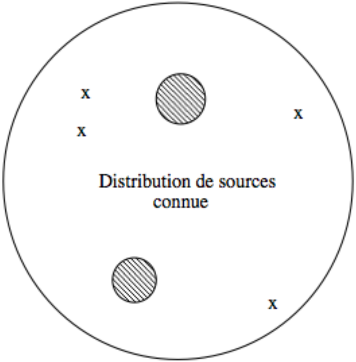
Inverse problems: a sparse synthesis approach.

Matthieu Kowalski

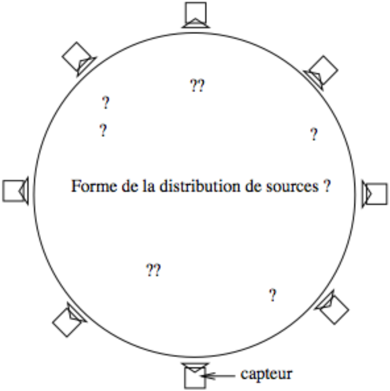
Univ Paris-Sud
L2S (GPI)

April 9, 2014

Inverse problem : *a contrario definition*

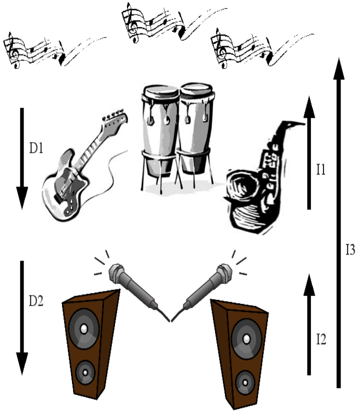


Direct problem



Inverse problem

Example: audio inverse problems



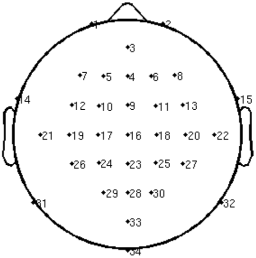
Direct problems

1. instrument synthesis
2. signal mixtures

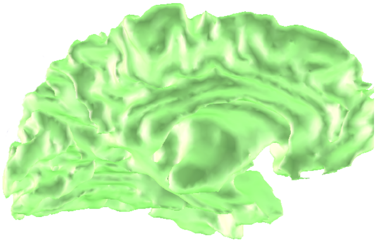
Inverse problems

1. automatique transcription
2. source separation
3. audio restauration

Example: MEEG inverse problem



? ⇓?



How to localize neuronal sources from M/EEG records ?

Inverse problem: formalization

$\mathcal{A}($	$\mathbf{y},$	$\mathbf{x},$	$\boldsymbol{\alpha},$	$\mathbf{b})$	$= 0$
model	observations/measures	unknown signal	hidden coefficients	errors and noise	

Explicit relation :	$\mathbf{y} = \mathcal{A}(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{b})$
Output error:	$\mathbf{y} = \mathcal{A}(\mathbf{x}, \boldsymbol{\alpha}) \diamond \mathbf{b}$
Additive error :	$\mathbf{y} = \mathcal{A}(\mathbf{x}, \boldsymbol{\alpha}) + \mathbf{b}$
Relation between $\mathbf{x}$ et $\boldsymbol{\alpha}$:	$\begin{cases} \mathbf{y} = \mathcal{A}_1(\mathbf{x}, \boldsymbol{\alpha}) + \mathbf{b} \\ \mathcal{A}_2(\mathbf{x}, \boldsymbol{\alpha}) = 0 \end{cases}$
nonlinear model:	$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b}$
Linear model :	$\mathbf{y} = \mathcal{A}\mathbf{x} \diamond \mathbf{b}$
Linear model + additive noise:	$\mathbf{y} = \mathcal{A}\mathbf{x} + \mathbf{b}$

Well posed inverse problem

A problem is well-posed in Hadamard sense [Hadamard 1923] if the following holds :

1. Existence: there is at least one solution.
2. Uniqueness: the set of solutions converge to a unique solution.
3. Stability: the solution depends continuously on the measurements.

- The problem is *overdetermined* if there are more measurements than sources
- The problem is *underdetermined* if one looks for more sources than measurements

Frameworks

Mathematical framework

- $\mathbf{y} \in \mathbb{R}^M$
- $\mathbf{x} \in \mathbb{R}^N$
- $A \in \mathbb{R}^{M.N}$

Optimization framework

$$\mathbf{x} = \operatorname{argmin} \mathcal{L}(\mathbf{y}, A, \mathbf{x}) + P(\mathbf{x}; \lambda)$$

1. A convex loss or data term $\mathcal{L}(\mathbf{y}, A, \mathbf{x})$ measuring the fit between the observed mixture $\mathbf{y}$ and the source signal $\mathbf{x}$ given the mixing system A;
2. A regularization term P modeling the assumptions about the sources,
3. An hyperparameter $\lambda \in \mathbb{R}_+$ governing the balance between the data term and the regularization term.

The Loss

Traditional assumption: Gaussian noise

$$\mathcal{L}(\mathbf{y}, A, \mathbf{x}) = \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_2^2$$

But other possible choices

- Impulsive noise:

$$\mathcal{L}(\mathbf{y}, A, \mathbf{x}) = \frac{1}{2}\|\mathbf{y} - A\mathbf{x}\|_1$$

- Poisson noise:

$$\mathcal{L}(\mathbf{y}, A, \mathbf{x}) = A\mathbf{x} - \mathbf{y} + \mathbf{y}\ln\left(\frac{\mathbf{y}}{A\mathbf{x}}\right)$$

The Penalty

Goal: Model the prior on the sources.

"Analysis" prior

Models the "physical" assumptions on the sources

- Minimum energy : $\frac{1}{2}\|\mathbf{x}\|_2^2$ [Tikhonov, 77]
- Total variation (images) : $\|\nabla \mathbf{x}\|_1$ [ROF, 92]

Sometimes, we need more flexibility: priors are not always in the "samples" domain

Optimization framework with dictionary

1. A Dictionary $\mathbf{\Phi}$
2. A convex loss or data term $\mathcal{L}(\mathbf{y}, A, \boldsymbol{\alpha})$ measuring the fit between the observed mixture $\mathbf{y}$ and some synthesis coefficients $\boldsymbol{\alpha}$, such that $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}$, given the mixing system A;
3. A regularization term P modeling the assumptions about the sources, in the synthesis coefficient domain
4. An hyperparameter $\lambda \in \mathbb{R}_+$ governing the balance between the data term and the regularization term.

The Dictionary

Synthesis point of view

Assume $\mathbf{x}$ can be written as

$$\begin{aligned}\mathbf{x} &= \sum_{k=1}^{K} \alpha_k \boldsymbol{\varphi}_k \\ &= \boldsymbol{\Phi}\boldsymbol{\alpha}\end{aligned}$$

with

$$\boldsymbol{\Phi} \in \mathbb{C}^{N.K}, \quad k \geq N$$

Examples

- Gabor
- wavelets
- Union of Gabor (hybrid model or Morphological Component Analysis): $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 = \boldsymbol{\Phi}_1\boldsymbol{\alpha}_1 + \boldsymbol{\Phi}_2\boldsymbol{\alpha}_2$
- Frames ([Balazs *et al.*, 2013])

The penalty (returns)

Sparse approximation: key idea
$\mathbf{x} \in \mathbb{R}^N$ admits a sparse decomposition inside a dictionnary of waveforms $\{\varphi_k\}_{k=1}^K$:

$$\mathbf{x} = \sum_{k \in \Lambda} \alpha_k \boldsymbol{\varphi}_k$$

with $\Lambda \subset \{1, \ldots, K\}$

Given a (noisy) observation $\mathbf{y} = A\mathbf{x} + \mathbf{n}$, the Lasso/Basis Pursuit Denoising [Tibshirani, 96], [Chen *et al.* 98] estimate reads:

$$\hat{\boldsymbol{\alpha}} = \operatorname*{argmin}_{\boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y} - A\boldsymbol{\Phi}\boldsymbol{\alpha}\|^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

and

$$\hat{\mathbf{x}} = \boldsymbol{\Phi}\hat{\boldsymbol{\alpha}}$$

The penalty (returns)

Structured penalties

- Structured sparsity via mixed norm [K,Torrésani 2008], [K, 2009]:
  - Group-Lasso [Yuan, Lin 2006]
    $P(\boldsymbol{\alpha}; \lambda) = \lambda \|\boldsymbol{\alpha}\|_{2;1} = \lambda \sum_g \sqrt{\sum_m |\alpha_{g,m}|^2}$
  - Elitist-Lasso [K,Torrésani 2008]
    $P(\boldsymbol{\alpha}; \lambda) = \lambda \|\boldsymbol{\alpha}\|_{1;2}^2 = \lambda \sum_g \left(\sum_m |\alpha_{g,m}|\right)^2$
- Hi-Lasso [Jenatton *et al.* 2011], [Sprechmann *et al.* 2011]
  $P(\boldsymbol{\alpha}; \lambda) = \lambda \left((1-\nu)\|\boldsymbol{\alpha}\|_{2;1} + \nu\|\boldsymbol{\alpha}\|_1\right)$
- sub-modular functions etc. [Bach 2012]

$$\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\alpha}}_2 = \operatorname*{argmin}_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{y} - A(\boldsymbol{\Phi}_1\boldsymbol{\alpha}_1 + \boldsymbol{\Phi}_2\boldsymbol{\alpha}_2\|^2 + P(\boldsymbol{\alpha}_1; \lambda_1) + P(\boldsymbol{\alpha}_2; \lambda_2)$$

and

$$\hat{\mathbf{x}} = \boldsymbol{\Phi}_1\hat{\boldsymbol{\alpha}}_1 + \boldsymbol{\Phi}_2\hat{\boldsymbol{\alpha}}_2$$

Proximity operators

we suppose that Φ is *orthogonal*. We denote by $\tilde{y} = \Phi^T y$

LASSO solution $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_1$

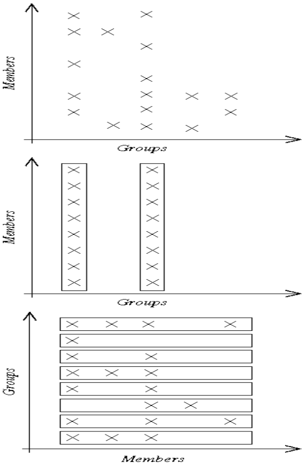
$$\hat{\alpha}_{g,m} = \arg(\tilde{y}_{g,m}) \left(|\tilde{y}_{g,m}| - \lambda\right)^+$$

G-LASSO solution $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_{2,1}$

$$\hat{\alpha}_{g,m} = \tilde{y}_{g,m} \left(1 - \frac{\lambda}{\|\tilde{y}_g\|_2}\right)^+$$

E-LASSO solution $\min_{\alpha} \|y - \Phi\alpha\|_2^2 + \lambda\|\alpha\|_{1,2}^2$

$$\hat{\alpha}_{g,m} = \arg(\tilde{y}_{g,m}) \left(|\tilde{y}_{g,m}| - \frac{\lambda}{1 + \lambda L_g} \|\|\tilde{y}_g\|\|\right)^+$$



(Relaxed) ISTA

- Let $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$, $L \geq \frac{1}{\|\boldsymbol{\Phi}^*\boldsymbol{\Phi}\|}$, $0 \leq \mu < 1$, and $t_{max} \in \mathbb{N}$.
- **For** $t = 0$ to t_{max}

$$\begin{aligned}
\boldsymbol{\alpha}^{(t+1/2)} &= \boldsymbol{\gamma}^{(t)} + \boldsymbol{\Phi}^*(\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\gamma}^{(t)})/L \\
\boldsymbol{\alpha}^{(t+1)} &= \mathbb{S}(\boldsymbol{\alpha}^{(t+1/2)}, \lambda/L) \\
\boldsymbol{\gamma}^{t+1} &= \boldsymbol{\alpha}^{(t+1)} + \mu^{(t+1)}(\boldsymbol{\alpha}^{(t+1)} - \boldsymbol{\alpha}^{(t)})
\end{aligned}$$

End For

with $\mathbb{S}$ a proximity operator (soft thresholding for ℓ_1).

Convergence proved by several authors

- [Combettes & Wajs 05] forward-backward (proximity operators);
- [Daubechies & al 04] Opial's fixed point theorem;
- [Figuereido & Nowak 03] EM algorithm;

Accelerated version by [Nesterov 07], [Beck & Teboulle 09] (FISTA).

Limitations

- Biased coefficients: large coefficients are shrinked [Gao, Bruce 97]
- Lake of flexibility for structures: needs to define an adequate convex penalty (not always simple)

Could we play directly on the thresholding step ?

Thresholding rules

Definition [Antoniadis 07]

1. $\mathbb{S}(.;\lambda)$ is an odd function. ($\mathbb{S}_+(.;\lambda)$ is used to denote the $\mathbb{S}(.;\lambda)$ restricted to $\mathbb{R}_+$.)
2. $\mathbb{S}(.;\lambda)$ is a shrinkage rule: $0 \leq \mathbb{S}_+(t;\lambda) \leq t, \ \forall t \in \mathbb{R}_+$.
3. $\mathbb{S}_+$ is nondecreasing on $\mathbb{R}_+$, and $\lim\limits_{t \to +\infty} \mathbb{S}(t;\lambda) = +\infty$

Examples

- Soft [Donoho, Johnstone 94]

$$\mathbb{S}(x;\lambda) = x\left(1 - \frac{\lambda}{|x|}\right)^+$$

- Hard [Donoho, Johnstone 94]

$$\mathbb{S}(x;\lambda) = x\mathbf{1}_{|x|>\lambda}$$

- NonNegativeGarrote (NNGarrote) [Gao 98]

$$\mathbb{S}(x;\lambda) = x(1 - \frac{\lambda}{|x|^2})^+$$

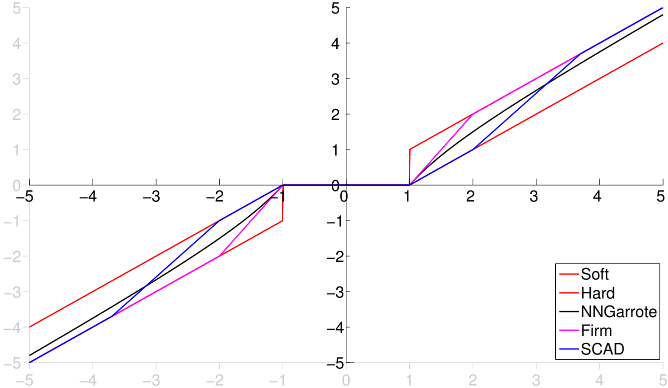
- Firm [Gao, Bruce 97]

$$\mathbb{S}(x;\lambda_1;\lambda 2) = \begin{cases} 0 & \text{if } |x| < \lambda_1 \\ \frac{x\lambda_2\left(1-\frac{\lambda_1}{|x|}\right)}{\lambda_2-\lambda 1} & \text{if } \lambda_1 \leq |x| < \lambda_2 \\ x & |x| > \lambda_2 \end{cases}$$

- SCAD [Antoniadis, Fan 01]

$$\mathbb{S}(x;\lambda;a) = \begin{cases} x(1 - \frac{\lambda}{|x|})^+ & \text{if } |x| < 2\lambda \\ \frac{x\left(a-1-\frac{a\lambda}{|x|}\right)}{a-2} & \text{if } 2\lambda \leq |x| < a\lambda \\ x & \text{if } |x| > a\lambda \end{cases}$$

Examples



Properties of Thresholding rules

Definition: semi-convex fonction

A function f is said to be semi-convex, iff there exists c such that

$$x \mapsto f(x) + \frac{c}{2}\|x\|^2$$

is convex

Proposition

We can associate a semi-convex penalty $P(.;\lambda)$, with $c \leq 1$ to any thresholding rules. Moreover, $\frac{1}{1-c}$ is an upper-bound of $\mathbb{S}'(.;\lambda)$.

Convergence results

Theorem

- ISTA converges with any thresholding rules
- Relaxed ista converges for $0 \leq \mu < 1 - c$

Examples

- NNGarrote ($c = 1/2$)

$$P(x;\lambda) = \lambda^2 + \operatorname{asinh}\left(\frac{|x|}{2\lambda}\right) + \lambda^2 \frac{|x|}{\sqrt{x^2 + 4\lambda^2} + |x|}$$

- SCAD ($c = a - 1$)

$$P(x;\lambda) = \begin{cases} \lambda x & \text{if } x \leq \lambda \\ \frac{(a\lambda x - x^2/2)}{a-1} & \text{if } \lambda < x \leq a\lambda \\ a\lambda & \text{if } x > a\lambda \end{cases}$$

1. Introduction: inverse problems
   - Exemples et formalisation
   - Formalisation
2. An optimization framework
3. Iterative Thresholding
   - Thresholding functions
   - Neighborhood thresholding
4. Numerical results
   - Audio declipping
5. Conclusion

Windowed Group-LASSO

Back to the model $\mathbf{y} = \mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{b}$, with $\mathbf{\Phi}$ orthonormal. Back to a simple indexing, and for each index k, we define a neighborhood $g(k)$.

Windowed G-Lasso [MK & BT 09], [K *et al.* 13]

$$\hat{\alpha}_k = \tilde{y}_k \left(1 - \frac{\lambda}{\sqrt{\sum_{m \in g(k)} |\tilde{y}_m|^2}}\right)^+$$

$$= \tilde{y}_k \left(1 - \frac{\lambda}{\|\tilde{y}_{g(k)}\|_2}\right)^+$$

with $\tilde{y} = \Phi^* y$

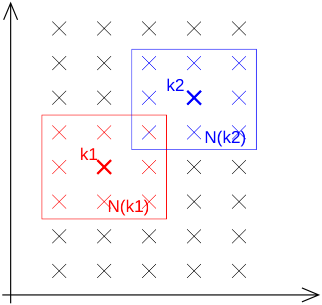


FIGURE : *WG-LASSO. Two overlapping groups: neighborhood of k_1 and k_2.*

Similar thresholding rules introduced by [Cai & Silvermanss 01] for wavelet thresholding.

Neighborhood with latents variables

Can we define the WG-Lasso by using proximity operator ?

thanks to the following strategy

- map the original coefficients into a bigger space;
- define independent groups over the neighborhood of the coefficients;
- apply the (group-lasso) proximity operator;
- go back to the original space.

Moreover, can we use the WG-Lasso inside ISTA ?

Expended operators

Definition : Expanding operator

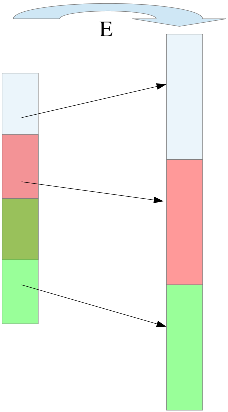
Let $\boldsymbol{\alpha} \in \mathbb{C}^N$. Let $\mathbf{E} : \mathbb{C}^N \to \mathbb{C}^{N\times N}$ be an expanded operator such that

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N) \mapsto (w_1^1\alpha_1, w_2^1\alpha_2, \dots, w_N^1\alpha_N, \dots, w_1^N\alpha_1, \dots, w_N^N\alpha_N)$$

with $w_i^j \geq 0$, $\sum_j |w_i^j|^2 = 1$ and $w_i^i > 0$

proposition

$\mathbf{E}$ is isometrical, and then $\mathbf{E}^T\mathbf{E} = \mathbf{I}$.



A left inverse

Definition : a natural left inverse

$$\begin{aligned} \mathbf{D} \ &: \ \mathbb{C}^{N\times N} \to \mathbb{C}^N \\ \mathbf{z} = (z_1^1, \dots, z_N^1, \dots, z_1^N, \dots, z_N^N) &\mapsto \mathbf{x} \\ \text{such that } \forall k,\ x_k &= \frac{1}{w_k^k} z_k^k \end{aligned} \tag{1}$$

$\mathbf{DE} = \mathbf{I}$ and then $\mathbf{DE}$ is a bi-orthogonal (oblique) projection.

Structured shrinkage and proximity operators

proposition

Let $\mathbb{S}$ be the shrinkage operator of the WG-Lasso and $\Omega = \|.\|_{21}$ the regularizer of the G-lasso. Let $\mathbf{E}$ be the expanded operator as previously defined and $\mathbf{D}$ its left inverse. Then

$$\mathbb{S}(., \lambda) = \mathbf{D} \circ \mathrm{prox}_{\lambda\Omega} \circ \mathbf{E}$$

$$\hat{\alpha}_k = \tilde{y}_k \left(1 - \frac{\lambda}{\sqrt{\sum_{m \in g(k)} |\tilde{y}_m|^2}} \right)^+ = \tilde{y}_k \left(1 - \frac{\lambda}{\|\tilde{y}_{g(k)}\|_2} \right)^+$$

Remark

$\mathbb{S}$ cannot be a proximity operator (it is even not a nonexpansive operator).

Neighborhood as a convex prior

social sparsity convex regularizers

Let $\boldsymbol{\alpha} \in \mathbb{C}^N$ and let $\mathbf{E}$ be the expanded operator.

cvx windowed group lasso:

$$\Omega_{wgl}(\boldsymbol{\alpha}) = \sum_{k=1}^{N} \sqrt{\sum_{\ell \in \mathcal{N}(k)} w_\ell^{(k)} |\alpha_\ell|^2} = \|\mathbf{E}\boldsymbol{\alpha}\|_{21}$$

cvx windowed elitist lasso:

$$\Omega_{wel}(\boldsymbol{\alpha}) = \sum_{k=1}^{N} \left(\sum_{\ell \in \mathcal{N}(k)} w_\ell^{(k)} |\alpha_\ell| \right)^2 = \|\mathbf{E}\boldsymbol{\alpha}\|_{12}^2$$

A convex functional for social sparsity

A natural convex functional is (aka group-Lasso with overlaps [Bayram 11])

$$F(\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|^2 + \lambda\|\mathbf{E}\boldsymbol{\alpha}\|_{21}$$

one can look for

$$\begin{aligned}\hat{\alpha} &= \operatorname*{argmin}_{\boldsymbol{\alpha}\in\mathbb{C}^N} F(\boldsymbol{\alpha}) \\ &= \mathbf{E}^T \operatorname*{argmin}_{\mathbf{u}} \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{E}^T\mathbf{z}\|^2 + \lambda\|\mathbf{z}\|_{21} \\ &\quad \text{s.t } \mathbf{E}\mathbf{E}^T\mathbf{z} = \mathbf{z}\end{aligned}$$

- Similar functional introduced by [Peyré & Fadili 11].
- several approach can be used to minimize F (ISTA + Douglas Rachford, augmented lagrangian...)

But: this penalty acts as a *discarding* procedure, not a *selection*.

G-Lasso with overlaps VS latent-G-Lasso

Instead of

$$F(\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|^2 + \lambda\|\mathbf{E}\boldsymbol{\alpha}\|_{21}$$

[Jacob & al. 09] propose to minimize

$$F(\tilde{\boldsymbol{\alpha}}) = \frac{1}{2}\|\mathbf{y} - \boldsymbol{\Phi}\mathbf{E}^T\tilde{\boldsymbol{\alpha}}\|^2 + \lambda\|\tilde{\boldsymbol{\alpha}}\|_{21}$$

to obtain a *selection* of active groups.

Curse of dimension in both cases !

Link between the convex functional and our shrinkages

ISTA with WG-Lasso becomes:

$$\mathbf{z}^{(k)} = \mathbf{ED}\ \mathrm{prox}_{\frac{\lambda}{\gamma}\|.\|_*}\left(\left(\tilde{\mathbf{z}}^{(k-1)}\right)\right)$$

$$\boldsymbol{\alpha}^k = \mathbf{D}\mathbf{z}^k$$

$$\text{where} \quad \tilde{\mathbf{z}}^{(k-1)} = \mathbf{z}^{(k-1)} + \frac{\mathbf{E}}{\gamma}\boldsymbol{\Phi}^*(\mathbf{y} - \boldsymbol{\Phi}\mathbf{E}^T\mathbf{z}^{(k-1)})$$

It is a proximal descent followed by an *oblique* projection on $\mathrm{Im}(E)$.

conjecture

ISTA with WG-Lasso converges to a fixed point.

Orthogonal social sparsity

An Orthogonal version

$$\mathbf{z}^{(k)} = \mathbf{E}\mathbf{E}^T \operatorname{prox}_{\frac{\lambda}{\gamma}\|.\|_*}\left(\left(\tilde{\mathbf{z}}^{(k-1)}\right)\right)$$

$$\text{where} \quad \tilde{\mathbf{z}}^{(k-1)} = \mathbf{z}^{(k-1)} + \frac{\mathbf{E}}{\gamma}\mathbf{\Phi}^*(\mathbf{y} - \mathbf{\Phi}\mathbf{E}^T\mathbf{z}^{(k-1)})$$

orth-WG-Lasso

$$\alpha_k = \tilde{y}_k \sum_j \frac{1}{w_j^j}\left(1 - \frac{\lambda}{\sqrt{\sum_{j' \in \mathcal{N}(j)} w_{j'}^{(j)} |\tilde{y}_{k'}|^2}}\right)^+ .$$

$$\text{WG-Lasso}: \quad \hat{\alpha}_k = \tilde{y}_k \left(1 - \frac{\lambda}{\sqrt{\sum_{m \in g(k)} |\tilde{y}_m|^2}}\right)^+$$

A family of shrinkage operators

$\boldsymbol{\alpha} = \mathbb{S}(\mathbf{y})$ is given coordinatewise:

- Lasso:

$$\alpha_k = y_k \left(1 - \frac{\lambda}{|y_k|}\right)^+$$

- NNGarrote / Empirical Wiener

$$\alpha_k = y_k \left(1 - \frac{\lambda}{|y_k|^2}\right)^+$$

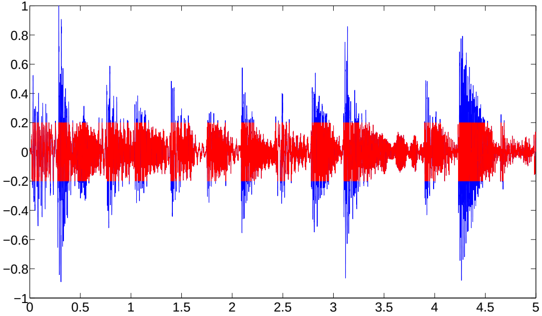
- Windowed Group Lasso

$$\alpha_k = \tilde{y}_k \left(1 - \frac{\lambda}{\|\tilde{y}_{g(k)}\|_2}\right)^+$$

- Empirical Persistent Wiener [Siedenburg 13]

$$\alpha_k = \tilde{y}_k \left(1 - \frac{\lambda}{\|\tilde{y}_{g(k)}\|_2^2}\right)^+$$

Audio Declipping



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

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

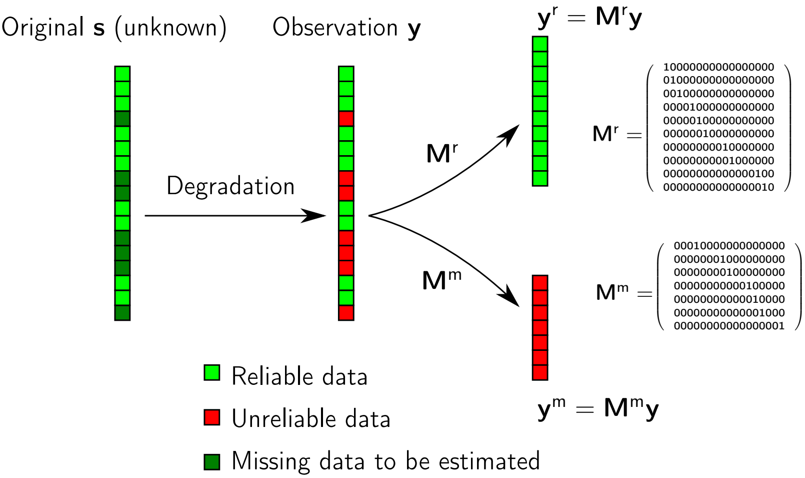
where

- $\mathbf{x} \in \mathbb{R}^N$ 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

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

Reliable vs Unreliable coeff.



Audio declipping: (constrained and convex) inverse problem

For audio declipping, we can add the following constraint

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

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

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

where $\mathbf{M}^{m^+}$ (resp. $\mathbf{M}^{m^-}$) select the positive (resp. negative) samples.

Problem: cannot be solved “efficiently” with (F)ISTA

Audio declipping: (convex unconstrained) inverse problem

Let

$$[\boldsymbol{\theta}^{clip} - \mathbf{x}]_+^2 = \sum_{k:\theta_k^{clip}>0} (\theta_k^{clip} - x_k)_+^2 + \sum_{k:\theta_k^{clip}<0} (-\theta_k^{clip} + x_k)_+^2$$

We consider the following unconstrained convex problem:

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

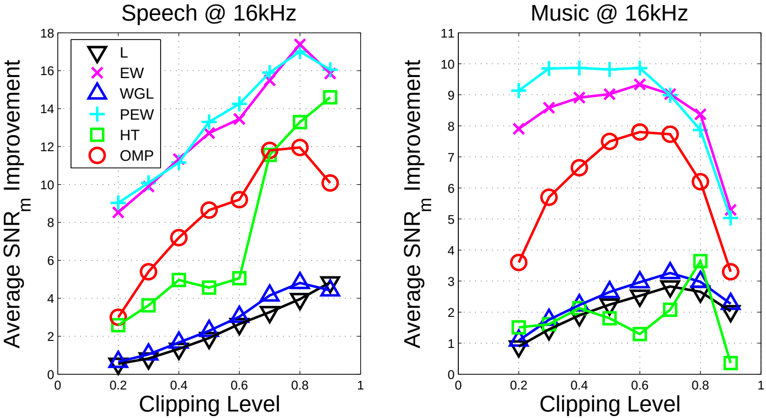
which is under the form

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

with f_1 Lipschitz-differentiable and f_2 semi-convex.

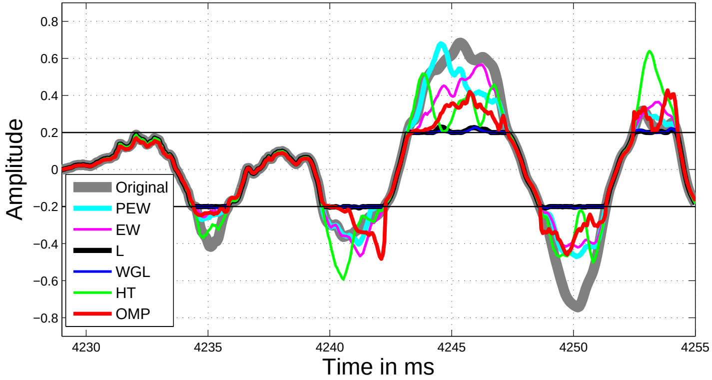
We can apply (relaxed)-ISTA directly !

Numerical results



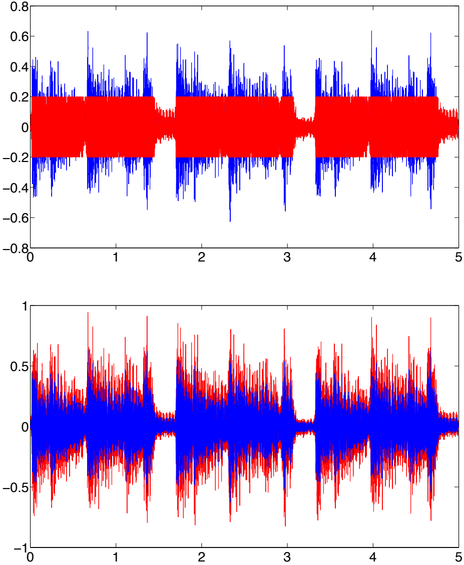
Average SNR$_{miss}$ for 10 speech (left) and music (right) signals over different clipping levels and operators. Neighborhoods extend 3 and 7 coefficients in time for speech and music signals, respectively.

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



Conclusion

Take home messages

- Use dictionary to get sparsity
- Play on thresholding rules in ISTA
- Define some neighborhoods for “flexible” structures

Next. . .

- Some practical issues (warm start: how many iterations, λ)
- Some theoretical issues (more on convergence)