SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 7, No. 1, pp. 324–382

© 2019 Society for Industrial and Applied Mathematics
and American Statistical Association

Constructing Sampling Schemes via Coupling: Markov Semigroups and Optimal Transport*

N. Nüsken† and G. A. Pavliotis‡

Abstract. In this paper we develop a general framework for constructing and analyzing coupled Markov chain Monte Carlo samplers, allowing for both (possibly degenerate) diffusion and piecewise deterministic Markov processes. For many performance criteria of interest, including the asymptotic variance, the task of finding efficient couplings can be phrased in terms of problems related to optimal transport theory. We investigate general structural properties, proving a singularity theorem that has both geometric and probabilistic interpretations. Moreover, we show that those problems can often be solved approximately and support our findings with numerical experiments. For the particular objective of estimating the variance of a Bayesian posterior, our analysis suggests using novel techniques in the spirit of antithetic variates. Addressing the convergence to equilibrium of coupled processes we furthermore derive a modified Poincaré inequality.

Key words. sampling, optimal transport, particle methods, Markov semigroups, MCMC

AMS subject classifications. 60J35, 47D07, 35Q62, 65C05, 65C35

DOI. 10.1137/18M119896X

1. Introduction and motivation.

Many computational problems arising in Bayesian statistics, machine learning, molecular dynamics, and various other fields require the approximation of probability distributions (in the following denoted by π) on a high-dimensional space E. In particular, uncertainty quantification in a Bayesian framework is intimately related to the evaluation of appropriate summary statistics such as the variance of the posterior [37, Chapter 10; 87, Chapter 8; 88, Chapter 6].Often, this task is approached by considering empirical measures associated to an ensemble of n particles, i.e., approximations of the form

$$\pi \approx \frac{1}{n}\sum_{i=1}^{n} \delta_{X^{(i)}} =: \tilde{\pi}, \tag{1}$$

where $X^{(i)}$ stands for the location of the ith particle and δ_x denotes the Dirac measure centered at $x \in E$. Usually, the particles are moved according to some (more often than not) stochastic dynamics, judiciously crafted in order for the empirical measure $\tilde{\pi}_t := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_t^{(i)}}$

*Received by the editors July 5, 2018; accepted for publication (in revised form) January 18, 2019; published electronically March 26, 2019.

http://www.siam.org/journals/juq/7-1/M119896.html

Funding: The work of the first author was supported by the EPSRC through a Roth Departmental Scholarship. The work of the second author was supported by the EPSRC through grants EP/P031587/1, EP/L024926/1, and EP/L020564/1.

†Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom. Current address: Institut für Mathematik, Universität Potsdam, D-14476 Potsdam, Germany (nuesken@uni-potsdam.de).

‡Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom (g.pavliotis@imperial.ac.uk).

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

to approach π when t reaches a terminal value (finite or infinite). This methodology has been particularly influential in statistical inference of hidden-state Markov models (stochastic filtering or sequential Monte Carlo; see for instance [80, 26] and references therein). Ensemble based methods have also been employed in the contexts of optimization [78, 81], molecular dynamics [85], Markov chain Monte Carlo (MCMC) [59, 71], or variational Bayesian inference [64]. Let us also mention the works [2] and [44], combining different aspects of various sampling strategies. The increasing availability of parallel-processing computational architectures has further encouraged the development and analysis of similar methodologies.

From an abstract perspective, many of the aforementioned algorithms targeting a probability measure π on some state space E naturally produce probability measures $\bar{\pi}$ on the product space $\bar{E} = \prod_{i=1}^{n} E_i$, where E_i is an identical copy of E, standing for the state space of the ith particle. Denoting by $P_i : \mathcal{P}(\bar{E}) \to \mathcal{P}(E_i)$ the mappings that send probability measures on $\bar{E}$ to their marginals on E_i, one then obtains the measure

$$\frac{1}{n}\sum_{i=1}^{n} P_i(\bar{\pi}) \tag{2}$$

as an approximation for π. Clearly, the map

$$\Pi : \mathcal{P}(\bar{E}) \to \mathcal{P}(E), \quad \bar{\pi} \mapsto \frac{1}{n}\sum_{i=1}^{n} P_i(\bar{\pi}) \tag{3}$$

is far from injective since $\Pi(\bar{\pi})$ only depends on the marginals of $\bar{\pi}$. This viewpoint shows that there is a considerable flexibility when generating the joint measure $\bar{\pi}$, immediately suggesting fruitful connections to the theory of couplings of probability measures [63, 90] prominently encountered, for instance, in relation to optimal transport problems [92, 94] or decay estimates in Wasserstein distances (see, for instance, [32]). Since in applications $\Pi(\bar{\pi})$ is only an approximation of the target measure of interest, the freedom to design appropriate couplings can be used to suppress bias, variance, and discretization errors. This general idea has proved to be very versatile, leading to powerful simulation techniques such as multilevel Monte Carlo [38], coupling from the past [79], and antithetic variates [54, section 9.2].

1.1. Couplings and MCMC.

In this paper we focus on coupling techniques in the context of MCMC simulations. Assume that we are interested in computing the expectation

$$\mathbb{E}_\pi f = \int_E f \mathrm{d}\pi \tag{4}$$

of a given test function (henceforth called *observable*) $f : E \to \mathbb{R}$ with respect to some probability measure π on E. As approximations relying on quadratures tend to be computationally infeasible in high dimensions, a standard approach is to construct a Markov process $(X_t)_{t\geq 0}$ on E such that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(X_s)\,\mathrm{d}s = \int_E f \mathrm{d}\pi, \tag{5}$$

i.e., the process $(X_t)_{t\geq 0}$ is supposed to be ergodic with respect to π. More generally, one often constructs a Markov process $(\bar{X}_t)_{t\geq 0}$ on an extended state space $\bar{E}$, ergodic with respect to a measure $\bar{\pi}$ that has π as its marginal,

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

$$\int \bar{\pi}(x,y)\,\mathrm{d}y = \pi(x), \tag{6}$$

where $(x,y) \in \bar{E}$ and $x \in E$. This idea is used, for instance, in Hamiltonian Monte Carlo [72] or sampling schemes based on underdamped Langevin dynamics [60, Chapter 2]. We also refer to the introduction in [29] for a more general perspective. In this work we take this approach further, in the sense that we consider extended measures $\bar{\pi}$ that have fixed marginals with respect to (multiple) complimentary subspaces of $\bar{E}$. Immediately, this viewpoint suggests fruitful connections to theory of optimal (multimarginal) transportation.

To explain our approach, let us consider n identical copies of E, $(X_t)_{t\geq 0}$, and π, denoted by E_i, $(X_t^i)_{t\geq 0}$, and π_i, for $i \in \{1,\dots,n\}$. Our main object of study is the class of Markovian couplings $(\bar{X}_t)_{t\geq 0}$ of $\{(X_t^i)_{t\geq 0} : i \in \{1,\dots,n\}\}$ on the product space $\bar{E} = \prod_{i=1}^n E_i$ that obey certain mild regularity assumptions. In particular, we characterize those couplings in terms of their infinitesimal generators in section 2 (see Proposition 8). One of the recurring themes of this work is the use of the latter in the analysis of coupled processes. From the coupling property of $(\bar{X}_t)_{t\geq 0}$ it follows immediately that if this process is ergodic, then its invariant measure (denoted by $\bar{\pi}$) is a coupling of the n copies of π.

For an observable $f \in L^1(\pi)$, we can define the extended observable $F : \bar{E} \to \mathbb{R}$ by

$$F(x_1,\dots,x_n) = \frac{1}{n}\sum_{i=1}^n f(x_i). \tag{7}$$

From (5) it is then immediate that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F(\bar{X}_t)\,\mathrm{d}t = \int_E f\,\mathrm{d}\pi, \tag{8}$$

i.e., the coupled process $(\bar{X}_t)_{t\geq 0}$ in conjunction with the observable (7) provides a valid sampling scheme. Let us remark that the framework we develop in section 2 accommodates the case when the spaces E_i, the processes $(X_t^i)_{t\geq 0}$ and the measures π_i are not identical, allowing for considerable flexibility in the construction of coupled samplers.

The study of observables of the form (7) provides a compelling dual perspective on the "sum of marginals" operator (3). Denoting by $\mathcal{B}_b(E)$ the space of bounded measureable functions we can consider the "extension operator"

$$\Pi^* : \mathcal{B}_b(E) \to \mathcal{B}_b(\bar{E}), \quad f \mapsto \frac{1}{n}\sum_{i=1}^n f_i, \tag{9}$$

provided by (7), using the notation $f_i(x_1,\dots,x_n) = f(x_i)$. For $\bar{\pi} \in \mathcal{P}(\bar{E})$ and $f \in \mathcal{B}_b(E)$ we clearly have $(\Pi\bar{\pi})(f) = \bar{\pi}(\Pi^* f)$, showing that understanding the class of observables given by (7) is sufficient for analyzing the properties of measures of the form (2). This idea features in particular in section 6 in the analysis of the exponential convergence to equilibrium for coupled processes.

Clearly, it is desirable to choose the coupling in such a way that the convergence in (8) is as fast as possible. Reasonable criteria involve the asymptotic variance (related to appropriate central limit theorems) and the spectral gap (related to the speed of convergence to equilbrium), both of which will be addressed in the present paper. We refer the reader to [29, section 1] for a more detailed discussion of these quantities.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Similar constructions to ours have been considered in the literature, in particular in a discrete time setting. In [36], the authors construct coupled Gibbs samplers using a very related rationale (see also [46] and [70]). Coupled Metropolis–Hastings samplers have been put forward in [21]. The work [22] provides a theoretical framework that, however, is quite different from the one developed in the present paper. Further algorithmic ideas related to coupled samplers can also be found in [57] and [83].

We would like to comment on the fact that a considerable amount of existing work proposes couplings that encourage multiple chains to coalesce as fast as possible, with high probability. The motivations for this include the possibility to implement convergence diagnostics [49, 50, 71], debiasing strategies [79, 40, 45], methods using control variates [73, 41], and convergence proofs [14]. In contrast to this, most of our coupling schemes obtained in section 5 are (in some form) aimed at steering the particles away from each other. The benefits of this might be twofold. Firstly, producing anticorrelated samples often results in cancellations, leading to a decrease in the variance of the corresponding estimator. This idea is developed thoroughly in sections 4 and 5. Secondly, increasing the average distance between particles might facilitate overall exploration of the state space. Experiments show that this is indeed possible and promising (see, for instance, Figure 5(c) for the case of two coupled zigzag samplers), but a more detailed theoretical understanding is yet to be developed.

1.2. Overview of the main results by means of a simple example. In this section we present our main findings informally by means of a very simple example, pointing to the exact statements in the forthcoming sections. Let us stress that our results hold in much greater generality, in particular also including the recently fashionable piecewise deterministic Markov processes (PDMPs).

Let us consider $n = 2$ particles (the locations of which are denoted by X_t and Y_t) moving each in one dimension according to the overdamped Langevin dynamics defined by the SDEs

$$\mathrm{d}X_t = -V'(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^x, \tag{10a}$$

$$\mathrm{d}Y_t = -V'(Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^y, \tag{10b}$$

where $V \in C^\infty(\mathbb{R})$ is a fixed potential such that

$$Z := \int_{\mathbb{R}} e^{-V}\mathrm{d}x < \infty, \tag{11}$$

and $(B_t^x)_{t\geq 0}$, $(B_t^y)_{t\geq 0}$ denote standard one-dimensional Brownian motions. As is wellknown, each of these processes considered separately is ergodic with respect to $\pi = \frac{1}{Z}e^{-V}\mathrm{d}x$, i.e., (5) holds for an appropriate class of observables. Note that we have deliberately refrained from stating that the Brownian motions $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$ are independent. Indeed, notwithstanding any dependence between these, it is immediate that (8) holds for the extended observable $F(x, y) = \frac{1}{2}(f(x) + f(y))$, as defined in (7). One of the main objectives of our analysis is to find couplings between $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$ such that the induced joint process $(X_t, Y_t)_{t\geq 0}$ has favorable properties, in terms of the asymptotic variance associated to (8) as well as in terms of convergence to equilibrium of $\Pi(\bar{\pi}_t)$, where $\bar{\pi}_t$ denotes the joint law of (X_t, Y_t) and Π has been defined in (3). The dependence between $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$ can be conveniently encoded in a suitable matrix-valued function $G : \mathbb{R}^2 \to \mathbb{R}^{2\times 2}$, writing

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$\mathrm{d}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} -V'(X_t) \\ -V'(Y_t) \end{pmatrix} \mathrm{d}t + \sqrt{2} G(X_t, Y_t) \begin{pmatrix} \mathrm{d}W_t^x \\ \mathrm{d}W_t^y \end{pmatrix} \tag{12}$$

for two independent Brownian motions $(W_t^x)_{t\geq 0}$ and $(W_t^y)_{t\geq 0}$. In this sense, the optimization problem alluded to above is naturally posed over an appropriate set of matrix-valued functions.

In section 2 we introduce the general framework, leading to a characterization of possible couplings in terms of infinitesimal generators of the dynamics. In the present example, the generators of the one-particle dynamics are given by

$$\mathcal{L}_x = -V'(x)\partial_x + \partial_x^2, \quad \mathcal{L}_y = -V'(y)\partial_y + \partial_y^2. \tag{13}$$

The generators of possible couplings $(X_t, Y_t)_{t\geq 0}$ turn out to be of the form

$$\bar{\mathcal{L}}_\Gamma := \mathcal{L}_x + \mathcal{L}_y + \Gamma, \quad \Gamma = 2\alpha \partial_x \partial_y, \tag{14}$$

where $\alpha : \mathbb{R}^2 \to [-1, 1]$ is a function with suitable regularity properties. The connection between α and G will be made precise in section 3. We use the term "coupling operator" when referring to Γ and denote the set of such operators by $\mathcal{G}$. Note that Γ as defined in (14) vanishes on functions that depend either only on x or only on y. In Proposition 8 we will see that this property essentially characterizes coupling operators in general. As it turns out (see the discussion in section 2.3), not every coupling of ergodic Markov processes is such that the joint process is ergodic. Hence, we introduce the subset $\mathcal{G}^0 \subset \mathcal{G}$ of ergodic coupling operators that do preserve ergodicity. In the present example, $\bar{\mathcal{L}}_\Gamma$ is elliptic whenever $-1 < \alpha < 1$ pointwise and therefore the corresponding coupling operators are ergodic. Intuitively, the nonergodic coupling operators in $\mathcal{G} \setminus \mathcal{G}^0$ can hence be thought of as lying "at the boundary" of $\mathcal{G}^0$. Although we have not been successful in proving a rigorous version of this statement in a general context, the reader is encouraged to keep this picture in mind.

On $\mathcal{G}^0$ we can consider the map $\Gamma \mapsto \bar{\pi}_\Gamma$, where $\bar{\pi}_\Gamma$ stands for the unique invariant measure associated with $\bar{\mathcal{L}}_\Gamma$. It is immediately clear from the construction that any $\bar{\pi}_\Gamma$ arising in this way is a coupling of π to itself (i.e., $\bar{\pi}_\Gamma$ has marginal π in both directions). We argue in section 4 that a wide range of optimization problems in our context can be cast in the following form, very closely linked to the theory of optimal transportation:

$$\min_{\Gamma \in \mathcal{G}^0} \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma, \tag{15}$$

where c is an appropriate cost function. Indeed, we show in section 4.1 that the task of optimizing the asymptotic variance of a coupled process with respect to a given observable is equivalent to (15), for a cost function that is constructed from the solution of a related Poisson equation. Addressing the problem (15), we first note that the dependence $\Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ is highly nonlinear; in particular, for $\Gamma \in \mathcal{G}^0$ and $\lambda \in [0, 1]$, the mapping $\lambda \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_{\lambda\Gamma}$ generally exhibits many local minima and maxima.[1] Nevertheless, we find that under suitable conditions the function $\Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ does not attain its extrema on interior points. This

[1] This claim is made assuming that $\lambda\Gamma \in \mathcal{G}^0$ for all $\lambda \in [0, 1]$.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

is the main result of section 4.2 and is stated rigorously in Theorem 49. In the example under consideration, this implies that optimal couplings necessarily satisfy $\|\alpha\|_\infty = 1$, leading to singular (i.e., degenerately elliptic) generators $\bar{\mathcal{L}}_\Gamma$. This conclusion is interesting in two respects: Firstly, it complements standard results from optimal transport theory showing that optimal couplings are typically singular in a certain sense. We stress, however, that the problem (15) is genuinely different from problems occurring in optimal transport theory and that our proof uses fundamentally different techniques. Secondly, this result supports the folklore that optimal MCMC samplers use the least amount of noise necessary to guarantee their ergodicity.

While the results from section 4 indicate the possible locations of optimal coupling operators Γ in the set $\mathcal{G}$, they do not help to actually find them in practice. In section 5 we address this problem by considering small perturbations around the trivial coupling $\mathbf{0} \in \mathcal{G}$ corresponding to independent Brownian motions. This leads to a much more tractable optimization problem that can be solved explicitly in concrete examples and gives promising results in our numerical experiments. In the present example, "mirror coupling" ($B_t^x = -B_t^y$) turns out to be optimal in terms of reducing the asymptotic variance of monotone observables, in the sense of the optimization problem just referred to. However, for different observables (perhaps exhibiting other types of symmetries) more intricate coupling strategies turn out to be advisable. We wish to stress that those observables are of particular relevance for the quantification of uncertainty in a Bayesian framework, for instance, in the computation of the variance or related quantities of a posterior distribution.

In section 6 we analyze the rate of convergence to equilibrium for coupled processes. As we will see, the former can be characterized in terms of an inequality of Poincaré type that is in turn related to an appropriate Hilbert space constructed in terms of the coupling. Applied to the present example, this result shows that the rate of convergence can be improved relative to the one-particle dynamics if the potential V is symmetric, i.e., $V(x) = V(-x)$. In general, the speed of convergence to equilibrium can also be slower, in the sense that there might appear a constant $C > 1$ in front of the exponential decay estimate. We leave a more detailed exploration of this phenomenon for future study.

The structure of the paper is as follows: In section 2, we introduce our framework in a general setting. In particular, we fix the notation (section 2.1), characterize coupled processes in terms of their generators (section 2.2), discuss ergodic properties (section 2.3), and provide a means of construction coupling operators given the generators of the marginal processes (section 2.4). In section 3, we illustrate our theory with concrete examples, namely diffusion processes such as overdamped and underdamped Langevin dynamics (see sections 3.1 and 3.2, respectively), as well as the zigzag process (see section 3.3), representing the class of PDMPs. In section 4.1 we derive a central theorem for coupled processes. The ensuing expression for the asymptotic variance is connected to the theory of optimal transportation, as exhibited and analyzed in section 4.2. In section 5 we take a perturbative approach towards the solutions of the aforementioned optimal transport problems and exemplify our results in the context of the examples presented in section 3. Finally, in section 6 we analyze the convergence of coupled processes to equilibrium relying on a suitable functional inequality of Poincaré type. The appendix comprises additional material required for some of the proofs throughout the article.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. Coupled processes and coupling operators. This section is devoted to the interplay between couplings of Markov processes and their infinitesimal generators. We start by specifying the setting and notations.

2.1. Preliminaries, notation, and setting.

2.1.1. Feller semigroups. For a given locally compact Polish space E we will denote the space of bounded, Borel measurable functions by $\mathcal{B}_b(E)$, the space of bounded continuous functions by $C_b(E)$, and the space of continuous functions vanishing at infinity[2] by $C_0(E)$. All of theses spaces become Banach spaces when equipped with the supremum norm, denoted by $\|\cdot\|_\infty$. The space of probability measures on E (equipped with the Borel σ-algebra $\mathcal{B}(E)$) will be denoted by $\mathcal{P}(E)$. An E-valued Markov process $(X_t)_{t\geq 0}$ induces a semigroup of linear operators $(S_t)_{t\geq 0}$ on $\mathcal{B}_b(E)$ via

$$(S_t f)(x) = \mathbb{E}[f(X_t)|X_0 = x], \quad f \in \mathcal{B}_b(E),\, x \in E. \tag{16}$$

Since the terminology varies slightly across the literature, we next give the definition of Feller processes used in this paper, mostly adopting the notations and conventions from [13, Chapter 1]. For more details we furthermore refer to [51, Chapter 17].

Definition 1 (Feller processes). *A Markov process $(X_t)_{t\geq 0}$ satisfies the* Feller property *if the following hold for the corresponding semigroup $(S_t)_{t\geq 0}$:*

1. *$(S_t)_{t\geq 0}$ leaves $C_0(E)$ invariant, i.e., $S_t f \in C_0(E)$ for all $f \in C_0(E)$ and $t \geq 0$.*
2. *$(S_t)_{t\geq 0}$ is strongly continuous on $C_0(E)$, i.e.,*

$$\|S_t f - f\|_\infty \xrightarrow{t\to 0} 0 \tag{17}$$

for all $f \in C_0(E)$.

Provided that $(S_t)_{t\geq 0}$ is a Feller semigroup as specified above, we define its generator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$ in the usual way [33, Chapter 2]. Throughout this paper, we will assume for convenience that the state space E has a differential structure such that the space $C_c^\infty(E)$ of compactly supported smooth functions is meaningfully defined. We can then make the following assumption on the domain of the generator $\mathcal{L}$:

Assumption 1. All considered Feller processes are rich,[3] i.e., $C_c^\infty(E) \subset \mathcal{D}(\mathcal{L})$.

The state spaces encountered in the examples in section 3 naturally admit differentiable structures, and the corresponding generators fulfill Assumption 1. Let us remark, however, that our framework can be extended to more general scenarios (including, for instance, infinite dimensional examples), replacing $C_c^\infty(E)$ by suitable function spaces adapted to the particular setting.

Remark 2. Clearly, $C_0(E)$ does not contain constant functions (apart from the zero function) if E is not compact. In preparation for Definition 6, we mention that $\mathcal{D}(\mathcal{L})$ can naturally

[2] Recall that a function $f : E_i \to \mathbb{R}$ vanishes at infinity if for all $\varepsilon > 0$ there exists a compact set $K \subset E_i$ such that $|f(x)| \leq \varepsilon$ for all $x \in E_i \setminus K$.

[3] We adopt this terminology, following, for instance, [55, section 1.5] and references therein.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

be extended to a subset of $C_b(E)$ by endowing the latter with the topology of uniform convergence on compact subsets of E. Following [86] (see also [48, section 4.8]), the extended generator $(\tilde{\mathcal{L}}, \mathcal{D}(\tilde{\mathcal{L}}))$ can then be defined by

$$\mathcal{D}(\tilde{\mathcal{L}}) = \left\{ f \in C_b(E) : \lim_{t\to 0} \frac{S_t f - f}{t} \quad \text{exists uniformly on compact sets} \right\}, \tag{18a}$$

$$\tilde{\mathcal{L}} f = \lim_{t\to 0} \frac{S_t f - f}{t}, \quad f \in \mathcal{D}(\tilde{\mathcal{L}}). \tag{18b}$$

Since $(S_t)_{t\geq 0}$ is conservative[4] we immediately see that $\mathbf{1} \in \mathcal{D}(\tilde{\mathcal{L}})$ and $\tilde{\mathcal{L}}\mathbf{1} = 0$, i.e., $\tilde{\mathcal{L}}$ vanishes on constant functions. Moreover, $(\tilde{\mathcal{L}}, \mathcal{D}(\tilde{\mathcal{L}}))$ is an extension of $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$, i.e., $\mathcal{D}(\mathcal{L}) \subset \mathcal{D}(\tilde{\mathcal{L}})$ and $\tilde{\mathcal{L}}|_{\mathcal{D}(\mathcal{L})} = \mathcal{L}$. Henceforth we will thus drop the tilde when no confusion is possible.

2.1.2. Product spaces. We will be dealing with a collection of locally compact Polish spaces E_i, indexed by $i \in \{1, \dots, n\}$, and denote their cartesian product by $\bar{E} := E_1 \times \cdots \times E_n$. For $f \in \mathcal{B}_b(E_i)$, it is of course understood that also $f \in \mathcal{B}_b(\bar{E})$, then depending only on the coordinate x_i in $\bar{x} \equiv (x_1, \dots, x_n)$. To a given function $f \in C_c^\infty(E_i)$ or $f \in C_0(E_i)$, we will also associate the canonical element in $\mathcal{B}_b(\bar{E})$, but we wish to emphasize that clearly f does not in general have compact support or does not vanish at infinity when considered as a function on $\bar{E}$. Frequently, the spaces E_i will be identical copies of each other, i.e., $\bar{E} = E^n$. Given $f \in \mathcal{B}_b(E)$, we will then write $f_i \in \mathcal{B}_b(\bar{E})$ for the function given by

$$f_i(x_1, \dots, x_n) = f(x_i), \quad (x_1, \dots, x_n) \in \bar{E}. \tag{19}$$

Sums of unbounded operators are defined in the usual way: For two operators $(A, \mathcal{D}(A))$ and $(B, \mathcal{D}(B))$ defined on the same Banach space X (i.e., $\mathcal{D}(A) \subset X$ and $\mathcal{D}(B) \subset X$), their sum is defined via

$$(A + B)f := Af + Bf, \quad f \in \mathcal{D}(A + B) := \mathcal{D}(A) \cap \mathcal{D}(B); \tag{20}$$

see, for instance, [33, Chapter III]. In the case when $(A, \mathcal{D}(A))$ and $(B, \mathcal{D}(B))$ are defined on two distinct spaces $\mathcal{B}_b(E_i)$ and $\mathcal{B}_b(E_j)$, $i \neq j$, (i.e., $\mathcal{D}(A) \subset \mathcal{B}_b(E_i)$ and $\mathcal{D}(B) \subset \mathcal{B}_b(E_j)$), their sum is defined as

$$A + B := A \otimes I + I \otimes B, \quad \mathcal{D}(A + B) := \mathcal{D}(A) \widehat{\otimes} \mathcal{D}(B), \tag{21}$$

where $\widehat{\otimes}$ denotes the canonical topological tensor product on $\mathcal{B}_b(E_i \times E_j)$ (see [1] and [4, A-I 3.7]).

2.2. Coupled processes. Assume that for $i \in \{1, \dots, n\}$, we are given locally compact Polish spaces E_i, representing the state spaces of n distinct particles. Furthermore, for $i \in \{1, \dots, n\}$, let us fix Feller semigroups $(S_t^i)_{t\geq 0}$ on E_i with generators $(\mathcal{L}_i, \mathcal{D}(\mathcal{L}_i))$ and associated Feller processes $(X_t^i)_{t\geq 0}$ on appropriate stochastic bases $(\Omega_i, \mathbb{P}_i, (\mathcal{F}_t^i)_{t\geq 0})$, representing the dynamics of those particles (in the following these processes will be referred to as the "one-particle dynamics"). Let us also assume that the spaces $C_c^\infty(E_i)$ are cores for the semigroups $(S_t^i)_{t\geq 0}$.

[4] Conservativeness of the semigroup $(S_t)_{t\geq 0}$ means that $S_t\mathbf{1} = \mathbf{1}$ for all $t \geq 0$, encoding the conservation of total probability mass.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 3. By Watanabe's theorem (see, for instance, [51, Proposition 17.9]), $C_c^\infty(E_i)$ is a core for $\mathcal{L}_i$ if it is dense in $\mathcal{D}(\mathcal{L}_i)$ and invariant under $(S_t^i)_{t\geq 0}$. It is possible to extend our framework by exchanging $C_c^\infty(E_i)$ for other cores, say, $\mathcal{D}_i$, as long as the first condition in Definition 6 is altered accordingly.

Consider now a Feller process $(\bar{X}_t)_{t\geq 0}$ on the product space $\bar{E} := E_1 \times \cdots \times E_n$, together with its associated semigroup $(\bar{S}_t)_{t\geq 0}$ on $\mathcal{B}_b(\bar{E})$ and generator $(\bar{\mathcal{L}}, \mathcal{D}(\bar{\mathcal{L}}))$ in $C_0(\bar{E})$. We will denote the E_i-valued coordinate processes of $(\bar{X}_t)_{t\geq 0}$ by $(\bar{X}_t^i)_{t\geq 0}$.

Definition 4 (Feller couplings). *The process $(\bar{X}_t)_{t\geq 0}$ is called a* Feller coupling *of the processes $(X_t^i)_{t\geq 0}$, if its marginals are given by these processes, i.e., if for all $i \in \{1, \ldots, n\}$, the processes $(\bar{X}_t^i)_{t\geq 0}$ and $(X_t^i)_{t\geq 0}$ induce the same law on the space of càdlàg functions $D([0,\infty), E_i)$.*[5]

Our aim in this section is to characterize the infinitesimal generators of such coupled processes.

Remark 5. We are making the following two assumptions when considering the class of processes described above. Firstly, we assume certain continuity properties of the process $(\bar{X}_t)_{t\geq 0}$, encoded mainly in the fact that the space $C_0(\bar{E})$ is invariant under the action of the corresponding semigroup (see, for instance, [13, Lemma 1.4] for more details). Restricting our attention to the class of Feller processes allows us to use the theory of strongly continuous semigroups on Banach spaces [33] for the development of the theory in this section. In examples and applications, however (see sections 3 and 5), we will relax this assumption a bit, allowing for more general processes.

Secondly, we consider processes $(\bar{X}_t)_{t\geq 0}$ that are Markovian. Obviously there are many non-Markovian couplings of the underlying processes $(X_t^i)_{t\geq 0}$, and indeed those might be of particular interest for applications. Hence we plan to investigate the possibility of extending our framework in this direction in a forthcoming project.

We now proceed to introduce a class of linear (unbounded) operators $(\Gamma, \mathcal{D}(\Gamma))$ on $\mathcal{B}_b(\bar{E})$.

Definition 6 (coupling operators). *Let $(\Gamma, \mathcal{D}(\Gamma))$ be a (possibly unbounded) linear operator on $\mathcal{B}_b(\bar{E})$. Then $(\Gamma, \mathcal{D}(\Gamma))$ is called a* coupling operator *if the following conditions are satisfied:*

1. *Test functions that depend on only one component of $\bar{x} = (x_1, \ldots, x_n)$ are in the kernel of Γ:*
   For all $i \in \{1, \ldots, n\}$ and $f \in C_c^\infty(E_i)$ it holds that $f \in \mathcal{D}(\Gamma)$ and

$$\Gamma f = 0. \tag{22}$$

2. *The operator*

$$\bar{\mathcal{L}}_\Gamma := \sum_{i=1}^n \mathcal{L}_i + \Gamma \tag{23}$$

with domain $\mathcal{D}(\bar{\mathcal{L}}_\Gamma) = \bigotimes_{i=1}^n \mathcal{D}(\mathcal{L}_i) \cap \mathcal{D}(\Gamma)$ is closable, and its closure is the infinitesimal generator of a Feller process on $\bar{E}$.

[5] Every Feller process has a càdlàg modification; see [13, Theorem 1.19].

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

The Feller semigroup corresponding to a coupling operator Γ *will be referred to by* $(\bar{S}_t^\Gamma)_{t\geq 0}$. *Furthermore, the set of coupling operators will be denoted by* $\mathcal{G}$, *i.e.,*

$$\mathcal{G} = \{\Gamma : \mathcal{D}(\Gamma) \subset \mathcal{B}_b(\bar{E}) \to \mathcal{B}_b(\bar{E}) : \textit{ conditions 1 and 2 are satisfied.}\}.$$

Remark 7. We will not distinguish (notationally) between $\bar{\mathcal{L}}_\Gamma$ and its closure. Notice also that the first condition in Definition 6 necessitates to think of Γ as an operator defined on (a subspace) of $C_b(\bar{E})$ (rather than $C_0(\bar{E})$), because of $C_c^\infty(E_i) \not\subset C_0(\bar{E})$. The second condition is naturally concerned with $\bar{\mathcal{L}}_\Gamma$ being the generator of a semigroup on $C_0(\bar{E})$ (and hence with the appropriate restriction of Γ). We refer to Remark 2 for a discussion about the extended generator on $C_b(\bar{E})$.

We have the following result, characterizing completely the set of rich Feller couplings in terms of the coupling operators $\mathcal{G}$.

Proposition 8. *For any* $\Gamma \in \mathcal{G}$, *the Feller process generated by* $\bar{\mathcal{L}}_\Gamma$ *as defined in* (23) *is a coupling of the processes* $((X_t^i)_{t\geq 0}, i \in \{1, \dots n\})$. *Conversely, if* $(\bar{X}_t)_{t\geq 0}$ *is a Feller coupling of the processes* $((X_t^i)_{t\geq 0}, i \in \{1, \dots n\})$, *then its generator is of the form* (23) *with* $\Gamma \in \mathcal{G}$.

Proof. Let $\Gamma \in \mathcal{G}$, and consider the process $(\bar{X}_t)_{t\geq 0}$ generated by the corresponding operator $\bar{\mathcal{L}}_\Gamma$ as defined in (23). Let us fix $i \in \{1, \dots, n\}$. Clearly, $C_c^\infty(E_i) \in \mathcal{D}(\bar{\mathcal{L}}_\Gamma)$ and $\bar{\mathcal{L}}_\Gamma f = \mathcal{L}_i f$ for $f \in C_c^\infty(E_i)$. Hence, for all $f \in C_c^\infty(E_i)$, the process

$$f(\bar{X}_t) - f(\bar{X}_0) - \int_0^t (\mathcal{L}_i f)(\bar{X}_s)\,\mathrm{d}s, \quad t \geq 0, \tag{24}$$

is a martingale with respect to the natural filtration $(\mathcal{F}_t^{\bar{X}})_{t\geq 0}$ generated by $(\bar{X}_t)_{t\geq 0}$. From the uniqueness of the martingale problem for the generator $\mathcal{L}_i$ (see, for instance, [34, section 4.4]) and the fact that $C_c^\infty(E_i)$ is a core for $\mathcal{L}_i$, it follows that $(\bar{X}_t^i)_{t\geq 0}$ has indeed the same law as $(X_t^i)_{t\geq 0}$.

Conversely, assume that $(\bar{X}_t)_{t\geq 0}$ is a Feller coupling of the processes $((X_t^i)_{t\geq 0}$, $i \in \{1, \dots n\})$, and denote its generator by $(\tilde{\mathcal{L}}, \mathcal{D}(\tilde{\mathcal{L}}))$. For fixed $i \in \{1, \dots, n\}$, we first argue that $C_c^\infty(E_i) \subset \mathcal{D}(\tilde{\mathcal{L}})$, referring to the domain of the extended generator defined in (18a). Indeed, this amounts to showing that for all $f \in C_c^\infty(E_i)$ the limit

$$\lim_{t\to 0} \frac{1}{t}(\tilde{S}_t f - f) \tag{25}$$

exists uniformly on compact sets. By the coupling property (Definition 4), we have that $\tilde{S}_t f = S_t^i f$ for $f \in C_c^\infty(E_i)$. Therefore (and since $C_c^\infty(E_i) \subset \mathcal{D}(\mathcal{L}_i)$ by assumption), it follows that the limit (25) even exists uniformly on the whole of $\bar{E}$.

We can now define $\tilde{\Gamma} := \tilde{\mathcal{L}} - \sum_{i=1}^n \mathcal{L}_i$ on $\mathcal{D}(\tilde{\Gamma}) := \mathcal{D}(\tilde{\mathcal{L}}) \cap \bigotimes_{i=1}^n \mathcal{D}(\mathcal{L}_i)$. It is then sufficient to show that Γ satisfies the first condition of Definition 6. To this end, take $f \in C_c^\infty(E_i)$ in the martingale problem for $\tilde{\mathcal{L}}$ to see that

$$f(\bar{X}_t^i) - f(\bar{X}_0^i) - \int_0^t (\mathcal{L}_i f)(\bar{X}_s^i)\,\mathrm{d}s - \int_0^t (\tilde{\Gamma} f)(\bar{X}_s)\,\mathrm{d}s, \quad t \geq 0, \tag{26}$$

is a martingale, again with respect to the natural filtration $(\mathcal{F}_t^{\bar{X}})_{t\geq 0}$ generated by $(\bar{X}_t)_{t\geq 0}$. Since $(\bar{X}_t^i)_{t\geq 0}$ and $(X_t^i)_{t\geq 0}$ are equal in law by assumption, it follows that $((\bar{X}_t^i)_{t\geq 0}, (\mathcal{F}_t^{\bar{X}})_{t\geq 0})$

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

is a solution to the martingale problem for $\mathcal{L}_i$. Hence, $\int_0^t (\tilde{\Gamma} f)(\bar{X}_s) \mathrm{d}s$ has to be a martingale as well. Since this process is of finite variation (and the initial condition for the process $(\bar{X}_t)_{t\geq 0}$ can be chosen arbitrarily), this implies $\tilde{\Gamma} f = 0$. ∎

Remark 9. Similar approaches, describing couplings in terms of coupling operators, are known from the literature. See, for instance, [18, Chapter 2, Definition 2.7] and references therein. The exact result of Proposition 8 and its proof using martingale problems seems to be new and in particular relevant for Conjecture 2.18 and Open Problem 2.19 in [18].

Example 10 (independent (or trivial) coupling). The zero operator $\Gamma = \mathbf{0}$ is always in $\mathcal{G}$, as the conditions of Definition 6 clearly hold. Indeed, consider the operator

$$\bar{\mathcal{L}}_0 := \sum_{i=1}^{n} \mathcal{L}_i \tag{27}$$

on the domain $\mathcal{D}(\bar{\mathcal{L}}_0) := \widehat{\bigotimes}_i \mathcal{D}(\mathcal{L}_i)$. It is straightforward (see, for instance, [4, A-I 3.7]) to show that $\bar{\mathcal{L}}_0$ is the generator of a Feller semigroup $(\bar{S}_t^0)_{t\geq 0}$ given by

$$\bar{S}_t^0 f = \left(\prod_{i=1}^{n} S_t^i \right) f, \quad f \in \mathcal{B}_b(\bar{E}),\, t \geq 0, \tag{28}$$

and that the associated Feller process is just $(\bar{X}_t^0)_{t\geq 0} = (X_t^1, \dots, X_t^n)_{t\geq 0}$, i.e., it is obtained from independent copies of the underlying processes.

Let us briefly discuss some of the implications of the conditions in Definition 6. As can be seen from the proof of Proposition 8, the first condition is instrumental in guaranteeing that the coupled process $(\bar{X}_t)_{t\geq 0}$ has the correct marginals. To put the second condition into context, we remark that generators of Feller semigroups can be characterized by means of the Hille–Yosida–Ray theorem in terms of the positive maximum principle (see [13, Lemma 1.28] and [13, Theorem 1.30]). As we will see in the examples in section 3, the latter often restricts the "size" of coupling operators, so that the set $\mathcal{G}$ usually turns out to be "bounded" in a certain sense. Let us close this section by mentioning the following conjecture.

Conjecture 11. The set $\mathcal{G}$ is convex.

Clearly, the first condition in Definition 6 is closed under convex combinations, so the challenge in proving the above conjecture rests in addressing the generator property required by the second condition (for instance, by appealing to the Hille–Yosida–Ray theorem [13, Theorem 1.30]). At this moment, we are not able to verify the third condition in the aforementioned reference, in particular due to issues concerning the domains of definition of the respective operators. Resolving Conjecture 11 would shed further light on the structure of $\mathcal{G}$, especially in connection with the results obtained in section 4.

2.3. Ergodicity and regularity of couplings. From here on, let us make the following assumption, natural in the context of MCMC samplers.

Assumption 2. The underlying one-particle processes $(X_t^i)_{t\geq 0}$ are ergodic; i.e., for every $i \in \{1, \dots, n\}$ there exists a unique probability measure $\pi_i \in \mathcal{P}(E_i)$ on E_i such that

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

$$\int_{E_i} (\mathcal{L}_i f)\,\mathrm{d}\pi_i = 0, \quad f \in \mathcal{D}(\mathcal{L}_i), \tag{29}$$

and, furthermore,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(X_t^i)\,\mathrm{d}t = \int_{E_i} f\,\mathrm{d}\pi_i, \quad f \in C_b(E_i), \tag{30}$$

for π_i-almost all initial conditions $X_0^i \in E_i$.

Following up on Example 10, we see that the semigroup $(\bar{S}_t^0)_{t\geq 0}$ as given in (28) with generator $\bar{\mathcal{L}}_0$ as defined in (27) is ergodic with respect to the product measure

$$\bar{\pi}_0 := \bigotimes_{i=1}^n \pi_i \tag{31}$$

on $\bar{E}$. Unfortunately, it turns out that not all coupling operators $\Gamma \in \mathcal{G}$ induce ergodic coupled processes, even under Assumption 2 (for an example, see [62, section 3.1][6]). We therefore make the following definition.

Definition 12 (ergodic couplings). *A coupling operator $\Gamma \in \mathcal{G}$ is called* ergodic *if the Feller process generated by $\bar{\mathcal{L}}_\Gamma$ is ergodic. The corresponding subset of ergodic coupling operators will be denoted by $\mathcal{G}^0$. The unique invariant measure associated to $\Gamma \in \mathcal{G}^0$ will be denoted by $\bar{\pi}_\Gamma$.*

Remark 13. By construction, the measures $\bar{\pi}_\Gamma$ are couplings of the one-particle invariant measures $(\pi_i)_{i=1}^n$.

Remark 14. For the analysis, ergodicity of the coupling is a crucial requirement (although with more work it might be possible to extend some of the results to the case when ergodicity fails to hold). Let us emphasize, however, that the validity of (8) does not depend on this, as only the marginal property of the coupling is used in its derivation. Hence in practice it is harmless to use nonergodic couplings, and in fact our results obtained in section 4 (in particular, Theorem 49) suggest using couplings that are at least not straightforwardly seen to be ergodic. In this case, quantities measuring the performance of the sampler (such as the asymptotic variance corresponding to certain observables) might be undefined or depend on the initial condition.

In general, ergodicity might fail in various ways. For instance, the process might not admit any invariant measure at all, or convergence of ergodic averages (in the sense of (30)) might not hold. The following result shows that the situation is simpler in our context.

Lemma 15. *Let Assumption 2 be satisfied. Then the following hold:*

1. *Every Feller coupling admits at least one invariant measure.*

[6] As this example shows, in specific situations it is possible to couple two ergodic processes in such a way that they stay at a fixed distance almost surely, for all time (at least for a certain class of initial conditions). Clearly a joint process with this property cannot be ergodic since the initial condition (more precisely, the relative initial condition of the two particles) is not forgotten.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

2. *If a Feller coupling $(\bar{X}_t)_{t\geq 0}$ admits a unique invariant measure $\bar{\pi}$, then it is ergodic, i.e.,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T F(\bar{X}_t)\,\mathrm{d}t = \int_{\bar{E}} F\,\mathrm{d}\bar{\pi}, \quad F \in C_b(\bar{E}), \tag{32}$$

for $\bar{\pi}$-almost all initial conditions $\bar{X}_0 \in \bar{E}$.

Proof. We proceed along the lines of the proof of the Krylov–Bogolyubov theorem [23, section 3.1]. Let $X_0^i \in E_i$ be initial conditions for the processes $(X_t^i)_{t\geq 0}$ such that (30) holds. Then, the families $(\tilde{\pi}_t^i)_{t\geq 0}$ of Cesàro averages

$$\tilde{\pi}_t^i(A) = \frac{1}{t}\int_0^t \left((S_s^i)^*\delta_{X_0^i}\right)(A)\,\mathrm{d}s, \quad A \in \mathcal{B}(E_i),$$

are convergent and therefore tight. Let $(\bar{X}_t)_{t\geq 0}$ be a Feller coupling, and denote the corresponding Cesàro averages by $(\tilde{\bar{\pi}}_t)_{t\geq 0}$. For any $t \geq 0$, $\tilde{\bar{\pi}}_t$ is a coupling of $(\tilde{\pi}_t^i)_{i=1}^n$. Using an obvious extension of [94, Lemma 4.4] to the multimarginal case, we see that $(\tilde{\bar{\pi}}_t)_{t\geq 0}$ is tight. By Prokhorov's theorem, there exists a weakly converging subsequence, the limit of which is an invariant measure (as in the proof of the Krylov–Bogolyubov theorem). This proves the first claim. Now let us assume that there exists a unique invariant measure. Since any convergent subsequence of $(\tilde{\bar{\pi}}_t)_{t\geq 0}$ has to converge to the same limit, and the sequence is tight, the second claim follows. ∎

By the results obtained in [52], uniqueness of the invariant measure is implied by certain regularity properties of the process. This leads to the following convenient criterion.

Corollary 16 (regular couplings). *Let Assumption 2 be satisfied, and consider a Feller coupling $(\bar{X}_t)_{t\geq 0}$. If the corresponding transition functions $(\rho_t(x,\cdot))_{t\geq 0, x\in\bar{E}}$ are mutually absolutely continuous (i.e., if the process is* regular*), then $(\bar{X}_t)_{t\geq 0}$ is ergodic.*

The measures π_i, as well as $\bar{\pi}_\Gamma$ (for $\Gamma \in \mathcal{G}^0$) induce the usual Hilbert spaces $L^2(\pi_i)$ and $L^2(\bar{\pi}_\Gamma)$ of square-integrable functions. A crucial role will be played furthermore by the corresponding subspaces of centered functions, defined by

$$L_0^2(\bar{\pi}_\Gamma) = \{f \in L^2(\bar{\pi}_\Gamma) \,|\, \bar{\pi}_\Gamma(f) = 0\}, \tag{33}$$

and analogously for $L_0^2(\pi_i)$. Since any Feller process has a right-continuous version, the semigroups $(\bar{S}_t^\Gamma)_{t\geq 0}$ as well as the corresponding generators $(\bar{\mathcal{L}}_\Gamma, \mathcal{D}(\bar{\mathcal{L}}_\Gamma))$ have unique extensions to strongly continuous semigroups on $L^2(\bar{\pi}_\Gamma)$ by Jensen's inequality. Slightly abusing the notation, we will denote those semigroups and their generators by the same letters. Before moving on to a somewhat more explicit description of coupling operators, let us mention the following open question, related to Conjecture 11.

Conjecture 17. The set $\mathcal{G}^0$ is convex.

2.4. A general way of constructing coupling operators. In this section we describe an approach to construct coupling operators explicitly in applications. The particular form presented here is also theoretically important since some of the calculations in later sections depend on it (especially the proof of Theorem 49).

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

As in the previous section, we assume that the marginal processes are ergodic with respect to invariant measures π_i (see Assumption 2). Furthermore, let us assume that the generators $\mathcal{L}_i$ can be written as

$$\mathcal{L}_i = -\sum_{k=1}^{K_i} (A_k^i)^* A_k^i + B^i, \tag{34}$$

where $(A_k^i)^*$ denotes the adjoint of A_k^i in $L^2(\pi_i)$ and B^i is antisymmetric in $L^2(\pi_i)$. Clearly, this decomposition into symmetric and antisymmetric parts is always possible, and in many cases the operators A_k^i and B^i can be chosen to have convenient forms. Note, however, that the decomposition (34) is not unique, since there are (infinitely) many ways of choosing the operators A_k^i. A particular choice of decomposing the generators $\mathcal{L}_i$ as in (34) hence essentially amounts to the choice of square-roots for the symmetric parts. We remark here that naturally $C_c^\infty(\bar{E}) \subset \mathcal{D}(A_k^i)$ and $C_c^\infty(\bar{E}) \subset \mathcal{D}(B^i)$ are implicitly assumed, authorizing the computations in later sections. The following lemma is essential for the construction in this subsection.

Lemma 18. *Let A_k^i and B^i be given as in (34). Then*

$$\operatorname{span} \mathbf{1} \subset \ker A_k^i, \quad \operatorname{span} \mathbf{1} \subset \ker B^i \tag{35}$$

for all $i \in \{1, \dots, n\}$, $k \in \{1, \dots K_i\}$.

Proof. See [93, Proposition 2]. ∎

We may now set

$$\Gamma = \sum_{(i,j,k,l)\in\mathcal{J}} \alpha_{ijkl}(x_1, \dots, x_n) A_k^i A_l^j \tag{36}$$

for appropriate measurable functions $\alpha_{ijkl} : \bar{E} \to \mathbb{R}$ and where we have introduced the set of *admissible indices*

$$\mathcal{J} = \{(i,j,k,l) \in \mathbb{N}^4 : 1 \le i, j \le n,\ i \ne j, 1 \le k \le K_i,\ 1 \le l \le K_j\} \tag{37}$$

associated to the decomposition (34). Applying Lemma 18, we see immediately that the first condition of Definition 6 is satisfied. The second condition will typically enforce certain bounds on the functions α_{ijkl} via the positive maximum principle as well as regularity constraints if we are interested in Feller couplings. Those properties will have to be determined according to the particular form of the generators $\mathcal{L}_i$. Furthermore, whether Γ as defined in (36) belongs to $\mathcal{G}^0$ will also depend on the choice of the functions α_{ijkl}.

It is not clear whether the construction presented in this section exhausts the class of coupling operators $\mathcal{G}$. We present this problem as a conjecture.

Conjecture 19. For $i \in \{1, \dots, n\}$, assume that we are given generators $(\mathcal{L}_i, \mathcal{D}(\mathcal{L}_i))$ of ergodic Feller semigroups. Then there exist decompositions of the form (34) and a set of functions

$$\mathcal{U} = \{(\alpha_{ijkl})_{(i,j,k,l)\in\mathcal{J}} : \bar{E} \to \mathbb{R}\} \tag{38}$$

such that

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$\mathcal{G} = \left\{ \Gamma = \sum_{(i,j,k,l)\in\mathcal{J}} \alpha_{ijkl} A^i_k A^j_l : \quad \alpha_{ijkl} \in \mathcal{U} \right\}. \tag{39}$$

In the case when the underlying processes $(X^i_t)_{t\geq 0}$ are $\mathbb{R}^{d_i}$-valued (i.e., $E_i = \mathbb{R}^{d_i}$) and have continuous paths almost surely, Courrège's theorem ([20, Theorem 0.1]; see also [48, section 4.5] for a more recent account) provides an explicit characterization of Feller generators. If, furthermore, these processes are ergodic with respect to given invariant measures, the decomposition of their generators into symmetric and antisymmetric part can be made explicit (see [29, Theorem 1]). Combining these theorems, we obtain the following partial result.

Proposition 20. *Let $E_i = \mathbb{R}^{d_i}$ for positive integers $d_i \in \mathbb{N}$, and assume that the processes $(X^i_t)_{t\geq 0}$ are ergodic and solve the Itô SDEs*

$$\mathrm{d}X^i_t = b^i(X^i_t)\,\mathrm{d}t + \sqrt{2}\sigma^i(X^i_t)\,\mathrm{d}W^i_t, \tag{40}$$

where $b^i \in C^1(\mathbb{R}^{d_i}, \mathbb{R}^{d_i})$, $\sigma \in C^1(\mathbb{R}^{d_i}, \mathbb{R}^{d_i\times m_i})$, and $(W^i_t)_{t\geq 0}$ are standard m_i-dimensional Brownian motions. Then the conclusion of Conjecture 19 holds.

3. Examples of coupled processes. Here we will illustrate the framework developed in the last section with concrete examples. Throughout we consider the task of sampling from the measure

$$\pi(\mathrm{d}x) = \frac{1}{Z} e^{-V(x)} \mathrm{d}x, \quad x \in \mathbb{R}^d, \tag{41}$$

where $V \in C^\infty(\mathbb{R}^d)$ is a potential satisfying

$$Z := \int_{\mathbb{R}^d} e^{-V(x)} \mathrm{d}x < \infty. \tag{42}$$

3.1. Overdamped Langevin dynamics. Our first group of examples is concerned with the overdamped Langevin dynamics [76, section 4.5]. Let us start with the one-dimensional case, already encountered in the introduction.

3.1.1. Two particles in one dimension. We consider $n = 2$ particles moving in dimension $d = 1$, each of them according to the dynamics

$$\mathrm{d}X_t = -V'(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t. \tag{43}$$

Note that in order to precisely fit into our framework developed in the previous section, the process $(X_t)_{t\geq 0}$ is required to be a Feller process according to Definition 1. This property can be guaranteed by imposing certain growth conditions on the potential V; see [3, Proposition 5.9, Example 5.4] and [65, Theorem 5.3.2, Example 5.3.3]. However, we wish to remark that the Feller property is not crucial in practice, and dispensing with this regularity requirement still leads to perfectly well-defined couplings as will become clear in Lemma 22 below. The generator of (43) is given by

$$\mathcal{L} = -V'(x)\partial_x + \partial_x^2 = -\partial_x^*\partial_x, \tag{44}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where the adjoint is taken in $L^2(\pi)$. In particular, $\mathcal{L}$ can naturally be written in the form (34), with $A = \partial_x$ and $B = 0$. To illustrate trivial couplings (see Example 10), consider first two independent identical copies of (43), denoted by $(X_t, Y_t)_{t\geq 0}$, hence evolving according to the dynamics

$$\mathrm{d}X_t = -V'(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^x, \tag{45a}$$

$$\mathrm{d}Y_t = -V'(Y_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^y \tag{45b}$$

on the product space $\bar{E} = \mathbb{R}^2$. Since for now the processes $(W_t^x)_{t\geq 0}$ and $(W_t^y)_{t\geq 0}$ are supposed to be two independent standard Brownian motions, the generator of (45) is given by

$$\bar{\mathcal{L}}_0 = \mathcal{L}_x + \mathcal{L}_y \tag{46}$$

with $\mathcal{L}_x = -V'(x)\partial_x + \partial_x^2 = -\partial_x^*\partial_x$ and $\mathcal{L}_y = -V'(y)\partial_y + \partial_y^2 = -\partial_y^*\partial_y$, in agreement with Example 10, (27). The invariant measure of (45) is given by the product

$$\bar{\pi}_0 = \pi_x \otimes \pi_y = \frac{1}{Z^2} e^{-(V(x)+V(y))}\mathrm{d}x\mathrm{d}y. \tag{47}$$

Now let us consider nontrivial couplings. Following section 2.4, we may set

$$\Gamma = 2\alpha(x,y)\partial_x\partial_y \tag{48}$$

for an appropriate measurable function $\alpha : \mathbb{R}^2 \to \mathbb{R}$ (we have inserted a factor of 2 for convenience). According to Proposition 20, the set of operators of the form (48) exhausts the set of coupling operators $\mathcal{G}$. Clearly, the first condition of Definition 6 is satisfied for this set of operators (this is already guaranteed by using the construction from section 2.4). The second condition enforces

$$-1 \leq \alpha(x,y) \leq 1 \quad \text{for all } x, y \in \mathbb{R}. \tag{49}$$

Indeed, observe that

$$\bar{\mathcal{L}}_\Gamma = \bar{\mathcal{L}}_0 + \Gamma = \nabla_z U(z) \cdot \nabla_z + Q(z) : \nabla_z \nabla_z, \tag{50}$$

where $z = (x, y)$, $U(z) = U(x, y) = V(x) + V(y)$,

$$Q(z) = Q(x,y) = \begin{pmatrix} 1 & \alpha(x,y) \\ \alpha(x,y) & 1 \end{pmatrix}, \tag{51}$$

and where : denotes the Frobenius inner product of matrices. According to Courrège's theorem, $\bar{\mathcal{L}}_\Gamma$ satisfies the positive maximum principle (required by the Hille–Yosida–Ray theorem) only if $Q(z)$ is nonnegative definite for every $z \in \mathbb{R}^2$. From this, we immediately deduce the constraint (49).

Remark 21. We are deliberately vague about the regularity properties of α. If we restrict our attention to Feller processes, α certainly has to be at least continuous, and there are multiple results in the literature guaranteeing the Feller property under mild further assumptions on α, in particular, Hölder regularity [3, 65]. For a discussion of the martingale problem for

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

generators with discontinuous coefficients see the recent preprint [56] and references therein. Note that even in this simple case, it is very challenging to characterize exactly the set $\mathcal{G}$ as introduced in Definition 6. In applications, however, the Feller property is not crucial. Lemma 22 below shows that measurability of α is sufficient to ensure that a reasonable coupled process can be constructed.

Assuming that (49) is satisfied, the dynamics induced by the generator (50) are (at least formally) given by

$$\mathrm{d}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} -V'(X_t) \\ -V'(Y_t) \end{pmatrix} \mathrm{d}t + \sqrt{2} G(X_t, Y_t) \begin{pmatrix} \mathrm{d}W_t^x \\ \mathrm{d}W_t^y \end{pmatrix} \tag{52}$$

where $G(x,y)G(x,y)^T = Q(x,y)$, for instance

$$G(x,y) = \begin{pmatrix} \cos\beta(x,y) & g(x,y)\sin\beta(x,y) \\ g(x,y)\sin\beta(x,y) & \cos\beta(x,y) \end{pmatrix} \tag{53}$$

with $\beta(x,y) = \frac{1}{2}\arcsin|\alpha(x,y)|$ and $g(x,y) = \operatorname{sgn}\alpha(x,y)$. We have chosen this parametrization since it generalizes readily to higher dimensions (see below). Let us stress that writing the dynamics in the form (52) is vital for applications, since it enables its simulation in a straightforward manner. The following lemma shows that the process constructed in this way is indeed a coupling in the sense of Definition 4. Furthermore, it turns out that only minimal regularity of α is required.

Lemma 22. *Let α be measurable. Then the dynamics (52) can be written as*

$$\mathrm{d}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} -V'(X_t) \\ -V'(Y_t) \end{pmatrix} \mathrm{d}t + \sqrt{2} \begin{pmatrix} \mathrm{d}B_t^x \\ \mathrm{d}B_t^y \end{pmatrix} \tag{54}$$

with two Brownian motions $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$ that are in general not independent.

Proof. The claim follows from applying Lemma 74 in Appendix A to the components of the SDE

$$\begin{pmatrix} \mathrm{d}B_t^x \\ \mathrm{d}B_t^y \end{pmatrix} = G \begin{pmatrix} \mathrm{d}W_t^x \\ \mathrm{d}W_t^y \end{pmatrix}, \tag{55}$$

where G is a 2×2-matrix with rows $G^x := (g\cos\beta\ \sin\beta)$ and $G^y := (g\sin\beta\ \cos\beta)$. It is straightforward to verify that those indeed satisfy condition (191) with $N = 1$ and $M = 2$. ∎

Note that the function α (equivalently the pair β and g) encodes the coupling between the Brownian motions $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$. The parameter $\beta \in [0, \frac{\pi}{4}]$ is related to the *strength* of the coupling, whereas $g \in \{-1, 1\}$ is related to its *direction*. Indeed, if $\beta \equiv 0$, then $(B_t^x)_{t\geq 0}$ and $(B_t^y)_{t\geq 0}$ are independent ("trivial coupling"; see (45) and Example 10). If $\beta \equiv \frac{\pi}{4}$ and $g \equiv 1$, then $B_t^x = B_t^y$ for $t \geq 0$, almost surely ("synchronous coupling"). Likewise, if $\beta \equiv \frac{\pi}{4}$ and $g \equiv -1$, then $B_t^x = -B_t^y$ ("mirror coupling").

Remark 23 (ergodic couplings). If the bound (49) is satisfied with strict inequalities, then the generator (50) is elliptic, and hence, by Corollary 16, the coupled process is ergodic. Let us mention that this condition is not necessary for ergodicity. Indeed, consider the coupling

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

operator $\Gamma = -2\partial_x\partial_y$, inducing the so-called "two-point motion" [7], i.e., $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are driven by the same Brownian motion, only differing by their initial laws. Under mild regularity conditions (i.e., Lipschitz continuity of the coefficients), it can be shown that $(X_t, Y_t)_{t\geq 0}$ is ergodic with respect to $\bar{\pi}_\Gamma = \frac{1}{Z}e^{-V(x)}\delta_{x-y}(\mathrm{d}x\mathrm{d}y)$; see, for instance, [62, Theorem 2.1].

3.1.2. The general case. Here, we will extend the discussion from the previous section to the general case of n particles moving in d dimensions; i.e., we are concerned with couplings of the dynamics

$$\mathrm{d}X_t^i = -\nabla V(X_t^i)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^i, \quad i = 1, \dots, n, \tag{56}$$

the processes $(X_t^i)_{t\geq 0}$ being $\mathbb{R}^d$-valued. If the Brownian motions $(W_t^i)_{t\geq 0}$ are independent, then the joint process $(\bar{X}_t)_{t\geq 0} = (X_t^1, \dots, X_t^n)_{t\geq 0}$ is ergodic with respect to the product measure $\bar{\pi}_0 = \bigotimes_{i=1}^n \pi_i$ on $\mathbb{R}^{nd}$ (see Example 10), and the corresponding generator is given by

$$\bar{\mathcal{L}}_0 = \sum_{i=1}^n \mathcal{L}_i, \quad \mathcal{L}_i = -\nabla V(x_i)\cdot\nabla_{x_i} + \Delta_{x_i} = -\sum_{k=1}^d (\partial_k^i)^*\partial_k^i. \tag{57}$$

Here, ∂_k^i denotes the derivative with respect to the kth component of x_i, and the adjoints are taken in the spaces $L^2(\pi_i)$. Clearly, the generators $\mathcal{L}_i$ are decomposed as in (34) with $A_k^i = \partial_k^i$ and $B^i = 0$.

Remark 24. Instead of (56), we can also consider the more general dynamics

$$\mathrm{d}X_t^i = -Q_i(X_t^i)\nabla V(X_t^i)\,\mathrm{d}t + (\nabla\cdot Q_i)(X_t^i)\,\mathrm{d}t + J_i\nabla V(X_t^i)\,\mathrm{d}t + \sqrt{2Q_i(X_t^i)}\,\mathrm{d}W_t^i \tag{58}$$

with $i = 1, \dots, n$, $J_i \in \mathbb{R}^{d\times d}_{\mathrm{skew}}$ being skew-symmetric matrices and $Q_i : \mathbb{R}^d \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ being positive definite matrix-valued functions, as discussed in [29, section 2]. Note that in this case the processes $(X_t^i)_{t\geq 0}$ are not copies of each other, since Q_i and J_i may not be the same for different particles.

To construct nontrivial couplings, we may set

$$\Gamma = \sum_{i,j=1, i\neq j}^n \sum_{k,l=1}^d \alpha_{ijkl}\partial_k^i\partial_l^j \tag{59}$$

for appropriate[7] functions $\alpha_{ijkl} : \mathbb{R}^{nd} \to \mathbb{R}$, following section 2.4. Note that $\partial_k^i\partial_l^j$ is symmetric with respect to the interchange of indices $(i,k) \leftrightarrow (j,l)$, and so we may assume that $\alpha_{ijkl} = \alpha_{jilk}$. As in the one-dimensional case, the generator of the coupled system $\bar{\mathcal{L}}_\Gamma = \bar{\mathcal{L}}_0 + \Gamma$ is a second order differential operator which we require to be (possibly degenerately) elliptic in order for the second condition of Definition 6 to be satisfied (again with reference to Courrège's theorem).

[7] Concerning the regularity of these functions, the discussion from the previous section applies, see in particular Remark 21.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

To derive more easily verifiable conditions on the functions α_{ijkl}, let us introduce a matrix-valued function (or *matrix field*) $Q : (\mathbb{R}^d)^n \to \mathbb{R}^{nd\times nd}$ as follows. Firstly, it is helpful to view the target space of Q as $\mathbb{R}^{nd\times nd} \equiv (\mathbb{R}^{d\times d})^{n\times n}$; i.e., we think of $Q(x_1,\dots,x_n)$ as an $n\times n$-matrix the entries of which are themselves $d\times d$ matrices. In other words, $Q_{ij}(x_1,\dots,x_n)$ is a $d\times d$ matrix for every pair $(i,j)\in\{1,\dots,n\}^2$. The matrix field Q can then be defined by

$$Q_{ij}(x_1,\dots,x_n) = \begin{cases} I_{d\times d}, & i=j,\\ \alpha_{ij}(x_1,\dots,x_n), & i\neq j, \end{cases} \tag{60}$$

where $\alpha_{ij}(x_1,\dots,x_n)\in\mathbb{R}^{d\times d}$ denotes the matrix with entries $(\alpha_{ijkl}(x_1,\dots,x_n))_{k,l=1,\dots,d}$. Observe that $Q(x_1,\dots,x_n)$ as defined in (60) is symmetric as a matrix in $\mathbb{R}^{nd\times nd}$ by our assumption that $\alpha_{ijkl}=\alpha_{jilk}$.

Remark 25. The matrices Q_{ij} can now be thought of as describing the coupling between the particles i and j.

As in the one-dimensional case, the generator of the fully coupled system can be written as

$$\bar{\mathcal{L}}_\Gamma = \bar{\mathcal{L}}_0 + \Gamma = \nabla_z U(z)\cdot\nabla_z + Q(z) : \nabla_z\nabla_z \tag{61}$$

introducing the notation $z\equiv(x_1,\dots,x_n)\in\mathbb{R}^{nd\times nd}$ and $U(z)=\sum_{i=1}^n V(x_i)$. To make the connection to SDEs and arrive at a description analogous to (52), let us consider matrix fields $G : (\mathbb{R}^d)^n \to (\mathbb{R}^{d\times d})^{n\times n}$ satisfying

$$\sum_{j=1}^n G_{ij}(x_1,\dots,x_n)G_{ij}(x_1,\dots,x_n)^T = I_{d\times d} \tag{62}$$

for all $i=1,\dots,n$. Here the transposition ^{T} is taken in $\mathbb{R}^{d\times d}$. Note that matrix fields of this form give rise to the matrix fields $Q : (\mathbb{R}^d)^n \to \mathbb{R}^{nd\times nd}$ defined in (60) via $Q=GG^T$ (where the transposition is taken in $\mathbb{R}^{nd\times nd}$).

Remark 26. The advantage of constructing the coupling in terms of the matrix field G is that pointwise positive semidefiniteness of Q is automatically satisfied. A practical way to fulfill the constraint (62) is to choose matrix fields $g_{ij} : (\mathbb{R}^d)^n \to \mathbb{R}^{d\times d}$ that are orthogonal pointwise, i.e.,

$$g_{ij}(x_1,\dots,x_n)g_{ij}(x_1,\dots,x_n)^T = g_{ij}(x_1,\dots,x_n)^T g_{ij}(x_1,\dots,x_n) = I_{d\times d} \tag{63}$$

for all $i,j=1,\dots,n$ and $(x_1,\dots,x_n)\in(\mathbb{R}^d)^n$, as well as weights $w_{ij} : (\mathbb{R}^d)^n \to \mathbb{R}$ satisfying

$$\sum_{j=1}^n w_{ij}^2(x_1,\dots,x_n) = 1 \tag{64}$$

for all $i=1,\dots,n$. Then, setting $G_{ij}=w_{ij}g_{ij}$, condition (62) holds. Intuitively, the orthogonal matrices g_{ij} encode the coupling between particle i and j through a rotation of the noise. The weights w_{ij} can be interpreted as the relative coupling strengths between the particles. Observe that both g_{ij} and w_{ij} may depend on $(x_1,\dots,x_n)$, i.e., on the locations of all the particles. This construction is a direct generalization of (53).

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Assuming $GG^T = Q$, the dynamics associated to the generator (61) is (again, at least formally) given by

$$\mathrm{d}X_t^i = -\nabla V(X_t^i)\,\mathrm{d}t + \sqrt{2}\sum_{j=1}^{n} G_{ij}(X_t^1,\dots,X_t^n)\,\mathrm{d}W_t^j, \quad i = 1,\dots,n, \tag{65}$$

where $((W_t^j)_{t\geq 0})_{j=1}^n$ are assumed to be independent standard Brownian motions. As in the one-dimensional case, we have the following lemma.

Lemma 27. *There exist $\mathbb{R}^d$-valued standard Brownian motions $((B_t^j)_{t\geq 0})_{j=1}^n$, not necessarily independent, such that the dynamics (65) can be written as*

$$\mathrm{d}X_t^i = -\nabla V(X_t^i)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^i, \quad i = 1,\dots,n. \tag{66}$$

Proof. The argument is identical to the one used in the proof of Lemma 22. ∎

Concerning the ergodicity of couplings, we have similar findings to those of Remark 23. If the matrix field (60) is positive definite at every point (i.e., the generator (61) is elliptic), then Corollary 16 implies that the coupled process is ergodic. An example of nonergodic couplings can be found in [62, section 3.1].

Example 28 (two particles). The foregoing constructions become more explicit when considering only $n = 2$ particles. To simplify the notation, we denote their positions by $x \equiv x_1$ and $y \equiv x_2$. The diffusion matrices (60) reduce to

$$Q(x,y) = \begin{pmatrix} I_{d\times d} & \alpha(x,y) \\ \alpha^T(x,y) & I_{d\times d} \end{pmatrix} \tag{67}$$

with $\alpha : (\mathbb{R}^d)^2 \to \mathbb{R}^{d\times d}$. Using Schur complements [17, Appendix 5.5], we see that $Q(x,y)$ is positive semidefinite if and only if

$$\alpha(x,y)^T\alpha(x,y) \leq I_{d\times d}, \tag{68}$$

in the sense of symmetric matrices (Loewner ordering). The corresponding coupling operator is given by

$$(\Gamma f)(x,y) = 2\,\mathrm{Tr}\left(\alpha^T(x,y)\nabla_{xy}^2 f(x,y)\right), \quad f \in C_c^\infty(\mathbb{R}^d\times\mathbb{R}^d), \tag{69}$$

where the matrix of mixed derivatives $\nabla_{xy}^2\phi$ is given by $\left(\nabla_{xy}^2\phi\right)_{ij} = (\partial_{x_i}\partial_{y_j}\phi)_{ij}$. In order to illustrate the construction from Remark 26, let $g : (\mathbb{R}^d\times\mathbb{R}^d) \to \mathbb{R}^{d\times d}$ be a field of orthogonal matrices (see (63)) and put

$$G(x,y) = \begin{pmatrix} \cos\beta(x,y)I_{d\times d} & \sin\beta(x,y)g(x,y) \\ \sin\beta(x,y)g^T(x,y) & \cos\beta(x,y)I_{d\times d} \end{pmatrix}, \tag{70}$$

the function $\beta : \mathbb{R}^d\times\mathbb{R}^d \to [0,\frac{\pi}{4}]$ again regulating the strength of the coupling, and being associated with the weights w_{ij} in (64). From $GG^T = Q$ it follows that α and g are connected via

$$\alpha = 2\cos\beta\sin\beta\cdot g. \tag{71}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

3.2. Underdamped Langevin dynamics. For fixed $\gamma > 0$ ("friction") and symmetric positive definite $M \in \mathbb{R}^{d\times d}_{\mathrm{sym}}$ ("mass"), the dynamics

$$\mathrm{d}q_t = M^{-1}p_t\,\mathrm{d}t, \tag{72a}$$

$$\mathrm{d}p_t = -\nabla V(q_t)\,\mathrm{d}t - \gamma p_t\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}W_t \tag{72b}$$

is ergodic with respect to the measure

$$\pi = \frac{1}{\bar{Z}} e^{-\left(V(q)+\frac{1}{2}p^T M^{-1}p\right)}\mathrm{d}q\mathrm{d}p, \quad (q,p) \in \mathbb{R}^{2d}, \tag{73}$$

$\bar{Z}$ being an appropriate normalization constant (see [76, Chapter 6] for details). The generator is given by

$$\mathcal{L} = M^{-1}p\cdot\nabla_q - \nabla_q V(q)\cdot\nabla_p + \gamma(-p\cdot\nabla_p + \Delta_p) \tag{74a}$$

$$= B - \sum_{k=1}^{d} A_k^* A_k, \tag{74b}$$

where $\mathcal{T} = M^{-1}p\cdot\nabla_q - \nabla_q V(q)\cdot\nabla_p$ is skew-symmetric in $L^2(\pi)$, $A_k = \sqrt{\gamma}\partial_{p_k}$, and the adjoint is taken in $L^2(\pi)$. To construct a coupled sampler of n processes, we may proceed as in the overdamped case and set

$$\Gamma = \sum_{i,j=1, i\neq j}^{n}\sum_{k,l=1}^{d} \alpha_{ijkl}\partial^i_{p_k}\partial^j_{p_l}, \tag{75}$$

for appropriate functions $\alpha_{ijkl} : (\mathbb{R}^{2d})^n \to \mathbb{R}$, denoting by $\partial^i_{p_k}$ the derivative with respect to the kth component of p of the ith particle. Following very closely the discussion in section 3.1, we can introduce matrix fields $Q : (\mathbb{R}^{2d})^n \to (\mathbb{R}^{d\times d})^{n\times n}$ and $G : (\mathbb{R}^{2d})^n \to (\mathbb{R}^{d\times d})^{n\times n}$ satisfying (60) and (62) (with $(x_1,\dots,x_n)$ replaced by $(q_1,p_1,\dots,q_n,p_n)$) such that the generator of the coupled system is given by

$$\bar{\mathcal{L}}_\Gamma = \sum_{i=1}^{n}\left(M^{-1}p_i\cdot\nabla_{q_i} - \nabla_q V(q_i)\cdot\nabla_{p_i} - \gamma p_i\cdot\nabla_{p_i}\right) + Q(q_1,p_1,\dots,q_n,p_n) : \nabla_p\nabla_p \tag{76}$$

with the associated dynamics being given by

$$\mathrm{d}q^i_t = M^{-1}p^i_t\,\mathrm{d}t, \tag{77a}$$

$$\mathrm{d}p^i_t = -\nabla_q V(q^i_t)\,\mathrm{d}t - \gamma p^i_t\,\mathrm{d}t + \sqrt{2\gamma}\sum_{j=1}^{n} G_{ij}(q_1,p_1,\dots,q_n,p_n)\,\mathrm{d}W^j_t \tag{77b}$$

for $i = 1,\dots,n$. Also in this case, it is straightforward to see that an appropriate version of Lemmas 22 and 27 holds.

Remark 29. Generalizing the above to the case where the particles have different frictions γ_i and masses M_i or some or all of them move according to perturbed versions of underdamped Langevin dynamics as considered in [29] is straightforward (see also Remark 24).

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Example 30 (overdamped and underdamped Langevin dynamics). It is also possible to couple different types of dynamics (such as overdamped and underdamped Langevin dynamics). For instance, consider the generators

$$\mathcal{L}_1 = -\nabla V(x) \cdot \nabla_x + \Delta_x, \tag{78a}$$

$$\mathcal{L}_2 = M^{-1} p \cdot \nabla_q - \nabla_q V(q) \cdot \nabla_p + \gamma(-p \cdot \nabla_p + \Delta_p), \tag{78b}$$

as in (57) and (74a). Setting

$$\Gamma = \sum_{k,l=1}^{d} \alpha_{kl}(x,q,p) \partial_{x_k} \partial_{p_l} \tag{79}$$

and following along the lines of sections 3.1 and 3.2 will result in the coupled dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\mathrm{d}B_t^{(1)}, \tag{80a}$$

$$\mathrm{d}q_t = M^{-1} p_t\,\mathrm{d}t, \tag{80b}$$

$$\mathrm{d}p_t = -\nabla_q V(q)\mathrm{d}t - \gamma p_t\,\mathrm{d}t + \sqrt{2\gamma}\,\mathrm{d}B_t^{(2)}, \tag{80c}$$

where the Brownian motions $(B_t^{(1)})_{t\geq 0}$ and $(B_t^{(2)})_{t\geq 0}$ are in general not independent (and the exact dependence results from the choice of the functions α_{kl}).

3.3. The zigzag process. In recent years, there has been a growing interest in using PDMPs [24] in the context of sampling problems. These are processes that move deterministically between random events, usually along the trajectories of an ODE. At those events, a random transition (e.g., a "jump") occurs. Both the deterministic dynamics as well as the random transitions can be chosen with a great deal of flexibility, resulting in a range of possible PDMP algorithms. Let us mention here the bouncy particle sampler [16], the zigzag sampler [10], randomized Hamiltonian Monte Carlo [15], and event-chain Monte Carlo techniques [67, 68]. The recent papers [35] and [91] provide good overviews in a general framework. For technical details, we would like to refer the reader to [30].

The objective of this section is to show how the framework from section 2 can be employed in the construction of coupled samplers from PDMPs, using the example of the zigzag process. For ease of exposition, we furthermore restrict our attention to the one-dimensional case. The treatment here follows [10] and [11] in style and notation.

The state space under consideration is $E = \mathbb{R} \times \{-1, +1\}$, and the generator of the zigzag process reads

$$\mathcal{L}f(x,\theta) = \theta \partial_x f(x,\theta) + \lambda(x,\theta)\left(f(x,-\theta) - f(x,\theta)\right), \quad f \in C_c^\infty(E), \tag{81}$$

where the *switching rate* λ is given by

$$\lambda(x,\theta) = \max(0, \theta V'(x)) + \gamma(x). \tag{82}$$

Here, $\gamma : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a nonnegative continuous function, called the *excess switching rate*. Roughly speaking, the zigzag process moves along straight lines in the direction determined

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

by $\theta \in \{-1, 1\}$. At random times a switch occurs, i.e., θ is replaced by $-\theta$. Those events are sampled according to the switching rate λ; i.e., at a point $(x, \theta) \in E$, the probability for the switch $\theta \mapsto -\theta$ in the time span $[t, t + \varepsilon]$ is given by $\lambda(x, t)\varepsilon + o(\varepsilon)$. For more details on the construction and simulation of zigzag processes we refer the reader to [9] and [10], as well as to [24] for more general PDMPs. According to [9, Proposition 1] and [91, Appendix B.2], the zigzag process satisfies the Feller property; for more general PDMPs this topic has been studied in [25, Theorem 27.6]. The measure

$$\pi = \frac{1}{2Z} e^{-V(x)} \mathrm{d}x \otimes (\delta_{-1} + \delta_{+1}) \tag{83}$$

is invariant and, under some additional assumptions,[8] ergodic. The generator (81) can be decomposed in the form

$$\mathcal{L} = -A^* A + B, \tag{84}$$

where

$$A = \left(\frac{1}{4} |V'| + \frac{1}{2} \gamma \right)^{1/2} (R - 1), \quad B = \theta \partial_x + \frac{1}{2} \theta V'(R - 1), \tag{85}$$

and the "flip operator" R is given by

$$(Rf)(x, \theta) = f(x, -\theta), \quad f \in C_c^\infty(\bar{E}). \tag{86}$$

A short calculation shows that indeed B is antisymmetric in $L^2(\pi)$, whereas A is symmetric.

To construct a coupled sampler from two zigzag processes, let us introduce the following notation: We consider the state space $\bar{E} = \mathbb{R}^2 \times \{-1, +1\}^2$, denoting its elements by $(x, y, \theta_x, \theta_y)$. Furthermore, we will make use of the flip operators

$$(R_x f)(x, y, \theta_x, \theta_y) = f(x, y, -\theta_x, \theta_y), \quad (R_y f)(x, y, \theta_x, \theta_y) = f(x, y, \theta_x, -\theta_y). \tag{87}$$

Following section 2.4, let us set

$$\Gamma = \alpha(x, y, \theta_x, \theta_y)(R_x - 1)(R_y - 1) \tag{88}$$

for an appropriate function $\alpha : \bar{E} \to \mathbb{R}$, i.e., Γ acts as

$$\begin{aligned} (\Gamma f)(x, y, \theta_x, \theta_y) = \alpha(x, y, \theta_x, \theta_y) \cdot \\ \cdot (f(x, y, \theta_x, \theta_y) - f(x, y, -\theta_x, \theta_y) - f(x, y, \theta_x, -\theta_y) + f(x, y, -\theta_x, -\theta_y)) \end{aligned} \tag{89}$$

on test functions $f \in C_c^\infty(\bar{E})$. Note that Γ vanishes on functions that either depend on only x and θ_x or only on y and θ_y. The next task is to obtain bounds on α that ensure that the second condition in Definition 6 is satisfied. To this end, let us expand

[8] See [9] and [11]. Let us mention in particular that ergodicity is guaranteed whenever the excess switching rate γ is strictly positive.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

$$(\bar{\mathcal{L}}_\Gamma f)(x,y,\theta_x,\theta_y) = (\mathcal{L}_x + \mathcal{L}_y + \Gamma) f(x,y,\theta_x,\theta_y) \tag{90a}$$
$$= \theta_x \partial_x f(x,y,\theta_x,\theta_y) + \theta_y \partial_y f(x,y,\theta_x,\theta_y) \tag{90b}$$
$$- \left(\lambda(x,\theta_x) + \lambda(y,\theta_y) - \alpha(x,y,\theta_x,\theta_y)\right) f(x,y,\theta_x,\theta_y) \tag{90c}$$
$$+ \left(\lambda(x,\theta_x) - \alpha(x,y,\theta_x,\theta_y)\right) f(x,y,-\theta_x,\theta_y) \tag{90d}$$
$$+ \left(\lambda(y,\theta_y) - \alpha(x,y,\theta_x,\theta_y)\right) f(x,y,\theta_x,-\theta_y) \tag{90e}$$
$$+ \alpha(x,y,\theta_x,\theta_y) f(x,y,-\theta_x,-\theta_y). \tag{90f}$$

For (90) to be the generator of a Markov process (in particular, for it to satisfy the positive maximum principle), the following inequalities have to be satified:

$$\lambda(x,\theta_x) + \lambda(y,\theta_y) - \alpha(x,y,\theta_x,\theta_y) \geq 0, \tag{91a}$$
$$\lambda(x,\theta_x) - \alpha(x,y,\theta_x,\theta_y) \geq 0, \tag{91b}$$
$$\lambda(y,\theta_y) - \alpha(x,y,\theta_x,\theta_y) \geq 0, \tag{91c}$$
$$\alpha(x,y,\theta_x,\theta_y) \geq 0, \qquad (x,y,\theta_x,\theta_y) \in \bar{E}. \tag{91d}$$

These conditions can be interpreted as saying that the transition probabilities for the coupled PDMP cannot be negative. Clearly, the conditions (91) are equivalent to

$$0 \leq \alpha(x,y,\theta_x,\theta_y) \leq \min\left(\lambda(x,\theta_x), \lambda(y,\theta_y)\right), \quad (x,y,\theta_x,\theta_y) \in \bar{E}. \tag{92}$$

Let us briefly comment on the dynamical behavior that the coupling operator (89) introduces. As can be seen from (90f), α is connected to "double flips"; i.e., the event that both particles change their directions at the same time. Setting α to either the lower or the upper bound in (92) will either discourage or encourage those double flips. As in the case of the overdamped and underdamped Langevin dynamics, the coupling behavior (encoded in α) is allowed to depend on the point $(x,y,\theta_x,\theta_y) \in \bar{E}$. We also remark that the process generated by $\bar{\mathcal{L}}_\Gamma$ as in (90) can be simulated conveniently by using the methods summarized in [9, Appendix B].

Remark 31. The construction in this section can be generalized to couplings of multiple zigzag processes in arbitrary dimensions by following a similar approach to the one taken in section 3.1.

4. Asymptotic variance and optimal transport.

In the following we analyze the asymptotic variance associated to estimators based on coupled processes (section 4.1) and connect the result to the theory of optimal transportation (section 4.2).

4.1. A central limit theorem for coupled processes.

The objective of this section is to establish a central limit theorem characterizing the convergence in (8) and to find an expression for the associated asymptotic variance in terms of ergodic coupling operators $\Gamma \in \mathcal{G}^0$ and invariant measures $\bar{\pi}_\Gamma$. In particular, our aim is to compare between estimators based on couplings (as in (8)) and the one-particle estimators (30). Naturally, Assumption 2 is still in force. Moreover, let us assume the following.

Assumption 3 (invertibility of the one-particle generators). The generators $\mathcal{L}_i$ are invertible on $L_0^2(\pi_i)$; i.e., for all $f \in L_0^2(\pi_i)$ there exists $\phi \in \mathcal{D}(\mathcal{L}_i) \cap L_0^2(\pi_i)$ such that

$$-\mathcal{L}_i \phi = f. \tag{93}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

It is well known that the validity of the foregoing assumption is guaranteed by sufficiently fast decay of the semigroups $(S_t^i)_{t\geq 0}$ in $L^2(\pi_i)$; see, for instance, [53]. In the following, let us fix observables of interest $f_i \in L_0^2(\pi_i)$ and denote the corresponding solutions to the Poisson equations (93) by ϕ_i. Supposing $X_0^i \sim \pi_i$, Assumption 3 implies the central limit theorems

$$\sqrt{T}\left(\frac{1}{T}\int_0^T f_i(X_t^i)\,\mathrm{d}t\right) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, 2\sigma_{f_i}^2), \tag{94}$$

where the asymptotic variances are given by

$$\sigma_{f_i}^2 = \langle f_i, \phi_i\rangle_{L^2(\pi_i)}; \tag{95}$$

see [8, 53].

We will now establish a similar central limit theorem for the coupled process $(\bar{X}_t)_{t\geq 0}$ induced by ergodic coupling operators $\Gamma \in \mathcal{G}^0$ and associated to extended observables of the form

$$F = \frac{1}{n}\sum_{i=1}^n f_i. \tag{96}$$

The fact that the observables f_i are assumed to be centered with respect to the measures π_i (i.e., $\pi_i(f_i) = 0$) implies that F is centered with respect to $\bar{\pi}_\Gamma$ (i.e., $\bar{\pi}_\Gamma(F) = 0$) since $\bar{\pi}_\Gamma$ is a coupling of the marginals $(\pi_i)_{i=1}^n$.

Theorem 32 (central limit theorem for coupled processes). *Let Assumption 3 be satisfied and assume that $f_i \in L_0^2(\pi_i)$ for $i \in \{1, \dots, n\}$, as well as $\Gamma \in \mathcal{G}^0$. Furthermore, let $\bar{X}_0 \sim \bar{\pi}_\Gamma$. Then*

$$\sqrt{T}\left(\frac{1}{T}\int_0^T F(\bar{X}_t)\,\mathrm{d}t\right) \xrightarrow[T\to\infty]{d} \mathcal{N}(0, 2\sigma_F^2). \tag{97}$$

The asymptotic variance σ_F^2 is given by

$$\sigma_F^2 = \frac{1}{n^2}\sum_{i=1}^n \sigma_{f_i}^2 - \int_{\bar{E}} \bar{\mathcal{L}}_0 \xi\,\mathrm{d}\bar{\pi}_\Gamma, \tag{98}$$

where

$$\xi = \frac{1}{n^2}\sum_{i<j} \phi_i\phi_j, \tag{99}$$

the functions $(\phi_i)_{i=1}^n$ being the solutions to the Poisson equations (93).

Remark 33. Note that in the case of the trivial coupling $\Gamma = \mathbf{0}$ (see Example 10), we have that

$$\int_{\bar{E}} \bar{\mathcal{L}}_0 \xi\,\mathrm{d}\bar{\pi}_0 = 0, \tag{100}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

since $\bar{\pi}_0$ is the invariant measure associated to the process generated by $\bar{\mathcal{L}}_0$. Therefore in this case, we obtain the result that the asymptotic variance σ_F^2 is given by the arithmetic mean of the asymptotic variances $\sigma_{f_i}^2$ of the one-particle processes, divided by n. This result is expected, since the computational cost of computing the evolution of the processes $(X_t^i)_{t\geq 0}$ is likewise increased by a factor of n.

Remark 34. Observe furthermore that

$$\int_{\bar{E}} \bar{\mathcal{L}}_\Gamma \xi \, \mathrm{d}\bar{\pi}_\Gamma = 0; \tag{101}$$

hence, using $\bar{\mathcal{L}}_\Gamma = \bar{\mathcal{L}}_0 + \Gamma$ we can equivalently express the asymptotic variance as

$$\sigma_F^2 = \frac{1}{n^2} \sum_{i=1}^n \sigma_{f_i}^2 + \int_{\bar{E}} \Gamma \xi \, \mathrm{d}\bar{\pi}_\Gamma \tag{102}$$

Proof of Theorem 32. First observe that by the fact that $\bar{\pi}_\Gamma$ is a coupling of $(\pi_i)_{i=1}^n$, we have that $\bar{\pi}_\Gamma(F) = 0$. The Poisson equation

$$-\bar{\mathcal{L}}_\Gamma \Phi = F, \quad \bar{\pi}_\Gamma(\Phi) = 0 \tag{103}$$

has a solution given by

$$\Phi = \frac{1}{n} \sum_{i=1}^n \phi_i, \tag{104}$$

resting on the fact that $\Gamma\Phi = 0$ by the first condition of Definition 6. Again, the condition $\bar{\pi}_\Gamma(\Phi) = 0$ is satisfied by the coupling property of $\bar{\pi}_\Gamma$. Using [8, Theorem 2.1], we see that the central limit theorem (97) holds with asymptotic variance

$$\sigma_F^2 = \langle F, \Phi \rangle_{L^2(\bar{\pi}_\Gamma)}. \tag{105}$$

Expanding the above yields

$$\begin{aligned} \sigma_F^2 &= \int_{\bar{E}} \left(\frac{1}{n} \sum_{i=1}^n f_i \right) \left(\frac{1}{n} \sum_{j=1}^n \phi_j \right) \mathrm{d}\bar{\pi}_\Gamma = \frac{1}{n^2} \sum_{i=1}^n \int_{\bar{E}} f_i \phi_i \, \mathrm{d}\bar{\pi}_\Gamma + \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^n \int_{\bar{E}} f_i \phi_j \, \mathrm{d}\bar{\pi}_\Gamma \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma_{f_i}^2 - \int_{\bar{E}} \bar{\mathcal{L}}_0 \xi \, \mathrm{d}\bar{\pi}_\Gamma, \end{aligned} \tag{106}$$

where in the last equation we used the fact that $\bar{\pi}_\Gamma$ has marginal π_i in the ith coordinate, expression (95), as well as the definition of ξ in (99). ∎

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

4.2. Connections to the theory of optimal transportation. In this section we will always assume that Assumption 3 is satisfied, so that the central limit theorems from the previous section hold. Theorem 32 then shows that, in order to reduce the asymptotic variance, we are led to the problem of minimizing the expression

$$\int_{\bar{E}} (-\bar{\mathcal{L}}_0 \xi)\, \mathrm{d}\bar{\pi}_\Gamma.$$

Remarkably, this expression depends on Γ through the measure $\bar{\pi}_\Gamma$ only.[9] We provide a sketch of this situation in Figure 1. Since $\bar{\pi}_\Gamma$ has fixed marginals (i.e., they do not depend on Γ), this task is very reminiscent of the Kantorovich problem [94, Chapter 1] appearing in the theory of optimal transportation [92, 94]. To make this connection more precise, let us introduce the following terminology.

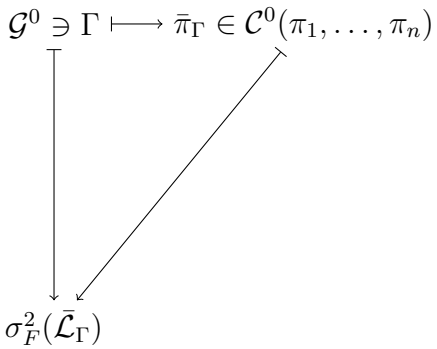


Figure 1. *Relationship between ergodic coupling operators, admissible couplings between the marginal invariant measures, and the associated asymptotic variance. The diagram commutes, in particular, all the information relevant for computing the asymptotic variance is contained in the invariant measure.*

Definition 35 (admissible couplings). *The set of couplings of the marginal invariant measures $(\pi_i)_{i=1}^n$ will be denoted by $\mathcal{C}$. A coupling $\bar{\pi} \in \mathcal{C}$ is called* admissible *if it arises as the invariant measure of an ergodically coupled process, i.e., if there exists $\Gamma \in \mathcal{G}^0$ such that*

$$\int_{\bar{E}} \bar{\mathcal{L}}_\Gamma f\, \mathrm{d}\bar{\pi} = 0, \tag{107}$$

for all $f \in \mathcal{D}(\bar{\mathcal{L}}_\Gamma)$. The set of admissible measures will be denoted by $\mathcal{C}^0$ or, stressing the dependence on the marginal measures, by $\mathcal{C}^0(\pi_1, \dots, \pi_n)$.

Our aim in this section can be summarized in the following form, only replacing $\mathcal{C}$ by the subset $\mathcal{C}^0$ in the standard formulation of the Kantorovich problem.

Problem 1. *For a fixed cost function $c \in C_b(\bar{E})$, find $\bar{\pi} \in \mathcal{C}^0$ such that*

$$\bar{\pi} \in \operatorname*{arg\,min}_{\bar{\pi} \in \mathcal{C}^0} \int_{\bar{E}} c\, \mathrm{d}\bar{\pi}. \tag{108}$$

[9] Another way of saying this is that the map $\Gamma \mapsto \sigma_F^2$ factors through the map $\Gamma \mapsto \bar{\pi}_\Gamma$, i.e., with respect to the asymptotic variance, no information is lost by considering only the invariant measure of the joint process.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Equivalently, find minimizers of the function

$$\mathcal{G}^0 \to \mathbb{R}, \quad \Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma. \tag{109}$$

Remark 36. For later reference, let us recall that the "usual Kantorovich problem" is obtained by considering all possible couplings $\bar{\pi} \in \mathcal{C}$ in (108); see [94, Chapter 1].

Remark 37. As already pointed out, setting $c = -\bar{\mathcal{L}}_0 \xi$, with ξ as defined in (99), is equivalent to the problem of optimizing the asymptotic variance for a particular observable. Other choices for c might be of interest. For instance, one might aim to optimize the asymptotic variance across a set of observables simultaneously. In this case, it seems reasonable to consider cost functions of the form $c = \sum_j c_j$, where c_j is the cost function associated with the jth observable. Assuming that all the particles evolve in the same state space E, another natural objective would be to maximize the average distance of the particles at equilibrium, leading to a cost function of the type

$$c(x_1, \dots, x_n) = -\sum_{\substack{i,j=1 \\ i \neq j}}^{n} d(x_i, x_j) \tag{110}$$

for some metric d on E. More generally, for some function $g : \mathbb{R} \to \mathbb{R}$ it might be worthwhile to consider

$$c(x_1, \dots, x_n) = -\sum_{\substack{i,j=1 \\ i \neq j}}^{n} g(d(x_i, x_j)). \tag{111}$$

A cost function of this type would be reasonable if one aims to use the empirical measure of an ensemble of particles in order to precondition the dynamics (see [59]), in which case the particles should neither be too close nor too far away from each other. We emphasize that since in our framework the marginal processes are held fixed, our results are not directly applicable to the algorithm presented in [59]. However, we expect that the results might be generalized to this context.

Let us also remark that the assumption $c \in C_b(\bar{E})$ is mainly for technical convenience, and both the continuity and the boundedness assumption can probably be weakened in applications by performing appropriate density or truncation arguments. Since we are interested in situations where the process $(\bar{X})_{t \geq 0}$ takes values in a compact set with high probability (i.e., the target measures π_i are concentrated in a compact set), boundedness of c is mainly a technical restriction.

Remark 38. Consider the scenario where one is interested in a target measure $\pi \equiv \pi_1$, and some approximate target $\pi_2 \approx \pi_1$ can be constructed with respect to which expectations are easy to compute. Given two ergodic Markov processes $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$ with invariant measures π_1 and π_2, respectively, the authors of [73] (see also [41]) proposed to couple $(X_t^1)_{t \geq 0}$ and $(X_t^2)_{t \geq 0}$ in such a way that their samples are highly correlated and $(X_t^2)_{t \geq 0}$ can be used as a control variate for $(X_t^1)_{t \geq 0}$. In this context it might be possible to frame the task of finding appropriate couplings in terms of Problem 1, with cost function (for instance) $c(x_1, x_2) = (x_1 - x_2)^2$. We leave exploration of this connection for future work.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 39. Clearly, it holds that

$$\inf_{\bar{\pi}\in\mathcal{C}^0}\int_{\bar{E}} c\,\mathrm{d}\bar{\pi} \geq \inf_{\bar{\pi}\in\mathcal{C}}\int_{\bar{E}} c\,\mathrm{d}\bar{\pi}, \tag{112}$$

so the solutions to the usual optimal transport problems provide lower bounds for Problem 1.

In the Kantorovich formulation, the cost function c is often induced by a distance (for instance, $c(x,y) = (d(x,y))^p$ for $n = 2$, $1 \leq p < \infty$), penalizing couplings that put probability mass on pairs of points (x,y) where x and y are far apart from each other (hence the name optimal transport). In the setting of MCMC (in particular in the context of variance reduction), it is plausible to encourage the particles to stay away from each other, leading to sample diversity and improved exploration of the state space. In this respect, our setting bears certain similarities with the use of optimal transport problems in functional density theory; see [19].

The set $\mathcal{C}^0$ depends on the generators $\mathcal{L}_i$. Furthermore, $\mathcal{C}^0$ is a strict subset of $\mathcal{C}$. The support of an ergodic invariant measure for a Markov process with continuous paths is necessarily connected, for instance, while in general the support of a coupling is not. The following example illustrates that $\mathcal{C}^0$ is indeed usually significantly smaller than $\mathcal{C}$.

Example 40 (the set $\mathcal{C}^0$ contains only few singular measures). Consider the setting from section 3.1.1, i.e., two particles moving in one dimension according to overdamped Langevin dynamics. Let us fix a coupling $\Gamma \in \mathcal{G}^0$ and assume that the invariant measure $\bar{\pi}_\Gamma$ is supported on the zero set of a smooth function $H : \bar{E} \to \mathbb{R}$ with nowhere vanishing gradient, i.e.,

$$\operatorname{supp}\bar{\pi}_\Gamma \subseteq \{(x,y) \in \bar{E} : \quad H(x,y) = 0\}. \tag{113}$$

This implies that $\bar{\pi}_\Gamma$ is supported on a submanifold of $\bar{E}$ and is hence necessarily singular with respect to the Lebesgue measure. Frequently, optimizers of standard optimal transport problems are of this type (see, for instance, [66, Theorem 1.2]). Itô's formula implies that

$$H(\bar{X}_t) = H(\bar{X}_0) + \int_0^t (G^T\nabla H)(\bar{X}_s)\cdot \mathrm{d}W_s + \int_0^t (\bar{\mathcal{L}}_\Gamma H)(\bar{X}_s)\,\mathrm{d}s, \quad t \geq 0, \tag{114}$$

where G is given in (53). Choosing the initial condition $\bar{X}_0 \sim \bar{\pi}_\Gamma$ results in $H(\bar{X}_t) = H(\bar{X}_0) = 0$ almost surely, for all $t \geq 0$. It then follows that both of the remaining integral terms individually have to be zero (owing to the decomposition into martingale and bounded variation part). The quadratic variation of the martingale part is given by

$$\int_0^t w(\bar{X}_s)\,\mathrm{d}s, \quad w = (\partial_x H)^2 + 4\cos\beta\sin\beta\cdot(\partial_x H)(\partial_y H) + (\partial_y H)^2. \tag{115}$$

Since the quadratic variation has be to be zero for all $t \geq 0$, it follows that $\partial_x H = \pm\partial_y H$ on $\operatorname{supp}\bar{\pi}_\Gamma$. Since H is smooth with nonvanishing gradient, it turns out that $\operatorname{supp}\bar{\pi}_\Gamma$ is contained in either one of the diagonals $x = y$ or $x = -y$, in fact either $\bar{\pi}_\Gamma = \frac{1}{Z}e^{-V(x)}\delta_{x-y}(\mathrm{d}x\mathrm{d}y)$ or $\bar{\pi}_\Gamma = \frac{1}{Z}e^{-V(x)}\delta_{x+y}(\mathrm{d}x\mathrm{d}y)$, noting that the latter is only possible if V has the symmetry property $V(x) = V(-x)$. We conclude that, at least in the example considered here, $\mathcal{C}^0$ contains only very few singular measures.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

From the theory of optimal transportation it is known that solutions of the Kantorovich problem (see Remark 36) are typically quite singular, in the sense that they are supported on small sets (see, for instance, [66, Theorem 1.2]). As Example 40 shows, these measures often do not belong to $\mathcal{C}^0$. The aim of this section is to show a similar singularity property for Problem 1. Informally speaking, we will see that under reasonable conditions, the optimizers of (109) are not attained for coupling operators in the interior of $\mathcal{G}^0$. To make this statement precise, let us fix the decompositions

$$\mathcal{L}_i = -\sum_{k=1}^{K_i} (A_k^i)^* A_k^i + B^i \tag{116}$$

of the underlying generators (see section 2.4) and lay the focus on coupling operators of the form (36), denoting this set by $\mathcal{G}(A)$:

$$\mathcal{G}(A) = \left\{ \Gamma = \sum_{(i,j,k,l)\in\mathcal{J}} \alpha_{ijkl} A_k^i A_l^j : \quad \alpha_{ijkl} : \bar{E} \to \mathbb{R} \right\} \cap \mathcal{G}, \tag{117}$$

where we recall the set $\mathcal{J}$ of admissible indices, defined in (37). We wish to stress, however, that the distinction between $\mathcal{G}(A)$ and $\mathcal{G}$ is often obsolete (see Proposition 20). The subset of ergodic coupling operators will similarly be denoted by $\mathcal{G}^0(A)$. Let us now introduce the "tangent space" to $\mathcal{G}(A)$:

$$T\mathcal{G}(A) = \left\{ \sum_{(i,j,k,l)\in\mathcal{J}} \alpha_{ijkl} A_k^i A_l^j \,|\, \alpha_{ijkl} \in C_c^\infty(\bar{E}) \right\}. \tag{118}$$

Remark 41. The definition of $T\mathcal{G}(A)$ encapsulates the first condition of Definition 6 in the sense that elements of $T\mathcal{G}(A)$ vanish on functions that only depend on one variable, whereas the second condition is not accounted for.

In order to state our main result, we need the following definition.

Definition 42 (interior points). *An operator $\Gamma \in \mathcal{G}^0(A)$ is called an* interior point *if*

1. *for all $\mathrm{d}\Gamma \in T\mathcal{G}(A)$ there exists $C > 0$ such that $\bar{\mathcal{L}}_\Gamma + \varepsilon \mathrm{d}\Gamma \in \mathcal{G}^0(A)$ for all $\varepsilon \in (-C, C)$,*
2. *the operator $\bar{\mathcal{L}}_\Gamma$ is invertible on $L_0^2(\bar{\pi}_\Gamma)$,*
3. *$\mathrm{d}\Gamma \bar{\mathcal{L}}_\Gamma^{-1} f \in C(\bar{E})$ for all $\mathrm{d}\Gamma \in T\mathcal{G}(A)$ and all $f \in C(\bar{E}) \cap L_0^2(\bar{\pi}_\Gamma)$.*

Remark 43. The first condition is the essence of the foregoing definition, describing the geometric intuition of interior points. The third condition is mostly technical, since in applications $\mathcal{L}_\Gamma^{-1}$ usually possesses sufficient smoothing properties in order for the composition $\mathrm{d}\Gamma \mathcal{L}_\Gamma^{-1}$ to preserve continuity.

Remark 44 (Lyapunov functions). It is possible and often convenient to replace the second condition by the weaker requirement that invertibility holds on a suitable subspace $\mathcal{V}_\Gamma$ of $L_0^2(\bar{\pi}_\Gamma)$. Our results in this section will then continue to hold, provided that the cost function c satisfies $c - \bar{\pi}_\Gamma(c) \in \mathcal{V}_\Gamma$ for all interior points Γ. As an example, assume that there exist Lyapunov functions $\mathcal{K}_i : E_i \to [1, \infty)$ for the one-particle dynamics, i.e.,

$$\mathcal{L}_i \mathcal{K}_i \leq -a_i \mathcal{K}_i + b_i \tag{119}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for suitable constants $a_i > 0$, $b_i \geq 0$. Defining $\bar{\mathcal{K}} = \sum_{i=1}^n \mathcal{K}_i$, it follows immediately from $\Gamma\bar{\mathcal{K}} = 0$ that $\bar{\mathcal{K}}$ is a Lyapunov function for $\bar{\mathcal{L}}_\Gamma$, independently of the coupling operator Γ. Under certain minorization (irreducibility) conditions (see [42], [61, Chapter 2.4]), one can show that $\mathcal{L}_\Gamma^{-1}$ is invertible on

$$\mathcal{V}_\Gamma = \left\{ f \in L_0^2(\bar{\pi}_\Gamma) : \left\| \frac{f}{\bar{\mathcal{K}}} \right\|_\infty < \infty \right\}. \tag{120}$$

See also [39, Theorem 3.2].

Example 45. In the setting of section 3.1.1 (overdamped Langevin dynamics), it is straightforward to see that Γ as defined in (48) satisfies condition 1 of Definition 42 if and only if $-1 < \alpha(x,y) < 1$ for all $(x,y) \in \mathbb{R}^2$, i.e., if and only if the bound (49) is strict. More generally, Γ as defined in (59) satisfies condition 1 if and only if the matrix Q as defined in (60) is (strictly) positive definite pointwise. Those conditions are clearly equivalent to the (pointwise) ellipticity of the corresponding generators $\bar{\mathcal{L}}_\Gamma$. In the case of underdamped Langevin dynamics (section 3.2), analogous statements are valid. Similarly, couplings of zigzag processes (section 3.3) satisfy condition 1 if and only if the bound (92) is strictly satisfied.

For our further discussion, we will need the following derivative formula.

Proposition 46. *Let $\Gamma \in \mathcal{G}^0(A)$ be an interior point, $\mathrm{d}\Gamma \in T\mathcal{G}(A)$, and consider the family of operators $\bar{\mathcal{L}}_\Gamma + \varepsilon \mathrm{d}\Gamma \in \mathcal{G}^0$ for ε small enough. Let the associated family of invariant measures be denoted by $\bar{\pi}_\Gamma^\varepsilon$, and fix $c \in C_b(\bar{E})$. Then the function $\varepsilon \mapsto \int_{\bar{E}} c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon$ is differentiable at $\varepsilon = 0$, and the derivative is given by*

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \left(\int_{\bar{E}} c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon \right) = -\int_{\bar{E}} c \left[\bar{\mathcal{L}}_\Gamma^*\right]^{-1} (\mathrm{d}\Gamma^* \mathbf{1}) \,\mathrm{d}\bar{\pi}_\Gamma^0, \tag{121}$$

where the adjoints are taken in $L^2(\bar{\pi}_\Gamma^0)$.

Proof. The proof can essentially be found in [58] (see also [61, Remark 5.5]) and is given for completeness in Appendix B. ∎

Remark 47. Notice that our notation entails that $\bar{\pi}_\Gamma \equiv \bar{\pi}_\Gamma^0$. Moreover, the right-hand side of (121) is well defined. Indeed, by the second condition in Definition 42, $\left[\bar{\mathcal{L}}_\Gamma^*\right]^{-1}$ is well defined on $L_0^2(\bar{\pi}_\Gamma^0)$, and furthermore $\operatorname{Ran} \mathrm{d}\Gamma^* \subseteq L_0^2(\bar{\pi}_\Gamma^0)$ due to

$$\int_{\bar{E}} \mathrm{d}\Gamma^* f \,\mathrm{d}\bar{\pi}_\Gamma^0 = \int_{\bar{E}} (\mathrm{d}\Gamma \mathbf{1}) f \,\mathrm{d}\bar{\pi}_\Gamma^0 = 0, \quad f \in C_c^\infty(\bar{E}), \tag{122}$$

using that $\mathrm{d}\Gamma \mathbf{1} = 0$ according to Lemma 18.

For an interior point $\Gamma \in \mathcal{G}^0(A)$ and $\mathrm{d}\Gamma \in T\mathcal{G}(A)$ let us introduce the suggestive notation

$$\frac{\mathrm{d}}{\mathrm{d}\Gamma} \int_{\bar{E}} c \,\mathrm{d}\bar{\pi}_\Gamma := \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \left(\int_{\bar{E}} c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon \right), \tag{123}$$

where, as in Proposition 46, $\bar{\pi}_\Gamma^\varepsilon$ denotes the family of invariant measures associated to the operators $\bar{\mathcal{L}}_\Gamma + \varepsilon \mathrm{d}\Gamma \in \mathcal{G}^0$ for ε small enough. Moreover, we will make use of the following terminology.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Definition 48 (critical points). *Let $\Gamma \in \mathcal{G}(A)$ be an interior point. Then Γ is called* critical *if*

$$\frac{\mathrm{d}}{\mathrm{d}\Gamma} \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma = 0 \tag{124}$$

for all $\mathrm{d}\Gamma \in T\mathcal{G}(A)$.

In our aim to find minimizers of the function $\Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$, it is natural to seek critical points. The following is our main result in this section.

Theorem 49. *Let $c \in C_b(\bar{E})$. Then either all interior points are critical, or no interior point is critical.*

Example 50. Let c be of the form

$$c(x_1, \ldots, x_n) = g_1(x_1) + \ldots + g_n(x_n) \tag{125}$$

for appropriate functions $g_i : E_i \to \mathbb{R}$. Then, since $\bar{\pi}_\Gamma$ is a coupling of the fixed marginals $(\pi_i)_{i=1}^n$, the function $\Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma = \sum_{i=1}^n \int_{E_i} g_i \, \mathrm{d}\pi_i$ is constant, and hence all interior points are critical.

Before proceeding to the proof of the theorem, let us give a few remarks.

Remark 51. Informally, Theorem 49 states that the mapping $\mathcal{G}^0(A) \ni \Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ is either locally constant or does not attain its extrema on interior points. In other words, if $\Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ is not constant, then its extrema lie "at the boundary" of $\mathcal{G}^0(A)$, although we have not rigorously defined this term, and moreover, $\int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ is not even well defined for nonergodic couplings.

Remark 52. A striking consequence of Theorem 49 is that under mild conditions, independent coupling (associated to $\mathbf{0} \in \mathcal{G}^0$) of overdamped or underdamped Langevin dynamics is not optimal for *any* criterion of the form $\int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$. Theorem 49 complements results from the theory of optimal transportation that state that optimal couplings are generically singular in terms of their support. Indeed, considering the example of overdamped or underdamped Langevin dynamics, the "boundary of $\mathcal{G}^0(A)$" consists of couplings that lead to degenerately elliptic generators that are in general not hypoelliptic. In particular, the corresponding invariant measures are not in general absolutely continuous with respect to the Lebesgue measure.

Remark 53. Theorem 49 also supports the folklore that optimal MCMC samplers use as little noise as possible to guarantee ergodicity, as degenerately elliptic operators correspond to dynamics where noise only acts in certain directions. For example, it is by now well documented that nonreversible samplers outperform their reversible counterparts in various settings (see, for instance, [28, 47, 75, 82]). The process of making a reversible sampler nonreversible can be thought of informally as decreasing the ratio between random and deterministic behavior.

Remark 54. Let us examine the function $\mathcal{G}^0(A) \ni \Gamma \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\Gamma$ along a ray. More precisely, fix $\mathrm{d}\Gamma \in T\mathcal{G}(A)$, set $\bar{\mathcal{L}}_\varepsilon := \bar{\mathcal{L}}_0 + \varepsilon \mathrm{d}\Gamma$ for ε small enough, and consider the function $\varepsilon \mapsto \int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\varepsilon$, where $(\bar{\pi}_\varepsilon)_\varepsilon$ denotes the corresponding family of invariant measures. Since $\mathrm{d}\Gamma$ is relatively bounded with respect to $\bar{\mathcal{L}}_0$ in $L^2(\bar{\pi}_0)$, we have the following Neumann power expansion for ε small enough:

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$\int_{\bar{E}} c \, \mathrm{d}\bar{\pi}_\varepsilon = \int_{\bar{E}} c \left(1 + \sum_{j=1}^{\infty} (-\varepsilon)^j [(\mathrm{d}\Gamma \bar{\mathcal{L}}_0^{-1} \Pi^\perp)^*]^j \right) \mathbf{1} \, \mathrm{d}\bar{\pi}_0, \tag{126}$$

where $\Pi^\perp$ is the orthogonal projection onto $L_0^2(\pi)$ in $L^2(\pi)$. For details, see [61, Theorem 5.2]. The factor of $(-\varepsilon)^j$ in expression (126) signals oscillatory behavior, and indeed it is straightforward to construct examples (for instance, in the Gaussian case), where (126) exhibits multiple local minima and maxima as a function of ε (see, for instance, the graph related to mirror coupling in Figure 3(c) below). This finding is not in contradiction with Theorem 49. Indeed, as Theorem 49 shows, at those extrema there are directions of ascent (or descent) in $T\mathcal{G}(A)$ not aligned with the considered ray, and thus, those extrema turn out not to be critical when considered in the whole of $\mathcal{G}^0(A)$.

Let us now prove Theorem 49 and start with the following key lemma. Its significance derives from the fact that the second statement manifestly does not depend on Γ.

Lemma 55. *Let $c \in C_b(\bar{E}) \cap L_0^2(\bar{\pi}_0)$ and $\Gamma \in \mathcal{G}^0(A)$ be an interior point. Then the following conditions are equivalent:*

1. *The following holds for all admissible indices $(i,j,k,l) \in \mathcal{J}$:*

$$A_k^i A_l^j \bar{\mathcal{L}}_\Gamma^{-1} \left(c - \bar{\pi}_\Gamma(c) \right) = 0. \tag{127}$$

2. *There exists $f \in \mathcal{D}(\bar{\mathcal{L}}_0)$ such that both of the following hold:*

(a) *for all admissible indices $(i,j,k,l) \in \mathcal{J}$ it holds that*

$$A_k^i A_l^j f = 0, \tag{128}$$

(b)

$$\bar{\mathcal{L}}_0 f = c. \tag{129}$$

Proof. First assume that (127) holds for all $(i,j,k,l) \in \mathcal{J}$. Then setting

$$f = \bar{\mathcal{L}}_\Gamma^{-1} \left(c - \bar{\pi}_\Gamma(c) \right)$$

immediately implies (128). Furthermore, from (128) and (117) it follows that $\Gamma f = 0$, implying $\bar{\mathcal{L}}_\Gamma f = \bar{\mathcal{L}}_0 f$ (in particular $f \in \mathcal{D}(\bar{\mathcal{L}}_0)$), as well as

$$\bar{\mathcal{L}}_0 f = \left(c - \bar{\pi}_\Gamma(c) \right). \tag{130}$$

Equation (130) clearly implies that $\bar{\pi}_0(c) = \bar{\pi}_\Gamma(c)$, and hence $\bar{\pi}_\Gamma(c) = 0$, leading to (129).

The reverse implication follows similarly by first observing that (128) and (129) imply that $\bar{\mathcal{L}}_\Gamma f = c$, and hence $\bar{\pi}_\Gamma(c) = 0$. Combining this with (128) shows that (127) holds. ∎

Proof of Theorem 49. Clearly, we can without loss of generality assume that $\bar{\pi}_0(c) = 0$. According to Definition 48 and Proposition 46, an interior point $\Gamma \in \mathcal{G}^0(A)$ is critical if and only if

$$\int_{\bar{E}} c \left[\bar{\mathcal{L}}_\Gamma^* \right]^{-1} (\mathrm{d}\Gamma^* 1) \, \mathrm{d}\bar{\pi}_\Gamma = 0 \tag{131}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

for all $\mathrm{d}\Gamma \in T\mathcal{G}(A)$, which is equivalent to

$$\sum_{(i,j,k,l)\in\mathcal{J}} \int_{\bar{E}} \left(A^i_k A^j_l \bar{\mathcal{L}}^{-1}_\Gamma \left(c - \bar{\pi}_\Gamma(c)\right)\right) \cdot \alpha_{ijkl}\, \mathrm{d}\bar{\pi}_\Gamma = 0 \tag{132}$$

for all $(\alpha_{ijkl})_{(i,j,k,l)\in\mathcal{J}} \subset C^\infty_c(\bar{E})$. The latter statement is clearly equivalent to the first statement in Lemma 55. The result now follows by noting that the second statement in Lemma 55 does not depend on Γ. ∎

5. A perturbative approach for the study of the asymptotic variance. Informally speaking, Theorem 49 shows that the objective of optimizing the asymptotic variance σ^2_F leads to the requirement that $\Gamma \in \mathcal{G}^0(A)$ should be chosen to be a "boundary point" (see the discussion in Remark 51). While being an interesting theoretical result, it does not give much guidance about how to choose a suitable coupling in practice (after all, both minima and maxima are obtained "at the boundary" of $\mathcal{G}^0(A)$). In this section we therefore develop a perturbative approach, based on operators of the form

$$\bar{\mathcal{L}}^\varepsilon_\Gamma = \bar{\mathcal{L}}_0 + \varepsilon \mathrm{d}\Gamma, \quad \mathrm{d}\Gamma \in T\mathcal{G}(A), \tag{133}$$

for ε small enough. In the following, we will assume that $\varepsilon \in \mathcal{I}$, where $\mathcal{I} \subset \mathbb{R}$ is an appropriate interval such that $\bar{\mathcal{L}}^\varepsilon_\Gamma \in \mathcal{G}^0$ for all $\varepsilon \in \mathcal{I}$. As usual, we consider observables of the form $F = \frac{1}{n}\sum_{i=1}^n f_i$, for some $f_i \in L^2_0(\pi_i)$, and suppose that Assumption 3 is satisfied. To stress the dependence of the asymptotic variance on the parameter ε we will write $\sigma^2_F(\varepsilon)$. Note that a similar setting has already been considered in Remark 54. There, we investigated the dependence of the asymptotic variance (or more generally, of the quantity $\int_{\bar{E}} c\, \mathrm{d}\bar{\pi}_\varepsilon$) on the parameter ε. Here, we are rather interested in the choice of the "direction" $\mathrm{d}\Gamma \in T\mathcal{G}(A)$, starting from the trivial (independent) coupling $\bar{\mathcal{L}}_0$.

Combining the expression (98) with either (126) or (121) we see that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\sigma^2_F\big|_{\varepsilon=0} = \int_{\bar{E}} \mathrm{d}\Gamma\xi\, \mathrm{d}\bar{\pi}_0, \tag{134}$$

where ξ is given by

$$\xi = \frac{1}{n^2}\sum_{i<j} \phi_i\phi_j, \tag{135}$$

in terms of the solutions to the Poisson equations (93). The benefit of (134) is that its right-hand side consists of expressions that are known in principle, as the measure $\bar{\pi}_0$ is given by the product $\bar{\pi}_0 = \bigotimes_{i=1}^n \pi_i$. It therefore serves as a starting point for finding a suitable coupling operator $\Gamma \in \mathcal{G}$. Let us summarize our approach in this section in the following form.

Problem 2. *Given invariant measures $\pi_i \in \mathcal{P}(E_i)$ and observables $f_i \in L^2_0(\pi_i)$, find a coupling operator $\Gamma \in \mathcal{G}$ such that*

$$\int_{\bar{E}} \Gamma\xi\, \mathrm{d}\bar{\pi}_0 \tag{136}$$

is minimized.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Problem 2 can be thought of in two different ways: Firstly, it can be interpreted as a linearization of Problem 1. Indeed, (136) depends linearly on Γ, whereas (109) is highly nonlinear (for an illustration of this fact, see the power expansion (126)). Another way of seeing this is by noting the similarity between (136) and the second term appearing on the right-hand side of (102). Not surprisingly, Problem 2 turns out to be much easier to (approximately) solve in practice. Note that by linearity, properties similar to the one expressed in Theorem 49 hold for Problem 2 (at least if $\mathcal{G}$ is convex). Choosing a coupling $\Gamma \in \mathcal{G}$ according to the formulation of Problem 2 is clearly heuristic. However, we have had good results with it in numerical experiments (see below).

Secondly, when a solution of Problem 2 is available, it is reasonable in practice to only implement a small perturbation of the independent sampler (i.e., choose ε to be small). Such a choice will not be optimal over all couplings in $\mathcal{G}$ according to Theorem 49. However, it is then guaranteed that the performance of the sampler is at least slightly improved. Let us also note that the formulation of Problem 2 does not require the coupling to be ergodic, as opposed to the formulation of Problem 1.

The aim of this section is to analyze Problem 2 for some of the examples presented in section 3 and to present some numerical experiments. To this end, let us introduce the shorthand notation

$$\delta\sigma_F^2(\Gamma) := \int_{\bar{E}} \Gamma\xi \, \mathrm{d}\bar{\pi}_0, \tag{137}$$

stressing the infinitesimal (approximate) nature of the objective in Problem 2. In what follows, Γ will be given in terms of a function α, belonging to a set $\mathcal{A}$. To emphasize this dependence we will write Γ_α. We will not impose regularity constraints on the function α (beyond measurability), so that the operators Γ_α will in general not induce couplings that satisfy the Feller property (see Remark 21) and hence strictly speaking do not belong to $\mathcal{G}$. We mention that in particular, the optimal couplings obtained in the following sections turn out to introduce discontinuities that are not covered by our general theory but pose no difficulties for the numerics and work well in the experiments.

5.1. Overdamped Langevin dynamics in one dimension with two particles. Consider the setting from the example presented in section 3.1.1. Then, (137) takes the form

$$\delta\sigma_F^2(\Gamma_\alpha) = \int_{\mathbb{R}^2} \alpha(x,y)\phi'(x)\phi'(y)e^{-(V(x)+V(y))} \, \mathrm{d}x\mathrm{d}y, \tag{138}$$

where ϕ is the solution to the Poisson equation

$$-\left(-V'\phi' + \phi''\right) = f, \quad \pi(\phi) = 0, \tag{139}$$

and $f \in L_0^2(\pi)$ is an observable of interest. Furthermore, Γ is given as in (48) with

$$\alpha \in \mathcal{A} := \{\alpha : \mathbb{R}^2 \to \mathbb{R} \,|\, \alpha \text{ measurable}, \ -1 \le \alpha \le 1\}. \tag{140}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Recall from section 3.1.1 that $\alpha \in \mathcal{A}$ induces a well-defined coupled process (Lemma 22) that, however, does not satisfy the Feller property in general (further regularity assumptions would be required). The following optimality result is immediate from an inspection of (138).

Proposition 56. *Let $\alpha^* \in \mathcal{A}$ be given by*

$$\alpha^*(x,y) = \begin{cases} 1 & \text{if } \phi'(x)\phi'(y) \le 0, \\ -1 & \text{if } \phi'(x)\phi'(y) > 0. \end{cases} \tag{141}$$

Then Γ_{α^*} solves Problem 2 in the sense that $\delta\sigma_F^2(\Gamma_{\alpha^*}) \le \delta\sigma_F^2(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

Remark 57. Clearly, we have

$$\int_{\bar{E}} \Gamma_{\alpha^*} \xi \, \mathrm{d}\bar{\pi}_\Gamma \le 0 \tag{142}$$

for α^* as defined in (141). We stress the difference between this expression and (138), where we compute the same integral but with respect to $\bar{\pi}_0$. By comparison with (102), we see that $\sigma_F^2(\Gamma_{\alpha^*}) \le \sigma_F^2(\mathbf{0})$, i.e., Γ_{α^*} always improves on independent coupling. However, we do not know whether Γ_{α^*} is optimal in the sense of Problem 1.

It is instructive to compare the solution of Problem 2 found in Proposition 56 to the solution of the usual Kantorovich problem (see Remark 36), the corresponding cost function $c = -\bar{\mathcal{L}}_0\xi$ being given in section 4.2.[10] Recall that Problem 2 can be considered to be a linearization of Problem 1, which in turn is related to the Kantorovich problem in the sense that the minimization is carried out over a smaller set of couplings (namely, those couplings that are invariant measures of coupled processes; see Definition 35). For our experiments, we choose the quadratic potential $V(x) = \frac{1}{2}x^2$, i.e., the task of sampling from a Gaussian measure. Furthermore, we consider the linear observable $f_1(x) = x$, the quadratic observable $f_2(x) = x^2$, and the "mixed" observable $f_3(x) = x^2 - x$. In Figure 2, we plot the invariant measures[11] of the coupled processes induced by (141) (left-hand side) and compare them to the solutions of the Kantorovich problem[12] (right-hand side), with the appropriate cost function $c = -\bar{\mathcal{L}}_0\xi$ as given in section 4.2. As it turns out, the solutions to Problem 2 and the standard Kantorovich problem look remarkably similar (at least in shape). We hence conclude that in the example considered here, Problem 2 is a good approximation of Problem 1, keeping in mind that the solution of the Kantorovich problem provides a lower bound for the objective function of Problem 1 (see Remark 39).

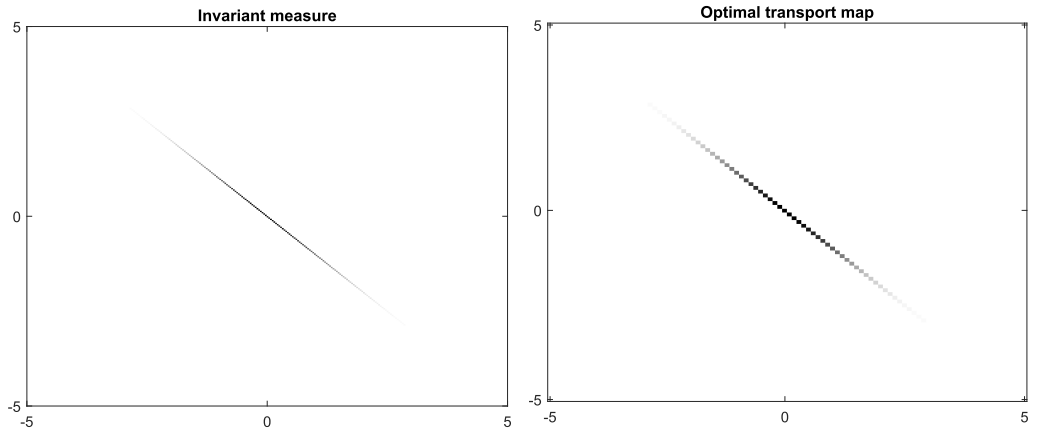
The following lemma serves to examine a few test cases and gain further intuition. For convenience, let us assume that f, and therefore, by elliptic regularity, ϕ, are smooth.

[10] Note that this requires the solution to the Poisson equation (139), as ξ is defined in terms of ϕ. In our case, these solutions can be obtained analytically, for example by expanding f into Hermite polynomials (after adding an appropriate constant ensuring that $\pi(f) = 0$). To be specific, those solutions turn out to be given by $\phi_1(x) = x$, $\phi_2(x) = \frac{1}{2}(x^2 - 1)$ and $\phi_3(x) = \frac{1}{2}(x^2 - 1) - x$.

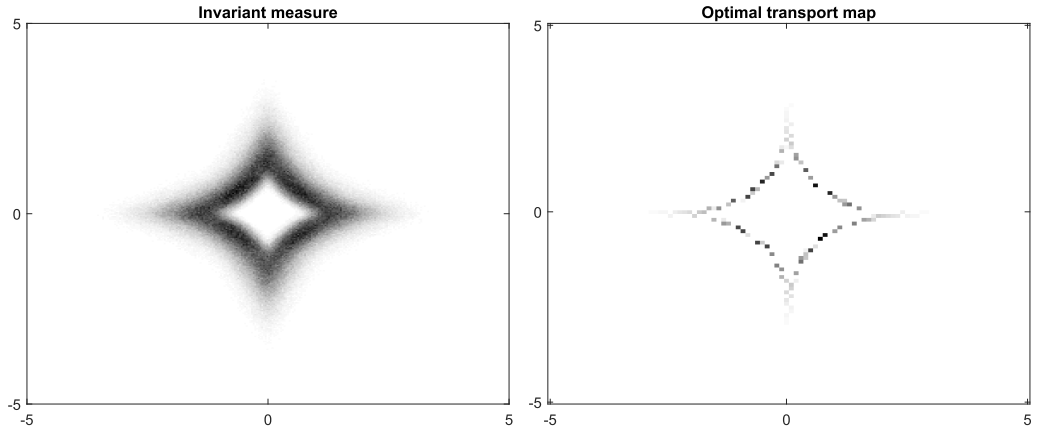
[11] Approximations of the invariant measures have been obtained from samples of longtime simulations of (52) (after a burn-in time), using an Euler–Mayurama discretization with timestep $\Delta t = 10^{-3}$.

[12] The optimal transport map was computed using the Python library POT 0.4.0 (accessible from https://pypi.python.org/pypi/POT/0.4.0) which is based on the algorithm proposed in [12].

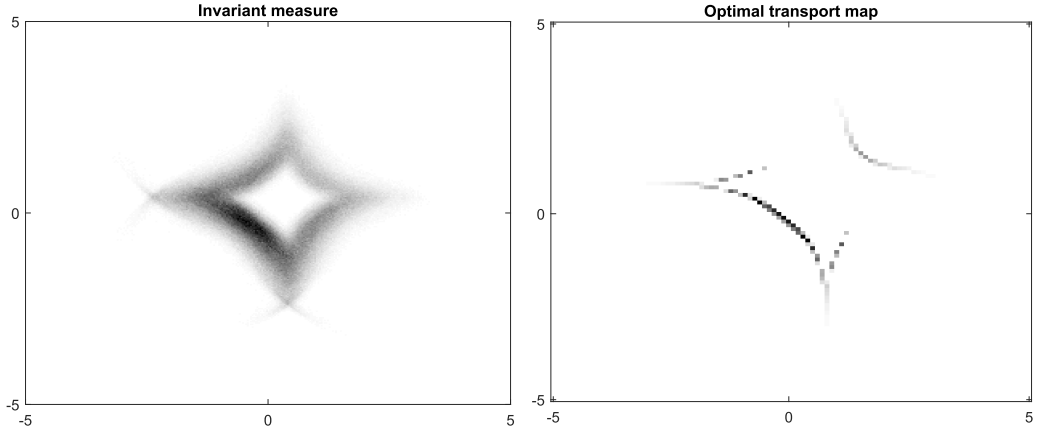
Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.



(a) Linear observable: $f_1(x) = x$.



(b) Quadratic observable: $f_2(x) = x^2$.



(c) Mixed observable: $f_3(x) = x^2 - x$.

Figure 2. *Comparison of invariant measures associated to solutions of Problem 2 and optimal transport maps as solutions to the Kantorovich problem for the example of overdamped Langevin dynamics in dimension one with quadratic potential.*

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Lemma 58. *Let $\phi \in L^2_0(\pi) \cap C^\infty(\mathbb{R})$ solve the Poisson equation (139).*

1. *Assume that $f \in L^2_0(\pi) \cap C^\infty(\mathbb{R})$ is monotonically increasing (decreasing). Then ϕ' is nonnegative (nonpositive).*
2. *Assume that V and f are symmetric, i.e., $V(-x) = V(x)$ and $f(-x) = f(x)$ for all $x \in \mathbb{R}$. Furthermore, let f be monotonically decreasing (increasing) on $(-\infty, 0]$. Then $\phi'(x) \cdot x \leq 0$ ($\phi'(x) \cdot x \geq 0$) for all $x \in \mathbb{R}$.*

The proof can be found in Appendix C. The following two corollaries are direct consequences of Lemma 58 and Proposition 56.

Corollary 59 (mirror coupling). *In the setting from the first part of Lemma 58,*

$$\alpha^* \equiv -1 \tag{143}$$

solves Problem 2, in the sense that $\delta\sigma^2_F(\Gamma_{\alpha^*}) \leq \delta\sigma^2_F(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

Note that the choice $\alpha^* \equiv -1$ corresponds to $\mathrm{d}B^x_t = -\mathrm{d}B^y_t$, hence the name "mirror coupling," which will be used throughout the following discussion. Likewise, we will use the term "symmetric coupling" for the coupling strategy detailed in the following corollary since it is appropriate in the symmetric setting of the second part of Lemma 58.

Corollary 60 (symmetric coupling). *In the setting from the second part of Lemma 58,*

$$\alpha^*(x,y) = \begin{cases} 1 & \text{if } x \cdot y \leq 0, \\ -1 & \text{if } x \cdot y > 0 \end{cases} \tag{144}$$

solves Problem 2, in the sense that $\delta\sigma^2_F(\Gamma_{\alpha^*}) \leq \delta\sigma^2_F(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

A few comments on the findings from Corollaries 59 and 60 are in order. If the observable is monotone (Corollary 59), then it turns out that choosing the mirror coupling $\mathrm{d}B^x_t = -\mathrm{d}B^y_t$ in (54) is optimal in the sense of Problem 2. This result has a clear connection to popular variance reduction techniques such as "antithetic variates" [54, Chapter 9.2], where correlations between random variables are used to produce cancellations. In the case of symmetric observables (Corollary 60), optimal coupling in the sense of Problem 2 leads to a more sophisticated strategy: When the two particles (the locations of which are again denoted by x and y) are on the same side of the potential (meaning that $x \geq 0$ and $y \geq 0$ or $x \leq 0$ and $y \leq 0$), then the Brownian motions should be coupled according to $\mathrm{d}B^x_t = -\mathrm{d}B^y_t$, as in the case of monotone observables. When the particles are on opposite sides ($x \geq 0$ and $y \leq 0$ or $x \leq 0$ and $y \geq 0$), according to Corollary 60 it is best to switch to synchronous coupling, $\mathrm{d}B^x_t = \mathrm{d}B^y_t$. Intuitively this can be understood as follows: By symmetry, the situation where $x \geq 0$ and $y \leq 0$ with synchronous coupling ($\mathrm{d}B^x_t = \mathrm{d}B^y_t$) is equivalent to $x \geq 0$ and $y \geq 0$, with mirror coupling ($\mathrm{d}B^x_t = -\mathrm{d}B^y_t$). Since f is monotone on $[0, \infty)$, this argument provides a plausible explanation for optimality by appealing to Corollary 59. Finally, let us mention that numerical experiments show that using mirror coupling in the case of observables of the type encountered in Corollary 60 ("naive antithetic variates") actually leads to a less effective sampler in terms of the asymptotic variance (see Figure 3(b)).

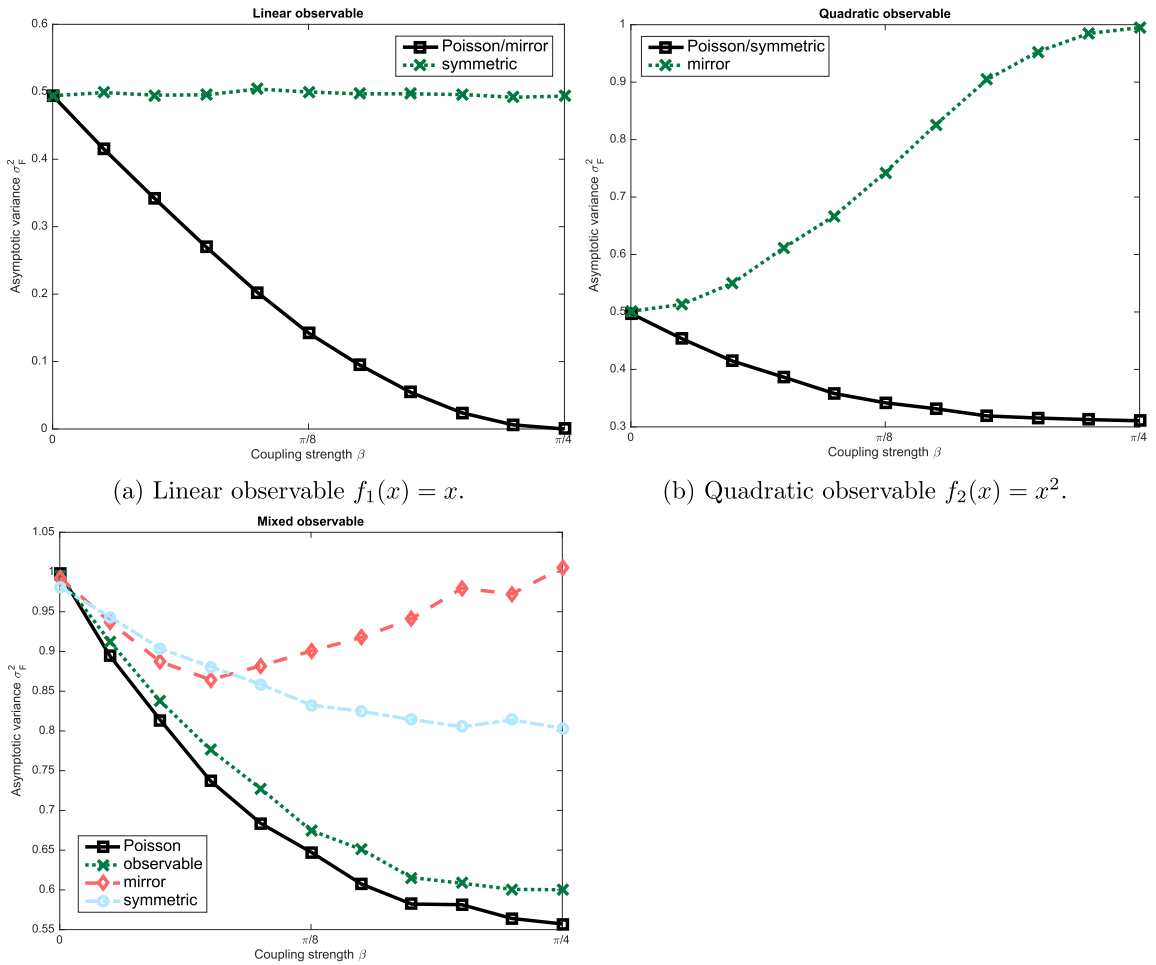
Let us consider now the same set-up as in the numerical experiments presented in Figure 2, i.e., we consider a Gaussian target measure ($V(x) = \frac{1}{2}x^2$), and the observables $f_1(x) = x$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

(linear), $f_2(x) = x^2$ (quadratic), and $f_3(x) = x^2 - x$ (mixed). In Figure 3, we plot[13] the asymptotic variances for f_1, f_2, and f_3, associated to different coupling schemes as a function of the coupling strength β. We employ mirror and and symmetric coupling, as detailed in Corollaries 59 and 60, modulated by β. More precisely, we let

$$\alpha_{\text{mirror}} \equiv 2 \sin \beta \cos \beta \tag{145}$$

and



(a) Linear observable $f_1(x) = x$.

(b) Quadratic observable $f_2(x) = x^2$.

(c) Mixed observable: $f_3(x) = x^2 - x$.

Figure 3. *Dependence of the asymptotic variance on different coupling schemes and coupling strengths. Mirror coupling corresponds to* α_{mirror} *in* (145)*, symmetric coupling to* $\alpha_{\text{symmetric}}$ *in* (146)*, Poisson coupling to* α_{Poisson} *in* (147)*, and observable coupling to* $\alpha_{\text{observable}}$ *in* (148)*. Note that Poisson and mirror coupling coincide for* f_1*, and Poisson and symmetric coupling coincide for* f_2*.*

[13] We estimated the asymptotic variance from $5 \cdot 10^4$ independent realizations of (52), discretized according to the Euler–Mayurama scheme with timestep $\Delta t = 10^{-3}$.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$\alpha_{\text{symmetric}}(x,y) = \begin{cases} 2\sin\beta\cos\beta & \text{if } x\cdot y \le 0, \\ -2\sin\beta\cos\beta & \text{if } x\cdot y > 0, \end{cases} \tag{146}$$

where $\beta \in [0, \frac{\pi}{4}]$ denotes the coupling strength.[14] Furthermore, we consider the optimal Poisson coupling, defined by

$$\alpha_{\text{Poisson}}(x,y) = \begin{cases} 2\sin\beta\cos\beta & \text{if } \phi'(x)\phi'(y) \le 0, \\ -2\sin\beta\cos\beta & \text{if } \phi'(x)\phi'(y) > 0, \end{cases} \tag{147}$$

ϕ being the solution to the Poisson equation (139) for the corresponding observable. For $\beta = 0$, we recover independent coupling, whereas $\beta = \frac{\pi}{4}$ leads to the optimal coupling from Proposition 56. According to Corollaries 59 and 60, Poisson coupling coincides with mirror coupling for f_1 and with symmetric coupling for f_2. For f_3, however, mirror, symmetric and Poisson couplings are different. To illustrate the effect of couplings that are not tailored to the observable of interest, we also plot the asymptotic variances associated to symmetric coupling for f_1, mirror coupling for f_2, and both mirror and symmetric coupling for f_3. For f_3, we furthermore consider a coupling strategy that uses the derivative of the observable instead of the derivative of the solution to the Poisson equation, specifically, the coupling induced by

$$\alpha_{\text{observable}}(x,y) = \begin{cases} 2\sin\beta\cos\beta & \text{if } f'(x)f'(y) \le 0, \\ -2\sin\beta\cos\beta & \text{if } f'(x)f'(y) > 0. \end{cases} \tag{148}$$

The motivation for this is that in applications, the solution to the Poisson equation is often hard to obtain,[15] whereas the gradient of the observable is readily available. By integration by parts we have

$$\int_{\mathbb{R}^d} \nabla f \cdot \nabla \phi \, \mathrm{d}\pi = \int_{\mathbb{R}^d} f^2 \, \mathrm{d}\pi \ge 0, \tag{149}$$

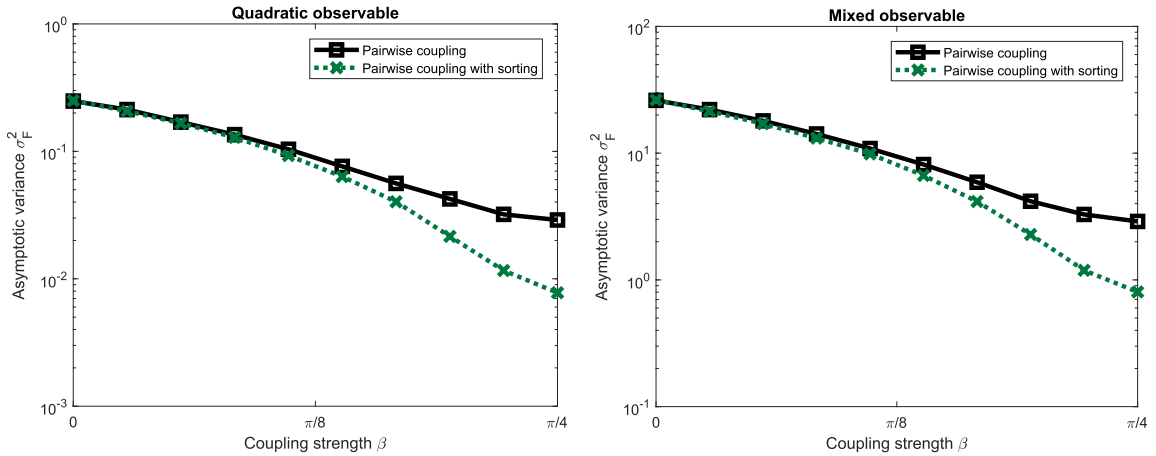
suggesting that we use ∇f as a surrogate for $\nabla\phi$ (at equilibrium, the scalar product of ∇f and $\nabla\phi$ is positive on average).

In all the cases considered, the Poisson coupling turns out be the most efficient, uniformly in the coupling strength β. The fact that the absolute value of the derivative $\frac{\mathrm{d}\sigma_F^2}{\mathrm{d}\beta}|_{\beta=0}$ is maximal for Poisson coupling is precisely the content of Proposition 56, whereas the fact that the asymptotic variance for Poisson coupling is maximal at $\beta = 0$ follows from Remark 57. It is interesting to note the monotonity of the asymptotic variance associated to Poisson coupling with respect to the coupling strength β; this phenomenon is not covered by our theory. Importantly, the efficiency of a certain coupling strongly depends on the considered observable. Indeed, the mirror coupling (which is excellent for the linear observable; see

[14] For $\beta \in [0, \frac{\pi}{4}]$, the function $2\sin\beta\cos\beta$ is monotone, taking values in $[0,1]$. We chose this parametrization in order for it to be consistent with (53).

[15] However, often one aims to approximate the solution to the Poisson equation in order to use it as a control variate; see, for instance, [27, 69, 84]. It suggests itself to use those approaches in conjunction with the coupling strategy developed here.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.



(a) Quadratic observable: $f(x) = \frac{1}{2}|x|^2$.

(b) Mixed observable: $f(x) = 5|x|^2 + l \cdot x$.

Figure 4. *Comparison between pairwise couplings with and without sorting according to $|\nabla\phi|$ for a Gaussian target measure in $d = 10$ dimensions and with $n = 10$ particles.*

Figure 3(a)) leads to an increase of the asymptotic variance for the quadratic observable (see Figure 3(b)). Similarly, the symmetric coupling (suited for the quadratic observable) does not improve the performance for the linear observable (but the performance is also not worsened). In Figure 3(c), we observe that the coupling based on the derivative of the observable (see (148)) works almost as well as the Poisson coupling, so this might be a reasonable choice in applications, although further studies are needed. For a comment about the minimum of the graph associated to mirror coupling for the mixed observable (see Figure 3(c)) we refer to Remark 54.

5.2. Overdamped Langevin dynamics with multiple particles in arbitrary dimensions. Let us extend the discussion from the previous section to arbitrary dimensions, first considering the case of two particles (as done in Example 28). Using the expression (69), we see that

$$\delta\sigma_F^2(\Gamma_\alpha) = \int_{\mathbb{R}^d\times\mathbb{R}^d} (\nabla\phi(y)\cdot\alpha(x,y)\nabla\phi(x))\, e^{-(V(x)+V(y))}\,\mathrm{d}x\mathrm{d}y, \tag{150}$$

where $\alpha : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is a matrix-valued function satisfying (68), i.e.,

$$\alpha \in \mathcal{A} := \left\{\alpha : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d\times d} \text{ measurable, } \alpha(x,y)^T\alpha(x,y) \le I_{d\times d} \text{ for all } x, y \in \mathbb{R}^d\right\}, \tag{151}$$

and ϕ is the solution to the Poisson equation

$$-(-\nabla V\cdot\nabla + \Delta)\phi = f, \quad \pi(\phi) = 0. \tag{152}$$

Since (68) implies

$$|\nabla\phi(y)\cdot\alpha(x,y)\nabla\phi(x)| \le |\nabla\phi(x)||\nabla\phi(y)|, \quad x, y \in \mathbb{R}, \tag{153}$$

we get the following optimality result.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proposition 61. *Assume that $\alpha^* \in \mathcal{A}$ is chosen such that*

$$\nabla\phi(y) \cdot \alpha^*(x,y)\nabla\phi(x) = -|\nabla\phi(x)||\nabla\phi(y)|, \quad x, y \in \mathbb{R}^d. \tag{154}$$

Then α^* solves Problem 2, i.e., $\delta\sigma_F^2(\Gamma_{\alpha^*}) \leq \delta\sigma_F^2(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

In the case when the solution ϕ to the Poisson equation is known it is straightforward to construct a matrix-valued function α^* such that both (68) and (154) are satisfied. For instance, any orthogonal matrix trivially satisfies (68), and (154) can be dealt with by choosing an appropriate rotation or reflection. As an example let us mention the following reflection in the plane spanned by $\nabla\phi(x)$ and $\nabla\phi(y)$:

$$\alpha^*(x,y) = \begin{cases} I_{d\times d} - 2\dfrac{\left(\widehat{\nabla\phi(x)}+\widehat{\nabla\phi(y)}\right)\left(\widehat{\nabla\phi(x)}+\widehat{\nabla\phi(y)}\right)^T}{\left(\widehat{\nabla\phi(x)}+\widehat{\nabla\phi(y)}\right)^2} & \text{if } \nabla\phi(x) \neq 0, \nabla\phi(y) \neq 0, \\ & \widehat{\nabla\phi(x)} + \widehat{\nabla\phi(y)} \neq 0, \\ I_{d\times d} & \text{otherwise.} \end{cases} \tag{155}$$

Here, $\widehat{\nabla\phi} = \frac{\nabla\phi}{|\nabla\phi|}$ is used to denote the normalized gradient of ϕ. As mentioned in section 5.1, the solution to the Poisson equation is usually hard to obtain in applications, but often approximations can be obtained, for instance, by Galerkin methods. Inspired by the integration by parts formula (149), it seems reasonable instead to use the normalized gradient of the observable $\widehat{\nabla f}$ as a surrogate for $\widehat{\nabla\phi}$, i.e.,

$$\alpha_f(x,y) = \begin{cases} I_{d\times d} - 2\dfrac{\left(\widehat{\nabla f(x)}+\widehat{\nabla f(y)}\right)\left(\widehat{\nabla f(x)}+\widehat{\nabla f(y)}\right)^T}{\left(\widehat{\nabla f(x)}+\widehat{\nabla f(y)}\right)^2} & \text{if } \nabla f(x) \neq 0, \nabla f(y) \neq 0, \\ & \widehat{\nabla f(x)} + \widehat{\nabla f(y)} \neq 0, \\ I_{d\times d} & \text{otherwise.} \end{cases} \tag{156}$$

We recall that a comparison of the couplings associated to (155) and (156) was performed in the one-dimensional case (see Figure 3(c)) where α_f almost achieved the same reduction of the asymptotic variance as α^*. Based on (149) and numerical experiments we conjecture that choosing α_f guarantees an improvement in terms of the asymptotic variance for small perturbations.

Conjecture 62. Let $f \in L_0^2(\pi)$ and $\phi \in L_0^2(\pi)$ be the corresponding solution to the Poisson equation (152). Then

$$\delta\sigma_F^2(\Gamma_\alpha) = \int_{\mathbb{R}^d\times\mathbb{R}^d} (\nabla\phi(y) \cdot \alpha_f(x,y)\nabla\phi(x))\, e^{-(V(x)+V(y))}\, \mathrm{d}x\mathrm{d}y \leq 0. \tag{157}$$

The complexity of the foregoing optimization problems is increased substantially when considering more than two particles. From a practical perspective, it is desirable to specify the coupling in terms of the matrix-valued function G appearing in (65) since this formulation is needed for the implementation of the numerical scheme. The linearized optimization objective (Problem 2), however, is formulated in terms of the coupling operator Γ. Passing from the

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

latter to G involves the computationally expensive task of computing the square root of the matrix Q defined in (60). The construction and effective implementation of optimally coupled samplers with multiple particles therefore remains a subject for future work, but could be based on the results for the case of two particles. To give an impression, let us outline an idea based on the notation introduced in Remark 24. It is natural to choose the orthogonal matrices g_{ij} describing the coupling between the ith and the jth particle according to (155), i.e.,

(158)
$$g_{ij}(x_1,\dots,x_n) = \begin{cases} I_{d\times d} - 2\dfrac{\left(\widehat{\nabla\phi(x_i)}+\widehat{\nabla\phi(x_j)}\right)\left(\widehat{\nabla\phi(x_i)}+\widehat{\nabla\phi(x_j)}\right)^T}{\left(\widehat{\nabla\phi(x_i)}+\widehat{\nabla\phi(x_j)}\right)^2} & \text{if } \nabla\phi(x_i)\neq 0, \nabla\phi(x_j)\neq 0, \\ & \widehat{\nabla\phi(x_i)}+\widehat{\nabla\phi(x_j)}\neq 0, \\ I_{d\times d} & \text{otherwise,} \end{cases}$$

or, when the solution ϕ to the Poisson equation (or an approximation thereof) is not available, according to (156). Since the benefit of the coupling in terms of reducing the asymptotic variance is directly related to the value of the expression in (154), it is plausible to choose the weights w_{ij} (see (64)) in such a way that particle x_i is preferentially coupled to particle x_j if $|\nabla\phi(x_i)|$ and $|\nabla\phi(x_j)|$ are similar in magnitude. To make this precise, denote by $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$ the permutation that orders the particles according to $|\nabla\phi|$, i.e.,

(159)
$$|\nabla\phi(x_{\sigma(1)})| \geq |\nabla\phi(x_{\sigma(2)})| \geq \dots \geq |\nabla\phi(x_{\sigma(n-1)})| \geq |\nabla\phi(x_{\sigma(n)})|.$$

Then, denoting the coupling strength by $\beta \in [0, \frac{\pi}{4}]$, we can set the weights as follows:

(160)
$$\begin{aligned} & w_{ii} = \cos\beta, \qquad i = 1,\dots,n, \\ w_{\sigma(1)\sigma(2)} = \sin\beta, \quad & w_{\sigma(3)\sigma(4)} = \sin\beta, \quad \dots \quad w_{\sigma(n-1)\sigma(n)} = \sin\beta, \\ & w_{ij} = 0 \quad \text{otherwise.} \end{aligned}$$

Let us emphasize that the sorting of the particles according to $|\nabla\phi|$ is supposed to be performed at every timestep. We have compared this coupling strategy to simple pairwise coupling without sorting[16] in terms of asymptotic variance,[17] i.e., replacing the second line of (160) by

(161)
$$w_{12} = \sin\beta, \quad w_{34} = \sin\beta, \quad \dots \quad w_{n-1,n} = \sin\beta,$$

for the example of sampling a standard Gaussian measure ($V = \frac{1}{2}|x|^2$) in $d = 10$ dimensions with $n = 10$ particles for the quadratic observable $f_1(x) = \frac{1}{2}|x|^2$ and the mixed observable $f_2(x) = 5|x|^2 + l \cdot x$, where $l = (1,\dots,1)$. As Figure 4 shows, the sorting strategy as detailed in (160) leads to a smaller asymptotic variance in comparison to simple pairwise couplings.

[16] This is equivalent to running $n/2$ two-particle samplers independently in parallel.

[17] We estimated the asymptotic variance from $5 \cdot 10^4$ independent realizations of (65), discretized according to the Euler–Mayurama scheme with timestep $\Delta t = 10^{-3}$.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

5.3. The zigzag process. Recall the setting from section 3.3, and fix an observable of interest $f \in L^2_0(\tilde{\pi})$ [18]. For notational convenience, let us introduce the shorthands

$$\begin{aligned} \alpha_{++}(x,y) &:= \alpha(x,y,+1,+1), & \alpha_{+-}(x,y) &:= \alpha(x,y,+1,-1), \\ \alpha_{-+}(x,y) &:= \alpha(x,y,-1,+1), & \alpha_{--}(x,y) &:= \alpha(x,y,-1,-1). \end{aligned} \tag{162}$$

Taking the constraint (92) into account, we will optimize over the set

$$\mathcal{A} = \{\alpha : \mathbb{R}^2 \times \{-1,1\}^2 \to \mathbb{R} \,|\, \alpha \text{ measurable}, \, 0 \le \alpha \le \min\left(\lambda(x,\theta_x), \lambda(y,\theta_y)\right)\}.$$

The corresponding coupling operators (see (89)) will be denoted by Γ_α. We have the following lemma, the proof of which can be found in Appendix C.

Lemma 63. *The zigzag process satisfies*

$$\begin{aligned} \delta\sigma_F^2(\Gamma_\alpha) &= \frac{1}{4}\int_{\mathbb{R}^2} \tilde{\alpha}(x,y) \cdot \tilde{\phi}'(x)\tilde{\phi}'(y) e^{-(V(x)+V(y))} \,\mathrm{d}x\mathrm{d}y, \\ \tilde{\alpha}(x,y) &= \alpha_{++}(x,y) + \alpha_{--}(x,y) - \alpha_{+-}(x,y) - \alpha_{-+}(x,y), \end{aligned} \tag{163}$$

where $\tilde{\phi} \in L^2(\tilde{\pi})$ is a solution to

$$-(-V'\tilde{\phi}' + \tilde{\phi}'') = f. \tag{164}$$

Remark 64. Observe the remarkable coincidence that (164) coincides with the Poisson equation (139) for the overdamped Langevin dynamics. We employ the notation $\tilde{\phi}$ to distinguish (164) from the Poisson equation (211) in the whole space $\bar{E} = \mathbb{R}^2 \times \{-1,1\}^2$.

The following result is immediate from the expression (163).

Proposition 65. *Let $\alpha^* \in \mathcal{A}$ be given by*

$$\alpha^*(x,y,\theta_x,\theta_y) = \begin{cases} \min\left(\lambda(x,\theta_x), \lambda(y,\theta_y)\right) & \text{if } \tilde{\phi}'(x)\tilde{\phi}'(x)\theta_x\theta_y \le 0, \\ 0 & \text{otherwise.} \end{cases} \tag{165}$$

Then Γ_{α^*} solves Problem 2 in the sense that $\delta\sigma_F^2(\Gamma_{\alpha^*}) \le \delta\sigma_F^2(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

Remark 66. The comment from Remark 57 applies here as well.

Combining Lemma 58 with Proposition 65 immediately yields the following corollaries.

Corollary 67. *In the setting from the first part of Lemma 58,*

$$\alpha^*(x,y,\theta_x,\theta_y) = \begin{cases} \min\left(\lambda(x,\theta_x), \lambda(y,\theta_y)\right) & \text{if } \theta_x\theta_y \le 0, \\ 0 & \text{otherwise} \end{cases} \tag{166}$$

solves Problem 2, in the sense that $\delta\sigma_F^2(\Gamma_{\alpha^*}) \le \delta\sigma_F^2(\Gamma_\alpha)$ for all $\alpha \in \mathcal{A}$.

[18] We use the notation $\tilde{\pi}(\mathrm{d}x) = \frac{1}{Z}e^{-V(x)}\mathrm{d}x$ to distinguish it from the invariant measure π on the full space $\mathbb{R} \times \{-1,1\}$, given in (83).

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Corollary 68. *In the setting from the second part of Lemma 58,*

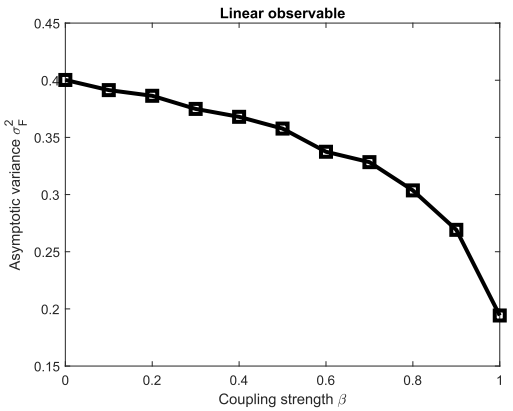
$$\alpha^*(x,y,\theta_x,\theta_y) = \begin{cases} \min\left(\lambda(x,\theta_x),\lambda(y,\theta_y)\right) & \text{if } xy\theta_x\theta_y \leq 0, \\ 0 & \text{otherwise} \end{cases} \tag{167}$$

solves Problem 2, in the sense that $\delta\sigma_F^2(\Gamma_{\alpha^*}) \leq \delta\sigma_F^2(\Gamma_\alpha)$ *for all* $\alpha \in \mathcal{A}$.

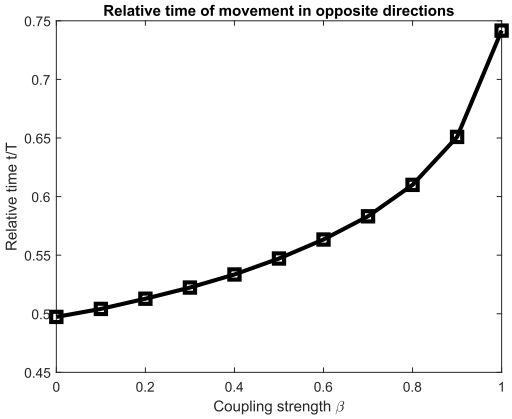
The results from Corollaries 67 and 68 can be interpreted intuitively in the following way. As already pointed out in section 3.3, setting $\alpha(x,y,\theta_x,\theta_y) = \min\left(\lambda(x,\theta_x),\lambda(y,\theta_y)\right)$ encourages simultaneous flips of the velocities θ_x and θ_y (when the particles are at locations x and y, with velocities θ_x and θ_y), whereas the flips occur independently if $\alpha(x,y,\theta_x,\theta_y) = 0$. The coupling associated to (166) therefore leads to an increased probability of simultaneous flips precisely when the two particles move in opposite directions. Observe that simultaneous flips preserve the value of $\theta_x\theta_y$, while single flips change its sign. As a consequence, the relative amount of time during which the two particles move in opposite directions is increased by the coupling associated to (166). Similarly to the case of mirror coupling for overdamped Langevin diffusions (see the discussion following Corollary 60), it is plausible that this dynamics leads to cancellations for monotone observables in the spirit of antithetic variates. The interpretation of Corollary 68 is analogous to the one of Corollary 60. For illustration, we consider again the case of a quadratic potential $V(x) = \frac{1}{2}x^2$ (i.e., a Gaussian target measure) and a linear observable $f(x) = x$. The coupling is chosen according to Corollary 67, i.e., in a suitable manner for the linear observable, modulated by a parameter $\beta \in [0,1]$, analogously to (147) and (148). In Figure 5(a–c) we plot the associated asymptotic variance, the relative time the particles move in opposite directions, as well as the average distance between the particles. Those graphs support the foregoing intuitive arguments. The fact that the average distance between the particles increases with the strength of the coupling is interesting, since it suggests that the state space can be explored more efficiently by using appropriate couplings. In Figure 5(d) we plot a typical trajectory of the joint system. Comparing this graph with the optimal transport map depicted in Figure 2(a), we conclude that the solution to Problem 2 found in Proposition 65 is somewhat close to the solution of the Kantorovich problem, but not nearly as much as the corresponding solution in the case of overdamped Langevin dynamics. Interestingly, the aforementioned similarity is much more pronounced in the case when the target distribution is heavy-tailed. As an example, we plotted a typical trajectory of a mirror-coupled zigzag process targeting a Cauchy distribution in Figure 5(e). We did perform numerical experiments for quadratic observables. For them, an improvement in the asymptotic variance is hardly noticeable. Furthermore, a typical trajectory for the coupling induced by (167) very much resembles the typical trajectories for the independent coupling. As it seems, couplings of zigzag processes are not very efficient in the setting of Lemma 58.2. A possible explanation is that PDMPs are more rigid than diffusions (in fact, by definition, they move deterministically during a considerable time span), allowing less flexibility in terms of couplings.

6. A remark on the rate of convergence to equilibrium. In this section, we study the rate of convergence to equilibrium for coupled processes. For convenience, let us assume that the spaces E_i and the operators $\mathcal{L}_i$ are identical copies of each other. When addressing the

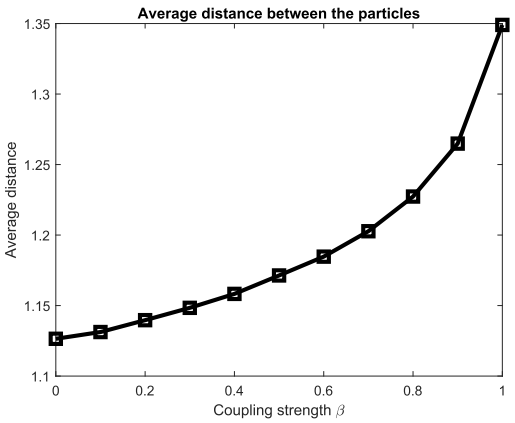
Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.



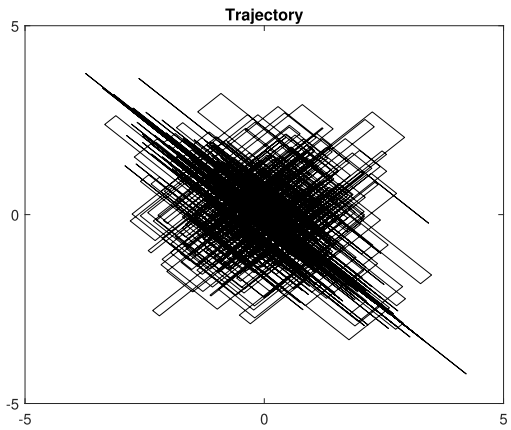
(a) Asymptotic variance associated to the linear observable $f(x) = x$, depending on the coupling strength β.



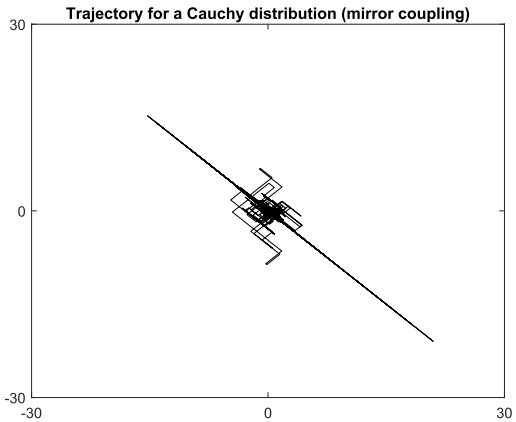
(b) Relative amount of time that the particles move in opposite directions, depending on the coupling strength β.



(c) Average distance between the two particles, depending on the coupling strength β.



(d) Part of the trajectory for a Gaussian target with mirror coupling.



(e) Part of the trajectory for a Cauchy target with mirror coupling.

Figure 5. *Coupling for the zigzag process according to Corollary 67.*

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

marginal process(es), we will usually suppress the indices and write $\mathcal{L}$ and E. Furthermore, let us fix an ergodic coupling operator $\Gamma \in \mathcal{G}^0$ and denote as usual the corresponding generator and semigroup by $\bar{\mathcal{L}}_\Gamma = \bar{\mathcal{L}}_0 + \Gamma$ and $(\bar{S}_t^\Gamma)_{t\geq 0}$, respectively. In what follows, we will make use of the following subspace of centered observables in $L_0^2(\bar{\pi}_\Gamma)$:

$$\tilde{L}_0^2(\bar{\pi}_\Gamma) := \left\{ F \in L_0^2(\bar{\pi}_\Gamma) \,|\, \text{there exists } f \in L_0^2(\pi) \text{ such that } F = \frac{1}{n}\sum_{i=1}^n f_i \right\} \subseteq L_0^2(\bar{\pi}_\Gamma). \tag{168}$$

Clearly, the space $\tilde{L}_0^2(\bar{\pi}_\Gamma)$ comprises the observables of interest in our framework. By using the extension operator Π^* from (9), $\tilde{L}_0^2(\bar{\pi}_\Gamma)$ can equivalently be defined via $\tilde{L}_0^2(\bar{\pi}_\Gamma) = \Pi^* L_0^2(\bar{\pi}_\Gamma)$. The main result of this section is the following characterization of exponential convergence to equilibrium.

Theorem 69. *For $\Lambda > 0$, the following are equivalent:*

1. *Poincaré inequality:*

$$\langle F, F\rangle_{L^2(\bar{\pi}_\Gamma)} \leq \frac{1}{\Lambda}\langle F, (-\bar{\mathcal{L}}_\Gamma)F\rangle_{L^2(\bar{\pi}_\Gamma)} \tag{169}$$

   for all $F \in \mathcal{D}(\bar{\mathcal{L}}_\Gamma) \cap \tilde{L}_0^2(\bar{\pi}_\Gamma)$.

2. *Exponential decay:*

$$\|\bar{S}_t^\Gamma F\|^2_{L^2(\bar{\pi}_\Gamma)} \leq e^{-2\Lambda t}\|F\|^2_{L^2(\bar{\pi}_\Gamma)}, \quad t \geq 0, \tag{170}$$

   for all $F \in \tilde{L}_0^2(\bar{\pi}_\Gamma)$.

Remark 70. Theorem 69 is well known if $\tilde{L}_0^2(\bar{\pi}_\Gamma)$ is replaced by the whole space $L_0^2(\bar{\pi}_\Gamma)$, see [6, Theorem 4.2.5]. For our purposes, however, it is natural to restrict attention to the smaller space $\tilde{L}_0^2(\bar{\pi}_\Gamma)$. In particular, by the duality $(\Pi\bar{\pi})(f) = \bar{\pi}(\Pi^* f)$ explained in the introduction, the decay estimate (170) implies exponential convergence of the laws $(\Pi\bar{\pi}_t)_{t\geq 0}$ with the same rate.

The proof of Theorem 69 relies on the following lemma.

Lemma 71. *Let $F \in \tilde{L}_0^2(\bar{\pi}_\Gamma)$ with $F = \frac{1}{n}\sum_{i=1}^n f_i$, $f \in L_0^2(\pi)$. Then*

$$\bar{S}_t^\Gamma F = \frac{1}{n}\sum_{i=1}^n (S_t f)_i, \quad t \geq 0. \tag{171}$$

In particular, $\tilde{L}_0^2(\bar{\pi}_\Gamma)$ is invariant under the flow of $(\bar{S}_t^\Gamma)_{t\geq 0}$.

Proof. For $F \in C_b(\bar{E}) \cap \tilde{L}_0^2(\bar{\pi}_\Gamma)$ we have

$$\begin{aligned}(\bar{S}_t^\Gamma F)(x_1,\dots,x_n) &= \mathbb{E}[F(\bar{X}_t)|\bar{X}_0 = (x_1,\dots,x_n)] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[f(X_t^i)|\bar{X}_0 = (x_1,\dots,x_n)] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[f(X_t^i)|X_0^i = x_i] = \frac{1}{n}\sum_{i=1}^n (S_t f)(x_i).\end{aligned} \tag{172}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Between the second and the third line, we used the fact that the process $(\bar{X}_t)_{t\geq 0}$ has $(X_t^i)_{t\geq 0}$ as its ith marginal, so in particular, the law of $f(X_t^i)$ depends on the initial condition $\bar{X}_0$ only through $X_0^i = x_i$. For arbitrary $F \in \tilde{L}_0^2(\bar{\pi}_\Gamma)$ the result follows by a standard density argument. ∎

Proof of Theorem 69. The proof is verbatim the same as for the usual result. However, the fact that $\tilde{L}_0^2(\bar{\pi}_\Gamma)$ is invariant under the flow $(\bar{S}_t^\Gamma)_{t\geq 0}$ is crucial. For completeness let us sketch the proof: Let $F \in \tilde{L}_0^2(\bar{\pi}_\Gamma)$, and assume that the Poincaré inequality (169) holds for some constant $\Lambda > 0$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\|\bar{S}_t^\Gamma F\|^2_{L^2(\bar{\pi}_\Gamma)}\right) = \langle \bar{S}_t^\Gamma F, \bar{\mathcal{L}}_\Gamma \bar{S}_t^\Gamma F\rangle_{L^2(\bar{\pi}_\Gamma)} \leq -\Lambda \langle \bar{S}_t^\Gamma F, \bar{S}_t^\Gamma F\rangle_{L^2(\bar{\pi}_\Gamma)}, \tag{173}$$

where the last inequality uses the fact that $\bar{S}_t^\Gamma f \in \tilde{L}_0^2(\bar{\pi}_\Gamma)$ according to Lemma 71. Exponential decay as in (170) follows by Gronwall's lemma. The converse direction follows by performing a Taylor expansion of the decay estimate (170) around $t = 0$. ∎

To explain the significance of Theorem 69, let us start by writing (169) in the form

$$\langle f, f\rangle_{L^2(\pi)} + \frac{1}{n}\sum_{\substack{i,j=1\\ i\neq j}}^n \langle f_i, f_j\rangle_{L^2(\bar{\pi}_\Gamma)} \leq \frac{1}{\Lambda}\left(\langle f, (-\mathcal{L})f\rangle_{L^2(\pi)} + \frac{1}{n}\sum_{\substack{i,j=1\\ i\neq j}}^n \langle f_i, (-\mathcal{L}_j) f_j\rangle_{L^2(\bar{\pi}_\Gamma)}\right), \tag{174}$$

using the marginal property of $\bar{\pi}_\Gamma$. Clearly, (174) deviates from the usual one-particle Poincaré inequality by the additional terms involving summation over pairs of particles. To make this more precise and analyze the impact of these terms, let us define the following bilinear form on $L_0^2(\pi)$:

$$\langle\!\langle f, g\rangle\!\rangle := \langle f, g\rangle_{L^2(\pi)} + \frac{1}{n}\sum_{\substack{i,j=1\\ i\neq j}}^n \langle f_i, g_j\rangle_{L^2(\bar{\pi}_\Gamma)}, \quad f, g \in L_0^2(\pi). \tag{175}$$

For $F = \sum_{i=1}^n f_i$ and $G = \sum_{i=1}^n g_i$ we have that $\langle F, G\rangle_{L^2(\bar{\pi}_\Gamma)} = n\langle\!\langle f, g\rangle\!\rangle$. Hence, $\langle\!\langle \cdot, \cdot\rangle\!\rangle$ is both symmetric and nonnegative definite, but $\langle\!\langle f, f\rangle\!\rangle = 0$ is possible for $f \neq 0$. It is therefore natural to define the equivalence relation

$$f \sim g :\Longleftrightarrow \langle\!\langle f - g, f - g\rangle\!\rangle = 0, \quad f, g \in L_0^2(\pi) \tag{176}$$

and the corresponding Hilbert space

$$\mathcal{H} := L_0^2(\pi)/\sim . \tag{177}$$

For $f \in L_0^2(\pi)$, the corresponding equivalence class will be denoted by $[f] \in \mathcal{H}$. Using again the correspondence $F = \sum_{i=1}^n f_i$, we see that $\langle\!\langle f, f\rangle\!\rangle = 0$ if and only if $F = 0$ $\bar{\pi}_\Gamma$-almost surely. By ergodicity, this is also equivalent to $\bar{\mathcal{L}}_\Gamma F = 0$, $\bar{\pi}_\Gamma$-almost surely. We hence see that $\mathcal{L}$ respects $\sim$-equivalence classes, i.e., $f \sim g$ if and only if $\mathcal{L}f \sim \mathcal{L}g$. Denoting the induced operator on $\mathcal{H}$ by $\mathcal{L}_\mathcal{H}$ (i.e., $\mathcal{L}_\mathcal{H}[f] = [\mathcal{L}f]$), it is then immediate that each of (169) and (174) is equivalent to

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$$\langle\!\langle f, f \rangle\!\rangle \leq \frac{1}{\Lambda} \langle\!\langle f, (-\mathcal{L}_{\mathcal{H}}) f \rangle\!\rangle, \quad f \in \mathcal{H} \cap \mathcal{D}(\mathcal{L}). \tag{178}$$

By its similarity to the one-particle Poincaré inequality, the formulation (178) is convenient for the comparison between the spectral gaps of the underlying and the coupled dynamics.

Let us assume from now on that $\mathcal{L}$ is self-adjoint in $L^2_0(\pi)$ with discrete spectrum, with

$$-\mathcal{L} e_i = \mu_i e_i, \quad 0 < \mu_1 \leq \mu_2 \leq, \dots \tag{179}$$

where the eigenvectors $(e_i)_{i \in \mathbb{N}}$ form an orthonormal basis in $L^2_0(\pi)$. The optimal constant in the one-particle Poincaré inequality is then clearly given by $\lambda = \mu_1$. In the study of the coupled Poincaré inequality (178), two interesting effects might occur. Firstly, the spectrum of $\mathcal{L}_{\mathcal{H}}$ might be different from the spectrum of $\mathcal{L}$. Secondly, $\mathcal{L}_{\mathcal{H}}$ might not be symmetric with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Let us start with the first point. Clearly, $\sigma(\mathcal{L}_{\mathcal{H}}) \subseteq \sigma(\mathcal{L})$; more precisely

$$\sigma(\mathcal{L}_{\mathcal{H}}) = \{\mu_i : \langle\!\langle e_i, e_i \rangle\!\rangle \neq 0\}. \tag{180}$$

Example 72. Consider the dynamics

$$\mathrm{d}X_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \tag{181a}$$

$$\mathrm{d}Y_t = -\nabla V(Y_t)\,\mathrm{d}t - \sqrt{2}\,\mathrm{d}B_t \tag{181b}$$

with a standard $\mathbb{R}^d$-valued Brownian motion $(B_t)_{t \geq 0}$. Let us assume that the potential V grows sufficiently fast at infinity such that the one-particle generator $\mathcal{L}$ has compact resolvent and hence discrete spectrum in $L^2_0(\pi)$. Furthermore, suppose that the eigenvalues and eigenfunctions are labelled and ordered as in (179). Let us now assume that V is even, i.e., $V(x) = V(-x)$, and that the process is ergodic. The invariant measure is then given by

$$\bar{\pi}_\Gamma(\mathrm{d}x\mathrm{d}y) = \frac{1}{Z} e^{-V(x)} \delta_{x+y}\,\mathrm{d}x\mathrm{d}y, \tag{182}$$

and the corresponding new (degenerate) scalar product in $L^2_0(\pi)$ turns out to be

$$\langle\!\langle f, g \rangle\!\rangle = \frac{1}{Z} \int_{\mathbb{R}^d} f(x) g(x) e^{-V(x)}\,\mathrm{d}x \tag{183a}$$

$$+ \frac{1}{2Z} \left(\int_{\mathbb{R}^d} f(x) g(-x) e^{-V(x)} \mathrm{d}x + \int_{\mathbb{R}^d} f(-x) g(x) e^{-V(x)} \mathrm{d}x \right). \tag{183b}$$

Notice that by the symmetry of V, all the eigenfunctions of $\mathcal{L}$ are either even or odd. Moreover, a short calculation shows that $\langle\!\langle f, f \rangle\!\rangle = 0$ if and only if f is odd (meaning that $-f(x) = f(-x)$). Using (180), we see that

$$\sigma(\mathcal{L}_{\mathcal{H}}) = \{\lambda_i : e_i \text{ is odd}\}. \tag{184}$$

Another short calculation shows that $\mathcal{L}_{\mathcal{H}}$ is symmetric with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$, i.e.,

$$\langle\!\langle f, \mathcal{L}_{\mathcal{H}} g \rangle\!\rangle = \langle\!\langle \mathcal{L}_{\mathcal{H}} f, g \rangle\!\rangle, \quad f \in \mathcal{H} \cap \mathcal{D}(\mathcal{L}). \tag{185}$$

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

If the first eigenfunction e_1 is odd,[19] it therefore follows that the coupled Poincaré inequality (174) holds with the constant $\Lambda = \mu_2$, showing an improved rate of convergence for the coupled dynamics.

Example 73. Let us examine the second point, i.e., the possibility of $\mathcal{L}_{\mathcal{H}}$ not being symmetric with respect to $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. For simplicity, assume that $\sigma(\mathcal{L}_{\mathcal{H}}) = \sigma(\mathcal{L})$, i.e., $\langle\!\langle e_i, e_i \rangle\!\rangle \neq 0$ for all $i \in \mathbb{N}$. Consider the case when the measures $\bar{\pi}_0$ and $\bar{\pi}_\Gamma$ have densities with respect to a common dominating measure m (for convenience denoted by the same symbols), and suppose there exist constants $c_1, c_2 > 0$ such that

$$c_1 \bar{\pi}_0(x) \leq \bar{\pi}_\Gamma(x) \leq c_2 \bar{\pi}_0(x), \quad x \in E. \tag{186}$$

This is the case precisely when the norms in $L^2(\bar{\pi}_0)$ and $L^2(\bar{\pi}_\Gamma)$ are equivalent. For $F = \sum_{i=1}^n f_i$, we have that $n\langle\!\langle f, f \rangle\!\rangle = \langle F, F \rangle_{L^2(\bar{\pi}_\Gamma)}$ as well as $n\langle f, f \rangle_{L^2(\pi)} = \langle F, F \rangle_{L^2(\bar{\pi}_0)}$. Using (186), we hence conclude that

$$c_1 \langle f, f \rangle_{L^2(\pi)} \leq \langle\!\langle f, f \rangle\!\rangle \leq c_2 \langle f, f \rangle_{L^2(\pi)}, \quad f \in L^2(\pi). \tag{187}$$

By assumption, the marginal process satisfies a Poincaré inequality as well as the equivalent decay estimate

$$\|S_t f\|^2_{L^2(\pi)} \leq e^{-2\lambda t} \|f\|^2_{L^2(\pi)}, \quad f \in L^2_0(\pi), \tag{188}$$

with $\lambda = \mu_1$. By the equivalence (187), we conclude that

$$\|\bar{S}^\Gamma_t F\|^2_{L^2(\bar{\pi}_\Gamma)} \leq C e^{-2\lambda t} \|F\|^2_{L^2(\bar{\pi}_\Gamma)}, \quad t \geq 0, \quad F \in \tilde{L}^2_0(\bar{\pi}_\Gamma), \tag{189}$$

with $C = \frac{c_2}{c_1}$. Comparing (188) and (189), we see that the coupled process achieves the same exponential rate of convergence as the one-particle processes, but possibly with a worse constant C in front of the exponential. The latter can be characterized in terms of the equivalence estimate (186).

Conclusively, the speed of convergence to equilibrium can be both faster (as in Example 72) and slower (as in Example 73) for coupled processes, in comparison with the underlying one-particle processes. We leave a more thorough investigation of the Poincaré inequality (174) for future work.

7. Outlook and future work. In this paper we have introduced a general framework for the construction and analysis of coupled MCMC samplers. Formulating the results in an abstract setting has allowed us to address both (possibly degenerate) diffusion processes as well as PDMPs, emphasizing common structural properties. The analysis of appropriate central limit theorems has exposed notable connections to the theory of optimal transportation. We showed that the ensuing optimization problem has singularity properties akin to those

[19] In one dimension, it can be proved that the first eigenfunction is always odd by appealing to the node theorem for Schrödinger operators in Sturm–Liouville theory [89, Chapter 9]. We conjecture that this fact might also be true in higher dimensions but are not able to give a proof or a reference. We are grateful to Dr. S. Bögli and Professor A. Laptev for useful discussions on this issue.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

appearing in the usual Kantorovich formulation. We then studied a surrogate problem, leading to novel coupling strategies that seem promising for applications. Finally, we derived a functional inequality of Poincaré type suitable for the study of the exponential convergence to equilibrium for coupled processes.

Our work can be extended in several directions. On the theoretical side, proving or disproving the Conjectures 11, 17, and 19 would further illuminate the structural properties of the developed theory. Moreover, establishing a more rigorous connection between the optimal transport problems 1 and 2 with the usual Kantorovich formulation might lead to further developments bridging the theories of Markov processes and optimal transportation.

In terms of applications in sampling, a more detailed study of the couplings between many particles is needed, a starting point being the results in section 5.2. Furthermore, it would be desirable to relax our assumption that the laws of the marginal processes remain unchanged, as this would allow for more pronounced interactions between the particles. In this regard, the inclusion of the methodology put forward in [59] in our framework would be of particular interest for practitioners.

In the broader context of statistical computation, it seems that coupling approaches along the lines developed here could be fruitfully applied in the context of the calculation of transport coefficients and sensitivities [5, 43]. More speculatively, it would be interesting to investigate the use of our ideas in the context of multilevel Monte Carlo [38] or computational optimal transport [77]. We leave these directions for future investigations.

Appendix A. Random orthogonal transformations of Brownian motions. The following lemma has been extracted from [31, page 56]; see also [74, Theorem 8.4.2]. This result states that the set of Brownian motions is preserved under possibly time-dependent linear transformations possessing certain orthogonality properties. Importantly, no regularity constraints with regard to the time dependence are required beyond measurability. This fact is crucial in the proofs of Lemmas 22 and 27.

Lemma 74 (random orthogonal transformations). *Suppose that the $\mathbb{R}^N$-valued stochastic process $(X_t)_{t\geq 0}$ is a solution to the SDE*

$$\mathrm{d}X_t = O_t \mathrm{d}W_t, \quad X_0 = x_0, \tag{190}$$

where $(W_t)_{t\geq 0}$ is an M-dimensional standard Brownian motion generating the filtration $(\mathcal{F}_t)_{t\geq 0}$, and $(O_t)_{t\geq 0}$ is a product-measurable $(\mathcal{F}_t)_{t\geq 0}$-adapted process taking values in $\mathbb{R}^{N\times M}$. Assume furthermore that

$$O_t O_t^T = I_{N\times N} \tag{191}$$

for all $t \geq 0$, almost surely. Then $(X_t)_{t\geq 0}$ is an N-dimensional standard Brownian motion.

Appendix B. The derivative formula for invariant measures. Here, we provide the proof of the derivative formula (121) that allows us to compute the change of the invariant measure under an infinitesimal change of the coupling.

Proof of Proposition 46. The idea of the proof stems from [58] in the context of invariant measures for discretized SDEs and was also advertized in [61, Remark 5.5].

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For convenience, let us first introduce the notation $g = -(\bar{\mathcal{L}}_\Gamma^*)^{-1}\mathrm{d}\Gamma^*\mathbf{1}$. Furthermore, we will make use of the projection operators

$$\Pi\phi = \int_{\bar{E}} \phi \,\mathrm{d}\bar{\pi}_\Gamma^0, \quad \Pi^\perp\phi = \phi - \Pi\phi \tag{192}$$

acting on $L^2(\bar{\pi}_\Gamma)$. Using $c = (\Pi + \Pi^\perp)c$ in (121), we see that (121) is equivalent to

$$\left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \int_{\bar{E}} \Pi^\perp c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon = -\int_{\bar{E}} \left(\Pi^\perp c\right) \left[\bar{\mathcal{L}}_\Gamma^*\right]^{-1} (\mathrm{d}\Gamma^* 1)\,\mathrm{d}\bar{\pi}_\Gamma^0. \tag{193}$$

We may thus without loss of generality assume that $\Pi^\perp c = c$ (i.e., $\Pi c = 0$), and will do so in the following. Furthermore, let us also assume that $c \in C_b^\infty(\bar{E})$ such that the calculations in what follows are justified.[20] The general case then follows by a standard approximation argument, using a sequence $(c_j)_j \subset C_b^\infty(\bar{E})$ that converges uniformly towards $c \in C_b(\bar{E})$. A short calculation (using the fact that $\bar{\pi}_\Gamma^0(\bar{\mathcal{L}}_\Gamma c) = 0$) shows that

$$\int_{\bar{E}} \bar{\mathcal{L}}_\Gamma^\varepsilon c \cdot (1 + \varepsilon g)\,\mathrm{d}\bar{\pi}_\Gamma^0 = \varepsilon^2 \int_{\bar{E}} (\mathrm{d}\Gamma c) g \,\mathrm{d}\bar{\pi}_\Gamma^0 = \mathcal{O}(\varepsilon^2). \tag{194}$$

Inserting $\Pi + \Pi^\perp = I$, we see that the above is equivalent to

$$\int_{\bar{E}} \Pi^\perp \bar{\mathcal{L}}_\Gamma^\varepsilon \Pi^\perp c \cdot (1 + \varepsilon g)\,\mathrm{d}\bar{\pi}_\Gamma^0 + \varepsilon \int_{\bar{E}} \Pi \mathrm{d}\Gamma c \cdot (1 + \varepsilon g)\,\mathrm{d}\bar{\pi}_\Gamma^0 = \mathcal{O}(\varepsilon^2). \tag{195}$$

At the same time, we have that

$$\int_{\bar{E}} \bar{\mathcal{L}}_\Gamma^\varepsilon c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon = 0. \tag{196}$$

Using again $\Pi + \Pi^\perp = I$ and $\Pi\bar{\mathcal{L}}_\Gamma = 0$, (196) can be expressed as

$$\int_{\bar{E}} \Pi^\perp \bar{\mathcal{L}}_\Gamma^\varepsilon \Pi^\perp c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon = -\varepsilon \int_{\bar{E}} \Pi \mathrm{d}\Gamma c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon. \tag{197}$$

We can now combine (195) and (197) to arrive at

$$\int_{\bar{E}} \Pi^\perp \bar{\mathcal{L}}_\Gamma^\varepsilon \Pi^\perp c \cdot (1 + \varepsilon g)\,\mathrm{d}\bar{\pi}_\Gamma^0 - \int_{\bar{E}} \Pi^\perp \bar{\mathcal{L}}_\Gamma^\varepsilon \Pi^\perp c \,\mathrm{d}\bar{\pi}_\Gamma^\varepsilon = \mathcal{O}(\varepsilon^2). \tag{198}$$

Let us introduce the pseudo-inverse

$$Q_\varepsilon = \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp - \varepsilon \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp \mathrm{d}\Gamma \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp \tag{199}$$

acting on $L_0^2(\bar{\pi}_\Gamma^0)$. We have that

$$\Pi^\perp \bar{\mathcal{L}}_\Gamma^\varepsilon \Pi^\perp Q_\varepsilon = \Pi^\perp - \varepsilon^2 \Pi^\perp \mathrm{d}\Gamma \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp \mathrm{d}\Gamma \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp, \tag{200}$$

[20] We mention that in particular, the Feller property guarantees that $\bar{\mathcal{L}}_\Gamma^\varepsilon c \in C_b(\bar{E})$ (see Remark 2) and so the integrals under consideration are finite.

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

i.e., in $L^2_0(\bar{\pi}^0_\Gamma)$, Q_ε inverts $\bar{\mathcal{L}}^\varepsilon_\Gamma$ up to an error of order ε^2. Upon replacing c by $Q_\varepsilon c$ in (198), it follows that

$$\int_{\bar{E}} c(1+\varepsilon g)\,\mathrm{d}\bar{\pi}^0_\Gamma - \int_{\bar{E}} c\,\mathrm{d}\bar{\pi}^\varepsilon_\Gamma = \mathcal{O}(\varepsilon^2), \tag{201}$$

recalling that $\Pi^\perp c = c$ by assumption. In the last step, we have used the fact that there exists a constant $C > 0$ such that

$$\left| \int_{\bar{E}} \Pi^\perp \mathrm{d}\Gamma \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp \mathrm{d}\Gamma \Pi^\perp \bar{\mathcal{L}}_\Gamma^{-1} \Pi^\perp c\,\mathrm{d}\bar{\pi}^\varepsilon_\Gamma \right| \leq C \tag{202}$$

uniformly in ε. Indeed, the integrand is bounded by the third condition of Definition 42 and the fact that the coefficients of $\mathrm{d}\Gamma$ have compact support. The bound (202) is required to ensure that the corresponding integral expression hidden on the right-hand side of (201) is indeed of order ε^2. Finally, dividing by ε in (201) and letting $\varepsilon \to 0$ yields the desired result. ∎

Appendix C. Properties of the solutions to one-dimensional Poisson equations. The proofs in this section essentially leverage the fact that the Poisson equations under consideration can be solved up to quadratures in one dimension.

Proof of Lemma 58. Variation of constants shows that ϕ' is given by

$$\phi'(x) = \left(- \int_{-\infty}^{x} f(s) e^{-V(s)} \mathrm{d}s + C \right) e^{V(x)} \tag{203}$$

for some constant $C \in \mathbb{R}$. The requirement that $\pi(\phi) = 0$ necessitates $C = 0$. Indeed, from $\pi(f) = 0$ it follows that the integral term in (203) goes to zero as $x \to \pm\infty$, and therefore

$$\lim_{x \to \pm\infty} \frac{\phi'(x)}{e^{V(x)}} = C. \tag{204}$$

By l'Hôpital's rule, we have that

$$\lim_{x \to \pm\infty} \frac{\phi(x) e^{-V(x)}}{\int_0^x e^{V(s)}\,\mathrm{d}s \cdot e^{-V(x)}} = C. \tag{205}$$

The requirement that ϕ is integrable with respect to $\pi(\mathrm{d}x) \propto e^{-V(x)}\,\mathrm{d}x$ implies that

$$\lim_{x \to \pm\infty} \phi(x) e^{-V(x)} = 0.$$

Furthermore (again by l'Hôpital's rule),

$$\lim_{x \to \pm\infty} \int_0^x e^{V(s)}\,\mathrm{d}s \cdot e^{-V(x)} = - \lim_{x \to \pm\infty} V'(x), \tag{206}$$

which cannot be zero since $\int_{-\infty}^{\infty} e^{-V(x)}\,\mathrm{d}x < \infty$. Hence, $C = 0$.

To prove the first part, notice that from $\pi(f) = 0$ and continuity, it follows that there exists $x^* \in \mathbb{R}$ such that $f(x^*) = 0$. Let us assume that f is monotonically increasing (for

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

monotonically decreasing f the reasoning is analogous). We then have that $f \leq 0$ on $(-\infty, x^*]$ and $f \geq 0$ on $[x^*, \infty)$. Consider now the function

$$\Phi(x) = -\int_{-\infty}^{x} f(s)e^{-V(s)}\,\mathrm{d}s, \quad x \in \mathbb{R}. \tag{207}$$

Clearly Φ is increasing on $(-\infty, x^*]$ and decreasing on $[x^*, \infty)$. From $\pi(f) = 0$ it follows that $\lim_{x\to\pm\infty} \Phi(x) = 0$ and hence $\Phi(x) \geq 0$ for all $x \in \mathbb{R}$. This proves the claim since $\phi'(x) = \Phi(x)e^{V(x)}$.

To prove the second part, first observe that $\pi(f) = 0$ implies

$$\int_{-\infty}^{0} f(s)e^{-V(s)}\mathrm{d}s + \int_{0}^{\infty} f(s)e^{-V(s)}\mathrm{d}s = 0. \tag{208}$$

Furthermore, the symmetry properties of f and V show that

$$\int_{-\infty}^{0} f(s)e^{-V(s)}\mathrm{d}s - \int_{0}^{\infty} f(s)e^{-V(s)}\mathrm{d}s = 0, \tag{209}$$

using the substitution $s \mapsto -s$. Equations (208) and (209) together imply that $\Phi(0) = 0$ for Φ as defined in (207). The claim now follows using an analogous argument to the one used in the proof of 1.). ∎

Proof of Lemma 63. Recall from (136) that

$$\delta\sigma_F^2(\Gamma_\alpha) = \int_{\bar{E}} \xi\,\mathrm{d}\bar{\pi}_0, \tag{210}$$

where $\xi(x, y) = \phi(x)\phi(y)$ and $\phi : \mathbb{R} \times \{-1, +1\} \to \mathbb{R}$ is the solution to the (one-particle) Poisson equation

$$\theta\partial_x\phi(x, \theta) + \lambda(x, \theta)\,(\phi(x, -\theta) - \phi(x, \theta)) = f(x), \quad \pi(\phi) = 0. \tag{211}$$

Note that for convenience, we have assumed without loss of generality that $\tilde{\pi}(f) = 0$. Let us now calculate

$$\begin{aligned}
\int_{\bar{E}} \Gamma\xi\,\mathrm{d}\pi_0 &= \frac{1}{4}\sum_{\theta_x=\pm1,\,\theta_y=\pm1}\int_{\mathbb{R}^2}\Bigg(\alpha(x, y, \theta_x, \theta_y)\cdot\Big[\phi(x, \theta_x)\phi(y, \theta_y) - \phi(x, \theta_x)\phi(y, -\theta_y) - \\
&\quad - \phi(x, -\theta_x)\phi(y, \theta_y) + \phi(x, -\theta_x)\phi(y, -\theta_y)\Big]\Bigg)e^{-(V(x)+V(y))}\,\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4}\int_{\mathbb{R}^2}\Bigg(\Big[\alpha_{++}(x, y) + \alpha_{--}(x, y) - \alpha_{+-}(x, y) - \alpha_{-+}(x, y)\Big]\cdot \\
&\quad \cdot\Big[\phi_+(x)\phi_+(y) + \phi_-(x)\phi_-(y) - \phi_+(x)\phi_-(y) - \phi_-(x)\phi_+(y)\Big]\Bigg)e^{-(V(x)+V(y))}\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4}\int_{\mathbb{R}^2}\Bigg(\Big[\alpha_{++}(x, y) + \alpha_{--}(x, y) - \alpha_{+-}(x, y) - \alpha_{-+}(x, y)\Big]\cdot \\
&\quad \cdot\Big[\phi_+(x) - \phi_-(x)\Big]\cdot\Big[\phi_+(y) - \phi_-(y)\Big]\Bigg)e^{-(V(x)+V(y))}\,\mathrm{d}x\mathrm{d}y,
\end{aligned} \tag{212}$$

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

where again we employed the notation $\phi_\pm(x) = \phi(x, \pm 1)$. Observe now that equation (211) can be recast as

$$\partial_x \phi_+(x) + \lambda_+(x)\,(\phi_-(x) - \phi_+(x)) = f(x), \tag{213a}$$

$$-\partial_x \phi_-(x) + \lambda_-(x)\,(\phi_+(x) - \phi_-(x)) = f(x), \tag{213b}$$

where both ϕ_+ and ϕ_- have to be integrable with respect to the measure $\frac{1}{Z} e^{-V(x)} \mathrm{d}x$ and satisfy

$$\int_{\mathbb{R}} (\phi_+(x) + \phi_+(x))\, e^{-V(x)} \mathrm{d}x = 0. \tag{214}$$

Adding (213a) and (213b) leads to

$$\left(\partial_x - V'\right)(\phi_+ - \phi_-) = 2f, \tag{215}$$

using $\lambda_+(x) - \lambda_-(x) = V'(x)$. Finally setting $\partial_x \tilde{\phi} = \frac{1}{2}(\phi_+ - \phi_-)$ and comparing with (212) leads to the desired result. Note that as in the proof of Lemma 58, (215) determines $\phi_+ - \phi_-$ uniquely under the condition that $\phi_+ - \phi_-$ is integrable with respect to $e^{-V(x)}\mathrm{d}x$. ∎

Acknowledgments. The authors would like to thank Pedro Aceves Sanchez, Sabine Bögli, Andreas Eberle, Julien Roussel, Ari Laptev, Gabriel Stoltz, and Urbain Vaes for stimulating discussions. We wish to furthermore thank the referees for their very careful reading of our manuscript and numerous suggestions as well as pointers to the literature that have helped to improve the presentation and opened up new perspectives.

REFERENCES

[1] I. Amemiya and K. Shiga, *On tensor products of Banach spaces*, Kōdai Math. Sem. Rep., 9 (1957), pp. 161–178, http://projecteuclid.org/euclid.kmj/1138843934.

[2] C. Andrieu, A. Doucet, and R. Holenstein, *Particle Markov chain Monte Carlo methods*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 72 (2010), pp. 269–342.

[3] L. Angiuli, G. Metafune, and C. Spina, *Feller semigroups and invariant measures*, Riv. Math. Univ. Parma (N.S.), 1 (2010), pp. 347–406.

[4] W. Arendt, A. Grabosch, G. Greiner, U. Groh, H. P. Lotz, U. Moustakas, R. Nagel, F. Neubrander, and U. Schlotterbeck, *One-Parameter Semigroups of Positive Operators*, Lecture Notes in Math. 1184, Springer-Verlag, Berlin, 1986, https://doi.org/10.1007/BFb0074922.

[5] R. Assaraf, B. Jourdain, T. Lelièvre, and R. Roux, *Computation of sensitivities for the invariant measure of a parameter dependent diffusion*, Stoch. Partial Differ. Equ. Anal. Comput., 6 (2018), pp. 125–183.

[6] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, Springer Science & Business Media, New york, 2013.

[7] P. H. Baxendale, *Statistical equilibrium and two-point motion for a stochastic flow of diffeomorphisms*, in Spatial Stochastic Processes, Progr. Probab. 19, Birkhäuser, Boston, MA, 1991, pp. 189–218.

[8] R. N. Bhattacharya, *On the functional central limit theorem and the law of the iterated logarithm for Markov processes*, Z. Wahrsch. Verw. Gebiete, 60 (1982), pp. 185–201, https://doi.org/10.1007/BF00531822.

[9] J. Bierkens and A. Duncan, *Limit theorems for the zig-zag process*, Adv. in Appl. Probab., 49 (2017), pp. 791–825, https://doi.org/10.1017/apr.2017.22.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

[10] J. Bierkens, P. Fearnhead, and G. Roberts, *The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data*, arXiv:1607.03188, 2016.
[11] J. Bierkens, G. Roberts, and P.-A. Zitt, *Ergodicity of the Zigzag Process*, arXiv:1712.09875, 2018.
[12] N. Bonneel, M. Van De Panne, S. Paris, and W. Heidrich, *Displacement interpolation using Lagrangian mass transport*, ACM Trans. Graphics, 30 (2011), 158.
[13] B. Böttcher, R. Schilling, and J. Wang, *Lévy Matters.* III, Lecture Notes in Math., 2099 Springer, Cham, 2013, https://doi.org/10.1007/978-3-319-02684-8.
[14] N. Bou-Rabee, A. Eberle, and R. Zimmer, *Coupling and Convergence for Hamiltonian Monte Carlo*, arXiv:1805.00452, 2018.
[15] N. Bou-Rabee and J. M. Sanz-Serna, *Randomized Hamiltonian Monte Carlo*, Ann. Appl. Probab., 27 (2017), pp. 2159–2194, https://doi.org/10.1214/16-AAP1255.
[16] A. Bouchard-Côté, S. J. Vollmer, and A. Doucet, *The bouncy particle sampler: A non-reversible rejection-free Markov chain Monte Carlo method*, J. Amer. Statist. Assoc. 113 (2018), pp. 855–867, https://doi.org/10.1080/01621459.2017.1294075.
[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, UK, Cambridge, 2004, https://doi.org/10.1017/CBO9780511804441.
[18] M.-F. Chen, *Eigenvalues, Inequalities, and Ergodic Theory*, Probab. Appl., Springer-Verlag, London, UK, 2005.
[19] C. Cotar, G. Friesecke, and C. Klüppelberg, *Density functional theory and optimal transportation with Coulomb cost*, Comm. Pure Appl. Math., 66 (2013), pp. 548–599, https://doi.org/10.1002/cpa.21437.
[20] P. Courrège, *Sur la forme intégro-différentielle des opérateurs de C_k^∞ dans C satisfaisant au principe du maximum*, Séminaire Brelot-Choquet-Deny. Théorie du potentiel, 10 (1965–1966), pp. 1–38, http://eudml.org/doc/109338.
[21] R. V. Craiu and C. Lemieux, *Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling*, Stat. and Comput., 17 (2007), p. 109.
[22] R. V. Craiu and X.-L. Meng, *Multiprocess parallel antithetic coupling for backward and forward Markov chain Monte Carlo*, Ann. Statist., 33 (2005), pp. 661–697, https://doi.org/10.1214/009053604000001075.
[23] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite-Dimensional Systems*, London Math. Soc. Lecture Note Series 229, Cambridge University Press, Cambridge, UK, 1996, https://doi.org/10.1017/CBO9780511662829.
[24] M. H. A. Davis, *Piecewise-deterministic Markov processes: A general class of nondiffusion stochastic models*, J. Roy. Statist. Soc. Ser. B, 46 (1984), pp. 353–388.
[25] M. H. A. Davis, *Markov Models and Optimization*, Monogr. Statist. Appl. Probab. 49, Chapman & Hall, London, UK, 1993, https://doi.org/10.1007/978-1-4899-4483-2.
[26] P. Del Moral, *Mean Field Simulation for Monte Carlo Integration*, Monogr. Statist. Appl. Probab. 126, CRC Press, Boca Raton, FL, 2013.
[27] P. Dellaportas and I. Kontoyiannis, *Control variates for estimation based on reversible Markov chain Monte Carlo samplers*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 74 (2012), pp. 133–161.
[28] A. B. Duncan, T. Lelièvre, and G. A. Pavliotis, *Variance reduction using nonreversible Langevin samplers*, J. Stat. Phys., 163 (2016), pp. 457–491, https://doi.org/10.1007/s10955-016-1491-2.
[29] A. B. Duncan, N. Nüsken, and G. A. Pavliotis, *Using perturbed underdamped Langevin dynamics to efficiently sample from probability distributions*, J. Stat. Phys., 169 (2017), pp. 1098–1131, https://doi.org/10.1007/s10955-017-1906-8.
[30] A. Durmus, A. Guillin, and P. Monmarché, *Piecewise Deterministic Markov Processes and Their Invariant Measure*, arXiv:1807.05421, 2018.
[31] A. Eberle, *Stochastic Analysis.* Lecture Notes, 2015. https://wt.iam.uni-bonn.de/fileadmin/WT/Inhalt/people/Andreas_Eberle/StoAn15/StochasticAnalysis2015.pdf, accessed 12/03/2018.
[32] A. Eberle, *Reflection couplings and contraction rates for diffusions*, Probab. Theory Related Fields, 166 (2016), pp. 851–886.
[33] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Grad. Texts, in Math. 194, Springer-Verlag, New York, 2000.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

[34] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Ser. Probab. Statist., John Wiley & Sons, New York, 1986, http://dx.doi.org/10.1002/9780470316658.
[35] P. Fearnhead, J. Bierkens, M. Pollock, and G. O. Roberts, *Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo*, arXiv:1611.07873, 2016.
[36] A. Frigessi, J. Gasemyr, and H. Rue, *Antithetic coupling of two Gibbs sampler chains*, Ann. Statist., 28 (2000), pp. 1128–1149.
[37] R. Ghanem, D. Higdon, and H. Owhadi, *Handbook of Uncertainty Quantification*, Springer, Cham, 2017.
[38] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer., 24 (2015), pp. 259–328, https://doi.org/10.1017/S096249291500001X, https://doi.org/10.1017/S096249291500001X.
[39] P. W. Glynn and S. P. Meyn, *A Liapounov bound for solutions of the Poisson equation*, Ann. Probab. 24 (1996), https://doi.org/10.1214/aop/1039639370.
[40] P. W. Glynn and C. Rhee, *Exact estimation for Markov chain equilibrium expectations*, J. Appl. Probab., 51 (2014), pp. 377–389.
[41] J. B. Goodman and K. K. Lin, *Coupling control variates for Markov chain Monte Carlo*, J. Comput. Phys., 228 (2009), pp. 7127–7136.
[42] M. Hairer and J. C. Mattingly, *Yet another look at Harris' ergodic theorem for Markov chains*, in Seminar on Stochastic Analysis, Random Fields and Applications VI, Progr. Probab. 63, Birkhäuser, Basel, 2011, pp. 109–117, https://doi.org/10.1007/978-3-0348-0021-1_7.
[43] E. J. Hall, M. A. Katsoulakis, and L. Rey-Bellet, *Uncertainty quantification for generalized Langevin dynamics*, J. Chem. Phys., 145 (2016), 224108.
[44] J. Heng, A. Doucet, and Y. Pokern, *Gibbs Flow for Approximate Transport with Applications to Bayesian Computation*, arXiv:1509.08787, 2015.
[45] J. Heng and P. E. Jacob, *Unbiased Hamiltonian Monte Carlo with Couplings*, arXiv:1709.00404, 2017.
[46] C. Holmes and A. Jasra, *Antithetic methods for Gibbs samplers*, J. Comput. Graph. Statist., 18 (2009), pp. 401–414.
[47] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu, *Accelerating diffusions*, Ann. Appl. Probab., 15 (2005), pp. 1433–1444, http://dx.doi.org/10.1214/105051605000000025.
[48] N. Jacob, *Pseudo Differential Operators and Markov Processes. Vol.* I, Imperial College Press, London, UK, 2001, https://doi.org/10.1142/9781860949746.
[49] V. E. Johnson, *Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths*, J. Amer. Statist. Assoc., 91 (1996), pp. 154–166.
[50] V. E. Johnson, *A coupling-regeneration scheme for diagnosing convergence in Markov chain Monte Carlo algorithms*, J. Amer. Statist. Assoc., 93 (1998), pp. 238–248.
[51] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer-Verlag, New York, 2002, https://doi.org/10.1007/978-1-4757-4015-8.
[52] W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*, Ann. Probab., 15 (1987), pp. 690–707.
[53] T. Komorowski, C. Landim, and S. Olla, *Fluctuations in Markov processes*, Grundlehren Math. Wiss. 345, Springer, Heidelberg, 2012, http://dx.doi.org/10.1007/978-3-642-29880-6.
[54] D. P. Kroese, T. Taimre, and Z. I. Botev, *Handbook of Monte Carlo Methods*, John Wiley & Sons, New York, 2013.
[55] F. Kühn, *Lévy Matters.* VI, Lecture Notes in Math., Springer, Cham, 2017.
[56] F. Kühn, *Existence of (Markovian) Solutions to Martingale Problems Associated with Lévy-Type Operators*, arXiv:1803.05646, 2018.
[57] J. Kwak et al., *An antithetic coupling approach to multi-chain based CSMA scheduling algorithms*, in Proceedings of the 35th Annual IEEE International Conference on Computer Communications, IEEE, 2016, pp. 1–9.
[58] B. Leimkuhler, C. Matthews, and G. Stoltz, *The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics*, IMA J. Numer. Anal., 36 (2016), pp. 13–79, https://doi.org/10.1093/imanum/dru056.
[59] B. Leimkuhler, C. Matthews, and J. Weare, *Ensemble preconditioning for Markov chain Monte Carlo simulation*, Stat. Comput., 28 (2018), pp. 277–290, https://doi.org/10.1007/s11222-017-9730-1.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

[60] T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations*, Imperial College Press, London, UK, 2010, http://dx.doi.org/10.1142/9781848162488.
[61] T. Lelièvre and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numer., 25 (2016), pp. 681–880, http://dx.doi.org/10.1017/S0962492916000039.
[62] V. Lemaire, G. Pagès, and F. Panloup, *Invariant measure of duplicated diffusions and application to Richardson-Romberg extrapolation*, Ann. Inst. Henri Poincaré Probab. Stat., 51 (2015), pp. 1562–1596, https://doi.org/10.1214/13-AIHP591.
[63] T. Lindvall, *Lectures on the Coupling Method*, Dover Publications, Mineola, NY, 2002.
[64] Q. Liu and D. Wang, *Stein variational gradient descent: A general purpose Bayesian inference algorithm*, in Advances In Neural Information Processing Systems 29, 2016, pp. 2378–2386.
[65] L. Lorenzi and M. Bertoldi, *Analytical Methods for Markov Semigroups*, Pure Appl. Math. (Boca Raton) 283, Chapman & Hall/CRC, Boca Raton, FL, 2007.
[66] R. J. McCann, B. Pass, and M. Warren, *Rectifiability of optimal transportation plans*, Canad. J. Math., 64 (2012), pp. 924–934, https://doi.org/10.4153/CJM-2011-080-6.
[67] M. Michel, S. C. Kapfer, and W. Krauth, *Generalized event-chain Monte Carlo: Constructing rejection-free global-balance algorithms from infinitesimal steps*, J. Chem. Phys., 140 (2014), p. 054116, https://doi.org/10.1063/1.4863991.
[68] M. Michel and S. Sénécal, *Forward Event-Chain Monte Carlo: A General Rejection-Free and Irreversible Markov Chain Simulation Method*, arXiv:1702.08397v2, 2017.
[69] A. Mijatović and J. Vogrinc, *On the Poisson equation for Metropolis–Hastings chains*, Bernoulli, 24 (2018), pp. 2401–2428.
[70] R. M. Neal, *Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation*, in Learning in Graphical Models, Springer, 1998, pp. 205–228.
[71] R. M. Neal, *Circularly-Coupled Markov Chain Sampling*, arXiv:1711.04399, 2017.
[72] R. M. Neal, *MCMC using Hamiltonian dynamics*, in Handbook of Markov Chain Monte Carlo, Chapman & Hall/CRC, Boca Raton, FL, 2011, pp. 113–162.
[73] R. M. Neal and R. L. Pinto, *Improving Markov Chain Monte Carlo Estimators by Coupling to an Approximating Chain*, Technical Report No. 0101, Department of Statistics, University of Toronto, Toronto, Canada, 2001.
[74] B. Øksendal, *Stochastic Differential Equations*, 6th ed., Universitext, Springer-Verlag, Berlin, 2003, https://doi.org/10.1007/978-3-642-14394-6.
[75] M. Ottobre, *Markov chain Monte Carlo and irreversibility*, Rep. Math. Phys., 77 (2016), pp. 267–292.
[76] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Springer, Cham, 2014.
[77] G. Peyré and M. Cuturi, *Computational optimal transport*, Found. Trends Mach. Learn., 11 (2019), pp. 355–607.
[78] R. Pinnau, C. Totzeck, O. Tse, and S. Martin, *A consensus-based model for global optimization and its mean-field limit*, Math. Models Methods Appl. Sci., 27 (2017), pp. 183–204.
[79] J. Propp and D. Wilson, *Coupling from the past: a user's guide*, in Microsurveys in Discrete Probability American Mathematical Society, Providence, RI, 1998, pp. 181–192.
[80] S. Reich and C. Cotter, *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press, New York, 2015, https://doi.org/10.1017/CBO9781107706804.
[81] C. R. Reeves and J. E. Rowe, *Genetic Algorithms–Principles and Perspectives*, Springer, Cham, 2004.
[82] L. Rey-Bellet and K. Spiliopoulos, *Improving the convergence of reversible samplers*, J. Stat. Phys., 164 (2016), pp. 472–494, https://doi.org/10.1007/s10955-016-1565-1.
[83] F. Rigat and A. Mira, *Parallel hierarchical sampling: A general-purpose interacting markov chains monte carlo algorithm*, Comput. Statist. Data Anal., 56 (2012), pp. 1450–1467.
[84] J. Roussel and G. Stoltz, *Spectral methods for Langevin dynamics and associated error estimates*, ESAIM Math. Model. Numer. Anal., 52 (2018), pp. 1051–1083.
[85] M. Rousset and G. Stoltz, *An Interacting Particle System Approach for Molecular Dynamics*. preprint, Aug. 2005, https://hal.archives-ouvertes.fr/hal-00008276.
[86] R. Schilling, *Conservativeness and extensions of Feller semigroups*, Positivity, 2 (1998), pp. 239–256, https://doi.org/10.1023/A:1009748105208.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.

[87] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, SIAM, Philadelphia PA, 2013.
[88] T. J. Sullivan, *Introduction to Uncertainty Quantification*, Springer, Cham, 2015.
[89] G. Teschl, *Mathematical Methods in Quantum Mechanics*, Grad. Stud. Math., American Mathematical Society, Providence, RI, 2014.
[90] H. Thorisson, *Coupling, Stationarity, and Regeneration*, Probab. Appl. (N. Y.), Springer-Verlag, New York, 2000, https://doi.org/10.1007/978-1-4612-1236-2.
[91] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, and A. Doucet, *Piecewise Deterministic Markov Chain Monte Carlo*, arXiv:1707.05296, 2017.
[92] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, American Mathematical Society, Providence, RI, 2003, https://doi.org/10.1007/b12016.
[93] C. Villani, *Hypocoercivity*, American Mathematical Society, Providence, RI, 2009.
[94] C. Villani, *Optimal Transport*, Grundlehren Math. Wiss. 338, Springer-Verlag, Berlin, 2009, https://doi.org/10.1007/978-3-540-71050-9.

Downloaded 05/15/19 to 155.198.12.56. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM and ASA. Unauthorized reproduction of this article is prohibited.