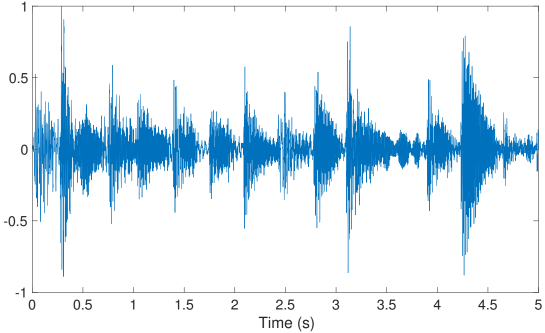
Audio declipping

Matthieu Kowalski

Univ Paris-Sud
L2S (GPI)

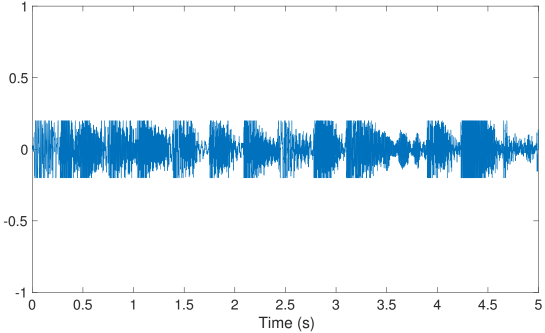
Audio Declipping

Original signal:

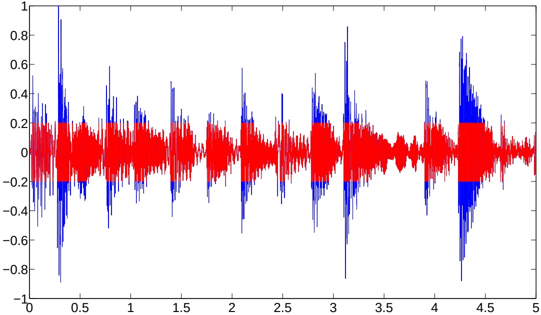


Audio Declipping

Clipped signal:



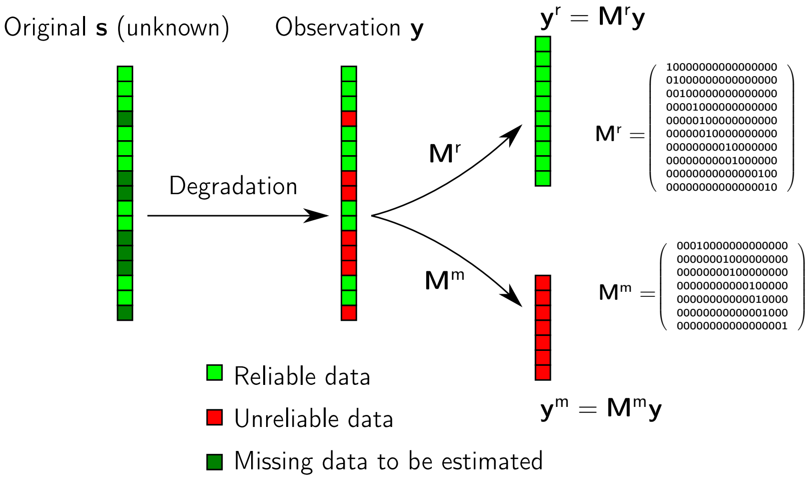
Audio Declipping



Goal:

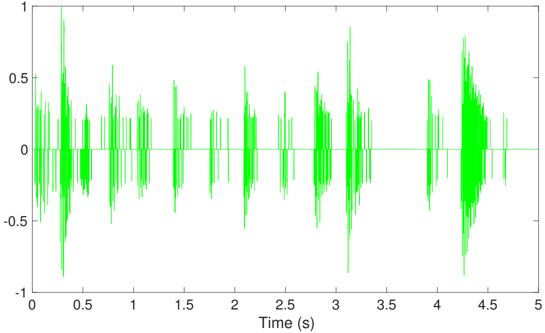
Can we get a good estimation of the original signal (blue) from the clipped one (red) ?

Reliable vs Unreliable coeff.



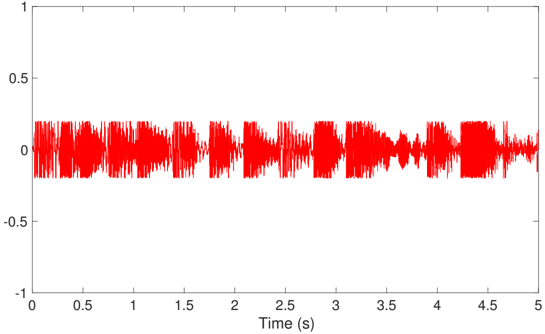
Reliable vs Unreliable coeff.

Reliable samples: $\mathbf{y}^r = \mathbf{M}^r\mathbf{y}$



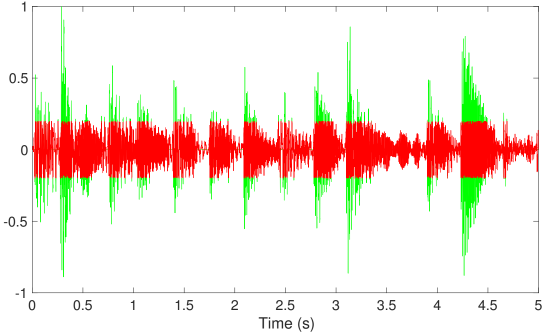
Reliable vs Unreliable coeff.

Unreliable (clipped) samples: $\mathbf{y}^m = \mathbf{M}^m \mathbf{y}$



Reliable vs Unreliable coeff.

Reliable + Unreliable (clipped) samples:



Audio inpainting: forward problem [A. Adler, V. Emiya *et Al*]

We have then:

$$\mathbf{y}^r = \mathbf{M}^r \mathbf{y} = \mathbf{M}^r \mathbf{s}$$

where

- $\mathbf{s} \in \mathbb{R}^N$ is the unknown "clean" signal;
- $\mathbf{y}^r \in \mathbb{R}^M$ are the "reliable" sample of the observed signal
- $\mathbf{M}^r \in \mathbb{R}^{M \times N}$ is the matrix of the reliable support of $\mathbf{x}$

we can also define the "missing" samples as

$$\mathbf{y}^m = \mathbf{M}^m \mathbf{y} = \mathbf{M}^m \mathbf{s}$$

Inverse problem: data term

Using the reliable coefficients, we must have

$$\mathbf{y}^r = \mathbf{M}^r \mathbf{s}$$

where $\mathbf{M}^r$ select the reliable samples. We can use a simple ℓ_2 loss

$$\hat{\mathbf{s}} = \operatorname*{argmin}_{\mathbf{s}} \frac{1}{2} \|\mathbf{y}^r - \mathbf{M}^r \mathbf{s}\|_2^2$$

We must take the clipped samples into account

Inverse problem: clipping constraints

For audio declipping, we can add the following constraint

$$\hat{\mathbf{s}} = \operatorname*{argmin}_{\mathbf{s}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\mathbf{s}\|_2^2$$

$$\text{s.t.} \quad \mathbf{M}^{m^+}\mathbf{\Phi}\boldsymbol{\alpha} > \theta^{clip}$$

$$\mathbf{M}^{m^-}\mathbf{\Phi}\boldsymbol{\alpha} < -\theta^{clip}$$

where

- $\mathbf{M}^{m^+}$ (resp. $\mathbf{M}^{m^-}$) select the positive (resp. negative) **clipped** samples.
- θ^{clip} is the clip threshold (here $\theta^{clip} = 0.2$)

Problem: infinite solutions! We must add some constraints on $\mathbf{s}$

Audio declipping: use a dictionnary

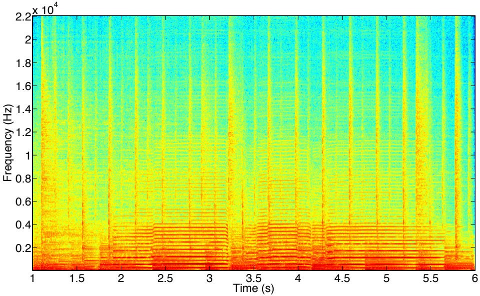
Let $\mathbf{\Phi}$ a dictionnary such that:

$$\mathbf{s} = \mathbf{\Phi}\boldsymbol{\alpha}$$

where $\boldsymbol{\alpha}$ are **sparse** synthesis coefficients

Audio signal: use the short time Fourier transform

$$s(t) = \mathbf{\Phi}\boldsymbol{\alpha} = \sum_{n,f} \alpha_{n,f}\varphi_{n,f}(t)$$



Inverse problem: a constrained sparse problem

Using the dictionnary $\mathbf{\Phi}$ + sparsity

$$\hat{\boldsymbol{\alpha}} = \underset{\mathbf{s}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\mathbf{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

$$\text{s.t.} \quad \mathbf{M}^{m^+}\mathbf{\Phi}\boldsymbol{\alpha} > \theta^{clip}$$

$$\mathbf{M}^{m^-}\mathbf{\Phi}\boldsymbol{\alpha} < -\theta^{clip}$$

where

- $\mathbf{M}^{m^+}$ (resp. $\mathbf{M}^{m^-}$) select the positive (resp. negative) **clipped** samples.
- θ^{clip} is the clip threshold (here $\theta^{clip} = 0.2$)
- $\hat{\mathbf{s}} = \hat{\boldsymbol{\alpha}}$

Problems:

- the proximity operator has no closed form
- Cannot use simple algorithms such as (F)ISTA

Rewrite the constraints

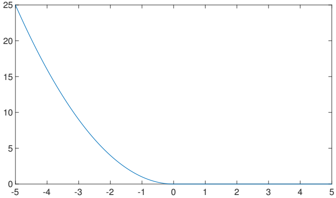
Idea: use a ℓ_2 loss on the clipped samples **if** the constraint is not respected

If $y^m(t) > \theta^{clip}$

$$\textbf{then } \mathcal{L}(\theta^{clip} - y^m(t)) = 0$$

If $\hat{y}^m(t) < \theta^{clip}$

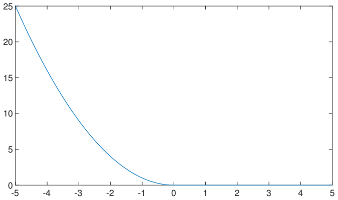
$$\textbf{else } \mathcal{L}(\theta^{clip} - y^m(t)) = (\theta^{clip} - y^m(t))^2$$



Rewrite the constraints

The squared hinge loss:

$$\begin{aligned}
\mathcal{L}(\theta^{clip} - \mathbf{y}^m) &= [\theta^{clip} - \mathbf{y}^m]_+^2 \\
&= \sum_{t:y^m(t)>0} (\theta^{clip} - y^m(t))_+^2 + \sum_{t:y^m(t)<0} (-\theta^{clip} + y^m(t))_+^2 \\
&= [\theta^{clip} - \mathbf{M}^m \boldsymbol{\Phi} \boldsymbol{\alpha}]_+^2
\end{aligned}$$



Audio declipping: (convex unconstrained) inverse problem

We consider the following **unconstrained** convex problem:

$$\boldsymbol{\alpha} = \operatorname*{argmin}_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r \boldsymbol{\Phi} \boldsymbol{\alpha}\|_2^2 + \frac{1}{2}[\boldsymbol{\theta}^{clip} - \mathbf{M}^m \boldsymbol{\Phi} \boldsymbol{\alpha}]_+^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

which is under the form

$$f_1(\boldsymbol{\alpha}) + f_2(\boldsymbol{\alpha})$$

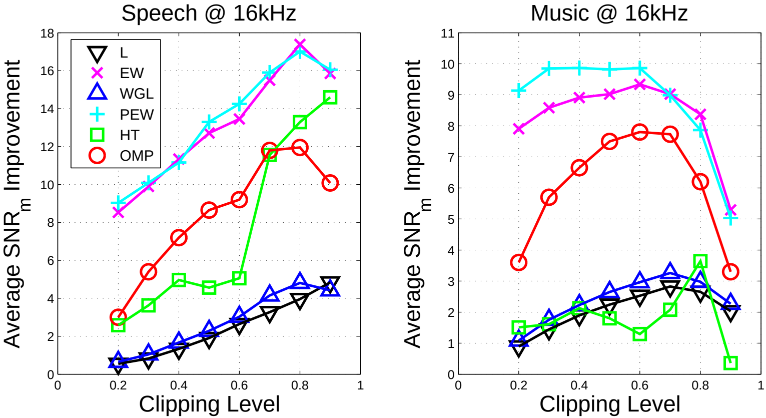
with f_1 Lipschitz-differentiable and f_2 semi-convex.

We can apply (relaxed)-ISTA directly !

FISTA for declipping

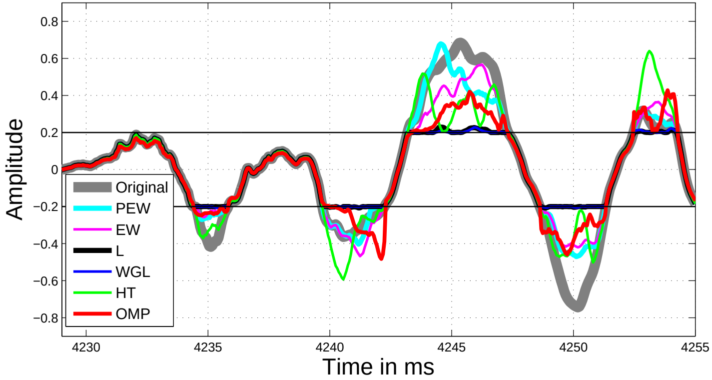
Thresolding operators

Numerical results



Average SNR$_{miss}$ for 10 speech (left) and music (right) signals over different clipping levels and operators. Neighborhoods extend 3 and 7 coefficients in time for speech and music signals, respectively.

Numerical results: zoom on reconstructions



Declipped music signal using different operators for clip level $\theta^{clip} = 0.2$ using the Lasso, WGL, EW, PEW, HT, and OMP operators. Neighborhood size for WGL and PEW was 7.

Original Vs clipped Vs declipped Signal