

IMA203:MATIM

Méthodes variationnelles:

said.ladjal@telecom-paris.fr

Introduction des méthodes variationnelles

- Pour corriger le défaut subi par une image on utilise une énergie qui sera faible pour une image « bonne » et grande pour une image « mauvaise »
- Il y a souvent deux termes dans une telle énergie:
 - Un terme « attache aux données » qui signifie que l'image reconstruite explique l'observation.
 - Un terme « régularité » qui privilégie les images qui ont des statistiques d'images naturelles.

Un exemple de construction d'une méthode variationnelle

- Nous allons rappeler le raisonnement qui mène à la restauration par norme du gradient au carré.
- D'abord, nous allons voir pourquoi il est nécessaire de régulariser.
- Ensuite nous rappelons une justification probabiliste de la construction d'une énergie à minimiser.

Modélisation

Le processus direct qui mène à la formation de l'image est modélisé par:

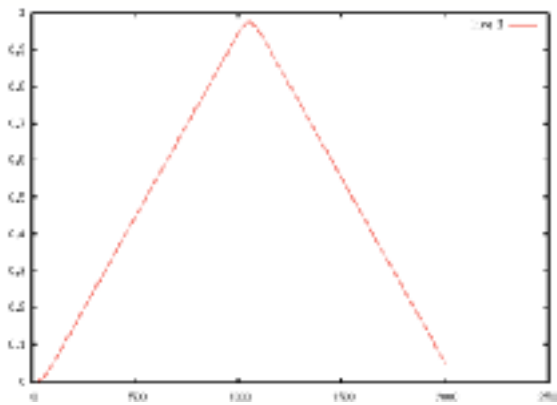
$$g = Af + b$$

L'opérateur A est linéaire et b représente le bruit.

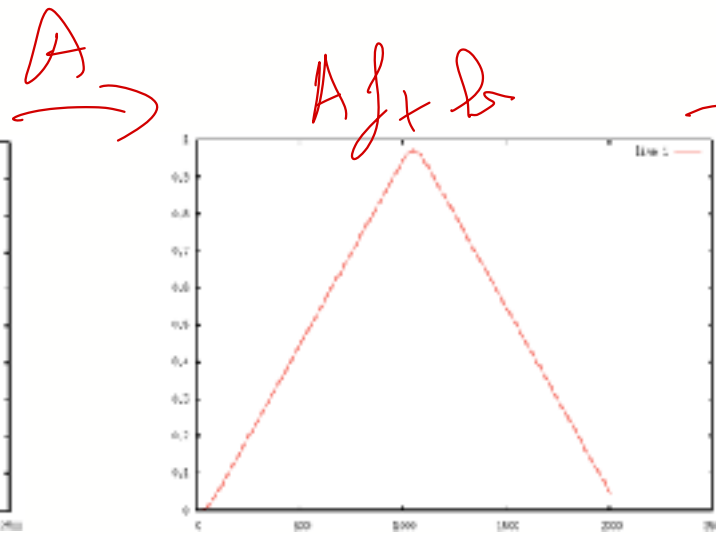
Problème inverse: Nécessité de la régularisation

- Une première idée consiste à appliquer l'opérateur inverse de A à l'équation précédente. Ici A est une petite convolution.

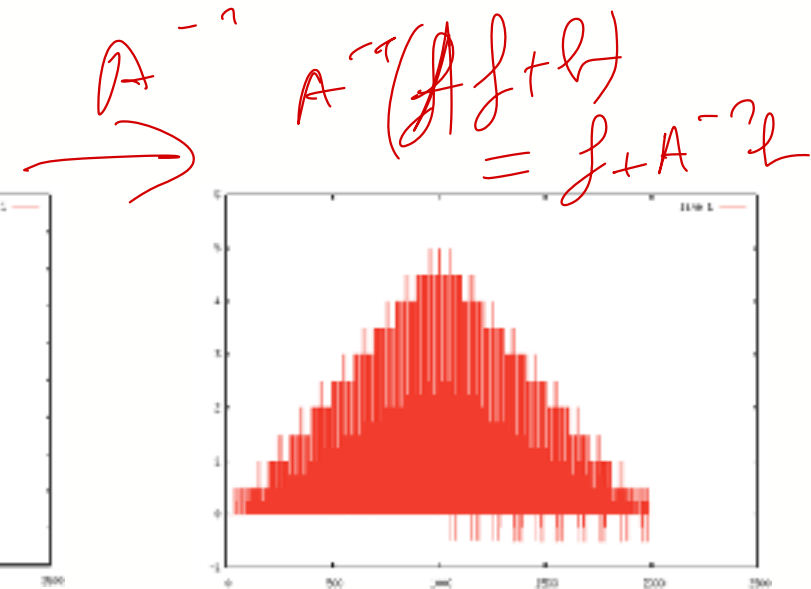
$$\tilde{f} = A^{-1}g$$



Original



Bruité



Inverse

Modélisation pour la suite

$$g = f * K + b$$

f : image parfaite

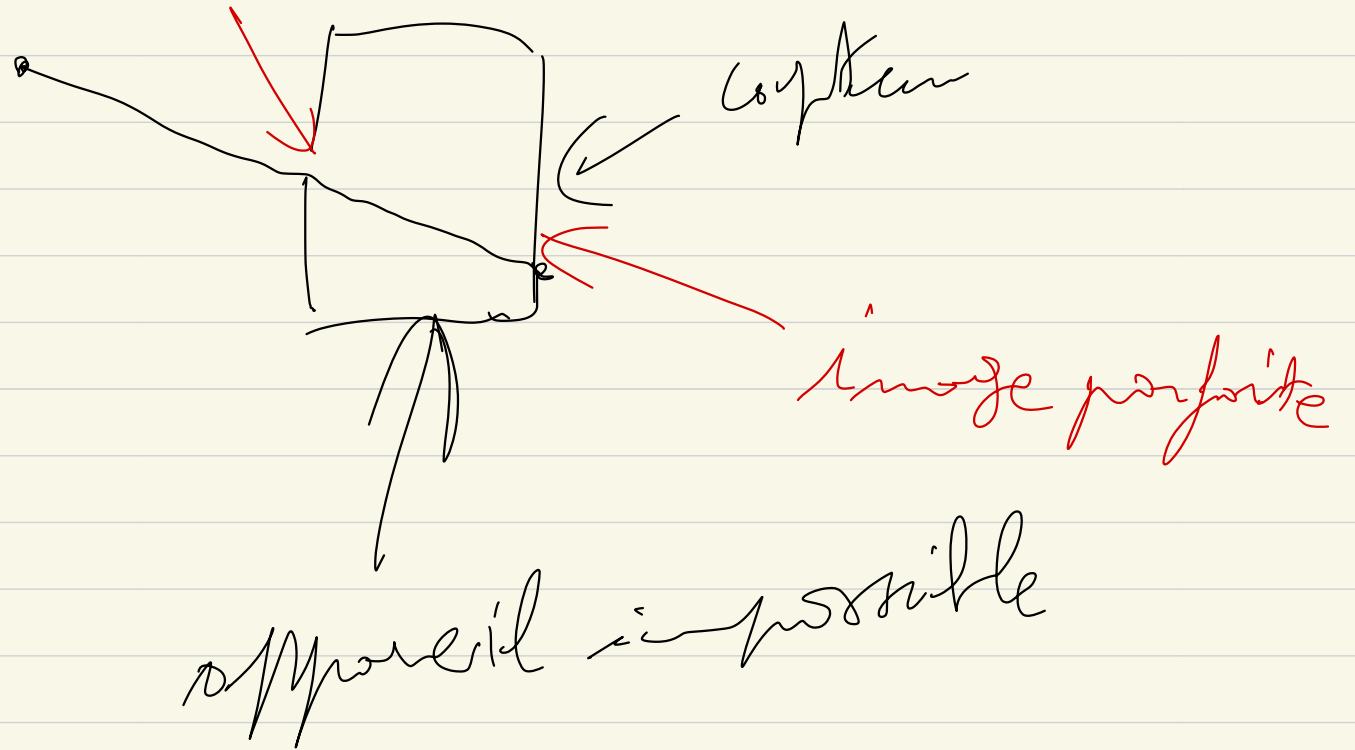
g : observation

K : Noyau de flou

b : bruit, supposé uniforme et indépendant du signal

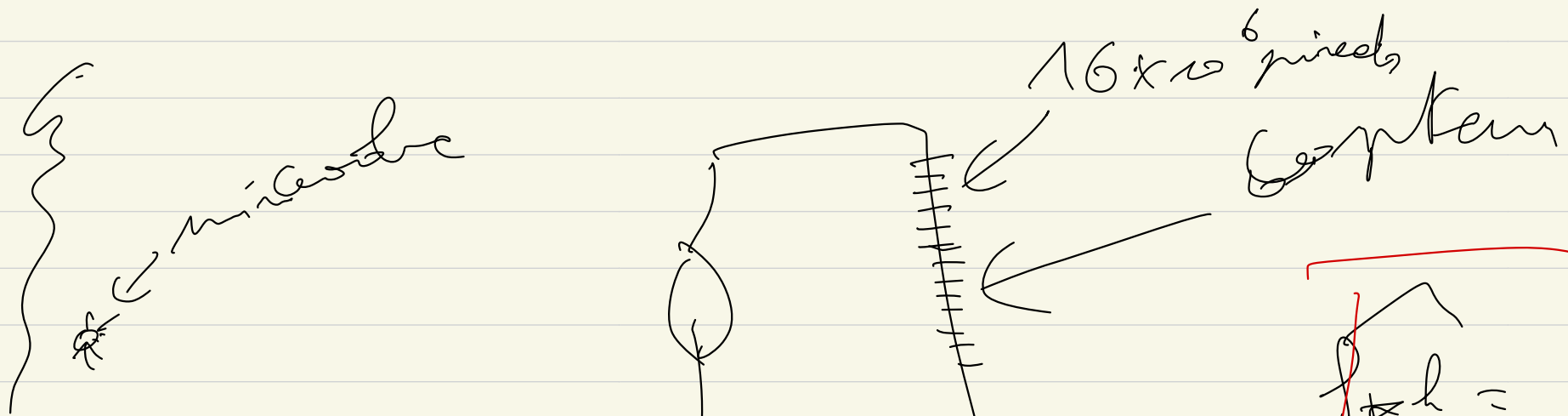
Qu'est-ce que l'image parfaite?

Kann sie Konille o.



Modélisation pour la suite: Continu/discret

- Le théorème de Shannon de bon échantillonnage permet un passage sans perte du discret au continu.
- Le caractère fini des images traitées implique de faire une hypothèse sur la continuation des images hors de leur support connu: On fera l'hypothèse de périodicité pour simplifier la présentation.



f l'image obscur

h = taille de diffraction

$$f \star h = \vec{f} \times \vec{h} = \vec{f} \times \vec{h}$$

$f \star h$ de plus $\vec{h} = \vec{1}$ garantie

Modélisation pour la suite: Continu/discret

On peut donc, en négligeant le sous-échantillonnage, dire que le problème se présente comme:

*Si $N = 1000$
 A de taille 10^{12} éléments*

$$g = Af + b$$

$f \in \mathbb{R}^{N^2}$: image de taille $N \times N$ inconnue

$g \in \mathbb{R}^{N^2}$: image de taille $N \times N$ connue

A : matrice de convolution carrée de taille $N^2 \times N^2$

$b \in \mathbb{R}^{N^2}$: réalisation d'un bruit gaussien

L'hypothèse de périodicité implique que la matrice A est circulante
et diagonalisable dans la base de Fourier

Résolution: Wiener

- Le signal est supposé être un vecteur gaussien:

$$X \text{ v.a. } \in \mathbb{R}^N$$

$$\mathbb{P}(X) \sim e^{-\frac{1}{2} X^T C^{-1} X}$$

$$C = E(X^T X) \text{ matrice de covariance déf. pos.}$$

$$\mathbb{P}(B) \sim e^{-\frac{1}{2} \frac{\|B\|^2}{\sigma_b^2}}$$

$$cov = \begin{pmatrix} Id. & 0 \\ 0 & \sigma_b^2 \end{pmatrix}$$

$$Y = AX + B$$

trouver $\tilde{X} = DY$ t.q.

$E(\|\tilde{X} - X\|^2)$ soit minimale

maximiser

i.e. minimiser

$$e^{-\frac{1}{2} \tilde{X}^T C^{-1} \tilde{X}} e^{-\frac{1}{2} \frac{\|A\tilde{X} - Y\|^2}{\sigma_b^2}}$$

$$\frac{1}{\sigma_b^2} \|A\tilde{X} - Y\|^2 + \tilde{X}^T C^{-1} \tilde{X}$$

Covariance

$$C = \mathbb{E}(X^T X)$$

~~X~~

soit X_1, \dots, X_n une base de données

X^T

$$C_{\text{empirique}} = \frac{1}{n} \sum_{k=1}^n X_k^T X_k$$

$X_k \in \mathbb{R}^N$: Hypothèse X_1, X_2, \dots, X_n

$$X_1 = (\alpha, \beta, \dots, \chi, \xi)$$

$$X_2 = (\xi, \alpha, \beta, \dots, \chi)$$

$$X_3 = (\chi, \xi, \alpha, \beta, \dots)$$

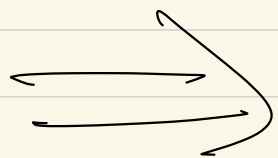
$$X_{N+1} = (\alpha, \dots, \alpha_N)$$

$$X_{N+2} = (\alpha_N, \alpha_1, \dots, \alpha_{N-1})$$

\vdots

\vdots

$X_{N+N} \# N$



C

est diagonalisable sur la base de Fourier.

Résolution: Wiener

$$\hat{K}(\omega)$$

$$\hat{g}(\omega)$$

$$|\hat{K}(\omega)|^2 + \frac{\sigma_b^2}{\sigma_b^2} \hat{g}(\omega)$$

- Le signal est supposé être un vecteur gaussien:

$$X \text{ v.a. } \in \mathbb{R}^N$$

$$\mathbb{P}(X) \sim e^{-\frac{1}{2} X^T C^{-1} X}$$

$$C = E(X^T X) \text{ matrice de covariance déf. pos.}$$

$$\mathbb{P}(B) \sim e^{-\frac{1}{2} \frac{\|B\|^2}{\sigma_b^2}}$$

$$Y = AX + B$$

$$\text{trouver } \tilde{X} = DY \text{ t.q.}$$

$$E(\|\tilde{X} - X\|^2) \text{ soit minimale}$$

$$\tilde{X} = \underbrace{(A^T A + \sigma_b^2 C^{-1})^{-1} A^T}_{D} Y$$

$$\|A\tilde{X} - AX\|^2 + \dots$$

Conclusions provisoires

- Il n'est pas praticable d'inverser simplement le défaut qui affecte une image pour améliorer la qualité de celle-ci.
- Divers moyens (nous avons montré l'utilisation d'une hypothèse statistique) permettent de mettre en place une énergie dont le minimiseur est une image satisfaisante.

Méthodes de minimisation

- La méthode la plus naturelle pour minimiser une énergie est la descente de gradient:
 - Il s'agit de trouver une direction dans laquelle l'énergie décroît le plus vite.
 - Elle revient à dériver l'énergie suivant par rapport à sa variable (qui peut être en grande dimension).
- Exemples: $E(f) = \|f - g\|^2 \quad \nabla E = (f - g)$

$$E(f) = \iint \|\nabla f\|^2 \quad \nabla E = \Delta f$$

$$E(f) = \sum_k \sum_l \left((f(k+1, l) - f(k, l))^2 + (f(k, l+1) - f(k, l))^2 \right)$$

$$K_u = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

$$K_y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbb{E}(f) = \frac{\|K_x * f\|_2^2}{\|K_x * f\|_2^2 + \|K_y * f\|_2^2}$$

$$E_K(f) = \|K * f\|_2^2$$

$$\|K * (f + \delta f)\|_2^2 - \|K * f\|_2^2$$

$$F_K(f) = \| (K * f) \|_L^2$$

con la TF \hat{f} in \mathbb{R}^n

$$E_{\langle} (f) = \sum_{\omega} |\hat{K}(\omega)|^2 |\hat{f}(\omega)|^2$$

$$|\alpha|^2 \quad |t+\delta|^2 - |\alpha|^2 \quad |t|^2$$

↑
spekt

$$121^2 \left(\overbrace{(t+s)(t+s)}^{1t^2} \right)$$

$$= |\alpha|^2 (|\beta|^2 + 2 \operatorname{Re}(\delta t))$$

~~$\frac{1}{10} \frac{1}{10}$~~ \rightarrow nicht

$$\sim |\alpha|^2 2\operatorname{Re}(\delta t)$$

$$\text{Si } E_K(f) = \|K * f\|_2^2$$

$$\text{donc } \nabla E_K(f) =$$

$$(K * \tilde{K}) * f$$

$$\tilde{K}(h, l) = K(-h, l)$$

$$K = (1, -1) \quad \tilde{K} = (-1, 1)$$

$$K * \tilde{K} = 1, 2, -1$$

$$E(f) = \|Af - g\|^2 + \int \|f\|^2$$

$$\Leftrightarrow C = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$E(f) = \|K_n * f\|_2^2$$

$$+ \|K_y * f\|_2^2$$

$$\nabla E(f) = \begin{pmatrix} 1, 4, -1 \\ -1 \end{pmatrix} * f$$

$$= \Delta f$$

$$E(f) = \int \| \nabla f \|^2$$

$$E(f) = \int_{\omega} \omega^2 |\hat{f}(\omega)|^2$$

ω vector fréquence.

Descente de gradient

- Le schéma général est le suivant:

$$f^{n+1} = f^n - \rho_n \nabla E(f^n)$$

Le pas de descente ρ_n peut être constant ou variable (voir plus loin pour un exemple)



$$F'(t) \equiv \text{signe}(t)$$

$$t_{n+1} = t_n - \int \text{signe}(t_n)$$

$$t_{n+1} = t_n - f(2t_n) \\ = (1-2f)t_n$$

$$\rightarrow (1-2f) \leq 1$$

$$1-2f > 1$$

$$\boxed{f \leq 1}$$

La variation totale

- Comme nous l'avons vu les principaux termes de régularisation qui ont été utilisés sont des termes quadratiques.
- Ils présentent des avantages essentiellement calculatoires. (Fourier, descente de gradient maîtrisée).
- Cependant, ils ont tendance à dégrader les bords de l'image en les floutant.

Définition

En continu: $TV(f) = \int \|\nabla f(x)\| dx$

En discret:

$$TV(f) = \sum_{i,j} \sqrt{(f(i+1,j) - f(i,j))^2 + (f(i,j+1) - f(i,j))^2}$$

Le gradient de la variation totale

- Remarquer que celui-ci n'est pas bien défini dans les zones où l'image est de gradient nul

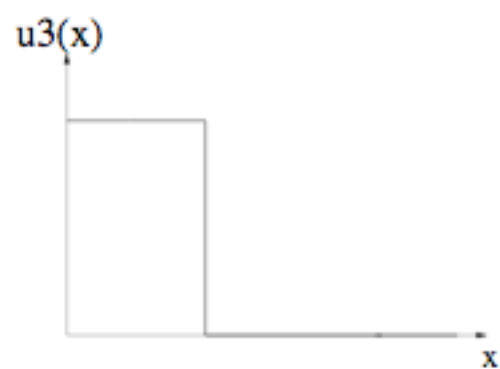
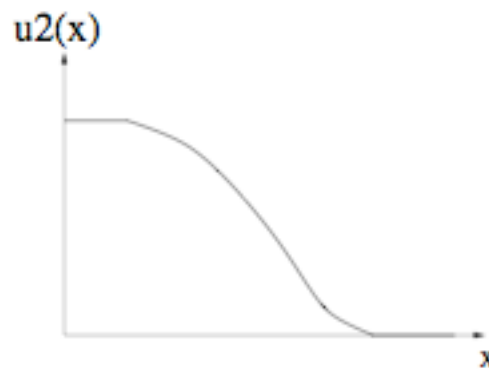
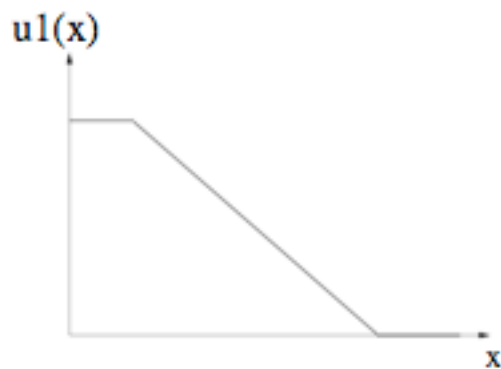
$$\nabla TV(f) = -\operatorname{div} \left(\frac{\nabla f}{\|\nabla f\|} \right)$$

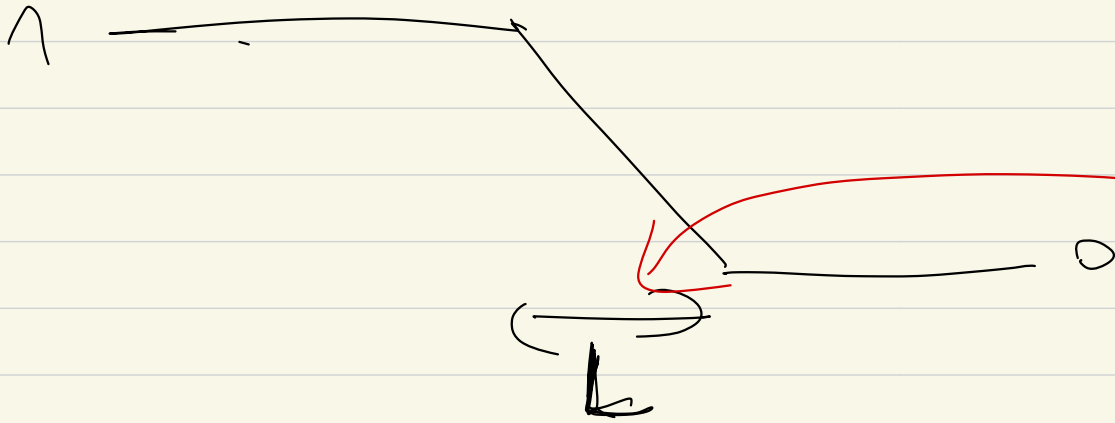
$$E(f) = \iint (\|\nabla f\|)^2$$

$$\nabla E(f) = \Delta f$$

Justification de l'utilisation de la variation totale comme fonctionnelle de régularisation

- Contrairement aux fonctionnelles quadratiques l'utilisation de la variation totale n'implique pas des solutions lisses.
- Parmi les trois fonctions ci-dessous, la variation totale est identique, contrairement à l'intégrale du carré de la dérivée.
- D'autres exemples suivront parmi les applications.





$$\|f - g\|^2 + TV(f)$$

$$\|f - g\|^2 + \int |\sigma f|^2$$

L gewant

$$\int_{-1}^{+1} |f'(t)| dt = 1$$

$$\int |f'(t)|^2 dt \stackrel{L}{=} \int_0^L \left(\frac{1}{L}\right)^2 = \frac{1}{L}$$

Ex: Zoom par minimisation de la VT

- François Malgouyres propose une méthode de zoom par minimisation de la variation totale.
- Elle se résout par descente de gradient sous contrainte.

$$u_0 = QSu$$

u_0 est l'image originale qui doit être obtenue par sous-échantillonnage (Q) d'une version filtrée de u (Su).

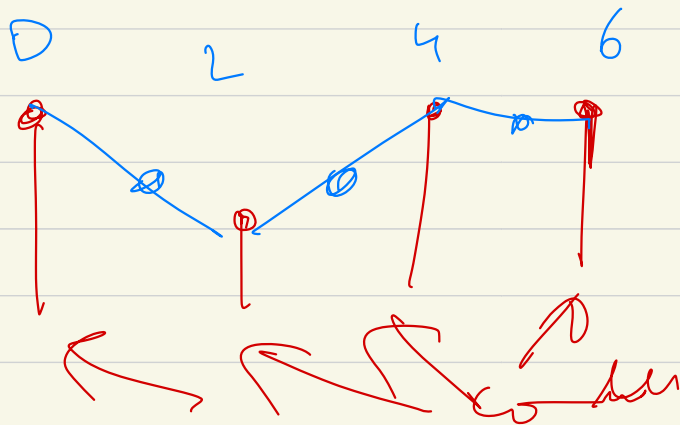
u_0 de taille $N \times N$ observée

u de taille $2N \times 2N$ qui zoom u_0 ?

$$\left\{ \begin{array}{l} u(2k, 2l) = u_0(k, l) \text{ contrainte} \end{array} \right.$$

$E(u)$ soit faible

En 1D: $E(u) = \sum |u_{n+1} - u_n|^2$





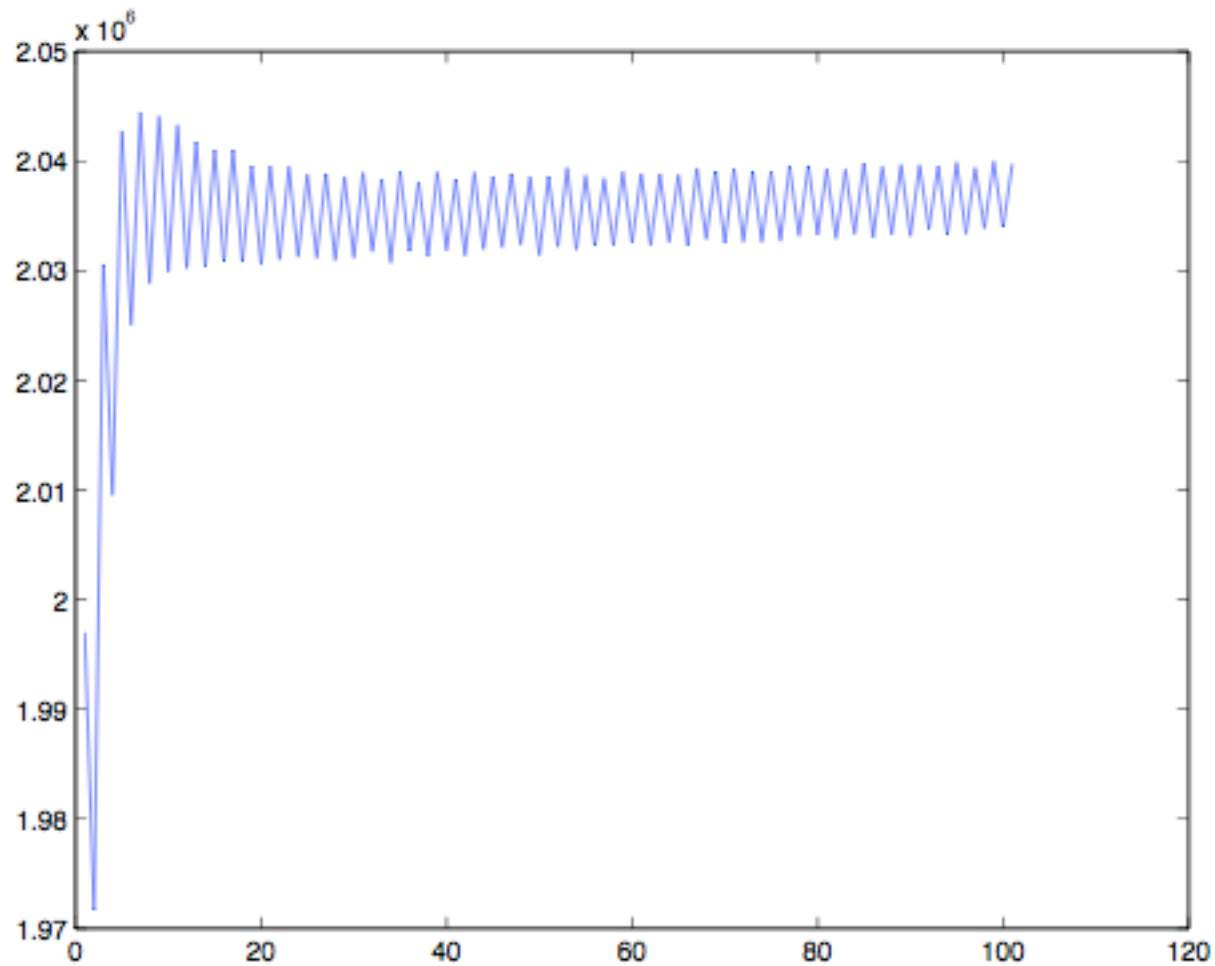






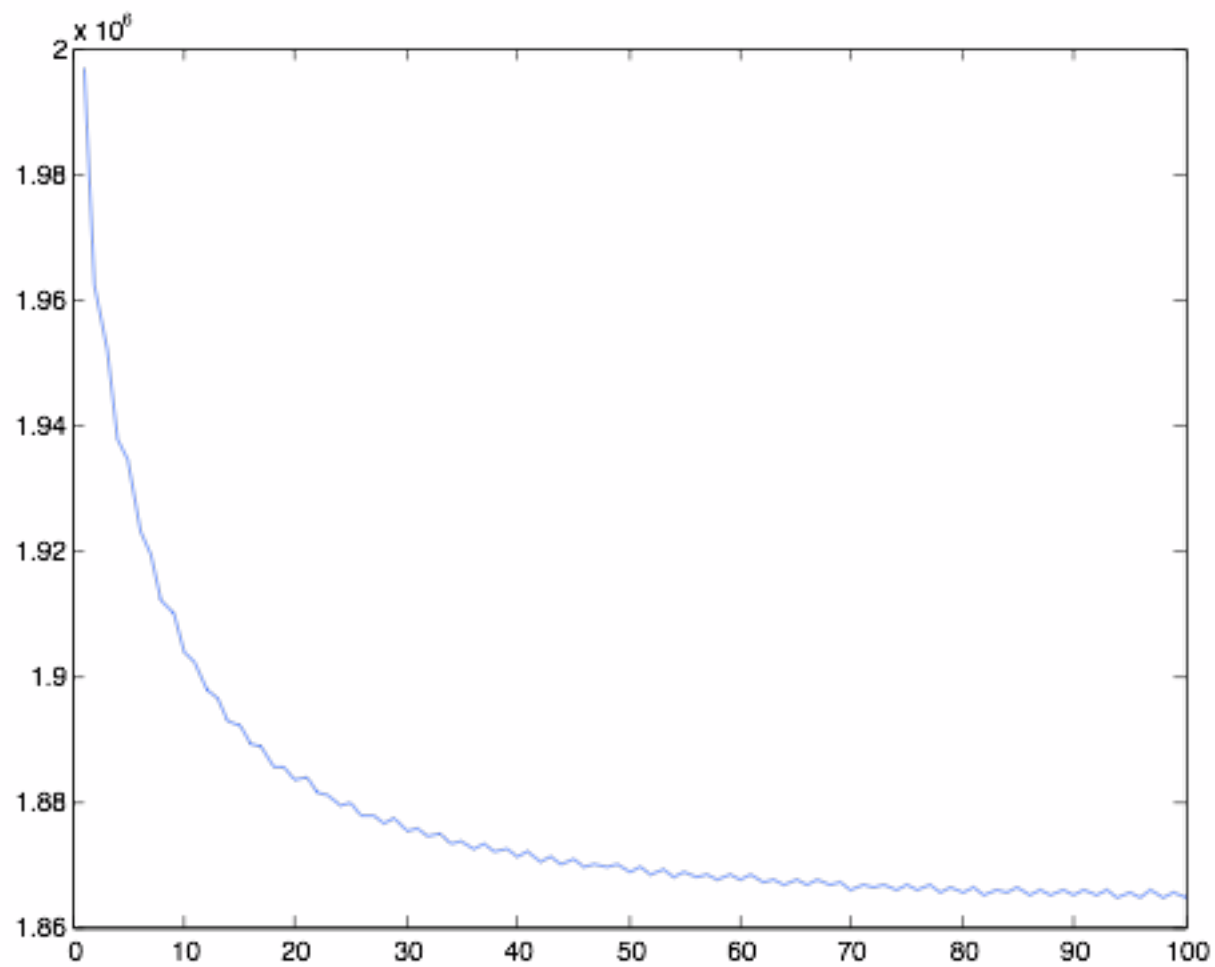
Problèmes numériques de minimisation de la VT par descente
de gradient.

$$Pas = I$$



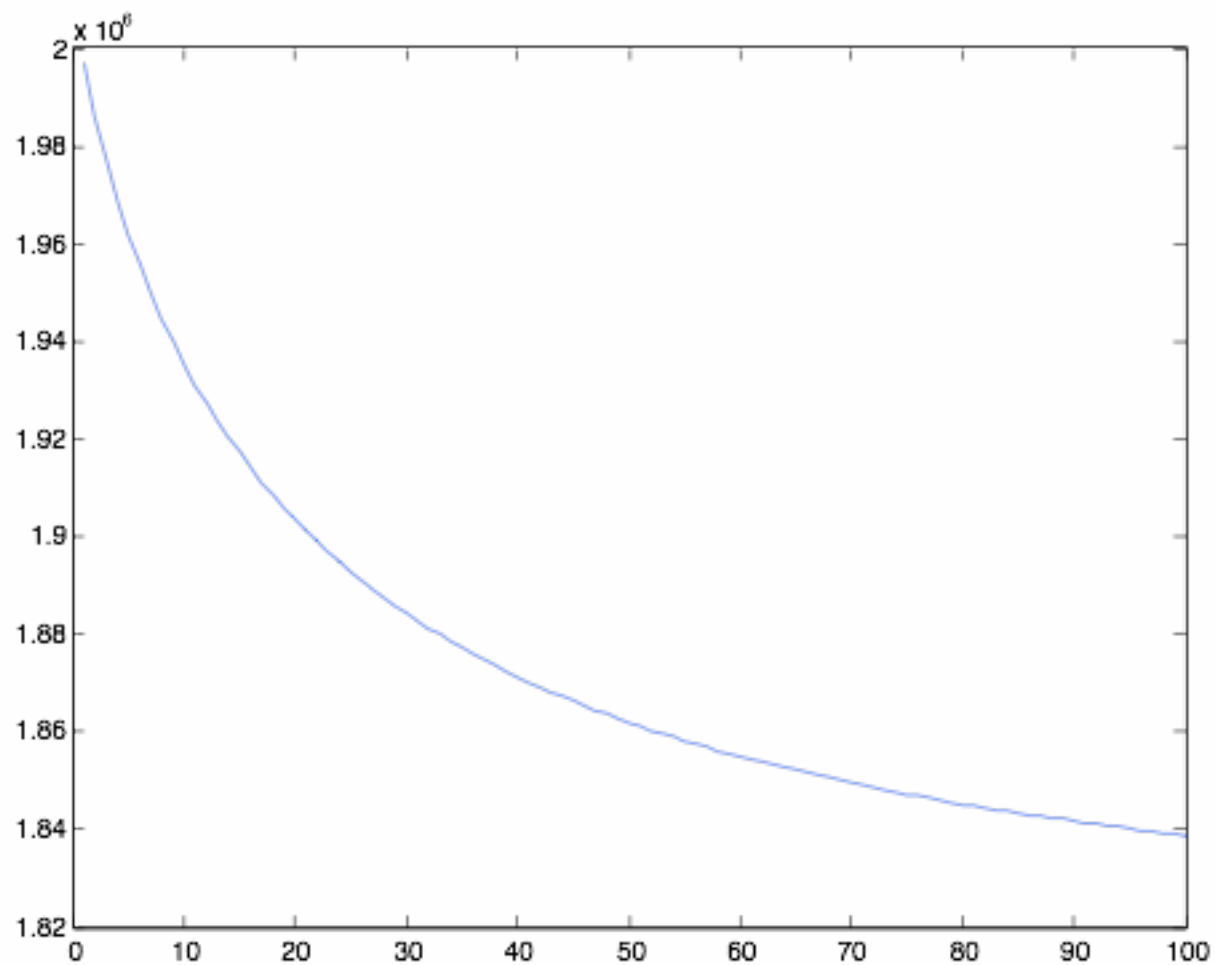
La variation totale augmente!!!

Pas=0.5



Cela semble bien se passer.

Pas=0.1



On pouvait faire mieux!

Débruitage par TV (ROF)

$$Y = \underbrace{Id X + b}_{\text{bruit}}$$

- Rudin, Osher et Fatemi ont proposé un schéma de minimisation de la variation totale sous contrainte de bruit gaussien connu.

Minimiser:

$$\int \|\nabla u\|$$

$$+ \|u - u_0\|_2^2$$

Sous contrainte: $\|u - u_0\|_2^2 = \sigma^2$

Cela aboutit à une équation d'Euler-Lagrange:

$$\frac{\partial u}{\partial t} = \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) - \lambda(u - u_0)$$

Les auteurs proposent une détermination de paramètre dynamiquement :

$$\lambda = -\frac{1}{2\sigma^2} \int \|\nabla u\| - \frac{1}{\|\nabla u\|} \times (\nabla u_0 \cdot \nabla u)$$

Une méthode de minimisation de la variation totale:

Transformée de
Fenchel:

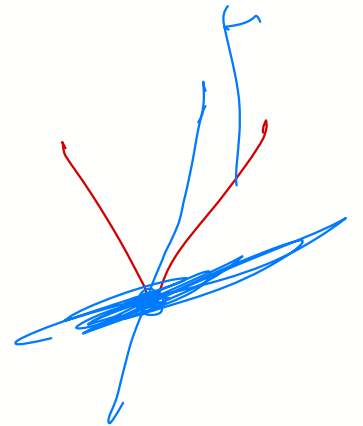
$$J^*(v) = \sup_u \langle u | v \rangle - J(u)$$

J^* est convexe.

$$J^{**} = J$$

Si: $J(\alpha u) = |\alpha|J(u)$ alors:

$$J^* = \begin{cases} 0 & \text{sur un fermé } K \\ +\infty & \text{ailleurs} \end{cases}$$

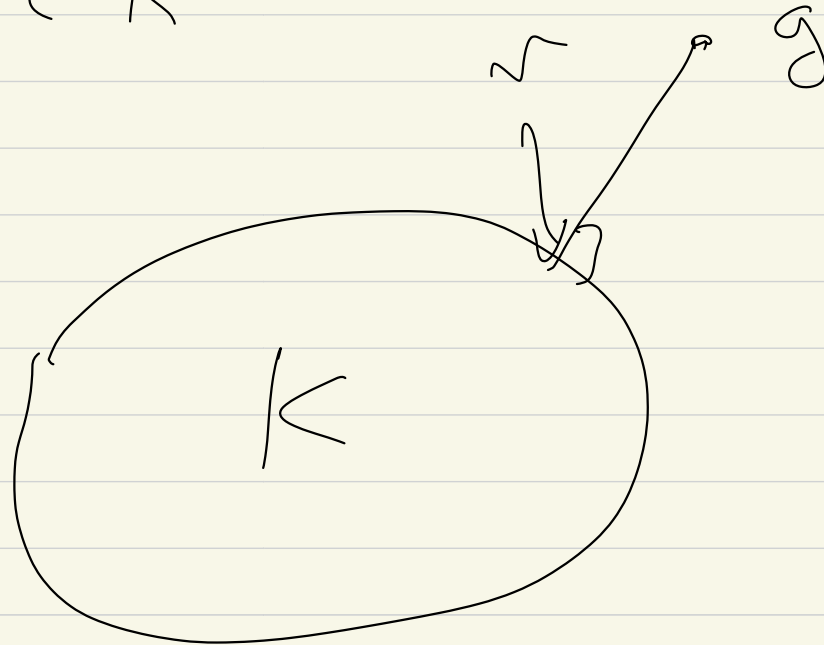


$$J^*(v) = \begin{cases} 0 & \text{si } v \in K \\ +\infty & \text{si } v \notin K \end{cases}$$

K convexe fermé

$$\min_v J^*(v) + \|v - g\|^2 \stackrel{?}{=}$$

$$v \in K$$

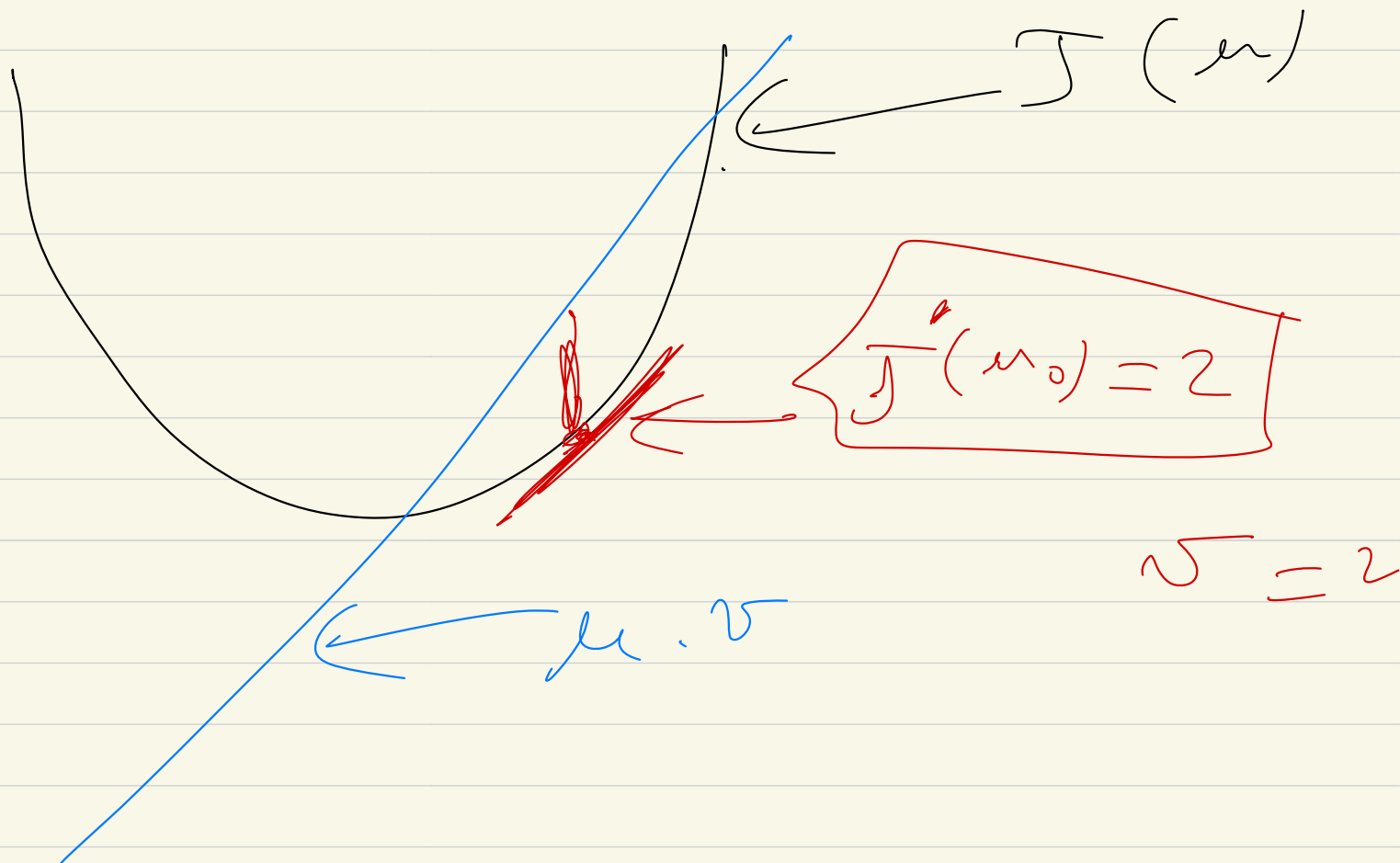


On veut minimiser : $TV(u) + \|u - g\|^2$

on montre \Leftrightarrow minimiser $TV^*(\frac{g}{\sqrt{2}}) + \|v - \frac{g}{\sqrt{2}}\|^2$

pour une certaine g_1

\Rightarrow projeter g_1 sur K



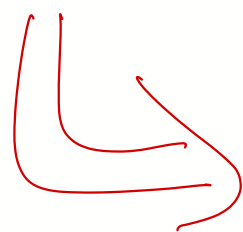
Sous-gradient

- Le sous-gradient remplace le gradient lorsqu'une fonctionnelle n'est pas différentiable.
- C'est un ensemble de vecteurs qui vérifient:

$$\forall u \in \partial J(x), \forall y, J(y) \geq J(x) + \langle u | y - x \rangle$$

Lorsque J est différentiable le sous-gradient est réduit à un singleton: le gradient.

On a: $u \in \partial J(x) \Leftrightarrow x \in \partial J^*(u)$



minimises $J(u) + \|u - g\|^2$

(\Leftrightarrow) minimises $J^*(v) + \|v - g_1\|^2$

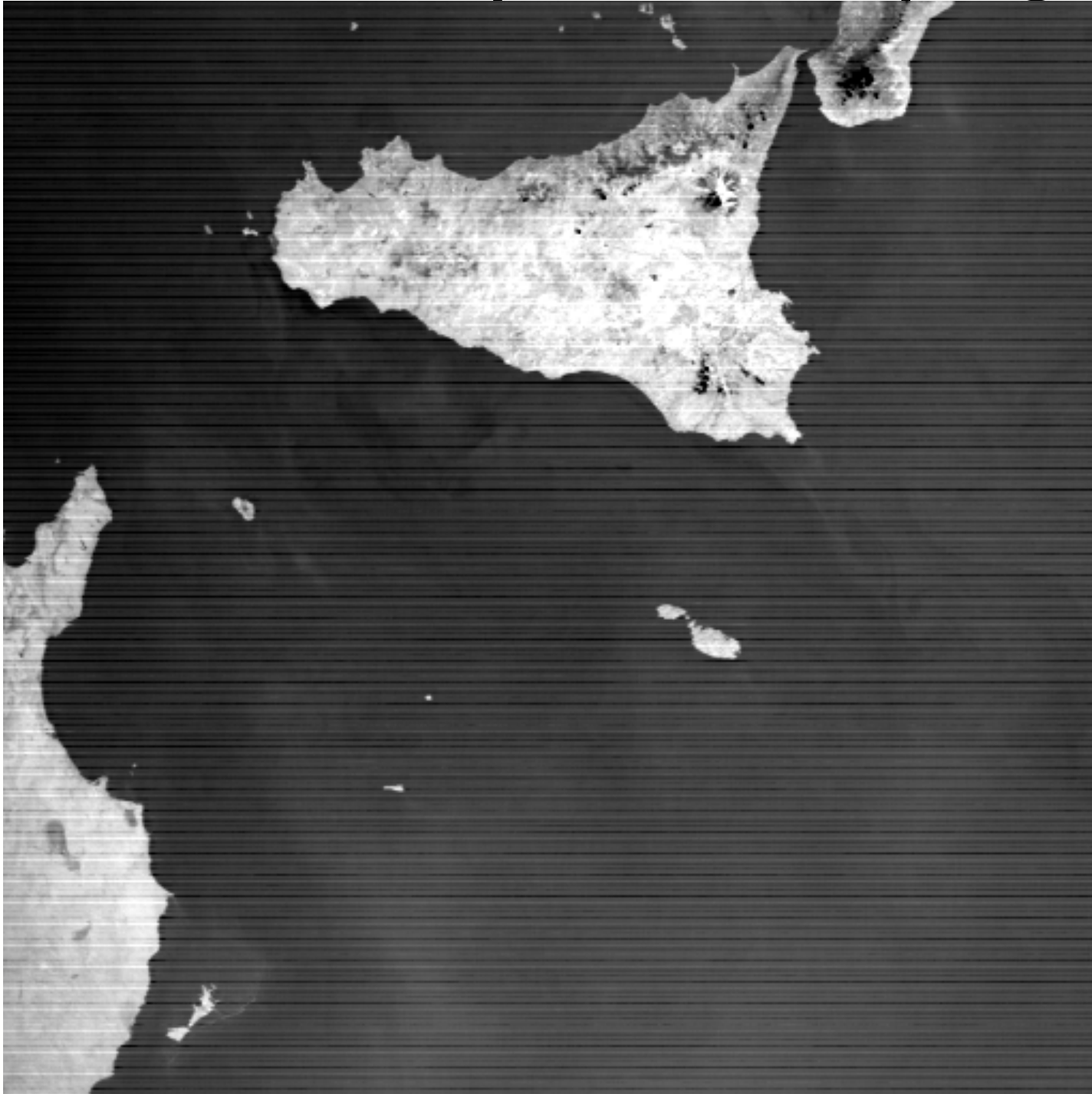
On montre que minimiser ROF revient à minimiser une fonctionnelle équivalente à base de J^* . Or J^* est très simple et la minimisation d'une fonctionnelle dans laquelle elle intervient revient à une projection sur le ferme K .

voir le TP.

projection sur K :

$$p_{ij}^{n+1} = \frac{p_{ij}^n + \tau (\nabla(\operatorname{div}(p^n) - g))_{ij}}{1 + \tau \|\nabla(\operatorname{div}(p^n) - g)\|_{ij}}$$

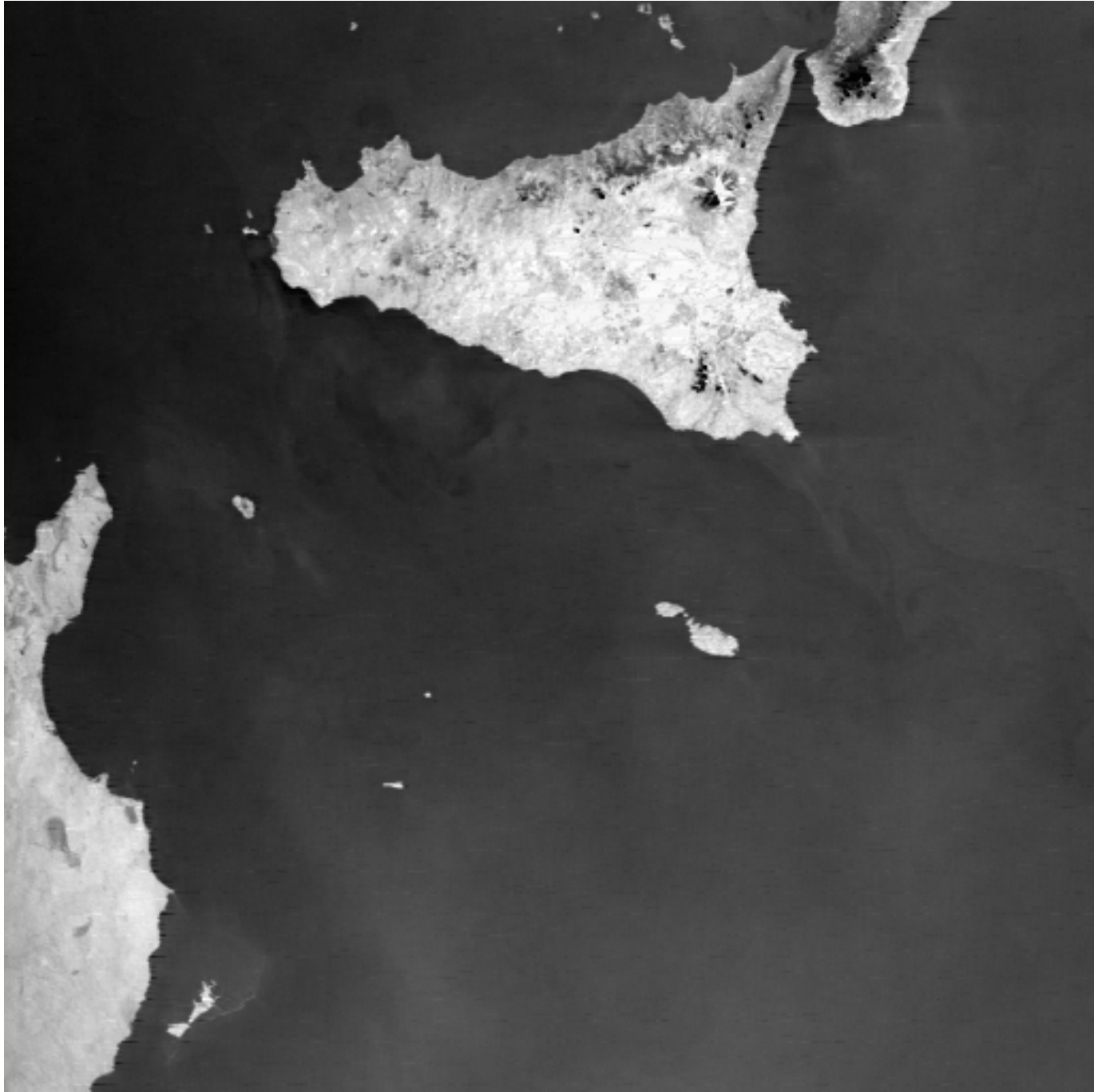
Autre exemple: Destripping



On va modéliser:

$$g = a + b. \text{ en minimisant } TV_x(b) + TV_y(a)$$





Déconvolution par TV



Déconvolution par TV

Minimisation par « split »

$$E(d, f) = TV(d) + \lambda \|g - Af\|^2 + \gamma \|d - \nabla f - b\|^2$$

d : représente le gradient de f

b : c'est un accumulateur $b_{n+1} = b_n + \nabla f - d$

On minimise alternativement par rapport à d et f .

Voir: <http://www.ipol.im/pub/art/2012/g-tvdc/>

$$\mathbb{E} (\|D(AX+B) - X\|^2)$$

$$= \frac{\|A\tilde{X} - Y\|^2}{\sigma_e^2} + \frac{\tilde{X}^T C^{-1} \tilde{X}}{2}$$

$$= \left(\frac{\|A\tilde{X} - Y\|^2}{\sigma_e^2} + \tilde{X}^T C^{-1} \tilde{X} \right)$$

$$E(f) = TV(f) + \|Af - g\|^2 \quad \text{SPLIT}$$

$$\rightarrow E(d, f) = TV(d) + \|Af - g\|^2 + \gamma \|d - \nabla f\|^2$$

$\gamma \rightarrow +\infty$
 d characte vector

$\min E(d, f)$ par rapport à f :

$$\|Af - g\|^2 + \gamma \|d - \nabla f\|^2 \rightarrow \text{quadratique}$$

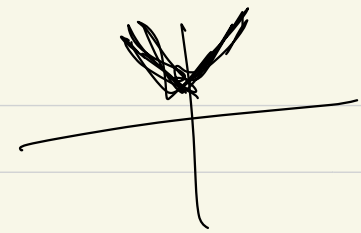
$\min E(d, f)$ par rapport à d :

$$TV(d) + \gamma \|d - \nabla f\|^2 \rightarrow \text{facile}$$

$$\sum_{i \in \Omega} \|\vec{d}_i\| + \gamma \sum_i \|\vec{d}_i - \nabla f_i\|^2$$

$$h(t) = |t| + (t - t_0)^2$$

minimiser h par rapport à t .



$$h'(t) = 1 + 2(t - t_0) \quad \text{si } t \geq 0$$

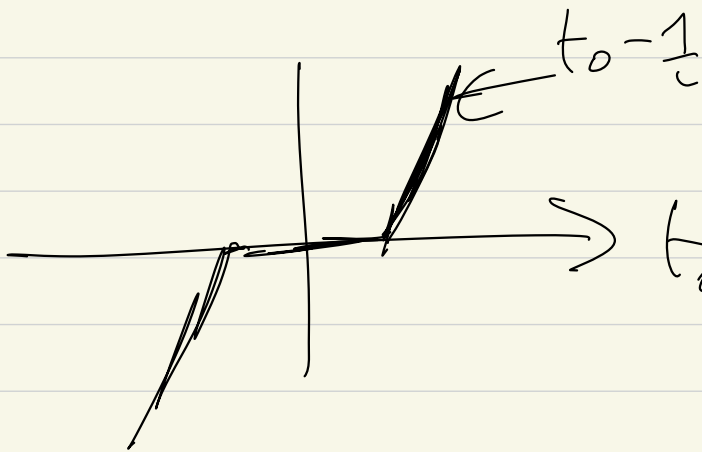
$$\text{si le minimum} \geq 0 \Rightarrow 1 + 2(t - t_0) = 0$$

$$t = t_0 - \frac{1}{2}$$

$$\text{si } t_0 > \frac{1}{2} \rightarrow t = t_0 - \frac{1}{2}$$

$$t_0 < -\frac{1}{2} \rightarrow t = t_0 + \frac{1}{2}$$

$$\text{si } - \rightarrow t = 0$$



$$h(\vec{v}) = \|\vec{v}\| + \|\vec{v} - \vec{v}_0\|^2$$

$$\|\vec{v}_0\| \geq 1 \rightarrow \vec{v} = \vec{v}_0 - \frac{1}{\|\vec{v}_0\|} \vec{v}_0$$

$$\|\vec{v}_0\| \leq \frac{1}{2} \rightarrow \vec{v} = 0$$