# On the convergence time of some non-reversible Markov chain Monte-Carlo methods

Marie Vialaret

ENSAE ParisTech

July 6, 2018

Joint work with **Florian Maire** at *Insight Center for Data Analytics – University College Dublin*

# Outline


1. Introduction – General framework and notations
2. Introduction of an auxiliary variable in a non-reversible MH algorithm – NRMHAV algorithm and a demonstrative example
3. Skew-detailed balance in practice
4. Discussion – Limits of the NRMHAV algorithm

# General framework

- MCMC $\rightarrow$ estimate intractable expectations

$$\mathcal{I} = \int_{\mathcal{X}} f(x) d\pi(x)$$

by simulating a $\pi$-invariant Markov chain on some state-space $\mathcal{X}$

- "Metropolisation" of a proposal $Q$ (MH algorithm [Metropolis et al., 1953, Hastings, 1970]) :
  - starting from $X_k = x$, propose a move $y \sim Q(x, .)$
  - accept $X_{k+1} = y$ w.p. $A(x, y) = \{1 \wedge \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\}$
  - if rejected, set $X_{k+1} = x$

# General framework
## Efficiency of MCMC

- **Finite-time efficiency :** how much time do we need to get close to the equilibrium (target distribution) ? $\leftrightarrow$ mixing time
  TV distance $||\mu_0 P^t - \pi||_{TV} = \frac{1}{2}\sum_{x\in\mathcal{S}} |\mu_0 P^t(x) - \pi(x)|$
- **Asymptotic efficiency :** once we are at the steady-state, how efficient is the journey of the chain through the state space ($\Rightarrow$ accurate estimate for $\mathcal{I}$) ?

# Why irreversibility ?

- A $\pi$-reversible MC with transition kernel $P$ satisfies :

$$\pi(x)P(x,y) = \pi(y)P(y,x)\ \forall(x,y) \in S^2,\ x \neq y. \quad (1)$$

- $\pi$-reversible $\Rightarrow$ $\pi$-invariant
- Peskun's theorem [Peskun, 1973]
- Standard MCMC algorithms are $\pi$-reversible (MH, standard Gibbs)
- **Advantages of irreversibility :**
  - reversibility constrains the dynamics of the chain ;
  - more accurate estimate for $\mathcal{I}$ (smaller variance) ($\leftrightarrow$ [Sun et al., 2010, Ottobre, 2016]...).

**Problems :** what is the most efficient irreversible MCMC method ? how can we solve the conflict between finite-time and asymptotic efficiency raised by some irreversible algorithms ?

# Starting point [Bierkens, 2016]

**Guiding principle :** avoid backtracking [Neal, 2004]

Non-reversible Metropolis-Hastings algorithm (NRMH) [Bierkens, 2016] $\rightarrow$ skew-symetric perturbation of the Hastings ratio :

- suppose the chain is in $X_k = x$
- propose $y \sim Q(x, .)$
- accept $X_{k+1} = y$ w.p. $A_\Gamma(x, y) = \{1 \wedge \frac{\Gamma(x,y) + \pi(y)Q(y,x)}{\pi(x)Q(x,y)}\}$

| Discrete case | Continuous case |
|---|---|
| $\Gamma$ vorticity matrix | $\gamma$ vorticity density |
| $\Gamma = -\Gamma'$ | $\gamma(x, y) = -\gamma(y, x)\ \forall (x, y)$ |
| $\Gamma \mathbb{1} = 0$ | $\int_{A \times \mathbb{R}^d} \gamma(x, y) dxdy = 0\ \forall A \in \mathcal{B}(\mathbb{R}^d)$ |
| $Q(x, y) = 0 \Rightarrow Q(y, x) = 0$ | $q(x, y) = 0 \Rightarrow q(y, x) = 0$ |
| $Q(x, y) = 0 \Rightarrow \Gamma(x, y) = 0$ | $q(x, y) = 0 \Rightarrow \gamma(x, y) = 0$ |
| $\Gamma(x, y) + \pi(y)Q(y, x) \geqslant 0$ $\forall (x, y)$ s.t. $\pi(x)Q(x, y) \neq 0$ | $\gamma(x, y) + \pi(y)q(y, x) \geqslant 0$ $\forall (x, y)$ s.t. $\pi(x)q(x, y) \neq 0$ |

# Illustration

## Set-up

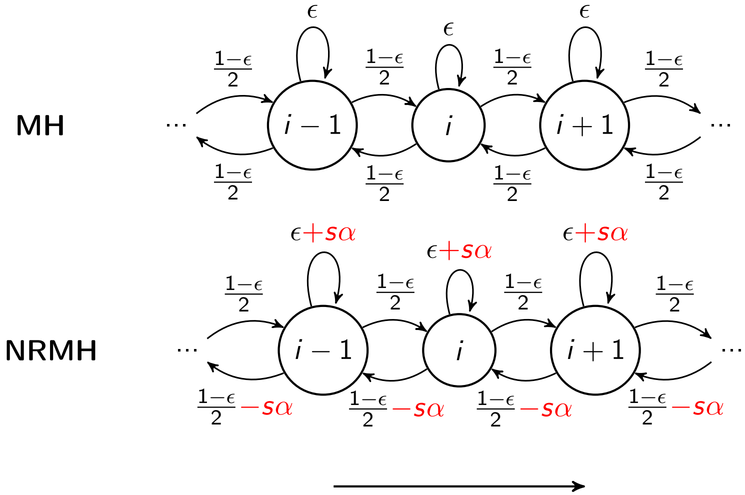
- Random walk on a circle with $s$ states : $S = \{1, ..., s\}$
- **Proposal :**

$$Q(x,y) = \begin{cases} \epsilon & \text{if } y = x, \\ \frac{1-\epsilon}{2} & \text{if } y = x \pm 1 \forall (x,y) \in \{2, ..., s\}^2, \\ \frac{1-\epsilon}{2} & \text{if } (x,y) \in \{1, s\} \text{ and } x \neq y, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

  with $\epsilon > 0$ small.
- **Target :** $\pi \sim \mathcal{U}\{1, ..., s\}$
- **Vorticity matrix :**

$$\Gamma = \begin{pmatrix} 0 & \alpha_\Gamma & 0 & 0 & \dots & 0 & -\alpha_\Gamma \\ -\alpha_\Gamma & 0 & \alpha_\Gamma & 0 & \dots & 0 & 0 \\ 0 & -\alpha_\Gamma & 0 & \alpha_\Gamma & 0 & \dots & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & \dots & 0 & -\alpha_\Gamma & 0 & \alpha_\Gamma & 0 \\ 0 & 0 & \dots & 0 & -\alpha_\Gamma & 0 & \alpha_\Gamma \\ \alpha_\Gamma & 0 & \dots & 0 & 0 & -\alpha_\Gamma & 0 \end{pmatrix} \tag{3}$$

# Illustration



# Illustration

## Results

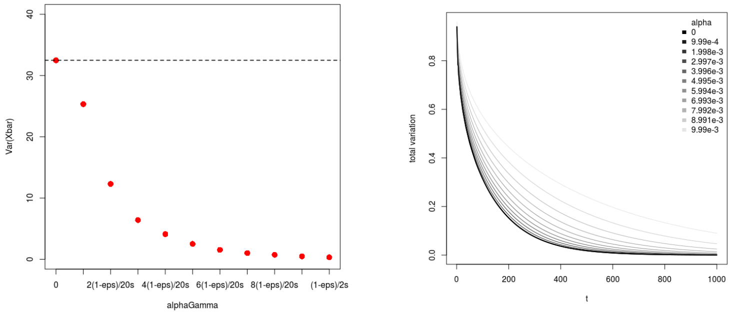



Figure 1: **Left.** Variance of the MC-estimate $\bar{X}$ of $\mathbb{E}(X)$ obtained using NRMH with different values of the vorticity parameter $\alpha_\Gamma$. The dotted line shows the value of this variance for reversible MH. **Right.** Evolution of the TV distance along the iterations of the NRMH algorithm for different values of $\alpha_\Gamma$. In black is the evolution of this distance for reversible MH.

# NRMHAV algorithm

## Discrete set-up – Hypotheses

**Intuitions :**

- improve the variance with the skew-symetric perturbations ;
- allow the chain to switch its direction sometimes $\rightarrow$ we expect a faster convergence ($\leftrightarrow$ **lifting** methods [Sakai and Hukushima, 2015, Vucelja, 2016, Ma et al., 2016, Poncet, 2017],etc.).

Enlarged state space $S \times \{-,+\} \rightarrow$ the chain is $X_k = (x_k, \xi_k)_k$
and the target $\tilde{\pi}(x_+) = \tilde{\pi}(x_-) = \pi(x)/2$
Consider $\pi$, $Q$ and non-zero matrices $\Gamma^+$, $\Gamma^-$ s.t.

- $\Gamma^\xi$ is a vorticity matrix i.e. $\Gamma^\xi = -(\Gamma^\xi)'$ and $\Gamma^\xi \mathbb{1} = 0$ ;
- $Q(x,y) = 0 \Rightarrow Q(y,x) = 0, \forall x, y \in S$ ;
- $\Gamma^\xi(x,y) \geqslant -\pi(y)Q(y,x)$ ;
- **SDBC :** $\pi(x)Q(x,y)A_{\Gamma^+}(x,y) = \pi(y)Q(y,x)A_{\Gamma^-}(y,x)\ \forall(x,y)$.

# NRMHAV algorithm

## Discrete set-up – Algorithm

- **Initialization :** let $X_1 = (x_0, \xi_0)$ for some $x_0 \in S$ and $\xi_0 \in \{-,+\}$
- Suppose $X_k = (x, \xi)$ and propose $y \sim Q(x,.)$
- **AR step 1 :** accept $X_{k+1} = (y, \xi)$ w.p. $A_{\Gamma^\xi} = \{1 \wedge \frac{\Gamma^\xi(x,y)+\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\}$
- **AR step 2 (switching step) :** if $(y, \xi)$ is rejected, accept $X_{k+1} = (x, -\xi)$ w.p. $\delta \in (0,1)$
- If $(x, -\xi)$ is rejected, set $X_{k+1} = (x, \xi)$

# Asymptotic efficiency

## Random walk on a circle

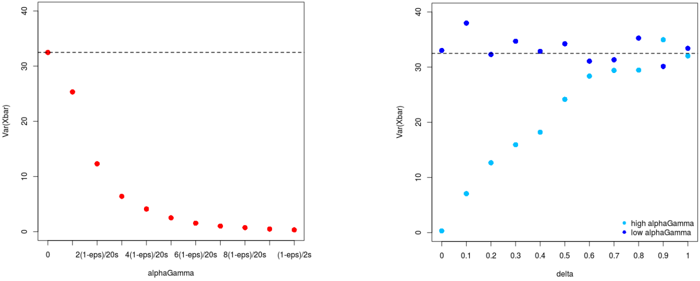



Figure 2: **Left.** Variance of the MC-estimate $\bar{X}$ of $\mathbb{E}(X)$ obtained using NRMH with different values of the vorticity parameter $\alpha_\Gamma$. **Right.** Variance of the same MC-estimate obtained using NRMHAV with fixed $\alpha_\Gamma$, for several values of the switching parameter $\delta$. The dotted lines show the value of this variance for reversible MH.

- NRMH-estimate has smaller variance when the vorticity is high ;
- NRMHAV-estimate has smaller variance when the vorticity is high and $\delta \to 0$ ($\to$ NRMH).

# Finite-time efficiency

## Random walk on a circle

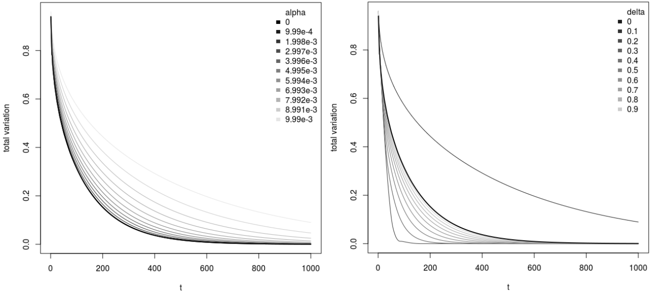



Figure 3: Evolution of the total variation distance along the iterations of the NRMH (left) and NRMHAV (right) algorithms on a circle with $s$ states for different values of the vorticity- and switching-parameters ; $\epsilon = 10^{-1}$. In black is the evolution of this distance for reversible MH. Initialisation is a Dirac in state 1. For NRMHAV, $\alpha_\Gamma$ is fixed close to its upper-bound.

- NRMH converges faster when the vorticity is low ($\rightarrow$ MH) ;
- NRMHAV converges faster when the vorticity is high and $\delta \sim 0.1$.

# NRMHAV algorithm

## Modification for practical issues

- **Initialization :** let $X_1 = (x_0, \xi_0)$ for some $x_0 \in S$ and $\xi_0 \in \{-,+\}$
- Suppose $X_k = (x, \xi)$ and propose $y \sim Q(x,.)$
- **Switching step :** set $\xi' = \xi$ w.p. $\delta' \in (0,1)$, otherwise $\xi' = -\xi$
- **AR step :** accept $X_{k+1} = (y, \xi')$ w.p. $A_{\Gamma^{\xi'}} = \{1 \wedge \frac{\Gamma^{\xi'}(x,y)+\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\}$
- If $(y, \xi')$ is rejected, set $X_{k+1} = (x, \xi')$

$\Rightarrow$ No need of SDBC **BUT** $(\xi_k)_k$ does not depend upon $(x_k)_k$ anymore...

# Discussion

**Main difficulties arising :**

- **SDBC** : It is in general not easy to satisfy a skew-detailed balance $\Rightarrow$ less satisfying algorithms (disconnection of the main chain and the auxiliary variable).
- **Condition** $\gamma(x,y) + \pi(y)q(y,x) \geqslant 0 \ \forall (x,y)$ : harder to satisfy in the continuous case $\Rightarrow$ the algorithm might be biased.

**Perspectives :**

- Time-adaptive vorticity parameter
- Use several proposals so as to avoid SDBC ($\leftrightarrow$ GMALA algorithms [Ma et al., 2016, Poncet, 2017])

Introduction

NRMHAV algorithm
NRMH
Intuitions and hypotheses
The algorithm
Example

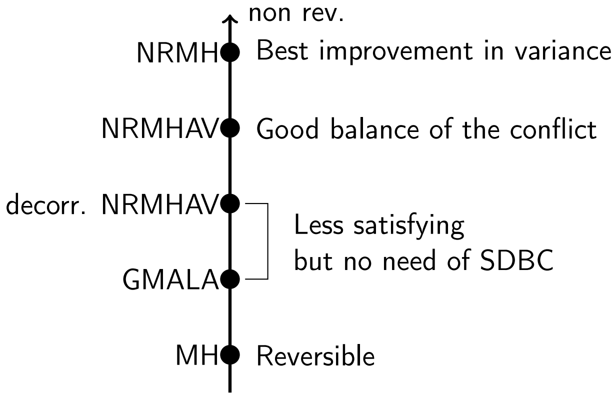
In practice

Discussion

# Discussion

Let

$$\phi = \int \int |\pi(x)K(x,y) - \pi(y)K(y,x)| dxdy$$



# References I

Bierkens, J. (2016).
Non-reversible metropolis-hastings.
*Statistics and Computing*, 26(6):1213–1228.

Hastings, W. (1970).
Monte carlo sampling methods using markov chains and their applications.
*Biometrika*, 57(1):97–109.

Ma, Y.-A., Chen, T., Wu, L., and Fox, E. B. (2016).
A unifying framework for devising efficient and irreversible mcmc samplers.
*arXiv preprint arXiv:1608.05973*.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953).
Equation of state calculations by fast computing machines.
*Journal of Chemical Physics*, 21(6).

Neal, R. M. (2004).
Improving asymptotic variance of mcmc estimators: Non-reversible chains are better.
*arXiv preprint math/0407281*.

Ottobre, M. (2016).
Markov chain monte carlo and irreversibility.
*Reports on Math. Phys.*

Peskun, P. (1973).
Optimum monte-carlo sampling using markov chains.
*Biometrika*, 60(3):607–612.

# References II

Poncet, R. (2017).
Generalized and hybrid metropolis-hastings overdamped langevin algorithms.
*arXiv preprint arXiv:1701.05833.*

Sakai, Y. and Hukushima, K. (2015).
Eigenvalue analysis of an irreversible random walk with skew detailed balance conditions.
*arXiv preprint arXiv:1511.08100.*

Sun, Y., Schmidhuber, J., and Gomez, F. J. (2010).
Improving the asymptotic performance of markov chain monte-carlo by inserting vortices.
In *Advances in Neural Information Processing Systems*, pages 2235–2243.

Vucelja (2016).
Lifting - a nonreversible markov chain monte carlo algorithm.
*American Journal of Physics*, 84(958).