

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

Základy numerické matematiky I.

Miroslav Tůma

`mirektuma@karlin.mff.cuni.cz`

Rozšířené sylaby a doplňující poznámky k předmětu, draft

7. ledna 2019

Obsah

1	Numerická matematika a její problémy	7
1.1	Čím se zabývá numerická matematika	7
1.1.1	Metodologie numerické matematiky	8
1.2	Nejdůležitější typy numerických problémů	9
1.3	Obsah tohoto kursu	9
1.4	Návaznost na předmět Lineární algebra.	10
2	Aproximace a klasifikace chyb výpočtu	15
2.1	Klasifikace chyb podle původu	15
2.2	Druhy chyb z pohledu numerického počítání	16
2.3	Způsoby (módy) analýzy/odvozování chyb	18
2.4	Shrnutí	19
3	Schurův rozklad, soustavy rovnic a problém vlastních čísel	21
3.1	Problém vlastních čísel versus rozklady	21
3.1.1	Vztah mezi rozklady matic a problémy vlastních čísel	21
3.2	Schurova věta	22
3.3	Obecně diagonalizovatelné matice	22
3.3.1	Unitární transformovatelnost	23
3.4	Tvrzení Schurovy věty	25
3.4.1	Schurova věta pro hermitovské matice	25
3.4.2	Schurova věta pro normální matice	26
3.4.3	Důsledky Schurovy věty	28
3.5	Normální matice, dyadický rozklad a aproximace	29
3.6	Rozklady a speciální matice	30
3.7	Diagonalizovatelné matice s navzájem různými vlastními čísly a vztahy některých tříd matic	30
3.7.1	Třídy matic a jejich srovnání: pouze množiny unitárních a hermitovských matic se protínají.	31
3.8	Schurova věta a reálné matice	31
3.8.1	Obecné reálné matice	31
3.8.2	Reálné normální matice	32
3.8.3	Zmínka o konstrukci Schurova rozkladu, QR algoritmus	32
3.9	Shrnutí	34

4	QR rozklad	37
4.1	Givensovy rotace	37
4.1.1	Pootočení v \mathbb{R}^2	37
4.1.2	Vnoření rotace do \mathbb{R}^n	38
4.1.3	QR rozklad pomocí Givensových rotací	39
4.1.4	Givensovy rotace a QR rozklad matice	40
4.1.5	Givensovy rotace a řídké matice	41
4.2	Householderovy reflexe	43
4.2.1	Princip zrcadlení a zobrazování vektorů zrcadlením	43
4.2.2	QR rozklad pomocí reflexí	45
4.2.3	Komplexní rozšíření Givensových rotací a Householderových reflexí	45
4.3	QR rozklad z pohledu matice jako celku	47
4.4	Gram-Schmidtův algoritmus	49
4.4.1	Základní odvození Gram-Schmidtova algoritmu	49
4.4.2	Dvě varianty Gram-Schmidtova algoritmu: CGS a MGS	50
4.4.3	Klasický Gram-Schmidtův algoritmus (CGS)	51
4.4.4	Modifikovaný Gram-Schmidtův algoritmus (MGS)	51
4.4.5	Zpřesňování algoritmu - reortogonalizace (ICGS)	52
4.5	Řešení soustavy pomocí QR rozkladu	53
4.6	QR rozklad a jeho stabilita	53
4.6.1	Aplikace rotací a reflexí	53
4.6.2	Ortogonalita spočtené matice	54
4.6.3	Stabilita výpočtu faktoru R	54
4.6.4	Norma rezídua	54
4.7	Počty operací pro jednotlivé druhy QR rozkladu	55
4.8	Arnoldiho algoritmus	55
4.8.1	Definice Krylovovy posloupnosti	55
4.8.2	Hledání ortonormální báze prostoru $K_m(A, b)$	55
4.9	Shrnutí	56
5	LU rozklad a Gaussova eliminace.	59
5.1	Standardní LU rozklad	59
5.1.1	Klíčové body v historii	59
5.1.2	Sloupcové eliminační matice a jejich inverze	59
5.1.3	Použití eliminačních matic: Gaussova eliminace a LU rozklad	60
5.1.4	Přímý a zpětný chod v řešení soustav rovnic pomocí LU rozkladu	62
5.2	LU rozklad řídkých matic	63
5.3	Rozklad s částečnou pivotací	66
5.3.1	Silná regularita	66
5.3.2	Citlivost na změny vstupů v LU rozkladu	68
5.3.3	Zpětná chyba a stabilita LU rozkladu	69
5.3.4	Zpětná stabilita LU rozkladu s pivotací (GEPP)	70
5.3.5	Škálování	71
5.3.6	Úplná pivotace	71
5.4	Choleského rozklad hermitovské PD (HPD) matice	72

5.4.1	Vztah mezi faktory L a U v LU rozkladu HPD matice	72
5.4.2	Existence Choleského rozkladu a jeho některé vlastnosti	74
5.4.3	Algoritmus Choleského rozkladu	75
5.4.4	Zpětná stabilita Choleského rozkladu	75
5.4.5	Iterační zpřesnění	76
5.4.6	Výpočetní náklady GE	77
5.5	Zmínka o srovnání LU a QR rozkladů pro řešení soustav	77
5.6	Shrnutí	77
6	Singulární rozklad matice	79
6.1	Matice jako lineární zobrazení a fundamentální podprostory	79
6.1.1	Spektrální rozklad hermitovské pozitivně semidefinitní matice	80
6.1.2	Vztah mezi spektrálním rozkladem matice A^*A a AA^*	81
6.1.3	Zápis singulárního rozkladu pomocí matic	83
6.1.4	Ekonomický tvar singulárního rozkladu	84
6.1.5	Singulární rozklad a podmíněnost	84
6.1.6	Poznámky k výpočtu singulárního rozkladu	86
6.1.7	Inverze a pseudoinverze matice	86
6.1.8	Normy a podmíněnost vyjádřené s pomocí singulárních čísel	88
6.1.9	Aproximace pomocí SVD s využitím dyadického rozvoje	89
7	Problém nejmenších čtverců	93
7.1	Dvě základní úlohy: problém a úplný problém nejmenších čtverců	93
7.1.1	Definice dvou problémů	93
7.1.2	Unitární invariance problému	94
7.2	LS a problémy ve statistice	94
7.3	Charakterizace řešení LS	94
7.4	Řešení problému nejmenších čtverců minimální v normě	95
7.4.1	Normální rovnice	96
7.5	Způsoby řešení problému nejmenších čtverců	97
7.5.1	QR rozklad	97
7.5.2	Použití normálních rovnic	98
7.5.3	Řešení rozšířeného systému	98
7.5.4	Singulární rozklad	99
8	Problém vlastních čísel	101
8.1	Částečný problém vlastních čísel	101
8.1.1	Mocninná metoda	101
8.1.2	Mocninná metoda a algoritmus pagerank	103
8.1.3	Krylovův prostor a Arnoldiho algoritmus	105
8.1.4	Lanczosův algoritmus	107
8.1.5	Porovnání Lanczosova a Arnoldiho algoritmu	109
9	Přímé metody a hermitovské indefinitní matice	111

10 Iterační metody pro řešení soustav rovnic	115
10.1 Klasické stacionární iterační metody	115
10.1.1 Základní fakta	115
10.1.2 Chyba a konvergence stacionárních iteračních metod	116
10.1.3 Jakobiho metoda	117
10.1.4 Gauss-Seidelova metoda	119
10.1.5 SOR (successive over-relaxation) metoda	121

Kapitola 1

Numerická matematika a její problémy

1.1 Čím se zabývá numerická matematika

- Vývoj a analýza (matematických) postupů pro vyřešení problémů tak, aby poskytovaly **(1) spolehlivě**, **(2) robustně** a **(3) za přijatelnou cenu aproximaci** řešení v toleranci dané zadáním.
- Velký rozvoj numerické matematiky souvisí s rozvojem výkonu a dostupnosti **výpočetní techniky**. Nicméně, mnoho aspektů numerické matematiky je teoretických a jejich studium není nutně závislé na výpočetní technice. Nicméně výpočetní technika je nutně za úvahami o metodách / postupech / algoritmech. V této souvislosti dnes hovoříme spíše o **výpočtové matematice** než o numerické matematice.
- – **(1)** Řešit problémy **spolehlivě** vzhledem k postupům (algoritmům) či modelu reality znamená znát **meze použitelnosti takových postupů**, jejich potenciální problémy a umět dobře interpretovat výsledky použitých postupů.
 - * Pro základní problémy jako je řešení soustav lineárních/nelineárních rovnic máme mnoho různých metod. Obecně výpočetní software umožňuje volbu mnoha parametrů. Musíme být schopni **zvolit spolehlivou metodu** podle vlastností soustavy jako její velikost, specifických rysů matice soustavy, podle aplikace, ve které soustava vznikla a podobně.
- – **(2)** Řešit problémy **robustně** budeme chápat jako nalezení postupů řešení pro velký rozsah možných vstupních dat.
- – **(3)** K tomu, abychom řešili problémy **za přijatelnou cenu** musíme vědět, co ona přijatelná cena postupů vlastně je. Tento aspekt tedy vede k studiu **složitosti** použitých postupů.
- Výpočtová matematika dneška představuje fúzi vědění toho, **co platí a proč to platí** a toho, **jak to nalézt**. Týká se jak matematických objektů, tak jejich

souvislosti s aplikacemi.

1.1.1 Metodologie numerické matematiky

- Metodologickým klíčem je nalezení takového **modelu problému/reality**, ze kterého je řešení odvoditelné. V praktických situacích je řešící postup obvykle **přibližný (aproximativní)**.
 - Je třeba počítat úlohy z modelů, které realitu vystihují.
 - Je třeba vyvíjet modely reality, které jsou **počítatelné** ve smyslu dříve zmíněné spolehlivosti, robustnosti a ceny.
- Analyticky je řešitelná pouze velmi malá část praktických problémů. Ale, **analytické postupy** jsou v numerické matematice podstatně přítomny jako její komponenta.

Jako příklad duality mezi modelem a výpočtovým postupem uveďme vývoj **modelů populační dynamiky**.

$$x^+ = xR(x),$$

kde x^+ je velikost populace v nové generaci, x je velikost populace v současné generaci a $R(x)$ je nějaká modelová funkce, která dotváří vztah.

- Malthusův model (Thomas Robert Malthus, An Essay on the Principle of Population (1798))

$$R(x) = r > 0.$$

- Verhulstovo započtení omezených zdrojů (Pierre-François Verhulst: Notice sur la loi que la population poursuit dans son accroissement (1838))

$$R(x) = \frac{r}{1 + xK}, \quad r > 0, K > 0.$$

- Model predátor/kořist (např. V. Volterra, Fluctuations in the abundance of a species considered mathematically (1926))

$$R(x) = \frac{rx}{1 + (x/K)^2}, \quad r > 0, K > 0.$$

S čím souvisí zmíněná přibližnost / aproximace

- Aproximace souvisí s **chybami**, které očekáváme při zadání a kterých se dopouštíme při samotném řešení řešícími postupy. Chyby při zadání očekáváme v rámci určitých **tolerancí**.
 - **Chyba** často vzniká tím, že místo nekonečně dimenzionálního popisu problému se používá jeho **konečnorozměrná reprezentace**. Typickou datovou strukturou ve výpočtech se pak stávají matice a grafy, ale v teoretických úvahách se využívá zhusta i obecnějších postupů.

- Velkým problémem aproximace je určit samotnou **toleranci**, kterou očekáváme v řešení. Ta je často dána tím, jak přesně dokážeme naměřit vstupní data. Tato tolerance na základě vstupních dat odpovídá jednomu typu chyby, který brzy zmíníme a který se nazývá zpětná chyba.

1.2 Nejdůležitější typy numerických problémů

- Výpočet **funkcionálu** $l : \mathcal{F} \rightarrow \mathbb{R}$ (komponenty derivací, integrály, normy).
- Řešení **analytických rovnic** (určení funkcí či funkčních hodnot pro dané operátorové rovnice jako jsou například obyčejné diferenciální rovnice (ODE), parciální diferenciální rovnice (PDE) či integrální rovnice).
- Řešení **optimalizačních problémů**: určení specifických funkcí nebo jejich hodnot vyhovujících předepsaným omezením, které optimalizují danou účelovou funkci.
- **Numerická lineární algebra** tvoří často základ vlastního numerického počítání v mnoha oblastech matematiky. Významná část problémů se převádí na problémy v numerické lineární algebře. Tyto postupy se někdy shrnují pojmem **matricové výpočty**. Mohutný rozvoj tohoto oboru se datuje do doby počátků rozvoje výpočetní techniky, kdy vznikly základní texty klasiků, jako byl J. von Neumann, H. Goldstine a C. Lanczos v čtyřicátých a padesátých letech dvacátého století.
- Obsahem numerické lineární algebry je především řešení **soustav algebraických rovnic** a řešení částečného či úplného **problému vlastních čísel**, jak podrobněji probereme v následující sekci.

1.3 Obsah tohoto kursu

- I. část:
 - Numerická lineární algebra: čtyři standardní **problémy**
 - * **Řešení soustav lineárních algebraických rovnic**

$$Ax = b, A \in \mathbb{C}^{n \times n}(\mathbb{R}^{n \times n}), b \in \mathbb{C}^n(\mathbb{R}^n), x \in \mathbb{C}^n(\mathbb{R}^n) \quad (1.1)$$

- Různé aplikace generují matice s různými vlastnostmi
- Problémy pak vyžadují naprosto různé přístupy rozlišované hlavně podle vlastností matice soustavy

- * **Řešení problému nejmenších čtverců**

$$\text{minimize } \|Ax - b\|, A \in \mathbb{C}^{n \times m}(\mathbb{R}^{n \times m}), b \in \mathbb{C}^n(\mathbb{R}^n), x \in \mathbb{C}^m(\mathbb{R}^m) \quad (1.2)$$

- Přeurčený problém, nedourčený problém
- Úplný problém nejmenších čtverců

* Řešení problému vlastních čísel

$$Ax = \lambda x, A \in \mathbb{C}^{n \times n}(\mathbb{R}^{n \times n}), \lambda \in C(R), x \in \mathbb{C}^n(\mathbb{R}^n), x \neq 0 \quad (1.3)$$

- Mohou být **částečné** nebo **úplné**.

Definice 1.3.1 V **částečném** problému hledáme několik, obvykle *extremálních vlastních čísel a vektorů*. V **úplném** problému hledáme *všechny*.

* Řešení problému singulárních čísel

$$A^*Ax = \lambda x, A \in \mathbb{C}^{n \times m}(\mathbb{R}^{n \times m}), \lambda \in C(R), x \in \mathbb{C}^m(\mathbb{R}^m), x \neq 0 \quad (1.4)$$

- Opět mohou být **částečné** nebo **úplné**.

– Numerická lineární algebra: dvě základní **metody**

- * Základem **iteračních** metod je vytváření **iterační posloupnosti**

$$x_0, x_1, \dots$$

do té doby, než získáme přijatelnou aproximaci řešení. Vyhodnocení, zda aproximace je přijatelná, je složitý problém, kterým se zde nebudeme zabývat.

- * **Přímé** metody umožňují zapsat řešení algebraickým výrazem.
- * Naprosto podstatnou roli hrají dva termíny zmíněné v úvodní přednášce, které mohou v různých situacích odlišný význam: **konvergence** a **stabilita**.
- * Častá je řešící **kombinace** přímé a iterační metody.

• II. část:

- Algebra nelineární, **optimalizace**: $f(x) = 0, \min f(x)$
- Konečnodimenzionální **aproximace funkcí** - interpolace
- Úlohy obsahující **integrály** – numerická integrace
- Úlohy obsahující **derivace** – numerická derivace, ODEs, PDEs item Někdy je produktem těchto úloh problém numerické lineární algebry-

1.4 Návaznost na předmět Lineární algebra.

V této sekci zmíníme a připomeneme některé **pojmy** a **postupy** z lineární algebry. Budeme si přitom odkazovat na poslední verzi skript Lineární algebra od L. Barto a J. Tůmy [1]. Ačkoli mnoho z těchto termínů bude dále osvětlováno na cvičeních, některé pojmy zde máme nashromážděny také jako výzvu k zopakování obsahu předmětu Lineární algebra. V přednášce předpokládáme jen letmý / částečný průchod tohoto seznamu.

- **Odstupňovaný tvar** ([1], str. 34) je tvarem, na který převádíme matici soustavy v Gaussově eliminaci pomocí **ekvivalentních úprav** (prohození dvou rovnic, vynásobení nějaké rovnice nenulovým číslem t , přičtení t -násobku jedné rovnice k jiné rovnici); Počet operací Gaussovy eliminace, který zmíníme, je uveden v ([1], Tvzení 2.19)
 - Pojmy pivot, hodnost matice soustavy, hodnost rozšířené matice soustavy
 - Pojmům pivotace (částečná, úplná, ([1], str. 49)) a stability řešení ([1], 2.6.1) se budeme podrobněji věnovat
 - V části naší přednášky, věnované Gaussově eliminaci, ozřejmíme pojmy špatné a dobré podmíněnosti
- **Tělesa**, nad kterými budeme diskutovat problémy numerické lineární algebry budou tělesa reálných a komplexních čísel.
- V **typech matic** budeme navazovat na ([1], kap. 4) (diagonální, permutační, trojúhelníkové) matice.
- **Otočení a zrcadlení** jako maticová zobrazení v rovině ([1], kap. 4.3). Například, bez detailů, matice určující rotaci v rovině kolem počátku o úhel ϕ v kladném směru (proti směru hodinových ručiček) je

$$\begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}.$$

- Matice **regulární, invertovatelné**, hledání inverzních matic.
- **Vektorové prostory**, direktní součet podprostorů
- **Lineární obal vektorů: span, LO**
- **Normy vektorů a matic**
- **Sloupcový prostor matice** (obor hodnot matice)

$$Im(A) \equiv Range(A) \equiv \mathcal{R}(A).$$

- **Nulový prostor matice** (jádro matice)

$$ker(A) \equiv \mathcal{N}(A).$$

- **Věta o dimenzi jádra a obrazu** ([1], Věta 5.99)
- **Jádro a obraz (obor hodnot)** lineárního operátoru
- **Determinanty** a jejich vlastnosti.
 - Determinanty jsou užitečným teoretickým nástrojem

- **Cramerovo pravidlo** vypadá pěkně, ale v praktickém životě na ně zapomeňte.
- **Skalární součin**, vyjádření projekce na podprostor, Gram-Schmidtovu ortogonalizaci budeme rozvíjet a dáme je do souvislosti s QR rozkladem
- **Unitární a ortogonální matice** ([1], Tvzení 8.82) a jejich příklady jako jsou rotace a zrcadlení
- **Metoda nejmenších čtverců** bude také probrána podrobněji. Viz ale v ([1]) souvislost s regresí ve statistických problémech a souvislost s pseudoinverzí, ke které se vrátíme.
- **Vlastní čísla a vlastní vektory** V tomto kursu navážeme na základní fakta algoritmy hledání vlastních čísel a vektorů či diskusí o vztazích mezi vlastními čísly a rozklady matic.
- **Číslo podmíněnosti** matice ([1], str. 414)
- **Podobnost** matic. Podobné matice mají stejný charakteristický polynom ([1], Tvzení 9.36).
- **Algebraickou násobností** vlastního čísla rozumíme jeho násobnost jako kořene charakteristického polynomu příslušné matice ([1], Definice 9.48)
- Čtvercová matice A řádu n je **diagonalizovatelná** právě tehdy, když existuje báze prostoru tvořená vlastními vektory matice A . ([1], Důsledek 9.60). Také: Má-li čtvercová matice A dimenze n celkem navzájem různých vlastních čísel, pak je diagonalizovatelná.
- **Geometrickou násobností** vlastního čísla čtvercové matice rozumíme dimenzi podprostoru vlastních vektorů matice A příslušných tomuto vlastnímu číslu. ([1], Definice 9.67)
- **Jordanův kanonický tvar**
 - Matice v Jordanově tvaru. Matice v Jordanově tvaru je blokově diagonální matice, jejíž bloky tvoří Jordanovy buňky. ([1], Sekce 9.4.2) Připomeňme, aniž bychom šli do detailů, že Jordanova buňka vypadá následovně

$$J_\lambda = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ & \lambda & 1 & \dots & 0 \\ & & \ddots & \ddots & 0 \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}.$$

- Pro každý operátor na konečně generovaném prostoru nad C existuje Jordanův kanonický tvar.

– zobecněné vlastní vektory

- Cayleyho-Hamiltonova věta ([1], Věta 9.119)
- Unitární a ortogonální diagonalizace. Unitární matice.
- Hermitovské a symetrické matice
 - Necht' A je komplexní čtvercová matice. Pak A je unitárně diagonalizovatelná a všechna její vlastní čísla jsou reálná právě tehdy, je-li A hermitovská. ([1], Věta 10.15)
 - Je-li A reálná, pak Matice A je ortogonálně diagonalizovatelná právě tehdy je-li symetrická. ([1], Důsledek 10.16).
- Pozitivně definitní a semidefinitní matice A , jejich vztah k A^*A .
- Singulární rozklad
- Vztah mezi Choleského rozkladem a LDL^T rozkladem.

Kapitola 2

Aproximace a klasifikace chyb výpočtu

V přibližném řešení problému (aproximaci tohoto řešení) se dopouštíme **chyb** různého typu a na různých úrovních. V dalším textu budeme tyto chyby klasifikovat nejprve podle **původu** v celém procesu řešení **problému od modelu k výpočtu** a potom z pohledu **samotného numerického počítání**.

2.1 Klasifikace chyb podle původu

- Chyba modelu

- Model obvykle vystihuje realitu jen přibližně
- Příklad: matematické kyvadlo: hmotný bod, nehmotné propojení

$$\phi'' = -\omega^2 \sin(\phi(t)), \quad \omega^2 = g/l \quad (2.1)$$

(vztah mezi úhlovou odchylkou ϕ a úhlovou frekvencí ω)

- Zanedbává se tření o vzduch a tření v uchycení (ložisku)

Existuje celý matematicko-fyzikální obor nazývaný **matematické modelování**, který se zabývá korektním návrhem modelu. Samozřejmě, návrh modelu souvisí se spoustou dalších souvisejících matematických oborů, numerickou matematiku nevyjímaje.

- Chyba dat

- Data nejsou obvykle naměřena přesně
- Příklad: Měření délky l u modelu kyvadla, gravitační konstanta různá na různých místech

Chyba v datech je často podstatná pro určení **přesnosti** výpočtu, jakkoli vágně teď o přesnosti hovoříme.

- **Chyba algoritmu**

- **Přesný algoritmus pro nalezení řešení v konečném počtu kroků nemusí existovat**
- A to i v případě postupů založených na konečnorozměrných aproximacích.
- Příklady
 - * u iteračních metod se tvoří **posloupnost přiblížení řešení** a ta se zastaví někde podle dané tolerance, aby platilo

$$Ax \approx b,$$

- * **diskretizační chyba**: zkracuje se počet členů nekonečné řady,
- * **lineární aproximace** nelineárních problémů.

S takovou to chybou samozřejmě musíme počítat a dokázat ji **pokud možno** odhadnout. Zde je prostor pro analytické postupy.

- **Zaokrouhlovací chyba** způsobená reálnými výpočetními prostředky. Souvisí s reprezentací dat v počítači.

Tuto chybu bychom mohli nazvat také chybou **praktického provedení algoritmu**. Tím, že model počítání je obvykle standardizován, její analýza je častou součástí textů, týkajících se numerické matematiky.

2.2 Druhy chyb z pohledu numerického počítání

Uvažujme následující jednoduché symbolické schéma matematizované úlohy s třemi obecnými parametry.

$$Fx = y, F : X \rightarrow Y \quad (2.2)$$

Podle toho, kterou ze tří složek vztahu x, z, F považujeme za neznámou a které zbylé považujeme za známé, zavedeme následující značení.

- Neznámou budeme označovat **s** a nazývat řešení (solution).
- Zbylé dvě složky budeme označovat **d** a nazývat datové vstupy nebo data.

S využitím tohoto značení pak budeme rozlišovat následující tři problémy:

- **Přímý problém**: Jsou dána data: $\mathbf{d} : \mathbf{F}, \mathbf{x}$ Řešení **s** pak odpovídá složka **y**. Příkladem je výpočet určitého integrálu.
- **Inverzní problém**: Jsou dána data $\mathbf{d} : \mathbf{F}, \mathbf{y}$. Příkladem je řešení **x** soustavy lineárních algebraických rovnic.
- **Identifikační problém**: Jsou dána data $\mathbf{d} : \mathbf{x}, \mathbf{y}$. Příkladem hledání zbylé složky **F** je prokládání křivky.

Abychom **formálně** popsali druhy chyb, vyjádřeme si vztah $Fx = y$ z (2.2) pouze implicitně jako vztah mezi danými **datovými vstupy** d a **řešením** s bez ohledu na to, jedná-li se o přímý problém, inverzní problém nebo problém identifikační následujícím způsobem:

$$d \rightarrow \{s \mid r(d, s) = 0\}. \quad (2.3)$$

- **Přímou chybou f (forward error)** nazveme odchylku Δs spočítaného řešení od přesného řešení s . Formálně můžeme psát

$$\hat{s} = s + \Delta s.$$

Tato chyba je nejbližší běžné intuitivní představě o chybě i názoru na to, jak chybu počítat. Jednoduše uvažujeme možné odchylky výpočtového postupu a získáme odhad přímé chyby. Ten ale může být **extrémně pesimistický** nebo obtížně získatelný. I to je důvod, proč nám přímá chyba “nestačí”.

- Příkladem přímé chyby (v inverzním problému) může být odchylka řešení soustavy $Ax = b$, kde A, b jsou maticové a vektorové datové vstupy. Vektor x je v našem značení řešení s . Přímá chyba odpovídá rozdílu mezi spočítaným \hat{x} a x pro

$$x = \hat{x} + \Delta x.$$

- **Zpětná chyba b (backward error)** je chyba řešícího postupu promítnutá do dat. Spočítáme-li místo přesného řešení s přibližné řešení \hat{s} , pak zpětná chyba odpovídá vztahu

$$r(d + b, \hat{s}) = 0.$$

Samozřejmě, nemusí být jednoduché zpětnou chybu určit. To, co je podstatné je

- nalézt **horní odhad její velikosti** (obvykle v nějaké normě, tedy v nějaké formě, která tento odhad převádí na skalár),
- určit tento odhad jako funkci **vstupních dat** na základě spočítaného přibližného řešení.

Zpětná chyba nebo alespoň její odhad dávají odpověď na otázku, jaký problém byl vlastně vyřešen. Je-li tato zpětná chyba menší než **neurčitost** v datech, například tedy, je-li menší než chyba modelování či experimentu, pak jsme získali přijatelné řešení \hat{s} . To ale předpokládá, že jsme schopni odhad zpětné chyby vyjádřit jako neurčitost v datech. Odhad zpětné chyby se získá obvykle snadněji než odhad přímé chyby, ale hlavně převádí problém chyby na něco, co známe, což jsou datové vstupy.

- Příklad: řešení soustavy $Ax = b$, A, b jsou datové vstupy. Zpětná chyba odpovídá v našem značení neurčitostem v datech A a b , které vyjádříme jako $\hat{A} = A + \Delta A$ a $\hat{b} = b + \Delta b$ a odhadneme je na základě spočítaného řešení tak, aby mohlo platit

$$\hat{A}\hat{x} = \hat{b}.$$

Tyto chyby

V souvislosti se zpětnou chybou budeme hovořit o zpětné stabilitě algoritmu. Budeme říkat, že algoritmus je **zpětně stabilní**, má-li malou zpětnou chybu. Ale to, kdy je zpětná chyba malá závisí na konkrétní úloze. Také, abychom dokázali říci, že je zpětná chyba malá, musíme ji vztáhnout k daným datům úlohy. Proto budeme koncept zpětné stability diskutovat samostatně u některých diskutovaných algoritmů a ne v této obecné části.

- Chyba **rezídua** (rezíduum) je odchylka samotného implicitního vztahu $r(d, \hat{s})$ od nuly, kterou jsme požadovali v (2.3). Často je tato chyba ta úplně nejjednodušší, která se dá spočítat.

– Příklad: řešení soustavy $Ax = b$, A, b jsou datové vstupy, chyba rezídua je dána rozdílem $A\hat{x} - b \equiv A(x + \Delta x) - b$ měřeným v nějaké normě.

Z velikosti rezídua můžeme pak často odhadnout velikost chyby, například přímé chyby.

- Speciální možnost, která ukazuje, že je vždy důležité rozlišit, co jsou data a co je hledané řešení, které je spočítáno přibližně: rezíduum může být někdy považováno za zpětnou chybu. Uvažujme řešení soustavy lineárních algebraických rovnic s výsledkem, kterým je přibližné řešení \hat{x} . Protože platí

$$A\hat{x} = b + r,$$

pak rezíduum r může být **také** považováno za zpětnou chybu, promítnutou pouze do změny datového vstupu pravé strany, ale to je pouze za velmi silného předpokladu, že matice A je přesná a je tedy pouze **konstantním členem procesu a nikoli daty**. V tomto případě by pro nás vstupní data byla pouze b .

2.3 Způsoby (módy) analýzy/odvozování chyb

Důležité rozlišování chyb je často založeno na tom, zdali je odhadujeme před výpočtem nebo až po výpočtu, kdy máme k dispozici výsledek řešení přímého, inverzního nebo identifikačního problému. Zmiňme následující dva způsoby odvozování.

- **Apriorní** odhad chyby. Je dán dopředu před výpočtem jako funkce $B(\mathbf{d})$, která je odvozená z dat, přesnosti počítače, detailů algoritmů a implementace.
 - Přímá a zpětná chyba jsou často (ale ne vždy) počítány (odhadovány) apriorně.
- **Aposteriorní** odhad chyby: Je dán aposteriorně po výpočtu jako funkce $B(\mathbf{d}, \hat{\mathbf{s}})$ odvozením z dat a dosaženého řešení s přihlédnutím k přesnosti počítače, detailům algoritmů a implementaci.
 - V aposteriorních odhadech se často používají spočtená **rezídua**, protože to jsou veličiny, které zároveň se spočteným řešením $\hat{\mathbf{s}}$ míváme k dispozici.

2.4 Shrnutí

- VZNIK CHYB V NUMERICKÉM ŘEŠENÍ MATEMATICKÝCH PROBLÉMŮ.
- PŘÍMÝ, INVERZNÍ A IDENTIFIKAČNÍ PROBLÉM.
- PŘÍMÁ CHYBA, ZPĚTNÁ CHYBA A REZÍDUUM. PŘÍKLADY NA TYTO DRUHY CHYB.
- APRIORNÍ A APOSTERIORNÍ ODHADY.

Kapitola 3

Schurův rozklad, soustavy rovnic a problém vlastních čísel

Dříve než v následujícím textu probereme jeden z centrálních klasických výsledků numerické lineární algebry, Schurovu větu (Schurův rozklad), zmíníme principiální rozdíl mezi výše zavedenými kategoriemi problémů. Konkrétně mezi rozklady na straně jedné a problémy vlastních nebo singulárních čísel na straně druhé.

3.1 Problém vlastních čísel versus rozklady

Rozdíl mezi těmito dvěma druhy problémů je možné také charakterizovat mezi řešením problémů **přímo** nebo **iteračně**.

Konvence 3.1.1 *Pokud neřekneme jinak, předpokládáme dále konvenci, že diskutované matice jsou obecně komplexní. V našem značení jsou tedy formálně prvky prostoru $\mathbb{C}^{n \times m}$ o příslušných dimenzích. Nezdůrazníme-li jinak, budeme o čtvercových maticích dále předpokládat, že jsou prvky $\mathbb{C}^{n \times n}$*

3.1.1 Vztah mezi rozklady matic a problémy vlastních čísel

- **Rozklady** jako tvoří základ Gaussovy eliminace nebo QR rozklad jsou základem přímého řešení soustav lineárních rovnic. Zároveň mají tyto rozklady v **přesné aritmetice** deterministicky omezené počty operací k jejich dosažení s horní hranicí danou polynomem nízkého stupně v dimenzi (dimenzích) matice. To ale můžeme o řadě iteračních metod říci také. Hlavní rozdíl je v tom, že u přímých metod předpokládáme, že počítané řešení všechny tyto předpokládané kroky potřebuje a řešení se dá vyjádřit algebraicky. Řešení soustav rovnic pomocí rozkladů je tedy považováno za přímou metodu.
- S **problémem vlastních čísel**, ale i se souvisejícím problémem singulárních čísel, to je složitější. Platí totiž následující věta.

Věta 3.1.1 (Abel (1824) – Ruffini (1799), moderní popis důkazu pomocí Galoi-sovy teorie) Řešení polynomiální rovnice s obecnými koeficienty a stupněm 5 nebo vyšším se nedá popsat algebraicky (v radikálech - odmocninách).

- Obecně tedy v řešení problémů vlastních čísel musíme konstruovat **přibližné rozklady** a měnit je iteračně.
- Jak se dají hledat rozklady? Pro problémy vlastních čísel například můžeme nejprve problém transformovat s využitím podobnosti.
- Jak uvidíme dále, některé rozklady pro řešení problému vlastních čísel jsou preferovanější než jiné.

3.2 Schurova věta

3.3 Obecně diagonalizovatelné matice

Definice 3.3.1 Čtvercovou matici nazveme (obecně) **diagonalizovatelnou**, je-li podobná diagonální matici.

Vztah diagonalizovatelnosti a podobnostní transformace je dán následujícím tvrzením.

Věta 3.3.1 Čtvercová matice $A \in \mathbb{C}^{n \times n}$ je **diagonalizovatelná** právě tehdy, existuje-li **ne nutně ortogonální** báze prostoru \mathbb{C}^n složená z vlastních vektorů této matice.

Důkaz: Je-li $A \in \mathbb{C}^{n \times n}$ diagonalizovatelná, pak existují diagonální matice $D = \text{diag}(d_i) \in \mathbb{C}^{n \times n}$ a regulární matice $S \in \mathbb{C}^{n \times n}$ takové, že je

$$D = S^{-1}AS. \quad (3.1)$$

Označíme-li sloupce matice S jako s_1, \dots, s_n , které tvoří bázi prostoru \mathbb{C}^n , pak můžeme rozepsat vztah diagonalizace

$$(As_1 \ \dots \ As_n) = AS = SD = (d_1s_1 \ \dots \ d_ns_n). \quad (3.2)$$

Po vyškálování sloupců matice S jejich normami dává příslušné vztahy mezi maticí A , jejími vlastními čísly a vlastními vektory a v prostoru \mathbb{C}^n tedy existuje báze z vlastních vektorů.

Opačně, vezměme bázi prostoru složenou z vlastních vektorů s_1, \dots, s_n . Pak platí

$$AS = (As_1 \ \dots \ As_n) = (\lambda_1s_1 \ \dots \ \lambda_ns_n) = (s_1 \ \dots \ s_n) \text{diag}(\lambda_i) = SD. \quad (3.3)$$

Následně, $D = S^{-1}AS$. ■

Postačující podmínka diagonalizovatelnosti je uvedena v následující větě.

Věta 3.3.2 Má-li čtvercová matice $A \in \mathbb{C}^{n \times n}$ n vzájemně různých vlastních čísel, pak je diagonalizovatelná.

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 23

Důkaz: Má-li matice $A \in \mathbb{C}^{n \times n}$ n vzájemně různých vlastních čísel $\lambda_1, \dots, \lambda_n$ se svými vlastními vektory x_1, \dots, x_n , pak musí být lineárně nezávislé. Předpokládejme, že tomu tak není a že platí

$$\dim(\text{span}(x_1, \dots, x_n)) = r < n. \quad (3.4)$$

Přechodem k vektorům docílíme toho, že prvních r vektorů z x_1, \dots, x_n je lineárně nezávislých a všechny ostatní jsou tedy jejich lineární kombinací. Platí tedy i vztah

$$x_{r+1} = \alpha_1 x_1 + \dots + \alpha_r x_r. \quad (3.5)$$

Odsud

$$Ax_{r+1} = \alpha_1 Ax_1 + \dots + \alpha_r Ax_r \quad (3.6)$$

a tedy

$$\lambda_{r+1} x_{r+1} = \alpha_1 \lambda_1 x_1 + \dots + \alpha_r \lambda_r x_r \quad (3.7)$$

Odečtením λ_{r+1} -násobku (3.5) od (3.7) získáme lineární kombinaci lineárně nezávislých vlastních vektorů x_1, \dots, x_r . To znamená, že všechny příslušné koeficienty jsou nulové. Následně, $x_{r+1} = 0$ a to je spor. Proto, $r = n$ a existuje tedy báze \mathbb{C}^n z vlastních vektorů matice. Podle Věty 3.3.1 je tedy matice diagonalizovatelná. ■

Poznámka 3.3.1 *Výše uvedená podmínka nesterlnosti vlastních čísel není nutná. Diagonalizovatelné matice mohou být i takové, které mají stejná vlastní čísla, jako je například jednotková matice. Důležité je, aby charakteristický polynom byl součinem lineárních členů, tedy, aby neexistoval žádný větší Jordanův blok v Jordanově kano-nickém rozkladu matice.*

Příslušná báze transformace může být ale **velmi špatně podmíněná** ve smyslu čísla podmíněnosti.

Poznámka 3.3.2 *Matice, které nemají dost vlastních vektorů k vytvoření báze celého prostoru, nazýváme defektní matice*

3.3.1 Unitární transformovatelnost

Viděli jsme výše, že obecná diagonalizovatelnost nemusí být nejlepším základem pro transformaci matice. Jak uvidíme, možným řešením problému s podmíněností je hledání unitární podobnosti. Důvod, proč je unitární podobnost matice velmi důležitá z numerického hlediska, si nyní vysvětlíme. Nechť $A \in \mathbb{C}^{n \times n}$ a aplikujme na tuto matici **podobnostní transformaci**.

$$B = S^{-1}AS \quad (3.8)$$

Tato podobnostní transformace může být motivována tím, že matice $B \in \mathbb{C}^{n \times n}$ je jednodušší a uvidíme v ní víc. Napišme formálně transformaci, kde počítáme s chybou v matici A . Tuto chybu můžeme formálně vyjádřit maticí $E \in \mathbb{C}^{n \times n}$ následovně

$$A = \bar{A} + E \quad (3.9)$$

24 3. Schurův rozklad, soustavy rovnic a problém vlastních čísel

Pak můžeme psát

$$B = S^{-1}AS = S^{-1}(\bar{A} + E)S = S^{-1}\bar{A}S + S^{-1}ES \quad (3.10)$$

No a teď zkusme odhadnout, jak se nám chyba transformovala podobnostní transformací. Proto, abychom ji odhadli, si nejprve připomeneme pojem maticové normy (viz [1]). Frobeniova norma matice $S \in \mathbb{C}^{n \times m}$ je definována výrazem

$$\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |a_{ij}|^2} = \sqrt{\text{trace}(A^*A)}. \quad (3.11)$$

Euklidovská (spektrální) norma je pak dána vztahem

$$\|A\|_2 = \|A\| = \max_{\|x\|_2=1} \|Ax\|_2, \quad (3.12)$$

tedy na základě euklidovských vektorových norm. Pak můžeme pro normu matice B psát

$$\|S^{-1}ES\| \leq \|S\| \|S^{-1}\| \|E\| \quad (3.13)$$

Je-li součin $\|S\| \|S^{-1}\|$ velký, chyba se nám může **výrazně zvětšit** nebo ji máme naodhadnutou, ale nevíme, který případ nastane a rozhodně nemůžeme vždy očekávat ten nejoptimističtější scénář. Navíc, výpočtový postup by pokud možno **neměl zvyšovat neurčitost**, která odpovídá vstupním datům nebo zavádět nové zdroje neurčitosti.

Definice 3.3.2 *Součin $\|S\| \|S^{-1}\|$ se nazývá číslo podmíněnosti a značí se někdy $\kappa(A)$. Platí*

$$\kappa(A) = \|A\| \|A^{-1}\| \geq \|A^{-1}A\| = 1 \quad (3.14)$$

Je-li matice S **unitární** (budeme často značit pomocí U), pak platí

$$\kappa(S) = \|S\| \|S^{-1}\| = 1. \quad (3.15)$$

V tomto případě tedy platí, že unitární podobnost podle Schurovy věty neurčitost v odhadu nezvyšuje. Číslo podmíněnosti můžeme vyjadřovat v různých normách. Uvidíme dále, že toto číslo hraje důležitou roli v řešení soustav lineárních rovnic.

Poznámka 3.3.3 *Pojem podobnosti má zajímavou geometrickou interpretaci. Konkrétně, podobnostní transformace zobrazuje objekty na objekty jim podobné, tedy takové, že zachovává úhly (**konformní zobrazení**) a všechny vzdálenosti mění ve stejném poměru.*

Nejprve uvedeme (zopakujeme) definici unitární matice:

Definice 3.3.3 *Čtvercovou matici $U \in \mathbb{C}^{n \times n}$ nazveme **unitární**, platí-li $UU^* = U^*U = I$, kde I je jednotková matice příslušné dimenze.*

Obecný tvar unitární matice nemusí být úplně zřejmý. Uveďme si dva příklady. Podle definice se dá ukázat, že unitární matice $U \in \mathbb{C}^{2 \times 2}$ má obecný tvar

$$\begin{pmatrix} a & b \\ -e^{i\phi}b^* & e^{i\phi}a^* \end{pmatrix}, \quad |a|^2 + |b|^2 = 1, \quad a, b \in \mathbb{C}, \quad \phi \in \mathbb{R}.$$

Obecná diagonální unitární matice $U \in \mathbb{C}^{n \times n}$ má pak tvar

$$U = \text{diag}(e^{i\phi_j}), \quad \phi_j \in \mathbb{R}, \quad (3.16)$$

ale můžeme se omezit na $0 \leq \phi_j < 2\pi$.

3.4 Tvrzení Schurovy věty

Věta 3.4.1 (Schurova věta) Pro libovolnou matici $A \in \mathbb{C}^{n \times n}$ existuje taková unitární matice $U \in \mathbb{C}^{n \times n}$, že matice

$$R = U^*AU \quad (3.17)$$

je horní trojúhelníková s vlastními čísly matice A na diagonále v libovolném předepsaném pořadí. Říkáme, že matice A a R , pro které platí vztah (3.17) jsou **unitárně podobné**. Rozklad (3.17) nazveme **Schurovou formou matice**.

Důkaz: Větu dokážeme indukcí podle dimenze n matice A . Pro $k = 1$ je tvrzení zřejmé. Nechť tvrzení Věty platí až do dimenze k . Uvažujme čtvercovou matici A dimenze $k + 1$ a uvažujme nějaké předepsané uspořádání jejích vlastních čísel. Nechť λ je první vlastní číslo v tomto uspořádání a nechť tedy

$$Ax = \lambda x, \|x\| = 1. \quad (3.18)$$

Nechť je dále matice $H = \begin{pmatrix} x & X \end{pmatrix} \in \mathbb{C}^{(k+1) \times (k+1)}$ unitární. Pak můžeme psát

$$H^*AH = \begin{pmatrix} x^*Ax & x^*AX \\ X^*Ax & X^*AX \end{pmatrix} = \begin{pmatrix} \lambda & x^*AX \\ 0 & X^*AX \end{pmatrix} \equiv \begin{pmatrix} \lambda & y^* \\ 0 & Y \end{pmatrix}. \quad (3.19)$$

Aplikujme nyní indukční předpoklad o Schurově rozkladu na matici Y (včetně předepsaného pořadí). Nechť V^*YV je rozklad, který splňuje tento indukční předpoklad. Pak pro

$$U = \begin{pmatrix} x & XV \end{pmatrix} = H \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix}$$

máme

$$R = U^*AU = \begin{pmatrix} 1 & 0 \\ 0 & V^* \end{pmatrix} \begin{pmatrix} \lambda & y^* \\ 0 & Y \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} \lambda & y^*V \\ 0 & V^*YV \end{pmatrix} \quad (3.20)$$

a Schurova věta je dokázána. ■

Poznámka 3.4.1 Schurův rozklad ze Schurovy věty **není jednoznačný**, což je ve Větě vyjádřeno možností libovolného uspořádání vlastních čísel na diagonále R . Také ale platí

$$A = URU^* = (UD)(D^*RD)(UD)^* \quad (3.21)$$

pro jakoukoli diagonální unitární matici.

3.4.1 Schurova věta pro hermitovské matice

Připomeňme si dvě speciální třídy matic a odpovídající varianty Schurovy věty. Nejprve to jsou matice hermitovské.

Definice 3.4.1 Čtvercovou matici $A \in \mathbb{C}^{n \times n}$ nazveme **hermitovskou**, pokud $A = A^*$. Výraz $A^* = \overline{A^T} = (\overline{A})^T$ označuje **matici konjugovanou k matici A** . Čtvercovou matici $A \in \mathbb{R}^{n \times n}$ nazveme **symetrickou**, platí-li $A = A^T$.

Pro hermitovské matice platí následující varianta Schurovy věty.

Věta 3.4.2 (Schurova věta pro hermitovské matice, První spektrální věta.) Hermitovská matice $A \in \mathbb{C}^{n \times n}$ je unitárně podobná reálné diagonální matici, to jest existují taková unitární matice $U \in \mathbb{C}^{n \times n}$ a reálná diagonální matice $D \in \mathbb{R}^{n \times n}$, že platí

$$U^*AU = D \quad (3.22)$$

Navíc, všechna vlastní čísla matice A jsou diagonální prvky D a mohou zde být obsažena v libovolném předepsaném pořadí. To mimo jiné znamená, že diagonální podobnost (diagonalizace) není jednoznačná.

Důkaz: Podle obecné Schurovy věty je matice A unitárně podobná horní trojúhelníkové matici R . Matice R je také hermitovská, protože platí

$$A = URU^*, A^* = (URU^*)^* = UR^*U^*. \quad (3.23)$$

Pak je tedy také $R = R^*$. Tuto matici, která je diagonální a reálná označíme D a věta je dokázána. ■

Uvedme příklad jednoduché komplexní hermitovské matice a jejího spektra. Matice

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

je hermitovská. Její charakteristický polynom je

$$p(\lambda) = \det(\lambda I - A) = (\lambda - 1)(\lambda + 2) - 4 = \lambda^2 + \lambda - 6 = (\lambda - 2)(\lambda + 3). \quad (3.24)$$

Spektrum matice je tedy reálné $\sigma(A) = \{-3, 2\}$. Vlastní vektory, které tvoří unitární bázi v $\mathbb{C}^{2 \times 2}$ jsou

$$\begin{pmatrix} \frac{-i}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \text{ a } \begin{pmatrix} \frac{2i}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}. \quad (3.25)$$

3.4.2 Schurova věta pro normální matice

V tomto oddíle uvedeme Schurovu větu (Druhou spektrální větu) pro matice normální. Nejprve definici.

Definice 3.4.2 Čtvercová matice $A \in \mathbb{C}^{n \times n}$ se nazývá **normální**, platí-li

$$A^*A = AA^*. \quad (3.26)$$

Schurova věta pro normální matice má následující tvar

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 27

Věta 3.4.3 (Schurova věta pro normální matice, Druhá spektrální věta.) Čtvercová matice $A \in \mathbb{C}^{n \times n}$ je **normální** právě tehdy, když je unitárně diagonalizovatelná, to jest právě tehdy, existují-li unitární matice $U \in \mathbb{C}^{n \times n}$ a diagonální matice $D \in \mathbb{C}^{n \times n}$ takové, že platí (Schurův rozklad pro normální matice)

$$U^*AU = D \quad (3.27)$$

Navíc, všechna vlastní čísla matice A jsou diagonální prvky D a mohou zde být obsažena v libovolném předepsaném pořadí.

Důkaz: Nechť je matice A **unitárně diagonalizovatelná**. Pak můžeme psát $A = UDU^*$ a dále

$$AA^* = UDU^*UD^*U^* = UDD^*U = UD^*DU^* = UD^*U^*UDU^* = A^*A \quad (3.28)$$

a matice A je tedy také normální.

Opačně, nechť je matice A **normální**. Přepišme podmínku normality s použitím Schurova rozkladu $A = URU^*$ neboli $R = U^*RU$ následovně:

$$R^*R = U^*A^*UU^*AU = U^*A^*AU = U^*AA^*U = U^*AUU^*A^*U = RR^*, \quad (3.29)$$

což dává vztah normality $RR^* = R^*R$ pro trojúhelníkovou matici R ze Schurovy věty. Rozepišme si tyto součiny

$$R^*R = \begin{pmatrix} r_{11}^* & & & \\ r_{12}^* & r_{22}^* & & \\ \vdots & \vdots & \ddots & \\ r_{1n}^* & r_{2n}^* & \dots & r_{nn}^* \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & \dots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} \quad (3.30)$$

a

$$RR^* = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & \dots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{pmatrix} \begin{pmatrix} r_{11}^* & & & \\ r_{12}^* & r_{22}^* & & \\ \vdots & \vdots & \ddots & \\ r_{1n}^* & r_{2n}^* & \dots & r_{nn}^* \end{pmatrix} \quad (3.31)$$

Porovnáním diagonálních prvků obou matic dostaneme

$$\begin{aligned} |r_{11}|^2 &= |r_{11}|^2 + |r_{12}|^2 + \dots + |r_{1n}|^2, \\ |r_{12}|^2 + |r_{22}|^2 &= |r_{22}|^2 + |r_{23}|^2 + \dots + |r_{2n}|^2, \\ &\vdots \\ |r_{1n}|^2 + |r_{2n}|^2 + \dots + |r_{nn}|^2 &= |r_{nn}|^2. \end{aligned}$$

Tyto vztahy implikují, že $r_{ij} = 0$ pro $i \neq j$ a R je tedy diagonální. ■

3.4.3 Důsledky Schurovy věty

Normální matice zahrnují dvě velmi důležité speciální třídy matic. Následující tvrzení jsou součástí základního kursu lineární algebry, ale patří k tomu nejpodstatnějším k zapamatování a tak je uvedeme. Jejich důkazy patří spíše na cvičení.

Věta 3.4.4 Čtvercová matice $A \in \mathbb{C}^{n \times n}$ je **unitární** právě tehdy, je-li normální a její vlastní čísla leží na jednotkové kružnici.

Důkaz: Je-li A unitární, je samozřejmě normální. Uvažujme pro nenulový vektor x vztah

$$Ax = \lambda x. \quad (3.32)$$

Unitární ekvivalence norem pak implikuje

$$\|x\|^2 = \|Ax\|^2 = |\lambda|^2 \|x\|^2, \quad (3.33)$$

odkud plyne, že $|\lambda| = 1$.

Opačně, z normality matice A plyne její diagonalizovatelnost, kterou můžeme zapsat ve tvaru

$$A = U\Lambda U^*.$$

Předpoklad, že její vlastní čísla, která jsou na diagonále diagonální matice Λ , leží na jednotkové kružnici můžeme zapsat

$$\Lambda^* \Lambda = I.$$

Z těchto dvou faktů plyne

$$A^* A = U\Lambda^* U^* U\Lambda U^* = U\Lambda^* \Lambda U^* = UU^* = I = U^* U = U\Lambda \Lambda^* U^* = U\Lambda U^* U\Lambda^* U^* = AA^*. \quad (3.34)$$

To vyjadřuje unitárnost matice A a tvrzení je dokázáno. ■

Přepisem diagonalizačního vztahu

$$U^* A U = D \quad (3.35)$$

na

$$AU = UD \quad (3.36)$$

získáváme následující důsledek

Důsledek 3.4.1 Čtvercová matice $A \in \mathbb{C}^{n \times n}$ je normální právě tehdy, existuje-li **ortogonální báze** prostoru \mathbb{C}^n složená z vlastních vektorů této matice.

Věta 3.4.5 Čtvercová matice A je **hermitovská** právě tehdy, je-li normální a všechna její vlastní čísla jsou reálná.

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 29

Důkaz: Nechť A je **hermitovská**, pak můžeme podle Schurovy věty psát $A = URU^*$ a tedy $A^* = UR^*U^*$. Z toho plyne, nejenom že $R = R^*$ a matice A je tedy zjevně normální, ale také její vlastní čísla, která jsou na hlavní diagonále R , jsou reálná. Tento závěr můžeme ale učinit i bez použití Schurova rozkladu následujícím způsobem. Předně, snadno vidíme, že hermitovská matice je normální. Dále, uvažujme následující sadu ekvivalencí pro vlastní číslo λ matice A a vlastní vektor x .

$$\lambda x^*x = x^*Ax = (Ax)^*x = \bar{\lambda}x^*x, \quad (3.37)$$

což implikuje, že $\lambda = \bar{\lambda}$ a λ je tedy reálné. Opačnou implikaci věty dokážeme následovně. Je-li matice A normální se všemi vlastními čísly reálnými máme podle Schurovy věty (diagonalizovatelnost)

$$A^* = UR^*U^* = URU^* = A, \quad (3.38)$$

a tvrzení je dokázáno, neboť vidíme, že A je také hermitovská. ■

3.5 Normální matice, dyadický rozklad a aproximace

Spektrální rozklad normální (tedy unitárně diagonalizovatelné) matice $A \in \mathbb{C}^{n \times n}$ lze také napsat ve tvaru

$$\begin{aligned} A = UDU^* &= U \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} U^* \\ &= \begin{pmatrix} \lambda_1 u_1 & \lambda_2 u_2 & \dots & \lambda_n u_n \end{pmatrix} \begin{pmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_n^* \end{pmatrix} = \sum_{i=1}^n \lambda_i u_i u_i^* \end{aligned}$$

Uveďme si zde některé vlastnosti, které se týkají členů dyadického rozkladu. Předně je zřejmé, že platí

$$\|\lambda_i u_i u_i^*\| = \max_{\|y\|=1} \lambda_i u_i u_i^* y = |\lambda_i|, \quad (3.39)$$

protože se maximum mezi vektory s normou 1 realizuje pro $y = u_i$.

Pro diskusi o Frobeniově normě nejprve připomeňme, že tato norma je unitárně invariantní. Konkrétně, platí

Věta 3.5.1 *Uvažujme matici $A \in \mathbb{C}^{n \times n}$ a unitární matice $U, V \in \mathbb{C}^{n \times n}$. Pak platí*

$$\|UAV\|_F = \|A\|_F. \quad (3.40)$$

30 3. Schurův rozklad, soustavy rovnic a problém vlastních čísel

Důkaz: Připomeňme, že Frobeniova norma matice $A \in \mathbb{C}^{n \times n}$ je definována následovně

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \sqrt{\text{trace}(A^*A)}. \quad (3.41)$$

Platí

$$\begin{aligned} \|UAV\|_F^2 &= \text{trace}(V^*A^*U^*UAV) = \text{trace}(V^*A^*AV) = \|AV\|_F^2 = \|V^*A^*\|_F^2 \\ &= \text{trace}(AVV^*A^*) = \text{trace}(AA^*) = \|A^*\|_F^2 = \|A\|_F^2 \end{aligned}$$

a odsud máme tvrzení. ■

Druhou mocninu Frobeniovy normy normální (tedy unitárně diagonalizovatelné) matice A pak můžeme vyjádřit následujícím způsobem

$$\|A\|_F^2 \equiv \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 = \|UDU^*\|_F^2 = \|D\|_F^2 = \sum_{i=1}^n |\lambda_i|^2 \quad (3.42)$$

jak snadno plyne právě z její unitární invariance.

Poznámka 3.5.1 Vidíme, že v případě, kdy jsou vlastní čísla na diagonále seřazena podle velikosti jejich absolutních hodnot největším počínaje, nabízí se možnost, že matici A můžeme **aproximovat** tak, že ji nahradíme součtem prvních členů jejího dyadického rozvoje. Tuto myšlenku zde nebudeme dále rozvíjet a vrátíme se k ní v kapitole o singulárním rozkladu matice.

3.6 Rozklady a speciální matice

Následující tabulku uvedeme pro připomenutí souvislostí mezi speciálními typy matic na jedné straně a některými rozklady na straně druhé.

typ matice \ typ rozkladu	Jordanův rozklad	SVD	Schurův rozklad
normální matice	UDU^*	$U\tilde{D}U^*$	UDU^*
diagonalizovatelné matice	SDS^*	$U\Sigma V^*$	URU^*
obecné matice	SJS^*	$U\Sigma V^*$	URU^*

Poznamenejme, že u SVD musí mít diagonální matice kladné prvky na diagonální matici. To vede k mírné změně faktorů, označené zde jako vztah $D \rightarrow \tilde{D}$. Blíže bude připomenuto v tomto textu později.

3.7 Diagonalizovatelné matice s navzájem různými vlastními čísly a vztahy některých tříd matic

Věta 3.7.1 Množina všech diagonalizovatelných matic s navzájem různými vlastními čísly je hustá v celém (komplexním) prostoru matic. Jinými slovy, pro každou čtvercovou matici $A \in \mathbb{C}^{n \times n}$ a každé $\epsilon > 0$ existuje diagonalizovatelná matice $A_\epsilon \in \mathbb{C}^{n \times n}$ s navzájem různými vlastními čísly splňující

$$\|A - A_\epsilon\| < \epsilon. \quad (3.43)$$

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 31

Důkaz: Podle Schurovy věty leží vlastní čísla na diagonále příslušné horní trojúhelníkové matice R Schurova rozkladu $A = URU^*$. Zvolme diagonální matici D_ϵ tak, že platí

$$\|D_\epsilon\| < \epsilon, \text{ a definujme } R_\epsilon = R + D_\epsilon. \quad (3.44)$$

a že tato matice má na diagonále různá vlastní čísla. Pak

$$\|A - UR_\epsilon U^*\| = \|URU^* - UR_\epsilon U^*\| = \|UD_\epsilon U^*\| = \|D_\epsilon\| < \epsilon \quad (3.45)$$

a vidíme, že můžeme položit $A_\epsilon = UR_\epsilon U^*$. ■

3.7.1 Třídy matic a jejich srovnání: pouze množiny unitárních a hermitovských matic se protínají.

- Obecné
- Diagonalizovatelné
- Normální (unitárně diagonalizovatelné)
- Unitární
- Hermitovské

3.8 Schurova věta a reálné matice

V této kapitole si zmíníme některé souvislosti Schurova rozkladu v reálném oboru. Nejprve ale připomeňme definici ortogonální matice.

Definice 3.8.1 Řekneme, že matice $U \in \mathbb{R}^{n \times n}$ je **ortogonální**, jestliže platí

$$U^T U = U U^T = I. \quad (3.46)$$

Ortogonální transformace jsou transformace realizované prostřednictvím ortogonálních matic.

3.8.1 Obecné reálné matice

Věta 3.8.1 (Schurova věta pro obecné reálné matice) Nechť $A \in \mathbb{R}^{n \times n}$ je obecná reálná čtvercová matice. Potom existují (reálná) ortogonální matice $U \in \mathbb{R}^{n \times n}$ a reálná blokově trojúhelníková matice $T \in \mathbb{R}^{n \times n}$ s bloky o velikosti 1×1 a 2×2 takové, že platí $T = U^T A U$.

3.8.2 Reálné normální matice

Věta 3.8.2 (Schurova věta pro reálné symetrické matice) *Nechť $A \in \mathbb{R}^{n \times n}$ je reálná symetrická matice. Potom existují (reálná) ortogonální matice $U \in \mathbb{R}^{n \times n}$ a reálná diagonální matice $D \in \mathbb{R}^{n \times n}$ takové, že platí $D = U^T A U$.*

Věta 3.8.3 (Schurova věta pro reálné normální matice) *Nechť $U \in \mathbb{R}^{n \times n}$ je reálná a obecně normální matice. To jest platí $A^T A = A A^T$. Potom existují (reálná) ortogonální matice $U \in \mathbb{R}^{n \times n}$ a reálná blokově diagonální matice $B \in \mathbb{R}^{n \times n}$ s bloky o velikosti 1×1 a 2×2 takové, že platí $B = U^T A U$.*

Bloky 1×1 a 2×2 odpovídají reálným vlastním číslům respektive dvojicím komplexně sdružených vlastních čísel. Sloupce ortogonální matice U odpovídající reálným vlastním číslům představují příslušné vlastní vektory, dvojice sloupců odpovídající párům komplexně sdružených vlastních čísel představují reálné báze odpovídajících invariantních podprostorů.

3.8.3 Zmínka o konstrukci Schurova rozkladu, QR algoritmus

Poznámka 3.8.1 *Jak bylo zmíněno výše, není přímá cesta ke Schurovu rozkladu vyjádřením v radikálech. Algoritmy mohou najít pouze přibližné řešení a to **iteračně**. K hledání Schurova rozkladu máme specifické algoritmy. Jeden z nich za chvíli zmíníme.*

Poznámka 3.8.2 *Zdůrazněme, že zde dohromady působí dvě aproximace:*

- Přibližný algoritmus pro Schurův rozklad
- Přibližné počítání na výpočetní technice

a to vede k netriviálním matematickým problémům.

Jedním ze způsobů nalezení Schurova rozkladu je QR algoritmus. Nepůjdeme zde do podrobností, ale QR rozklad známe a tím pádem konstrukce následující posloupnosti (v jednoduchém speciálním případě s ortogonálními maticemi) nám není neblízká.

$$A_0, A_1, A_2, \dots \tag{3.47}$$

následujícím způsobem

Algoritmus 3.8.1 QR algoritmus

Input: Matice $A \in \mathbb{R}^{n \times n}$.

Output: Přibližný trojúhelníkový faktor R .

1. Polož $A_1 = A$, $k = 1$
2. **while** (A_k není “dostatečně trojúhelníková”) **do**
2. Rozlož $A_k = Q_k R_k$ QR rozkladem
3. Polož $A_{k+1} = R_k Q_k$
10. **end while**
11. **return** A_k

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 33

Vztah mezi po sobě následujícími členy posloupnosti (3.47) můžeme vidět z následujícího rozpisu

$$\begin{aligned} A_{k+1} &= R_k Q_k = Q_k^* Q_k R_k Q_k = Q_k^* A_k Q_k \\ &= Q_k^* \dots Q_2^* Q_1^* A Q_1 Q_2 \dots Q_k. \end{aligned}$$

Také platí

$$\begin{aligned} Q_1 Q_2 \dots Q_{k-1} Q_k R_k R_{k-1} \dots R_1 &= Q_1 Q_2 \dots Q_{k-1} (Q_k R_k) R_{k-1} \dots R_1 \\ &= Q_1 Q_2 \dots Q_{k-1} A_k R_{k-1} \dots R_1 \\ &= (Q_1 Q_2 \dots Q_{k-1} A_k) R_{k-1} \dots R_1 \\ &= A Q_1 Q_2 \dots Q_{k-1} R_{k-1} \dots R_1 \\ &= A^k \end{aligned}$$

Označíme-li

$$\hat{Q}_k = Q_1 Q_2 \dots Q_{k-1} Q_k, \quad \hat{R}_k = R_k R_{k-1} \dots R_1, \quad (3.48)$$

pak můžeme předcházející vztahy shrnout

$$A_k = \hat{Q}_{k-1}^* A \hat{Q}_{k-1}, \quad \hat{Q}_k \hat{R}_k = A^k. \quad (3.49)$$

To mimo jiné znamená, že matice v posloupnosti (3.47) jsou navzájem **podobné** a jsou postupně ortogonálně transformované matice.

Za určitých podmínek tato posloupnost konverguje k trojúhelníkové matici. Příklad konvergence je uveden v následující Větě.

Věta 3.8.4 *Nechť matice $A \in \mathbb{R}^{n \times n}$ má vlastní čísla $\lambda_1, \dots, \lambda_n$, kde*

$$|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|. \quad (3.50)$$

a zapišme její diagonalizaci ve tvaru

$$X^{-1} A X = \Lambda \equiv \text{diag}(\lambda_1, \dots, \lambda_n). \quad (3.51)$$

Předpokládejme dále, že existuje LU rozklad $X^{-1} = LU$, kde matice L je dolní trojúhelníkový faktor s jednotkovou diagonálou. Označme $X = QR$ QR rozklad matice X . Označíme-li dále podle předcházejícího $A^k = \hat{Q}_k \hat{R}_k$ QR rozklad matice A^k , pak existují diagonální matice D_k s $|D_k| = I$ takové, že platí

$$\hat{Q}_k D_k \xrightarrow[k \rightarrow \infty]{} Q. \quad (3.52)$$

a iterace $A_k = \hat{Q}_{k-1}^ A \hat{Q}_{k-1}$ esenciálně konvergují k trojúhelníkovému faktoru Schurova rozkladu. Pojem esenciální konvergence si vysvětlíme v důkaze.*

Důkaz:

$$A^k = X\Lambda^k X^{-1} = QR\Lambda^k LU = QR(\Lambda^k L\Lambda^{-k})\Lambda^k U = QR(I + F_k)\Lambda^k U, \quad (3.53)$$

kde $F_k \xrightarrow[k \rightarrow \infty]{} 0$. protože poddiagonální prvky matice $\Lambda^k L\Lambda^{-k}$ na pozicích $(i, j), i > j$ jsou rovny $l_{ij}(\lambda_i/\lambda_j)^k$ pro $L = (l_{ij})$ a konvergují k nule pro rostoucí k . Dále můžeme psát

$$A^k = Q(I + RF_k R^{-1})(R\Lambda^k U) = Q(I + G_k)R\Lambda^k U, \quad (3.54)$$

kde $G_k \xrightarrow[k \rightarrow \infty]{} 0$. Necht' $\tilde{Q}_k \tilde{R}_k$ je QR rozklad matice $I + G_k$. Ze spojitosti QR rozkladu víme, že $\tilde{Q}_k \xrightarrow[k \rightarrow \infty]{} I$. Pak

$$A^k = Q\tilde{Q}_k \tilde{R}_k R\Lambda^k U \quad (3.55)$$

Označme $\delta_1, \dots, \delta_n$ diagonální prvky horní trojúhelníkové matice $\tilde{R}_k R\Lambda^k U$ a položme

$$D_k = \text{diag}(\delta_1^*/|\delta_1|, \dots, \delta_n^*/|\delta_n|).$$

Trojúhelníkový faktor v rozkladu

$$A^k = (Q\tilde{Q}_k D_k^{-1})(D_k \tilde{R}_k R\Lambda^k U)$$

má pak kladné diagonální prvky. Z jednoznačnosti QR rozkladu pak vidíme, že musí být

$$\hat{Q}_k D_k = Q\tilde{Q}_k D_k^{-1} D_k$$

$\hat{Q}_k D_k$ tedy konverguje k Q . Snadno nahlédneme, že Q je unitární matice Schurova rozkladu, protože platí

$$X = QR \Leftrightarrow R^{-1}Q^T AQR = \Lambda \Leftrightarrow A = QR\Lambda R^{-1}Q^T.$$

Pak ovšem iterace QR algoritmu $A_k = \hat{Q}_k^* A \hat{Q}_k$ konvergují k trojúhelníkovému faktoru Schurova rozkladu až na ono diagonální škálování. Tuto konvergenci nazveme esenciální konvergencí. Všimněme si, že v samotné esenciální konvergenci vhodné škálování nemění diagonální prvky trojúhelníkového faktoru, ale může měnit v každé iteraci mimodiagonální prvky násobky čísel s absolutní hodnotou 1. ■

3.9 Shrnutí

- PRINCIPIÁLNÍ POTŘEBA ITERAČNÍ KONSTRUKCE ROZKLADŮ PRO ŘEŠENÍ PROBLÉMŮ VLASTNÍCH ČÍSEL.
- SCHUROVA VĚTA V OBEČNÉM TVARU, SCHUROVA VĚTA PRO NORMÁLNÍ MATICE, SCHUROVA VĚTA PRO HERMITOVSKÉ MATICE.
- DŮSLEDKY SCHUROVY VĚTY: CHARAKTERIZACE HERMITOVSKÝCH A NORMÁLNÍCH MATIC.
- ZÁSADNÍ ROZDÍLY MEZI ROZKLADY, KTERÉ SE DAJÍ POUŽÍT PRO ŘEŠENÍ PROBLÉMU VLASTNÍCH ČÍSEL. DŮLEŽITOST UNITÁRNÍ TRANSFORMOVATELNOSTI.

3. Schurův rozklad, soustavy rovnic a problém vlastních čísel 35

- ROLE ČÍSLA PODMÍNĚNOSTI VE VÝBĚRU TĚCHTO MATICOVÝCH ROZKLADŮ Z NUMERICKÉHO POHLEDU.
- SCHUROVA VĚTA A REÁLNÉ MATICE.
- ZÁKLADNÍ SCHÉMA QR ALGORITMU.

Kapitola 4

QR rozklad

Tato kapitola je věnována QR rozkladu obecně obdélníkové matice A . Na rozdíl od většinu textu budeme zde nejprve uvažovat reálné matice $A \in \mathbb{R}^{n \times m}$. Důvodem je jednak velké množství aplikací v reálném oboru a jednak geometrická názornost transformací, na kterých je v tomto oboru QR rozklad založen. V následujícím textu nejprve zavedeme **dva základní typy** ortogonálních transformací. První z nich jsou **Givensovy rotace**. Druhým diskutovaným typem ortogonálních transformací jsou **Householderovy reflexe**. Obě tyto transformace byly nazvány na počest významných numerických matematiků minulého století.

4.1 Givensovy rotace

4.1.1 Pootočení v \mathbb{R}^2

Uvažujme nejprve rotace v prostoru \mathbb{R}^2 . Jak víme, **matice rotace** $G(\phi)$, která realizuje pootočení o kladný úhel ϕ (proti směru hodinových ručiček), se dá zapsat jako

$$G(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (4.1)$$

Jak tuto rotaci aplikujeme jako transformaci na vektor $x \in \mathbb{R}^2$? Uvažujme takový vektor x , který rozepíšeme po komponentách s využitím jednotkových vektorů ve směrech souřadných os.

$$x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \xi_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \xi_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \xi_1 e_1 + \xi_2 e_2 \quad (4.2)$$

Aplikace matice rotace na tento vektor pak vypadá

$$G(\phi)x = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \xi_1 G(\phi)e_1 + \xi_2 G(\phi)e_2 = \xi_1 \begin{pmatrix} \cos \phi \\ \sin \phi \end{pmatrix} + \xi_2 \begin{pmatrix} -\sin \phi \\ \cos \phi \end{pmatrix}, \quad (4.3)$$

kde jsme opět zdůraznili rozpis po komponentách. Příslušný úhel mezi danými vektory můžeme získat inverzní transformací, jak je uvedeno v následujícím tvrzení.

$$\begin{aligned}
G_{ij}(\phi)x &= G_{ij}(\phi) \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \\
&= \begin{matrix} & & i & & & & j & & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \\ i & & & & & & & & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \\ j & & & & & & & & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \\ & & & & & & & & & & \end{matrix} \begin{pmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & \cos \phi & & & & -\sin \phi & & & & \\ & & & 1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & 1 & & & & & \\ & & \sin \phi & & & & \cos \phi & & & & \\ & & & & & & & 1 & & & \\ & & & & & & & & \ddots & & \\ & & & & & & & & & 1 & \end{pmatrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \\
&= \begin{matrix} & & & & \\ i & & & & \\ & & & & \\ & & & & \\ j & & & & \\ & & & & \\ & & & & \\ & & & & \end{matrix} \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_i \cos \phi - \xi_j \sin \phi \\ \vdots \\ \xi_i \sin \phi + \xi_j \cos \phi \\ \vdots \\ \xi_n \end{pmatrix}
\end{aligned}$$

4.1.3 QR rozklad pomocí Givensových rotací

QR rozklad reálné a obecně obdélníkové matice $A \in \mathbb{R}^{n \times m}$ pomocí Givensových rotací je založen na jejich postupné aplikaci zleva na matici A takovým způsobem, že se touto aplikací nulují prvky v A . Podívejme se pro jednoduchost nejprve na situaci v \mathbb{R}^2 . Konkrétně, chceme, aby při aplikaci rotace na vektor $x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$, platil následující vztah:

$$y = G(\phi)x = G(\phi) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} \xi_1 \cos \phi - \xi_2 \sin \phi \\ \xi_1 \sin \phi + \xi_2 \cos \phi \end{pmatrix} = \begin{pmatrix} \sqrt{\xi_1^2 + \xi_2^2} \\ 0 \end{pmatrix} \quad (4.7)$$

Protože unitární transformace **nemění normu**, tak ten nevynulovaný prvek **jí musí být roven**. První otázka, kterou si musíme položit, je způsob, jak získáme **úhel** rotace odpovídající tomuto vynulování. Uvažujme následující sadu úprav, kde předpokládáme nenulovost normy vektoru x a nenulovost jeho první komponenty ξ_1 . Z rovnice pro druhé komponenty máme

$$\xi_1 \sin \phi + \xi_2 \cos \phi = 0 \quad (4.8)$$

Z toho postupně získáme i s použitím vztahu pro první komponenty

$$\begin{aligned} \tan \phi &= -\frac{\xi_2}{\xi_1} \\ \sin \phi &= -\frac{\xi_2}{\sqrt{\xi_1^2 + \xi_2^2}} \\ \cos \phi &= \frac{\xi_1}{\sqrt{\xi_1^2 + \xi_2^2}} \end{aligned}$$

Jinou možností je předpokládat ve výsledku první komponentu negativní a napsat

$$\begin{aligned} G(\phi) \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} &= \begin{pmatrix} -\sqrt{\xi_1^2 + \xi_2^2} \\ 0 \end{pmatrix} \\ \sin \phi &= \frac{\xi_2}{\sqrt{\xi_1^2 + \xi_2^2}} \\ \cos \phi &= -\frac{\xi_1}{\sqrt{\xi_1^2 + \xi_2^2}} \end{aligned}$$

Vidíme, že rotace, která nuluje příslušný prvek, není nutně jednoznačně určená. Jak dále uvidíme, použitím **(i)** Givensových rotací v \mathbb{R}^n a **(ii)** jejich aplikací k matici A v **přípustném pořadí**, je možné získat QR rozklad této matice. Co je to přípustné pořadí budeme demonstrovat dále.

4.1.4 Givensovy rotace a QR rozklad matice

Ačkoli bychom QR rozklad zde mohli diskutovat pro obecně obdélníkové matice, pro názornost se soustředíme na matice čtvercové, tedy případ $n = m$.

QR rozklad matice $A \in \mathbb{R}^{n \times n}$ založený na Givensových rotacích si můžeme představit následujícím způsobem. Uvažujme nejprve v \mathbb{R}^n rotaci v rovině určené jednotkovými vektory e_1 a e_2 . Tuto transformaci označíme $G_{1,2}$ a její aplikaci si schématicky zapíšeme jako

$$G_{1,2}A = G_{1,2} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} \bar{a}_{11} & \bar{a}_{12} & \dots & \bar{a}_{1n} \\ 0 & \bar{a}_{22} & \dots & \bar{a}_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \quad (4.9)$$

V případě, že platilo $a_{11} = 0$, matice transformace (redukována jen na dva směry) je

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (4.10)$$

a jejím výsledkem je pouze přehození příslušných řádků. Tím vlastně řešíme speciální případ vyloučený výše. Jsou-li nulové oba prvky a_{11} i a_{22} , transformací je identita, kterou v praxi nemusíme samozřejmě vůbec aplikovat.

Poznámka 4.1.1 *Vidíme, že $G_{1,2}$ ovlivní prvky v řádcích 1 a 2, ale ne v jiných řádcích.*

QR rozklad matice (zatím jej chápeme v intuitivním smyslu, jak si jej vybavujeme z kursu lineární algebry) je založen na tom, že postupnou aplikací rotací na matici získáme matici, která má všechny prvky pod hlavní diagonálou nulové. Přípustné pořadí je takové, které toho docílí. V praxi to může znamenat například takové pořadí, které nemění dříve vynulované prvky zpět na nenulové. Označíme-li výslednou matici $R \in \mathbb{R}^{n \times n}$ a označíme-li rotace v pořadí, jak je aplikujeme, Q_1, \dots, Q_s , $s \leq (n-1)n/2$, pak QR rozklad můžeme psát

$$Q_s \dots Q_1 A = R. \quad (4.11)$$

Má-li R všechny prvky pod hlavní diagonálou nulové, pak součin unitárních matic $Q_s \dots Q_1$ je unitární a můžeme jej označit Q^T . Ve výsledku jsme takto našli rozklad

$$A = QR. \quad (4.12)$$

Zastavme se ještě u možných přípustných pořadí. Jedním z nich je například postupné nulování prvků matice A na pozicích

$$(2, 1), (3, 1), \dots, (n, 1), (3, 2), \dots, (n, 2), \dots, (n, n-1). \quad (4.13)$$

Toto ale není ale jediné možné pořadí. Klíčem k tomu, která pořadí jsou přípustná, je, abychom dospěli k matici R , která má všechny prvky pod hlavní diagonálou nulové.

4.1.5 Givensovy rotace a řídké matice

Situace je mnohem zajímavější v případě, že již matice A má mnoho prvků nulových. Matice, kde jsme schopni využít faktu nulovosti prvků matice k zefektivnění výpočtových postupů, nazýváme matice **řídké**. V tomto případě je možné řídkosti využít například k výraznému zmenšení počtu s použitých rotací. Zajímavé je také studovat, **jak a kde** se změní nulovost a nenulovost prvků matice po aplikaci jedné nebo více Givensových rotací.

Pozorování 4.1.1 *Při aplikaci Givensovy rotace, která působí v rovině určené vektory e_i a e_j na matici $A \in \mathbb{R}^{n \times m}$, je struktura řádků i a j ve výsledné matici dána sjednocením struktur v obou z nich.*

Toto pozorování se dá využít k posouzení řídkosti QR rozkladu řídké matice A . Diskutujme tento rozklad podrobněji. V QR rozkladu řídké matice hrají roli dva následující základní aspekty.

- Za prvé, řídkost výsledného faktoru R ,
- za druhé, počet operací během QR rozkladu.

První z těchto aspektů je jasný. Faktor R může obecně obsahovat některé nenulové prvky na pozicích, které byly v matici A původně nenulové. Takové prvky budeme nazývat **prvky zaplnění**. V praxi bude těchto prvků obecně mnoho a pozorování výše ukazuje princip zvětšování jejich množství sjednocováním struktur řádků zapojených v rotacích. Tento aspekt QR rozkladu řídkých matic se dá formalizovat ještě jiným způsobem. Uvažujme permutační matici $P \in \mathbb{R}^{n \times n}$ a QR rozklad matice PA , kde $A \in \mathbb{R}^{n \times m}$ má pro jednoduchost lineárně nezávislé sloupce. Pak platí

$$A^T A = A^T P^T P A = (QR)^T QR = R^T R. \quad (4.14)$$

Vzhledem k jednoznačnosti QR rozkladu za předpokladu kladných diagonálních prvků R vidíme následující

- Faktor R je při přesném počítání roven Choleskému faktoru matice $A^T A$.
- Permutace řádků, kterou vyjadřuje permutační matice P , nehraje ve **výsledném** faktoru R žádnou roli.

Problém faktoru R v QR rozkladu, který je nahlížen jako faktor Choleského faktoru matice $A^T A$, můžeme vidět na extrémním příkladě, kdy matice A obsahuje husté řádky. Uvažujme následující strukturu matice A

$$\begin{pmatrix} * & & & & & \\ & * & * & & & \\ & & * & & & \\ & & & * & & \\ * & * & * & * & * & \end{pmatrix}. \quad (4.15)$$

Je zřejmé, že matice $A^T A$ není obecně řídká, neboť je úplně zaplněná. To může vést k velké výpočetní složitosti ve výpočtu QR rozkladu. A to také snadno vidíme, například, z aplikace Givensových rotací.

Ačkoli tedy rozklad je nezávislý na řádkové permutaci matice A , tedy na permutační matici P , tato matice hraje velkou roli v praktické aplikaci Givensových rotací. Konkrétně, promyšlená volba pořadí Givensových rotací umožňuje zmenšit počet operací během QR rozkladu, což se týká druhého zmíněného aspektu. Obecnou důležitou otázkou je ale jaké přípustné pořadí použít, aby počet aritmetických operací v QR rozkladu byl co nejmenší. Nepůjdeme zde do detailů, ale existenci rozdílů v počtu operací ukážeme na příkladě následující struktury řídké matice, kde hvězdičky značí nenulové prvky.

$$\begin{pmatrix} * & * & * \\ & * & * \\ * & & * \\ * & & * \\ * & & * \\ * & & * \\ * & & * \end{pmatrix} \quad (4.16)$$

Je jasné, že když pro eliminaci nenulových prvků v prvním sloupci nejprve použijeme rotace s využitím třetího řádku

$$G_{34}, G_{35}, G_{36}, G_{37}, G_{38}, G_{13}$$

bude celkový potřebný počet operací menší, než když použijeme nejprve rotace s použitím prvního řádku

$$G_{13}, G_{14}, G_{15}, G_{16}, G_{17}, G_{18}.$$

Použití prvního řádku pro eliminaci totiž (jak je vidět z výše uvedeného pozorování) dá vzniknout novým nenulovým prvkům ve druhém sloupci, které pak musí být dále eliminovány.

4.2 Householderovy reflexe

Stejně jako v případě Givensových rotací se zde omezíme na těleso reálných čísel a budeme uvažovat transformace pouze v prostoru \mathbb{R}^n . Jak uvidíme, na rozdíl od Givensových rotací uvažujeme tyto nové transformace rovnou v obecné dimenzi n .

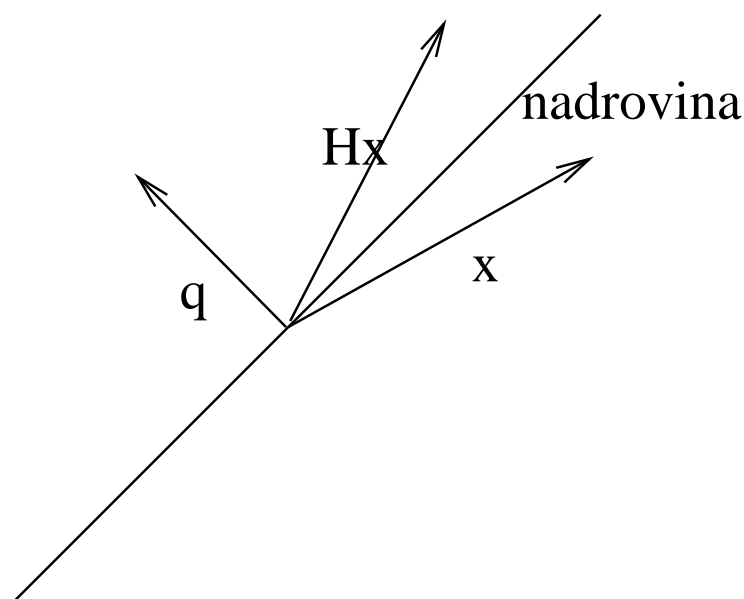
4.2.1 Princip zrcadlení a zobrazování vektorů zrcadlením

Uvažujme nadrovinu \mathcal{H} v \mathbb{R}^n . Víme, že nadrovina má dimenzi $n - 1$ a popíšeme ji jejím normálovým vektorem q , $\|q\| = 1$ následujícím způsobem.

$$\mathcal{H}(q) = \{z \in \mathbb{R}^n \mid z \perp q\} \quad (4.17)$$

Pro obecný vektor $x \in \mathbb{R}^n$ si vyjádříme jeho obraz při zrcadlení podle této nadroviny, kterou budeme také nazývat nadrovinou zrcadlení. Příslušný postup teď popíšeme v několika logických krocích.

- Nejprve si vektor x rozložíme na složku ve směru normálového vektoru, která je rovna $x_q = qq^T x$. Matice qq^T přitom reprezentuje projekci na normálový vektor q nadroviny \mathcal{H} .
- Složka kolmá na vektor q (čili rovnoběžná s nadrovinou) je $x - x_q$.
- **Zrcadlený** vektor bude mít druhou z těchto složek, to jest $x - x_q$, shodnou s původním vektorem, ale ve směru rovnoběžném s q má místo x_q vektor $-x_q$.



Obrázek 4.2.1: Ilustrace Householderovy reflexe vektoru x označené zde jako Hx nadrovinou s normálovým vektorem q v \mathbb{R}^2 .

Dohromady tedy lze zrcadlený vektor a příslušnou transformaci $H(q)$ zapsat následovně

$$y = H(q)x = (x - x_q) - x_q = (I - 2qq^T)x. \quad (4.18)$$

Snadno nahlédneme, že matice $H(q)$ je ortogonální, symetrická a také splňuje

$$H^2(q) = I, \quad (4.19)$$

protože, celkem logicky, zrcadlíme-li dvakrát, jako kdybychom nezrcadlili vůbec. Schéma zrcadlení je znázorněno následujícím schématem.

Proces zrcadlení mezi jedním a druhým vektorem o stejné normě je formálně popsán v následující větě tak, že příslušné nadroviny, reprezentované svými normálovými vektory, jsou zde zkonstruovány.

Věta 4.2.1 (Věta o zrcadlení) *Mějme dva různé vektory $x, y \in \mathbb{R}^n$ se stejnou normou. Nechť*

$$q_1 = \frac{x - y}{\|x - y\|}, \quad q_2 = \frac{x + y}{\|x + y\|} \quad (4.20)$$

Pak $H(q_1)x = y$, $H(q_2)x = -y$.

Důkaz: Vektor $x - y$ je kolmý k nadrovině zrcadlení vektoru x na vektor y . Normalizací získáme q_1 . Podobně, vektor $x + y$ je kolmý k nadrovině zrcadlení vektoru x na vektor $-y$. ■

Stejně jako v případě Givensových rotací, zrcadlení lze použít k hromadnému nulování prvků matice a následně tak získat QR rozklad.

Uvažujme vektor $x \in \mathbb{R}^n$ a zrcadlení takové, že jej zobrazí na vektor $\pm\|x\|e_1$. To tedy znamená, že chceme najít nadrovinu zrcadlení, kde zrcadlený vektor bude splňovat

$$H(x) = \pm\|x\|e_1 \quad (4.21)$$

Příslušné normálové vektory z předcházející věty potom jsou dány vztahy

$$q_1 = \frac{x - \|x\|e_1}{\|x - \|x\|e_1\|}, \quad q_2 = \frac{x + \|x\|e_1}{\|x + \|x\|e_1\|}. \quad (4.22)$$

Otázkou zůstává, kterou z těchto variant v praxi zvolit. Volba v praxi musí vzít do úvahy fakt, že výpočty se provádí v aritmetice, která má pouze konečnou přesnost. V ní platí, že přesnost výsledku můžeme významně ztratit **vyrušením** při odečítání blízkých čísel. Aby k tomu nedošlo, volíme znaménko podle komponenty ξ_1 vektoru x . To jest, nadrovinu zrcadlení určenou vektorem q_1 použijeme v případě, kdy $\xi_1 < 0$ a q_2 použijeme v případě, kdy $\xi_1 > 0$.

4.2.2 QR rozklad pomocí reflexí

Stejně jako v případě Givensových rotací je QR rozklad pomocí Householderových reflexí založen na jejich postupné aplikaci zleva na matici A takovým způsobem, že se touto aplikací nulují prvky této matice. Na rozdíl od nich je však přípustné pořadí obvykle zřetelněji určeno. Standardní postup nuluje postupně poddiagonální prvky ve sloupcích $1, \dots, n-1$ matice A tak, abychom získali matici R . Použili-li jsme $n-1$ reflexí Q_1, \dots, Q_{n-1} , pro každý sloupec jednu, QR rozklad pak můžeme psát

$$Q_{n-1} \dots Q_1 A = R. \quad (4.23)$$

Součin unitárních matic reflexe $Q_{n-1} \dots Q_1$ je unitární a můžeme jej označit Q^T a stejně jako v případě rotací jsme našli rozklad

$$A = QR. \quad (4.24)$$

V případě, kdy je matice A řídká, je problém nulovosti a nenulovosti prvků jak v R tak i v mezivýsledcích QR rozkladu, kdy jsme na A aplikovali pouze část rotací, také velmi zajímavý. Obecně je vliv reflexí na řídké matice jasný z jejich 'matematického vyjádření.

4.2.3 Komplexní rozšíření Givensových rotací a Householderových reflexí

- Tato rozšíření nejsou jednoznačná, ale jsou. Je tedy možné je aplikovat na obecně komplexní matice. Jedno z možných zobecnění Givensových rotací není založeno na geometrickém pojmu rotace, ale využívá matici

$$\begin{matrix} & & & i & & j \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ i & & & & & & \\ & & & & & & \\ & & & & & & \\ j & & & & & & \\ & & & & & & \end{matrix} \begin{pmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & c & & & & & & & & -s \\ & & & 1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & 1 & & & & & \\ & & & \bar{s} & & & \bar{c} & & & & \\ & & & & & & & 1 & & & \ddots \\ & & & & & & & & & & & 1 \end{pmatrix}, \quad (4.25)$$

kde $|c|^2 + |s|^2 = 1$. Za účelem nulování pak můžeme položit

$$s = -\frac{\bar{\xi}_2}{\sqrt{|\xi_1|^2 + |\xi_2|^2}}$$

$$c = \frac{\bar{\xi}_1}{\sqrt{|\xi_1|^2 + |\xi_2|^2}}$$

a dostaneme

$$y = G_{1,2}x = G_{1,2} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} \xi_1 c \phi - \xi_2 s \phi \\ \xi_1 \bar{s} \phi + \xi_2 \bar{c} \phi \end{pmatrix} = \begin{pmatrix} \sqrt{|\xi_1|^2 + |\xi_2|^2} \\ 0 \end{pmatrix} \equiv \begin{pmatrix} \|x\| \\ 0 \end{pmatrix}. \quad (4.26)$$

$$\mathcal{H}(q) = \{z \in \mathbb{C}^n \mid z^* q = 0\} \quad (4.27)$$

$$H(q) = (I - 2qq^*). \quad (4.28)$$

- $x_q = qq^*x$ leží ve směru normály $\Leftrightarrow q^*x$ je reálné (že v tom směru tedy obecně neleží vždycky)

Věta 4.2.2 (Věta o komplexním zrcadlení) *Mějme dva různé vektory x a y se stejnou normou v \mathbb{C}^n . Nechť*

$$q_1 = \frac{x - y}{\|x - y\|}, \quad q_2 = \frac{x + y}{\|x + y\|} \quad (4.29)$$

*Pak $H(q_1)x = y$, $H(q_2)x = -y$ právě tehdy, když y^*x je reálné.*

No ale toho můžeme vždycky docílit. Je-li $\xi_1 = |\xi_1|e^{i\alpha}$, a zvolíme-li $y = -e^{i\alpha}\|x\|e_1$, pak je

$$y^*x = -|\xi_1|\|x\|$$

a skalární součin je přitom reálný. Pak tedy pro

$$q = \frac{x + e^{i\alpha}\|x\|e_1}{\|x + e^{i\alpha}\|x\|e_1\|} \quad (4.30)$$

platí

$$H(q)x = e^{-i\alpha}\|x\|e_1 \quad (4.31)$$

- Výpočet je numericky stabilní, protože první prvek vektoru q je

$$\frac{(|\xi_1| + \|x\|)e^{i\alpha}}{\|x + e^{i\alpha}\|x\|e_1\|}. \quad (4.32)$$

4.3 QR rozklad z pohledu matice jako celku

Jak jsme viděli výše, QR rozklad je mimo jiné **základním kamenem** nalezení Schurova rozkladu, ale jeho role je významná i v dalších úlohách. Ačkoli jsme jej výše popsali jeho dva základní přístupy, zmiňme si ještě QR rozklad pro obecnou obdélníkovou matici. Třetí přístup, který zde uvedeme, je přirozeně spjat s obdélníkovými maticemi.

Definice 4.3.1 *Nechť $A \in \mathbb{C}^{n \times m}$ je obecná obdélníková matice. Rozklad tvaru $A = QR$, kde Q je matice s ortonormálními (ON) sloupci a R má všechny prvky pod hlavní diagonálou nulové, to jest kde $R = [r_{ij}]$, $r_{ij} = 0$ pro $i > j$, nazýváme QR rozkladem matice A .*

V praxi můžeme rozlišovat více případů podle vztahu dimenzí n a m . V případě, kdy $n \geq m$, může mít matice R pod svou hlavní diagonálou řádky se samými nulovými prvky. Následující obrázek nejprve znázorňuje schématicky tvar rozkladu pro $n \geq m$ a potom pro případ $n \leq m$. Samozřejmě, v případě rovnosti n a m oba případy splývají.

Někdy hovoříme o **ekonomickém** QR rozkladu, kde místo matice R máme ve výsledku čtvercovou matici $R_1 \in \mathbb{R}^{m \times m}$ bez oněch nulových řádků a místo matice $Q \in \mathbb{R}^{n \times n}$ uvažujeme matici $Q_1 \in \mathbb{R}^{n \times m}$ složenou z prvních n sloupců Q . V případě $n \leq m$ získáme takzvaný široký rozklad, kde speciální ekonomický tvar není třeba uvažovat.. Schéma ekonomického QR rozkladu pro $n \geq m$ je na následujícím obrázku.

Poznámka 4.3.1 *QR rozklad není obecně jednoznačný, neboť pro každou diagonální matici D , pro kterou $D^*D = I$, a pro libovolný QR rozklad matice A této dimenze platí*

$$A = QR = QD^*DR = (QD^*)(DR)$$

Když však povolíme na diagonále matice R pouze kladné prvky, je ekonomický QR rozklad jednoznačný a tvrzení hned formulujeme jako větu.

Věta 4.3.1 *Nechť $A \in \mathbb{C}^{n \times m}$ je obecná obdélníková matice s lineárně nezávislými sloupci. Pak existuje jediná dvojice matic $Q \in \mathbb{C}^{n \times m}$ a $R \in \mathbb{C}^{m \times m}$ taková, že Q je matice s ortonormálními sloupci a R je horní trojúhelníková matice s kladnými diagonálními prvky splňující*

$$A = QR.$$

$$\begin{array}{c}
 \boxed{\text{A}} = \boxed{\text{Q}} \begin{array}{|c|} \hline \text{R} \\ \hline \text{0} \\ \hline \end{array} \\
 \\
 \boxed{\text{A}} = \boxed{\text{Q}} \begin{array}{|c|} \hline \text{R} \\ \hline \text{0} \\ \hline \end{array}
 \end{array}$$

Obrázek 4.3.2: Ilustrace QR rozkladu pro případ $n \geq m$ (nahore) a $n \leq m$ (dole).

$$\boxed{\text{A}} = \boxed{\text{Q}} \begin{array}{|c|} \hline \text{R} \\ \hline \end{array}$$

Obrázek 4.3.3: Ilustrace ekonomického QR rozkladu pro případ $n \geq m$.

4.4 Gram-Schmidtův algoritmus

Gram-Schmidtův algoritmus je proces, kterým můžeme také nalézt QR rozklad matice. Tím, že jej obvykle popisujeme jinak než jako aplikaci transformační matice, věnujeme mu speciální pozornost a popisujeme jej samostatně. Zde jej uvažujeme pro obecně komplexní vektory. Navíc, jeho provedení má více variant, které se výrazně liší numerickými vlastnostmi a tak jeho podrobnější algoritmický popis je rozumné provést samostatně.

4.4.1 Základní odvození Gram-Schmidtova algoritmu

Definice 4.4.1 *Gram-Schmidtův algoritmus je proces, který nalezne ortonormální bázi podprostoru generovaného danými a obecně komplexními vektory a_1, \dots, a_m . Schématicky budeme psát pro $k = 1, \dots, m$ přerod z daných vektorů a_1, \dots, a_k na ortonormální bázi*

$$\text{span}\{a_1, \dots, a_k\} \rightarrow \text{span}\{q_1, \dots, q_k\}, \quad (4.33)$$

kde q_1, \dots, q_k jsou ortonormální.

- $r_{11} = \|a_1\|$, $q_1 = a_1/r_{11}$ (inicializace, $\|q_1\| = 1$)
- Jak nalézt druhý vektor? Má být

$$\begin{aligned} \text{span}\{a_1, a_2\} &= \text{span}\{q_1, q_2\} \\ q_2 &\perp q_1 \end{aligned}$$

- Promítněme a_2 na q_1 a tu projekci, která je ve směru vektoru q_1 , odečteme od a_2 . Výsledek pak normalizujeme a dostaneme

$$a_2 - q_1 q_1^* a_2 = (I - q_1 q_1^*) a_2 \quad (4.34)$$

- Normalizace a normalizovaný vektor

$$r_{22} = \|(I - q_1 q_1^*) a_2\|, \quad q_2 = (I - q_1 q_1^*) a_2 / \|(I - q_1 q_1^*) a_2\|, \quad r_{12} = q_1^* a_2 \quad (4.35)$$

- No a takhle pokračujeme s tím, že promítáme na podprostor z předcházejících ortonormálních vektorů q_1, \dots, q_{k-1} následujícím způsobem

$$\begin{aligned} Q_{k-1} &= [q_1, \dots, q_{k-1}] \\ z &= (I - Q_{k-1} Q_{k-1}^*) a_k = a_k - \sum_{i=1}^{k-1} (q_i^* a_k) q_i \\ r_{kk} &= \|z\|, \quad q_k = z / \|z\|, \quad r_{ik} = q_i^* a_k \end{aligned}$$

- Lineárně nezávislé vektory a_1, \dots, a_m implikují $\|z\| \neq 0$, protože vždycky odečítáme jen složky ve směru předchozích vektorů.
- Z hodnot r_{ik} se dá vytvořit horní trojúhelníková matice

$$A = [a_1, \dots, a_m], Q = [q_1, \dots, q_m], R = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ & r_{22} & \dots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{pmatrix} \quad (4.36)$$

- No a máme zase QR rozklad

4.4.2 Dvě varianty Gram-Schmidtova algoritmu: CGS a MGS

V praxi rozlišujeme dvě základní varianty Gram-Schmidtovy ortogonalizace podle toho, jakým způsobem jsou nové sloupcové vektory ortogonalizovány. Tyto varianty dávají v **přesné počítačové aritmetice** shodné výsledky, ale v konečné přesnosti se chovají velmi odlišně. První z těchto variant, kterou nazveme **klasickou Gram-Schmidtovou** ortogonalizací (CGS) postupuje tak, jak jsme o ortogonalizaci hovořili výše. Konkrétně, nový vektor q_{k+1} vznikne normalizací vektoru z definovaného takto

$$z = \left(I - \sum_{i=1}^{k-1} q_i q_i^* \right) a_k \equiv \left(I - Q_{k-1} Q_{k-1}^* \right) a_k \equiv C_Q^k a_k \quad (4.37)$$

V tomto případě ortogonalizujeme nový vektor **nezávisle** proti předcházejícím ortonormálním vektorům. Aplikace modifikované Gram-Schmidtovy ortogonalizace (MGS), která je v přesné aritmetice matematicky ekvivalentní CGS, se zapíše

$$\left(I - q_{k-1} q_{k-1}^* \right) \dots \left(I - q_1 q_1^* \right) a_k \equiv M_Q^k a_k. \quad (4.38)$$

Proces tedy probíhá tak, že **mezivýsledek** ortogonalizace nějakým vektorem tvořené báze je dále postupně ortogonalizován proti dalším vektorům báze.

Věta 4.4.1 Pro výše zavedené operátory C_Q^k a M_Q^k platí pro $k \geq 2$

$$C_Q^k = M_Q^k. \quad (4.39)$$

Důkaz: Rovnost je jasná pro $k = 2$ neboť obě strany tvrzení vyjadřují totéž. Necht' $k > 2$, pak uvažujme indukčního předpoklad

$$M_Q^{k-1} = C_Q^{k-1},$$

který můžeme také zapsat

$$\left(I - q_{k-2} q_{k-2}^* \right) \dots \left(I - q_1 q_1^* \right) = \left(I - \sum_{i=1}^{k-2} q_i q_i^* \right). \quad (4.40)$$

Úpravou levé strany a použitím indukčního předpokladu postupně dostáváme

$$\begin{aligned}
 M_Q^k &= (I - q_{k-1}q_{k-1}^*) \dots (I - q_1q_1^*) \\
 &= (I - q_{k-1}q_{k-1}^*) \left(I - \sum_{i=1}^{k-2} q_iq_i^* \right) \\
 &= \left(I - \sum_{i=1}^{k-2} q_iq_i^* \right) - q_{k-1}q_{k-1}^* - q_{k-1}q_{k-1}^* \sum_{i=1}^{k-2} q_iq_i^* \\
 &= \left(I - \sum_{i=1}^{k-2} q_iq_i^* \right) - q_{k-1}q_{k-1}^* = C_Q^k.
 \end{aligned}$$

Z ortogonality q_{k-1} vůči předchozím vektorům tedy dostáváme platnost indukčního kroku a lemma je dokázána. ■

V dalším textu tyto dva způsoby ortogonalizace popíšeme i formou algoritmických schémat.

4.4.3 Klasický Gram-Schmidtův algoritmus (CGS)

Algoritmus 4.4.1 CGS

Input: $A = [a_1, \dots, a_m]$

Output: QR CGS rozklad.

0. **Inicializace:** $r_{11} = \|a_1\|$, $q_1 = a_1/r_{11}$, $Q_1 = [q_1]$
1. **for** $k = 2 : m$ **do**
2. $z = a_k$
3. $[r_{1k}, \dots, r_{k-1,k}]^T = Q_{k-1}^* z$ (nezávislé počítání koeficientů)
4. $z = z - Q_{k-1}[r_{1k}, \dots, r_{k-1,k}]^T$
5. $r_{kk} = \|z\|$
6. $q_k = z/r_{kk}$
7. $Q_k = [Q_{k-1}, q_k]$
8. **end do**

4.4.4 Modifikovaný Gram-Schmidtův algoritmus (MGS)

Algoritmus 4.4.2 MGS

Input: $A = [a_1, \dots, a_m]$

Output: QR MGS rozklad.

0. **Inicializace:** $r_{11} = \|a_1\|$, $q_1 = a_1/r_{11}$, $Q_1 = [q_1]$
1. **for** $k = 2 : m$ **do**
2. $z = a_k$
3. **for** $i = 1 : k - 1$ **do**
4. $r_{ik} = q_i^* z$
5. $z = z - r_{ik}q_i$

6. **end do**
7. $r_{kk} = \|z\|$
8. $q_k = z/r_{kk}$
9. $Q_k = [Q_{k-1}, q_k]$
10. **end do**

4.4.5 Zpřesňování algoritmu - reortogonalizace (ICGS)

S kvalitou ortogonalizace může být problém v tom, že výsledné vektory q_1, \dots, q_m nemusí být příliš ortogonální při implementaci v aritmetice s konečnou přesností. Jak dále uvidíme, problém vzniká především u varianty CGS. Kvalitu vytvořené báze můžeme vylepšit, když částečně ortogonalizované mezivýsledky ortogonalizujeme proti předchozím vektorům. Tento postup, který nazýváme **reortogonalizací**, nyní schématicky uvedeme pro k -tý krok procesu, kdy máme k dispozici prvních $k - 1$ normalizovaných vektorů.

- Uvažujme klasickou variantu ortogonalizace CGS

$$z_k = a_k - \left((q_{k-1}^* a_k) q_{k-1} + \dots + (q_1^* a_k) q_1 \right) \quad (4.41)$$

- Aplikujme další ortogonalizaci ještě na výsledek z_k

$$w = z_k - \left((q_{k-1}^* z_k) q_{k-1} + \dots + (q_1^* z_k) q_1 \right) \quad (4.42)$$

- Dosazením

$$w = a_k - \left((q_{k-1}^* a_k + q_{k-1}^* z_k) q_{k-1} + \dots + (q_1^* a_k + q_1^* z_k) q_1 \right) \quad (4.43)$$

Následující algoritmus implementuje CGS s touto dodatečnou ortogonalizací (reortogonalizací).

Algoritmus 4.4.3 ICGS

Input: $A = [a_1, \dots, a_m]$

Output: QR ICGS rozklad.

0. **Inicializace:** $r_{11} = \|a_1\|$, $q_1 = a_1/r_{11}$, $Q_1 = [q_1]$
1. **for** $k = 2 : m$ **do**
2. $z_k = a_k$
3. $r = 0$
4. **for** $l = 1 : k-1$ **do**
5. $\tilde{r} = Q_{k-1}^* z_k$
6. $z_k = z_k - Q_{k-1} \tilde{r}$
7. $r = r + \tilde{r}$
8. **end do**

9. $[r_{1k}, \dots, r_{k-1,k}]^T = r$
10. $r_{kk} = \|z_k\|$
11. $q_k = z_k / r_{kk}$
12. $Q_k = [Q_{k-1}, q_k]$
13. **end do**

4.5 Řešení soustavy pomocí QR rozkladu

Uvažujme čtvercovou matici $A \in \mathbb{R}^{n \times n}$ soustavy lineárních rovnic. Její řešení pomocí QR rozkladu vyplývá z následující soustavy ekvivalencí

$$Ax = b \Leftrightarrow QRx = b \Leftrightarrow QRx = QQ^*b \Leftrightarrow Rx = Q^*b \quad (4.44)$$

4.6 QR rozklad a jeho stabilita

Připomeňme nejprve, že ve standardním modelu výpočtu v aritmetice s konečnou přesností (bez uvažování podtečení či přetečení) se uvažuje následující model operací

$$fl(x \text{ op } y) = (1 + \delta)(x \text{ op } y), \quad |\delta| \leq \epsilon/2, \quad \text{op} = +, -, *, / \quad (4.45)$$

Jak jsme již uvedli výše, říkáme, že algoritmus je **zpětně stabilní**, má-li malou zpětnou chybu. Samozřejmě, tato definice je závislá na kontextu a v konkrétních případech si řekneme, co zpětná stabilita algoritmu bude znamenat.

4.6.1 Aplikace rotací a reflexí

Je-li $U \in \mathbb{C}^{n \times n}$ (reálná nebo obecně komplexní analogie) matice Givensovy rotace nebo Householderovy reflexe, pak platí

$$fl(UA) = U(A + E), \quad \|E\|/\|A\| \leq \gamma n^2 \epsilon + O(\epsilon^2). \quad (4.46)$$

Předcházející výraz zkracujeme na přibližnou rovnost

$$\|E\| \approx \epsilon \|A\| \quad (4.47)$$

V tomto případě říkáme, že aplikace rotací a reflexí je **zpětně stabilní**, neboť spočtené řešení této úlohy je přesným řešením modifikované úlohy se vstupními daty, která jsou blízka původním datům dané úlohy. To jsme výše nazvali malou zpětnou chybou. To, že je chyba skutečně malá plyne z toho, že v ní jsou jen členy, které nám přirozeně musí vzniknout z důvodu

- konečné přesnosti počítače (ϵ),
- jednoduchých operací v pohyblivé řádové čárce (norma matice A , skalární faktor, polynom malého stupně dimenze matice).

Jiné členy, jako je například číslo podmíněnosti, tu nejsou.

4.6.2 Ortogonalita spočtené matice

Předpokládejme, že jsme spočítali v QR rozkladu matici \hat{Q} místo přesné matice Q . Zajímá nás chyba v ortogonalizaci E_Q matice A , kterou můžeme zavést vztahem

$$\hat{Q}^* \hat{Q} = I + E_Q. \quad (4.48)$$

Tuto chybu budeme nazývat **ztráta ortogonality** matice Q a budeme ji měřit normou $\|E_Q\|$. Výsledkem jsou následující závislosti pro různé typy QR rozkladu, kde $C(n, m)$ s příslušným indexem je nějaká funkce dimenze (dimenzí) problému (obecně různá v různých situacích shrnutých níže), která neroste příliš prudce.

$$\begin{aligned} \text{Householderovy reflexe} &\rightarrow \|E_Q\| \leq C_H(n, m)\epsilon \\ \text{Givensovy rotace} &\rightarrow \|E_Q\| \leq C_G(n, m)\epsilon \\ \text{CGS} &\rightarrow \|E_Q\| \leq C_C(n, m)\kappa^2(A)\epsilon \\ \text{MGS} &\rightarrow \|E_Q\| \leq C_M(n, m)\kappa(A)\epsilon \\ \text{ICGS} &\rightarrow \|E_Q\| \leq C_I(n, m)\epsilon \end{aligned}$$

Poznamenejme, že u MGS předpokládáme ortogonalizaci vůči předcházejícím vektorům báze v původním pořadí. S jiným pořadím v MGS může být výsledek méně přesný, až jako CGS. CGS navíc předpokládá mírně pozměněný výpočet diagonálních prvků.

4.6.3 Stabilita výpočtu faktoru R

Uvažujme QR rozklad, který vypočítá \hat{Q} a \hat{R} . Při výpočtu faktoru R s použitím Householderovy reflexí, Givensových rotací či **MGS** existuje taková přesně unitární matice Q taková, že

$$Q^*(A + E) = \hat{R} \quad (4.49)$$

kde pro zpětnou chybu E zároveň platí $\|E\|_F \leq C(n, m)\epsilon\|A\|_F$. Jinými slovy, v těchto případech je výpočet R zpětně stabilní.

V případě Householderových reflexí navíc platí, že \hat{Q} je blízko Q v následujícím smyslu:

$$\|\hat{Q} - Q\|_F \leq C(m, n)\epsilon. \quad (4.50)$$

4.6.4 Norma rezídua

Zde se ještě pozastavme u normy rezídua dané následujícím rozdílem.

$$A - \hat{Q}\hat{R} \quad (4.51)$$

Pak platí $\|A - \hat{Q}\hat{R}\| \leq C(n, m)\epsilon\|A\|$ pro Householderovy reflexe, Givensovy rotace, MGS, ICGS i **CGS**.

4.7 Počty operací pro jednotlivé druhy QR rozkladu

Zde uvedeme jen jednoduché návodné počty pro čtvercové matice.

- Householder: $4/3n^3$
- Givens: $2n^3$
- CGS, MGS: $2n^3$

4.8 Arnoldiho algoritmus

4.8.1 Definice Krylovovy posloupnosti

Definice 4.8.1 Předpokládejme, že $A \in \mathbb{C}^{n \times n}$, $b \in \mathbb{C}^n$. Posloupnost

$$b, Ab, A^2b, \dots \quad (4.52)$$

nazýváme **Krylovovou posloupností** příslušnou matici A a vektoru b .

V další definici zavedeme Krylovův prostor, který se hojně využívá jak v iteračních algoritmech řešení soustav lineárních rovnic, tak i při řešení problému vlastních čísel.

Definice 4.8.2

$$K_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\} \quad (4.53)$$

nazýváme m -tým **Krylovovým prostorem** pro $m \leq n$.

Důležitý postup, který nalezne ON bázi Krylovova prostoru, je uveden níže.

4.8.2 Hledání ortonormální báze prostoru $K_m(A, b)$

Hledání ON báze Krylovova prostoru je z formálního hlediska založeno na modifikaci Gram-Schmidtovy ortogonalizace, ve které používáme jako výchozí vektory vektory generující tento prostor.

Nalezení ortonormální báze prostoru $K_m(A, b)$

Schématicky je postup založen na opakování následujících třech kroků

- Zvolíme $q_1 = b/\|b\|$
- Pro $k > 1$ postupujeme Gram-Schmidtovou ortogonalizací (zde demonstrujeme s použitím CGS)

$$z = a_k - \sum_{i=1}^{k-1} (q_i^* a_k) q_i \longrightarrow z = Aq_{k-1} - \sum_{i=1}^{k-1} (q_i^* Aq_{k-1}) q_i \quad (4.54)$$

Na levé straně je **krok standardního GS procesu**, na pravé straně je **krok Arnoldiho algoritmu**, jak tento proces nazýváme.

- normalizace vektoru z

$$r_{kk} = \|z\|, \quad q_k = z/r_{kk}. \quad (4.55)$$

Hledání ON báze prostoru $K_m(A, b)$ jako QR rozklad

Proces získání ON báze Krylovova prostoru můžeme také chápat jako QR rozklad. Konkrétně, přepišme si vztah pro výpočet vektoru z následovně

$$Aq_{k-1} = z + \sum_{i=1}^{k-1} (q_i^* Aq_{k-1}) q_i \quad (4.56)$$

tedy

$$Aq_{k-1} = r_{kk} q_k + \sum_{i=1}^{k-1} (q_i^* Aq_{k-1}) q_i. \quad (4.57)$$

Vidíme, že každý sloupec Aq_j , $j = 1, k-1$ je lineární kombinací předcházejících ON vektorů a nového sloupce, což spolu s prvním sloupcem dává pro $k \geq 2$ následující rozklad

$$[b \ AQ_{k-1}] = Q_k R_k. \quad (4.58)$$

Neuvažujeme-li první sloupec v tomto rozkladu, můžeme Arnoldiho algoritmus také chápat jako převedení matice A na horní Hessenbergův tvar.

$$AQ_{k-1} = Q_k H_{k,k-1} \quad (4.59)$$

Matice $H_{k,k-1} \in \mathbb{C}^{k \times (k-1)}$ je vlastně matice R_k , které jsme odebrali první sloupec. Horní Hessenbergův tvar matice je strukturálně znázorněn následujícím schématem

$$\begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ & * & * & * & * & * \\ & & * & * & * & * \\ & & & * & * & * \\ & & & & * & * \end{pmatrix} \quad (4.60)$$

Předcházející vztah můžeme napsat i jinak, kde počítaný k -tý sloupec je explicitně vyjádřen následujícím způsobem

$$AQ_{k-1} = Q_{k-1} H_{k-1,k-1} + r_{k,k} q_k e_{k-1}^T. \quad (4.61)$$

4.9 Shrnutí

- GIVENSOVY ROTACE A HOUSEHOLDEROVY REFLEXE Z NUMERICKÉHO POHLEDU.
- APLIKACE GIVENSOVÝCH ROTACÍ A HOUSEHOLDEROVÝCH REFLEXÍ NA ŘÍDKÉ MATICE.
- GRAM-SCHMIDTŮV ALGORITMUS A NUMERICKÉ VLASTNOSTI JEHO VARIANT.

- STABILITA QR ROZKLADU:
 - APLIKACE ROTACÍ A REFLEXÍ,
 - ORTOGONALITA FAKTORU Q ,
 - CHYBA REZÍDUA,
 - CHYBA VÝPOČTU FAKTORU R .
- BÁZE KRYLOVSKÉHO PROSTORU A JEJÍ NALEZENÍ. INTERPRETACE JEJÍHO HLEDÁNÍ JAKO PROVEDENÍ QR ROZKLADU.

Kapitola 5

LU rozklad a Gaussova eliminace.

Uvažujme řešení soustavy lineárních algebraických rovnic

$$Ax = b, \quad (5.1)$$

kde $A \in \mathbb{C}^{n \times n}$, $b \in \mathbb{C}^n$ a hledáme řešení $x \in \mathbb{C}^n$. Předpokládáme, že matice A je regulární.

5.1 Standardní LU rozklad

5.1.1 Klíčové body v historii

- Nine chapters on arithmetics (Čína, 200 B.C.): eliminace soustavy o třech neznámých
- Seki Kowa, G.W. Leibniz (17. století), pojetí determinantu
- Gabriel Cramer (začátek 18. století)
- Gauss (1777 - 1855)
- Hotelling, Goldstine a von Neumann (1947) - výpočty chyby rezidua pro inverzi matice
- Turing (1948) - podmíněnost soustavy

5.1.2 Sloupcové eliminační matice a jejich inverze

Sloupcovou eliminační matici M_k pro nějaký index $k \leq n$ a pro daná čísla $m_{k+1,k}, \dots, m_{n,k}$ si definujeme následovně

$$M_k = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & m_{k+1,k} & \ddots & & \\ & & \dots & & 1 & \\ & & m_{n,k} & & & 1 \end{pmatrix} = I + m_k e_k^T, \quad m_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ m_{k+1,k} \\ \vdots \\ m_{n,k} \end{pmatrix} \quad (5.2)$$

Jedna z nejkrásnějších vlastností těchto jednoduchých matic je, že i jejich inverze se snadno vyjádří pomocí stejných prvků. Konkrétně, inverze regulární matice M_k je dána výrazem $M_k^{-1} = I - m_k e_k^T$, což můžeme rozepsat následovně

$$M_k^{-1} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & -m_{k+1,k} & \ddots & & \\ & & \dots & & 1 & \\ & & -m_{n,k} & & & 1 \end{pmatrix} \quad (5.3)$$

Použití sloupcových eliminačních matic pro řešení soustavy rovnic uvidíme níže.

5.1.3 Použití eliminačních matic: Gaussova eliminace a LU rozklad

Kroky procesu **Gaussovy eliminace** se dají vyjádřit aplikacemi (inverzí) sloupcových eliminačních matic na rozkládanou matici A zleva. To si znázorníme následovně.

$$A^{(0)} \equiv A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \rightarrow A^{(1)} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ 0 & a_{32}^{(1)} & \dots & a_{3n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{pmatrix} = M_1^{-1} A \quad (5.4)$$

je-li

$$M_1 = \begin{pmatrix} 1 & & & & \\ a_{21}/a_{11} & 1 & & & \\ a_{31}/a_{11} & & 1 & & \\ \vdots & & & \ddots & \\ a_{n1}/a_{11} & & & & 1 \end{pmatrix} \quad (5.5)$$

Všimněme si přitom značení, které jsme takto zavedli a budeme nadále používat. Aplikací inverze sloupcové eliminační matice na matice A zleva nám vznikla následující ekvivalentní soustava rovnic.

$$Ax = b \rightarrow A^{(1)}x = M_1^{-1}b \quad (5.6)$$

Vidíme, že matice transformované soustavy je v určitém smyslu jednodušší. **Nejznámější** vyjádření Gaussovy eliminace je založeno na převádění matice soustavy do odstupňovaného tvaru opakováním tohoto postupu aplikovaným postupně na podmatice výsledku. Zvolíme-li

$$M_2 = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & a_{32}^{(1)}/a_{22}^{(1)} & 1 & \\ & \vdots & & 1 \\ & a_{n2}^{(1)}/a_{22}^{(1)} & & 1 \end{pmatrix}, \quad (5.7)$$

pak aplikací M_2^{-1} na transformovanou soustavu zleva získáme

$$M_2^{-1}A^{(1)}x = M_2^{-1}M_1^{-1}b \quad (5.8)$$

Opakováním postupu se dostaneme k soustavě

$$M_{n-1}^{-1} \dots M_1^{-1}Ax = M_{n-1}^{-1} \dots M_1^{-1}b \quad (5.9)$$

Součin inverzních sloupcových eliminačních matic, které jsme na A aplikovali zleva, společně s touto maticí si označíme U a vidíme, že tento součin je v horním trojúhelníkovém tvaru

$$U = M_{n-1}^{-1} \dots M_1^{-1}A \quad (5.10)$$

Samotný součin těchto matic si označíme výrazem L^{-1} , tedy

$$L^{-1} = M_{n-1}^{-1} \dots M_1^{-1} \quad (5.11)$$

Dostaneme tak rovnost

$$Ux = L^{-1}b \Leftrightarrow LUx = b. \quad (5.12)$$

Uveďme si dvě podstatné vlastnosti matice L

$$L = M_1 \dots M_{n-1} \quad (5.13)$$

Za prvé, matice je samozřejmě **dolní trojúhelníková**, neboť vznikla znásobením dolních trojúhelníkových matic. Za druhé, L je explicitně definována pomocí prvků matic $A \equiv A^{(0)}, \dots, A^{(n-1)}$. Konkrétně platí

$$L = \begin{pmatrix} 1 & & & \\ a_{21}^{(0)}/a_{11}^{(0)} & 1 & & \\ a_{31}^{(0)}/a_{11}^{(0)} & & \ddots & \\ \vdots & \vdots & \ddots & \\ a_{n1}^{(0)}/a_{11}^{(0)} & & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & a_{32}^{(1)}/a_{22}^{(1)} & & \\ \vdots & \vdots & \ddots & \\ & a_{n2}^{(1)}/a_{22}^{(1)} & \dots & 1 \end{pmatrix} \dots = \begin{pmatrix} 1 & & & \\ a_{21}^{(0)}/a_{11}^{(0)} & 1 & & \\ a_{31}^{(0)}/a_{11}^{(0)} & a_{32}^{(1)}/a_{22}^{(1)} & & \\ \vdots & \vdots & \ddots & \\ a_{n1}^{(0)}/a_{11}^{(0)} & a_{n2}^{(1)}/a_{22}^{(1)} & \dots & 1 \end{pmatrix} \quad (5.14)$$

Tohle je jeden z klíčových momentů této kapitoly a formulujeme si jej jako důsledek našich úvah.

Důsledek 5.1.1 *Postupem z této kapitoly, kde jsme uvažovali o nulování poddiagonální částí zároveň získali i rozklad $A = LU$ s **explicitním** vyjádřením prvků v **obou** faktorech L i U , z nichž první je dolní trojúhelníkový s jednotkami na diagonále a druhý je v horním trojúhelníkovém tvaru.*

Konkrétně U vypadá následovně

$$U = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \dots & a_{3n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & \dots & a_{nn}^{(n-1)} \end{pmatrix} \quad (5.15)$$

Můžeme tedy mluvit záměnně o LU rozkladu nebo Gaussově eliminaci.

Definice 5.1.1 *Nechť $A \in \mathbb{C}^{n \times n}$ je regulární. Odvozený rozklad tvaru LU , kde L je dolní trojúhelníkovou maticí s jednotkovou diagonálou a U je horní trojúhelníkovou maticí nazýváme LU rozkladem matice A .*

Platí

$$Ax = b \Leftrightarrow L^{-1}Ax = L^{-1}b \Leftrightarrow Ux = L^{-1}b \equiv b^{(n-1)} \quad (5.16)$$

5.1.4 Přímý a zpětný chod v řešení soustav rovnic pomocí LU rozkladu

V této podkapitole zdůrazníme význam odděleného získání faktorů L a U a dalšího postupu řešení soustavy lineárních algebraických rovnic, které nazveme přímým a zpětným chodem.

Algoritmus 5.1.1 LU rozklad

Input: $A = [a_{ij}]$

Output: LU rozklad.

0. **Inicializace:** $A^{(0)} = A$
1. **for** $k = 1 : n - 1$ **do**
2. **for** $i = k + 1 : n$ **do**
3. $m_{ik} = a_{ik}^{(k-1)} / a_{kk}^{(k-1)}$
4. **for** $j = k + 1 : n$ **do**
5. $a_{ij}^{(k)} = a_{ij}^{(k-1)} - m_{ik}a_{kj}^{(k-1)}$
6. **end j**
8. **end i**
9. **end k**

Někdy se tento rozklad doplňuje ještě o aplikaci rozkladu přímo na pravou stranu b jako je to uvedeno ve skriptech. V tomto textu LU rozklad striktně oddělujeme od jeho aplikace na pravou stranu. Proces aplikace tady uvádíme jako dva kroky, nazývané **přímá** a **zpětná** substituce (přímý a zpětný chod).

Algoritmus 5.1.2 Přímý chod (substituce) při řešení soustavy LU rozkladem

Input: $A = [a_{ij}]$, LU rozklad této matice, vektor pravé strany soustavy b .

Output: Vektor $z = L^{-1}b$.

0. **Inicializace:** $b^{(0)} = b = [\beta_1, \dots, \beta_n]^T$

1. **for** $k = 1 : n - 1$ **do**

2. **for** $i = k + 1 : n$ **do**

3. $m_{ik} = a_{ik}^{(k-1)} / a_{kk}^{(k-1)} \equiv l_{ik}$

4. $\beta_i^{(k)} = \beta_i^{(k-1)} - m_{ik} \beta_k^{(k-1)}$

5. **end** i

6. **end** k

Algoritmus 5.1.3 Zpětný chod při řešení soustavy LU rozkladem

Input: $A = [a_{ij}]$, LU rozklad této matice, vektor $z = L^{-1}b$

Output: Řešení soustavy rovnic $x = [\xi_1, \dots, \xi_n]^T$.

0. **Inicializace:** $A^{(n-1)}$, $z = [z_1, \dots, z_n]^T$

1. $\xi_n = z_n / u_{nn}$

2. **for** $i = n - 1 : -1 : 1$ **do**

3. $\xi_i = z_i$

4. **for** $j = i + 1 : n$ **do**

5. $\xi_i = \xi_i - u_{ij} \xi_j$

6. **end** j

7. $\xi_i = \xi_i / u_{ii}$

8. **end** i

Význam oddělení přímého a zpětného chodu od samotného rozkladu je založen na tom, že pravou stranu nebo více pravých stran nemusíme mít v době rozkladu k dispozici. Dále, moderní algoritmy používají pro rozklady a následující substituční chody obvykle úplně jiné datové struktury s jinými nároky na výpočetní prostředky.

5.2 LU rozklad řídkých matic

Na rozdíl od QR rozkladu se LU rozklad chová vstřícněji k **řídkým** maticím, tedy k maticím, kde nulovost části jejich prvků hraje podstatnou roli. Konkrétně, vynásobení matice sloupcovou eliminační maticí zleva znamená, že **struktura řádku pivota** (množina pozic nenulových prvků řádku pivota) bude sjednocována se strukturou ostatních řádků, ale **sama nebude ovlivněna**. Podrobněji, je-li index řádku pivota i , pak pro každé $j > i$ je struktura j -tého řádku výsledku dána sjednocením původní struktury j -tého řádku a struktury řádku pivota na všech sloupcových pozicích sloupcem $i+1$ počínaje. Samozřejmým předpokladem je, že neuvažujeme **náhodné vyrušení** nenulových prvků během operace rozkladu. Struktura řádku i pivota se nezmění.

Uvažujme následující příklad matice, jejíž strukturu řídkosti znázorníme hvězdičkami.

$$A = \begin{pmatrix} * & * & * & * & * \\ * & * & & & \\ * & & * & & \\ * & & & * & \\ * & & & & * \end{pmatrix}. \quad (5.17)$$

Uvažme její LU rozklad a použijme na první krok rozkladu následující sloupcovou eliminační matici M_1 , kde její obecně nenulové prvky znázorníme zase hvězdičkami

$$M_1 = \begin{pmatrix} 1 & & & & \\ * & 1 & & & \\ * & & 1 & & \\ * & & & 1 & \\ * & & & & 1 \end{pmatrix}. \quad (5.18)$$

Snadno nahlédneme, že struktura matice M_1A bude obecně hustá, jak můžeme následně znázornit, neuvažujeme-li náhodné numerické vyrušení prvků

$$M_1A = \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}. \quad (5.19)$$

Na rozdíl od QR rozkladu je ale praktická procedura, která se vyhne podobné situaci, výrazně jednodušší. Stačí totiž uvažovat řešení soustavy PA s permutovanými řádky, nebo i soustavy se symetrickou permutací řádků i sloupců. Uvažujme nejprve následující permutační matici P , která první řádek matice A přesune na poslední.

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (5.20)$$

Permutovanou matici pak můžeme znázornit následovně

$$A = \begin{pmatrix} * & * & & & \\ * & & * & & \\ * & & & * & \\ * & & & & * \\ * & * & * & * & * \end{pmatrix}. \quad (5.21)$$

V tomto případě se vyhneme zaplnění ve faktoru U , ale faktor L bude hustý, protože v každém kroku se nám zaplní další sloupec v poddiagonální části rozkládané matice

podle následujícího schématu pro struktury částečně eliminovaných matic.

$$\begin{pmatrix} * & * & & & \\ * & & * & & \\ * & & & * & \\ * & & & & * \\ * & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & & & \\ & * & * & & \\ & * & & * & \\ & * & & & * \\ * & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & & & \\ & * & * & & \\ & & * & * & \\ & & * & & * \\ * & * & * & * & * \end{pmatrix} \rightarrow \dots \rightarrow \begin{pmatrix} * & * & & & \\ & * & * & & \\ & & * & * & \\ & & & * & * \\ * & & & & * \end{pmatrix} \quad (5.22)$$

Formálně lze schéma řešení soustavy rovnic s permutovanými řádky vyjádřit následovně

$$\begin{aligned} PA &= LU \\ Ax &= b \Leftrightarrow PAx = Pb \Leftrightarrow x = U^{-1}L^{-1}Pb. \end{aligned}$$

Ještě lepší je ale pomocí výše uvedené permutace P matici permutovat symetricky, tedy permutovat jak její řádky, tak i její sloupce. Struktura symetricky permutované matice je na následujícím obrázku a při jejím rozkladu žádné zaplnění nevznikne.

$$A = \begin{pmatrix} * & & & & * \\ & * & & & * \\ & & * & & * \\ & & & * & * \\ * & * & * & * & * \end{pmatrix}. \quad (5.23)$$

Řešení soustavy rovnic s permutovanými řádky i sloupci se dá opět jednoduše vyjádřit následujícím způsobem. Symetricky permutovaná matice s permutační maticí P se dá zapsat následovně

$$PAP^T = LU. \quad (5.24)$$

Řešení soustavy rovnic $Ax = b$ pak postupuje v následujících krocích.

$$\begin{aligned} PAP^T &= LU \text{ (rozklad)} \\ Lz &= Pb \text{ (substituce)} \\ Uy &= z \text{ (substituce)} \\ x &= P^T y \text{ (permutace vektoru)} \end{aligned}$$

Nalezení permutace matice A takové, aby se v rozkladu nebo jeho části snížila velikost zaplnění, čili počet nově vzniklých nenulových prvků, je možné algoritmicky formalizovat. Důležitou třídou algoritmů, které zaplnění snižují představují postupy, které postupně vybírají řádky pomocí **lokálních kritérií**. Pro každý krok nejprve vybírají řádek pivota, který zaplnění minimalizuje, což je často řádek s nejmenším počtem nenulových prvků v rozkládané matici. Existuje ale ještě druhá důležitá motivace výběru řádku pivota. Ta není motivována úvahami o struktuře nenulových prvků řídkých matic, ale problémem **proveditelnosti a stability** rozkladu. Jí bude věnována velká část následujícího textu. V praxi je nutné ohledy na proveditelnost a stabilitu rozkladu na jedné straně a na velikost zaplnění na straně druhé kombinovat.

5.3 Rozklad s částečnou pivotací

Výsledky, které se týkají stability LU rozkladu, jsou obecně slabší než analogické výsledky pro QR rozklad a obdobné tvrzení platí i pro **praktické provedení** LU rozkladu. V následujícím textu budeme diskutovat variantu přístupu k LU rozkladu, kde do tohoto procesu zapojíme permutaci řádků rozkládané matice.

5.3.1 Silná regularita

Důležitým pojmem, který se týká LU rozkladu matice, je **silná regularita**, který definujeme následujícím způsobem.

Definice 5.3.1 Matice $A = (a_{ij})$, $A \in \mathbb{C}^{n \times n}$ je silně regulární právě tehdy, platí-li

$$\det \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \dots & \vdots \\ a_{k1} & \dots & a_{kk} \end{pmatrix} \neq 0, \quad k = 1, \dots, n. \quad (5.25)$$

Podmínku z této definice vyjadřujeme slovně, že všechny hlavní minory matice A jsou nenulové.

Věta 5.3.1 Pro pivot $a_{kk}^{(k-1)}$ příslušné eliminační matice platí $a_{kk}^{(k-1)} \neq 0$ právě tehdy, když A je silně regulární.

Důkaz: Zde uvedeme jen poznámku k odvození tvrzení věty. $A^{(k)}$ je singulární právě tehdy, je-li $a_{k+1,k+1}^{(k)} = 0$, což znamená, že jsme narazili na lineárně závislý sloupec, protože v algoritmu se provádí jen takové operace, které nemění hodnotu matice. ■

V praxi nemusí být předpoklad silné regularity pro regulární matici A splněn. Příkladem je matice

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad (5.26)$$

která je regulární, ale nikoli silně regulární. Nicméně, v praktickém provádění LU rozkladu nejsou problémem jen matice, které nejsou silně regulární, ale i matice, které jsou mají některý ze svých hlavních minorů **blízko** nule. To znamená, že příslušná podmatice je blízko množiny singulárních matic. To si demonstrujeme na následujícím příkladě LU rozkladu

$$\begin{pmatrix} \epsilon & 1 \\ 1 & \epsilon \end{pmatrix} = \begin{pmatrix} 1 & \\ 1/\epsilon & 1 \end{pmatrix} \begin{pmatrix} \epsilon & 1 \\ \epsilon - 1/\epsilon & \end{pmatrix} \quad (5.27)$$

V této matici máme pro malé ϵ pivot s **malou absolutní hodnotou**. V důsledku dělení prvků matice ve sloupci pivota tímto pivotem získáváme v průběhu rozkladu mezivýsledky s velkou absolutní hodnotou (úměrné $1/\epsilon$), kde je při práci v aritmetice v konečné přesnosti značná pravděpodobnost ztráty přesnosti výpočtu. Následující věta neřeší úplně problém blízkosti hlavních minorů nule, tak jako jsme v minulém příkladě mohli za matici **numericky singulární** považovat matici tvořenou jejím prvním řádkovým a sloupcovým prvkem pro ϵ s malou absolutní hodnotou. Nicméně praktické hledání matice P uvedené níže bude často řešit i tento problém.

$$P_t M_k^{-1} P_t = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \dots & 1 & & \\ & & -m_{t,k} & & 1 & \\ & & \dots & & & 1 \\ & & -m_{l,k} & & & \\ & & \dots & & & 1 \end{pmatrix} \quad (5.30)$$

Výše uvedená transformace dá použít na **přesunování** permutačních matic ve vyjádření pro maticový zápis GEPP výše. Konkrétně platí pro $M_k \equiv M_1$ a $P_t \equiv P_2$.

$$P_2 M_1^{-1} P_1 A = (P_2 M_1^{-1} P_2) P_2 P_1 A = \tilde{M}_1^{-1} P_2 P_1 A \quad (5.31)$$

Aplikujeme tuto myšlenku opakovaně a dostaneme rozklad s permutací P a trochu jinou maticí L . Uvažujme výše uvedený horní trojúhelníkový faktor

$$U = M_{n-1}^{-1} P_{n-1} \dots P_2 M_1^{-1} P_1 A \quad (5.32)$$

Pak postupně dostaneme

$$\begin{aligned} U &= M_{n-1}^{-1} P_{n-1} M_{n-2}^{-1} P_{n-2} M_{n-3}^{-1} P_{n-3} \dots M_2^{-1} P_2 M_1^{-1} P_1 A \\ U &= M_{n-1}^{-1} (P_{n-1} M_{n-2}^{-1} P_{n-1}) P_{n-1} P_{n-2} M_{n-3}^{-1} P_{n-3} \dots P_2 M_1^{-1} P_1 A \\ U &= M_{n-1}^{-1} (P_{n-1} M_{n-2}^{-1} P_{n-1}) (P_{n-1} P_{n-2} M_{n-3}^{-1} P_{n-2} P_{n-1}) P_{n-1} P_{n-2} \dots P_2 M_1^{-1} P_1 A \\ U &= \tilde{M}_{n-1}^{-1} \tilde{M}_{n-2}^{-1} \tilde{M}_{n-3}^{-1} P_{n-1} P_{n-2} \dots P_2 M_1^{-1} P_1 A \\ U &= \dots \\ U &= (\tilde{M}_{n-1}^{-1}) (\tilde{M}_{n-2}^{-1}) \dots P_{n-1} P_{n-2} \dots P_2 P_1 A \\ U &= L^{-1} P A \end{aligned}$$

Věta 5.3.3 Gaussovu eliminaci s částečnou pivotací (GEPP) lze napsat jako $PA = LU$ tak, jak jsme to viděli výše při diskusi o struktuře rozkladu řídké matice, ale tvrzení formulujeme i zde.

$$Ax = b \Leftrightarrow PAx = Pb \Leftrightarrow Pb = LUx \quad (5.33)$$

5.3.2 Citlivost na změny vstupů v LU rozkladu

V této podkapitole si ukážeme transformaci změn na vstupu na změny v řešení. Uvažujme pro jednoduchost pouze změnu (perturbaci) na pravé straně soustavy rovnic.

Základní vztah

$$A(x + \Delta x) = b + \Delta b \quad (5.34)$$

Postupnými úpravami dostáváme

$$\begin{aligned} A\Delta x &= \Delta b \\ \|\Delta x\| &= \|A^{-1}\Delta b\| \leq \|A^{-1}\|\|\Delta b\| \\ \|b\| &\leq \|A\|\|x\| \Rightarrow \|x\| \geq \|b\|/\|A\| \end{aligned}$$

Dohromady tedy získáme následující vztah, který ukazuje transformaci perturbace.

$$\frac{\|\Delta x\|}{\|x\|} \leq \frac{\|A^{-1}\|\|A\|\|\Delta b\|}{\|b\|} = \kappa(A) \frac{\|\Delta b\|}{\|b\|} \quad (5.35)$$

Vidíme, že číslo podmíněnosti $\kappa(A)$ je mírou citlivosti soustavy vzhledem k perturbacím. Analogický vztah se dá odvodit když uvažujeme i/nebo změny v matici A .

Kvalita spočteného přibližného řešení

Nechť je spočtené řešení \hat{x} . Ptáme se na velikost přímé chyby $\|x - \hat{x}\|$ a zkusme ji odhadnout a posteriori. Můžeme psát

$$\|x - \hat{x}\| = \|A^{-1}(b - A\hat{x})\| \leq \|A^{-1}\|\|r\| \Rightarrow \quad (5.36)$$

$$\frac{\|x - \hat{x}\|}{\|x\|} \leq \frac{\|A^{-1}\|\|A\|\|r\|}{\|b\|} = \kappa(A) \frac{\|r\|}{\|b\|} \quad (5.37)$$

Důsledkem je, že je-li $\kappa(A)$ malé, pak z malého rezídua plyne i malá velikost přímé chyby. Obecně ale z malého rezídua malá velikost přímé chyby neplyne.

5.3.3 Zpětná chyba a stabilita LU rozkladu

Věta 5.3.4 *Nechť \hat{L} a \hat{U} jsou vypočtené faktory LU rozkladu obecně bez pivotace na počítači se strojovou přesností ϵ . Pak platí*

$$\begin{aligned} A + E &= \hat{L}\hat{U} \\ \|E\|_{\infty} &\leq 2n\epsilon\|\hat{L}\|_{\infty}\|\hat{U}\|_{\infty} + O(\epsilon^2) \end{aligned}$$

Abychom dokázali říci, zdali je zpětná chyba malá, čili zdali je LU rozklad zpětně stabilní, museli bychom vyjádřit odhad zpětné chyby jako funkci vstupních dat. Jinak máme zpětnou chybu, danou v tomto případě jako rezíduum $A - \hat{L}\hat{U}$, ale o zpětné stabilitě nejsme schopni mnoho říci. Teprve vyjádření E jako neurčitosti v datech A dělá ze zpětné chyby užitečnou veličinu. Problém se stabilitou může být skryt právě ve velikostech norem faktorů $\|\hat{L}\|_{\infty}$ a $\|\hat{U}\|_{\infty}$. Normy těchto faktorů mohou být mnohem větší než norma datového vstupu. Následující věta ukazuje, jak je to se zpětnou chybou

řešení soustavy rovnic. V tomto případě zpětná chyba se nedá napsat jen jako nějaké rezíduum. V obou uvedených případech je ale potřeba odhadnout zpětnou chybu jako neurčitost v původních datech, aby bylo možné posoudit zpětnou stabilitu diskutovaných úloh.

Věta 5.3.5 *Nechť \hat{L} a \hat{U} jsou vypočtené faktory LU rozkladu (bez pivotace) na počítači se strojovou přesností ϵ , $n\epsilon \ll 1$. Nechť \hat{x} je vypočtené řešení soustavy $Ax = b$ pomocí tohoto LU rozkladu. Pak je \hat{x} přesné řešení úlohy*

$$\begin{aligned}(A + \Delta A)\hat{x} &= b \\ \|\Delta A\|_\infty &\leq 6n\epsilon\|\hat{L}\|_\infty\|\hat{U}\|_\infty + O(\epsilon^2)\end{aligned}$$

5.3.4 Zpětná stabilita LU rozkladu s pivotací (GEPP)

Uvažujme nyní GEPP. Platí stále tvrzení z minulé sekce, ale vliv pivotace se projeví v odhadech norem faktorů. Abychom měli tvrzení, které o GEPP říká co nejvíce, musím se snažit neznámé veličiny odhadnout pomocí známých, například pomocí velikostí vstupních dat. Snadno nahlédneme, že platí následující věta.

Věta 5.3.6 *V GEPP máme $\|\hat{L}\|_\infty \leq n$.*

Důkaz: Protože diagonální prvek je největší v absolutní hodnotě příslušné eliminační matice a prvky faktoru L získáme vydělením tímto prvkem, pak jsou všech poddiagonální prvky L nejvýše rovny jedné a odtud plyne tvrzení. ■

Problémem je náhrada výrazu $\|\hat{U}\|_\infty$ výrazem $\|\hat{A}\|_\infty$, který potřebujeme, abychom mohli něco říci o zpětné stabilitě. Předpokládejme, že poměr

$$\frac{\|\hat{U}\|_\infty}{\|\hat{A}\|_\infty} = g \tag{5.38}$$

je nějak omezen. Tento poměr g nazveme **růstový faktor**. Můžeme jej definovat pro různé maticové normy příslušných matic. Nahradíme-li ve formuli výše normu faktoru součinem normy matice A a růstového faktoru a zároveň dosadíme omezení $\|\hat{L}\|_\infty \leq n$, získáme vztah

$$\|\Delta A\|_\infty \leq 6n^2g\epsilon\|\hat{A}\|_\infty + O(\epsilon^2) \tag{5.39}$$

Na základě tohoto budeme říkat, že GEPP je **podmíněně zpětně stabilní**. Tou podmíněností se myslí, že pro stabilitu je potřeba malý růstový faktor.

Velikost růstového faktoru

Jak velký vlastně může růstový faktor být? Zkusme si napsat odhady pro klíčovou operaci z LU rozkladu. Uvažujme vytvoření prvků $a_{ij}^{(k)}$ matice $A^{(k)}$ pro $k \geq 1$, $i, j > k$. V GEPP platí

$$m_{ik} \leq 1 \tag{5.40}$$

a proto můžeme psát

$$|a_{ij}^{(k)}| \leq |a_{ij}^{(k-1)}| + |a_{kj}^{(k-1)}| \leq 2 \max_{i,j} |a_{ij}^{(k-1)}| \leq 2 \times 2 \max_{i,j} |a_{ij}^{(k-2)}| \leq \dots$$

Celkem tedy dostáváme maximální růst

$$|u_{ij}| \leq 2^{n-1} \max_{i,j} |a_{ij}|. \quad (5.41)$$

Může tento pesimistický scénář nastat? Ano. Pro následující matici bude růstový faktor opravdu exponenciální. Naneštěstí může dojít k velkému růstu i v důležitých praktických úlohách.

$$\begin{pmatrix} 1 & & & & 1 \\ -1 & 1 & & & 1 \\ \vdots & \ddots & \ddots & & \vdots \\ -1 & \dots & -1 & 1 & 1 \\ -1 & \dots & -1 & -1 & 1 \end{pmatrix} \quad (5.42)$$

5.3.5 Škálování

Uvažujme následující transformaci soustavy lineárních rovnic s maticí $A \in \mathbb{C}^{n \times n}$ pomocí diagonálních matic D_1 a D_2 .

$$D_1 A D_2 y = D_1 b, \quad x = D_2 y \quad (5.43)$$

Spočítáme-li LU rozklad matice $D_1 A D_2$, pak se výpočet řešení x nové soustavy liší od původního postupu jen tak, že nejprve přenásobíme původní pravou stranu maticí D_1 (n operací) a na závěr vynásobíme vektor y maticí D_2 . Vidíme, že komplikace řešícího postupu vnesená škálováním je naprosto minimální. Tato transformace, kterou nazýváme **škálování** nebo také **vyvažování**, je schopna výrazně snížit velikost čísla podmíněnosti, ale důvodů pro škálování může být více. Prvním nich je **normalizace naměřených dat**. Dalším důvodem bývá **přizpůsobení absolutním velikostem konstant používaným v algoritmu**. Třetím důvodem může být výše zmíněná **minimalizace čísla podmíněnosti**, ale to je samozřejmě obtížný úkol. Platí ale například následující věta.

Věta 5.3.7 *Zvolíme-li diagonální matice dimenze n takové, že $D_2 = I$ a D_1 má jako diagonální prvky inverze euklidovských norem příslušných řádků, pak je číslo podmíněnosti $\kappa(D_1 A D_2)$ nejvýše \sqrt{n} -násobek nejmenšího čísla podmíněnosti dosažitelného libovolnou jinou volbou diagonální matice D_1 .*

5.3.6 Úplná pivotace

Úplná pivotace znamená, že se nevybírám jenom prvek který má co největší hodnotu ve sloupci matice, ale hledá se v celé matici. Kroky rozkladu pak můžeme zapsat následovně

$$PAQ = LU \quad (5.44)$$

pro permutační matice P a Q . Řešení soustavy rovnic $Ax = b$ pak probíhá v následujících krocích.

$$\begin{aligned} PAQ &= LU \text{ (rozklad)} \\ Lz &= Pb \text{ (substituce)} \\ Uy &= z \text{ (substituce)} \\ x &= Qy \text{ (permutace vektoru)} \end{aligned}$$

Problém úplné pivotace je v tom, že datové struktury, které je nutno použít pro implementaci, obvykle neumožňují efektivní výpočet. Hledání permutace, která může zmenšit zaplnění v LU rozkladu i při případné další částečné pivotaci pak může být výrazně efektivnější.

5.4 Choleského rozklad hermitovské PD (HPD) matice

Zahajme větou, kterou budeme předpokládat, ale která se dá snadno dokázat

Věta 5.4.1 *Každá hermitovská pozitivně definitní matice je silně regulární.*

Její formální důkaz pro hermitovskou a pozitivně definitní matici $A \in \mathbb{C}^{n \times n}$ se dá založit na tom, že podmatice

$$A_{2:n,2:n}^{(1)}, A_{3:n,3:n}^{(2)}, A_{4:n,4:n}^{(3)}, \dots$$

jsou také pozitivně definitní a to tedy znamená, že nemohou mít žádný diagonální prvek nulový. Podle Věty 5.3.1 pak víme, že matice je silně regulární. Tato věta mimo jiné znamená, že v rozkladu hermitovské PD matice **nejsou potřeba permutace** pro **existenci** rozkladu při počítání v přesné aritmetice. V této kapitole si mimo jiné ukážeme výsledek, který se týká stability LU rozkladu HPD matice v aritmetice v konečnou přesností. V následujícím ukážeme, že v případě rozkladu HPD matic platí vztah mezi jejími faktory L a U a výslednou faktorizaci budeme nazývat Choleského rozkladem.

5.4.1 Vztah mezi faktory L a U v LU rozkladu HPD matice

Věta 5.4.2 *LU rozkladu HPD matice $A \in \mathbb{C}^{n \times n} = \tilde{L}U$ je korektně definován a platí následující vztah*

$$\tilde{L} = [\text{diag}(u_{11}, \dots, u_{nn})]^{-1}U^* \quad (5.45)$$

Důkaz: Uvažujme jednoznačně určený LU rozklad matice A

$$\begin{aligned} A &= \tilde{L}U \\ A &= \tilde{L} \text{diag}(u_{11}, \dots, u_{nn}) \tilde{U} \end{aligned}$$

Matice \tilde{U} má tedy také jednotkovou diagonálu stejně jako \tilde{L} v naší konvenci. Využijme toho, že A je HPD a že tedy musí platit pro rozklad následující vztah

$$\tilde{L} \operatorname{diag}(u_{11}, \dots, u_{nn}) \tilde{U} = \tilde{U}^* \operatorname{diag}(\bar{u}_{11}, \dots, \bar{u}_{nn}) \tilde{L}^*. \quad (5.46)$$

Označme nyní pro jednoduchost mimodiagonální prvky matic \tilde{L} a \tilde{U} symboly l_{ij} respektive u_{ij} a porovnávejme matice v předcházejícím vztahu po sloupcích. Aplikujeme obě ekvivalentní strany v tomto vztahu na jednotkový vektor e_1 , což můžeme zapsat následovně

$$\tilde{L} \operatorname{diag}(u_{11}, \dots, u_{nn}) \tilde{U} e_1 \equiv \tilde{L} \begin{pmatrix} u_{11} & & \\ & \ddots & \\ & & u_{nn} \end{pmatrix} \tilde{U} e_1 = \tilde{L} u_{11} e_1 \quad (5.47)$$

$$\tilde{U}^* \operatorname{diag}(\bar{u}_{11}, \dots, \bar{u}_{nn}) \tilde{L}^* e_1 \equiv \tilde{U}^* \begin{pmatrix} \bar{u}_{11} & & \\ & \ddots & \\ & & \bar{u}_{nn} \end{pmatrix} \tilde{L}^* e_1 = \tilde{U}^* \bar{u}_{11} e_1. \quad (5.48)$$

Porovnejme nyní první sloupec a první diagonální prvek obou matic a dostaneme

$$\begin{aligned} u_{11} &= \bar{u}_{11} \\ \tilde{L} e_1 &= \tilde{U}^* e_1 \end{aligned}$$

Vidíme, že první sloupec \tilde{L} a první sloupec \tilde{U}^* se rovnají.

Uvažujme nyní nějaké $1 < i \leq n$ a induktivně předpokládejme, že platí

$$u_{jj} = \bar{u}_{jj}, \quad j = 1, \dots, i-1 \quad (5.49)$$

a také

$$\tilde{L} e_j = \tilde{U}^* e_j, \quad j = 1, \dots, i-1. \quad (5.50)$$

Po vynásobení obou rozkladů jednotkovým vektorem e_i získáme vztahy, které by měly vyjadřovat totéž.

$$\tilde{L} \operatorname{diag}(u_{11}, \dots, u_{nn}) \tilde{U} e_i = \tilde{L} \operatorname{diag}(u_{11}, \dots, u_{nn}) \begin{pmatrix} u_{1i} \\ \vdots \\ u_{i-1,i} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \tilde{L} \begin{pmatrix} u_{11} u_{1i} \\ \vdots \\ u_{i-1,i-1} u_{i-1,i} \\ u_{ii} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (5.51)$$

$$\tilde{U}^* \text{diag}(\bar{u}_{11}, \dots, \bar{u}_{nn}) \tilde{L}^* e_i = \tilde{U}^* \text{diag}(\bar{u}_{11}, \dots, \bar{u}_{nn}) \begin{pmatrix} \bar{l}_{i1} \\ \vdots \\ \bar{l}_{i,i-1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \tilde{U}^* \begin{pmatrix} \bar{u}_{11} \bar{l}_{i1} \\ \vdots \\ \bar{u}_{i-1,i-1} \bar{l}_{i,i-1} \\ \bar{u}_{ii} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (5.52)$$

Podle indukčního předpokladu máme rovnost prvních komponent sloupce. Konkrétně máme

$$u_{11} = \bar{u}_{11}, \dots, u_{i-1,i-1} = \bar{u}_{i-1,i-1} \quad (5.53)$$

a také

$$u_{1i} = \bar{l}_{i1}, \dots, u_{i-1,i} = \bar{l}_{i,i-1}. \quad (5.54)$$

A tak dostáváme rovnost i pro i -tý sloupec \tilde{L} a i -tý sloupce \tilde{U}^* a můžeme psát

$$A = \tilde{L} \text{diag}(u_{11}, \dots, u_{nn}) \tilde{L}^* \quad (5.55)$$

což je ekvivalentní vztahu

$$\text{diag}(\bar{u}_{11}, \dots, \bar{u}_{nn}) = \tilde{L}^{-1} A \tilde{L}^{-*} \quad (5.56)$$

■

Diagonální prvky můžeme odmocnit, protože platí $u_{ii} > 0$ pro všechny indexy i . Kdyby totiž pro příslušný diagonální prvek platilo

$$(e_i^T \tilde{L}^{-1}) A (\tilde{L}^{-*} e_i) \leq 0 \quad (5.57)$$

a to by byl spor s pozitivní definitností matice A . Podotkněme, že inverze trojúhelníkové matice s jednotkovou diagonálou je opět trojúhelníková matice, stejného typu, s jednotkovou diagonálou.

5.4.2 Existence Choleského rozkladu a jeho některé vlastnosti

Věta 5.4.3 Pro HPD matici $A \in \mathbb{C}^{n \times n}$ existuje jednoznačný rozklad $A = LL^*$, kde L je dolní trojúhelníková matice s kladnými prvky na diagonále.

Věta 5.4.4 Pro Choleského faktor napsaný po řádcích $L = (l_1^T, \dots, l_n^T)^T$ platí následující vztah

$$l_i l_i^* = \|l_i\|^2 = a_{ii}. \quad (5.58)$$

Důkaz: Snadno nahlédneme, protože L je dolní trojúhelníková matice. ■

Následující tvrzení vyplývá z předchozího

Věta 5.4.5

$$\|L\|_F^2 \equiv \text{trace}(A) = \sum_{i=1}^n a_{ii}. \quad (5.59)$$

Důkaz: $\|L\|_F^2 = \sum_{i=1}^n \sum_{j=1}^n |l_{ij}|^2 = \sum_{i=1}^n l_i l_i^* = \sum_{i=1}^n a_{ii} = \text{trace}(A)$ ■

5.4.3 Algoritmus Choleského rozkladu

Algoritmus Choleského rozkladu pro HPD matici je uveden v následujícím schématu. Ačkoli je jeho schéma odvozeno ze schémat pro LU faktorizaci s využitím výše uvedeného vztahu mezi faktory, zkusíme si nejprve ukázat, jak vypadá tento rozklad pro HPD matici $A \in \mathbb{C}^{2 \times 2}$. Konkrétně,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} l_{11} & \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} l_{11} & l_{12} \\ & l_{22} \end{pmatrix} \quad (5.60)$$

kde

$$\begin{aligned} l_{11} &= \sqrt{a_{11}^{(0)}} \equiv a_{11} \\ l_{12} &= \bar{l}_{21} \\ l_{22} &= \sqrt{a_{22} - l_{21} \bar{l}_{21}} \equiv \sqrt{a_{22}^{(1)}} \end{aligned}$$

Algoritmus 5.4.1 Přímý chod Choleského rozkladu

Input: $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, A je HPD

Output: Choleského rozklad.

```

0. Inicializace:  $A^{(0)} = A$ 
1. for  $k = 1 : n$  do
2.    $l_{kk} = \sqrt{a_{kk}^{(k-1)}}$ 
3.   for  $i = k + 1 : n$  do
4.      $l_{ik} = a_{ik}^{(k-1)} / l_{kk}$ 
5.     for  $j = k + 1 : i$  do
6.        $a_{ij}^{(k)} = a_{ij}^{(k-1)} - l_{ik} \bar{l}_{jk}$ 
7.     end  $j$ 
8.   end  $i$ 
9. end  $k$ 

```

5.4.4 Zpětná stabilita Choleského rozkladu

Přepišme si nejprve vztah pro zpětnou chybu uvedenou výše pro LU rozklad.

Věta 5.4.6 *Nechť \hat{L} je vypočtený faktor Choleského rozkladu HPD matice na počítači se strojovou přesností ϵ . Pak pro Frobeniovu normu chyby platí následující vztah.*

$$\begin{aligned} A + E &= \hat{L} \hat{L}^* \\ \|E\|_F &\leq 2n\epsilon \|\hat{L}\|_F^2 + O(\epsilon^2) \end{aligned}$$

K tomu, abychom dokázali zpětnou stabilitu Choleského rozkladu, stačí získat například vztah tvaru

$$\|E\|_F \sim \epsilon \|A\|_F + O(\epsilon^2) \quad (5.61)$$

Analogické tvrzení v případě LU rozkladu využilo růstového faktoru. V případě Choleského rozkladu je možné vztah tohoto typu skutečně odvodit.

Věta 5.4.7 *Nechť \hat{L} je vypočtený faktor HPD matice $A \in \mathbb{C}^{n \times n}$ se strojovou přesností ϵ . Nechť dále $2n^{3/2}\epsilon \ll 1$. Potom pro matici E takovou, že platí $A + E = \hat{L}\hat{L}^*$ máme*

$$\|E\|_F \leq \left(\frac{2n^{3/2}}{1 - 2n^{3/2}\epsilon} \right) \epsilon \|A\|_F + O(\epsilon^2) \quad (5.62)$$

Důkaz:

Porovnejme nejprve stopy matic následujícím způsobem pro vztah se zpětnou chybou v rozkladu

$$\|\hat{L}\|_F^2 = \text{trace}(A + E) = \text{trace}(A) + \text{trace}(E) \quad (5.63)$$

Protože pro libovolnou matici $B \in \mathbb{C}^{n \times n}$ platí vztah

$$|\text{trace}(B)| \leq \sqrt{n} \|B\|_F, \quad (5.64)$$

jehož platnost můžeme vidět z následujícího obecného vztahu, kde matici C vezmeme jako jednotkovou

$$\text{trace}(CB) = \langle \text{vec}(C^T), \text{vec}(B) \rangle \leq \|\text{vec}(C^T)\| \|\text{vec}(B)\| = \|C\|_F \|B\|_F \quad (5.65)$$

Pak dostaneme jeho dosazením do předchozího pro odhad stop matic A a E následující

$$\|\hat{L}\|_F^2 \leq \sqrt{n}(\|A\|_F + \|E\|_F) \quad (5.66)$$

Neboť ale víme, že platí

$$\|E\|_F \leq 2n\epsilon \|\bar{L}\|_F^2 + O(\epsilon^2) \quad (5.67)$$

pak dostáváme

$$\|E\|_F \leq 2n\sqrt{n}\epsilon(\|A\|_F + \|E\|_F) + O(\epsilon^2), \quad (5.68)$$

z čehož už je výsledek zřejmý. ■

5.4.5 Iterační zpřesnění

Iterační zpřesnění je jednoduchá iterační metoda, která může pomoci zlepšit řešení soustavy lineárních rovnic v níže uvedeném smyslu. Nejprve uveďme samotný postup, který konstruuje posloupnost

$$x_1, \dots \quad (5.69)$$

podle následujícího schématu

$$\begin{aligned} Ax &= b \longrightarrow x_1 \\ r_1 &= b - Ax_1 = A(x - x_1) = Ae^{(1)} \\ Ae^{(1)} &= r_1 \\ x_2 &= x_1 + e^{(1)} \\ &\dots \end{aligned}$$

Teorie v pozadí iteračního zpřesnění soustav říká, že se dá

- zmenšit **přímá chyba** při počítání r_i ve **zvýšené přesnosti**
- zmenšit **zpětná chyba**: jedna iterace iteračního zpřesnění zajistí zpětnou stabilitu řešení, což můžeme konkrétně zapsat

$$(A + \Delta A)x_2 = b, \quad \frac{\|\Delta A\|}{\|A\|} \approx \epsilon + O(\epsilon^2) \quad (5.70)$$

5.4.6 Výpočetní náklady GE

Výpočetní náklady budou probrány na cvičení.

5.5 Zmínka o srovnání LU a QR rozkladů pro řešení soustav

V předcházejícím textu jsme viděli, že pro QR faktorizaci platí výrazně silnější záruky na stabilitu než se dají odvodit pro LU faktorizaci. Ačkoliv jsme o tom v textu nehovořili, uvedené horní meze na stabilitu obvykle dobře vystihují nejhorší, ale zároveň prakticky dosažitelné, chování těchto rozkladů. Na druhé straně, LU rozklad obvykle vede k výrazně rychlejším výpočtům. V tomto zrychlení hrají roli dva faktory. První z nich je počet operací LU rozkladu obecně. Druhý z těchto efektů je výrazně větší šance udržet strukturu faktorů L a U řídkou na rozdíl od faktorů R QR rozkladu. Právě tento druhý důvod je v praxi **extrémně důležitý**.

5.6 Shrnutí

- GAUSSOVA ELIMINACE JAKO LU ROZKLAD A VLIV PIVOTACE.
- LU ROZKLAD ŘÍDKÉ MATICE - ZMĚNY VE STRUKTUŘE.
- ŠKÁLOVÁNÍ, ÚPLNÁ PIVOTACE, ITERAČNÍ ZPŘESNĚNÍ.
- CHOLESKÉHO ROZKLAD HERMITOVSKÉ MATICE.
- STABILITA LU ROZKLADU S ŘÁDKOVOU PIVOTACÍ.
- ZPĚTNÁ STABILITA CHOLESKÉHO ROZKLADU.
- POROVNÁNÍ LU A QR ROZKLADU Z HLEDISKA ROZKLADU ŘÍDKÉ MATICE A STABILITY.

Kapitola 6

Singulární rozklad matice

Dříve, než se zde začneme zabývat singulárním rozkladem matice, připomeneme si některé vlastnosti matic chápaných jako reprezentace lineárních zobrazení mezi prostory.

6.1 Matice jako lineární zobrazení a fundamentální podprostory

Uvažujme obecně obdélníkovou komplexní matici $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = r$. Tato matice definuje lineární zobrazení, které prvku x z prostoru \mathbb{C}^m přiřazuje prvek Ax z prostoru \mathbb{C}^n . Následující věta zavádí fundamentální prostory lineárního zobrazení A .

Věta 6.1.1 *Uvažujme lineární zobrazení $\mathcal{A} : \mathbb{C}^m \rightarrow \mathbb{C}^n$. Napíšeme-li rozklad prostoru $\mathbb{C}^m = \mathcal{N}(A) \oplus \mathcal{N}(A)^\perp$, pak platí*

$$\mathcal{R}(A^*) = \mathcal{N}(A)^\perp \tag{6.1}$$

a také

$$\mathcal{N}(A^*) = \mathcal{R}(A)^\perp. \tag{6.2}$$

Prostory $\mathcal{R}(A^*)$, $\mathcal{R}(A)$, $\mathcal{N}(A^*)$, $\mathcal{N}(A)$ nazýváme **fundamentální prostory lineárního zobrazení A** .

Důkaz: Důkaz vyplývá z následující sady ekvivalencí

$$z \in \mathcal{R}(A^*)^\perp \Leftrightarrow \forall x \langle A^*x, z \rangle = 0 \Leftrightarrow \forall x \langle x, Az \rangle = 0 \Leftrightarrow Az = 0 \Leftrightarrow z \in \mathcal{N}(A),$$

$$z \in \mathcal{R}(A)^\perp \Leftrightarrow \forall x \langle Ax, z \rangle = 0 \Leftrightarrow \forall x \langle x, A^*z \rangle = 0 \Leftrightarrow A^*z = 0 \Leftrightarrow z \in \mathcal{N}(A^*).$$

■

Shrňme tedy, že platí

$$\mathcal{N}(A) \oplus \mathcal{R}(A^*) = \mathbb{C}^m, \mathcal{N}(A) \perp \mathcal{R}(A^*) \tag{6.3}$$

a také

$$\mathcal{N}(A^*) \oplus \mathcal{R}(A) = \mathbb{C}^n, \mathcal{N}(A^*) \perp \mathcal{R}(A) \quad (6.4)$$

V následujícím textu uvedeme další speciálnější tvrzení, která se týkají fundamentálních prostorů pozitivně semidefinitních matic AA^* a A^*A a která použijeme při odvození singulárního rozkladu.

6.1.1 Spektrální rozklad hermitovské pozitivně semidefinitní matice

Matice AA^* a A^*A jsou čtvercové a pozitivně semidefinitní. Zapamatujme si, že v našem označení máme

$$AA^* \in \mathbb{C}^{n \times n}, \quad A^*A \in \mathbb{C}^{m \times m}. \quad (6.5)$$

Následující věta obsahuje tvrzení pro součiny matic A a A^* , které hrají podstatnou roli v odvozování singulárního rozkladu.

Věta 6.1.2 *Pro fundamentální prostory zavedených matic platí*

$$\begin{aligned} \mathcal{N}(A) &= \mathcal{N}(A^*A), \quad \mathcal{R}(A^*) = \mathcal{R}(A^*A), \\ \mathcal{N}(A^*) &= \mathcal{N}(AA^*), \quad \mathcal{R}(A) = \mathcal{R}(AA^*). \end{aligned}$$

Důkaz:

Ukážeme, že pro $x \in \mathbb{C}^m$ platí vztah

$$x \in \mathcal{N}(A^*A) \Rightarrow x \in \mathcal{N}(A)$$

pomocí následující sady implikace a ekvivalencí

$$\begin{aligned} x \in \mathcal{N}(A^*A) &\Leftrightarrow A^*Ax = 0 \\ &\Leftrightarrow \forall z \in \mathbb{C}^m : \langle z, A^*Ax \rangle = 0 \\ &\Leftrightarrow \forall z \in \mathbb{C}^m : \langle Az, Ax \rangle = 0 \\ &\Rightarrow \langle Ax, Ax \rangle = 0 \\ &\Leftrightarrow x \in \mathcal{N}(A) \end{aligned}$$

Opačná implikace je zřejmá a celkem tedy máme

$$\mathcal{N}(A) = \mathcal{N}(A^*A). \quad (6.6)$$

Druhé tvrzení odpovídá doplňkům těchto množin do příslušného prostoru. Konkrétně můžeme psát

$$\begin{aligned}
x \in \mathcal{R}(A^*A) &\Leftrightarrow x \perp \mathcal{N}((A^*A)^*) \equiv \mathcal{N}(A^*A) \\
&\Leftrightarrow x \perp \mathcal{N}(A) \\
&\Leftrightarrow x \in \mathcal{R}(A^*).
\end{aligned}$$

Druhá sada tvrzení (třetí a čtvrté) snadno vyplyne, když budeme uvažovat matici A^* místo matice A . ■

Připomeňme si, že platí následující vztah pro dimenze diskutovaných podprostorů

$$\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^*))$$

a stejně tak

$$\dim(\mathcal{R}(AA^*)) = \dim(\mathcal{R}(A^*A)). \quad (6.7)$$

6.1.2 Vztah mezi spektrálním rozkladem matice A^*A a AA^*

Předpokládejme, že známe spektrální rozklad matice $A^*A \in \mathbb{C}^{m \times m}$. Označme si λ_j její vlastní čísla a v_j její ortonormální vlastní vektory matice pro $j = 1, \dots, m$. Tato matice je pozitivně semidefinitní a tedy všechna její vlastní čísla jsou nezáporná. Rozepišme si vztahy mezi jejími vlastními čísly a vlastními vektory následujícím způsobem

$$A^*Av_j = \lambda_j v_j, \quad \|v_j\| = 1, \quad j = 1, \dots, m. \quad (6.8)$$

Předpokládejme dále, že tato vlastní čísla jsou uspořádána tak, že platí

$$\lambda_1 \geq \dots \geq \lambda_r > 0, \quad \lambda_{r+1} = \dots = \lambda_m = 0 \quad (6.9)$$

Maticově můžeme **diagonalizaci matice** A^*A zapsat pro $V = [v_1, \dots, v_m]$ následovně

$$V^*A^*AV = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0), \quad V^*V = I_m \quad (6.10)$$

Vidíme ale také

$$\text{span}\{v_1, \dots, v_r\} = \mathcal{R}(A^*A) = \mathcal{R}(A^*) \quad (6.11)$$

a tedy také

$$\text{span}\{v_{r+1}, \dots, v_m\} = \mathcal{N}(A^*A) = \mathcal{N}(A). \quad (6.12)$$

Zkusme teď najít vztah mezi vlastními čísly a vektory této matice A^*A a vlastními čísly a vektory matice AA^* .

Věta 6.1.3 *Uvažujme výše uvedený spektrální rozklad matice A^*A . Pak platí*

$$AA^*u_j = \lambda_j u_j, \quad \|u_j\| = 1, \quad j = 1, \dots, r.$$

Dále, vektory $u_j = Av_j/\sqrt{\lambda_j}$ pro $j = 1, \dots, r$ jsou ortonormální vlastní vektory AA^* .

Důkaz: Rekapitulace a stručný důkaz: Podle zadání máme v každém případě

$$A^*Av_j = \lambda_j v_j, \quad j = 1, \dots, r. \quad (6.13)$$

a tedy také

$$AA^*Av_j = \lambda_j Av_j, \quad \|v_j\| = 1, \quad j = 1, \dots, r \quad (6.14)$$

což jest

$$AA^*u_j = \lambda_j u_j, \quad j = 1, \dots, r, \quad (6.15)$$

kde jsou vektory $u_j, j = 1, \dots, r$ nenulové, protože příslušné vektory v_j nejsou v $\mathcal{N}(A)$.

Dále ukážeme, že vektory Av_j jsou ortogonální pro $j = 1, \dots, r$. Uvažujme indexy $j, k \in \{1, \dots, r\}, j \neq k$. Z ortogonality vektorů $v_j, j = 1, \dots, r$ pak dostaneme

$$(Av_j)^*(Av_k) = v_j^* A^* Av_k = \lambda_k v_j^* v_k = 0. \quad (6.16)$$

Navíc norma Av_j je rovna $\sqrt{\lambda_j}$ jak vidíme z úpravy

$$(Av_j)^*(Av_j) = v_j^* A^* Av_j = \lambda_j v_j^* v_j = \lambda_j. \quad (6.17)$$

Tudíž

$$\|u_j\| = 1, \quad (6.18)$$

■

Vidíme tedy, že u_1, \dots, u_r tvoří ortogonální bázi prostoru $\mathcal{R}(AA^*) = \mathcal{R}(A)$. Tuto bázi lze doplnit na bázi \mathbb{C}^n pomocí **nějakých** vektorů u_{r+1}, \dots, u_n z $\mathcal{N}(AA^*) = \mathcal{N}(A^*)$. Výsledek po tomto doplnění můžeme zapsat s použitím matice $U = [u_1, \dots, u_n]$ následovně

$$U^* AA^* U = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0), \quad U^* U = I_n \quad (6.19)$$

Pro následující odmocniny nenulových vlastních čísel si zavedeme následující speciální označení

Definice 6.1.1 *Odmocniny nenulových vlastních čísel matice A^*A pro $A \in \mathbb{C}^{n \times m}, \text{rank}(A) = r$ nazveme **singulární čísla** a budeme psát*

$$\sigma_j = \sqrt{\lambda_j}, \quad j = 1, \dots, r. \quad (6.20)$$

Rekapitulujme si nyní všechny čtyři fundamentální prostory spojené s lineárním zobrazením. Máme všechny fundamentální prostory: A prostřednictvím jejich bází následujícím přehledem.

$$\mathcal{R}(A^*) = \text{span}\{v_1, \dots, v_r\}, \quad \mathcal{N}(A) = \text{span}\{v_{r+1}, \dots, v_m\}, \quad (6.21)$$

$$\mathcal{R}(A) = \text{span}\{u_1, \dots, u_r\}, \quad \mathcal{N}(A^*) = \text{span}\{u_{r+1}, \dots, u_n\}. \quad (6.22)$$

6.1.3 Zápis singulárního rozkladu pomocí matic

Vraťme se k zobrazování prvků báze fundamentálních prostorů. Platí následující vztahy.

$$\begin{aligned}
 Av_1 &\rightarrow \sqrt{\lambda_1} \frac{Av_1}{\sqrt{\lambda_1}} \equiv \sigma_1 u_1 \\
 Av_2 &\rightarrow \sigma_2 u_2 \\
 &\vdots \rightarrow \vdots \\
 Av_r &\rightarrow \sigma_r u_r \\
 Av_{r+1} &\rightarrow 0 \\
 &\vdots \rightarrow \vdots \\
 Av_m &\rightarrow 0
 \end{aligned}$$

Příčemž

$$\mathcal{R}(A) = \text{span}\{u_1, \dots, u_r\}, \mathcal{N}(A) = \text{span}\{v_{r+1}, \dots, v_m\} \quad (6.23)$$

Maticově můžeme tedy psát následující vztah, kde $U \in \mathbb{C}^{n \times n}$, $V \in \mathbb{C}^{m \times m}$ jsou unitární matice příslušných dimenzí a matice $\Sigma \in \mathbb{R}^{n \times m}$ je diagonální.

$$AV = U\Sigma, \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}, \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \quad (6.24)$$

Tento vztah ale můžeme vidět i následujícím způsobem, pomocí lineárního zobrazení $A^* \in \mathbb{C}^{m \times n}$.

$$\begin{aligned}
 A^*u_1 &\rightarrow \sigma_1 v_1 \\
 A^*u_2 &\rightarrow \sigma_2 v_2 \\
 &\vdots \rightarrow \vdots \\
 A^*u_r &\rightarrow \sigma_r v_r \\
 A^*u_{r+1} &\rightarrow 0 \\
 &\vdots \rightarrow \vdots \\
 A^*u_n &\rightarrow 0
 \end{aligned}$$

Analogicky:

$$\mathcal{R}(A^*) = \text{span}\{v_1, \dots, v_r\}, \mathcal{N}(A^*) = \text{span}\{u_{r+1}, \dots, u_m\} \quad (6.25)$$

Maticově můžeme tedy opět psát analogicky pro U, V unitární a Σ diagonální a reálnou.

$$V^*A^* = \Sigma^T U, \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}, \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \quad (6.26)$$

Zrekapitulujme si to, co jsme nyní získali, ve formě věty.

Věta 6.1.4 Každou matici $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = r$ lze rozložit na součin

$$A = U\Sigma V^*, U \in \mathbb{C}^{n \times n}, V \in \mathbb{C}^{m \times m}, \Sigma \in \mathbb{R}^{n \times m}, \quad (6.27)$$

kde $U = (u_1 \ \dots \ u_n)$, $V = (v_1 \ \dots \ v_m)$ jsou unitární matice, $\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$, $\sigma_1 \geq \dots \geq \sigma_r > 0$. Tomuto rozkladu říkáme *singulární rozklad*.

Někdy se používá následující terminologie.

Definice 6.1.2 Vektory v_j , $j = 1, \dots, m$ se nazývají *pravé singulární vektory* příslušné výše uvedenému singulárnímu rozkladu. Vektory u_j , $j = 1, \dots, n$ se nazývají *levé singulární vektory*.

Odpověď na otázku o jednoznačnosti singulárního rozkladu je shrnuta v následující větě.

Věta 6.1.5 Uvažujme singulární rozklad matice $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = r$. Její singulární čísla $\sigma_1 \geq \dots \geq \sigma_r > 0$ jsou tímto rozkladem jednoznačně určena. Jsou-li navíc, například, všechna její singulární čísla navzájem různá, pak **jejich** levé i pravé singulární vektory jsou jednoznačně určeny až na násobení komplexními skalárními faktory, které mají absolutní hodnotu 1. Přitom volbou levého (pravého) singulárního vektoru určíme jednoznačně i ten druhý.

6.1.4 Ekonomický tvar singulárního rozkladu

Singulární rozklad můžeme zapsat v ekonomickém tvaru, který použije místo matice $\Sigma \in \mathbb{R}^{n \times m}$ jen její podmatici $\Sigma_r \in \mathbb{R}^{r \times r}$ a odpovídající podmatice matic U a V . Výsledný ekonomický tvar můžeme zapsat

$$A = U_r \Sigma_r V_r^*, U_r \in \mathbb{C}^{n \times r}, V_r \in \mathbb{C}^{m \times r}, \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathbb{R}^{r \times r}, \quad (6.28)$$

Grafické znázornění singulárního rozkladu a jeho ekonomického tvaru je na Obrázku 6.1.1.

Poznámka 6.1.1 Je-li matice A reálná, je i její singulární rozklad reálný. Toto tvrzení snadno plyne z faktu, že v reálném případě, tedy je-li $A \in \mathbb{R}^{n \times m}$, jsou matice A^*A i AA^* reálné symetrické pozitivně semidefinitní s reálnými ortonormálními vlastními vektory (speciální případ diagonalizace matice).

6.1.5 Singulární rozklad a podmíněnost

Zmiňme zde vztah singulárního rozkladu ke spektrálnímu rozkladu matice. Je-li matice $B \in \mathbb{C}^{n \times n}$ normální, pak můžeme psát

$$B = Q\Lambda Q^*, \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (6.29)$$

$$\begin{array}{c}
 \boxed{A} = \boxed{U} \begin{array}{|c|} \hline \boxed{\Sigma_r} \\ \hline \end{array} \boxed{\Sigma} \boxed{V^*} \\
 \\
 \boxed{A} = \boxed{U_r} \boxed{\Sigma_r} \boxed{V_r^*}
 \end{array}$$

Obrázek 6.1.1: Ilustrace singulárního rozkladu a jeho ekonomické varianty.

Pro B regulární pak můžeme psát

$$D = \text{diag}(\lambda_1/|\lambda_1|, \dots, \lambda_n/|\lambda_n|), DD^* = D^*D = I \quad (6.30)$$

a tedy

$$D^*\Lambda = \text{diag}(|\lambda_1|, \dots, |\lambda_n|) = \Lambda D^*. \quad (6.31)$$

Singulární rozklad normální a regulární matice B se tedy dá zapsat ve tvaru

$$B = (QD)(D^*\Lambda)Q^* = U\Sigma V^* \quad (6.32)$$

Vidíme, že singulární čísla jsou absolutní hodnoty vlastních čísel rozkládané matice B . Tato poznámka přímo naznačuje, že k výpočtu singulárního rozkladu můžeme použít spektrální rozklad symetrických a pozitivně semidefinitních matic $B = AA^*$ a $B = A^*A$, o kterých jsme hovořili výše. Konkrétně, je-li

- $n \gg m$, k výpočtu SVD lze použít spektrální rozklad matice $A^*A \in \mathbb{R}^{m \times m}$.
- $m \gg n$, k výpočtu SVD lze použít spektrální rozklad matice $AA^* \in \mathbb{R}^{n \times n}$.

Rozšířme nyní definici čísla podmíněnosti následujícím způsobem pro obdélníkové matice s plnou řádkovou nebo sloupcovou hodnotí, jak bude diskutováno na cvičeních.

$$\kappa(A) = \frac{\sigma_1}{\sigma_{\min(m,n)}} \quad (6.33)$$

Potom platí pro matici A s plnou sloupcovou hodnotí

$$\kappa(B) = \kappa(A^*A) = \kappa^2(A). \quad (6.34)$$

Analogický vztah platí pro matici s plnou řádkovou hodnotí, pro kterou A^* má plnou sloupcovou hodnotí. Toto zvětšení čísla podmíněnosti může významně zkomplikovat kvalitu výpočtu singulárního rozkladu založeného na diagonalizaci, a tedy hledání vlastních čísel příslušných hermitovských pozitivně definitních matic.

6.1.6 Poznámky k výpočtu singulárního rozkladu

Standardní výpočet singulárního rozkladu matice A je založen na dvou krocích. **První krok** tuto matici převede na bidiagonální tvar. Následující schéma ukazuje transformovanou matici v horním bidiagonálním tvaru.

$$\begin{pmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & * & \\ & & & & * & * \\ & & & & & * \end{pmatrix} \quad (6.35)$$

Původně hustou matici můžeme na tento tvar převést pomocí unitárních transformací **zleva** i **zprava**. Následující obrázek tento postup schématicky ukazuje. Konkrétně vidíme, že aplikací unitárních transformací zleva i zprava sice nedokážeme matici diagonalizovat, protože neexistuje žádné přípustné pořadí, které by zároveň nezaplňovalo dříve vynulované prvky matice. Nicméně, bidiagonalizovat matici je takto možné.

$$\begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & & & & \\ & * & * & * & * & * \\ & * & * & * & * & * \\ & * & * & * & * & * \\ & * & * & * & * & * \\ & * & * & * & * & * \end{pmatrix} \rightarrow \quad (6.36)$$

$$\begin{pmatrix} * & * & & & & \\ & * & * & * & * & * \\ & & * & * & * & * \\ & & * & * & * & * \\ & & * & * & * & * \\ & & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & * & * \\ & & * & * & * & * \\ & & * & * & * & * \\ & & * & * & * & * \end{pmatrix} \rightarrow \dots \rightarrow \begin{pmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & * & \\ & & & & * & * \\ & & & & & * \end{pmatrix} \quad (6.37)$$

Druhým krokem rozkladu je nalezení singulárního rozkladu této transformované bidiagonální matice. Jednou z nejpoužívanějších procedur, která počítá všechna singulární čísla, je varianta QR algoritmu. Dodejme, že tento postup je navíc velmi přesný, neboť singulární čísla spočítá s relativní přesností na úrovni strojové přesnosti. Z předcházejících úvah víme, že takový algoritmus je iterační.

Souhrnně, **počet operací** singulárního rozkladu je **přibližně** (spodní odhad) $8/3n^3$, počítáme-li jen diagonální matici Σ , čili pouze singulární čísla. Počítáme-li také unitární matice, pak je spodním odhadem počtu operací výraz $16/3n^3$

6.1.7 Inverze a pseudoinverze matice

Definice 6.1.3 Uvažujme matici $A \in \mathbb{C}^{n \times m}$ hodnosti r se singulárním rozkladem $A = U\Sigma V^*$. Matici

$$A^\dagger = V\Sigma^\dagger U^* = V_r \Sigma_r^{-1} U_r^* \quad (6.38)$$

nazveme její Moore-Penroseovou zobecněnou inverzí (pseudoinverzí), kde

$$\Sigma^\dagger = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} \quad (6.39)$$

Tato pseudoinverze je jednoznačně určená.

Dá se ukázat, že tato matice splňuje Moore-Penroseovy vztahy

$$AA^\dagger A = A, (AA^\dagger)^* = AA^\dagger, A^\dagger AA^\dagger = A^\dagger, (A^\dagger A)^* = A^\dagger A, \quad (6.40)$$

kterými se někdy definují různé **zobecněné maticové inverze** včetně pseudoinverze. V případě plné sloupcové nebo plné řádkové hodnosti matice A můžeme tvar Moore-Penroseovy pseudoinverze explicitně napsat pomocí dané matice A a její konjugované transpozice A^* způsobem, který popíšeme v následujícím textu.

Věta 6.1.6 (Případ plné sloupcové hodnosti matice) Necht' $A \in \mathbb{C}^{n \times m}$, $n \geq m$ a také $\text{rank}(A) = m$. Pak platí

$$A^\dagger = (A^*A)^{-1}A^* \quad (6.41)$$

Důkaz: Všimněme si, že samozřejmě $n \geq m$ platit musí. Pak máme

$$\begin{aligned} A^\dagger &= V_m \Sigma_m^{-1} U_m^* \\ &= (V_m \Sigma_m^{-2} V_m^*) (V_m \Sigma_m U_m^*) \\ &= (V_m \Sigma_m^{-2} V_m^*) A^* \\ &= (V_m \Sigma_m^2 V_m^*)^{-1} A^* \\ &= (A^*A)^{-1} A^*, \end{aligned}$$

což jsme měli ukázat. ■

Věta 6.1.7 (Případ plné řádkové hodnosti matice) Necht' $A \in \mathbb{C}^{n \times m}$, $n \leq m$ a také $\text{rank}(A) = n$. Pak platí

$$A^\dagger = A^*(AA^*)^{-1} \quad (6.42)$$

Důkaz: Je analogický důkazu předcházejícího tvrzení.

$$\begin{aligned} A^\dagger &= V_n \Sigma_n^{-1} U_n^* \\ &= (V_n \Sigma_n U_n^*) (U_n \Sigma_n^{-2} U_n^*) \\ &= (V_n \Sigma_n U_n^*) (U_n \Sigma_n^2 U_n^*)^{-1} \\ &= (V_n \Sigma_n U_n^*) ((U_n \Sigma_n V_n^*) (V_n \Sigma_n U_n^*))^{-1} \\ &= A^*(AA^*)^{-1}, \end{aligned}$$

což jsme měli ukázat. ■

Věta 6.1.8 Pro $A \in \mathbb{C}^{n \times m}$ čtvercovou a regulární matici platí

$$A^\dagger = A^{-1} \quad (6.43)$$

Všimněme si dále následujícího faktu.

Věta 6.1.9 Pro matici $A \in \mathbb{C}^{n \times m}$ platí, že

$$AA^\dagger = U_r U_r^* \quad (6.44)$$

je ortogonálním projektorem na $\mathcal{R}(A)$ a

$$A^\dagger A = V_r V_r^* \quad (6.45)$$

je ortogonálním projektorem na $\mathcal{R}(A^*)$.

Důkaz: Snadno ověříme obě vlastnosti ortogonálního projektoru, tedy jeho **idempotenci** a **samosdruženost**. Ukážeme to pouze pro první tvrzení. Důkaz druhého tvrzení je analogický.

To, že matice $U_r U_r^*$ zobrazuje z \mathbb{C}^n do $\mathcal{R}(A)$ je zřejmé. I proto, že $U_r \in \mathbb{C}^{n \times r}$ a $U_r^* \in \mathbb{C}^{r \times n}$. Pro $P_r = U_r U_r^*$ platí

$$P_r^2 = U_r U_r^* U_r U_r^* = P_r$$

a $U_r U_r^*$ je tedy idempotentní. Samosdruženost matice $U_r U_r^*$ plyne z následujících ekvivalencí:

$$\langle P_r x, y \rangle = \langle U_r U_r^* x, y \rangle = \langle U_r^* x, U_r^* y \rangle = \langle x, U_r U_r^* y \rangle = \langle x, P_r y \rangle$$

pro příslušné vektory. ■

6.1.8 Normy a podmíněnost vyjádřené s pomocí singulárních čísel

Věta 6.1.10 Pro matici $A \in \mathbb{C}^{n \times m}$ hodnosti r se singulárními čísly $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ máme

$$\begin{aligned} \|A\| &= \sigma_1 \\ \|A\|_F &= (\sigma_1^2 + \dots + \sigma_r^2)^{1/2} \\ \|A\| &\leq \|A\|_F \leq \sqrt{r} \|A\| \end{aligned}$$

Důkaz: Platnost prvních dvou vztahů je jasná z následujícího rozepsání

$$\begin{aligned} \|A\| &= (\rho(A^* A))^{1/2} \Rightarrow \|A\| = \sigma_1 \\ \|A\|_F &= \|U \Sigma V^*\|_F = \|\Sigma\|_F = (\sigma_1^2 + \dots + \sigma_r^2)^{1/2} \end{aligned}$$

a pak je zřejmý i ten poslední vztah. ■

Připomeňme si nyní ještě dříve probírané číslo podmíněnosti a související veličiny

Věta 6.1.11 Pro regulární čtvercovou matici $A \in \mathbb{C}^n$ platí kromě výše uvedeného $\|A\| = \sigma_1$ také

$$\|A^{-1}\| = \sigma_n^{-1} \quad (6.46)$$

Číslo podmíněnosti tak můžeme vyjádřit vztahem

$$\kappa(A) = \frac{\sigma_1}{\sigma_n} \quad (6.47)$$

Poznámka 6.1.2 Zavedeme-li následující značení pro regulární čtvercovou matici $A \in \mathbb{C}^n$

$$\maxmag(A) = \max_{\|x\|=1} \|Ax\| \equiv \|A\|, \quad \minmag(A) = \min_{\|x\|=1} \|Ax\|, \quad (6.48)$$

pak můžeme psát

$$\kappa(A) = \maxmag(A) / \minmag(A) \quad (6.49)$$

Ačkoli to vypadá, že nové označení nepřináší nic nového, nově vyjádřené číslo podmíněnosti naznačuje, jak souvisí s maximální a minimální dilatací (**magnifikací, zvětšením**) jednotkových vektorů.

6.1.9 Aproximace pomocí SVD s využitím dyadického rozvoje

Singulární rozklad obecné matice můžeme napsat ve tvaru dyadického rozvoje následujícím způsobem

$$A = \sum_{j=1}^r \sigma_j u_j v_j^* = \sum_{j=1}^r A_j \quad (6.50)$$

a tím se vracíme k dřívější vyjádřením, které se týkaly normálních matic. Analogicky, jako pro dyadický rozvoj diagonalizovatelných matic máme pro normy matic hodnoti 1

$$\|A_j\| = \|A_j\|_F = \sigma_j \quad (6.51)$$

a tudíž také

$$\|A_1\| \geq \dots \geq \|A_r\| > 0. \quad (6.52)$$

Dyadický rozvoj lze použít k aproximaci matice A . Praktičnost takového vyjádření vyplývá z faktu, že singulární čísla jsou v singulárním rozkladu seřazena podle velikosti od největších k nejmenším. Konkrétně, označíme-li

$$A^{(k)} = \sum_{j=1}^k \sigma_j u_j v_j^*, \quad (6.53)$$

pak můžeme matici A aproximovat nějakým částečným součtem tohoto tvaru. Nejprve si dokažme pomocné lemma.

Lemma 6.1.1 Nechť $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = r$. Pak pro každou matici $X \in \mathbb{C}^{n \times m}$ hodnosti k , kde $k < r$ platí

$$\|A - X\| \geq \sigma_{k+1}. \quad (6.54)$$

Důkaz: Uvažujme matici X splňující podmínky lemmatu. Dimenze jejího nulového prostoru $\mathcal{N}(X)$ je tedy rovna $m - k$. Uvažujme prostor složený z $k + 1$ pravých singulárních vektorů matice A , tedy prostor

$$V = \text{span}\{v_1, \dots, v_{k+1}\}, \quad (6.55)$$

který je zobrazen na podprostor z příslušných levých singulárních vektorů. Oba prostory V a $\mathcal{N}(X)$ jsou podprostory \mathbb{C}^m a součet jejich dimenzí je $m + 1$. Musí tedy existovat jejich netriviální průnik. Jinými slovy, existuje nenulový vektor y takový, že

$$y \in V \cap \mathcal{N}(X) \quad (6.56)$$

a tedy

$$\begin{aligned} y \in V &\Rightarrow y = \sum_{i=1}^{k+1} \alpha_i v_i, \\ y \in \mathcal{N}(X) &\Rightarrow Xy = 0 \end{aligned}$$

Zvolíme-li y s jednotkovou normou, což pro nenulové y můžeme, pak pro jeho koeficienty získáme implikaci

$$\|y\| = 1 \Rightarrow 1 = \sum_{i=1}^{k+1} |\alpha_i|^2. \quad (6.57)$$

Upravme nyní vztah pro $(A - X)y$ následovně

$$(A - X)y = Ay - Xy = Ay - 0 = \sum_{i=1}^{k+1} \alpha_i Av_i = \sum_{i=1}^{k+1} \alpha_i \sigma_i u_i. \quad (6.58)$$

Pro eukleidovskou normu z toho dostaneme

$$\|(A - X)y\|^2 = \sum_{i=1}^{k+1} |\alpha_i|^2 |\sigma_i|^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} |\alpha_i|^2 = \sigma_{k+1}^2. \quad (6.59)$$

Protože ale pro maticovou normu máme

$$\|A - X\|^2 \geq \|(A - X)y\|^2, \quad (6.60)$$

dostáváme tvrzení věty. ■

Následující věta vztah optimality aproximace formalizuje.

Věta 6.1.12 (*Eckart-Young-Mirsky*) *Nechť $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = r$ a také $k < r$. Potom platí, že matice $A^{(k)}$ realizuje minimum vzdálenosti od matice A v eukleidovské normě. Tedy,*

$$\|A - A^{(k)}\| = \min_{X \in \mathbb{C}^{n \times m}, \text{rank}(X)=k} \|A - X\|. \quad (6.61)$$

Přitom platí

$$\|A - A^{(k)}\| = \sigma_{k+1}. \quad (6.62)$$

Důkaz: Protože matice $A^{(k)} \in \mathbb{C}^{n \times m}$ definovaná výše jako

$$A^{(k)} = \sum_{j=1}^k \sigma_j u_j v_j^*, \quad (6.63)$$

splňuje předpoklady Věty a přitom platí

$$\|A - A^{(k)}\| = \left\| \sum_{i=1}^r \sigma_i u_i v_i^* - \sum_{i=1}^k \sigma_i u_i v_i^* \right\| = \sigma_{k+1}, \quad (6.64)$$

tvrzení plyne z předchozího lemmatu. ■

Kapitola 7

Problém nejmenších čtverců

Tato kapitola se věnuje problému nejmenších čtverců, který byl již zmíněn výše. Motivační formulace tohoto problému je snaha o dobrou aproximaci řešení soustavy v případě, kdy je matice soustavy obecně obdélníková. Uvidíme, že tento problém přirozeně souvisí s problémem regrese v matematické statistice.

7.1 Dvě základní úlohy: problém a úplný problém nejmenších čtverců

Rozlišme dva základní a související problémy. Uvažujme $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$. Neexistuje-li přesné řešení problému a chceme-li řešit soustavu

$$Ax \approx b \tag{7.1}$$

alespoň přibližně, máme dvě základní možnosti, jak definovat **chybu splnění** této soustavy.

7.1.1 Definice dvou problémů

První možnost je dána předpokladem, že nesplníme úplně soustavu a budeme se snažit o co nejmenší **odchylku pravé strany** b , jinými slovy o co nejmenší rezíduum. Druhá možnost je, že připustíme i **odchylku v matici** A . Oba dva základní případy tedy nesplnění formulují jako odchylku v zadaných **datech** problému. Tedy je formulují jako přesné řešení změněného problému podle definice zpětné chyby.

Definice 7.1.1 *Problémem nejmenších čtverců (LS) pro $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$ je nalézt vektor $x \in \mathbb{C}^m$ a příslušné rezíduum $r \in \mathbb{C}^n$ tak, že je splněn vztah*

$$\min_{x,r} \|r\|, Ax = b + r \quad \text{nebo (jinak)} \quad \min_x \|b - Ax\| \tag{7.2}$$

Definice 7.1.2 *Úplný problém nejmenších čtverců (TLS) pro $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$ je nalézt vektor $x \in \mathbb{C}^m$, příslušné rezíduum $r \in \mathbb{C}^n$ a chybovou matici $E \in \mathbb{C}^{n \times m}$ splňující*

$$\min_{x,E,r} \|(Er)\|_F, (A + E)x = b + r \tag{7.3}$$

(Er) zde přitom reprezentuje matici o $m + 1$ sloupcích složenou z rezídua a chybové matice.

7.1.2 Unitární invariance problému

Důležitým pozorováním je, že oba zavedené problémy jsou **unitárně invariantní**. Je-li matice Q odpovídající dimenze unitární, pak platí pro LS, že $\|Qr\| = \|r\|$. Analogicky můžeme znázornit unitární invarianci v případě TLS. Je totiž $\|Q[r E]\|_F = \|[r E]\|_F$.

7.2 LS a problémy ve statistice

Uvažujme následující problém v reálném oboru. Mějme n pozorování tvaru (x_i, y_i) pro $i = 1, \dots, n$ a předpokládejme lineární závislost mezi vstupy x_i a výstupy y_i tvaru

$$y_i = \alpha + \beta x_i + r_i, \quad i = 1, \dots, n. \quad (7.4)$$

Tento systém můžeme formálně zapsat také

$$Y = X[\alpha \beta]^T + R \quad (7.5)$$

pro daná $Y \in \mathbb{R}^n$, $R \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times 2}$. Problému nalézt dvě proměnné α a β minimalizující

$$\|Y - X[\alpha \beta]^T\| \quad (7.6)$$

se říká ve statistice problém **jednoduché regrese**. Zobecněním tohoto problému je problém obecné **lineární regrese**, kde hledáme více neznámých. Matice X je obecně z prostoru $\mathbb{R}^{n \times m}$ a říká se jí někdy návrhová matice (matice regresorů). Problém TLS se nazývá problém **ortogonální regrese** a je zobecněním problému lineární regrese.

7.3 Charakterizace řešení LS

Řešení problému zde uvedeme nejprve ve formě několika tvrzení.

Věta 7.3.1 Pro $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$ platí, že $x \in \mathbb{C}^m$ je řešení LS právě tehdy, platí-li

$$Ax = b \mid_{\mathcal{R}(A)}, \quad \|b - Ax\| = \|b \mid_{\mathcal{N}(A^*)}\| \quad (7.7)$$

Důkaz: Neboť můžeme psát

$$\mathbb{C}^n = \mathcal{R}(A) \oplus \mathcal{N}(A^*) \quad (7.8)$$

pak lze také pravou stranu LS problému b rozložit na složky v těchto dvou fundamentálních prostorech následujícím způsobem

$$b = b \mid_{\mathcal{N}(A^*)} + b \mid_{\mathcal{R}(A)}. \quad (7.9)$$

Pro obecné $z \in \mathbb{R}^m$ pak můžeme postupně psát

$$\begin{aligned} b - Az &= b|_{\mathcal{N}(A^*)} + b|_{\mathcal{R}(A)} - Az, \\ \|b - Az\|^2 &= \|b|_{\mathcal{N}(A^*)}\|^2 + \|b|_{\mathcal{R}(A)} - Az\|^2. \end{aligned}$$

Druhá z těchto ekvivalencí je vyjádřením Pythagorovy věty. Druhý člen na její pravé straně lze vhodnou volbou z položit roven nule, protože odpovídající rovnice je řešitelná. První člen na pravé straně naopak nejsme schopni volbou z ovlivnit. Norma rezídua je tedy nejmenší právě tehdy, je-li ten druhý člen roven nule. ■

Poznámka 7.3.1 *Je-li $b \perp \mathcal{R}(A)$, pak pravá strana LS není korelována s daty a jediným rozumným praktickým řešením je $x = 0$.*

Důkaz: Je zřejmé, že $b|_{\mathcal{R}(A)}$ je nulový vektor, výraz $\|b|_{\mathcal{R}(A)} - Az\|^2$ nabízí triviální řešení $x = 0$ (v případě LN sloupců je to jediné takové řešení). Pouze další informace nebo podmínky by mohly motivovat výběr jiného řešení. ■

Věta 7.3.2 *Jsou-li sloupce A jsou lineárně nezávislé, pak lze řešení LS jednoznačně vyjádřit.*

Poznámka 7.3.2 *Pokud jsou sloupce A lineárně závislé, pak je fundamentální prostor $\mathcal{N}(A)$ netriviální. Pak platí, že $x + z$ pro $z \in \mathcal{N}(A)$ je také řešení. Protože z ale ničím nepřispívá, je logické hledat řešení LS minimální v normě. Uvidíme, že takové řešení je jednoznačné.*

7.4 Řešení problému nejmenších čtverců minimální v normě

Věta 7.4.1 *Pro $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$ platí, že existuje právě jedno řešení $x \in \mathbb{C}^m$ problému LS, které je minimální v normě. Toto řešení je dáno vztahy*

$$Ax = b|_{\mathcal{R}(A)}, x \in \mathcal{R}(A^*). \quad (7.10)$$

Důkaz:

Zvolme $x \in \mathbb{C}^m$. Uvažujme rozklad \mathbb{C}^m na fundamentální prostory. Pak můžeme tento vektor rozepsat jako následující direktní součet

$$x = x|_{\mathcal{N}(A)} + x|_{\mathcal{R}(A^*)}. \quad (7.11)$$

Pro LS řešení platí

$$b|_{\mathcal{R}(A)} = Ax = Ax|_{\mathcal{N}(A)} + Ax|_{\mathcal{R}(A^*)} = 0 + Ax|_{\mathcal{R}(A^*)} \quad (7.12)$$

Je tedy jasné, že $x|_{\mathcal{R}(A^*)}$ je řešením LS problému, Protože ale také platí, že

$$\|x\|^2 = \|x|_{\mathcal{N}(A)}\|^2 + \|x|_{\mathcal{R}(A^*)}\|^2, \quad (7.13)$$

tato norma je minimální právě tehdy, platí-li $x|_{\mathcal{N}(A)} = 0$. Řešení LS problému minimální v normě tedy musí být v $\mathcal{R}(A^*)$. Zbývá ještě ukázat, že toto řešení je jednoznačné. Mějme dvě taková řešení, která si označíme x a y . Pak můžeme psát

$$\begin{aligned} A(x - y) &= b|_{\mathcal{R}(A)} - b|_{\mathcal{R}(A)} = 0, \\ x - y &\in \mathcal{N}(A) \end{aligned}$$

ale platí také

$$x - y \in \mathcal{R}(A^*)$$

Jediný vektor, který mají společný oba fundamentální podprostory, je vektor nulový a proto

$$x = y. \quad (7.14)$$

■

7.4.1 Normální rovnice

Soustava **normálních rovnic** je pojmem, který s LS problémem velmi souvisí. Ne vždy je tato soustava cestou k nejlepšímu praktickému postupu řešení ve smyslu rychlosti nebo stability, ale v teorii nejmenších čtverců hraje velmi významnou roli.

Věta 7.4.2 *Nechť $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$. Potom platí, že $x \in \mathbb{C}^m$ je řešením LS problému právě tehdy, je-li také řešením soustavy normálních rovnic*

$$A^*Ax = A^*b \quad (7.15)$$

Důkaz: Je-li x řešením LS problému, pak platí $Ax = b|_{\mathcal{R}(A)}$ a vynásobením této soustavy maticí A^* zleva dostaneme

$$A^*Ax = A^*b|_{\mathcal{R}(A)}. \quad (7.16)$$

Tuto rovnost rozepíšeme následujícím způsobem

$$A^*Ax = A^*b|_{\mathcal{R}(A)} = A^*(b|_{\mathcal{R}(A)} + b|_{\mathcal{N}(A^*)}) = A^*b \quad (7.17)$$

a první implikace je ukázána.

Opačně, nechť je soustava normálních rovnic splněna. Pak platí také $A^*Ax = A^*b|_{\mathcal{R}(A)}$ a tedy i $A^*(Ax - b|_{\mathcal{R}(A)}) = 0$. To ale implikuje rovnost

$$Ax - b|_{\mathcal{R}(A)} \in \mathcal{N}(A^*) \quad (7.18)$$

Protože ale platí zároveň

$$Ax - b|_{\mathcal{R}(A)} \in \mathcal{R}(A), \quad (7.19)$$

pak musí být $Ax = b|_{\mathcal{R}(A)}$, neboť rozdíl vektorů na levé straně posledních dvou rovností je v průniku dvou komplementárních fundamentálních podprostorů matice A a musí tedy být roven nule. ■

7.5 Způsoby řešení problému nejmenších čtverců

Tato kapitola se bude věnovat praktickým procedurám řešení LS problému, tedy problému nejmenších čtverců. Uvedeme si několik postupů, přičemž budeme v prvních třech případech předpokládat, že matice soustavy $A \in \mathbb{C}^{n \times m}$ má **plnou sloupcovou hodnotnost**. Předpokládáme tedy $n \geq m$. V případě, že matice nemá plnou sloupcovou hodnotnost, řešící algoritmy bývají komplikovanější, protože musí používat obvykle kombinaci více technik.

7.5.1 QR rozklad

Uvažujme QR rozklad ve tvaru

$$A = QR = Q \begin{pmatrix} R_m \\ 0 \end{pmatrix}. \quad (7.20)$$

Pak můžeme psát

$$\min_x \|b - Ax\| = \min_x \|b - QRx\| = \min_x \|Q^*(b - QRx)\| = \min_x \|Q^*b - Rx\| \quad (7.21)$$

Výraz na pravé straně můžeme podrobněji rozepsat s použitím následujícího značení

$$\begin{aligned} Q &= [q_1, \dots, q_n] = \begin{pmatrix} Q_m & Q_m^\perp \end{pmatrix} \\ \text{span}([q_1, \dots, q_m]) &= \mathcal{R}(A) \\ \text{span}([q_{m+1}, \dots, q_n]) &= \mathcal{R}(A)^\perp \equiv \mathcal{N}(A^*) \end{aligned}$$

Pak vidíme, že

$$Q^*b - Rx = \begin{pmatrix} Q_m^*b \\ (Q_m^\perp)^*b \end{pmatrix} - \begin{pmatrix} R_m \\ 0 \end{pmatrix} x = \begin{pmatrix} Q_m^*b - R_mx \\ (Q_m^\perp)^*b \end{pmatrix} \quad (7.22)$$

Vztah musí platit po blokových komponentách, což dává

$$\begin{aligned} R_mx &= Q_m^*b \\ \min_x \|b - Ax\| &= \|(Q_m^\perp)^*b\| \end{aligned}$$

Pouze první bloková rovnice obsahuje vektor x . Navíc, vzhledem k tomu, že jsme předpokládali plnou sloupcovou hodnotnost matice A , pak je matice R_m regulární, to jest, má na diagonále nenulová čísla. Vektor Q_m^*b je vektor dimenze m a tato první bloková rovnice má tedy právě jedno řešení. Druhá bloková komponenta obsahuje reziduum LS problému. V praxi se matice Q^* může konstruovat pouze implicitně a aplikovat zároveň s rozkladem i na pravou stranu. Má-li matice A dobře určenou plnou sloupcovou hodnotnost, je R_m regulární a použití QR rozkladu pro řešení LS problému je vhodné.

7.5.2 Použití normálních rovnic

Snadno nahlédneme, že v případě plné sloupcové hodnosti matice A je postup dobře definovaný a ekvivalentní použití QR rozkladu, jak vidíme z následující sady ekvivalencí.

$$A^*Ax = A^*b \Leftrightarrow R^*Rx = R^*Q^*b \Leftrightarrow R_m^*R_mx = R_m^*Q_m^*b \Leftrightarrow R_mx = Q_m^*b \quad (7.23)$$

Samotný termín řešení soustavy normálních rovnic, ale obvykle naznačuje, že použijeme Choleského faktorizaci. To lze učinit na základě následujících ekvivalencí, přičemž plná sloupcová hodnost ukazuje, že Choleského faktorizaci lze korektně použít.

$$A^*Ax = A^*b, A^*A = LL^* \Leftrightarrow x = (LL^*)^{-1}A^*b \quad (7.24)$$

Postup může být výhodný, je-li $n \gg m$. Dimenze soustavy normálních rovnic je totiž právě m . Na druhou stranu, nevýhoda postupu je v následující vlastnosti týkající se podmíněnosti matice soustavy normálních rovnic.

$$\kappa(A^*A) = \frac{\sigma_1(A^*A)}{\sigma_m(A^*A)} = \frac{\sigma_1^2(A)}{\sigma_m^2(A)} = \kappa^2(A). \quad (7.25)$$

Vidíme, že číslo podmíněnosti systému normálních rovnic je kvadrátem původního čísla podmíněnosti, což znamená, že řešení je mnohem citlivější na změny v rezíduu.

7.5.3 Řešení rozšířeného systému

Napišme si nejprve vyjádření soustavy, ze které získáme řešení LS problému a definujme výsledné rezíduum výrazem δ .

$$\begin{aligned} Ax &= b \mid_{\mathcal{R}(A)} \\ \delta &= b - Ax = b - b \mid_{\mathcal{R}(A)} \in \mathcal{N}(A^*) \\ &\Downarrow \\ A^*\delta &= 0 \end{aligned}$$

Jak jsme viděli dostali jsme soustavu dvou soustav rovnic, kterou zapíšeme kompaktně následujícím způsobem.

$$\begin{pmatrix} I & A \\ A^* & 0 \end{pmatrix} \begin{pmatrix} \delta \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \quad (7.26)$$

Tomuto ekvivalentnímu tvaru řešení LS problému se říká **rozšířená soustava** rovnic. Příslušná matice soustavy se nazývá **rozšířená** matice. Klíčem k použití této soustavy v řešení LS problému je následující věta, kde uvažujeme mírně obecnější verzi rozšířené matice.

Lemma 7.5.1 *Nechť $A \in \mathbb{C}^{n \times m}$, $\text{rank}(A) = m$, $n \geq m$, $B \in \mathbb{C}^{n \times n}$ HPD. Pak je matice*

$$C = \begin{pmatrix} B & A \\ A^* & 0 \end{pmatrix} \quad (7.27)$$

regulární, hermitovská a indefinitní.

Důkaz: Rozšířená matice je zřejmě **hermitovská**. To, že je také **regulární**, dokážeme sporem. Uvažujme rozšířenou soustavu s nulovou pravou stranou. Neboť $B\delta + Ax = 0$ a matice B je regulární, pak dostáváme

$$\delta + B^{-1}Ax = 0. \quad (7.28)$$

Z druhé rovnice v rozšířeném systému plyne $A^*\delta = 0$ a dosazením do předchozího vztahu vynásobeného maticí A^* zleva pak dostaneme

$$A^*\delta + A^*B^{-1}Ax = A^*B^{-1}Ax = 0. \quad (7.29)$$

Ale $A^*B^{-1}A$ je regulární (A má podle zadání plnou sloupcovou hodnost) a proto $x = 0$, tudíž i $\delta = 0$. Jinak lze vidět regularitu i z toho, že $-A^*B^{-1}A$ je Schurův doplněk regulární matice B v C . Protože je B regulární, je-li i C regulární.

Třetí vlastnost, kterou máme dokázat, je **indefinitnost** matice C . Jinými slovy, abychom dokázali, že matice C je indefinitní, musíme dokázat, že má kladná i záporná vlastní čísla. Uvažujme **nenulový** vektor $(0 \ \dots \ 0 \ v_{n+1} \ \dots \ v_{m+n})^T$. Platí $v^*Cv = 0$. Uvažujme Schurův rozklad hermitovské matice (Λ je diagonální) v následujícím tvaru

$$C = S\Lambda S^* = \begin{pmatrix} s_1 & \dots & s_{n+m} \end{pmatrix} \text{diag}(\lambda_1, \dots, \lambda_{n+m}) \begin{pmatrix} s_1 & \dots & s_{n+m} \end{pmatrix}^* \quad (7.30)$$

$$0 = v^*S\Lambda S^*v = \sum_{i=1}^{n+m} \lambda_i |s_i^*v|^2 \quad (7.31)$$

Protože je C regulární, pak jsou všechna vlastní čísla nenulová. Dále, protože je v nenulový, pak musí platit $S^*v \neq 0$, protože vektory v S tvoří bázi celého prostoru $\mathbb{C}^{(n+m) \times (n+m)}$. Dále tedy musí existovat **kladná i záporná** nenulová vlastní čísla, aby ta suma dala dohromady nulu. ■

7.5.4 Singulární rozklad

Uvažujme ekonomický singulární rozklad $A = U_r \Sigma_r V_r^*$. Řešení problému nejmenších čtverců minimální v normě je dáno vztahy

$$Ax = b \mid_{\mathcal{R}(A)}, \quad x \in \mathcal{R}(A^*). \quad (7.32)$$

Protože transformace pravé strany je dána projekcí na $\mathcal{R}(A)$, platí $b \mid_{\mathcal{R}(A)} = U_r U_r^* b$ a můžeme psát

$$U_r \Sigma_r V_r^* x = U_r U_r^* b \quad (7.33)$$

Matice U_r reprezentuje ortonormální bázi $\mathcal{R}(A)$ a proto se rovnají i souřadnice v této bázi. To jest,

$$V_r^* x = \Sigma_r^{-1} U_r^* b \quad (7.34)$$

Protože hledáme $x \in \mathcal{R}(A^*)$, pak je můžeme vyjádřit ve tvaru $x = V_r y$, protože V_r je báze fundamentálního prostoru $\mathcal{R}(A^*)$. Tedy

$$y = V_r^* x = \Sigma_r^{-1} U_r^* b. \quad (7.35)$$

Tvrzení, které jsme odvodili, můžeme zformulovat jako následující větu.

Věta 7.5.1 *Nechť $A \in \mathbb{C}^{n \times m}$, $b \in \mathbb{C}^n$ a $A = U_r \Sigma_r V_r^*$ je ekonomický singulární rozklad matice A . Pak platí, že*

$$x = V_r \Sigma_r^{-1} U_r^* b = A^\dagger b \quad (7.36)$$

je řešení problému nejmenších čtverců minimální v normě.

Kapitola 8

Problém vlastních čísel

V této kapitole budeme diskutovat řešení problému vlastních čísel, který v obecné formulaci hledá vlastní čísla a jim příslušné vlastní vektory. Obecně se hovoří o dvou typech tohoto problému: o **částečném** problému vlastních čísel a **úplném** problému vlastních čísel. Pro řešení prvního typu problému, jehož úkolem je najít nějaká, často extrémální, vlastní čísla a jim příslušné vlastní vektory, zde uvedeme několik možných postupů. Různé přístupy, které zde zmíníme, se rozlišují mimo jiné podle typu matice, které se problém vlastních čísel týká.

8.1 Částečný problém vlastních čísel

Základním algoritmem této skupiny je mocinná metoda, která v nejjednodušším případě hledá pouze jedno **dominantní** vlastní číslo a příslušný vlastní vektor dané matice. Dominantním vlastním číslem myslíme to vlastní číslo, které je v absolutní hodnotě největší. V následujícím algoritmu existenci jediného dominantního vlastního čísla předpokládáme.

8.1.1 Mocinná metoda

Algoritmus mocinné metody je uveden níže. Pro matici $A \in \mathbb{C}^{n \times n}$ a počáteční vektor $v_1 \in \mathbb{C}^n$ spočívá tento algoritmus v tvoření posloupnosti

$$v_1, v_2 = Av_1, \dots, v_k = Av_{k-1}, k = 1, \dots \quad (8.1)$$

Tyto vektory jsou v každém průchodu cyklem algoritmu normalizovány. Poměrový vztah na řádku 5 algoritmu, kterým se počítá odhad vlastního čísla, přepsaný pro nenormalizovaný vektor w má tvar

$$\mu = \frac{w^T Aw}{w^T w} \quad (8.2)$$

se nazývá **Rayleighův kvocient**.

Algoritmus 8.1.1 Mocinná metoda

Input: Matice $A \in \mathbb{C}^{n \times n}$ s vlastními čísly $\lambda_1, \dots, \lambda_n$, $|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|$,

počáteční vektor $v_0 \in \mathbb{C}^n$

Output: Aproximace μ_k , v_k jediného dominantního vlastního čísla λ_1 a jemu příslušného vlastního vektoru.

1. $v_0 = v_0 / \|v_0\|$
2. **for** $k = 1, 2, \dots$ **do**
3. $w = Av_{k-1}$
4. $v_k = w / \|w\|$
5. $\mu_k = v_k^* Av_k$
6. **end** k

Následující věta obsahuje základní výsledek, který se týká konvergence mocninné metody pro jednoduchý speciální případ matice A . Konkrétně předpokládáme, že matice A je hermitovská, protože to je v souladu s dalším standardním dělením problémů vlastních čísel, a sice na problémy hermitovské a nehermitovské.

Věta 8.1.1 *Nechť $A \in \mathbb{R}^{n \times n}$ je hermitovská matice s vlastními čísly $\lambda_1, \dots, \lambda_n$, pro která platí*

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n|. \quad (8.3)$$

Nechť u_1 je vlastní vektor matice A takový, že $Au_1 = \lambda_1 u_1$. Uvažujme algoritmus mocninné metody. Nechť startovací normalizovaný vektor $v_0 \in \mathbb{C}^n$ splňuje $\alpha_1 = u_1^ v_0 \neq 0$. Pak platí*

$$\lim_{k \rightarrow \infty} \frac{v_k}{\operatorname{sgn}(\alpha_1 \lambda_1^k)} = \frac{u_1}{\|u_1\|} \quad (8.4)$$

a také

$$\lim_{k \rightarrow \infty} \mu_k \equiv v_k^* Av_k = \lambda_1. \quad (8.5)$$

Důkaz: Uvažujme Schurův rozklad hermitovské matice ve tvaru $A = U\Lambda U^*$. Pak můžeme postupně psát

$$\begin{aligned} v_0 &= \sum_{i=1}^n \alpha_i u_i \\ A^k v_0 &= \sum_{i=1}^n \alpha_i \lambda_i^k u_i, \quad k \geq 1 \\ A^k v_0 &= \alpha_1 \lambda_1^k \left(u_1 + \sum_{i=2}^n \frac{\alpha_i}{\alpha_1} \left(\frac{\lambda_i}{\lambda_1} \right)^k u_i \right) = \alpha_1 \lambda_1^k (u_1 + y_k), \quad k \geq 1 \end{aligned}$$

Protože $\alpha_1 \equiv u_1^* v_0 \neq 0$ a dále také platí

$$v_1 = \frac{Av_0}{\|Av_0\|}, \quad v_2 = \frac{Av_1}{\|Av_1\|} = \frac{AAv_0 / \|Av_0\|}{\|AAv_0 / \|Av_0\|\|} = \frac{AAv_0}{\|AAv_0\|} = \frac{A^2 v_0}{\|A^2 v_0\|}, \quad (8.6)$$

a tedy obecně pro $k \geq 2$

$$v_k = \frac{Av_{k-1}}{\|Av_{k-1}\|} = \frac{AAv_{k-2} / \|Av_{k-2}\|}{\|AAv_{k-2} / \|Av_{k-2}\|\|} = \frac{AAv_{k-2}}{\|AAv_{k-2}\|} = \dots = \frac{A^k v_0}{\|A^k v_0\|}, \quad (8.7)$$

pak můžeme psát pro $k \geq 1$

$$v_k = \frac{\alpha_1 \lambda_1^k (u_1 + y_k)}{\|\alpha_1 \lambda_1^k (u_1 + y_k)\|} = \operatorname{sgn}(\alpha_1 \lambda_1^k) \frac{(u_1 + y_k)}{\|(u_1 + y_k)\|}, \quad (8.8)$$

kde $y_k \rightarrow 0$. Pak tedy platí

$$\lim_{k \rightarrow \infty} (u_1 + y_k) / \operatorname{sgn}(\alpha_1 \lambda_1^k) = u_1$$

a tedy také

$$\lim_{k \rightarrow \infty} v_k / \operatorname{sgn}(\alpha_1 \lambda_1^k) = u_1 / \|u_1\|, \quad (8.9)$$

což je první hledaný vztah. Dále platí

$$\lim_{k \rightarrow \infty} \mu_k \equiv \lim_{k \rightarrow \infty} v_k^* A v_k = \lambda_1,$$

jak plyne z definice Rayleighova kvocientu, který jsme ztotožnili s odhadem vlastního čísla. Poznamenejme, že rychlost konvergence výrazu $(u_1 + y_k) / \operatorname{sgn}(\alpha_1 \lambda_1^k)$ k vlastnímu vektoru u_1 je lineární a je dána podílem $|\lambda_2 / \lambda_1|$. ■

Počáteční vektor se často volí náhodně. Předpoklad nenulovosti α_1 z předchozí věty je tak splněna s velkou pravděpodobností. Věta se dá formulovat i za slabších předpokladů, například pro matice obecně diagonalizovatelné nebo ještě obecnější, kde se v důkaze použijí vlastnosti Jordanovy kanonické formy. Další zobecnění je možné získat obměnou vztahu mezi absolutními hodnotami vlastních čísel dané matice. Všechny tyto změny ale přispívají k hlavní linii důkazu především technicky.

8.1.2 Mocinná metoda a algoritmus pagerank

V této kapitole budeme pracovat s objekty v reálném oboru. Uvažujme systém n webových stránek. Zaveďme si následující ohodnocení důležitosti webové stránky P_i nazvané pagerank a označené jako $r(P_i)$.

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|}, \quad (8.10)$$

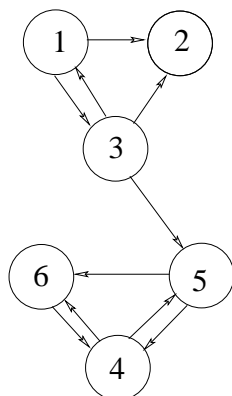
kde B_{P_i} je množina stránek, které na P_i ukazují a $|P_j|$ označuje celkový počet odkazů z P_j na jiné webové stránky. Tím, že tato ohodnocení jsou vzájemně závislá, Brin a Page navrhli iterační proceduru, jak je spočítat. Konkrétně, počítali $r_k(P_i)$ pro $i = 1, \dots, n$; $k = 1, \dots$, kde

$$r_{k+1}(P_i) = \sum_{P_j \in B_{P_i}} \frac{r_k(P_j)}{|P_j|}, \quad k = 1, \dots \quad (8.11)$$

s počátečními hodnotami $r_0(P_i) = 1/n$, kde P_1, \dots, P_n jsou uvažované webové stránky.

Uvažujme následující orientovaný graf, který ukazuje vzájemné odkazování šesti webových stránek P_1, \dots, P_6 .

Zvolíme-li na počátku všechna ohodnocení (pagerank) rovna $1/6$, pak postupně získáváme



Obrázek 8.1.1: Orientovaný graf, který ukazuje vzájemné odkazování šesti webových stránek.

iterace 0	iterace 1	iterace 2
1/6	1/18	1/36
1/6	5/36	1/18
1/6	1/12	1/36
1/6	1/4	17/72
1/6	5/36	11/72
1/6	1/6	14/72

Na základě velikosti ohodnocení soudíme na důležitost stránky. Zkusme si nyní formulovat problém maticově. Zaveďme matici, ve které budou převrácené hodnoty počtu odkazů na jiné webové stránky. Konkrétně, zaveďme pro náš příklad matici H následovně tak, že

$$H_{ij} = 1/|P_i|, \quad (8.12)$$

je-li odkaz z i na j .

$$H = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad (8.13)$$

a dále zaveďme vektor

$$\pi_k^T = \left(r_k(P_1) \quad r_k(P_2) \quad r_k(P_3) \quad r_k(P_4) \quad r_k(P_5) \quad r_k(P_6) \right). \quad (8.14)$$

Pak můžeme pro iteraci výše uvedeného algoritmu psát

$$\pi_{k+1} = H^T \pi_k. \quad (8.15)$$

No a to není nic jiného než mocninná metoda aplikovaná na matici H^T . Matice H^T je velmi blízká matici **stochastické** (matice s nezápornými prvky, jejíž sloupcové součty

jsou jedna. Mohou se v ní vyskytovat sloupce se samými nulami, které odpovídají vrcholům, ze kterých **nevedou žádné odkazy**. Takové matice se říká substochastická.

Věta 8.1.2 *Je-li matice H^T stochastická a nerozložitelná (H^T se nedá převést symetrickou permutací na blokově trojúhelníkový tvar, který má více než jeden diagonální blok s pozitivní dimenzí), existuje právě jeden vektor ohodnocení (pagerank). Je-li navíc aperiodická (neexistuje takové $k > 1$, že $H^k = H$), iterační metoda k němu konverguje nezávisle na startovacím vektoru.*

Tyto vlastnosti se získávají modifikacemi původní matice. Například, stochasticita se získá nahrazením nulových sloupců vektory s prvky $1/n$.

8.1.3 Krylovův prostor a Arnoldiho algoritmus

Budeme předpokládat, že matice $A \in \mathbb{C}^{n \times n}$ je čtvercová. Jedním z nejdůležitějších iteračních postupů na řešení částečného problému vlastních čísel je Arnoldiho algoritmus, který pro $k = 1, \dots$ konstruuje vektory $v_1, \dots, v_{k+1} \in \mathbb{C}^n$, matice $H_k \in \mathbb{C}^{k \times k}$ v horním Hessenbergově tvaru a skalár $h_{k+1,k}$ které splňují vztah

$$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T \quad (8.16)$$

a kde $V_k = (v_1, \dots, v_k)$ je báze k -tého Krylovova prostoru

$$K_k(A, b) = \text{span}\{b, Ab, \dots, A^{k-1}b\} \quad (8.17)$$

Algoritmus 8.1.2 Arnoldiho algoritmus

Input: $A = [a_{ij}] \in \mathbb{C}^{n \times n}$, $v \in \mathbb{C}^n$

Output: Matice $H_k \in \mathbb{C}^{k \times k}$ v horním Hessenbergově tvaru a ON báze V_k k -tého Krylovova prostoru takové, že platí $V_k^* A V_k = H_k$

1. $v_1 = v / \|v\|$
2. **for** $k = 1, 2, \dots$ **do**
3. $w = Av_k$
4. **for** $i = 1 : k$ **do**
5. $h_{ik} = v_i^* w$
6. $w = w - h_{ik} v_i$
7. **end** i
8. $h_{k+1,k} = \|w\|$
9. $v_{k+1} = w / h_{k+1,k}$
10. **end** k

Lemma 8.1.1 *Poddiagonální prvky matice H_k jsou reálné a kladné.*

Důkaz: Poddiagonální prvky matice H_k jsou normalizační koeficienty a musí tedy být reálné a kladné. ■

Poznámka 8.1.1 Připomeňme si, že

- $V_k V_k^*$ je ortogonální projektor na k -tý Krylovův prostor $K_k(A, v)$.
- $V_k V_k^* A$ je **ortogonální zúžení** operátoru reprezentovaného maticí A na $K_k(A, v)$ ve standardní bázi. To jest, například platí

$$x \in K_k(A, v) \implies V_k V_k^* A x \in K_k(A, v) \quad (8.18)$$

Pozorování 8.1.1 Existuje $d(A, v) \in \mathbb{N}$, $1 \leq d \leq n$ takové, že pro $d > 1$ platí

$$\begin{aligned} \text{span}\{v\} \equiv K_1(A, v) &\subsetneq \dots \subsetneq K_{d-1}(A, v) \subsetneq K_d(A, v) \\ &\equiv \{v, Av, \dots, A^{d-1}v\} \\ &\equiv \{v, Av, \dots, A^d v\} \\ &= K_{d+1}(A, v) = \dots = K_n(A, v), \end{aligned}$$

případně pro triviální případ $d = 1$ platí

$$\begin{aligned} \text{span}\{v\} \equiv K_1(A, v) &\equiv \{v, \dots, A^{d-1}v\} \\ &\equiv \{v, \dots, A^d v\} \\ &= K_{d+1}(A, v) = \dots = K_n(A, v). \end{aligned}$$

Tomuto $d(A, v)$ se říká **maximální dimenze** (stupeň) Krylovova prostoru generovaného uspořádanou dvojicí (A, v) .

Lemma 8.1.2 Necht' $A \in \mathbb{C}^{n \times n}$, $v \in \mathbb{C}^n$, $k < d(A, v)$. Necht' jsou dále matice $V_k \in \mathbb{C}^{n \times k}$, $H_k \in \mathbb{C}^{k \times k}$ matice z Arnoldiho algoritmu, $v_{k+1} \in \mathbb{C}^n$ a necht' $\{\mu, y\}$ je vlastní pár matice H^k . Pak $\{\mu, x\}$, $x = V_k y$ splňují

$$\|Ax - \mu x\| = h_{k+1,k} |e_k^T y|, \quad Ax - \mu x \perp K_k(A, v). \quad (8.19)$$

Důkaz: Aplikujeme-li obě strany následující rovnosti na vlastní vektor y matice H_k

$$AV_k y = V_k H_k y + h_{k+1,k} v_{k+1} e_k^T y \quad (8.20)$$

dostaneme

$$AV_k y = V_k H_k y + h_{k+1,k} v_{k+1} e_k^T y = \mu V_k y + h_{k+1,k} v_{k+1} e_k^T y. \quad (8.21)$$

Pro $x = V_k y$ pak plyne

$$Ax - \mu x = h_{k+1,k} v_{k+1} e_k^T y, \quad (8.22)$$

z čehož plynou obě dvě tvrzení. První z nich tím, že **aplikujeme na poslední vztah normu**, druhé z nich je důsledkem toho, že $Ax - \mu x$ je násobkem v_{k+1} a tudíž **kolmé** na vektory báze Krylovova prostoru V_k . ■

Speciální případ tvrzení je uveden v následující poznámce.

Poznámka 8.1.2 *Nechť $k = d \equiv d(A, v)$. Pak ve vytvářeném Krylovovu prostoru nelze najít vektor lineárně nezávislý na vektorech dosud zkonstruované báze V_k . Platí tedy*

$$AV_k = V_k H_k \quad (8.23)$$

Je-li $\{\mu, y\}$ je vlastní pár matice H_k , pak je $\{\mu, V_k y\}$ vlastní pár matice A neboť platí

$$AV_k y = V_k H_k y = \mu V_k y. \quad (8.24)$$

*Vektory y nazýváme **Ritzovy vektory** a čísla μ nazýváme **Ritzova čísla** příslušného problému vlastních čísel.*

8.1.4 Lanczosův algoritmus

V případě, že matice $A \in \mathbb{C}^{n \times n}$ je čtvercová a hermitovská, pak je speciálním případem Arnoldiho algoritmu **Lanczosův algoritmus**, který má mnoho zajímavých souvislostí v různých oblastech matematiky. Níže uvidíme výrazně silnější tvrzení o kvalitě dosažené aproximace problému vlastních čísel než v případě obecného Arnoldiho algoritmu.

Lemma 8.1.3 *Matice H_k v Lanczosově algoritmu je tridiagonální.*

Důkaz: Ze vztahu

$$AV_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T \quad (8.25)$$

snadno dostaneme přenásobením maticí V_k^* zleva a přeuspořádáním členů

$$H_k = V_k^* AV_k = V_k^* A^* V_k = (V_k^* AV_k)^* = H_k^*, \quad (8.26)$$

z čehož plyne tvrzení, protože H_k je v horním Hessenbergově tvaru. ■

Lemma 8.1.4 *Matice H_k v Lanczosově algoritmu je reálná a symetrická.*

Důkaz: To, že jsou diagonální prvky reálné plyne z faktu, že H_k je hermitovská. Poddiagonální prvky H_k a tím pádem i naddiagonální prvky H_k jsou reálné a kladné, protože to jsou **normalizační koeficienty** i v obecnější proceduře – Arnoldiho algoritmu. ■ Speciální značení, které budeme používat pro základní vztah v případě Lanczosova algoritmu je následující

$$AV_k = V_k T_k + \beta_{k+1} v_{k+1} e_k^T, \quad (8.27)$$

kde

$$T_k = \begin{pmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \ddots & \ddots & \\ & & \ddots & \ddots & \beta_k \\ & & & \beta_k & \alpha_k \end{pmatrix}. \quad (8.28)$$

Rozepsáním vztahu pro jednotlivé sloupce dostaneme

$$\beta_{j+1}v_{j+1} = Av_j - \alpha_jv_j - \beta_jv_{j-1}, \quad j = 1, \dots, k, \quad v_0 = 0, \quad \beta_1 = 0. \quad (8.29)$$

Předpokládáme přitom, že v_1 je normalizovaný startovací vektor, často zvolený náhodně. Postup pak můžeme zapsat ve formě algoritmu, kde podstatnou součástí postupu je normalizace vektorů ve V_k .

Algoritmus 8.1.3 Lanczosův algoritmus

Input: Hermitovská matice $A = [a_{ij}]$, počáteční vektor $v \in \mathbb{C}^n$

Output: Reálná symetrická tridiagonální matice T_k a ON báze k -tého Krylovova prostoru V_k takové, že platí $V_k^*AV_k = T_k$

1. $v_0 = 0, v_1 = v/\|v\|, \beta_1 = 0$
2. **for** $k = 1, 2, \dots$ **do**
3. $w = Av_k - \beta_kv_{k-1}$
4. $\alpha_k = v_k^*w$
5. $w = w - \alpha_kv_k$
6. $\beta_{k+1} = \|w\|$
7. $v_{k+1} = w/\beta_{k+1}$
8. **end** k

Následující lemma ukazuje, že Lanczosův algoritmus dává mnohem lepší záruky na výsledek než Arnoldiho algoritmus.

Lemma 8.1.5 *Nechť $A \in \mathbb{C}^{n \times n}$ s vlastními čísly $\lambda_1, \dots, \lambda_n$ je hermitovská matice. Nechť dále $v \in \mathbb{C}^n$ je nenulový počáteční vektor Lanczosova algoritmu a uvažujme $k < d(A, v)$. Nechť $\{\mu_i, y_i\}$ jsou příslušné vlastní páry matice T_k generované tímto algoritmem pro $i = 1, \dots, k$. Pak*

$$\min_{j=1, \dots, n} |\lambda_j - \mu_i| \leq \beta_{k+1} |e_k^T y_i|. \quad (8.30)$$

Důkaz: Z obecného tvrzení pro Arnoldiho algoritmus víme, že platí $\|Ax_i - \mu_i x_i\| = \beta_{k+1} |e_k^T y_i|$ pro $x_i = V_k y_i$. Dále zdůrazněme, že v tomto textu pokládáme vlastní vektory vždy normalizované, to jest s normou rovnou jedné. Nechť $A = U\Lambda U^*$ je Schurův rozklad s diagonální maticí Λ (diagonalizace normální matice A). Definujme $w = U^* x_i$. Pak máme

$$\begin{aligned} \|Ax_i - \mu_i x_i\| &= \|U\Lambda U^* x_i - \mu_i x_i\| = \|\Lambda U^* x_i - \mu_i U^* x_i\| = \\ &= \|\Lambda w - \mu_i w\| = \|\text{diag}(\lambda_1 - \mu_i, \dots, \lambda_n - \mu_i)w\| \geq \\ \min_{j=1, \dots, n} |\lambda_j - \mu_i| \|w\| &= \min_{j=1, \dots, n} |\lambda_j - \mu_i| \|U^* x_i\| = \min_{j=1, \dots, n} |\lambda_j - \mu_i| \|x_i\| \end{aligned}$$

Následně,

$$\min_{j=1, \dots, n} |\lambda_j - \mu_i| \leq \frac{\|Ax_i - \mu_i x_i\|}{\|x_i\|} = \beta_{k+1} \frac{|e_k^T y_i|}{\|x_i\|} \quad (8.31)$$

Protože jsou vektory y_i normalizovány, tak platí

$$\|x_i\| = \|V_k y_i\| = \|y_i\| = 1 \quad (8.32)$$

a tvrzení je tak dokázáno. ■

8.1.5 Porovnání Lanczosova a Arnoldiho algoritmu

Lanczosův algoritmus má nižší počet operací než Arnoldiho algoritmus, protože v každém kroku ortogonalizujeme jen vůči jednomu vektoru. Nejnáročnější operace v každém kroku je tak pouze násobení maticí A .

Na druhé straně, teoretické záruky Lanczosova algoritmu jsou výrazně větší než u Arnoldiho algoritmu, jak plyne z porovnání příslušných uvedených tvrzení.

Kapitola 9

Přímé metody a hermitovské indefinitní matice

Důvodem, proč přímé metody pro tyto matice je vhodné diskutovat samostatně, je obvykle požadovaná **symetrie rozkladu**. Abychom si demonstrovali složitost problému, uvažujme následující symetrický rozklad. Rozkládaná matice je zřejmě indefinitní, protože má vlastní čísla $\epsilon - 1$ a $\epsilon + 1$.

$$A = \begin{pmatrix} \epsilon & 1 \\ 1 & \epsilon \end{pmatrix} = \begin{pmatrix} 1 & \\ & 1/\epsilon \end{pmatrix} \begin{pmatrix} \epsilon & \\ & \epsilon - 1/\epsilon \end{pmatrix} \begin{pmatrix} 1 & 1/\epsilon \\ & 1 \end{pmatrix} \quad (9.1)$$

V tomto rozkladu vidíme zřejmou nestabilitu danou růstem velikosti (absolutních hodnot) prvků. Jednostrannou řádkovou permutaci z GEPP nemůžeme pro kontrolu růstu prvků uvažovat, nechceme-li ztratit symetrii, která může být podstatná pro aplikace, ze kterých matice pochází. Permutovat tedy musíme **symetricky** podle schématu

$$A \rightarrow PAP^T,$$

jak jsme o tom již hovořili výše. Bude-li místo veličiny ϵ nula, LDL^T rozklad s diagonální maticí D nebude existovat. Dá se ukázat, že v takovém případě se standardní rozklad může trochu zobecnit a použít rozklad

$$LDL^T, \quad (9.2)$$

kde matice D je **blokově diagonální** s bloky o velikosti 1×1 a 2×2 . Pro pivotaci pak existují různá **pivotační schémata**, která se nazývají podle těch, co je navrhli, například Bunchův-Kaufmannův algoritmus a Bunchův-Parlettův algoritmus. Motivací pro užitečnost bloků velikosti 2×2 je fakt, že v případě, že velikosti jeho diagonálních prvků jsou malé a velikosti jeho mimodiagonálních prvků jsou velké (ve svých absolutních hodnotách), pak blokový rozklad může přispět k malému růstu nenulových prvků v průběhu rozkladu. Konkrétně, inverze matice A z (9.1) je rovna

$$\frac{1}{\epsilon^2 - 1} \begin{pmatrix} \epsilon & -1 \\ -1 & \epsilon \end{pmatrix} \quad (9.3)$$

a násobení touto maticí, což je operace rozkladu analogická dělení diagonálním prvkem, nevede k rychlému růstu prvků. Růst v rozkladu hermitovských indefinitních matic se

dá stejně jako v případě LU rozkladu charakterizovat růstovým faktorem. Postupem ztedy ískáme rozklad s obecně symetrickou permutací aplikovanou zleva na řádky a analogicky na sloupce, kde matice D je blokově diagonální. Řešení je pak počítáno v několika krocích a odvozeno podobně, jako v předcházejících variantách Gaussovy eliminace.

$$PAP^T = LDL^*$$

$$Lz = Pb, Dw = z, L^*y = v, x = Py$$

Příklady algoritmů pro taková schémata nyní uvedeme ve formě algoritmů.

Algoritmus 9.0.1 Krok úplné pivotace podle Bunche a Parletta (1971)

1. Polož $\alpha = (1 + \sqrt{17})/8 \approx 0.64$
2. Nalezni a_{kk} : diagonální prvek maximální velikosti
3. Nalezni a_{ij} : mimodiagonální prvek maximální velikosti ($i < j$)
4. **if** $|a_{kk}| \geq \alpha|a_{ij}|$ **then**
5. použij a_{kk} jako 1×1 pivot (**hotovo** pro $a_{kk} = 0$)
6. **else**
7. použij $\begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix}$ jako blokový 2×2 pivot
8. **end if**

Algoritmus, který provádí úplnou diagonální pivotaci je velmi stabilní, ale jeho implementace na rozklad řídkých indefinitních matic může být velmi drahá. Problém efektivního postupu v případě řídkých matic řeší do značné míry pivotace podle Bunche a Kaufmannové, jejíž jeden krok je uveden v následujícím algoritmu.

Algoritmus 9.0.2 Krok úplné pivotace podle Bunche a Kaufmannové (1977)

1. $\alpha = (1 + \sqrt{17})/8 \approx 0.64$
2. $i = 1$ (možná varianta, kde i splnuje $|a_{ii}| \geq \alpha|a_{kk}|$ mezi všemi k)
3. Nalezni $j \neq i$ takové, že $a_{ji} = \max\{|a_{ki}|, k \neq i\} =: \lambda$
4. **if** $|a_{ii}| \geq \alpha\lambda$ **then**
5. použij a_{ii} jako 1×1 pivot
6. **else**
7. $\sigma = \max\{|a_{kj}|, k \neq j\}$
8. **if** $|a_{ii}|\sigma \geq \alpha\lambda^2$ **then**
9. use a_{ii} jako 1×1 pivot
10. **else if** $|a_{jj}| \geq \alpha\sigma$ **then**
11. použij a_{jj} as a 1×1 pivot
12. **else**
13. použij $\begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix}$ jako blokový 2×2 pivot

14. **end if**
15. **end if**

Schématicky si můžeme znázornit, které prvky jsou v jednom kroku Algoritmu 9.0.2 testovány, následujícím způsobem.

$$\begin{pmatrix} d & . & . & \lambda & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \lambda & . & . & c & . & \sigma & . \\ . & . & . & . & . & . & . \\ . & . & . & \sigma & . & . & . \\ . & . & . & . & . & . & . \end{pmatrix}$$

Růstový faktor je v tomto algoritmu omezen hodnotou $(2.57)^{n-1}$, což je o něco víc než v případě standardní částečné pivotace LU rozkladu.

Kapitola 10

Iterační metody pro řešení soustav rovnic

Uvažujme řešení soustavy lineárních rovnic

$$Ax = b \tag{10.1}$$

s regulární maticí soustavy $A \in \mathbb{C}^{n \times n}$, pravou stranou $b \in \mathbb{C}^n$ a hledaným řešením $x \in \mathbb{C}^n$. Zatímco rozklady, které jsme v textu uvedli, vedou na skupinu metod **přímých**, v této kapitole se budeme věnovat některým metodám **iteračním**.

10.1 Klasické stacionární iterační metody

10.1.1 Základní fakta

Stacionární iterační metody jsou založeny na statickém štěpení matice A následujícího typu

$$A = M - N, \tag{10.2}$$

kde předpokládáme, že M je regulární a snadno invertovatelná. Pak můžeme psát následující posloupnost úprav.

$$\begin{aligned} Ax &= b \\ (M - N)x &= b \\ (M - (M - A))x &= b \\ Mx &= b + (M - A)x \\ x &= x + M^{-1}(b - Ax) \end{aligned}$$

Budeme-li poslední uvedenou rovnost (případně předposlední uvedenou rovnost) považovat za **definiční transformaci**, kde aproximaci řešení na pravé straně přiřadíme novou

aproximaci na levé straně, pak máme celý proces konstrukce posloupnosti aproximací $x_0, x_1, \dots, x_k, \dots$ shrnutý následovně

$$x_{k+1} = x_k + M^{-1}(b - Ax_k) = (I - M^{-1}A)x_k + M^{-1}b; \quad k = 0, \dots$$

Matici $I - M^{-1}A$ budeme nazývat **iterační matice**.

10.1.2 Chyba a konvergence stacionárních iteračních metod

Uvažujme zavedenou posloupnost iterací a přesné řešení soustavy x . Pak můžeme psát pro chybu v kroku $k + 1$ následující vztah

$$x - x_{k+1} = (I - M^{-1}A)(x - x_k) = (I - M^{-1}A)^{k+1}(x - x_0). \quad (10.3)$$

Zavedeme-li výraz pro chybu řešení $e_k = x - x_k$ pak máme

$$e_{k+1} = (I - M^{-1}A)^{k+1}e_0. \quad (10.4)$$

Máme tedy vztah pro změnu chyby, ale nikoli chybu samotnou. Přejdem k normě získáme

$$\frac{\|e_{k+1}\|}{\|e_0\|} = \frac{\|x - x_{k+1}\|}{\|x - x_0\|} \leq \|(I - M^{-1}A)^{k+1}\| \leq \|(I - M^{-1}A)\|^{k+1}. \quad (10.5)$$

Pro posuzování konvergence těchto iteračních metod je důležitý následující vztah.

Věta 10.1.1 *Nutnou a postačující podmínkou, aby iterační posloupnost definovaná v (10.3) konvergovala pro libovolnou volbu počátečního přiblížení je vztah*

$$\rho(I - M^{-1}A) < 1, \quad (10.6)$$

kde $\rho(I - M^{-1}A)$ nazýváme *poloměr konvergence matice $I - M^{-1}A$ a je roven maximální absolutní hodnotě nějakého vlastního čísla matice $I - M^{-1}A$. Věta se dá snadno dokázat pomocí podobnostní transformace iterační matice na blokově diagonální matici složenou z Jordanových bloků a rozepsáním mocnin těchto Jordanových bloků.*

Poznámka 10.1.1 *Poznamenejme, že poloměr konvergence iterační matice $\rho(I - M^{-1}A)$ může být také zapsán ve tvaru*

$$\rho(I - M^{-1}A) = \lim_{k \rightarrow \infty} \|((I - M^{-1}A)^k)\|^{1/k} < 1 \quad (10.7)$$

Následující postačující podmínka konvergence je často snadno ověřitelná.

Věta 10.1.2 *Postačující podmínkou pro to, aby iterační posloupnost definovaná v (10.3) konvergovala pro libovolnou volbu počátečního přiblížení je platnost vztahu*

$$\|I - M^{-1}A\|_P < 1 \quad (10.8)$$

v některé maticové (tedy, sub-multiplikativní) normě $\|\cdot\|_P$.

Důkaz: Ukážeme, že absolutní hodnota libovolného vlastního čísla dané matice $C \in \mathbb{C}^{n \times n}$ je rovna nanejvýš libovolné maticové normě $\|\cdot\|_P$ matice C . Zvolme $\epsilon > 0$ a sestrojme matici

$$B = \frac{1}{\|C\|_P + \epsilon} C. \quad (10.9)$$

Její norma zřejmě splňuje

$$\|B\|_P = \frac{\|C\|_P}{\|C\|_P + \epsilon} < 1. \quad (10.10)$$

To ale znamená, že

$$\lim_{k \rightarrow \infty} B^k = 0, \quad (10.11)$$

neboť ze sub-multiplikativity maticových norem plyne $\|B^k\|_P \leq \|B\|_P^k$. S použitím Jordanova normálního tvaru se dá ukázat, že všechna vlastní čísla matice B jsou v absolutní hodnotě menší než 1. Nechť λ je vlastní číslo matice C a x příslušný vlastní vektor. Pak platí

$$Cx = \lambda x \Rightarrow Bx = \frac{\lambda}{\|C\|_P + \epsilon} x. \quad (10.12)$$

Proto platí

$$\frac{|\lambda|}{\|C\|_P + \epsilon} < 1 \quad (10.13)$$

a odsud $|\lambda| < \|C\|_P + \epsilon$. Protože ϵ bylo zvoleno libovolně, plyne z toho i

$$|\lambda| \leq \|C\|_P. \quad (10.14)$$

Aplikací výsledku na matici $C = I - M^{-1}A$ získáme tvrzení. ■

Poznámka 10.1.2 *Může se stát, že platí $\rho(I - M^{-1}A) < 1 < \|(I - M^{-1}A)\|_P$ pro nějakou maticovou normu $\|\cdot\|_P$. Pak může být křivka konvergence lokálně konkávní (po nějakou dobu se chyba zvětšuje).*

10.1.3 Jakobiho metoda

Uvažujme matici A štěpenou následujícím způsobem (konvenčně) na diagonální část D , striktní dolní trojúhelníkovou část L a striktní horní trojúhelníkovou část U .

$$A = D - L - U$$

Jakobiho metoda je založena na štěpení

$$A = M - N \equiv D - (L + U), \quad (10.15)$$

to jest na iteracích

$$Dx_{k+1} = Lx_k + Ux_k + b \Rightarrow M \equiv D, N \equiv L + U. \quad (10.16)$$

Iterační matice Jakobiho metody je $D^{-1}(L + U)$ neboť

$$I - M^{-1}A = I - M^{-1}(D - L - U) = I - D^{-1}D + D^{-1}(L + U). \quad (10.17)$$

To můžeme zapsat pro $i = 1, \dots, n$ po složkách následovně

$$\begin{aligned} \xi_i^{(k+1)} &= \frac{1}{a_{ii}} \left(- \sum_{j=1, j \neq i}^n a_{ij} \xi_j^{(k)} + b_i \right) \\ \text{pro } x^{(k)} &= (\xi_1^{(k)}, \dots, \xi_n^{(k)})^T \end{aligned}$$

Věta 10.1.3 *Nutná a postačující podmínka konvergence Jakobiho metody pro libovolné počáteční přiblížení x_0 je tedy samozřejmě*

$$\rho(D^{-1}(L + U)) < 1 \quad (10.18)$$

Pro konvergenci postačuje splnění $\|D^{-1}(L + U)\|$ v některé maticové normě.

Definujme si nejprve **ostře diagonálně dominantní** matici.

Definice 10.1.1 *Řekneme, že matice $A \in \mathbb{C}^{n \times n}$ je ostře diagonálně dominantní, platí-li pro prvky na jejím řádku vztah*

$$\sum_{j=1, j \neq i}^n |a_{ij}| < |a_{ii}|, \quad i = 1, \dots, n \quad (10.19)$$

Věta 10.1.4 *Je-li matice $A \in \mathbb{C}^{n \times n}$ ostře diagonálně dominantní, pak Jakobiho metoda konverguje.*

Důkaz: Pro ostře diagonálně dominantní matici máme podle definice

$$\sum_{j=1, j \neq i}^n |a_{ij}| < |a_{ii}|, \quad i = 1, \dots, n$$

tedy

$$\sum_{j=1, j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|} < 1, \quad i = 1, \dots, n \quad (10.20)$$

a tedy

$$\max_{i=1, \dots, n} \sum_{j=1, j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|} < 1, \quad (10.21)$$

což je postačující podmínka pro konvergenci v maximové maticové normě. ■

Věta 10.1.5 *Je-li matice $A \in \mathbb{C}^{n \times n}$ hermitovská s diagonálou D a kladnými diagonálními prvky, pak Jakobiho metoda konverguje **právě tehdy**, jsou-li A a $2D - A$ pozitivně definitní.*

Důkaz: Nejprve si všimněme, že matice $2D - A$ se liší od matice A jen změnou znamének u mimodiagonálních prvků. Nechť $D_1 = \text{diag}(\sqrt{a_{11}}, \dots, \sqrt{a_{nn}})$, to jest $D = D_1^2$. Pak dostaneme podobnostní transformaci

$$I - D^{-1}A = D^{-1}(L + U) = D_1^{-1}[D_1^{-1}(L + U)D_1^{-1}]D_1. \quad (10.22)$$

Jakobiho metoda tedy konverguje právě tehdy, je-li $\rho(D_1^{-1}(L + U)D_1^{-1}) < 1$. Matice $D_1^{-1}(L + U)D_1^{-1} \equiv I - D_1^{-1}AD_1^{-1}$ je hermitovská.

Protože hermitovská matice má všechna vlastní čísla reálná, pak Jakobiho metoda konverguje právě tehdy, jsou-li tato vlastní čísla z otevřeného intervalu $(-1, 1)$. To platí právě tehdy, mají-li obě matice $M_1 = I + D_1^{-1}(L + U)D_1^{-1}$ a $M_2 = I - D_1^{-1}(L + U)D_1^{-1}$ **zároveň** všechna vlastní čísla kladná. Protože jsou obě tyto matice také hermitovské, je to právě tehdy, když jsou pozitivně definitní. Platí ovšem, že

$$M_1 = I + D_1^{-1}(L + U)D_1^{-1} = D_1^{-1}(D_1^2 + L + U)D_1^{-1} = D_1^{-1}(2D - A)D_1^{-1} \quad (10.23)$$

a také

$$M_2 = I - D_1^{-1}(L + U)D_1^{-1} = D_1^{-1}(D_1^2 - L - U)D_1^{-1} = D_1^{-1}AD_1^{-1}. \quad (10.24)$$

Tvrzení je tak dokázáno, protože je-li obecně matice B pozitivně definitní a T regulární, pak je T^*BT také pozitivně definitní. ■

10.1.4 Gauss-Seidelova metoda

Uvažujme stejně jako výše matici $A \in \mathbb{C}^{n \times n}$ štěpenou následujícím způsobem (konvenčně) na diagonální část D , striktní dolní trojúhelníkovou část L a striktní horní trojúhelníkovou část U .

$$A = D - L - U.$$

Gauss-Seidelova metoda je založena na štěpení

$$A = M - N \equiv (D - L) - U, \quad (10.25)$$

to jest na iteracích

$$(D - L)x_{k+1} = Ux_k + b \Rightarrow M \equiv D - L, N \equiv U. \quad (10.26)$$

Iterační matice Gauss-Seidelovy metody je $(D - L)^{-1}U$ neboť

$$I - M^{-1}A = I - M^{-1}(D - L - U) = I - (D - L)^{-1}(D - L) + (D - L)^{-1}U = (D - L)^{-1}U. \quad (10.27)$$

To můžeme zapsat pro $i = 1, \dots, n$ po složkách následovně

$$\xi_i^{(k+1)} = \frac{1}{a_{ii}} \left(- \sum_{j=1}^{i-1} a_{ij} \xi_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} \xi_j^{(k)} + b_i \right)$$

pro $x^{(k)} = (\xi_1^{(k)}, \dots, \xi_n^{(k)})^T$

První věta je následující analogie tvrzení pro Jakobiho metodu.

Věta 10.1.6 *Je-li matice $A \in \mathbb{C}^{n \times n}$ ostře diagonálně dominantní, Gauss-Seidelova metoda konverguje.*

Důkaz: Pro konvergenci stačí, aby platilo $\|(D - L)^{-1}U\|_\infty < 1$. Tato maticová norma je generována vektorovou maximovou normou a proto musí existovat vektor u takový, že platí

$$\|(D - L)^{-1}U\|_\infty = \|(D - L)^{-1}Uu\|_\infty, \|u\|_\infty = 1. \quad (10.28)$$

Označme $v = (D - L)^{-1}Uu$. Pro složky vektoru v , kde $\|v\|_\infty = \|(D - L)^{-1}Uu\|_\infty$ platí $|v_i| \leq \|v\|_\infty$. Nechť $|v_s| = \|v\|_\infty$. Napišme nyní vztah mezi vektory u a v

$$(D - L)v = Uu.$$

to jest taky pro s -tý řádek

$$a_{s1}v_1 + \dots + a_{ss}v_s = -a_{s,s+1}u_{s+1} - \dots - a_{sn}u_n.$$

Odtud

$$v_s = -\sum_{i=1}^{s-1} \frac{a_{si}}{a_{ss}} v_i - \sum_{i=s+1}^n \frac{a_{si}}{a_{ss}} u_i$$

Označme

$$a = \sum_{i=1}^{s-1} \left| \frac{a_{si}}{a_{ss}} \right|, \quad b = \sum_{i=s+1}^n \left| \frac{a_{si}}{a_{ss}} \right|.$$

Pak $a + b < 1$, neboť matice A je ostře diagonálně dominantní. Dále

$$\|v\|_\infty = |v_s| \leq a\|v\|_\infty + b\|u\|_\infty = a\|v\|_\infty + b$$

a z toho plyne

$$\|(D - L)^{-1}U\|_\infty = \|v\|_\infty \leq \frac{b}{1 - a} < 1, \quad (10.29)$$

jak jsme měli dokázat. ■

Věta 10.1.7 *Je-li matice $A \in \mathbb{C}^{n \times n}$ hermitovská a pozitivně definitní, Gauss-Seidelova metoda konverguje.*

Důkaz: Uvažujme matici $A = D - L - U$, $L = U^*$ a ukažme, že $\rho(D - U^*)^{-1}U < 1$. Nechť λ je vlastní číslo matice $(D - U^*)^{-1}U$ a u vlastní vektor odpovídající tomu vlastnímu číslu. Pak platí mimo jiné následující vztahy

$$(D - U^*)^{-1}Uu = \lambda u \Rightarrow Uu = \lambda Du - \lambda U^*u = \lambda Au + \lambda Uu.$$

Pro skalární součin pak dostaneme

$$(Uu, u) = \lambda(Au, u) + \lambda(Uu, u).$$

Označme $(Au, u) = p$. Z pozitivní definitnosti matice A víme, že $p > 0$. Uvažujme dále označení $(Uu, u) = a + ib$, kde p, a, b jsou reálná čísla. Pak můžeme psát

$$a + ib = \lambda p + \lambda(a + ib) \Rightarrow \lambda = \frac{a + ib}{p + a + ib}.$$

Dále uvidíme, že $p + a$ je nenulové a zlomek je tedy korektně definován. Vynásobením tohoto zlomku číslem $\bar{\lambda}$ získáme

$$|\lambda|^2 = \frac{a^2 + b^2}{(p + a)^2 + b^2}.$$

Předcházející vztah také můžeme vidět z následujícího odvození

$$\left| \frac{a + ib}{p + a + ib} \right|^2 = \left| \frac{\rho_1 e^{i(\phi_1 - \phi_2)}}{\rho_2} \right|^2 = \left| \frac{\rho_1}{\rho_2} \right|^2 = \frac{|\rho_1|^2}{|\rho_2|^2} = \frac{|a + ib|^2}{|p + a + ib|^2} = \frac{a^2 + b^2}{(p + a)^2 + b^2}.$$

Neboť

$$(U^*u, u) = (u, Uu) = \overline{(Uu, u)} = a - ib,$$

pak

$$p = (Du, u) - (Uu, u) - (U^*u, u) = (Du, u) - (a + ib + a - ib) = (Du, u) - 2a.$$

(Du, u) je podle předpokladu kladné, neboť pozitivně definitní matice A musí mít kladné diagonální prvky a také

$$(Du, u) = \sum_{i=1}^n a_{ii} u_i^* u_i = \sum_{i=1}^n a_{ii} |u_i|^2.$$

Dále

$$(p + a)^2 = p(p + 2a) + a^2 = p(Du, u) + a^2.$$

Všimněme si, že $p(Du, u)$ musí být kladné, neboť je součinem kladných čísel. Dostáváme tedy

$$|\lambda|^2 = \frac{a^2 + b^2}{p(Du, u) + a^2 + b^2}$$

a neboť jsme neměli omezení na výběr vlastního čísla, tvrzení věty je dokázáno. ■

Poznámka 10.1.3 Všimněme si, že Gauss-Seidelova metoda je stejně časově náročná jako Jakobiho metoda, ale má menší paměťové nároky než Jakobiho metoda, protože vektor nové iterace může přepisovat vektor předcházející iterace.

10.1.5 SOR (successive over-relaxation) metoda

Tato metoda je založena na následujícím parametrizovaném vztahu.

$$\begin{aligned} A &= (D + \omega L) - (\omega - 1)D - \omega U \\ 0 &< \omega < 2 \end{aligned}$$

Iterační matice SOR metody je $(D + \omega L)^{-1}((\omega - 1)D + \omega U)$. Po složkách pak můžeme psát

$$\xi_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \omega \left[-\frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} \xi_j^{(k+1)} \right) - \frac{1}{a_{ii}} \left(\sum_{j=i+1}^n a_{ij} \xi_j^{(k)} \right) + \frac{1}{a_{ii}} b_i \right]$$

pro $x^{(k)} = (\xi_1^{(k)}, \dots, \xi_n^{(k)})^T$

Literatura

- [1] L. Barto and J. Tůma. *Lineární algebra.* skriptum,
(pro jistotu zkopírované, aby se nám nezměnily odkazy:)
http://www.karlin.mff.cuni.cz/~mirektuma/znm/Barto_Tuma_skripta_la5.pdf, 20xx.