Lecture 3: Models and algorithms for ℓ_2-ℓ_0 optimisation problems

Luca Calatroni
CR CNRS, Laboratoire I3S
CNRS, UCA, Inria SAM, France

MIVA ERASMUS BIP PhD winter school

Advanced methods for mathematical image analysis

University of Bologna, IT

January 18-20 2022

Table of contents

Introduction

Why ℓ_0?

Many problems in signal/image processing are concerned with **sparse recovery**: compressed sensing, variable selection, source separation, learning...

$$d = Ax + n$$

- $d \in \mathbb{R}^m$: observed data (signal processing notation)
- $x \in \mathbb{R}^n$ unknown solution to be estimated
- $A \in \mathbb{R}^{m \times n}$ observation matrix,

- Few observations y and large explicative unknown variables x, with $m \ll n$. Undertermined system! A is ill-conditioned, noise is present.
- **Regularisation**: assume the signal is sparse by considering ℓ_1-norm or ℓ_0 pseudo-norm constraints:

$$\|x\|_1 \leq K, \qquad \|x\|_0 \leq K$$

with $\|x\|_0 := \#\{x_i,\ i = 1, \ldots, n : x_i \neq 0\} = \sum_{i=1}^n |x_i|_0$, with

$$|z|_0 = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

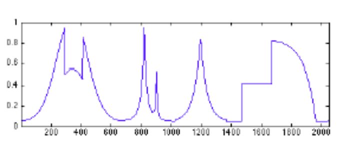
NB: ℓ_0-norm is NOT a norm as $\|\lambda x\|_0 = \|x\|_0 \neq \lambda \|x\|_0$.

Dictionary representation in imaging

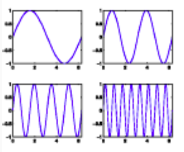
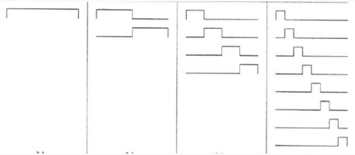
Image are heterogeneous signals, with smooth (homogeneous) areas, edges, texture,...

Take $d \in \mathbb{R}^m$ be a patch of an image or a signal





Each d is represented by given waveforms whose shape matches the image structure. Standard choices of a_i vectors come from Haar, smooth wavelets, sine/cosine transform...

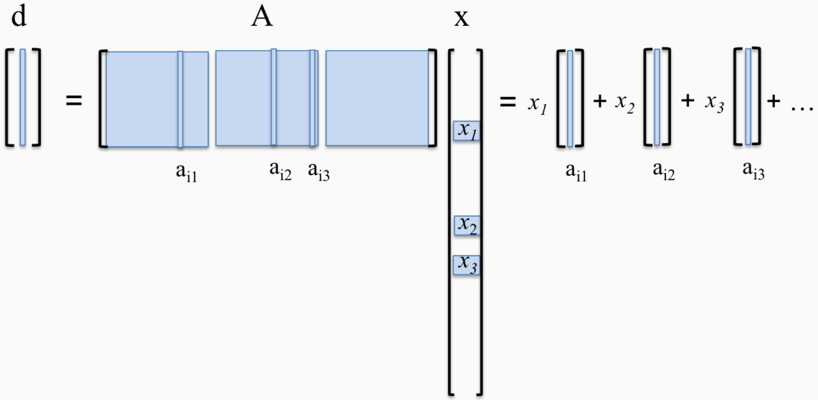




Take $A = [a_1, ..., a_n] \in \mathbb{R}^{m \times n}$ to be a set of normalised (basis) vectors.

Dictionary representation in imaging

- Such A is a redundant dictionary (sequence of representative waveforms)
- The dictionary A is **adapted** to the signal d if d can be represented by a few number of vectors a_i (atoms) of A, that is $d \approx Ax$ with x **sparse**, that is

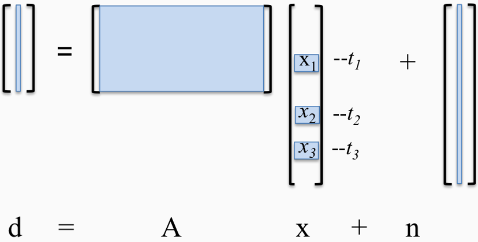
$$\|x\|_0 \leq K, \qquad K << n$$



Examples

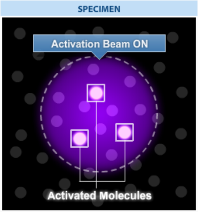
- signal is a sum of spikes, modelled by a sum of Dirac $\sum_{r=1}^{K} x_r \delta_{t_r}$.
- acquisition system is modelled as a convolution with a Gaussian function: $d(\cdot) = h * \sum_{r=1}^{K} x_r \delta_{t_r} = \sum_{r=1}^{K} x_r h(\cdot - t_r)$.

Assume that 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 Localisation in super-resolution microscopy

Single Molecule Localisation in super-resolution microscopy I

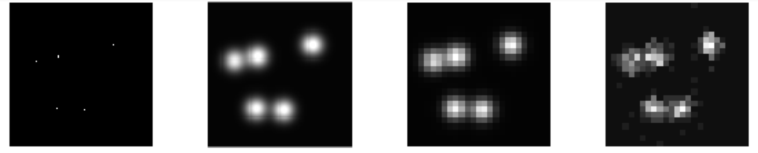


SMLM idea

Modelling: for $t \in \{1, \dots, T\}$, given a blurry, undersampled and noisy image $d_t \in \mathbb{R}^m$, consider the problem:

$$\text{find sparse} \quad x_t \qquad \text{s.t.} \qquad d_t = Ax_t + n_t, \quad \forall t \in \{1, \dots, T\}$$

$A := SH \in \mathbb{R}^{m \times n}$ with $H \in \mathbb{R}^{n \times n}$ convolution and $S \in \mathbb{R}^{m \times n}$ undersampling , $n = Lm$, $L > 1$.

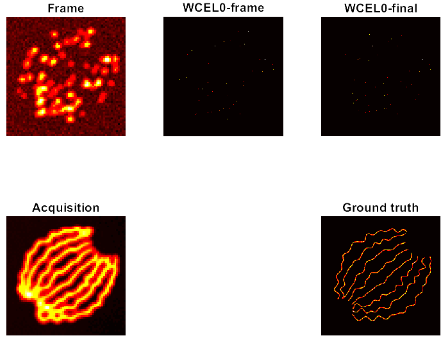


Single Molecule Localisation in super-resolution microscopy II

Regularisation approach: look for sparse solutions at each time $t \in \{1, \dots, T\}$

$$x_t^* \in \arg\min_x \ \frac{1}{2}\|Ax - d_t\|^2 + \lambda\|x\|_0 + \iota_{x\geq 0}(x), \qquad \lambda > 0$$

Final reconstruction obtained simply by $x = \sum_{i=1}^{T} x_t^*$ (Gazagnes, Soubies, Blanc-Féraud, Schaub, '15, Lazzaretti, Calatroni, Estatico, '21)



ℓ_2-ℓ_0 minimisation

ℓ_2-ℓ_0 minimisation

ℓ_2-ℓ_0: problem forms

For $A \in \mathbb{R}^{m\times n}$, $m \leq n$ consider the following formulations:

- **Exact recovery**:

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

- **Approximation problem in constrained forms** ($\epsilon > 0, K > 0$)

$$\hat{x} \in \arg\min_{x\in\mathbb{R}^n} \; \frac{1}{2}\|Ax - d\|_2^2 \; \text{ subject to } \; \|x\|_0 \leq K$$

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

- **Approximation problem in penalised form** ($\lambda > 0$)

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

- non-continuous, non-convex and NP-hard optimisation problem (Natarajan, '95, Davies et al., '97): *a solution cannot be verified in polynomial time w.r.t the dimension of the problem*
- Non equivalent formulations
- Existence of optimal solutions and relations between formulations in Nikolova, '16
- Very active field of research in signal and image processing, and in statistics.

How people do: ℓ_2-ℓ_1 minimisation

A popular way to deal with this problem consists in considering the ℓ_1-norm instead

ℓ_2-ℓ_1 problem formulations

- **Constrained formulation** ($K > 0$):

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

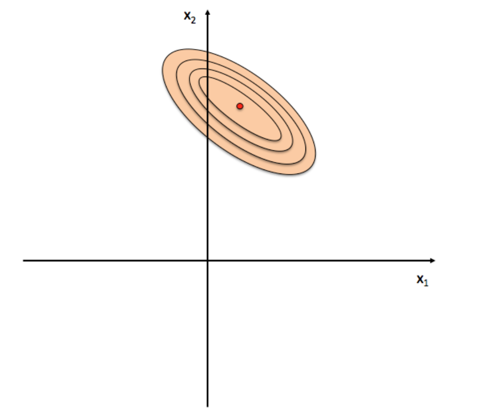
- **Penalised formulation** ($\lambda > 0$):

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

- Easier optimization problems: convex and continuous (but non smooth) $\rightarrow$ available solvers (see **previous courses**)!
- The two formulations are equivalent
- Under some conditions involving A, solving these problems allows to find a solution of the ℓ_2-ℓ_0 problem (Candès, Romberg, Tao, '05)
- They are known as Basis Pursuit De-Noising (BPDN) Chen et al., '98, or LASSO (Tibshirani, '96) problems, respectively.

ℓ_1 norm promotes sparsity

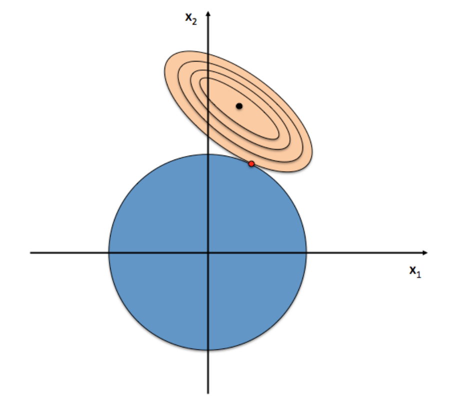
Standard example in $\mathbb{R}^2$.



Level lines of $\|Ax - d\|_2^2$.

ℓ_1 norm promotes sparsity

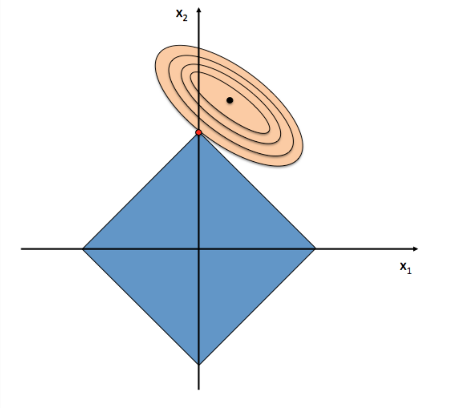
Standard example in $\mathbb{R}^2$.



Level lines of $\|Ax - d\|_2^2$ with ℓ_2 constraint $\|x\|_2 \leq K \ \rightarrow \ (x_1, x_2) \neq (0, 0)$.

ℓ_1 norm promotes sparsity

Standard example in $\mathbb{R}^2$.



Level lines of $\|Ax - d\|_2^2$ with ℓ_1 constraint $\|x\|_1 \leq K \ \rightarrow \ x_1 = 0$.

Sparsity through sof-thresholding

Recall that in 1D:

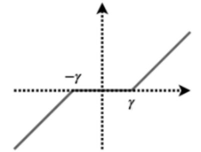
$$\hat{x} = \arg\min_{x\in\mathbb{R}} \left\{ \frac{1}{2}(d-x)^2 + \lambda|x| \right\} = \mathrm{prox}_{\lambda|\cdot|}(d)$$

is reached at

$$\hat{x} = \mathcal{T}_\lambda(d) = \begin{cases} d - \mathrm{sign}(d)\lambda & \text{if } |d| > \lambda \\ 0 & \text{if } |d| \leq \lambda \end{cases}$$

By, separability, this is then used for defining $\mathrm{prox}_{\lambda\|\cdot\|_1}(\cdot)$.

. . . many zeros!



Note: using ℓ_2 norm we get instead

$$\hat{x} = \arg\min_{x\in\mathbb{R}} \left\{ \frac{1}{2}(d-x)^2 + \lambda x^2 \right\}.$$

$\hat{x} = \frac{d}{1+2\lambda}$ which is different from 0 as soon as $d \neq 0$.

Algorithmic advantages in solving ℓ_2-ℓ_1 problems

You now know how to solve the problem:

$$\arg\min_x \frac{1}{2}\|Ax - d\|^2 + \lambda\|x\|_1, \qquad \lambda > 0$$

- ISTA (Combettes, Wajs, '05)
- FISTA (Beck, Teboulle, '09)
- If A is positive definite $\rightarrow$ strongly convex problem, hence V-FISTA can be used (Beck, '17)

For analysis approaches, i.e. when sparsity is assumed w.r.t. to some basis $W \in \mathbb{R}^{N\times n}$ (gradient, wavelets. . .)

$$\arg\min_x \frac{1}{2}\|Ax - d\|^2 + \lambda\|Wx\|_1, \qquad \lambda > 0$$

you can use, e.g., ADMM (Glowinski, Marroco, '75, Boyd et al, '11).

So... why just not solving ℓ_2-ℓ_1?

Compressed Sensing Theory

- A sparse signal ($\|x\|_0 \leq K$) can be exactly reconstructed by solving the constrained ℓ_1 problem when Restricted Isometry Property (RIP) of matrix A (Donoho et al., Candès et al. '06)
- Roughly speaking A satisfies the RIP if $A^T A \approx Id$.

- Under RIP conditions on A, ℓ_0 can be replaced by ℓ_1.
- Otherwise (frequent cases in inverse problems), the two optimisation problems give different solutions.
- ℓ_1 promotes sparsity but introduces biases, since in correspondence of the actual non-zeros the magnitude is lowered.
- ℓ_0 better promotes sparsity than ℓ_1 in the general case.

Algorithms for ℓ_2-ℓ_0 minimisation

Algorithms for ℓ_2-ℓ_0 minimisation

Iterative Hard Thresholding

Non-convex proximal gradient: iterative hard thresholding

Consider the penalised form of the problem:

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

- $\frac{1}{2}\|Ax-d\|^2$ is L-smooth ($L = \|A\|^2$)
- The proximal operator of $\|\cdot\|_0$ is the hard thresholding operator

Algorithm: Iterative hard thresholding (IHT)

Input: $x_0 \in \mathbb{R}^n$, $\tau \in \left(0, \frac{1}{L}\right)$.
for $k \geq 0$ **do**

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{\tau\lambda\|\cdot\|_0}\left(x_k - \tau A^T(Ax_k - d)\right) \\ &= \mathcal{H}_{\sqrt{2\lambda\tau}}\left(x_k - \tau A^T(Ax_k - d)\right) \end{aligned}$$

end for

- IHT converges to a critical point (in Blumensath, Davies, '09 for $\tau = 1$ and $\|A\| < 1$, in Attouch et al., '13 general FB-type result)
- As always for non convex problems, **initialisation is crucial**! One good idea is to initialise with the solution of

$$\arg\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-d\|_2^2 + \lambda\|x\|_1 \qquad \rightarrow \text{computed by FISTA}$$

IHT: ideas

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

Introduce the surrogate function for all $z \in \mathbb{R}^n$:

$$C^S_{\ell_0}(x,z) := \frac{1}{2}\|Ax-d\|_2^2 + \lambda\|x\|_0 - \frac{1}{2}\|Ax-Az\|_2^2 + \|x-z\|_2^2$$

It can be shown that if $\|A\| < 1$, then $C^S_{\ell_0}(x,z)$ majorises $G_{\ell_0}(x)$:

$$G_{\ell_0}(x) \leq C^S_{\ell_0}(x,z), \quad \forall z \in \mathbb{R}^n.$$

Note, moreover, that $G_{\ell_0}(x) = C^S_{\ell_0}(x,x)$. We can thus **optimise** $C^S_{\ell_0}(x,z)$ **with respect to** x. We can rewrite:

$$C^S_{\ell_0}(x,z) = \frac{1}{2}\sum_{i=1}^n \left(x_i^2 - 2x_i\left(z_i + a_i^T d - a_i^T Az\right) + \lambda|x_i|_0\right) + \frac{1}{2}\left(\|d\|^2 + \|z\|^2 - \|Az\|^2\right)$$

By treating the case $x_i = 0$ and $x_i \neq 0$ separately and comparing we get:

$$x = \mathcal{H}_{\sqrt{2\lambda}}(z - A^T(Az-d)), \quad \forall z$$

IHT obtained by setting $z = x_k$ and $x = x_{k+1}$.

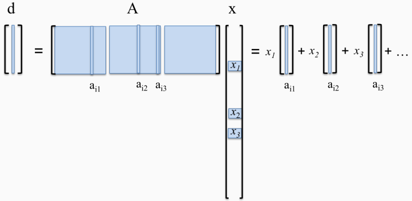
Algorithms for ℓ_2-ℓ_0 minimisation

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* (Herzen et al., '10), *Single Best Replacement* (Soussen et al, '11).

Matching Pursuit

$d \in \mathbb{R}^m$ is the signal to represent with a limited number of $K \ll n$ of **atoms** of dictionary $A \in \mathbb{R}^{m \times n}$, i.e. of columns a_i of A, $i = 1, \ldots, n$.



MP considers the constrained formulation:

$$\underset{x \in \mathbb{R}^n}{\arg\min} \; \|Ax - d\|^2, \quad \text{subject to} \quad \|x\|_0 \leq K$$

and try to add **one component at a time**.

Matching pursuit: main ideas

Assumption: A has unit column norms, i.e. $\|a_i\| = 1$ for all $i = 1, \ldots, n$.

Algorithm: Matching pursuit

Input: A s.t. $\|a_i\| = 1$, d, $K \ll n$.
Initialise: $r_0 = d$, $\sigma_0 = \emptyset$, $x_0 = 0$.
while $\#\sigma_k \leq K$ **do**

$$i_k = \operatorname*{arg\,max}_{j \in \{1,\ldots,n\}} |\langle r_k, a_j \rangle|$$

$$\sigma_{k+1} = \sigma_k \cup \{i_k\}$$

$$x_{k+1} = x_n + \langle a_{i_k}, r_k \rangle e_{i_k}$$

$$r_{k+1} = r_k - \langle r_k, a_{i_k} \rangle a_{i_k}$$

end while

- The quantity $\|r_k\|$ converges exponentially to 0 (Mallat et al, '93)
- In Gribonval et al., '96, a different correlation function (not $|\langle \cdot, \cdot \rangle|$) is considered.

Orthogonal Matching Pursuit

OMP idea (Pati et al. '93, Tropp, '04): at each iteration of MP optimally estimate the intensity values having **the current support fixed** by solving

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \|Ax - d\|^2, \quad \text{subject to } x_i = 0 \; \forall i \notin \omega := \sigma(x_k) \cup i_{k+1}$$

Algorithm: Orthogonal matching pursuit

Input: A s.t. $\|a_i\| = 1$, d, $K \ll n$.
Initialise: $r_0 = d$, $\sigma_0 = \emptyset$, $x_0 = 0$.
while $\#\sigma_k \leq K$ **do**

$$i_k = \arg\max_{j \in \{1,\ldots,n\}} |\langle r_k, a_j \rangle|$$

$$\sigma_{k+1} = \sigma_k \cup \{i_k\}$$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \|Ax - d\|^2, \quad \text{subject to } x_i = 0 \; \forall i \notin \sigma(x_{k+1})$$

$$r_{k+1} = d - Ax_{k+1}$$

end while

- "Orthogonal" as by definition at each $k \geq 0$ the residual belongs to the orthogonal space of the current support
- Convergence in n iterations at most (new component at each iteration)
- Exact sparse recovery results (under some conditions on A) (Tropp, '04)

Further greedy algorithms

The main idea of the other existing greedy algorithms is that at each iteration one component is:

- added
- removed
- replaced

The more complex is the strategy, the best is the solution, but the largest is the computing time. . .

Continuous relaxations

Continuous relaxation idea

Think of a different idea for solving the problem:

$$\frac{1}{2}\|Ax-d\|^2+\lambda\|x\|_0 \quad \Longrightarrow \quad \frac{1}{2}\|Ax-d\|^2+\sum_{i=1}^{n}\phi_i(x_i)$$

Idea: use continuous and separable functions $\phi_i(x_i)$ (convex and non-convex).

- ℓ_1 norm: LASSO (Tibshirani, '96), Basis Pursuit (Chen, '98), Compressed Sensing (Donoho, '06, Candès et al., '06)
- Adaptive LASSO (Zou, '06)
- Exponential approximation (Mangasarian, '96)
- Log-sum penalty (Candès, '08)
- Smoothly Clipped Absolute Deviation (SCAD) (Fan, Liu, '01) and Minimax Concave Penalty (MCP) (Zhang, '10
- ℓ_p "norms", $p<1$ (Chartrand, '07, Foucart, Lai, '09)
- **Beautiful review** (Soubies, Blanc-Féraud, Aubert, '17)

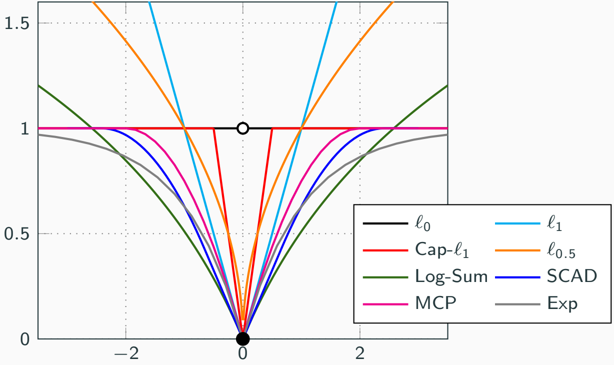
Which approximation should we use?

Continuous relaxation idea

Think of a different idea for solving the problem:

$$\frac{1}{2}\|Ax - d\|^2 + \lambda\|x\|_0 \quad \Longrightarrow \quad \frac{1}{2}\|Ax - d\|^2 + \sum_{i=1}^{n} \phi_i(x_i)$$

Idea: use continuous and separable functions $\phi_i(x_i)$ (convex and non-convex).



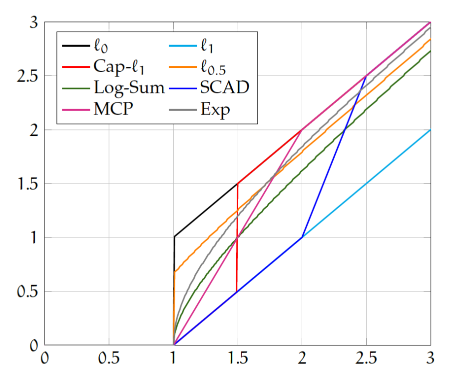
Which approximation should we use?

Continuous relaxation idea

Think of a different idea for solving the problem:

$$\frac{1}{2}\|Ax - d\|^2 + \lambda\|x\|_0 \quad \Longrightarrow \quad \frac{1}{2}\|Ax - d\|^2 + \sum_{i=1}^{n} \phi_i(x_i)$$

Idea: use continuous and separable functions $\phi_i(x_i)$ (convex and non-convex).



Thresholding on $\mathbb{R}_+$

Which approximation should we use?

Continuous relaxations

Exactness

What is a good relaxation?

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

Good (exact) relaxation

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

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

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

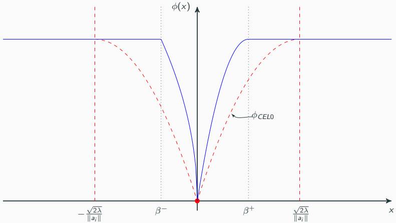
$$x^* \text{ minimiser of } \tilde{G} \Rightarrow x^* \text{ minimiser of } G_{\ell_0} \tag{P2}$$

The continuous exact ℓ_0 relaxation (CEL0) penalty

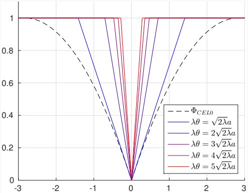
In Soubies, Aubert, Blanc-Féraud, '15-'17 a particular choice of $\phi : \mathbb{R} \to \mathbb{R}_+$ is studied. By convex conjugation, the penalty **removing most of the local minimisers** is:

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

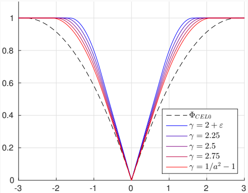
where $\mathbf{1}_C(x) = 1$ if $x \in C$ and $\mathbf{1}_C(x) = 0$ otherwise.



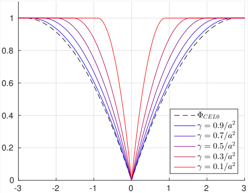
Good relaxations: examples



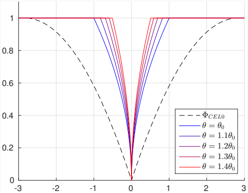
Capped-ℓ_1, Zhang, '09



SCAD, Fan, Li, '01



MCP, Zhang, '01



Truncated-ℓ_p

Examples of penalties for which (P1) (top) or (P1) and (P2) (bottom) hold for $a = 0.5$, $\lambda = 1$ and $d = 1.8$ in the 1D case.

The CEL0 relaxation

$$G_{CEL0}(x) := \frac{1}{2}\|Ax-d\|^2 + \underbrace{\sum_{i=1}^{n} \phi_{CEL0}(\|a_i\|,\lambda,x_i)}_{\Phi_{CEL0}:=}$$

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

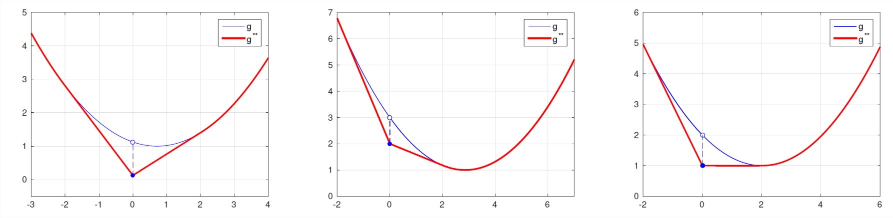
Properties of G_{CEL0}:

- Inferior limit of all functions satisfying (P1) and (P2)
- Convex envelope of G_{ℓ_0} if A diagonal or $A^T A = s\text{Id}$, $s > 0$
- Continuous
- Non convex for general operators A
- Convexity w.r.t. each component x_i, $i = 1, \dots, n$

Thanks to its continuity we can resort to *nonsmooth, nonconvex* algorithms such as, e.g., forward-backward and *majorisation-minimisation* (MM) algorithms (e.g., iterative reweighted ℓ_1 Ochs et al., '15).

Understanding the relaxation

1D example: $G_{\ell_0}(x) := \frac{1}{2}(ax - y)^2 + \lambda|x|_0$ for $a, \lambda > 0$.



Blue lines: plots of G_{ℓ_0} for different values of d (note discontinuity in $x = 0$). Red lines: plots of G_{CEL0} (convex biconjugate).

In 1D G_{CEL0} is **always** a convex function, in the multi-dimensional case it depends on the operator A. Generally, it is non-convex with convex 1D restrictions.

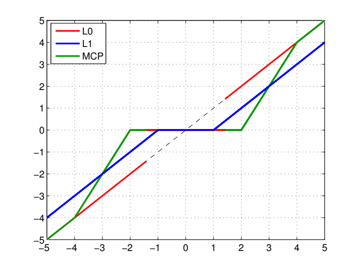
Forward-backward splitting for ℓ_2-$CEL0$

Iterate for $k \geq 0$ and $\tau \in (0, \frac{1}{\|A\|^2})$

$$x_{k+1} \in \text{prox}_{\tau \Phi_{CEL0}} \left(x_k - \tau A^T (Ax_k - d) \right)$$

where, by separability, we can look at the prox of the 1D components:

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



Dependence of ϕ_{CEL0} on $a = \|a_i\|$ at component $u = x_i$.

Convergence to a critical point under Kurdyka-Łojaseiwicz (KL) property (Attouch et al, '13).

Continuous relaxations

Iteratively reweighted algorithms

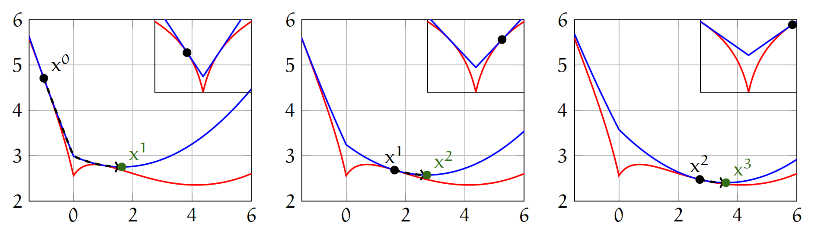
Key idea

$$\min_{x \in \mathbb{R}^n} \; F(x) := f(x) + g(x)$$

for g proper, l.s.c. and bounded from below but generally **non-convex**

Majorisation-minimisation technique

Construct **a sequence** of easier (convex) functions majorising F and minimise them to simplify the problem.



Minimisation of a non-convex function (red) using MM techniques. Non-convexity induced by $g(x) = \log(1 + 2|x|)$. Majorant functions in blue.

Key idea

$$\min_{x \in \mathbb{R}^n} \; F(x) := f(x) + g(x)$$

for g proper, l.s.c. and bounded from below but generally **non-convex**

Majorisation-minimisation technique

Construct **a sequence** of easier (convex) functions majorising F and minimise them to simplify the problem.

Pseudocode: general idea for MM algorithms

Input: $x_0 \in \mathbb{R}^n$.
while not converging **do**
Build a majorising function $M_{x_k} : \mathbb{R}^n \to \mathbb{R}$ such that:

- $\forall x \in \mathbb{R}^n$: $F(x) \leq M_{x_k}(x)$
- $F(x_k) = M_{x_k}(x_k)$
- $M_{x_k}(x_k) \in \Gamma_0(\mathbb{R}^n)$

Define $x_{k+1} \in \arg\min_x \; M_{x_k}(x)$
end while

MM approaches

Several approaches for building such functions:

- Iterative least-squares (IRLS) (Daubechies et al. '10, Gorodnitsky, Rao, '97):

$$M_{x_k}(x) = \sum (w_{x_k})_i x_i^2$$

- MM approaches for inverse problems (Chouzenoux et al., '10 -...)
- **Iterative reweighted ℓ_1 algorithms**: better suited to construct majorants of functions which are not sufficiently smooth of the form:

$$F(x) = \frac{1}{2}\|Ax - d\|^2 + \sum \phi(|x_i|)$$

with $\phi : \mathbb{R}_+ \to \mathbb{R}$ continuous, concave and non-decreasing (Ochs et al, '15.)

Algorithm: IRℓ_1 (Ochs et al, '15)

Input: $x_0 \in \mathbb{R}^n$.
while not converging **do**
  $(w_{x_k})_i \in \partial^+\phi_i(|(x_k)_i|)$
  $x_{k+1} \in \arg\min_x \ \frac{1}{2}\|Ax - d\|^2 + \sum_{i=1}^n (w_{x_k})_i |x_i| \to$ solve with FISTA
end while

$\partial^+\phi_i(|(x_k)_i|)$ extends the notion of subdifferentials to the non-convex case (Clarke, '90, Rockfellar, Wets, '09)

IRℓ_1 for G_{CEL0} minimisation

Weights can be computed in an explicit form:

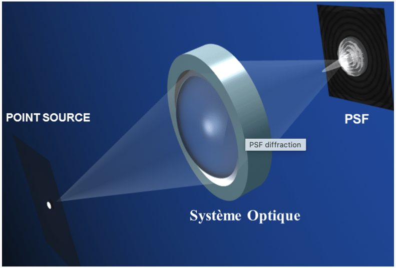
$$(w_{x_k})_i := \begin{cases} \sqrt{2\lambda}\|a_i\| - \|a_i\|^2|(x_k)_i| & \text{if } 0 \leq |(x_k)_i| < \sqrt{2\lambda}/\|a_i\| \\ 0 & \||(x_k)_i| \geq \sqrt{2\lambda}/\|a_i\| \end{cases}$$

Convergence of IRℓ_1 to critical points can be proved for general class of functions satisfying the so-called Kurdyka-Łojasiewicz property (Ochs et al, '15).

Application to super-resolution microscopy

Super-resolution microscopy

Spatial resolution is limited by light diffraction phenomena.

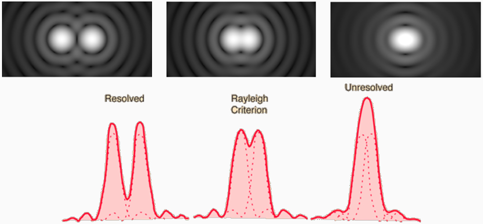


Rayleigh criterion

$$d = \frac{0.61\lambda}{NA} \approx 200nm$$

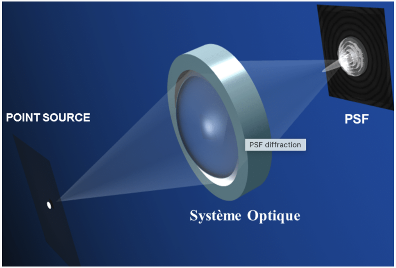
- λ: emission wavelength
- NA: microscope numerical aperture

Point Spread Function: Gaussian, Airy disk. . .



Super-resolution microscopy

Spatial resolution is limited by light diffraction phenomena.

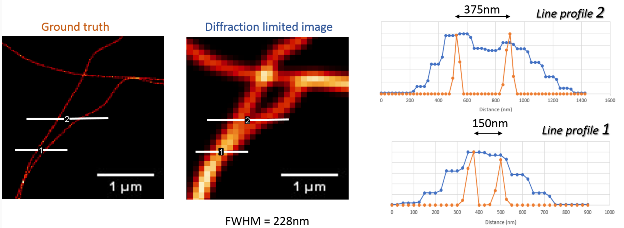


Rayleigh criterion

$$d = \frac{0.61\lambda}{NA} \approx 200nm$$

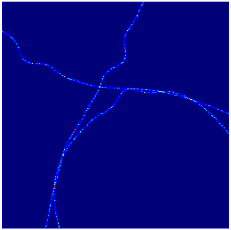
- λ: emission wavelength
- NA: microscope numerical aperture

Point Spread Function: Gaussian, Airy disk...

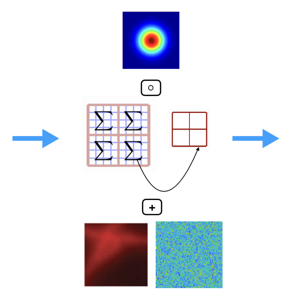


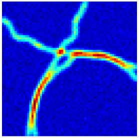
Resolvable VS. non-resolvable line profiles

Discrete mathematical modelling



$\boldsymbol{X}$



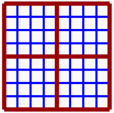


$\boldsymbol{Y}$

Image formation model

$$\boldsymbol{Y} = \mathcal{P}(M_q(H(\boldsymbol{X})) + \boldsymbol{B}) + \boldsymbol{N}$$

- $\boldsymbol{Y} \in \mathbb{R}^{N \times N}$: LR acquisition
- $\boldsymbol{X} \in \mathbb{R}^{L \times L}$: HR image ($L = qN$, $q \in \mathbb{N}$)
- $\mathcal{P}(\cdot)$: Poisson r.v.
- $M_q \in \mathbb{R}^{N \times L}$: down-sampling matrix
- $H \in \mathbb{R}^{N \times N}$: convolution matrix
- $\boldsymbol{N}$: additive white Gaussian noise
- $\boldsymbol{B}$: background



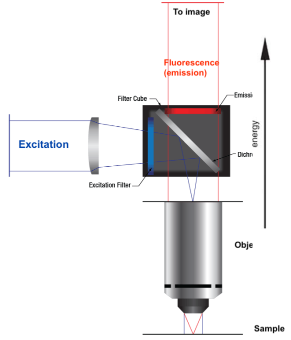
$q = 4$

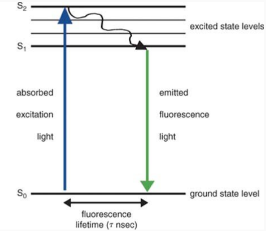
State-of-the-art methods in fluorescence microscopy

Key idea

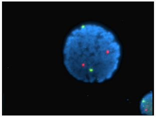
In microscopy imaging, the experimental setup and the sample preparation can be used to ‘sparsify’ the measurements.

Fluorescence microscopy





Absorption/emission diagram



Fluorescent molecules

Nobel prize in chemistry in 2008.

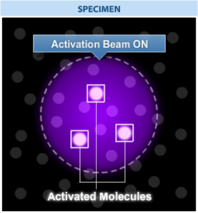
State-of-the-art methods in fluorescence microscopy

Key idea

In microscopy imaging, the experimental setup and the sample preparation can be used to ‘sparsify’ the measurements.

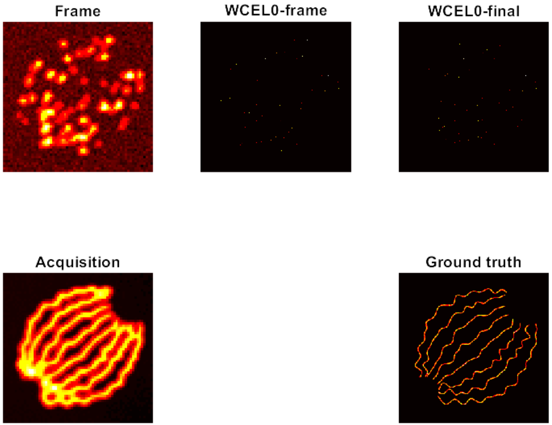
Example: **S**ingle **M**olecule **L**ocalization **M**icroscopy (Betzig, Zhuang, Hess, ’06, Rust, Bates, Zhuang, ’06)

- Specific fluorescent molecules activating with low probability in a sequential way
- Improved sparsity!



http://zeiss-campus.magnet.fsu.edu/

Spoiler



$$\boldsymbol{y}_t = \mathcal{P}(\boldsymbol{\Psi}\boldsymbol{x}_t + \boldsymbol{b}) + \boldsymbol{n}_t, \quad \boldsymbol{\Psi} := \boldsymbol{M}_q\boldsymbol{H}, \quad \boldsymbol{n}_t \sim \mathcal{N}(0, \sigma^2\boldsymbol{Id}), \quad \bar{\boldsymbol{y}} := \sum_{t=1}^{T} \boldsymbol{y}_t / T$$

Weighted CEL0

To incorporate signal-dependence (modelling Poisson photon counting) in Lazzaretti, Calatroni, Estatico, '21 we considered a weighted ℓ_2 fidelity term.

Weighted-ℓ_2-ℓ_0 problem

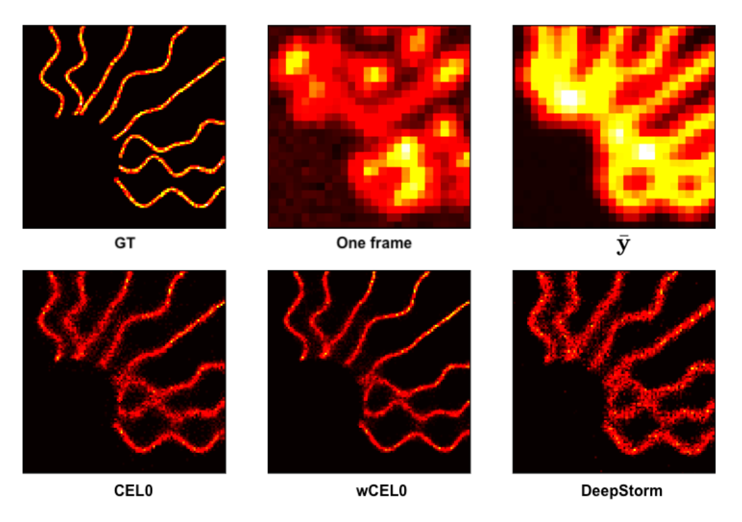
$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^{L^2}} \left\{ G_{w\ell_0}(\boldsymbol{x}) := \frac{1}{2}\sum_{j=1}^{N^2} \frac{((\boldsymbol{\Psi x})_j - y_j - b_j)^2}{y_j + b_j} + \lambda\|\boldsymbol{x}\|_0 + \iota_{\geq 0}(\boldsymbol{x}) \right\}, \quad \lambda > 0$$

Theorem

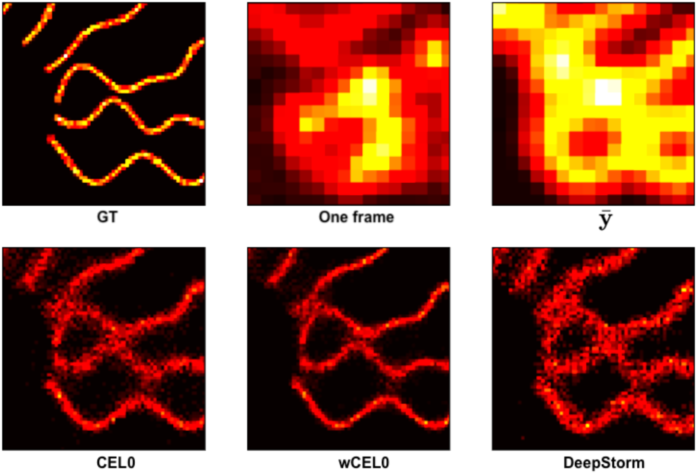
- If $\boldsymbol{\Psi}^T \boldsymbol{W} \boldsymbol{\Psi} = \boldsymbol{D}^2$ with $\boldsymbol{D} = \operatorname{diag}(\|\psi_{\boldsymbol{i}}\|_W) \in \mathbb{R}^{L^2 \times L^2}$, then $G_{\mathtt{wCEL0}} = G^{**}_{\mathtt{w}\ell_0}$.
- $\arg\min G_{\mathtt{wCEL0}} = \arg\min G_{w\ell_0}$ (same global minimisers)
- $\boldsymbol{x}$ minimiser of $G_{\mathtt{wCEL0}} \Rightarrow \boldsymbol{x}$ minimiser of $G_{w\ell_0}$ (less local minimisers).

+ Minimisation with IRℓ_1.

Zoom on a detail



Zoom on a detail



Conclusions

We focused on models and algorithms tackling the ℓ_2-ℓ_0 minimisation problem.

- NP-hardness is avoided by alternative formulations
- Greedy approaches provide interesting results, at the price of increased complexity
- **Continuous relaxations** (both convex and non-convex) ease the problem
- CEL0 is the "best" (liminf) continuous, non-convex relaxation, and it is exact.
- A MM strategy such as IRℓ_1 can be used. Fast convex optimisation is here essential for solving inner problems with high precision.
- Application areas are vast: inverse problems in imaging, vision, variable selection in machine learning. . .

Interested in a PostDoc (or PhD) in optimisation?

AGENCE NATIONALE DE LA RECHERCHE

ANR

Task-adaptive bilevel learning of flexible statistical models for imaging and vision
(2023-2027)

- 2-year post-doctoral position (open)
- 1 PhD position (from October 2023)

Announcement II: SSVM 2023

- **What?** IX conference on Scale Space and Variational Methods in Computer Vision (SSVM).
- **Where?** Hotel Flamingo, Santa Margherita di Pula, Sardegna, IT.
- **When?** May 21-25 2023
- **Who?** Giunta Gruppo UMI MIVA + G. Rodriguez (local organiser)
- **Why** Oral + poster session of selected papers (published in Springer LNCS)

Website: SSVM 2023

NEW DEADLINE for submissions: January 30 2023

Questions?

calatroni@i3s.unice.fr