On sparse-ℓ_0 solutions of least-square fitting: on-grid methods, algorithms, and some results on image processing

Laure Blanc-Féraud

Projet MORPHEME - UCA, CNRS, INRIA -

Sparsity4PSL - June 24th-27th 2019

Outline of the talk

Many signal processing areas are concerned with **sparse solution recovery**: compressed sensing, variable selection, source separation, learning...

- Linear observation : $A\mathrm{x} = d$
  - d : observed data, vector in $\mathbb{R}^M$
  - x unknown data to be estimated in $\mathbb{R}^N$
  - A observation matrix, $M \times N$ matrix.

  usually $M < N$, the system is undertermined, A is ill-conditioned, observations are noisy
- Least square solution $\hat{\mathrm{x}} = \arg \min_{\mathrm{x} \in \mathbb{R}^N} \|A\mathrm{x} - d\|_2^2$
- Regularization: sparse signal hypothesis modeled by considering "ℓ_0-norm" constraints:

$$\|\mathrm{x}\|_0 \leq K \quad \text{where} \quad \|\mathrm{x}\|_0 = \#\{\mathrm{x}_i,\ i = 1, \ldots, N : \mathrm{x}_i \neq 0\}$$

NB: ℓ_0-norm **is NOT** a norm as $\|\lambda \mathrm{x}\|_0 = \|\mathrm{x}\|_0 \neq \lambda \|\mathrm{x}\|_0$.

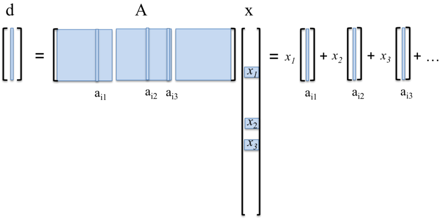
1.0 Dictionary representation in image processing

- Image are non-stationary, they exhibit smooth areas, oscillations, edges, textures,...
- Each part is represented by given waveforms which best match the image structure, for example Basis B_i as Haar, smooth wavelets, sine/cosine,...
- Construct a redundant dictionary with all these representative waveforms, possibly by a succession of bases
- An image d will be represented in this over-complete dictionary, if we find

$$\arg\min_{\mathrm{x}\in\mathbb{R}^N} \|A\mathrm{x} - d\|_2^2 + \lambda\|x\|_0$$

or

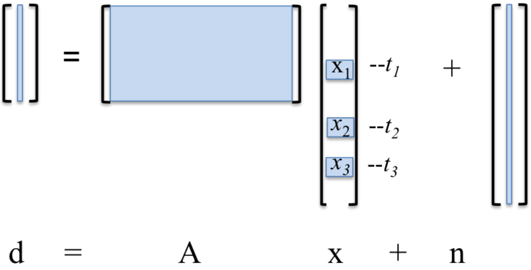
$$\arg\min_{\mathrm{x}\in\mathbb{R}^N} \|A\mathrm{x} - d\|_2^2 \text{ subject to } \|\mathrm{x}\|_0 \leq K$$



1.1 Examples in Signal/image Processing

- signal is a sum of pulses, spikes, modeled by a sum of Dirac $\sum_{r=1}^{K} x_r \delta_{t_r}$.
- acquisition system, channel, is modeled as a linear system, e.g. convolution by a Gaussian function: $d(.) = h * \sum_{r=1}^{K} x_r \delta_{t_r} = \sum_{r=1}^{K} x_r h(. - t_r)$.

By assuming the Dirac locations t_r are on a regular grid indexed by $i = 1, \ldots N$

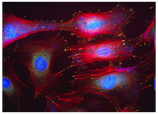


- 1D example: Channel estimation in communications -
- 2D example: Single Molecule Localization in super-resolution microscopy -

2D example in Super-resolution microscopy: SMLM (Single Molecule Localization Microscopy)

Fluorescence microscopy

- Genes of fluorescent molecules are combined with genes of proteins of structure we want to study
  Nobel Prize of chemistry 2008
- Illumination by a laser causes the fluorophores to emit photons
- structure of interest can be imaged through the microscope



It allows

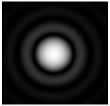
- **living** cell imaging
- 3D imagery
- Resolution 200 nm in lateral direction, around 400 axial direction (depth)

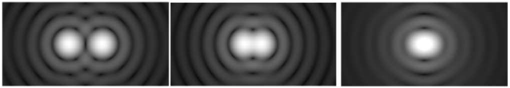
Approximate sizes : cell 10 -100 μm, nucleus 4 -7 μm, proteins 10 -100 μm, molecules few nm.

2D example in Super-resolution microscopy: SMLM (continued)

Conventional fluorescence microscopy limits

- physical diffraction limit of optical systems
- Airy patch = impulse response of the microscope (PSF: *Point Spread Function*)
- overlapping patches limit at ≈200nm the distance between two molecules to be resolved (Rayleigh limit)





Super-resolution by single molecule localization

- **Photo-activable molecules**: PALM *Photo Activated Localisation Microscopy* ([Betzig & al 06, Hess & al, 2006]) et STORM *STochastic Optical Reconstruction Microscopy* ([Rust & al, 2006])
- Sequentially activate and image a small random set of fluorescent molecules.

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

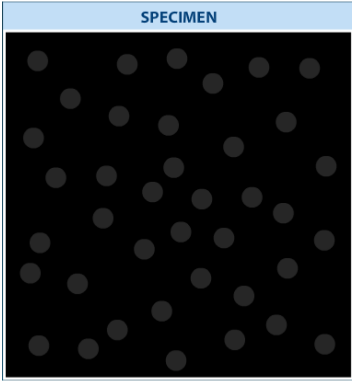


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

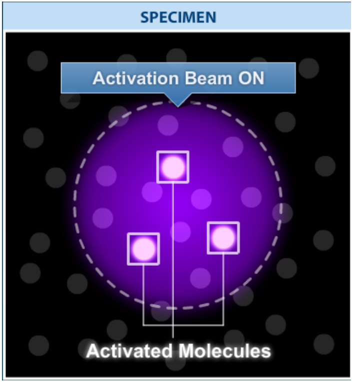


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

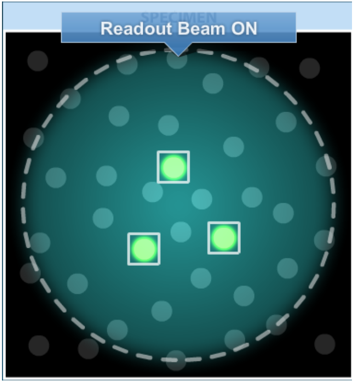


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

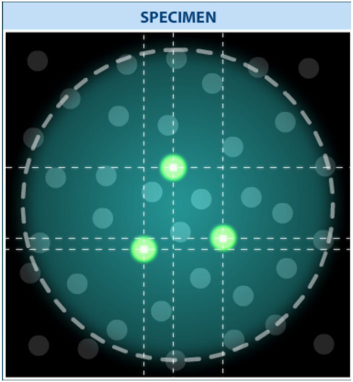


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

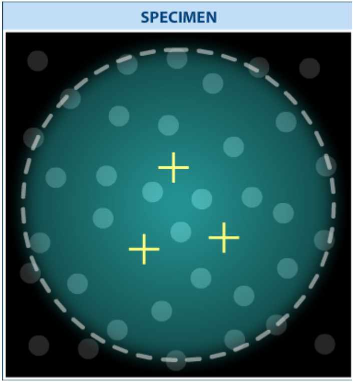


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

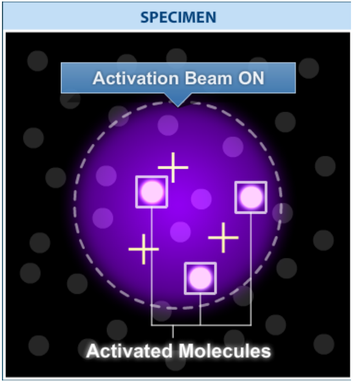


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

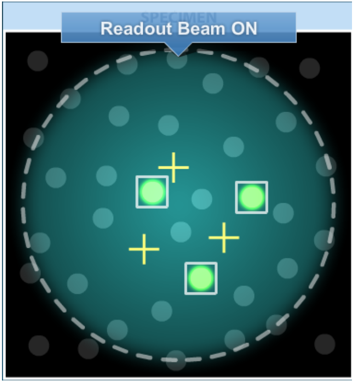


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

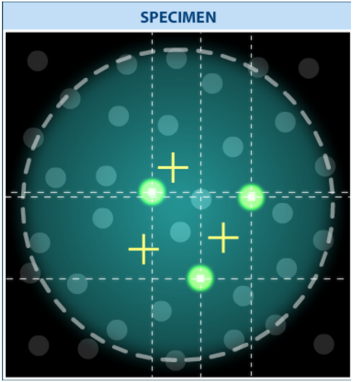


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

- activation
- imaging
- localization
- assembling

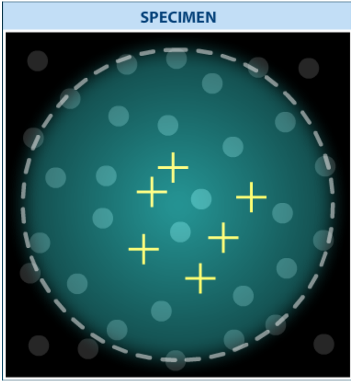
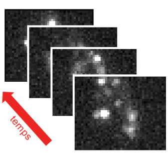


Figure: PALM microscopy principle. From Zeiss tutorials [http://zeiss-campus.magnet.fsu.edu/tutorials/index.html]

2D example in Super-resolution microscopy: SMLM (continued)

Limitations: number of acquisition needed to obtain the super-resolved image

- cost time and memory
- temporal resolution restricted (motion)

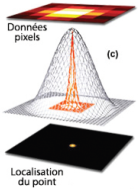


→ **Increase molecule density**

- Localization more difficult due to **more overlapping**

Localization algorithms

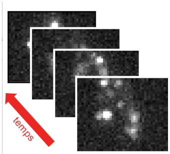
- Challenge ISBI 2013 [Sage et al 15]
- PSF fitting, and derived methods for high density molecule localization (e.g. DAOSTORM, [Holden & al 11]).
- **Deconvolution and reconstruction on a finer grid** (e.g. FALCON, [Min & al, 2014])



2D example in Super-resolution microscopy: SMLM (continued)

Limitations: number of acquisition needed to obtain the super-resolved image

- cost time and memory
- temporal resolution restricted (motion)

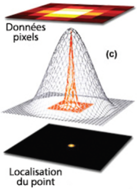


→ **Increase molecule density**

- Localization more difficult due to **more overlapping**

Localization algorithms

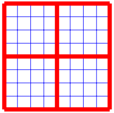
- Challenge ISBI 2013 [Sage et al 15]
- PSF fitting, and derived methods for high density molecule localization (e.g. DAOSTORM, [Holden & al 11]).



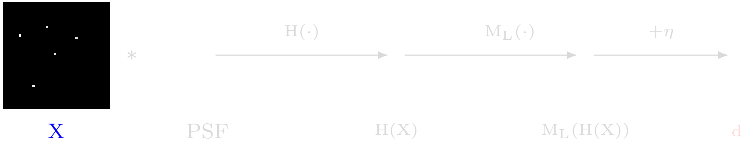
- **Deconvolution and reconstruction on a finer grid** (e.g. FALCON, [Min & al, 2014])

Image formation model PALM / STORM

d $\in \mathbb{R}^{M\times M}$ one acquisition.
X $\in \mathbb{R}^{ML\times ML}$ an image where each pixel of **d** is divided in **L**$\times$**L** pixels.



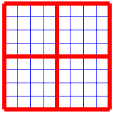
L=4



2D example in Super-resolution microscopy: SMLM (continued)

Image formation model PALM / STORM

$\mathbf{d} \in \mathbb{R}^{M \times M}$ one acquisition.
$\mathbf{X} \in \mathbb{R}^{ML \times ML}$ an image where each pixel of $\mathbf{d}$ is divided in $\mathbf{L}\times\mathbf{L}$ pixels.



L=4

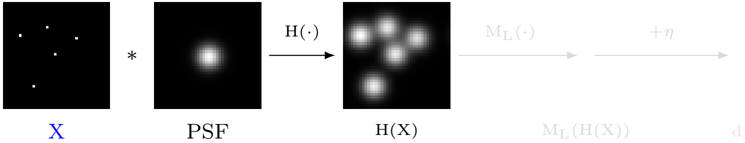
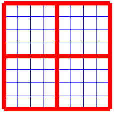


Image formation model PALM / STORM

$\mathbf{d} \in \mathbb{R}^{M \times M}$ one acquisition.
$\mathbf{X} \in \mathbb{R}^{ML \times ML}$ an image where each pixel of $\mathbf{d}$ is divided in $\mathbf{L} \times \mathbf{L}$ pixels.



L=4

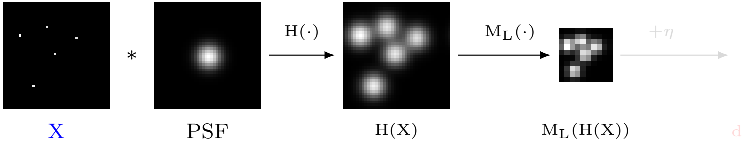
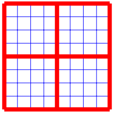


Image formation model PALM / STORM

$\mathbf{d} \in \mathbb{R}^{M\times M}$ one acquisition.
$\mathbf{X} \in \mathbb{R}^{ML\times ML}$ an image where each pixel of $\mathbf{d}$ is divided in $\mathbf{L}\times\mathbf{L}$ pixels.



L=4

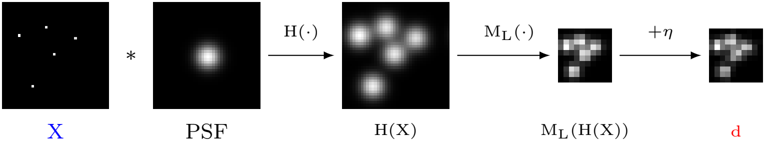
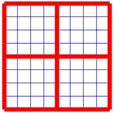
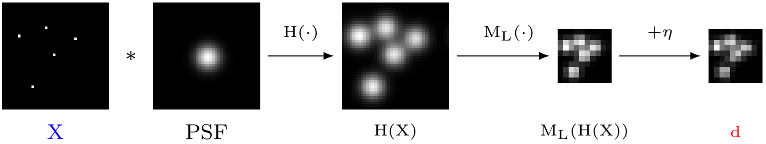


Image formation model PALM / STORM

$\mathbf{d} \in \mathbb{R}^{M \times M}$ one acquisition.
$\mathbf{X} \in \mathbb{R}^{ML \times ML}$ an image where each pixel of $\mathbf{d}$ is divided in $\mathbf{L}\times\mathbf{L}$ pixels.



L=4



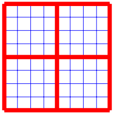
Model

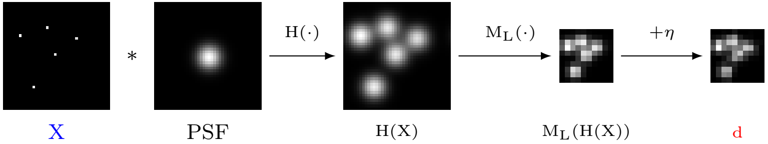
$$\mathrm{d} = \mathrm{M_L}(\mathrm{H}(\mathrm{X})) + \eta,$$

2D example in Super-resolution microscopy: SMLM (continued)

Image formation model PALM / STORM

$\mathbf{d} \in \mathbb{R}^{M \times M}$ one acquisition.
$\mathbf{X} \in \mathbb{R}^{ML \times ML}$ an image where each pixel of $\mathbf{d}$ is divided in $\mathbf{L} \times \mathbf{L}$ pixels.





Problem $\ell_2 - \ell_0$

$$\hat{\mathrm{X}} \in \arg\min_{\mathrm{X}} \frac{1}{2}\|\mathrm{d} - \mathrm{M_L}(\mathrm{H}(\mathrm{X}))\|_2^2 + \lambda \|X\|_0$$

1.3 ℓ_2-ℓ_0 optimization problems

Exact Recovery problem

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_0 \;\text{ subject to }\; Ax = d$$

Approximation problem: two constrained forms

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|Ax - d\|_2^2 \;\text{ subject to }\; \|x\|_0 \leq K$$

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_0 \;\text{ subject to }\; \|Ax - d\|_2^2 \leq \epsilon$$

Approximation problem : penalized form

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_0$$

$$A \in \mathbb{R}^{M \times N} \text{ with } M \ll N$$

- Non equivalent formulations
- Existence of an optimal solution and relationships between optimal solutions in [Nikolova 16]
- Intensive work in signal and image processing, and in statistics.
- **non-continuous**, **non-convex** and **NP-hard** optimization problem. [Natarajan 95] [Davis & al 97]. Rouhgly speaking, *a solution cannot be verified in polynomial time w.r.t the dimension of the problem*

Outline of the talk

1. Introduction and examples
2. **Iterative Hard Thresholding** (IHT): Forward-Backward Splitting (FBS) algorithm [Blumensath and Davies 08]
3. Greedy algorithms
4. Continuous relaxation
5. Exact reformulation
6. Some results on super-resolution microscopy
7. Conclusion

Penalized form

$$\hat{\mathrm{x}} = \arg\min_{\mathrm{x}\in\mathbb{R}^N} \frac{1}{2}\|A\mathrm{x}-d\|_2^2 + \lambda\|\mathrm{x}\|_0$$

- $\frac{1}{2}\|A\mathrm{x}-d\|_2^2$ is L-gradient Lipschitz ($L = \|A\|^2$)
- Proximal of $\|.\|_0$ has explicit expression, this is the Hard Threshold

Iterative Hard Thresholding

(IHT): Forward-Backward Splitting (FBS) algorithm

$$\mathrm{x}^{k+1} = \mathrm{prox}_{\gamma\lambda\|.\|_0}\left(\mathrm{x}^k - \gamma A^t\left(A\mathrm{x}^k - d\right)\right)$$

$\gamma < \frac{1}{L}$ is the gradient step.

Computation of $\mathrm{prox}_{\gamma\lambda\|.\|_0}$:

$$\mathrm{prox}_{\gamma\lambda\|.\|_0}(\mathrm{y}) = \arg\min_{\mathrm{x}\in\mathbb{R}^N}\left\{\frac{1}{2}\|\mathrm{x}-\mathrm{y}\|^2 + \gamma\lambda\|\mathrm{x}\|\|_0\right\}$$

$$\frac{1}{2}(\mathrm{x}-\mathrm{y})^2 + \gamma\lambda\|\mathrm{x}\|\|_0 = \sum_{i=1}^N (x_i - y_i)^2 + \gamma\lambda|x_i|_0$$

where $|u|_0 = 1$ if $u \neq 0$, 0 elsewhere.

Then it is sufficient to compute in 1D $\arg\min_{u\in\mathbb{R}}\left\{g(u) := \frac{1}{2}(u-y)^2 + \gamma\lambda|u|_0\right\}$

Computation of $\arg\min_{u\in\mathbb{R}} \left\{ g(u) := \frac{1}{2}(u-y)^2 + \gamma\lambda|u|_0 \right\}$

- if $u = 0$ then $g(0) = \frac{1}{2}(y)^2$
- The minimum could be reached at $\hat{u} = 0$, the value is $g(\hat{u}) = \frac{1}{2}(y)^2$

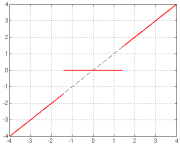
- if $u \neq 0$ then $g(u) = \frac{1}{2}(u-y)^2 + \lambda$
- The minimum is reached at $\hat{u} = y$ and the value is $g(\hat{u}) = \lambda$

if $|y| \leq \sqrt{2\lambda}$ then $\hat{u} = 0$

if $|y| \geq \sqrt{2\lambda}$ then $\hat{u} = y$

The solution is given by the Hard Threshold function

$$\hat{u} = \begin{cases} y & \text{if } |y| > \sqrt{2\lambda}, \\ 0 & \text{if } |y| \leq \sqrt{2\lambda}. \end{cases}$$



Find the solution of the optimal problem

$$\hat{x} = \arg \min_{x \in \mathbb{R}^N} \frac{1}{2} \|Ax - d\|_2^2 + \lambda \|x\|_0$$

by Forward Backward Splitting algorithm (Iterative Hard Thresholding)

$$x^{k+1} = \text{prox}_{\gamma\lambda\|.\|_0} \left(x^k - \gamma A^t \left(Ax^k - d \right) \right)$$

- IHT algorithm converges to a critical point [Blumensath and Davies 08, Attouch et al 13].
- **Initialization** point is important, for example initialize with the solution with the ℓ_1-norm problem: $\arg \min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|Ax - y\|^2 + \gamma\lambda \|x\|\|_1 \right\}$. It is not guaranty that this solution is sparse.

Outline of the talk

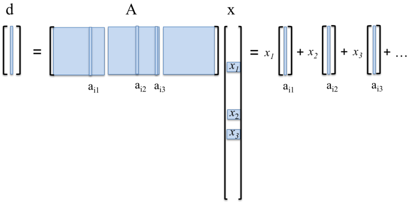
1. Introduction and examples
2. Iterative Hard Thresholding
3. **Greedy algorithms**, *Matching Pursuit (MP)* [Mallat et al 93], *Orthogonal MP* [Pati et al 93], *Orthogonal Least Squares (OLS)* [Chen et al 89], *Bayesian OMP* [Herzet et al 10], *Single Best Replacement* [Soussen et al 11] and further variants.
4. Continuous relaxation
5. Exact reformulation
6. Some results on super-resolution microscopy
7. Conclusion

3. Greedy algorithms

Greedy algorithms, *Matching Pursuit (MP)* [Mallat et al 93], *Orthogonal MP* [Pati et al 93], *Orthogonal Least Squares (OLS)* [Chen et al 89], *Bayesian OMP* [Herzet et al 10], *Single Best Replacement* [Soussen et al 11] and further variants.

Matching Pursuit:

d is the signal we want to represent with the a limited number $K << N$ of waveforms or atoms of dictionary A, one atom is one column of A, *i.e.* $A_{.,i} = \mathrm{a}_i$, $i = 1, ..N$.



For that we have to solve

$$\hat{\mathrm{x}} = \arg\min_{\mathrm{x}\in\mathbb{R}^N} \|A\mathrm{x} - d\|_2^2 \text{ subject to } \|\mathrm{x}\|_0 \leq K.$$

$$\text{(or } \hat{\mathrm{x}} = \arg\min_{\mathrm{x}\in\mathbb{R}^N} \|\mathrm{x}\|_0 \text{ subject to } \|A\mathrm{x} - d\|_2^2 \leq \epsilon\text{)}$$

Matching Pursuit algorithm add one component at a time.

Matching Pursuit principle

It is assumed without loss of generality that A has unit norm columns, $\|A_{.,i}\| = \|a_i\| = 1$.

The **first component** $i^1 \in \{1, ..., N\}$ will be such that the **correlation** between d and atom i is maximum: $i^1 = \arg \max_{j \in \{1,..,N\}} |\langle a_j, d\rangle|$.

Then the **optimal solution** is $\mathrm{x}^1 = (0, 0, .., \langle a_{i^1}, d\rangle, 0, .., 0)$, where the non null component is at index i^1, which is written as $\mathrm{x}^1 = \langle a_{i^1}, d\rangle.e_{i^1}$, $e_i \in \mathbb{R}^N$, $i \in \{1, .., N\}$ is the canonical basis in $\mathbb{R}^N$.

The criterion is $\|A.\mathrm{x}^1 - d\|^2 = \|d\|^2 - (\langle a_{i^1}, d\rangle)^2$.

The **residual** is $r = d - A.\mathrm{x}^1 = d - \langle a_{i^1}, d\rangle a_{i^1}$, and the process is repeated.

Matching Pursuit Algorithm
Input: A (with unit norm column), d, K.

Initialize: $r^0 = d, \sigma^0 = \varnothing, (x^0 = 0)$.

Repeat, while $\#\sigma^k \leq K$: (or while $\|r^k\| > \epsilon$)

$$\begin{aligned} i^k &= \arg \max_{j \in \{1,..,N\}} |\langle r^k, a_j \rangle| \\ \sigma^{k+1} &= \sigma^k \cup \{i^k\} \\ r^{k+1} &= r^k - \langle r^k, a_{i^k} \rangle . a_{i^k} \end{aligned} \tag{1}$$

σ^k is the support of the current solution x^k, that is the indexes of the non-zero components. $\#\sigma^k$ is the cardinal of σ^k. The initial value of $\#\sigma^0$ is 0 and it increases by 1 at each iteration.
The optimal solution at current iteration is $x^{k+1} = x^k + \langle r^k, a_{i^k} \rangle . e_{i^k}$.

- The residual $\|r^k\|$ converges exponentially to 0 [Mallat et al 93].
- Sub-optimal solution: retro-project the residual onto $Span\{(a_i)_{i \in \sigma^K}\}$ reduce the approximation error ($\|A.\mathrm{x}^K - d\|^2$).

Orthogonal Matching Pursuit [Pati et al 93, Tropp 04]: at each iteration, optimally estimate the intensities with the current support of the solution fixed, by $\mathrm{x}^{k+1} = \arg \min_{\{\mathrm{x}/\sigma_\mathrm{x} \subset \sigma^{k+1}\}} \|A\mathrm{x} - d\|^2$.

Orthogonal Matching Pursuit (OMP) Algorithm Input: A (with unit

norm column), d, K.

Initialize: $r^0 = d, \sigma^0 = \varnothing$

Repeat, while $\#\sigma^k \leq K$:

$$\begin{aligned}
i^k &= \arg\max_{j \notin \sigma^k} |\langle r^k, a_j \rangle| \\
\sigma^{k+1} &= \sigma^k \cup \{i^k\} \\
\mathrm{x}^{k+1} &= \arg \min_{\{\mathrm{x}/\sigma_\mathrm{x} \subset \sigma^{k+1}\}} \|A\mathrm{x} - d\|^2 \\
r^{k+1} &= d - A\mathrm{x}^{k+1}
\end{aligned}$$

- Convergence in N iterations at most (at each iteration a **new** component is selected),
- Exact sparse recovery results (under conditions on A) [Tropp 04].

Further algorithms:

At each iteration, several strategies for one component to be

- added,
- removed,
- replaced.

Orthogonal Least Squares (OLS) [Chen et al 89], *Bayesian OMP* [Herzet et al 10], *Single Best Replacement* [Soussen et al 11] and further variants [Jain & al 11, Soussen et al 15]...

The more complex is the strategy, the best is the solution and the longest is the computing time.

Outline of the talk

1. Introduction and examples
2. Iterative Hard Thresholding
3. Greedy algorithms
4. **Continuous relaxation**,
   - **convex** ℓ_1 relaxation (LASSO [Tibshirani 96], Basic Pursuit [Chen et al 98], Compressed Sensing [Donoho et al 06, Candès et al 06]), reweighted ℓ_1 [Candès et al 08].
   - **Non-convex** Adaptive Lasso [Zou 06], Nonnegative Garrote [Breiman 95], Exponential approximation [Mangasarian 96], Log-Sum Penalty [Candès et al 08], Smoothly Clipped Absolute Deviation (SCAD) [Fan and Li 01], Minimax Concave Penalty (MCP) [Zhang 10], ℓ_p-norms $0 < p < 1$ [Chartrand 07, Foucart and Lai 09], Smoothed ℓ_0-norm Penalty (SL0) [Mohimani et al 09], Continuous Exact ℓ_0 relaxation (CEL0) [Soubies et al 17],...
5. Exact reformulation
6. Some results on super-resolution microscopy
7. Conclusion

Continuous separable relaxation (convex and non-convex)

$$\frac{1}{2}\|Ax-d\|_2^2+\lambda\|x\|_0 \quad \rightarrow \quad \frac{1}{2}\|Ax-d\|_2^2+\lambda\sum_{i\in\mathbb{I}_N}\phi(x_i)$$

Continuous approximation of the ℓ_0-norm function:

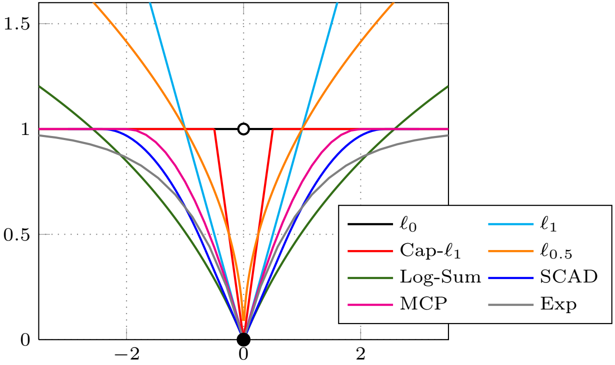
- ℓ_1-norm: Lasso [Tibshirani 96] ; Basic Pursuit [Chen et al 98] ; Compressed Sensing [Donoho et al 06, Candès et al 06])
- Adaptive Lasso [Zou 06] ;
- Nonnegative Garrote [Breiman 95] ;
- Exponential approximation [Mangasarian 96] ;
- Log-Sum Penalty [Candès et al 08] ;
- Smoothly Clipped Absolute Deviation (SCAD) [Fan and Li 01] ;
- Minimax Concave Penalty (MCP) [Zhang 10] ;
- ℓ_p-norms $0<p<1$ [Chartrand 07, Foucart and Lai 09] ;
- Smoothed ℓ_0-norm Penalty (SL0) [Mohimani et al 09] ;

Are they *good* approximations?
Which one to use?

Continuous separable relaxation (convex and non-convex)

$$\frac{1}{2}\|A\mathrm{x}-d\|_2^2+\lambda\|\mathrm{x}\|_0 \quad \rightarrow \quad \frac{1}{2}\|A\mathrm{x}-d\|_2^2+\lambda\sum_{i\in\mathbb{I}_N}\phi(\mathrm{x}_i)$$

Continuous approximation of the ℓ_0-norm function:



Are they *good* approximations?
Which one to use?

4.0 ℓ_1 convex relaxation: a specific case

Replacing ℓ_0-norm with ℓ_1-norm gives **convex** problems. Non differentiability in 0 of the ℓ_1 norm enforces sparsity.

Basis Pursuit (BP) [Chen et al 98]

$$\arg\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } Ax = d$$

- Compresssed Sensing reconstruction problems [Donoho et al 06, Candès et al 06]
- Results of exact recovery of a sparse solution using ℓ_1 minimization rather than ℓ_0 minimization have been shown, under quite restrictive conditions on matrix A (Restrictive Isometry Property RIP, incoherence...) [Donoho Elad 03, Gribonval Nielsen 03, Candès Wakin 08]

Basis Pursuit De-Noising (BPDN) [Chen et al 98], LASSO [Tibshirani 96]

Noisy version

$$\arg\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } \|Ax - d\|_2^2 \leq \epsilon$$

or

$$\arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_1$$

- Sparse signal recovery under conditions on A [Candès et al 06, Candès Wakin 08].

LASSO = Least Absolute Shrinkage and Selection Operator

$$G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_0 \quad \rightarrow \quad \tilde{G}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \sum_{i=1}^{N} \phi(x_i)$$

Definition of a *good* continuous approximation

- $G_{\ell_0}(x)$ and $\tilde{G}(x)$ have **same global** minimizers

$$\arg\min_{x \in \mathbb{R}^N} \tilde{G}(x) = \arg\min_{x \in \mathbb{R}^N} G_{\ell_0}(x) \qquad \text{(P1)}$$

- $\tilde{G}(x)$ has **less local** minimizers than $G_{\ell_0}(x)$

$$\hat{x} \text{ minimiseur de } \tilde{G} \implies \hat{x} \text{ minimiseur de } G_{\ell_0} \qquad \text{(P2)}$$

Question:

Can we derive necessary and suffisant conditions on $\phi(.)$ such that $\tilde{G}(x)$ is a good approximation of G_{ℓ_0}, with **no conditions on** A and $\forall d \in \mathbb{R}^M$?

$$G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_0 \quad \to \quad \tilde{G}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \sum_{i=1}^N \phi(x_i)$$

Definition of a *good* continuous approximation

- $G_{\ell_0}(x)$ and $\tilde{G}(x)$ have **same global** minimizers

$$\arg\min_{x \in \mathbb{R}^N} \tilde{G}(x) = \arg\min_{x \in \mathbb{R}^N} G_{\ell_0}(x) \quad \text{(P1)}$$

- $\tilde{G}(x)$ has **less local** minimizers than $G_{\ell_0}(x)$

$$\hat{x} \text{ minimiseur de } \tilde{G} \implies \hat{x} \text{ minimiseur de } G_{\ell_0} \quad \text{(P2)}$$

Question:

Can we derive necessary and suffisant conditions on $\phi(.)$ such that $\tilde{G}(x)$ is a good approximation of G_{ℓ_0}, with **no conditions on A** and $\forall d \in \mathbb{R}^M$?

4. ℓ_2-ℓ_0 optimization by continuous relaxation

Notations

- $G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_0$
- $\tilde{G}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \sum_{i=1}^N \phi(x_i)$

- (P1) $\arg\min_{x\in\mathbb{R}^N} \tilde{G}(x) = \arg\min_{x\in\mathbb{R}^N} G_{\ell_0}(x)$
- (P2) $\hat{x}$ minimizer of $\tilde{G} \implies \hat{x}$ minimizer of G_{ℓ_0}

- B : a finite subset of points of $\mathbb{R}$ on which ϕ is not differentiable.
- $\|a_i\|$ column i of matrix A ($\|a_i\| \neq 0$).

Additional assumptions

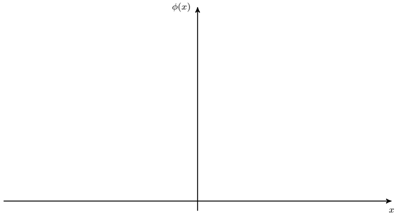
- $\min_{x\in\mathbb{R}} G_{\ell_0}(x) = \min_{x\in\mathbb{R}} \tilde{G}(x)$,
- ϕ is locally Lipschitz on $\mathbb{R}$,
- ϕ is twice differentiable on $\mathbb{R} \setminus B$,
- ϕ is not differentiable on B.

Theorem (NS conditions for (P1))

$\tilde{G}$ *has property (P1)* $\forall d \in \mathbb{R}$ *iff* ϕ *verifies:*

- $\phi(0) = 0$,
- $\forall x \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right)$,
  $\phi(x) = \lambda |x|_0 = \lambda$,
- $\forall x \in \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right) \setminus \{0\}$,
  $\phi(x) > \phi_{\text{CEL0}}(\|a_i\|, \lambda; x)$

$\mathbb{1}_{\{x \in D\}} = 1$ if $x \in D$; 0 otherwise.

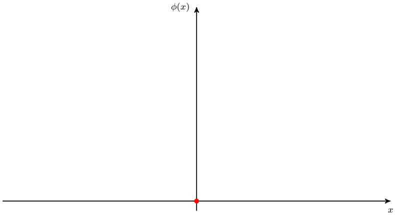


Theorem (NS conditions for (P1))

$\tilde{G}$ *has property (P1)* $\forall d \in \mathbb{R}$ *iff* ϕ *verifies:*

- $\phi(0) = 0$,
- $\forall x \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right)$, $\phi(x) = \lambda|x|_0 = \lambda$,
- $\forall x \in \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right) \setminus \{0\}$, $\phi(x) > \phi_{\text{CEL0}}(\|a_i\|, \lambda; x)$

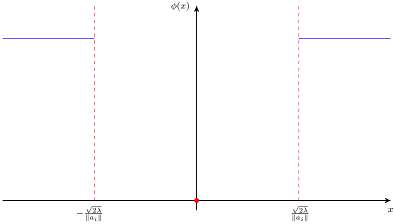
$\mathbb{1}_{\{x \in D\}} = 1$ if x $\in D$; 0 otherwise.



Theorem (NS conditions for (P1))

$\tilde{G}$ *has property (P1)* $\forall d \in \mathbb{R}$ *iff* ϕ *verifies:*

- $\phi(0) = 0$,
- $\forall x \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right)$, $\phi(x) = \lambda|x|_0 = \lambda$,
- $\forall x \in \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right) \setminus \{0\}$, $\phi(x) > \phi_{\text{CEL0}}(\|a_i\|, \lambda; x)$

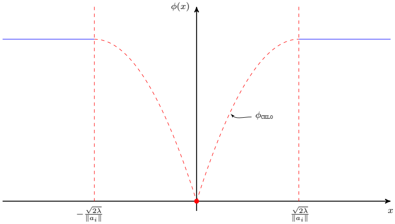


$\mathbb{1}_{\{x \in D\}} = 1$ if $x \in D$; 0 otherwise.

Theorem (NS conditions for (P1))

$\tilde{G}$ has property (P1) $\forall d \in \mathbb{R}$ iff ϕ verifies:

- $\phi(0) = 0$,
- $\forall x \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right)$,
  $\phi(x) = \lambda |x|_0 = \lambda$,
- $\forall x \in \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right) \setminus \{0\}$,
  $\phi(x) > \phi_{\text{CEL0}}(\|a_i\|, \lambda; x)$



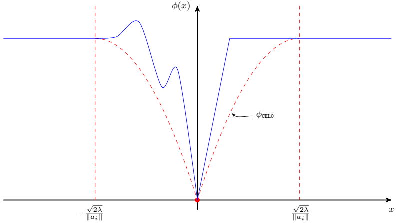
$$\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$$

$\mathbb{1}_{\{x \in D\}} = 1$ if x $\in D$; 0 otherwise.

Theorem (NS conditions for (P1))

$\tilde{G}$ *has property (P1)* $\forall d \in \mathbb{R}$ *iff* ϕ *verifies:*

- $\phi(0) = 0$,
- $\forall x \in \mathbb{R} \setminus \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right)$, $\phi(x) = \lambda|x|_0 = \lambda$,
- $\forall x \in \left(-\frac{\sqrt{2\lambda}}{\|a_i\|}, \frac{\sqrt{2\lambda}}{\|a_i\|}\right) \setminus \{0\}$, $\phi(x) > \phi_{\text{CEL0}}(\|a_i\|, \lambda; x)$



$$\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$$

$\mathbb{1}_{\{\text{x} \in D\}} = 1$ if x $\in D$; 0 otherwise.

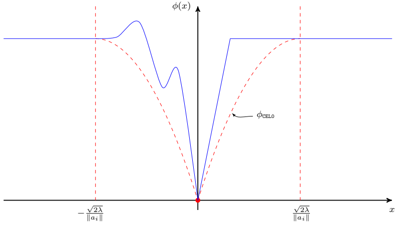
Theorem (NS conditions for (P1)-(P2))

$\tilde{g}$ has property (P1) and (P2) $\forall d \in \mathbb{R}$ iff in addition to the previous conditions, ϕ verifies:

- $\forall x \in B \setminus \{0\}, \ \lim\limits_{\substack{v \to x \\ v<x}} \phi'(v) > \lim\limits_{\substack{v \to x \\ v>x}} \phi'(v)$
- $\forall x \in (\beta^-, \beta^+) \setminus B, \ \phi''(x) \leq -\|a_i\|^2$

  $\exists v \in \mathcal{V}(x), \ \phi''(v) < -\|a_i\|^2$

for $\beta^- \in \left[-\frac{\sqrt{2\lambda}}{\|a_i\|}, 0\right)$ *and* $\beta^+ \in \left(0, \frac{\sqrt{2\lambda}}{\|a_i\|}\right]$.



$$\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$$

$\mathbb{1}_{\{\text{x} \in D\}} = 1$ if x $\in D$; 0 otherwise.

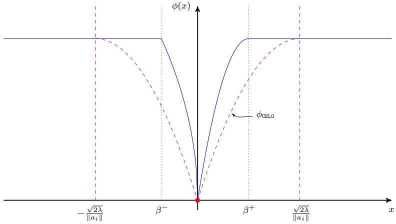
Theorem (NS conditions for (P1)-(P2))

$\tilde{g}$ has property (P1) and (P2) $\forall d \in \mathbb{R}$ iff in addition to the previous conditions, ϕ verifies:

- $\forall x \in B \setminus \{0\}, \ \lim\limits_{\substack{v \to x \\ v < x}} \phi'(v) > \lim\limits_{\substack{v \to x \\ v > x}} \phi'(v)$
- $\forall x \in (\beta^-, \beta^+) \setminus B, \ \phi''(x) \leq -\|a_i\|^2$

  $\exists v \in \mathcal{V}(x), \ \phi''(v) < -\|a_i\|^2$

for $\beta^- \in \left[-\frac{\sqrt{2\lambda}}{\|a_i\|}, 0\right)$ and $\beta^+ \in \left(0, \frac{\sqrt{2\lambda}}{\|a_i\|}\right]$.



$$\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$$

$\mathbb{1}_{\{x \in D\}} = 1$ if x $\in D$; 0 otherwise.

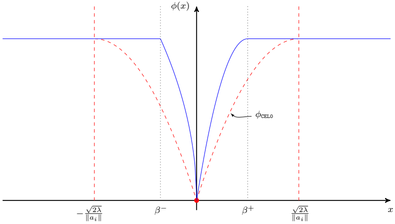
Theorem (NS conditions for (P1)-(P2))

$\tilde{g}$ has property (P1) and (P2) $\forall d \in \mathbb{R}$ iff in addition to the previous conditions, ϕ verifies:

- $\forall x \in B \setminus \{0\}$, $\lim\limits_{\substack{v \to x \\ v < x}} \phi'(v) > \lim\limits_{\substack{v \to x \\ v > x}} \phi'(v)$
- $\forall x \in (\beta^-, \beta^+) \setminus B$, $\phi''(x) \leq -\|a_i\|^2$

  $\exists v \in \mathcal{V}(x)$, $\phi''(v) < -\|a_i\|^2$

for $\beta^- \in \left[-\frac{\sqrt{2\lambda}}{\|a_i\|}, 0\right)$ *and* $\beta^+ \in \left(0, \frac{\sqrt{2\lambda}}{\|a_i\|}\right]$.



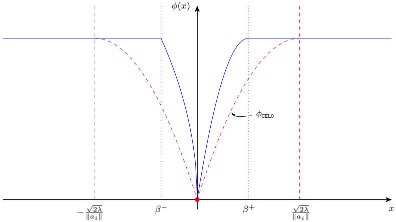
$$\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x| \leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$$

$\mathbb{1}_{\{\mathrm{x} \in D\}} = 1$ if x $\in D$; 0 otherwise.

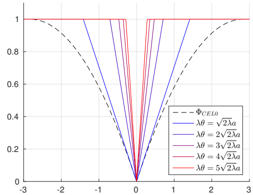
Proof is based on characterization of minimizers of G_{ℓ_0} [Nikolova 13] and critical points of $\tilde{\mathrm{G}}$.

With conditions (P1) and (P2), ϕ **depends on** $\|a_i\|$ and λ when applied on x_i:

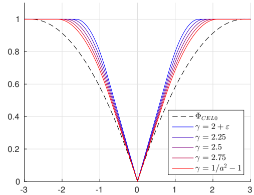
$$\tilde{G}(x) := \frac{1}{2}\|Ax - d\|_2^2 + \sum_{i \in \mathbb{I}_N} \phi(\|a_i\|, \lambda, x_i)$$



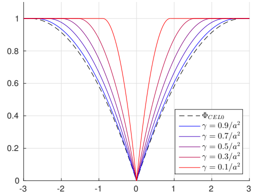
4. ℓ_2-ℓ_0 optimization by continuous relaxation



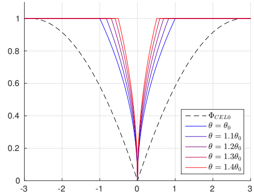
(a) Capped-ℓ_1 [Zhang 09]



(b) SCAD [Fan and Li 01]



(c) MCP [Zhang 10]



(d) Truncated-ℓ_p

Figure: Examples of penalties for which (P1) (Top) or (P1) and (P2) (Bottom) hold for $a = 0.5$, $\lambda = 1$ and $d = 1.8$.

The function $\phi_{\texttt{CEL0}}$ is a Minimax Concave Penalty (MCP) [Zhang 10].

Examples using state of the art penalties

Penalty	Def $\phi(u)$	P1	P2	Conditions
Cap-ℓ_1 [Zhang 09]	$\lambda \min\{\theta\lvert u\rvert, 1\}$	✓	✗	$\lambda\theta \geq \sqrt{2\lambda}\lVert a_i\rVert$
SCAD [Fan and Li 01]	$\begin{cases} \tilde\lambda\lvert u\rvert & \text{if } \lvert u\rvert \leq \tilde\lambda, \\ \frac{2\gamma\tilde\lambda\lvert u\rvert-\tilde\lambda^2-u^2}{2(\gamma-1)} & \text{if } \tilde\lambda < \lvert u\rvert \leq \gamma\tilde\lambda, \\ \frac{(\gamma+1)\tilde\lambda^2}{2} & \text{if } \lvert u\rvert > \gamma\tilde\lambda \end{cases}$	✓	✗	$\frac{(\gamma+1)\tilde\lambda^2}{2} = \lambda$ $2 < \gamma \leq \frac{1}{\lVert a_i\rVert} - 1$
MCP [Zhang 10]	$\begin{cases} \lambda & \text{if } \lvert u\rvert > \sqrt{2\lambda\gamma_i} \\ \left(\sqrt{\frac{2\lambda}{\gamma_i}}\lvert u\rvert - \frac{u^2}{2\gamma_i}\right) & \text{if } \lvert u\rvert \leq \sqrt{2\lambda\gamma_i} \end{cases}$	✓	✓	$\gamma_i < \frac{1}{\lVert a_i\rVert^2}$
Trunc-ℓ_p	$\lambda \min\{\theta_i\lvert u\rvert^{p_i}, 1\}$	✓	✓	$\theta_i \geq \left(\frac{\lVert a_i\rVert^2}{p_i(1-p_i)\lambda}\right)^{p_i/2}$

$$\tilde{\mathrm{G}}(x) := \frac{1}{2}\lVert A\mathrm{x} - d\rVert_2^2 + \sum_{i\in\mathbb{I}_N} \phi(\lVert a_i\rVert, \lambda, \mathrm{x}_i)$$

$$\phi_{\texttt{CEL0}}(\lVert a_i\rVert, \lambda, \mathrm{x}_i) = \phi_{\texttt{MCP}}(\gamma_i, \lambda, \mathrm{x}_i) \text{ for } \gamma_i = \frac{1}{\lVert a_i\rVert^2}$$

The $\ell_2 - \ell_0$ and ℓ_2- CEL0 functionals :

$$G_{\ell_0}(x) := \frac{1}{2}\|Ax - d\|^2 + \lambda\|x\|_0$$

$$G_{\text{CEL0}}(x) = \frac{1}{2}\|Ax - d\|^2 + \sum_{i\in\mathbb{I}_N} \phi_{\text{CEL0}}(\|a_i\|, \lambda, x_i)$$

where $\phi_{\text{CEL0}}(\|a_i\|, \lambda, x) = \lambda - \frac{\|a_i\|^2}{2}\left(|x| - \frac{\sqrt{2\lambda}}{\|a_i\|}\right)^2 \mathbb{1}_{\left\{|x|\leq \frac{\sqrt{2\lambda}}{\|a_i\|}\right\}}$

Properties of $G_{\text{CEL0}}(x)$

- **Limit inf** of the functions satisfying (P1) and (P2)
- **Convex hull** if A diagonal or orthogonal ($A^T A$ diagonal)
- **Continuity**
- **Non convex** in the general case (for any A)
- but **convexity** with respect to each **component**

Nonsmooth nonconvex algorithms

The **continuity of** G_{CEL0} allows to use recent *nonsmooth nonconvex* algorithms to minimize (indirectly) G_{ℓ_0},

- *Difference of Convex* (DC) functions programming [Gasso et al 09]
- *Majorization-Minimization*(MM) algorithms (*e.g.* Iteratively Reweighted ℓ_1 (IRL1) [Ochs et al 2015])
- *Forward-Backward splitting* (GIST [Gong et al 13], [Attouch et al 13])

Forward-Backward Splitting Algorithm

$$\mathrm{x}^{k+1} \in \mathrm{prox}_{\gamma \Phi_{\mathrm{CEL0}}(\cdot)}\left(\mathrm{x}^k - \gamma^k A^T (A\mathrm{x}^k - d)\right),$$

where $0 < \gamma < \frac{1}{\|A\|^2}$ and

$$\mathrm{prox}_{\gamma \phi_{\mathrm{CEL0}}(a,\lambda;\cdot)}(u) = \begin{cases} \mathrm{sign}(u) \min\left(|u|, (|u| - \sqrt{2\lambda}\gamma a)_+/(1 - a^2\gamma)\right) & \text{if } a^2\gamma < 1 \\ u\mathbb{1}_{\{|u|>\sqrt{2\gamma\lambda}\}} + \{0, u\}\mathbb{1}_{\{|u|=\sqrt{2\gamma\lambda}\}} & \text{if } a^2\gamma \geq 1 \end{cases}$$

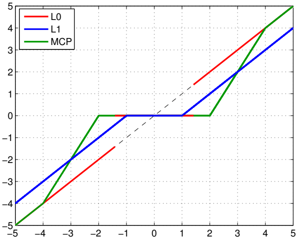


Figure: Proximal operators. Red: ℓ_0, Blue: ℓ_1, Green: Φ_{CEL0} (depends on $a = \|a_i\|$ at component $u = \mathrm{x}_i$).

Forward-Backward Splitting Algorithm

$$\mathrm{x}^{k+1} \in \mathrm{prox}_{\gamma\Phi_{\mathtt{CEL0}}(\cdot)}\left(\mathrm{x}^k - \gamma^k A^T(A\mathrm{x}^k - d)\right),$$

where $0 < \gamma < \frac{1}{\|A\|^2}$ and

$$\mathrm{prox}_{\gamma\phi_{\mathtt{CEL0}}(a,\lambda;\cdot)}(u) = \begin{cases} \mathrm{sign}(u)\min\left(|u|, (|u| - \sqrt{2\lambda}\gamma a)_+/(1-a^2\gamma)\right) & \text{if } a^2\gamma < 1 \\ u\mathbb{1}_{\{|u|>\sqrt{2\gamma\lambda}\}} + \{0,u\}\mathbb{1}_{\{|u|=\sqrt{2\gamma\lambda}\}} & \text{if } a^2\gamma \geq 1 \end{cases}$$

- Convergence to a critical point under Kurdyka-Lojaseiwicz (KL) property [Attouch et al 13].
- Accelerated algorithm in the non convex case [Li Lin 15]

Outline of the talk

1. Introduction and examples
2. Iterative Hard Thresholding
3. Greedy algorithms
4. Continuous relaxation
5. **Exact reformulation** ([Bi et al 14, Yuan & Ghanem 16, Liu et al 18], ,...)
6. Some results on super-resolution microscopy
7. Conclusion

Exact reformulation

- *Class of continuous nonconvex penalties* → asymptotic connections with the ℓ_2-ℓ_0 criteria [Chouzenoux et al 13]
- *Reformulation using Difference of Convex functions* → asymptotic or local minimizer results [Le Thi et al 14, Le Thi et al 15]
- *Equivalence of ℓ_0- and ℓ_p-norm ($0 < p \leq 1$) minimization under linear equalities or inequalities (e.g. exact reconstruction problem)* [Fung and Mangasarian 11]
- *Reformulation and optimization through Mixed-Integer Programs (MIPs)* → global optimum for problems of reasonable size (a few hundred variables) [Bourguignon et al 15]
- **Exact reformulation** ([Bi et al 14, Yuan & Ghanem 16, Liu et al 18], ,...)

5. Exact reformulation of ℓ_0: Penalized reformulation

Lemma 1 [Liu et al 18, Yuan & Ghanem 16]

$$\|x\|_0 = \min_{-1 \leq u \leq 1} \|u\|_1 \text{ s.t } \|x\|_1 = < u, x >$$

Exact reformulation for the $\ell_2 - \ell_0$ penalized problem

Initial problem:

$$\min_{x} \frac{1}{2}\|Ax - d\|_2^2 + \lambda\|x\|_0$$

Penalized reformulation:

$$\min_{x,u} G_\rho(x, u) := \frac{1}{2}\|Ax - d\|^2 + \iota_{\{-1 \leq \cdot \leq 1\}}(u) + \lambda\|u\|_1 + \rho(\|x\|_1 - < x, u >)$$

with $\iota_{\{x \in D\}}(x) = 0$ if $x \in D$, $+\infty$ otherwise.

Theorem [Bechensteen,et al.]

If $\rho > \sigma_{max}(A)\|d\|_2$, and A is of full rank. Then:

1. If (x_ρ, u_ρ) is a local (respectively global) minimizer of G_ρ, then x_ρ is a local (respectively global) minimizer of the initial problem.
2. If $\hat{x}$ is a global minimizer of the initial problem, then $(\hat{x}, \hat{u})$ is a global minimizer of G_ρ with $\hat{u}$ associated with Lemma 1.

5. Exact reformulation of ℓ_0: Constrained reformulation

Lemma 1 [Liu et al 18, Yuan & Ghanem 16]

$$\|\mathrm{x}\|_0 = \min_{-1\leq \mathrm{u} \leq 1} \|\mathrm{u}\|_1 \text{ s.t } \|\mathrm{x}\|_1 = <\mathrm{u},\mathrm{x}>$$

Exact reformulation for the $\ell_2 - \ell_0$ constrained problem

Initial problem:

$$\min_{\mathrm{x}} \frac{1}{2}\|A\mathrm{x}-d\|_2^2 + \iota_{\{\|\cdot\|_0 \leq K\}}(\mathrm{x})$$

Constrained reformulation:

$$\min_{\mathrm{x},\mathrm{u}} G_\rho(\mathrm{x},\mathrm{u}) := \frac{1}{2}\|A\mathrm{x}-d\|^2 + \iota_{\{\cdot\geq 0\}}(\mathrm{x}) + \iota_{\{-1\leq\cdot\leq 1\}}(\mathrm{u}) + \iota_{\{\|\cdot\|_1\leq K\}}(u) + \rho(\|\mathrm{x}\|_1 - <\mathrm{x},\mathrm{u}>)$$

Theorem [Bechensteen,et al.]

If $\rho > \sigma_{max}(A)\|d\|_2$, and A is of full rank. Then:

1. If (x_ρ, u_ρ) is a local (respectively global) minimizer of G_ρ, then x_ρ is a local (respectively global) minimizer of the initial problem.
2. If $\hat{x}$ is a global minimizer of the initial problem, then $(\hat{x}, \hat{u})$ is a global minimizer of G_ρ with $\hat{u}$ associated with Lemma 1.

Why minimize the constrained or penalized reformulation instead of their initial formulation?

Constrained reformulation:

$$\min_{x,u} \frac{1}{2}\|Ax - d\|^2 + \iota_{\{\cdot \geq 0\}}(x) + \iota_{\{-1 \leq \cdot \leq 1\}}(u) + \iota_{\{\|\cdot\|_1 \leq K\}}(u) + \rho(\|x\|_1 - < x, u >)$$

Penalized reformulation:

$$\min_{x,u} \frac{1}{2}\|Ax - d\|^2 + \iota_{\{\cdot \geq 0\}}(x) + \iota_{\{-1 \leq \cdot \leq 1\}}(u) + \lambda\|u\|_1 + \rho(\|x\|_1 - < x, u >)$$

- Biconvex
- Non-convexity linked to the coupling term $< x, u >$
- Minimizing the reformulation is equivalent to minimize the initial problem regarding local and global minimizers

We add a positivity constraint on x and we finally define

$$G_\rho(\mathrm{x},\mathrm{u}) = \frac{1}{2}\|A\mathrm{x}-d\|^2 + \iota_{\{\cdot\geq 0\}}(\mathrm{x}) + \rho\|\mathrm{x}\|_1 + \iota_{\{\|\cdot\|_1\leq K\}}(\mathrm{u}) + \iota_{\{-1\leq\cdot\leq 1\}}(\mathrm{u}) - \rho < \mathrm{x},\mathrm{u} >$$

The global optimization scheme is (continuation method)

Initialize: $\rho^0 > 0, n = 0$

Repeat: Solve the problem G_{ρ^n}:

$$\{\mathrm{x}^{n+1},\mathrm{u}^{n+1}\} = \arg\min_{\mathrm{x},\mathrm{u}} G_{\rho^n}(\mathrm{x},\mathrm{u})$$

Update: $\rho^{n+1} = \alpha\rho^n$, $\alpha > 1$

Until: $\rho^{n+1} > \sigma_{max}(A)\|d\|_2$

$$G_{\rho^n}(\mathrm{x},\mathrm{u}) = \frac{1}{2}\|A\mathrm{x}-d\|^2 + \iota_{\{\cdot\geq 0\}}(\mathrm{x}) + \rho^n\|\mathrm{x}\|_1 + \iota_{\{\|\cdot\|_1\leq K\}}(\mathrm{u}) + \iota_{\{-1\leq\cdot\leq 1\}}(\mathrm{u}) - \rho^n < \mathrm{x},\mathrm{u} >$$

At fixed ρ^n we apply the Proximal Alternate Minimization (PAM) algorithm [Attouch & al 10]

Initialize: $\mathrm{u}^0 = 0 \in \mathbb{R}^M$

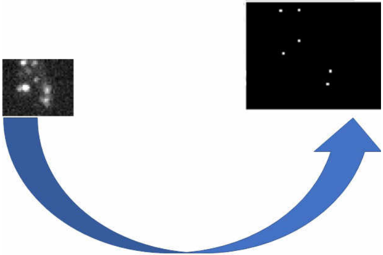
Repeat: $\arg\min G_{\rho^n}$ using alternate minimizations

- $\{\mathrm{x}^{n+1}\} = \arg\min_{\mathrm{x}} G_{\rho^n}(\mathrm{x},\mathrm{u}^n) + \frac{1}{2c^n}\|\mathrm{x}-\mathrm{x}^n\|^2$
  $\rightarrow$ FISTA Algorithm [Beck et al 09]

- $\{\mathrm{u}^{n+1}\} = \arg\min_{\mathrm{u}} G_{\rho^n}(\mathrm{x}^{n+1},\mathrm{u}) + \frac{1}{2d^n}\|\mathrm{u}-\mathrm{u}^n\|^2$
  $\rightarrow$ Algorithm [Stefanov, 2004]

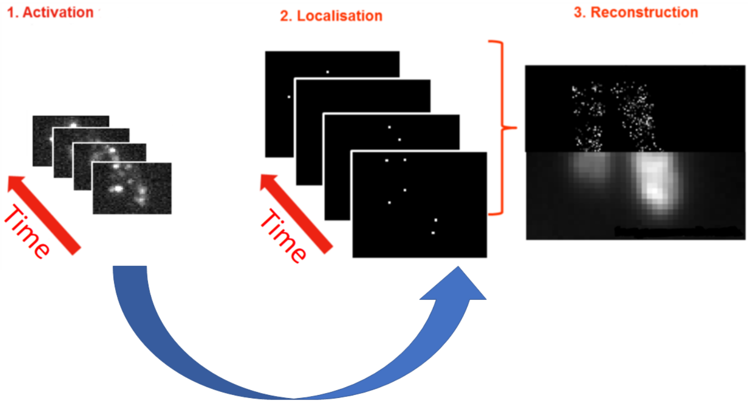
Until: convergence

Convergence of the algorithm towards a critical point of G_{ρ^n} for c^n and d^n such that $0 < r_- < c^n, d^n < r_+$ and under KL condition on G_{ρ^n} and assuming that x_n and u_n are bounded [Attouch & al 10].



$$\hat{x} \in \arg\min_{x} \frac{1}{2}\|Ax - d\|_2^2 + \iota_{\{\cdot \geq 0\}}(x) + R(x)$$

6. Results: Single-Molecule Localization Microscopy



$$\hat{x} \in \arg\min_{x} \frac{1}{2}\|Ax - d\|_2^2 + \iota_{\{\cdot \geq 0\}}(x) + R(x)$$

6. Results, ISBI challenge 2013, simulated dataset

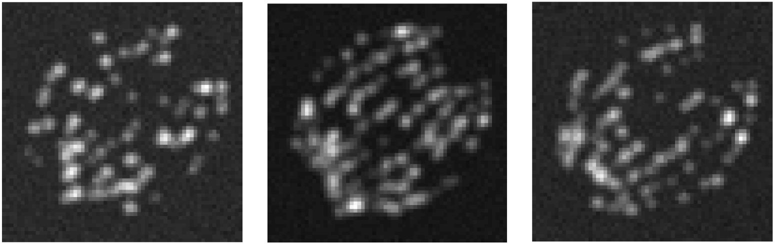


Figure: Simulated images (among the 361 simulated high density images for this sample). Data from IEEE ISBI Challenge 2013.
http://bigwww.epfl.ch/smlm/datasets/index.html

8 simulated tubes of 30nm diameter
Camera of 64×64 pixels of size 100nm.
Gaussian PSF, FWHM = 258.21 nm (full width at half maximum)
80932 molecules activated on 361 frames.

6. Results, ISBI challenge 2013, simulated dataset

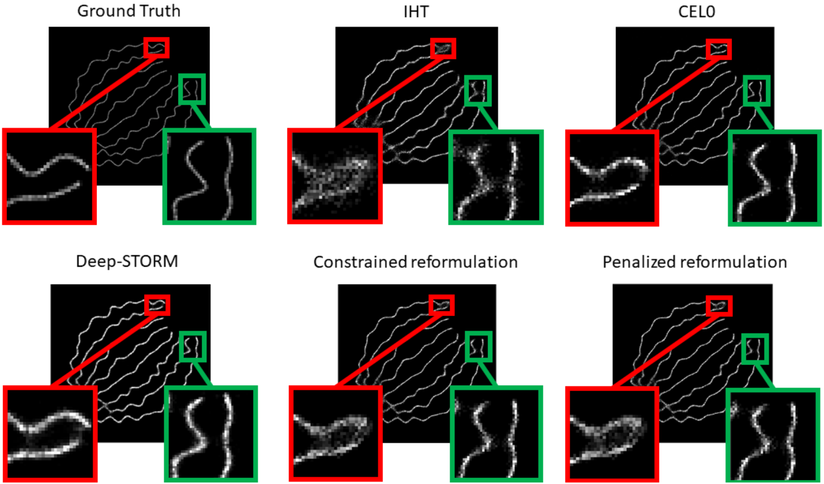
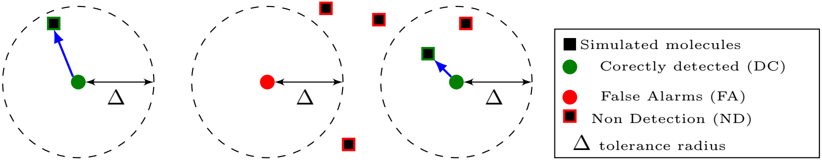


Figure: Reconstruction from simulated data set, reduction ratio $L = 4$.

6. Results, ISBI challenge 2013, simulated dataset

Jaccard index calculus



$$\text{Jaccard index} = \frac{\text{DC}}{\text{DC} + \text{FA} + \text{ND}}$$

Jaccard index results

	Jaccard index (%)			
Method - Tolerance (nm)	50	100	150	200
IHT	20.1	35.9	40.4	41.3
CEL0	**29.3**	**41.3**	42.4	42.6
Constrained reformulation	25.2	40.0	**43.2**	**43.9**
Penalized reformulation	25.0	39.3	42.2	42.8
Deep-STORM	×	×	×	×

Table: The jaccard index obtained and the tolerance

6. Results, ISBI challenge 2013, Real dataset

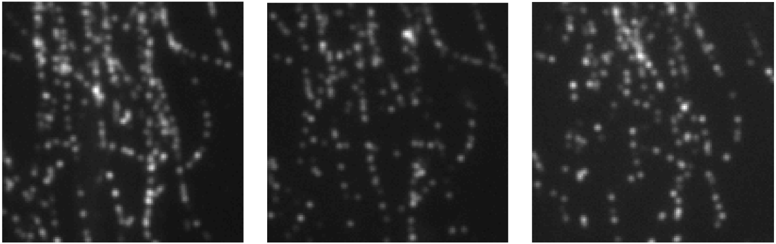


Figure: Real images (among the 500 real high density images for this sample). Data from IEEE ISBI Challenge 2013.
http://bigwww.epfl.ch/smlm/datasets/index.html

Camera of 128×128 pixels of size 100nm.
Gaussian PSF, FWHM = 358.1 nm (full width at half maximum)

6. Results, ISBI challenge 2013, Real dataset

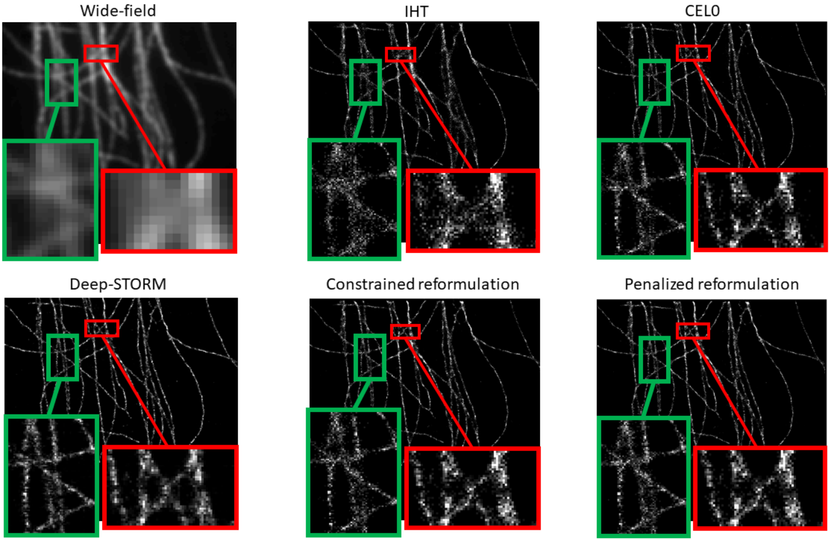


Figure: Reconstruction from the real data set, reduction ratio $L = 4$.

7. Concluding remarks

Synthesis

- IHT: simple, but bad local minimizer.
- Greedy: advanced versions can be efficient but complexity increased
- Continuous relaxation:
  - Penalized problem
  - Continuous Exact ℓ_0: preserve global minimizers, can remove local ones, non convex optimization,
- Exact reformulation:
  - Penalized and constrained problems
  - Double size problem: biconvex optimization, can be applied with any data term (not only least square).

Still active research topic

- Exact continuous relaxation for the **constraint problem**,
- More studies on **non-quadratic** data fidelity terms,
- Efficient algorithms are still needed for non convex continuous optimization,
- **Gridless** method [Catala, Duval, Peyre 2019].

Thanks to

- **Gilles Aubert**, Professor of Mathematics, emeritus, UCA, Nice France.
- **Emmanuel Soubies**, young researcher at CNRS, IRIT Toulouse, France.
- **Arne Bechensteen**, PhD student, Morpheme, I3S-INRIA SAM, Sophia Antipolis France.
- **Eric Debreuve**, researcher at CNRS, Morpheme, I3S-INRIA SAM, Sophia Antipolis France.
- **Simon Bahadoran**, past intern, Morpheme, I3S-INRIA SAM, Sophia Antipolis France.

References I

Arne Bechensteen, Laure Blanc-Féraud and Gilles Aubert, *Reformulation of l_2-l_0 constrained criterion for SMLM*, preprint, 2018.

Hedy Attouch, Jérôme Bolte, Patrick Redont and Antoine Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Lojasiewicz inequality*, Mathematics of operation research, 35(2), (2010), pp. 438–457.

Hedy Attouch, Jérôme Bolte, and Benar Fux Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods*, Mathematical Programming, 137 (2013), pp. 91–129.

Amir Beck and Marc Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202.

S. Bi, X. Liu, and S. Pan, *Exact penalty decomposition method for zero-norm minimization based on MPEC formulation*, SIAM Journal on Scientific Computing, 36(4) (2014).

Thomas Blumensath and Mike E Davies, *Iterative thresholding for sparse approximations*, Journal of Fourier Analysis and Applications, 14 (2008), pp. 629–654.

Sébastien Bourguignon, Jordan Ninin, Hervé Carfantan and Marcel Mongeau, *Optimisation exacte de critères parcimonieux en norme ℓ_0 par programmation mixte en nombres entiers*, Colloque GRETSI 2015.

Leo Breiman, *Better subset regression using the nonnegative garrote*, Technometrics, 37 (1995), pp. 373–384.

Betzig, Eric and Patterson, George H and Sougrat, Rachid and Lindwasser, O Wolf and Olenych, Scott and Bonifacino, Juan S and Davidson, Michael W and Lippincott-Schwartz, Jennifer and Hess, Harald F, *Imaging intracellular fluorescent proteins at nanometer resolution*, Science, 5793 (2006), pp. 1642-1645.

References II

Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on, 52 (2006), pp. 489–509.

Emmanuel J Candès, Michael B Wakin, and Stephen P Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, Journal of Fourier analysis and applications, 14 (2008), pp. 877–905.

Emmanuel J Candès, and Michael B Wakin, *An introduction to compressive sampling*, IEEE Signal Processing Magazine, 25(2), (2008), pp. 21–30.

P. Catala, V. Duval and G. Peyre, *A low - rank approach to off - the - grid sparse deconvolution*, NCMIP conf 2017, arXiv 2019.

Rick Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, Signal Processing Letters, IEEE, 14 (2007), pp. 707–710.

S. Chen, S. Billings and W. Luo, *Orhogonal least squares methods and their application to non-linear system identification*, International journal of Control, 50(5) (1989), pp. 1873–1896.

Scott Shaobing Chen, David L Donoho, and Michael A Saunders, *Atomic decomposition by Basis Pursuit*, SIAM journal on scientific computing, 20 (1998), pp. 33–61.

Emilie Chouzenoux, Anna Jezierska, Jean-Christophe Pesquet and Hugues Talbot, *A majorize-minimize subspace approach for \ell_ 2-\ell_ 0 image regularization*, SIAM Journal on Imaging Sciences, 6 (2013), pp.563–591.

Frank H Clarke, *Optimization and nonsmooth analysis*, vol. 5, Siam, 1990.

References III

Ingrid Daubechies, Michel Defrise, and Christine De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1413–1457.

Geoff Davis, Stéphane Mallat and Marco Avellaneda *Adaptive greedy approximations*, Constructive approximation, 13 (1997), pp. 57–98.

David L Donoho and Micael Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, in Proceedings of the National Academy of Sciences 100(5) (2003), pp. 72–76.

David L Donoho, *For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution*, Communications on Pure and Applied Mathematics, 59 (2006), pp. 797–829.

Jianqing Fan and Runze Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, 96 (2001), pp. 1348–1360.

Simon Foucart and Ming-Jun Lai, *Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$*, Applied and Computational Harmonic Analysis, 26 (2009), pp. 395–407.

G.M. Fung and O.L. Mangasarian, *Equivalence of minimal ℓ_0- and ℓ_p-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p*, Journal of optimization theory and applications, 151 (2011), pp. 1–10.

Gilles Gasso, Alain Rakotomamonjy, and Stéphane Canu, *Recovering sparse signals with a certain family of nonconvex penalties and DC programming*, Signal Processing, IEEE Transactions on, 57 (2009), pp. 4686–4698.

References IV

Pinghua Gong, Changshui Zhang, Zhaosong Lu, Jianhua Huang, and Jieping Ye, *A General Iterative Shrinkage and Thresholding Algorithm for Non-convex Regularized Optimization Problems*, in Proceedings of The 30th International Conference on Machine Learning, 2013, pp. 37–45.

Remi Gribonval and Morten Nielsen, *Sparse representation in unions of bases*, IEEE Transactions on Information Theory, 49(12), (2003), pp. 73–76.

Hess, Samuel T and Girirajan, Thanu PK and Mason, Michael D *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy* Biophysical journal, 11 (2006), Elsevier, pp. 4258-4272.

Cédric Herzet and Angélique Drémeau, *Bayesian Pursuit Algorithms* , in Proceedings of European Signal Processing conference (EUSIPCO), Aalborg, Danemark, August 2010.

Seamus J Holden, Stephan Uphoff and Achillefs N Kapanidis *DAOSTORM: an algorithm for high-density super-resolution microscopy* Nature Methods, 8 (2011), pp. 279–280.

P. Jain, A. Tewari and I.S.Dhillon *Orthogonal Matching Pursuit with Replacement* Advanced in Neural Information Processing Systems, 24 (2011), pp. 1215–1223.

Hoai An Le Thi, Hoai Minh Le and Tao Pham Dinh, *Feature selection in machine learning: an exact penalty approach using a Difference of Convex function Algorithm*, Machine Learning, 2014, pp. 1–24.

Hoai An Le Thi, Tao Pham Dinh, Hoai Minh Le and Xuan Thanh Vo, *DC approximation approaches for sparse optimization*, European Journal of Operational Research, 244 (2015), pp. 26–46.

References V

Huan Li and Zhouchen Lin, *Accelerated Proximal Gradient Methods for Nonconvex Programming*, Part of: Advances in Neural Information Processing Systems 28 (NIPS 2015), pp. 679?704.

Yulan Liu, Shujun. Bi, and Shoahua Pan, *Equivalent Lipschitz surrogates for zero-norm and rank optimization problems.*, Journal of Global Optimization, 72 (4), (2018), pp. 679?704.

Stéphane G Mallat and Zhifeng Zhang, *Matching pursuits with time-frequency dictionaries*, Signal Processing, IEEE Transactions on, 41 (1993), pp. 3397–3415.

OL Mangasarian *Machine learning via polyhedral concave minimization* Applied Mathematics and Parallel Computing (1996), pp. 175–188.

Min, J. and Vonesch, C. and Kirshner, H. and Carlini, L. and Olivier, N. and Holden, S. and Unser, M. *FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data.* Scientific Reports, 4 (2014), pp. 4577.

Hosein Mohimani, Massoud Babaie-Zadeh and Christian Jutten, *A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm*, Signal Processing, IEEE Transactions on, 57 (2008), pp. 289–301.

Balas Kausik Natarajan *Sparse approximate solutions to linear systems*, SIAM journal on computing, 24 (1995), pp. 227–234.

Mila Nikolova, *Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 904–937.

Mila Nikolova, *Relationship between the optimal solutions of least squares regularized with L0-norm and constrained by k-sparsity*, Appl. Comput. Harmon. Anal., 41(1), (2016), pp. 237–265.

References VI

P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, *An iteratively reweighted Algorithm for Non-smooth Non-convex Optimization in Computer Vision*, SIAM Journal on Imaging Sciences, 8(1), 2015.

Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad, *Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition*, in Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, IEEE, 1993, pp. 40–44.

Rust, Michael J and Bates, Mark and Zhuang, Xiaowei *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*, Nature methods, 10 (2006), pp. 793-796.

Daniel Sage, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley and Michael Unser, *Quantitative evaluation of software packages for single-molecule localization microscopy*, in Nature methods, 2015, pp. 717–724.

Stefan M. Stefanov *Convex quadratic minimization subject to a linear constraint and box constraints* Applied Mathematics Research Express, 1 (2004), pp. 17-42.

Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert *A Continuous Exact l0 Penalty (CEL0) for Least Squares Regularized Problem*, SIAM Journal on Imaging Sciences, 8(3), 2015.

Emmanuel Soubies, Laure Blanc-Féraud and Gilles Aubert *A Unified View of Exact Continuous Penalties for l2-l0 Minimization*, SIAM Journal of Optimization, 27(3), 2017.

Charles Soussen, Jérôme Idier, David Brie, and Junbo Duan, *From Bernoulli–Gaussian deconvolution to sparse signal restoration*, Signal Processing, IEEE Transactions on, 59 (2011), pp. 4572–4584.

References VII

Charles Soussen, Jérôme Idier, Junbo Duan and David Brie, *Homotopy Based Algorithms for ℓ_0-Regularized Least-Squares*, Signal Processing, IEEE Transactions on, 63(13) (2015), pp. 3301–3316.

R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, 46 (1996), pp. 431–439.

Joel A Tropp, *Greed is good: Algorithmic results for sparse approximation*, Information Theory, IEEE Transactions on, 50 (2004), pp. 2231–2242.

Ganzhao Yuan and Bernard Ghanem, *Sparsity Constrained Minimization via Mathematical Programming with Equilibrium Constraints*, arXiv:1608.04430 (2016).

Cun-Hui Zhang, *Multi-stage convex relaxation for learning with sparse regularization*, Advances in Neural Information Processing Systems, (2009), pp. 1929–1936.

Cun-Hui Zhang, *Nearly unbiased variable selection under minimax concave penalty*, The Annals of Statistics, (2010), pp. 894–942.

Hui Zou, *The adaptive lasso and its oracle properties*, Journal of the American statistical association, 101 (2006), pp. 1418–1429.