_k و v_k دو بردار هستند. در این صورت، جواب تقریبی در تکرار بعدی به صورت

$$x_{k+1} = x_k + \alpha_k v_k,$$

خواهد بود، که در آن $\alpha_k \in \mathbb{R}$. از این رو،

$$b - Ax_{k+1} = b - Ax_k - \alpha_k Av_k = r_k - \alpha_k Av_k.$$

لذا، با استفاده از شرط پتروف-گالرکین $b - Ax_{k+1} \perp \mathcal{L}_k$ ، اگر $w_k^T Av_k \neq 0$ ، آنگاه خواهیم داشت

$$w_k^T (r_k - \alpha_k Av_k) = 0 \implies \alpha_k = \frac{w_k^T r_k}{w_k^T Av_k} = \frac{\langle r_k, w_k \rangle}{\langle Av_k, w_k \rangle}.$$

سه انتخاب مهم برای w_k و v_k وجود دارد که در ادامه به شرح آنها می‌پردازیم.

۵-۴-۱ روش سریع‌ترین کاهش

در این روش قرار می‌دهیم $v_k = w_k = r_k$. در این صورت، گام‌های اصلی روش به صورت

$$r_k := b - Ax_k,$$

$$\alpha_k := \langle r_k, r_k \rangle / \langle Ar_k, r_k \rangle,$$

$$x_{k+1} := x_k + \alpha_k r_k,$$

خواهند بود. همان‌گونه که می‌بینیم این روند شامل دو ضرب ماتریس در بردار Ax_k و Ar_k است. اما با توجه به

$$\begin{aligned} r_{k+1} &= b - Ax_{k+1} \\ &= b - A(x_k + \alpha_k r_k) \\ &= r_k - \alpha_k Ar_k, \end{aligned}$$

می‌توان آن را به صورت الگوریتم ۲.۵ جمع‌بندی نمود به طوری که تنها شامل یک ضرب ماتریس در بردار باشد.

الگوریتم ۲.۵ سریع‌ترین کاهش (SD)

1. Choose an initial guess x_0 , $r_0 := b - Ax_0$ and $p_0 := Ar_0$
2. For $k = 0, 1, 2, \dots$, until convergence, Do
3. $\alpha_k := \langle r_k, r_k \rangle / \langle p_k, r_k \rangle$
4. $x_{k+1} := x_k + \alpha_k r_k$
5. $r_{k+1} := r_k - \alpha_k p_k$
6. $p_{k+1} := Ar_{k+1}$
7. EndDo

ملاحظه ۱.۵ فرض کنید

$$f(x) = \frac{1}{2} \|x - x_*\|_A^2 = \frac{1}{2} \langle A(x - x_*), (x - x_*) \rangle.$$

به سادگی می‌توان دید که الگوریتم سریع‌ترین کاهش در تکرار $(k+1)$ ام برداری مثل

$$x_{k+1} = x_k + \alpha_k p_k,$$

محاسبه می‌کند به‌طوری‌که

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}} f(x_k + \alpha p_k),$$

$$p_k = -\nabla f(x)|_{x_k}$$

در ادامه همگرایی روش سریع‌ترین کاهش را مطالعه می‌کنیم.

لم ۱.۵ (نامساوی کانترولیچ) فرض کنید ماتریس B معین مثبت متقارن و $\lambda_{\min}$ و $\lambda_{\max}$ به‌ترتیب کوچک‌ترین و بزرگ‌ترین مقدار ویژه B باشند. در این صورت، به‌ازای هر $x \in \mathbb{R}^n$ ، داریم

$$\frac{\langle Bx, x \rangle \langle B^{-1}x, x \rangle}{\langle x, x \rangle^2} \leq \frac{(\lambda_{\min} + \lambda_{\max})^2}{4\lambda_{\min}\lambda_{\max}}.$$

برهان. بدون از دست دادن کلیت، می‌توان فرض کرد $\|x\|_2 = 1$. از اینکه ماتریس B متقارن است، یک ماتریس متعامد مثل Q وجود دارد به‌طوری‌که $B = Q^T D Q$ ، که در آن $D = \text{diag}_n(\lambda_1, \lambda_2, \dots, \lambda_n)$ و λ_i ها مقادیر ویژه B هستند به‌طوری‌که

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n.$$

داریم

$$\begin{aligned} \langle Bx, x \rangle \langle B^{-1}x, x \rangle &= \langle Q^T D Q x, x \rangle \langle Q^T D^{-1} Q x, x \rangle \\ &= \langle D Q x, Q x \rangle \langle D^{-1} Q x, Q x \rangle. \end{aligned}$$

قرار می‌دهیم $z = Qx = (z_1, z_2, \dots, z_n)^T$. در این صورت،

$$\|z\|_2 = \|Qx\|_2 = \|x\|_2 = 1.$$

بنابراین، با فرض $i = 1, 2, \dots, n$ ، $\beta_i = z_i^2$ داریم $\sum_{i=1}^n \beta_i = 1$ و

$$\begin{aligned} \langle Bx, x \rangle \langle B^{-1}x, x \rangle &= \langle D z, z \rangle \langle D^{-1} z, z \rangle \\ &= \sum_{i=1}^n \beta_i \lambda_i \sum_{i=1}^n \beta_i \frac{1}{\lambda_i} \\ &= \sum_{i=1}^n \beta_i \frac{1}{\lambda_i}, \end{aligned}$$

که در آن

$$\xi = \sum_{i=1}^n \beta_i \lambda_i.$$

روشن است که $\lambda_1 \leq \xi \leq \lambda_n$. فرض کنید $d(\lambda)$ خط‌گذرنده از دو نقطه $(\lambda_1, 1/\lambda_1)$ و $(\lambda_n, 1/\lambda_n)$ باشد. باتوجه به اینکه تابع $f(\lambda) = 1/\lambda$ در بازه $[\lambda_1, \lambda_n]$ محدب است، داریم (شکل ۳-۵ را ببینید)

$$f(\lambda) \leq d(\lambda), \quad \forall \lambda \in [\lambda_1, \lambda_n].$$

بنابراین،

$$\begin{aligned} \sum_{i=1}^n \beta_i \frac{1}{\lambda_i} &= \sum_{i=1}^n \beta_i f(\lambda_i) \\ &\leq \sum_{i=1}^n \beta_i d(\lambda_i) \\ &= \sum_{i=1}^n \beta_i \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{\lambda_i}{\lambda_1 \lambda_n} \right) \\ &= \frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{1}{\lambda_1 \lambda_n} \sum_{i=1}^n \beta_i \lambda_i \\ &= \frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{\xi}{\lambda_1 \lambda_n}. \end{aligned}$$

درنتیجه،

$$(Bx, x)(B^{-1}x, x) \leq \xi \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_n} - \frac{\xi}{\lambda_1 \lambda_n} \right).$$

اما بیشینه سمت راست این رابطه، به‌ازای $\xi = (\lambda_1 + \lambda_n)/2$ به‌دست می‌آید و درنتیجه،

$$(Bx, x)(B^{-1}x, x) \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n},$$

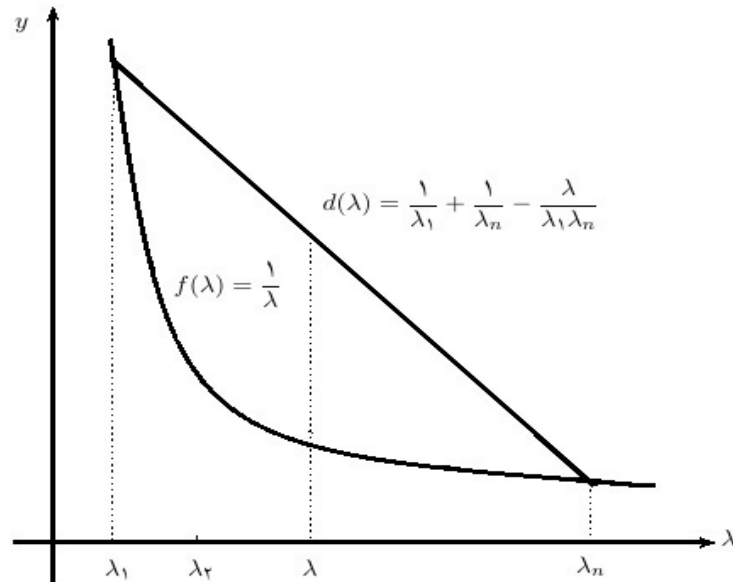
□

که اثبات لم کامل می‌شود.

قضیه ۶.۵ فرض کنید ماتریس A معین مثبت متقارن باشد. در این صورت، A -نرم بردارهای خطا، یعنی

$$d_k = x_* - x_k$$

$$\|d_{k+1}\|_A \leq \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} \|d_k\|_A,$$



شکل ۳-۵ نمایش تابع $f(\lambda)$ و خط $d(\lambda)$ در لم ۱.۵.

صدق می‌کند. این رابطه نشان می‌دهد که الگوریتم سریع‌ترین کاهش به‌ازای هر حدس اولیه دلخواه همگرا است.

برهان. ابتدا توجه کنید که

$$\begin{aligned} Ad_{k+1} &= A(x^* - x_{k+1}) = Ax^* - Ax_{k+1} \\ &= b - Ax_{k+1} = r_{k+1}, \end{aligned}$$

و $d_{k+1} = d_k - \alpha_k r_k$ بنابراین، با توجه به اینکه

$$\langle r_k - \alpha_k Ar_k, r_k \rangle = \langle r_{k+1}, r_k \rangle = 0,$$

داریم

$$\begin{aligned} \|d_{k+1}\|_A^2 &= \langle Ad_{k+1}, d_{k+1} \rangle = \langle r_{k+1}, d_{k+1} \rangle \\ &= \langle r_k - \alpha_k Ar_k, d_k - \alpha_k r_k \rangle \\ &= \langle r_k, d_k \rangle - \alpha_k \langle r_k - \alpha_k Ar_k, r_k \rangle - \alpha_k \langle Ar_k, d_k \rangle \end{aligned}$$

$$\begin{aligned}
 &= \|d_k\|_A^2 \left(1 - \alpha_k \frac{\langle Ar_k, d_k \rangle}{\langle r_k, d_k \rangle} \right) \\
 &= \|d_k\|_A^2 \left(1 - \frac{\langle r_k, r_k \rangle}{\langle Ar_k, r_k \rangle} \frac{\langle Ar_k, d_k \rangle}{\langle r_k, d_k \rangle} \right) \\
 &= \|d_k\|_A^2 \left(1 - \frac{\langle r_k, r_k \rangle}{\langle Ar_k, r_k \rangle} \frac{\langle r_k, Ad_k \rangle}{\langle r_k, A^{-1}r_k \rangle} \right) \\
 &= \|d_k\|_A^2 \left(1 - \frac{\langle r_k, r_k \rangle}{\langle Ar_k, r_k \rangle} \frac{\langle r_k, r_k \rangle}{\langle r_k, A^{-1}r_k \rangle} \right).
 \end{aligned}$$

اکنون، با استفاده از لم ۱.۵ نتیجه لازم به دست می‌آید. با توجه به

$$\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} < 1,$$

□

قسمت دوم قضیه برقرار است.

توجه کنید که داریم

$$\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}} = \frac{\text{cond}_2(A) - 1}{\text{cond}_2(A) + 1}.$$

لذا، اگر $\text{cond}_2(A)$ بزرگ باشد (ماتریس A بدحالت باشد)، آنگاه سرعت همگرایی روش سریع‌ترین کاهش بسیار کند خواهد بود.

مثال ۳.۵ دستگاه معرفی شده در بخش ۱-۸ را در نظر بگیرید که در آن $\sigma = \tau = 0$. می‌توان دید که ماتریس ضرایب دستگاه معین مثبت متقارن است. این دستگاه را به‌ازای مقادیر $N = 10, 20, 30$ حل کرده و نتایج را در جدول ۵-۱ ارائه کرده‌ایم. در این جدول $time$ و $iters$ به‌ترتیب تعداد تکرارها و زمان لازم (برحسب ثانیه) برای همگرایی است. حدس اولیه را بردار $x_0 = 0$ ، بردار سمت راست را به‌صورت

$$b = A(1, 1, \dots, 1)^T,$$

و شرط توقف را نیز به‌صورت $\|r_k\|_2 < 10^{-7} \|r_0\|_2$ گرفته‌ایم. در این جدول علاوه بر نتایج مربوط به روش سریع‌ترین کاهش (SD)، نتایج مربوط به روش‌های تکراری ژاکوبی (Jacobi)، و گاوس-سایدل (GS) را نیز ارائه کرده‌ایم. این جدول نشان می‌دهد که از لحاظ تعداد تکرار، روش‌های ژاکوبی و سریع‌ترین کاهش شبیه به هم هستند و روش گاوس-سایدل از هردوی آنها بهتر است. با این وجود، می‌بینیم که روش سریع‌ترین کاهش از لحاظ زمانی از هردو روش ژاکوبی و گاوس-سایدل بهتر است.

جدول ۱-۵ نتایج عددی برای مثال ۳.۵.

$n = N^2$	nnz	Jacobi		GS		SD	
		$iters$	$time$	$iters$	$time$	$iters$	$time$
۱۰۰	۴۶۰	۳۵۳	۰/۰۱	۱۷۸	۰/۱۵	۳۳۸	۰/۰۱
۴۰۰	۱۹۲۰	۱۲۱۲	۰/۰۶	۶۰۸	۱/۷۵	۱۲۲۱	۰/۰۲
۹۰۰	۴۳۸۰	۲۵۳۴	۰/۹۴	۱۲۶۹	۸/۷۸	۲۵۷۷	۰/۰۷
۱۶۰۰	۷۸۴۰	۴۲۹۴	۵/۸۳	۲۱۴۹	۲۹/۷۵	۴۳۸۳	۰/۱۶

۲-۴-۵ مانده مینیمال

در این روش فرض می‌کنیم ماتریس A معین مثبت باشد (نه لزوماً متقارن)، به عبارت دیگر، ماتریس $A + A^T$ معین مثبت متقارن است. ساختار روش کاملاً مشابه روش سریع‌ترین کاهش است، اما با این تفاوت که در این روش قرار می‌دهیم $v_k = r_k$ و $w_k = Ar_k$. در این صورت، گام‌های اصلی روش به صورت

$$r_k := b - Ax_k,$$

$$\alpha_k := \langle Ar_k, r_k \rangle / \langle Ar_k, Ar_k \rangle,$$

$$x_{k+1} := x_k + \alpha_k r_k,$$

خواهند بود. این روند شامل دو ضرب ماتریس در بردار Ar_k و Ax_k است. اما با توجه به

$$r_{k+1} = b - Ax_{k+1} = b - A(x_k + \alpha_k r_k) = r_k - \alpha_k Ar_k,$$

می‌توان آن را به صورت الگوریتم ۳.۵ جمع‌بندی نمود به‌طوری‌که تنها شامل یک ضرب ماتریس در بردار باشد.

الگوریتم ۳.۵ مانده مینیمال (MR)

1. Choose an initial guess x_0 , $r_0 := b - Ax_0$ and $p_0 = Ar_0$
2. For $k = 0, 1, 2, \dots$, until convergence, Do
3. $\alpha_k := \langle p_k, r_k \rangle / \langle p_k, p_k \rangle$
4. $x_{k+1} := x_k + \alpha_k r_k$
5. $r_{k+1} := r_k - \alpha_k p_k$
6. $p_{k+1} := Ar_{k+1}$
7. EndDo

ملاحظه ۲.۵ فرض کنید $f(x) = \frac{1}{2} \|b - Ax\|_2^2$. نشان دهید که روش تکراری مانده مینیمال در تکرار $(k+1)$ ام برداری مثل $x_{k+1} = x_k + \alpha_k r_k$ محاسبه می‌کند، که در آن

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}} f(x_k + \alpha r_k).$$

در ادامه همگرایی روش مانده مینیمال را بررسی می‌کنیم.

قضیه ۷.۵ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ معین مثبت باشد و

$$\mu = \frac{1}{2} \lambda_{\min}(A + A^T), \quad \sigma = \|A\|_2.$$

در این صورت، بردارهای مانده الگوریتم مانده مینیمال در رابطه

$$\|r_{k+1}\|_2 \leq \left(1 - \frac{\mu}{\sigma^2}\right)^{1/2} \|r_k\|_2,$$

صدق می‌کند و الگوریتم ۳.۵ به‌ازای هر حدس اولیه دلخواه x_0 همگراست.

برهان. نخست توجه کنید که $\langle r_{k+1}, Ar_k \rangle = 0$. بنابراین،

$$\begin{aligned} \|r_{k+1}\|_2^2 &= \langle r_{k+1}, r_{k+1} \rangle = \langle r_{k+1}, r_k - \alpha_k Ar_k \rangle = \langle r_{k+1}, r_k \rangle \\ &= \langle r_k - \alpha_k Ar_k, r_k \rangle = \langle r_k, r_k \rangle - \alpha_k \langle Ar_k, r_k \rangle \\ &= \|r_k\|_2^2 \left(1 - \frac{\langle Ar_k, r_k \rangle}{\langle Ar_k, Ar_k \rangle} \frac{\langle Ar_k, r_k \rangle}{\langle r_k, r_k \rangle}\right) \\ &= \|r_k\|_2^2 \left(1 - \frac{\langle Ar_k, r_k \rangle^2}{\langle r_k, r_k \rangle^2} \frac{\|r_k\|_2^2}{\|Ar_k\|_2^2}\right). \end{aligned} \quad (9-5)$$

اکنون با استفاده از سازگاری نرم، داریم

$$\frac{\|r_k\|_2^2}{\|Ar_k\|_2^2} \geq \frac{\|r_k\|_2^2}{\|A\|_2^2 \|r_k\|_2^2} = \frac{1}{\sigma^2}. \quad (10-5)$$

همچنین، با استفاده از قضیه ۱۱.۱، داریم

$$\frac{\langle Ar_k, r_k \rangle}{\langle r_k, r_k \rangle} = \frac{1}{2} \frac{\langle (A + A^T)r_k, r_k \rangle}{\langle r_k, r_k \rangle} \geq \frac{1}{2} \lambda_{\min}(A + A^T) = \mu > 0. \quad (11-5)$$

حال با جایگذاری (۱۰-۵) و (۱۱-۵) در (۹-۵)، نتیجه لازم به‌دست می‌آید.

جدول ۲-۵ نتایج عددی برای مثال ۴.۵.

$n = N^2$	nnz	MR	
		$iters$	$time$
۳۶۰۰۰۰	۱۷۹۷۶۰۰	۱۷۲۰	۱۹/۸۳
۶۴۰۰۰۰	۳۱۹۶۸۰۰	۳۱۰۴	۶۲/۹۵
۱۰۰۰۰۰۰	۴۹۹۶۰۰۰	۴۸۷۳	۱۵۲/۶۹

برای اثبات قسمت آخر قضیه، توجه کنید که

$$\begin{aligned}\mu &= \frac{1}{2} \lambda_{\min}(A + A^T) \leq \frac{1}{2} \|A + A^T\|_2 \leq \frac{1}{2} (\|A\|_2 + \|A^T\|_2) \\ &= \frac{1}{2} (2\|A\|_2) = \|A\|_2 = \sigma.\end{aligned}$$

□ لذا، داریم $1 \leq \mu^2/\sigma^2 < \infty$ ، که همگرایی روش را نشان می‌دهد.

مثال ۴.۵ دستگاه معرفی‌شده در بخش ۱-۸ را در نظر بگیرید که در آن $\sigma = 10$ و $\tau = -10$. می‌توان دید که ماتریس A معین مثبت است (متقارن نیست). این دستگاه را به‌ازای مقادیر $N = 600, 800, 1000$ و با استفاده از روش ماندهٔ مینیمال (MR) حل کرده و نتایج را در جدول ۲-۵ ارائه کرده‌ایم. سایر فرض‌ها همانند مثال ۳.۵ می‌باشند. با توجه به قضیهٔ ۷.۵ روش ماندهٔ مینیمال برای حل این دستگاه همگراست. جدول ۲-۵ همگرایی روش را تأیید می‌کند.

۳-۴-۵ سریع‌ترین کاهش نرم مانده

در این روش، فرض می‌کنیم ماتریس A تنها نامنفرد باشد و قرار می‌دهیم $w_k = Av_k$ و $v_k = A^T r_k$ در این صورت، گام‌های اصلی روش به‌صورت

$$r_k := b - Ax_k,$$

$$v_k := A^T r_k,$$

$$\alpha_k := \|v_k\|_2 / \|Av_k\|_2,$$

$$x_{k+1} := x_k + \alpha_k v_k,$$

خواهند بود. همانند دو روش سریع‌ترین کاهش و ماندهٔ مینیمال، داریم

$$r_{k+1} = b - Ax_{k+1} = b - A(x_k + \alpha_k v_k) = r_k - \alpha_k Av_k.$$

اکنون، می‌توانیم الگوریتم سریع‌ترین کاهش نرم مانده را به صورت الگوریتم ۴.۵ جمع‌بندی کنیم.

الگوریتم ۴.۵ سریع‌ترین کاهش نرم مانده (RNSD)

1. Choose an initial guess $x_0, r_0 := b - Ax_0$
2. For $k = 0, 1, 2, \dots$, until convergence, Do
3. $v_k := A^T r_k$
4. $q_k := Av_k$
5. $\alpha_k := \|v_k\|_2^2 / \|q_k\|_2^2$
6. $x_{k+1} := x_k + \alpha_k v_k$
7. $r_{k+1} := r_k - \alpha_k q_k$
8. EndDo

ملاحظه ۳.۵ فرض کنید $\|b - Ax\|_2^2 = f(x)$. نشان دهید که روش سریع‌ترین کاهش نرم مانده در تکرار $(k+1)$ ام برداری مثل $x_{k+1} = x_k + \alpha_k p_k$ محاسبه می‌کند، به طوری که

$$\alpha_k = \arg \min_{\alpha \in \mathbb{R}} f(x_k + \alpha p_k),$$

که در آن $p_k = -\nabla f(x)|_{x_k}$.

قضیه ۸.۵ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ نامنفرد باشد. در این صورت، بردارهای مانده در الگوریتم سریع‌ترین کاهش نرم مانده در رابطه

$$\|r_{k+1}\|_2 \leq \|r_k\|_2 \left(1 - \frac{1}{\text{cond}_2(A)} \right), \quad (12-5)$$

صدق می‌کند. این رابطه نشان می‌دهد که الگوریتم سریع‌ترین کاهش نرم مانده با هر حدس اولیه دلخواه همگراست.

برهان. با توجه به نحوه ساخت الگوریتم سریع‌ترین کاهش نرم مانده، داریم $\langle r_{k+1}, Av_k \rangle = 0$. بنابراین،

$$\begin{aligned} \|r_{k+1}\|_2^2 &= \langle r_{k+1}, r_{k+1} \rangle = \langle r_{k+1}, r_k - \alpha_k Av_k \rangle \\ &= \langle r_{k+1}, r_k \rangle = \langle r_k - \alpha_k Av_k, r_k \rangle \\ &= \langle r_k, r_k \rangle - \alpha_k \langle Av_k, r_k \rangle \\ &= \|r_k\|_2^2 \left(1 - \frac{\|v_k\|_2^2 \langle Av_k, r_k \rangle}{\|Av_k\|_2^2 \langle r_k, r_k \rangle} \right). \end{aligned} \quad (13-5)$$

جدول ۳-۵ نتایج عددی برای مثال ۵.۵.

$n = N^2$	nnz	RNSD	
		$iters$	$time$
۹۰۰	۴۳۸۰	۵۱۹۹	۰/۱۷
۱۶۰۰	۷۸۴۰	۸۰۰۴	۰/۴۱
۲۵۰۰	۱۲۳۰۰	۱۱۴۴۰	۰/۹۱
۳۶۰۰	۱۷۷۶۰	۱۵۷۵۲	۱/۷۵

اکنون، با استفاده از سازگاری نرم، داریم

$$\frac{\|v_k\|_2^2}{\|Av_k\|_2^2} \geq \frac{\|v_k\|_2^2}{\|A\|_2^2 \|v_k\|_2^2} = \frac{1}{\|A\|_2^2} = \frac{1}{\sigma_{\max}^2(A)}. \quad (۱۴-۵)$$

همچنین، داریم

$$\frac{\langle Av_k, r_k \rangle}{\langle r_k, r_k \rangle} = \frac{\langle AA^T r_k, r_k \rangle}{\langle r_k, r_k \rangle} \geq \lambda_{\min}(AA^T) = \sigma_{\min}^2(A). \quad (۱۵-۵)$$

اکنون، با جایگذاری (۱۴-۵) و (۱۵-۵) در (۱۳-۵) و با توجه به اینکه

$$\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} = \text{cond}_2(A),$$

نتیجه لازم به دست می‌آید. سرانجام، با توجه به اینکه $\text{cond}_2(A) \geq 1$ ، رابطه (۱۲-۵) همگرایی روش را نتیجه می‌دهد. $\square$

یک نکته قابل ذکر درباره روش سریع‌ترین کاهش نرم مانده این است که سرعت همگرایی آن اغلب کند است؛ زیرا

$$1 - \frac{1}{\text{cond}_2(A)} \approx 1,$$

و با توجه به قضیه ۱۲-۵ سرعت همگرایی کند خواهد بود.

مثال ۵.۵ مثال ۴.۵ را به‌ازای $N = 30, 40, 50, 60$ و با استفاده از روش سریع‌ترین کاهش نرم مانده (RNSD) حل می‌کنیم. نتایج عددی در جدول ۳-۵ ارائه شده است. همان‌گونه که جدول نشان می‌دهد سرعت همگرایی روش بسیار کند است.

تمرینات

۱.۵ فرض کنید $A \in \mathbb{C}^{m \times n}$. نشان دهید $\text{cond}_2(A^H A) = (\text{cond}_2(A))^2$. این تمرین نشان می‌دهد که حل دستگاه معادلات خطی $A^H A x = A^H b$ به جای $Ax = b$ ممکن است مناسب نباشد. چه زمانی این کار می‌تواند مناسب باشد؟

۲.۵ نشان دهید که در مثال ۳.۵ ماتریس A معین مثبت متقارن و در مثال ۴.۵ معین مثبت است.

۳.۵ دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن ماتریس A معین مثبت متقارن است. (الف) فرض کنید e_i ستون i ام ماتریس همانی باشد. دنباله‌ای از روش‌های تصویری یک‌بُعدی را در نظر بگیرید که در آن $\mathcal{K} = \mathcal{L} = \text{span}\{e_i\}$ ، به‌طوری‌که اندیس i به‌طور دلخواه انتخاب می‌شود. فرض کنید x_{new} تقریب جدیدی باشد که بعد از یک گام روش تصویری به‌دست می‌آید و $r = b - Ax$ ، $d = A^{-1}b - x$ و $d_{\text{new}} = A^{-1}b - x_{\text{new}}$ نشان دهید

$$\langle Ad_{\text{new}}, d_{\text{new}} \rangle = \langle Ad, d \rangle - \frac{\langle r, e_i \rangle^2}{a_{ii}}.$$

آیا این رابطه همگرایی روش را نتیجه می‌دهد؟

(ب) فرض کنید در هر گام از روش تصویری i اندیس بزرگ‌ترین درایه از حیث قدرمطلق در بردار r باشد. نشان دهید

$$\|d_{\text{new}}\|_A \leq \left(1 - \frac{1}{n \text{cond}_2(A)}\right)^{1/2} \|d\|_A.$$

آیا این رابطه همگرایی روش را نشان می‌دهد؟ راهنمایی: داریم $\|r\|_2 / \sqrt{n} \geq |e_i^T r|$.

۴.۵ در این تمرین می‌خواهیم روش معرفی‌شده در تمرین ۳.۵ را تعمیم دهیم. فرض کنید ماتریس A معین مثبت متقارن باشد و

$$\mathcal{K} = \mathcal{L} = \text{span}\{e_{i_1}, e_{i_2}, \dots, e_{i_k}\},$$

به‌طوری‌که $1 \leq i_1 < i_2 < \dots < i_k \leq n$. یک روش تکراری می‌سازیم به‌طوری‌که هر تکرار آن از یک روش تصویری متعامد روی $\mathcal{K}$ به‌دست می‌آید. نشان دهید اگر در هر تکرار $\mathcal{J} = \{i_1, i_2, \dots, i_k\}$ مجموعه اندیس‌های مربوط به بزرگ‌ترین درایه‌های بردار r از حیث قدرمطلق باشند، آنگاه خواهیم داشت

$$\|d\|_A^2 - \|d_{\text{new}}\|_A^2 \geq \frac{\sum_{k \in \mathcal{J}} r_k^2}{\sum_{k \in \mathcal{J}} a_k},$$

و

$$\|d_{new}\|_A \leq \left(1 - \left(\frac{\lambda_{\min}(A)}{\sum_{k \in \mathcal{J}} a_{kk}}\right) \left(\frac{\sum_{k \in \mathcal{J}} r_k^2}{\sum_{k \in \mathcal{J}} r_k^2}\right)\right)^{1/2} \|d\|_A.$$

سپس همگرایی روش را نتیجه بگیرید. راهنمایی: مرجع [۶۰] را ببینید.

۵.۵ دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن ماتریس A معین مثبت متقارن است. (الف) فرض کنید v یک بردار ناصفر باشد. دنباله‌ای از روش‌های تصویری یک‌بُعدی را در نظر بگیرید که در آن $\mathcal{K} = \mathcal{L} = \text{span}\{v\}$. فرض کنید x_{new} یک تقریب جدید باشد که بعد از یک گام روش تصویری به دست می‌آید و $r = b - Ax$ ، $d = A^{-1}b - x$ و $d_{new} = A^{-1}b - x_{new}$. نشان دهید

$$\langle Ad_{new}, d_{new} \rangle = \langle Ad, d \rangle - \frac{\langle r, v \rangle^2}{\langle Av, v \rangle}.$$

آیا این رابطه همگرایی روش را نتیجه می‌دهد؟

(ب) فرض کنید بردار v طوری انتخاب شود که $\|r\|_1 = (v, r)$. برای مثال می‌توان فرض کرد $i = 1, \dots, n$ ، $v_i = \text{sign}(e_i^T r)$ نشان دهید

$$\|d_{new}\|_A \leq \left(1 - \frac{1}{n \text{cond}_2(A)}\right)^{1/2} \|d\|_A.$$

آیا این رابطه همگرایی روش را نشان می‌دهد؟

۶.۵ دستگاه $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} 2 & 6 & -3 & 1 \\ -6 & 5 & -2 & 2 \\ 3 & 1 & 8 & 1 \\ -1 & 1 & 1 & 10 \end{pmatrix}.$$

با ذکر دلیل بیان کنید که چرا روش ماندهٔ مینیمال (MR) برای حل این دستگاه همگرا است.

۷.۵ فرض کنید ماتریس A نامنفرد باشد. یک روش تکراری را در نظر بگیرید که در آن

$$x_{k+1} = x_k + \alpha_k d_k,$$

به‌طوری‌که d_k بردار جهت نامیده می‌شود و α_k یک اسکالر است. فرض کنید بردار d_k برداری است که

۲۰۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

به‌صورت $d_k = A^T f_k$ انتخاب می‌شود که در آن f_k یک بردار ناصفر است. فرض کنید $x^* = A^{-1}b$ جواب واقعی دستگاه $Ax = b$ باشد. حال روشی را در نظر بگیرید که در آن بردار x_{k+1} طوری محاسبه می‌شود که $\|x^* - x_{k+1}\|_2$ کمینه گردد.

(الف) α_k را طوری محاسبه کنید که $\|x^* - x_{k+1}\|_2$ کمینه گردد و نشان دهید که بردار خطای $e_{k+1} = x^* - x_{k+1}$ بر d_k عمود است. توجه کنید که α_k نبایستی شامل مجهولات (مثلاً x^* یا e_k) باشد.

(ب) نشان دهید

$$\|e_{k+1}\|_2 \leq \|e_k\|_2 \sin \angle(e_k, d_k),$$

که در آن $\angle(e_k, d_k)$ زاویه بین e_k و d_k است.

(ج) به‌ازای هر حدس اولیه x_0 و زمانی که $f_k = r_k$ ، همگرایی روش را ثابت کنید.

روش‌های مبتنی بر زیرفضای کرایلف (قسمت ۱)

در فصل ۵ چند روش تصویری یک‌بُعدی برای حل دستگاه معادلات خطی ارائه شد. همان‌گونه که دیدیم انتخاب زیرفضاها در روش‌های تصویری بسیار با اهمیت است. در این فصل، نخست زیرفضای کرایلف را معرفی نموده و سپس ایده اصلی روش‌های تکراری مبتنی بر زیرفضای کرایلف را بیان کرده و در ادامه چند روش تکراری حاصل را معرفی می‌کنیم.

۱-۶ معرفی زیرفضای کرایلف

دستگاه معادلات خطی

$$Ax = b, \quad (1-6)$$

را در نظر بگیرید که در آن ماتریس $A \in \mathbb{R}^{n \times n}$ نامنفرد است، $b \in \mathbb{R}^n$ و $x \in \mathbb{R}^n$. فرض کنید p چندجمله‌ای مشخصه ماتریس A باشد. بنا به قضیه کیلی-همیلتون ۸.۱، داریم

$$p(A) = 0.$$

بنابراین، اگر

$$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0,$$

آنگاه

$$A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0I = 0.$$

از اینکه A نامنفرد است، داریم

$$a_{\circ} = (-1)^n \det(A) \neq 0. \quad (2-6)$$

لذا،

$$A^{-1} = -\frac{1}{a_{\circ}}(A^{n-1} + a_{n-1}A^{n-2} + \dots + a_1 I) = q_{n-1}(A), \quad (3-6)$$

که در آن

$$q_{n-1}(x) = -\frac{1}{a_{\circ}}(x^{n-1} + a_{n-1}x^{n-2} + \dots + a_1).$$

ازاین‌رو، اگر x^* جواب واقعی دستگاه (۱-۶) باشد، آنگاه

$$x^* = q_{n-1}(A)b \in \text{span}\{b, Ab, \dots, A^{n-1}b\}. \quad (4-6)$$

تعریف ۱.۶ به‌ازای هر $m \in \mathbb{N}$ و بردار $v \in \mathbb{R}^n$ ، زیرفضای کرایلف m -بُعدی $\mathcal{K}_m(A, v)$ به‌صورت

$$\mathcal{K}_m(A, v) = \text{span}\{v, Av, \dots, A^{m-1}v\},$$

تعریف می‌شود.

اکنون، با توجه به رابطه (۴-۶)، می‌بینیم که $x^* \in \mathcal{K}_n(A, b)$. بنابراین، برای محاسبه x^* می‌توان این بردار را در $\mathcal{K}_n(A, b)$ جست‌وجو کرد. فرض کنید $x_{\circ}$ یک حدس اولیه برای جواب دستگاه (۱-۶) باشد. دراین‌صورت، داریم $Ax_{\circ} = b - Ax_{\circ}$.

با فرض $r_{\circ} = b - Ax_{\circ}$ و $x - x_{\circ} = \delta$ ، دستگاه (۱-۶) معادل با

$$A\delta = r_{\circ}, \quad (5-6)$$

خواهد بود. لذا، اگر δ از معادله (۵-۶) محاسبه شود، آنگاه جواب واقعی دستگاه (۱-۶) به‌صورت $x = x_{\circ} + \delta$ خواهد بود. همان‌گونه که دیدیم $\delta \in \mathcal{K}_n(A, r_{\circ})$. اما به‌ازای مقادیر بزرگ n این جست‌جو بسیار پرهزینه (از نظر حجم محاسبات و اشغال حافظه) می‌باشد. بنابراین، سعی می‌کنیم یک جواب تقریبی برای دستگاه محاسبه کنیم. برای این کار، یک m کوچک (معمولاً $m \ll n$ ، مثلاً $m = 20$) انتخاب می‌شود و یک جواب تقریبی مثل δ_m در $\mathcal{K}_m(A, r_{\circ})$ جست‌وجو می‌گردد. این یعنی، به دنبال یک جواب تقریبی به‌صورت

$$x_m = x_{\circ} + \delta_m \in x_{\circ} + \mathcal{K}_m(A, r_{\circ}), \quad (6-6)$$

هستیم. اما برای تعیین x_m از دو روش تصویری کلی معرفی شده در فصل ۵ استفاده می‌کنیم. برای این کار، کفایت زیرفضای $\mathcal{L}$ را $\mathcal{K}_m(A, r_0)$ یا $AK_m(A, r_0)$ انتخاب کنیم. با انتخاب $\mathcal{L} = \mathcal{K}_m(A, r_0)$ ، الگوریتم متعامدسازی کامل (FOM) و $\mathcal{L} = AK_m(A, r_0)$ ، الگوریتم مانده مینیمال تعمیم یافته (GMRES) به دست می‌آید. این دو الگوریتم را در ادامه معرفی می‌کنیم. اما پیش از معرفی این دو روش، برخی از خواص زیرفضای کرایلف و الگوریتم محاسبه یک پایه برای این زیرفضا را معرفی می‌کنیم.

۲-۶ خواص زیرفضای کرایلف و محاسبه پایه برای آن

فرض $w \in \mathcal{K}_m(A, r_0)$. در این صورت، اعدادی مثل $\alpha_i \in \mathbb{R}$ ، $i = 0, 1, \dots, m$ وجود دارند به طوری که

$$w = \alpha_0 r_0 + \alpha_1 A r_0 + \dots + \alpha_{m-1} A^{m-1} r_0 = q_{m-1}(A) r_0,$$

که در آن

$$q_{m-1}(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_{m-1} x^{m-1}.$$

از این رو، اگر $w \in \mathcal{K}_m(A, r_0)$ و تنها اگر یک چندجمله‌ای، از درجه حداکثر $m-1$ مثل q_{m-1} وجود داشته باشد به طوری که $w = q_{m-1}(A) r_0$.

لم ۱.۶ اگر به ازای یک $j \geq 1$ داشته باشیم $A^j r_0 \in \mathcal{K}_j(A, r_0)$ ، آنگاه به ازای هر $k \geq j$ خواهیم داشت $A^k r_0 \in \mathcal{K}_j(A, r_0)$.

برهان. اگر $A^j r_0 \in \mathcal{K}_j(A, r_0)$ ، آنگاه یک چندجمله‌ای مثل q_{j-1} از درجه حداکثر $j-1$ وجود دارد به طوری که $A^j r_0 = q_{j-1}(A) r_0$. اکنون، داریم

$$A^{j+1} r_0 = A(A^j r_0) = A q_{j-1}(A) r_0 = q_j(A) r_0, \quad (7-6)$$

که در آن $q_j(t) = t q_{j-1}(t)$ یک چندجمله‌ای از درجه حداکثر j است. لذا، اگر چندجمله‌ای q_j را به صورت

$$q_j(t) = s_{j-1}(t) + a_j t^j,$$

نشان دهیم، آنگاه از رابطه (۷-۶) خواهیم داشت

$$A^{j+1} r_0 = s_{j-1}(A) r_0 + a_j A^j r_0 \in \mathcal{K}_j(A, r_0),$$

۲۰۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

زیرا $A^j r_0, s_{j-1}(A)r_0 \in \mathcal{K}_j(A, r_0)$ به‌همین ترتیب و با استقرا می‌توان نشان داد که به‌ازای هر $k \geq j$ داریم $A^k r_0 \in \mathcal{K}_j(A, r_0)$. $\square$

ملاحظه ۱.۶ لم ۱.۶ نشان می‌دهد که اگر به‌ازای یک j که $1 \leq j \leq n$ داشته باشیم $A^j r_0 \in \mathcal{K}_j(A, r_0)$ آنگاه

$$x^* - x_0 \in \mathcal{K}_j(A, r_0), \quad (۸-۶)$$

یا به عبارت دیگر، $x^* \in x_0 + \mathcal{K}_j(A, r_0)$.

تعریف ۲.۶ فرض کنید $A \in \mathbb{R}^{n \times n}$ و $v \in \mathbb{R}^n$. چندجمله‌ای مینیمال بردار v نسبت به A یک چندجمله‌ای تکین p با کوچک‌ترین درجه است به‌طوری‌که $p(A)v = 0$. درجه چندجمله‌ای مینیمال بردار v نسبت به ماتریس A ، گرید بردار v نسبت به A نامیده می‌شود و با $\text{grade}_A(v)$ نشان می‌دهند. اگر p چندجمله‌ای مینیمال A باشد، آنگاه $p(A) = 0$ و ازاین‌رو، $p(A)v = 0$. لذا،

$$\text{grade}_A(v) \leq \deg(p) \leq n.$$

قضیه ۱.۶ بُعد زیرفضای کرایلف $\mathcal{K}_m(A, v)$ برابر با m است اگر و تنها اگر گرید بردار v ، یعنی $\mu = \text{grade}_A(v)$ ، کوچک‌تر از m نباشد. به عبارت دیگر،

$$\dim(\mathcal{K}_m) = m \iff \text{grade}_A(v) \geq m.$$

ازاین‌رو،

$$\dim(\mathcal{K}_m) = \min\{m, \text{grade}_A(v)\}.$$

برهان. داریم

$$\begin{aligned} \dim(\mathcal{K}_m) = m &\iff Av, v, \dots, v, A^{m-1}v \text{ مستقل خطی باشند} \\ &\iff \sum_{i=0}^{m-1} \alpha_i A^i v \neq 0, \quad \forall (\alpha_0, \alpha_1, \dots, \alpha_{m-1}) \neq 0 \\ &\iff q_{m-1}(A)v \neq 0, \quad \forall q_{m-1} \in \Pi_{m-1} \\ &\iff \text{grade}_A(v) \geq m, \end{aligned}$$

که در آن Π_{m-1} مجموعه چندجمله‌ای‌های از درجه حداکثر $m-1$ می‌باشند. برای اثبات قسمت دوم، اگر $\text{grade}_A(v) \geq m$ ، آنگاه با توجه به قسمت اول قضیه خواهیم داشت

$$\dim(\mathcal{K}_m) = m = \min\{m, \text{grade}_A(v)\}.$$

اما اگر $\mu = \text{grade}_A(v) < m$ ، آنگاه یک چندجمله‌ای از درجه μ وجود دارد به طوری که $p_\mu(A)v = 0$ و به ازای همه چندجمله‌ای‌هایی مثل q_k که $k \leq \mu-1$ ، داریم $q_k(A)v \neq 0$. بنابراین، مجموعه بردارهای $\{v, Av, A^2v, \dots, A^{\mu-1}v\}$ وابسته خطی و مجموعه بردارهای $\{v, Av, A^2v, \dots, A^{\mu-1}v\}$ مستقل خطی هستند. با توجه به لم ۱۰۶، این نتیجه نشان می‌دهد که $\dim(\mathcal{K}_m(A, v)) = \mu$ و بنابراین،

$$\dim(\mathcal{K}_m) = \mu = \text{grade}_A(v) = \min\{m, \text{grade}_A(v)\}.$$

□

بدین ترتیب اثبات قضیه کامل می‌شود.

مثال ۱۰۶ ماتریس نامنفرد

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & 2 \end{pmatrix},$$

را در نظر بگیرید. فرض کنید e_i ، $i = 1, 2, 3$ ، ستون i ام ماتریس همانی 3×3 باشد. در این صورت، داریم

$$\begin{aligned} \mathcal{K}_3(A, e_1) &= \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix} \right\}, & \dim(\mathcal{K}_3(A, e_1)) &= 3, \\ \mathcal{K}_3(A, e_2) &= \text{span} \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 7 \\ 8 \end{pmatrix} \right\}, & \dim(\mathcal{K}_3(A, e_2)) &= 2, \\ \mathcal{K}_3(A, e_3) &= \text{span} \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 6 \end{pmatrix} \right\}, & \dim(\mathcal{K}_3(A, e_3)) &= 3. \end{aligned}$$

۲۰۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

همان‌گونه که می‌بینیم انتخاب بردار v روی بُعد فضای کرایلف $\mathcal{K}_3(A, v)$ اثر دارد. توجه کنید که چندجمله‌ای مشخصه و مینیمال ماتریس A برابر با

$$p_3(x) = x^3 - 6x^2 + 9x - 2,$$

است، درحالی‌که چندجمله‌ای مینیمال بردار e_2 نسبت به ماتریس A برابر با

$$q(x) = x^2 - 4x + 1,$$

می‌باشد. درحقیقت، $\text{grade}_A(e_2) = 2$.

اکنون، ببینیم که چگونه می‌توان یک پایهٔ یکامتعامد برای زیرفضای $\mathcal{K}_m(A, r_0)$ یافت. برای این کار، از الگوریتم گرام-اشمیت استفاده می‌کنیم. ابتدا این الگوریتم را مرور می‌کنیم.

فرض کنید $B_m = \{x_0, x_1, \dots, x_{m-1}\}$ یک مجموعه از بردارهای مستقل خطی در $\mathbb{C}^n$ باشد. الگوریتم گرام-اشمیت یک مجموعهٔ یکامتعامد $\mathcal{O}_m = \{v_1, v_2, \dots, v_m\}$ تولید می‌کند به‌طوری‌که $\text{span}(\mathcal{O}_m) = \text{span}(B_m)$. برای این کار، دنباله‌ای مثل $\mathcal{O}_k = \{v_1, v_2, \dots, v_k\}$ ، $k = 1, 2, \dots, m$ به‌عنوان یک پایهٔ یکامتعامد برای $\mathcal{S}_k = \text{span}(B_k)$ ، $k = 1, 2, \dots, m$ قرار می‌دهیم که در آن $B_k = \{x_0, x_1, \dots, x_{k-1}\}$. به‌ازای $k = 1$ ، قرار می‌دهیم

$$v_1 = \frac{x_0}{\|x_0\|_2}.$$

دراین صورت، به‌وضوح داریم

$$\mathcal{S}_1 = \text{span}\{x_0\} = \text{span}\{v_1\}.$$

ازاین‌رو، $\mathcal{O}_1 = \{v_1\}$. فرض کنید $\mathcal{O}_k$ یک پایهٔ یکامتعامد برای $\mathcal{S}_k = \text{span}(B_k)$ باشد. بردار v_{k+1} را طوری می‌سازیم که مجموعهٔ $\mathcal{O}_{k+1}$ یک پایهٔ یکامتعامد برای $\mathcal{S}_{k+1}$ باشد. برای این منظور، باید داشته باشیم

$$x_k \in \text{span}\{v_1, v_2, \dots, v_{k+1}\}.$$

با نوشتن بسط فوریهٔ بردار x_k برحسب پایهٔ یکامتعامد $\text{span}\{v_1, v_2, \dots, v_{k+1}\}$ خواهیم داشت

$$x_k = \sum_{i=1}^{k+1} \langle x_k, v_i \rangle v_i.$$

بنابراین،

$$\langle x_k, v_{k+1} \rangle v_{k+1} = x_k - \sum_{i=1}^k \langle x_k, v_i \rangle v_i =: w_{k+1}.$$

حال با توجه به اینکه باید داشته باشیم $\|v_{k+1}\|_2 = 1$ ، می‌توانیم قرار دهیم

$$v_{k+1} = \frac{w_{k+1}}{\|w_{k+1}\|_2}.$$

با جمع‌بندی این مطالب الگوریتم گرام-اشمیت برای محاسبه پایه $\mathcal{O}_k$ را می‌توان به صورت الگوریتم ۱.۶ نوشت.

الگوریتم ۱.۶ گرام-اشمیت

1. Set $\beta = \|x_0\|_2$ and compute $v_1 = x_0/\beta$
2. For $j = 1, \dots, m-1$, Do:
 3. Compute $h_{ij} = \langle x_j, v_i \rangle$ for $i = 1, 2, \dots, j$
 4. $w_j := x_j - \sum_{i=1}^j h_{ij} v_i$
 5. $h_{j+1,j} = \|w_j\|_2$
 6. $v_{j+1} := w_j / h_{j+1,j}$
7. EndDo

اگر الگوریتم ۱.۶ شکست نخورد، یعنی در گام ۵ مقادیر $h_{j+1,j}$ ، $j = 2, 3, \dots, m$ ، هیچ‌گاه صفر نشوند، آنگاه مجموعه $\{v_1, v_2, \dots, v_m\}$ یک مجموعه یکامتعامل است و

$$\text{span}\{x_0, x_1, \dots, x_{m-1}\} = \text{span}\{v_1, v_2, \dots, v_m\}.$$

اگر به ازای یک j ، داشته باشیم $h_{j+1,j} = 0$ ، آنگاه $w_j = 0$ و از گام ۴ الگوریتم، خواهیم داشت

$$x_j = \sum_{i=1}^j h_{i,j} v_i \in \text{span}\{v_1, v_2, \dots, v_j\} = \text{span}\{x_0, x_1, \dots, x_{j-1}\}.$$

بنابراین، بردار x_j ، به بردارهای $x_0, x_1, \dots, x_{j-1}$ وابسته است و نمی‌تواند در پایه باشد.

اکنون، اگر بخواهیم از الگوریتم ۱.۶ برای محاسبه یک پایه یکامتعامل برای $\mathcal{K}_m(A, r_0)$ ، کافیت در سراسر الگوریتم به جای بردار x_j ، بردار $A^j r_0$ را جایگزین کنیم. اما آرنودی پیشنهاد می‌کند که به جای بردار x_j ، $j = 1, 2, \dots, m-1$ در الگوریتم ۱.۶ بردار Av_j را قرار دهیم. در این صورت، الگوریتم

۲۱۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

گرام-اشمیت برای محاسبه یک پایه یکامتعامد برای $\mathcal{K}_m(A, r_0)$ به صورت الگوریتم ۲.۶ خواهد بود. این الگوریتم به خاطر ایده به کاررفته آرنولدی به الگوریتم گرام-اشمیت آرنولدی معروف شده است.

الگوریتم ۲.۶ گرام-اشمیت آرنولدی

1. Set $\beta = \|r_0\|_2$ and $v_1 = r_0/\beta$
2. For $j = 1, \dots, m$ Do:
3. Compute $h_{ij} = \langle Av_j, v_i \rangle$ for $i = 1, 2, \dots, j$
4. Compute $w_j = Av_j - \sum_{i=1}^j h_{ij}v_i$
5. $h_{j+1,j} := \|w_j\|_2$
6. If $h_{j+1,j} = 0$, then Stop
7. $v_{j+1} := w_j/h_{j+1,j}$
8. EndDo

در اینجا ذکر این نکته ضروری است که الگوریتم ۲.۶ علاوه بر مجموعه بردارهای $\{v_1, v_2, \dots, v_m\}$ بردار v_{m+1} را نیز محاسبه می‌کند که نقش مهمی در روش‌های مبتنی بر زیرفضای کرایلف دارد.

قضیه ۲.۶ اگر الگوریتم ۲.۶ پیش از گام m ام متوقف نشود، آنگاه مجموعه بردارهای $\mathcal{B} = \{v_1, v_2, \dots, v_m\}$ یک پایه یکامتعامد برای فضای کرایلف $\mathcal{K}_m(A, r_0)$ تشکیل می‌دهد. برهان. با توجه به نحوه ساختن v_j ها، مجموعه $\mathcal{B}$ یکامتعامد است. نشان می‌دهیم مجموعه $\mathcal{B}$ ، فضای کرایلف $\mathcal{K}_m(A, r_0)$ را تولید می‌کند. هر عضو $\delta \in \mathcal{K}_m(A, r_0)$ را می‌توان به صورت

$$\delta = q_{m-1}(A)r_0,$$

نمایش داد که در آن $q_{m-1}(t)$ یک چندجمله‌ای از درجه حداکثر $m-1$ است. لذا، کافیت نشان دهیم که هر بردار v_j را می‌توان به صورت $q_{j-1}(A)r_0$ نمایش داد که در آن q_{j-1} یک چندجمله‌ای از درجه حداکثر $j-1$ است. این موضوع را به کمک استقرا روی j نشان می‌دهیم.

از آنجایی که $v_1 = r_0/\beta$ ، با فرض $q_0(t) = 1/\beta$ ، نتیجه به‌ازای $j=1$ درست است؛ چرا که داریم

$$q_0(A)r_0 = (I/\beta)r_0 = r_0/\beta = v_1.$$

فرض کنید نتیجه برای همه اعداد صحیح کوچکتر یا مساوی j درست باشد. با استفاده از الگوریتم ۲.۶، داریم

$$\begin{aligned} h_{j+1,j}v_{j+1} &= Av_j - \sum_{i=1}^j h_{ij}v_i \\ &= Aq_{j-1}(A)r_0 - \sum_{i=1}^j h_{ij}q_{i-1}(A)r_0 \\ &= \tilde{q}_j(A)r_0 - \tilde{q}_{j-1}(A)r_0 \\ &= (\tilde{q}_j(A) - \tilde{q}_{j-1}(A))r_0, \end{aligned}$$

که در آن $\tilde{q}_j(t) = tq_{j-1}(t)$ و $\tilde{q}_{j-1}(t) = \sum_{i=1}^j h_{ij}q_{i-1}(t)$. روشن است که چندجمله‌ای‌های $\tilde{q}_j$ و $\tilde{q}_{j-1}$ به ترتیب از درجه‌های حداکثر j و $j-1$ هستند. از این رو، می‌بینیم که می‌توان بردار v_{j+1} را به صورت $q_j(A)r_0$ نشان داد که در آن $q_j = \tilde{q}_j - \tilde{q}_{j-1}$ یک چندجمله‌ای از درجه حداکثر j است. $\square$

قضیه ۳.۶ اگر $V_m = (v_1, v_2, \dots, v_m)$ و ماتریس بالاهسنبرگی $(m+1) \times m$ با درایه‌های h_{ij} تعریف شده در الگوریتم ۲.۶ را با $\bar{H}_m$ و ماتریس حاصل از حذف سطر آخر $\bar{H}_m$ را با H_m نشان دهیم، آنگاه روابط

$$AV_m = V_m H_m + w_m e_m^T \quad (9-6)$$

$$= V_{m+1} \bar{H}_m, \quad (10-6)$$

$$V_m^T AV_m = H_m, \quad (11-6)$$

برقرارند، که در آن $V_{m+1} = (V_m, v_{m+1})$.

برهان. با توجه به گام‌های ۴ و ۷ الگوریتم ۲.۶، داریم

$$h_{j+1,j}v_{j+1} = w_j = Av_j - \sum_{i=1}^j h_{ij}v_i,$$

و در نتیجه،

$$Av_j = \sum_{i=1}^j h_{ij}v_i + h_{j+1,j}v_{j+1} = \sum_{i=1}^{j+1} h_{ij}v_i, \quad j = 1, \dots, m.$$

لذا،

$$\begin{cases} Av_1 &= h_{11}v_1 + h_{21}v_2, \\ Av_2 &= h_{12}v_1 + h_{22}v_2 + h_{32}v_3, \\ &\vdots \\ Av_{m-1} &= h_{1,m-1}v_1 + h_{2,m-1}v_2 + h_{3,m-1}v_3 + \cdots + h_{m,m-1}v_m, \\ Av_m &= h_{1m}v_1 + h_{2m}v_2 + h_{3m}v_3 + \cdots + h_{m+1,m}v_{m+1}. \end{cases}$$

اگر این روابط را به شکل ماتریسی بنویسیم، خواهیم داشت

$$\begin{aligned} AV_m &= A(v_1, v_2, \dots, v_m) \\ &= (Av_1, Av_2, \dots, Av_m) \\ &= (v_1, v_2, \dots, v_m, v_{m+1}) \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1,m-1} & h_{1m} \\ h_{21} & h_{22} & \cdots & h_{2,m-1} & h_{2m} \\ \circ & h_{32} & \ddots & h_{3,m-1} & h_{3m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \circ & \circ & \cdots & h_{m,m-1} & h_{mm} \\ \circ & \circ & \cdots & \circ & h_{m+1,m} \end{pmatrix} \\ &= V_{m+1} \bar{H}_m. \end{aligned}$$

از طرفی، با توجه به رابطه به دست آمده، داریم

$$\begin{aligned} AV_m &= \begin{pmatrix} H_m \\ h_{m+1,m} e_m^T \end{pmatrix} \\ &= V_m H_m + h_{m+1,m} v_{m+1} e_m^T \\ &= V_m H_m + w_m e_m^T. \end{aligned}$$

سرانجام، از اینکه $V_m^T V_m = I$ و $V_m^T w_m = \circ$ می‌بینیم که

$$V_m^T AV_m = V_m^T V_m H_m + V_m^T w_m e_m^T = H_m,$$

□

که اثبات قضیه را کامل می‌کند.

$$A \quad V_m \quad = \quad V_m \quad \begin{matrix} H_m \\ \boxed{\begin{matrix} \diagup \end{matrix}} \end{matrix} + w_m e_m^T$$

شکل ۱-۶ ماتریس AV_m برابر است با $V_m H_m$ بعلاوه یک ماتریس با رتبه ۱.

قضیه ۴.۶ الگوریتم گرام-اشمیت آرنولدی در گام j ام شکست می‌خورد، یعنی $h_{j+1,j} = 0$ ، اگر و تنها اگر $\text{grade}_A(r_0) = j$.

برهان. فرض کنید $\text{grade}_A(r_0) = j$. به‌برهان خلف، فرض کنید $h_{j+1,j} \neq 0$. در این صورت، بردار v_{j+1} ساخته می‌شود. لذا، $\dim(\mathcal{K}_{j+1}) = j + 1$. از این رو، با استفاده از قضیه ۱.۶ خواهیم داشت $\text{grade}_A(r_0) \geq j + 1$ که یک تناقض است. برعکس، فرض کنید $h_{j+1,j} = 0$. در این صورت، $w_j = 0$ و در نتیجه،

$$Av_j - \sum_{i=1}^j h_{ij} v_i = 0 \implies Av_j \in \text{span}\{v_1, v_2, \dots, v_j\}.$$

این رابطه نشان می‌دهد که $\{r_0, Ar_0, \dots, A^j r_0\}$ یک مجموعه وابسته خطی است. یا به عبارت دیگر، یک چندجمله‌ای ناصفر مثل q_j از درجه حداکثر j وجود دارد، به طوری که

$$q_j(A)r_0 = 0.$$

این نتیجه نشان می‌دهد که $\text{grade}_A(r_0) \leq j$. اما حالت $\text{grade}_A(r_0) < j$ ناممکن است؛ زیرا در این صورت، الگوریتم پیش از گام j متوقف می‌شد. $\square$

قضیه ۴.۶ نشان می‌دهد که در حالت $h_{j+1,j} = 0$ ، الگوریتم متوقف می‌شود، اما در این حالت بنا به یادداشت ۱.۶ خواهیم داشت

$$x^* \in x_0 + \mathcal{K}_j(A, r_0),$$

که این خود یک موفقیت است. توجه کنید که اگر j عدد کوچکی باشد، آنگاه $x_0 + \mathcal{K}_j(A, r_0)$ یک زیرفضای کوچکی خواهد بود و می‌توانیم جواب واقعی مسئله را در یک فضای کوچک جستجو کنیم. در حالت کلی، الگوریتم گرام-اشمیت آرنولدی پایدار نیست. لازم به ذکر است که یک الگوریتم برای حل یک مسئله پایدار گفته می‌شود هرگاه یک خطای کوچک در یکی از مراحل حل مسئله به‌طور اساسی رشد نکرده و در دقت نتیجه نهایی تأثیر زیادی نگذارد. اما می‌توان این الگوریتم را به‌صورت الگوریتم ۳.۶ اصلاح نمود که نسبت به الگوریتم گرام-اشمیت آرنولدی پایدارتر باشد. می‌توان دید که این دو الگوریتم با یکدیگر معادل هستند.

الگوریتم ۳.۶ گرام-اشمیت آرنولدی اصلاح‌شده

1. Set $\beta = \|r_0\|_2$ and $v_1 = r_0/\beta$
2. For $j = 1, \dots, m$ Do:
3. Compute $w_j = Av_j$
4. For $i = 1, \dots, j$ Do:
5. $h_{ij} := \langle w_j, v_i \rangle$
6. $w_j := w_j - h_{ij}v_i$
7. EndDo
8. $h_{j+1,j} := \|w_j\|_2$. If $h_{j+1,j} = 0$, Stop.
9. $v_{j+1} = w_j/h_{j+1,j}$
10. EndDo

۳-۶ الگوریتم متعامدسازی کامل

همان‌گونه که پیش‌تر ذکر کردیم، در FOM یک جواب تقریبی به صورت

$$x_m = x_0 + \delta_m \in x_0 + \mathcal{K}_m(A, r_0), \quad (۱۲-۶)$$

جستجو می‌کنیم به‌طوری‌که

$$b - Ax_m \perp \mathcal{K}_m(A, r_0). \quad (۱۳-۶)$$

فرض کنید V_m ماتریس یک پایهٔ یکامتعامد برای $\mathcal{K}_m(A, r_0)$ باشد که توسط الگوریتم گرام-اشمیت آرنولدی محاسبه شده است. در این صورت،

$$x_m = x_0 + \delta_m = x_0 + V_m y_m,$$

که در آن $y_m \in \mathbb{R}^m$ داریم

$$b - Ax_m = b - A(x_0 + V_m y_m) = r_0 - AV_m y_m.$$

لذا، با اعمال شرط تعامد (۱۳-۶)، خواهیم داشت

$$r_0 - AV_m y_m \perp \mathcal{K}_m(A, r_0),$$

که معادل است با $V_m^T(r_0 - AV_m y_m) = 0$ ، که خود معادل است با

$$V_m^T AV_m y_m = V_m^T r_0. \quad (۱۴-۶)$$

اکنون، با استفاده از قضیه ۳.۶، داریم $V_m^T AV_m = H_m$. از طرفی، داریم $v_1 = r_0/\beta$ به‌طوری‌که $\beta = \|r_0\|_2$. لذا، $r_0 = \beta v_1$ و $V_m^T r_0 = V_m^T(\beta v_1) = \beta e_1$ ، که در آن e_1 ستون اول ماتریس همانی $m \times m$ است. بنابراین، می‌توان رابطه (۱۴-۶) را به‌صورت $H_m y_m = \beta e_1$ نوشت. با محاسبه y_m خواهیم داشت $y_m = H_m^{-1}(\beta e_1)$. با جمع‌بندی مطالب ارائه‌شده، دو گام اصلی FOM به‌صورت

$$x_m = x_0 + V_m y_m, \quad y_m = H_m^{-1}(\beta e_1),$$

خواهد بود. این روند در الگوریتم ۴.۶ خلاصه می‌شود.

الگوریتم ۴.۶ روش متعامدسازی کامل (FOM)

1. Compute $r_0 := b - Ax_0$, $\beta := \|r_0\|_2$, and $v_1 := r_0/\beta$.
2. Define the $m \times m$ matrix $H_m = \{h_{ij}\}_{i,j=1,\dots,m}$ set $H_m = 0$
3. For $j = 1, \dots, m$, Do:
 4. Compute $w_j := Av_j$
 5. For $i = 1, \dots, j$ Do:
 6. $h_{ij} := \langle w_j, v_i \rangle$
 7. $w_j := w_j - h_{ij}v_i$
 8. End Do
 9. $h_{j+1,j} := \|w_j\|_2$. If $h_{j+1,j} = 0$, set $m := j$ and Goto 12
 10. $v_{j+1} := w_j/h_{j+1,j}$.
11. EndDo
12. Compute $y_m = H_m^{-1}(\beta e_1)$ and $x_m = x_0 + V_m y_m$

۲۱۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

ملاحظه ۲.۶ هرگاه الگوریتم ۴.۶ در گام ۹ متوقف شود (یعنی، $h_{j+1,j} = 0$)، جواب دقیق دستگاه به‌دست خواهد آمد. این‌گونه توقف، به قطع موفق معروف است.

قضیه ۵.۶ فرض کنید x_m جواب تقریبی به‌دست آمده توسط الگوریتم ۴.۶ باشد. در این صورت، بردار مانده x_m در رابطه

$$b - Ax_m = -h_{m+1,m} e_m^T y_m v_{m+1},$$

صدق می‌کند و بنابراین،

$$\|b - Ax_m\|_2 = h_{m+1,m} |e_m^T y_m|.$$

برهان. با استفاده رابطه (۹-۶)، داریم

$$\begin{aligned} b - Ax_m &= b - A(x_* + V_m y_m) \\ &= r_* - AV_m y_m \\ &= \beta v_1 - (V_m H_m + w_m e_m^T) y_m \\ &= \beta v_1 - V_m H_m y_m - h_{m+1,m} e_m^T y_m v_{m+1}. \end{aligned} \quad (۱۵-۶)$$

اما از اینکه $H_m y_m = \beta e_1$ می‌بینیم که

$$\begin{aligned} \beta v_1 - V_m H_m y_m &= \beta v_1 - V_m (\beta e_1) \\ &= \beta v_1 - \beta v_1 = 0. \end{aligned}$$

□ با استفاده از این رابطه و رابطه (۱۵-۶)، نتیجه لازم به‌دست می‌آید.

در عمل ممکن است دقت جواب تقریبی x_m که توسط الگوریتم ۴.۶ محاسبه می‌شود، مناسب نباشد. برای محاسبه یک جواب دقیق‌تر می‌توان از دو نسخه مختلف FOM استفاده کرد: نسخه با شروع مجدد و نسخه کامل. این دو نسخه را در ادامه معرفی می‌کنیم.

۱-۳-۶ الگوریتم FOM با شروع مجدد

در نسخه با شروع مجدد الگوریتم FOM، با استفاده از یک زیرفضای کرایلف m -بُعدی و حدس اولیه x_* ، جواب تقریبی x_m را به‌دست می‌آوریم. سپس جواب تقریبی x_m را به‌عنوان حدس اولیه x_* جدید

قرار داده و الگوریتم را برای به‌دست آوردن x_m جدید، دوباره اجرا می‌کنیم. این روند را آن‌قدر تکرار می‌کنیم تا به یک جواب تقریبی با یک دقت مناسب دست پیدا کنیم. این روند در الگوریتم ۵.۶ جمع‌بندی می‌شود. در این الگوریتم و دیگر الگوریتم‌های مشابه که در ادامه معرفی می‌شوند منظور از “If satisfied” این است که جواب تقریبی محاسبه‌شده در شرط توقف صدق کند.

الگوریتم ۵.۶ FOM با شروع مجدد

1. Compute $r_0 := b - Ax_0$, $\beta := \|r_0\|_2$, and $v_1 := r_0/\beta$.
2. Generate the Arnoldi basis and the matrix H_m using the Arnoldi algorithm starting with v_1 .
3. Compute $y_m = H_m^{-1}(\beta e_1)$ and $x_m = x_0 + V_m y_m$. If satisfied, then Stop.
4. Set $x_0 := x_m$ and go to 1.

گاهی برای اینکه مشخص شود که بعد فضای کرایلف در نسخه با شروع مجدد چه عددی است، الگوریتم ۵.۶ را به صورت “FOM(m) با شروع مجدد” نیز نشان می‌دهند.

۲-۳-۶ الگوریتم FOM کامل

در تکرار k ام از نسخه کامل الگوریتم FOM، بُعد فضای کرایلف را k می‌گیریم. در واقع، در تکرار اول بُعد فضای کرایلف را یک گرفته و جواب تقریبی به‌دست‌آمده را به‌عنوان حدس اولیه برای تکرار بعدی می‌گیریم. سپس یک واحد به بُعد فضای کرایلف اضافه کرده و روند بالا را ادامه می‌دهیم. اگر

$$\dim(\mathcal{K}_n(A, r_0)) = s,$$

آنگاه به‌روشنی در تکرار s ام به جواب واقعی (در حساب دقیق) می‌رسیم. در عمل معمولاً s بزرگ است و روش پُرهزینه می‌باشد.

۴-۶ الگوریتم مانده مینیمال تعمیم‌یافته

در روش تکراری GMRES یک جواب تقریبی به‌صورت

$$x_m = x_0 + \delta_m \in x_0 + \mathcal{K}_m(A, r_0), \quad (۱۶-۶)$$

محاسبه می‌شود به‌طوری‌که

$$b - Ax_m \perp \mathcal{L} = A\mathcal{K}_m(A, r_0). \quad (۱۷-۶)$$

۲۱۸ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

همانند الگوریتم FOM، اگر V_m یک ماتریس پایهٔ یکامتعامد برای $\mathcal{K}_m(A, r_0)$ باشد، آنگاه یک بردار مثل $y_m \in \mathbb{R}^m$ وجود دارد به‌طوری‌که

$$x_m = x_0 + \delta_m = x_0 + V_m y_m$$

مشابه با الگوریتم FOM، داریم

$$b - Ax_m = r_0 - AV_m y_m.$$

لذا، با اعمال شرط تعامد (۶-۱۷)، خواهیم داشت

$$r_0 - AV_m y_m \perp A\mathcal{K}_m(A, r_0),$$

که معادل است با

$$(AV_m)^T(r_0 - AV_m y_m) = 0,$$

که معادل است با

$$(AV_m)^T(AV_m)y_m = (AV_m)^T r_0. \quad (۶-۱۸)$$

اکنون، با استفاده از قضیهٔ ۳.۶ داریم $AV_m = V_{m+1}\bar{H}_m$. لذا،

$$\begin{aligned} (AV_m)^T(AV_m) &= (V_{m+1}\bar{H}_m)^T(V_{m+1}\bar{H}_m) = \bar{H}_m^T \bar{H}_m, \\ (AV_m)^T r_0 &= (V_{m+1}\bar{H}_m)^T r_0 = \bar{H}_m^T V_{m+1}^T r_0 = \bar{H}_m^T(\beta e_1). \end{aligned}$$

با جایگذاری این دو رابطه در رابطه (۶-۱۸)، خواهیم داشت

$$\bar{H}_m^T \bar{H}_m y_m = \bar{H}_m^T(\beta e_1),$$

که این خود معادل است با (قضیهٔ ۲۴.۱ را ببینید)

$$y_m = \arg \min_{y \in \mathbb{R}^m} \|\bar{H}_m y - \beta e_1\|_2$$

سعد و شولتز در مقالهٔ معروف [۸۸]، الگوریتم GMRES را به‌صورت زیر معرفی کردند. با استناد به قضیهٔ ۳.۵، جواب تقریبی

$$x_m = x_0 + \delta_m = x_0 + V_m y_m,$$

از یک روش تصویری مورب به‌دست می‌آید اگر و تنها اگر بردار x_m نرم-۲ بردار مانده را روی $x_0 + K_m(A, r_0)$ کمینه کند. بنابراین، با استفاده از قضیه ۳.۶، داریم

$$\begin{aligned} \|b - Ax_m\|_2 &= \|b - A(x_0 + V_m y_m)\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|b - A(x_0 + V_m y)\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|r_0 - AV_m y\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|\beta v_1 - AV_m y\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|\beta V_{m+1} e_1 - V_{m+1} \bar{H}_m y\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|V_{m+1}(\beta e_1 - \bar{H}_m y)\|_2 \\ &= \min_{y \in \mathbb{R}^m} \|\beta e_1 - \bar{H}_m y\|_2. \end{aligned}$$

که در آن $r_0 = b - Ax_0$ از این‌رو، برای محاسبه y_m کافیت مسئله

$$y_m = \arg \min_{y \in \mathbb{R}^m} \|\beta e_1 - \bar{H}_m y\|_2, \quad (۱۹-۶)$$

را حل کنیم. در نتیجه، الگوریتم GMRES را می‌توان به‌صورت الگوریتم ۶.۶ جمع‌بندی نمود.

الگوریتم ۶.۶ ماندهٔ مینیمال تعمیم یافته (GMRES)

1. Compute $r_0 = b - Ax_0$, $\beta := \|r_0\|_2$ and $v_1 := r_0/\beta$.
2. Define the matrix $\bar{H}_m = (h_{ij}) \in \mathbb{R}^{(m+1) \times m}$. Set $\bar{H}_m = 0$.
3. For $j = 1, \dots, m$, Do:
 4. Compute $w_j := Av_j$
 5. For $i = 1, \dots, j$, Do:
 6. $h_{ij} = \langle w_j, v_i \rangle$
 7. $w_j := w_j - h_{ij}v_i$
 8. EndDo
 9. Compute $h_{j+1,j} = \|w_j\|_2$. if $h_{j+1,j} = 0$ set $m := j$ and Goto 12
 10. $v_{j+1} = w_j/h_{j+1,j}$.
 11. EndDo
 12. Compute y_m the minimizer of $\|\beta e_1 - \bar{H}_m y\|_2$ and $x_m = x_0 + V_m y_m$.

ملاحظه ۳.۶ هرگاه الگوریتم ۶.۶ در گام ۹ متوقف شود، یعنی $h_{j+1,j} = 0$ ، جواب دقیق دستگاه به‌دست خواهد آمد. در واقع یک قطع موفق خواهیم داشت.

۲۲۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

همانند الگوریتم FOM دو نسخه مختلف برای الگوریتم GMRES وجود دارد، الگوریتم GMRES با شروع مجدد و الگوریتم GMRES کامل. در اینجا الگوریتم GMRES با شروع مجدد را ارائه می‌کنیم.

الگوریتم ۷.۶ GMRES با شروع مجدد

1. Compute $r_0 = b - Ax_0$, $\beta = \|r_0\|_2$, and $v_1 = r_0/\beta$
2. Generate the Arnoldi basis and the matrix $\bar{H}_m$ using the Arnoldi algorithm starting with v_1
3. Compute y_m which minimizes $\|\beta e_1 - \bar{H}_m y\|_2$ and $x_m = x_0 + V_m y_m$
4. If satisfied then stop, else set $x_0 := x_m$ and goto 1.

گاهی برای اینکه مشخص شود که بعد فضای کرایلف در نسخه با شروع مجدد GMRES چه عددی است، الگوریتم ۷.۶ را به صورت $\text{GMRES}(m)$ با شروع مجدد نیز نشان می‌دهند.

۵-۶ به کارگیری الگوریتم‌های FOM و GMRES

در این بخش ابتدا روش به کارگیری الگوریتم GMRES را ارائه کرده سپس بیان می‌کنیم که چگونه می‌توان با یک تغییر کوچک، این روش را برای الگوریتم FOM نیز به کار برد.

۱-۵-۶ به کارگیری الگوریتم GMRES

یک مشکل آشکار در الگوریتم ۶.۶، این است که این الگوریتم جواب تقریبی x_m را در هر گام به طور صریح فراهم نمی‌کند. همچنین تعیین زمان توقف آسان نیست. مسئله کمترین توان‌های دوم

$$\min_{y \in \mathbb{R}^m} \|\beta e_1 - \bar{H}_m y\|_2,$$

را می‌توان با استفاده از روش‌های مختلفی حل کرد. در این بخش ایده سعد و شولتز که در [۸۸] ارائه شده است را مرور می‌کنیم.

ماتریس‌های گایونز $\Omega_1, \Omega_2, \dots, \Omega_m$ را طوری می‌سازیم که

$$\Omega_m \Omega_{m-1} \cdots \Omega_1 \bar{H}_m = \begin{pmatrix} \bar{R}_m \\ 0 \end{pmatrix},$$

که در آن $\bar{R}_m$ یک ماتریس بالامثلی باشد. اگر تعریف کنیم

$$\bar{H}_m^{(j-1)} = \Omega_{j-1} \Omega_{j-2} \cdots \Omega_1 \bar{H}_m,$$

آنگاه Ω_j ماتریس گیونزی خواهد بود که درایه $h_{j+1,j}^{(j-1)}$ در ماتریس $\overline{H}_m^{(j-1)}$ را با استفاده از درایه $h_{j,j}^{(j-1)}$ صفر می‌کند. شکل کلی ماتریس گیونز Ω_j به صورت

$$\Omega_j = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & c_j & s_j & \\ & & -s_j & c_j & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix},$$

است که در آن $(\Omega_j)_{j+1,j+1} = c_j$, $(\Omega_j)_{j+1,j} = -s_j$, $(\Omega_j)_{j,j+1} = s_j$, $(\Omega_j)_{jj} = c_j$ و

$$s_j = \frac{h_{j+1,j}}{\sqrt{(h_{jj}^{(j-1)})^2 + h_{j+1,j}^2}}, \quad c_j = \frac{h_{jj}^{(j-1)}}{\sqrt{(h_{jj}^{(j-1)})^2 + h_{j+1,j}^2}}. \quad (۲۰-۶)$$

لازم به ذکر است که در عمل، نیازی به محاسبه صریح Ω_j ها نیست و کافی است که c_j و s_j ها محاسبه شوند. قرار می‌دهیم

$$\bar{g}_m = \Omega_m \Omega_{m-1} \cdots \Omega_1 (\beta e_1) = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \\ \gamma_{m+1} \end{pmatrix} = \begin{pmatrix} g_m \\ \gamma_{m+1} \end{pmatrix}.$$

دراین صورت، داریم

$$\begin{aligned} g(y) &= \|\beta e_1 - \overline{H}_m y\|_r^2 \\ &= \|\Omega_m \Omega_{m-1} \cdots \Omega_1 (\beta e_1 - \overline{H}_m y)\|_r^2 \\ &= \left\| \begin{pmatrix} g_m \\ \gamma_{m+1} \end{pmatrix} - \begin{pmatrix} \overline{R}_m \\ \circ \end{pmatrix} y \right\|_r^2 \\ &= \|g_m - \overline{R}_m y\|_r^2 + |\gamma_{m+1}|^2. \end{aligned}$$

بنابراین، کمینه $g(y)$ وقتی به دست می‌آید که $g_m - \bar{R}_m y = 0$ و در این صورت، خواهیم داشت

$$y = y_m = \bar{R}_m^{-1} g_m.$$

لذا، به عنوان یک نتیجه، داریم

$$\|b - Ax_m\|_2 = \|\beta e_1 - \bar{H}_m y\|_2 = |\gamma_{m+1}|.$$

بنابراین، از $|\gamma_{m+1}| < \epsilon$ می‌توان به عنوان یک شرط توقف استفاده کرد.

روند ارائه‌شده را برای حالت خاص $m = 5$ ، شرح می‌دهیم. به ازای $m = 5$ ، داریم

$$\bar{H}_5 = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} & h_{15} \\ h_{21} & h_{22} & h_{23} & h_{24} & h_{25} \\ & h_{32} & h_{33} & h_{34} & h_{35} \\ & & h_{43} & h_{44} & h_{45} \\ & & & h_{54} & h_{55} \\ & & & & h_{65} \end{pmatrix}, \quad \bar{g}_5 = \begin{pmatrix} \beta \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

در این صورت، با ضرب ماتریس بالاهسنبرگی $\bar{H}_5$ در

$$\Omega_1 = \begin{pmatrix} c_1 & s_1 & & & \\ -s_1 & c_1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \\ & & & & & 1 \end{pmatrix},$$

با

$$s_1 = \frac{h_{21}}{\sqrt{h_{11}^2 + h_{21}^2}}, \quad c_1 = \frac{h_{11}}{\sqrt{h_{11}^2 + h_{21}^2}},$$

ماتریس و بردار سمت راست

$$\overline{H}_\delta^{(1)} = \begin{pmatrix} h_{11}^{(1)} & h_{12}^{(1)} & h_{13}^{(1)} & h_{14}^{(1)} & h_{15}^{(1)} \\ & h_{22}^{(1)} & h_{23}^{(1)} & h_{24}^{(1)} & h_{25}^{(1)} \\ & & h_{33} & h_{34} & h_{35} \\ & & & h_{44} & h_{45} \\ & & & & h_{55} \\ & & & & & h_{65} \end{pmatrix}, \quad \overline{g}_1 = \begin{pmatrix} c_1 \beta \\ -s_1 \beta \\ \circ \\ \circ \\ \circ \\ \circ \end{pmatrix}.$$

به‌دست می‌آیند. اکنون، برای حذف h_{32} ، ماتریس $\overline{H}_\delta^{(1)}$ و بردار $\overline{g}_1$ را در ماتریس گیونز

$$\Omega_2 = \begin{pmatrix} 1 & & & & \\ & c_2 & s_2 & & \\ & -s_2 & c_2 & & \\ & & & 1 & \\ & & & & 1 \\ & & & & & 1 \end{pmatrix},$$

ضرب می‌کنیم، که در آن

$$s_2 = \frac{h_{32}}{\sqrt{(h_{22}^{(1)})^2 + h_{32}^2}}, \quad c_2 = \frac{h_{22}^{(1)}}{\sqrt{(h_{22}^{(1)})^2 + h_{32}^2}}.$$

این فرایند را ادامه می‌دهیم تا مسئله، به یک مسئله با ماتریس و بردار سمت راست

$$\overline{H}_\delta^{(\delta)} = \begin{pmatrix} h_{11}^{(\delta)} & h_{12}^{(\delta)} & h_{13}^{(\delta)} & h_{14}^{(\delta)} & h_{15}^{(\delta)} \\ & h_{22}^{(\delta)} & h_{23}^{(\delta)} & h_{24}^{(\delta)} & h_{25}^{(\delta)} \\ & & h_{33}^{(\delta)} & h_{34}^{(\delta)} & h_{35}^{(\delta)} \\ & & & h_{44}^{(\delta)} & h_{45}^{(\delta)} \\ & & & & h_{55}^{(\delta)} \\ & & & & & \circ \end{pmatrix}, \quad \overline{g}_\delta = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{pmatrix},$$

تبدیل شود.

۶-۵-۲ به‌کارگیری الگوریتم FOM

همان‌گونه که دیدیم بردار y_m در الگوریتم FOM از حل دستگاه $H_m y_m = \beta e_1$ به‌دست می‌آید که در آن $e_1 \in \mathbb{R}^m$. نشان می‌دهیم که با یک تغییر کوچک در شیوه به‌کارگیری الگوریتم GMRES، می‌توانیم آن را برای الگوریتم FOM نیز به‌کار بگیریم. دستگاه $\bar{H}_m y_m = \beta e_1$ را در نظر بگیرید که در آن $e_1 \in \mathbb{R}^{m+1}$. داریم

$$\Omega_{m-1} \cdots \Omega_1 \bar{H}_m y_m = \Omega_{m-1} \cdots \Omega_1 (\beta e_1),$$

که در آن $i = 1, 2, \dots, m-1$ همان ماتریس‌هایی هستند که در شیوه به‌کارگیری الگوریتم GMRES استفاده می‌شوند. این دستگاه معادل با دستگاهی به‌صورت

$$\begin{pmatrix} R_m \\ \circ \quad \cdots \quad \circ \quad h_{m+1,m} \end{pmatrix} y_m = \begin{pmatrix} g_m \\ \circ \end{pmatrix},$$

می‌باشد، که در آن R_m یک ماتریس بالامثلثی است. روشن است که بردار y_m در الگوریتم FOM از حل دستگاه بالامثلثی $R_m y_m = g_m$ به‌دست می‌آید. این نتیجه نشان می‌دهد که اگر در الگوریتم GMRES، بعد از به‌کارگیری ماتریس گیونز Ω_{m-1} ، سطر آخر دستگاه $\bar{H}_m y_m = \beta e_1$ را حذف و دستگاه حاصل را حل کنیم و از y_m حاصل به‌جای y_m به‌دست‌آمده در روش GMRES استفاده کنیم، جواب x_m در FOM نتیجه می‌شود.

۶-۶ همگرایی الگوریتم $\text{GMRES}(m)$ با شروع مجدد

در این بخش همگرایی الگوریتم مانده مینیمال تعمیم‌یافته با شروع مجدد را مطالعه می‌نماییم.

قضیه ۶.۶ فرض کنید $A \in \mathbb{R}^{n \times n}$ معین مثبت باشد ($A + A^T$ معین مثبت متقارن باشد).

در این صورت، به ازای هر $m \geq 1$ ، الگوریتم $\text{GMRES}(m)$ با شروع مجدد همگراست.

برهان. در هر تکرار از الگوریتم $\text{GMRES}(m)$ با شروع مجدد، فضای کرایلف $\mathcal{K}_m(A, r_0)$ شامل بردار r_0 است. با توجه به اینکه الگوریتم $\text{GMRES}(m)$ نرم مانده را در فضای کرایلف $\mathcal{K}_m(A, r_0)$ کمینه می‌کند، نتیجه می‌گیریم که برای ماتریس‌های معین مثبت، کاهش نرم مانده دست‌کم به اندازه روش مانده مینیمال می‌باشد. بنابراین، با توجه به همگرایی روش مانده مینیمال (قضیه ۷.۵) نتیجه می‌گیریم که الگوریتم $\text{GMRES}(m)$ نیز همگراست. $\square$

اکنون، همگرایی الگوریتم $\text{GMRES}(m)$ با شروع مجدد را در حالت کلی‌تر بررسی می‌کنیم.

لم ۲.۶ فرض کنید x_k و x_{k-1} جواب تقریبی محاسبه‌شده توسط الگوریتم $\text{GMRES}(m)$ با شروع مجدد به ترتیب در تکرار k ام و $(k-1)$ ام باشند. در این صورت،

$$x_k = x_{k-1} + q_m(A)r_{k-1},$$

که در آن q_m یک چندجمله‌ای از درجه حداکثر $m-1$ است و بردار مانده تکرار $(k-1)$ ام می‌باشد و

$$\|r_k\|_2 = \|(I - Aq_m(A))r_{k-1}\|_2 = \min_{q \in \mathbb{P}_{m-1}} \|(I - Aq(A))r_{k-1}\|_2. \quad (21-6)$$

که در آن $r_k = b - Ax_k$.

برهان. همان‌گونه که دیدیم، در تکرار k ام، الگوریتم $\text{GMRES}(m)$ با شروع مجدد، نرم-۲ بردار مانده را در زیرفضای $x_{k-1} + \mathcal{K}_m(A, r_{k-1})$ کمینه می‌کند. هر عضو از این زیرفضا به صورت

$$x = x_{k-1} + q(A)r_{k-1},$$

است که در آن q یک چندجمله‌ای از درجه حداکثر $m-1$ است. اما داریم

$$\begin{aligned} b - Ax &= b - A(x_{k-1} + q(A)r_{k-1}) \\ &= r_{k-1} - Aq(A)r_{k-1} \\ &= (I - Aq(A))r_{k-1}. \end{aligned}$$

این رابطه نشان می‌دهد که r_k در رابطه (۲۱-۶) صدق می‌کند. $\square$

قضیه ۷.۶ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ قطری‌شدنی باشد و $A = X\Lambda X^{-1}$ که در آن $\Lambda = \text{diag}_n(\lambda_1, \lambda_2, \dots, \lambda_n)$ قرار دهید

$$\epsilon^{(m)} = \min_{p \in \mathbb{P}_m, p(\circ) = 1} \max_{i=1, \dots, n} |p(\lambda_i)|.$$

اگر r_k و r_{k-1} بردارهای مانده مربوط به جواب‌های محاسبه‌شده توسط الگوریتم $\text{GMRES}(m)$ با شروع مجدد، در تکرار k ام و $(k-1)$ ام باشند، آنگاه خواهیم داشت

$$\|r_k\|_2 \leq \text{cond}_2(X) \epsilon^{(m)} \|r_{k-1}\|_2.$$

برهان. با توجه به رابطه (۶-۲۱)، داریم

$$(I - Aq(A))r_{k-1} = p(A)r_{k-1},$$

که در آن p یک چندجمله‌ای از درجه حداکثر m است به‌طوری‌که $p(\circ) = 1$. اما داریم

$$\begin{aligned} \|p(A)r_{k-1}\|_2 &= \|p(X\Lambda X^{-1})r_{k-1}\|_2 \\ &= \|Xp(\Lambda)X^{-1}r_{k-1}\|_2 \\ &\leq \|X\|_2\|X^{-1}\|_2\|p(\Lambda)\|_2\|r_{k-1}\|_2 \\ &= \text{cond}_2(X)\|p(\Lambda)\|_2\|r_{k-1}\|_2. \end{aligned}$$

از طرفی،

$$\|p(\Lambda)\|_2 = \max_{i=1,\dots,n} |p(\lambda_i)|.$$

بنابراین، از رابطه (۶-۲۱)، داریم

$$\begin{aligned} \|r_k\|_2 &= \min_{q \in \mathbb{P}_{m-1}} \|(I - Aq(A))r_{k-1}\|_2 \\ &= \min_{p \in \mathbb{P}_m, p(\circ)=1} \|p(A)r_{k-1}\|_2 \\ &\leq \min_{p \in \mathbb{P}_m, p(\circ)=1} \text{cond}_2(X)\|p(\Lambda)\|_2\|r_{k-1}\|_2 \\ &= \min_{p \in \mathbb{P}_m, p(\circ)=1} \text{cond}_2(X) \max_{i=1,\dots,n} |p(\lambda_i)| \|r_{k-1}\|_2 \\ &\leq \text{cond}_2(X) \min_{p \in \mathbb{P}_m, p(\circ)=1} \max_{i=1,\dots,n} |p(\lambda_i)| \|r_{k-1}\|_2 \\ &= \text{cond}_2(X)\epsilon^{(m)}\|r_{k-1}\|_2. \end{aligned}$$

□

بنابراین، اثبات قضیه کامل می‌شود.

این قضیه نشان می‌دهد اگر

$$\rho = \text{cond}_2(X)\epsilon^{(m)} < 1,$$

آنگاه همگرایی روش $\text{GMRES}(m)$ تضمین خواهد شد. در حالتی که ماتریس A نرمال باشد، به‌طور یکانی قطری‌شدنی خواهد بود (ماتریس X یکانی است). در این صورت، خواهیم داشت $\text{cond}_2(X) = 1$

و ρ به

$$\rho = \epsilon^{(m)} = \min_{p \in \mathbb{P}_m, p(\circ) = 1} \max_{i=1, \dots, n} |p(\lambda_i)|,$$

کاهش می‌یابد. درواقع، توزیع مقادیر ویژه ماتریس ضرایب، همگرایی روش GMRES را توصیف می‌کند.

۷-۶ الگوریتم GMRES در متلب

به‌کارگیری الگوریتم $\text{GMRES}(m)$ با شروع مجدد در متلب با آنچه که پیش‌تر ذکر شد، اندکی متفاوت است. در نسخه متلب، ابتدا جواب تقریبی در $x_0 + \mathcal{K}_1(A, r_0)$ جستجو می‌شود. اگر جواب تقریبی به‌دست‌آمده مناسب باشد الگوریتم متوقف می‌شود؛ درغیراین‌صورت، جواب تقریبی به‌دست‌آمده به‌عنوان حدس اولیه برای تکرار بعدی در نظر گرفته می‌شود و جواب تقریبی بعدی در $x_0 + \mathcal{K}_2(A, r_0)$ جستجو می‌گردد. این روند به‌همین ترتیب ادامه می‌یابد و در هر تکرار بُعد فضای کرایلف یک واحد افزایش می‌یابد تا به $x_0 + \mathcal{K}_m(A, r_0)$ برسیم. در صورتی که جواب مناسب نباشد، جواب تقریبی به‌دست‌آمده به‌عنوان حدس اولیه در نظر گرفته می‌شود و روند بالا تکرار می‌شود (به عبارت دیگر شروع مجدد می‌شود).

به نظر می‌رسد که نسخه $\text{GMRES}(m)$ با شروع مجدد در متلب هزینه بسیار زیادتری در مقایسه با الگوریتمی که پیش‌تر معرفی کردیم، متحمل می‌شود. اما درحقیقت این طور نیست، زیرا هر زمان که بُعد فضای کرایلف یک واحد افزایش پیدا می‌کند، یک بردار به مجموعه بردارهای پایه و یک سطر و ستون به ماتریس $\bar{H}_m$ اضافه می‌شوند که بایستی محاسبه شوند. در این صورت، برای حل مسئله کمترین توان‌های دوم متناظر، تنها ماتریس‌های گیونز پیشین را روی ستون آخر $\bar{H}_m$ و بردار سمت راست تکرار قبل اعمال می‌کنیم.

تابع `gmres` در متلب برای الگوریتم $\text{GMRES}(m)$ با شروع مجدد به همراه ورودی‌ها و خروجی‌های آن به صورت

`[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,[],[],x0)`

است. ورودی‌های تابع `gmres` عبارتند از

۱. `A`: ماتریس ضرایب دستگاه؛

۲. `b`: بردار سمت راست؛

۳. `m`: بُعد فضای کرایلف؛

۴. `tol`: شرط توقف به صورت $\|r_k\|_2 \leq \text{tol} \|r_0\|_2$ است؛

۵. `maxit`: حداکثر تعداد شروع مجدد؛

۶. `x0`: حدس اولیه؛

لازم به ذکر است که دو ورودی ششم و هفتم تابع gmres به صورت [] هستند. در این دو مکان، می‌توان از ماتریس پیش‌شرط‌ساز استفاده کرد که در فصل‌های بعد جزئیات آن ذکر می‌شود.

خروجی‌های تابع gmres عبارتند از

۱. x : جواب تقریبی به دست آمده؛

۲. flag : اگر flag=0، یعنی الگوریتم در حداکثر maxit شروع مجدد به جواب تقریبی با دقت خواسته شده همگرا شده است، اگر flag=1 یعنی روش در maxit شروع مجدد همگرا نشده است، مقدار flag=2 مربوط به ماتریس پیش‌شرط‌ساز می‌شود (در حال حاضر مورد بحث ما نمی‌باشد)، و اگر flag=3 در این صورت، ایستایی در کاهش مانده رخ داده است. ایستایی به این معنا است که از یک تکرار به بعد هیچ کاهشی در نرم مانده حاصل نمی‌شود.

۳. relres : این مقدار برابر با $\|r_k\|_2 / \|b\|_2$ است که در آن r_k بردار مانده مربوط به

جواب تقریبی به دست آمده x_k می‌باشد؛

۴. iter : تعداد شروع مجدد‌های لازم برای همگرایی؛

۵. resvec : برداری است که درایه‌های آن نرم-۲ بردار مانده هستند. مؤلفه اول بردار نرم-۲ مانده r_0 و بقیه مؤلفه‌ها نرم-۲ مانده سایر تکرارها هستند؛

با توجه به توضیحات ارائه شده درباره نسخه $GMRES(m)$ با شروع مجدد در متلب، متذکر می‌شویم که خروجی iter یک بردار دوتایی به صورت (iter(1), iter(2)) است به طوری که iter(1)-1 تعداد شروع مجدد‌های کامل است و مؤلفه دوم آن، یعنی iter(2)، بعد فضای کرایلف مربوط به آخرین شروع مجدد است که جواب دستگاه در آن جست‌وجو شده است. توجه کنید که در آخرین شروع مجدد ممکن است پیش از رسیدن به بعد فضای کرایلف m ، به جواب تقریبی خواسته شده دست یابیم. با این توضیحات، معمولاً تعداد تکرارهای روش $GMRES(m)$ با شروع مجدد به صورت

$$iters = (iter(1) - 1) \times m + iter(2),$$

گزارش می‌شود.

برای استفاده از تابع gmres متلب برای نسخه کامل GMRES، کافیسست به جای ورودی سوم، یعنی

m، عبارت [] را وارد کنیم، به عبارت دیگر از دستور

$$[x, flag, relres, iter, resvec] = gmres(A, b, [], tol, maxit, [], [], x0)$$

استفاده کنیم.

مثال ۲.۶ ماتریس ضرایب دستگاه حاصل از گسسته‌سازی مدل معرفی شده در بخش ۱-۸ به ازای $\sigma = 40$ و $\tau = -20$ را در نظر بگیرید. دستگاهی را در نظر می‌گیریم که ماتریس ضرایب آن A است و بردار سمت

جدول ۱-۶ نتایج عددی روش GMRES با شروع مجدد و کامل برای مثال ۲.۶.

n	$m = 20$		$m = 40$		نسخه کامل	
	$iters$	$relres$	$iters$	$relres$	$iters$	$relres$
۱۰۰۰۰	۲۹۲	$9.8e-8$	۳۶۵	$8.9e-8$	۲۲۲	$8.5e-8$
۲۲۵۰۰	۵۵۶	$9.7e-8$	۴۷۴	$9.7e-8$	۳۲۸	$9.8e-8$
۴۰۰۰۰	۹۱۹	$9.6e-8$	۵۵۷	$9.2e-8$	۴۳۴	$9.2e-8$
۶۲۵۰۰	۱۴۰۵	$9.9e-8$	۷۶۶	$9.9e-8$	۵۳۸	$9.7e-8$

راست آن به صورت

$$b = A(1, 1, \dots, 1)^T \in \mathbb{R}^n,$$

می‌باشد، که در آن بعد ماتریس برابر با $n = N^2$ است. قرار می‌دهیم

$$x_0 = 0, \quad tol = 10^{-7}, \quad maxit = 1000.$$

نتایج عددی به‌ازای دو مقدار $m = 20, 40$ و مقادیر $200, 250, 150, 100$ در جدول ۱-۶ ارائه شده‌اند. در این جدول مقادیر $iters$ و $relres$ گزارش شده‌اند. همان‌گونه که ملاحظه می‌شود روش تکراری $GMRES(m)$ با شروع مجدد برای حل این دستگاه همگرا شده است. در دو ستون آخر این جدول نیز نتایج عددی مربوط به نسخه کامل GMRES ارائه شده است. اگرچه تعداد تکرارهای نسخه کامل از نسخه با شروع مجدد کمتر است، اما باید توجه داشت که در نسخه کامل وقتی که تعداد تکرارها افزایش می‌یابد، زمان حل دستگاه نیز به شدت افزایش می‌یابد. دلیل آن این است که ابعاد ماتریس $\bar{H}_m$ و در نتیجه ابعاد مسئله کمترین توان‌های دوم به شدت افزایش می‌یابد.

۸-۶ الگوریتم لانکزوس و گرادیان مزدوج

در این بخش فرض می‌کنیم ماتریس A متقارن باشد. نشان می‌دهیم ماتریس H_m که در الگوریتم گرام-اشمیت آرنولدی تشکیل می‌شود یک ماتریس سه‌قطری است و این الگوریتم را می‌توان به شکل ساده‌تری نوشت که به نام الگوریتم لانکزوس معروف است. در این صورت، الگوریتم FOM نیز شکل ساده‌تری به خود خواهد گرفت. در صورتی که ماتریس A علاوه بر متقارن بودن معین مثبت متقارن نیز باشد، خواهیم دید که الگوریتم FOM، الگوریتم معروف گرادیان مزدوج (CG) که توسط هستینز و استایفل در سال ۱۹۵۲ ارائه شده است را نتیجه می‌دهد [۴۹].

۲۳۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

۱-۸-۶ الگوریتم لانکزوس

برای ارائه الگوریتم لانکزوس به لم زیر نیاز داریم.

لم ۳.۶ فرض کنید الگوریتم گرام-اشمیت آرنولدی برای ماتریس متقارن A به کار رود. در این صورت، h_{ij} هایی که توسط این الگوریتم تولید می‌شوند در رابطه

$$h_{ij} = 0, \quad 1 \leq i < j-1,$$

$$h_{j,j+1} = h_{j+1,j}, \quad j = 1, 2, \dots, m,$$

صدق می‌کنند. به عبارت دیگر، ماتریس H_m ای که در الگوریتم گرام-اشمیت آرنولدی تولید می‌شود یک ماتریس سه‌قطری متقارن است.

برهان. با توجه به قضیه ۳.۶، داریم $H_m = V_m^T A V_m$. لذا،

$$\begin{aligned} H_m^T &= (V_m^T A V_m)^T \\ &= V_m^T A^T V_m \\ &= V_m^T A V_m = H_m. \end{aligned}$$

اکنون، چون H_m یک ماتریس بالاهسنبرگی است، نتیجه می‌گیریم که H_m یک ماتریس سه‌قطری است. $\square$

برای سادگی در نمایش نمادها فرض می‌کنیم

$$\alpha_j = h_{jj}, \quad \theta_j = h_{j-1,j}.$$

همچنین، ماتریس H_m را با T_m نشان می‌دهیم. از این رو، ماتریس T_m به صورت

$$T_m = \begin{pmatrix} \alpha_1 & \theta_2 & & & \\ \theta_2 & \alpha_2 & \theta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \theta_{m-1} & \alpha_{m-1} & \theta_m \\ & & & \theta_m & \alpha_m \end{pmatrix},$$

نمایش داده می‌شود. با توجه به اینکه ماتریس T_m یک ماتریس سه قطری متقارن است، الگوریتم گرام-اشمیت آرنولد و در نتیجه، الگوریتم FOM شکل خاصی به خود می‌گیرد که در آینده برای به دست آوردن الگوریتم گرادیان مزدوج بسیار مفید خواهد بود.

نخست توجه کنید که داریم $\beta = \|r_0\|_2$ و

$$v_1 = r_0 / \beta.$$

فرض کنید بردارهای یکامتعامل $v_k, k = 1, 2, \dots, j$ محاسبه شده‌اند و می‌خواهیم بردار v_{j+1} را محاسبه کنیم. با استفاده از قضیه ۳.۶، داریم $AV_m = V_{m+1}\bar{T}_m$. با برابر قرار دادن ستون j ام دو طرف این رابطه خواهیم داشت

$$Av_j = \theta_j v_{j-1} + \alpha_j v_j + \theta_{j+1} v_{j+1}.$$

با توجه به ساختار T_m ، می‌بینیم که مقدار θ_j پیش‌تر محاسبه شده است. بنابراین، داریم

$$Av_j - \theta_j v_{j-1} = \alpha_j v_j + \theta_{j+1} v_{j+1}.$$

سمت چپ این رابطه کاملاً معلوم است. از این رو، با فرض

$$w_j = Av_j - \theta_j v_{j-1},$$

خواهیم داشت

$$w_j = \alpha_j v_j + \theta_{j+1} v_{j+1}. \quad (22-6)$$

اگر طرفین این رابطه در بردار v_j ضرب داخلی کنیم، آنگاه خواهیم داشت

$$\alpha_j = \langle w_j, v_j \rangle.$$

اکنون، قرار می‌دهیم

$$w_j := w_j - \alpha_j v_j.$$

در این صورت، با توجه به رابطه (۲۲-۶)، اگر قرار دهیم $\theta_{j+1} = \|w_j\|_2$ ، آنگاه در صورتی که $\theta_{j+1} \neq 0$ ، با قرار دادن

$$v_{j+1} = w_j / \theta_{j+1},$$

مجموعه بردارهای $\{v_1, v_2, \dots, v_{j+1}\}$ یک مجموعه یک‌معامد خواهد بود. با توجه به نتایج به‌دست‌آمده، الگوریتم لانکزوس که نسخه‌ای از الگوریتم گرام-اشمیت آرنولدی اصلاح‌شده برای ماتریس متقارن A است، به‌صورت الگوریتم ۸.۶ جمع‌بندی می‌شود. یادآوری می‌شود که این الگوریتم یک پایه یک‌معامد برای فضای کرایلف $K_m(A, r_0)$ محاسبه می‌کند.

الگوریتم ۸.۶ الگوریتم لانکزوس

1. Set $\beta = \|r_0\|_2$, $v_1 = r_0/\beta$, $\theta_1 = 0$ and $v_0 = 0$
2. For $j = 1, 2, \dots, m$, Do
3. $w_j := Av_j - \theta_j v_{j-1}$
4. $\alpha_j := \langle w_j, v_j \rangle$
5. $w_j := w_j - \alpha_j v_j$
6. $\theta_{j+1} := \|w_j\|_2$. If $\theta_{j+1} = 0$, then Stop
7. $v_{j+1} := w_j/\theta_{j+1}$
8. EndDo

بعد از محاسبه ماتریس پایه $V_m = (v_1, v_2, \dots, v_m)$ توسط الگوریتم لانکزوس، جواب تقریبی به‌دست‌آمده به‌صورت

$$x_m = x_0 + V_m y_m, \quad y_m = T_m^{-1}(\beta e_1),$$

خواهد بود. در این صورت، الگوریتم لانکزوس برای حل دستگاه $Ax = b$ به‌صورت الگوریتم ۹.۶ خواهد بود، که در آن A یک ماتریس متقارن است.

الگوریتم ۹.۶ الگوریتم لانکزوس برای حل دستگاه $Ax = b$

1. Compute $r_0 = b - Ax_0$, $\beta := \|r_0\|_2$, $v_1 := r_0/\beta$
2. For $j = 1, 2, \dots, m$, Do
3. $w_j := Av_j - \theta_j v_{j-1}$ (If $j = 1$ set $\theta_1 v_0 = 0$)
4. $\alpha_j := \langle w_j, v_j \rangle$
5. $w_j := w_j - \alpha_j v_j$
6. $\theta_{j+1} := \|w_j\|_2$. If $\theta_{j+1} = 0$, set $m := j$ and go to 9
7. $v_{j+1} := w_j/\theta_{j+1}$
8. EndDo
9. Set $T_m = \text{tridiag}_m(\theta_i, \alpha_i, \theta_{i+1})$, and $V_m = (v_1, v_2, \dots, v_m)$
10. Compute $y_m = T_m^{-1}(\beta e_1)$ and $x_m = x_0 + V_m y_m$

همانند الگوریتم FOM، در الگوریتم ۹.۶، داریم

$$r_m = b - Ax_m = -\theta_{m+1} e_m^T y_m v_{m+1}. \quad (۲۳-۶)$$

از گام ۹ الگوریتم ۹.۶ می‌بینیم که برای محاسبه x_m همه v_i ها $i = 1, 2, \dots, m$ بایستی ذخیره شوند. خوشبختانه، برای رفع این مشکل، نسخه‌ای از این الگوریتم به نام الگوریتم D-Lanczos وجود دارد که در آن جواب تقریبی فعلی x_m برحسب جواب محاسبه شده قبلی x_{m-1} به هنگام می‌شود و تعداد کمی از بردارها نیز در هر گام به هنگام می‌شوند. برای نوشتن این الگوریتم، تجزیه LU ماتریس T_m ، یعنی $T_m = L_m U_m$ ، را می‌نویسیم. در اینجا، ماتریس L_m یک ماتریس دوقطری پائین مثلثی با درایه‌های قطری ۱ و ماتریس U یک ماتریس دوقطری بالامثلثی است. بنابراین، تجزیه LU ماتریس T_m به ازای $m = 5$ به صورت

$$T_5 = \begin{pmatrix} 1 & & & & \\ \lambda_2 & 1 & & & \\ & \lambda_3 & 1 & & \\ & & \lambda_4 & 1 & \\ & & & \lambda_5 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 & \theta_2 & & & \\ & \eta_2 & \theta_3 & & \\ & & \eta_3 & \theta_4 & \\ & & & \eta_4 & \theta_5 \\ & & & & \eta_5 \end{pmatrix},$$

است. در این صورت، جواب تقریبی x_m به صورت

$$x_m = x_0 + V_m U_m^{-1} L_m^{-1} (\beta e_1),$$

خواهد بود. با فرض

$$P_m = V_m U_m^{-1}, \quad (۲۴-۶)$$

و

$$z_m = L_m^{-1} (\beta e_1), \quad (۲۵-۶)$$

خواهیم داشت

$$x_m = x_0 + P_m z_m. \quad (۲۶-۶)$$

۲۳۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

با استفاده از رابطه (۶-۲۴)، داریم

$$P_m U_m = V_m.$$

اگر ستون آخر دو طرف این رابطه را با هم برابر قرار دهیم، خواهیم داشت

$$p_m = \frac{1}{\eta_m} (v_m - \theta_m p_{m-1}).$$

این رابطه نشان می‌دهد که ستون آخر P_m را می‌توان برحسب ستون آخر V_m و ستون $(m-1)$ ام ماتریس P_m به‌هنگام کرد. اگر در رابطه $T_m = L_m U_m$ ، درایه‌های $(m, m-1)$ از دو طرف رابطه را با هم برابر قرار دهیم، خواهیم داشت

$$\lambda_m = \frac{\theta_m}{\eta_{m-1}},$$

و اگر درایه‌های (m, m) از دو طرف رابطه را با هم برابر قرار دهیم، می‌بینیم که

$$\eta_m = \alpha_m - \lambda_m \theta_m.$$

ازطرفی، فرض کنید

$$z_m = \begin{pmatrix} z_{m-1} \\ \zeta_m \end{pmatrix}.$$

دراین صورت، از رابطه (۶-۲۵) داریم $L_m z_m = \beta e_1$ ، که با برابر قرار دادن درایه‌های m ام از دو طرف رابطه، نتیجه می‌شود

$$\zeta_m = -\lambda_m \zeta_{m-1}.$$

همچنین از رابطه (۶-۲۶)، داریم

$$\begin{aligned} x_m &= x_0 + \begin{pmatrix} P_{m-1} & p_m \end{pmatrix} \begin{pmatrix} z_{m-1} \\ \zeta_m \end{pmatrix} \\ &= x_0 + P_{m-1} z_{m-1} + p_m \zeta_m \\ &= x_{m-1} + \zeta_m p_m. \end{aligned}$$

با جمع‌بندی این مطالب، الگوریتم D-Lanczos به‌دست می‌آید که به‌صورت الگوریتم ۱۰.۶ جمع‌بندی می‌شود.

الگوریتم ۱۰.۶ D-Lanczos

1. Compute $r_0 = b - Ax_0$, $\zeta_1 := \beta := \|r_0\|_2$, $v_1 := r_0/\beta$
2. Set $\lambda_1 = \theta_1 = 0$, $p_0 = 0$
3. For $m = 1, 2, \dots$, until convergence Do:
 4. $w := Av_m - \theta_m v_{m-1}$ and $\alpha_m = \langle w, v_m \rangle$
 5. If $m > 1$, then compute $\lambda_m = \theta_m/\eta_{m-1}$ and $\zeta_m = -\lambda_m \zeta_{m-1}$
 6. $\eta_m = \alpha_m - \lambda_m \theta_m$
 7. $p_m = (v_m - \theta_m p_{m-1})/\eta_m$
 8. $x_m := x_{m-1} + \zeta_m p_m$
 9. If x_m has converged then Stop
10. $w := w - \alpha_m v_m$
11. $\theta_{m+1} := \|w\|_2$ and $v_{m+1} = w/\theta_{m+1}$
12. EndDo

روشن است که این الگوریتم جواب دستگاه سه‌قطری $T_m y_m = \beta e_1$ را با استفاده از روش حذفی گاوس بدون محورگیری به‌دست می‌آورد. می‌توان این دستگاه را با استفاده از محورگیری جزئی نیز حل نمود که با هزینه ذخیره‌سازی یک بردار اضافی همراه است. الگوریتم پیچیده‌تر دیگری نیز وجود دارد که از تجزیه LQ استفاده می‌نماید و به الگوریتم SYMMLQ معروف است [۷۹].

۲-۸-۶ الگوریتم گرادیان مزدوج

در این زیربخش، نشان می‌دهیم زمانی که ماتریس A علاوه بر متقارن بودن، معین مثبت نیز باشد، می‌توان الگوریتم معروف گرادیان مزدوج (CG) را از الگوریتم D-Lanczos استخراج کرد. ابتدا توجه کنید که الگوریتم‌های ۹.۶ و ۱۰.۶ معادل یکدیگرند. بنابراین، با توجه به رابطه (۶-۲۳) می‌بینیم که بردار مانده $r_m = b - Ax_m$ هم‌راستا با بردار v_{m+1} است. از این رو، بردارهای مانده در الگوریتم D-Lanczos بر یکدیگر عمودند.

قضیه ۸.۶ فرض کنید ماتریس A معین مثبت متقارن باشد و $r_m = b - Ax_m$ ، $m = 0, 1, \dots$ بردارهای مانده و $p_m = 0, 1, \dots$ بردارهایی باشند که توسط الگوریتم D-Lanczos تولید می‌شوند. در این صورت،

(الف) هر بردار مانده r_m به صورت $r_m = \sigma_m v_{m+1}$ نوشته می‌شود که در آن σ_m یک اسکالر است. به عنوان یک نتیجه، بردارهای مانده دوبه‌دو متعامد هستند.

۲۳۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

(ب) بردارهای $p_i, i = 0, 1, \dots$ یک مجموعه A -متعامد می‌باشند؛ یعنی اگر $i \neq j$ ، آنگاه $(Ap_i, p_j) = 0$.

برهان. قسمت (الف) از رابطه (۲۳-۶) نتیجه می‌شود. برای اثبات قسمت دوم، قرار می‌دهیم

$$P_m = V_m U_m^{-1}.$$

در این صورت،

$$\begin{aligned} P_m^T A P_m &= U_m^{-T} V_m^T A V_m U_m^{-1} \\ &= U_m^{-T} T_m U_m^{-1} \\ &= U_m^{-T} L_m. \end{aligned}$$

اکنون، توجه کنید که ماتریس $U_m^{-T} L_m$ یک ماتریس پائین‌مثلثی و از طرف دیگر متقارن است؛ زیرا ماتریس $P_m^T A P_m$ متقارن است. بنابراین، ماتریس $U_m^{-T} L_m$ یک ماتریس قطری است که اثبات قسمت (ب) را کامل می‌کند. $\square$

اینک، با استفاده از اینکه بردارهای مانده بر هم عمود و p_j ها بر هم A -عمود هستند، الگوریتم گرادیان مزدوج را استخراج می‌کنیم. با توجه به گام ۷ از الگوریتم D-Lanczos، داریم

$$p_{j+1} = \frac{1}{\eta_{j+1}}(v_{j+1} - \theta_{j+1} p_j).$$

اما از قضیه ۸.۶، داریم $v_{j+1} = r_j / \sigma_j$. با جایگذاری در رابطه بالا، خواهیم داشت

$$\sigma_j \eta_{j+1} p_{j+1} = r_j - \theta_{j+1} \sigma_j p_j.$$

اکنون، با فرض $\hat{p}_j = \sigma_{j-1} \eta_j p_j$ ، خواهیم داشت

$$\hat{p}_{j+1} = r_j - \frac{\theta_{j+1} \sigma_j}{\sigma_{j-1} \eta_j} \hat{p}_j.$$

بنابراین، با فرض

$$\beta_j = -\frac{\theta_{j+1} \sigma_j}{\sigma_{j-1} \eta_j},$$

نتیجه می‌گیریم که

$$\hat{p}_{j+1} = r_j + \beta_j \hat{p}_j. \quad (۲۷-۶)$$

از طرفی، از گام ۸ الگوریتم D-Lanczos، داریم

$$\begin{aligned} x_j &= x_{j-1} + \xi_j p_j \\ &= x_{j-1} + \frac{\xi_j}{\sigma_{j-1} \eta_j} \hat{p}_j \\ &= x_{j-1} + \alpha_j \hat{p}_j, \end{aligned} \quad (28-6)$$

که در آن

$$\alpha_j = \frac{\xi_j}{\sigma_{j-1} \eta_j}.$$

با توجه به گام ۲ از الگوریتم D-Lanczos، داریم $p_{\circ} = p$. بنابراین، با توجه به تعریف $\hat{p}_{\circ}$ خواهیم داشت $\hat{p}_{\circ} = p$ و در نتیجه، از رابطه (۲۷-۶) داریم $p_1 = r$. در همه الگوریتم‌های مشابهی که تاکنون ارائه نموده‌ایم (مثل سریع‌ترین کاهش، مانده مینیمال، و ...) بردار جهت آغازین، یعنی $\hat{p}_{\circ}$ ، برداری است که متناظر با بردار مانده مربوط حدس اولیه تعریف می‌شود. از این رو، در الگوریتم گرادیان مزدوج نیز جهت اولیه را r اختیار می‌کنیم؛ یعنی قرار می‌دهیم $\hat{p}_{\circ} = r$. همچنین، برای یک‌دست بودن الگوریتم‌ها، با توجه به رابطه‌های (۲۷-۶) و (۲۸-۶)، بردارهای جهت و جواب‌های تقریبی را به صورت زیر می‌نویسیم

$$x_{j+1} = x_j + \alpha_j \hat{p}_j, \quad (29-6)$$

$$\hat{p}_{j+1} = r_{j+1} + \beta_j \hat{p}_j. \quad (30-6)$$

این دو رابطه به همراه A -عمود بودن بردارهای جهت و عمود بودن بردارهای مانده، اساس الگوریتم CG هستند. برای سادگی از اینجا به بعد علامت “ $\hat{\cdot}$ ” را هم می‌توانیم نادیده بگیریم. اکنون، به محاسبه α_j ها و β_j ها می‌پردازیم. با استفاده از رابطه (۲۹-۶)، بردارهای مانده در رابطه

$$\begin{aligned} r_{j+1} &= b - Ax_{j+1} \\ &= b - A(x_j + \alpha_j p_j) \\ &= r_j - \alpha_j Ap_j, \end{aligned} \quad (31-6)$$

صدق می‌کنند. از متعامد بودن بردارهای مانده، داریم

$$\circ = \langle r_{j+1}, r_j \rangle = \langle r_j - \alpha_j Ap_j, r_j \rangle,$$

که نتیجه می‌دهد

$$\alpha_j = \frac{\langle r_j, r_j \rangle}{\langle Ap_j, r_j \rangle}. \quad (۳۲-۶)$$

از طرفی، با توجه به اینکه $\langle Ap_j, p_{j-1} \rangle = 0$ ، داریم

$$\begin{aligned} \langle Ap_j, r_j \rangle &= \langle Ap_j, p_j - \beta_{j-1} p_{j-1} \rangle \\ &= \langle Ap_j, p_j \rangle - \beta_{j-1} \langle Ap_j, p_{j-1} \rangle \\ &= \langle Ap_j, p_j \rangle. \end{aligned}$$

در نتیجه، از رابطه (۳۲-۶) خواهیم داشت

$$\alpha_j = \frac{\langle r_j, r_j \rangle}{\langle Ap_j, p_j \rangle}.$$

از طرفی،

$$\begin{aligned} 0 &= \langle p_{j+1}, Ap_j \rangle \\ &= \langle r_{j+1} + \beta_j p_j, Ap_j \rangle \\ &= \langle r_{j+1}, Ap_j \rangle + \beta_j \langle p_j, Ap_j \rangle, \end{aligned}$$

و بنابراین،

$$\beta_j = -\frac{\langle r_{j+1}, Ap_j \rangle}{\langle p_j, Ap_j \rangle}. \quad (۳۳-۶)$$

اما با استفاده از رابطه (۳۱-۶)، داریم

$$Ap_j = -\frac{1}{\alpha_j} (r_{j+1} - r_j), \quad (۳۴-۶)$$

که با جایگذاری این رابطه در رابطه (۳۳-۶) و اینکه $\langle r_{j+1}, r_j \rangle = 0$ ، خواهیم داشت

$$\begin{aligned} \beta_j &= \frac{1}{\alpha_j} \frac{\langle r_{j+1}, (r_{j+1} - r_j) \rangle}{\langle Ap_j, p_j \rangle} \\ &= \frac{\langle Ap_j, p_j \rangle}{\langle r_j, r_j \rangle} \frac{\langle r_{j+1}, r_{j+1} \rangle}{\langle Ap_j, p_j \rangle} \\ &= \frac{\langle r_{j+1}, r_{j+1} \rangle}{\langle r_j, r_j \rangle}. \end{aligned}$$

با جمع‌بندی این مطالب الگوریتم گرادیان مزدوج (CG) را می‌توان به صورت الگوریتم ۱۱.۶ ارائه نمود.

الگوریتم ۱۱.۶ گرادیان مزدوج (CG)

1. Choose an initial guess x_0 and $r_0 := b - Ax_0$, $p_0 := r_0$
2. For $j = 0, 1, 2, \dots$, until convergence, Do
3. $\alpha_j := \langle r_j, r_j \rangle / \langle Ap_j, p_j \rangle$
4. $x_{j+1} := x_j + \alpha_j p_j$
5. $r_{j+1} := r_j - \alpha_j Ap_j$
6. $\beta_j := \langle r_{j+1}, r_{j+1} \rangle / \langle r_j, r_j \rangle$
7. $p_{j+1} := r_{j+1} + \beta_j p_j$
8. EndDo

همان‌گونه که ملاحظه می‌شود در هر تکرار از این الگوریتم تنها یک ضرب ماتریس در بردار لازم است. اکنون، یکی از قضایای مهم درباره همگرایی الگوریتم گرادیان مزدوج را ارائه می‌کنیم. برای این کار ابتدا لم بعدی را ذکر می‌کنیم.

لم ۴.۶ در الگوریتم گرادیان مزدوج، داریم

$$\langle r_k, p_j \rangle = 0, \quad j = 0, 1, \dots, k-1.$$

برهان. به استقرا این لم را ثابت می‌کنیم. به‌ازای $k=1$ ، داریم

$$\langle r_1, p_0 \rangle = \langle r_1, r_0 \rangle = 0.$$

اینک، فرض کنید حکم به‌ازای $k=m$ برقرار باشد. اگر $j \leq m-1$ ، آنگاه

$$\begin{aligned} \langle r_{m+1}, p_j \rangle &= \langle r_m - \alpha_m Ap_m, p_j \rangle \\ &= \langle r_m, p_j \rangle - \alpha_m \langle Ap_m, p_j \rangle = 0. \end{aligned}$$

اگر $j=m$ ، آنگاه

$$\begin{aligned} \langle r_{m+1}, p_m \rangle &= \langle r_m - \alpha_m Ap_m, p_m \rangle \\ &= \langle r_m, p_m \rangle - \frac{\langle r_m, r_m \rangle}{\langle Ap_m, p_m \rangle} \langle Ap_m, p_m \rangle \\ &= \langle r_m, p_m \rangle - \langle r_m, r_m \rangle. \end{aligned}$$

۲۴۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

اما از گام ۷ داریم $p_m = r_m + \beta_{m-1}p_{m-1}$ با جایگذاری p_m در رابطه بالا، خواهیم داشت

$$\begin{aligned}\langle r_{m+1}, p_m \rangle &= \langle r_m, p_m \rangle - \langle r_m, r_m \rangle \\ &= \langle r_m, r_m + \beta_{m-1}p_{m-1} \rangle - \langle r_m, r_m \rangle \\ &= \langle r_m, r_m \rangle + \beta_{m-1} \langle r_m, p_{m-1} \rangle - \langle r_m, r_m \rangle \\ &= \langle r_m, r_m \rangle - \langle r_m, r_m \rangle = 0.\end{aligned}$$

□ بدین ترتیب اثبات لم کامل می‌گردد.

قضیه ۹.۶ اگر ماتریس $A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن باشد، آنگاه در حساب دقیق الگوریتم گرادیان مزدوج حداکثر در n تکرار به جواب دستگاه همگرا می‌شود.

برهان. ابتدا توجه کنید که اگر به‌ازای یک $k \leq n-2$ ، داشته باشیم $p_k = 0$ ، آنگاه با توجه به گام ۵ الگوریتم ۱۱.۶، خواهیم داشت $r_{k+1} = r_k$. در نتیجه، با استفاده از عمود بودن r_{k+1} بر r_k ، داریم

$$\|r_k\|_2^2 = \langle r_k, r_k \rangle = \langle r_k, r_{k+1} \rangle = 0 \implies r_k = 0.$$

این رابطه نشان می‌دهد که $Ax_k = b$ اکنون، اگر $i = 0, 1, \dots, n-1$ ، آنگاه با توجه به اینکه این بردارها A -متعامد هستند، نتیجه می‌گیریم که مستقل خطی نیز هستند و یک پایه برای $\mathbb{R}^n$ تشکیل می‌دهند. در نتیجه، اعدادی حقیقی مثل γ_i ، $i = 0, 1, \dots, n-1$ موجودند به‌طوری‌که

$$r_n = \sum_{i=0}^{n-1} \gamma_i p_i.$$

اکنون، با استفاده از لم ۴.۶، داریم $\langle p_i, r_n \rangle = 0$ ، $i = 0, 1, \dots, n-1$. از این‌رو،

$$\begin{aligned}\|r_n\|_2^2 &= \langle r_n, r_n \rangle = \left\langle \sum_{i=0}^{n-1} \gamma_i p_i, r_n \right\rangle \\ &= \sum_{i=0}^{n-1} \gamma_i \langle p_i, r_n \rangle = \sum_{i=0}^{n-1} \gamma_i \times 0 = 0.\end{aligned}$$

□ بنابراین، $r_n = 0$ و در نتیجه، $Ax_n = b$ که اثبات قضیه را کامل می‌کند.

این قضیه نشان می‌دهد که در حساب دقیق، الگوریتم گرادیان مزدوج حداکثر در n تکرار (تعداد متناهی تکرار) به جواب واقعی دستگاه همگرا می‌شود. در عمل برای ماتریس‌های بزرگ یا ماتریس

بدحالت، باتوجه به انباشتگی خطاها معمولاً روش گرادیان مزدوج در تعداد متناهی تکرار به جواب واقعی همگرا نمی‌شود. بنابراین، تکرار آن قدر ادامه می‌یابد تا به یک جواب تقریبی با دقت کافی برسیم. همان‌گونه که دیدیم الگوریتم گرادیان مزدوج از یک روش تصویری متعامد به‌دست می‌آید. بنابراین، در تکرار m ام، یک تقریب به‌صورت

$$x_m = x_0 + \delta_m \in x_0 + \mathcal{K}_m(A, r_0),$$

به‌دست می‌آید به‌طوری‌که

$$\|x_* - x_m\|_A = \min_{x \in x_0 + \mathcal{K}_m(A, r_0)} \|x_* - x\|_A. \quad (۳۵-۶)$$

هر عضو زیرفضای $x_0 + \mathcal{K}_m(A, r_0)$ به‌صورت $x = x_0 + q(A)r_0$ است که در آن q یک چندجمله‌ای از درجه حداکثر $1 - m$ است. داریم

$$r_k = b - Ax_k = A(x_* - x_k) = Ae_k,$$

که در آن $e_k = x_* - x_k$ لذا

$$\begin{aligned} e &= x_* - x \\ &= x_* - x_0 - q(A)r_0 \\ &= e_0 - q(A)Ae_0 \\ &= (I - Aq(A))e_0. \end{aligned}$$

درنتیجه، از رابطه (۳۵-۶) خواهیم داشت

$$\|e_m\|_A = \min_{q \in \mathbb{P}_{m-1}} \|(I - Aq(A))e_0\|_A = \min_{r \in \mathbb{P}_m, r(\cdot) \neq 1} \|r(A)e_0\|_A, \quad (۳۶-۶)$$

که در آن $\mathbb{P}_k$ مجموعه چندجمله‌ای‌ها از درجه حداکثر k است. از اینکه ماتریس A متقارن است، با توجه به قضیه ۱۷.۱ یک ماتریس متعامد مثل U وجود دارد به‌طوری‌که

$$A = U^T D U,$$

که در آن $D = \text{diag}(\lambda_1, \dots, \lambda_n)$. بنابراین،

$$r(A) = U^T r(D) U,$$

و

$$Ar(A) = U^T D U U^T r(D) U = U^T D r(D) U.$$

در نتیجه، با فرض $\xi = Ue_\circ = (\xi_1, \dots, \xi_n)^T$ خواهیم داشت

$$\begin{aligned} \|r(A)e_\circ\|_A^2 &= \langle Ar(A)e_\circ, r(A)e_\circ \rangle \\ &= \langle U^T D r(D) U e_\circ, U^T r(D) U e_\circ \rangle \\ &= \langle D r(D) U e_\circ, U U^T r(D) U e_\circ \rangle \\ &= \langle D r(D) \xi, r(D) \xi \rangle \\ &= \sum_{i=1}^n \lambda_i r(\lambda_i)^2 \xi_i^2 \\ &\leq \max_i r(\lambda_i)^2 \sum_{i=1}^n \lambda_i \xi_i^2 \\ &= \max_i r(\lambda_i)^2 \|e_\circ\|_A^2 \\ &\leq \max_{\lambda \in [\lambda_1, \lambda_n]} r(\lambda)^2 \|e_\circ\|_A^2, \end{aligned}$$

که در آن λ_1 و λ_n به ترتیب کوچک‌ترین و بزرگ‌ترین مقدار ویژه ماتریس A هستند. بنابراین، با استفاده از این رابطه و رابطه (۴۵-۶) داریم

$$\|e_m\|_A \leq \min_{r \in \mathbb{P}_m, r(\circ)=1} \max_{\lambda \in [\lambda_1, \lambda_n]} |r(\lambda)| \|e_\circ\|_A. \quad (۳۷-۶)$$

برای بررسی خواص سمت راست این رابطه از چند جمله‌ای‌های چبیشف استفاده می‌کنیم. با استفاده از قضیه ۳۲.۲ و با فرض $\alpha = \lambda_1$, $\beta = \lambda_n$ و $\gamma = \circ$ رابطه (۳۷-۶) به صورت

$$\|e_m\|_A \leq \frac{1}{\left| T_m \left(\frac{\lambda_1 + \lambda_n}{\lambda_1 - \lambda_n} \right) \right|} \|e_\circ\|_A, \quad (۳۸-۶)$$

نوشته می‌شود. با فرض $\eta = (\lambda_1 + \lambda_n)/(\lambda_1 - \lambda_n)$ داریم

$$\begin{aligned} \eta + \sqrt{\eta^2 - 1} &= \frac{\lambda_1 + \lambda_n + 2\sqrt{\lambda_1 \lambda_n}}{\lambda_1 - \lambda_n} \\ &= \frac{(\sqrt{\lambda_1} + \sqrt{\lambda_n})^2}{(\sqrt{\lambda_1} + \sqrt{\lambda_n})(\sqrt{\lambda_1} - \sqrt{\lambda_n})} \end{aligned}$$

$$= \frac{\sqrt{\lambda_1} + \sqrt{\lambda_n}}{\sqrt{\lambda_1} - \sqrt{\lambda_n}}.$$

در نتیجه، با استفاده از رابطه (۲-۵) می‌بینیم که

$$T_m \left(\frac{\lambda_1 + \lambda_n}{\lambda_1 - \lambda_n} \right) = \frac{1}{2} \left[\left(\frac{\sqrt{\lambda_1} + \sqrt{\lambda_n}}{\sqrt{\lambda_1} - \sqrt{\lambda_n}} \right)^m + \left(\frac{\sqrt{\lambda_1} + \sqrt{\lambda_n}}{\sqrt{\lambda_1} - \sqrt{\lambda_n}} \right)^{-m} \right].$$

در رابطه اخیر، عبارت‌های داخل کروشه هم علامت هستند. بنابراین، می‌توان نوشت

$$\left| T_m \left(\frac{\lambda_1 + \lambda_n}{\lambda_1 - \lambda_n} \right) \right| \geq \frac{1}{2} \left| \frac{\sqrt{\lambda_1} + \sqrt{\lambda_n}}{\sqrt{\lambda_1} - \sqrt{\lambda_n}} \right|^m = \frac{1}{2} \left(\frac{\sqrt{\lambda_n} + \sqrt{\lambda_1}}{\sqrt{\lambda_n} - \sqrt{\lambda_1}} \right)^m.$$

با استفاده از این رابطه، رابطه (۶-۳۸) و اینکه $\text{cond}_2(A) = \lambda_n/\lambda_1$ خواهیم داشت

$$\|e_m\|_A \leq 2 \left(\frac{\sqrt{\text{cond}_2(A)} - 1}{\sqrt{\text{cond}_2(A)} + 1} \right)^m \|e_0\|_A. \quad (۶-۳۹)$$

این رابطه نشان می‌دهد که اگر $\text{cond}_2(A)$ عدد بزرگی باشد، آنگاه سرعت همگرایی روش ممکن است کند باشد.

اکنون به مقایسه روش گرادیان مزدوج و روش سریع‌ترین کاهش می‌پردازیم. همان‌گونه که پیش‌تر دیدیم، در روش سریع‌ترین کاهش داریم (قضیه ۶.۵ را ببینید)

$$\|e_m\|_A \leq \left(\frac{\text{cond}(A) - 1}{\text{cond}(A) + 1} \right)^m \|e_0\|_A. \quad (۶-۴۰)$$

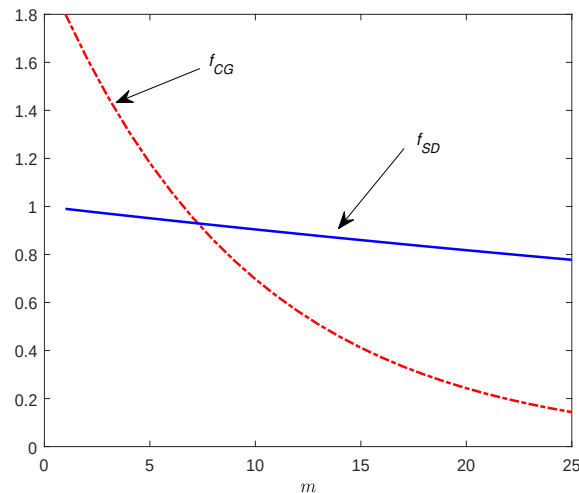
برای مقایسه این دو روش، توابع زیر را تعریف می‌کنیم

$$f_{CG}(m) = 2 \left(\frac{\sqrt{\text{cond}(A)} - 1}{\sqrt{\text{cond}(A)} + 1} \right)^m,$$

$$f_{SD}(m) = \left(\frac{\text{cond}(A) - 1}{\text{cond}(A) + 1} \right)^m.$$

کافی است توابع f_{SD} و f_{CG} را با یکدیگر مقایسه کنیم. برای این کار، فرض می‌کنیم $\text{cond}(A) = 100$ و این دو تابع را برای مقادیر $m = 1, 2, \dots, 25$ در شکل ۶-۲ رسم کرده‌ایم. همان‌گونه که ملاحظه می‌شود سرعت کاهش f_{CG} به مراتب از f_{SD} بیشتر است.

تابع pcg در متلب برای الگوریتم گرادیان مزدوج فراهم شده است. این تابع به همراه ورودی و خروجی‌های آن به صورت



شکل ۲-۶ مقایسه روند همگرایی روش سریع‌ترین کاهش و گرادیان مزدوج

$$[x, \text{flag}, \text{relres}, \text{iter}, \text{resvec}] = \text{pcg}(A, b, \text{tol}, \text{maxit}, [], [], x_0)$$

است. متغیرهای ورودی و خروجی تابع `pcg` کاملاً شبیه به تابع `gmres` است و چیزی برای توضیح وجود ندارد.

مثال ۳.۶ دستگاه معرفی‌شده در بخش ۱-۸ را در نظر بگیرید که در آن $\tau = \sigma = 0$. می‌توان دید که ماتریس A معین مثبت متقارن است. دستگاه $Ax = b$ را در نظر بگیرید که در آن

$$b = A(1, 1, \dots, 1)^T \in \mathbb{R}^n,$$

و $n = N^2$. این دستگاه را به‌ازای مقادیر مختلف N ، یعنی $N = 200, 400, 600, 800, 1000$ و با استفاده از روش CG (با کمک تابع `pcg` در متلب) حل کرده و نتایج را در جدول ۲-۶ ارائه کرده‌ایم. شرط توقف را

$$\frac{\|r_k\|_2}{\|r_0\|_2} \leq 10^{-7},$$

و حدس اولیه را بردار صفر فرض کرده‌ایم. در جدول ۲-۶، اعداد n و nnz به‌ترتیب بعد ماتریس A و تعداد درایه‌های ناصفر آن می‌باشند. این جدول به‌خوبی توانایی الگوریتم CG برای حل دستگاه‌های بزرگ با ماتریس ضرایب معین مثبت متقارن را نشان می‌دهد.

جدول ۶-۲ نتایج عددی برای مثال ۳.۶.

n	nnz	CG	
		$Iters$	$relres$
۴۰۰۰۰	۱۹۹۲۰۰	۳۳۶	$9.7e-8$
۱۶۰۰۰۰	۷۹۸۴۰۰	۶۵۹	$9.9e-8$
۳۶۰۰۰۰	۱۷۹۷۶۰۰	۹۷۷	$9.8e-8$
۶۴۰۰۰۰	۳۱۹۶۸۰۰	۱۲۹۰	$9.7e-8$
۱۰۰۰۰۰۰	۴۹۹۶۰۰۰	۱۵۹۹	$9.8e-8$

۹-۶ گرادیان مزدوج برای حل دستگاه معادلات نرمال

در حالتی که ماتریس‌های $A^T A$ یا AA^T معین مثبت متقارن هستند، می‌توان الگوریتم گرادیان مزدوج را برای حل دستگاه‌هایی که ماتریس ضرایب آنها این دو ماتریس هستند، به‌کار برد که در ادامه جزئیات این دو الگوریتم را بررسی می‌کنیم.

۹-۶-۱ الگوریتم CGNR

فرض کنید $A \in \mathbb{R}^{m \times n}$ ، $m \geq n$ و $\text{rank}(A) = n$. در این صورت، دستگاه ماتریس $A^T A$ معین مثبت متقارن است و دستگاه

$$A^T A x = A^T b,$$

جواب یکتا دارد. بنابراین، می‌توان الگوریتم گرادیان مزدوج را برای حل این دستگاه به‌کار برد. برای این کار، کفایت در الگوریتم ۱۱.۶ به‌جای ماتریس A و بردار b به‌ترتیب $A^T A$ و $A^T b$ را جایگزین کنیم. باید توجه کنیم که در این صورت بردار مانده r_j بایستی با

$$\begin{aligned} z_j &= A^T b - A^T A x_j \\ &= A^T (b - A x_j) \\ &= A^T r_j, \end{aligned}$$

جایگزین شود. در این صورت، گام‌های الگوریتم ۱۱.۶ به‌صورت زیر خواهند بود

- $\alpha_j := \langle z_j, z_j \rangle / \langle A^T A p_j, p_j \rangle = \langle z_j, z_j \rangle / \langle A p_j, A p_j \rangle = \|z_j\|_2^2 / \|A p_j\|_2^2$,
- $x_{j+1} := x_j + \alpha_j p_j$,

- $r_{j+1} := r_j - \alpha_j A p_j$,
- $z_{j+1} := A^T r_{j+1}$,
- $\beta_j := \langle z_{j+1}, z_{j+1} \rangle / \langle z_j, z_j \rangle = \|z_{j+1}\|_2^2 / \|z_j\|_2^2$,
- $p_{j+1} := z_{j+1} + \beta_j p_j$.

با توجه به اینکه بردار $A p_j$ در چند گام از الگوریتم استفاده شده است، قرار می‌دهیم $w_j = A p_j$. الگوریتمی که به این ترتیب به دست می‌آید الگوریتم CGNR نامیده می‌شود و به صورت زیر جمع‌بندی می‌شود.

الگوریتم ۱۲.۶ الگوریتم CGNR

1. Choose an initial guess x_0 and $r_0 := b - A x_0$, $z_0 = A^T r_0$ and $p_0 := z_0$
2. For $j = 0, 1, 2, \dots$, until convergence, Do
3. $w_j = A p_j$
4. $\alpha_j := \|z_j\|_2^2 / \|w_j\|_2^2$
5. $x_{j+1} := x_j + \alpha_j p_j$
6. $r_{j+1} := r_j - \alpha_j w_j$
7. $z_{j+1} := A^T r_{j+1}$
8. $\beta_j := \|z_{j+1}\|_2^2 / \|z_j\|_2^2$
9. $p_{j+1} := z_{j+1} + \beta_j p_j$
10. EndDo

با استناد به رابطه (۳۵-۶)، اگر x_m تقریب به دست آمده در تکرار m ام از الگوریتم CGNR باشد، آنگاه

$$\|x^* - x_m\|_{A^T A} = \min_{x \in x_* + \mathcal{K}_m(A^T A, A^T r_*)} \|x^* - x\|_{A^T A}, \quad (۴۱-۶)$$

که در آن x^* جواب یکتای دستگاه $A^T A x = A^T b$ است. اما، داریم

$$\begin{aligned} \|x^* - x\|_{A^T A} &= \langle A^T A(x^* - x), x_* - x \rangle \\ &= \langle A(x^* - x), A(x_* - x) \rangle \\ &= \langle b - A x, b - A x \rangle \\ &= \|b - A x\|^2. \end{aligned}$$

بنابراین، رابطه (۴۵-۶) را می‌توان به صورت زیر نوشت

$$\|b - Ax_m\|_2 = \min_{x \in x_0 + \mathcal{K}_m(A^T A, A^T r_0)} \|b - Ax\|_2. \quad (42-6)$$

درحقیقت، این رابطه نشان می‌دهد که بردار x_m ، نرم مانده دستگاه $Ax = b$ را روی زیرفضای $x_0 + \mathcal{K}_m(A^T A, A^T r_0)$ کمینه می‌کند. به همین دلیل است که این الگوریتم، CGNR نام‌گذاری شده است (حرف “N” برای کلمه Normal و حرف “R” برای کلمه Residual). چند نکته مهم دیگر درباره الگوریتم CGNR قابل ذکر است. همانند رابطه (۸-۱) می‌توان دید که

$$\text{cond}_2(A^T A) = \text{cond}_2^*(A).$$

بنابراین، حتی اگر $\text{cond}_2(A)$ خیلی بزرگ نباشد، ممکن است مقدار $\text{cond}_2(A^T A)$ بزرگ خواهد بود. برای مثال، اگر $\text{cond}_2(A) = 10^8$ ، آنگاه $\text{cond}_2(A^T A) = 10^{16}$ ، که عدد بسیار بزرگی است. درحقیقت هرگونه بهبود در جواب در یک تکرار روش، ممکن است بر اثر خطاهای گردکردن از بین برود. ازاین‌رو، در حالت کلی، استفاده از الگوریتم CGNR برای حل یک دستگاه مربعی بد حالت توصیه نمی‌شود. در هر صورت، اگر عدد حالت ماتریس A به اندازه کافی کوچک باشد، آنگاه روش CGNR مشکلی نخواهد داشت. در حالت خیلی خاص که ماتریس A متعامد است ($A^T A = I$)، بهترین روش، حل دستگاه معادلات نرمال است (الگوریتم CGNR در صفر تکرار همگرا می‌شود).

مثال ۴.۶ چهار ماتریس WELL1033 ، WELL1850 ، ILLC1033 ، و ILLC1850 را از بازار ماتریس با وبگاه اشاره‌شده در [۷۵] (یا از [۳۳]) انتخاب نموده‌ایم. این ماتریس‌ها با خواص عمومی‌شان از قبیل تعداد سطرها (m)، تعداد ستون‌ها (n) و تعداد درایه‌های ناصفر (nnz) در جدول ۳-۶ ارائه شده‌اند. چهار دستگاه با این ماتریس‌ها ساختیم که بردار سمت راست هریک از آنها به صورت $b = Ae$ است که در آن $e \in \mathbb{R}^n$ برداری است که همه مؤلفه‌های آن برابر با ۱ می‌باشد. الگوریتم CGNR را برای حل دستگاه معادلات نرمال

$$A^T Ax = A^T b,$$

به‌کار برده‌ایم. حدس اولیه را بردار صفر گرفته و تکرار را زمانی متوقف می‌کنیم که

$$\text{relres} = \frac{\|r_k\|_2}{\|r_0\|_2} < 10^{-7},$$

جدول ۳-۶ نتایج عددی برای مثال ۴.۶.

نام ماتریس	m	n	nnz	$iters$	$relres$
WELL1033	۱۰۳۳	۱۵۶	۴۷۳۲	۱۵۶	$8/28e - 8$
WELL1850	۱۸۵۰	۷۱۲	۸۷۵۸	۴۰۳	$9/77e - 8$
ILLC1033	۱۰۳۳	۳۲۰	۴۷۳۲	۳۰۵۶	$9/87e - 8$
ILLC1850	۱۸۵۰	۷۱۲	۸۷۵۸	۱۵۷۹	$9/98e - 8$

که در آن $r_k = b - Ax_k$. نتایج عددی نیز در جدول ۳-۶ نشان داده شده‌اند. در این جدول، تعداد تکرارها برای همگرایی ($Iters$) و مقدار $relres$ نشان داده شده‌اند. گرچه برای هر چهار دستگاه روش تکراری CGNR همگرا شده است، می‌بینیم که با وجود اینکه ابعاد ماتریس‌ها کوچک هستند، تعداد تکرارها نسبتاً بزرگ می‌باشند.

۲-۹-۶ الگوریتم CGNE

فرض کنید $A \in \mathbb{R}^{m \times n}$ ، $m \leq n$ و $\text{rank}(A) = m$. در این صورت، ماتریس AA^T معین مثبت متقارن است و دستگاه

$$AA^T u = b, \quad (۴۳-۶)$$

جواب یکتا دارد، که در آن $x = A^T u$. در نتیجه، می‌توان الگوریتم گرادیان مزدوج را برای حل دستگاه بالا به‌کار برد. برای این کار، کافیت در الگوریتم ۱۱.۶ به‌جای ماتریس A ، ماتریس AA^T و به‌جای متغیر x ، متغیر u را جایگزین کنیم. ابتدا توجه کنید که اگر u_j جواب تقریبی به‌دست‌آمده توسط این الگوریتم باشد، آنگاه جواب تقریبی برای دستگاه اصلی، یعنی $Ax = b$ ، به‌صورت $x_j = A^T u_j$ خواهد بود. همچنین، بردار مانده این تقریب برای دستگاه نرمال به‌صورت زیر خواهد بود

$$r_j = b - AA^T u_j = b - Ax_j.$$

در حقیقت، بردار مانده دستگاه نرمال (۴۳-۶)، با بردار مانده دستگاه اولیه، یعنی $Ax = b$ ، یکسان خواهد بود. اکنون، با توجه به نکته ذکرشده، گام‌های الگوریتم ۱۳.۶ برای حل دستگاه (۴۳-۶) به‌صورت زیر خواهند بود

- $\alpha_j := \langle r_j, r_j \rangle / \langle AA^T q_j, q_j \rangle = \langle r_j, r_j \rangle / \langle A^T q_j, A^T q_j \rangle = \|r_j\|_2^2 / \|A^T q_j\|_2^2$,
- $u_{j+1} := u_j + \alpha_j q_j$,

- $r_{j+1} := r_j - \alpha_j A A^T q_j$,
- $\beta_j := \langle r_{j+1}, r_{j+1} \rangle / \langle r_j, r_j \rangle = \|r_{j+1}\|_2^2 / \|r_j\|_2^2$,
- $q_{j+1} := r_{j+1} + \beta_j q_j$.

با فرض $p_j = A^T q_j$ گام‌های بالا را می‌توان به صورت زیر بازنویسی کرد

- $\alpha_j := \|r_j\|_2^2 / \|p_j\|_2^2$,
- $x_{j+1} := x_j + \alpha_j p_j$,
- $r_{j+1} := r_j - \alpha_j A p_j$,
- $\beta_j := \|r_{j+1}\|_2^2 / \|r_j\|_2^2$,
- $p_{j+1} := A^T r_{j+1} + \beta_j p_j$.

با جمع‌بندی این نتایج، الگوریتم CGNE (یا الگوریتم کرایگ [۲۹]) به دست می‌آید.

الگوریتم ۱۳.۶ الگوریتم CGNE (الگوریتم کرایگ)

1. Choose an initial guess x_0 and $r_0 := b - Ax_0$ and $p_0 := A^T r_0$
2. For $j = 0, 1, 2, \dots$, until convergence, Do
3. $\alpha_j := \|r_j\|_2^2 / \|p_j\|_2^2$
4. $x_{j+1} := x_j + \alpha_j p_j$
5. $r_{j+1} := r_j - \alpha_j A p_j$
6. $\beta_j := \|r_{j+1}\|_2^2 / \|r_j\|_2^2$
7. $p_{j+1} := A^T r_{j+1} + \beta_j p_j$
8. EndDo

باتوجه به رابطه (۶-۳۵)، اگر u_m تقریب به دست آمده در تکرار m ام از الگوریتم CGNE باشد، آنگاه

$$\|u^* - u_m\|_{AA^T} = \min_{u \in u_* + \mathcal{K}_m(AA^T, r_*)} \|u^* - u\|_{AA^T}, \quad (۴۴-۶)$$

که در آن u^* جواب یکتای دستگاه $AA^T u = b$ است. اما، داریم

$$\begin{aligned} \|u^* - u\|_{AA^T} &= \langle AA^T(u^* - u), u^* - u \rangle \\ &= \langle A^T(u^* - u), A^T(u^* - u) \rangle \\ &= \langle A^T u^* - A^T u, A^T u^* - A^T u \rangle \\ &= \langle x^* - x, x^* - x \rangle \\ &= \|x^* - x\|_V^2, \end{aligned}$$

جدول ۴-۶ نتایج عددی برای مثال ۵.۶.

نام ماتریس	m	n	nnz	$iters$	$relres$
WM1	۲۰۷	۲۷۷	۲۹۰۹	۶۸۶	$5/28e - 8$
WM2	۲۰۷	۲۶۰	۲۹۴۲	۶۵۷	$8/99e - 8$
WM3	۲۰۷	۲۶۰	۲۹۴۸	۸۴۴	$6/84e - 8$

که در آن $x = A^T u$ از طرفی، از اینکه $u \in u_0 + \mathcal{K}_m(AA^T, r_0)$ نتیجه می‌گیریم که

$$x = A^T u \in A^T u_0 + A^T \mathcal{K}_m(AA^T, r_0) = x_0 + \mathcal{K}_m(A^T A, A^T r_0).$$

در نتیجه، می‌توان رابطه (۴۴-۶) را به صورت زیر نوشت

$$\|x^* - x_m\|_2 = \min_{x \in x_0 + \mathcal{K}_m(A^T A, A^T r_0)} \|x^* - x\|_2. \quad (۴۵-۶)$$

در حقیقت، این رابطه نشان می‌دهد که بردار x_m نرم خطای دستگاه $Ax = b$ را روی زیرفضای $x_0 + \mathcal{K}_m(A^T A, A^T r_0)$ کمینه می‌کند. به همین دلیل است که این الگوریتم، CGNE نامگذاری شده است (حرف “N” برای کلمه Normal و حرف “E” برای کلمه Error).

مثال ۵.۶ سه ماتریس WM1، WM2، و WM3 را از بازار ماتریس انتخاب نموده‌ایم. همانند مثال ۴.۶، خواص این ماتریس‌ها در جدول ۴-۶ ارائه شده‌اند. سایر شرایط همانند مثال ۴.۶ هستند، با این تفاوت که در اینجا دستگاه

$$AA^T u = b,$$

را حل کرده‌ایم، که در آن $x = A^T u$. نتایج عددی نیز در جدول ۴-۶ نشان داده شده‌اند. همان‌گونه می‌بینیم روش تکراری CGNE برای هر سه دستگاه همگرا شده است، اما با اینکه ابعاد ماتریس‌ها کوچک می‌باشند، تعداد تکرارها نسبتاً بزرگ هستند.

تمرینات

۱.۶ رابطه (۲-۶) را ثابت کنید.

۲.۶ فرض کنید $v \in \mathbb{R}^n$ و $A \in \mathbb{R}^{n \times n}$. نشان دهید که زیرفضای کرایلف $\mathcal{K}_m(A, v)$ یک زیرفضای برداری از $\mathbb{R}^n$ است.

۳.۶ فرض کنید

$$A = \begin{pmatrix} ۱ & ۱ & ۱ \\ ۱ & -۱ & ۱ \\ ۰ & ۱ & ۱ \end{pmatrix}, \quad v = \begin{pmatrix} ۱ \\ ۱ \\ -۱ \end{pmatrix}.$$

گراید بردار v نسبت به ماتریس A را به دست آورید.

۴.۶ فرض کنید

$$A = \begin{pmatrix} ۲ & ۰ & ۰ & ۰ \\ ۱ & ۴ & ۱ & ۲ \\ ۰ & -۱ & ۱ & -۱ \\ -۱ & -۲ & -۱ & ۰ \end{pmatrix}.$$

نشان دهید که چندجمله‌ای مشخصه ماتریس A به صورت

$$p_۴(x) = x^۴ - ۷x^۳ + ۱۸x^۲ - ۲۰x + ۸,$$

و چندجمله‌ای مینیمال آن به صورت

$$p_۳(x) = x^۳ - ۵x^۲ + ۸x - ۴,$$

است. نشان دهید

$$\text{grade}_A(v) = ۱,$$

که در آن $v = (۱, ۰, ۱, -۱)^T$ ؛ به عبارت دیگر، گراید بردار v نسبت به ماتریس A برابر با ۱ است.

۵.۶ آیا روش تکراری GMRES(2) با شروع مجدد برای حل دستگاه

$$\begin{pmatrix} ۲ & -۱ & ۱ & ۲ \\ ۱ & ۳ & ۳ & -۲ \\ -۱ & -۳ & ۴ & ۲ \\ -۲ & ۲ & ۲ & ۵ \end{pmatrix} \begin{pmatrix} x_۱ \\ x_۲ \\ x_۳ \\ x_۴ \end{pmatrix} = \begin{pmatrix} ۴ \\ ۱۰ \\ ۱ \\ ۶ \end{pmatrix},$$

همگراست؟ چرا؟

۶.۶ دستگاه $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

نشان دهید که الگوریتم GMRES(1) با شروع مجدد به‌ازای حدس اولیه $x_0 = 0$ برای حل این دستگاه هیچ وقت همگرا نخواهد شد.

۷.۶ نشان دهید که چگونه نسخه کامل الگوریتم‌های FOM و GMRES با حدس اولیه $x_0 = 0$ برای حل دستگاه $Ax = b$ همگرا می‌شود که در آن

$$A = \begin{pmatrix} & & & & 1 \\ & & & & \\ 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

۸.۶ فرض کنید ماتریس A به‌صورت

$$A = \begin{pmatrix} I & Y \\ 0 & I \end{pmatrix},$$

تعریف شده باشد، که در آن I ماتریس همانی $m \times m$ و $Y \in \mathbb{R}^{m \times m}$ یک ماتریس دلخواه است. الگوریتم GMRES کامل برای حل یک دستگاه مثل $Ax = b$ به‌کار برده می‌شود، که در آن b یک بردار دلخواه است. حداکثر تکرار لازم برای همگرایی چیست؟

۹.۶ فرض کنید ماتریس A به‌صورت

$$A = \begin{pmatrix} I & Y \\ 0 & S \end{pmatrix},$$

تعریف شده باشد، که در آن I ماتریس همانی $m \times m$ و $Y, S \in \mathbb{R}^{m \times m}$ دو ماتریس دلخواه هستند. نسخه کامل الگوریتم GMRES را برای حل دستگاه معادلات خطی $Ax = b$ به‌کار می‌بریم که در آن b یک بردار دلخواه است. فرض کنید

$$r^{(\circ)} = \begin{pmatrix} r_{\circ}^{(1)} \\ r_{\circ}^{(2)} \end{pmatrix},$$

بردار مانده اولیه و چندجمله‌ای مینیمال $r_{\circ}^{(2)}$ نسبت به S برابر با k باشد، یعنی

$$\text{grade}_S(r_{\circ}^{(2)}) = k.$$

چه تعداد تکرار برای همگرایی نسخه کامل روش GMRES نیاز است؟

راهنمایی: مجموع $\sum_{i=0}^k \beta_i (A^{i+1} - A^i) r_{\circ}$ را محاسبه کنید که در آن $\sum_{i=0}^k \beta_i t^i$ چندجمله‌ای مینیمال $r_{\circ}^{(2)}$ نسبت به S است.

۱۰.۶ فرض کنید

$$A = \begin{pmatrix} I & Y_2 & & & \\ & I & Y_3 & & \\ & & I & \ddots & \\ & & & I & Y_k \\ & & & & I \end{pmatrix},$$

که در آن I ماتریس همانی $m \times m$ و $Y_i \in \mathbb{R}^{m \times m}$ ، $i = 2, 3, \dots, k$ ماتریس‌های دلخواه باشند.
(الف) نشان دهید

$$(I - A)^k = \circ.$$

(ب) فرض کنید الگوریتم GMRES کامل برای حل یک دستگاه به صورت $Ax = b$ به کار برده شود، که در آن b یک بردار دلخواه است. حداکثر تکرار لازم برای اینکه این روش برای حل این دستگاه همگرا شود چیست؟

۱۱.۶ فرض $u \in \mathbb{R}^n$ یک بردار ناصفر باشد. نشان دهید مجموعه مقادیر ویژه ماتریس $A = I + uu^T$ به صورت $\sigma(A) = \{1, 1 + u^T u\}$ است، که در آن I نشان دهنده ماتریس همانی است. ثابت کنید نسخه کامل الگوریتم GMRES در حداکثر دو تکرار (در حساب دقیق) به جواب دقیق دستگاه $Ax = b$ همگرا می‌شود، که در آن b یک بردار دلخواه در $\mathbb{R}^n$ است.

۲۵۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

۱۲.۶ فرض کنید ماتریس A به صورت $A = I + \alpha B$ باشد که در آن B یک ماتریس متقارن کج است، یعنی $B^T = -B$. نشان دهید اگر الگوریتم آرنولد را برای حل دستگاه $Ax = b$ به کار ببریم، ماتریس بالاهسنبرگی حاصل به صورت

$$H_m = \begin{pmatrix} 1 & -\eta_2 & & & \\ \eta_2 & 1 & -\eta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \eta_{m-1} & 1 & -\eta_m \\ & & & \eta_m & 1 \end{pmatrix}$$

خواهد بود.

روش‌های مبتنی بر زیرفضای کرایلف (قسمت ۲)

در فصل ۶ تعدادی از روش‌های تکراری مبتنی بر زیرفضای کرایلف را معرفی نمودیم که بر مبنای متعامدسازی بردارهای زیرفضای کرایلف استوار بودند. در این فصل رده‌ای از روش‌های مبتنی بر زیرفضای کرایلف را معرفی می‌کنیم که به جای متعامدسازی در زیرفضای کرایلف، از دو متعامدسازی لانکزوس استفاده می‌کند. از مهم‌ترین روش‌هایی که در این بخش معرفی می‌شوند عبارتند از روش تکراری لانکزوس دوطرفه، الگوریتم گرادیان دومزدوج (BiCG)، الگوریتم مانده شبه‌مینیمال (QMR)، الگوریتم گرادیان مزدوج مربع‌شده (CGS)، و الگوریتم BiCG پایدارشده (BiCGSTAB).

۷-۱ الگوریتم لانکزوس برای ماتریس‌های نامتقارن

الگوریتم دو متعامدسازی لانکزوس یک تعمیم از الگوریتم لانکزوس متقارن است که در فصل پیش مطالعه کردیم. این الگوریتم یک زوج پایه برای زیرفضاهای کرایلف

$$\begin{aligned}\mathcal{K}_m(A, v_1) &= \text{span}\{v_1, Av_1, \dots, A^{m-1}v_1\}, \\ \mathcal{K}_m(A^T, w_1) &= \text{span}\{w_1, A^T w_1, \dots, A^{m-1}w_1\},\end{aligned}$$

می‌سازد به طوری که

$$\langle v_i, w_j \rangle = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

الگوریتم دو متعامدسازی لانکزوس به صورت زیر نوشته می‌شود.

الگوریتم ۱.۷ الگوریتم دومتعامدسازی لانکروس

1. Choose two vectors v_1 and w_1 such that $\langle v_1, w_1 \rangle = 1$.
2. Set $\beta_1 = \delta_1 = 0$, and $w_0 = v_0 = 0$.
3. For $j = 1, 2, \dots, m$ Do:
4. $\alpha_j := \langle Av_j, w_j \rangle$
5. $\hat{v}_{j+1} := Av_j - \alpha_j v_j - \beta_j v_{j-1}$
6. $\hat{w}_{j+1} := A^T w_j - \alpha_j w_j - \delta_j w_{j-1}$
7. $\delta_{j+1} = |\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle|^{\frac{1}{2}}$. If $\delta_{j+1} = 0$, then stop.
8. $\beta_{j+1} = \langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle / \delta_{j+1}$.
9. $w_{j+1} := \hat{w}_{j+1} / \beta_{j+1}$
10. $v_{j+1} := \hat{v}_{j+1} / \delta_{j+1}$
11. EndDo.

قضیه ۱.۷ فرض کنید الگوریتم ۱.۷ پیش از گام m متوقف نشود. در این صورت، مجموعه‌های $S_m^{(v)} = \{v_1, v_2, \dots, v_m\}$ و $S_m^{(w)} = \{w_1, w_2, \dots, w_m\}$ دومتعامد هستند. به علاوه، $S_m^{(w)}$ به ترتیب یک پایه برای $\mathcal{K}_m(A, v_1)$ و $\mathcal{K}_m(A^T, w_1)$ تشکیل می‌دهند. همچنین، داریم

$$AV_m = V_m T_m + \delta_{m+1} v_{m+1} e_m^T, \quad (۱-۷)$$

$$A^T W_m = W_m T_m^T + \beta_{m+1} w_{m+1} e_m^T, \quad (۲-۷)$$

$$W_m^T AV_m = T_m, \quad (۳-۷)$$

$$AV_m = V_{m+1} \bar{T}_m, \quad (۴-۷)$$

که در آن $W_m = (w_1, w_2, \dots, w_m)$ ، $V_{m+1} = (V_m, v_{m+1})$ ، $V_m = (v_1, v_2, \dots, v_m)$ و

$$T_m = \begin{pmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \end{pmatrix}, \quad \bar{T}_m = \begin{pmatrix} T_m \\ \delta_{m+1} e_m^T \end{pmatrix},$$

به طوری که $e_m = (1, 0, \dots, 0)^T \in \mathbb{R}^m$

برهان. به استقرا روی m نشان می‌دهیم مجموعه‌های $S_m^{(w)}$ و $S_m^{(v)}$ دومتعامد هستند. به ازای $m = 1$ ، داریم $S_1^{(w)} = \{w_1\}$ و $S_1^{(v)} = \{v_1\}$. لذا، از اینکه $\langle v_1, w_1 \rangle = 1$ ، چیزی برای اثبات وجود ندارد.

حال فرض می‌کنیم دو مجموعه $S_j^{(v)}$ و $S_j^{(w)}$ دومتعامد باشند و نشان می‌دهیم که دو مجموعه $S_{j+1}^{(w)}$ و $S_{j+1}^{(v)}$ نیز دومتعامد هستند. نخست، نشان می‌دهیم که به‌ازای هر $i \leq j$ ، $\langle v_{j+1}, w_i \rangle = 0$. به‌ازای $i = j$ داریم

$$\begin{aligned} \langle v_{j+1}, w_j \rangle &= \frac{1}{\delta_{j+1}} \langle \hat{v}_{j+1}, w_j \rangle \\ &= \frac{1}{\delta_{j+1}} \langle Av_j - \alpha_j v_j - \beta_j v_{j-1}, w_j \rangle \\ &= \frac{1}{\delta_{j+1}} \left(\langle Av_j, w_j \rangle - \alpha_j \langle v_j, w_j \rangle - \beta_j \langle v_{j-1}, w_j \rangle \right) \\ &= \frac{1}{\delta_{j+1}} \left(\langle Av_j, w_j \rangle - \alpha_j \right) = 0. \quad (\text{گام ۴}) \end{aligned}$$

توجه کنید که تساوی ماقبل آخر با توجه به رابطه‌های $\langle v_{j-1}, w_j \rangle = 0$ و $\langle v_j, w_j \rangle = 1$ ، نتیجه شده است. به‌ازای $j < i$ ، داریم

$$\begin{aligned} \langle v_{j+1}, w_i \rangle &= \frac{1}{\delta_{j+1}} \left(\langle Av_j, w_i \rangle - \alpha_j \langle v_j, w_i \rangle - \beta_j \langle v_{j-1}, w_i \rangle \right) \\ &= \frac{1}{\delta_{j+1}} \left(\langle v_j, A^T w_i \rangle - \beta_j \langle v_{j-1}, w_i \rangle \right) \\ &= \frac{1}{\delta_{j+1}} \left(\langle v_j, \beta_{i+1} w_{i+1} + \alpha_i w_i + \delta_i w_{i-1} \rangle - \beta_j \langle v_{j-1}, w_i \rangle \right). \quad (\text{گام ۶}) \end{aligned}$$

اگر $i \leq j-2$ ، آنگاه تمام جملات داخل پرانتز برابر با صفر خواهند بود و درنتیجه، $\langle v_{j+1}, w_i \rangle = 0$. در صورتی که $i = j-1$ ، داریم

$$\begin{aligned} \langle v_{j+1}, w_{j-1} \rangle &= \frac{1}{\delta_{j+1}} \left(\langle v_j, \beta_j w_j + \alpha_{j-1} w_{j-1} + \delta_{j-1} w_{j-2} \rangle - \beta_j \langle v_{j-1}, w_{j-1} \rangle \right) \\ &= \frac{1}{\delta_{j+1}} \left(\beta_j \langle v_j, w_j \rangle - \beta_j \langle v_{j-1}, w_{j-1} \rangle \right) = 0. \end{aligned}$$

به‌طور مشابه، می‌توان نشان داد که به‌ازای هر $i \leq j$ ، $\langle v_i, w_{j+1} \rangle = 0$.

برای اینکه نشان دهیم $S_m^{(v)}$ یک پایه برای فضای کرایلف $\mathcal{K}_m(A, v_1)$ است، ابتدا نشان می‌دهیم که این مجموعه، یک مجموعه مستقل خطی است. فرض کنید

$$\sum_{i=1}^n \alpha_i v_i = 0.$$

طرفین این رابطه را در w_j ضرب داخلی می‌کنیم، که در آن $1 \leq j \leq m$. در این صورت،

$$\left\langle \sum_{i=1}^n \alpha_i v_i, w_j \right\rangle = 0 \implies \sum_{i=1}^n \alpha_i \langle v_i, w_j \rangle = 0 \implies \alpha_j \langle v_j, w_j \rangle = 0 \implies \alpha_j = 0.$$

لذا مجموعه $S_m^{(v)}$ ، یک مجموعه مستقل خطی است. اکنون، به استقرا نشان می‌دهیم که هر عضو v_j از این مجموعه را می‌توان به صورت $v_j = q_{j-1}(A)v_1$ نوشت که در آن q_{j-1} یک چندجمله‌ای از درجه حداکثر $j-1$ است. به ازای $j=1$ ، اگر تعریف کنیم $q_0(t) = 1$ ، آنگاه خواهیم داشت $v_1 = q_0(A)v_1$. اکنون، فرض کنید به ازای هر $j \leq m$ ، داشته باشیم $v_i = q_{i-1}(A)v_1$ در این صورت، از گام‌های ۵ و ۱۰ الگوریتم، داریم

$$v_{j+1} = \frac{1}{\delta_{j+1}} (Av_j - \alpha_j v_j - \beta_j v_{j-1}).$$

بنابراین، اگر چندجمله‌ای q_j را به صورت

$$q_j = \frac{1}{\delta_{j+1}} (tq_{j-1} - \alpha_j q_{j-1} - \beta_j q_{j-2}),$$

تعریف کنیم، آنگاه q_j یک چندجمله‌ای از درجه حداکثر j است و داریم $v_{j+1} = q_j(A)v_1$. به طور مشابه می‌توان نشان داد که $S_m^{(w)}$ یک پایه برای $\mathcal{K}_m(A^T, v_1)$ است. برای اثبات رابطه (۷-۲۵)، از گام‌های ۵ و ۱۰ الگوریتم داریم

$$Av_j = \beta_j v_{j-1} + \alpha_j v_j + \delta_{j+1} v_{j+1}, \quad j = 1, 2, \dots, m.$$

لذا،

$$\begin{cases} Av_1 &= \alpha_1 v_1 + \delta_2 v_2, \\ Av_2 &= \beta_2 v_1 + \alpha_2 v_2 + \delta_3 v_3, \\ &\vdots \\ Av_{m-1} &= \beta_{m-1} v_{m-2} + \alpha_{m-1} v_{m-1} + \delta_m v_m, \\ Av_m &= \beta_m v_{m-1} + \alpha_m v_m + \delta_{m+1} v_{m+1}. \end{cases}$$

اگر این روابط را به شکل ماتریسی بنویسیم، خواهیم داشت

$$\begin{aligned} AV_m &= A(v_1, v_2, \dots, v_m) \\ &= (Av_1, Av_2, \dots, Av_m) \end{aligned}$$

$$\begin{aligned}
 &= (v_1, v_2, \dots, v_m, v_{m+1}) \begin{pmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \\ & & & & \delta_{m+1} \end{pmatrix} \\
 &= (V_m, v_{m+1}) \begin{pmatrix} T_m \\ \delta_{m+1} e_m^T \end{pmatrix} \\
 &= V_{m+1} \bar{T}_m, \\
 &= V_m T_m + \delta_{m+1} v_{m+1} e_m^T.
 \end{aligned}$$

همچنین، داریم

$$\begin{aligned}
 AV_m &= (V_m, v_{m+1}) \begin{pmatrix} T_m \\ \delta_{m+1} e_m^T \end{pmatrix} \\
 &= V_m T_m + \delta_{m+1} v_{m+1} e_m^T.
 \end{aligned}$$

به‌طور مشابه رابطه (۲-۷) ثابت می‌شود. سرانجام، با ضرب ماتریس W_m^T در طرفین رابطه (۲۵-۷)،
 رابطه (۳-۷) به‌دست می‌آید. $\square$

چند نکته مهم درباره الگوریتم ۱.۷ قابل ذکر است. انتخاب β_j ها و δ_j ها می‌تواند به‌صورت‌های مختلف انجام گیرد، اما بایستی به‌گونه‌ای انتخاب شوند که $\langle v_{j+1}, w_{j+1} \rangle = 1$. برای مثال، با انتخاب مقادیر $\delta_{j+1} = |\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle|^{\frac{1}{2}}$ و $\beta_{j+1} = \langle \hat{w}_{j+1}, \hat{v}_{j+1} \rangle / \delta_{j+1}$ ، خواهیم داشت

$$\begin{aligned}
 \langle v_{j+1}, w_{j+1} \rangle &= \langle \hat{v}_{j+1} / \beta_{j+1}, \hat{w}_{j+1} / \delta_{j+1} \rangle \\
 &= \langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle / (\beta_{j+1} \delta_{j+1}) = 1.
 \end{aligned}$$

آخرین تساوی از رابطه اخیر همچنین نشان می‌دهد که اگر $\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle = \beta_{j+1} \delta_{j+1}$ ، آنگاه خواهیم داشت $\langle v_{j+1}, w_{j+1} \rangle = 1$.

توقف در الگوریتم ۱.۷ با آنچه که در فرایند لانکزوس مقارن (الگوریتم ۸.۶) داشتیم، کاملاً متفاوت است. اگر در گام ۷ داشته باشیم $\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle = 0$ ، آنگاه ممکن است $\hat{v}_{j+1} = 0$ یا $\hat{w}_{j+1} = 0$ ، یا

۲۶۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

اینکه هیچ‌کدام از این دو بردار صفر نباشند. در حالتی که $\hat{v}_{j+1} = 0$ ، نتیجه می‌شود $\mathcal{K}_m(A, v_1)$ ناوردا است. در این صورت، اگر x_0 یک حدس اولیه برای جواب دستگاه $Ax = b$ باشد، $r_0 = b - Ax_0$ و $v_1 = r_0 / \|r_0\|_2$ ، آنگاه

$$x^* \in x_0 + \mathcal{K}_m(A, v_1) = x_0 + \mathcal{K}_m(A, r_0),$$

که در آن x^* جواب واقعی $Ax = b$ است. در صورتی که $\hat{w}_{j+1} = 0$ ، می‌بینیم که $\mathcal{K}_m(A^T, w_1)$ ناوردا است. در این صورت، اگر $\tilde{x}_0$ یک حدس اولیه برای جواب دستگاه $A^T \tilde{x} = \tilde{b}$ باشد، $\tilde{r}_0 = \tilde{b} - A^T \tilde{x}_0$ و $\tilde{w}_1 = \tilde{r}_0 / \|\tilde{r}_0\|_2$ ، آنگاه

$$\tilde{x}^* \in \tilde{x}_0 + \mathcal{K}_m(A^T, w_1) = \tilde{x}_0 + \mathcal{K}_m(A^T, \tilde{r}_0),$$

که در آن $\tilde{x}^*$ جواب واقعی دستگاه $A^T \tilde{x} = \tilde{b}$ است.

اما اگر هیچ‌کدام از دو بردار $\hat{v}_{j+1}$ و $\hat{w}_{j+1}$ صفر نباشند، ولی $\langle \hat{v}_{j+1}, \hat{w}_{j+1} \rangle = 0$ ، آنگاه این شکست، جدی خواهد بود. خوشبختانه برای رفع این مشکل، راهکاری وجود دارد که به راهکار پیش‌نگرش معروف است (مرجع [۸۱] را ببینید).

۲-۷ الگوریتم لانکزوس دوطرفه برای حل دستگاه معادلات خطی

در این بخش، الگوریتم لانکزوس دوطرفه برای حل دستگاه معادلات خطی $Ax = b$ را ارائه می‌کنیم. به‌طور مشابه، این الگوریتم را می‌توان برای حل دستگاه $A^T \tilde{x} = \tilde{b}$ نوشت.

با فرض $v_1 = r_0 / \|r_0\|_2$ ، الگوریتم لانکزوس دوطرفه برای حل دستگاه $Ax = b$ از روش تصویری

$$b - Ax_m \perp \mathcal{K}_m(A^T, w_1) \quad \text{که} \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

استفاده می‌کند. از اینکه $x_m \in x_0 + \mathcal{K}_m(A, r_0)$ ، بردار x_m را می‌توان به‌صورت

$$x_m = x_0 + V_m y_m,$$

نوشت که در آن $V_m = (v_1, v_2, \dots, v_m)$ و $y_m \in \mathbb{R}^m$. از طرفی، داریم

$$\begin{aligned} b - Ax_m &= b - A(x_0 + V_m y_m) \\ &= r_0 - AV_m y_m \end{aligned}$$

$$\begin{aligned} &= \beta v_1 - AV_m y_m \\ &= \beta V_m e_1 - AV_m y_m, \end{aligned} \quad (۵-۷)$$

که در آن $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^m$. بنابراین، از $r_0 - Ax_m \perp \mathcal{K}_m(A^T, w_1)$ نتیجه می‌گیریم که

$$W_m^T(\beta V_m e_1 - AV_m y_m) = 0,$$

که این خود معادل با

$$\beta e_1 = T_m y_m,$$

است، چرا که $W_m^T V_m = I_m$ و $W_m^T A V_m = T_m$. لذا برای محاسبه بردار y_m ، کافیست دستگاه

آخر را حل کنیم، به عبارت دیگر، داریم

$$y_m = T_m^{-1}(\beta e_1). \quad (۶-۷)$$

با جمع‌بندی مطالب ارائه‌شده الگوریتم لانکزوس دوطرفه برای حل دستگاه معادلات خطی $Ax = b$ به صورت الگوریتم ۲.۷ درمی‌آید.

الگوریتم ۲.۷ لانکزوس دوطرفه برای حل دستگاه معادلات خطی

1. Compute $r_0 = b - Ax_0$ and $\beta := \|r_0\|_2$.
2. Run m steps of the nonsymmetric Lanczos algorithm, i.e.,
start with $v_1 = r_0/\beta$ and any vector w_1 such that $\langle v_1, w_1 \rangle = 1$,
generate the Lanczos vectors $v_1, \dots, v_m, w_1, \dots, w_m$
and the tridiagonal matrix T_m .
3. Compute $y_m = T_m^{-1}(\beta e_1)$ and $x_m = x_0 + V_m y_m$.

همانند الگوریتم GMRES با شروع مجدد و کامل، می‌توان نسخه با شروع مجدد و کامل الگوریتم ۲.۷ را تعریف کرد. از روابط (۵-۷) و (۲۵-۷)، داریم

$$\begin{aligned} b - Ax_m &= \beta V_m e_1 - AV_m y_m \\ &= \beta V_m e_1 - (V_m T_m + \delta_{m+1} v_{m+1} e_m^T) y_m \\ &= V_m(\beta e_1 - T_m y_m) - \delta_{m+1} v_{m+1} e_m^T y_m \end{aligned} \quad (۷-۷)$$

باتوجه به گام ۳ از الگوریتم ۲.۷، داریم

$$\beta e_1 - T_m y_m = 0.$$

ازاین‌رو، از رابطه (۷-۷)، خواهیم داشت

$$b - Ax_m = -\delta_{m+1} v_{m+1} e_m^T y_m. \quad (۸-۷)$$

لذا،

$$\|b - Ax_m\|_2 = |\delta_{m+1} e_m^T y_m| \|v_{m+1}\|_2. \quad (۹-۷)$$

درنتیجه، بدون اینکه بردار مانده را به‌صورت مستقیم محاسبه کنیم، می‌توانیم از رابطه (۹-۷) برای محاسبه نرم-۲ بردار مانده استفاده کنیم.

۳-۷ الگوریتم گرادیان دومزدوج (BiCG)

الگوریتم گرادیان دومزدوج (BiCG) از الگوریتم ۲.۷ و کاملاً شبیه به الگوریتم گرادیان مزدوج استخراج می‌شود. این الگوریتم ابتدا توسط لانگزوس در [۶۷] پیشنهاد شد و سپس به‌صورت کاملاً متفاوت (به‌شکل الگوریتم گرادیان مزدوج) توسط فلچر [۳۷] معرفی شد. الگوریتم BiCG نه‌تنها دستگاه $Ax = b$ را حل می‌کند، بلکه به‌طور ضمنی دستگاه $A^T \tilde{x} = \tilde{b}$ را نیز حل می‌کند. حل دستگاه $A^T \tilde{x} = \tilde{b}$ معمولاً در فرمول‌بندی الگوریتم ارایه نمی‌شود. در این بخش، بردارها و ضرایب ظاهرشده در روش حل دستگاه $A^T \tilde{x} = \tilde{b}$ را با یک علامت تیلدا ($\sim$) همراه می‌کنیم.

فرض کنید

$$T_m = L_m U_m,$$

تجزیه LU ماتریس T_m باشد. دراین‌صورت، با توجه به الگوریتم ۲.۷، داریم

$$\begin{aligned} x_m &= x_0 + V_m T_m^{-1} (\beta e_1) \\ &= x_0 + V_m U_m^{-1} L_m^{-1} (\beta e_1) \\ &= x_0 + P_m z_m, \end{aligned}$$

که در آن $P_m = V_m U_m^{-1}$ و $z_m = L_m^{-1} (\beta e_1)$.

همانند الگوریتم ۱۰.۶، فرض کنید

$$T_m = L_m U_m = \begin{pmatrix} 1 & & & \\ \lambda_2 & 1 & & \\ & \lambda_3 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_m & 1 \end{pmatrix} \begin{pmatrix} \eta_1 & \beta_2 & & \\ & \eta_2 & \beta_3 & \\ & & \ddots & \ddots \\ & & & \eta_{m-1} & \beta_m \\ & & & & \eta_m \end{pmatrix}.$$

از اینکه $P_m = V_m U_m^{-1}$ داریم

$$P_m U_m = V_m.$$

از برابری ستون آخر ماتریس V_m و ستون آخر $P_m U_m$ ، خواهیم داشت

$$p_m = \frac{1}{\eta_m} (v_m - \beta_m p_{m-1}). \quad (۱۰-۷)$$

از طرفی، همانند الگوریتم D-Lanczos (الگوریتم ۱۰.۶)، داریم

$$\begin{aligned} x_m &= x_0 + P_m z_m \\ &= x_0 + \begin{pmatrix} P_{m-1} & p_m \end{pmatrix} \begin{pmatrix} z_{m-1} \\ \zeta_m \end{pmatrix} \\ &= x_0 + P_{m-1} z_{m-1} + p_m \zeta_m \\ &= x_{m-1} + \zeta_m p_m. \end{aligned} \quad (۱۱-۷)$$

اما برای محاسبه یک رابطه بازگشتی برای ζ_m ، از رابطه $z_m = L_m^{-1}(\beta e_1)$ استفاده می‌کنیم. این رابطه معادل با رابطه $L_m z_m = \beta e_1$ است که از برابری آخرین درایه طرفین آن خواهیم داشت

$$\lambda_m \zeta_{m-1} + \zeta_m = 0.$$

لذا،

$$\zeta_m = -\lambda_m \zeta_{m-1}. \quad (۱۲-۷)$$

بنابراین، لازم است یک رابطه بازگشتی برای λ_m بیابیم. برای این کار، از رابطه $T_m = L_m U_m$ داریم

$$\delta_m = L_m \text{ ام } m \text{ سطر} \times U_m \text{ ام } (m-1) \text{ ستون} = \lambda_m \eta_{m-1},$$

که رابطه

$$\lambda_m = \frac{\delta_m}{\eta_{m-1}}, \quad (13-7)$$

را نتیجه می‌دهد. اکنون با استفاده از رابطه‌های (۷-۱۱)، (۷-۱۲) و (۷-۱۳) می‌توان الگوریتمی مشابه با الگوریتم D-Lanczos ارائه کرد. با کمک نتایج به‌دست‌آمده و مشابه با الگوریتم گرادیان مزدوج، الگوریتم گرادیان دومزدوج (BiCG) را برای حل دستگاه نامتقارن $Ax = b$ ارائه می‌کنیم. دستگاه $A^T \tilde{x} = \tilde{b}$ را در نظر بگیرید. در الگوریتم لانکزوس دوطرفه (الگوریتم ۲.۷)، برای حل این دستگاه از روش تصویری

$$\tilde{b} - A^T \tilde{x}_m \perp \mathcal{K}_m(A, v_1) \text{ را به قسمی پیدا کنید که } \tilde{x}_m \in \tilde{x}_0 + \mathcal{K}_m(A^T, \tilde{r}_0)$$

استفاده می‌شود. در این صورت، همانند روش تصویری ارائه‌شده برای دستگاه $Ax = b$ ، خواهیم داشت

$$\tilde{x}_m = \tilde{x}_0 + W_m \tilde{y}_m,$$

که در آن

$$\tilde{y}_m = T_m^{-T}(\tilde{\beta} e_1),$$

به‌طوری‌که $\tilde{\beta} = \|\tilde{r}_0\|_2$ و $\tilde{r}_0 = \tilde{b} - A^T \tilde{x}_0$. اکنون، با استفاده از تجزیه LU ماتریس T_m ، داریم

$$\begin{aligned} \tilde{x}_m &= \tilde{x}_0 + W_m T_m^{-T}(\tilde{\beta} e_1) \\ &= \tilde{x}_0 + W_m L_m^{-T} U_m^{-T}(\tilde{\beta} e_1). \end{aligned}$$

با فرض $\tilde{P}_m = W_m L_m^{-T}$ و $\tilde{z}_m = U_m^{-T}(\tilde{\beta} e_1)$ می‌توان بردار $\tilde{x}_m$ را به‌صورت

$$\tilde{x}_m = \tilde{x}_0 + \tilde{P}_m \tilde{z}_m, \quad (14-7)$$

نوشت. اینک، با توجه به این رابطه می‌توان الگوریتمی مشابه با الگوریتم D-Lanczos برای حل دستگاه $A^T \tilde{x} = \tilde{b}$ ارائه کرد.

در ادامه، نشان می‌دهیم که ستون‌های ماتریس $P_m = V_m U_m^{-1}$ نسبت به ستون‌های ماتریس $\tilde{P}_m = W_m L_m^{-T}$ متعامد است. برای نشان دادن این نتیجه، داریم

$$\begin{aligned} (\tilde{P}_m)^T A P_m &= (W_m L_m^{-T})^T A V_m U_m^{-1} \\ &= L_m^{-1} W_m^T A V_m U_m^{-1} \\ &= L_m^{-1} T_m U_m^{-1} \\ &= L_m^{-1} L_m = I. \end{aligned}$$

از طرفی، با استفاده از رابطه (۷-۸) روشن است که بردارهای مانده تولیدشده توسط الگوریتم لانکزوس دوطرفه برای حل دستگاه $Ax = b$ و $A^T \tilde{x} = \tilde{b}$ در رابطه زیر صدق می‌کنند

$$\langle \tilde{r}_j, r_j \rangle = 0, \quad \forall i \neq j,$$

زیرا بردار r_j با v_{j+1} هم‌راستا و بردار $\tilde{r}_j$ با w_{j+1} هم‌راستا است.

اکنون، با کمک نتایج ارائه‌شده، آماده‌ایم که الگوریتم BiCG را معرفی کنیم. فرض کنید x_j جواب تقریبی به دست آمده در تکرار j ام الگوریتم لانکزوس دوطرفه باشد. در این صورت، با استفاده از رابطه (۷-۱۱) می‌توان نوشت

$$x_{j+1} = x_j + \alpha_j p_j. \quad (۷-۱۵)$$

با محاسبه بردار مانده متناظر با x_{j+1} ، داریم

$$r_{j+1} = b - A x_{j+1} = r_j - \alpha_j A p_j,$$

اما از اینکه $\langle \tilde{r}_j, r_{j+1} \rangle = 0$ ، نتیجه می‌گیریم

$$0 = \langle \tilde{r}_j, r_j - \alpha_j A p_j \rangle = \langle \tilde{r}_j, r_j \rangle - \alpha_j \langle \tilde{r}_j, A p_j \rangle.$$

در نتیجه،

$$\alpha_j = \frac{\langle \tilde{r}_j, r_j \rangle}{\langle \tilde{r}_j, A p_j \rangle} = \frac{\langle r_j, \tilde{r}_j \rangle}{\langle A p_j, \tilde{r}_j \rangle}. \quad (۷-۱۶)$$

از رابطه $\tilde{P}_m = W_m L_m^{-T}$ ، داریم $\tilde{P}_m L_m^T = W_m$. اگر ستون‌های آخر دو طرف این رابطه را با هم برابر قرار دهیم، خواهیم داشت

$$\tilde{p}_m = w_m - \lambda_m \tilde{p}_{m-1}.$$

بنابراین، می‌توانیم شبیه به الگوریتم D-Lanczos، بردار $\tilde{p}_{j+1}$ را به صورت زیر بنویسیم

$$\tilde{p}_{j+1} = \tilde{r}_{j+1} + \tilde{\beta}_j \tilde{p}_j. \quad (17-17)$$

بنابراین، از این رابطه و اینکه $\langle Ap_j, \tilde{p}_{j-1} \rangle = 0$ ، داریم

$$\begin{aligned} \langle Ap_j, \tilde{r}_j \rangle &= \langle Ap_j, \tilde{p}_j - \tilde{\beta}_j \tilde{p}_{j-1} \rangle \\ &= \langle Ap_j, \tilde{p}_j \rangle - \tilde{\beta}_j \langle Ap_j, \tilde{p}_{j-1} \rangle \\ &= \langle Ap_j, \tilde{p}_j \rangle. \end{aligned} \quad (18-17)$$

با جایگذاری این رابطه در (۱۶-۷) خواهیم داشت

$$\alpha_j = \frac{\langle r_j, \tilde{r}_j \rangle}{\langle Ap_j, \tilde{p}_j \rangle}. \quad (19-17)$$

به طور مشابه، رابطه (۱۰-۷) نشان می‌دهد که می‌توان بردار جهت در تکرار $(j+1)$ ام را به صورت

$$p_{j+1} = r_{j+1} + \beta_j p_j. \quad (20-17)$$

نوشت. همچنین، به سادگی از رابطه (۱۴-۷) می‌توان دید که بردار جواب دستگاه $A^T \tilde{x} = \tilde{b}$ در تکرار $(j+1)$ ام به صورت

$$\tilde{x}_{j+1} = \tilde{x}_j + \tilde{\alpha}_j \tilde{p}_j, \quad (21-17)$$

نوشته می‌شود.

نشان می‌دهیم که مقادیر α_j و $\tilde{\alpha}_j$ در رابطه‌های (۱۵-۷) و (۲۱-۷) با هم برابر هستند. برای

نشان دادن این حقیقت، با توجه به رابطه (۱۸-۷)، داریم

$$\tilde{\alpha}_j = \frac{\langle \tilde{r}_j, r_j \rangle}{\langle A^T \tilde{p}_j, r_j \rangle} = \frac{\langle \tilde{r}_j, r_j \rangle}{\langle A^T \tilde{p}_j, p_j \rangle} = \frac{\langle \tilde{r}_j, r_j \rangle}{\langle \tilde{p}_j, Ap_j \rangle} = \alpha_j. \quad (22-17)$$

برای محاسبه β_j ابتدا توجه کنید که

$$r_{j+1} = r_j - \alpha_j Ap_j, \quad (23-17)$$

$$\tilde{r}_{j+1} = \tilde{r}_j - \alpha_j A^T \tilde{p}_j. \quad (24-17)$$

اینک، داریم

$$\circ = \langle A^T \tilde{p}_j, p_{j+1} \rangle = \langle A^T \tilde{p}_j, r_{j+1} + \beta_j p_j \rangle = \langle A^T \tilde{p}_j, r_{j+1} \rangle + \beta_j \langle A^T \tilde{p}_j, p_j \rangle.$$

لذا، با استفاده از رابطه (۲۴-۷)، داریم

$$\begin{aligned} \beta_j &= -\frac{\langle A^T \tilde{p}_j, r_{j+1} \rangle}{\langle A^T \tilde{p}_j, p_j \rangle} \\ &= -\frac{\langle A^T \tilde{p}_j, r_{j+1} \rangle}{\langle \tilde{p}_j, Ap_j \rangle} \\ &= -\frac{1}{\langle \tilde{p}_j, Ap_j \rangle} \langle \frac{1}{\alpha_j} (\tilde{r}_j - \tilde{r}_{j+1}), r_{j+1} \rangle \\ &= \frac{1}{\alpha_j \langle \tilde{p}_j, Ap_j \rangle} (\langle \tilde{r}_{j+1}, r_{j+1} \rangle - \langle \tilde{r}_j, r_{j+1} \rangle) \\ &= \frac{\langle \tilde{r}_{j+1}, r_{j+1} \rangle}{\alpha_j \langle \tilde{p}_j, Ap_j \rangle}. \end{aligned}$$

حال با جایگذاری مقدار α_j از رابطه (۲۲-۷) در رابطه اخیر، خواهیم داشت

$$\beta_j = \frac{\langle \tilde{r}_{j+1}, r_{j+1} \rangle}{\langle \tilde{r}_j, r_j \rangle}.$$

به‌طور مشابه می‌توان نشان داد که $\tilde{\beta}_j = \beta_j$. با جمع‌بندی نتایج به‌دست‌آمده، الگوریتم BiCG را به‌صورت الگوریتم ۴.۷ می‌نویسیم.

الگوریتم ۴.۷ الگوریتم BiCG برای حل دستگاه $Ax = b$

1. Choose an initial guess x_0 and $r_0 := b - Ax_0$.
Choose $\tilde{r}_0$ such that $\langle r_0, \tilde{r}_0 \rangle \neq 0$.
2. Set $p_0 = r_0$ and $\tilde{p}_0 = \tilde{r}_0$.
3. For $j = 0, 1, 2, \dots$, until convergence, Do
 4. $\alpha_j := \langle r_j, \tilde{r}_j \rangle / \langle Ap_j, \tilde{p}_j \rangle$
 5. $x_{j+1} := x_j + \alpha_j p_j$
 6. $r_{j+1} := r_j - \alpha_j Ap_j$
 7. $\tilde{r}_{j+1} := \tilde{r}_j - \alpha_j A^T \tilde{p}_j$
 8. $\beta_j := \langle r_{j+1}, \tilde{r}_{j+1} \rangle / \langle r_j, \tilde{r}_j \rangle$
 9. $p_{j+1} := r_{j+1} + \beta_j p_j$
 10. $\tilde{p}_{j+1} := \tilde{r}_{j+1} + \beta_j \tilde{p}_j$
11. EndDo

ملاحظه ۱.۷ اگر بخواهیم دستگاه $A^T \tilde{x} = \tilde{b}$ را حل کنیم، کافی است در گام ۱ حدس اولیه $\tilde{x}_0$ را انتخاب کرده و قرار دهیم $\tilde{r}_0 = \tilde{b} - A^T \tilde{x}_0$. همچنین، در گام ۵ قرار دهیم $\tilde{x}_{j+1} := \tilde{x}_j + \alpha_j \tilde{p}_j$.

۴-۷ الگوریتم مانده شبه‌مینیمال (QMR)

در بخش ۱-۷ دیدیم که

$$AV_m = V_{m+1} \bar{T}_m, \quad (۲۵-۷)$$

که در آن

$$\bar{T}_m = \begin{pmatrix} T_m \\ \delta_{m+1} e_m^T \end{pmatrix}.$$

فرض کنید $v_1 = r_0 / \beta$ که در آن $\beta = \|r_0\|_2$. همچنین، فرض کنید $x \in x_0 + \mathcal{K}_m(A, r_0)$. در این صورت، یک $y \in \mathbb{R}^m$ وجود دارد به‌طوری‌که

$$x = x_0 + V_m y.$$

لذا از رابطه (۲۵-۷)، داریم

$$\begin{aligned} b - Ax &= b - A(x_0 + V_m y) \\ &= r_0 - AV_m y \\ &= \beta v_1 - V_{m+1} \bar{T}_m y \\ &= \beta V_{m+1} e_1 - V_{m+1} \bar{T}_m y \\ &= V_{m+1} (\beta e_1 - \bar{T}_m y). \end{aligned}$$

بنابراین،

$$\|b - Ax\|_2 = \|V_{m+1} (\beta e_1 - \bar{T}_m y)\|_2.$$

اگر ستون‌های ماتریس V_{m+1} یک‌امتعامد باشند، آنگاه

$$\|b - Ax\|_2 = \|\beta e_1 - \bar{T}_m y\|_2.$$

در این صورت، بردار y را به گونه‌ای تعیین می‌کنیم که مقدار $\|\beta e_1 - \bar{T}_m y\|_2$ را کمینه کند. الگوریتمی که به این طریق به دست می‌آید، همان الگوریتم GMRES است. اما در اینجا متأسفانه ستون‌های V_{m+1} یک‌متعامد نیستند، اما با این حال معقول به نظر می‌رسد که عبارت

$$J(y) = \|\beta e_1 - \bar{T}_m y\|_2, \quad (۲۶-۷)$$

را روی y کمینه کرده و اگر y_m کمینه‌کننده $J(y)$ باشد، آنگاه

$$x_m = x_0 + V_m y_m,$$

را به عنوان یک جواب تقریبی بپذیریم.

ماتریس $\bar{T}_m$ در رابطه ۲۶-۷، یک ماتریس $(m+1) \times m$ و به صورت

$$\bar{T}_m = \begin{pmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \\ & & & & \delta_{m+1} \end{pmatrix},$$

است. ماتریس‌های گایونز Ω_i ، $i = 1, 2, \dots, m$ را به گونه‌ای می‌سازیم که ماتریس

$$\bar{R}_m = \Omega_m \Omega_{m-1} \cdots \Omega_1 \bar{T}_m,$$

یک ماتریس شبه‌بالامثلثی باشد (در حقیقت درایه‌های مکان $(i+1, i)$ ، $i = 1, 2, \dots, m$ ، صفر شوند). در این صورت، خواهیم داشت

$$\begin{aligned} J(y) &= \|\beta e_1 - \bar{T}_m y\|_2 \\ &= \|\Omega_m \Omega_{m-1} \cdots \Omega_1 (\beta e_1 - \bar{T}_m y)\|_2 \\ &= \|\bar{g}_m - \bar{R}_m y\|_2, \end{aligned}$$

که در آن

$$\bar{g}_m = \Omega_m \Omega_{m-1} \cdots \Omega_1 (\beta e_1).$$

۲۷۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

به‌ازای $m = 5$ ماتریس $\bar{R}_m$ و بردار $\bar{g}_m$ به‌صورت زیر خواهند بود

$$\bar{R}_m = \begin{pmatrix} r_{11} & r_{12} & r_{13} & & \\ & r_{22} & r_{23} & r_{24} & \\ & & r_{33} & r_{34} & r_{35} \\ & & & r_{44} & r_{45} \\ & & & & r_{55} \\ & & & & & 0 \end{pmatrix}, \quad \bar{g}_m = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_5 \\ \gamma_6 \end{pmatrix}.$$

درحقیقت، $\bar{R}_m$ ماتریس شبه‌بالامثلثی موجود در تجزیه QR ماتریس $\bar{T}_m$ است. درنتیجه، خواهیم داشت

$$J(y) = \|\bar{g}_m - \bar{R}_m y\|_2 = \sqrt{\|g_m - R_m y\|_2^2 + |\gamma_{m+1}|^2},$$

که در آن R_m و g_m به‌ترتیب از حذف سطر آخر $\bar{R}_m$ و $\bar{g}_m$ به‌دست می‌آیند. بنابراین، زمانی $J(y)$ کمینه می‌شود که $g_m - R_m y = 0$ ، یا به‌عبارت‌دیگر،

$$y_m := y = R_m^{-1} g_m.$$

دراین‌صورت، خواهیم داشت

$$\min_{y \in \mathbb{R}^m} J(y) = J(y_m) = |\gamma_{m+1}|.$$

ازاین‌رو، داریم

$$x_m = x_0 + V_m y_m. \quad (27-7)$$

با جمع‌بندی مطالب ارائه‌شده، می‌توان گفت که اساس کار الگوریتم QMR به این صورت است که ابتدا دستگاه بالامثلثی

$$R_m y_m = g_m,$$

را حل کرده و سپس بردار x_m را از رابطه (27-7) محاسبه می‌کنیم.

در ادامه، نسخه رایج الگوریتم QMR را ارائه می‌کنیم. در این نسخه، ابتدا یک جواب تقریبی در فضای کرایلف یک‌بُعدی به‌دست می‌آید و در هر تکرار با افزایش یک واحد به بعد فضای کرایلف، جواب تقریبی

به‌هنگام می‌شود. در تکرار m ام، با فرض

$$P_m = V_m R_m^{-1}, \quad (۲۸-۷)$$

داریم

$$\begin{aligned} x_m &= x_{\circ} + V_m y_m \\ &= x_{\circ} + V_m R_m^{-1} g_m \\ &= x_{\circ} + P_m g_m. \end{aligned}$$

اکنون، فرض کنید

$$g_m = \begin{pmatrix} g_{m-1} \\ \gamma_m \end{pmatrix},$$

که در آن

$$\gamma_m = c_m \gamma_m^{(m-1)},$$

به‌طوری‌که $\gamma_m^{(m-1)}$ آخرین مؤلفه بردار $\bar{g}_{m-1}$ است، یعنی آخرین مؤلفه بردار سمت راست، پیش از اینکه ماتریس دوران m ام بر آن اثر کرده باشد. در این صورت، اگر ماتریس P_m را به‌صورت $P_m = (p_1, p_2, \dots, p_m)$ بنویسیم، آنگاه خواهیم داشت

$$\begin{aligned} x_m &= x_{\circ} + \begin{pmatrix} P_{m-1} & p_m \end{pmatrix} \begin{pmatrix} g_{m-1} \\ \gamma_m \end{pmatrix} \\ &= x_{\circ} + P_{m-1} g_{m-1} + \gamma_m p_m \\ &= x_{m-1} + \gamma_m p_m. \end{aligned} \quad (۲۹-۷)$$

بنابراین، لازم است فقط بردار p_m و مقدار γ_m را محاسبه کنیم. ابتدا شیوه محاسبه p_m را شرح می‌دهیم. از رابطه (۲۸-۷)، داریم

$$P_m R_m = V_m,$$

اگر ستون آخر دوطرف این رابطه را با هم برابر قرار دهیم، خواهیم داشت

$$p_m = \frac{1}{r_{mm}} (v_m - r_{m-2,m} p_{m-2} - r_{m-1,m} p_{m-1}).$$

بنابراین، لازم است مقادیر $r_{i,m}$ را به‌ازای $m-1, m-2, \dots, m$ محاسبه کنیم. بعد از محاسبه جواب تقریبی x_{m-1} ، اگر در تکرار m ام، بعد فضای کرایلف را یک واحد افزایش دهیم، ماتریس $\bar{T}_m$ تولید می‌شود که ستون آخر آن به‌صورت

$$\begin{pmatrix} \circ \\ \vdots \\ \circ \\ \beta_m \\ \alpha_m \\ \delta_{m+1} \end{pmatrix},$$

است. اکنون برای به‌هنگام کردن ماتریس $\bar{R}_{m-1}$ ، ماتریس‌های گیونز Ω_{m-1} و Ω_{m-2} (به‌ترتیب) و سپس ماتریس Ω_m (برای صفرکردن درایه δ_{m+1}) را روی بردار بالا اثر می‌دهیم. در این صورت، درایه‌های مکان $m-2, m-1, m$ و m به‌ترتیب همان درایه‌های $r_{m-2,m}, r_{m-1,m}$ و $r_{m,m}$ خواهد بود. در ادامه، نحوه محاسبه γ_m را شرح می‌دهیم. در تکرار m ام و بعد از افزایش یک‌واحدی در بُعد فضای کرایلف، بردار βe_1 و بعد از اعمال ماتریس‌های گیونز Ω_i ، $i = 1, 2, \dots, m$ ، به‌صورت زیر درمی‌آید

$$\begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{m-1} \\ \gamma_m \\ \circ \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{m-1} \\ \gamma_m \\ \circ \end{pmatrix},$$

که پس از به‌کار بردن ماتریس گیونز Ω_m روی این بردار، خواهیم داشت

$$\bar{g}_m = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{m-1} \\ c_m \gamma_m \\ -s_m \gamma_m \end{pmatrix}.$$

در نتیجه، مقدار γ_m در رابطه (۷-۲۹) برابر با درایه m ام بردار اخیر است.

با جمع‌بندی مطالب ذکر شده، اکنون می‌توانیم نسخه متداول الگوریتم QMR را ارائه کنیم. پیش از ارائه الگوریتم متذکر می‌شویم که در عمل نیازی به ذخیره‌سازی ماتریس $\bar{T}_k$ نیست و در هر تکرار، تنها کافی است دو ماتریس گیونز مربوط به دو تکرار قبل ذخیره شوند.

الگوریتم ۴.۷ الگوریتم QMR

1. $r_0 := b - Ax_0$, $\gamma_1 := \|r_0\|_2$ and $w_1 := v_1 := r_0/\gamma_1$
2. For $m = 1, 2, \dots$, until convergence, Do
3. Compute α_m, δ_{m+1} and v_{m+1}, w_{m+1}
4. Update the QR factorization of $\bar{T}_m$; i.e.,
5. Apply $\Omega_i, i = m - 2, m - 1$, to the m th column of $\bar{T}_m$
6. Compute the rotation coefficients c_m, s_m
7. Apply rotation Ω_m to the last column of $\bar{T}_m$ and to $\bar{g}_m$; i.e.,
8. $\gamma_{m+1} := -s_m \gamma_m$
9. $\gamma_m := c_m \gamma_m$
10. $\alpha_m := c_m \alpha_m + s_m \delta_{m+1} \left(= \sqrt{\delta_{m+1}^2 + \alpha_m^2} \right)$
11. $p_m = (v_m - t_{m-2,m} p_{m-2} - t_{m-1,m} p_{m-1}) / t_{mm}$.
12. $x_m := x_{m-1} + \gamma_m p_m$
13. If $|\gamma_{m+1}|$ is small enough then stop.
14. EndDo

در گام ۲ از الگوریتم QMR، مقادیر α_m, δ_{m+1} و بردارهای v_{m+1}, w_{m+1} توسط الگوریتم ۱.۷ محاسبه می‌شوند. به علاوه، در گام ۶، مقادیر c_m و s_m از رابطه‌های زیر به دست می‌آیند

$$s_m = \frac{t_{m+1,m}}{\sqrt{t_{mm}^2 + t_{m+1,m}^2}}, \quad c_m = \frac{t_{mm}}{\sqrt{t_{mm}^2 + t_{m+1,m}^2}}.$$

قضیه ۲.۷ نرم‌مانده تقریب x_m در الگوریتم QMR در رابطه زیر صدق می‌کند

$$\|b - Ax_m\|_2 \leq \|V_{m+1}\|_2 |s_1 s_2 \cdots s_m| \|r_0\|_2.$$

برهان. دیدیم که

$$b - Ax_m = V_{m+1}(\beta e_1 - \bar{T}_m y_m).$$

از طرفی، با توجه به نتایج به دست آمده، داریم

$$\|\beta e_1 - \bar{T}_m y_m\|_2^2 = |\gamma_{m+1}|^2 + \|g_m - R_m y_m\|_2^2.$$

۲۷۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

اما، بردار y_m طوری اختیار می‌شود که $g_m - R_m y_m = 0$. بنابراین،

$$\|\beta e_1 - \bar{T}_m y_m\|_2 = |\gamma_{m+1}|.$$

در نتیجه،

$$\begin{aligned} \|b - Ax_m\|_2 &= \|V_{m+1}(\beta e_1 - \bar{T}_m y_m)\|_2 \\ &\leq \|V_{m+1}\|_2 \|\beta e_1 - \bar{T}_m y_m\|_2 \\ &= \|V_{m+1}\|_2 |\gamma_{m+1}| \end{aligned}$$

اما، داریم

$$\begin{aligned} |\gamma_{m+1}| &= |-s_m \gamma_m| \\ &= |s_m s_{m-1} \gamma_{m-1}| \\ &\vdots \\ &= |s_m s_{m-1} \cdots s_1 \gamma_0| \\ &= |s_m s_{m-1} \cdots s_1 \|r_0\|_2| \end{aligned}$$

□ که نتیجه لازم را به دست می‌دهد.

بدون از دست دادن کلیت، می‌توان بردارهای v_i ها را یک‌ه در نظر گرفت؛ به عبارت دیگر، $\|v_i\|_2 = 1$ ، $i = 1, 2, \dots, m+1$. در این صورت، با توجه به تمرین ۱.۱، خواهیم داشت

$$\begin{aligned} \|b - Ax_m\|_2 &\leq \|V_{m+1}\|_2 |\gamma_{m+1}| \\ &\leq \|V_{m+1}\|_F |\gamma_{m+1}| \\ &= \sqrt{m+1} |\gamma_{m+1}|. \end{aligned}$$

در گام ۱۳ از الگوریتم QMR، متذکر می‌شویم که برخلاف الگوریتم GMRES مقدار $|\gamma_{m+1}|$ اندازه نرم بردار مانده نیست. بنابراین، با توجه به رابطه اخیر می‌توان از شرط

$$\frac{|\gamma_{m+1}|}{\|r_0\|_2} < \frac{1}{\sqrt{m+1}} \epsilon,$$

به‌عنوان یک شرط توقف استفاده کرد. درحقیقت، این شرط توقف، رابطه زیر را ارضاء خواهد کرد

$$\frac{\|r_{m+1}\|_2}{\|r_0\|_2} < \epsilon.$$

۵-۷ الگوریتم گرادیان مزدوج مربع‌شده (CGS)

در هر تکرار از الگوریتم‌های BiCG و QMR به حاصل ضرب هر دو ماتریس A و A^T در یک بردار نیاز است. بنابراین، لازم است هر دو ماتریس A و A^T به‌صورت تنک ذخیره شوند. با مراجعه به الگوریتم‌های BiCG و QMR درمی‌یابیم که اگر بخواهیم دستگاه $Ax = b$ را حل کنیم، ماتریس A^T فقط در محاسبه α_j و β_j ظاهر می‌شود. در ادامه الگوریتم CGS را معرفی می‌کنیم که در آن نیازی به ماتریس A^T نیست و مقادیر α_j و β_j بدون مراجعه به ماتریس A^T محاسبه می‌شوند. این الگوریتم توسط سونیولد در سال ۱۹۸۴ ارائه شده است [۸۹].

قضیه ۳.۷ در الگوریتم BiCG، داریم

$$r_j = \varphi_j(A)r_0, \quad (۳۰-۷)$$

$$p_j = \pi_j(A)r_0 = \pi_j(A)p_0, \quad (۳۱-۷)$$

که در آن φ و π چندجمله‌ای‌هایی از درجه j هستند به‌طوری‌که $\varphi_j(0) = 1$. برهان. به استقرا روی j ، این قضیه را ثابت می‌کنیم. فرض کنید $j = 1$. دراین‌صورت، از گام ۶ الگوریتم BiCG، داریم

$$r_1 = r_0 - \alpha_0 A p_0 = r_0 - \alpha_0 A r_0 = (I - \alpha_0 A) r_0.$$

بنابراین، با فرض $\varphi_1(t) = 1 - \alpha_0 t$ ، داریم $\varphi_1(0) = 1$ و $r_1 = \varphi_1(A)r_0$. ازطرفی، از گام ۹ الگوریتم، داریم

$$p_1 = r_1 + \beta_0 p_0 = \varphi_1(A)r_0 + \beta_0 r_0 = (\beta_0 I + \varphi_1(A)) r_0.$$

درنتیجه، با فرض $\pi_1(t) = \beta_0 + \varphi_1(t)$ ، خواهیم داشت $p_1 = \pi_1(A)r_0$. اکنون، فرض کنید قضیه برای $j = k$ درست باشد. دراین‌صورت، از گام ۶ الگوریتم BiCG، داریم

$$r_{k+1} = r_k - \alpha_k A p_k$$

$$\begin{aligned} &= \varphi_k(A)r_0 - \alpha_k A\pi_k(A)r_0 \\ &= (\varphi_k(A) - \alpha_k A\pi_k(A))r_0. \end{aligned}$$

بنابراین، با فرض $\varphi_{k+1}(t) = \varphi_k(t) - \alpha_k t \pi_k(t)$ خواهیم داشت $\varphi_{k+1}(0) = 1$ و $r_{k+1} = \varphi_{k+1}(A)r_0$. همچنین، از گام ۹ الگوریتم، داریم

$$\begin{aligned} p_{k+1} &= r_{k+1} + \beta_k p_k \\ &= \varphi_{k+1}(A)r_0 + \beta_k \pi_k(A)r_0 \\ &= (\varphi_{k+1}(A) + \beta_k \pi_k(A))r_0. \end{aligned}$$

درنتیجه، با فرض $\pi_{k+1}(t) = \varphi_{k+1}(t) + \beta_k \pi_k(t)$ خواهیم داشت $p_{k+1} = \pi_{k+1}(A)r_0$. که اثبات قضیه را کامل می‌کند. $\square$

به‌طور مشابه، می‌توان دید که

$$\tilde{r}_j = \varphi_j(A^T)\tilde{r}_0, \quad \tilde{p}_j = \pi_j(A^T)\tilde{r}_0, \quad (32-7)$$

که در آن φ_j و π_j همان چندجمله‌ای‌هایی هستند که در رابطه‌های (۴۶-۷) و (۴۷-۷) ظاهر شده‌اند. اکنون، با توجه به قضیه ۳.۷ و رابطه (۳۲-۷)، داریم

$$\alpha_j = \frac{\langle r_j, \tilde{r}_j \rangle}{\langle Ap_j, \tilde{p}_j \rangle} = \frac{\langle \varphi_j(A)r_0, \varphi_j(A^T)\tilde{r}_0 \rangle}{\langle A\pi_j(A)r_0, \pi_j(A^T)\tilde{r}_0 \rangle} = \frac{\langle \varphi_j^\top(A)r_0, \tilde{r}_0 \rangle}{\langle A\pi_j^\top(A)r_0, \tilde{r}_0 \rangle}.$$

بنابراین، اگر بتوانیم یک رابطه بازگشتی برای $\varphi_j(A)r_0$ و $\pi_j^\top(A)r_0$ بیابیم، آنگاه بدون اینکه از A^T استفاده کنیم، می‌توانیم α_j را محاسبه کنیم. از گام‌های ۶ و ۹ الگوریتم BiCG، داریم

$$\varphi_{j+1}(t) = \varphi_j(t) - \alpha_j t \pi_j(t) \quad (33-7)$$

$$\pi_{j+1}(t) = \varphi_{j+1}(t) + \beta_j \pi_j(t). \quad (34-7)$$

اگر طرفین این دو رابطه را به توان ۲ برسانیم، خواهیم داشت

$$\varphi_{j+1}^2(t) = \varphi_j^2(t) - 2\alpha_j t \pi_j(t) \varphi_j(t) + \alpha_j^2 t^2 \pi_j^2(t), \quad (35-7)$$

$$\pi_{j+1}^2(t) = \varphi_{j+1}^2(t) + 2\beta_j \varphi_{j+1}(t) \pi_j(t) + \beta_j^2 \pi_j^2(t). \quad (36-7)$$

اما داریم

$$\varphi_j(t)\pi_j(t) = \varphi_j(t) (\varphi_j(t) + \beta_{j-1}\pi_{j-1}(t)) = \varphi_j^2(t) + \beta_{j-1}\varphi_j(t)\pi_{j-1}(t).$$

با جایگذاری این رابطه در رابطه‌های (۳۵-۷) و (۳۶-۷)، خواهیم داشت (در این رابطه‌ها متغیر t نمایش داده نشده‌اند)

$$\varphi_{j+1}^2 = \varphi_j^2 - \alpha_j t \left(2\varphi_j^2 + 2\beta_{j-1}\varphi_j\pi_{j-1} - \alpha_j t \pi_j^2 \right), \quad (37-7)$$

$$\varphi_{j+1}\pi_j = \varphi_j^2 + \beta_{j-1}\varphi_j\pi_{j-1} - \alpha_j t \pi_j^2, \quad (38-7)$$

$$\pi_{j+1}^2 = \varphi_{j+1}^2 + 2\beta_j\varphi_{j+1}\pi_j + \beta_j^2\pi_j^2. \quad (39-7)$$

رابطه‌های بازگشتی بالا اساس الگوریتم CGS است. اگر بردارهای

$$r_j = \varphi_j^2(A)r_0,$$

$$p_j = \pi_j^2(A)r_0,$$

$$q_j = \varphi_{j+1}(A)\pi_j(A)r_0,$$

را تعریف کنیم، آنگاه از رابطه‌های (۳۷-۷)، (۳۸-۷) و (۳۹-۷) خواهیم داشت

$$r_{j+1} = r_j - \alpha_j A \left(2r_j + 2\beta_{j-1}q_{j-1} - \alpha_j A p_j \right), \quad (40-7)$$

$$q_j = r_j + \beta_{j-1}q_{j-1} - \alpha_j A p_j, \quad (41-7)$$

$$p_{j+1} = r_{j+1} + 2\beta_j q_j + \beta_j^2 p_j. \quad (42-7)$$

برای اینکه الگوریتم را به صورت مناسب‌تری نمایش دهیم، قرار می‌دهیم

$$d_j = 2r_j + 2\beta_{j-1}q_{j-1} - \alpha_j A p_j.$$

در این صورت، از رابطه (۴۰-۷) خواهیم داشت

$$r_{j+1} = r_j - \alpha_j A d_j.$$

با جمع‌بندی مطالب بالا، با شروع از حدس اولیه x_0 و با فرض $r_0 = b - Ax_0$ ، $p_0 = 0$ ، $q_0 = 0$ و $\beta_0 = 0$ رابطه‌های بازگشتی زیر را برای محاسبه جواب تقریبی خواهیم داشت

- $\alpha_j := \langle r_j, \tilde{r}_0 \rangle / \langle Ap_j, \tilde{r}_0 \rangle,$
- $d_j := 2r_j + 2\beta_{j-1}q_{j-1} - \alpha_j Ap_j,$
- $q_j := r_j + \beta_{j-1}q_{j-1} - \alpha_j Ap_j,$
- $x_{j+1} := x_j + \alpha_j d_j,$
- $r_{j+1} := r_j - \alpha_j Ad_j,$
- $\beta_j := \langle r_{j+1}, \tilde{r}_0 \rangle / \langle r_j, \tilde{r}_0 \rangle,$
- $p_{j+1} := r_{j+1} + \beta_j (2q_j + \beta_j p_j).$

با فرض

$$u_j = r_j + \beta_{j-1}q_{j-1},$$

خواهیم داشت

- $d_j := u_j + q_j,$
- $q_j := u_j - \alpha_j Ap_j,$
- $p_{j+1} := r_{j+1} + \beta_j (q_j + \beta_j p_j).$

با جمع‌بندی مطالب بالا الگوریتم CGS را به‌صورت زیر جمع‌بندی می‌کنیم.

الگوریتم ۵.۷ الگوریتم CGS برای حل دستگاه $Ax = b$

1. Choose an initial guess $x_0, r_0 := b - Ax_0$ and $\tilde{r}_0$ arbitrary
2. Set $p_0 := u_0 := r_0$
3. For $j = 0, 1, 2, \dots$, until convergence, Do
 4. $\alpha_j := \langle r_j, \tilde{r}_0 \rangle / \langle Ap_j, \tilde{r}_0 \rangle$
 5. $q_j := u_j - \alpha_j Ap_j$
 6. $x_{j+1} := x_j + \alpha_j (u_j + q_j)$
 7. $r_{j+1} := r_j - \alpha_j A(u_j + q_j)$
 8. $\beta_j := \langle r_{j+1}, \tilde{r}_0 \rangle / \langle r_j, \tilde{r}_0 \rangle$
 9. $u_{j+1} := r_{j+1} + \beta_j q_j$
 10. $p_{j+1} := u_{j+1} + \beta_j (q_j + \beta_j p_j)$
11. EndDo

همان‌گونه که می‌بینیم در این الگوریتم هیچ محاسبه‌ای با ماتریس A^T انجام نمی‌گیرد و در هر تکرار تنها دو ضرب ماتریس در بردار لازم است. بنابراین، از لحاظ حجم محاسبات تقریباً با الگوریتم BiCG

یکسان است. معمولاً الگوریتم CGS برای حل تعداد زیادی از دستگاه‌های معادلات خطی با ماتریس نامتقارن همگرا می‌شود و سرعت همگرایی آن تقریباً دو برابر سرعت همگرایی BiCG باشد. در هر صورت، چون چندجمله‌ای مانده در الگوریتم CGS به توان ۲ می‌رسد، اگر در یک تکرار از روش BiCG نرم مانده افزایش یابد، آنگاه نرم مانده CGS افزایش خیلی بیشتری خواهد داشت. از این رو، منحنی نرم مانده روش CGS نوسانات زیادی خواهد داشت که معمولاً منجر به ناپایداری خواهد شد. برای رفع این مشکل، فان در فورست در [۹۴] الگوریتم BiCG پایدار شده (BiCGSTAB) را معرفی کرد که در ادامه آن را مطالعه می‌کنیم.

۶-۷ الگوریتم BiCG پایدار شده (BiCGSTAB)

برای استخراج الگوریتم BiCGSTAB به جای اینکه بردار مانده به صورت $r'_j = \varphi_j^2(A)r_0$ اختیار شود، به صورت

$$r'_j = \psi_j(A)\varphi_j(A)r_0,$$

انتخاب می‌شود که در آن φ_j چندجمله‌ای مربوط به بردار مانده BiCG در تکرار j ام و چندجمله‌ای ψ_j یک چندجمله‌ای جدید است که به طور بازگشتی و به منظور هموارسازی یا پایداری رفتار همگرایی الگوریتم به کار می‌رود. در حالت خاص، چندجمله‌ای ψ_j به صورت

$$\psi_0(t) = 1, \quad \psi_{j+1}(t) = (1 - w_j t)\psi_j(t), \quad j = 0, 1, 2, \dots, \quad (۴۳-۷)$$

تعریف می‌شود که در آن w_j یک اسکالر است و باید محاسبه شود.

به دست آوردن روابط بازگشتی در الگوریتم BiCGSTAB، شبیه به الگوریتم CGS است. ابتدا توجه کنید که داریم

$$\psi_{j+1}\varphi_{j+1} = (1 - w_j t)\psi_j\varphi_{j+1} = (1 - w_j t)(\psi_j\varphi_j - \alpha_j t\psi_j\pi_j). \quad (۴۴-۷)$$

اگر بتوانیم یک رابطه بازگشتی برای $\psi_j\pi_j$ بیابیم، آنگاه $\psi_{j+1}\varphi_{j+1}$ به هنگام خواهد شد. برای این منظور، از گام ۹ الگوریتم BiCG، داریم

$$\begin{aligned} \psi_j\pi_j &= \psi_j(\varphi_j + \beta_{j-1}\pi_{j-1}) \\ &= \psi_j\varphi_j + \beta_{j-1}(1 - w_{j-1}t)\psi_{j-1}\pi_{j-1}. \end{aligned} \quad (۴۵-۷)$$

۲۸۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

اکنون، بردارهای r_j و p_j را به صورت

$$r_j = \psi_j(A)\varphi_j(A)r_0, \quad (۴۶-۷)$$

$$p_j = \psi_j(A)\pi_j(A)r_0, \quad (۴۷-۷)$$

تعریف می‌کنیم. در نتیجه، از رابطه‌های (۴۴-۷) و (۴۵-۷)، خواهیم داشت

$$r_{j+1} = (I - w_j A)(r_j - \alpha_j A p_j),$$

$$p_{j+1} = r_{j+1} + \beta_j (I - w_j A) p_j,$$

در ادامه، نحوه محاسبه ثابت‌ها در رابطه‌های بازگشتی بالا را شرح می‌دهیم. با توجه به الگوریتم BiCG، داریم

$$\beta_j = \frac{\rho_{j+1}}{\rho_j},$$

که در آن

$$\begin{aligned} \rho_\ell &= \langle \varphi_\ell(A)r_0, \varphi_\ell(A^T)\tilde{r}_0 \rangle \\ &= \langle \varphi_\ell^\top(A)r_0, \tilde{r}_0 \rangle, \quad \ell = j, j+1. \end{aligned}$$

متأسفانه با استفاده از این فرمول‌ها، مقادیر ρ_j و ρ_{j+1} قابل محاسبه نیستند، زیرا هیچ‌کدام از بردارهای $\varphi_\ell(A)r_0$ ، $\varphi_\ell(A^T)\tilde{r}_0$ و $\varphi_\ell^\top(A)r_0$ معلوم نیستند. باین حال، بین ρ_ℓ و

$$\tilde{\rho}_\ell = \langle \varphi_\ell(A)r_0, \psi_\ell(A^T)\tilde{r}_0 \rangle,$$

رابطه مفیدی وجود دارد. داریم

$$\tilde{\rho}_\ell = \langle \psi_\ell(A)\varphi_\ell(A)r_0, \tilde{r}_0 \rangle = \langle r_\ell, \tilde{r}_\ell \rangle.$$

برای یافتن رابطه بین ρ_ℓ و $\tilde{\rho}_\ell$ ، نخست، فرض کنید $\ell = j$ و

$$\varphi_j(t) = \gamma_1^{(j)}t^j + \gamma_2^{(j)}t^{j-1} + \dots + \gamma_{j+1}^{(j)},$$

$$\psi_j(t) = \eta_1^{(j)}t^j + \eta_2^{(j)}t^{j-1} + \dots + \eta_{j+1}^{(j)}.$$

داریم

$$\begin{aligned}\tilde{\rho}_j &= \langle \varphi_j(A)r_0, \psi_j(A^T)\tilde{r}_0 \rangle \\ &= \langle \varphi_j(A)r_0, \eta_{\lambda}^{(j)} A^j \tilde{r}_0 + \eta_{\lambda}^{(j)} A^{j-1} \tilde{r}_0 + \dots + \eta_{j+1}^{(j)} \tilde{r}_0 \rangle \\ &= \langle \varphi_j(A)r_0, \eta_{\lambda}^{(j)} A^j \tilde{r}_0 \rangle,\end{aligned}$$

زیرا،

$$\langle \varphi_\ell(A)r_0, A^s \tilde{r}_0 \rangle = 0, \quad s = 1, 2, \dots, j-1,$$

به این دلیل که، $r_j = \varphi_\ell(A)r_0$ در راستای بردار v_{j+1} است و

$$A^s \tilde{r}_0 \in \mathcal{K}_j(A^T, \tilde{r}_0), \quad s = 1, 2, \dots, j-1.$$

ازاین‌رو، داریم

$$\begin{aligned}\tilde{\rho}_j &= \langle \varphi_j(A)r_0, \eta_{\lambda}^{(j)} A^j \tilde{r}_0 \rangle \\ &= \langle \varphi_j(A)r_0, \frac{\eta_{\lambda}^{(j)}}{\gamma_{\lambda}^{(j)}} \varphi_j(A^T) \tilde{r}_0 \rangle \\ &= \frac{\eta_{\lambda}^{(j)}}{\gamma_{\lambda}^{(j)}} \rho_j.\end{aligned}$$

اکنون، با توجه به رابطه (۷-۳۳)، داریم $\eta_{\lambda}^{(j+1)} = -w_j \eta_{\lambda}^{(j)}$. همچنین، با توجه به

$$\pi_{j+1} = \varphi_{j+1} + \beta_j \pi_j \text{ و } \varphi_{j+1} = \varphi_j - \alpha_j t \pi_j$$

$$\begin{aligned}\varphi_{j+1} &= \varphi_j - \alpha_j t \pi_j \\ &= \varphi_j - \alpha_j t (\varphi_j + \beta_{j-1} \pi_{j-1}) \\ &= -\alpha_j t \varphi_j + \varphi_j - \alpha_j \beta_{j-1} t \pi_{j-1}.\end{aligned}$$

بنابراین، نتیجه می‌گیریم که $\gamma_{\lambda}^{(j+1)} = -\alpha_j \gamma_{\lambda}^{(j)}$. با جمع‌بندی مطالب به‌دست‌آمده، داریم

$$\frac{\tilde{\rho}_{j+1}}{\tilde{\rho}_j} = \left(\frac{\eta_{\lambda}^{(j+1)} \rho_{j+1}}{\gamma_{\lambda}^{(j+1)}} \right) / \left(\frac{\eta_{\lambda}^{(j)} \rho_j}{\gamma_{\lambda}^{(j)}} \right)$$

$$= \left(\frac{-w_j \eta_{\lambda}^{(j)} \rho_{j+1}}{-\alpha_j \gamma_{\lambda}^{(j)}} \right) / \left(\frac{\eta_{\lambda}^{(j)} \rho_j}{\gamma_{\lambda}^{(j)}} \right) \\ = \frac{w_j}{\alpha_j} \times \frac{\rho_{j+1}}{\rho_j} = \frac{w_j}{\alpha_j} \beta_j.$$

بنابراین،

$$\beta_j = \frac{\tilde{\rho}_{j+1}}{\tilde{\rho}_j} \times \frac{\alpha_j}{w_j} = \frac{\langle r_{j+1}, \tilde{r}_j \rangle}{\langle r_j, \tilde{r}_j \rangle} \times \frac{\alpha_j}{w_j}.$$

از اینکه ضریب t^j در چندجمله‌ای φ_j و π_j ثابت هستند، داریم

$$\alpha_j = \frac{\langle r_j, \tilde{r}_j \rangle}{\langle Ap_j, \tilde{p}_j \rangle} = \frac{\langle \varphi_j(A)r_0, \psi_j(A^T)\tilde{r}_0 \rangle}{\langle A\pi_j(A)r_0, \psi_j(A^T)\tilde{r}_0 \rangle} \\ = \frac{\langle \psi_j(A)\varphi_j(A)r_0, \tilde{r}_0 \rangle}{\langle A\psi_j(A)\pi_j(A)r_0, \tilde{r}_0 \rangle} \\ = \frac{\tilde{\rho}_j}{\langle Ap_j, \tilde{r}_0 \rangle} = \frac{\langle r_j, \tilde{r}_j \rangle}{\langle Ap_j, \tilde{r}_0 \rangle}.$$

با فرض $s_j = r_j - \alpha_j Ap_j$ رابطه (۷-۴۶) به صورت

$$r_{j+1} = (I - w_j A)s_j,$$

نوشته می‌شود. اکنون، پارامتر w_j را به گونه‌ای محاسبه می‌کنیم که $\|r_{j+1}\|_2^2$ کمینه شود. داریم

$$\|r_{j+1}\|_2^2 = \langle (I - w_j A)s_j, (I - w_j A)s_j \rangle \\ = \langle s_j, s_j \rangle - 2w_j \langle As_j, s_j \rangle + w_j^2 \langle As_j, As_j \rangle.$$

بنابراین، اگر مشتق $\|r_{j+1}\|_2^2$ نسبت به w_j را برابر با صفر قرار دهیم، خواهیم داشت

$$\frac{d}{dw_j} \|r_{j+1}\|_2^2 = -2 \langle As_j, s_j \rangle + 2w_j \langle As_j, As_j \rangle = 0,$$

که رابطه زیر را نتیجه می‌دهد

$$w_j = \frac{\langle As_j, s_j \rangle}{\langle As_j, As_j \rangle}.$$

ازطرفی، داریم

$$r_{j+1} = (I - w_j A)s_j = s_j - w_j As_j = r_j - \alpha_j Ap_j - w_j As_j.$$

درنتیجه، خواهیم داشت

$$b - Ax_{j+1} = b - Ax_j - \alpha_j Ap_j - w_j As_j \implies x_{j+1} = x_j + \alpha_j p_j + w_j s_j.$$

اکنون، با توجه به نتایج به‌دست آمده، الگوریتم BiCGSTAB را به‌صورت زیر جمع‌بندی می‌کنیم.

الگوریتم ۶.۷ الگوریتم BiCGSTAB برای حل دستگاه $Ax = b$ (نسخه اولیه)

1. Choose an initial guess x_0 , $r_0 := b - Ax_0$ and $\tilde{r}_0$ arbitrary
2. Set $p_0 := r_0$
3. For $j = 0, 1, 2, \dots$, until convergence, Do
4. $\alpha_j := \langle r_j, \tilde{r}_0 \rangle / \langle Ap_j, \tilde{r}_0 \rangle$
5. $s_j := r_j - \alpha_j Ap_j$
6. $w_j = \langle As_j, s_j \rangle / \langle As_j, As_j \rangle$
7. $x_{j+1} = x_j + \alpha_j p_j + w_j s_j$
8. $r_{j+1} = s_j - w_j As_j$
9. $\beta_j = \frac{\langle r_{j+1}, \tilde{r}_j \rangle}{\langle r_j, \tilde{r}_j \rangle} \times \frac{\alpha_j}{w_j}$
10. $p_{j+1} := r_{j+1} + \beta_j(p_j - w_j Ap_j)$
11. EndDo

در گام ۵ از الگوریتم BiCGSTAB، اگر $s_j = 0$ ، آنگاه خواهیم داشت

$$r_j - \alpha_j Ap_j = 0 \implies b - Ax_j = \alpha_j Ap_j \implies A(x_j + \alpha_j p_j) = b.$$

این رابطه نشان می‌دهد که بردار

$$\hat{x}_{j+1} = x_j + \alpha_j p_j,$$

جواب واقعی دستگاه است. از این‌رو، بعد از گام پنجم، اگر نرم بردار s_j به اندازه کافی کوچک بود توقف کرده و $\hat{x}_{j+1}$ را به‌عنوان یک جواب تقریبی می‌پذیریم. درغیراین‌صورت، فرایند تکرار ادامه پیدا کرده و در گام ۷ بردار x_{j+1} را از رابطه زیر محاسبه می‌کنیم

$$x_{j+1} = \hat{x}_{j+1} + w_j s_j.$$

همچنین، در گام ۸، اگر نرم بردار مانده r_{j+1} به اندازه کافی کوچک بود، توقف کرده و x_{j+1} را به‌عنوان جواب تقریبی می‌پذیریم. با توجه به نکات ذکرشده، در هر تکرار از الگوریتم ۶.۷ دو شرط توقف مورد آزمایش قرار می‌گیرد. بنابراین، الگوریتم BiCGSTAB را می‌توان به‌صورت زیر بازنویسی کرد.

الگوریتم ۷.۷ الگوریتم BiCGSTAB برای حل دستگاه $Ax = b$

1. Choose an initial guess x_0 , $r_0 := b - Ax_0$ and $\tilde{r}_0$ arbitrary
2. Set $p_0 := r_0$
3. For $j = 0, 1, 2, \dots$, Do
4. $\hat{p}_j = Ap_j$
5. $\alpha_j := \langle r_j, \tilde{r}_0 \rangle / \langle \hat{p}_j, \tilde{r}_0 \rangle$
6. $s_j := r_j - \alpha_j \hat{p}_j$
7. $\hat{x}_{j+1} := x_j + \alpha_j p_j$
8. If $\|s_j\|$ is small enough, then stop. $\hat{x}_{j+1}$ is the approximate solution
9. $\hat{s}_j = As_j$
10. $w_j = \langle \hat{s}_j, s_j \rangle / \langle \hat{s}_j, \hat{s}_j \rangle$
11. $x_{j+1} = \hat{x}_{j+1} + w_j s_j$
12. $r_{j+1} = s_j - w_j \hat{s}_j$
13. If $\|r_{j+1}\|$ is small enough, then stop. x_{j+1} is the approximate solution
14. $\beta_j = \frac{\langle r_{j+1}, \tilde{r}_j \rangle}{\langle r_j, \tilde{r}_j \rangle} \times \frac{\alpha_j}{w_j}$
15. $p_{j+1} := r_{j+1} + \beta_j(p_j - w_j \hat{p}_j)$
16. EndDo

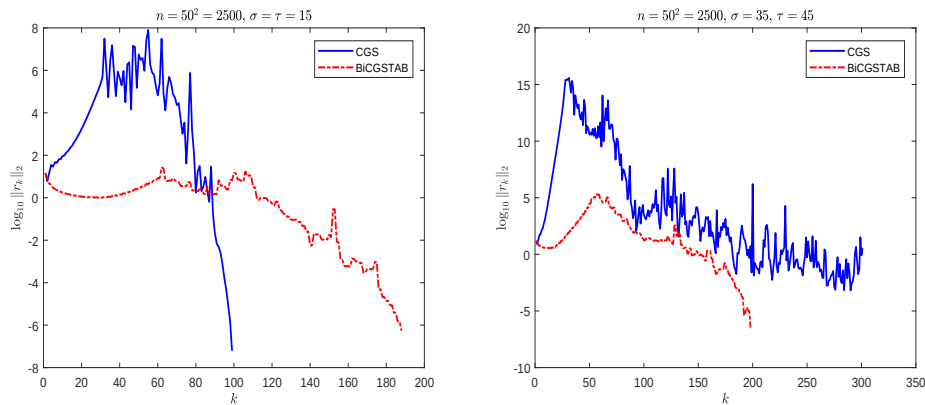
در نرم‌افزار متلب، توابع bicg، qmr، cgs و bicgstab، به ترتیب، برای الگوریتم‌های BiCG، QMR، CGS و BiCGSTAB فراهم شده است. ورودی و خروجی‌های هر چهار تابع، همانند تابع pcg است و کافی است به جای pcg هر یک از چهار تابع مذکور را قرار دهید. برای مثال، برای حل دستگاه $Ax = b$ با استفاده از الگوریتم BiCGSTAB می‌توانیم از دستور

$$[x, \text{flag}, \text{relres}, \text{iter}, \text{resvec}] = \text{bicgstab}(A, b, \text{tol}, \text{maxit}, [], [], x_0)$$

در متلب استفاده کنیم. لازم به ذکر است که گاهی تکرارهای الگوریتم BiCGSTAB به صورت $k + 0.5$ گزارش می‌شود که در آن k یک عدد طبیعی است (مثلاً ۲۷/۵)، که معنای آن این است که الگوریتم در تکرار $(k + 1)$ ام با شرط توقف اول به جواب تقریبی مطلوب رسیده است.

مثال ۱.۷ ماتریس ضرایب دستگاه حاصل از گسسته‌سازی مدل معرفی‌شده در بخش ۱-۸ به ازای $N = 50$ ، $\sigma = 35$ و $\tau = 45$ را در نظر بگیرید. دستگاهی را در نظر می‌گیریم که ماتریس ضرایب آن برابر با A است و بردار سمت راست آن به صورت

$$b = A(1, 1, \dots, 1)^T \in \mathbb{R}^n,$$



شکل ۱-۷ منحنی $\log_{10} \|r_k\|_2$ نسبت به k برای مثال $N = 50$ به ازای $\sigma = \tau = 15$ سمت راست به ازای $\sigma = \tau = 45$ و سمت چپ به ازای $\sigma = 35$ و $\tau = 45$.

می‌باشد، که در آن بعد ماتریس برابر با $n = N^2 = 2500$ است. قرار می‌دهیم

$$x_0 = 0, \quad tol = 10^{-7}, \quad maxit = 300.$$

در شکل ۱-۷ (نمودار سمت چپ)، تابع $\log_{10} \|r_k\|_2$ بر حسب k را برای دو روش CGS و BiCGSTAB رسم کرده‌ایم (توجه کنید که مقادیر $\|r_k\|_2$ در `resvec` ذخیره شده‌اند). همان‌گونه که می‌بینیم، منحنی مربوط به روش CGS نوسانات زیادی دارد و به علاوه، در ۳۰۰ تکرار همگرا نشده است. اما می‌بینیم که روش BiCGSTAB نه تنها همگرا شده است، بلکه منحنی نرم-۲ بردارهای مانده آن نیز بسیار هموارتر از روش CGS است. البته باید متذکر شویم که همیشه روش BiCGSTAB سریع‌تر از CGS همگرا نمی‌شود. برای مثال، به ازای $\sigma = \tau = 15$ این نمودار در شکل ۱-۷ (سمت راست) نمایش داده شده است. اگرچه می‌بینیم که نوسانات در روش CGS بسیار زیاد است، اما همگرایی آن سریع‌تر از BiCGSTAB است.

مثال ۲.۷ ماتریس ضرایب دستگاه حاصل از گسسته‌سازی مدل معرفی‌شده در بخش ۱-۸ به ازای $N = 50, 100, 150$ ، $\sigma = -30$ و $\tau = 5$ را در نظر بگیرید. سایر مشخصات را همانند دستگاه‌های مثال قبل در نظر می‌گیریم. نتایج عددی روش‌های CGS، QMR، BiCG و BiCGSTAB در جدول ۱-۷ ارائه شده‌اند. در این جدول، تعداد تکرارها برای همگرایی (*iters*) و مقدار *relres* گزارش شده است. نتایج عددی ارائه‌شده، کارایی هر چهار روش را در این مثال نشان می‌دهد.

جدول ۱-۷ نتایج عددی برای مثال ۲.۷.

	$N = 50$		$N = 100$		$N = 150$	
Method	<i>iters</i>	<i>relres</i>	<i>iters</i>	<i>relres</i>	<i>iters</i>	<i>relres</i>
QMR	۱۴۳	$5/6e-8$	۲۷۰	$2/6e-8$	۴۰۳	$2/6e-8$
BiCG	۱۴۳	$5/7e-8$	۲۷۰	$2/6e-8$	۴۰۰	$7/3e-8$
CGS	۸۷	$5/6e-8$	۱۷۹	$4/5e-8$	۲۵۵	$4/1e-8$
BiCGSTAB	۸۱/۵	$8/0e-8$	۱۶۷/۵	$3/4e-8$	۲۴۶/۵	$9/0e-8$

تمرینات

۱.۷ فرض کنید $V_m = [v_1, v_2, \dots, v_m]$ و $W_m = [w_1, w_2, \dots, w_m]$ ماتریس‌هایی باشند که توسط الگوریتم دو متعامدسازی لانکزوس (الگوریتم ۱.۷) تولید می‌شوند. ماتریس نمایش عملگر تصویر مورب روی فضای کرایلف $\mathcal{K}(A, v_1)$ و عمود بر $\mathcal{K}(A^T, w_1)$ را بنویسید. همچنین، ماتریس نمایش عملگر تصویر مورب روی فضای کرایلف $\mathcal{K}(A^T, w_1)$ و عمود بر $\mathcal{K}(A, v_1)$ را بنویسید. تحت چه شرایطی یک عملگر تصویر روی K و عمود بر L وجود دارد؟ این شرایط برای بردارهای لانکزوس و الگوریتم لانکزوس چگونه خواهد بود.

۲.۷ دو تساوی در رابطه (۷-۳۲) را ثابت کنید.

۳.۷ نشان دهید که در الگوریتم BiCG به ازای هر $i \neq j$ بردارهای r_j و $\tilde{r}_i$ بر هم عمودند و بردارهای p_j و $\tilde{p}_i$ بر هم A -عمودند، یعنی $\langle A\tilde{p}_i, p_j \rangle = 0$.

۴.۷ الگوریتم BiCGSTAB را برای حل دستگاه پیش‌شرط‌سازی شده

$$AP^{-1}y = b, \quad x = P^{-1}y, \quad (48-7)$$

بنویسید.

۵.۷ با کمک تابع bicgstab در متلب دستگاه پیش‌شرط‌سازی شده (۷-۴۸) را برای حل دستگاه‌های مثال (۲.۷) به کار ببرید. ماتریس پیش‌شرط‌ساز P را $ILU(0)$ بگیرید.

۶.۷ مثال ۱.۷ را برای $\sigma = -\tau = 10$ حل کنید. علاوه بر مقایسه روش‌های تکراری CGS و BiCGSTAB نسخه کامل الگوریتم GMRES را نیز با آنها مقایسه کنید.

پیش شرط‌سازی روش‌های تکراری

یکی از راه‌های متداول برای بالا بردن سرعت همگرایی روش‌های تکراری مبتنی بر زیرفضای کرایلِف برای حل دستگاه

$$Ax = b, \quad (۱-۸)$$

استفاده از راه‌کار پیش شرط‌سازی می‌باشد. اگر ماتریس ضرایب دستگاه معین مثبت متقارن باشد، آنگاه بهترین روش برای آن روش گرادیان مزدوج است. در این صورت، می‌توان دید که هرچه مقادیر ویژه ماتریس حول یک نقطه و دور از نقطه $(0, 0)$ بهتر خوشه شده باشد، آنگاه سرعت همگرایی روش بیشتر خواهد بود [۱۵]. این حقیقت از آنجا نتیجه می‌شود که اگر مقادیر یک ماتریس معین مثبت متقارن مثل A حول یک نقطه خوشه شده باشد، آنگاه بزرگ‌ترین مقدار ویژه (λ_1) و کوچک‌ترین مقدار ویژه (λ_n) ماتریس به هم نزدیک خواهد بود و در نتیجه، عدد حالت ماتریس A ، یعنی

$$\text{cond}_2(A) = \frac{\lambda_1}{\lambda_n},$$

کوچک خواهد بود. در این صورت، ماتریس A خوش حالت است و انتظار می‌رود که همگرایی روش گرادیان مزدوج، سریع باشد.

اگر ماتریس ضرایب متقارن نباشد، موضوع خیلی پیچیده است و توزیع مقادیر ویژه ماتریس ضرایب لزوماً نرخ همگرایی یک روش تکراری مبتنی بر زیرفضای کرایلِف مثل GMRES را توصیف نمی‌کند؛ با این وجود، هرچه مقادیر ویژه حول یک نقطه به دور از $(0, 0)$ بهتر خوشه شده باشد، برای مسائل کاربردی

زیادی سرعت همگرایی زیاد خواهد بود، به‌ویژه، برای ماتریس‌هایی که به ماتریس نرمال نزدیک هستند (مقدار $\|A^H A - A A^H\|$ کوچک باشد) [۱۵]. بنابراین، طبیعی است که سعی کنیم دستگاه را به دستگاه معادل دیگری تبدیل کنیم به‌طوری‌که ماتریس ضرایب آن از طیف مناسب‌تری برخوردار باشد. یک ماتریس پیش‌شرط‌ساز ماتریسی است که این تبدیل را انجام می‌دهد. یک ماتریس پیش‌شرط‌ساز مناسب، سرعت همگرایی روش تکراری را افزایش می‌دهد و بدون استفاده از ماتریس پیش‌شرط‌ساز ممکن است حتی روش با شکست مواجه شود. در این فصل، مطالب جامع‌تری از یک پیش‌شرط‌ساز و روش‌های تکراری پیش‌شرط‌سازی شده ارائه می‌شود و در فصل بعد چند شیوه ساختن پیش‌شرط‌ساز را معرفی می‌کنیم.

۱-۸ روش‌های تکراری پیش‌شرط‌سازی شده

به‌طورکلی، دو نوع ماتریس پیش‌شرط‌ساز وجود دارد:

(الف) ماتریس پیش‌شرط‌ساز M تقریبی از معکوس ماتریس A باشد؛ یعنی $M \approx A^{-1}$ یا $MA \approx I$.

(ب) ماتریس پیش‌شرط‌ساز M تقریبی از ماتریس A باشد؛ یعنی $M \approx A$ یا $M^{-1}A \approx I$.

در هر دو حالت (الف) و (ب) ماتریس پیش‌شرط‌ساز M بایستی نامنفرد باشد.

در حالت اول به دو صورت می‌توان دستگاه (۱-۸) را پیش‌شرط‌سازی نمود. در حالت اول که پیش‌شرط‌سازی از چپ نامیده می‌شود، به‌جای دستگاه (۱-۸)، دستگاه

$$MAx = Mb, \quad (2-8)$$

را حل می‌کنیم و در حالت دوم که پیش‌شرط‌سازی از راست نامیده می‌شود، نخست دستگاه

$$AMu = b, \quad (3-8)$$

را حل کرده و سپس قرار می‌دهیم $x = Mu$. لازم به ذکر است که اگر ماتریس M به صورت حاصلضرب دو ماتریس M_1 و M_2 تجزیه شده باشد، یعنی $M = M_1 M_2$ ، آنگاه

$$M = M_1 M_2 \approx A^{-1} \implies M_2^{-1} M_1^{-1} \approx A \implies M_2 A M_1 \approx I.$$

در این صورت، پیش‌شرط‌سازی می‌تواند از هر دو طرف چپ و راست و به‌طور هم‌زمان انجام شود که به پیش‌شرط‌سازی دوطرفه معروف است. در این حالت، به‌جای حل دستگاه (۱-۸)، دستگاه

$$M_2 A M_1 u = M_2 b,$$

را حل کرده و سپس قرار می‌دهیم $x = M^{-1}u$.

در پیش‌شرط‌سازی نوع (الف)، اگر از یک روش مبتنی بر زیرفضای کرایلف برای حل دستگاه پیش‌شرط‌سازی شده استفاده کنیم، فقط از ضرب ماتریس (ماتریس A یا ماتریس پیش‌شرط‌ساز) در بردار استفاده می‌شود. بنابراین، به‌کارگیری این‌گونه از پیش‌شرط‌سازها بسیار ساده است.

همانند حالت (الف) دستگاه‌های پیش‌شرط‌سازی شده در حالت (ب)، به‌صورت

$$M^{-1}Ax = M^{-1}b, \quad (۴-۸)$$

$$AM^{-1}u = b, \quad x = M^{-1}u, \quad (۵-۸)$$

$$M_1^{-1}AM_2^{-1}u = M_1^{-1}b, \quad x = M_2^{-1}u, \quad (۶-۸)$$

خواهند بود که به‌ترتیب پیش‌شرط‌سازی از چپ، راست و دوطرفه نامیده می‌شوند. توجه کنید در حالت دوطرفه، داریم $M = M_1M_2 \approx A$ و بنابراین، $M_1^{-1}AM_2^{-1} \approx I$. در اینجا لازم به ذکر است اگر از یک روش مبتنی بر زیرفضای کرایلف برای حل هریک از دستگاه‌های (۴-۸)، (۵-۸) یا (۶-۸) استفاده کنیم، آنگاه در هر ضرب ماتریس ضرایب دستگاه معادلات خطی پیش‌شرط‌سازی شده در یک بردار، نیاز به حل یک یا دو دستگاه می‌باشد. برای روشن شدن مطلب، فرض کنید به‌ازای بردار داده‌شده w می‌خواهیم

$$z = M^{-1}Aw,$$

را محاسبه کنیم. برای این کار، کافیهست دستگاه $Mz = Aw$ را برای محاسبه بردار z حل کنیم. بنابراین، در حالت (ب) ماتریس پیش‌شرط‌ساز باید طوری محاسبه شود که این‌گونه دستگاه‌ها به‌سادگی حل شوند. یک پیش‌شرط‌ساز خوب (در هر دو حالت (الف) و (ب)) بایستی دو خاصیت مهم داشته باشد. نخست اینکه به‌سادگی محاسبه شوند، دوم اینکه به‌سادگی به‌کار گرفته شوند. درواقع، ساختن و به‌کارگیری یک پیش‌شرط‌ساز سبب تحمیل هزینه اضافی خواهد بود و استفاده از یک پیش‌شرط‌ساز بایستی مقرون به صرفه باشد.

۲-۸ الگوریتم گرادیان مزدوج پیش‌شرط‌سازی شده

فرض کنید ماتریس‌های A و M هر دو معین مثبت متقارن باشند و $M \approx A$. در این صورت، ماتریس ضرایب دستگاه پیش‌شرط‌سازی شده $M^{-1}Ax = M^{-1}b$ یا $AM^{-1}u = b$ لزوماً متقارن نیستند. بنابراین، الگوریتم گرادیان مزدوج (CG) را نمی‌توان به‌طور مستقیم برای حل این دو دستگاه استفاده کرد.

۲۹۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

یکی از راه‌های فائق آمدن بر این مشکل این است که M را به صورت تجزیه شده در نظر بگیریم. برای مثال، فرض کنید ماتریس M به صورت $M = LL^T$ تجزیه شده باشد. در این صورت، می‌توان دستگاه پیش‌شرط‌سازی شده

$$L^{-1}AL^{-T}u = L^{-1}b, \quad x = L^{-T}u.$$

را حل کرد که در آن ماتریس ضرایب دستگاه معین مثبت متقارن است. فرض می‌کنیم $\hat{A} = L^{-1}AL^{-T}$ و $\hat{b} = L^{-1}b$. در این صورت، در الگوریتم گرادیان مزدوج برای حل دستگاه $\hat{A}u = \hat{b}$ ، خواهیم داشت

$$\begin{aligned} u_{j+1} &= u_j + \alpha_j \hat{p}_j \implies L^T x_{j+1} = L^T x_j + \alpha_j \hat{p}_j \\ &\implies x_{j+1} = x_j + \alpha_j L^{-T} \hat{p}_j. \end{aligned}$$

با فرض $p_j = L^{-T} \hat{p}_j$ ، داریم

$$x_{j+1} = x_j + \alpha_j p_j.$$

از طرف دیگر، داریم

$$\begin{aligned} \hat{p}_{j+1} &= \hat{r}_{j+1} + \beta_j \hat{p}_j \implies L^T p_{j+1} = \hat{r}_{j+1} + \beta_j L^T p_j \\ &\implies p_{j+1} = L^{-T} \hat{r}_{j+1} + \beta_j p_j. \end{aligned}$$

همچنین، $\hat{r}_{j+1}$ بردار مانده الگوریتم گرادیان مزدوج برای حل دستگاه $\hat{A}\hat{x} = \hat{b}$ است. از این رو،

$$\begin{aligned} \hat{r}_{j+1} &= \hat{b} - \hat{A}\hat{x}_{j+1} \\ &= L^{-1}b - L^{-1}AL^{-T}u_{j+1} \\ &= L^{-1}(b - Ax_{j+1}) \\ &= L^{-1}r_{j+1}. \end{aligned}$$

از سوی دیگر،

$$\begin{aligned} \alpha_j &= \frac{\langle \hat{r}_j, \hat{r}_j \rangle}{\langle \hat{A}\hat{p}_j, \hat{p}_j \rangle} \\ &= \frac{\langle \hat{r}_j, \hat{r}_j \rangle}{\langle L^{-1}AL^{-T}\hat{p}_j, \hat{p}_j \rangle} \end{aligned}$$

$$\begin{aligned} &= \frac{\langle \hat{r}_j, \hat{r}_j \rangle}{\langle AL^{-T}\hat{p}_j, L^{-T}\hat{p}_j \rangle} \\ &= \frac{\langle \hat{r}_j, \hat{r}_j \rangle}{\langle Ap_j, p_j \rangle}. \end{aligned}$$

به‌طور مشابه، داریم

$$\beta_j = \frac{\langle \hat{r}_{j+1}, \hat{r}_{j+1} \rangle}{\langle \hat{r}_j, \hat{r}_j \rangle}.$$

با جمع‌بندی این مطالب، الگوریتم گرادیان مزدوج با پیش‌شرط‌ساز دوطرفه به‌صورت الگوریتم ۱.۸ نوشته می‌شود.

الگوریتم ۱.۸ گرادیان مزدوج با پیش‌شرط‌ساز دوطرفه

1. Compute $r_0 = b - Ax_0$, $\hat{r}_0 = L^{-1}r_0$ and $p_0 = L^{-T}\hat{r}_0$.
2. For $j = 0, 1, 2, \dots$, until convergence Do:
3. $\alpha_j := \langle \hat{r}_j, \hat{r}_j \rangle / \langle Ap_j, p_j \rangle$
4. $x_{j+1} = x_j + \alpha_j p_j$
5. $\hat{r}_{j+1} := \hat{r}_j - \alpha_j L^{-1}Ap_j$
6. $\beta_j := \langle \hat{r}_{j+1}, \hat{r}_{j+1} \rangle / \langle \hat{r}_j, \hat{r}_j \rangle$
7. $p_{j+1} := L^{-T}\hat{r}_{j+1} + \beta_j p_j$
8. EndDo

همان‌گونه که اشاره کردیم ماتریس $M^{-1}A$ لزوماً متقارن نیست و بنابراین، الگوریتم گرادیان مزدوج را نمی‌توان به‌طور مستقیم برای حل دستگاه $M^{-1}Ax = M^{-1}b$ به‌کار برد. اما می‌دانیم که ماتریس $M^{-1}A$ نسبت به M -ضرب داخلی

$$\langle x, y \rangle_M = \langle Mx, y \rangle = \langle x, My \rangle,$$

خودالحاق است؛ زیرا

$$\begin{aligned} \langle M^{-1}Ax, y \rangle_M &= \langle Ax, y \rangle = \langle x, Ay \rangle \\ &= \langle x, M(M^{-1}A)y \rangle \\ &= \langle x, M^{-1}Ay \rangle_M. \end{aligned}$$

توجه کنید که خودالحاقی، درحقیقت، یک تعمیم از تقارن است. بنابراین، کافیت که در الگوریتم گرادیان مزدوج برای دستگاه پیش‌شرط‌سازی شده از چپ، به‌جای ضرب داخلی اقلیدسی، M -ضرب داخلی را جایگزین کنیم.

فرض کنید $r_j = b - Ax_j$ بردار ماندهٔ دستگاه اولیه و $z_j = M^{-1}r_j$ بردار ماندهٔ دستگاه پیش‌شرط‌سازی شده باشد. در این صورت،

1. $\alpha_j := \frac{\langle z_j, z_j \rangle_M}{\langle M^{-1}Ap_j, p_j \rangle_M} = \frac{\langle Mz_j, z_j \rangle}{\langle Ap_j, p_j \rangle} = \frac{\langle r_j, z_j \rangle}{\langle Ap_j, p_j \rangle},$
2. $x_{j+1} := x_j + \alpha_j p_j,$
3. $r_{j+1} := r_j - \alpha_j Ap_j, \quad z_{j+1} := M^{-1}r_{j+1},$
4. $\beta_j := \frac{\langle z_{j+1}, z_{j+1} \rangle_M}{\langle z_j, z_j \rangle_M} = \frac{\langle Mz_{j+1}, z_{j+1} \rangle}{\langle Mz_j, z_j \rangle} = \frac{\langle r_{j+1}, z_{j+1} \rangle}{\langle r_j, z_j \rangle},$
5. $p_{j+1} := z_{j+1} + \beta_j p_j.$

با توجه به این نتایج، روش گرادیان مزدوج با پیش‌شرط‌ساز چپ به صورت الگوریتم ۲.۸ به دست می‌آید.

الگوریتم ۲.۸ گرادیان مزدوج با پیش‌شرط‌ساز چپ

1. Compute $r_0 = b - Ax_0, z_0 = M^{-1}r_0$ and $p_0 = z_0.$
2. For $j = 0, 1, 2, \dots$, until convergence Do:
3. $\alpha_j := \langle r_j, z_j \rangle / \langle Ap_j, p_j \rangle$
4. $x_{j+1} = x_j + \alpha_j p_j$
5. $r_{j+1} := r_j - \alpha_j Ap_j$
6. $z_{j+1} := M^{-1}r_{j+1}$
7. $\beta_j := \langle r_{j+1}, z_{j+1} \rangle / \langle r_j, z_j \rangle$
8. $p_{j+1} := z_{j+1} + \beta_j p_j$
9. EndDo

۳-۸ الگوریتم $\text{GMRES}(m)$ پیش‌شرط‌سازی شده

در روش تکراری $\text{GMRES}(m)$ و سایر روش‌های تکراری مشابه برای حل دستگاه‌های نامتقارن، هر سه نوع پیش‌شرط‌سازی (چپ، راست و دوطرفه) مورد استفاده قرار می‌گیرند. در اینجا ما نسخهٔ $\text{GMRES}(m)$ با پیش‌شرط‌سازی از چپ و راست را ارائه می‌نمایم و ارائهٔ نسخهٔ دوطرفه را به خواننده واگذار می‌کنیم.

۱-۳-۸ پیش‌شرط‌سازی از چپ

دستگاه پیش‌شرط‌سازی شده از چپ زیر را در نظر بگیرید

$$M^{-1}Ax = M^{-1}b.$$

روشن است که برای نوشتن الگوریتم $\text{GMRES}(m)$ برای حل دستگاه کافیهست در الگوریتم $\text{GMRES}(m)$ ماتریس A و بردار b به‌ترتیب با ماتریس $M^{-1}A$ و بردار $M^{-1}b$ جایگزین شوند. در این صورت، روش $\text{GMRES}(m)$ با پیش‌شرط‌ساز چپ به‌صورت الگوریتم ۳.۸ جمع‌بندی می‌شود.

الگوریتم ۳.۸ $\text{GMRES}(m)$ با پیش‌شرط‌ساز چپ

1. Compute $r_0 = M^{-1}(b - Ax_0)$, $\beta = \|r_0\|_2$, and $v_1 = r_0/\beta$.
2. For $j = 1, \dots, m$ Do:
3. Compute $w := M^{-1}Av_j$
4. For $i = 1, \dots, j$ Do:
5. $h_{ij} = \langle w, v_i \rangle$
6. $w := w - h_{ij}v_i$
7. EndDo
8. Compute $h_{j+1,j} = \|w\|_2$ and $v_{j+1} = w/h_{j+1,j}$
9. EndDo
10. Define $V_m = [v_1, \dots, v_m]$, $\bar{H}_m = \{h_{ij}\}_{1 \leq i \leq j+1, 1 \leq j \leq m}$.
11. Compute $y_m = \arg\min_y \|\beta e_1 - \bar{H}_m y\|_2$ and $x_m = x_0 + V_m y_m$.
12. If satisfied Stop, else set $x_0 := x_m$ and Go To 1

حلقه گرام-اشمیت آرنولدی در الگوریتم ۳.۸ یک پایه یک‌معامد برای زیرفضای کرایلف

$$\text{span} \{r_0, M^{-1}Ar_0, \dots, (M^{-1}A)^{m-1}r_0\}.$$

می‌سازد، که در آن $r_0 = M^{-1}(b - Ax_0)$. بردارهای مانده در ۳.۸ به‌صورت

$$z_m = M^{-1}(b - Ax_m),$$

هستند. بنابراین، همه نرم‌های مانده در این الگوریتم، نرم بردارهایی از این نوع هستند. توجه کنید که دسترسی آسان به مانده دستگاه اصلی، یعنی $r_m = b - Ax_m$ ، وجود ندارد، جز آنکه به‌طور صریح محاسبه شود، یعنی با ضرب مستقیم ماتریس M در بردارهای مانده دستگاه پیش‌شرط‌سازی شده.

۲-۳-۸ پیش‌شرط‌سازی از راست

دستگاه از راست پیش‌شرط‌سازی شده

$$AM^{-1}u = b, \quad u = Mx, \quad (7-8)$$

را در نظر بگیرید. در اینجا نیازی به محاسبه مستقیم متغیر u نیست. به محض اینکه مانده $u_0 = b - Ax_0 = b - AM^{-1}u_0$ محاسبه شود، تمام بردارهای زیرفضای کرایلف بدون مراجعه به متغیر u می‌توانند محاسبه شوند. بردار مانده اولیه دستگاه پیش‌شرط‌سازی شده را می‌توان از رابطه $r_0 = b - Ax_0$ به دست آورد که برابر با $b - AM^{-1}u_0$ است. در عمل بردار x_0 در دست است نه بردار u_0 . در پایان جواب تقریبی دستگاه (۷-۸) برحسب متغیر u به صورت

$$u_m = u_0 + \sum_{i=1}^m v_i \eta_i,$$

خواهد بود، که در آن $u_0 = Mx_0$. با ضرب طرفین رابطه اخیر در M^{-1} جواب تقریبی دستگاه اولیه به صورت

$$x_m = x_0 + M^{-1} \sum_{i=1}^m v_i \eta_i,$$

نوشته می‌شود. بنابراین، با جمع‌بندی این مطالب، الگوریتم $\text{GMRES}(m)$ با پیش‌شرط‌ساز راست به صورت الگوریتم ۴.۸ خواهد بود.

الگوریتم ۴.۸ $\text{GMRES}(m)$ با پیش‌شرط‌ساز راست

1. Compute $r_0 = b - Ax_0$, $\beta := \|r_0\|_2$, and $v_1 := r_0/\beta$.
2. For $j = 1, \dots, m$ Do:
 3. Compute $w := AM^{-1}v_j$
 4. For $i = 1, \dots, j$ Do:
 5. $h_{ij} = \langle w, v_i \rangle$
 6. $w := w - h_{ij}v_i$
 7. EndDo
 8. Compute $h_{j+1,j} = \|w\|_2$ and $v_{j+1} = w/h_{j+1,j}$
 9. Define the $V_m = [v_1, \dots, v_m]$, $\bar{H}_m = \{h_{ij}\}_{1 \leq i \leq j+1, 1 \leq j \leq m}$.
10. EndDo
11. Compute $y_m = \arg\min_y \|\beta e_1 - \bar{H}_m y\|_2$ and $x_m = x_0 + M^{-1}V_m y_m$.
12. If satisfied Stop, else set $x_0 := x_m$ and Go To 1

در این الگوریتم، حلقه آرنولدی یک پایه یک‌متعامد برای زیرفضای کرایلف

$$\text{span} \{r_0, AM^{-1}r_0, \dots, (AM^{-1})^{m-1}r_0\},$$

می‌سازد، که در آن $r_0 = b - Ax_0$.

۸-۳-۳ پیش‌شرط‌سازی دوطرفه

در بسیاری از حالت‌ها ماتریس M به صورت تجزیه‌شده

$$M = LU,$$

می‌باشد. در این صورت، دستگاه پیش‌شرط‌سازی شده دوطرفه به صورت

$$L^{-1}AU^{-1}u = L^{-1}b, \quad x = U^{-1}u, \quad (۸-۸)$$

خواهد بود. انتظار می‌رود که خواننده بتواند با بهره‌گیری از دو الگوریتم $\text{GMRES}(m)$ با پیش‌شرط‌ساز راست و چپ، الگوریتم $\text{GMRES}(m)$ برای دستگاه پیش‌شرط‌سازی شده دوطرفه (۸-۸) را بنویسد.

۸-۳-۴ مقایسه پیش‌شرط‌ساز چپ و راست

نخست توجه کنید که هر سه ماتریس $M^{-1}A$ ، AM^{-1} و $L^{-1}AU^{-1}$ با یکدیگر متشابه هستند، و بنابراین، مقادیر ویژه آنها یکسانند. از این رو، می‌توان انتظار داشت که سرعت همگرایی یک روش تکراری مبتنی بر زیرفضای کرایلف در هر سه حالت تقریباً یکی باشد، هرچند باید توجه کرد که همیشه توزیع مقادیر ویژه یک دستگاه غیرهرمیتی، سرعت همگرایی یک روش تکراری برای حل آن دستگاه را توصیف نمی‌کند. دستگاه پیش‌شرط‌سازی شده $M^{-1}Ax = M^{-1}b$ را در نظر بگیرید. می‌دانیم که الگوریتم GMRES ،

نرم مانده این دستگاه، یعنی $\|M^{-1}(b - Ax)\|_p$ را روی زیرفضای

$$x_0 + \mathcal{K}_m^L(M^{-1}A, z_0) = x_0 + \text{span}\{z_0, M^{-1}Az_0, \dots, (M^{-1}A)^{m-1}z_0\},$$

کمینه می‌کند، که در آن $z_0 = M^{-1}r_0$ بردار مانده تقریب اولیه برای دستگاه پیش‌شرط‌سازی شده است (بالانویس L در $\mathcal{K}^L$ برای پیش‌شرط‌سازی از چپ استفاده شده است). بنابراین، جواب تقریبی x_m را می‌توان به صورت

$$x_m = x_0 + s_{m-1}(M^{-1}A)z_0,$$

نوشت که در آن s_{m-1} یک چندجمله‌ای از درجه $m-1$ است به طوری که مقدار

$$\begin{aligned} \|M^{-1}(b - Ax)\|_p &= \|M^{-1}(b - A(x_0 + s(M^{-1}A)z_0))\|_p \\ &= \|z_0 - M^{-1}As(M^{-1}A)z_0\|_p, \end{aligned}$$

را روی مجموعه همه چندجمله‌ای‌های s از درجه حداکثر $m - 1$ کمینه می‌کند. این شرط بهینگی را می‌توان برحسب بردار مانده دستگاه اولیه نوشت. برای این کار، داریم

$$\begin{aligned} z_0 - M^{-1}As(M^{-1}A)z_0 &= M^{-1} \left(r_0 - As(M^{-1}A)M^{-1}r_0 \right) \\ &= M^{-1} \left(r_0 - AM^{-1}s(AM^{-1})r_0 \right), \end{aligned} \quad (9-8)$$

که در آن آخرین تساوی از برابری $s(M^{-1}A)M^{-1} = M^{-1}s(AM^{-1})$ حاصل شده است. اکنون، دستگاه پیش‌شرط‌سازی شده از راست $AM^{-1}u = b$ را در نظر بگیرید، که در آن $x = M^{-1}u$. اگر حدس اولیه برای این دستگاه را بردار u_0 بگیریم، آنگاه بردار مانده این دستگاه به صورت زیر خواهد بود

$$b - AM^{-1}u = b - Ax;$$

یعنی بردار مانده دستگاه پیش‌شرط‌سازی شده با بردار مانده دستگاه اولیه یکسان است. در نتیجه، الگوریتم GMRES برای حل دستگاه $AM^{-1}u = b$ ، مقدار $\|b - AM^{-1}u\|_2$ را روی زیرفضای

$$u_0 + \mathcal{K}_m^R(AM^{-1}, r_0) = u_0 + \text{span}\{r_0, AM^{-1}r_0, \dots, (AM^{-1})^{m-1}r_0\},$$

کمینه می‌کند. به عبارت دیگر، الگوریتم GMRES مقدار $\|b - Ax\|_2$ را روی زیرفضای

$$M^{-1}u_0 + M^{-1}\mathcal{K}_m^R(AM^{-1}, r_0) = x_0 + \text{span}\{z_0, M^{-1}Az_0, \dots, (M^{-1}A)^{m-1}z_0\},$$

کمینه می‌کند (بالانویس R در $\mathcal{K}^R$ برای پیش‌شرط‌سازی از راست استفاده شده است). این خود معادل است با اینکه الگوریتم GMRES برای حل دستگاه پیش‌شرط‌سازی شده $AM^{-1}u = b$ برداری مثل

$$x_m = x_0 + s_{m-1}(M^{-1}A)z_0,$$

را محاسبه می‌کند که در آن s_{m-1} یک چندجمله‌ای از درجه $m - 1$ است به طوری که مقدار

$$\begin{aligned} \|b - Ax\|_2 &= \left\| b - A \left(x_0 + s_{m-1}(M^{-1}A)z_0 \right) \right\|_2 \\ &= \left\| r_0 - AM^{-1}s(AM^{-1})r_0 \right\|_2, \end{aligned} \quad (10-8)$$

را روی مجموعه همه چندجمله‌ای‌های s از درجه حداکثر $m - 1$ کمینه می‌کند. اکنون، با توجه به رابطه‌های (۹-۸) و (۱۰-۸) می‌توانیم قضیه زیر را بیان کنیم.

قضیه ۱.۸ جواب تقریبی محاسبه‌شده توسط روش‌های GMRES پیش‌شرط‌سازی‌شده از چپ و راست، برداری به‌صورت

$$x_m = x_0 + s_{m-1}(M^{-1}A)z_0 = x_0 + M^{-1}s_{m-1}(AM^{-1})r_0,$$

است که در آن $z_0 = M^{-1}r_0$ و s_{m-1} یک چندجمله‌ای از درجه $m-1$ است. چندجمله‌ای s_{m-1} مقدار $\|b - Ax\|_2$ را در پیش‌شرط‌سازی از راست و مقدار $\|M^{-1}(b - Ax)\|_2$ را در پیش‌شرط‌سازی از چپ کمینه می‌کند.

۵-۳-۸ الگوریتم GMRES انعطاف‌پذیر (FGMRES)

در همه روش‌های تکراری پیش‌شرط‌سازی‌شده‌ای که پیش از این معرفی شده‌اند، ماتریس پیش‌شرط‌ساز M در تمام تکرارها ثابت است. اما ممکن است در تکرارهای مختلف ماتریس پیش‌شرط‌ساز M تغییر کند یا تنها عمل $M^{-1}x$ مشخص باشد، برای مثال از یک روش تکراری دیگر برای محاسبه آن استفاده شود. در این حالت، روش‌های پیش‌شرط‌سازی‌شدهٔ مرور شده همگرا نخواهند بود. اما می‌توان برخی از این روش‌ها را برای این حالت سازگار کرد. در اینجا نسخهٔ پیش‌شرط‌سازی‌شده از راست GMRES را در نظر گرفته و طوری اصلاح می‌کنیم که امکان تغییر پیش‌شرط‌ساز در هر تکرار وجود داشته باشد. الگوریتمی که به این ترتیب به دست می‌آید، الگوریتم GMRES انعطاف‌پذیر می‌گویند که به‌طور خلاصه با FGMRES نشان می‌دهند [۸۶].

فرض کنید در تکرار j ام، ماتریس پیش‌شرط‌ساز برابر با M_j باشد. در این صورت، در گام سوم الگوریتم ۴.۸ برداری به‌صورت $w_j = AM_j^{-1}v_j$ محاسبه می‌شود. با فرض

$$z_j = M_j^{-1}v_j,$$

خواهیم داشت $w_j = Az_j$. با توجه به اینکه ماتریس پیش‌شرط‌ساز، تکرار به تکرار، تغییر می‌کند، بردار z_j نیز تغییر خواهد کرد. بنابراین، طبیعی است گام ۱۱ الگوریتم را به‌صورت زیر بنویسیم

$$\begin{aligned} x_m &= x_0 + (M_1^{-1}v_1, M_2^{-1}v_2, \dots, M_m^{-1}v_m)y_m \\ &= x_0 + (z_1, z_2, \dots, z_m)y_m = x_0 + Z_my_m, \end{aligned}$$

که در آن $Z_m = (z_1, z_2, \dots, z_m)$. با جمع‌بندی این مطالب، الگوریتم FGMRES به‌صورت زیر جمع‌بندی می‌شود.

الگوریتم ۵.۸ FGMRES

1. Compute $r_0 = b - Ax_0$, $\beta := \|r_0\|_2$, and $v_1 := r_0/\beta$.
2. For $j = 1, \dots, m$, Do:
 3. Compute $z_j := M_j^{-1}v_j$
 4. Compute $w := Az_j$
 5. For $i = 1, \dots, j$ Do:
 6. $h_{ij} = \langle w, v_i \rangle$
 7. $w := w - h_{ij}v_i$
 8. EndDo
 9. Compute $h_{j+1,j} = \|w\|_2$ and $v_{j+1} = w/h_{j+1,j}$
10. Define the $V_m = [v_1, \dots, v_m]$, $\overline{H}_m = \{h_{ij}\}_{1 \leq i \leq j+1, 1 \leq j \leq m}$.
11. EndDo
12. Set $Z_m = (z_1, z_2, \dots, z_m)$
13. Compute $y_m = \underset{y}{\operatorname{argmin}} \|\beta e_1 - \overline{H}_m y\|_2$ and $x_m = x_0 + Z_m y_m$
14. If satisfied Stop, else set $x_0 := x_m$ and Go To 1

اگر همه M_j ها در این الگوریتم، یکسان و برابر با M باشند، آنگاه این الگوریتم معادل با الگوریتم ۴.۸ خواهد بود و در نتیجه، با استناد به قضیه ۳.۶ خواهیم داشت

$$AM^{-1}V_m = V_{m+1}\overline{H}_m.$$

اما اگر ماتریس پیش‌شرط‌ساز، تکرار به تکرار تغییر کند، آنگاه این رابطه برقرار نخواهد بود و به جای آن خواهیم داشت

$$AZ_m = V_{m+1}\overline{H}_m.$$

بنابراین، الگوریتم GMRES برداری مثل x_m در $x_0 + \operatorname{span}\{z_1, z_2, \dots, z_m\}$ جست‌وجو می‌کند به‌طوری‌که

$$x_m = \underset{x \in x_0 + \operatorname{span}\{z_1, z_2, \dots, z_m\}}{\operatorname{argmin}} \|b - Ax\|_2.$$

هر بردار $x \in x_0 + \operatorname{span}\{z_1, z_2, \dots, z_m\}$ به‌صورت

$$x = x_0 + Z_m y,$$

نوشته می‌شود، که در آن $y \in \mathbb{R}^m$. بنابراین، داریم

$$\begin{aligned}\|b - Ax\|_2 &= \|b - A(x_0 + Z_my)\|_2 \\ &= \|r_0 - AZ_my\|_2 \\ &= \|r_0 - V_{m+1}\overline{H}_my\|_2 \\ &= \|V_{m+1}(\beta e_1 - \overline{H}_my)\|_2 \\ &= \|\beta e_1 - \overline{H}_my\|_2.\end{aligned}$$

ازاین‌رو، خواهیم داشت $x_m = x_0 + Z_my_m$ که در آن

$$y_m = \arg \min_{y \in \mathbb{R}^m} \|\beta e_1 - \overline{H}_my\|_2.$$

با جمع‌بندی نتایج به‌دست‌آمده، می‌توانیم قضیه زیر را بیان کنیم.

قضیه ۲.۸ جواب تقریبی x_m به‌دست‌آمده در گام m ام از روش FGMRES نرم-۲ بردار مانده، یعنی $\|b - Ax\|_2$ را روی زیرفضای آفین $x_0 + \text{span}\{z_1, z_2, \dots, z_m\}$ کمینه می‌کند. قضیه بعد نشان می‌دهد که تحت شرایطی الگوریتم FGMRES جواب واقعی دستگاه را محاسبه می‌کند.

قضیه ۳.۸ فرض کنید $\beta = \|r_0\|_2 \neq 0$ و به‌ازای هر $i < j$ ، $h_{i+1,i} \neq 0$. اگر ماتریس H_j نامنفرد باشد، آنگاه x_j جواب دستگاه $Ax = b$ است اگر و تنها اگر $h_{j+1,j} = 0$. برهان. فرض کنید $h_{j+1,j} = 0$ و برای هر $i < j$ ، $h_{i+1,i} \neq 0$. دراین‌صورت، خواهیم داشت

$$AZ_j = V_{j+1}\overline{H}_j = V_j H_j.$$

درنتیجه، برای هر بردار

$$x = x_0 + Z_j y \in x_0 + \text{span}\{z_1, z_2, \dots, z_j\},$$

داریم

$$\|b - Ax\|_2 = \|r_0 - AZ_j y\|_2 = \|\beta v_1 - V_j H_j y\|_2 = \|\beta e_1 - H_j y\|_2,$$

۳۰۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

که در آن $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^j$. اکنون، اگر ماتریس H_j نامنفرد باشد، آنگاه جواب مسئله کمترین توان‌های دوم

$$\min_{y \in \mathbb{R}^j} \|b - A(x_0 + Z_j y)\|_2 = \min_{y \in \mathbb{R}^j} \|\beta e_1 - H_j y\|_2,$$

برابر است با

$$\tilde{y} = H_j^{-1}(\beta e_1).$$

بنابراین، جواب تقریبی به صورت

$$\tilde{x} = x_0 + Z_j \tilde{y},$$

و نرم بردار مانده نیز برابر با صفر خواهد بود:

$$\|b - A\tilde{x}\|_2 = \|\beta e_1 - H_j \tilde{y}\|_2 = \|\beta e_1 - H_j(H_j^{-1}(\beta e_1))\|_2 = 0.$$

این رابطه نشان می‌دهد که بردار $\tilde{x}$ جواب واقعی دستگاه $Ax = b$ است.

برعکس، فرض کنید $\tilde{x} = x_0 + Z_j \tilde{y}$ جواب واقعی دستگاه $Ax = b$ باشد. در این صورت، داریم

$$\begin{aligned} 0 &= b - A\tilde{x} = b - A(x_0 + Z_j \tilde{y}) = r_0 - AZ_j \tilde{y} \\ &= \beta v_1 - V_{j+1} \bar{H}_j \tilde{y} = V_{j+1}(\beta e_1 - \bar{H}_j \tilde{y}). \end{aligned}$$

اگر $0 \neq h_{j+1,j}$ ، آنگاه بردار v_{j+1} قابل محاسبه و ستون‌های ماتریس V_{j+1} یک‌معامد خواهد بود. بنابراین، از رابطه اخیر خواهیم داشت

$$0 = \beta e_1 - \bar{H}_j \tilde{y}.$$

اکنون، با توجه به اینکه ماتریس $\bar{H}_j$ یک ماتریس بالاهسنبرگی است و $h_{i+1,i} \neq 0$ ، به‌ازای هر $j \leq i$ ، با استفاده از جایگذاری پس‌رو نتیجه می‌شود که $\tilde{y} = 0$. بنابراین، خواهیم داشت $\beta = 0$ ، که یک تناقض است. در نتیجه، $h_{j+1,j} = 0$. □

۴-۸ پیش‌شرط‌سازی دستگاه معادلات نرمال

با توجه به اینکه ماتریس ضرایب دستگاه معادلات نرمال معین مثبت متقارن است، می‌توان روش‌های مختلفی از جمله گرادیان مزدوج را برای حل آن به‌کار برد. دستگاه معادلات نرمال را می‌توان از راست، چپ

یا دوطرفه پیش‌شرط‌سازی کرد. بنابراین، به‌سادگی می‌توان الگوریتم گرادیان مزدوج پیش‌شرط‌سازی شده را برای حل دستگاه معادلات نرمال توسعه داد. این کار کاملاً شبیه به الگوریتم گرادیان مزدوج پیش‌شرط‌سازی شده برای حل دستگاه معین مثبت متقارن $Ax = b$ انجام می‌پذیرد. در ادامه، نسخه از چپ پیش‌شرط‌سازی شده الگوریتم‌های CGNE و CGNR را ارائه می‌کنیم.

دستگاه معادلات نرمال پیش‌شرط‌سازی شده

$$M^{-1}A^T Ax = M^{-1}b,$$

را در نظر بگیرید که در آن M ماتریس پیش‌شرط‌ساز است. همانند الگوریتم ۲.۸، فرض می‌کنیم

$$r_j = b - Ax_j, \quad \tilde{r}_j = A^T r_j, \quad z_j = M^{-1} \tilde{r}_j.$$

در این صورت، مقدار α_j از رابطه زیر به دست می‌آید

$$\alpha_j = \frac{\langle \tilde{r}_j, z_j \rangle}{\langle A^T A p_j, p_j \rangle} = \frac{\langle \tilde{r}_j, z_j \rangle}{\langle A p_j, A p_j \rangle} = \frac{\langle \tilde{r}_j, z_j \rangle}{\langle w_j, w_j \rangle} = \frac{\langle \tilde{r}_j, z_j \rangle}{\|w_j\|_2^2},$$

که در آن $w_j = A p_j$. با جمع‌بندی مطالب ذکر شده، الگوریتم از چپ پیش‌شرط‌سازی شده CGNR به صورت زیر نوشته می‌شود.

الگوریتم ۶.۸ الگوریتم CGNR پیش‌شرط‌سازی شده از چپ

1. Compute $r_0 = b - Ax_0$, $\tilde{r}_0 = A^T r_0$, $z_0 = M^{-1} \tilde{r}_0$ and $p_0 = z_0$
2. For $j = 0, 1, 2, \dots$, until convergence Do:
 3. $w_j := A p_j$
 4. $\alpha_j = \langle \tilde{r}_j, z_j \rangle / \|w_j\|_2^2$
 5. $x_{j+1} = x_j + \alpha_j p_j$
 6. $r_{j+1} := r_j - \alpha_j w_j$
 7. $\tilde{r}_{j+1} := A^T r_{j+1}$
 8. $z_{j+1} := M^{-1} \tilde{r}_{j+1}$
 9. $\beta_j := \langle z_{j+1}, \tilde{r}_{j+1} \rangle / \langle z_j, \tilde{r}_j \rangle$
 10. $p_{j+1} := z_{j+1} + \beta_j p_j$
 11. EndDo

به‌طور مشابه، دستگاه معادلات نرمال پیش‌شرط‌سازی شده $M^{-1}AA^T u = M^{-1}b$ را در نظر بگیرید که در آن $x = A^T u$. در این صورت، با فرض $r_j = b - Ax_j$ و $z_j = M^{-1}r_j$ در الگوریتم CG برای

حل این دستگاه خواهیم داشت

$$\alpha_j = \frac{\langle r_j, z_j \rangle}{\langle AA^T q_j, q_j \rangle} = \frac{\langle r_j, z_j \rangle}{\langle A^T q_j, A^T q_j \rangle} = \frac{\langle r_j, z_j \rangle}{\langle p_j, p_j \rangle},$$

که در آن $p_j = A^T q_j$ داریم همچنین،

$$u_{j+1} = u_j + \alpha_j q_j \iff A^T u_{j+1} = A^T u_j + \alpha_j A^T q_j \iff x_{j+1} = x_j + \alpha_j p_j,$$

$$r_{j+1} = r_j - \alpha_j AA^T q_j = r_j - \alpha_j A p_j,$$

$$z_{j+1} = M^{-1} r_{j+1},$$

$$q_{j+1} = z_{j+1} + \beta_j q_j \iff A^T q_{j+1} = A^T z_{j+1} + \beta_j A^T q_j$$

$$\iff p_{j+1} = A^T z_{j+1} + \beta_j p_j.$$

اکنون با توجه به نتایج ذکرشده، الگوریتم CGNE پیش‌شرط‌سازی‌شده از چپ به‌صورت زیر جمع‌بندی می‌شود.

الگوریتم ۷.۸ الگوریتم CGNE پیش‌شرط‌سازی‌شده از چپ

1. Compute $r_0 = b - Ax_0$, $z_0 = M^{-1}r_0$ and $p_0 = A^T z_0$
2. For $j = 0, 1, 2, \dots$, until convergence Do:
3. $w_j := Ap_j$
4. $\alpha_j = \langle r_j, z_j \rangle / \|p_j\|_2^2$
5. $x_{j+1} = x_j + \alpha_j p_j$
6. $r_{j+1} := r_j - \alpha_j w_j$
7. $z_{j+1} := M^{-1}r_{j+1}$
8. $\beta_j := \langle r_{j+1}, z_{j+1} \rangle / \langle r_j, z_j \rangle$
9. $p_{j+1} := A^T z_{j+1} + \beta_j p_j$
10. EndDo

تاکنون، چندین روش تکراری پیش‌شرط‌سازی‌شده برای حل دستگاه معادلات خطی (نرمال و غیرنرمال) ارائه شد. سایر الگوریتم‌ها، مثل الگوریتم‌های QMR، BiCGSTAB، و ...، را نیز می‌توان به‌طور مشابه پیش‌شرط‌سازی کرد.

۵-۸ روش‌های تکراری پیش‌شرط‌سازی شده در متلب

همان‌گونه که پیش از این دیدیم بسیاری از الگوریتم‌های معرفی شده برای حل دستگاه معادلات خطی، مثل GMRES، CG، BiCGSTAB، و ... به صورت توابعی در متلب فراهم شده‌اند. همه این توابع را می‌توان برای حل دستگاه‌های پیش‌شرط‌سازی شده نیز به کار برد. در اینجا شیوه استفاده از الگوریتم GMRES را برای حل یک دستگاه پیش‌شرط‌سازی شده ارائه می‌کنیم و سایر توابع را نیز می‌توان به طور مشابه استفاده کرد.

۱-۵-۸ پیش‌شرط‌سازی از چپ

تابع gmres در حالت کلی دستگاه پیش‌شرط‌سازی شده از چپ به صورت

$$M^{-1}Ax = M^{-1}b,$$

را حل می‌کند که در آن $M = M_1 M_2$. در حقیقت، این تابع، الگوریتم ۳.۸ را برای حل این دستگاه به کار می‌برد. در آغاز الگوریتم، برداری به صورت $r_0 = M^{-1}(b - Ax_0)$ و سپس، در هر تکرار از این الگوریتم و در گام ۳، برداری به صورت $w = M^{-1}Av_j$ محاسبه می‌شود. برای محاسبه بردارهای r_0 و w باید دو دستگاه زیر حل شوند

$$Mr_0 = b - Ax_0, \quad Mw = Av_j.$$

برای انجام این کار، بردارهای Av_j و $b - Ax_0$ به راحتی محاسبه می‌شوند. اما آنچه که در اینجا از اهمیت بیشتری دارد، حل دو دستگاه به صورت $Mz = s$ است که باید بردار z محاسبه شود. ماتریس M معمولاً طوری انتخاب می‌شود که این دستگاه با هزینه کمی حل شود. برای مثال، ماتریس M ، بالامثلی، پائین‌مثلثی یا حاصل ضرب این دو ماتریس انتخاب می‌شود. همچنین، ماتریس M ممکن است طوری انتخاب شود که دستگاه $Mz = s$ به سادگی با یک روش تکراری حل شود.

در حالتی که، ماتریس M به صورت $M = M_1 M_2$ باشد، برای حل دستگاه $Mz = s$ ، فرض می‌کنیم $M_2 z = r$. سپس، نخست دستگاه $M_1 r = s$ را حل کرده، و بعد با حل دستگاه $M_2 z = r$ بردار z را محاسبه می‌کنیم.

هرآنچه که در بالا ذکر شد، توسط تابع gmres در متلب انجام می‌شود. این تابع به صورت زیر است.

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,M1,M2,x0);
```

درحقیقت، کافی است ماتریس‌های M_1 و M_2 را به‌عنوان ورودی‌های ششم و هفتم در تابع gmres وارد کنیم. در حالتی که ماتریس M به‌صورت تجزیه‌شده نباشد، می‌توان از هریک از توابع زیر استفاده کرد.

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,M,[],x0);
```

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,[],M,x0);
```

پیش‌شرط‌ساز M را می‌توان به‌صورت یک تابع به‌شکل زیر نیز معرفی کرد.

```
fm=@(x) M\ x;
```

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,fm,[],x0);
```

در حقیقت، پیش از اجرای الگوریتم GMRES برای یک بردار دلخواه x ، دستور محاسبه $M^{-1}x$ را به آن می‌دهیم و در طول اجرای الگوریتم، در هر گام که اثر ماتریس M^{-1} بر یک بردار نیاز شد، به دستور داده‌شده مراجعه کرده و آن را پیاده‌سازی می‌کند. اگر ماتریس M به صورت تجزیه‌شده $M = M_1 M_2$ باشد، می‌توان از دستورات زیر استفاده کرد.

```
fm=@(x) M2\ (M1\ x);
```

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,fm,[],x0);
```

در صورتی که ماتریس M یک تقریب از A^{-1} باشد، نمی‌توان مستقیماً از تابع gmres در متلب استفاده کرد و بایستی پیش‌شرط‌ساز را به‌صورت تابع وارد کرد. برای مثال، برای حل دستگاه $MAx = Mb$ توسط تابع gmres باید از دستورات زیر استفاده کنیم.

```
fm=@(x) M*x;
```

```
[x,flag,relres,iter,resvec]=gmres(A,b,m,tol,maxit,fm,[],x0);
```

توجه کنید که در تابع gmres بالا، می‌توان جای fm و $[]$ را عوض کرد.

یک نکته بسیار مهم قابل ذکر در اینجا وجود دارد. زمانی که تابع gmres برای حل دستگاه $M^{-1}Ax = M^{-1}b$ به‌کار برده می‌شود، خروجی‌های $relres$ و $resvec$ اطلاعات مربوط به دستگاه پیش‌شرط‌سازی شده را برمی‌گرداند، نه اطلاعات مربوط به دستگاه اولیه. در حقیقت، مقدار $relres$ برابر است با

$$relres = \frac{\|M^{-1}(b - Ax_k)\|_2}{\|M^{-1}b\|_2},$$

حتی اگر حدس اولیه x ناصفر باشد. همچنین، مؤلفه‌های بردار $resvec$ برابر با

$$\|M^{-1}(b - Ax_k)\|_2,$$

هستند. برای مقایسه کارایی دو پیش‌شرط‌ساز مختلف برای حل یک دستگاه معادلات خطی، معقول به نظر می‌رسد (یک مقایسه عادلانه) که مانده دستگاه اولیه را برای دو پیش‌شرط‌ساز مقایسه کنیم. برای این کار، می‌توانیم از پیش‌شرط‌ساز راست استفاده کنیم، زیرا مانده دستگاه پیش‌شرط‌سازی شده با مانده دستگاه اولیه یکسان است. متأسفانه نمی‌توان از تابع gmres در متلب مستقیماً برای حل دستگاه از راست پیش‌شرط‌سازی شده استفاده کرد. اما در ادامه خواهیم دید که می‌توان با یک ترفند مناسب از تابع gmres برای حل یک دستگاه از راست پیش‌شرط‌سازی شده استفاده کرد.

۸-۵-۲ پیش‌شرط‌سازی راست

دستگاه پیش‌شرط‌سازی شده

$$AM^{-1}u = b, \quad x = M^{-1}u, \quad (۱۱-۸)$$

را در نظر بگیرید. برای حل این دستگاه، ابتدا دستگاه $AM^{-1}u = b$ را با تابع gmres حل کرده و سپس با حل دستگاه $Mx = u$ بردار x را محاسبه می‌کنیم. برای حل دستگاه $AM^{-1}u = b$ می‌توانیم ورودی ماتریس ضرایب در تابع gmres را به صورت یک تابع به شکل زیر وارد کنیم.

```
fa=@(x) A*(M\ x);
```

```
[x,flag,relres,iter,resvec]=gmres(fa,b,m,tol,maxit,[],[],x0);
```

```
x=M\u;
```

در اینجا توجه کنید که داریم

$$r_k = b - AM^{-1}u_k = b - Ax_k,$$

و

$$relres = \frac{\|b - AM^{-1}u_k\|_2}{\|b\|_2} = \frac{\|b - Ax_k\|_2}{\|b\|_2}.$$

به طور مشابه، برای حل دستگاه

$$AMu = b, \quad x = Mu,$$

می‌توان از دستورات زیر استفاده کرد.

```
fa=@(x) A*(M*x);
```

```
[x,flag,relres,iter,resvec]=gmres(fa,b,m,tol,maxit,[],[],x0);
```

```
x=M*u;
```

با کمک راه‌کارهای ذکرشده در زیربخش‌های ۸-۵-۱ و ۸-۵-۲ می‌توان به‌سادگی تابع gmres را برای حل دستگاه‌هایی که از دو طرف پیش‌شرط‌سازی شده‌اند، به‌کار گرفت (تمرین ۵.۸ را ببینید). همان‌گونه که پیش از این ذکر کردیم، برای حل دستگاه (۸-۱۱)، اگر بخواهیم بردار $z = M^{-1}x$ را در گام ۳ الگوریتم GMRES از حل تقریبی $x = Mz$ به‌دست آوریم (مثلاً با کمک یک روش تکراری)، باید از نسخهٔ انعطاف‌پذیر GMRES، یعنی FGMRES، استفاده کنیم. متأسفانه تابع مربوط به نسخهٔ FGMRES در متلب وجود ندارد و لازم است کُد متلب آن تهیه شود.

ملاحظه ۱.۸ توابع pcg، qmr، bicg، cgs و bicgstab در متلب که به‌ترتیب برای روش‌های تکراری CG، QMR، BiCG، CGS و BiCGSTAB فراهم شده‌اند، همگی اطلاعات مربوط به دستگاه اولیه (نه دستگاه پیش‌شرط‌سازی‌شده) را در relres و resvec ذخیره می‌کنند. از این رو، برای مقایسهٔ چند پیش‌شرط‌ساز برای حل یک دستگاه معادلات خطی با استفاده از این روش‌ها، می‌توان از شیوهٔ پیش‌شرط‌سازی از راست یا چپ استفاده کرد.

تمرینات

۱.۸ فرض کنید ماتریس‌های $M, A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن باشند. نشان دهید که ماتریس $M^{-1}A$ نسبت به ضرب داخلی $\langle x, y \rangle_A = \langle Ax, y \rangle = \langle x, Ay \rangle$ خودالحاق است. سپس با استفاده از آن، الگوریتم گرادیان مزدوج برای حل دستگاه معادلات خطی $M^{-1}Ax = M^{-1}b$ را بنویسید.

۲.۸ الگوریتم $\text{GMRES}(m)$ برای حل دستگاه پیش‌شرط‌سازی‌شده $L^{-1}AU^{-1}u = L^{-1}b$ بنویسید که در آن $Ux = u$.

۳.۸ الگوریتم BiCGSTAB از راست پیش‌شرط‌سازی‌شده را بنویسید.

۴.۸ الگوریتم CGS از دو طرف پیش‌شرط‌سازی‌شده را بنویسید.

۵.۸ چگونه می‌توان از تابع gmres در متلب برای حل دستگاه از دو طرف پیش‌شرط‌سازی‌شده

$$M_1^{-1}AM_2^{-1}u = M_1^{-1}b, \quad x = M_2^{-1}u,$$

استفاده کرد.

شیوه‌های ساخت پیش‌شرط‌ساز

در فصل پیش، دیدیم که چگونه می‌توان یک ماتریس پیش‌شرط‌ساز را برای حل دستگاه معادلات خطی به‌وسیله روش‌های تکراری مبتنی بر زیرفضای کرایلف به‌کار گرفت. این بخش به معرفی چند روش ساخت ماتریس‌های پیش‌شرط‌ساز اختصاص دارد. نخست به معرفی پیش‌شرط‌سازهایی که توسط یک روش تکراری ایستا تولید می‌شوند، می‌پردازیم. سپس، شیوه‌های ساخت ماتریس پیش‌شرط‌ساز مبتنی بر تجزیه LU ناقص و مباحث نظری مربوطه معرفی می‌شوند. روش‌های تکراری که با یک ماتریس پیش‌شرط‌ساز مبتنی بر تجزیه LU ناقص همراه می‌شوند، جزو روش‌های کارا برای حل دستگاه معادلات خطی هستند. در ادامه، به معرفی روش‌هایی برای ساخت پیش‌شرط‌ساز از نوع تقریب معکوس که به‌صورت تجزیه‌شده است را معرفی کرده و خواص آنها را بررسی می‌کنیم.

۹-۱ پیش‌شرط‌سازهای تولیدشده توسط روش‌های تکراری ایستا

فرض کنید روش تکراری ایستای

$$x^{(k+1)} = M^{-1} N x^{(k)} + M^{-1} b,$$

توسط شکافت $A = M - N$ تولید شده باشد. به‌علاوه، فرض کنید این روند تکراری به‌ازای هر حدس اولیه دلخواه $x^{(0)}$ همگرا باشد، یعنی $\rho(M^{-1}N) < 1$. همان‌گونه که در فصل ۲ دیدیم مقادیر ویژه ماتریس $M^{-1}A$ دروی یک دیسک به مرکز $(1, 0)$ و به شعاع ۱ خوشه می‌شوند. بنابراین، همان‌گونه که در ابتدای فصل پیش ذکر کردیم، انتظار می‌رود که همگرایی روش‌های تکراری نامتقارن مثل GMRES

برای حل هریک از دستگاه‌های پیش‌شرط‌سازی شده

$$M^{-1}Ax = M^{-1}b,$$

یا

$$AM^{-1}u = b, \quad x = M^{-1}u,$$

سریع باشد. لازم به ذکر است که دو ماتریس $M^{-1}A$ و M^{-1} متشابه هستند (چرا؟) و در نتیجه، توزیع مقادیر ویژه آنها یکسان است. اکنون، با توجه به نکات اشاره‌شده، با کمک دو گام زیر می‌توان یک پیش‌شرط‌ساز برای بهبود سرعت همگرایی یک روش تکراری غیرایستا (مثلاً، مبتنی بر زیرفضای کرایلف) ساخت

(۱) یک شکافت مثل $A = M - N$ بسازید به طوری که به ازای هر حدس اولیه دلخواه $x^{(0)}$ روش تکراری $x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b$ همگرا باشد.

(۲) دستگاه پیش‌شرط‌سازی شده $M^{-1}Ax = M^{-1}b$ را با استفاده از یک روش تکراری غیرایستا (مثلاً، مبتنی بر زیرفضای کرایلف مثل CG یا GMRES) حل کنید.

۹-۱-۱ پیش‌شرط‌ساز AOR

همان‌گونه که در فصل ۲ دیدیم روش تکراری AOR توسط شکافت

$$M_{\gamma,\omega} = \frac{1}{\omega}(D - \gamma E), \quad N_{\gamma,\omega} = \frac{1}{\omega}[(1 - \omega)D + (\omega - \gamma)E + \omega F],$$

تولید می‌شود. بنابراین، ماتریس $M_{\gamma,\omega}$ می‌تواند به عنوان یک پیش‌شرط‌ساز مورد استفاده قرار گیرد. توجه کنید که اگر قرار دهیم

$$P_{\gamma} = D - \gamma E,$$

آنگاه خواهیم داشت

$$\text{cond}(P_{\gamma}^{-1}A) = \text{cond}(M_{\gamma,\omega}^{-1}A).$$

بنابراین، ضریب $1/\omega$ تأثیری در پیش‌شرط‌سازی ندارد و می‌توان آن را نادیده گرفت. توجه کنید که در این صورت، پیش‌شرط‌ساز AOR به پیش‌شرط‌ساز SOR تبدیل می‌شود.

مثال ۱.۹ ماتریس ضرایب دستگاه حاصل از گسسته‌سازی مدل معرفی‌شده در بخش ۱-۸ به‌ازای $N = 60$ ، $\sigma = 50$ و $\tau = -5$ را در نظر بگیرید. بردار b را طوری می‌گیریم که جواب واقعی دستگاه، برداری باشد که همه درایه‌های آن برابر با ۱ هستند. نسخه کامل روش تکراری GMRES و نسخه از راست پیش‌شرط‌سازی‌شده آن با پیش‌شرط‌ساز AOR را برای حل دستگاه استفاده می‌کنیم. مقدار γ را برابر با $1/32$ می‌گیریم. برنامه متلب استفاده‌شده به‌صورت زیر است که سایر اطلاعات ورودی در این برنامه دیده می‌شوند.

```
maxit=1000; tol=1e-7; N=60; sigma=50; tau=-5;
A=kron_conv_diff(sigma,tau,N); b=sum(A,2); n=N^2;
x0=zeros(n,1);
[x,flag,relres,iter,resvec]=gmres(A,b,[],tol,maxit,[],[],x0);
iter(2)
gamma=1.32; D=diag(diag(A)); E=D-tril(A); P=D-gamma*E;
x0=zeros(n,1);
fa=@(x) A*(P\ x);
[u,flag,relres,iter,resvec]=gmres(fa,b,[],tol,maxit,[],[],x0);
x=P\u;
iter(2)
```

با اجرای این برنامه، خواهیم دید که الگوریتم GMRES بدون پیش‌شرط‌سازی در ۱۲۲ تکرار و با پیش‌شرط‌ساز AOR در ۶۴ تکرار همگرا می‌شود. لازم به ذکر است که مقدار $\gamma = 1/32$ پارامتر بهینه (پارامتری که کمترین تکرار را به دست می‌دهد) در بازه $(0, 2)$ است. همان‌گونه که می‌بینیم پیش‌شرط‌ساز AOR تعداد تکرارهای GMRES را تقریباً به میزان پنجاه درصد کاهش داده است.

۲-۱-۹ پیش‌شرط‌ساز SSOR

در فصل ۲ دیدیم که اگر قرار دهیم

$$M_{\omega} = \frac{1}{\omega(2-\omega)}(D - \omega E)D^{-1}(D - \omega F),$$

$$N_{\omega} = \frac{1}{\omega(2-\omega)}((1-\omega)D + \omega E)D^{-1}((1-\omega)D + \omega F),$$

۳۱۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

آنگاه خواهیم $A = M_\omega - N_\omega$ و $G_\omega = M_\omega^{-1}N_\omega$ ، که در آن ماتریس تکرار روش تکراری SSOR است. بنابراین، ماتریس M_ω می‌تواند به‌عنوان یک پیش‌شرط‌ساز برای حل دستگاه $Ax = b$ مورد استفاده قرار گیرد (پیش‌شرط‌ساز SSOR). همانند پیش‌شرط‌ساز AOR، می‌توان از ضریب $1/(\omega(2-\omega))$ چشم‌پوشی کرد و ماتریس

$$P_\omega = (D - \omega E)D^{-1}(D - \omega F),$$

را به‌عنوان ماتریس پیش‌شرط‌ساز در نظر گرفت. در به‌کارگیری این پیش‌شرط‌ساز در الگوریتم GMRES، لازم است بردارهایی به‌صورت

$$z = P_\omega^{-1}x = (D - \omega F)^{-1}D(D - \omega E)^{-1}x,$$

محاسبه شود که این کار توسط دستور

$$z = (D - \omega F) \setminus (D * ((D - \omega E) \setminus x));$$

در متلب انجام می‌پذیرد. توجه کنید که ماتریس‌های $D - \omega E$ و $D - \omega F$ مثلثی هستند و به‌سادگی به‌ترتیب با استفاده از جایگذاری پیشرو و جایگذاری پسرو حل می‌شوند.

مثال ۲.۹ دستگاه مثال ۱.۹ با روش از راست پیش‌شرط‌سازی‌شده GMRES با پیش‌شرط‌ساز SSOR حل می‌کنیم. پارامتر ω را برابر با ۱/۵۸ می‌گیریم (این پارامتر در بازه $(0, 2)$ بهینه است). این روش در ۲۲ تکرار همگرا می‌شود. با مقایسه پیش‌شرط‌ساز SSOR با پیش‌شرط‌ساز AOR، برتری پیش‌شرط‌ساز SSOR را در این مثال نشان می‌دهد. برنامه متلب مورد استفاده به‌صورت زیر است.

```
w=1.58;
D=diag(diag(A)); E=D-tril(A); F=D-triu(A);
P1=D-w*E; P2=D-w*F;
x0=zeros(n,1);
fa=@(x) A*(P2\ (D*(P1\ x)));
[u,flag,relres,iter,resvec]=gmres(fa,b,[],tol,maxit,[],[],x0);
x=P2\ (D*(P1\ u));
iter(2)
```

۲-۹ پیش‌شرط‌سازهای مبتنی بر تجزیه ناقص LU

فرض کنید همه زیرماتریس‌های اصلی پیشروی ماتریس $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ نامنفرد باشند. در این صورت، یک ماتریس پائین‌مثلثی L با درایه‌های قطری ۱ و یک ماتریس بالامثلثی با درایه‌های قطری ناصفر مثل U وجود دارد به طوری که

$$A = LU.$$

به علاوه، این تجزیه یکتا است. در این بخش، روش‌هایی را مرور می‌کنیم که با کمک آنها می‌توان ماتریس پائین‌مثلثی $\tilde{L}$ و بالامثلثی $\tilde{U}$ را طوری محاسبه کرد که $\tilde{L} \approx L$ و $\tilde{U} \approx U$. بنابراین، ماتریس

$$M = \tilde{L}\tilde{U},$$

را می‌توان به عنوان یک پیش‌شرط‌ساز برای دستگاه معادلات خطی $Ax = b$ به کار برد.

۱-۲-۹ محاسبه تجزیه LU با کمک الگوریتم حذفی گاوس

الگوریتم ۱.۹، که به الگوریتم حذفی گاوس معروف است، تجزیه LU ماتریس A را در صورت وجود، محاسبه می‌کند. این الگوریتم همه محاسبات را روی خود ماتریس A انجام می‌دهد و در پایان کار، قسمت پائین‌مثلثی اکید A با قسمت پائین‌مثلثی اکید L و قسمت بالامثلثی U با قسمت بالامثلثی A یکسان است. توجه کنید که درایه‌های روی قطر اصلی L برابر با ۱ است و نیاز به ذخیره‌سازی نیست. به سادگی می‌توان دید که این الگوریتم با الگوریتم ۱.۱ معادل است. با این وجود، ذکر این نکته ضروری به نظر می‌رسد که ترتیب حلقه‌ها در الگوریتم ۱.۹ با الگوریتم ۱.۱ متفاوت است.

الگوریتم ۱.۹ روش حذفی گاوس

1. For $i = 2, \dots, n$, Do:
 2. For $k = 1, 2, \dots, i - 1$ Do:
 3. $a_{ik} := a_{ik}/a_{kk}$
 4. For $j = k + 1, \dots, n$ Do:
 5. $a_{ij} := a_{ij} - a_{ik}a_{kj}$
 6. EndDo
 7. EndDo
 8. EndDo

در حلقه اصلی الگوریتم ۱.۹ و در مرحله i ام، سطر i ام از L و U محاسبه می‌شوند. در واقع، در گام i ام تنها سطر i ام و سطر مربوط به درایه محوری در محاسبات درگیر هستند.

ماتریس تُنک $A \in \mathbb{R}^{n \times n}$ را در نظر بگیرید. در روش تجزیه ناقص LU، یک ماتریس پائین‌مثلثی تُنک مثل L و یک ماتریس بالامثلثی تُنک مثل U محاسبه می‌شوند به‌طوری‌که

$$A \approx LU.$$

می‌توان از الگوریتم ۱۰۹ برای محاسبه تجزیه ناقص LU استفاده کرد. اگر ماتریس A یک ماتریس تُنک باشد، لزوماً عامل‌های تجزیه LU، یعنی ماتریس‌های L و U ، تُنک نیستند. در نتیجه، باید راه‌کاری برای تُنک نگه داشتن عامل‌های L و U ارائه کرد. یک راه‌کار برای حفظ تُنکی این است که پیش از شروع محاسبه تجزیه، یک الگوی صفر برای عامل‌های L و U در نظر بگیریم. قرار می‌دهیم

$$P \subset \{(i, j) \mid i \neq j; 1 \leq i, j \leq n\}.$$

اکنون، در الگوریتم ۱۰۹، اگر $(i, j) \notin P$ ، درایه متناظر در عامل L یا U را محاسبه کرده و در غیر این صورت، درایه متناظر را برابر با صفر قرار می‌دهیم. این روند به‌صورت الگوریتم ۲۰۹ جمع‌بندی می‌شود.

الگوریتم ۲۰۹ شکل کلی تجزیه ناقص LU

1. For $i = 2, \dots, n$, Do:
 2. For $k = 1, 2, \dots, i - 1$ and $(i, k) \notin P$ Do:
 3. $a_{ik} := a_{ik} / a_{kk}$
 4. For $j = k + 1, \dots, n$ and for $(i, j) \notin P$, Do:
 5. $a_{ij} := a_{ij} - a_{ik} a_{kj}$
 6. EndDo
 7. EndDo
 8. EndDo

مشکل اصلی در الگوریتم ۲۰۹ انتخاب الگوی صفر P است. یک انتخاب این است که P را دقیقاً الگوی صفر ماتریس A بگیریم که در این صورت، این الگوریتم را الگوریتم ILU(0) می‌گویند. ایده اصلی تجزیه ILU(0) در ابتدا برای صورت‌های خاص تفاضلات متناهی برای حل معادلات دیفرانسیل جزئی به‌کار گرفته شد و سپس این روش با استفاده از حذف کردن درایه‌های درج‌شده تعمیم یافت. یک درایه درج‌شده، درایه‌ای است که در L یا U و در مکانی ظاهر می‌شود که در ماتریس اولیه و در آن مکان یک درایه صفر قرار دارد. بنابراین، اگر مجموعه (i, j) ‌هایی که $a_{ij} \neq 0$ را با $NZ(A)$ نشان دهیم، آنگاه الگوریتم ILU(0) به‌صورت الگوریتم ۳۰۹ خواهد بود.

الگوریتم ۳.۹ ILU(0)

1. For $i = 2, \dots, n$, Do:
2. For $k = 1, 2, \dots, i - 1$ and $(i, k) \in NZ(A)$ Do:
3. $a_{ik} := a_{ik}/a_{kk}$
4. For $j = k + 1, \dots, n$ and for $(i, j) \in NZ(A)$, Do:
5. $a_{ij} := a_{ij} - a_{ik}a_{kj}$
6. EndDo
7. EndDo
8. EndDo

همان‌گونه که از الگوریتم ۳.۹ مشخص است، درایه‌هایی که در طی فرایند ILU(0) حذف می‌شوند به الگوی ناصفر ماتریس A بستگی دارند، نه به بزرگی درایه‌های ماتریس A . نسخه دیگری از تجزیه ناقص LU ماتریس A این است که از ساختار کلی الگوریتم حذفی گاوس استفاده شود و در طول فرایند حذفی گاوس درایه‌های کوچک از لحاظ قدرمطلق حذف شوند. این الگوریتم به الگوریتم ILUT مشهور است که توسط سعد در [۸۵] ارائه شده است. این نوع پیش‌شرط‌ساز بسیار کاراست.

الگوریتم ۴.۹ ILUT

1. Set $L = (\ell_{ij}) = I \in \mathbb{R}^{n \times n}$, $U = (u_{ij}) = 0 \in \mathbb{R}^{n \times n}$ and $w = 0 \in \mathbb{R}^n$
2. For $i = 1, \dots, n$ Do:
3. $w := A(i, :)$
4. For $k = 1, 2, \dots, i - 1$ and when $w_k \neq 0$ Do:
5. $w_k := w_{ik}/a_{kk}$
6. Apply a dropping rule to w_k
7. If $w_k \neq 0$, then
8. $w := w - w_k * U(k, :)$
9. EndIf
10. EndDo
11. Apply a dropping rule to row w
12. $\ell_{ij} := w_j$ for $j = 1, \dots, i - 1$
13. $u_{ij} := w_j$ for $j = i, \dots, n$
14. $w := 0$
15. EndDo

درگام‌های ۶ و ۱۱، شیوه‌های مختلف برای عمل چکاندن (با صفر جایگزین کردن) درایه‌ها وجود دارد که نسخه‌های مختلف از الگوریتم ILUT را به وجود می‌آورد. مهم‌ترین نسخه، الگوریتم ILUT(p, μ) است

که در آن p یک عدد طبیعی و μ یک عدد حقیقی کوچک و مثبت است. در این نسخه از الگوریتم، از قاعده زیر برای چکاندن استفاده می‌شود.

(۱) در گام ۵، اگر درایه w_k در شرط $|w_k| \leq \mu \|A(i, :)\|_1$ صدق کند، آنگاه این درایه چکانده می‌شود.
(۲) در گام ۱۱، نخست درایه‌هایی از w که از لحاظ قدرمطلق کوچک‌تر یا مساوی $\mu \|A(i, :)\|_1$ هستند، چکانده می‌شوند. سپس p تا از بزرگ‌ترین (از لحاظ قدرمطلق) درایه‌های w که مربوط به ماتریس L و p تا از بزرگ‌ترین (از لحاظ قدرمطلق) درایه‌ها در w که مربوط به ماتریس U می‌باشند، به‌علاوه درایه قطری نگهداری و سایر درایه‌ها چکانده می‌شوند.

در نسخه $ILUT(p, \mu)$ ، هرچه پارامتر μ کوچک‌تر یا پارامتر p بزرگ‌تر اختیار شوند، ماتریس‌های L و U چگال‌تر و کیفیت پیش‌شرط‌ساز بیشتر خواهد بود، اما هزینه محاسباتی و ذخیره‌سازی افزایش می‌یابد. پارامتر p معمولاً طوری انتخاب می‌شود که

$$\rho = \frac{nnz(L) + nnz(U) - n}{nn(A)} \approx 1,$$

که در آن $nnz(\cdot)$ بیانگر تعداد درایه‌های ناصفر ماتریس است. درخصوص انتخاب پارامتر μ هیچ مطلب دقیقی را نمی‌توان بیان کرد. با این وجود، تجربه نشان می‌دهد که به‌ازای $0.1 \leq \mu \leq 0.5$ معمولاً جواب‌های مناسبی به‌دست می‌آید.

چند نکته در خصوص الگوریتم $ILUT$ قابل ذکر است. در حالت کلی، حتی اگر تجزیه LU ماتریس A موجود باشد، دلیلی وجود ندارد که تجزیه ناقص LU ماتریس نیز وجود داشته باشد. یعنی حتی اگر ماتریس A دارای تجزیه LU باشد، آنگاه ممکن است در محاسبه تجزیه ناقص ماتریس A ، یک درایه محوری صفر شود. در عمل دو طریق برای فائق آمدن بر این مشکل وجود دارد. یکی اینکه اگر یک درایه محوری صفر شد، آن را با یک عدد کوچک جایگزین کنیم و دیگر اینکه از عمل محورگیری (جزئی یا کلی) استفاده کنیم. در روش اول، کیفیت پیش‌شرط‌ساز به‌دست‌آمده ممکن است مناسب نباشد، و در روش دوم، پیچیدگی محاسباتی افزایش می‌یابد.

یکی از مشکلات الگوریتم $ILUT$ این است که اگر A متقارن باشد، آنگاه در حالت کلی ماتریس

$$M = LU,$$

متقارن نمی‌باشد؛ به عبارت دیگر، الگوریتم $ILUT$ تقارن را لزوماً حفظ نمی‌کند. نسخه‌های دیگری از تجزیه LU با نام‌های $ILUS$ و $BILUS$ وجود دارد که تقارن را حفظ می‌کنند، که به‌زودی الگوریتم $ILUS$ را معرفی می‌کنیم [۶۱، ۲۷].

نسخه‌ای از ILUT که متلب از آن استفاده می‌کند کمی متفاوت است. در نسخه متلب، درایه‌هایی که در شرایط

$$|U(i, j)| \geq \mu \|A(:, j)\|_1, \quad |L(i, j)| \geq \frac{\mu \|A(:, j)\|_1}{|U(j, j)|},$$

صدق کنند نگهداری و سایر درایه‌ها چکانده می‌شوند. اگر یک درایه محوری در طول محاسبه تجزیه ناقص صفر شود، این درایه با μ جایگزین می‌شود. به علاوه، این امکان هم وجود دارد که کاربر بتواند از محورگیری جزئی نیز استفاده کند.

تعریف ۱.۹ ماتریس $A \in \mathbb{R}^{n \times n}$ یک H-ماتریس نامیده می‌شود هرگاه ماتریس مقایسه آن، یعنی $\langle A \rangle = (b_{ij})$ یک M-ماتریس باشد که در آن

$$b_{ij} = \begin{cases} -|a_{ij}|, & i \neq j, \\ |a_{ii}|, & i = j. \end{cases}$$

مثال ۳.۹ ماتریس

$$A = \begin{pmatrix} -4 & 1 & -3 \\ -2 & 3 & 0 \\ -2 & 4 & 5 \end{pmatrix},$$

یک H-ماتریس است؛ زیرا ماتریس

$$\langle A \rangle = \begin{pmatrix} 4 & -1 & -3 \\ -2 & 3 & 0 \\ -2 & -4 & 5 \end{pmatrix},$$

یک M-ماتریس است، چون اولاً یک L-ماتریس است و ثانیاً داریم

$$A^{-1} = \frac{1}{8} \begin{pmatrix} 15 & 17 & 9 \\ 10 & 14 & 6 \\ 14 & 18 & 10 \end{pmatrix} \geq 0.$$

در [۷۴] نشان داده شده است که تجزیه ناقص LU برای H-ماتریس‌ها شکست نمی‌خورد. می‌توان دید که هر ماتریس غالب قطری اکید نیز یک H-ماتریس است. همچنین، روشن است هر M-ماتریس

یک H-ماتریس است. لذا، تجزیه ناقص LU برای ماتریس‌هایی که غالب قطری هستند یا یک M-ماتریس هستند نیز شکست نمی‌خورد ([۷۶] را هم ببینید).
در متلب، برای به‌کارگیری الگوریتم ILU(0) می‌توان از دستور

```
setup.type='nofill';  
[L,U]=ilu(A,setup);
```

استفاده کرد. لازم به ذکر است که به‌جای دو دستور بالا، می‌توان از دستور معادل با آن، یعنی

```
[L,U]=ilu(A);
```

استفاده کرد. درحقیقت، اگر پارامتر setup مشخص نشود، متلب به‌طور پیش‌فرض از الگوریتم ILU(0) استفاده می‌کند. برای به‌کارگیری ILUT با پارامتر μ ، می‌توان از دستورات زیر استفاده کرد (در اینجا، باید پارامتر $\mu = \mu$ وارد شود).

```
setup.type='ilutp';  
setup.droptol=mu;  
[L,U]=ilu(A,setup);
```

مثال ۴.۹ دستگاه مورد استفاده در مثال ۱.۹ را به‌ازای $N = 100, 150, 200$ در نظر بگیرید. سایر فرض‌ها نیز همانند مثال ۱.۹ هستند. این دستگاه‌ها را با نسخه کامل الگوریتم GMRES بدون پیش‌شرط‌ساز، با پیش‌شرط‌ساز ILU(0)، و پیش‌شرط‌ساز ILUT به‌ازای $\mu = 0.05, 0.1$ حل کرده‌ایم. پیش‌شرط‌سازی از راست انجام شده است. نتایج به‌دست آمده در جدول ۹-۱ ارائه شده‌اند. در این جدول، زمان محاسبه پیش‌شرط‌ساز با $ptime$ ، تعداد تکرارها برای همگرایی با $iters$ و زمان لازم برای همگرایی با $ittime$ نشان داده شده‌اند. این جدول به‌خوبی تأثیر پیش‌شرط‌ساز مبتنی بر تجزیه ناقص LU را در کاهش تعداد تکرارهای نسخه کامل GMRES نشان می‌دهد.

مثال ۵.۹ در این مثال، سه ماتریس ADD20، ADD32، و MEMPLUS را از بازار ماتریس با وبگاه اشاره‌شده در [۷۵] (یا از [۳۳]) انتخاب نمودیم. این ماتریس‌ها با خواص عمومی‌شان در جدول ۹-۲ ارائه شده‌اند. سه دستگاه با این سه ماتریس می‌سازیم. تمام فرض‌ها (حدس اولیه، بردار سمت راست، تولرانس و ...) همانند مثال قبل هستند. این دستگاه‌ها را با استفاده نسخه کامل GMRES حل کرده‌ایم. نتایج به‌دست آمده نیز در جدول ۹-۲ گزارش شده‌اند. نسخه کامل GMRES با پیش‌شرط‌ساز ILUT به‌ازای $\mu = 0.1, 0.01$ برای حل این سه دستگاه استفاده شده است. پیش‌شرط‌سازی از راست انجام گرفته

جدول ۹-۱ نتایج عددی برای مثال ۴.۹.

ILU(0)							
$n = N^2$	nnz	$ptime$	ρ	No Precon.		Precon.	
				$iters$	$itime$	$iters$	$itime$
۱۰۰۰۰	۴۹۶۰۰	۰/۰۰	۱	۱۹۸	۱/۳۴	۵۸	۰/۱۶
۲۲۵۰۰	۱۱۱۹۰۰	۰/۰۰	۱	۲۹۵	۵/۱۴	۸۶	۰/۵۷
۴۰۰۰۰	۱۹۹۲۰۰	۰/۰۱	۱	۳۹۱	۱۴/۵۸	۱۱۴	۱/۵۶
ILUT, $\mu = ۰/۰۵$							
$n = N^2$	nnz	$ptime$	ρ	$iters$	$itime$	$iters$	$itime$
۱۰۰۰۰	۴۹۶۰۰	۰/۰۳	۱/۲۰	۱۹۸	۱/۳۴	۴۴	۰/۱۱
۲۲۵۰۰	۱۱۱۹۰۰	۰/۰۸	۱/۴۰	۲۹۵	۵/۱۴	۵۲	۰/۲۵
۴۰۰۰۰	۱۹۹۲۰۰	۰/۱۹	۱/۴۰	۳۹۱	۱۴/۵۸	۶۹	۰/۶۷
ILUT, $\mu = ۰/۰۱$							
$n = N^2$	nnz	$ptime$	ρ	$iters$	$itime$	$iters$	$itime$
۱۰۰۰۰	۴۹۶۰۰	۰/۰۴	۲/۳۵	۱۹۸	۱/۳۴	۲۱	۰/۰۵
۲۲۵۰۰	۱۱۱۹۰۰	۰/۰۹	۲/۱۸	۲۹۵	۵/۱۴	۳۲	۰/۱۴
۴۰۰۰۰	۱۹۹۲۰۰	۰/۲۱	۲/۱۸	۳۹۱	۱۴/۵۸	۴۳	۰/۳۳

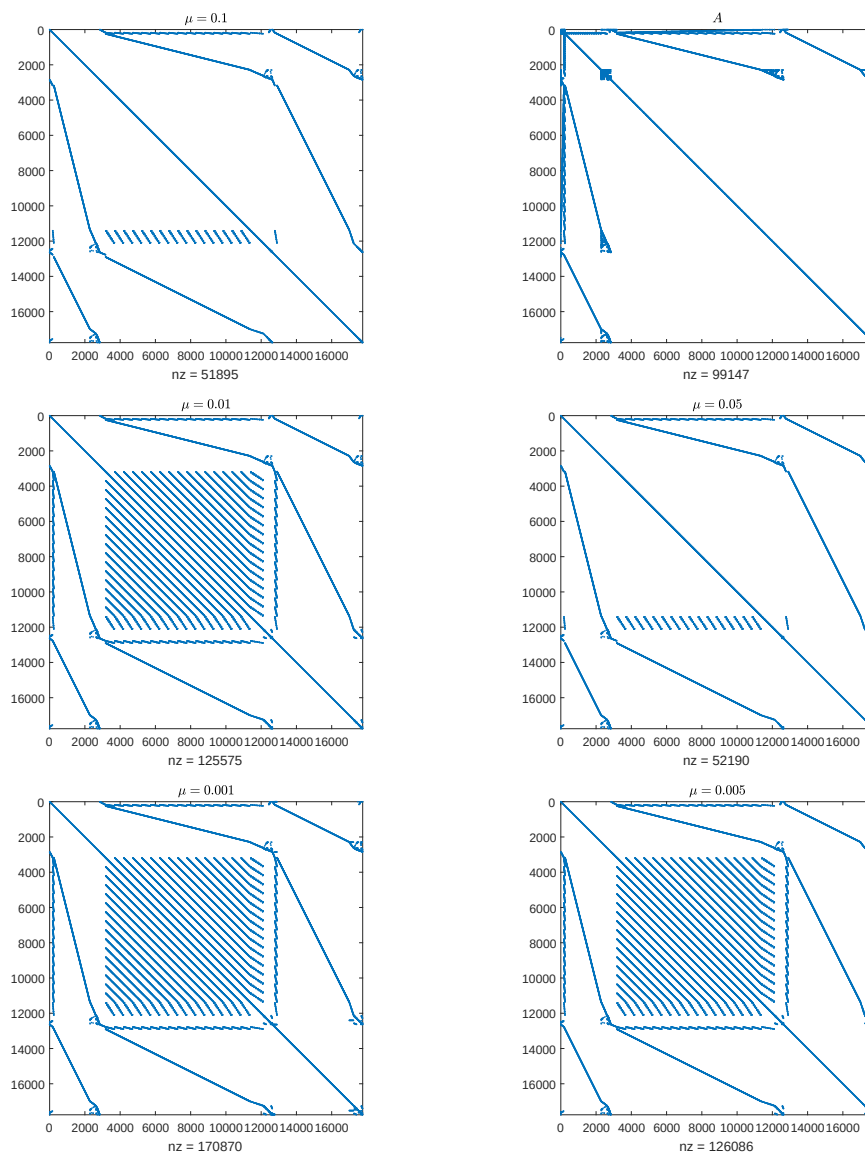
جدول ۹-۲ خواص عمومی ماتریس‌های مثال ۵.۹.

نام ماتریس	n	nnz	$iters$	$itime$
ADD20	۲۳۹۵	۱۳۱۵۱	۲۵۳	۰/۴۶
ADD32	۴۹۶۰	۱۹۸۴۸	۶۸	۰/۲۰
MEMPLUS	۱۷۷۵۸	۹۹۱۴۷	۵۴۱	۱۳/۶۲

جدول ۹-۳ نتایج عددی برای مثال ۵.۹.

نام ماتریس	ILUT, $\mu = ۰/۱$				ILUT, $\mu = ۰/۰۱$			
	ρ	$ptime$	$iters$	$itime$	ρ	$ptime$	$iters$	$itime$
ADD20	۰/۴۷	۰/۰۰	۳۵	۰/۰۳	۰/۵۱	۰/۰۰	۱۱	۰/۰۲
ADD32	۰/۶۴	۰/۰۰	۹	۰/۰۲	۱/۱۷	۰/۰۱	۷	۰/۰۲
MEMPLUS	۰/۵۲	۰/۰۱	۱۳۴	۱/۰۱	۱/۲۷	۰/۰۳	۲۴	۰/۰۹

است. نتایج به‌دست‌آمده در جدول ۹-۳ ارائه شده‌اند. تمام مفروضات همانند مثال قبل و نتایج به‌ازای $\mu = ۰/۱$ و $\mu = ۰/۰۱$ ارائه شده‌اند. این جدول به خوبی کارایی پیش‌شرط‌ساز ILUT را نشان می‌دهد. در شکل ۹-۱، تُنکی ماتریس ADD20 و $L + U$ به‌ازای μ های مختلف نمایش داده شده است. این شکل به خوبی تأثیر پارامتر μ بر تُنکی عامل‌های تجزیه ناقص LU را نمایش می‌دهد.



شکل ۹-۱ ساختار تنگی ماتریس MEMPLUS و ماتریس $L + U$ مربوطه به‌ازای مقادیر مختلف μ برای مثال ۵.۹.

۹-۲-۲ تجزیه چولسکی ناقص

فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن باشد. با استفاده از قضیه ۱۳.۱ می‌بینیم که می‌توان ماتریس A را به صورت

$$A = LL^T,$$

نوشت که در آن L یک ماتریس پائین‌مثلثی با درایه‌های قطری مثبت است (تجزیه چولسکی). الگوریتم ۵.۹ ماتریس L را به دست می‌دهد [۱].

الگوریتم ۵.۹ تجزیه چولسکی

1. Set $L = (\ell_{ij}) = 0 \in \mathbb{R}^{n \times n}$
2. For $j = 1, \dots, n$, Do:
3.
$$\ell_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} \ell_{jk}^2 \right)^{1/2}$$
4. For $i = j + 1, \dots, n$, Do:
5.
$$\ell_{ij} := \left(a_{ij} - \sum_{k=1}^{j-1} \ell_{ik} \ell_{jk} \right) / \ell_{jj}$$
6. EndIf
7. EndDo

برای محاسبه تجزیه ناقص چولسکی، کافی است عمل چکاندن به الگوریتم ۵.۹ اضافه شود. به ازای $\mu > 0$ ، الگوریتم تجزیه چولسکی ناقص به صورت الگوریتم ۶.۹ نوشته می‌شود.

الگوریتم ۶.۹ تجزیه چولسکی ناقص ($\text{ICT}(\mu)$)

1. Set $L = (\ell_{ij}) = 0 \in \mathbb{R}^{n \times n}$
2. For $j = 1, \dots, n$, Do:
3.
$$\ell_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} \ell_{jk}^2 \right)^{1/2}$$
4. For $i = j + 1, \dots, n$ Do:
5.
$$t := \left(a_{ij} - \sum_{k=1}^{j-1} \ell_{ik} \ell_{jk} \right) / \ell_{jj}$$
6. If $|t| \geq \mu \|A(j:n, j)\|_1$, then $\ell_{ij} := t$
7. EndIf
8. EndDo

همانند الگوریتم $\text{ILU}(0)$ ، اگر در گام ۶ از الگوریتم ۶.۹ شرط

6. If $(i, j) \in \text{NZ}(A)$, then $\ell_{ij} := t$

جایگزین شود، آنگاه الگوریتم حاصل را الگوریتم $\text{IC}(0)$ می‌نامیم.

در متلب، برای به کارگیری الگوریتم $\text{IC}(0)$ ، می‌توان از دستورات

```
setup.type='nofill';
```

```
L=ichol(A,setup);
```

جدول ۴-۹ خواص ماتریس‌ها و نتایج مربوط به گرادیان مزدوج برای مثال ۶.۹.

نام ماتریس	خواص ماتریس		بدون پیش‌شرط‌سازی	
	n	nnz	$iters$	$time$
NOS5	۴۶۸	۵۱۷۲	۳۶۸	۰/۱۱
NOS7	۷۲۹	۴۶۱۷	۹۴۱	۰/۱۵
APACHE1	۸۰۸۰۰	۵۴۲۱۸۴	۹۹۵	۳/۵۱

استفاده کرد. توجه کنید که به‌جای دو دستور بالا، می‌توان از دستور معادل آن، یعنی

`L=ichol(A);`

نیز استفاده کرد. درحقیقت، اگر پارامتر `setup` مشخص نشود، متلب به‌طور پیش‌فرض از الگوریتم `IC(0)` استفاده می‌کند. برای به‌کارگیری `ICT( $\mu$ )`، می‌توان از دستورات زیر استفاده کرد (در اینجا، باید پارامتر $\mu = \mu$ وارد شود).

```
setup.type='ict';
setup.droptol=mu;
L=ichol(A,setup);
```

مثال ۶.۹ در این مثال، سه ماتریس NOS5، NOS7، و APACHE1 از مجموعهٔ ماتریس‌های «بازار ماتریس» [۷۵] (یا از [۳۳]) انتخاب نمودیم. تمام این ماتریس‌ها معین مثبت متقارن هستند. سه دستگاه معادلات خطی با این ماتریس‌ها ساخته و بردار سمت راست دستگاه‌ها را $b = A(1, 1, \dots, 1)^T$ گرفتیم. از روش گرادیان مزدوج و گرادیان مزدوج پیش‌شرط‌سازی‌شده با پیش‌شرط‌ساز چولسکی ناقص (نسخهٔ متلب با $\mu = 0.05$) استفاده کرده‌ایم. حدس اولیه را بردار صفر، شرط توقف را

$$\|r_k\|_2 \leq 10^{-7} \|r_0\|_2,$$

و حداکثر تعداد تکرارها را ۱۰۰۰ گرفته‌ایم. خواص ماتریس‌ها و نتایج مربوط به الگوریتم گرادیان مزدوج در جدول ۴-۹ گزارش شده‌اند. همچنین، نتایج حاصله برای دستگاه پیش‌شرط‌سازی‌شده در جدول ۵-۹ ارائه شده‌اند. همان‌گونه که دیده می‌شود تجزیهٔ چولسکی ناقص یک پیش‌شرط‌ساز مناسب برای حل این دستگاه‌هاست.

جدول ۹-۵ نتایج عددی مربوط به دستگاه پیش‌شرط‌سازی شده برای مثال ۶.۹.

نام ماتریس	$ptime$	ρ	$iters$	$time$
NOS5	۰/۰۰	۱/۵۱	۲۵	۰/۰۳
NOS7	۰/۰۰	۱/۲۷	۱۸	۰/۰۳
APACHE1	۰/۰۳	۱/۴۱	۲۷۰	۲/۶۰

۹-۲-۳ بهبود کارایی تجزیه چولسکی با کمک بازمرتب‌سازی

یکی از شیوه‌های کاهش حافظه مورد نیاز برای محاسبه تجزیه چولسکی یک ماتریس (یا نسخه ناقص آن)، استفاده از راه‌کار بازمرتب‌سازی است. دستگاه $Ax = b$ را در نظر بگیرید. در راه‌کار بازمرتب‌سازی، یک ماتریس جایگشت مثل P محاسبه می‌شود به‌طوری‌که تعداد درایه‌های عامل تجزیه چولسکی ماتریس PAP^T از تعداد درایه‌های ناصفر عامل تجزیه چولسکی ماتریس A کمتر باشد. سپس، به جای حل دستگاه $Ax = b$ ، دستگاه

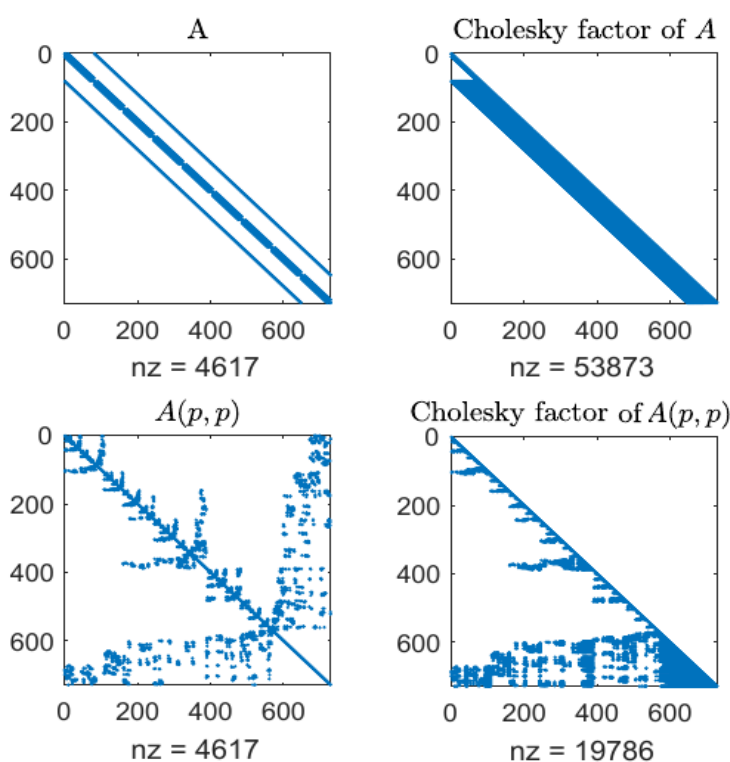
$$PAP^T u = Pb, \quad x = P^T u, \quad (9-1)$$

حل می‌شود. مهم‌ترین الگوریتم برای بازمرتب‌سازی یک دستگاه با ماتریس ضرایب معین مثبت متقارن، الگوریتم مینیمم درجه تقریبی متقارن (SYMAMD) است. این الگوریتم به‌طور تقریبی بهترین جایگشت P را به دست می‌دهد (جایگشتی که عامل چولسکی ماتریس PAP^T تقریباً کمترین درایه را از بین همه جایگشت‌ها دارد). بنابراین، نه تنها حافظه ذخیره‌سازی عامل تجزیه چولسکی ماتریس کاهش می‌یابد، بلکه حل دستگاه با کمک تجزیه چولسکی ماتریس ضرایب با هزینه کمتری انجام می‌گیرد. برای مطالعه جزئیات این الگوریتم به [۳۸] مراجعه کنید. در متلب، برای الگوریتم SYMAMD یک تابع با همین نام (symamd) وجود دارد.

در هر تکرار از یک روش تکراری مبتنی بر زیرفضای کرایلف برای حل یک دستگاه معادلات خطی پیش‌شرط‌سازی شده، ممکن است نیاز به حل یک یا چند زیردستگاه با ماتریس ضرایب معین مثبت متقارن باشد. در این صورت، استفاده از بازمرتب‌سازی به‌طور قابل توجهی زمان اجرای برنامه را کاهش می‌دهد. برای محاسبه ماتریس P (به‌طور ضمنی) و حل دستگاه (۹-۱) با کمک تجزیه چولسکی ماتریس PAP^T ، می‌توانیم از دستورات زیر استفاده کنیم.

```
p=symamd(A);
L=chol(A(p,p),'lower');
x(p,:)=L'\(L\b(p));
```

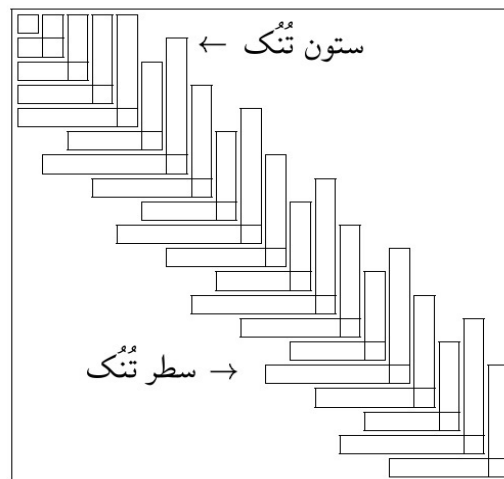
برای مثال ماتریس NOS7 از مجموعه ماتریس‌های مثال ۶.۹ را در نظر بگیرید. در شکل ۹-۲، الگوی ناصفر ماتریس‌های $A = \text{NOS7}$ و $A(p, p)$ و عامل‌های تجزیه چولسکی A و $A(p, p)$ نمایش داده شده‌اند. همان‌گونه که می‌بینیم عامل تجزیه چولسکی ماتریس $A(p, p)$ به مراتب تُنک‌تر از عامل تجزیه چولسکی ماتریس A است. لازم به ذکر است که تابع `symamd` را می‌توان برای چولسکی ناقص (دستور `ichol`) نیز به کار برد.



شکل ۹-۲ ساختار تُنکی ماتریس NOS7 (پیش و بعد از بازمرتب‌سازی)، و ساختار تُنکی عامل تجزیه چولسکی آنها.

۹-۲-۴ الگوریتم ILUS

یکی از مشکلات الگوریتم ILUT این است که اگر A متقارن باشد، لزوماً $L = U^T$ نخواهد بود؛ به عبارت دیگر، الگوریتم ILUT تقارن را لزوماً حفظ نمی‌کند. برای دسته‌ای از ماتریس‌ها این مشکل را می‌توان حل کرد. برای حل بسیاری از مسائل عملی (به‌ویژه در دینامیک سیالات) ماتریس‌هایی



شکل ۹-۳ نمایی از یک ماتریس در قالب اسکی‌لاین

به‌دست می‌آیند که در قالب اسکی‌لاین ذخیره می‌شوند. ساده‌ترین نوع قالب اسکی‌لاین برای ذخیره‌سازی ماتریس‌های متقارن (SSK) این است که سطرهای ماتریس، از سطر اول تا سطر n ام را در یک بردار حقیقی ذخیره کرده و بردار صحیح دیگر را که درایه‌های آن نشان دهنده ابتدای هر سطر می‌باشد را در نظر بگیریم [۲۴]. مؤلفه ستونی درایه‌های ماتریس به‌سادگی به‌دست می‌آیند و نیازی به ذخیره‌سازی آنها نیست. ساختمان داده برای ماتریس‌هایی که در قالب اسکی‌لاین نامتقارن (NSK) ذخیره می‌شوند شامل دو قسمت می‌باشد. در اینجا قسمت پائین‌مثلثی ماتریس A و ترانزپوز قسمت بالامثلثی اکید آن را در فرمت اسکی‌لاین متقارن ذخیره می‌کنند. شکل ۹-۳، نمایی از یک ماتریس در قالب اسکی‌لاین را نشان می‌دهد. در [۲۷]، چو و سعد الگوریتم تجزیه ILUS را ارائه کردند که برای ماتریس‌هایی که در قالب اسکی‌لاین

ذخیره می‌شوند مناسب است و به‌علاوه، تقارن را نیز حفظ کند. در ادامه این الگوریتم را مطالعه می‌کنیم. الگوریتم ILUS، تجزیه ناقص LDU ماتریس A را به‌صورت بازگشتی و با محاسبه تجزیه LDU زیرماتریس‌های اصلی پیشروی A محاسبه می‌کند. نخست، قرار می‌دهیم $A_1 = (a_{11})$. تجزیه LDU ماتریس A_1 به‌صورت $A_1 = L_1 D_1 U_1$ است که در آن

$$L_1 = (1), \quad D_1 = (d_1), \quad U_1 = (1).$$

فرض کنید

$$A_k = L_k D_k U_k,$$

۳۲۴ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

تجزیه LDU زیرماتریس اصلی پیشروی مرتبه k باشد، که در آن D_k یک ماتریس قطری است و ماتریس‌های L_k و U_k^T پائین‌مثلثی با درایه‌های قطری ۱ هستند. در این صورت، یک تجزیه LDU برای ماتریس

$$A_{k+1} = \begin{pmatrix} A_k & v_k \\ w_k & \alpha_{k+1} \end{pmatrix}, \quad (2-9)$$

می‌سازیم. تجزیه LDU ماتریس A_{k+1} به صورت

$$A_{k+1} = L_{k+1} D_{k+1} U_{k+1} = \begin{pmatrix} L_k & \circ \\ y_k & 1 \end{pmatrix} \begin{pmatrix} D_k & \circ \\ \circ & d_{k+1} \end{pmatrix} \begin{pmatrix} U_k & z_k \\ \circ & 1 \end{pmatrix}, \quad (3-9)$$

است که در آن بردارهای z_k و y_k و عدد d_{k+1} بایستی محاسبه شوند. این رابطه معادل است با

$$A_{k+1} = \begin{pmatrix} A_k & v_k \\ w_k & \alpha_{k+1} \end{pmatrix} = \begin{pmatrix} L_k D_k U_k & L_k D_k z_k \\ y_k D_k U_k & y_k D_k z_k + d_{k+1} \end{pmatrix}.$$

بنابراین،

$$L_k D_k z_k = v_k,$$

$$y_k D_k U_k = w_k,$$

$$y_k D_k z_k + d_{k+1} = \alpha_{k+1},$$

که رابطه‌های زیر را برای محاسبه z_k ، y_k و δ_{k+1} نتیجه می‌دهد

$$z_k = D_k^{-1} L_k^{-1} v_k, \quad (4-9)$$

$$y_k = w_k U_k^{-1} D_k^{-1}, \quad (5-9)$$

$$d_{k+1} = \alpha_{k+1} - y_k D_k z_k. \quad (6-9)$$

بنابراین، هر سطر و ستون از عامل‌ها را می‌توان با حل یک دستگاه مثلثی به دست آورد. روشن است

که با استفاده از روند بالا و به صورت بازگشتی می‌توان تجزیه LDU ماتریس $A = A_n$ را به دست آورد.

برای محاسبه تجزیه ناقص LDU کافی است بردارهای z_k و y_k به صورت تقریبی و تنگ محاسبه

شوند. برای این کار، ابتدا فرض کنید $E_k = I - L_k$. در این صورت، E_k یک ماتریس پائین‌مثلثی اکید

است و در نتیجه، داریم $\rho(E_k) = 0$. از این رو، با استفاده از لم ۲.۲، خواهیم داشت

$$L_k^{-1} = (I - E_k)^{-1} = \sum_{j=0}^{\infty} E_k^j.$$

در نتیجه، برای محاسبه تقریبی z_k ، به ازای یک مقدار کوچک از p ، قرار می‌دهیم

$$z_k = D_k^{-1} L_k^{-1} v_k = D_k^{-1} (I + E_k + E_k^2 + \dots + E_k^p) v_k. \quad (7-9)$$

توجه کنید که در اینجا نیازی به محاسبه صریح E_k^j نیست و بردار z_k را می‌توان با استفاده از الگوریتم هورنر و تنها با استفاده از ضرب ماتریس در بردار محاسبه نمود. برای مثال، به ازای $p = 3$ ، داریم

$$z_k = D_k^{-1} L_k^{-1} v_k = D_k^{-1} (v_k + E_k(v_k + E_k(v_k + E_k v_k))),$$

که با کمک روند زیر، می‌توان بردار z_k را محاسبه کرد

$$s_1 = v_k + E_k v_k,$$

$$s_2 = v_k + E_k s_1,$$

$$s_3 = v_k + E_k s_2,$$

$$z_k = D_k^{-1} s_3.$$

برای حفظ تنگی عامل L از یک قاعده چکاندن برای z_k استفاده می‌کنیم. برای مثال، ابتدا درایه‌هایی که از μ کوچک‌تر هستند چکانده می‌شوند و سپس به تعداد ℓ تا از بزرگ‌ترین درایه‌ها (از لحاظ قدرمطلق) نگهداری شده و سایر درایه‌ها نیز چکانده می‌شوند. به طور مشابه، روند بالا را می‌توان برای محاسبه بردار y_k نیز به کار برد. اکنون، می‌توانیم الگوریتم ILUS را به صورت زیر جمع‌بندی کنیم.

الگوریتم ۷.۹ تجزیه ILUS

1. Set $D_1 := (a_{11})$ and $L_1 = U_1 := (1)$
2. For $k = 1, \dots, n - 1$ Do:
3. Compute a sparse $z_k \approx D_k^{-1} L_k^{-1} v_k$
4. Compute a sparse $y_k \approx w_k U_k^{-1} D_k^{-1}$
5. Compute $d_{k+1} := \alpha_{k+1} - y_k D_k z_k$
6. Form L_{k+1} , D_{k+1} and U_{k+1}
7. EndDo

زمانی‌که ماتریس A متقارن است داریم $U = L^T$ و بنابراین، محاسبه y_k (و در نتیجه، ماتریس U_k) حذف می‌شود. در این صورت، محاسبات تقریباً نصف خواهد شد.

۳-۹ پیش‌شرط‌سازهای مبتنی بر معکوس تقریبی تُنک

روش‌های پیش‌شرط‌سازی بر پایه معکوس تقریبی تُنک در سال‌های اخیر بسیار توسعه یافته است. همان‌گونه که پیش‌تر اشاره شد، ایده اصلی در این روش این است که یک ماتریس تُنک مثل M به دست آید به‌طوری‌که یک تقریب از معکوس A باشد. ماتریس M به‌طور صریح محاسبه می‌شود و به‌عنوان ماتریس پیش‌شرط‌ساز استفاده می‌گردد. مزیت اصلی این روش بر روش‌های دیگر این است که می‌توان آن را به‌صورت موازی به‌کار گرفت، زیرا به‌کارگیری آن فقط شامل ضرب ماتریس در بردار است. به‌علاوه، نشان داده شده است که پیش‌شرط‌سازهایی که بر پایه معکوس تقریبی تُنک می‌باشند در عمل بسیار کارا هستند. معکوس تقریبی تُنک بر این نکته تکیه دارد که برای یک ماتریس تُنک می‌توان ماتریس تُنک دیگری مثل M یافت به‌طوری‌که یک تقریب از A^{-1} باشد. این مسئله در حالت کلی روشن نیست، زیرا معکوس یک ماتریس، معمولاً چگال است. به‌طور دقیق‌تر می‌توان گفت که اگر A تحویل‌ناپذیر و تُنک باشد، آنگاه معکوس آن یک ماتریس پُر است. یعنی، برای الگوی ناصفر تُنک تحویل‌ناپذیر داده‌شده از A ، می‌توان درایه‌های ناصفر را به‌گونه‌ای تعریف کرد که همه درایه‌های ماتریس معکوس آن ناصفر باشند [۱۵]. با این وجود، حالت‌های زیادی وجود دارد که بسیاری از درایه‌های ماتریس معکوس از لحاظ قدرمطلق کوچک هستند و بنابراین، تقریب‌زدن A^{-1} با یک ماتریس تُنک امکان‌پذیر است. برای مثال، نتیجه‌ای که دمکو و همکارانش [۳۰] به دست آوردند بیانگر این مطلب است که اگر ماتریس A معین مثبت متقارن و به‌صورت نواری باشد آنگاه درایه‌های معکوس آن در طول هر سطر یا ستون به‌صورت نمایی کاهش می‌یابند. به‌طور دقیق‌تر، ثابت‌هایی مثل $1 > \rho > 0$ و C وجود دارند به‌طوری‌که

$$|(A^{-1})_{ij}| \leq C\rho^{|i-j|}, \quad i, j = 1, 2, \dots, n,$$

که در آن، اعداد ρ و C به پهنای نوار و عدد حالت ماتریس A بستگی دارند. برای ماتریس‌هایی که پهنای نوار بزرگ یا عدد حالت بزرگ دارند ممکن است که C خیلی بزرگ و ρ به یک نزدیک باشد و در نتیجه، سرعت کاهش خیلی کند باشد.

همان‌گونه که در ابتدای فصل ۸ اشاره شد، به دو صورت می‌توان معکوس یک ماتریس را به‌طور تقریبی محاسبه کرد. صورت اول این است که ماتریس پیش‌شرط‌ساز $M \approx A^{-1}$ را به‌صورت یک

ماتریس ساده به‌دست آوریم و دیگری این است که معکوس تقریبی را به‌صورت حاصل‌ضرب دو یا چند ماتریس بنویسیم. حالت دوم، به‌نام معکوس تقریبی تجزیه‌شده مشهور است و به‌صورت

$$M = M_U M_L,$$

است که در آن $M_U \approx U^{-1}$ ، $M_L \approx L^{-1}$ و U و L به‌ترتیب عامل‌های پائین و بالامثلثی ماتریس A هستند. در ادامه، چند روش رایج برای محاسبه معکوس تقریبی $\tilde{M}$ را مطالعه می‌کنیم.

۹-۳-۱ تقریب مستقیم معکوس ماتریس

ایده اصلی محاسبه ماتریس $M \approx A^{-1}$ جواب مسئله کمینه‌سازی

$$\min_{M \in S} \|I - AM\|_F,$$

است که در آن S مجموعه ماتریس‌های $\tilde{M}$ ، I ماتریس همانی و $\|\cdot\|_F$ نرم فروبنیوس است. از اینکه

$$\|I - AM\|_F^2 = \sum_{j=1}^n \|e_j - Am_j\|^2,$$

که در آن e_j ستون j ام ماتریس همانی است، محاسبه ماتریس M تبدیل به n مسئله کمترین توان‌های دوم مستقل از هم با قیدهای $\tilde{M}$ می‌شود. توجه کنید که روش بالا ماتریس پیش‌شرط‌ساز از راست را تولید می‌کند. یک ماتریس پیش‌شرط‌ساز چپ می‌تواند از کمینه‌سازی با محدودیت

$$\|I - MA\|_F = \|I - A^T M^T\|_F,$$

به‌دست آید. این عمل را می‌توان با به‌دست آوردن ماتریس پیش‌شرط‌ساز راست برای A^T و سپس با اختیارکردن ترانزپوز آن، انجام داد. فرض کنید S مجموعه ماتریس‌های $\tilde{M}$ باشد که دارای الگوی ناصفر خاصی هستند و این الگو از پیش تعیین شده باشد. بعد از اینکه این الگو داده شد، محاسبه ماتریس پیش‌شرط‌ساز بسیار ساده است و این عمل برای کامپیوترهای موازی بسیار مناسب است. زمانی که الگوی ناصفر از پیش معلوم باشد، محاسبه ماتریس پیش‌شرط‌ساز می‌تواند به‌صورت زیر انجام شود. الگوی ناصفر، یک مجموعه به‌صورت

$$G \subseteq \{(i, j) \mid 1 \leq i, j \leq n\},$$

است به‌گونه‌ای که به‌ازای هر $(i, j) \notin G$ داریم $m_{ij} = 0$. بنابراین، S مجموعه همه ماتریس‌های

$n \times n$ است به‌طوری‌که الگوی ناصفر آنها در G باشد. فرض کنید m_j ستون j ام ماتریس M و برای یک j ثابت،

$$J = \{i \mid (i, j) \in G\},$$

مجموعه اندیس‌های متناظر با درایه‌های ناصفر m_j باشد. فرض کنید $A(:, J)$ زیرماتریسی از A باشد که اندیس ستون‌های آن در J هستند. اکنون، فرض کنید I مجموعه اندیس سطرهاى ناصفر $A(:, J)$ باشد. توجه خود را به ماتریس $\hat{A} = A(I, J)$ ، به بردار نامشخص $\hat{m}_j = m_j(I)$ و به بردار طرف دوم $\hat{e}_j = e_j(I)$ معطوف می‌کنیم. در این صورت، برای به‌دست آوردن درایه‌های ناصفر m_j می‌توان مسئله کمترین توان‌های دوم زیر را حل کرد که یک مسئله با ابعاد کوچک است

$$\|\hat{e}_j - \hat{A}\hat{m}_j\|_2 = \min_{\hat{m}} \|\hat{e}_j - \hat{A}\hat{m}\|_2.$$

به‌طور مثال می‌توان این مسئله را با استفاده از تجزیه QR ماتریس $\hat{A}$ حل کرد. واضح است که هر ستون m_j می‌تواند مستقل از ستون‌های دیگر محاسبه شود. در اینجا باید توجه کرد که معمولاً $\hat{A}$ ماتریسی با بُعد کوچک است و بنابراین، هرکدام از مسئله‌های کمترین توان‌های دوم، با بعد کوچک هستند و می‌توانند با استفاده از روش‌های مستقیم حل شوند (قضیه ۲۴.۱ را ببینید). اما مسئله اصلی این است که چگونه الگوی ناصفر را انتخاب کنیم. به بیان دقیق‌تر، الگوی ناصفر M را چگونه اختیار کنیم تا یک ماتریس پیش‌شرط‌ساز خوب داشته باشیم. یکی از راه‌ها این است که الگوی ناصفر M را همانند A بگیریم، اما این انتخاب همیشه جواب‌های خوبی به‌دست نمی‌دهد.

با توجه به اینکه انتخاب یک الگوی ناصفر برای M در حالت کلی مشکل است، چندین مؤلف یک الگوی ناصفر کوچک (مثلاً، یک ماتریس قطری) را انتخاب کرده و به‌تدریج این الگو را بهبود می‌بخشند تا اینکه معیار توقف

$$\|e_j - Am_j\|_2 < \epsilon,$$

برای ϵ داده‌شده، برقرار شود، یا به حداکثر درایه‌های ناصفر مجاز در m_j برسیم. مهم‌ترین الگوریتم ارائه‌شده از گروت و هاگل است که به‌نام الگوریتم SPAI مشهور است [۲۴]. برای جزئیات بیشتر درباره این الگوریتم به [۲۴] مراجعه کنید.

روش‌های دیگری در [۲۸] توسط چو و سعد ارائه شده‌اند. در این روش‌ها کمینه‌سازی $\|I - AM\|_F$ با حل تقریبی n دستگاه معادلات خطی

$$Am_j = e_j, \quad j = 1, \dots, n, \quad (8-9)$$

توسط یک روش تکراری انجام می‌شود. درحقیقت، برای این کار آنها تعداد کمی تکرار از یک روش از نوع مینیمم مانده را استفاده می‌کنند. مهم‌ترین روش بر پایه الگوریتم MR (الگوریتم ۳.۵) است. در اینجا به الگوی ناصفر نیاز نمی‌باشد و درایه‌های بزرگی که در تکرارها ظاهر می‌شوند، نگهداری شده و سایر درایه‌ها حذف می‌شوند.

در روش MR، ابتدا یک حدس اولیه تُنک انتخاب شده و سپس تعداد کمی تکرار از الگوریتم MR برای حل دستگاه‌های معادلات خطی (۸-۹) به کار می‌رود. این الگوریتم به‌طور خلاصه به‌صورت زیر نوشته می‌شود.

الگوریتم ۸.۹ الگوریتم MR برای حل (۸-۹) (نسخه اولیه)

1. Set $M := M_0$
2. For each column $j = 1, \dots, n$, Do
3. Define $m_j = Me_j$
4. For $i = 1, \dots, n_j$ Do
5. $r_j := e_j - Am_j$
6. $\alpha_j := \langle r_j, Ar_j \rangle / \langle Ar_j, Ar_j \rangle$
7. $m_j := m_j + \alpha_j r_j$
8. Apply numerical dropping to m_j
9. EndDo
10. EndDo

در گام ۸ از الگوریتم عمل چکاندن درایه‌های کوچک برای حفاظت از تُنکی انجام می‌شود. این عمل ممکن است نرم مانده مربوط به m_j را افزایش دهد. برای حل این مشکل الگوریتم جدیدی در [۲۸] ارائه شده است. در این الگوریتم ابتدا یک حدس اولیه انتخاب می‌شود و در هر تکرار یک درایه درج شده در بردار جواب x اضافه می‌گردد. بردار جهت d به‌وسیله چکاندن برخی از درایه‌های r به‌دست می‌آید. برای حفظ تُنکی بردار جواب x ، بردار d طوری انتخاب می‌شود که الگوی تُنکی آن همانند x ، بعلاوه یک درایه جدید می‌باشد که آن‌هم بزرگ‌ترین درایه از لحاظ قدرمطلق می‌باشد. در این صورت، طول گام به‌صورت

$$\alpha = \frac{(r, Ad)}{(Ad, Ad)},$$

۳۳۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

می‌باشد. بنابراین، تضمین می‌شود که نرم ماندهٔ جواب جدید بیشتر از نرم ماندهٔ جواب قبلی نباشد. این الگوریتم را می‌توان به صورت زیر جمع‌بندی کرد.

الگوریتم ۹.۹ الگوریتم MR برای حل (۸-۹)

1. Set $M = M_0$
2. For each column $j = 1, \dots, n$, Do
3. $m_j := Me_j$
4. $r_j = e_j - Am_j$
5. For $k = 1, 2, \dots, lfil$, Do
6. Choose d to be r_j with the same pattern as m_j ;
 If $nnz(m_j) < lfil$ then add one entry which is the
 largest remaining entry in absolute value
7. $q := Ad$
8. $\alpha_j := \langle r_j, q \rangle / \langle q, q \rangle$
9. $r_j := r_j - \alpha_j q$
10. $m_j := m_j + \alpha d$
11. EndDo
12. EndDo

متأسفانه روش‌هایی که معکوس تقریبی تُنک یک ماتریس را با کمک یک روش تکراری محاسبه می‌کنند، به‌طور گسترده مورد توجه قرار نگرفت. دلیل اصلی آن این است که اگر یک دستگاه بدحالت باشد، حل آن با یک روش تکراری مبتنی بر زیرفضای کرایلف یک‌بعدی مثل MR ساده نخواهد بود. در عوض، روش‌هایی که معکوس ماتریس را به صورت تجزیه شده محاسبه می‌کنند بسیار مورد توجه محققین قرار گرفته است. در ادامه، یکی از این روش‌ها را معرفی می‌کنیم.

۹-۳-۲ معکوس تقریبی تُنک تجزیه شده

همان‌گونه که پیش‌تر اشاره کردیم، می‌توانیم پیش‌شرط‌ساز M را طوری بیابیم که مستقیماً یک تقریب از A^{-1} باشد؛ به عبارت دیگر، $M \approx A^{-1}$. در این حالت، در عمل نیازی به محاسبهٔ مستقیم AM (یا MA) نیست، زیرا وقتی که یک روش تکراری بر پایهٔ زیرفضای کرایلف مورد استفاده قرار می‌گیرد، فقط ضرب ماتریس در بردار My لازم است. توجه کنید که ضرب ماتریس در بردار را به سادگی می‌توان در کامپیوترهای موازی انجام داد. یکی از ایرادهای روش محاسبهٔ معکوس تقریبی تُنک این است که هزینهٔ

ساخت M معمولاً زیاد است، مگر اینکه محاسبه، روی کامپیوترهای موازی به‌طور مناسب قابل اجرا باشد. یکی از روش‌ها در اینجا این است که معکوس تقریبی $\tilde{A}$ به‌صورت تجزیه‌شده محاسبه شود. فرض کنید

$$A = LDU,$$

تجزیه LDU ماتریس A باشد. در این صورت، داریم

$$A^{-1} = W D^{-1} Z,$$

که در آن $W = U^{-1}$ و $Z = L^{-1}$. حال اگر

$$G = \tilde{W} \tilde{D} \tilde{Z},$$

که در آن $\tilde{Z} \approx Z$ ، $\tilde{W} \approx W$ ، $\tilde{D} \approx D^{-1}$ ، ماتریس‌های $\tilde{W}$ و $\tilde{Z}^T$ پائین‌مثلثی با عناصر قطری یک و $\tilde{D}$ قطری باشد، آنگاه G را می‌توان به‌عنوان یک ماتریس پیش‌شرط‌ساز برای $Ax = b$ به‌کار برد. در اینجا G را معکوس تقریبی $\tilde{A}$ تجزیه‌شده و $\tilde{W}$ و $\tilde{Z}$ را عامل‌های معکوس تقریبی $\tilde{A}$ می‌گویند.

روش‌های مختلفی برای محاسبه عامل‌های معکوس تقریبی $\tilde{A}$ یک ماتریس وجود دارد. از بین روش‌های موجود، الگوریتم FSAI ارائه‌شده توسط کلوتلینا و یرمین [۶۴، ۶۵]، الگوریتم AIB ارائه‌شده توسط سعد [۸۷] و شکل بلوکی‌آن که در [۶۲]، الگوریتم AINV ارائه‌شده توسط بنزی و توما [۱۸، ۱۹] و الگوریتم FAPINV که ابتدا توسط لو ارائه [۷۱-۷۳] و سپس توسط ژانگ و لی تعمیم داده شد [۶۹، ۹۸، ۹۹]، از مهم‌ترین الگوریتم‌ها هستند. الگوریتم‌های دیگری نیز برای محاسبه یک پیش‌شرط‌ساز که به‌طور مستقیم ماتریس M^{-1} را تقریب می‌زنند، وجود دارد ([۸۷] را ببینید). از بین الگوریتم‌هایی که معکوس ماتریس A را به‌طور مستقیم یا تجزیه‌شده تقریب می‌زنند، دو الگوریتم AINV و FAPINV نسبت به سایر پیش‌شرط‌سازها، کاراتر می‌باشند.

ایده اصلی الگوریتم FAPINV توسط لو در چند مقاله پی‌درپی ارائه شد [۷۱-۷۳]. سپس این الگوریتم توسط ژانگ در [۹۹] مورد بررسی بیشتری قرار گرفت. چون در این الگوریتم روند محاسبه عامل‌ها به‌صورت پسرو انجام می‌شود (سطرهای W و ستون‌های Z از انتها به ابتدا محاسبه می‌شوند)، این الگوریتم را BFAPINV می‌نامند. در [۹۸]، ژانگ الگوریتم دیگری ارائه نمود که محاسبه عامل‌های معکوس به‌صورت پیشرو انجام می‌گیرد، و به این دلیل، به الگوریتم FFAPINV معروف شد. در ادامه، الگوریتم FFAPINV را معرفی کرده و خواص آن را بررسی می‌کنیم.

۳-۳-۹ الگوریتم FFAPINV

فرض کنید W و Z عامل‌های معکوس ماتریس $A = (a_{ij})$ باشند، یعنی

$$WAZ = D, \quad (9-9)$$

که در آن $D = \text{diag}(d_1, d_2, \dots, d_n)$ و $Z = (z_1, z_2, \dots, z_n)$ ، $W = (w_1^T, w_2^T, \dots, w_n^T)^T$ به‌طوری‌که w_i ها و z_i ها به‌ترتیب سطرها و ستون‌های W و Z هستند. در این صورت، با توجه به (۹-۹)، داریم

$$w_i Az_j = \begin{cases} d_i, & i = j, \\ 0, & i \neq j. \end{cases} \quad (10-9)$$

با توجه به ساختار ماتریس‌های Z و W می‌بینیم که

$$z_1 = e_1, \quad z_j = e_j - \sum_{i=1}^{j-1} \alpha_{ij} z_i, \quad j = 2, \dots, n, \quad (11-9)$$

$$w_1 = e_1^T, \quad w_j = e_j^T - \sum_{i=1}^{j-1} \beta_{ji} w_i, \quad j = 2, \dots, n, \quad (12-9)$$

که در آن α_{ij} ها و β_{ji} ها اعدادی حقیقی و e_j ، ستون j ام ماتریس همانی است. ابتدا توجه کنید که داریم

$$d_1 = z_1^T A z_1 = e_1^T A e_1 = a_{11}.$$

اکنون، فرض کنید $2 \leq j \leq n$ عددی طبیعی و ثابت باشد. با استفاده از روابط (۱۱-۹) و (۱۲-۹) و برای $k = 1, \dots, j-1$ ، داریم

$$\begin{aligned} 0 &= w_k A z_j \\ &= w_k A e_j - \sum_{i=1}^{j-1} \alpha_{ij} w_k A z_i \\ &= w_k A_{*j} - \alpha_{kj} w_k A z_k \\ &= w_k A_{*j} - \alpha_{kj} d_k, \end{aligned}$$

که در آن A_{*j} ستون j ام ماتریس A است. بنابراین،

$$\alpha_{ij} = \frac{1}{d_i} w_i A_{*j}, \quad i = 1, \dots, j-1.$$

به‌طور مشابه، داریم

$$\beta_{ji} = \frac{1}{d_i} A_{j*} z_i, \quad i = 1, \dots, j-1,$$

که در آن A_{j*} سطر j ام ماتریس A است. با جمع‌بندی روابط بالا الگوریتم زیر برای محاسبهٔ عامل‌های معکوس یک ماتریس نوشته می‌شود.

الگوریتم ۱۰.۹ محاسبهٔ عامل‌های معکوس

1. $z_1 := e_1, w_1 := e_1^T$ and $d_1 := a_{11}$
2. For $j = 2, \dots, n$, Do
3. $z_j := e_j; \quad w_j := e_j^T$
4. For $i = 1, \dots, j-1$, Do
5. $\alpha_{ij} := (w_i A_{*j})/d_i; \quad \beta_{ji} := (A_{j*} z_i)/d_i$
6. $z_j := z_j - \alpha_{ij} z_i; \quad w_j := w_j - \beta_{ji} w_i$
7. EndDo
8. $d_j := w_j A z_j$
9. EndDo

چند نکتهٔ مهم در این جا قابل ذکر است. اگر از سمت چپ، رابطهٔ (۹-۱۱) را در $w_j A$ ضرب کنیم، داریم

$$\begin{aligned} d_j &= w_j A z_j \\ &= w_j A e_j - \sum_{i=1}^{j-1} \alpha_{ij} w_j A z_i \\ &= w_j A_{*j}. \end{aligned}$$

بنابراین، در گام ۸ از الگوریتم ۱۰.۹، می‌توان از $d_j = w_j A_{*j}$ استفاده کرد. در این صورت، از یک ضرب ماتریس در بردار پرهیز می‌شود. برای ماتریس‌های متقارن داریم $W = Z^T$. از این رو، محاسبات نصف خواهد شد. در صورتی که ماتریس A معین مثبت متقارن باشد، آنگاه

$$d_j = z_j^T A z_j > 0.$$

لذا، اگر برای ماتریس‌های معین مثبت متقارن از $d_j = z_j^T A z_j$ در گام ۸ از الگوریتم ۱۰.۹ استفاده شود، الگوریتم بدون شکست انجام خواهد گرفت.

برای محاسبهٔ عامل‌های تقریبی تُنکِ ماتریس A ، کافیت الگوریتم ۱۰.۹ با روشی مناسب برای تُنک‌سازی بردارهای z_j و w_j در گام ۶ الگوریتم همراه شود. برای این کار، از یک شیوهٔ چکاندن در دو قسمت استفاده می‌کنیم. فرض کنید که $\mu > 0$ داده شده باشد. اگر $|\alpha_{ij}| < \mu$ ، $|\beta_{ji}| < \mu$ ، آنگاه از به‌هنگام‌سازی z_j (w_j) در گام ۶ پرهیز می‌شود. درغیراین‌صورت، بعد از به‌هنگام‌سازی z_j (w_j) در گام ۶، درایه‌هایی که قدرمطلق آنها از μ کوچک‌تر باشند با صفر جایگزین می‌شوند. براساس نتایج عددی که در مقالات متعدد ارائه شده‌اند، مقدار $\mu = 0.1$ معمولاً نتایج مناسبی می‌دهد [۹۹، ۹۸، ۱۹]. با توجه به این روش چکاندن، الگوریتم ۱۱.۹ که به الگوریتم FFAPINV معروف است را می‌توان جمع‌بندی کرد.

الگوریتم ۱۱.۹ الگوریتم FFAPINV

1. $z_1 := e_1, w_1 := e_1^T$ and $d_1 := a_{11}$
2. For $j = 2, \dots, n$, Do
3. $z_j := e_j; \quad w_j := e_j^T$
4. For $i = 1, \dots, j - 1$, Do
5. $\alpha_{ij} := (w_i A_{*j})/d_i; \quad \beta_{ji} := (A_{j*} z_i)/d_i$
6. If $|\alpha_{ij}| > \mu$, then $z_j := z_j - \alpha_{ij} z_i$
7. If $|\beta_{ji}| > \mu$, then $w_j := w_j - \beta_{ji} w_i$
8. Drop entries of z_j and w_j whose absolute values are smaller than μ
9. EndDo
10. $d_j := w_j A z_j$
11. EndDo

این الگوریتم برای ماتریس‌های معین مثبت متقارن خوش‌تعریف است، چون همهٔ درایه‌های محوری مثبت هستند، یعنی

$$d_j = z_j^T A z_j > 0.$$

توجه کنید که برای هر j ، داریم $z_j \neq 0$. برای بررسی خواص دیگری از الگوریتم، چند نتیجه از منابع [۵۹، ۵۴] را بیان می‌کنیم.

لم ۱۰.۹ فرض کنید W و Z^T ماتریس‌های پائین‌مثلثی با درایه‌های قطری یک باشند به‌طوری‌که

$$WAZ = D = \text{diag}(d_1, d_2, \dots, d_n).$$

در این صورت،

$$d_j = z_j^T A z_j = w_j A w_j^T. \quad (۱۳-۹)$$

برهان. از رابطه $WAZ = D$ داریم

$$Z^T A Z = Z^T W^{-1} D.$$

به‌وضوح ماتریس $Z^T W^{-1} D$ پائین‌مثلثی است و

$$\text{diag}(Z^T W^{-1} D) = D.$$

در نتیجه،

$$d_j = z_j^T A z_j, \quad j = 1, 2, \dots, n.$$

□ به‌طور مشابه، رابطه دوم ثابت می‌شود.

توجه کنید که در این لم ماتریس A دارای هیچ خاصیت اضافی مثل معین مثبت بودن یا تقارن ندارد. اکنون، با توجه به این لم، قضیه بعدی را می‌توان نوشت.

قضیه ۱۰.۹ فرض کنید A یک ماتریس معین مثبت (نه لزوماً متقارن) و همه مفروضات لم ۱۰.۹ برقرار باشند. در این صورت،

$$d_j > 0, \quad j = 1, 2, \dots, n.$$

برهان. با توجه به لم ۱۰.۹، داریم

$$d_j = z_j^T A z_j, \quad j = 1, 2, \dots, n.$$

□ بنابراین، با توجه به اینکه A معین مثبت است، نتیجه لازم به‌دست می‌آید.

با توجه به این قضیه، الگوریتم ۱۲.۹ برای ماتریس‌های معین مثبت جمع‌بندی می‌شود. برای پرهیز از ضرب ماتریس در بردار در محاسبه $d_j := z_j^T A z_j$ ، ابتدا d_j را با استفاده از $d_j := A_j * z_j$ محاسبه می‌کنیم. اگر $d_j = 0$ ، آنگاه از $d_j = z_j^T A z_j$ استفاده می‌کنیم. به‌وضوح این الگوریتم برای ماتریس‌های معین مثبت بدون شکست انجام می‌شود. در هر صورت، ممکن است درایه‌های بسیار کوچک محوری تولید شود. در عمل اگر قدرمطلق d_j از 10^{-15} کوچک‌تر باشد، همان‌گونه که در [۲۰، ۵۲] پیشنهاد شده است، مقدار $\text{sgn}(d_j) \times 10^{-1}$ به‌جای d_j جایگزین می‌شود.

الگوریتم ۱۲.۹ FFAPINV-NSPD

1. $z_1 := e_1, w_1 := e_1^T$ and $d_1 := a_{11}$
2. For $j = 2, \dots, n$, Do
3. $z_j := e_j; \quad w_j := e_j^T$
4. For $i = 1, \dots, j-1$, Do
5. $\alpha_{ij} := (w_i A_{*j})/d_i; \quad \beta_{ji} := (A_{j*} z_i)/d_i$
6. If $|\alpha_{ij}| > \mu$, then $z_j := z_j - \alpha_{ij} z_i$
7. If $|\beta_{ji}| > \mu$, then $w_j := w_j - \beta_{ji} w_i$
8. Drop entries of z_j and w_j whose absolute values are smaller than μ
9. EndDo
10. $d_j := A_{jj} z_j$
11. If $d_j = 0$, then $d_j := z_j^T A z_j$
12. EndDo

۴-۳-۹ محاسبه تجزیه ILU بر پایه الگوریتم FFAPINV

ابتدا قضیه بعدی که در [۵۴] ارائه شده است را بیان می‌کنیم. این قضیه ارتباط بین الگوریتم FFAPINV و تجزیه ILU را نشان می‌دهد.

قضیه ۲.۹ فرض کنید W و Z عامل‌های معکوس ماتریس A باشد. اگر $A = LDU$ تجزیه LDU ماتریس A باشد، آنگاه $L_{ji} = \beta_{ji}$ و $U_{ij} = \alpha_{ij}$ که در آن β_{ji} و α_{ij} در گام ۵ الگوریتم ۱۰.۹ محاسبه می‌شوند.

برهان. فرض کنید $WAZ = D$. در این صورت، $A = W^{-1} D Z^{-1}$ تجزیه LDU ماتریس A خواهد بود. با توجه به یکتا بودن تجزیه LDU، داریم $W^{-1} = L$ و $Z^{-1} = U$. از $Z^{-1} = U$ ، رابطه $ZU = I$ نتیجه می‌شود که در آن I نشانگر ماتریس همانی است. با برابر قرار دادن ستون‌های Z و U ، داریم $ZU = I$

$$z_j = e_j - \sum_{i=1}^{j-1} U_{ij} z_i,$$

که در آن e_j ستون j ام ماتریس I است. از اینکه مجموعه $\{z_1, z_2, \dots, z_{j-1}, e_j\}$ مستقل خطی هستند، با استفاده از رابطه بالا و رابطه (۹-۱۱)، خواهیم داشت $\alpha_{ij} = U_{ij}$. به‌طور مشابه، می‌توان دید که $L_{ji} = \beta_{ji}$. $\square$

نتیجه ۱.۹ فرض کنید $A = (a_{ij})$. اگر برای $i < c_j$, $a_{ij} = 0$, آنگاه برای $i < c_j$, $\alpha_{ij} = 0$ که α_{ij} در گام ۵ از الگوریتم ۱۰.۹ محاسبه می‌شود. به‌طور مشابه، اگر برای $i < r_j$, $a_{ji} = 0$, آنگاه برای $i < r_j$, $\beta_{ji} = 0$ که β_{ji} در گام ۵ از الگوریتم ۱۰.۹ محاسبه می‌شود.

بر همان. همان‌گونه که می‌دانیم نیمه بالایی پهنای نوار ستون A و ستون U یکی هستند (مرجع [۷۷]، صفحه ۱۵۷). بنابراین، با توجه به قضیه ۲.۹ داریم $\alpha_{ij} = U_{ij} = 0$, $i < c_j$. به‌طور مشابه، قسمت دوم این نتیجه را می‌توان ثابت کرد. □

این نتیجه به ما کمک می‌کند تا از محاسبه بسیاری از α_{ij} ها و β_{ij} ها که برابر با صفر هستند، جلوگیری کنیم. اکنون، با جمع‌بندی مطالب به‌دست‌آمده می‌توانیم الگوریتم ۱۳.۹ با نام FAPINV-ILU برای محاسبه تجزیه ناقص LDU یک ماتریس را ارائه کنیم.

الگوریتم ۱۳.۹ FFAPINV-ILU

1. Set $L = U = I_{n \times n}$, $z_1 := e_1$, $w_1 := e_1^T$ and $d_1 := a_{11}$
2. For $j = 2, \dots, n$, Do
3. $z_j := e_j$; $w_j := e_j^T$
4. For $i = c_j, \dots, j-1$, Do
5. $U_{ij} := \frac{1}{d_i} w_i A_{*j}$
6. If $|U_{ij}| > \mu$, then $z_j := z_j - U_{ij} z_i$
7. Drop entries of z_j whose absolute values are smaller than μ
8. EndDo
9. For $i = r_j, \dots, j-1$, Do
10. $L_{ji} := \frac{1}{d_i} A_{j*} z_i$
11. If $|L_{ji}| > \mu$, then $w_j := w_j - L_{ji} w_i$
12. Drop entries of w_j whose absolute values are smaller than μ
13. EndDo
14. $d_j := w_j A z_j$
15. EndDo
16. Return L, U and $D = \text{diag}(d_1, d_2, \dots, d_n)$ ($A \approx LDU$)

در ادامه، موضوع شکست یا عدم شکست این الگوریتم برای M-ماتریس‌ها را بررسی می‌کنیم. همان‌گونه که دیدیم اگر A معین مثبت متقارن یا معین مثبت (نه لزوماً متقارن) و از $d_j = z_j^T A z_j$ استفاده کنیم، آنگاه الگوریتم شکست نخواهد خورد. حال فرض کنید که A یک M-ماتریس باشد. در این صورت، در گام ۱۴ از این الگوریتم از $d_j = w_j A_{*j}$ استفاده کرده و نشان می‌دهیم که الگوریتم بدون

شکست انجام می‌گیرد. با توجه به این مفروضات، قضیه بعدی را ثابت می‌کنیم که مشابه قضیه ۲.۱ در [۶۹] برای الگوریتم BFAPINV است.

قضیه ۳.۹ فرض کنید که A یک M -ماتریس و درگام ۱۴ از الگوریتم FAPINV-ILU از $d_j = w_j A_{*j}$ استفاده شود. به علاوه، فرض کنید w_j ها (z_j ها) سطرها (ستون‌های) عامل معکوس W (Z) باشند که توسط الگوریتم FAPINV-ILU بدون چکاندن ($\mu = 0$) تولید می‌شوند. در این صورت،

$$w_j \geq 0, \quad z_j \geq 0, \quad d_j > 0, \quad j = 1, 2, \dots, n. \quad (14-9)$$

برهان. نشان می‌دهیم به ازای $j = 1, 2, \dots, n$ داریم $d_j > 0$. نخست یادآوری می‌کنیم که اگر A یک M -ماتریس باشد، آنگاه دترمینان تمام زیرماتریس‌های اصلی پیشروی A مثبت خواهد بود [۱۸]. فرض کنید Δ_j دترمینان j امین زیرماتریس اصلی پیشروی A باشد. در این صورت، با استفاده از $AZ = W^{-1}D$ داریم

$$d_j = \frac{\Delta_j}{\Delta_{j-1}} > 0, \quad j = 1, 2, \dots, n,$$

فرض کنید $w_j = (w_{j1}, \dots, w_{j,j-1}, 1, 0, \dots, 0)^T$ و $z_j = (z_{1j}, \dots, z_{j-1,j}, 1, 0, \dots, 0)^T$ به استقرا نشان می‌دهیم که $w_j \geq 0, z_j \geq 0$ ، برای $j = 1, 2, \dots, n$. برای $j = 1$ داریم $w_1 = e_1^T \geq 0$ و $z_1 = e_1 \geq 0$. فرض کنید نامساوی‌های موجود در (۱۴-۹) برای هر $k \leq j-1$ برقرار باشد، یعنی

$$w_k \geq 0, \quad z_k \geq 0, \quad k = 1, 2, \dots, j.$$

برای $j = 1, 2, \dots, n$ داریم

$$U_{i,j+1} = \frac{1}{d_i} w_i A_{*j+1} = \frac{1}{d_i} (a_{i,j+1} + \sum_{k=1}^{j-1} w_{ik} a_{k,j+1}) \leq 0.$$

این رابطه از آنجا نتیجه می‌شود که A یک M -ماتریس است و به ازای هر j هر $1 \leq \ell \leq j$ ، داریم $a_{\ell,j+1} \leq 0$. به طور مشابه، می‌توان نشان داد که $a_{\ell,j+1} \leq 0$ و $i = 1, 2, \dots, j, L_{j+1,i} \leq 0$ و $w_{j+1} \geq 0$. □

قضیه ۴.۹ فرض کنید که A یک M -ماتریس و درگام ۱۴ از الگوریتم FAPINV-ILU از $d_j = w_j A_{*j}$ استفاده شود. به علاوه، فرض کنید $\tilde{W}$ و $\tilde{Z}$ عامل‌های معکوس A باشند که توسط الگوریتم FFAPINV-ILU محاسبه می‌شوند، یعنی $\tilde{W}A\tilde{Z} \approx \tilde{D}$. همچنین، فرض کنید W و Z عامل‌های

معکوس A باشند که توسط الگوریتم FFAPINV-ILU بدون چکاندن تولید می‌شوند، یعنی $WAZ = D$ در این صورت،

$$W \geq \tilde{W} \geq 0, \quad Z \geq \tilde{Z} \geq 0, \\ \tilde{d}_j \geq d_j > 0, \quad j = 1, 2, \dots, n,$$

که در آن $\tilde{D} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_n)$

برهان. اثبات با استقرا انجام می‌شود. فرض کنید

$$\tilde{W} = (\tilde{w}_1^T, \dots, \tilde{w}_n^T)^T, \\ \tilde{Z} = (\tilde{z}_1, \dots, \tilde{z}_n), \\ \tilde{D} = \text{diag}(\tilde{d}_1, \dots, \tilde{d}_n).$$

برای $j = 1$ داریم

$$w_1 = \tilde{w}_1 = e_1^T, \quad z_1 = \tilde{z}_1 = e_1, \quad d_1 = \tilde{d}_1 = a_{11}.$$

اکنون، فرض کنید

$$w_k \geq \tilde{w}_k \geq 0, \quad z_k \geq \tilde{z}_k \geq 0, \quad \tilde{d}_k \geq d_k \geq 0, \quad k = 1, 2, \dots, j.$$

با توجه به مفروضات استقرا برای j ، $i = 1, 2, \dots, j$ داریم

$$U_{i,j+1} = \frac{1}{d_i} w_i A_{*,j+1} = \frac{1}{d_i} (a_{i,j+1} + \sum_{k=1}^{i-1} w_{ik} a_{k,j+1}) \\ \leq \frac{1}{\tilde{d}_i} (a_{i,j+1} + \sum_{k=1}^{i-1} \tilde{w}_{ik} a_{k,j+1}) = \frac{1}{\tilde{d}_i} \tilde{w}_i A_{*,j+1} \\ = \tilde{U}_{i,j+1} \leq 0,$$

که در آن $\tilde{U}$ ماتریس بالامثلثی با درایه‌های قطری یک است که در تجزیه LDU ماتریس A ظاهر شده و توسط الگوریتم FFAPINV-ILU محاسبه می‌شود و

$$\tilde{w}_i = (\tilde{w}_{i1}, \dots, \tilde{w}_{i,i-1}, 1, 0, \dots, 0).$$

۳۴۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

بنابراین،

$$z_{j+1} = e_{j+1} - \sum_{i=1}^j U_{i,j+1} z_i \geq e_{j+1} - \sum_{i=1}^j \tilde{U}_{i,j+1} \tilde{z}_i = \hat{z}_{j+1}.$$

حال توجه کنید که $\tilde{z}_{j+1}$ از $\hat{z}_{j+1}$ با چکاندن بعضی از درایه‌های $\tilde{U}_{i,j+1}$ ، $i = 1, \dots, j$ ، به دست می‌آید. بنابراین، $z_{j+1} \geq \hat{z}_{j+1} \geq \tilde{z}_{j+1} \geq 0$. به طور مشابه، می‌توان دید که

$$w_k \geq \tilde{w}_k \geq 0.$$

سرانجام، داریم

$$\begin{aligned} d_{j+1} &= w_{j+1} A_{*,j+1} = a_{j+1,j+1} + \sum_{i=1}^j w_{j+1,i} a_{i,j+1} \\ &\leq a_{j+1,j+1} + \sum_{i=1}^j \tilde{w}_{j+1,i} a_{i,j+1} = \tilde{w}_{j+1} A_{*,j+1} = \tilde{d}_{j+1}, \end{aligned}$$

□ که این رابطه اثبات را کامل می‌کند.

این قضیه نشان می‌دهد اگر A یک M -ماتریس باشد و در گام ۱۴ از الگوریتم FAPINV-ILU از $d_j = w_j A_{*,j}$ استفاده شود آنگاه الگوریتم در حساب دقیق شکست نخواهد خورد. نکته قابل ذکر دیگر این است که از قضیه ۴.۹ می‌توان نتیجه گرفت که الگوریتم FFAPINV نیز برای M -ماتریس‌ها خوش تعریف است. همچنین، با استفاده از قضیه ۴.۹، نتیجه بعدی را می‌توان بیان کرد.

نتیجه ۲.۹ فرض کنید A یک M -ماتریس باشد. در این صورت، تجزیه ناقص LDU ماتریس A ، یعنی $A \approx \tilde{L} \tilde{D} \tilde{U}$ دارای خاصیت زیر است

$$L \leq \tilde{L}, \quad U \leq \tilde{U},$$

$$d_j > 0, \quad j = 1, 2, \dots, n,$$

که در آن $A = LDU$ تجزیه LDU ماتریس A است. به علاوه،

$$\tilde{L}_{ij} \leq 0, \quad \tilde{U}_{ij} \leq 0, \quad i \neq j.$$

□ برهان. این نتیجه، چیزی است که در روند اثبات قضیه قبل به دست آمده است.

در [۵۹] نشان داده شده است که اگر ماتریس A یک H -ماتریس باشد، آنگاه الگوریتم FFAPINV (در نتیجه، الگوریتم FAPINV-ILU) بدون شکست انجام می‌شود.

مثال ۷.۹ برای بررسی توانایی الگوریتم FFAPINV-NSPD، این الگوریتم را برای چند دستگاه معین مثبت نامتقارن به کار می‌بریم. روش تکراری مورد استفاده الگوریتم $GMRES(m)$ و از پیش‌شرط‌سازی از چپ استفاده شده است. حدس اولیه را بردار صفر گرفته و بردار b طوری انتخاب شده است که جواب واقعی دستگاه برابر با $x = (1, 1, \dots, 1)^T$ باشد؛ یعنی بردار سمت راست به صورت

$$b = A(1, 1, \dots, 1)^T$$

است. از شرط توقف $\|r_i\|_2 < 10^{-10} \|r_0\|_2$ استفاده شده است که در آن $r_i = b - Ax_i$. ماتریس‌هایی که از گسسته‌سازی پنج‌نقطه‌ای معادله دیفرانسیل

$$-(bu_x)_x - (cu_x)_x + du_x + (du)_x + eu_y + (eu)_y + fu = g, \quad (۱۵-۹)$$

در مربع واحد $(0, 1) \times (0, 1)$ ، با استفاده از

$$b(x, y) = e^{-xy}, \quad c(x, y) = e^{xy}, \quad d(x, y) = \beta(x + y),$$

$$e(x, y) = \gamma(x + y), \quad f(x, y) = 1/(1 + x + y),$$

و پارامترهای $\beta = 20$ و $\gamma = 0$ ، به دست آمده‌اند، استفاده می‌کنیم. تمام ماتریس‌هایی که به این طریق به دست می‌آیند معین مثبت نامتقارن هستند [۸۸]. در منزلگاه بازار ماتریس‌ها [۷۵]، برنامه‌ای به زبان فترن به نام MATPDE وجود دارد که این ماتریس‌ها را تولید می‌کند و برای تولید ماتریس‌های مورد استفاده از این برنامه استفاده کرده‌ایم. فرض کنید m تعداد گره‌های درونی روی هر ضلع این مربع باشد که در شبکه‌بندی آن استفاده می‌شوند. با استفاده از MATPDE و $110, 100, 90, 80, 70$ $m =$ ماتریس‌های PDE4900، PDE6400، PDE8100، PDE10000 و PDE12100 را تولید می‌کنیم که به ترتیب با ابعاد ۴۹۰۰، ۶۴۰۰، ۸۱۰۰، ۱۰۰۰۰ و ۱۲۱۰۰ هستند. برای حفظ تئگی ماتریس پیش‌شرط‌ساز از پارامترهای $\mu = 0.1$ و $\mu = 0.2$ استفاده می‌کنیم. نتایج عددی در جدول ۹-۶ داده شده‌اند. در این جدول، برای هر ماتریس تعداد درایه‌های ناصفر ماتریس (nnz) ، و تعداد تکرارها $(iters)$ برای همگرایی روش $GMRES(5)$ بدون پیش‌شرط‌سازی داده شده‌اند. در شش ستون آخر از جدول ۹-۶، نتایج عددی مربوط به $GMRES(5)$ با پیش‌شرط‌سازی از چپ با استفاده از پیش‌شرط‌ساز

جدول ۹-۶ نتایج عددی برای پیش‌شرط‌ساز FFAPINV-NSPD در مثال ۷.۹.

		بدون پیش شرط سازی		دستگاه پیش شرط سازی شده					
ماتریس	nnz	$iters$	$time$	μ	ρ	P-T	I-T	T-T	$iters$
PDE4900	۲۴۲۲۰	۱۷۳	۳/۳۳	۰/۱	۲/۲۹	۰/۳۴	۰/۸۳	۱/۱۷	۳۵
				۰/۲	۰/۸۸	۰/۳۱	۰/۹۳	۱/۲۴	۴۴
PDE6400	۳۱۶۸۰	۱۹۲	۴/۸۶	۰/۱	۲/۱۹	۰/۵۶	۱/۳۳	۱/۸۹	۴۳
				۰/۲	۰/۸۸	۰/۵۳	۱/۳۴	۱/۸۷	۴۸
PDE8100	۴۰۱۴۰	۲۲۶	۷/۱۹	۰/۱	۲/۰۹	۰/۸۹	۲/۲۲	۳/۱۱	۵۷
				۰/۲	۰/۸۷	۰/۸۶	۲/۰۱	۲/۸۷	۵۳
PDE10000	۴۹۶۰۰	۲۵۸	۱۰/۰۹	۰/۱	۲/۰۰	۱/۳۲	۲/۴۷	۳/۷۹	۵۱
				۰/۲	۰/۸۶	۱/۳۱	۳/۷۰	۵/۰۱	۷۹
PDE12100	۶۰۰۶۰	۳۱۹	۱۵/۶۴	۰/۱	۱/۹۲	۱/۹۳	۳/۴۵	۵/۳۸	۵۹
				۰/۲	۰/۸۴	۱/۹۰	۵/۱۸	۷/۰۸	۹۷

محاسبه شده توسط الگوریتم FFAPINV-NSPD داده شده‌اند. در اینجا P-T، I-T، T-T و *iters* به ترتیب برای نمایش زمان ساخت برای ماتریس پیش شرط ساز، زمان لازم برای تعداد تکرارها برای همگرایی، زمان کل و تعداد تکرارها برای همگرایی را نمایش می‌دهند. لازم به ذکر است که زمان کل برابر است با مجموع I-T و P-T. مقدار ρ نیز همانند قبل است. نتایج عددی ارائه شده در جدول ۹-۶، نشان می‌دهند که FFAPINV-NSPD یک پیش شرط ساز کارا برای دستگاه‌هایی است که ماتریس ضرایب آنها معین مثبت نامتقارن هستند. همان گونه که می‌بینیم، این پیش شرط ساز تعداد تکرارها را تقریباً به میزان پنج برابر برای $\mu = 0.1$ و چهار برابر برای $\mu = 0.2$ ، کاهش می‌دهد.

مثال ۸.۹ در این مثال چندین ماتریس از بازار ماتریس‌ها [۷۵] یا منزلگاه مشخص شده در [۳۳] انتخاب شده‌اند. همانند قبل بردار سمت راست به صورت $b = A(1, 1, \dots, 1)^T$ فرض شده است. نتایج عددی مربوط به الگوریتم $GMRES(m)$ برای حل دستگاه اولیه (بدون پیش شرط‌سازی شده) در جدول ۹-۷ نشان داده شده‌اند. فقط برای ماتریس‌های SHERMAN2 و SHERMAN3، فرض کرده ایم $m = 50$ و برای سایر ماتریس‌ها $m = 10$. در این جدول نماد FEM-3D برای FEM_3D_THERMAL1 به کار برده شده است و علامت † به این معنا است که روش در ۱۰۰۰ تکرار همگرا نشده است. حدس اولیه، برداری است که مؤلفه‌های آن به طور تصادفی با توزیع یکنواخت در $(0, 1)$ تولید شده‌اند. سایر نمادها و مفروضات همانند قبل هستند.

برای بررسی کارایی پیش‌شرط‌ساز FFAPINV-ILU ابتدا نتایج عددی مربوط پیش‌شرط‌ساز ILUT را ارائه کرده و نتایج مربوط به پیش‌شرط‌ساز FFAPINV-ILU را با آن مقایسه می‌کنیم. در اینجا از نسخهٔ

جدول ۷-۹ نتایج عددی برای حل دستگاه‌های مثال ۸.۹ بدون پیش‌شرط‌سازی.

ماتریس	n	nnz	$time$	$iters$
ADD20	۲۳۹۵	۱۳۱۵۱	۴/۲۰	۳۵۵
ADD32	۴۹۶۰	۱۹۸۴۸	۰/۴۲	۱۶
RAEFSKY1	۳۲۴۲	۲۹۳۴۰۹	—	†
RAEFSKY2	۳۲۴۲	۲۹۳۵۵۱	۳۳/۱۶	۷۸۲
SHERMAN1	۱۰۰۰	۳۷۵۰	۴/۳۰	۹۴۴
SHERMAN2	۱۰۸۰	۲۳۰۹۴	—	†
SHERMAN3	۵۰۰۵	۲۰۰۳۳	—	†
SHERMAN4	۱۱۰۴	۳۷۸۶	۰/۴۸	۸۸
SHERMAN5	۳۳۱۲	۲۰۷۹۳	—	†
MEMPLUS	۱۷۷۵۸	۹۹۱۴۷	—	†
FEM_3D	۱۷۸۸۰	۴۳۰۷۴۰	۷/۰۳	۶۱
KAUFHOLD	۸۷۶۵	۴۲۴۷۱	—	†
CHIPCOOL0	۲۰۰۸۲	۲۸۱۱۵۰	—	†
CHIPCOLL1	۲۰۰۸۲	۲۸۱۱۵۰	—	†
EPB1	۱۴۷۳۴	۹۵۰۵۳	۳۱/۰۰	۴۳۱
EPB2	۲۵۲۲۸	۱۷۵۰۲۷	۲۵/۴۷	۲۰۵
VENKAT01	۶۲۴۲۴	۱۷۱۷۷۹۲	—	†
ASIC_100KS	۹۹۱۹۰	۵۷۸۸۹۰	—	†

$ILUT(p, \tau)$ استفاده می‌کنیم. نتایج عددی مربوط پیش‌شرط‌ساز $ILUT(p, \tau)$ به همراه $GMRES(m)$ در جدول ۸-۹ داده شده‌اند. برای همه مثال‌ها پیش‌شرط‌سازی از راست انجام شده است. سایر نمادها و مفروضات همانند قبل هستند.

جدول ۹-۹ نتایج مربوط به پیش‌شرط‌ساز $FFAPINV-ILU$ به همراه $GMRES(m)$ را گزارش می‌کند. تمام مفروضات همانند قبل هستند. پارامتر μ طوری انتخاب شده است که چگالی (ρ) پیش‌شرط‌ساز $FFAPINV-ILU$ تقریباً با پیش‌شرط‌ساز $ILUT(p, \mu)$ برابر باشد. برنامه‌ای که برای الگوریتم $FFAPINV-ILU$ نوشته شده بهینه نیست، به این معنا که بهترین زمان برای تولید پیش‌شرط‌ساز را نمی‌دهد. با این وجود همان‌گونه که جدول ۹-۹ نشان می‌دهد، پیش‌شرط‌ساز $FFAPINV-ILU$ برای همه ماتریس‌ها، نه تنها به‌طور قابل توجه تعداد تکرارها، بلکه زمان کل را نیز کاهش می‌دهد. در حالت کلی، از لحاظ زمانی پیش‌شرط‌ساز $FFAPINV-ILU$ نمی‌تواند با پیش‌شرط‌ساز $ILUT$ رقابت کند، حتی اگر الگوریتم به‌صورت بهینه نوشته شود. نتایج جدول‌های ۸-۹ و ۹-۹ نشان می‌دهند که

جدول ۸-۹ نتایج عددی روش $GMRES(m)$ با پیش‌شرط‌ساز ILUT برای مثال ۸.۹.

ماتریس	p, τ	ρ	P-T	I-T	T-T	iters
ADD20	۱۰, ۰/۰۰۱	۰/۸۲	۰/۰۲	۰/۰۳	۰/۰۵	۱
ADD32	۱۰, ۰/۰۰۹	۰/۷۳	۰/۰۰	۰/۰۶	۰/۰۶	۲
RAEFSKY1	۲۷, ۰/۰۰۰۱	۰/۵۸	۰/۵۲	۰/۷۸	۱/۳۰	۱۴
RAEFSKY2	۴۲, ۰/۰۰۰۱	۰/۹۰	۰/۶۹	۰/۶۱	۱/۳۰	۹
SHERMAN1	۲۰, ۰/۰۰۵	۱/۰۱	۰/۰۰	۰/۰۶	۰/۰۶	۴
SHERMAN2	۶۰, ۰/۰۰۰۰۱	۰/۹۹	۰/۰۲	۰/۱۳	۰/۱۴	۲
SHERMAN3	۱۰, ۰/۱	۰/۹۴	۰/۰۰	۱/۰۸	۱/۰۸	۵
SHERMAN4	۵, ۰/۱۵	۰/۹۸	۰/۰۰	۰/۰۵	۰/۰۵	۱۱
SHERMAN5	۱۰, ۰/۱	۰/۶۳	۰/۰۲	۰/۴۲	۰/۴۴	۱۹
MEMPLUS	۱۵, ۰/۰۰۱	۱/۳۸	۰/۰۳	۰/۷۰	۰/۷۳	۷
FEM_3D	۵, ۰/۱	۰/۳۲	۰/۰۵	۱۸/۸۴	۱۸/۸۹	۱۴۷
KAUFHOLD	۵, ۰/۱	۰/۲۶	۰/۰۰	۰/۰۵	۰/۰۵	۱
CHIPCOOL0	۵, ۰/۱	۰/۵۶	۰/۰۵	۳/۵۳	۳/۵۸	۲۷
CHIPCOLL1	۵, ۰/۱	۰/۵۵	۰/۰۳	۳/۸۰	۳/۸۳	۲۹
EPB1	۵, ۰/۱	۰/۹۷	۰/۰۲	۱/۷۰	۱/۷۲	۲۰
EPB2	۵, ۰/۱	۰/۸۱	۰/۰۲	۱/۰۳	۱/۰۵	۷
VENKAT01	۹, ۰/۰۰۷	۰/۵۹	۰/۳۸	۲/۸۰	۳/۱۷	۵
ASIC_100KS	۵, ۰/۱	۰/۸۴	۰/۰۹	۰/۵۹	۰/۶۹	۱

هر دو پیش‌شرط‌ساز قوی و کارا هستند. به‌طور کلی، الگوریتم $ILUT(p, \tau)$ نتایج بهتری نسبت به FFAPIN-ILU را می‌دهد، اگرچه مثال‌های نقض نیز وجود دارد.

یادآوری می‌کنیم که، اگر چه نتایج الگوریتم ILUT از الگوریتم FFAPINV-ILU بهتر است، اما الگوریتم FFAPINV-ILU مزایای مهمی بر الگوریتم ILUT دارد. اولاً الگوریتم FFAPINV-ILU برای ماتریس‌های معین مثبت متقارن بدون شکست انجام می‌گیرد، اما الگوریتم ILUT ممکن است شکست بخورد. ثانیاً الگوریتم FFAPINV-ILU برای کامپیوترهای موازی نسبت به ILUT مناسب‌تر است.

تمرینات

۱.۹ فرض کنید $A, M \in \mathbb{C}^{n \times n}$ به‌طوری‌که M نامنفرد است. نشان دهید که ماتریس‌های AM^{-1} و $M^{-1}A$ متشابه هستند.

جدول ۹-۹ نتایج عددی روش $GMRES(m)$ با پیش‌شرط‌ساز FAPINV-ILU برای مثال ۸.۹.

ماتریس	τ	ρ	P-T	I-T	T-T	iters
ADD20	۰/۰۰۱	۰/۹۵	۰/۳۳	۰/۰۳	۰/۳۶	۲
ADD32	۰/۰۵	۰/۷۳	۱/۰۶	۰/۰۵	۱/۱۱	۲
RAEFSKY1	۰/۰۱	۰/۵۷	۱/۵۲	۰/۸۰	۲/۳۱	۱۳
RAEFSKY2	۰/۰۱	۰/۹۱	۳/۰۰	۰/۵۰	۳/۵۰	۷
SHERMAN1	۰/۰۵	۱/۰۰	۰/۰۳	۰/۰۶	۰/۰۹	۱۳
SHERMAN2	۰/۰۰۰۱	۱/۱۲	۰/۱۱	۴/۰۲	۴/۱۳	۶۱
SHERMAN3	۰/۰۵	۱/۰۵	۰/۳۹	۱۵/۳۹	۱۵/۷۸	۷۴
SHERMAN4	۰/۱	۰/۹۳	۰/۰۳	۰/۰۹	۰/۱۳	۱۸
SHERMAN5	۰/۰۲	۰/۹۴	۰/۳۴	۰/۴۴	۰/۷۸	۲۴
MEMPLUS	۰/۰۰۵	۱/۴۰	۲۰/۴۱	۰/۶۶	۲۱/۰۶	۶
FEM_3D	۰/۱	۰/۲۸	۴/۸۹	۰/۵۶	۵/۴۵	۴
KAUFHOLD	۰/۱	۰/۲۵	۰/۸۴	۰/۰۵	۰/۸۹	۱
CHIPCOOL0	۰/۱	۰/۴۸	۵/۷۸	۶/۲۸	۱۲/۰۶	۴۷
CHIPCOLL1	۰/۰۸۵	۰/۵۵	۵/۸۳	۴/۸۶	۱۰/۶۹	۳۶
EPB1	۰/۰۷	۰/۹۷	۲/۵۶	۴/۰۳	۶/۵۹	۴۶
EPB2	۰/۰۶	۰/۷۹	۷/۷۷	۲/۲۷	۱۰/۰۳	۱۵
VENKAT01	۰/۱	۰/۵۷	۵۴/۷۰	۳/۹۵	۵۸/۶۶	۷
ASIC_100KS	۰/۱	۰/۸۰	۱۱۴/۹۲	۰/۶۱	۱۱۵/۵۳	۱

۲.۹ فرض کنید ماتریس‌های $M, A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن باشند. نشان دهید که ماتریس $M^{-1}A$ نسبت به ضرب داخلی $(x, y)_A = (Ax, y) = (x, Ay)$ خودالحاق است. سپس با استفاده از آن، الگوریتم گرادیان مزدوج برای حل دستگاه معادلات خطی $M^{-1}Ax = M^{-1}b$ را بنویسید.

۳.۹ الگوریتم $GMRES(m)$ برای حل دستگاه پیش‌شرط سازی شده $L^{-1}AU^{-1}u = L^{-1}b$ بنویسید که در آن $Ux = u$.

۴.۹ فرض کنید $A = M - N$ یک شکافت از A باشد و $T = M^{-1}N$. نشان دهید اگر $\rho(T) < ۱$ ، آنگاه

$$A^{-1} = \left(\sum_{k=0}^{\infty} T^k \right) M^{-1}.$$

قرار دهید

$$P = \left(\sum_{k=0}^p T^k \right) M^{-1}.$$

که در آن p یک عدد کوچک (مثلاً ۱ یا ۲). حال می‌توان P را به‌عنوان یک پیش‌شرط‌ساز برای حل دستگاه $Ax = b$ به‌کار برد. الگوریتم $\text{GMRES}(m)$ با پیش‌شرط‌ساز چپ P را در نظر بگیرید. در گام‌هایی که پیش‌شرط‌ساز P ظاهر می‌شوند، نحوه به‌کارگیری P را مشخص کنید.

۵.۹ مثال ۱۹.۲ را در نظر بگیرید. توزیع مقادیر ویژه ماتریس ضرایب هریک از دستگاه‌های $P_i Ax = P_i b$ ، $i = 0, 1, \dots, 5$ را به‌ازای $n = 10^5, 25$ نمایش دهید. کدامیک از ماتریس‌ها مقادیر ویژه‌شان بهتر خوشه شده است. به ازای $n = 10000$ و $n = 22500$ هر یک از این دستگاه‌ها را با روش $\text{GMRES}(10)$ حل نمایید و تکرارها را با هم مقایسه کنید. کدام پیش‌شرط‌ساز مناسب‌تر است؟

۶.۹ معادله انتقال حرارت سه‌بعدی

$$\begin{cases} -(u_{xx} + u_{yy} + u_{zz}) + q(u_x + u_y + u_z) = f(x, y, z), & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega, \end{cases} \quad (۱۶-۹)$$

را در نظر بگیرید به‌طوری‌که $\Omega = [0, 1] \times [0, 1] \times [0, 1]$ و q یک عدد حقیقی مثبت است. بازه $[0, 1]$ را در هر یک از جهت‌های x, y و z به $m + 1$ قسمت مساوی تقسیم کرده و قرار می‌دهیم $h = 1/(m + 1)$. سپس از طرح تفاضل مرکزی برای مشتقات اول و دوم استفاده کرده و معادله دیفرانسیل (۱۶-۹) را گسسته‌سازی می‌کنیم. در این صورت، دستگاهی مثل $Ax = b$ تولید خواهد شد که ماتریس ضرایب آن به‌صورت

$$A = T_x \otimes I \otimes I + I \otimes T_y \otimes I + I \otimes I \otimes T_z, \quad (۱۷-۹)$$

است که در آن

$$T_x = \text{tridiag}_n(t_2, t_1, t_3), \quad T_y = \text{tridiag}_n(t_2, 0, t_3), \quad T_z = \text{tridiag}_n(t_2, 0, t_3),$$

به‌طوری‌که $t_2 = -1 + r$ و $t_3 = -1 - r$ ، $t_1 = 6$ (مراجع [۷۰، ۱۱]) $r = (qh)/2$ که در آن q و h به‌زای $n = 10^5, 25, 30$ و $q = 1, 10^5, 10^6$ را ببینید). سپس این دستگاه‌ها را با $\text{GMRES}(20)$ حل کنید. سپس این دستگاه‌ها را با $\text{GMRES}(20)$ پیش‌شرط‌سازی شده با پیش‌شرط‌ساز ILUT متلب با پارامترهای $\mu = 0.5, 0.1, 0.05$ حل کرده و نتایج را با هم مقایسه نمایید.

۷.۹ نشان دهید اگر A یک M -ماتریس باشد، آنگاه درمیان همه زیرماتریس‌های اصلی پیشروی A مثبت خواهند بود.

روش تکراری شکافت هرمیتی و هرمیتی کج

دستگاه معادلات خطی

$$Ax = b,$$

را در نظر بگیرید که در آن $A \in \mathbb{C}^{n \times n}$ یک ماتریس تُنک با ابعاد بزرگ و نامنفرد است. همان‌گونه که دیدیم روش گرادیان مزدوج (CG) مهم‌ترین روش برای حل دستگاه معادلات خطی با ماتریس ضرایب معین مثبت هرمیتی است. همچنین روش ماندهٔ مینیمال تعمیم‌یافته (GMRES) یک روش کارا برای حل دستگاه معادلات خطی با ماتریس ضرایب دلخواه است. اگر ماتریس A حقیقی و معین مثبت باشد (نه لزوماً متقارن) آنگاه نسخهٔ با شروع مجدد این الگوریتم همیشه همگرا است [۸۷، ۸۸]. در سال ۲۰۰۳، بای، گلوب و انجی در [۱۱] روش تکراری شکافت هرمیتی و هرمیتی کج (HSS) را ارائه کردند. نشان داده شده است که این الگوریتم برای هر دستگاه که ماتریس ضرایب آن معین مثبت (هرمیتی یا غیرهرمیتی) باشد، همگراست [۱۱]. این الگوریتم به سرعت مورد توجه محققین قرار گرفت و نسخه‌های مختلف از آن و همچنین، چندین کاربرد آن ارائه شده است. در این فصل مروری بر برخی از این نتایج خواهیم داشت.

۱-۱۰ روش تکراری شکافت هرمیتی و هرمیتی کج

فرض کنید ماتریس نامنفرد A دو شکافت به صورت

$$A = M - N = P - Q,$$

داشته باشد. با کمک این دو شکافت، دستگاه $Ax = b$ را به دو صورت زیر می‌نویسیم

$$\begin{cases} Mx = Nx + b, \\ Px = Qx + b. \end{cases}$$

اکنون، برای حل دستگاه $Ax = b$ روند دومرحله‌ای زیر را می‌سازیم

$$\begin{cases} Mx^{(k+\frac{1}{r})} = Nx^{(k)} + b, \\ Px^{(k+1)} = Qx^{(k+\frac{1}{r})} + b. \end{cases} \quad k = 0, 1, \dots \quad (1-10)$$

با محاسبه $x^{(k+\frac{1}{r})}$ از رابطه نخست و جایگذاری در رابطه دوم خواهیم داشت

$$x^{(k+1)} = P^{-1}QM^{-1}Nx^{(k)} + P^{-1}(I + QM^{-1})b. \quad (2-10)$$

این روند تکراری (و درنتیجه، روند تکراری دومرحله‌ای) به‌ازای هر حدس اولیه $x^{(0)}$ به جواب دستگاه $Ax = b$ همگراست اگر و تنها اگر

$$\rho(P^{-1}QM^{-1}N) < 1.$$

فرض کنید ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت باشد. می‌دانیم که ماتریس A را می‌توان به‌صورت

$$A = H + S, \quad (3-10)$$

شکافت که در آن

$$H = \frac{1}{r}(A + A^H), \quad S = \frac{1}{r}(A - A^H).$$

با توجه به قضیه ۱۵.۱، روشن است که ماتریس H معین مثبت هرمیتی است و ماتریس S هرمیتی‌کج می‌باشد. شکافت (۳-۱۰) به شکافت هرمیتی و هرمیتی‌کج معروف است. به‌ازای $\alpha > 0$ ، دو شکافت

$$\begin{cases} A = (\alpha I + H) - (\alpha I - S), \\ A = (\alpha I + S) - (\alpha I - H), \end{cases}$$

را ساخته و با توجه به روش تکراری دومرحله‌ای، روش تکراری HSS را به‌صورت زیر می‌سازیم

$$\begin{cases} (\alpha I + H)x^{(k+\frac{1}{r})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - H)x^{(k+\frac{1}{r})} + b. \end{cases} \quad k = 0, 1, 2, \dots$$

روشن است که ماتریس $\alpha I + H$ معین مثبت هرمیتی است و بنابراین، نامنفرد است. از طرفی، با توجه به تمرین ۱۲.۱، مقادیر ویژه $\alpha I + S$ به صورت $\alpha + i\beta$ هستند که در آن $\alpha > 0$. لذا، مقادیر ویژه ماتریس $\alpha I + S$ ناصفر می باشند و در نتیجه، ماتریس $\alpha I + S$ نامنفرد است.

در هر مرحله از روند تکراری HSS، جواب هر یک از دستگاه ها را با استفاده از روش های مستقیم به دست می آوریم. در مرحله نخست از $x^{(0)}$ به عنوان یک حدس اولیه برای دستگاه اول استفاده کرده، و سپس تقریب حاصل از این دستگاه را در دستگاه دوم به عنوان حدس اولیه به کار می بریم. در نهایت تقریب به دست آمده از دستگاه دوم را، در مرحله بعد به عنوان جواب اولیه استفاده می کنیم. در هر تکرار از روش تکراری HSS، ماتریس های دو تکرار داخلی بین بخش های هرمیتی و هرمیتی کج ماتریس A تغییر می کند. این روش شبیه به روش جهت های متناوب ضمنی (ADI) می باشد، که برای حل معادلات دیفرانسیل جزئی به کار می رود. برای جریات بیشتر درباره روش ADI به [۳۱، ۸۲] مراجعه کنید. پیش از اینکه به همگرایی روش تکراری HSS بپردازیم، لم زیر را بیان می کنیم که اثبات آن ساده است و به خواننده واگذار می شود.

لم ۱.۱۰ فرض کنید ماتریس $M \in \mathbb{C}^{n \times n}$ نامنفرد باشد. اگر به ازای هر $x \in \mathbb{C}^n$ تعریف کنیم $\|x\|_M = \|Mx\|_2$ ، آنگاه $\|\cdot\|_M$ یک نرم روی $\mathbb{C}^n$ است. نرم ماتریسی تولید شده روی $\mathbb{C}^{n \times n}$ که توسط نرم برداری $\|\cdot\|_2$ تولید می شود به صورت $\|X\|_M = \|MXM^{-1}\|_2$ است.

قضیه ۱.۱۰ فرض کنید ماتریس A معین مثبت و $H = \frac{1}{4}(A + A^H)$ و $S = \frac{1}{4}(A - A^H)$ به ترتیب بخش های هرمیتی و هرمیتی کج آن باشند و $\alpha > 0$. در این صورت، $G(\alpha)$ ، ماتریس تکرار روش HSS به صورت

$$G(\alpha) = (\alpha I + S)^{-1}(\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S), \quad (4-10)$$

است و کران بالای $\rho(G(\alpha))$ ، به صورت

$$\delta(\alpha) = \max_{\lambda_i \in \sigma(H)} \left| \frac{\alpha - \lambda_i}{\alpha + \lambda_i} \right|,$$

خواهد بود که $\sigma(H)$ طیف ماتریس H است. به علاوه،

$$\rho(G(\alpha)) \leq \delta(\alpha) < 1.$$

۳۵۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

برهان. می‌دانیم که روش HSS یک روند تکرار دومارحله‌ای به صورت (۱۰-۱) است که در آن

$$M = H + \alpha I, \quad N = S - \alpha I, \quad (۵-۱۰)$$

$$P = S + \alpha I, \quad Q = H - \alpha I. \quad (۶-۱۰)$$

بنابراین، با توجه به رابطه (۲-۱۰) ماتریس تکرار روش به صورت رابطه (۴-۱۰) نوشته می‌شود. از آنجایی که برای هر دو ماتریس مربعی A و B با شرط اینکه ماتریس‌های AB و BA قابل محاسبه باشند، داریم $\rho(AB) = \rho(BA)$ (تمرین (۵.۱) را ببیند)، نتیجه می‌شود که

$$\begin{aligned} \rho(G(\alpha)) &= \rho\left((\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)(\alpha I + S)^{-1}\right) \\ &\leq \|(\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)(\alpha I + S)^{-1}\|_2 \\ &\leq \|(\alpha I - H)(\alpha I + H)^{-1}\|_2 \|(\alpha I - S)(\alpha I + S)^{-1}\|_2. \end{aligned}$$

اکنون، قرار می‌دهیم $Q(\alpha) = (\alpha I - S)(\alpha I + S)^{-1}$. با توجه به اینکه $S^H = -S$ ، داریم

$$\begin{aligned} Q^H(\alpha)Q(\alpha) &= (\alpha I + S)^{-H}(\alpha I - S)^H(\alpha I - S)(\alpha I + S)^{-1} \\ &= (\alpha I - S)^{-1}(\alpha I + S)(\alpha I - S)(\alpha I + S)^{-1} \\ &= (\alpha I - S)^{-1}(\alpha^2 I - \alpha S + \alpha S - S^2)(\alpha I + S)^{-1} \\ &= (\alpha I - S)^{-1}(\alpha^2 I - S^2)(\alpha I + S)^{-1} \\ &= (\alpha I - S)^{-1}(\alpha I - S)(\alpha I + S)(\alpha I + S)^{-1} = I. \end{aligned}$$

بنابراین، $Q(\alpha)$ یک ماتریس یکانی است و در نتیجه، $\|Q(\alpha)\|_2 = 1$. پس

$$\rho(G(\alpha)) \leq \|(\alpha I - H)(\alpha I + H)^{-1}\|_2 = \max_{\lambda_i \in \sigma(H)} \left| \frac{\alpha - \lambda_i}{\alpha + \lambda_i} \right|.$$

از اینکه $\lambda_i > 0$ ، $i = 1, 2, \dots, n$ و $\alpha > 0$ خواهیم داشت

$$\rho(G(\alpha)) \leq \delta(\alpha) < 1.$$

□

بدین ترتیب اثبات قضیه کامل می‌شود.

این قضیه نشان می‌دهد که نرخ همگرایی مجانبی روش تکراری HSS به وسیله $\delta(\alpha)$ کراندار می‌شود و این کران تنها به طیف H بستگی دارد و به طیف S ، و به بردارهای ویژه H ، S و A بستگی ندارد. در

اینجا لازم به ذکر است که اگر کران‌های بالا و پائین مقادیر ویژه H معلوم باشند، مقدار α را می‌توان طوری محاسبه کرد که $\delta(\alpha)$ کمینه شود.

نتیجه ۱۰.۱۰ فرض کنید ماتریس A یک ماتریس معین مثبت، H و S به ترتیب بخش‌های هرمیتی و هرمیتی کج آن و $\lambda_{\min}$ و $\lambda_{\max}$ به ترتیب کوچک‌ترین و بزرگ‌ترین مقادیر ویژه H باشند. اگر $\alpha > 0$ ، آنگاه

$$\alpha^* = \arg \min_{\alpha} \left\{ \max_{\lambda_{\min} \leq \lambda \leq \lambda_{\max}} \left| \frac{\alpha - \lambda}{\alpha + \lambda} \right| \right\} = \sqrt{\lambda_{\min} \lambda_{\max}},$$

و

$$\sigma(\alpha^*) = \frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} = \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1},$$

که در آن $\kappa(H)$ عدد حالت طیفی H می‌باشد ($\kappa(H) = \|H\|_2 \|H^{-1}\|_2$).

برهان. فرض کنید

$$\eta(\alpha) = \max_{\lambda_{\min} \leq \lambda \leq \lambda_{\max}} \left| \frac{\alpha - \lambda}{\alpha + \lambda} \right|.$$

تابع f را به صورت

$$f_{\alpha}(\lambda) = \frac{\alpha - \lambda}{\alpha + \lambda}, \quad \lambda > 0,$$

تعریف می‌کنیم. داریم

$$f'_{\alpha}(\lambda) = \frac{-2\alpha}{(\alpha + \lambda)^2} < 0.$$

بنابراین، تابع $f_{\alpha}(\lambda)$ به طور اکید نزولی است و در نتیجه،

$$\eta(\alpha) = \max \left\{ \left| \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}} \right|, \left| \frac{\alpha - \lambda_{\max}}{\alpha + \lambda_{\max}} \right| \right\}. \quad (10-7)$$

نخست، فرض کنید $\lambda_{\min} \leq \alpha \leq \lambda_{\max}$. در این صورت،

$$\eta(\alpha) = \max \left\{ \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}}, \frac{\lambda_{\max} - \alpha}{\lambda_{\max} + \alpha} \right\}.$$

اما

$$\begin{aligned} \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}} &\leq \frac{\lambda_{\max} - \alpha}{\lambda_{\max} + \alpha} &\iff &\lambda_1 \leq \alpha \leq \sqrt{\lambda_{\min} \lambda_{\max}}, \\ \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}} &\geq \frac{\lambda_{\max} - \alpha}{\lambda_{\max} + \alpha} &\iff &\sqrt{\lambda_{\min} \lambda_{\max}} \leq \alpha \leq \lambda_n. \end{aligned}$$

از این رو،

$$\eta(\alpha) = \begin{cases} \frac{\lambda_{\max} - \alpha}{\lambda_{\max} + \alpha}, & \lambda_1 \leq \alpha \leq \sqrt{\lambda_{\min} \lambda_{\max}}, \\ \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}}, & \sqrt{\lambda_{\min} \lambda_{\max}} \leq \alpha \leq \lambda_n. \end{cases}$$

این تابع را می‌توان برای هر $\alpha > 0$ به صورت زیر گسترش داد (چرا؟)

$$\eta(\alpha) = \begin{cases} f_1(\alpha) = \frac{\lambda_{\max} - \alpha}{\lambda_{\max} + \alpha}, & 0 < \alpha \leq \sqrt{\lambda_{\min} \lambda_{\max}}, \\ f_2(\alpha) = \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}}, & \alpha \geq \sqrt{\lambda_{\min} \lambda_{\max}}. \end{cases}$$

توابع $f_1(\alpha)$ و $f_2(\alpha)$ در بازه‌های متناظرشان به ترتیب اکیداً نزولی و اکیداً صعودی هستند. در نتیجه، کمینه تابع η به ازای $\alpha^* = \sqrt{\lambda_{\min} \lambda_{\max}}$ به دست می‌آید.

با استفاده از ایده‌های هندسی نیز می‌توان پارامتر بهینه را محاسبه کرد. به علاوه، بررسی هندسی نکات ارزشمندی را به ما خواهد آموخت. توابع $f_{\min}$ و $f_{\max}$ را به صورت زیر تعریف می‌کنیم

$$f_{\min}(\alpha) = \left| \frac{\alpha - \lambda_{\min}}{\alpha + \lambda_{\min}} \right|, \quad f_{\max}(\alpha) = \left| \frac{\alpha - \lambda_{\max}}{\alpha + \lambda_{\max}} \right|.$$

به ازای $\alpha \geq 0$ ، این دو تابع در شکل ۱-۱ رسم شده‌اند.

با توجه به شکل ۱-۱، روشن است که $\eta(\alpha)$ پیش از نقطه M شامل نمودار $f_{\max}(\alpha)$ و بعد از نقطه M شامل نمودار $f_{\min}(\alpha)$ می‌باشد. بنابراین، نقطه M ، نقطه کمینه $\eta(\alpha)$ خواهد بود. اینک، اگر α^* طول نقطه کمینه باشد، واضح است که $\lambda_{\min} < \alpha^* < \lambda_{\max}$. بنابراین، داریم $\alpha^* - \lambda_{\min} > 0$ و $\alpha^* - \lambda_{\max} < 0$. و با توجه به شکل، کمینه در نقطه‌ای اتفاق می‌افتد که

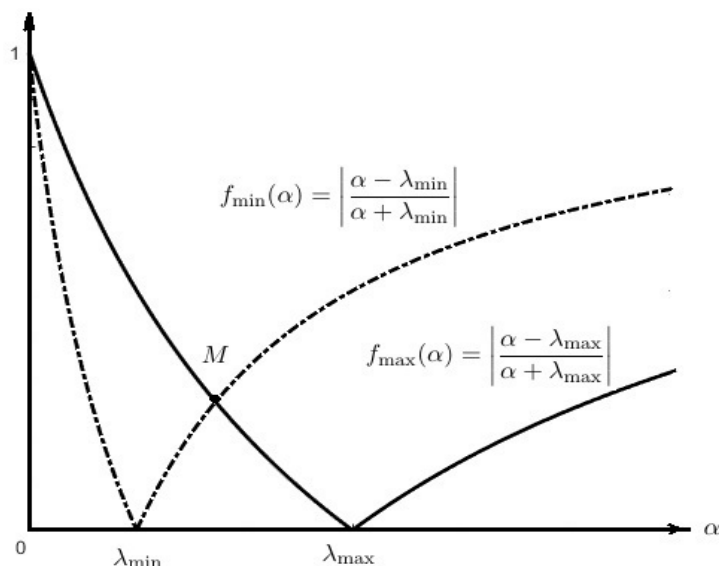
$$-\frac{\alpha^* - \lambda_{\min}}{\alpha^* + \lambda_{\min}} = \frac{\alpha^* - \lambda_{\max}}{\alpha^* + \lambda_{\max}}.$$

با حل این معادله، خواهیم داشت

$$\alpha^* = \sqrt{\lambda_{\min} \lambda_{\max}}.$$

اکنون، با قرار دادن α^* در $\eta(\alpha)$ ، خواهیم داشت

$$\sigma(\alpha^*) = \frac{\alpha^* - \lambda_{\min}}{\alpha^* + \lambda_{\min}} = \frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1}.$$



شکل ۱۰-۱ توابع $f_{\min}(\alpha)$ ، $f_{\max}(\alpha)$ و نقطه بهینه M .

در آخرین تساوی از اینکه ماتریس H معین مثبت هرمیتی است و عدد حالت ماتریس H به صورت

$$\kappa(H) = \|H\|_2 \|H^{-1}\|_2 = \frac{\lambda_{\max}}{\lambda_{\min}},$$

□

می باشد، استفاده شده است.

تأکید می کنیم که در نتیجه قبل، شعاع طیفی ماتریس تکرار، یعنی $\rho(G(\alpha))$ را کمینه نمی کند، بلکه $\delta(\alpha)$ ، که یک کران بالا برای $\rho(G(\alpha))$ می باشد را کمینه می کند. فرض کنید

$$x^{(k+1)} = G(\alpha)x^{(k)} + f, \quad (10-8)$$

بیانگر روند تکراری HSS باشد. به علاوه، فرض کنید که شرایط قضیه ۱۰-۱ برقرار باشد. روشن است اگر $x^* = A^{-1}b$ ، آنگاه

$$x^* = G(\alpha)x^* + f. \quad (10-9)$$

با تفريق طرفین رابطه‌های (۱۰-۸) و (۱۰-۹)، خواهیم داشت

$$x^{(k+1)} - x^* = G(\alpha)(x^{(k)} - x^*). \quad (10-10)$$

نرم‌های برداری و ماتریسی

$$|||x||| = \|(\alpha I + S)x\|_2, \quad \forall x \in \mathbb{C}^n,$$

$$|||X||| = \|(\alpha I + S)X(\alpha I + S)^{-1}\|_2, \quad \forall X \in \mathbb{C}^{n \times n},$$

را تعريف می‌کنیم. اکنون، از رابطه (۱۰-۹)، داریم

$$\begin{aligned} |||x^{(k+1)} - x^*||| &= |||G(\alpha)(x^{(k)} - x^*)||| = \|(\alpha I + S)G(\alpha)(x^{(k)} - x^*)\|_2 \\ &= \|(\alpha I + S)G(\alpha)(\alpha I + S)^{-1}(\alpha I + S)(x^{(k)} - x^*)\|_2 \\ &\leq \|(\alpha I + S)G(\alpha)(\alpha I + S)^{-1}\|_2 \|(\alpha I + S)(x^{(k)} - x^*)\|_2 \\ &= \|(\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)(\alpha I + S)^{-1}\|_2 |||x^{(k)} - x^*||| \\ &\leq \|(\alpha I - H)(\alpha I + H)^{-1}\|_2 |||x^{(k)} - x^*||| \\ &= \delta(\alpha) |||x^{(k)} - x^*|||. \end{aligned}$$

بنابراین، با توجه به نتیجه ۱۰.۱، اگر از α^* در روش تکراری HSS استفاده کنیم، آنگاه خواهیم داشت

$$|||x^{(k)} - x^*||| \leq (\sigma(\alpha^*))^k |||x^{(0)} - x^*||| = \left(\frac{\sqrt{\kappa(H)} - 1}{\sqrt{\kappa(H)} + 1} \right)^k |||x^{(0)} - x^*|||.$$

این رابطه نشان می‌دهد که سرعت همگرایی روش تکراری HSS تقریباً همانند روش تکراری گرادیان مزدوج است.

فرض کنید ماتریس A نرمال باشد، یعنی $A^H A = A A^H$. در این صورت، بنا به تمرین ۷.۱ داریم $HS = SH$. به سادگی می‌توان دید که ماتریس

$$\tilde{M}(\alpha) = (\alpha I + S)G(\alpha)(\alpha I + S)^{-1}, \quad (11-10)$$

نیز نرمال است (تمرین ۲.۱ را ببینید). اکنون، از قضیه ۱۸.۱ نتیجه می‌گیریم که

$$\rho(G(\alpha)) = \rho(\tilde{M}(\alpha)) = |||G(\alpha)|||.$$

با فرض $P(\alpha) = (\alpha I - H)(\alpha I + H)^{-1}$ و $Q(\alpha) = (\alpha I - S)(\alpha I + S)^{-1}$ ، به سادگی می‌توان نشان داد که $P(\alpha)Q(\alpha) = Q(\alpha)P(\alpha)$. لذا، با توجه به اینکه ماتریس $Q(\alpha)$ یکانی است، خواهیم داشت

$$\begin{aligned} |||G(\alpha)|||^2 &= ||P(\alpha)Q(\alpha)||^2 \\ &= \rho((P(\alpha)Q(\alpha))^H P(\alpha)Q(\alpha)) \\ &= \rho(Q(\alpha)^H P(\alpha)^H P(\alpha)Q(\alpha)) \\ &= \rho(P(\alpha)^H P(\alpha)Q(\alpha)Q(\alpha)^H) \\ &= \rho(P(\alpha)^H P(\alpha)) \\ &= \rho(P(\alpha))^2 = \delta(\alpha)^2. \end{aligned}$$

با جمع‌بندی این مطالب، می‌بینیم که اگر ماتریس A نرمال باشد، آنگاه

$$\rho(G(\alpha)) = |||G(\alpha)||| = \delta(\alpha).$$

در این صورت، α^* هر سه مقدار $\delta(\alpha)$ ، $|||G(\alpha)|||$ و $\rho(G(\alpha))$ را کمینه می‌کند.

مثال ۱۰.۱۰ دستگاه حاصل از گسسته‌سازی مدل معرفی شده در بخش ۸-۱ را با مقادیر $\sigma = -\tau = 10$ در نظر می‌گیریم. بردار سمت راست را طوری می‌گیریم که جواب دقیق دستگاه برابر با بردار $x^* = (1, 1, \dots, 1)^T \in \mathbb{R}^n$ باشد. دستگاه حاصل را با استفاده از روش HSS با حدس اولیه $x^{(0)} = 0$ و شرط توقف

$$\frac{\|r^{(k)}\|_2}{\|r^{(0)}\|_2} < 10^{-7},$$

حل کردیم که در آن $r^{(j)} = b - Ax^{(j)}$ بردار مانده تکرار j ام است. نتایج عددی در جدول ۱۰-۱ ارائه شده‌اند. به ازای هر n ، روش HSS را یک بار با پارامتر $\alpha^* = \sqrt{\lambda_{\max}\lambda_{\min}}$ و یک بار با پارامتر بهینه (α_{opt}) که به صورت آزمون و خطا به دست آمده است، اجرا شده است. در جدول ۱۰-۱ تعداد تکرارها برای همگرایی با $iters$ نشان داده شده است. همان گونه که ملاحظه می‌شود روش HSS روش کارایی برای حل این دستگاه می‌باشد.

از مزایای روش HSS این است که اگر ماتریس ضرایب دستگاه معین مثبت باشد (نه لزوماً هرمیتی) آنگاه روش برای هر $\alpha > 0$ همگرا است. ضعف اصلی آن این است که در هر تکرار، باید دو دستگاه

جدول ۱-۱۰ نتایج عددی برای مثال ۱.۱۰.

$n = N^2$	نتایج به‌ازای $\alpha^* = \sqrt{\lambda_{\max}\lambda_{\min}}$		نتایج به‌ازای پارامتر بهینه	
	α^*	$iters$	α_{opt}	$iters$
۹۰۰	۰/۴۰۵	۹۲	۰/۵۸۰	۷۱
۱۶۰۰	۰/۳۰۶	۱۲۰	۰/۴۴۰	۹۳
۲۵۰۰	۰/۲۴۶	۱۴۹	۰/۳۷۰	۱۱۴

معادلات خطی حل شود، که این کار در اکثر موارد مقرون به‌صرفه نخواهد بود. برای رفع این ضعف روند تکراری دیگری معرفی می‌گردد که به روند تکراری HSS نادقیق (IHSS) معروف است. این روش را در ادامه معرفی می‌کنیم.

۱-۲ روش شکافت هرمیتی و هرمیتی‌کج نادقیق

در هر تکرار از روش HSS، باید دو دستگاه با ماتریس ضرایب $\alpha I + H$ و $\alpha I + S$ حل شوند. همان‌گونه که ذکر گردید، این روند در مسائل واقعی، پرهزینه و غیرعملی خواهد بود. برای حل این مشکل و افزایش کارایی روش معرفی‌شده، هر کدام از دستگاه‌ها را با روش‌های تکراری مناسب، به‌صورت تقریبی حل می‌کنیم. به این ترتیب روش جدیدی ساخته می‌شود که روش شکافت هرمیتی و هرمیتی‌کج نادقیق (IHSS) نامیده می‌شود.

با توجه به اینکه ماتریس ضرایب دستگاه اول، یعنی $\alpha I + H$ ، معین مثبت هرمیتی است، این دستگاه را با روش گرادیان مزدوج حل می‌کنیم. روش گرادیان مزدوج مناسب‌ترین روش برای حل دستگاه‌های معین مثبت هرمیتی است. این ویژگی یکی از مزایای روش IHSS است و به این دلیل این روش بسیار مورد توجه قرار گرفته است. ماتریس ضرایب دستگاه دوم، انتقالی از یک ماتریس هرمیتی‌کج است. بنابراین، می‌توان از الگوریتم ارائه‌شده در [۳۶] استفاده کرد. لازم به‌ذکر است که از الگوریتم GMRES نیز می‌توان استفاده کرد.

روش تکراری IHSS: حدس اولیه $\bar{x}^{(0)}$ را در نظر گرفته و به‌ازای $k = 0, 1, \dots$ ، دنباله $\{\bar{x}^{(k)}\}$ را تا زمانی که همگرا شود به‌صورت زیر تولید کنید.

با استفاده از روش گرادیان مزدوج بردار $\bar{x}^{(k+\frac{1}{2})}$ را طوری محاسبه کنید که

$$(\alpha I + H)\bar{x}^{(k+\frac{1}{2})} \approx (\alpha I - S)\bar{x}^{(k)} + b. \quad (10-12)$$

با کمک یک روش مبتنی بر زیرفضای کرایلف مثل $GMRES$ بردار $\bar{x}^{(k+1)}$ را طوری محاسبه کنید که

$$(\alpha I + S)\bar{x}^{(k+1)} \approx (\alpha I - H)\bar{x}^{(k+\frac{1}{p})} + b. \quad (10-13)$$

برای سادگی در به کارگیری عددی و آنالیز همگرایی الگوریتم IHSS، این الگوریتم را به صورت زیر بازنویسی می کنیم. در گام نخست، از تکرار $(k+1)$ ام قرار می دهیم

$$\bar{z}^{(k)} := \bar{x}^{(k+\frac{1}{p})} - \bar{x}^{(k)}.$$

لذا،

$$\bar{x}^{(k+\frac{1}{p})} = \bar{x}^{(k)} + \bar{z}^{(k)}.$$

اکنون، با جایگذاری $\bar{x}^{(k+\frac{1}{p})}$ در گام نخست از روش تکراری IHSS خواهیم داشت

$$(\alpha I + H)(\bar{x}^{(k)} + \bar{z}^{(k)}) \approx (\alpha I - S)\bar{x}^{(k)} + b,$$

که معادل با

$$(\alpha I + H)\bar{z}^{(k)} \approx b - (H + S)\bar{x}^{(k)} = b - A\bar{x}^{(k)} = \bar{r}^{(k)},$$

است. در تکرار $(k+1)$ ام از روش IHSS و در گام نخست، یک تقریب مثل $\bar{z}^{(k)}$ برای دستگاه

$$(\alpha I + H)z^{(k)} = \bar{r}^{(k)},$$

محاسبه می کنیم به طوری که بردار مانده

$$\bar{p}^{(k)} = \bar{r}^{(k)} - (\alpha I + H)\bar{z}^{(k)},$$

در شرط توقف $\|\bar{p}^{(k)}\| \leq \varepsilon_k \|\bar{r}^{(k)}\|$ صدق کند. سپس، قرار می دهیم

$$\bar{x}^{(k+\frac{1}{p})} = \bar{x}^{(k)} + \bar{z}^{(k)},$$

به طور مشابه و در گام دوم، یک تقریب مثل $\bar{z}^{(k+\frac{1}{p})}$ برای دستگاه

$$(\alpha I + S)z^{(k+\frac{1}{p})} = \bar{r}^{(k+\frac{1}{p})} = b - A\bar{x}^{(k+\frac{1}{p})},$$

محاسبه می کنیم به طوری که بردار مانده

$$\bar{q}^{(k+\frac{1}{p})} = \bar{r}^{(k+\frac{1}{p})} - (\alpha I + S)\bar{z}^{(k+\frac{1}{p})},$$

در شرط توقف $\|\bar{q}^{(k+\frac{1}{2})}\| \leq \eta_k \|\bar{r}^{(k+\frac{1}{2})}\|$ صدق کند. سپس، قرار می‌دهیم

$$\bar{x}^{(k+1)} = \bar{x}^{(k+\frac{1}{2})} + \bar{z}^{(k+\frac{1}{2})}.$$

بنابراین، الگوریتم روش IHSS را به‌صورت زیر جمع‌بندی می‌کنیم. لازم به ذکر است که در اینجا $\|\cdot\|$ یک نرم برداری دلخواه روی $\mathbb{C}^n$ است.

الگوریتم ۱۰.۱۰ الگوریتم IHSS

1. Choose and initial guess $\bar{x}^{(0)}$ and $\bar{r}^{(0)} = b - A\bar{x}^{(0)}$.
2. For $k = 0, 1, \dots$, until convergence Do:
 3. Solve $(\alpha I + H)z^{(k)} = \bar{r}^{(k)}$, by the CG method to compute $\bar{z}^{(k)}$ satisfying $\|\bar{p}^{(k)}\| \leq \varepsilon_k \|\bar{r}^{(k)}\|$ where $\bar{p}^{(k)} = \bar{r}^{(k)} - (\alpha I + H)\bar{z}^{(k)}$.
 4. $\bar{x}^{(k+\frac{1}{2})} := \bar{x}^{(k)} + \bar{z}^{(k)}$
 5. $\bar{r}^{(k+\frac{1}{2})} := b - A\bar{x}^{(k+\frac{1}{2})}$
 6. Solve $(\alpha I + S)z^{(k+\frac{1}{2})} = \bar{r}^{(k+\frac{1}{2})}$, by the GMRES method satisfying $\|\bar{q}^{(k+\frac{1}{2})}\| \leq \eta_k \|\bar{r}^{(k+\frac{1}{2})}\|$ where $\bar{q}^{(k+\frac{1}{2})} = \bar{r}^{(k+\frac{1}{2})} - (\alpha I + S)\bar{z}^{(k+\frac{1}{2})}$.
 7. $\bar{x}^{(k+1)} := \bar{x}^{(k+\frac{1}{2})} + \bar{z}^{(k+\frac{1}{2})}$
 8. $\bar{r}^{(k+1)} := b - A\bar{x}^{(k+1)}$
 9. EndDo

قضیه بعد، روند IHSS را در حالت کلی بیان کرده و به بررسی همگرایی آن می‌پردازد. در این قضیه،

به‌ازای هر بردار $x \in \mathbb{C}^n$ ، نرم برداری $\|\cdot\|_{M_T}$ را به‌صورت

$$\|x\|_{M_T} = \|M_T x\|,$$

تعریف می‌کنیم، که در آن $\|\cdot\|$ نرمی است که در الگوریتم ۱۰.۱۰ استفاده شده است. در این صورت، نرم

طبیعی تولیدشده توسط این نرم به‌صورت زیر خواهد بود (به‌ازای هر $X \in \mathbb{R}^{n \times n}$)

$$\begin{aligned} \|X\|_{M_T} &= \max_{x \neq 0} \frac{\|Xx\|_{M_T}}{\|x\|_{M_T}} \\ &= \max_{x \neq 0} \frac{\|M_T Xx\|}{\|M_T x\|} \\ &= \max_{x \neq 0} \frac{\|M_T X M_T^{-1} M_T x\|}{\|M_T x\|} \\ &= \max_{y \neq 0} \frac{\|M_T X M_T^{-1} y\|}{\|y\|} \\ &= \|M_T X M_T^{-1}\|. \end{aligned}$$

اکنون با در نظر گرفتن نرم‌های برداری و ماتریسی $\|\cdot\|_2$ قضیه زیر را بیان و ثابت می‌کنیم.

قضیه ۲.۱۰ فرض کنید $A \in \mathbb{C}^{n \times n}$ و $A = M_i - N_i$ ، $i = 1, 2$ ، دو شکافت برای ماتریس A باشند. اگر $\{\bar{x}^{(k)}\}$ دنباله تکرار دومرحله‌ای باشد که به صورت زیر تعریف شود

$$M_1 \bar{z}^{(k)} = \bar{r}^{(k)} + \bar{p}^{(k)}, \quad \bar{x}^{(k+1)} = \bar{x}^{(k)} + \bar{z}^{(k)}, \quad (14-10)$$

با شرط $\|\bar{r}^{(k)}\| = b - A\bar{x}^{(k)}$ که در آن $\|\bar{p}^{(k)}\| \leq \varepsilon_k \|\bar{r}^{(k)}\|$ و

$$M_2 \bar{z}^{(k+1)} = \bar{r}^{(k+1)} + \bar{q}^{(k+1)}, \quad \bar{x}^{(k+1)} = \bar{x}^{(k+1)} + \bar{z}^{(k+1)}, \quad (15-10)$$

با شرط $\|\bar{q}^{(k+1)}\| \leq \eta_k \|\bar{r}^{(k+1)}\|$ که در آن $\bar{r}^{(k+1)} = b - A\bar{x}^{(k+1)}$ ، آنگاه خواهیم داشت

$$\begin{aligned} \bar{x}^{(k+1)} &= M_1^{-1} N_1 M_1^{-1} N_1 \bar{x}^{(k)} + M_1^{-1} (I + N_2 M_1^{-1}) b \\ &\quad + M_1^{-1} (N_2 M_1^{-1} \bar{p}^{(k)} + \bar{q}^{(k+1)}). \end{aligned}$$

به علاوه، اگر $x^* \in \mathbb{C}^n$ جواب واقعی دستگاه $Ax = b$ باشد، آنگاه به ازای $k = 0, 1, 2, \dots$ داریم

$$\|\bar{x}^{(k+1)} - x^*\|_{M_1} \leq (\bar{\sigma} + \bar{\mu} \bar{\theta} \varepsilon_k + \bar{\theta} (\bar{\rho} + \bar{\theta} \bar{\nu} \varepsilon_k) \eta_k) \|\bar{x}^{(k)} - x^*\|_{M_1}, \quad (16-10)$$

که در آن

$$\begin{aligned} \bar{\sigma} &= \|N_2 M_1^{-1} N_1 M_1^{-1}\|, \quad \bar{\rho} = \|M_2 M_1^{-1} N_1 M_1^{-1}\|, \\ \bar{\mu} &= \|N_2 M_1^{-1}\|, \quad \bar{\theta} = \|A M_1^{-1}\|, \quad \bar{\nu} = \|M_2 M_1^{-1}\|. \end{aligned}$$

در حالت خاص، اگر

$$\bar{\sigma} + \bar{\mu} \bar{\theta} \varepsilon_{\max} + \bar{\theta} (\bar{\rho} + \bar{\theta} \bar{\nu} \varepsilon_{\max}) \eta_{\max} < 1,$$

آنگاه دنباله $\{\bar{x}^{(k)}\}$ به x^* همگرا خواهد بود، که در آن $\varepsilon_{\max} = \max_k \{\varepsilon_k\}$ و $\eta_{\max} = \max_k \{\eta_k\}$. برهان. از رابطه (۱۴-۱۰)، داریم

$$\begin{aligned} \bar{x}^{(k+1)} &= \bar{x}^{(k)} + M_1^{-1} (\bar{r}^{(k)} + \bar{p}^{(k)}) \\ &= \bar{x}^{(k)} + M_1^{-1} (b - A\bar{x}^{(k)}) + M_1^{-1} \bar{p}^{(k)} \\ &= (I - M_1^{-1} A) \bar{x}^{(k)} + M_1^{-1} b + M_1^{-1} \bar{p}^{(k)} \end{aligned}$$

۳۶۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

$$\begin{aligned} &= (I - M_1^{-1}(M_1 - N_1))\bar{x}^{(k)} + M_1^{-1}b + M_1^{-1}\bar{p}^{(k)} \\ &= M_1^{-1}N_1\bar{x}^{(k)} + M_1^{-1}b + M_1^{-1}\bar{p}^{(k)}. \end{aligned} \quad (۱۷-۱۰)$$

به‌طور مشابه از رابطه (۱۵-۱۰)، داریم

$$\begin{aligned} \bar{x}^{(k+1)} &= \bar{x}^{(k+\frac{1}{r})} + M_1^{-1}(\bar{r}^{(k+\frac{1}{r})} + \bar{q}^{(k+\frac{1}{r})}) \\ &= (I - M_1^{-1}A)\bar{x}^{(k+\frac{1}{r})} + M_1^{-1}b + M_1^{-1}\bar{q}^{(k+\frac{1}{r})} \\ &= M_1^{-1}N_1\bar{x}^{(k+\frac{1}{r})} + M_1^{-1}b + M_1^{-1}\bar{q}^{(k+\frac{1}{r})}. \end{aligned}$$

با جایگذاری $x^{(k+\frac{1}{r})}$ از (۱۷-۱۰) در تساوی اخیر، خواهیم داشت

$$\begin{aligned} \bar{x}^{(k+1)} &= M_1^{-1}N_1(M_1^{-1}N_1\bar{x}^{(k)} + M_1^{-1}b + M_1^{-1}\bar{p}^{(k)}) + M_1^{-1}b + M_1^{-1}\bar{q}^{(k+\frac{1}{r})} \\ &= M_1^{-1}N_1M_1^{-1}N_1\bar{x}^{(k)} + M_1^{-1}N_1M_1^{-1}b + M_1^{-1}N_1M_1^{-1}\bar{p}^{(k)} \\ &\quad + M_1^{-1}b + M_1^{-1}\bar{q}^{(k+\frac{1}{r})} \\ &= M_1^{-1}N_1M_1^{-1}N_1\bar{x}^{(k)} + M_1^{-1}(I + N_1M_1^{-1})b \\ &\quad + M_1^{-1}(N_1M_1^{-1}\bar{p}^{(k)} + \bar{q}^{(k+\frac{1}{r})}). \end{aligned} \quad (۱۸-۱۰)$$

از آنجایی که $x^* \in \mathbb{C}^n$ جواب واقعی دستگاه $Ax = b$ است و $A = M_1 - N_1$ ، خواهیم داشت

$$x^* = M_1^{-1}N_1x^* + M_1^{-1}b. \quad (۱۹-۱۰)$$

به‌طور مشابه، با توجه به $A = M_2 - N_2$ ، می‌بینیم که

$$x^* = M_2^{-1}N_2x^* + M_2^{-1}b.$$

حال با جایگذاری x^* از (۱۹-۱۰) در طرف راست تساوی اخیر، خواهیم داشت

$$x^* = M_2^{-1}N_2M_1^{-1}N_1x^* + M_2^{-1}(I + N_2M_1^{-1})b. \quad (۲۰-۱۰)$$

با تفریق طرفین رابطه (۲۰-۱۰) از طرفین (۱۸-۱۰) و تفریق طرفین رابطه (۱۹-۱۰) از طرفین رابطه (۱۸-۱۰)، خواهیم داشت

$$\bar{x}^{(k+\frac{1}{r})} - x^* = M_1^{-1}N_1(\bar{x}^{(k)} - x^*) + M_1^{-1}\bar{p}^{(k)}, \quad (۲۱-۱۰)$$

و

$$\bar{x}^{(k+1)} - x^* = M_{\Upsilon}^{-1} N_{\Upsilon} M_{\Upsilon}^{-1} N_{\Upsilon} (\bar{x}^{(k)} - x^*) + M_{\Upsilon}^{-1} (N_{\Upsilon} M_{\Upsilon}^{-1} \bar{p}^{(k)} + \bar{q}^{(k+\frac{1}{\Upsilon})}). \quad (22-10)$$

با گرفتن نرم از طرفین رابطه‌های (21-10) و (22-10) داریم

$$\begin{aligned} \|\bar{x}^{(k+\frac{1}{\Upsilon})} - x^*\|_{M_{\Upsilon}} &\leq \|M_{\Upsilon}^{-1} N_{\Upsilon} (\bar{x}^{(k)} - x^*)\|_{M_{\Upsilon}} + \|M_{\Upsilon}^{-1} \bar{p}^{(k)}\|_{M_{\Upsilon}} \\ &\leq \|M_{\Upsilon}^{-1} N_{\Upsilon}\|_{M_{\Upsilon}} \|\bar{x}^{(k)} - x^*\|_{M_{\Upsilon}} + \|M_{\Upsilon}^{-1} \bar{p}^{(k)}\|_{M_{\Upsilon}} \\ &\leq \|M_{\Upsilon} M_{\Upsilon}^{-1} N_{\Upsilon} M_{\Upsilon}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\Upsilon}} \\ &\quad + \|M_{\Upsilon} M_{\Upsilon}^{-1}\| \cdot \|\bar{p}^{(k)}\|, \end{aligned}$$

و

$$\begin{aligned} \|\bar{x}^{(k+1)} - x^*\|_{M_{\Upsilon}} &\leq \|M_{\Upsilon}^{-1} N_{\Upsilon} M_{\Upsilon}^{-1} N_{\Upsilon}\|_{M_{\Upsilon}} \|\bar{x}^{(k)} - x^*\|_{M_{\Upsilon}} \\ &\quad + \|M_{\Upsilon}^{-1} (N_{\Upsilon} M_{\Upsilon}^{-1} \bar{p}^{(k)} + \bar{q}^{(k+\frac{1}{\Upsilon})})\|_{M_{\Upsilon}} \\ &= \|N_{\Upsilon} M_{\Upsilon}^{-1} N_{\Upsilon} M_{\Upsilon}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\Upsilon}} \\ &\quad + \|N_{\Upsilon} M_{\Upsilon}^{-1} \bar{p}^{(k)} + \bar{q}^{(k+\frac{1}{\Upsilon})}\| \\ &\leq \|N_{\Upsilon} M_{\Upsilon}^{-1} N_{\Upsilon} M_{\Upsilon}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\Upsilon}} \\ &\quad + \|N_{\Upsilon} M_{\Upsilon}^{-1}\| \cdot \|\bar{p}^{(k)}\| + \|\bar{q}^{(k+\frac{1}{\Upsilon})}\|. \quad (23-10) \end{aligned}$$

توجه کنید که

$$\begin{aligned} \|\bar{r}^{(k)}\| &= \|b - A\bar{x}^{(k)}\| = \|A(x^* - \bar{x}^{(k)})\| \\ &= \|M_{\Upsilon}^{-1} A(x^* - \bar{x}^{(k)})\|_{M_{\Upsilon}} \\ &\leq \|M_{\Upsilon}^{-1} A\|_{M_{\Upsilon}} \|x^* - \bar{x}^{(k)}\|_{M_{\Upsilon}} \\ &= \|AM_{\Upsilon}^{-1}\| \cdot \|x^* - \bar{x}^{(k)}\|_{M_{\Upsilon}}. \quad (24-10) \end{aligned}$$

به همین ترتیب، داریم

$$\|\bar{r}^{(k+\frac{1}{\Upsilon})}\| \leq \|AM_{\Upsilon}^{-1}\| \cdot \|x^* - \bar{x}^{(k+\frac{1}{\Upsilon})}\|_{M_{\Upsilon}}. \quad (25-10)$$

از تعریف دنباله $\{\bar{p}^{(k)}\}$ و رابطه (۲۴-۱۰)، خواهیم داشت

$$\|\bar{p}^{(k)}\| \leq \varepsilon_k \|\bar{r}^{(k)}\| \leq \varepsilon_k \|AM_{\gamma}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}}. \quad (۲۶-۱۰)$$

همچنین از تعریف دنباله $\{\bar{q}^{(k+\frac{1}{\gamma})}\}$ و رابطه‌های (۲۱-۱۰) و (۲۵-۱۰) نتیجه می‌گیریم که

$$\begin{aligned} \|\bar{q}^{(k+\frac{1}{\gamma})}\| &\leq \eta_k \|\bar{r}^{(k+\frac{1}{\gamma})}\| \\ &\leq \eta_k \|AM_{\gamma}^{-1}\|_{\gamma} \|x^* - \bar{x}^{(k+\frac{1}{\gamma})}\|_{M_{\gamma}} \\ &\leq \eta_k \|AM_{\gamma}^{-1}\| \left(\|M_{\gamma}^{-1} N_{\gamma} (\bar{x}^{(k)} - x^*)\|_{M_{\gamma}} + \|M_{\gamma}^{-1} \bar{p}^{(k)}\|_{M_{\gamma}} \right) \\ &\leq \eta_k \|AM_{\gamma}^{-1}\| \left(\|M_{\gamma} M_{\gamma}^{-1} N_{\gamma} M_{\gamma}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}} \right. \\ &\quad \left. + \|M_{\gamma} M_{\gamma}^{-1}\| \cdot \|\bar{p}^{(k)}\| \right) \\ &\leq \eta_k \|AM_{\gamma}^{-1}\| \left(\|M_{\gamma} M_{\gamma}^{-1} N_{\gamma} M_{\gamma}^{-1}\| \right. \\ &\quad \left. + \varepsilon_k \|M_{\gamma} M_{\gamma}^{-1}\| \cdot \|AM_{\gamma}^{-1}\| \right) \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}}. \end{aligned} \quad (۲۷-۱۰)$$

سرانجام با جایگذاری رابطه‌های (۲۷-۱۰) و (۲۶-۱۰) در رابطه (۲۳-۱۰) خواهیم داشت

$$\begin{aligned} \|\bar{x}^{(k+1)} - x^*\|_{M_{\gamma}} &\leq \|N_{\gamma} M_{\gamma}^{-1} N_{\gamma} M_{\gamma}^{-1}\| \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}} \\ &\quad + \|N_{\gamma} M_{\gamma}^{-1}\|_{\gamma} \cdot \varepsilon_k \|AM_{\gamma}^{-1}\|_{\gamma} \cdot \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}} \\ &\quad + \eta_k \|AM_{\gamma}^{-1}\|_{\gamma} \left(\|M_{\gamma} M_{\gamma}^{-1} N_{\gamma} M_{\gamma}^{-1}\|_{\gamma} \right. \\ &\quad \left. + \varepsilon_k \|M_{\gamma} M_{\gamma}^{-1}\| \cdot \|AM_{\gamma}^{-1}\|_{\gamma} \right) \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}} \\ &\leq (\bar{\sigma} + \bar{\mu} \bar{\theta} \varepsilon_k + \bar{\theta} (\bar{\rho} + \bar{\theta} \bar{\nu} \varepsilon_k) \eta_k) \|\bar{x}^{(k)} - x^*\|_{M_{\gamma}}, \end{aligned}$$

□ که اثبات قضیه را کامل می‌کند.

با توجه به قضیه قبل، اگر در روش IHSS، دستگاه‌های داخلی به‌طور دقیق حل شوند، مقادیر ε_k و η_k و ازاین‌رو، $\varepsilon_{\max}$ و $\eta_{\max}$ برابر با صفر خواهند بود. در این حالت، سرعت همگرایی روش‌های تکراری IHSS و HSS یکسان خواهد بود.

دیدیم که در روش‌های تکراری HSS و IHSS شکافت ماتریس A به‌صورت‌های

$$\begin{aligned} A &= M_{\gamma} - N_{\gamma} = (\alpha I + H) - (\alpha I - S) \\ &= M_{\gamma} - N_{\gamma} = (\alpha I + S) - (\alpha I - H), \end{aligned}$$

می‌باشند. بنابراین، از قضیه قبل نتیجه زیر حاصل می‌شود.

قضیه ۳.۱۰ فرض کنید ماتریس A معین مثبت، $H = (A + A^H)/2$ و $S = (A - A^H)/2$ به ترتیب بخش‌های هرمیتی و هرمیتی کج آن باشند و $\alpha > 0$. اگر $\{x^{(k)}\}$ دنباله تولید شده توسط روند تکراری IHSS برای دستگاه $Ax = b$ باشد، آنگاه

$$\|\bar{x}^{(k+1)} - x^*\| \leq (\delta(\alpha) + \theta \rho \eta_k)(1 + \theta \varepsilon_k) \|\bar{x}^{(k)} - x^*\|, \quad k = 0, 1, 2, \dots,$$

که در آن

$$\rho = \|(\alpha I + S)(\alpha I + H)^{-1}\|_2, \quad \theta = \|A(\alpha I + S)^{-1}\|_2. \quad (28-10)$$

به علاوه، اگر

$$(\delta(\alpha) + \theta \rho \eta_{\max})(1 + \theta \varepsilon_{\max}) < 1,$$

آنگاه دنباله $\{\bar{x}^{(k)}\}$ به جواب واقعی $x^* \in \mathbb{C}^n$ همگرا خواهد بود که در آن $\varepsilon_{\max} = \max_k \{\varepsilon_k\}$ و $\eta_{\max} = \max_k \{\eta_k\}$. لازم به ذکر است که در اینجا M_2 با $\alpha I + S$ جایگزین شده است. به همین دلیل از نرم $\|\cdot\|_{M_2}$ به جای $\|\cdot\|$ استفاده شده است. **برهان.** این قضیه، به سادگی با جایگذاری

$$M_1 = \alpha I + H, \quad M_2 = \alpha I + S,$$

$$N_1 = \alpha I - S, \quad N_2 = \alpha I - H.$$

در قضیه قبل، حکم ثابت می‌شود. $\square$

با دقت در قضیه ۲.۱۰، می‌توانیم $\{\varepsilon_k\}$ و $\{\eta_k\}$ را طوری انتخاب کنیم که محاسبات در روش تکراری IHSS تا حد امکان کاهش یابد. قضیه بعد روش مفیدی برای انتخاب مقادیر $\{\varepsilon_k\}$ و $\{\eta_k\}$ معرفی می‌کند که سرعت همگرایی طرح تکراری IHSS بهبود یابد.

قضیه ۴.۱۰ با فرضیات قضیه ۲.۱۰، فرض کنید $\{\tau_1(k)\}$ و $\{\tau_2(k)\}$ دنباله‌های مثبت و غیرکاهشی، با شرایط $\tau_1(k) \geq 1$ ، $\tau_2(k) \geq 1$ و

$$\limsup \tau_1(k) = \limsup \tau_2(k) = +\infty,$$

باشند. به علاوه، فرض کنید δ_1 و δ_2 ثابت‌های حقیقی در بازه $(0, 1)$ باشند به طوری که

$$\varepsilon_k \leq c_1 \delta_1^{\tau_1(k)}, \quad \eta_k \leq c_2 \delta_2^{\tau_2(k)}, \quad k = 0, 1, 2, \dots, \quad (29-10)$$

که در آن c_1 و c_2 ثابت‌های نامنفی هستند. در این صورت، خواهیم داشت

$$\|\bar{x}^{(k+1)} - x^*\|_{M_T} \leq (\sqrt{\sigma} + \bar{\omega} \bar{\theta} \delta^{\tau(k)})^2 \|\bar{x}^{(k)} - x^*\|_{M_T}, \quad k = 0, 1, 2, \dots,$$

که

$$\tau(k) = \min\{\tau_1(k), \tau_2(k)\}, \quad \delta = \max\{\delta_1, \delta_2\}, \quad (30-10)$$

و

$$\bar{\omega} = \max\left\{\sqrt{c_1 c_2 \bar{\nu}}, \frac{1}{2\sqrt{\sigma}}(c_1 \bar{\mu} + c_2 \bar{\rho})\right\}.$$

همچنین

$$\limsup_{k \rightarrow \infty} \frac{\|\bar{x}^{(k+1)} - x^*\|_{M_T}}{\|\bar{x}^{(k)} - x^*\|_{M_T}} = \sigma.$$

بنابراین، به طور مجانبی سرعت همگرایی روش تکراری دومرحله‌ای با روش تکراری دومرحله‌ای نادقیق یکسان است.

برهان. با استفاده از رابطه‌های $(16-10)$ و $(29-10)$ ، داریم

$$\begin{aligned} \|\bar{x}^{(k+1)} - x^*\|_{M_T} &\leq (\bar{\sigma} + \bar{\mu} \bar{\theta} \varepsilon_k + \bar{\theta}(\bar{\rho} + \bar{\theta} \bar{\nu} \varepsilon_k) \eta_k) \|\bar{x}^{(k)} - x^*\|_{M_T} \\ &\leq (\bar{\sigma} + \bar{\mu} \bar{\theta} c_1 \delta_1^{\tau_1(k)} + \bar{\theta}(\bar{\rho} + \bar{\theta} \bar{\nu} c_1 \delta_1^{\tau_1(k)}) c_2 \delta_2^{\tau_2(k)}) \|\bar{x}^{(k)} - x^*\|_{M_T} \\ &\leq (\bar{\sigma} + \bar{\mu} \bar{\theta} c_1 \delta^{\tau(k)} + \bar{\theta}(\bar{\rho} + \bar{\theta} \bar{\nu} c_1 \delta^{\tau(k)}) c_2 \delta^{\tau(k)}) \|\bar{x}^{(k)} - x^*\|_{M_T} \\ &= (\sigma + (c_1 \bar{\mu} + c_2 \bar{\rho}) \bar{\theta} \delta^{\tau(k)} + c_1 c_2 \bar{\nu} \bar{\theta}^2 \delta^{2\tau(k)}) \|\bar{x}^{(k)} - x^*\|_{M_T} \\ &\leq (\sqrt{\sigma} + \bar{\omega} \bar{\theta} \delta^{\tau(k)})^2 \|\bar{x}^{(k)} - x^*\|_{M_T}. \end{aligned}$$

از نامساوی اخیر خواهیم داشت

$$\frac{\|\bar{x}^{(k+1)} - x^*\|_{M_T}}{\|\bar{x}^{(k)} - x^*\|_{M_T}} \leq (\sqrt{\sigma} + \bar{\omega} \bar{\theta} \delta^{\tau(k)})^2.$$

اکنون، با توجه به

$$\lim_{k \rightarrow \infty} (\sqrt{\bar{\sigma}} + \bar{\omega} \bar{\theta} \delta^{\tau(k)})^2 = \bar{\sigma},$$

اثبات قضیه کامل می‌شود. $\square$

قضیه قبل در حالت کلی بیان و از ماتریس M_2 استفاده شده است. با وارد کردن ماتریس‌های H و S در روند تکراری، قضیه بعد نتیجه می‌شود.

قضیه ۵.۱۰ فرض کنید ماتریس A معین مثبت، H و S به ترتیب بخش‌های هرمیتی و هرمیتی کج آن باشند و $\alpha > 0$. اگر $\{\bar{x}^{(k)}\}$ دنباله تولید شده در روش IHSS و $\{\tau_1(k)\}$ و $\{\tau_2(k)\}$ دنباله‌های مثبت و ناکاهشی، با شرایط $\tau_1(k) \geq 1$ ، $\tau_2(k) \geq 1$ و

$$\limsup \tau_1(k) = \limsup \tau_2(k) = +\infty,$$

و δ_1 و δ_2 ثابت‌هایی باشند که در رابطه (۲۹-۱۰) معرفی شدند، آنگاه خواهیم داشت

$$\|\bar{x}^{(k+1)} - x^*\| \leq (\sqrt{\delta(\alpha)} + \omega \theta \delta^{\tau(k)})^2 \|\bar{x}^{(k)} - x^*\|, \quad k = 0, 1, \dots,$$

که ρ و θ در رابطه (۲۸-۱۰) و $\tau(k)$ و δ در رابطه (۳۰-۱۰) تعریف شده‌اند و

$$\omega = \max\left\{\sqrt{c_1 c_2 \rho}, \frac{1}{2\sqrt{\delta(\alpha)}}(c_1 \delta(\alpha) + c_2 \rho)\right\}.$$

به‌ویژه،

$$\limsup_{k \rightarrow \infty} \frac{\|\bar{x}^{(k+1)} - x^*\|}{\|\bar{x}^{(k)} - x^*\|} = \delta(\alpha).$$

این رابطه نشان می‌دهد که به‌طور مجانبی سرعت همگرایی روش IHSS با سرعت همگرایی HSS یکسان است.

برهان. با استفاده از قضایای ۳.۱۰ و ۴.۱۰ این قضیه به‌سادگی ثابت می‌شود. $\square$

مثال ۲.۱۰ دستگاه معادلات خطی مثال ۱.۱۰ را در نظر بگیرید. این دستگاه را با روش IHSS حل می‌کنیم. برای این کار، در هر تکرار دستگاه متناظر با ماتریس ضرایب $\alpha I + H$ را با روش گرادین مزدوج و دستگاه متناظر با $\alpha I + S$ را با مانده مینیمال تعمیم‌یافته با شروع مجدد و با $restart = 10$ حل می‌کنیم. در هر دو روش، حداکثر تعداد تکرارهای درونی را 10^6 گرفته و تکرارها را تا جایی ادامه می‌دهیم که نرم مانده جواب به‌نسبت $\{10^{-7}, 10^{-8}, \dots, 10^{-12}\}$ کاهش یابد، که در آن k نشان دهنده

جدول ۲-۱۰ نتایج عددی برای مثال ۲.۱۰.

نتایج به‌ازای پارامتر بهینه روش HSS		
$n = N^2$	α_{opt}	$iters$
۹۰۰	۰.۵۸۰	۸۳
۱۶۰۰	۰.۴۴۰	۹۷
۲۵۰۰	۰.۳۷۰	۱۲۷

تکرار است. سایر مفروضات از قبیل شرط توقف تکرار خارجی و حدس اولیه را همانند مثال ۱.۱۰ می‌گیریم. نتایج عددی در جدول ۲-۱۰ ارائه شده‌اند. پارامتر بهینه را پارامتر بهینه روش HSS می‌گیریم. با مقایسه نتایج جدول ۱-۱۰ و ۲-۱۰ می‌بینیم که از لحاظ تعداد تکرار تفاوت چندانی بین روش HSS و IHSS وجود ندارد.

۳-۱۰ پیش‌شرط‌ساز HSS

همان‌گونه، که دیدیم اگر $A = M - N$ یک شکافت برای ماتریس نامنفرد A باشد و روند تکراری

$$x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b,$$

به جواب دستگاه $Ax = b$ همگرا باشد، یعنی $\rho(M^{-1}N) < 1$ ، آنگاه ماتریس M یک پیش‌شرط‌ساز مناسب برای سرعت بخشیدن همگرایی یک روش تکراری مبتنی بر زیرفضای کرایلف مثل GMRES برای حل دستگاه $Ax = b$ خواهد بود. به عبارت دیگر، مقادیر ویژه $M^{-1}A$ در دایره‌ای به شعاع ۱ و مرکز $(1, 0)$ قرار دارد و سرعت همگرایی روش تکراری GMRES برای حل دستگاه

$$M^{-1}Ax = M^{-1}b,$$

یا

$$AM^{-1}y = b, \quad x = M^{-1}y,$$

مناسب خواهد بود. لازم به ذکر است که دو ماتریس $M^{-1}A$ و AM^{-1} متشابه هستند و از این رو، مجموعه مقادیر ویژه آنها یکسان است.

در این بخش، پیش‌شرط‌سازی که توسط روش تکراری HSS تولید می‌شود را به دست می‌آوریم. داریم

$$G(\alpha) = (\alpha I + S)^{-1}(\alpha I - H)(\alpha I + H)^{-1}(\alpha I - S)$$

$$\begin{aligned} &= (\alpha I + S)^{-1}(\alpha I + H)^{-1}(\alpha I - H)(\alpha I - S) \\ &= ((\alpha I + H)(\alpha I + S))^{-1}(\alpha I - H)(\alpha I - S) \\ &= \left(\frac{1}{2\alpha}(\alpha I + H)(\alpha I + S) \right)^{-1} \frac{1}{2\alpha}(\alpha I - H)(\alpha I - S). \end{aligned}$$

بنابراین، با فرض

$$M(\alpha) = \frac{1}{2\alpha}(\alpha I + H)(\alpha I + S), \quad N(\alpha) = \frac{1}{2\alpha}(\alpha I - H)(\alpha I - S),$$

خواهیم داشت

$$G(\alpha) = (M(\alpha))^{-1}N(\alpha).$$

همچنین، به سادگی می‌توان دید که

$$A = M(\alpha) - N(\alpha). \quad (31-10)$$

لذا این رابطه، شکافت مربوط به روش تکراری HSS را معرفی می‌کند. در نتیجه، می‌توان از ماتریس $M(\alpha)$ به عنوان یک پیش شرط ساز برای دستگاه $Ax = b$ استفاده کرد.

۱۰-۴ برای مطالعه بیشتر

پس از ارائه روش تکراری HSS، این روش به طور گسترده مورد توجه ریاضی دانان قرار گرفت. برخی از آنها در خصوص بهبود روش و برخی در زمینه کاربرد آن برای مسائل مختلف تمرکز کرده‌اند. روش تکراری شکافت نرمال و هرمیتی کج (NSS) توسط بای و همکارانش در [۱۳] ارائه شده است (تمرین ۵.۱۰ را ببینید). سپس، بای و همکارانش در [۱۴] روش شکافت معین مثبت و هرمیتی کج (PSS) را ارائه کردند (تمرین ۶.۱۰ را ببینید). اخیراً یک تعمیم از همه روش‌های بالا در [۶۳] با عنوان روش تکراری شکافت نیمه معین مثبت و نیمه معین مثبت (PPS) ارائه شده است. در [۱۲] بای و همکارانش روش HSS اصلاح شده را برای حل دستگاه مختلط متقارن $(W + iT)x = b$ ارائه کردند که در آن ماتریس‌های W و T نیمه معین مثبت متقارن هستند به گونه‌ای دست کم یکی از آنها معین مثبت هستند. سپس، نسخه‌های پیش شرط سازی شده روش MHSS در [۹، ۸] ارائه شده‌اند. بنزی و گلوب در [۱۶] پیش شرط ساز HSS را برای مسئله نقطه زینی به کار بردند که در فصل بعد ارائه می‌شود. با استناد به این مقاله، پژوهش‌های متعددی توسط محققین انجام شده است.

تمرینات

۱.۱۰ دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

نشان دهید که روند تکراری HSS بدون هیچ شرطی برای حل این دستگاه همگرا است. مقدار بهینه پارامتر α را بیابید.

۲.۱۰ نشان دهید ماتریس $\tilde{M}(\alpha)$ تعریف‌شده در رابطه (۱۰-۱۱) نرمال است.

۳.۱۰ دستگاه $(W + iT)x = b$ را در نظر بگیرید که در آن $W, T \in \mathbb{R}^{n \times n}$. ثابت کنید اگر ماتریس W معین مثبت متقارن و ماتریس T نیمه‌معین مثبت متقارن باشند، آنگاه روش تکراری HSS برای حل این دستگاه بدون هیچ شرطی همگراست. روش تکراری HSS را برای حل این دستگاه بنویسید. راهنمایی: مرجع [۱۲] را ببینید.

۴.۱۰ دستگاه

$$(W + iT)x = b, \quad (32-10)$$

را در نظر بگیرید که در آن $W, T \in \mathbb{R}^{n \times n}$. نشان دهید که روند تکراری

$$\begin{cases} (\alpha I + W)x^{(k+1)} = (\alpha I - iT)x^{(k)} + b, \\ (\alpha I + T)x^{(k+1)} = (\alpha I + iW)x^{(k+1)} - ib, \end{cases} \quad k = 0, 1, 2, \dots$$

یک روند تکراری دومرحله‌ای برای حل دستگاه (۳۲-۱۰) است. نشان دهید که این روند تکراری را می‌توان به صورت روند تکراری تک‌مرحله‌ای

$$x^{(k+1)} = G(\alpha)x^{(k)} + c(\alpha),$$

نوشت که در آن

$$G(\alpha) = (\alpha I + T)^{-1}(\alpha I + iW)(\alpha I + W)^{-1}(\alpha I - iT),$$

$$c(\alpha) = (I - i)\alpha(\alpha I + T)^{-1}(\alpha I + W)^{-1}b.$$

ثابت کنید

$$\rho(G(\alpha)) \leq \delta(\alpha) = \max_{\lambda \in \sigma(W)} \frac{\sqrt{\alpha^2 + \lambda^2}}{\alpha + \lambda}.$$

نشان دهید این روش به ازای هر $\alpha > 0$ به جواب این دستگاه همگراست و مقدار کمینه $\delta(\alpha)$ به ازای $\alpha^* = \sqrt{\lambda_{\min} \lambda_{\max}}$ به دست می آید، که در آن $\lambda_{\min}$ و $\lambda_{\max}$ به ترتیب کوچک ترین و بزرگ ترین مقدار ویژه W هستند. به علاوه، نشان دهید که روند تکراری بالا توسط شکافت $A = M(\alpha) - N(\alpha)$ تولید می شود که در آن

$$M(\alpha) = \frac{1+i}{2\alpha}(\alpha I + W)(\alpha I + T),$$

$$N(\alpha) = \frac{1+i}{2\alpha}(\alpha I + iW)(\alpha I - iT).$$

راهنمایی: مرجع [۱۲] را ببینید.

۵.۱۰ دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت است. فرض کنید ماتریس A یک شکافت به صورت $A = N + S$ دارد که در آن ماتریس N نرمال و ماتریس S هرمیتی کج است. نخست، نشان دهید ماتریس N معین مثبت است. به علاوه، نشان دهید که به ازای هر $\alpha > 0$ روند تکراری زیر به ازای هر حدس اولیه دلخواه $x^{(0)}$ بدون هیچ شرطی همگراست

$$\begin{cases} (\alpha I + N)x^{(k+\frac{1}{2})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - N)x^{(k+\frac{1}{2})} + b, \end{cases} \quad k = 0, 1, 2, \dots$$

این روند تکراری به روش تکراری NSS معروف است.

۶.۱۰ (الف) دستگاه معادلات خطی $Ax = b$ را در نظر بگیرید که در آن ماتریس $A \in \mathbb{C}^{n \times n}$ معین مثبت است. فرض کنید ماتریس A یک شکافت به صورت $A = P + S$ دارد که در آن ماتریس P معین مثبت و ماتریس S هرمیتی کج است. نشان دهید که به ازای هر $\alpha > 0$ روند تکراری زیر به ازای هر حدس اولیه دلخواه $x^{(0)}$ بدون هیچ شرطی همگراست

$$\begin{cases} (\alpha I + P)x^{(k+\frac{1}{2})} = (\alpha I - S)x^{(k)} + b, \\ (\alpha I + S)x^{(k+1)} = (\alpha I - P)x^{(k+\frac{1}{2})} + b, \end{cases} \quad k = 0, 1, 2, \dots$$

این روند تکراری به روش تکراری PSS معروف است.

۳۷۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

(ب) دستگاه معادلات خطی $A\mathbf{x} = \mathbf{b}$ را در نظر بگیرید که در آن

$$A = \begin{pmatrix} G & W \\ -W & G \end{pmatrix}$$

به‌طوری‌که ماتریس G معین مثبت است و ماتریس W نیمه‌معین مثبت متقارن است. نشان دهید که روش تکراری PSS با شکافت زیر برای حل دستگاه $A\mathbf{x} = \mathbf{b}$ بدون هیچ شرطی همگراست

$$A = \begin{pmatrix} G & \circ \\ \circ & G \end{pmatrix} + \begin{pmatrix} \circ & W \\ -W & \circ \end{pmatrix}$$

مسئله نقطه زینی

یک مسئله نقطه زینی دستگاهی به صورت

$$\begin{pmatrix} A & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \quad (1-11)$$

است که در آن $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{m \times m}$, $f \in \mathbb{R}^n$, $g \in \mathbb{R}^m$ و $m \leq n$. مسئله نقطه زینی در حل بسیاری از مسائل علوم و مهندسی مثل بهینه‌سازی، حل دستگاه معادلات خطی مختلط، مکانیک سیالات و پردازش تصویر ظاهر می‌شود. در این فصل، نخست چند کاربرد مسئله را مطالعه کرده و سپس به معرفی چند روش حل و راهکارهای پیش‌شرط‌سازی دستگاه می‌پردازیم. در این فصل، برای هر دو بردار x و y از نماد $(x; y)$ برای نمایش $(x^T, y^T)^T$ استفاده می‌کنیم.

۱-۱۱ برخی از کاربردهای مسئله نقطه زینی

در این بخش به برخی از کاربردهای مسائل نقطه زینی اشاره می‌کنیم.

۱-۱-۱۱ مسائل بهینه‌سازی نقطه زینی

مسئله برنامه‌ریزی درجه دوم مقید

$$\begin{aligned} \min_{x \in \mathbb{R}^n} f(x) &= \frac{1}{2} x^T A x - c^T x, \\ \text{s.t.} \quad B^T x &= d, \end{aligned} \quad (2-11)$$

۳۷۲ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تُنک

را در نظر بگیرید که در آن $A \in \mathbb{R}^{n \times n}$ یک ماتریس معین مثبت متقارن، $B \in \mathbb{R}^{n \times m}$ یک ماتریس با رتبه کامل سطری است، $c \in \mathbb{R}^n$ و $d \in \mathbb{R}^m$. با تعریف تابع لاگرانژ مربوط به این مسئله خواهیم داشت [۸۴]

$$\begin{aligned}\mathcal{L}(x, y) &= f(x) + y^T (B^T x - d) \\ &= \frac{1}{2} x^T A x + x^T B y - x^T c - y^T d \\ &= \frac{1}{2} \begin{pmatrix} x^T & y^T \end{pmatrix} \begin{pmatrix} A & B \\ B^T & \circ \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x^T & y^T \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} \quad (3-11) \\ &= \frac{1}{2} \mathbf{x}^T \mathcal{A} \mathbf{x} - \mathbf{x}^T \mathbf{e} := \mathcal{G}(\mathbf{x}),\end{aligned}$$

که در آن

$$\mathcal{A} = \begin{pmatrix} A & B \\ B^T & \circ \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} c \\ d \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

با محاسبه تابع گرادیان $\mathcal{G}$ نسبت به $\mathbf{x}$ ، داریم

$$\nabla \mathcal{G} = \mathcal{A} \mathbf{x} - \mathbf{e}.$$

بنابراین، نقاط بحرانی $\mathcal{G}$ از حل دستگاه

$$\mathcal{A} \mathbf{x} = \mathbf{e}, \quad (4-11)$$

به دست می‌آیند. اگر $\mathbf{x}^*$ جواب دستگاه (۴-۱۱) باشد، آنگاه

$$\begin{aligned}\mathcal{G}(\mathbf{x}^*) &= \frac{1}{2} (\mathbf{x}^*)^T \mathcal{A} \mathbf{x}^* - (\mathbf{x}^*)^T \mathbf{e} \\ &= -\frac{1}{2} (\mathbf{x}^*)^T \mathcal{A} \mathbf{x}^*.\end{aligned} \quad (5-11)$$

در نتیجه، با استفاده از رابطه‌های (۳-۱۱) و (۵-۱۱)، به ازای هر $\mathbf{x} = (x; y)$ ، داریم

$$\begin{aligned}\mathcal{G}(\mathbf{x}) - \mathcal{G}(\mathbf{x}^*) &= \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \mathcal{A} (\mathbf{x} - \mathbf{x}^*) \\ &= \frac{1}{2} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x^* \\ y^* \end{pmatrix} \right)^T \begin{pmatrix} A & B^T \\ B & \circ \end{pmatrix} \left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x^* \\ y^* \end{pmatrix} \right).\end{aligned}$$

از این رابطه نتیجه می‌گیریم که اگر $y = y^*$ ، آنگاه

$$\mathcal{L}(x, y^*) - \mathcal{L}(x^*, y^*) = \frac{1}{2}(x - x^*)^T A(x - x^*) \geq 0,$$

و اگر $x = x^*$ ، آنگاه

$$\mathcal{L}(x^*, y) - \mathcal{L}(x^*, y^*) = 0.$$

بنابراین، برای هر $x \in \mathbb{R}^n$ و $y \in \mathbb{R}^m$ داریم

$$\mathcal{L}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*).$$

از این رو، هر جواب دستگاه (۱۱-۴) به صورت $(x^*; y^*)$ یک نقطه زینی برای تابع لاگرانژ $\mathcal{L}(x, y)$ است.

۱۱-۲ مسئله کمترین توان‌های دوم

فرض کنید $B \in \mathbb{R}^{n \times n}$ و $b \in \mathbb{R}^n$. مسئله کمترین توان‌های دوم

$$\|Bx^* - b\|_2 = \min_{x \in \mathbb{R}^n} \|Bx - b\|_2,$$

را در نظر بگیرید. با توجه به قضیه ۲۴.۱، بردار x^* جواب این مسئله است اگر و تنها اگر دستگاه معادلات خطی نرمال

$$B^T Bx = B^T b,$$

صدق کند. این دستگاه معادل است با

$$B^T(b - Bx) = 0.$$

با فرض $r = b - Bx$ ، داریم $B^T r = 0$. در نتیجه،

$$\begin{cases} r + Bx = b, \\ B^T r = 0, \end{cases}$$

که معادل با مسئله نقطه‌زینی زیر است

$$\begin{pmatrix} I & B \\ B^T & \circ \end{pmatrix} \begin{pmatrix} r \\ x \end{pmatrix} = \begin{pmatrix} b \\ \circ \end{pmatrix}.$$

۱۱-۳ حل مسئله ناویر-استوکس به روش عناصر متناهی

مسئله ناویر-استوکس یکی از مسائل مهم و کاربردی در مکانیک سیالات است. این مسئله حرکت سیال را مدل‌سازی می‌کند. فرض کنید Ω یک زیرمجموعه باز، پیوسته و کراندار از $\mathbb{R}^2$ باشد به طوری که مرز آن، یعنی $\partial\Omega$ ، به اندازه کافی هموار می‌باشد. مسئله ناویر-استوکس یک دستگاه معادلات دیفرانسیل غیرخطی به صورت زیر است

$$\begin{cases} -\nu \Delta \mathbf{u} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \nabla p = \mathbf{f}, & \text{in } \Omega, \\ \nabla \cdot \mathbf{u} = \circ, & \text{in } \Omega, \\ \mathbf{u} = \circ, & \text{on } \partial\Omega, \\ \int_{\Omega} p dx = \circ, \end{cases} \quad (6-11)$$

را در نظر بگیرید که در آن $\nu > \circ$ و $p: \mathbb{R}^2 \rightarrow \mathbb{R}$. به علاوه، $\mathbf{f} = (f_1, f_2)$ و $\mathbf{u} = (u_1, u_2)$ ، به طوری که $f_1, f_2, u_1, u_2: \mathbb{R}^2 \rightarrow \mathbb{R}$ یادآوری می‌کنیم که

$$\begin{aligned} \Delta \mathbf{u} &= (\Delta u_1, \Delta u_2) = \left(\frac{\partial^2 u_1}{\partial x^2} + \frac{\partial^2 u_1}{\partial y^2}, \frac{\partial^2 u_2}{\partial x^2} + \frac{\partial^2 u_2}{\partial y^2} \right), \\ (\mathbf{u} \cdot \nabla) \mathbf{u} &= \left(u_1 \frac{\partial u_1}{\partial x} + u_2 \frac{\partial u_1}{\partial y}, u_1 \frac{\partial u_2}{\partial x} + u_2 \frac{\partial u_2}{\partial y} \right), \\ \nabla \cdot \mathbf{u} &= \text{div}(\mathbf{u}) = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y}. \end{aligned}$$

با توجه به جمله $(\mathbf{u} \cdot \nabla) \mathbf{u}$ در مسئله (۶-۱۱)، روشن است که این مسئله، غیرخطی است. یکی از روش‌های عددی برای حل آن، روش پیکارد است [۳۵]. در این روش، نخست یک تقریب اولیه، مثل $\mathbf{u}^{(\circ)}$ ، برای تابع $\mathbf{u}$ در نظر گرفته می‌شود. سپس، با معلوم بودن تقریب $\mathbf{u}^{(k)}$ ، روند تکراری زیر برای محاسبه

$(\mathbf{u}^{(k+1)}, p^{(k+1)})$ تعریف می‌شود

$$\left\{ \begin{array}{ll} -\nu \Delta \mathbf{u}^{(k+1)} + (\mathbf{u}^{(k)} \cdot \nabla) \mathbf{u}^{(k+1)} + \nabla p^{(k+1)} = \mathbf{f}, & \text{in } \Omega, \\ \nabla \cdot \mathbf{u}^{(k+1)} = 0, & \text{in } \Omega, \\ \mathbf{u}^{(k+1)} = 0, & \text{on } \partial\Omega, \\ \int_{\Omega} p^{(k+1)} dx = 0, & \end{array} \right. \quad k = 0, 1, 2, \dots$$

این روند تکراری تا جایی ادامه می‌یابد که $\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}\|$ و $\|p^{(k+1)} - p^{(k)}\|$ به اندازه کافی کوچک باشند. در هر تکرار از روش پیکارد مسئله اخیر حل می‌شود که به مسئله اوسین معروف است. این مسئله، یک مسئله خطی است که با روش اجزای محدود قابل حل است. از حل این مسئله به روش عناصر متناهی یا تفاضلات متناهی، یک مسئله نقطه زینی به صورت (۱-۱۱) به دست می‌آید.

در برخی از مسئله‌های کاربردی می‌توان از جمله غیرخطی $(\mathbf{u} \cdot \nabla) \mathbf{u}$ چشم‌پوشی کرد، که در این صورت مسئله استوکس به دست می‌آید. این مسئله به صورت زیر نوشته می‌شود

$$\left\{ \begin{array}{ll} -\nu \Delta \mathbf{u} + \nabla p = \mathbf{f}, & \text{in } \Omega, \\ \nabla \cdot \mathbf{u} = 0, & \text{in } \Omega, \\ \mathbf{u} = 0, & \text{on } \partial\Omega, \\ \int_{\Omega} p dx = 0, & \end{array} \right. \quad (Y-11)$$

گسسته‌سازی این معادله نیز به روش عناصر متناهی یا تفاضلات متناهی منجر به حل یک دستگاه به صورت (۱-۱۱) می‌شود.

۲-۱۱ وجود جواب برای مسئله نقطه زینی

مسئله نقطه زینی (۱-۱۱) را در نظر بگیرید. از اینجا به بعد فرض می‌کنیم ماتریس C نیمه‌معین مثبت متقارن باشد. وقتی که ماتریس A معین (یا نیمه‌معین) مثبت متقارن است، ماتریس ضرایب دستگاه (۱-۱۱) متقارن است و بنابراین، مقادیر ویژه آن حقیقی است. با این وجود ماتریس ضرایب دستگاه

نامعین است. اما دستگاه (۱۱-۱) را به صورت

$$\mathcal{A}\mathbf{x} = \begin{pmatrix} A & B^T \\ -B & C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ -g \end{pmatrix} = \mathbf{b}, \quad (۸-۱۱)$$

می‌نویسیم. در این صورت، دستگاه (۱۱-۹) خواص بسیار خوبی دارد که در لم زیر بیان می‌شود.

لم ۱۰.۱۱ فرض کنید ماتریس $H = \frac{1}{\gamma}(A + A^T)$ نیمه‌معین مثبت، ماتریس B با رتبه کامل و ماتریس C نیمه‌معین مثبت متقارن باشند. به علاوه، فرض کنید $\mathcal{N}(H) \cap \mathcal{N}(B) = \{\mathbf{0}\}$. در این صورت (الف) ماتریس $\mathcal{A}$ نامنفرد است.

(ب) ماتریس $\mathcal{A}$ حقیقی نیمه‌مثبت است، یعنی به ازای هر $\mathbf{v} \in \mathbb{R}^{n+m}$ ، $\langle \mathcal{A}\mathbf{v}, \mathbf{v} \rangle \geq 0$.

(ج) ماتریس $\mathcal{A}$ نیمه‌پایدار مثبت است، یعنی قسمت حقیقی مقادیر ویژه $\mathcal{A}$ نامنفی هستند.

(د) به علاوه، اگر ماتریس $H = \frac{1}{\gamma}(A + A^T)$ معین مثبت باشد، آنگاه ماتریس $\mathcal{A}$ پایدار مثبت است، یعنی قسمت حقیقی مقادیر ویژه $\mathcal{A}$ مثبت هستند.

برهان. برای اثبات (الف) فرض کنید $\mathbf{x} = (u; p)$ به طوری که $\mathcal{A}\mathbf{x} = \mathbf{0}$ که معادل با

$$Au + B^T p = \mathbf{0}, \quad (۹-۱۱)$$

$$-Bu + Cp = \mathbf{0}, \quad (۱۰-۱۱)$$

است. از طرفی، از $\mathcal{A}\mathbf{x} = \mathbf{0}$ داریم $\mathbf{x}^T \mathcal{A}\mathbf{x} = 0$ که این خود معادل با

$$u^T Au + p^T Cp = 0,$$

است. از اینکه ماتریس A نیمه‌معین مثبت و ماتریس C نیمه‌معین مثبت متقارن است نتیجه می‌گیریم که

$$u^T Au = p^T Cp = 0.$$

از $p^T Cp = 0$ می‌بینیم که $Cp = \mathbf{0}$ که با جایگذاری در رابطه (۱۱-۱۰) خواهیم داشت $Bu = \mathbf{0}$. همچنین، از $u^T Au = 0$ داریم

$$\frac{1}{\gamma} u^T (A + A^T) u = u^T H u = 0.$$

لذا $u \in \mathcal{N}(H) \cap \mathcal{N}(B)$ ، بنابراین، $u \in \mathcal{N}(H) \cap \mathcal{N}(B)$ که نتیجه می‌دهد $u = 0$. با جایگذاری $u = 0$ در رابطه (۹-۱۱)، داریم $B^T p = 0$. اکنون، با توجه به اینکه ماتریس B با رتبه کامل است نتیجه می‌گیریم که $p = 0$. از این رو، $x = 0$ که نامنفرد بودن A را ثابت می‌کند.

برای اثبات (ب) می‌بینیم که بخش متقارن ماتریس A به صورت

$$\mathcal{H} = \frac{1}{2}(A + A^T) = \begin{pmatrix} H & 0 \\ 0 & C \end{pmatrix},$$

است. روشن است که ماتریس $\mathcal{H}$ نیمه‌معین مثبت متقارن است. بنابراین، $\langle A\mathbf{v}, \mathbf{v} \rangle \geq 0$.

برای اثبات قسمت (ج)، فرض کنید $(\lambda, \mathbf{v})$ یک زوج ویژه A باشد به طوری که $\|\mathbf{v}\|_2 = 1$. بنابراین، $\lambda = \mathbf{v}^H A \mathbf{v}$. اگر مزدوج طرفین این رابطه را محاسبه کنیم خواهیم داشت

$$\bar{\lambda} = (\mathbf{v}^H A \mathbf{v})^H = \mathbf{v}^H A^T \mathbf{v}.$$

لذا،

$$\begin{aligned} \Re(\lambda) &= \frac{1}{2}(\lambda + \bar{\lambda}) = \frac{1}{2}\mathbf{v}^H (A + A^T) \mathbf{v} = \mathbf{v}^H \mathcal{H} \mathbf{v} \\ &= \Re(\mathbf{v})^T \mathcal{H} \Re(\mathbf{v}) + \Im(\mathbf{v})^T \mathcal{H} \Im(\mathbf{v}) \geq 0. \end{aligned}$$

برای اثبات قسمت (د)، فرض کنید $(\lambda, \mathbf{v})$ یک زوج ویژه A باشد به طوری که $\|\mathbf{v}\|_2 = 1$. با استناد به قسمت (ج)، داریم $\Re(\lambda) = \mathbf{v}^H \mathcal{H} \mathbf{v}$ و $\mathbf{v} = (u; p)$ خواهیم داشت

$$\begin{aligned} \Re(\lambda) &= \mathbf{v}^H \mathcal{H} \mathbf{v} = u^* H u + p^* C p \\ &= \Re(u)^T H \Re(u) + \Im(u)^T H \Im(u) + \Re(p)^T C \Re(p) + \Im(p)^T C \Im(p) \geq 0. \end{aligned}$$

نشان می‌دهیم که $\Re(\lambda)$ نمی‌تواند برابر با صفر باشد. به برهان خلف، اگر $\Re(\lambda)$ ، آنگاه

$$\Re(u)^T H \Re(u) = \Im(u)^T H \Im(u) = \Re(p)^T C \Re(p) = \Im(p)^T C \Im(p) = 0.$$

از اینکه H معین مثبت متقارن، این رابطه نتیجه می‌دهد که $H u = 0$ و در نتیجه، $u = 0$. اکنون، از $A u + B^T p = \lambda u$ داریم $A u = \lambda u$. با جایگذاری $u = 0$ در این رابطه نتیجه می‌گیریم که $B^T p = 0$. رتبه ستونی ماتریس B^T کامل است، لذا $p = 0$. پس $\mathbf{v} = 0$ که یک تناقض است، زیرا $\mathbf{v}$ یک بردار ویژه است. $\square$

نتیجه ۱۰.۱۱ فرض کنید ماتریس A معین مثبت و ماتریس B با رتبه کامل، و C نیمه‌معین مثبت متقارن باشند. در این صورت، با استفاده از لم ۱۰.۱۱ ماتریس A نامنفرد خواهد بود.

۳-۱۱ روش تکراری اوزاوا و پیش‌شرط‌ساز اوزاوا

اگرچه روش اوزاوا یکی از روش‌های تکراری قدیمی برای حل مسئله نقطه زینی است، این روش همچنان مورد توجه محققین قرار دارد؛ زیرا پیش‌شرط‌سازهای کارایی را ارائه می‌کنند. فرض کنید ماتریس A معین مثبت متقارن، رتبه B کامل و ماتریس C نیمه‌معین مثبت متقارن باشد. روش اوزاوا برای حل مسئله (۱۱-۸) به صورت

$$\begin{cases} u^{(k+1)} = u^{(k)} + A^{-1} (f - Au^{(k)} - B^T p^{(k)}), \\ p^{(k+1)} = p^{(k)} - \tau (g - Bu^{(k+1)} + Cp^{(k)}), \end{cases} \quad (11-11)$$

تعریف می‌شود که در آن τ یک عدد مثبت است [۲، ۲۵]. گام اول از روند تکراری اوزاوا را می‌توان به صورت

$$u^{(k+1)} = A^{-1} (f - B^T p^{(k)}),$$

نوشت. بنابراین، برای محاسبه $u^{(k+1)}$ باید دستگاه $Au = f^{(k)}$ را حل کنیم که در آن $f^{(k)} = f - B^T p^{(k)}$. با توجه به اینکه ماتریس A معین مثبت متقارن است، این دستگاه را می‌توان به طور دقیق با استفاده از تجزیه چولسکی و به طور نادقیق با استفاده از الگوریتم CG حل کرد.

با جایگذاری $u^{(k+1)}$ از رابطه اول در رابطه دوم، خواهیم داشت

$$p^{(k+1)} = p^{(k)} + \tau ((BA^{-1}f - g) - (C + BA^{-1}B^T)p^{(k)}). \quad (12-11)$$

این روند تکراری همان روش تکراری ریچاردسون برای حل دستگاه

$$(C + BA^{-1}B^T)p^{(k)} = BA^{-1}f - g,$$

است. ماتریس ضرایب این دستگاه، ماتریس شور A ، یعنی $S = C + BA^{-1}B^T$ ، است. به سادگی می‌توان دید که ماتریس S معین مثبت متقارن است (تمرین ۳.۱۱ را ببینید). روند تکراری (۱۲-۱۱) به ازای

$$0 < \tau < \frac{2}{\lambda_{\max}(S)},$$

همگرا است که در آن $\lambda_{\max}(S)$ بزرگترین مقدار ویژه S است. به علاوه، پارامتر بهینه در این بازه برابر با

$$\tau_{\text{opt}} = \frac{2}{\lambda_{\min}(S) + \lambda_{\max}(S)},$$

است.

روند تکراری (۱۱-۱۱) را می‌توان به صورت

$$\begin{pmatrix} u^{(k+1)} \\ p^{(k+1)} \end{pmatrix} = \mathcal{G}_\tau \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \mathcal{M}_\tau^{-1} b, \quad (11-13)$$

نوشت، که در آن

$$\mathcal{M}_\tau = \begin{pmatrix} A & \circ \\ -B & \frac{1}{\tau} I \end{pmatrix}, \quad \mathcal{N}_\tau = \begin{pmatrix} \circ & -B^T \\ \circ & \frac{1}{\tau} I - C \end{pmatrix},$$

و $\mathcal{G}_\tau = \mathcal{M}_\tau^{-1} \mathcal{N}_\tau$. به سادگی می‌توان دید که $\mathcal{A} = \mathcal{M}_\tau - \mathcal{N}_\tau$. بنابراین، روند تکراری اوزاوا یک روند تکراری ایستا برای حل مسئله نقطه زینی (۱۱-۸) است و ماتریس تکرار روش برابر با $\mathcal{G}_\tau = \mathcal{M}_\tau^{-1} \mathcal{N}_\tau$ می‌باشد. در نتیجه، می‌توان از ماتریس $\mathcal{M}_\tau$ به عنوان یک پیش شرط ساز برای مسئله (۱۱-۸) استفاده کرد.

۱۱-۴ پیش شرط سازهای قطری بلوکی و بالامثلثی بلوکی

با توجه به اینکه ماتریس A بزرگ و تنک است، از روش‌های تکراری برای حل مسئله (۱۱-۸) استفاده می‌شود. اما سرعت روش‌های تکراری برای حل این دستگاه کند است. لذا، برای بهبود سرعت همگرایی از پیش شرط ساز استفاده می‌شود. در ادامه، برخی از پیش شرط سازهای ارائه شده را مطالعه می‌کنیم. ایده اصلی پیش شرط سازهای قطری بلوکی و بالامثلثی بلوکی در دو مقاله [۵۱، ۷۸] ارائه شده است. در این بخش به معرفی آنها می‌پردازیم.

ابتدا فرض می‌کنیم که ماتریس A معین مثبت (نه لزوماً متقارن) و رتبه ماتریس B کامل باشد. به علاوه، فرض می‌کنیم که $C = \circ$. با توجه به نتیجه ۱.۱۱ ماتریس A نامنفرد است. به سادگی می‌توان دید که

$$A = \begin{pmatrix} A & B^T \\ -B & \circ \end{pmatrix} = \begin{pmatrix} I & \circ \\ -BA^{-1} & I \end{pmatrix} \begin{pmatrix} A & B^T \\ \circ & S \end{pmatrix} =: \mathcal{LU}, \quad (11-14)$$

۳۸۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

که در آن $S = BA^{-1}B^T$ ماتریس شور A می‌باشد. با توجه به اینکه ماتریس A نامنفرد است، نتیجه می‌گیریم که ماتریس S نیز نامنفرد است.

نخست، در تجزیه بلوکی $A = LU$ ، بلوک $(2, 1)$ در ماتریس L و بلوک $(1, 2)$ در ماتریس U را برابر با ماتریس صفر قرار می‌دهیم. در این صورت، داریم

$$A \approx \begin{pmatrix} A & \circ \\ \circ & S \end{pmatrix} = P_1. \quad (15-11)$$

درواقع، ماتریس A با یک ماتریس قطری بلوکی تقریب زده شده است. همچنین اگر بلوک $(2, 1)$ در ماتریس L را برابر با صفر قرار دهیم، آنگاه خواهیم داشت

$$A \approx \begin{pmatrix} A & B^T \\ \circ & S \end{pmatrix} = P_2. \quad (16-11)$$

در ادامه، توزیع مقادیر ویژه ماتریس‌های $P_1^{-1}A$ و $P_2^{-1}A$ را بررسی می‌کنیم. نخست، ماتریس

$$T = P_1^{-1}A,$$

را در نظر می‌گیریم. فرض کنید $(\lambda, (x; y))$ یک زوج ویژه ماتریس T باشد، یعنی

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \lambda P_1 \begin{pmatrix} x \\ y \end{pmatrix},$$

که معادل با

$$\begin{cases} B^T y = (\lambda - 1)Ax, \\ -Bx = \lambda Sy. \end{cases} \quad (17-11)$$

است. اگر $\lambda = 1$ ، آنگاه از رابطه نخست در رابطه (۱۷-۱۱) خواهیم داشت $B^T y = \circ$. چون رتبه ستونی ماتریس B^T کامل است، نتیجه می‌گیریم $y = \circ$. با جایگذاری $y = \circ$ در رابطه دوم (۱۷-۱۱) خواهیم داشت $Bx = \circ$. بنابراین، $x \in \mathcal{N}(B)$. از این رو، به ازای هر $x \in \mathcal{N}(B)$ و $x \neq \circ$ زوج $((x; \circ), 1)$ یک زوج ویژه ماتریس T است. اکنون، فرض می‌کنیم $\lambda \neq 1$. در این صورت، از رابطه نخست (۱۷-۱۱)، داریم

$$x = \frac{1}{\lambda - 1} A^{-1} B^T y.$$

با جایگذاری این رابطه در رابطه دوم (۱۷-۱۱)، خواهیم داشت

$$\frac{1}{1-\lambda} Sy = \lambda Sy.$$

اکنون، با توجه به اینکه $y \neq 0$ (چرا؟) و ماتریس S نامنفرد است، از رابطه اخیر نتیجه می‌گیریم که

$$\frac{1}{1-\lambda} = \lambda.$$

با حل این معادله، می‌بینیم که $\frac{1}{\lambda}(1 \pm \sqrt{3}i)$ دو ریشه معادله هستند. با جمع‌بندی مطالب ارائه‌شده

می‌بینیم که ماتریس T دارای سه مقدار ویژه متمایز 1 و $\frac{1}{\lambda}(1 \pm \sqrt{3}i)$ است.

نشان می‌دهیم که چندجمله‌ای مینیمال ماتریس T به صورت

$$p(t) = (t-1)(t^2 - t + 1),$$

است. برای نشان دادن این مطلب داریم

$$T = \begin{pmatrix} A^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} \begin{pmatrix} A & B^T \\ -B & 0 \end{pmatrix} = \begin{pmatrix} I & A^{-1}B^T \\ -S^{-1}B & 0 \end{pmatrix}.$$

با کمی محاسبه می‌توان دید که

$$T^2 - T = - \begin{pmatrix} Q & 0 \\ 0 & I \end{pmatrix},$$

که در آن $Q = A^{-1}B^TS^{-1}B$. می‌توان دید که $Q^2 = Q$. لذا،

$$(T^2 - T)^2 = T^2 - T,$$

که معادل با

$$T(T-I)(T^2 - T + I) = 0,$$

است. اینک، از اینکه ماتریس T نامنفرد است، نتیجه می‌گیریم که $(T-I)(T^2 - T + I) = 0$.

بنابراین، $p(t)$ چندجمله‌ای مینیمال ماتریس T است. لازم به ذکر است که ماتریس T قطری‌شدنی نیز

می‌باشد (چرا؟). چندجمله‌ای مینیمال T از درجه ۳ است. بنابراین، یک روش تکراری مبتنی بر زیرفضای

کرایلف مثل نسخه کامل GMRES حداکثر در ۳ تکرار همگرا می‌شود.

در ادامه، خواص ماتریس $\mathcal{G} = \mathcal{P}_\tau^{-1} \mathcal{A}$ را بررسی می‌کنیم. اگر $(\lambda, (x; y))$ یک زوج ویژه ماتریس $\mathcal{G}$ باشد، یعنی

$$\mathcal{A} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \mathcal{P}_\tau \begin{pmatrix} x \\ y \end{pmatrix},$$

که معادل با

$$\begin{cases} (1 - \lambda)B^T y = (\lambda - 1)Ax, \\ -Bx = \lambda Sy, \end{cases} \quad (18-11)$$

است. به‌ازای $\lambda = 1$ رابطه نخست از (18-11) برقرار است. از رابطه دوم (18-11) داریم $y = -S^{-1}Bx$. بنابراین، به‌ازای هر $x \notin \mathcal{N}(B)$ ، $(x; -S^{-1}Bx)$ یک بردار ویژه متناظر با مقدار ویژه ۱ است. نشان می‌دهیم که ماتریس $\mathcal{G}$ مقدار ویژه دیگری ندارد. به‌برهان خلف، فرض کنید $\lambda \neq 1$ مقدار ویژه دیگری برای $\mathcal{G}$ باشد. در این صورت، از رابطه اول (18-11) خواهیم داشت $x = -A^{-1}B^T y$. با جایگذاری x در رابطه دوم (18-11) خواهیم داشت $(1 - \lambda)Sy = 0$. بنابراین، $y = 0$ با جایگذاری $y = 0$ در رابطه اول از (18-11) خواهیم داشت $x = 0$. بنابراین، بردار ویژه متناظر با λ برابر با صفر خواهد بود که یک تناقض است. به‌راحتی می‌توان دید که

$$\mathcal{P}_\tau^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}B^T S^{-1} \\ 0 & S^{-1} \end{pmatrix},$$

و

$$\mathcal{G} = \mathcal{P}_\tau^{-1} \mathcal{A} = \begin{pmatrix} I + A^{-1}B^T S^{-1}B & A^{-1}B^T \\ -S^{-1}B & 0 \end{pmatrix}.$$

با استفاده از این رابطه، با کمی محاسبه می‌توان نشان داد که

$$(\mathcal{G} - I)^2 = 0.$$

بنابراین، چندجمله‌ای مینیمال $\mathcal{G}$ به‌صورت $q(t) = (t - 1)^2$ است. لذا، نسخه کامل الگوریتم GMRES برای حل دستگاه $\mathcal{G}^{-1}Ax = \mathcal{G}^{-1}b$ در دو تکرار همگرا می‌شود. لازم به ذکر است که ماتریس $\mathcal{G}$ قطری‌شدنی نیست (چرا؟).

اگر علامت بلوک $(2, 2)$ در ماتریس‌های P_1 و P_2 را تغییر دهیم پیش‌شرط‌سازهای مفیدی به دست می‌آیند. این دو پیش‌شرط‌ساز را با P_3 و P_4 نشان می‌دهیم که به صورت

$$P_3 = \begin{pmatrix} A & \circ \\ \circ & -S \end{pmatrix}, \quad P_4 = \begin{pmatrix} A & B^T \\ \circ & -S \end{pmatrix},$$

می‌باشند. همانند پیش‌شرط‌ساز P_1 می‌توان دید که $P_3^{-1}A$ دارای سه مقدار ویژه متمایز 1 و $(1 \pm \sqrt{5})/4$ است، چندجمله‌ای مینیمال ماتریس برابر با $q(t) = (t-1)(t^2 - t - 1)$ است و قطری‌شدنی است. همچنین همانند P_2 می‌توان دید که ماتریس $P_4^{-1}A$ دارای دو مقدار ویژه $1 \pm \sqrt{5}$ است و چندجمله‌ای مینیمال آن به صورت $q(t) = t^2 - 1$ است. به علاوه، ماتریس $P_4^{-1}A$ قطری‌شدنی است. نکته قابل ذکر در اینجا این است که توزیع مقادیر ویژه ماتریس $P_i^{-1}A$ و AP_i^{-1} یکسان است، زیرا با توجه به

$$P_i^{-1}A = P_i^{-1}(AP_i^{-1})P_i,$$

این دو ماتریس متشابه هستند و مقادیر ویژه یکسانی دارند. بنابراین، زمانی که یک روش تکراری مبتنی بر زیرفضای کرایلف مثل GMRES برای حل دستگاه پیش‌شرط‌سازی شده به کار می‌رود، تفاوت چندانی بین پیش‌شرط‌سازی از چپ یا راست وجود ندارد. اگرچه پیش‌شرط‌سازی از راست مزایایی بر پیش‌شرط‌سازی از چپ دارد که پیش‌تر اشاره شد.

وقتی که یک روش مبتنی بر زیرفضای کرایلف برای حل دستگاه پیش‌شرط‌سازی استفاده می‌شود، در تشکیل پایه یکامتعامل برای فضای کرایلف نیاز به محاسبه بردارهایی به صورت $\mathbf{v} = P_i^{-1}A\mathbf{u}$ (در پیش‌شرط‌سازی از چپ) یا $\mathbf{v} = AP_i^{-1}\mathbf{u}$ (در پیش‌شرط‌سازی از راست) است. در اینجا نحوه محاسبه $\mathbf{v} = AP_i^{-1}\mathbf{u}$ را توضیح می‌دهیم و محاسبه $\mathbf{v} = P_i^{-1}A\mathbf{u}$ به طور مشابه انجام می‌شود. با فرض $\mathbf{w} = P_i^{-1}\mathbf{u}$ داریم $\mathbf{v} = A\mathbf{w}$. در این صورت، کفایت ابتدا دستگاه $P_i\mathbf{w} = \mathbf{u}$ را حل کرده و سپس بردار $\mathbf{v}$ را با استفاده از $\mathbf{v} = A\mathbf{w}$ محاسبه کنیم.

فرض کنید $\mathbf{u} = (u_1; u_2)$ ، $\mathbf{w} = (w_1; w_2)$ و $\mathbf{v} = (v_1; v_2)$ که در آن $u_1, v_1, w_1 \in \mathbb{R}^n$ و $u_2, v_2, w_2 \in \mathbb{R}^m$. برای ماتریس پیش‌شرط‌ساز P_1 بردارهای w_1 و w_2 با حل دستگاه‌های $Aw_1 = u_1$ و $Sw_2 = u_2$ بدست می‌آیند. بعد از محاسبه بردارهای w_1 و w_2 ، قرار می‌دهیم $\mathbf{v} = A\mathbf{w}$.

به‌طور مشابه، برای پیش‌شرط‌ساز P_2 ، نخست دستگاه $Sw_2 = u_2$ و سپس دستگاه

$$Aw_1 = u_1 - B^T w_2,$$

را حل کرده و در نهایت قرار می‌دهیم $v = Aw$. به‌طور مشابه، به‌کارگیری پیش‌شرط‌سازهای P_3 و P_4 را می‌توان ارائه کرد.

همان‌گونه که دیدیم در به‌کارگیری هرکدام از پیش‌شرط‌سازهای ارائه‌شده، دو دستگاه با ماتریس ضرایب S و A حل می‌شوند. اگر ماتریس‌های A و S معین مثبت متقارن باشند، آنگاه می‌توان دستگاه متناظر را با استفاده از تجزیه چولسکی حل کرد و در صورتی که معین مثبت (نامتقارن) باشند می‌توان از تجزیه LU استفاده کرد.

در صورتی که A و S معین مثبت متقارن باشند، می‌توان دستگاه متناظر را به‌صورت تقریبی و با استفاده از یک روش تکراری مثل CG حل کرد و در صورتی که معین مثبت باشند با استفاده از یک روش تکراری مثل GMRES حل کرد. اگر این دستگاه‌ها با دقت زیاد حل نشوند، دستگاه اصلی $AP_i^{-1}y = b$ بایستی با استفاده از الگوریتم FGMRES حل شود.

همان‌گونه که دیدیم در به‌کارگیری هریک از پیش‌شرط‌سازهای ارائه‌شده، باید یک دستگاه با ماتریس شور $S = BA^{-1}B^T$ و یک دستگاه با ماتریس ضرایب A حل شوند. در صورتی که ماتریس A معین مثبت متقارن باشد، می‌توان از تجزیه چولسکی ماتریس A برای حل دستگاه متناظر استفاده کرد و در غیراین صورت، از تجزیه LU بهره گرفت. همچنین، می‌توان از یک روش تکراری نیز برای حل دستگاه با ماتریس ضرایب A استفاده کرد: درحالتی که ماتریس A معین مثبت متقارن باشد، از روش گرادیان مزدوج و در غیراین صورت، از الگوریتم GMRES استفاده می‌شود. مشکل اصلی در به‌کارگیری پیش‌شرط‌سازهای P_i ، $i = 1, 2, 3, 4$ ، حل دستگاه با ماتریس شور $S = BA^{-1}B^T$ است. محاسبه معکوس ماتریس A و سپس محاسبه ضرب سه ماتریس بسیار پرهزینه است. از طرفی، ماتریس‌های A و B تُنک هستند، اما ماتریس S لزوماً تُنک نیست. بنابراین، در حل یک دستگاه ماتریس S با استفاده از یک روش مستقیم، هزینه زیادی متحمل خواهیم شد.

یکی از راه‌های فائق آمدن بر مشکلات ارائه‌شده این است که ماتریس‌های A و S را به‌ترتیب با تقریب‌هایی مثل $\hat{A}$ و $\hat{S}$ جایگزین کنیم، به‌گونه‌ای که حل دستگاه با ماتریس‌های $\hat{A}$ و $\hat{S}$ ساده باشد. باید توجه کرد که در این صورت، ممکن است از کیفیت پیش‌شرط‌سازها به‌طور قابل توجهی کم شود.

متداولترین روش در به‌کارگیری P_i ها این است که ماتریس A با یک تجزیه ناقص آن جایگزین شود. برای مثال اگر ماتریس A معین مثبت متقارن باشد، قرار می‌دهیم

$$\hat{A} = \hat{L}\hat{L}^T \approx A,$$

و

$$\hat{S} = B\hat{A}^{-1}B^T = B(\hat{L}\hat{L}^T)^{-1}B^T.$$

به‌همین ترتیب، اگر ماتریس A معین مثبت متقارن نباشد، می‌توان از تجزیه ناقص LU به‌عنوان یک تقریب برای ماتریس A استفاده کرد. در این صورت، دستگاه متناظر با $\hat{A} = \hat{L}\hat{L}^T$ به‌طور مستقیم و دستگاه متناظر با ماتریس ضرایب $\hat{S}$ توسط یک روش تکراری حل می‌شود. این دستگاه را می‌توان به‌صورت تقریبی با استفاده از روش گرادیان مزدوج حل کرد. در هر تکرار از این روش، دو ضرب ماتریس در بردار با ماتریس‌های B و B^T ، و حل دستگاه مثلثی با ماتریس‌های L و L^T خواهیم داشت.

به‌طور مشابه وقتی که $C \neq 0$ ، پیش‌شرط‌سازهای ارائه‌شده را می‌توان اصلاح کرد، به این ترتیب که تنها ماتریس مکمل شور به‌صورت $S = C + BA^{-1}B^T$ تغییر می‌کند. با توجه به مشکلات ذکرشده، تاکنون پیش‌شرط‌سازهای متنوعی برای مسایل نقطه زینی ارائه شده است. از مهم‌ترین مراجع برای مرور این پیش‌شرط‌سازها مقالات و کتاب‌های [۱۰، ۱۷، ۸۴] هستند. در ادامه چند پیش‌شرط‌ساز دیگر را معرفی می‌کنیم.

۱۱-۵ پیش‌شرط‌ساز HSS

دستگاه (۱۱-۸) را در نظر بگیرید که در آن ماتریس A معین مثبت (نه لزوماً متقارن) و C نیمه‌معین مثبت متقارن است. ماتریس A را به‌صورت $A = \mathcal{H} + \mathcal{S}$ می‌نویسیم که در آن

$$\mathcal{H} = \frac{1}{2}(A + A^T), \quad \mathcal{S} = \frac{1}{2}(A - A^T).$$

در این صورت، روش تکراری HSS برای حل دستگاه $A\mathbf{x} = \mathbf{b}$ را می‌توان به‌صورت

$$\begin{cases} (\alpha\mathcal{I} + \mathcal{H})\mathbf{x}^{(k+\frac{1}{2})} = (\alpha\mathcal{I} - \mathcal{S})\mathbf{x}^{(k)} + \mathbf{b}, \\ (\alpha\mathcal{I} + \mathcal{S})\mathbf{x}^{(k+1)} = (\alpha\mathcal{I} - \mathcal{H})\mathbf{x}^{(k+\frac{1}{2})} + \mathbf{b}, \end{cases} \quad (11-19)$$

نوشت که در آن $\alpha > 0$ و $\mathcal{I}$ ماتریس همانی از $(m+n) \times (m+n)$ است. در اینجا توجه کنید که ماتریس A حقیقی، و در نتیجه، ماتریس‌های $\mathcal{H}$ و $\mathcal{S}$ حقیقی هستند. روشن است که ماتریس‌های $\mathcal{H}$ و $\mathcal{S}$

به صورت

$$\mathcal{H} = \begin{pmatrix} H & \circ \\ \circ & C \end{pmatrix}, \quad S = \begin{pmatrix} S & B^T \\ -B & \circ \end{pmatrix},$$

هستند که در آن $H = \frac{1}{\gamma}(A + A^T)$ و $S = \frac{1}{\gamma}(A - A^T)$. همان‌گونه که دیدیم اگر ماتریس $\mathcal{H}$ معین مثبت باشد، آنگاه روش تکراری HSS بدون هیچ شرطی همگراست. ماتریس $\mathcal{H}$ معین مثبت است اگر و تنها اگر ماتریس‌های H و C معین مثبت متقارن باشند. اما این اتفاق در مسائل عملی به ندرت رخ می‌دهد. در ادامه، خواهیم دید که تحت شرایطی ملایم‌تر، روش تکراری (۱۱-۱۹) همگرا خواهد بود.

نیم‌گام اول از روند تکراری (۱۱-۱۹) را می‌توان به صورت

$$\begin{cases} (\alpha I + H)u^{(k+\frac{1}{\gamma})} = \alpha u^{(k)} - Su^{(k)} + f - B^T p^{(k)}, \\ (\alpha I + C)p^{(k+\frac{1}{\gamma})} = \alpha p^{(k)} - g + Bu^k, \end{cases} \quad (۱۱-۲۰)$$

نوشت. هر دو ماتریس $\alpha I + H$ و $\alpha I + C$ معین مثبت متقارن هستند. بنابراین، هر دو دستگاه در رابطه (۱۱-۲۰) را می‌توان به صورت دقیق با استفاده از تجزیه چولسکی یا به صورت نادقیق با استفاده از الگوریتم CG حل کرد. نیم‌گام دوم از روند تکراری (۱۱-۱۹) را می‌توان به صورت

$$\begin{cases} (\alpha I + S)u^{(k+1)} + B^T p^{(k+1)} = (\alpha I - H)u^{(k+\frac{1}{\gamma})} + f =: f^{(k)}, \\ -Bu^{(k+1)} + \alpha p^{(k+1)} = (\alpha I - C)p^{(k+\frac{1}{\gamma})} - g =: g^{(k)}, \end{cases} \quad (۱۱-۲۱)$$

نوشت. روشن است که حل این دو دستگاه به سادگی حل دو دستگاه در (۱۱-۲۰) نیست. با محاسبه $u^{(k+1)}$ از رابطه اول و جایگذاری در رابطه دوم (۱۱-۲۱)، خواهیم داشت

$$\left(\alpha^2 I + B(I + \frac{1}{\alpha}S)^{-1}B^T \right) p^{(k+1)} = B(I + \frac{1}{\alpha}S)^{-1}f^{(k)} + \alpha g^{(k)}. \quad (۱۱-۲۲)$$

می‌توان این دستگاه را به طور دقیق با استفاده از تجزیه LU یا به صورت نادقیق با استفاده از الگوریتم GMRES حل کرد. با محاسبه بردار $p^{(k+1)}$ ، بردار $u^{(k+1)}$ از رابطه

$$u^{(k+1)} = (\alpha I + S)^{-1}(f^{(k)} - B^T p^{(k+1)}),$$

به دست می‌آید. اگر $S = \circ$ ، آنگاه دستگاه (۱۱-۲۲) به صورت

$$(\alpha^2 I + BB^T)p^{(k+1)} = Bf^{(k)} + \alpha g^{(k)},$$

نوشته می‌شود. بعد از حل این دستگاه، بردار $u^{(k+1)}$ از رابطه

$$u^{(k+1)} = \frac{1}{\alpha}(f^{(k)} - B^T p^{(k+1)}),$$

به دست می‌آید. لازم به ذکر است که ماتریس $\alpha I + BB^T$ معین مثبت متقارن و دستگاه مربوطه را می‌توان با استفاده از تجزیه چولسکی یا الگوریتم CG حل کرد.

در ادامه، در خصوص همگرایی روند تکراری (۱۱-۱۹) بحث می‌کنیم. با محاسبه بردار $x^{(k+\frac{1}{2})}$ از رابطه اول در (۱۱-۱۹) و جایگذاری در رابطه دوم خواهیم داشت

$$\mathbf{x}^{(k+1)} = \mathcal{T}_\alpha \mathbf{x}^{(k)} + \mathbf{c}, \quad (11-23)$$

که در آن ماتریس تکرار روش به صورت

$$\mathcal{T}_\alpha = (\alpha I + \mathcal{S})^{-1}(\alpha I - \mathcal{H})(\alpha I + \mathcal{H})^{-1}(\alpha I - \mathcal{S}),$$

است و

$$\mathbf{c} = (\alpha I + \mathcal{S})^{-1}(\mathcal{I} + (\alpha I - \mathcal{H})(\alpha I + \mathcal{H})^{-1})\mathbf{b}.$$

قضیه ۱.۱۱ فرض کنید ماتریس A معین مثبت باشد، ماتریس C نیمه معین مثبت متقارن و ماتریس B با رتبه کامل باشند. در این صورت، روش تکراری (۱۱-۱۹) برای حل دستگاه (۱۱-۸) بدون هیچ شرطی همگرا است، یعنی به ازای هر $\alpha > 0$ ، $\rho(\mathcal{T}_\alpha) < 1$.

برهان. فرض کنید

$$\hat{\mathcal{T}}_\alpha = (\alpha I + \mathcal{S})\mathcal{T}_\alpha(\alpha I + \mathcal{S})^{-1}.$$

بنابراین، ماتریس $\mathcal{T}_\alpha$ متشابه با ماتریس

$$\hat{\mathcal{T}}_\alpha = (\alpha I - \mathcal{H})(\alpha I + \mathcal{H})^{-1}(\alpha I - \mathcal{S})(\alpha I + \mathcal{S})^{-1} = \mathcal{R}\mathcal{U},$$

است، که در آن ماتریس $\mathcal{R} = (\alpha I - \mathcal{H})(\alpha I + \mathcal{H})^{-1}$ متقارن است و ماتریس $\mathcal{U} = (\alpha I - \mathcal{S})(\alpha I + \mathcal{S})^{-1}$ متعامد می‌باشد. به سادگی می‌توان دید که ماتریس $\mathcal{R}$ را می‌توان

به صورت

$$\mathcal{R} = \begin{pmatrix} R_{11} & \circ \\ \circ & R_{22} \end{pmatrix},$$

نوشت، که در آن

$$R_{11} = (\alpha I - H)(\alpha I + H)^{-1}, \quad R_{22} = (\alpha I - C)(\alpha I + C)^{-1}. \quad (24-11)$$

هر دو ماتریس R_{22} و R_{11} متقارن هستند. لذا ماتریس‌های متعامد $V_{11} \in \mathbb{R}^{n \times n}$ و $V_{22} \in \mathbb{R}^{m \times m}$ و ماتریس‌های قطری $D_1 \in \mathbb{R}^{n \times n}$ و $D_2 \in \mathbb{R}^{m \times m}$ وجود دارند به‌طوری‌که

$$V_{11}^T R_{11} V_{11} = D_1, \quad V_{22}^T R_{22} V_{22} = D_2,$$

که در آن درایه‌های قطری D_1 ، مقادیر ویژه R_{11} و درایه‌های قطری D_2 ، مقادیر ویژه R_{22} هستند. با فرض

$$\mathcal{V} = \begin{pmatrix} V_{11} & \circ \\ \circ & V_{22} \end{pmatrix}, \quad \mathcal{D} = \begin{pmatrix} D_1 & \circ \\ \circ & D_2 \end{pmatrix},$$

خواهیم داشت

$$\mathcal{V}^T \mathcal{R} \mathcal{V} = \mathcal{D} = \begin{pmatrix} D_1 & \circ \\ \circ & D_2 \end{pmatrix}.$$

فرض کنید μ_i ، $i = 1, \dots, n$ مقادیر ویژه H و ν_i ، $i = 1, \dots, m$ مقادیر ویژه C باشند. روشن است که به‌ازای هر i ، $\mu_i > 0$ و $\nu_i \geq 0$. ماتریس $\mathcal{D}$ را می‌توان به‌صورت

$$\mathcal{D} = \begin{pmatrix} \frac{\alpha - \mu_1}{\alpha + \mu_1} & & & & & & \\ & \frac{\alpha - \mu_2}{\alpha + \mu_2} & & & & & \\ & & \ddots & & & & \\ & & & \frac{\alpha - \mu_n}{\alpha + \mu_n} & & & \\ & & & & \frac{\alpha - \nu_1}{\alpha + \nu_1} & & \\ & & & & & \frac{\alpha - \nu_2}{\alpha + \nu_2} & \\ & & & & & & \ddots & \\ & & & & & & & \frac{\alpha - \nu_m}{\alpha + \nu_m} \end{pmatrix},$$

نوشت. توجه کنید که

$$\left| \frac{\alpha - \mu_i}{\alpha + \mu_i} \right| < 1, \quad i = 1, \dots, n, \quad \left| \frac{\alpha - \nu_i}{\alpha + \nu_i} \right| \leq 1, \quad i = 1, \dots, m. \quad (25-11)$$

روشن است که ماتریس RU به طور متعامد با ماتریس $V^T RU V$ متشابه است و داریم

$$V^T RU V = (V^T R V)(V^T U V) = DQ,$$

که در آن $Q = V^T U V$. بنابراین، ماتریس T_α با ماتریس DQ متشابه است و از این رو، داریم

$$\rho(T_\alpha) = \rho(\hat{T}_\alpha) = \rho(DQ) = \rho(QD).$$

ادعا می‌کنیم که به ازای هر $\alpha > 0$,

$$\rho(QD) < 1.$$

فرض کنید $\lambda \in \mathbb{C}$ یک مقدار ویژه QD و $\mathbf{x} \in \mathbb{C}^{n+m}$ بردار ویژه متناظر باشد به طوری که $\|\mathbf{x}\|_2 = 1$. به علاوه، فرض کنید $\lambda \neq 0$ ، درغیراین صورت، چیزی برای اثبات باقی نمی‌ماند. نشان می‌دهیم که $|\lambda| < 1$. از $QD\mathbf{x} = \lambda\mathbf{x}$ داریم $D\mathbf{x} = \lambda Q^T \mathbf{x}$ و در نتیجه،

$$\|D\mathbf{x}\|_2 = |\lambda| \|Q^T \mathbf{x}\|_2 = |\lambda| \|\mathbf{x}\|_2 = |\lambda|.$$

بنابراین، با توجه به رابطه (25-11)، داریم

$$\begin{aligned} |\lambda|^2 &= \|D\mathbf{x}\|_2^2 \\ &= \sum_{i=1}^n \left(\frac{\alpha - \mu_i}{\alpha + \mu_i} \right)^2 |\mathbf{x}_i|^2 + \sum_{i=1}^m \left(\frac{\alpha - \nu_i}{\alpha + \nu_i} \right)^2 |\mathbf{x}_{n+i}|^2 \\ &\leq \sum_{i=1}^n |\mathbf{x}_i|^2 + \sum_{i=1}^m |\mathbf{x}_{n+i}|^2 = \|\mathbf{x}\|_2^2 = 1. \end{aligned}$$

برای نشان دادن $|\lambda| < 1$ ، کافی است نشان دهیم که به ازای دستکم یک $1 \leq i \leq n$ ، داریم $\mathbf{x}_i \neq 0$. زیرا، اگر به ازای هر $1 \leq i \leq n$ ، $\mathbf{x}_i = 0$ ، آنگاه بردار ویژه $\mathbf{x}$ به صورت $(0; \hat{x})$ خواهد بود که در آن $\hat{x} \in \mathbb{C}^m$. ماتریس Q را به صورت

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix},$$

افرازبندی می‌کنیم. در این صورت، از $QD\mathbf{x} = \lambda\mathbf{x}$ داریم

$$QDx = \begin{pmatrix} Q_{11}D_1 & Q_{12}D_2 \\ Q_{21}D_1 & Q_{22}D_2 \end{pmatrix} \begin{pmatrix} 0 \\ \hat{x} \end{pmatrix} = \begin{pmatrix} Q_{12}D_2\hat{x} \\ Q_{22}D_2\hat{x} \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda\hat{x} \end{pmatrix}.$$

از این رابطه نتیجه می‌شود که

$$Q_{12}D_2\hat{x} = 0, \quad (26-11)$$

$$Q_{22}D_2\hat{x} = \lambda\hat{x}. \quad (27-11)$$

در ادامه، نشان خواهیم داد که رتبه ماتریس Q_{12} کامل است. بنابراین، با آگاهی از کامل بودن رتبه Q_{12} ، از رابطه (۲۶-۱۱) نتیجه می‌گیریم که $D_2\hat{x} = 0$ و با جایگذاری آن در رابطه (۲۷-۱۱) می‌بینیم که $\hat{x} = 0$. لذا $\mathbf{x} = 0$ که یک تناقض با ناصفر بودن بردار ویژه $\mathbf{x}$ است. اکنون، نشان می‌دهیم که رتبه ماتریس Q_{12} کامل است. داریم

$$\begin{aligned} \mathcal{U} &= (\alpha I - S)(\alpha I + S)^{-1} \\ &= \begin{pmatrix} \alpha I - S & -B^T \\ B & \alpha I \end{pmatrix} \begin{pmatrix} \alpha I + S & B^T \\ -B & \alpha I \end{pmatrix}^{-1} \\ &=: \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}. \end{aligned}$$

با استفاده از تمرین (۱۰.۱۱) و با کمی محاسبه می‌توان دید که

$$U_{12} = - \left(I + (\alpha I - S)(\alpha I + S)^{-1} \right) B^T \left(\alpha I + B(\alpha I + S)^{-1} B^T \right)^{-1}. \quad (28-11)$$

ابتدا، نشان می‌دهیم که ماتریس

$$L = I + (\alpha I - S)(\alpha I + S)^{-1},$$

نامنفرد است. به‌برهان خلف، اگر منفرد باشد، آنگاه یک بردار ناصفر مثل z وجود دارد به‌طوری‌که $Lz = 0$. بنابراین، $(\alpha I - S)(\alpha I + S)^{-1}z = -z$. با فرض $0 \neq z = (\alpha I + S)^{-1}w$ ، این رابطه را می‌توان به‌صورت

$$(\alpha I - S)w = -(\alpha I + S)w,$$

نوشت. از این رابطه نتیجه می‌گیریم که $w = 0$ ، که یک تناقض است. از طرفی، ماتریس

$$Y = \alpha I + B(\alpha I + S)^{-1} B^T,$$

معین مثبت است؛ زیرا داریم

$$\begin{aligned} Y + Y^T &= 2\alpha I + B \left((\alpha I + S)^{-1} + (\alpha I + S)^{-T} \right) B^T \\ &= 2\alpha I + B(\alpha I + S)^{-1} \left((\alpha I + S)^T + (\alpha I + S) \right) (\alpha I + S)^{-T} B^T \\ &= 2\alpha I + B(\alpha I + S)^{-1} (\alpha I - S + \alpha I + S) (\alpha I + S)^{-T} B^T \\ &= 2\alpha \left(I + (B(\alpha I + S)^{-1}) (B(\alpha I + S)^{-1})^T \right). \end{aligned}$$

این رابطه نشان می‌دهد که $Y + Y^T$ معین مثبت متقارن است، لذا Y معین مثبت و نامنفرد است. با جمع‌بندی مطالب ارائه‌شده و اینکه رتبه ستونی ماتریس B^T کامل است، نتیجه می‌گیریم که رتبه ماتریس U_{12} نیز کامل است. سرانجام، داریم

$$Q = V^T U V = \begin{pmatrix} V_{11}^T U_{11} V_{11} & V_{11}^T U_{12} V_{22} \\ V_{22}^T U_{21} V_{11} & V_{22}^T U_{22} V_{22} \end{pmatrix},$$

و بنابراین،

$$Q_{12} = V_{11}^T U_{12} V_{22}.$$

این رابطه نشان می‌دهد که رتبه ماتریس Q_{12} کامل است، زیرا ماتریس‌های V_{11} و V_{22} متعامد هستند و رتبه ستونی ماتریس U_{12} کامل است. به این ترتیب، اثبات قضیه کامل می‌شود. $\square$

همان‌گونه که در [۱۷] ذکر شد سرعت همگرایی روش تکراری HSS برای حل مسئله نقطه زینی بسیار کند است. از این رو، در عمل از پیش‌شرط‌ساز HSS برای تسریع سرعت همگرایی یک روش تکراری مبتنی بر زیرفضای کرایلف برای مسئله نقطه زینی استفاده می‌شود. همانند روش HSS، با فرض

$$\mathcal{M}_\alpha = \frac{1}{2\alpha}(\alpha \mathcal{I} + \mathcal{H})(\alpha \mathcal{I} + \mathcal{S}), \quad \mathcal{N}_\alpha = \frac{1}{2\alpha}(\alpha \mathcal{I} - \mathcal{H})(\alpha \mathcal{I} - \mathcal{S}),$$

داریم

$$\mathcal{A} = \mathcal{M}_\alpha - \mathcal{N}_\alpha,$$

و

$$\mathcal{T}_\alpha = \mathcal{M}_\alpha^{-1} \mathcal{N}_\alpha = \mathcal{I} - \mathcal{M}_\alpha^{-1} \mathcal{A}.$$

اگر شرایط قضیه ۱.۱۱ برقرار باشند، با توجه به اینکه $\rho(\mathcal{T}_\alpha) < ۱$ نتیجه می‌گیریم که مقادیر ویژه $\mathcal{T}_\alpha$ درون دایره‌ای به مرکز $(۰, ۰)$ و شعاع ۱ قرار دارند. بنابراین، همه مقادیر ویژه $\mathcal{M}_\alpha^{-1} \mathcal{A}$ درون دایره‌ای به مرکز $(۱, ۰)$ و به شعاع ۱ قرار دارند. از این رو، استفاده از روش‌های مبتنی بر زیرفضای کرایلف مثل GMRES برای حل دستگاه

$$\mathcal{M}_\alpha^{-1} \mathcal{A}x = \mathcal{M}_\alpha^{-1} b, \quad (۲۹-۱۱)$$

مناسب خواهد بود.

لزوم معین مثبت بودن ماتریس A در قضیه ۱.۱۱ اساسی نیست. اگر ماتریس A نیمه‌معین مثبت باشد، آنگاه روش تکراری (۱۱-۱۹) خوش‌تعریف است (هر دو ماتریس $\alpha \mathcal{I} + \mathcal{H}$ و $\alpha \mathcal{I} + \mathcal{S}$ نامنفرد هستند)، اما ممکن است به‌ازای هر $\alpha > ۰$ ، $\rho(\mathcal{T}_\alpha) = ۱$ به مثال زیر توجه کنید.

مثال ۱.۱۱ فرض کنید $n = ۲$ ، $m = ۱$ و

$$A = \begin{pmatrix} ۱ & ۰ \\ ۰ & ۰ \end{pmatrix}, \quad B = \begin{pmatrix} ۰ & ۱ \end{pmatrix}, \quad C = (۰).$$

به‌ازای مقدار دلخواه α ، با محاسبه $\mathcal{M}_\alpha$ خواهیم داشت

$$\mathcal{M}_\alpha = \frac{۱}{۲} \begin{pmatrix} \alpha + ۱ & ۰ & ۰ \\ ۰ & \alpha & ۱ \\ ۰ & -۱ & \alpha \end{pmatrix}.$$

همچنین، داریم

$$\mathcal{M}_\alpha^{-1} \mathcal{A} = \frac{۱}{۱ + \alpha^۲} \begin{pmatrix} \frac{۲(۱ + \alpha^۲)}{\alpha + ۱} & ۰ & ۰ \\ ۰ & ۲ & ۲\alpha \\ ۰ & -۲\alpha & ۲ \end{pmatrix}.$$

با محاسبه مقادیر ویژه ماتریس $M_{\alpha}^{-1}A$ خواهیم داشت

$$\sigma(M_{\alpha}^{-1}A) = \left\{ \frac{2}{1+\alpha}, \frac{2i}{\alpha+i}, \frac{2}{1+\alpha i} \right\}.$$

بنابراین،

$$\sigma(T_{\alpha}) = \sigma(I - M_{\alpha}^{-1}A) = \left\{ \frac{\alpha-1}{\alpha+1}, \frac{\alpha-i}{\alpha+i}, \frac{\alpha i-1}{\alpha i+1} \right\}.$$

ازاین رو،

$$\rho(T_{\alpha}) = \max \left\{ \left| \frac{\alpha-1}{\alpha+1} \right|, 1, 1 \right\} = 1.$$

دراین صورت، مقادیر ویژه ماتریس $M_{\alpha}^{-1}A$ درون و روی دایره‌ای به مرکز $(1, 0)$ و شعاع ۱ قرار دارند و می‌توان از M_{α} به عنوان یک پیش شرط ساز استفاده کرد.

در ادامه، نشان می‌دهیم که با یک ترفند مناسب می‌توان یک روند تکراری همگرا ساخت. ابتدا توجه کنید که $1 \notin \sigma(T_{\alpha})$ ، زیرا هر دو ماتریس M_{α} و A نامنفرد هستند. روند تکراری $(11-23)$ از رابطه نقطه ثابت

$$x = T_{\alpha}x + c, \quad (11-30)$$

به دست می‌آید. فرض کنید $0 < \omega < 1$. رابطه $(11-30)$ را به صورت

$$x = ((1-\omega)I + \omega T_{\alpha})x + \omega c,$$

نوشته و روند تکراری

$$x^{(k+1)} = ((1-\omega)I + \omega T_{\alpha})x^{(k)} + \omega c,$$

را می‌سازیم. دراین صورت، ماتریس تکرار روش به دست آمده به صورت

$$T_{\alpha,\omega} = (1-\omega)I + \omega T_{\alpha},$$

است. مقادیر ویژه این ماتریس به صورت

$$\lambda_{\alpha,\omega} = (1-\omega) + \omega \lambda_{\alpha},$$

هستند، که در آن $\lambda_{\alpha} \in \sigma(T_{\alpha})$. اکنون، از اینکه $1 \neq \lambda_{\alpha}$ ، $|\lambda_{\alpha}| \leq 1$ و $0 < \omega < 1$ نتیجه می‌گیریم $|\lambda_{\alpha,\omega}| < 1$ ، که همگرایی روش را نتیجه می‌دهد.

در به‌کارگیری پیش‌شرط‌ساز HSS در یک روش تکراری مبتنی بر زیر فضای کرایلف برای حل مسئله نقطه‌زینی، بایستی بردارهایی مثل

$$v = \mathcal{M}_\alpha^{-1} \mathcal{A}u,$$

محاسبه شوند. برای محاسبه v ابتدا بردار $w = \mathcal{A}u$ را محاسبه کرده و سپس دستگاه $\mathcal{M}_\alpha v = w$ را حل می‌کنیم. برای حل این دستگاه، ابتدا دستگاه $(\alpha \mathcal{I} + \mathcal{H})z = w$ ، و سپس دستگاه $(\alpha \mathcal{I} + \mathcal{S})v = z$ را حل می‌کنیم. با جمع‌بندی این مطالب، محاسبه بردار v توسط گام‌های زیر انجام می‌شود.

1. Set $w = \mathcal{A}u$.
2. Solve $(\alpha \mathcal{I} + \mathcal{H})z = w$ for z .
3. Solve $(\alpha \mathcal{I} + \mathcal{S})v = z$ for v .

در گام دوم باید دو دستگاه با ماتریس ضرایب $\alpha I + H$ و $\alpha I + C$ حل شوند. با توجه به اینکه این دو ماتریس معین مثبت متقارن هستند، می‌توان این دو دستگاه را به‌طور دقیق با استفاده از تجزیه چولسکی و به‌طور نادقیق با استفاده از الگوریتم گرادیان مزدوج حل کرد. در گام سوم، باید یک دستگاه با ماتریس ضرایب

$$\mathcal{E} = \begin{pmatrix} \alpha I + S & B^T \\ -B & \alpha I \end{pmatrix}, \quad (31-11)$$

حل شود که شبیه به دستگاه (31-11) است. بنابراین، می‌توان همانند (31-11) این دستگاه را حل کرد. اکنون، حالتی را در نظر می‌گیریم که در آن ماتریس A معین مثبت متقارن است. در این صورت، ماتریس‌های $\mathcal{H}$ و $\mathcal{S}$ به‌صورت زیر تعریف می‌شوند

$$\mathcal{H} = \begin{pmatrix} A & \circ \\ \circ & C \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} \circ & B^T \\ -B & \circ \end{pmatrix}. \quad (32-11)$$

از این رو، ماتریس پیش‌شرط‌ساز $\mathcal{M}_\alpha$ به‌صورت زیر خواهد بود

$$\begin{aligned} \mathcal{M}_\alpha &= \frac{1}{2\alpha} (\alpha \mathcal{I} + \mathcal{H})(\alpha \mathcal{I} + \mathcal{S}) \\ &= \frac{1}{2\alpha} \begin{pmatrix} \alpha I + A & \circ \\ \circ & \alpha I + C \end{pmatrix} \begin{pmatrix} \alpha I & B^T \\ -B & \alpha I \end{pmatrix} \\ &= \frac{1}{2\alpha} \begin{pmatrix} \alpha I + A & \circ \\ \circ & \alpha I + C \end{pmatrix} \begin{pmatrix} I & \circ \\ -\frac{1}{\alpha} B & I \end{pmatrix} \begin{pmatrix} \alpha I & \circ \\ \circ & F_\alpha \end{pmatrix} \begin{pmatrix} I & \frac{1}{\alpha} B^T \\ \circ & I \end{pmatrix}, \end{aligned}$$

که در آن $F_\alpha = \alpha I + \frac{1}{\alpha} BB^T$. این تجزیه نشان می‌دهد که در به‌کارگیری پیش‌شرط‌ساز M_α در هر تکرار از یک روش تکراری مبتنی بر زیرفضای کرایلف لازم است سه دستگاه با ماتریس ضرایب $\alpha I + A$ ، $\alpha I + C$ و F_α حل شوند. با توجه به اینکه هر سه ماتریس معین مثبت متقارن هستند، این سه دستگاه را می‌توان به‌طور دقیق با استفاده از تجزیه چولسکی یا به‌طور نادقیق با استفاده از روش تکراری گرادیان مزدوج حل کرد.

در حالتی که ماتریس A معین مثبت است، استفاده از پیش‌شرط‌ساز HSS برای مسئله نقطه زینی (۸-۱۱) با شکافت $A = \mathcal{H} + \mathcal{S}$ که در آن ماتریس‌های $\mathcal{H}$ و $\mathcal{S}$ توسط (۳۳-۱۱) تعریف شده‌اند در [۸۰] با نام روش شکافت معین مثبت و هرمیتی‌کج تباهیده (DPSS) مورد بررسی قرار گرفته است. در [۲۴]، کائو یک شکافت به‌صورت $A = \mathcal{H} + \mathcal{S}$ را در نظر گرفت که در آن

$$\mathcal{H} = \begin{pmatrix} A & \circ \\ \circ & \circ \end{pmatrix}, \quad \mathcal{S} = \begin{pmatrix} \circ & B^T \\ -B & C \end{pmatrix}, \quad (۳۳-۱۱)$$

و نشان داد که روش تکراری (۱۱-۱۹) به‌ازای این شکافت بدون هیچ شرطی همگراست. سپس، از پیش‌شرط‌ساز تولیدشده برای بهبود سرعت همگرایی روش تکراری GMRES برای حل مسئله نقطه زینی (۸-۱۱) استفاده کرد.

یک نسخه منظم‌سازی‌شده DPSS در [۲۳] و نسخه‌های منظم‌سازی‌شده HSS در مقالات [۶، ۷] ارائه شده‌اند.

۱۱-۶ روش لاگرانژی افزوده

مسئله نقطه زینی (۱-۱۱) را در نظر بگیرید که در آن $\text{rank}(B) = m$ ، $C = \circ$ ، و ماتریس A نیمه‌معین مثبت متقارن است به‌گونه‌ای که $\mathcal{N}(A) \cap \mathcal{N}(B) = \{\circ\}$. با مفروضات بالا مسئله (۱-۱۱) به‌صورت زیر نوشته می‌شود

$$\begin{pmatrix} A & B^T \\ -B & \circ \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ -g \end{pmatrix}. \quad (۳۴-۱۱)$$

با استفاده از لم ۱۱.۱ ماتریس ضرایب این دستگاه نامنفرد است. این‌گونه مسائل در بهینه‌سازی و حل مسئله استوکس با روش اجزای محدود ظاهر می‌شوند. با توجه به اینکه ماتریس A ممکن است منفرد باشد، هیچ‌کدام از پیش‌شرط‌سازهای P_i ، $i = 1, 2, 3, 4$ که در بخش ۱۱-۴ معرفی شدند را نمی‌توان

۳۹۶ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تَنگ

برای این دستگاه به‌کار برد زیرا در همه آنها بلوک $(1, 1)$ منفرد و در نتیجه، $\mathcal{P}_i$ ها منفرد خواهند بود. اما با یک ترفند مناسب می‌توان این مشکل را حل کرد که به روش لاگرانژی افزوده معروف است. در ادامه، این روش را معرفی می‌کنیم.

فرض کنید $W \in \mathbb{R}^{m \times m}$ یک ماتریس معین مثبت متقارن دلخواه باشد و $\gamma > 0$. با ضرب طرفین معادله دوم در $(11-34)$ در ماتریس $-\gamma B^T W^{-1}$ و افزودن طرفین رابطه به دست آمده به طرفین رابطه اول در $(11-34)$ ، خواهیم داشت

$$\begin{pmatrix} A + \gamma B^T W^{-1} B & B^T \\ -B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f + \gamma B^T W^{-1} g \\ -g \end{pmatrix}. \quad (11-35)$$

روشن است که دستگاه‌های $(11-34)$ و $(11-35)$ معادل هم هستند، با این تفاوت که بلوک $(1, 1)$ در دستگاه $(11-35)$ معین مثبت متقارن است؛ زیرا نخست اینکه ماتریس $A_\gamma = A + \gamma B^T W^{-1} B$ متقارن است و به علاوه، به ازای هر بردار ناصفر x ، داریم

$$\begin{aligned} x^T A_\gamma x &= x^T A x + \gamma x^T B^T W^{-1} B x \\ &= x^T A x + \gamma (Bx)^T W^{-1} (Bx) \geq 0 \end{aligned}$$

از طرفی، اگر $x^T A_\gamma x = 0$ ، آنگاه

$$x^T A x = \gamma (Bx)^T W^{-1} (Bx) = 0.$$

اکنون، با توجه به قضیه ۲۰.۱، داریم

$$Ax = 0, \quad Bx = 0.$$

در نتیجه، داریم $x \in \mathcal{N}(A) \cap \mathcal{N}(B)$. از این رو، $x = 0$ که نشان می‌دهد که ماتریس A_γ معین مثبت متقارن است.

در مقالات [۴۱، ۴۲] نشان داده شده است که به‌کاربردن روش‌های تکراری و پیش‌شرط‌سازها برای مسئله $(11-35)$ مزایایی بر $(11-34)$ دارد؛ درحقیقت، اگر ماتریس A بد حالت باشد، با انتخاب مناسب ماتریس W ، بلوک $(1, 1)$ در دستگاه $(11-35)$ خوش حالت خواهد بود. برای مثال، یک انتخاب مناسب به صورت

$$\gamma = \frac{\|A\|_2}{\|B\|_2}, \quad W = I,$$

است، که تقریباً عدد حالت بلوک $(1, 1)$ و کل ماتریس ضرایب در دستگاه $(11-35)$ را کمینه می‌کند $[41, 42]$.

از اینجا به بعد، برای سادگی فرض کنید

$$A_\gamma = A + \gamma B^T W^{-1} B, \quad f_\gamma = f + \gamma B^T W^{-1} g.$$

در این صورت، دستگاه $(11-35)$ به صورت زیر خواهد بود

$$A_\gamma \mathbf{x} \equiv \begin{pmatrix} A_\gamma & B^T \\ -B & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f_\gamma \\ -g \end{pmatrix} = \mathbf{b}_\gamma. \quad (11-36)$$

اکنون، یک پیش شرط ساز باین مثلثی بلوکی کارا برای مسئله $(11-37)$ به صورت زیر تعریف می‌شود $[43]$

$$\mathcal{P}_\gamma = \begin{pmatrix} A_\gamma & 0 \\ -B & \frac{1}{\gamma} W \end{pmatrix}. \quad (11-37)$$

برای بررسی طیف ماتریس $A_\gamma \mathcal{P}_\gamma^{-1}$ ، نخست، لم زیر را بیان می‌کنیم.

لم ۲.۱۱ فرض کنید ماتریس‌های A و W معین مثبت متقارن باشد و رتبه ماتریس B کامل باشد. داریم

$$(i) \quad \left(B (I + B^T B)^{-1} B^T \right)^{-1} = I + (B B^T)^{-1};$$

$$(ii) \quad \left(B (A + B^T W^{-1} B)^{-1} B^T \right)^{-1} = W^{-1} + (B A^{-1} B^T)^{-1}.$$

برهان. برای اثبات قسمت (i) ، ابتدا توجه کنید که

$$(I + B B^T) B = B + B B^T B = B (I + B^T B).$$

در نتیجه، $B (I + B^T B)^{-1} = (I + B B^T)^{-1} B$. اگر طرفین این رابطه را از راست در B^T ضرب کرده و سپس از طرفین رابطه به دست آمده معکوس بگیریم، آنگاه خواهیم داشت

$$(B (I + B^T B)^{-1} B^T)^{-1} = (B B^T)^{-1} (I + B B^T) = I + (B B^T)^{-1}.$$

برای اثبات قسمت (ii) ، با فرض $\hat{B} = W^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ داریم

$$\left(B (A + B^T W^{-1} B)^{-1} B^T \right)^{-1} = \left(B (A^{\frac{1}{2}} A^{\frac{1}{2}} + B^T W^{-\frac{1}{2}} W^{-\frac{1}{2}} B)^{-1} B^T \right)^{-1}$$

$$\begin{aligned}
 &= \left(BA^{-1}(I + \hat{B}^T \hat{B})^{-1} A^{-1} B^T \right)^{-1} \\
 &= \left(W^{-1} \hat{B} (I + \hat{B}^T \hat{B})^{-1} \hat{B}^T W^{-1} \right)^{-1} \\
 &= W^{-1} \left(\hat{B} (I + \hat{B}^T \hat{B})^{-1} \hat{B}^T \right)^{-1} W^{-1} \\
 &\stackrel{(i)}{=} W^{-1} \left(I + (\hat{B} \hat{B}^T)^{-1} \right) W^{-1} \\
 &= W^{-1} + W^{-1} (\hat{B} \hat{B}^T)^{-1} W^{-1} \\
 &= W^{-1} + (BA^{-1} B^T)^{-1},
 \end{aligned}$$

□ که اثبات لم را کامل می‌کند.

برای مطالعه توزیع مقادیر ویژه ماتریس $\mathcal{P}_\gamma^{-1} A_\gamma$ ، فرض می‌کنیم شرایط لم ۲.۱۱ برقرار باشد. نخست، توجه کنید که با استفاده از تمرین ۱.۱۱، داریم

$$\mathcal{P}_\gamma^{-1} = \begin{pmatrix} A_\gamma^{-1} & \circ \\ \gamma W^{-1} B A_\gamma^{-1} & \gamma W^{-1} \end{pmatrix}.$$

اکنون با ضرب ماتریس $\mathcal{P}_\gamma^{-1}$ در ماتریس A_γ خواهیم داشت

$$\mathcal{P}_\gamma^{-1} A_\gamma = \begin{pmatrix} I & A_\gamma^{-1} B^T \\ \circ & \gamma W^{-1} B A_\gamma^{-1} B^T \end{pmatrix}. \quad (38-11)$$

این رابطه نشان می‌دهد که ماتریس $\mathcal{P}_\gamma^{-1} A_\gamma$ دارای n مقدار ویژه ۱ است و سایر مقادیر ویژه آن مقادیر ویژه ماتریس

$$G_\gamma = \gamma W^{-1} B A_\gamma^{-1} B^T,$$

هستند. با توجه به قسمت دوم لم ۲.۱۱، داریم

$$\left(B A_\gamma^{-1} B^T \right)^{-1} = \gamma W^{-1} + (B A^{-1} B^T)^{-1},$$

یا

$$\begin{aligned}
 G_\gamma &= \gamma W^{-1} B A_\gamma^{-1} B^T \\
 &= \gamma W^{-1} (\gamma W^{-1} + (B A^{-1} B^T)^{-1})^{-1}
 \end{aligned}$$

$$= \gamma(\gamma I + (W^{-1}BA^{-1}B^T)^{-1})^{-1}.$$

ازاین رو، مقادیر ویژه G_γ به صورت زیر هستند

$$\lambda_i(\gamma) = \frac{\gamma}{\gamma + \mu_i^{-1}},$$

که در آن μ_i ها مقادیر ویژه $W^{-1}BA^{-1}B^T$ هستند. لازم به ذکر است که داریم

$$W^{-1}BA^{-1}B^T = W^{-\frac{1}{2}}(W^{-\frac{1}{2}}BA^{-1}B^TW^{-\frac{1}{2}})W^{\frac{1}{2}}.$$

لذا، ماتریس های $W^{-1}BA^{-1}B^T$ و $W^{-\frac{1}{2}}BA^{-1}B^TW^{-\frac{1}{2}}$ متشابه هستند و مقادیر این دو ماتریس یکسان هستند. بنابراین، از اینکه ماتریس $W^{-\frac{1}{2}}BA^{-1}B^TW^{-\frac{1}{2}}$ معین مثبت متقارن است نتیجه می گیریم که $\mu > 0$. ازاین رو،

$$0 < \lambda(\gamma) = \frac{\gamma}{\gamma + \mu^{-1}} < 1.$$

ازطرفی، وقتی که $\gamma \rightarrow \infty$ ، آنگاه $\lambda(\gamma)$ به سمت عدد ۱ میل می کند.

در به کارگیری پیش شرط ساز $\mathcal{P}_\gamma$ در یک روش تکراری مبتنی بر زیرفضای کرایلف، در هر تکرار باید دو دستگاه با ماتریس ضرایب W و A_γ حل شوند. جزئیات حل این گونه دستگاه ها پیش تر بیان شده است.

تمرینات

۱.۱۱ فرض کنید $A \in \mathbb{R}^{r \times r}$ ، $B \in \mathbb{R}^{s \times s}$ ، $C \in \mathbb{R}^{r \times s}$ و $D \in \mathbb{R}^{s \times r}$. نشان دهید اگر ماتریس های A و $S = B - RA^{-1}C$ نامنفرد باشند، آنگاه

$$\begin{pmatrix} A & C \\ R & B \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}CS^{-1}RA^{-1} & -A^{-1}CS^{-1} \\ -S^{-1}AR^{-1} & S^{-1} \end{pmatrix}.$$

۲.۱۱ نشان دهید که ماتریس R_{11} (همچنین، ماتریس R_{22}) که در توسط رابطه (۲۴-۱۱) تعریف شده است، یک ماتریس متقارن است.

۳.۱۱ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن، رتبه ماتریس $B \in \mathbb{R}^{m \times n}$ کامل باشد ($m \leq n$ و $\text{rank}(B) = m$)، ماتریس $C \in \mathbb{R}^{m \times m}$ نیمه معین مثبت متقارن باشد. نشان دهید ماتریس $S = C + BA^{-1}B^T$ معین مثبت متقارن است.

۴۰۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

۴.۱۱ نشان دهید که روند تکراری اوزاوا (۱۱-۱۱) را می‌توان به صورت (۱۱-۱۳) نوشت.

۵.۱۱ فرض کنید ماتریس A نیمه‌معین مثبت متقارن و رتبهٔ سطری ماتریس B کامل است. مستقیماً (بدون استناد به لم ۱.۱۱) ثابت کنید ماتریس

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ -B & \circ \end{pmatrix}, \quad (۳۹-۱۱)$$

نامنفرد است اگر و تنها اگر $\{ \circ \} = \mathcal{N}(A) \cap \mathcal{N}(B)$.

۶.۱۱ فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ معین مثبت متقارن است و رتبهٔ سطری ماتریس $B \in \mathbb{R}^{m \times n}$ کامل باشد ($\text{rank}(B) = m$ و $m \leq n$). مسئلهٔ

$$\|c - B^T y^*\|_{A^{-1}} = \min_{y \in \mathbb{R}^m} \|c - B^T y\|_{A^{-1}}, \quad (۴۰-۱۱)$$

را در نظر بگیرید که در آن $c \in \mathbb{R}^n$ یک بردار دلخواه است و $\|\cdot\|_{A^{-1}} = \sqrt{\langle A^{-1} \cdot, \cdot \rangle}$. نشان دهید جواب مسئلهٔ (۴۰-۱۱) در دستگاه معادلات خطی $BA^{-1}B^T y = BA^{-1}c$ صدق می‌کند. با فرض $x = A^{-1}(c - B^T y)$ ، نشان دهید که x و y در دستگاه معادلات نقطهٔ زینی زیر صدق می‌کنند

$$\begin{pmatrix} A & B^T \\ -B & \circ \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ \circ \end{pmatrix}.$$

۷.۱۱ فرض کنید $\mathcal{P}_1$ و $\mathcal{P}_2$ ماتریس‌هایی باشند که به ترتیب در رابطه‌های (۱۱-۱۵) و (۱۱-۱۶) تعریف شده باشند. نشان دهید که ماتریس $\mathcal{P}_1^{-1} \mathcal{A}$ قطری‌شدنی است، اما ماتریس $\mathcal{P}_2^{-1} \mathcal{A}$ قطری‌شدنی نیست، که در آن ماتریس $\mathcal{A}$ در رابطهٔ (۱۱-۱۴) تعریف شده است.

۸.۱۱ فرض کنید

$$\mathcal{P}_3 = \begin{pmatrix} A & \circ \\ \circ & -S \end{pmatrix}, \quad \mathcal{P}_4 = \begin{pmatrix} A & B^T \\ \circ & -S \end{pmatrix},$$

که در آن $S = BA^{-1}B^T$. نشان دهید اگر $\mathcal{A}$ ماتریس تعریف‌شده در (۱۱-۴۱) باشد، آنگاه (الف) ماتریس $\mathcal{P}_3^{-1} \mathcal{A}$ دارای سه مقدار ویژهٔ متمایز ۱ و $\frac{1}{2}(1 \pm \sqrt{5})$ است، چندجمله‌ای مینیمال ماتریس برابر با $q(t) = (t-1)(t^2 - t - 1)$ است و قطری‌شدنی است؛

(ب) ماتریس $\mathcal{A} \mathcal{P}_\tau^{-1}$ دارای دو مقدار ویژه ± 1 است، چندجمله‌ای مینیمال آن به صورت $q(t) = t^2 - 1$ است و ماتریس $\mathcal{A} \mathcal{P}_\tau^{-1}$ قطری‌شدنی است.

۹.۱۱ نشان دهید که می‌توان دستگاه (۱۱-۲۱) را به گونه‌ای حل کرد که به جای حل یک دستگاه با ماتریس ضرایب $\alpha^2 I + B(I + \frac{1}{\alpha} S)^{-1} B^T$ ، یک دستگاه با ماتریس ضرایب $\alpha I + S + \frac{1}{\alpha} B^T B$ حل شود. از لحاظ هزینه محاسباتی این دو دستگاه را با هم مقایسه کنید.

۱۰.۱۱ فرض کنید $m \leq n$ و رتبه ماتریس $B \in \mathbb{R}^{m \times n}$ کامل باشد ($\text{rank}(B) = m$). به علاوه، فرض کنید ماتریس $A \in \mathbb{R}^{n \times n}$ روی فضای پوچ ماتریس B معین مثبت باشد، یعنی به ازای هر بردار ناصفر $x \in \mathcal{N}(B)$ داریم $x^T A x > 0$. نشان دهید که ماتریس

$$\mathcal{A} = \begin{pmatrix} A & B^T \\ -B & 0 \end{pmatrix}, \quad (11-41)$$

نامنفرد است.

منابع و مأخذ

- [۱] د. خجسته سالکویه، روش‌های عددی در جبر خطی، انتشارات دانشگاه گیلان، چاپ دوم، ۱۳۹۶.
- [2] K. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Nonlinear Programming*, Stanford University Press, Stanford, CA, 1958.
- [3] O. Axelsson, *Iterative Solution Methods*, Cambridge university press, 1996.
- [4] O. Axelsson, *Unified analysis of preconditioning methods for saddle point matrices*, Numer. Linear Algebra Appl. 22 (2015) 233-253.
- [5] U.M. Aschur and C. Greif, *A First Course in Numerical Methods*, SIAM, Philadelphia, 2011.
- [6] Z.-Z. Bai, *Regularized HSS iteration methods for stabilized saddle-point problems*, IMA J. Numer. Anal. 39 (2019) 1888-1923.
- [7] Z.-Z. Bai and M. Benzi, *Regularized HSS iteration methods for saddle-point linear systems*, BIT 57 (2017) 287-311.
- [8] Z.-Z. Bai, M. Benzi, and F. Chen, *On preconditioned MHSS iteration methods for complex symmetric linear systems*, Numer. Algor. 56 (2011) 297-317.
- [9] Z.-Z. Bai, M. Benzi, F. Chen, and Z.-Q. Wang, *Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems*, IMA J. Numer. Anal. 33 (2013) 343-369.

- [10] Z.-Z. Bai and J.-Y. Pan, *Matrix Analysis and Computations*, SIAM, 2021.
- [11] Z.-Z. Bai, G.H. Golub, and M.K. Ng, *Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems*, SIAM J. Matrix Anal. Appl. 24 (2003) 603-626.
- [12] Z.-Z. Bai, G.H. Golub, and M.K. Ng, *Modified HSS iteration methods for a class of complex symmetric linear systems*, Computing 87 (2010) 93-111.
- [13] Z.-Z. Bai, G.H. Golub, and M.K. Ng, *On successive-overrelaxation acceleration of the Hermitian and skew-Hermitian splitting iterations*, Numer. Linear Algebra Appl. 14 (2007) 319-335.
- [14] Z.-Z. Bai, G.H. Golub, L.-Z. Lu, and J.-F. Yin, *Block triangular and skew-Hermitian splitting methods for positive definite linear systems*, SIAM J. Sci. Comput. 26 (2005) 844-863.
- [15] M. Benzi, *Preconditioning techniques for large linear systems: a survey*, J. Comput. Phys. 182 (2002) 418-477.
- [16] M. Benzi and G.H. Golub, *A preconditioner for generalized saddle point problems*, SIAM J. Matrix Anal. Appl. 26 (2004) 20-41.
- [17] M. Benzi, G.H. Golub, and J. Liesen, *Numerical solution of saddle point problems*, Acta Numerica 14 (2005), pp. 1-137.
- [18] M. Benzi, C.D. Meyer, and M. Tuma, *A sparse approximate inverse preconditioner for the conjugate gradient method*, SIAM J. Sci. Comput. 17 (1996) 1135-1149.
- [19] M. Benzi and M. Tuma, *A sparse approximate inverse preconditioner for nonsymmetric linear systems*, SIAM J. Sci. Comput. 19 (1998) 968-994.

- [20] M. Benzi and M. Tuma, *A comparative study of sparse approximate inverse preconditioners*, Appl. Numer. Math. 30 (1999) 305-340.
- [21] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979.
- [22] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
- [23] Y. Cao, *Regularized DPSS preconditioner for non-Hermitian saddle point problems*, Appl. Math. Lett. 84 (2018) 96-102.
- [24] Y. Cao, *A block positive-semidefinite splitting preconditioner for generalized saddle point linear systems*, J. Comput. Appl. Math. 374 (2020) 112787.
- [25] Z.-H. Cao, *Fast Uzawa algorithm for generalized saddle point problems*, Appl. Numer. Math. 46 (2003) 157–171.
- [26] C.C. Cheney, *Introduction to Approximation Theory*, McGraw Hill, NY, 1966.
- [27] E. Chow and Y. Saad, *ILUS: an incomplete LU factorization for matrices in sparse skyline format*, Int. J. Numer. Meth. Fluids 25 (1997) 739-748.
- [28] E. Chow and Y. Saad, *Approximate inverse preconditioners via sparse-sparse Iterations*, SIAM J. Sci. Comput. 19 (1998) 995–1023.
- [29] E.J. Craig, *The N -Step iteration procedures*, J. Math. Phys. 34 (1955) 64-73.
- [30] S. Demko, W.F. Moss and P.W. Smith, *Decay rates for inverses of band matrices*, Math. Comput. 43 (1984) 491–499.
- [31] J. Douglas and H. Rachford, *Alternating direction methods for three space variables*, Numer. Math. 4 (1955) 41-63.
- [32] B.N. Datta, *Numerical Linear Algebra and Applications*, Brooks/Cole Publishing Company, 1995.

- [33] T. Davis, *University of Florida sparse matrix collection*, NA Digest, 92 (1994), <http://www.cise.ufl.edu/research/sparse/matrices>.
- [34] I.S. Duff, *A survey of sparse matrix research*, Proceedings of the IEEE 65 (1977) 500-535.
- [35] H.C. Elman, D.J. Silvester and A.J. Wathen, *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*, Oxford University Press, 2006.
- [36] J. Erxiong, *Algorithm for solving shifted skew-symmetric linear system*, Front. Math. China 2 (2007) 227-242.
- [37] R. Fletcher, *Conjugate gradient methods for indefinite systems*, in Proceedings of the Dundee Biennial Conference on Numerical Analysis 1974, G. A. Watson, ed., New York, Springer Verlag, (1975) 73-89.
- [38] A. George and J.W. Liu, *The evolution of the minimum degree ordering algorithm*, SIAM Review 31 (1989) 1-19.
- [39] W. Gander, M.J. Gander, and F. Kwok, *Scientific Computing - An Introduction Using Maple and MATLAB*, Springer, 2014.
- [40] G.H. Golub, *The Use of Chebyshev Matrix Polynomials in the Iterative Solution of Linear Equations Compared to the Method of Successive Overrelaxation*, PhD thesis, University of Illinois at Urbana-Champaign, 1959.
- [41] G.H. Golub and C. Greif, *On solving block-structured indefinite linear systems*, SIAM J. Sci. Comput. 24 (2003) 2076-2092.
- [42] G.H. Golub, C. Greif, and J.M. Varah, *An algebraic analysis of a block diagonal preconditioner for saddle point systems*, SIAM J. Matrix Anal. Appl. 27 (2005) 779-792.

- [43] G.H. Golub, C.F. Van Loan, *Matrix Computations*, Fourth Edition, The Johns Hopkins University Press, 2013.
- [44] M.J. Grote and T. Huckle, *Parallel preconditioning with sparse approximate inverses*, SIAM J. Sci. Comput. 18 (1997) 838-853.
- [45] A. Hadjidimos, *Accelerated overrelaxation method*, Math. Comput. 32 (1978) 149-157.
- [46] P.C. Hansen, *Discrete Inverse Problems*, SIAM, Philadelphia, 2010.
- [47] P.C. Hansen, *Regularization tools version 4.0 for Matlab 7.3*, Numer. Algor. 46 (2007) 189-194.
- [48] M. Hasani and D.K. Salkuyeh, *Improvements of two preconditioned AOR iterative methods for Z-Matrices*, Bull. Iranian Math. Soc. 40 (2014) 357-371.
- [49] M.R. Hestenes and E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards: Section B 49 (1952) 409-436.
- [50] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [51] I.C.F. Ipsen, *A note on preconditioning nonsymmetric matrices*, SIAM J. Sci. Comput. 23 (2001) 1050-1051.
- [52] S.A. Kharchenko, L.Y. Kolotilina, A.A. Nikishin, and A.Y. Yeremin, *A robust AINV-type method for constructing sparse approximate inverse preconditioners in factored form*, Numer. Linear Algebra Appl. 8 (2001) 165-179.
- [53] D.K. Salkuyeh, *An iterative method for symmetric positive semidefinite linear system of equations*, Dem. Math. 47 (2014) 482-492.
- [54] D.K. Salkuyeh, *A sparse approximate inverse preconditioner for nonsymmetric positive definite matrices*, J. Appl. Math. & Informatics 28 (2010) 1131-1141.

- [55] D.K. Salkuyeh, *Generalized Jacobi and the Gauss-Seidel methods for solving linear system of equations*, Numer. Math. J. Chinese Univ. 16 (2007) 164-170.
- [56] D.K. Salkuyeh, *Stationary iteration methods from preconditioning point of view*, Submitted.
- [57] D.K. Salkuyeh, M. Hasani and F.P.A. Beik, *On the preconditioned AOR iterative method for Z-matrices*, Comp. Appl. Math. 36 (2017) 877-883.
- [58] D. Khojasteh Salkuyeh and A. Fahim, *A new iterative refinement of the solution of ill-conditioned linear system of equations*, Inter. J. Comput. Math. 88 (2011) 950-956.
- [59] D.K. Salkuyeh, A. Rafiei, and H. Roohani, *ILU preconditioning based on the FAPINV algorithm*, Opuscula Math. 35 (2015) 235-250.
- [60] D.K. Salkuyeh and F. Toutounian, *A sparse-sparse iteration for computing a sparse incomplete factorization factorization of the inverse of an SPD matrix*, Appl. Numer. Math. 59 (2009) 1265-1273.
- [61] D.K. Salkuyeh and F. Toutounian, *BILUS: A block version of ILUS factorization*, J. Appl. Math. Comput. 15 (2004) 299-312.
- [62] D.K. Salkuyeh and F. Toutounian *A block version algorithm to approximate inverse factors*, Appl. Math. Comput. 162 (2005) 1499-1509.
- [63] D.K. Salkuyeh and M. Masoudi *Positive semidefinite/positive semidefinite splitting iteration methods for solving nonsingular non-Hermitian positive semidefinite systems*, <https://arxiv.org/abs/2503.02361>, 2025, submitted.
- [64] L.Y. Kolotilina and A.Y. Yeremin, *Factorized sparse approximate inverse preconditioning I. Theory*, SIAM J. Matrix Anal. Appl. 14 (1993) 45-58.

- [65] L.Y. Kolotilina and A.Y. Yeremin, *Factorized sparse approximate inverse preconditioning II: Solution of 3D FE systems on massively parallel computers*, Int. J. High Speed Comput. 7 (1995) 191-215.
- [66] R. Kress, *Numerical Analysis*, Springer-Verlag, New York, 1998.
- [67] C. Lanczos, *Solution of systems of linear equations by minimized iterations*, J. Res. National Bureau Standards 49 (1952) 33-53.
- [68] V.I. Lebedev and S.A. Finogenov, *On the order of choice of the iteration parameters in the Chebyshev cyclic iteration method* Zhur. Vych. Mat. i Mat. Fiz. 11 (1971) 425-438.
- [69] E.-J. Lee and J. Zhang, *Factored approximate inverse preconditioners with dynamic sparsity patterns*, Comput. Math. Appl. 62 (2011) 235-242.
- [70] L. Li, T.Z. Huang, X.P. Liu, *Modified Hermitian and skew-Hermitian splitting methods for non-Hermitian positive-definite linear systems*, Numer. Linear Algebra Appl. 14 (2007) 217-235.
- [71] J.-G. Luo, *An incomplete inverse as a preconditioner for the conjugate gradient method*, Comput. Math. Appl. 25 (1993) 73-79.
- [72] J.-G. Luo, *A new class of decomposition for inverting asymmetric and indefinite matrices*, Comput. Math. Appl. 25 (1993) 95-104.
- [73] J.-G. Luo, *A new class of decomposition for symmetric systems*, Mech. Res. Commun. 19 (1992) 159-166.
- [74] T.A. Manteuffel, *An incomplete factorization technique for positive definite linear systems*, Math. Comput. 34 (1980) 473-497.
- [75] Matrix Market page, <http://math.nist.gov/MatrixMarket>.

- [76] J.A. Meijjerink and H.A. van der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comput. 31 (1977) 148-162.
- [77] C.D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
- [78] M.F. Murphy, G.H. Golub, and A.J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput. 21 (2000) 1969-1972.
- [79] C.C. Paige and A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal. 12 (1975) 617-624.
- [80] J.-Y. Pan, M.K. Ng, and Z.-Z. Bai, *New preconditioners for saddle point problems*, Appl. Math. Comput. 172 (2006) 762-771.
- [81] B.N. Parlett, D.R. Taylor, and Z.S. Liu, *A look-ahead Lanczos algorithm for nonsymmetric matrices*, Math. Comput. 44 (1985) 105-124.
- [82] D. Peaceman and H. Rachford, *The Numerical Solution of Parabolic and Elliptic differential*, Equations. J. Soc. Indust. Appl. math. 3 (1995) 28-41.
- [83] R. Piziak and P.L. Odell, *Matrix Theory: From Generalized Inverses to Jordan Form*, Chapman & Hall/CRC, 2007.
- [84] M. Rozložník, *Saddle-Point Problems and Their Iterative Solution*, Springer, 2018.
- [85] Y. Saad, *ILUT: A dual threshold ILU factorization*, Numer. Linear. Algebra Appl. 1 (1994) 307-402.
- [86] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput. 14 (1993) 461-469.
- [87] Y. Saad, *Iterative Methods for Sparse linear Systems*, Second Edition, SIAM, Philadelphia, 2003.

- [88] Y. Saad and M.H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput. 7 (1986) 856-869.
- [89] P. Sonneveld, *CGS: A fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput. 10 (1989) 36-52.
- [90] Y. Song, *Semiconvergence of extrapolated iterative methods for singular linear systems*, J. Comput. Appl. Math. 106 (1999) 117-129.
- [91] Y. Song, *Semiconvergence of nonnegative splittings for singular matrices*, Numer. Math. 85 (2000) 109-127.
- [92] Y. Song and L. Wang, *On the semiconvergence of extrapolated iterative methods for singular linear systems*, Appl. Numer. Math. 44 (2003) 404-413.
- [93] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Second Edition, Springer-Verlag, New York, 1980.
- [94] H.A. van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput. 13 (1992) 631-644.
- [95] L. Wang and Y. Song, *Preconditioned AOR iterative methods for M-matrices*, J. Comput. Appl. Math. 226 (2009) 114-124.
- [96] X. Wu, R. Shao, and Y. Zhu, *New iterative improvement of solution for an ill-conditioned system of equations based on a linear dynamic system*, Comput. Math. Appl. 44 (2002) 917-923.
- [97] M. Wu, L. Wang, and Y. Song, *Preconditioned AOR iterative method for linear systems*, Appl. Numer. Math. 57 (2007) 672-685.
- [98] J. Zhang, *A procedure for computing factored approximate inverse*, M.Sc. dissertation, Department of Computer Science, University of Kentucky, 1999.

- [99] J. Zhang, *A sparse approximate inverse technique for parallel preconditioning of general sparse matrices*, Appl. Math. Comput. 130 (2002) 63-85.
- [100] F. Zhang, *Linear Algebra, Challenging Problems for Students*, The Johns Hopkins University Press, 2009.

نمایه

گرام-اشمیت آرنولد، ۲۱۰	۱
لانکزوس، ۲۳۰، ۲۳۲	آرنولد، ۲۰۹
لانکزوس دوطرفه، ۲۶۱	آنالیز همگرایی، ۳۵۷
مانده شبه‌مینیمال، ۲۶۸	اثر، ۷
مانده شبه‌مینیمال، ۲۵۵	اجزای محدود، ۳۷۵
متعامدسازی کامل، ۲۱۴، ۲۱۷	استاندارد، ۹
مینیمم درجه تقریبی متقارن، ۳۲۱	استایفل، ۲۲۹
BiCG، ۲۶۷، ۲۶۲، ۲۵۵	استوکس، ۳۷۵
BiCG پایدارشده، ۲۷۹	اسکی‌لین، ۳۲۳
BiCGSTAB، ۲۷۹، ۲۸۳	اشغال حافظه، ۵۷
CG، ۲۳۵	اصلاح‌شده، ۲۱۴
CGNE، ۲۴۹، ۳۰۱	اعمال سطری مقدماتی، ۴۰
CGNR، ۲۴۶، ۳۰۱	افرازبندی، ۳۹۰
CGS، ۲۷۵، ۲۷۸	افزوده، ۳۹۵
D-Lanczos، ۲۳۳	الگو، ۳۱۲
FFAPINV، ۳۳۴	الگوریتم
FFAPINV-ILU، ۳۳۷	تجزیه چولسکی ناقص، ۳۱۹
FFAPINV-NSPD، ۳۳۵	چولسکی ناقص، ۳۲۰
FFINV، ۳۳۳	دومتعامدسازی لانکزوس، ۲۵۵
FGMRES، ۲۹۷	کرایگ، ۲۴۹
FOM، ۲۰۵، ۲۱۴	گرادیان دومزدوج، ۲۵۵، ۲۶۲
GMRES، ۲۰۵، ۲۱۷، ۲۹۳، ۳۱۰	گرادیان مزدوج، ۲۲۹، ۲۹۱
GMRES انعطاف‌پذیر، ۲۹۷	گرادیان مزدوج مربع‌شده، ۲۵۵، ۲۷۵
GMRES با شروع مجدد، ۲۲۰	گرام-اشمیت، ۲۱۰

پ	GMRES کامل، ۲۲۰
پارامتر	IC(0)، ۳۱۹
بهینه، ۸۵	ILU(0)، ۳۱۳
شتاب، ۷۴	ILUS، ۳۲۳
فوق تخفیف، ۷۴	ILUT، ۳۱۳
پایا، ۱۴۲	MR، ۳۳۰
پایدار، ۴۱، ۲۱۴	QMR، ۲۵۵، ۲۶۸، ۲۷۳
پایدار شده، ۲۷۹	SYMMMLQ، ۲۳۵
پایه، ۲۰۸	الگوی صفر، ۳۱۲
پائین مثلثی، ۳۹	انباشتگی خطا، ۵۷
پتروف-گالرکین، ۱۷۹، ۱۸۰	انعطاف‌پذیر، ۲۹۷
پردازش تصویر، ۳۷۱	اوزاوا، ۳۷۸
پرون-فروبنیوس، ۱۰۳	اوسین، ۳۷۵
پسرو، ۳۳۱	ایستا، ۵۹
پواسون، ۴۳	ایستایی، ۲۲۸
پوچ‌توان، ۱۴۷	
پهنای بالائی نوار، ۳	ب
پهنای پائینی نوار، ۳	بازمرتب‌سازی، ۳۲۱
پهنای نوار، ۳، ۳۲۶، ۳۳۷	بالا مثلثی، ۳۸
پیشرو، ۳۳۱	بالا مثلثی بلوکی، ۳۷۹
پیش‌شرط‌ساز، ۶۶، ۱۱۸، ۲۸۸	بالاهسنبرگی، ۲، ۲۱۱
اوزاوا، ۳۷۸	بخش مقارن، ۳۷۷
HSS، ۳۶۶	بد حالت، ۳۰، ۵۸، ۱۹۴
پیش‌شرط‌سازی، ۶۶، ۲۸۷	برد، ۶، ۲۸
پیش‌شرط‌سازی دوطرفه، ۲۸۸	بردار تکین، ۲۸
پیش‌شرط‌سازی شده، ۶۶	بسط فوری، ۱۱
پیش‌نگرش، ۲۶۰	بعد، ۴
پیکارد، ۳۷۴	بلوک جردن، ۲۵، ۱۴۰
ت	به‌طور سازگار مرتب‌شده، ۸۷
تابع	بهینه‌سازی، ۳۷۱
علامت، ۱۳۵	

- جایگشت، ۱
جمع مستقیم، ۵، ۱۴۳
جهت متناوب، ۳۴۹
- چ
چبیشف، ۱۲۳، ۱۲۵، ۲۴۲
چکاندن، ۳۱۳، ۳۳۴
چگال، ۴۶، ۴۹، ۳۱۴، ۳۲۶
چگالی، ۳۴۳
چندجمله‌ای
تکین، ۲۰۶
چبیشف، ۱۲۵، ۲۴۲
مشخصه، ۱۲
مینیمال، ۱۵، ۲۰۶، ۳۸۱، ۳۸۲
چولسکی، ۱۹
- ح
حافظه، ۲۰۴
حذفی گاوس، ۵۸، ۵۹، ۳۱۱
حساب دقیق، ۲۴۰
حسابی، ۵۴
- خ
خاصیت سازگاری، ۷
خاصیت ضربی، ۷، ۶۱
خاصیت A، ۸۵
خودالحاق، ۳۰۶، ۳۴۵
خودتوان، ۱۶۵
خوش حالت، ۳۰، ۲۸۷
خوشه، ۳۰۷
خوشه شدن، ۷۲
خوشه شده، ۲۸۷
- گرایان، ۳۷۲
لاگرانژ، ۳۷۲
محدب، ۱۹۲
تباهیده، ۳۹۵
تجزیه
چولسکی، ۱۹، ۳۱۹، ۳۸۴
چولسکی ناقص، ۳۱۹، ۳۲۰
مقدار تکین، ۲۸
ناقص، ۳۱۲
هسته-پوچ توان، ۱۴۷
LDU، ۴۳
LQ، ۲۳۵
LU، ۴۲، ۲۳۳
QR، ۲۶
تحویل پذیر، ۹۶
تحویل ناپذیر، ۹۶، ۱۰۳، ۳۲۶
ترانهاده، ۱
ترانهاده هرمیتی، ۱
ترکیب خطی، ۴
تفاضلات متناهی، ۳۷۵
تقریب، ۲۴۱
تکرار جبری، ۱۳۹
تکرار هندسی، ۱۳۹
تکراری، ۴۸
تکین، ۱۵
تُنک، ۴۵، ۳۴۷
تنکی، ۳۱۲، ۳۱۷
توزیع مقادیر ویژه، ۳۰۸، ۳۹۸
- ج
جایگذاری پسرو، ۳۹
جایگذاری پیشرو، ۳۹

ضربی، ۷، ۶۱

ضمنی، ۳۴۹

ط

طیف، ۱۲، ۳۴۹

ع

عاد، ۱۵

عامل معکوس، ۳۳۱

عدد حالت، ۲۸۷

عدد حالت طیفی، ۳۵۱

عملگر الحاقی، ۱۷۳

عملگر تصویر، ۱۶۵، ۱۶۹

عملگر تصویر متعامد، ۱۷۲، ۱۷۶

عناصر متناهی، ۱۶۵، ۳۷۴

غ

غالب قطری، ۹۶

غالب قطری اکید، ۹۹

غالب قطری تحویل ناپذیر، ۹۹

غالب قطری ضعیف، ۹۹

غیرایستا، ۵۹

ف

فرامعین، ۳۴

فرم جردن، ۶۰

فروبنیوس، ۷

فرومعین، ۳۵

فضای

بردار، ۳

پوچ، ۶، ۲۸

تولیدشده، ۴

ضرب داخلی، ۸

ستونی تَنگ فشرده، ۵۱

سرعت همگرایی مجانبی، ۱۱۶

سریع ترین کاهش، ۱۹۰، ۱۹۴

سه قطری، ۲

سیال، ۳۷۴

ش

شاخص، ۱۳۸

شاخص پوچ توانی، ۱۴۷

شبه بالامثلشی، ۲، ۲۱، ۲۷

شتاب چبیشف، ۱۲۳

شرایط پتروف-گالرکین، ۱۷۹، ۱۸۰

شرایط گالرکین، ۱۸۰

شرط توقف، ۲۱۷

شرط مرزی، ۱۲۲

شروع مجدد، ۲۱۶، ۲۱۷، ۲۲۴

شعاع شبه طیفی، ۱۵۶

شعاع طیفی، ۱۴، ۳۵۳

شکافت، ۶۴، ۳۴۸

شکافت منظم، ۱۱۲

شکافت هرمیتی و هرمیتی کج، ۳۴۷

شکست، ۲۱۳

شکل شور، ۲۱

شکل شور حقیقی، ۲۱

شکل شور مختلط، ۲۱

شکل کانونی جردن، ۲۵

شور، ۲۱

ض

ضرب داخلی، ۸

ضرب داخلی استاندارد، ۹

ضرب داخلی، ۲۹۱

- فوق تخفیف متوالی، ۷۴
- گرشگورین، ۲۶، ۹۹
- گرید، ۲۰۶
- گسسته‌سازی، ۳۷۵
- ق
- قالب
- اسکی لاین، ۳۲۳
- ستونی تُنک فشرده، ۵۱
- سطری تُنک فشرده، ۴۹
- مختصاتی، ۴۷
- قطری، ۲
- قطری بلوکی، ۳۷۹
- قطری شدنی، ۱۳، ۳۸۲
- قطع موفق، ۲۱۶، ۲۱۹
- قویاً همبند، ۱۰۰
- ک
- کراندار، ۳۷۴
- کرایگ، ۲۴۹
- کرایلف، ۲۰۳، ۲۸۷
- کرونکر، ۴۵
- کمترین توان‌های دوم، ۲۷، ۳۷۳
- کمینه، ۱۸۴
- کوشی-شوارتز، ۹
- کیلی، ۱۵
- گ
- گاس-سایدل، ۷۲، ۱۸۲
- گرادیان دومزدوج، ۲۶۲، ۲۶۴
- گرادیان مزدوج، ۲۲۹، ۲۳۵، ۳۴۷
- گراف، ۹۷
- گراف جهت‌دار، ۹۷
- گرام-اشمیت، ۲۷، ۲۰۸، ۲۱۰
- گرام-اشمیت آرنولدی، ۲۱۰
- ل
- لانکروس، ۲۲۹، ۲۵۵
- لانکروس دوطرفه، ۲۶۰
- لانکروس متقارن، ۲۵۵
- م
- ماتریس
- افزوده، ۴۰
- پایه، ۱۷۰
- تحویل ناپذیر، ۱۰۳
- شور، ۳۷۸
- گیونز، ۲۷
- مثبت، ۱۰۳
- منفرد، ۱۳۷
- نامنفی، ۱۰۲
- نواری، ۳
- هوسهولدر، ۲۷
- هیلبرت، ۵۸
- یکانی، ۱، ۳۵۰
- مارکف، ۱۰۲
- ماندهٔ مینیمال، ۱۹۶
- ماندهٔ مینیمال تعمیم‌یافته، ۲۰۵، ۲۱۷
- ماندهٔ مینیمال تعمیم‌یافته، ۳۴۷
- متشابه، ۱۳
- متعامد، ۱
- متعامدسازی کامل، ۲۰۵، ۲۱۴
- مقارن‌کج، ۱۳
- مقارن، ۱، ۱۳

- مقارن کج، ۱
 متلب، ۴۵
 متناوب، ۳۴۹
 مجانبی، ۳۵۰
 محدب، ۱۹۲
 محورگیری، ۴۱، ۲۳۵
 محورگیری جزئی، ۴۱، ۲۳۵
 مربع شده، ۲۷۵
 مزدوج، ۲۲۹
 مستقل خطی، ۴
 مسئله استوکس، ۳۷۵
 مسئله اوسین، ۳۷۵
 مسئله برنامه ریزی، ۳۷۱
 مسئله نقطه زینی، ۳۷۱
 معادلات خطی نرمال، ۳۷۳
 معادله پواسون، ۴۳
 معکوس تعمیم یافته، ۲۹
 معکوس تقریبی، ۳۲۶
 معکوس تقریبی تَنک، ۳۳۱
 معین مثبت، ۲۰، ۲۲۴، ۳۴۷
 معین مثبت مقارن، ۱۶، ۵۸
 معین مثبت هرمیتی، ۱۶، ۸۱
 مقدار تکین، ۲۸
 مقدار ویژه ساده، ۱۳۹
 مقدار ویژه نیمه ساده، ۱۳۹
 مقید، ۳۷۱
 مکانیک سیالات، ۳۷۱، ۳۷۴
 مکمل، ۱۱
 مکمل متعامد، ۱۱
 منظم سازی شده، ۳۹۵
 مورب، ۱۷۲
 مور-پنرز، ۲۹
 مولد، ۷
 میدان، ۳
 مینیمم درجه تقریبی مقارن، ۳۲۱
 ن
 ناپایدار، ۴۱
 نادقیق، ۳۵۶
 نامساوی کانتروویچ، ۱۹۱
 نامساوی مثلث، ۶
 نامساوی هندسی-حسابی، ۵۴
 نامعین، ۳۷۶
 نامنفی، ۱۰۲
 ناورداء، ۲۶۰
 ناویر-استوکس، ۳۷۴
 نرخ همگرایی، ۶۸، ۳۵۰
 نرخ همگرایی مجانبی، ۷۱
 نرم، ۶
 -۱، ۷
 -۲، ۷
 اقلیدسی، ۷
 برداری، ۶
 بینهایت، ۷
 طبیعی، ۷، ۳۵۸
 فروبنیوس، ۷
 ماتریسی، ۶
 نرم طبیعی، ۶۹
 نرمال، ۱، ۳۵۴، ۳۷۳
 نظریه بازی ها، ۱۰۲
 نقطه بحرانی، ۳۷۲
 نقطه زینی، ۳۷۱، ۳۷۳
 نمایش اقتصادی، ۳۲

۴۲۰ _____ روش‌های تکراری و پیش‌شرط‌سازی برای حل دستگاه معادلات خطی بزرگ و تنگ

نواری، ۳، ۳۲۶

نوسانات، ۲۷۹، ۲۸۵

نیمه‌پایدار، ۳۷۶

نیمه‌ساده، ۱۴۰

نیمه‌مثبت، ۳۷۶

نیمه‌معین مثبت متقارن، ۱۶

نیمه‌معین مثبت هرمیتی، ۱۶

نیمه‌همگرا، ۱۳۸، ۱۵۳

نیمه‌همگرایی، ۱۳۸

و

وابسته خطی، ۴

ه

هرمیتی، ۱، ۱۳

هرمیتی‌کج، ۱، ۱۳، ۳۴۸

هستینز، ۲۲۹

همبند، ۹۷

همگرایی، ۲۲۴

همیلتون، ۱۵

هندسی، ۵۴

هورنر، ۳۲۵

هیلبرت، ۵۸

ی

یال، ۹۷

یکامتعامد، ۱۰، ۲۰۸

یکانی، ۱

یکنوا، ۶۹

واژه‌نامه فارسی به انگلیسی

range	برد	mathematical induction	استقرای ریاضی
vector	بردار	perturbation	اختلال
left singular vector	بردار تکین چپ	Stokes	استوکس
right singular vector	بردار تکین راست	scaler	اسکالر
eigenvector	بردار ویژه	intersection	اشتراک
dimension	بُعد	arithmetic operations	اعمال حسابی
nullity	بعد فضای پوچ، پوچی	row elementary operations	اعمال سطری مقدماتی
block	بلوک	augmented	افزوده
maximum	بیشینه، ماکسیمم	nonzero pattern	الگوی ناصفر
invariant	پایا	error propagation	انباشتگی خطا
stable	پایدار	index	اندیس
stabilized	پایدارشده	flexible	انعطاف‌پذیر
stability	پایداری	Uzawa	اوزاوا
base	پایه	Useen	اوسین
lower triangular	پائین‌مثلثی	stationary	ایستا
Petrov	پتروف	stagnation	ایستایی
backward	پسرو	upper triangular	بالامثلثی
Perron	پرون	reordering	بازمرتب‌سازی
image processing	پردازش تصویر	upper Hessenberg	بالاهسنبرگی
Poisson	پواسون	ill-conditioned	بدحالت

algebraic	جبری	nilpotent	پوچ‌توان
Jordan	جردن	forward	پیشرو
partial	جزئی	preconditioner	پیش‌شرط‌ساز
direct sum	جمع مستقیم	preconditioning	پیش‌شرط‌سازی
exact solution	جواب واقعی	look-ahead	پیش‌نگرش
چ			
dropping	چکاندن	sign function	تابع علامت
dense	چگال	decomposition	تجزیه
density	چگالی	LU factorization	تجزیه LU
polynomial	چندجمله‌ای	singular value decomposition	تجزیه مقدار تکین
Cholesky	چولسکی	reducible	تحویل‌پذیر
ح			
memory	حافظه	irreducible	تحویل‌ناپذیر
initial guess	حدس اولیه	relaxation	تخفیف
Gaussian elimination	حذفی گاوس	transpose	ترانهاده
exact arithmetic	حساب دقیق	orthogonality	تعامد
exact arithmetic	حقیقی	generalized	تعمیم‌یافته
خ			
error	خطا	finite differences	تفاضلات متناهی
linear	خطی	iteration	تکرار
idempotent	خودالحاق	algebraic multiplicity	تکرار جبری
well-conditioned	خوش‌حالت	geometric multiplicity	تکرار هندسی
clustered	خوشه‌شده	iterative	تکراری
د			
Gerschgorin circles	دایره‌های گرشگورین	singular	تکین
determinant	دترمینان	sparse	تُنک
entry	درايه	backward substitution	جایگذاری پسرو
Drazin	درازین	forward substitution	جایگذاری پیشرو
fill-in	درج‌شده	permutation	جایگشت
		algebra	جبر

leading principle submatrix	زیرماتریس اصلی پیشرو	quadratic system	درجه دوم دستگاه
Jacobi	ژاکوبی	system of linear equations	دستگاه معادلات خطی
consistent	سازگار	two-sided	دوطرفه
steepest descent	سریع‌ترین کاهش	biconjugate	دومزدوج
steepest descent residual norm	سریع‌ترین کاهش نرم مانده	Dirichlet disk	دیریکله دیسک
tridiagonal	سه قطری	storage	ذخیره‌سازی
fluid	سیال	vertex	رأس
acceleration	شتاب	rank	رتبه
boundary conditions	شرایط مرزی	PageRank	رتبه صفحه
pseudo spectral radius	شعاع شبه طیفی	full rank	رتبه کامل
spectral radius	شعاع طیفی	rank deficient	رتبه ناقص
splitting	شکافت	trace	رد، اثر
regular splitting	شکافت منظم	projection method	روش تصویری
Schur form	شکل شور	iterative method	روش تکراری
Jordan Canonical form	شکل کانونی جردن	power method	روش توانی
Schur	شور	inverse power method	روش توانی معکوس
inner product	ضرب داخلی	direct method	روش مستقیم
multiplicative	ضربی	R-method	روش R
spectrum	طیف	Richardson	ریچاردسون
natural	طبیعی	eigenpair	زوج ویژه
condition number	عدد حالت	subspace	زیرفضا
inverse factor	عامل معکوس	Krylov subspace	زیرفضای کرایلف
		principle submatrix	زیرماتریس اصلی

Cauchy-Schwarz	کوشی-شوارتز	sign	علامت
minor	کهاد	science	علوم
	گ	adjoint operator	عملگر الحاقی
Galerkin	گالرکین	projection operator	عملگر تصویر
Gauss-Seidel	گوس-سایدل	finite elements	عناصر متناهی
conjugate gradient	گرادیان مزدوج	pivot entry	عنصر محوری
graph	گراف		غ
Gram-Schmidt	گرام-اشمیت	diagonally dominant	غالب قطری
modified Gram-Schmidt	گرام-اشمیت اصلاح‌شده	over-determined	ف dominant diagonally
Gershgorin	گرشگورین	Frobenius	فروبونیوس
group	گروه	under-determined	فرومعین
grade	گرید	vector space	فضای برداری
discrete	گسسته	nullspace	فضای پوچ
Givens	گیونز	inner product space	فضای حاصلضرب داخلی
	ل	successive overrelation	فوق تخفیف متوالی
Lagrangian	لاگرانژی		ق
Lanczos	لانکزوس	coordinate format	قالب مختصاتی
	م	diagonal	قطری
matrix	ماتریس	diagonalizable	قطری‌شدنی
augmented matrix	ماتریس افزوده	lucky breakdown	قطع موفق
block matrix	ماتریس بلوکی		ک
iteration matrix	ماتریس تکرار	descent	کاهش
Givens matrix	ماتریس گیونز	bounded	کراندار
elementary matrix	ماتریس مقدماتی	Krylov	کرایلف
nonnegative matrix	ماتریس نامنفی	Kronecker	کرونکر
Hilbert matrix	ماتریس هیلبرت	least squares	کمترین توان‌های دوم، کمترین مربعات
Markov	مارکوف	minimum	کمینه، مینیمم

engineering	مهندسی	residual	مانده
field	میدان	similar	متشابه
minimizer	مینیم‌کننده	orthogonal	متعامد
minor	مینور	symmetric	متقارن
minimal	مینیمال	skew-symmetric	متقارن کج، متقارن اریب
	ن	positive	مثبت
nonempty	نا تهی	triangle	مثلث
incomplete	ناقص	spanning set	مجموعه گسترنده
nonsingular	نامنفرد	convex	محدب
nonnegative	نامنفی	pivoting	محورگیری
norm	نرم	partial pivoting	محورگیری جزئی
Euclidean norm	نرم اقلیدسی	complex	مختلط
normal	نرمال	conjugate	مزدوج
induced norm	نرم القایی	linearly independent	مستقل خطی
infinity norm	نرم بینهایت	characteristic	مشخصه
natural norm	نرم طبیعی	equations	معادلات
matrix norm	نرم ماتریسی	inverse	معکوس
maximum norm	نرم ماکزیمم	nonsingular	معکوس‌پذیر
critical point	نقطه بحرانی	invertible	معکوس‌پذیر
saddle point	نقطه زینی	generalized inverse	معکوس تعمیم‌یافته
map	نگاشت	ordinary	معمولی
banded	نواری	definite	معین
semi-stable	نیمه پایدار	symmetric positive definite	معین مثبت متقارن
semi-simple	نیمه ساده	singular value	مقدار تکین
semi-positive	نیمه مثبت	eigenvalue	مقدار ویژه
		singular	منفرد
symmetric positive semidefinite	نیمه معین مثبت متقارن	Moore-Penrose	مور-پن‌رز
		imaginary	موهومی

semi-convergence	نیمه‌همگرایی
linearly dependent	وابسته خطی
Hermitian	هرمیتی
skew-Hermitian	هرمیتی کج ، هرمیتی اریب
identity	همانی
connected	همبند
convergent	همگرا
convergence	همگرایی
homogeneous	همگن
congruent	همنهشت
geometrical	هندسی
Householder	هوسهولدر
Hilbert	هیلبرت
edge	یال
orthonormal	یکامتعامد
unitary	یکانی
unique	یکتا
unit	یکه

واژه‌نامه انگلیسی به فارسی

A		condition number	عدد حالت
acceleration	شتاب	congruent	همنهشت
adjoint operator	عملگر الحاقی	conjugate	مزدوج
algebra	جبر	conjugate gradient	گرادیان مزدوج
algebraic	جبری	connected	همبند
algebraic multiplicity	تکرار جبری	consistent	سازگار
arithmetic operations	اعمال حسابی	convergence	همگرایی
augmented	افزوده	convergent	همگرا
augmented matrix	ماتریس افزوده	convex	محدب
B		coordinate format	قالب مختصاتی
backward	پسرو	critical point	نقطه بحرانی
backward substitution	جایگذاری پسرو	D	
base	پایه	decomposition	تجزیه
banded	نواری	definite	معین
biconjugate	دومزدوج	dense	چگال
block	بلوک	density	چگالی
block matrix	ماتریس بلوکی	descent	کاهش
boundary conditions	شرایط مرزی	determinant	دترمینان
bounded	کراندار	diagonal	قطری
C		diagonalizable	قطری‌شدنی
Cauchy-Schwarz	کوشی-شوارتز	diagonally dominant	غالب قطری
characteristic	مشخصه	dimension	بُعد
Cholesky	چولسکی	direct method	روش مستقیم
clustered	خوشه‌شده	direct sum	جمع مستقیم
complex	مختلط	discrete	گسسته

Dirichlet	دیریکله	Gaussian elimination	حذفی گاوس
disk	دیسک	Gauss-Seidel	گاوس-سایدل
Drazin	درازین	generalized	تعمیم‌یافته
dropping	چکاندن	generalized inverse	معکوس تعمیم‌یافته
E		geometrical	هندسی
edge	یال	geometric multiplicity	تکرار هندسی
eigenpair	زوج ویژه	Gershgorin	گرشگورین
eigenvalue	مقدار ویژه	Gerschgorin circles	دایره‌های گرشگورین
eigenvector	بردار ویژه	Givens	گیونز
elementary matrix	ماتریس مقدماتی	Givens matrix	ماتریس گیونز
engineering	مهندسی	grade	گرید
entry	درایه	Gram-Schmidt	گرام-اشمیت
equations	معادلات	graph	گراف
error	خطا	group	گروه
error propagation	انباشتگی خطا	H	
Euclidean norm	نرم اقلیدسی	Hermitian	هرمیتی
exact arithmetic	حساب دقیق	Hilbert	هیلبرت
exact solution	جواب واقعی	Hilbert matrix	ماتریس هیلبرت
F		homogeneous	همگن
field	میدان	Householder	هوسهولدر
fill-in	درج‌شده	I	
finite differences	تفاضلات متناهی	idempotent	خودالحاق
finite elements	عناصر متناهی	identity	همانی
flexible	انعطاف‌پذیر	ill-conditioned	بدحالت
fluid	سیال	image processing	پردازش تصویر
forward	پیشرو	incomplete	ناقص
forward substitution	جایگذاری پیشرو	index	اندیس
Frobenius	فروبنیوس	induced norm	نرم القایی
full rank	رتبه کامل	infinity norm	نرم بینهایت
G		initial guess	حدس اولیه
Galerkin	گالرکین		

inner product	ضرب داخلی	linearly independent	مستقل خطی
inner product space	فضای حاصلضرب داخلی	look-ahead	پیش‌نگرش
intersection	اشتراک	lower triangular	پائین‌مثلثی
invariant	پایا	lucky breakdown	قطع موفق
inverse	معکوس	LU factorization	تجزیه LU
inverse factor	عامل معکوس	M	
inverse power method	روش توانی معکوس	map	نگاشت
invertible	معکوس‌پذیر	Markov	مارکوف
irreducible	تحویل‌ناپذیر	mathematical induction	استقرا ریاضی
iteration	تکرار	matrix	ماتریس
iteration matrix	ماتریس تکرار	matrix norm	نرم ماتریسی
iterative	تکراری	maximum	بیشینه، ماکسیمم
iterative method	روش تکراری	maximum norm	نرم ماکزیمم
J		memory	حافظه
Jacobi	ژاکوبی	minimal	مینیمال
Jordan	جردن	minimizer	مینیم‌کننده
Jordan Canonical form	شکل کانونی جردن	minimum	کمینه، مینیمم
K		minor	کهاد
Kronecker	کرونکر	modified Gram-Schmidt	گرام-اشمیت اصلاح‌شده
Krylov	کرایلف	Moore-Penrose	مور-پن‌رز
Krylov subspace	زیرفضای کرایلف	multiplicative	ضربی
L		N	
Lagrangian	لاگرانژی	natural	طبیعی
Lanczos	لانکزوس	natural norm	نرم طبیعی
leading principle submatrix	زیرماتریس اصلی پیشرو	nilpotent	پوچ‌توان
least squares	کمترین توان‌های دوم، کمترین مربعات	nonempty	نا تهی
left singular vector	بردار تکین چپ	nonnegative	نامنفی
linear	خطی	nonnegative matrix	ماتریس نامنفی
linearly dependent	وابسته خطی	nonsingular	معکوس‌پذیر، نامنفرد

nonzero pattern	الگوی ناصفر	pseudo spectral radius	شعاع شبه‌طیفی
norm	نرم	purturbation	اختلال
normal	نرمال	Q	
nullity	بعد فضای پوچ، پوچی	quadratic	درجه دوم
nullspace	فضای پوچ	R	
O		range	برد
ordinary	معمولی	rank	رتبه
orthogonal	متعامد	rank deficient	رتبه ناقص
orthogonality	تعامد	reducible	تحویل‌پذیر
orthonormal	یکامتعامد	regular splitting	شکافت منظم
Oseen	اوسین	relaxation	تخفیف
over-determined	فرامعین	reordering	بازمرتب‌سازی
P		residual	مانده
PageRank	رتبه‌صفحه	Richardson	ریچاردسون
partial	جزئی	right singular vector	بردار تکین راست
partial pivoting	محورگیری جزئی	R-method	روش R
permutation	جایگشت	row elementary operations	اعمال سطری مقدماتی
Perron	پرون	S	
Petrov	پتروف	saddle point	نقطه زینی
Picard	پیکارد	scaler	اسکالر
pivot entry	عنصر محوری	Schur	شور
pivoting	محورگیری	Schur form	شکل شور
Poisson	پواسون	science	علوم
polynomial	چندجمله‌ای	semi-convergence	نیمه‌همگرا
positive	مثبت	semi-positive	نیمه‌مثبت
power method	روش توانی	semi-simple	نیمه‌ساده
preconditioner	پیش‌شرط‌ساز	semi-stable	نیمه‌پایدار
preconditioning	پیش‌شرط‌سازی	sign	علامت
principle submatrix	زیرماتریس اصلی	sign function	تابع علامت
projection method	روش تصویری	similar	متشابه
projection operator	عملگر تصویر		

singular	تکین، منفرد	transpose	ترانهاد
singular value	مقدار تکین	triangle	مثلث
singular value decomposition	تجزیه مقدار تکین	tridiagonal	سه قطری
skew-symmetric	مقارن کج، مقارن اریب	two-sided	دو طرفه
skew-Hermitian	هرمیتی کج، هرمیتی اریب	U	
spanning set	مجموعه گسترده	under-determined	فرومعین
sparse	تُنک	unique	یکتا
spectral radius	شعاع طیفی	unit	یکه
spectrum	طیف	unitary	یکانی
splitting	شکافت	upper Hessenberg	بالاهسنبرگی
stable	پایدار	upper triangular	بالامثلثی
stabilized	پایدار شده	Uzawa	اوزاوا
stability	پایداری	V	
stagnation	ایستایی	vector	بردار
stationary	ایستا	vector space	فضای برداری
steepest descent	سریع‌ترین کاهش	vertex	رأس
steepest descent residual norm	سریع‌ترین کاهش نرم مانده	W	
Stokes	استوکس	well-conditioned	خوش حالت
storage	ذخیره‌سازی		
successive overrelation	فوق تخفیف متوالی		
subspace	زیرفضا		
symmetric	مقارن		
symmetric positive definite	معین مثبت مقارن		
symmetric positive semidefinite	نیمه‌معین مثبت مقارن		
system	دستگاه		
system of linear equations	دستگاه معادلات خطی		
T			
trace	رد، اثر		

University of Guilan Press

**Iterative Methods and Preconditioning
For Solving Large and Sparse
System of Linear Equations**

By:

Davod Khojasteh Salkuyeh, Ph.D