Efficient sampling of high-dimensional Gaussian fields: the non-stationary / non-sparse case

F. Orieux*, O. Féron and J.-F. Giovannelli*

January 26, 2013

Abstract

This paper is devoted to the problem of sampling Gaussian fields in high dimension. Solutions exist for two specific structures of inverse covariance: sparse and circulant. The proposed approach is valid in a more general case and especially as it emerges in inverse problems. It relies on a perturbation-optimization principle: adequate stochastic perturbation of a criterion and optimization of the perturbed criterion. It is shown that the criterion minimizer is a sample of the target density. The motivation in inverse problems is related to general (non-convolutive) linear observation models and their resolution in a Bayesian framework implemented through sampling algorithms when existing samplers are not feasible. It finds a direct application in myopic and/or unsupervised inversion as well as in some non-Gaussian inversion. An illustration focused on hyperparameter estimation for super-resolution problems assesses the effectiveness of the proposed approach.

1 Introduction

This work deals with simulation of high-dimensional Gaussian and conditional Gaussian fields. The problem difficulty is directly related to handling high-dimensional covariances $R$ and precision matrices $Q = R^{-1}$. Inversion and factorization of these matrices can be very costly in terms of time and

*F. Orieux is with Pasteur Institute, 25 rue du Dr Roux, 75015 Paris, France, orieux@pasteur.fr. O. Féron is with EDF Research & Developments, Dpt OSIRIS, 92140 Clamart, France, olivier-2.feron@edf.fr. J.-F. Giovannelli is with the Laboratoire de l’Intégration du Matériau au Système, 33405 Talence, France, Giova@IMS-Bordeaux.fr.
memory, if not impossible. General tools [1, 2] provide pixel-by-pixel sequential Gibbs or Hastings-Metropolis algorithms but they are not practicable in high dimension. This problem is old and solutions exist in two cases.

- When \( Q \) is sparse, two strategies have been proposed. The first one [3, chap. 8], relies on a parallel Gibbs sampler based on a chessboard-like decomposition. It takes advantage of the sparsity of \( Q \) to allow large blocks of variables to be simultaneously updated. The second strategy [4, 5] relies on a Cholesky decomposition \( Q = L^T L \): a sample \( x \) can be obtained by solving the linear system \( Lx = \varepsilon \), where \( \varepsilon \) is a zero-mean white Gaussian vector. The sparsity of \( L \) ensures feasible numerical resolution of the linear system.

- In [6, 7] the authors pointed out an efficient solution for the case of circulant matrix \( Q \), even non-sparse. In this case, the covariance is diagonal in the Fourier domain: the sampling is based on independent sampling of the Fourier coefficients. Finally, the sampling is efficiently computed by FFT and it has been used in [8–10].

To our knowledge there is no solution for sampling more general high-dimensional Gaussian fields. In this paper we propose an efficient algorithm for a more general case where \( Q \) is non-sparse, non-circulant and very large. The proposed approach is applicable to any precision matrix of the form

\[
Q = \sum_{k=1}^{K} M_k^T R_k^{-1} M_k
\]

for which, to the best of our knowledge, no practical solution exists. A recent paper [11] briefly describes a similar algorithm for a compress sensing problem in signal processing. Our paper deepens and generalizes this contribution.

The problem of sampling such fields is commonly encountered in Bayesian approaches for inverse problems and especially in high dimension like in image reconstruction. Indeed, let us consider the general linear forward model

\[
y = Hx + n,
\]

where \( y \), \( n \) and \( x \) denote the observations, the noise and the unknown image and \( H \) is a linear operator. Consider, again, two prior densities for \( n \) and \( x \) that are Gaussian conditionally to a set of parameters \( \theta \) and focus on the joint estimation of \( x \) and \( \theta \) from the posterior density \( p(x, \theta|y) \). This framework is very general and can be used in many applications. In image
reconstruction, it covers a majority of current problems such as unsupervised [9] or myopic [10] inversion, since acquisition (or instrument) parameters and hyperparameters can be included in $\theta$. The framework also covers some non-Gaussian priors involving auxiliary/hidden variables [8, 9, 12–14] (location mixture or scale mixture of Gaussian), by including these variables in $\theta$.

The joint estimation of $x$ and $\theta$ from the posterior density $p(x, \theta|y)$ commonly requires the handling of the posterior conditional probability $p(x|\theta, y)$. Under the above assumption, this density is Gaussian with precision matrix $Q$ of the form (1), as shown in section 2.2. The capability to sample from this density makes it possible to propose, for instance, stochastic optimization [12] or Gibbs sampler [10, 13]. In the general case of inverse problems, $Q$ is neither sparse nor circulant so existing sampling methods fail whereas the proposed sampling method is effective.

Subsequently, section 2 presents the proposed algorithm and its direct application to general inverse problems. Section 3 illustrates the algorithm through an academic inverse problem in super-resolution imaging. Section 4 concludes and presents some perspectives.

2 Perturbation-optimization algorithm

2.1 Description

Here we focus on the problem of sampling from a target Gaussian density $\mathcal{N}(0, Q^{-1})$ where $Q$ is in the form (1). When $Q$ is neither sparse nor circulant, existing methods fail in high dimension and we propose an efficient solution based on the Perturbation-Optimization (PO) algorithm described by Algorithm 1 and Proposition 1.

Algorithm 1: Perturbation-Optimization algorithm.

1. **Step P (Perturbation):** Generate independent Gaussian variables $\eta_k$, $k = 1, \ldots, K$ following

   \[ \eta_k \sim \mathcal{N}(0, R_k), \quad \forall k = 1, \ldots K \quad (3) \]

2. **Step O (Optimization):** Compute $\hat{x}$ as the minimizer of the criterion

   \[ J(x|\eta_1, \ldots, \eta_K) = \sum_{k=1}^{K} (\eta_k - M_k x)^t R_k^{-1} (\eta_k - M_k x) \quad (4) \]
**Proposition 1.** The minimizer $\hat{x}$ of criterion (4) resulting from Algorithm 1 is Gaussian

$$\hat{x} \sim \mathcal{N}(0, Q^{-1}).$$

**Proof.** The minimizer $\hat{x}$ of criterion (4) has an analytical expression:

$$\hat{x} = \left[ \sum_{k=1}^{K} M_k R_k^{-1} M_k^T \right]^{-1} \left( \sum_{k=1}^{K} M_k R_k^{-1} \eta_k \right)$$

$$= Q^{-1} \left( \sum_{k=1}^{K} M_k R_k^{-1} \eta_k \right).$$

It is clearly a zero-mean Gaussian vector as a linear combination of $K$ zero-mean Gaussian vectors. The covariance is calculated below using elementary algebra: from (3) and (6), we have

$$\mathbb{V}[\hat{x}] = Q^{-1} \left[ \sum_{k,k'=1}^{K} M_k R_k^{-1} \mathbb{E} \left[ \eta_k \eta_{k'} \right] R_{k'}^{-1} M_{k'} \right] Q^{-1}$$

$$= Q^{-1} \left[ \sum_{k=1}^{K} M_k R_k^{-1} \mathbb{E} \left[ \eta_k \eta_k \right] R_k^{-1} M_k \right] Q^{-1}$$

$$= Q^{-1} \left[ \sum_{k=1}^{K} M_k R_k^{-1} M_k \right] Q^{-1} = Q^{-1}$$

that completes the proof.

The criterion $J(x|\eta_1, \ldots, \eta_K)$ being quadratic, we have access to the whole available literature on efficient numerical optimization tools, e.g. iterative techniques such as gradient based ones (standard, corrected, conjugate, optimal step size, ...). We have to highlight that in theory the sample of the target density is the exact optimum of the perturbed criterion. Therefore the optimization step may require as much descent steps as the dimension of the problem. However, the optimization procedure can be stopped more rapidly without practical loss of efficiency.

Obviously, the efficiency of the algorithm depends on the capability to easily sample from Gaussian densities $\mathcal{N}(0, R_k)$. This will be actually the case in inverse problem applications as shown in section 2.2.

Moreover, we can actually extend Proposition 1 and Algorithm 1 when the mean of the target Gaussian density is not zero, by proposing Algorithm 2 above and Corrollary 1 below.
Algorithm 2: Perturbation-Optimization algorithm.

1: Step P (Perturbation): Generate independent Gaussian variables $\zeta_k$, $k = 1, \ldots, K$ following
   \begin{equation}
   \zeta_k \sim \mathcal{N}(m_k, R_k), \quad \forall k = 1, \ldots K
   \end{equation}

2: Step O (Optimization): Compute $\bar{x}$ as the minimizer of the criterion
   \begin{equation}
   J(x|\zeta_1, \ldots, \zeta_K) = \sum_{k=1}^{K} (\zeta_k - M_kx)^t R_k^{-1} (\zeta_k - M_kx)
   \end{equation}

Corollary 1: The solution $\bar{x}$ resulting from Algorithm 2 is Gaussian
   \begin{equation}
   \bar{x} \sim \mathcal{N}\left( Q^{-1} \left( \sum_{k=1}^{K} M_k^t R_k^{-1} m_k \right), Q^{-1} \right).
   \end{equation}

Proof – Consider $\eta_k = \zeta_k - m_k$, $k = 1, \ldots, K$, and the minimizer $\hat{x}$ of the criterion (4). Hence it is trivial to show that $\bar{x} = \hat{x} + Q^{-1}\left( \sum_{k=1}^{K} M_k^t R_k^{-1} m_k \right)$.

Using the results of Proposition 1 on $\hat{x}$, we can show
   \begin{align*}
   \mathbb{E}[\bar{x}] &= Q^{-1} \left( \sum_{k=1}^{K} M_k^t R_k^{-1} m_k \right) \\
   \mathbb{V}[\bar{x}] &= \mathbb{V}[\hat{x}] = Q^{-1}
   \end{align*}

and that completes the proof.

2.2 Application to inverse problems

The purpose is to solve an inverse problem, stated by the forward model (2), in a Bayesian framework based on the following models:

- $H$ describes an observation system that can depend on unknown acquisition parameters,
- prior densities for the observation noise and for the object are Gaussian $n \sim \mathcal{N}(m_n, R_n)$ and $x \sim \mathcal{N}(m_x, R_x)$, conditionally on a set of auxiliary variables.
In a general statement, \( \theta \) collects acquisition parameters, hyperparameters and auxiliary variables. This framework covers myopic (semi-blind) and unsupervised inversion, non-stationary or inhomogeneous Gaussian priors and non-Gaussian priors involving auxiliary variables.

The general inversion problem then consists in estimating \( x \) and \( \theta \) through the density \( p(x, \theta | y) \). The posterior mean can be approximated using a Gibbs sampler. It is an iterative algorithm which alternately samples from 

\[
p(\theta | x, y) \quad \text{and} \quad p(x | \theta, y)
\]

The conditional posterior \( p(x | y, \theta) \) is a correlated Gaussian field: 

\[
x \sim \mathcal{N}(m_{\text{post}}^x, R_{\text{post}}^x)
\]

where \( \theta \) is embedded in \( H, R_n \) and \( R_x \) for simpler notations.

If \( H \) has no particular properties then the precision matrix \( Q = (R_x^{\text{post}})^{-1} \) is neither sparse nor circulant, and existing sampling methods are not applicable. The Perturbation-Optimization algorithm makes it possible to efficiently sample from \( \mathcal{N}(m_{\text{post}}^x, R_{\text{post}}^x) \). In particular, applying Algorithm 2 with \( K = 2, M_1 = H, M_2 = I, R_1 = R_n, R_2 = R_x, m_1 = m_n \) and \( m_2 = m_x \), directly gives a sample from this density. Then, this algorithm ensures that correct posterior mean and covariance are obtained, at the same time. This increases the usefulness of this method for inverse problems.

3 Illustration

The proposed PO algorithm is an effective tool for high dimensional inverse problems, e.g. image reconstruction. In this context, it opens up the possibility to resort to stochastic sampling algorithms (MCMC, Gibbs, Metropolis-Hastings,...) providing two main advantages:

- the capability to jointly estimate several unknowns when the global modelization is more natural through conditional distributions (hierarchical structure),

- in addition, the access to the entire distribution of the unknowns providing uncertainties (standard deviations, confidence intervals,...).

3.1 Two examples: electromagnetics and fluorescent microscopy

For example, the proposed PO algorithm has been applied to an electromagnetic inverse scattering problem by one of the authors [13]. In a domain integral representation, the forward model expresses observed data as
a bi-linear function of unknown object and unknown induced current. The
bi-linear structure leads to a modelization with conditional Gaussians: the
prior for the induced current is Gaussian given the object, and the prior for
the object is Gaussian given the induced current. The joint estimation of
the object and the current is tackled in a Bayesian framework and computed
by means of a Gibbs sampler in which the sampling of the current is made
possible thanks to the proposed PO algorithm.

In [15] it has been applied by another one of the authors to process
data in biology imaging to achieve super-resolution in fluorescent microscopy
trough Structured Illumination. The problem is also tackled in a Bayesian
framework and implemented by means of a Gibbs sampler. The density for
the object given the other variables is Gaussian with non-invariant covariance
(due to non-invariant illumination of the biological sample) making the use of
existing techniques impossible. Again, the proposed PO algorithm overcomes
this difficulty and results in the capability to estimate hyperparameters and
acquisition parameters, while also providing uncertainties.

3.2 Unsupervised super-resolution

In the following, we detail an application of the proposed PO algorithm to the
super-resolution (SR) academic problem: several blurred and down-sampled
(low resolution) images of a scene are available in order to retrieve the original
(high resolution) scene [16, 17]. It is shown that the crucial novelty, enabled
by the proposed PO algorithm, is to allow the use of sampling algorithms
in SR methods and to provide joint image and hyperparameters estimation
including uncertainties.

The usual forward model writes $y = Hx + n = PCx + n$, where $y \in \mathbb{R}^M$
collects the low resolution images (5 images of 128 × 128 pixels), $x \in \mathbb{R}^N$
is the original image (256 × 256 pixels), $n$ is the noise, $C$ and $P$ are circulant
convolution and decimation matrices. The prior density for $n$ is $N(0, \gamma_{n}^{-1}I)$
and the one for $x$ is $N(0, \gamma_{x}^{-1}D^t D)$ where $D$ is the Laplacian operator. The
hyperparameters $\gamma_{n}$ and $\gamma_{x}$ are unknown and their prior law are Jeffreys’.
The posterior density is then

$$p(x, \gamma_{n}, \gamma_{x}|y) \propto \gamma_{n}^{M/2-1}\gamma_{x}^{(N-1)/2-1} \cdot \exp\left[-\frac{\gamma_{n}}{2}\|y - PCx\|^2 - \frac{\gamma_{x}}{2}\|Dx\|^2\right]. \quad (9)$$

It is explored by a Gibbs sampler: iteratively sampling $\gamma_{n}$, $\gamma_{x}$ and $x$ under
Figure 1: Chains and histograms of hyperparameters $\gamma_n$ and $\gamma_x$. 
Figure 2: Image reconstruction: true image 2(a), one of the low resolution images 2(b) and the proposed estimate 2(c). The plot 2(d) is a true image slice inside the 99% confidence interval around the estimate.
their respective conditional probabilities

\[
p(\gamma_n^{(k)}|\gamma_x, y) = \mathcal{G}\left(1 + M/2, 2/\|y - PC x^{(k-1)}\|^2\right)
\]

\[
p(\gamma_x^{(k)}|\gamma_n, y) = \mathcal{G}\left(1 + (N - 1)/2, 2/\|D x^{(k-1)}\|^2\right)
\]

\[
p(x^{(k)}|\gamma_x, \gamma_n, y) = \mathcal{N}(m_x^{\text{post}}, R_x^{\text{post}})
\]

with

\[
R_x^{\text{post}} = \left(\gamma_n^{(k)} C^t P^t P C + \gamma_x^{(k)} D^t D\right)^{-1}
\]

\[
m_x^{\text{post}} = \gamma_n^{(k)} R_x^{\text{post}} P^t C^t y.
\]

The conditional posteriors for the hyperparameters are Gamma laws and consequently, easy to sample. The conditional posterior for \( x \) is Gaussian, but existing sampling approaches are not operational due to the structure of the covariance \( R_x^{\text{post}} \), as explained in Section 2.2 with \( H = PC \); \( H \) is non-circulant due to the decimation and \( H \) is not sparse especially in the case of large support. In this case, the PO algorithm 2 directly provides a desired sample (with both correct mean and correct covariance).

It is important to keep in mind that the proposed PO algorithm does not improve image quality itself (w.r.t. other SR methods) but the crucial novelty is to allow for hyperparameter estimation. In this sense, Fig. 1 shows the hyperparameter iterates (that illustrate the operation and convergence) and histograms (that approximate marginal posteriors); the posterior means are \( \hat{\gamma}_n \approx 8 \) and \( \hat{\gamma}_x \approx 2 \times 10^{-3} \). Concerning the images themselves, results are shown in Fig. 2: estimated image in 2(c) clearly shows a better resolution than data in Fig. 2(b) and it is visually close to the original image of 2(a). It is then clear that the approach produces correct hyperparameters i.e. correct balance between data and prior. Moreover, uncertainties are derived from the samples through the posterior standard deviation. It is illustrated in Fig. 2(d) which shows that the true image is inside the 99% confidence interval around the estimate. As a conclusion, the proposed PO algorithm makes it possible to include sampling algorithms in SR method whereas it was not possible before. It enables to provide joint image and hyperparameters estimation as well as uncertainties computations.
4 Conclusion

This paper presents a novel approach for sampling high-dimensional Gaussian fields when usual approaches are ineffective. A sample of the target density is produced as the minimizer of a precisely designed quadratic criterion. It relies on a perturbation-optimization principle: adequate stochastic perturbation of a criterion and optimization of the perturbed criterion. It is shown that the criterion minimizer is a sample of the target density. The approach is applicable as soon as a particular factorization of the precision matrix is available, and it is usually the case in inverse problems. There is a wide class of applications, in particular any data processing problem based on a linear forward model and conditional Gaussian prior for noise and object. The effectiveness of the proposed algorithm has been illustrated in [13, 15] and in this paper on a more academic super-resolution imaging problem allowing automatic tuning of hyperparameters.

5 Acknowledgment

The authors would like to thank Jérôme Idier (IRCyN) for inspiration of this work [18], Thomas Rodet and Ali Mohammad–Djafari (L2S), for fruitful discussions, and Cornelia Vacar (IMS) for carefully reading the paper.

References

[1] C. P. Robert and G. Casella, Monte-Carlo Statistical Methods, ser. Springer Texts in Statistics. New York, NY: Springer, 2000.

[2] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Markov Chain Monte Carlo in practice. Boca Raton: Chapman & Hall/CRC, 1996.

[3] G. Winkler, Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer Verlag, BerlinGermany, 2003.

[4] H. Rue, “Fast sampling of Gaussian Markov random fields,” J. R. Statist. Soc B, vol. 63, no. 2, 2001.

[5] P. Lalanne, D. Prévost, and P. Chavel, “Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation,” Appl. Opt, vol. 40, 2001.
[6] 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, August 1985.

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

[8] D. Geman and C. Yang, “Nonlinear image recovery with half-quadratic regularization,” *IEEE Trans. Im. Proc.*, vol. 4, no. 7, July 1995.

[9] J.-F. Giovannelli, “Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function,” *IEEE Trans. Im. Proc.*, vol. 17, no. 1, January 2008.

[10] F. Orieux, J.-F. Giovannelli, and T. Rodet, “Bayesian estimation of regularization and point spread function parameters for wiener–hunt deconvolution,” *J. Opt. Soc. Am. A*, vol. 27, no. 7, pp. 1593–1607, 2010.

[11] X. Tan, J. Li, and P. Stoica, “Efficient sparse Bayesian learning via Gibbs sampling,” in *Acoustics Speech and Signal Processing (ICASSP)*, 2010 *IEEE International Conference on*, March 2010, pp. 3634 –3637.

[12] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, 1984.

[13] O. Féron, B. Duchêne, and A. Mohammad-Djafari, “Microwave imaging of piecewise constant objects in a 2D-TE configuration,” *Intern. Journ. App. Electr. Mec.*, vol. 26, no. 3-4, 2007.

[14] 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.

[15] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, and J.-C. Olivo-Marin, “Bayesian estimation for optimized structured illumination microscopy,” *IEEE Trans. Image Processing*, 2011, in revision.

[16] S. C. Park, M. K. Park, and M. G. Kang, “Super-resolution image reconstruction: a technical overview,” *IEEE Sig. Proc. Mag.*, May 2003.
[17] G. Rochefort, F. Champagnat, G. L. Besnerais, and J.-F. Giovannelli, “An improved observation model for super-resolution under affine motion,” *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3325–3337, November 2006.

[18] J. Idier, informal discussion, 2009.