Consistent dictionary learning for signal declipping

Lucas Rencker[1★(✉)], Francis Bach[2], Wenwu Wang[1], and Mark D. Plumbley[1]

[1]Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
[2]SIERRA-project team, INRIA, Paris, France
{l.rencker,w.wang,m.plumbley}@surrey.ac.uk
francis.bach@inria.fr

Abstract. Clipping, or saturation, is a common nonlinear distortion in signal processing. Recently, declipping techniques have been proposed based on sparse decomposition of the clipped signals on a fixed dictionary, with additional constraints on the amplitude of the clipped samples. Here we propose a dictionary learning approach, where the dictionary is directly learned from the clipped measurements. We propose a soft-consistency metric that minimizes the distance to a convex feasibility set, and takes into account our knowledge about the clipping process. We then propose a gradient descent-based dictionary learning algorithm that minimizes the proposed metric, and is thus consistent with the clipping measurement. Experiments show that the proposed algorithm outperforms other dictionary learning algorithms applied to clipped signals. We also show that learning the dictionary directly from the clipped signals outperforms consistent sparse coding with a fixed dictionary.

1 Introduction

Clipping is a common nonlinear distortion in digital or analog systems, that often occurs due to dynamic range limitations. When a signal reaches a certain maximum allowed amplitude, the waveform is truncated and samples are said to be clipped. Declipping is the task of recovering the clipped samples from the surrounding, unclipped samples. Early strategies to recover a clipped signal include autoregressive modelling [1], bandwidth limited models [2], or Bayesian estimation [3]. More recently, sparsity-based declipping techniques have attracted a lot of interest. The idea is that the original signal can be sparsely represented using a known dictionary of atoms. Declipping can be treated as a simple signal *inpainting* problem, i.e. by discarding the clipped samples and solving a sparse decomposition problem on the unclipped samples [4]. However it was noted in [4, 5] that the reconstruction can be greatly improved by using extra information in the reconstruction process: indeed, we know that the clipped samples should have an amplitude that is *greater* than the clipping threshold. Several

★ The research leading to these results has received funding from the European Union's H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no 642685 MacSeNet.

approaches have been proposed in the literature in order to enforce clipping consistency, i.e. taking into account the clipping threshold. Sparse decomposition with amplitude constraints were proposed in [4–9], and solved using a two-step algorithm [4], Alternating Direction Method of Multipliers (ADMM) [6, 10], or using general purpose convex optimization toolboxes [7–9, 11]. However these approaches can be computationally intensive, and possibly non robust to measurement noise. Smooth regularizers were proposed in [12, 13], which lead to simple unconstrained cost functions, and can be optimized using variants of well known algorithms such as Iterative Hard Thresholding (IHT) or Iterative Shrinkage/Thresholding (ISTA). Additional information has also been used, such as perceptual weights [8], social sparsity priors [13], or multichannel data [14].

Sparsity-based declipping techniques proposed in the literature use fixed dictionaries such as discrete cosine transform (DCT) or Gabor. However, dictionary learning has proved to perform better in a variety of signal reconstruction tasks, such as denoising [15] or inpainting [16]. Well known dictionary learning algorithms have been proposed for denoising or inpainting [15–17], however dictionary learning from clipped measurements has not been addressed in the literature. In this paper we propose a dictionary learning algorithm that is able to learn directly from nonlinearly clipped measurements. We formulate the declipping problem as a problem of minimizing the distance between a sparse signal and a convex feasibility set. This provides a convex and smooth cost function which generalizes the Euclidean distance commonly used in sparse coding and dictionary learning. We then propose a gradient-descent based sparse coding and dictionary learning algorithm, that takes into account our knowledge about the clipping process. Experiments show that the proposed consistent dictionary learning algorithm performs better on the task of declipping than state-of-the art dictionary learning algorithms for signal inpainting. We also show that the proposed consistent dictionary learning improves the reconstruction, compared to consistent sparse coding with a fixed dictionary.

The paper is organized as follows: in Section 2 we briefly give an overview of sparsity-based declipping techniques, and of dictionary learning. In Section 3 we propose a new formulation of the declipping problem, and a consistent dictionary learning algorithm for signal declipping. Experiments are presented in Section 4, before the conclusion is drawn.

2 Background

2.1 Signal declipping

Let $\mathbf{x} \in \mathbb{R}^N$ be a clean input signal, and $\mathbf{y} \in \mathbb{R}^N$ its clipped measurement. In this paper we consider the case of *hard* clipping, where each sample y_i is measured as:

$$y_i = \begin{cases} \theta^+ & \text{if} \quad x_i \geq \theta^+ \\ \theta^- & \text{if} \quad x_i \leq \theta^- \\ x_i & \text{otherwise,} \end{cases} \tag{1}$$

where $\theta^+ > 0$ and $\theta^- < 0$ are positive and negative clipping thresholds respectively, and x_i is the input sample. This can be written in vector form as:

$$\mathbf{y} = \mathbf{M^r}\,\mathbf{x} + \theta^+\,\mathbf{M^{c+}}\,\mathbf{1} + \theta^-\,\mathbf{M^{c\text{-}}}\,\mathbf{1}, \tag{2}$$

where $\mathbf{1}$ is the all-ones vector in $\mathbb{R}^N$, and $\mathbf{M^r}, \mathbf{M^{c+}}$ and $\mathbf{M^{c\text{-}}}$ are diagonal *sensing* matrices in $\{0,1\}^{N\times N}$ that define the *reliable*, positive and negative clipped samples respectively. These matrices can be estimated by detecting samples that have reached the clipping threshold, e.g. $[\mathbf{M^{c+}}]_{i,i} = 1$ if $y_i = \theta^+$, or 0 otherwise. Since the clipped samples are missing, a simple way to treat declipping is to formulate it as an *inpainting* problem, i.e. a problem of interpolating missing samples [4]. Assuming that the original signal can be sparsely represented in a known dictionary $\mathbf{D} \in \mathbb{R}^{N\times M}$, the inpainting problem can be formulated as:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^M} \|\,\mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \quad \|\,\boldsymbol{\alpha}\,\|_0 \le K, \tag{3}$$

where $\|.\|_0$ is the ℓ_0 pseudo-norm, and K is a parameter that controls the sparsity level. Eqn. (3) is a classical sparse coding problem, which can be solved using well known algorithms like IHT [18]. However, we can use extra information about the clipping process. Indeed, we know that the clipped samples should have an amplitude that is above (resp. below) the clipping threshold θ^+ (resp. θ^-). This can be enforced using amplitude constraints in the reconstruction process [4, 5]:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^M} \|\,\mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \begin{cases} \|\,\boldsymbol{\alpha}\,\|_0 \le K \\ \mathbf{M^{c+}}\,\mathbf{D}\,\boldsymbol{\alpha} \succeq \theta^+\,\mathbf{M^{c+}}\,\mathbf{1} \\ \mathbf{M^{c\text{-}}}\,\mathbf{D}\,\boldsymbol{\alpha} \preceq \theta^-\,\mathbf{M^{c\text{-}}}\,\mathbf{1} \end{cases} \tag{4}$$

Eqn. (4) is a difficult non-convex and constrained optimization problem, which cannot be readily solved using off-the-shelf sparse decomposition solvers such as IHT. A two-step algorithm was proposed in [4], where the support of non-zero atoms is first estimated using (3), and the signal is then estimated using a constrained least squares on the estimated support. However, the support selection does not take into account the clipping constraints and is thus suboptimal. A similar constraint-based formulation was proposed in [6]:

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^M} \|\,\boldsymbol{\alpha}\,\|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}), \tag{5}$$

where $\mathbb{1}_{\mathcal{C}(\mathbf{y})}$ is the indicator function of the set $\mathcal{C}(\mathbf{y})$, and:

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} |\, \mathbf{M^r}\,\mathbf{y} = \mathbf{M^r}\,\mathbf{x}, \mathbf{M^{c+}}\,\mathbf{x} \succeq \mathbf{M^{c+}}\,\mathbf{y}, \mathbf{M^{c\text{-}}}\,\mathbf{x} \preceq \mathbf{M^{c\text{-}}}\,\mathbf{y}\} \tag{6}$$

is the set of *feasible* signals, i.e. the set of signals that are consistent with the observation $\mathbf{y}$. The authors in [6] proposed an ADMM based algorithm to solve (5). The ADMM-based declipper [6] leads to good performance, but proves to be computationally expensive since it involves non-orthogonal projections which need to be computed iteratively[1]. Similar ℓ_1-based constrained formulation were also proposed in [7–9], and solved using general purpose convex optimization toolboxes [11], which can also be time consuming. Moreover, constrained formulation like (5) might not be robust to measurement noise, as will be discussed in the experimental section. Several authors proposed to enforce consistency with the clipped samples in a more tractable way. A smooth regularizer that penalizes clipped samples was proposed in [12]:

$$\begin{aligned} \min_{\boldsymbol{\alpha}\in\mathbb{R}^M} &\| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M^{c+}}(\theta^+\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 \\ &+\| \mathbf{M^{c\text{-}}}(\theta^-\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2 \quad \text{s.t.} \quad \| \boldsymbol{\alpha} \|_0 \leq K, \end{aligned} \tag{7}$$

where $(u)_+ = \max(0, u)$ and $(u)_- = -(-u)_+$. Since the cost in (7) is smooth, gradient-based sparse coding algorithms can easily be extended to the clipping consistent model (7). A *consistent* IHT was proposed in [12] in order to enforce clipping consistency. A similar formulation with an ℓ_1 norm was proposed in [13] along with ISTA-like algorithms. Although the algorithm in [12] did not perform as well as the ADMM based declipper [6], the soft consistency metric in (7) provides a simple, unconstrained way to enforce consistency with the clipped samples. Moreover, simple iterative thresholding algorithms can be derived, which are computationally faster than solving constrained optimization problems like (5).

2.2 Dictionary learning

Previously mentioned declipping techniques use fixed dictionaries, such as DCT or Gabor. However in many applications, learning a dictionary that is adaptive to the data has proved to lead to much better signal estimates [15,16]. A dictionary learning problem (from clean signals) is often formulated as [19]:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 \quad \text{s.t.} \quad \forall t, \| \boldsymbol{\alpha}_t \|_0 \leq K \tag{8}$$

where $\{\mathbf{x}_t\}_{1\ldots T}$ is a collection of T signals in $\mathbb{R}^N$. The dictionary is often constrained to be in $\mathcal{D} = \{\mathbf{D} \in \mathbb{R}^{N\times M} | \forall i, \|\mathbf{d}_i\|_2 \leq 1\}$ in order to avoid scaling ambiguity [19]. Many dictionary learning algorithms have been proposed to learn from clean or noisy data, such as MOD [17] or K-SVD [15]. In the case of inpainting, a weighted K-SVD (wK-SVD) has been proposed in order to deal with missing samples [16]. Dictionary learning from nonlinearly clipped data has not been addressed in the literature. Since dictionary learning usually alternates between several iterations of sparse coding and dictionary update over large datasets, a computationally tractable and stable formulation is needed. In the next section, we propose a soft data-consistency metric, that provides a simple optimization problem for dictionary learning. We then propose a consistent dictionary learning algorithm that is able to learn from the clipped measurements.

[1] An *analysis* sparsity version of (5) was also proposed in [6], which proved to be computationally more tractable. In this paper we focus on the synthesis sparsity model, and leave the analysis sparsity counterpart for future work.

3 Consistent dictionary learning for signal declipping

3.1 Proposed problem formulation

We first reformulate declipping as a problem of minimizing the distance between the approximated signal, and the feasible set $\mathcal{C}(\mathbf{y}_t)$ defined in (6):

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}_t, \mathcal{C}(\mathbf{y}_t))^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K, \tag{9}$$

where $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))$ is the Euclidean distance between $\mathbf{x}$ and the set $\mathcal{C}(\mathbf{y})$, defined as:

$$\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y})) = \min_{\mathbf{z}\in\mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2. \tag{10}$$

The formulation (9) thus enforces the estimated signals to be “close” to their feasibility sets $\mathcal{C}(\mathbf{y}_\mathrm{t})$ in a Euclidean-distance sense, unlike the formulation in (5) which constrains the signals to be exactly in $\mathcal{C}(\mathbf{y}_\mathrm{t})$. We thus have proposed here a problem of minimizing the distance to a set, which differs from classical sparse coding and dictionary learning approaches which minimize the distance to a point in $\mathbb{R}^N$. Using (10), (9) can further be reformulated as a “min-min” problem:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \min_{\mathbf{z}\in\mathcal{C}(\mathbf{y}_t)} \|\mathbf{D}\,\boldsymbol{\alpha}_t - \mathbf{z}\|_2^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K. \tag{11}$$

Note that as a minimum of a family of convex functions $\|.\|_2$ over a non-empty and convex set $\mathcal{C}(\mathbf{y})$, $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))$ is a convex cost function [20, Section 3.2.5]. Moreover, using Danskin’s Min-Max theorem ([21, Theorem 4.1], originally proposed in [22]), it can be shown that $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$ is differentiable with gradient [23]:

$$\nabla_{\mathbf{x}}\, \mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = 2(\mathbf{x} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x})), \tag{12}$$

where $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x})$ is the Euclidean projection of $\mathbf{x}$ onto $\mathcal{C}(\mathbf{y})$. The proposed formulation in (9) is thus a problem of minimizing a smooth and convex cost function, with a sparsity constraint, which is similar to the classical dictionary learning problem (8). The proposed cost function thus generalizes the linear least-squares commonly used in sparse coding and dictionary learning.

3.2 Algorithm

We propose a simple gradient descent-based algorithm, which we present in Algorithm 1. The proposed algorithm alternates between a sparse coding step and a dictionary update step. Similarly to IHT, the sparse coding step alternates between gradient descent (13) and a hard thresholding (14). The dictionary is updated using projected gradient descent (15). n_i and μ_i ($i = 1, 2$) are parameters that control the number of gradient descent steps and step sizes respectively.

Algorithm 1 Dictionary learning for declipping

Require: $\{\mathbf{y}_t\}_{1...T}$, $\mathbf{D}^0$, n_1, n_2, μ_1, μ_2
  initialize: $\mathbf{D} \leftarrow \mathbf{D}^0$, $\boldsymbol{\alpha}_t \leftarrow \mathbf{0}$
  while stopping criterion not reached **do**
    for $t = 1...T$ **do** ▷ Sparse coding step
      for $i = 1, ..., n_1$ **do**

$$\boldsymbol{\alpha}_t \leftarrow \boldsymbol{\alpha}_t + \mu_1 \mathbf{D}^T (\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t) \tag{13}$$

$$\boldsymbol{\alpha}_t \leftarrow \mathcal{H}_K(\boldsymbol{\alpha}_t) \tag{14}$$

    for $j = 1, ..., n_2$ **do** ▷ Dictionary update step

$$\mathbf{D} \leftarrow \Pi_{\mathcal{D}}\big(\mathbf{D} + \mu_2 \sum_t (\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t)\,\boldsymbol{\alpha}_t^T\,\big) \tag{15}$$

  return $\hat{\mathbf{D}}, \{\hat{\boldsymbol{\alpha}}_t\}_{1...T}$

3.3 Computation of the residuals, and interpretation

Alg. 1 involves the computation of residuals $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) - \mathbf{D}\,\boldsymbol{\alpha}$ at every step. These residuals can be easily computed in closed form. It can be easily verified that the projection operator $\Pi_{\mathcal{C}(\mathbf{y})}$ is computed as:

$$\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x}) = \mathbf{M}^{\mathbf{r}}\,\mathbf{y} + \mathbf{M}^{\mathbf{c}+}\max(\mathbf{y}, \mathbf{x}) + \mathbf{M}^{\mathbf{c}\text{-}}\min(\mathbf{y}, \mathbf{x}). \tag{16}$$

Note that this is a simple 1-dimensional orthogonal projection on each sample, that can be computed at a negligible cost. The residuals can be computed as:

$$\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) - \mathbf{D}\,\boldsymbol{\alpha} = \mathbf{M}^{\mathbf{r}}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha}) + \mathbf{M}^{\mathbf{c}+}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})_+ + \mathbf{M}^{\mathbf{c}\text{-}}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})_-. \tag{17}$$

This also shows that the proposed soft-consistency metric (9) can be written in closed form as:

$$\begin{aligned} \mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))^2 = &\| \mathbf{M}^{\mathbf{r}}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M}^{\mathbf{c}+}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 \\ &+ \| \mathbf{M}^{\mathbf{c}\text{-}}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2, \end{aligned} \tag{18}$$

which (noticing that $\mathbf{M}^{\mathbf{c}+}\,\mathbf{y} = \theta^+\,\mathbf{M}^{\mathbf{c}+}\,\mathbf{1}$) is the same as the cost (7) used in [12, 13]. The proposed approach is thus a different way to motivate the soft-consistency metric (7), and the sparse coding step in Alg. 1 is equivalent to the "consistent IHT" [12]. Note also that when no sample is clipped, we have $\mathcal{C}(\mathbf{y}) = \{\mathbf{y}\}$, $\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))^2 = \|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2$, and $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) = \mathbf{y}$. Thus (9) becomes a classical dictionary learning problem, and Alg. 1 a classical dictionary learning algorithm. The proposed method is thus a generalization of dictionary learning to nonlinearly clipped measurements.

4 Evaluation

We evaluate the performance of the proposed algorithm on audio declipping tasks [2]. The test set consists of 10 speech and 10 music signals of 10s each, sampled at 16kHz. The signals were processed with Hamming windows of size $N = 256$ samples, with 75% overlap, for a total of approximately $T = 2500$ frames per signal. The dictionary learning algorithm was initialized with a DCT dictionary of $M = 512$ atoms, and the sparse coefficients initialized to zero. Each gradient descent step was then initialized with a *warm restart* strategy, i.e. using the estimate from the previous iteration [19]. We performed 50 iterations of gradient descent, with 20 iterations for each inner sparse coding and dictionary update step. The gradient descent steps were chosen as $\mu_1 = 1/\|\mathbf{D}\|_2^2$ and $\mu_2 = 1/\|\mathbf{A}\|_2^2$ (with $\mathbf{A} = [\boldsymbol{\alpha}_1, ..., \boldsymbol{\alpha}_T]$), and updated at each iteration using the current estimates $\mathbf{D}$ and $\mathbf{A}$. When no noise is present, the estimated signals $\hat{\mathbf{x}}$ can be re-projected on the set $\{\mathbf{x}|\, \mathbf{M^r}\, \mathbf{x} = \mathbf{M^r}\, \mathbf{y}\}$ as a final step, in order to avoid approximation errors. The quality of the estimated signal can then be evaluated using the signal to distortion ratio (SDR) computed on the *clipped* samples: $\mathrm{SDR_c}(\hat{\mathbf{x}}, \mathbf{x}) = 20 \log \frac{\|(\mathbf{M^{c+}} + \mathbf{M^{c-}})\mathbf{x}\|_2}{\|(\mathbf{M^{c+}} + \mathbf{M^{c-}})(\mathbf{x} - \hat{\mathbf{x}})\|_2}$.

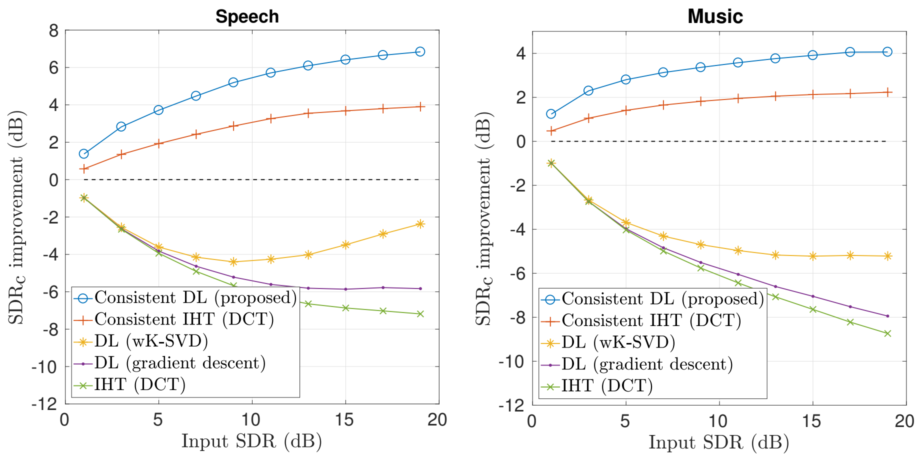

Fig. 1: Comparison with state-of-the-art dictionary learning algorithms

Figure 1 shows the performance of the proposed consistent dictionary learning (DL) algorithm compared to other dictionary learning algorithms for inpainting. We show the average performance for different clipping levels, ranging from severe clip (SDR = 1dB) to light clip (SDR = 19dB). As a baseline, we show the performance of IHT, computed on the unclipped samples and with a fixed

[2] The MATLAB code and some examples are available at `http://www.cvssp.org/Personal/LucasRencker/software.html`

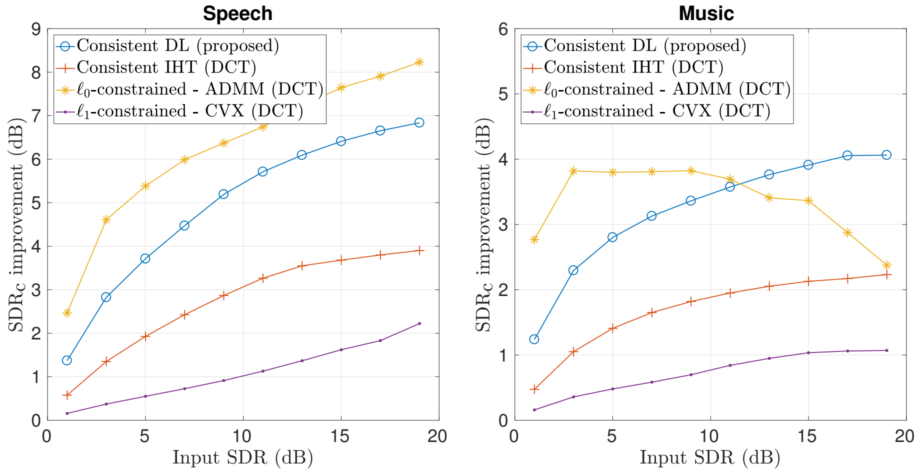

Fig. 2: Comparison with state-of-the-art declipping algorithms

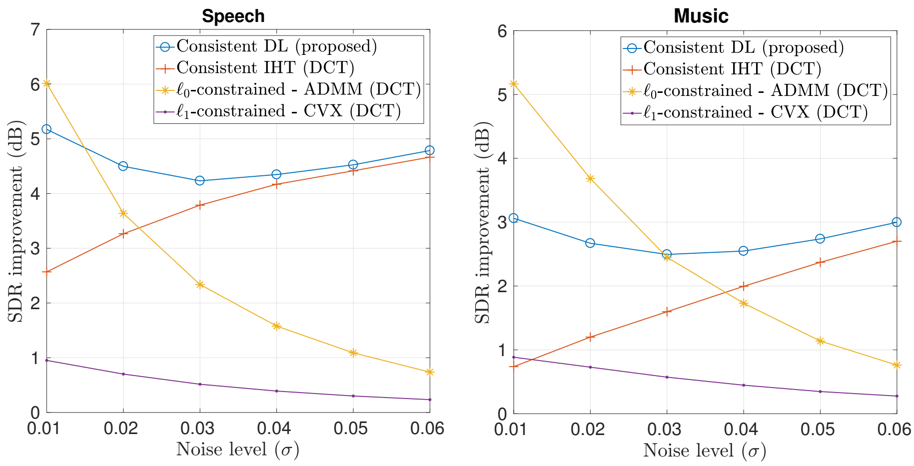

Fig. 3: Robustness to measurement noise

DCT dictionary. We show the performance of two dictionary learning algorithms computed on the unclipped samples: a gradient descent-based algorithm similar to Alg. 1, and wK-SVD which achieved state-of-the art performance in signal inpainting [16]. Although these two algorithms slightly improve the reconstruction compared to IHT, the overall performance is quiet poor. Consistent IHT [12] clearly outperforms methods that discard clipped samples. The proposed consistent DL algorithm further improves the reconstruction, with an improvement of up to 3dB in the case of speech signals. This shows that learning the dictionary directly from the clipped signals outperforms fixed dictionaries, and that the learned dictionary generalizes well to the clipped samples. Note also that while standard dictionary learning algorithms fail to improve the reconstruction when the data is heavily clipped (SDR $\leq$ 5dB), the proposed dictionary learning algorithm is still able to improve compared to consistent IHT with fixed DCT.

Figure 2 shows the performance comparison with other declipping algorithms proposed in the literature. We compare with consistent IHT, the ℓ_1-constrained formulation proposed in [9] solved using CVX [11], and the ℓ_0-constrained formulation (5) solved using the ADMM-based algorithm proposed in [6], considered as the current state-of-the-art. All ℓ_0-based algorithms were computed with a fixed $K = 32$, except the ADMM algorithm which we have found does not converge when K is fixed. We have thus implemented ADMM with the adaptive sparsity strategy proposed in [6], which might favor it. Although the proposed consistent DL algorithm does not match ADMM's performance on average, our algorithm bridges the gap between consistent IHT and ADMM, and outperforms ADMM in the case of music signals when SDR $\geq$ 12dB. However as shown in the next experiment, the proposed algorithm is more robust to measurement noise. Figure 3 shows the reconstruction performance for signals contaminated with additive Gaussian noise with variance σ^2, and clipped at $\theta = 0.3$. Figure 3 shows that algorithms based on soft-consistency metric such as consistent IHT or the proposed algorithm are more robust to noise than constrained-based formulation. In particular, the proposed algorithm outperforms every other algorithms for noise levels above 0.01 in speech, and 0.03 in music. From a computational point of view, consistent IHT takes about 5s to process a 10s signal, the proposed consistent DL and ADMM about 2-3 min, and CVX about an hour.

5 Conclusion

We proposed a smooth and convex cost function for signal declipping, and a dictionary learning algorithm that is able to learn from clipped measurements. The proposed algorithm outperforms classical dictionary learning algorithms, and improves the declipping performance compared to consistent sparse coding with a fixed dictionary. The proposed algorithm is simple and efficient, and the model proposed in (9) could potentially be applied to other nonlinear measurements, such as quantization or 1-bit measurements, which will be addressed in a future publication. Analysis sparsity has shown promising results in [6], so future work will also investigate analysis dictionary learning for declipping.

References

1. A. Janssen, R. N. Veldhuis, and L. B. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 2, pp. 317–330, 1986.
2. J. S. Abel and J. O. Smith, "Restoring a clipped signal," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Toronto, Canada, 1991, pp. 1745–1748.
3. S. J. Godsill, P. J. Wolfe, and W. N. Fong, "Statistical model-based approaches to audio restoration and analysis," *Journal of New Music Research*, vol. 30, no. 4, pp. 323–338, 2001.

4. A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 3, pp. 922–932, 2012.
5. A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A constrained matching pursuit approach to audio declipping," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 329–332.
6. S. Kitić, N. Bertin, and R. Gribonval, "Sparsity and cosparsity for audio declipping: a flexible non-convex approach," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.*, Liberec, Czech Republic, Aug. 2015, pp. 243–250.
7. H. Mansour, R. Saab, P. Nasiopoulos, and R. Ward, "Color image desaturation using sparse reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 778–781.
8. B. Defraene, N. Mansour, S. D. Hertogh, T. van Waterschoot, M. Diehl, and M. Moonen, "Declipping of audio signals using perceptual compressed sensing," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 12, pp. 2627–2637, Dec. 2013.
9. S. Foucart and T. Needham, "Sparse recovery from saturated measurements," *Information and Inference: A Journal of the IMA*, vol. 6, no. 2, pp. 196–212, 2016.
10. 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.
11. M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.
12. S. Kitić, L. Jacques, N. Madhu, M. P. Hopwood, A. Spriet, and C. D. Vleeschouwer, "Consistent iterative hard thresholding for signal declipping," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 5939–5943.
13. K. Siedenburg, M. Kowalski, and M. Dörfler, "Audio declipping with social sparsity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Florence, Italy, May 2014, pp. 1577–1578.
14. A. Ozerov, Ç. Bilen, and P. Pérez, "Multichannel audio declipping," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 659–663.
15. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
16. J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.
17. K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, 1999, pp. 2443–2446.
18. T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Computat. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
19. J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," *Found. Trends Comput. Graph. Vis.*, vol. 8, no. 2-3, pp. 85–283, 2014.
20. S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.
21. J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM review*, vol. 40, no. 2, pp. 228–264, 1998.
22. J. M. Danskin, *The theory of max-min and its application to weapons allocation problems.* Ökonometrie und Unternehmensforschung, 1967.
23. R. B. Holmes, "Smoothness of certain metric projections on Hilbert space," *Transactions of the American Mathematical Society*, vol. 184, pp. 87–100, 1973.