

Úvod do řídkých reprezentací signálů a komprimovaného snímání

Autoři textu:

Mgr. Pavel Rajmic, Ph.D.

Ing. Marie Daňková

BRNO 2014



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Vznik těchto skript byl podpořen projektem č. CZ.1.07/2.2.00/28.0096
Evropského sociálního fondu a státním rozpočtem České republiky.

Autoři Mgr. Pavel Rajmic, Ph.D.
 Ing. Marie Daňková
Název Úvod do řídkých reprezentací signálů a komprimovaného snímání
Vydavatel Vysoké učení technické v Brně
 Fakulta elektrotechniky a komunikačních technologií
 Ústav telekomunikací
 Technická 12, 616 00 Brno
Vydání první
Rok vydání 2014
Náklad elektronicky
ISBN 978-80-214-5169-8

Tato publikace neprošla redakční ani jazykovou úpravou.

Kontakt: Pavel Rajmic, Ústav telekomunikací, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Technická 12, 616 00 Brno.
Tel. 541 146 937, email rajmic@feec.vutbr.cz

Obsah

Úvod	7
1 Značení a výchozí znalosti	9
1.1 Normy vektorů, řídkost	9
1.2 Normy matic	11
1.3 Pseudoinverze (Moore-Penroseova)	12
1.4 Bilineární a kvadratické formy	12
2 Báze a framy	15
2.1 Vektorový prostor konečné dimenze	15
2.1.1 Generátory vektorového prostoru	15
2.1.2 Báze vektorového prostoru	16
2.1.3 Framy	17
2.1.4 Dualita framů a bází	20
2.2 Vektorový prostor nekonečné dimenze	21
2.3 Kategorizace framů	27
3 Řídké reprezentace signálů	29
3.1 Řídkost signálů	29
3.2 Řídká řešení systémů lineárních rovnic	29
3.2.1 Jednoduché příklady	31
3.2.2 Postačující podmínky pro jednoznačnost řešení	32
3.2.3 Výpočetní složitost problému (P0)	35
3.3 ℓ_1 -relaxace	35
3.3.1 Podmínky ekvivalence řešení ℓ_0 - a ℓ_1 -minimalizace	35
3.3.2 Matice vyhovující RIP	40
3.4 Učení slovníku	40
4 Komprimované snímání	43
4.1 Klasické vzorkování	43
4.2 Komprimované snímání	44
4.3 Simulace	46
4.3.1 Vliv měřicí matice na úspěšnost ℓ_1 -rekonstrukce	47
4.3.2 Vliv délky nosiče a koherence mezi Φ a Ψ na míru úspěšnosti ℓ_1 -rekonstrukce	48
4.3.3 Přejít fází	51
4.4 Komprimované snímání založené na nízké hodnotě	52
5 Proximální algoritmy pro řešení konvexních optimalizačních úloh	53
5.1 Proximální algoritmy	53
5.1.1 Optimalizačních úloh s omezeními a bez omezení	54
5.2 Proximální operátor	55
5.2.1 Proximální operátor ℓ_1 -normy	57
5.2.2 Proximální operátor pro nukleární normu	58

5.3	Proximální gradientní metoda (dopředně-zpětné dělení)	59
-----	---	----

SEZNAM OBRÁZKŮ

1.1	Ilustrace hranic jednotkových koulí pro různé normy	10
1.2	Ilustrace účinku lineárních zobrazení na příkladu tří různých matic 3×3 . .	13
2.1	Šedesátčtyři prvků DCT ortogonální báze prostoru $\mathbb{R}^{8 \times 8}$	17
2.2	Příklad jednoduchého framu a jeho duálního framu	19
2.3	Mercedes-Benz frame.	19
2.4	Vlnka Daubechies 2 jakožto mateřská funkce ψ a funkce z ní odvozená dilatací a posunem	22
2.5	Ukázka několika atomů Gaborova framu složeného z posunutých a modulovaných B-splinů	23
2.6	Příklady vlnkových bazových poslopností v diskrétním případě	24
2.7	Několik prvků biortogonální báze pro $\mathbb{R}^{128 \times 128}$ s použitím vlnky <code>bior4.4</code> .	25
2.8	Příklad jednoho diskrétního „mateřského“ curveletu	25
2.9	Několik bazových funkcí contourletové transformace	26
2.10	Několik bazových funkcí biortogonální vlnky s označením <code>bior2.4</code>	26
2.11	Diagram rozdělení framů	27
3.1	Spektrogramy úryvků hudebních signálů	30
3.2	Řídkost obrazu ve vlnkové transformaci	30
3.3	Schéma nedourčeného systému rovnic $\mathbf{Ax} = \mathbf{y}$	31
3.4	Vizualizace soustav lineárních rovnic v \mathbb{R}^3 – nadroviny přípustných řešení a řídká řešení.	33
3.5	„Nafukující se“ koule v různých normách	36
3.6	Vizualizace norem: $\ell_{0,5}$, ℓ_1 , ℓ_2 a řešení dvou odlišných soustav (3.12) a (3.13)	37
4.1	Příklad matice \mathbf{R} , která vznikla z jednotkové matice $N \times N$ rovnoměrně-náhodným zachováním pouze m řádků	45
4.2	Ilustrace situace při komprimovaném snímání (bez přítomnosti šumu) . . .	45
4.3	Situace z obr. 4.2, ale se zdůrazněním procesu měření	45
4.4	Vliv použité měřicí matice na úspěšnost rekonstrukce signálu při použití signálu řídkého v bázi DCT.	47
4.5	Ukázka jedné realizace náhodného podvzorkování signálu délky $N = 256$, snímáno $m = 25$ vzorků. Signál je řídký ve frekvenční oblasti	48
4.6	Vliv použité měřicí matice na úspěšnost rekonstrukce signálu při použití signálu řídkého v bázi waveletu Daubechies 3.	49
4.7	Ukázka jedné realizace náhodného podvzorkování signálu a neúspěchu ℓ_1 -rekonstrukce v několika úsecích	49
4.8	Vliv délky nosiče na úspěšnost rekonstrukce po náhodném podvzorkování .	50
4.9	Srovnání atomů waveletu Daubechies 1, 4, 7 a 10 odpovídajících hloubce dekompozice $j = 3$. Je zde jasně patrná rostoucí délka nosiče.	51
4.10	Relativní četnost úspěšné rekonstrukce při použití matice s Bernoulliho rozdělením	51
5.1	Demonstrace účinku regularizačního parametru λ na rekonstrukci	55
5.2	Znázornění subdiferenciálu funkce	56
5.3	Funkce měkkého prahování	58
5.4	Průběh proximální gradientní metody při řešení úlohy (5.17). Nalezené minimum je vyznačeno jako červený puntík.	61

ÚVOD

Signály s tzv. *řídkou reprezentací* přitahují pozornost odborníků již bezmála třicet let, přičemž hlavní vlna zájmu přišla kolem přelomu tisíciletí. Pojem řídká reprezentace znamená, že daný signál je možné vyjádřit přesně či velmi dobře aproximovat lineární kombinací velmi malého počtu vektorů ze zvoleného reprezentačního systému.

Historické kořeny této tématické oblasti sahají do počátku 90. let 20. století k článkům [55, 16, 77]. Paradoxně působí, že poslední dva citované články byly publikovány autory, kteří pracovali ve stejném ústavu na Stanfordově univerzitě [32]. Zájem komunity zpracování signálů o tuto oblast v poslední dekádě zažívá srovnatelný boom jako způsobil v 80. a 90. letech objev vynikajících vlastností vlnkové transformace. Paralelně ovšem lze vystopovat druhou větev, kterou rozvíjejí zejména statistici a vědci z komunity strojového učení a dolování dat. Hlavní dosud známé výsledky z teorie řídkých reprezentací na straně zpracování signálů byly publikovány v letech 2000 až 2006, přičemž tahouny oblasti jsou vesměs kapacity, které jsou známé již z pionýrské činnosti ve výzkumu vlnkové transformace. Z literatury je zřejmé, že hlavní teoretické výsledky i první aplikace publikovali vesměs matematici, což lze jednoduše vysvětlit tím, že toto zaměření vyžaduje dobré znalosti hned v několika matematických disciplínách současně – prolínají se zde zejména lineární algebra, funkcionální analýza, teorie optimalizace a teorie pravděpodobnosti.

Aplikací je dnes již nepřeberné množství, s převahou v oboru zpracování obrazu (neboť přirozené obrazy mají obvykle velmi řídkou reprezentaci ve vlnkové transformaci). Obecně lze říci, že aplikace lze hledat všude tam, kde je možné signál řídkce vyjádřit ať už v nějaké „standardní“ bázi nebo v systému, který byl nalezen signálu „na míru“. Za nejatraktivnější aplikaci řídkých reprezentací lze s jistotou označit tzv. komprimované snímání (compressed sensing) [13, 9, 27, 38, 39, 33], které umožnilo obejít limity klasického shannonovského vzorkování a významně ovlivnilo i oblast teorie informace. Mezi hlavní aplikace zde patří biomedicína [53], nové metody radiolokace [36], optimalizace ekvalizace v bezdrátových OFDM kanálech [76], nebo korekční kódování v komunikačních technologiích [13].

Zájem o řídké reprezentace se v poslední době částečně přesunul také k tzv. nízkohodnostním reprezentacím. To znamená, že signál lze uspořádat do matice, která má nízkou hodnotu (přesně či přibližně) – takových signálů je překvapivě hodně. Aplikace tohoto zaměření směřují k doplňování chybějících údajů do datových polí a předpovědi volby zákazníka [12], ale i zpracování videa, robustní analýze hlavních komponent [11], hyperspektrálnímu zobrazování [40] či zrychlení snímání v magnetické rezonanci [61, 24]. Teoretický aparát však zůstává velmi podobný jako u řídkých reprezentací.

V kapitole 1 zavedeme značení a uvedeme některé teoretické pasáže, které v textu budou používány. Kapitola 2 pak připomíná některé vlastnosti bází vektorových prostorů, důležité pro pochopení následných konceptů, a rozšiřuje pole pozornosti o redundantní reprezentační systémy, tzv. framy, ve kterých je možné signály reprezentovat. Kapitola 3 pak seznamuje se základy řídkých reprezentací signálů, přičemž v kapitole 4 se věnujeme komprimovanému snímání na bázi řídkých a nízkohodnostních reprezentací. Kapitola 5 se věnuje zejména tzv. proximálním algoritmům pro nalezení požadovaných reprezentací/signálů.

Práce s tímto tématem na bakalářském stupni je z důvodů nároků na matematické

základy poněkud omezená, nikoliv však nemožná. Na FEKT VUT v Brně mohou být tato skripta použita jako doplňující studijní materiál pro kurzy BASS (Analýza signálů a soustav) a zejména BCZS (Číslicové zpracování signálů), a dále pro talentované studenty píšící bakalářskou práci na toto téma. Na magisterském stupni jsou možnosti již širší; kromě diplomových prací je možné užitek těchto skript očekávat pro studenty kurzů věnovaných zpracování signálů, obrazů – např. MCSI (Číslicové zpracování signálů), MZPR (Zpracování řeči), MCAS (Číslicové zpracování akustických signálů). Samozřejmě uplatnění leží v oblasti návazného doktorského studia.

Autoři děkují Vítězslavu Veselému, Radkovi Hrbáčkovi, Janu Špiříkovi, Václavovi Machovi a Zdeňkovi Průšovi za konzultace a pomoc s přípravou některých materiálů.

1 ZNAČENÍ A VÝCHOZÍ ZNALOSTI

Počet prvků množiny (tzv. kardinalitu) budeme značit stejně jako absolutní hodnotu; tj. např. $|\{-3, 2, 8, 9\}| = 4$.

Skalární veličiny budeme značit kurzívou, jako např. m, N . Komplexní sdružení čísla c bude označováno pruhem jako \bar{c} .

Vektory budeme značit tučným řezem, jako $\mathbf{x}, \mathbf{y}, \boldsymbol{\phi}$ atd. Konečněrozměrné vektory uvažujeme sloupcové, pokud nebude řečeno jinak. Předpokládáme indexování prvků vektorů počínaje jedničkou, tzn. $\mathbf{x} = [x_1, \dots, x_N]^\top$.

Definice 1.1. Nosičem vektoru \mathbf{x} myslíme množinu jeho indexů, v nichž má vektor nenulové hodnoty. Tuto množinu značíme $\text{supp}(\mathbf{x})$. Tedy $\text{supp}(\mathbf{x}) = \{i \mid x_i \neq 0\}$.

Např. pro signál $\mathbf{x} = [x_1, \dots, x_8]^\top = [0, 0, 3, 0, 4, 2, 0, 5]^\top$ máme $\text{supp}(\mathbf{x}) = \{3, 5, 6, 8\}$ a $|\text{supp}(\mathbf{x})| = 4$.

Matice značíme rovněž tučně, ale velkým písmenem, jako např. $\mathbf{A}, \boldsymbol{\Psi}$. Jejich prvky pak příslušnými malými písmeny, tj. $a_{i,j}, \psi_{i,j}$. Řádek i matice \mathbf{A} bude zastoupen symbolem $\mathbf{a}_{i,:}$, obdobně symbol $\mathbf{a}_{:,j}$ bude zastupovat sloupec j . Symbol $*$ značí u matic tzv. hermitovskou transpozici: \mathbf{A}^* je matice vzniklá složením transpozice matice a komplexního sdružení každého jejího prvku. Především platí $(\mathbf{AB})^* = \mathbf{B}^* \mathbf{A}^*$. Hodnost matice \mathbf{A} budeme značit $\text{rank } \mathbf{A}$ a jádro lineárního zobrazení určeného maticí \mathbf{A} (tj. podprostor tvořený všemi řešeními homogenní soustavy lineárních rovnic $\mathbf{Ax} = \mathbf{0}$) budeme značit $\text{ker } \mathbf{A}$.

Operátor $\text{vec}(\cdot)$ aplikovaný na matici má funkci přeskládání všech sloupců matice postupně pod sebe do jediného vektoru. Obdobně pokud je aplikován na vícerozměrný tenzor.

Vektorové prostory budeme označovat jako $\mathbb{V}, \mathbb{R}^N, \mathbb{C}^N$ atd. Všude budeme vektorové prostory uvažovat jako neprázdné. Systémy generátorů vektorových prostorů (tedy včetně bází) budou označovány tučnými verzálkami $\mathbf{B}, \mathbf{E}, \mathbf{F}$ atd., \mathbf{I} značí jednotkovou matici.

Na řadě obrázků je použita konvence zobrazování prvků matic a vektorů barvou podle jejich hodnoty (tzv. pseudobarevné schéma). Nejčastěji je použito schéma „jet“, kdy tmavě modrá barva reprezentuje nejnižší hodnotu, tmavě červená nejvyšší hodnotu a čísla mezi nimi jsou postupně znázorněny barevným přechodem mezi těmito dvěma barvami. Viz např. obr. 3.3 na straně 31.

1.1 Normy vektorů, řídkost

Norma vektoru je číslo, které vyjadřuje jeho „velikost“. V inženýrské praxi je zdaleka nejčastější normou norma eukleidovská, ale to není případ oblasti, které je věnována tato práce.

Definice 1.2. ℓ_p -norma vektoru $\mathbf{x} \in \mathbb{C}^N$ je definována jako

$$\begin{aligned} \|\mathbf{x}\|_p &:= \left(\sum_{i=1}^N |x_i|^p \right)^{1/p} \quad \text{pro } 1 \leq p < \infty, \\ \|\mathbf{x}\|_p &:= \sum_{i=1}^N |x_i|^p \quad \text{pro } 0 < p < 1, \\ \|\mathbf{x}\|_\infty &:= \max_i |x_i|, \\ \|\mathbf{x}\|_0 &:= |\text{supp}(\mathbf{x})|. \end{aligned} \tag{1.1}$$

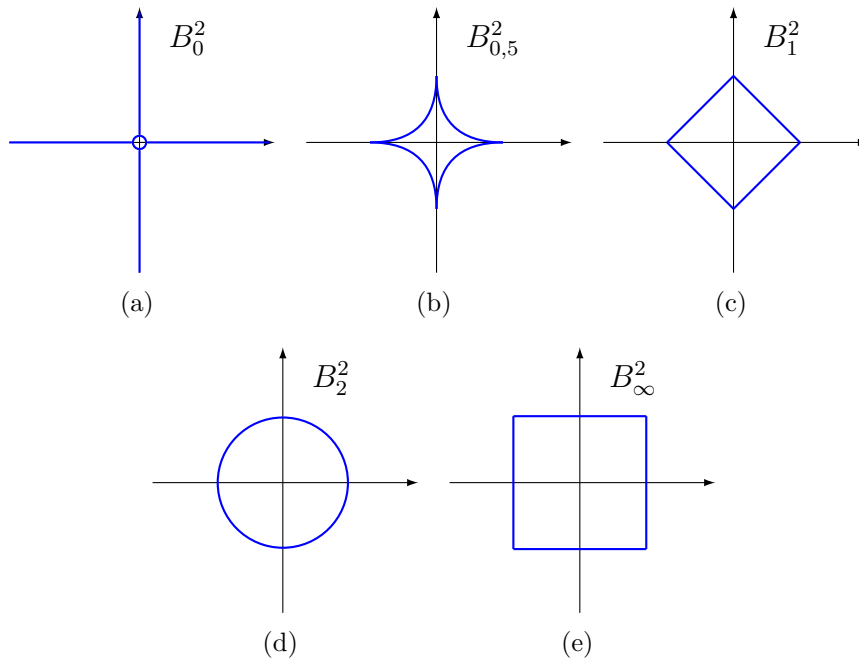
O normu se přísně vzato jedná pouze v případě $1 \leq p \leq \infty$. Pro zjednodušení však bude pro všechna p použito jednotně označení ℓ_p -norma.

Norma $\|\cdot\|_1$ představuje součet absolutních hodnot prvků vektoru (tj. $\|\mathbf{x}\|_1 = \sum_i |x_i|$) a norma $\|\cdot\|_0$ představuje počet nenulových složek vektoru. Pokud budeme používat zjednodušené označení $\|\cdot\|$, budeme tím mít na mysli eukleidovskou normu $\|\mathbf{x}\|_2 = \sqrt{\sum_i |x_i|^2}$. Abychom si dokázali lépe představit, jak se normy od sebe liší, zobrazíme jednotkové koule ve smyslu jednotlivých norem.

Definice 1.3. Jednotková koule B_p^N v normě ℓ_p je definována jako

$$B_p^N := \{ \mathbf{x} \in \mathbb{C}^N \mid \|\mathbf{x}\|_p \leq 1 \}. \tag{1.2}$$

Ilustrace jednotkových koulí v různých normách je na obrázku 1.1, jednotková koule v normě ℓ_0 kopíruje osy souřadného systému (s výjimkou počátku).



Obrázek 1.1: Ilustrace hranic jednotkových koulí (a) B_0^2 , (b) $B_{0,5}^2$, (c) B_1^2 , (d) B_2^2 a (e) B_∞^2 .

Definice 1.4. Vektor $\mathbf{x} \in \mathbb{C}^N$ nazveme k -řídkým (k -sparse), pokud platí

$$\|\mathbf{x}\|_0 \leq k. \quad (1.3)$$

Tedy k -řídký vektor je takový, který má nejvýše k nenulových složek. Relativní řídkostí vektoru \mathbf{x} délky N pak budeme rozumět poměr $\frac{k}{N}$. Dále označíme $\Sigma_k := \Sigma_k^N := \{\mathbf{x} \in \mathbb{C}^N \mid \|\mathbf{x}\|_0 \leq k\}$ množinu všech k -řídkých vektorů délky N .

Reálné signály ovšem nebývají řídké ve striktním smyslu, tak jak jsme definovali výše, ale místo nulových složek obsahují malé nenulové hodnoty. Proto je vhodné definovat chybu aproximace řídkým vektorem:

Definice 1.5. Chyba nejlepší aproximace vektoru $\mathbf{x} \in \mathbb{C}^N$ k -řídkým vektorem v normě ℓ_p (best k -term approximation error) je definována jako

$$\sigma_k(\mathbf{x})_p := \sigma_k^N(\mathbf{x})_p := \inf_{\mathbf{z} \in \Sigma_k^N} \|\mathbf{x} - \mathbf{z}\|_p. \quad (1.4)$$

Chyba tedy nemůže klesnout pod tuto hodnotu. Příklad: Pro $\mathbf{x} = [5, 2, 3]$ máme $\sigma_2(\mathbf{x})_0 = 1$, $\sigma_2(\mathbf{x})_1 = 2$ a $\sigma_2(\mathbf{x})_2 = 2$. Pokud speciálně máme k -řídký vektor \mathbf{x} , pak $\sigma_k(\mathbf{x})_p$ bude jistě nulové pro jakékoliv p .

Pro $T \subset \{1, \dots, N\}$ označíme $\mathbf{x}_T \in \mathbb{C}^N$ vektor odvozený z $\mathbf{x} \in \mathbb{C}^N$ tak, že prvky na pozicích patřících do množiny T zachováme a ostatní vynulujeme. Komplement T označíme $T^c = \{1, \dots, N\} \setminus T$.

Jednoduchou úvahou lze dojít k tomu, že $\sigma_k(\mathbf{x})_p$ lze také vyjádřit jako p -normu vektoru, který vznikne z \mathbf{x} odstraněním k složek s největší velikostí [38, 19]. Lze tedy psát také $\sigma_k(\mathbf{x})_p = \min_{T \subset \{1, \dots, N\}, |T| \leq k} \|\mathbf{x}_{T^c}\|_p$. Vzhledem k tomu, že určení chyby vždy závisí na konkrétním \mathbf{x} , lze to považovat za *adaptivní* záležitost.

1.2 Normy matic

Také na matice lze aplikovat normy. Nejjednodušší je matici vektorizovat a použít na výsledek normu pro vektory. Proto pokud v textu budeme aplikovat některou z vektorových norem na matici, bude se tím mít na mysli $\|\mathbf{A}\| := \|\text{vec } \mathbf{A}\|$. Předpokládáme $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$, případně $\mathbf{A} \in \mathbb{C}^{n_1 \times n_2}$, neboli matice o n_1 řádcích a n_2 sloupcích.

Definice 1.6. Frobeniova norma matice \mathbf{A} je definována jako $\|\mathbf{A}\|_F = \|\text{vec } \mathbf{A}\|_2 = \sqrt{\sum_i \sum_j |a_{ij}|^2}$.

Frobeniova norma je nejběžnější maticovou normou; jde o energii prvků matice \mathbf{A} . Existují také normy matic, které nelze vyjádřit použitím vektorových norem. Takovým příkladem normy je tzv. *nukleární norma* matice, známá též pod několika dalšími názvy. Pro její zavedení nyní uvažujme singulární rozklad (SVD) matice \mathbf{A} , známý z lineární algebry:

$$\mathbf{A} = \sum_{l=1}^n \sigma_l \mathbf{u}_l \mathbf{v}_l^* = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \quad (1.5)$$

kde $n = \min\{n_1, n_2\}$ je menší z rozměrů matice, dále $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ jsou singulární čísla matice \mathbf{A} , $\mathbf{\Sigma} = \text{diag}(\sigma_l)$ a $\mathbf{u}_l \in \mathbb{R}^{n_1}$, $\mathbf{v}_l \in \mathbb{R}^{n_2}$ jsou příslušné levé a pravé

singulární vektory, které tvoří sloupce unitární matice \mathbf{U} , resp. \mathbf{V} . Singulární čísla jsou odmocninami vlastních čísel symetrických a pozitivně semidefinitních matic $\mathbf{A}\mathbf{A}^*$ a $\mathbf{A}^*\mathbf{A}$. Je známým faktem, že \mathbf{A} má hodnotu $\text{rank } \mathbf{A} = r$ právě tehdy, když vektor $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{A})$ singulárních čísel je r -řádký, tj. platí $\text{rank } \mathbf{A} = \|\boldsymbol{\sigma}(\mathbf{A})\|_0$.

Definice 1.7. Nukleární normu *definujeme jako ℓ_1 -normu singulárních čísel, tj.*

$$\|\mathbf{A}\|_* := \|\boldsymbol{\sigma}(\mathbf{A})\|_1 := \sum_{l=1}^n \sigma_l. \quad (1.6)$$

Absolutní hodnotu u σ_l není třeba psát, neboť singulární čísla jsou vždy nezáporná.

Jako poslední uvedeme tzv. spektrální maticovou normu, která je definovaná pomocí eukleidovské (vektorové) normy.

Definice 1.8. Spektrální (operátorová) norma *matice je definována jako*

$$\|\mathbf{A}\| = \|\mathbf{A}\|_2 = \max_{\mathbf{x} \in \mathbb{C}^N, \mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}. \quad (1.7)$$

Lze ukázat, že $\|\mathbf{A}\|_2 = \max(\sigma_i) = \|\boldsymbol{\sigma}(\mathbf{A})\|_\infty$.

1.3 Pseudoinverze (Moore-Penroseova)

Definice 1.9. Matice pseudoinverzní \mathbf{A}^+ k matici \mathbf{A} je taková, pro kterou platí:

- $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$
- $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$
- $(\mathbf{A}\mathbf{A}^+)^* = \mathbf{A}\mathbf{A}^+$
- $(\mathbf{A}^+\mathbf{A})^* = \mathbf{A}^+\mathbf{A}$.

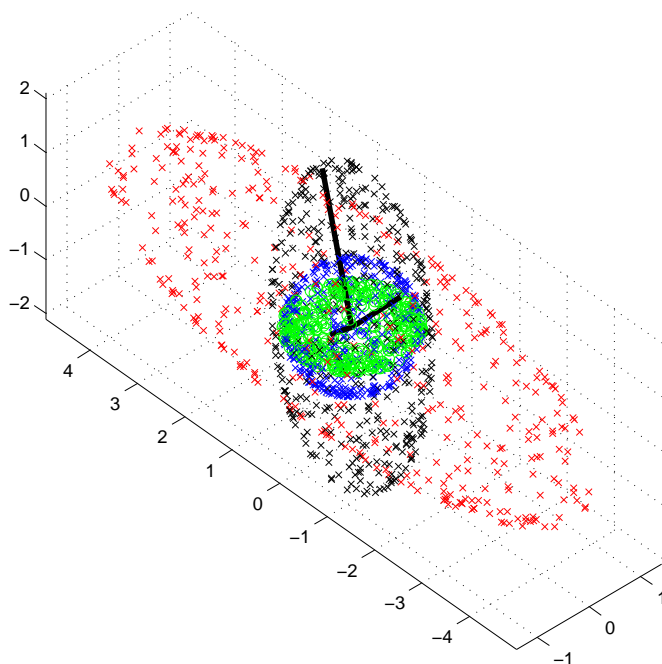
Tedy se jedná o zobecnění inverze matice na všechny matice: $\mathbf{A}\mathbf{A}^+\mathbf{x} = \mathbf{x}$. Zobecnění by mohlo být nekonečně mnoho [4], avšak jmenované podmínky zajišťují *jedinečnou volbu*: Řešíme-li systém lineárních rovnic $\mathbf{A}\mathbf{x} = \mathbf{b}$, který je nedourčený (méně lineárně nezávislých podmínek než prvků vektoru \mathbf{x}), pak $\mathbf{A}^+\mathbf{b}$ je mezi všemi řešeními to s nejnižší energií (nejmenší eukleidovskou normou); pokud je systém naopak přeурčený, řešení neexistuje, \mathbf{b} leží mimo prostor generovaný sloupci \mathbf{A} , a $\mathbf{A}\mathbf{A}^+\mathbf{b}$ má tu vlastnost, že má ze všech $\mathbf{A}\mathbf{x}$ nejbližší k \mathbf{b} ve smyslu eukleidovské normy. To odpovídá aproximaci při metodě nejmenších čtverců.

Především lze ukázat, že platí $\mathbf{A}^+ = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^+$, což v případě plné řádkové hodnosti \mathbf{A} přechází v

$$\mathbf{A}^+ = \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}. \quad (1.8)$$

1.4 Bilineární a kvadratické formy

Bilineární forma určená maticí \mathbf{A} je součin $\mathbf{y}^*\mathbf{A}\mathbf{x}$, kde rozměry součinitelů musejí být kompatibilní. Lze zapsat $\mathbf{y}^*\mathbf{A}\mathbf{x} = (\mathbf{A}^*\mathbf{y})^*\mathbf{x}$. Hermitovsky transponovaná matice \mathbf{A}^* je jediná s touto vlastností a nazývá se také (maticovým) *operátorem adjungovaným* k operátoru \mathbf{A} .



Obrázek 1.2: Ilustrace účinku lineárních zobrazení. Jedná se o aplikaci tří různých matic typu 3×3 , které realizují zobrazení $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Modrou barvou jsou vykresleny vektory na povrchu jednotkové koule. Červeně, černě a zeleně obarvené vektory jsou příslušné vektory po aplikování matic. Úsečkami jsou zvýrazněny hlavní poloosy černého elipsoidu.

Pokud je \mathbf{A} čtvercová, pak výraz $\mathbf{x}^* \mathbf{A} \mathbf{x}$ nazýváme *kvadratickou formou*. Je známo, že v případě symetrie $\mathbf{A} = \mathbf{A}^*$ jsou její vlastní čísla reálná. Uspořádáme-li je $\lambda_1 \leq \dots \leq \lambda_n$, tak platí

$$\lambda_1 = \min_i \lambda_i = \min_{\|\mathbf{x}\|=1} \mathbf{x}^* \mathbf{A} \mathbf{x} \leq \max_{\|\mathbf{x}\|=1} \mathbf{x}^* \mathbf{A} \mathbf{x} = \max_i \lambda_i = \lambda_n. \quad (1.9)$$

Když tedy bereme \mathbf{A} za matici transformace, pak (1.9) znamená, že jednotková kružnice ($\|\mathbf{x}\| = 1$) se při transformaci změní na (hyper)elipsoid s maximální poloosou dlouhou právě λ_n a minimální poloosou dlouhou právě λ_1 . Hlavní osy takového elipsoidu korespondují s vlastními vektory \mathbf{A} . Viz ilustrační obrázek 1.2.

2 BÁZE A FRAMY

Reprezentace signálů pomocí nedourčených systémů lineárních rovnic patří k moderním směrům ve zpracování signálů. V nedourčených systémech existuje více neznámých než lineárních vztahů mezi nimi, proto reprezentace signálu není jednoznačně určena, a tudíž je možné (a vhodné) hledat takovou reprezentaci, která je z nějakého pohledu optimální. Dnes je velká pozornost věnována algoritmům, které mezi všemi hledají tzv. *řídkou* reprezentaci signálu, tzn. takovou, že signál je v nějakém systému určen pouze několika málo prvky. Takovým reprezentacím je věnována i tato práce.

Účelem této kapitoly je ale nejprve ujasnit poznatky týkající se reprezentačních systémů. Připomeneme důležité pojmy (vektorový prostor, báze, skalární součin) a vlastnosti, ze kterých budeme vycházet. Pak budeme diskutovat redundantní reprezentační systémy – framy – a uvedeme několik jejich příkladů včetně konstrukce tzv. duálních systémů. Nakonec je uvedena kategorizace framů.

2.1 Vektorový prostor konečné dimenze

Vektorový prostor (VP) je algebraická struktura splňující známé axiomy [73]. Prvky vektorového prostoru jsou vektory. Dimenze vektorového prostoru bude v této části konečné číslo $0 < n \in \mathbb{N}$, v části 2.1 budeme pracovat s konečnými množinami vektorů. Zavedeme konvenci, že takovou množinu lze ztotožnit s maticí, která vznikne seřazením vektorů do jejích sloupců. Symbol pro obojí bude shodný.

2.1.1 Generátory vektorového prostoru

Systém generátorů vektorového prostoru \mathbb{V} je podmnožina vektorů \mathbf{E} ve \mathbb{V} , která jej generuje, čímž se rozumí, že každý vektor $\mathbf{x} \in \mathbb{V}$ je nějakou (zde konečnou) lineární kombinací generátorů. Generátorů \mathbb{V} může být i více než je dimenze n , a pak vektor $\mathbf{x} \in \mathbb{V}$ lze napsat jako vzájemně *různé* lineární kombinace, a tak jeden a tentýž vektor může mít více reprezentací. Tomu říkáme *nedourčenost*. Pokud \mathbf{x} lze v systému jeho generátorů $\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m\}$ zapsat jako

$$\mathbf{x} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \dots + c_m\mathbf{e}_m = \mathbf{E}\mathbf{c}, \quad (2.1)$$

pak skalárům c_i říkáme *souřadnice* \mathbf{x} v \mathbf{E} . Ve vztahu (2.1) se objevuje zápis $\mathbf{E}\mathbf{c}$, což je násobení vektoru $\mathbf{c} = [c_1, \dots, c_m]^\top$ maticí \mathbf{E} vzniklou seřazením sloupců \mathbf{e}_i vedle sebe. To koresponduje s nejběžnějším, tzv. aditivním, způsobem modelování signálu, kdy signál \mathbf{x} je považován za superpozici (lineární kombinaci) dílčích signálů:

$$\mathbf{x} = \sum_{i=1}^m c_i \mathbf{e}_i. \quad (2.2)$$

Zdůrazňujeme, že i -tý koeficient odpovídá i -tému sloupci matice \mathbf{E} . Tuto skutečnost využijeme v dalších kapitolách.

Jako příklady nejobvyklejších vektorových prostorů konečné dimenze uvedme n -tice reálných či komplexních čísel, \mathbb{R}^n a \mathbb{C}^n .

Vektorové prostory mohou být opatřeny normou (která nás informuje o „velikosti“ vektorů) a skalárním součinem (měří „úhel“ mezi dvěma vektory). Pak mluvíme o normovaných prostorech, resp. o prostorech s vnitřním součinem (tzv. unitární prostory) [73]. Často se norma zavádí tzv. indukovaně pomocí skalárního součinu: $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Prostor \mathbb{C}^N s vektory tvaru $\mathbf{x} = [x_1, \dots, x_N]^\top$ nejčastěji doplňujeme o skalární součin

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^N x_i \bar{y}_i = \mathbf{y}^* \mathbf{x} \quad (2.3)$$

a indukovanou normu

$$\|\mathbf{x}\|_2 := \|\mathbf{x}\| = \sqrt{\sum_{i=1}^N |x_i|^2} = \sqrt{\mathbf{x}^* \mathbf{x}}. \quad (2.4)$$

Pro vektorové prostory vektorů reálných čísel je v předchozích vztazích nadbytečné komplexní sdružení.

2.1.2 Báze vektorového prostoru

Bází vektorového prostoru se rozumí minimální systém jeho generátorů. Je to množina *lineárně nezávislých* vektorů, jejichž lineární kombinací lze dosáhnout na libovolný vektor v uvažovaném VP. V konečněrozměrném prostoru dimenze N je bází každá množina obsahující N lineárně nezávislých vektorů. Je-li $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_N\}$ bází, pak každý prvek $\mathbf{x} \in \mathbb{V}$ lze vyjádřit pomocí *jednoznačných* souřadnic c_i v oné bázi jako

$$\mathbf{x} = \sum_{i=1}^N c_i \mathbf{b}_i = \mathbf{B} \mathbf{c}. \quad (2.5)$$

Má-li jeden prostor více bází, potom všechny mají stejný počet generujících vektorů.

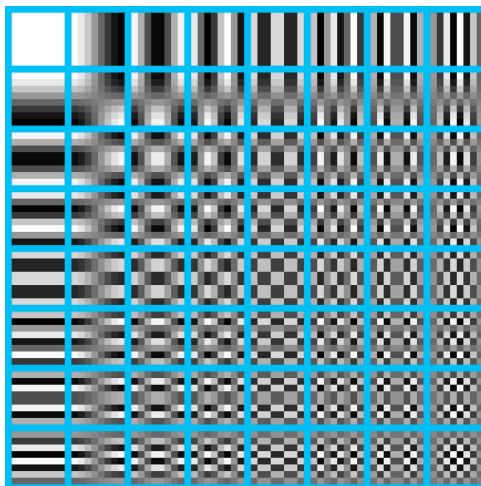
Ortogonální a ortonormální báze

Nejsnadněji se pracuje s bázemi *ortogonálními* a *ortonormálními*. Ortogonální báze je taková, kdy pro libovolné dva vektory z báze $\mathbf{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ platí

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = 0 \quad \text{pro } i \neq j, \quad \langle \mathbf{b}_i, \mathbf{b}_i \rangle \neq 0. \quad (2.6)$$

Tedy všechny dvojice bázových vektorů jsou na sebe kolmé. V ortonormální bázi navíc k (2.6) platí pro všechny její prvky $\|\mathbf{b}_i\| = 1$, neboli $\mathbf{B}^* = \mathbf{B}^{-1}$. Například, pokud definujeme $\mathbf{x}_1 = (1, 0)^\top$, $\mathbf{x}_2 = (0, 2)^\top$, pak tyto dva vektory tvoří ortogonální bázi v \mathbb{R}^2 . Vektory \mathbf{x}_1 a $\frac{1}{2}\mathbf{x}_2$ tvoří ortonormální bázi.

Dalším příkladem může být například prostor \mathbb{R}^{64} s bází založenou na DCT – diskrétní kosinové transformaci [68], kde je skalární součin definován dle (2.3). V takto definované bázi se vyjadřují bloky obrazu 8×8 pixelů, jako je tomu u obrazového komprimačního standardu JPEG [81], viz obr. 2.1.



Obrázek 2.1: Šedesátčtyři prvků DCT ortogonální báze prostoru $\mathbb{R}^{8 \times 8}$, obrázek převzat z [57].

Dvojice biortogonálních a biortonormálních bází

Mějme dvě různé báze $\mathbf{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ a $\mathbf{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_N\}$ ve stejném vektorovém prostoru. Říkáme, že jsou navzájem *biortogonální*, pokud splňují podmínku

$$\langle \mathbf{e}_i, \mathbf{f}_j \rangle = \begin{cases} 1 & \text{pro } i = j, \\ 0 & \text{pro } i \neq j \end{cases} \quad (2.7)$$

neboli $\mathbf{F}^* \mathbf{E} = \mathbf{I}$. Tedy není požadováno, aby dvojice vektorů z \mathbf{E} byly na sebe kolmé, a podobně pro \mathbf{F} , ale každý vektor z \mathbf{E} musí být kolmý na všechny vektory z \mathbf{F} vyjma jediného. Existují i báze *biortonormální*. Ty samozřejmě navíc musí splňovat podmínku $\|\mathbf{e}_i\| = \|\mathbf{f}_j\| = 1$ pro všechna i, j .

2.1.3 Framy

Bázi v konečněrozměrném prostoru \mathbb{V} tvoří lineárně nezávislé vektory, jejichž počet je roven dimenzi prostoru. Pokud je počet generátorů vyšší než dimenze, můžeme stále reprezentovat jakýkoliv vektor ve VP (konečným součtem typu (2.1)), generátory však již budou nutně lineárně závislé. Takovou množinu vektorů nazýváme *frame*, nebo přesněji (*konečný*) *frame* ve \mathbb{V} . Tento pojem poprvé definoval článek [30]. Uvedená nadbytečnost vyjádření může být pro reprezentaci signálu i výhodou. Framy jsou méně omezené než báze, a proto se používají pro jejich flexibilitu. Nevýhodou jsou náročnější výpočtové metody a riziko numerické nestability.

Nyní uvedeme definici framu, poté bude následovat několik technických poznámek a po něm geometrická interpretace definice, což je pro pochopení konceptu velmi důležité.

Definice 2.1. Množina (nejvýše spočetná) vektorů $\{\phi_k\}_{k \in \mathbb{I}}$ v prostoru \mathbb{V} v něm tvoří frame, pokud existují konstanty $0 < A \leq B < \infty$ takové, že platí

$$A \|\mathbf{x}\|^2 \leq \sum_{k \in \mathbb{I}} |\langle \mathbf{x}, \phi_k \rangle|^2 \leq B \|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathbb{V}. \quad (2.8)$$

Prvky framu ϕ_k se obvykle nazývají *atomy*. Konstanty A, B se nazývají *mezemi framu*. Optimální dolní mez pak definujeme jako supremum ze všech dolních mezí a optimální horní jako infimum z horních mezí.

Pokud systém $\{\phi_k\}_{k \in \mathbb{I}}$ generuje \mathbb{V} , nemusí to ještě nutně znamenat, že ve \mathbb{V} tvoří frame – např. systém $\left\{(1, 0), (0, 1), \left(0, \frac{1}{\sqrt{2}}\right), \left(0, \frac{1}{\sqrt{3}}\right), \dots\right\}$ jistě generuje \mathbb{R}^2 , ale konstanta B je nekonečná, protože výraz $\sum_{k \in \mathbb{I}} |\langle \mathbf{x}, \phi_k \rangle|^2 = |x_1|^2 + |x_2|^2 \cdot \sum_{k=1}^{\infty} \frac{1}{k}$ nekonverguje. Tedy frame není pouhým zobecněním báze, kde lineární nezávislost byla prostě vypuštěna.

Pokud ovšem uvažujeme v případě prostoru dimenze N množinu \mathbb{I} konečnou, $\mathbb{I} = \{1, \dots, m\}$, $m \geq N$, pak i B je konečné číslo, což lze snadno vyvodit použitím Cauchy-Schwarzovy nerovnosti [18, kap. 1]:

$$\sum_{k=1}^m |\langle \mathbf{x}, \phi_k \rangle|^2 \leq \sum_{k=1}^m \|\phi_k\|^2 \cdot \|\mathbf{x}\|^2, \quad \forall \mathbf{x} \in \mathbb{V}. \quad (2.9)$$

Co se týče dolní meze, $A > 0$ splňující (2.8) existuje, pokud $\{\phi_k\}_{k=1}^m$ generuje \mathbb{V} . Pokud by totiž nastala situace, že největší ze všech A je nulové, znamenalo by to, že ve \mathbb{V} existuje nějaké $\mathbf{x} \neq \mathbf{0}$ takové, že suma v (2.8) je nulová, tudíž \mathbf{x} by bylo kolmé na celý prostor generovaný $\{\phi_k\}_{k \in \mathbb{I}}$, což je spor. Lze dokázat, že to platí i obráceně.

Každá konečná množina vektorů je frame pro jeho lineární obal. Také platí, že v konečné dimenzi z každého framu lze vybrat bázi [17, 18].

Podívejme se nyní, co vyjadřuje výraz $\sum_k |\langle \mathbf{x}, \phi_k \rangle|^2$ v (2.8). Skalární součin $\langle \mathbf{x}, \phi_k \rangle$ lze přepsat jako $\phi_k^* \mathbf{x}$, a to znamená, že lze posloupnost $\{\langle \mathbf{x}, \phi_k \rangle\}_k$ zapsat jako $\Phi^* \mathbf{x}$. Matice Φ obsahuje ϕ_k jako své sloupce a adjungovaná matice Φ^* je jednoznačně určená. Na základě toho tedy zapíšeme

$$\sum_k |\langle \mathbf{x}, \phi_k \rangle|^2 = (\Phi^* \mathbf{x})^* \Phi^* \mathbf{x} = \mathbf{x}^* \Phi \Phi^* \mathbf{x}, \quad (2.10)$$

což je kvadratická forma určená symetrickou maticí $\mathbf{S} = \Phi \Phi^*$. Tak můžeme definici (2.8) přepsat do tvaru

$$A \leq \mathbf{x}^* \Phi \Phi^* \mathbf{x} \leq B \quad \text{pro všechna } \mathbf{x} \in \mathbb{V} \text{ taková, že } \|\mathbf{x}\| = 1. \quad (2.11)$$

Srovnáním se vztahem (1.9) ihned vidíme, že optimální dolní mez A je rovna nejmenšímu vlastnímu číslu $\Phi \Phi^*$ a optimální horní mez B tomu největšímu. Ekvivalentně můžeme psát

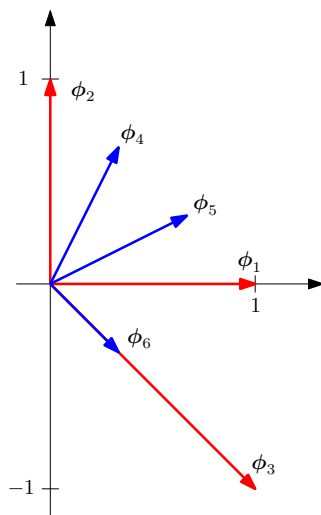
$$A \leq \frac{\|\Phi^* \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \leq B \quad \text{pro všechna nenulová } \mathbf{x} \in \mathbb{V}, \quad (2.12)$$

což interpretujeme tak, že operátor Φ^* deformuje eukleidovskou jednotkovou kouli ve \mathbb{V} na elipsoid s maximální poloosou délky \sqrt{A} a minimální poloosou délky \sqrt{B} . Viz také souvislost s operátorovou normou z definice 1.8.

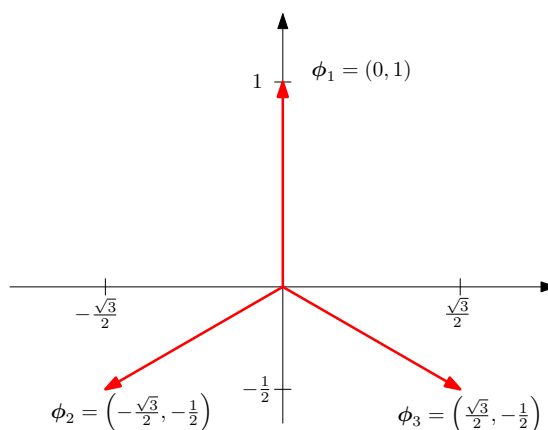
Framy, pro které platí $A = B$, nazýváme *těsné*. Táž rovnost platí i pro ortogonální báze se shodně velkými prvky, takže z tohoto pohledu jsou (možná překvapivě) rovnocenné, což mj. přináší jednu obrovskou výhodu oproti framům, které těsné nejsou. K tomu se vrátíme v části 2.1.4.

Na obr. 2.2 vidíme frame $\{\phi_1, \phi_2, \phi_3\}$ v \mathbb{R}^2 . Na první pohled je patrné, že vektory ϕ_1, ϕ_2 tvoří ortonormální bázi a vektor ϕ_3 je jejich lineární kombinací. Jako další příklad může sloužit tzv. Mercedes-Benz frame v \mathbb{R}^2 , který zapíšeme ve formě matice takto:

$$\Phi = \begin{pmatrix} 0 & -\sqrt{3}/2 & \sqrt{3}/2 \\ 1 & -1/2 & -1/2 \end{pmatrix}. \quad (2.13)$$



Obrázek 2.2: Příklad framu (červeně) a jeho duálního framu v \mathbb{R}^2 (modře).



Obrázek 2.3: Mercedes-Benz frame.

Je to těsný frame (viz výpočet v části 2.3) a jeho grafické znázornění je vidět na obr. 2.3.

Klasickým příkladem báze pro prostor \mathbb{C}^N posloupností (diskrétních signálů) délky N je fourierovská ortonormální báze $\{\mathbf{e}_k\}_{k=0}^{N-1}$, kde \mathbf{e}_k je signál (vektor) obsahující prvky

$$\frac{1}{\sqrt{N}} e^{j2\pi k\ell/N} \quad \text{pro } \ell = 0, \dots, N-1.$$

Lze ukázat [18, část 1.3], že pro $m > N$, přeparametrizovaný systém $\{\mathbf{f}_k\}_{k=0}^{m-1}$, kde \mathbf{f}_k odpovídá $\frac{1}{\sqrt{m}} e^{j2\pi k\ell/m}$ pro $\ell = 0, \dots, N-1$, tvoří v \mathbb{C}^N frame, který je dokonce těsný s $A = B = 1$. Z inženýrského pohledu to znamená, že se jedná o harmonické signály, jejichž kmitočty ovšem nemusí být celočíselnými násobky.

2.1.4 Dualita framů a bázi

Již uvedená biortogonalita úzce souvisí s konceptem tzv. duality vektorových prostorů [73]. Necht je dán frame $\mathbf{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ pro prostor \mathbb{V} dimenze $N < m$. Tedy jeden konkrétní vektor z \mathbb{V} může být vyjádřen nekonečným množstvím kombinací $\mathbf{x} = \sum_k c_k \mathbf{e}_k$, a to proto, že vektory framu jsou nutně lineárně závislé. Frame $\mathbf{F} = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, generující stejný prostor se nazývá *duální* k \mathbf{E} , pokud pro každé $\mathbf{x} \in \mathbb{V}$ platí, že

$$\mathbf{x} = \sum_k \langle \mathbf{x}, \mathbf{f}_k \rangle \mathbf{e}_k = \sum_k \langle \mathbf{x}, \mathbf{e}_k \rangle \mathbf{f}_k \quad (2.14)$$

neboli $\mathbf{x} = \mathbf{E}\mathbf{F}^*\mathbf{x} = \mathbf{F}\mathbf{E}^*\mathbf{x}$.

Podmínka (2.8) je nutná a postačující k tomu, aby duální frame existoval. Duálních framů k \mathbf{E} ale může být více.¹ Pouze jeden duální frame \mathbf{F} však má vlastnost, že posloupnost tzv. framových koeficientů (souřadnic) $\{c_k\}_{k \in \mathbb{I}}$ má mezi všemi posloupnostmi splňujícími $\mathbf{x} = \sum_k c_k \mathbf{f}_k$ minimální energii (tj. normu (2.4)) pro jakékoliv \mathbf{x} . Nazývá se *kanonický duální frame* k \mathbf{E} a lze ukázat, že je definován pomocí vztahu

$$\mathbf{f}_k = \mathbf{S}^{-1} \mathbf{e}_k \quad (2.15)$$

neboli $\mathbf{F} = \mathbf{S}^{-1}\mathbf{E}$, kde tzv. *framový operátor* \mathbf{S} je definovaný tak, že $\mathbf{S}\mathbf{x} = \sum_k \langle \mathbf{x}, \mathbf{e}_k \rangle \mathbf{e}_k$ [18, část 1.1]. V konečněrozměrném případě lze tento operátor ztotožnit s maticí $\mathbf{S} = \mathbf{E}\mathbf{E}^*$.

Ze vztahu (2.14) plyne, že souřadnice c_k vektoru v primárním framu hledáme pomocí duálního: $\langle \mathbf{x}, \mathbf{f}_k \rangle = c_k$, a naopak: $\langle \mathbf{x}, \mathbf{e}_k \rangle = d_k$, neboli $\mathbf{F}^*\mathbf{x} = \mathbf{c}$ a $\mathbf{E}^*\mathbf{x} = \mathbf{d}$. Toto je velmi důležité z hlediska zpracování signálu.

Příklad duálního framu vidíme na obrázku 2.2. Červené vektory ϕ_1, ϕ_2, ϕ_3 tvoří původní frame v \mathbb{R}^2 a modré vektory ϕ_4, ϕ_5, ϕ_6 tvoří k původnímu framu frame kanonický duální.

Kanonický duální frame je možné vytvořit za použití Moore-Penroseovy pseudoinverze [2]. Vyjdeme z vlastnosti (1.8):

$$\mathbf{E}^+ = \mathbf{E}^* (\mathbf{E}\mathbf{E}^*)^{-1}, \quad (2.16)$$

kde \mathbf{E} je původní frame – tedy matice plně řádkové hodnosti. Z definice framového operátoru můžeme rovnici (2.16) upravit na tvar

$$\mathbf{E}^+ = \mathbf{E}^* \mathbf{S}^{-1}. \quad (2.17)$$

Toto pouze dále upravíme pomocí hermitovské transpozice a dostaneme tak ekvivalenci s definicí (2.15):

$$(\mathbf{E}^+)^* = (\mathbf{E}^* \mathbf{S}^{-1})^* = \mathbf{S}^{-1} \mathbf{E} = \mathbf{F}. \quad (2.18)$$

Hledání duálního framu, tedy zde hledání inverze \mathbf{S} , může být často náročnou úlohou, ať už z důvodu špatné podmíněnosti matice nebo kvůli jejím velkým rozměrům. V případě *těsných* framů, pro které máme $A = B$, lze ukázat, že \mathbf{S} je diagonální matice (operátor), $\mathbf{S} = A\mathbf{I}$, čímž se situace podstatně zjednodušuje, neboť duální frame (je jediný) vznikne z primárního pouze přeškálováním: $\mathbf{f}_k = \mathbf{e}_k/A$. Proto se věnuje značné úsilí hledání právě těsných framů v různých prostorech.

¹ve skutečnosti jakmile existují dva, pak je jich nekonečně mnoho

Výše uvedené zákonitosti lze samozřejmě přenést i na případ $N = m$, tedy nechť \mathbf{E} je nyní bázi. Pak pro bázi duální \mathbf{F} platí (2.14), tj. $\{\mathbf{f}_k\} = \{\mathbf{S}^{-1}\mathbf{e}_k\}$ je jediná duální báze. Ta je zároveň biortogonální k \mathbf{E} , jak lze snadno ukázat. Zároveň z (2.7) vyplývá, že ortonormální báze je speciálním případem, kdy je biortogonální sama k sobě: platí totiž $\mathbf{E}^* = \mathbf{E}^{-1}$ a díky tomu $\mathbf{E}\mathbf{E}^* = \mathbf{I}$, a tedy $\mathbf{E} = \mathbf{F}$.

2.2 Vektorový prostor nekonečné dimenze

V aplikacích se sice obvykle pracuje jen v konečněrozměrných prostorech (ačkoliv vysoké dimenze), ale pro lepší souvislosti a upozornění na rozdíly na tomto místě uvádíme stručné pojednání o prostorech nekonečné dimenze.

Jakmile přejdeme do prostorů nekonečných (ať už spočetných nebo ne), platnost některých dosud samozřejmých a přirozených záležitostí pomíjí. Týká se to zejména otázek na konvergenci řad typu

$$\mathbf{x} = \sum_{k=-\infty}^{\infty} c_k \mathbf{f}_k, \quad (2.19)$$

přičemž aby vše fungovalo obdobně jako v konečné dimenzi, požadujeme nepodmíněnou (unconditional) konvergenci, kdy součet nezávisí na pořadí sčítanců. Konvergenci řady typu (2.19) přitom chápeme tak, že $\|\mathbf{x} - \sum_{k=-N}^N c_k \mathbf{f}_k\| \rightarrow 0$ pro $N \rightarrow \infty$. Vzhledem k tomu, že na některé vektory \mathbf{x} lze „dosáhnout“ pouze pomocí skutečně nekonečného počtu sčítanců v (2.19) (např. obdélníková funkce ve fourierovské bázi), do hry vstupuje také pojem uzávěr množiny (přidáním limit konvergentních posloupností).

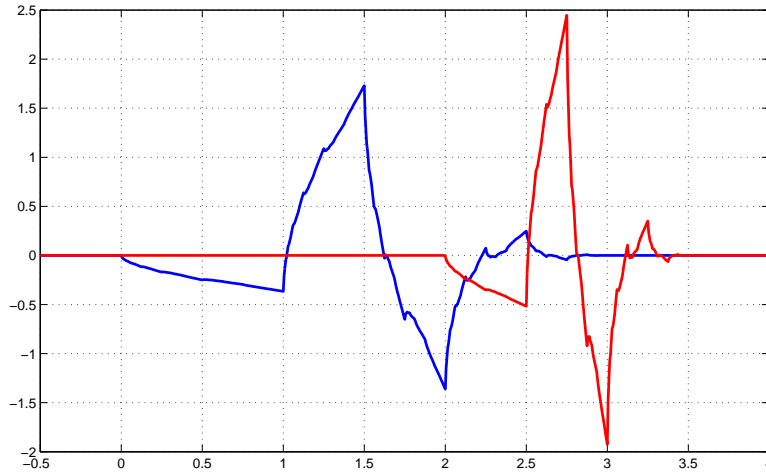
V souladu s většinou literatury budeme nadále pracovat se separabilním Hilbertovým prostorem, pro který vyhradíme symbol \mathcal{H} , a který má tyto vlastnosti:

- *unitárnost* – v prostoru je definovaný skalární součin, který na něm indukuje normu a z ní odvozenou metriku,
- *úplnost* – každá cauchyovská posloupnost vektorů je konvergentní k prvku z tohoto prostoru; limita „nevybíhá“ z prostoru ven,
- *separabilnost* – prostor obsahuje alespoň jednu nejvýše spočetnou podmnožinu, která je schopna generovat podprostor hustý v původním prostoru (jeho uzávěr je celé \mathcal{H}).

Pojem *báze* naráží na lineární nezávislost, kterou lze obtížně definovat pro nekonečně mnoho vektorů. Pojem lineární nezávislosti je proto zde potřeba přeformulovat: množina $\{\mathbf{f}_k\}_{k=-\infty}^{\infty}$ bude považována za lineárně nezávislou, pokud každá její konečná podmnožina bude lineárně nezávislá.

Byly definovány různé typy bází, nejčastěji se pracuje s tzv. Schauderovou, což je podmnožina vektorového prostoru taková, že každý vektor \mathbf{f} lze ve tvaru (2.19) vyjádřit pouze jediným způsobem [82]. Tzv. Rieszovu bázi v \mathcal{H} lze definovat tak, že vznikne jako transformace ortonormální báze (ta je vždy Schauderova) pomocí jakéhokoliv bijektivního ohraničeného operátoru, jehož inverze je též ohraničená [25], kde ohraničenost je ekvivalentní s jeho spojitostí.

Definice frame v nekonečné dimenzi je shodná s (2.8) až na to, že prvků v množině \mathbb{I} je nyní spočetně mnoho. Lze ukázat, že každá Rieszova báze \mathcal{H} tvoří v \mathcal{H} frame. V obecném případě však existence konečných $0 < A \leq B$ není zajištěna. Nerovnost $0 < A$ přitom znamená, že tzv. analyzující (diskretizační) operátor $T^*: \mathcal{H} \rightarrow \ell^2(\mathbb{Z})$, definovaný $T^*\mathbf{x} = \{\langle \mathbf{x}, \mathbf{f}_k \rangle\}_k$, je bijektivní se spojitou inverzí [18]. Nerovnost $B < \infty$ lze zase interpre-



Obrázek 2.4: Vlnka Daubechies 2 jakožto mateřská funkce ψ (modře) a funkce z ní odvozená dilatací s $a = 1/2$ a posunem s $b = 2$ (červeně). Jsou navzájem ortogonální.

toovat jako spojitost lineárního operátoru T^* , což je zde ekvivalentní s jeho ohraničeností. Poznamenejme, že operátor T^* nyní převzal roli matice Φ^* z textu výše.

Známým příkladem úplného separabilního Hilbertova prostoru nekonečné dimenze je prostor funkcí (analogových signálů) s konečnou energií $L^2(\mathbb{R})$ [73]. Definujeme jej jako

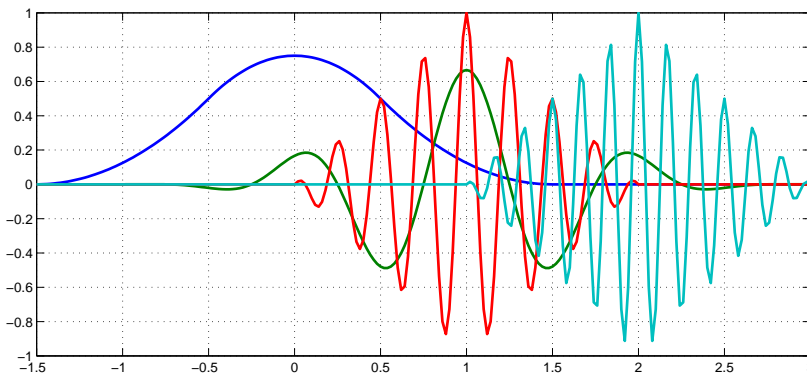
$$L^2(\mathbb{R}) = \left\{ f : \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dt} < \infty \right\}, \quad (2.20)$$

kde f je komplexní funkce reálné proměnné (přesněji ztotožňujeme funkce, které se rovnají skoro všude, protože jinak by uvedený integrál nebyl normou). Skalární součin funkcí (vektorů) f a g je definován jako $\langle f, g \rangle = \int_{-\infty}^{\infty} f(t)\overline{g(t)}dt$ a indukovaná norma $\|f\|_2 = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dt}$ je úměrná energii signálu.

Podprostor v $L^2(\mathbb{R})$ tvoří funkce, které mají omezené kmitočtové spektrum Fourierovy transformace. Lze ukázat, že pokud spektrum signálu f má nejvyšší kmitočť ω_{\max} , má tento prostor $PW_{\omega_{\max}}$ (tzv. Paley-Wienerův) ortonormální bázi $\{\text{sinc}(2\omega_{\max}x - k)\}_{k=-\infty}^{\infty}$, kde je definováno $\text{sinc}(x) = \sin(\pi x)/\pi x$ [18, část 3.8]. To koresponduje se notoricky známým postupem rekonstrukce signálu z jeho vzorků, pokud ty byly pořízeny tzv. ideálním vzorkováním s kmitočtem vyšším než $2\omega_{\max}$.

Systémy generátorů pro prostory nekonečné dimenze jsou často stavěny systematicky, na principu dilatace a translace. Pro $L^2(\mathbb{R})$ např. existují „mateřské“ funkce $\psi \in L^2(\mathbb{R})$, ze kterých jsou odvozeny ostatní funkce ve formě $|a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right)$. Parametr b koresponduje s posunutím v čase a a odpovídá změně měřítka. Tento koncept je nazýván mnohoměřítková analýza (MRA) [54], protože signál je možné vyjádřit jako součet dílčích signálů, které mají svůj základ v různých měřítkách, tak jak je to známé z vlnkové (waveletové) transformace, kde a, b jsou vhodně diskretizovány, aby vytvořily frame, biortogonální či dokonce ortonormální bázi [54, 25, 66] Příklad dvou ortogonálních bázevých vlnkových funkcí je na obrázku 2.4.

Obdobné systémy pro $L^2(\mathbb{R})$ existují v časově-kmitočtovém pojetí signálů, kde jsou prvky framu konstruovány za pomoci translace a modulace fixní funkce $g \in L^2(\mathbb{R})$. Soubor



Obrázek 2.5: Ukázka několika atomů Gaborova framu složeného z posunutých a modulovaných B-splinů. Zobrazena je reálná část. B-splínová okénková funkce g je modře.

takovýchto funkcí

$$\{e^{j2\pi mbx}g(x - na)\}_{m,n \in \mathbb{Z}} \quad (2.21)$$

pro zvolené dvě hodnoty a a b se nazývá *Gaborův (či gaborovský) systém*, podle D. Gabora, který se jako první věnoval jejich matematické analýze. V kontextu zpracování signálů se lze také často setkat s pojmem *krátkodobá Fourierova analýza* či *transformace*, pod zkratkou STFT.

Motivací pro zavedení systému tohoto typu je fakt, že klasická Fourierova transformace pracuje s harmonickými funkcemi, které mají globální dosah – jeden významný koeficient ovlivní podobu celého signálu. To však neodpovídá tomu, jak funguje lidské vnímání zvuků – zjednodušeně, člověk vnímá jejich kmitočtovou strukturu, ale tu dokáže zároveň umístit v čase [44]. Napodobit tento způsob potom přirozeně znamená analyzovat *časově lokální* spektra signálu. Odtud pramení pojem *časově-kmitočtová reprezentace*.

Základní otázkou je, jak zvolit funkci $g \in L^2(\mathbb{R})$ a parametry translace a modulace a, b tak, aby Gaborův systém (2.21) tvořil frame v prostoru $L^2(\mathbb{R})$. Touto problematikou se zabývala a zabývá řada prací [37] a dodnes není tato otázka vyčerpávajícím způsobem vyřešena. Zcela obecně platí pouze následující tvrzení:

Tvrzení 2.2 ([18]). *Nechť $a, b > 0$ a $g \in L^2(\mathbb{R})$. Gaborovský systém*

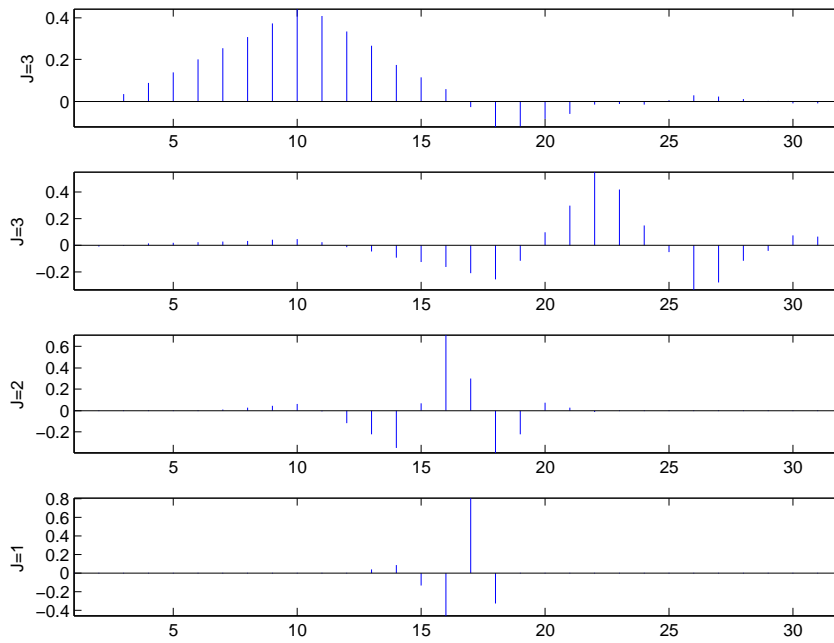
$$\{e^{j2\pi mbx}g(x - na)\}_{m,n \in \mathbb{Z}}$$

může být frame pro $L^2(\mathbb{R})$, pouze pokud platí $ab \leq 1$. Pokud je tento systém framem pro $L^2(\mathbb{R})$ a platí $ab = 1$, pak je tento systém dokonce Rieszovou bází.

Pokud je gaborovský systém frame, pak je možné každý signál rekonstruovat z jeho časově-kmitočtových koeficientů. Často se však Gaborova transformace používá pouze pro účely analýzy signálu. Gaborovy koeficienty tvoří mřížkovou strukturu, což umožňuje časově-kmitočtový obsah signálu vizualizovat v podobě tzv. *spektrogramu*, což je dvojrozměrné pole modulů Gaborových koeficientů a jako takové se obvykle zobrazuje formou obrázku s nepravými barvami, viz např. obr. 3.1.

Dalším příkladem nekonečněrozměrného prostoru je prostor T -periodických funkcí s konečnou energií na jedné periodě, kde klasická fourierovská teorie pracuje s ortogonální bází

$$\exp\left(jk\frac{2\pi t}{T}\right) \quad \text{pro } k \in \mathbb{Z}. \quad (2.22)$$



Obrázek 2.6: Příklady bázových posloupností v $\ell^2(\mathbb{Z})$. Jsou to posloupnosti vycházející z mnohoměřítkové analýzy (MRA) s použitím vlnky s označením `sym3`. V rámci pevného měřítka J (v obr. odpovídá jednomu podgrafu) jsou ortogonální všechna posunutí vyobrazených posloupností o $2^J k$, $k \in \mathbb{Z}$. Navíc každá taková posloupnost z jednoho měřítka je ortogonální vůči jakékoliv v jiném měřítku.

Signál lze rozvinout do Fourierovy řady $f(t) = \sum_{k \in \mathbb{Z}} c_k \exp(jk \frac{2\pi t}{T})$, kde koeficienty (souřadnice) c_k získáme (v teorii) „snadno“ díky ortogonalitě báze.

Dosud jsme mluvili o prostorech definovaných na reálné ose či jejím intervalu. Běžným příkladem Hilbertova prostoru spočetné dimenze je $\ell^2(\mathbb{Z})$. Jedná se o diskrétní signály (posloupnosti) s konečnou energií [73], kam spadají všechny signály

$$\mathbf{x} = [\dots, x_{-1}, x_0, x_1, x_2, \dots]^\top, \quad x_i \in \mathbb{C}, \quad \sum_{i=-\infty}^{\infty} |x_i|^2 < \infty \quad (2.23)$$

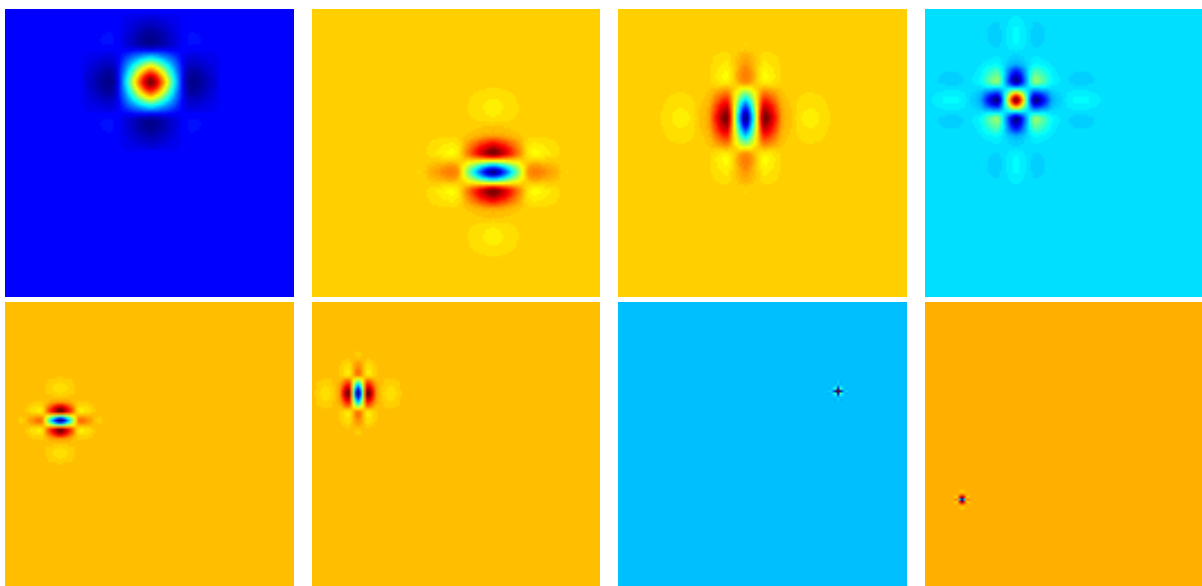
a kde skalární součin je definován vztahem

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=-\infty}^{\infty} x_i \bar{y}_i. \quad (2.24)$$

Tzv. kanonická báze v tomto prostoru je $[\dots, 0, 0, 1, 0, 0, \dots]^\top$, což je systém posouvaných Diracových impulzů. Složitějším příkladem ortonormální báze v $\ell^2(\mathbb{Z})$ jsou např. diskrétní podoby vlnek, viz obr. 2.6.

Diskrétní obrazové signály lze považovat za vektory z prostoru $\ell^2(\mathbb{Z} \times \mathbb{Z})$. Na obr. 2.7 je ukázka několika málo bázových vektorů tohoto Hilbertova prostoru, kde ovšem omezujeme nosič na velikost obrázku.

V oblasti zpracování obrazů bylo v poslední dekádě či dvou vyvinuto mnoho nových transformací, které často dostávají jména podle charakteru svých funkcí a povětšinou

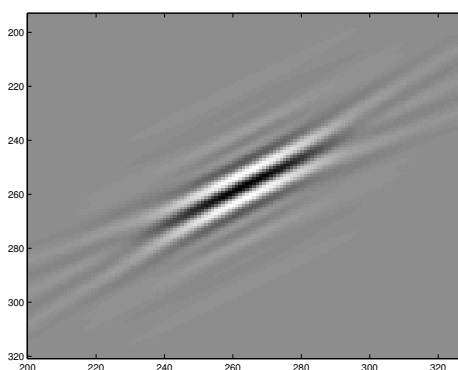


Obrázek 2.7: Několik prvků biortogonální báze pro $\mathbb{R}^{128 \times 128}$ (obrázek). Byla použita vlnka s označením `bior4.4`, jiným označením CDF9/7 (Cohen, Daubechies, Feauveau), která se využívá v kompresním formátu JPEG2000 [79] a jiných [74]. Pro zvýraznění průběhu je použita nepravá barevná škála „jet“ z MATLABu, pro každý obrázek zvlášť.

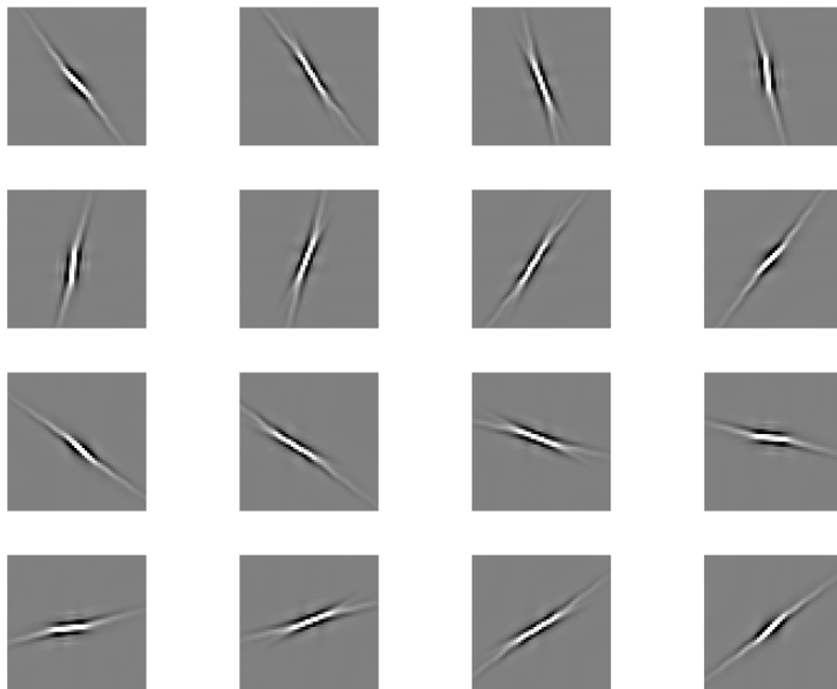
vycházejí z myšlenek vlnkové transformace. Je to například *diskrétní curveletová transformace* [7], která dokáže řídkce reprezentovat obrazy s výraznými hranami, viz příklad na obr. 2.8.

Další diskretní transformací pro reprezentaci obrazu pomocí těsných framů je *nepodvzorkovaná contourletová transformace* [22]. Tato transformace se zaměřuje především na reprezentaci hran (kontur) v obraze. Příklad jejích funkcí je na obr. 2.9.

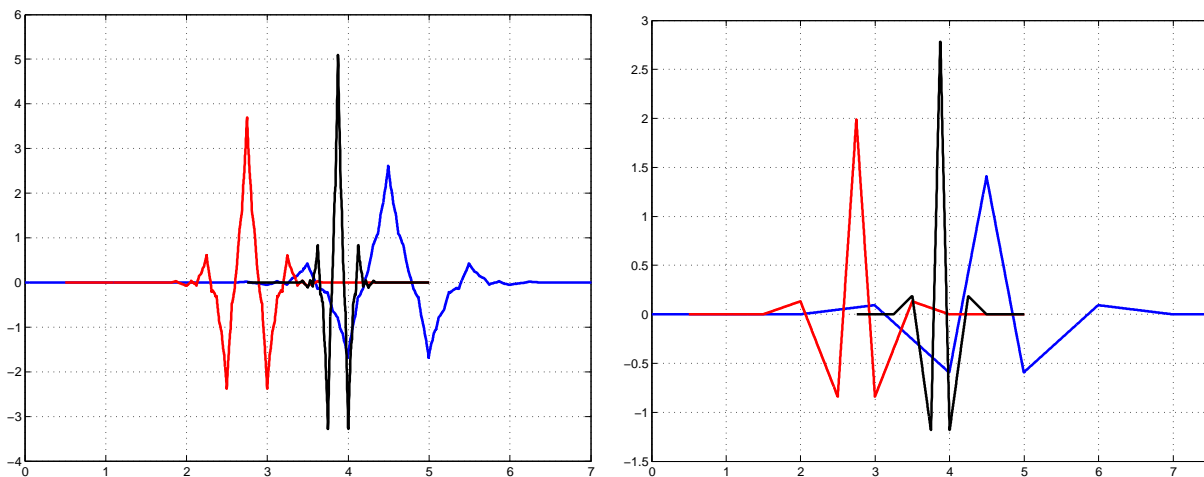
Jako třetí alternativu uvádíme transformaci *shearlet* [46, 52, 41], která se v poslední době těší největšímu zájmu pro jednoduchost konstrukce a vlastnosti pro zpracování obrazů.



Obrázek 2.8: Příklad jednoho diskretního „mateřského“ curveletu, převzato z [7].



Obrázek 2.9: Několik bázových funkcí contourletové transformace, převzato z [22].



Obrázek 2.10: Několik bázových funkcí biortogonální vlnky s označením `bior2.4`. Na prvním obrázku je několik vektorů vlnkové báze pro $L^2(\mathbb{R})$, jejíž prvky nejsou vzájemně ortogonální. Na druhém je jiná báze tohoto prostoru, rovněž není ortogonální. Avšak kterýkoliv vektor z báze v jednom obrázku je ortogonální ke každému vektoru z druhého, vyjma jediného, viz rovnici (2.7).

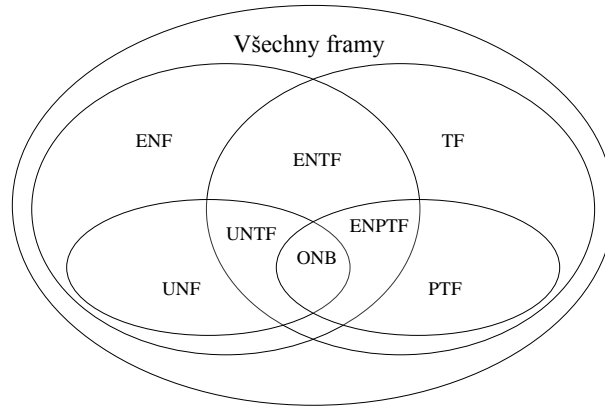
2.3 Kategorizace framů

Framy můžeme rozdělit do skupin podle jejich vlastností, jichž lze využít i při konkrétních aplikacích. Při rozdělení vycházíme především z definice (2.8).

Pokud $A = B$, pak takový frame nazýváme *těsný* (tight). Např. výše zmiňovaný Mercedes-Benz frame uvedený ve vztahu (2.13) je těsný a lze jednoduše spočítat, že jeho optimální meze jsou $A = B = 3/2$:

$$\begin{aligned} \sum_{k=1}^3 |\langle \mathbf{x}, \phi_k \rangle|^2 &= |x_2|^2 + \left| -\frac{\sqrt{3}}{2}x_1 \right|^2 + \left| -\frac{1}{2}x_2 \right|^2 + \left| \frac{\sqrt{3}}{2}x_1 \right|^2 + \\ &+ \left| -\frac{1}{2}x_2 \right|^2 = \frac{3}{2}(x_1^2 + x_2^2) = \frac{3}{2} \|\mathbf{x}\|^2. \end{aligned} \quad (2.25)$$

Zvláštním případem je pak tzv. 1-těsný frame, který se často nazývá *Parsevalův (těsný) frame*, kdy platí dokonce $A = B = 1$. Další skupinou jsou framy, pro něž platí, že všechny jejich prvky mají stejnou normu (Equal-norm frames – ENF), $\|\phi_i\| = \|\phi_j\|, i, j \in I$. Unitární framy (Unit-norm frames – UNF) jsou pak framy, v nichž mají všechny prvky normu rovnou 1, $\|\phi_i\| = 1, i \in I$. Viz obr. 2.11. Poznamenejme, že uvedená kategorizace je spíše základní, rozhodně není vyčerpávající [80, 17].



Obrázek 2.11: Diagram rozdělení framů. ENF: Equal norm frames, TF: Tight frames, ENTF: Equal-norm tight frames, UNF: Unit-norm frames, PTF: Parseval tight frames, UNTF: Unit-norm tight frames, ENPTF: Equal norm Parseval tight frames, ONB: Ortonormální báze [15].

3 ŘÍDKÉ REPREZENTACE SIGNÁLŮ

Mezi nekonečně mnoha řešeními nedourčených systémů lineárních rovnic hrají zvláštní roli ta, která mají co nejvíce neznámých současně nulových (tzv. řídká řešení), a to z důvodů usnadnění interpretace dat, snadné a silné komprese, množství dalších aplikací a rovněž z pohledu numerické stability.

Část 3.1 má za úkol motivovat předpoklad řídkosti u signálů. V části 3.2 formalizujeme základní úlohu a diskutujeme podmínky pro jednoznačnost řídkého řešení. Část 3.3 pak mluví o konvexní aproximaci základní úlohy.

3.1 Řídkost signálů

V této části uvedeme několik příkladů, které mají za úkol zdůvodnit použitelnost řídkých reprezentací v praxi.

Na obrázku 3.1 vidíme spektrogramy hudebních signálů. Je zřejmé, že převážná část gaborovských koeficientů je na úrovni -40 dB či méně, což jsou hodnoty téměř nulové či zanedbatelné vůči ostatním, významným koeficientům. Dále obrázek 3.2 ukazuje obrazový signál o velikosti 256×256 pixelů, který je rozložen na vlnkové koeficienty (ortogonální transformací), poté jsou koeficienty seřazeny podle velikosti. Pro účely rekonstrukce obrazu je ponecháno pouze 7% nejvýznamnějších koeficientů, zbytek je nahrazen nulami. Je evidentní, že rozdíl mezi originálem a rekonstrukcí je až překvapivě malý.

Podstatnou informaci o signálu tedy v uvedených ukázkách nese pouze zlomek koeficientů, a to platí u řady reálných signálů. Nekoresponduje to však exaktně s pojetím řídkosti, jak byla definována vztahem (1.3), tedy jako počet nenulových koeficientů. V našich příkladech je totiž většina signálových koeficientů sice zanedbatelných, nicméně nenulových. Z toho důvodu se pracuje také s pojmem *aproximativní řídkost* či *kompresibilita* signálu [19]. Vágní, ale postačující charakterizace takového signálu $\mathbf{y} \in \mathbb{R}^N$ je, že $\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon}$, přičemž $\|\boldsymbol{\epsilon}\|_2 \ll \|\mathbf{y}\|_2$ a $\|\mathbf{x}\|_0 \ll N$, tedy že signál lze až na malou chybu reprezentovat řídkou lineární kombinací sloupců \mathbf{A} .

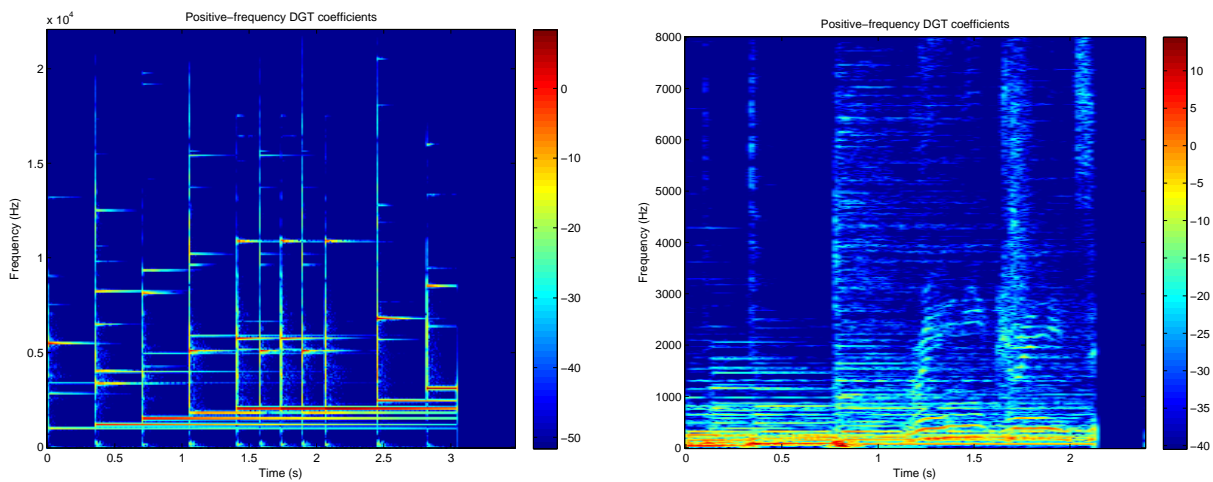
3.2 Řídká řešení systémů lineárních rovnic

Úkolem je řešit obvyklou soustavu lineárních rovnic $\mathbf{A}\mathbf{x} = \mathbf{y}$, ovšem s tím, že neznámý a hledaný vektor \mathbf{x} má být co nejřidší, tzn. má obsahovat pokud možno co největší počet nulových složek. Tedy jde o tuto úlohu:

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{vzhledem k} \quad \mathbf{A}\mathbf{x} = \mathbf{y}, \quad (\text{P0})$$

kde známe vektor $\mathbf{y} \in \mathbb{C}^m$ (pozorování, měření, signál) a matici $\mathbf{A} \in \mathbb{C}^{m \times N}$, která definuje lineární zobrazení. Zde předpokládáme pouze případy, kdy $m < N$, resp. $m \ll N$, a \mathbf{A} je plné hodnosti (řádkové). Ilustrativní schéma úlohy je na obr. 3.3.

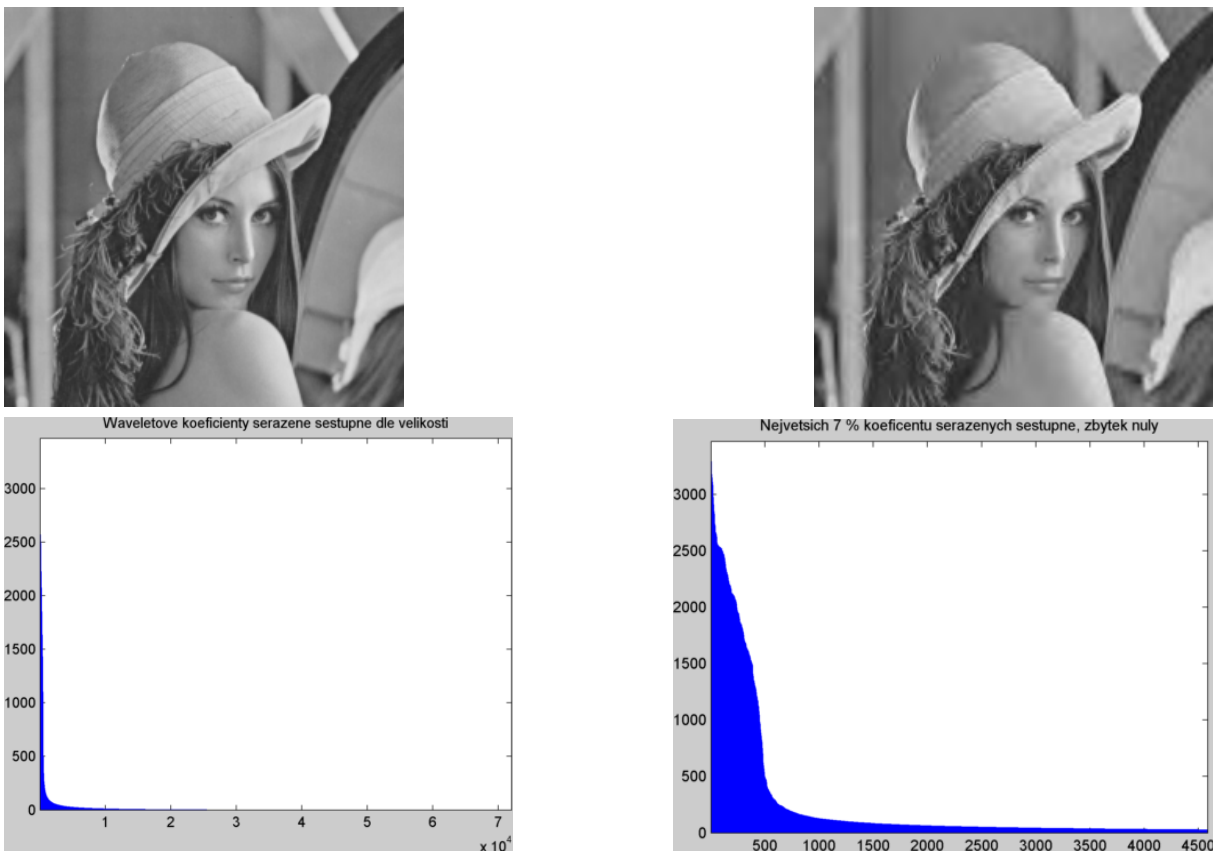
Matice \mathbf{A} se nejčastěji nazývá *slovník* (dictionary), přičemž se sloupce matice nazývají *atomy* (atoms). Můžeme se setkat i s názvem *reprezentační systém* nebo *měřicí matice* (measurement matrix), přičemž poslední z výrazů považujeme za zavádějící, což bude ještě ozřejmeno v kapitole 4 věnované komprimovanému snímání. Matice \mathbf{A} v této kapitole bude



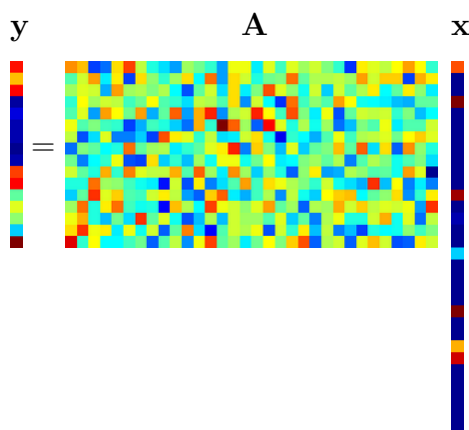
(a) Signál „Glockenspiel“ (zvonkohra). Zde je evidentní harmonická struktura signálu, jsou zde vidět i vyšší harmonické a nástup každého tónu.

(b) Signál poprockové hudby. Obsahuje výrazné nízké kmitočty, ale současně jsou patrné údery činelů a zpěvová linka.

Obrázek 3.1: Spektrogramy úryvků hudebních signálů. Vyobrazení jsou získána redundantní Gaborovou transformací. „Prázdný“ prostor na konci signálu je dán výpočetními nároky DGT na dělitelnost délky okna, počtu kanálů a délky signálu.



Obrázek 3.2: Řídkost obrazu ve vlnkové transformaci. Originál a jeho vlnkové koeficienty nalevo, oříznuté koeficienty a rekonstrukce z nich napravo.



Obrázek 3.3: Schéma nedourčeného systému rovnic $\mathbf{Ax} = \mathbf{y}$. Vektor \mathbf{x} obsahuje pouze několik nenulových prvků (souřadnic), a tedy pozorovaný vektor \mathbf{y} je lineární kombinací pouze několika málo sloupců matice \mathbf{A} .

obsahovat po sloupcích jednotlivé vektory z reprezentační báze nebo framu (viz kap. 2), a odpovídající vektor \mathbf{x} bude obsahovat souřadnice těchto vektorů pro signál \mathbf{y} .

Všechna \mathbf{x} , která splňují $\mathbf{Ax} = \mathbf{y}$, nazýváme přípustná řešení (přípustné reprezentace vektoru \mathbf{y}). Z lineární algebry je známo, že za výše uvedených podmínek na matici \mathbf{A} je přípustných řešení nekonečně mnoho a tvoří afinní prostor. V následujícím se budeme mj. zabývat požadavky kladenými na matici \mathbf{A} a na řídkost \mathbf{x} , které zaručí možnost „řídké rekonstrukce“ (sparse recovery) původního \mathbf{x} z \mathbf{y} .

Na tomto místě je vhodné zdůraznit, že požadavek řídkosti odpovídá pouze *počtu* nenulových prvků ve vektoru; nezahrnuje v sobě informaci o pozicích těchto prvků ani o žádných vztazích (např. korelacích) mezi nimi. Koncept tzv. *strukturované řídkosti* [50, 51] se snaží zahrnout tyto další předpoklady, často bližší praxi.

3.2.1 Jednoduché příklady

Mějme soustavu dvou lineárních rovnic

$$\begin{aligned} \frac{4}{5}x_1 + 2x_2 + x_3 &= 4 \\ \frac{7}{5}x_1 + \frac{7}{2}x_2 + x_3 &= 7, \end{aligned} \quad (3.1)$$

která je vizualizována na obrázku 3.4(a) (modrá resp. červená plocha). Řešením této soustavy je přímka v rovině $x_3 = 0$ určená rovnicí $x_2 = 2 - \frac{2}{5}x_1$ (znázorněna sytou červenou). Soustava má dvě řešení o řídkosti $k = 1$, která jsou vyznačena modrými body ($[0, 2, 0]$ a $[5, 0, 0]$).

Malou úpravou koeficientů v soustavě (3.1) získáme odlišnou soustavu, zobrazenou na obrázku 3.4(b):

$$\begin{aligned} 2x_1 + 2x_2 + x_3 &= 4 \\ \frac{7}{2}x_1 + \frac{7}{2}x_2 + x_3 &= 7. \end{aligned} \quad (3.2)$$

Tato soustava má opět nekonečně mnoho řešení (přímka $x_3 = 0$, $x_2 = 2 - x_1$), dvě z nich jsou 1-řádká ($[0, 2, 0]$ a $[2, 0, 0]$). Na první pohled se soustavy (3.2) a (3.1) liší jen zanedbatelně, v části 3.3.1 se však ukáže zásadní rozdíl, který mezi nimi je.

Poslední příklad soustavy rovnic je

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ \frac{1}{2}x_1 + \frac{1}{10}x_2 + x_3 &= \frac{1}{2}. \end{aligned} \quad (3.3)$$

Ta má pouze jediné řešení s řádkostí 1: $[1, 0, 0]$, viz obr. 3.4(c).

Je snadné si představit, že drobná změna koeficientů v (3.3) (na levé či pravé straně) povede k tomu, že budou existovat pouze 2-řádká a 3-řádká řešení. Tato úvaha přirozeně vede k aproximační úloze, kde splnění požadavku $\mathbf{Ax} = \mathbf{y}$ není striktní, ale povolí se malá odchylka:

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{vzhledem k} \quad \|\mathbf{Ax} - \mathbf{y}\|_p \leq \delta, \quad (\text{P0}\delta)$$

kde nejčastěji uvažujeme $p = 2$. Tento typ úlohy se obvykle objevuje v případech, kdy je signál pod vlivem šumu či chyb měření, protože to v signálu může způsobit odchýlení od přesného řešení.

3.2.2 Postačující podmínky pro jednoznačnost řešení

Do češtiny těžko přeložitelný pojem *spark* (doslova jiskra) je důležitou vlastností pro rozhodování o existenci a vlastnostech řešení.

Definice 3.1 ([28]). Číslo $\text{spark}(\mathbf{A})$ definujeme jako nejmenší počet sloupců matice \mathbf{A} , které jsou lineárně závislé. Formálně

$$\text{spark}(\mathbf{A}) = \min_{\substack{\mathbf{z} \in \ker \mathbf{A}, \\ \mathbf{z} \neq \mathbf{0}}} \|\mathbf{z}\|_0. \quad (3.4)$$

Kupříkladu, vezměme matici

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 3 & 3 \end{bmatrix}. \quad (3.5)$$

Pro ni platí $\text{spark}(\mathbf{A}_1) = 4$, zatímco pro

$$\mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 3 & 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 3 & 0 & 1 \end{bmatrix} \quad (3.6)$$

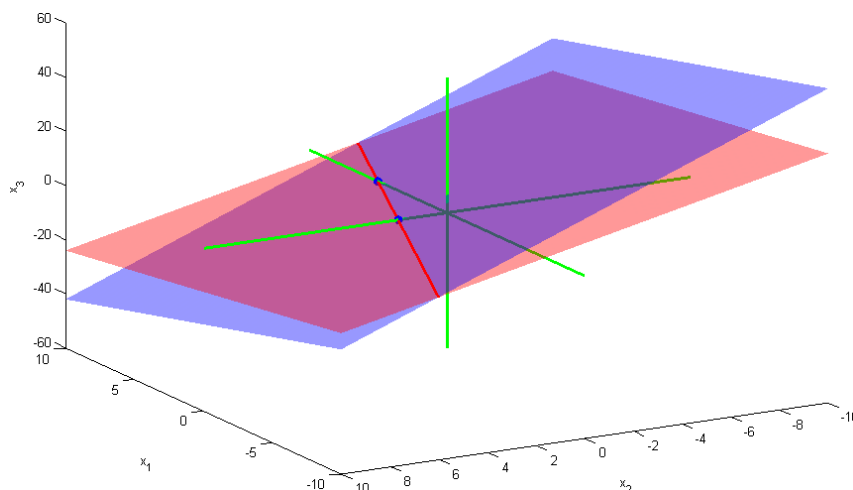
máme $\text{spark}(\mathbf{A}_2) = \text{spark}(\mathbf{A}_3) = 3$.

Pro nenulovou matici $\mathbf{A} \in \mathbb{C}^{m \times N}$, kde $m < N$, platí $\text{spark}(\mathbf{A}) \in \{2, \dots, m+1\}$, přičemž hodnoty 2 je dosaženo, když jeden sloupec je přímo násobkem jiného. Čím menší spark je, tím řidší musí vektor \mathbf{x} být, aby byla zajištěna jedinečnost tohoto řešení.

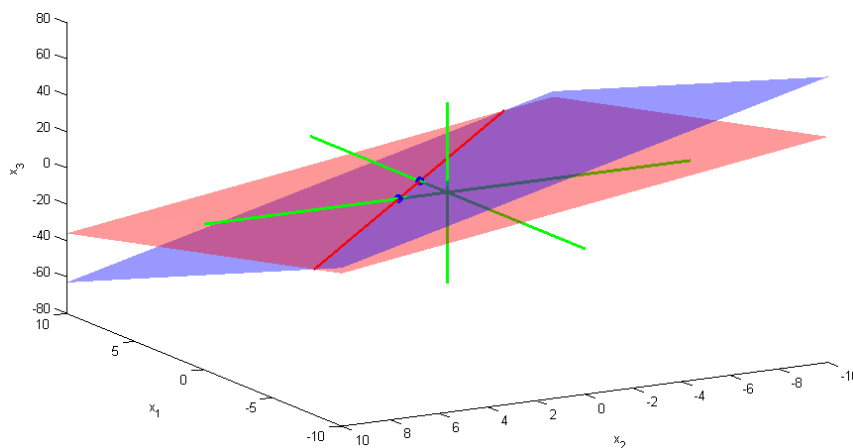
Tvrzení 3.2 ([6]). Pokud má soustava $\mathbf{Ax} = \mathbf{b}$ řešení \mathbf{x} splňující

$$\|\mathbf{x}\|_0 < \frac{\text{spark}(\mathbf{A})}{2}, \quad (3.7)$$

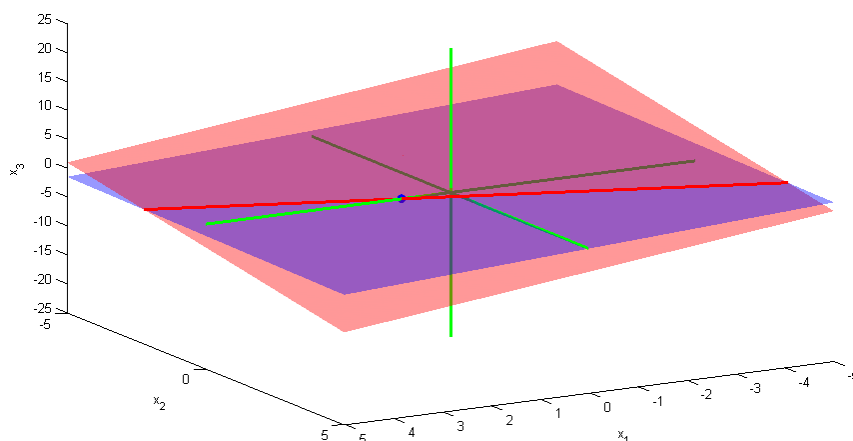
pak \mathbf{x} je nutně nejřidší možné řešení a žádné jiné řešení se stejnou řádkostí neexistuje.



(a) Vizualizace soustavy (3.1) se dvěma 1-řídkými řešeními. Přímka možných řešení se protíná s osami ve dvou bodech.



(b) Vizualizace soustavy (3.2) se dvěma 1-řídkými řešeními. Přímka možných řešení protíná osy ve dvou bodech, přičemž úhel, který svírá s osami x_1 a x_2 je $\pi/4$.



(c) Vizualizace soustavy (3.3) s jediným 1-řídkým řešením.

Obrázek 3.4: Vizualizace soustav lineárních rovnic v \mathbb{R}^3 – nadroviny přípustných řešení a řídká řešení.

Jaký je důsledek tohoto tvrzení? Pokud se podaří nalézt (jakýmkoliv způsobem!) řešení s kardinalitou vyhovující (3.7), pak víme, že zároveň bylo nalezeno řešení původního problému (P0). Bohužel nalezení $\text{spark}(\mathbf{A})$ je srovnatelně výpočetně náročné jako řešení problému (P0), proto je nutné hledat jednodušší způsob ověření jedinečnosti řešení. Navíc, předchozí tvrzení poskytuje pouze postačující podmínku, takže může existovat nejřidší řešení nesplňující (3.7).

Definice 3.3 ([6]). *Vzájemná koherence (mutual coherence) matice \mathbf{A} je definována jako největší absolutní normalizovaný skalární součin dvou různých sloupců matice \mathbf{A} ,*

$$\mu(\mathbf{A}) = \max_{\substack{1 \leq j, k \leq N, \\ j \neq k}} \frac{|\mathbf{a}_j^\top \mathbf{a}_k|}{\|\mathbf{a}_j\|_2 \cdot \|\mathbf{a}_k\|_2}, \quad (3.8)$$

kde \mathbf{a}_j označuje j -tý sloupec matice \mathbf{A} .

Pomocí vzájemné koherence můžeme zjistit „míru lineární závislosti“ mezi různými sloupci matice. Jedině unitární matice má koherenci nulovou, neboť její sloupce jsou po dvou ortogonální a tedy čítec v (3.8) je roven nule. Zde však uvažujeme pouze případy $m < N$ a pro matice $m \times N$ plné hodnosti vždy dle [38] platí

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(\mathbf{A}) \leq 1, \quad (3.9)$$

přičemž nalevo je rovnosti dosaženo pro tzv. *grassmanovské framy* [75]. V případě matic rozměru 100×200 tedy máme $\mu \geq 0,071$.

Tvrzení 3.4 ([6]). *Pro libovolnou matici \mathbf{A} platí*

$$\text{spark}(\mathbf{A}) \geq 1 + \frac{1}{\mu(\mathbf{A})}. \quad (3.10)$$

Ohraničením $\text{spark}(\mathbf{A})$ dostaneme tvrzení podobné 3.2:

Tvrzení 3.5 ([6]). *Pokud má soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ řešení \mathbf{x} splňující*

$$\|\mathbf{x}\|_0 < \frac{1}{2} \left(1 + \frac{1}{\mu(\mathbf{A})} \right), \quad (3.11)$$

pak \mathbf{x} je nutně nejřidší možné a je jediné takové. Navíc toto řešení lze dosáhnout ℓ_1 -minimalizací (ta je osvětlena v části 3.3).

Cenou za jednodušší výpočet je tedy silnější nerovnost. Blíží-li se koherence nule, pravá strana (3.11) roste nad všechny meze, avšak naopak pokud koherence se blíží jedné, pravá strana jde k jedné. Je tedy přirozené, že existuje snaha hledat a používat maximálně nekoherentní slovníky (tj. ty, které mají co nejnižší koherenci).

Jednoduše lze ukázat, že pokud $\mu(\mathbf{A}) > 1/3$, nerovnost (3.11) vnucuje podmínku pro řešení s kardinalitou pouze 1. Vraťme se k příkladu s grassmanovskými framy: pro ně dle (3.11) bude platit při volbě rozměru matice 100×200 , že nalezneme-li řešení \mathbf{x} kardinality 7 nebo méně, pak si můžeme být jisti, že jde o nejřidší možné. Matice uvedené výše mají vzájemnou koherenci přibližně $\mu(\mathbf{A}_1) = 0,785$, $\mu(\mathbf{A}_2) = 0,970$, $\mu(\mathbf{A}_3) = 0,971$.

3.2.3 Výpočetní složitost problému (P0)

Předpokládejme, že platí $\text{spark}(\mathbf{A}) > 2k_0$ a existuje nějaké k_0 -řídké řešení soustavy; tedy podle tvrzení 3.2 je to řešení nejřidší možné a jednoznačné. Pokud trváme na nalezení přesného řešení, nezbyvá než projít všech $\binom{N}{k_0}$ kombinací podmnožin sloupců matice. To odpovídá NP-složitosti, což v praxi, kde N je vysoké číslo, není přijatelné a výpočetní čas by se se současnou technikou počítal v týdnech, měsících či rocích [6]. Chceme-li výsledek v rozumném čase, nezbyvá než směnit rychlost za riziko úbytku přesnosti. Aproximativní metody můžeme zhruba rozdělit do dvou skupin podle principu, na kterém se zakládají, více viz kapitolu 5.

3.3 ℓ_1 relaxace

(Pseudo)norma ℓ_0 není konvexní funkce a tudíž není možné pro problém (P0) použít žádnou z řady metod a algoritmů konvexní optimalizace (konvexního programování), které jsou dnes k dispozici [5].

Protože ale normy ℓ_p jsou konvexní pro $p \geq 1$, viz obr. 1.1, nabízí se otázka, zda by nebylo možné využít „nejbližší“ konvexní normu, tj. ℓ_1 , k alespoň přibližnému, ale časově dosažitelnému řešení problému (P0). Tedy řešit úlohu

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \quad (\text{P1})$$

Odpověď je překvapivě příznivá, v některých případech se řešení úloh (P0) a (P1) shodují. Dokonce si ukážeme, že takových případů je vlastně „většina“. V přítomnosti zašuměných dat je odpovídající úlohou

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \delta, \quad (\text{P1}\delta)$$

tedy povoluje se odchylka vlivem šumu/nepřesností. Problém je znám pod zkratkou LASSO (Least Absolute Shrinkage and Selection Operator). Je vhodné upozornit, že původní úloha [77] byla formulována sice odlišně, nicméně ekvivalentně.

3.3.1 Podmínky ekvivalence řešení ℓ_0 - a ℓ_1 -minimalizace

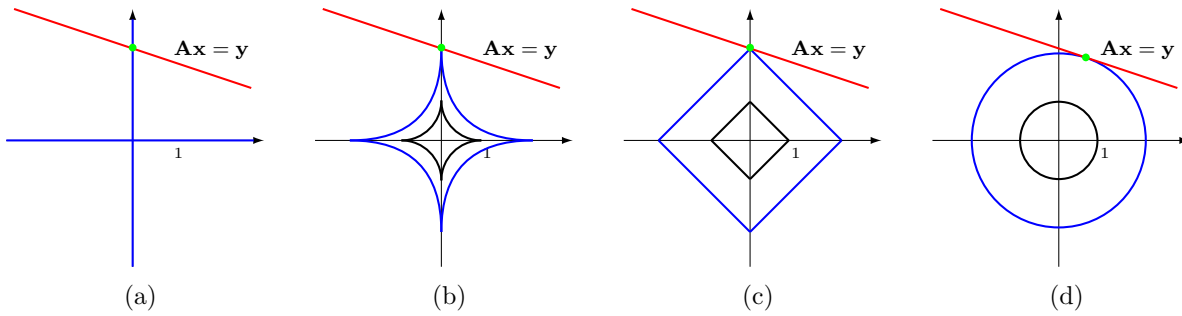
Obr. 3.5 ilustruje řešení úlohy $\min_{\mathbf{x}} \|\mathbf{x}\|_p$ vzhledem k $\mathbf{A}\mathbf{x} = \mathbf{y}$ v prostoru \mathbb{R}^2 , postupně pro $p = 0; 0,5; 1; 2$. Přípustná řešení reprezentuje červená přímka. Z obrázku je patrné, že řešení problémů (P0) a (P1) je shodné. Řešení v případě, kdy použijeme eukleidovskou normu, $p = 2$, je však odlišné (jedná se o řešení s minimální energií, což má úzkou spojitost s metodou nejmenších čtverců a Moore-Penroseovou pseudoinverzí).

Na vizualizacích v obrázcích 3.6(a) až 3.6(c) jsou vidět vždy dvě různé množiny přípustných řešení $\mathbf{A}\mathbf{x} = \mathbf{y}$ v prostoru \mathbb{R}^2 . Dvě matice soustav \mathbf{A} byly zvoleny tak, aby přípustná řešení byla charakterizována přímkami

$$x_1 - \frac{3}{10}x_2 = -30, \quad (3.12)$$

resp.

$$x_1 - x_2 = 50, \quad (3.13)$$



Obrázek 3.5: Vrstevnice norem (a) ℓ_0 , (b) $\ell_{0,5}$, (c) ℓ_1 a (d) ℓ_2 a jejich dotyk s nadrovinou určenou soustavou $\mathbf{Ax} = \mathbf{y}$. Koule v normě ℓ_0 zůstává vlastně stejná, neboť nezávisí na velikosti nenulových prvků.

přičemž se provádí postupně $\ell_{0,5}$ - , ℓ_1 - a ℓ_2 -minimalizace. Velikost normy pro vektory $[x_1, x_2]$ je zde vykreslena jako třetí souřadnice. Množiny přípustných řešení a průběh příslušných norem vidíme jako červenou (3.12), resp. modrou (3.13) čáru.

Obrázek 3.6(b) mimo jiné vysvětluje, kdy ℓ_1 -minimalizace může selhat v hledání řídkého řešení. Modrá přímka přípustných řešení (3.13) je zde rovnoběžná s hranou jednotkové kružnice normy ℓ_1 a tedy existuje nekonečně mnoho řešení s minimální normou, přičemž ovšem jenom ty „krajní“ jsou řídké.

Vizualizace 3.6(c) ilustruje normu ℓ_2 a stejné dvě úlohy. Nalezené řešení nebude nikdy řídké. Proto se ℓ_2 -minimalizace (která je výpočetně velice snadná) nehodí pro rekonstrukci řídkých signálů.

Vrátíme-li se v podobném duchu k systému (3.1), z jeho ilustrace na obr. 3.4(a) je patrné, že ℓ_1 -optimalizace zaručeně povede i k ℓ_0 minimu. Naopak 3.4(b) prezentující soustavu (3.2) ukazuje, že ℓ_1 -optimalizace může nalézt i neřídké řešení (záleží na použitém algoritmu).

Mnoho prací se zabývalo a stále zabývá nalezením podmínek, které zajišťují ekvivalenci problémů (P0) a (P1). Jako příklad uveďme, že v článku [43] je dokázáno, že výše uvedená podmínka na řídkost řešení (3.11) zaručuje dokonce, že takovéto řešení nalezneme ℓ_1 -minimalizací.

V dalším textu uvedeme několik dalších podmínek, které se vztahují na matici \mathbf{A} a na míru řídkosti vektoru \mathbf{x} .

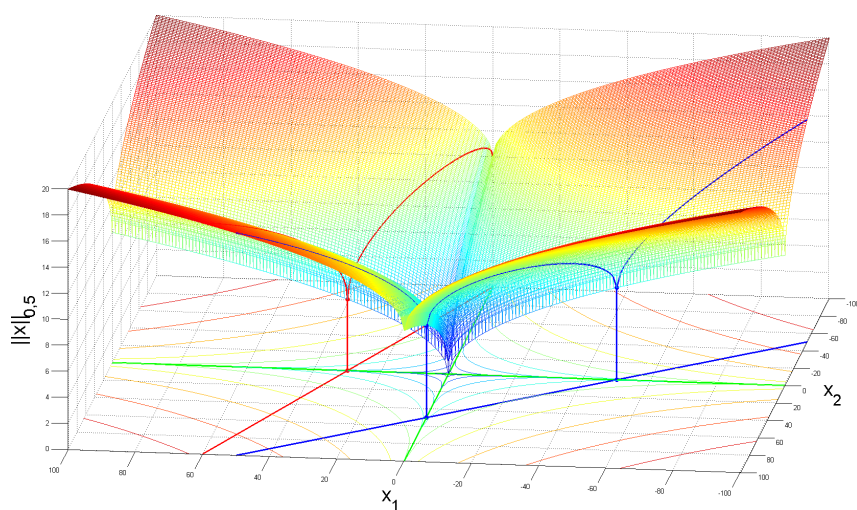
NSP – vlastnost nulového prostoru

Definice 3.6 ([38]). Řekneme, že matice $\mathbf{A} \in \mathbb{C}^{m \times N}$ má vlastnost nulového prostoru (Null Space Property, NSP) řádu k s konstantou $\gamma \in (0, 1)$, pokud platí

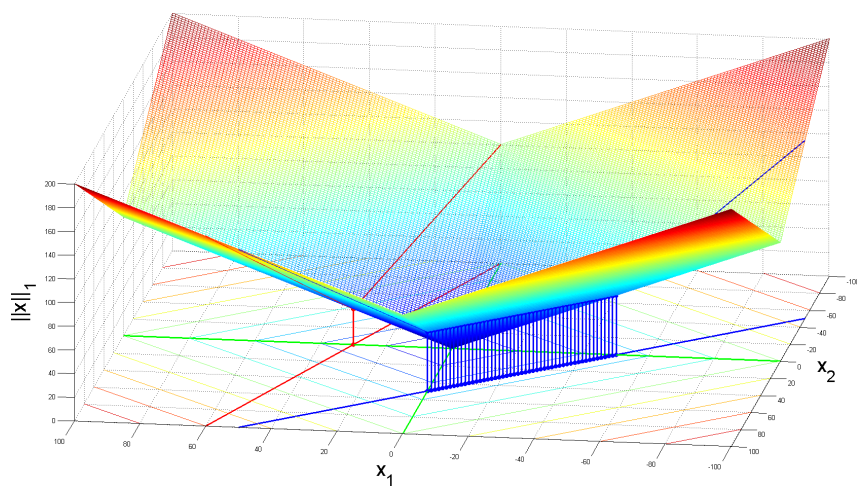
$$\|\boldsymbol{\eta}_T\|_1 \leq \gamma \|\boldsymbol{\eta}_{T^c}\|_1, \quad (3.14)$$

pro všechny množiny $T \subset \{1, \dots, N\}$, $|T| \leq k$ a pro všechny vektory $\boldsymbol{\eta} \in \ker \mathbf{A}$.

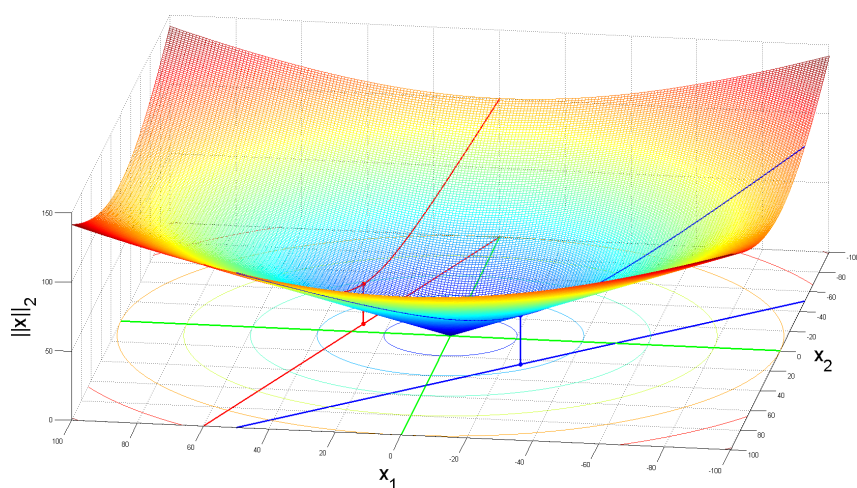
Splnění NSP zajišťuje pro libovolný vektor z jádra $\ker(\mathbf{A})$, že v něm bude ℓ_1 -norma koncentrována v „malém počtu“ prvků. Na první pohled je vidět, že ověření NSP není jednoduché.



(a)



(b)



(c)

Obrázek 3.6: Vizualizace norem: (a) $\ell_{0,5}$, (b) ℓ_1 , (c) ℓ_2 a řešení dvou odlišných soustav (3.12) a (3.13).

NSP zajišťuje, že k -řádké řešení je jednoznačné a je možné jej nalézt ℓ_1 -minimalizací [28, 43]. Následující tvrzení dává horní odhad chyby i v ostatních případech:

Tvrzení 3.7 ([38]). *Nechť matice $\mathbf{A} \in \mathbb{C}^{m \times N}$ splňuje NSP řádu k s konstantou $\gamma \in (0, 1)$. Nechť $\mathbf{x} \in \mathbb{C}^N$, $\mathbf{y} = \mathbf{A}\mathbf{x}$ a $\mathbf{x}^* \in \mathbb{C}^N$ je řešení ℓ_1 -minimalizace a necht $\sigma_k(\mathbf{x})_1$ je chyba nejlepší aproximace \mathbf{x} k -řádkým vektorem v ℓ_1 normě. Potom*

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq \frac{2(1 + \gamma)}{1 - \gamma} \sigma_k(\mathbf{x})_1. \quad (3.15)$$

Pokud za splnění předpokladů tvrzení existuje nějaké nejvýše k -řádké řešení soustavy $\mathbf{y} = \mathbf{A}\mathbf{x}$, pak $\sigma_k(\mathbf{x})_1 = 0$ a nula na pravé straně rovnice vynutí $\mathbf{x} = \mathbf{x}^*$, neboť $\|\mathbf{x} - \mathbf{x}^*\|_1 = 0$. To znamená, že ono řádké řešení nalezneme i ℓ_1 -optimalizací.

Zároveň také platí naopak, že pokud je možné ze soustavy $\mathbf{A}\mathbf{x} = \mathbf{y}$ rekonstruovat všechny k -řádké vektory \mathbf{x} pomocí ℓ_1 -minimalizace, pak matice \mathbf{A} splňuje NSP řádu k s nějakou (jinou) konstantou $\gamma \in (0, 1)$ [19]. NSP s větší z těchto dvou konstant je tak ekvivalentní řídké ℓ_1 -rekonstrukci [38].

RIP – vlastnost zeslabené isometrie

Vlastnost zeslabené isometrie (Restricted Isometry Property, RIP) nabízí oproti NSP výpočetně přijatelnější alternativu, která je navíc stabilní i pod vlivem šumu.

Definice 3.8 ([38]). *Konstanta omezené isometrie δ_k matice $\mathbf{A} \in \mathbb{C}^{m \times N}$ je nejmenší číslo takové, že platí*

$$(1 - \delta_k) \leq \frac{\|\mathbf{A}\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2} \leq (1 + \delta_k) \quad (3.16)$$

pro všechny nenulové vektory $\mathbf{z} \in \Sigma_k^N = \{\mathbf{x} \in \mathbb{C}^N \mid \|\mathbf{x}\|_0 \leq k\}$. Řekneme, že matice \mathbf{A} splňuje RIP řádu k s konstantou δ_k , pokud $\delta_k \in (0, 1)$.

Připomeňme, že isometrie znamená, že lineární zobrazení zachovává délku vektorů neboli je unitární. Zeslabení (restrikce) v případě (3.16) je dvojí: jednak se omezujeme pouze na všechny podmatice \mathbf{A} o nejvýše k sloupcích, a jednak nepožadujeme přesnou isometrii, ale povolujeme malou odchylku δ_k – jednoduše řečeno, všechny takové podmatice musejí být přibližně ortogonální.

Je nutné do definice RIP zahrnout *všechny* podmatice o maximálně k sloupcích, protože obecně dopředu nevíme, které prvky vektoru \mathbf{x} budou nenulové a tedy nevíme, které sloupce matice \mathbf{A} se podílejí na signálu \mathbf{y} .

Podobně jako u NSP lze určit i u RIP chybu aproximace ℓ_1 minimalizací. Stejně tak lze nalézt vztah mezi RIP a NSP.

Tvrzení 3.9 ([38]). *Nechť $\mathbf{A} \in \mathbb{C}^{m \times N}$ splňuje RIP řádu $K = k + h$ s konstantou $\delta_K \in (0, 1)$. Pak \mathbf{A} splňuje NSP řádu k s konstantou*

$$\gamma = \sqrt{\frac{k}{h} \cdot \frac{1 + \delta_K}{1 - \delta_K}}. \quad (3.17)$$

Zvolíme-li v tomto tvrzení $h = 2k$, pak nám vyjde, že RIP s konstantou δ_{3k} zaručuje splnění NSP řádu k s konstantou $\gamma = \frac{\sqrt{2}}{2} \sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}}$. Tudíž $\delta_{3k} < 1/3$ zaručuje $\gamma < 1$ a podle tvrzení 3.7 je v takovém případě pro všechny k -řídké vektory bezchybná rekonstrukce signálu pomocí ℓ_1 -technik zaručena.

Nyní sledujme jednoduchou úvahu [19, 14], kterou uvádíme před představením několika podmínek pro řešení, které využívají vlastnost RIP: Předpokládejme, že matice \mathbf{A} splňuje RIP řádu $2k$ s konstantou $\delta_{2k} \in (0, 1)$. Předpokládejme dále, že soustava $\mathbf{y} = \mathbf{A}\mathbf{x}$ má dvě k -řídká řešení, která označíme \mathbf{x}_1 a \mathbf{x}_2 . Je zřejmé, že vektor $(\mathbf{x}_1 - \mathbf{x}_2)$ má řídkost nejvýše $2k$ a tedy měl by vyhovovat nerovnici

$$(1 - \delta_{2k}) \leq \frac{\|\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)\|_2^2}{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2} \leq (1 + \delta_{2k}). \quad (3.18)$$

Dostáváme se však do sporu s faktem, že $\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$ (protože obě řešení musí vyhovovat původní rovnici), což celkově znamená, že řešení řídkosti k za uvedených podmínek musí být jediné.

To samo však ještě nezaručuje, že toto řešení bude možné nalézt „rozumnými“ prostředky, s přijatelnou výpočetní náročností a numerickou přesností. Následující tvrzení proto uvádějí silnější podmínky a ohraničují shora odchylku řešení.

Tvrzení 3.10 ([38]). *Nechť $\mathbf{A} \in \mathbb{C}^{m \times N}$ splňuje RIP řádu $3k$ s konstantou $\delta_{3k} < \frac{1}{3}$. Pro $\mathbf{x} \in \mathbb{C}^N$, necht $\mathbf{y} = \mathbf{A}\mathbf{x}$ a $\mathbf{x}^* \in \mathbb{C}^N$ je řešení ℓ_1 -minimalizace. Pak*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}}, \quad (3.19)$$

kde C je konstanta (kterou zde neuvádíme) závisající pouze na δ_{3k} .

Jiný odhad uvádí [14]:

Tvrzení 3.11. *Nechť \mathbf{A} splňuje RIP řádu $2k$ s konstantou $\delta_{2k} < \sqrt{2} - 1 \approx 0,4142$. Pak*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}} \quad (3.20)$$

a

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq C \sigma_k(\mathbf{x})_1 \quad (3.21)$$

pro nějakou konstantu C .

Následující tvrzení specifikuje vlastnosti řešení v případě šumu (šumem se zde myslí rušení signálu jakékoliv typu, není zde omezení pouze řečneme na bílý šum).

Tvrzení 3.12 ([38]). *Předpokládejme, že matice $\mathbf{A} \in \mathbb{C}^{m \times N}$ vyhovuje RIP řádu $2k$ s konstantou*

$$\delta_{2k} < \frac{2}{3 + \sqrt{7/4}} \approx 0,4627.$$

Následující platí pro všechna $\mathbf{x} \in \mathbb{C}^N$. Necht měření jsou zatížena šumem: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, $\|\mathbf{e}\|_2 \leq \epsilon$, \mathbf{x}^ je řešením úlohy*

$$\arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{vzhledem k} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \epsilon. \quad (3.22)$$

Potom

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_1 \frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}} + C_2 \epsilon \quad (3.23)$$

pro kladné konstanty C_1, C_2 závislé pouze na δ_{2k} .

Stojí za povšimnutí, že celková chyba je velmi prostě rozdělena na část, jež závisí na řídkosti, a část podmíněnou energií rušení. V článku [14] lze najít příklad: v případě $\delta_{2k} = 1/4$ jsou tyto konstanty $C_1 \leq 5,5$ a $C_2 \leq 6$.

3.3.2 Matice vyhovující RIP

Z předchozího víme, jaké požadavky jsou kladeny na matici \mathbf{A} , aby byla zaručena řídká řešitelnost úlohy. V této chvíli by nás mohlo zajímat, jak vypadají konkrétní matice, které vyhovují RIP.

V prvé řadě, dosud se nikomu nepodařilo zkonstruovat deterministické matice, které by splňovaly RIP s předem definovanými parametry. Jinak je to ale s maticemi, které vznikají za přispění náhody. Lze ukázat, že matice splňující RIP jsou s vysokou pravděpodobností např. ty, které vzniknou náhodným výběrem m řádků, přičemž

$$m \geq C \cdot k \cdot \ln(N/k), \quad (3.24)$$

z těchto $N \times N$ matic [14], kde C je opět vhodná konstanta:

- *Gaussovske matice*, tj. matice, jejichž prvky jsou nezávisle generovány z normálního rozdělení se střední hodnotou 0 a rozptylem $1/m$. O těchto maticích víme, že splňují RIP řádu k s konstantou δ s pravděpodobností $1 - \vartheta$, když se volí $m \geq c \cdot \delta^{-2}(k \ln(N/k) + \ln(1/\vartheta))$.
- *Bernoullijske matice*, kde každý prvek je náhodnou veličinou nabývající $\pm 1/\sqrt{m}$ se stejnou pravděpodobností.¹ O těchto maticích platí obdobné závěry jako u gaussovských matic.

Matice Gaussova nebo Bernoulliho typu spadají do obecnější kategorie tzv. sub-gaussovských rozdělení, která mají ojedinělou výhodu, a sice jejich rekonstrukční schopnosti jsou invariantní vůči unitární transformaci, což bude např. s výhodou využito v kapitole o komprimovaném snímání.

Podmínku RIP, i když s většími nároky na počet řádků m , splňují také *částečné náhodné Fourierovy matice* (partial random Fourier matrices). Ty vznikají tak, že z deterministické unitární matice DFT rozměru $N \times N$ vybereme náhodně pouze m řádků (to odpovídá náhodně vybraným kmitočtům). Přesněji, bylo ukázáno [38, 69], že takové matice „s vysokou pravděpodobností“ splňují RIP s konstantou δ za podmínky $m \geq C\delta^{-2}k \ln^\alpha k \ln(N)$.

3.4 Učení slovníku

Dosud jsme pracovali s předpokladem, že reprezentační slovník je fixní, předem daný. Takovými slovníky může být libovolný zástupce bází či framů, kterým byla věnována kapitola 2. Přirozenou otázkou je, zda by bylo možné pro určitou kategorii signálů slovník

¹V české literatuře se toto rozdělení obvykle označuje jako *alternativní*.

vytvořit či *naučit* tak, aby bylo zajištěno co nejřidší vyjádření takovýchto signálů? Literatura věnovaná podobným otázkám se do značné míry prolíná s oblastmi matematické statistiky [31], strojového učení a dolování dat (data-mining) [47].

Co se týká samotného *učení slovníku* (dictionary learning), mezi nejčastěji využívané algoritmy patří MOD (method of directions) [34, 35] a zejména různé varianty algoritmu K-SVD [1, 32], který je zobecněním algoritmu *K-means*, známého z oblasti strojové klasifikace.

4 KOMPRIMOVANÉ SNÍMÁNÍ

Tato kapitola se věnuje tzv. komprimovanému snímání (compressed sensing, compressive sampling), což je jedna z důležitých, pravděpodobně nejatraktivnější, aplikací v oblasti řídkých reprezentací signálů. Myšlenka tohoto přístupu spočívá v neadaptivním snímání signálu jen tolika „vzorky“, kolik je jich skutečně potřeba. Teoretické zázemí je obsahuje část 4.2 a numerické experimenty jsou v části 4.3.

V předchozí kapitole bylo cílem nalezení či aproximování tzv. řídkého řešení soustavy lineárních rovnic, která modeluje signál, čili

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{vzhledem k} \quad \mathbf{A}\mathbf{x} = \mathbf{y}. \quad (\text{P0})$$

Jak se ukáže později, v úloze komprimovaného snímání bude mít matice \mathbf{A} speciální tvar.

4.1 Klasické vzorkování

Revidujme na tomto místě krátce, jak je vzorkování klasicky chápáno.

Představme si, že máme k dispozici vzorky signálu se spojitým časem. Jak z nich rekonstruujeme původní signál? Je zřejmé, že bez dodatečné informace o charakteru signálu to je nemožné.

Shannon, Whittaker, Nyquist, Kotelnikov, Raabe, Someya a další nezávisle na sobě objevili, že pokud signál vzorkujeme tak, že

$$f_{\text{vz}} > 2f_{\text{max}}, \quad (4.1)$$

kde f_{max} je nejvyšší nenulový kmitočet obsažený v signálu, pak je možné (v teorii) původní signál *přesně* rekonstruovat z jeho vzorků, a to jednoduchou interpolací mezi vzorky pomocí funkce sinc. Pokud uvedená podmínka není splněna, není zaručena přesná rekonstrukce a hrozí efekt překrývání spekter, nazývaný *aliasing*.

Pro možnost rekonstrukce je tedy potřeba o signálu mít nějakou *apriorní* informaci, v uvedeném případě je to jeho omezené spektrum. Lze vzorkovací kmitočet stlačit níže, pokud bychom o signálu věděli ještě něco více? Ano, další poměrně známá poučka říká, že pokud signál má spektrum rozkládající se pouze na intervalu (f_1, f_2) , pak pro rekonstrukce bude dostačovat odebrat vzorky kmitočtem splňujícím

$$\frac{2f_1}{n-1} > f_{\text{vz}} > \frac{2f_2}{n}, \quad (4.2)$$

pokud takové vhodné celé n existuje. Rekonstrukce se provede opět interpolací.

Tyto znalosti nás ujišťují, že vzorků je možné odebrat tím méně, čím více o signálu víme. V dalším ukážeme, že apriorní vlastnost může být právě řídkost a že množství vzorků může ještě klesnout, pokud upustíme od tradičního pojetí *vzorku* jako hodnoty signálu se spojitým časem v nekonečně krátkém okamžiku (konvoluce s Diracovým impulzem).

Ačkoliv práce ohledně komprimovaného snímání signálů *se spojitým časem* existují [58, 59], my se budeme zabývat pouze signály *diskrétními*, kde tedy půjde o volnou analogii k operaci *podvzorkování*. Pro tu samozřejmě platí stejné zákonitosti jako ve výše uvedených příkladech.

4.2 Komprimované snímání

Obvyklý způsob komprimování dat byl až donedávna tento: nasbírat všechna data (tj. celý signál), provést vhodnou transformaci, vyhodnotit získané koeficienty a většinu z nich nevyužít („vyhodit“), protože z nějakého důvodu nesou málo informace. Například v kompresním formátu JPEG tomu odpovídá získání hodnot všech pixelů fotografie (řekněme 10 milionů), provedení dvojrozměrné diskretní kosinové transformace – DCT – a kvantování vzniklých koeficientů. Ponechání pouze k koeficientů (nejčastěji se jedná o ty s největší velikostí) a zakódování pouze jejich pozice/indexu a velikosti tak vlastně odpovídá *adaptivnímu* přístupu k signálu, který úzce souvisí s chybou nejlepší aproximace, viz vztah (1.4). Úspěch tohoto přístupu je umožněn právě tím, že k určitému typu signálu (třeba fotografie) je možné nalézt takovou reprezentaci (zde DCT po blocích), ve které je signál řídký nebo přibližně řídký.

Komprimované snímání (compressed sensing, compressive sampling, CS) přichází s jinou strategií: za předpokladu řídkosti ve vhodné reprezentaci rovnou snímat signál *lineárně a neadaptivně*, a to pouze tolikrát, kolik je skutečně třeba!

Signál tedy předpokládáme řídký v nějakém slovníku (po vzoru [14] se omezíme pouze na ortonormální báze). Bázi označíme Ψ a tedy signál \mathbf{z} má vyjádření $\mathbf{z} = \Psi\mathbf{x}$, kde \mathbf{x} je k -řídký vektor souřadnic. Cílem je provést „malý počet“ neadaptivních měření, které budou mít charakter skalárních součinů se signálem, což lze zapsat jako $\mathbf{y} = \mathbf{P}\mathbf{z} = \mathbf{P}\Psi\mathbf{x}$. Zde \mathbf{P} je tzv. *měřicí* (snímací) *matice* rozměru $m \times N$ a jednotlivé složky vektoru \mathbf{y} jsou výsledky měření, které vznikají jako lineární kombinace vzorků signálu. Počet měření je $m \ll N$.

Tedy zde vidíme, že pojem vzorek je zde zobecněn, v komprimovaném snímání se vzorkem myslí měření – skalár, který vznikne *lineární kombinací původních hodnot signálu*. Pokud by koeficienty lineární kombinace byly samé nuly vyjma jedné jedničky (Dirac), dostali bychom se zpět na vzorkování v obvyklém slova smyslu.

Fáze rekonstrukce signálu je zde ovšem už nelineární (a časově mnohem náročnější). Rekonstrukce technikami ℓ_1 -minimalizace probíhá takto:

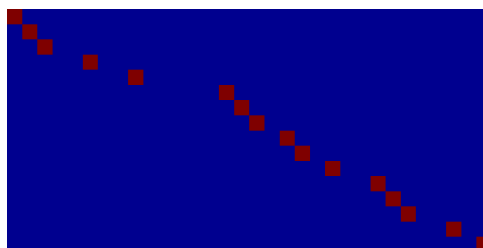
$$\mathbf{x}^{\text{OPT}} := \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{vzhledem k } \mathbf{y} = \mathbf{P}\Psi\mathbf{x}. \quad (\text{P1U})$$

Níže bude uvedeno, za jakých podmínek se tímto dosáhne právě nejřidšího řešení. Rekonstrukci samotného signálu \mathbf{z} jednoduše získáme jako $\mathbf{z} = \Psi\mathbf{x}^{\text{OPT}}$.

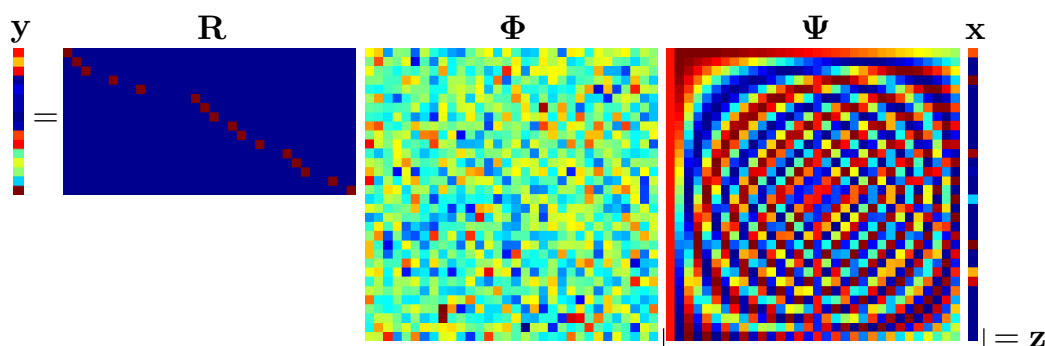
Nyní tedy otázka zní, jaké jsou požadavky na měřicí matici \mathbf{P} , aby bylo možné z tohoto „malého“ počtu měření zpátky rekonstruovat signál? V literatuře lze najít, že vhodné měřicí matice se uvažují ve tvaru $\mathbf{P} = \mathbf{R}\Phi$. Zde Φ je (zatím nespecifikovaná) matice $N \times N$ a \mathbf{R} je matice, která vznikne z jednotkové matice $N \times N$ ponecháním pouze m náhodně vybraných řádků, a tedy funguje jako náhodný výběr řádků z Φ . Náhodný výběr v \mathbf{R} se přitom řídí rovnoměrným rozdělením pravděpodobnosti. Konkrétní příklad realizace matice \mathbf{R} je na obr. 4.1 a ilustrace celkové situace je na obrázcích 4.2 a 4.3.

Tedy celkově, roli matice \mathbf{A} z kapitoly 3 hraje nyní matice $\mathbf{A} := \mathbf{R}\Phi\Psi$, a znovu pokládáme klíčovou otázku: jak zvolit Φ a m , aby \mathbf{A} umožňovala rekonstrukci signálu ℓ_1 -technikami?

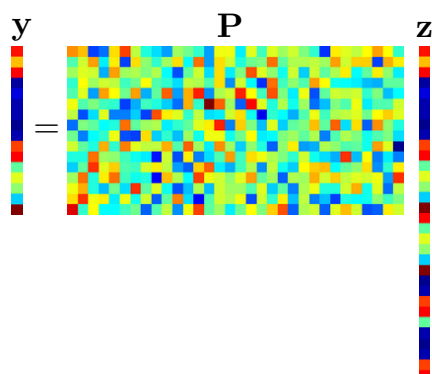
Na tuto otázku existuje řada odpovědí, nejprve uvedeme podmínku zahrnující vzájemnou koherenci μ . Ve speciálním případě, kdy je matice složena ze dvou ortonormálních



Obrázek 4.1: Příklad matice \mathbf{R} , která vznikla z jednotkové matice $N \times N$ rovnoměrně-náhodným zachováním pouze m řádků (zde $N = 32$, $m = 16$).



Obrázek 4.2: Ilustrace situace při komprimovaném snímání (bez přítomnosti šumu): naměřený vektor \mathbf{y} je roven součinu měřicí matice $\mathbf{P} = \mathbf{R}\Phi$, unitární matice Ψ a řídkého vektoru \mathbf{x} . Do procesu snímání vstupuje vektor \mathbf{z} , který je pozorovatelný a není sám řídký, zato je řídký v bázi Ψ , zde ilustrativně v ortonormální bázi zpětné DCT. Matice Φ je v tomto obrázku gaussovská.



Obrázek 4.3: Situace z obr. 4.2, ale se zdůrazněním procesu měření: násobení signálu \mathbf{z} maticí \mathbf{P} představuje m lineárních měření signálu \mathbf{z} .

bázi Φ a Ψ , tedy $[\Phi, \Psi]$, máme

$$\mu([\Phi, \Psi]) = \max_{1 \leq i, j \leq N} |\psi_i^\top \phi_j| \quad (4.3)$$

a hodnota (4.3) se pohybuje mezi $\frac{1}{\sqrt{N}}$ a 1. Následující tvrzení uvádí podmínku, za jaké je zaručena přesná rekonstrukce z m měření. (Tvrzení je záměrně formulované vágně, aby byla uchopitelnější podstata; detailněji viz [14, 8].)

Tvrzení 4.1 ([14], Theorem 1). *Nechť je dán signál \mathbf{z} , který má v Ψ k -řádkou reprezentaci \mathbf{x} . Pak řešení ℓ_1 -minimalizace (P1U), kde \mathbf{y} jsou měření, je současně s vysokou pravděpodobností nejřidší možné, pokud je zvoleno*

$$m \geq C \cdot \mu^2([\Phi, \Psi]) \cdot k \cdot N \cdot \ln N, \quad (4.4)$$

pro specifickou kladnou konstantu C .

Z nerovnice (4.4) je vidět, že počet měření závisí na řídkosti *pouze lineárně!* Role koherence mezi ortobázemi Φ a Ψ je zde zcela zřejmá: *kvadraticky* ovlivňuje nutný počet měření. To vysvětluje snahu hledat takové dvojice, jejichž koherence je minimální. Jak již bylo uvedeno dříve, je například známo, že minimální koherence $1/\sqrt{N}$ dosahuje dvojice $[\mathbf{I}, \mathbf{F}]$ – v takovém případě (4.4) říká, že počet měření musí být řádově pouze $k \cdot \ln N$. Naopak, jak koherence roste, přestává být měření dle podmínky (4.4) výhodné, až v určitém momentu přestává mít smysl, neboť m přeroste počet vzorků signálu N . (To nastane např. když bychom zvolili $\Phi = \Psi$.)

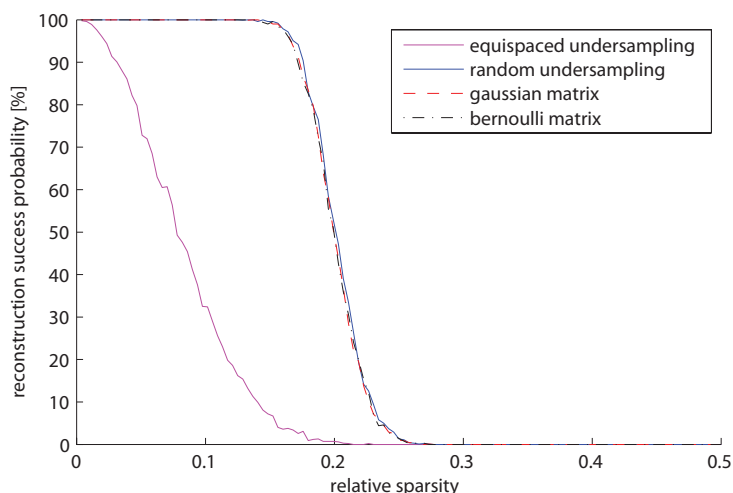
Např. v aplikacích pro zobrazování pomocí magnetické rezonance (MRI) máme z fyzikální podstaty problému rovnost $\Phi = \mathbf{F}$ a reprezentacní bázi volíme podle charakteru snímané oblasti, nejčastěji $\Psi = \mathbf{W}$. Tedy měříme náhodně vybrané spektrální koeficienty signálu, který je řídký ve vlnkové oblasti [14, 8, 53].

Nyní dáme odpověď na výše položenou otázku prostřednictvím vlastnosti zeslabené isometrie – RIP. Omezíme se přitom pouze na matice $\mathbf{P} = \mathbf{R}\Phi$, kde Φ jsou subgaussovského typu, ačkoliv to nejsou jediné možné konstrukce [14]. Víme, že takové matice \mathbf{P} s vysokou pravděpodobností splňují RIP, pokud $m \geq C \cdot k \cdot \ln(N/k)$. Subgaussovské matice mají důležitou vlastnost – invarianci vůči unitární transformaci [38]. Tato vlastnost znamená, že pokud \mathbf{P} vyhovuje RIP, pak pro $\mathbf{P}\Psi$ platí totéž; proto též celá \mathbf{A} vyhovuje RIP. V tomto smyslu tedy uvedené matice jsou „univerzální“, neboť nezávisejí na ortonormální bázi, ve které je signál řídký. Poznamenejme také, že je zajímavé, že matici Ψ vůbec nemusíme znát při snímání, nýbrž pouze při rekonstrukci.

V případě, že v matici $\mathbf{A} = \mathbf{R}\Phi\Psi$ vystupují dvě (deterministické) ortonormální báze Φ a Ψ , dle [14, 72] platí (zjednodušeně), že pro úspěšnou rekonstrukci je potřeba $m \geq C \cdot k \cdot (\ln N)^4$ měření. Zatím nedokázanou hypotézou však je, že platí mnohem slabší podmínka, kdy mocninu v nerovnosti vypustíme.

4.3 Simulace

Tato kapitola se věnuje experimentům, které do značné míry objasňují princip, na kterém je komprimované snímání založeno. Simulace zároveň potvrzují některé teoretické



Obrázek 4.4: Vliv použité měřicí matice na úspěšnost rekonstrukce signálu při použití signálu řídkého v bázi DCT.

poznatky uvedené výše. Pro testy byl použit rekonstrukční algoritmus z knihovny ℓ_1 -magic [71] využívající lineárního programování k řešení úlohy (P1) a algoritmus *Orthogonal Matching Pursuit* (OMP) [55, 62, 63]. Poznamenáváme, že konkrétním algoritmům je věnována kapitola 5.

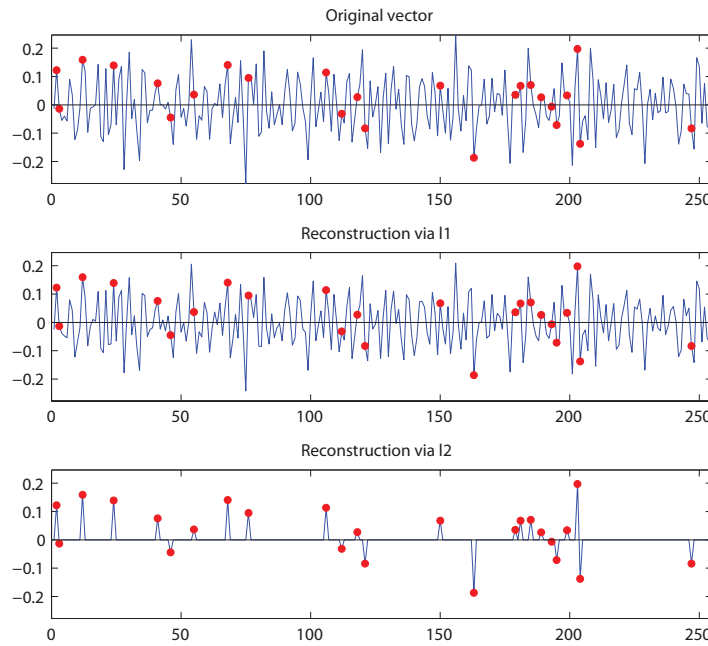
4.3.1 Vliv měřicí matice na úspěšnost ℓ_1 -rekonstrukce

První z testů zkoumá vliv použité měřicí matice na úspěšnost rekonstrukce signálu řídkého v bázi DCT. Tedy Ψ v (P1U) je matice zpětné DCT. Náhodně vygenerovaný signál byl snímán v časové oblasti za použití jedné ze čtyř měřících matic \mathbf{P} (rovnoměrné a náhodné podvzorkování, gaussovské a bernoulliovské matice) a následně byla řešena úloha (P1).

Výsledky experimentu jsou uvedeny v grafu na obr. 4.4. Závislost úspěchu rekonstrukce na relativní řídkosti signálu byla získána tak, že délka vektoru byla zvolena $N = 256$, provedeno bylo vždy $m = 128$ měření, signály měly postupně $k = 1$ až $k = 128$ nenulových DCT koeficientů. Jejich hodnoty byly generovány náhodně z rovnoměrného rozdělení od -1 do 1 . Pro každou řídkost vektoru a každý typ měřicí matice bylo provedeno 1000 pokusů. Rekonstrukce byla považována za úspěšnou, pokud byla relativní odchylka v ℓ_2 -normě nalezeného řešení $\tilde{\mathbf{x}}$ od řídkého vektoru maximálně 5 %, tedy $\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 \leq 0,05 \cdot \|\mathbf{x}\|_2$.

Překvapivě nejlepšího výsledku bylo dosaženo náhodným podvzorkováním signálu (tedy měřicí $\mathbf{P} = \mathbf{R}\mathbf{I} = \mathbf{R}$). Naopak při rovnoměrném podvzorkování byla úspěšnost rekonstrukce velmi malá, což se ale dalo očekávat, neboť díky podvzorkování pro většinu k není splněn vzorkovací teorém, a tedy dochází k aliasingu. (U ostatních metod pojem aliasing nemá smysl, neboť se již nejedná o klasické vzorkování.) Gaussovské a bernoulliovské náhodné matice vedou obě k podobným výsledkům jako náhodné vzorkování.

Konkrétní ukázka rekonstrukce z náhodných vzorků signálu je na obrázku 4.5. Vidíme zde signál generovaný řídkce v oblasti DCT, ze kterého je vybráno náhodně několik vzorků v časové oblasti. Rekonstrukce je nejprve provedena ℓ_1 -minimalizací a poté pro srovnání ℓ_2 -minimalizací (což odpovídá výpočtu pseudoinverzí [65]). Srovnáním rekonstruovaných signálů s původním zjistíme, že rekonstrukce pomocí ℓ_1 -minimalizace je téměř dokonalá,



Obrázek 4.5: Ukázka jedné realizace náhodného podvzorkování signálu délky $N = 256$, snímáno $m = 25$ vzorků (červeně). Signál je řídký ve frekvenční oblasti, zde $k = 8$ nenulových DCT koeficientů.

zatímco pomocí ℓ_2 -techniky získáme signál s nejnižší energií, který se s původním signálem shoduje pouze v místech vzorkování.

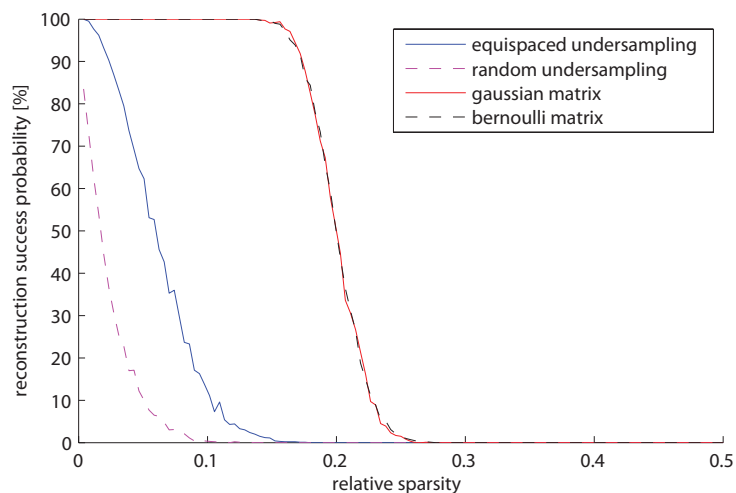
Výborný výsledek náhodného vzorkování, $\mathbf{P} = \mathbf{R}$, lze vysvětlit tím, že prvky báze DCT mají *nosič přes celou délku signálu* a tudíž každý vzorek obsahuje současně informaci o všech komponentách signálu. Je v tom možné vidět analogii případu, který již byl uveden, a sice že časově-kmitočtová dvojice $[\mathbf{I}, \mathbf{F}]$ má minimální možnou koherenci. Jinak tomu bude v následujícím příkladě.

4.3.2 Vliv délky nosiče a koherence mezi Φ a Ψ na míru úspěšnosti ℓ_1 -rekonstrukce

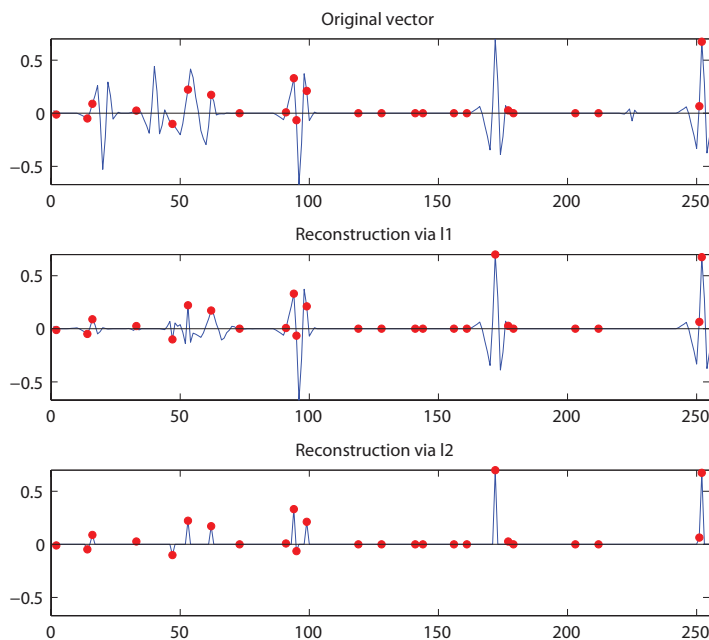
Shodný test byl proveden se signály, které byly generovány náhodně jako řídké v bázi Ψ ortonormálních waveletů třídy Daubechies. (Ty v diskrétním čase odpovídají filtrům FIR s konečnou impulzní odezvou.)

V testu byla použita nejvyšší hloubka dekompozice $J = 5$. Z obr. 4.6 a 4.7 je zřejmé, že v případě waveletu Daubechies 3 dává náhodné vzorkování nejhorší výsledky. Vysvětlení tohoto úkazu je prosté: tento wavelet má velmi krátký nosič (v hloubce dekompozice j je to $5(2^j - 1) + 1$, a tedy nejkratší atom má nosič délky 6 a nejdelší 156 vzorků [67]), a proto náhodné vzorky se nemusejí „trefit“ do nosiče těchto elementů a tedy o něm ani neunesou žádnou informaci, využitelnou při rekonstrukci. Ovšem u gaussovských a bernoulliovských matic je tomu jinak – každé měření nese informaci o celém signálu, což vysvětluje jejich úspěšnost.

V této souvislosti vypadá tedy jako přirozené, že existuje pozitivní závislost mezi délkou nosiče bazového prvku a šancí na rekonstrukci. Mohl by se zdát přirozený i fakt, že



Obrázek 4.6: Vliv použité měřicí matice na úspěšnost rekonstrukce signálu při použití signálu řídkého v bázi waveletu Daubechies 3.



Obrázek 4.7: Ukázka jedné realizace náhodného podvzorkování signálu délky $N = 256$, počet vzorků $m = 25$. Signál je řídký v bázi waveletu Daubechies 3 ($k = 8$ nenulových koeficientů). Neúspěch ℓ_1 -rekonstrukce je evidentní v několika úsecích.

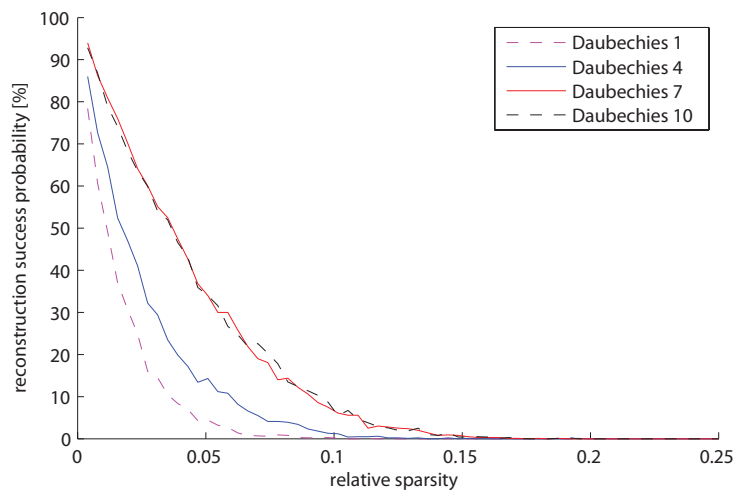
Tabulka 4.1: Vzájemné koherence waveletů Daubechies s Diracovou bází, tj. $\mu([\mathbf{W}, \mathbf{I}])$.

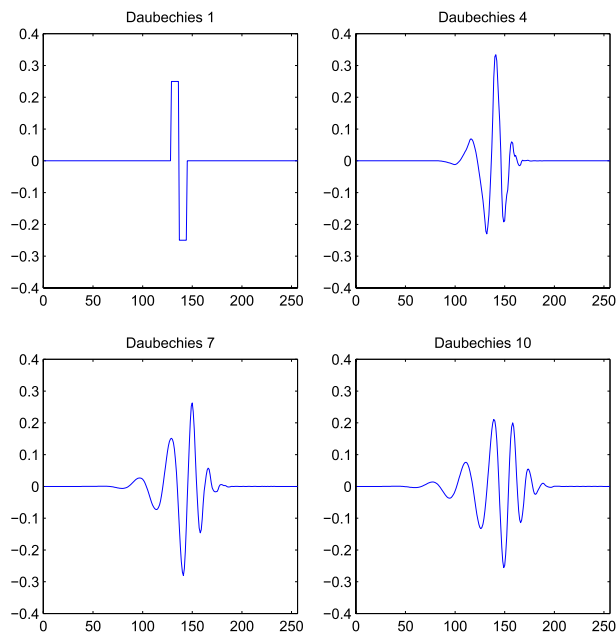
Wavelet	Hodnota μ
Daubechies 1	0,70711
Daubechies 2	0,83652
Daubechies 3	0,80689
Daubechies 4	0,71485
Daubechies 5	0,72431
Daubechies 6	0,75113
Daubechies 7	0,72913
Daubechies 8	0,67563
Daubechies 9	0,65729
Daubechies 10	0,68846

čím delší nosič, tím menší vzájemná koherence s Diracovou bází; avšak to obecně neplatí, protože koherence jakékoliv báze Ψ s Diracovou bází \mathbf{I} je jednoduše rovna největšímu prvku v Ψ . Situaci ukazuje tab. 4.1.

Abychom tuto teorii prověřili, byl proveden experiment nastavený stejně jako výše, ale použity byly postupně ortogonální wavelety Daubechies různých řádů (viz obr. 4.9). Na obr. 4.8 jsou jeho výsledky. Jak je vidět, koherence při použití Daubechies 1 je srovnatelná s koherencí v případě řádu 4, avšak úspěšnost rekonstrukce v druhém případě je výrazně vyšší. Daubechies 7 má koherenci mírně vyšší než Daubechies 4, takže bychom dle teorie očekávali horší výsledky, avšak experiment ukazuje přesně opačný fakt. Konečně, řád 10 má koherenci nižší než řád 7, nicméně výsledky jsou téměř totožné.

Shrnuto, tento pokus ukazuje a upozorňuje, že vzájemná koherence pomáhá pouze vytyčit hranici pro „nejhorší možný případ“ (worst case), ale nedává nám celou informaci o charakteru v „průměrném případě“.

**Obrázek 4.8:** Vliv délky nosiče na úspěšnost rekonstrukce signálu po náhodném podvzorkování.

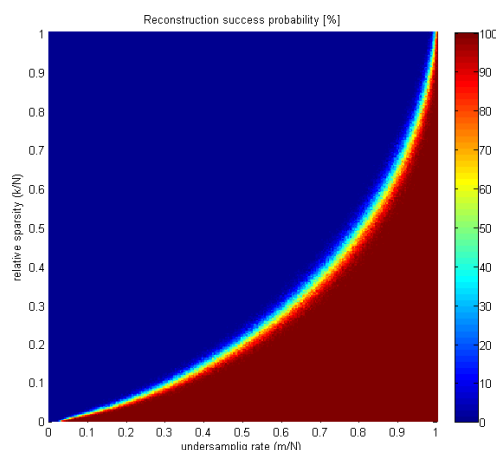


Obrázek 4.9: Srovnání atomů waveletu Daubechies 1, 4, 7 a 10 odpovídajících hloubce dekompozice $j = 3$. Je zde jasně patrná rostoucí délka nosiče.

4.3.3 Přechod fází

Hranice mezi oblastmi s „jistou“ a „nemožnou“ rekonstrukcí signálu \mathbf{x} z měření \mathbf{y} se nazývá přechod fází (*Phase Transition*). Typický tvar tohoto přechodu v závislosti na míře podvzorkování a řídkosti vektoru je vidět na obr. 4.10.

Šířka fázového přechodu závisí na délce vektoru \mathbf{x} – s rostoucí délkou N klesá šířka fázového přechodu, až v limitě je tento přechod zcela ostrý [6]. Důležité však je, že se s rostoucí délkou N nemění pozice teoretického středu přechodu a tedy relativní řídkost nutná k rekonstrukci signálu nezávisí na N .



Obrázek 4.10: Relativní četnost úspěšné rekonstrukce při použití matice s Bernoulliho rozdělením ($N = 300$). Na vodorovné ose je „míra podvzorkování“ $\frac{m}{N}$, na svislé pak relativní řídkost vektoru $\frac{k}{N}$.

4.4 Komprimované snímání založené na nízké hodnotě

Komprimované snímání lze aplikovat také pro signály, které lze chápat jako matice (tedy nejčastěji obrazy, datové záznamy). Místo řídkosti se zde pracuje s předpokladem nízké hodnoty matice $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$, příp. $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$. Měření, tak jak jej známe z předchozích částí, má opět podobu lineární kombinace, nyní prvků z \mathbf{X} . Formálně je to lineární zobrazení $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$, kde $m \ll n_1 n_2$, tedy měření \mathbf{y} dostáváme jako

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) \in \mathbb{R}^m.$$

Abychom byli schopni rekonstruovat \mathbf{X} z \mathbf{y} , budeme předpokládat, že \mathbf{X} má hodnot nejvýše $\text{rank } \mathbf{X} = r \ll \min\{n_1, n_2\}$. Naivní přístup k řešení optimalizačního problému

$$\arg \min_{\mathbf{X}} \text{rank } \mathbf{X} \quad \text{vzhledem k } \mathbf{y} = \mathcal{A}(\mathbf{X}) \quad (4.5)$$

je opět NP-těžký, ale můžeme ho relaxovat podobně jako v předchozím případě [70].

V části 1.1 jsme uvedli SVD rozklad matice a konstatovali jsme, že matice \mathbf{X} má hodnot r právě tehdy, když vektor $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\mathbf{X})$ singulárních čísel má řídkost r , tj. když platí $\text{rank } \mathbf{X} = \|\boldsymbol{\sigma}(\mathbf{X})\|_0$. Zde již vidíme příležitost k relaxaci analogické s relaxací na ℓ_1 -normu pro vektory. Nukleární normu jsme definovali jako ℓ_1 -normu singulárních čísel, $\|\mathbf{X}\|_* = \|\boldsymbol{\sigma}(\mathbf{X})\|_1$, tudíž můžeme uvažovat relaxovaný minimalizační problém

$$\arg \min_{\mathbf{X}} \|\mathbf{X}\|_* \quad \text{vzhledem k } \mathbf{y} = \mathcal{A}(\mathbf{X}). \quad (4.6)$$

Tento problém je konvexní, je tedy efektivně řešitelný a v poslední době existuje řada algoritmů pro získání optima [12, 45].

Analýze podmínek pro shodu s řešením původní úlohy (4.5) byla opět věnována řada prací, viz např. [70], přičemž se nejčastěji uvažují náhodná zobrazení \mathcal{A} . Ukazuje se, že matice \mathbf{X} s hodnotí nejvýše r mohou být s vysokou pravděpodobností rekonstruovány z m subgaussovských měření, jestliže [39]

$$m \geq Cr \max\{n_1, n_2\}. \quad (4.7)$$

5 PROXIMÁLNÍ ALGORITMY PRO ŘEŠENÍ KON- VEXNÍCH OPTIMALIZAČNÍCH ÚLOH

V předchozích kapitolách jsme mluvili o řešení optimalizačních úloh, vedoucích k nalezení řídkých reprezentací signálů. Na jejich řešení je dnes možné použít celou škálu algoritmů, jejichž základní přehled zde uvádíme, a posléze se detailněji věnujeme tzv. proximálním algoritmům, které pro praktické výpočty v aplikacích používáme.

Jak už bylo řečeno, žádná metoda nemůže najít *vždy* přesné řešení v polynomiálním čase, a proto se uchylujeme k metodám aproximativním. S trochou zjednodušení lze říct, že algoritmy lze rozdělit do tří kategorií [32, 48, 63, 64, 23], ačkoliv stálý vzestup počtu nových algoritmů ukazuje, že toto základní dělení je pouze hrubé:

- *Hladové (greedy) algoritmy*, jejichž hlavní princip spočívá v tom, že v každé iteraci najdou jeden (případně více) „nejvýznamnějších“ atomů. Důležité je, že v dalším průběhu algoritmu vybraný atom už podílí na konečném řešení nemůže být zbaven. Výhodou těchto metod je nízká složitost, nevýhodou pak, že není zaručeno dosažení globálního optima; dokonce jsou známy příklady, kdy tyto algoritmy zcela selhávají. Zástupci této kategorie jsou *MP* (Matching Pursuit) [55], v současnosti často používaný *OMP* (Orthogonal Matching Pursuit) [62, 78] a různé další odvozeniny.
- Další kategorie metod je založená na ℓ_1 -relaxaci, a proto se tyto algoritmy nazývají *relaxační*. Tyto algoritmy spoléhají na to, že se za určitých podmínek (viz část 3.3) dostaneme k řešení přesnému nebo alespoň relativně blízkému. Z těchto metod jmenujme *BP* (Basis Pursuit) [16], modifikovanou *LARS* (Least Angle Regression, homotopy method) [31, 38], *IRLS* (Iterative Reweighted Least Squares, někdy také jako *FOCUSS* – FOCal Underdetermined System Solver) [42, 38] nebo *Dantzig Selector* [10]. Jelikož ℓ_1 -norma je konvexní funkce, na tyto úlohy lze použít i tzv. proximální algoritmy, kterým se věnuje zbytek této kapitoly.
- Mimo tyto kategorie stojí algoritmy založené na *prahování* (thresholding) [32] a další, které využívají jednotlivých předností z obou skupin algoritmů. Taková skupina algoritmů se pak nazývá *hybridní*. Jedním z dalších algoritmů je algoritmus *A*OMP* [49], využívající A^* -algoritmu prohledávání informačních stromů. V neposlední řadě musíme jmenovat algoritmy vycházející z myšlenky aproximativního předávání zpráv v grafovém modelu (*approximate message passing*) [29].

5.1 Proximální algoritmy

Tzv. *proximální algoritmy* jsou postupy z teorie optimalizace, které zahrnují celou škálu konvexních úloh, včetně relaxovaných úloh s ℓ_1 -normou. Je vhodné podotknout, že tyto metody, ačkoliv byly známé již dříve [60, 56, 20, 5], získaly si značnou pozornost až díky řídkým reprezentacím.

Optimalizace je založena na iterativní minimalizaci funkcionalů, přičemž jsou známy podmínky, za kterých algoritmus konverguje k optimu úlohy. U proximálních metod není rychlost tak vysoká jako u některých specializovaných algoritmů, ale to je cenou za flexibilitu, kterou využíváme.

V této kapitole ukážeme algoritmus, který umí nalézt řešení poměrně obecné, tzv.

konvexní úlohy bez omezujících podmínek

$$\arg \min_{\mathbf{x}} g(\mathbf{x}) + h(\mathbf{x}), \quad (5.1)$$

přičemž přesné podmínky, které musejí g a h splňovat, uvedeme až později.

5.1.1 Optimalizační úlohy s omezeními a bez omezení

Výše uvedená úloha

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \mathbf{Ax} = \mathbf{y}. \quad (\text{P1})$$

je optimalizační úlohou s omezujícími podmínkami (constrained), který neodpovídá (5.1). Na tvar bez omezujících podmínek lze však převést za pomoci indikátorové funkce.

Definice 5.1. *Nechť C je neprázdňá podmnožina \mathbb{R}^N . Indikátorová (charakteristická) funkce množiny C je*

$$\iota_C : \mathbf{x} \mapsto \begin{cases} 0 & \text{pro } \mathbf{x} \in C, \\ \infty & \text{jinak.} \end{cases} \quad (5.2)$$

Místo (P1) se pak řeší nová úloha

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 + \iota_{\{\mathbf{x}: \mathbf{Ax}=\mathbf{y}\}}, \quad (5.3)$$

která je již v požadovaném tvaru a druhý sčítanec vynucuje náležitost řešení do přípustné množiny. Obdobně druhý typ úlohy

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \|\mathbf{Ax} - \mathbf{y}\|_2 \leq \delta, \quad (\text{P1}\delta)$$

lze převést na

$$\arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 + \iota_{\{\mathbf{x}: \|\mathbf{Ax}-\mathbf{y}\|_2 \leq \delta\}}. \quad (5.4)$$

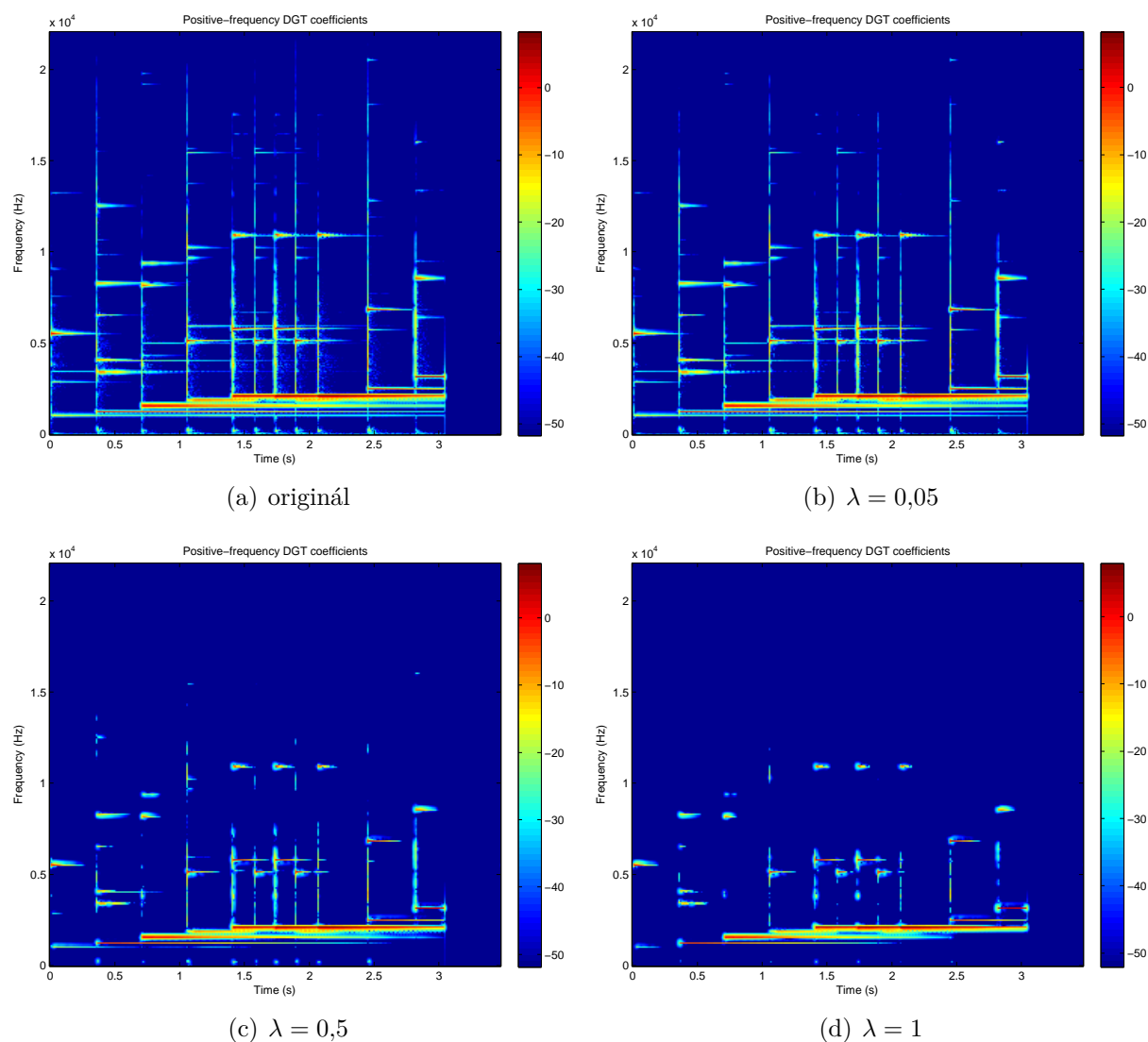
Lze však také ukázat, že existuje ekvivalence mezi omezenou úlohou (P1 δ) a neomezenou úlohou, která je dnes (zčásti nesprávně) nazývána LASSO:

$$\arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{Ax} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (5.5)$$

přičemž ekvivalence je míněna tak, že ke každému δ lze nalézt $\lambda > 0$ tak, že obě úlohy dávají stejný vektor optimálních \mathbf{x} .

Úloha posledního typu bývá také nazývána *regularizovaná lineární inverzní úloha*, protože dodatečným členem (nemusí to být zrovna ℓ_1 -norma) penalizujeme určitá řešení z nekonečně mnoha, čímž úlohu tzv. *regularizujeme*. Demonstrujme nyní krátce vliv regularizačního parametru na výsledek optimalizace. Čím vyšší bude λ , tím více penalizujeme neřídká řešení, ale na druhou stranu při příliš vysokém λ tím můžeme způsobit velkou odchylku od dat. Volba parametru je tedy citlivá záležitost a zcela obecné doporučení pro ni neexistuje. Obr. 5.1 ukazuje spektrogramy signálu vzešlé z řešení úlohy (5.5). Jedná se o Gaborovu reprezentaci signálu \mathbf{y} . V roli \mathbf{A} tedy vystupuje diskrétní Gaborův frame. Vykreslené spektrogramy nejsou nic jiného než moduly Gaborových koeficientů \mathbf{x} získaných pro uvedená různá λ .

Než se dostaneme k algoritmům, které úlohy bez omezujících podmínek řeší, musíme se nejprve věnovat tzv. *proximálnímu operátoru*, který je základním kamenem algoritmů.



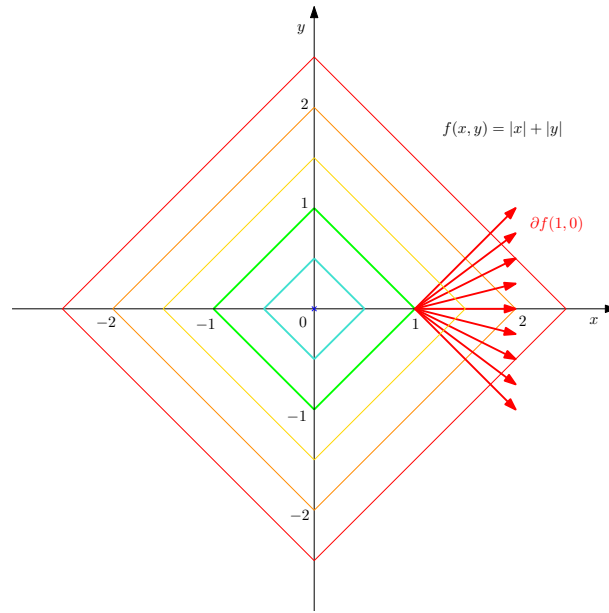
Obrázek 5.1: Demonstrace účinku regularizačního parametru λ na rekonstrukci.

5.2 Proximální operátor

Projekční operátor na konvexní množinu nachází bod v této konvexní množině ležící nejblíže výchozímu bodu. Proximální operátor konvexní funkce je přirozeným rozšířením pojmu projekčního operátoru na konvexní množinu. Úlohou této části je jej definovat, interpretovat a ukázat jeho vlastnosti.

Definice 5.2 ([73]). Řekneme, že funkce $f(\mathbf{x})$ je zdola polospojité na otevřené podmnožině U metrického prostoru, jestliže pro libovolné $\alpha \in \mathbb{R}$ je množina $\{\mathbf{x} \in U : f(\mathbf{x}) > \alpha\}$ otevřená.

Funkci zdola polospojitou si můžeme představit tak, že je spojitá až na možné výjimky izolovaných skoků, ve kterých tato funkce nabývá vždy nižší hodnoty, než bychom dostali limitami zprava a zleva v tomto bodě (nebo může nabývat hodnoty menší z těchto limit).



Obrázek 5.2: Znázornění subdiferenciálu funkce $f(x, y) = |x| + |y|$ v bodě $[1, 0]$. Barevně vrstevnice, červené vektory představují subgradienty.

Definice 5.3 ([20]). *Nechť f je zdola polospojité konvexní funkce s neprázdným definičním oborem, $f : \mathbb{R}^N \rightarrow (-\infty, \infty)$. Subdiferenciál funkce f je operátor $\partial f : \mathbb{R}^N \rightarrow 2^{\mathbb{R}^N}$, který definujeme jako*

$$\mathbf{x} \mapsto \left\{ \mathbf{u} \in \mathbb{R}^N : \left(\forall \tilde{\mathbf{x}} \in \mathbb{R}^N \right) (\tilde{\mathbf{x}} - \mathbf{x})^\top \mathbf{u} + f(\mathbf{x}) \leq f(\tilde{\mathbf{x}}) \right\}. \quad (5.6)$$

Jednotlivé prvky \mathbf{u} subdiferenciálu se nazývají subgradienty.

Subgradient ∂f je zobecněný pojem ke gradientu $\nabla f = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_N} \right]$ pro nediferencovatelné funkce. Je zřejmé, že v bodech, kde je f diferencovatelná, tyto pojmy splývají. Subgradient ∂f je operátor, který přiřazuje funkci v bodě množinu.

Pro lepší pochopení uvedeme jako příklad subdiferenciál funkce dvou proměnných $f(x, y) = |x| + |y|$, konkrétně v bodě $[x, y] = [1, 0]$. Podmínka (5.6) pro tento případ má tvar

$$[\tilde{x} - x, \tilde{y} - y] \begin{bmatrix} u \\ v \end{bmatrix} + |x| + |y| \leq |\tilde{x}| + |\tilde{y}|, \quad (5.7)$$

což v našem případě znamená $(\tilde{x} - 1)u + \tilde{y}v + 1 \leq |\tilde{x}| + |\tilde{y}|$ pro každé $\tilde{x}, \tilde{y} \in \mathbb{R}$. Tato nerovnost musí platit pro všechna u, v , a proto rozбором možných kombinací znamének \tilde{x}, \tilde{y} dostaneme, že subgradienty jsou vektory tvaru $[u, v] = [1, v]$, přičemž $-1 \leq v \leq 1$. Subdiferenciál funkce $f(x, y) = |x| + |y|$ v bodě $[1, 0]$ je znázorněn na obr. 5.2.

Definice 5.4 ([20]). *Nechť $f : \mathbb{R}^N \rightarrow (-\infty, \infty)$ je zdola polospojité konvexní funkce s neprázdným definičním oborem. Pro každé $\mathbf{x} \in \mathbb{R}^N$ má minimalizační problém*

$$\arg \min_{\mathbf{y} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + f(\mathbf{y}) \quad (5.8)$$

jednoznačné řešení, které budeme značit $\text{prox}_f \mathbf{x}$. Tento operátor $\text{prox}_f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ budeme nazývat proximální operátor funkce f .

Z pohledu zpracování signálů se na vztah (5.8) lze dívat jako na regularizované potlačování šumu (regularized denoising). Předpokládáme-li, že daný signál \mathbf{x} vznikl aditivním zašuměním čistého signálu, který bychom chtěli získat zpět, pak člen $\|\mathbf{x} - \mathbf{y}\|_2^2$ zajišťuje, že výsledek odšumování nebude „příliš daleko“ od známého signálu \mathbf{x} (šum je rozumně velký). Člen $f(\mathbf{y})$ představuje *regularizér*, tedy výraz, který penalizuje hledané \mathbf{y} pomocí matematicky formulované konvexní funkce, a tím vynucuje některé apriorní vlastnosti \mathbf{y} ; jako příklady uveďme energii $f(\mathbf{y}) = \|\mathbf{y}\|_2$, relaxovanou řídkost $f(\mathbf{y}) = \|\mathbf{y}\|_1$ nebo totální variaci $f(\mathbf{y}) = \|\mathbf{y}\|_{\text{TV}}$. Formulace typu (5.8) lze rovněž odvodit přes bayesovský přístup k odstraňování šumu [32, 6].

Bylo zmíněno hned na počátku této části, že proximální operátor je obecnějším pojmem než běžná projekce P_C bodu \mathbf{x} na neprázdnou uzavřenou konvexní množinu $C \subset \mathbb{R}^N$; vskutku, běžnou projekci P_C je možné vidět jako optimum problému

$$\arg \min_{\mathbf{y} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \iota_C(\mathbf{y}). \quad (5.9)$$

pro stanovený bod \mathbf{x} , tedy roli f z definice (5.8) zde hraje indikátorová funkce ι_C .

Proximální operátor f můžeme charakterizovat touto inkluzí:

$$(\forall (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^n \times \mathbb{R}^n) \quad \mathbf{p} = \text{prox}_f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{p} \in \partial f(\mathbf{p}), \quad (5.10)$$

kteřá vychází z podmínek pro minimum: $\mathbf{x}^* = \arg \min_{\mathbf{x}} f(\mathbf{x}) \Rightarrow \mathbf{0} \in \partial f$. Inkluzi je možno redukovat na

$$(\forall (\mathbf{x}, \mathbf{p}) \in \mathbb{R}^n \times \mathbb{R}^n) \quad \mathbf{p} = \text{prox}_f(\mathbf{x}) \Leftrightarrow \mathbf{x} - \mathbf{p} = \nabla f(\mathbf{p}),$$

když je f diferencovatelná.

Nyní uveďme některé základní vlastnosti [20].

Tvrzení 5.5. *Nechť φ je zdola polospojité konvexní funkce s neprázdným definičním oborem, $\varphi : \mathbb{R}^N \rightarrow (-\infty, \infty)$. Pak platí následující pro posunutí, resp. změnu měřítka:*

- (i) *Nechť $\psi = \varphi(\cdot - \mathbf{z})$, kde $\mathbf{z} \in \mathbb{R}^N$. Potom $\text{prox}_\psi(\mathbf{x}) = \mathbf{z} + \text{prox}_\varphi(\mathbf{x} - \mathbf{z})$.*
- (ii) *Nechť $\psi = \varphi\left(\frac{\cdot}{\rho}\right)$, kde $\rho \in \mathbb{R} \setminus \{0\}$. Potom $\text{prox}_\psi(\mathbf{x}) = \rho \text{prox}_{\frac{\varphi}{\rho^2}}\left(\frac{\mathbf{x}}{\rho}\right)$.*

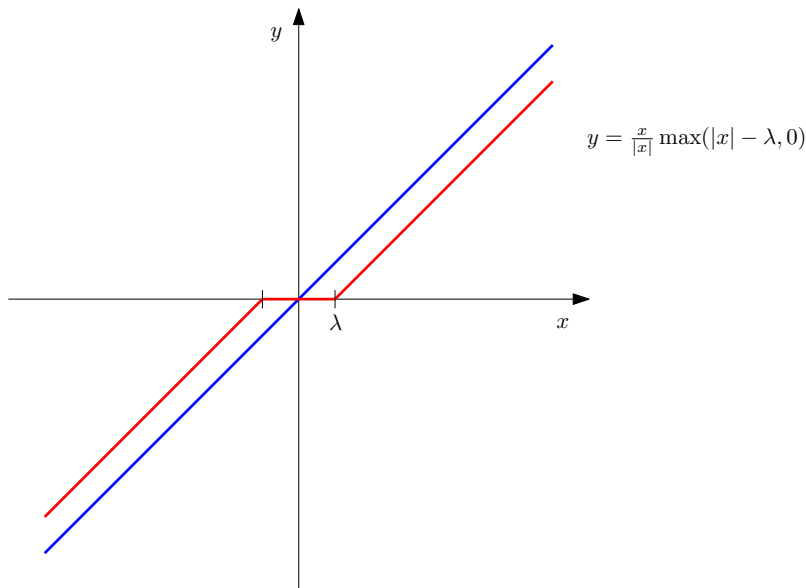
Důkaz. Položíme $\mathbf{p} = \text{prox}_\psi(\mathbf{x})$. Podle (5.10) je to ekvivalentní s $\mathbf{x} - \mathbf{p} \in \partial\psi(\mathbf{p})$. Postupně můžeme psát:

- (i) $\mathbf{x} - \mathbf{p} \in \partial\psi(\mathbf{p}) \Rightarrow \mathbf{x} - \mathbf{p} \in \partial\varphi(\mathbf{p} - \mathbf{z}) \Rightarrow (\mathbf{x} - \mathbf{z}) - (\mathbf{p} - \mathbf{z}) \in \partial\varphi(\mathbf{p} - \mathbf{z}) \Rightarrow \mathbf{p} - \mathbf{z} = \text{prox}_\varphi(\mathbf{x} - \mathbf{z}) \Rightarrow \mathbf{p} = \text{prox}_\psi(\mathbf{x}) = \mathbf{z} + \text{prox}_\varphi(\mathbf{x} - \mathbf{z})$.
- (ii) $\mathbf{x} - \mathbf{p} \in \partial\psi(\mathbf{p}) \Rightarrow \mathbf{x} - \mathbf{p} \in \frac{1}{\rho} \partial\varphi\left(\frac{\mathbf{p}}{\rho}\right) \Rightarrow \frac{\mathbf{x}}{\rho} - \frac{\mathbf{p}}{\rho} \in \partial\left(\frac{\varphi}{\rho^2}\right)\left(\frac{\mathbf{p}}{\rho}\right) \Rightarrow \mathbf{p} = \text{prox}_\psi(\mathbf{x}) = \rho \text{prox}_{\frac{\varphi}{\rho^2}}\left(\frac{\mathbf{x}}{\rho}\right)$.

□

5.2.1 Proximální operátor ℓ_1 -normy

Norma ℓ_1 hraje klíčovou roli při hledání řídkých reprezentací signálů a v komprimovaném snímání, proto uveďme proximální operátor speciálně pro tento případ $f(\mathbf{y}) = \lambda \|\mathbf{y}\|_1$,



Obrázek 5.3: Funkce měkkého prahování pro skalární veličinu. Modře je znázorněna identita a červeně tato funkce s obecným prahem λ . Vidíme, že pokud je velikost x v absolutní hodnotě vyšší než práh λ , pak je velikost x o tuto hodnotu posunuta, naopak pokud velikost x je již podprahová, koeficient je vynulován.

kde λ je regularizační skalár, který přidáváme bez újmy na obecnosti, ale jehož přítomnost oceníme v aplikacích. Podle definice můžeme napsat

$$\text{prox}_{\lambda \|\cdot\|_1}(\mathbf{x}) = \arg \min_{\mathbf{y}} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{y}\|_1, \quad (5.11)$$

což snadno rozepíšeme jako

$$\arg \min_{\mathbf{y}} \frac{1}{2} \sum_{i=1}^N (x_i - y_i)^2 + \lambda \sum_{i=1}^N |y_i|. \quad (5.12)$$

Pro daný vektor \mathbf{x} proximální operátor vrací takovou množinu $\{y_i\}$, že (5.12) je minimální. Poměrně jednoduchou analýzou [23], postupně po jednotlivých intervalech, kde je funkce $|\cdot|$ diferencovatelná, dospějeme k závěru, že

$$y_i = \frac{x_i}{|x_i|} \max(|x_i| - \lambda, 0) \quad (5.13)$$

(pro nulová x_i se uvedený výraz považuje za nulový). Tato funkce se nazývá *měkké prahování* (soft thresholding) [26, 16] a budeme ji značit $y_i = \text{soft}_\lambda(x_i)$, případně vektorově $\mathbf{y} = \text{soft}_\lambda(\mathbf{x})$, přičemž zdůrazňujeme, že operátor prahuje *každou složku vektoru \mathbf{x} zvlášť*. Měkké prahování je znázorněno na obr. 5.3.

5.2.2 Proximální operátor pro nukleární normu

Nukleární normu matice \mathbf{X} jsme definovali jako součet singulárních čísel z jejího SVD rozkladu, tedy jako ℓ_1 -normu vektoru singulárních čísel, $\|\mathbf{X}\|_* = \|\boldsymbol{\sigma}(\mathbf{X})\|_1$, viz část 1.1.

Obdobnou analýzou jako výše je možné dojít k závěru, že proximální operátor nukleární normy stojí na měkkém prahování singulárních čísel, formálně

$$\text{svt}(\mathbf{X}, \lambda) := \text{prox}_{\lambda \|\cdot\|_*}(\mathbf{X}) = \sum_{l=1}^n \text{soft}_{\lambda}(\sigma_l) \mathbf{u}_l \mathbf{v}_l^*, \quad (5.14)$$

kde \mathbf{u}_l a \mathbf{v}_l jsou singulární vektory \mathbf{X} . Čísla σ_l jsou nezáporná, takže prahování je jednodušší.

5.3 Proximální gradientní metoda (dopředně-zpětné dělení)

Pokud známe proximální operátor, můžeme se posunout dále a ukázat si algoritmy, které původní minimalizační úlohu numericky vyřeší. Omezíme se zde pouze na algoritmus tzv. *dopředně-zpětného dělení* (forward-backward splitting). Budeme tedy řešit konvexní úlohu

$$\arg \min_{\mathbf{x}} g(\mathbf{x}) + h(\mathbf{x}), \quad (5.1)$$

přičemž $h : \mathbb{R}^N \rightarrow \mathbb{R}$ je zdola polospojité konvexní funkce a $g : \mathbb{R}^N \rightarrow \mathbb{R}$ je konvexní funkce, která je diferencovatelná s β -lipschitzovským spojitým gradientem ∇g , tj. pro každé $[\mathbf{x}, \mathbf{y}] \in \mathbb{R}^N \times \mathbb{R}^N$ platí

$$\|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq \beta \|\mathbf{x} - \mathbf{y}\|, \quad (5.15)$$

kde $\beta \in (0, \infty)$. Úloha (5.1) má alespoň jedno řešení, pokud $g + h$ je koercivní, tj. $\lim_{\|\mathbf{x}\| \rightarrow \infty} (g(\mathbf{x}) + h(\mathbf{x})) = \infty$. Pokud navíc $g + h$ je *ryze* konvexní (tedy g nebo h je *ryze* konvexní), pak je toto řešení jediné. Platí tato důležitá ekvivalence:

Tvrzení 5.6. *Nechť $\mathbf{x} \in \mathbb{R}^N$. Pak*

$$\mathbf{x} \text{ je řešením úlohy (5.1)} \Leftrightarrow \mathbf{x} = \text{prox}_{th}(\mathbf{x} - t \cdot \nabla g(\mathbf{x}))$$

pro všechna $t \in (0, \infty)$.

Tato charakterizace pomocí pevného bodu implikuje možnost numerického řešení (5.1) iterativně takto:

$$\mathbf{x}^{k+1} = \text{prox}_{t^k h}(\mathbf{x}^k - t^k \cdot \nabla g(\mathbf{x}^k)) \quad (5.16)$$

pro vhodnou hodnotu parametru t^k . Horní index k zastupuje číslo iterace. Tato metoda se nazývá *proximální gradientní metoda* nebo také *dopředně-zpětné dělení* (forward-backward splitting). Skládá se ze dvou kroků – dopředného, explicitního gradientního kroku, ve kterém nevyužíváme nic než g na výpočet mezikroku $\mathbf{x}^{k+\frac{1}{2}} = \mathbf{x}^k - t^k \cdot \nabla g(\mathbf{x}^k)$, a zpětného, obecně implicitního proximálního kroku využívajícího pouze h k výpočtu $\mathbf{x}^{k+1} = \text{prox}_{t^k h}(\mathbf{x}^{k+\frac{1}{2}})$. Pokud je zvoleno $t^k < \frac{2}{\beta}$, posloupnost $\{\mathbf{x}^k\}$ generovaná iteracemi dle (5.16) konverguje k řešení problému (5.1), viz [20, 21, 61].

Ukažme fungování metody na jednoduchém příkladu nalezení řešení úlohy

$$\arg \min_x \frac{1}{2}(x - 2)^2 + \frac{1}{2}|2x + 1|. \quad (5.17)$$

Označíme-li $g(x) = \frac{1}{2}(x-2)^2$ a $h(x) = \frac{1}{2}|2x+1|$, můžeme konstatovat, že obě jsou konvexní a spojité a g je navíc diferencovatelná. Lipschitzovská konstanta musí být takové β , aby platilo (5.15), tedy konkrétně $|g'(x) - g'(y)| \leq \beta|x-y|$ pro všechna x a y . Gradient g je jednoduše $g'(x) = x-2$, což po dosazení dává podmínku

$$|x-2 - (y-2)| = |x-y| \leq 1 \cdot |x-y|.$$

Z toho ihned plyne, že $\beta = 1$. Funkce $g + h$ je evidentně koercivní a funkce g je navíc ryze konvexní – lze tedy použít proximální gradientní metodu k nalezení *jednoznačného* řešení.

Zvolíme libovolný počáteční bod – například $x_0 = 3$. Dále zvolíme pro jednoduchost konstantní krok $t^k = t = 0,3 < \frac{2}{\beta} = 2$. V každé iteraci se nejprve posuneme v t -násobku gradientu g :

$$x^{k+\frac{1}{2}} = x^k - t \cdot \nabla g(x^k) = x^k - 0,3 \cdot (x^k - 2) = 0,7x^k + 0,6.$$

Dále na tento bod aplikujeme proximální operátor, tedy $x^{k+1} = \text{prox}_{th}(x^{k+\frac{1}{2}})$. Absolutní hodnotu lze vidět jako ℓ_1 -normu skalární veličiny. Využijeme-li vlastnosti pro posunutí a změnu měřítka proximálního operátoru (tvrzení 5.5) v kombinaci s měkkým prahováním jakožto proximálním operátorem ℓ_1 -normy, dostaneme:

$$x^{k+1} = \frac{1}{2} \text{soft} \left(2x^{k+\frac{1}{2}} + 1; 0,3 \cdot 2^2 \cdot \frac{1}{2} \right) - \frac{1}{2} = \frac{1}{2} \text{soft} \left(2x^{k+\frac{1}{2}} + 1; 0,6 \right) - \frac{1}{2}.$$

V iteracích pokračujeme tak dlouho, dokud dva po sobě jdoucí body jsou od sebe dostatečně blízko, což znamená, že se nalézáme poblíž řešení. Pokud konkrétně ukončíme iterace, pokud $|x^{k+1} - x^k| < 10^{-5} \cdot |x^k|$, pak algoritmus zastaví v nalezeném minimu – bodě $x = 1$ – po 32 iteracích. Viz pro ilustraci obrázek 5.4.

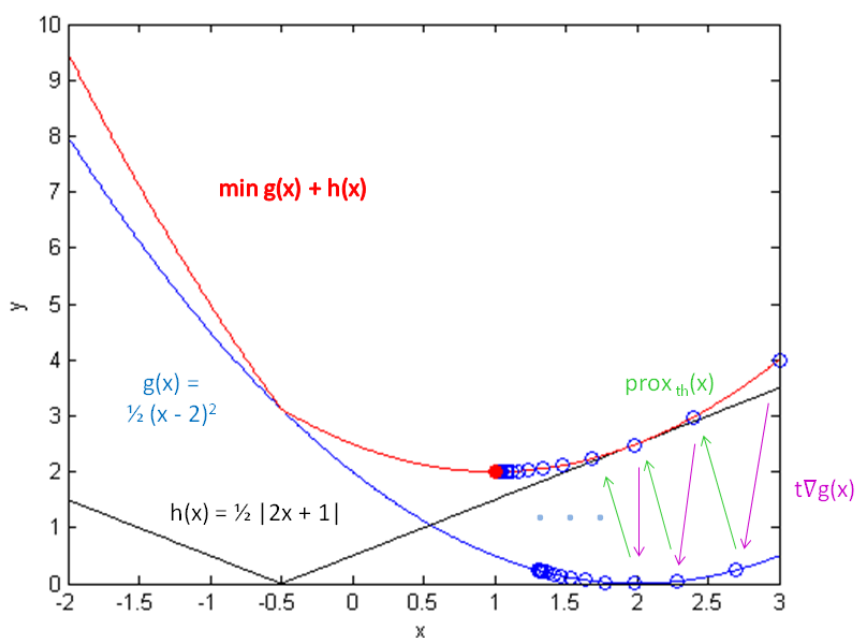
Zformálněme nyní algoritmus s použitím obecného, proměnlivého kroku t^k , a s využitím parametru λ^k , který umožní řízení směru v proximálním kroku:

Algoritmus 5.7 (Proximální gradientní algoritmus).

1. Zvolíme $\epsilon \in (0, \min\{1, 1/\beta\})$.
2. Zvolíme počáteční bod $\mathbf{x}_0 \in \mathbb{R}^N$.
3. Dokud není splněno kritérium zastavení iterací, provádíme pro $k = 0, 1, \dots$
 - Zvolíme $t^k \in [\epsilon, 2/\beta - \epsilon]$
 - Vypočítáme $x^{k+1/2} = x^k - t^k \nabla g(x^k)$
 - Zvolíme $\lambda^k \in [\epsilon, 1]$
 - Vypočítáme $x^{k+1} = x^k + \lambda^k (\text{prox}_{t^k h} x^{k+1/2} - x^k)$

Gradientní krok přibližuje aktuální iterované koeficienty datům, čímž však může dočasně dojít ke ztrátě vlastnosti požadované regularizačním členem. Proximální operátor proto ve svém dílčím kroku nalezne nové koeficienty, které nejsou příliš daleko od těch aktuálních, ale zároveň nejsou tolik penalizovány regularizačním členem.

V řadě případů je tento algoritmus pomalý, zvláště když se pracuje se stovkami tisíc proměnných. Často se jako rychlejší používá akcelerovaný algoritmus, v komunitě zpracování signálů známý jako Beck-Teboulleho algoritmus [3]. Je založený na dynamické optimální volbě skalárů t^k, λ^k během iterací. Přehledový článek [20] ukazuje, že také jiné algoritmy, známé izolovaně z vybraných oblastí zpracování signálů, lze formulovat jako proximální. Jedná se např. o Landweberovy iterace, (Fast) iterative shrinkage/thresholding algorithm – (F)ISTA, Alternating-direction method of multipliers – ADMM a další.



Obrázek 5.4: Průběh proximální gradientní metody při řešení úlohy (5.17). Nalezené minimum je vyznačeno jako červený puntík.

Reference

- [1] Aharon, M.; Elad, M.; Bruckstein, A. M.: K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representations. *IEEE Transactions on Signal Processing*, ročník 54, 2006: s. 4311–4322.
- [2] Balazs, P.: Frames and Finite Dimensionality: Frame Transformation, Classification and Algorithms. *Applied Mathematical Sciences*, ročník 2, č. 43, 2008: s. 2131–2144.
- [3] Beck, A.; Teboulle, M.: A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, ročník 2, č. 1, 2009: s. 183–202, doi:10.1137/080716542.
- [4] Ben-Israel, A.; Greville, T. N.: *Generalized Inverses: Theory and Applications*. Springer-Verlag, 2003, ISBN 0-387-00293-6.
- [5] Boyd, S. P.; Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, 2004, ISBN 0521833787.
- [6] Bruckstein, A. M.; Donoho, D. L.; Elad, M.: From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Review*, ročník 51, č. 1, 2009: s. 34–81, ISSN 0036-1445.
- [7] Candes, E.; Demanet, L.; Donoho, D.; aj.: Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, ročník 5, č. 3, 2006: s. 861–899, ISSN 1540-3459.
- [8] Candes, E.; Romberg, J.: Sparsity and incoherence in compressive sampling. *Inverse Problems*, ročník 23, 2006: s. 969–985.
- [9] Candes, E.; Romberg, J.; Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, ročník 52, č. 2, Feb 2006: s. 489–509, ISSN 0018-9448, doi: 10.1109/TIT.2005.862083.
- [10] Candes, E.; Tao, T.: The Dantzig selector: Statistical estimation when p is much larger than n . *Annals of Statistics*, ročník 35, č. 6, 2007: s. 2313–2351, ISSN 0090-5364.
- [11] Candès, E. J.; Li, X.; Ma, Y.; aj.: Robust Principal Component Analysis? *J. ACM*, ročník 58, č. 3, Červen 2011: s. 11:1–11:37, ISSN 0004-5411, doi:10.1145/1970392.1970395.
URL <http://doi.acm.org/10.1145/1970392.1970395>
- [12] Candes, E. J.; Recht, B.: Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, ročník 9, č. 6, 2009: s. 717–772, ISSN 1615-3375, doi:10.1007/s10208-009-9045-5.
URL <http://dx.doi.org/10.1007/s10208-009-9045-5>
- [13] Candes, E. J.; Tao, T.: Decoding by Linear Programming. *IEEE Transactions on Information Theory*, ročník 51, 2005: s. 4203–4215.

-
- [14] Candes, E. J.; Wakin, M. B.: An introduction to compressive sampling. *IEEE Signal Processing Magazine*, ročník 25, č. 2, 2008: s. 21–30, ISSN 1053-5888.
- [15] Chebira, A.; Kovacevic, J.: *An Introduction to Frames*. Now Publishers Inc, 2008, ISBN 160198068X.
- [16] Chen, S.; Donoho, D.; Saunders, M.: *Atomic decomposition by basis pursuit*. *SIAM J. Sci Comput.* 20 (1998), no.1, reprinted in *SIAM Review*, 2001.
- [17] Christensen, O.: *An Introduction to Frames nad Riesz Bases*. Boston-Basel-Berlin: Birkhäuser, 2003, ISBN 0817642951.
- [18] Christensen, O.: *Frames and Bases, An Introductory Course*. Boston: Birkhäuser, 2008, ISBN 9780817646776.
- [19] Cohen, A.; Dahmen, W.; DeVore, R.: Compressed sensing and best k -term approximation. *Journal of The American Mathematical Society*, ročník 22, 2009: s. 211–231.
- [20] Combettes, P.; Pesquet, J.: Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 2011: s. 185–212, doi:10.1007/978-1-4419-9569-8_10.
- [21] Combettes, P.; Wajs, V.: Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, ročník 4, č. 4, 2005: s. 1168–1200.
- [22] da Cunha, A. L.; Zhou, J.; Do, M.: The Nonsampled Contourlet Transform: Theory, Design, and Applications. *IEEE Transactions on Image Processing*, ročník 15, č. 10, 2006: s. 3089–3101.
- [23] Daňková, M.: *Komprimované snímání v perfuzním zobrazování pomocí magnetické rezonance*. Diplomová práce, Vysoké učení technické v Brně, 2014.
- [24] Daňková, M.; Rajmic, P.; Jiřík, R.: Compressed sensing of MR perfusion imaging. In *New trends in biomedical engineering*, Brno University of Technology, 2013, s. 63–70, in Czech.
- [25] Daubechies, I.: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992, ISBN 0898712742.
- [26] Donoho, D.: De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, ročník 41, č. 3, 1995: s. 613–627, ISSN 0018-9448.
- [27] Donoho, D.: Compressed sensing. *IEEE Transactions on Information Theory*, ročník 52, č. 4, 2006: s. 1289–1306, ISSN 0018-9448.
- [28] Donoho, D. L.; Elad, M.: Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proceedings of The National Academy of Sciences*, ročník 100, č. 5, 2003: s. 2197–2202.
- [29] Donoho, D. L.; Maleki, A.; Montanaria, A.: Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA*, ročník 106, č. 45, 2009: s. 18914–18919, doi:10.1073/pnas.0909892106.

- [30] Duffin, R. J.; Schaeffer, A. C.: A Class of Nonharmonic Fourier Series. *Transactions of the American Mathematical Society*, ročník 72, č. 2, 1952: s. 341–366.
- [31] Efron, B.; Hastie, T.; Johnstone, I.; aj.: Least angle regression. *Annals of Statistics*, ročník 32, 2004: s. 407–499.
- [32] Elad, M.: *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010, ISBN 9781441970107.
- [33] Eldar, Y. C.; Kutyniok, G.: *Compressed sensing: theory and applications*. Cambridge University Press, 2012.
- [34] Engan, K.; Aase, S.; Hakon Husoy, J.: Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, ročník 5, 1999, ISSN 1520-6149, s. 2443–2446 vol.5, doi:10.1109/ICASSP.1999.760624.
- [35] Engan, K.; Aase, S. O.; Husoy, J. H.: Multi-frame compression: theory and design. *Signal Processing*, ročník 80, č. 10, 2000: s. 2121–2140, ISSN 0165-1684, doi:10.1016/S0165-1684(00)00072-4.
URL <http://www.sciencedirect.com/science/article/pii/S0165168400000724>
- [36] Fannjiang, A. C.; Strohmer, T.; Yan, P.: Compressed Remote Sensing of Sparse Objects. *SIAM Journal on Imaging Sciences*, ročník 3, č. 3, 2010: s. 595–618.
- [37] Feichtinger, H. G.; Strohmer, T.: *Advances in Gabor Analysis*. Birkhäuser, 2001, ISBN 978-0-8176-4239-6, 376 s.
- [38] Fornasier, M.; Rauhut, H.: *Handbook of Mathematical Methods in Imaging*, kapitola Compressive Sensing. Springer, 2011, ISBN 978-0-387-92920-0.
- [39] Foucard, S.; Rauhut, H.: *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis, Springer New York, 2013, doi:10.1007/978-0-8176-4948-7.
- [40] Golbabaee, M.; Arberet, S.; Vandergheynst, P.: Compressive Source Separation: Theory and Methods for Hyperspectral Imaging. *Image Processing, IEEE Transactions on*, ročník 22, č. 12, Dec 2013: s. 5096–5110, ISSN 1057-7149, doi:10.1109/TIP.2013.2281405.
- [41] Goossens, B.; Aelterman, J.; Luong, H.; aj.: Efficient design of a low redundant Discrete Shearlet Transform. In *Local and Non-Local Approximation in Image Processing, 2009. LNLA 2009. International Workshop on*, aug. 2009, s. 112–124, doi:10.1109/LNLA.2009.5278394.
- [42] Gorodnitsky, I. F.; Rao, B. D.: Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *IEEE Transactions on Signal Processing*, ročník 45, č. 3, 1997: s. 600–616.

-
- [43] Gribonval, R.; Nielsen, M.: Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, ročník 49, 2003: s. 3320–3325.
- [44] Gröchenig, K.: *Foundations of time-frequency analysis*. Birkhäuser, 2001.
- [45] Gross, D.: Recovering Low-Rank Matrices From Few Coefficients in Any Basis. *Information Theory, IEEE Transactions on*, ročník 57, č. 3, March 2011: s. 1548–1566, ISSN 0018-9448, doi:10.1109/TIT.2011.2104999.
- [46] Guo, K.; Labate, D.: Optimally Sparse Multidimensional Representation Using Shearlets. *SIAM Journal on Mathematical Analysis*, ročník 39, č. 1, 2007: s. 298–318, doi:10.1137/060649781, <http://epubs.siam.org/doi/pdf/10.1137/060649781>. URL <http://epubs.siam.org/doi/abs/10.1137/060649781>
- [47] Hastie, T.; Tibshirani, R.; Friedman, J.: *The Elements of Statistical Learning*. Springer Series in Statistics, New York, NY, USA: Springer New York Inc., 2001.
- [48] Henzlová, P.: *Přehled metod komprimovaného snímání*. Bakalářská práce, ČVUT, 2010.
- [49] Karahanoglu, N. B.; Erdogan, H.: A* Orthogonal Matching Pursuit: Best-First Search for Compressed Sensing Signal Recovery. *Elsevier Digital Signal Processing*, 2011.
- [50] Kowalski, M.; Siedenburg, K.; Dörfler, M.: Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators. *Signal Processing, IEEE Transactions on*, ročník 61, č. 10, 2013: s. 2498–2511, ISSN 1053-587X, doi:10.1109/TSP.2013.2250967.
- [51] Kowalski, M.; Torrèsani, B.: Structured Sparsity: from Mixed Norms to Structured Shrinkage. In *SPARS'09 – Signal Processing with Adaptive Sparse Structured Representations*, editace R. Gribonval, Inria Rennes – Bretagne Atlantique, 2009, s. 1–6. URL <http://hal.inria.fr/inria-00369577/en/>
- [52] Kutyniok, G.; Lim, W.-Q.: Compactly supported shearlets are optimally sparse. *Journal of Approximation Theory*, ročník 163, č. 11, 2011: s. 1564–1589, ISSN 0021-9045, doi:<http://dx.doi.org/10.1016/j.jat.2011.06.005>. URL <http://www.sciencedirect.com/science/article/pii/S0021904511001067>
- [53] Lustig, M.; Donoho, D.; Santos, J.; aj.: Compressed Sensing MRI. *IEEE Signal Processing Magazine*, ročník 25, č. 2, 2008: s. 72–82, ISSN 1053-5888.
- [54] Mallat, S.: A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, ročník 11, 1989: s. 674–693, ISSN 0162-8828.
- [55] Mallat, S.; Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, ročník 41, č. 12, 1993: s. 3397–3415, ISSN 1053-587X.

- [56] Martinet, B.: Regularization d'inequations variationelles par approximations successive. *Revue Francaise de Informatique at Recherche Operationele*, ročník 4, 1970: s. 154–158.
- [57] Wikipedia contributors: JPEG. In *Wikipedia: the free encyclopedia*, St. Petersburg (Florida): Wikipedia Foundation, [cit. 6. 11. 2010].
URL <http://en.wikipedia.org/wiki/JPEG>.
- [58] Mishali, M.; Eldar, Y. C.: Sub-Nyquist Sampling. *Signal Processing Magazine, IEEE*, ročník 28, č. 6, 11 2011: s. 98–124, ISSN 1053-5888, doi:10.1109/MSP.2011.942308.
- [59] Mishali, M.; Eldar, Y. C.; Tropp, J. A.: Efficient sampling of sparse wideband analog signals. In *IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, 2008, ISBN 978-1-4244-2481-8, ISSN 0899-6156, s. 290–294.
- [60] Moreau, J. J.: Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, , č. 93, 1965: s. 273–299.
- [61] Otazo, R.; Candes, E. J.; Sodickson, D.: Low-rank and sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 2014, submitted to Magnetic Resonance in Medicine.
- [62] Pati, Y.; Rezaifar, R.; Krishnaprasad, P.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, ISSN 1058-6393, s. 40–44.
- [63] Špiřík, J.: Algorithms for computing sparse solutions. In *Proceedings of the 17th conference STUDENT EEICT Volume 3*, 2011, s. 123–127.
- [64] Špiřík, J.: Performance analysis of Matching Pursuit algorithm modifications. In *Proceedings of the 13th International Conference on Research in Telecommunication Technologies*, 2011.
- [65] Špiřík, J.; Rajmic, P.; Veselý, V.: Reprezentace signálů: od bází k framům. *Elektrorevue – Internetový časopis*, 2010: str. 11.
URL <http://elektrorevue.cz/cz/download/reprezentace-signalu--od-bazi-k-framum/>
- [66] Rajmic, P.; Průša, Z.: Discrete wavelet transform of finite signals: detailed study of the algorithm. *International Journal of Wavelets, Multiresolution and Information Processing*, ročník 12, č. 01, 2014: str. 1450001, doi:10.1142/S0219691314500015.
URL <http://www.worldscientific.com/doi/abs/10.1142/S0219691314500015>
- [67] Rajmic, P.; Prusa, Z.: Podrobná studie algoritmu pro výpočet waveletové transformace v diskrétním čase. *Elektrorevue – Internetový časopis*, ročník 113, 2010: s. 1–18.
URL <http://elektrorevue.cz/files/200000574-a3ef3a4e91>
- [68] Rao, K. R.; Yip, P.: *Discrete cosine transform: algorithms, advantages, and applications*. Academic Press, 1990, ISBN 0-12-580203-X.

- [69] Rauhut, H.; Romberg, J.; Tropp, J. A.: Restricted isometries for partial random circulant matrices. *Applied and Computational Harmonic Analysis*, ročník In Press, 2011, ISSN 1063-5203.
- [70] Recht, B.; Fazel, M.; Parrilo, P.: Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, ročník 52, č. 3, 2010: s. 471–501, doi:10.1137/070697835.
URL <http://epubs.siam.org/doi/abs/10.1137/070697835>
- [71] Romberg, J.: ℓ_1 -magic.
URL <http://users.ece.gatech.edu/~justin/l1magic/>
- [72] Rudelson, M.; Vershynin, R.: Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Information Sciences and Systems, 2006 40th Annual Conference on*, march 2006, s. 207–212, doi:10.1109/CISS.2006.286463.
- [73] Rudin, W.: *Analýza v reálném a komplexním oboru*. Praha: Academia, druhé vydání, 2003.
- [74] Said, A.; Pearlman, W.: A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *IEEE Transactions on Circuits and Systems for Video Technology*, ročník 6, č. 3, 1996: s. 243–250, ISSN 1051-8215.
- [75] Strohmer, T.; Heath, R.: Grassmannian Frames with Applications to Coding and Communication. *Applied and Computational Harmonic Analysis*, ročník 14, č. 3, 2003: s. 257–275.
- [76] Tauböck, G.; Hlawatsch, F.; Eiwien, D.; aj.: Compressive Estimation of Doubly Selective Channels in Multicarrier Systems: Leakage Effects and Sparsity-Enhancing Processing. *IEEE Journal of Selected Topics in Signal Processing*, ročník 4, č. 2, 2010: s. 255–271, ISSN 1932-4553.
- [77] Tibshirani, R.: Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, ročník 58, č. 1, 1996: s. 267–288.
- [78] Tropp, J.: Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, ročník 50, 2004: s. 2231–2242.
- [79] Unser, M.; Blu, T.: Mathematical properties of the JPEG2000 wavelet filters. *IEEE Transactions on Image Processing*, ročník 12, 2003: s. 1080–1090.
- [80] Veselý, V.; Rajmic, P.: *Funkcionální analýza s aplikacemi ve zpracování signálů*. Brno: Vysoké učení technické v Brně, 2014.
- [81] Wallace, G.: The JPEG Still Picture Compression Standard. *IEEE Transactions on Consumer Electronics*, ročník 38, 1991: s. 18–34.
- [82] Wickerhauser, M. V.: *Mathematics for Multimedia*. Birkhäuser, 2009, ISBN 978-0-8176-4879-4.