ANNALES
DE L'INSTITUT
HENRI
POINCARÉ
PROBABILITÉS
ET STATISTIQUES

Metropolis–Hastings transition kernel couplings

John O'Leary[1,a] and Guanyang Wang[2,b]

[1]*Tudor Investment Corporation, Boston, USA,* [a]*john.oleary@tudor.com*
[2]*Department of Statistics, Rutgers University, Piscataway, USA,* [b]*guanyang.wang@rutgers.edu*

Abstract. Couplings play a central role in the analysis of Markov chain convergence and in the construction of novel Markov chain Monte Carlo estimators, diagnostics, and variance reduction techniques. The set of possible couplings is often intractable, frustrating the search for tight bounds and efficient estimators. To address this challenge for algorithms in the Metropolis–Hastings (MH) family, we establish a simple characterization of the set of MH transition kernel couplings. We then extend this result to describe the set of maximal couplings of the MH kernel, resolving an open question of O'Leary, Wang and Jacob (In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (2021) 1225–1233 PMLR). Our results represent an advance in understanding the MH transition kernel and a step forward for coupling this popular class of algorithms.

Résumé. Les techniques de couplage jouent un rôle central dans l'analyse de la convergence des chaînes de Markov et dans la construction de nouveaux estimateurs à partir de Chaînes de Markov par Monte Carlo, ainsi que de diagnostiques et techniques de réduction de la variance. Souvent, l'ensemble des couplages possibles n'est pas calculable et la recherche de bornes précises et d'estimateurs efficaces semble hors d'atteinte. Pour aborder un tel défi pour les algorithmes de la famille Metropolis–Hastings (MH), nous établissons une caractérisation simple de l'ensemble des couplages des noyaux de transition MH. Nous étendons ensuite ce résultat en décrivant l'ensemble des couplages maximaux des noyaux MH, résolvant une question ouverte de O'Leary, Wang and Jacob (In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* (2021) 1225–1233 PMLR). Nos résultats représentent un progrès dans la compréhension des noyaux de transition MH et des couplages pour cette classe d'algorithmes populaire.

1. Introduction

Couplings have played an important role in the analysis of Markov chain convergence since the early days of the field [19, 34]. Beyond their role as a proof technique, couplings have also been used as a basis for sampling [25,26,59,68], convergence diagnosis [11,42,43], variance reduction [31,57,66], and unbiased estimation [29,37–39,41,54,55]. Couplings that deliver smaller meeting times are associated with better results, and the design of such couplings is a central problem for theoretical and computational statisticians who use these methods.

Given a pair of Markov chains, the distribution of meeting times that can occur under any coupling is constrained by the total variation distance between the marginal distributions of the chains at each iteration [1,50]. Couplings that achieve the fastest meeting times allowed by this bound are said to be efficient or maximal and exist in some generality [30,32,67]. These couplings are rarely Markovian, whereas this is a practical requirement in many of the applications above. Efficient Markovian couplings are known in a few cases, but they remain elusive for discrete-time MCMC algorithms [8,12,17, 18,40,46,48]. For techniques such as the Metropolis–Hastings (MH) algorithm [52,56], it is difficult even to describe or parameterize the set of possible couplings, much less to identify an optimum within this set. Despite its practical importance, coupling design is often an unsystematic task that relies on qualitative insight, experience, and luck.

In this paper, we aim to address this challenge by deriving a simple characterization of the set of possible MH kernel couplings. We show that every such coupling can be represented as a coupled proposal step followed by a coupled acceptance step with a short list of properties. We also show the converse: that any coupled proposal and acceptance steps with these properties combine to yield a coupling of MH transition kernels. Despite its simplicity, this result requires a

clear understanding of the mechanics of MH coupling and has proved elusive since at least Johnson [43]. Yet it is of some theoretical and practical value, as described below.

On the theory side, our result addresses an important open question on the structure of MH kernel couplings, which is whether all such couplings have a 'two-step' representation. So far, two approaches to constructing MH kernel couplings have appeared in the literature: in the first (one-step) method, a coupling is defined directly in terms of the probability density or mass function of the marginal transition kernel [63,69,74,75]. In the other (two-step) method, one first defines a coupling at the MH proposal step followed by a second coupling at the acceptance step [13,41,43]. One-step couplings seem more general, but they can be difficult to analyze. In contrast, two-step couplings have proven amenable to methods like those of Bou-Rabee, Eberle and Zimmer [13], which yield explicit bounds on the contraction rate between chains. Our results show that all MH kernel couplings have a two-step representation, resolving the apparent tension between these two approaches.

On the practical side, our results make it easier to identify high-performance couplings of the MH transition kernel. First, they reframe the problem of transition kernel coupling design into the simpler tasks of selecting proposal and acceptance step couplings. Our results show that any coupling can be expressed in a two-step form, so there is no need to consider complex one-step couplings such as those in O'Leary [62] in order to achieve rapid convergence between chains. This is an additional benefit since two-step couplings are often easier to implement than one-step couplings. Finally, proposal and acceptance step couplings often have natural parameterizations, and in this case, our results support the use of numerical methods such as gradient descent to optimize the performance of MH kernel couplings.

With the above results in hand, we turn to maximal couplings of the MH transition kernel, which are sometimes called greedy couplings [2,35], one-step maximal couplings [36,45,70], or step-by-step maximal couplings [65] of the associated Markov chains. While these couplings do not typically achieve the fastest meeting times allowed by the coupling inequality, they often perform well and serve as a valuable reference point for further analysis. Building on the results described above, we characterize maximal couplings in terms of the properties of the associated proposal and acceptance couplings. We also resolve an open question of O'Leary, Wang and Jacob [63] on the structure of maximal transition kernel couplings and their relationship with maximal couplings of the proposal distributions.

We note that our results hold for simple MH algorithms as well as refinements such as Hamiltonian Monte Carlo [22, 58,60], the Metropolis-adjusted Langevin algorithm [72], and particle MCMC [4]. They also hold on both continuous and discrete state spaces, and for 'lazy' implementations where the proposal distribution is not absolutely continuous with respect to the base measure. Finally, they hold for methods like Barker's algorithm [9] where the acceptance rate function differs from the usual MH form. Thus our characterizations of kernel couplings and maximal kernel couplings apply to most Markov chains which involve a sequence of proposal and acceptance steps.

The rest of this paper is organized as follows. In Section 2, we describe our notation and setting, state our main result on transition kernel couplings, and set out a series of definitions and lemmas to prove it. In Section 3, we highlight a few important properties of maximal couplings, consider the relationship between maximal proposal and transition kernel couplings, and prove our main result on the structure of maximal couplings of the MH transition kernel. In Section 4, we analyze the kernel couplings of several widely-used MH algorithms such as the random-walk MH algorithm and the Metropolis-adjusted Langevin algorithm. In Section 5 we discuss our results and outline future directions. Finally, in the appendix, we illustrate our definitions and results with a series of simple examples.

2. Metropolis–Hastings kernel couplings

In this section, we show that every MH-like transition kernel coupling can be represented as a proposal coupling followed by an acceptance indicator coupling (Theorem 2.2). After establishing our setting and notation, we build up to this proof through several auxiliary results (Lemmas 2.4–2.10). The key step is Lemma 2.8, which shows how to map from an arbitrary MH-like transition kernel coupling to a proposal coupling together with a structure we call a 'coupled acceptance mechanism.' The representation of transition kernel couplings in terms of coupled proposal and acceptance steps follows from this result. Finally, Corollary 2.11 in Section 2.7 gives a simplified form of Theorem 2.2 under mild assumptions on the proposal kernel.

2.1. *Notation and setting*

Let $(\mathcal{Z}, \mathcal{G})$ be a measurable space. We say $\Theta : \mathcal{Z} \times \mathcal{G} \to [0, 1]$ is a Markov kernel if $\Theta(z, \cdot)$ is a probability measure on $(\mathcal{Z}, \mathcal{G})$ for each $z \in \mathcal{Z}$ and if $\Theta(\cdot, A) : \mathcal{Z} \to [0, 1]$ is a measurable function for each $A \in \mathcal{G}$. We say that Θ is a sub-probability kernel if it satisfies the same conditions but with $\Theta(z, \cdot)$ a sub-probability measure for each $z \in \mathcal{Z}$ rather than a probability measure. Finally, let 2^S denote the power set of a set S, let $\text{Bern}(\alpha)$ denote the Bernoulli distribution on $\{0, 1\}$ with $\mathbb{P}(\text{Bern}(\alpha) = 1) = \alpha$ for $\alpha \in [0, 1]$, and let $a \wedge b := \min(a, b)$ for any $a, b \in \mathbb{R}$.

Suppose that $(\mathcal{X}, \mathscr{F})$ is a Polish space with base measure λ. We take $\mathcal{X}$ to be our state space. Fix a Markov kernel $Q : \mathcal{X} \times \mathscr{F} \to [0, 1]$, which we call the proposal kernel, and a function $a : \mathcal{X} \times \mathcal{X} \to [0, 1]$, which we call the acceptance rate function. To be an acceptance rate function, we require that $a(x, \cdot)$ is $Q(x, \cdot)$-measurable and that $a(x, x) = 1$ for all $x \in \mathcal{X}$. The second condition simplifies many proofs and involves no loss of generality, as we discuss below. Given a current state $x \in \mathcal{X}$ and a measurable set $A \in \mathscr{F}$, we think of $Q(x, A)$ as the probability of proposing a move from x to some $x' \in A$, and we think of $a(x, x')$ as the probability of accepting such a proposal. We call the probability measure $B(x, x') := \mathrm{Bern}(a(x, x'))$ the acceptance indicator distribution associated with the acceptance rate function a, given a proposed move from x to x'.

In this paper, we consider transition kernels that generalize the MH algorithm according to the following definition:

Definition 2.1 (MH-like transition kernel). We say that a Markov kernel P is *generated by Q and a* if for all $x \in \mathcal{X}$, $x' \sim Q(x, \cdot)$ and $b_x \sim B(x, x') = \mathrm{Bern}(a(x, x'))$ imply $X := b_x x' + (1 - b_x)x \sim P(x, \cdot)$. We say P is *generated by Q and B* if P is generated by Q and a, and if B is the acceptance indicator distribution associated with a. Finally, we say that a Markov transition kernel P is *MH-like* if there exists a Markov kernel $Q : \mathcal{X} \times \mathscr{F} \to [0, 1]$ and an acceptance rate function $a : \mathcal{X} \times \mathcal{X} \to [0, 1]$ such that P is generated by Q and a.

Most MCMC algorithms that involve alternating proposal and acceptance steps are MH-like. The MH case arises when we fix a target distribution $\pi \ll \lambda$, assume $Q(x, \cdot) \ll \lambda$ for all $x \in \mathcal{X}$, and set $a(x, x') := 1 \wedge (\pi(x')q(x', x)/(\pi(x)q(x, x')))$, where $\pi(\cdot) := \mathrm{d}\pi/\mathrm{d}\lambda$ and $q(x, \cdot) := \mathrm{d}Q(x, \cdot)/\mathrm{d}\lambda$. Our analysis also holds for alternative forms of $a(x, x')$ and for proposal kernels that are not absolutely continuous with respect to λ. The requirement that $a(x, x) = 1$ for all $x \in \mathcal{X}$ involves no loss of generality since any transition kernel P that can be expressed in terms of some $\tilde{a}$ without this property can also be expressed in terms of another acceptance function a that has it. Thus our analysis applies to continuous and discrete state spaces, to lazy algorithms in which $Q(x, \{x\}) > 0$, and to the transition kernel of Barker's algorithm [9] where $a(x, x') := \pi(x')q(x', x)/(\pi(x)q(x, x') + \pi(x')q(x', x))$ for each $x \neq x'$. We refer the reader to Andrieu, Lee and Livingstone [5] for a general framework for thinking about MH-like transition kernels.

Next, let μ and ν be any finite measures on $(\mathcal{X}, \mathscr{F})$. A measure γ on $(\mathcal{X} \times \mathcal{X}, \mathscr{F} \otimes \mathscr{F})$ is called a coupling of μ and ν if $\gamma(A \times \mathcal{X}) = \mu(A)$ and $\gamma(\mathcal{X} \times A) = \nu(A)$ for all $A \in \mathscr{F}$. We write $\Gamma(\mu, \nu)$ for the set of couplings of μ and ν, sometimes called the Fréchet class of these measures [47]. When μ and ν are probability measures, the coupling inequality states that $\mathbb{P}(X = Y) \leq 1 - \|\mu - \nu\|_{\mathrm{TV}}$ for $(X, Y) \sim \gamma \in \Gamma(\mu, \nu)$. Here $\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \in \mathscr{F}} |\mu(A) - \nu(A)|$ is the total variation distance. See e.g. Lindvall [50, Chapter 1.2] or Levin, Peres and Wilmer [49, Chapter 4] for a further discussion of this important inequality. A coupling γ that achieves the coupling inequality bound is said to be maximal, and we write $\Gamma^{\mathrm{max}}(\mu, \nu)$ for the set of such couplings. We note that there are important similarities between maximal couplings and optimal transport couplings, whose goal is to minimize the expected distance between draws according to some metric. The need for exact meeting depends on the context. For most of the applications described in Section 1, the advantage of maximal couplings is that the pairs of chains can be kept together indefinitely once meeting occurs.

Suppose that Θ is a Markov kernel on $(\mathcal{Z}, \mathscr{G})$. Following Douc et al. [20, Chapter 19], we call $\bar{\Theta} : (\mathcal{Z} \times \mathcal{Z}) \times (\mathscr{G} \otimes \mathscr{G}) \to [0, 1]$ a kernel coupling based on Θ if $\bar{\Theta}$ is a Markov kernel on $(\mathcal{Z} \times \mathcal{Z}, \mathscr{G} \otimes \mathscr{G})$ and if $\bar{\Theta}((z_1, z_2), \cdot) \in \Gamma(\Theta(z_1, \cdot), \Theta(z_2, \cdot))$ for all $z_1, z_2 \in \mathcal{Z}$. We write $\Gamma(\Theta, \Theta)$ for the set of all such kernel couplings. Likewise, we say that $\bar{\Theta}$ is a maximal kernel coupling if each $\bar{\Theta}((z_1, z_2), \cdot)$ is a maximal coupling of $\Theta(z_1, \cdot)$ and $\Theta(z_2, \cdot)$, and we write $\Gamma^{\mathrm{max}}(\Theta, \Theta)$ for the set of these. It is important to remember that a maximal kernel coupling $\bar{P} \in \Gamma^{\mathrm{max}}(P, P)$ does not usually correspond to a maximal coupling of Markov chains in the sense of Aldous [1]. The former produces the highest probability of meeting at each individual step, while the latter produces the smallest meeting times allowed by the coupling inequality.

Suppose that an MH-like kernel P is generated by a proposal kernel Q and an acceptance rate function a. Then we call any $\bar{P} \in \Gamma(P, P)$ a transition kernel coupling and any $\bar{Q} \in \Gamma(Q, Q)$ a proposal kernel coupling. Let $\mathcal{B}$ be the power set of $\{0, 1\}^2$. We call $\bar{B} : \mathcal{X}^2 \times \mathcal{X}^2 \times \mathcal{B} \to [0, 1]$ an acceptance indicator coupling if $\bar{B}((x, y), (x', y'), \cdot)$ is a measure on $\{0, 1\}^2$ for all $x, y, x', y' \in \mathcal{X}$ and if $\bar{B}(\cdot, \cdot, S) : \mathcal{X}^2 \times \mathcal{X}^2 \to [0, 1]$ is measurable for all $S \in \mathcal{B}$. Abusing notation, we drop the third argument of $\bar{B}$ and write $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$ whenever a pair of random variables $(b_x, b_y) \in \{0, 1\}^2$ follows the acceptance indicator coupling $\bar{B}$. Thus we say that a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$ generate a kernel coupling $\bar{P}$ if for all $x, y \in \mathcal{X}$, $(x', y') \sim \bar{Q}((x, y), \cdot)$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$ imply that $(X, Y) := (b_x x' + (1 - b_x)x, b_y y' + (1 - b_y)y) \sim \bar{P}((x, y), \cdot)$. The first challenge before us is to show that every coupling $\bar{P} \in \Gamma(P, P)$ arises in this way. The second is to determine conditions on $\bar{Q}$ and $\bar{B}$ such that pairs (X, Y) defined as above follow a kernel coupling $\bar{P} \in \Gamma(P, P)$.

2.2. *Main result*

The literature on MH transition kernel couplings has so far considered individual examples of couplings or at most small classes of them. Johnson [43, Section 3.4] showed that one can start with a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and construct an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate an MH transition kernel coupling $\bar{P} \in \Gamma(P, P)$. $\bar{Q}$ is taken to be the γ-coupling of Lindvall [50] and $\bar{B}$ is obtained by setting $b_x = 1(U \leq a(x, x'))$ and $b_y = 1(U \leq a(y, y'))$ using the same $U \sim \text{Unif}$. Jacob, O'Leary and Atchadé [41] maintains this $\bar{B}$ while modifying $\bar{Q}$ to obtain faster convergence between chains, employing a reflection strategy that yields dependence between x' and y' even when $x' \neq y'$. Most recently, O'Leary, Wang and Jacob [63] defined a conditional $\bar{B}$ to obtain a maximal transition kernel coupling $\bar{P}$ from any maximal proposal kernel coupling $\bar{Q}$. Such couplings $\bar{B}$ accept proposals with $x' = y'$ at more than the MH rate in exchange for a lower acceptance rate when $x' \neq y'$, such that each chain continues to marginally evolve according to P.

In contrast with these previous studies, our aim in this paper is to fully characterize the set of MH transition kernel couplings $\Gamma(P, P)$ in terms of proposal and acceptance indicator couplings. In particular, we prove the following:

Theorem 2.2. *Let P be the MH-like transition kernel on $(\mathcal{X}, \mathcal{F})$ generated by a proposal kernel Q and an acceptance rate function a. A joint kernel $\bar{P} \in \Gamma(P, P)$ if and only if it is generated by $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ such that for any $x, y \in \mathcal{X}$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, we have*

1. *$\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for $Q(x, \cdot)$-almost all x', and*
2. *$\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all y'.*

This result shows that all MH-like transition kernel couplings $\bar{P}$ are 'natural' in the sense that they arise from coupled proposals $(x', y') \sim \bar{Q}((x, y), \cdot)$ that are accepted or rejected according to coupled indicators $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$. The $\bar{B}$ conditions say that from any current state pair (x, y), the x' and y' proposals are marginally accepted at the rate given by a, potentially with complicated joint behavior. Finally, Theorem 2.2 confirms that simple conditions on the joint proposal and acceptance steps suffice to yield a kernel coupling $\bar{P}$ in $\Gamma(P, P)$.

It is simple to show that the combination of a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$ with the properties above yields a kernel coupling $\bar{P} \in \Gamma(P, P)$. The converse requires a deeper understanding of the structure of transition kernel couplings. An example of the difficulty is as follows. A transition from the state pair (x, y) to a measurable rectangle $A_x \times A_y$ with $x \notin A_x$ and $y \notin A_y$ can only occur if points $x' \in A_x$ and $y' \in A_y$ are proposed and accepted. This makes it simple to work out the behavior of a hypothetical $\bar{Q}$ and $\bar{B}$ on these sets. However, transitions to $A_x \times \{y\}$, $\{x\} \times A_y$, and $\{x\} \times \{y\}$ can arise from the partial or full rejection of proposed moves to measurable subsets of $A_x \times \mathcal{X}$, $\mathcal{X} \times A_y$, and $\mathcal{X} \times \mathcal{X}$. It is challenging to relate all of these transition probabilities to joint proposal and acceptance measures in a consistent way.

Nevertheless, we prove Theorem 2.2 by constructing a mapping from an arbitrary kernel $\bar{P} \in \Gamma(P, P)$ to a joint distribution $\bar{Q}$ on proposals (x', y') and a coupling $\bar{B}$ of acceptance indicators (b_x, b_y), such these reproduce $\bar{P}$ and agree with the original proposal kernel Q and acceptance rate function a. To do so, we define a collection of joint measures Φ on (x', y') and (b_x, b_y) with certain decomposition properties. This Φ implies a proposal coupling $\bar{Q}$, and a Radon–Nikodym argument then yields an acceptance indicator coupling $\bar{B}$ with the desired properties.

2.3. *Coupled acceptance mechanisms*

To begin, we define a relationship that can exist between an MH-like transition kernel coupling and a proposal coupling:

Definition 2.3. We say that a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and a transition kernel coupling $\bar{P} \in \Gamma(P, P)$ are related by a *coupled acceptance mechanism* $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$ if each $\Phi_{ij}((x, y), \cdot)$ is a sub-probability on $(\mathcal{X} \times \mathcal{X}, \mathcal{F} \otimes \mathcal{F})$, and if for all $x, y \in \mathcal{X}$, $A_x, A_y \in \mathcal{F}$ we have

1. $\bar{Q}((x, y), A_x \times A_y) = (\Phi_{11} + \Phi_{10} + \Phi_{01} + \Phi_{00})((x, y), A_x \times A_y)$

2. $\bar{P}((x, y), A_x \times A_y) = \Phi_{11}((x, y), A_x \times A_y) + \Phi_{10}((x, y), A_x \times \mathcal{X})1(y \in A_y)$

$\quad\quad\quad\quad + \Phi_{01}((x, y), \mathcal{X} \times A_y)1(x \in A_x) + \Phi_{00}((x, y), \mathcal{X} \times \mathcal{X})1(x \in A_x)1(y \in A_y)$

3. $Q(x, \{x\}) = (\Phi_{11} + \Phi_{10})((x, y), \{x\} \times \mathcal{X})$ and $Q(y, \{y\}) = (\Phi_{11} + \Phi_{01})((x, y), \mathcal{X} \times \{y\})$.

We will see that the existence of a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$ is equivalent to the existence of an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. In Lemma 2.4 we show that if $(X, Y) \sim \bar{P}((x, y), \cdot)$ is
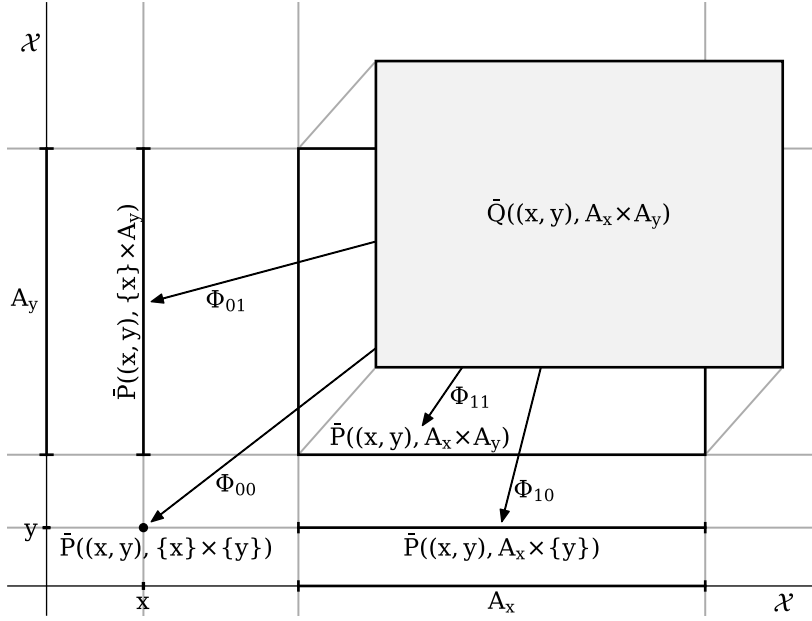
Fig. 1. Schematic diagram of a coupled acceptance mechanism Φ relating a proposal coupling $\bar{Q}$ and a transition kernel coupling $\bar{P}$. Here (x, y) represents the current state and $A_x \times A_y$ is a measurable rectangle in $\mathcal{X} \times \mathcal{X}$. $\bar{Q}((x, y), A_x \times A_y)$ gives the probability of a proposal $(x', y') \in A_x \times A_y$. The coupled acceptance mechanism $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$ distributes this probability into contributions to the probability $\bar{P}((x, y), \cdot)$ of transitions to the sets $A_x \times A_y$, $A_x \times \{y\}$, $\{x\} \times A_y$, and $\{x\} \times \{y\}$. The conditions of Definition 2.3 ensure that Φ agrees with both $\bar{Q}$ and $\bar{P}$.

generated by $(x', y') \sim \bar{Q}((x, y), \cdot)$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, then a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$ exists and can be defined by $\Phi_{ij}((x, y), \cdot) := \mathbb{P}((x', y') \in \cdot, b_x = i, b_y = j)$ for $i, j \in \{0, 1\}$. Conversely, in Lemma 2.9, we show that if a coupled acceptance mechanism Φ relates $\bar{Q}$ and $\bar{P}$, then there exists an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$.

Thus Definition 2.3 captures a set of relationships between a proposal coupling $\bar{Q}$ and a transition kernel coupling $\bar{P}$ required for the former to potentially generate the latter. We see that Φ divides the probability in $\bar{Q}((x, y), \cdot)$ according to scenarios in which one, both, or neither proposal is accepted, as illustrated in Figure 1. Condition 1 requires the total probability over these scenarios to sum to $\bar{Q}$ for all measurable subsets of $\mathcal{X} \times \mathcal{X}$. Condition 2 says that the resulting distribution over transitions must also agree with $\bar{P}$. Finally, Condition 3 addresses proposals with $x' = x$ or $y' = y$ and requires Φ to agree with the assumption that such proposals are always accepted. In Appendix A.1, we show that Condition 3 is independent of Conditions 1 and 2. We also show that more than one coupled acceptance mechanism Φ can relate the same pair $\bar{Q} \in \Gamma(Q, Q)$ and $\bar{P} \in \Gamma(P, P)$.

The crux of Theorem 2.2 is to prove that every coupling of MH-like transition kernels $\bar{P}$ arises from a proposal and acceptance coupling. Thus, we begin by looking for coupled acceptance mechanisms Φ that distribute the probability $\bar{Q}((x, y), A_x \times A_y)$ of a proposal from (x, y) to $(x', y') \in A_x \times A_y$ into contributions to the probability of a transition from (x, y) to $(X, Y) \in A_x \times A_y$, $A_x \times \{y\}$, $\{x\} \times A_y$, and $\{x\} \times \{y\}$ in a consistent way. The following result establishes the necessity of the conditions in Definition 2.3 for this task.

Lemma 2.4. *Let $\bar{P} \in \Gamma(P, P)$ be a coupling of MH-like transition kernels. If a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$ generate $\bar{P}$, then there exists a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$.*

Proof. Fix (x, y), let $(x', y') \sim \bar{Q}((x, y), \cdot)$, and let $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$. For $A \in \mathcal{F} \otimes \mathcal{F}$ and $i, j \in \{0, 1\}$, define $\Phi_{ij}((x, y), A) := \mathbb{P}((x', y') \in A, b_x = i, b_y = j \mid x, y)$. This Φ satisfies Condition 1, since for all $A_x, A_y \in \mathcal{F}$ we have $\bar{Q}((x, y), A_x \times A_y) = \mathbb{P}((x', y') \in A_x \times A_y | x, y)$. It also satisfies Condition 2, as

$$\bar{P}\big((x, y), A_x \times A_y\big) = \mathbb{P}(X \in A_x, Y \in A_y \mid x, y)$$

$$= \mathbb{P}\big(x' \in A_x, y' \in A_y, b_x = 1, b_y = 1 \mid x, y\big) + \mathbb{P}\big(x' \in A_x, y \in A_y, b_x = 1, b_y = 0 \mid x, y\big)$$

$$+ \mathbb{P}\big(x \in A_x, y' \in A_y, b_x = 0, b_y = 1 \mid x, y\big) + \mathbb{P}(x \in A_x, y \in A_y, b_x = 0, b_y = 0 \mid x, y)$$

$$= \Phi_{11}\big((x, y), A_x \times A_y\big) + \Phi_{10}\big((x, y), A_x \times \mathcal{X}\big)1(y \in A_y)$$
$$+ \Phi_{01}\big((x, y), \mathcal{X} \times A_y\big)1(x \in A_x) + \Phi_{00}\big((x, y), \mathcal{X} \times \mathcal{X}\big)1(x \in A_x)1(y \in A_y).$$

Finally, for Condition 3 we have

$$(\Phi_{11} + \Phi_{10})\big((x, y), \{x\} \times \mathcal{X}\big) = \mathbb{P}\big(x' = x, b_x = 1\big) = a(x, x)Q\big(x, \{x\}\big) = Q\big(x, \{x\}\big)$$
$$(\Phi_{11} + \Phi_{01})\big((x, y), \mathcal{X} \times \{y\}\big) = \mathbb{P}\big(y' = y, b_y = 1\big) = a(y, y)Q\big(y, \{y\}\big) = Q\big(y, \{y\}\big).$$

Thus Φ is a coupled acceptance mechanism relating $\bar{Q}$ and $\bar{P}$. $\qquad\square$

The $\bar{P}$ condition of Definition 2.3 takes a more intuitive, if less compact, form when expressed in terms of measurable sets $A_x, A_y \in \mathcal{F}$ with $x \notin A_x$ and $y \notin A_y$:

Lemma 2.5. *Let $\bar{P} \in \Gamma(P, P)$ and suppose $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$ are a collection of measures on $(\mathcal{X} \times \mathcal{X}, \mathcal{F} \otimes \mathcal{F})$. Then Condition 2 of Definition 2.3 holds if and only if for all $A_x, A_y \in \mathcal{F}$ with $x \notin A_x$ and $y \notin A_y$, we have*

$$1. \qquad \bar{P}\big((x, y), A_x \times A_y\big) = \Phi_{11}\big((x, y), A_x \times A_y\big)$$

$$2. \qquad \bar{P}\big((x, y), A_x \times \{y\}\big) = \Phi_{11}\big((x, y), A_x \times \{y\}\big) + \Phi_{10}\big((x, y), A_x \times \mathcal{X}\big)$$

$$3. \qquad \bar{P}\big((x, y), \{x\} \times A_y\big) = \Phi_{11}\big((x, y), \{x\} \times A_y\big) + \Phi_{01}\big((x, y), \mathcal{X} \times A_y\big)$$

$$4. \qquad \bar{P}\big((x, y), \{x\} \times \{y\}\big) = \Phi_{11}\big((x, y), \{x\} \times \{y\}\big) + \Phi_{10}\big((x, y), \{x\} \times \mathcal{X}\big)$$
$$+ \Phi_{01}\big((x, y), \mathcal{X} \times \{y\}\big) + \Phi_{00}\big((x, y), \mathcal{X} \times \mathcal{X}\big).$$

Proof. The equalities above follow by evaluating Condition 2 of Definition 2.3 at the four possible combinations of $\{x\}$ or $A_x \setminus \{x\}$ and $\{y\}$ or $A_y \setminus \{y\}$. For the converse, note that for any $A_x, A_y \in \mathcal{F}$ we have

$$\bar{P}\big((x, y), A_x \times A_y\big) = \bar{P}\big((x, y), \big(A_x \setminus \{x\}\big) \times \big(A_y \setminus \{y\}\big)\big) + \bar{P}\big((x, y), \big(A_x \setminus \{x\}\big) \times \{y\}\big)1(y \in A_y)$$
$$+ \bar{P}\big((x, y), \{x\} \times \big(A_y \setminus \{y\}\big)\big)1(x \in A_x) + \bar{P}\big((x, y), \{x\} \times \{y\}\big)1(x \in A_x)1(y \in A_y).$$

Replacing these terms with the equalities above yields Condition 2 of Definition 2.3. $\qquad\square$

Condition 1 of Lemma 2.5 reflects the fact that the only way for a joint chain to transition from (x, y) to $A_x \times A_y$ with $x \notin A_x$ and $y \notin A_y$ is to propose and accept a transition to some $(x', y') \in A_x \times A_y$. Condition 2 says that the probability of transitioning to a point in $A_x \times \{y\}$ amounts to the combined probability of accepting a proposal directly to such a point or by proposing a move to some $(x', y') \in A_x \times \mathcal{X}$, with x' accepted and y' rejected. The other two conditions have similar interpretations.

2.4. *Existence of a coupled acceptance mechanism*

To prove Theorem 2.2 we must show that any MH-like transition kernel coupling $\bar{P} \in \Gamma(P, P)$ arises from a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ with certain properties. In this subsection we show that for any $\bar{P} \in \Gamma(P, P)$, there exists a kernel coupling $\bar{Q} \in \Gamma(Q, Q)$ and a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$. In the next subsection we show that we can transform this Φ into an acceptance coupling $\bar{B}$. Finally, we use these results to prove our main theorem. We will make use of the following objects, which are defined for $x \in \mathcal{X}$ and $A \in \mathcal{F}$:

$$\alpha_0(x, A) := Q\big(x, A \setminus \{x\}\big) - P\big(x, A \setminus \{x\}\big) \qquad \alpha_1(x, A) := Q\big(x, A \cap \{x\}\big) + P\big(x, A \setminus \{x\}\big)$$

$$\beta(x) := \begin{cases} \dfrac{Q(x, \{x\})}{P(x, \{x\})} & \text{if } P\big(x, \{x\}\big) > 0 \\ 1 & \text{otherwise} \end{cases} \qquad \mu(x, A) := \begin{cases} \dfrac{\alpha_0(x, A)}{\alpha_0(x, \mathcal{X})} & \text{if } \alpha_0(x, \mathcal{X}) > 0 \\ 1(x \in A) & \text{otherwise.} \end{cases}$$

Note that these definitions depend only on the marginal kernels Q and P. Each of the above has a simple interpretation in terms of the underlying chains:

Lemma 2.6. *Let $x \in \mathcal{X}$, $A \in \mathscr{F}$, $x' \sim Q(x, \cdot)$, $b_x \sim B(x, x') = \mathrm{Bern}(a(x, x'))$, and $X := b_x x' + (1 - b_x) x$, so that by definition $X \sim P(x, \cdot)$. Then $\alpha_i(x, A) = \mathbb{P}(x' \in A, b_x = i|x)$ for $i \in \{0, 1\}$, $\beta(x) = \mathbb{P}(b_x = 1|x, X = x)$ when $P(x, \{x\}) > 0$, and $\mu(x, A) = \mathbb{P}(x' \in A|b_x = 0, x)$ when $\alpha_0(x, \mathcal{X}) > 0$.*

Proof. By assumption, the acceptance rate function has $a(x, x) = 1$ for all x. Thus $\mathbb{P}(x' \in A \cap \{x\}, b_x = 0|x) = 0$, and

$$\alpha_0(x, A) = Q(x, A \setminus \{x\}) - P(x, A \setminus \{x\}) = \mathbb{P}(x' \in A \setminus \{x\}|x) - \mathbb{P}(X \in A \setminus \{x\}|x)$$

$$= \mathbb{P}(x' \in A|x) - \mathbb{P}(x' \in A, b_x = 1|x) = \mathbb{P}(x' \in A, b_x = 0|x).$$

Similarly,

$$\alpha_1(x, A) = Q(x, A \cap \{x\}) + P(x, A \setminus \{x\}) = \mathbb{P}(x' \in A \cap \{x\}|x) + \mathbb{P}(X \in A \setminus \{x\}|x)$$

$$= \mathbb{P}(x' \in A \cap \{x\}, b_x = 1|x) + \mathbb{P}(x' \in A \setminus \{x\}, b_x = 1|x) = \mathbb{P}(x' \in A, b_x = 1|x).$$

Suppose $P(x, \{x\}) > 0$. Then since $a(x, x) = \mathbb{P}(b_x = 1|x, x' = x) = 1$ we have

$$\beta(x) = \frac{Q(x, \{x\})}{P(x, \{x\})} = \frac{\mathbb{P}(b_x = 1, x' = x|x)}{\mathbb{P}(X = x|x)} = \frac{\mathbb{P}(b_x = 1, X = x|x)}{\mathbb{P}(X = x|x)} = \mathbb{P}(b_x = 1|x, X = x).$$

Finally, say $\alpha_0(x, \mathcal{X}) > 0$. Then

$$\mu(x, A) = \frac{\alpha_0(x, A)}{\alpha_0(x, \mathcal{X})} = \frac{\mathbb{P}(x' \in A, b_x = 0|x)}{\mathbb{P}(b_x = 0|x)} = \mathbb{P}(x' \in A|b_x = 0, x).$$

Thus, each of $\alpha_0, \alpha_1, \beta$, and μ has a simple meaning in terms of the joint probability of x' and b_x induced by the proposal kernel Q and the acceptance indicator distribution B. $\square$

We will also need the following properties:

Lemma 2.7. *Let $x \in \mathcal{X}$ and $A \in \mathscr{F}$. Then $Q(x, A \cap \{x\}) = P(x, A \cap \{x\})\beta(x)$, $Q(x, A) = \alpha_1(x, A) + \mu(x, A)\alpha_0(x, \mathcal{X})$, $\beta(x) \in [0, 1]$, and $\mu(x, \cdot)$ is a measure on $(\mathcal{X}, \mathscr{F})$,*

Proof. When $P(x, \{x\}) > 0$, $\beta(x) \in [0, 1]$ by Lemma 2.6. Otherwise $\beta(x) = 1$, so in either case $\beta(x) \in [0, 1]$. Also by Lemma 2.6, we observe that $\mu(x, \cdot)$ is the indicator of a measurable set or a well-defined conditional probability, and so in either case it defines a measure on $(\mathcal{X}, \mathscr{F})$.

For the first Q equality, if $x \notin A$, then the $Q(x, A \cap \{x\}) = 0 = P(x, A \cap \{x\})\beta(x)$. If $x \in A$ then we want to show $Q(x, \{x\}) = P(x, \{x\})\beta(x)$. This holds by the definition of $\beta(x)$ when $P(x, \{x\}) > 0$. Since $a(x, x) = 1$ for all $x \in \mathcal{X}$, $P(x, \{x\}) = 0$ implies $Q(x, \{x\}) = 0$, so the result also holds in this case.

When $\alpha_0(x, \mathcal{X}) > 0$, the second Q equality holds by the definitions of μ, α_0, and α_1. This leaves the $\alpha_0(x, \mathcal{X}) = 0$ case. In general, $\alpha_0(x, \mathcal{X}) = Q(x, \{x\}^c) - P(x, \{x\}^c) = \int (1 - a(x, x'))Q(x, dx')$, since $a(x, x) = 1$ for all $x \in \mathcal{X}$. Thus $\alpha_0(x, \mathcal{X}) = 0$ implies $a(x, x') = 1$ for $Q(x, \cdot)$-almost all x', so $\alpha_1(x, A) = Q(x, A \cap \{x\}) + \int_{A \setminus \{x\}} a(x, x')Q(x, dx') = Q(x, A)$. Thus the second equality also holds when $\alpha_0(x, \mathcal{X}) = 0$. $\square$

We now describe the main lemma used in our proof of Theorem 2.2.

Lemma 2.8. *For any coupling $\bar{P} \in \Gamma(P, P)$ of MH-like transition kernels, there exists a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$.*

Proof. The proof proceeds in three steps. First we define a collection of sub-probability kernels $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$ in terms of $\bar{P}$. Then we use Φ to define a kernel coupling $\bar{Q} \in \Gamma(Q, Q)$. Finally, we show that Φ satisfies the conditions of Definition 2.3, making it a coupled acceptance mechanism relating $\bar{Q}$ and $\bar{P}$.

We begin with an explicit formula for Φ_{11} in terms of $\bar{P}$. For $x, y \in \mathcal{X}$ and $A \in \mathscr{F} \otimes \mathscr{F}$, let

$$\Phi_{11}((x, y), A) := \bar{P}((x, y), A \cap (\{x\}^c \times \{y\}^c)) + \bar{P}((x, y), A \cap (\{x\} \times \{y\}^c))\beta(x)$$

$$+ \bar{P}((x, y), A \cap (\{x\}^c \times \{y\}))\beta(y) + \bar{P}((x, y), A \cap (\{x\} \times \{y\}))\beta(x)\beta(y).$$

Here Φ_{11} is a sub-probability kernel, since $\beta(\cdot) \in [0, 1]$ by Lemma 2.7. We use a product measure construction to define the other three components of Φ. For $x, y \in \mathcal{X}$ and $A_x, A_y \in \mathcal{F}$, let

$$\Psi_{10}((x, y), A_x) := \alpha_1(x, A_x) - \Phi_{11}((x, y), A_x \times \mathcal{X}) \quad \text{and} \quad \Psi_{01}((x, y), A_y) := \alpha_1(y, A_y) - \Phi_{11}((x, y), \mathcal{X} \times A_y).$$

We claim that $\Psi_{10}((x, y), \cdot)$ and $\Psi_{01}((x, y), \cdot)$ are sub-probabilities on $(\mathcal{X}, \mathcal{F})$. Indeed, by Lemma 2.7 and the fact that $\bar{P} \in \Gamma(P, P)$, we have

$$\alpha_1(x, A_x) = P(x, A_x \setminus \{x\}) + P(x, A_x \cap \{x\})\beta(x)$$
$$= \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}^c) + \bar{P}((x, y), (A_x \cap \{x\}) \times \{y\}^c)\beta(x)$$
$$+ \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}) + \bar{P}((x, y), (A_x \cap \{x\}) \times \{y\})\beta(x).$$

Plugging $A_x \times \mathcal{X}$ into the definition of Φ_{11} yields analogous terms:

$$\Phi_{11}((x, y), A_x \times \mathcal{X}) = \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}^c) + \bar{P}((x, y), (A_x \cap \{x\}) \times \{y\}^c)\beta(x)$$
$$+ \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\})\beta(y) + \bar{P}((x, y), (A_x \cap \{x\}) \times \{y\})\beta(x)\beta(y).$$

Thus $\Psi_{10}((x, y), A_x) = (\bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}) + \bar{P}((x, y), (A_x \cap \{x\}) \times \{y\})\beta(x))(1 - \beta(y))$, and similarly for $\Psi_{01}((x, y), A_y)$. It follows that $\Psi_{10}((x, y), \cdot)$ and $\Psi_{01}((x, y), \cdot)$ are sub-probabilities. We also set

$$\Psi_{00}(x, y) := 1 - \alpha_1(x, \mathcal{X}) - \alpha_1(y, \mathcal{X}) + \Phi_{11}((x, y), \mathcal{X} \times \mathcal{X}).$$

Algebraic manipulations show that $\Psi_{00}(x, y) = \bar{P}((x, y), \{x\} \times \{y\})(1 - \beta(x))(1 - \beta(y)) \in [0, 1]$.

With these results in hand, we define $\Phi_{10}((x, y), \cdot)$, $\Phi_{01}((x, y), \cdot)$, and $\Phi_{00}((x, y), \cdot)$ as the product measures extending the following equalities to $(\mathcal{X} \times \mathcal{X}, \mathcal{F} \otimes \mathcal{F})$:

$$\Phi_{10}((x, y), A_x \times A_y) = \Psi_{10}((x, y), A_x)\mu(y, A_y), \qquad \Phi_{01}((x, y), A_x \times A_y) = \mu(x, A_x)\Psi_{01}((x, y), A_y),$$
$$\Phi_{00}((x, y), A_x \times A_y) = \mu(x, A_x)\mu(y, A_y)\Psi_{00}(x, y).$$

Thus each component of $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$ is a sub-probability kernel.

Next, we set $\bar{Q}((x, y), A) := (\Phi_{11} + \Phi_{10} + \Phi_{01} + \Phi_{00})((x, y), A)$ for all $A \in \mathcal{F} \otimes \mathcal{F}$. To see that $\bar{Q} \in \Gamma(Q, Q)$, note that for any $x, y \in \mathcal{X}$ and $A_x \in \mathcal{F}$ we have

$$\bar{Q}((x, y), A_x \times \mathcal{X}) = (\Phi_{11} + \Phi_{10} + \Phi_{01} + \Phi_{00})((x, y), A_x \times \mathcal{X})$$
$$= (\Phi_{11} + \Phi_{10})(A_x \times \mathcal{X}) + \mu(x, A_x)(\Phi_{01} + \Phi_{00})(\mathcal{X} \times \mathcal{X})$$
$$= \alpha_1(x, A_x) + \mu(x, A_x)\alpha_0(x, \mathcal{X}) = Q(x, A_x).$$

The second-to-last equality follows from the definitions of Ψ_{10}, Ψ_{01} and Ψ_{00}, while the last equality is due to Lemma 2.7. A similar argument shows that $\bar{Q}((x, y), \mathcal{X} \times A_y) = Q(y, A_y)$ for any $x, y \in \mathcal{X}$ and $A_y \in \mathcal{F}$. Thus $\bar{Q} \in \Gamma(Q, Q)$.

Finally, we show that Φ satisfies the conditions of Definition 2.3. The $\bar{Q}$ condition is automatically satisfied given the definitions above. For the $\bar{P}$ condition, we check the four cases described in Lemma 2.5. For the $(A_x \setminus \{x\}) \times (A_y \setminus \{y\})$ case, we have $\Phi_{11}((x, y), (A_x \setminus \{x\}) \times (A_y \setminus \{y\})) = \bar{P}((x, y), (A_x \setminus \{x\}) \times (A_y \setminus \{y\}))$. For the $(A_x \setminus \{x\}) \times \{y\}$ case,

$$\Phi_{11}((x, y), (A_x \setminus \{x\}) \times \{y\}) + \Phi_{10}((x, y), (A_x \setminus \{x\}) \times \mathcal{X})$$
$$= P(x, A_x \setminus \{x\}) - \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}^c) = \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}).$$

Here we have used the definitions of Φ_{11} and Φ_{10} and the fact that $\alpha_1(x, A \setminus \{x\}) = P(x, A \setminus \{x\})$ for the first equality. The second equality follows from the definition of $\bar{P} \in \Gamma(P, P)$. By a similar argument

$$\Phi_{11}((x, y), \{x\} \times (A_y \setminus \{y\})) + \Phi_{10}((x, y), \mathcal{X} \times (A_y \setminus \{y\})) = \bar{P}((x, y), \{x\} \times (A_y \setminus \{y\})).$$

For the $\{x\} \times \{y\}$ case we have

$$\Phi_{11}((x, y), \{x\} \times \{y\}) + \Phi_{10}((x, y), \{x\} \times \mathcal{X}) + \Phi_{01}((x, y), \mathcal{X} \times \{y\}) + \Phi_{00}((x, y), \mathcal{X} \times \mathcal{X})$$
$$= \Phi_{11}((x, y), \{x\} \times \{y\}) + \Psi_{10}((x, y), \{x\}) + \Psi_{01}((x, y), \{y\}) + \Psi_{00}(x, y)$$
$$= 1 + \Phi_{11}((x, y), \{x\}^c \times \{y\}^c) - \alpha_1(x, \{x\}^c) - \alpha_1(y, \{y\}^c) = \bar{P}((x, y), \{x\} \times \{y\}).$$

For the last condition of Definition 2.3, note that by the definitions of Φ and α_1,

$$(\Phi_{11} + \Phi_{10})\big((x, y), \{x\} \times \mathcal{X}\big) = \alpha_1(x, \{x\}) = Q(x, \{x\})$$

$$(\Phi_{11} + \Phi_{01})\big((x, y), \mathcal{X} \times \{y\}\big) = \alpha_1(y, \{y\}) = Q(y, \{y\}).$$

We conclude that Φ is a coupled acceptance mechanism relating $\bar{P}$ and $\bar{Q}$. $\qquad\square$

2.5. *Existence of an acceptance indicator coupling*

Next, we show that if we have a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$, then there exists an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. In the following, we write Δ^{n-1} for the set of multinomial distributions on a set with n elements.

Lemma 2.9. *Say $\bar{P} \in \Gamma(P, P)$, $\bar{Q} \in \Gamma(Q, Q)$, and Φ is a coupled acceptance mechanism relating $\bar{P}$ and $\bar{Q}$. Then there exist $\bar{Q}((x, y), \cdot)$-measurable functions $\phi_{ij}((x, y), \cdot)$ for $i, j \in \{0, 1\}$ with $\phi = (\phi_{11}, \phi_{10}, \phi_{01}, \phi_{00})((x, y), (x', y')) \in \Delta^3$ for $\bar{Q}((x, y), \cdot)$-almost all (x', y'). If we define an acceptance indicator coupling $\bar{B}$ so that $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$ implies $\mathbb{P}(b_x = i, b_y = j | x, y, x', y') = \phi_{ij}((x, y), (x', y'))$, then $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$.*

Proof. For each $i, j \in \{0, 1\}$, we have $\Phi_{ij}((x, y), \cdot) \ll \bar{Q}((x, y), \cdot)$ by the $\bar{Q}$ condition of Definition 2.3. Thus we can form Radon–Nikodym derivatives $\phi_{ij}((x, y), \cdot) := d\Phi_{ij}((x, y), \cdot)/d\bar{Q}((x, y), \cdot)$. The RN derivative is linear and $d\bar{Q}((x, y), \cdot)/d\bar{Q}((x, y), \cdot) = 1$, so for $\bar{Q}((x, y), \cdot)$-almost all (x', y') we have $\sum_{i, j \in \{0, 1\}} \phi_{ij}((x, y), (x', y')) = 1$. Since $\phi_{ij} \geq 0$, this implies that each $\phi_{ij} \leq 1$. Thus $\phi((x, y), (x', y')) \in \Delta^3$ for $\bar{Q}((x, y), \cdot)$-almost all (x', y').

For points (x', y') where the above holds, let $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$ be the multinomial random variable on $\{0, 1\}^2$ with $\mathbb{P}(b_x = i, b_y = j | x, y, x', y') = \phi_{ij}((x, y), (x', y'))$ for $i, j \in \{0, 1\}$. Suppose that $(x', y') \sim \bar{Q}((x, y), \cdot)$ and let $X := x' b_x + x(1 - b_x)$ and $Y := y' b_y + y(1 - b_y)$. Then for any $A_y, A_y \in \mathscr{F}$ we have

$$\mathbb{P}\big((X, Y) \in A_x \times A_y, b_x = 1, b_y = 1 | x, y\big) = \Phi_{11}\big((x, y), A_x \times A_y\big)$$

$$\mathbb{P}\big((X, Y) \in A_x \times A_y, b_x = 1, b_y = 0 | x, y\big) = \Phi_{10}\big((x, y), A_x \times \mathcal{X}\big) 1(y \in A_y)$$

$$\mathbb{P}\big((X, Y) \in A_x \times A_y, b_x = 0, b_y = 1 | x, y\big) = \Phi_{01}\big((x, y), \mathcal{X} \times A_y\big) 1(x \in A_x)$$

$$\mathbb{P}\big((X, Y) \in A_x \times A_y, b_x = 0, b_y = 0 | x, y\big) = \Phi_{00}\big((x, y), \mathcal{X} \times \mathcal{X}\big) 1(x \in A_x) 1(y \in A_y).$$

It follows from these expressions together with Definition 2.3 that $\mathbb{P}((X, Y) \in A_x \times A_y) = \bar{P}((x, y), A_x \times A_y)$ on all measurable rectangles $A_x \times A_y$, and hence $\mathbb{P}((X, Y) \in A) = \bar{P}((x, y), A)$ for all $A \in \mathscr{F} \otimes \mathscr{F}$. We conclude that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. $\qquad\square$

Next, we show that when a joint proposal kernel $\bar{Q}$ and a joint transition kernel $\bar{P}$ are related by a coupled acceptance mechanism Φ, then the acceptance indicator coupling derived above satisfies the conditions of Theorem 2.2.

Lemma 2.10. *Let Φ be a coupled acceptance mechanism relating $\bar{P} \in \Gamma(P, P)$ and $\bar{Q} \in \Gamma(Q, Q)$, and let ϕ and $\bar{B}$ be as in the proof of Lemma 2.9. If $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, then*

1. *$\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for $Q(x, \cdot)$-almost all x', and*
2. *$\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all y'.*

Proof. Let $a_x := \phi_{11} + \phi_{10}$. Then $\mathbb{P}(b_x = 1 | x, y, x') = \mathbb{E}_{\bar{Q}((x, y), \cdot)}[a_x((x, y), (x', y')) | x, y, x']$, and $a_x((x, y), \cdot)$ is $\bar{Q}((x, y), \cdot)$-measurable. For all $A \in \mathscr{F}$, we have

$$\int 1\big(x' \in A\big) \mathbb{P}\big(b_x = 1 | x, y, x'\big) \bar{Q}\big((x, y), (dx', dy')\big)$$

$$= \int 1\big(x' \in A\big) a_x\big((x, y), (x', y')\big) \bar{Q}\big((x, y), (dx', dy')\big)$$

$$= (\Phi_{11} + \Phi_{10})\big((x, y), A \times \mathcal{X}\big) = Q\big(x, \{x\} \cap A\big) + P\big(x, A \setminus \{x\}\big) = \int_A a\big(x, x'\big) Q\big(x, dx'\big)$$

$$= \int 1\big(x' \in A\big) a\big(x, x'\big) \bar{Q}\big((x, y), (dx', dy')\big).$$

The first equality follows from the defining property of conditional expectations. Condition 2 of Definition 2.3 implies $(\Phi_{11} + \Phi_{10})((x, y), (A \setminus \{x\}) \times \mathcal{X}) = \bar{P}((x, y), (A \setminus \{x\}) \times \mathcal{X})$, while Condition 3 implies that $Q(x, A \cap \{x\}) = (\Phi_{11} + \Phi_{10})((x, y), (A \cap \{x\}) \times \mathcal{X})$. These combine to yield the third equality, above.

By the essential uniqueness of the Radon–Nikodym derivative, $\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for all x' in a measurable set $\tilde{A} \in \mathcal{F}$ with $\bar{Q}((x, y), \tilde{A} \times \mathcal{X}) = Q(x, \tilde{A}) = 1$. Thus this equality holds for $Q(x, \cdot)$-almost all $x' \in \mathcal{X}$. A similar argument shows that $\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all $y' \in \mathcal{X}$. $\qquad\square$

2.6. *Main result*

Having established the lemmas above, we can now prove the main result of this section.

Proof of Theorem 2.2. For the 'if' case, assume $\bar{Q} \in \Gamma(Q, Q)$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, an acceptance indicator coupling with $\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for $Q(x, \cdot)$-almost all x' and $\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all y'. Let $\bar{P}((x, y), \cdot)$ be the law of $(X, Y) := (b_x x' + (1 - b_x)x, b_y y' + (1 - b_y)y)$. Then for any $A \in \mathcal{F}$,

$$\bar{P}((x, y), A \times \mathcal{X}) = \mathbb{P}(X \in A | x, y) = \mathbb{P}(x' \in A, b_x = 1 | x, y) + \mathbb{P}(x \in A, b_x = 0 | x, y)$$

$$= \int 1(x' \in A)a(x, x')Q(x, dx') + 1(x \in A)\int (1 - a(x, x'))Q(x, dx') = P(x, A).$$

Similarly, $\bar{P}((x, y), \mathcal{X} \times A) = P(y, A)$ for any $A \in \mathcal{F}$. Thus $\bar{P} \in \Gamma(P, P)$.

For the 'only if' case, take any $\bar{P} \in \Gamma(P, P)$. By Lemma 2.8, there exists a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and a coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$. Then by Lemma 2.9, there exists an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. Finally, by Lemma 2.10, this $\bar{B}$ will be such that if $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, then $\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for $Q(x, \cdot)$-almost all x' and $\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all y'. Thus we conclude that $\bar{P}$ is generated by a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$ with the desired properties. $\qquad\square$

2.7. *Simplified characterization of couplings*

To prove the 'only if' part of Theorem 2.2 we constructed a map from an arbitrary coupling $\bar{P} \in \Gamma(P, P)$ to a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. Lemmas 2.6–2.10 suggest a potentially complicated relationship between $\bar{P}$, $\bar{Q}$, and $\bar{B}$, at least when P is allowed to be any MH-like transition kernel. As we show below in Corollary 2.11, this relationship takes a more intuitive form if we require $Q(x, \cdot)$ to be absolutely continuous with respect to the base measure and require $Q(x, \{x\}) = 0$ for all $x \in \mathcal{F}$. These assumptions often hold, for instance when the state space is continuous and λ is non-atomic, when Q represents a non-lazy random walk on a discrete state space, and in many other cases.

We need to make a few definitions and analytical observations before stating the main result of this section. First, for all $A_x \in \mathcal{F}$ let $\bar{P}_y(x, A_x) := \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\})$, so that $\bar{P}_y(x, \cdot)$ is a sub-probability on $(\mathcal{X}, \mathcal{F})$. Then for all $x, y \in \mathcal{X}$, $Q(x, \cdot) \ll \lambda$ implies $\bar{P}_y(x, \cdot) \ll \lambda$, since

$$\bar{P}_y(x, A_x) = \bar{P}((x, y), (A_x \setminus \{x\}) \times \{y\}) \leq \bar{P}((x, y), (A_x \setminus \{x\}) \times \mathcal{X})$$

$$= P(x, A_x \setminus \{x\}) = \int_{A_x \setminus \{x\}} a(x, x')q(x, x')\lambda(dx').$$

Here $q(x, \cdot) = dQ(x, \cdot)/d\lambda$. Therefore $\bar{P}_y(x, \cdot)$ will have density $\bar{p}_y(x, \cdot)$ with respect to λ by the Radon–Nikodym theorem. Similarly, $\bar{P}_x(y, A_y) := \bar{P}((x, y), \{x\} \times (A_y \setminus \{y\}))$ will have density $\bar{p}_x(y, \cdot)$ with respect to λ.

Let $r(x) := P(x, \{x\})$, the probability of a transition from x to itself under P. From the definitions in Section 2.4, we have $\beta(x) = 1(r(x) = 0)$ and

$$\mu(x, A_x) := \begin{cases} \dfrac{Q(x, A_x) - P(x, A_x \setminus \{x\})}{r(x)} & \text{if } r(x) > 0 \\ 1(x \in A_x) & \text{if } r(x) = 0. \end{cases}$$

It follows that $\mu(x, A_x)(1 - \beta(x)) = (Q(x, A_x) - P(x, A_x \setminus \{x\}))/r(x)$ if $r(x) > 0$, and 0 otherwise. Thus $\mu(x, \cdot)(1 - \beta(x))$ has density $m(x, x') := q(x, x')(1 - a(x, x'))/r(x)$ when $r(x) > 0$, and 0 otherwise.

Next, we separate $\bar{P}$ into a part that is absolutely continuous with respect to $\lambda \times \lambda$ and another part that is singular to it. By the Lebesgue decomposition theorem (e.g. [23, Chapter 5.5]) for each $x, y \in \mathcal{X}$ we have $\bar{P}((x, y), \cdot) = \bar{P}_{\ll}((x, y), \cdot) + \bar{P}_{\perp}((x, y), \cdot)$ with $\bar{P}_{\ll}((x, y), \cdot) \ll \lambda \times \lambda$ and $\bar{P}_{\perp}((x, y), \cdot) \perp \lambda \times \lambda$. We then define $p((x, y), \cdot)$ to be the density of $\bar{P}_{\ll}((x, y), \cdot)$ with respect to $\lambda \times \lambda$. Finally, let $\bar{r}(x, y) := \bar{P}((x, y), \{x\} \times \{y\})$, the probability of a transition from (x, y) to itself under $\bar{P}$. With these definitions, we can now give an explicit characterization of the couplings $\bar{P} \in \Gamma(P, P)$ under mild conditions on the proposal distribution.

Corollary 2.11. *Let P be the MH-like transition kernel on $(\mathcal{X}, \mathscr{F})$ generated by a proposal kernel Q and an acceptance rate function a, and suppose that $Q(x, \cdot) \ll \lambda$ with $Q(x, \{x\}) = 0$ for all $x \in \mathcal{X}$. If a joint kernel $\bar{P} \in \Gamma(P, P)$, then it is generated by a proposal coupling $\bar{Q} = \bar{Q}_{\ll} + \bar{Q}_{\perp} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ with the following properties for all $x, y \in \mathcal{X}$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$:*

1. $\bar{Q}_{\perp}((x, y), A) = \bar{P}_{\perp}((x, y), A \cap (\{x\}^c \times \{y\}^c))$ *for* $A \in \mathscr{F}$,
2. $\bar{Q}_{\ll}((x, y), \cdot)$ *has the following density with respect to* $\lambda \times \lambda$:

$$q\big((x, y), (x', y')\big) = p\big((x, y), (x', y')\big) + \bar{p}_y(x, x')m(y, y') + m(x, x')\bar{p}_x(y, y') + m(x, x')m(y, y')\bar{r}(x, y),$$

3. $b_x = b_y = 1$ *almost surely when (x', y') is in the support of* $\bar{Q}_{\perp}((x, y), \cdot)$,

and for $\bar{Q}_{\ll}((x, y), \cdot)$-almost all (x', y'),

4. $\mathbb{P}(b_x = 1 | x, y, x', y')q((x, y), (x', y')) = p((x, y), (x', y')) + \bar{p}_y(x, x')m(y, y')$,
5. $\mathbb{P}(b_y = 1 | x, y, x', y')q((x, y), (x', y')) = p((x, y), (x', y')) + \bar{p}_x(y, y')m(x, x')$.

Proof. Let $B_{xy} := \{x\}^c \times \{y\}^c$. We may focus attention on the behavior of $\bar{Q}((x, y), \cdot)$ on B_{xy}, since $\bar{Q}((x, y), B_{xy}) = 1$. This follows since for any $x, y \in \mathcal{X}$, $B_{xy}^c = (\{x\} \times \mathcal{X}) \cup (\mathcal{X} \times \{y\})$, $\bar{Q}((x, y), \{x\} \times \mathcal{X}) = Q(x, \{x\}) = 0$, and $\bar{Q}((x, y), \mathcal{X} \times \{y\}) = Q(y, \{y\}) = 0$.

When $\bar{P} \in \Gamma(P, P)$, Lemma 2.8 tells us that there exists a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and a coupled acceptance mechanism Φ relating $\bar{P}$ and $\bar{Q}$. In particular, for $x, y \in \mathcal{X}$ and $A_x, A_y \in \mathscr{F}$ with $x \notin A_x$ and $y \notin A_y$, the proof of that Lemma shows that we can use the following measures:

$$\bar{Q}\big((x, y), A_x \times A_y\big) = (\Phi_{11} + \Phi_{10} + \Phi_{01} + \Phi_{00})\big((x, y), A_x \times A_y\big)$$

$$\Phi_{11}\big((x, y), A_x \times A_y\big) = \bar{P}\big((x, y), A_x \times A_y\big)$$

$$\Phi_{10}\big((x, y), A_x \times A_y\big) = \bar{P}\big((x, y), A_x \times \{y\}\big)\big(1 - \beta(y)\big)\mu(y, A_y)$$

$$\Phi_{01}\big((x, y), A_x \times A_y\big) = \big(1 - \beta(x)\big)\mu(x, A_x)\bar{P}\big((x, y), \{x\} \times A_y\big)$$

$$\Phi_{00}\big((x, y), A_x \times A_y\big) = \big(1 - \beta(x)\big)\mu(x, A_x)\big(1 - \beta(y)\big)\mu(y, A_y)\bar{P}\big((x, y), \{x\} \times \{y\}\big).$$

From the Lebesgue decomposition of $\bar{P}$, we have

$$\bar{P}\big((x, y), A_x \times A_y\big) = \bar{P}_{\perp}\big((x, y), A_x \times A_y\big) + \int_{A_x \times A_y} p\big((x, y), (x', y')\big)\mathrm{d}(x', y').$$

The definitions above imply that $\Phi_{10}((x, y), \cdot)$ has density $p_y(x, x')m(y, y')$, $\Phi_{01}((x, y), \cdot)$ has density $p_x(y, y')m(x, x')$, and $\Phi_{00}((x, y), \cdot)$ has density $\bar{r}(x, y)m(x, x')m(y, y')$. Thus $\bar{Q}((x, y), \cdot)$ is the sum of a singular part $\bar{Q}_{\perp}((x, y), A) = \bar{P}((x, y), A \cap B_{xy})$ and an absolutely continuous part $\bar{Q}_{\ll}((x, y), \cdot)$ with the density function specified in Condition 2, above.

Next by Lemma 2.9, we know that there exists an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. This $\bar{B}$ has the property that $\bar{Q}((x, y), \cdot)$-almost all values of (x', y'), if $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, then

$$\mathbb{P}\big(b_x = 1 | x, y, x', y'\big) = \mathrm{d}(\Phi_{11} + \Phi_{10})\big((x, y), \cdot\big) / \mathrm{d}\bar{Q}\big((x, y), \cdot\big)$$

$$\mathbb{P}\big(b_y = 1 | x, y, x', y'\big) = \mathrm{d}(\Phi_{11} + \Phi_{01})\big((x, y), \cdot\big) / \mathrm{d}\bar{Q}\big((x, y), \cdot\big).$$

For $A \in \mathscr{F}$ contained in the support of $\bar{Q}_{\perp}((x, y), \cdot)$, we have $\Phi_{11}((x, y), A) = \bar{Q}((x, y), A)$ and $\Phi_{10}((x, y), A) = \Phi_{01}((x, y), A) = 0$, so the Radon–Nikodym derivatives above equal 1 almost surely. This proves Condition 3. Otherwise,

for $\bar{Q}((x, y), \cdot)$-almost all points (x', y') with $q((x, y), (x', y')) > 0$ we have

$$\mathbb{P}\big(b_x = 1 | x, y, x', y'\big) = \frac{p((x, y), (x', y')) + \bar{p}_y(x, x')m(y, y')}{q((x, y), (x', y'))}$$

and similarly for $\mathbb{P}(b_y | x, y, x', y')$. Finally, $p((x, y), (x', y')) = \bar{p}_y(x, x') = \bar{p}_x(y, y') = 0$ whenever $q((x, y), (x', y')) = 0$, thus proving Conditions 4 and 5. $\square$

Corollary 2.11 applies to MH-like algorithms on both continuous and discrete spaces, as long as $Q(x, \cdot)$ has a density or mass function and as long as the proposal x' almost surely differs from the current point x. It provides an explicit way to represent any kernel coupling in terms of proposal and acceptance indicator couplings.

3. Maximal kernel couplings

We now apply the results above to characterize the maximal couplings of a general MH-like transition kernel P. The identification of couplings $\bar{P} \in \Gamma(P, P)$ that induce fast meeting between chains is an important question for both theoretical analysis and in applied work. Maximal couplings $\bar{P} \in \Gamma^{\max}(P, P)$ represent myopically optimal solutions to this problem, in the sense that they achieve the highest one-step meeting probability $\mathbb{P}(X = Y | x, y)$ from each state pair (x, y). Understanding the structure of $\Gamma^{\max}(P, P)$ can aid in coupling design and serves as a reference point in the search for efficient Markovian couplings of MH-like chains.

The literature so far has said little about $\Gamma^{\max}(P, P)$. O'Leary, Wang and Jacob [63] showed that when the state space $\mathcal{X}$ is continuous, this set contains at least a few computationally feasible couplings. That paper showed that for any maximal proposal coupling $\bar{Q} \in \Gamma^{\max}(Q, Q)$, there exists an acceptance indicator coupling $\bar{B}$ such that the resulting transition kernel coupling $\bar{P} \in \Gamma^{\max}(P, P)$. Maximal proposal couplings are often easy to construct, so this observation provides a way to obtain a range of couplings in $\Gamma^{\max}(P, P)$. It also suggests a closer relationship between $\Gamma^{\max}(Q, Q)$ and $\Gamma^{\max}(P, P)$ than exists in general, as we discuss below.

In the following analysis we write $\Delta := \{(z, z) : z \in \mathcal{X}\}$ for the diagonal of $\mathcal{X} \times \mathcal{X}$, $\delta : \mathcal{X} \to \Delta$ for the map $z \mapsto (z, z)$, and $A_\Delta := \delta(A) = \{(z, z) : z \in A\}$. We continue to assume that $(\mathcal{X}, \mathscr{F})$ is a Polish space, so that $\Delta \in \mathscr{F} \otimes \mathscr{F}$ and so that δ is a measurable function. As noted above, the coupling inequality states that if μ and ν are probability measures on $(\mathcal{X}, \mathscr{F})$ and if $\gamma \in \Gamma(\mu, \nu)$, then $\mathbb{P}_{(X,Y) \sim \gamma}(X = Y) \leq 1 - \|\mu - \nu\|_{\mathrm{TV}}$. A maximal coupling is one that achieves this bound. In the following section we see that this is equivalent to a measure-theoretic condition, which we use to characterize maximal couplings of MH-like kernels in Section 3.3.

3.1. The Hahn maximality condition

Given probability measures μ and ν on $(\mathcal{X}, \mathscr{F})$, the Hahn-Jordan theorem (e.g., [23, Chapter 5.6]) states that there exists a measurable set $S \in \mathscr{F}$ and sub-probability measures μ^r and ν^r such that $\mu - \nu = \mu^r - \nu^r$ and $\mu^r(S^c) = \nu^r(S) = 0$. The pair (S, S^c) is called a Hahn decomposition for $\mu - \nu$, and it is essentially unique in the sense that if $R \in \mathscr{F}$ is another measurable set with $\mu^r(R^c) = \nu^r(R) = 0$, then $(\mu - \nu)(S \triangle R) = 0$. Here $A \triangle B = (A \setminus B) \cup (B \setminus A)$ indicates the symmetric difference of measurable sets $A, B \in \mathscr{F}$. The pair (μ^r, ν^r) is called the Jordan decomposition of $\mu - \nu$, and it is uniquely characterized by the above.

In the Jordan decomposition $\mu - \nu = \mu^r - \nu^r$, μ^r and ν^r are called the upper and lower variation of $\mu - \nu$, and $\mu \wedge \nu := \mu - \mu^r = \nu - \nu^r$ is called the meet or infimum measure of μ and ν. Here $\mu \wedge \nu$ is non-negative and has the defining property that if η is another measure on $(\mathcal{X}, \mathscr{F})$ with $\eta(A) \leq \mu(A) \wedge \nu(A)$ for all $A \in \mathscr{F}$, then $\eta(A) \leq (\mu \wedge \nu)(A)$ for all $A \in \mathscr{F}$. By the definition of total variation, $\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \in \mathscr{F}} |\mu(A) - \nu(A)| = \mu^r(\mathcal{X}) = \nu^r(\mathcal{X}) = 1 - (\mu \wedge \nu)(\mathcal{X})$. See e.g. Dshalalow [21, Chapter 5] or Aliprantis and Burkinshaw [3, Section 36] for more on the lattice-theoretic properties of the set of measures on $(\mathcal{X}, \mathscr{F})$.

For any measure μ on $(\mathcal{X}, \mathscr{F})$ let $\delta_\star \mu$ be the pushforward of μ by the diagonal map δ, so that $\delta_\star \mu(A) = \mu(\delta^{-1}(A))$ for $A \in \mathscr{F} \otimes \mathscr{F}$. This makes $\delta_\star \mu$ a measure on $(\mathcal{X} \times \mathcal{X}, \mathscr{F} \otimes \mathscr{F})$ with support contained in Δ and with $\delta_\star \mu(B_\Delta) = \mu(B)$ for any $B \in \mathscr{F}$. In this notation, maximal couplings can be characterized as follows:

Lemma 3.1 (Douc et al. [20], Theorem 19.1.6). *Let $\mu - \nu = \mu^r - \nu^r$ be the Jordan decomposition of a pair of probability measures μ and ν on $(\mathcal{X}, \mathscr{F})$. A coupling $\gamma \in \Gamma(\mu, \nu)$ is maximal if and only if there exists a $\gamma^r \in \Gamma(\mu^r, \nu^r)$ such that $\gamma(A) = \gamma^r(A) + \delta_\star(\mu \wedge \nu)(A)$ for all $A \in \mathscr{F} \otimes \mathscr{F}$.*

Note that we must have $\mu^r(\mathcal{X}) = \nu^r(\mathcal{X})$ for $\Gamma(\mu^r, \nu^r)$ to be nonempty. This follows from the Jordan decomposition, since $0 = (\mu - \nu)(\mathcal{X}) = (\mu^r - \nu^r)(\mathcal{X})$. Also $\gamma^r(\Delta) = 0$, as

$$1 - \|\mu - \nu\|_{\mathrm{TV}} = \gamma(\Delta) = \gamma^r(\Delta) + \delta_\star(\mu \wedge \nu)(\Delta) = \gamma^r(\Delta) + (\mu \wedge \nu)(\mathcal{X}) = \gamma^r(\Delta) + 1 - \|\mu - \nu\|_{\mathrm{TV}}.$$

Finally, we observe that Lemma 3.1 implies the maximal coupling recognition result of Ernst et al. [24, Lemma 20]. Thus we have the following characterization of maximal couplings based on the Hahn decomposition:

Corollary 3.2 (Hahn Maximality Condition). *Let μ and ν be measures on $(\mathcal{X}, \mathcal{F})$. A coupling $\gamma \in \Gamma(\mu, \nu)$ is maximal if and only if there is an $S \in \mathcal{F}$ with $\gamma((S^c \times \mathcal{X}) \setminus \Delta) = \gamma((\mathcal{X} \times S) \setminus \Delta) = 0$. Any (S, S^c) with this property will be a Hahn decomposition for $\mu - \nu$.*

Proof. Let $\mu - \nu = \mu^r - \nu^r$ be a Jordan decomposition, so that we have $\mu^r(S^c) = \nu^r(S) = 0$ for some $S \in \mathcal{F}$. If $\gamma \in \Gamma(\mu, \nu)$ is maximal, Lemma 3.1 implies that $\gamma(A) = \gamma^r(A) + \delta_\star(\mu \wedge \nu)(A)$ for all $A \in \mathcal{F} \otimes \mathcal{F}$. Therefore $\gamma((S^c \times \mathcal{X}) \setminus \Delta) = \gamma^r(S^c \times \mathcal{X}) = \mu^r(S^c) = 0$. By a similar argument, $\gamma((\mathcal{X} \times S) \setminus \Delta) = 0$. For the converse, let $\gamma \in \Gamma(\mu, \nu)$ and $\gamma((S^c \times \mathcal{X}) \setminus \Delta) = \gamma((\mathcal{X} \times S) \setminus \Delta) = 0$. Then $\mu(B) = \gamma(B \times \mathcal{X}) = \gamma((B \times \mathcal{X}) \setminus \Delta) + \gamma(B_\Delta)$ and $\nu(B) = \gamma(\mathcal{X} \times B) = \gamma((\mathcal{X} \times B) \setminus \Delta) + \gamma(B_\Delta)$ for any $B \in \mathcal{F}$.

By assumption, S contains the support of $\gamma((\cdot \times \mathcal{X}) \setminus \Delta)$ and S^c contains the support of $\gamma((\mathcal{X} \times \cdot) \setminus \Delta)$. Thus $\mu(\cdot) - \nu(\cdot) = \gamma((\cdot \times \mathcal{X}) \setminus \Delta) - \gamma((\mathcal{X} \times \cdot) \setminus \Delta)$ is the Jordan decomposition of $\mu - \nu$ and (S, S^c) is a Hahn decomposition. The uniqueness of the Jordan decomposition implies $\gamma((B \times \mathcal{X}) \setminus \Delta) = \mu^r(B)$ and $\gamma((\mathcal{X} \times B) \setminus \Delta) = \nu^r(B)$, which in turn yields $\gamma^r(\cdot) := \gamma(\cdot \setminus \Delta) \in \Gamma(\mu^r, \nu^r)$. We also have $(\mu \wedge \nu)(B) = \mu(B) - \mu^r(B) = \nu(B) - \nu^r(B)$, so the above implies $\gamma(B_\Delta) = (\mu \wedge \nu)(B)$ for all $B \in \mathcal{F}$. We conclude that $\gamma(A) = \gamma(A \setminus \Delta) + \gamma(A \cap \Delta) = \gamma^r(A) + \delta_\star(\mu \wedge \nu)(A)$ for any $A \in \mathcal{F} \otimes \mathcal{F}$. $\qquad\square$

In Section 3.3, we use this result to establish conditions for $\bar{P} \in \Gamma^{\mathrm{max}}(P, P)$ in terms of proposal and acceptance couplings. Before doing so, we need to clarify the relationship between the maximality of a transition kernel coupling $\bar{P}$ and the maximality of a proposal coupling $\bar{Q}$ that generates it.

3.2. *Maximal proposal kernel couplings*

It seems reasonable to guess that if $\bar{P} \in \Gamma^{\mathrm{max}}(P, P)$ is generated by a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$, then $\bar{Q}$ must itself be maximal. The proposal-based maximal couplings of O'Leary, Wang and Jacob [63] have this property, and it seems plausible that in order to maximize the probability of $X = Y$ one might need to start by maximizing the probability of $x' = y'$. However, the following shows that no special relationship exists between maximal proposal and transition couplings.

Lemma 3.3. *Suppose the transition kernel coupling $\bar{P} \in \Gamma^{\mathrm{max}}(P, P)$ is generated by a coupling $\bar{Q}_m \in \Gamma^{\mathrm{max}}(Q, Q)$ and an acceptance coupling $\bar{B}_m$. Suppose that $\bar{Q}_m((x, y), \Delta^c) > 0$ for some (x, y), and at that (x, y), $\mathbb{P}(b_x = b_y = 1 | x, y) < 1$ where $(x'_m, y'_m) \sim \bar{Q}_m((x, y), \cdot)$ and $(b_x, b_y) \sim \bar{B}_m((x, y), (x'_m, y'_m))$. Then there exists a non-maximal coupling $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ also generate $\bar{P}$.*

In the following proof, we use $\bar{Q}_m$ to construct a $\bar{Q}$ that agrees with $\bar{Q}_m$ on accepted proposals and independently redraws rejected ones. The hypotheses on $\bar{Q}_m$ and on (b_x, b_y) are needed to ensure that the resulting $\bar{Q}$ is not maximal.

Proof. Let $\bar{Q}((x, y), \cdot)$ be the distribution of the (x', y') output of Algorithm 1. We claim that $\bar{Q} \in \Gamma(Q, Q)$. For $A \in \mathcal{F}$,

$$\bar{Q}\big((x, y), A \times \mathcal{X}\big) = \mathbb{P}\big(x'_m \in A, b_x = 1 | x, y\big) + \sum_{j \in \{0,1\}} \mathbb{P}(b_x = 0, b_y = j | x, y)\, \mathbb{P}(\tilde{x} \in A | \tilde{b}_x = 0, \tilde{b}_y = j, x, y)$$

$$= \mathbb{P}\big(x'_m \in A, b_x = 1 | x, y\big) + \sum_{j \in \{0,1\}} \mathbb{P}\big(x'_m \in A, b_x = 0, b_y = j | x, y\big) = \mathbb{P}\big(x'_m \in A | x\big) = Q(x, A).$$

Similarly, $\bar{Q}((x, y), \mathcal{X} \times A) = Q(y, A)$. Let (x, y) be such that $\bar{Q}_m((x, y), \Delta^c) > 0$. We have assumed that there is a positive probability of rejecting either x'_m or y'_m at (x, y). Thus $\mathbb{P}(x' = y' | x, y) < \mathbb{P}(x'_m = y'_m | x, y)$, and so we conclude that $\bar{Q}$ is not a maximal coupling.

Algorithm 1 Construction of $\bar{Q}$ for Lemma 3.3

---

1. Draw $(x'_m, y'_m) \sim \bar{Q}_m((x, y), \cdot)$ and $(b_x, b_y) \sim \bar{B}_m((x, y), (x'_m, y'_m))$
2. For $z \in \{x, y\}$:
   (a) If $b_z = 1$, set $z' = z'_m$. Else:
   (b) Draw $(\tilde{x}, \tilde{y}) \sim \bar{Q}_m((x, y), \cdot)$ and $(\tilde{b}_x, \tilde{b}_y) \sim \bar{B}_m((x, y), (\tilde{x}, \tilde{y}))$
   (c) If $(\tilde{b}_x, \tilde{b}_y) = (b_x, b_y)$, set $z' = \tilde{z}$
   (d) Else: go to 2(b)
3. Return (x', y') and (b_x, b_y)

Next, for $i, j \in \{0, 1\}$ and $A \in \mathscr{F} \otimes \mathscr{F}$, define $\Phi_{ij}((x, y), A) := \mathbb{P}((x', y') \in A, b_x = i, b_y = j)$ using the full output of Algorithm 1. We observe that this is a coupled acceptance mechanism relating $\bar{Q}$ and $\bar{P}$. The first condition of Definition 2.3 is satisfied by construction. For the second condition, define $X_m = b_x x'_m + (1 - b_x)x$ and $Y_m = b_y y'_m + (1 - b_y)y$. Since $\bar{Q}_m$ and $\bar{B}_m$ generate $\bar{P}$, we must have $(X_m, Y_m) \sim \bar{P}((x, y), \cdot)$. Thus for any $A_x \times A_y \in \mathscr{F} \otimes \mathscr{F}$,

$$\Phi_{11}\big((x, y), A_x \times A_y\big) + \Phi_{10}\big((x, y), A_x \times \mathcal{X}\big)1(y \in A_y)$$

$$+ \Phi_{01}\big((x, y), \mathcal{X} \times A_y\big)1(x \in A_x) + \Phi_{00}\big((x, y), \mathcal{X} \times \mathcal{X}\big)1(x \in A_x)1(y \in A_y)$$

$$= \mathbb{P}(X_0 \in A_x, Y_0 \in A_y, b_x = 1, b_y = 1|x, y) + \mathbb{P}(X_0 \in A_x, Y_0 = y, b_x = 1, b_y = 0|x, y)$$

$$+ \mathbb{P}(X_0 = x, Y_0 \in A_y, b_x = 0, b_y = 1|x, y) + \mathbb{P}(X_0 = x, Y_0 = y, b_x = 0, b_y = 0|x, y) = \bar{P}\big((x, y), A_x \times A_y\big).$$

The third condition on Φ follows from the fact that $b_x = 1$ if $x'_m = x$ and $b_y = 1$ if $y'_m = y$. Since Φ is a coupled acceptance mechanism relating $\bar{Q}$ and $\bar{P}$, Lemma 2.9 ensures the existence of a $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$. □

Lemma 3.3 shows that if a maximal transition kernel coupling $\bar{P} \in \Gamma^{\max}(P, P)$ is generated by a proposal coupling $\bar{Q}$ and an acceptance coupling $\bar{B}$, then $\bar{Q}$ need not be maximal. In Appendix A.2 we observe that some maximal couplings $\bar{P}$ cannot be generated from any maximal coupling $\bar{Q} \in \Gamma^{\max}(Q, Q)$. Although O'Leary, Wang and Jacob [63] showed that we can usually derive *some* maximal coupling $\bar{P}$ from a maximal coupling $\bar{Q}$, we conclude that there is no general relationship between the maximality of a proposal kernel coupling and that of an associated transition kernel coupling.

3.3. *Characterization of maximal kernel couplings*

We now turn to the main result of this section, which extends Theorem 2.2 to characterize the maximal couplings of an MH-like transition kernel in terms of proposal and acceptance indicator couplings. For each $x, y \in \mathcal{X}$, let (S_{xy}, S_{xy}^c) be a Hahn decomposition for $P(x, \cdot) - P(y, \cdot)$. Thus $S_{xy} \in \mathscr{F}$ and $P(x, A) \geq P(y, A)$ for any $A \in \mathscr{F}$ with $A \subset S_{xy}$. Note that if $P(x, \cdot)$ and $P(y, \cdot)$ have Radon–Nikodym derivatives $p(x, \cdot)$ and $p(y, \cdot)$ with respect to a common dominating measure, then we can use $S_{xy} := \{z : p(x, z) \geq p(y, z)\}$ to form these sets.

Theorem 3.4. *Let P be the MH-like transition kernel on $(\mathcal{X}, \mathscr{F})$ generated by a proposal kernel Q and an acceptance rate function a. Then $\bar{P} \in \Gamma^{\max}(P, P)$ if and only if $\bar{P}$ is generated by $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ with the following properties: if $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, then for all $x, y \in \mathcal{X}$:*

1. $\mathbb{P}(b_x = 1|x, y, x') = a(x, x')$ *for $Q(x, \cdot)$-almost all x'*
2. $\mathbb{P}(b_y = 1|x, y, y') = a(y, y')$ *for $Q(y, \cdot)$-almost all y'*

and for $\bar{Q}((x, y), \cdot)$-almost all (x', y'),

3. $\mathbb{P}(b_x = b_y = 1|x, y, x', y') = 0$ *if $x' \neq y'$ and either $x' \in S_{xy}^c$ or $y' \in S_{xy}$*
4. $\mathbb{P}(b_x = 1, b_y = 0|x, y, x', y') = 0$ *if $x' \neq y$ and either $x' \in S_{xy}^c$ or $y \in S_{xy}$*
5. $\mathbb{P}(b_x = 0, b_y = 1|x, y, x', y') = 0$ *if $y' \neq x$ and either $x \in S_{xy}^c$ or $y' \in S_{xy}$*
6. $\mathbb{P}(b_x = 0, b_y = 0|x, y, x', y') = 0$ *if $x \neq y$ and either $x \in S_{xy}^c$ or $y \in S_{xy}$.*

Recall that by Corollary 3.2, the maximality of a coupling $\bar{P} \in \Gamma(P, P)$ is equivalent to a condition on the support of each $\bar{P}((x, y), \cdot)$. Conditions 3-6 relate these support constraints to the behavior of a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$. See Figure 2 for an illustration of the acceptance scenarios considered in these conditions and a visual intuition for why certain cases must be ruled out for $\bar{Q}$ and $\bar{B}$ to generate a maximal $\bar{P}$.
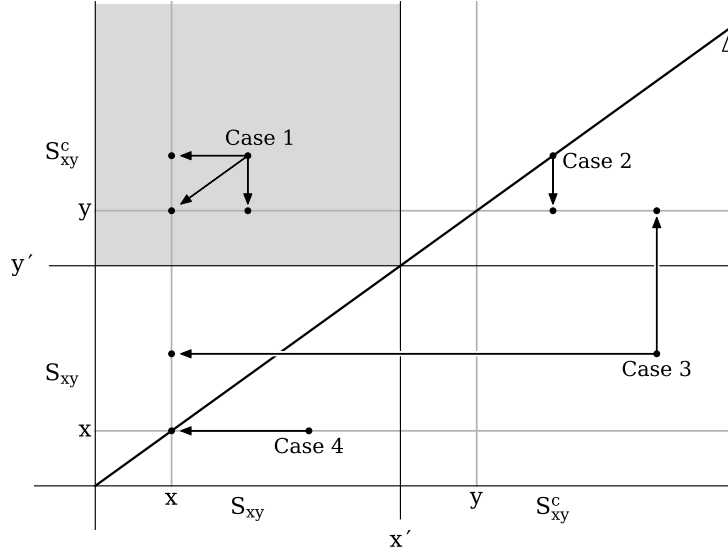
Fig. 2. Diagram of acceptance scenarios considered in Theorem 3.4. The support of a maximal coupling $\bar{P}((x, y), \cdot)$ is contained in the union of $S_{xy} \times S_{xy}^c$ (gray box) and Δ (the diagonal). Arrows illustrate the relationship of proposals (x', y') to transitions (X, Y) under different accept/reject combinations, with transitions outside the support of $\bar{P}((x, y), \cdot)$ forbidden almost surely. Case 1: the maximality of $\bar{P}$ does not constrain the acceptance pattern of proposals $(x', y') \in S_{xy} \times S_{xy}^c$. Case 2: proposals in Δ can be fully accepted ($b_x = b_y = 1$) or fully rejected ($b_x = b_y = 0$), but y' must be accepted if $x' \in S_{xy}^c$, and x' must be accepted if $y' \in S_{xy}$. Case 3: proposals in $(S_{xy} \times S_{xy}^c)^c \cap \Delta^c$ must be fully rejected unless $y' = x$ or $x' = y$. Case 4: a proposal (x', y') outside the support of $\bar{P}((x, y), \cdot)$ may be partially accepted ($b_x \neq b_y$) if it yields a transition to (x, x) or (y, y).

Proof of Theorem 3.4. Suppose $\bar{P} \in \Gamma^{\max}(P, P)$. By Theorem 2.2, there exists a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ and an acceptance indicator coupling $\bar{B}$ such that for any $x, y \in \mathcal{X}$ and $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, we have $\mathbb{P}(b_x = 1 | x, y, x') = a(x, x')$ for $Q(x, \cdot)$-almost all x', and $\mathbb{P}(b_y = 1 | x, y, y') = a(y, y')$ for $Q(y, \cdot)$-almost all y'. Thus Conditions 1 and 2 directly follow from Theorem 2.2.

Since $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$, we have $(X, Y) \sim \bar{P}((x, y), \cdot)$ where $X = b_x x' + (1 - b_x)x$ and $Y = b_y y' + (1 - b_y)y$. Since $\bar{P}$ is maximal, Corollary 3.2 implies $0 = \bar{P}((x, y), (S_{xy}^c \times \mathcal{X}) \setminus \Delta) = \mathbb{P}(X \in S_{xy}^c, X \neq Y | x, y)$. Breaking this up into the four possible acceptance scenarios $(b_x, b_y) = (1, 1), (1, 0), (0, 1),$ and $(0, 0)$ yields

$$0 = \mathbb{E}\left[1\left(x' \neq y', x' \in S_{xy}^c\right)\mathbb{P}(b_x = b_y = 1 | x, y, x', y')\right] \qquad 0 = \mathbb{E}\left[1\left(x' \neq y, x' \in S_{xy}^c\right)\mathbb{P}(b_x = 1, b_y = 0 | x, y, x', y')\right]$$

$$0 = \mathbb{E}\left[1\left(y' \neq x, x \in S_{xy}^c\right)\mathbb{P}(b_x = 0, b_y = 1 | x, y, x', y')\right] \qquad 0 = \mathbb{E}\left[1\left(x \in S_{xy}^c \setminus \{y\}\right)\mathbb{P}(b_x = b_y = 0 | x, y, x', y')\right].$$

In turn, these equations imply that for $\bar{Q}((x, y), \cdot)$-almost all (x', y'),

$$\mathbb{P}(b_x = b_y = 1 | x, y, x', y') = 0 \text{ if } x' \neq y' \text{ and } x' \in S_{xy}^c \qquad \mathbb{P}(b_x = 1, b_y = 0 | x, y, x', y') = 0 \text{ if } x' \neq y \text{ and } x' \in S_{xy}^c$$

$$\mathbb{P}(b_x = 0, b_y = 1 | x, y, x', y') = 0 \text{ if } y' \neq x \text{ and } x \in S_{xy}^c \qquad \mathbb{P}(b_x = b_y = 0 | x, y, x', y') = 0 \text{ if } x \neq y \text{ and } x \in S_{xy}^c.$$

This shows that the first either/or case of each of Conditions 3-6 are satisfied. Since $\bar{P}$ is maximal, Corollary 3.2 also implies $0 = \bar{P}((x, y), (\mathcal{X} \times S_{xy}) \setminus \Delta)$. Proceeding as above shows that the second either/or cases are also satisfied. So we conclude that if $\bar{P} \in \Gamma^{\max}(P, P)$, then $\bar{P}$ is generated by a $\bar{Q}$ and $\bar{B}$ satisfying the six conditions stated above.

For the converse, suppose that $\bar{Q} \in \Gamma(Q, Q)$ and a proposal coupling $\bar{B}$ generate $\bar{P}$ and satisfy the given hypotheses. Since Conditions 1 and 2 are equivalent to the conditions of Theorem 2.2, we have $\bar{P} \in \Gamma(P, P)$. Now let $(x', y') \sim \bar{Q}((x, y), \cdot)$, $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$, $X = b_x x' + (1 - b_x)x$, and $Y = b_y y' + (1 - b_y)y$. $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$, so

$$\bar{P}\left((x, y), \left(S_{xy}^c \times \mathcal{X}\right) \setminus \Delta\right) = \mathbb{P}\left(X \in S_{xy}^c, X \neq Y | x, y\right)$$

$$= \mathbb{P}\left(x' \in A, x' \neq y', b_x = 1, b_y = 1 | x, y\right) + \mathbb{P}\left(x' \in A, x' \neq y, b_x = 1, b_y = 0 | x, y\right)$$

$$+ \mathbb{P}\left(y' \neq x, b_x = 0, b_y = 1 | x, y\right)1(x \in A) + \mathbb{P}(b_x = 0, b_y = 0 | x, y)1(x \in A)1(x \neq y) = 0.$$

The last equality follows directly from Conditions 3-6 of the Theorem, with Condition 3 ensuring that the first term equals zero, Condition 4 ensuring that the second term equals zero, and so on. A similar argument yields

$\bar{P}((x, y), (\mathcal{X} \times S_{xy}) \setminus \Delta) = 0$. By Corollary 3.2, $\bar{P}$ is maximal if and only if there is a measurable set $S \in \mathcal{F}$ such that $\bar{P}((x, y), (S^c \times \mathcal{X}) \setminus \Delta) = \bar{P}((x, y), (\mathcal{X} \times S) \setminus \Delta) = 0$. The argument above shows that S_{xy} has these properties, so we conclude that $\bar{P} \in \Gamma^{\max}(P, P)$. $\qquad\square$

In Lemma 3.3 and Appendix A.2, we observed that maximal proposal couplings only weakly relate to maximal transition kernel couplings. Theorem 3.4 shows that any proposal coupling can give rise to a maximal coupling, as long as it satisfies the conditions of the theorem for some acceptance indicator coupling $\bar{B}$. In particular, such a $\bar{Q}$ must distribute probability over Δ so that $P(x, A) \wedge P(y, A) \leq \bar{Q}((x, y), A_\Delta)$ for all $A \in \mathcal{F}$.

As noted above, most of the MH couplings considered so far in the literature use a maximal proposal coupling $\bar{Q} \in \Gamma^{\max}(Q, Q)$. However, the examples of Jacob, O'Leary and Atchadé [41] and O'Leary [62] show that some of these perform better than others, and that the distribution of meeting times often reflects the degree of contraction between chains when a meeting is not proposed rather than the frequency with which meetings are proposed. These examples suggest that non-maximal proposal couplings cannot be dismissed out of hand. The results above sharpen this conclusion by observing that with the right acceptance indicator coupling, a non-maximal proposal coupling can still produce transition kernel meetings at the optimal rate.

4. Two-step representations of common MH algorithms

In Theorem 2.2 and Corollary 2.11, we showed that any coupling of MH-like transition kernels can be expressed as a proposal coupling followed by an acceptance indicator coupling. In Sections 4.1 and 4.2, we apply these results to couplings of the MH algorithms on a discrete state space, the random walk Metropolis algorithm, and the Metropolis-adjusted Langevin algorithm. In Section 4.3 we consider a two-step representation of the transition kernel coupling used in Nummelin splitting [44,73]. An overall impression from these examples is that our results are simple to apply and can sometimes recast complex kernel couplings in a more approachable, two-step form.

4.1. The Metropolis–Hastings algorithm on a finite state space

We begin by considering a finite state space $\mathcal{X}$ with an MH-like transition kernel P based on a proposal kernel Q with $Q(x, \{x\}) = 0$ for all $x \in \mathcal{X}$. Fix any transition kernel coupling $\bar{P} \in \Gamma(P, P)$. We write $r(x) = P(x, \{x\})$ for the probability of a transition from x to itself, and in this subsection we also write $\bar{\Theta}((x, y), (x', y'))$ in place of $\bar{\Theta}((x, y), \{x'\} \times \{y'\})$, where $\bar{\Theta}$ is any joint kernel in $\Gamma(Q, Q)$ or $\Gamma(P, P)$.

Corollary 2.11 shows that $\bar{P}$ can be generated by a particular proposal coupling $\bar{Q}$ and acceptance indicator coupling $\bar{B}$. Condition 1 of that result implies that $\bar{Q}_\perp((x, y), \cdot) = \bar{P}_\perp((x, y), \cdot) = 0$, since the base measure on $\mathcal{X} \times \mathcal{X}$ is counting measure. Condition 2 then shows that we can express $\bar{Q}((x, y), (x', y'))$ as the sum of four terms, Φ_{11}, Φ_{10}, Φ_{01}, and Φ_{00}. These all equal zero if $x' = x$ or $y' = y$, since then $0 \leq \Phi_{ij}((x, y), (x', y')) \leq \bar{Q}((x, y), (x', y')) = 0$ by our 'non-laziness' condition on Q. For $x' \neq x$ and $y' \neq y$, we have

$$\Phi_{11}((x, y), (x', y')) = \bar{P}((x, y), (x', y'))$$

$$\Phi_{10}((x, y), (x', y')) = \bar{P}((x, y), (x', y))\mu(y, y')\mathbf{1}(r(y) > 0)$$

$$\Phi_{01}((x, y), (x', y')) = \mu(x, x')\mathbf{1}(r(x) > 0)\bar{P}((x, y), (x, y'))$$

$$\Phi_{00}((x, y), (x', y')) = \mu(x, x')\mu(y, y')\mathbf{1}(r(x) > 0)\mathbf{1}(r(y) > 0)\bar{P}((x, y), (x, y)).$$

If $x' \neq x$ and $r(x) > 0$, then $\mu(x, x') = (Q(x, \{x'\}) - P(x, \{x'\}))/r(x)$. From here, it is easy to confirm that $\bar{Q} \in \Gamma(Q, Q)$. Summing $\bar{Q}((x, y), (x', y'))$ over y' yields $Q(x, \{x'\})$, since

$$\bar{P}((x, y), \{x'\} \times \{y\}^c) + \bar{P}((x, y), (x', y))\mathbf{1}(r(y) > 0)$$

$$+ \mu(x, x')\mathbf{1}(r(x) > 0)r(x) + \mu(x, x')\mathbf{1}(r(x) > 0)\mathbf{1}(r(y) > 0)\bar{P}((x, y), (x, y))$$

$$= P(x, \{x'\}) + (Q(x, \{x'\}) - P(x, \{x'\})) = Q(x, \{x'\}).$$

A similar argument shows that $\sum_{x'} \bar{Q}((x, y), (x', y')) = Q(y, \{y'\})$.

We can also derive the necessary acceptance indicator coupling by using the Φ formulas above with the proof method of Lemma 2.9. Let (x, y) be the current state and let (x', y') be a proposed state pair with $\bar{Q}((x, y), (x', y')) > 0$. Then the probability of accepting both proposals will be $\Phi_{11}((x, y), (x', y'))/\bar{Q}((x, y), (x', y'))$, the probability of accepting

x' and rejecting y' will be $\Phi_{10}((x, y), (x', y'))/\bar{Q}((x, y), (x', y'))$, and so on. These acceptance probabilities are defined in terms of Radon–Nikodym derivatives, which are simple to compute in this case due to the discrete state space.

The current example represents a significant specialization of the hypotheses of Theorem 2.2 and Corollary 2.11. Still, it provides a useful intuition for how our methods work in more general situations. In this case, we see that $\bar{Q}((x, y), (x', y'))$ arises as a simple combination of the values of $\bar{P}((x, y), \cdot)$ under the four possible acceptance scenarios, using weights derived from μ. We recall from Lemma 2.6 that $\mu(x, x') = \mathbb{P}(x' = x | b_x = 0, x)$ under suitable conditions. Thus we see that the form of $\bar{Q}$ in our existence results is the one that assumes the least possible dependence between x' and y' when one or both of these are rejected. This observation holds in the general case, as well.

4.2. *The random walk Metropolis and Metropolis-adjusted Langevin algorithms*

Next, we consider two popular MH kernels on the continuous state space $\mathcal{X} = \mathbb{R}^n$. We write λ for Lebesgue measure on $\mathcal{X}$, and we take π to be a target distribution on $\mathcal{X}$ with density $\pi(\cdot)$ with respect to λ. We will also write Q for any proposal kernel on $\mathcal{X}$ with density $q(x, \cdot)$, a for the associated MH acceptance rate, P for the resulting transition kernel, and $\bar{P} \in \Gamma(P, P)$ for some fixed kernel coupling based on P.

For the Random Walk Metropolis (RWM) algorithm, we use a proposal density q such that $q(x, x') = g(\|x - x'\|)$ for some $g : \mathbb{R}^+ \to \mathbb{R}^+$. We also consider the Metropolis-adjusted Langevin algorithm (MALA) [72], which uses a discretized approximation to the Langevin diffusion $\mathrm{d}X_t = \frac{1}{2}\nabla \log \pi(X_t)\mathrm{d}t + \mathrm{d}W_t$, where W_t represents a Brownian motion. For MALA, we use the proposal distribution $Q(x, \cdot) = \mathrm{N}(x + \tau\nabla\pi(x), 2\tau\mathbb{I})$, where $\tau > 0$ serves as a tuning parameter. This distribution has density $q(x, y) \propto \exp(-\frac{1}{4\tau}\|y - x - \tau\nabla\log\pi(x)\|^2)$. MALA's use of gradient information makes it more computationally expensive, but it tends to outperform RWM in high dimensions [71].

For both algorithms, Corollary 2.11 implies the existence of a proposal coupling $\bar{Q} \in \Gamma(Q, Q)$ that can be written as $\bar{Q} = \bar{Q}_\perp + \bar{Q}_\ll$, where $\bar{Q}_\perp$ is singular and $\bar{Q}_\ll$ is absolutely continuous with respect to the base measure $\lambda \times \lambda$. Condition 1 of Corollary 2.11 tells us that $\bar{Q}_\perp((x, y), \cdot) = \bar{P}_\perp((x, y), \cdot \cap (\{x\}^c \times \{y\}^c))$, the singular part of the transition kernel coupling after removing the probability at the current state pair (x, y). Condition 2 of that result tells us that the continuous part $\bar{Q}_\ll((x, y), \cdot)$ of the proposal coupling will have density

$$\bar{q}\big((x, y), (x', y')\big) = p\big((x, y), (x', y')\big) + \bar{p}_y(x, x')m(y, y') + m(x, x')\bar{p}_x(y, y') + m(x, x')m(y, y')\bar{r}(x, y).$$

Here $\bar{p}_x$, $\bar{p}_y$, m, and $\bar{r}$ are all as defined in Section 2.7.

By Condition 3, proposals (x', y') in the support of $\bar{Q}_\perp((x, y), \cdot)$ are automatically accepted. Otherwise, we can determine the probabilities of accepting one, both, or neither of the proposals by dividing the relevant term of the density above by $\bar{q}$. Thus the probability of accepting both x' and y' will be $p((x, y), (x', y'))/\bar{q}((x, y), (x', y'))$, the probability of accepting x' and rejecting y' will be $\bar{p}_y((x, y)x')m(y, y')/\bar{q}((x, y), (x', y'))$, and so on.

We conclude that the differences between the two-step representation of couplings of these algorithms are contained entirely within the functions that make up the expression for $\bar{q}$. A similar conclusion also holds for more complicated algorithms such as Hamiltonian Monte Carlo [22,58,60] and Metropolis-within-Gibbs [27,52], for which the functions $\bar{p}$ and m require a correspondingly greater effort to derive.

4.3. *Coupling with a minorization condition*

Let P be an MH-like transition kernel generated by a proposal kernel Q and an acceptance rate function a, and assume that P has stationary distribution π. As above, we assume that $Q(x, \cdot) \ll \lambda$ and $Q(x, \{x\}) = 0$ for all $x \in \mathcal{X}$. The kernel P satisfies a minorization condition corresponding to a set $\mathcal{C} \subseteq \mathcal{X}$ if there exists an $\epsilon > 0$ and a probability measure ν on $(\mathcal{X}, \mathscr{F})$ such that $P(x, A) \geq \epsilon\nu(A)$ for all $x \in \mathcal{C}$ and all $A \in \mathscr{F}$. If $\mathcal{C} = \mathcal{X}$, then a simple argument shows that $\|P^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq M(1 - \epsilon)^t$ for some $M > 0$ and each positive integer t [53]. If $\mathcal{C}$ is a proper subset of $\mathcal{X}$, further assumptions are needed to control the distribution of return times of the chain to $\mathcal{C}$ and so to obtain convergence rate results.

Minorization arguments often use a coupling technique called Nummelin splitting [7,28,44,61] to bound the convergence rate of such a chain to its stationary distribution. Suppose that we run two identically distributed Markov chains (X_t) and (Y_t), with one initialized arbitrarily and the other initialized from its stationary distribution π. If both chains occupy $\mathcal{C}$ at the same iteration t, then we flip a coin with probability ϵ of heads. On heads, we draw $X_{t+1} = Y_{t+1} \sim \nu$, and otherwise we draw X_{t+1} and Y_{t+1} independently from their residual distributions so that $X_{n+1} \sim \nu_x^r := (P(X_n, \cdot) - \epsilon\nu(\cdot)/(1 - \epsilon)$ and $Y_{n+1} \sim \nu_y^r := (P(Y_n, \cdot) - \epsilon\nu(\cdot))/(1 - \epsilon)$. Once the chains meet, we update both chains using the same transitions from P, so that they remain together at all subsequent iterations. Otherwise, we repeat this process whenever both chains occupy $\mathcal{C}$, until meeting occurs.

The procedure above yields a coupling $\bar{P} \in \Gamma(P, P)$, defined without direct reference to the proposal or acceptance distributions. Alternatively, we can use Theorem 2.2 to represent this coupling in terms of a proposal coupling $\bar{Q}$ and an acceptance indicator coupling $\bar{B}$. For every $x, y \in C$, the kernel coupling $\bar{P}((x, y), \cdot)$ described above is a mixture of a measure on the diagonal $\Delta = \{(x, x) : x \in \mathcal{X}\}$ and a measure on its complement in $\mathcal{X} \times \mathcal{X}$. In particular, for any measurable set $S \in \mathcal{F} \otimes \mathcal{F}$ we have $\bar{P}((x, y), S) = \epsilon \nu(S \cap \Delta) + (1 - \epsilon)(\nu_x^r \times \nu_y^r)(S \setminus \Delta)$.

The key step in finding a two-step representation of $\bar{P}$ is to identify an appropriate coupled acceptance mechanism $\Phi = (\Phi_{11}, \Phi_{10}, \Phi_{01}, \Phi_{00})$. Due to our assumptions on Q, we have $\beta(x) = 0$ if $r(x) > 0$ and 1 otherwise. We also have

$$\mu(x, A_x) = \begin{cases} \dfrac{Q(x, A_x) - P(x, A_x \setminus \{x\})}{r(x)} & \text{if } r(x) > 0 \\ 1(x \in A_x) & \text{otherwise.} \end{cases}$$

A similar description holds for $\mu(y, A_y)$. From the proof of Lemma 2.8, we can write the components of Φ as follows:

$$\Phi_{11}((x, y), A_x \times A_y) = \bar{P}((x, y), A_x \times A_y \cap (\{x\}^c \times \{y\}^c))$$

$$\Phi_{10}((x, y), A_x \times A_y) = \bar{P}((x, y), (A_x \cap \{x\}^c) \times \{y\}) \times \mu(y, A_y)$$

$$\Phi_{01}((x, y), A_x \times A_y) = \mu(x, A_x) \times \bar{P}((x, y), \{x\} \times (A_y \cap \{y\}^c))$$

$$\Phi_{00}((x, y), A_x \times A_y) = r(x, y)\mu(x, A_x) \times \mu(y, A_y).$$

Here $r(x, y) = \bar{P}((x, y), \{x\} \times \{y\})$ is the probability that the joint chain stays put at (x, y). In this example the two chains evolve independently independent when meeting does not occur, and so we have $r(x, y) = r(x)r(y)$. With Φ defined as above, we take $\bar{Q} = \Phi_{00} + \Phi_{01} + \Phi_{10} + \Phi_{11}$ to be our proposal coupling.

Next we consider the acceptance indicator coupling $\bar{B}$. Lemma 2.9 implies the existence of acceptance indicators $(b_x, b_y) \sim \bar{B}((x, y), (x', y'))$ such that $\mathbb{P}(b_x = i, b_y = j | x, y, x', y') = \phi_{ij}((x, y), (x', y'))$ where $\phi_{ij}((x, y), \cdot) = d\Phi_{ij}((x, y), \cdot)/dQ((x, y), \cdot)$. If we write each term of Φ_{ij} in its integral form, then ϕ_{ij} can also be evaluated at every point as a ratio of densities. For example, let $x' \neq x$, $y' \neq y$ be two different points in $\mathcal{X}$. Then the probability density of $\bar{Q}$ for a move from (x, y) to (x', y') will be

$$q((x, y), (x', y')) = (p(x, x') - \epsilon \nu(x'))(p(y, y') - \epsilon \nu(y')) + (p(x, x') - \epsilon \nu(x'))(q(y, y') - p(y, y'))$$
$$+ (p(x, x') - q(x, x'))(q(y, y') - \epsilon \nu(y')) + (q(x, x') - p(x, x'))(q(y, y') - p(y, y')).$$

Here $p(x, x') = q(x, x')a(x, x')$ is the transition kernel density from x to x'. Thus, when $x' \neq x$, $y' \neq y$, and $x' \neq y'$, we have the following expressions for the distribution of the acceptance indicator pair (b_x, b_y) conditional on the current state (x, y) and the proposal state (x', y'):

$$p_{11} := \mathbb{P}(b_x = 1, b_y = 1 | x, y, x', y') = \frac{(p(x, x') - \epsilon \nu(x'))(p(y, y') - \epsilon \nu(y'))}{q((x, y), (x', y'))}$$

$$p_{10} := \mathbb{P}(b_x = 1, b_y = 0 | x, y, x', y') = \frac{(p(x, x') - \epsilon \nu(x'))(q(y, y') - p(y, y'))}{q((x, y), (x', y'))}$$

$$p_{01} := \mathbb{P}(b_x = 0, b_y = 1 | x, y, x', y') = \frac{(p(x, x') - q(x, x'))(q(y, y') - \epsilon \nu(y'))}{q((x, y), (x', y'))}$$

$$p_{00} := \mathbb{P}(b_x = 0, b_y = 0 | x, y, x', y') = \frac{(q(x, x') - p(x, x'))(q(y, y') - p(y, y'))}{q((x, y), (x', y'))}.$$

The marginal probability of accepting the 'x' move is $p_{10} + p_{11}$ and the probability of accepting the 'y' move is $p_{01} + p_{11}$, consistent with Lemma 2.10. As a sanity check, the probability density of moving from (x, y) to (x', y') according to our coupling and acceptance mechanism is $q((x, y), (x', y')) \times p_{11} = (p(x, x') - \epsilon \nu(x')) \cdot (p(y, y') - \epsilon \nu(y'))$, which is indeed the probability density of the minorization coupling $\bar{P}((x, y), \cdot)$ at point (x', y'). The cases with $y' = x'$, $x' = x$, and $y' = y$ can be calculated in the same way.

5. Discussion

Couplings are a topic of great theoretical and practical interest. In theoretical work, couplings provide a powerful tool for proving results on Markov chains and other stochastic processes. They are also related to assignment problems under

marginal constraints, a topic of interest since at least Birkhoff [10]. In applications, couplings provide a foundation for techniques related to convergence diagnosis, variance reduction, and unbiased estimation, as described in Section 1.

In this paper we considered kernel and maximal kernel couplings of the MH algorithm and related methods. In Theorem 2.2, we showed that any transition kernel coupling can be expressed in terms of a proposal coupling and an acceptance indicator coupling with certain properties. Conversely, any joint kernel generated this way must be a legitimate transition kernel coupling. In Theorem 3.4 we took this a step further, showing that maximal kernel couplings correspond to pairs of proposal and acceptance couplings subject to a short list of additional conditions. These results provide a unified approach for describing kernel couplings of MH-like algorithms. In principle, our results could be extended to other algorithms, such as multiple-try MH [51], the guided walk of Gustafson [33], and other irreversible MCMC methods. However, it may be more difficult to derive simple two-step expressions like those described in Sections 2.7 and 4 for these other algorithms.

Throughout this study we have viewed the transition kernel P on a state space $(\mathcal{X}, \mathcal{F})$ as arising from a fixed proposal kernel Q and acceptance rate function a. Alternatively, we could have started with any kernel P that is weakly dominated by any other kernel Q, in the sense that $P(x, A) \leq Q(x, A)$ for all $x \in \mathcal{X}$ and all $A \in \mathcal{F}$ with $x \notin A$. This is essentially the ordering condition of Peskun [64], as generalized to infinite state spaces by Tierney [76]. In light of the above, our Theorem 2.2 can be interpreted as stating that given a kernel P weakly dominated by a kernel Q, and given a Radon–Nikodym derivative $a(x, \cdot) = \mathrm{d}P(x, \cdot)/\mathrm{d}Q(x, \cdot)$ away from the current state x, then every kernel coupling $\bar{P} \in \Gamma(P, P)$ can be generated by a coupling $\bar{Q} \in \Gamma(Q, Q)$ followed by an acceptance indicator coupling $\bar{B}$ based on a. This interpretation turns our result into a joint probability existence and characterization theorem that includes MH-like transition kernel couplings as a special case.

Looking forward, we believe that our results can be used to design new and more efficient couplings for the MH algorithm. We also expect them to support theoretical work on kernel couplings for this broad class of discrete-time Markov chains. It may be possible to derive meeting time bounds or information on the spectrum of $\bar{P}$ based on data about $\bar{Q}$ and $\bar{B}$, using a version of the drift and minorization arguments of Rosenthal [73,75] or the techniques of Atchadé and Perron [6]. A deeper understanding of kernel couplings may also support optimization arguments like those of Boyd et al. [14–16]. Finally, in analogy with our description of maximal couplings in Theorem 3.4, it may be possible to characterize the optimal transport couplings in $\Gamma(P, P)$ in terms of the properties of proposal and acceptance indicator couplings.

This study represents a step toward understanding the properties of $\Gamma(P, P)$ and $\Gamma^{\max}(P, P)$ in terms of those of $\Gamma(Q, Q)$ and the associated acceptance indicator couplings. Markovian couplings which are efficient in the sense of Aldous [1] are well understood only in special cases. Beyond that, few other bounds on the efficiency of Markovian couplings are well-established [17]. There remain much to learn about such couplings for MH-like chains. A clearer understanding of the set of Markovian couplings will make it possible to identify better options for use in practice and to determine how efficient couplings can be for this important class of algorithms.

Appendix A: Simple examples

In this appendix we present two simple examples to illustrate concepts from the main text. In the first example, we explore the definitions and proof methods of Section 2 using an MH kernel on a two-point state space. In the second example, we consider a three-point state space to show that some maximal kernel couplings can only be generated from non-maximal proposal couplings, as discussed in Section 3.

A.1. *Two-step representation of a transition kernel coupling*

Let $\mathcal{X} = \{1, 2\}$, $\mathcal{F} = 2^{\mathcal{X}}$, and assume a current state pair $(x, y) = (1, 2)$. For a small finite state space like this, it is convenient to represent a kernel $\Theta : \mathcal{X} \times \mathcal{F} \to [0, 1]$ by $\Theta(z, \cdot) = (\Theta(z, \{1\}), \Theta(z, \{2\}))$ and to represent a coupling $\bar{\Theta} : (\mathcal{X} \times \mathcal{X}) \times (\mathcal{F} \otimes \mathcal{F}) \to [0, 1]$ using a table like the following:

$$
\bar{\Theta}((x, y), \cdot) =
\begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & \Theta((x,y),(1,1)) & \Theta((x,y),(2,1)) \\
\hline
2 & \Theta((x,y),(1,2)) & \Theta((x,y),(2,2)) \\
\hline
 & 1 & 2
\end{array} \leftarrow y
$$

The columns of this table represent the destination values for the x chain (e.g. the values of x' or X) while the rows represent the equivalent for the y chain. The small 'x' and 'y' serve as a reminder of the current state pair.

We assume a uniform proposal distribution and an MH acceptance rate function a based on a target distribution $\pi = (1/3, 2/3)$, so that $a(1, \cdot) = (1, 1)$ and $a(2, \cdot) = (1/2, 1)$. Together these imply a transition kernel P with $P(1, \cdot) = (1/2, 1/2)$ and $P(2, \cdot) = (1/4, 3/4)$. Simple algebra shows that any proposal and transition coupling based on the kernels above must take following forms, for some $\rho \in [0, 1/2]$ and $\lambda \in [0, 1/4]$:

$$
\bar{Q}((1,2), \cdot) = \begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & \rho & 1/2 - \rho \\
\hline
2 & 1/2 - \rho & \rho \\
\hline
 & 1 & 2
\end{array} \leftarrow y
\qquad
\bar{P}((1,2), \cdot) = \begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & \lambda & 1/4 - \lambda \\
\hline
2 & 1/2 - \lambda & \lambda + 1/4 \\
\hline
 & 1 & 2
\end{array} \leftarrow y
$$

Definition 2.3, on coupled acceptance mechanisms, raises two questions: whether its Condition 3 is independent of its Conditions 1 and 2, and whether a given $\bar{P} \in \Gamma(P, P)$ can be related to $\bar{Q} \in \Gamma(Q, Q)$ by more than one coupled acceptance mechanism. We answer both questions in the affirmative.

Suppose that we have the following proposal and transition kernel couplings, corresponding to the case of $\rho = 1/4$ and $\lambda = 0$ above:

$$
\bar{Q}((1,2), \cdot) = \begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & 1/4 & 1/4 \\
\hline
2 & 1/4 & 1/4 \\
\hline
 & 1 & 2
\end{array} \leftarrow y
\qquad
\bar{P}((1,2), \cdot) = \begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & 0 & 1/4 \\
\hline
2 & 1/2 & 1/4 \\
\hline
 & 1 & 2
\end{array} \leftarrow y
$$

By Lemma 2.5, we observe that the first two conditions of Definition 2.3 require any coupled acceptance mechanism Φ relating $\bar{Q}$ and $\bar{P}$ to take the following form for real numbers a, b, c, d, and e:

$$
\Phi_{11}((1,2), \cdot) = \begin{array}{c|c|c|}
1 & 0 & 1/4 \\
\hline
2 & a & b \\
\hline
 & 1 & 2
\end{array}
\qquad
\Phi_{10}((1,2), \cdot) = \begin{array}{c|c|c|}
1 & d & 0 \\
\hline
2 & c & 1/4 - b \\
\hline
 & 1 & 2
\end{array}
$$

$$
\Phi_{01}((1,2), \cdot) = \begin{array}{c|c|c|}
1 & 0 & 0 \\
\hline
2 & e & 0 \\
\hline
 & 1 & 2
\end{array}
\qquad
\Phi_{00}((1,2), \cdot) = \begin{array}{c|c|c|}
1 & 1/4 - d & 0 \\
\hline
2 & 1/4 - a - c - e & 0 \\
\hline
 & 1 & 2
\end{array}
$$

Thus $(\Phi_{11} + \Phi_{10})((x, y), \{x\} \times \mathcal{X}) = a + c + d$, and $(\Phi_{11} + \Phi_{01})((x, y), \mathcal{X} \times \{y\}) = a + b + e$. Observe that $a + c + d = Q(x, \{x\}) = 1/2$ and $a + b + e = Q(y, \{y\}) = 1/2$ do not automatically hold without imposing the third condition of Definition 2.3. Therefore, this condition is not implied by the other two. Also note that even with that condition, any choice of a, b, c with $0 \le c \le b \le 1/4$ and $0 \le a \le 1/4 - b \vee c$ will produce a valid Φ relating $\bar{Q}$ and $\bar{P}$. Thus we see that more than one coupled acceptance mechanism can relate the same proposal coupling $\bar{Q}$ and transition kernel coupling $\bar{P}$. Consequently, there can be more than one choice of acceptance indicator coupling $\bar{B}$ such that $\bar{Q}$ and $\bar{B}$ generate $\bar{P}$.

We continue with our two-state example to demonstrate the construction of a coupled acceptance mechanism Φ and proposal coupling $\bar{Q}$ as in the proof of Lemma 2.8. Under the assumptions above, $\alpha_0, \alpha_1, \beta, \mu$ will take the following values:

$$\alpha_0(1, \cdot) = (0, 0) \qquad \alpha_1(1, \cdot) = (1/2, 1/2) \qquad \mu(1, \cdot) = (1, 0)$$

$$\alpha_0(2, \cdot) = (1/4, 0) \qquad \alpha_1(2, \cdot) = (1/4, 1/2) \qquad \mu(2, \cdot) = (1, 0)$$

$$\beta(\cdot) = (1, 2/3).$$

We continue to use a transition kernel coupling $\bar{P}$ with the following values:

$$
\bar{P}((1,2), \cdot) = \begin{array}{c|c|c|}
 & \multicolumn{2}{c}{\overset{x}{\downarrow}} \\
\hline
1 & 0 & 1/4 \\
\hline
2 & 1/2 & 1/4 \\
\hline
 & 1 & 2
\end{array} \leftarrow y
$$

Following the construction in the proof of Lemma 2.8 and the values of α_0, α_1, μ, and β shown above, we have

$$\Phi_{11}((1,2),\cdot) = \begin{array}{c|c|c|} & 1 & 2 \\ \hline 1 & 0 & 1/4 \\ \hline 2 & 1/3 & 1/6 \\ \hline \end{array} \qquad \Phi_{10}((1,2),\cdot) = \begin{array}{c|c|c|} & 1 & 2 \\ \hline 1 & 1/6 & 1/12 \\ \hline 2 & 0 & 0 \\ \hline \end{array}$$

Similar calculations show that both $\Phi_{01}((1,2),\cdot)$ and $\Phi_{00}((1,2),\cdot)$ consist entirely of zeros. Finally, since $\bar{Q}((x,y),\cdot) = (\Phi_{11} + \Phi_{10} + \Phi_{01} + \Phi_{00})((x,y),\cdot)$, we have

$$\bar{Q}((1,2),\cdot) = \begin{array}{cc} & \begin{array}{cc} & x \\ & \downarrow \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{array}{|c|c|} \hline 1/6 & 1/3 \\ \hline 1/3 & 1/6 \\ \hline \end{array} \leftarrow y \\ & \begin{array}{cc} 1 & 2 \end{array} \end{array}$$

We see that this proposal coupling has the marginal distributions $Q(1,\cdot) = Q(2,\cdot) = (1/2, 1/2)$, as desired.

Finally, we derive an acceptance indicator coupling $\bar{B}$ such that it and $\bar{Q}$ will generate the given $\bar{P}$. We proceed according to the construction in the proof of Lemma 2.9. There, we defined the acceptance indicator coupling so that if $(b_x, b_y) \sim \bar{B}((x,y),(x',y'))$, then for $i, j \in \{0,1\}$,

$$\mathbb{P}\big(b_x = i, b_y = j \,|\, x, y, x', y'\big) = \phi_{ij}\big((x,y),(x',y')\big) = \frac{d\Phi_{ij}((x,y),\cdot)}{d\bar{Q}((x,y),\cdot)}(x', y').$$

Due to the discrete state space, we can evaluate this Radon–Nikodym derivative by dividing the values of $\Phi_{ij}((x,y), (x',y'))$ by those of $\bar{Q}((x,y),(x',y'))$. This yields the following joint acceptance probabilities:

$$\mathbb{P}(b_x = b_y = 1 \,|\, x, y, \cdot) = \begin{array}{c|c|c|} & 1 & 2 \\ \hline 1 & 0 & 3/4 \\ \hline 2 & 1 & 1 \\ \hline \end{array} \qquad \mathbb{P}(b_x = 1, b_y = 0 \,|\, x, y, \cdot) = \begin{array}{c|c|c|} & 1 & 2 \\ \hline 1 & 1 & 1/4 \\ \hline 2 & 0 & 0 \\ \hline \end{array}$$

$$\mathbb{P}(b_x = 0, b_y = 1 \,|\, x, y, \cdot) = \mathbb{P}(b_x = b_y = 0 \,|\, x, y, \cdot) = \begin{array}{c|c|c|} & 1 & 2 \\ \hline 1 & 0 & 0 \\ \hline 2 & 0 & 0 \\ \hline \end{array}$$

Thus we see that the methods of Section 2 provide a simple way to derive a proposal coupling $\bar{Q}$ and an acceptance coupling $\bar{B}$ which reproduce the given MH transition kernel coupling $\bar{P}$.

A.2. *Maximal kernel coupling generated by non-maximal proposal coupling*

We conclude with a slightly more complicated example to demonstrate the properties of maximal couplings of the MH transition kernel. Assume $\mathcal{X} = \{1, 2, 3\}$, $\mathcal{F} = 2^{\mathcal{X}}$, and $(x, y) = (1, 2)$. From these states, we assume the following proposal and transition kernel distributions:

$$Q(x = 1, \cdot) = (0, 1/2, 1/2) \qquad P(x = 1, \cdot) = (1/2, 1/2, 0)$$
$$Q(y = 2, \cdot) = (1/2, 0, 1/2) \qquad P(y = 2, \cdot) = (1/2, 0, 1/2).$$

It is easy to confirm that these distributions correspond to an MH transition kernel obtained when $Q(3, \cdot) = (0, 1, 0)$ and $\pi = (2/5, 2/5, 1/5)$. Simple algebra shows that any coupling in $\Gamma^{\max}(Q, Q)$ and any coupling in $\Gamma^{\max}(P, P)$ must have the following properties:

$$\bar{Q}((1,2),\cdot) = \begin{array}{cc} & \begin{array}{ccc} & x & \\ & \downarrow & \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{array}{|c|c|c|} \hline 0 & 1/2 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 0 & 1/2 \\ \hline \end{array} \leftarrow y \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array} \qquad \bar{P}((1,2),\cdot) = \begin{array}{cc} & \begin{array}{ccc} & x & \\ & \downarrow & \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{array}{|c|c|c|} \hline 1/2 & 0 & 0 \\ \hline 0 & 0 & 0 \\ \hline 0 & 1/2 & 0 \\ \hline \end{array} \leftarrow y \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \end{array}$$

There exists no indicator coupling $\bar{B}$ such that it and the unique maximal $\bar{Q}$ given above generate $\bar{P}$, since the combination of $(x', y') \sim \bar{Q}((1, 2), \cdot), (b_x, b_y) \sim \bar{B}((1, 2), (x', y')), (X, Y) = (b_x x' + (1 - b_x)x, b_y y' + (1 - b_y)y)$, and $(X, Y) \sim \bar{P}((1, 2), \cdot)$ yield a contradiction:

$$1/2 = \mathbb{P}(X = 2, Y = 3|x = 1, y = 2) \le \mathbb{P}(x' = 2, y' = 3|x = 1, y = 2) = 0.$$

Note that in line with Theorem 2.2, we can still generate $\bar{P}((1, 2), \cdot)$ using the following non-maximal proposal coupling and the coupled acceptance indicators b_x, b_y:

$$
\tilde{Q}((1, 2), \cdot) =
\begin{array}{c}
\\
1 \\
2 \\
3 \\
\\
\end{array}
\begin{array}{|c|c|c|}
\hline
0 & 0 & 1/2 \\
\hline
0 & 0 & 0 \\
\hline
0 & 1/2 & 0 \\
\hline
\end{array}
$$

with columns $1, 2, 3$ ($x \downarrow$) and rows $1, 2, 3$ ($\leftarrow y$)

$$\mathbb{P}(b_x = 1, b_y = 1|(x, y) = (1, 2), (x', y') = (2, 3)) = 1$$

$$\mathbb{P}(b_x = 0, b_y = 1|(x, y) = (1, 2), (x', y') = (3, 1)) = 1.$$

A qualitative conclusion from Lemma 3.3 and the example above is that maximal couplings $\bar{P} \in \Gamma^{\max}(P, P)$ may require a certain amount of proposal probability on the diagonal, but the maximality of $\bar{Q}$ is neither necessary nor sufficient for $\bar{Q}$ to be able to generate $\bar{P}$.

Acknowledgements

Funding

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82* 243–297. Springer, Berlin, 1983. MR0770418 https://doi.org/10.1007/BFb0068322

[2] D. Aldous and J. Fill. *Reversible Markov Chains and Random Walks on Graphs*. University of California Press, Berkeley, 1995.

[3] C. D. Aliprantis and O. Burkinshaw. *Principles of Real Analysis*. Gulf Pub., Houston, 1998. MR1669668

[4] C. Andrieu, A. Doucet and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** (3) (2010) 269–342. MR2758115 https://doi.org/10.1111/j.1467-9868.2009.00736.x

[5] C. Andrieu, A. Lee and S. Livingstone. A general perspective on the Metropolis-Hastings kernel, 2020. Preprint. Available at arXiv:2012.14881.

[6] Y. F. Atchadé and F. Perron. On the geometric ergodicity of Metropolis-Hastings algorithms. *Statistics* **41** (1) (2007) 77–84. MR2303970 https://doi.org/10.1080/10485250601033214

[7] K. B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** (1978) 493–501. MR0511425 https://doi.org/10.2307/1998882

[8] S. Banerjee and W. S. Kendall. Rigidity for Markovian maximal couplings of elliptic diffusions. *Probab. Theory Related Fields* **168** (1–2) (2017) 55–112. MR3651049 https://doi.org/10.1007/s00440-016-0706-4

[9] A. A. Barker. Monte Carlo calculations of the radial distribution functions for a proton–electron plasma. *Aust. J. Phys.* **18** (2) (1965) 119–134.

[10] G. Birkhoff. Three observations on linear algebra. *Univ. Nac. Tacuman, Rev. Ser. A* **5** (1946) 147–151. MR0020547

[11] N. Biswas, P. E. Jacob and P. Vanetti. Estimating convergence of Markov chains with L-lag couplings. In *Advances in Neural Information Processing Systems* 7391–7401, 2019.

[12] B. Böttcher Markovian maximal coupling of Markov processes. Preprint, 2017. Available at arXiv:1710.09654.

[13] N. Bou-Rabee, A. Eberle and R. Zimmer. Coupling and convergence for Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **30** (3) (2020) 1209–1250. MR4133372 https://doi.org/10.1214/19-AAP1528

[14] S. Boyd, P. Diaconis, P. Parrilo and L. Xiao. Fastest mixing Markov chain on graphs with symmetries. *SIAM J. Optim.* **20** (2) (2009) 792–819. MR2515797 https://doi.org/10.1137/070689413

[15] S. Boyd, P. Diaconis, J. Sun and L. Xiao. Fastest mixing Markov chain on a path. *Amer. Math. Monthly* **113** (1) (2006) 70–74. MR2202924 https://doi.org/10.2307/27641840

[16] S. Boyd, P. Diaconis and L. Xiao. Fastest mixing Markov chain on a graph. *SIAM Rev.* **46** (4) (2004) 667–689. MR2124681 https://doi.org/10.1137/S0036144503423264

[17] K. Burdzy and W. S. Kendall. Efficient Markovian couplings: Examples and counterexamples. *Ann. Appl. Probab.* (2000) 362–409. MR1768241 https://doi.org/10.1214/aoap/1019487348

[18] S. Connor and S. Jacka. Optimal co-adapted coupling for the symmetric random walk on the hypercube. *J. Appl. Probab.* **45** (3) (2008) 703–713. MR2455179 https://doi.org/10.1239/jap/1222441824

[19] W. Doeblin. Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Mathématique de l'Union Interbalkanique* **2** (77–105) (1938) 78–80.

[20] R. Douc, E. Moulines, P. Priouret and P. Soulier. *Markov Chains*. Springer, Berlin, 2018. MR3889011 https://doi.org/10.1007/978-3-319-97704-1

[21] J. H. Dshalalow. *Foundations of Abstract Analysis*. Springer, Berlin, 2012. MR3012328 https://doi.org/10.1007/978-1-4614-5962-0

[22] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B* **195** (2) (1987) 216–222. MR3960671 https://doi.org/10.1016/0370-2693(87)91197-x

[23] R. Dudley. *Real Analysis and Probability*. Cambridge University Press, New York, 2002. MR1932358 https://doi.org/10.1017/CBO9780511755347

[24] P. A. Ernst, W. S. Kendall, G. O. Roberts and J. S. Rosenthal. MEXIT: Maximal un-coupling times for stochastic processes. *Stochastic Process. Appl.* **129** (2) (2019) 355–380. MR3907003 https://doi.org/10.1016/j.spa.2018.03.001

[25] J. A. Fill. An interruptible algorithm for perfect sampling via Markov chains. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* 688–695, 1997. MR1753391

[26] J. M. Flegal and R. Herbei. Exact sampling for intractable probability distributions via a Bernoulli factory. *Electron. J. Stat.* **6** (2012) 10–37. MR2879671 https://doi.org/10.1214/11-EJS663

[27] A. E. Gelfand and A. F. Smith. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** (410) (1990) 398–409. MR1141740

[28] A. Gelman, S. Brooks, G. Jones and X. Meng. *Handbook of Markov Chain Monte Carlo: Methods and Applications. Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. CRC Press, Boca Raton, 2010. MR2742422 https://doi.org/10.1201/b10905

[29] P. W. Glynn and C.-H. Rhee. Exact estimation for Markov chain equilibrium expectations. *J. Appl. Probab.* **51** (A) (2014) 377–389. MR3317370 https://doi.org/10.1239/jap/1417528487

[30] S. Goldstein. Maximal coupling. *Probab. Theory Related Fields* **46** (2) (1979) 193–204. MR0516740 https://doi.org/10.1007/BF00533259

[31] J. B. Goodman and K. K. Lin. Coupling control variates for Markov chain Monte Carlo. *J. Comput. Phys.* **228** (19) (2009) 7127–7136. MR2568586 https://doi.org/10.1016/j.jcp.2009.03.043

[32] D. Griffeath. A maximal coupling for Markov chains. *Probab. Theory Related Fields* **31** (2) (1975) 95–106. MR0370771 https://doi.org/10.1007/BF00539434

[33] P. Gustafson. A guided walk Metropolis algorithm. *Stat. Comput.* **8** (4) (1998) 357–364.

[34] T. E. Harris. On chains of infinite order. *Pacific J. Math.* **5** (Suppl. 1) (1955) 707–724. MR0075482

[35] T. P. Hayes and A. Sinclair. A general lower bound for mixing of single-site dynamics on graphs. In *46th Annual IEEE Symposium on Foundations of Computer Science* 511–520. IEEE, Pittsburgh, PA, USA, 2005. MR2326236 https://doi.org/10.1214/105051607000000104

[36] T. P. Hayes and E. Vigoda. A non-Markovian coupling for randomly sampling colorings. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings* 618–627. IEEE, Cambridge, MA, USA, 2003.

[37] J. Heng, J. Houssineau and A. Jasra. On unbiased score estimation for partially observed diffusions, 2021. Preprint. Available at arXiv:2105.04912.

[38] J. Heng and P. E. Jacob. Unbiased Hamiltonian Monte Carlo with couplings. *Biometrika* **106** (2) (2019) 287–302. MR3949304 https://doi.org/10.1093/biomet/asy074

[39] J. Heng, A. Jasra, K. J. Law and A. Tarakanov. On unbiased estimation for discretized models, 2021. Preprint. Available at arXiv:2102.12230.

[40] E. P. Hsu and K.-T. Sturm. Maximal coupling of Euclidean Brownian motions. *Commun. Math. Stat.* **1** (1) (2013) 93–104. MR3197874 https://doi.org/10.1007/s40304-013-0007-5

[41] P. E. Jacob, J. O'Leary and Y. F. Atchadé. Unbiased Markov chain Monte Carlo methods with couplings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** (3) (2020) 543–600. MR4112777 https://doi.org/10.1111/rssb.12336

[42] V. E. Johnson. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. *J. Amer. Statist. Assoc.* **91** (433) (1996) 154–166. MR1394069 https://doi.org/10.2307/2291391

[43] V. E. Johnson. A coupling-regeneration scheme for diagnosing convergence in Markov chain Monte Carlo algorithms. *J. Amer. Statist. Assoc.* **93** (441) (1998) 238–248. MR1614640 https://doi.org/10.2307/2669620

[44] G. L. Jones and J. P. Hobert. Honest exploration of intractable probability distributions via Markov chain. *Monte Carlo. Statistical Science* (2001) 312–334. MR1888447 https://doi.org/10.1214/ss/1015346317

[45] M. Kartashov and V. Golomozyi. Maximal coupling procedure and stability of discrete Markov chains. II. *Theory Probab. Math. Statist.* **87** (2013) 65–78. MR3241447 https://doi.org/10.1090/S0094-9000-2014-00905-9

[46] W. S. Kendall. Coupling, local times, immersions. *Bernoulli* **21** (2) (2015) 1014–1046. MR3338655 https://doi.org/10.3150/14-BEJ596

[47] W. S. Kendall. *Lectures on Probabilistic Coupling*, 2017. MR3338655 https://doi.org/10.3150/14-BEJ596

[48] K. Kuwada. Characterization of maximal Markovian couplings for diffusion processes. *Electron. J. Probab.* **14** (2009) 633–662. MR2486817 https://doi.org/10.1214/EJP.v14-634

[49] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*, **107**. Am. Statist. Assoc., Alexandria, 2017. MR3726904 https://doi.org/10.1090/mbk/107

[50] T. Lindvall. *Lectures on the Coupling Method*. Dover, New York, 1992. MR1180522

[51] J. S. Liu, F. Liang and W. H. Wong. The multiple-try method and local optimization in Metropolis sampling. *J. Amer. Statist. Assoc.* **95** (449) (2000) 121–134. MR1803145 https://doi.org/10.2307/2669532

[52] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** (6) (1953) 1087–1092.

[53] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, Berlin, 2012. MR2509253 https://doi.org/10.1017/CBO9780511626630

[54] L. Middleton, G. Deligiannidis, A. Doucet and P. E. Jacob. Unbiased smoothing using particle independent Metropolis-Hastings. In *Proceedings of the 22th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research. PMLR* 16–18, 2019.

[55] L. Middleton, G. Deligiannidis, A. Doucet and P. E. Jacob. Unbiased Markov chain Monte Carlo for intractable target distributions. *Electron. J. Stat.* **14** (2) (2020) 2842–2891. MR4132645 https://doi.org/10.1214/20-EJS1727

[56] W. K. H. Monte. Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1) (1970) 97–109. MR3363437 https://doi.org/10.1093/biomet/57.1.97

[57] R. Neal and R. Pinto. Improving Markov chain Monte Carlo estimators by coupling to an approximating chain. Technical report, Department of Statistics, University of Toronto, 2001.

[58] R. M. Neal. Bayesian learning via stochastic dynamics. In *Advances in Neural Information Processing Systems* 475–482, 1993.

[59] R. M. Neal. Circularly-coupled Markov chain sampling. Technical report, Department of Statistics, University of Toronto, 1999.

[60] R. M. Neal. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo* 2, **2**, 2011. MR2858447

[61] E. Nummelin. Uniform and ratio limit theorems for Markov renewal and semi-regenerative processes on a general state space. *Ann. Inst. Henri Poincaré Probab. Stat.* **14** (2) (1978) 119–143. MR0507729

[62] J. O'Leary, 2021. Couplings of the Random-Walk Metropolis algorithm. Preprint. Available at arXiv:2101.01790.

[63] J. O'Leary, G. Wang and P. E. Jacob. Maximal couplings of the Metropolis-Hastings algorithm. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics* 1225–1233. *Proceedings of Machine Learning Research* **130**. PMLR, Virtual Event, 2021.

[64] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60** (3) (1973) 607–612. MR0362823 https://doi.org/10.1093/biomet/60.3.607

[65] N. S. Pillai and A. Smith. Mixing times for a constrained Ising process on the two-dimensional torus at low density. *Ann. Inst. Henri Poincaré Probab. Stat.* **55** (2019) 1649–1678.

[66] D. Piponi, M. Hoffman and P. Sountsov. Hamiltonian Monte Carlo swindles. In *International Conference on Artificial Intelligence and Statistics* 3774–3783. PMLR, Palermo, Sicily, Italy, 2020.

[67] J. Pitman. On coupling of Markov chains. *Probab. Theory Related Fields* **35** (4) (1976) 315–322. MR0415775 https://doi.org/10.1007/BF00532957

[68] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** (1–2) (1996) 223–252. MR1611693 https://doi.org/10.1002/(SICI)1098-2418(199608/09)9:1/2<223::AID-RSA14>3.3.CO;2-R

[69] Q. Qin and J. P. Hobert. Wasserstein-based methods for convergence complexity analysis of MCMC with applications. *Ann. Appl. Probab.* (2021). MR4386523 https://doi.org/10.1214/21-aap1673

[70] A. Reutter and V. E. Johnson. General strategies for assessing convergence of MCMC algorithms using coupled sample paths. Technical report, Institute of Statistics and Decision Sciences, Duke University, 1995.

[71] G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** (1) (1998) 255–268. MR1625691 https://doi.org/10.1111/1467-9868.00123

[72] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** (4) (1996) 341–363. MR1440273 https://doi.org/10.2307/3318418

[73] J. S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **90** (430) (1995) 558–566. MR1340509

[74] J. S. Rosenthal. Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Stat. Comput.* **6** (3) (1996) 269–275.

[75] J. S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electron. Commun. Probab.* **7** (2002) 123–128. MR1917546 https://doi.org/10.1214/ECP.v7-1054

[76] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability* (1998) 1–9. MR1620401 https://doi.org/10.1214/aoap/1027961031