Enhancing sparsity by reweighted ℓ_1 minimization

Emmanuel Candès - Mickael Wakin - Stephen Boyd

Presented by Jérome Lesaint

Thursday 5th June 2014

Sparse recovery : Basics I

The problem under study

Recover original signal $x_0 \in \mathbb{R}^n$ from measurements $y \in \mathbb{R}^m$ where :

$$y = \phi x_0$$

ϕ being an $m \times n$ matrix with $m < n$

This problem has, of course, infinitely many solutions.
Under additional sparsity assumption on x_0 one has to solve the optimization problem :

The crude problem :

$$\arg\min_x \quad ||x||_0 \quad \text{subject to } y = \phi x$$

where $||x||_0 = |\{i : x_i \neq 0\}|$

Sparse recovery : Basics II

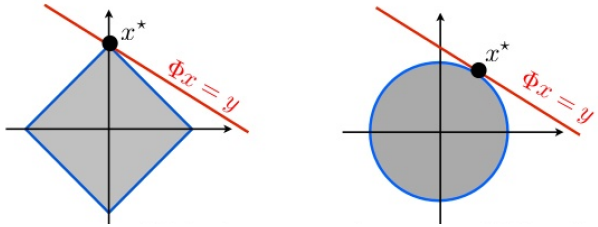
This problem is known to be NP-hard. Hence the relaxed convex optimization problem :

The reasonnable problem :

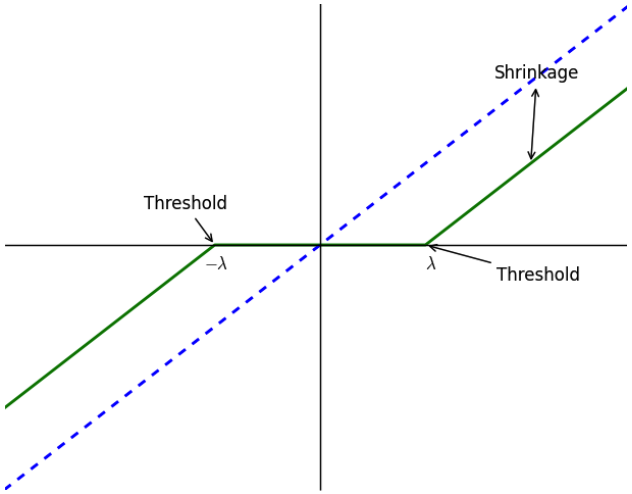
$$\arg\min_x \quad ||x||_1 \quad \text{subject to } y = \phi x \qquad (P)$$

This problem is efficiently solved using Basis Pursuit ([Chen et al., 1998]) or LASSO ([Tibshirani, 1994]).

Two intuitive reasons why the ℓ_1-norm is sparsity-inducing



Clear ! !



Consider :

$$x^\star(y) = \arg\min_x \quad \frac{1}{2}\|y - x\|^2 + \lambda\|x\|_1$$

*(Map $y \mapsto x^\star(y)$ is the **proximal operator**).*
Solution is as indicated.

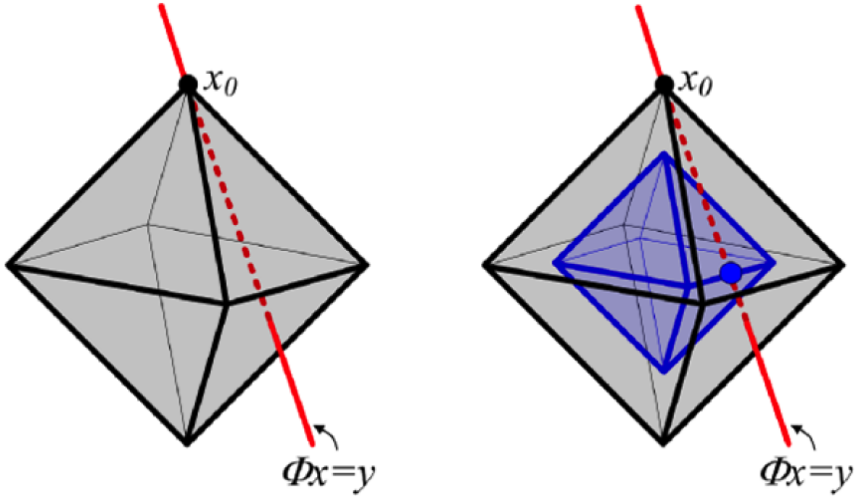
Not 100% reliable !

Consider the simple problem (P) with :

$$x_0 = (0, 1, 0)$$

$$\phi = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

Easy to see that the solution is $(\frac{1}{3}, 0, \frac{1}{3})$ which is not the sparsest one.



Make ℓ_1-norm more democratic... I

- Non-zero entries are considered equivalently in ℓ_0-norm.
- With ℓ_1-norm, large coefficients penalize more the objective than smaller ones.
- Prevents null entries from arising in the solution
- ... and small non-zero entries that are not in the solution from vanishing.

Hence the idea to push forward the optimization program by solving a reweighted problem as follows :

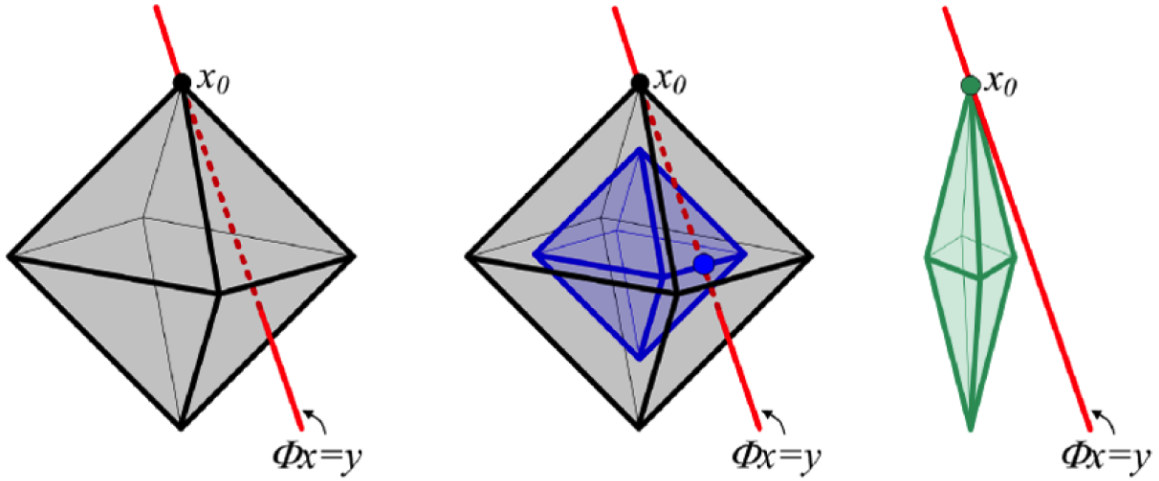
Reweighted ℓ_1 minimization problem

$$\arg\min_x \sum \omega_i |x_i| \quad \text{subject to } y = \phi x \qquad (WP_1)$$

where $\omega_i = \begin{cases} \frac{1}{|x_i|} & x_i \neq 0 \\ \infty & x_i = 0 \end{cases}$

Make ℓ_1-norm more democratic... II

Geometrically, this is equivalent to rescaling the vector space and the ℓ_1-balls :



The iterative reweighted algorithm

1. Set $W = Id_{\mathbb{R}^n}$
2. Solve the weighted ℓ_1 minimization problem :

$$x^{(l)} = \arg\min_x \quad ||W^{(l)}x||_{\ell_1} \quad \text{subject to } y = \phi x \qquad (WP_1)$$

3. Update weights :

$$w_i^{(l+1)} = \frac{1}{|x_i^{(l)}| + \epsilon}$$

4. Terminate if $l_{\max}$ iterations or on convergence.
   Otherwise, go to step 2.

Expected benefits I

At first glance, we have multiplied computation load by a factor $l_{\max}$! ! !

Enhancing sparsity should be understood as :

the reduction of the oversampling ratio m/k that allows for exact recovery.

Demo

An MM algorithm I

Principle of Majorize-Minimize algorithm : iteratively minimize a function that majorizes the objective. Consider the log-sum penalty problem :

$$\arg\min_{x} \quad \sum_{i=0}^{n} \log(|x_i| + \epsilon) \qquad \text{subject to} \quad y = \phi x$$

It is equivalent to :

$$\arg\min_{x,u} \quad \sum_{i=0}^{n} \log(u_i + \epsilon) \qquad \text{subject to} \quad \left\{ \begin{array}{l} y = \phi x \\ |x_i| \leq u_i, \forall i = 1, ..., n \end{array} \right.$$

which in turn is of the general form :

$$\arg\min_{v} \quad g(v) \quad \text{subject to} \quad v \in \mathcal{C}$$

where g is concave and differentiable.

An MM algorithm II

g being differentiable, it can be locally approximated by its tangent. And since it is concave, the tangent lies above the graph of g. We have the majorizing function. Hence the iterative algorithm :

$$v^{(l+1)} = \arg\min_{v} \quad \sum_{i=1}^{n} g(v^{(l)}) + \nabla g(v^{(l)}) \cdot (v - v^{(l)}) \quad \text{subject to} \quad v \in \mathcal{C}$$

Omitting the constant term in this expression, one now has to solve :

$$(x^{(l+1)}, u^{(l+1)}) = \arg\min_{v} \quad \sum_{i=1}^{n} \frac{u_i}{u_i^{(l)} + \epsilon} \quad \text{subject to} \quad \begin{cases} y = \phi x \\ |x_i| \leq u_i \end{cases}$$

This is again equivalent to :

$$x^{(l+1)} = \arg\min_{x} \quad \sum_{i=1}^{n} \frac{|x_i|}{|x_i^{(l)}| + \epsilon} \quad \text{subject to} \quad y = \phi x$$

which is the reweighted ℓ_1 minimization algo ! ! !

Summary

$$\arg\min_{x} \quad \sum_{i=0}^{n} \log(|x_i| + \epsilon) \quad \text{s. t.} \quad y = \phi x$$

$$x^{(l+1)} = \arg\min_{x} \quad \sum_{i=0}^{n} \frac{|x_i|}{|x_i^{(l)}| + \epsilon} \quad \text{s. t.} \quad y = \phi x$$

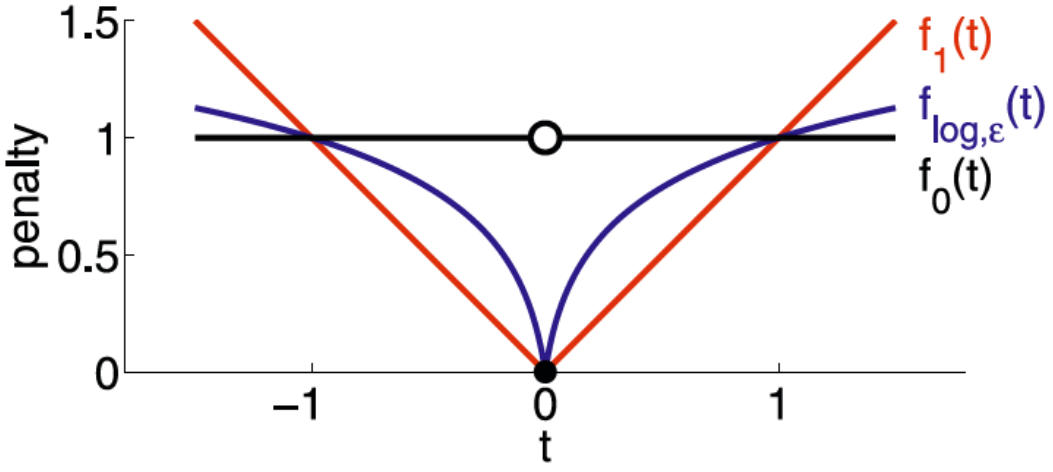
$$\arg\min_{u,x} \quad \sum_{i=0}^{n} \log(u_i + \epsilon) \quad \text{s. t.} \quad \begin{array}{l} y = \phi x \\ |x_i| \leq u_i \end{array}$$

$$\xrightarrow{\text{MM}}$$

$$(u^{(l+1)}, x^{(l+1)}) = \arg\min_{u,x} \quad \sum_{i=0}^{n} \frac{u_i}{u_i^{(l)} + \epsilon} \quad \text{s. t.} \quad \begin{array}{l} y = \phi x \\ |x_i| \leq u_i \end{array}$$

Why log-sum penalty ?

Just because it approximates much better the ℓ_0-norm and thus is much more sparsity-inducing than the ℓ_1-norm.

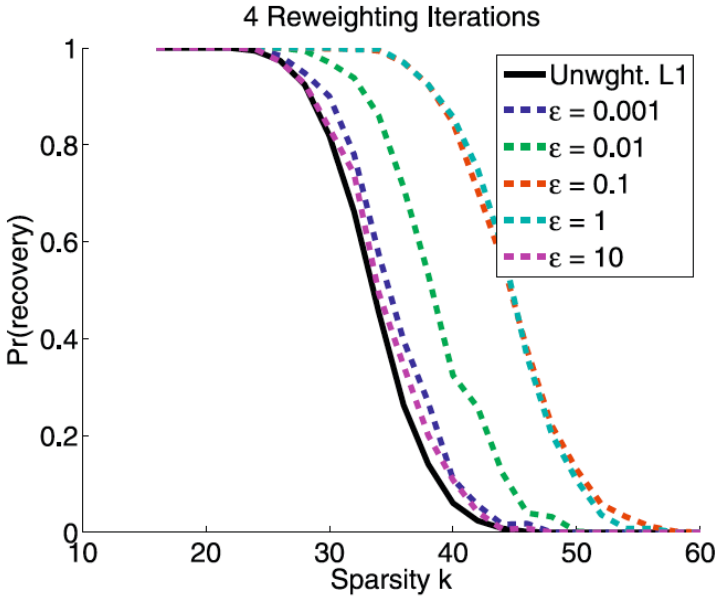


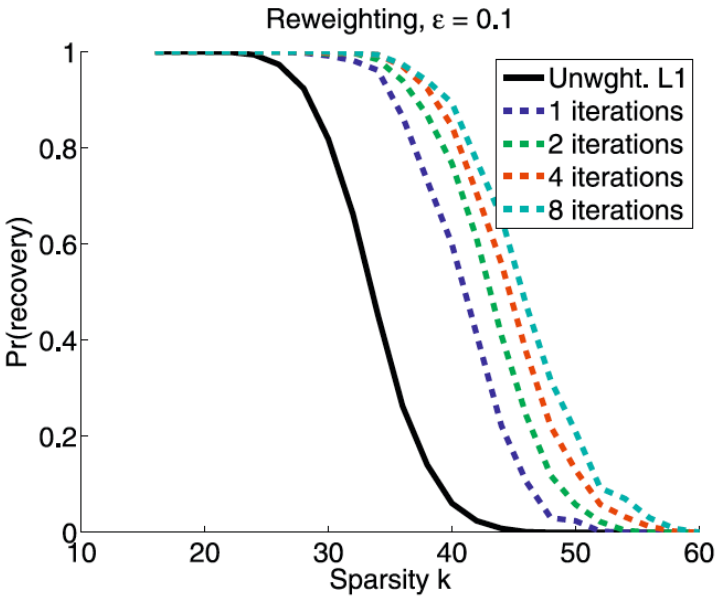
Parallel with Least Squares

- Least Square minimizes the ℓ_2-norm of the residual $Ax - b$.
- Outliers sensitive.
- Solve reweighted to better approximate an ℓ_1 criterion.

Num Exp. 1 : sparse recovery

A : iid gaussian entries.
$n = 256$; $m = 100$
Probability of perfect recovery, over 500 trials !





The choice of ϵ

- A rough approximation for choosing ϵ : 10% of std deviation of non-zero coefficients.
- A modified algorithm with adaptative choice of ϵ.

Add to step 3 of the algo :

Reorder in decreasing order of magnitude coefficients of $x^{(l)}$. Set :

$$\epsilon = \max(|x^{(l)}|_{i_0}, 10^{-3})$$

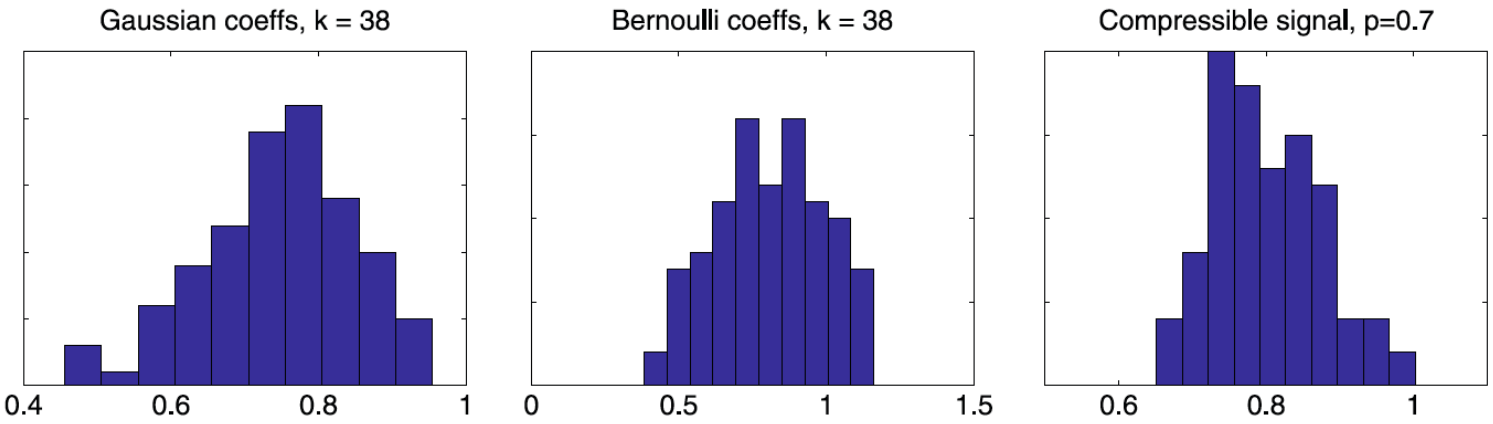
where $i_0 = \dfrac{m}{4\log(\frac{n}{m})}$.

Num Exp. 3 : Denoising

- Observed data : $y = \phi x_0 + z$.
- Solve :

$$x^{(l)} = \arg\min_x \quad \|W^{(l)}x\|_{\ell_1} \qquad \text{subject to} \qquad \|y - \phi x\|_{\ell_2} \leq \delta$$

- Results :



Num Exp. 4 : Statistical estimation

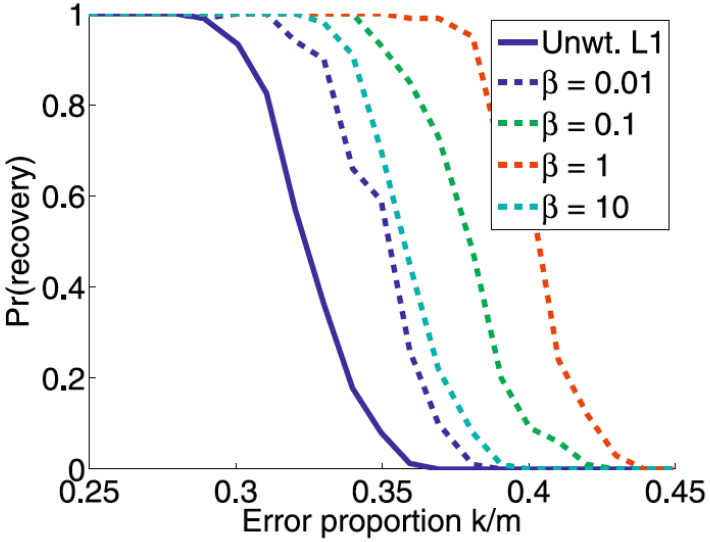
qsdfsd

Num Exp. 5 : Error correction [Candes and Tao, 2005]

- Let $x_0 \in \mathbb{R}^n$ be a signal to be transmitted. It is not sparse.
- Encode the message with matrix $\phi \in \mathbb{R}^{m \times n}$, $m \geq n$.
- Due to tranmission errors, receive $y = \phi x_0 + e$ where e is the corruption vector, which is sparse.
- Apply reweighted ℓ_1 minimization to :

$$\arg\min_x \quad \|y - \phi x\|_{\ell_1}$$

- Set the ϵ parameter to $\beta \times$ sd. of corrupted$y$



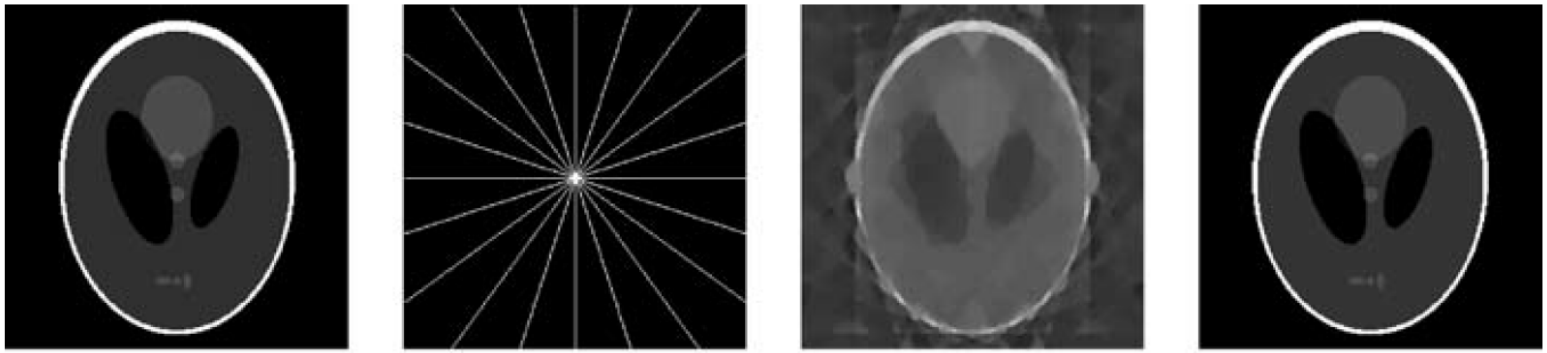
Conclusion :

Reweighted ℓ_1 allows a larger corrupted entries proportion to be overcome (from 28% to 35% approx.)

Num Exp. 6 : Sparse image gradient reconstruction via TV minimization

- Let $x_0 \in \mathbb{R}^n$ be an image whose gradient is sparse ($n = 256 \times 256$, and gradient has 2184 non-zero entries).
- Sample the Fourier transform of x_0 along 10 radial lines in the Fourier space ($m = 2521$ real-valued measurments) and observe $y = \phi x_0$ where ϕ is a subset of the Fourier coefficients.
- Set ϵ to 0.1 and apply reweighted ℓ_1 minimization to :

$$\arg\min_x \quad \|x\|_{TV} \quad \text{subject to} \quad y = \phi x$$



It would require 4257 measurements to achieve a perfect recovery with unweighted minimization. Oversampling ratio reduction from 1.05 (4257)

Num Exp. 7 : Where compressive sensing comes up ! I

The signal x_0 may not be sparse but in a overcomplete dictionnary ψ ie : $x_0 = \psi\alpha$ where α is sparse.
Two ways of adressing the reconstruction problem :

1. The **synthesis-based recovery**, which solves :

$$\arg\min_{\alpha} \quad \|\alpha\|_{\ell_1} \quad \text{subject to} \quad y = \phi\psi\alpha$$

2. The **analysis-based recovery**, which solves :

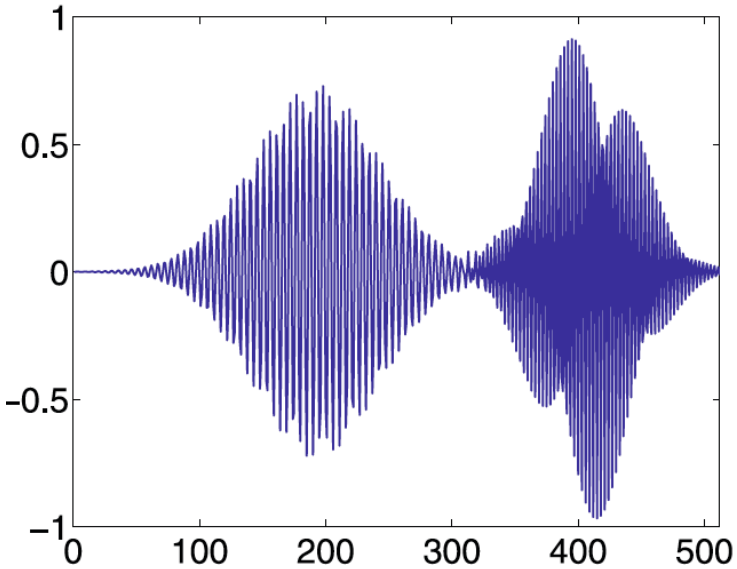
$$\arg\min_{x} \quad \|\psi^* x\|_{\ell_1} \quad \text{subject to} \quad y = \phi x$$

Both can be applied reweighting.

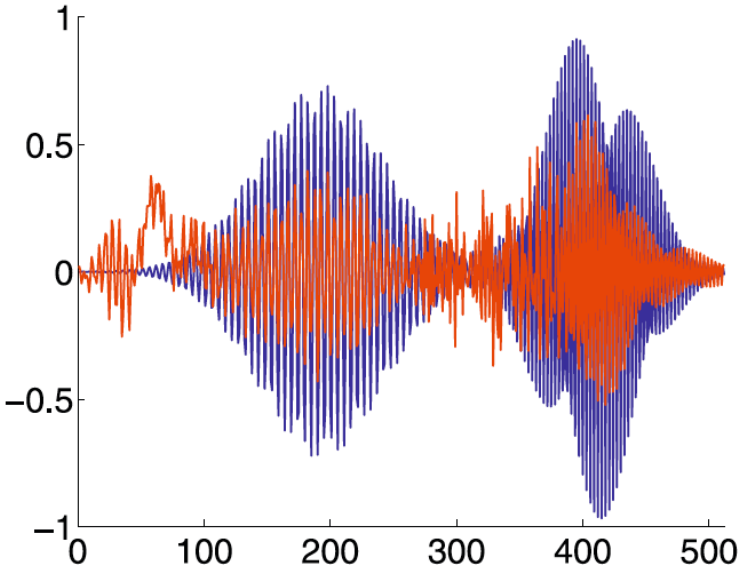
Num Exp. 7 : Where compressive sensing comes up ! II

- Let $x_0 \in \mathbb{R}^n$ with $n = 512$ be the superposition of 2 radar modulated pulses.
- Collect 30 measurements form an iid ± 1 random matrix (undersampling factor ≥ 17 ! !).
- Reconstruct signal with a time-freq Gabor dictionnary ($43\times$ overcomplete) not containing the 2 pulses.
- Results :

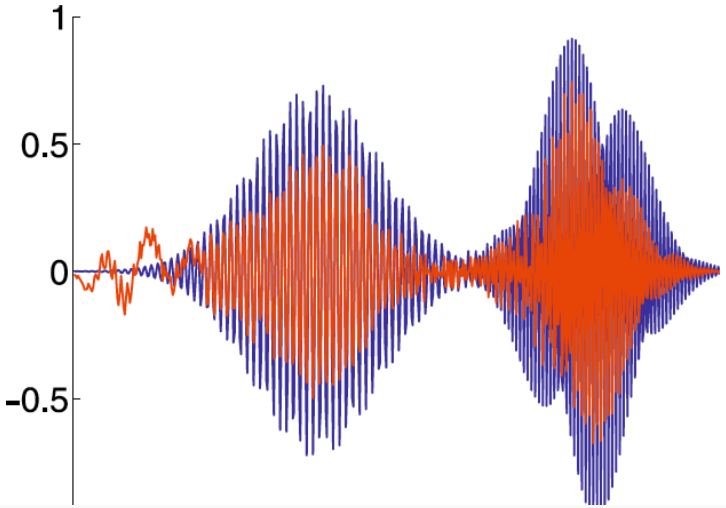
Num Exp. 7 : Where compressive sensing comes up ! III

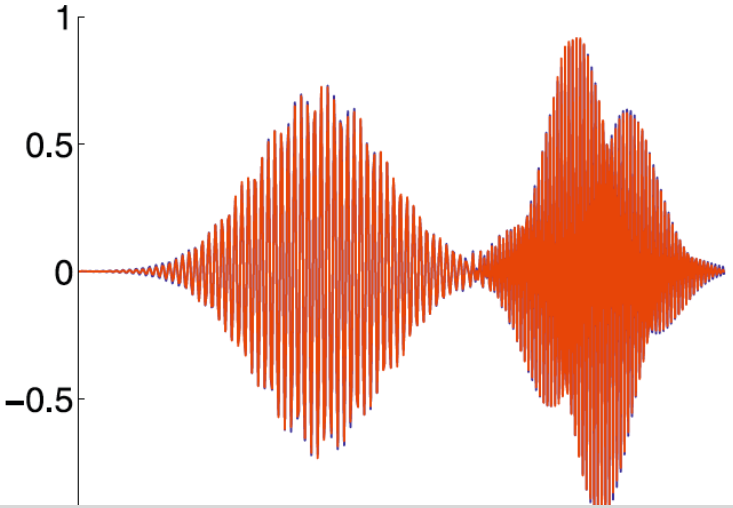


(a)



(b)





Bibliography I

Boyd, S. and Vandenberghe, L. (2004).
Convex Optimization.
Cambridge University Press, NY, USA.

Candes, E. J. and Tao, T. (2005).
Decoding by linear programming.
IEEE Trans. Inf. Theor., 51(12) :4203–4215.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998).
Atomic decomposition by basis pursuit.
SIAM Journal on Scientific Computing, 20 :33–61.

Tibshirani, R. (1994).
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society, Series B, 58 :267–288.