Consistent dictionary learning for signal declipping

LVA/ICA 2018, Guildford, UK

Lucas Rencker[1] Francis Bach[2] Wenwu Wang[1] Mark D. Plumbley[1]

[1]CVSSP, University of Surrey
[2]SIERRA-team, INRIA

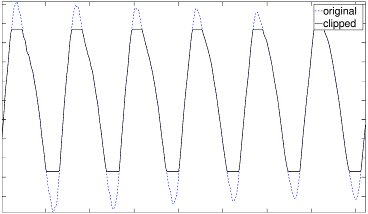
lucas.rencker@surrey.ac.uk

05/07/2018

Table of Contents

Table of Contents

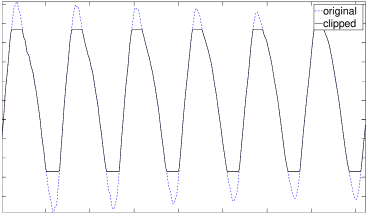
Background: declipping



Clipping:

- Common distortion in signal processing
- Signal saturates above a certain threshold

Background: declipping



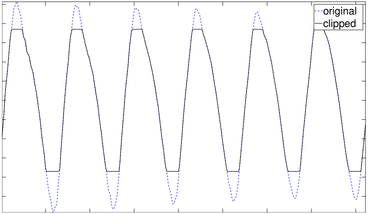
Clipping:

- Common distortion in signal processing
- Signal saturates above a certain threshold

Declipping:

- Recovering original signal from clipped signal
- Non-linear, highly under-determined inverse problem (only low energy samples are available)

Background: declipping



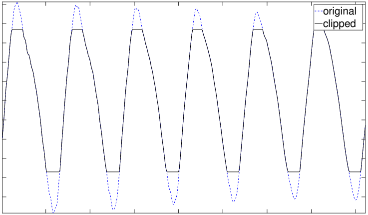
Clipping:

- Common distortion in signal processing
- Signal saturates above a certain threshold

Declipping:

- Recovering original signal from clipped signal
- Non-linear, highly under-determined inverse problem (only low energy samples are available)
- Declipping strategies: AR modelling, bandwidth-limited models, Bayesian approaches

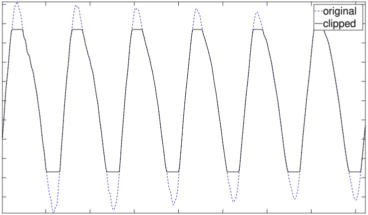
More recently: **sparsity based approaches**:



- **y**: measured clipped signal
- **x**: original clean signal

- Assume original signal is sparse $\mathbf{x} = \mathbf{D}\,\boldsymbol{\alpha}$, where $\mathbf{D} \in \mathbb{R}^{N \times M}$ $(N \leq M)$ **overcomplete dictionary** and $\|\,\boldsymbol{\alpha}\,\|_0 \leq K$.

More recently: **sparsity based approaches**:



- $\mathbf{y}$: measured clipped signal
- $\mathbf{x}$: original clean signal

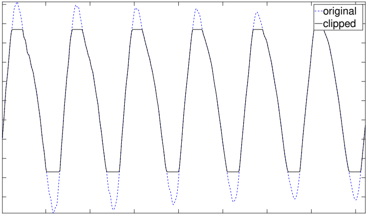
- Assume original signal is sparse $\mathbf{x} = \mathbf{D}\,\boldsymbol{\alpha}$, where $\mathbf{D} \in \mathbb{R}^{N\times M}$ $(N \leq M)$ **overcomplete dictionary** and $\|\,\boldsymbol{\alpha}\,\|_0 \leq K$.

- "Straighforward" declipping formulation:

$$\min_{\boldsymbol{\alpha}} \|\,\mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \quad \|\,\boldsymbol{\alpha}\,\|_0 \leq K, \tag{1}$$

where $\mathbf{M^r}$ is a binary sensing matrix defining the *reliable* (i.e. unclipped) samples.

More recently: **sparsity based approaches**:



- $\mathbf{y}$: measured clipped signal
- $\mathbf{x}$: original clean signal

- Assume original signal is sparse $\mathbf{x} = \mathbf{D}\,\boldsymbol{\alpha}$, where $\mathbf{D} \in \mathbb{R}^{N \times M}$ $(N \leq M)$ **overcomplete dictionary** and $\|\,\boldsymbol{\alpha}\,\|_0 \leq K$.

- "Straighforward" declipping formulation:

$$\min_{\boldsymbol{\alpha}} \|\,\mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \quad \|\,\boldsymbol{\alpha}\,\|_0 \leq K, \tag{1}$$

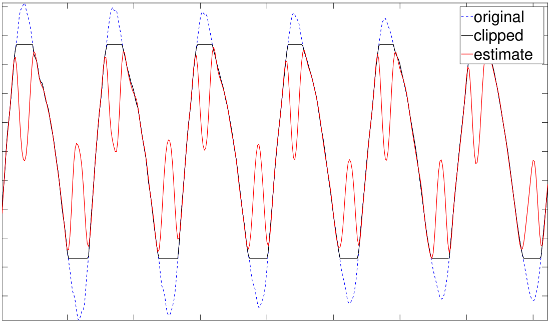
  where $\mathbf{M^r}$ is a binary sensing matrix defining the *reliable* (i.e. unclipped) samples.
- Many well-known algorithms to solve (1), e.g. (Orthogonal) Matching Pursuit, Iterative Hard Thresholding (IHT), etc...

Example:

- $\hat{\boldsymbol{\alpha}} = \text{argmin}_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \quad \| \boldsymbol{\alpha} \|_0 \leq K$
- estimate full clean signal $\hat{\mathbf{x}} = \mathbf{D}\hat{\boldsymbol{\alpha}}$:

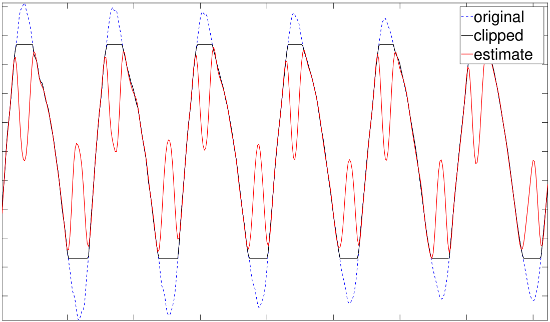
Example:

- $\hat{\alpha} = \text{argmin}_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 \quad \text{s.t.} \quad \|\,\boldsymbol{\alpha}\,\|_0 \leq K$
- estimate full clean signal $\hat{\mathbf{x}} = \mathbf{D}\hat{\alpha}$:



Example:

- $\hat{\alpha} = \text{argmin}_{\alpha} \| \mathbf{M}^{\mathbf{r}}(\mathbf{y} - \mathbf{D}\,\alpha)\|_2^2 \quad \text{s.t.} \quad \|\, \alpha \,\|_0 \leq K$
- estimate full clean signal $\hat{\mathbf{x}} = \mathbf{D}\hat{\alpha}$:



- **“classical” well known sparse recovery algorithms do not perform well on declipping!**

- Strategy: enforce reconstructed samples to be **above/below the clipping threshold** [Adler2012]:

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

where $\mathbf{M}^{\mathbf{c}+}$ and $\mathbf{M}^{\mathbf{c}\text{-}}$ define the position of the positive/negative clipped samples, and θ^+/θ^- positive/negative clipping thresholds.

- Strategy: enforce reconstructed samples to be **above/below the clipping threshold** [Adler2012]:

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

  where $\mathbf{M^{c+}}$ and $\mathbf{M^{c-}}$ define the position of the positive/negative clipped samples, and θ^+/θ^- positive/negative clipping thresholds.

- Formulation is consistent with the clipping process (fully models our knowledge about the clipping process)

- Strategy: enforce reconstructed samples to be **above/below the clipping threshold** [Adler2012]:

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

where $\mathbf{M}^{\mathbf{c+}}$ and $\mathbf{M}^{\mathbf{c-}}$ define the position of the positive/negative clipped samples, and θ^+/θ^- positive/negative clipping thresholds.

- Formulation is consistent with the clipping process (fully models our knowledge about the clipping process)
- **Difficult constrained, high-dimensional, non-convex optimization problem!**

ADMM-based sparse declipper: (SPADE) [Kitic,2015]

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) \tag{3}$$

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

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} |\, \mathbf{M}^{\mathbf{r}}\,\mathbf{y} = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}\} \tag{4}$$

the **constraint set** associated with clipped signal **y**.

ADMM-based sparse declipper: (SPADE) [Kitic,2015]

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) \tag{3}$$

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

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} | \, \mathbf{M}^{\mathbf{r}}\,\mathbf{y} = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}\} \tag{4}$$

the **constraint set** associated with clipped signal **y**.

- Alternates between hard-thresholding, and **non-orthogonal projection**:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{u} - \boldsymbol{\alpha}\|_2^2 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}), \quad \mathbf{u} \in \mathbb{R}^M \tag{5}$$

ADMM-based sparse declipper: (SPADE) [Kitic,2015]

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) \tag{3}$$

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

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} | \, \mathbf{M^r}\,\mathbf{y} = \mathbf{M^r}\,\mathbf{x}, \mathbf{M^{c+}}\,\mathbf{x} \succeq \mathbf{M^{c+}}\,\mathbf{y}, \mathbf{M^{c\text{-}}}\,\mathbf{x} \preceq \mathbf{M^{c\text{-}}}\,\mathbf{y}\} \tag{4}$$

the **constraint set** associated with clipped signal $\mathbf{y}$.

- Alternates between hard-thresholding, and **non-orthogonal projection**:

$$\underset{\boldsymbol{\alpha}}{\text{argmin}} \, \|\mathbf{u} - \boldsymbol{\alpha}\, \|_2^2 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}), \quad \mathbf{u} \in \mathbb{R}^M \tag{5}$$

- **Hard to compute when D is not a tight frame!** $(\mathbf{D}^T\mathbf{D} \neq \xi\mathbf{I})$

ADMM-based sparse declipper: (SPADE) [Kitic,2015]

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) \tag{3}$$

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

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} | \, \mathbf{M^r}\,\mathbf{y} = \mathbf{M^r}\,\mathbf{x}, \mathbf{M^{c+}}\,\mathbf{x} \succeq \mathbf{M^{c+}}\,\mathbf{y}, \mathbf{M^{c\text{-}}}\,\mathbf{x} \preceq \mathbf{M^{c\text{-}}}\,\mathbf{y}\} \tag{4}$$

the **constraint set** associated with clipped signal **y**.

- Alternates between hard-thresholding, and **non-orthogonal projection**:

$$\underset{\boldsymbol{\alpha}}{\text{argmin}} \, \|\mathbf{u} - \boldsymbol{\alpha}\,\|_2^2 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}), \quad \mathbf{u} \in \mathbb{R}^M \tag{5}$$

- **Hard to compute when D is not a tight frame!** ($\mathbf{D}^T\mathbf{D} \neq \xi \mathbf{I}$)
- Needs to be computed iteratively, using (e.g.) another nested ADMM (**Heavy computational cost!**)

ADMM-based sparse declipper: (SPADE) [Kitic,2015]

$$\min_{\boldsymbol{\alpha}} \| \boldsymbol{\alpha} \|_0 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}) \tag{3}$$

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

$$\mathcal{C}(\mathbf{y}) \triangleq \{\mathbf{x} |\, \mathbf{M}^{\mathbf{r}}\,\mathbf{y} = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}\} \tag{4}$$

the **constraint set** associated with clipped signal $\mathbf{y}$.

- Alternates between hard-thresholding, and **non-orthogonal projection**:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{u} - \boldsymbol{\alpha}\|_2^2 + \mathbb{1}_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}), \quad \mathbf{u} \in \mathbb{R}^M \tag{5}$$

- **Hard to compute when D is not a tight frame!** ($\mathbf{D}^T\mathbf{D} \neq \xi\mathbf{I}$)
- Needs to be computed iteratively, using (e.g.) another nested ADMM (**Heavy computational cost!**)
- **Unstable** (does not converge when sparsity level K is fixed)

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016] $\Rightarrow$ low performance, still extremely slow

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016]$\Rightarrow$ low performance, still extremely slow
- Smooth regularizers to enforce clipping consistency [Kitic,2013],[Siedenburg,2014]:

$$\min_{\boldsymbol{\alpha}} \| \mathbf{M}^{\mathbf{r}}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M}^{\mathbf{c}+}(\theta^{+}\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_{+}\|_2^2 + \| \mathbf{M}^{\mathbf{c}-}(\theta^{-}\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_{-}\|_2^2 \quad \text{s.t.} \quad \| \boldsymbol{\alpha} \|_0 \leq K, \tag{6}$$

with $(u)_+ = \max(0, u)$ and $(u)_- = -(u)_+$.

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016]$\Rightarrow$ low performance, still extremely slow
- Smooth regularizers to enforce clipping consistency [Kitic,2013],[Siedenburg,2014]:

$$\min_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M^{c+}}(\theta^+\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 + \| \mathbf{M^{c-}}(\theta^-\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq K, \tag{6}$$

with $(u)_+ = \max(0, u)$ and $(u)_- = -(u)_+$.

  - Quadratic cost when clipping constraint is violated ($\theta^+\,\mathbf{M^{c+}}\,\mathbf{1} \succeq \mathbf{M^{c+}}\,\mathbf{D}\,\boldsymbol{\alpha}$)

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016]$\Rightarrow$ low performance, still extremely slow
- Smooth regularizers to enforce clipping consistency [Kitic,2013],[Siedenburg,2014]:

$$\min_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M^{c+}}(\theta^+\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 + \| \mathbf{M^{c-}}(\theta^-\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}\|_0 \leq K, \tag{6}$$

with $(u)_+ = \max(0, u)$ and $(u)_- = -(u)_+$.

  - Quadratic cost when clipping constraint is violated ($\theta^+\,\mathbf{M^{c+}}\,\mathbf{1} \succeq \mathbf{M^{c+}}\,\mathbf{D}\,\boldsymbol{\alpha}$)
  - **Smooth** cost function

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016]$\Rightarrow$ low performance, still extremely slow
- Smooth regularizers to enforce clipping consistency [Kitic,2013],[Siedenburg,2014]:

$$\min_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M^{c+}}(\theta^+\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 + \| \mathbf{M^{c-}}(\theta^-\mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2 \quad \text{s.t.} \quad \| \boldsymbol{\alpha} \|_0 \leq K, \tag{6}$$

with $(u)_+ = \max(0, u)$ and $(u)_- = -(u)_+$.

  - Quadratic cost when clipping constraint is violated ($\theta^+ \mathbf{M^{c+}}\,\mathbf{1} \succeq \mathbf{M^{c+}}\,\mathbf{D}\,\boldsymbol{\alpha}$)
  - **Smooth** cost function
  - Gradient descent based algorithms can be extended (**"Consistent IHT for signal declipping"** [Kitic,2013])

Alternative consistent strategies:

- Analysis sparsity models [Kitic,2015], [Gaultier,2017]
- ℓ_1-based constrained formulations [Foucart,2016]$\Rightarrow$ low performance, still extremely slow
- Smooth regularizers to enforce clipping consistency [Kitic,2013],[Siedenburg,2014]:

$$\min_{\boldsymbol{\alpha}} \| \mathbf{M^r}(\mathbf{y} - \mathbf{D}\,\boldsymbol{\alpha})\|_2^2 + \| \mathbf{M^{c+}}(\theta^+ \mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_+\|_2^2 \\ + \| \mathbf{M^{c-}}(\theta^- \mathbf{1} - \mathbf{D}\,\boldsymbol{\alpha})_-\|_2^2 \quad \text{s.t.} \quad \| \boldsymbol{\alpha} \|_0 \leq K, \tag{6}$$

with $(u)_+ = \max(0, u)$ and $(u)_- = -(u)_+$.

  - Quadratic cost when clipping constraint is violated ($\theta^+ \mathbf{M^{c+}}\, \mathbf{1} \succeq \mathbf{M^{c+}}\, \mathbf{D}\,\boldsymbol{\alpha}$)
  - **Smooth** cost function
  - Gradient descent based algorithms can be extended (**"Consistent IHT for signal declipping"** [Kitic,2013])
  - Computationally simple

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)
- Dictionary learning has proved to perform better in many inverse problems (denoising, inpainting, deblurring).

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)
- Dictionary learning has proved to perform better in many inverse problems (denoising, inpainting, deblurring).
- Dictionary learning from clean data $\{\mathbf{x}_t\}_{t=1,\dots,T}$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K \tag{7}$$

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)
- Dictionary learning has proved to perform better in many inverse problems (denoising, inpainting, deblurring).
- Dictionary learning from clean data $\{\mathbf{x}_t\}_{t=1,\dots,T}$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K \tag{7}$$

  - **Adapt the dictionary to the observed data**
  - **Make use of similarities/correlation between frames** $\{\mathbf{x}_t\}_{t=1,\dots,T}$

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)
- Dictionary learning has proved to perform better in many inverse problems (denoising, inpainting, deblurring).
- Dictionary learning from clean data $\{\mathbf{x}_t\}_{t=1,\ldots,T}$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K \tag{7}$$

  - **Adapt the dictionary to the observed data**
  - **Make use of similarities/correlation between frames** $\{\mathbf{x}_t\}_{t=1,\ldots,T}$
- Many algorithms to solve (7) (MOD, K-SVD, ...) in the context of clean/noisy data

Dictionary learning

- All declipping approaches use **fixed dictionaries** (DCT, Gabor)
- Dictionary learning has proved to perform better in many inverse problems (denoising, inpainting, deblurring).
- Dictionary learning from clean data $\{\mathbf{x}_t\}_{t=1,\ldots,T}$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{D}\,\boldsymbol{\alpha}_t\|_2^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K \qquad (7)$$

  - **Adapt the dictionary to the observed data**
  - **Make use of similarities/correlation between frames** $\{\mathbf{x}_t\}_{t=1,\ldots,T}$
- Many algorithms to solve (7) (MOD, K-SVD, ...) in the context of clean/noisy data
- Not addressed in the context of clipped data

Dictionary learning for declipping?

Dictionary learning often performs many iterations over large datasets, so we need a formulation that is:

- **computationally tractable**
- **stable**
- **does not make any assumption on the dictionary** (tightness etc...)

Table of Contents

Proposed problem formulation

- Reformulate declipping as a problem of **minimizing the distance between the approximated signals $\mathbf{D}\,\boldsymbol{\alpha}_t$, and their feasibility sets** $\mathcal{C}(\mathbf{y}_t)$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}_t, \mathcal{C}(\mathbf{y}_t))^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K, \tag{8}$$

with:

$$\mathcal{C}(\mathbf{y}_t) \triangleq \{\mathbf{x} |\, \mathbf{M}^{\mathbf{r}}\,\mathbf{y}_t = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}_t, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}_t\}, \tag{9}$$

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

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

Proposed problem formulation

- Reformulate declipping as a problem of **minimizing the distance between the approximated signals $\mathbf{D}\,\boldsymbol{\alpha}_t$, and their feasibility sets** $\mathcal{C}(\mathbf{y}_t)$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}_t, \mathcal{C}(\mathbf{y}_t))^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K, \tag{8}$$

with:

$$\mathcal{C}(\mathbf{y}_t) \triangleq \{\mathbf{x} | \, \mathbf{M}^{\mathbf{r}}\,\mathbf{y}_t = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}_t, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}_t\}, \tag{9}$$

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

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

- **Enforces signals to be "close" to their feasibility sets, instead of being exactly in the set.**

Proposed problem formulation

- Reformulate declipping as a problem of **minimizing the distance between the approximated signals $\mathbf{D}\,\boldsymbol{\alpha}_t$, and their feasibility sets** $\mathcal{C}(\mathbf{y}_t)$:

$$\min_{\mathbf{D}\in\mathcal{D},\boldsymbol{\alpha}_t} \sum_t \mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}_t, \mathcal{C}(\mathbf{y}_t))^2 \quad \text{s.t.} \quad \forall t, \|\,\boldsymbol{\alpha}_t\,\|_0 \leq K, \tag{8}$$

with:

$$\mathcal{C}(\mathbf{y}_t) \triangleq \{\mathbf{x} | \, \mathbf{M}^{\mathbf{r}}\,\mathbf{y}_t = \mathbf{M}^{\mathbf{r}}\,\mathbf{x}, \mathbf{M}^{\mathbf{c}+}\,\mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+}\,\mathbf{y}_t, \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}}\,\mathbf{y}_t\}, \tag{9}$$

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

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

- **Enforces signals to be "close" to their feasibility sets, instead of being exactly in the set.**
- Minimize distance to a **set**, instead of minimizing distance to a point!

Properties of $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$:

$\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$ so:

Properties of $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$:

$\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$ so:

- $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$ is **continuous**

Properties of $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$:

$\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$ so:

- $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$ is **continuous**

Moreover since $\mathcal{C}(\mathbf{y})$ is convex:

- $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$ is **convex**, as a minimum of convex functions over a convex set [Boyd,2004].

Differentiability of $d(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$:

Danskin's Min-Max theorem [Bonnans,1998]:

- $\mathcal{C}$ a compact set
- $g(\mathbf{x}) = \min_{\mathbf{z} \in \mathcal{C}} \phi(\mathbf{x}, \mathbf{z})$
- $\forall \mathbf{z} \in \mathbb{R}^N$, $\phi(., \mathbf{z})$ is **differentiable** with gradient $\nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z})$
- $\phi(\mathbf{x}, \mathbf{z})$ and $\nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z})$ are **continuous** on $\mathbb{R}^N \times \mathbb{R}^N$

If:

- $\text{argmin}_{\mathbf{z} \in \mathcal{C}} \phi(\mathbf{x}, \mathbf{z}) = \{\mathbf{z}^*\}$ is **unique**

Then:

- **$g(.)$ is differentiable with gradient:**

$$\nabla g(\mathbf{x}) = \nabla_{\mathbf{x}} \phi(\mathbf{x}, \mathbf{z}^*). \tag{11}$$

Differentiability of $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$:

Here:

- $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$
- $\nabla_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 = \mathbf{x} - \mathbf{z}$
- $\operatorname{argmin}_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2 \triangleq \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x})$ **orthogonal projection** of $\mathbf{x}$ onto set $\mathcal{C}(\mathbf{y})$.

Differentiability of $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$:

Here:

- $\mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \min_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2$
- $\nabla_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 = \mathbf{x} - \mathbf{z}$
- $\operatorname{argmin}_{\mathbf{z} \in \mathcal{C}(\mathbf{y})} \|\mathbf{x} - \mathbf{z}\|_2^2 \triangleq \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x})$ **orthogonal projection** of $\mathbf{x}$ onto set $\mathcal{C}(\mathbf{y})$.
- $\Rightarrow \mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2$ is differentiable with gradient:

$$\nabla_{\mathbf{x}} \frac{1}{2} \mathrm{d}(\mathbf{x}, \mathcal{C}(\mathbf{y}))^2 = \mathbf{x} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{x}) \tag{12}$$

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\alpha, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\alpha - \mathbf{y}\|_2^2 \quad (13)$$

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous
- Convex

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous
- Convex
- Differentiable with gradient:
  $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\alpha, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\alpha - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\alpha}\mathcal{L}(\mathbf{D}\,\alpha, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\alpha - \mathbf{y})$
- Lipschitz gradient

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous
- Convex
- Differentiable with gradient:
  $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\,\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))^2 \quad (14)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}))$
- Lipschitz gradient

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \qquad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \frac{1}{2}\,\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))^2 \qquad (14)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}))$
- Lipschitz gradient

When $\mathcal{C}(\mathbf{y}) = \{\mathbf{y}\}$ (unclipped signal), the two models are equivalent!

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\,\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha},\mathcal{C}(\mathbf{y}))^2 \quad (14)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}))$
- Lipschitz gradient

When $\mathcal{C}(\mathbf{y}) = \{\mathbf{y}\}$ (unclipped signal), the two models are equivalent!

- Generalizes the Linear Least Squares cost

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\,\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha},\mathcal{C}(\mathbf{y}))^2 \quad (14)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}))$
- Lipschitz gradient

When $\mathcal{C}(\mathbf{y}) = \{\mathbf{y}\}$ (unclipped signal), the two models are equivalent!

- Generalizes the Linear Least Squares cost
- Minimizing the proposed cost (14) is as simple as minimizing (13)

Summary/Comparison with Linear Least Squares:

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\|\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y}\|_2^2 \quad (13)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \mathbf{y})$
- Lipschitz gradient
- (Closed-form solution)

$$\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \frac{1}{2}\,\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha},\mathcal{C}(\mathbf{y}))^2 \quad (14)$$

- Continuous
- Convex
- Differentiable with gradient: $\nabla_{\boldsymbol{\alpha}}\mathcal{L}(\mathbf{D}\,\boldsymbol{\alpha},\mathbf{y}) = \mathbf{D}^T(\mathbf{D}\,\boldsymbol{\alpha} - \Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha}))$
- Lipschitz gradient

When $\mathcal{C}(\mathbf{y}) = \{\mathbf{y}\}$ (unclipped signal), the two models are equivalent!

- Generalizes the Linear Least Squares cost
- Minimizing the proposed cost (14) is as simple as minimizing (13)
- **Performing consistent sparse declipping is as simple as doing (regular) sparse coding!**

Table of Contents

Algorithm

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

Algorithm

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

Alternate minimization between sparse coefficients $\boldsymbol{\alpha}_t$ and dictionary $\mathbf{D}$:

Proposed Consistent dictionary learning algorithm:

Iterate until convergence:

Sparse coding step:

for $t = 1, \ldots, T$:

$$\boldsymbol{\alpha}_t \leftarrow \boldsymbol{\alpha}_t + \mu_1 \mathbf{D}^T(\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t) \qquad \triangleright \text{ Gradient descent step}$$

$$\boldsymbol{\alpha}_t \leftarrow \mathcal{H}_K(\boldsymbol{\alpha}_t) \qquad \triangleright \text{ Hard-thresholding}$$

Dictionary update step:

$$\mathbf{D} \leftarrow \Pi_{\mathcal{D}}\Big(\mathbf{D} + \mu_2 \sum_t (\Pi_{\mathcal{C}(\mathbf{y}_t)}(\mathbf{D}\,\boldsymbol{\alpha}_t) - \mathbf{D}\,\boldsymbol{\alpha}_t)\,\boldsymbol{\alpha}_t^T\Big) \qquad \triangleright \text{ Gradient desc.}$$

Explicit computation of projection operator

- The algorithm requires the computation of projections $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha})$ at each iteration.

Explicit computation of projection operator

- The algorithm requires the computation of projections $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha})$ at each iteration.
- The projection operator $\Pi_{\mathcal{C}(\mathbf{y})}(.)$ can be computed in closed form as:

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

Explicit computation of projection operator

- The algorithm requires the computation of projections $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha})$ at each iteration.
- The projection operator $\Pi_{\mathcal{C}(\mathbf{y})}(.)$ can be computed in closed form as:

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

- Simple elementwise maxima (negligible computational cost)

Explicit computation of projection operator

- The algorithm requires the computation of projections $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha})$ at each iteration.
- The projection operator $\Pi_{\mathcal{C}(\mathbf{y})}(.)$ can be computed in closed form as:

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

- Simple elementwise maxima (negligible computational cost)
- This also shows that the cost $\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))$ can be written explicitly as:

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

Explicit computation of projection operator

- The algorithm requires the computation of projections $\Pi_{\mathcal{C}(\mathbf{y})}(\mathbf{D}\,\boldsymbol{\alpha})$ at each iteration.
- The projection operator $\Pi_{\mathcal{C}(\mathbf{y})}(.)$ can be computed in closed form as:

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

- Simple elementwise maxima (negligible computational cost)
- This also shows that the cost $\mathrm{d}(\mathbf{D}\,\boldsymbol{\alpha}, \mathcal{C}(\mathbf{y}))$ can be written explicitly as:

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

- Equivalent to regularization-based methods
- Sparse coding step is equivalent to Consistent IHT [Kitic2013]!

Table of Contents

Experiments

- Tested on audio signals, $T = 2500$ time frames of size $N = 256$, and dictionaries of size $M = 512$.
- Signal-to-Distortion ratio (SDR), computed on the clipped samples, at different clipping levels

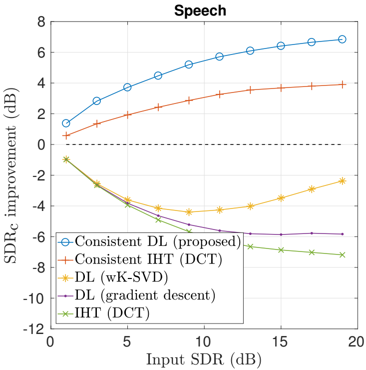


Figure: Comparison with state-of-the-art dictionary learning algorithms

Experiments

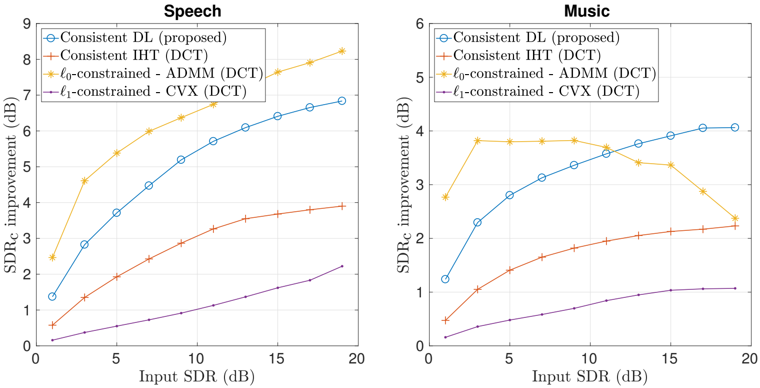


Figure: Comparison with state-of-the-art declipping algorithms

Table of Contents

Conclusion

- Re-formulate the declipping problem as minimizing the distance to a convex feasibility set
- Convex and differentiable cost function, generalizes linear least squares $\Rightarrow$ simple optimization problem.
- Consistent dictionary learning improves compared to consistent sparse coding with fixed dictionary.

Conclusion

- Re-formulate the declipping problem as minimizing the distance to a convex feasibility set
- Convex and differentiable cost function, generalizes linear least squares $\Rightarrow$ simple optimization problem.
- Consistent dictionary learning improves compared to consistent sparse coding with fixed dictionary.

Future work:

$$\begin{aligned}\mathcal{C}(\mathbf{y}) &= \{\mathbf{x} | \, \mathbf{M^r} \, \mathbf{y} = \mathbf{M^r} \, \mathbf{x}, \mathbf{M^{c+}} \, \mathbf{x} \succeq \mathbf{M^{c+}} \, \mathbf{y}, \mathbf{M^{c\text{-}}} \, \mathbf{x} \preceq \mathbf{M^{c\text{-}}} \, \mathbf{y}\} \\ &= \{\mathbf{x} | f(\mathbf{x}) = \mathbf{y}\} \\ &= f^{-1}(\mathbf{y})\end{aligned}$$

where f is the nonlinear clipping function.

Conclusion

- Re-formulate the declipping problem as minimizing the distance to a convex feasibility set
- Convex and differentiable cost function, generalizes linear least squares ⇒ simple optimization problem.
- Consistent dictionary learning improves compared to consistent sparse coding with fixed dictionary.

Future work:

$$\begin{aligned}\mathcal{C}(\mathbf{y}) &= \{\mathbf{x} | \, \mathbf{M}^{\mathbf{r}} \, \mathbf{y} = \mathbf{M}^{\mathbf{r}} \, \mathbf{x}, \mathbf{M}^{\mathbf{c}+} \, \mathbf{x} \succeq \mathbf{M}^{\mathbf{c}+} \, \mathbf{y}, \mathbf{M}^{\mathbf{c}\text{-}} \, \mathbf{x} \preceq \mathbf{M}^{\mathbf{c}\text{-}} \, \mathbf{y}\} \\ &= \{\mathbf{x} | f(\mathbf{x}) = \mathbf{y}\} \\ &= f^{-1}(\mathbf{y})\end{aligned}$$

where f is the nonlinear clipping function.

- extend the proposed method to other nonlinear functions (ex: quantization, 1-bit sensing)

Conclusion

- Re-formulate the declipping problem as minimizing the distance to a convex feasibility set
- Convex and differentiable cost function, generalizes linear least squares ⇒ simple optimization problem.
- Consistent dictionary learning improves compared to consistent sparse coding with fixed dictionary.

Future work:

$$\begin{aligned}\mathcal{C}(\mathbf{y}) &= \{\mathbf{x}|\, \mathbf{M^r}\,\mathbf{y} = \mathbf{M^r}\,\mathbf{x}, \mathbf{M^{c+}}\,\mathbf{x} \succeq \mathbf{M^{c+}}\,\mathbf{y}, \mathbf{M^{c\text{-}}}\,\mathbf{x} \preceq \mathbf{M^{c\text{-}}}\,\mathbf{y}\} \\ &= \{\mathbf{x}|f(\mathbf{x}) = \mathbf{y}\} \\ &= f^{-1}(\mathbf{y})\end{aligned}$$

where f is the nonlinear clipping function.

- extend the proposed method to other nonlinear functions (ex: quantization, 1-bit sensing)
- extend other sparse coding/dictionary learning algorithms to optimize the proposed cost function

Consistent dictionary learning for signal declipping
L. Rencker, F. Bach, W. Wang, Mark D. Plumbley

Code and audio examples available at:

`http://www.cvssp.org/Personal/LucasRencker/software.html`

or:

`https://github.com/LucasRr`

MacSeNet
Machine Sensing Training Network

UNIVERSITY OF SURREY

Consistent dictionary learning for signal declipping
L. Rencker, F. Bach, W. Wang, Mark D. Plumbley

Code and audio examples available at:
http://www.cvssp.org/Personal/LucasRencker/software.html
or:
https://github.com/LucasRr

Thank you!