

Un algorithme MCMC distribué pour la résolution de problèmes inverses de grande dimension

P.-A. Thouvenin*, A. Repetti†‡ et P. Chainais*

GRETSI'22

* Université de Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France [†]Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom [‡]Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom.

1. Contexte et enjeux

Problème inverse : estimer $\boldsymbol{x} \in \mathbb{R}^N$ à partir d'observations $\boldsymbol{y} \in \mathbb{R}^M$

$$y = \mathcal{D}(Ax) \leadsto \pi(x \mid y) \propto \exp(-f_{\mathbf{v}}(Ax) - g(Bx))$$

a priori sur **x**

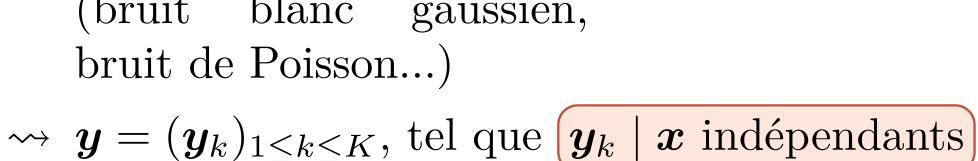
 $oldsymbol{A} \in \mathbb{R}^{M imes N}, \ \mathcal{D} : \mathbb{R}^M
ightarrow \mathbb{R}^M$ $f_{oldsymbol{y}} \colon \mathbb{R}^M o]-\infty, +\infty]$ $g \colon \mathbb{R}^P \to]-\infty, +\infty], \ \boldsymbol{B} \in \mathbb{R}^{P imes N}$ modèles d'acquisition et de bruit attache aux données

- Problèmes de grande dimension : $10^3 \le N \le 10^8$, $M \approx N$
- \Rightarrow décomposer le problème sur K noeuds + limiter les communications
- + quantification d'incertitude \Rightarrow algorithmes MCMC
- Architectures distribuées : client serveur vs. SPMD [1]

Client-serveur Single Program Multiple Data comms: goulet d'étranglement ✓ comms : entre noeuds *voisins* ▲ ajustement charge des noeuds? ✓ localité des données Worker 1 Worker 2 Worker 3

2. Structure du problème et approche proposée

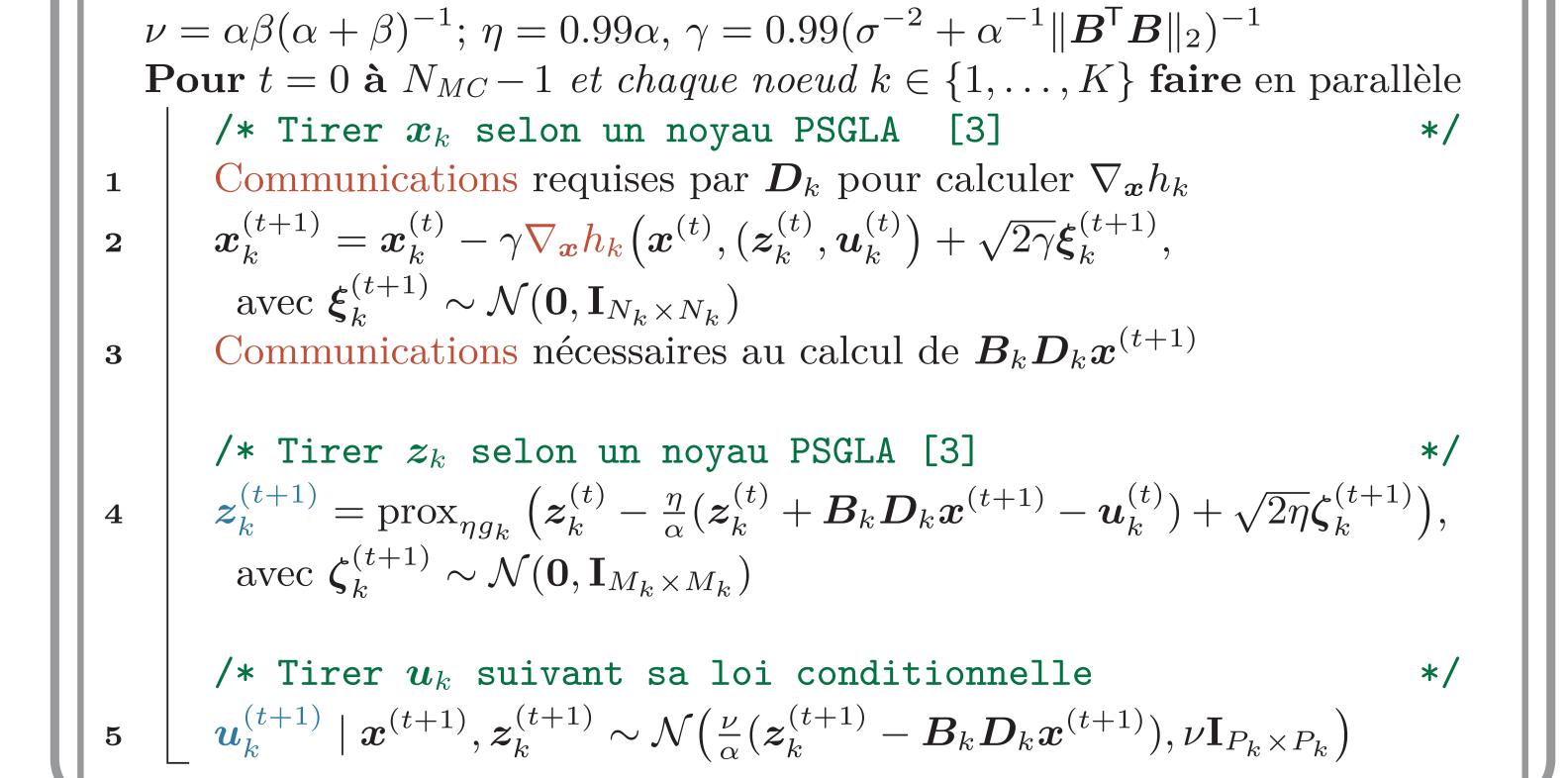
- Structure considérée
- 1. (A, B parcimonieuses)par blocs (convolution, laplacien, masque...)
- \leadsto couplage localisé des x_n recouvrement $\ll \lfloor N/K \rfloor$
- 2. $\mathcal{D}(\boldsymbol{z}) = (\mathcal{D}_k(\boldsymbol{z}_k))_{1 \leq k \leq K}$ (bruit blanc gaussien, bruit de Poisson...)



- 3. éclatement par augmentation approchée AXDA [2]. Par exemple :

$$\widetilde{\pi}_{\alpha}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{u} \mid \boldsymbol{y}, \beta) \propto \exp\left(-f_{\boldsymbol{y}}(\boldsymbol{A}\boldsymbol{x}) - g(\boldsymbol{z}) - \frac{1}{2\alpha} \|\boldsymbol{B}\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{u}\|_{2}^{2} - \frac{1}{2\beta} \|\boldsymbol{u}\|_{2}^{2}\right)$$

- Formulation du problème
- $f_{\boldsymbol{y}}(\boldsymbol{A}\boldsymbol{x}) = \sum_{k=1}^{\infty} f_{\boldsymbol{y}_k}(\boldsymbol{A}_k \boldsymbol{C}_k \boldsymbol{x}), \, \boldsymbol{C}_k \in \mathbb{R}^{\widetilde{N}_k \times N} \text{ sélection, } \boldsymbol{A}_k \in \mathbb{R}^{M_k \times \widetilde{N}_k}$
- idem, g choisie telle que : $g(\boldsymbol{B}\boldsymbol{x}) = \sum g_k(\boldsymbol{B}_k\boldsymbol{D}_k\boldsymbol{x})$
- Approche proposée
 - 1. éclatement approprié basé sur AXDA [2]
 - 2. algo. MCMC distribué avec architecture SPMD
 - noyaux de transition : Gibbs ou PSGLA [3]
 - \rightarrow exemple: inpainting sous bruit blanc gaussien, a priori TV [4]



3. Expériences sur données synthétiques

- Cadre des expériences
 - problème : M = |0.6N|, SNR = 40 dB, $(\alpha, \beta, \tau) = (9, 1, 0.2)$
 - * passage à l'échelle : $N = 512 \times 512, K \in \{1, 2, 4, 8, 16\};$
 - * inpainting en grande dimension: $N=1024\times 1024,\, K=16.$
 - $-N_{\rm MC}=10^4$ échantillons, dont $N_{\rm bi}=5\times10^3$ burn-in;
 - comparaison avec l'échantillonneur [2], basé sur AXDA.
- Résultats
- estimateur de qualité comparable à [2];
- bonne performance de parallélisation (accélération $\approx K$);
- inpainting en grande dimension:
 - * 4h nécessaires avec [2] (1.35 s / itér.)
 - * 90s pour l'approche proposée (K = 16: 17.93 ms / itér).

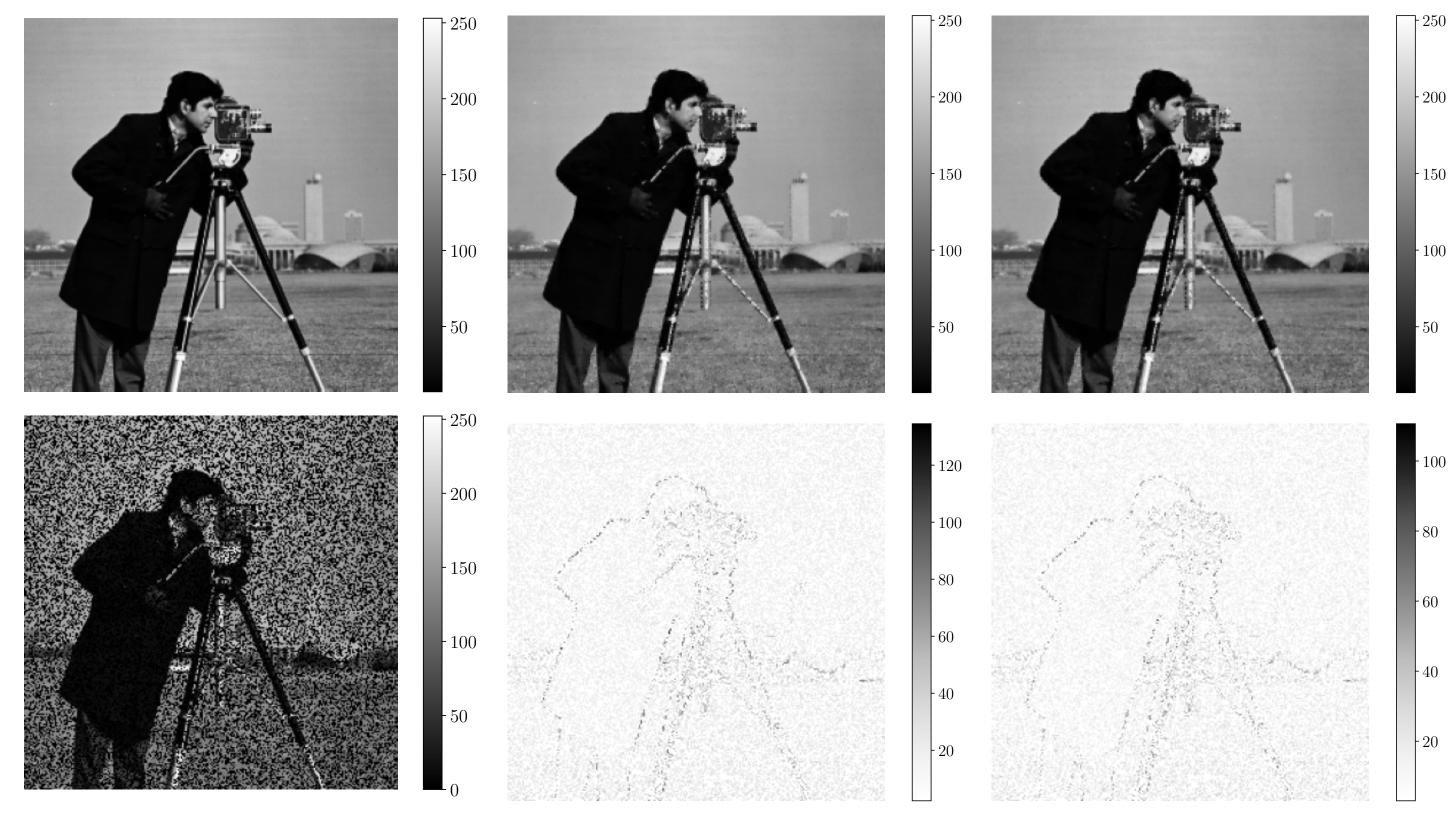


Figure 1: Ligne 1: vérité terrain, estimateur MMSE pour [2] et l'approche proposée. Ligne 2: observations, intervalles de crédibilité à 95% pour [2] et l'approche proposée.

K	SNR (MMSE)	SNR (MAP)	Temps / iter. $(\times 10^{-3} \text{ s})$	Accélération	Temps total (s)
1 (algo. [2])	23.33	22.45	65.56 (2.08)	0.19	262.20
1	23.45	22.95	$12.21 \ (0.63)$	1.00	61.04
2	23.46	22.88	6.07(0.42)	2.01	30.37
4	23.48	22.88	3.50(0.21)	3.49	17.50
8	23.44	22.86	1.93(0.77)	$\boldsymbol{6.33}$	9.63
16	23.48	22.90	1.08(2.35)	11.30	5.38

4. Conclusions et perspectives

- Conclusions: échantillonneur SPMD-distribué (PSGLA within Gibbs)
- ✓ qualité d'estimation comparable à [2];
- forte réduction du temps d'inférence + flexibilité de parallélisation;
- ✓ passage à l'échelle (strong scaling).
- Perspectives
- \rightarrow estimation des hyper-paramètres α, β ;
- → inférence pour des problèmes inverses sur hypergraphes;
- → extension de l'approche : communications asynchrones.
- F. Darema, "The SPMD model: Past, present and future," in Recent Advances in Parallel Virtual Machine and Message Passing Interface, Y. Cotronis and J. Dongarra, Eds., Berlin, Heidelberg, 2001, pp. 1–1, ISBN: 978-3-540-45417-5.
- M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler application to large-scale inference problems," IEEE Trans. Signal Process., vol. 67, no. 6, pp. 1648–1661, 2019.
- A. Salim and P. Richtàrik, "Primal dual interpretation of the proximal stochastic gradient langevin algorithm," in Adv. in Neural Information Processing Systems, vol. 33, 2020, pp. 3786–3796.
- A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," J. Math. Imag. Vision, vol. 40, no. 1, pp. 120–145, 2011.





Travail soutenu par la Chaire IA Sherlock ANR-20-CHIA-0031-01 portée par P. Chainais, par le programme national d'investissement d'avenir ANR-16-IDEX-0004 ULNE et la Région Hauts-de-France.