Audio inpainting by sparse regularization methods

Journées bordelaises d'analyse mathématique des images

EQUIPE-PROJET
PANAMA, Inria Rennes

Emetteur
Srđan Kitić

April 8, 2014

OUTLINE

General introduction

The audio declipping inverse problem

Sparse and cosparse regularization

Declipping algorithms

Declipping results

Blind decompression

Decompression results

Conclusions

1
AUDIO INPAINTING
General introduction

Inpainting problems in audio signal processing

Recovery of audio signals corrupted by:

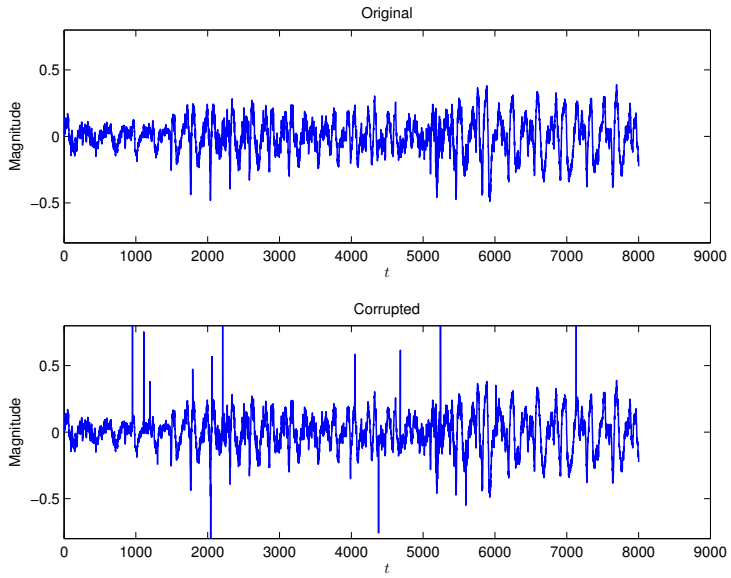
- Impulsive noise / clicks,
- Clipping / magnitude saturation,
- Packet loss,
- CD/DVD scratches,
- Source separation and more.

Different approaches, depending on the context:

- AR modeling [JVV86],
- Bayesian estimation [GR95, MG14],
- Neural networks [Unc03, Czy97],
- Bandwidth replication [LA05],
- Sparse recovery [PBD$^+$10, MG14].

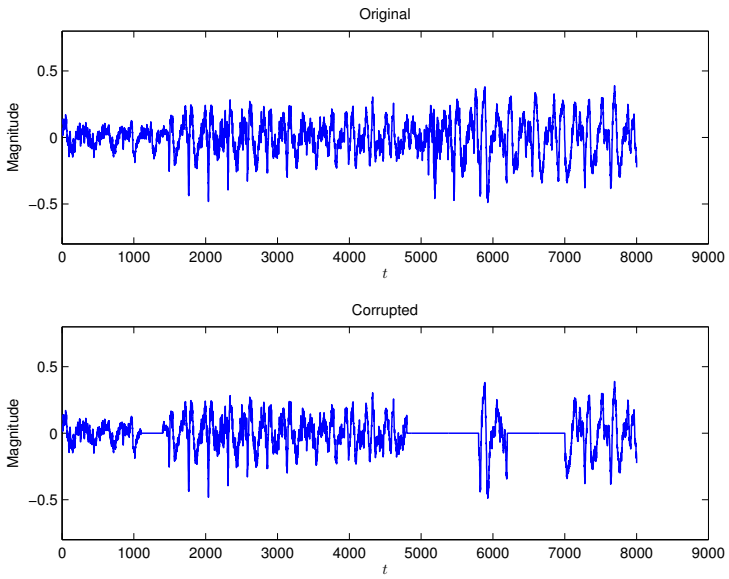
Impulsive noise

- A localized, impulsive degradation, at random position in the signal.
- Duration of the degradation is between 20μs and 4ms.
- Many interpolation approaches, such as median filtering, ”splicing“ etc.
- The most effective is a model-based approach based on AR-process.



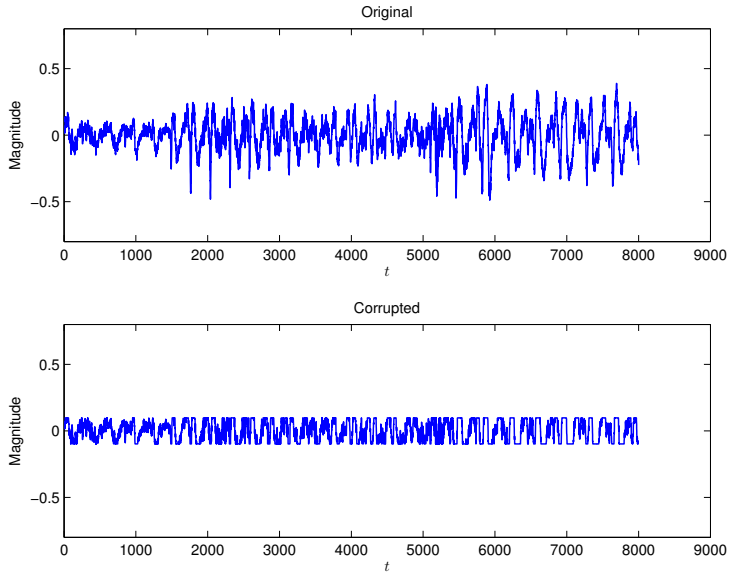
Packet loss

- Very difficult scenario - entire blocks of data are completely lost.
- Duration of the "gap" depends on the packet size and may be over 100ms.
- Packet Loss Concealment (PLC) techniques [WSL00] based on insertion, waveform substitution and model-based methods.
- Typically, speech signals can be recovered if the gap is smaller than a phoneme duration (less than 40ms).



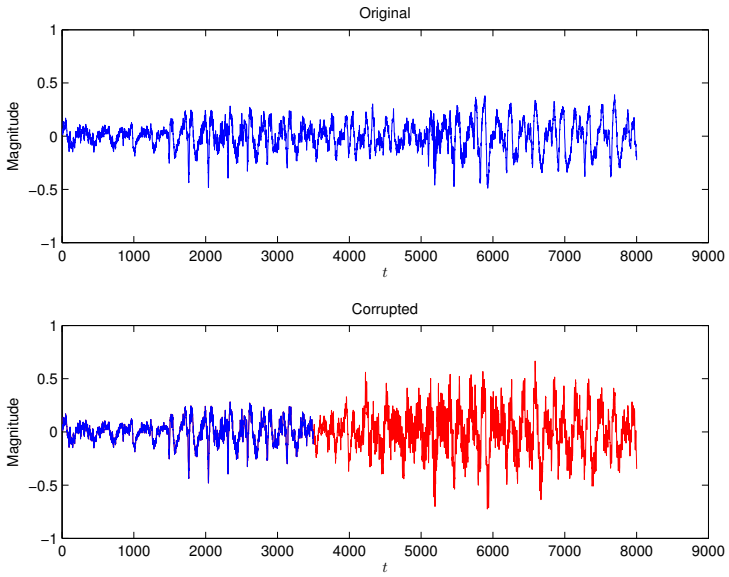
Clipping

- Another difficult scenario - magnitude information above the threshold is completely lost.
- Duration of the ”gap“ depends on the threshold.
- Declipping techniques based on interpolation and signal models.
- Recovery performance depends on the clipping threshold and the audio content.



Source separation

- Specific case where one source is desired and the rest are considered as noise.
- Duration of the ”gap“ depends on the period during which only the desired source is active.
- Standard separation methods based on ICA [CJ10], for example.
- In the case of multichannel audio, pattern matching techniques [SLOVB14] may be effective.



2
AUDIO DECLIPPING
The inverse problem

Mathematical formulation of audio clipping

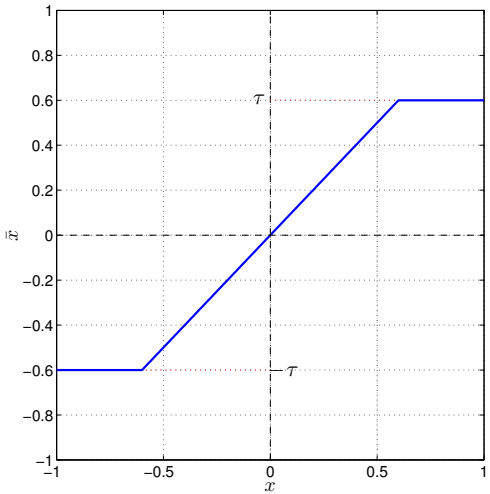
Let $x \in \mathbb{R}^n$ be the single channel, discrete time audio signal and $\mathcal{C}_\tau(x) = \bar{x} \in \mathbb{R}^n$ its clipped version.

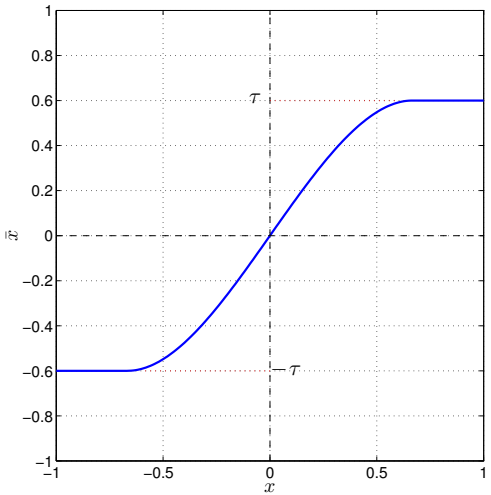
Hard clipping:

$$\bar{x}_i = \begin{cases} x_i & \text{if } |x_i| < \tau, \\ \operatorname{sgn}(x_i)\tau & \text{otherwise.} \end{cases}$$

Soft clipping (by cubic nonlinearity):

$$\bar{x}_i = \begin{cases} \left(\frac{9x_i}{4} - \frac{27x_i^3}{16}\right)\tau & \text{if } |x_i| < \frac{2}{3}\tau, \\ \operatorname{sgn}(x_i)\tau & \text{otherwise.} \end{cases}$$

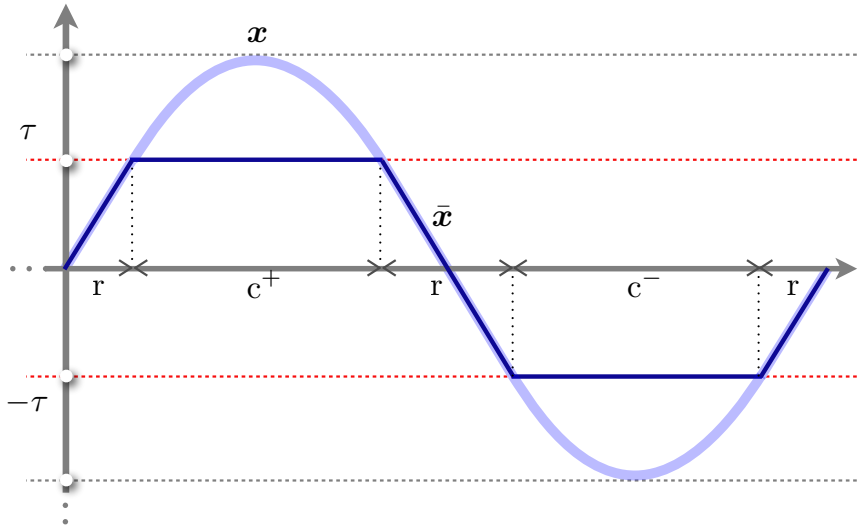




Declipping *hard*-clipped signal

- $\bar{x} = \mathcal{C}_\tau(x)$, a hard-clipped signal.
- $\mathsf{M}_\mathsf{r}, \mathsf{M}_\mathsf{c}^+, \mathsf{M}_\mathsf{c}^-$ extract "reliable", clipped-positive and clipped-negative samples.
- The goal is to find an estimate $\hat{x}$ such that:

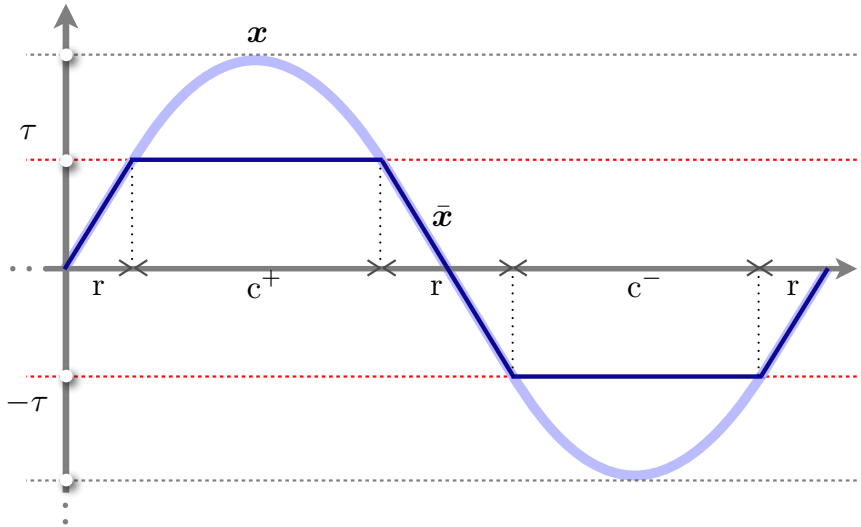
$$\mathsf{M}_\mathsf{r}\,\hat{x} = \mathsf{M}_\mathsf{r}\,\bar{x}$$
$$\mathsf{M}_\mathsf{c}^+\,\hat{x} \geq \mathsf{M}_\mathsf{c}^+\,\bar{x}$$
$$\mathsf{M}_\mathsf{c}^-\,\hat{x} \leq \mathsf{M}_\mathsf{c}^-\,\bar{x}$$



Declipping *hard*-clipped signal

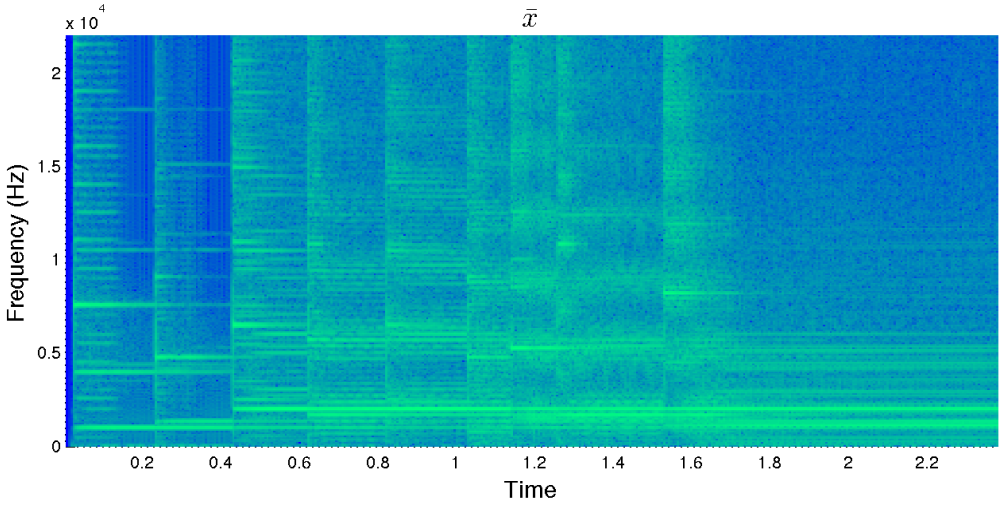
- $\bar{x} = \mathcal{C}_\tau(x)$, a hard-clipped signal.
- $\mathrm{M_r}, \mathrm{M_c^+}, \mathrm{M_c^-}$ extract "reliable", clipped-positive and clipped-negative samples.
- The goal is to find an estimate $\hat{x}$ such that:

$$\mathrm{M_r}\,\hat{x} = \mathrm{M_r}\,\bar{x}$$
$$\mathrm{M_c^+}\,\hat{x} \geq \mathrm{M_c^+}\,\bar{x}$$
$$\mathrm{M_c^-}\,\hat{x} \leq \mathrm{M_c^-}\,\bar{x}$$

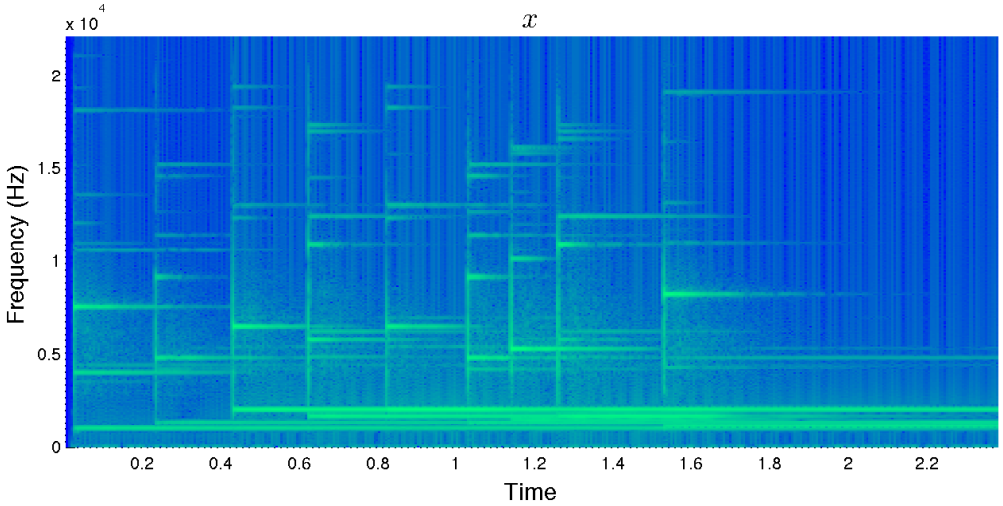


An ill-posed problem!

Time-frequency visualization

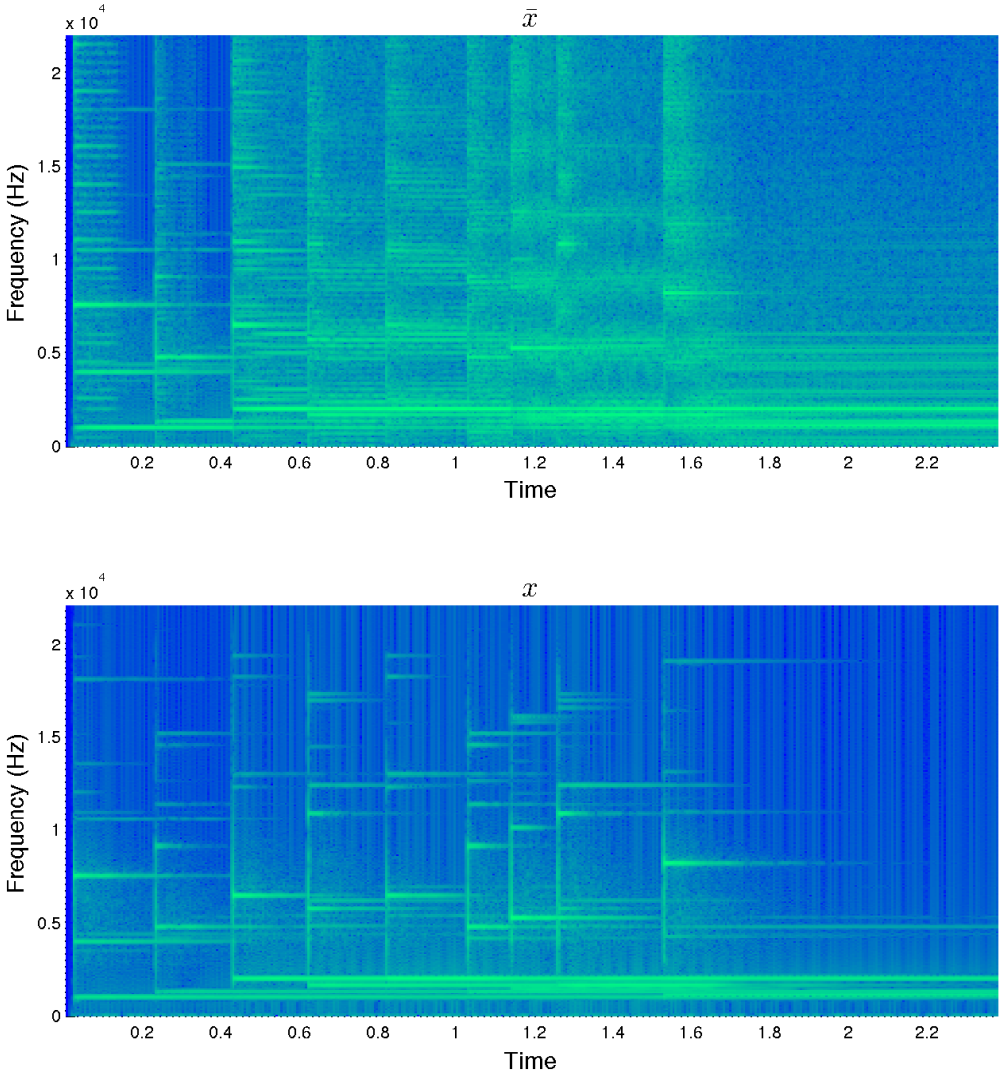


Clipped: spectrum spreading due to introduced discontinuities in the waveform.



Original: in the time-frequency plane, the energy of audio signals is *mostly* concentrated!

Time-frequency visualization



The idea: regularize the ill-posed declipping problem by enforcing the energy compactness in an estimate.

3
SPARSE REGULARIZATION

Synthesis *vs* analysis approach

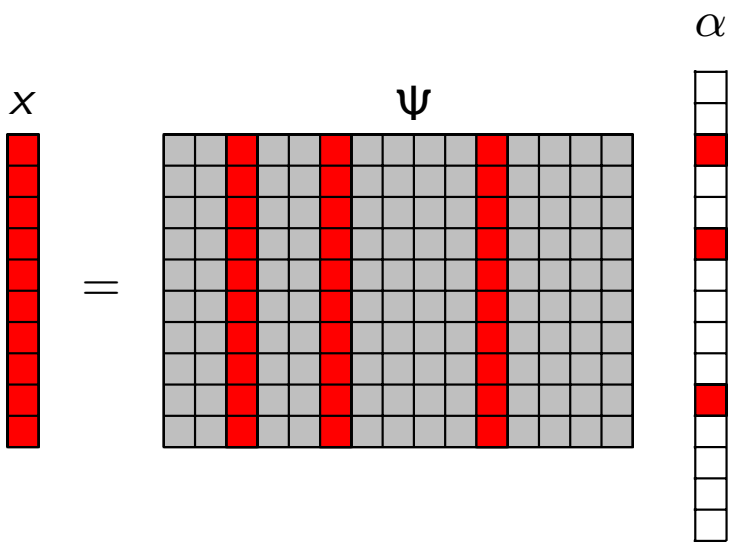
Sparse synthesis framework

The assumption: signal $x \in \mathbb{R}^n$ can be approximated by a linear combination of *atoms* taken from a *dictionary* $\Psi \in \mathbb{R}^{n \times m}$, $n \leq m$:

$$x = \Psi\alpha$$

The number of atoms (eq. non-zero weights in $\alpha \in \mathbb{R}^m$) k needed for the approximation is relatively small compared to N:

$$\#\{\alpha\} = \|\alpha\|_0 = k \ll n, m$$



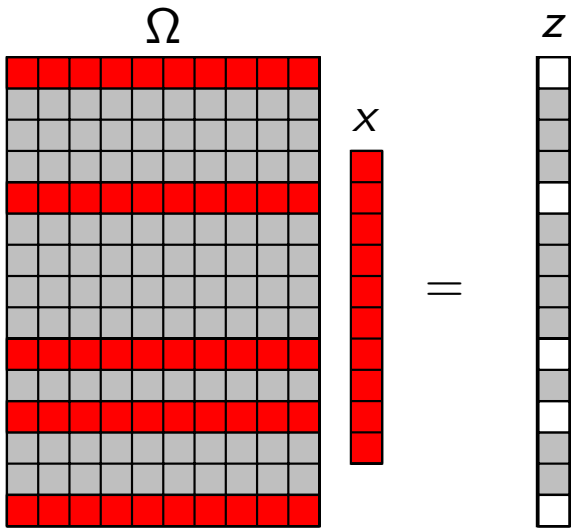
Sparse analysis framework

The assumption [NDEG13]: signal $x \in \mathbb{R}^n$ can be sparsified by applying a suitable *analysis operator* $\Omega \in \mathbb{R}^{p \times n}$, $p \geq n$:

$$z = \Omega x$$

The number of zero-elements ℓ in the product $z \in \mathbb{R}^p$ is relatively large:

$$p - \|\Omega x\|_0 = \ell$$



Comparison of the two methodologies

Sparse synthesis

- Dictionary: $\Psi \in \mathbb{R}^{n \times m}$.
- Constructive model, atomic composition.
- Support: column vectors of Ψ corresponding to non-zeros in α.
- Non-unique representation.
- Number of subspaces: $\binom{m}{k}$, dimension: k.

Sparse analysis

- Operator: $\Omega \in \mathbb{R}^{p \times n}$.
- Descriptive model, constrained decomposition.
- Cosupport: row vectors of Ω orthogonal to x.
- Unique representation.
- Number of subspaces: $\binom{p}{\ell}$, dimension $n - \ell$.

Nominal equivalence: only if $\Psi = \Omega^{-1}$.

Regularized declipping problem

Sparse synthesis

$$\begin{aligned} \text{minimize } & {}_{\alpha}\|\alpha\|_0 \\ \text{subject to } & \mathsf{M}_\mathsf{r}\,\Psi\alpha = \mathsf{M}_\mathsf{r}\,\bar{x} \\ & \mathsf{M}_\mathsf{c}^{+}\,\Psi\alpha \geq \mathsf{M}_\mathsf{c}^{+}\,\bar{x} \\ & \mathsf{M}_\mathsf{c}^{-}\,\Psi\alpha \leq \mathsf{M}_\mathsf{c}^{-}\,\bar{x} \end{aligned}$$

Sparse analysis

$$\begin{aligned} \text{minimize } & {}_{x}\|\Omega x\|_0 \\ \text{subject to } & \mathsf{M}_\mathsf{r}\,x = \mathsf{M}_\mathsf{r}\,\bar{x} \\ & \mathsf{M}_\mathsf{c}^{+}\,x \geq \mathsf{M}_\mathsf{c}^{+}\,\bar{x} \\ & \mathsf{M}_\mathsf{c}^{-}\,x \leq \mathsf{M}_\mathsf{c}^{-}\,\bar{x} \end{aligned}$$

Ω and Ψ are some *overcomplete* transform matrices known for compacting the energy of audio signals.

Choice of the dictionary and the operator

- Modulated complex lapped transform (MCLT) [DD06]:

$$\Psi = \begin{bmatrix} \psi_0^c & \psi_1^c & \dots & \psi_{m/2-1}^c & \psi_0^s & \psi_1^s & \dots & \psi_{m/2-1}^s \end{bmatrix},$$

$$\psi_j^c(t) = \cos\left(\frac{\pi}{m}(t+1/2)(j+1/2)\right), \psi_j^s(t) = \sin\left(\frac{\pi}{m}(t+1/2)(j+1/2)\right),$$

  where $t = [0, n-1]$ and $j = [0, m-1]$.

- Two-times redundant ($m = 2n$) DCT-DST dictionary.
- The atoms are chosen according to recommendations in [Gri01] to use the transform as Gabor-like dictionary.
- The analysis operator is the transpose $\Omega = \Psi^T$, ($p = m$), for consistency.

Computational perspective

Minimizing either $\|\alpha\|_0$ or $\|\Omega x\|_0$ is NP-hard!

Sparse synthesis

- Convex relaxation: minimize $\|\alpha\|_1$ or some other convex objective, if applicable.
- Greedy: MP, OMP, IHT, HTP etc.

Sparse analysis

- Convex relaxation: minimize $\|\Omega x\|_1$ or some other convex objective, if applicable.
- Greedy: GAP, analysis IHT/HTP etc.

Important: model is rarely a perfect reflection of reality
(assume "$\approx$" rather than "$=$")!

4
DECLIPPING ALGORITHMS
based on sparse and cosparse prior

Constrained Matching Pursuit for Audio Declipping

- Two-stage algorithm [AEJ+12]:
  1. Orthogonal Matching Pursuit for CS:
     - Initialize the support $\Lambda = \{\emptyset\}$ and residual $r^{(0)} = \mathrm{M_r}\,\bar{x}$.
     - Select atom: $j = \arg\max_j \langle r^{k-1}, \psi_j \rangle$,
     - Update support: $\Lambda \leftarrow \Lambda \cup j$; $\Psi_\Lambda = \left[\psi_i\right], \{\psi_i \in \Psi \mid i \in \Lambda\}$,
     - Compute the estimate: $\alpha^{(k)} = \arg\min_\alpha \| \mathrm{M_r}\,\bar{x} - \mathrm{M_r}\,\Psi_\Lambda \alpha \|_2^2$,
     - Compute new residual: $r^{(k)} = \mathrm{M_r}\,\bar{x} - \mathrm{M_r}\,\Psi_\Lambda \alpha^{(k)}$,
     - Termination criterion: $\|r^{(k)}\|_2 \leq \epsilon$.
  2. Refinement by clipping constraints:

$$\begin{aligned} \hat{\alpha} = \arg\min_\alpha & \| \mathrm{M_r}\,\bar{x} - \mathrm{M_r}\,\Psi_\Lambda \alpha \|_2^2 \\ \text{subject to } & \mathrm{M_c^+}\,\Psi_\Lambda \alpha \geq \mathrm{M_c^+}\,\bar{x} \\ & \mathrm{M_c^-}\,\Psi_\Lambda \alpha \leq \mathrm{M_c^-}\,\bar{x} \end{aligned}$$

- Final estimate: $\hat{x} = \Psi_\Lambda \hat{\alpha}$.

Consistent Iterative Hard Thresholding

- Algorithm [KJM+13] based on IHT by Blumensath et al. Objective:

  $\min_\alpha \| \mathrm{M_r}\,\Psi\alpha - \mathrm{M_r}\,\bar{x}\|_2^2 + \| \left(\mathrm{M_c^+}\,\bar{x} - \mathrm{M_c^+}\,\Psi\alpha\right)_+ \|_2^2 + \| \left(\mathrm{M_c^-}\,\bar{x} - \mathrm{M_c^-}\,\Psi\alpha\right)_- \|_2^2$

  subject to α being sparse and $(u_i)_\pm = \pm\max(0, \pm u_i)$.
- Define:

$$\mathcal{B}(u_i) = \begin{cases} u_i & \forall i \in S_r, \\ (u_i)_+ & \forall i \in S_p, \\ (u_i)_- & \forall i \in S_n. \end{cases}$$

- Iterative update: $\alpha^{(k+1)} = \mathcal{H}_{k+1}\left(\alpha^{(k)} + \mu\Psi^T\mathcal{B}(\bar{x} - \Psi\alpha^{(k)})\right)$.
- $\mathcal{H}_K(\cdot)$ is the hard-thresholding operator.
- $K \leftarrow k+1$: sparsity relaxed per iteration.
- Stepsize μ estimated through line-search.
- Termination criterion: $\|r^{(k)}\|_2 = \|\mathcal{B}(\bar{x} - \Psi\alpha^{(k)})\|_2 \leq \epsilon$.

Analysis Hard Thresholding for Audio Declipping

- Ideas from Consistent IHT cannot be readily applied, since:

  $$\text{minimize}_x \|y - x\|_2^2 \text{ subject to } \|\Omega x\|_0 \leq k \text{ is NP-hard [TGP14]!}$$

- Instead, we enforce *approximate* cosparsity through ADMM approach [KBG14].
- Reformulate the problem by splitting variables ($z \in \mathbb{R}^p$):

$$\begin{aligned} \text{minimize}_{\ z,x} & \|\Omega x - z\|_2^2 \\ \text{subject to } & \|z\|_0 \ll p, \\ & \mathrm{M_r}\, x = M_r \bar{x} \\ & \mathrm{M_c^+}\, x \geq \mathrm{M_c^+}\, \bar{x} \\ & \mathrm{M_c^-}\, x \leq \mathrm{M_c^-}\, \bar{x}. \end{aligned}$$

Analysis Hard Thresholding for Audio Declipping

1. Initialize: $x^{(0)} = \bar{x}, u^{(0)} = 0$.
2. $z^{(k)} = \arg\min_z \|z - \Omega x^{(k-1)} - u^{(k-1)}\|_2^2$ s. t. $\|z\|_0 \leq k = \mathcal{H}_k\left(\Omega x^{(k-1)} + u^{(k-1)}\right)$.
3. x-update:

$$x^{(k)} = \arg\min_x \|\Omega x - z^{(k)} + u^{(k-1)}\|_2^2$$
$$\text{subject to } \mathrm{M_r}\, x = \mathrm{M_r}\, \bar{x}$$
$$\mathrm{M_c^+}\, x \geq \mathrm{M_c^+}\, \bar{x}$$
$$\mathrm{M_c^-}\, x \leq \mathrm{M_c^-}\, \bar{x}$$

4. Lagrangian variable update: $u^{(k)} = u^{(k-1)} + \Omega x^{(k)} - z^{(k)}$.
5. Termination criterion: $\|r^{(k)}\|_\infty = \|\Omega x^{(k)} - z^{(k)} + u^{(k)}\|_\infty \leq \epsilon$.

Analysis Hard Thresholding for Audio Declipping

- Computing the exact x-update is expensive!
- Instead we first solve for the equality constraints only:

$$\begin{aligned}\text{Let: } \hat{x}^{(k)} &= (\mathrm{I} - \mathrm{M_r}^{\dagger}\,\mathrm{M_r})x_{\text{null}} + \mathrm{M_r}^{\dagger}\,\mathrm{M_r}\,\bar{x} = \Pi x_{\text{null}} + \mathrm{M_r}^{\dagger}\,\mathrm{M_r}\,\bar{x}.\\ \text{Solve: } x_{\text{null}} &= \arg\min_x \|\Omega(\Pi x + \mathrm{M_r}^{\dagger}\,\mathrm{M_r}\,\bar{x}) - z^{(k)} + u^{(k-1)}\|_2^2.\\ &= \arg\min_x \|\tilde{\Omega}_{\bar{r}} x - q\|_2^2.\end{aligned}$$

- Then we project the solution to clipping (box) constraints:

$$x^{(k)} = \hat{x}^{(k)} + \mathcal{B}(\bar{x} - \hat{x}^{(k)})$$

.

- Suboptimal, but sufficient for the convergence.

Linear prediction declipping

- Adaptation of the interpolation method proposed by Janssen [JVV86].
- The signals are modeled as autoregressive (AR) processes of finite order $r = 3c + 2$, where c is number of clipped samples.
- The objective is "whitening" the signal ($a \in \mathbb{R}^r$ are the filter coefficients, $a_0 = -1$):

$$Q(a, x) = \sum_{i=r}^{n-1} \left(\sum_{j=0}^{r} a_j x_{i-j} \right)^2 = \sum_{i=r}^{n-1} e_i^2$$

- Vectors a and x are estimated by alternating minimization of $Q(a, x)$ and projecting the estimate x to clipping constraints.
- Potential downfalls: sensitive to initialization, computational time proportional to filter order.

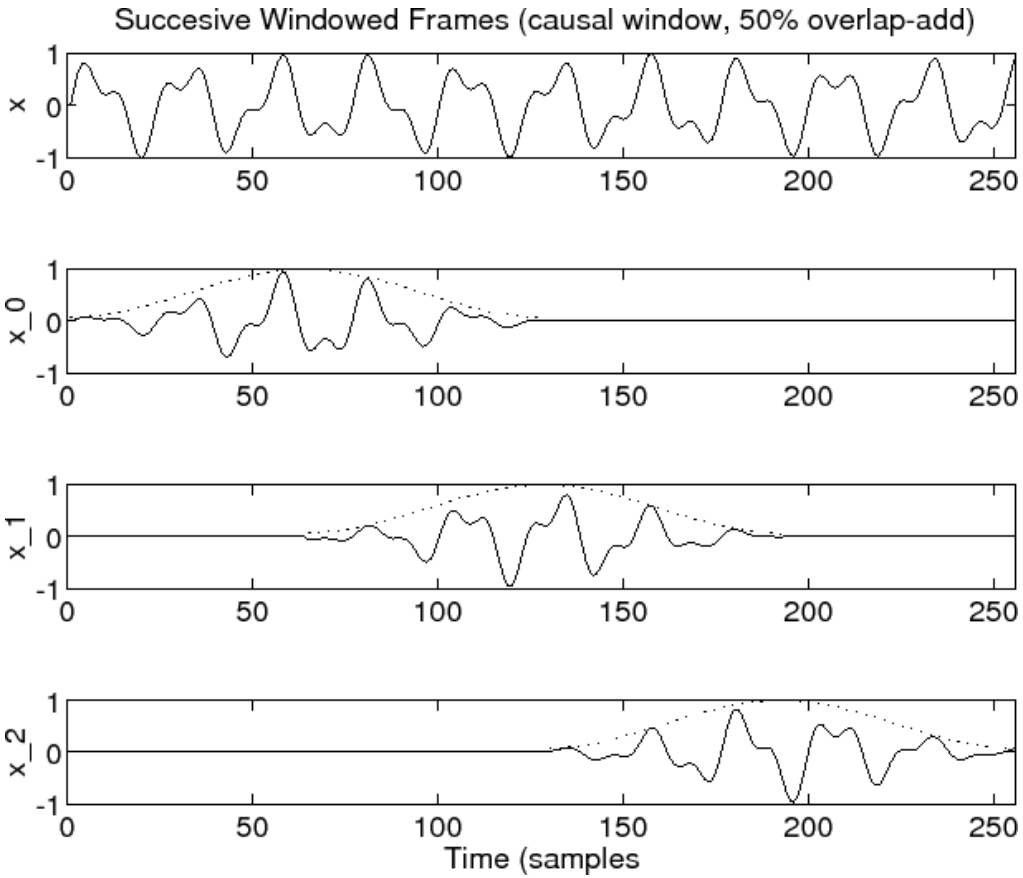
Conceptual analysis

- Constrained MP: support is chosen without clipped observations in $\bar{x}$!
- Constrained MP and Consistent IHT: sensitive trade-off between good fit and overfitting.
- Analysis HT and Constrained MP: potentially slow due to intermediate constrained minimization steps.
- AR declipping will be slow for severely clipped signals.
- Efficant computation of $\Psi(\cdot)$, $\Psi^T(\cdot)$, $\Omega(\cdot)$ and $\Omega^T(\cdot)$ is highly recommended!
- All algorithms are non-convex heuristics and only locally optimal.

5
AUDIO BENCHMARKS
Declipping wideband audio data

Frame-based processing

- Constant Overlap-Add (COLA) scheme.
- Overlap stepsize: 75%.
- Weighting function: Hamming (square rooted for the analysis and synthesis window).
- Frame duration: ∼20ms.



Benchmark

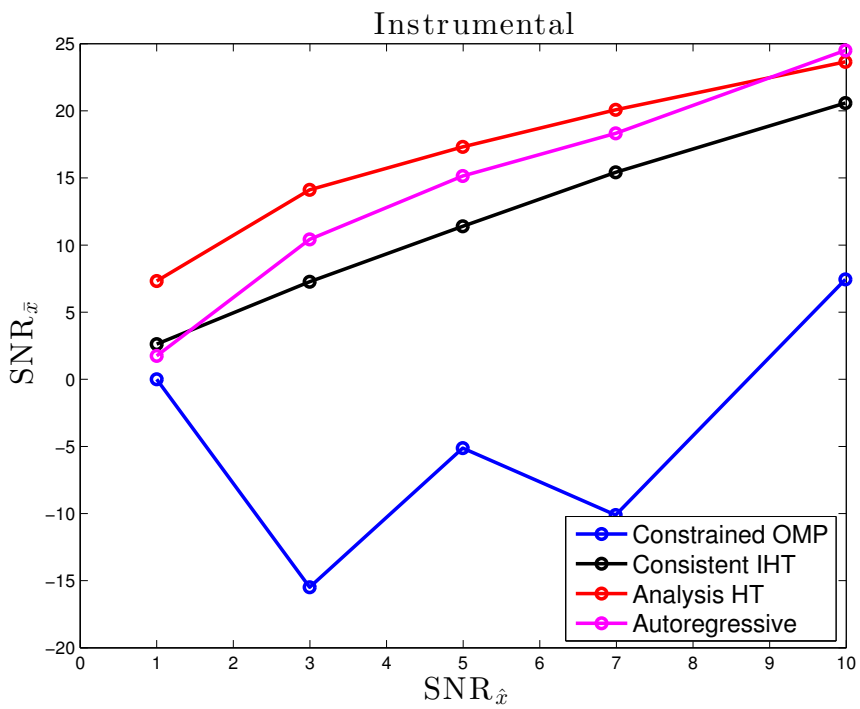
- Two wideband audio tracks (sampling: 16kHz, encoding: 16bit).
- DCT and Gabor for the dictionary/operator.
- Performance criterion: SNR difference between the input and post-processed data:

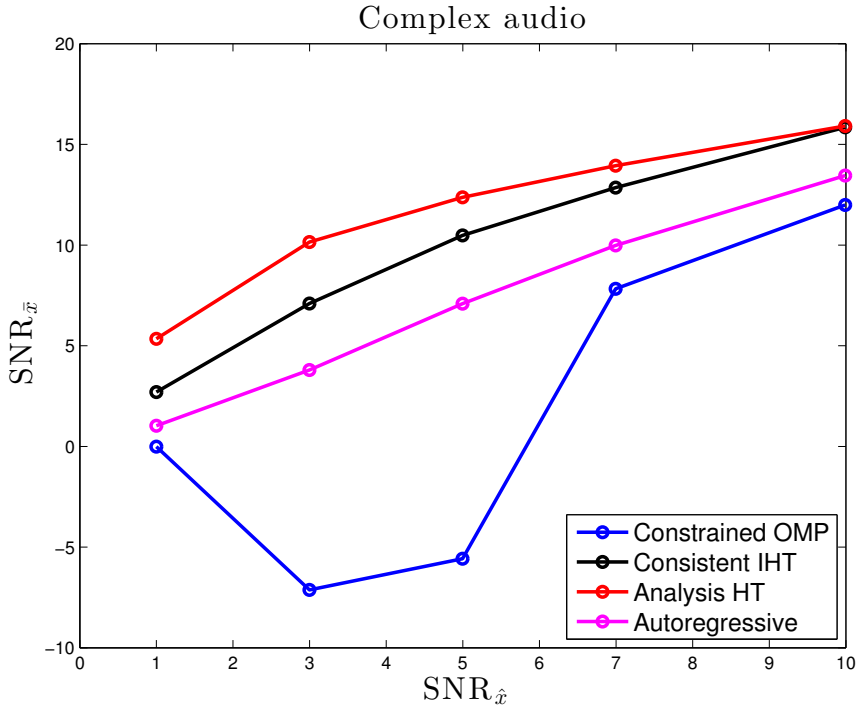
$$\mathrm{SNR}_{\bar{x}} = 20 \log_{10} \frac{\|x\|_2}{\|x - \bar{x}\|_2}$$

$$\mathrm{SNR}_{\hat{x}} = 20 \log_{10} \frac{\|x\|_2}{\|x - \hat{x}\|_2}$$

- Clipping range: from $\mathrm{SNR}_{\bar{x}} = 1\mathrm{dB}$ to $\mathrm{SNR}_{\bar{x}} = 10\mathrm{dB}$.

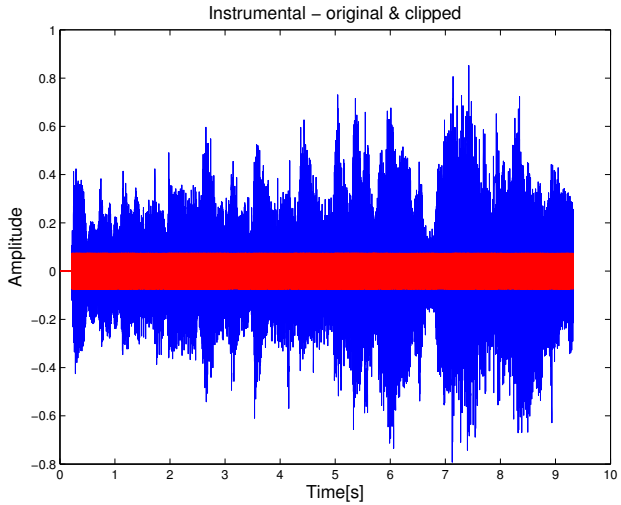
Results - recovery performance

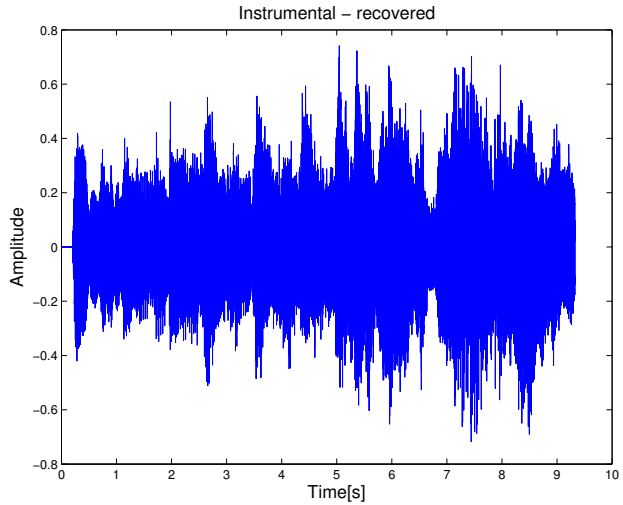


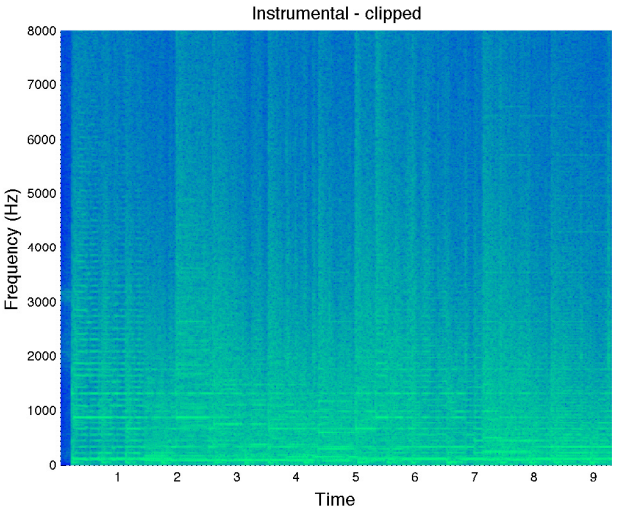


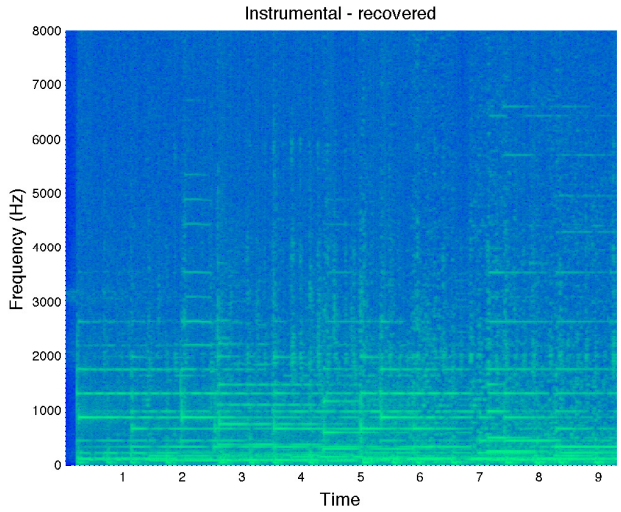
Output *vs* input SNR for the benchmarked algorithms.

Results - audio preview

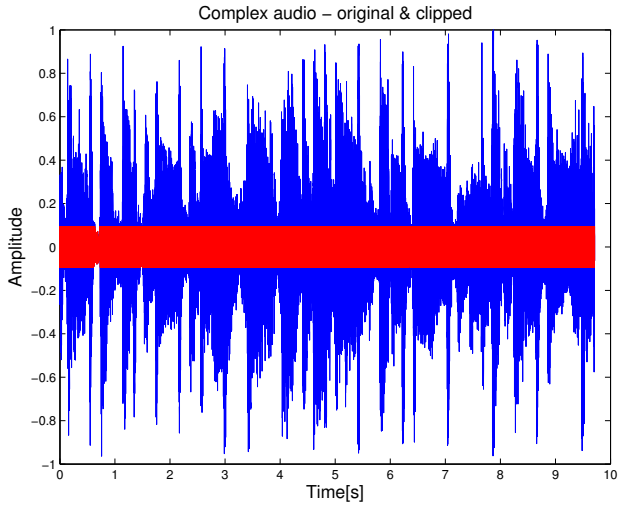


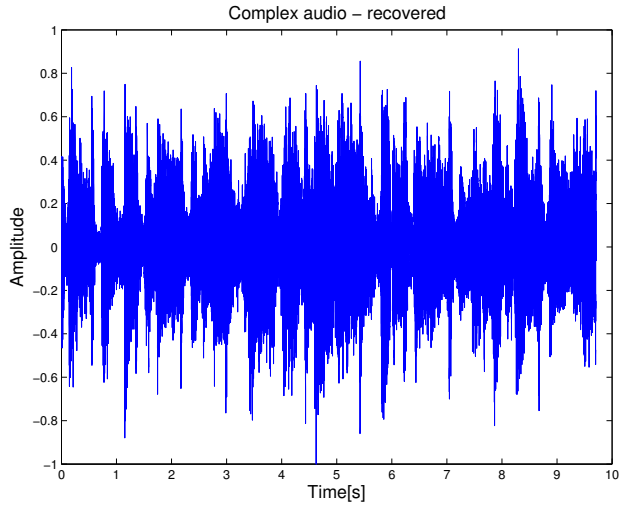


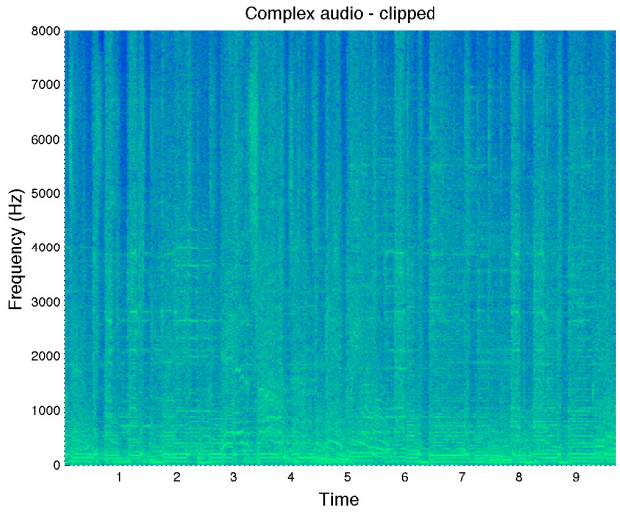


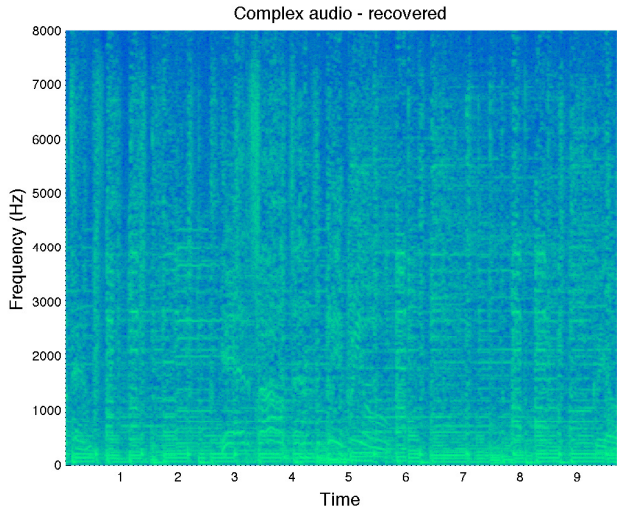


Results - audio preview





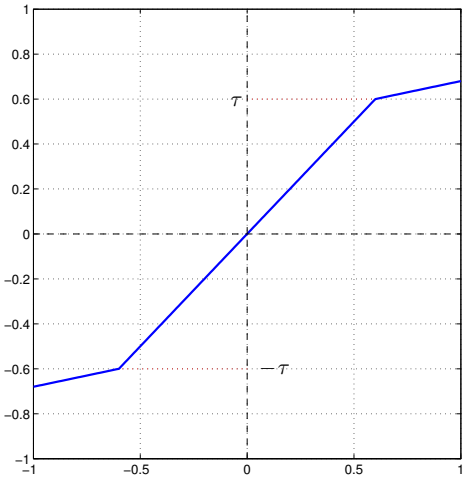




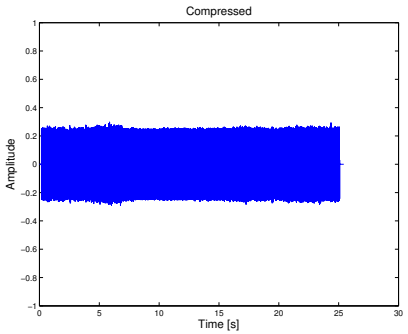
Blind decompression

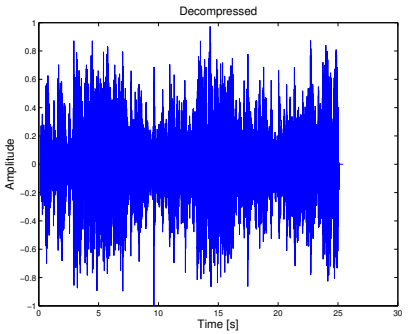
- More realistic scenario [MGS03]: data is not "perfectly" clipped; instead, high magnitudes are gradually compressed.
- The threshold τ and compression coefficient γ are unknown.

$$\bar{x}_i = \begin{cases} x_i & \text{if } |x_i| < \tau, \\ \operatorname{sgn}(x_i)\tau(1-\gamma) + \gamma x_i & \text{otherwise.} \end{cases}$$



- Arbitrary declipping is not entirely satisfactory:



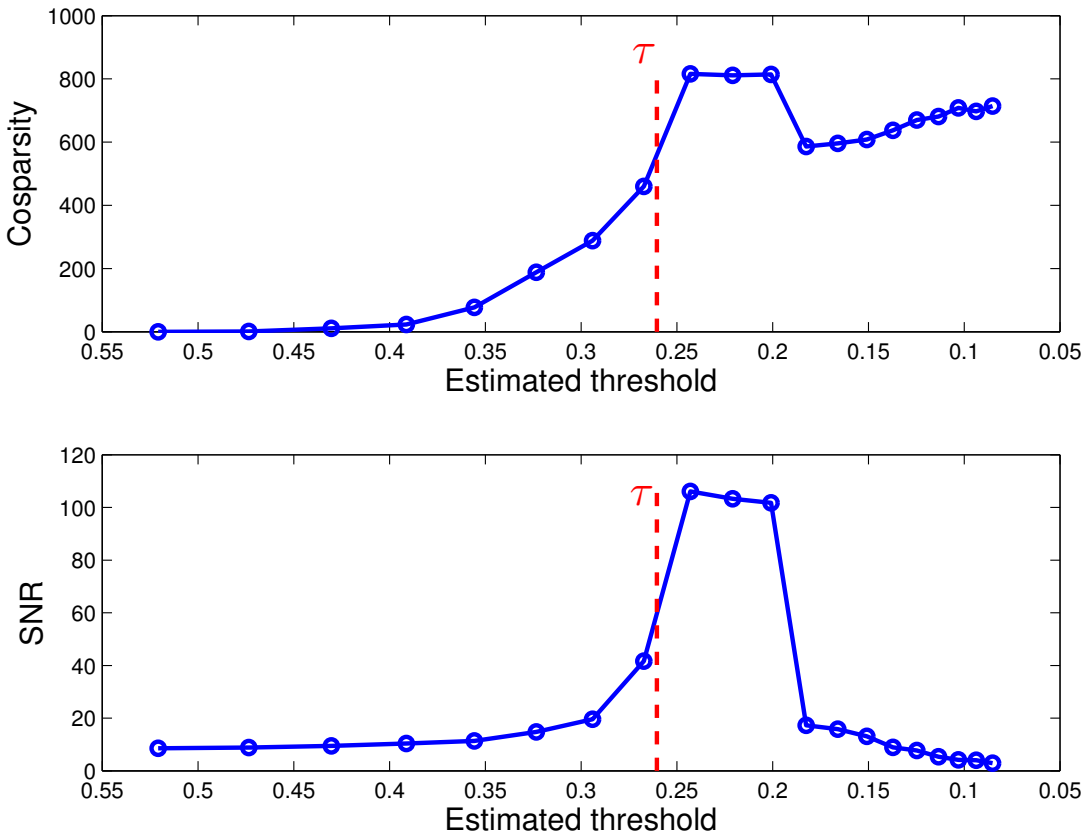


Blind decompression

More systematic approach:

1. Initialization: $\hat{\tau} = \|\bar{x}\|_\infty$.
2. Backtracking: $\hat{\tau} \leftarrow \frac{1}{b}\hat{\tau}, b > 1$.
3. Define the measurement matrices $\mathsf{M}_\mathsf{r}(\hat{\tau}), \mathsf{M}_\mathsf{c}^+(\hat{\tau}), \mathsf{M}_\mathsf{c}^-(\hat{\tau})$, assuming that samples $|\bar{x}_i| < \hat{\tau}$ are reliable.
4. Compute the estimate $\hat{x}^{(k)}$ by analysis HT declipping.
5. Evaluate approximate cosparsity of the estimate: $\hat{\ell}^{(k)} = p - \|\hat{z}^{(k)}\|_0$.
6. Stopping criterion based on the cosparsity decrease: $\hat{\ell}^{(k+1)} \leq \hat{\ell}^{(k)}$.

Blind decompression - results



Cosparsity of the estimate and decompressed SNR *vs* estimated threshold for the ℓ-cosparse signal x.

- Exhibits a "phase transition" behavior.
- Crude scheme - works, but computationally expensive.
- Impact of τ and γ on the performance?
- Performance on the real audio data?

Conclusions

- Analysis Hard Thresholding outperforms all the others for most of the given clipping range.
- Consistent IHT and AR declipping offer good trade-off between processing time and quality of reconstruction.
- Constrained OMP fails to recover severely clipped signals due to inaccurate support estimation.
- Blind decompression / declipping seems possible.
- Envisioned improvements: enforcing structure in sparse and cosparse estimation (some existing approaches - check [SKD14]).
- Coupling AR model with sparse/cosparse regularization?

MERCI

Inria Rennes

PANAMA team

team.inria.fr/panama

References I

[AEJ+12] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley.
Audio inpainting.
Audio, Speech, and Language Processing, IEEE Transactions on, 20(3):922–932, 2012.

[CJ10] P. Comon and C. Jutten.
Handbook of Blind Source Separation: Independent component analysis and applications.
Academic press, 2010.

[Czy97] A. Czyzewski.
Learning algorithms for audio signal enhancement, part 1: Neural network implementation for the removal of impulse distortions.
Journal of the Audio Engineering Society, 45(10):815–831, 1997.

[DD06] M. E. Davies and L. Daudet.
Sparse audio representations using the mclt.
Signal processing, 86(3):457–470, 2006.

[GR95] S. J. Godsill and P. J. W. Rayner.
A bayesian approach to the restoration of degraded audio signals.
Speech and Audio Processing, IEEE Transactions on, 3(4):267–278, 1995.

[Gri01] R. Gribonval.
Fast matching pursuit with a multiscale dictionary of gaussian chirps.
Signal Processing, IEEE Transactions on, 49(5):994–1001, 2001.

[JVV86] A. Janssen, R. Veldhuis, and L. Vries.
Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes.
Acoustics, Speech and Signal Processing, IEEE Transactions on, 34(2):317–330, 1986.

References II

[KBG14] S. Kitić, N. Bertin, and R. Gribonval.
Audio declipping by cosparse hard thresholding.
In *iTwist '14 - international Traveling Workshop on Interactions between Sparse models and Technology*, 2014.

[KJM+13] S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. De Vleeschouwer.
Consistent iterative hard thresholding for signal declipping.
In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5939–5943. IEEE, 2013.

[LA05] E. Larsen and R. M. Aarts.
Audio bandwidth extension: application of psychoacoustics, signal processing and loudspeaker design.
John Wiley & Sons, 2005.

[MG14] J. Murphy and S. Godsill.
Structured sparse bayesian modelling for audio restoration.
In *Compressed Sensing & Sparse Filtering*, pages 423–453. Springer, 2014.

[MGS03] B. C. J. Moore, B. R. Glasberg, and M. A. Stone.
Why are commercials so loud? perception and modeling of the loudness of amplitude-compressed speech.
Journal of the Audio Engineering Society, 51(12):1123–1132, 2003.

[NDEG13] S. Nam, M. E. Davies, M. Elad, and R. Gribonval.
The Cosparse Analysis Model and Algorithms.
Applied and Computational Harmonic Analysis, 34(1):30–56, 2013.

[PBD+10] M. D. Plumbley, T. Blumensath, L. Daudet, R. Gribonval, and M. E. Davies.
Sparse representations in audio and music: from coding to source separation.
Proceedings of the IEEE, 98(6):995–1005, 2010.

References III

[SKD14] K. Siedenburg, M. Kowalski, and M. Dörfler.
Audio declipping with social sparsity.
Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014.

[SLOVB14] N. Souviraà-Labastie, A. Olivero, E. Vincent, and F. Bimbot.
Audio source separation using multiple deformed references.
In *European Signal Processing Conference (EUSIPCO)*, 2014.
Submitted.

[TGP14] A. M. Tillmann, R. Gribonval, and M. E. Pfetsch.
Projection Onto The k-Cosparse Set is NP-Hard.
Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014.

[Unc03] A. Uncini.
Audio signal processing by neural networks.
Neurocomputing, 55(3):593–625, 2003.

[WSL00] B. Wah, X. Su, and D. Lin.
A survey of error-concealment schemes for real-time audio and video transmissions over the internet.
In *Multimedia Software Engineering, 2000. Proceedings. International Symposium on*, pages 17–24.
IEEE, 2000.