



HAL
open science

Audio Declipping by Cosparse Hard Thresholding

Srđan Kitić, Nancy Bertin, Rémi Gribonval

► **To cite this version:**

Srđan Kitić, Nancy Bertin, Rémi Gribonval. Audio Declipping by Cosparse Hard Thresholding. .
hal-00922497v2

HAL Id: hal-00922497

<https://inria.hal.science/hal-00922497v2>

Submitted on 13 Jan 2014 (v2), last revised 1 Mar 2014 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AUDIO DECLIPPING BY COSPARSE HARD THRESHOLDING

S. Kitić[†], N. Bertin^{*}, R. Gribonval[†]

[†]Inria , ^{*}IRISA - CNRS UMR 6074

ABSTRACT

Recovery of clipped wideband audio signals is a very challenging inverse problem. Recently, it has been successfully addressed by several methods based on the sparse synthesis data model. In this work we propose an algorithm for enhancement of clipped wideband audio signals that exploits the sparse analysis (*cosparse*) data model. Experiments on real audio data indicate that the algorithm has better signal restoration performance than state-of-the-art sparse synthesis declipping methods.

Index Terms— clipping, audio, sparse, cosparse, inverse problems

1. INTRODUCTION

Audio signals are prone to various degradations and clipping is among the most common ones. It may arise during recording due to poor dynamic range of the microphone or analog-to-digital circuits, but it may also be seen as the extreme case of magnitude compression [1].

Clipping inevitably causes an information loss: in the time domain the signal’s amplitude above some threshold is being “cut-off”, which manifests as the expansion (spread) in the frequency domain. The intuitive approach to de-clip the signal may be to “clean” the signal’s frequency spectrum and preserve only those frequencies needed to represent the original signal. However, the estimation of the correct support in the frequency domain is also a difficult problem, hence some prior information about the underlying signal is necessary.

We formalize clipping as the following element-wise operation on the original signal $\mathbf{x} \in \mathbb{R}^n$ (yielding a clipped signal $\bar{\mathbf{x}}$):

$$\bar{\mathbf{x}} = \text{sign}(\mathbf{x}) \min(|\mathbf{x}|, \tau) \quad (1)$$

The threshold τ is the amplitude saturation level (here, we assume *symmetric* clipping, *i.e.* the same threshold is applied to both positive and negative samples of \mathbf{x}).

We can then introduce the following sampling operators: $\mathcal{M}_r(\bar{\mathbf{x}})$ (which extracts the unclipped, “reliable”, samples), $\mathcal{M}_c^+(\bar{\mathbf{x}})$ and $\mathcal{M}_c^-(\bar{\mathbf{x}})$ (which extract the clipped samples with

positive and negative amplitude, respectively). Knowing that the amplitude of the original signal has to be higher (respectively, lower) than the threshold for the clipped samples, we can formulate the inverse problem of signal de-clipping in the following way:

$$\begin{aligned} \text{Find } \hat{\mathbf{x}} \in \mathbb{R}^n \text{ such that: } & \mathcal{M}_r(\hat{\mathbf{x}}) = \mathcal{M}_r(\bar{\mathbf{x}}) \\ & \mathcal{M}_c^+(\hat{\mathbf{x}}) \geq \mathcal{M}_c^+(\bar{\mathbf{x}}) \\ & \mathcal{M}_c^-(\hat{\mathbf{x}}) \leq \mathcal{M}_c^-(\bar{\mathbf{x}}) \end{aligned} \quad (2)$$

This is an ill-posed inverse problem, since there are infinitely many possible solutions for $\hat{\mathbf{x}}$. Therefore, we choose to regularize the problem by introducing *the sparsity hypothesis*. The linear constraints in (2) will henceforth be shortly termed as the constraint set Γ .

2. SPARSE ANALYSIS DATA MODEL

The premise of sparse data models is that many signals have low-dimensional representation, but not necessarily in their original domain. This assumption is sometimes sufficient to mitigate the ill-posedness of the original problem.

Sparse analysis, also known as *cosparse* data model assumes that the signal in question can be “sparsified” by applying an adequate linear transform, here represented as a matrix operator $\Omega \in \mathbb{R}^{p \times n}$, $p \geq n$. Keeping the clipping constraints Γ , we have the following optimization problem:

$$\text{minimize}_{\hat{\mathbf{x}}} \|\Omega \hat{\mathbf{x}}\|_0 \text{ s.t. } \hat{\mathbf{x}} \text{ respects } \Gamma \quad (3)$$

Minimizing the ℓ_0 -norm in the cosparse case is NP-hard, but there are convex relaxation and greedy algorithms [2] which may be used to approximate its solution.

The cosparse data model only recently attracted the attention of scientific community, and to our best knowledge, it has not yet been exploited to solve the declipping problem.

3. COSPARSE DECLIPPING

Our approach is based on transposing the *Consistent Iterative Hard Thresholding* (Consistent IHT) algorithm [3] to the cosparse data model. Consistent IHT relies on applying hard

This work was supported in part by the European Research Council, PLEASE project (ERC-StG-2011-277906).

[†]Centre Inria Rennes - Bretagne Atlantique, France

^{*}Rennes, France

thresholding $H_K(\mathbf{s})$ ¹ to transform coefficient vector \mathbf{s} in order to approximate sparse solutions. The transform coefficients are estimated by performing a gradient descent step of an objective that incorporates the constraints expressed in Γ .

The pragmatical issue is that computing signal's k -cosparse projection is proven to be NP-hard [4]. To circumvent this problem, we take a different approach, largely based on the *Alternating Direction Method of Multipliers* [5] framework. The resulting algorithm is presented as pseudocode in algorithm 1 and will be referred to as Cosparse Declipping by Hard Thresholding (CoDec-HT).

Algorithm 1 COsparse DEClipping by Hard Thresholding

Require: $\Omega, \bar{\mathbf{x}}, \Gamma, \varepsilon$

- 1: $\hat{\mathbf{x}}^{(0)} = \bar{\mathbf{x}}, \mathbf{u}^{(0)} = \mathbf{0}, k = 1$
- 2: $\hat{\mathbf{z}}^{(k)} = H_k(\Omega\hat{\mathbf{x}}^{(k-1)} + \mathbf{u}^{(k-1)})$
- 3: $\hat{\mathbf{x}}^{(k)} = \arg \min_{\tilde{\mathbf{x}}} \|\Omega\tilde{\mathbf{x}} - \hat{\mathbf{z}}^{(k)} + \mathbf{u}^{(k-1)}\|_2^2$ s.t. $\tilde{\mathbf{x}}$ respects Γ
- 4: **if** $\|\Omega\hat{\mathbf{x}}^{(k)} - \hat{\mathbf{z}}^{(k)} + \mathbf{u}^{(k-1)}\|_\infty \leq \varepsilon$ **then**
- 5: terminate
- 6: **else**
- 7: $\mathbf{u}^{(k)} = \mathbf{u}^{(k-1)} + \Omega\hat{\mathbf{x}}^{(k)} - \hat{\mathbf{z}}^{(k)}$
- 8: $k \leftarrow k + 1$
- 9: go to 2
- 10: **end if**
- 11: **return** $\hat{\mathbf{x}}^{(k)}$

Mimicking the heuristics used in Consistent IHT, we relax the sparsity k of \mathbf{z} through iterations, which allows to “learn” the unknown sparsity level of a given signal. Through iterations, the vectors $\Omega\hat{\mathbf{x}}$ and $\hat{\mathbf{z}}$ eventually get close to each other, thus the signal estimate $\hat{\mathbf{x}}$ becomes approximately cosparse.

The analysis operator Ω can be chosen as any (possibly overcomplete) transform which sparsifies the audio signal. In our experiments, we used a concatenation of Discrete Cosine Transform and Discrete Sine Transform matrices.

4. EXPERIMENTS

We carried declipping experiments with 10-second excerpts from two wideband music pieces, sampled at 16kHz with 16bit encoding. Audio 1 is a rock piece with vocals, drums, two electric guitars and bass, while Audio 2 is an instrumental piece performed on an analogue synthesizer. The signal is divided into 1024-sample long overlapping frames, which are, after processing, re-synthesized by an overlap-add scheme.

Before processing, the audio is clipped given a predefined SNR_{inp} (signal-to-noise ratio of the input) value. The restoration performance is measured as the output signal-to-noise ra-

¹ $H_K(\mathbf{s})$ keeps K highest in absolute magnitude elements of \mathbf{s} and sets the remaining to zero.

tio (SNR_{out}):

$$\text{SNR}_{\text{inp}} = 20 \log_{10} \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \bar{\mathbf{x}}\|_2}, \text{SNR}_{\text{out}} = 20 \log_{10} \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2} \quad (4)$$

The algorithm is compared against [3] and [6]², which serve as a reference. Consistent IHT (with dictionary $\Psi = \Omega^H$) and the proposed algorithm are based on a similar algorithmic methodology and differ only by the underlying data model. Algorithm [6] exploits a structured prior termed *social sparsity* [7].

The results presented in Figure 1 suggest that cosparsity-based hard thresholding outperforms synthesis-based one for all clipping levels. Interestingly, the SNR improvement rate of CoDec HT is roughly of the same order as for the social sparsity based algorithm, despite the fact that the latter uses a stronger prior.

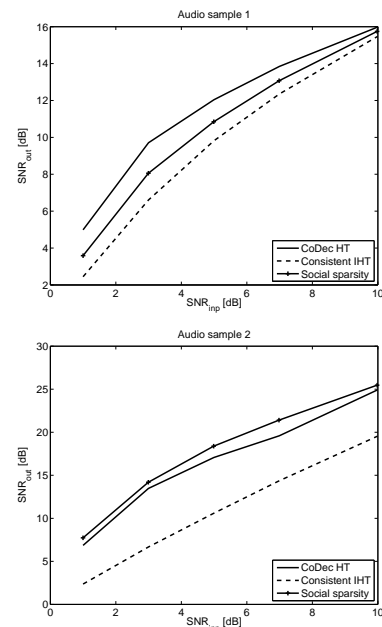


Fig. 1: Declipping results for two sample wideband audio tracks

5. CONCLUSION

Experimental results indicate that sparse analysis regularization can be successfully applied to the audio declipping inverse problem. Future work will aim at incorporating structured cosparsity in the data model to improve declipping performance, and at addressing extended magnitude corruption scenarios, such as dynamic range compression or soft clipping.

²We acknowledge M.Kowalski who provided his original code.

6. REFERENCES

- [1] J. O. Smith, *Physical audio signal processing: For virtual musical instruments and audio effects*, W3K Publishing, 2010.
- [2] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, “The Cospase Analysis Model and Algorithms,” *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.
- [3] 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*, 2013.
- [4] A. M. Tillmann, R. Gribonval, and M. E. Pfetsch, “Projection onto the k-cospase set is NP-hard,” *CoRR*, vol. abs/1303.5305, 2013.
- [5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [6] K. Siedenburg, M. Kowalski, and M. Dörfler, “Audio declipping with social sparsity,” *Submitted to ICASSP, 2014*, 2013.
- [7] M. Kowalski, K. Siedenburg, and M. Dörfler, “Social sparsity! neighborhood systems enrich structured shrinkage operators,” 2012.