A Proper Version of Synthesis-Based Sparse Audio Declipper

Pavel Záviška*, Pavel Rajmic*, Ondřej Mokrý* and Zdeněk Průša†

*Signal Processing Laboratory, Brno University of Technology, Brno, Czech Republic
†Acoustic Research Institute, Austrian Academy of Science, Vienna, Austria

Objectives

The state of the art within the sparsity based approaches to audio declipping has been achieved by the SPADE algorithm by Kitić et. al. The algorithm comes in synthesis and analysis variant, but the respective problems solved by the two variants do not correspond. We propose a new synthesis variant that outperforms the old one in terms of both the restoration quality and speed.

Introduction

Hard clipping as one of the most common audio signal degradations is modeled as

$$\mathbf{y}[n] = \begin{cases} \mathbf{x}[n] & \text{for } |\mathbf{x}[n]| < \theta_c, \\ \theta_c \cdot \operatorname{sgn}(\mathbf{x}[n]) & \text{for } |\mathbf{x}[n]| \geq \theta_c, \end{cases} \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^N$ is the original signal, $\mathbf{y} \in \mathbb{R}^N$ denotes the clipped signal and θ_c represents the symmetrical clipping threshold. The goal of declipping is to find a suitable signal from the set Γ defined as

$$\{\tilde{\mathbf{x}} \mid M_R\tilde{\mathbf{x}} = M_R\mathbf{y},\ M_H\tilde{\mathbf{x}} \geq \theta_c,\ M_L\tilde{\mathbf{x}} \leq -\theta_c\}. \quad (2)$$

An effective approach is to exploit the sparsity. One thus seeks the "sparsest" signal $\hat{\mathbf{x}}$ from the set Γ.

SPADE

SPADE algorithms approximate the solution of the non-convex inverse problems

$$\arg\min_{\mathbf{x},\mathbf{z}} \|\mathbf{z}\|_0 \text{ s.t. } \mathbf{x} \in \Gamma,\ \|A\mathbf{x} - \mathbf{z}\|_2 \leq \epsilon, \quad (3a)$$

$$\arg\min_{\mathbf{x},\mathbf{z}} \|\mathbf{z}\|_0 \text{ s.t. } \mathbf{x} \in \Gamma,\ \|\mathbf{x} - D\mathbf{z}\|_2 \leq \epsilon, \quad (3b)$$

where $\mathbf{z} \in \mathbb{C}^P$ are the signal coefficients with $A\colon \mathbb{R}^N \to \mathbb{C}^P$ and $D\colon \mathbb{C}^P \to \mathbb{R}^N$ being the analysis and synthesis operators respectively, where $D = A^*$ and $DD^* = A^*A = \mathrm{Id}$. Using the indicator functions, the problems can be recast as

$$\arg\min_{\mathbf{x},\mathbf{z}} \iota_\Gamma(\mathbf{x}) + \iota_{\ell_0 \leq k}(\mathbf{z}) \text{ s.t. } \begin{cases} \|A\mathbf{x} - \mathbf{z}\|_2 \leq \epsilon, \\ \|\mathbf{x} - D\mathbf{z}\|_2 \leq \epsilon, \end{cases} \quad (4)$$

for sufficiently small sparsity k. The solution to (4) is then approximated using the ADMM.

Proposed S-SPADE scheme

The ADMM solves the given problem by splitting the minimization of the Augmented Lagrangian, which is formed for the synthesis variant of (4) as

$$L_\rho = \iota_{\ell_0 \leq k}(\mathbf{z}) + \iota_\Gamma(\mathbf{x}) + \frac{\rho}{2}\|D\mathbf{z} - \mathbf{x} + \mathbf{u}\|_2^2 - \frac{\rho}{2}\|\mathbf{u}\|_2^2. \quad (5)$$

The minimization steps to be iterated are then

$$\mathbf{z}^{(i+1)} = \arg\min_{\|\mathbf{z}\|_0 \leq k} \|D\mathbf{z} - \mathbf{x}^{(i)} + \mathbf{u}^{(i)}\|_2^2, \quad (6a)$$

$$\mathbf{x}^{(i+1)} = \arg\min_{\mathbf{x} \in \Gamma} \|D\mathbf{z}^{(i+1)} - \mathbf{x} + \mathbf{u}^{(i)}\|_2^2, \quad (6b)$$

together with the dual variable update

$$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + D\mathbf{z}^{(i+1)} - \mathbf{x}^{(i+1)}. \quad (6c)$$

Note that the projection (6b) can be performed exactly, whereas (6a) can be only approximated.

Original S-SPADE

Similarly, the original S-SPADE can be derived based on the minimization of the Augmented Lagrangian. The non-consistency lies in the observation that the originally approximated problem is

$$\arg\min_{\mathbf{w},\mathbf{z}} \|\mathbf{z}\|_0 \text{ s.t. } D\mathbf{w} \in \Gamma,\ \|\mathbf{w} - \mathbf{z}\|_2 \leq \epsilon \quad (7)$$

instead of (3b), thus it does not correspond to the analysis variant (3a) as the proposed version does.

Experiments

Experiments were performed on 5 diverse audio files sampled at 16 kHz. All audio samples were processed frame-wise using a Hann window 1024 samples long with a 75 % overlap. The oversampled DFT was used as the time-frequency transformation. The relaxation parameters were set to $r = 1, s = 1$ and $\epsilon = 0.1$.

The restoration quality was evaluated using ΔSDR, which expresses the signal-to-distortion improvement in dB, defined as

$$\Delta\mathrm{SDR} = \mathrm{SDR}(\mathbf{x}, \hat{\mathbf{x}}) - \mathrm{SDR}(\mathbf{x}, \mathbf{y}), \quad (8)$$

where $\mathbf{y}$ represents the clipped signal, $\mathbf{x}$ is the original undistorted signal and $\hat{\mathbf{x}}$ denotes the restored signal. The SDR itself is computed as:

$$\mathrm{SDR}(\mathbf{u}, \mathbf{v}) = 10 \log \frac{\|\mathbf{u}\|_2^2}{\|\mathbf{u} - \mathbf{v}\|_2^2} \text{ [dB]}. \quad (9)$$

For different redundancies of the oversampled DFT, we compare A-SPADE, S-SPADE_O (the original) and S-SPADE_DP (the proper version), see Fig. 1 for comparison in terms of restoration quality and Fig. 2 for comparison in terms of the performance speed.

The source codes for MATLAB are available at the link in the QR-code below.

SPADE Algorithms

Alg. 1: A-SPADE original
Require: $A, \mathbf{y}, M_R, M_H, M_L, s, r, \epsilon$
$\hat{\mathbf{x}}^{(0)} = \mathbf{y}, \mathbf{u}^{(0)} = \mathbf{0}, i = 0, k = s$
$\bar{\mathbf{z}}^{(i+1)} = \mathcal{H}_k\left(A\hat{\mathbf{x}}^{(i)} + \mathbf{u}^{(i)}\right)$
$\hat{\mathbf{x}}^{(i+1)} = \arg\min_{\mathbf{x}} \|A\mathbf{x} - \bar{\mathbf{z}}^{(i+1)} + \mathbf{u}^{(i)}\|_2^2$ s.t. $\mathbf{x} \in \Gamma$
if $\|A\hat{\mathbf{x}}^{(i+1)} - \bar{\mathbf{z}}^{(i+1)}\|_2 \leq \epsilon$ **then**
terminate
else
$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + A\hat{\mathbf{x}}^{(i+1)} - \bar{\mathbf{z}}^{(i+1)}$
$i \leftarrow i + 1$
if $i \bmod r = 0$ **then**
$k \leftarrow k + s$
go to 2
return $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(i+1)}$

Alg. 2: S-SPADE original
Require: $D, \mathbf{y}, M_R, M_H, M_L, s, r, \epsilon$
$\hat{\mathbf{z}}^{(0)} = D^*\mathbf{y}, \mathbf{u}^{(0)} = \mathbf{0}, i = 0, k = s$
$\bar{\mathbf{z}}^{(i+1)} = \mathcal{H}_k\left(\hat{\mathbf{z}}^{(i)} + \mathbf{u}^{(i)}\right)$
$\hat{\mathbf{z}}^{(i+1)} = \arg\min_{\mathbf{z}} \|\mathbf{z} - \bar{\mathbf{z}}^{(i+1)} + \mathbf{u}^{(i)}\|_2^2$ s.t. $D\mathbf{z} \in \Gamma$
if $\|\hat{\mathbf{z}}^{(i+1)} - \bar{\mathbf{z}}^{(i+1)}\|_2 \leq \epsilon$ **then**
terminate
else
$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + \hat{\mathbf{z}}^{(i+1)} - \bar{\mathbf{z}}^{(i+1)}$
$i \leftarrow i + 1$
if $i \bmod r = 0$ **then**
$k \leftarrow k + s$
go to 2
return $\hat{\mathbf{x}} = D\hat{\mathbf{z}}^{(i+1)}$

Alg. 3: S-SPADE proposed
Require: $D, \mathbf{y}, M_R, M_H, M_L, s, r, \epsilon$
$\hat{\mathbf{x}}^{(0)} = \mathbf{y}, \mathbf{u}^{(0)} = \mathbf{0}, i = 0, k = s$
$\bar{\mathbf{z}}^{(i+1)} = \mathcal{H}_k\left(D^*(\hat{\mathbf{x}}^{(i)} - \mathbf{u}^{(i)})\right)$
$\hat{\mathbf{x}}^{(i+1)} = \arg\min_{\mathbf{x}} \|D\bar{\mathbf{z}}^{(i+1)} - \mathbf{x} + \mathbf{u}^{(i)}\|_2^2$ s.t. $\mathbf{x} \in \Gamma$
if $\|D\bar{\mathbf{z}}^{(i+1)} - \hat{\mathbf{x}}^{(i+1)}\|_2 \leq \epsilon$ **then**
terminate
else
$\mathbf{u}^{(i+1)} = \mathbf{u}^{(i)} + D\bar{\mathbf{z}}^{(i+1)} - \hat{\mathbf{x}}^{(i+1)}$
$i \leftarrow i + 1$
if $i \bmod r = 0$ **then**
$k \leftarrow k + s$
go to 2
return $\hat{\mathbf{x}} = \hat{\mathbf{x}}^{(i+1)}$

Results

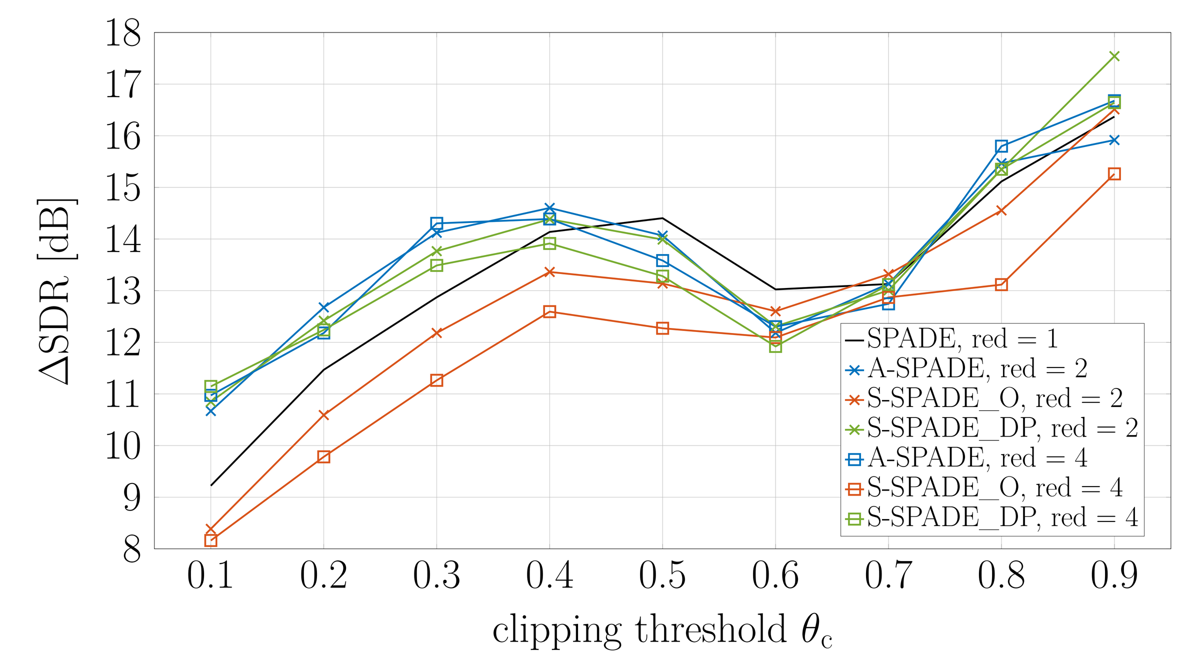

Figure 1: Average performance in terms of ΔSDR for all three algorithms. Notation "red" denotes redundancy of the DFT.

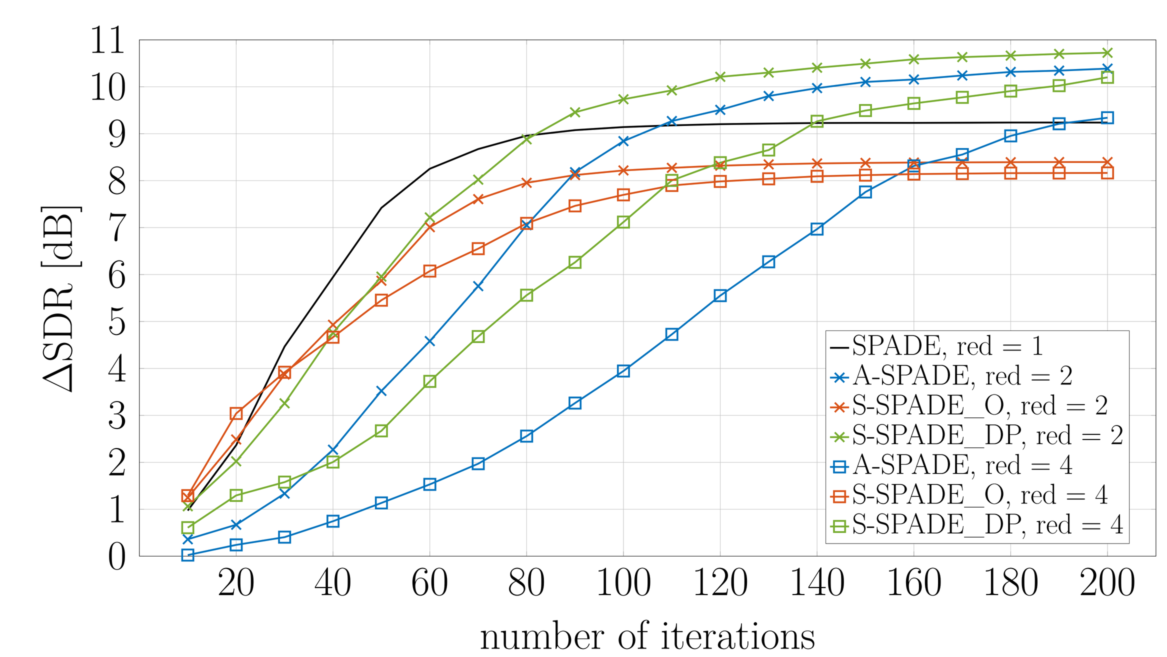

Figure 2: Average ΔSDR versus the number of iterations.

Conclusion

A novel algorithm for audio declipping based on the sparse synthesis model was introduced. Unlike original S-SPADE, the proposed version indeed solves the problem formulation (3b). The restoration performance is significantly better than with the original version of S-SPADE, and it is comparable with the analysis variant. The experiments also show that the new S-SPADE converges faster than A-SPADE.