ÖAW AUSTRIAN ACADEMY OF SCIENCES

ARI – ACOUSTICS RESEARCH INSTITUTE

Dictionary Learning for Sparse Audio Inpainting

Georg Tauböck

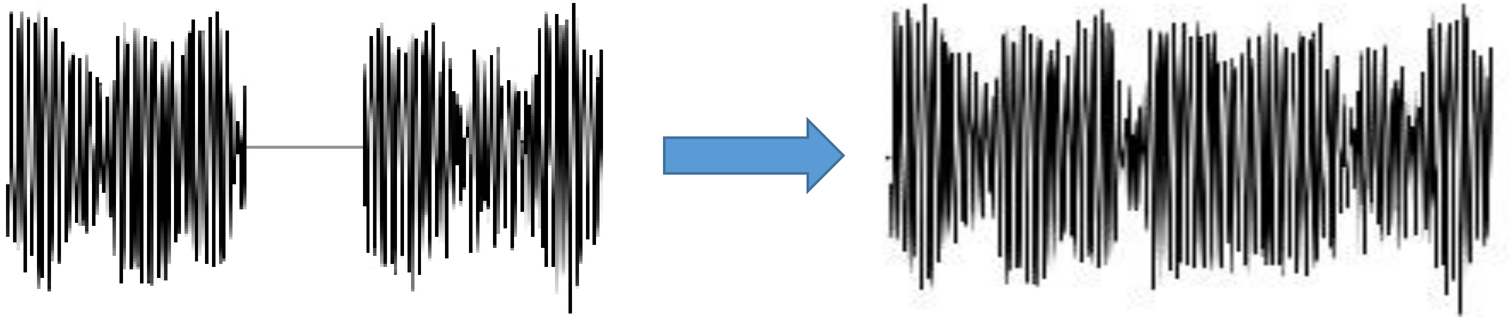
joint work with

Shristi Rajbamshi, Nicki Holighaus, and Peter Balazs

Deep Learning Seminar, Vienna, April 26, 2023

Audio Inpainting

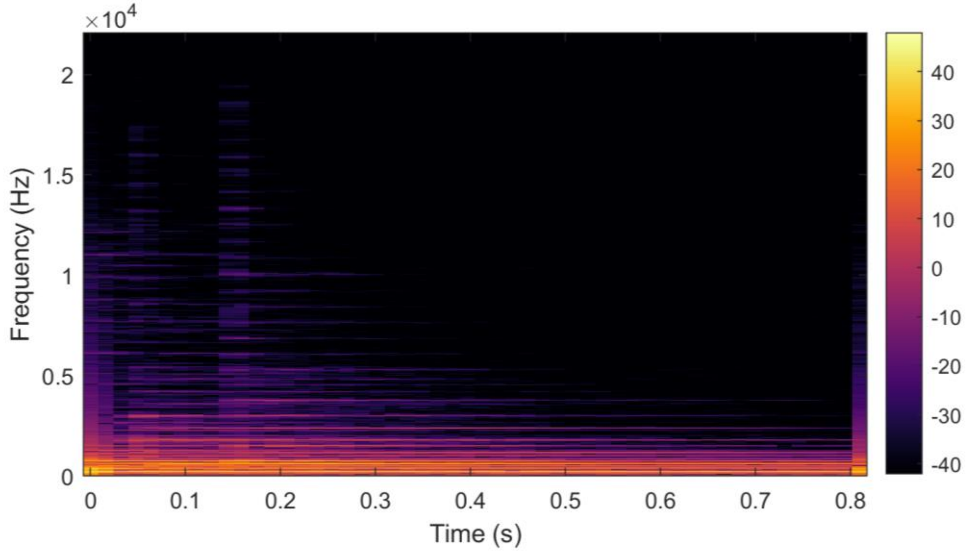
- Assume: audio signal has gap (missing or unreliable consecutive samples)
- Problem: reconstruct original signal within gap



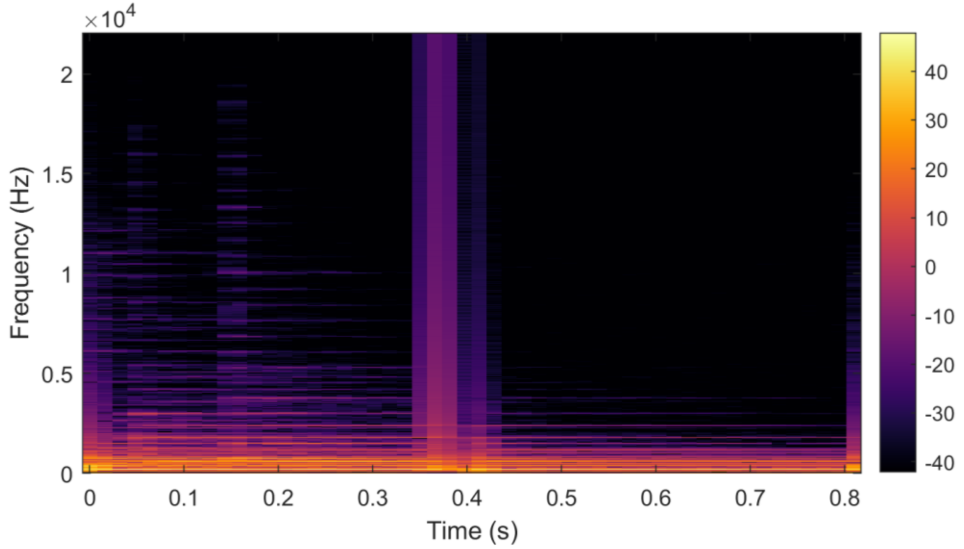
signal with gap signal with filled gap

Sparse Audio Inpainting

- Key approach: exploit time-frequency (TF) sparsity of real-world audio signals



original signal



gapped signal

Sparse Audio Inpainting

- Idea: use reliable samples from adjacent parts of the gap

↓

set of all feasible signals: $\Gamma_{\mathrm{x}} \triangleq \{\mathrm{y} \in \mathbb{R}^N \colon \mathrm{M_R y} = \mathrm{M_R x}\}$

$\mathrm{x} \in \mathbb{R}^N$ … time-domain signal

$\mathrm{M_R} \colon \mathbb{R}^N \to \mathbb{R}^N$ … binary “reliable mask” projection operator (keeps signal samples corresponding to reliable part; sets others to zero)

SPAIN Algorithm

- ℓ_0-minimization in TF domain subject to feasible set:

$$\min_{b,y} \|b\|_0 \quad \text{s.t.} \quad y \in \Gamma_x \text{ and } \|Ay - b\|_2 \leq \epsilon \qquad \text{(P0)}$$

$\|\cdot\|_0$ … number of non-zeros

A … frame analysis operator (used to compute TF coefficients)

- Applies Alternating Direction Method of Multipliers (ADMM) to solve (P0)
- Segment-wise implementation using redundant DFT dictionary and overlap-add method
- Similar to SPADE-algorithm for de-clipping

Modified SPAIN Algorithm

- $\ell_{0,\infty}$-minimization in TF domain, where the supremum is taken over time:

$$\min_{\mathrm{B,y}} \|\mathrm{B}\|_{0,\infty} \quad \mathrm{s.t.} \quad \mathrm{y} \in \Gamma_{\mathrm{x}} \ \text{and} \ \|\eta(\mathrm{A_G y}) - \mathrm{B}\|_{\mathrm{F}} \leq \epsilon \qquad (\mathrm{P0}\infty)$$

$$\|\mathrm{X}\|_{0,\infty} = \max\{\|\mathrm{x}_0\|_0, \|\mathrm{x}_1\|_0, \dots, \|\mathrm{x}_{K-1}\|_0\}, \quad \text{where } \mathrm{X} = [\mathrm{x}_0 \ \mathrm{x}_1 \cdots \mathrm{x}_{\mathrm{K}-1}]$$

$\|\cdot\|_{\mathrm{F}}$ … Froebenius norm $\qquad \eta(\mathrm{A_G} \ \cdot)$ … real discrete Gabor transform (DGT)

- Applies ADMM to solve (P0∞)
- Exploits conjugate-symmetry in TF domain
- Works without segment-wise implementation
- Not only for real DGT, but for arbitrary painless frames ($\Leftrightarrow$ $\mathrm{A^H A}$ is diagonal)

Alternating Direction Method of Multipliers (ADMM)

- Solves:

$$\min_{\mathrm{y}} f(\mathrm{y}) + g(\mathrm{Ay}),$$

where $\mathrm{y} \in \mathbb{C}^N$, $\mathrm{A}: \mathbb{C}^N \to \mathbb{C}^P$ linear operator, f and g are real convex functions

- Reformulation:

$$\min_{\mathrm{y,b}} f(\mathrm{y}) + g(\mathrm{b}) \quad \mathrm{s.t.} \quad \mathrm{Ay} - \mathrm{b} = 0$$

- Fix sparsity paramter k, let $\mathcal{S}_k \triangleq \{\mathrm{X} : \|\mathrm{X}\|_{0,\infty} \leq k\}$, and rewrite (P0∞) with $\epsilon = 0$ as:

$$\min_{\mathrm{B,y}} \ell_{\mathcal{S}_k}(\mathrm{B}) + \ell_{\Gamma_{\mathrm{x}}}(\mathrm{y}) \ \ \mathrm{s.t.} \quad \mathrm{A_G y} - \eta^{\dagger}(\mathrm{B}) = 0$$

Alternating Direction Method of Multipliers (ADMM)

- Augmented Lagrangian in *scaled form*:

$$\mathcal{L}_\delta(\mathrm{y}, \mathrm{r}, \mathrm{B}) = \ell_{\mathcal{S}_k}(\mathrm{B}) + \ell_{\Gamma_\mathrm{x}}(\mathrm{y}) + \frac{\delta}{2}\left\|\mathrm{A}_\mathrm{G}\mathrm{y} - \eta^\dagger(\mathrm{B}) + \mathrm{r}\right\|_2^2 - \frac{\delta}{2}\|\mathrm{r}\|_2^2$$

- Update rules:

$$\mathrm{B}^{(i+1)} = \underset{\mathrm{B}:\|\mathrm{B}\|_{0,\infty}\leq k}{\operatorname{argmin}} \left\|\mathrm{A}_\mathrm{G}\mathrm{y}^{(i)} - \eta^\dagger(\mathrm{B}) + \mathrm{r}^{(i)}\right\|_2$$

$$\mathrm{y}^{(i+1)} = \underset{\mathrm{y}\in\Gamma_\mathrm{x}}{\operatorname{argmin}} \left\|\mathrm{A}_\mathrm{G}\mathrm{y} - \eta^\dagger(\mathrm{B}^{(i+1)}) + \mathrm{r}^{(i)}\right\|_2$$

$$\mathrm{r}^{(i+1)} = \mathrm{r}^{(i)} + \mathrm{A}_\mathrm{G}\mathrm{y}^{(i+1)} - \eta^\dagger(\mathrm{B}^{(i+1)})$$

Modified SPAIN Algorithm

Algorithm 1: A-SPAIN-MOD

Input: A_G, D_G^{cd}, $M_R x$, Γ_x, s, t, ϵ

Output: $\hat{x}$

1 $y^{(0)} = M_R x$, $R^{(0)} = 0$, $i = 0$, $k = s$

2 $B^{(i+1)} = \mathcal{H}_k^{TF}\left(\eta\left(A_G y^{(i)}\right) + R^{(i)}\right)$

3 $y^{(i+1)} = \arg\min_{y \in \Gamma_x} \left\| y - D_G^{cd}\left(\eta^\dagger\left(B^{(i+1)} - R^{(i)}\right)\right)\right\|_2$

4 **if** $\left\|\eta\left(A_G y^{(i+1)}\right) - B^{(i+1)}\right\|_F \leq \epsilon$ **then**

5 terminate

6 **else**

7 $R^{(i+1)} = R^{(i)} + \eta\left(A_G y^{(i+1)}\right) - B^{(i+1)}$

8 $i \leftarrow i + 1$

9 **if** $i \mod t = 0$ **then**

10 $k \leftarrow k + s$

11 go to 2

12 **return** $\hat{x} = y^{(i+1)}$

time-frequency hard-thresholding operator $\mathcal{H}_k^{TF}(\,\cdot\,)$

$$\mathcal{H}_k^{TF}([x_0\ x_1 \cdots x_{K-1}]) = [\mathcal{H}_k(x_0)\ \mathcal{H}_k(x_1) \cdots \mathcal{H}_k(x_{K-1})]$$

inverse real DGT $D_G^{cd}(\eta^\dagger(\,\cdot\,))$

Dictionary Learning

- Learns frame from neighborhood of gap, such that TF sparsity is maximized:

$$\min_{\mathrm{A}} \|(\eta(\mathrm{Ax}))|\mathcal{N}\|_{0,\infty} \quad \mathrm{s.t.} \quad \mathrm{A}^{\mathrm{H}}\mathrm{A} \text{ is diagonal}$$

- Deforms given Gabor frame by unitary transformation:

$$\min_{\mathrm{U}\in\mathcal{U}} \|\mathrm{U}(\eta(\mathrm{A_G x}))|\mathcal{N}\|_{0,\infty}\,, \qquad \mathcal{U} \ldots \text{ subset of unitary matrices}$$

- Relaxation:

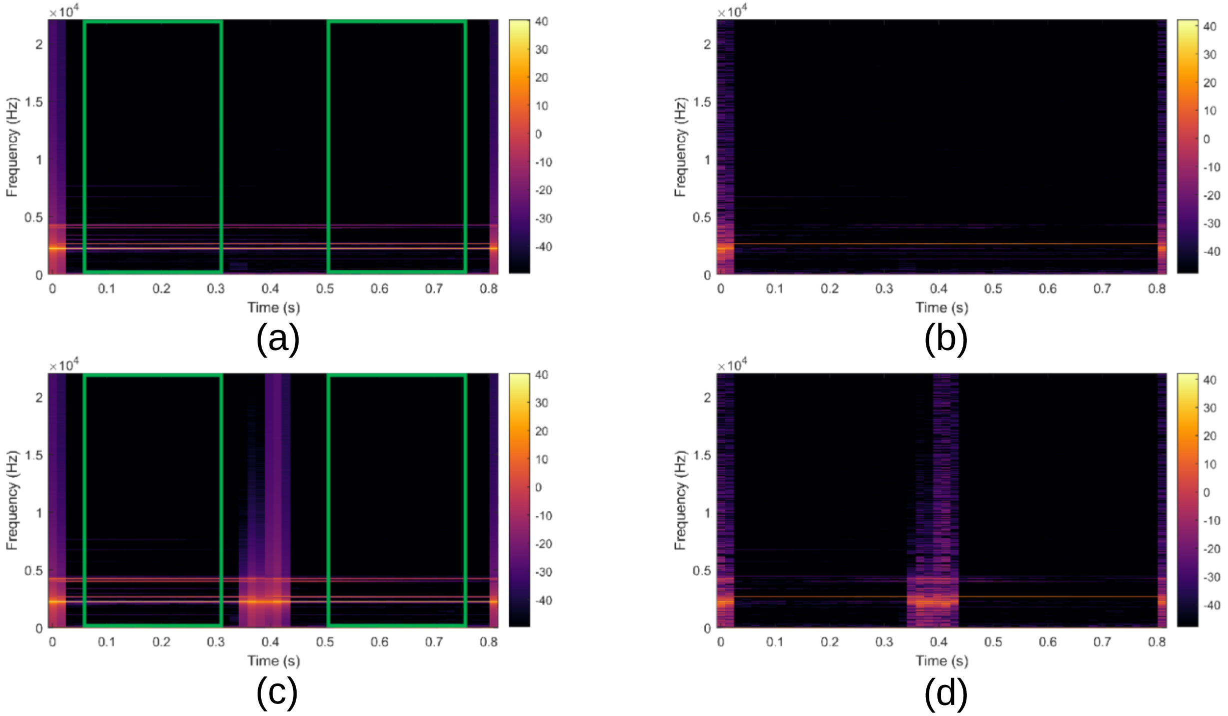
$$\min_{\mathrm{U}\in\mathcal{U}} \sum_{q\in\mathcal{N}} \|\mathrm{U}(\eta(\mathrm{A_G x}))|q\|_1$$

- Solved by basis optimization technique originally developed for wireless communications

AUSTRIAN ACADEMY OF SCIENCES

ARI – ACOUSTICS RESEARCH INSTITUTE

Dictionary Learning



analysis coefficients of signal “a35_glockenspiel”

(a) … Gabor dictionary without gap

(b) … learned dictionary without gap

(c) … Gabor dictionary with gap

(d) … learned dictionary with gap

Coefficients within green rectangles are used for training

Learned SPAIN Algorithm

Algorithm 3: A-SPAIN-LEARNED

Input: $\mathrm{A_G}$, $\mathrm{D_G^{cd}}$, $\mathrm{U_{e/o}}$, $\mathrm{M_R x}$, Γ_{x}, s, t, ϵ

Output: $\hat{\mathrm{x}}$

1 $\mathrm{y}^{(0)} = \mathrm{M_R x}$, $\mathrm{R}^{(0)} = 0$, $i = 0$, $k = s$

2 $\mathrm{B}^{(i+1)} = \mathcal{H}_k^{\mathrm{TF}}\left(\mathrm{U_{e/o}}\eta(\mathrm{A_G y}^{(i)}) + \mathrm{R}^{(i)}\right)$

3 $\mathrm{y}^{(i+1)} = \underset{\mathrm{y}\in\Gamma_{\mathrm{x}}}{\arg\min}\left\|\mathrm{y} - \mathrm{D_G^{cd}}\left(\eta^{\dagger}(\mathrm{U_{e/o}^H B}^{(i+1)} - \mathrm{U_{e/o}^H R}^{(i)})\right)\right\|_2$

4 **if** $\left\|\mathrm{U_{e/o}}\eta(\mathrm{A_G y}^{(i+1)}) - \mathrm{B}^{(i+1)}\right\|_{\mathrm{F}} \leq \epsilon$ **then**

5 terminate

6 **else**

7 $\mathrm{R}^{(i+1)} = \mathrm{R}^{(i)} + \mathrm{U_{e/o}}\eta(\mathrm{A_G y}^{(i+1)}) - \mathrm{B}^{(i+1)}$

8 $i \leftarrow i + 1$

9 **if** $i \mod t = 0$ **then**

10 $k \leftarrow k + s$

11 go to 2

12 **return** $\hat{\mathrm{x}} = \mathrm{y}^{(i+1)}$

$\mathrm{U}_{e/o} \triangleq$ U obtained by basis optimization technique

Basis optimization technique

- Aims at solving:

$$\min_{\mathrm{U}\in\mathcal{U}} \sum_{q\in\mathcal{N}} \|\mathrm{U}(\eta(\mathrm{A_G x}))|q\|_1$$

- Idea:

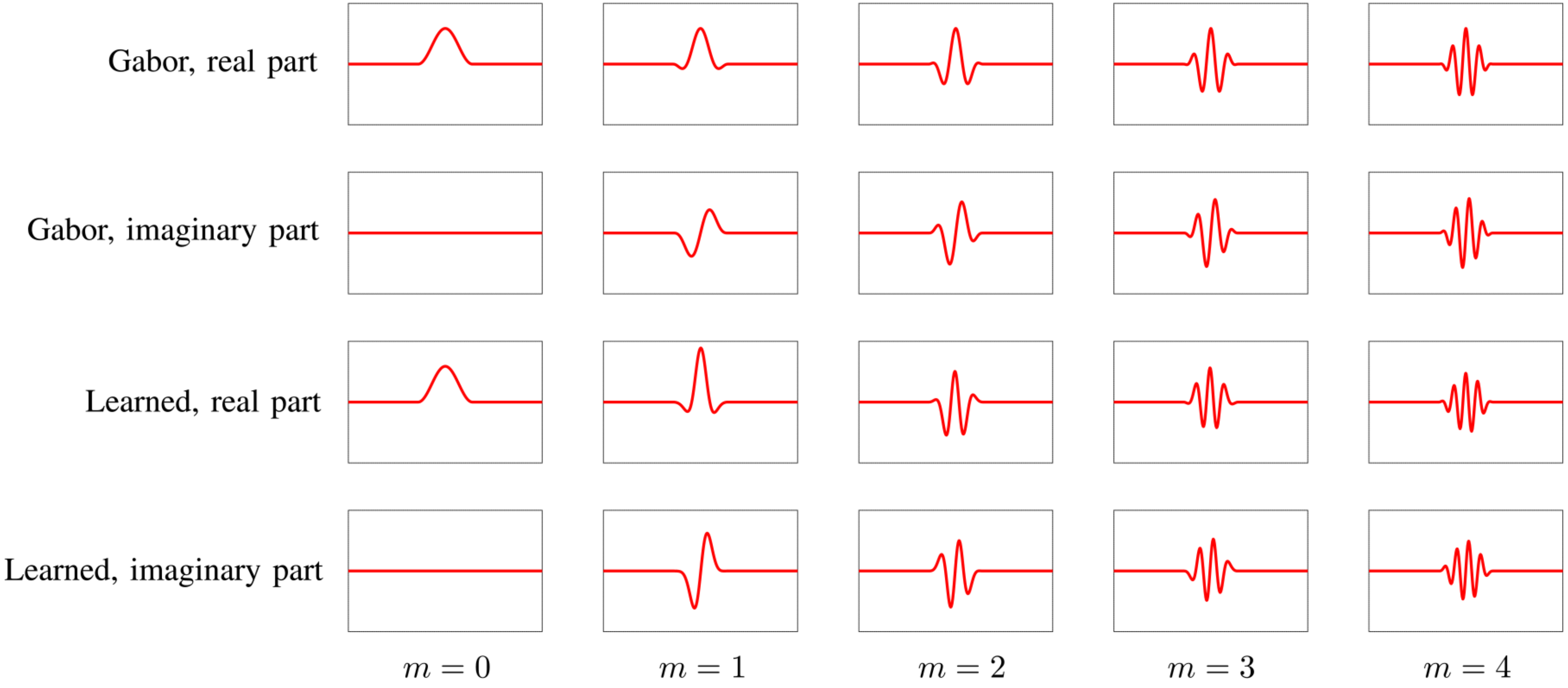
$$\mathrm{U} = e^{\mathrm{iH}} \approx \mathrm{I} + \mathrm{iH}, \qquad \mathrm{H}\ \text{Hermitian},\ \|\mathrm{H}\|\ \text{small}$$

- At rth iteration:

$$\mathrm{H}^{(r)} = \operatorname*{argmin}_{\mathrm{H}\in\mathcal{H}_r} \sum_{q\in\mathcal{N}} \left\|(\mathrm{I}+\mathrm{iH})\ \mathrm{U}^{(r)}\ (\eta(\mathrm{A_G x}))|q\right\|_1 \quad \text{and} \quad \mathrm{U}^{(r+1)} = e^{\mathrm{iH}^{(r)}}\ \mathrm{U}^{(r)}$$

$$\mathcal{H}_r \quad \ldots \quad \text{set of all Hermitian matrices with } \|\mathrm{H}\| \le \rho_r$$

Dictionary Learning



Audio Inpainting via Weighted ℓ_1-Minimization

- ℓ_0-minimization in TF domain subject to feasible set:

$$\min_{y} \|\eta(A_G y)\|_0 \quad \text{s.t.} \quad y \in \Gamma_x \qquad \text{(P0)} \qquad \text{non-convex}$$

- Convex relaxation via weighted ℓ_1-minimization:

$$\min_{y} \|w \odot \eta(A_G y)\|_1 + \ell_{\Gamma_x}(y) \qquad \text{(WP1)}$$

$w \odot$ … element-wise multiplication with weighting vector w

- (WP1) is solved by using a proximal algorithm (Chambolle Pock)

Learned Weighted ℓ_1-Minimization

- Weighted ℓ_1-minimization using learned dictionary:

$$\min_{y} \| w_L \odot U\, \eta(A_G y) \|_1 + \ell_{\Gamma_x}(y) \qquad \text{(WL1)}$$

w_L … new weighting vector adapted to learned dictionary

- (WL1) is solved by using a proximal algorithm (Chambolle Pock)

Simulations – Setup

- Performance measures:
  - signal-to-noise ratio (SNR): $\mathrm{snr}(\mathrm{x}_{\mathrm{orig}}, \mathrm{x}_{\mathrm{inp}}) = 10 \log_{10} \frac{\|\mathrm{x}_{\mathrm{orig}}\|_2^2}{\|\mathrm{x}_{\mathrm{orig}} - \mathrm{x}_{\mathrm{inp}}\|_2^2}$ [dB]
  - PEMO-Q criterion: objective difference grade (ODG)
- Test signals: 10 music recordings from the EBU dataset (sampled with 44.1 kHz)
- Gaps: 5 gaps at random positions; lengths: 5 ms – 85 ms
- TF dictionary: tight Gabor frame with Hann window
  - window length: $w_{\mathrm{g}} = 2800$ samples $= \sim 64$ ms
  - window shift: $a = 700$ samples
  - M = 2800 frequency channels

Simulations – Performance Comparison