GRADIENT SCAN GIBBS SAMPLER : AN EFFICIENT HIGH-DIMENSIONAL SAMPLER APPLICATION IN INVERSE PROBLEMS

F. Orieux[1], O. Féron[2] and J.-F. Giovannelli[3]

[1] Univ. Paris-Sud 11, L2S, UMR8506, 91 192 Gif-sur-Yvette, France
[2] EDF Research & Developments, 92 140 Clamart, France
[2] Univ. Paris Dauphine, Finance for Energy Market Research Centre, 75116 Paris, France
[3] Univ. Bordeaux, IMS, UMR 5218, F-33 400 Talence, France

ABSTRACT

The paper deals with Gibbs samplers that include high-dimensional conditional Gaussian distributions. It proposes an efficient algorithm that only requires a scalar Gaussian sampling. The algorithm relies on a random excursion along a random direction. It is proved to converge, *i.e.* the drawn samples are asymptotically under the target distribution. Our original motivation is in unsupervised inverse problems related to general linear observation models and their solution in a hierarchical Bayesian framework implemented through sampling algorithms. The paper provides an illustration focused on 2-D simulations and on the super-resolution problem.

Index Terms — High-dimensional sampling, Big Data, Gibbs sampling, Bayesian strategy, inverse problem.

1. INTRODUCTION

Gaussian distributions are common throughout statistical signal and image processing, nevertheless their efficient sampling may be a cumbersome problem in high-dimension and the current paper deals with this question.

Our original motivation is in inverse problems [2], especially advanced questions and hierarchical Bayesian strategies, implemented through Monte-Carlo Markov Chain and more specifically Gibbs Sampler (GS). Indeed, consider the general linear direct operator A and data y modeled by $y = Ax + n$. The unknown object x and the noise n are modeled as independent Gaussian vectors conditionally on a vector θ. This framework encompass a variety of problems by including diverse parameters in θ: unsupervised [3] and semi-blind [4] problems, by including hyperparameters and acquisition parameters; non-Gaussian models based on conditionally Gaussian ones involving latent variables (*e.g.*, location or scale mixtures of Gaussian) for edge preserving [5, 6, 3]; hierarchical models [7, 8, 9] involving labels for inversion-segmentation. In this framework, the estimation of both x and θ relies on the sampling of the joint posterior $f(x, \theta|y)$, and this is the original motivation of our work. It requires the handling of the high-dimensional conditional posterior $f(x|\theta, y)$ that is Gaussian. It is denoted $p(x)$ with mean m and precision Q and its sampling is the core question of the paper.

The difficulty is directly related to handling high-dimensional precision Q. Its diagonalization, inversion and factorization (square root, Cholesky,...) are generally unfeasible in high-dimension due to both computational cost and memory footprint. Nevertheless, such solutions are practicable in two cases: Q circulant (strategy [10, 11] relies on diagonalization computed by FFT) and Q sparse (algorithm [12, 13, 14] is based on a Cholesky decomposition and a linear system[1]). For more general cases, solutions founded on iterative optimization algorithm have been recently proposed.

1. A conjugate gradient sampler [16] relies on two properties: (i) a Gaussian distribution has Gaussian conditional distributions and (ii) a set of mutually conjugate directions w.r.t. Q is available.

2. Paper [17] (also [18, 19]) proposes a Perturbation-Optimization principle (previously applied in [7, 8]): optimization of an adequately perturbed criterion.

In both cases, generally, an exact sample is not rigorously tractable. In the former, it requires specific characteristics regarding eigenvalues of the precision matrix. In the latter, an exact optimization is needed but practically a truncation of the iterations is implemented. So, the convergence is no more guaranteed except in [20] that introduces a Metropolis step.

Our main contribution is to promote directional sampling: given a direction δ, the keystone of the advance is to exactly sample the component of x along the direction δ. This must be achieved under the appropriate conditional distribution $p(x_\delta|x_{\setminus\delta})$, where $x_{\setminus\delta}$ denotes the complementary part of x (see Sect. 2.1 for precise definition). The algorithm takes advantage of the ease of calculating the pdf for the directional component of a multivariate Gaussian. These ideas are strongly related to different existing works.

- If the direction δ is the canonical coordinate, the algorithm amounts to a pixel-by-pixel GS [5].

- The algorithm can be viewed as a specific random scan sampler [21, 22, 23].

---

[1] Another strategy [15] simultaneously updates large blocks of variables.

- A similar idea is at work in Langevin and Hybrid Monte Carlo [24, 25, 26] (also [27]): the proposal law relies on an ascent direction of the target to increase the acceptation probability. Here, the exact distribution is sampled, so the proposal is always accepted.

Convergence is guaranteed, *i.e.* samples are asymptotically distributed according to the pdf $p(\boldsymbol{x})$. The proof is valid for any direction $\boldsymbol{\delta}$, but practically, the direction is the (possibly preconditioned) gradient of $p(\boldsymbol{x})$.

Subsequently, Sect. 2 presents the proposed algorithm. Sect. 3 gives an illustration through a simple 2-D problem and an academic problem in super-resolution. Sect. 4 presents conclusions and perspectives.

2. GRADIENT SCAN SAMPLING

The objective is to generate samples from a high-dimensional Gaussian distribution $\mathcal{N}(\boldsymbol{m}, \boldsymbol{Q}^{-1})$, for $\boldsymbol{x} \in \mathbb{R}^N$ written:

$$p(\boldsymbol{x}) = (2\pi)^{-N/2}(\det \boldsymbol{Q})^{1/2} \exp\left\{-J(\boldsymbol{x})\right\}$$

with potential: $J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{m})^{\mathrm{t}}\boldsymbol{Q}(\boldsymbol{x} - \boldsymbol{m})/2$.

2.1. Preliminary results

This section gives usual definitions and results based on [28] needed to provide convergence proof and links between matrix factorization and optimization / sampling procedures.

Definition 1. *A set $\{\boldsymbol{d}_n, n = 1, \ldots, N\}$ of non-zero vectors in $\mathbb{R}^N$ such that: $\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_m = 0 \quad for\ n, m = 1, \ldots, N,\ n \neq m$ is said mutually conjugate w.r.t. $\boldsymbol{Q}$.* △

A mutually conjugate set $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N\}$ w.r.t. $\boldsymbol{Q}$ is a basis of $\mathbb{R}^N$, then, for all $\boldsymbol{x} \in \mathbb{R}^N$:

$$\boldsymbol{x} = \sum_{n=1}^{N} \alpha_n \boldsymbol{d}_n \quad \text{with} \quad \alpha_n = \frac{\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{x}}{\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_n}.$$

So, if $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{Q}^{-1})$ is a Gaussian random vector with mean $\boldsymbol{m}$ and precision $\boldsymbol{Q}$, then the α_n are also Gaussian:

$$\alpha_n \sim \mathcal{N}\left(\frac{\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{m}}{\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_n} \ ; \ \frac{1}{\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_n}\right) \tag{1}$$

and reciprocally if the α_n are under (1) then $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{Q}^{-1})$.

In particular, let $\boldsymbol{x}^{\mathrm{c}} \in \mathbb{R}^N$ be a "current" point and $\boldsymbol{d}_1 \in \mathbb{R}^N$ a given "direction". One can find $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_N$ such that $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N\}$ is mutually conjugate w.r.t. $\boldsymbol{Q}$ and $\boldsymbol{x}^{\mathrm{c}}$ writes:

$$\boldsymbol{x}^{\mathrm{c}} = \sum_{n} \alpha_n^{\mathrm{c}} \boldsymbol{d}_n$$

Consider now the one-dimensional subset

$$D_{d_1}(\boldsymbol{x}^{\mathrm{c}}) = \left\{\sum_{n=1}^{N} \alpha_n \boldsymbol{d}_n, \ \alpha_1 \in \mathbb{R}, \alpha_n = \alpha_n^{\mathrm{c}} \text{ for } n \neq 1\right\}$$

$$= \left\{\boldsymbol{x}^{\mathrm{c}} + (\alpha_1 - \alpha_1^{\mathrm{c}})\boldsymbol{d}_1, \ \alpha_1 \in \mathbb{R}\right\}$$

We are interested in the conditional pdf $p(\boldsymbol{x}|\boldsymbol{x} \in D_{d_1}(\boldsymbol{x}^{\mathrm{c}}))$.

Proposition 1. *A sample $\widetilde{\boldsymbol{x}}$ according to $p(\boldsymbol{x}|\boldsymbol{x} \in D_{d_1}(\boldsymbol{x}^{\mathrm{c}}))$ can be obtained by:*

1. sample $\widetilde{\alpha} \sim \mathcal{N}\left(\dfrac{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}(\boldsymbol{x}^{\mathrm{c}} - \boldsymbol{m})}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1} \ ; \ \dfrac{1}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1}\right)$

2. compute $\widetilde{\boldsymbol{x}} = \boldsymbol{x}^{\mathrm{c}} - \widetilde{\alpha}\,\boldsymbol{d}_1$

Proof. The point is to show that the coordinate of $\widetilde{\boldsymbol{x}}$ on the direction $\boldsymbol{d}_1$ follows the Gaussian distribution (1). To this end, let us write $\widetilde{\boldsymbol{x}}$:

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x}^{\mathrm{c}} - \widetilde{\alpha}\,\boldsymbol{d}_1 = (\alpha_1^{\mathrm{c}} - \widetilde{\alpha})\boldsymbol{d}_1 + \sum_{n=2}^{N} \alpha_n^{\mathrm{c}} \boldsymbol{d}_n$$

The coordinates of the current point $\boldsymbol{x}^{\mathrm{c}}$ reads:

$$\alpha_n^{\mathrm{c}} = \boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{x}^{\mathrm{c}}/\boldsymbol{d}_n^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_n$$

So, the distribution of the component on $\boldsymbol{d}_1$ is:

$$\begin{aligned}
\alpha_1^{\mathrm{c}} - \widetilde{\alpha} \ &\sim \ \mathcal{N}\left(\alpha_1^{\mathrm{c}} - \frac{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}(\boldsymbol{x}^{\mathrm{c}} - \boldsymbol{m})}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1} \ ; \ \frac{1}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1}\right) \\
&= \ \mathcal{N}\left(\frac{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{m}}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1} \ ; \ \frac{1}{\boldsymbol{d}_1^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{d}_1}\right) \qquad \square
\end{aligned}$$

2.2. Perturbed Gradient Scan Gibbs Sampling

The principle of the proposed algorithm is to sample, at each iteration, one direction of $\mathbb{R}^N$. The chosen direction will be the (possibly preconditioned) gradient of the potential of $p(\boldsymbol{x})$, with a stochastic perturbation as an additional Gaussian component in $\mathbb{R}^N$ with mean $\boldsymbol{0}$ and covariance $\sigma^2 \boldsymbol{I}_N$. The proposed algorithm is named Perturbed Gradient Scan Gibbs Sampler (PGSGS) and it is described by algorithm 1.

Algorithm 1 : PGSGS

Define an initial point $\boldsymbol{x}^{(0)}$ and a variance $\sigma^2 > 0$.
For $t = 1, 2, \ldots$ iterates

1: compute $\boldsymbol{d} = \nabla J(\boldsymbol{x}^{(t-1)}) = \boldsymbol{Q}(\boldsymbol{x}^{(t-1)} - \boldsymbol{m})$

2: sample perturbed gradient $\tilde{\boldsymbol{d}} \sim \mathcal{N}(\boldsymbol{d}, \sigma^2 \boldsymbol{I}_N)$

3: sample $\alpha^{(t)} \sim \mathcal{N}\left(\dfrac{-\tilde{\boldsymbol{d}}^{\mathrm{t}}\boldsymbol{Q}(\boldsymbol{x}^{(t-1)} - \boldsymbol{m})}{\tilde{\boldsymbol{d}}^{\mathrm{t}}\boldsymbol{Q}\tilde{\boldsymbol{d}}} \ ; \ \dfrac{1}{\tilde{\boldsymbol{d}}^{\mathrm{t}}\boldsymbol{Q}\tilde{\boldsymbol{d}}}\right)$

4: compute $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} + \alpha^{(t)}\tilde{\boldsymbol{d}}$

5: quit when the stopping criterion is reached

Proposition 2. *For any $\sigma > 0$, the Markov chain produced by algorithm 1 is aperiodic and irreducible, in addition it admits $p(\boldsymbol{x})$ as an invariant distribution.*

Proof. The algorithm 1 can be viewed as a random scan Gibbs sampler as studied in [22] (see also [21, 23]), where the random choice is not according to the coordinates in the

canonical basis, but according to a given direction $\boldsymbol{d}_1$. Therefore the same arguments as detailed in [22] can be used to prove the proposition: the fact that algorithm 1 is sampling from the right conditional distribution ensures that $p(\boldsymbol{x})$ is an invariant distribution of the chain. And the fact that if $\sigma > 0$ all the direction can be explored with a strictly positive probability ensures the chain to be irreducible. $\square$

2.3. Remarks and extensions

Proposition 2 states that $p(\boldsymbol{x})$ is an invariant density if $\sigma > 0$. In fact, the direction perturbation is needed to ensure the chain's irreducibility. If $\sigma = 0$, there is no perturbation, hence the sampling direction is a deterministic function of $\boldsymbol{x}^{(t-1)}$. More assumption are then needed to ensure the irreducibility. This point will be addressed in a future paper.

A possible extension is to include conjugate directions. This comes down to integrate a "conjugate gradient sampling algorithm" [16] into the Gibbs loop. Another possible extension relies on preconditioned directions as shown in Sect. 3.

More generally, one can consider other directions, as a function $\Phi(\boldsymbol{x}^{(t-1)})$ of current states $\boldsymbol{x}^{(t-1)}$. The perturbation applied in step 2 may be unnecessary, depending on Φ. For example if $\Phi(\boldsymbol{x}^{(t-1)})$ is the preconditioned gradient direction the perturbation is necessary in the general case. If $\Phi(\boldsymbol{x}^{(t-1)})$ is one particular pixel (and ensures that all pixels are infinitely sampled) then the algorithm comes down to the pixel-by-pixel Gibbs sampler, and the perturbation step is unnecessary.

3. NUMERICAL RESULTS

The paper details an application of the proposed PGSGS algorithm in two problems. The first one is an illustration in 2D for quantitative and qualitative results. The second is a super-resolution problem (as in [17] and [20]).

3.1. Results on 2D Gaussian law

In the 2D case, samples are generated by several algorithms:

- Cholesky (`Chol`)
- PGSGS (`GSS`) possibly with preconditioning (`PreGSS`). The preconditionner is taken as the inverse of the co-variance diagonal.
- Truncated and Exact Perturbation Optimization (`TPO` and `EPO`) [17].

For comparison, inefficient algorithms have also been tested:

- PGSGS with null (or too small) perturbation (`NoPer`)
- the iid direction sampler (`IIDDS`): a PGSGS with perturbation simulated from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

Fig. 1 shows the histograms for each algorithm. The `Chol` and `EPO` correctly simulate the law as expected. Concerning `PreGSS` the algorithm provides as good as `Chol` simulation.

GSS has more difficulties and often need more iterations as illustrated here where the exploration is not yet sufficient. The Kullback-Leiber divergence (KLD) between Gaussians, along the iterations, for each algorithm are in Fig. 2 and 4. We separate the KLD into a mean and a covariance term. As expected, `Chol` performs better than any other, followed by our algorithm `PreGSS`, `EPO` and finally `GSS`. `TPO` is the worst and the effect of truncation is blatant. The main difference concerns the covariance since the mean seems well estimated by all algorithms (all the KLD are in the same range).

For inefficient algorithms of Fig. 3 and 4, the mean seems also correctly estimated, but the covariance is defective. It illustrates the difficulty of full Gaussian simulation when the covariance is important. It is also surprising that inefficient algorithms still correctly estimate the mean. The Fig. 3 is a good illustration of the phenomena in particular for `IIDDS`. For this algorithm, the coefficients are concentrated around the mean but the exploration is deffient by choosing an isotropic direction around the origin. The covariance is biased even with a huge number of iteration. This explains the bad KLD reported in Tab. 1. The `IIDDS` illustrates the inefficiency of PGSGS when a defective direction is chosen.
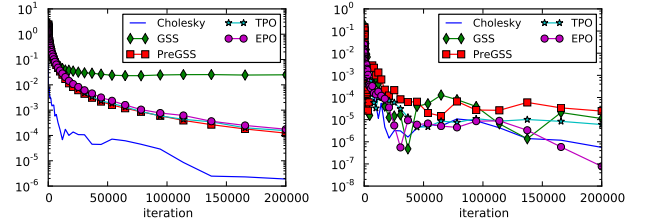

Fig. 2. KLD of the correct algorithms. left: the full KLD. right: KLD of the mean. The main difference concerns the covariance estimation.
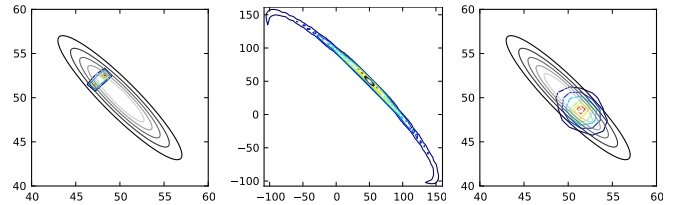

Fig. 3. Histograms of samples of deficient algorithm from left to right: `GSS NoPer`, `IIDDS` and `TPO`.

3.2. Super resolution as a large scale problem

As a more realistic question we address a super-resolution problem. Due to high-dimension `Chol` and `EPO` are *not* feasible. The feasible algorithms are all variants of PGSGS and `TPO` (and [20] partially inspired by `TPO`).

The usual forward model reads $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$. In this equation, $\boldsymbol{y} \in \mathbb{R}^M$ collects the pixels of the low resolution

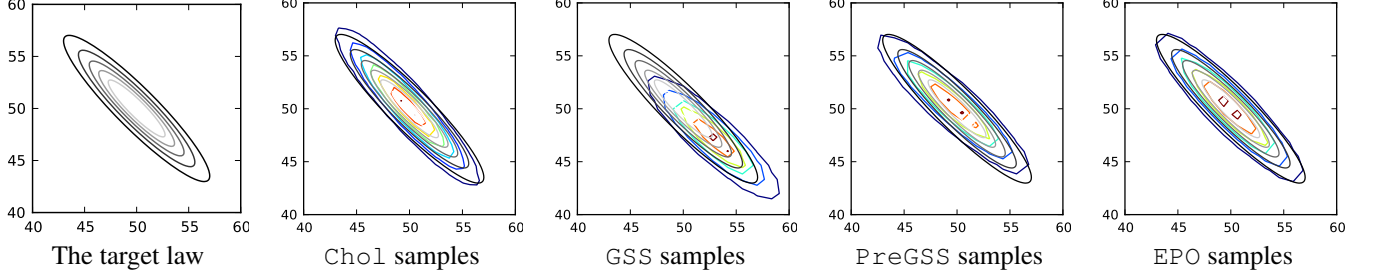| The target law | Chol samples | GSS samples | PreGSS samples | EPO samples |

Fig. 1. Histograms from correct algorithms. The target law is correctly explored by all algorithms.

Chol	EPO	GSS	PreGSS	GSS NoPer	PreGSS NoPer	TPO
$6{,}51 \cdot 10^{-6}$	$1{,}63 \cdot 10^{-4}$	$2{,}93$	$1{,}93 \cdot 10^{-4}$	$17\,069$	$204\,070$	$1{,}05$

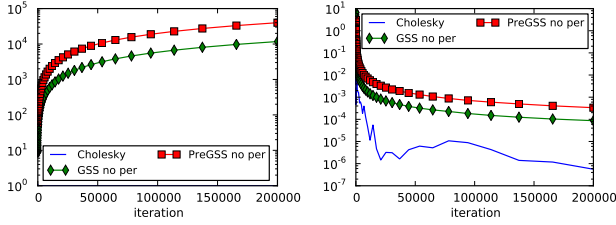**Table 1**. The Kullback-Leiber divergence

Fig. 4. KLD of the deficient algorithms. left: the covariance part of the KLD. right: the mean part of the KLD. The main difference concerns the covariance estimation.

data (five 128×128 images) and $\boldsymbol{x} \in \mathbb{R}^N$ collects the pixels of the original scene (one 256×256 image). The noise $\boldsymbol{n} \in \mathbb{R}^M$ accounts for measurement and modeling errors. $\boldsymbol{H}$ is a $N \times N$ circulant-block-circulant convolution matrix modelling the observation system (here it is a 5×5 window). $\boldsymbol{S}$ is a $M \times N$ matrix modelling motion (here translation) and decimation: it is a down-sampling binary matrix indicating which pixel of the blurred image is observed. This operator break the invariance of the forward model.

The chosen noise prior is uncorrelated $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \gamma_{\boldsymbol{n}}^{-1}\boldsymbol{I})$ ans the chosen object prior accounts for smoothness: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, (\gamma_{\boldsymbol{x}}\boldsymbol{D}^{\mathrm{t}}\boldsymbol{D})^{-1})$ where $\boldsymbol{D}$ is the circulant convolution matrix of the Laplacian filter. The full posterior pdf writes

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\frac{\gamma_{\boldsymbol{n}}}{2}\|\boldsymbol{y} - \boldsymbol{SHx}\|^2 - \frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{Dx}\|^2\right].$$

Because of $\boldsymbol{S}$, the covariance can not be diagonalized by Fourier transform. However we have a diagonalizable preconditioner $\boldsymbol{M} = (\gamma_{\boldsymbol{n}}\boldsymbol{H}^t\boldsymbol{H} + \gamma_{\boldsymbol{x}}\boldsymbol{D}^t\boldsymbol{D})^{-1}$, hence a direct application of our `PreGSS`: at iteration t

1. draw $\boldsymbol{d}^{(t)} \sim p\left(\boldsymbol{d}|\boldsymbol{x}^{(t-1)}\right) = \mathcal{N}\left(\boldsymbol{M}\boldsymbol{\nabla}J(\boldsymbol{x}), \epsilon\boldsymbol{I}\right)$

2. draw $\alpha^{(t)} \sim p\left(\alpha|\boldsymbol{d}^{(t)}\right) = \mathcal{N}\left(\mu_\alpha, \sigma_\alpha^2\right)$ with

$$\mu_\alpha = -\frac{\boldsymbol{d}^{\mathrm{t}}\boldsymbol{\nabla}J(\boldsymbol{x})}{\boldsymbol{d}^{\mathrm{t}}\boldsymbol{\Sigma}^{-1}\boldsymbol{d}} \quad \text{and} \quad \sigma_\alpha^2 = \frac{1}{\boldsymbol{d}^{\mathrm{t}}\boldsymbol{\Sigma}^{-1}\boldsymbol{d}}$$

and update $\boldsymbol{x}^{(t)} = \boldsymbol{x}^{(t-1)} + \alpha^{(t)}\boldsymbol{d}^{(t)}$.
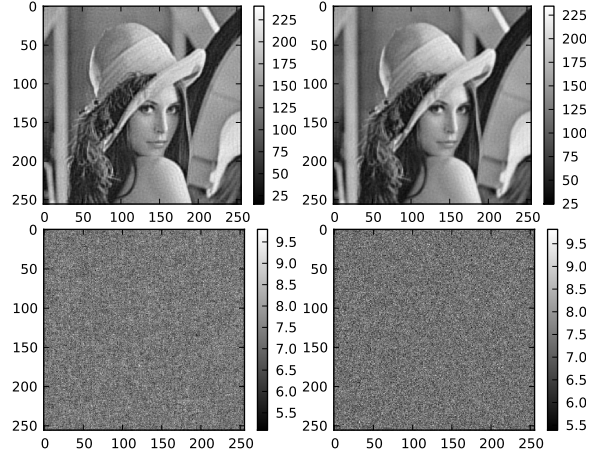

Fig. 5. Top: estimated mean with `TPO` (left) and `PreGSS` (right). Bottom: estimated standard deviation.

We compare to the `TPO` algorithm [17]. In both cases 100 samples have been simulated. The truncation for `TPO` is sufficiently late (more than 50 iterations) so no difference is visible with more iterations. We can see Fig. 5 that `PreGSS` provides the same quality estimation for both mean and standard deviation than `TPO` *while being a guaranteed algorithm*.

4. CONCLUSIONS AND PERSPECTIVES

As a conclusion, we provide a new algorithm for simulation of high-dimensional Gaussians with guaranteed convergence. We illustrate in a 2D problem where everything is under control and in a super-resolution problem. A direct perspective is the addition of such algorithm in a Gibbs sampler that includes hyper-parameter simulation. In this case, our algorithm still simulate the correct target posterior without approximation.

5. REFERENCES

[1] F. Orieux, J.-F. Giovannelli, T. Rodet, and A. Abergel, "Estimating hyperparameters and instrument parameters in regularized inversion. illustration for Herschel/SPIRE map making," *Astron. Astrophys.*, vol. 549, Jan. 2013.

[2] J. Idier, Ed., *Bayesian Approach to Inverse Problems.* London: ISTE Ltd and John Wiley & Sons Inc., 2008.

[3] J.-F. Giovannelli, "Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function," *IEEE Trans. Image Processing*, vol. 17, no. 1, pp. 16–26, Jan. 2008.

[4] F. Orieux, J.-F. Giovannelli, and T. Rodet, "Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution," *J. Opt. Soc. Amer.*, vol. 27, no. 7, pp. 1593–1607, July 2010.

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[6] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, no. 7, pp. 932–946, July 1995.

[7] O. Féron, B. Duchêne, and A. Mohammad-Djafari, "Microwave imaging of piecewise constant objects in a 2D-TE configuration," *International Journal of Applied Electromagnetics and Mechanics*, vol. 26, no. 6, pp. 167–174, IOS Press 2007.

[8] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss-Markov-Potts prior models and variational Bayesian computation," *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2265–2277, 2010.

[9] C. Vacar and J.-F. Giovannelli, "Unsupervised joint deconvolution segmentation for textured images," *submitted to IEEE Trans. Signal Processing*, 2015.

[10] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, pp. 959–963, Aug. 1985.

[11] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Application.* Academic Press Inc, 1992.

[12] H. Rue, "Fast sampling of Gaussian Markov random fields," *J. R. Statist. Soc. B*, vol. 63, no. 2, 2001.

[13] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*, ser. Monographs on Statistics and Applied Probability. Chapman & Hall, 2005, vol. 104.

[14] P. Lalanne, D. Prévost, and P. Chavel, "Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation," *Applied Optics*, vol. 40, 2001.

[15] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods.* Springer Verlag, Berlin Germany, 2003.

[16] A. Parker and C. Fox, "Sampling gaussian distributions in krylov spaces with conjugategradients," *SIAM J. Sci. Comput.*, vol. 34, no. 3, 2012.

[17] F. Orieux, O. Féron, and J.-F. Giovannelli, "Sampling high-dimensional Gaussian fields for general linear inverse problem," *IEEE Signal Proc. Let.*, vol. 19, no. 5, pp. 251–254, May 2012.

[18] X. Tan, J. Li, and P. Stoica, "Efficient sparse Bayesian learning via Gibbs sampling," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3634 –3637.

[19] G. Papandreou and A. Yuille, "Gaussian sampling by local perturbations," in *Proc. Int. Conf. on Neural Information Processing Systems (NIPS)*, Vancouver, Canada, Dec. 2010, pp. 1858–1866.

[20] C. Gilavert, S. Moussaui, and J. Idier, "Rééchantillonnage gaussien en grande dimension pour les problèmes inverses," in *Actes 24e coll. GRETSI*, Brest, France, Sep. 2013.

[21] R. A. Levine, Z. Yu, W. G. Hanley, and J. J. Nitao, "Implementing random scan Gibbs samplers," *Computational Statistics*, vol. 20, pp. 177–196, 2005.

[22] R. A. Levine and G. Casella, "Optimizing random scan Gibbs samplers," *Journal of Multivariate Analysis*, vol. 97, pp. 2071–2100, 2006.

[23] K. Latuszynski, G. O. Roberts, and J. Rosenthal, "Adaptive gibbs samplers and related mcmc methods," *Annals of Applied Probability*, vol. 23, pp. 66–98, 2013.

[24] O. Stramer and R. L. Tweedie, "Langevin-type models i: Diffusions with given stationary distributions, and their discretizations," *Methodology and Computing in Applied Probability*, vol. 1, pp. 283–306, 1999.

[25] ——, "Langevin-type models ii: Self-targeting candidates for mcmc algorithms," *Methodology and Computing in Applied Probability*, vol. 1, pp. 307–328, 1999.

[26] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Physics Letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[27] C. Vacar, J.-F. Giovannelli, and Y. Berthoumieu, "Langevin and Hessian with Fisher approximation stochastic sampling for parameter estimation of structured covariance," in *Proc. IEEE ICASSP*, Prague, Czech Republic, May 2011, pp. 3964–3967.

[28] C. Fox, "A conjugate direction sampler for normal distributions with a few computed examples," Electronics Technical Report No. 2008-1, University of Otago, Dunedin, New Zealand, Tech. Rep., 2008.