Sampling high-dimensional Gaussian distributions for general linear inverse problems

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

***Abstract*—This paper is devoted to the problem of sampling Gaussian distributions in high dimension. Solutions exist for two specific structures of inverse covariance: sparse and circulant. The proposed algorithm is valid in a more general case especially as it emerges in linear inverse problems as well as in some hierarchical or latent Gaussian models. It relies on a perturbation-optimization principle: adequate stochastic perturbation of a criterion and optimization of the perturbed criterion. It is proved that the criterion optimizer is a sample of the target distribution. The main motivation is in inverse problems related to general (non-convolutive) linear observation models and their solution in a Bayesian framework implemented through sampling algorithms when existing samplers are infeasible. It finds a direct application in myopic / unsupervised inversion methods as well as in some non-Gaussian inversion methods. An illustration focused on hyperparameter estimation for super-resolution method shows the interest and the feasibility of the proposed algorithm.**

***Index Terms*—Stochastic sampling, high-dimensional sampling, inverse problem, Bayesian strategy, unsupervised, myopic**

I. Introduction

This work deals with simulation of high-dimensional Gaussian and conditional Gaussian distributions. The difficulty of the problem is directly related to handling high-dimensional covariances $\boldsymbol{R}$ and precision matrices $\boldsymbol{Q} = \boldsymbol{R}^{-1}$. The problem has already been investigated and solutions exist in two cases.

- When $\boldsymbol{Q}$ is sparse, two strategies are available. The first one [1, chap. 8], relies on a parallel Gibbs sampler based on a chessboard-like decomposition. It takes advantage of the sparsity of $\boldsymbol{Q}$ to update simultaneously large blocks of variables. The second strategy [2, 3] relies on a Cholesky decomposition $\boldsymbol{Q} = \boldsymbol{L}^{\mathrm{t}}\boldsymbol{L}$: a sample $\boldsymbol{x}$ is obtained by solving the linear system $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a zero-mean white Gaussian vector. The sparsity of $\boldsymbol{Q}$ ensures feasible numerical factorization and the sparsity of $\boldsymbol{L}$ ensures feasible numerical resolution of the linear system.
- [4, 5] propose a solution for circulant matrix $\boldsymbol{Q}$, 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 sample is computed by FFT and it has been used in [6–10].

To our knowledge there is no solution for more general structure in high dimension because factorization (Cholesky, QR, square root,...), diagonalization and inversion of $\boldsymbol{Q}$ and $\boldsymbol{R}$ are numerically infeasible. The obstacle is due to both computational cost and memory footprint. The proposed algorithm overcomes this obstacle when $\boldsymbol{Q}$ is of the form

$$\boldsymbol{Q} = \sum_{k=1}^{K} \boldsymbol{M}_k^{\mathrm{t}} \boldsymbol{R}_k^{-1} \boldsymbol{M}_k \tag{1}$$

as it appears in inverse problems [11]. Indeed, let us consider the general linear forward model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, where $\boldsymbol{y}$, $\boldsymbol{n}$ and $\boldsymbol{x}$ are the observation, the noise and the unknown image and $\boldsymbol{A}$ is a linear operator. Consider, again, two prior distributions for $\boldsymbol{n}$ and $\boldsymbol{x}$ that are Gaussian conditionally on a parameter $\boldsymbol{\theta}$. This framework is very general: it includes linear inverse problems [11] as well as some hierarchical or latent Gaussian models [12] and it can be used in many applications. In image reconstruction, it covers a majority of current problems, e.g. unsupervised [8] or myopic (semi-blind) [9] inverse problems, by including acquisition parameters and hyperparameters in $\boldsymbol{\theta}$. Moreover, the framework also includes non-linear models, based on conditional linear models such as bilinear or multilinear ones (see Section III-B). The framework also covers some non-stationary or inhomogeneous Gaussian priors and non-Gaussian priors involving auxiliary / latent variables [6, 8, 13–15] (e.g., location or scale mixtures of Gaussian), by including these variables in $\boldsymbol{\theta}$.

Let us focus on the joint estimation of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ from the posterior $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. It commonly requires the handling of the conditional posterior $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ that is Gaussian with precision matrix $\boldsymbol{Q}$ of the form (1), as will be shown in section II-B. In the general case, $\boldsymbol{Q}$ is neither sparse nor circulant so existing sampling algorithms fail when the dimension of $\boldsymbol{x}$ is very large while the proposed one handles this case. It relies on a perturbation-optimization principle: adequate stochastic perturbation of a quadratic criterion and optimization of the perturbed criterion. A recent paper [16] briefly describes a similar algorithm for compressed sensing in signal processing. Our paper deepens and generalizes this contribution.

Subsequently, Section II presents the proposed algorithm and its direct application to linear inverse problems. Section III gives an illustration through an academic problem in super-resolution. Section IV presents conclusions and perspectives.

II. Perturbation-optimization algorithm

A. Description

We focus on the problem of sampling from a target Gaussian distribution whose precision matrix $\boldsymbol{Q}$ is in the form (1). When $\boldsymbol{Q}$ is neither sparse nor circulant, existing algorithms fail in high dimension because of an excessive memory footprint as illustrated in section III. We propose a solution based on the Perturbation-Optimization (PO) algorithm described hereafter, whose memory footprint is far smaller.

F. Orieux is with Pasteur Institute, 75015 Paris, France, orieux@pasteur.fr. O. Féron is with EDF Research & Developments, OSIRIS, 92140 Clamart, France, olivier-2.feron@edf.fr. J.-F. Giovannelli is with Univ. Bordeaux, IMS, 33400 Talence, France, Giova@IMS-Bordeaux.fr.

Proposition 1: The optimizer $\widehat{\boldsymbol{x}}$ of criterion (5) resulting from Algorithm 1 is Gaussian

$$\widehat{\boldsymbol{x}} \sim \mathcal{N}\left(\boldsymbol{Q}^{-1}\left(\sum_{k=1}^{K}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\boldsymbol{m}_k\right), \boldsymbol{Q}^{-1}\right) . \tag{2}$$

Proof: The optimizer $\widehat{\boldsymbol{x}}$ of criterion (5) is explicit:

$$\begin{aligned}\widehat{\boldsymbol{x}} &= \left[\sum_{k=1}^{K}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\right]^{-1}\left(\sum_{k=1}^{K}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\boldsymbol{\eta}_k\right)\\ &= \boldsymbol{Q}^{-1}\left(\sum_{k=1}^{K}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\boldsymbol{\eta}_k\right) .\end{aligned} \tag{3}$$

It is clearly a Gaussian vector as a linear combination of K Gaussian vectors. Its expectation and covariance are calculated below using elementary algebra: from (4) and (3), we have

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{x}}\right] &= \boldsymbol{Q}^{-1}\left(\sum_{k}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\mathbb{E}\left[\boldsymbol{\eta}_k\right]\right)\\ &= \boldsymbol{Q}^{-1}\left(\sum_{k}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\boldsymbol{m}_k\right)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\widehat{\boldsymbol{x}}\widehat{\boldsymbol{x}}^{\mathrm{t}}\right] &= \boldsymbol{Q}^{-1}\left(\sum_{k,k'}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\mathbb{E}\left[\boldsymbol{\eta}_k\boldsymbol{\eta}_{k'}^{\mathrm{t}}\right]\boldsymbol{R}_{k'}^{-1}\boldsymbol{M}_{k'}\right)\boldsymbol{Q}^{-1}\\ &= \boldsymbol{Q}^{-1}\left(\sum_{k}\boldsymbol{M}_k^{\mathrm{t}}\boldsymbol{R}_k^{-1}\left(\boldsymbol{R}_k+\boldsymbol{m}_k\boldsymbol{m}_k^{\mathrm{t}}\right)\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\right)\boldsymbol{Q}^{-1}\\ &= \boldsymbol{Q}^{-1}+\mathbb{E}\left[\widehat{\boldsymbol{x}}\right]\mathbb{E}\left[\widehat{\boldsymbol{x}}\right]^{\mathrm{t}}\end{aligned}$$

that completes the proof. ∎

The feasibility of Step P clearly depends on the capability to sample from Gaussian distributions $\mathcal{N}(\boldsymbol{m}_k, \boldsymbol{R}_k)$. It is usually the case in inverse problems and it will be actually the case in super-resolution applications shown in section III-A and in other contributions shortly described in section III-B.

Regarding Step O, J being quadratic, a large literature [17] is available about its numerical optimization, e.g. gradient procedure (standard, corrected, conjugate, optimal step size…). Such algorithms require the computation of criterion (5) and its gradient. The feasibility of Step O clearly depends on the capability to compute that without the storage of large matrices. It is usually the case in inverse problems and it will be actually the case in applications shown in section III-A and described in section III-B.

However, the desired sample is the exact optimizer, so, Step O could require N iterations of a conjugate gradient algorithm for a problem of dimension N. Therefore the complexity could be $O(N^3)$ that is equivalent to the one of a Cholesky decomposition. However, the optimization procedure can be stopped earlier without practical loss of precision and the complexity falls down to $O(PN^2)$ for P iterations. In addition, for a band matrix, the complexity of the proposed algorithm becomes $O(MPN)$ and the one of the Cholesky decomposition becomes $O(MN^2)$. Anyway, the main advantage of the proposed algorithm is its reduced memory footprint: it avoids the storage of neither $\boldsymbol{Q}$ nor its (Cholesky, QR, square root,…) factors.

Algorithm 1 : Perturbation-Optimization algorithm.

1: **Step P (Perturbation)**: Generate independent vectors

$$\boldsymbol{\eta}_k \sim \mathcal{N}(\boldsymbol{m}_k, \boldsymbol{R}_k), \quad \text{for } k = 1, \dots K \tag{4}$$

2: **Step O (Optimization)**: Compute $\widehat{\boldsymbol{x}}$ as the minimizer of

$$J(\boldsymbol{x}) = \sum_{k=1}^{K}\left(\boldsymbol{\eta}_k - \boldsymbol{M}_k\boldsymbol{x}\right)^{\mathrm{t}}\boldsymbol{R}_k^{-1}\left(\boldsymbol{\eta}_k - \boldsymbol{M}_k\boldsymbol{x}\right) \tag{5}$$

Remark 1: Still regarding Step O, it would be awkward if $\boldsymbol{Q}$ was badly scaled, but it is not the case here for the following reason. In usual ill-conditioned inverse problems, $\boldsymbol{A}$ is badly-scaled but the aim of regularization is precisely to overcome this difficulty and to produce a well-scaled matrix $\boldsymbol{Q}$.

B. Application to inverse problems

The purpose is to solve an inverse problem, stated by the forward model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$, in a Bayesian framework based on the following models:

- $\boldsymbol{A}$ describes any observation system that can depend on unknown acquisition parameters,
- priors for the noise $\boldsymbol{n}$ and the object $\boldsymbol{x}$ are Gaussian $\mathcal{N}(\boldsymbol{m}_n, \boldsymbol{R}_n)$ and $\mathcal{N}(\boldsymbol{m}_x, \boldsymbol{R}_x)$, conditionally on a set of hyperparameters and auxiliary variables.

In a general statement, acquisition parameters, hyperparameters and auxiliary variables are collected in $\boldsymbol{\theta}$. The general inverse problem then consists in estimating $\boldsymbol{x}$ and $\boldsymbol{\theta}$ through the posterior $p(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$. Its exploration can be achieved by means of a Gibbs sampler which iteratively samples from $p(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y})$ and $p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$. The conditional posterior $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\theta})$ is a correlated Gaussian distribution: $\mathcal{N}(\boldsymbol{m}_x^{\mathrm{post}}, \boldsymbol{R}_x^{\mathrm{post}})$ with

$$\begin{aligned}\boldsymbol{R}_x^{\mathrm{post}} &= \left(\boldsymbol{A}^t\boldsymbol{R}_n^{-1}\boldsymbol{A} + \boldsymbol{R}_x^{-1}\right)^{-1}\\ \boldsymbol{m}_x^{\mathrm{post}} &= \boldsymbol{R}_x^{\mathrm{post}}\left(\boldsymbol{A}^t\boldsymbol{R}_n^{-1}\left[\boldsymbol{y} - \boldsymbol{m}_n\right] + \boldsymbol{R}_x^{-1}\boldsymbol{m}_x\right)\end{aligned}$$

where $\boldsymbol{\theta}$ is embedded in $\boldsymbol{A}, \boldsymbol{R}_n$ and $\boldsymbol{R}_x$ for simpler notations.

If $\boldsymbol{A}$ has no particular properties, $\boldsymbol{Q} = (\boldsymbol{R}_x^{\mathrm{post}})^{-1}$ is neither sparse nor circulant, and existing sampling algorithms are not applicable. The PO algorithm makes it possible to sample from $\mathcal{N}(\boldsymbol{m}_x^{\mathrm{post}}, \boldsymbol{R}_x^{\mathrm{post}})$ by applying Algorithm 1 with $K = 2$, $\boldsymbol{M}_1 = \boldsymbol{A}$, $\boldsymbol{M}_2 = \boldsymbol{I}$, $\boldsymbol{R}_1 = \boldsymbol{R}_n$, $\boldsymbol{R}_2 = \boldsymbol{R}_x$, $\boldsymbol{m}_1 = \boldsymbol{m}_n$ and $\boldsymbol{m}_2 = \boldsymbol{m}_x$. In this context, it can be said that the optimization procedure converts prior samples into a posterior one.

III. Illustration

The proposed PO algorithm makes it possible to resort to stochastic sampling algorithms in inverse problems providing two main advances:

- capability to jointly estimate extra unknowns included in $\boldsymbol{\theta}$ (acquisition parameters, hyperparameters, …),
- access to the entire unknown distribution providing uncertainties (standard deviation, credibility interval,…).

These advances are illustrated in the present section.

A. Unsupervised super-resolution

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 [18, 19]. 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 thus to provide hyperparameter estimation.

We resort to the standard forward model in SR: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} = \boldsymbol{P}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$. In this equation, $\boldsymbol{y} \in \mathbb{R}^M$ collects the pixels of the low resolution images, here 5 images of 128×128 pixels ($M = 81920$) and $\boldsymbol{x} \in \mathbb{R}^N$ collects the pixels of the original image, here 256×256 pixels ($N = 65536$). The noise $\boldsymbol{n} \in \mathbb{R}^M$ accounts for measurement and modeling errors. $\boldsymbol{H}$ is a $N \times N$ circulant convolution matrix that accounts for the convolution part of the observation system. Practically, the impulse response is a Laplace shape with FWHM of 4 pixels. $\boldsymbol{P}$ is a $M \times N$ decimation matrix: it is a binary matrix indicating which pixel is observed. Finally, $\boldsymbol{A}$ is a $M \times N$ matrix (that is to say 81920×65536). The prior distribution for $\boldsymbol{n}$ is $\mathcal{N}(\boldsymbol{0}, \gamma_n^{-1}\boldsymbol{I})$ and the one for $\boldsymbol{x}$ is $\mathcal{N}(\boldsymbol{0}, \gamma_x^{-1}\boldsymbol{D}^t\boldsymbol{D})$ where $\boldsymbol{D}$ is the $N \times N$ circulant convolution matrix of the Laplacian filter. The hyperparameters γ_n and γ_x are *unknown* and their prior law are Jeffreys'. The posterior [9] is

$$p(\boldsymbol{x}, \gamma_n, \gamma_x | \boldsymbol{y}) \propto \gamma_n^{M/2-1} \gamma_x^{(N-1)/2-1} \exp\left[-\gamma_n \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{H}\boldsymbol{x}\|^2/2 - \gamma_x \|\boldsymbol{D}\boldsymbol{x}\|^2/2\right]. \quad (6)$$

It is explored by a Gibbs sampler: iteratively sampling γ_n, γ_x and $\boldsymbol{x}$ under their respective posterior conditional distribution

$$p(\gamma_n^{(k)} | \boldsymbol{x}, \gamma_x, \boldsymbol{y}) = \mathcal{G}\left(1 + M/2, 2/\left\|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{H}\boldsymbol{x}^{(k-1)}\right\|^2\right)$$

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

$$p(\boldsymbol{x}^{(k)} | \gamma_x, \gamma_n, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{m}_x^{\text{post}}, \boldsymbol{R}_x^{\text{post}})$$

$$\text{with} \quad \boldsymbol{R}_x^{\text{post}} = \left(\gamma_n^{(k)} \boldsymbol{H}^t \boldsymbol{P}^t \boldsymbol{P} \boldsymbol{H} + \gamma_x^{(k)} \boldsymbol{D}^t \boldsymbol{D}\right)^{-1}$$

$$\text{and} \quad \boldsymbol{m}_x^{\text{post}} = \gamma_n^{(k)} \boldsymbol{R}_x^{\text{post}} \boldsymbol{P}^t \boldsymbol{H}^t \boldsymbol{y}.$$

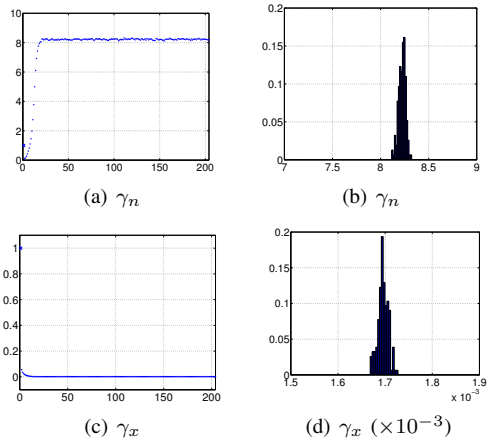
The conditional posteriors for the hyperparameters are Gamma distributions so they are easy to sample.

The conditional posterior for $\boldsymbol{x}$ is Gaussian, but the use of existing algorithms is impossible due to the structure and the size of $\boldsymbol{R}_x^{\text{post}}$. Regarding the structure, according to Section II-B, with $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{H}$: $\boldsymbol{A}$ is non-circulant due to the decimation and $\boldsymbol{A}$ is non-sparse due to large support of the impulse response. Regarding the size, $\boldsymbol{R}_x^{\text{post}}$ (and its Cholesky factor) is a huge $N \times N$ matrix, that is to say 65536×65536 and its footprint in memory would be 32 GB. As a consequence, neither the precision matrix nor its Cholesky factor can be stored on standard computers.

On the contrary, the proposed PO algorithm only requires the storage of four 256×256 matrices and its footprint in memory is only 2 MB that is easy to manage on standard computers. Regarding the computational cost:

- Step P requires a sample under each prior distribution: $\boldsymbol{x}$ is computed by FFT (see item 2 of Section I) and $\boldsymbol{n}$ is trivially computed since it is a white noise.
- Step O is achived by a conjugate gradient procedure with optimal step size. It only requires computations of convolutions (by FFT), decimation and zero-padding.

So, the proposed PO algorithm is feasible and it easily provides a desired sample. Practically, it takes[1] about one second (i.e. around $P = 50$ gradient iterations) to obtain one sample.



(a) γ_n (b) γ_n (c) γ_x (d) γ_x ($\times 10^{-3}$)

Fig. 1. Chains and histograms of hyperparameters γ_n and γ_x.

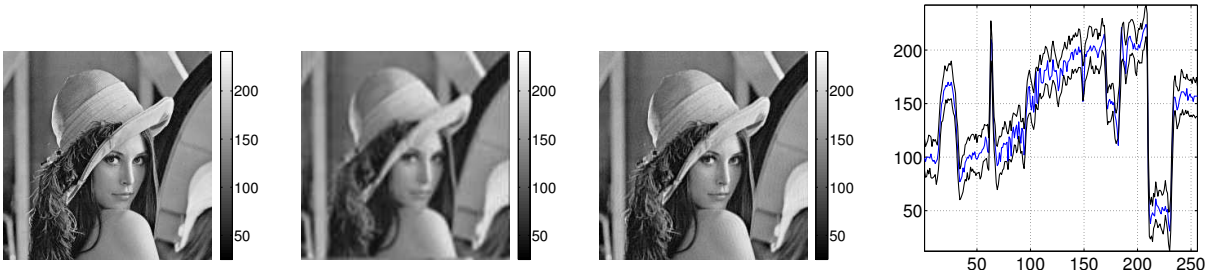
Fig. 1 shows the iterates and illustrates the operation and convergence. After a burn-in period of about 25 iterations, the algorithm is in its converged state and the total number of iterations is 59 to ensure a good exploration of the distribution. Histograms approximating marginal posteriors are also given and the posterior means are $\widehat{\gamma}_n \approx 7.7$ and $\widehat{\gamma}_x \approx 2.2 \times 10^{-3}$.

Concerning the images themselves, results are shown in Fig. 2: the estimated image in 2(c) clearly shows a better resolution than the data in 2(b) and it is visually close to the original image in 2(a). Nevertheless, it is important to keep in mind that, w.r.t. other SR methods, the proposed PO algorithm does not improve image quality itself but the crucial novelty is to allow for hyperparameter estimation. In this sense, it is 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): the true image is inside the 99% credibility interval around the estimate. As a conclusion, the proposed PO algorithm makes it possible to resort to sampling algorithms in SR method whereas it was not possible before. It then enables hyperparameter estimation while other SR methods require hand-made hyperparameter tuning. In addition, it enables to compute uncertainties based on posterior standard deviation.

B. Three other examples

The PO algorithm has been used in three other contexts: electromagnetic inverse scattering [14], fluorescent microscopy trough structured illumination [20] and super-resolution from data provided the Herschel observatory in astronomy [**?**].

[1]The algorithm is implemented within the computing environment Matlab on a PC with a 3 GHz CPU and 3 GB of RAM.



(a) True (b) Data (c) Estimate (d) Uncertainty

Fig. 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% credibility interval around the estimate.

The problems are tackled in a Bayesian framework and implemented by means of stochastic sampling. In these contexts, the distribution for the object given the other variables is Gaussian with large size precision matrix. Its structure is neither sparse nor circulant making the use of existing algorithms impossible. This is due to non-linearity and label variables in [14] and non-invariance of the observation model in [20, **?**]. Nevertheless, the precision matrix is in the form (1), so, the proposed PO is applicable.

IV. Conclusion

The paper presents an algorithm for sampling high-dimensional Gaussian distributions when existing algorithms are infeasible. 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 optimizer is a sample of the target distribution. The algorithm is applicable for a particular decomposition of the precision matrix that emerges in general linear inverse problems.

There is a wide class of applications, in particular any processing problem based on a conditional linear forward model and conditional Gaussian priors for noise and object. The interest and the feasibility of the proposed algorithm have been illustrated in [14, 20, **?**] and in this paper on a more academic super-resolution problem allowing automatic tuning of hyperparameters.

An interesting perspective deals with the case of stopped optimization procedure. It is a question under consderation to prove that, embedded in a Gibbs loop, a finite number (maybe one) of iteration of the optimization step is enough to guarantee convergence towards the target law.

V. Acknowledgment

The authors would like to thank J. Idier (IRRCyN) for inspiration of this work [21], T. Rodet and A. Djafari (L2S), for fruitful discussions, C. Vacar (IMS) and P. Szacherski (IMS and CEA-LETI) for carefully reading the paper.

References

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

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

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

[4] 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.

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

[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] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[8] 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.

[9] 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, 2010.

[10] H. Helgason, V. Pipiras, and P. Abry, "Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding," *Signal Processing*, 2011.

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

[12] H. Rue, S. Martino, and N. Chopin, "Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations," *J. R. Statist. Soc. B*, vol. 71, pp. 1–35, 2009.

[13] 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.

[14] 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.

[15] 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.

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

[17] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Series in Operations Research. New York: Springer Verlag, 2000.

[18] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Trans. Signal Processing Mag.*, pp. 21–36, May 2003.

[19] G. Rochefort, F. Champagnat, G. Le Besnerais, and J.-F. Giovannelli, "An improved observation model for SR under affine motion," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3325–3337, Nov. 2006.

[20] 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 press.

[21] J. Idier, "Optimization and sampling in the Gaussian case," Personnal communication, 2006.