



HAL
open science

A Data Augmentation Approach for Sampling Gaussian Models in High Dimension

Yosra Marnissi, Dany Abboud, Emilie Chouzenoux, Jean-Christophe Pesquet,
Mohamed El-Badaoui, Amel Benazza-Benyahia

► **To cite this version:**

Yosra Marnissi, Dany Abboud, Emilie Chouzenoux, Jean-Christophe Pesquet, Mohamed El-Badaoui, et al.. A Data Augmentation Approach for Sampling Gaussian Models in High Dimension. EUSIPCO 2019 - 27th European Signal Processing Conference, Sep 2019, La Corogne, Spain. 10.23919/eusipco.2019.8902496 . hal-02314418

HAL Id: hal-02314418

<https://hal.science/hal-02314418v1>

Submitted on 12 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Data Augmentation Approach for Sampling Gaussian Models in High Dimension

Y. Marnissi⁽¹⁾, D. Abboud⁽¹⁾, E. Chouzenoux⁽²⁾, J-C. Pesquet⁽²⁾, M. El-Badaoui^(1,3) and A. Benazza-Benyahia⁽⁴⁾

¹ SAFRAN TECH, Groupe Safran, 78772 Magny Les Hameaux Cedex, France.

² Center for Visual Computing, INRIA, CentraleSupélec, University Paris-Saclay, 91190 Gif sur Yvette, France.

³ University of Lyon, UJM-St-Etienne, LASPI, EA3059, F-42023, Saint-Etienne, France.

⁴ University of Carthage, SUP'COM, COSIM Lab., 2083 El Ghazala Ariana, Tunisia.

Abstract—Recently, methods based on Data Augmentation (DA) strategies have shown their efficiency for dealing with high-dimensional Gaussian sampling within Gibbs samplers compared to iterative-based sampling (e.g., Perturbation-Optimization). However, they are limited by the feasibility of the direct sampling of the auxiliary variable. This paper reviews DA sampling algorithms for Gaussian sampling and proposes a DA method which is especially useful when direct sampling of the auxiliary variable is not straightforward from a computational viewpoint. Experiments in two vibration analysis applications show the good performance of the proposed algorithm.

Index Terms—Data augmentation, Auxiliary variables, MCMC, Gaussian, Correlation, Bayesian.

I. INTRODUCTION

This paper deals with the problem of sampling from a high dimensional Gaussian distribution¹ with mean $\mathbf{m} \in \mathbb{R}^Q$ and precision matrix $\mathbf{G} = \sum_{j=1}^J \mathbf{G}_j \in \mathbb{R}^{Q \times Q}$ such that

$$(\forall j \in \{1, \dots, J\}) \quad \mathbf{G}_j = \mathbf{H}_j^\top \mathbf{\Lambda}_j \mathbf{H}_j, \quad (1)$$

where for $j \in \{1, \dots, J\}$, $\mathbf{\Lambda}_j \in \mathbb{R}^{N_j \times N_j}$ is a positive semi-definite matrix and $\mathbf{H}_j \in \mathbb{R}^{N_j \times Q}$. Very often, we do not have direct access to the mean \mathbf{m} but only to the potential vector $\mathbf{p} = \mathbf{G}\mathbf{m}$. We further assume that the latter reads $\mathbf{p} = \sum_{j=1}^J \mathbf{G}_j \mathbf{m}_j$ where for $j \in \{1, \dots, J\}$, $\mathbf{m}_j \in \mathbb{R}^{Q}$. Gaussian sampling arises in linear inverse problems involving Gaussian or hierarchical Gaussian models. In such situations, Gaussian simulation is mostly needed as a sampling step at each iteration of a Markov Chain Monte Carlo algorithm e.g., Gibbs samplers. Typical applications are image deconvolution [1], super-resolution [2], inpainting [3], [4], weather forecasting [5], etc.

The problem of high dimensional Gaussian sampling has been widely addressed in the literature and several solutions have been proposed. Typical approaches for large scale Gaussian simulation are inspired from deterministic iterative optimization techniques. The most known ones are samplers

derived from the Perturbation-Optimization strategy [2], [3], [6] and matrix splitting optimization [4], [7]. These two families of methods are good candidates for efficient Gaussian sampling in large scale problems since they avoid the storage and the factorization of large matrices. However, they both require solving a linear system at each iteration which can be computationally expensive. Thereby, when implemented through a Gibbs sampler, they may turn out to be less efficient; in particular, when \mathbf{G} depends on some target parameters evolving along the algorithm. Recently, new sampling strategies have been proposed as alternatives to optimization based Gaussian sampling [8], [9], [10]. By adding some auxiliary variables, the authors demonstrate, in several inverse problems applications, that sampling becomes much easier in the new augmented space. In the following, we will use the term Data Augmentation (DA) sampling to designate any sampling algorithm that introduces auxiliary variables [11].

In this paper, we are interested in DA strategies for Gaussian sampling in large scale problems. Section II reviews recent DA Gibbs algorithms that allow to separate heterogeneous correlations in the covariance matrix. The DA Gibbs sampler, initially introduced in [9], is extended to handle the limitation of state-of-the-art techniques. Section III compares the performance of the different DA samplers in vibration analysis applications. Finally, some conclusions are drawn in Section IV.

II. DATA AUGMENTATION STRATEGY FOR GAUSSIAN SAMPLING

A. Principle

The difficulty of Gaussian sampling arises particularly when the matrices $(\mathbf{H}_j, \mathbf{\Lambda}_j)_{1 \leq j \leq J}$ cannot be diagonalized in the same basis. Therefore, it is desirable to separate heterogeneous matrices in order to facilitate sampling. This has been successfully achieved using DA strategies [8], [9], [10]. Specifically, auxiliary variables $\mathbf{u} \in \mathbb{R}^P$, are added to the model² with a predefined joint distribution with density $q(\mathbf{x}, \mathbf{u})$. The key requirement is that $q(\mathbf{x}, \mathbf{u})$ should define a valid probability density function (i.e. non-negative whose integral with respect

¹Note that the degenerate Gaussian distribution case (i.e., whose covariance matrix is positive semi-definite but not with full rank) is considered as a proper distribution w.r.t. the restriction of the Lebesgue measure to the image subspace of this covariance matrix. This amounts to replacing inverse with generalized inverse and determinant with pseudo-determinant in the computation of the density function. Throughout this article, for deficient rank positive semi-definite matrices, $(\cdot)^{-1}$ will denote the generalized inverse.

²Initially, the model contains the main variable $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{G}^{-1})$ and possibly other latent variables controlling \mathbf{m} and/or \mathbf{G} .

TABLE I: Conditional distributions of \mathbf{x} and of the auxiliary variables.

DA	Conditions on parameters	First auxiliary variable $\mathbf{u} \mathbf{v}, \mathbf{x} \sim \mathcal{N}(\mathbf{m}_u, \mathbf{Q}_u)$	Second auxiliary variable $\mathbf{v} \mathbf{u}, \mathbf{x} \sim \mathcal{N}(\mathbf{m}_v, \mathbf{Q}_v)$	Resulting conditional distribution $\mathbf{x} \mathbf{u}, \mathbf{v} \sim \mathcal{N}(\mathbf{G}_x^{-1}\mathbf{p}_x, \mathbf{G}_x^{-1})$
SP	$0 < \mu$	$\mathbf{Q}_u = \left(\frac{1}{\mu}\mathbf{Id} + \mathbf{G}_1\right)^{-1}$ $\mathbf{m}_u = \mathbf{Q}_u \left(\frac{1}{\mu}\mathbf{x} + \mathbf{G}_1\mathbf{m}_1\right)$		$\mathbf{G}_x = \frac{1}{\mu}\mathbf{Id} + \sum_{j=2}^J \mathbf{G}_j$ $\mathbf{p}_x = \frac{1}{\mu}\mathbf{u} + \sum_{j=2}^J \mathbf{G}_j\mathbf{m}_j$
SPA	$0 < \eta < \mu$	$\mathbf{Q}_u = \left(\frac{1}{\eta}\mathbf{Id} + \mathbf{G}_1\right)^{-1}$ $\mathbf{m}_u = \mathbf{Q}_u \left(\frac{1}{\eta}(\mathbf{x} + \mathbf{v}) + \mathbf{G}_1\mathbf{m}_1\right)$	$\mathbf{Q}_v = \frac{\eta(\mu-\eta)}{\mu}\mathbf{Id}$ $\mathbf{m}_v = \frac{\mu-\eta}{\mu}(\mathbf{u} - \mathbf{x})$	$\mathbf{G}_x = \frac{1}{\eta}\mathbf{Id} + \sum_{j=2}^J \mathbf{G}_j$ $\mathbf{p}_x = \frac{1}{\eta}(\mathbf{u} - \mathbf{v}) + \sum_{j=2}^J \mathbf{G}_j\mathbf{m}_j$
EDA	$0 < \mu\ \mathbf{G}_1\ < 1$	$\mathbf{Q}_u = \frac{1}{\mu}\mathbf{Id} - \mathbf{G}_1$ $\mathbf{m}_u = \frac{1}{\mu}\mathbf{x} - \mathbf{G}_1\mathbf{x}$		$\mathbf{G}_x = \frac{1}{\mu}\mathbf{Id} + \sum_{j=2}^J \mathbf{G}_j$ $\mathbf{p}_x = \mathbf{u} + \sum_{j=1}^J \mathbf{G}_j\mathbf{m}_j$
GEDA	$0 < \mu\ \mathbf{G}_1\ < 1$	$\mathbf{Q}_u = \mu\mathbf{Id}$ $\mathbf{m}_u = (\mathbf{Id} - \mu\mathbf{G}_1)\mathbf{x} + \mu\mathbf{H}_1^T\Lambda_1\mathbf{v}$	$\mathbf{Q}_v = \Lambda_1^{-1}$ $\mathbf{m}_v = \mathbf{H}_1\mathbf{u}$	$\mathbf{G}_x = \frac{1}{\mu}\mathbf{Id} + \sum_{j=2}^J \mathbf{G}_j$ $\mathbf{p}_x = \frac{1}{\mu}\mathbf{u} - \mathbf{G}_1\mathbf{u} + \sum_{j=1}^J \mathbf{G}_j\mathbf{m}_j$

to \mathbf{u} and \mathbf{x} equals 1). This is ensured if the marginalization of $q(\mathbf{x}, \mathbf{u})$ with respect to \mathbf{u} and \mathbf{x} gives rise to valid marginal probability density functions $q(\mathbf{x}) = \int_{\mathbb{R}^M} q(\mathbf{x}, \mathbf{u}) d\mathbf{u}$ and $q(\mathbf{u}) = \int_{\mathbb{R}^Q} q(\mathbf{x}, \mathbf{u}) d\mathbf{x}$ [11]. At each iteration of the new Gibbs sampler, the Gaussian sampling step is replaced by two sampling steps from the conditional distributions of the two variables and in an arbitrary order. The DA strategy is said to be exact if the introduction of auxiliary variables does not alter the initial model. This is achieved if the marginal distribution $q(\mathbf{x})$ is equal to the target distribution (being in our case the Gaussian distribution of mean \mathbf{m} and covariance matrix \mathbf{G}^{-1}). Hereafter, examples of recently proposed DA methods are reviewed. These methods can be categorized into approximate and exact DA samplers.

B. Related works

Without loss of generality, we consider the problem of eliminating the coupling induced by \mathbf{G}_1 in the precision matrix (the methods can be easily generalized to other coupling matrices following the same lines). To this end, $q(\mathbf{x}, \mathbf{u})$ is selected, so as, in the new augmented space, \mathbf{G}_1 is no more coupled directly with \mathbf{x} but only intervenes through auxiliary variables. Table I summarizes the state-of-the-art DA strategies proposed for the aforementioned purpose.

1) *Approximate DA*: The split (SP) sampler [10], is derived from the deterministic variable splitting optimization strategy. The main idea is to split the initial model into the product of a pair of density functions, for example the likelihood and the prior distribution. Each function is expressed with respect to one variable namely the main variable \mathbf{x} and the splitting variable \mathbf{u} . A quadratic function $\phi(\mathbf{x}, \mathbf{u}; \mu) = \frac{1}{\mu}\|\mathbf{x} - \mathbf{u}\|^2$ with $\mu > 0$, is added to the model to control the discrepancy between the two parameters. This function can be particularly interpreted as the minus-logarithm of their joint distribution. Consequently, for the purpose of eliminating \mathbf{G}_1 in the distribution of \mathbf{x} , the resulting conditional distributions are summarized in Table I. It can be noticed that the additional splitting variables make the different precision matrices appear in two separate distributions $q(\mathbf{x}|\mathbf{u})$ and $q(\mathbf{u}|\mathbf{x})$. However, this splitting strategy is not exact which means that the resulting distribution of \mathbf{x} is only an approximation of the target one.

The two distributions only coincide in the limiting case when the variance μ of the Gaussian splitting variable goes to zero. However, the smaller μ , the higher the correlation between samples. Such a scenario can jeopardize the mixing properties of the samples. In [10], the authors propose to introduce an additional auxiliary variable \mathbf{v} in the SP Gibbs sampler to decrease the correlation between \mathbf{x} and \mathbf{u} . This allows the use of higher values for μ , with the aim to better approach the exact target distribution. The resulting sampler is known as the Split And Augmented (SPA) sampler. It is worth noting that SPA Gibbs sampler reduces to SP Gibbs sampler by integrating out \mathbf{v} . Further discussions and theoretical results on approximate DA Bayesian approaches can be found in recent works [12], [13].

2) *Exact DA*: Exact DA (EDA) strategies have been derived in [8], [9], [14] to cope with both the high dimensionality and the strong correlation existing between the target parameters in high dimensional Gaussian models. These methods are related to half-quadratic optimization approaches proposed in [15], as established in [8]. The resulting hierarchical Gibbs scheme is summarized in the third row of Table I. Note that only the matrices $(\mathbf{G}_j)_{j \neq 1}$ are still directly coupled to the main variable in the new augmented space, similarly to the splitting DA methods. Eventually, the advantage of this DA method compared to SP and SPA, is that the former is exact. Nevertheless, contrary to the EDA sampler, the splitting DA strategies are not restricted to Gaussian models (see [10] for examples).

C. A generalized exact DA strategy for Gaussian models

The main interest of DA strategies is that, in the new augmented space, the sampling task is much easier than direct sampling from the initial model. This can be achieved provided that the sampling step from $q(\mathbf{u}|\mathbf{x})$ does not introduce a too high computational cost in the algorithm. Ultimately, the reviewed DA methods fail when direct sampling is not feasible. The feasibility depends on the structure of matrix \mathbf{G}_1 . To alleviate these limitations, we propose a new DA Gibbs sampler that will be designated as the GEDA algorithm. More precisely, two auxiliary variables $\mathbf{u} \in \mathbb{R}^Q$ and $\mathbf{v} \in \mathbb{R}^{N_1}$ are introduced according to the following hierarchical model:

- $\mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{G}^{-1})$,
- $\mathbf{u}|\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mathbf{\Gamma}^{-1})$,
- $\mathbf{v}|\mathbf{u} \sim \mathcal{N}(\mathbf{H}_1\mathbf{u}, \mathbf{\Lambda}_1^{-1})$,

where \mathbf{G} and \mathbf{m} take the form presented in Section I and

$$\mathbf{\Gamma} = \frac{1}{\mu}\mathbf{Id} - \mathbf{G}_1, \quad 0 < \mu\|\mathbf{G}_1\| < 1. \quad (2)$$

Since \mathbf{v} is independent from \mathbf{x} conditionally to \mathbf{u} , the joint density distribution of these variables reads $q(\mathbf{x}, \mathbf{u}, \mathbf{v}) = q(\mathbf{v}|\mathbf{u})q(\mathbf{u}|\mathbf{x})q(\mathbf{x})$. In particular, its minus logarithm can be expressed up to an additive constant as follows:

$$\begin{aligned} \mathcal{J}(\mathbf{x}, \mathbf{u}, \mathbf{v}) &= \frac{1}{2\mu}\|\mathbf{x} - \mu\mathbf{G}_1\mathbf{m}_1\|^2 + \frac{1}{2}\sum_{j=2}^J \mathbf{x}^\top \mathbf{G}_j (\mathbf{x} - \mathbf{m}_j) \\ &+ \frac{1}{2\mu}\|\mathbf{u}\|^2 - \mathbf{x}^\top \mathbf{\Gamma}\mathbf{u} - \mathbf{v}^\top \mathbf{\Lambda}_1 \mathbf{H}_1 \mathbf{u} + \frac{1}{2}\mathbf{v}^\top \mathbf{\Lambda}_1 \mathbf{v}. \end{aligned} \quad (3)$$

The resulting conditional distributions of \mathbf{x} , \mathbf{u} and \mathbf{v} are given in Table I. It can be seen that the interest behind introducing the variable \mathbf{u} is to eliminate \mathbf{G}_1 in the covariance matrix of \mathbf{x} conditionally to \mathbf{u} while the introduction of the variable \mathbf{v} aims at facilitating the sampling of \mathbf{u} so that this variable can be drawn directly without requiring intensive computations. Similarly, the sampling step for \mathbf{v} can be performed efficiently for a large instance of inverse problems for which $\mathbf{\Lambda}_1$ is diagonal or has a simple structure.

Table II compares the different DA methods with respect to the feasibility of a direct sampling of the auxiliary variables when $\mathbf{\Lambda}_1 = \alpha\mathbf{Id}$ with $\alpha > 0$. This may particularly arises in linear inverse problems with decorrelated Gaussian noise. It can be noted that, in contrast with the approximate DA samplers, EDA sampler makes direct sampling possible when the matrix \mathbf{H}_1 is the product of two matrices belonging to diagonal, tight frame [16] or circulant families [9, Section 3.4]. This may arise for example in image recovery applications when $\mathbf{H}_1 = \mathbf{P}\mathbf{M}$ where \mathbf{P} is a tight frame analysis operator and \mathbf{M} is a convolution matrix with periodic boundary condition. Non-trivial forms of \mathbf{H}_1 arise in several applications such as compressive sensing [17], spectroscopy [18], image reconstruction [16], etc. In such situations, only the GEDA algorithm enables efficient sampling of all variables (see Section III-B for an example).

TABLE II: The feasibility of direct sampling if $\mathbf{\Lambda}_1 = \alpha\mathbf{Id}$.

\mathbf{H}_1	SP	SPA	EDA	GEDA
Diagonal	✓	✓	✓	✓
Tight frame	✓	✓	✓	✓
Circulant	✓	✓	✓	✓
Product			✓	✓
Non-trivial form				✓

It is worth noting that, the EDA algorithm can be viewed as a particular instance of the proposed Gibbs algorithm by marginalizing with respect to \mathbf{v} . Thus, one might prefer to use the EDA sampler rather than the GEDA if direct sampling of the auxiliary variables is tractable. Indeed, it is expected

that this marginalization improves the mixing properties of the samples. The proposed GEDA method can be used as an alternative to the EDA algorithm when direct sampling of the auxiliary variable is not feasible. In Section III, we show that the GEDA sampler still performs well when compared to the reviewed DA strategies even if direct sampling of the auxiliary variable is feasible. Interestingly and following the same lines as for the EDA sampler in [9], the GEDA algorithm can be also easily generalized to cases when the distribution of interest is non-Gaussian but its minus-logarithm density comprises a quadratic function with respect to some variables controlling the mean and/or the variance (e.g., location or scale mixture of Gaussian, Gaussian Markov random fields etc) by including these variables in the Gibbs scheme.

III. APPLICATION TO VIBRATION ANALYSIS

A. Order tracking

In a rotating machine, each mechanical component generates unique vibration patterns as the machine operates. It is a common practice to monitor these components by analyzing their vibratory level, provided that the system kinematic is known (i.e. the frequencies of the monitored components). This reduces to an amplitude and phase modulations estimation problem. Such problem is known in the literature as order tracking [19].

1) *Problem formulation:* The order tracking problem can be addressed with the following dynamic model [20]:

$$\mathbf{y}(n) = \sum_{k=1}^K \mathbf{s}_k(n) + \mathbf{w}(n) \quad (4)$$

where $n \in \{1, \dots, N\}$ denotes time index, $k \in \{1, \dots, K\}$ labels the sinusoidal components \mathbf{s}_k . In model (4), $\mathbf{y}(n) \in \mathbb{R}^2$ contains the real and the imaginary part of the Hilbert transform of the measured vibration data, $\mathbf{s}_k(n) = \mathbf{C}_k(n)\mathbf{x}_k(n)$ where $\mathbf{C}_k(n) \in \mathbb{R}^{2 \times 2}$ is given by $\mathbf{C}_k(n) = [\Psi(\phi_k(n)), \Psi(\phi_k(n) + \frac{\pi}{2})]$, $\phi_k(n)$ is the instantaneous phase of the k^{th} component, $\Psi(\cdot) = [\cos(\cdot), \sin(\cdot)]^\top$ and $\mathbf{w}(n)$ is assumed to be a zero-mean Gaussian noise with variance σ^2 . Let $\mathbf{y} = [\mathbf{y}(1)^\top, \dots, \mathbf{y}(N)^\top]^\top$ and $\mathbf{x} = [\mathbf{x}_1^\top, \dots, \mathbf{x}_K^\top]^\top$ where $\mathbf{x}_k = [\mathbf{x}_k(1)^\top, \dots, \mathbf{x}_k(N)^\top]^\top$. Furthermore, let $\mathbf{a}_k = [[\mathbf{x}_k(1)]_1, \dots, [\mathbf{x}_k(N)]_1]^\top$ and $\mathbf{b}_k = [\mathbf{x}_k(1)]_2, \dots, [\mathbf{x}_k(N)]_2]^\top$. Note that \mathbf{a}_k and \mathbf{b}_k can be extracted from \mathbf{x}_k using some suitable sparse matrices \mathbf{P}_1 and \mathbf{P}_2 i.e. $\mathbf{a}_k = \mathbf{P}_1\mathbf{x}_k$ and $\mathbf{b}_k = \mathbf{P}_2\mathbf{x}_k$. To perform the estimation, it is advantageous to consider the low-frequency or, equivalently, the slow-varying part of the amplitude and phase modulations profiles. This can be practically made by adding the following smoothing prior:

$$(\forall k \in \{1, \dots, K\}) \quad g(\mathbf{x}_k) \propto \gamma_k^N \exp(-\gamma_k \mathbf{x}_k^\top \mathbf{B} \mathbf{x}_k) \quad (5)$$

where $\mathbf{B} = \mathbf{P}_1^\top \mathbf{L}^\top \mathbf{L} \mathbf{P}_1 + \mathbf{P}_2^\top \mathbf{L}^\top \mathbf{L} \mathbf{P}_2$ such that $\mathbf{L} = \delta \mathbf{Id} + \nabla$, $\nabla \in \mathbb{R}^{N \times N}$ is a circulant matrix associated to a discrete Laplacian filter, $\delta > 0$ is a small constant that ensures the positive definiteness of $\mathbf{L}^\top \mathbf{L}$, and $\gamma_k > 0$ is a regularization parameter. Furthermore, we use an inverse Gamma prior

distribution for σ^2 and Gamma prior distributions for the regularization parameters i.e., $\sigma^2 \sim \mathcal{IG}(a, b)$ and for every $k \in \{1, \dots, K\}$, $\gamma_k \sim \mathcal{G}(a_k, b_k)$ where a, b, a_k, b_k are positive constants that are set in practice to small values to ensure weakly informative priors. From the observation and the prior models, the posterior of the target signal \mathbf{x} reduces to a Gaussian one with precision matrix and potential given respectively by

$$\mathbf{G} = \frac{1}{\sigma^2} \mathbf{C}^\top \mathbf{C} + \mathbf{M}, \quad \mathbf{p} = \frac{1}{\sigma^2} \mathbf{C}^\top \mathbf{y} \quad (6)$$

where $\mathbf{C} = [\mathbf{C}_1, \dots, \mathbf{C}_K]$, \mathbf{C}_k is a block matrix formed by the matrices $\mathbf{C}_k(1) \dots, \mathbf{C}_k(N)$ and \mathbf{M} is a block matrix formed by the matrices $\gamma_1 \mathbf{B}, \dots, \gamma_K \mathbf{B}$. Note that the precision matrix in (6) reduces to (1) by setting $J = 3$, $\mathbf{H}_1 = \mathbf{C}$, $\mathbf{\Lambda}_1 = \frac{1}{\sigma^2} \mathbf{Id}$, and, for every $j \in \{2, 3\}$, $\mathbf{\Lambda}_j$ is the diagonal matrix containing γ_k and \mathbf{H}_j is the block matrix formed by K blocks of $\mathbf{L}\mathbf{P}_j$. The posterior distributions of the remaining parameters are given by:

- $\sigma^2 | \mathbf{x}, \mathbf{y} \sim \mathcal{IG}(a + N, b + \frac{1}{2} \|\mathbf{C}\mathbf{x} - \mathbf{y}\|^2)$,
- $\forall k \in \{1, \dots, K\} \quad \gamma_k | \mathbf{x}_k \sim \mathcal{G}(a_k + N, b_k + \mathbf{x}_k^\top \mathbf{B} \mathbf{x}_k)$

2) *Gibbs samplers*: Since \mathbf{M} and \mathbf{C} cannot be diagonalized in the same basis, direct sampling from the Gaussian distribution with parameters (6) is intractable. Thus, we propose to resort to DA strategies. In particular, we aim at eliminating the coupling induced by $\mathbf{C}^\top \mathbf{C}$ in the posterior precision matrix of \mathbf{x} . Since $\mathbf{C}\mathbf{C}^\top = \sum_{k=1}^K \mathbf{C}_k \mathbf{C}_k^\top = K \mathbf{Id}$, $\|\mathbf{C}\mathbf{C}^\top\| = K$ and direct sampling of the Gaussian auxiliary variables in the EDA Gibbs sampler is straightforward. For the SP and SPA algorithms, an explicit expression of the covariance matrix of the auxiliary variable \mathbf{u} can be found by using the Woodbury matrix identity so that, similarly to EDA, direct sampling of the auxiliary variables can be fulfilled easily. Regarding the sampling step of the target signal, it can be noted that for all the DA strategies, the different components $\mathbf{a}_1, \dots, \mathbf{a}_K$, $\mathbf{b}_1, \dots, \mathbf{b}_K$ are uncorrelated given the auxiliary variables and their covariance matrices are circulant so that they can be drawn easily, in an independent manner, in the Fourier domain. As it is complicated to sample from the conditional distributions of the parameters $\sigma^2, \gamma_1, \dots, \gamma_K$ subject to the auxiliary variables, we follow [9] i.e., we instead sample from their marginalized distributions by partially collapsing all the auxiliary variables.

3) *Performance comparison on synthetic data*: We consider a synthetic signal containing 3 components with time-varying instantaneous amplitudes and frequencies and their 4 first integer multiple harmonics over a duration of 4 seconds and at a sampling frequency of 3,000Hz. Thus, $K = 15$ and $N = 12,000$. A Gaussian noise with variance $\sigma^2 = 7605$ is artificially added to the signal so that the initial signal-to-noise ratio is equal to 0 dB. The hyperparameters $a, b, a_k, b_k, k \in \{1, \dots, K\}$ are set to zero to ensure non-informative priors. Simulations were performed on an Intel(R) Core(TM) i5-6300U CPU @ 2.40 GHz, using a Matlab 2014 implementation. The spectrogram of the noisy signal and the estimated one (using the empirical average of 1,000 samples

generated by the GEDA algorithm after convergence) are shown in Figure 1.

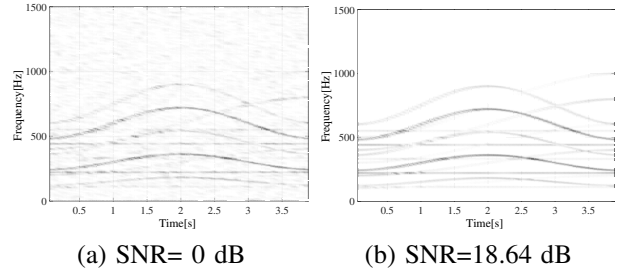


Fig. 1: Spectrogram. (a) Noisy signal (b) Estimated signal.

Figure 2 shows the evolution of the parameter σ^2 with respect to the computational time for the considered DA sampling algorithms for different values of μ and η (given here up to a multiplicative factor σ^2). One can see that the best convergence speed is achieved by the different samplers for very small value of μ while SP appears to converge towards a wrong distribution for higher values of μ (i.e., $\mu \geq 0.1$). These results are consistent with the findings highlighted in [10]. The remaining samplers share a roughly similar convergence speed.

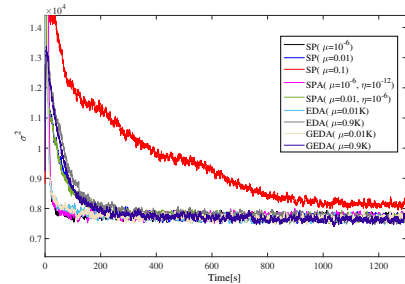


Fig. 2: Evolution of σ^2 with respect to time.

Table III shows the mixing results for the DA algorithms in terms of time per iteration after the burn-in period, Mean Square Jump (MSJ) in stationarity and MSJ per second. Note that the MSJ is estimated with an empirical average over 1,000 samples after convergence similarly to [9]. It can be noted that all algorithms have the same iteration cost and share good mixing properties except SP and SPA for low values of μ . Compared to SP, the addition of auxiliary variables in SPA enhances slightly the mixing but the two algorithms still explore less efficiently the parameter space compared to exact DA samplers. In particular EDA is twice more efficient than SP and SPA. Moreover, it appears that the best mixing properties for EDA and GEDA are achieved for large values of μ . It is worth noting that EDA shows mixing properties slightly better than GEDA which is expected since EDA is a marginalized version of GEDA. However, one should recall that GEDA covers a wider scope of Gaussian sampling problems than EDA. It follows that GEDA seems to be a good candidate to replace EDA in Gaussian sampling when the covariance matrix does not satisfy the requirements in [9].

TABLE III: Mixing results of DA samplers.

	T[s]	MSJ	MSJ/T
SP ($\mu = 10^{-6}$)	0.14	12.86	86.39
SP ($\mu = 10^{-2}$)	0.14	415.48	2836.19
SPA ($\mu = 10^{-6}, \eta = 10^{-12}$)	0.15	12.93	80.63
SPA ($\mu = 10^{-2}, \eta = 10^{-6}$)	0.15	433.64	2776.62
EDA ($\mu = 0.01K$)	0.12	153.02	1275.16
EDA ($\mu = 0.9K$)	0.12	817.32	6438.12
GEDA ($\mu = 0.01K$)	0.14	149.48	1067.71
GEDA ($\mu = 0.9K$)	0.13	598.01	4326.98

B. A more complex illustrative scenario

For illustrative purpose, we consider the compressive sensing model in Section 4.2 of [17] to reconstruct a vibration signal of length $Q = 30,000$ acquired in a spur Gearbox from a low number of measurements $N = 5,000$. The reconstruction requires to specify a sparse representation operator for the vibration signal which is here achieved by the Fast Fourier transform. Following [17], to promote compressible solutions, the Fourier coefficients of the signal are assumed to be i.i.d according to a zero mean scale mixture of Gaussian distributions with a Gamma mixing density, which is equivalent to the Student's t distribution. Thus, the problem amounts to a Gaussian sampling problem where the precision matrix is of the form (1) with $J = 2$, $\mathbf{H}_1 = \mathbf{S}\Psi$ where Ψ is the inverse Fast Fourier transform operator and \mathbf{S} is a random Gaussian projection matrix with a reduced dimension $N = 5,000$, and $\mathbf{\Lambda}_1$ and \mathbf{G}_2 are diagonals. It is clear that as \mathbf{H}_1 has a non-trivial form, in particular because \mathbf{S} is neither diagonal, nor circulant nor a tight frame, the state-of-the-art DA techniques fail to sample directly the target variables, while the proposed GEDA method can still be applied. Figure 3 shows the initial and the reconstructed spectra with the GEDA sampler ($\mu = 0.9\|\mathbf{G}_1\|^{-2}$) in the frequency band (1,000, 4,000)Hz. The algorithm needs about 1,000 iterations to converge which is equivalent approximately to 500 seconds.

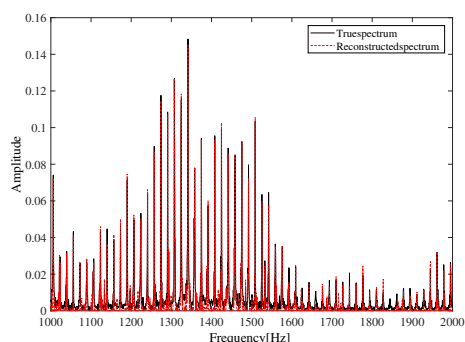


Fig. 3: Target and reconstructed spectra.

IV. CONCLUSION

This paper has reviewed recent DA strategies for Gaussian sampling and proposed a new one that can be used as an alternative to the method introduced in [9] when direct sampling of

the auxiliary variable is not feasible. It relies on adding two auxiliary variables: while the first auxiliary variable aims at facilitating the sampling of the target signal, the second one enables direct sampling of the first auxiliary variable. Simulation results in two vibration analysis applications indicate the good performance of the considered DA Gibbs samplers.

REFERENCES

- [1] C. Fox and R. A. Norton, "Fast sampling in a linear-Gaussian inverse problem," *SIAM-ASA J. Uncertain.*, vol. 4, no. 1, pp. 1191–1218, 2016.
- [2] C. Gilavert, S. Moussaoui, and J. Idier, "Efficient Gaussian sampling for solving large-scale inverse problems using MCMC," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 70–80, 2015.
- [3] G. Papandreou and A. L. Yuille, "Gaussian sampling by local perturbations," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS 2010)*, (Vancouver, Canada), pp. 1858–1866, 6–11 Dec 2010.
- [4] A.-C. Barros, F. Caron, J.-F. Giovannelli, and A. Doucet, "Clone MCMC: Parallel High-Dimensional Gaussian Gibbs Sampling," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS 2017)*, (Long Beach, CA), pp. 5020–5028, 4–9 Dec 2017.
- [5] Y. Gel, A. E. Raftery, and T. Gneiting, "Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method," *J. A. Stat. Assoc.*, vol. 99, no. 467, pp. 575–583, 2004.
- [6] O. Féron, F. Orieux, and J.-F. Giovannelli, "Gradient Scan Gibbs Sampler: an efficient algorithm for high-dimensional Gaussian distributions," *IEEE Journal of Selected Topics in Signal Process.*, vol. 10, no. 2, pp. 343–352, 2016.
- [7] M. Johnson, J. Saunderson, and A. Willsky, "Analyzing Hogwild Parallel Gaussian Gibbs sampling," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS 2013)*, (Harrah and Harveys, Lake Tahoe), pp. 2715–2723, 5–10 Dec 2013.
- [8] Y. Marnissi, E. Chouzenoux, J.-C. Pesquet, and A. Benazza-Benyahia, "An auxiliary variable method for langevin based MCMC algorithms," in *Proc. IEEE Stat. Signal Process. Workshop (SSP 2016)*, (Palma de Mallorca, Spain), pp. 297–301, 26–29 Jun. 2016.
- [9] Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, and J.-C. Pesquet, "An Auxiliary Variable Method for Markov Chain Monte Carlo Algorithms in High Dimension," *Entropy*, vol. 20, no. 2, 2018.
- [10] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler-Application to large-scale inference problems," *arXiv preprint arXiv:1804.05809*, 2018.
- [11] D. A. Van Dyk and X.-L. Meng, "The art of data augmentation," *J. Comput. Graph. Stat.*, vol. 10, no. 1, pp. 1–50, 2001.
- [12] M. Vono, N. Dobigeon, and P. Chainais, "Asymptotically exact data augmentation: models, properties and algorithms," *arXiv preprint arXiv:1902.05754*, 2019.
- [13] S. Sisson, Y. Fan, and M. Beaumont, "Overview of approximate Bayesian computation," *arXiv preprint arXiv:1802.09720*, 2018.
- [14] R. Cavicchioli, C. Chaux, L. Blanc-Féraud, and L. Zanni, "ML estimation of wavelet regularization hyperparameters in inverse problems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2013)*, (Vancouver, Canada), pp. 1553–1557, 26–31 May 2013.
- [15] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A 11-unified variational framework for image restoration," in *Proc. European Conference on Computer Vision (ECCV 2004)*, (Prague, Czech Republic), pp. 1–13, 11–14 May 2004.
- [16] N. Pustelnik, A. Benazza-Benyahia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based image deconvolution and reconstruction," *Wiley Encyclopedia of Electrical and Electronics Engineering*, 2016.
- [17] R. Fuentes, C. Mineo, S. G. Pierce, K. Worden, and E. J. Cross, "A probabilistic compressive sensing framework with applications to ultrasound signal processing," *Mech. Syst. Signal Process.*, vol. 117, pp. 383–402, 2019.
- [18] P. J. Brown, T. Fearn, and M. Vannucci, "Bayesian wavelet regression on curves with application to a spectroscopic calibration problem," *J. A. Stat. Assoc.*, vol. 96, no. 454, pp. 398–408, 2001.
- [19] H. Vold, H. Herlufsen, M. Mains, and D. Corwin-Renner, "Multi axle order tracking with the Vold-Kalman tracking filter," *Sound Vib.*, vol. 31, no. 5, pp. 30–35, 1997.
- [20] R. Turner and M. Sahani, "Probabilistic amplitude and frequency demodulation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS 2011)*, (Granada Spain), pp. 981–989, 12–17 Dec 2011.