Large scale distributed optimization

Lab Session 2

5th November 2020

Emilie CHOUZENOUX / Jean-Baptiste FEST

jean-baptiste.fest@centralesupelec.fr

Centrale-Supelec,Center for Visual Computing

université PARIS-SACLAY

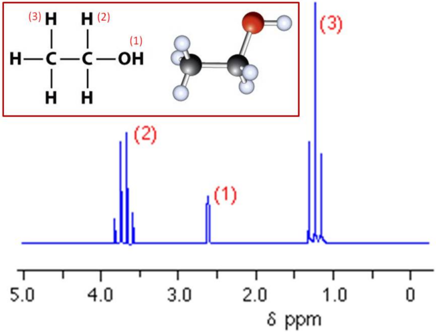
CentraleSupélec

Few words about the NMR process

- **NMR (Nuclear magnetic Resonance) spectroscopy :** a technique based on magnetic properties of few atomic nucleus. Use in many fields to find the component of complex chemical substances.
- **General principle :** disrupting the molecule by applying a magnetic field and then analysing the atomic process until the reach of an equilibrium state.

Few words about the NMR process

- **NMR (Nuclear magnetic Resonance) spectroscopy :** a technique based on magnetic properties of few atomic nucleus. Use in many fields to find the component of complex chemical substances.
- **General principle :** disrupting the molecule by applying a magnetic field and then analysing the atomic process until the reach of an equilibrium state.



A simple NMR example : the ethanol's spectrum (source : SDSB)

Around DOSY experiment

- **DOSY (Diffusion Ordered SpectroscopY) :** proposed by Morris & Johnson in 1992 is a particular RMN technique. Focus on diffusion coefficients to make the separation between each molecule's component.
- **Concretely :** Consists to make M experiments for which a (gradient) pulse field is applied and measured at a certain acquisition time.
- **Mathematical model:** The pulse intensities $y = \left(y^{(m)}\right)_{1 \leq m \leq M}$ and the acquisition time $t = \left(t^{(m)}\right)_{1 \leq m \leq M}$ are linked by a Laplace transform through the diffusion T :

$$\forall m \quad y^{(m)} = \int \chi(T) \exp(-t^{(m)} T) \mathrm{d} T$$

Around DOSY experiment

- **DOSY (Diffusion Ordered SpectroscopY) :** proposed by Morris & Johnson in 1992 is a particular RMN technique. Focus on diffusion coefficients to make the separation between each molecule's component.
- **Concretely :** Consists to make M experiments for which a (gradient) pulse field is applied and measured at a certain acquisition time.
- **Mathematical model:** The pulse intensities $y = \left(y^{(m)}\right)_{1 \leq m \leq M}$ and the acquisition time $t = \left(t^{(m)}\right)_{1 \leq m \leq M}$ are linked by a Laplace transform through the diffusion T :

$$\forall m \quad y^{(m)} = \int \chi(T) \exp(-t^{(m)} T) \mathrm{d}T$$

Where χ is the diffusion distribution to be estimated.

Around DOSY experiment

$$\boxed{\forall m \quad y^{(m)} = \int \chi(T)\exp(-t^{(m)}T)\mathrm{d}T.}$$

How to move from a continuous form to a discrete one ?

Around DOSY experiment

$$\boxed{\forall m \quad y^{(m)} = \int \chi(T)\exp(-t^{(m)}T)\mathrm{d}T.}$$

How to move from a continuous form to a discrete one ?

Answer : Using a quadrature method for the integral by considering a grid diffusion of length M : $T = \left(T^{(n)}\right)_{1\leq n\leq N}$. Then :

$$\forall m \quad y^{(m)} = \int \chi(T)\exp\left(-t^{(m)}T\right)\mathrm{d}T \simeq \sum_{n=1}^{N} \underbrace{\Delta T^{(m)} \ \chi(T^{(m)})}_{x^{(n)}} \underbrace{\exp\left(-t^{(m)}T_n\right)}_{K_{m,n}}$$

Under a matrix form : $y \simeq Kx$ with $x = \left(x^{(n)}\right)_{1\leq n\leq N}$.

Considering a general additive noise $w \in \mathbb{R}^N$, this finally leads to the inverse problem :

$$\boxed{y = Kx + w}$$

General Optimization problem

Let $y \in \mathbb{R}^N$ the noisy DOSY NMR data. Our goal is to find a $\widehat{x} \in \mathbb{R}^N$ s.t

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \quad F(x) = \frac{1}{2}\|Kx - y\|^2 + \beta g(x)$$

- $K \in \mathbb{R}^{N \times N}$: discretization matrix
- g : penalization's function / β : regularization parameter.

General Optimization problem

Let $y \in \mathbb{R}^N$ the noisy DOSY NMR data. Our goal is to find a $\hat{x} \in \mathbb{R}^N$ s.t

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \quad F(x) = \frac{1}{2}\|Kx - y\|^2 + \beta g(x)$$

- $K \in \mathbb{R}^{N \times N}$: discretization matrix
- g : penalization's function / β : regularization parameter.

Through this lab session :

- Few penalization functions g will be taken into consideration by order of complexity.
- Suggest preliminary theoretical approaches to find the most appropriate algorithm in every cases.
- Performances analysis of every penalization's strategy.

General Optimization problem

Let $y \in \mathbb{R}^N$ the noisy DOSY NMR data. Our goal is to find a $\hat{x} \in \mathbb{R}^N$ s.t

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \quad F(x) = \frac{1}{2}\|Kx - y\|^2 + \beta g(x)$$

- $K \in \mathbb{R}^{N \times N}$: discretization matrix
- g : penalization's function / β : regularization parameter.

Through this lab session :

- Few penalization functions g will be taken into consideration by order of complexity.
- Suggest preliminary theoretical approaches to find the most appropriate algorithm in every cases.
- Performances analysis of every penalization's strategy.

Optimization problem with smoothness prior

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \| Kx - y \|^2 + \beta \; \frac{1}{2} \| Dx \|^2$$

Where $D \in \mathbb{R}^{N \times N}$: Matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$. and the convention $x^{(0)} = x^{(n)}$.

Optimization problem with smoothness prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Kx - y\|^2 + \beta \, \frac{1}{2} \|Dx\|^2$$

Where $D \in \mathbb{R}^{N \times N}$: Matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$. and the convention $x^{(0)} = x^{(n)}$.

Here the easiest way : calculate the gradient and the hessian of F

$$\forall x \in \mathbb{R}^N \qquad \begin{aligned} \nabla F(x) &= K^\top K \; x - K^\top y + 2\beta \; D^\top D \; x, \\ \nabla^2 F(x) &= K^\top K + \beta \; D^\top D. \end{aligned}$$

Moreover it's simple to see that $\nabla^2 F(x) \succeq 0$ and then F is convex it follows :

$$\widehat{x} \text{ minimize } f \iff \nabla F(\widehat{x}) = 0 \iff \left(K^\top K + \beta D^T D\right)\widehat{x} = K^\top y.$$

Optimization problem with smoothness prior

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Kx - y\|^2 + \beta \ \frac{1}{2}\|Dx\|^2$$

Where $D \in \mathbb{R}^{N\times N}$: Matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$. and the convention $x^{(0)} = x^{(n)}$.

Here the easiest way : calculate the gradient and the hessian of F

$$\forall x \in \mathbb{R}^N \qquad \nabla F(x) = K^\top K \ x - K^\top y + 2\beta \ D^\top D \ x,$$
$$\nabla^2 F(x) = K^\top K + \beta \ D^\top D.$$

Moreover it's simple to see that $\nabla^2 F(x) \succeq 0$ and then F is convex it follows :

$$\hat{x} \text{ minimize } f \iff \nabla F(\hat{x}) = 0 \iff \left(K^\top K + \beta D^T D\right)\hat{x} = K^\top y.$$

With $\left(K^\top K + \beta D^T D\right)$ invertible since :

$$\begin{aligned} v^\top \ \left(K^\top K + \beta D^T D\right) \ v = 0 &\iff v^\top K^\top K v = \|Kv\|^2 = 0 \text{ and } v^\top D^\top D v = \|Dv\|^2 = 0, \\ &\iff Kv = 0 \text{ and } Dv = 0, \\ &\iff Kv = 0 \text{ and } v \in Vect\left((1,\dots,1)\right), \\ &\iff \boxed{v = 0} \ \left(\text{with } K_{m,n} = \exp(-T^{(n)} T^{(m)}) > 0 \ \right). \end{aligned}$$

Optimization problem with smoothness prior

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Kx - y\|^2 + \beta\,\frac{1}{2}\|Dx\|^2$$

Where $D \in \mathbb{R}^{N\times N}$: Matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$. and the convention $x^{(0)} = x^{(n)}$.

Conclusion :

$$\hat{x} \text{ minimize } f \iff \boxed{\hat{x} = \left(K^\top K + \beta D^T D\right)^{-1} K^\top y.}$$

Optimization problem with smoothness prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \| Kx - y \|^2 + \beta \, \frac{1}{2} \| Dx \|^2$$

Where $D \in \mathbb{R}^{N \times N}$: Matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$. and the convention $x^{(0)} = x^{(n)}$.

Conclusion :

$$\widehat{x} \text{ minimize } f \iff \boxed{\widehat{x} = \left(K^\top K + \beta D^T D\right)^{-1} K^\top y.}$$

Implementation strategy :

Not necessary a specific algorithm here. Since the dimension of the problem are not so high, the most simple consists of directly inversing the Hessian.

Optimization problem with smoothness prior + constraints

Find $\hat{x}$ s.t :

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \|Kx - y\|^2 + \beta \left(\frac{1}{2} \|Dx\|^2 + \iota_C(x) \right)$$

- $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \;\; / \;\; x_{min} \leq \; x^{(i)} \; \leq x_{max} \quad \forall i \right\}$.

universite PARIS-SACLAY

CentraleSupélec

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \| Kx - y \|^2 + \beta \left(\frac{1}{2} \| Dx \|^2 + \iota_C(x) \right)$$

- $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \;\; / \;\; x_{min} \leq \; x^{(i)} \; \leq x_{max} \quad \forall i \right\}$.

Rewriting as a constraints problem :

$$\widehat{x} = \arg\min_{x \in C} \frac{1}{2} \| Kx - y \|^2 + \beta \; \frac{1}{2} \| Dx \|^2 \tag{1}$$

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Kx - y\|^2 + \beta \left(\frac{1}{2}\|Dx\|^2 + \iota_C(x) \right)$$

- $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \ / \ x_{min} \leq x^{(i)} \leq x_{max} \quad \forall i \right\}$.

Rewriting as a constraints problem :

$$\widehat{x} = \arg\min_{x \in C} \frac{1}{2}\|Kx - y\|^2 + \beta \, \frac{1}{2}\|Dx\|^2 \tag{1}$$

- C : closed and bounded set in a finite dimension space. C is consequently compact. By continuity of (1), it follows that set of minimizer is non empty.
- C is non empty convex. Since the problem is strictly convex (same function as in the the first problem), it has at most one minimizer.

Conclusion :

One unique minimizer for the smoothness prior constrainted problem.

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \left(\frac{1}{2}\|Dx\|^2 + \iota_C(x)\right)$$

- ▷ $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- ▷ $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \;\; / \;\; x_{min} \leq \; x^{(i)} \; \leq x_{max} \quad \forall i \right\}$.

Implementation strategy :

- Since F is proper, convex and continuous : $f \in \Gamma_0(H)$.
- F ν-Lipschitzian gradient with ν the largest eigenvalue of $K^\top K + \beta D^\top D$ since :

$$\forall x, y \in C \quad \|\nabla F(x) - \nabla F(y)\| = \left\|\left(K^\top K + \beta D^\top D\right)(x-y)\right\| \leq \nu\|x-y\|.$$

- C is closed with the set of minimizers on C is non empty

université PARIS-SACLAY

CentraleSupélec

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \left(\frac{1}{2}\|Dx\|^2 + \iota_C(x) \right)$$

- ▷ $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- ▷ $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \ / \ x_{min} \leq x^{(i)} \leq x_{max} \quad \forall i \right\}$.

Implementation strategy :

- Since F is proper, convex and continuous : $f \in \Gamma_0(H)$.
- F ν-Lipschitzian gradient with ν the largest eigenvalue of $K^\top K + \beta D^\top D$ since :

$$\forall x, y \in C \quad \|\nabla F(x) - \nabla F(y)\| = \left\| \left(K^\top K + \beta D^\top D \right)(x - y) \right\| \leq \nu \|x - y\|.$$

- C is closed with the set of minimizers on C is non empty

Best candidate : Projected Gradient algorithm

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \left(\frac{1}{2}\|Dx\|^2 + \iota_C(x)\right)$$

- $D \in \mathbb{R}^{N \times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \;\; / \;\; x_{min} \leq \; x^{(i)} \; \leq x_{max} \quad \forall i \right\}$.

Best candidate : Projected Gradient algorithm

Here, we'll consider the case with a fixed-step size.

$$\forall n \quad \begin{cases} y_n = x_n - \gamma\left(K^\top K x_n - K^\top y + \beta D^\top D x_n\right) \\ x_{n+1} = x_n + \lambda\left(P_C(y_n) - x_n\right). \end{cases}$$

With the choices $\gamma \in]0, 2/\nu[$ and $\lambda \in]0, 2 - \nu\gamma/2[$.

Optimization problem with smoothness prior + constraints

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \left(\frac{1}{2}\|Dx\|^2 + \iota_C(x) \right)$$

- $D \in \mathbb{R}^{N\times N}$: matrix of differences : $\forall n \quad [Dx]_n = x^{(n)} - x^{(n-1)}$ and the convention $x^{(0)} = x^{(n)}$.
- $C = [x_{min}, x_{max}] = \left\{ x \in \mathbb{R}^N \;\; / \;\; x_{min} \leq \; x^{(i)} \; \leq x_{max} \quad \forall i \right\}$.

Best candidate : Projected Gradient algorithm

Here, we'll consider the case with a fixed-step size.

$$\forall n \quad \begin{cases} y_n = x_n - \gamma\left(K^\top K x_n - K^\top y + \beta D^\top D x_n\right) \\ x_{n+1} = x_n + \lambda\left(P_C(y_n) - x_n\right). \end{cases}$$

With the choices $\gamma \in]0, 2/\nu[$ and $\lambda \in]0, 2 - \nu\gamma/2[$.

To calculate P_C :

Since $C = [x_{min}, x_{max}]$ we simply have :

$$\forall i = 1, \ldots N \quad (P_C(y))_i = \begin{cases} x_{min} & \text{if} \quad x_i < x_{min}, \\ x_{max} & \text{if} \quad x_i > x_{max}, \\ x_i & \text{if} \quad x_{min} \leq x_i \leq x_{max}. \end{cases}$$

Optimization problem with sparsity prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \|Kx - y\|^2 + \beta \, \|x\|_1$$

Optimization problem with sparsity prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \| Kx - y \|^2 + \beta \, \|x\|_1$$

We use here general properties relative to unscontrained coercive optimization :

- Just note that :

$$\forall x \in \mathbb{R}^N \quad F(x) \geq \|x\|_1 \geq C\|x\|.$$

  Where C constant which comes from norms' equivalence in finite dimension. This proves the coercivity of F and then it guarantees the existence of at least one minimizer.

- However, lack of argument to conclude on the uniqueness...

Optimization problem with sparsity prior

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Kx - y\|^2 + \beta \, \|x\|_1$$

We use here general properties relative to unscontrained coercive optimization :

- Just note that :

$$\forall x \in \mathbb{R}^N \quad F(x) \geq \|x\|_1 \geq C\|x\|.$$

  Where C constant which comes from norms' equivalence in finite dimension. This proves the coercivity of F and then it guarantees the existence of at least one minimizer.

- However, lack of argument to conclude on the uniqueness...

Conclusion : At least one minimizer for the sparsity's problem .

Optimization problem with sparsity prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \|x\|_1$$

Implementation strategy :

We can write $F = F_1 + F_2$ with : $F_1 : x \in \mathbb{R}^N \longmapsto \frac{1}{2}\|Kx - y\|^2$ and $F_2 = \beta\|.\|_1$

- Since F_1, F_2 are proper continuous and convex (easy to check) then $F \in \Gamma_0(\mathscr{H})$.
- F_1 is ν-Lipschitzian gradient with ν the largest eigenvalue of $K^\top K$ since :

$$\forall x, y \in C \quad \|\nabla F_1(x) - \nabla F_1(y)\| = \left\|\left(K^\top K\right)(x - y)\right\| \leq \nu\|x - y\|.$$

- F admits at least one minimizer.

Optimization problem with sparsity prior

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Kx - y\|^2 + \beta \|x\|_1$$

Implementation strategy :

We can write $F = F_1 + F_2$ with : $F_1 : x \in \mathbb{R}^N \longmapsto \frac{1}{2}\|Kx - y\|^2$ and $F_2 = \beta \|.\|_1$

- Since F_1, F_2 are proper continuous and convex (easy to check) then $F \in \Gamma_0(\mathcal{H})$.
- F_1 is ν-Lipschitzian gradient with ν the largest eigenvalue of $K^\top K$ since :

$$\forall x, y \in C \quad \|\nabla F_1(x) - \nabla F_1(y)\| = \left\| \left(K^\top K\right)(x - y) \right\| \leq \nu \|x - y\|.$$

- F admits at least one minimizer.

Best candidate : Forward-Backward algorithm

Optimization problem with sparsity prior

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Kx - y\|^2 + \beta \, \|x\|_1$$

Best candidate : Forward-Backward algorithm

Here, we'll consider the case with a fixed-step size.

$$\forall n \quad \begin{cases} y_n = x_n - \gamma K^\top (Kx_n - y), \\ x_{n+1} = x_n + \lambda \left(prox_{\gamma\beta\|.\|_1}(y_n) - x_n \right). \end{cases}$$

With the choices $\gamma \in]0, 2/\nu[$ and $\lambda \in]0, 2 - \nu\gamma/2[$.

Optimization problem with sparsity prior

Find $\hat{x}$ s.t :

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx - y\|^2 + \beta \|x\|_1$$

Best candidate : Forward-Backward algorithm

Here, we'll consider the case with a fixed-step size.

$$\forall n \quad \begin{cases} y_n = x_n - \gamma K^\top (Kx_n - y), \\ x_{n+1} = x_n + \lambda \left(prox_{\gamma\beta\|.\|_1}(y_n) - x_n \right). \end{cases}$$

With the choices $\gamma \in]0, 2/\nu[$ and $\lambda \in]0, 2 - \nu\gamma/2[$.

To calculate $prox_{\gamma\beta\|.\|_1}$ **:**

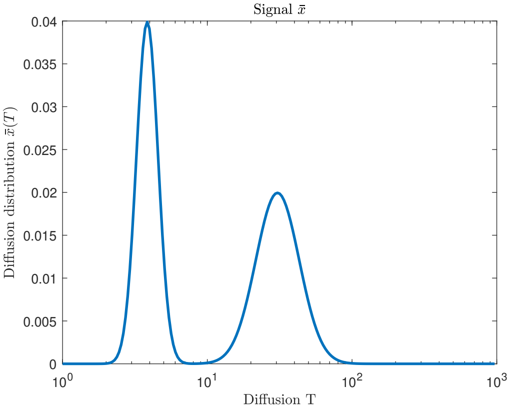
Note that $\gamma\beta\|.\|_1 = \sum_{n=1}^{N} \gamma\beta|\ .\ |$ and then following a proximal's property (slide 12 part IV) with the table (slide 8 part IV), it follows :

$$prox_{\gamma\beta\|.\|_1}(y) = \left(prox_{\gamma\beta|.|_1}(y^{(n)}) \right)_{n=1,\ldots,N} = \boxed{\left(sign(y^{(n)})\ max\left\{ \left| y^{(n)} \right| - \gamma\beta,\ 0 \right\} \right)_{n=1,\ldots,N}}.$$

Start-up support

Start-up support

You should find a similar figure to below for $\overline{x}$:



Start-up support

Noisy data's simulation :

$$\boxed{y = K\overline{x} + w}$$

With : $w \sim \mathcal{N}(0,\ \sigma^2 Id)$ and $\sigma = 0{,}01\,(K\overline{x})^{(1)}$.

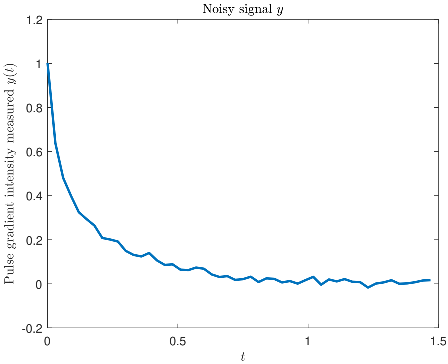
Start-up support

Noisy data's simulation :

$$\boxed{y = K\overline{x} + w}$$

With : $w \sim \mathcal{N}(0,\ \sigma^2 Id)$ and $\sigma = 0{,}01\,(K\overline{x})^{(1)}$.

You should find a similar figure to below for y :



About matrix D, few possibility of structure one which works well is to considerate D as the difference between and identity the first order circulant one :

```
[[ 1.   0.   0.  ...   0.   0.  -1.]
 [-1.   1.   0.  ...   0.   0.   0.]
 [ 0.  -1.   1.  ...   0.   0.   0.]
 ...
 [ 0.   0.   0.  ...   1.   0.   0.]
 [ 0.   0.   0.  ...  -1.   1.   0.]
 [ 0.   0.   0.  ...   0.  -1.   1.]]
```

Python :

```
np.eye   np.roll
```

Matlab :

```
toeplitz
```

Suggested structure for your solvers

For each case of penalization (except the first one with the analytic solution) :

$\hat{x}$, err = your_solver(beta)

Ini : x_0 (ex : $(x_0)_i = \frac{x_{min}+x_{max}}{2}$)
Stop_crits : iterations nb & tolerance $\frac{\|x_{n+1}-x_n\|}{\|x_n\|}$
Others para : ν, λ...

While Stop_crits == False :
apply scheme (Gradient proj, FB...)
End While

Return $\hat{x}$, $err = \frac{\|\hat{x}-\overline{x}\|}{\|\overline{x}\|}$

Suggested structure for your solvers

To find the best β parameter

- *Python* : Define a list of β and then you can use the following command

```
beta_Opt = min(List_beta, key = b :your_solver(b)[1])
```

- *Matlab* : Define a list of β and then loop uper the solver :

```
Err =[ ];
For beta in List_beta
    [~,err] = your_solver(beta);
    Err = [Err, err];
End
[~, imin] = min (Err);
beta_Opt = List_beta(imin);
```

Of course you can directly write your solver in the main script without building a specific function !

université PARIS-SACLAY

CentraleSupélec

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \|Kx - y\|^2 + \beta \; ent(x)$$

Where $ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$

Optimization problem with entropy regularization

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \| Kx - y \|^2 + \beta \ ent(x)$$

Where $ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$

First $f \in \Gamma_0\left(\mathbb{R}^N\right)$? :

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Kx-y\|^2 + \beta \; ent(x)$$

$$\text{Where} \quad ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right) \text{ and } \phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x>0 \\ 0 & \text{if } x=0 \\ +\infty & \text{otherwise} \end{array} \right.$$

First $f \in \Gamma_0\left(\mathbb{R}^N\right)$? :

- It's easy to prove that ϕ is proper and then *ent* is also proper.
- ϕ is continuous on $\mathbb{R}_+^*$ and tends positively to $\phi(0)=0$. Moreover using the fact $f=+\infty$ on $\mathbb{R}_-^*$ if follows that ϕ s.c.i. And then *ent* s.c.i (see penalization function of TP1 for a similar proof).

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Kx - y\|^2 + \beta \; ent(x)$$

$$\text{Where} \quad ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right) \text{ and } \phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } \; x > 0 \\ 0 & \text{if } \; x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$$

First $f \in \Gamma_0\left(\mathbb{R}^N\right)$? :

- It's easy to prove that ϕ is proper and then *ent* is also proper.
- ϕ is continuous on $\mathbb{R}_+^*$ and tends positively to $\phi(0) = 0$. Moreover using the fact $f = +\infty$ on $\mathbb{R}_-^*$ if follows that ϕ s.c.i. And then *ent* s.c.i (see penalization function of TP1 for a similar proof).
- By twice-derivating on $\mathbb{R}_+^*$, we clearly have the strict convexity of ϕ on this set. Then, by decomposing the different subcases you can show that ϕ is convex on $\mathbb{R}$ (similarly to exo 1 TD1 on Huber function). Finally, by positive summation it follows that *ent* is also strictly convex convex.

Optimization problem with entropy regularization

Find $\widehat{x}$ s.t :

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \| Kx - y \|^2 + \beta \; ent(x)$$

Where $ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$

Minimizers's study

université PARIS-SACLAY

CentraleSupélec

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|Kx-y\|^2 + \beta \; ent(x)$$

Where $ent :\in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x>0 \\ 0 & \text{if } x=0 \\ +\infty & \text{otherwise} \end{array} \right.$

Minimizers's study

- Since ϕ is coercive, it follows that is also coercive. By posivity of $\frac{1}{2}\|K.-y\|^2$, it follows that F coercive.

  $\Longrightarrow$ At least one minimizer
- F is strictly convex as a sum of strictly convex and $\frac{1}{2}\|K.-y\|^2$ convex.

Optimization problem with entropy regularization

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|Kx-y\|^2 + \beta\ ent(x)$$

Where $ent :\in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } x>0 \\ 0 & \text{if } x=0 \\ +\infty & \text{otherwise} \end{array} \right.$

Minimizers's study

- Since ϕ is coercive, it follows that is also coercive. By posivity of $\frac{1}{2}\|K.-y\|^2$, it follows that F coercive.

  $\Longrightarrow$ At least one minimizer
- F is strictly convex as a sum of strictly convex and $\frac{1}{2}\|K.-y\|^2$ convex.

Conclusion : F admit one unique minimizer.

Optimization problem with entropy regularization

Find $\widehat{x}$ s.t :

$$\widehat{x} = \underset{x \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \|Kx - y\|^2 + \beta \; ent(x)$$

Where $ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right)$ and $\phi : x \in \mathbb{R} \longmapsto \begin{vmatrix} x\ln(x) & \text{if } \; x > 0 \\ 0 & \text{if } \; x = 0 \\ +\infty & \text{otherwise} \end{vmatrix}$

Proximal calculus to apply a Forward-Backward algorithm :

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Kx - y\|^2 + \beta \; ent(x)$$

$$\text{Where} \quad ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right) \text{ and } \phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } \; x > 0 \\ 0 & \text{if } \; x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$$

Proximal calculus to apply a Forward-Backward algorithm :

Using the proximal property (slide 12 part IV) we have :
$prox_{\gamma ent}(x) = \left(prox_{\gamma\phi}(x^{(n)})\right)_{1 \le n}$. And then for all n :

$$p^{(n)} = prox_{\phi}(x^{(n)}) = \arg\min_{y \in \mathbb{R}} \frac{1}{2}\|y - x\|^2 - \gamma\phi(y) = \arg\min_{y \in \mathbb{R}_+} \underbrace{\frac{1}{2}(y-x)^2 + \gamma \; y ln(y)}_{\text{st convex, coercive \& diff}} \; .$$

Therefore, by anulating the gradient $p^{(n)}$ verifies : $(y - p^{(n)}) + \gamma \ln\left(p^{(n)}\right) + \gamma = 0.$

Optimization problem with entropy regularization

Find $\hat{x}$ s.t :

$$\hat{x} = \arg\min_{x \in \mathbb{R}^N} \frac{1}{2}\|Kx - y\|^2 + \beta \; ent(x)$$

$$\text{Where} \quad ent : \in \mathbb{R}^N \longmapsto \sum_{n=1}^{N} \phi\left(x^{(n)}\right) \text{ and } \phi : x \in \mathbb{R} \longmapsto \left| \begin{array}{ll} x\ln(x) & \text{if } \; x > 0 \\ 0 & \text{if } \; x = 0 \\ +\infty & \text{otherwise} \end{array} \right.$$

Proximal calculus to apply a Forward-Backward algorithm :

The trick to conclude : Pass to the exponential to have : $\frac{p^{(n)}}{\gamma} \exp\left(\frac{p^{(n)}}{2}\right) = \exp\left(\frac{x}{\gamma} - 1 - \ln(\gamma)\right)$ and then use the Lambert function W the inverse of $x \in \mathbb{R}_+ \mapsto x\ln x$ we finally have :

$$\boxed{p^{(n)} = \gamma W\left(\frac{x}{\gamma} - 1 - \ln(\gamma)\right)}.$$

Pas

General Intructions :

In your report :

- Don't forget to add title and legend to your figures ☺!
- Write some commentaries on your results. You can add additional figures to illustrate your own interpretation (evolution of error as a function of β...). Don't hesitate to be a little talkative ☺!

General Intructions :

In your report :

- Don't forget to add title and legend to your figures ☺!
- Write some commentaries on your results. You can add additional figures to illustrate your own interpretation (evolution of error as a function of β...). Don't hesitate to be a little talkative ☺!

You can send :

- A commented Jupyter notebook.
- A zip file included your code and an additional report. If you use *Matlab* you can send a publish file as a report if you're familiar with it.
- My mail address :

  jean-baptiste.fest@centralesupelec.fr

  (easiest with it compare with the one on the course's website)