A Partially Collapsed Gibbs Sampler for Unsupervised Nonnegative Sparse Signal Restoration

M.C. Amrouche[1], H. Carfantan[1] and J. Idier[2]

[1]Institut de Recherche en Astrophysique et Planétologie, Université de Toulouse, CNRS/UPS/CNES, Toulouse, France
[2]Laboratoire des Sciences du Numérique de Nantes, CNRS/ECN, Nantes, France

Abstract

We introduce a new strategy, based on the **Bernoulli-Generalized-Hyperbolic** prior, to reconcile **nonnegativity** constraint with **efficient sampling** methods using partially collapsed Gibbs sampling (**Marginalisation**) for **unsupervised** nonnegative sparse signal restoration.

1. Nonnegative Sparse Signal Restoration

Problem statement

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\epsilon}$$

Goal Find the sparse nonnegative vector $\boldsymbol{x}$

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \mathbf{H}\boldsymbol{x}\|^2 \text{ s.c } \|\boldsymbol{x}\|_0 \leq S,\ x_k \geq 0\ \forall k$$

$||\boldsymbol{x}||_0$ is the ℓ_0 pseudo-norm, *i.e.*, Card$(k|x_k \neq 0)$

→ **Stochastic sampling:** probabilistic hierarchical models.

★ **Unsupervised** case, and $\mathbf{H}$ is highly correlated.

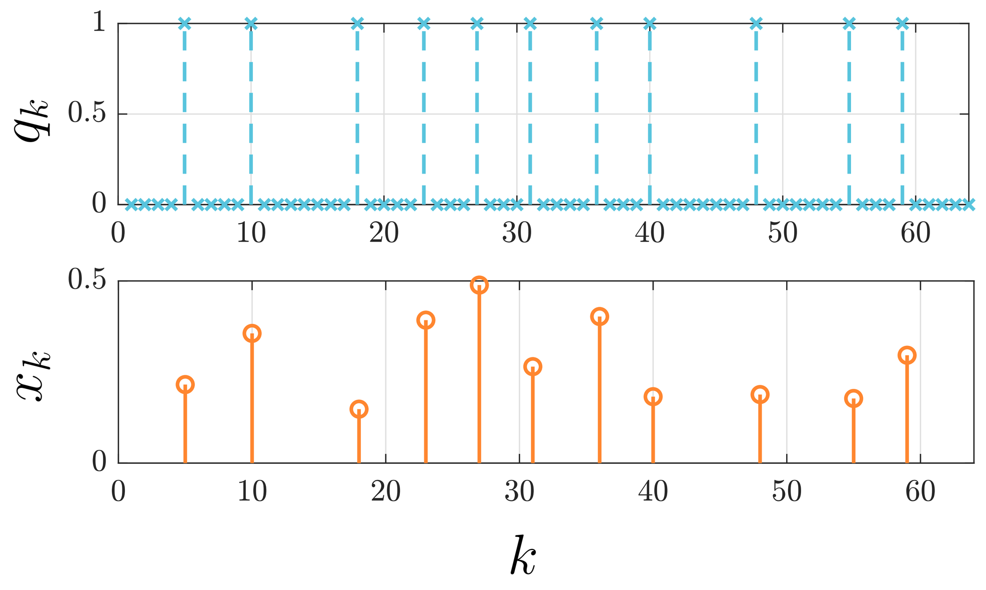
2. Available method : BTG Sampler

Bernoulli-Truncated-Gaussian prior [1]

Let $\boldsymbol{q}$ binary variables such that $\sum q_k = ||\boldsymbol{x}||_0$.

$$\begin{cases} q_k \in \{0,1\} \\ \Pr(q_k = 1) = \xi \end{cases}$$

$$\begin{cases} x_k | q_k = 1 \sim \mathcal{N}^+(0, \sigma_x^2) \\ x_k | q_k = 0 \sim \quad \delta(x_k) \end{cases}$$



where $\mathcal{N}^+$ is the truncated Gaussian.

Posterior distribution noise distribution (i.i.d): $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$

$$\underbrace{p(\boldsymbol{x},\boldsymbol{q},\boldsymbol{\theta}|\boldsymbol{y})}_{\text{Posterior}} \propto \underbrace{\exp\left(-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||^2\right)}_{\text{Likelihood}} \underbrace{p(\boldsymbol{x}|\boldsymbol{q})P(\boldsymbol{q}|\xi)p(\boldsymbol{\theta})}_{\text{Priors}}$$

$p(\boldsymbol{\theta})$ is the prior of the hyper-parameters $\boldsymbol{\theta} = \{\xi, \sigma_x^2, \sigma_\epsilon^2\}$.

BTG-Gibbs

for each k
  Sample $q_k | \boldsymbol{q}_{-k}, \boldsymbol{x}_{-k}, \boldsymbol{\theta}, \boldsymbol{y}$ (Bernoulli)
  Sample $x_k | \boldsymbol{q}, \boldsymbol{x}_{-k}, \boldsymbol{\theta}, \boldsymbol{y}$ (Truncated-Gaussian)
end
Sample $\boldsymbol{\theta} | \boldsymbol{q}, \boldsymbol{x}, \boldsymbol{y}$ (unsupervised case)

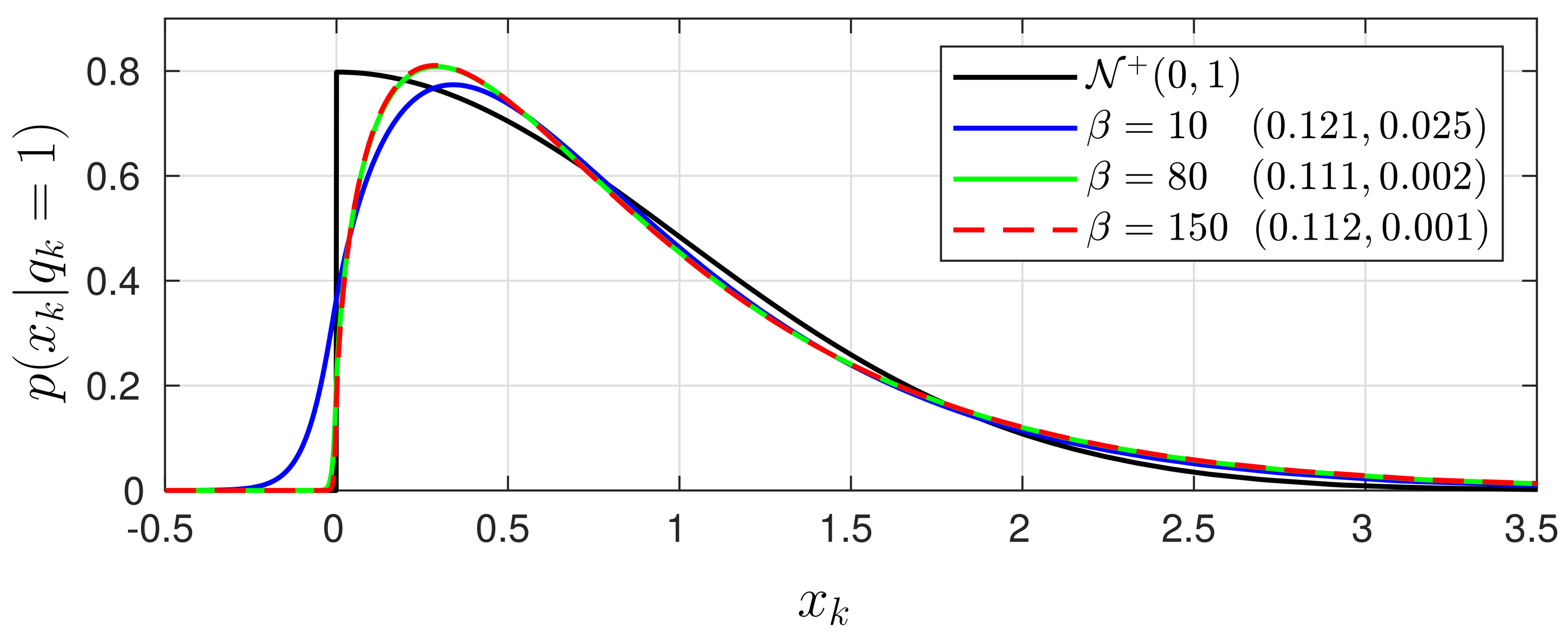
✗ Samples q_k and x_k are highly correlated. Slow convergence!

3. Contribution : BGH Sampler

Bernoulli-Generalized-Hyperbolic prior (BGH)

$$\begin{cases} q_k \in \{0,1\} \\ \Pr(q_k = 1) = \xi \end{cases} \text{ and } \begin{cases} x_k | q_k = 1 \sim GH(\boldsymbol{\nu}_{-\beta}, \beta) \\ x_k | q_k = 0 \sim \quad \delta(x_k) \end{cases}$$

where $\boldsymbol{\nu}_{-\beta} = \arg\min_{\boldsymbol{\nu}_{-\beta}} \text{TV}(GH(\boldsymbol{\nu}_{-\beta}, \beta), \mathcal{N}^+(0,1))$



β controls the skewness of the distribution (*i.e.*, $P(x_k \leq 0 | q_k = 1)$).

GH distributions [2] **continuous Gaussian mixture**:

if $X \sim GH(\boldsymbol{\nu})$ then,

$$p_X(x) = \int_{\mathbb{R}^+} p_{X|W}(x|w)\, p_W(w)\, dw \qquad \begin{array}{c} W \sim GIG(\boldsymbol{\kappa}) \\ X|W \sim \mathcal{N}(\mu + \beta W, W) \end{array}$$

Posterior with additional variables $\boldsymbol{w}$

$$p(\boldsymbol{q},\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||^2\right) p(\boldsymbol{x}|\boldsymbol{q}) P(\boldsymbol{q}|\xi)$$

$$p(\boldsymbol{x},\boldsymbol{w},\boldsymbol{q}|\boldsymbol{y},\boldsymbol{\theta}) \propto \overbrace{\exp\left(-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y} - \mathbf{H}\boldsymbol{x}||^2\right) p(\boldsymbol{x}|\boldsymbol{q},\boldsymbol{w})}^{\text{Gaussian / } \boldsymbol{x}} p(\boldsymbol{w}|\boldsymbol{q}) P(\boldsymbol{q}|\xi)$$

✓ $\boldsymbol{x}$ is easily marginalizable from $p(\boldsymbol{x},\boldsymbol{w},\boldsymbol{q}|\boldsymbol{y},\boldsymbol{\theta})$.

BGH-PCGS

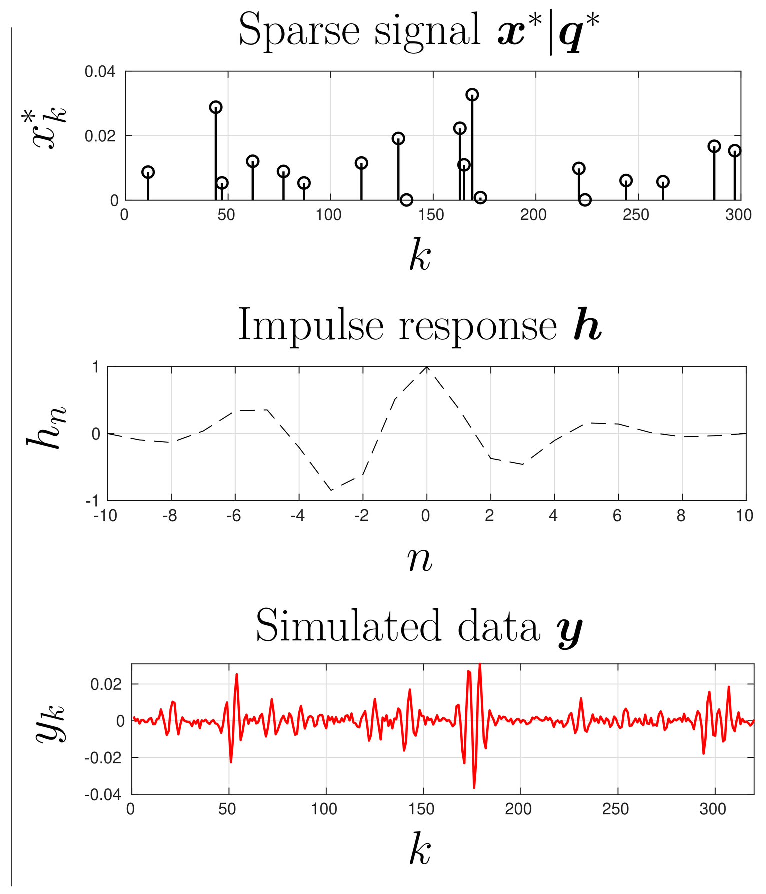
for each k
  Sample $q_k, w_k | \boldsymbol{q}_{-k}, \boldsymbol{w}_{-k}, \boldsymbol{\theta}, \boldsymbol{y}$ (from $p(\boldsymbol{q},\boldsymbol{w}|\boldsymbol{y},\boldsymbol{\theta})$)
end
Sample $\boldsymbol{x} | \boldsymbol{q}, \boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{y}$ (Gaussian, of size $L = ||\boldsymbol{z}||_0$)
Sample $\boldsymbol{\theta} | \boldsymbol{q}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{y}$

✓ Efficient sampling using PCGS [3]

4. Experiment

Sparse Deconvolution

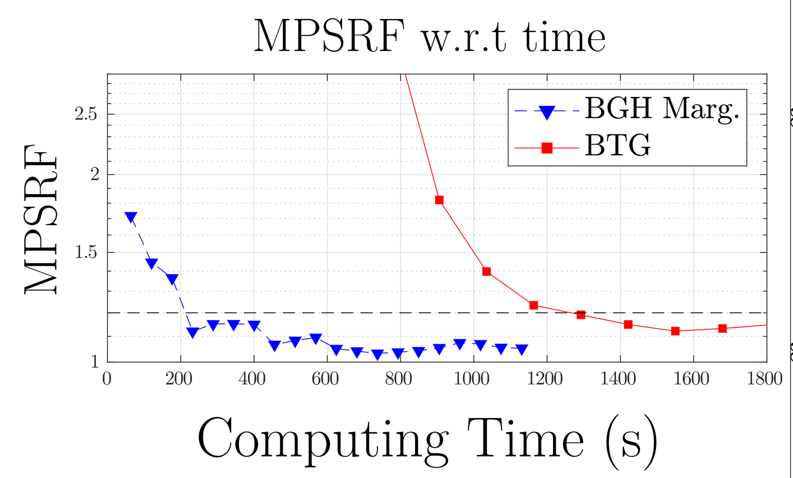
- $\boldsymbol{x}^*|\boldsymbol{q}^*$ BTG sequence,
- Noise level (SNR = 12 dB),
- Unsupervised scenario
- For the BGH: $\beta = 150$.

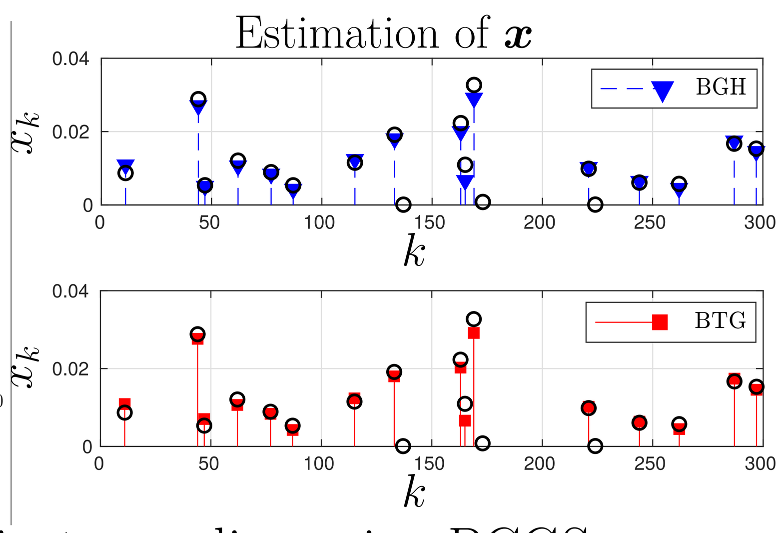


Convergence Monitoring [4]

MPSRF using $J = 10$ independent Markov chains.

Results





✓ Nonnegative restoration + Efficient sampling using PCGS.

5. Future Work

★ Approximate other models for nonnegativity:
 Bernoulli-Exponential.

★ Exact decomposition: unconstrained case
 Bernoulli-Laplace, Bernoulli-Cauchy.

★ Automatic tuning of parameter β.

References

[1] Vincent Mazet, Jérôme Idier, and David Brie, "Déconvolution impulsionnelle positive myope," in *20 Colloque sur le traitement du signal et des images*. GRETSI, 2005.

[2] O. Barndorff-Nielsen and David George Kendall, "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, Mar. 1977.

[3] David A van Dyk and Taeyoung Park, "Partially collapsed Gibbs samplers," *Journal of the American Statistical Association*, June 2008.

[4] Stephen P. Brooks and Andrew Gelman, "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, 1998.