Modèles convexes et algorithmes d'optimisation en imagerie.

Pierre Weiss.

October 25, 2011

III.2/ Théorie de la complexité en optimisation convexe. Applications à l'imagerie.

Algorithmes d'optimisation... Deuxième partie.

Jusqu'à présent, on n'est pas capable de résoudre les problèmes annoncés au départ !

- Les descentes de sous-gradient sont trop lentes.
- Les schémas accélérés ne s'adaptent pas aux non-différentiabilités.

Conclusion :
Pour obtenir des algorithmes plus efficaces, il faut :

- Se concentrer sur des classes de fonctions plus restreintes que les fonctions convexes quelconques.
- Utiliser d'autres propriétés que les gradients et sous-gradients (opérateurs proximaux, dualité).

Quelques algorithmes d'optimisation non lisse efficaces.

- Deux mots sur les opérateurs proximaux (Moreau 1960).
- Algorithmes primaux quand une partie du problème est différentiable.
- Algorithmes duaux quand une partie du problème est fortement convexe.
- Algorithmes primaux-duaux dans le cas général.

Opérateurs proximaux.

[Combettes, Wajs 03], Signal recovery by proximal forward-backward splitting. SIAM MMS

Définition : Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ un fonction convexe s.c.i. On appelle opérateur proximal ou résolvante de f l'opérateur :

$$\operatorname{prox}_f(x_0) = \arg\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2}\|x - x_0\|^2.$$

Notes :

- Cet opérateur est bien défini car $f(x) + \frac{1}{2}\|x - x_0\|^2$ est strictement convexe et coercive.
- Il est parfois noté $(\mathrm{Id} + \partial f)^{-1}(x_0)$ car les conditions d'optimalité mènent à

$$0 \in \partial f(\operatorname{prox}_f(x_0)) + \operatorname{prox}_f(x_0) - x_0.$$

Opérateurs proximaux.

Intuition :

C'est une généralisation de la projection.

Si $f(x) = \begin{cases} 0 & \text{si } x \in X \\ +\infty & \text{sinon} \end{cases}$, alors :

$$\begin{aligned} \operatorname{prox}_f(x_0) &= \arg\min_{x \in X} \frac{1}{2}\|x - x_0\|^2 & (1) \\ &= \Pi_X(x_0). & (2) \end{aligned}$$

Propriétés élémentaires :

- L'opérateur proximal est non expansif :

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,\ \|\operatorname{prox}_f(x) - \operatorname{prox}_f(y)\| \leq \|x - y\|.$$

- Il caractérise les minimiseurs :

$$x \in \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} f(x) \Leftrightarrow 0 \in \partial f(x) \Leftrightarrow \operatorname{prox}_f(x) = x \text{ (point fixe).}$$

Une classe de problèmes non différentiable :

L'opérateur proximal est central pour la résolution - à l'aide de méthodes de premier ordre - des problèmes non différentiables. Nous allons montrer son utilisation pour la résolution de quelques classes de probèmes :

1. Somme d'une fonction convexe différentiable et d'une fonction convexe (Primal).
2. Somme d'une fonction convexe et d'une fonction fortement convexe (Dual).
3. Autres problèmes de type point selle.

Une classe de problèmes non différentiable Auslender70:

On considère désormais et jusqu'à nouvel ordre le problème suivant :

$$\min_{x\in\mathbb{R}^n} f_1(x) + f_2(x)$$

- f_1 est convexe, différentiable, à gradient L-Lispchitz.
- f_2 est convexe, s.c.i telle que prox_{f_2} peut être calculé.

Propriété : Les minimiseurs x^* sont caractérisés par

$$x^* = \mathrm{prox}_{\tau f_2}\left(x^* - \tau\nabla f_1(x^*)\right).$$

Preuve :

$$\begin{aligned} & 0 \in \nabla f_1(x^*) + \partial f_2(x^*) \\ \Leftrightarrow\ & x^* - \tau\nabla f_1(x^*) \in x^* + \tau\partial f_2(x^*) \\ \Leftrightarrow\ & x^* = \mathrm{prox}_{\tau f_2}\left(x^* - \tau\nabla f_1(x^*)\right). \end{aligned}$$

Somme d'une fonction différentiable et d'une fonction convexe..

Le gradient proximal : la caractérisation de point-fixe précédente donne envie d'utiliser le schéma :

$$x_{k+1} = \mathrm{prox}_{\tau f_2}(x_k - \tau \nabla f_1(x_k))$$

Exemple (compressive sensing) :

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2 + \|x\|_1$$

$$x_{k+1} = shrink_\tau(x_k - \tau A^T(Ax - b))$$

[I. Daubechie etal 03] An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, CPAM.

Convergence linéaire : Si $\mu = \lambda_{min}(A^T A) > 0$, alors le schéma précédent converge linéairement en choisissant $\tau = \frac{2}{\mu + L}$.

Somme d'une fonction différentiable et d'une fonction convexe..

Preuve :

$$\begin{aligned}
\|x_{k+1} - x^*\| &= \|\mathrm{prox}_{\tau f_2}(x_k - \tau A^T(Ax_k - b)) - x^*\| \\
&= \|\mathrm{prox}_{\tau f_2}(x_k - \tau A^T(Ax_k - b)) \\
&\quad -\mathrm{prox}_{\tau f_2}(x^* - \tau A^T(Ax^* - b))\| \\
&\leq \|(I - \tau A^T A)(x_k - x^*)\| \text{ (non expansif)} \\
&\leq |||(I - \tau A^T A)||| \cdot \|(x_k - x^*)\|
\end{aligned}$$

avec : $|||(I - \tau A^T A)||| = \max(|1 - \tau\mu|, |1 - \tau L|)$.
Enfin, on a

$$\min_{\tau} \max(1 - \tau\mu, 1 - \tau L) = \frac{2}{\mu + L}.$$

Que se passe-t-il dans le cas non fortement convexe ?

Somme d'une fonction différentiable et d'une fonction convexe.

f_1 différentiable + gradient Lipschitz $\Rightarrow$

$$f_1(x) \leq f_1(x_k) + \langle \nabla f_1(x_k), x - x_k \rangle + \frac{L}{2} ||x - x_k||_2^2$$

Somme d'une fonction différentiable et d'une fonction convexe.

f_1 différentiable + gradient Lipschitz $\Rightarrow$

$$f_1(x) \leq f_1(x_k) + \langle \nabla f_1(x_k), x - x_k \rangle + \frac{L}{2}||x - x_k||_2^2$$

$$f_1(x) + f_2(x) \leq \underbrace{f_1(x_k) + \langle \nabla f_1(x_k), x - x_k \rangle + \frac{L}{2}||x - x_k||_2^2 + f_2(x)}_{\Psi_k(x)}$$

En prenant $x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \Psi_k(x)$ on assure :

$$f(x_{k+1}) \leq f(x_k) - \clubsuit$$

où $\clubsuit$ est une "généralisation" de $\frac{\|\nabla f(x_k)\|^2}{2L}$.

Somme d'une fonction différentiable et d'une fonction convexe.

L'itération précédente est équivalente à :

$$x_{k+1} = \underset{x \in \mathbb{R}^n}{\arg\min} \frac{1}{L} f_2(x) + \frac{1}{2} \left\| x - \left(x_k - \frac{\nabla f_1(x_k)}{L} \right) \right\|$$

Soit encore :

$$x_{k+1} = \operatorname{prox}_{f_2/L} \left(x^k - \frac{\nabla f_1(x^k))}{L} \right)$$

Somme d'une fonction différentiable et d'une fonction convexe.

Résultat de convergence 1 [Nesterov 07]

Sous les hypothèses précédentes, pour d admissible tel que $||d|| \leq 1$

$$Df(x^k)(d) \geq -O\left(L\frac{f(x^0) - \min f}{\sqrt{k}}\right)$$

Résultat de convergence 2

Si de plus f_1 est convexe alors :

$$f(x^k) - f(x^*) \leq \frac{L||y^0 - y^*||^2}{k} \quad \text{(variables duales)}$$

Nette amélioration par rapport aux techniques de sous-gradient!

Schéma multi-pas [Nesterov 07].

- **In:** Nombre d' itérations N, point initial $x_0 \in \mathbb{R}^n$.
- **Out:** x_N une estimée de x^*.
- **Init:** Poser $t_1 = 1$, $y_1 = x_0$.

Pour k allant de 0 à N:

- Poser $x_k = \mathrm{prox}_{f_2/L}\left(y_k - \frac{\nabla f_1(y_k)}{L}\right)$.
- Calculer $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$.
- Poser $y_{k+1} = x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})$.

Schéma multi-pas [Nesterov 07].

Résultat de convergence

L'algorithme assure que :

$$f(x^k) - f(x^*) \leq \frac{L||x^0 - x^*||^2}{k^2}$$

C'est un taux de convergence optimal.

Résultats pratiques...

- Pour les transformées "simples" de l'analyse harmonique, 30 itérations mènent à une précision suffisante pour le système visuel.
- Vers du temps réel ? De l'ordre de la seconde pour une image 1000×1000 (architecture GPU).

Minimisation d'une classe de fonctions fortement convexes.

Le cadre précédent ne permet pas de traiter des problèmes de type :

$$\min_{x\in\mathbb{R}^n} \|\nabla x\|_1 + \frac{1}{2}\|x-x_0\|^2$$

en effet :

- $x \mapsto \|\nabla x\|_1$ est trop complexe pour en calculer la résolvante.
- $x \mapsto \|\nabla x\|_1$ n'est pas différentiable...

On va voir que la dualité permet de contourner ces problèmes.

Une classe de fonctions fortement convexes.

$$\min_{x\in\mathbb{R}^n} P(x) := f_1(Ax) + f_2(x)$$

- $f_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convexe, s.c.i.
- $A : \mathbb{R}^n \to \mathbb{R}^m$ linéaire (ou affine).
- $f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ μ-fortement convexe.
- $A \cdot ri(\mathrm{dom}(f_2)) \cap ri(\mathrm{dom}(f_1)) \neq \emptyset$.

Exemple typique (TV + terme l^2) :

$$\min_{x\in\mathbb{R}^n} ||\nabla x||_1 + \frac{\lambda}{2}||x - x^0||_2^2$$

Compléments d'analyse convexe...

Définition (polaire)
Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe s.c.i. La transformée de Fenchel ou polaire de f est définie par

$$f^*(x') = \sup_{x \in \mathbb{R}^n} \langle x', x \rangle - f(x)$$

Propriétés

- f^* est convexe s.c.i.
- On a $f^{**} = f$ ce qui implique que :

$$\begin{aligned} & f(x') \\ =& f^{**}(x') \\ =& \sup_{x \in \mathbb{R}^n} \langle x', x \rangle - f^*(x) \end{aligned}$$

On peut redéfinir f à partir de sa polaire.

Compléments d'analyse convexe...

Dualité de Fenchel-Rockafellar : On a donc :

$$
\begin{aligned}
&\underbrace{\min_{x\in\mathbb{R}^n} f_1(Ax) + f_2(x)}_{\text{Problème primal}} \\
&= \min_{x\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^m} \langle Ax, y\rangle - f_1^*(y) + f_2(x) \\
&= \min_{x\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^m} \langle x, A^*y\rangle - f_1^*(y) + f_2(x) \\
&= \sup_{y\in\mathbb{R}^m} \inf_{x\in\mathbb{R}^n} \langle x, A^*y\rangle - f_1^*(y) + f_2(x) \quad \text{(Voir schéma)} \\
&= \sup_{y\in\mathbb{R}^m} -f_1^*(y) - \sup_{x\in\mathbb{R}^n} -\langle x, A^*y\rangle - f_2(x) \\
&= \sup_{y\in\mathbb{R}^m} -f_1^*(y) - f_2^*(-A^*x) \\
&= -\underbrace{\inf_{y\in\mathbb{R}^m} f_1^*(y) + f_2^*(-A^*x)}_{\text{Problème dual}}
\end{aligned}
$$

Compléments d'analyse convexe...

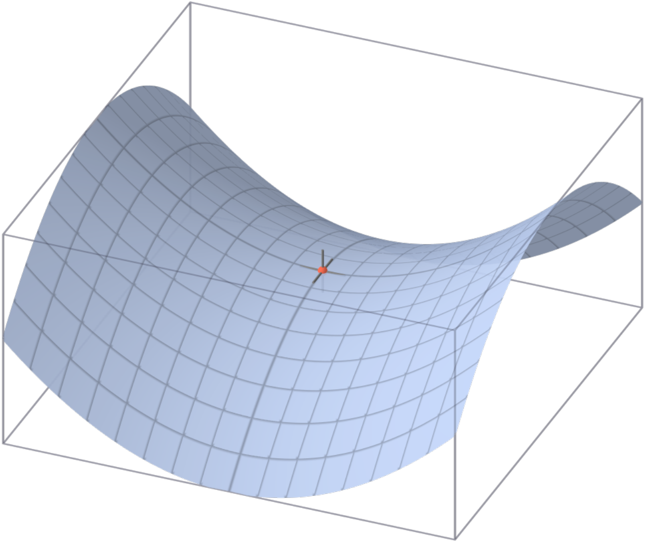


Figure: Une fonction convexe-concave et son point-selle.

Relations primales-duales.
Soit (x^*, y^*) un point selle, on a :

$$0 \in Ax^* - \partial f_1^*(y^*)$$

$$0 \in -A^*y^* - \partial f_2(x^*)$$

Soit encore :

$$y^* \in (\partial f_1)(Ax^*)$$

$$x^* \in (\partial f_2^*)(-A^*y^*)$$

Car :

$$x_2 \in \partial f(x_1) \Leftrightarrow x_1 \in \partial f^*(x_2)$$

Les solutions du problème dual apportent des informations sur le primal et vice-versa !

Compléments d'analyse convexe...

Théorème (dernières pages du livre de J.B. Hiriart-Urruty et C. Lemaréchal): Soit $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ une fonction convexe, s.c.i. Les deux propositions suivantes sont équivalentes :

1. f est différentiable et ∇f est L-Lipschitz.
2. f^* est fortement convexe de module $\frac{1}{L}$.

Minimisation d'une classe de fonctions fortement convexes.

Conclusion

Si f_2 est μ-fortement convexe, on a :

$$\min_{x\in\mathbb{R}^n} f_1(Ax) + f_2(x) = -\min_{y\in\mathbb{R}^m} f_2^*(-A^*y) + f_1^*(y)$$

- f_2^* est différentiable avec un gradient $\frac{1}{\mu}$-Lipschitz.
- f_1^* est convexe.
- Pour tout $y^* \in Y^*$, $x^* = \nabla f_2^*(-A^*y^*)$.

Un algorithme naturel : gradient proximal

$$y^{k+1} = \text{prox}_{f_1^*/L}\left(y^k + A\nabla f_2^*(-A^*y^k)\right)$$
$$x^{k+1} = \nabla f_2^*(-A^*y^{k+1}).$$

Minimisation d'une classe de fonctions fortement convexes.

Analyse du taux de convergence.

On a :

$$\min_{x\in\mathbb{R}^n} \underbrace{f_1(Ax) + f_2(x)}_{P(x)} = -\min_{y\in\mathbb{R}^m} \underbrace{f_2^*(-A^*y) + f_1^*(y)}_{D(y)}$$

et pour tout $y^* \in Y^*$, $x^* = \nabla f_2^*(-A^*y^*)$.

Proposition [Peyré, Fadili, Weiss] (sur demande)

Soit

$$x(y) = \nabla f_2^*(-A^*y)$$

Alors

$$\|x(y) - x^*\|^2 \leq \frac{2}{\mu}|D(y) - D(y^*)|$$

Un taux de CV sur le dual ⇒ taux de CV sur le primal.

Minimisation d'une classe de fonctions fortement convexes.

Résultat de convergence (gradient proximal dual)

Sous les hypothèses précédentes :

$$||x^k - x^*||^2 \leq \frac{|||A|||^2 ||y^0 - y^*||^2}{\mu k}$$

En notant $\Delta(x^k, y^k) = P(x^k) - D(y^k)$, le saut de dualité on a :

$$||x^k - x^*||^2 \leq \Delta(x^k, y^k).$$

La distance au minimiseur peut être évaluée à chaque itération !

Minimisation d'une classe de fonctions fortement convexes..

Exemple : le problème TV-L2

$$\min_{x\in\mathbb{R}^n} ||\nabla x||_1 + \frac{\lambda}{2}||x - x^0||^2 = - \min_{y\in\mathbb{R}^{2n}, ||y||_\infty \leq 1} ||\lambda\nabla^* y - x^0||^2$$

Similaire à :

- [A. Chambolle 04] An algorithm for TV minimization.

Le schéma assure que :

$$||x^k - x^*||^2 \leq \frac{8n}{\lambda k}.$$

La complexité augmente linéairement avec la dimension !

Minimisation d'une classe de fonctions fortement convexes.

Taux de convergence

Les schémas accélérés de Nesterov appliqués au dual assurent :

$$||x^k - x^*||^2 \leq \frac{2|||A|||^2||y^0 - y^*||^2}{\mu^2 k^2}$$

et si $dom(f_1) = \mathbb{R}^m$:

$$f(x^k) - f(x^*) \leq \frac{2||y^0 - y^*||^2|||A|||^2}{\mu k^2}.$$

(Sinon, rien n'assure que les itérées primales soient admissibles...)

Quelques applications pratiques des algorithmes précédents

On considère le problème :

$$\min_{x\in\mathbb{R}^n} \|\nabla x\|_1 + \frac{\lambda}{2}\|x - x_0\|_2^2$$

Débruitage de x_0 sous l'hypothèse de bruit gaussien additif.
[FILM, AUTRES EXEMPLES]

Quelques applications pratiques des algorithmes précédents

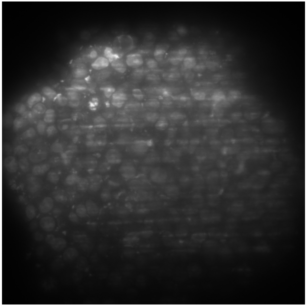
Débruitage d'images SPIM.

Travail en collaboration avec Jérôme Fehrenbach (Institut de Mathématiques de Toulouse) et le cancéropole.

- L'existence de schémas efficaces doit guider la modélisation.

Motivation et présentation de la méthode

Images SPIM: les images sont affectées par des raies sombres (ou parfois brillantes).



Motivation et présentation de la méthode

On se propose d'éliminer ces raies, pour obtenir des images plus lisibles.

- L'origine physique de ces raies est peu claire. On va donc adopter une méthode de traitement d'image sans modélisation de la physique.

Motivation et présentation de la méthode

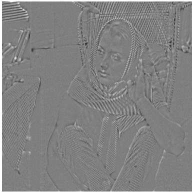
Notre algorithme s'inspire des méthodes récentes de décomposition d'images en :

- $u_0 = u + v$ (structure + texture),.
- $u_0 = u + v + b$ (structure + texture + bruit).

[Meyer 01, Starck Donoho 05, Aujol 05, Fadili 10,]

$=$

$+$



Motivation et présentation de la méthode

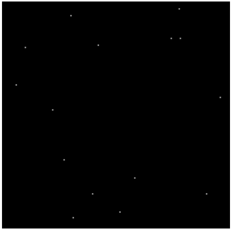
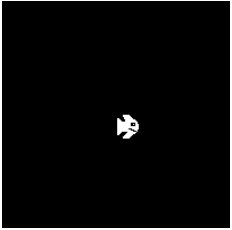
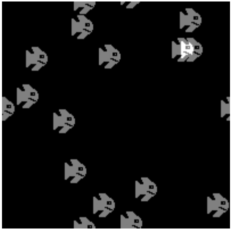
- Difficulté : grand nombre d'inconnues (plus de 500 images $1000 \times 1000 \sim$ 1-50Go.).
- Pour s'en sortir : formuler le problème comme un problème de minimisation convexe. Cela permet le développement d'algorithmes efficaces, qui convergent vers un minimum global.

Modélisation du bruit

On décrit le bruit comme étant composé de raies + bruit blanc (deux composantes).
La raie est un “motif élémentaire” qui se retrouve à plusieurs endroits dans l’image.
Pour déplacer et répliquer un motif ψ dans l’espace on utilise un produit de convolution :

$$\lambda * \psi(x) = \sum_y \lambda(y)\psi(x-y).$$

Hypothèse sous-jacente : le bruit est stationnaire.


$*$

$=$


Modélisation du bruit

De façon plus générale, le bruit est modélisé comme le produit de convolution entre un bruit blanc (pas forcément gaussien) et un noyau donné.

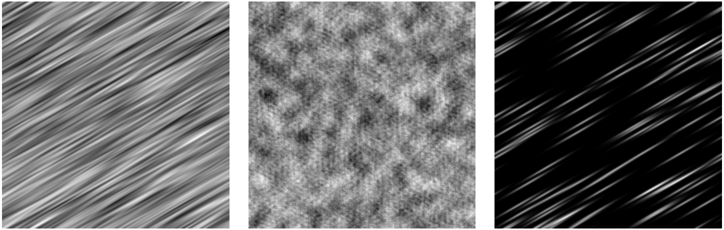
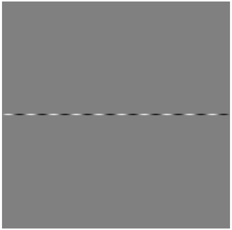
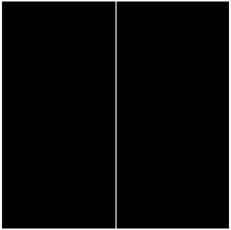
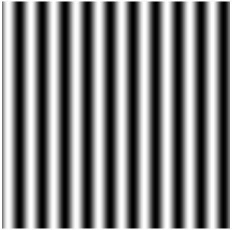
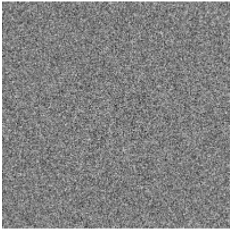
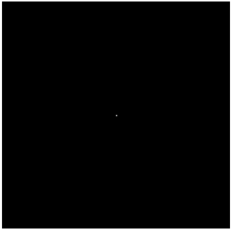
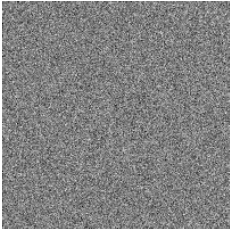


Figure: Exemples de bruits stationaires obtenus par convolution d'un noyau avec un bruit blanc.

Modélisation du bruit

Autres exemples de produit de convolution:

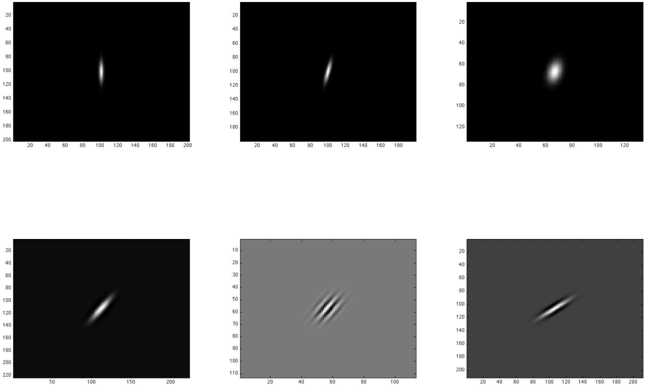
 $*$  $=$ 

 $*$  $=$ 

Le produit de convolution s'exprime facilement en Fourier.

Modélisation du bruit

Modélisation d'une raie : un *filtre de Gabor* (=une gaussienne anisotrope modulée par une sinusoïde).



Modélisation du bruit

Le bruit est décrit par plusieurs types de raies (longueur/largeur différentes). Il s'écrit

$$b = \sum_{i=1}^{m} \lambda_i * \psi_i.$$

La loi de probabilité de chaque λ_i dépend de la modélisation du bruit. D'une manière générale on écrit:

$$\mathcal{P}(\lambda_i) = \exp(-\phi_i(\lambda_i)).$$

L'hypothèse que le bruit est blanc permet de choisir des fonctions ϕ_i séparables (ce qui simplifie l'analyse numérique et est justifié par l'hypothèse de stationnarité).

Modéle de restauration

On utilise le principe du Maximum A Posteriori. On cherche l'image u la plus probable étant données :

1) la forme et l'orientation de chaque raie ψ_i
(en pratique : une/deux raie + un bruit Gaussien),

2) une description de la probabilité de chaque bruit ϕ_i,

3) une description d'un modèle d'image a priori.

La recherche de l'image u la plus probable peut s'exprimer comme un problème de minimisation convexe.

Modéle de restauration

Principe du Maximum A Posteriori (MAP) :

Connaissant $u^0 \in \mathbb{R}^n$, trouver $\operatorname*{Arg\,max}_{\lambda \in \mathbb{R}^{m \times n}} P(\lambda | u^0)$.

$$\text{Loi de Bayes :} \quad P(\lambda|u^0) = \frac{P(u^0|\lambda)P(\lambda)}{P(u^0)},$$

donc sous l'hypothèse d'indépendance de u et de λ :

$$\operatorname*{Arg\,max}_{\lambda \in \mathbb{R}^{m \times n}} P(\lambda|u^0) = \operatorname*{Arg\,min}_{\lambda \in \mathbb{R}^{m \times n}} (-\log(P(u)) - \log(P(\lambda)))$$

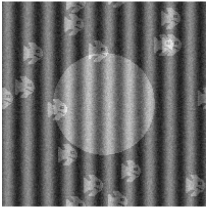
Soit encore :

$$\operatorname*{Arg\,min}_{\lambda \in \mathbb{R}^{m \times n}} \left(\left\| \nabla \left(u_0 - \sum_{i=1}^{m} \lambda_i * \Psi_i \right) \right\|_{1,\epsilon} + \sum_{i=1}^{m} \phi_i(\lambda_i) \right)$$

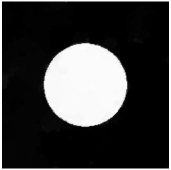
Résultats (vérification du principe)

$$\operatorname*{Arg\,min}_{\lambda \in \mathbb{R}^{3n}} \left\| \nabla \left(u - \sum_{i=1}^{3} \lambda_i \star \psi_i \right) \right\|_1 + \|\lambda_1\|_\infty + \|\lambda_2\|_2^2 + \|\lambda_3\|_1$$

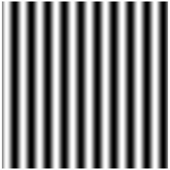
Noyaux : “raie verticale (l^∞)”, “Dirac (l^2)”, ”poisson (l^1)”.



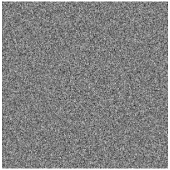
=



+



+



+

Résultats (vérification du principe)

 $\rightarrow$ 

 $\rightarrow$ 

Résultats (vérification du principe)

Résultats (vérification du principe)

 $\rightarrow$ 

Figure: Le filtre naturel : un filtre en forme de gradient topologique... Mais mauvaise résolution.

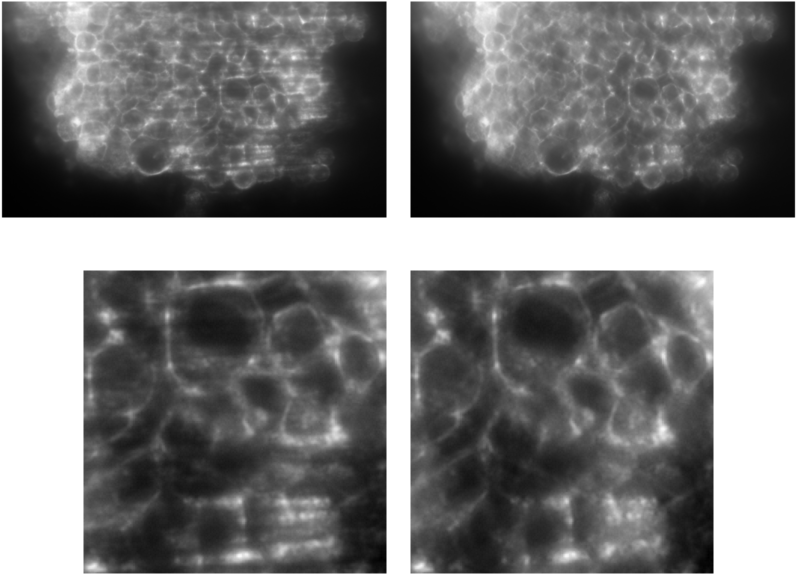
Résultats (vérification du principe)

 → 

Figure: Le filtre naturel : un filtre en forme de gradient topologique... Mais mauvaise résolution.

Figure: Super-résolution... Le cerveau de l'ombre.

Résultats



Résultats

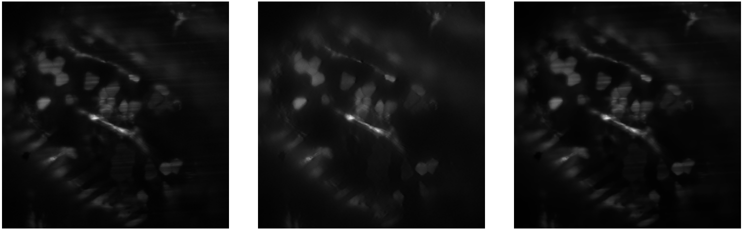
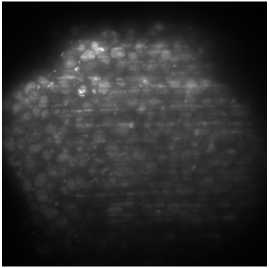


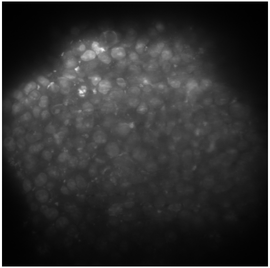
Figure: de gauche à droite : originale; 2 filtres (raie de Gabor + Dirac pour le bruit gaussien) ; méthode TV-L^2

Résultats

Sur une autre image :



$=$



$+$

Remarque sur le choix des fonctions ϕ_i

Le modèle de décomposition est le suivant :

$$\operatorname*{Arg\,min}_{\lambda \in \mathbb{R}^{m \times n}} \left\| \nabla \left(u - \sum_{i=1}^{m} \psi_i \star \lambda_i \right) \right\|_1 + \sum_{i=1}^{m} \phi_i(\lambda_i)$$

Comment choisir les fonctions ϕ_i ?

Les choix classiques sont :

- $\phi_i(\cdot) = \| \cdot \|_1$: bruit laplace ou de type impulsionnel.
- $\phi_i(\cdot) = \frac{1}{2} \| \cdot \|_2^2$: bruit Gaussien.
- $\phi_i(\cdot) = \| \cdot \|_\infty$: bruit uniforme/amplitude bornée.

Est-ce bien nécessaire ?

Remarque sur le choix des fonctions ϕ_i

Théorème de Lindeberg-Feller :
Soit

$$b = \lambda \star \psi$$

- $\lambda \in \mathbb{R}^{\mathbb{N}}$ est un bruit blanc.
- $\psi \in \mathbb{R}^{\mathbb{N}}$ est un noyau de convolution.

Si la décroissance de ψ est suffisamment lente
$\rightarrow$ b_i suit une loi gaussienne pour tout i.
Conclusion : si les noyaux ψ sont "suffisamment" étalés, la norme l^2 suffit.
Intérêt : si $\phi_i(\cdot) = \frac{1}{2}\|\cdot\|_2^2, \;\; \forall i$, le problème est fortement convexe.

Liens avec les modèles à base de normes négatives

Objectif des planches à venir :
Les modèles de décomposition en texture+ structure avec normes négatives sont un cas particulier du formalisme présenté.

Liens avec les modèles à base de normes négatives

[Meyer 01] Oscillating patterns in image processing and nonlinear evolution equations

Une texture est d'autant plus "probable" qu'elle est oscillante.

Les normes négatives ($W^{-1,p}$) permettent de détecter les fonctions oscillantes.

Soit :

$$||v||_{-1,p} = \min_{|q| \in L^p(\Omega), \mathrm{div}(q)=v} ||q||_p$$

[Aubert, Aujol, 2004] : Sur un ensemble borné $v_n \rightharpoonup 0 \Rightarrow ||v_n||_{-1,p} \to 0$ (pour $p \geq 1$).

Liens avec les modèles à base de normes négatives

[Meyer 01, Osher-Vese 03, Aujol 05, Vese 03-10...]
Décomposer u_0 en structure + texture:

$$u_0 = u + v.$$

En résolvant :

$$\underset{u+v=u_0, u_0 \in \mathbb{R}^n}{\arg\min} \|\nabla u\|_1 + \alpha \|v\|_{-1,p}$$

En posant $v = \mathrm{div} g$, ce problème peut être réécrit :

$$\underset{g \in \mathbb{R}^n \times \mathbb{R}^n}{\arg\min} \|\nabla (u_0 - \mathrm{div} g)\|_1 + \alpha \|g\|_p.$$

Liens avec les modèles à base de normes négatives

On a :

$$\text{div} g = \partial_1 g_1 + \partial_2 g_2$$

Et après discrétisation:

$$\partial_1 g_1 + \partial_2 g_2 = g_1 \star h_1 + g_2 \star h_2$$

où h_1 et h_2 dépendent du choix de discrétisation.
Par exemple :

$$h_1 = [1, -1] \quad h_2 = \left[\begin{array}{c} 1 \\ -1 \end{array} \right].$$

Les coefficients g_1 et g_2 repésentent :
les coefficients de la texture v dans un dictionnaire composé des fonctions h_1 et h_2 translatées dans l'espace.

Liens avec les modèles à base de normes négatives

Interprétation bayésienne du modèle de Y. Meyer.

- Soit Γ un v.a. de $\mathbb{R}^n$ de densité $P(\Gamma) \propto \exp(-\alpha\|\Gamma\|_p)$.
- Soit θ un v.a. de $\mathbb{R}^n$ de densité uniforme sur $[0, 2\pi]$.
- Soit $G = \begin{bmatrix} \Gamma\cos(\theta) \\ \Gamma\sin(\theta) \end{bmatrix}$.

Alors la décomposition $u_0 = u + v$ du problème de Y.Meyer
correspond à une solution MAP

- où u est la réalisation d'un v.a. de densité $\exp(-\|\nabla u\|_1)$.
- où v est la réalisation d'un v.a. $V = \mathrm{div} G$.

Liens avec les modèles à base de normes négatives

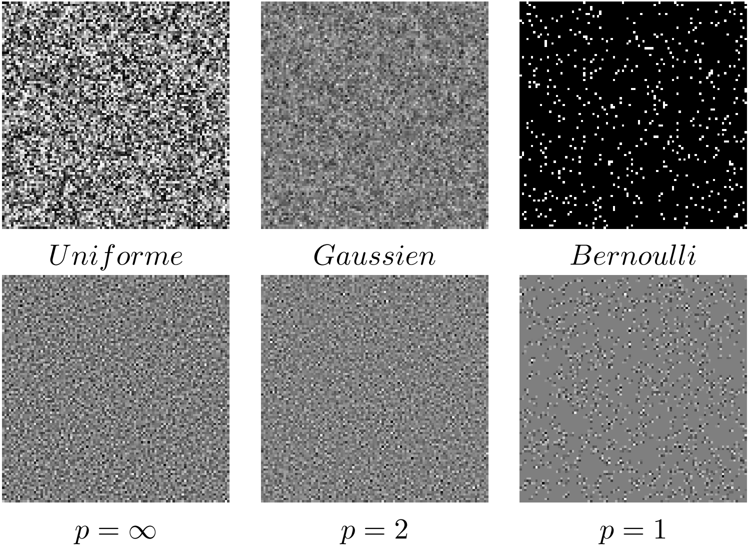


Figure: Haut : bruits standards. Bas : bruits synthétisés correspondant aux modèles $W^{-1,p}$.

Conclusion SPIM

- Suppression des raies par post-traitement des images SPIM.
- Des techniques d'optimisation convexe sont utilisées et développées.
- Une modélisation originale du bruit.

- Pour accélérer : programmation sous CUDA (facteur 100/Matlab).
- Traitement en un temps inférieur à la seconde pour une image 1000x1000.

Justification des techniques de lissage...

Un nouveau cadre de travail :

$$\min_{x\in\mathbb{R}^n} P(x) := f_1(Ax) + f_2(x)$$

- f_1 et f_2 convexes, s.c.i. et “simples”.
- $A : \mathbb{R}^n \to \mathbb{R}^m$ linéaire.
- $A \cdot ri(\mathrm{dom}(f_2)) \cap ri(\mathrm{dom}(f_1)) \neq \emptyset$.

Remarque centrale : On sait résoudre rapidement

$$\min_{x\in\mathbb{R}^n} f_1(Ax) + f_2(x) + \frac{\epsilon}{2} f_3(x)$$

où f_3 est fortement convexe ou

$$\min_{x\in\mathbb{R}^n} f_{1,\epsilon}(Ax) + f_2(x)$$

où $f_{1,\epsilon}$ est une régularisée de f_1 (e.g. Moreau-Yosida).
La rapidité des schémas peut-elle compenser les erreurs d’approximation ?

Justification des techniques de lissage...

Oui ! Idée de la preuve :
On considère le problème :

$$f_\epsilon^* = \min_{x \in \mathbb{R}^n} f_1(Ax) + \frac{\epsilon}{2}\|x - x_0\|^2 \quad (\mathcal{P}_\epsilon)$$

On génère une suite (x_k) avec un algorithme optimal sur $\mathcal{P}_\epsilon$.

$$0 \leq P(x_k) - P^* \leq \underbrace{\frac{\epsilon}{2}\|x^* - x_0\|^2}_{\text{approximation}} + \underbrace{\frac{|||A||| \cdot \|y_0 - y_\epsilon^*\|^2}{\epsilon k^2}}_{\text{schéma}}$$

Puis à k fixé, on minimise le membre de droite par rapport à ϵ.

Justification des techniques de lissage...

Lissage primal (Nesterov 2003)

Résultat

Si μ est proportionel à $1/k$, le schéma précédent assure :

$$P(x^k) - P(x^*) \leq \frac{||A|| \cdot ||x^0 - x^*_\mu||_2 \cdot ||y^0 - y^*||_2}{k}.$$

Optimalité

Le schéma est “optimal”.

- Nemirovski 1992 *Information-based complexity of linear operator equations.* Journal of Complexity.

Justification des techniques de lissage...

Lissage dual

Résultat

Si ϵ est proportionel à $1/k$, le schéma assure :

$$P(x^k) - P(x^*) \leq \frac{||A|| \cdot ||x^0 - x^*||_2 \cdot ||y^0 - y^*_\epsilon||_2}{k}.$$

Avantages du dual

- La solution du problème régularisé est unique.
- On obtient un taux de convergence en norme.

Justification des techniques de lissage...

C'est un résultat profond.

- Meilleur que les descentes de sous-gradient en $O\left(\frac{1}{k}\right)$.
- Méthode optimale pour une vaste classe d'inégalité variationnelles.

Cependant...

- Dépendance de ϵ et de k à fixer au départ.
- Allez voir le dernier preprint de A. Chambolle et T. Pock !

Design de bobines de gradient pour l'IRM (En collaboration avec M. Poole, U. Brisbane).

Problèmes liés aux bobines de gradient.

- Les champs magnétiques varient rapidement dans le temps $\Rightarrow$ courants de Foucault induits, vibrations des structures mécaniques, bruits $(> 130dB)$.
- Pour améliorer la qualité, courants importants $\Rightarrow$ échauffements locaux des bobines, détériorations.

$\Rightarrow$ Trouver un design qui réduit ces effets.

Design de bobines pour l'IRM : modélisation

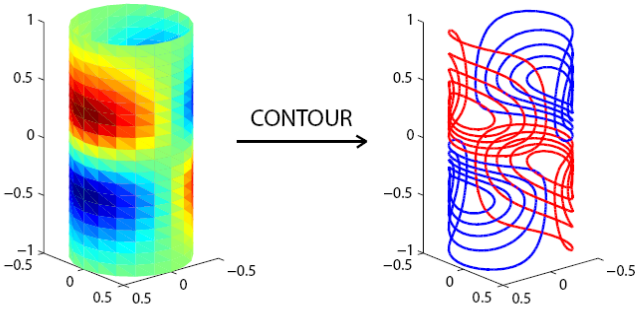
Problème inverse :

Trouver un arrangement de fils qui produise un champs magnétique donné après mise sous tension.

Solutions:

- Optimisation de forme classique et topologiques inefficaces.
  - Des fils adjacents pourraient être reliés.
  - On ne cherche qu'un seul fil continu.
- Utilisation d'une méthode de type level set.

Design de bobines pour l'IRM : modélisation



Design de bobines pour l'IRM : modélisation

Régime statique.
Le champs magnétique B est relié au champs de courant J par la loi de Biot-Savart :

$$B(r) = \frac{\mu_0}{4\pi} \int J(r') \times \frac{r - r'}{|r - r'|^3} dv$$

$\Rightarrow$ Problème inverse :
trouver J pour obtenir un B donné.

Design de bobines pour l'IRM : modélisation

Régime statique.
Le champs magnétique B est relié au champs de courant J par la loi de Biot-Savart :

$$B(r) = \frac{\mu_0}{4\pi} \int J(r') \times \frac{r - r'}{|r - r'|^3} dv$$

$\Rightarrow$ Problème inverse :
trouver J pour obtenir un B donné.
C'est un problème inverse mal-posé.

Design de bobines pour l'IRM : modélisation

Termes d'énergie classiques :

- La distance au champ cible :

$$\|B(r) - \frac{\mu_0}{4\pi} \int J(r') \times \frac{r - r'}{|r - r'|^3} dv\|_2^2$$

- L'énergie stockée :

$$W = \frac{\mu_0}{8\pi} \int \int \frac{J(r) \cdot J(r')}{|r - r'|} dv dv'$$

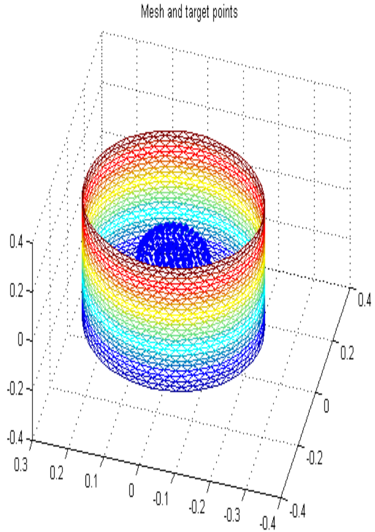
- La chaleur (effet Joule) :

$$R = \rho_C \int |J(r)|^2 dv$$

- Des forces de couple de Lorentz :

$$\tau = \int r \times J(r) \times B_0 dv.$$

Design de bobines pour l'IRM : modélisation et discrétisation



Design de bobines pour l'IRM : discrétisation

On discrétise :

$$J(r) \sim \sum_{i=1}^{n} x_n j_n(r)$$

avec $div(j_n) = 0$.
De la même façon :

$$B(r) \sim \sum_{i=1}^{n} x_n b_n(r)$$

Puis :

$$W(x) = \langle x, M_W x \rangle \text{ énergie stockée}$$

$$P(x) = \langle x, M_R x \rangle \text{ résistance}$$

$$Tx = M_T x \text{ couple magnétique}$$

Design de bobines pour l'IRM

Finalement le problème inverse standard est :

$$\text{Trouver } x \in \underset{x \in \mathbb{R}^n, Tx=0}{\arg\min} \|Bx - b\|_2^2 + \alpha W(x) + \beta P(x)$$

Peut être résolu par des techniques d'algèbre linéaire.

Design de bobines pour l'IRM

Finalement le problème inverse standard est :

$$\text{Trouver } x \in \underset{x \in \mathbb{R}^n, Tx = 0}{\arg\min} \|Bx - b\|_2^2 + \alpha W(x) + \beta P(x)$$

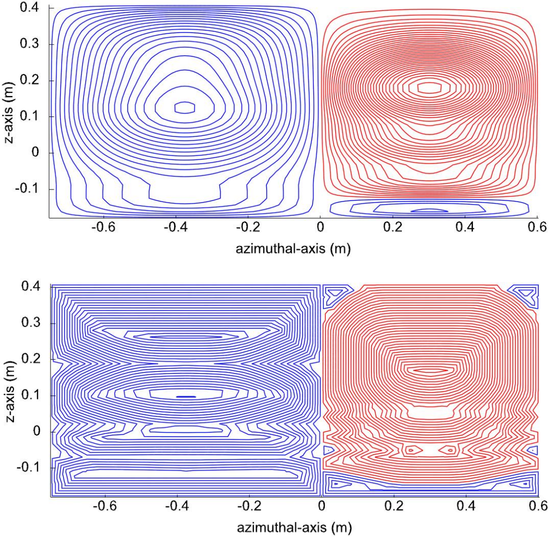
Peut être résolu par des techniques d'algèbre linéaire.

Nouveauté : minimiser les échauffements locaux de la bobine

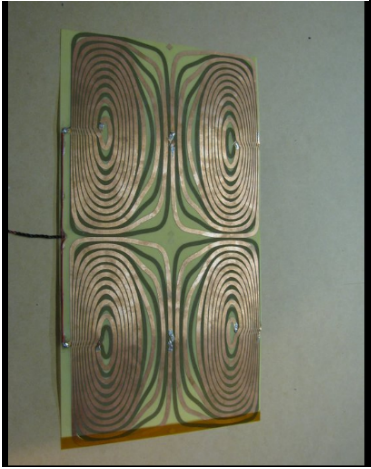
$$\text{Trouver } x \in \underset{x \in \mathbb{R}^n, Tx = 0}{\arg\min} \|Bx - b\|_2^2 + \alpha W(x) + \beta P(x) + \gamma \|j(x)\|_\infty$$

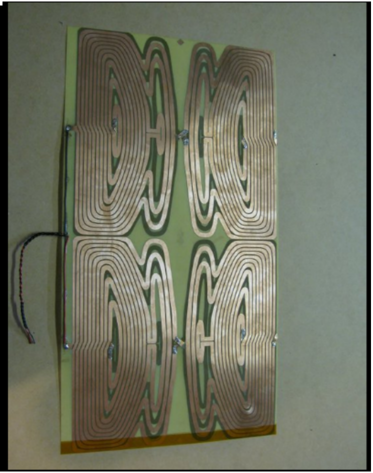
C'est un problème d'optimisation convexe (programmation quadratique).

Design de bobines pour l'IRM : résultats.

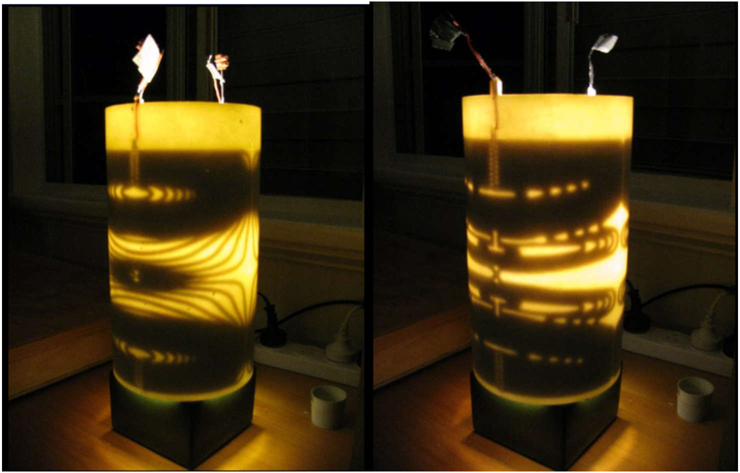


Design de bobines pour l'IRM : résultats.

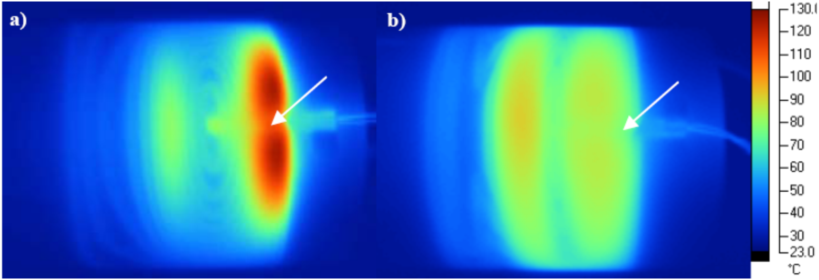




Design de bobines pour l'IRM : résultats.



Design de bobines pour l'IRM : résultats.



Comparaisons chaleur. Design classique - Nouveau Design.

Ce n'est qu'un au revoir !

- Merci énormément à Maïtîne et Pierre pour l'invitation et l'organisation !