ADAPTED WASSERSTEIN DISTANCE BETWEEN THE LAWS OF SDES

JULIO BACKHOFF-VERAGUAS, SIGRID KÄLLBLAD, AND BENJAMIN A. ROBINSON

Abstract. We consider the adapted optimal transport problem between the laws of Markovian stochastic differential equations (SDE) and establish the optimality of the synchronous coupling between these laws. The proof of this result is based on time-discretisation and reveals an interesting connection between the synchronous coupling and the celebrated discrete-time Knothe–Rosenblatt rearrangement. We also prove a result on equality of topologies restricted to a certain subset of laws of continuous-time processes.

1. Introduction

For all their merits, the concepts of weak convergence and Wasserstein distances have proven to be insufficient for applications involving stochastic processes where filtrations and the flow of information play a pivotal role. For instance, neither usual stochastic optimisation problems (such as optimal stopping or utility maximisation) nor Doob–Meyer decompositions behave continuously with respect to these topologies. Over the last decades, several approaches have been proposed to overcome these shortcomings; in this paper our focus is on one such notion, namely the so-called adapted Wasserstein distance.

More precisely, we study the adapted Wasserstein distance between the laws of solutions of one-dimensional stochastic differential equations (SDE) when the space of continuous functions is equipped with the L^p-metric. Imposing fairly general conditions on the coefficients of the SDEs — typically amounting to Markovianity and mild regularity assumptions — our contribution can be summarised as follows:

(i) a characterisation of the optimal coupling attaining the adapted Wasserstein distance;
(ii) a time-discretisation method allowing us to derive most continuous-time statements from their more elementary discrete-time counterparts;
(iii) a result stating that the topology induced by the adapted Wasserstein distance coincides with several topologies (including the weak topology) when restricting the coefficients of the SDEs to belong to certain equicontinuous families.

At a conceptual level, our main contribution is to connect two hitherto unrelated objects: on the one hand, the *synchronous coupling* of SDEs, which is the coupling arising when letting a single Wiener process drive two SDEs; on the other hand, the *Knothe–Rosenblatt* rearrangement, which is a celebrated discrete-time adapted coupling that preserves the lexicographical order. We argue that, in a certain sense, the synchronous coupling is the continuous-time counterpart of the Knothe–Rosenblatt rearrangement.

Concerning the contributions (i) and (iii) above, similar statements have been made in the pioneering work of Bion-Nadal and Talay [11] for the problem of optimally controlling the correlation between SDEs with smooth coefficients. Since we will prove that the adapted Wasserstein distance admits such a control reformulation, our results will contain those of [11] as particular cases.

Date: September 29, 2022.
Key words and phrases. Knothe–Rosenblatt rearrangement, optimal transport, stochastic differential equations, optimal couplings.
BR acknowledges the support of the Austrian Science Fund (FWF) projects Y782-N25 and P35519.

Adapted Wasserstein distance. In order to make our contributions precise, let us formally define our key object of study, the adapted Wasserstein distance, and give some motivation for its introduction.

To this end, consider $\Omega = C([0,1],\mathbb{R})$ and endow this space with the sup-norm and its corresponding Borel σ-field, and $\Omega \times \Omega$ with the corresponding product σ-field; we write ω and $\bar{\omega}$ for the first and second components of the canonical process on $\Omega \times \Omega$. For any two probability measures μ, ν on Ω, the set $\mathrm{Cpl}(\mu,\nu)$ of *couplings* between μ and ν consists of all probability measures π on $\Omega \times \Omega$ with *marginals* μ, ν; that is, $\int_{A\times\Omega} \pi(\mathrm{d}x, \mathrm{d}y) = \mu(A)$ and $\int_{\Omega\times B} \pi(\mathrm{d}x, \mathrm{d}y) = \nu(B)$, for any measurable sets $A, B \subset \Omega$. For $p \geq 1$, we also write $\mathcal{P}_p$ for the set of probability measures μ on Ω such that the canonical process has finite L^p moments with respect to $\mathrm{d}t \times \mu$, where $\mathrm{d}t$ denotes the Lebesgue measure on $[0,1]$. The classical p-Wasserstein distance $\mathcal{W}^p$ with respect to the L^p-distance on the underlying space Ω then takes the following form (see, for example, Villani [34]):

$$\mathcal{W}_p^p(\mu,\nu) := \inf_{\pi\in\mathrm{Cpl}(\mu,\nu)} \mathbb{E}^\pi \left[\int_0^1 |\omega_t - \bar{\omega}_t|^p \, \mathrm{d}t \right], \qquad \mu,\nu \in \mathcal{P}_p.$$

This distance notably fails to take the flow of information into account. For example, the values of optimisation problems for continuous-time stochastic processes may not be continuous in Wasserstein distance with respect to the reference measure; see Example 6.1. As a remedy, we define the adapted Wasserstein distance by restricting to couplings that respect the asymmetric flow of information originating from the processes. To formalise this, let $(\mathcal{F}_t)_{t\in[0,1]}$ be the canonical filtration on Ω and write $\mathcal{F}_t^\mu$ for the completion of $\mathcal{F}_t$ under a probability measure μ on Ω; i.e. $\mathcal{F}_t^\mu$ is the sigma-algebra generated by $\mathcal{F}_t$ and the null sets for μ. For $\pi \in \mathrm{Cpl}(\mu,\nu)$, let $\pi_x(\mathrm{d}y)$ be a regular disintegration kernel for which $\pi(\mathrm{d}x, \mathrm{d}y) = \mu(dx)\pi_x(\mathrm{d}y)$, and let $S(x,y) := (y,x)$, for $(x,y) \in \Omega \times \Omega$. Following [4], we define bi-causal couplings as follows.

Definition 1.1 (bi-causal couplings)**.** The set of *causal couplings* $\mathrm{Cpl}_{\mathrm{c}}(\mu,\nu)$ consists of all $\pi \in \mathrm{Cpl}(\mu,\nu)$ such that, for all $t \in [0,1]$ and $A \in \mathcal{F}_t$,

$$\omega \mapsto \pi_\omega(A) \text{ is } \mathcal{F}_t^\mu\text{-measurable.}$$

The set of *bi-causal couplings* $\mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)$ consists of all $\pi \in \mathrm{Cpl}_{\mathrm{c}}(\mu,\nu)$ with $S_\#\pi \in \mathrm{Cpl}_{\mathrm{c}}(\nu,\mu)$.

The intuition behind the concept of causality is perhaps most easily grasped in a discrete-time setup; i.e. when $t \in \{1,2,\dots,T\}$. The defining property of causality can then be phrased as requiring, with obvious adaptation of notation, that $\pi((\bar{\omega}_1,...,\bar{\omega}_n) \in A|\omega_1,...,\omega_T) = \pi((\bar{\omega}_1,...,\bar{\omega}_n) \in A|\omega_1,...,\omega_n)$, for all $A \in \mathcal{B}(\mathbb{R}^n)$, $n = 1,...,T$. In this discrete-time setting, if the coupling is further supported on the graph of a function, say $\varphi : \mathbb{R}^T \to \mathbb{R}^T$ (i.e. a Monge map), then causality amounts to $\varphi(x_1,...,x_T) = (\varphi_1(x_1), \varphi_2(x_1,x_2), \dots, \varphi_T(x_1,\dots,x_T))$, for some functions $\varphi_n : \mathbb{R}^n \to \mathbb{R}$, $n = 1,...,T$. Put into words: 'one cannot look into the future when deciding where to allocate mass at a given time'. This emphasises the role played by the flow of information; i.e. filtrations. We refer to Beiglböck and Lacker [9] for a detailed exposition of how Monge maps relate to general transport plans in the presence of causality constraints.

With the above notation at hand, we are now ready to define our version of the *adapted* (bi-causal) *Wasserstein distance*[1] $\mathcal{AW}_p$, $p \geq 1$:

$$\mathcal{AW}_p^p(\mu,\nu) := \inf_{\pi\in\mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \mathbb{E}^\pi \left[\int_0^1 |\omega_t - \bar{\omega}_t|^p \, \mathrm{d}t \right], \qquad \mu,\nu \in \mathcal{P}_p. \tag{1.1}$$

[1] We note that in [4] the terminology adapted Wasserstein distance was used for a bi-causal Wasserstein distance featuring a different cost function compared to ours. The cost function used in [4] was based on comparing the drift and martingale parts of Doob–Meyer decompositions separately and was tailor-made for financial applications.

Optimality of the synchronous coupling. Throughout this article, we consider probability measures that are the laws of solutions of stochastic differential equations of the following type:

$$\mathrm{d}X_t^{b,\sigma} = b(X_t^{b,\sigma})\mathrm{d}t + \sigma(X_t^{b,\sigma})\mathrm{d}W_t, \quad X_0^{b,\sigma} = x_0, \quad t \in [0,1], \tag{1.2}$$

where $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}_+$ are measurable functions satisfying some general conditions to be specified. If a unique strong solution $X^{b,\sigma}$ exists, we denote by $\mu^{b,\sigma} := \mathrm{Law}(X^{b,\sigma})$ the induced probability measure on Ω. Throughout, without loss of generality, we suppose that all SDEs are equipped with the same fixed initial condition $x_0 \in \mathbb{R}$ and omit it from the notation.

Now, given two such measures, $\mu^{b,\sigma}$ and $\mu^{\bar{b},\bar{\sigma}}$, we can construct a specific coupling between them as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a Wiener process W is supported, and let $X^{b,\sigma}$ and $X^{\bar{b},\bar{\sigma}}$ be solutions of (1.2) with coefficients (b,σ) and $(\bar{b},\bar{\sigma})$, respectively, where the same Wiener process W is driving both SDEs. This naturally defines the coupling $\mathbb{P} \circ (X^{b,\sigma}, X^{\bar{b},\bar{\sigma}})^{-1} \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{\mu},\bar{\sigma}})$ — this specific coupling will play a pivotal role throughout the article and we name it the *synchronous coupling.*

Our first main result establishes general conditions under which this coupling is optimal.

Assumption 1.2. The coefficients $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}_+$ in (1.2) are continuous, have linear growth, and are such that pathwise uniqueness holds for (1.2).

Theorem 1.3. *Suppose that* (b,σ) *and* $(\bar{b},\bar{\sigma})$ *satisfy Assumption 1.2. Then, for any* $p \geq 1$, *the synchronous coupling attains the infimum in* (1.1) *defining* $\mathcal{AW}_p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$.

We also prove the conclusion of Theorem 1.3 in Proposition 3.28 under a different set of assumptions, which allow for the drift coefficients to be discontinuous.

It follows from the above result and [11] that, for SDEs with sufficiently regular coefficients, the adapted Wasserstein distance coincides with the distance obtained when optimising the cost in (1.1) over the smaller class of couplings induced by solutions of the pair of SDEs equipped with *correlated* Wiener processes (see Definition 2.1). Notably, the distance in [11] was introduced in this way. The a priori establishment of this reformulation of the adapted Wasserstein distance in terms of an associated control problem (where one controls the degree of correlation) constitutes a crucial part of our analysis (see Proposition 2.2).

Discrete approximation methods and stability. A key object of study in this paper is the Knothe–Rosenblatt rearrangement [22], which is a multi-dimensional generalisation of the classical monotone rearrangement; see Figure 1 for an illustration. Consider now the discrete problem of optimally coupling two laws on $\mathbb{R}^n$, $n \in \mathbb{N}$. When restricting to bi-causal couplings and imposing certain monotonicity properties on the marginal laws, it is known that the Knothe–Rosenblatt rearrangement is optimal for this problem; see Rüschendorf [31], the more recent work [7], and the more detailed account given below. This fact lies at the heart of our proof; combined with a careful approximation argument, it yields the optimality of the synchronous coupling. Indeed, in contrast with [11], our proof relies on an approximation procedure where we first solve the associated discrete-time problem and then pass to the limit. Our method of proof thus unveils the informational and structural similarities between the Knothe–Rosenblatt and synchronous couplings. For this reason we advocate the interpretation of the synchronous coupling as the continuous analogue of the Knothe–Rosenblatt rearrangement.

As a by-product, we also obtain explicit approximation results on bi-causal couplings and stochastic differential equations. First and foremost, we provide results on how to explicitly approximate bi-causal couplings. The approximation notably relies on the classical Euler–Maruyama scheme for solutions of stochastic differential equations; specifically, given coefficients (b,σ) and $(\bar{b},\bar{\sigma})$, the Euler–Maruyama scheme $(X^n, \bar{X}^n)$ is given by $(X_0^n, \bar{X}_0^n) = (x_0, x_0)$ and, for $h = 1/n$,

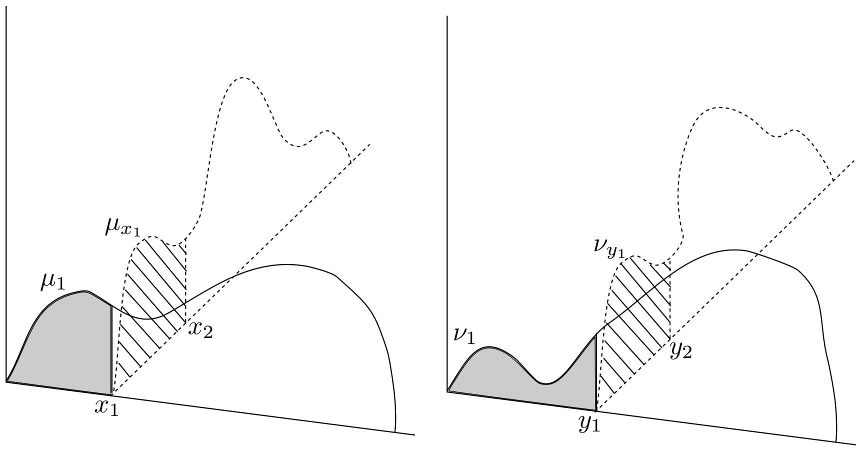

FIGURE 1. Illustration of the Knothe–Rosenblatt rearrangement in two dimensions. The first marginals of μ, ν are denoted μ_1, ν_1, and the conditional distributions by μ_{x_1}, ν_{y_1}. Similarly shaded regions have the same area.

$k = 0, ..., n-1$,

$$\left\{ \begin{array}{lcl} X_t^n & = & X_{kh}^n + b(X_{kh}^n)\,(t-kh) + \sigma(X_{kh}^n)\,(W_t - W_{kh}) \\ \bar{X}_t^n & = & \bar{X}_{kh}^n + \bar{b}(\bar{X}_{kh}^n)\,(t-kh) + \bar{\sigma}(\bar{X}_{kh}^n)\,\big(\bar{W}_t - \bar{W}_{kh}\big) \end{array} \right. , \quad t \in (kh, (k+1)h]. \tag{1.3}$$

With the adapted Wasserstein distance defined analogously to (1.1) also for marginal distributions on $\Omega \times \Omega$ (see (2.3) below), we then have the following result:

Theorem 1.4. *Let $b, \bar{b} : \mathbb{R} \to \mathbb{R}$, $\sigma, \bar{\sigma} : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz, and let $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$. Then there exists a probability space supporting two correlated Wiener processes W and $\bar{W}$ such that the joint law of the processes $(X^n, \bar{X}^n)$ given by (1.3) satisfies $\mathcal{AW}_p(\mathrm{Law}(X^n, \bar{X}^n), \pi) \xrightarrow{n\to\infty} 0$, $p \geq 1$.*

We also obtain stability results for the parameter dependence of stochastic differential equations driven by correlated Wiener processes (see Proposition 4.6), even in the presence of path-dependent coefficients. We note that these results on discretisation and stability are crucial for our analysis and also of independent interest. For instance, the discretisation scheme that we introduce could potentially be used to compute the adapted Wasserstein distance numerically; however, we do not investigate this in the present paper.

The synchronous distance and the associated topology. A related and relevant object of study is the topology induced by the adapted Wasserstein distance. In particular, since there are alternative distances with which the space of (laws of) stochastic processes could be equipped, a central question is how the topologies induced by such distances are related.

We have already mentioned the classical Wasserstein distance. Naturally, the classical (metrizable) weak convergence topology could also be used. Moreover, for $p \geq 1$, one may consider the (asymmetric) *causal* Wasserstein distance $\mathcal{CW}_p(\mu, \nu)$, $\mu, \nu \in \mathcal{P}_p$, defined analogously to the adapted (bi-causal) Wasserstein distance, by replacing $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ with $\mathrm{Cpl}_{\mathrm{c}}(\mu, \nu)$ in (1.1). In this asymmetric setting, we say that μ_n converges to μ in $\mathcal{CW}_p$, if $\mathcal{CW}_p(\mu, \mu_n) \to 0$. We also consider its symmetrised version $\mathcal{SCW}_p(\mu, \nu) = \max(\mathcal{CW}_p(\mu, \nu), \mathcal{CW}_p(\nu, \mu))$, $\mu, \nu \in \mathcal{P}_p$. Finally, inspired by the pivotal role played by the synchronous coupling, we introduce the *synchronous distance*[2] $\mathcal{SW}_p$, defined by

$$\mathcal{SW}_p^p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}}) := \mathbb{E}\left[\int_0^1 \big|X_t^{b,\sigma} - X_t^{\bar{b},\bar{\sigma}}\big|^p \mathrm{d}t\right], \quad p \geq 1,$$

[2]We acknowledge private communication with Mathias Beiglböck, Gudmund Pammer and Alexander Posch, from which we learned that, in discrete time, the distance induced by the *Knothe–Rosenblatt rearrangement* is topologically equivalent to the adapted Wasserstein distance. Our Theorem 1.5 can be seen as a continuous-time analogue of this result, with the *synchronous coupling* in place of the Knothe–Rosenblatt rearrangement.

where $X^{b,\sigma}$ and $X^{\bar{b},\bar{\sigma}}$ are the p-integrable solutions of the SDE (1.2) with coefficients (b,σ) and $(\bar{b},\bar{\sigma})$, evaluated on some probability space with respect to the *same* Wiener process W (c.f. the definition of the synchronous coupling). This distance is notably stronger than all of the above-mentioned distances.

Coming back to the question of how topologies induced by various distances are related, for discrete-time processes [5] established that a number of well-studied distances generate the same topology; see below for more on this and further related references. We provide a result in this spirit for a certain class of continuous-time processes. Specifically, we consider solutions of the SDE (1.2) with coefficients belonging to the following set:

$$\mathcal{A}^{\Lambda} = \{\varphi \in C(\mathbb{R},\mathbb{R}) : |\varphi(x)-\varphi(y)| \le \Lambda|x-y| \text{ and } |\varphi(0)| \le \Lambda,\ x,y \in \mathbb{R}\}\,, \quad \Lambda > 0.$$

Theorem 1.5. *Restricted to the set* $\mathcal{P}^{\Lambda} = \{\mu^{b,\sigma} : b,\sigma \in \mathcal{A}^{\Lambda}\}$, $\Lambda > 0$, *the topologies induced by the following metrics all coincide and are independent of* $p \in [1,\infty)$*:*

— $\mathcal{SW}_p$*, the synchronous distance;*
— $\mathcal{AW}_p$*, the adapted Wasserstein distance;*
— $\mathcal{SCW}_p$*, the symmetrised causal Wasserstein distance;*
— $\mathcal{W}_p$*, the Wasserstein distance.*

This common topology is further equal to the topology of $\mathcal{CW}_p$ *convergence, the topology of weak convergence when we equip* Ω *with the* $L^p(\mathrm{d}t)$ *norm, for arbitrary* $p \in [1,\infty]$*, and also to the topology of convergence in finite-dimensional distributions. Moreover,* $\mathcal{P}^{\Lambda}$ *is compact in this common topology.*

We remark that the above topologies remain equal when, in the definition of any of the metrics, we replace the cost $\int_0^1 |\omega_t - \bar{\omega}_t|^p \mathrm{d}t$ by $\sup_{t\in[0,1]} |\omega_t - \bar{\omega}_t|^p$.

Related literature. The condition of *causality* has a long history and can be traced back, at least, to the work on existence of solutions of stochastic differential equations by Yamada and Watanabe [35]; it has also appeared under the name of *compatibility* in Kurtz [23]. The concept was recently popularised and studied in a continuous-time framework by Lassalle [24], and systematically investigated for discrete-time processes using dynamic programming arguments in [7] (see also [18] for a recursive approach to a closely related optimal stopping problem). We refer to Beiglböck and Lacker [9] for further historical remarks and for an account of the connections to the filtering literature. To the best of our knowledge, the symmetric condition of *bi-causality* first appeared in Rüschendorf [31], and for a more recent account we refer again to [7]; see below for more on these two articles. A distance based on the bi-causality condition was independently introduced and studied, under the name of *nested distance*, in a series of papers by Pflug and Pichler; see, for example, [27] and [28] and the references therein. The concept of causality aside, numerous alternative approaches to incorporating the flow of information into process distances can be found in the literature. Most notably, albeit in different ways, the seminal works of Aldous [3] and Hellwig [16] both rely on incorporating the distance between certain conditional disintegration kernels of the processes. The above-mentioned article [5] established that, for discrete-time processes, both of these distances as well as the adapted Wasserstein distance and the nested distance generate the same topology; we refer to [6] and Eder [15] for further properties of this common topology.

In continuous time, for diffusion processes, the adapted Wasserstein distance was first introduced in [11]. For general continuous semi-martingales, an adapted Wasserstein distance was introduced and studied in [4], although they notably consider a different cost function compared to the one studied in this paper. We refer to Acciaio et al. [1] and Acciaio et al. [2] for further studies of adapted distances in continuous time; see also [7, Section 2] for a detailed exposition of the related literature within mathematical finance.

The *Knothe–Rosenblatt rearrangement*, also referred to as the Knothe–Rosenblatt coupling, was introduced in [30] and [22] as a multi-dimensional generalisation of the monotone rearrangement. The optimality of the Knothe–Rosenblatt rearrangement, under certain conditions on the marginal distributions, was first established by Rüschendorf [31] for a two-step discrete bi-causal optimal transport problem (although the terminology was not yet in place). The proof relies on well-known optimality properties of the monotone rearrangement. In [7], based on a recursive argument, the optimality of the Knothe–Rosenblatt rearrangement was then generalised to a multi-stage discrete problem; this result underpins our analysis.

In continuous time, when restricting to marginal laws corresponding to solutions of certain one-dimensional SDEs, the optimality of the coupling that we refer to as the *synchronous coupling* was first established in [11]. Although they consider the optimisation problem in general dimensions, they only identify an optimiser in dimension one. In [11] the authors take a stochastic control approach and their proofs rely on verification arguments for the associated Hamilton–Jacobi–Bellman (HJB) equation. We note that such stochastic control arguments provide the natural continuous-time analogue of the above-mentioned recursive arguments used to prove optimality of the Knothe–Rosenblatt rearrangement in discrete time. Indeed, the crucial observation in the verification argument of [11] is that, under sufficient regularity assumptions, the second order cross-derivative of the value function is negative, while the algebraic analogue of this conditions is key to the discrete-time proofs in [31] and [7]. However, the use of classical solutions of the HJB equation, as employed in [11], inevitably requires the coefficients of the SDEs to be smooth enough for the associated stochastic flows to be differentiable. In this paper, we provide a probabilistic proof of the optimality of the synchronous coupling in dimension one. This enables us to relax the assumptions on the coefficients and establish this optimality property in its natural generality.

Our time-discretisation method also suggests a possible approach to numerical approximation of adapted Wasserstein distances between the laws of solutions of SDEs. However, we do not explore this direction further. For existing numerical methods for computing adapted Wasserstein distances, see the recent work of Eckstein and Pammer [14], and the references therein.

As for the study of different topologies, we have already mentioned the comprehensive account given in [5], where a number of distances were shown to generate the same topology on the space of discrete-time stochastic processes. In fact a stronger result is true. In Bartl et al. [8] and Pammer [26], the authors go beyond the convention of identifying a process with its law and instead consider processes equipped with a filtration. In discrete time, it is shown that all topologies that are strong enough to encode the information of the filtration still coincide, even in this generalised setting. In [11], a result of a similar flavour was provided within a continuous-time set-up when restricting to processes being solutions of SDEs with sufficiently smooth coefficients; see Propositions 1.8 and 1.9 therein. Our Theorem 1.5 generalises the results in [11]; we note that our proof is remarkably simple as it is a straightforward application of our stability result in Proposition 4.6.

Structure of the article. The remainder of the article is organised as follows.

In Section 2, we establish that the distance studied in [11] coincides with the adapted Wasserstein distance that we consider. We then prove Theorem 1.4, showing that bi-causal couplings between laws of SDEs can be approximated in adapted Wasserstein distance on the product space by the joint laws of time-discretisations of the SDEs.

In Section 3, we prove Theorem 1.3 under the additional assumption that the coefficients of the SDEs are Lipschitz. We discuss the optimality of the Knothe–Rosenblatt rearrangement in discrete time in Section 3.1, before defining a variation of the classical Euler–Maruyama scheme for SDEs and proving its convergence in Section 3.2. We combine the discrete-time optimality and convergence results to prove optimality of the synchronous coupling under Lipschitz conditions in

Section 3.3. Finally, in Section 3.4, we apply the same techniques to prove a variation of Theorem 1.3 under a different set of assumptions that allow for discontinuities in the drift coefficients.

We complete the proof of Theorem 1.3 in Section 4, by proving a stability result for SDEs in Section 4.1, and then combining this with the results for Lipschitz coefficients in Section 4.2.

In Section 5, we prove Theorem 1.5 on the coinciding topologies, employing again the stability result of the previous section.

Finally, in Section 6, we collect some examples. The first example motivates the introduction of the adapted Wasserstein distance. We then consider possible extensions of Theorem 1.3; we present two examples of SDEs with non-Markovian coefficients, illustrating that the synchronous coupling may not be optimal for such SDEs but that there exist cases where optimality is preserved.

2. Preliminary results on bi-causal couplings and approximation in $\mathcal{AW}_p$

Throughout this paper we work in dimension one. Define the space of continuous paths $\Omega := C([0,1],\mathbb{R})$ with the canonical filtration $(\mathcal{F}_s)_{s\in[0,1]}$ and $\mathcal{F} = \mathcal{F}_1$, the corresponding Borel sigma-field. For functions $b:\mathbb{R}\to\mathbb{R}$, $\sigma:\mathbb{R}\to\mathbb{R}_+$ and a constant $x_0\in\mathbb{R}$, if there exists a unique strong solution of the SDE (1.2), then we write $\mu^{b,\sigma}$ for its law; without loss of generality, we suppose that all SDEs start from the same initial value x_0 and so we omit x_0 from any notation. The set $\mathcal{P}^*$ is the subset of probability measures on $(\Omega,\mathcal{F})$ which consists of all such probability measures $\mu^{b,\sigma}$.

We will repeatedly make use of the following definition of correlated Wiener processes.

Definition 2.1 (correlated Wiener process)**.** Let $W,\bar W$ be standard real-valued Wiener processes on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ and let $\mathcal{G}$ be the completion under $\mathbb{P}$ of the filtration jointly generated by $W,\bar W$.

Given a $\mathcal{B}([0,1])\otimes\mathcal{G}$-progressively measurable function $\rho:[0,1]\times\Omega\times\Omega\to[-1,1]$, we say that the two-dimensional process $(W,\bar W)$ is a *ρ-correlated Wiener process* if the cross-variation satisfies

$$\mathrm{d}\langle W,\bar W\rangle_t = \rho(t,W,\bar W)\mathrm{d}t;$$

we say that it is a correlated Wiener process if it is a ρ-correlated Wiener process for some ρ.

We now introduce the subset of couplings between elements of $\mathcal{P}^*$ that Bion-Nadal and Talay study in [11]. Given $\mu^{b,\sigma},\mu^{\bar b,\bar\sigma}\in\mathcal{P}^*$, consider the set of couplings of the form $\pi = \mathrm{Law}(X,\bar X)$, where there exists some correlated Wiener process $(W,\bar W)$ such that $(X,\bar X)$ is $\mathcal{G}$-adapted and satisfies

$$\begin{cases} X_t = x_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s, & t\in[0,1],\\ \bar X_t = x_0 + \int_0^t \bar b(\bar X_s)\mathrm{d}s + \int_0^t \bar\sigma(\bar X_s)\mathrm{d}\bar W_s, & t\in[0,1]. \end{cases} \tag{2.1}$$

In this case, we say that $(X,\bar X)$ is a *strong solution of the system* (2.1) with respect to the correlated Wiener process $(W,\bar W)$.

For any $p\ge 1$, one can define a metric $\bar{\mathcal{W}}_p$ on $\mathcal{P}^*\cap\mathcal{P}_p$ in the same way as the adapted Wasserstein distance $\mathcal{AW}_p$ is defined in (1.1), but with the set $\mathrm{Cpl_{bc}}(\mu,\nu)$ of bi-causal couplings between measures $\mu,\nu\in\mathcal{P}^*$ replaced by the set of joint laws of strong solutions of (2.1) driven by some correlated Wiener process. For $p=2$, $\bar{\mathcal{W}}_2$ is the modified Wasserstein distance introduced by Bion-Nadal and Talay in [11].

We claim that, for any $p\ge1$, the metric $\bar{\mathcal{W}}_p$ coincides with the adapted Wasserstein distance $\mathcal{AW}_p$. Indeed, the next result shows that, for $\mu,\nu\in\mathcal{P}^*$, the set of joint laws of strong solutions of (2.1) driven by some correlated Wiener process is equal to the set of bi-causal couplings $\mathrm{Cpl_{bc}}(\mu,\nu)$.

Proposition 2.2. *Let $\mu^{b,\sigma},\mu^{\bar b,\bar\sigma}\in\mathcal{P}^*$. Then the set of bi-causal couplings $\mathrm{Cpl_{bc}}(\mu^{b,\sigma},\mu^{\bar b,\bar\sigma})$ is equal to the set of joint laws of strong solutions of the system* (2.1) *driven by some correlated Wiener* *process. In particular, for any $p \geq 1$, if $\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}} \in \mathcal{P}_p$, then*

$$\bar{\mathcal{W}}_p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}}) = \mathcal{AW}_p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}}).$$

Proof. First suppose that $\tilde{\pi}$ is the joint law of the strong solution of the system of SDEs (2.1) driven by some correlated Wiener process. Then $\tilde{\pi}$ is a coupling between $\mu^{b,\sigma}$ and $\mu^{\bar{b},\bar{\sigma}}$. Since $(X, \bar{X})$ is adapted to the completed natural filtration $\mathcal{G}$ of $(W, \bar{W})$, we also have that $\tilde{\pi}$ is bi-causal.

To show the converse, suppose now that $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$. Let $(\omega, \bar{\omega})$ be the canonical process on the product space, and let $\mathcal{F}, \bar{\mathcal{F}}$ be the natural filtrations of $\omega, \bar{\omega}$, respectively. Since the coupling π is bi-causal, we have the following independence property: for any $t > 0$, conditional on $\mathcal{F}_t$, $(\omega_s)_{s \in (t,1]}$ is independent of $\bar{\mathcal{F}}_t$ under π. Therefore, for any $t \in (0, 1]$, we have

$$\mathbb{E}^\pi \left[\omega_{t+h} - \omega_t - \int_t^{t+h} b(\omega_u) \mathrm{d}u | \mathcal{F}_t \otimes \bar{\mathcal{F}}_t \right] = \mathbb{E}^\pi \left[\omega_{t+h} - \omega_t - \int_t^{t+h} b(\omega_u) \mathrm{d}u | \mathcal{F}_t \right] = 0,$$

where the latter equality is a consequence of

$$\begin{aligned} \mathbb{E}^\pi &\left[\left(\omega_{t+h} - X_t - \int_t^{t+h} b(\omega_u) \mathrm{d}u \right) (H((\omega_s)_{s \leq t})) \right] \\ &= \mathbb{E} \left[\left(X^{b,\sigma}_{t+h} - X^{b,\sigma}_t - \int_t^{t+h} b(X^{b,\sigma}_u) \mathrm{d}u \right) \left(H((X^{b,\sigma}_s)_{s \leq t}) \right) \right] \\ &= \mathbb{E} \left[\left(\int_t^{t+h} \sigma(X^{b,\sigma}_u) \mathrm{d}B_u \right) \left(H((X^{b,\sigma}_s)_{s \leq t}) \right) \right] = 0, \end{aligned}$$

for $X^{b,\sigma}$ the strong solution of (1.2) driven by some Wiener process B, and any bounded measurable function $H : C([0, t], \mathbb{R}) \to \mathbb{R}$.

We deduce that $\omega_t = x_0 + \int_0^t b(\omega_s) \mathrm{d}s + M_t$ under π, where M is a martingale with respect to $\mathcal{F} \otimes \bar{\mathcal{F}}$, and a fortiori M must be continuous. Similarly $\bar{\omega}_t = x_0 + \int_0^t \bar{b}(\bar{\omega}_s) \mathrm{d}s + \bar{M}_t$ under π, for $\bar{M}$ a continuous martingale with respect to $\mathcal{F} \otimes \bar{\mathcal{F}}$.

Since the quadratic variation of $X^{b,\sigma}$ coincides with that of the canonical process ω under π, we deduce from the definition of M that $\langle M \rangle_t = \int_0^t \sigma(\omega_t)^2 \mathrm{d}t$. Similarly, $\langle \bar{M} \rangle_t = \int_0^t \bar{\sigma}(\bar{\omega}_t)^2 \mathrm{d}t$. Hence $t \mapsto \int_0^t 1_{\sigma(\omega_t)=0} \mathrm{d}M_t$ is the constant (null) martingale and $M_t = \int_0^t 1_{\sigma(\omega_t) \neq 0} \mathrm{d}M_t$, with a similar statement for $\bar{M}$. After enlarging the probability space, we introduce $\hat{W}$ a standard Wiener process independent of all pre-existing random variables, and then define processes $W, \bar{W}$ via

$$\mathrm{d}W_t = 1_{\sigma(\omega_t)=0} \mathrm{d}\hat{W}_t + 1_{\sigma(\omega_t) \neq 0} \frac{\mathrm{d}M_t}{\sigma(\omega_t)}, \quad \mathrm{d}\bar{W}_t = 1_{\bar{\sigma}(\bar{\omega}_t)=0} \mathrm{d}\hat{W}_t + 1_{\bar{\sigma}(\bar{\omega}_t) \neq 0} \frac{\mathrm{d}\bar{M}_t}{\bar{\sigma}(\bar{\omega}_t)}.$$

Then $(\omega, \bar{\omega})$ satisfies (2.1). Further, if we denote by $\mathcal{G}$ the completion of the natural filtration of $(W, \bar{W})$, then both W and $\bar{W}$ are $\mathcal{G}$-martingales under π, which follows by bi-causality and the independence of $\hat{W}$. Hence both $W, \bar{W}$ are $\mathcal{G}$-Wiener processes, by Lévy's characterisation. By the Kunita-Watanabe inequality, we have that $\mathrm{d}\langle W, \bar{W} \rangle$ is a.s. absolutely continuous with respect to Lebesgue measure, and so there exists a $\mathcal{B}([0, 1]) \otimes \mathcal{G}$-progressively measurable function $\rho : [0, 1] \times \Omega \times \Omega \to [-1, 1]$ such that $\rho_t \mathrm{d}t = \mathrm{d}\langle W, \bar{W} \rangle_t$. Therefore $(W, \bar{W})$ is a ρ-correlated Wiener process, as in Definition 2.1.

By assumption on b, σ, the SDE with these coefficients driven by the Wiener process W admits a unique strong solution $\tilde{X}$. Letting $\mathcal{H}$ be the natural filtration of $\hat{W}$, we have that both $\tilde{X}$ and ω are adapted to $\mathcal{F} \otimes \bar{\mathcal{F}} \otimes \mathcal{H}$. By pathwise uniqueness, we deduce that $\tilde{X} = \omega$ and, in particular, ω is adapted to the filtration of W and hence to $\mathcal{G}$. The analogous statement applies to $\bar{\omega}$, and so we have that $(\omega, \bar{\omega})$ is a strong solution of the system (2.1) with respect to $(W, \bar{W})$, as required. □

Remark 2.3. In our definition of (bi-)causal couplings we make use of the canonical filtration, while in [4] the right-continuous filtration is used. In general, our definition gives a smaller set of couplings. However, when the marginal processes are strong Markov both sets coincide following an application of [20, Ch. 2, Proposition 7.7]. A sufficient condition for this to hold is that the coefficients of the SDEs are locally bounded and that the associated martingale problems are well-posed for all initial conditions; see [20, Ch. 5, Theorem 4.20]. In Example 6.1, we also encounter a situation where the above sets of couplings coincide, although one of the marginal processes is not strong Markov.

We next make use of the above result to prove a more general version of Theorem 1.4 on the approximation of the laws of SDEs in adapted Wasserstein distance. For $b:\mathbb{R}\to\mathbb{R}$ and $\sigma:\mathbb{R}\to\mathbb{R}_+$, consider the SDE (1.2). Also, for $h>0$, $b^h:[0,1]\times C([0,1],\mathbb{R})\to\mathbb{R}$, and $\sigma^h:[0,1]\times C([0,1],\mathbb{R})\to\mathbb{R}_+$ progressively measurable, consider the SDE

$$\mathrm{d}X_t^h = b^h(t,X^h)\mathrm{d}t+\sigma^h(t,X^h)\mathrm{d}W_t,\quad X_0^h=x_0,\quad t\in[0,1]. \tag{2.2}$$

Assumption 2.4. Given $p\geq 1$, suppose that, for $b:\mathbb{R}\to\mathbb{R}$ and $\sigma:\mathbb{R}\to\mathbb{R}_+$, the SDE (1.2) has a p-integrable unique strong solution and that, for $b^h:[0,1]\times C([0,1],\mathbb{R})\to\mathbb{R}$ and $\sigma^h:[0,1]\times C([0,1],\mathbb{R})\to\mathbb{R}_+$, the SDE (2.2) has a p-integrable unique strong solution, for all $h>0$. Moreover, suppose that the solution of (2.2) converges to the solution of (1.2) in L^p.

Remark 2.5. Under Lipschitz conditions on the coefficients of the SDE (1.2), the Euler–Maruyama scheme (1.3) converges to the unique solution of (1.2) in L^p, for all $p\geq 1$ (see, e.g. [21]), and so the below theorem will imply the result of Theorem 1.4. Under the same conditions, the monotone Euler–Maruyama scheme that we develop below in Definition 3.11 also converges to the solution of (1.2) in L^p, for all $p\geq 1$.

For $p\geq 1$, and $\alpha,\beta\in\mathcal{P}(\Omega\times\Omega)$ with finite p^{th} moment, we define the p-adapted Wasserstein distance between α and β, analogously to (1.1), by

$$\mathcal{AW}_p^p(\alpha,\beta):=\inf_{\pi\in\mathrm{Cpl}_{\mathrm{bc}}(\alpha,\beta)}\mathbb{E}^\pi\left[\int_0^1|\omega_t-\bar{\omega}_t|^p\mathrm{d}t\right], \tag{2.3}$$

where $\mathrm{Cpl}_{\mathrm{bc}}(\alpha,\beta)\subset\mathcal{P}(\Omega\times\Omega\times\Omega\times\Omega)$ is the set of bi-causal couplings between α and β, defined analogously to Definition 1.1, and $(\omega,\bar{\omega})$ now denotes the canonical process on $\Omega\times\Omega\times\Omega\times\Omega$.

Theorem 2.6. *Suppose that for some $p\geq 1$, $(b,\sigma,(b^h)_{h>0},(\sigma^h)_{h>0})$ and $(\bar{b},\bar{\sigma},(\bar{b}^h)_{h>0},(\bar{\sigma}^h)_{h>0})$ satisfy Assumption 2.4. Then, for any bi-causal coupling $\pi\in\mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma},\mu^{\bar{b},\bar{\sigma}})$, there exists a probability space supporting a correlated Wiener process $(W,\bar{W})$ such that*

$$\lim_{h\to 0}\mathcal{AW}_p(\mathrm{Law}(X^h,\bar{X}^h),\pi)=0,$$

where X^h (resp. $\bar{X}^h$) solves (2.2) driven by W (resp. $\bar{W}$) with coefficients (b^h,σ^h) (resp. $(\bar{b}^h,\bar{\sigma}^h)$), for $h>0$.

Proof. Take $\pi\in\mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma},\mu^{\bar{b},\bar{\sigma}})$. By Proposition 2.2, there exists a correlated Wiener process $(W,\bar{W})$ such that $\pi=\mathrm{Law}(X,\bar{X})$, where the SDE for X (resp. $\bar{X}$) is driven by W (resp. $\bar{W}$). Then $\mathrm{Law}((X^h,\bar{X}^h),(X,\bar{X}))\in\mathrm{Cpl}_{\mathrm{bc}}(\mathrm{Law}(X^h,\bar{X}^h),\pi)$. Indeed, for $t\in[0,1]$, and any bounded $f:\Omega\times\Omega\to\mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}\left[f((X_s^h,\bar{X}_s^h)_{s\in[0,t]})|\,\mathcal{F}_1^X\vee\mathcal{F}_1^{\bar{X}}\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[f((X_s^h,\bar{X}_s^h)_{s\in[0,t]})|\,\mathcal{F}_t^X\vee\mathcal{F}_t^{\bar{X}}\vee\sigma\{(W_u-W_t,\bar{W}_u-\bar{W}_t):u\in(t,1]\}\right]|\,\mathcal{F}_1^X\vee\mathcal{F}_1^{\bar{X}}\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[f((X_s^h,\bar{X}_s^h)_{s\in[0,t]})|\,\mathcal{F}_t^X\vee\mathcal{F}_t^{\bar{X}}\right]|\,\mathcal{F}_1^X\vee\mathcal{F}_1^{\bar{X}}\right]\\
&\quad=\mathbb{E}\left[f((X_s^h,\bar{X}_s^h)_{s\in[0,t]})|\,\mathcal{F}_t^X\vee\mathcal{F}_t^{\bar{X}}\right],
\end{aligned}$$

where the second equality follows from the fact that $\sigma\{(W_u - W_t, \bar{W}_u - \bar{W}_t) : u \in (t,1]\}$ is independent of $(X^h_s, \bar{X}^h_s)_{s\in[0,t]}$ and $\mathcal{F}^X_t \vee \mathcal{F}^{\bar{X}}_t$. This implies causality in one direction. Since the roles of $(X^h, \bar{X}^h)$ and $(X, \bar{X})$ are symmetric in this calculation, we have bi-causality.

Hence, by convergence in L^p, we have

$$\lim_{h\to 0} \mathcal{AW}^p_p(\mathrm{Law}(X^h, \bar{X}^h), \pi) \le \lim_{h\to 0} \mathbb{E}\left[\int_0^1 |X_t - X^h_t|^p \mathrm{d}t\right] + \lim_{h\to 0} \mathbb{E}\left[\int_0^1 |\bar{X}_t - \bar{X}^h_t|^p \mathrm{d}t\right] = 0.$$

□

3. The synchronous coupling: Properties and optimality

The main aim of this section is to prove Theorem 1.3 on the optimality of the synchronous coupling for the adapted Wasserstein distance between laws of SDEs under the additional assumption that the coefficients are Lipschitz. We also establish the result for a particular class of coefficients which allows for discontinuities in the drift coefficient.

For functions $b, \bar{b} : \mathbb{R} \to \mathbb{R}$, $\sigma, \bar{\sigma} : \mathbb{R} \to \mathbb{R}_+$, we consider the SDEs

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t; \quad X_0 = x_0, \\ \mathrm{d}\bar{X}_t &= \bar{b}(\bar{X}_t)\mathrm{d}t + \bar{\sigma}(\bar{X}_t)\mathrm{d}\bar{W}_t; \quad \bar{X}_0 = x_0, \end{aligned} \tag{3.1}$$

for $t \in [0,1]$, under sufficient conditions to guarantee existence and uniqueness of strong solutions $X, \bar{X}$. We denote by μ and ν the laws of X and $\bar{X}$, respectively.

We define the *synchronous coupling* $\pi^{\mathrm{sync}}_{\mu,\nu} \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)$ as follows.[3] Set $\bar{W} = W$ and let $(X^{\mathrm{sync}}, \bar{X}^{\mathrm{sync}})$ be the solutions of the SDEs (3.1) driven by the common Wiener process W. Then the synchronous coupling is defined as the joint law $\pi^{\mathrm{sync}}_{\mu,\nu} := \mathrm{Law}(X^{\mathrm{sync}}, \bar{X}^{\mathrm{sync}})$.

In [11], Bion-Nadal and Talay show that, under certain smoothness conditions, the synchronous coupling is optimal for the adapted Wasserstein distance between the laws of solutions of (3.1). In the following, we provide an alternative proof based on discrete-time approximations, which allows us to significantly extend the result of [11].

3.1. The Knothe–Rosenblatt rearrangement.

The Knothe–Rosenblatt rearrangement (also known as the Knothe–Rosenblatt coupling or quantile transformation) was introduced independently by Rosenblatt [30] and Knothe [22], and can be seen as a multidimensional extension of the monotone rearrangement. We illustrate this coupling in Figure 1. We define the Knothe–Rosenblatt rearrangement in terms of conditional cumulative distribution functions, as in [7].

Definition 3.1. Let μ be a probability measure on $\mathbb{R}^n$, $n \in \mathbb{N}$. For $x \in \mathbb{R}^k$, $k < n$, we define the *conditional cumulative distribution function* $F_{\mu_{x_1,\dots,x_k}} : \mathbb{R} \to [0,1]$ by

$$F_{\mu_{x_1,\dots,x_k}}(a) := \mu_{x_1,\dots,x_k}(-\infty, a], \quad a \in \mathbb{R},$$

where $\mu_{x_1,\dots,x_k}$ is the one-dimensional conditional distribution in the $(k+1)^{\mathrm{th}}$ coordinate under μ, given the first k coordinates $x_1, \dots, x_k$. Similarly, define $F_{\mu_1}(a) = \mu_1(-\infty, a]$, $a \in \mathbb{R}$, where μ_1 is the first marginal of μ.

For any cumulative distribution function F, we write F^{-1} for its left-continuous inverse, the *quantile function*, defined by $F^{-1}(y) = \inf\{u \in \mathbb{R} \colon F(u) \ge y\}$.

Definition 3.2 (Knothe–Rosenblatt rearrangement)**.** Given probability measures μ, ν on $\mathbb{R}^n$, let $U_1, \dots, U_n$ be independent uniform random variables on $[0,1]$. Define $X_1 = F^{-1}_{\mu_1}(U_1)$, $Y_1 = F^{-1}_{\nu_1}(U_1)$ and, for $k = 2, \dots, n$, define inductively the random variables

$$X_k = F^{-1}_{\mu_{X_1,\dots,X_{k-1}}}(U_k), \quad Y_k = F^{-1}_{\nu_{Y_1,\dots,Y_{k-1}}}(U_k).$$

[3] If one drops the assumption that $\sigma, \bar{\sigma}$ are both positive, one could recover the results of this paper by redefining the synchronous coupling to be the one induced by $(W, \bar{W})$ with correlation $\mathrm{sign}(\sigma\bar{\sigma})$.

Then the Knothe–Rosenblatt rearrangement between the marginals μ and ν is given by $\pi^{\mathrm{KR}}_{\mu,\nu} := \mathrm{Law}(X_1, \dots, X_n, Y_1, \dots, Y_n)$.

If μ is absolutely continuous with respect to the Lebesgue measure, then $\pi^{\mathrm{KR}}_{\mu,\nu}$ is induced by the Monge map $(x_1, \dots, x_n) \mapsto T(x_1, \dots, x_n) = (T^1(x_1), T^2(x_2; x_1), \dots, T^n(x_n; x_1 \dots, x_{n-1}))$ given by $T^1(x_1) = F^{-1}_{\nu_1} \circ F_{\mu_1}(x_1)$ and, for $k = 2, \dots, n$,

$$T^k(x_k; x_1, \dots, x_{k-1}) = F^{-1}_{\nu_{T^1(x_1),\dots,T^{k-1}(x_{k-1};x_1,\dots,x_{k-2})}} \circ F_{\mu_{x_1,\dots,x_{k-1}}}(x_k).$$

In the following definition and throughout the paper, we use the terms increasing and decreasing in a weak sense; we do not require strict monotonicity.

Definition 3.3 (co-monotonicity)**.** Two functions $f, g : \mathbb{R} \to \mathbb{R}$ are called *co-monotone* in each of the following three cases: f and g are both increasing, f and g are both decreasing, or one of f and g is constant and the other is arbitrary.

Remark 3.4. The following condition is equivalent to the condition in Definition 3.2. For independent uniform random variables $U_1, \dots, U_n$ on $[0, 1]$, we have the representation

$$X = (T_1(U_1), T_2(U_2; X_1), \dots, T_n(U_n; X_1, \dots, X_{n-1})),$$
$$Y = (S_1(U_1), S_2(U_2; Y_1), \dots, S_n(U_n; Y_1, \dots Y_{n-1})),$$

where the functions T_i, S_i are co-monotone in their first argument, for all $i = 1, \dots, n$.

The following result from [7, Proposition 5.3], which generalises [31, Corollary 2], gives monotonicity conditions on the marginals μ, ν that guarantee optimality of the Knothe–Rosenblatt rearrangement for an associated discrete-time bi-causal transport problem. This result forms the basis for our proof of optimality of the synchronous coupling in continuous time.

Proposition 3.5 (optimality of Knothe–Rosenblatt [7, 31])**.** *For probability measures μ and ν on $\mathbb{R}^n$, define $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ analogously to Definition 1.1. Suppose that the functions $x_k \mapsto F_{\mu_{x_1,\dots,x_{k-1},x_k}}(u)$ and $y_k \mapsto F_{\nu_{y_1,\dots,y_{k-1},y_k}}(u)$ are co-monotone for each $u \in \mathbb{R}$, $k = 1, \dots, n-1$, and $(x_1, \dots, x_{k-1}), (y_1, \dots, y_{k-1}) \in \mathbb{R}^{k-1}$. Then the Knothe–Rosenblatt rearrangement between μ and ν is optimal for the bi-causal transport problem*

$$\inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \int c \mathrm{d}\pi,$$

for any c of the form

$$c(x_1, \dots, x_n, y_1, \dots, y_n) = \sum_{k=1}^{n} c_k(x_k - y_k),$$

where each c_k is convex and finite.

Remark 3.6. A well-studied example of a cost function c_k is given by $c_k(x_k, y_k) = |x_k - y_k|^p$, for some $p \geq 1$, $x_k, y_k \in \mathbb{R}$, $k = 1, \dots, n$. In this case, the infimum in Proposition 3.5 is the *discrete-time adapted Wasserstein distance*

$$\mathcal{AW}^p_p(\mu, \nu) := \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \int \sum_{k=1}^{n} |x_k - y_k|^p \, \pi(\mathrm{d}x, \mathrm{d}y),$$

for μ, ν measures on $\mathbb{R}^n$ with finite p^{th} moment. In this paper we will apply Proposition 3.5 to marginals arising from a numerical scheme for SDEs in order to deduce that the Knothe–Rosenblatt rearrangement attains the associated discrete-time adapted Wasserstein distance.

Remark 3.7. The proof of Proposition 3.5 can be seen as a discrete-time analogue of Bion-Nadal and Talay's proof of optimality of the synchronous coupling in [11]. Indeed, both proofs are based on verification arguments; given sufficient regularity conditions, the key condition for the continuous-time case in [11] is that the second order cross-derivative of the value function is negative, while the analogous algebraic condition is used in [7] to prove Proposition 3.5.

Definition 3.8 (first order stochastic dominance)**.** For probability measures μ, ν on $\mathbb{R}$, we say that ν *dominates* μ *in first order* if $\mu(-\infty, a] \geq \nu(-\infty, a]$, for every $a \in \mathbb{R}$.

For a probability measure μ on $\mathbb{R}^n$, we say that the measure μ (or a stochastic process with law μ) is *increasing in first order stochastic dominance* if, for all $(x_1, \dots, x_{k-1}) \in \mathbb{R}^{k-1}$, $k = 1, \dots, n-1$, $\mu_{x_1,\dots,x_{k-1},x'}$ dominates $\mu_{x_1,\dots,x_{k-1},x}$ in first order, for any $x < x'$, or equivalently, if the map $x_k \mapsto F_{\mu_{x_1,\dots,x_{k-1},x_k}}(u)$ is decreasing for all $u \in \mathbb{R}$.

Remark 3.9. Note that, if μ and ν are each increasing in first order stochastic dominance, then the conditions of Proposition 3.5 are satisfied. For example, the finite dimensional distributions of any continuous strong Markov process are increasing in first order stochastic dominance, as was shown in [10, Proposition 5.2] by use of a coupling argument which originates from [17].

In this paper, we will be interested in time-discretisation schemes for SDEs that satisfy this monotonicity condition. For example, for SDEs with constant volatility and Lipschitz drift, the classical discrete-time Euler–Maruyama scheme is increasing in first order stochastic dominance, when the step-size is small enough. The discrete-time monotone Euler–Maruyama scheme that we develop below is increasing in first order stochastic dominance for a more general class of SDEs.

3.2. **A monotone numerical scheme.** In this section, we will show that the continuous-time adapted Wasserstein distance between laws of solutions of (3.1) with Lipschitz coefficients is the limit of discrete-time Wasserstein distances, each of which are attained by the Knothe–Rosenblatt rearrangement. As a consequence, we will deduce that the synchronous coupling attains the continuous-time adapted Wasserstein distance. For this reason, we argue that we can view the synchronous coupling as a continuous-time analogue of the Knothe–Rosenblatt rearrangement.

We first seek a discrete-time bi-causal transport problem, whose value converges to that of the continuous-time problem. In order to achieve this, we discretise the SDEs (3.1) via a numerical scheme. In light of Proposition 3.5, we seek a numerical scheme which is increasing in first order stochastic dominance. We make the following observation.

Remark 3.10. Recall the Euler–Maruyama scheme for (1.2), as defined in (1.3). Fix $N \in \mathbb{N}$, let $h = \frac{1}{N}$, set $X^h_0 = x_0$, and take W to be a Wiener process. Then, for each $k = 0, \dots, N-1$ and $t \in (kh, (k+1)h]$, define

$$X^h_t = X^h_{kh} + (t - kh)b(X^h_{kh}) + \sigma(X^h_{kh})(W_t - W_{kh}). \tag{3.2}$$

We call $(X^h_t)_{t\in[0,1]}$ the *Euler–Maruyama scheme* for X, and $(X^h_k)_{k=0,\dots,N}$ the *discrete-time Euler–Maruyama scheme* for X.

Now suppose that b is Lipschitz, with Lipschitz constant C, and that σ is *constant*. Then, for $h < C^{-1}$, the process $(X^h_k)_{k=0,\dots,N}$ is increasing in first order stochastic dominance. Indeed, since the function $x \mapsto x + hb(x)$ is then increasing, we have

$$\mathbb{P}\left[x + hb(x) + \sigma(W_{(k+1)h} - W_{kh}) \leq a\right] \geq \mathbb{P}\left[x' + hb(x') + \sigma(W_{(k+1)h} - W_{kh}) \leq a\right],$$

for $x < x'$, $a \in \mathbb{R}$.

When we take a non-constant diffusion coefficient σ, the above discrete-time Euler–Maruyama scheme is no longer increasing in first order stochastic dominance. We are therefore led to introduce a new numerical scheme where we truncate the Brownian increments in the Euler–Maruyama scheme, as first proposed for the Milstein scheme in [25]. While favourable convergence properties of this Milstein scheme are proved in [25], the required regularity on the coefficients of the SDE is stronger than we would like; for example, the first derivative of the drift is required. The modified scheme that we introduce will enable us to relax those regularity assumptions so as to allow for general Lipschitz coefficients. In [25], the authors also consider another numerical scheme, namely a fully implicit Euler scheme with a truncated Brownian increment. They apply this scheme to a geometric Brownian motion, but they do not give any general convergence results. In the present paper, we rather apply a similar truncation method to the explicit Euler–Maruyama scheme (3.2). For the scheme that we construct, we also seek a continuous interpolation that can be readily associated to the choice of the Wiener process driving the SDE. All in all, this motivates us to consider the following truncated scheme.

Let W be a standard one-dimensional Wiener process. For $N \in \mathbb{N}$, let $h = \frac{1}{N}$ and, following [25], fix the truncation level to be $A_h := 4\sqrt{-h \log h}$. Define the stopping time τ_0^h to be the first time that W leaves the interval $(-A_h, A_h)$ and define $W_t^h := W_{t \wedge \tau_0^h}$, for $t \in [0, h]$. Then, in turn, for each $k = 1, \ldots, N-1$, define the stopping time τ_k^h to the be first time after kh that the Brownian increment $(W_t - W_{kh})$ leaves the interval $(-A_h, A_h)$ and define $W_t^h := W_{kh}^h + (W_\cdot - W_{kh})_{t \wedge \tau_k^h}$, for $t \in (kh, (k+1)h]$.

Definition 3.11 (monotone Euler–Maruyama scheme)**.** Consider the SDE (1.2) with Lipschitz coefficients $b : \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \to \mathbb{R}_+$. Fix $N \in \mathbb{N}$, $h = \frac{1}{N}$, let $X_0^h = x_0$ and, for each $k = 0, \ldots, N-1$ and $t \in (kh, (k+1)h]$, define

$$X_t^h := X_{kh}^h + (t - kh) b(X_{kh}^h) + \sigma(X_{kh}^h)(W_t^h - W_{kh}^h).$$

We call the process $(X_t^h)_{t \in [0,1]}$ the *monotone Euler–Maruyama scheme* for X.

We will also refer to the *discrete-time monotone Euler–Maruyama scheme*, which we define as the process $(X_{kh}^h)_{k=0,\ldots,N}$.

Remark 3.12. Note that the process $(W_t^h)_{t \in [0,1]}$ is a martingale with respect to $\mathcal{F}^W$, the filtration generated by the Wiener process W satisfying the usual conditions.

The following lemma, which we adapt from [25], gives a bound on the fourth moment of the error created by the truncation and thus justifies the choice of the truncation level A_h.

Lemma 3.13. *For $N \in \mathbb{N}$, $h = 1/N$, and fixed $k \in \{1, \ldots, N\}$, we have the fourth moment bound*

$$\mathbb{E}\left|W_{(k+1)h} - W_{kh} - (W_{(k+1)h}^h - W_{kh}^h)\right|^4 \le C h^{10}.$$

Proof. First note that

$$\begin{aligned}\mathbb{E}\left|W_{(k+1)h} - W_{kh} - (W_{(k+1)h}^h - W_{kh}^h)\right|^4 &= \mathbb{E}\left|W_{(k+1)h} - W_{(k+1)h \wedge \tau_k^h}\right|^4 \\ &= 3\mathbb{E}|h - h \wedge \tau_0^h|^4 \le 3h^2 \mathbb{P}[\tau_0^h \le h],\end{aligned}$$

since W has identically distributed increments, and calculate

$$\mathbb{P}[\tau_0^h \le h] = 2\mathbb{P}\left[\sup_{t \in [0,h]} W_t \ge A_h\right] = 4\mathbb{P}[W_h \ge A_h],$$

using the reflection principle. Then

$$\mathbb{E}\left|W_{(k+1)h} - W_{kh} - (W_{(k+1)h}^h - W_{kh}^h)\right|^4 \le \frac{12h^2}{\sqrt{2\pi h}} \int_0^\infty e^{-\frac{(x+A_h)^2}{2h}} \mathrm{d}x < 6h^2 e^{-\frac{A_h^2}{2h}}.$$

Recalling the definition $A_h = 4\sqrt{-h \log h}$, we conclude. □

Remark 3.14. From the above proof, we see that, for an arbitrary $K \in \mathbb{N}$, we can redefine $A_h := K\sqrt{-h \log h}$ and achieve a fourth moment bound of $Ch^{2+\frac{K^2}{2}}$ in Lemma 3.13.

We now verify that, for SDEs with Lipschitz coefficients, the monotone Euler–Maruyama scheme exhibits certain key properties. First we show that, for sufficiently small $h > 0$, the discrete-time monotone Euler–Maruyama scheme is increasing in first order stochastic dominance. Then we show that, for two SDEs driven by a common Wiener process, the joint law of the discrete-time monotone Euler–Maruyama schemes coincides with the Knothe–Rosenblatt rearrangement between the laws of the two schemes.

Lemma 3.15. *Suppose that the coefficients b and σ in* (1.2) *are Lipschitz. Then, for sufficiently small $h > 0$, the discrete-time monotone Euler–Maruyama scheme $(X^h_{kh})_{k=0,\dots,N}$ for* (1.2) *is increasing in first order stochastic dominance.*

Proof. Let $x, x' \in \mathbb{R}$ such that $x < x'$, and define the random variables $Y = x + hb(x) + \sigma(x)(W^h_{(k+1)h} - W^h_{kh})$, $Y' = x' + hb(x') + \sigma(x')(W^h_{(k+1)h} - W^h_{kh})$. Then, letting C_0 and C_1 be the Lipschitz constants of b and σ, respectively, and using the bound on the truncated Brownian increment, we have

$$Y' - Y \geq (1 - hC_0 - A_h C_1)(x' - x).$$

Noting that $\lim_{h \to 0} A_h = 0$, we can choose h sufficiently small that $1 - hC_0 - A_h C_1 > 0$, and conclude that we have the desired ordering in first order stochastic dominance. □

In light of Proposition 3.5, Lemma 3.15 implies that, for the adapted Wasserstein distance between the laws of two discrete-time monotone Euler–Maruyama schemes, the Knothe–Rosenblatt rearrangement is an optimiser, when all coefficients are Lipschitz.

Lemma 3.16. *Fix a Wiener process W on some probability space and let $X, \bar{X}$ be the unique strong solutions of the SDEs* (3.1) *driven by the common Wiener process W — i.e. $\bar{W} = W$ in* (3.1)*. For $h > 0$, let $X^h, \bar{X}^h$ be the associated discrete-time monotone Euler–Maruyama schemes, and write μ^h, ν^h for the laws of $X^h, \bar{X}^h$, respectively. Then the joint law* $\mathrm{Law}(X^h, \bar{X}^h)$ *is equal to the Knothe–Rosenblatt rearrangement $\pi^{\mathrm{KR}}_{\mu^h,\nu^h}$ between μ^h and ν^h.*

Proof. Let $N \in \mathbb{N}$ and $h = \frac{1}{N}$. Fix $k \in \{0, \dots, N-1\}$ and write $\Delta_k W^h$ for the truncated increment $W^h_{(k+1)h} - W^h_{kh}$ of the Wiener process. Then, since $\sigma, \bar{\sigma}$ are non-negative functions, the maps $\Delta_k W^h \mapsto X^h_{(k+1)h}$, given by $X^h_{(k+1)h} = X^h_{kh} + hb(X^h_{kh}) + \sigma(X^h_{kh})\Delta_k W^h$, and $\Delta_k W^h \mapsto \bar{X}^h_{(k+1)h}$, given by $\bar{X}^h_{(k+1)h} = \bar{X}^h_{kh} + h\bar{b}(\bar{X}^h_{kh}) + \bar{\sigma}(\bar{X}^h_{kh})\Delta_k W^h$, are both increasing. Since we take the same random variables $\Delta_k W^h$, $k = 0, \dots, N-1$, for both X^h and $\bar{X}^h$, the characterisation of the Knothe–Rosenblatt rearrangement given in Remark 3.4 implies that the joint law of X^h and $\bar{X}^h$ coincides with the Knothe–Rosenblatt rearrangement between μ^h and ν^h. □

The remainder of this section is devoted to establishing the following result which verifies that the scheme converges in the L^p-norm to the solution of the SDE.

Lemma 3.17. *Suppose that the coefficients b and σ in* (1.2) *are Lipschitz. Then, for the monotone Euler–Maruyama scheme X^h given in Definition 3.11, we have for any $p \geq 1$, the L^p-convergence*

$$\lim_{h \to 0} \mathbb{E}\left[\int_0^1 |X^h_t - X_t|^p \mathrm{d}t\right] = 0.$$

The proof of this result proceeds as follows. We first establish L^2-convergence by showing that the monotone Euler–Maruyama scheme is close in the L^2-norm to the standard Euler–Maruyama scheme. We then deduce L^p-convergence from the L^2-convergence and a bound on the p^{th} moments of the monotone scheme. Throughout the following proofs, we make use of generic constants, which may change from one line to the next.

Remark 3.18. Under the assumption that the coefficients b, σ in (1.2) are Lipschitz, there exists a unique strong solution X to (1.2) according to a classical result of Itô (e.g. [12, Proposition 1.9]). Moreover, for $p \geq 1$, the process X satisfies the moment bound

$$\mathbb{E}\left[\sup_{0 \leq t \leq 1} |X_t|^p\right] \leq C_p, \quad t \in [0, 1], \tag{3.3}$$

for some constant $C_p > 0$. This follows from a standard application of Jensen's inequality, the BDG inequality and Grönwall's lemma (see e.g. [29, Theorem 3.15] and the discussion thereafter).

In order to prove the L^2-convergence of the monotone Euler–Maruyama scheme X^h to the unique strong solution X of (1.2), when the coefficients are Lipschitz, we first recall the following estimates for the standard Euler–Maruyama scheme $\tilde{X}^h$ defined in (3.2). First (from e.g. the proof of [21, Theorem 10.2.2]) there exists a constant $\tilde{C}_0$ such that, for any $h > 0$, we have the L^2 estimate

$$\mathbb{E}\left[\sup_{0\le s\le 1}|\tilde{X}^h_s - X_s|^2\right] \le \tilde{C}_0 h. \tag{3.4}$$

Similarly to Remark 3.18, one can also derive the following moment bound. For any $p \ge 1$, there exists a constant $\tilde{C}_p$ such that, for any $h > 0$,

$$\mathbb{E}\left[\sup_{0\le s\le 1}|\tilde{X}^h_s|^p\right] \le \tilde{C}_p. \tag{3.5}$$

Proposition 3.19. *Suppose that the coefficients b and σ in* (1.2) *are Lipschitz. Then there exists a constant $C > 0$ such that, for any $N \in \mathbb{N}$ and $h = \frac{1}{N}$,*

$$\mathbb{E}\left[\sup_{0\le s\le 1}|\tilde{X}^h_s - X^h_s|^2\right] \le Ch^3.$$

Proof. Fix $N \in \mathbb{N}$ and $h = \frac{1}{N}$. For $s \in [0,1]$, introduce the notation $t_{n_s} := \sup\{t \le s \colon t = kh,\ \text{for some } k = 0,\dots,N\}$ and, for $t \in [0,1]$, define the remainder terms

$$R_t := \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \left(b(\tilde{X}^h_{t_{n_r}}) - b(X^h_{t_{n_r}})\right)\mathrm{d}r\right|^2\right], \quad S_t := \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \left(\sigma(\tilde{X}^h_{t_{n_r}}) - \sigma(X^h_{t_{n_r}})\right)\mathrm{d}W^h_r\right|^2\right],$$

$$U_t := \mathbb{E}\left[\sup_{0\le s\le t}\left|\int_0^s \sigma(\tilde{X}^h_{t_{n_r}})\mathrm{d}W_r - \int_0^s \sigma(\tilde{X}^h_{t_{n_r}})\mathrm{d}W^h_r\right|^2\right],$$

so that

$$Z_t := \mathbb{E}\left[\sup_{0\le s\le t}|\tilde{X}^h_s - X^h_s|^2\right] \le C(R_t + S_t + U_t).$$

Fix $t \in [0,1]$. By Jensen's inequality, we can bound

$$R_t \le \int_0^t \mathbb{E}\left[\sup_{0\le s\le u}\left|b(\tilde{X}^h_{t_{n_s}}) - b(X^h_{t_{n_s}})\right|^2\right]\mathrm{d}u.$$

Then, using the Lipschitz property of b and expanding the set of times over which we take the supremum, we can find a constant C^R such that

$$R_t \le C^R \int_0^t \mathbb{E}\left[\sup_{0\le s\le u}|\tilde{X}^h_s - X^h_s|^2\right]\mathrm{d}u = C^R \int_0^t Z_u \mathrm{d}u.$$

We now consider the term S_t. Since W^h is a martingale with respect to the completion of the natural filtration of W and since $\mathrm{d}\langle W^h\rangle_t \le \mathrm{d}t$, by Doob's martingale inequality,

$$S_t \le 4\int_0^t \mathbb{E}\left[\sup_{0\le s\le u}|\sigma(\tilde{X}^h_{t_{n_s}}) - \sigma(X^h_{t_{n_s}})|^2\right]\mathrm{d}u.$$

In the same way as for R_t, we now use the Lipschitz property of σ to find a constant C^S such that $S_t \le C^S \int_0^t Z_u \mathrm{d}u$.

Finally, we bound the term U_t. For each $k = 0, \dots, N-1$, let us write $\Delta_k W = W_{(k+1)h} - W_{kh}$ and $\Delta_k W^h = W^h_{(k+1)h} - W^h_{kh}$. Then, applying Doob's inequality, we get

$$\begin{aligned} U_t &\le 4\mathbb{E}\left[\left|\sum_{k=0}^{n_t-1} \sigma(\tilde{X}^h_{kh})\left[\Delta_k W - \Delta_k W^h\right] + \sigma(\tilde{X}^h_{t_{n_t}})\left[W_t - W_{t_{n_t}} - (W^h_t - W^h_{t_{n_t}})\right]\right|^2\right] \\ &\le 4N \sum_{k=0}^{N-1} \mathbb{E}\left[\sigma(\tilde{X}^h_{kh})^2(\Delta_k W - \Delta_k W^h)^2\right], \end{aligned}$$

and, by the Cauchy–Schwarz inequality,

$$U_t \le 4N \sum_{k=0}^{N} \sqrt{\mathbb{E}[\sigma(\tilde{X}^h_{kh})^4]}\sqrt{\mathbb{E}\left[(\Delta_k W - \Delta_k W^h)^4\right]}.$$

Applying Lemma 3.13, we can bound each term $\mathbb{E}\left[(\Delta_k W - \Delta_k W^h)^4\right] \le Ch^{10}$. Using the Lipschitz property of σ and the L^p bound (3.5) for the Euler–Maruyama scheme, we can also bound $\mathbb{E}[\sigma(\tilde{X}^h_{kh})^4] \le C$, for any $k = 0, \dots, N$. Therefore we have

$$U_t \le \bar{C} N^2 h^5 = \bar{C} h^3.$$

Combining the bounds on R_t, S_t and U_t, and defining $C = C^R + C^S$, we can bound Z_t by

$$Z_t \le \bar{C} h^3 + C \int_0^t Z_t \mathrm{d}u,$$

and by Grönwall's inequality we conclude that $Z_t \le \tilde{C} h^3$, for some constant $\tilde{C} > 0$. □

Remark 3.20. Similarly to Remark 3.14, the power in the bound in Proposition 3.19 can be made arbitrarily large, by multiplying the truncation level A_h by a sufficiently large constant.

The following immediate corollary now gives a rate for the L^2-convergence of the monotone Euler–Maruyama scheme X^h to the solution X of the SDE (1.2).

Corollary 3.21. *Suppose that the coefficients b and σ in* (1.2) *are Lipschitz. Then there exists a constant $C > 0$ such that, for any $h > 0$ sufficiently small,*

$$\mathbb{E}\left[\sup_{0 \le s \le 1} |X^h_s - X_s|^2\right] \le Ch.$$

Proof. Combining the rate of L^2 convergence of the Euler–Maruyama scheme given in (3.4) with the estimate of the L^2-error between the Euler–Maruyama scheme and the monotone Euler–Maruyama scheme given in Proposition 3.19, we can conclude via a simple application of the triangle inequality that

$$\mathbb{E}\left[\sup_{0 \le s \le 1} |X_s - X^h_s|^2\right] \le 2\mathbb{E}\left[\sup_{0 \le s \le 1} |\tilde{X}^h_s - X_s|^2\right] + 2\mathbb{E}\left[\sup_{0 \le s \le 1} |\tilde{X}^h_s - X^h_s|^2\right] \le Ch.$$

□

In order to obtain L^p-convergence, we make use of the following bounds on the p^{th} moments of the monotone Euler–Maruyama scheme X^h, for $h > 0$.

Lemma 3.22. *Suppose that the coefficients b and σ in* (1.2) *are Lipschitz. Then, for $p > 1$, there exists a constant $C_p > 0$, depending only on the initial condition x_0 and the Lipschitz constants of the coefficients b and σ, such that for any $t \in [0, 1]$,*

$$\mathbb{E}\left[\sup_{0 \le t \le 1} \left|X^h_t\right|^p\right] \le C_p.$$

Proof. Follows from a standard application of martingale inequalities, Jensen's inequality and Grönwall's lemma, similarly to Remark 3.18. □

Convergence in L^p now follows immediately.

Proof of Lemma 3.17. By Corollary 3.21, X^h converges to X in L^2, and hence in L^q for all $q \in [1,2]$. For fixed $p \geq 2$, Lemma 3.22 gives a bound on the $(p+1)^{\text{th}}$ moment of X^h. Moreover, the $(p+1)^{\text{th}}$ moment of X is bounded by (3.3). Combining the L^2-convergence with the bounds in L^{p+1} implies L^p-convergence, as required. □

Remark 3.23. Let the coefficients of (3.1) be Lipschitz, and recall the notation $\pi^{\mathrm{KR}}_{\mu^h,\nu^h}$ for the Knothe-Rosenblatt rearrangement between the laws of the monotone Euler–Maruyama schemes for (3.1), with step-size $h > 0$. Also write $\pi^{\mathrm{sync}}_{\mu,\nu}$ for the synchronous coupling between μ and ν. Then, combining Lemma 3.17 and Theorem 2.6, for any $p \geq 1$, we have the convergence

$$\lim_{h\to 0} \mathcal{AW}_p(\pi^{\mathrm{KR}}_{\mu^h,\nu^h}, \pi^{\mathrm{sync}}_{\mu,\nu}) = 0.$$

3.3. **Proof of optimality under Lipschitz conditions.** We are now ready to prove the following result, which implies the conclusion of Theorem 1.3 under Lipschitz conditions on the coefficients in (3.1).

Proposition 3.24. *Let $b, \bar{b} : \mathbb{R} \to \mathbb{R}$ and $\sigma, \bar{\sigma} : \mathbb{R} \to \mathbb{R}_+$ be Lipschitz functions, and let $X, \bar{X}$ be the unique strong solutions of* (3.1) *with laws μ and ν, respectively.*

For $N \in \mathbb{N}$, set $h = \frac{1}{N}$, and define the monotone Euler–Maruyama schemes X^h and $\bar{X}^h$ for X and $\bar{X}$, respectively, as in Definition 3.11. Let μ^h be the law of $(X^h_{kh})_{k=0,\ldots,N}$ and ν^h the law of $(\bar{X}^h_{kh})_{k=0,\ldots,N}$. Then, for $p \geq 1$,

$$\lim_{h\to 0} h^{\frac{1}{p}} \mathcal{AW}_p(\mu^h, \nu^h) = \mathcal{AW}_p(\mu, \nu).$$

Moreover, the distance on the right hand side is attained by the synchronous coupling between μ and ν.

Remark 3.25. Note that, under the Lipschitz conditions on the coefficients, there exist p-integrable unique strong solutions of the SDEs (3.1), for $p \geq 1$; see Remark 3.18. Also, according to Lemma 3.22, the associated monotone Euler–Maruyama schemes are bounded in L^p, $p \geq 1$, since the L^2-bound implies an L^1-bound. The adapted Wasserstein distances in the statement of Proposition 3.24 are therefore well-defined.

Proof of Proposition 3.24. Let $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$. Then, by Proposition 2.2, there exists a process ρ and a ρ-correlated Wiener process $(W, \bar{W})$ such that π is the joint law of $(X^\rho, \bar{X}^\rho)$, the strong solution of the system

$$\begin{aligned} \mathrm{d}X^\rho_t &= b(X^\rho_t)\mathrm{d}t + \sigma(X^\rho_t)\mathrm{d}W_t, \quad X^\rho_0 = x_0, \\ \mathrm{d}\bar{X}^\rho_t &= \bar{b}(\bar{X}^\rho_t)\mathrm{d}t + \bar{\sigma}(\bar{X}^\rho_t)\mathrm{d}\bar{W}_t, \quad \bar{X}^\rho_0 = x_0, \end{aligned} \tag{3.6}$$

driven by $(W, \bar{W})$. Now fix $N \in \mathbb{N}$ and $h = \frac{1}{N}$, and let $X^{\rho,h}$ and $\bar{X}^{\rho,h}$ be the monotone Euler–Maruyama schemes for X^ρ and $\bar{X}^\rho$, respectively, as defined by Definition 3.11. Note that $\mathrm{Law}((X^{\rho,h}_{kh})_{k=0,\ldots,N}, (\bar{X}^{\rho,h}_{kh})_{k=0,\ldots,N}) \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)$, and so

$$\inf_{\pi' \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h,\nu^h)} \mathbb{E}^{\pi'}\left[h\sum_{k=1}^N |\omega_{kh} - \bar{\omega}_{kh}|^p\right] \leq \mathbb{E}\left[h\sum_{k=1}^N |X^{\rho,h}_{kh} - Y^{\rho,h}_{kh}|^p\right]. \tag{3.7}$$

We can write

$$\left|\mathbb{E}\left[\int_0^1 |X_t^\rho - \bar{X}_t^\rho|^p \mathrm{d}t\right]^{\frac{1}{p}} - \mathbb{E}\left[h\sum_{k=1}^N |X_{kh}^{\rho,h} - \bar{X}_{kh}^{\rho,h}|^p\right]^{\frac{1}{p}}\right|$$
$$\leq \mathbb{E}\left[\sum_{k=1}^N \int_{(k-1)h}^{kh} |X_{kh}^{\rho,h} - X_t^{\rho,h}|^p \mathrm{d}t\right]^{\frac{1}{p}} + \mathbb{E}\left[\sum_{k=1}^N \int_{(k-1)h}^{kh} |\bar{X}_{kh}^{\rho,h} - \bar{X}_t^{\rho,h}|^p \mathrm{d}t\right]^{\frac{1}{p}}$$
$$+ \mathbb{E}\left[\int_0^1 |X_t^\rho - X_t^{\rho,h}|^p \mathrm{d}t\right]^{\frac{1}{p}} + \mathbb{E}\left[\int_0^1 |\bar{X}_t^\rho - \bar{X}_t^{\rho,h}|^p \mathrm{d}t\right]^{\frac{1}{p}} \xrightarrow{h\to 0} 0,$$

where we bound the final two terms by Lemma 3.17, and for the first two terms we make the following estimates. Fixing $k = 1, \dots, N$ and $t \in ((k-1)h, kh]$, we bound

$$\begin{aligned}
\mathbb{E}\left[\left|X_{kh}^{\rho,h} - X_t^{\rho,h}\right|\right]^p &\leq p(kh-t)\mathbb{E}\left[\left|b(X_{(k-1)h}^{\rho,h})\right|\right]^p + p\mathbb{E}\left[\left|W_{kh}^h - W_t^h\right|^p \left|\sigma(X_{(k-1)h}^{\rho,h}\right|^p\right] \\
&\leq p(kh-t)\mathbb{E}\left[\left|b(X_{(k-1)h}^{\rho,h})\right|\right]^p + p\mathbb{E}\left[\left|W_{kh}^h - W_t^h\right|^{2p}\right]^{\frac{1}{2}} \mathbb{E}\left[\left|\sigma(X_{(k-1)h}^{\rho,h})\right|^{2p}\right]^{\frac{1}{2}} \\
&\leq C(h^{\frac{p}{2}} + h),
\end{aligned}$$

applying the triangle and Cauchy–Schwarz inequalities in the first two lines, and then bounding the moment of the increment of W^h by comparison with moments of a Gaussian, and using the Lipschitz property of b and σ to bound the remaining expectations. The same bound holds when we replace $X^{\rho,h}$ by $\bar{X}^{\rho,h}$.

Taking the infimum over ρ, or equivalently over $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$, in the right-hand side of (3.7) and taking the limit as $h \to 0$, we have

$$\limsup_{h\to 0} \inf_{\pi' \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)} \mathbb{E}^{\pi'}\left[h\sum_{k=1}^N |\omega_{kh} - \bar{\omega}_{kh}|^p\right] \leq \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \mathbb{E}^\pi\left[\int_0^1 |\omega_t - \bar{\omega}_t|^p \mathrm{d}t\right]. \tag{3.8}$$

To prove the reverse inequality, take $\rho \equiv 1$ in the above arguments, so that the ρ-correlated Wiener process $(W, \bar{W})$ is in fact (W, W). Writing $(X^{\mathrm{sync}}, \bar{X}^{\mathrm{sync}})$ for the strong solution of the system (3.6) driven by the perfectly correlated Wiener process (W, W), and $X^{\mathrm{sync},h}, \bar{X}^{\mathrm{sync},h}$ for the associated monotone Euler–Maruyama schemes as defined in Definition 3.11, we have

$$\lim_{h\to 0} \mathbb{E}\left[h\sum_{k=0}^N |X_{kh}^{\mathrm{sync},h} - \bar{X}_{kh}^{\mathrm{sync},h}|^p\right] = \mathbb{E}\left[\int_0^1 |X_t^{\mathrm{sync}} - \bar{X}_t^{\mathrm{sync}}|^p \mathrm{d}t\right].$$

For sufficiently large $N \in \mathbb{N}$ and $h = \frac{1}{N}$, Lemma 3.15 and Proposition 3.5 imply that the adapted Wasserstein distance $\mathcal{AW}_p(\mu^h, \nu^h)$ between μ^h and ν^h is attained by the Knothe–Rosenblatt rearrangement $\pi_{\mu^h,\nu^h}^{\mathrm{KR}}$. According to Lemma 3.16, this coupling $\pi_{\mu^h,\nu^h}^{\mathrm{KR}}$ coincides with the coupling induced by the discrete-time monotone Euler–Maruyama schemes driven by one common Wiener process; i.e. $\pi_{\mu^h,\nu^h}^{\mathrm{KR}} = \mathrm{Law}((X_{kh}^{\mathrm{sync},h})_{k=0,\dots,N}, (\bar{X}_{kh}^{\mathrm{sync},h})_{k=0,\dots,N})$. This gives us

$$\inf_{\pi' \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)} \mathbb{E}^{\pi'}\left[h\sum_{k=0}^N |\omega_{kh} - \bar{\omega}_{kh}|^p\right] = \mathbb{E}\left[h\sum_{k=1}^N |X_{kh}^{\mathrm{sync},h} - \bar{X}_{kh}^{\mathrm{sync},h}|^p\right].$$

Taking the limit as $h \to 0$, we have

$$\lim_{h\to 0} \inf_{\pi' \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)} \mathbb{E}^{\pi'}\left[h\sum_{k=0}^N |\omega_{kh} - \bar{\omega}_{kh}|^p\right] = \mathbb{E}\left[\int_0^1 |X_t^{\mathrm{sync}} - \bar{X}_t^{\mathrm{sync}}|^p \mathrm{d}t\right],$$

and combining this with (3.8), we can conclude. □

In Section 4, we will conclude the proof of Theorem 1.3 — treating the case where the coefficients of the SDEs are merely required to be continuous — by proving a stability result for SDEs that enables us to relax the Lipschitz conditions that were imposed in this section. First we will extend the Lipschitz result in a different direction, where we can apply the same discretisation scheme as employed above, after first applying a transformation to the SDEs.

3.4. **Extension to discontinuous drifts.** We now prove the conclusion of Theorem 1.3 under a different set of assumptions, which allow for discontinuities in the drift. We first apply a Zvonkin-type transformation (see [36]) to remove the drift and then use the monotone Euler–Maruyama scheme introduced in Definition 3.11 for the resulting martingale with Lipschitz diffusion coefficient.

Assumption 3.26. Suppose that the coefficients $b, \bar{b}, \sigma, \bar{\sigma}$ of the SDEs (3.1) satisfy the following conditions:

(i) $b, \bar{b}$ are bounded and measurable;
(ii) $\sigma, \bar{\sigma}$ are bounded, uniformly positive and Lipschitz continuous; *and*
(iii) $b/\sigma^2, \bar{b}/\bar{\sigma}^2$ are Lebesgue-integrable.

Remark 3.27. By Zvonkin's theorem [36], there exist unique strong solutions $(X_t)_{t\geq 0}, (\bar{X}_t)_{t\geq 0}$ of the SDEs (3.1) under Assumption 3.26 (i)–(ii). In fact, for well-posedness of the SDEs, the Lipschitz continuity can be weakened to $\frac{1}{2}$-Hölder continuity, but we will make use of the Lipschitz condition later on in order to apply the monotone Euler–Maruyama scheme.

Proposition 3.28. *Under Assumption 3.26, the conclusion of Theorem 1.3 holds. Namely, for any $p \geq 1$, the synchronous coupling $\pi^{\mathrm{sync}}_{\mu,\nu}$ attains the adapted Wasserstein distance $\mathcal{AW}_p(\mu, \nu)$ between the laws μ, ν of the unique strong solutions of* (3.1).

To prove this result, we use exactly the drift-removing transformation introduced by Zvonkin in [36] to prove existence and uniqueness of strong solutions. Define the increasing map $T : \mathbb{R} \to \mathbb{R}_+$ by

$$T(x) := \int_{x_0}^{x} \exp\left\{-2\int_{x_0}^{z} \frac{b(y)}{\sigma^2(y)}\mathrm{d}y\right\}\mathrm{d}z, \quad x \in \mathbb{R},$$

and let $Y_t := T(X_t)$, for $t \in [0, 1]$, where X is the unique strong solution of (1.2) with coefficients b, σ. Then, by Itô's formula, Y solves the SDE

$$\mathrm{d}Y_t = (\sigma T') \circ T^{-1}(Y_t)\mathrm{d}W_t; \quad Y_0 = T(x_0). \tag{3.9}$$

Lemma 3.29. *Suppose that $b : \mathbb{R} \to \mathbb{R}$ is bounded and measurable, and that $\sigma : \mathbb{R} \to \mathbb{R}_+$ is bounded, uniformly positive, and Lipschitz. Then the map $(\sigma T') \circ T^{-1} : \mathbb{R} \to \mathbb{R}_+$ is Lipschitz.*

Proof. For $x_1, x_2 \in \mathbb{R}$, we are required to show that there is some constant $K > 0$ such that

$$|\sigma(x_2)T'(x_2) - \sigma(x_1)T'(x_1)| \leq K|T(x_2) - T(x_1)|.$$

Since σ is Lipschitz, there exists a Lebesgue-almost everywhere derivative σ' that is Lebesgue-almost surely bounded by the Lipschitz constant K^σ of σ. Hence $\sigma T'$ is also Lebesgue-almost surely differentiable, and its derivative satisfies

$$(\sigma T')'(x) = \sigma'(x)\exp\left\{-2\int_{x_0}^{x} \frac{b(y)}{\sigma^2(y)}\mathrm{d}y\right\} - 2\frac{b(x)}{\sigma(x)}\exp\left\{-2\int_{x_0}^{x} \frac{b(y)}{\sigma^2(y)}\mathrm{d}y\right\},$$

for Lebesgue-almost every $x \in \mathbb{R}$. Then, integrating, we have

$$|\sigma(x_2)T'(x_2) - \sigma(x_1)T'(x_1)| \leq \left(K^\sigma + 2\frac{\|b\|_\infty}{\inf_{y\in\mathbb{R}}\sigma(y)}\right)|T(x_2) - T(x_1)|,$$

using the Lebesgue-almost sure bound on σ' and the assumption that σ is bounded away from zero. □

In light of Lemma 3.29, we can apply the monotone Euler–Maruyama scheme defined in Definition 3.11 to the transformed SDE (3.9). Fix $N \in \mathbb{N}$ and $h = \frac{1}{N}$. The full scheme for the SDE (1.2) in this case is then as follows. Let $X^h_0 = x_0$ and, for $k = 0, \dots, N-1$ and $t \in (kh, (k+1)h]$, define

$$X^h_t := T^{-1}\left[T(X^h_{kh}) + T'(X^h_{kh})\sigma(X^h_{kh})(W^h_t - W^h_{kh})\right]. \tag{3.10}$$

The map T is increasing and invertible with increasing inverse. Therefore, for $h > 0$ sufficiently small and $k = 0, \dots, N-1$, we can use the same arguments as in the proof of Lemma 3.15, along with the fact that $\sigma T' \circ T^{-1}$ is Lipschitz, to see that $X^h_{kh} \mapsto X^h_{(k+1)h}$ is a concatenation of three increasing maps. Hence the process $(X^h_{kh})_{k=0,\dots,N}$ is increasing in first order stochastic dominance.

Remark 3.30. Assumption 3.26.(iii) guarantees that T^{-1} is Lipschitz, with some Lipschitz constant $C > 0$. Indeed, if $x < x'$, then

$$|T(x') - T(x)| = \int_x^{x'} \exp\left\{\int_{x_0}^z -2\frac{b(y)}{\sigma^2(y)}\mathrm{d}y\right\}\mathrm{d}z \geq (x' - x)\exp\left\{-2\sup_{z\in\mathbb{R}}\int_{x_0}^z \frac{b(y)}{\sigma^2(y)}\mathrm{d}y\right\}.$$

Applying the moment bound (3.3) to the solution Y of (3.9) thus gives us that the solution X of (1.2) is also p-integrable, for $p \geq 1$. Similarly, applying Lemma 3.22 to the monotone Euler–Maruyama scheme Y^h for (3.9) yields an L^p-bound for X^h defined by (3.10), for $h > 0$ and $p \geq 1$. Hence all required adapted Wasserstein distances are well-defined.

Furthermore, for any $p \geq 1$, $h > 0$ and $s \in [0, 1]$, we can write

$$|X^h_s - X_s|^p = |T^{-1}(Y^h_s) - T^{-1}(Y_s)|^p \leq C|Y^h_s - Y_s|^p.$$

Hence, from the L^p-convergence of Y^h to Y given by Lemma 3.17, we can deduce L^p-convergence of X^h to X.

Proof of Proposition 3.28. We follow the proof of Proposition 3.24, using the L^p-convergence of the scheme defined by (3.10) together with the Lipschitz continuity of T^{-1} to conclude. □

4. $\mathcal{AW}_p$ between laws of SDEs with continuous coefficients

In this section, we complete the proof of Theorem 1.3 under Assumption 1.2; that is, we relax the above Lipschitz assumption and show optimality of the synchronous coupling for the adapted Wasserstein distance between laws of SDEs for which pathwise uniqueness holds and whose coefficients are continuous and have linear growth.

We proceed by first establishing a general stability result for SDEs in Section 4.1. In Section 4.2, we then make use of this stability result to deduce the full result of Theorem 1.3, by making uniform approximations and applying the result for SDEs with Lipschitz coefficients from Proposition 3.24.

We start by making some remarks on Assumption 1.2 and providing examples of coefficients that satisfy this assumption.

Remark 4.1. Under Assumption 1.2, strong existence is guaranteed for the SDEs (3.1). Indeed, by a result of Skorokhod [32], there exist weak solutions of the SDEs under the given continuity and linear growth assumptions on the coefficients. Then, by the Yamada-Watanabe criterion [20, Ch. 5, Corollary 3.23], the combination of pathwise uniqueness and weak existence implies the existence of strong solutions. We refer to [13] for an example of a Markovian SDE for which strong existence does not hold.

Remark 4.2. Assumption 1.2 is satisfied, for example, in the following cases:

(i) $b, \bar{b}, \sigma, \bar{\sigma}$ are Lipschitz [12, Proposition 1.9 (Itô)] — see Proposition 3.24;

(ii) $b, \bar{b}, \sigma, \bar{\sigma}$ are continuous and bounded, $\sigma, \bar{\sigma}$ are $\frac{1}{2}$-Hölder continuous and bounded below by a positive constant [12, Proposition 1.10 (Zvonkin)] — see Proposition 3.28;

(iii) $b, \bar{b}, \sigma, \bar{\sigma}$ are continuous with linear growth, $\sigma, \bar{\sigma}$ are strictly positive and $\frac{1}{2}$-Hölder continuous, and $b/\sigma^2, \bar{b}/\bar{\sigma}^2$ are locally Lebesgue-integrable [12, Proposition 1.11 (Engelbert–Schmidt)];

(iv) $b, \bar{b}$ are Lipschitz, $\sigma, \bar{\sigma}$ have linear growth and are uniformly continuous with a strictly increasing modulus of continuity $h : \mathbb{R}_+ \to \mathbb{R}$ satisfying $\int_0^{0+} h^{-2}(x)\mathrm{d}x = +\infty$ [12, Proposition 1.12 (Yamada–Watanabe)].

4.1. **A stability result for SDEs.** We now provide our stability result for SDEs. We prove this result in a more abstract setting which allows for path-dependent coefficients.

To this end, recall the notation $\Omega = C([0,1],\mathbb{R})$ and also denote the sup-norm $\|\omega\|_\infty := \sup_{s\in[0,1]} |\omega_s|$, $\omega \in \Omega$. As before, Ω is equipped with the canonical filtration and the uniform topology. We also equip $\Omega \times \Omega$ etc. with the product filtration and product topology. Moreover, for probability measures μ, ν on $\Omega \times \Omega$, which we write as $\mu, \nu \in \mathcal{P}(\Omega \times \Omega)$, and for any $p \geq 1$, we define the p-Wasserstein distance between μ and ν to be

$$\inf_{\pi\in \mathrm{Cpl}(\mu,\nu)} \mathbb{E}^\pi \left[\|\omega - \bar{\omega}\|_\infty^p + \|\omega' - \bar{\omega}'\|_\infty^p\right], \tag{4.1}$$

when μ, ν have finite p^{th} moment, where ω', $\bar{\omega}'$ are canonical processes on the second coordinate of $\Omega \times \Omega$.

Remark 4.3. For any $p \geq 1$, μ_n converges to μ with respect to the p-Wasserstein distance on $\mathcal{P}(\Omega \times \Omega)$, as defined in (4.1), if and only if, for any continuous function $\phi : \Omega \times \Omega \to \mathbb{R}$ with at most polynomial growth of order p — i.e. $|\phi(\omega, \omega')| \leq C(1 + \|\omega\|_\infty^p + \|\omega'\|_\infty^p)$, $(\omega, \omega') \in \Omega \times \Omega$ — it holds that $\mathbb{E}^{\mu^n}[\phi(\omega, \omega')] \to \mathbb{E}^{\mu}[\phi(\omega, \omega')]$ (see e.g. [34, Definition 6.8]).

We will work under the following assumption.

Assumption 4.4. Suppose that $x_0, \bar{x}_0 \in \mathbb{R}$ and $b, \bar{b} : [0,1] \times \Omega \to \mathbb{R}$, $\sigma, \bar{\sigma} : [0,1] \times \Omega \to \mathbb{R}_+$ satisfy the following:

(i) $b, \bar{b}, \sigma, \bar{\sigma}$ are progressively measurable;
(ii) for each $t \in [0,1]$, the functions $b(t,\cdot), \bar{b}(t,\cdot), \sigma(t,\cdot), \bar{\sigma}(t,\cdot)$ are continuous w.r.t. $\|\cdot\|_\infty$;
(iii) there exists $K > 0$ such that, for all $t \in [0,1]$, $\omega \in \Omega$,

$$|b(t,\omega)| \vee |\bar{b}(t,\omega)| \vee |\sigma(t,\omega)| \vee |\bar{\sigma}(t,\omega)| \leq K(1 + \|\omega\|_\infty); \tag{4.2}$$

(iv) there exists a unique strong solution of each of the SDEs

$$\begin{aligned} dX_t &= b(t,X)dt + \sigma(t,X)dW_t, \quad X_0 = x_0, \\ d\bar{X}_t &= \bar{b}(t,\bar{X})dt + \bar{\sigma}(t,\bar{X})d\bar{W}_t, \quad \bar{X}_0 = \bar{x}_0. \end{aligned} \tag{4.3}$$

Recall from Definition 2.1 that $(W, \bar{W})$ is a ρ-correlated Wiener process, if $d\langle W, \bar{W}\rangle_t = \rho(t, W, \bar{W})dt$, for some progressively measurable $\rho : [0,1] \times \Omega \times \Omega \to [-1,1]$.

Remark 4.5. Whenever $(W, \bar{W})$ is a ρ-correlated Wiener process defined on some stochastic basis, under Assumption 4.4 (iv), one can uniquely construct a strong solution $(X, \bar{X})$ of the system (4.3) driven by $(W, \bar{W})$ on the same stochastic basis. We also note that 4.4 (iv) can be weakened to pathwise uniqueness only, similarly to Assumption 1.2. Indeed, weak existence is guaranteed already by Assumption 4.4 (i)–(iii) and a classical result of Skorokhod (e.g. adapting the proof of [19, Theorem 21.9] to coefficients with linear growth) and so the Yamada-Watanabe criterion [19, Lemma 21.17] applies.

Proposition 4.6. *Let $(W, \bar{W})$ be a ρ-correlated Wiener process, for some progressively measurable process ρ.*

Suppose that $(x_0, \bar{x}_0, b, \bar{b}, \sigma, \bar{\sigma})$ satisfies Assumption 4.4, and write $(X, \bar{X})$ for the unique strong solution of (4.3) driven by $(W, \bar{W})$.

For $n \in \mathbb{N}$, consider also $(x_0^n, \bar{x}_0^n, b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)$ satisfying Assumption 4.4.(i) and (iii), with a uniform slope constant K in (4.2), and such that strong existence holds for (4.3); let $(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n})$ be one such strong solution, when (4.3) is driven by $(W, \bar{W})$.

Suppose also that, as $n \to \infty$, $(x_0^n, \bar{x}_0^n) \to (x_0, \bar{x}_0)$ and the following convergence holds:

$$\|\omega^n - \omega\|_\infty \to 0 \implies (b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)(t, \omega^n) \to (b, \bar{b}, \sigma, \bar{\sigma})(t, \omega), \text{ for each } t \in [0,1]. \tag{4.4}$$

Then, for any $p \geq 1$,

$$\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}) \xrightarrow{n\to\infty} \text{Law}(X, \bar{X}),$$

with respect to the p-Wasserstein distance on $\mathcal{P}(\Omega \times \Omega)$.

Remark 4.7. In fact continuity of $b, \bar{b}, \sigma, \bar{\sigma}$ is implied by the convergence (4.4) and so we could drop Assumption 4.4.(ii) from our assumptions. We keep this assumption, however, to make it transparent that the stability result of Proposition 4.6 cannot be applied to coefficients with discontinuities. As such, Proposition 3.28 from the preceding section is the most general result that we are able to obtain for SDEs with discontinuous coefficients using the methods of this paper.

Remark 4.8. In the case that $W, \bar{W}$ are independent Wiener processes and all coefficients are Markovian, then the above stability result can be found, for example, in Stroock and Varadhan [33, Theorem 11.1.4]. For the case of correlated Wiener processes, however, we are not aware of an existing result in the literature.

Proof of Proposition 4.6. Similarly to Remark 3.18, standard SDE estimates based on the BDG inequality, Jensen's inequality, and Grönwall's lemma show the existence of $K_p < \infty$ such that $\mathbb{E}[\|X^{b^n,\sigma^n}\|_\infty^p] \leq K_p(1 + |x_0^n|^p)$, with similar bounds for $\bar{X}^{\bar{b}^n,\bar{\sigma}^n}$. As $(x_0^n, \bar{x}_0^n)_{n\in\mathbb{N}}$ converges, this shows that, for all $p \geq 1$, the p^{th} moments of $\|X^{b^n,\sigma^n}\|_\infty$ and $\|\bar{X}^{\bar{b}^n,\bar{\sigma}^n}\|_\infty$ are uniformly bounded in $n \in \mathbb{N}$. On the one hand, this implies that $(\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}))_{n\in\mathbb{N}}$ is tight, and on the other hand that it suffices to prove that $\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}) \to \text{Law}(X, \bar{X})$ weakly on $\mathcal{P}(\Omega \times \Omega)$. Thanks to Assumption 4.4.(iv), this can be achieved if we prove that each weak accumulation point of $(\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}))_{n\in\mathbb{N}}$ solves the martingale problem associated to the system for $(X, \bar{X})$.

Let $\text{Law}(Y, \bar{Y})$ be one such weak accumulation point, and consider the enlarged joint law $(\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}, W, \bar{W}))_{n\in\mathbb{N}}$, which is also tight. Then, after possibly passing to a subsequence, Skorokhod's representation theorem ensures the existence of a sequence of stochastic processes $(X^n, \bar{X}^n, W^n, \bar{W}^n)_{n\in\mathbb{N}}$ defined on a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

$$(\text{Law}(X^n, \bar{X}^n, W^n, \bar{W}^n))_{n\in\mathbb{N}} = (\text{Law}(X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}, W, \bar{W}))_{n\in\mathbb{N}},$$

and $(X^n, \bar{X}^n, W^n, \bar{W}^n) \to (Y, \bar{Y}, W^\infty, \bar{W}^\infty)$ almost surely in $\Omega \times \Omega \times \Omega \times \Omega$. Moreover, for each $n \in \mathbb{N}$, there exist deterministic maps $F^n, \bar{F}^n$ such that $X^{b^n,\sigma^n} = F^n(W)$ and $\bar{X}^{\bar{b}^n,\bar{\sigma}^n} = \bar{F}^n(\bar{W})$, and so $X^n = F^n(W^n)$ and $\bar{X}^n = \bar{F}^n(\bar{W}^n)$. Therefore $(X^n, \bar{X}^n)$ is a strong solution of the system (4.3), with coefficients $b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n$, driven by $(W^n, \bar{W}^n)$. By the equality in law, we can also verify that $(W^n, \bar{W}^n)$ is a ρ-correlated Wiener process in its own filtration.

Let $\varepsilon > 0$. By Lusin's theorem applied to the measurable function ρ, we can find a closed set $E \subset [0,1] \times \Omega \times \Omega$ with $m_\varepsilon := (dt \times \mathbb{P})(\{(t, \omega) : (t, W^n(\omega), \bar{W}^n(\omega)) \notin E\}) \leq \varepsilon$ and $\rho|_E$ continuous. We remark that m_ε is independent of $n \in \mathbb{N} \cup \{\infty\}$ as it only depends on the joint law of $(W^n, \bar{W}^n)$. By Tietze's theorem there exists a continuous function $\rho^\varepsilon : [0,1] \times \Omega \times \Omega \to [-1, 1]$, which coincides with ρ on E. For $n \in \mathbb{N}$, the martingale problem associated with the system for $(X^n, \bar{X}^n)$ reads as follows: for every bounded $f : [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ which is differentiable in time, twice differentiable in space, and with corresponding bounded and continuous derivatives, it holds that $R_f^n = 0$, where

$$R_f^n := \mathbb{E}\Big[f(T, X_1^n, \bar{X}_1^n) - f(0, x_0^n, \bar{x}_0^n) - \int_0^1 \{\partial_t f + b^n \partial_x f + \bar{b}^n \partial_{\bar{x}} f + \frac{1}{2}(\sigma^n)^2 \partial_{xx} f + \frac{1}{2}(\bar{\sigma}^n)^2 \partial_{\bar{x}\bar{x}} f + \rho(t, W^n, \bar{W}^n)\sigma^n \bar{\sigma}^n \partial_{x\bar{x}} f\}(t, X^n, \bar{X}^n) dt\Big],$$

and we identify $f(t, X, Y) \equiv f(t, X_t, Y_t)$. On the other hand, we may also define

$$R_f^\infty := \mathbb{E}\Big[f(T, Y_1, \bar{Y}_1) - f(0, x_0, \bar{x}_0) - \int_0^1 \{\partial_t f + b\partial_x f + \bar{b}\partial_{\bar{x}} f + \frac{1}{2}\sigma^2\partial_{xx} f + \frac{1}{2}\bar{\sigma}^2\partial_{\bar{x}\bar{x}} f + \rho(t, W^\infty, \bar{W}^\infty)\sigma\bar{\sigma}\partial_{x\bar{x}} f\}(t, Y, \bar{Y})dt\Big],$$

and so our goal is to show that $R_f^\infty = 0$ for all f in the aforementioned class of functions. To this end, for $n \in \mathbb{N}$, we introduce $R_f^{n,\varepsilon}$ and $R_f^{\infty,\varepsilon}$, defined analogously to R_f^n and R_f^∞ with ρ replaced by ρ^ε. We claim that $\lim_{n\to\infty} R_f^{n,\varepsilon} = R_f^{\infty,\varepsilon}$. To see this, note first that since ρ^ε as well as f and its partial derivatives are continuous, and since $b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n$ converge in the sense of (4.4), the integrand converges $\mathrm{d}t \times \mathbb{P}$-a.s.. In turn, since ρ^ε as well as f and its partial derivatives are bounded, and since $b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n$ have uniform linear growth, for $n \in \mathbb{N}$, we can leverage the uniform moment estimates given at the start of the proof to apply dominated convergence and conclude the desired claim. Next, note that $|R_f^n - R_f^{n,\varepsilon}| \le C\varepsilon$, for all $n \in \mathbb{N} \cup \{\infty\}$, with a constant C depending on f but, crucially, independent of n and ε; this follows again by the uniform linear growth assumption and the uniform moment estimates, which extend to Y and $\bar{Y}$. To conclude, we write

$$|R_f^\infty| \le |R_f^\infty - R_f^{\infty,\varepsilon}| + \lim_{n\to\infty} |R_f^{\infty,\varepsilon} - R_f^{n,\varepsilon}| + \lim_{n\to\infty} |R_f^{n,\varepsilon} - R_f^n| \le 2C\varepsilon.$$

□

We return now to the setting of SDEs with time-homogeneous Markovian coefficients, as in the rest of the paper. In order to apply the stability result Proposition 4.6, we identify $b(t, \omega) = b(\omega_t)$, $t \in [0, 1]$, $\omega \in \Omega$, for a Markovian coefficient $b : \mathbb{R} \to \mathbb{R}$, for example.

The following corollary shows how we can apply the above stability result to find the adapted Wasserstein distance between laws of SDEs by approximating their coefficients.

Corollary 4.9. *For $n \in \mathbb{N}$, let $b, \bar{b}, b^n, \bar{b}^n : \mathbb{R} \to \mathbb{R}$, $\sigma, \bar{\sigma}, \sigma^n, \bar{\sigma}^n : \mathbb{R} \to \mathbb{R}_+$ and suppose that the assumptions of Proposition 4.6 are satisfied for the Markovian coefficients $(b, \bar{b}, \sigma, \bar{\sigma})$ and $(b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)_{n\in\mathbb{N}}$.*

Let $(W, \bar{W})$ be a ρ-correlated Wiener process and, maintaining the notation of Proposition 4.6, write μ, ν for the laws of $X, \bar{X}$ and, for $n \in \mathbb{N}$, write μ^n, ν^n for the laws of $X^{b^n,\sigma^n}, \bar{X}^{\bar{b}^n,\bar{\sigma}^n}$.

Let $p \ge 1$ and suppose that, for each $n \in \mathbb{N}$, the synchronous coupling $\pi^{\mathrm{sync}}_{\mu^n,\nu^n}$ is optimal for $\mathcal{AW}_p(\mu^n, \nu^n)$. Then the synchronous coupling $\pi^{\mathrm{sync}}_{\mu,\nu}$ is optimal for $\mathcal{AW}_p(\mu, \nu)$.

Proof. For $n \in \mathbb{N}$, we have, by the assumption of optimality of $\pi^{\mathrm{sync}}_{\mu^n,\nu^n}$ and Proposition 2.2, that

$$\mathbb{E}^{\pi^{\mathrm{sync}}_{\mu^n,\nu^n}}\left[\int_0^1 |\omega_t - \bar{\omega}_t|^p dt\right] \le \mathbb{E}\left[\int_0^1 \left|X_t^{b^n,\sigma^n} - \bar{X}_t^{\bar{b}^n,\bar{\sigma}^n}\right|^p dt\right].$$

From Proposition 4.6 we have convergence of $\pi^{\mathrm{sync}}_{\mu^n,\nu^n}$ to $\pi^{\mathrm{sync}}_{\mu,\nu}$, with respect to the p-Wasserstein distance on $\mathcal{P}(\Omega \times \Omega)$. As the function $(\omega, \bar{\omega}) \mapsto \int_0^1 |\omega_t - \bar{\omega}_t|^p dt$ is continuous on $\Omega \times \Omega$, and has polynomial growth at rate p, we conclude by Remark 4.3 that

$$\mathbb{E}^{\pi^{\mathrm{sync}}_{\mu^n,\nu^n}}\left[\int_0^1 |\omega_t - \bar{\omega}_t|^p dt\right] \to \mathbb{E}^{\pi^{\mathrm{sync}}_{\mu,\nu}}\left[\int_0^1 |\omega_t - \bar{\omega}_t|^p dt\right].$$

By the same reasoning,

$$\mathbb{E}\left[\int_0^1 \left|X_t^{b^n,\sigma^n} - \bar{X}_t^{\bar{b}^n,\bar{\sigma}^n}\right|^p dt\right] \to \mathbb{E}\left[\int_0^1 |X_t - \bar{X}_t|^p dt\right].$$

Thanks to Proposition 2.2, we can conclude. □

4.2. **Proof of Theorem 1.3.** We have now developed all of the tools required to prove the main theorem.

Proof of Theorem 1.3. Under Lipschitz conditions on the coefficients of the SDEs (3.1), we have already proved the conclusion of the theorem in Proposition 3.24. Now suppose that the more general condition of Assumption 1.2 is satisfied, namely assume that the coefficients are continuous with linear growth, and that pathwise uniqueness holds. Then, according to Remark 4.5, the SDEs (3.1) satisfy Assumption 4.4, where notably the coefficients $b, \bar{b}, \sigma, \bar{\sigma}$ are Markovian and time-homogeneous. We can approximate $(b, \bar{b}, \sigma, \bar{\sigma})$ locally uniformly by a sequence of Lipschitz functions $(b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)_{n\in\mathbb{N}}$ on $\mathbb{R}$, which all satisfy the same linear growth bound. Note that locally uniform convergence of the Markovian coefficients implies convergence in the sense of (4.4). This guarantees all the conditions of Proposition 4.6. Defining μ^n, ν^n as the laws of the solutions of (3.1) with coefficients $(b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)$, for $n \in \mathbb{N}$, we have by Proposition 3.24 that the synchronous coupling $\pi^{\text{sync}}_{\mu^n,\nu^n}$ is optimal for $\mathcal{AW}_p(\mu^n, \nu^n)$. Hence, by Corollary 4.9, the synchronous coupling $\pi^{\text{sync}}_{\mu,\nu}$ is optimal for $\mathcal{AW}_p(\mu, \nu)$. □

Remark 4.10. We note that we can further extend Theorem 1.3 by combining different sets of assumptions. If the coefficients (b, σ) satisfy Assumption 1.2 and the coefficients $(\bar{b}, \bar{\sigma})$ satisfy Assumption 3.26 (or vice-versa), then the conclusion of Theorem 1.3 still holds, by the following reasoning.

Suppose that $(\bar{b}, \bar{\sigma})$ satisfy Assumption 3.26. If (b, σ) are Lipschitz, then by examining the proofs of Proposition 3.24 and Proposition 3.28, it is straightforward to see that the result still holds.

Now suppose that (b, σ) satisfy Assumption 1.2 but are not Lipschitz. We then need to adapt the stability result of Proposition 4.6 because the coefficients $(\bar{b}, \bar{\sigma})$ may not be continuous. As before, we can approximate (b, σ) by Lipschitz functions in the sense of (4.4). On the other hand, we fix the constant sequence $(\bar{b}^n, \bar{\sigma}^n) = (\bar{b}, \bar{\sigma})$, for all $n \in \mathbb{N}$. Then, after applying Skorokhod's representation theorem, we apply Lusin's theorem for a second time, in order to find a continuous function that coincides with $\bar{b}$ on a set of arbitrarily large measure, with respect to the law of the fixed process $X^{\bar{b},\bar{\sigma}}$. We can then follow the remainder of the proof as above.

Remark 4.11 (time-dependent coefficients)**.** One can also extend Theorem 1.3 to the time-inhomogeneous case. In particular, assuming that the coefficients are Lipschitz in space uniformly in time, Lipschitz in time uniformly in space, and have linear growth in space, the proofs in Section 2 and Section 3 remain valid with only minor modifications. In this case, the monotone Euler–Maruyama scheme should be modified such that the expression for X^h_t in Definition 3.11 is replaced by $X^h_t = X^h_{kh} + (t - kh)b(kh, X^h_{kh}) + \sigma(kh, X^h_{kh})(W^h_t - W^h_{kh})$. Under the given conditions on the coefficients, the corresponding standard Euler–Maruyama scheme converges in L^2 by [21, Theorem 10.2.2] and one can deduce L^p convergence of the monotone scheme, for any $p \geq 1$, as in Section 3.2. Further, the above stability argument still applies and also allows us to pass to coefficients that are only continuous in time. Therefore Theorem 1.3 holds under Assumption 1.2 for time-dependent coefficients that are also continuous in time.

5. On the topology induced by the adapted Wasserstein distance

In this section, we apply the stability result of Proposition 4.6 to prove Theorem 1.5. This theorem states that, restricted to a particular subset of probability measures, the topology induced by the adapted Wasserstein distance coincides with the topologies induced by the synchronous distance, the (symmetric) causal Wasserstein distance and the classical Wasserstein distance, as defined in the introduction, as well as with the topologies of weak convergence and convergence in finite-dimensional distributions.

We recall the following notation. When strong existence and pathwise uniqueness hold for the SDE (1.2) with coefficients (b, σ), we write $X^{b,\sigma}$ for the unique strong solution, and $\mu^{b,\sigma}$ for its law. We then define $\mathcal{P}^*$ to be the set of all such laws. For $\Lambda > 0$, we define $\mathcal{A}^\Lambda$ to be the set of Λ-Lipschitz functions whose absolute value at zero is also bounded by Λ, and then define $\mathcal{P}^\Lambda := \{\mu^{b,\sigma}\colon\ (b,\sigma) \in \mathcal{A}^\Lambda \times \mathcal{A}^\Lambda\} \subset \mathcal{P}^*$. A further result of Theorem 1.5 is that the set $\mathcal{P}^\Lambda$ is compact with respect to the topology induced by the adapted Wasserstein distance.

We introduce a further subset $\bar{\mathcal{P}} := \{\mu^{b,\sigma} : (b,\sigma) \text{ satisfies Assumption 1.2}\} \subset \mathcal{P}^*$. Also recall the set $\mathcal{P}_p$ of measures on Ω with finite p^{th} moment, for $p \geq 1$. We note that $\sup_{t\in[0,1]} \omega_t \in L^p(\mu)$, for any $\mu \in \bar{\mathcal{P}}$ and $p \geq 1$, by the estimates given in Remark 3.18. For $p \geq 1$, the synchronous distance $\mathcal{SW}_p$ is therefore well-defined on $\bar{\mathcal{P}}$, and $(\bar{\mathcal{P}}, \mathcal{SW}_p)$ is a metric space.

Proof of Theorem 1.5. Note first that, for $p \geq 1$, we have the inclusion $\mathcal{P}^\Lambda \subset \bar{\mathcal{P}} \subset \mathcal{P}_p$. The subspace topologies obtained by restricting the topologies listed in the statement of the theorem to $\mathcal{P}^\Lambda$ are thus well-defined.

Let us define another (a priori stronger) topology τ_p on $\mathcal{P}^\Lambda$ to be the topology induced by the distance $\mathcal{SW}_p^\infty$, defined by

$$\mathcal{SW}_p^\infty(\mu,\nu) := \mathbb{E}^{\pi^{\text{sync}}_{\mu,\nu}}\left[\sup_{0\leq t\leq 1} |\omega_t - \bar{\omega}_t|^p\right]^{1/p}, \quad \mu,\nu \in \bar{\mathcal{P}}, \quad p \in [1,\infty).$$

We first show that, for $p \in [1,\infty)$, $\mathcal{P}^\Lambda$ is (sequentially) compact with respect to τ_p and that τ_p is independent of p. To this end, note that, viewed as a subset of continuous functions, $\mathcal{A}^\Lambda$ is (sequentially) compact with respect to the topology of local uniform convergence. Indeed, consider a sequence $(\varphi_n)_{n\in\mathbb{N}} \subset \mathcal{A}^\Lambda$. For every $K > 0$, by the Arzelà-Ascoli theorem, there exists a subsequence that converges uniformly on $[-K, K]$; it follows by a diagonalisation argument that the sequence $(\varphi_n)_{n\in\mathbb{N}}$ converges locally uniformly and its limit belongs to $\mathcal{A}^\Lambda$. Consider a sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}^\Lambda$ and let $(b_n, \sigma_n)_{n\in\mathbb{N}} \subset \mathcal{A}^\Lambda \times \mathcal{A}^\Lambda$ be such that $\mu_n = \mu^{b_n,\sigma_n}$. By the above compactness, an equally denoted subsequence $(b_n, \sigma_n)_{n\in\mathbb{N}}$ converges locally uniformly to some $(b,\sigma) \subset \mathcal{A}^\Lambda \times \mathcal{A}^\Lambda$; we write $\mu = \mu^{b,\sigma}$. Now fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a Wiener process W. Let $X^{b,\sigma}$ and X^{b_n,σ_n} be the unique strong solutions of the SDE (1.2) driven by the same Wiener process W, on the same probability space, with coefficients (b,σ) and (b^n, σ^n), respectively, for $n \in \mathbb{N}$. Applying Proposition 4.6, we obtain that $(\text{Law}(X^{b_n,\sigma_n}, X^{b,\sigma}))_{n\in\mathbb{N}}$ converges in the p-Wasserstein distance on $\mathcal{P}(\Omega\times\Omega)$ to $\text{Law}(X^{b,\sigma}, X^{b,\sigma})$, for any $p \geq 1$. Noting that, for any $p \in [1,\infty)$, the function $(\omega,\bar{\omega}) \mapsto \sup_{0\leq t\leq 1} |\omega_t - \bar{\omega}_t|^p$ is continuous and has at most rate p polynomial growth, we have by Remark 4.3 that,

$$\mathcal{SW}_p^\infty(\mu_n,\mu) = \mathbb{E}\left[\sup_{0\leq t\leq 1}\left|X_t^{b_n,\sigma_n} - X_t^{b,\sigma}\right|^p\right]^{1/p} \xrightarrow[n\to\infty]{} 0,$$

for any $p \in [1,\infty)$. Therefore $\mathcal{P}^\Lambda$ is (sequentially) compact w.r.t. τ_p for every $p \in [1,\infty)$ and, by the same argument, all topologies τ_p on $\mathcal{P}^\Lambda$ coincide. Let us call τ this common topology.

We now use the following well-known fact for topological spaces (A,τ^A), (B,τ^B). If $I : (A,\tau^A) \to (B,\tau^B)$ is continuous and invertible, A is τ^A-compact, and τ^B is Hausdorff, then I^{-1} is continuous. Applied to I being the identity map, $A = B$, τ^B being Polish, and τ^B weaker than τ^A, this argument shows that if A is τ^A-compact then $\tau^A = \tau^B$.

As $\mathcal{P}^\Lambda$ is τ-compact, and by the previous paragraph, it now only remains to argue that convergence in each of the topologies listed in the theorem is implied by convergence in $\mathcal{SW}_p^\infty$, for some $p \in [1,\infty)$. It is clear that $\mathcal{SW}_p^\infty(\mu,\nu) \geq \mathcal{SW}_p(\mu,\nu)$, for any $p \in [1,\infty)$, $\mu,\nu \in \bar{\mathcal{P}}$. Now note that, for $\mu, \mu_n \in \bar{\mathcal{P}}$, $n \in \mathbb{N}$, we have $\pi^{\text{sync}}_{\mu_n,\mu} \in \text{Cpl}_{\text{bc}}(\mu_n,\mu)$, and therefore, for $p \in [1,\infty)$,

$$\lim_{n\to\infty} \mathcal{SW}_p^\infty(\mu_n,\mu) = 0 \implies \lim_{n\to\infty} \mathcal{SW}_p(\mu_n,\mu) = 0 \implies \lim_{n\to\infty} \mathcal{AW}_p(\mu_n,\mu) = 0.$$

Further, since $\text{Cpl}_{\text{bc}}(\mu,\nu) \subseteq \text{Cpl}_{\text{c}}(\mu,\nu) \subseteq \text{Cpl}(\mu,\nu)$, and since $\mathcal{AW}_p$ and $\mathcal{W}_p$ are both symmetric, we immediately get that for $p \geq 1$,

$$\mathcal{AW}_p(\mu,\nu) \geq \mathcal{SCW}_p(\mu,\nu) \geq \mathcal{W}_p(\mu,\nu), \quad \mu,\nu \in \mathcal{P}_p,$$

which yields the corresponding ordering for the topologies induced by these metrics. By the same token, convergence in $\mathcal{CW}_p$ is also implied by convergence in $\mathcal{SW}_p^\infty$. The convergence of μ_n to μ in $\mathcal{W}_1$ implies weak convergence of μ_n to μ with respect to the uniform topology on $C([0,1],\mathbb{R})$, which in turn implies that convergence holds also for the weak topology associated with the L^p-topology on $C([0,1],\mathbb{R})$, for any $p \in [1,\infty]$. Finally, convergence in finite-dimensional distributions is implied by convergence in the weak topologies above. □

Remark 5.1. The result of Theorem 1.5 clearly applies also to sets of the form $\{\mu^{b,\sigma} : (b,\sigma) \in \mathcal{A}^\Lambda \times \mathcal{A}^{\tilde{\Lambda}}\}$, $\Lambda, \tilde{\Lambda} > 0$. An inspection of the proof shows that the result also applies to the set $\{\mu^{b,\sigma} :\ b, \sigma \in \mathcal{A}^{\kappa,\Lambda},\ \sigma > 0\}$, where $\Lambda > 0$, $\kappa \in [1/2, 1]$ and

$$\mathcal{A}^{\kappa,\Lambda} = \{\varphi \in C(\mathbb{R},\mathbb{R}) : |\varphi(x) - \varphi(y)| \le \Lambda |x-y|^\kappa \text{ and } |\varphi(x)| \le \Lambda(1+|x|),\ x, y \in \mathbb{R}\},$$

applying the same Arzelà-Ascoli argument. In this case, existence and uniqueness of strong solutions is guaranteed by a result of Engelbert and Schmidt [12, Proposition 1.11] (c.f. Remark 4.2), as the continuity of the coefficients and the strict positivity of σ implies that the local integrability condition required for this result is satisfied.

6. Examples

In this final section, we collect some examples. We first provide one that motivates the use of the adapted Wasserstein distance when considering distances between processes. This is a continuous-time analogue of the example given in [4].

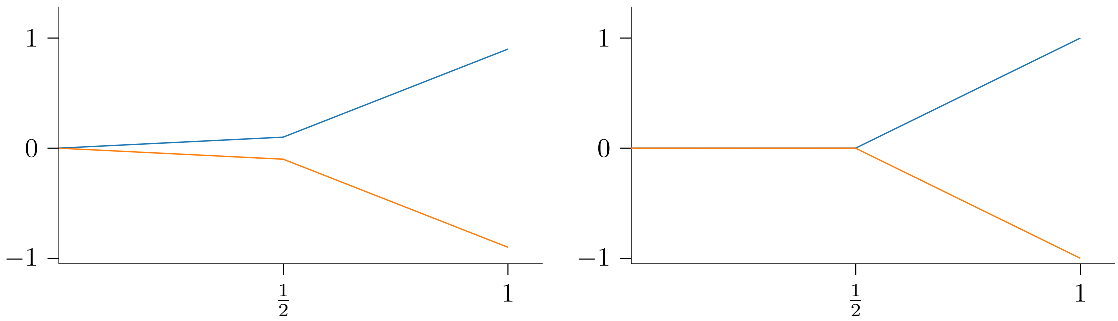

FIGURE 2. The two possible trajectories of X^n, for some $n \in \mathbb{N}$, are shown on the left, and the two possible trajectories of X^∞ on the right.

Example 6.1 (motivating example)**.** For $n \in \mathbb{N} \cup \{\infty\}$, define the process $(X_t^n)_{t\in[0,1]}$ with $X_0^n = 0$ such that

$$X_{\frac{1}{2}}^n = \frac{1}{n} \quad \text{and} \quad X_1^n = 1, \quad \text{with probability } \frac{1}{2},$$
$$X_{\frac{1}{2}}^n = -\frac{1}{n} \quad \text{and} \quad X_1^n = -1, \quad \text{with probability } \frac{1}{2},$$

linearly interpolated for intermediate times, and define $\mu^n := \mathrm{Law}(X^n)$. The two possible trajectories of X^n, $n \in \mathbb{N}$ are shown on the left-hand side of Figure 2, and the trajectories of X^∞ are shown on the right-hand side of Figure 2.

One can see that the behaviour of the approximating processes after time $\frac{1}{2}$ is completely determined by the history of the process up to that time, whereas the behaviour of the limiting process after time $\frac{1}{2}$ is independent of the past. The classical Wasserstein distance cannot distinguish these differing information structures.

For each $n \in \mathbb{N}$, it is possible to couple μ^n and μ^∞ in such a way that paths that terminate at a positive value are mapped onto each other, and likewise for negative values. Thus we can see that the Wasserstein distance between μ^n and μ^∞ converges to 0 as $n \to \infty$. Such a coupling is not bi-causal, however. In fact the only bi-causal coupling is the product coupling, which maps paths that terminate at a positive value onto those that terminate at a negative value with probability $\frac{1}{2}$. We can thus bound the adapted Wasserstein distance $\mathcal{AW}_p(\mu^n, \mu^\infty)$ from below by a positive constant, for any $p \geq 1$. Note that if we take the right-continuous version of the filtration in the definition of causality, Definition 1.1, then the previous argument still holds.

Consider finally the problem of finding $V^n := \inf_{Z\ \mathcal{F}^{X^n}_{\frac{1}{2}}\text{-measurable}} \mathbb{E}|X^n_1 - Z|^2$, for each $n \in \mathbb{N} \cup \{\infty\}$, with $\mathcal{F}^{X^n}$ denoting the raw natural filtration of X^n. Since X^n_1 is $\mathcal{F}^{X^n}_{\frac{1}{2}}$-measurable, for each $n \in \mathbb{N}$, we have $V^n = 0$. On the other hand, $V^\infty = 1$, since the sigma-algebra $\mathcal{F}^{X^\infty}_{\frac{1}{2}}$ is trivial. Thus we see that V^n does not converge to V^∞ as $n \to \infty$. This exemplifies how the classical Wasserstein distance fails to capture the role of information in dynamic decision problems.

We now show that, if the coefficients of the SDEs (3.1) are non-Markovian, then the synchronous coupling may fail to attain the adapted Wasserstein distance between the laws of the solutions of (3.1). In this example, the diffusion coefficients will be constant and the drift coefficients will be discontinuous. This leaves open the question of whether the synchronous coupling may still be optimal for continuous non-Markovian coefficients.

Example 6.2 (non-Markovian counterexample)**.** This example already appears in [4], as a counter-example in a different setting.

Let $C \in \mathbb{R}$, $h > 0$, and define $b : [0,1] \times \Omega \to \mathbb{R}$ by $b(t,\omega) := C \operatorname{sign}(\omega_h) \mathbb{1}_{t>h}$, $t \in [0,1]$, $\omega \in \Omega$. Then let X^b be the unique strong solution of

$$\mathrm{d}X^b_t = b(t, X^b)\mathrm{d}t + \mathrm{d}W_t, \quad X^b_0 = 0,$$

and $\mu := \mathrm{Law}(X^b)$, where strong existence and pathwise uniqueness are guaranteed by Zvonkin [36]. Similarly, define $\bar{b} := -b$, and let $X^{\bar{b}}$ be the unique strong solution of

$$\mathrm{d}X^{\bar{b}}_t = \bar{b}(t, X^{\bar{b}})\mathrm{d}t + \mathrm{d}W_t, \quad X^{\bar{b}}_0 = 0,$$

and $\nu := \mathrm{Law}(X^{\bar{b}})$. Now consider the couplings

$$\begin{aligned}\pi^{\text{sync}} &:= \mathrm{Law}\left(W_\cdot + C \operatorname{sign}(W_h)[\cdot - h]_+ \ , \ W_\cdot - C \operatorname{sign}(W_h)[\cdot - h]_+\right);\\ \pi^{\text{async}} &:= \mathrm{Law}\left(W_\cdot + C \operatorname{sign}(W_h)[\cdot - h]_+ \ , \ -W_\cdot + C \operatorname{sign}(W_h)[\cdot - h]_+\right).\end{aligned}$$

Note that $\pi^{\text{sync}}, \pi^{\text{async}} \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)$. In fact, π^{sync} is the synchronous coupling between its marginals, whereas arguably the *anti-synchronous* coupling π^{async} is the opposite of the synchronous coupling (cf. the relationship between the monotone and antitone couplings between measures on $\mathbb{R}$). A few computations reveal that, for the quadratic cost, we have

$$\begin{aligned}\mathbb{E}^{\pi^{\text{sync}}}\left[\int_0^1 |\omega_t - \bar{\omega}_t|^2 \mathrm{d}t\right] &= \mathbb{E}\left[\int_h^1 (2C \operatorname{sign}(W_h)[t-h])^2 \mathrm{d}t\right] = \frac{4}{3}C(1-h)^3, \quad \text{while}\\ \mathbb{E}^{\pi^{\text{async}}}\left[\int_0^1 |\omega_t - \bar{\omega}_t|^2 \mathrm{d}t\right] &= \mathbb{E}\left[\int_0^1 (2W_t)^2 \mathrm{d}t\right] = 2.\end{aligned}$$

Choosing h sufficiently small and C sufficiently large, we have that the expected cost of the anti-synchronous coupling is strictly less than the expected cost of the synchronous coupling. Hence, the synchronous coupling does not attain the adapted Wasserstein distance $\mathcal{AW}_2(\mu,\nu)$.

As a counterpoint to the previous example, we highlight that we may still derive optimality of the synchronous coupling in certain non-Markovian cases.

Example 6.3 (kinetic SDEs)**.** Let $b, \bar{b} : \mathbb{R} \to \mathbb{R}$ be Lipschitz continuous, with Lipschitz constants C, $\bar{C}$, respectively, and consider, for $t \in [0,1]$, the kinetic equations

$$\begin{aligned} dX_t &= b\left(\int_0^t X_s \mathrm{d}s\right)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0; \\ d\bar{X}_t &= \bar{b}\left(\int_0^t \bar{X}_s \mathrm{d}s\right)\mathrm{d}t + \mathrm{d}\bar{W}_t, \quad \bar{X}_0 = 0. \end{aligned} \tag{6.1}$$

For $N \in \mathbb{N}$, $h = \frac{1}{N}$, a reasonable time-discretisation of such SDEs could be $(X_k^h)_{k=0,\ldots,N}$, $(\bar{X}_k^h)_{k=0,\ldots,N}$, where $X_0^h = 0$ and, for $k = 0, \ldots, N-1$, the increments of X^h satisfy $X_{k+1}^h - X_k^h = b\left(h \sum_{i \le k} X_i^h\right) h + W_{(k+1)h} - W_{kh}$, and similarly for $\bar{X}^h$.

It is then possible to check that, if h satisfies $C, \bar{C} < 1/h^2$, then this discretisation scheme is increasing in first order stochastic dominance. Therefore, by Proposition 3.5, the Knothe–Rosenblatt rearrangement is optimal among bicausal couplings between the laws of X^h and $\bar{X}^h$. Similarly to Lemma 3.16, by choosing a common Wiener process $W = \bar{W}$ in the definitions of the discretisations X^h and $\bar{X}^h$, we have that the Knothe–Rosenblatt rearrangement then coincides with the joint law $\mathrm{Law}(X^h, \bar{X}^h)$. Following the same arguments as in the Markovian case in Section 3, we expect to obtain the optimality of the synchronous coupling for the continuous-time problem of finding the adapted Wasserstein distance between the laws of solutions of (6.1). However, we do not carry out this analysis rigorously. Moreover, we conjecture that the synchronous coupling should be optimal between the laws of SDEs whose coefficients (say, at time t) depend on the current position (say, ω_t) and a reasonable additive functional of the path up to that time (say, $\Theta_t(\omega)$). Beyond the kinetic case (6.1), examples include $\Theta_t = t$ or, more challengingly, $\Theta_t(\omega)$ equal to the local time spent by ω on a given set up to time t.

The kinetic system (6.1) in the previous example can be made Markovian if we enlarge the state space to $\mathbb{R}^2$, by introducing the additional state variables $V_t = \int_0^t X_s \mathrm{d}s$ and $\bar{V}_t = \int_0^t \bar{X}_s \mathrm{d}s$, $t \in [0,1]$. In contrast to this example, we finally present a strong Markov two-dimensional example where the synchronous coupling is not optimal.

Example 6.4 (two-dimensional counterexample)**.** Let $A > 0$ be small and $C > 0$ large, let B, W be independent Brownian motions and consider the SDEs

$$\begin{aligned} \mathrm{d}X_t &= C\left(\mathbb{1}_{\{Y_t > -A\}} - \mathbb{1}_{\{Y_t < A\}}\right)\mathrm{d}t + \mathrm{d}B_t, \quad X_0 = 0, \\ \mathrm{d}Y_t &= \mathbb{1}_{\{|Y_t| < A\}}\mathrm{d}W_t, \quad Y_0 = 0. \end{aligned}$$

Also let $\bar{W}, \bar{B}$ be independent Brownian motions and consider the SDEs

$$\begin{aligned} \mathrm{d}\bar{X}_t &= -C\left(\mathbb{1}_{\{\bar{Y}_t > -A\}} - \mathbb{1}_{\{\bar{Y}_t < A\}}\right)\mathrm{d}t + \mathrm{d}\bar{B}_t, \quad \bar{X}_0 = 0, \\ \mathrm{d}\bar{Y}_t &= \mathbb{1}_{\{|\bar{Y}_t| < A\}}\mathrm{d}\bar{W}_t, \quad \bar{Y}_0 = 0. \end{aligned}$$

The process Y (resp. $\bar{Y}$) is a Brownian motion until hitting $-A$ or A, where it freezes. The process X (resp. $\bar{X}$) is also a Brownian motion until the aforementioned hitting time, after which a drift C or $-C$ is added depending on whether Y had hit A or $-A$ (resp. $\bar{Y}$ had hit $-A$ or A).

Suppose that $\bar{B} = B$ and $\bar{W} = W$. We call this the synchronous coupling in dimension two. Then the discrepancy $\int_0^1 |X_t - \bar{X}_t|^2 \mathrm{d}t$ is very large, since X and $\bar{X}$ have large drifts in different directions after the hitting time.

On the other hand, take $\bar{B} = B$ and $\bar{W} = -W$. This creates a small discrepancy $\int_0^1 |Y_t - \bar{Y}_t|^2 \mathrm{d}t$, but now $X_t = \bar{X}_t$, for all $t \in [0,1]$. Hence, for A sufficiently small and C sufficiently large, this second coupling has a lower L^2 cost than the synchronous coupling.

References

[1] B. Acciaio, J. Backhoff-Veraguas, and R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM Journal on Control and Optimization*, 57(6):3666–3693, 2019.

[2] B. Acciaio, J. Backhoff-Veraguas, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Process. Appl.*, 130(5):2918–2953, 2020.

[3] D. J. Aldous. Weak convergence and general theory of processes. Unpublished monograph; Department of Statistics, University of California, Berkeley, CA 94720, July 1981.

[4] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted Wasserstein distances and stability in mathematical finance. *Finance Stoch.*, 24(3):601–632, 2020.

[5] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probab. Theory Related Fields*, 178(3-4):1125–1172, 2020.

[6] J. Backhoff Veraguas, M. Beiglböck, M. Eder, and A. Pichler. Fundamental properties of process distances. *Stochastic Process. Appl.*, 130(9):5575–5591, 2020.

[7] J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

[8] D. Bartl, M. Beiglböck, and G. Pammer. The Wasserstein space of stochastic processes. *arXiv:2104.14245 [math]*, April 2021.

[9] M. Beiglböck and D. Lacker. Denseness of adapted processes among causal couplings. *arXiv:1805.03185 [math]*, May 2020.

[10] M. Beiglböck, G. Pammer, and W. Schachermayer. From Bachelier to Dupire via Optimal Transport. *arXiv:2106.12395 [q-fin]*, June 2021.

[11] J. Bion-Nadal and D. Talay. On a Wasserstein-type distance between solutions to stochastic differential equations. *Ann. Appl. Probab.*, 29(3):1609–1639, 2019.

[12] A. S. Cherny and H.-J. Engelbert. *Singular Stochastic Differential Equations*, volume 1858 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[13] A. M. G. Cox and B. A. Robinson. SDEs with no strong solution arising from a problem of stochastic control. *arXiv:2205.02519 [math]*, May 2022.

[14] S. Eckstein and G. Pammer. Computational methods for adapted optimal transport. *arXiv:2203.05005 [math]*, March 2022.

[15] M. Eder. Compactness in adapted weak topologies. *arXiv:1905.00856*, May 2019.

[16] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *J. Math. Econom.*, 25(4):443–464, 1996.

[17] D. G. Hobson. Volatility misspecification, option pricing and superreplication via coupling. *Ann. Appl. Probab.*, 8(1):193–205, 1998.

[18] S. Källblad. A dynamic programming approach to distribution-constrained optimal stopping. *The Annals of Applied Probability*, 32(3):1902–1928, 2022.

[19] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[20] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[21] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

[22] H. Knothe. Contributions to the theory of convex bodies. *Michigan Math. J.*, 4:39–52, 1957.

[23] T. Kurtz. The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities. *Electron. J. Probab*, 12:951–965, 2007.

[24] R. Lassalle. Causal transference plans and their Monge-Kantorovich problems. *Stochastic Analysis and Applications*, 36(3):452–484, 2018.

[25] G. N. Milstein, Y. M. Repin, and M. V. Tretyakov. Numerical Methods for Stochastic Systems Preserving Symplectic Structure. *SIAM Journal on Numerical Analysis*, 40(4):1583–1604, Jan. 2002.

[26] G. Pammer. A note on the adapted weak topology in discrete time. *arXiv:2205.00989*, May 2022.

[27] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012.

[28] G. C. Pflug and A. Pichler. *Multistage stochastic optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[29] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

[30] M. Rosenblatt. Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, 1952.

[31] L. Rüschendorf. The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete*, 70(1):117–129, 1985.

[32] A. V. Skorokhod. *Studies in the theory of random processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[33] D. Stroock and S. Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1979.

[34] C. Villani. *Optimal Transport. Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[35] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167, 1971.

[36] A. K. Zvonkin. A transformation of the phase space of a diffusion process that removes the drift. *Mathematics of the USSR-Sbornik*, 22(1):129–149, Feb. 1974.

Universität Wien, Vienna, Austria
Email address: julio.backhoff@univie.ac.at

KTH Royal Institute of Technology, Stochkolm, Sweden
Email address: sigrid.kallblad@math.kth.se

Universität Wien, Vienna, Austria
Email address: ben.robinson@univie.ac.at