



HAL
open science

A Partially Collapsed Gibbs Sampler for Unsupervised Nonnegative Sparse Signal Restoration

Mehdi Amrouche, Hervé Carfantan, Jérôme Idier

► **To cite this version:**

Mehdi Amrouche, Hervé Carfantan, Jérôme Idier. A Partially Collapsed Gibbs Sampler for Unsupervised Nonnegative Sparse Signal Restoration. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Jun 2021, Toronto, Canada. pp.5519-5523, 10.1109/icassp39728.2021.9414293 . hal-03227140

HAL Id: hal-03227140

<https://hal.science/hal-03227140>

Submitted on 17 May 2021

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 PARTIALLY COLLAPSED GIBBS SAMPLER FOR UNSUPERVISED NONNEGATIVE SPARSE SIGNAL RESTORATION

M.C. Amrouche¹, H. Carfantan¹ and J. Idier²

¹Institut de Recherche en Astrophysique et Planétologie,
Université de Toulouse, CNRS/UPS/CNES, Toulouse, France

²Laboratoire des Sciences du Numérique de Nantes, CNRS/ECN, Nantes, France
mamrouche@irap.omp.eu, hcarfantan@irap.omp.eu, jerome.idier@ls2n.fr

Abstract— In this paper the problem of restoration of unsupervised nonnegative sparse signals is addressed in the Bayesian framework. We introduce a new probabilistic hierarchical prior, based on the Generalized Hyperbolic (GH) distribution, which explicitly accounts for sparsity. On the one hand, this new prior allows us to take into account the non-negativity. On the other hand, thanks to the decomposition of GH distributions as continuous Gaussian mean-variance mixture, a partially collapsed Gibbs sampler (PCGS) implementation is made possible, which is shown to be more efficient in terms of convergence time than the classical Gibbs sampler.

1. INTRODUCTION

This paper tackles the restoration of a sparse nonnegative signal, observed through a linear operator and corrupted by additive white noise. The $N \times 1$ observed signal \mathbf{y} can be written as $\mathbf{y} = \mathbf{H}\mathbf{z} + \epsilon$, where \mathbf{z} is a $K \times 1$ nonnegative sparse signal, \mathbf{H} is a $N \times K$ matrix and ϵ models the perturbations.

In the literature, sparse signal restoration problems arise in different fields such as reflection seismology, astronomy and compressed sensing. The objective is to find a sparse representation of a signal \mathbf{y} that is a linear combination of a limited number of elements (atoms) taken from a given dictionary \mathbf{H} . This problem is often referred to as subset selection because it consists in selecting a subset of columns of \mathbf{H} . Mathematically, this can be formulated as the minimization of the squared error $\|\mathbf{y} - \mathbf{H}\mathbf{z}\|^2$ subject to $\|\mathbf{z}\|_0 < S$, where $\|\cdot\|$ and $\|\cdot\|_0$ respectively stands for the Euclidean norm, and the ℓ_0 pseudo-norm. This yields a combinatorial discrete problem known to be NP-hard [1].

One alternative is the *convex relaxation* of the problem which substitutes the ℓ_0 pseudo-norm by the ℓ_1 norm [2, 3], the sparsity of the solutions coming from the non-smooth character of the ℓ_1 norm at zero. *Greedy algorithms*, such as *Matching Pursuit* and its improved versions *Orthogonal Matching Pursuit* and *Orthogonal Least Squares* [2, 4], form another class of methods. Here the main idea is to iteratively recover the set of active atoms by incremental selection. Nonnegative adaptations have been proposed for both *convex relaxation* and *greedy algorithms*. For example, the ℓ_1 relaxation can be easily extended to the nonnegative setting [5, 6]. Nonnegative extensions of

greedy algorithms have been also introduced [7, 8].

On the other hand, a hierarchical Bayesian model which explicitly accounts for sparsity has been proposed for spike deconvolution, namely the Bernoulli-Gaussian (BG) model [9, 10]. Deterministic optimization algorithms [10] and Markov chain Monte Carlo techniques (MCMC) [11] are used to compute, respectively, the Maximum a Posteriori (MAP) and the Posterior Mean (PM) estimators. In the sparse train deconvolution context, where the dictionary is often strongly correlated, the PM estimator has been shown empirically to give better results than greedy algorithms and convex relaxation [12]. Moreover, the Bayesian framework allows one to estimate the model hyperparameters (unsupervised case).

In the MCMC framework, a nonnegative adaptation of the BG model was studied in [13], based on a Bernoulli-Truncated-Gaussian (BTG) model. Vector \mathbf{z} is modelled as a couple of variables (\mathbf{q}, \mathbf{x}) , where \mathbf{q} is a vector of independent, identically distributed (i.i.d.) Bernoulli variables, and x_k is distributed according to a centered truncated Gaussian when $q_k = 1$, otherwise $x_k = 0$. Posterior Mean estimation of \mathbf{q} and \mathbf{x} is then obtained using an MCMC method. The latter yields satisfactory results, but it requires a high computational cost. In [14, 15], a partially collapsed Gibbs sampler (PCGS) inspired from [16] is proposed to obtain a faster convergence rate in the case of BG deconvolution. It is based on marginalizing out the amplitudes \mathbf{x} from the joint posterior, which is practically possible because their posterior conditional is Gaussian. Unfortunately, partially collapsed sampling cannot be adapted as easily to the context of nonnegative restoration.

The aim of this paper is to propose a partially collapsed scheme in the context of nonnegative sparse restoration. Our main contribution is a new prior based on the *Generalized Hyperbolic* (GH) distribution. On the one hand, the proposed prior mimics the BTG prior to take the non-negativity into account. On the other hand, the decomposition of GH distributions as continuous Gaussian mean-variance mixtures allows us to marginalize the amplitudes, enabling the construction of a partially collapsed Gibbs sampler which is shown hereafter to converge more rapidly than the BTG version.

The organization of this paper is as follows. An overview of the Generalized Hyperbolic distribution and its properties are given in Section 2. Section 3 introduces the Bernoulli-

Generalized-Hyperbolic (BGH) model in the Bayesian framework. Its corresponding partially collapsed Gibbs sampler is presented in Section 4. Simulation results are given in Section 5 to compare the BGH and the BTG samplers both in terms of signal restoration and convergence. Finally, conclusions are drawn in Section 6.

2. GENERALIZED HYPERBOLIC DISTRIBUTIONS

2.1. Description and properties

Generalized Hyperbolic (GH) distributions form a five parameters family $\nu = (\lambda, \alpha, \beta, \delta, \mu)$ introduced by Barndorff-Nielsen in [17]. For a random variable $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ the probability density is given by

$$p_X(x) = \frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)} \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{(\sqrt{\delta^2 + (x-\mu)^2}/\alpha)^{\frac{1}{2}-\lambda}} e^{\beta(x-\mu)},$$

where $\gamma^2 = \alpha^2 - \beta^2$ and K_λ is the modified Bessel function of third kind. We will take advantage of three interesting properties of the GH distributions:

P1) GH distributions are invariant under affine transformations. If $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$, then $Y = aX + b \sim GH(\lambda, \alpha/a, \beta/a, a\delta, a\mu + b)$.

P2) They can be expressed as continuous normal mean-variance mixtures $p_X(x) = \int_0^\infty p_{X|W}(x|w) p_W(w) dw$ where $X|W \sim \mathcal{N}(\mu + \beta W, W)$ and each random variance W has a *Generalized Inverse Gaussian* [18] distribution $GIG(\lambda, \gamma, \delta)$.

Therefore, a GH random variable can be seen as a hierarchical model: in order to generate a GH sample, a random variance $W \sim GIG(\lambda, \gamma, \delta)$ is first drawn, and then $X|W \sim \mathcal{N}(\mu + \beta W, W)$ is drawn.

P3) The GIG distribution is a limiting case of the GH family [19]: if $X \sim GH\left(\lambda, \beta + \frac{\psi^2}{2}, \beta, \tau\left(\beta + \frac{\psi^2}{2}\right)^{-\frac{1}{2}}, 0\right)$ and $\beta \rightarrow \infty$, then $X \sim GIG(\lambda, \psi, \tau)$ i.e., β is a skewness parameter with $P(X \leq 0) = 0$ when $\beta \rightarrow \infty$. Also it is worth mentioning that the variances of the Gaussian involved in the mixture are inversely proportional to β . Meaning that large values of β implies highly correlated X and W variables.

2.2. Truncated Gaussian approximation

Here, we would like to approximate a truncated Gaussian distribution $\mathcal{N}^+(0, 1)$ with a GH one, while satisfying a trade-off. On the one hand, a distance between the two distributions must be made small. Here, we consider the total variation

$$TV(p, q) = \int_{-\infty}^{\infty} |p(x) - q(x)| dx. \quad (1)$$

In particular, small values of $TV(GH(\nu), \mathcal{N}^+(0, 1))$ impose that the GH law has a small mass on the negative half-line \mathbb{R}^- . Obviously, this is a desirable feature with respect to the non-negativity constraint. However, putting a too small mass on \mathbb{R}^-

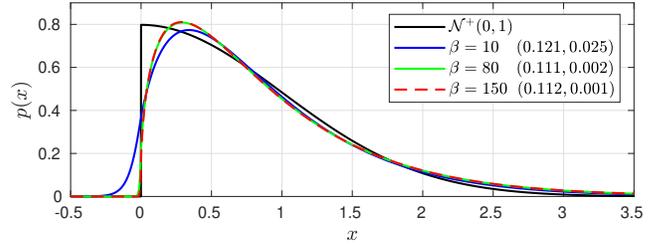


Fig. 1: Approximation of the standard truncated Gaussian with a GH distributions. Total variation between each approximation and the truncated Gaussian and $P(X \leq 0)$ are given in the following fashion ($TV, P(X \leq 0)$).

happens to be counterproductive in terms of algorithmic efficiency, because it induces strongly coupled pairs (W, X) , which annihilates the benefit from marginalizing the amplitude variables.

In practice, we propose to proceed as follows. We consider that parameter β plays a distinct role, since it allows us to control the mass put on \mathbb{R}^- . For fixed values of β , we numerically minimize $TV(GH(\nu), \mathcal{N}^+(0, 1))$ with respect to the remaining parameters $\nu_{-\beta} = (\lambda, \alpha, \delta, \mu)$. Figure 1 illustrates the results of such approximations for $\beta \in \{10, 80, 150\}$ and gives the associated values of TV and $P(X \leq 0)$. We will discuss the value of β in Section 5.

Finally, let $\nu_{\mathcal{N}} = (\lambda_{\mathcal{N}}, \alpha_{\mathcal{N}}, \beta_{\mathcal{N}}, \delta_{\mathcal{N}}, \mu_{\mathcal{N}})$ be the parameters of the GH approximation of $\mathcal{N}^+(0, 1)$. From P1, we can deduce a GH approximation of $\mathcal{N}^+(0, \sigma_x^2)$ for any $\sigma_x > 0$ according to:

$$GH(\lambda_{\mathcal{N}}, \alpha_{\mathcal{N}}/\sigma_x, \beta_{\mathcal{N}}/\sigma_x, \delta_{\mathcal{N}}\sigma_x, \mu_{\mathcal{N}}\sigma_x). \quad (2)$$

For the sake of simplicity, as the parameters are fixed given σ_x , we will write hereafter $GH_{\mathcal{N}}(\sigma_x^2)$ instead of (2) and $GIG_{\mathcal{N}}(\sigma_x^2)$ instead of $GIG(\lambda_{\mathcal{N}}, \gamma_{\mathcal{N}}/\sigma_x, \delta_{\mathcal{N}}\sigma_x)$.

3. BAYESIAN FRAMEWORK

3.1. Bernoulli-Generalized-Hyperbolic prior

We consider that $\mathbf{z} = (\mathbf{q}, \mathbf{x}, \mathbf{w})$ where $\forall k = 1, \dots, K, q_k \sim \mathcal{B}(\xi)$, and:

$$\begin{aligned} \text{when } q_k = 1, & \begin{cases} w_k \sim GIG_{\mathcal{N}}(\sigma_x^2) \\ x_k|w_k \sim \mathcal{N}(\sigma_x\mu_{\mathcal{N}} + w_k\beta_{\mathcal{N}}/\sigma_x, w_k), \end{cases} \\ \text{when } q_k = 0, & \quad x_k = 0 \end{aligned}$$

such that $x_k|q_k = 1 \sim GH(\nu_{\mathcal{N}})$, and the sparsity level of \mathbf{z} is tuned by the parameter $\xi = P(q_k = 1)$,

3.2. Posterior distribution

We consider an i.i.d. noise vector $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$. The posterior distribution can be written as follows:

$$p(\mathbf{q}, \mathbf{x}, \mathbf{w}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left(-\frac{\|\mathbf{y} - \bar{\mathbf{H}}\bar{\mathbf{x}}\|^2}{2\sigma^2}\right) p(\mathbf{x}, \mathbf{w}|\mathbf{q}, \boldsymbol{\theta}) P(\mathbf{q}|\boldsymbol{\theta}) \quad (3)$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{H}}$ respectively gathers the entries x_k and the columns \mathbf{H}_k for which $q_k = 1$, and $\boldsymbol{\theta} = [\xi, \sigma^2, \sigma_x^2]^T$ are the hyper-parameters.

3.3. Partially marginalized posterior distribution

From (3), one can deduce that $\bar{\mathbf{x}}|\mathbf{q}, \mathbf{w}, \mathbf{y}, \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Gamma})$, with

$$\boldsymbol{\eta} = \sigma^{-2} \boldsymbol{\Gamma} \bar{\mathbf{H}}^T \mathbf{y} + \boldsymbol{\Gamma} \mathbf{W}^{-1} \boldsymbol{\mu}_x, \boldsymbol{\Gamma} = (\sigma^{-2} \bar{\mathbf{H}}^T \bar{\mathbf{H}} + \mathbf{W}^{-1})^{-1},$$

where $\boldsymbol{\mu}_x = \sigma \mu_{\mathcal{N}} \mathbf{1}_L + \frac{\beta_{\mathcal{N}}}{\sigma_x} \mathbf{w}$, $L = \sum_k q_k$ and $\mathbf{W} = \text{diag}\{\mathbf{w}\}$. As a consequence, one can easily calculate the marginalized posterior distribution with respect to \mathbf{x}

$$p(\mathbf{w}|\mathbf{q}, \boldsymbol{\theta}, \mathbf{y}) P(\mathbf{q}|\boldsymbol{\theta}, \mathbf{y}) \propto |\mathbf{B}|^{-1/2} P(\mathbf{q}|\xi) \prod_k p(w_k) \times \exp\left(-\frac{1}{2}(\mathbf{y}^T \mathbf{B}^{-1} \mathbf{y} + \boldsymbol{\mu}_x^T \mathbf{C}^{-1} \boldsymbol{\mu}_x - 2\mathbf{y}^T \mathbf{D} \boldsymbol{\mu}_x)\right)$$

$$\begin{aligned} \text{with } \mathbf{B}^{-1} &= \sigma^{-2} \mathbf{I} - \sigma^{-4} \bar{\mathbf{H}} \boldsymbol{\Gamma} \bar{\mathbf{H}}^T & (N \times N) \\ \mathbf{C}^{-1} &= \mathbf{W}^{-1} - \mathbf{W}^{-1} \boldsymbol{\Gamma} \mathbf{W}^{-1} & (L \times L) \\ \mathbf{D} &= \sigma^{-2} \bar{\mathbf{H}} \boldsymbol{\Gamma} \mathbf{W}^{-1} & (N \times L) \end{aligned}$$

4. PARTIALLY COLLAPSED GIBBS SAMPLER

The sampling strategy adopted here is summarized in Algorithm 1. At each iteration i , first sample $q_k^{(i)}$ and $w_k^{(i)}$ for each k , then sample $\mathbf{x}^{(i)}|\mathbf{q}^{(i)}, \mathbf{w}^{(i)}$, and finally sample the hyperparameters $\boldsymbol{\theta}$ according to their posterior.

Algorithm 1: Partially collapsed Gibbs sampler

- At each iteration i :
1. for all k in $1, \dots, K$:
 - (a) draw $q_k, w_k | \mathbf{q}_{-k}, \mathbf{w}_{-k}, \boldsymbol{\theta}, \mathbf{y}$ according to Algorithm 2
 2. draw $\mathbf{x} | \mathbf{q}, \mathbf{w}, \boldsymbol{\theta}, \mathbf{y}$ (Gaussian distribution)
 3. draw $\boldsymbol{\theta} | \mathbf{q}, \mathbf{w}, \mathbf{x}, \mathbf{y}$ according to their conditional posterior distributions

While sampling $\mathbf{x}|\mathbf{q}, \mathbf{w}$ is an easy task ($\bar{\mathbf{x}}$ is Gaussian of size L), sampling q_k and w_k need to be considered in a special framework as the variable w_k is defined only if $q_k = 1$ which implies that the posterior distribution is defined in a space with a varying dimension. Reversible-Jump (RJ) MCMC methods, introduced in [20], are able to manage jumps between subspaces of different dimensions.

For the problem considered here, two states can be distinguished; whether $q_k = 1$ and $w_k \in \mathbb{R}^+$, or $q_k = 0$ and w_k is not defined. The RJ-MCMC framework allows to jump between these two states using the moves hereafter:

- Birth*: from $q_k = 0$, propose $(q'_k = 1, w'_k)$,
- Death*: from $(q_k = 1, w_k)$, propose $q'_k = 0$,
- Update*: from $(q_k = 1, w_k)$, propose $(q'_k = 1, w'_k)$.

Note that the move from $q_k = 0$ to $q'_k = 0$ is not of interest as this proposal is always accepted, and introduces no change. In the following, we denote by $p_{uu'}$ the probability of proposing a move from the state u to u' . Since we have chosen to systematically propose a *birth* move when $q_k = 0$ then $p_{01} = 1$. Otherwise, when $q_k = 1$, it is reasonable to randomly propose either a *death* or an *update* move with equal probability $p_{10} = p_{11} = \frac{1}{2}$.

Algorithm 2: Sampling q_k and w_k using the RJ framework

```

if  $q_k = 0$  then
  - birth:
    - propose  $q'_k = 1$  and  $w'_k \sim q_{01}(w'_k)$ 
    - accept with probability  $\alpha_{01}$ .
else
  with probability  $p_{10}$ 
  - death:
    - propose  $q'_k = 0$  and  $w'_k = \emptyset$ ,
    - accept with probability  $\alpha_{10}$ .
  with probability  $p_{11} = 1 - p_{10}$ 
  - update:
    - propose  $q'_k = 1$  and  $w'_k \sim q_{11}^{(1)}(w'_k)$  or
       $w'_k \sim q_{11}^{(2)}(w'_k)$  with probability  $\frac{1}{2}$ ,
    - accept with probability  $\alpha_{11}$ .
end

```

The w'_k candidates are proposed according to the following proposal distributions. When a *birth* move is chosen, it seems natural to propose w'_k according to its prior distribution $q_{01}(w'_k) \sim GIG_{\mathcal{N}}(\sigma_x^2)$. Also, if a *death* is selected the proposal is deterministic, hereafter for notational convenience we will write $q_{10}(w'_k) = \delta_{\emptyset}(w'_k)$ (for more details, one may refer to [21, Remark 4.2]). Finally, for the *update* move, a mix between two proposals is considered: the first one is the prior distribution $q_{11}^{(1)}(w'_k) \sim GIG_{\mathcal{N}}(\sigma_x^2)$ allowing a better exploration of the feasible domain of w_k , while the second is done via a random-walk Metropolis Hastings where the proposal is a truncated Gaussian $q_{11}^{(2)}(w'_k) \sim \mathcal{N}^+(w_k, \rho^2)$ enabling a local exploration of the posterior [22, 23]. Such a RJ-MCMC step is summarized in Algorithm 2

To ensure the reversibility, and thus the invariance of the Markov chain with respect to the posterior distribution (see [20, 21]), the candidates are accepted according to $\alpha_{uu'} = \min\{1, r_{uu'}\}$, where

$$r_{uu'} = \frac{p(\mathbf{w}'|\mathbf{q}', \boldsymbol{\theta}, \mathbf{y}) P(\mathbf{q}'|\boldsymbol{\theta}, \mathbf{y}) p_{u'u} q_{u'u}(w'_k)}{p(\mathbf{w}|\mathbf{q}, \boldsymbol{\theta}, \mathbf{y}) P(\mathbf{q}|\boldsymbol{\theta}, \mathbf{y}) p_{uu'} q_{uu'}(w'_k)} \quad (4)$$

Direct evaluation of (4) is inefficient from the computational viewpoint. Therefore, we have adapted and extended the numerical implementation technique introduced in [15] to reduce the computation and memory load. A key point is to recursively handle Cholesky factors instead of matrices \mathbf{B} , \mathbf{C} and \mathbf{D} . Due to lack of space, calculation details cannot be given here.

5. SIMULATION TESTS

In this section, our goal is to compare the BTG and BGH samplers, both in terms of signal restoration and convergence rate. We have designed a classical spike train deconvolution problem similar to [13]. Figure 2 shows the data \mathbf{y} and the impulse response \mathbf{h} , where the observed signal \mathbf{y} (of length $N = 300$) corresponds to the convolution of the spike train \mathbf{x} with \mathbf{h} (of length $P = 21$) and corrupted by an additive Gaussian noise

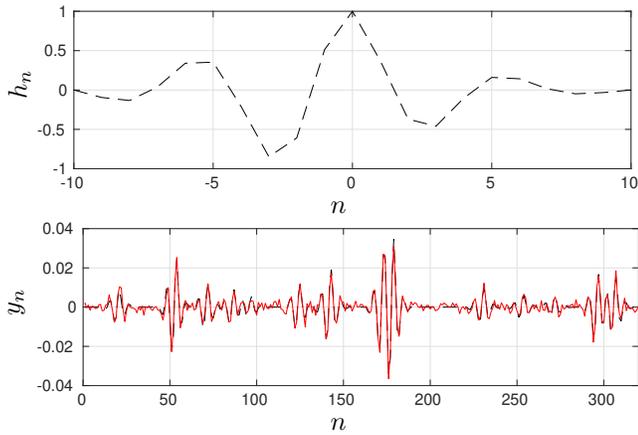


Fig. 2: Impulse response \mathbf{h} and simulated data \mathbf{y} for the test scenario.

(signal to noise ratio SNR = 12dB). The explicit expression of impulse response \mathbf{h} is:

$$h_n = \cos\left((n-10)\frac{\pi}{4}\right) e^{-|0.225n-2|^{1.5}}, \quad n = 0, \dots, 20.$$

In other words, the dictionary \mathbf{H} corresponds to a convolution matrix where the entries are shifted versions of \mathbf{h} . Furthermore, we consider a fully unsupervised situation where the hyperparameters $\boldsymbol{\theta} = [\xi, \sigma^2, \sigma_x^2]^T$ are also unknown and sampled. It is worth mentioning the simulations were run using MATLAB on a computer with Intel Xeon E5-2680 processors with a CPU clocked at 2.8 GHz.

To empirically compare the convergence speed of the different samplers, we have used Brooks and Gelman's graphical method to assess convergence [24]. This diagnostic method is based upon the covariance estimation of J independent Markov chains of equal length I by calculating the multivariate potential scale reduction factor (MPSRF). Convergence is diagnosed when the MPSRF is close to one. As suggested in [24], we have chosen MPSRF < 1.2. The signal restoration quality have been assessed by the signal-to-noise ratio (SNR) $20 \log(\|\mathbf{x}\|_2 / \|\mathbf{x} - \hat{\mathbf{x}}\|_2)$, where $\hat{\mathbf{x}}$ refers to the posterior mean estimate of \mathbf{x} .

Figure 3 shows the Posterior Mean estimates of \mathbf{x} produced by both samplers, which are of similar quality (an SNR of 9.40 dB for the BGH and 9.39 dB for the BTG). This indicates that the approximation method of the truncated Gaussian by the GH distribution is effective. These results are also compared to the Bernoulli-Gaussian (BG) model in order to emphasize the importance of the non-negativity of the prior.

Figure 4 gives the MPSRF of the variables \mathbf{x} in logarithmic scale for both samplers, as a function of time. Whereas the BTG sampler needs around 10^5 iterations to converge, the proposed BGH converges in about 10^4 iterations, thanks to partially collapsed sampling. In terms of computing time, the acceleration factor is larger than five. Finally, let us stress that the considered numerical test corresponds to a problem of moderate complexity. For problems of increasing complexity, the relative efficiency of the partially collapsed sampler is expected to increase compared to the standard sampler, according to [15].

As mentioned above, the tuning of parameter $\beta_{\mathcal{N}}$ is crucial to ensure the good performance of the BGH. On the one hand,

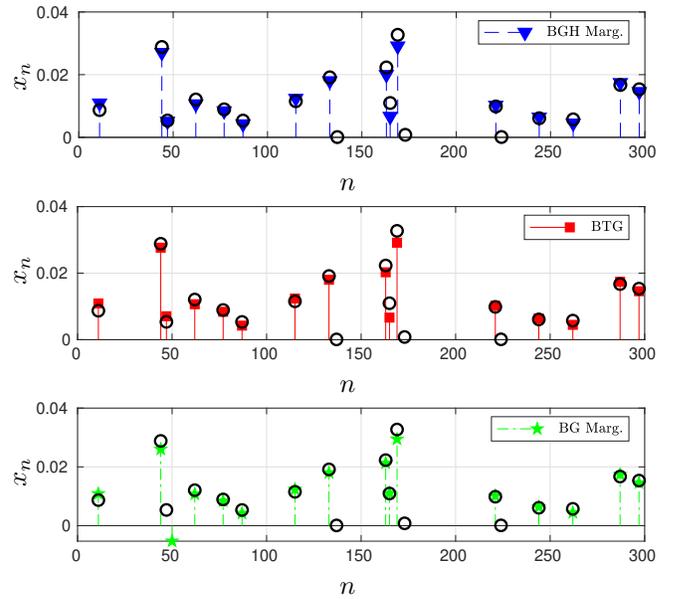


Fig. 3: Restored signal $\hat{\mathbf{x}}$ at convergence for the BGH, BTG and BG compared to its true value (black circles).

$\beta_{\mathcal{N}}$ has to be large enough to enforce the non-negativity, but on the other hand, too large values of $\beta_{\mathcal{N}}$ degrade the mixing properties of the PCGS sampler. For this specific experiment, the parameter $\beta_{\mathcal{N}}$ was empirically set to 150 in order to achieve the best trade-off between the quality of the restored signal and convergence properties of the BGH sampler.

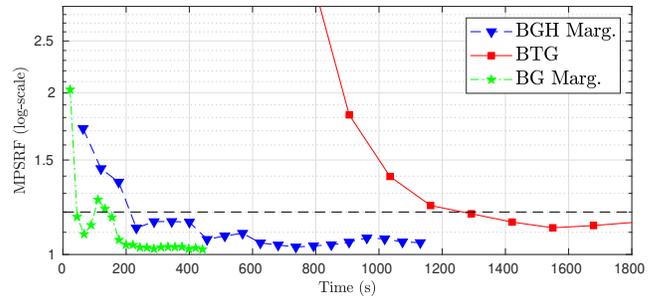


Fig. 4: Evolution of MPSRF (in log scale) of the BGH, BTG and BG samplers. The horizontal dashed line corresponds to the 1.2 threshold.

6. CONCLUSION

We have introduced a partially collapsed Gibbs strategy to accelerate nonnegative sparse signal restoration based on Bernoulli Gaussian model. Our contribution is based on an approximation of the truncated Gaussian using a Generalized Hyperbolic distribution, and on the Gaussian mixture decomposition property of the latter. Nevertheless, this PCGS strategy remains valid for any given distribution with Gaussian mixture decomposition property.

An interesting perspective will be to automatically tune parameter β of the GH distribution, to adjust the trade-off between the convergence rate of the proposed sampler, and its capability to enforce the nonnegativity constraint.

7. REFERENCES

- [1] B. K. Natarajan, “Sparse approximate solutions to linear systems”, *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
- [2] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification”, *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, May 1989.
- [3] D. L. Donoho and Y. Tsaig, “Fast solution of ℓ_1 -norm minimization problems when the solution may be sparse”, *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
- [4] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition”, in *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, Nov. 1993, pp. 40–44 vol.1.
- [5] I. Barbu and C. Herzet, “A new approach for volume reconstruction in TomoPIV with the alternating direction method of multipliers”, *Measurement Science and Technology*, vol. 27, no. 10, pp. 104002, Sept. 2016.
- [6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression”, *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.
- [7] M. Yaghoobi and M. E. Davies, “Fast non-negative orthogonal least squares”, in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug. 2015, pp. 479–483.
- [8] T. T. Nguyen, J. Idier, C. Soussen, and E. Djermoune, “Non-negative orthogonal greedy algorithms”, *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5643–5658, Nov. 2019.
- [9] J. Kormylo and J. Mendel, “Maximum likelihood detection and estimation of Bernoulli-Gaussian processes”, *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 482–488, May 1982.
- [10] F. Champagnat, Y. Goussard, and J. Idier, “Unsupervised deconvolution of sparse spike trains using stochastic approximation”, *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2988–2998, Dec. 1996.
- [11] Q. Cheng, R. Chen, and T.-H. Li, “Simultaneous wavelet estimation and deconvolution of reflection seismic signals”, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 2, pp. 377–384, Mar. 1996.
- [12] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier, “Sparse deconvolution: Comparison of statistical and deterministic approaches”, in *IEEE Statistical Signal Processing Workshop (SSP)*, June 2011, pp. 317–320.
- [13] V. Mazet, J. Idier, and D. Brie, “Déconvolution impulsionnelle positive myope”, in *20 Colloque sur le traitement du signal et des images*. GRETSI, 2005.
- [14] M. Boudineau, H. Carfantan, S. Bourguignon, and M. Bazot, “Sampling schemes and parameter estimation for nonlinear Bernoulli-Gaussian sparse models”, in *IEEE Statistical Signal Processing Workshop (SSP)*, June 2016, pp. 1–5.
- [15] D. Ge, J. Idier, and E. Le Carpentier, “Enhanced sampling schemes for MCMC based blind Bernoulli-Gaussian deconvolution”, *Signal Processing*, vol. 91, no. 4, pp. 759–772, Apr. 2011.
- [16] D. A. van Dyk and T. Park, “Partially collapsed Gibbs samplers”, *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 790–796, June 2008.
- [17] O. Barndorff-Nielsen and D. G. Kendall, “Exponentially decreasing distributions for the logarithm of particle size”, *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 353, no. 1674, pp. 401–419, Mar. 1977.
- [18] B. Jrgensen, *Statistical properties of the generalized inverse Gaussian distribution*, Springer New York, 1982.
- [19] E. Eberlein and E. A. v. Hammerstein, “Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes”, in *Seminar on Stochastic Analysis, Random Fields and Applications IV*, R. C. Dalang, M. Dozzi, and F. Russo, Eds., Basel, 2004, pp. 221–264, Birkhäuser Basel.
- [20] P. J. Green, “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination”, *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.
- [21] R. Waagepetersen and D. Sorensen, “A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping”, *International Statistical Review*, vol. 69, no. 1, pp. 49–61, May 2001.
- [22] S. Bourguignon and H. Carfantan, “Spectral analysis of irregularly sampled data using a Bernoulli-Gaussian model with free frequencies”, in *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, May 2006, vol. 3, pp. 516–519.
- [23] C. Andrieu and A. Doucet, “Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC”, *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2667–2676, Oct. 1999.
- [24] S. P. Brooks and A. Gelman, “General methods for monitoring convergence of iterative simulations”, *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998.