On variable splitting for Markov chain Monte Carlo

Maxime Vono(1), Nicolas Dobigeon(1) and Pierre Chainais(2)
(1)University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France
(2)Ecole Centrale de Lille, CRIStAL, Lille, France
{maxime.vono,nicolas.dobigeon}@irit.fr, pierre.chainais@centralelille.fr

***Abstract*—Variable splitting is an old but widely used technique which aims at dividing an initial complicated optimization problem into simpler sub-problems. In this work, we take inspiration from this variable splitting idea in order to build efficient Markov chain Monte Carlo (MCMC) algorithms. Starting from an initial complex target distribution, auxiliary variables are introduced such that the marginal distribution of interest matches the initial one asymptotically. In addition to have theoretical guarantees, the benefits of such an asymptotically exact data augmentation (AXDA) are fourfold: (i) easier-to-sample full conditional distributions, (ii) possibility to embed while accelerating state-of-the-art MCMC approaches, (iii) possibility to distribute the inference and (iv) to respect data privacy issues. The proposed approach is illustrated on classical image processing and statistical learning problems.**

I. Motivations

Numerous machine learning, signal and image processing problems involve the estimation of a hidden object of interest $\mathbf{x} \in \mathbb{R}^d$ based on (noisy) observations $\mathbf{y} \in \mathbb{R}^n$. This unknown object of interest can stand for parameters of a given model in machine learning [1] or may represent a signal or image to be recovered within an inverse problem. The main approaches to solve these problems can be casted into the class of optimization-based methods. The latter are known to be fast, efficient and might scale into big data and high-dimensional settings. [2]. A widely used optimization-based approach is the alternating direction method of multipliers (ADMM) [3]–[5] which is based on a technique called variable splitting. By the introduction of auxiliary variables, the ADMM simplifies, accelerates and can distribute the inference task [6], [7]. When the log-likelihood is supposed differentiable, optimization algorithms can provide confidence intervals on the pointwise estimation. However, this is not the case in general and people often resort to simulation-based methods such as Markov chain Monte Carlo (MCMC) [8] to quantify this estimation uncertainty. The price to pay for the latter can be high and even computationally prohibitive in high-dimensional settings since the Markov chain might fail to explore efficiently the parameter space.

To deal with these issues, we propose to rely on variable splitting to build novel MCMC algorithms, as detailed in the next section.

II. Proposed approach

Starting from an initial complicated target distribution with density (w.r.t. the Lebesgue measure)

$$\pi(\mathbf{x}) \propto \exp(-f_1(\mathbf{x}) - f_2(\mathbf{A}\mathbf{x})), \tag{1}$$

we introduce an auxiliary variable $\mathbf{z}$ such that the new target density becomes

$$\pi_\rho(\mathbf{x}, \mathbf{z}) \propto \exp(-f_1(\mathbf{x}) - f_2(\mathbf{z}) - \phi_\rho(\mathbf{A}\mathbf{x}, \mathbf{z})), \tag{2}$$

where ϕ_ρ stands for a divergence measuring the discrepancy between $\mathbf{x}$ and $\mathbf{z}$ [9], [10]. The marginal of interest under (2) is assumed to match (1) when $\rho \to 0$ leading to an asymptotically exact data augmentation (AXDA) scheme [11]. When $\phi_\rho(\mathbf{A}\mathbf{x}, \mathbf{z}) = (2\rho^2)^{-1} \|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2^2$, the distance in total variation between the marginal under (2) and (1) can be controlled exactly for a fixed $\rho > 0$ [11].

Interestingly, inferring from this AXDA model with a simulation-based method can be undertaken naturally with a special instance of a Gibbs sampler [8], [12] described in Algorithm 1. Similarly to the ADMM, the two functionals f_1 and f_2 are now dealt with separately which preludes simpler sampling steps which can be parallelized in some cases [10], [13]. If a conditional distribution cannot be sampled easily even after the splitting step, one can embed efficient existing MCMC algorithms within Algorithm 1 such as proximal MCMC ones [14], [15]. In the following, we describe and illustrate the main benefits of using the proposed approach on two image processing problems.

III. Illustrations

Image deblurring with total variation prior – We consider an image deconvolution problem where an original image $\mathbf{x}$ of size 256×256 ($d = 65536$) is blurred via a 5×5 Gaussian blur kernel with standard deviation equal to 2, see Figure 1. The likelihood has been supposed to be Gaussian while the total variation (TV) prior has been considered to model spatial constraints. The potential ϕ_ρ has been chosen to be quadratic leading to a Gaussian $\mathbf{x}$-conditional which can be sampled efficiently in the Fourier domain. The $\mathbf{z}$-conditional distribution has been dealt with by embedding the proximal Moreau-Yosida unadjusted Langevin algorithm (P-MYULA) [15] since this distribution is not differentiable. The results are shown in Table I and Figure 2 where the proposed approach has been also compared to the deterministic approaches of [6] and [16]. Note that the proposed approach leads to reconstruction results similar to optimization-based methods, can accelerate the convergence of state-of-the-art algorithms with a well chosen parameter ρ while providing at the same time uncertainty quantification.

Poisson image restoration under log-concave prior – Another application of the proposed approach that has been considered is the restoration of images contaminated with Poisson noise [17]. The main difficulty of this problem is that the posterior distribution obtained with a Poisson likelihood and a log-concave and possibly non-differentiable prior cannot be sampled with state-of-the-art algorithms. Indeed, P-MYULA assumes that the smooth potential of (1) is gradient-Lipschitz which is not the case with the Poisson log-likelihood. Fortunately, the proposed approach, by introducing appropriate auxiliary variables and by embedding P-MYULA, is able to sample efficiently from the posterior distribution. Illustrations and results can be found in Figure 3.

IV. Conclusion

We proposed a novel MCMC approach which takes inspiration from variable splitting, a widely-used optimization technique. This leads to simpler sampling steps which can be addressed efficiently by embedding state-of-the-art approaches while accelerating the latter in some cases.

Algorithm 1: Split Gibbs sampler

Input: Functions f, g, ϕ_ρ, parameter ρ, total number of iterations T_{MC}, number of burn-in iterations T_{bi}, initialization $\mathbf{z}^{(0)}$

for $t \leftarrow 1$ **to** T_{MC} **do**

  % Drawing the variable of interest

  Sample $\mathbf{x}^{(t)}$ according to $\pi_\rho\left(\mathbf{x}|\mathbf{z}^{(t-1)}\right)$;

  % Drawing the splitting variable

  Sample $\mathbf{z}^{(t)}$ according to according to $\pi_\rho\left(\mathbf{z}|\mathbf{x}^{(t)}\right)$;

end

Output: Collection of samples $\left\{\mathbf{x}^{(t)}, \mathbf{z}^{(t)}\right\}_{t=T_{\mathrm{bi}}+1}^{T_{\mathrm{MC}}}$ asymptotically distributed according to (2).

Fig. 1. Image deblurring with TV prior. (left) Original image, (middle) noisy and blurred image and (right) MMSE estimate computed with Algorithm 1.

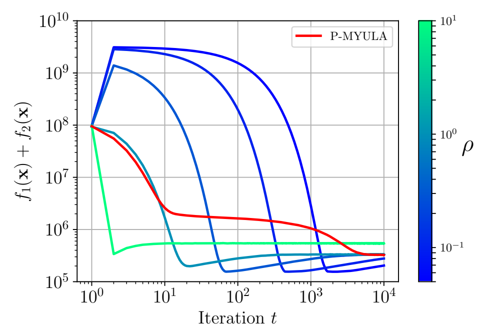

Fig. 2. Image deblurring with TV prior. Convergence of the Markov chains associated to Algo. 1 w.r.t ρ (from guppie green to blue) and P-MYULA (red) toward the typical set of π.

TABLE I

IMAGE DEBLURRING WITH TV PRIOR. PERFORMANCE RESULTS FOR BOTH OPTIMIZATION AND SIMULATION-BASED ALGORITHMS AVERAGED OVER 10 RUNS. FOR MCMC ALGORITHMS, THE SNR HAS BEEN CALCULATED WITH MMSE ESTIMATES.

	SALSA	**FISTA**	**Algo. 1**	**P-MYULA**
time (s)	1	10	470	3600
time ($\times$ var. split.)	1	10	1	7.7
nb. iterations	22	214	$\sim 10^4$	10^5
SNR (dB)	17.87	17.86	18.36	17.97

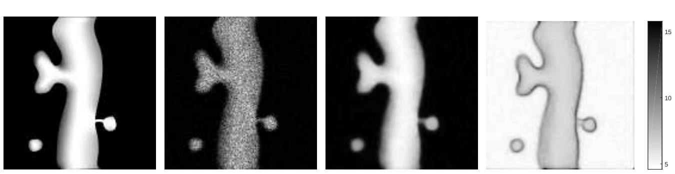

Fig. 3. From left to right: original image, noisy and blurred observation, MMSE estimate computed with Algo 1 and associated 95% credibility intervals.

REFERENCES

[1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[2] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, March 2016.

[3] R. Courant, "Variational methods for the solution of problems of equilibrium and vibrations," *Bull. Amer. Math. Soc.*, vol. 49, pp. 1–23, 1943.

[4] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17 – 40, 1976.

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[6] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2345–2356, Sept. 2010.

[7] ——, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, March 2011.

[8] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.

[9] M. Vono, N. Dobigeon, and P. Chainais, "Split-and-augmented Gibbs sampler - Application to large-scale inference problems," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1648–1661, Jan. 2019.

[10] L. J. Rendell, A. M. Johansen, A. Lee, and N. Whiteley, "Global consensus Monte Carlo," 2018. [Online]. Available: https://arxiv.org/abs/1807.09288/

[11] M. Vono, N. Dobigeon, and P. Chainais, "Asymptotically exact data augmentation: models, properties and algorithms," submitted. [Online]. Available: https://arxiv.org/abs/1902.05754/

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov 1984.

[13] M. Vono, N. Dobigeon, and P. Chainais, "Sparse Bayesian binary logistic regression using the split-and-augmented Gibbs sampler," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2018, Aalborg, Denmark*, 2018.

[14] M. Pereyra, "Proximal Markov chain Monte Carlo algorithms," *Stat. Comput.*, vol. 26, no. 4, pp. 745–760, July 2016.

[15] A. Durmus, E. Moulines, and M. Pereyra, "Efficient Bayesian computation by proximal Markov chain Monte Carlo: When Langevin meets Moreau," *SIAM J. Imag. Sci.*, vol. 11, no. 1, pp. 473–506, 2018.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[17] M. Vono, N. Dobigeon, and P. Chainais, "Bayesian image restoration under Poisson noise and log-concave prior," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Brighton, U.K., May 2019.