A simplified variable metric hybrid Monte Carlo method

M. P. Calvo, I. Rodrigo, and J. M. Sanz-Serna

Citation: AIP Conference Proceedings **1558**, 14 (2013); doi: 10.1063/1.4825409
View online: http://dx.doi.org/10.1063/1.4825409
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1558?ver=pdfcov
Published by the AIP Publishing

A Simplified Variable Metric Hybrid Monte Carlo Method

M.P. Calvo[1], I. Rodrigo and J.M. Sanz-Serna

IMUVA. Universidad de Valladolid, Spain.

Abstract. We present a variable metric Hybrid Monte Carlo method following the ideas in [3], and propose a choice of such a metric which results efficient in the case of the sampling from the potential of a stiff spring. This is the first step in the extension of these ideas to deal with more general potentials appearing in Molecular Dynamics.

Keywords: Hybrid Monte Carlo, Hamiltonian problems, sampling methods
PACS: 02.50.Ng, 02.70.Tt, 02.70.Ns

INTRODUCTION

The Hybrid Monte Carlo method, first introduced in [2], allows to sample from probability density functions of the form

$$\pi(\mathbf{x}) \propto \exp(-V(\mathbf{x})), \quad \mathbf{x} \in R^d.$$

The variable $\mathbf{x} \in R^d$ can be interpreted as the configuration variable of a mechanical system and V as the corresponding potential function. The idea behind the Hybrid Monte Carlo method is to consider the Hamiltonian function

$$H(\mathbf{x},\mathbf{p}) = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} + V(\mathbf{x}), \tag{1}$$

which defines a canonical probability distribution in phase space R^{2d}

$$\propto \exp(-H(\mathbf{x},\mathbf{p})) = \exp(-V(\mathbf{x})) \times \exp(-\frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p}),$$

whose marginal distribution for $\mathbf{x}$ is the target $\pi(\mathbf{x})$, while the momenta follow a multivariate Gaussian density with zero mean and covariance matrix M (M is a constant mass matrix, frequently the identity matrix).

It is well known that the canonical density $\propto \exp(-H(\mathbf{x},\mathbf{p}))$ is preserved by the exact flow of the Hamiltonian system and almost preserved when the Hamiltonian system is integrated using an appropriate numerical method. The Hybrid Monte Carlo algorithm defines transitions $\mathbf{x}_n$ to $\mathbf{x}_{n+1}$ as follows:

1. Draw a momentum $\mathbf{p}_n$ from a Gaussian with zero mean and covariance matrix M (i.e. compute $\mathbf{p}_n = M^{1/2}\mathbf{Z}$ where $\mathbf{Z}$ denotes a vector of independent standard Gaussians).
2. Find an approximation $(\mathbf{x}^*,\mathbf{p}^*)$ to the exact solution of the Hamiltonian system after T units of time, starting from $(\mathbf{x}_n,\mathbf{p}_n)$. For instance, this can be done by advancing K steps of length h ($T = K\cdot h$) with the Stormer-Verlet method as follows: set $(\mathbf{x}_n^0,\mathbf{p}_n^0) = (\mathbf{x}_n,\mathbf{p}_n)$ and for $k = 0,\dots,K-1$ repeat
   - $\mathbf{p}_n^{k+1/2} = \mathbf{p}_n^k + \frac{h}{2}M^{-1}\mathbf{f}(\mathbf{x}_n^k)$,
   - $\mathbf{x}_n^{k+1} = \mathbf{x}_n^k + h\mathbf{p}_n^{k+1/2}$,
   - $\mathbf{p}_n^{k+1} = \mathbf{p}_n^{k+1/2} + \frac{h}{2}M^{-1}\mathbf{f}(\mathbf{x}_n^{k+1})$,

   where $\mathbf{f}(\mathbf{x}) = -\nabla V(\mathbf{x})$.
3. Set $(\mathbf{x}^*,\mathbf{p}^*) = (\mathbf{x}_n^K,\mathbf{p}_n^K)$ and set $\mathbf{x}_{n+1} = \mathbf{x}^*$ (acceptance) with probability

$$a = \min(1, \exp[-(H(\mathbf{x}^*,\mathbf{p}^*) - H(\mathbf{x}_n,\mathbf{p}_n))]).$$

[1] Corresponding author e-mail: maripaz@mac.uva.es

11th International Conference of Numerical Analysis and Applied Mathematics 2013
AIP Conf. Proc. 1558, 14-17 (2013); doi: 10.1063/1.4825409

4. If the proposal $\mathbf{x}^*$ is rejected, set $\mathbf{x}_{n+1} = \mathbf{x}_n$.

The transitions $\mathbf{x}_n \to \mathbf{x}_{n+1}$ defined in this way generate a Markov chain reversible with respect to the target density.

A VARIABLE METRIC HYBRID MONTE CARLO METHOD

The standard Hybrid Monte Carlo method uses a constant mass matrix in the Hamilton equations. The idea now is to consider variable mass matrices to improve, if possible, the efficiency of the sampling. In [3] the authors propose to replace the Hamiltonian function H given in (1) by $H^+(\mathbf{x},\mathbf{p}) = T(\mathbf{x},\mathbf{p}) + V^+(\mathbf{x})$, where

$$V^+(\mathbf{x}) = V(\mathbf{x}) + \frac{1}{2}\log\det M(\mathbf{x}) \quad \text{and} \quad T(\mathbf{x},\mathbf{p}) = \frac{1}{2}\mathbf{p}^T M(\mathbf{x})^{-1}\mathbf{p},$$

which has invariant density $\propto \exp(-H^+(\mathbf{x},\mathbf{p}))$. The marginal density with respect to $\mathbf{x}$ is again the target $\pi(\mathbf{x})$ and, at each transition, the initial momenta must be chosen as $\mathbf{p}_n = M(\mathbf{x}_n)^{1/2}\mathbf{Z}$ where $\mathbf{Z}$ denotes a vector of independent standard Gaussians. The resulting Hamiltonian system is no longer separable, the Stormer-Verlet scheme can not be used for its numerical integration and, in general, implicit schemes are required to define transitions $\mathbf{x}_n \to \mathbf{x}_{n+1}$. As a consequence, the computational cost per step of the new method is much higher than the computational cost of the standard Hybrid Monte Carlo. In order to reduce it, after introducing velocities $\mathbf{v} = M(\mathbf{x})^{-1}\mathbf{p}$, the same authors [4] propose the following explicit algorithm to define transitions from $\mathbf{x}_n$ to $\mathbf{x}_{n+1}$:

1. Draw a velocity $\mathbf{v}_n$ from a Gaussian with zero mean and covariance matrix $M(\mathbf{x}_n)^{-1}$ (i.e. compute $\mathbf{v}_n = M(\mathbf{x}_n)^{-1/2}\mathbf{Z}$ where $\mathbf{Z}$ denotes a vector of independent standard Gaussians).
2. Find an approximation $(\mathbf{x}^*,\mathbf{v}^*)$ to the exact solution of the Hamiltonian system by means of K steps of length h ($T = K\cdot h$) as follows: set $(\mathbf{x}_n^0,\mathbf{v}_n^0) = (\mathbf{x}_n,\mathbf{v}_n)$ and for $k = 0,\ldots,K-1$ repeat
   - $\mathbf{v}_n^{k+1/2} = \mathbf{v}_n^k + \frac{h}{2}M(\mathbf{x}_n^k)^{-1}\mathbf{f}(\mathbf{x}_n^k)$,
   - $\mathbf{x}_n^{k+1} = \mathbf{x}_n^k + h\mathbf{v}_n^{k+1/2}$,
   - $\mathbf{v}_n^{k+1} = \mathbf{v}_n^{k+1/2} + \frac{h}{2}M(\mathbf{x}_n^{k+1})^{-1}\mathbf{f}(\mathbf{x}_n^{k+1})$,

   where, as above, $\mathbf{f}(\mathbf{x}) = -\nabla V(\mathbf{x})$.
3. Set $(\mathbf{x}^*,\mathbf{v}^*) = (\mathbf{x}_n^K,\mathbf{v}_n^K)$ and set $\mathbf{x}_{n+1} = \mathbf{x}^*$ with probability

$$a = \min\left(1,\exp\left[-(H^+(\mathbf{x}^*,\mathbf{v}^*) - H^+(\mathbf{x}_n,\mathbf{v}_n))\right]\right) = \min\left(1,\sqrt{\frac{|M(\mathbf{x}^*)|}{|M(\mathbf{x}_n)|}}\exp\left[-(H(\mathbf{x}^*,\mathbf{v}^*) - H(\mathbf{x}_n,\mathbf{v}_n))\right]\right),$$

   where $|M(\mathbf{x})|$ denotes the determinant of matrix $M(\mathbf{x})$ and $H(\mathbf{x},\mathbf{v}) = V(\mathbf{x}) + \frac{1}{2}\mathbf{v}^T M(\mathbf{x})\mathbf{v}$.
4. If the proposal $\mathbf{x}^*$ is rejected, set $\mathbf{x}_{n+1} = \mathbf{x}_n$.

Although the numerical integrator is explicit, we observe that in the new method it is necessary to evaluate the mass matrix $M(\mathbf{x})$ (or its inverse) at each position where the gradient of the potential is required and it is also necessary to compute the square root of $M(\mathbf{x})$ and its determinant at each transition. This increases the overall computational cost of the method when compared with the standard Hybrid Monte Carlo.

The Choice of the Variable Mass Matrix

Let us assume that the target density is $\pi(\mathbf{x}) \propto \exp(\mathscr{L}(\mathbf{x}))$ with

$$\mathscr{L}(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mathbf{m})^T C^{-1}(\mathbf{x}-\mathbf{m}),$$

with C a given symmetric positive definite matrix. If $\mathbf{x}_n$ is the current position,

$$\mathscr{L}(\mathbf{x}) = \mathscr{L}(\mathbf{x}_n) + (\mathbf{m}-\mathbf{x}_n)^T C^{-1}(\mathbf{x}-\mathbf{x}_n) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_n)^T C^{-1}(\mathbf{x}-\mathbf{x}_n).$$

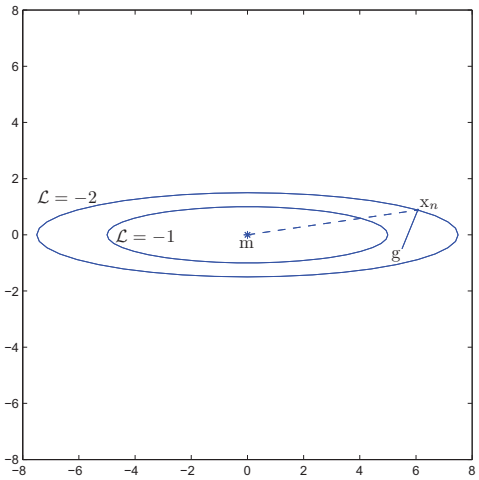

FIGURE 1. Picture to illustrate the choice of the variable mass matrix

This means that $\mathbf{m} - \mathbf{x}_n$ is the gradient of $\mathscr{L}$ at $\mathbf{x}_n$ with respect to the metric defined by C^{-1} and provides the best direction from $\mathbf{x}_n$ to reach the local maximum at $\mathbf{m}$. (Notice that $C^{-1}(\mathbf{m} - \mathbf{x}_n)$ is the gradient of $\mathscr{L}$ with respect to the Euclidean metric.) Figure 1 shows a picture illustrating this facts.

Then it is natural to take $M(\mathbf{x}) = -\mathscr{H}_{\mathscr{L}}$, where $\mathscr{H}_{\mathscr{L}}$ denotes the Hessian matrix of $\mathscr{L}$. To avoid situations where this Hessian is not positive definite, we take

$$M(\mathbf{x}) = \chi(-\mathscr{H}_{\mathscr{L}}(\mathbf{x})),$$

with χ an appropriate function satisfying $\chi(\xi) > 0$ and $\chi(\xi)/\xi \to 1$ as $\xi \to +\infty$.

In the particular case of a radial potential $V = V(r)$, with $r = |\mathbf{x}|$, then

$$\mathscr{H}_V = \frac{1}{r^2}\left[V''(r) - \frac{1}{r}V'(r)\right]\mathbf{x}\mathbf{x}^T + \frac{1}{r}V'(r)I_d = V''(r)\underbrace{\frac{\mathbf{x}\mathbf{x}^T}{r^2}}_{P_{\mathbf{x}}} + \frac{1}{r}V'(r)\underbrace{\left(I_d - \frac{\mathbf{x}\mathbf{x}^T}{r^2}\right)}_{P_\perp}, \tag{2}$$

where V' and V'' denote the first and second derivative of V with respect to r and $P_{\mathbf{x}}$ and $P_\perp$ denote the orthogonal projections onto the radial direction and the orthogonal hyperplane, respectively. Hence, taking $\mathscr{L} = -V$ one gets

$$M(\mathbf{x})^{-1} = \frac{1}{\chi(V''(r))}P_{\mathbf{x}} + \frac{1}{\chi\left(V'(r)/r\right)}P_\perp, \qquad M(\mathbf{x})^{-1/2} = \frac{1}{\sqrt{\chi(V''(r))}}P_{\mathbf{x}} + \frac{1}{\sqrt{\chi\left(V'(r)/r\right)}}P_\perp,$$

where it can also be used that $P_\perp \mathbf{y} = \mathbf{y} - P_{\mathbf{x}}\mathbf{y}$. Furthermore, the computation of products $M(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x})$ appearing in the algorithm can be efficiently done as $\mathbf{f}(\mathbf{x})$ only has radial component and, therefore $M(\mathbf{x})^{-1}\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x})/\chi(V''(r))$.

A Numerical Experiment

Let us consider the canonical density associated to the potential of the motion of a stiff spring in R^d

$$V(\mathbf{x}) = \frac{1}{2}k(r-\ell)^2, \quad r = |\mathbf{x}|, \tag{3}$$

where k is the elastic constant and ℓ is the length of the spring at equilibrium. The canonical density $\propto \exp(-V(\mathbf{x}))$ is essentially concentrated on a region of width $1/\sqrt{k}$ around the hypersphere $|\mathbf{x}| = \ell$.

The force at point $\mathbf{x}$ is

$$\mathbf{f}(\mathbf{x}) = \nabla_x(-V(\mathbf{x})) = -k(r-\ell)\frac{\mathbf{x}}{r}$$

and we define $M(\mathbf{x}) = \chi(\mathscr{H}_V)$ with $\chi(\xi) = \sqrt{k_0^2 + \xi^2}$, being $k_0 = d\sqrt{k}$, and $\mathscr{H}_V$ as in (2) for the potential (3).

We have considered different values of the dimension $d = 2, 3, 10$, two different values of the elastic constant $k = 1000$ and $k = 100000$ and different combinations (h, K) for the numerical integration of the Hamiltonian systems, leading to the same global computational time. The most remarkable fact observed in the experiments is that when increasing the value of the elastic constant k (by a factor of 100) the value of the step size h used in the Hybrid Monte Carlo method must be reduced (approximately by a factor of 10) in order to get positive acceptance ratios. On the contrary, when using the variable metric algorithm, the same step-size h can be used with both values of k producing very similar acceptance ratios in both cases. On the other hand, for a given value T used to generate elements in the Markov chain, decreasing the step size h implies that the number K of steps in the numerical integration must increase, leading to a higher computational cost. Thus, for $k = 1000$, the performances of the standard Hybrid Monte Carlo method and the simplified variable metric algorithm are very similar but for $k = 100000$ the last method, although computationally more expensive per integration step, becomes more efficient when sampling from the potential (3).

ACKNOWLEDGMENTS

The authors have been supported by project MTM2010-18246-C03-01 from Ministerio de Ciencia e Innovación

REFERENCES

1. E. Cancès, F. Legoll, and G. Stoltz, *ESAIM: Mathematical Modelling and Numerical Analysis* **41**, 351–389 (2007).
2. S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, *Phys. Lett. B* **55**, 2774–2777 (1987).
3. M. Girolami, and B. Calderhead, *J. R. Statist. Soc. B* **73**, 1–37 (2011).
4. M. Girolami, *private communication* (Oct. 2012).