



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA STROJNÍHO INŽENÝRSTVÍ**

FACULTY OF MECHANICAL ENGINEERING

**ÚSTAV MATEMATIKY**

INSTITUTE OF MATHEMATICS

**MODERNÍ METODY RESTAURACE POŠKOZENÝCH  
AUDIOSIGNÁLŮ**

MODERN METHODS FOR RESTORATION OF DEGRADED AUDIOSIGNALS

**DIPLOMOVÁ PRÁCE**

MASTER'S THESIS

**AUTOR PRÁCE**

AUTHOR

**Bc. Ondřej Mokrý**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**doc. Mgr. Pavel Rajmic, Ph.D.**

**BRNO 2019**



# Zadání diplomové práce

Ústav:	Ústav matematiky
Student:	<b>Bc. Ondřej Mokrý</b>
Studijní program:	Aplikované vědy v inženýrství
Studijní obor:	Matematické inženýrství
Vedoucí práce:	<b>doc. Mgr. Pavel Rajmic, Ph.D.</b>
Akademický rok:	2018/19

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

## **Moderní metody restaurace poškozených audiosignálů**

### **Stručná charakteristika problematiky úkolu:**

Práce se týká problematiky restaurace audiosignálů, a to v případě znehodnocení výpadkem souvislého úseku vzorků. Práce bude využívat zejména metody založené na tzv. řídkosti signálů. V rámci práce budou navržena zlepšení stávajících metod za účelem zlepšení kvality rekonstrukce, jmenovitě např. zavedení tzv. analyzujícího modelu, váhování transformačního operátoru, váhování v časové oblasti, strukturovaná řídkost, nekonvexní optimalizace. Proběhne porovnání se staršími metodami pro restauraci signálů.

### **Cíle diplomové práce:**

Student prostuduje modely znehodnocení audiosignálu. Student analyzuje tento problém a seznámí se s obvyklými metodami pro jeho řešení, jejich vlastnostmi. Seznámí se rovněž s moderními metodami založenými na tzv. řídkosti signálů a navrhne související metodu restaurace a její modifikace. Metody poté implementuje v MATLABu a provede objektivní a subjektivní testy na reálných nahrávkách.

### **Seznam doporučené literatury:**

HRBÁČEK, R., RAJMIC, P., VESELÝ, V. a ŠPIŘÍK, J. Řídké reprezentace signálů: úvod do problematiky, Elektrevue, 2011. ISSN 1213-1539.

ŠPIŘÍK, J., RAJMIC, P. a VESELÝ, V. Reprezentace signálů: od bází k framům, Elektrevue, 2010. ISSN 1213-1539.

ADLER, A., EMIYA, V., JAFARI, M. G., ELAD, M., GRIBONVAL, R. a PLUMBLEY, M. D. Audio Inpainting. Audio, Speech, and Language Processing, IEEE Transactions on [online]. USA: IEEE, 1203, 20(3), DOI: 10.1109/TASL.2011.2168211. ISSN 1558-7916.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2018/19

V Brně, dne

L. S.

---

prof. RNDr. Josef Šlapal, CSc.  
ředitel ústavu

---

doc. Ing. Jaroslav Katolický, Ph.D.  
děkan fakulty

## ABSTRAKT

Diplomová práce se zabývá problémem doplňování souvislého úseku chybějících vzorků v digitálním audiosignálu. Tento problém je formulován jako optimalizační úloha, při níž v množině přípustných rekonstruovaných signálů hledáme takový, jehož reprezentace při vhodně zvolené transformaci je co nejvíce řídká. Popsány jsou různé konkrétní formulace, a to analyzující i syntetizující model jak pro konvexní, tak i pro nekonvexní úlohu. Pro uvedené formulace jsou navrženy algoritmy k jejich řešení a v konvexním případě je navíc metoda doplněna o různé postupy pro kompenzaci poklesu energie v doplněném úseku signálu. Všechny navržené algoritmy jsou otestovány na reálných nahrávkách a je ukázána jejich konkurenceschopnost ve srovnání se současným stavem problematiky (state-of-the-art).

## KLÍČOVÁ SLOVA

audiosignál, doplňování chybějících dat, časově-kmitočtové reprezentace, framy, řídké reprezentace, proximální operátory, konvexní a nekonvexní optimalizace, kompenzace poklesu energie

## ABSTRACT

The master's thesis deals with the problem of restoring a block of missing samples in a digital audio signal. This problem is formulated as an optimization task, which seeks the sparsest time-frequency representation of a signal within the set of feasible reconstructed signals. Several particular formulations are discussed, namely the analyzing and the synthesizing model, both for convex and non-convex approaches. Suitable algorithms are proposed for solving these formulations, and in the convex case, the method is further enhanced by various procedures to compensate for the energy drop in the inpainted signal segment. The proposed algorithms are tested on real recordings, and their performance is shown to be competitive with the state-of-the-art.

## KEYWORDS

audiosignal, inpainting, time-frequency representations, frames, sparse representations, proximal operators, convex and non-convex optimization, energy decrease compensation

MOKRÝ, Ondřej. *Moderní metody restaurace poškozených audiosignálů*. Brno, 2019, 75 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav matematiky. Vedoucí práce: doc. Mgr. Pavel Rajmic, Ph.D.



## PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Moderní metody restaurace poškozených audi-  
osignálů“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím  
odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny  
v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplo-  
mové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným  
způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom  
následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu  
autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský  
zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících  
z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno .....

.....  
podpis autora



## PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu doc. Mgr. Pavlu Rajmicovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci. Za méně odborné, leč neméně prospěšné poznámky děkuji také Ing. Janu Mokrému a Mgr. Tereze Lerchové. Všeobecný dík za příspěví k dalším detailům patří též Ing. Pavlu Záviškovi.

Brno .....

.....

podpis autora



# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Teoretické základy</b>	<b>7</b>
2.1	Potřebná teorie vektorových prostorů . . . . .	7
2.2	(Ne)ortogonální reprezentace . . . . .	8
2.2.1	Framy . . . . .	8
2.2.2	Gaborovy systémy a framy . . . . .	10
2.3	Konvexní optimalizace . . . . .	12
<b>3</b>	<b>Přehled metod pro audio inpainting</b>	<b>17</b>
3.1	Definice problému . . . . .	17
3.2	Modelování signálu jako AR procesu . . . . .	18
3.3	Metody založené na řídkosti . . . . .	20
3.4	Další metody . . . . .	21
3.4.1	Metody založené na soběpodobnosti . . . . .	21
3.4.2	Neuronové sítě . . . . .	23
<b>4</b>	<b>Konvexní metody založené na řídkosti</b>	<b>25</b>
4.1	Analyzující a syntetizující model $\ell_1$ relaxace . . . . .	25
4.1.1	Algoritmus pro řešení analyzujícího modelu . . . . .	26
4.1.2	Algoritmus pro řešení syntetizujícího modelu . . . . .	28
4.2	Modifikace modelu . . . . .	30
4.2.1	Váhování atomů . . . . .	30
4.2.2	Offset . . . . .	32
4.3	Modifikace v časové oblasti . . . . .	36
4.3.1	Váhování doplněného úseku v čase . . . . .	36
4.3.2	Gradual inpainting . . . . .	38
<b>5</b>	<b>Nekonvexní metody založené na řídkosti</b>	<b>41</b>
5.1	Orthogonal Matching Pursuit . . . . .	41
5.2	Sparse Audio Inpainter . . . . .	43
5.2.1	Princip algoritmu . . . . .	43
5.2.2	A-SPAIN . . . . .	45
5.2.3	S-SPAIN . . . . .	47
<b>6</b>	<b>Experimenty</b>	<b>51</b>
6.1	Hodnocení kvality rekonstrukce . . . . .	51
6.2	Testovací vzorky . . . . .	51
6.3	Schéma experimentu . . . . .	52
6.4	Konvexní metody . . . . .	52
6.4.1	Offset . . . . .	54
6.4.2	Váhování atomů . . . . .	55
6.4.3	Váhování doplněného úseku v čase . . . . .	57
6.4.4	Gradual inpainting . . . . .	58
6.5	Nekonvexní metody . . . . .	60
6.6	Celkové porovnání . . . . .	62

7 Závěr	65
Literatura	67
Seznam symbolů, veličin a zkratk	71
Seznam příloh	73
A Obsah přiloženého CD	75

# Seznam obrázků

2.1	Příklad atomů Gaborova framu. . . . .	11
2.2	Ukázka spektrogramu. . . . .	12
2.3	Ilustrace k definici polospojivosti zdola. . . . .	13
3.1	Ilustrace audio inpaintingu pro základní model poruchy. . . . .	17
3.2	Ilustrace křížového prolnutí. . . . .	19
3.3	Spektrogramy vybraných signálů. . . . .	20
3.4	Schematický zakres metody podobnostních grafů. . . . .	22
4.1	Ilustrace poklesu energie v doplněném úseku signálu. . . . .	26
4.2	Atomy Gaborova systému $\mathbf{g}^n$ , které zasahují do chybějícího úseku. . . . .	32
4.3	Ilustrace poklesu energie v doplněném úseku při použití offsetu. . . . .	33
4.4	Vizualizace Gaborova systému a značení. . . . .	34
4.5	Návrh variant offsetu pro Gaborův systém. . . . .	35
4.6	Ilustrace rozdílnosti jednotlivých variant offsetu na rekonstruovaném signálu. . . . .	35
4.7	Vizualizace postupu pro váhování doplněného úseku v čase. . . . .	38
5.1	Porovnání tvrdého prahování a OMP. . . . .	49
6.1	Testovací vzorky, první část. . . . .	53
6.1	Testovací vzorky, druhá část. . . . .	54
6.2	Experiment na offset. . . . .	55
6.3	Ilustrace variant váhování atomů. . . . .	56
6.4	Experiment na váhování atomů. . . . .	57
6.5	Vliv délky a počtu segmentů na křivku energie signálu. . . . .	58
6.6	Experiment na váhování doplněného úseku v čase. . . . .	59
6.7	Experiment na gradual inpainting. . . . .	60
6.8	Experiment na nekonvexní metody. . . . .	61
6.9	Porovnání S-SPAIN H a S-SPAIN OMP. . . . .	62
6.10	Ukázka vlivu váhování v časové oblasti. . . . .	63
6.11	Celkové porovnání metod. . . . .	64



# 1 Úvod

Zpracování digitálních signálů je v dnešní době aktuální problematikou. Signály, tedy z matematického pohledu číselné vektory určitých vlastností, potřebujeme zaznamenávat, zpracovávat, analyzovat, ale také přenášet a uchovávat, a to ať už se jedná o data vizuálního, zvukového či jiného charakteru.

V této práci se soustředíme právě na audiosignály a na problém jejich restaurace v případě specifického poškození. Nejčastěji se v literatuře setkáváme se třemi restauračními úlohami – denoisingem, declippingem a inpaintingem.

První jmenovaná úloha, odšumování, se zabývá situací, kdy je k signálu přičten náhodný šum (například vlivem přenosu v neideálním prostředí) a ten chceme odstranit. Declipping, též desaturace, se zabývá audiosignály, jejichž amplituda byla při záznamu či přenosu omezena takovým způsobem, že vzorky převyšující jistou povolenou úroveň jsou saturovány, tedy nabývají právě této omezující hodnoty. Poslední jmenovaný audio inpainting se zabývá doplňováním zcela chybějících vzorků. Tato situace může nastat při přenosu, kdy dochází ke ztrátám paketů, při digitalizaci poškozených nahrávek či přítomností impulzního šumu, který znehodnocuje prvky natolik, že je můžeme považovat za chybějící.

V této práci se budeme zabývat právě doplňováním chybějících vzorků. Jedná se o odvětví poměrně staré, první úspěšné metody byly publikovány již v 80. letech minulého století, za všechny jmenujme například Janssenův algoritmus v článku [20], jenž nebyl dosud přesvědčivě překonán. Článek [3] z roku 2012 byl později jedním z impulzů pro nový přístup, jímž je požadavek řídkosti audiosignálu ve vhodné reprezentaci. Motivací je fyzikální podstata hudebního tónu, který lze vyjádřit jako součet nemnoha harmonických funkcí odpovídajících základní frekvenci tónu a jeho harmonickým. Fourierova transformace audiosignálu<sup>1</sup> nám proto poskytuje řídký vektor koeficientů, tedy takový, který obsahuje nemnoho významných složek. Audio inpainting (jakožto i ostatní jmenované problémy) pak hledá řešení – rekonstruovaný signál – v množině signálů s řídkým spektrem, což se ukazuje být velmi účinným předpokladem.

Poznamenejme ještě, že tato práce se výhradně soustředí na problém audio inpaintingu. Popsané principy a algoritmy by však byly snadno adaptovatelné i pro další úlohy zpracování audiosignálů.

Členění práce je následující: V kapitole 2 se budeme věnovat matematickým základům využitým ve zbytku práce. Kapitoulou 3 začíná ústřední téma práce, jímž je doplňování chybějících vzorků audiosignálu. Po obecném přehledu dosavadních metod v kapitole 3 následují dvě kapitoly, v nichž jsou odvozeny a rozvinuty moderní metody založené na řídkosti. Konkrétně v kapitole 4 vycházíme z formulace úlohy doplňování vzorků jakožto konvexního optimalizačního problému, proto hovoříme o konvexních metodách. Naopak kapitola 5 se zabývá dvěma heuristickými algoritmy pro přibližné řešení nekonvexní optimalizační úlohy, tudíž hovoříme o metodách nekonvexních. Odvozené či popsání algoritmy jsou poté na reálných nahrávkách otestovány a vyhodnoceny, což popisuje kapitola 6.

---

<sup>1</sup>Dále popíšeme obdobnou, leč vhodnější transformaci.



## 2 Teoretické základy

Cílem této kapitoly je korektně zavést potřebné pojmy a také formulovat některá užitečná tvrzení pro pozdější odvozování algoritmů pro restauraci audiosignálů v kapitolách 4 a 5. Jednotlivé komentované pojmy se mohou jevit jako příliš nesouvisející, v dalších částech práce však budou mít svůj význam.

### 2.1 Potřebná teorie vektorových prostorů

Obecné pojmy jako vektorový prostor, norma či skalární (vnitřní) součin budeme považovat za obecně známé. V této části však definujeme některé konkrétní pojmy, které budou dále využívány, a také zavedeme dále používané značení.

**Definice 2.1.** *Nechť  $1 \leq p < \infty$ . Prostor  $\ell_p(\mathbb{N})$  definujeme jako prostor posloupností  $\mathbf{c} = \{c_n\}_{n \in \mathbb{N}}$ , pro které platí  $\sum_{n \in \mathbb{N}} |c_n|^p < \infty$ .*

**Definice 2.2.** *Jako normu v prostoru  $\ell_p(\mathbb{N})$ ,  $1 \leq p < \infty$ , zkráceně  $\ell_p$  normu, označujeme zobrazení přiřazující posloupnosti  $\mathbf{c} \in \ell_p(\mathbb{N})$  nezáporné číslo*

$$\|\mathbf{c}\|_p = \left( \sum_{n \in \mathbb{N}} |c_n|^p \right)^{1/p}. \quad (2.1)$$

**Poznámka 2.3.** Ačkoliv  $\ell_p$  normy jsou definovány na prostorech posloupností, značení přebíráme do Eukleidovských prostorů, kde suma v rovnici (2.1) obsahuje pouze konečný počet sčítanců, který odpovídá dimenzi prostoru.

**Definice 2.4.** *Jako řídkost, nebo též  $\ell_0$  pseudonormu, označujeme zobrazení, které posloupnosti  $\mathbf{c} = \{c_n\}_{n \in \mathbb{N}}$  přiřazuje počet jejích nenulových členů, tedy číslo*

$$\|\mathbf{c}\|_0 = |\{c_n \mid c_n \neq 0, n = 1, 2, \dots\}|, \quad (2.2)$$

kde  $|\cdot|$  značí mohutnost množiny.

**Poznámka 2.5.** Řídkost je někdy nazývána neformálně též  $\ell_0$  normou a odpovídá tomu i symbolické značení, neboť jistým způsobem navazuje na zavádění  $\ell_p$  norem. O normu se však nejedná, neboť obecně neplatí axiom  $\|k\mathbf{c}\|_0 = |k| \cdot \|\mathbf{c}\|_0$  pro nenulový skalár  $k$ . Proto hovoříme o pseudonormě.

Velký význam bude dále mít takzvaná váhovaná  $\ell_1$  norma, kterou zde proto zvlášť definujeme. Pro zjednodušení se však omezíme na konečněrozměrný prostor  $\mathbb{C}^N$ .

**Definice 2.6.** *Mějme dán prostor  $\mathbb{C}^N$  a na něm definovanou  $\ell_1$  normu. Dále necht  $\mathbf{w} \in \mathbb{R}^N$ , přičemž  $w_n > 0$ ,  $n = 1, 2, \dots, N$ . Váhovanou  $\ell_1$  normu (s váhami  $\mathbf{w}$ ) pak zavádíme jako zobrazení  $\mathbb{C}^N \ni \mathbf{c} \mapsto \|\mathbf{w} \odot \mathbf{c}\|_1$ , kde  $\odot$  značí násobení vektorů po složkách, tj.  $\|\mathbf{w} \odot \mathbf{c}\|_1 = \sum_{n=1}^N w_n |c_n|$ .*

**Poznámka 2.7.** Výhodou  $\ell_1$  normy je, že se na rozdíl od řídkosti jedná o konvexní funkci, což je stěžejní vlastnost pro její minimalizaci, k níž později dospějeme. V kontextu  $\ell_p$  norem je navíc řídkosti jistým způsobem nejbližší a existují dokonce tvrzení, která za určitých předpokladů zaručují ekvivalenci minimalizace řídkosti<sup>1</sup> a  $\ell_1$  normy vektoru. Blíže je toto téma zpracováno například v publikaci [33].

<sup>1</sup>Zde je situace jazykově mírně nejasná, neboť minimalizaci  $\ell_0$  pseudonormy bychom spíše nazvali hledáním maximálně řídkého vektoru, současně však chceme  $\ell_0$  pseudonormu také nazývat řídkostí. Dále se proto pojmu řídkost v souvislosti s minimalizací budeme spíše vyhýbat.

Protože budeme často pracovat s vektory a maticemi (reálných či komplexních čísel), zavádíme jejich značení po řadě následovně<sup>2</sup>:

$$\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^\top \in \mathbb{R}^{N \times 1} \text{ nebo } \mathbb{C}^{N \times 1}, \quad (2.3)$$

kde namísto  $\mathbb{R}^{N \times 1}$ , resp.  $\mathbb{C}^{N \times 1}$  píšeme pouze  $\mathbb{R}^N$ , resp.  $\mathbb{C}^N$ ,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M1} & a_{M2} & a_{M3} & \dots & a_{MN} \end{bmatrix} \in \mathbb{R}^{M \times N} \text{ nebo } \mathbb{C}^{M \times N}. \quad (2.4)$$

Kromě transpozice použité v zápisu vektoru používáme pro matice či vektory s komplexními prvky též hermitovskou transpozici, což pro matici  $\mathbf{A}$  značíme jako  $\mathbf{A}^*$ . Stejně značení používáme pro adjungované operátory<sup>3</sup>. Poznamenejme však, že u maticových operátorů, které s maticemi ztotožňujeme, se nejedná o konflikt značení, neboť hermitovská transpozice a adjungovanost jsou v tomto případě ekvivalentní.

## 2.2 (Ne)ortogonální reprezentace

V úvodu této části připomeneme pojem ortonormální báze, jejímuž zobecňování se budeme dále věnovat.

**Definice 2.8.** *Ortonormální bázi prostoru  $X$  se skalárním součinem  $\langle \cdot, \cdot \rangle$  je množina vektorů  $B = \{u_n\}_{n \in \mathbb{N}} \subset X$  taková, že generuje celý prostor  $X$  a platí*

$$\langle u_m, u_n \rangle = \begin{cases} 0 & \text{pro } m \neq n, \\ 1 & \text{pro } m = n, \end{cases} \quad (2.5)$$

tedy  $\|u_n\| = \sqrt{\langle u_n, u_n \rangle} = 1 \ \forall n \in \mathbb{N}$ .

V případě prostoru  $X$  nekonečné dimenze pod pojmem generování celého prostoru v předchozí definici rozumíme, že  $X$  je uzávěrem množiny všech (konečných) lineárních kombinací prvků  $B$ . V prostoru konečné dimenze  $N$  má přirozeně báze  $B$  právě  $N$  prvků a  $X$  je přímo množinou všech lineárních kombinací prvků  $B$ , v tomto případě tedy není nutný uzávěr.

### 2.2.1 Framy

Vybrané pojmy funkcionální analýzy budeme definovat v obecném Hilbertově prostoru  $H$ . Budeme tak činit v souladu s obvyklým zavedením, jak jej lze pro obecné Hilbertovy prostory nalézt například v článku [11] nebo ve skriptu funkcionální analýzy [37]. Pro aplikace je pak možné použít stejné definice, neboť používané reálné či komplexní Eukleidovské prostory jsou speciálním příkladem prostorů Hilbertových, pouze zde namísto posloupností vystupují vektory.

<sup>2</sup>Vektory dle konvence uvažujeme v tzv. sloupcovém tvaru, kdy se nejedná o uspořádanou  $N$ -tici, ale o matici rozměrů  $N \times 1$ .

<sup>3</sup>Na prostoru  $X$  se skalárním součinem  $\langle \cdot, \cdot \rangle$  pro operátory  $A$ ,  $A^*$  platí  $\langle Ax, y \rangle = \langle x, A^*y \rangle \ \forall x, y \in X$ .

**Definice 2.9** (podle [11]). *Nechť  $\{f_n\}_{n \in \mathbb{N}}$  je posloupnost prvků Hilbertova prostoru  $H$ . Tato posloupnost se nazývá*

(a) *frame, jestliže existují konstanty  $0 < A \leq B < \infty$  takové, že*

$$A \|f\|^2 \leq \sum_{n \in \mathbb{N}} |\langle f, f_n \rangle|^2 \leq B \|f\|^2 \quad \forall f \in H, \quad (2.6)$$

*kde  $A, B$  nazýváme mezemi framu a jednotlivé prvky  $f_n$  označujeme jako atomy;*

(b) *Rieszova báze, jestliže uzávěrem lineárního obalu  $\{f_n\}_{n \in \mathbb{N}}$  je celý prostor  $H$  a existují konstanty  $0 < A \leq B < \infty$  takové, že*

$$A \sum_{n \in \mathbb{N}} |c_n|^2 \leq \left\| \sum_{n \in \mathbb{N}} c_n f_n \right\|^2 \leq B \sum_{n \in \mathbb{N}} |c_n|^2 \quad (2.7)$$

*pro každou posloupnost  $\{c_n\}_{n \in \mathbb{N}}$  s pouze konečným počtem nenulových prvků.*

**Definice 2.10.** *Pokud pro daný frame existují meze  $A = B = \alpha$  splňující (2.6), nazýváme tento frame  $\alpha$ -těsným. Je-li navíc  $\alpha = 1$ , pak hovoříme o Parsevalově framu.*

V definici 2.9 je podstatné, že Rieszova báze je zobecněním ortonormální báze dle definice 2.8 a frame je dalším zobecněním Rieszovy báze. Pro aplikace ve zpracování signálů je pak specifickou vlastností framu, že může být systémem redundantním, tedy obsahovat více prvků, než je nutné k vygenerování celého prostoru. Reprezentace pomocí framů pak nejsou jednoznačné, což umožňuje požadování vybraných vlastností, například v této práci je to řídkost reprezentace.

S reprezentací také souvisí operátory analýzy a syntézy, které definujeme následovně:

**Definice 2.11** (podle [37]). *Syntetizujícím operátorem příslušným framu  $\{f_n\}_{n \in \mathbb{N}}$  rozumíme operátor  $D: \ell_2(\mathbb{N}) \rightarrow H$  definovaný předpisem*

$$D\mathbf{c} = \sum_{n \in \mathbb{N}} c_n f_n \quad \forall \mathbf{c} = \{c_n\}_{n \in \mathbb{N}} \in \ell_2(\mathbb{N}). \quad (2.8)$$

*Odpovídajícím analyzujícím operátorem je operátor  $A = D^*: H \rightarrow \ell_2(\mathbb{N})$ , který můžeme zapsat jako*

$$Af = \{\langle f, f_n \rangle\}_{n \in \mathbb{N}} \quad \forall f \in H. \quad (2.9)$$

Užitečnost Parsevalových framů se ukáže zejména v platnosti následujícího tvrzení:

**Věta 2.12.** *Pro Parsevalovy framy platí*

$$DA = \text{Id}, \quad AD = \mathcal{P}_{\mathcal{R}(A)}, \quad (2.10)$$

*kde jako Id značíme identitu (bez ohledu na prostor) a  $\mathcal{P}_{\mathcal{R}(A)}$  označuje operátor ortogonální projekce na obor hodnot  $\mathcal{R}(A)$  operátoru  $A$ . Dále*

$$\|Af\| = \|f\| \quad \forall f \in H, \quad (2.11a)$$

$$\|\mathbf{c}\| \geq \|D\mathbf{c}\| \quad \forall \mathbf{c} \in \ell_2(\mathbb{N}), \quad (2.11b)$$

$$\|\mathbf{c}\| = \|D\mathbf{c}\| \quad \forall \mathbf{c} \in \mathcal{R}(A). \quad (2.11c)$$

*Důkaz.* Platnost rovností (2.10) není zcela triviální a korektní důkaz by vyžadoval zavádění neúměrného množství dalších pojmů, proto odkazujeme na literaturu, například [37]. Dokážeme však vlastnosti (2.11).

V rovnici (2.11a) na levé straně máme prvek prostoru  $\ell_2(\mathbb{N})$ . Dosadíme tedy za  $Af$  vyjádření (2.9) a příslušnou normu rozepíšeme jako

$$\|Af\|^2 = \|\{\langle f, f_n \rangle\}_{n \in \mathbb{N}}\|_2^2 = \sum_{n \in \mathbb{N}} |\langle f, f_n \rangle|^2. \quad (2.12)$$

Protože se však jedná o Parsevalův frame, platí (2.6) pro  $A = B = 1$ , tedy  $\sum_{n \in \mathbb{N}} |\langle f, f_n \rangle|^2 = \|f\|^2$ , čímž dostáváme platnost (2.11a).

Platnost (2.11b) a (2.11c) dostaneme, pokud provedeme rozklad  $\mathbf{c} = \mathbf{a} + \mathbf{b}$ , kde  $\mathbf{a} \in \mathcal{R}(A)$  a  $\mathbf{b} \perp \mathcal{R}(A)$ . Potom platí

$$\|D\mathbf{c}\|^2 = \|AD\mathbf{c}\|^2 = \|\mathcal{P}_{\mathcal{R}(A)}(\mathbf{c})\|^2 = \|\mathbf{a}\|^2 \leq \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 = \|\mathbf{c}\|^2, \quad (2.13)$$

kde jsme v první a druhé úpravě využili po řadě vlastností (2.11a) a (2.10). Tím máme dokázánu nerovnost (2.11b), přičemž vidíme, že rovnost nastává právě pro  $\mathbf{c} = \mathbf{a} = \mathcal{P}_{\mathcal{R}(A)}(\mathbf{c})$ , tedy pro  $\mathbf{c} \in \mathcal{R}(A)$ , což dokazuje platnost (2.11c).  $\square$

## 2.2.2 Gaborovy systémy a framy

Speciálním případem framu může být tzv. *Gaborův systém*, který je založený na operátorech translace a modulace. Aplikacím ve zpracování signálů se přiblížíme tím, že tyto pojmy budeme definovat již v konkrétním Hilbertově prostoru  $L^2(\mathbb{R})$  komplexních funkcí reálné proměnné.

**Definice 2.13** (podle [11]). *Na prostoru  $L^2(\mathbb{R})$  definujeme pro  $a, b \in \mathbb{R}$  operátory translace a modulace po řadě následovně:*

$$T_a g(t) = g(t - a), \quad (2.14a)$$

$$M_b g(t) = \exp(2\pi i b t) \cdot g(t), \quad (2.14b)$$

kde  $g \in L^2(\mathbb{R})$  a  $t \in \mathbb{R}$ .

**Definice 2.14.** *Gaborův systém pro danou funkci  $g \in L^2(\mathbb{R})$  a parametry  $a, b > 0$  je množina*

$$\{M_{mb}(T_{na}g)\}_{m,n \in \mathbb{Z}}. \quad (2.15)$$

*Je-li navíc systém (2.15) frame, nazýváme jej Gaborovým framem. Funkci  $g$  pak nazýváme oknem Gaborova framu.*

Obecně není zcela prozkoumáno, jaké podmínky zaručují, že je Gaborův systém frame. Platí ovšem následující věta, která udává nutnou podmínku.

**Věta 2.15** (podle [11]). *Nechť  $a, b > 0$  a  $g \in L^2(\mathbb{R})$ . Příslušný Gaborův systém může být frame pro  $L^2(\mathbb{R})$ , pouze pokud platí  $ab \leq 1$ . Pokud je tento systém framem a platí  $ab = 1$ , pak je tento systém dokonce Rieszovou bází.*

V aplikační části práce budeme pracovat výhradně s diskrétními signály, tedy pouze s konečněrozměrnými vektorovými prostory  $\mathbb{C}^N$ ,  $N < \infty$ . Zde je skalární součin tvaru

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{n=1}^N u_n \overline{v_n} = \mathbf{v}^* \mathbf{u}. \quad (2.16)$$

Výše zavedené operátory syntézy a analýzy pak budou tvaru  $D: \mathbb{C}^P \rightarrow \mathbb{C}^N$  a  $A: \mathbb{C}^N \rightarrow \mathbb{C}^P$ , kde pro redundantní frame platí  $N < P$ . Můžeme je tedy ztotožnit s operátory maticovými nebo přímo maticemi, přičemž atomy framu budou sloupce matice odpovídající syntetizujícímu operátoru. Prostor  $\mathbb{C}^N$  konvenčně nazýváme časovou oblastí,  $\mathbb{C}^P$  je pak prostorem či oblastí koeficientů.

Speciálně zavedení Gaborova framu je v konečné dimenzi možné s tím rozdílem, že namísto proměnné  $t$  ve spojitém případě máme pouze konečný počet indexů. S indexy prvků proto počítáme modulárně vzhledem k délce vektoru.

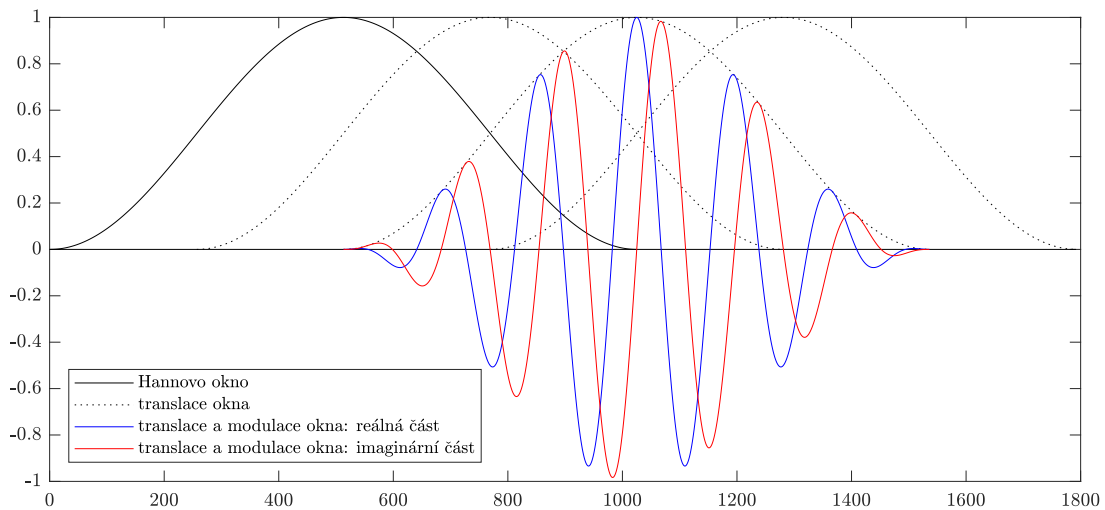
Pro bližší představu o Gaborově framu zařazujeme příklad jeho atomů. Jako funkci  $g$  uvažujeme Hannovo okno<sup>4</sup>, které (již v diskretní podobě, tedy jako vektor  $\mathbf{g}$ ) definujeme podle [17] jako

$$g_n = \sin^2\left(\frac{n-1}{w}\pi\right) \quad n = 1, \dots, w, \quad (2.17)$$

kde hodnotu  $w$  označujeme jako délku okna.

**Poznámka 2.16.** Můžeme se setkat i s alternativními definicemi Hannova okna. Například v toolboxu pro MATLAB LTFAT [31], který budeme hojně využívat v implementaci, je jako první prvek okna chápán jeho vrchol, což při definici dle rovnice (2.17) odpovídá prvku prostřednímu.

Vizualizaci Hannova okna a pomocí něj definovaných atomů Gaborova framu zobrazuje obrázek 2.1.



Obr. 2.1: Příklad atomů Gaborova framu. Zobrazeno je Hannovo okno  $\mathbf{g}$  délky  $w$  (které je samo jedním z atomů) a jeho vybrané translace a modulace. Pro přehlednost nejsou v obrázku označeny délka okna  $w$  (1024 vzorků) a posun sousedních oken  $a$  (256 vzorků).

## Gaborovy framy ve zpracování signálů

Význam Gaborových framů pro zpracování signálů je ten, že příslušný analyzující operátor realizuje tzv. krátkodobou Fourierovu transformaci (short-time Fourier transform, STFT, někdy též Gaborova transformace). Můžeme totiž tuto transformaci chápat jako aplikaci Fourierovy transformace na část signálu „vyříznutou“ oknem  $g$ . Získáváme tak

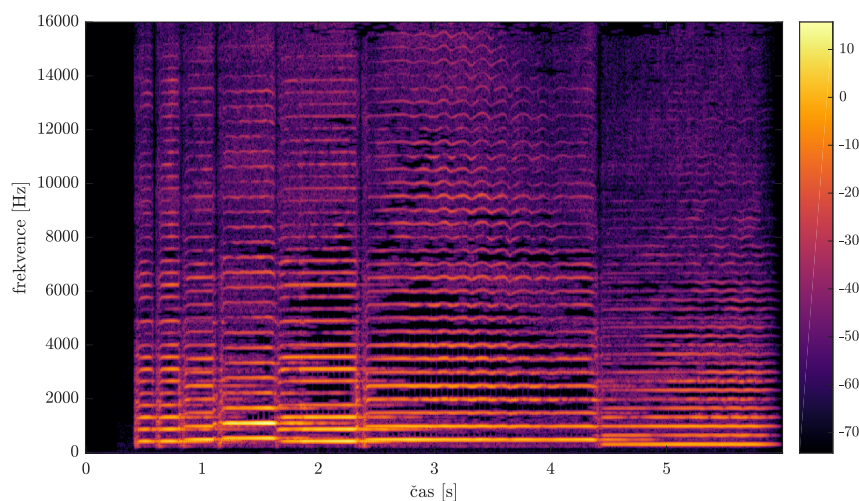
<sup>4</sup>Podle rakouského meteorologa Julia von Hanna. Setkat se můžeme též s označením Hanningovo okno.

koeficienty, které vyjadřují spektrum signálu v daném čase, proto hovoříme též o časově-frekvenční analýze. Na rozdíl od Fourierovy transformace pak samotný koeficient nese jistým způsobem lokální informaci, neovlivňuje však podobu celého signálu, což blíže odpovídá lidskému vnímání zvuku [33].

Navíc, jak již bylo řečeno, se v případě Gaborových framů jedná o nejednoznačnou reprezentaci signálu, což nám dále otevře možnost pro hledání řídkých reprezentací.

## Spektrogram

Pojem krátkodobé Fourierovy transformace přirozeně vybízí k zobrazení fourierovského spektra signálu, které však počítáme pouze pro určitý časový úsek, jak bylo popsáno. Právě takovýmto nástrojem je *spektrogram*, který zobrazuje vývoj spektra v závislosti na čase. Ukázka spektrogramu je na obrázku 2.2, na němž je zobrazen záznam zvuku houslí. Na vodorovné ose vidíme čas, na svislé frekvence a barva udává absolutní hodnotu příslušného koeficientu po Gaborově transformaci (škála pro tyto hodnoty je v dB).



Obr. 2.2: Ukázka spektrogramu.

Poznamenejme, že atomy Gaborova framu jsou komplexní (což je vidět z rovnic (2.14b) a (2.15)), tudíž i koeficienty jsou obecně komplexní. Obrázek 2.2 je proto takzvaným amplitudovým spektrogramem, neboť zobrazuje amplitudu (absolutní hodnotu) koeficientů. Obdobně bychom mohli vytvořit spektrogram fázový<sup>5</sup>. Dále v textu však budeme používat jednotný termín spektrogram, přičemž nejčastěji budeme mít na mysli amplitudový spektrogram, výjimečně pak přímo matici komplexních koeficientů.

## 2.3 Konvexní optimalizace

V této části zavedeme pojmy z oblasti konvexní optimalizace, které budeme dále používat, a uvedeme některá užitečná tvrzení. Začít přirozeně musíme definicí konvexní funkce:

**Definice 2.17.** *Nechť  $f: X \rightarrow \mathbb{R}$  je reálná funkce definovaná na vektorovém prostoru  $X$ . Funkci  $f$  nazýváme konvexní, jestliže platí*

$$\forall x_1, x_2 \in X, \forall t \in \langle 0, 1 \rangle: f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2). \quad (2.18)$$

<sup>5</sup>Fází zde myslíme argument (úhel) v kontextu goniometrického tvaru komplexního čísla.

Jestliže předchozí nerovnost je splněna ostře (vyjma hodnot  $t = 0$  a  $t = 1$ ), hovoříme o funkci striktně konvexní.

Nyní již můžeme zavést základní úlohu konvexní optimalizace, kterou je minimalizace konvexní funkce  $f$  (tzv. účelové funkce) na konvexní množině  $Y \subseteq X$ . V kontextu této práce pro nás není důležitá hodnota tohoto minima, ale především bod, ve kterém funkce této hodnoty nabývá, tedy takzvaný argument minima. Formálně tuto úlohu zapisujeme

$$\hat{x} = \arg \min_x f(x) \quad \text{vzhledem k } x \in Y \subseteq X. \quad (2.19)$$

Poznamenejme, že jestliže úloha (2.19) neobsahuje žádné omezení (tedy  $Y = X$ ) a funkce  $f$  je konvexní, ale nikoliv striktně konvexní, pak řešení nemusí být jediné. Dokonce máme-li dva různé body  $\hat{x}_1, \hat{x}_2$ , v nichž funkce  $f$  nabývá svého minima, pak v důsledku nerovnosti (2.18) nabývá této hodnoty i na všech bodech úsečky s krajními body  $\hat{x}_1$  a  $\hat{x}_2$ . Je-li však funkce  $f$  striktně konvexní, snadno bychom sporem ukázali, že minima může nabývat v nejvýše jednom bodě.

V případě více řešení zřejmě zápis (2.19) není korektní, protože řešením úlohy je množina a  $\hat{x}$  je jejím prvkem. Jako zjednodušení však užíváme jednotný zápis a v případě více řešení rozumíme zápisem (2.19) vybrání některého z nich.

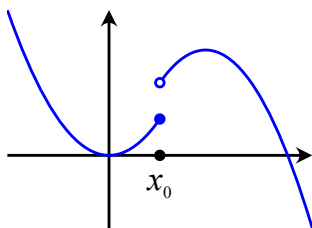
Následují definice některých pro aplikace stěžejních pojmů konvexní optimalizace.

**Definice 2.18.** Indikátorovou funkci  $\iota_Y(x)$  obecné podmnožiny  $Y$  prostoru  $X$  definujeme předpisem

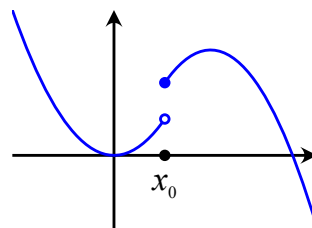
$$\iota_Y(x) = \begin{cases} 0 & \text{pro } x \in Y, \\ \infty & \text{pro } x \in X \setminus Y. \end{cases} \quad (2.20)$$

**Definice 2.19** (podle [13]). Reálnou funkci  $f$  na prostoru  $X$  nazýváme zdola polospojitou v bodě  $x_0 \in X$ , jestliže  $\forall \varepsilon \in (-\infty, f(x_0))$  existuje okolí  $\Omega$  bodu  $x_0$  takové, že  $\forall x \in \Omega: f(x) \in (\varepsilon, \infty)$ . Je-li funkce zdola polospojita v každém bodě prostoru  $X$ , hovoříme o funkci zdola polospojité.

Význam předchozí definice si můžeme představit na příkladu funkce, která má jediný bod nespojitosti  $x_0$ , což vidíme na obrázku 2.3. Funkce na obrázku 2.3a je v bodě  $x_0$  zdola polospojita – jestliže si představíme  $\varepsilon$  jako hladinu nižší než  $f(x_0)$ , pak zřejmě funkční hodnoty na dostatečně malém okolí  $x_0$  leží nad touto hladinou. To však není případ funkce na obrázku 2.3b, kde si stačí představit  $\varepsilon$  ležící mezi limitou funkce zleva a zprava v bodě  $x_0$ . Poznamenejme také, že spojitá funkce  $f$  je vždy též polospojita.



(a) funkce, která je zdola polospojita



(b) funkce, která není zdola polospojita

Obr. 2.3: Ilustrace k definici polospojivosti zdola. Převzato z <https://en.wikipedia.org/wiki/Semi-continuity>.

**Definice 2.20.** Necht  $Y$  je neprázdná uzavřená konvexní podmnožina prostoru  $X$ , na němž je definována norma. Projekcí na množinu  $Y$  rozumíme zobrazení  $\mathcal{P}_Y: X \rightarrow X$  dané pro každé  $x \in X$  předpisem

$$x \mapsto \arg \min_{y \in Y} \|y - x\|. \quad (2.21)$$

Je-li speciálně  $Y$  podprostorem  $X$ , hovoříme o ortogonální projekci.

**Definice 2.21.** Necht  $f: X \rightarrow \mathbb{R}$  je konvexní funkce na prostoru  $X$ , na němž je definována norma. Proximálním operátorem funkce  $f$  rozumíme zobrazení  $\text{prox}_f: X \rightarrow X$  dané pro každé  $x \in X$  předpisem

$$x \mapsto \arg \min_{y \in X} f(y) + \frac{1}{2} \|y - x\|^2. \quad (2.22)$$

Díky tomu, že v předchozí definici uvažujeme konvexní funkci  $f$  v součtu s kvadrátem normy, což je funkce striktně konvexní, řešení minimalizační úlohy (2.22) je jediné. Proximální operátor je tudíž dobře definován.

V aplikační části práce, zejména v kapitole 4, budeme hojně využívat proximální operátory dvou funkcí – (váhované)  $\ell_1$  normy a indikátorové funkce konvexní množiny. Formou lemmatu proto tyto operátory zavedeme.

**Lemma 2.22.** Proximálním operátorem váhované  $\ell_1$  normy s váhami  $\mathbf{w}$  je operátor

$$\mathcal{S}_{\mathbf{w}}(\mathbf{x}) = \arg(\mathbf{x}) \odot \max(|\mathbf{x}| - \mathbf{w}, 0), \quad (2.23)$$

kde argument (komplexního čísla), maximum i absolutní hodnotu vnímáme jako operace po složkách. Tento operátor nazýváme měkké prahování.

Měkké prahování (v anglické literatuře *soft thresholding*) jsme zde zavedli pro vektorový parametr  $\mathbf{w}$ . Pokud bychom uvažovali neváhovanou  $\ell_1$  normu, či přesněji její  $\tau$ -násobek, dostali bychom známější měkké prahování s (konstantním) prahem  $\tau$  [12]. Důkaz lemmatu 2.22 vynecháváme, nalézt jej lze například v [27].

**Lemma 2.23.** Necht  $Y$  je neprázdná uzavřená konvexní podmnožina prostoru  $X$ , na němž je definována norma. Proximálním operátorem funkce  $\iota_Y$  pak je operátor  $\mathcal{P}_Y$ .

*Důkaz.* Zvolme  $x \in X$  jako libovolný pevný bod a zavedme funkci  $g: X \rightarrow \mathbb{R}$  jako  $g(y) = \iota_Y(y) + \frac{1}{2} \|y - x\|^2$ . Zřejmě hledaný proximální operátor je zobrazení  $x \mapsto \arg \min_{y \in X} g(y)$ . Současně však po rozepsání indikátorové funkce platí

$$g(y) = \begin{cases} \frac{1}{2} \|y - x\|^2 & \text{pro } y \in Y, \\ \infty & \text{pro } y \notin Y. \end{cases} \quad (2.24)$$

Protože požadujeme neprázdnou množinu  $Y$ , alespoň pro nějaká  $y \in X$  nastává první možnost, proto druhou možnost vůbec nemusíme při minimalizaci uvažovat a platí

$$x \mapsto \arg \min_{y \in Y} g(y) = \arg \min_{y \in Y} \frac{1}{2} \|y - x\|^2. \quad (2.25)$$

Protože však argument minima nezávisí na násobení kladnou konstantou a nemění jej ani druhá mocnina normy, dostáváme tímto definici projekce 2.20 a důkaz je hotov.  $\square$

Následující definice existuje v mnohem obecnějším tvaru, který by však vyžadoval zavedení dalších pojmů, které bychom dále nevyužili. Omezíme se tedy na případ Hilbertova prostoru. Jako  $\overline{\mathbb{R}}$  značíme množinu rozšířených reálných čísel, tj.  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ .

**Definice 2.24** (podle [2]). *Nechť  $H$  je Hilbertův prostor se skalárním součinem  $\langle \cdot, \cdot \rangle$  a  $f: H \rightarrow \overline{\mathbb{R}}$ . Funkci konvexně sdruženou<sup>6</sup> k funkci  $f$  značíme  $f^*$  a definujeme*

$$f^*(x) = \sup_{u \in H} \langle x, u \rangle - f(u). \quad (2.26)$$

**Lemma 2.25.** *Nechť  $f$  je konvexní funkce a  $\alpha > 0$ . Pak platí*

$$\text{prox}_{\alpha f^*}(x) = x - \alpha \text{prox}_{f/\alpha}(x/\alpha). \quad (2.27)$$

Předchozí tvrzení zde nebudeme dokazovat, pouze zmíníme, že se jeho platnost opírá o tzv. Moreauovu identitu [22, 13].

Proximální operátory jsou základním stavebním prvkem (iterativních) proximálních algoritmů, které se ukazují být efektivním nástrojem pro řešení úloh konvexní optimalizace. V této práci budeme využívat dva takovéto algoritmy, konkrétně Douglasův-Rachfordův a Chambollův-Pockův algoritmus [12, 10]. Pro lepší návaznost textu jsou konkrétní tvary těchto algoritmů i oblast jejich možného použití uvedeny až v kapitole 4, kde jsou využity.

Mezi další algoritmy pro konvexní optimalizaci patří tzv. *Alternating Direction Method of Multipliers* (ADMM) [8]. I tento postup bude přiblížen až dále v této práci, konkrétně v kapitole 5.

---

<sup>6</sup>Kromě termínu konvexní sdružení se užívá též pojmu Fenchelova–Legendrova transformace.

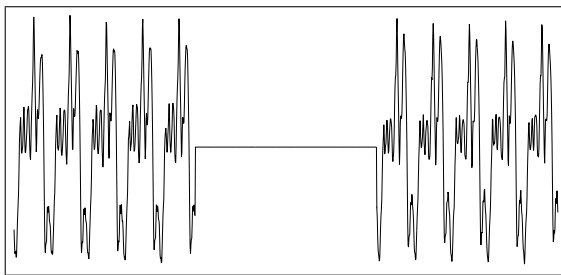


## 3 Přehled metod pro audio inpainting

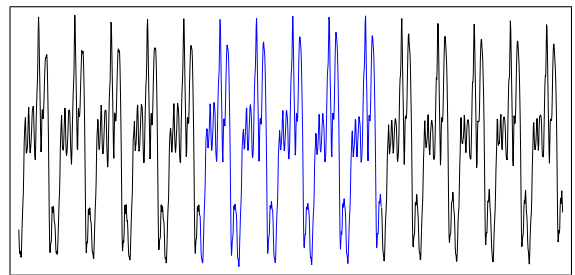
### 3.1 Definice problému

Doplňování chybějících vzorků audiosignálu je jedním z klasických problémů číslicového zpracování signálů. Po vzoru anglické literatury jej budeme dále označovat jako (*audio*) *inpainting*. Výchozím bodem je pro nás digitalizovaný audiosignál, jehož některé vzorky chybí a pozice těchto vzorků jsou předem známy. Cílem konzistentního zpracování je doplnit tyto vzorky při zachování neporušených částí vstupního signálu tak, aby doplněná místa v restaurovaném signálu nebyla sluchem rozpoznatelná. Naším cílem však není pouze zamaskovat vzniklé díry, ale máme ambice doplnit informaci co nejbližší původnímu (nepoškozenému) signálu. Poznamenejme, že při experimentech budeme vycházet z neporušených signálů, ve kterých uměle vytvoříme chybějící úseky, proto bude možné splnění těchto cílů hodnotit.

Struktura chybějících dat může být různá. My budeme především uvažovat chybějící vzorky koncentrované do jednoho delšího úseku, který budeme označovat jako *díru*. Tuto situaci ilustruje obrázek 3.1. Dalším obvykle používaným modelem poruchy je náhodná ztráta jednotlivých vzorků s danou pravděpodobností, tomuto modelu však bude věnována menší pozornost; mimo jiné proto, že se jedná o podstatně méně náročnou úlohu.



(a) pozorovaný signál, chybějící úsek je zobrazen jako nulový



(b) restaurovaný signál, zvýrazněný je doplněný úsek

Obr. 3.1: Ilustrace audio inpaintingu pro základní model poruchy, kterým je jediný souvislý úsek chybějících vzorků. Jedná se o záznam zvuku houslí a vidíme, že chybějící úsek má délku jednotek period, což odpovídá řádově desítkám milisekund. Zobrazeno je pouze malé okolí díry.

Pro úplnost budeme zmiňovat také metody, které nejsou konzistentní, tedy dovolují odchylku řešení od pozorovaných dat. Jednotně však budeme značit:

- $\mathbf{y}$  pozorovaný (poškozený) signál,  $\mathbf{y} \in \mathbb{C}^N$ ,
- $M_R$  operátor, který na pozice chybějících vzorků dosadí nuly,  $M_R: \mathbb{C}^N \rightarrow \mathbb{C}^N$ ,
- $\mathbf{y}_R$  nepoškozená část pozorovaného signálu doplněná nulami,  $\mathbf{y}_R = M_R \mathbf{y}$ ,
- $\hat{\mathbf{x}}$  řešení problému audio inpaintingu v časové oblasti,  $\hat{\mathbf{x}} \in \mathbb{C}^N$ ,
- $\Gamma$  množina přípustných řešení,  $\Gamma = \{\mathbf{x} \in \mathbb{C}^N \mid M_R \mathbf{x} = \mathbf{y}_R\} \subset \mathbb{C}^N$ .

Poznamenejme, že signály budeme uvažovat komplexní, což nám dále zjednoduší práci s některými operátory. V praxi se však omezujeme na signály reálné, proto nepsaným, ale přirozeným požadavkem je nulová imaginární část  $\hat{\mathbf{x}}$ . Nejjobecnější formulaci konzistent-

ního audio inpaintingu je pro nás tedy

$$\text{najít vhodné } \hat{\mathbf{x}} \text{ takové, že } \hat{\mathbf{x}} \in \Gamma, \text{ tedy } M_{\mathbf{R}}\hat{\mathbf{x}} = \mathbf{y}_{\mathbf{R}}. \quad (3.1)$$

Když mluvíme o formulaci nekonzistentní, rozumíme tím nahrazení rovnosti v definici množiny  $\Gamma$  pouze přibližnou rovností. Formulace (3.1) zřejmě neposkytuje informaci o tom, co je vhodné řešení, ani jak jej najít. Níže proto uvedeme přehled některých možných přístupů, které kladou na hledaný signál  $\hat{\mathbf{x}}$  dodatečné požadavky tak, aby se úloha (3.1) stala jednoznačně řešitelnou.

## 3.2 Modelování signálu jako AR procesu

Modelování signálu jako autoregresního (AR) stochastického procesu je jedna z nejstarších metod pro audio inpainting. Vychází z pozorování, že AR procesy zdárně modelují harmonické chování hudebních signálů či lidský hlas. Právě harmonické složky signálu v okolí díry nám poskytují informaci, kterou lze doplnit do chybějícího úseku, neboť je-li tento dostatečně krátký, můžeme předpokládat, že jeho spektrum bude obdobné jako spektrum sousedních (nepoškozených) úseků.

AR modelování lze využít více způsoby. V obecném případě modelujeme signál  $\mathbf{x}$  pomocí koeficientů  $\mathbf{a}$ , které dále uvažujeme řádu  $p$ , tedy zjednodušeně

$$x_n + a_1x_{n-1} + \dots + a_px_{n-p} = e_n, \quad (3.2)$$

kde  $\mathbf{e}$  je vektor odchylek AR modelu od signálu  $\mathbf{x}$ . Poznamenejme, že předchozí rovnice, ač to z jejího tvaru není ihned patrné, předepisuje prvek  $x_n$  jako lineární kombinaci  $p$  předchozích prvků, a to až na chybový člen  $e_n$ .

Nejstarší využití AR modelu pro audio inpainting publikovali autoři Janssen, Veldhuis a Vries v roce 1986 v článku [20]. Cílem jejich algoritmu je řešit optimalizační úlohu

$$(\hat{\mathbf{a}}, \hat{\mathbf{x}}) = \arg \min_{\mathbf{a}, \mathbf{x}} Q(\mathbf{a}, \mathbf{x}), \quad (3.3)$$

kde  $Q(\mathbf{a}, \mathbf{x}) = \|\mathbf{e}\|_2$ .

Úloha (3.3) není jednoduše řešitelná, neboť účelová funkce je polynomem čtvrtého stupně – obsahuje členy typu  $a_i^2 x_j^2$ , kde  $a_i$  i  $x_j$  jsou neznámé. Proto autoři navrhují přibližné řešení, ve kterém dvojici  $(\mathbf{a}^{(k+1)}, \mathbf{x}^{(k+1)})$  definují lineární rovnice, jejichž matice i pravé strany závisí na  $(\mathbf{a}^{(k)}, \mathbf{x}^{(k)})$ . Jedná se tedy o iterativní algoritmus (dále budeme hovořit zjednodušeně o Janssenově algoritmu), u kterého není sice zaručeno, že konverguje k řešení úlohy (3.3), v [20] je však ukázáno, že hodnota objektivní funkce  $Q$  se těmito iteracemi snižuje.

Popsaná metoda se vyznačuje tím, že v každé iteraci pracuje s AR koeficienty celého signálu. Alternativní přístup lze nalézt například v [16], kde je prezentován jednorázový (neiterativní) postup pro doplnění kompaktního úseku chybějících vzorků, který můžeme stručně shrnout třemi kroky:

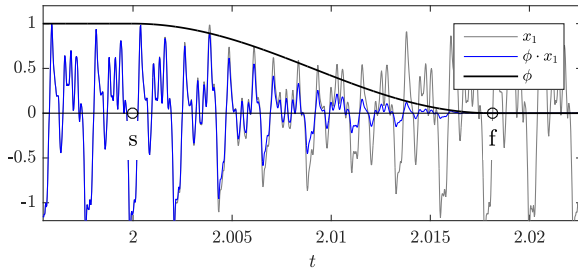
1. vypočítáme AR koeficienty z levého a pravého okolí díry, přičemž na pravé straně pracujeme s časově invertovaným signálem,
2. pomocí těchto koeficientů predikujeme doplnění vzorků zleva a zprava do díry,
3. provedeme křížové prolnutí řešení zleva a zprava (komentář k tomuto obecnému principu je v poznámce 3.1).

Výhodou této metody je, že dokáže částečně zachytit situaci, kdy se signál nalevo a napravo od díry od sebe liší. Nevýhodou pak je, že chybějící úsek musí být obklopen dostatečně dlouhými neporušenými úseky signálu.

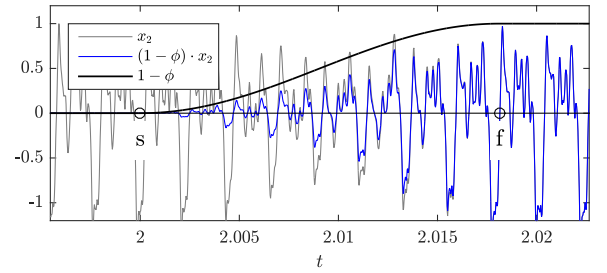
Mezi metody založené na AR modelování můžeme zařadit také restauraci dlouhých výpadků signálu pomocí sinusoidálního modelování, která byla navržena v [24] a zde ji pouze okrajově zmíníme. Při tomto přístupu syntetizujeme signál pomocí sinusoid závislých na třech parametrech – amplitudě, frekvenci a fázi. Základní myšlenka přístupu popsaného v [24] je využití AR modelu nikoliv k predikci chybějících vzorků v časové oblasti, jak je tomu ve dvou výše popsaných metodách, ale k predikci parametrů sinusoid v průběhu chybějícího úseku.

**Poznámka 3.1** (křížové prolnutí). Křížové prolnutí, v anglické literatuře *crossfade*, je obecný postup kombinace dvou signálů. Signály v tuto chvíli chápeme jako spojitě reálné či komplexní funkce reálné proměnné. Využívá se v situaci, kdy na intervalu, který pracovním označíme  $\langle s, f \rangle$ , jako je tomu v obrázku 3.2, chceme plynule přejít od signálu  $x_1$  k signálu  $x_2$ . Předpokládáme tedy, že mimo interval  $\langle s, f \rangle$  je výsledný signál  $x(t)$  definován

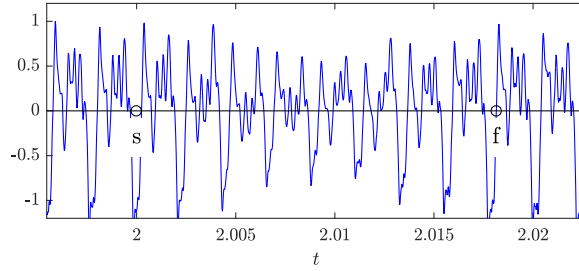
$$x(t) = \begin{cases} x_1(t) & \text{pro } t < s, \\ x_2(t) & \text{pro } t > f. \end{cases} \quad (3.4)$$



(a) signál  $x_1$  a funkce  $\phi$



(b) signál  $x_2$  a funkce  $1 - \phi$



(c) kombinace signálů z obrázků 3.2a a 3.2b křížovým prolnutím

Obr. 3.2: Ilustrace křížového prolnutí. Kombinovány jsou dva návrhy doplnění úseku mezi body  $s$  a  $f$ , zleva (signál  $x_1$ ) a zprava (signál  $x_2$ ).

Zbývá určit, jak volit kombinaci  $x_1$  a  $x_2$  na intervalu  $\langle s, f \rangle$ . Důležité je povšimnout si, že kdybychom mezi  $x_1$  a  $x_2$  přepínali v konkrétní časový okamžik, nemusí být výsledný signál spojitý, což je nežádoucí. Budeme tedy uvažovat, že signál  $x(t)$  je dodefinován jako

$$x(t) = \phi(t) \cdot x_1(t) + (1 - \phi(t)) \cdot x_2(t), \quad s \leq t \leq f, \quad (3.5)$$

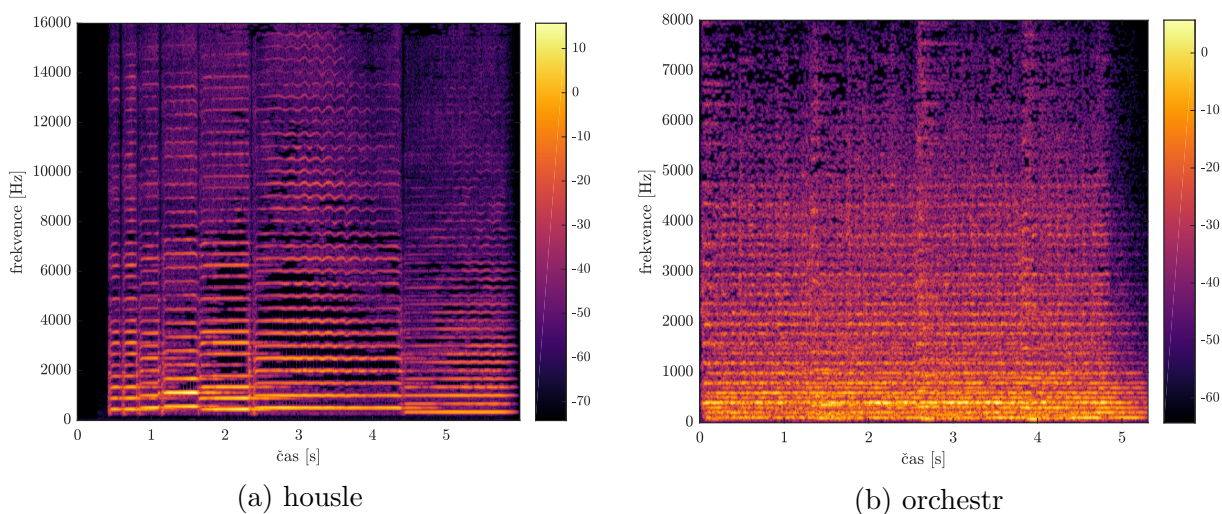
kde funkce  $\phi: \langle s, f \rangle \rightarrow \langle 0, 1 \rangle$  má následující vlastnosti:

$$\phi(s) = 1, \quad \phi(f) = 0, \quad \phi \text{ je spojitá a nerostoucí.} \quad (3.6)$$

Všechny tři požadavky jsou přirozené, neboť požadujeme, aby přechod mezi  $x_1$  a  $x_2$  byl plynulý. Navíc zřejmě funkce, kterými násobíme signály  $x_1$  a  $x_2$ , jsou v součtu konstantní jednička, tedy křížovým prolutím nepřidáváme ani neubíráme žádnou energii. Pro co nejhladší přechod ještě můžeme dále vyslovit podmínku, aby byla funkce  $\phi$  hladká (nebo dokonce nekonečně diferencovatelná) na  $(s, f)$  a v bodech  $s$  a  $f$  měla nulové jednostranné derivace. Právě použití takovéto funkce ilustruje obrázek 3.2, kde je v pozici  $\phi$  použita část periody druhé mocniny funkce kosinus.

### 3.3 Metody založené na řídkosti

Modernější třídou metod pro restaurace audiosignálů jsou metody založené na řídkosti. Řídkost se zde objevuje ve chvíli, kdy pomocí vhodného operátoru provedeme analýzu audiosignálu a získáme vektor koeficientů s velkým počtem nulových (nebo málo významných) složek. Ukázkou této situace je obrázek 3.3, který dva různé signály zobrazuje jako spektrogramy, tedy zachycuje vývoj jejich fourierovského spektra v čase. Zejména na obrázku 3.3a vidíme, že spektru dominuje malý počet výrazných složek – jedná se o tóny houslí; v dolní části spektrogramu je základní frekvence tónu a další výrazné složky spektra jsou frekvence vyšších harmonických. Obrázek 3.3b již zachycuje komplexnější signál, konkrétně záznam orchestru. Harmonickou strukturu (výraznější horizontální linie) vidíme i v tomto signálu, ve srovnání se záznamem houslí se však jedná o signál podstatně méně řídký.



Obr. 3.3: Spektrogramy dvou signálů – záznamu samotných houslí a orchestru.

S využitím řídkosti pro bližší specifikaci základního problému (3.1) přišel kolektiv autorů v čele s Adlerem v článku [3]. Zde použitou formulaci získáme z (3.1) tak, že řešení hledáme takové, jehož koeficienty ve vhodně zvolené reprezentaci (jako je například Gaborův systém) jsou nejřidší možné. Formálně tedy autoři formulovali úlohu audio inpaintingu následovně:

$$\hat{\mathbf{z}} = \arg \min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{vzhledem k} \quad \|\mathbf{y}_R - M_R D \mathbf{z}\|_2^2 \leq \varepsilon. \quad (3.7)$$

Oproti dosud zavedenému značení zde navíc vystupuje syntetizující operátor  $D$ , který může odpovídat například násobení maticí obsahující jako sloupce právě Gaborův systém. Poznamenejme, že řešením formulace (3.7) jsou koeficienty signálu, samotný restaurovaný

signál bychom tudíž získali jako  $\hat{\mathbf{x}} = D\hat{\mathbf{z}}$ . Navíc ještě dodejme, že oproti naší formulaci (3.1) je zde rozdíl v tom, že nepožadujeme rovnost řešení a nepoškozené části vstupního signálu  $\hat{\mathbf{z}}$ , ale postačuje nám dostatečně malá norma rozdílu, limitovaná v případě zápisu (3.7) kladným parametrem  $\varepsilon$ .

Ačkoliv formulace (3.7) není jediný způsob, jak do úlohy audio inpaintingu integrovat řídkost (jak uvidíme v následujících částech), věnujme jí nyní více pozornosti. Jedná se o takzvaný *syntetizující model*, právě proto, že formulace obsahuje syntetizující operátor  $D$ . Alternativní *analyzující model* bude představen níže v části 4.1. Důležitou vlastností je, že úloha (3.7) je NP-těžký problém, řešitelný pouze přibližně – v [3] byl využit algoritmus *Orthogonal Matching Pursuit* (OMP), kterému se bude více věnovat část 5.1.

Na tomto místě ještě zmiňme, že existují i složitější formy řídkosti, které mohou lépe zachycovat strukturu audiosignálu, takzvaná strukturovaná řídkost. Tento aparát se snaží zachytit například jev pozorovaný v obrázku 3.3, kdy signál po analýze obsahuje výrazné skupiny koeficientů s jasně patrnou strukturou, v tomto případě se po vizualizaci jedná o horizontální linie. Intuitivně tedy můžeme formulovat úlohu tak, že konkrétnímu koeficientu přisuzujeme větší váhu, pokud odpovídající v čase předcházející koeficient byl významný. Bližší rozbor aparátu strukturované řídkosti lze nalézt například v článku [23], využití speciálně pro audio inpainting je pak například v [25].

Okrajově ještě zmiňme jiný způsob využití struktury audiosignálu, kterým je neložkální řídkost. Tento koncept, prezentovaný teprve nedávno v [36], má za cíl v prvním kroku vyhledat podobné úseky signálu a následně nevyžadovat podobnost koeficientů ve dvou bezprostředně následujících úsecích, jak bylo výše naznačeno, ale právě u nalezených podobných úseků, které mohou být z pohledu času v signálu více vzdálené.

Audio inpainting založený na řídkosti je hlavním tématem této práce, budeme se mu tedy věnovat podrobněji v kapitolách 4 a 5. Konkrétně kapitola 4 je věnována tzv.  $\ell_1$  relaxaci, kdy není cílem minimalizovat počet nenulových složek vektoru koeficientů (jak je tomu například ve formulaci (3.7)), ale  $\ell_1$  normu vektoru koeficientů. Tím dostaneme úlohu konvexní optimalizace, pro jejíž řešení existují efektivní numerické nástroje. Kapitola (5) je pak věnována bližšímu rozboru dvou nekonvexních metod, a to již zmíněného OMP a nového algoritmu SPAIN.

## 3.4 Další metody

### 3.4.1 Metody založené na soběpodobnosti

V této části představíme zcela odlišný přístup k audio inpaintingu, než bylo prezentováno dosud. Následující dvě metody totiž využívají k doplnění informace do díry v signálu neporušené úseky téhož signálu, které jsou poškozenému místu v jistém smyslu podobné a mohou tudíž být (s případnými drobnými úpravami) použity namísto úseku poškozeného. Také se liší primární cíl, kterým je spíše díru zamaskovat, nežli co nejpřesněji doplnit chybějící informaci. Tento přístup však přináší výhodu v možnosti doplňování děr délky i v řádu sekund.

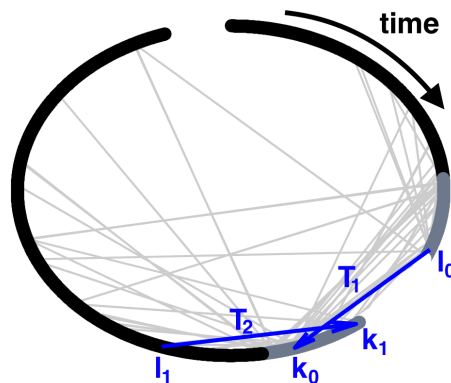
První metoda, označená v originálním článku [4] jako *self-content-based audio inpainting*, pracuje se signálem rozděleným na bloky stejné délky. Každý blok je transformován na vektor charakteristik (v původním textu *feature vector*), který audiosignál vhodně reprezentuje – bližší popis lze najít v dodatku článku [4]. V případě, že se v signálu objeví blok s chybějícími vzorky, je k němu v nepoškozené části signálu dohledán blok podobný, a to na základě dvou kritérií:

1. blízkost v prostoru vektorů charakteristik,
2. vysoká pravděpodobnost vzniklé sekvence audio bloků.

Druhé kritérium znamená, že po nahrazení bloku poškozeného blokem vybraným dostaneme sekvenci, která má jako celek na základě analýzy zbytku signálu vysokou pravděpodobnost výskytu. Výběrem bloku však autoři nekončí – navrhují dále zpracování vybraného bloku tak, aby se ještě více přiblížil novému kontextu (tedy okolí díry). Konkrétně navrhují úpravu frekvence a fáze, pokud se jedná o hudební signál nebo samohlásku v řeči, úpravu intenzity a v závěru také volbu vhodného napojení nového bloku na okolí díry. Zde se jedná o skutečnost, že bloky v optimálním případě volíme delší než chybějící úsek, tudíž můžeme ovlivnit, jak velkou část vybraného bloku chceme využít (počínaje délkou díry a konče délkou celého bloku).

Druhou metodou založenou na soběpodobnosti je audio inpainting s využitím *podobnostních grafů*, prezentovaný v článku [30]. Zásadním rozdílem oproti předchozí metodě je předem neurčená délka vybíraného úseku, neboť zde se nepracuje s audio bloky pevné délky. Cílem je najít v neporušené části signálu takový úsek (ne nutně odpovídající délce díry), kterým nahradíme úsek porušený a pro který bude především splněna podmínka, že po jeho zkopírování na místo porušeného úseku nebude rozpoznatelný ani začátek, ani konec doplněné části. Právě pro zajištění přirozeného přechodu potřebujeme využít podobnost jednotlivých míst v signálu, kterou autoři vyhodnocují na základě spektra okolí daného vzorku a s využitím derivace fáze. Takováto podobnost je pak reprezentována grafem, jehož uzly jsou jednotlivé vzorky signálu a hrany jsou ohodnoceny právě bezrozměrnou podobností. Následně s pomocí grafu hledáme takové dvojice vzorků  $(l_0, k_0)$  a  $(l_1, k_1)$  splňující podmínky:

1.  $k_0 < l_1$ ,
2.  $l_0$  je poblíž začátku chybějícího úseku,  $k_1$  je poblíž jeho konce,
3. platí  $k_1 - l_0 \approx l_1 - k_0$ ,
4. ohodnocení hran  $(l_0, k_0)$  a  $(l_1, k_1)$  je vysoké.



Obr. 3.4: Schematický záznam metody podobnostních grafů. Obvodová linie zde symbolizuje signál, šedé úseky jsou blízké okolí díry (díra samotná je zobrazena pouze jako přerušení černé linie signálu). Tětiny jsou vybrané hrany grafu, vyznačující místa s vysokou podobností. Obrázek je převzatý z [30].

Význam hodnot  $k_0, k_1, l_0, l_1$  ilustruje obrázek 3.4. Poškozený úsek, který se nachází mezi body  $l_0$  a  $k_1$ , nahradíme tak, že v místě  $l_0$  dojde k přechodu na místo  $k_0$ , pokračujeme

neporušeným signálem a poté z místa  $l_1$  se vrátíme na místo  $k_1$ . Zřejmě tedy první z výše formulovaných podmínek udává, že přechody dávají smysl z hlediska směru času, druhá podmínka nás omezuje na nahrazování díry a jejího blízkého okolí, třetí podmínka zajišťuje přibližné zachování délky signálu a poslední podmínka nás omezuje na výběr takových přechodů, které budou dostatečně přirozené.

Závěrem ještě poznamenejme, že práce s úplným podobnostním grafem (tedy grafem obsahujícím hrany mezi všemi vzorky signálu) by byla neúnosně pomalá či dokonce kvůli paměťové náročnosti nereálná. Graf je proto od počátku konstruován pouze částečný, kdy se jednak omezujeme na hrany končící v okolí díry a navíc jsou vždy ze svazku hran vybírání jen reprezentanti s lokálně nejvyšší váhou.

Již z přibližně popsaneho algoritmu vidíme, že z popsaných metod je tato schopna v dostatečně repetitivním signálu doplnit nejdelší díry, jak také na experimentech ukázali autoři [30]. Také je však nejméně konzistentní, neboť dovoluje měnit nejen jednotlivé vzorky v okolí díry, které nebyly poškozené, ale dokonce může dojít ke změně délky celého signálu.

### 3.4.2 Neuronové sítě

Úspěchy neuronových sítí a hlubokého učení v rozmanitých oborech byly motivací pro jejich nasazení i na problém audio inpaintingu. První prací na toto téma je článek [26] z roku 2018. Autoři zde aplikují neuronovou síť nikoliv na signál v časové oblasti, ale na jeho spektrogram. Základním principem popsané metody je použití dvou neuronových sítí ve dvou krocích:

1. kodér – konvoluční neuronová síť – vytvoří na základě spektrogramu z levého a pravého okolí díry reprezentaci signálu,
2. dekodér – neformálně řečeno dekonvoluční neuronová síť – na základě reprezentace z prvního kroku zpětně vytvoří spektrogram restaurovaného úseku.

Známostí neuronových sítí je, že reprezentaci získanou v prvním kroku nelze snadno interpretovat, proto ji zde blíže nepopisujeme. Z teoretického hlediska je však vhodné zmínit, že autoři [26] pracují se dvěma variantami spektrogramu, reálným a komplexním. Reálný spektrogram zobrazuje pouze absolutní hodnoty časově-frekvenčních koeficientů, na rozdíl od komplexního, který obsahuje koeficienty neupravené. Proto je zřejmé, že postup pracující s reálným spektrogramem je závislý ještě na závěrečném kroku rekonstrukce fáze, bez které není možná syntéza restaurovaného signálu. Přes tuto zdánlivou nevýhodu však autoři [26] ukazují, že právě postup pracující s reálným spektrogramem dosahuje v objektivním hodnocení<sup>1</sup> rekonstrukce lepších výsledků.

---

<sup>1</sup>O metodách hodnocení rekonstrukce bude bližší zmínka v úvodu kapitoly 6.



## 4 Konvexní metody založené na řídkosti

V této kapitole se budeme věnovat konvexnímu přístupu k audio inpaintingu. Základní myšlenkou je tzv.  $\ell_1$  relaxace, což označuje nahrazení řídkosti ve formulaci úlohy v jistém smyslu nejbližší konvexní funkcí, kterou je  $\ell_1$  norma.

V obecné matematické podobě můžeme výchozí (nerelaxovanou) úlohu psát jako

$$\arg \min_{\mathbf{x}} \|\mathbf{z}\|_0 \quad \text{vzhledem k} \quad F(\mathbf{x}, \mathbf{z}) = 0, \quad \mathbf{x} \in \Gamma, \quad \mathbf{z} \in \mathbb{C}^P. \quad (4.1)$$

Jak bylo dříve zavedeno,  $\mathbf{x}$  označuje hledaný signál a  $\mathbf{z}$  jsou časově-frekvenční koeficienty. Abychom mohli úlohu formulovat v požadované obecnosti, zavádíme zde navíc zatím nespecifikovanou funkci  $F$ , která dává do souvislosti právě signál v časové oblasti a koeficienty.

**Poznámka 4.1.** Značením v rovnici (4.1) zavádíme konvenci, kdy jako argument minima vystupuje proměnná, která (poněkud netypicky) nevystupuje v účelové funkci. Chceme tím vyjádřit, že nalezneme-li optimální koeficienty  $\hat{\mathbf{z}}$ , pak řešením celé úlohy je signál  $\hat{\mathbf{x}}$ , který splňuje  $F(\hat{\mathbf{x}}, \hat{\mathbf{z}}) = 0$ , nikoliv tyto koeficienty.

Relaxovanou úlohu pak dostáváme ve tvaru

$$\arg \min_{\mathbf{x}} \|\mathbf{z}\|_1 \quad \text{vzhledem k} \quad F(\mathbf{x}, \mathbf{z}) = 0, \quad \mathbf{x} \in \Gamma, \quad \mathbf{z} \in \mathbb{C}^P. \quad (4.2)$$

Protože současně množina přípustných řešení  $\Gamma$  vystupující v rovnici (4.1) i v relaxované variantě (4.2) je konvexní (jedná se o podprostor prostoru všech signálů  $\mathbb{C}^N$ ), audio inpainting s využitím  $\ell_1$  relaxace získal podobu úlohy konvexní optimalizace, na jejíž řešení jsou známé algoritmy, které níže popíšeme.

Dále se v této kapitole budeme zabývat volbou funkce  $F$  v rovnici (4.2) a zejména budeme řešit problém, který vzniká v případě, kdy pomocí  $\ell_1$  relaxace doplňujeme delší chybějící úsek. V takovém případě totiž dochází v rekonstruovaném úseku k poklesu energie, který způsobuje v signálu nežádoucí slyšitelné artefakty. Ačkoliv jsme dosud nepopsali přesný tvar konvexní úlohy (4.2) ani algoritmy sloužící k jejímu řešení, můžeme již na tomto místě obrázkem 4.1 ilustrovat tuto problematickou situaci.

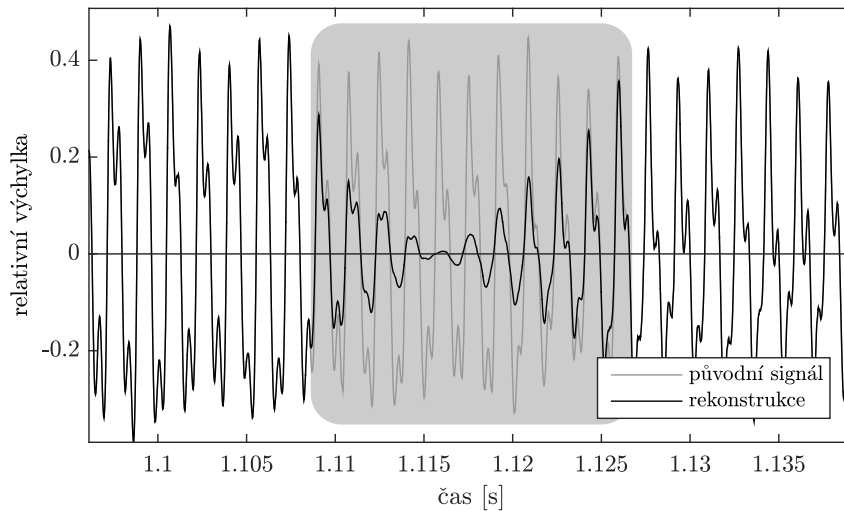
Pokles energie v rekonstrukci na obrázku 4.1 je zcela evidentní a nejedná se o ojedinělý případ. V této kapitole proto budeme navrhopvat metody, které vycházejí z konvexní formulace (4.2), ale snaží se poklesu energie ve výsledném signálu předcházet.

### 4.1 Analyzující a syntetizující model $\ell_1$ relaxace

V této části blíže rozebereme dva různé tvary relaxované úlohy (4.2), a to tzv. analyzující a syntetizující model. Stručně řečeno se jedná o volbu funkce  $F(\mathbf{x}, \mathbf{z})$ , která byla v úloze (4.2) použita a která dává do souvislosti signál a časově-frekvenční koeficienty. Nejjednodušší možnosti reprezentují následující funkce

$$F_1(\mathbf{x}, \mathbf{z}) = A\mathbf{x} - \mathbf{z}, \quad F_2(\mathbf{x}, \mathbf{z}) = \mathbf{x} - D\mathbf{z}. \quad (4.3)$$

Právě podle toho, zda využíváme analyzující operátor  $A$  (varianta  $F_1$ ) nebo syntetizující operátor  $D$  (varianta  $F_2$ ) pak hovoříme o analyzujícím, resp. syntetizujícím modelu.



Obr. 4.1: Ilustrace poklesu energie v doplněném úseku signálu. Doplněný úsek indikuje šedá oblast. Šedě je též jako reference vykreslen původní (neporušený) signál, tedy signál, ve kterém jsme díru uměle vytvořili, abychom mohli následně pozorovat zřetelný pokles energie v černě zobrazené rekonstrukci.

Pro zřetelnější zápis využijeme specifického tvaru podmínky  $F(\mathbf{x}, \mathbf{z}) = 0$ . Analyzující model nám pak zřejmě udává podmínku  $A\mathbf{x} = \mathbf{z}$ , což můžeme chápat jako substituci a úlohu tak dostáváme ve tvaru obsahujícím již pouze jednu (vektorovou) proměnnou:

$$\arg \min_{\mathbf{x}} \|A\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \mathbf{x} \in \Gamma. \quad (4.4)$$

Obdobnou úpravou dostaneme syntetizující model

$$\arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{vzhledem k} \quad D\mathbf{z} \in \Gamma. \quad (4.5)$$

Hledaný signál bychom v tomto modelu získali teprve aplikací syntetizujícího operátoru na výsledek úlohy (4.5), který je v prostoru koeficientů. Také zdůrazněme, že v případě, kdy operátory  $A$  a  $D = A^*$  odpovídají ortonormální bázi, tedy  $AD = DA = \text{Id}$ , jsou úlohy (4.4) a (4.5) ekvivalentní. Ovšem pak se také ztrácí manévrovací prostor ve smyslu hledání řídkého řešení, neboť reprezentace signálu je jednoznačná. V případě redundantní reprezentace, což je příklad framů, se však jedná o různé úlohy.

#### 4.1.1 Algoritmus pro řešení analyzujícího modelu

Pro řešení analyzujícího modelu (4.4) využijeme iterativní Chambollův-Pockův (CP) algoritmus. Jedná se o tzv. proximální algoritmus, neboť je založen na opakovaném užití proximálních operátorů. Konkrétně CP algoritmus, prezentovaný v [10], poskytuje posloupnost  $\{\mathbf{x}^{(n)}\}$  konvergující k řešení problému

$$\arg \min_{\mathbf{x}} f(K\mathbf{x}) + g(\mathbf{x}), \quad (4.6)$$

kde na funkce  $f, g$  a operátor  $K$  jsou kladeny následující požadavky:

- $f$  konvexní, zdola polospojitá,
- $g$  konvexní, zdola polospojitá, nezáporná,
- $K$  spojitý lineární operátor.

Obecný tvar Chambolla-Pockova algoritmu nalezneme v algoritmu 1.

---

**Algoritmus 1:** Chambollův-Pockův algoritmus v obecném tvaru

---

**input:** operátory  $K: X \rightarrow Y$  a  $K^*: Y \rightarrow X$ , proximální operátory funkcí  $f^*$  a  $g$   
**1** nastav parametry  $\sigma, \tau > 0$  splňující  $\sigma\tau \|K\|^2 < 1$  a  $\theta \in \langle 0, 1 \rangle$   
**2** nastav hodnotu primární proměnné  $\mathbf{p}^{(0)} \in X$  a duální proměnné  $\mathbf{q}^{(0)} \in Y$   
**3** nastav hodnotu výstupní proměnné  $\mathbf{x}^{(0)} = \mathbf{p}^{(0)}$   
**4 for**  $n = 0, 1, 2, \dots$  **do**  
**5**      $\mathbf{q}^{(n+1)} = \text{prox}_{\sigma f^*}(\mathbf{q}^{(n)} + \sigma K \mathbf{p}^{(n)})$   
**6**      $\mathbf{x}^{(n+1)} = \text{prox}_{\tau g}(\mathbf{x}^{(n)} - \tau K^* \mathbf{q}^{(n+1)})$   
**7**      $\mathbf{p}^{(n+1)} = \mathbf{x}^{(n+1)} + \theta(\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)})$   
**8 end**  
**9 return**  $\mathbf{x}^{(n+1)}$

---

Úlohu (4.6) dáme do souvislosti s analyzujícím modelem pro audio inpainting tak, že tento model formulujeme v neomezeném tvaru s využitím indikátorové funkce dle definice 2.18 jako

$$\arg \min_{\mathbf{x}} \|A\mathbf{x}\|_1 + \iota_{\Gamma}(\mathbf{x}). \quad (4.7)$$

V roli funkce  $f$  máme v rovnici (4.7)  $\ell_1$  normu, funkcí  $g$  je indikátorová funkce množiny přípustných řešení  $\Gamma$  a lineárním operátorem vystupujícím v kompozici s funkcí  $f$  je analyzující operátor  $A$ . Podle [10] nyní k sestavení algoritmu potřebujeme znát proximální operátory funkcí  $\sigma f^*(\mathbf{x})$  a  $\tau g(\mathbf{x})$  (konstanty  $\sigma$  a  $\tau$  blíže určíme níže). Pro funkci  $\tau g = \tau \iota_{\Gamma} = \iota_{\Gamma}$  je proximálním operátorem ortogonální projekce signálu na množinu  $\Gamma$ , kterou jednoduše obdržíme tak, že vzorky signálu odpovídající nepoškozené části nahradíme vzorky  $\mathbf{y}_R$ . K vyjádření proximálního operátoru funkce  $\sigma f^*$  využijeme vzorce (2.27) uvedeného již v kapitole 2, který nám poskytuje rovnost

$$\text{prox}_{\sigma f^*}(\mathbf{x}) = \mathbf{x} - \sigma \cdot \text{prox}_{f/\sigma}(\mathbf{x}/\sigma). \quad (4.8)$$

Proximálním operátorem funkce  $f(\mathbf{x}) = \|\mathbf{x}\|_1$  je také již dříve zavedené měkké prahování, zde s parametrem  $1/\sigma$ . Následující lemma nám umožní vyjádřit  $\text{prox}_{\sigma f^*}(\mathbf{x})$  v kompaktním tvaru:

**Lemma 4.2.** *Nechť  $\alpha \in \mathbb{R}, \alpha \geq 0$  je práh měkkého prahování. Pak*

$$\mathcal{S}_{\alpha}(\alpha \mathbf{x}) = \alpha \cdot \mathcal{S}_1(\mathbf{x}). \quad (4.9)$$

*Důkaz.* Důkaz provedeme přímými úpravami, přičemž vycházíme z definice měkkého prahování. Připomeňme, že  $\arg(c)$  označuje argument komplexního čísla  $c$ ; je-li vstupem vektor, výstupem je vektor argumentů jednotlivých jeho složek. Symbolem  $\odot$  pak rozumíme násobení vektorů po složkách. Platí

$$\begin{aligned} \mathcal{S}_{\alpha}(\alpha \mathbf{x}) &= \arg(\alpha \mathbf{x}) \odot \max(|\alpha \mathbf{x}| - \alpha, 0) \\ &= \arg(\mathbf{x}) \odot \alpha \max(|\mathbf{x}| - 1, 0) \\ &= \alpha \cdot \mathcal{S}_1(\mathbf{x}), \end{aligned}$$

což je tvar, ke kterému jsme měli dojít. □

Uvažujeme-li nyní místo obecného prahu  $\alpha$  hodnotu  $1/\sigma$ , jak je tomu v rovnici (4.8), dostaneme vyjádření

$$\text{prox}_{\sigma f^*}(\mathbf{x}) = \mathbf{x} - \mathcal{S}_1(\mathbf{x}) = \mathcal{C}_1(\mathbf{x}). \quad (4.11)$$

Použili jsme nově definovaný operátor  $\mathcal{C}_\alpha(\mathbf{x}) = 1 - \mathcal{S}_\alpha(\mathbf{x})$ , jehož explicitní vyjádření udává následující lemma:

**Lemma 4.3.** *Operátor  $\mathcal{C}_\alpha(\mathbf{x}) = \mathbf{x} - \mathcal{S}_\alpha(\mathbf{x})$  může být explicitně vyjádřen jako*

$$\mathcal{C}_\alpha(\mathbf{x}) = \arg(\mathbf{x}) \odot \min(\alpha, |\mathbf{x}|). \quad (4.12)$$

*Důkaz.* Opět s využitím definice měkkého prahování platí

$$\begin{aligned} \mathcal{C}_\alpha(\mathbf{x}) &= \mathbf{x} - \mathcal{S}_\alpha(\mathbf{x}) \\ &= \mathbf{x} - \arg(\mathbf{x}) \odot \max(|\mathbf{x}| - \alpha, 0) \\ &= \arg(\mathbf{x}) \odot (|\mathbf{x}| - \max(|\mathbf{x}| - \alpha, 0)) \\ &= -\arg(\mathbf{x}) \odot \max(-\alpha, -|\mathbf{x}|) \\ &= \arg(\mathbf{x}) \odot \min(\alpha, |\mathbf{x}|), \end{aligned}$$

čímž je tvrzení dokázáno. □

Dosavadní úvahy shrnuje algoritmus 2.

---

**Algoritmus 2:** Chambollův-Pockův algoritmus pro audio inpainting

---

**input:** operátory  $A$  a  $D = A^*$ , množina  $\Gamma$ , nepoškozená část signálu  $\mathbf{y}_R$

- 1 nastav parametry  $\sigma, \tau > 0$  splňující  $\sigma\tau \|A\|^2 < 1$
- 2 nastav hodnotu primární proměnné  $\mathbf{p}^{(0)} \in \mathbb{C}^N$  a duální proměnné  $\mathbf{q}^{(0)} \in \mathbb{C}^P$
- 3 nastav hodnotu výstupní proměnné  $\mathbf{x}^{(0)} = \mathbf{p}^{(0)}$
- 4 **for**  $n = 0, 1, 2, \dots$  **do**
- 5      $\mathbf{q}^{(n+1)} = \mathcal{C}_1(\mathbf{q}^{(n)} + \sigma A\mathbf{p}^{(n)})$
- 6      $\mathbf{x}^{(n+1)} = \mathcal{P}_\Gamma(\mathbf{x}^{(n)} - \tau D\mathbf{q}^{(n+1)})$
- 7      $\mathbf{p}^{(n+1)} = 2\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}$
- 8 **end**
- 9 **return**  $\hat{\mathbf{x}} = \mathbf{x}^{(n+1)}$

---

Obecná forma CP algoritmu 1 obsahuje ještě navíc parametr  $\theta \in \langle 0, 1 \rangle$ , který by se uplatnil v sedmém kroku algoritmu 2. V [10] však je uvedena postačující podmínka konvergence algoritmu, jak ji máme nastíněnou v kroku prvním, a to pouze pro hodnotu  $\theta = 1$ . Algoritmus 2 již proto tuto volbu zahrnuje.

Zbývá určit, kdy chod algoritmu ukončit. Používat budeme dvě varianty – buď omezíme maximální počet iterací, nebo zvolíme hodnotu relativní normy řešení, tedy podmínku

$$\|\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}\|_2 < \varepsilon \|\mathbf{x}^{(n)}\|_2, \quad (4.14)$$

kde  $\varepsilon > 0$  je zvolená konstanta.

### 4.1.2 Algoritmus pro řešení syntetizujícího modelu

Pro řešení syntetizujícího modelu (4.5) využijeme Douglasův-Rachfordův (DR) algoritmus, jehož obecnou podobu lze najít např. v [12]. Stejně jako CP algoritmus je i tento

z kategorie proximálních algoritmů, jedná se však o přístup jednodušší. Pro analyzující model jsme jej využít nemohli, protože proximální operátor  $\ell_1$  normy v kompozici s lineárním analyzujícím operátorem neznáme.

Je dokázáno, že DR algoritmus poskytuje posloupnost  $\{\mathbf{x}^{(n)}\}$  konvergující k řešení úlohy

$$\arg \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}), \quad (4.15)$$

kde funkce  $f, g$  jsou konvexní a zdola polospojité, vnitřky jejich definičních oborů mají neprázdný průnik a navíc platí podmínka  $f(\mathbf{x}) + g(\mathbf{x}) \rightarrow \infty$  pro  $\|\mathbf{x}\| \rightarrow \infty$ . Obecný tvar tohoto algoritmu shrnuje algoritmus 3.

---

**Algoritmus 3:** Douglasův-Rachfordův algoritmus v obecném tvaru

---

**input:** proximální operátory funkcí  $f$  a  $g$

- 1 nastav parametry  $\tau > 0$  a  $\delta \in (0, 1)$
- 2 nastav hodnotu pomocné proměnné  $\mathbf{q}^{(0)}$
- 3 **for**  $n = 0, 1, 2, \dots$  **do**
- 4      $\mathbf{z}^{(n)} = \text{prox}_{\tau f}(\mathbf{q}^{(n)})$
- 5     zvol  $\lambda \in \langle \delta, 2 - \delta \rangle$
- 6      $\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} + \lambda (\text{prox}_{\tau g}(2\mathbf{z}^{(n)} - \mathbf{q}^{(n)}) - \mathbf{z}^{(n)})$
- 7 **end**
- 8 **return**  $D\mathbf{z}^{(n)}$

---

Aby byla zřejmá souvislost obecného tvaru úlohy (4.15) a syntetizujícího modelu (4.5), vyjádříme náš problém v neomezeném tvaru s využitím indikátorové funkce jako

$$\arg \min_{\mathbf{z}} \|\mathbf{z}\|_1 + \iota_{\Gamma}(D\mathbf{z}). \quad (4.16)$$

Vidíme, že změnou značení proměnné a položením  $f(\mathbf{z}) = \|\mathbf{z}\|_1$  a  $g(\mathbf{z}) = \iota_{\Gamma}(D\mathbf{z})$  dostáváme skutečně syntetizující model ve tvaru (4.15) a navíc funkce  $f, g$  splňují podmínky pro použití DR algoritmu.

Pro sestavení algoritmu již stačí znát pouze proximální operátory funkcí  $\tau f(\mathbf{z})$  a  $\tau g(\mathbf{z})$  (kladná konstanta  $\tau$  značí parametr DR algoritmu). Pro  $\ell_1$  normu je proximálním operátorem měkké prahování, pro kompozici syntetizujícího operátoru  $D$  a indikátorové funkce  $\iota_{\Gamma}$  se jedná o jistou formu projekce, kterou lze na základě [35] vypočítat jako

$$\text{prox}_{\tau \iota_{\Gamma} \circ D}(\mathbf{z}) = \mathbf{z} - AM_{\text{R}}D\mathbf{z} + A\mathbf{y}_{\text{R}}. \quad (4.17)$$

Podmínkou je, že operátory  $A$  a  $D$  přísluší Parsevalovu framu. Výsledný DR algoritmus – již adaptovaný pro řešení úlohy (4.16) – nalezneme v algoritmu 4. Šestý krok algoritmu jsme ponechali neupravený, aby bylo zřetelné použití rovnice (4.17).

---

**Algoritmus 4:** Douglasův-Rachfordův algoritmus pro audio inpainting

---

**input:** operátory  $A$  a  $D = A^*$ , množina  $\Gamma$ , nepoškozená část signálu  $\mathbf{y}_R$

- 1 nastav parametry  $\tau > 0$  a  $\delta \in (0, 1)$
- 2 nastav hodnotu pomocné proměnné  $\mathbf{q}^{(0)} \in \mathbb{C}^P$
- 3 **for**  $n = 0, 1, 2, \dots$  **do**
- 4      $\mathbf{z}^{(n)} = \mathcal{S}_\tau(\mathbf{q}^{(n)})$
- 5     zvol  $\lambda \in \langle \delta, 2 - \delta \rangle$
- 6      $\mathbf{q}^{(n+1)} = \mathbf{q}^{(n)} + \lambda \left( (2\mathbf{z}^{(n)} - \mathbf{q}^{(n)}) - AM_R D(2\mathbf{z}^{(n)} - \mathbf{q}^{(n)}) + A\mathbf{y}_R - \mathbf{z}^{(n)} \right)$
- 7 **end**
- 8 **return**  $\hat{\mathbf{x}} = D\mathbf{z}^{(n)}$

---

Chod algoritmu můžeme ukončit obdobně, jako tomu bylo u analyzujícího modelu v algoritmu CP. Používat proto budeme opět omezení maximálního počtu iterací, nebo podmínku pro relativní normu řešení

$$\|\mathbf{z}^{(n)} - \mathbf{z}^{(n-1)}\|_2 < \varepsilon \|\mathbf{z}^{(n-1)}\|_2, \quad (4.18)$$

kde  $\varepsilon > 0$  je zvolená konstanta.

Jak bylo nastíněno, dále se budeme věnovat úpravám analyzujícího i syntetizujícího modelu s cílem zabránit poklesu energie v rekonstruovaném úseku.

## 4.2 Modifikace modelu

V této části se začneme věnovat modifikacím algoritmů pro audio inpainting s využitím  $\ell_1$  relaxace, které si kladou za cíl zmírnit, nejlépe zcela eliminovat pokles energie v doplněném úseku signálu. Takovou situaci ilustroval obrázek 4.1, kde byla doplňovaná díra tak dlouhá, že amplituda rekonstruovaného signálu lokálně klesla až k nule.

Příčinou zřejmě může být proces minimalizace  $\ell_1$  normy vektoru koeficientů místo řídkosti, neboť přitom kromě vynulování některých složek tohoto vektoru dochází také k nežádoucímu snižování velikosti složek ostatních, které se uplatní v syntéze rekonstruovaného signálu. Tento nežádoucí jev se objevující problém je zejména v oblasti statistiky nazýván *bias* [18, 9, 32, 14].

Dále proto navrhneme dvě úpravy formulace modelu inpaintingu, které budou popsány bias snižovat. Část 4.3 se pak bude věnovat alternativnímu přístupu, jímž jsou úpravy již rekonstruovaného signálu v časové oblasti.

### 4.2.1 Váhování atomů

Váhováním atomů rozumíme přiřazení vah jednotlivým časově-frekvenčním koeficientům signálu tak, že nejsou všechny stejně penalizovány  $\ell_1$  normou. Zavedeme-li váhy jako vektor  $\mathbf{w} \in \mathbb{R}^P$  (odpovídá délce vektoru koeficientů) s přirozenou podmínkou  $w_n > 0$ ,  $n = 1, 2, \dots, P$ , pak váhování znamená minimalizaci funkce  $\|\mathbf{w} \odot \cdot\|_1$  namísto  $\|\cdot\|_1$ . Analyzující model (4.4) pak nabude tvaru

$$\arg \min_{\mathbf{x}} \|\mathbf{w} \odot A\mathbf{x}\|_1 \quad \text{vzhledem k} \quad \mathbf{x} \in \Gamma. \quad (4.19)$$

Obdobně je váhovaný syntetizující model tvaru

$$\arg \min_{\mathbf{z}} \|\mathbf{w} \odot \mathbf{z}\|_1 \quad \text{vzhledem k} \quad D\mathbf{z} \in \Gamma. \quad (4.20)$$

Podstatné je, že popsané algoritmy CP a DR mohou být opět použity. Minimalizujeme totiž váhovanou  $\ell_1$  normu, pro niž dle lemmatu 2.22 opět známe proximální operátor, jímž je  $\mathcal{S}_{\mathbf{w}}$ , tedy měkké prahování s vektorovým parametrem, které jsme definovali vzorcem

$$\mathcal{S}_{\mathbf{w}}(\mathbf{x}) = \arg(\mathbf{x}) \odot \max(|\mathbf{x}| - \mathbf{w}, 0). \quad (2.23)$$

Obdobně můžeme pro vektorový parametr adaptovat funkci  $\mathcal{C}_{\mathbf{w}}$ . Vzorec (2.23) nám také nabízí interpretaci váhování – je-li  $n$ -tému atomu přiřazena nízká hodnota  $w_n$ , pak v kroku algoritmu, který zajišťuje snižování  $\ell_1$  normy koeficientů, snižujeme příslušný koeficient pouze o nízkou hodnotu.

Nyní chceme navrhnout jednotlivé váhy tak, abychom dosáhli menšího úbytku energie v rekonstruovaném úseku. Několik variant navrhneme pro speciální případ Gaborova systému. Situaci, kdy prvky Gaborova slovníku – atomy – zasahují do díry v signálu, ilustruje obrázek 4.2. Schematicky je zvláště zobrazena část atomů v díře a část odpovídající neporušenému signálu, kterou pro atom  $\mathbf{g}^n$ ,  $n = 1, 2, \dots, P$ , označíme jako  $\mathbf{g}_{\mathbf{R}}^n = M_{\mathbf{R}}\mathbf{g}^n$ .

Uvažujme následovně – jestliže má daný atom malý podíl v neporušené části signálu, může být tato část syntetizována (díky redundanci systému) bez jeho využití, příslušný koeficient tedy bude v minimalizaci  $\ell_1$  normy potlačen. Takový atom ale přitom významně ovlivňuje energii doplněného úseku. Proto odpovídající váhu zvolíme nízkou, čímž docílíme toho, že koeficient bude při minimalizaci (váhované)  $\ell_1$  normy méně penalizován.

S touto motivací navrhuje následující varianty, jak jednotlivé váhy volit. Neváhouvanou variantu jsme zařadili jako speciální případ.

$$(a) \quad w_n = 1, \text{ neboli } \mathbf{w} \odot \mathbf{z} = \text{Id } \mathbf{z} = \mathbf{z} \quad (\text{základní varianta bez váhování}) \quad (4.21a)$$

$$(b) \quad w_n = \frac{|\text{supp}(\mathbf{g}_{\mathbf{R}}^n)|}{|\text{supp}(\mathbf{g}^n)|} \quad (\text{dle nosiče}) \quad (4.21b)$$

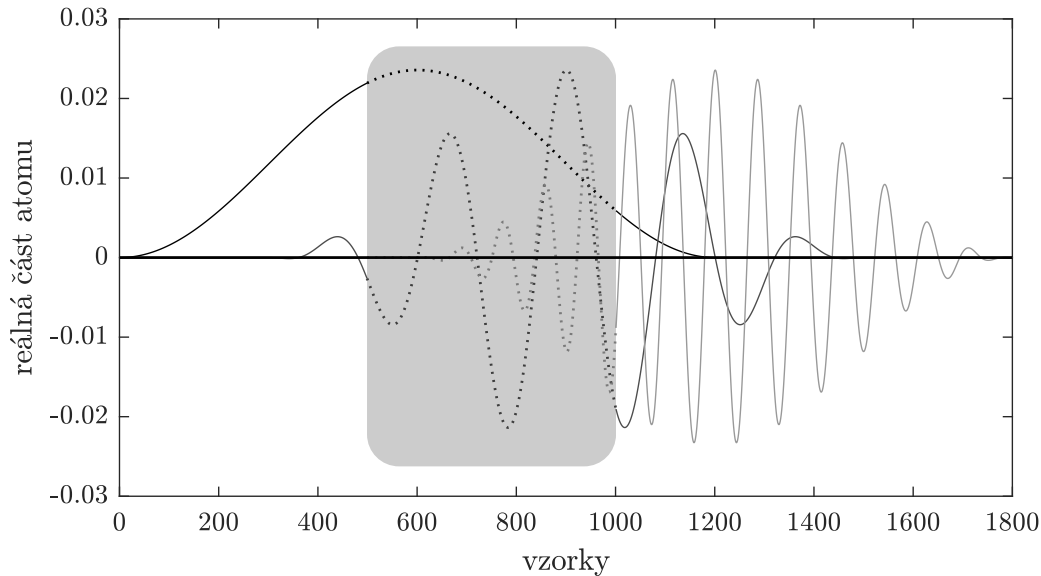
$$(c) \quad w_n = \frac{\|\mathbf{g}_{\mathbf{R}}^n\|_1}{\|\mathbf{g}^n\|_1} \quad (\text{dle } \ell_1 \text{ normy}) \quad (4.21c)$$

$$(d) \quad w_n = \frac{\|\mathbf{g}_{\mathbf{R}}^n\|_2}{\|\mathbf{g}^n\|_2} \quad (\text{dle } \ell_2 \text{ normy}) \quad (4.21d)$$

$$(e) \quad w_n = \frac{\|\mathbf{g}_{\mathbf{R}}^n\|_2^2}{\|\mathbf{g}^n\|_2^2} \quad (\text{dle energie}) \quad (4.21e)$$

**Poznámka 4.4.** V případě, kdy pracujeme s Gaborovým framem, nejsou v předcházejících vzorcích nutné jmenovatele, neboť se jedná o hodnoty pro všechny atomy stejné; můžeme je tedy chápat jako kladný skalární násobek funkce  $\|\mathbf{w} \cdot \odot\|_1$ , což na hledaný argument minima nemá vliv. Tuto skutečnost snadno nahlédneme z podoby jednotlivých atomů Gaborova framu – jedná se totiž vždy o modulaci (a translaci, což nyní nemusíme uvažovat) okna  $\mathbf{g}$ , tato modulace ovšem nemá vliv na absolutní hodnoty jednotlivých vzorků<sup>1</sup>. Proto normy celých atomů použité v (4.21c), (4.21d) a (4.21e) jsou neměnné. Stejná situace je i ve variantě (4.21b), neboť modulace okna  $\mathbf{g}$  (jehož nosičem je uzavřený interval, což je náš případ) zřejmě nesníží amplitudu vzorků na 0.

<sup>1</sup>Ve spojitém případě je modulací okna  $g(x)$  s parametrem  $b$  funkce  $M_b g(x) = \exp(2\pi i b x) \cdot g(x)$ . Ve vzorcích (4.21) však při rozepsání jednotlivých norem vystupují moduly složek vektorů. Spočteme-li modul pro hodnotu  $M_b g(x)$ , dostáváme  $|M_b g(x)| = |\exp(2\pi i b x) \cdot g(x)| = |\exp(2\pi i b x)| \cdot |g(x)| = |g(x)|$ . Stejná vlastnost přirozeně platí i v diskrétním případě, proto skutečně modulace nemá na absolutní hodnotu vzorků vliv.



Obr. 4.2: Tři příklady atomů Gaborova systému  $\mathbf{g}^n$ , které zasahují do chybějícího úseku (symbolizován šedou oblastí). Zobrazena je reálná část atomů. Části atomů odpovídající nepoškozenému signálu, tedy  $\mathbf{g}_R^n$ , jsou zobrazeny plnou čarou.

Jmenovatele však ve vzorcích ponecháváme, neboť ilustrují myšlenku, že měříme podíl atomu v nepoškozené části. Navíc vzorce (4.21) jsou použitelné i pro jiné reprezentace, než jsou Gaborovy framy.

**Poznámka 4.5.** V případě, kdy délka chybějícího úseku přesáhne délku nosiče okna  $\mathbf{g}$ , může přirozeně nastat situace, kdy by některá z vah byla nulová. To by bylo v rozporu s úvodním předpokladem, že všechny váhy jsou kladné. Tento patologický případ však nebudeme dále uvažovat.

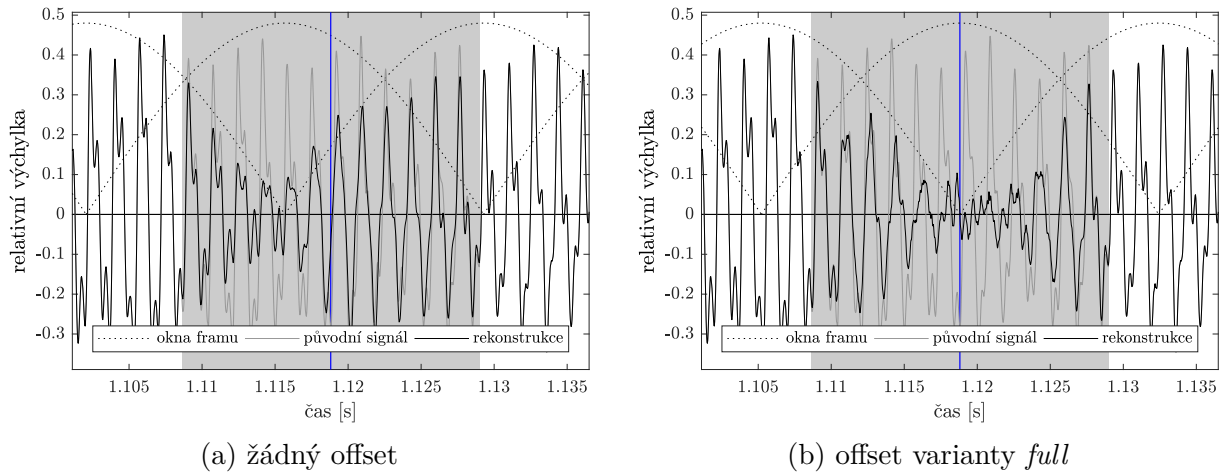
## 4.2.2 Offset

Pod pojmem *offset* rozumíme nastavení počátečního bodu transformace signálu – protože časově-frekvenční transformace vnímají signály jako periodické, můžeme zvolit, který bod v čase budeme považovat pro tuto periodu za počáteční. Uvažujeme-li speciálně zpracování signálu s využitím Gaborova framu konstruovaného s oknem  $\mathbf{g}$ , můžeme si alternativně představit, že offsetem nastavujeme posunutí všech translací<sup>2</sup> okna  $\mathbf{g}$  vůči signálu.

Z hlediska problému poklesu energie v doplněném úseku signálu je pro nás důležité pozorování, že průběh energie se řídí vzájemnou pozicí oken a díry v signálu, jak ilustruje obrázek 4.3. Především vidíme, že je-li posloupnost oken symetrická vzhledem ke středu díry (jako je tomu na obrázku 4.3b), pak je stejně symetrický i pokles energie (a naopak, což je případ obrázku 4.3a).

Cílem dále bude navrhnout počáteční bod Gaborovy transformace tak, aby pokles energie byl největší právě ve středu doplněného úseku. Tímto nebude pokles energie kompenzován, přesto je však tento krok významný, neboť v části 4.3 navrhne metody pro kompenzaci poklesu energie v časové oblasti (tedy na již rekonstruovaném signálu), kde předpoklad symetrie průběhu energie s výhodou využijeme.

<sup>2</sup>Tyto translace budeme dále označovat jednoduše jako okna.



Obr. 4.3: Ilustrace poklesu energie v doplněném úseku při použití offsetu. Vlevo je původní průběh, vpravo varianta s offsetem, který přesouvá místo minimální energie ke středu doplněné díry. Díru symbolizuje šedá oblast, její střed je zobrazen modře.

### Zkrácení signálu

V implementaci audio inpaintingu s offsetem využijeme algoritmus pro zkrácení nosiče signálu na nutné minimum, který byl prezentovaný v [34]. Základní myšlenkou tohoto algoritmu je, že okna dostatečně vzdálená od poškozeného úseku nejsou při rekonstrukci potřebná; pro doplnění jediné díry v signálu proto stačí zpracovávat pouze jeho krátký úsek. Tento přístup nám umožňuje významně zrychlit výpočet audio inpaintingu, protože nevyžaduje výpočetně náročnou analýzu a syntézu částí signálu, které k rekonstrukci dané díry nepřispívají. Navíc jedním ze vstupů algoritmu je právě hodnota *offset* (ve výpočtu budeme tuto hodnotu značit *off*), která ovlivňuje určení počátečního bodu zkráceného signálu, což odpovídá posunu počátečního bodu (Gaborovy) transformace.

Pro úplnost zde část odvození minimálního nosiče nutného pro rekonstrukci jedné díry uvedeme (značení a zkrácené odvození přebíráme z [34]). Protože signál zpracováváme jako periodický, musíme brát v úvahu, aby okna s nenulovým překryvem s dírou nepřesáhla hranice zkráceného signálu. V opačném případě bychom po periodizaci mohli část díry mylně považovat za nepoškozený signál. Označme

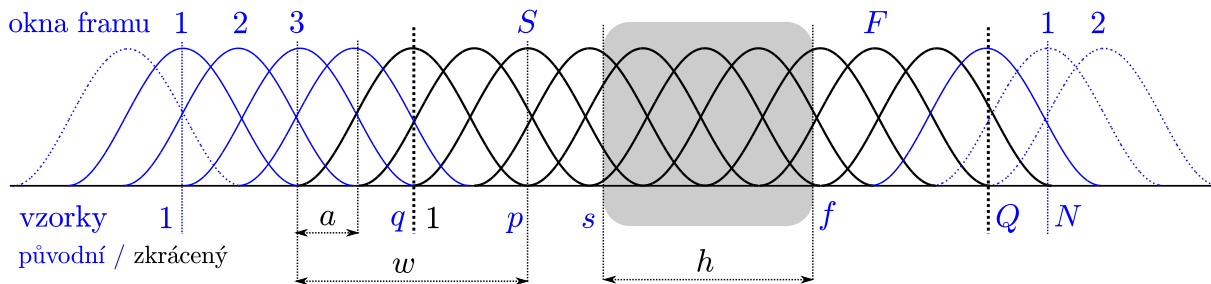
- $s$  první index díry,
- $f$  poslední index díry,
- $q$  první index zkráceného signálu,
- $Q$  poslední index zkráceného signálu,

to vše vzhledem k indexování původního signálu počínaje indexem 1 (situaci ilustruje obrázek 4.4).

Dále budeme uvažovat, že okno s indexem  $(k + 1)$  má svůj centrální prvek na pozici  $1 + k \cdot a$ , první prvek na pozici  $1 + k \cdot a - \lfloor w/2 \rfloor$  a poslední prvek na pozici  $k \cdot a + \lceil w/2 \rceil$ . Centrální prvky prvního a posledního okna potřebného pro rekonstrukci jsou  $q$  a  $(Q + 1)$ . Popsané obecné podmínky na zkrácený signál můžeme nyní formalizovat jako

$$q \leq s - \left\lceil \frac{w}{2} \right\rceil, \quad Q \geq f + \left\lfloor \frac{w}{2} \right\rfloor, \quad a \mid ((Q + 1) - q). \quad (4.22)$$

Poslední podmínka, že  $a$  dělí  $((Q + 1) - q)$ , musí být splněna proto, že v Gaborově systému



Obr. 4.4: Vizualizace Gaborova systému a značení. Šedá oblast představuje chybějící úsek. Modré označení je vzhledem k původnímu signálu, černé vzhledem k signálu zkrácenému. Indexy  $q$  a (černá) 1 označují první vzorek zkráceného signálu vzhledem k původnímu, resp. zkrácenému signálu. Pro doplnění díry nám postačí úsek černě zobrazených oken. Okna v obrázku, která jsou ve skutečnosti identická s jinými okny (vzhledem k tomu, že signál je zpracováván jako periodický), jsou vykreslena tečkovanou čarou (například okna 1 a 2).

jsou centrální indexy oken vždy vzdáleny o násobek parametru  $a$ .

Pro integraci offsetu potřebujeme dále označit

- $p$  centrální index prvního okna potřebného pro rekonstrukci,
- $S$  index tohoto okna; tedy platí vazba  $p = 1 + (S - 1) \cdot a$ ,

což opět ilustruje obrázek 4.4. Parametr *off* nyní chápeme jako hodnotu, o kterou při výpočtu zkrácení signálu měníme  $p$ . Po přepočtu hodnoty  $p$  bychom však měli ověřit, že okno s indexem  $S$  (tato hodnota je v algoritmu v [34] vypočítána dříve) je stále první relevantní okno pro rekonstrukci díry – volbou příliš vysoké hodnoty *off* bychom totiž mohli dosáhnout toho, že díra nyní zasáhne i do předcházejícího okna. Navíc můžeme bez újmy na obecnosti předpokládat  $0 \leq \text{off} < a$ , neboť posunem oken o násobek hodnoty  $a$  dostáváme ekvivalentní situaci.

Tyto úvahy zařadíme do algoritmu prezentovaného v [34] bezprostředně za výpočet hodnoty  $p$  v podobně následujících kroků:

---

**Algoritmus 5:** Doplnění algoritmu pro zkrácení signálu prezentovaného v [34]

---

```

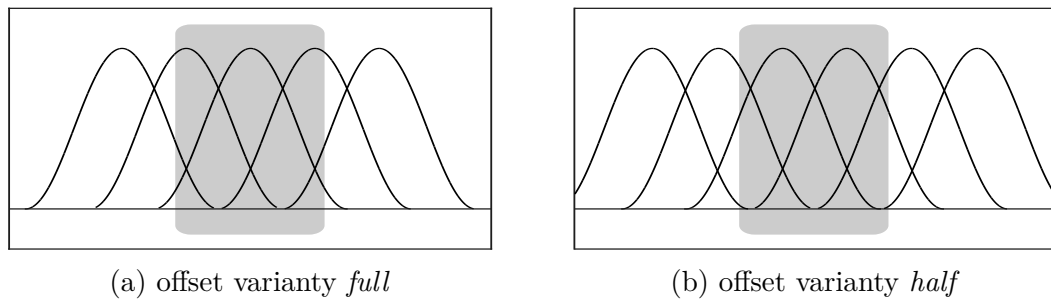
1  $\text{off} \leftarrow \text{off} \bmod a$ 
2  $p \leftarrow p + \text{off}$ 
3 if  $p - a + \lceil w/2 \rceil - 1 \geq s$  then
   | // hodnota na levé straně podmínky je poslední index okna  $S - 1$ 
4   |  $S \leftarrow S - 1$ 
5   |  $p \leftarrow p - a$ 
6 end

```

---

### Návrh variant offsetu

Cílem nyní je vypočítat hodnotu *off* z předchozího odvození tak, abychom splnili původní požadavek, tedy že posloupnost oken bude symetrická vzhledem k centrálnímu indexu díry. Obrázek 4.5 ilustruje, že tohoto lze dosáhnout dvěma způsoby – buď centrální index díry odpovídá centrálnímu indexu některého okna (což budeme označovat jako *full offset*), nebo je centrální index díry na ose dvou sousedních oken (*half offset*).



Obr. 4.5: Návrh variant offsetu pro Gaborův systém. Při variantě *full* centrální index díry splývá s centrálním prvkem okna, při variantě *half* pak s osou dvou sousedních oken.

Výpočet hodnoty *off*, o kterou posunujeme počáteční bod zkráceného signálu, udává pro obě varianty offsetu algoritmus 6.

---

**Algoritmus 6:** Výpočet offsetu

---

**input:** indexy  $s$  a  $f$  prvního, resp. posledního chybějícího vzorku, posun sousedních oken  $a$ , varianta offsetu (*full/half*)

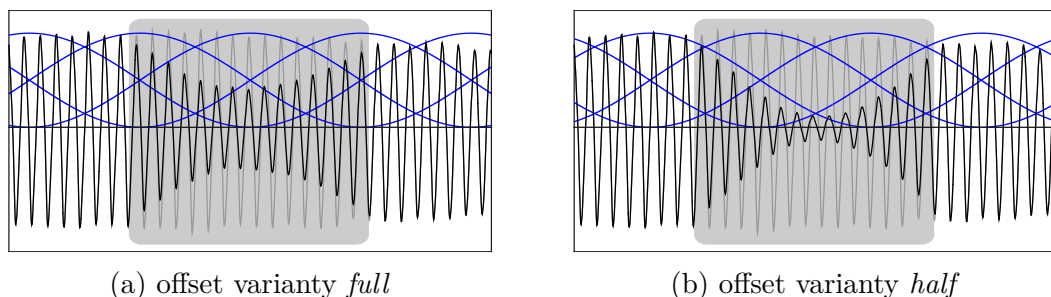
```

1 centrální index díry  $c = \lceil (s + f)/2 \rceil$ 
2 najdi  $k \in \mathbb{N}$ :  $1 + k \cdot a \leq c \leq 1 + (k + 1) \cdot a$  // tedy  $k = \lfloor (c - 1)/a \rfloor$ 
3 switch varianta offsetu do
4   | case full
5   |    $d = 1 + k \cdot a$ 
6   | end
7   | case half
8   |    $d = 1 + k \cdot a + \lceil a/2 \rceil$ 
9   | end
10 end
11 return  $off = c - d$ 

```

---

Lze tušit, že jednotlivé přístupy nepovedou ke stejným výsledkům rekonstrukce. Skutečně, obrázek 4.6 na rekonstrukci velice řídkého signálu ilustruje, že obě varianty dosahují minima energie ve středu doplněné díry, ovšem zřetelně se liší v tom, jak nízké úrovně energie ve středu díry dosáhne.



Obr. 4.6: Ilustrace rozdílnosti jednotlivých variant offsetu na rekonstruovaném signálu. Šedá oblast vyznačuje doplněný úsek, černě je rekonstrukce, šedě jako reference nepoškozený signál. Modře jsou symbolicky znázorněna okna framů.

Dále bychom proto chtěli být schopni předvídat, s kterým přístupem dosáhneme lepší rekonstrukce. Zde navrhuje predikci pomocí míry překryvu oken framu s chybějícím úsekem. Označme  $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_L$  všechna okna použitá v analýze signálu (tedy jednotlivé translace okna  $\mathbf{g}$ ). Míru<sup>3</sup>  $\mu$  definujeme jako

$$\mu = \sum_{l=1}^L \|\mathbf{g}_l - M_{\mathbf{R}}\mathbf{g}_l\|_2^2, \quad (4.23)$$

kde normu počítáme vždy z části okna příslušející chybějícímu úseku a neuvažujeme zde žádnou modulaci okna. Hodnota  $\mu$  je závislá na tvaru okna  $\mathbf{g}$ , délce okna  $w$ , posunu sousedních oken  $a$ , délce díry  $h$  a konečně také variantě offsetu. V kapitole 6 otestujeme různé kombinace těchto vstupů pro rekonstrukci signálu a ukážeme, nakolik rozdíl hodnot  $\mu$  dle vzorce (4.23) počítaných pro *full* a *half* offset koreluje se SNR rekonstrukce.

## 4.3 Modifikace v časové oblasti

V této části pokračujeme v návrzích možností kompenzace poklesu energie rekonstruovaného signálu v místě doplněné díry. Metody dosud uvedené se zaměřovaly na úpravu formulace úlohy a časově-frekvenční reprezentace signálu. Metody níže volí přístup opačný – pracují se signálem v časové oblasti. Konkrétně v části 4.3.1 již doplněný úsek násobíme vhodnou funkcí tak, abychom zvýšili jeho amplitudu. V části 4.3.2 představujeme metodu, při níž jednu díru doplňujeme postupně od okrajů, nikoliv celou naráz.

### 4.3.1 Váhování doplněného úseku v čase

Váhováním doplněného signálu v čase rozumíme postup, kdy již máme k dispozici doplněný úsek, v němž však klesá energie – proto jej násobíme vhodnou funkcí tak, abychom energii signálu podpořili. Uvažujeme-li nyní signál jako spojitou funkci na  $\langle 0, T \rangle$  s doplněným úsekem v intervalu  $\langle s, f \rangle \subset \langle 0, T \rangle$ , pak vhodná funkce  $q(t)$  splňuje následující předpoklady:

$$q(t) \text{ je spojitá a diferencovatelná na } \langle 0, T \rangle, \quad (4.24a)$$

$$\frac{dq}{dt} = 0 \quad \text{pro } t \in \{s, f\}, \quad (4.24b)$$

$$q(t) = 1 \quad \text{pro } t \notin \langle s, f \rangle, \quad (4.24c)$$

$$q(t) \text{ je neklesající na } (s, (s+f)/2) \text{ a neroustoucí na } ((s+f)/2, f). \quad (4.24d)$$

První tři podmínky zajišťují, že průběh amplitudy signálu upravujeme plynule. Poslední podmínka pak navíc zajišťuje, že energii doplňujeme dle předpokládaného tvaru jejího poklesu; ten je s použitím dříve popsáného offsetu největší ve středu doplněného úseku.

V praxi pracujeme s diskrétním signálem, tedy funkci  $q$  na úseku  $\langle s, f \rangle$  diskretizujeme tak, že dostaneme vektor  $\mathbf{q}$  délky odpovídající délce původní díry. Výslednou rekonstrukci získáme tak, že doplněný úsek násobíme po složkách vektorem  $\mathbf{q}$ .

### Volba váhovací funkce

Nyní se budeme věnovat volbě váhovací funkce  $q(t)$ , která bude splňovat podmínky (4.24) a navíc bude reflektovat konkrétní signál a doplňovaný úsek. Toho dosáhneme tak, že

<sup>3</sup>Mírou zde rozumíme obecný pojem, nikoliv funkci zavedenou korektně v souladu s teorií míry.

vhodnou funkci  $q(t)$  získáme z blízkého okolí díry, které předpokládáme, že není porušené. V těchto blízkých úsecích uměle vytvoříme další díry, které doplníme, čímž získáme rekonstrukce i s referenčními úseky nepoškozeného signálu. Na těchto uměle vytvořených dírách tedy víme, jaký je průběh poklesu energie, což přeneseme do místa díry, jejíž rekonstrukce byla původním cílem.

Obecnou úvahu zpřesňují následující kroky, podle kterých algoritmus postupuje. Předpokládáme, že původně chybějící úsek již máme doplněný. Pro ilustraci je k dispozici též grafické znázornění na obrázku 4.7.

1. Vytvoříme uměle díry v okolí původně chybějícího úseku.
2. Tyto díry doplníme.
3. V každé doplněné díře a v příslušných úsecích neporušeného signálu vypočteme průběh energie (energii počítáme po segmentech zvolené délky se středy v časech  $t_1, t_2, \dots, t_m$ , průběhy energie jsou tedy vektory délky  $m$ ).
4. Nalezneme vektor součinitelů  $\mathbf{m} \in \mathbb{R}^m$ , který bude minimalizovat rozdíl energie v originálních a doplněných úsecích (přesněji viz rovnici (4.26)).
5. Součinitele  $\mathbf{m}$  odmocníme<sup>4</sup>, čímž dostaneme vektor  $\mathbf{n}$ .
6. Provedeme symetrizaci<sup>5</sup> vektoru  $\mathbf{n}$  jako

$$n_i \leftarrow \frac{n_i + n_{m+1-i}}{2}, \quad 1 \leq i \leq m. \quad (4.25)$$

7. Funkci  $q(t)$  získáme jako kubický spline<sup>6</sup> procházející v bodech  $t_1, t_2, \dots, t_m$  hodnotami vektoru  $\mathbf{n}$  a splňující podmínku nulové derivace a jednotkové funkční hodnoty v časech  $s$  a  $f$ .
8. Vektor  $\mathbf{q}$  získáme tak, že diskretizujeme funkci  $q(t)$  na intervalu  $\langle s, f \rangle$  potřebným počtem bodů.

Blíže nyní popíšeme, jak je realizován čtvrtý krok popsaného postupu. Uvažujme, že vektory průběhů energie pro jednotlivé díry uspořádáme jako sloupce matice. Matici  $\mathbf{X}$  budou tvořit průběhy z doplněných úseků, matici  $\mathbf{Y}$  pak průběhy z příslušných úseků neporušeného signálu. Součinitele  $\mathbf{m}$  pak udává rovnice

$$\mathbf{m} = \arg \min_{\mathbf{a}} \|\mathbf{Y} - \text{diag}(\mathbf{a}) \cdot \mathbf{X}\|_2^2. \quad (4.26)$$

Zde vidíme, že vektor  $\mathbf{m}$  bude obsahovat takové koeficienty, že jimi vynásobený průběh energie doplněné díry bude v kontextu všech doplněných děr nejbližší možný energii neporušeného signálu. Úloha (4.26) je zápisem metody nejmenších čtverců, řešení proto získáme po vzorcích pomocí jednotlivých prvků matic  $\mathbf{X}$  a  $\mathbf{Y}$  jako

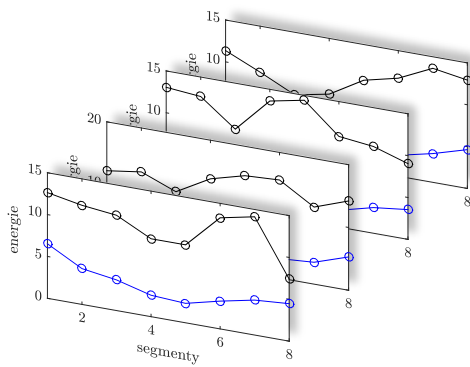
$$m_i = \frac{\sum_j y_{ij} x_{ij}}{\sum_j x_{ij}^2}. \quad (4.27)$$

**Poznámka 4.6.** Rovnice (4.26) představuje zjednodušený model, který neklade žádné podmínky na multiplikátory  $\mathbf{m}$ , například že by měly být nezáporné, případně i shora

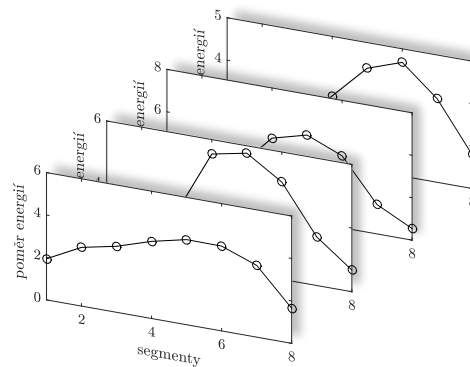
<sup>4</sup>Vektor  $\mathbf{m}$  reprezentuje poměr energií, my však hledáme poměr vzorků signálu, což rozměrově odpovídá druhé odmocnině z poměru energií.

<sup>5</sup>Zde využíváme výsledků části 4.2.2, kde jsme navrhli offset, díky kterému je průběh poklesu energie symetrický.

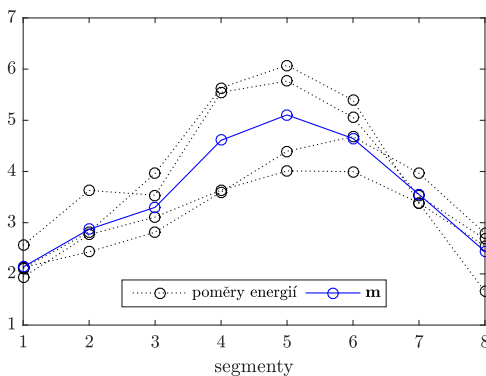
<sup>6</sup>Teoretický rozbor interpolace splinem je nad rámec této práce, odkazujeme proto na použití příslušné funkce v MATLABu (<https://www.mathworks.com/help/matlab/ref/spline.html>), případně literaturu z oblasti numerických metod [15].



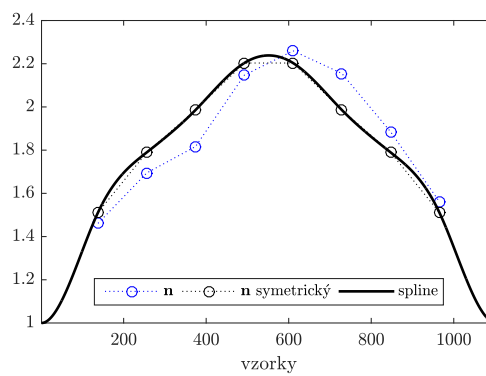
(a) průběhy energie v originálním signálu (černá) a doplněných úsecích (modrá)



(b) poměr energie původního signálu vzhledem k energii doplněného úseku pro jednotlivé rekonstruované díry



(c) poměry energií a vektor  $\mathbf{m}$  dle rovnice (4.26)



(d) symetrizace vektoru  $\mathbf{n}$  a výsledný vektor  $\mathbf{q}$  jako spline

Obr. 4.7: Vizualizace postupu pro váhování doplněného úseku v čase.

omezené vybranou hodnotou (bez tohoto omezení může být problematická situace, kdy je energie doplněného úseku blízká nule). Řešení úlohy s omezením by však bylo neúměrně náročnější než explicitní řešení pomocí vzorce (4.27) a v experimentální části ukážeme, že i se zjednodušením je popsáný koncept účinný.

**Poznámka 4.7.** Místo popsáného adaptivního způsobu výpočtu váhovací křivky  $\mathbf{q}$  by bylo možné volit i univerzálnější postup, kdy bychom se vhodnou křivku předem „naučili“ na databázi signálů a tuto pak používali. Takový přístup by byl nepochybně rychlejší a nabízí prostor pro využití metod hlubokého učení<sup>7</sup>. Jedná se však o rozšíření, které již je nad rámec tohoto textu.

### 4.3.2 Gradual inpainting

Pojmem *gradual inpainting*<sup>8</sup> rozumíme zpracování signálů v jistých stupních. Vycházíme z pozorování, že energie rekonstruovaného úseku a obecně také přesnost rekonstrukce (ve smyslu blízkosti nepoškozenému originálu) je nejvyšší při okrajích díry. Omezíme-li se na

<sup>7</sup>I zde se jedná o celé odvětví, pro které v této práci není prostor. Základní povědomí čtenář jistě získá z encyklopedického článku na [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning), případně lze odkázat na tam uvedené reference a další odkazy.

<sup>8</sup>Českým ekvivalentem by bylo *doplňování chybějícího úseku po stupních*, vzhledem k používání cizího termínu *inpainting* však ponecháváme celé označení metody nepřeložené.

jediný chybějící úsek, základní myšlenkou je v prvním stupni rekonstruovat celý tento úsek, ovšem jako řešení přijmout pouze krátké části z jeho okrajů, o kterých předpokládáme, že se nejvíce blíží původnímu signálu. Zbytek doplněného úseku pro nás zůstává neznámý (nalezené řešení neuvažujeme), což je výchozí stav pro následující stupeň. Takto pokračujeme, dokud nedoplníme celý původně chybějící úsek. Tím, že v každém stupni rekonstruujeme kratší díru než ve stupni předcházejícím, bychom měli dosahovat lepší rekonstrukce, než při jednorázovém zpracování celé díry.

Popsaný postup shrnuje algoritmus 7.

---

**Algoritmus 7:** Gradual inpainting

---

**input:** pozorovaný signál  $\mathbf{y}$ , indexy  $s$  a  $f$  prvního, resp. posledního chybějícího vzorku

- 1 nastav index stupně  $k = 0$
- 2 nastav řešení  $\mathbf{x}^{(0)} = \mathbf{y}$
- 3 **while**  $s \leq f$  **do**
- 4     najdi  $\mathbf{x}^{(k)}$  pomocí audio inpaintingu s daty  $\mathbf{x}^{(k-1)}$  a dírou mezi indexy  $s$  a  $f$
- 5     zvol parametr  $0 < r \in \mathbb{N}$
- 6      $s \leftarrow s + r$  // zkrácení díry zleva
- 7      $f \leftarrow f - r$  // zkrácení díry zprava
- 8      $k \leftarrow k + 1$
- 9 **end**
- 10 **return**  $\hat{\mathbf{x}} = \mathbf{x}^{(k)}$

---

**Poznámka 4.8.** V šestém a sedmém kroku algoritmu 7 vidíme, že zkracování díry probíhá symetricky. Toto zjednodušení můžeme využít, neboť uvažujeme zpracování jediné díry a používáme offset navržený v části 4.2.2. Řešíme tedy symetrický pokles energie a smysluplné je proto zpracovávat díru také symetricky.

V algoritmu 7 jsme neuvažovali konkrétní způsob řešení audio inpaintingu v jednotlivých stupních. Přesto však není každá volba zcela smysluplná, což ilustrujeme na váhovaném syntetizujícím modelu, který je v  $k$ -tém stupni tvaru

$$\mathbf{x}^{(k)} = D\mathbf{z}^{(k)}, \quad \mathbf{z}^{(k)} = \arg \min_{\mathbf{z}} \|\mathbf{w}^{(k)} \odot \mathbf{z}\|_1 \quad \text{vzhledem k } D\mathbf{z} \in \Gamma^{(k)}. \quad (4.28)$$

Množina přípustných řešení  $\Gamma^{(k)}$  (bez ohledu na použitý model) je dána pomocí řešení  $\mathbf{x}^{(k-1)}$  z předcházejícího stupně a aktuálních hodnot  $s, f$  jako

$$\Gamma^{(k)} = \{\mathbf{x} \in \mathbb{C}^N \mid M_{\mathbf{R}}^{(k)} \mathbf{x} = M_{\mathbf{R}}^{(k)} \mathbf{x}^{(k-1)}\}. \quad (4.29)$$

S množinou  $\Gamma^{(k)}$  se musí měnit i odpovídající operátor pro výběr nepoškozených vzorků určený indexy  $s$  a  $f$ , který jsme označili  $M_{\mathbf{R}}^{(k)}$ . Protože v důsledku šestého a sedmého kroku algoritmu 7 odpovídá operátor  $M_{\mathbf{R}}^{(k)}$  užšímu výběru vzorků než operátor  $M_{\mathbf{R}}^{(k-1)}$  (klademe totiž stejnou podmínku na vzorky určené  $M_{\mathbf{R}}^{(k-1)}$ , navíc však ještě na  $2r$  vzorků nově považovaných za rekonstruované), pro množiny přípustných řešení po sobě následujících stupňů platí

$$\Gamma^{(k)} \subset \Gamma^{(k-1)}. \quad (4.30)$$

Opačná inkluze zřejmě nemůže platit, podstatná však pro nás je příslušnost

$$\mathbf{x}^{(k-1)} \in \Gamma^{(k)}, \quad (4.31)$$

kterou snadno ověříme dosazením  $\mathbf{x}^{(k-1)}$  do podmínky v rovnici (4.29).

Uvažujme nyní, že atomy v úloze (4.28) neváhujeme, tedy  $\mathbf{w}^{(k)} \odot \mathbf{z} = \mathbf{z} \forall k$ . Potom ve dvou po sobě následujících stupních se příslušné úlohy ve tvaru (4.28) liší pouze v množině přípustných řešení, ovšem podle (4.31) platí  $\mathbf{x}^{(k)} = D\mathbf{z}^{(k)} \in \Gamma^{(k+1)}$ . Celkově tedy dostáváme, že výsledné řešení gradual inpaintingu je rovno řešení  $\mathbf{x}^{(1)}$  z prvního stupně, neboli nijak se neliší od inpaintingu bez zavedení stupňů.

Z praktického hlediska tedy předchozí úvaha implikuje, že gradual inpainting využívající  $\ell_1$  minimalizaci musí ve čtvrtém kroku dle algoritmu 7 využívat buď skutečně váhovanou variantu (pak se v po sobě následujících stupních sice neliší množiny přípustných řešení, ale liší se váhovací vektor  $\mathbf{w}^{(k)}$  v účelové funkci), nebo jinou úpravu průběžného řešení, kterou může být například váhování signálu v časové oblasti dle části 4.3.1.

## 5 Nekonvexní metody založené na řídkosti

V této kapitole se budeme věnovat dvěma metodám pro audio inpainting založený na řídkosti, které vycházejí z řešení nekonvexních úloh. Jsou jimi algoritmus *Orthogonal Matching Pursuit* (OMP) [29], což byl vůbec první přístup využívající řídkost v audio inpaintingu [3], a nový algoritmus *Sparse Audio Inpainter* (SPAIN). Rozdílem oproti konvexním metodám je, že se jedná o postupy heuristické, které nezaručují konvergenci k optimálnímu řešení problému. Výhodou naopak je, že pracují přímo s řídkostí, nikoliv s  $\ell_1$  normou.

Pro bližší pochopení nekonvexních metod připomeňme obecnou formulaci audio inpaintingu využívajícího řídkost, jak byla vyslovena v úvodu kapitoly 4:

$$\arg \min_{\mathbf{x}} \|\mathbf{z}\|_0 \quad \text{vzhledem k} \quad F(\mathbf{x}, \mathbf{z}) = 0, \quad \mathbf{x} \in \Gamma, \quad \mathbf{z} \in \mathbb{C}^P. \quad (4.1)$$

Konvexní přístup byl založen na relaxaci účelové funkce v (4.1). Nekonvexní naopak relaxuje podmínku  $F(\mathbf{x}, \mathbf{z}) = 0$  na přibližný tvar  $F(\mathbf{x}, \mathbf{z}) \approx 0$ , což nám otevírá prostor pro hledání řídkého řešení. Detailní rozbor takto relaxované úlohy následuje u jednotlivých algoritmů v částech 5.1 a 5.2.

### 5.1 Orthogonal Matching Pursuit

Audio inpainting pomocí OMP můžeme formulovat jako optimalizační úlohu

$$\arg \min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{vzhledem k} \quad \|M_{\mathbf{R}}D\mathbf{z} - \mathbf{y}_{\mathbf{R}}\|_2 \leq \varepsilon. \quad (5.1)$$

Rozebereme-li blíže formulaci (5.1), zjistíme, že jsme nepoužili proměnnou pro rekonstruovaný signál  $\mathbf{x}$ , tudíž formulace zcela nezapadá do obecného rámce (4.1). Způsobeno je to tím, že výstupem OMP vůbec nemusí být přípustné řešení pro naši původní formulaci audio inpaintingu (4.1)<sup>1</sup>, tedy takové, aby platilo  $D\mathbf{z} \in \Gamma$ . Výsledné řešení  $\mathbf{x}$  proto získáme jako vektor  $\mathbf{y}$ , ve kterém pouze doplníme chybějící úseky vzorky z  $D\mathbf{z}$ . Potom zjevně  $M_{\mathbf{R}}\mathbf{x} = \mathbf{y}_{\mathbf{R}}$ , tudíž by problém (5.1) mohl být alternativně formulován i s použitím takto definovaného  $\mathbf{x} \in \Gamma$ .

Pro přiblížení tvaru algoritmu zavedeme nové označení  $D_{\mathbf{R}} = M_{\mathbf{R}}D$ . Cílem potom je najít co nejřidší a současně dostatečně přesnou reprezentaci  $\mathbf{y}_{\mathbf{R}}$  pomocí atomů  $D_{\mathbf{R}}$ . Iterativní hladový algoritmus OMP pracuje tak, že postupně vybírá jednotlivé atomy slovníku  $D_{\mathbf{R}}$ , přičemž v každé iteraci přibírá takový atom, který je nejvíce korelovaný s aktuálním reziduem. Následně provedeme ortogonální projekci  $\mathbf{y}_{\mathbf{R}}$  na lineární podprostor generovaný dosud vybranými atomy, čímž získáme řídkou aproximaci a nové reziduum a pokračujeme další iterací.

Pro formální zápis tohoto postupu zavedeme označení  $D_{\mathbf{R}}(\Omega)$ ,  $\Omega \subseteq \{1, 2, \dots, P\}$ , což reprezentuje syntézu pomocí takových atomů  $D_{\mathbf{R}}$ , jejichž indexy spadají do množiny  $\Omega$ . Dále pro  $n = 1, 2, \dots, P$  označíme  $\mathbf{d}_{\mathbf{R}}^n$   $n$ -tý atom slovníku  $D_{\mathbf{R}}$ . Algoritmus OMP pak shrnuje algoritmus 8.

---

<sup>1</sup>tj. konzistentní řešení

---

**Algoritmus 8:** Orthogonal Matching Pursuit (OMP) pro audio inpainting

---

**input:** množina  $\Gamma$ , nepoškozená část signálu  $\mathbf{y}_R$ , operátory syntézy s úplným slovníkem  $D$  a s redukováným slovníkem  $D_R$ , tolerance  $\varepsilon$ , maximální počet iterací  $K \leq P$

- 1 nastav čítač iterací  $k = 1$
- 2 nastav nosič  $\Omega^{(0)} = \emptyset$
- 3 nastav reziduum  $\mathbf{r}^{(0)} = \mathbf{y}_R$
- 4 **while**  $\|\mathbf{r}^{(k)}\|_2 > \varepsilon \wedge k \leq K$  **do**
- 5      $m = \arg \max_n |\langle \mathbf{r}^{(k-1)}, \mathbf{d}_R^n \rangle|$      // výběr nejvíce korelovaného atomu
- 6      $\Omega^{(k)} = \Omega^{(k-1)} \cup \{m\}$
- 7      $P^{(k)} = D_R(\Omega^{(k)}) \left( D_R(\Omega^{(k)}) \right)^+$      // operátor ortogonální projekce
- 8      $\mathbf{r}^{(k)} = (\text{Id} - P^{(k)})\mathbf{y}_R$
- 9      $k \leftarrow k + 1$
- 10 **end**
- 11  $\hat{\mathbf{z}} = \left( D_R(\Omega^{(k)}) \right)^+ \mathbf{y}_R$
- 12 **return**  $\hat{\mathbf{x}} = \mathcal{P}_\Gamma(D\hat{\mathbf{z}})$

---

Do závěru algoritmu 8 (kroky 11 a 12) bylo shrnuto i získání výsledné rekonstrukce signálu – nalezenou řídkou reprezentaci  $\hat{\mathbf{z}}$  zde použijeme k syntéze již pomocí celého slovníku  $D$  a poté přidáváme projekci na množinu přípustných řešení  $\Gamma$  v časové oblasti.

Pro OMP je podstatný osmý krok algoritmu 8, ve kterém vystupuje ortogonální projekce na podprostor generovaný vybranými atomy  $D_R(\Omega^{(k)})$ . V důsledku tohoto kroku je reziduum vždy kolmé ke všem dosud vybraným atomům, proto žádný atom není v průběhu algoritmu vybrán podruhé. Dalším důsledkem je, že v každém kroku roste o jedna dimenze prostoru, který je generován pomocí  $D_R(\Omega^{(k)})$ . To platí proto, že reziduum (a tudíž ani nově vybraný atom) nespadá do lineárního obalu atomů dosud vybraných. Je-li tedy cílem reprezentovat vektor  $\mathbf{y}_R \in \mathbb{C}^N$ , pak algoritmus 8 skončí nejpozději po  $N$  krocích, neboť  $D_R(\Omega^{(N)})$  dle předchozí úvahy odpovídá bázi prostoru  $\mathbb{C}^{N^2}$ .

Ve vstupech algoritmu 8 jsme uvažovali podmínku  $K \leq P$ , která je v případě redundantního slovníku méně přísná než právě odvozená konvergence v nejvýše  $N$  krocích. Je však opodstatněná tím, že slovník  $D_R$  obsahuje právě  $P$  atomů, tudíž vybírání  $(P + 1)$  atomů v rámci algoritmu 8 by nebylo přípustné.

Z důvodu výpočtu projekce v sedmém kroku, který můžeme rozepsat jako

$$P^{(k)} = D_R(\Omega^{(k)}) \left( D_R(\Omega^{(k)})^* D_R(\Omega^{(k)}) \right)^{-1} D_R(\Omega^{(k)})^*, \quad (5.2)$$

je algoritmus OMP výpočetně velice náročný. Například v toolboxu *Sparsify* [6] pro MATLAB je však k dispozici efektivní implementace OMP, která místo pseudoinverze využívá QR rozklad slovníku  $D_R(\Omega^{(k)})$  a tento rozklad aktualizuje s měnícím se nosičem  $\Omega^{(k)}$ .

**Poznámka 5.1** (pseudoinverze). Moore-Penroseovu pseudoinverzi operátoru, kterou využívá sedmý krok algoritmu 8, jsme v této práci korektně nedefinovali, neboť pro pochopení tématu není stěžejní. Jedná o zobecnění inverze na operátory, které nejsou bijektivní

---

<sup>2</sup>Teoreticky může být počet kroků i výrazně menší, pokud je možná signál  $\mathbf{y}_R$  přesně syntetizovat pomocí malého počtu atomů.

(v případě matic se jedná o matice singulární nebo obdélníkové). Definici a řadu vlastností lze nalézt například v [5]. Na tomto místě je pro nás důležitá vlastnost

$$M^+ = (M^*M)^{-1} M^*, \quad (5.3)$$

kteřou jsme využili v rovnici (5.2). Podstatné však je, že tato vlastnost platí pouze pokud  $M$  je injekce (v případě matic má lineárně nezávislé sloupce). To zřejmě nemůže platit v algoritmu 8 pro  $k > N$ , tudíž ačkoliv jeho sedmý krok je formulován dostatečně obecně, vyjádření (5.2) je platné pouze pro  $N$  iterací.

**Poznámka 5.2** (váhování atomů v OMP). V článku [3], kde byl algoritmus OMP použit poprvé pro audio inpainting, bylo zavedeno podobné váhování atomů jako v části 4.2.1. Konkrétně byly atomy  $\mathbf{d}_R^n$  redukovaného slovníku  $D_R$  v rámci inicializace algoritmu OMP normovány, což odpovídá našemu váhování podle  $\ell_2$  normy.

## 5.2 Sparse Audio Inpainter

Sparse Audio Inpainter (SPAIN) je nový algoritmus pro audio inpainting. Vzorem pro něj je Sparse Audio Declipper (SPADE) [21], který, jak název napovídá, je uzpůsoben problému audio declippingu. Souvislost mezi SPAIN a SPADE je v obecně shodné formulaci úlohy

$$\arg \min_{\mathbf{x}, \mathbf{z}} \|\mathbf{z}\|_0 \quad \text{vzhledem k } \mathbf{x} \in \Gamma \text{ a } \|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2 \leq \varepsilon, \quad (5.4a)$$

$$\arg \min_{\mathbf{x}, \mathbf{z}} \|\mathbf{z}\|_0 \quad \text{vzhledem k } \mathbf{x} \in \Gamma \text{ a } \|\mathbf{x} - D\mathbf{z}\|_2 \leq \varepsilon, \quad (5.4b)$$

kde (5.4a) nazýváme v souladu s metodami konvexními tzv. analyzující variantou, zatímco (5.4b) je varianta syntetizující. Formulacemi (5.4) také říkáme, že hledaný signál  $\mathbf{x}$  a řídké koeficienty  $\mathbf{z}$  si přímo neodpovídají. Rozdíl mezi SPAIN a SPADE pak je v množině přípustných signálů  $\Gamma$ , která v je případě audio declippingu více omezující, tudíž je též declipping snáze<sup>3</sup> řešitelným problémem.

Dále popíšeme obecný princip algoritmu a v částech 5.2.2 a 5.2.3 odvodíme konkrétní podoby analyzující (A-SPAIN), resp. syntetizující (S-SPAIN) varianty. Ačkoliv A-SPAIN je přímou adaptací A-SPADE, jak je prezentováno v [21], úplné odvození algoritmu bylo dosud publikováno pouze v [41]. Syntetizující varianta je pak odvozována nově v jiné podobě, než měl původní S-SPADE. Zde se budeme držet mimo [41] také článků [40, 28].

### 5.2.1 Princip algoritmu

Cílem bude využít *Alternating Direction Method of Multipliers* (ADMM) [7]. Tato metoda je primárně určena k řešení konvexního problému

$$\arg \min f(\mathbf{u}) + g(L\mathbf{u}), \quad (5.5)$$

kde  $f, g$  jsou konvexní funkce a  $L$  je lineární operátor. Problém (5.5) můžeme přepsat do omezeného tvaru zavedením pomocné proměnné jako

$$\arg \min f(\mathbf{u}) + g(\mathbf{v}) \quad \text{vzhledem k } \mathbf{v} = L\mathbf{u}. \quad (5.6)$$

<sup>3</sup>Snáze nikoliv z hlediska výpočetního či implementačního, ale protože o hledaném signálu máme a priori více informací.

Úloha (5.6) je pak řešitelná pomocí minimalizace tzv. *rozšířeného lagrangiánu* s (kladným) penalizačním parametrem  $\rho$ :

$$\mathcal{L}_\rho(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}) = f(\mathbf{u}) + g(\mathbf{v}) + \boldsymbol{\lambda}^\top (L\mathbf{u} - \mathbf{v}) + \frac{\rho}{2} \|L\mathbf{u} - \mathbf{v}\|_2^2. \quad (5.7)$$

Obecné schéma algoritmu ADMM pak sestává z následujících kroků:

$$\mathbf{u}^{(i+1)} = \arg \min_{\mathbf{u}} \mathcal{L}_\rho(\mathbf{u}, \mathbf{v}^{(i)}, \boldsymbol{\lambda}^{(i)}), \quad (5.8a)$$

$$\mathbf{v}^{(i+1)} = \arg \min_{\mathbf{v}} \mathcal{L}_\rho(\mathbf{u}^{(i+1)}, \mathbf{v}, \boldsymbol{\lambda}^{(i)}), \quad (5.8b)$$

$$\boldsymbol{\lambda}^{(i+1)} = \boldsymbol{\lambda}^{(i)} + \rho (L\mathbf{u}^{(i+1)} - \mathbf{v}^{(i+1)}). \quad (5.8c)$$

Alternativou je tzv. *scaled form*<sup>4</sup>, ve které definujeme *reziduum*  $\mathbf{r} = L\mathbf{u} - \mathbf{v}$  a novou duální proměnnou  $\boldsymbol{\sigma} = \boldsymbol{\lambda}/\rho$ . Poslední dva členy rozšířeného lagrangiánu pak přepíšeme jako

$$\boldsymbol{\lambda}^\top \mathbf{r} + \frac{\rho}{2} \|\mathbf{r}\|_2^2 = \frac{\rho}{2} \left\| \mathbf{r} + \frac{1}{\rho} \boldsymbol{\lambda} \right\|_2^2 - \frac{1}{2\rho} \|\boldsymbol{\lambda}\|_2^2 = \frac{\rho}{2} \|\mathbf{r} + \boldsymbol{\sigma}\|_2^2 - \frac{\rho}{2} \|\boldsymbol{\sigma}\|_2^2. \quad (5.9)$$

S takto zavedenou substitucí dostáváme nový tvar rozšířeného lagrangiánu

$$\mathcal{L}_\rho(\mathbf{u}, \mathbf{v}, \boldsymbol{\sigma}) = f(\mathbf{u}) + g(\mathbf{v}) + \frac{\rho}{2} \|\mathbf{r} + \boldsymbol{\sigma}\|_2^2 - \frac{\rho}{2} \|\boldsymbol{\sigma}\|_2^2. \quad (5.10)$$

Ve schématu ADMM (5.8) pak vynecháním členů, které jsou v jednotlivých krocích konstantní, dostaneme konkrétnější podobu

$$\mathbf{u}^{(i+1)} = \arg \min_{\mathbf{u}} f(\mathbf{u}) + \frac{\rho}{2} \|L\mathbf{u} - \mathbf{v}^{(i)} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.11a)$$

$$\mathbf{v}^{(i+1)} = \arg \min_{\mathbf{v}} g(\mathbf{v}) + \frac{\rho}{2} \|L\mathbf{u}^{(i+1)} - \mathbf{v} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.11b)$$

$$\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)} + L\mathbf{u}^{(i+1)} - \mathbf{v}^{(i+1)}. \quad (5.11c)$$

**Poznámka 5.3.** V komplexním případě je třeba rozšířený lagrangián přeformulovat tak, aby zůstal reálnou funkcí. Místo rovnice (5.7) proto používáme novou definici

$$\mathcal{L}_\rho(\mathbf{u}, \mathbf{v}, \boldsymbol{\lambda}) = f(\mathbf{u}) + g(\mathbf{v}) + \boldsymbol{\lambda}^\top \begin{bmatrix} \Re(L\mathbf{u} - \mathbf{v}) \\ \Im(L\mathbf{u} - \mathbf{v}) \end{bmatrix} + \frac{\rho}{2} \left\| \begin{bmatrix} \Re(L\mathbf{u} - \mathbf{v}) \\ \Im(L\mathbf{u} - \mathbf{v}) \end{bmatrix} \right\|_2^2, \quad (5.12)$$

kde pro  $\mathbf{v} \in \mathbb{C}^P$  je duální proměnná  $\boldsymbol{\lambda} \in \mathbb{R}^{2P}$ . Rovnice (5.9) pak opět vede na rozšířený lagrangián ve tvaru (5.10), pouze s tím rozdílem, že  $\mathbf{r}, \boldsymbol{\sigma} \in \mathbb{R}^{2P}$  namísto  $\mathbb{C}^P$ . V praxi není tato úprava nutná a můžeme pracovat s komplexními vektory  $\mathbf{r}$  a  $\boldsymbol{\sigma}$ , protože pro libovolný komplexní vektor  $\mathbf{c}$  platí

$$\|\mathbf{c}\|_2^2 = \left\| \begin{bmatrix} \Re(\mathbf{c}) \\ \Im(\mathbf{c}) \end{bmatrix} \right\|_2^2. \quad (5.13)$$

Podproblémy (5.11a) a (5.11b) tudíž mohou zůstat v komplexních proměnných beze změny a (5.11c) je pouze součtem  $\boldsymbol{\sigma}$  a  $\mathbf{r}$ , který je v  $\mathbb{C}^P$  ekvivalentní se sčítáním reálné a imaginární části zvlášť.

<sup>4</sup>Můžeme hovořit například o škálované formě, v české literatuře však uspokojivý překlad nebyl nalezen.

Nyní se můžeme vrátit k souvislosti ADMM s formulací (5.4). Zavedeme-li neznámou  $k$  pro řídkost vektoru  $\mathbf{z}$ , můžeme (5.4) ekvivalentně přepsat následovně (jedním zápisem sdružujeme analyzující i syntetizující variantu):

$$\arg \min_{\mathbf{x}, \mathbf{z}} k \quad \text{vzhledem k } \mathbf{x} \in \Gamma, \quad \|\mathbf{z}\|_0 \leq k \quad \text{a} \quad \begin{cases} \|A\mathbf{x} - \mathbf{z}\|_2 \leq \varepsilon, \\ \|\mathbf{x} - D\mathbf{z}\|_2 \leq \varepsilon. \end{cases} \quad (5.14)$$

Tento tvar ještě ADMM neodpovídá, ovšem pokud budeme uvažovat  $k$  pevné, můžeme s využitím indikátorových funkcí psát

$$\arg \min_{\mathbf{x}, \mathbf{z}} \iota_{\Gamma}(\mathbf{x}) + \iota_{\ell_0 \leq k}(\mathbf{z}) \quad \text{vzhledem k} \quad \begin{cases} A\mathbf{x} - \mathbf{z} = \mathbf{0}, \\ D\mathbf{z} - \mathbf{x} = \mathbf{0}, \end{cases} \quad (5.15)$$

kde jako  $\ell_0 \leq k$  zkráceně značíme množinu  $k$ -řídkých vektorů  $\{\mathbf{z}, \|\mathbf{z}\|_0 \leq k\}$ . Navíc jsme formulovali přísnější podmínku svazující dohromady  $\mathbf{x}$  a  $\mathbf{z}$ . Tato úprava názorně ukazuje, že úloha (5.15) nabývá přesně tvaru úlohy pro ADMM (5.6). Přístup algoritmu SPAIN je však heuristický v tom, že řešíme jednotlivé kroky ADMM pro úlohu (5.15) (pro pevné nízké  $k$ ), ačkoliv indikátorová funkce množiny  $k$ -řídkých vektorů není konvexní. Následně v závislosti na volbě jistých parametrů zvyšujeme  $k$  a pokračujeme opět s kroky ADMM, a to dokud není splněna podmínka  $\|\mathbf{r}\|_2 \leq \varepsilon$ , čímž se vracíme k původní podmínce pro vazbu  $\mathbf{x}$  a  $\mathbf{z}$ .

## 5.2.2 A-SPAIN

Pro odvození přesného tvaru algoritmu se vrátíme k analyzující formulaci (5.15), tedy

$$\arg \min_{\mathbf{x}, \mathbf{z}} \iota_{\Gamma}(\mathbf{x}) + \iota_{\ell_0 \leq k}(\mathbf{z}) \quad \text{vzhledem k} \quad A\mathbf{x} - \mathbf{z} = \mathbf{0}, \quad (5.16)$$

kde označíme  $f(\mathbf{x}) = \iota_{\Gamma}(\mathbf{x})$  a  $g(\mathbf{z}) = \iota_{\ell_0 \leq k}(\mathbf{z})$ . Kroky ADMM (5.11) pak budou tvaru

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \iota_{\Gamma}(\mathbf{x}) + \frac{\rho}{2} \|A\mathbf{x} - \mathbf{z}^{(i)} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.17a)$$

$$\mathbf{z}^{(i+1)} = \arg \min_{\mathbf{z}} \iota_{\ell_0 \leq k}(\mathbf{z}) + \frac{\rho}{2} \|A\mathbf{x}^{(i+1)} - \mathbf{z} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.17b)$$

$$\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)} + A\mathbf{x}^{(i+1)} - \mathbf{z}^{(i+1)}. \quad (5.17c)$$

Řešení jednotlivých minimalizačních podproblémů není zcela zřejmé, proto formulujeme jejich úpravu pomocí níže následujících lemmat. Podproblém (5.17a) uvažujeme v ekvivalentním omezeném tvaru

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{z}^{(i)} + \boldsymbol{\sigma}^{(i)}\|_2^2 \quad \text{vzhledem k } \mathbf{x} \in \Gamma, \quad (5.18)$$

kde jsme využili toho, že při minimalizaci indikátorové funkce nehraje žádnou roli kladný skalár  $\rho/2$ .

**Lemma 5.4.** *Předpokládejme, že operátory  $A$  a  $D$  přísluší Parsevalovu framu, tudíž  $DA = \text{Id}$ . Pak optimalizační úloha (5.18) je ekvivalentní úloze*

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|\mathbf{x} - D(\mathbf{z}^{(i)} - \boldsymbol{\sigma}^{(i)})\|_2^2 \quad \text{vzhledem k } \mathbf{x} \in \Gamma \quad (5.19)$$

$$= \mathcal{P}_{\Gamma} \left( D(\mathbf{z}^{(i)} - \boldsymbol{\sigma}^{(i)}) \right). \quad (5.20)$$

*Důkaz (podle [41]).* Označme  $\mathbf{s} = \mathbf{z}^{(i)} - \boldsymbol{\sigma}^{(i)}$  a zavedme (jednoznačný) rozklad

$$\mathbf{s} = A\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad (5.21)$$

kde  $A\boldsymbol{\xi} \in \mathcal{R}(A)$ ,  $\boldsymbol{\varepsilon} \perp \mathcal{R}(A)$  a jako  $\mathcal{R}(A)$  značíme obor hodnot operátoru  $A$ . S novým označením můžeme rovnici (5.18) přepsat jako

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|A(\mathbf{x} - \boldsymbol{\xi}) - \boldsymbol{\varepsilon}\|_2^2 \quad \text{vzhledem k } \mathbf{x} \in \Gamma. \quad (5.22)$$

Protože zřejmě  $A(\mathbf{x} - \boldsymbol{\xi}) \in \mathcal{R}(A)$  a  $\boldsymbol{\varepsilon} \perp \mathcal{R}(A)$ , s použitím Pythagorovy věty máme

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|A(\mathbf{x} - \boldsymbol{\xi})\|_2^2 + \|\boldsymbol{\varepsilon}\|_2^2 \quad \text{vzhledem k } \mathbf{x} \in \Gamma. \quad (5.23)$$

Protože hledáme argument minima, člen  $\|\boldsymbol{\varepsilon}\|_2^2$  nemusíme uvažovat, neboť je na  $\mathbf{x}$  nezávislý. Pokud dále využijeme vlastnosti Parsevalova framu (2.11), dostaneme pro (5.23) ekvivalentní vyjádření

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|DA(\mathbf{x} - \boldsymbol{\xi})\|_2^2 \quad \text{vzhledem k } \mathbf{x} \in \Gamma, \quad (5.24)$$

kde  $DA = \text{Id}$ . Závěrem ukážeme, že  $\boldsymbol{\xi} = D\mathbf{s}$ . Protože podle (5.21)  $A\boldsymbol{\xi} = \mathbf{s} - \boldsymbol{\varepsilon}$  a dle předpokladu  $DA = \text{Id}$ , platí

$$\boldsymbol{\xi} = D(\mathbf{s} - \boldsymbol{\varepsilon}) = D\mathbf{s} - D\boldsymbol{\varepsilon}. \quad (5.25)$$

Víme, že  $\boldsymbol{\varepsilon} \perp \mathcal{R}(A) \Leftrightarrow \langle \boldsymbol{\varepsilon}, A\boldsymbol{\omega} \rangle = 0 \forall \boldsymbol{\omega}$ . Protože operátor syntézy  $D$  je adjungovaným operátorem k operátoru analýzy  $A$ , platí  $\langle \boldsymbol{\varepsilon}, A\boldsymbol{\omega} \rangle = \langle D\boldsymbol{\varepsilon}, \boldsymbol{\omega} \rangle \forall \boldsymbol{\varepsilon} \forall \boldsymbol{\omega}$ , tudíž  $\langle D\boldsymbol{\varepsilon}, \boldsymbol{\omega} \rangle = 0 \forall \boldsymbol{\omega} \Leftrightarrow D\boldsymbol{\varepsilon} = 0$ , z čehož dosazením do (5.25) plyne  $\boldsymbol{\xi} = D\mathbf{s}$ . Potom  $\|\mathbf{x} - \boldsymbol{\xi}\|_2^2 = \|\mathbf{x} - D\mathbf{s}\|_2^2 = \|\mathbf{x} - D(\mathbf{z}^{(i)} - \boldsymbol{\sigma}^{(i)})\|_2^2$ , čímž je tvrzení dokázáno.  $\square$

Obdobně přepíšeme podproblém (5.17b) do omezeného tvaru jako

$$\mathbf{z}^{(i+1)} = \arg \min_{\mathbf{z}} \|A\mathbf{x}^{(i+1)} - \mathbf{z} + \boldsymbol{\sigma}^{(i)}\|_2^2 \quad \text{vzhledem k } \|\mathbf{z}\|_0 \leq k. \quad (5.26)$$

Řešením této úlohy je tvrdé prahování  $\mathcal{H}_k$ , které vynuluje všechny složky argumentu mimo  $k$  složek s největší absolutní hodnotou, tedy

$$\mathbf{z}^{(i+1)} = \mathcal{H}_k(A\mathbf{x}^{(i+1)} + \boldsymbol{\sigma}^{(i)}). \quad (5.27)$$

Tento výsledek plyne z následujícího obecného lemmatu.

**Lemma 5.5.** *Pro libovolné pevně zvolené  $\mathbf{s} \in \mathbb{C}^P$  a  $0 \leq k \leq P$  platí*

$$\arg \min_{\mathbf{z} \in \mathbb{C}^P, \|\mathbf{z}\|_0 \leq k} \|\mathbf{z} - \mathbf{s}\|_2^2 = \mathcal{H}_k(\mathbf{s}). \quad (5.28)$$

*Důkaz (podle [41]).* Je zřejmé, že řešení úlohy (5.28)  $\mathbf{z}_{\text{opt}}$  je řídkosti  $k$  (předpokládáme, že nenastává triviální případ, kdy  $\|\mathbf{s}\|_0 < k$  a tudíž  $\mathbf{z}_{\text{opt}} = \mathbf{s}$ ). Funkci  $\|\mathbf{z} - \mathbf{s}\|_2^2 = \sum_{i=1}^P |z_i - s_i|^2$  rozdělíme na dvě části – součet čtverců koeficientů  $\mathbf{s}$ , kterým odpovídají nulové koeficienty  $\mathbf{z}$ , jak vynucuje podmínka  $\|\mathbf{z}\|_0 \leq k$ , a zbytek (tedy část, kde složky  $\mathbf{z}$  jsou nenulové). První část zřejmě minimalizujeme tak, že nulové koeficienty  $\mathbf{z}_{\text{opt}}$  budou odpovídat  $(P - k)$  v absolutní hodnotě nejmenším koeficientům  $\mathbf{s}$ . Druhou část pak minimalizujeme tak, že zbývajících  $k$  koeficientů položíme rovných příslušným složkám  $\mathbf{s}$ . Výsledek je tudíž z definice roven  $\mathbf{z}_{\text{opt}} = \mathcal{H}_k(\mathbf{s})$ .  $\square$

Všechny dosavadní úpravy shrnuje algoritmus 9, ve kterém navíc zavádíme parametry  $r, s$ , které řídí zvyšování řidkosti  $k$  průběhu iterací ADMM. Kroky (5.17a) a (5.17b) zde byly zaměněny, což slouží k názornému porovnání s dále odvozeným algoritmem S-SPAIN. Podle [7] tato záměna v konvexním případě neovlivňuje konvergenci algoritmu, předpokládáme tedy, že nebude mít ani stěžejní vliv na výsledek tohoto heuristického algoritmu.

---

### Algoritmus 9: A-SPAIN

---

**input:** pozorovaný signál  $\mathbf{y}$ , operátory  $A, D$ , množina přípustných řešení  $\Gamma$ ,  
parametry  $s, r \in \mathbb{N}, \varepsilon > 0$

- 1  $\mathbf{x}^{(0)} = \mathbf{y}, \boldsymbol{\sigma}^{(0)} = \mathbf{0}, i = 0, k = s$
- 2  $\mathbf{z}^{(i+1)} = \mathcal{H}_k \left( A\mathbf{x}^{(i)} + \boldsymbol{\sigma}^{(i)} \right)$  // odpovídá (5.17b)
- 3  $\mathbf{x}^{(i+1)} = \mathcal{P}_\Gamma \left( D(\mathbf{z}^{(i+1)} - \boldsymbol{\sigma}^{(i)}) \right)$  // odpovídá (5.17a)
- 4 **if**  $\|A\mathbf{x}^{(i+1)} - \mathbf{z}^{(i+1)}\|_2 \leq \varepsilon$  **then**
- 5 | terminate
- 6 **else**
- 7 |  $\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)} + A\mathbf{x}^{(i+1)} - \mathbf{z}^{(i+1)}$  // odpovídá (5.17c)
- 8 |  $i \leftarrow i + 1$
- 9 | **if**  $i \bmod r = 0$  **then**
- 10 | |  $k \leftarrow k + s$
- 11 | **end**
- 12 | go to 2
- 13 **end**
- 14 **return**  $\hat{\mathbf{x}} = \mathbf{x}^{(i+1)}$

---

### 5.2.3 S-SPAIN

Jako v případě A-SPAIN se i nyní vrátíme k formulaci (5.15), tentokrát však k její syntetizující variantě, tedy

$$\arg \min_{\mathbf{x}, \mathbf{z}} \iota_\Gamma(\mathbf{x}) + \iota_{\ell_0 \leq k}(\mathbf{z}) \quad \text{vzhledem k} \quad D\mathbf{z} - \mathbf{x} = \mathbf{0}, \quad (5.29)$$

kde označíme  $g(\mathbf{x}) = \iota_\Gamma(\mathbf{x})$  a  $f(\mathbf{z}) = \iota_{\ell_0 \leq k}(\mathbf{z})$ . Označení funkcí je opačné oproti odvození A-SPAINu, protože v tomto případě je  $\mathbf{z}$  z pohledu ADMM původní proměnnou (tedy příslušná funkce je  $f$ ), zatímco  $\mathbf{x}$  je proměnná pomocná definovaná jako  $D\mathbf{z}$ .

Kroky ADMM (5.11) tentokrát mají tvar

$$\mathbf{z}^{(i+1)} = \arg \min_{\mathbf{z}} \iota_{\ell_0 \leq k}(\mathbf{z}) + \frac{\rho}{2} \|D\mathbf{z} - \mathbf{x}^{(i)} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.30a)$$

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \iota_\Gamma(\mathbf{x}) + \frac{\rho}{2} \|D\mathbf{z}^{(i+1)} - \mathbf{x} + \boldsymbol{\sigma}^{(i)}\|_2^2, \quad (5.30b)$$

$$\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)} + D\mathbf{z}^{(i+1)} - \mathbf{x}^{(i+1)}. \quad (5.30c)$$

V rozboru jednotlivých kroků tentokrát začneme problémem (5.30b), který přepíšeme do omezeného tvaru (opět můžeme zanedbat parametr  $\rho$ )

$$\mathbf{x}^{(i+1)} = \arg \min_{\mathbf{x}} \|D\mathbf{z}^{(i+1)} - \mathbf{x} + \boldsymbol{\sigma}^{(i)}\|_2^2 \quad \text{vzhledem k} \quad \mathbf{x} \in \Gamma. \quad (5.31)$$

Tento podproblém je pouze jiným zápisem projekce na množinu  $\Gamma$ , konkrétně

$$\mathbf{x}^{(i+1)} = \mathcal{P}_\Gamma \left( D\mathbf{z}^{(i+1)} + \boldsymbol{\sigma}^{(i)} \right). \quad (5.32)$$

Komplikovanější je podproblém (5.30a), který přepíšeme jako

$$\mathbf{z}^{(i+1)} = \arg \min_{\mathbf{z}} \left\| D\mathbf{z} - \mathbf{x}^{(i)} + \boldsymbol{\sigma}^{(i)} \right\|_2^2 \quad \text{vzhledem k} \quad \|\mathbf{z}\|_0 \leq k. \quad (5.33)$$

Zde se jedná o tzv. *sparse synthesis problem*, kde hledáme  $k$ -řídkou lineární kombinaci atomů  $D$ , která bude nejbližší vektoru  $(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)})$ . Na rozdíl od analyzujícího modelu a obdobně vypadající úlohy (5.26), kterou jsme vyřešili tvrdým prahováním, je úloha (5.33) NP-těžkým problémem. Příčinou je zde předpokládaná neortogonalita  $D$ . Uchýlíme se tedy k dvěma různým přibližným řešením, kterými budou

$$\mathbf{z}_H^{(i+1)} = \mathcal{H}_k \left( A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)}) \right) \quad (5.34)$$

a  $\mathbf{z}_{\text{OMP}}^{(i+1)}$ , které získáme  $k$  iteracemi algoritmu OMP se slovníkem  $D$  a aproximovaným signálem  $(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)})$ .

**Poznámka 5.6.** Konkrétní tvar takto použitého OMP dostaneme z algoritmu 8, kde namísto slovníku  $D_R$  vystupuje celý slovník  $D$  a namísto  $\mathbf{y}_R$  dosadíme  $(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)})$ . Navíc maximálním počtem iterací v této situaci je číslo  $k$  a ukončovací kritérium na základě normy rezidua vůbec neuvažujeme.

Opodstatněnost volby (5.34) ukazují s využitím (2.10) a (2.11) následující úpravy [41]:

$$\left\| D\mathbf{z}^{(i+1)} - \mathbf{x}^{(i)} + \boldsymbol{\sigma}^{(i)} \right\|_2^2 = \left\| D\mathbf{z}^{(i+1)} - DA(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)}) \right\|_2^2 \quad (5.35a)$$

$$= \left\| D \left( \mathbf{z}^{(i+1)} - A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)}) \right) \right\|_2^2 \quad (5.35b)$$

$$\leq \left\| \mathbf{z}^{(i+1)} - A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)}) \right\|_2^2. \quad (5.35c)$$

Za podmínky  $\|\mathbf{z}^{(i+1)}\|_0 \leq k$  je podle Lemmatu 5.5 norma (5.35c) minimalizována právě pro  $\mathbf{z}^{(i+1)} = \mathbf{z}_H^{(i+1)}$  podle rovnice (5.34). Současně porovnáním (5.35a) a (5.35c) vidíme, že objektivní funkce v původním problému (5.33) bude nutně nabývat nejvýše stejné hodnoty. Předpokládáme-li tedy, že  $\mathbf{z}_H^{(i+1)}$  je dostatečně přesnou aproximací  $A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)})$  (což je předpoklad s rostoucím  $k$  relevantní), pak se aproximací (5.33) pomocí  $\mathbf{z}_H^{(i+1)}$  dopustíme přijatelně malé chyby.

Použití OMP je oproti tvrdému prahování výpočetně značně náročnější (využívá analýzu i syntézu v každé iteraci). Jednoduchá ilustrace na obrázku 5.1 však ukazuje, že by aproximace pomocí  $\mathbf{z}_{\text{OMP}}^{(i+1)}$  v neortogonálním případě měla být lepší než  $\mathbf{z}_H^{(i+1)}$ . V experimentech však ukážeme, že tato lepší aproximace obvykle nevede k lepšímu výsledku celkové rekonstrukce signálu a zvýšená výpočetní náročnost použitím OMP tudíž není opodstatněná.

Celý algoritmus S-SPAIN, opět již i s parametry  $r, s$  pro řízení řídkosti, je v algoritmu 10. Mezi variantou využívající tvrdé prahování (S-SPAIN H) a variantou s vnořeným OMP (S-SPAIN OMP) se rozhoduje ve druhém kroku algoritmu.



Obr. 5.1: Porovnání tvrdého prahování a OMP pro aproximaci  $\mathbf{s} = A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)})$  v prostoru  $\mathbb{R}^2$ . Pro ortogonální slovník  $D$  (vlevo) vidíme, že jsou obě aproximace totožné. Pro neortogonální (vpravo) je však aproximace pomocí OMP k vektoru  $\mathbf{s}$  blíže.

---

**Algoritmus 10:** S-SPAIN H / S-SPAIN OMP, dle volby aproximace v kroku 2

---

**input:** pozorovaný signál  $\mathbf{y}$ , operátory  $A, D$ , množina přípustných řešení  $\Gamma$ ,  
parametry  $s, r \in \mathbb{N}, \epsilon > 0$

- 1  $\mathbf{x}^{(0)} = A\mathbf{y}$ ,  $\boldsymbol{\sigma}^{(0)} = \mathbf{0}$ ,  $i = 0$ ,  $k = s$
- 2  $\mathbf{z}^{(i+1)} = \mathcal{H}_k(A(\mathbf{x}^{(i)} - \boldsymbol{\sigma}^{(i)}))$  nebo OMP // aproximace (5.30a)
- 3  $\mathbf{x}^{(i+1)} = \mathcal{P}_\Gamma(D\mathbf{z}^{(i+1)} + \boldsymbol{\sigma}^{(i)})$  // odpovídá (5.30b)
- 4 **if**  $\|D\mathbf{z}^{(i+1)} - \mathbf{x}^{(i+1)}\|_2 \leq \epsilon$  **then**
- 5 | terminate
- 6 **else**
- 7 |  $\boldsymbol{\sigma}^{(i+1)} = \boldsymbol{\sigma}^{(i)} + D\mathbf{z}^{(i+1)} - \mathbf{x}^{(i+1)}$  // odpovídá (5.30c)
- 8 |  $i \leftarrow i + 1$
- 9 | **if**  $i \bmod r = 0$  **then**
- 10 | |  $k \leftarrow k + s$
- 11 | **end**
- 12 | go to 2
- 13 **end**
- 14 **return**  $\hat{\mathbf{x}} = \mathbf{x}^{(i+1)}$

---

**Poznámka 5.7** (případ invertibilních operátorů). Uvažujeme situaci, kdy pracujeme s ne-redundantní transformací signálu, tedy v pozici  $A$  a  $D$  máme invertibilní operátory a platí  $A = D^{-1}$ . Formulace problému pro A-SPAIN a S-SPAIN (5.15) se liší pouze v podmínce svazující  $\mathbf{x}$  a  $\mathbf{z}$ ; v analyzující variantě  $A\mathbf{x} - \mathbf{z} = \mathbf{0}$ , v syntetizující  $\mathbf{x} - D\mathbf{z} = \mathbf{0}$ . Jednu podmínku snadno přeformulujeme ekvivalentně jako druhou aplikací  $D = A^{-1}$  (z analyzující varianty na syntetizující) nebo  $A = D^{-1}$  (ze syntetizující varianty na analyzující) na obě strany podmínky. Dostáváme tudíž ekvivalenci formulací úloh pro A-SPAIN a S-SPAIN, tudíž i výsledné algoritmy jsou díky jednotnému postupu odvození v tomto případě ekvivalentní.



## 6 Experimenty

### 6.1 Hodnocení kvality rekonstrukce

Protože budeme dále pracovat s a priori nepoškozenými signály, které budeme uměle degradovat, můžeme evaluaci výsledků založit na porovnání rekonstrukce s původním nepoškozeným signálem.

Jak je v současné literatuře (například [3]) obvyklé, bude první objektivní mírou rekonstrukce *signal-to-noise ratio*<sup>1</sup> (SNR), definované vzorcem

$$\text{SNR}(\mathbf{s}, \hat{\mathbf{x}}) = 20 \log_{10} \frac{\|\mathbf{s}\|_2}{\|\mathbf{s} - \hat{\mathbf{x}}\|_2} \quad [\text{dB}], \quad (6.1)$$

kde  $\mathbf{s}$  je původní nepoškozený signál a  $\hat{\mathbf{x}}$  je rekonstrukce degradovaného signálu. V naší konzistentní formulaci audio inpaintingu se mimo doplněný úsek signály  $\mathbf{s}$  a  $\hat{\mathbf{x}}$  rovnají, SNR proto počítáme pouze z částí odpovídajících chybějícím úsekům. Protože ve vzorci (6.1) vystupuje ve jmenovateli norma rozdílu originálního signálu a doplněného úseku, zřejmě vyšší hodnoty SNR budou odpovídat přesnější rekonstrukci. Zde se však skrývá i slabina SNR, a to že ne vždy odpovídá tomu, jak audiosignál vnímá posluchač – například budeme-li uvažovat dvě sinusoidy o stejné frekvenci a amplitudě posunutě vzájemně o polovinu periody, bude SNR velice nízké, přesto člověk sluchem nepozná rozdíl.

V závěru proto srovnáme vybrané metody rekonstrukce i pomocí PEMO-Q [19], což je pokročilejší nástroj pro hodnocení audiosignálů, který uvažuje lidské vnímání zvuku. Měřenými veličinami budou *perceptual similarity measure* (konkrétně uvažujeme 5. percentil hodnot počítaných v závislosti na čase,  $\text{PSM}_t$ ) a *objective difference grade* (ODG). Obě veličiny měří podobnost signálů, budou tedy opět aplikovány na  $\mathbf{s}$  a  $\hat{\mathbf{x}}$ , tentokrát však simulujeme lidského posluchače, tudíž pracujeme s celými signály, nikoliv pouze s rekonstruovanými úseky. Hodnoty  $\text{PSM}_t$  jsou v rozmezí od 0 do 1 (vyšší hodnota značí vyšší perceptuální podobnost signálů), hodnoty ODG jsou od  $-4$  do 0 a jejich význam udává tabulka 6.1.

Tab. 6.1: Význam hodnot ODG. Artefaktem zde v našem případě rozumíme slyšitelný pozůstatek díry v signálu, kterým může být pokles energie, krátký výpadek či „kliknutí“.

Hodnocení artefaktu	ODG
nepostřehnutelný (imperceptible)	0
slyšitelný, ale neruší (perceptible, but not annoying)	-1
příliš neruší (slightly annoying)	-2
rušivý (annoying)	-3
velmi rušivý (very annoying)	-4

### 6.2 Testovací vzorky

K testování popsaných algoritmů bylo vybráno 10 hudebních vzorků z veřejné databáze EBU SQAM [1]. Tato databáze obsahuje ukázky v bezztrátovém formátu určené k sub-

<sup>1</sup>Zde se nejedná doslova o poměr signál/šum, dle konvence v literatuře však nebudeme zavádět jiné označení.

jektivnímu hodnocení kvality audiosignálu. Pro tuto práci z ní bylo vybráno 10 reálných nahrávek, všechny signály mají vzorkovací frekvenci 44,1 kHz a mají průměrnou délku přibližně 7 s. Vzorky byly vybrány tak, aby reprezentovaly různé stupně řídkosti STFT koeficientů, což ilustrují spektrogramy na obrázku 6.1.

Dle značení v databázi se jedná o tyto vzorky:

- (a) a08\_violin,
- (b) a16\_clarinet,
- (c) a18\_bassoon,
- (d) a25\_harp,
- (e) a35\_glockenspiel,
- (f) a41\_celesta,
- (g) a42\_accordion,
- (h) a58\_guitar\_sarasate,
- (i) a60\_piano\_schubert,
- (j) a66\_wind\_ensemble\_stravinsky.

Budeme-li se v rozboru experimentů vyjadřovat zvláště k vybraným nahrávkám, pak budeme spíše využívat pojmenování podle hudebního nástroje, nikoliv tyto původní názvy signálů.

## 6.3 Schéma experimentu

Za účelem srovnání popsaných metod budeme pracovat s jednotným modelem poruchy, pro který byly jednotlivé postupy navrženy, tedy s kompaktním chybějícím úsekem většího rozsahu (řádově desítky milisekund). Protože však jediná díra nese riziko příliš velkého vlivu pozice v signálu na výsledek rekonstrukce, budeme pracovat u každého vzorku s 8 dírami na částečně náhodných pozicích<sup>2</sup>. Protože délka díry má značný vliv na výsledek z hlediska hodnocení rekonstrukce, jednotlivé délky díry v dané realizaci experimentu nebudeme kombinovat a zvolíme vždy jednu délku mezi 5 a 50 ms.

Délce chybějícího úseku musí odpovídat i dostatečně dlouhé okno transformace, jímž bude obvykle Hannovo okno délky 2800 vzorků (přibližně 64 ms) s posunem sousedních oken 700 vzorků (přibližně 16 ms, tedy překryv 75 %). Přitom jsme si vědomi, že délka okna má stěžejní vliv na úspěšnost rekonstrukce, v této práci se však experimentování s tímto parametrem vyhneme.

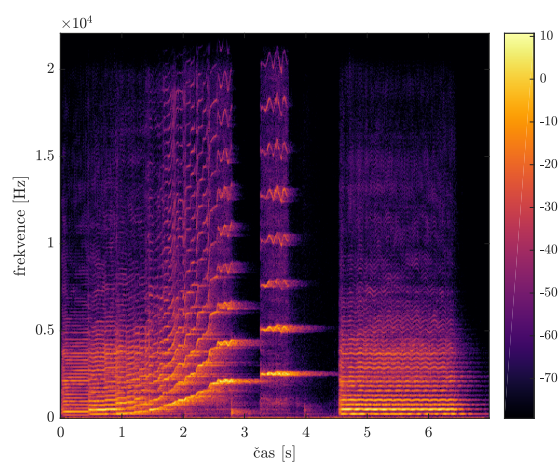
Konkrétní podoby použitých slovníků budeme komentovat u jednotlivých metod.

**Poznámka 6.1** (vizualizace výsledků). V následujících částech budeme srovnávat jednotlivé algoritmy převážně pomocí hodnot SNR, později též pomocí pokročilejších ODG a  $PSM_t$ . Hodnoty SNR v decibelech budou vždy průměrovány přes všech 10 signálů a 8 pozic díry (tj. průměrujeme již vypočítané hodnoty v dB) a budeme je zobrazovat v závislosti na délce chybějícího úseku.

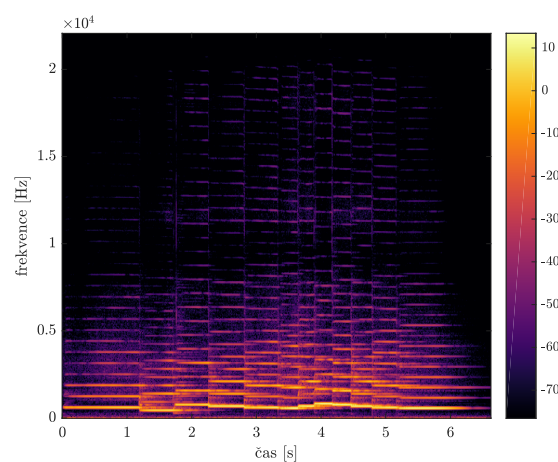
## 6.4 Konvexní metody

V této části budeme pracovat s analyzujícím a syntetizujícím modelem (váhované)  $\ell_1$  relaxace a jejich modifikacemi dle kapitoly 4. Pro analyzující model budeme výhradně využívat Chambollův-Pockův algoritmus 2, pro syntetizující model Douglasův-Rachfordův

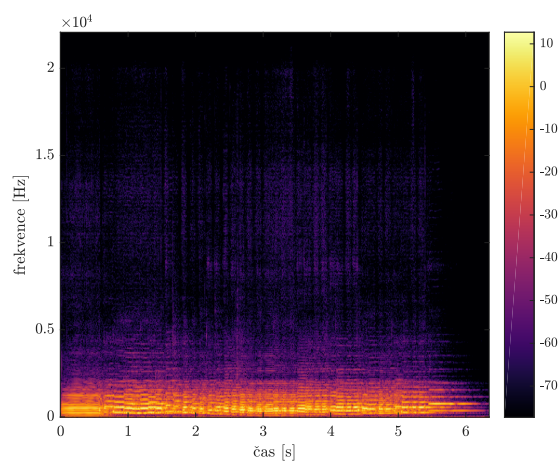
<sup>2</sup>Pozice jsou pouze částečně náhodné, neboť respektujeme minimální vzdálenost mezi dírami tak, aby rekonstrukce jedné neovlivňovala sousední, jak je popisováno v části 4.2.2.



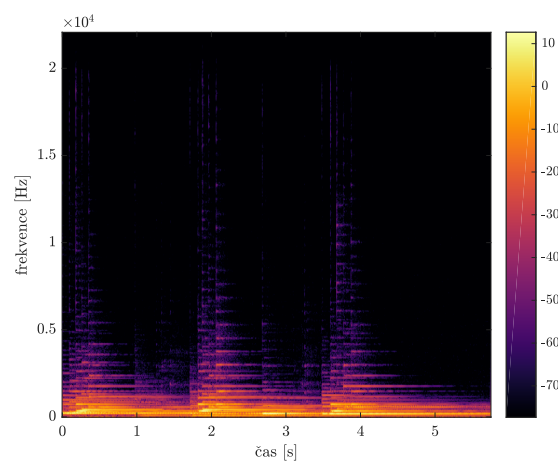
(a) a08\_violin



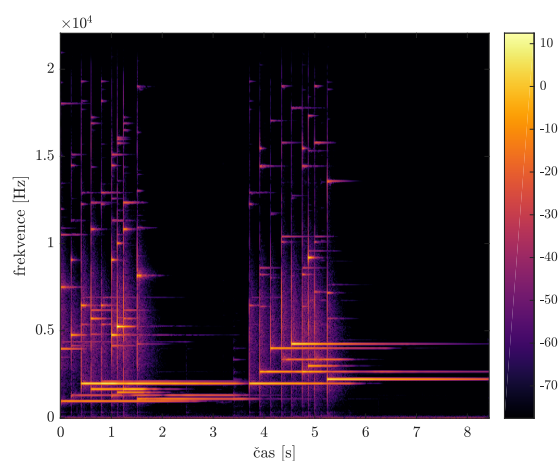
(b) a16\_clarinet



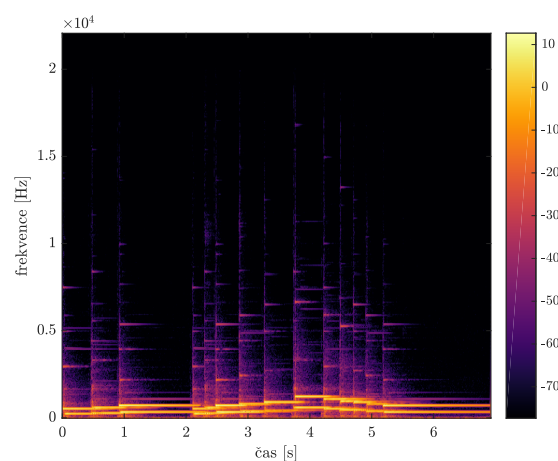
(c) a18\_bassoon



(d) a25\_harp

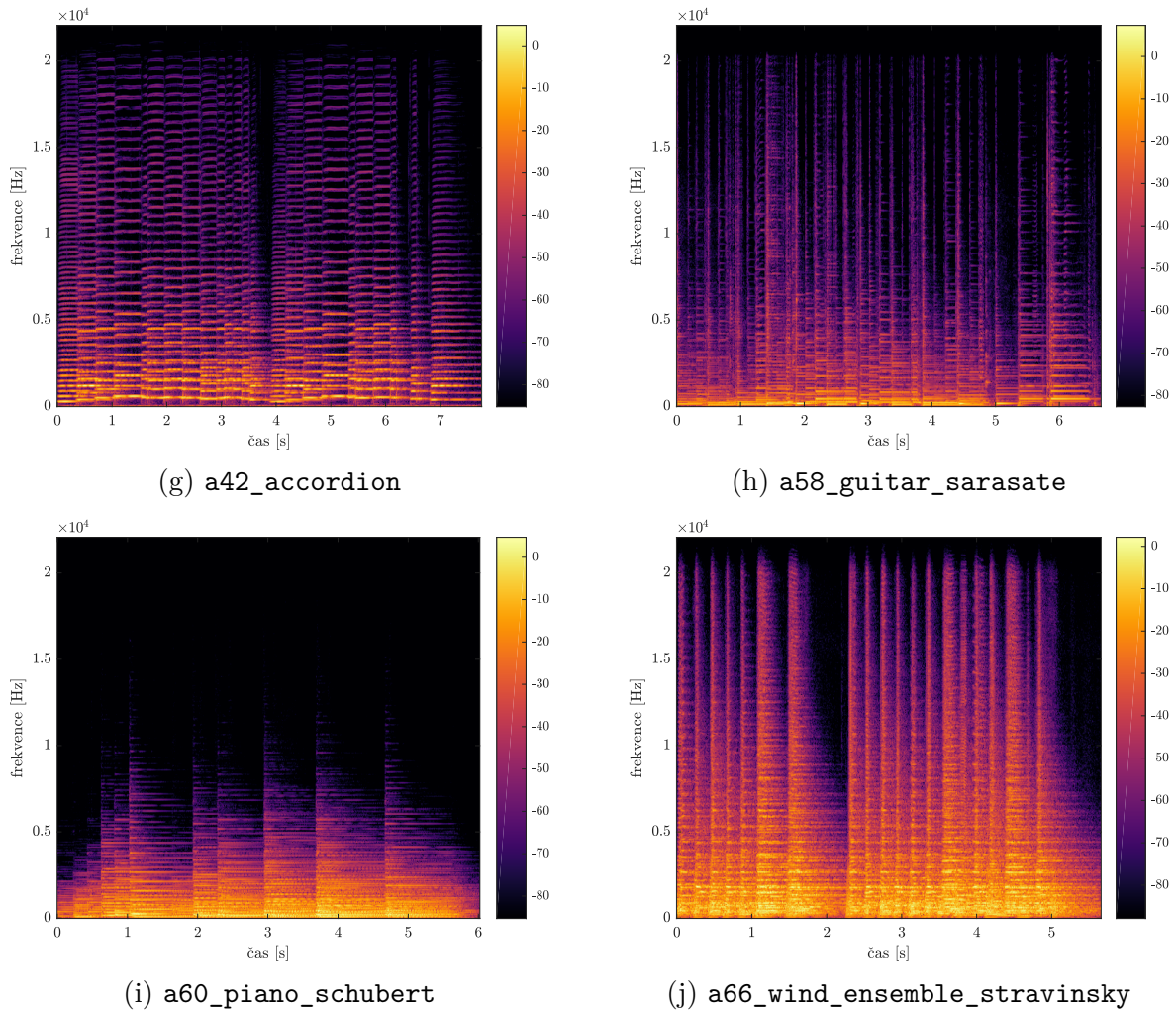


(e) a35\_glockenspiel



(f) a41\_celesta

Obr. 6.1: Testovací vzorky, první část.



Obr. 6.1: Testovací vzorky, druhá část.

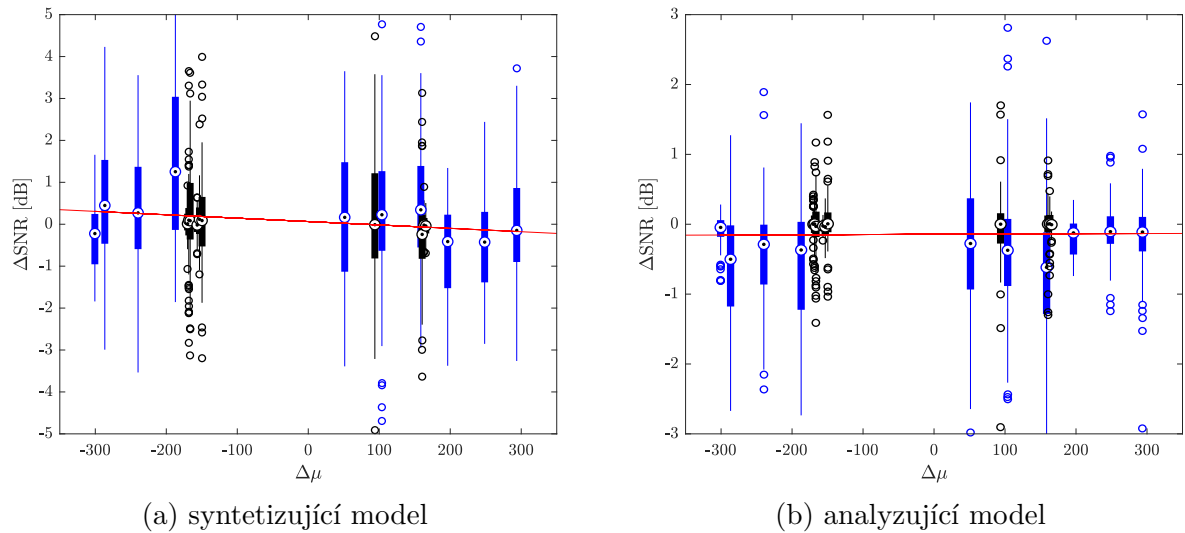
algoritmus 4. Jednotně také volíme ukončovací kritérium pro oba algoritmy – buď dosažení maximálního počtu 1000 iterací, nebo relativní změna normy průběžného řešení menší než  $10^{-4}$ . Toto odpovídá v nerovnicích (4.14) a (4.18) hodnotě  $\varepsilon = 10^{-4}$ .

### 6.4.1 Offset

Na rozdíl od pořadí metod v předchozích kapitolách začneme s výsledky experimentu pro offset (dle části 4.2.2). Motivací je to, že offset chceme používat i pro testování ostatních metod, abychom omezili náhodný vliv vzájemné pozice chybějícího úseku a počátečního bodu signálu pro STFT.

Konkrétně zkoumáme souvislost míry  $\mu$  dle vzorce (4.23) a SNR. Bezrozměrnou hodnotou  $\Delta\mu$  je pro nás rozdíl  $\mu$  pro offset varianty *full* a varianty *half* (v tomto pořadí). Obdobně zavádíme  $\Delta\text{SNR}$  [dB]. Navíc používáme dvě různé varianty transformace – v obou případech je to STFT s Hannovým oknem délky 2800 vzorků a překryv sousedních oken je buď 3/4 nebo 7/8 délky okna. Posledním parametrem je počet kmitočtových kanálů, který byl nastaven na 2800 (tedy odpovídá délce okna v počtu vzorků).

Výsledky experimentu ukazuje obrázek 6.2. Pro vizualizaci byl použit tzv. box plot, český někdy též krabicový diagram, k němuž se vyjadřuje poznámka 6.2.



Obr. 6.2: Experiment na offset. Vykresleno je rozložení  $\Delta\text{SNR}$  v závislosti na  $\Delta\mu$  pomocí box plotu (poznámka 6.2). Modré boxy odpovídají framu s překryvem sousedních oken 3/4, černé boxy jsou pro překryv 7/8. Červeně je navíc vykreslena lineární regrese pro mediány.

**Poznámka 6.2** (box plot). Box plot zobrazuje rozdělení tříděného statistického souboru. Pro každou třídu je vykreslen medián (jako bod), box zachycuje rozmezí 25. a 75. kvantilu. Svislá linie vyznačuje rozsah všech zbývajících hodnot, které nejsou považovány za odlehlé hodnoty, tzn. jejich vzdálenost od boxu není větší než 150% výšky boxu. Parametry, jako jsou právě hodnoty kvantilů či kritérium pro odlehlé hodnoty, přebíráme z výchozího nastavení funkce `boxplot` pro MATLAB (<https://www.mathworks.com/help/stats/box-plots.html>).

Z obrázku 6.2 vidíme, že pomocí míry dle vzorce (4.23) jsme nedosáhli přesvědčivého výsledku, který by nám poskytoval návod pro volbu varianty offsetu, neboť žádnou zřejmou závislost  $\Delta\text{SNR}$  na  $\Delta\mu$  nepozorujeme. Pouze v obrázku 6.2a pro syntetizující model je slabě znatelný klesající trend pro mediány hodnot  $\Delta\text{SNR}$ , který ukazuje na zápornou korelaci  $\Delta\text{SNR}$  a  $\Delta\mu$ . U analyzujícího modelu na obrázku 6.2b se však i tento mírný trend vytrácí.

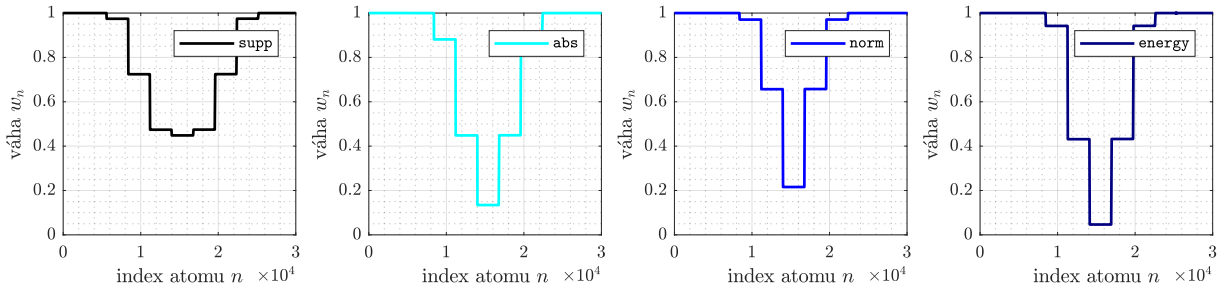
Závěrem tedy je, že jelikož výsledky s hodnotou  $\Delta\mu$  nejsou přesvědčivé, budeme v následujících experimentech používat jednotnou variantu offsetu (konkrétně volíme variantu *full*), čímž eliminujeme jeden parametr nastavení algoritmů.

## 6.4.2 Váhování atomů

Protože jsme neváhovanou variantu  $\ell_1$  relaxace řadili jako speciální variantu váhování (s konstantními váhami), nevěnujeme jí zvláštní část, ale srovnáme základní analyzující a syntetizující model zde.

Pro navržené varianty váhování atomů zavedeme pro přehlednost jednodušší značení, přičemž se držíme zkratk v příložených kódech pro MATLAB. Varianty dle rovnic (4.21) označíme následovně:

- (a) **none**: základní varianta bez váhování (4.21a),
- (b) **supp**: dle nosiče (4.21b),



Obr. 6.3: Ilustrace variant váhování atomů. Zobrazeny jsou všechny návrhy kromě neváhované varianty, jíž by odpovídala konstantní hodnota 1. Atomy odpovídají Gaborovu framu dle konvenčního nastavení, délka díry byla 35 ms.

- (c) **abs**: dle  $\ell_1$  normy (4.21c),
- (d) **norm**: dle  $\ell_2$  normy (4.21d),
- (e) **energy**: dle energie (4.21e).

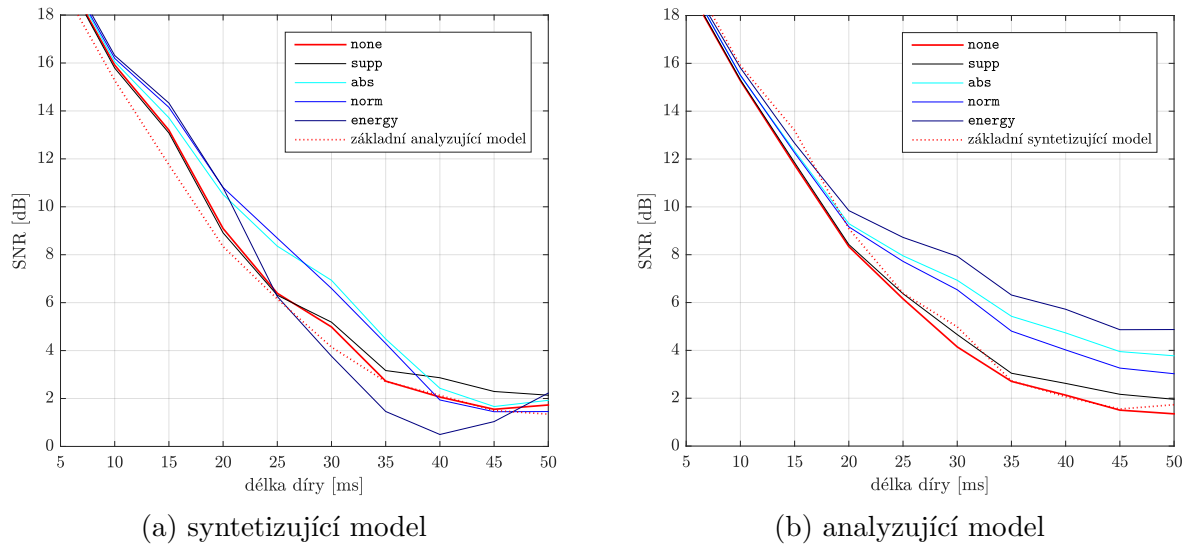
Jako časově-frekvenční transformaci používáme v tomto experimentu STFT s parametry dle konvence zavedené v části 6.3, přičemž počet kmitočtových kanálů je jako v experimentu na offset roven délce okna, tedy 2800.

Pro toto nastavení a délku díry 35 ms jsou nejprve v obrázku 6.3 zobrazeny jednotlivé složky vektoru vah  $\mathbf{w}$  pro navržené varianty váhování. Ukazuje se zde několik významných jevů:

- Průběhy vah jsou po částech konstantní, což koresponduje s poznámkou 4.4, která se vyjadřovala k tomu, že váhy jsou závislé pouze na modulu prvků jednotlivých atomů (nikoliv na fázi), a dále je to důsledkem pořadí atomů (nejdříve řadíme podle translace, poté podle modulace).
- Pozorujeme také vliv offsetu, neboť průběhy vah jsou symetrické, což potvrzuje symetrii oken vzhledem ke středu díry.
- Zřejmě se od sebe hodnotami vah liší znatelně jednotlivé přístupy, což opodstatňuje jejich testování.

Výsledky rekonstrukcí z hlediska SNR zobrazuje obrázek 6.4. Kromě váhovaných variant pro syntetizující (obrázek 6.4a) a analyzující (obrázek 6.4b) model je v obou grafech pro srovnání znázorněn i výsledek pro druhý model bez váhování.

V obrázku 6.4 v první řadě vidíme, že základní syntetizující model  $\ell_1$  relaxace dosahuje z hlediska SNR mírně lepších výsledků než model analyzující. Zcela odlišná je však situace s váhováním atomů. U analyzujícího modelu na obrázku 6.4b dosahujeme pomocí váhování soustavného zvýšení SNR, přičemž nejvyšší zlepšení nastává při váhování podle energie. Navíc pokud bychom dle obrázku 6.3 seřadili varianty váhování podle toho, jaké rozpětí vah poskytují, dostaneme pro výsledky v obrázku 6.4b odpovídající pořadí hodnot SNR. Jiná situace je však u syntetizujícího modelu na obrázku 6.4a. Mezi ostatními výsledky je pozoruhodné zejména váhování podle energie, které v rozmezí délek díry od 30 do 45 ms dokonce snižuje SNR rekonstrukce oproti neváhované variantě. Ostatní metody poskytují spíše mírné, nicméně konzistentní zlepšení. Pro díry od délky 30 ms však z hlediska SNR jednoznačně lepší výsledky poskytuje váhovaný analyzující model.



Obr. 6.4: Experiment na váhování atomů. Kromě variant váhování je vždy pro srovnání vykreslen i výsledek pro neváhovaný alternativní model.

### 6.4.3 Váhování doplněného úseku v čase

Před větším testováním této modifikace se stručně vyjádříme k parametrům algoritmu dle části 4.3.1, které jsou následující (v závorce je vždy zkratkové označení odpovídající implementaci v MATLABu):

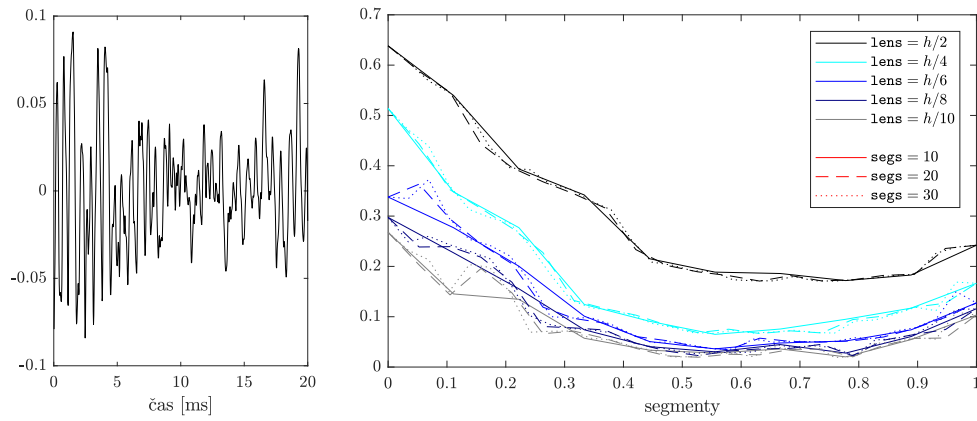
- délka vektoru **m**, tzn. počet segmentů, z nichž počítáme energii (**segs**),
- délka jednoho segmentu (**lens**),
- počet navíc doplněných děr (**gaps**),
- posun mezi sousedními navíc vytvořenými dírami (**shift**).

Počet děr navíc přitom volíme vždy sudý, tudíž vlevo a vpravo od původní díry jich je stejný počet. Pevný volíme parametr **shift** =  $w/2$ , přičemž však k původní díře nejbližší doplňkové díry jsou odsazeny o  $w$  vzorků. Na modelové situaci proto zbývá analyzovat vztah mezi parametry **segs** a **lens**, počet doplňkových děr bude poté součástí experimentu na celé použité databázi signálů.

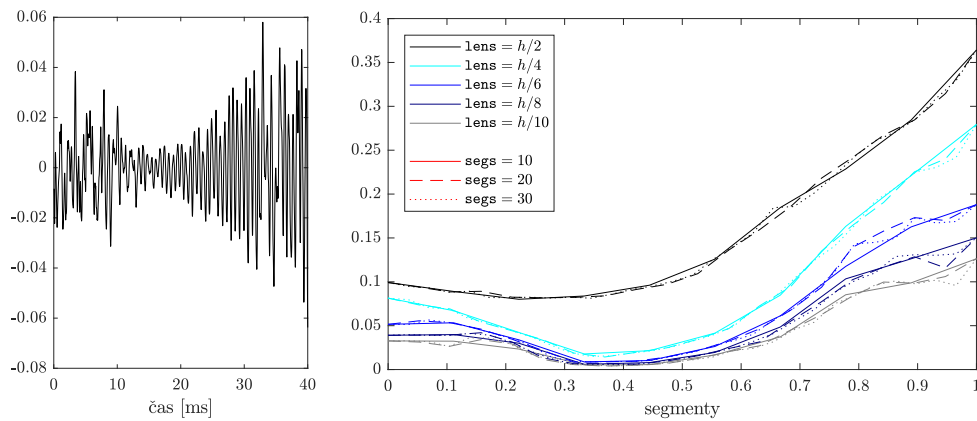
Z obrázku 6.5, ve kterém zobrazujeme průběh energie pro jednu doplněnou díru, vidíme, že jak zvyšování počtu segmentů (**segs**), tak snižování jejich délky (**lens**) má za důsledek „méně hladký“ průběh energie. Při výsledné kombinaci takovýchto průběhů z více děr by vzniklé výkyvy mohly mít nežádoucí vliv. Dále vidíme, že všechny křivky energie přibližně odpovídají představě o vývoji energie, kterou máme ze zobrazených úseků v časové oblasti, a především že použití délky segmentu relativní vůči délce díry nepřináší pro různě dlouhé díry rozdílné výsledky. Pro následující experiment je tudíž opodstatněná jednotná volba, kterou bude **segs** = 10 (nejvíce vyhlazený průběh) a kompromisní **lens** =  $h/4$ , kde  $h$  značí jako obvykle délku chybějícího úseku (v obrázku 6.5b nejlépe zachycuje pokles energie kolem středu úseku).

Výsledky experimentu vidíme na obrázku 6.6. Testujeme zde dva různé scénáře:

- váhování v čase vnímáme jako alternativu k váhování atomů (grafy 6.6a a 6.6c),
- váhování atomů a signálu v čase kombinujeme (grafy 6.6b a 6.6d).



(a) délka díry 20 ms



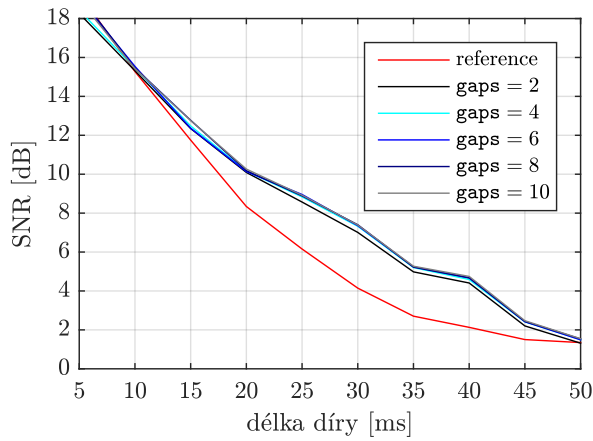
(b) délka díry 40 ms

Obr. 6.5: Vliv délky a počtu segmentů na křivku energie signálu. Pro dvě zvolené délky díry  $h$  je vždy zobrazen rekonstruovaný úsek signálu (používáme základní syntetizující model) a následně průběhy energie získané různou volbou parametrů **segs** a **lens**. Legenda vždy popisuje zvláště barvu a typ čáry. energii segmentů vykreslujeme pro lepší srovnání jednotně na intervalu  $\langle 0, 1 \rangle$ .

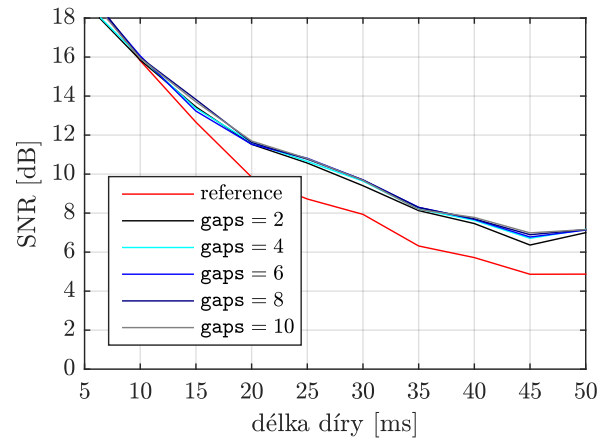
Vidíme, že mimo syntetizujícího modelu s váhováním atomů dosahujeme váhováním v čase zlepšení SNR v řádu jednotek dB. Výsledek je pozoruhodný zejména v případě analyzujícího modelu – z obrázku 6.6a sice vidíme, že váhování v čase nedosahuje přesvědčivě lepších výsledků než váhování atomů (srovnajme s referenčními hodnotami v obrázku 6.6b), v případě váhovaného analyzujícího modelu však váhování v čase přináší další významné zlepšení. Jednotným výsledkem ve všech modelech pak je, že počet doplňkových děr v rámci testovaného rozsahu není pro výsledek stěžejní, ačkoliv vyšší počet děr výsledné SNR mírně zvyšuje.

#### 6.4.4 Gradual inpainting

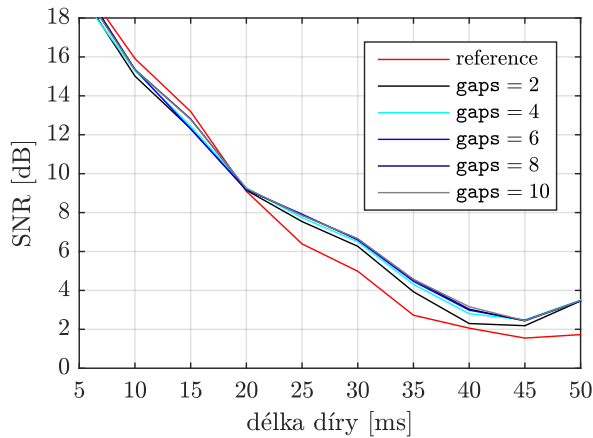
V testování gradual inpaintingu vycházíme z již provedeného experimentu na váhování atomů. Jak bylo dokázáno v části 4.3.2, základní  $\ell_1$  minimalizace ve spojení s přístupem po stupních nemůže přinést žádné zlepšení rekonstrukce, kombinujeme proto tento přístup právě s váhováním atomů. Protože cílem přístupu po stupních je výsledky s váhovanou  $\ell_1$  minimalizací dále vylepšit, volíme přirozeně způsob váhování, se kterým jsme v předchozím experimentu dosáhli nejlepších výsledků. Pro syntetizující model proto volíme



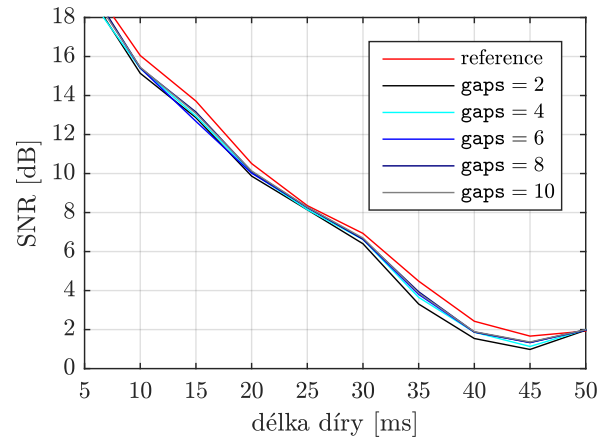
(a) analyzující model (**none**)



(b) analyzující model (**energy**)



(c) syntetizující model (**none**)



(d) syntetizující model (**abs**)

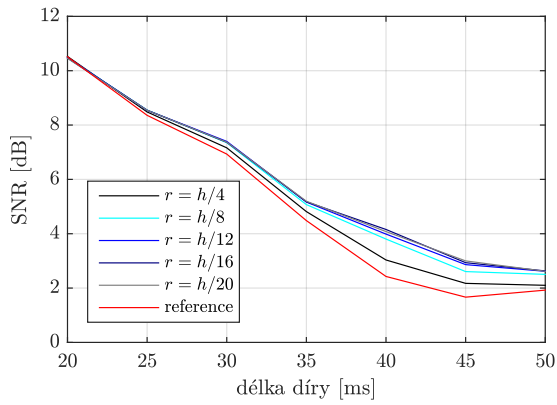
Obr. 6.6: Experiment na váhování doplněného úseku v čase. Výsledky jsou vykresleny pro různé volby parametru **gaps**, jako reference je vždy použit odpovídající model se stejným váhováním atomů, ovšem bez následného váhování v čase. Váhování atomů pro grafy 6.6b a 6.6d je voleno dle úspěšnosti v experimentu výše (obrázek 6.4).

variantu **abs** (ačkoliv dle obrázku 6.4a není volba jednoznačná), pro analyzující model variantu **energy**.

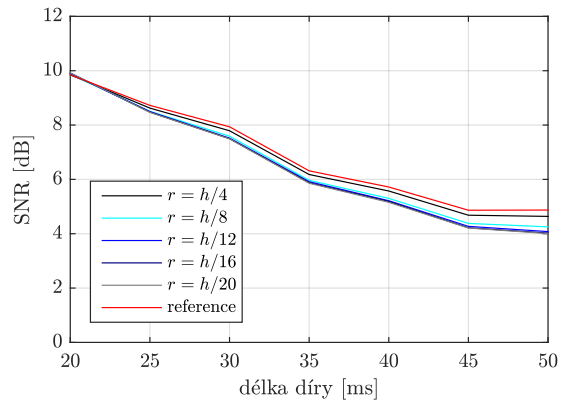
Dle algoritmu 7, který zde použijeme, má gradual inpainting jeden stěžejní parametr, kterým je počet vzorků  $r$ , o který v každém stupni zkracujeme díru zleva a zprava. Hodnoty  $r$  volíme jako podíly délky díry  $h$  (ve vzorcích). Díky tomu, že poměr  $r$  a  $h$  volíme sudý, je v každém stupni včetně posledního doplněno  $2r$  vzorků.

Výsledky experimentu ukazuje obrázek 6.7, ovšem pro lepší čitelnost pouze v redukované podobě – pro díry do 20 ms není znatelný rozdíl hodnot SNR, výsledky pro tyto kratší díry proto vůbec nezobrazujeme.

Opět se výrazně liší dopad metody po stupních na analyzující a syntetizující model. V obrázku 6.7b vidíme, že gradual inpainting v analyzujícím modelu nepřináší žádné další zlepšení oproti jednoduchému, dle energie váhovanému modelu, přičemž snižování hodnoty  $r$  způsobuje i snižování SNR rekonstrukce. Opačný trend vidíme na obrázku 6.7a pro syntetizující model. Postupný přístup zde rekonstrukci z hlediska SNR mírně, leč stabilně vylepšuje. Vzájemným srovnáním grafů 6.7a a 6.7b však vidíme, že ani s tímto pokrokem syntetizující model nedosahuje hodnot SNR analyzujícího modelu.



(a) syntetizující model (**abs**)



(b) analyzující model (**energy**)

Obr. 6.7: Experiment na gradual inpainting. Pro lepší čitelnost vykreslujeme hodnoty pouze pro chybějící úseky od délky 20 ms. Kromě variant s různou hodnotou parametru  $r$  je vykreslena jako reference i jednostupňová varianta s odpovídajícím váhováním atomů. Váhování atomů je voleno dle úspěšnosti v experimentu výše a v obrázcích 6.7a a 6.7b je vyznačeno v závorce.

## 6.5 Nekonvexní metody

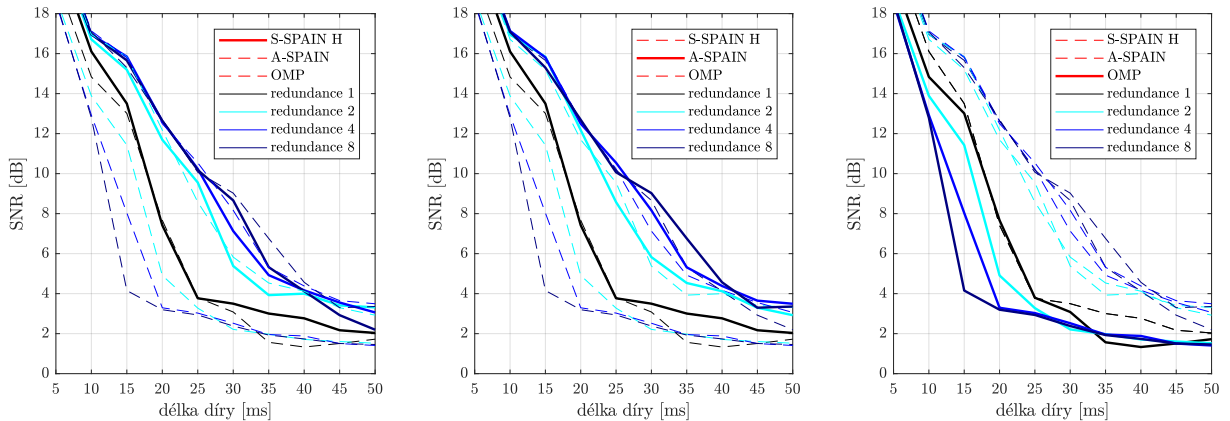
Na rozdíl od předchozích metod, kde jsme pomocí STFT počítali koeficienty celého signálu, u prezentovaných postupů vycházejících z nekonvexní formulace preferujeme jiný postup. Konkrétně pomocí zvoleného okna provádíme segmentaci signálu, jednotlivé segmenty rekonstruujeme zvoleným algoritmem a následně opravené segmenty opět složíme (inspirujeme se použitím tohoto postupu v [3] pro OMP).

Souvislost s STFT použitou při  $\ell_1$  relaxaci bude ta, že pro jednotlivé segmenty použijeme jako analyzující operátor diskrétní Fourierovu transformaci (DFT). Protože však chceme porovnat algoritmy OMP a SPAIN i pro redundantní slovníky, budeme navíc pracovat s tzv. redundantní DFT, kterou získáme analýzou signálu doplněného nulami na násobek délky (dostaneme tedy také odpovídající násobně větší počet koeficientů). Obdobné použití redundantní DFT lze najít v [38] pro audio declipping a příbuzný algoritmus SPADE, který byl vzorem pro SPAIN.

**Poznámka 6.3** (DFT pro reálný signál). Pokud aplikujeme DFT (i redundantní) na reálný signál, dostaneme koeficienty se specifickou strukturou, kdy ke každému koeficientu existuje komplexně sdružený protějšek. Způsobeno je to tím, že atomy pro frame, k němuž je DFT analyzujícím operátorem, jsou komplexní a také jsou po párech komplexně sdružené<sup>3</sup>. K syntéze reálného signálu tedy vždy potřebujeme oba sdružené atomy, a to se sdruženými koeficienty. K tomuto faktu přihlížíme v algoritmech OMP i SPAIN, neboť vždy pracujeme s koeficienty reálného signálu. Při OMP tedy vždy při výběru jednoho atomu automaticky vybereme i jemu komplexně sdružený. Stejně tak při tvrdém prahování v SPAIN nevybíráme  $k$  koeficientů, jak bylo obecně popsáno v části 5.2 a jak by tomu bylo buď u komplexního signálu, nebo u reálné transformace reálného signálu, ale vybíráme  $k$  komplexně sdružených párů.

<sup>3</sup>Vyjma tzv. stejnosměrné složky (a při sudé délce signálu  $N$  také složky  $N/2$ ), což je však na tomto místě nepodstatným detailem.

Pro algoritmy SPAIN a OMP bylo dále nastaveno obdobné ukončovací kritérium s  $\varepsilon = 1/10$ . Zbývající parametry algoritmu SPAIN byly nastaveny jako  $r = s = 1$ , tedy po každé iteraci ADMM zvyšujeme počet atomů o jedna.



Obr. 6.8: Experiment na nekonvexní metody. Všechny tři grafy zachycují stejné výsledky – průměrné hodnoty SNR pro tři algoritmy (S-SPAIN H, A-SPAIN, OMP) a čtyři různé redundance DFT (1, 2, 4, 8). Pouze v každém grafu je zvýrazněn jeden z algoritmů.

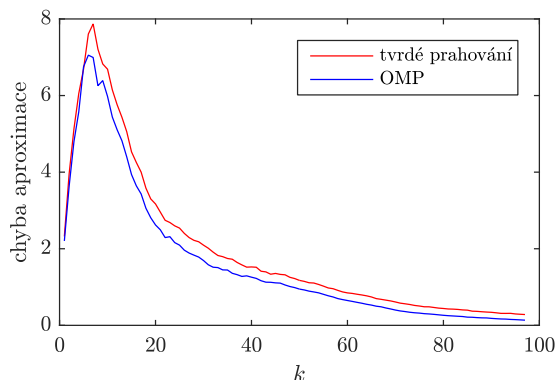
Srovnání SPAIN a OMP z hlediska SNR ukazují grafy na obrázku 6.8. Především vidíme, že nový algoritmus SPAIN téměř ve všech případech převyšuje výsledky OMP, výjimkou jsou délky díry v rozmezí 15 až 25 ms a DFT bez redundance. Z hlediska redundance také vidíme, že S-SPAIN H a A-SPAIN pro neredundantní transformaci dosahují identických výsledků, což potvrzuje teoretickou ekvivalenci obou dvou algoritmů v tomto případě (poznámka 5.7). Při zvyšování redundance zřejmě u algoritmu SPAIN dochází ke zlepšování (na rozdíl od OMP), například mezi redundancí 4 a 8 již však nejsou příliš výrazné rozdíly, zvláště s uvážením, že významně narůstá výpočetní náročnost.

Z důvodu extrémní výpočetní náročnosti se také samostatně věnujeme algoritmu S-SPAIN OMP<sup>4</sup>, a to pouze pro DFT s redundancí 2 (vyšší redundance náročnost dále zvyšuje). Navíc porovnání S-SPAIN H a S-SPAIN OMP provádíme na zmenšené úloze, kdy délky díry, okna a posun sousedních oken jsou vzhledem k předchozímu experimentu čtvrtinové.

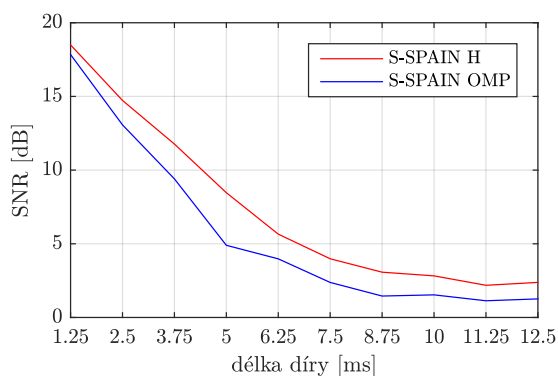
Nejdříve v grafu 6.9a ilustrujeme, že předpoklad lepší řídké aproximace pomocí OMP oproti tvrdému prahování je správný. Poznamenejme, že vykreslovaná hodnota  $\|D\mathbf{z}^{(k)} - \mathbf{x}^{(k-1)} + \boldsymbol{\sigma}^{(k-1)}\|_2$ , kde  $\mathbf{z}^{(k)}$  je získané pomocí tvrdého prahování (dle rovnice (5.34)) nebo pomocí OMP (poznámka 5.6), pro různá  $k$  není zcela srovnatelná. V jednotlivých iteracích S-SPAIN se totiž mění signál  $(\mathbf{x}^{(k-1)} - \boldsymbol{\sigma}^{(k-1)})$ , jehož řídkou aproximaci hledáme. Významné pozorování však je, že pro všechna  $k$  dosahujeme s pomocí OMP mírně lepší aproximace (hodnota chyby v grafu 6.9a je mírně menší).

Graf na obrázku 6.9b však ukazuje překvapivý výsledek, že S-SPAIN H dosahuje nezávadně lepších hodnot SNR než S-SPAIN OMP. Zřejmě tedy zlepšení v dílčím kroku (5.30a), které algoritmus OMP poskytuje oproti tvrdému prahování za cenu řádově vyššího výpočetního času celého algoritmu, není pro celkový výsledek významné.

<sup>4</sup>Výpočetní náročnost je způsobena iteračním algoritmem vloženým do iteračního algoritmu. Například v průběhu prvních 100 iterací S-SPAIN OMP při nastavení parametrů  $r = s = 1$  musí celkem proběhnout 5050 iterací OMP.



(a) porovnání z hlediska chyby aproximace v průběhu iterací



(b) porovnání z hlediska SNR v závislosti na délce díry

Obr. 6.9: Porovnání S-SPAIN H a S-SPAIN OMP. Chybou aproximace v levém obrázku rozumíme rozdíl mezi vektorem a jeho hledanou řídkou aproximací v úloze (5.30a), tedy normu  $\|D\mathbf{z}^{(k)} - \mathbf{x}^{(k-1)} + \boldsymbol{\sigma}^{(k-1)}\|_2$ , kde  $\mathbf{z}^{(k)}$  hledáme jako argument minima této normy. Hodnoty jsou získány z rekonstrukce první díry délky 7,5 ms v signálu houslí. V pravém obrázku vidíme srovnání z hlediska SNR.

## 6.6 Celkové porovnání

Dosud jsme srovnávali pouze varianty jednotlivých přístupů k audio inpaintingu, vzájemné porovnání následuje nyní. Metody zhodnotíme mimo SNR také pomocí indikátorů motivovaných lidským sluchem, což jsou míry ODG a  $PSM_t$  zavedené v úvodu kapitoly. Protože srovnání všech popsaných metod a variant by již bylo nadbytečné, vybíráme pouze následující přístupy:

- syntetizující a analyzující model neváhované  $\ell_1$  minimalizace,
- syntetizující a analyzující model  $\ell_1$  minimalizace s vybraným váhováním atomů,
- analyzující model  $\ell_1$  minimalizace s vybraným váhováním atomů a následným váhováním v časové oblasti,
- S-SPAIN H,
- A-SPAIN,
- Janssenův algoritmus.

Nastavení algoritmů založených na řídkosti je konvenční, pouze bylo třeba zvolit váhování atomů (**abs** pro syntetizující model, **energy** pro analyzující), parametry váhování v časové oblasti (**segs** = 10, **lens** =  $h/4$ , **gaps** = 4) a slovník pro SPAIN (redundance 4).

Zvláště se musíme věnovat parametrům Janssenova algoritmu, který dosud nebyl v našich experimentech použit. Pro ten jsme použili implementaci přiloženou k článku [3], v němž je algoritmus popsaný v kapitole 3 aplikovaný na segmentovaný signál, stejně jako náš SPAIN či OMP. Parametry okna tudíž volíme stejné jako u těchto nekonvexních metod. Navíc je třeba zvolit řád  $p$  autoregresního modelu, který je v převzaté implementaci ve výchozím nastavení

$$p = \min(3H + 2, w/3), \quad (6.2)$$

kde  $H$  je počet vzorků chybějících v daném okně a  $w$  je délka okna (hodnotu  $w/3$  přirozeně bereme zaokrouhlenou). Posledním parametrem je počet iterací cyklu, v jehož průběhu zpřesňujeme odhady autoregresních koeficientů; zde volíme jednotně 50 iterací.

Výsledky experimentu ukazují tři grafy na obrázku 6.11, které postupně rozebereme.

## Porovnání z hlediska SNR

Hodnoty SNR obvyklým způsobem ilustruje obrázek 6.11a. Mimo výsledků Janssenova algoritmu se nejedná o nové výsledky, pouze jsou dány do souvislosti hodnoty z předchozích experimentů.

Pro díry kratší než 25 ms vidíme, že A-SPAIN dosáhl lepších výsledků než ostatní metody založené na řídkosti a přiblížil se Janssenovu algoritmu. Od délky 30 ms však z pohledu SNR dosahuje jednoznačného prvenství analyzující model  $\ell_1$  minimalizace s váhováním atomů a následným váhováním v čase. Tento výsledek potvrzuje úspěšnost metod pro kompenzaci poklesu energie, neboť zřejmě pro kratší díry, kde pokles energie není významný, nedosahuje  $\ell_1$  minimalizace výsledků obou variant algoritmu SPAIN.

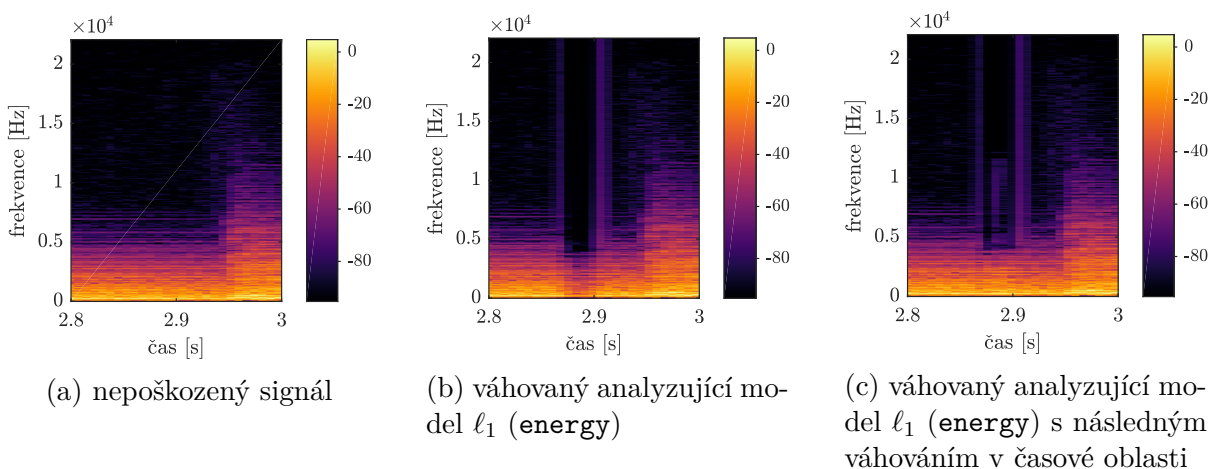
## Porovnání z hlediska ODG

Z obrázku 6.11b vidíme, že hodnocení pomocí SNR nebylo ani z hlediska vnímané kvality zvuku scestné, neboť výsledky si převážně odpovídají. Zásadní je však pohled na modifikované varianty  $\ell_1$  minimalizace. Rozevívá se zde totiž rozdíl mezi syntetizujícím a analyzujícím modelem, přičemž například váhovaný syntetizující model se pro kratší díry přiblížil výsledkům SPAIN. Také model s dodaným váhováním v čase již z hlediska ODG nedosahuje pro delší díry natolik pozoruhodného výsledku, jako je tomu z hlediska SNR na obrázku 6.11a.

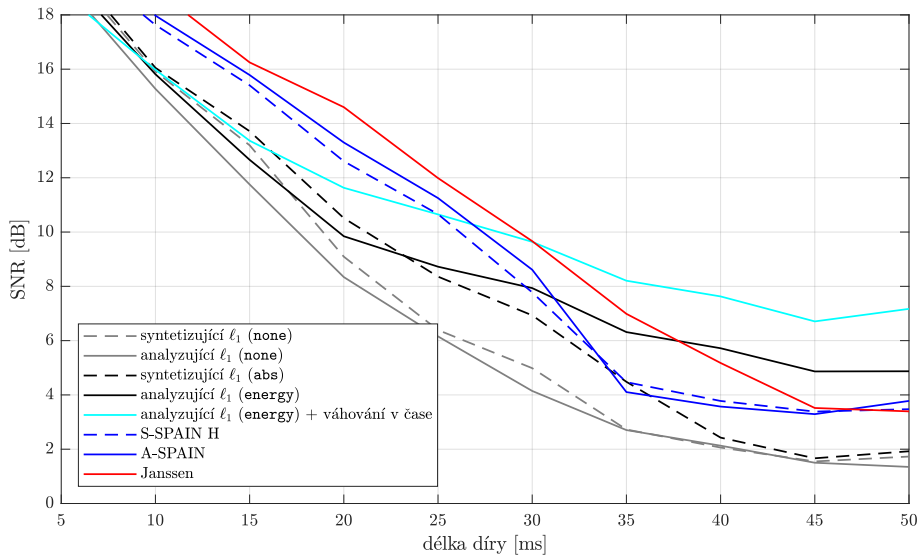
Přirozeným vysvětlením je, že váhování v časové oblasti sice zvyšuje energii doplněného úseku, ovšem způsobuje také přidání množství frekvencí, které se v původním signálu nevyskytovaly – tento jev ukazuje obrázek 6.10. Ve srovnání s nepoškozeným signálem (spektrogram 6.10a) obsahují spektrogramy rekonstrukce pomocí  $\ell_1$  minimalizace vertikální skupiny koeficientů, které zachycují nedostatečnou návaznost doplněného úseku na nepoškozené okolí. Významný je však spektrogram 6.10c, v němž doplněná díra obsahuje ve svém středu ještě třetí skupinu nenulových koeficientů navíc, což je právě vliv násobení vzorků v časové oblasti. Důsledkem ilustrované situace pak mohou být slyšitelné artefakty, které hodnocení z hlediska ODG zachytí.

## Porovnání z hlediska $PSM_t$

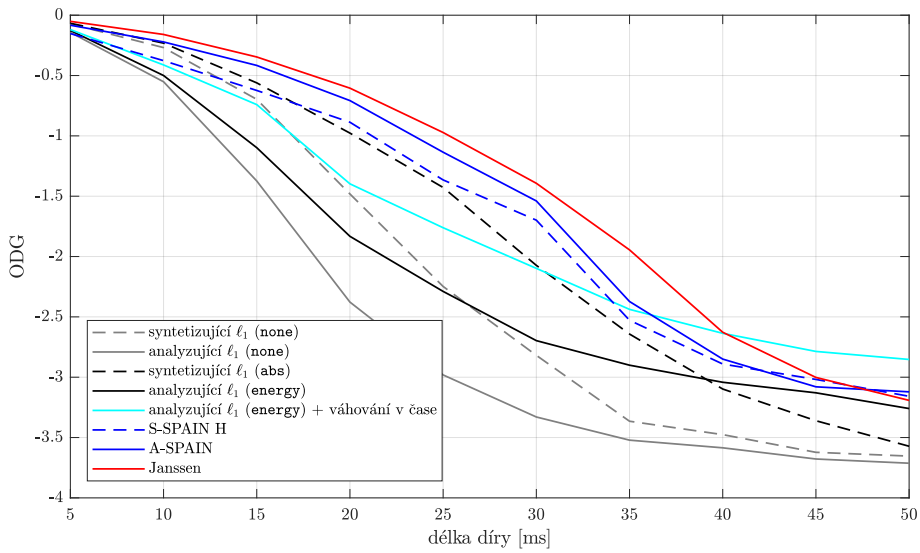
Míru  $PSM_t$  v obrázku 6.11c již ukazujeme jenom pro úplnost, i přes odlišný způsob výpočtu totiž poskytuje výsledky téměř shodné s hodnocením z hlediska ODG.



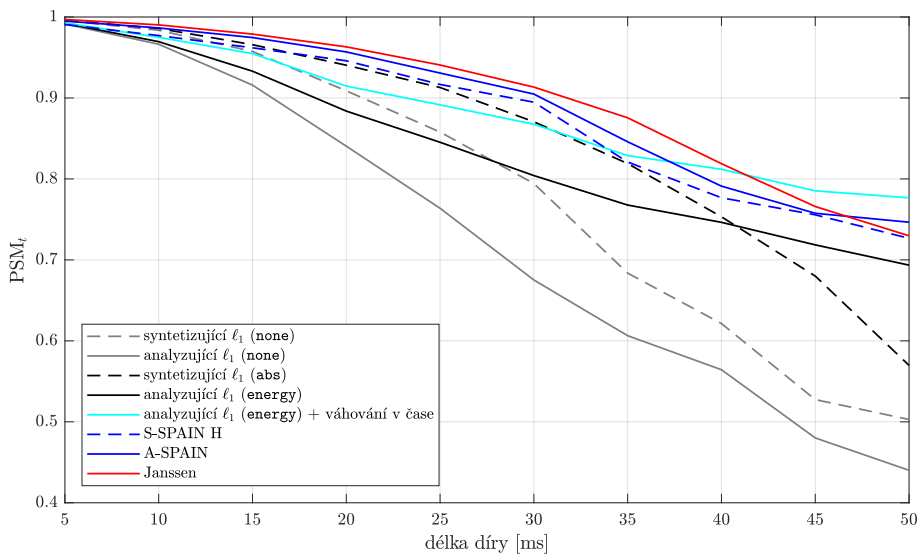
Obr. 6.10: Ukázka vlivu váhování v časové oblasti. Zobrazena je doplněná díra v záznamu pianu s blízkým okolím.



(a) porovnání z hlediska SNR



(b) porovnání z hlediska ODG



(c) porovnání z hlediska  $PSM_t$

Obr. 6.11: Celkové porovnání metod.

## 7 Závěr

Cílem této diplomové práce bylo prostudovat metody pro doplňování souvislého chybějícího úseku vzorků v digitálním audiosignálu, tedy problematiku tzv. audio inpaintingu. Zvláštní pozornost byla věnována moderním metodám založeným na řídkosti reprezentace audiosignálu, a to zejména při reprezentaci využívající krátkodobou Fourierovu transformaci. Právě předpoklad řídkosti posloužil jako základ pro formulaci audio inpaintingu jako optimalizační úlohy. Formulací bylo sestaveno několik, základní rozdělení bylo na formulaci konvexní (kapitola 4) a nekonvexní (kapitola 5).

Pro konvexní formulaci byl kromě obvyklého syntetizujícího modelu sestaven i model analyzující, který byl pro audio inpainting studován dosud pouze v článku [25], ovšem pro lehce odlišný tvar optimalizační úlohy. V kapitole 4 bylo navíc navrženo několik modifikací metody za účelem kompenzace poklesu energie, jenž se při aplikaci konvexní relaxace zákonitě projevuje v doplněném úseku signálu. Pod úpravu modelu byly zařazeny dva návrhy týkající se transformačního operátoru, a to váhování atomů, které bylo studováno již v bakalářské práci [27] a zde bylo nově zavedeno pro analyzující model, a tzv. offset, jehož účelem bylo dosáhnout symetrického poklesu energie vzhledem ke středu doplněné díry. Offset byl dále využit pro modifikace v časové oblasti, konkrétně váhování doplněného úseku v čase (opět v návaznosti na [27]) a tzv. gradual inpainting, tedy doplňování díry po částech směrem od okrajů. Pro oba tyto postupy byl využit právě předpoklad symetrie v průběhu energie, jehož splnění offset zajistil.

V kapitole 5 bylo popsáno využití algoritmu Orthogonal Matching Pursuit (OMP), což byl vůbec první přístup pro audio inpainting založený na řídkosti, prezentovaný v článku [3]. Především však byl odvozen nový heuristický algoritmus Sparse Audio Inpainter (SPAIN), vycházející z příbuzného algoritmu SPADE pro audio declipping. Přínosem této práce bylo v této části zejména detailní odvození analyzující i syntetizující varianty algoritmu (A-SPAIN, resp. S-SPAIN).

Experimentální část práce (kapitola 6) byla rozdělena na tři hlavní celky – analýza navržených modifikací konvexních metod, srovnání nekonvexních metod a následně celkové porovnání vybraných reprezentantů obou tříd se state-of-the-art algoritmem pro audio inpainting podle Janssenova [20], založeným na autoregresním modelování. Funkčnost algoritmů byla testována na sadě 10 reálných hudebních nahrávek a pro různé délky chybějícího úseku. Hodnocení rekonstrukce bylo prováděno mimo obvyklého odstupu signálu od šumu (signal-to-noise ratio, SNR) také pomocí algoritmu PEMO-Q, který rovněž poskytuje objektivní porovnání rekonstruovaného signálu se signálem nepoškozeným, na rozdíl od SNR však zahrnuje model lidského sluchu.

Významným výsledkem části 6.4 věnující se konvexním metodám bylo ukázání značného vlivu váhování atomů i váhování v časové oblasti na analyzující model  $\ell_1$  relaxace a také porovnání s modelem syntetizujícím, v němž bylo ve většině případů prokázáno znatelně menší zlepšení výsledků rekonstrukce. Obecně však byla potvrzena opodstatněnost navržených metod a jejich modifikací.

V části 6.5 pak bylo ukázáno, že mezi nekonvexními algoritmy SPAIN výrazně převyšuje výsledky staršího OMP. Pozornost byla také věnována přibližnému výpočtu jednoho z kroků v algoritmu S-SPAIN, kde bylo ukázáno, že bližší aproximace v tomto kroku překvapivě nevede k lepšímu výsledku celého iteračního algoritmu, naopak výsledky rekonstrukce z pohledu SNR soustavně zhoršuje.

Závěrečné srovnání metod v části 6.6 pak ukázalo, že navržený algoritmus SPAIN přibližně dosáhl úrovně rekonstrukce Janssenova algoritmu a jednoznačně přesáhl neupravený model konvexní relaxace, a to v syntetizujícím i analyzujícím případě. Pozo-

ruhodným výsledkem však je, že i přes jednoduchost analyzujícího modelu  $\ell_1$  relaxace s váhováním atomů (ve srovnání s mnohem sofistikovanějším Janssenovým algoritmem či SPAIN) tento přístup pro díry větší délky přesáhl z hlediska SNR všechny ostatní metody. S následným váhováním v časové oblasti pak je jeho prvenství z hlediska SNR pro delší díry jednoznačné, především však dosahuje i z hlediska ostatních kritérií výsledků lepších než ostatní testované algoritmy.

Popsané výsledky do budoucna vybízí k analýze, zda podobné váhování atomů lze zařadit do algoritmu SPAIN, pokud by se ukázalo, že pokles energie je stěžejním problémem i při tomto heuristickém přístupu. Alternativní a slibnou možností je pak váhování atomů na základě psychoakustického modelu, jehož cílem je podporovat koeficienty odpovídající perceptuálně významným harmonickým složkám rekonstrukce. Tento přístup se v nedávné době ukázal jako efektivní pro audio declipping [39] a nabízí se proto jeho testování pro příbuzné úlohy restaurace audiosignálů.

# Literatura

- [1] *EBU SQAM CD*. Sound Quality Assessment Material recordings for subjective tests [online]. Dostupné z: <https://tech.ebu.ch/publications/sqamcd>.
- [2] ABERGEL, R.: *Legendre-Fenchel and Rockafellar Duality* [online]. Dostupné z: <http://helios.mi.parisdescartes.fr/~rabergel/legendre-fenchel-rockafellar-duality-4tv.html>.
- [3] ADLER, A. et al.: Audio Inpainting. *IEEE Transactions on Audio, Speech, and Language Processing*. 2012, **20**(3), 922–932. ISSN 1558-7916, DOI: 10.1109/TASL.2011.2168211.
- [4] BAHAT, Y., SCHECHNER, Y. Y. a ELAD, M.: Self-content-based audio inpainting. *Signal Processing*. 2015, číslo 111, 61–72. ISSN 0165-1684. Dostupné z: <http://www.sciencedirect.com/science/article/pii/S0165168414005623>.
- [5] BEN-ISRAEL, A. a GREVILLE, T. N.: *Generalized Inverses: Theory and Applications*. Berlin: Springer, 2003. ISBN 0-387-00293-6.
- [6] BLUMENSATH, T.: *Sparsify Toolbox* [online]. Dostupné z: <https://www.southampton.ac.uk/engineering/about/staff/tb1m08.page#software>.
- [7] BOYD, S. P. et al.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*. 2011, **3**(1), 1–122. Dostupné z: <http://dblp.uni-trier.de/db/journals/ftml/ftml13.html#BoydPCPE11>.
- [8] BOYD, S. P. a VANDENBERGHE, L.: *Convex Optimization*. Cambridge University Press, 2004. ISBN 0521833787.
- [9] CANDÈS, E. J., WAKIN, M. B. a BOYD, S. P.: Enhancing Sparsity by Reweighted  $\ell_1$  Minimization. *Journal of Fourier Analysis and Applications*. 2008, **14**, 877–905.
- [10] CHAMBOLLE, A. a POCK, T.: A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*. 2011, **40**(1), 120–145. ISSN 1573-7683, DOI: 10.1007/s10851-010-0251-1.
- [11] CHRISTENSEN, O.: A short introduction to frames, Gabor systems, and wavelet systems. *Azerbaijan Journal of Mathematics*. 2014, **4**(1), 25–39. ISSN 2218-6816.
- [12] COMBETTES, P. a PESQUET, J.: Proximal splitting methods in signal processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. 2011. DOI: 10.1007/978-1-4419-9569-8\_10.
- [13] CONDAT, L.: A Generic Proximal Algorithm for Convex Optimization—Application to Total Variation Minimization. *IEEE Signal Processing Letters*. 2014, **21**(8), 985–989. ISSN 1070-9908, DOI: 10.1109/LSP.2014.2322123.
- [14] DAŇKOVÁ, M. a RAJMÍČ, P.: Low-rank model for dynamic MRI: joint solving and debiasing. In: *ESMRMB 2016, 33rd Annual Scientific Meeting, Vienna. Magnetic resonance materials in physics biology and medicine*. Berlin: Springer, 2016.

- [15] ČERMÁK, L. a HLAVIČKA, R.: *Numerické metody*. Brno: Vysoké učení technické v Brně, 2006.
- [16] ETTER, W.: Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters. *IEEE Transactions on Signal Processing*. 1996, **44**(5), 1124–1135. ISSN 1053-587X, DOI: 10.1109/78.502326.
- [17] HARRIS, F.: On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*. 1978, **66**(1), 51–83. DOI: 10.1109/proc.1978.10837.
- [18] HASTIE, T., TIBSHIRANI, R. a WAINWRIGHT, M.: *Statistical Learning with Sparsity*. Boca Raton: CRC Press, 2015. ISBN 978-1-4987-1216-3.
- [19] HUBER, R. a KOLLMEIER, B.: PEMO-Q—A New Method for Objective Audio Quality Assessment Using A Model of Auditory Perception. *IEEE Trans. Audio Speech Language Proc.* 2006, **14**(6), 1902–1911. DOI: 10.1109/TASL.2006.883259.
- [20] JANSSEN, A. J. E. M., VELDHUIS, R. N. J. a VRIES, L. B.: Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Trans. Acoustics, Speech and Signal Processing*. 1986, **34**(2), 317–330. DOI: 10.1109/TASSP.1986.1164824.
- [21] KITIĆ, S., BERTIN, N. a GRIBONVAL, R.: Sparsity and cosparsity for audio de-lipping: a flexible non-convex approach. In: *LVA/ICA*. 2015.
- [22] KOMODAKIS, N. a PESQUET, J.: Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems. *IEEE Signal Processing Magazine*. 2015, **32**(6), 31–54. ISSN 1053-5888, DOI: 10.1109/MSP.2014.2377273.
- [23] KOWALSKI, M., SIEDENBURG, K. a DÖRFLER, M.: Social Sparsity! Neighborhood Systems Enrich Structured Shrinkage Operators. *IEEE Transactions on Signal Processing*. 2013, **61**(10), 2498–2511. ISSN 1053-587X, DOI: 10.1109/TSP.2013.2250967.
- [24] LAGRANGE, M., MARCHAND, S. a RAULT, J.-B.: Long Interpolation of Audio Signals Using Linear Prediction in Sinusoidal Modeling. *J. Audio Eng. Soc.* 2005, **53**(10), 891–905. Dostupné z: <http://www.aes.org/e-lib/browse.cfm?elib=13390>.
- [25] LIEB, F. a STARK, H.-G.: Audio inpainting: Evaluation of time-frequency representations and structured sparsity approaches. *Signal Processing*. 2018, číslo 153, 291–299. DOI: 10.1016/j.sigpro.2018.07.012.
- [26] MARAFIOTI, A. et al.: *A context encoder for audio inpainting*. 2018. Dostupné z: <https://arxiv.org/abs/1810.12138>.
- [27] MOKRÝ, O.: Restaurace poškozených audiosignálů pomocí řídkých reprezentací. Brno, 2017. Bakalářská práce. Vysoké učení technické v Brně. Fakulta strojního inženýrství. Ústav matematiky. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/100257>.

- [28] MOKRÝ, O., ZÁVIŠKA, P., RAJMIC, P. a VESELÝ, V.: *Introducing SPAIN (SParse Audio INpainter)*. 2018. Dostupné z: <https://arxiv.org/abs/1810.13137>.
- [29] PATI, Y., REZAIIFAR, R. a 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.
- [30] PERRAUDIN, N. et al.: Inpainting of long audio segments with similarity graphs. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2018. DOI: 10.1109/taslp.2018.2809864.
- [31] PRŮŠA, Z. et al.: The Large Time-Frequency Analysis Toolbox 2.0. In: *Sound, Music, and Motion*. Berlin: Springer, 2014. ISBN 978-3-319-12975-4. Dostupné z: [http://dx.doi.org/10.1007/978-3-319-12976-1\\_25](http://dx.doi.org/10.1007/978-3-319-12976-1_25)
- [32] RAJMIC, P.: Exact risk analysis of wavelet spectrum thresholding rules. In: *Proceedings of the 2003 10th IEEE International Conference on Electronics, Circuits and Systems..* 2003, **2**, 455–458. DOI: 10.1109/ICECS.2003.1301820.
- [33] RAJMIC, P.: *Řídké a nízkohodnostní reprezentace signálů s aplikacemi*. Brno, 2014. Habilitační práce. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací.
- [34] RAJMIC, P. et al.: Acceleration of Audio Inpainting by Support Restriction. In: *7th International Congress on Ultra Modern Telecommunications and Control Systems*. 2015. Dostupné z: <http://www.utko.feec.vutbr.cz/~rajmic/papers/p325-rajmic-ICUMT2015.pdf>
- [35] RAJMIC, P., ZÁVIŠKA, P., VESELÝ, V. a MOKRÝ, O.: *New Generalized Projection Speeds Up Audio Declipping*. 2019. Dostupné z: <https://www.preprints.org/manuscript/201903.0093/v1>.
- [36] TOUMI, I. a EMIYA, V.: Sparse Non-Local Similarity Modeling for Audio Inpainting. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018. DOI: 10.1109/icassp.2018.8462187.
- [37] VESELÝ, V. a RAJMIC, P.: *Funkcionální analýza s aplikacemi ve zpracování signálů*. Brno: Vysoké učení technické v Brně, 2014. ISBN 978-80-214-5186-5.
- [38] ZÁVIŠKA, P. et al.: Revisiting Synthesis Model in Sparse Audio Declipper. *LVA/ICA*. Berlin: Springer, 2018. ISBN 978-3-319-93764-9.
- [39] ZÁVIŠKA, P., RAJMIC, P. a SCHIMMEL, J.: *Psychoacoustically motivated audio declipping based on weighted l1 minimization*. 2019. Dostupné z: <https://arxiv.org/abs/1905.00628>.
- [40] ZÁVIŠKA, P., MOKRÝ, O. a RAJMIC, P.: *A Proper Version of Synthesis-based Sparse Audio Declipper*. 2018. Dostupné z: <https://arxiv.org/abs/1810.12204>.
- [41] ZÁVIŠKA, P., MOKRÝ, O. a RAJMIC, P.: *S-SPADE Done Right: Detailed Study of the Sparse Audio Declipper Algorithms*. Brno: Vysoké učení technické v Brně, 2018. Dostupné z: <https://arxiv.org/abs/1809.09847>.



# Seznam symbolů, veličin a zkratek

## Číselné množiny

$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$	po řadě množiny všech přirozených, celých, reálných a komplexních čísel; přirozená čísla uvažujeme včetně nuly
$\overline{\mathbb{R}}$	množina rozšířených reálných čísel, $\mathbb{R} \cup \{-\infty, \infty\}$
$ \{\dots\} $	mohutnost množiny
$(a, b)$	otevřený interval daný čísly $a, b \in \overline{\mathbb{R}}, a < b$
$\langle a, b \rangle$	uzavřený interval daný čísly $a, b \in \mathbb{R}, a \leq b$

## Unární operace s čísly (případně s vektory po složkách)

$\lceil a \rceil, \lfloor a \rfloor$	horní a dolní celá část reálného čísla $a$
$\Re(z), \Im(z)$	reálná a imaginární část komplexního čísla $z$
$\bar{z}$	komplexně sdružené číslo k číslu $z$
$\arg(z)$	argument komplexního čísla $z$
$ z $	modul komplexního čísla $z$

## Vektorové prostory

$\mathbf{u}, \mathbf{v}, \dots$	vektory
$\mathbf{X}, \mathbf{Y}, \dots$	matice
$\mathbf{x}^\top, \mathbf{X}^\top$	transpozice vektoru či matice
$\mathbf{x}^*, \mathbf{X}^*$	hermitovská transpozice vektoru či matice
$\ \cdot\ $	norma na prostoru dle kontextu (norma vektoru, operátoru)
$\ \cdot\ _p$	$\ell_p$ norma (definice 2.1 a 2.2 a poznámka 2.3)
$\ \cdot\ _0$	řídkost, nebo též $\ell_0$ pseudonorma (definice 2.4)
$\text{supp}(\mathbf{x})$	nosič vektoru $\mathbf{x} \in \mathbb{C}^N$ , $\text{supp}(\mathbf{x}) = \{n = 1, 2, \dots, N \mid x_n \neq 0\}$
$\langle \mathbf{u}, \mathbf{v} \rangle$	skalární součin vektorů $\mathbf{u}$ a $\mathbf{v}$

## Operátory a funkce

$F, F^*$	operátor a příslušný adjungovaný operátor (mezi libovolnými prostory)
$f, g, \dots$	reálné funkce reálné či komplexní proměnné (případně více proměnných)
$f^*$	konvexně sdružená funkce k funkci $f$ (definice 2.24)
$\sup_{x \in X} f(x)$	supremum funkčních hodnot funkce $f$ na množině $X$
$\exp(x)$	exponenciální funkce, též $e^x$
$\iota_Y(x)$	indikátorová funkce množiny $Y$ (definice 2.18)
$\mathcal{R}(A)$	obor hodnot operátoru $A$
$\text{Id}$	operátor identity
$\text{prox}_f$	proximální operátor funkce $f$ (definice 2.21)
$\mathcal{P}_Y$	operátor projekce na množinu $Y$ (definice 2.20)
$\mathcal{H}_k$	tvrdé prahování s prahem $k$
$\mathcal{S}_\tau$	měkké prahování s prahem $\tau$
$\mathcal{C}_\tau$	„klipovací“ operátor s prahem $\tau$ (definovaný v rámci lemmatu 4.3)

## Zkratky

STFT	short-time Fourier transform (krátkodobá Fourierova transformace, též Gaborova transformace)
DFT	discrete Fourier transform (diskrétní Fourierova transformace)
OMP	orthogonal matching pursuit
SPADE	sparse audio declipper
SPAIN	sparse audio inpainter
A-SPAIN	analyzující varianta SPAIN
S-SPAIN H	syntetizující varianta SPAIN s tvrdým prahováním
S-SPAIN OMP	syntetizující varianta SPAIN s vloženým OMP
SNR	signal-to-noise ratio, jednotkou je decibel (dB)
$PSM_t$	perceptual similarity measure
ODG	objective difference grade

## Konvence značení pro úlohy audio inpaintingu

$\mathbf{y}$	pozorovaný (poškozený) signál, $\mathbf{y} \in \mathbb{C}^N$
$M_{\mathbf{R}}$	operátor, který na pozice chybějících vzorků dosadí nuly, $M_{\mathbf{R}}: \mathbb{C}^N \rightarrow \mathbb{C}^N$
$\mathbf{y}_{\mathbf{R}}$	nepoškozená část pozorovaného signálu doplněná nulami, $\mathbf{y}_{\mathbf{R}} = M_{\mathbf{R}}\mathbf{y}$
$\hat{\mathbf{x}}$	řešení problému audio inpaintingu v časové oblasti, $\hat{\mathbf{x}} \in \mathbb{C}^N$
$\Gamma$	množina přípustných řešení, $\Gamma = \{\mathbf{x} \in \mathbb{C}^N \mid M_{\mathbf{R}}\mathbf{x} = \mathbf{y}_{\mathbf{R}}\} \subset \mathbb{C}^N$
$A, D$	operátory analýzy a syntézy (definice 2.11)

## Značení parametrů Gaborových framů

$w$	délka okna, v této práci 2800 vzorků (přibližně 64 ms)
$a$	posun mezi sousedními okny, v této práci nejčastěji $w/4$
$M$	počet kmitočtových kanálů, v této práci 2800

## Zkratky použité v experimentech pro váhování atomů v $\ell_1$ minimalizaci a váhování doplněného úseku v čase

<b>none</b>	váhování konstantou (4.21a)
<b>supp</b>	... dle nosiče (4.21b)
<b>abs</b>	... dle $\ell_1$ normy (4.21c)
<b>norm</b>	... dle $\ell_2$ normy (4.21d)
<b>energy</b>	... dle energie (4.21e)
<b>segs</b>	počet segmentů, z nichž počítáme energii
<b>lens</b>	délka jednoho segmentu
<b>gaps</b>	počet navíc doplněných děr
<b>shift</b>	posun mezi sousedními navíc vytvořenými dírami

# Seznam příloh

A Obsah přiloženého CD

75



# A Obsah přiloženého CD

Přiložené CD obsahuje kódy pro MATLAB potřebné k realizaci všech popsaných metod. Soubory jsou organizovány následovně:

```
/ ..... kořenový adresář přiloženého CD
├── Janssen ..... funkce pro realizaci Janssenova algoritmu
│   ├── inpaintFrame_janssenInterpolation.m
│   └── janssen.m
├── OMP ..... funkce pro realizaci algoritmu OMP
│   ├── greed_omp_qr.m
│   └── omp_segmentation.m
├── PemoQ ..... kompletní balíček pro perceptuálně motivované hodnocení
├── reweighted l1 relaxation ..... funkce pro realizaci konvexních metod
│   ├── DouglasRachford.m
│   ├── ChambollePock.m
│   ├── reweighted.m
│   └── tdc.m
├── SPAIN ..... funkce pro realizaci algoritmu SPAIN
│   ├── aspain.m
│   ├── franamp.m
│   ├── hard_thresholding.m
│   ├── spain_segmentation.m
│   └── sspain.m
├── EBU_SQAM.mat
├── inpainting_comparison_multigap.m ..... hlavní skript určený ke spuštění
├── min_sig_supp_2.m
├── offset.m
├── proj_time.m
├── snr_n.m
├── syn_atoms.m
└── weights.m
```

Některé funkce byly převzaty z externích zdrojů, konkrétně:

- funkce `inpaintFrame_janssenInterpolation.m` byla převzata z [3], ke stažení v rámci Audio Inpainting Toolbox na <http://www.small-project.eu/software-data.html>,
- funkce `greed_omp_qr.m` byla převzata ze Sparsify Toolbox [6],
- funkce `franamp.m` byla převzata z LTFAT [31] a byla upravena pro potřeby této práce, zejména byla zajištěna kompatibilita s redundantní DFT,
- funkce `min_sig_supp_2.m` byla převzata z [34] a byla upravena podle algoritmu 5.

Všechny kódy byly testovány v prostředí MATLAB R2017b. Pro jejich funkčnost je navíc nutné mít spuštěný toolbox LTFAT (ke stažení na <http://ltfat.github.io/>).