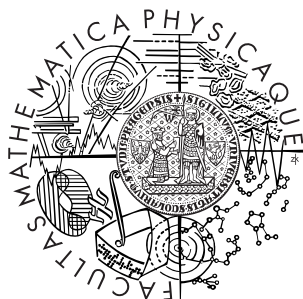


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Marie Michenková

Metoda úplných nejmenších čtverců s omezením hodnoti

Katedra numerické matematiky

Vedoucí bakalářské práce: RNDr. Iveta Hnětynková, Ph.D.

Studijní program: Matematika, Obecná matematika

2009

Na tomto místě bych chtěla poděkovat vedoucí bakalářské práce RNDr. Ivetě Hnětyňkové, Ph.D., za cenné návrhy a připomínky a rovněž za zapůjčení doplňkové literatury.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 6. 8. 2009

Marie Michenková

Obsah

Úvod	6
Značení	8
1 Lineární aproximační problémy	9
2 Metody nejmenších čtverců	11
2.1 Metoda nejmenších čtverců	11
2.2 Metoda úplných nejmenších čtverců	12
2.2.1 Existence a jednoznačnost řešení	13
2.2.2 Negerické řešení	15
3 Metody nejmenších čtverců s omezením hodnoti	17
3.1 Ill-posed problémy	17
3.2 Truncated SVD	19
3.3 Metoda úplných nejmenších čtverců s omezením hodnoti . .	20
4 Vlastnosti metody úplných nejmenších čtverců s omezením hodnoti	23
4.1 Regularizace pomocí T-TLS a T-SVD	23
4.2 Souvislost T-TLS a T-SVD	25
4.3 Vliv perturbací na řešení metodou T-TLS	28
5 Numerické experimenty	32
Závěr	38
Literatura	39

Název práce: Metoda úplných nejmenších čtverců s omezením hodnosti

Autor: Marie Michenková

Katedra (ústav): Katedra numerické matematiky

Vedoucí bakalářské práce: RNDr. Iveta Hnětynková, Ph.D.

e-mail vedoucího: hnetynkova@cs.cas.cz

Abstrakt: V celé řadě aplikací se setkáváme s potřebou řešit lineární aproximační úlohy $Ax \approx b$, kde matice A i vektor b jsou zatíženy chybami. Systavy často mívají neúplnou numerickou hodnost a klasická metoda úplných nejmenších čtverců mnohdy neposkytuje použitelné výsledky. Jednou z možností, jak uvedené úlohy v praxi řešit, je použití metody úplných nejmenších čtverců s omezením hodnosti, kdy původní systém aproximujeme dobře podmíněným systémem nižší hodnosti. V předložené práci se zabýváme metodou úplných nejmenších čtverců s omezením hodnosti v kontextu dalších metod určených k řešení lineárních aproximačních problémů. Zahrnutý jsou rovněž poznatky o vlastnostech metody, které jsou demonstrovány na numerických experimentech.

Klíčová slova: lineární algebraické rovnice, metody nejmenších čtverců, numerická hodnost, regularizace

Title: Truncated total least squares method
Author: Marie Michenková
Department: Department of Numerical Mathematics
Supervisor: RNDr. Iveta Hnětynková, Ph.D.
Supervisor's e-mail address: hnetynkova@cs.cas.cz

In a wide range of applications, we are facing the necessity to solve linear approximation problems $Ax \approx b$, where errors are confined both to A and b . Systems are often numerically ranked-deficient and classical total least squares method usually produces inutile results. One of the possibilities how to practically solve presented problems is utilization of truncated total least squares method, where the original system is approximated by a well-conditioned system of lower rank. In the presented work we study truncated total least squares method in the context of other methods designated for solving linear approximation problems. Properties of the method are further demonstrated on numerical experiments.

Keywords: linear algebraic equations, least squares methods, numerical rank, regularization

Úvod

Řešení systému lineárních algebraických rovnic

$$Ax = b$$

patří k jednomu ze základních problémů numerické matematiky i matematiky vůbec. V numerické matematice získáváme tyto soustavy z celé řady aplikací v oblastech jakými jsou například počítačová tomografie [10], radarové a sonarové zobrazování, zpracování obrazu, seizmologie, astronomický výzkum nebo statistické aplikace. Matice vzniklých systémů zpravidla bývají silně obdélníkové, jsou zatíženy chybami a často mívají neúplnou hodnotu. Existuje řada numerických metod, jimiž lze tyto soustavy řešit. V případě systémů s neúplnou hodnotou však některé klasické přístupy, jako metoda nejmenších čtverců [1] či metoda úplných nejmenších čtverců [13], selhávají a je nutné problémy jistým způsobem regularizovat. V práci se zaměříme na jednu z regularizačních metod a to na metodu *úplných nejmenších čtverců s omezením hodnoty* [3], [4], určenou k řešení soustav, kde matice A i vektor b obsahují chyby. Shrňme vlastnosti metody včetně vlastností regularizačních [5], [14], a popíšeme její souvislost s další regularizační metodou - metodou nejmenších čtverců s omezením hodnoty [13]. Dále se zaměříme na vliv perturbací na řešení diskutovanou metodou [3], [15], [16]. Budeme uvažovat pouze soustavy reálné, ale velkou část následující analýzy je možné rozšířit i na soustavy komplexní.

Práce je uspořádána následujícím způsobem. V kapitole 1 popisujeme blíže problémy, jejichž řešením se v této práci budeme zabývat. V kapitole 2 se věnujeme metodám nejmenších čtverců, přičemž se zaměřujeme na otázku existence a konstrukce řešení. Kapitola 3 je věnována metodě (úplných) nejmenších čtverců s omezením hodnoty. Kapitola 4 obsahuje stručný přehled konkrétních vlastností metody úplných nejmenších čtverců s omezením hodnoty v závislosti na vlastnostech soustavy. V kapitole 5 uvádíme

numerické experimenty, na kterých demonstrujeme některé teoretické poznatky obsažené v kapitole 4.

Značení

\mathbb{R}	množina reálných čísel
$A = (a_{i,j})_{i,j=1}^{m,n}$	matice $m \times n$
a_i, a_i^k	i -tý sloupec matice A , k -tá složka i -tého vektoru matice A
e_i	i -tý sloupec jednotkové matice
A^{-1}	inverze matice A
A^\dagger	Moore-Penroseova pseudoinverze matice A
A^T	matice transponovaná k matici A
$\text{diag}(\alpha, \beta, \dots)$	čtvercová matice s α, β, \dots na diagonále
$\ x\ $	euklidovská norma vektoru x , $\ x\ = \sqrt{x^T x}$
$\ A\ $	spektrální norma matice A , $\ A\ = \sup_{x \neq 0} \frac{\ Ax\ }{\ x\ }$
$\ \ \ _F$	Frobeniova norma mmatice A , $\ A\ _F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2}$
$\mathcal{R}(A)$	lineární obal sloupců matice A
$\mathcal{N}(A)$	jádro matice A
P_V	ortogonální projekce na podprostor V
$\text{dist}(V, W)$	$\ P_V - P_W\ $

Kapitola 1

Lineární aproximační problémy

Uvažujme soustavu

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

kde A je matice příslušného modelu a b vektor pozorování. Budeme navíc předpokládat, že matice A i vektor b jsou nenulové. Pokud $b \in \mathcal{R}(A)$, pak existuje přesné řešení x (obecně nejednoznačné). Takový systém se nazývá kompatibilní. V opačném případě, tedy v případě kdy $b \notin \mathcal{R}(A)$ (systém nazveme nekompatibilní), řešení neexistuje.

Pokud je soustava (1.2) zatížena chybami, píšeme

$$Ax \approx b. \quad (1.2)$$

Ve statistice jsou tyto modely známy jako *errors-in-variables models*. Cílem bude tedy najít nějakou vhodnou aproximaci přesného řešení. Aby něco takového mělo vůbec smysl, je zřejmě nutné vyloučit případ, kdy $b \perp \mathcal{R}(A)$ (ekvivalentně $A^T b = 0$), neboť v tomto případě nelze b aproximovat sloupci matice A (a tedy nejbližší aproximace přesného řešení x je nulový vektor). Pro soustavy vzniklé nějakým měřením, např. v počítačové tomografii, obvykle platí $m \geq n$ (často $m \gg n$). Pro jednoduchost budeme nadále uvažovat jen tyto soustavy, většinu úvah lze v tomto směru ale jednoduše zobecnit i na případ $m < n$.

Pokud známe model (tedy matici A) a chyby jsou obsaženy pouze ve vektoru pravé strany b , pak můžeme využít nejjednodušší metodu a to metodu nejmenších čtverců (viz kapitola 2.1), ve statistice známou jako *lineární regrese*. Často však matici A získáme také jako výsledek nějakého (nepřesného) měření a je tedy také zatížena chybami. V tomto případě lze jako

nejjednodušší metodu uplatnit metodu úplných nejmenších čtverců (viz kapitola 2.2), ve statistice *ortogonální regrese*. V případech, kdy je matice A špatně podmíněná, neposkytuje tento přístup dobré výsledky a prostor zde dostává právě metoda úplných nejmenších čtverců s omezením hodnosti (viz kapitola 3).

Singulární rozklad

Singulární rozklad (SVD) má zásadní teoretický i praktický význam pro řešení soustavy (1.2).

Nechť A má singulární rozklad

$$A = U\Sigma V^T = \sum_{i=1}^n u_i \sigma_i v_i^T, \quad (1.3)$$

$$U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m}, \quad V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n},$$

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

kde U a V jsou unitární, u_i , v_i jsou levé resp. pravé singulární vektory a

$$\sigma_1 \geq \dots \geq \sigma_n \geq 0$$

jsou singulární čísla matice A . Pro konstrukci tohoto rozkladu viz např. [5]. Označme $r \equiv \text{rank}(A)$ počet nenulových singulárních čísel σ_i matice A .

Obdobně nyní zavedme SVD matice $[b, A]$

$$[b, A] = \tilde{U} \tilde{\Sigma} \tilde{V}^T = \sum_{i=1}^{n+1} \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T. \quad (1.4)$$

Pro singulární čísla matic A a $[b, A]$ platí následující důležitá věta o prokládání.

Věta 1 (Interlacing theorem): Singulární čísla matic A a $[b, A]$ splňují

$$\tilde{\sigma}_1 \geq \sigma_1 \geq \tilde{\sigma}_2 \geq \dots \geq \sigma_n \geq \tilde{\sigma}_{n+1} \geq 0.$$

Důkaz lze nalézt např v [5].

Kapitola 2

Metody nejmenších čtverců

Abychom mohli lépe pochopit podstatu metody úplných nejmenších čtverců s omezením hodnoty, je třeba ji uvést do souvislosti s ostatními metodami nejmenších čtverců určenými k řešení systémů (1.2). Na dvě základní metody se zaměříme v této kapitole. Budeme uvažovat nekompatibilní systémy.

2.1 Metoda nejmenších čtverců

Metodu nejmenších čtverců (Ordinary least squares - LS) použijeme k řešení soustavy (1.2) v případě, kdy chyby jsou obsaženy pouze ve vektoru pravé strany (pozorování). Metoda LS hledá vektor b' tak, aby byla splněna následující podmínka

$$b' = \arg \min_{\bar{b} \in \mathcal{R}(A)} \|\bar{b} - b\|_2. \quad (2.1)$$

Jinak řečeno, hledáme minimální změnu, neboli perturbaci, $h \equiv b' - b$ vektoru b tak, aby soustava $Ax = b'$ byla řešitelná. Řešení metodou nejmenších čtverců je pak jakékoli $x \in \mathbb{R}^n$ splňující $Ax = b'$. Je zřejmé, že x' je řešení ve smyslu metody LS právě, když splňuje

$$x' = \arg \min_{\bar{x} \in \mathbb{R}^n} \|A\bar{x} - b\|_2. \quad (2.2)$$

Z podmínky (2.1) vyplývá, že vektor b' je ortogonální projekcí vektoru b do podprostoru $\mathcal{R}(A)$, tedy do prostoru generovaného sloupci matice A . Z vlastností ortogonální projekce vyplývá, že b' definované v (2.1) vždy existuje a je jednoznačné. Řešení metodou LS však jednoznačné není, protože k řešení x' lze přičíst libovolný vektor z jádra matice A a příslušný součet je opět řešením (1.2) ve smyslu nejmenších čtverců. Obvykle hledáme řešení s nejmenší

euklidovskou normou, tedy s nulovým průmětem do jádra $\mathcal{N}(A)$, označíme ho x^{LS} . Lze ukázat (viz [13], věta 2.1), že

$$x' \text{ je řešení metodou LS} \Leftrightarrow A^T(b - Ax') = 0.$$

Řešení x' metodou LS tedy splňuje

$$A^T Ax' = A^T b.$$

Pokud navíc předpokládáme, že $\text{rank}(A) = n$, pak

$$x^{LS} \equiv A^\dagger b, \quad (2.3)$$

kde $A^\dagger = (A^T A)^{-1} A^T$. V obecném případě využijeme singulární rozklad (1.3) matice A a podle [13], věta 2.5, platí

$$x^{LS} \equiv \sum_{i=1}^r \frac{u_i^T b}{\sigma_i} v_i, \quad (2.4)$$

což je konzistentní s (2.3).

Numerické metody výpočtu x^{LS} jsou jak přímé (založené na použití singulárního rozkladu nebo QR-rozkladu matice A), tak iterační, viz [1].

2.2 Metoda úplných nejmenších čtverců

Metodu úplných nejmenších čtverců (Total Least Squares - TLS) použijeme k řešení soustavy (1.2) v případě, kdy chyby jsou obsaženy jak ve vektoru pravé strany b (pozorování), tak v matici A (modelu). Metoda TLS hledá matici $[\tilde{b}, \tilde{A}]$ tak, aby byla splněna následující podmínka

$$[\tilde{b}, \tilde{A}] = \arg_{[\bar{b}, \bar{A}]} \min_{\bar{b} \in \mathcal{R}(\bar{A})} \|[\bar{b}, \bar{A}] - [b, A]\|_F. \quad (2.5)$$

Jinak řečeno, hledáme minimální změnu, neboli perturbaci, $[g, E] \equiv [\tilde{b}, \tilde{A}] - [b, A]$ matice $[b, A]$ tak, aby soustava $\tilde{A}x = \tilde{b}$ byla řešitelná. Podrobně se touto metodou zabývají např. autoři v [13]. Číslo $\|[g, E]\|_F$ nazýváme TLS vzdáleností. Podmínku (2.5) můžeme také přepsat do následující podoby

$$\begin{aligned} \text{a) } & [\tilde{b}, \tilde{A}] \begin{pmatrix} -1 \\ \tilde{x} \end{pmatrix} = ([b, A] + [g, E]) \begin{pmatrix} -1 \\ \tilde{x} \end{pmatrix} = 0 \\ \text{b) } & [g, E] = \arg_{[\bar{g}, \bar{E}]} \min \|[\bar{g}, \bar{E}]\|_F \text{ při splnění a).} \end{aligned} \quad (2.6)$$

S touto podmínkou budeme dále budeme dále pracovat. V této kapitole využijeme jeden důležitý výsledek, jehož důkaz lze nalézt např. v [2].

Věta 2 (Eckart-Young-Mirsky): Nechť $r = \text{rank}(A)$ a nechť A má SVD (1.3). Pak $\forall l < r$

$$\min_{B, \text{rank}(\bar{B})=l} \|B - \bar{B}\|_F = \|B - B_l\|_F = \sqrt{\sum_{i=l+1}^r \sigma_i^2},$$

kde $B_l = \sum_{i=1}^l u_i \sigma_i v_i^T$.

Věta říká, jakým způsobem sestrojít nejbližší aproximaci (ve smyslu Frobeniovy normy) matice A maticí nižší, předem dané, hodnosti.

2.2.1 Existence a jednoznačnost řešení

Existence a jednoznačnost řešení úlohy (2.5) resp. (2.6) je dána tvarem podprostoru generovaného pravými singulárními vektory příslušnými nejmenšímu sigulárnímu číslu matice $[b, A]$. V kapitole 2.2.1 a 2.2.2 budeme předpokládat, že matice A systému (1.2) má plnou sloupcovou hodnost, tedy $r = n$. V opačném případě řešení úlohy (2.5) neexistuje, jak bylo ukázáno v [13]. Sledujme nyní postup v [13]. Protože hodnost matice A je rovna n a systém (1.2) je nekompatibilní, je hodnost matice $[b, A]$ rovna $n + 1$. Je zřejmé, že k tomu, aby systém $\tilde{A}x = \tilde{b}$ byl řešitelný, musí hodnost matice $[\tilde{b}, \tilde{A}]$ být menší nebo rovna n . Nejbližší aproximace matice $[b, A]$ hodnosti n je podle věty 2 s $B \equiv [b, A]$ matice

$$[\tilde{b}, \tilde{A}] = \sum_{i=1}^n \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T. \quad (2.7)$$

Matice atice opravy je

$$[g, E] = [\tilde{b}, \tilde{A}] - [b, A] \equiv -\tilde{u}_{n+1} \tilde{\sigma}_{n+1} \tilde{v}_{n+1}^T \quad (2.8)$$

a norma této matice je

$$\|[g, E]\|_F = \tilde{\sigma}_{n+1}.$$

Systém $[\tilde{b}, \tilde{A}]y = 0$ je řešitelný a jeho řešením je \tilde{v}_{n+1} . Aby soustava $\tilde{A}x = \tilde{b}$ byla řešitelná, je nutné, aby byla splněna následující podmínka

$$\exists y \in \mathbb{R}^n : [\tilde{b}, \tilde{A}]y = 0 \quad \wedge \quad y_{n+1}^1 \neq 0. \quad (2.9)$$

Pokud $\tilde{\sigma}_{n+1}$ je jednoduché vlastní číslo, tak se podmínka (2.9) redukuje na podmínku

$$\tilde{v}_{n+1}^1 \neq 0. \quad (2.10)$$

Pokud je podmínka (2.10) splněna, pak řešení metodou úplných nejmenších čtverců existuje, je jednoznačné a má tvar

$$x^{TLS} = -\tilde{v}_{n+1}/\tilde{v}_{n+1}^1.$$

Pro podrobnosti viz [13] str. 33-37. Řešení metodou úplných nejmenších čtverců lze napsat podle [13], věta 3.7, také v tzv. *uzavřeném tvaru*

$$x^{TLS} = (A^T A - \tilde{\sigma}_{n+1} I)^{-1} A^T b.$$

Případu, kdy podmínka (2.10) není splněna, se budeme věnovat na konci této podkapitoly.

Pokud je $\tilde{\sigma}_{n+1}$ násobné, označme jeho násobost k , pak řešení není jednoznačné, protože není jednoznačně určena matice $[g, E]$ v (2.8). To plyne z nejednoznačnosti matice \tilde{V} v singulárním rozkladu (1.4). Označme V' pravý singulární podprostor příslušný $\tilde{\sigma}_{n+1}$. Aby existovalo řešení, je podle podmínky (2.9) nutné, aby aslepoň jeden vektor z V' měl nenulovou první složku, tedy

$$e_1^T V' \neq 0. \quad (2.11)$$

Předpokládáme, že podmínka (2.10) je splněna, pak podobně jako u metody nejmenších čtverců budeme hledat řešení s nejmenší euklidovskou normou. Můžeme ho získat například Hausholdervou transformací na V' , kdy maximalizujeme první složku vektoru \tilde{v}_{n+1} , tedy

$$V' Q = \begin{pmatrix} 0 & \cdots & 0 & \gamma \\ & & Y & z \end{pmatrix}, \quad (2.12)$$

kde Q má ortonormální sloupce, $\gamma \in \mathbb{R}$, $Y \in \mathbb{R}^{n \times (n-k)}$, $z \in \mathbb{R}^n$. Řešení s nejmenší normou pak získáme jako

$$x^{TLS} \equiv \frac{-z}{\gamma} \quad \text{a} \quad [g, E] = [b, A] \begin{pmatrix} \gamma \\ z \end{pmatrix} (\gamma, z^T). \quad (2.13)$$

Detailně opět viz [13] str. 57-60.

Případu, kdy podmínka (2.11) není splněna, se budeme věnovat na konci této podkapitoly.

Je zajímavé, že pokud platí $\sigma_n > \tilde{\sigma}_{n+1}$, pak jsou podmínky (2.10) resp. (2.11) jsou splněny, jak bylo dokázáno v [13], věta 3.2. Tato podmínka je postačující, nikoli nutná.

Vraťme se k případu, kdy $e_1^T V' = 0$. V takovém případě řešení neexistuje a tato problematika byla podrobně zkoumána v [12]. Tuto situaci názorně demonstruje následující příklad pro $n = 1$ a $m = 2$. Buď

$$[b, A] = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix},$$

pak $\sigma_2 = 1$ a $v_2 = (0, 1)^T$. Podle věty 2 je

$$[\tilde{b}, \tilde{A}] = \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix},$$

tedy soustava $\tilde{A}x = \tilde{b}$ není řešitelná. Na druhou stranu je řešitelná soustava

$$\begin{pmatrix} 0 \\ \epsilon \end{pmatrix} x = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

s řešením $x = 2/\epsilon$. Tedy libovolně málo změněná soustava už je řešitelná, ale pro normu řešení této soustavy platí

$$\epsilon \rightarrow 0 \Rightarrow \|x\| \rightarrow \infty.$$

Sabine Van Huffel pro tento případ definovala tzv. *nongeneric solution*, o kterém se podrobněji zmíníme v další podkapitole. Toto řešení však není řešením úlohy (2.5).

2.2.2 Negerické řešení

Pro případ $e_1^T V' = 0$, kde V' je jako v předchozí kapitole, zavedla Sabine Van Huffel tzv. *nongeneric solution* problému (2.6) (viz [13], str. 66-76). Celá konstrukce vychází z následující myšlenky: nechť například $\tilde{\sigma}_k > \tilde{\sigma}_{k+1}$ je nejmenší singulární číslo matice $[b, A]$ takové, že pravý podprostor příslušný tomuto singulárnímu číslu obsahuje vektor s nenulovou první složkou, pak pro jednoduché singulární číslo definujeme

$$[g, E] \equiv -\tilde{u}_l \tilde{\sigma}_l \tilde{v}_l^T, \quad [\tilde{b}, \tilde{A}] \equiv \sum_{i=1, i \neq l}^{n+1} \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T,$$

a TLS distance $\|[g, E]\|_F$ je v tomto případě rovna $\tilde{\sigma}_l$ a dále postupujeme jako v předchozí podkapitole. V případě násobného singulárního čísla využijeme k hledání řešení s minimální normou Hausholderovu transformaci podobně jako pro klasické řešení.

Kapitola 3

Metody nejmenších čtverců s omezením hodnoty

V mnoha aplikacích, např. v oblasti zpracování signálu nebo v počítačové tomografii, se setkáváme s takzvanými ill-posed problémy, tedy problémy (1.2), které jsou v nějakém smyslu špatně postaveny. Abychom získali nějaké smyslupné řešení těchto problémů, musíme příslušný problém regularizovat. Jednou z možností regularizace je nahradit původní systém systémem nižší hodnoty.

3.1 Ill-posed problémy

Ill-posed jsou většinou označovány problémy, kde malá změna vstupních dat vede k velkým změnám v řešení. Pro ilustraci vlastnosti "ill-posed" uvedeme následující příklad: Představme si, že řešíme problém s nepřesnou pravou stranou, tedy

$$Ax \approx b, \quad b = b_{exact} + b_{noise}, \quad Ax_{exact} = b_{exact}, \quad (3.1)$$

kde o b_{noise} víme pouze $\|b_{exact}\| \gg \|b_{noise}\|$. Pokud budeme soustavu (3.1) řešit metodou nejmenších čtverců, získáváme řešení

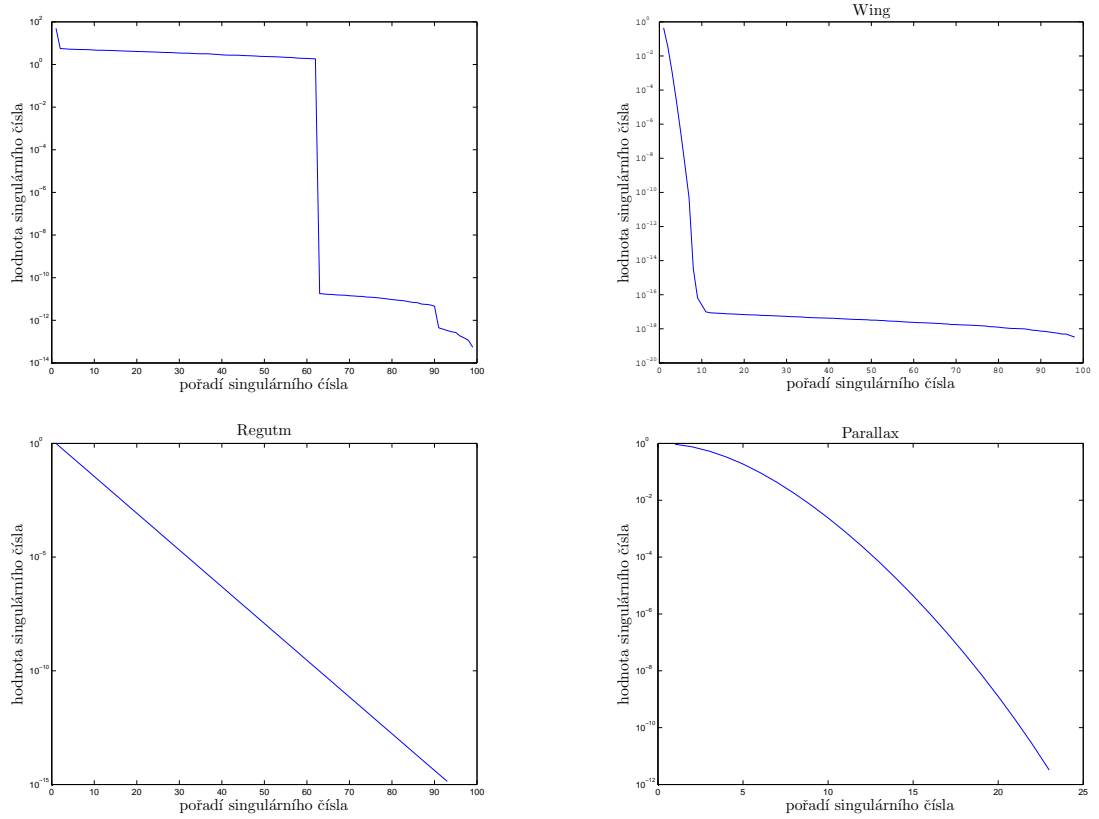
$$x = \sum_{i=1}^r \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^r \frac{u_i^T b_{exact}}{\sigma_i} v_i + \sum_{i=1}^r \frac{u_i^T b_{noise}}{\sigma_i} v_i. \quad (3.2)$$

První člen pravé strany tedy odpovídá přesnému řešení x_{exact} a podle diskrétní Picardovy podmínky pravé strany (viz [6]) součiny $u_i^T b_{exact}$ klesají

k nule rychleji než singulární čísla σ_i . Pro druhý člen však žádnou podobnou podmínkou použít nemůžeme a dělení malými singulárními čísly v kombinaci s neklesající nebo pomalu klesající velikostí projekcí vektoru b do podprostorů levých singulárních vektorů může vést k tomu, že člen odpovídající b_{noise} zcela zastíní přesné řešení. Z uvedeného příkladu vidíme, že v některých případech metoda LS není vhodným nástrojem k určení přibližného řešení. Podobný závěr bychom zřejmě mohli udělat i pro TLS.

Ořezávání, tzv. *truncation*, je určeno k řešení soustav typu (1.2), kde matice A resp. $[b, A]$ má neúplnou numerickou hodnotu, tedy jedno nebo více singulárních čísel v rozkladu (1.3) resp. (1.4) je velmi malých. Idea ořezávání je vytvořit z matice A resp. $[b, A]$ „blízkou“ matici s úplnou numerickou hodnotou. Abychom mohli aplikovat výše uvedenou úvahu, bude napřed nutné stanovit numerickou hodnotu matice A resp. $[b, A]$ - určit tzv. *truncation level*. Existuje více způsobů, kterými je možné truncation level určit, některé lze nalézt např. v [14], avšak neexistuje metoda, která by poskytovala dobré výsledky pro všechny matice. Jako nejjednodušší způsob se zřejmě nabízí určit předem toleranci, tedy číslo, které již budeme považovat za nulu (např. 10^{-16}) a singulární čísla menší než tato tolerance budeme brát za nulová. Tato metoda selhává např. v případě, kdy budou všechna singulární čísla nad nebo pod hranici tolerance nebo budou všechna koncentrována poblíž této hranice. Jinou možností je stanovit truncation level až na základě rozložení singulárních čísel. Toho využijeme především v případě, kdy jsou „velká“ a „malá“ singulární čísla dobře oddělena, tedy numerická hodnota je dobře definována. U některých typů matic optimální truncation level prakticky nelze zvolit.

Obrázek (3.1) ukazuje příklady rozložení singulárních čísel matic 100x100, první matice byla vytvořena pro demonstraci uměle s dobře oddělenými singulárními čísly, poslední tři matice pocházejí z [7]. Matice vlevo nahoře má dobře definovanou numerickou hodnotu 63, u matice Wing je vhodné zvolit truncation level 10, tedy v místě ohybu. U zbývajících 2 matic optimální truncation level neexistuje. Ill-posed problémy vedoucí k maticím, které nemají dobře určený numerický rank, nazýváme *diskrétní ill-posed problémy*. V následujícím textu budeme předpokládat, že jsme již nějakou vhodnou metodou truncation level určili a označíme ho l .



Obrázek 3.1: Singulární čísla vybraných matic

3.2 Truncated SVD

Metoda truncated SVD (dále T-SVD), nebo také *truncated least squares*, česky *metoda nejmenších čtverců s omezením hodnoti*, je prvním a nejjednodušším případem využití truncation. Stejně jako u metody LS předpokládáme, že chyby jsou obsaženy pouze ve vektoru pravé strany b systému (1.2). Tato metoda může být považována za nejjednodušší regularizační metodu založenou na SVD. U metody T-SVD nahradíme matici A nejbližší maticí hodností l , kde l je zvolený truncation level, a aplikujeme původní metodu nejmenších čtverců. Nejbližší aproximace matice A hodností l je podle věty 2

$$A_l = \sum_{i=1}^l u_i \sigma_i v_i^T. \quad (3.3)$$

Metodu T-SVD pak formulujeme pomocí podmínky

$$x' = \arg \min_{\bar{x} \in \mathbb{R}^n} \|A_l \bar{x} - b\|_2. \quad (3.4)$$

Použitím (2.4) pro $A \equiv A_l$ ihned vidíme, že řešení metodou T-SVD s nejmenší euklidovskou normou vždy existuje a má tvar

$$x_l \equiv A_l^\dagger b = \sum_{i=1}^l \frac{u_i^T b}{\sigma_i} v_i. \quad (3.5)$$

3.3 Metoda úplných nejmenších čtverců s omezením hodnoti

Podobně jako u metody T-SVD, budeme při metodě úplných nejmenších čtverců s omezením hodnoti (Truncated TLS - T-TLS) matici $[b, A]$ nahrazovat maticí s hodnotí l , kde l je zvolený truncation level. Metodu formulujeme tak, že k podmínkám (2.6) přidáme ještě dodatečnou podmínku

$$\text{rank}([\tilde{b}, \tilde{A}]) = l < r + 1. \quad (3.6)$$

Základní myšlenka je podobná jako u metody T-SVD, tedy že původní matici, v tomto případě $[b, A]$, nahradíme její nejbližší aproximací hodnoti l a aplikujeme původní metodu TLS. Nejbližší aproximace matice $[b, A]$ hodnoti l je podle věty 2

$$[\tilde{b}_l, \tilde{A}_l] = \sum_{i=1}^l \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T \quad (3.7)$$

a matice opravy je

$$[g, E] = - \sum_{i=l+1}^{n+1} \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T. \quad (3.8)$$

Tedy

$$\|[g, E]\|_F = \sqrt{\sum_{i=l+1}^{r+1} \tilde{\sigma}_i^2}. \quad (3.9)$$

Z vlastností metody TLS plyne, že T-TLS řešení nemusí být jednoznačné a nemusí ani existovat. My se zaměříme opět na hledání řešení s nejmenší euklidovskou normou. Můžeme ho získat Hausholedrovou transformací na

podprostor pravých singulárních vektorů příslušných posledním $n + 1 - l$ singulárním číslům (podobně jako v (2.12)), my ale uvedeme jiný způsob, převzatý z [4], který může být v některých případech výhodnější. Nechť (1.4) je singulární rozklad matice $[b, A]$. Označme nyní

$$\begin{aligned} \tilde{U} &= \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{m-l}{\longleftrightarrow} \\ \tilde{U}_1 & \tilde{U}_2 \end{pmatrix}, \quad \tilde{\Sigma} = \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{n+1-l}{\longleftrightarrow} \\ \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{pmatrix} \begin{array}{c} \updownarrow \\ \updownarrow \\ \updownarrow \end{array} \begin{array}{c} l \\ m-l \end{array}, \\ \tilde{V} &= \begin{pmatrix} \tilde{V}_1 & \tilde{V}_2 \end{pmatrix} = \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{n+1-l}{\longleftrightarrow} \\ \tilde{V}_{11} & \tilde{V}_{12} \\ \tilde{V}_{21} & \tilde{V}_{22} \end{pmatrix} \begin{array}{c} \updownarrow \\ \updownarrow \\ \updownarrow \end{array} \begin{array}{c} 1 \\ n \end{array}. \end{aligned} \quad (3.10)$$

Zápisy (3.7) a (3.8) pak přepíšeme na

$$[\tilde{b}_l, \tilde{A}_l] = \tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^T, \quad [g, E] = \tilde{U}_2 \tilde{\Sigma}_2 \tilde{V}_2^T.$$

Řešení bude existovat, pokud podprostor pravých singulárních vektorů příslušných posledním $n + 1 - l$ singulárních číslům bude obsahovat vektor s nenulovou první složkou. To můžeme zařídit tím, že zvolíme truncation level l tak, aby

- 1) $\tilde{\sigma}_l > \tilde{\sigma}_{l+1}$
 - 2) $l \leq \text{numRank}([b, A])$
 - 3) $\tilde{V}_{12} \neq 0$.
- (3.11)

Řešení ve smyslu metody T-TLS má tvar

$$\tilde{x}_l = (\tilde{V}_{21}^T)^\dagger \tilde{V}_{11}^T = -\tilde{V}_{22} \tilde{V}_{12}^\dagger. \quad (3.12)$$

Norma řešení a velikost T-TLS vzdálenosti jsou pak dány vztahy

$$\|\tilde{x}_l\| = \sqrt{\|\tilde{V}_{12}\|^{-2} - 1} \quad (3.13)$$

a

$$\|[g, E]\|_F = \sqrt{\sum_{i=l+1}^{r+1} \tilde{\sigma}_i^2}.$$

Odvození (3.12) lze najít v [8].

Někteří autoři, např. v [13] a [15] resp. [16], metodu neoznačují jako truncated TLS, ale pouze jako TLS pro násobná singulární čísla $\tilde{\sigma}_{l+1} = \dots = \tilde{\sigma}_{n+1}$ a tedy

$$[g, E] = - \sum_{i=l+1}^{n+1} \tilde{u}_i \tilde{\sigma}_i \tilde{v}_i^T \equiv -\tilde{u}_{l+1} \tilde{\sigma}_{l+1} \tilde{v}_{l+1}^T,$$

což však nemá vliv na tvar řešení.

Kapitola 4

Vlastnosti metody úplných nejmenších čtverců s omezením hodnosti

Z předchozích kapitole jsme se seznámili s metodou úplných nejmenších čtverců s omezením hodnosti. Zatím nevíme, jaké vlastnosti řešení touto metodou můžeme očekávat. V kapitole se zaměříme na regularizační vlastnosti metody a dáme ji do souvislosti s dalšími regularizačními metodami založenými na singulárním rozkladu. Pokusíme se také popsat systémy, při jejichž řešení poskytují metody T-TLS a T-SVD podobné výsledky. Z kapitoly 3 víme, že omezení hodnosti snižuje u uvažované metody citlivost na data úlohu, v této kapitole uvedeme konkrétní odhady pro změnu řešení při perturbaci dat.

4.1 Regularizace pomocí T-TLS a T-SVD

Vraťme se nyní k tomu, jak je možno ill-posed problémy regularizovat. Nejjednodušší regularizační metody vycházejí přímo ze singulárního rozkladu (1.3) matice A a řešení těmito regularizačními metodami lze zapsat ve tvaru

$$x = \sum_{i=1}^r f_i \frac{u_i^T b}{\sigma_i} v_i, \quad (4.1)$$

kde čísla f_i jsou tzv. *filter factors*. Srovnáme-li nyní (4.1) s (3.5), vidíme, že metoda T-SVD je příkladem tohoto typu regularizačních metod s $f_i = 1$ pro

$i = 1, \dots, l$ a $f_i = 0$ pro $i = l + 1, \dots, r$, přičemž l zde chápeme jako regularizační parametr. Další známou možností je použití Tichonovovy regularizace, která balancuje chybu rezidua a euklidovskou normu řešení, tedy hledáme

$$\arg_{x \in \mathbb{R}^n} \min x \in \mathbb{R}^n \{ \|Ax - b\|^2 + \lambda \|Lx\|^2 \}, \quad (4.2)$$

kde λ je "penalizační" parametr a matice L má plný řádkový rank. Často volíme za L jako matici jednotkovou a není těžké ukázat, že pro tento případ můžeme řešení zapsat ve tvaru

$$x^{Tik} = \sum_{i=1}^r \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{u_i^T b}{\sigma_i} v_i. \quad (4.3)$$

Tak je (při vhodném λ) odstraněn problém dělení malými singulárními čísly.

V následujícím textu se zaměříme na metodu T-TLS. Uvažujme řešení ve tvaru (3.12) a kromě podmínek (3.11) budeme navíc l volit takové, aby $\|\tilde{V}_{12}\| \geq \tau$, kde pomocí τ bude budeme kontrolovat normu řešení. Nejdříve se zaměříme na to, zda metoda T-TLS skutečně funguje jako filtrační metoda. Nabízí se srovnat řešení metodu T-SVD a T-TLS. Pro $\|x_l - \tilde{x}_l\|$, kde \tilde{x}_l a x_l jsou T-TLS resp. T-SVD řešení, máme podle kapitoly 4.2 různé odhady, avšak ve všech pravá strana závisí alespoň nepřímo na podílu $\left(\frac{\tilde{\sigma}_{l+1}}{\sigma_l}\right)$.

Aby pro nás měl tedy tento odhad význam, je zapotřebí aby $1 \gg \frac{\tilde{\sigma}_{l+1}}{\sigma_l}$. To by ovšem znamenalo, že $\sigma_l \gg \sigma_{l+1}$, což zřejmě neodpovídá diskretním ill-posed problémům. Jiná možnost je pokusit se převést tvar řešení (3.12) na tvar (4.1) pro vhodné f_i . Abychom mohli něco takového vůbec provést, je nejdříve nutné ověřit, zda \tilde{x}_l nemá nenulové projekce do prostorů generovaných pravými singulárními vektory matice A , které nejsou obsaženy s sumě (4.1). Jinak řečeno ověřit implikace

$$u_i^T b = 0 \Rightarrow v_i^T \tilde{x}_l = 0 \quad \text{a} \quad i > r \Rightarrow v_i^T \tilde{x}_l = 0.$$

Obě tyto implikace byly pro jednoduchá singulární čísla $\tilde{\sigma}_i$ a σ_i ověřeny v [4] sérií technických lemat. Tímto je zřejmě dokázáno, že \tilde{x}_l je tvaru (4.1). Nyní uvedeme důležitý výsledek o tvaru filtračních faktorů.

Věta 3 (Filter factors): Nechť (1.3) a (1.4) je singulární rozklad matice A resp. $[b, A]$ a nechť všechna singulární čísla obou matic jsou jednoduchá, pak filtrační faktory v (4.1) mají pro metodu T-TLS tvar

$$f_i = \sum_{j=l+1}^{n+1} \frac{\sigma_i^2}{\sigma_i^2 - \tilde{\sigma}_j^2} \frac{(\tilde{v}_j^1)^2}{\|\tilde{V}_{12}\|^2} = \sum_{j=1}^l \frac{\sigma_i^2}{\tilde{\sigma}_j^2 - \sigma_i^2} \frac{(\tilde{v}_j^1)^2}{\|\tilde{V}_{12}\|^2}.$$

Důkaz této věty lze nalézt v [4] stejně jako následující odhady. Při splnění předpokladů věty 3 platí pro f_i následující nerovnosti

$$0 \leq f_i - 1 \leq \frac{\tilde{\sigma}_{l+1}^2}{\sigma_i^2 - \tilde{\sigma}_{l+1}^2} \text{ a } f_i \text{ rostou monotónně pro } i \leq l, \quad (4.4)$$

$$0 \leq f_i \leq \|\tilde{V}_{12}\|^{-2} \frac{\sigma_i^2}{\tilde{\sigma}_l^2 - \sigma_i^2} \text{ pro } l < i \leq r. \quad (4.5)$$

Z odhadů rovněž plyne, že pro dobře definovaný numerický rank l budou pro $i = l + 1, \dots, r$ budou hodnoty $f_i \approx 0$ (využijeme-li navíc předpoklad kladený na $\|V_{12}\|$), nicméně i když neexistují mezi singulárními čísly velké mezery, bude zachována regularizační vlastnost metody T-TLS a významnou výhodou oproti metodě T-SVD je, že v takovém případě zůstává v omezené míře zachován i vliv prvků s indexem větším než l . Zůstává však důležitá otázka, jaký zvolit regularizační parametr l , aby na jedné straně bylo potlačeno destruktivní chování šumu a na druhé straně nebyly ztraceny užitečné informace z původní úlohy. Více o této problematice lze nalézt např. v [14].

4.2 Souvislost T-TLS a T-SVD

V této kapitole se zaměříme na zkoumání vztahů mezi dvěma na první pohled rozdílnými metodami. Sledujme postup uvedený v [3]. Protože tvary (3.12) a (3.5) nám příliš prostoru pro srovnávání těchto dvou přístupů nedávají, pokusíme se řešení metodou T-TLS převést do tvaru podobného T-SVD řešení. Truncation level bude stejný u obou metod, bude splňovat (3.11) a označíme ho opět l .

Připomeňme, že řešení metodou T-SVD splňuje rovnost

$$A_l^T A_l x_l = A_l^T b, \quad (4.6)$$

kde A_l nejbližší aproximace matice A hodnosti l definovaná ve větě 2. Nechť $[\tilde{b}_l, \tilde{A}_l]$ je stejně jako v předchozím textu nejbližší aproximace matice $[b, A]$ hodnosti l . Jednoduchou úvahou zjistíme, že T-TLS řešení \tilde{x}_l můžeme získat jako LS řešení s maticí \tilde{A}_l . \tilde{x}_l tedy splňuje

$$\tilde{A}_l^T \tilde{A}_l x_l = \tilde{A}_l^T b. \quad (4.7)$$

Nyní je zřejmě třeba jistým způsobem porovnat matice \tilde{A}_l a A_l . My to uděláme tak, že si matici \tilde{A}_l vyjádříme jako perturbaci matice A_l , tedy $\tilde{A}_l = A_l + E$. Pokud se budou sloupcové a řádkové vektory uvedených dvou matic výrazně lišit, budou metody poskytovat velmi rozdílná řešení a proto se v dalším textu zaměříme pouze na matice, které jsou si blízké ve smyslu následující definice uvedené v [3].

Definice (Acute perturbation): Řekneme, že matice A je *acute* perturbací matice B , pokud jsou splněny následující podmínky

$$\|P_A - P_B\| < 1$$

$$\|P_{A^T} - P_{B^T}\| < 1,$$

kde P_C je ortogonální projekce na $\mathcal{R}(C)$.

Není těžké ověřit, že A je acute perturbací B právě když úhly mezi prostory tvořenými sloupci resp. řádky matic jsou ostré. Zřejmě ale není příliš výhodné ověřovat acute podmínku tímto způsobem ani přímo z definice. Autoři v [3] dokázali, že pro to, aby matice A_l a \tilde{A}_l byly acute, je postačující, aby byla splněna podmínka $\tilde{\sigma}_{l+1} < \sigma_l$ (tedy stejná podmínka, která je postačující pro existenci T-TLS řešení). V dalším textu již budeme pracovat pouze s maticemi A , pro které je tato podmínka splněna.

Rozdělme nyní matici U , V a Σ podobně jako jsme rozdělili matice vzniklé singulárním rozkladem matice $[b, A]$.

$$\begin{aligned}
 U &= \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{m-l}{\longleftrightarrow} \\ U_1 & U_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{n-l}{\longleftrightarrow} \\ \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \quad \begin{matrix} \updownarrow \\ \updownarrow \end{matrix} \quad \begin{matrix} l \\ m-l \end{matrix} \\
 V &= (V_1 \quad V_2) = \begin{pmatrix} \overset{l}{\longleftrightarrow} & \overset{n-l}{\longleftrightarrow} \\ V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad \begin{matrix} \updownarrow \\ \updownarrow \end{matrix} \quad \begin{matrix} 1 \\ n \end{matrix} \tag{4.8}
 \end{aligned}$$

Pomocí nových matic pak přepíšeme rovnice (4.6) a (4.7) do tvaru

$$\begin{aligned}
 V_1 \Sigma_1^T \Sigma_1 V_1^T x_l &= A^T b - V_2 \Sigma_2^T U_2^T b \\
 \tilde{V}_{21} \tilde{\Sigma}_1^T \tilde{\Sigma}_1 \tilde{V}_{21}^T \tilde{x}_l &= A^T b - \tilde{V}_{22} \tilde{\Sigma}_2^T \tilde{U}_2^T b = A^T b - \tilde{V}_{22} \tilde{\Sigma}_2^T \tilde{\Sigma}_2 \tilde{V}_{12}^T
 \end{aligned} \tag{4.9}$$

a dosazením první rovnice do druhé, odečtením a zanedbáním rozdílových členů vyšších řádů získáme

$$V_1 \Sigma_1^T \Sigma_1 V_1^T (x_l - \tilde{x}_l) \approx V_1 \Sigma_1^T \Sigma_1 \Delta V^T + \Delta V \Sigma_1^T \Sigma_1 V_1^T + V_1 \Delta S V_1^T + \tilde{V}_{22} \tilde{\Sigma}_2^T \tilde{\Sigma}_2 \tilde{V}_{12}^T - V_2 \Sigma_2^T U_2^T b, \quad (4.10)$$

kde $\Delta V = \tilde{V}_{21} - V_1$ a $\Delta S = \tilde{\Sigma}_2^T \tilde{\Sigma}_2 - \Sigma_1^T \Sigma_1$. Tuto rovnici můžeme dále odhadnout (podle definice spektrální normy)

$$\|x_l - \tilde{x}_l\| \leq \frac{1}{\sigma_l^2} [(2\|\Delta V\| \tilde{\sigma}_1^2 + \|\Delta S\|) \|\tilde{x}_l\| + \tilde{\sigma}_{l+1}^2 + \sigma_{l+1} \|U_2^T b\|]. \quad (4.11)$$

Odhad není sice nejlepší možný, ale způsob jakým jsme ho odvodili nám napovídá, že řešení se budou velmi lišit v případech, kdy:

- neexistuje ztatečná mezera mezi malými a velkými singulárními čísly matice A
- systém je značně nekompatibilní, tedy $\tilde{\sigma}_{l+1}$ je velké
- A_l je špatně podmíněná a tedy σ_l je malé
- jsou velké rozdíly mezi singulárními čísly matice A a matice $[b, A]$, speciálně pokud $|u_l^T b|$ roste, pak roste i rozdíl $\tilde{\sigma}_l^2 - \sigma_l^2$, podrobně viz [3].

Nyní uvedeme několik konkrétních výsledků pro odhad normy $\|x_l - \tilde{x}_l\|$ vzhledem k $\|x_l\|$. Podrobnou analýzu tohoto problému lze nalézt v [15] a [16].

Věta 4: Buď x_l T-SVD řešení s minimální normou a \tilde{x}_l T-TLS řešení s minimální normou. Nechť $\tilde{\sigma}_{l+1} < \frac{1}{2}\sigma_l$, pak platí

$$\frac{\|x_l - \tilde{x}_l\|}{\|x_l\|} \leq \frac{\tilde{\sigma}_{l+1}}{\sigma_l - \tilde{\sigma}_{l+1}} \left[3 + \kappa(A_k) \frac{\|b - Ax_l\|}{\|b\|} \right].$$

Důkaz: viz [3], věta 3.2.

Věta 5: Nechť x_l a \tilde{x}_l jsou jako ve větě 4 a nechť platí $\tilde{\sigma}_l > \tilde{\sigma}_{l+1} = \dots = \tilde{\sigma}_{n+1}$, pak platí

$$\|x_l - \tilde{x}_l\| \leq \frac{\tilde{\sigma}_{n+1}^2}{\sigma_l^2 - \tilde{\sigma}_{n+1}^2} \|x_l\|$$

$$\|x_l - \tilde{x}_l\| \leq \frac{\tilde{\sigma}_{n+1}^2}{\sigma_l^2} \|\tilde{x}_l\|.$$

Důkaz: viz [16], věta 3.1.

Další výsledek, který zde uvedeme se už netýká řešení s minimální normou, ale libovolného řešení. Nechť S_{T-SVD} a S_{T-TLS} je množina všech T-SVD, resp. T-TLS řešení systému (1.2), tedy

$$S_{T-SVD} = \{x | x = A_l^\dagger b + (I - A_l^\dagger A_l)y; y \in \mathbb{R}^n\}$$

$$S_{T-TLS} = \{x | x = \tilde{A}_l^\dagger b + (I - \tilde{A}_l^\dagger \tilde{A}_l)y; y \in \mathbb{R}^n\},$$

přičemž $(I - C^\dagger C)$ je projekce na $\mathcal{N}(C)$.

Věta 6: Nechť x je libovolný prvek S_{T-SVD} a nechť platí $\tilde{\sigma}_l > \tilde{\sigma}_{l+1} = \dots = \tilde{\sigma}_{n+1}$, pak existuje takový prvek $\tilde{x} \in S_{T-TLS}$, že platí

$$\|x - \tilde{x}\| \leq \frac{\tilde{\sigma}_{n+1}^2}{\sigma_l^2} \|\tilde{x}_l\|,$$

kde \tilde{x}_l je jako ve větě 4.

Důkaz: viz [16], věta 3.2.

Podobným způsobem jako jsme zkoumali $\|x_l - \tilde{x}_l\|$ můžeme zkoumat také vztah reziduí $r_l = b - Ax_l$ a $\tilde{r}_l = b - A\tilde{x}_l$. Uvedeme pouze jeden výsledek týkající se tohoto problému, další odhady lze opět nalézt v [15] a [16].

Věta 7: Nechť x_l a \tilde{x}_l jsou jako ve větě 4 a nechť $\tilde{\sigma}_{l+1} < \sigma_l$, pak

$$\|r_l - \tilde{r}_l\| \leq 2 \frac{\tilde{\sigma}_{l+1}}{\sigma_l - \tilde{\sigma}_{l+1}} \|b\|.$$

Důkaz: viz [3], věta 3.3.

4.3 Vliv perturbací na řešení metodou T-TLS

V této kapitole se budeme zabývat především citlivostí řešení metodou T-TLS na data úlohy, tedy matici A a vektor pozorování b . Ještě před tím však

zmíníme jeden zajímavý a důležitý výsledek uvedený v [13] z jiné, příbuzné oblasti - aproximace T-TLS řešení T-SVD řešením.

Věta 8: Buď x_0 je LS řešení systému $A_0x = b_0$ s minimální euklidovskou normou a buď $\text{rank}(A_0) = l$. Nechť $[b, A]$ je matice stejného řádu jako $[b_0, A_0]$, \bar{A} matice s $\text{rank}(\bar{A}) = l$ a $[\bar{b}, \bar{A}]$ ortogonální projekce $[b, A]$ na $\mathcal{R}(\bar{A})$. Nechť navíc existuje ϵ takové, že $\|[b, A] - [b_0, A_0]\|_F \leq \epsilon$, $\|A - \bar{A}\|_F \leq \epsilon$ a $\epsilon < \frac{1}{2}\|A_0^\dagger\|^{-1}$. Buď $[\tilde{b}_l, \tilde{A}_l]$ jako v (3.7) a \tilde{x}_l buď řešení systému $\tilde{A}_l x = \tilde{b}_l$ s nejmenší euklidovskou normou. Pak pro \bar{x} řešení $\bar{A}x = \bar{b}$ s nejmenší normou platí následující odhad

$$\|\tilde{x}_l - \bar{x}\| \leq 2\epsilon^2\alpha^2(\sqrt{1 + \|x_0\|^2} + \|x_0\|), \text{ kde } \alpha = \frac{\|A_0^\dagger\|}{1 - 2\epsilon\|A_0^\dagger\|}$$

a tedy

$$\bar{x} = \tilde{x}_l + \mathcal{O}(\epsilon^2).$$

Věta říká, že pokud známe hodnotu matice A_0 , pak řešení, které spočítáme z libovolné matice \bar{A} splňující předpoklady věty 4, bude blízké T-TLS řešení systému $Ax \approx b$ s truncation levellem rovným hodnotě matice A_0 .

Vraťme se nyní k otázce citlivosti T-TLS řešení na změny v matici A a pravé straně b systému (1.2). Vyjdeme z analýzy, kterou provedli autoři v [13] pro metodu úplných nejmenších čtverců a využijeme konkrétních výsledků pro metodu T-TLS uvedených v [3], [15] a [16].

Označme

$$\bar{A}x \approx \bar{b} \tag{4.12}$$

soustavu vzniklou perturbací soustavy (1.2),

$$[\bar{b}, \bar{A}] = \sum_{i=1}^n \bar{u}_i \bar{\sigma}_i \bar{v}_i^T, \quad \bar{A} = \sum_{i=1}^n u'_i \sigma'_i v'^T_i \tag{4.13}$$

příslušný singulární rozklad a

$$[\Delta b, \Delta A] \equiv [\bar{b}, \bar{A}] - [b, A], \quad \|[b, A]\| \gg \|[\Delta b, \Delta A]\| \tag{4.14}$$

matici příslušné perturbace. U obou řešení (původního i perturbovaného) uvažujeme stejný truncation level a označíme ho l .

Může se stát, že

$$\exists i > l : \bar{\sigma}_i \geq \tilde{\sigma}_l \quad (4.15)$$

nebo

$$\exists i < l + 1 : \bar{\sigma}_i \leq \tilde{\sigma}_{l+1}, \quad (4.16)$$

tedy že se perturbováním "promíchají" malá a velká singulární čísla. Stejně číslo pak může být v singulárním rozkladu považováno za aproximaci nuly, zatímco v druhém ho považujeme za číslo nenulové. Taková situace zřejmě není žádoucí. Nadále budeme proto uvažovat, že perturbace (4.14) splňuje podmínku

$$\|[\Delta b, \Delta A]\| < \tilde{\sigma}_l - \tilde{\sigma}_{l+1}, \quad (4.17)$$

která zaručuje, že nenastane ani jeden z případů (4.15) a (4.16). Zaměříme se nyní na existenci T-TLS řešení perturbované soustavy (4.12). Abychom měli zaručenu existenci T-TLS řešení soustavy (1.2), budeme podle kapitoly 2.2.1 předpokládat pro další úvahy, že $\sigma_l > \tilde{\sigma}_{l+1}$. Pokud podmínku (4.17) zesílíme na podmínku

$$\|[\Delta b, \Delta A]\| < \sigma_l - \tilde{\sigma}_{l+1}, \quad (4.18)$$

pak $\sigma'_l > \tilde{\sigma}_{l+1}$ a tedy řešení perturbované soustavy (4.12) metodou T-TLS existuje. Podrobně viz [3]. V následujícím textu předpokládáme, že matice perturbace (4.14) splňuje podmínku (4.18).

Protože problematika citlivosti metody T-TLS na perturbace nebyla ještě zcela vyřešena, budeme se dále zabývat dvěma dílčími problémy

- citlivost řešení metodou T-TLS na speciální perturbace
- citlivost $\mathcal{N}([\tilde{b}_l, \tilde{A}_l])$ na perturbace

V této části bude \tilde{x}_l značit T-TLS řešení s nejmenší euklidovskou normou původní soustavy (1.2) a \bar{x}_l T-TLS řešení s nejmenší euklidovskou normou perturbované soustavy (4.12).

Citlivost řešení metodou T-TLS na speciální perturbace

V této sekci uvedeme výsledek, který prezentoval Wei v [15] a [16].

Věta 9: Nechť $\|[\Delta b, \Delta A]\| \leq \frac{1}{6}(\sigma_l - \tilde{\sigma}_{n+1})$, pak platí

$$\|\tilde{x}_l - \bar{x}_l\| \leq \frac{\|[\Delta b, \Delta A]\| + \tilde{\sigma}_{n+1}}{\sigma_l + \tilde{\sigma}_{n+1}} (3 + 5\|\tilde{x}_l\|) \leq \frac{6(\|[\Delta b, \Delta A]\| + \tilde{\sigma}_{n+1})}{\sigma_l + \tilde{\sigma}_{n+1}} \sqrt{1 + \|\tilde{x}_l\|^2}.$$

Autor v [15] rovněž ukázal, že pro libovolné T-TLS řešení \bar{x} perturbovaného problému (4.12) lze nalézt T-TLS řešení \tilde{x} původního problému (1.2) tak, že

$$\frac{\|\tilde{x} - \bar{x}\|}{\|\tilde{x}\|}$$

je stejného řádu jako

$$\frac{\|\tilde{x}_l - \bar{x}_l\|}{\|\tilde{x}_l\|}.$$

Citlivost $\mathcal{N}([\tilde{b}_l, \tilde{A}_l])$ na perturbace

Protože řešení \tilde{x}_l je lineární kombinací vektorů z $\mathcal{N}([\tilde{b}_l, \tilde{A}_l])$ a \bar{x}_l je lineární kombinací vektorů z $\mathcal{N}([\bar{b}_l, \bar{A}_l])$, kde

$$[\bar{b}_l, \bar{A}_l] \equiv \sum_{i=1}^l \bar{u}_i \bar{\sigma}_i \bar{v}_i^T,$$

nabízí se pro srovnání těchto řešení porovnat tyto prostory. Touto problematikou se podrobně pro metodu TLS zabývali autoři v [13] v kapitole 7.3. Důsledky pro metodu T-TLS lze najít v [3]. Zde uvedeme pouze jeden základní výsledek, jehož důkaz lze nalézt právě v [3].

Věta 10: Nechť $\|[\Delta b, \Delta A]\| \leq \sigma_l - \tilde{\sigma}_{l+1}$, pak platí

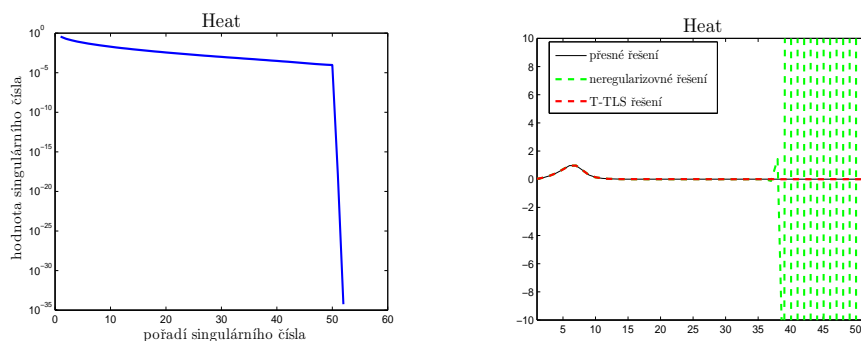
$$\text{dist} \left(\mathcal{N}([\tilde{b}_l, \tilde{A}_l]), \mathcal{N}([\bar{b}_l, \bar{A}_l]) \right) \leq \frac{\|[\Delta b, \Delta A]\|}{\sigma_l - \tilde{\sigma}_{l+1} - \|[\Delta b, \Delta A]\|}.$$

Kapitola 5

Numerické experimenty

V této kapitole prakticky demonstrujeme poznatky o regularizačních vlastnostech metod T-TLS a T-SVD uvedené v sekci 4.1 a současně poznatky o souvislosti těchto metod uvedené v sekci 4.2. Experimenty byly vyhotoveny v programu MATLAB R2008a s využitím balíčku Regularization Tools [7]. K určení optimálního truncation levelu jsme využili metodu L-křivky, více o této metodě lze nalézt v [14].

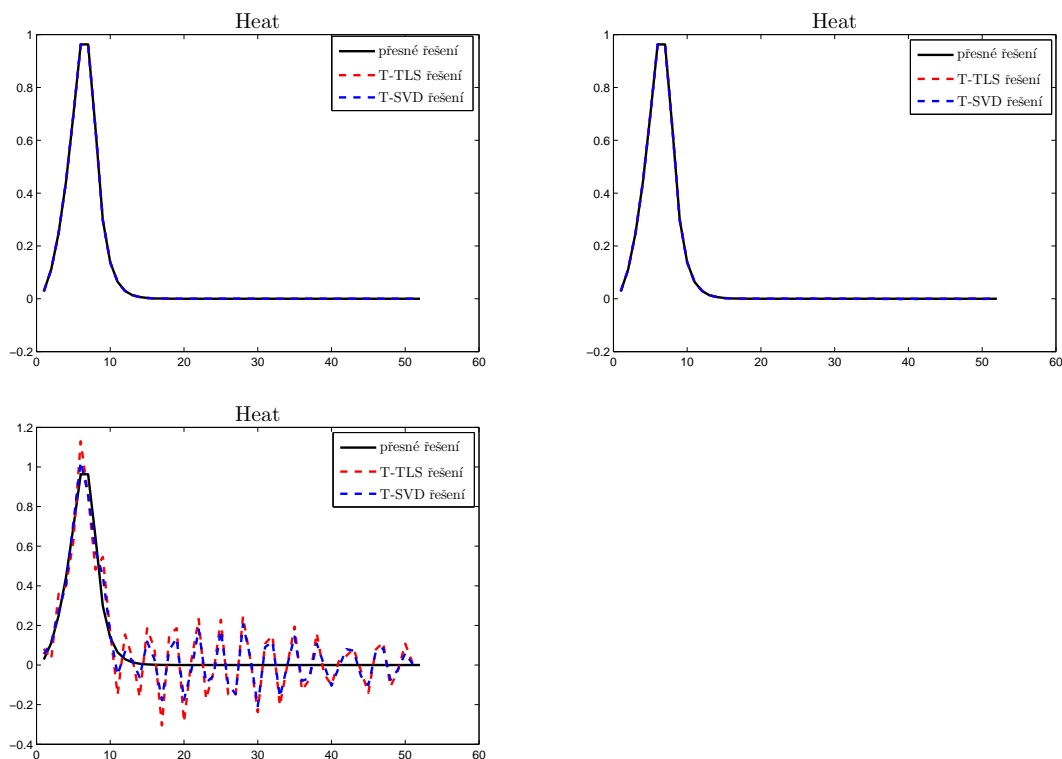
Experiment 1. V prvním experimentu ukážeme, jak se liší naivní řešení systému (1.2) získané jako $x = A^\dagger b$ a řešení získané metodami s omezením hodnoty, pokud matice A má neúplnou, ale dobře definovanou numerickou hodnotu. Pro demonstraci jsme vybrali problém *Heat* z balíčku [7]. Pomocí příkazu `[A, b, x]=heat(52)` jsme setrojili matici A velikosti 52×52 a řešení x . Vektor pravé strany b_{exact} jsme setrojili jako `b_exact=A*x`. Pomocí funkce `rand` jsme vytvořili vektor b_{noise} tak, aby $\frac{\|b_{noise}\|}{\|b_{exact}\|} = 10^{-12}$ a vektor b jsme pak získali jako $b = b_{exact} + b_{noise}$. Na obrázku 5.1 vlevo vidíme, že poslední dvě singulární čísla matice A jsou výrazně menší než čísla předchozí. Numerická hodnota matice je tedy 50 a je dobře definovaná. K řešení systému $Ax = b$ jsme použili metodu nejmenších čtverců a metodu úplných nejmenších čtverců s omezením hodnoty (T-TLS) s truncation levelem 50. Grafické porovnání těchto výsledků najdeme na obrázku 2.1 vpravo. Vidíme, že metoda T-TLS aproximuje přesné řešení x původního systému (nezatíženého šumem) velmi dobře, zatímco řešení získané metodou LS bez regularizace osciluje a k přesnému řešení se ani nepřiblížilo. Určitá regularizace je v tomto případě nutná.



Obrázek 5.1: Singulární čísla matice Heat (vlevo) a porovnání neregularizovaného řešení a T-TLS řešení se skutečným původním řešením (vpravo).

V druhé části experimentu se zaměříme na porovnání metody T-TLS s metodou T-SVD při různých hladinách šumu. Protože malá a velká singulární čísla matice A jsou dobře oddělena, lze předpokládat, že pokud nebude pravá strana příliš porušená, budou T-TLS a T-SVD řešení dobře aproximovat původní přesné řešení x . Podle (4.11) můžeme rovněž předpokládat, že metody T-SVD a T-TLS budou poskytovat řešení, která si budou blízká. Jako zkušební hladiny šumu $\frac{\|b_{noise}\|}{\|b_{exact}\|}$ jsme zvolili hodnoty 10^{-8} , 10^{-5} a 10^{-2} . Obrázek 5.2 ukazuje, že pro první dvě hodnoty šumu, kdy truncation level určený L-křivkou byl 50, jako v předchozí části, se T-TLS a T-SVD řešení téměř neliší. Při hladině šumu 10^{-2} a truncation levelu 33 (určeného pomocí L-křivky) už přesné řešení bylo aproximováno podstatně hůře a řešení pomocí metod T-TLS a T-SVD se lišila více. Odlišnost T-TLS a T-SVD řešení může být způsobena tím, že singulární čísla σ_{33} a σ_{34} jsou si blízká a rozdíl $\sigma_k - \tilde{\sigma}_{k+1}$ ve větě 4 je tedy malý. Výsledky numerického experimentu s maticí Heat jsou tedy konzistentní s odhady, které jsme učinili na začátku této části.

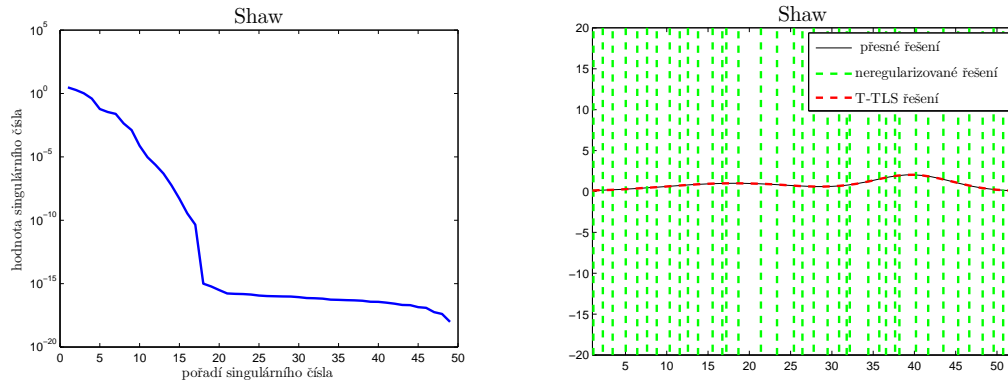
Experiment 2. V tomto uvažujeme soustavu s maticí se singulárními čísly rozloženými do tvaru „L“. Taková matice má dobře definovanou numerickou hodnotu, ale velká a malá singulární čísla nejsou oddělena tak dobře, jako v předchozím experimentu. Pro demonstraci jsme vybrali problém *Shaw* z balíčku [7]. Matici, přesné řešení a pravou stranu jsme sestrojili stejným způsobem jako pro matici Heat v experimentu 1. Hladina šumu je $\frac{\|b_{noise}\|}{\|b_{exact}\|} = 10^{-12}$. Na obrázku 5.3 vlevo vidíme, že k ohybu dochází v σ_{19} ,



Obrázek 5.2: Porovnání metod T-TLS a T-SVD při řešení problému Heat při hladině šumu 10^{-8} (vlevo nahoře), 10^{-5} (vpravo nahoře) a 10^{-2} (vlevo dole).

numerický rank této matice je tedy 19. K řešení systému $Ax = b$ jsme opět použili metodu nejmenších čtverců a metodu úplných nejmenších čtverců s omezením hodnoty s truncation levelem 19. Protože matice A má neúplnou numerickou hodnotu, očekáváme, že řešení metodou nejmenších čtverců bude oscilovat podobně jako v experimentu 1, tuto úvahu potvrzuje obrázek 5.3 vpravo. Opět je nutná regularizace. Protože matice Shaw má dobře určenou numerickou hodnotu řešení metodou T-TLS aproximuje při nízké hladině šumu přesné řešení dobře.

Druhá část experimentu bude totožná s druhou částí experimentu 1. V případě matice Shaw očekáváme, že k postřehnutelnému zhoršování vlastností aproximace bude docházet již u nižší hladiny šumu, protože přechod

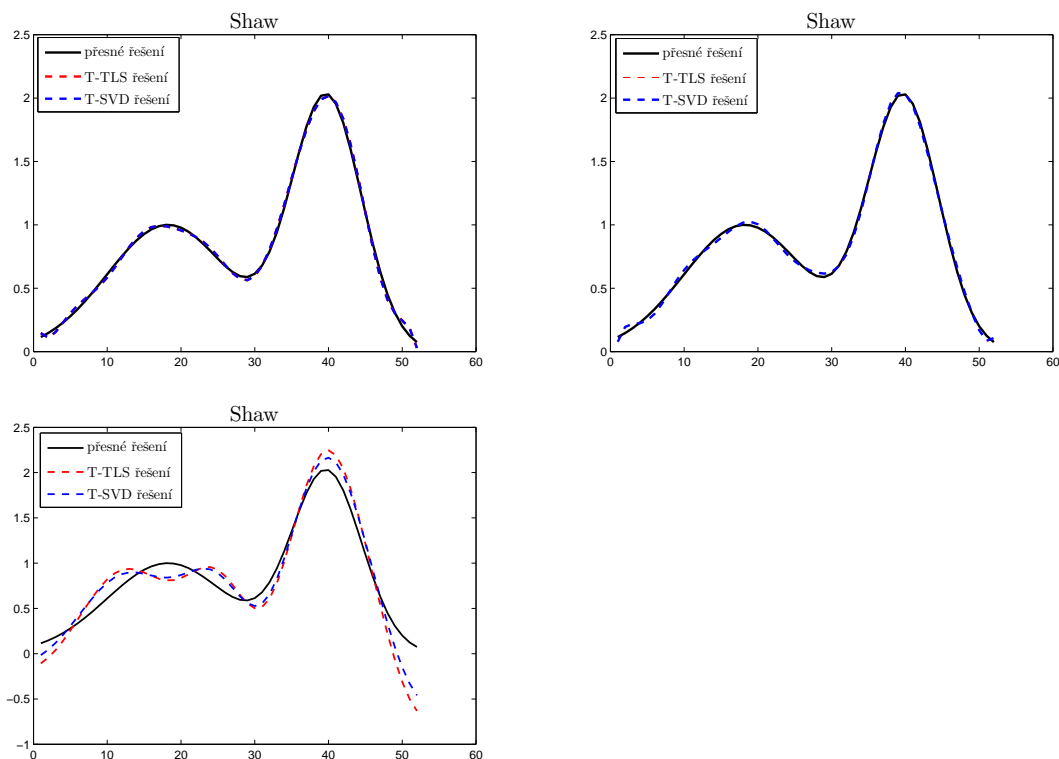


Obrázek 5.3: Singulární čísla matice Shaw (vlevo) a porovnání neregularizovaného řešení a T-TLS řešení se skutečným původním řešením (vpravo).

mezi velkými a malými singulárními čísly je hladší, ale zhoršování bude pozvolné, ne skokovité. To potvrzuje obrázek 5.4. Využijeme-li (4.11) můžeme se domnívat, že metody T-SVD a T-TLS budou poskytovat řešení, která si budou v případě vyšší hladiny šumu blíží než u matice Heat, neboť singulární čísla σ_i jsou zde lépe oddělena. To samozřejmě nemusí platit vždy. Zkušební hladiny budou opět 10^{-8} , 10^{-5} a 10^{-2} . Truncation level pro tyto hladiny byl postupně 14, 10, a 7. Při hodnotě šumu 10^{-8} již byly postřehnutelné odchylky od přesného řešení. Metody T-TLS a T-SVD poskytovaly podobná řešení.

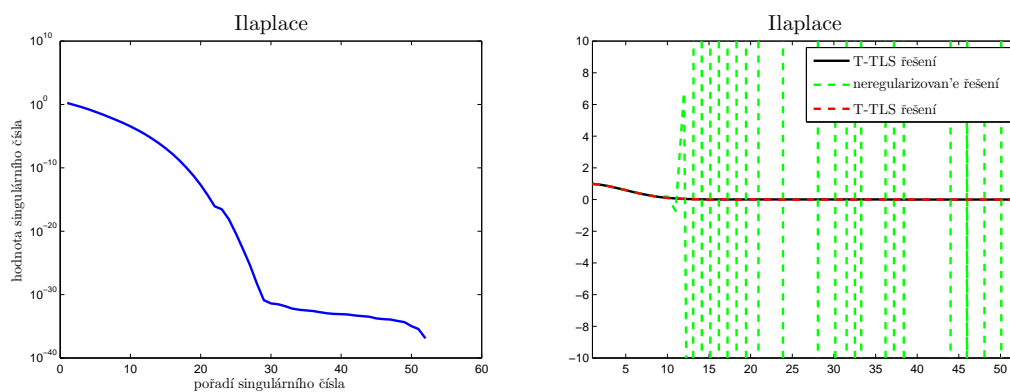
Experiment 3. V tomto experimentu se zaměříme na stejné otázky jako v experimentu 1, ale maticí soustavy bude matice Laplace z [7]. Rozložením singulárních čísel (viz obrázek 5.5 vlevo) se na první pohled podobá matici Shaw. Je třeba si však uvědomit, že u matice Laplace je rozdíl mezi σ_1 a σ_{27} (singulární číslo v „ohybu“) řádově větší, viz obrázky 5.3 a 5.5 vlevo. Je pravděpodobné, že i při malé perturbaci pravé strany (hladina šumu např. 10^{-20}), budou singulární čísla matice Laplace větší než σ_{27} „ořezána“ a truncation level bude menší než 27. Například při hladině šumu 10^{-12} byl truncation level pro metodu T-TLS roven 19. Srovnání T-TLS řešení s řešením metodou nejmenších čtverců a přesným řešením vidíme na obrázku 5.5 vpravo.

Protože singulární čísla $\sigma_1, \dots, \sigma_{26}$ klesají k nule postupně a nejsou dobře oddělena, lze podle (4.11) očekávat, že T-TLS a T-SVD řešení se budou lišit. Na druhé straně aproximační vlastnosti T-TLS a T-SVD řešení se nebudou

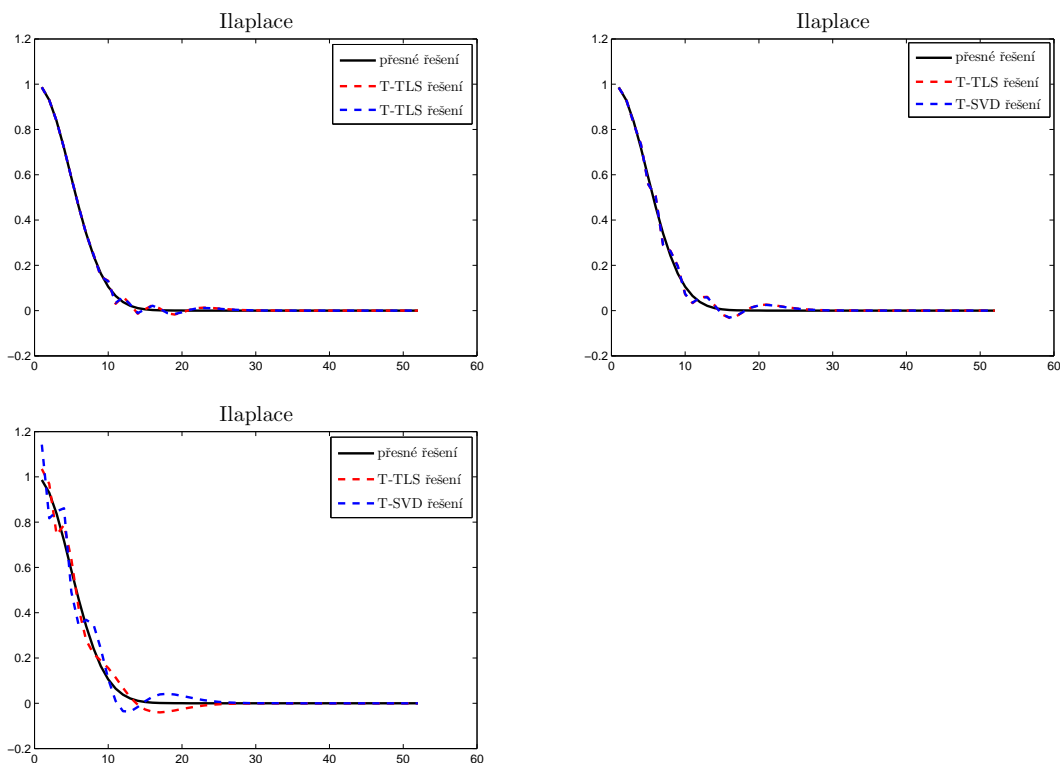


Obrázek 5.4: Porovnání metod T-TLS a T-SVD při řešení problému Shaw při hladině šumu 10^{-8} (vlevo nahoře), 10^{-5} (vpravo nahoře) a 10^{-2} (vlevo dole).

s rostoucí hladinou šumu příliš zhoršovat, protože mírná změna truncation levelu mezi hodnotami 1 a 27 nezpůsobí výraznější změnu aproximace. Jinými slovy nenastane situace, kdy po překročení optimálního truncation levelu (snížení) T-TLS a T-SVD řešení rychle ztrácejí aproximační schopnost (jako např. v experimentu 1). Pro hladiny šumu 10^{-8} , 10^{-5} a 10^{-2} určila L-křivka truncation levely postupně jako 16, 12 a 6. Řešení jsou vykreslena v obrázku 5.6.



Obrázek 5.5: Singulární čísla matice Π aplace (vlevo) a porovnání neregularizovaného řešení a T-TLS řešení se skutečným původním řešením (vpravo).



Obrázek 5.6: Porovnání metod T-TLS a T-SVD při řešení problému Shaw při hladině šumu 10^{-8} (vlevo nahoře), 10^{-5} (vpravo nahoře) a 10^{-2} (vlevo dole).

Závěr

V předložené práci jsme se seznámili s metodami nejmenších čtverců, vysvětlili jsme, proč neregularizované metody nejmenších čtverců mohou v případě ill-posed problémů poskytovat nepoužitelná řešení a jak lze tyto metody regularizovat nahrazením původního systému systémem nižší hodnosti. Zaměřili jsme se na regularizační vlastnosti těchto metod a na srovnání metody úplných nejmenších čtverců s omezením hodnosti s metodou truncated SVD. Jedna sekce práce byla dále věnována citlivosti metody úplných nejmenších čtverců s omezením hodnosti na perturbace matice soustavy (modelu) a pravé strany (vektoru pozorování). V závěrečné kapitole jsme demonstrovali výsledky teoretické analýzy sumarizované v předchozích kapitolách na numerických experimentech.

Literatura

- [1] A. Björck: *Numerical methods for total least squares*. SIAM second ed 1996.
- [2] G. Eckart, G. Young: The approximation of one matrix by another of lower rank. *Psychometrika* **1** (1936), str. 211–218.
- [3] R.D. Fierro, J.R. Bunch: Collinearity and total least squares. *SIAM J. Matrix Anal. Appl.* **15** (1994), str. 1167–1181.
- [4] R.D. Fierro, G.H. Golub, P.Ch. Hansen, D.P. O’Leary: Regularization by truncated total least squares. *SIAM J. Scient. Comp.* **18** (1997), str. 1223–1241.
- [5] G.H. Golub, Ch.F. Van Loan: *Matrix Computations*. Baltimore, The John Hopkins University Press 1984.
- [6] P.C. Hansen: *Rank-Deficient and Discrete Ill-Posed Problems – Numerical Aspects of Linear Inversion*. Philadelphia, SIAM 1997.
- [7] P. Ch. Hansen: Regularization Tools, a Matlab package for analysis of discrete regularization problems. *Numerical Algorithms* 6, 1994.
- [8] I. Hnětynková: Total Least Squares and Truncated Total Least Squares - v přípravě.
- [9] I. Hnětynková: *Krylov subspace approximations in linear algebraic problems* - doktorská práce. MFF UK (2006), str. 79–85.
- [10] F. Natterer: *The Mathematics of Computerized Tomography*. Philadelphia, SIAM Classics, 2001.
- [11] C.C. Paige, Z. Strakoš: Scaled total least squares fundamentals. *Numer. Math.* **91** (2002), str. 117–146.

- [12] Ch.C. Paige, Z. Strakoš: Core problem in linear algebraic systems. SIAM J. Matrix Anal. Appl. **27** (2006), str. 861–875
- [13] S. Van Huffel, J. Vandewalle: *The Total Least Squares Problem – Computational Aspects and Analysis*. Philadelphia, SIAM 1991.
- [14] K. Vasilík: *Zastavovací kritéria v ill-posed úlohách* - bakalářská práce. MFF UK (2008), str. 13–21.
- [15] M. Wei: The analysis for the total least squares problem with more than one solution. SIAM J. Matrix Anal. Appl. **13** (1992), str. 746–763.
- [16] M. Wei: Algebraic relations between the total least squares and least squares problems with more than one solution. Numer. Math. **62** (1992), str. 123–148.