

centralelille

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

P.-A. Thouvenin^{*}, A. Repetti^{\dagger ‡} et P. Chainais^{*}

* 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	3. Expéri
Problème inverse : estimer $\boldsymbol{x} \in \mathbb{R}^N$ à partir d'observations $\boldsymbol{y} \in \mathbb{R}^M$	• Cadre des expérie
$oldsymbol{y} = \mathcal{D}(oldsymbol{A}oldsymbol{x}) \rightsquigarrow \pi(oldsymbol{x} \mid oldsymbol{y}) \propto \exp\left(-f_{oldsymbol{y}}(oldsymbol{A}oldsymbol{x}) - g(oldsymbol{B}oldsymbol{x}) ight)$	- problème : $M = \lfloor$ * passage à l'éch
$\begin{array}{ll} \boldsymbol{A} \in \mathbb{R}^{M \times N}, \ \mathcal{D} : \mathbb{R}^{M} \to \mathbb{R}^{M} & \text{modèles d'acquisition et de bruit} \\ f_{\boldsymbol{y}} : \mathbb{R}^{M} \to] - \infty, +\infty] & \text{modèles d'acquisition et de bruit} \\ g : \mathbb{R}^{P} \to] - \infty, +\infty], \ \boldsymbol{B} \in \mathbb{R}^{P \times N} & a \ priori \ \text{sur } \boldsymbol{x} \end{array}$	* inpainting en – $N_{MC} = 10^4$ échant
Problèmes de grande dimension : $10^3 \le N \le 10^8$, $M \approx N$	 – comparaison avec • Bésultats

décomposer le problème sur K noeuds + limiter les communications + quantification d'incertitude \Rightarrow algorithmes MCMC

	3.	Expériences	sur	données	synthétiques	
lre	\mathbf{des}	expériences				

- $|0.6N|, \text{SNR} = 40 \text{ dB}, (\alpha, \beta, \tau) = (9, 1, 0.2)$
 - **helle** : $N = 512 \times 512, K \in \{1, 2, 4, 8, 16\};$
 - grande dimension: $N = 1024 \times 1024, K = 16.$
- tillons, dont $N_{bi} = 5 \times 10^3$ burn-in;
- l'échantillonneur [2], basé sur AXDA.

nesultats

estimateur de qualité comparable à [2];

Architectures distribuées : client serveur vs. SPMD [1]



2. Structure du problème et approche proposée

Structure considérée 1. **(***A*, *B* **parcimonieuses)** par blocs (convolution, laplacien, masque...)





- 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

- noyaux de transition : Gibbs ou PSGLA [3]
- \rightarrow exemple : inpainting sous bruit blanc gaussien, a priori TV [4]

l'approche proposée.

K	$\frac{SNR}{(MMSE)}$	${ m SNR}$ (MAP)	Temps / iter. (×10 ⁻³ 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)	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)
- \checkmark qualité d'estimation comparable à [2];
- **forte réduction** du temps d'inférence + flexibilité de parallélisation;
- \checkmark passage à l'échelle (strong scaling).
- Perspectives
- \rightarrow estimation des hyper-paramètres $\alpha, \beta;$
- \rightarrow inférence pour des problèmes inverses sur hypergraphes;

 $\nu = \alpha \beta (\alpha + \beta)^{-1}; \ \eta = 0.99 \alpha, \ \gamma = 0.99 (\sigma^{-2} + \alpha^{-1} \| \mathbf{B}^{\mathsf{T}} \mathbf{B} \|_2)^{-1}$ **Pour** t = 0 à $N_{MC} - 1$ et chaque noeud $k \in \{1, \ldots, K\}$ faire en parallèle /* Tirer x_k selon un noyau PSGLA [3] */ Communications requises par D_k pour calculer $\nabla_{\boldsymbol{x}} h_k$ 1 $x_{k}^{(t+1)} = x_{k}^{(t)} - \gamma \nabla_{x} h_{k} (x^{(t)}, (z_{k}^{(t)}, u_{k}^{(t)}) + \sqrt{2\gamma} \xi_{k}^{(t+1)},$ 2 avec $\boldsymbol{\xi}_{k}^{(t+1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{N_{k} \times N_{k}})$ Communications nécessaires au calcul de $B_k D_k x^{(t+1)}$ 3 /* Tirer \boldsymbol{z}_k selon un noyau PSGLA [3] $\boldsymbol{z}_{k}^{(t+1)} = \operatorname{prox}_{\eta q_{k}} \left(\boldsymbol{z}_{k}^{(t)} - \frac{\eta}{\alpha} (\boldsymbol{z}_{k}^{(t)} + \boldsymbol{B}_{k} \boldsymbol{D}_{k} \boldsymbol{x}^{(t+1)} - \boldsymbol{u}_{k}^{(t)}) + \sqrt{2\eta} \boldsymbol{\zeta}_{k}^{(t+1)} \right),$ 4 avec $\boldsymbol{\zeta}_{k}^{(t+1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{M_{k} \times M_{k}})$ /* Tirer u_k suivant sa loi conditionnelle $oldsymbol{u}_k^{(t+1)} \mid oldsymbol{x}^{(t+1)}, oldsymbol{z}_k^{(t+1)} \sim \mathcal{N}ig(rac{
u}{lpha}(oldsymbol{z}_k^{(t+1)} - oldsymbol{B}_koldsymbol{D}_koldsymbol{x}^{(t+1)}),
u \mathbf{I}_{P_k imes P_k}ig)$ 5

 \rightarrow 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 ap- $\left[2\right]$ plication 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 [3]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 $\left[4\right]$ 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-ERSITÉ 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.