Split-and-augmented Gibbs sampler
Application to large-scale inference problems

Maxime Vono[*], Nicolas Dobigeon[*], Pierre Chainais[†]

[*]Univ. Toulouse, IRIT/INP-ENSEEIHT, CNRS, Toulouse, France
[†]Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, Lille, France

Motivation: Bayesian inference

Contributions

- General Bayesian framework
- Variable splitting-inspired sampling
- Simple + Fast + efficient sampling
- Large scope of applications

Usual target density

$$\pi(\mathbf{x}) \propto \exp\left(-f(\mathbf{x}) - g(\mathbf{x})\right)$$

$f, g : \mathbb{R}^N \to \bar{\mathbb{R}}$
possibly ***non-smooth*** and/or ***non-convex***
e.g. machine learning, signal processing

Variable splitting in optimisation

$$\min_{\mathbf{x},\mathbf{z}} \quad f(\mathbf{x}) + g(\mathbf{z}) \qquad \text{s.t.} \quad \mathbf{x} = \mathbf{z}$$

Gibbs sampling

Split-and-augmented Gibbs sampler

% Drawing the variable of interest $\mathbf{x}$
$\mathbf{x}$ from $\propto \exp\left(-f(\mathbf{x}) - \phi_\rho(\mathbf{x}, \mathbf{z} - \mathbf{u})\right)$

% Drawing the splitting variable $\mathbf{z}$
$\mathbf{z}$ from $\propto \exp\left(-g(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z} - \mathbf{u})\right)$

% Drawing the auxiliary variable $\mathbf{u}$
$\mathbf{u}$ from $\propto \exp\left(-\psi_\alpha(\mathbf{u}) - \phi_\rho(\mathbf{x}, \mathbf{z} - \mathbf{u})\right)$

- Embedding of efficient existing samplers
  e.g. prox. MCMC, E-PO, AuxV1, ...
- Parallel with ADMM

Perspectives

- Theoretical analysis
- Non-convex potentials

References

Vono M., Dobigeon N., Chainais P., "Split-and-augmented Gibbs sampler – Application to large-scale inference problems", ***submitted***, 2018, https://arxiv.org/abs/1804.05809/

Variable splitting & data augmentation

Variable splitting

Idea: Introduce a splitting variable $\mathbf{z}$ + coupling function ϕ_ρ

e.g. $\phi_\rho(\mathbf{x}, \mathbf{z}) = \frac{1}{2\rho^2}\|\mathbf{x} - \mathbf{z}\|_2^2$

quadratic penalty function

Data augmentation

Idea: Less correlated MCMC draws with an auxiliary variable $\mathbf{u}$ + prior ψ_α

e.g. $\psi_\alpha(\mathbf{u}) = \frac{1}{2\alpha^2}\|\mathbf{u}\|_2^2$

conjugate density

Split density

$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp\left(-f(\mathbf{x}) - g(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z})\right)$$

Assumption:

$$\frac{\exp\left(-\phi_\rho(\mathbf{x}, \mathbf{z})\right)}{\int_{\mathbb{R}^N} \exp\left(-\phi_\rho(\mathbf{x}, \mathbf{z})\right) d\mathbf{z}} \xrightarrow[\rho \to 0]{} \delta_{\mathbf{x}}(\mathbf{z})$$

Limiting density of each marginal

Let p_ρ denote (by abuse of notation) any marginal under π_ρ. Then, it follows

$$\left\|p_\rho - \pi\right\|_{\mathrm{TV}} \xrightarrow[\rho \to 0]{} 0$$

Split-augmented density

$$\pi_{\rho,\alpha}(\mathbf{x}, \mathbf{z}, \mathbf{u}) \propto \exp\left(-f(\mathbf{x}) - g(\mathbf{z}) - \phi_\rho(\mathbf{x}, \mathbf{z} - \mathbf{u}) - \psi_\alpha(\mathbf{u})\right)$$

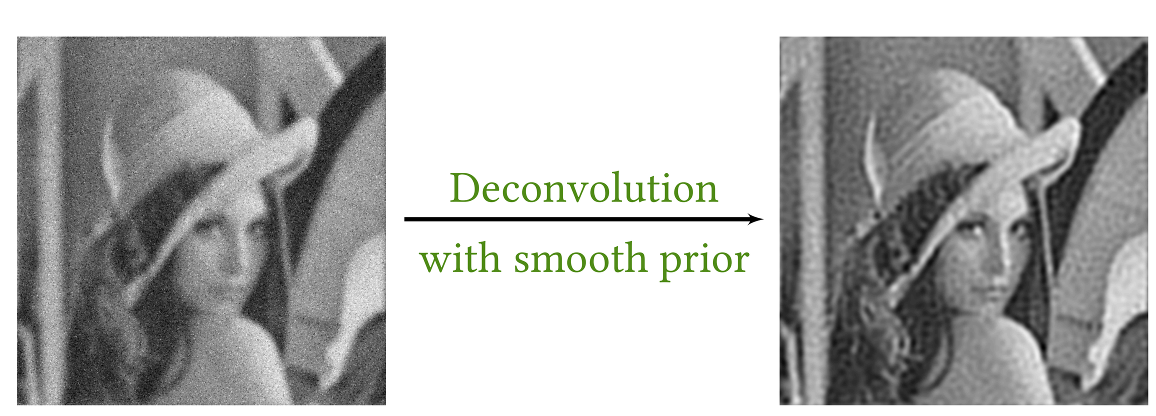
Probabilistic rationale: We must retrieve π_ρ by marginalizing $\mathbf{u}$ which is verified for quadratic forms for ϕ_ρ and ψ_α

Split-and-augmented scheme that can be generalized to

$$\pi(\mathbf{x}) \propto \exp\left(-\sum_{i=1}^{N_h} h_i\left(\mathbf{A}_i \mathbf{x}\right)\right)$$

Applications

Inverse problems



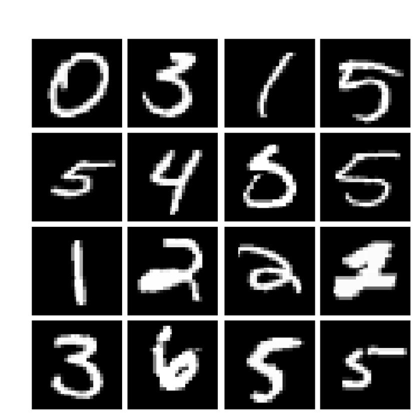
- High dimensional Gaussian sampling
- Covariance matrix split into two parts
- AuxV1 embedding

	computational complexity	# iterations	time (s)
RJ-PO	$\mathcal{O}(N_{\mathrm{CG}} N \log N) + 2N\mathcal{N}$	10^3	4192
AuxV1	$\mathcal{O}(N \log N) + 2N\mathcal{N}$	10^3	37
AuxV2	$\mathcal{O}(N \log N) + 4N\mathcal{N}$	3×10^3	209
SP	$\mathcal{O}(N \log N) + 3N\mathcal{N}$	10^3	62
SPA	$\mathcal{O}(N \log N) + 4N\mathcal{N}$	10^3	86

	SNR (dB)	PSNR (dB)
RJ-PO	19.58	25.24
AuxV1	19.58	25.24
AuxV2	19.60	25.26
SP	19.58	25.23
SPA	19.58	25.23

- Competing performances with AuxV1
- Faster than RJ-PO
- More efficient than AuxV2

Machine learning



Classification with Bayesian binary logistic regression

- Proximity operator of the logistic cost function
- Binomial likelihood split into M parts
  possible parallelisation of the M sampling steps
- P-MYULA embedding

	ADMM	P-MYULA	SPA
MNIST 1-vs-7	99.53	99.44	99.47
USPS 1-vs-7	99.18	99.06	99.11
MNIST 4-vs-6	99.06	98.88	99.12
USPS 4-vs-6	96.21	95.30	96.49
MNIST 3-vs-5	96.10	95.58	95.76
USPS 3-vs-5	97.83	97.08	97.47
MNIST one-vs-all	91.49	90.97	90.35

- Faster + more efficient than P-MYULA
- Reasonable cost w.r.t. ADMM
- Possible parallelisation