Note on the convergence time of some non-reversible Markov chain Monte Carlo methods

Marie Vialaret\textsuperscript{a,b}, Florian Maire\textsuperscript{a,*}

\textsuperscript{a}University College Dublin, The Insight Centre for Data Analytics
\textsuperscript{b}ENSAE ParisTech

Abstract

Introduced in statistical physics, non-reversible Markov chain Monte Carlo (MCMC) algorithms have recently received an increasing attention from the computational statistics community. The main motivation is that, in the context of MCMC algorithms, non-reversible Markov chains usually yield more accurate empirical estimators than their reversible counterparts. In this note, we study the efficiency of non-reversible MCMC algorithms according to their speed of convergence. In particular, we show that in addition to their variance reduction effect, some non-reversible MCMC algorithms have also the undesirable property to slow down the convergence of the Markov chain. This point, which has been overlooked by the literature, has obvious practical implications. We accompany our analysis with novel non-reversible MCMC algorithm extending the non-reversible Metropolis-Hastings (NRMH) approach proposed in Bierkens (2016) that aims at solving, in some capacity, this conflict. This is achieved by introducing different vorticity flows in the Metropolis-Hastings algorithm that avoid slow convergence while retaining NRMH appealing variance reduction property.

Keywords: MCMC, non-reversible Markov chain, variance reduction, speed of convergence

1. Introduction

Markov chain Monte Carlo (MCMC) methods enjoy a wide popularity in numerous fields of applied mathematics and are used for instance for parame-

*Corresponding author
Email address: florian.maire@ucd.ie (Florian Maire)
ter estimation or model validation. The purpose of MCMC is to approximate quantities of the form
\[ \pi f := \int_S f(x) d\pi(x), \] (1)
i.e. the expectation of some \( \pi \)-measurable function \( f \) with respect to a distribution \( \pi \) defined on a state space \( S \), when an analytic expression of \( \pi f \) is not available and direct simulation from \( \pi \) is not doable.

Notations. In the following, \( \pi \) will be referred to as the target distribution, \( \mathcal{S} \) will denote a sigma-algebra on \( S \) and \( \text{Pr} \) the probability distribution generated by the underlying random experiment, in absence of ambiguity. For a Markov chain \( \{X_t, t \in \mathbb{N}\} \) with transition kernel \( P \) operating on \( (S, \mathcal{S}) \), we denote by \( P^t \) the iterated kernel defined as \( P^t(x, A) = \text{Pr}(X_t \in A \mid X_0 = x) \), for all \((x, A) \in S \times \mathcal{S} \). For any measure \( \mu \) on \((S, \mathcal{S})\), \( \mu P \) defines the measure \( \mu P := \int_S P(x, \cdot) \mu(dx) \). Finally, we define by \( \mathcal{M}_1(S) \) the set of probability measures on \((S, \mathcal{S})\).

Reversible Markov chains. MCMC methods aim to simulating an ergodic Markov chain whose invariant distribution is \( \pi \). As the chain converges towards its stationary distribution, it is possible to compute an empirical average of \( f \), by using the sample path of the Markov chain, once at stationarity. Perhaps the most popular MCMC algorithm is the well-known Metropolis-Hastings (MH) algorithm [9, 7] (outlined in Algorithm 1), which consists in iterating between (i) proposing a new state using a proposal distribution \( Q \) and (ii) accepting-rejecting this proposed state. Step (ii) guarantees that, by construction, MH generates a \( \pi \)-reversible Markov kernel which, therefore, admits \( \pi \) as limiting distribution. Recall that a Markov chain on a state space \( S \) with transition kernel \( P \) is said to be time reversible (or simply reversible) with respect to \( \pi \) if \((\pi, P)\) satisfy the detailed balance condition (DBC):
\[ \forall (x, y) \in S^2, \ x \neq y, \ \pi(x) P(x, y) = \pi(y) P(y, x). \] (2)
or, in terms of measures, if for all \((A, B) \in \mathcal{S}^2\), we have:
\[ \int_A \pi(dx) P(x, B) = \int_B \pi(dx) P(x, A). \] (3)
The terminology “time-reversible” comes from the fact that the dynamic of the chain is the same whether the time flow goes in one direction or the other.
If there exists no distribution $\pi$ such that Eq. (2) holds, the chain is said to be non-reversible. Standard MCMC algorithms such as MH or the standard one-component Gibbs sampler [6] produce time-reversible Markov chains. Reversible chains present numerous advantages, as several theoretical results (rate of convergence, spectral analysis, etc.) make them easy to study. The main reason for their popularity is perhaps the property that a $\pi$-reversible Markov chain is necessarily $\pi$-invariant. Hence, constructing a Markov chain satisfying (2) avoids further questions regarding the existence of a stationary distribution.

**Algorithm 1** Metropolis-Hastings algorithm

- Initialize in $X_0 \sim \mu_0$ and let $X_t = x$
- Propose $Y \sim Q(x, \cdot) \rightsquigarrow y$
- Set $X_{t+1} = y$ with probability $A(x, y) = 1 \wedge R(x, y)$ where

$$R(x, y) := \begin{cases} 1 & \text{if } \pi(x)Q(y, x)/\pi(y)Q(x, y) \neq 0 \\ \pi(y)Q(y, x)/\pi(x)Q(x, y) & \text{otherwise} \end{cases} \quad (4)$$

If the proposal is rejected, set $X_{t+1} = x$

Nevertheless, as the DBC (2) imposes that the joint probabilities $\Pr(X_t \in A, X_{t+1} \in B)$ and $\Pr(X_t \in B, X_{t+1} \in A)$ are equal, reversibility may prevent the Markov chain from roaming efficiently through the state space, especially when $\pi$’s topology is irregular. This is illustrated by the following example.

**Example 1.** Let $s$ and $n$ be two integers such that $s \geq 4$ is even and $n > 1$. Define the discrete distribution on the circle $S = \{1, 2, \ldots, s\}$ ordered in the counterclockwise direction where $\pi_n(x) \propto 1$ if $x$ is odd and $\pi_n(x) \propto 1/n$ if $x$ is even. We consider the $\pi_n$-reversible Metropolis-Hastings Markov chain which attempts moving between neighbouring states, i.e. for all $(x, y) \in S^2 \setminus \{(1, s), (s, 1)\}$, we have $Q(x, y) = (1/2)\delta_{|x-y|=1}$ and $Q(1, s) = Q(s, 1) = 1/2$.

When $n$ is large, the $\pi_n$-reversibility and the fact that two consecutive modes are separated by a probability in $\mathcal{O}(1/n)$ make the chain reluctant to move quickly between them. In fact, the expected returning time to a given mode is of order $\mathcal{O}(n)$ implying that the Markov chain is mixing very slowly.
Efficiency of MCMC algorithms. Efficiency of a particular MCMC algorithm can be assessed from two points of view:

- **finite-time perspective:** starting from \( \mu \), an initial distribution on \((S, \mathcal{G})\), the Markov chain should converge towards its stationary distribution as fast as possible. In the following, the convergence speed will be assessed using the total variation distance:

\[
\sup_{\mu \in \mathcal{M}_1(S)} \left\| \mu P^t - \pi \right\| := \sup_{\mu \in \mathcal{M}_1(S)} \sup_{A \in \mathcal{G}} \left| \mu P^t(A) - \pi(A) \right|.
\] (5)

- **asymptotic perspective:** assuming that stationarity is reached, the algorithm should wander through the state space as efficiently as possible, so as to obtain a MC estimate of \( \pi f \) that is as accurate as possible. In particular, the variance of the empirical estimator \( \hat{\pi}_f = n^{-1} \sum_{t=1}^n f(X_t) \) should be as small as possible.

As explained in [14], those two measures of efficiency can sometimes be clashing. For reversible Markov chains, convergence and asymptotic efficiency are typically measured by two spectral quantities, the spectral gap and the spectral interval (see [14]), the larger the better. In the context of Example 1 with \( s = 4 \), it can be readily checked that the spectrum of the Metropolis-Hastings transition kernel is \( \{1, 1 - 1/n, 0, -1/n\} \) and thus the spectral gap and the spectral interval are both equal to \( 1/n \). Moreover, a careful derivation shows that the asymptotic variance is of order \( O(n) \), which illustrates the poor quality of the empirical estimator of this reversible Markov chain.

Recent works have shown that the asymptotic efficiency of MCMC algorithms using a non-reversible Markov chain is typically lower than those using reversible dynamic (see for instance [11, 14, 5, 10]). Several methods have been developed to construct such chains (see [1, 3, 8, 18], amongst others). In most approaches, the irreversibility results from the introduction of a vector field (e.g., cycles, vortices, etc.) in the chain dynamic. Intuitively, the variance reduction can be explained by those guiding features which reduce, to some extent, the uncertainty on the Markov chain sample paths. However, one can wonder if, as a by-product, the irreversibility does not slow down the convergence of the chain to its stationary distribution. Investigating this question is precisely the purpose of this note.

We first identify situations where the vorticity field making the dynamic irreversible slows down the convergence of the chain. Intuitively, assuming an
initial measure with mass on a subset $A \subset \mathcal{S}$, since the vector field imposes a privileged direction, a subset $B \subset \mathcal{S}$ which is in the opposite direction of the flow starting in $A$ will be slowly explored. In other words, one can construct a MC empirical estimator with a lower variance using a non-reversible chain but at the risk of waiting for a longer transient phase than for the reversible algorithms. We propose a novel non-reversible MCMC algorithm that aims at solving this conflict, at least to some extent.

**Brief state of the art.** Several methods have been developed to simulate non-reversible Markov chains, most of them consisting in a modification of standard reversible algorithms, designed so as to retain their $\pi$-invariance. The intuition is that, as reversible chains have a high probability to backtrack (because of the DBC, Eq. (2)), it is desirable to transform their transition kernel so as to get chains that are less likely to do so. A first family of non-reversible MCMC is obtained by introducing skew-symmetric perturbations in the transition probabilities (in the MH ratio) which ensures that one direction is privileged by the Markov chain. In particular, the probability of moving in the privileged direction can be increased in Eq. (4) by a quantity, say $\epsilon(x)$, that depends on the current state of the chain $x$, while the probability of a move in the opposite direction is decreased by the same quantity. Algorithms that follow this approach can be found in [1, 4, 10, 17]. As a result, the state space is roamed “more efficiently” as the chain is less likely to be “stuck” in a given area. In other words, the irreversibility effect can be thought of as giving the Markov chain some sort of inertia or momentum. More details on this type of non-reversible MCMC will be given at Section 2.

Another approach consists in enlarging the state space $\mathcal{S}$ to $\mathcal{S} \times \{-1, +1\}$, using an auxiliary variable $\xi$ to materialize changes in the privileged direction of the chain. These type of methods, sometimes referred to as lifting methods [16, 18], propose to switch the privileged direction when the proposal is rejected. Similarly, the generalized Metropolis-adjusted Langevin algorithm (GMALA) method [8, 13] uses several proposition kernels, according to the value of the auxiliary variable the chain is currently at. Finally, we mention the Zig-Zag process that builds a non-reversible Markov chain, see [2]. This is somewhat similar in essence to the lifting methods with the difference that between two changes of direction, the chain moves along a deterministic direction for a random time, whose distribution guarantees the $\pi$-invariance of the irreversible dynamic.
2. Non-reversible MH

For notational simplicity, we consider the case where \( \mathcal{S} \) is discrete; nevertheless the ideas and results presented below have a direct equivalent for general state spaces. Our starting point is the non-reversible MH (NRMH) algorithm proposed in [1] and outlined at Algorithm 2. A skew-symmetric perturbation referred to as a vorticity matrix/field, \( \Gamma : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R} \), is introduced in the MH ratio and should satisfy the following properties:

**Assumption 1.** The vector field \( \Gamma \) should satisfy a skew-symmetry condition

\[
\Gamma \neq 0, \quad \forall (x, y) \in \mathcal{S}^2, \quad \Gamma(x, y) = -\Gamma(y, x),
\]

and a non-explosion condition

\[
\forall x \in \mathcal{S}, \quad \int_{\mathcal{S}} \Gamma(x, dy) = 0.
\]

In addition, we assume that \((Q, \Gamma)\) satisfy jointly

**Assumption 2.** The proposal distribution satisfies a symmetric structure condition i.e. for all \((x, y) \in \mathcal{S}^2\), \(Q(x, y) = 0 \Rightarrow Q(y, x) = 0\) and the non-negativity of the MH acceptance probability imposes a lower bound condition on \(\Gamma\), i.e. for all \((x, y) \in \mathcal{S}^2\), \(\Gamma(x, y) \geq -\pi(y)Q(y, x)\).

Algorithm 2 Non-reversible Metropolis-Hastings algorithm (NRMH), from [1].

1: Initialize in \(X_0 \sim \mu_0\)  
   Transition \(X_t = x \rightarrow X_{t+1}\):  
2: Propose \(Y \sim Q(x, \cdot) \sim y\)  
3: Set \(X_{t+1} = y\) with probability \(A_\Gamma(x, y) = 1 \land R_\Gamma(x, y)\) where

\[
R_\Gamma(x, y) := \begin{cases} 
\frac{\Gamma(x, y) + \pi(y)Q(y, x)}{\pi(x)Q(x, y)} & \text{if } \pi(x)Q(x, y) \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]  

(6)

4: If the proposal is rejected, set \(X_{t+1} = x\)

If \(\Gamma\) and \(Q\) satisfy Assumptions [1] 2] the NRMH Markov chain admits \(\pi\) as invariant distribution (see [1, Theorem 2.5]). The intuition behind the non-explosion condition is that the non-reversibility introduced in the algorithm
must compensate overall through the state space. Similar conditions are also necessary to guarantee the validity of the the non-reversible GMALA methods presented in [8, 13].

Example 1 (continued). We implement Algorithm 2 to infer the distribution defined at Example 1. More precisely, we use the vector flow defined for all \((x,y) \in S^2\) as:

\[
\Gamma_\zeta(x,y) := \begin{cases}
\zeta & \text{if } y = x + 1 \text{ or } (x,y) = (s,1), \\
-\zeta & \text{if } y = x - 1 \text{ or } (x,y) = (1,s), \\
0 & \text{otherwise}
\end{cases}
\]  

(7)

where \(0 < \zeta \leq (1/4)(1+n)^{-1}\). This condition on \(\zeta\) and the structure of \(\Gamma_\zeta\) ensures that Assumptions 1 and 2 are both satisfied. Figure 1 gives an illustration of the efficiency of MH and NRMH. In particular, it shows that as expected, NRMH allows to reduce significantly the variance of the Monte Carlo estimate: the asymptotic variance of NRMH for the test function \(f \colon x \mapsto 1_{x=1}\) was in this case nearly 10 times less than MH. However, Figure 1 also shows a rather unexpected fact, namely that the non-reversibility slows down the convergence of the Markov chain. The intuition mentioned in the introduction is illustrated with the sample path of both chains: for the vorticity parameter \(\zeta = (1/4)(1+n)^{-1}\) all states \((x,y) \in S \times S\) with \(y < x\) satisfy \(P_{NRMH}^\zeta(x,y) = 0\). Hence, starting with a measure \(\mu_0 = \delta_1\), the larger states will be explored at a much slower rate than with the reversible MH. Quantitatively, the convergence rate of MH is \(1.2\) times as fast than NRMH.

We consider a second example that expands this paradox.

Example 2. Consider a random walk on a circle discretized in \(s\) states \(S = \{1, \ldots, s\}\), ordered following the counterclockwise direction. Contrarily to Example 1, \(\pi\) is now the uniform distribution on \(S\). The proposal distribution is defined for some \(\epsilon > 0\) and \((x,y) \in S\) as

\[
Q_\epsilon(x,y) = \begin{cases}
\epsilon & \text{if } x = y, \\
(1-\epsilon)/2 & \text{if } |x-y| = 1, \\
(1-\epsilon)/2 & \text{if } (x,y) \in (1,s) \cup (s,1).
\end{cases}
\]  

(8)

\(\text{Note that if } \epsilon = 0, \text{ the transition kernel } P_{\epsilon}^{MH} \text{ generated by MH does not satisfy } \lim_{t \to \infty} \sup_{x \in S} \left\| \delta_x P_{\epsilon}^{MH}^t - \pi \right\| = 0: \text{ the Hastings ratio is always equal to } 1 \text{ and thus the state at any time } t \text{ depends on the initialization i.e. the Markov chain is not irreducible.} \)
Figure 1: (Example 1) Illustration of MH (Alg. 1) and non-reversible MH (Alg. 2) with \( s = 50 \), \( n = 10 \) and \( \epsilon = 1/(1+n)^{-1} \). Note: the convergence plot is exact as the transition matrix of both algorithms is known, the distribution of the Monte Carlo estimate was obtained using \( L = 1,000 \) independent chains starting from \( \pi \) and length \( T = 10,000 \) for both algorithms. Other test functions of the type \( f : x \mapsto 1_{x \in i} \) for \( i \in S \) gave similar results. The last row illustrates a particular sample path of length \( T = 10,000 \) for both Markov chains. The left and centre plots represent the function \( \{(1+t/T) \cos(2\pi X_t/p), (1+t/T) \sin(2\pi X_t/p)\} \) for \( t = 1, \ldots, T \) for the MH and NRMH Markov chains, respectively. This shows that NRMH does explore the circle more efficiently.
Define for some $\zeta \in (0, \zeta_{\text{max}})$, the matrix $\Gamma_\zeta$ as in Example 1. It can be readily checked that setting $\zeta_{\text{max}} = (1 - \epsilon)/2s$ is sufficient to define a family of matrices $\{\Gamma_\zeta, \zeta \in (0, \zeta_{\text{max}})\}$ which satisfy Assumptions 1 and 2 and can thus serve as vorticity matrices in Algorithm 3. Note that when using $\zeta = 0$, NRMH corresponds to MH. When $\zeta$ increases, the probability of accepting moves in the counterclockwise direction is inflated while the probability to accept a clockwise bound move is decreased: introducing $\Gamma_\zeta$ results in defining a counterclockwise inertia to the Markov chain dynamic, whose intensity increases with $\zeta$.

Figure 2 reports the efficiency of NRMH in the context of Example 2 for different values of $\zeta \in [0, \zeta_{\text{max}}]$. As expected, imposing a privileged direction to the chain leads to a more accurate MC estimator of $\pi f$ when $f$ is the identity function, and the stronger the inertia the larger the variance reduction (see the table in Figure 2). On the other hand, the larger the inertia, the slower the algorithm converges, as shown by the plot in Figure 2. Thus it seems that in this example, the higher the inertia, the bigger the conflict between the speed of convergence and the accuracy of the estimators. The case $\zeta = 0$, corresponding to MH, yields to variance that can be up to 100 times larger than NRMH but converges more than 3 times faster. The slowness of NRMH implemented with $\zeta \approx \zeta_{\text{max}}$ results from the overwhelming counterclockwise flow that prevents a fast exploration of the larger states.

3. Two vorticity flows and a skew-detailed balance condition

Is it possible to modify NRMH so that the Markov chain $\{X_t, t > 0\}$ reaches equilibrium faster while retaining its good asymptotic properties? Motivated by the observations made at Examples 1 and 2 we introduce a skew-symmetric perturbation in the MH Markov chain associated with a large vorticity parameter in an attempt to reduce the variance of MC estimators while simultaneously allowing the non-reversible dynamic direction to switch throughout the algorithm so as to mitigate convergence issues. Borrowing from the lifting literature [16, 18], we introduce an auxiliary variable $\xi \in \{-, +\}$ that materializes the flow direction of the chain. We consider the enlarged state space $\mathcal{S} \times \{-, +\}$ and aim at constructing a $\tilde{\pi}$-invariant Markov chain, where $\tilde{\pi}$ is the distribution on the enlarged state space defined as:

$$\tilde{\pi}(x, +) = \tilde{\pi}(x, -) = \pi(x)/2, \quad \forall x \in \mathcal{S}.$$  (10)
Figure 2: (Example 2). Parameters: $s = 50$ and $\epsilon = 10^{-1}$. **Plot.** Evolution of the TV distance $\|\pi - \delta_1 \{ P^{NRMH}_{\epsilon,\zeta} \}^t \|$ along the iterations of NRMH on a circle with $s$ states for different values of $\zeta$. **Table.** Variance of the MC estimate of $\mathbb{E}(X)$ obtained with NRMH for different values of $\zeta$, started under $\pi$ and after $T = 1,000$ iterations. Variances were estimated using $L = 500$ i.i.d Markov chain simulations.

| $\zeta / \zeta_{\text{max}}$ | 0   | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1   |
|-------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Var($\bar{x}$)                | 34.65 | 22.73 | 14.08 | 6.37 | 4.05 | 2.44 | 1.82 | 1.14 | 0.77 | 0.50 | 0.36 |
Algorithm 3 Non-reversible Metropolis-Hastings algorithm with auxiliary variable.

1: Initialize in \((X_0, \xi_0) \sim \mu_0\)
   
   Transition \((X_t, \xi_t) = (x, \xi) \rightarrow (X_{t+1}, \xi_{t+1})\):

2: Propose \(Y \sim Q(x, \cdot) \sim y\)

3: Set \((X_{t+1}, \xi_{t+1}) = (y, \xi)\) with probability \(A_{\Gamma\xi}(x, y) = 1 \land R_{\Gamma\xi}(x, y)\) where

\[
R_{\Gamma}(x, y) := \begin{cases} 
\frac{(\Gamma(x, y) + \pi(y)Q(y, x))}{\pi(x)Q(x, y)} & \text{if } \pi(x)Q(x, y) \neq 0 \\
1 & \text{otherwise}
\end{cases}
\]  

(9)

4: If the move attempted at step 3: is rejected, set \((X_{t+1}, \xi_{t+1}) = (x, -\xi)\) with probability \(\delta\)

5: If the moves attempted at step 3: and 4: are both rejected, set \((X_{t+1}, \xi_{t+1}) = (x, \xi)\)

Consider two vorticity matrices, denoted \(\Gamma^+\) and \(\Gamma^-\), each corresponding to a flow direction. We propose Algorithm 3 referred to as non-reversible Metropolis-Hastings with auxiliary variable (NRMHAV). With some abuse of notation, we identify \(\xi\) with the sign of \(\xi a\), for any real number \(a > 0\). To ensure that NRMHAV is \(\tilde{\pi}\)-invariant, both vorticity matrices \(\Gamma^+\) and \(\Gamma^-\) need to satisfy Assumptions 1 and 2 along with a condition referred to as the skew-detailed balance condition (SDBC) defined as:

Assumption 3. For all \((x, y) \in S^2\),

\[
\pi(x)Q(x, y)A_{\Gamma^+}(x, y) = \pi(y)Q(y, x)A_{\Gamma^-}(y, x) .
\]

The terminology SDBC is borrowed from [6, Eq. 29] in which the authors specify a condition similar to Assumption 3. It can be thought as a constraint that compensates the flow intensity of the two skew-symmetric perturbations at a global level. Algorithm 3 simulates a Markov chain characterized by the
transition kernel

\[
K((x, \xi), (y, \xi')) = \begin{cases} 
Q(x, y)A_{\Gamma \xi}(x, y) & \text{if } x \neq y, \xi' = \xi, \\
Q(x, x)A_{\Gamma \xi}(x, x) + (1 - \delta) \int_{\mathcal{S}} Q(x, dy) \{1 - A_{\Gamma \xi}(x, y)\} & \text{if } x = y, \xi' = \xi, \\
\delta \int_{\mathcal{S}} Q(x, dy) \{1 - A_{\Gamma \xi}(x, y)\} & \text{if } x = y, \xi' = -\xi, \\
0 & \text{otherwise}.
\end{cases}
\]

(11)

Theorem 1. Let \(Q\) be a transition kernel on the marginal space \(\mathcal{S} \times \mathcal{S}\), \(\pi\) a distribution on \(\mathcal{S}\) that is nowhere zero and two vorticity matrices \(\Gamma^+\) and \(\Gamma^-\) defined such that Assumptions 1, 2 and 3 hold. Define the distribution \(\tilde{\pi}\) on the enlarged state space \(\tilde{\mathcal{S}} = \mathcal{S} \times \{-, +\}\) as in Eq. (10). Then the transition kernel \(K\) defined by Algorithm 3 is a \(\tilde{\pi}\)-invariant Markov kernel and is \(\tilde{\pi}\)-reversible if and only if \(\Gamma^+ = \Gamma^- = 0\).

This result is proved for a discrete state space \(\mathcal{S}\) in Appendix A but its extension to the general state space case is straightforward.

Example 2 (continued). Algorithm 3 (NRMH with auxiliary variable) is implemented to sample from the distribution \(\pi\) of Example 2. This illustrates the potential gains one can obtain from the introduction of such changes of direction in the vector field. First, define \(\Gamma^+ = \Gamma^- = -\Gamma\) where \(\Gamma\) has been introduced in Example 2 and \(\zeta \in (0, \zeta_{\text{max}})\). Figure 3 shows that when changes of direction are rare and the flow inertia is high (i.e. \(\delta \approx 0\) and \(\zeta \approx \zeta_{\text{max}}\), NRMHAV yields an estimator of \(\pi f\) (with \(f\) the identity function in this case) with a variance significantly reduced compared to MH but slightly larger than NRMH. Remarkably, the rate of convergence of NRMHAV is considerably faster than NRMH (which corresponds to the TV plot that dominates all the other ones) and MH (in bold). Hence the introduction of a vorticity matrix with a large inertia parameter coupled with a direction switching parameter with reasonably low intensity leads, in this example, to a better algorithm than MH (both in finite time and asymptotic regime). Moreover, it inherits from NRMH its variance reduction feature while avoiding its dramatically slow speed of convergence.
Figure 3: (Example 2). Parameters $s = 50$, $\epsilon = 10^{-1}$ and $\zeta = \zeta_{\text{max}} - 10^{-5}$. **Plot.** Evolution of the total variation distance along the iterations of NRMHAV for different values of $\delta$ and initial distribution $\mu_0 = \delta_1$. **Table.** Variance of the MC estimate of $E(X)$ obtained with NRMHAV for different values of $\delta$. Results obtained with $L = 500$ i.i.d Markov chains of length $T = 10^3$. 

| $\delta$ | 0   | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Var($\bar{X}$) | 0.41 | 7.14 | 10.97 | 15.96 | 22.22 | 26.19 | 29.19 | 31.73 | 29.77 | 32.18 |
4. Two vorticity flows without a skew-detailed balance condition

The previous section has shown that designing a Markov chain that browses the state space using several privileged directions may be highly beneficial. However, such Markov chains assume the existence of two vorticity matrices $\Gamma^+$ and $\Gamma^-$ that satisfy Assumption 3. When one considers increasing complexity examples, finding vorticity matrices that verify this assumption might be difficult – if not impossible. This is for instance the case in the following example.

Example 3. Consider a discrete two-dimensional random walk on a grid of size $10 \times 10$, i.e. the state space is $\mathcal{S} = \{1, 2, ..., 100\}$ and a target distribution $\pi$ specified by the “sigma” pattern $\Psi$ (in dark gray) presented at Figure 4. In particular, $\pi(x) \propto 1$ if $x \not\in \Psi$ and $\pi \propto 10$ if $x \in \Psi$. We consider a random walk algorithm whose proposal $Q$ allocates uniform weights on the neighbors of the current state (north, south, east, west).

Contrarily to Example 2, it is more difficult to find an algorithm that builds a vorticity matrix that satisfies Assumptions 1–3. Indeed, in Example 2 a vorticity matrix that satisfies Assumptions 1 and 2 automatically satisfies Assumption 3, but this is not the case in Example 3. We now present how a slight modification of Algorithm 3 can simulate a $\tilde{\pi}$-invariant Markov chain, even in absence of Assumption 3, by drawing the auxiliary variable $\xi_t$ independently of $X_t$. This algorithm is presented in Algorithm 4.
and referred to as the non-reversible MH with decorrelated auxiliary variable (NRMHAVd).

**Algorithm 4** Decorrelated NRMHAV (NRMHAVd).

1: Initialize in $(X_0, \xi_0) \sim \mu_0$
   Transition $(X_t, \xi_t) = (x, \xi) \rightarrow (X_{t+1}, \xi_{t+1})$
2: Refresh the direction, i.e. draw $\xi_{t+1} \sim \eta$ such that $\xi_{t+1} = \xi$ w.p. $\delta$ and $\xi_{t+1} = -\xi$ w.p. $1 - \delta$.
3: Set $\Gamma = \Gamma^\eta$
4: Propose $Y \sim Q(x, \cdot) \sim y$
5: Set $(X_{t+1}, \xi_{k+1}) = (y, \eta)$ w.p. $A_{\Gamma}(x, y) = 1 \land R_{\Gamma}(x, y)$ where
   \begin{equation}
   R_{\Gamma}(x, y) := \begin{cases}
   \delta A_{\Gamma}(x, y)Q(x, y) / \pi(x)Q(x, y) & \text{if } \pi(x)Q(x, y) \neq 0 \\
   1 & \text{otherwise}
   \end{cases}
   \end{equation}
6: If the proposal is rejected, set $(X_{t+1}, \xi_{t+1}) = (x, \eta)$

This algorithm simulates a Markov chain with transition kernel $K$, defined as:

$$K((x, \xi), (y, \eta)) = \begin{cases}
    \delta A_{\Gamma}(x, y)Q(x, y) & \text{if } x \neq y, \eta = \xi, \\
    (1 - \delta)A_{\Gamma-\xi}(x, y)Q(x, y) & \text{if } x \neq y, \eta = -\xi, \\
    \delta \left\{ Q(x, x) + \sum_{z \neq x} Q(x, z)A_{\Gamma}(x, z) \right\} & \text{if } x = y, \eta = \xi, \\
    (1 - \delta) \left\{ Q(x, x) + \sum_{z \neq x} Q(x, z) (1 - A_{\Gamma-\xi}(x, z)) \right\} & \text{if } x = y, \eta = -\xi.
\end{cases}$$

**Theorem 2.** Let $\Gamma^+$ and $\Gamma^-$ be two vorticity matrices, $Q$ a transition kernel on $(S, \mathcal{S})$ and $\pi$ a distribution on $S$ such that Assumptions 1 and 2 hold. Define the distribution $\bar{\pi}$ on the enlarged state space as in Eq. (10). Then the transition kernel $K$ defined by Algorithm 4 is a Markov kernel, is $\bar{\pi}$-invariant, and is $\bar{\pi}$-reversible if and only if $\Gamma^+ = \Gamma^- = 0$.

**Proof.** The kernel $K$ given by Eq. 13 is a valid Markov transition kernel as for all $(x, \xi), (y, \eta) \in \{S \times \{-, +\}\}^2$, $K(x, \xi; y, \eta) \geq 0$ and $\sum_{y \in S} K(x, \xi; y, \eta) +$
\[ K(x; y, -\eta) \} = 1. \] The invariance is a corollary of the \( \pi \)-invariance of the NRMH transition kernel [11, Theorem 2.5]. Indeed, denoting by \( K^+ \) and \( K^- \) the NRMH kernels associated to the vorticity matrix \( \Gamma^+ \) and \( \Gamma^- \) respectively, we have that:

\[
\Pr(X_{t+1} \in A, \xi = +) = \sum \pi(y, +) Pr(X_{t+1} \in A, \xi = + | X_t = y, \xi = +) \\
+ \sum \pi(y, -) Pr(X_{t+1} \in A, \xi = + | X_t = y, \xi = -),
\]

\[
= (1/2) \sum \pi(y) \delta K^+(y, A) \\
+ (1/2) \sum \pi(y) (1 - \delta) K^-(y, A),
\]

\[
= (1/2) \delta \pi(A) + (1/2)(1 - \delta) \pi(A) = \pi(A).
\]

Regarding the reversibility of NRMHAVd, taking \( x \neq y \), we have

\[
\tilde{\pi}(x, \xi) K(x, \xi; y, \xi') = \begin{cases} \\
(\delta/2) \pi(x) Q(x, y) A_{\Gamma^+}(x, y) & \text{if } \xi' = \xi \\
(1 - \delta)/2 \pi(x) Q(x, y) A_{\Gamma^-}(x, y) & \text{if } \xi' = -\xi \\
(\delta/2) \pi(y) Q(y, x) \left(A_{\Gamma^+}(y, x) + \Gamma^\xi(x, y)/\pi(y) Q(y, x)\right) & \text{if } \xi' = \xi \\
(1 - \delta)/2 \pi(y) Q(y, x) \left(A_{\Gamma^-}(y, x) + \Gamma^-\xi(x, y)/\pi(y) Q(y, x)\right) & \text{if } \xi' = -\xi
\end{cases}
\]

and NRMHAVd is \( \tilde{\pi} \)-reversible if and only if \( \Gamma^+ = \Gamma^- = 0 \).

Provided that \( \delta \) is large enough, Algorithm 4 simulates a Markov chain that has two privileged directions imposed by the vorticity matrices. Interestingly, this construction does not require \( (\Gamma^+, \Gamma^-) \) to satisfy the SDBC and therefore more complex distributions (such as Example 3) can be simulated using NRMHAVd. Nevertheless, Algorithm 4 drops the appealing feature of Algorithm 3 that a change of direction occurs with probability one as soon as the marginal chain stays put. In the construction of Algorithm 4, the trajectory of the direction spins \( \{\xi_t, t \in \mathbb{N}\} \) is decorrelated from the marginal chain \( \{X_t, t \in \mathbb{N}\} \), hence the acronym NRMHAVd. Turning back to Example 3, Figure 5 compares the NRMH (Alg. 2) and NRMHAVd (Alg. 4), for two different initializations of the marginal chain \( \{X_t, t \in \mathbb{N}\} \).
Remark 1. Example 2 features a high degree of symmetry and therefore using $\Gamma^+_{\zeta}$ or $\Gamma^-_{\zeta}$ in Algorithm 2 or initializing the auxiliary variable $\xi_0$ with $+$ or $-$ in Algorithm 3 do not have any influence on the results of the simulation. In the case of Example 3, the initialization of $\xi_0$ in Algorithm 4 and the choice of $\Gamma \in \{\Gamma^+_{\zeta}, \Gamma^-_{\zeta}\}$ in Algorithm 5 clearly matter.

Remark 2. Compared to Example 2, the vorticity matrices used in Example 3 do not have a straightforward geometric interpretation: their structure is such that the acceptance probabilities are not forcing the chain to move according to a global privileged direction but rather using a local privileged direction depending on the current location on the state space (see the description of the vorticity field designed for this Example in Appendix B and illustrated in Figure B.6).

Remark 3. Note that in the analysis of Example 3, we only focus on the convergence time of the algorithms. In this example, there is no significant difference between the variance of empirical estimators for the two algorithms implemented with different values of $\zeta$ and $\delta$.

The two rows of Figure 5 gives the convergence time of NRMH (left) and NRMHAVd (right). For NRMH, the convergence time is given for different values of the vorticity parameter $\zeta$ while for NRMHAVd, the convergence time is reported for different switching parameters $\delta$. For each plot, the two plain curves correspond to the implementation of the algorithm with the two opposite vorticity parameters $\zeta$ and $-\zeta$. We used NRMH- (resp. NRMHAVd-) and NRMH+ (resp. NRMHAVd+) for the shorthand notation of the two algorithms implemented with the vorticity parameter $\zeta$ and $-\zeta$.

On the north-western dial, one can see that using the vorticity matrix $\Gamma^+_{\zeta}$ in NRMH (red) leads to a faster convergence time compared to MH (black dashed line), regardless the value of $\zeta$. However, using $\Gamma^-_{\zeta}$ (blue) leads to much slower algorithms for all values of $\zeta$. As a consequence, the non-reversibility of NRMH carries an implicit risk on the finite-time efficiency of the algorithm: if the “wrong” privileged direction is selected (i.e. if, in this example, the vorticity matrix $\Gamma^-_{\zeta}$ is used), the time needed to reach the steady-state can be large, and even larger than MH. The north-eastern dial plots the convergence time of NRMHAVd for a fixed value of $\zeta$ and different values of $\delta$. For a well-chosen value of $\delta$, the convergence time of NRMHAVd

\[ \text{[2]} \text{The plot is displayed only for the highest values of } \delta, \text{which are actually the most} \]
can be smaller than NRMH using the same value of $\zeta$, the initialization of $\xi$ being either $+$ (red) or $-$ (blue). Note that MH converges faster than NRMHAVd initialized with $\Gamma^\zeta_\xi$. This means that, while mitigated compared to NRMH, NRMHAVd comprises the inherent risk to start with an unsuitable vorticity flow.

The initial measure using in the second row leads to more favorable results both for NRMH and NRMHAVd. We observe on the south-western dial that using $\Gamma^+\zeta$ or its opposite give better convergence times than reversible MH. The south-eastern dial shows that in that case, introducing an auxiliary variable can lead to better convergence times in both cases: if for instance we take $\delta = 0.99$, NRMHAVd is faster than MH and NRMH, regardless the initialisation of the direction spin.

This example does not lead to results as clear as in Example 2. This is mainly due to the more complex structure of $\pi$ that does not exhibit a particular direction. However, introducing an auxiliary variable in NRMH can, in the best case, beat the convergence time of MH and NRMH, and, in the worst cases, lower the risk of using NRMH with an inefficient vorticity flow.

5. Discussion

Considering a $\pi$-invariant Markov kernel $K$, it is possible to define a measure of non-reversibility $\rho(K)$ as the L1 distance between the measures $\pi(dx)K(x,dy)$ and $\pi(dy)K(y,dx)$. For some Markov chains, $\rho(K)$ is computable exactly:

- for NRMH, $\rho(K)$ depends on $\Gamma$ and is equal to
  \[
  \rho(K_{\text{NRMH}}) = \sum_{x \in S} \sum_{y \neq x} \left| \pi(x)Q(x,y)A_\Gamma(x,y) - \pi(y)Q(y,x)A_\Gamma(y,x) \right|
  \]

- for both NRMHAV and NRMHAVd, $\rho(K)$ is independent of the change

\begin{itemize}
  \item interesting ones: when $\delta$ is small, the convergence times are really close for NRMHAVd± and decrease when $\delta$ roams $(0,0.7)$.
\end{itemize}
of spin probability $\delta$ and equals

$$\rho(K_{\text{NRMH}}) = \rho(K_{\text{NRMHAV}})$$

$$= \left( \frac{1}{2} \right) \sum_{x \in \mathcal{S}} \sum_{y \neq x} \left| \pi(x) Q(x, y) \left( A_{\Gamma^+}(x, y) + A_{\Gamma^-}(x, y) \right) - \pi(y) Q(y, x) \left( A_{\Gamma^+}(y, x) + A_{\Gamma^-}(y, x) \right) \right|.$$

For the discrete examples 2 and 3 presented above, we observe that for NRMH, $\rho(K)$ increases with $\zeta$. Experimentally, we found that $\rho(K_{\text{NRMHAV}})$ is significantly smaller than $\rho(K_{\text{NRMH}})$. This result is somewhat intuitive: the most “irreversible” algorithms are the ones that impose one (and only one) privileged direction to the Markov chain. The second type of algorithms, which involve changes of direction, might lead to algorithms that are “less irreversible”. In some sense, the switching parameter compensates the introduction of the irreversible flow.

At this stage, one can question the rationale for introducing a skew-symmetric perturbation along with a strategy aimed to mitigate its effects. The examples studied in this paper confirm the utility of non-reversible algorithms so as to get more accurate Monte Carlo estimates (in stationary regime). However, since in practice the Markov chain does not start from stationarity, a practitioner may be also interested in speeding up the transient phase. In this respect, our work shows that a strong irreversibility (as measured by $\rho(K)$) tend to slow down the transient phase of the Markov chain. It might be therefore desirable to make some concessions on the strength of the non-reversibility, at least in the transient phase. We have presented a direct approach (NRMHAV) that consists in introducing different vorticity matrices so as to have multiple privileged directions. The price to pay is the skew-detailed balance condition (SDBC) that needs to be satisfied for the different flows and which may be challenging to guarantee in some scenario. We have proposed a simple adaptation of NRMHAV, namely NRMHAVd, that bypasses the need of this condition. Remarkably, the level of irreversibility of NRMHAV and NRMHAVd is the same, implying that their asymptotic efficiency is similar (as supported by our example).

This work, mostly experimental, open several research questions. First, its extension to continuous state spaces needs to be addressed. Theoretically the algorithm exists but its implementation is not straightforward.
The difficulty lies in the definition of the “vorticity density” which is more problematic to find than a vorticity matrix because an upper bound of the function \((x, y) \mapsto \pi(x)Q(x, y)\) is, in most settings, difficult to identify (Assumption 2). If an upper bound cannot be found, NRMH will not generate a \(\pi\)-stationary Markov chain. However, assuming that a lower bound on the probability of the MH ratio to be positive exists, \(\text{i.e.} \Pr\{\gamma(X, X') > -\pi(X')Q(X', X)\} \geq 1 - \varepsilon\), if \(\varepsilon\) is small enough, using NRMH with acceptance ratio \(\max(0, A_T(X, X'))\) would lead to an inexact algorithm but with stationary distribution (if it exists) probably close enough from \(\pi\). In this respect, recent developments on the perturbation theory of Markov chains (see \(e.g.\) [15]) will be useful.

It could also be interesting to study non-reversible chains that have more than two possible privileged directions, in order to lower the conflict between convergence time and precision of the Monte Carlo estimates. Designing an adaptive vorticity parameter that would learn the topology and the privileged directions of \(\pi\) on the fly (\(e.g.\) using a PCA online algorithm) could also be interesting. Finally, recent works that have been developed around generalized MALA algorithms [8, 13] open interesting perspectives in that framework. We expect this type of algorithms to be “less irreversible” than the non-reversible MH algorithms presented in this paper, but since the hypotheses that guarantee \(\pi\)-stationarity of the Markov chain are weaker than in our case, they might be more widely applicable.

At a more general level, the concept of irreversibility measure of a Markov chain deserves to be further developed at a theoretical level. In particular, one can wonder if a (partial) ordering of MCMC algorithms according to their irreversibility measure can be established, in a Peskun ordering style [12] for non-reversible Markov chains.

Acknowledgements. This research work was funded by ENSAE ParisTech and the Insight Center for Data Analytics – University College Dublin.

References

[1] Bierkens, J., 2016. Non-reversible Metropolis-Hastings. Statistics and Computing 26 (6), 1213–1228.
URL https://doi.org/10.1007/s11222-015-9598-x

[2] Bierkens, J., Roberts, G., 2016. A piecewise deterministic scaling limit of lifted Metropolis–Hastings in the Curie–Weiss model. The Annals of
[3] Bouchard-Côté, A., Vollmer, S. J., Doucet, A., 2017. The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method. Journal of the American Statistical Association.

[4] Chen, T.-L., Hwang, C.-R., 2013. Accelerating reversible Markov chains. Statistics & Probability Letters 83 (9), 1956–1962.

[5] Diaconis, P., Holmes, S., Neal, R. M., 2000. Analysis of a nonreversible Markov chain sampler. Annals of Applied Probability 10 (3), 726–752. URL http://www.jstor.org/stable/2667319

[6] Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (6), 721–741.

[7] Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.

[8] Ma, Y.-A., Fox, E. B., Chen, T., Wu, L., 2018. Irreversible samplers from jump and continuous Markov processes. Statistics and Computing, 1–26.

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

[10] Neal, R. M., 2004. Improving asymptotic variance of MCMC estimators: Non-reversible chains are better. arXiv preprint math/0407281.

[11] Ottobre, M., 2016. Markov chain Monte Carlo and irreversibility. Reports on Mathematical Physics 77 (3), 267–292.

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

[13] Poncet, R., 2017. Generalized and hybrid Metropolis-Hastings overdamped Langevin algorithms. arXiv preprint arXiv:1701.05833.
Appendix A. Proof of the invariance of NRMHAV

We prove Theorem 1 that states that the transition kernel \( K \) generated by Algorithm 3 is a proper transition kernel, is \( \bar{\pi} \)-invariant and is non-reversible if and only if \( \Gamma^+ = \Gamma^- = 0 \).

**Proof.** It is straightforward to prove that \( K \) is indeed a transition kernel. Note that it is not necessary to impose any condition on the parameter \( \delta \) to obtain a valid transition kernel, so the only constrained parameters is the vorticity matrix \( \Gamma \). To prove the invariance of \( K \), we need to prove that

\[
\sum_{y,\eta} \bar{\pi}(y, \eta)K(y, \eta; x, \xi) = \bar{\pi}(x, \xi).
\]

\[
\sum_{y,\eta} \bar{\pi}(y, \eta)K(y, \eta; x, \xi) = \sum_y \bar{\pi}(y, \xi)K(y, \xi; x, \xi) + \sum_y \bar{\pi}(y, -\xi)K(y, -\xi; x, \xi)
\]

\[
= \bar{\pi}(x, \xi)K(x, \xi; x, \xi) + \bar{\pi}(x, -\xi)K(x, -\xi; x, \xi) + \sum_{y \neq x} \bar{\pi}(y, \xi)K(y, \xi; x, \xi)
\]

\[
= \bar{\pi}(x, \xi)\left\{Q(x, x)A_{\Gamma\xi}(x, x) + (1 - \delta)\sum_z Q(x, z)(1 - A_{\Gamma\xi}(x, z))
\right. \\
+ \delta \sum_z Q(x, z)(1 - A_{\Gamma-\xi}(x, z)) \left. \right\} + \sum_{y \neq x} \bar{\pi}(y, \xi)Q(y, x)A_{\Gamma\xi}(y, x)
\]

(A.1)
the second equality coming from the fact that $K(y, -\xi; x, \xi) \neq 0$ iff $x = y$ and
the third from the fact that $\tilde{\pi}(x, \xi) = \tilde{\pi}(x, -\xi) = \pi(x)/2$. Now, let $A(x, \xi) := \sum_{y \neq x} \tilde{\pi}(y, \xi)Q(y, x)A_{\Gamma^\xi}(y, x)$ and $A(x) := \{y \in S \mid y \neq x \text{ and } \pi(y)Q(y, x) \neq 0\}$. By definition of $A_{\Gamma^\xi}$, we have:

$$A(x, \xi) = \sum_{\substack{y \neq x \in A(x) \setminus \{x\}}} (1/2)\pi(y)Q(y, x)$$

$$+ \sum_{y \in A(x)} (1/2)\pi(y)Q(y, x) \left\{ 1 \wedge \frac{\Gamma^\xi(y, x) + \pi(x)Q(x, y)}{\pi(y)Q(y, x)} \right\},$$

$$= (1/2) \sum_{y \in A(x)} \pi(y)Q(y, x) \wedge \{\Gamma^\xi(y, x) + \pi(x)Q(x, y)\},$$

$$= (1/2) \sum_{y \in A(x)} \{\pi(y)Q(y, x) - \Gamma^\xi(y, x)\} \wedge \pi(x)Q(x, y) + (1/2) \sum_{y \in A(x)} \Gamma^\xi(y, x),$$

$$= (1/2) \sum_{y \in A(x)} \pi(x)Q(x, y)A_{\Gamma^\xi}(x, y) + (1/2) \sum_{y \in A(x)} \Gamma^\xi(y, x),$$

$$= \tilde{\pi}(x, \xi) \sum_{y \in A(x)} Q(x, y)A_{\Gamma^\xi}(x, y) + (1/2) \sum_{y \in A(x)} \Gamma^\xi(y, x),$$

$$= \tilde{\pi}(x, \xi) \sum_{y \neq x} Q(x, y)A_{\Gamma^\xi}(x, y).$$

(A.2)

The fourth equality follows by skew-symmetry of $\Gamma^\xi$ and the last equality from $\Gamma^\xi 1 = 0$ along with the lower-bound condition on $\Gamma$ implying that
$\Gamma(x, y) = 0$ if $Q(x, y) = 0$ (Assumptions 1 and 2). Similarly, define $B(x, \xi) := \tilde{\pi}(x, \xi) \sum_{z} Q(x, z) \left\{ (1 - \delta)(1 - A_{\Gamma^\xi}(x, z)) + \delta(1 - A_{\Gamma^{-\xi}}(x, z)) \right\}$ and using Lemma
we have:

\[ B(x, \xi) = \tilde{\pi}(x, \xi) \sum_{z \in S} Q(x, z)(1 - A_{\Gamma \xi}(x, z)) \]

\[ + \delta \tilde{\pi}(x, \xi) \sum_{z \in S} Q(x, z) \{ A_{\Gamma \xi}(x, z) - A_{\Gamma - \xi}(x, z) \} , \]

\[ = \tilde{\pi}(x, \xi) \sum_{z \in S} Q(x, z)(1 - A_{\Gamma \xi}(x, z)), \]

\[ = \tilde{\pi}(x, \xi) \sum_{z \neq x} Q(x, z)(1 - A_{\Gamma \xi}(x, z)), \]

\[ = \tilde{\pi}(x, \xi) \sum_{z \neq x} Q(x, z) - A(x, \xi), \quad (A.3) \]

where the penultimate equality follows from \( A_{\Gamma \xi}(x, x) = 1 \) for all \( x \in S \) and the last one from Eq. \( (A.2) \). Finally, combining Eqs. \( (A.1) \) and \( (A.3) \), we obtain:

\[ \sum_{y, \eta} \tilde{\pi}(y, \eta) K(y, \eta; x, \xi) = \tilde{\pi}(x, \xi)Q(x, x)A_{\Gamma \xi}(x, x) + B(x, \xi) + A(x, \xi), \]

\[ = \tilde{\pi}(x, \xi)Q(x, x)A_{\Gamma \xi}(x, x) + \tilde{\pi}(x, \xi) \sum_{z \neq x} Q(x, z), \]

\[ = \tilde{\pi}(x, \xi). \]

We now study the \( \tilde{\pi} \)-reversibility of \( K \), i.e. conditions on \( \Gamma^\xi \) such that for all \( (x, y) \in S^2 \) and \( (\xi, \eta) \in \{-, +\}^2 \) such that \( (x, \xi) \neq (y, \eta) \), we have:

\[ \tilde{\pi}(x, \xi)K(x, \xi; y, \eta) = \tilde{\pi}(y, \eta)K(y, \eta; x, \xi). \quad (A.4) \]

First note that if \( x = y \) and \( \xi = -\eta \), then Eq. \( (A.4) \) is equivalent to

\[ \sum_{z \in S} Q(x, y) (A_{\Gamma \xi}(x, z) - A_{\Gamma \eta}(x, z)) = 0 \]

which is true from Lemma \( 3 \) and the fact that \( \pi \) is non-zero almost everywhere. Second, for \( x \neq y \) and \( \xi \neq \eta \), Eq. \( (A.4) \) is trivially true by definition of \( K \). Hence, condition(s) on the vorticity matrix to ensure \( \tilde{\pi} \)-reversibility are to be investigated only for the case \( \xi = \eta \) and \( x \neq y \). In such a case Eq. \( (A.4) \) is equivalent to

\[ \pi(x)Q(x, y)A_{\Gamma \xi}(x, y) = \pi(y)Q(y, x)A_{\Gamma \xi}(y, x) \Rightarrow \Gamma^\xi(x, y) = 0, \]
by direct application of Eq. (A.5). Hence $K$ is $\bar{\pi}$-reversible if and only if $\Gamma^+ = \Gamma^- = 0$. □

**Lemma 3.** In the context of Algorithm 3 (NRMHAV) satisfying Assumptions 1, 2 and 3, we have for all $x \in S$ and $\xi \in \{+, -\}$

$$\pi(x) \sum_{z \in S} Q(x, z) \{A_{\Gamma^\xi}(x, z) - A_{\Gamma^{-\xi}}(x, z)\} = 0.$$ 

**Proof.** Using that for three real numbers $a, b, c$, we have $a \wedge b = (a \wedge b - c) + c$, together with the fact that $\Gamma^\xi(x, y) = -\Gamma^\xi(y, x)$, we have:

$$\pi(x)Q(x, y)A_{\Gamma^\xi}(x, y) = \pi(x)Q(x, y)\left\{1 \wedge \frac{\Gamma^\xi(x, y) + \pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right\},$$

$$= \pi(y)Q(y, x)\left\{1 \wedge \frac{\Gamma^\xi(y, x) + \pi(x)Q(x, y)}{\pi(y)Q(y, x)}\right\} + \Gamma^\xi(x, y),$$

$$= \pi(y)Q(y, x)A_{\Gamma^\xi}(y, x) + \Gamma^\xi(x, y). \quad (A.5)$$

The proof follows from combining the skew-detailed balance equation (10) and Eq. (A.5):

$$\pi(x) \sum_{z \in S} Q(x, z) \{A_{\Gamma^\xi}(x, z) - A_{\Gamma^{-\xi}}(x, z)\}$$

$$= \sum_{z \in S} \{\pi(x)Q(x, z)A_{\Gamma^\xi}(x, z) - \pi(x)Q(x, z)A_{\Gamma^{-\xi}}(x, z)\},$$

$$= \sum_{z \in S} \{\pi(x)Q(x, z)A_{\Gamma^\xi}(x, z) - \pi(z)Q(z, x)A_{\Gamma^\xi}(z, x)\},$$

$$= \sum_{z \in S} \Gamma^\xi(x, z),$$

$$= 0. \quad \Box$$

**Appendix B. Generation of vorticity matrices on $s \times s$ grids**

The reader can find in Algorithm 5 the method used to generate vorticity matrices in the set-up of the Example 3. In the general case of a random walk on an $s \times s$ grid, $\Gamma^\xi$ is an $s^2 \times s^2$ matrix that can be constructed systematically.
Algorithm 5 Method for generating a vorticity matrix satisfying the necessary conditions for the invariance of the decorrelated NRMHAV algorithm, when the state space is a $s \times s$ grid.

Parameters:

- $s$ the size of the grid ($\Rightarrow s^2$ states, $s$ rows and $s$ columns)
- $\zeta$ the vorticity parameter s.t. $0 \leq \zeta \leq \min_x \pi(x)/s$

Let $\Gamma_\zeta$ be a $s^2 \times s^2$ matrix with null diagonal

1: **Perturbation of the acceptance probabilities for the states that have 4 neighbors**

- Start with filling the entries of $\Gamma_\zeta$ corresponding to states that are not located on a border of the grid
  
  - for all the states $i$ of the second column of the grid that have 4 neighbors (i.e. states 6 and 7 in a $4 \times 4$ grid), set $\Gamma_\zeta(i, i-1) = -\zeta$ and $\Gamma_\zeta(i, i+1) = \zeta$
  
  - for all the states $i$ of the third column of the grid that have 4 neighbors (i.e. states 10 and 11 in a $4 \times 4$ grid), set $\Gamma_\zeta(i, i-1) = \zeta$ and $\Gamma_\zeta(i, i+1) = -\zeta$
  
  - if $s$ is even, repeat the operation until the $(s-1)^{th}$ column of the grid
  
  - if $s$ is odd, repeat the operation until the $(s-2)^{th}$ column of the grid
  
  - for all the rows of $\Gamma_\zeta$ corresponding to a state with 4 neighbours on the grid, fill the empty entries with zeros
  
  - for the rows $i$ of $\Gamma_\zeta$ corresponding to a state with 4 neighbours on the grid, for $j \in \{1, ..., s^2\}$ do $\Gamma_\zeta(j, i) = -\Gamma_\zeta(i, j)$

2: **Borders of the grid – North-western corner**

- Complete the $(s+1)^{th}$ line of $\Gamma_\zeta$, corresponding to the eastern neighbour of the north-western corner of the grid: $\Gamma_\zeta(s+1, 1) = -\Gamma_\zeta(s+1, s+2)$, and apply the skew-symmetry property of $\Gamma_\zeta$: $\Gamma_\zeta(1, s+1) = -\Gamma_\zeta(s+1, 1)$

- Complete the 1$^{st}$ line of $\Gamma_\zeta$, corresponding to the north-western corner of the grid: $\Gamma_\zeta(1, 2) = -\Gamma_\zeta(1, s+1)$, and apply the skew-symmetry property of $\Gamma_\zeta$: $\Gamma_\zeta(2, 1) = -\Gamma_\zeta(1, 2)$

- Fill the empty entries of lines 1 and $s+1$ of $\Gamma_\zeta$ by zeros

- For $i \in \{1, s+1\}$, for $j \in \{1, ..., s^2\}$, do $\Gamma_\zeta(j, i) = -\Gamma_\zeta(i, j)$

3: **Borders of the grid – Rest of the borders**

- The rest of the matrix can be filled row by row; for each row $i$ that still has empty elements, do the following:

  - In the set $v_i$ of neighbours of $i$, take the first state $j$ such that $\Gamma_\zeta(i, j)$ is empty, and set $\Gamma_\zeta(i, j) = -\sum_k \text{s.t. } \Gamma_\zeta(i, k) \neq \text{NA} \Gamma_\zeta(i, j)$: doing this leads the entries of $\Gamma_\zeta$ to satisfy the skew-symmetry condition and $\Gamma_\zeta 1 = 0$

  - Then fill the rest of the line with zeros

  - For $j$ in $\{1, ..., s^2\}$ do $\Gamma_\zeta(j, i) = -\Gamma_\zeta(i, j)$
using the properties that $\Gamma_\zeta(x, y) = -\Gamma_\zeta(y, x)$ for all $(x, y) \in \mathbb{S}^2$ and $\Gamma_\zeta 1 = 0$. It has a block-diagonal structure:

$$\Gamma_\zeta = \begin{pmatrix}
B & 0 & 0 & \cdots & 0 \\
0 & B & 0 & \cdots & 0 \\
0 & 0 & B & 0 & \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\end{pmatrix}$$  \hspace{1cm} (B.1)$$

where each $2s \times 2s$ diagonal block $B$ has the following structure:

$$B = \begin{pmatrix}
B_D & B_{OD} \\
-B_{OD} & -B_D
\end{pmatrix}$$  \hspace{1cm} (B.2)$$

where

$$B_D = \begin{pmatrix}
0 & -\zeta & 0 & 0 & \cdots & 0 & 0 \\
\zeta & 0 & -\zeta & 0 & \cdots & 0 & 0 \\
0 & \zeta & 0 & -\zeta & \cdots & 0 & \cdots \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & \zeta & 0 & -\zeta & 0 \\
0 & 0 & \cdots & 0 & \zeta & 0 & -\zeta \\
0 & 0 & \cdots & 0 & 0 & \zeta & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \zeta
\end{pmatrix}$$

and

$$B_{OD} = \begin{pmatrix}
\zeta & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 0 \\
0 & \cdots & 0 & -\zeta
\end{pmatrix}$$

and $\zeta$ is such that the MH ratio (12) is always non-negative. The vorticity matrix is of size $(s^2 \times s^2)$, meaning that the number of diagonal blocks varies upon $s$:

- **if $s$ is even**: $\exists k \in \mathbb{N}$ s.t. $s = 2k \Rightarrow s^2 = 4k^2$ and each block $B$ is a square matrix of dimension $4k$, then there are exactly $k$ $B$-blocks in the vorticity matrix $\Gamma_\zeta$;

- **if $s$ is odd**: $\exists k \in \mathbb{N}$ s.t. $s = 2k + 1 \Rightarrow s^2 = (2k + 1)^2$ and each block $B$ is a square matrix of dimension $2(2k + 1)$, then as $\frac{(2k+1)^2}{2(2k+1)} = k + \frac{1}{2}$, $\Gamma_\zeta$ is made of $k$ $B$-blocks and the last terms of the diagonal are completed with zeros.
For instance, if $s = 3$ (resp. if $s = 4$), the vorticity matrix is given by $\Gamma^{(3)}_\zeta$ (resp. $\Gamma^{(4)}_\zeta$) as follows:

$$
\Gamma^{(3)}_\zeta = \begin{pmatrix}
0 & -\zeta & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
\zeta & 0 & -\zeta & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \zeta & 0 & 0 & 0 & -\zeta & 0 & 0 & 0 \\
-\zeta & 0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -\zeta & 0 & \zeta & 0 & 0 & 0 \\
0 & 0 & \zeta & 0 & -\zeta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},
$$

$$
\Gamma^{(4)}_\zeta = \begin{pmatrix}
B_4 & 0_8 \\
0_8 & B_4
\end{pmatrix},
$$

where

$$
B_4 = \begin{pmatrix}
0 & -\zeta & 0 & 0 & \zeta & 0 & 0 & 0 & 0 \\
\zeta & 0 & -\zeta & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \zeta & 0 & -\zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \zeta & 0 & 0 & 0 & 0 & -\zeta & 0 \\
-\zeta & 0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\zeta & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & 0 & -\zeta & 0 & \zeta & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & -\zeta & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \zeta & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

and $0_m$ stands for the zero-matrix of size $m \times m$. 
In all plots, the curves give the number of iterations $t$ needed for the algorithm to satisfy $\|\delta_{x_0} P^t - \pi\| \leq 10^{-3}$ ($x_0$ being the initialization of the algorithm), depending on the value of $\zeta$ and $\delta$. The two rows correspond to different initializations. In the four dials, the black dashed line gives the convergence time for MH. **Left-hand side – NRMH.** Convergence time for NRMH using $\Gamma_{\zeta}^+$ (red) or $\Gamma_{\zeta}^-$ (blue), according to the values of $\zeta$. The red (resp. blue) dashed line gives the minimum value of the convergence time for NRMH using $\Gamma_{\zeta}^+$ (resp. $\Gamma_{\zeta}^-$). **Right-hand side – NRMHAVd.** Convergence time for NRMHAVd initialized with $\Gamma_{\zeta}^+$ (red) or $\Gamma_{\zeta}^-$ (blue) according to the values of $\delta$, $\zeta$ being set close to its upper bound. The thin red (resp. blue) dashed line gives the minimum value of the convergence time of NRMHAVd initialized with $\Gamma_{\zeta}^+$ (resp. $\Gamma_{\zeta}^-$). The bold red (resp. blue) dashed line gives the value of the convergence time for NRMH parametrized with this particular $\zeta$ and using $\Gamma_{\zeta}^+$ (resp. $\Gamma_{\zeta}^-$).
Figure B.6: Illustration of the vorticity matrix specified by Algorithm 5 in the case $s = 4$. We note that the privileged direction of the flow is position specific.