

Échantillonnage en grande dimension sous contraintes

Groupe de Lecture Apprentissage et Statistique

Pierre MINIER

IMS, Groupe Signal-Image

28 mars 2024

Un (deux ?) article étudié

An MCMC Method for Uncertainty Quantification in Nonnegativity Constrained Inverse Problems

- Date: 2012 (+ compléments avec celui de 2020)
- Journal: Inverse Problems in Science and Engineering



(a) J. M. Bardsley, Montana



(b) C. Fox, Otago

Figure 1: Les auteurs

Sommaire

- 1 Contrainte de positivité vs non-négativité
- 2 Échantillonnage sans contrainte
- 3 Échantillonnage sous contrainte de positivité
- 4 Échantillonnage sous contrainte de non négativité
- 5 Résultats de l'étude

Positivité vs non négativité

Contrainte de positivité (stricte)

Une probabilité nulle d'avoir des pixels nuls

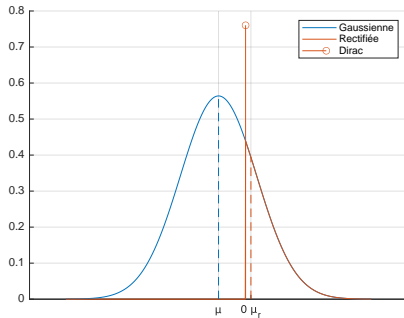
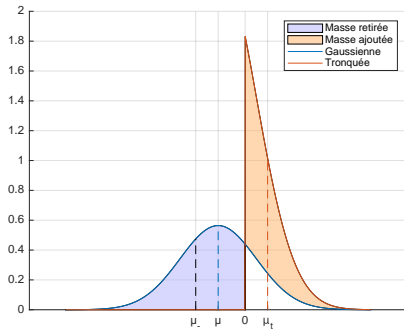
$$\{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} > 0\}$$

Contrainte de non négativité

Une probabilité non nulle d'avoir des pixels nuls

$$\{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq 0\}$$

Exemples gaussiens



(a) Normale tronquée – Positivité

(b) Normale rectifiée – Non négativité

Figure 2: Exemples possibles de modifications

Notations

Modèle d'observation linéaire

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta} \quad (1)$$

- $\mathbf{b} \in \mathbb{R}^m$: mesures / observations connues
- $\mathbf{A} \in \mathbb{R}^{m \times n}$: matrice de mélange connue
- $\mathbf{x} \in \mathbb{R}^n$: vecteur à estimer
- $\boldsymbol{\eta} \in \mathbb{R}^m$: bruit de mesure aléatoire

Vraisemblance

Modèle blanc et gaussien du bruit

$$\boldsymbol{\eta} \sim \mathcal{N}\left(0, \lambda^{-1} \mathbf{I}_m\right)$$

λ : précision pouvant être estimée

Vraisemblance (*likelihood*)

$$\mathbf{b} \sim \mathcal{N}\left(\mathbf{A}\mathbf{x}, \lambda^{-1} \mathbf{I}_m\right)$$

$$p(\mathbf{b} \mid \mathbf{x}, \lambda) \propto \lambda^{m/2} \exp\left(-\frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2\right) \quad (2)$$

[2012] : $\lambda^{n/2}$

[2020] : $\lambda^{m/2}$ (correct)

A priori (*prior*) 1/2

Sur \mathbf{x} : distribution gaussienne

$$p(\mathbf{x} \mid \delta) \propto \delta^{n/2} \exp\left(-\frac{\delta}{2} \mathbf{x}^t \mathbf{C} \mathbf{x}\right) \quad (3)$$

Matrice de régularisation \mathbf{C}

Soient \mathbf{x}_{∂_i} regroupant les n_i pixels voisin de x_i , et \bar{x}_{∂_i} leur moyenne.

$$x_i \mid \mathbf{x}_{\partial_i} \sim \mathcal{N}(\bar{x}_{\partial_i}, 1/n_i)$$

On obtient:

$$\mathbf{C}_{ij} = \begin{cases} n_i & i = j, \\ -1 & j \in \partial_i, \\ 0 & \text{sinon.} \end{cases}$$

A priori (*prior*) 2/2

Sur λ : distribution Gamma

$$p(\lambda) \propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda) \quad (4)$$

Sur δ : distribution Gamma

$$p(\delta) \propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta) \quad (5)$$

Pourquoi une Gamma ?

- 1 Conjuguée gaussienne \rightarrow Conditionnelles simples
- 2 Non informative pour $\alpha = 1$ et $\beta = 10^{-4}$

Modèle hiérarchique

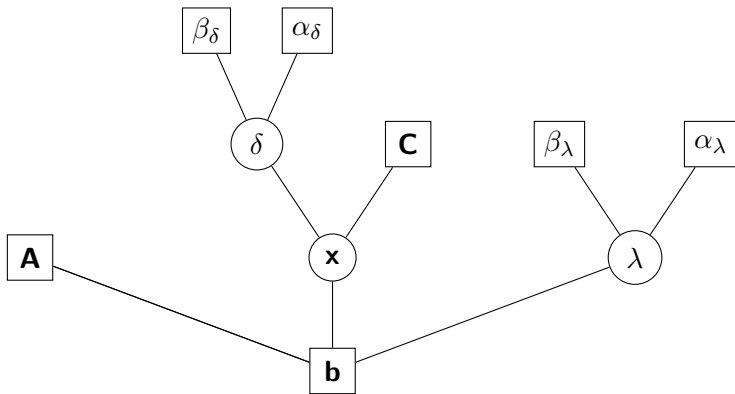


Figure 3: Les paramètres entourés sont estimés, les encadrés sont connus

A posteriori

Théorème de Bayes

$$\begin{aligned} p(\mathbf{x}, \lambda, \delta | \mathbf{b}) &= \frac{p(\mathbf{b} | \mathbf{x}, \lambda, \delta) p(\mathbf{x}, \lambda, \delta)}{p(\mathbf{b})} \\ &= \frac{p(\mathbf{b} | \mathbf{x}, \lambda) p(\mathbf{x} | \delta) p(\lambda) p(\delta)}{p(\mathbf{b})} \\ &\propto p(\mathbf{b} | \mathbf{x}, \lambda) p(\mathbf{x} | \delta) p(\lambda) p(\delta) \end{aligned}$$

Densité a posteriori

$$\begin{aligned} p(\mathbf{x}, \lambda, \delta | \mathbf{b}) &\propto \lambda^{m/2+\alpha_\lambda-1} \delta^{n/2+\alpha_\delta-1} \\ &\exp\left(-\frac{\lambda}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 - \frac{\delta}{2} \mathbf{x}^t \mathbf{Cx} - \beta_\lambda \lambda - \beta_\delta \delta\right) \quad (6) \end{aligned}$$

Conditionnelle

$$p(\mathbf{x}, \lambda, \delta \mid \mathbf{b}) \propto \lambda^{m/2+\alpha_\lambda-1} \delta^{n/2+\alpha_\delta-1} \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 - \frac{\delta}{2}\mathbf{x}^t\mathbf{C}\mathbf{x} - \beta_\lambda\lambda - \beta_\delta\delta\right)$$

Trois conditionnelles

$$\mathbf{x} \mid \lambda, \delta, \mathbf{b} \sim \mathcal{N}\left(\mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}, \mathbf{P}^{-1}\right) \quad (10)$$

$$\lambda \mid \mathbf{x}, \delta, \mathbf{b} \sim \Gamma\left(\frac{m}{2} + \alpha_\lambda, \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \beta_\lambda\right) \quad (11)$$

$$\delta \mid \mathbf{x}, \lambda, \mathbf{b} \sim \Gamma\left(\frac{n}{2} + \alpha_\delta, \frac{1}{2}\mathbf{x}^t\mathbf{C}\mathbf{x} + \beta_\delta\right) \quad (12)$$

Conditionnelle de \mathbf{x}

$$p(\mathbf{x} \mid \lambda, \delta, \mathbf{b})$$

$$\propto \exp\left(-\frac{\lambda}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2 - \frac{\delta}{2}\mathbf{x}^t\mathbf{Cx}\right)$$

$$\propto \exp\left(-\frac{\lambda}{2}\mathbf{x}^t\mathbf{A}^t\mathbf{Ax} + \frac{\lambda}{2}\mathbf{x}^t\mathbf{A}^t\mathbf{b} + \frac{\lambda}{2}\mathbf{b}^t\mathbf{Ax} - \frac{\lambda}{2}\mathbf{b}^t\mathbf{b} - \frac{\delta}{2}\mathbf{x}^t\mathbf{Cx}\right)$$

$$\propto \exp -\frac{1}{2}\left[\mathbf{x}^t(\lambda\mathbf{A}^t\mathbf{A} + \delta\mathbf{C})\mathbf{x} - 2\mathbf{x}^t(\lambda\mathbf{A}^t\mathbf{A} + \delta\mathbf{C})(\lambda\mathbf{A}^t\mathbf{A} + \delta\mathbf{C})^{-1}\lambda\mathbf{A}^t\mathbf{b}\right]$$

$$\propto \exp -\frac{1}{2}\left[\mathbf{x}^t\mathbf{Px} - 2\mathbf{x}^t\mathbf{PP}^{-1}\lambda\mathbf{A}^t\mathbf{b}\right]$$

$$\propto \exp -\frac{1}{2}\left[\left(\mathbf{x} - \mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}\right)^t \mathbf{P} \left(\mathbf{x} - \mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}\right)\right]$$

D'où:

$$\mathbf{x} \mid \lambda, \delta, \mathbf{b} \sim \mathcal{N}\left(\mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}, \mathbf{P}^{-1}\right) \quad (10)$$

[2020] \mathbf{P} est inversible en supposant $\ker(\mathbf{A}) \cap \ker(\mathbf{C}) = \{\mathbf{0}\}$.

Méthode de Monte-Carlo par chaînes de Markov

```
k = 0;  
Initialiser  $\lambda_0$  et  $\delta_0$ ;  
while  $k < K$  do  
  | Étape 1: Tirer  $\mathbf{x}^k$  selon  $\lambda_k$  et  $\delta_k$ ;  
  | Étape 2: Tirer  $\lambda_{k+1}$  selon  $\mathbf{x}^k$ ;  
  | Étape 3: Tirer  $\delta_{k+1}$  selon  $\mathbf{x}^k$ ;  
  |  $k \leftarrow k + 1$ ;  
end
```

Algorithm 1: MCMC

Formulation équivalente de l'étape 1 du MCMC

Gibbs

Les x_i sont tirés tour-à-tour selon leurs conditionnelles.

Conditionnelle des x_i

Une composante d'un vecteur gaussien est gaussien (la réciproque est fausse). Il suffit donc d'obtenir la moyenne et la variance de

$$x \mid \mathbf{x}_{-i}, \lambda, \delta$$

Notations utilisées

- \mathbf{P}_{-i} la matrice $\lambda \mathbf{A}^t \mathbf{A} + \delta \mathbf{C}$ sans la colonne i
- P_{ij} le coefficient (i, j) de la matrice $\lambda \mathbf{A}^t \mathbf{A} + \delta \mathbf{C}$
- \mathbf{x}_{-i} le vecteur \mathbf{x} sans la composante i

Formulation équivalente de l'étape 1 du MCMC

$$\mathbf{x} \sim \mathcal{N}(\mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}, \mathbf{P}^{-1})$$

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{x} - \mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b})^t \mathbf{P} (\mathbf{x} - \mathbf{P}^{-1}\lambda\mathbf{A}^t\mathbf{b}) \\ &= \mathbf{x}^t \mathbf{P} \mathbf{x} - 2\mathbf{x}^t \lambda \mathbf{A}^t \mathbf{b} + c \end{aligned}$$

$$= [x_1 \cdots x_i \cdots x_N] \begin{bmatrix} \sum_{j=1}^N P_{1j}x_j \\ \cdots \\ \sum_{j=1}^N P_{ij}x_j \\ \cdots \\ \sum_{j=1}^N P_{Nj}x_j \end{bmatrix} - 2\lambda \begin{bmatrix} \sum_{j=1}^N A_{1j}x_j \\ \cdots \\ \sum_{j=1}^N A_{ij}x_j \\ \cdots \\ \sum_{j=1}^N A_{Nj}x_j \end{bmatrix}^t \mathbf{b} + c$$

$$f(x_i) \sim x_i \sum_{j=1}^N P_{ij}x_j + x_{-i}^t [P_{ki}x_i]_{\substack{1 \leq k \leq N \\ k \neq i}} - 2\lambda [A_{ki}x_i]_{1 \leq k \leq N}^t \mathbf{b}$$

Formulation équivalente de l'étape 1 du MCMC

$$\begin{aligned}
 f(x_i) &\sim x_i \sum_{j=1}^N P_{ij} x_j + x_{-i}^t [P_{ki} x_i]_{\substack{1 \leq k \leq N \\ k \neq i}} - 2\lambda [A_{ki} x_i]_{1 \leq k \leq N}^t \mathbf{b} \\
 &= x_i P_{ii} x_i + x_i \sum_{\substack{k=1 \\ k \neq i}}^N P_{ik} x_k + \sum_{\substack{k=1 \\ k \neq i}}^N x_k P_{ki} x_i - 2\lambda x_i [A_{ki}]_{1 \leq k \leq N}^t \mathbf{b} \\
 &\sim P_{ii} x_i^2 + 2 \frac{P_{ii}}{P_{ii}} x_i (\mathbf{P}_{-i} \mathbf{x}_{-i} - \lambda [\mathbf{A}^t \mathbf{b}]_i) \\
 &\sim P_{ii} x_i^2 - 2 P_{ii} x_i \frac{\lambda [\mathbf{A}^t \mathbf{b}]_i - \mathbf{P}_{-i} \mathbf{x}_{-i}}{P_{ii}}
 \end{aligned}$$

Exposant d'une gaussienne univariée

$$g(x) = \gamma(x - \mu)^2 = \gamma(x^2 - 2x\mu + \mu^2) \sim \gamma x^2 - 2\gamma x \mu$$

Formulation équivalente de l'étape 1 du MCMC

Étape 1 du MCMC

Tirer \mathbf{x}^k selon:

$$\mathbf{x}^k \mid \lambda_k, \delta_k \sim \mathcal{N} \left(\mathbf{P}_k^{-1} \lambda_k \mathbf{A}^t \mathbf{b}, \mathbf{P}_k^{-1} \right)$$

Étape 1 du MCMC avec une boucle de Gibbs

Tirer tour-à-tour les composantes x_i^k selon:

$$x_i^k \mid \mathbf{x}_{-i}, \lambda_k, \delta_k \sim \mathcal{N} \left(\frac{\lambda [\mathbf{A}^t \mathbf{b}]_i - \mathbf{P}_{-i} \mathbf{x}_{-i}}{P_{ii}}, P_{ii} \right)$$

Avec $\mathbf{P} = \lambda \mathbf{A} \mathbf{A}^t + \delta \mathbf{C}$.

Ajout de la contrainte de positivité

$$p(x \mid \mathbf{x}_{-i}, \lambda_k, \delta_k) \propto \exp \left[-\frac{P_{ii}}{2} \left(x - \frac{\lambda[\mathbf{A}^t \mathbf{b}]_i - \mathbf{P}_{-i} \mathbf{x}_{-i}}{P_{ii}} \right)^2 \right]$$

Ajout de la contrainte

Tirer tour-à-tour les x_i^k selon:

$$p_+(x \mid \mathbf{x}_{-i}^k, \lambda_k, \delta_k) \propto p(x \mid \mathbf{x}_{-i}^k, \lambda_k, \delta_k) \mathbb{1}_+(x) \quad (15)$$

Échantillonner une loi tronquée univariée

Par rejet: en tirant parti de la proportionnalité

\mathcal{L} := un échantillonneur de la loi d'origine;

$x = -1$;

while $x < 0$ **do**

 | Tirer une proposition x selon \mathcal{L} ;

end

Algorithm 2: Procédure par rejet

(Dés)avantages

- Se généralise pour n'importe quelle dimension
- Coûteux selon l'importance de la troncature
Taux de rejet = aire de la partie tronquée de la densité

Échantillonner une loi tronquée univariée

Par rejet: en tirant parti de la proportionnalité

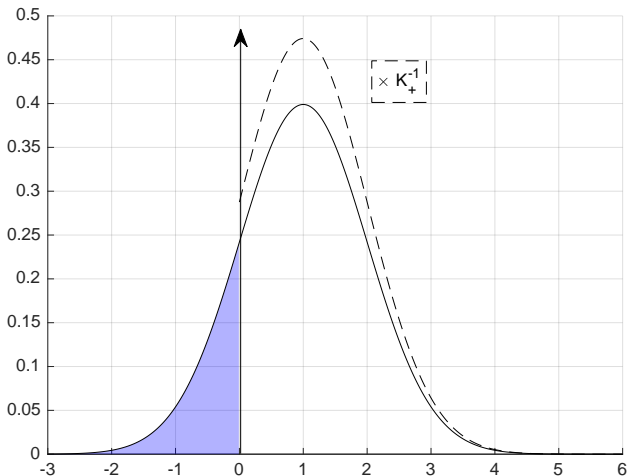


Figure 4: Rejet de la masse bleue → rehaussement de la partie positive

Échantillonner une loi tronquée univariée

Inversion de la fonction de répartition

Explication

On cherche T tel que $X = T(U)$ avec $U \sim \mathcal{U}([0, 1])$

$$\begin{aligned}\phi(x) &= \Pr(X \leq x) \\ &= \Pr(T(U) \leq x) \\ &= \Pr(U \leq T^{-1}(x)) \\ &= T^{-1}(x) \\ T &= \phi^{-1}\end{aligned}$$

Cas gaussien tronqué

$$\phi_+^{-1}(u) = \mu + \sqrt{\frac{2}{\gamma}} \operatorname{erfinv} \left\{ (u - 1) \operatorname{erf} \left(\sqrt{\frac{\gamma}{2}} \mu \right) + u \right\} \quad (16)$$

Échantillonner une loi tronquée univariée

Démonstration

Notations

$$p_+(x) = K_+^{-1} p(x) \mathbb{1}_+(x)$$

$$\phi_+(t) = \Pr(X_+ \leq t)$$

$$\phi(t) = \Pr(X \leq t)$$

$$\begin{aligned} K_+ &= \int_{-\infty}^{+\infty} p(x) \mathbb{1}_+(x) dx \\ &= \int_0^{+\infty} p(x) \mathbb{1}_+(x) dx \\ &= 1 - \phi(0) \end{aligned}$$

Échantillonner une loi tronquée univariée

Démonstration

$$\begin{aligned}\phi_+(t) &= \int_{-\infty}^t p_+(x) dx \\ &= K_+^{-1} \left[\int_{-\infty}^t p(x) dx - \int_{-\infty}^0 p(x) du \right] \\ &= K_+^{-1} [\phi(t) - \phi(0)]\end{aligned}$$

$$\begin{aligned}\phi_+(t) &= u \\ t &= \phi^{-1}[K_+ u + \phi(0)] \\ \phi_+^{-1}(u) &= \phi^{-1}[(1 - u)\phi(0) + u]\end{aligned}$$

Récap. sur la contrainte de positivité

Points clefs

- Modification du MCMC avec une étape de Gibbs
- Ajout de la contrainte sur les conditionnelles $x_i \mid \mathbf{x}_{-i}, \lambda, \delta$
- Échantillonnage par inversion de la fonction de répartition

Mais toujours pas de proportion non nulle de pixels nuls...

L'étape 1 du MCMC comme un système linéaire

Étape 1 du MCMC

On pose $\mathbf{P}_k = \lambda_k \mathbf{A}^t \mathbf{A} + \delta_k \mathbf{C}$ (diagonalisable avec Fourier).

Étape 1: $\mathbf{x}^k \sim \mathcal{N}(\mathbf{P}_k^{-1} \lambda_k \mathbf{A}^t \mathbf{b}, \mathbf{P}_k^{-1})$

Système à résoudre

Trouver \mathbf{x}^k solution de:

$$\mathbf{P}_k \mathbf{x}^k = \lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k \quad (13)$$

où $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{P}_k)$

L'étape 1 du MCMC comme un système linéaire

$$\mathbf{x}^k \sim \mathcal{N}(\mathbf{P}_k^{-1} \lambda_k \mathbf{A}^t \mathbf{b}, \mathbf{P}_k^{-1}) \quad (10)$$

$$\mathbf{P}_k \mathbf{x}^k = \lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k \quad (13)$$

\mathbf{w}_k est un vecteur gaussien: moyenne et covariance ?

$$\begin{aligned} \mathbb{E}[\mathbf{P}_k \mathbf{x}^k] &= \mathbf{P}_k \mathbb{E}[\mathbf{x}^k] = \mathbf{P}_k \mathbf{P}_k^{-1} \lambda_k \mathbf{A}^t \mathbf{b} = \lambda_k \mathbf{A}^t \mathbf{b} \\ &= \mathbb{E}[\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k] \end{aligned}$$

$$\begin{aligned} \mathbb{V}[\mathbf{P}_k \mathbf{x}^k] &= \mathbf{P}_k^t \mathbb{V}[\mathbf{x}^k] \mathbf{P}_k = \mathbf{P}_k^t \mathbf{P}_k^{-1} \mathbf{P}_k = \mathbf{P}_k^t = \mathbf{P}_k \\ &= \mathbb{V}[\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k] = \mathbb{V}[\mathbf{w}_k] \end{aligned}$$

$$\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{P}_k)$$

Un problème d'optimisation: *Perturbation Optimisation*

Formulation du problème d'optimisation

$$\mathbf{x}^k = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - \mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \right\} \quad (14)$$

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - \mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k)$$

$$f'(\mathbf{x}) = \mathbf{P}_k \mathbf{x} - (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k)$$

$$f''(\mathbf{x}) = \mathbf{P}_k : \text{semi-définie positive}$$

$$f'(\mathbf{x}) = 0 \Leftrightarrow \mathbf{P}_k \mathbf{x} = \lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k \quad (13)$$

Ajout de la contrainte de non négativité

$$\mathbf{x}_{CON}^k = \arg \min_{\mathbf{x} \geq 0} \left\{ \frac{1}{2} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - \mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \right\} \quad (23)$$

Pourquoi cela fonctionne ?

En optimisation, on peut toucher la bordure: avoir une contrainte active ($x_i^k = 0$), contrairement aux lois de probabilités continues.

Lien avec l'échantillonnage sans contrainte

Minimisation d'une norme

$$\mathbf{x}_{CON}^k = \arg \min_{\mathbf{x} \geq 0} \|\mathbf{x} - \mathbf{x}_{UC}^k\|_{\mathbf{P}_k}^2 \quad (22)$$

$$\|\square\|_{\mathbf{P}_k}^2 = \square^t \mathbf{P}_k \square$$

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x} \geq 0} \|\mathbf{x} - \mathbf{x}_{UC}^k\|_{\mathbf{P}_k}^2 \\ &= \arg \min_{\mathbf{x} \geq 0} (\mathbf{x} - \mathbf{x}_{UC}^k)^t \mathbf{P}_k (\mathbf{x} - \mathbf{x}_{UC}^k) \\ &= \arg \min_{\mathbf{x} \geq 0} [\mathbf{x}^t - (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k)^t \mathbf{P}_k^{-t}] \mathbf{P}_k [\mathbf{x} - \mathbf{P}_k^{-1} (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k)] \\ &= \arg \min_{\mathbf{x} \geq 0} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - 2\mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \\ &\quad + (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k)^t \mathbf{P}_k^{-1} (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \end{aligned}$$

Lien avec l'échantillonnage sans contrainte

$$\begin{aligned}\mathbf{x}^* &= \arg \min_{\mathbf{x} \geq 0} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - 2\mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \\ &= \arg \min_{\mathbf{x} \geq 0} \frac{1}{2} \mathbf{x}^t \mathbf{P}_k \mathbf{x} - \mathbf{x}^t (\lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k) \\ \mathbf{x}_{CON}^k &= \arg \min_{\mathbf{x} \geq 0} \|\mathbf{x} - \mathbf{x}_{UC}^k\|_{\mathbf{P}_k}^2\end{aligned}\tag{22}$$

Dégénérescence de la matrice de régularisation [2012]

A priori de \mathbf{x} sans contrainte

$$p(\mathbf{x} \mid \delta) \propto \delta^{n/2} \exp\left(-\frac{\delta}{2} \mathbf{x}^t \mathbf{C} \mathbf{x}\right) \quad (3)$$

Contrainte de non négativité

\mathbf{D}_x : matrice diagonale $[\mathbf{D}_x]_{ii} = \begin{cases} 1 & \text{si } x_i > 0 \\ 0 & \text{si } x_i = 0 \end{cases}$

$$\frac{\delta}{2} \mathbf{x}^t \mathbf{C} \mathbf{x} = \frac{\delta}{2} \mathbf{x}^t (\mathbf{D}_x \mathbf{C} \mathbf{D}_x) \mathbf{x}$$

$$p(\mathbf{x} \mid \delta) \propto \delta^{n_p/2} \exp\left(-\frac{\delta}{2} \mathbf{x}^t \mathbf{C} \mathbf{x}\right) \quad (18)$$

n_p est le nombre d'éléments positifs de \mathbf{x} .

Dégénérescence de la matrice de régularisation [2020]

Définitions

$$\mathcal{A}(\mathbf{x}) = \{i \mid x_i = 0\}$$

$$\mathbf{x}_{\mathcal{A}} = \mathbf{D}_{\mathcal{A}}\mathbf{x}$$

$$\mathbf{x}_{\mathcal{A}} = \mathbf{0}$$

$$\mathcal{I}(\mathbf{x}) = \{i \mid x_i > 0\}$$

$$\mathbf{x}_{\mathcal{I}} = \mathbf{D}_{\mathcal{I}}\mathbf{x}$$

$$\mathbf{x}_{\mathcal{I}} > \mathbf{0}$$

$\mathbf{D}_{\mathcal{I}}$ est la matrice identité I_N sans les lignes d'indices i de $\mathcal{A}(\mathbf{x})$.
Idem pour $\mathbf{D}_{\mathcal{A}}$ avec les indices de $\mathcal{I}(\mathbf{x})$...

L'ensemble \mathcal{I} était implicite en 2012

$$p(\mathbf{x}, \mathcal{I} \mid \mathbf{b}, \lambda, \delta) = p(\mathbf{x} \mid \mathbf{b}, \lambda, \delta, \mathcal{I})p(\mathcal{I} \mid \mathbf{b}, \lambda, \delta) \quad (17)$$

Dégénérescence de la matrice de régularisation [2020]

$$\begin{aligned}
 p(\mathbf{x} \mid \mathbf{b}, \lambda, \delta, \mathcal{I}) &\propto \mathbb{1}_{\Omega}(\mathbf{x}) p(\mathbf{b} \mid \mathbf{x}, \lambda, \mathcal{I}) p(\mathbf{x} \mid \delta, \mathcal{I}) \\
 &\propto \mathbb{1}_{\Omega}(\mathbf{x}) \exp\left(-\frac{\lambda}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 - \frac{\delta}{2} \mathbf{x}^t \mathbf{C}\mathbf{x}\right) \delta_0(\mathbf{D}_A \mathbf{x}) \\
 &\propto \mathbb{1}_{\Omega_{\mathcal{I}}}(\mathbf{x}_{\mathcal{I}}) \exp\left(-\frac{\lambda}{2} \|\mathbf{A}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}} - \mathbf{b}\|_2^2 - \frac{\delta}{2} \mathbf{x}_{\mathcal{I}}^t \mathbf{C}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}\right) \delta_0(\mathbf{D}_A \mathbf{x})
 \end{aligned}$$

On extrait $\mathbf{x} \mid \delta, \mathcal{I}$ et on calcule la constante de normalisation.

$$\begin{aligned}
 p(\mathbf{x} \mid \mathbf{b}, \lambda, \delta, \mathcal{I}) &\propto \exp\left(-\frac{\lambda}{2} \mathbf{x}_{\mathcal{I}}^t \mathbf{C}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}\right) \delta_0(\mathbf{D}_A \mathbf{x}) \\
 \int p(\mathbf{x} \mid \delta, \mathcal{I}) &= \sqrt{\frac{(2\pi)^{n_p}}{\det(\delta \mathbf{C}_{\mathcal{I}})}} \propto \delta^{n_p/2}
 \end{aligned}$$

Dégénérescence de la matrice de régularisation [2020]

A priori de \mathbf{x} avec contrainte

$$p(\mathbf{x} \mid \delta, \mathcal{I}) \propto \delta^{n_p/2} \exp\left(-\frac{\delta}{2} \mathbf{x}_{\mathcal{I}}^t \mathbf{C}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}}\right) \quad (18)$$

n_p est le nombre d'éléments (strictement positifs) de $\mathbf{x}_{\mathcal{I}}$.

Conditionnelle de δ (Étape 3 MCMC)

$$\delta \mid \mathbf{x}, \mathbf{b}, \mathcal{I} \sim \Gamma\left(\frac{n_p}{2} + \alpha_\delta, \frac{1}{2} \mathbf{x}_{\mathcal{I}}^t \mathbf{C}_{\mathcal{I}} \mathbf{x}_{\mathcal{I}} + \beta_\delta\right) \quad (21)$$

Les auteurs déclarent avoir de meilleurs intervalles de confiance après la correction en rouge (en 2012 et en 2020).

En 2012: $\mathbf{D}_{\mathcal{I}} = I_N$, même performance (plus coûteux en calculs).

Un algorithme d'optimisation: GPCG

Gradient Projection – Conjugate Gradient (GPCG)

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \geq 0} \left\{ q(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t \mathbf{B} \mathbf{x} - \mathbf{x}^t \mathbf{c} \right\} \quad (24)$$

$\mathbf{B} = \mathbf{P}_k = \lambda_k \mathbf{A}^t \mathbf{A} + \delta_k \mathbf{C}$: semi-définie positive

$$\mathbf{c} = \lambda_k \mathbf{A}^t \mathbf{b} + \mathbf{w}_k$$

Introduction d'un opérateur maximum avec 0 sur chaque composante $\max(0, x_j)$ dans les itérations.

- Algorithme de Moré et Torello (1991)
- Variante de *Conjugate Gradient*
- Grande dimension: nombre important de contraintes actives ajoutées ou supprimées à chaque itération

Expériences réalisées

- 3 expériences: reconstruction d'images en 1D, 2D, et tomographie
- 20 chaînes MCMC avec des départs aléatoires
- Critère d'arrêt selon des distances inter/intra-chaînes

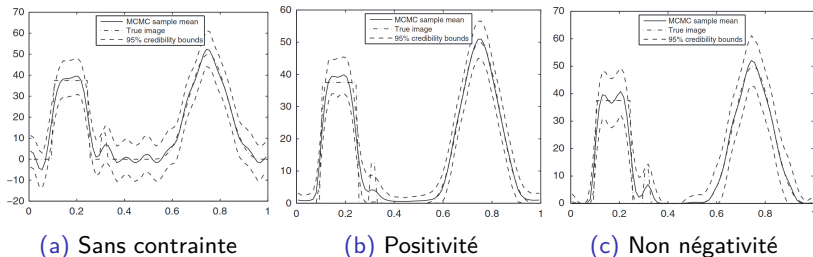


Figure 5: Expérience 1D

Résultats

1D	Vrai	UC	PC	NNC
λ	5.35	[3.94, 8.35]	[1.48, 3.73]	[3.61, 7.28]
Temps (sec)	×	36	52	49
2D	Vrai	UC	PC	NNC
λ	8.09	[7.59, 8.63]	×	[7.66, 8.14]
Temps (sec)	×	30	×	2 845
Tomographie	Vrai	UC	PC	NNC
λ	13.24	[13.85, 15.47]	×	[12.61, 13.53]
Temps (sec)	×	2 637	×	8 301

Conclusion

Points clefs

- Mise en place de contrainte de positivité / non négativité à partir d'un MCMC sans contrainte
- Estimation d'intervalle de confiance
- Positivité: avec un Gibbs
- Non négativité: avec de l'optimisation
- Meilleurs résultats avec la non négativité

Références

Article de 2020

J. M. Bardsley and P.C. Hansen. *MCMC Algorithms for Computational UQ of Nonnegativity Constrained Linear Inverse Problems*. SIAM Journal on Scientific Computing, 42(2):A1269-A1288, 2020. DOI: 10.1137/18M1234588

Article de 2012

J. M. Bardsley and C. Fox. *An MCMC Method for Uncertainty Quantification in Nonnegativity Constrained Inverse Problems*. Inverse Problems in Science and Engineering, vol. 20, no. 4, pp. 477-498, 2012. DOI: 17415977.2011.637208