A GENERAL NON-SMOOTH HAMILTONIAN MONTE CARLO SCHEME USING BAYESIAN PROXIMITY OPERATOR CALCULATION

Lotfi Chaari, Jean-Yves Tourneret and Hadj Batatia

University of Toulouse, IRIT - INP-ENSEEIHT, France
firstname.lastname@enseeiht.fr

ABSTRACT

Sampling from multi-dimensional and complex distributions is still a challenging issue for the signal processing community. In this research area, Hamiltonian Monte Carlo (HMC) schemes have been proposed several years ago, using the target distribution geometry to perform efficient sampling. More recently, a non-smooth HMC (ns-HMC) scheme has been proposed to generalize HMC for distributions having non-smooth energy functions. This new scheme relies on the use of a proximity operator, which cannot be explicitly calculated for a large class of energy functions. We propose in this paper a fast and more general ns-HMC scheme that can be applied to any energy function by using a Bayesian calculation of the proximity operator, which makes the proposed scheme applicable to any energy function. Moreover, the proposed scheme relies on an interesting property of the proximity operator avoiding heavy calculations at each sampling step. The proposed scheme is tested on different sampling examples involving ℓ_p and total variation energy functions.

Index Terms— MCMC, HMC, ns-HMC, proximity operator

1. INTRODUCTION

Sparse signal recovery is still a hot research topic especially when the signals and images of interest are of large dimensions. This is the case with several recent applications considered in remote sensing [1] or medical imaging [2]. To solve ill-posed inverse problems, Bayesian techniques have demonstrated their ability to provide accurate estimations by automatically estimating all the model parameters and hyperparameters from the data. These techniques generally rely on a maximum a *posteriori* (MAP) estimation built upon the signal/image likelihood and priors. The inherent hierarchical Bayesian models generally involve sparse priors either in the original domain or in a transform space (such as the wavelet domain [3]). Using these priors can make the analytic derivation of Bayesian estimators difficult. Numerical approximations of the MAP estimators can therefore be built, using samples generated according to the posterior of interest using Markov chain Monte Carlo (MCMC) sampling techniques [4]. When the problem dimensions are large, the sampling task becomes difficult due to the poor performance of standard algorithms such as the Metropolis-Hastings (MH) [5] or the random walk MH algorithms [4]. To address this problem, several efficient sampling algorithms have been proposed such as elliptical slice sampling [6] or the Metropolis-adjusted Langevin algorithm (MALA) [7, 8]. In the same direction, sampling using Hamiltonian dynamics has also been investigated in [9, 10] resulting in the Hamiltonian Monte Carlo (HMC) algorithm. However, HMC schemes cannot be used in the case of exponential target distribution with non-differentiable energy function. To alleviate this problem, a non-smooth HMC (ns-HMC) sampling scheme has been recently proposed to make sampling using Hamiltonian dynamics possible for any exponential distribution [11]. This has been made possible by using the concepts of sub-differential and proximity operators [12], allowing the leapfrog discretization scheme to be generalized using a proximity operator calculation step.

However, analytic calculation of the proximity operator for a wide class of energy functions is not possible. This is the case of the ℓ_p norm function for general values of p (except for some values such as 1, 1.5, 2, 3 and 4). This drawback prevents the use of the ns-HMC algorithm in the case of sparse target distributions whose proximity operator of the energy function cannot be computed.

In this paper, our contribution is twofold. We first propose a modified ns-HMC scheme, which can be used for any energy function even when the proximity operator cannot be explicitly calculated. The proposed scheme relies on a Bayesian calculation of the proximity operator. The second contribution relies on the shift property of the proximity operator allowing costly calculations inside every sampling step to be avoided. The proximity operator is hence calculated only once at the initialization step, and then deduced at different points from the initially calculated value.

The rest of this paper is organized as follows. The addressed problem is formulated in Section 2. The proposed efficient ns-HMC scheme is developed in Section 3 and validated in Section 4. Finally, some conclusions and perspectives are drawn in Section 5.

ISBN 978-0-9928626-7-1 © EURASIP 2017

2. PROBLEM FORMULATION

Akin to [11], we consider a signal of interest $\boldsymbol{x} \in \mathbb{R}^N$ having the probability density function (pdf) $f(\boldsymbol{x};\boldsymbol{\theta})$ belonging to the exponential family parametrized by a vector of hyperparameter $\boldsymbol{\theta}$

$$f(\boldsymbol{x};\boldsymbol{\theta}) \propto \exp\left(-E_{\boldsymbol{\theta}}(\boldsymbol{x})\right), \tag{1}$$

where $E_{\boldsymbol{\theta}}$ is the energy function. In the following, we are interested in generating samples according to f for a known hyperparameter vector $\boldsymbol{\theta}$.

2.1. Hamiltonian Monte Carlo

Hamiltonian dynamics are originally used in the classical mechanics theory in order to model the total energy of a moving particle. For multidimensional efficient sampling, HMC [9, 10] techniques proceed by introducing an auxiliary variable $\boldsymbol{q}$ and a function K to simulate the potential energy variation, $E_{\boldsymbol{\theta}}$ assumed to model the kinetic energy of a moving particle. The Hamiltonian H can be expressed as

$$H(\boldsymbol{x},\boldsymbol{q}) = E(\boldsymbol{x}) + K(\boldsymbol{q}). \tag{2}$$

The motion equations of this operator determine the temporal evolution of the position $\boldsymbol{x}(t)$

$$\frac{d\boldsymbol{q}}{dt} = \frac{\partial H}{\partial \boldsymbol{x}};\ \frac{d\boldsymbol{x}}{dt} = -\frac{\partial H}{\partial \boldsymbol{q}}. \tag{3}$$

The Hamiltonian in (2) can equivalently be expressed by the following pdf

$$\begin{aligned} f_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{q}) &\propto \exp\left(-H(\boldsymbol{x},\boldsymbol{q})\right) \\ &\propto f(\boldsymbol{x};\boldsymbol{\theta}) \exp\left(-K(\boldsymbol{q})\right). \end{aligned} \tag{4}$$

HMC sampling is performed by sequentially updating $\boldsymbol{x}$ and $\boldsymbol{q}$ by sampling according to appropriate probability distributions. More precisely, this sampling is performed in two steps. The first one consists of sampling $\boldsymbol{q}$ according to the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix. The second step aims at proposing candidates $\boldsymbol{x}^*$ and $\boldsymbol{q}^*$ by simulating the Hamiltonian dynamics which can be discretized using the leapfrog discretization scheme [9, 10].

2.2. Non-smooth Hamiltonian Monte Carlo

As an extension of the HMC scheme, ns-HMC has been proposed in [11] to make possible the use of Hamiltonian dynamics for efficient sampling even for target distributions with non-smooth energy functions. For instance, ns-HMC can be used efficiently to sample a generalized Gaussian (GG) distribution with pdf

$$\forall \boldsymbol{x} \in \mathbb{R}^N, \mathrm{GG}(\boldsymbol{x}|\lambda,p) = \frac{p}{2\lambda\Gamma(1/p)} \exp\left(-E_{\boldsymbol{\theta}}(x)\right) \tag{5}$$

where $E_{\boldsymbol{\theta}}(x) = \frac{\|x\|_p^p}{\lambda^p}$ and $\boldsymbol{\theta} = \{\lambda, p\}$, $\|.\|_p^p$ being the ℓ_p norm. The ns-HMC scheme modifies the leapfrog discretization scheme by introducing a step calculating the proximity operator of $E_{\boldsymbol{\theta}}$. In the general case, the ns-HMC sampling scheme is given by algorithm 1, where L_f and ϵ denote the number of leapfrog steps and the stepsize, respectively [11].

Algorithm 1: ns-HMC algorithm [11].

- Initialize with some $\boldsymbol{x}^{(0,0)}$, set the iteration number $r = 0$, L_f and ϵ;
- **for** $r = 1, \ldots, S$ **do**
  - Sample $\boldsymbol{q}^{(r,0)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$;
  - Compute $\boldsymbol{q}^{(r,\frac{1}{2}\epsilon)} = \boldsymbol{q}^{(r,0)} - \frac{\epsilon}{2}\left[\boldsymbol{x}^{(r-1,0)} - \mathrm{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(r-1,0)})\right]$;
  - Compute $\boldsymbol{x}^{(r,\epsilon)} = \boldsymbol{x}^{(r-1,0)} + \epsilon\boldsymbol{q}^{(r,\frac{1}{2}\epsilon)}$;
  - **for** $l_f = 1$ *to* $L_f - 1$ **do**
    - * Compute $\boldsymbol{q}^{(r,(l_f+\frac{1}{2})\epsilon)} = \boldsymbol{q}^{(r,l_f\epsilon)} - \frac{\epsilon}{2}\left[\boldsymbol{x}^{(r,l_f\epsilon)} - \mathrm{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(r,l_f\epsilon)})\right]$;
    - * Compute $\boldsymbol{x}^{(r,(l_f+1)\epsilon)} = \boldsymbol{x}^{(r,l_f\epsilon)} + \epsilon\boldsymbol{q}^{(r,(l_f+\frac{1}{2})\epsilon)}$;
  - **end**
  - Compute $\boldsymbol{q}^{(r,(L_f+\frac{1}{2})\epsilon)} = \boldsymbol{q}^{(r,L_f\epsilon)} - \frac{\epsilon}{2}\left[\boldsymbol{x}^{(r,L_f\epsilon)} - \mathrm{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(r,L_f\epsilon)})\right]$;
  - Apply standard MH acceptation rule by taking $\boldsymbol{q}^* = \boldsymbol{q}^{(r,\epsilon L_f)}$ and $\boldsymbol{x}^* = \boldsymbol{x}^{(r,\epsilon L_f)}$;
- **end**

However, calculating $\mathrm{prox}_{E_{\boldsymbol{\theta}}}$ is not always straightforward. For instance, the proximity operator of the ℓ_p norm can only be explicitly calculated for few values of p (such as 1, 1.5, 2, 3 and 4) [13]. For other values of p, an iterative calculation is the only way to approximate the proximity operator, which is not always easy to perform [13]. Consequently, using the ns-HMC scheme is not always possible. The following section introduces a new sampling technique resulting in a fast and more general ns-HMC scheme.

3. GENERAL ns-HMC

3.1. Bayesian proximity operator calculation

We first recall the definition of the proximity operator as stated in [12].

Definition 3.1 *Let $\Phi \in \Gamma_0(\mathbb{R})$. For every $x \in \mathbb{R}$, the function $\Phi + \|\cdot - x\|^2/2$ reaches its infimum at a unique point referred to as proximity operator and denoted by $\mathrm{prox}_{\Phi}(x)$.*

In order to calculate the proximity operator of a function Φ, some analytic tools are available in the literature [13, 14]. It is worth noticing that this calculation does not lead to closed-form expressions of the proximity operator for a wide panel of functions. However, iterative calculation is possible in many

ISBN 978-0-9928626-7-1 © EURASIP 2017

cases by numerically approximating the proximity operator. Alternatively, we propose in what follows a Bayesian technique calculating the proximity operator that can be applied to any convex function Φ.

Following Definition 3.1 for the multidimensional case, and when the infimum is reached, we can write

$$\text{prox}_{\Phi}(\boldsymbol{x}) = \arg\min_{\boldsymbol{y}\in\mathbb{R}} \left[\Phi(\boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2\right]. \quad (6)$$

This optimization problem can be reformulated as finding the maximum of the following exponential function $\exp\left(-\Phi(\boldsymbol{x}) - \frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{2}\right)$, which can be seen as a posterior distribution of the random variable $\boldsymbol{X}$ conditionally to the observed random variable $\boldsymbol{Y}$

$$f(\boldsymbol{x} \mid \boldsymbol{y}) \propto \exp\left(-\Phi(\boldsymbol{x}) - \frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{2}\right) \quad (7)$$

where the prior on $\boldsymbol{x}$ can be expressed as

$$f(\boldsymbol{x}) \propto \exp\left(-\Phi(\boldsymbol{x})\right) \quad (8)$$

and the likelihood as

$$f(\boldsymbol{y} \mid \boldsymbol{x}) \propto \exp\left(\frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{2}\right). \quad (9)$$

Under this construction, it is easy to show that calculating the proximity operator is equivalent to solving a denoising problem under the assumption of an additive Gaussian noise having a covariance matrix equal to the identity matrix. Maximizing the posterior in (7) can therefore be performed by calculating the MAP estimator $\widehat{\boldsymbol{x}}$ that corresponds exactly to the proximity operator calculation as defined in Definition 3.1.

An MH algorithm can be used to sample according to the posterior distribution in (7) using an appropriate proposal distribution κ sharing the same support as Φ. The MAP estimator $\widehat{\boldsymbol{x}} = \text{prox}_{\Phi}(x)$ can then be approximated using the sampled chain $\{\boldsymbol{x}^{(r)}\}_{r=1,\ldots}$ after withdrawing samples corresponding to the burn-in period.

3.2. The proposed general ns-HMC scheme

Each iteration of the ns-HMC scheme in Algorithm 1 involves $L_f + 1$ calculations of $\text{prox}_{E_{\boldsymbol{\theta}}}(x)$ at different points. When no closed-form expression can be obtained for the proximity operator, the ns-HMC scheme becomes useless. In this paper, we propose a fast ns-HMC scheme based on a Bayesian calculation of $\text{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(0,0)})$ for any energy function $E_{\boldsymbol{\theta}}$. More precisely, this calculation is performed at the initialization point $\boldsymbol{x}^{(0,0)}$. All other evaluations of the proximity operator are then performed using $\text{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(0,0)})$. This is possible thanks to the following proximity operator property [13].

Property 3.1 *Let* $\Psi = \Phi(. - z)$ *where* $z \in \mathbb{R}$. *Then* $\text{prox}_{\Psi}(x) = z + \text{prox}_{\Phi}(x),\ \forall x \in \mathbb{R}$.

Algorithm 2: MH sampler for proximity operator calculation.

- Initialize with some $\boldsymbol{x}^{(0)}$ and set $r = 0$;
while *not convergence* **do**
① Propose a candidate $\boldsymbol{x}^* \sim \kappa(. \mid \boldsymbol{x}^{(r)})$;
② Calculate the acceptation ratio $r = \dfrac{\kappa(\boldsymbol{x}^* \mid \boldsymbol{x}^{(r)}) \exp\left(-\Phi(\boldsymbol{x}^*) - \frac{\|\boldsymbol{y}-\boldsymbol{x}^*\|^2}{2}\right)}{\kappa(\boldsymbol{x}^{(r)} \mid \boldsymbol{x}^*) \exp\left(-\Phi(\boldsymbol{x}^{(r)}) - \frac{\|\boldsymbol{y}-\boldsymbol{x}^{(r)}\|^2}{2}\right)}$;
③ Calculate the acceptation probability $\alpha = \min\{r, 1\}$;
④ Accept the proposed candidate with the probability α;
⑤ Set $r \leftarrow r + 1$;
end

Using this property, and based on an initial calculation of the proximity operator at a point $\boldsymbol{x}^0$, calculating the proximity operator at a candidate point $\boldsymbol{x}^*$ can be performed by decomposing $\boldsymbol{x}^* = \boldsymbol{x}^0 - (\boldsymbol{x}^0 - \boldsymbol{x}^*)$ and using the following relation

$$\text{prox}_{\Phi}(\boldsymbol{x}^*) = \boldsymbol{x}^0 - \boldsymbol{x}^* + \text{prox}_{\Phi}(\boldsymbol{x}^0). \quad (10)$$

The resulting general ns-HMC sampling scheme is detailed in Algorithm 3.

Algorithm 3: Proposed general ns-HMC algorithm.

- Initialize with some $\boldsymbol{x}^{(0,0)}$, set the iteration number $r = 0$, L_f and ϵ;
- Compute $P_0 = \text{prox}_{E_{\boldsymbol{\theta}}}(\boldsymbol{x}^{(0,0)})$ using Algorithm 2;
for $r = 1, \ldots, S$ **do**
- Sample $\boldsymbol{q}^{(r,0)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$;
- Compute $\boldsymbol{q}^{(r,\frac{1}{2}\epsilon)} = \boldsymbol{q}^{(r,0)} - \frac{\epsilon}{2}\left[2\boldsymbol{x}^{(r-1,0)} - \boldsymbol{x}^{(0,0)} - P_0)\right]$;
- Compute $\boldsymbol{x}^{(r,\epsilon)} = \boldsymbol{x}^{(r-1,0)} + \epsilon\boldsymbol{q}^{(r,\frac{1}{2}\epsilon)}$;
for $l_f = 1$ *to* $L_f - 1$ **do**
* Compute $\boldsymbol{q}^{(r,(l_f+\frac{1}{2})\epsilon)} = \boldsymbol{q}^{(r,l_f\epsilon)} - \frac{\epsilon}{2}\left[2\boldsymbol{x}^{(r,l_f\epsilon)} - \boldsymbol{x}^{(0,0)} - P_0\right]$;
* Compute $\boldsymbol{x}^{(r,(l_f+1)\epsilon)} = \boldsymbol{x}^{(r,l_f\epsilon)} + \epsilon\boldsymbol{q}^{(r,(l_f+\frac{1}{2})\epsilon)}$;
end
- Compute $\boldsymbol{q}^{(r,(L_f+\frac{1}{2})\epsilon)} = \boldsymbol{q}^{(r,L_f\epsilon)} - \frac{\epsilon}{2}\left[2\boldsymbol{x}^{(r,L_f\epsilon)} - \boldsymbol{x}^{(0,0)} - P_0\right]$;
- Apply standard MH acceptation rule to $(\boldsymbol{x}^*, \boldsymbol{q}^*)$ with $\boldsymbol{q}^* = \boldsymbol{q}^{(r,\epsilon L_f)}$ and $\boldsymbol{x}^* = \boldsymbol{x}^{(r,\epsilon L_f)}$;
end

It is worth noticing that the Bayesian calculation of the proximity operator could be calculated at each step but with an increased computational cost with respect to the proposed

ISBN 978-0-9928626-7-1 © EURASIP 2017

solution. Indeed, in the proposed scheme, the Bayesian calculation of the proximity operator is only calculated at the initialization step. The calculated value is then used to update the proximity operator value at different points. Another interesting property of the proposed scheme is that it does not depend on the initial point where the proximity operator is calculated first.

4. EXPERIMENTAL VALIDATION

In this section, we validate the proposed fast ns-HMC scheme on two main experiments. The first experiment aims at evaluating the sampling performance for a GG distribution. The second experiment illustrates the sampling performance on a particular target distribution involving a total variation (TV) term [15] for which the proximity operator cannot be calculated.

4.1. Experiment 1: ℓ_p sampling

In this section, we investigate the sampling of a GG distribution whose pdf is (5). In this experiment, sampling according to a GG distribution is addressed. For the scalar case, Fig. 1 displays the histogram of the sampled chains (after convergence) using standard ns-HMC and the proposed scheme for $p = 1.5$, a value for which the proximity operator has a closed-form expression (the scale parameter was fixed to $\lambda = 1$). This figure shows that in this case the two methods perform similarly. As regards computational time, the two methods run in approximately 7.5 seconds for 10^4 iterations.

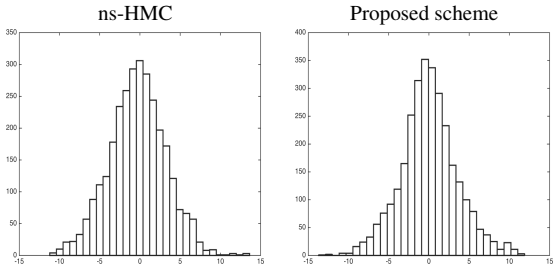

Fig. 1. Histograms of sampled chains using ns-HMC and fast ns-HMC for a univariate GG distribution with $p = 1.5$ and $\lambda = 1$.

The autocorrelations of the samples generated by the ns-HMC and fast ns-HMC are displayed in Fig. 2, showing that the the two methods generate samples with very similar correlation properties.

In a second scenario, we consider $p = 1.2$ and $\lambda = 1$. Fig. 3 shows the histogram of the sampled coefficients using the proposed scheme and a random walk MH algorithm (rw-MH). This figure shows that the proposed method provides accurate sampling since the histogram of the sampled chain is close to the one obtained by rw-MH. It is worth noticing that since the proximity operator cannot be explicitly calculated for this value of the shape parameter p, the standard ns-HMC scheme cannot be run in this case.

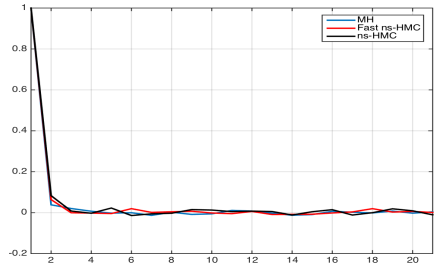

Fig. 2. ACFs of the sampled chains using MH, rw-MH and fast ns-HMC, for a 1d GG distribution with $p = 1.2$ and $\lambda = 1$.

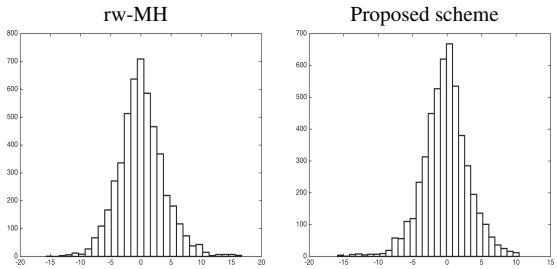

Fig. 3. Histograms of sampled chains using ns-HMC and fast ns-HMC for a univariate GG distribution with $p = 1.2$ and $\lambda = 1$.

As regards autocorrelation levels, Fig. 4 displays the ACF curves corresponding to the sampled chains using MH, rw-MH and the proposed scheme. These curves clearly show the low autocorrelation level obtained by our method, which is close to the one obtained by the standard MH algorithm.

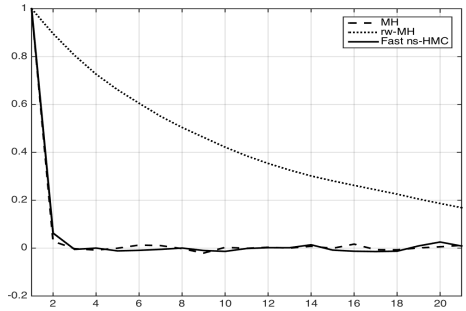

Fig. 4. ACFs of the sampled chains using MH, rw-MH and fast ns-HMC, for a univariate GG distribution with $p = 1.2$ and $\lambda = 1$.

In the next Section, a more challenging sampling example is proposed to better asses the performance of the proposed scheme on distributions whose proximity operator of the energy function cannot be explicitly calculated.

4.2. Experiment 2: TV sampling

This experiment illustrates the performance of the proposed scheme on sampling from an exponential distribution involving a TV term

$$f(\boldsymbol{x} \mid \alpha, \lambda) \propto \exp\left(-E_{\alpha,\lambda}(\boldsymbol{x})\right) \tag{11}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

where $E_{\alpha,\lambda}(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_2^2 + \lambda\|\nabla\boldsymbol{x}\|_1$, $\alpha \in \mathbb{R}_+$, $\lambda \in \mathbb{R}_+$ and $\nabla\boldsymbol{x}$ is the gradient of $\boldsymbol{x} \in \mathbb{R}^2$. This distribution is frequently encountered in image restoration problems involving TV regularization terms [16]. For this example, the proximity operator of $E_{\alpha,\lambda}$ cannot be explicitly calculated. Note also that even iterative calculations are difficult to conduct for this case unless splitting the TV term, which results in more complex resolution algorithms.

Fig. 5 displays the 2D histograms of the sampled chains using rw-MH and the proposed scheme, in addition to the target pdf. This figure shows that the target distribution is well approximated around the origin. This approximation is less precise for high values, and this with both methods. In order to

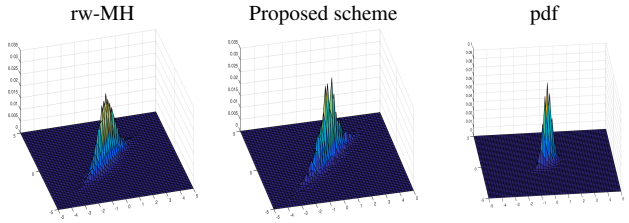

Fig. 5. 2D histograms of sampled chains using rw-MH and the proposed scheme ($\alpha = \lambda = 1$).

asses the convergence speed of the proposed scheme, Fig. 6 illustrates the Kullback-Leibler (KL) divergence between the target pdf and the histograms of the generated chains coefficients for the bivariate pdf (11). Note that the curves have been obtained by averaging the outputs of 50 Monte Carlo runs.

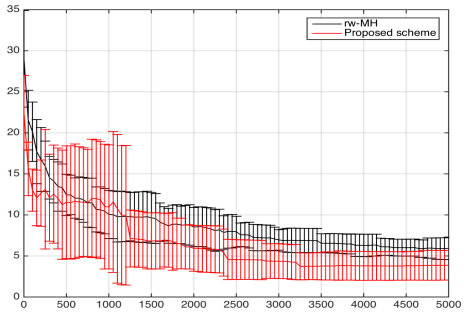

Fig. 6. Mean KL divergence (versus iteration number) between the target pdf and the histogram of the generated samples using rw-MH and the proposed scheme ($\alpha = \lambda = 1$).

This figure clearly shows the high acceleration rate obtained with the proposed technique w.r.t. rw-MH algorithm. This acceleration is coherent with the rates observed with the standard ns-HMC scheme [11].

5. CONCLUSION

This paper proposed a modified ns-HMC sampling scheme involving a Bayesian calculation of the proximity operator. This extension makes the ns-HMC algorithm applicable to more general distributions, in particular to the distributions for which the proximal operator cannot be computed analytically. An advantage of the proposed scheme is that it only requires an initial calculation of the proximity operator. The shift property is then used to deduce the proximity operator at any other point during the sampling steps. Sampling experiments demonstrated the efficiency of the proposed technique to perform fast and accurate sampling according to two specific target distributions. Future work will consider the application of the proposed scheme to large data restoration problems.

6. REFERENCES

[1] C. Chaux, P. L. Combettes, J.-C Pesquet, and V. Wajs, "Iterative image deconvolution using overcomplete representations," in *Proc. EUSIPCO*, Florence, Italy, Sep. 4-8 2006.

[2] L. Chaari, P. Ciuciu, S. Mériaux, and J.-C. Pesquet, "Spatio-temporal wavelet regularization for parallel MRI reconstruction: application to functional MRI," *Magn. Reson. Mater. in Phys., Bio. Med.*, vol. 27, no. 6, pp. 509–529, 2014.

[3] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, Ph. Ciuciu, and A. Benazza-Benyahia, "A hierarchical Bayesian model for frame representation," *IEEE Trans. Signal Process.*, vol. 18, no. 11, pp. 5560–5571, Nov. 2010.

[4] C. Robert and G. Castella, *Monte Carlo statistical methods*, Springer, New York, 2004.

[5] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[6] I. Murray, R. P. Adams, and D. MacKay, "Elliptical slice sampling," *J. Mach. Learn. Res.*, vol. 9, pp. 541–548, 2010.

[7] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli*, vol. 1, no. 4, pp. 341–363, 1996.

[8] M. Girolami and B. Calderhead, "Riemann manifold Langevin and Hamiltonian Monte Carlo methods," *J. R. Statist. Soc. B*, vol. 73, no. 2, pp. 123–214, 2011.

[9] K. M. Hanson, "Markov chain Monte Carlo posterior sampling with the Hamiltonian method," in *SPIE Medical Imaging: Image Processing, M. Sonka and K. M. Hanson, eds.*, 2001, pp. 456–467.

[10] R. M. Neal, "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo*, G. Jones X. L. Meng S. Brooks, A. Gelman, Ed., chapter 5. Chapman and Hall/CRC Press, 2010.

[11] L. Chaari, J.-Y. Tourneret, C. Chaux, and H. Batatia, "A Hamiltonian Monte Carlo method for non-smooth energy sampling," *IEEE Trans. on Signal Process.*, vol. 64, no. 21, pp. 5585 – 5594, Jun. 2016.

[12] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[13] C. Chaux, P. Combettes, J.-C. Pesquet, and V.R Wajs, "A variational formulation for frame-based inverse problems," *Inv. Prob.*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.

[14] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[15] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[16] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, and P. Ciuciu, "Parameter estimation for hybrid wavelet-total variation regularization," in *Proc. SSP*, Nice, France, June, 28-30 2011.

ISBN 978-0-9928626-7-1