ADAPTED WASSERSTEIN DISTANCE BETWEEN THE LAWS OF SDES

JULIO BACKHOFF-VERAGUAS[1], SIGRID KÄLLBLAD[2], AND BENJAMIN A. ROBINSON[3]

ABSTRACT. We consider the bicausal optimal transport problem between the laws of scalar time-homogeneous stochastic differential equations, and we establish the optimality of the synchronous coupling between these laws. The proof of this result is based on time-discretisation and reveals a novel 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. We complement our main results with examples showing how the optimal coupling may change in path-dependent and multidimensional settings.

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 we focus 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 (SDEs) when the space of continuous functions is equipped with the L^p-metric. We address this problem by embedding it into a class of bicausal optimal transport problems. Imposing fairly general conditions on the coefficients of the SDEs, typically amounting to time-homogeneity and mild regularity assumptions, our contribution can be summarised as follows:

  (i) characterisation of the optimal coupling, the so-called synchronous coupling, for a class of bicausal optimal transport problems, which notably includes 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 stability result for optimisers of bicausal optimal transport problems;
 (iv) a result stating that the topology induced by the adapted Wasserstein distance coincides with several topologies (including the weak topology) when restricting to SDEs whose coefficients belong to equicontinuous families;
  (v) examples illustrating the sub-optimality of the synchronous coupling for path-dependent SDEs and in higher dimensions.

[1]FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA, AUSTRIA
[2]KTH ROYAL INSTITUTE OF TECHNOLOGY, STOCKHOLM, SWEDEN
[3]DEPARTMENT OF STATISTICS, UNIVERSITY OF KLAGENFURT, AUSTRIA
E-mail addresses: julio.backhoff@univie.ac.at, sigrid.kallblad@math.kth.se, benjamin.robinson@aau.at.
Date: February 20, 2025.

A further significant contribution is the connection of two hitherto unrelated objects: the *synchronous coupling* of SDEs, which is the coupling that arises when a single Wiener process drives two SDEs; and the *Knothe–Rosenblatt* rearrangement, which is a celebrated discrete-time adapted coupling that preserves the lexicographical order. One key result that we establish is an optimality property of the Knothe–Rosenblatt rearrangement. We then argue that in a certain sense, the synchronous coupling is the continuous-time counterpart of the Knothe–Rosenblatt rearrangement.

The pioneering work of Bion-Nadal and Talay [16] contains a similar statement to our contribution (i) above; they use PDE techniques to prove optimality of the synchronous coupling for the problem of optimally controlling the correlation between one-dimensional SDEs with smooth coefficients. Our first main result establishes that, for general cost functions and path-dependent SDEs, bicausal optimal transport problems admit such a control reformulation. A posteriori, it is thus clear that [16] establishes (i) for the adapted 2-Wasserstein distance and smooth coefficients. Using probabilistic methods, we generalise the optimality result of [16] to more general cost functions and SDEs.

Adapted Wasserstein distance. We now define the adapted Wasserstein distance and give some motivation for its introduction. Let $\Omega = C([0,1], \mathbb{R})$ be the space of continuous paths from $[0,1]$ into $\mathbb{R}$ endowed with the uniform topology and corresponding Borel σ-field $\mathcal{B}(\Omega)$, and endow the product space $\Omega \times \Omega$ with the corresponding product σ-field. 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 of *couplings* between μ and ν, $\mathrm{Cpl}(\mu, \nu)$, consists of all probability measures π on $\Omega \times \Omega$ with *marginals* μ, ν; that is, $\pi(B, \Omega) = \mu(B)$ and $\pi(\Omega, B) = \nu(B)$, for all $B \in \mathcal{B}(\Omega)$. For $p \geq 1$, we write $\mathcal{P}_p$ for the set of probability measures μ on Ω with finite p^{th} moments with respect to the L^p-norm on Ω; i.e. $\int_\Omega \int_0^1 |\omega_t|^p \mathrm{d}t \, \mathrm{d}\mu < \infty$.

The classical p-Wasserstein distance $\mathcal{W}_p$ on $\mathcal{P}_p$ with respect to the L^p-norm on the underlying space Ω then takes the following form (see, for example, Villani [47, Definition 6.1]):

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

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 5.1.

As a remedy, the adapted Wasserstein distance is defined by restricting to couplings that respect the asymmetric flow of information originating from the processes. Following [6], we define bicausal couplings as follows: Let $\mathcal{F}$ and $\bar{\mathcal{F}}$ be the natural filtrations of ω and $\bar{\omega}$, respectively. For any $t \in [0,1]$ and any probability measure μ on Ω, write $\mathcal{F}_t^\mu$ for the completion[1] of $\mathcal{F}_t$ under μ. For any $\pi \in \mathrm{Cpl}(\mu, \nu)$, let $\pi_\omega(\mathrm{d}\bar{\omega})$ be the regular disintegration kernel for which $\pi(\mathrm{d}\omega, \mathrm{d}\bar{\omega}) = \mu(\mathrm{d}\omega)\pi_\omega(\mathrm{d}\bar{\omega})$.

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

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

The set of *bicausal couplings* $\mathrm{Cpl}_{bc}(\mu, \nu)$ consists of all $\pi \in \mathrm{Cpl}_c(\mu, \nu)$ with $S_\#\pi \in \mathrm{Cpl}_c(\nu, \mu)$, where $S(\omega, \bar{\omega}) = (\bar{\omega}, \omega)$, for all $(\omega, \bar{\omega}) \in \Omega \times \Omega$.

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.[2] With the

---

[1] I.e. $\mathcal{F}_t^\mu$ is the sigma-algebra generated by $\mathcal{F}_t$ and the null sets for μ.

[2] The intuition behind the concept of causality is perhaps most easily grasped in a discrete-time setup, i.e. when considering a finite set of time points, say $\{1, 2, \ldots, N\}$. The defining property of causality can then be phrased as requiring, with obvious adaptation of notation, that $\pi((\bar{\omega}_1, ..., \bar{\omega}_n) \in A | \omega_1, ..., \omega_N) = \pi((\bar{\omega}_1, ..., \bar{\omega}_n) \in A | \omega_1, ..., \omega_n)$,

above notation at hand, we are now ready to define the *adapted Wasserstein distance* $\mathcal{AW}_p$, $p \geq 1$:

$$(1.2) \qquad \mathcal{AW}_p^p(\mu, \nu) := \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)} \mathbb{E}^\pi \left[\int_0^1 |\omega_t - \bar{\omega}_t|^p \mathrm{d}t \right], \qquad \mu, \nu \in \mathcal{P}_p.$$

To give an historical account, the condition of causality can be traced back, at least, to the work on existence of solutions of SDEs by Yamada and Watanabe [48]; it has also appeared under the name of *compatibility* in Kurtz [32]. The concept was recently popularised and studied in a continuous-time framework by Lassalle [33], and systematically investigated for discrete-time processes using dynamic programming arguments in [5] (see also [27] for a recursive approach to a closely related optimal stopping problem). We refer to Beiglböck and Lacker [10] for further historical remarks and for an account of the connections to the filtering literature. We also refer to Beiglböck and Lacker [10], Beiglböck, Pammer, and Schrott [14] for a detailed exposition of how Monge maps relate to general transport plans in the presence of causality constraints.

To the best of our knowledge, the symmetric condition of bicausality first appeared in the work of Rüschendorf [44]; for a more recent account we refer again to [5]. A distance based on the bicausality condition was independently introduced and studied, under the name of *nested distance*, in a series of papers by Pflug and Pichler; see, for example, [38] and [39] 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 [23] both rely on incorporating the distance between certain conditional disintegration kernels of the processes. See also [7], where it was shown that these different distances all generate the same topology.

In continuous time, for diffusion processes, the modified Wasserstein distance was introduced in [16]; we show herein that this distance coincides with the adapted Wasserstein distance. For general continuous semi-martingales, motivated by financial applications, an adapted Wasserstein distance was defined in [6] with respect to a cost function that compares the drift and martingale parts of the Doob–Meyer decomposition separately. We refer to [1, 2] for further studies of adapted distances in continuous time; see also [5, Section 2] for an exposition of the related literature within mathematical finance.

Optimality of the synchronous coupling. Throughout this article, we consider the laws of solutions of SDEs of the following type:

$$(1.3) \qquad \mathrm{d}X_t^{b,\sigma} = b(X_t^{b,\sigma})\mathrm{d}t + \sigma(X_t^{b,\sigma})\mathrm{d}W_t, \quad X_0^{b,\sigma} = x_0, \quad t \in [0,1],$$

where $b \colon \mathbb{R} \to \mathbb{R}$ and $\sigma \colon \mathbb{R} \to \mathbb{R}_+ = [0, \infty)$ are some measurable functions such that a unique strong solution $X^{b,\sigma}$ exists; we write $\mu^{b,\sigma} := \mathrm{Law}(X^{b,\sigma})$ for the induced probability measure on Ω. We suppose that all SDEs are equipped with the same initial condition, $x_0 \in \mathbb{R}$, and omit it from the notation.

Given two such measures, $\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}}$, we can couple them as follows: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting a Wiener process W, and let $X^{b,\sigma}, X^{\bar{b},\bar{\sigma}}$ be the solutions of (1.3) with coefficients (b, σ) and $(\bar{b}, \bar{\sigma})$, respectively, when both SDEs are driven by W. In this way, we define a bicausal coupling $\mathbb{P} \circ (X^{b,\sigma}, X^{\bar{b},\bar{\sigma}})^{-1} \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$. This coupling plays 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 $b \colon \mathbb{R} \to \mathbb{R}$ and $\sigma \colon \mathbb{R} \to \mathbb{R}_+$ in (1.3) are continuous, have linear growth, and are such that pathwise uniqueness holds for (1.3).

---

for all $A \in \mathcal{B}(\mathbb{R}^n)$, $n = 1, ..., N$. In such a discrete-time setting, if the coupling is further supported on the graph of a function, say $\varphi \colon \mathbb{R}^N \to \mathbb{R}^N$ (i.e. a Monge map), then causality amounts to $\varphi(x_1, ..., x_N) = (\varphi_1(x_1), \varphi_2(x_1, x_2), \ldots, \varphi_N(x_1, \ldots, x_N))$, for some functions $\varphi_n \colon \mathbb{R}^n \to \mathbb{R}$, $n = 1, ..., N$.

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* (1.2) *defining* $\mathcal{AW}_p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$.

Our main result notably establishes the optimality of the synchronous coupling not only for the adapted Wasserstein distance, but also for the *bicausal optimal transport* problem with respect to a more general class of cost functions (see Theorem 3.24); we also provide this conclusion under a different set of assumptions allowing for the drift coefficients to be discontinuous (see Proposition 3.30). The follow-up work [41] further extends the main result in the direction of SDEs with irregular coefficients.

For one-dimensional SDEs with sufficiently regular coefficients, the optimality of the synchronous coupling was first established in [16] for the so-called modified Wasserstein distance, which is the distance obtained when optimising the cost in (1.2) over couplings induced by solutions of a pair of SDEs (1.3) driven by *correlated* Wiener processes. It then follows from the above result combined with [16, Section 2.1] that the adapted and modified Wasserstein distances coincide. A crucial part of our analysis is the a priori reformulation of the adapted Wasserstein distance in terms of an associated control problem (where one controls the degree of correlation); a similar result holds for a bicausal optimal transport problem with a general cost function when the marginals are given by possibly path-dependent SDEs (see Proposition 2.2).

In [16], the authors take a stochastic control approach and their proofs rely on a verification argument for the associated Hamilton–Jacobi–Bellman (HJB) equation. While such stochastic control arguments provide the natural continuous-time analogue of the recursive arguments used to prove optimality of the Knothe–Rosenblatt rearrangement in discrete time (see Remark 3.10), the use of classical solutions of the HJB equation, as employed in [16], inevitably requires the cost function as well as the coefficients of the SDEs to be smooth enough for the associated stochastic flows to be differentiable. Here, we rather take a probabilistic approach to prove the optimality of the synchronous coupling. This approach enables us to relax the assumptions on the cost function and the coefficients and establish this optimality property in its natural generality.

We clarify that, while the analysis in [16] also pertains to multidimensional diffusions, the authors only identify an explicit optimiser in dimension one. In fact, the optimality of the synchronous coupling does not generally extend to higher dimensions. Our Examples 5.4 and 5.5 demonstrate the sub-optimality of the synchronous coupling for certain multidimensional SDEs. We also note that the discrete approximation methods described below rely on one-dimensional results from optimal transport; see Remark 3.6.

Discrete approximation methods and stability. Consider now the problem of optimally coupling the laws of one-dimensional discrete-time stochastic processes with $n \in \mathbb{N}$ time steps, i.e. coupling probability measures on $\mathbb{R}^n$. A key object of study in this paper is the *Knothe–Rosenblatt rearrangement*, introduced in [42] and [31], which generalises the classical monotone rearrangement to the laws of such discrete-time processes; see Figure 1 for an illustration. When restricting to bicausal couplings and imposing certain monotonicity properties on the marginal laws, it turns out that the Knothe–Rosenblatt rearrangement is optimal for an L^p-cost. For a two-step discrete bicausal optimal transport problem, this was first shown by Rüschendorf [44]. In [5], based on a recursive argument, the result was then generalised to a multi-stage discrete problem with Markov marginal laws. We generalise this result to an even broader class of cost functions and marginal laws, and we link the assumptions on the marginals to the notion of stochastic monotonicity; see Proposition 3.5. We emphasise that the dimension $n \in \mathbb{N}$ here represents the number of time steps of the discrete-time process and that the state space is necessarily one-dimensional; see Remark 3.6.

This discrete-time optimality result underpins our analysis in continuous time; applying this result to a carefully chosen discretisation of the SDEs, we deduce the optimality of the synchronous coupling. Indeed, 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. In order to carry out the above procedure, for the class of bicausal transport problems that we study, we establish a stability result for optimal couplings (see Proposition 2.6).
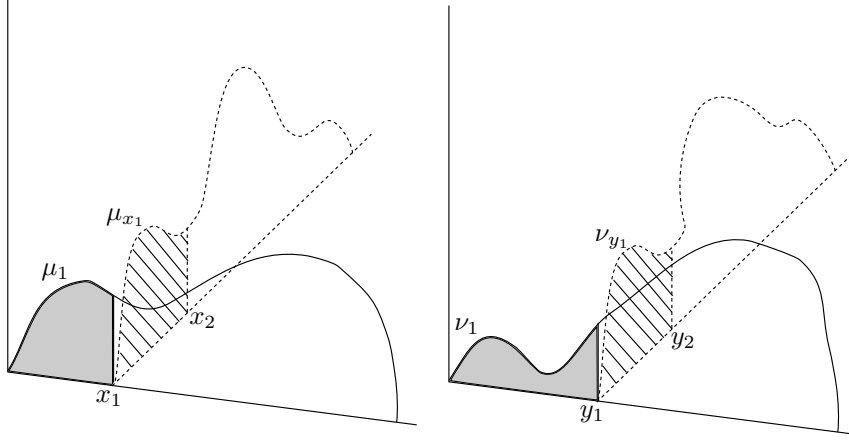

FIGURE 1. Illustration of the Knothe–Rosenblatt rearrangement in two dimensions. The first marginals of μ, ν are denoted μ_1, ν_1, and the conditional distributions by μ_{x_1}, ν_{y_1}. Similarly shaded regions have the same area.

We further obtain approximation results in the adapted Wasserstein distance for bicausal couplings between the laws of SDEs. Given coefficients (b, σ), $(\bar{b}, \bar{\sigma})$ and $n \in \mathbb{N}$, consider the *Euler–Maruyama scheme* $(X^n, \bar{X}^n)$, which is given by $(X_0^n, \bar{X}_0^n) = (x_0, x_0)$ and, for $h = 1/n$, $k = 0, ..., n-1$,

$$(1.4) \quad \begin{cases} X_t^n &= X_{kh}^n + b(X_{kh}^n)(t - kh) + \sigma(X_{kh}^n)(W_t - W_{kh}) \\ \bar{X}_t^n &= \bar{X}_{kh}^n + \bar{b}(\bar{X}_{kh}^n)(t - kh) + \bar{\sigma}(\bar{X}_{kh}^n)(\bar{W}_t - \bar{W}_{kh}) \end{cases}, \quad t \in (kh, (k+1)h].$$

With the adapted Wasserstein distance defined analogously to (1.2) for marginal distributions on $\Omega \times \Omega$, we have the following result:

Theorem 1.4. *Let* $b, \bar{b} \colon \mathbb{R} \to \mathbb{R}$, $\sigma, \bar{\sigma} \colon \mathbb{R} \to \mathbb{R}_+$ *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 (1.4) satisfies* $\mathcal{AW}_p(\mathrm{Law}(X^n, \bar{X}^n), \pi) \xrightarrow{n \to \infty} 0$, $p \geq 1$.

The above result holds also for SDEs with path-dependent coefficients (see Theorem 2.11). To obtain our main results, we will work with a modification of the Euler–Maruyama scheme in which the increments of the Wiener process are truncated; cf. Liu and Pagès [35], Milstein, Repin, and Tretyakov [36]. Adapting Theorem 1.4 for this modified scheme, we will prove convergence of the Knothe–Rosenblatt rearrangement to the synchronous coupling in a sense that is made precise in Proposition 3.21.

While these results on discretisation and stability are crucial for our analysis, they are also of independent interest. For instance, our time-discretisation method suggests a possible approach to numerical approximation of adapted Wasserstein distances between the laws of SDEs; however, we do not explore this direction further. For existing numerical methods for computing adapted

Wasserstein distances, we refer to Eckstein and Pammer [20], Pichler and Weinhardt [40], and the references therein.

The synchronous distance and the associated topology. We also study the topology induced by the adapted Wasserstein distance $\mathcal{AW}_p$. In particular, we investigate the relationship between the topologies induced by different distances on the spaces $\mathcal{P}_p$ of laws of continuous-time stochastic processes.

The classical Wasserstein distance (1.1) metrises the usual weak topology on the space $\mathcal{P}_p$ (see, e.g. [46, Theorem 7.12]). 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 (bicausal) Wasserstein distance, by replacing $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ with $\mathrm{Cpl}_{\mathrm{c}}(\mu, \nu)$ in (1.2). 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* $\mathcal{SW}_p$, defined by

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

where $X^{b,\sigma}, X^{\bar{b},\bar{\sigma}}$ are the p-integrable solutions of the SDE (1.3), with coefficients $(b, \sigma), (\bar{b}, \bar{\sigma})$, 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.

We show that all the distances discussed above induce the same topology when restricted to solutions of the SDE (1.3) for which the coefficients belong to the following set:

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

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.*

The above topologies also 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 $\sup_{t \in [0,1]} |\omega_t - \bar{\omega}_t|^p$.

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.

Note that, in contrast to the optimality results presented thus far, the topological equivalence of Theorem 1.5 carries over to higher dimensions; see Remark 4.2.

In discrete time, Beiglböck, Pammer, and Posch [13] studied a distance induced by the Knothe–Rosenblatt rearrangement and showed that it is topologically equivalent to the adapted Wasserstein distance. Our Theorem 1.5 can thus be seen as a continuous-time analogue of [13, Theorem 1.4], with the synchronous coupling in place of the Knothe–Rosenblatt rearrangement. For discrete-time processes, [7] also established that the same topology is generated by the adapted Wasserstein distance, the nested distance, and the distances introduced by Aldous [3] and Hellwig [23]; we refer to [8, 21] for further properties of this common topology. In fact, an even stronger result is true: Bartl, Beiglböck, and Pammer [9] and Pammer [37] go beyond the convention of identifying a process with its law and instead consider processes equipped with a filtration. It is shown that

all topologies that are strong enough to encode the information of the filtration still coincide, even in this generalised discrete-time setting. In a continuous-time setup, Bion-Nadal and Talay [16, Propositions 1.8, 1.9] give a result in a similar spirit to Theorem 1.5 for their modified Wasserstein distance. We note that our proof is remarkably simple as it is a straightforward application of stability results for SDEs.

Structure of the article. In Section 2, we establish the equality of the modified and adapted Wasserstein distances; we also provide a stability result and prove Theorem 1.4. We give our optimality result for the Knothe–Rosenblatt rearrangement in Section 3.1. In Section 3.2, we introduce a variation of the classical Euler–Maruyama scheme and use it to prove Theorem 1.3 under an additional Lipschitz assumption. We complete the proof of Theorem 1.3 in Section 3.3. In Section 3.4, we establish a variation of Theorem 1.3 which allows for discontinuities in the drift coefficients. In Section 4, we prove Theorem 1.5 on the equality of topologies. In Section 5, we collect some examples, the first of which motivates the introduction of the adapted Wasserstein distance. We then present counterexamples to the optimality of the synchronous coupling in non-Markovian and higher dimensional settings, as well as a counterexample to the optimality of the Knothe–Rosenblatt rearrangement for a particular choice of cost function. Various approximation and stability results are deferred to the appendix.

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

Throughout this paper we work in dimension one. Define the space of continuous paths $\Omega \coloneqq C([0,1],\mathbb{R})$ equipped with the uniform topology and corresponding Borel sigma-field $\mathcal{F} \coloneqq \mathcal{B}(\Omega)$. We also equip Ω with the canonical filtration $(\mathcal{F}_s)_{s\in[0,1]}$ and note that $\mathcal{F} = \mathcal{F}_1$.

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 $W, \bar{W}$. We say that the two-dimensional process $(W, \bar{W})$ is a *ρ-correlated Wiener process* if the cross-variation satisfies

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

for some $\mathcal{B}([0,1]) \otimes \mathcal{G}$-progressively measurable function $\rho\colon [0,1] \times \Omega \times \Omega \to [-1,1]$, which is then unique in the $(d\mathbb{P} \times dt)$-sense. We say that $(W, \bar{W})$ is a correlated Wiener process if it is a ρ-correlated Wiener process for some ρ.

Given $x_0, \bar{x}_0 \in \mathbb{R}$ and path-dependent coefficients $b, \bar{b}\colon [0,1] \times \Omega \to \mathbb{R}$, $\sigma, \bar{\sigma}\colon [0,1] \times \Omega \to \mathbb{R}_+$, which are progressively measurable, consider the system of SDEs

$$(2.1) \qquad \begin{cases} X_t = x_0 + \int_0^t b(s, X)\mathrm{d}s + \int_0^t \sigma(s, X)\mathrm{d}W_s, & t \in [0,1], \\ \bar{X}_t = \bar{x}_0 + \int_0^t \bar{b}(s, \bar{X})\mathrm{d}s + \int_0^t \bar{\sigma}(s, \bar{X})\mathrm{d}\bar{W}_s, & t \in [0,1]. \end{cases}$$

Suppose that there exists some correlated Wiener process $(W, \bar{W})$ and a $\mathcal{G}$-adapted process $(X, \bar{X})$ (cf. Definition 2.1) that satisfies (2.1). In this case, we say that $(X, \bar{X})$ is a *strong solution of the system* (2.1) driven by the correlated Wiener process $(W, \bar{W})$. Writing μ, ν for the marginal laws of $X, \bar{X}$, we denote the set of couplings of the form $\mathrm{Law}(X, \bar{X})$ by $\widetilde{\mathrm{Cpl}}(\mu, \nu)$.

2.1. Characterisation of bicausal couplings between SDEs.

Proposition 2.2. *Suppose that there exist unique strong solutions to the SDEs* (2.1) *and write* μ, ν *for their respective laws. Then the set of bicausal couplings* $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ *is equal to the set of couplings* $\widetilde{\mathrm{Cpl}}(\mu, \nu)$.

Proof. Suppose that $\tilde{\pi} \in \widetilde{\mathrm{Cpl}}(\mu, \nu)$. Then $\tilde{\pi} \in \mathrm{Cpl}(\mu, \nu)$. Moreover, there is some correlated Wiener process $(W, \bar{W})$ such that $\tilde{\pi} = \mathrm{Law}(X, \bar{X})$, where $(X, \bar{X})$ is a strong solution of the system (2.1) driven by $(W, \bar{W})$. Since $(X, \bar{X})$ is adapted to the completed filtration of $(W, \bar{W})$, we conclude that $\tilde{\pi}$ is bicausal.

To show the converse, suppose now that $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$. Let $(\omega, \bar{\omega})$ be the canonical process on the product space, and let $\mathcal{F}, \bar{\mathcal{F}}$ denote the canonical filtrations corresponding to $\omega, \bar{\omega}$, respectively. Note that, by definition of the measure μ, we may define a continuous process M such that, under π, M is an $\mathcal{F}^\mu$-martingale with quadratic variation $\langle M \rangle$ given by $\langle M \rangle_t = \int_0^t \sigma(s, \omega)^2 \mathrm{d}s$, for all $t \in [0, 1]$, and

$$(2.2) \qquad \omega_t = x_0 + \int_0^t b(s, \omega)\mathrm{d}s + M_t, \quad \text{for all } t \in [0, 1].$$

Since the coupling π is bicausal, we have the following independence property: for any $s \in [0, 1]$, conditional on $\mathcal{F}_s^\mu$, $(\omega_t)_{t \in [0,1]}$ is independent of $\bar{\mathcal{F}}_s$ under π. Therefore, for any $s, t \in [0, 1]$ with $s < t$, we have

$$\mathbb{E}^\pi \big[M_t - M_s \mid \mathcal{F}_s^\mu \otimes \bar{\mathcal{F}}_s \big] = \mathbb{E}^\pi \big[M_t - M_s \mid \mathcal{F}_s^\mu \big] = 0,$$

and by the tower property of conditional expectation, $\mathbb{E}^\pi \big[M_t - M_s \mid \mathcal{F}_s \otimes \bar{\mathcal{F}}_s \big] = 0$. Thus M is also an $\mathcal{F} \otimes \bar{\mathcal{F}}$-martingale under π.

Using the symmetry of the definition of bicausality, we find that $\bar{\omega}$ admits an analogous representation to (2.2) under π for some continuous $\mathcal{F} \otimes \bar{\mathcal{F}}$-martingale $\bar{M}$.

Enlarging the probability space as necessary, introduce a Wiener process $\hat{W}$ that is independent of M and $\bar{M}$ under π. Define processes $W, \bar{W}$ via

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

and let $\mathcal{G}$ denote the completion of the natural filtration of $(W, \bar{W})$ under π. Then $(\omega, \bar{\omega})$ satisfies (2.1) driven by $(W, \bar{W})$. Further, by the bicausality of π and the independence of $\hat{W}$, both W and $\bar{W}$ are $\mathcal{G}$-martingales under π. It then follows from Lévy's characterisation that both W and $\bar{W}$ are $\mathcal{G}$-Wiener processes. By the Kunita–Watanabe inequality (see, e.g. [34, Proposition 4.18]), we have that $\mathrm{d}\langle W, \bar{W} \rangle$ is almost surely absolutely continuous with respect to Lebesgue measure, and so there exists a $\mathcal{B}([0, 1]) \otimes \mathcal{G}$-progressively measurable function $\rho \colon [0, 1] \times \Omega \times \Omega \to [-1, 1]$ such that $\rho_t \mathrm{d}t = \mathrm{d}\langle W, \bar{W} \rangle_t$, for all $t \in [0, 1]$. Therefore $(W, \bar{W})$ is a ρ-correlated Wiener process, as in Definition 2.1.

By assumption, the SDE with coefficients b, σ 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 $\tilde{X} = \omega$ under π and, in particular, ω is adapted to the natural filtration of W and hence to $\mathcal{G}$. The analogous statement applies to $\bar{\omega}$, and so we have that the canonical process $(\omega, \bar{\omega})$ under the bicausal coupling π is a strong solution of the system (2.1) with respect to $(W, \bar{W})$, as required. $\qquad \square$

Remark 2.3 (modified Wasserstein distance)**.** For any $p \geq 1$, consider the subset $\bar{\mathcal{P}}_p \subset \mathcal{P}_p$ consisting of laws of SDEs of the form (2.1). On this set, one can define a metric $\bar{\mathcal{W}}_p$ in the same way as the adapted Wasserstein distance $\mathcal{AW}_p$ is defined in (1.2), but with the set $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ of bicausal couplings between $\mu, \nu \in \bar{\mathcal{P}}_p$ replaced by $\widetilde{\mathrm{Cpl}}(\mu, \nu)$. According to Proposition 2.2, for $\mu, \nu \in \bar{\mathcal{P}}_p$, we have $\bar{\mathcal{W}}_p(\mu, \nu) = \mathcal{AW}_p(\mu, \nu)$.

For $p = 2$, restricting to elements of $\bar{\mathcal{P}}_2$ that are laws of time-homogeneous SDEs of the form (1.3), $\bar{\mathcal{W}}_2$ thus defined is the *modified Wasserstein distance* introduced by Bion-Nadal and Talay in [16].

Remark 2.4. In our definition of (bi)causal couplings we make use of the canonical filtration, while in [6] 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 [29, 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 [29, Ch. 5, Theorem 4.20]. In Example 5.1, we also encounter a situation where the above sets of couplings coincide, although one of the marginal processes is not strong Markov.

2.2. **Stability of bicausal optimisers.** Our strategy for proving our main results in this paper is to approximate our problem of interest by problems for which the optimiser is known. A key ingredient in this approach is the stability of optimisers. We here establish such a stability result for a more general class of *bicausal optimal transport problems* obtained by replacing the p-norm in the definition (1.2) of $\mathcal{AW}_p$ with a general *cost function* $c\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$; we typically assume that there exists some $K > 0$ such that

$$(2.3) \qquad\qquad |c(x,y)| \leq K[1 + |x|^p + |y|^p], \quad \text{for all } x, y \in \mathbb{R}.$$

Since it will be useful to allow for approximation by processes that are not necessarily continuous, we take $\hat{\Omega} = \mathbb{D}([0,1], \mathbb{R})$ to be the Polish space of càdlàg functions equipped with the Skorokhod topology and corresponding Borel σ-field $\hat{\mathcal{F}}$. We define product spaces, the canonical process, the set of p-integrable probability measures, and the set of couplings in total analogy to the continuous case. In particular, for $p \geq 1$ and $\pi, \pi' \in \mathcal{P}(\hat{\Omega} \times \hat{\Omega})$ with finite p^{th} moment, in analogy to (1.1), the p-Wasserstein distance is given by

$$(2.4) \qquad\qquad \mathcal{W}_p^p(\pi, \pi') = \inf_{\alpha \in \mathrm{Cpl}(\pi,\pi')} \mathbb{E}^\alpha \left[\int_0^1 |(\omega_t, \bar{\omega}_t) - (\omega'_t, \bar{\omega}'_t)|^p \mathrm{d}t \right],$$

where $\mathrm{Cpl}(\pi, \pi')$ denotes the set of probability measures on $(\hat{\Omega} \times \hat{\Omega}) \times (\hat{\Omega} \times \hat{\Omega})$ with marginal distribution onto the first (resp. last) two coordinates given by π (resp. π'), and $((\omega, \bar{\omega}), (\omega', \bar{\omega}'))$ denotes the corresponding canonical process.

Remark 2.5. Although we consider Wasserstein distances with respect to an L^p-metric, we defined the Borel σ-fields $\mathcal{F}$ and $\hat{\mathcal{F}}$ on Ω and $\hat{\Omega}$ with respect to the uniform topology and Skorokhod topology, respectively. We now verify that these Borel σ-fields are equal to the Borel σ-fields corresponding to the topology of L^p-convergence. Indeed, both $\mathcal{F}$ and $\hat{\mathcal{F}}$ are generated by the coordinate mapping and are thus included in the respective Borel σ-field for L^p-convergence; see, for example, [15, Example 1.3 and Theorem 12.5]. On the other hand, convergence in the uniform topology (resp. Skorokhod topology) implies L^p-convergence on Ω (resp. $\hat{\Omega}$) and so we have equality of the Borel σ-fields.

We start with an auxiliary result.

Proposition 2.6. *Let $p \geq 1$. Suppose that there exist unique strong solutions to the SDEs (2.1) and write μ, ν for their respective laws. Let $X^n, \bar{X}^n \colon \Omega \to \hat{\Omega}$, $n \in \mathbb{N}$, be measurable and suppose that for any ρ-correlated Wiener process $(W, \bar{W})$,*

$$(2.5) \qquad\qquad \mathcal{W}_p(\pi^{\rho,n}, \pi^\rho) \xrightarrow{n \to \infty} 0,$$

where $\pi^{\rho,n} = \mathrm{Law}(X^n \circ W, \bar{X}^n \circ \bar{W})$ and $\pi^\rho = \mathrm{Law}(X, \bar{X})$ with $X, \bar{X}$ solving (2.1) with respect to $(W, \bar{W})$. Let $c\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and satisfy (2.3) for some $K > 0$ and suppose that there exists some $\hat{\rho}\colon [0,1] \times \Omega \times \Omega \to [-1,1]$ such that for each $n \in \mathbb{N}$, $\pi^{\hat{\rho},n}$ attains

$$\inf_{\pi \in \widetilde{\mathrm{Cpl}}(\mu^n, \nu^n)} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t\, \mathrm{d}\pi,$$

where $\widetilde{\mathrm{Cpl}}(\mu^n, \nu^n)$ denotes all couplings of the form $\pi^{\rho,n}$ for some correlation process ρ (writing μ^n, ν^n for their marginals). Then

$$(2.6) \qquad \lim_{n \to \infty} \inf_{\pi \in \widetilde{\mathrm{Cpl}}(\mu^n, \nu^n)} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \, \mathrm{d}\pi = \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \, \mathrm{d}\pi;$$

moreover, the right hand side is attained by $\pi^{\hat{\rho}}$.

Proof. By assumption, there exists some $\hat{\rho}$-correlated Wiener process such that, for each $n \in \mathbb{N}$, and for any admissible correlation process ρ,

$$(2.7) \qquad \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \mathrm{d}\pi^{\hat{\rho},n} \leq \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \mathrm{d}\pi^{\rho,n}.$$

We now argue that the functional $(x(\cdot), y(\cdot)) \mapsto \int_0^1 c(x(t), y(t)) \mathrm{d}t$ is continuous with at most polynomial growth of order p on with respect to the L^p-metric on $\hat{\Omega} \times \hat{\Omega}$. The growth claim is immediate since

$$\left| \int_0^1 c(x(t), y(t)) \mathrm{d}t \right| \leq \int_0^1 K(1 + |x(t)|^p + |y(t)|^p) \mathrm{d}t = K(1 + \|x\|_{L^p}^p + \|y\|_{L^p}^p).$$

Suppose now that $\|x^n - x\|_{L^p} + \|y^n - y\|_{L^p} \to 0$. The sequences $\{|x^n(\cdot)|^p\}_n$ and $\{|y^n(\cdot)|^p\}_n$ are then uniformly integrable w.r.t. Lebesgue measure on $[0, 1]$. By the growth assumption on c, the sequence $\{|c(x^n(\cdot), y^n(\cdot))|\}_n$ is similarly uniformly integrable. Define $a_n := \int_0^1 c(x^n(t), y^n(t)) \mathrm{d}t$. For any subsequence $\{a_{n_k}\}_k$ we can find a sub-subsequence $\{a_{n_{k_j}}\}_j$ such that $c(x^{n_{k_j}}(\cdot), y^{n_{k_j}}(\cdot)) \to c(x(\cdot), y(\cdot))$ almost surely, as c is continuous. Then by uniform integrability we infer that $a_{n_{k_j}} \to a := \int_0^1 c(x(t), y(t)) \mathrm{d}t$. As the limit does not depend on the subsequence, we conclude that $a_n \to a$.

Equipping $\hat{\Omega}$ with the topology induced by the L^p-metric, it follows from Remark 2.5 that the product space $\hat{\Omega} \times \hat{\Omega}$ is a separable Radon space. Given this property, the continuity and polynomial growth shown above, and the $\mathcal{W}_p$-convergence (2.5), we can apply [4, Lemma 5.1.7 and Proposition 7.1.5] to take the limit in (2.7) as $n \to \infty$, and we deduce that, for any admissible correlation process ρ,

$$\iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \mathrm{d}\pi^{\hat{\rho}} \leq \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \mathrm{d}\pi^{\rho}.$$

By Proposition 2.2, $\pi^{\hat{\rho}}$ thus attains the infimum on the right-hand side of (2.6). Once again using the convergence of $\pi^{\hat{\rho},n}$ to $\pi^{\hat{\rho}}$ in $\mathcal{W}_p$, together with the optimality of these couplings for their respective transport problems, we conclude that the value of the problem also converges. $\square$

We now make use of the above result to obtain two stability results where the assumptions are placed directly on the marginals; they will be crucial for our upcoming analysis. The first one allows for càdlàg approximations.

Corollary 2.7. *Suppose that there exist unique strong solutions $X, \bar{X}$ to the SDEs (2.1). Let $p \geq 1$ and $X^n, \bar{X}^n \colon \Omega \to \hat{\Omega}$, $n \in \mathbb{N}$, be measurable and such that $X^n \circ W$ (resp. $\bar{X}^n \circ \bar{W}$) converges in L^p to X (resp. $\bar{X}$). Let $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and satisfy (2.3), and suppose that there exists some $\hat{\rho}$ as in Proposition 2.6. Then condition (2.5) is satisfied and the conclusions of Proposition 2.6 hold true.*

Proof. For any ρ-correlated Wiener process $(W, \bar{W})$, let $(X, \bar{X})$ be the solution of the SDE (2.1) with respect to $(W, \bar{W})$ and define $(X^n, \bar{X}^n) = (X^n \circ W, \bar{X}^n \circ \bar{W})$. Then $\mathrm{Law}((X^n, \bar{X}^n), (X, \bar{X})) \in \mathrm{Cpl}(\pi^{\rho,n}, \pi^{\rho})$. Moreover, by the assumption of L^p-convergence,

$$\mathbb{E}\left[\int_0^1 |(X_t^n, \bar{X}_t^n) - (X_t, \bar{X}_t)|^p \mathrm{d}t \right]^{\frac{1}{p}} \leq \mathbb{E}\left[\int_0^1 |X_t^n - X_t|^p \mathrm{d}t \right]^{\frac{1}{p}} + \mathbb{E}\left[\int_0^1 |\bar{X}_t^n - \bar{X}_t|^p \mathrm{d}t \right]^{\frac{1}{p}} \xrightarrow{n \to \infty} 0,$$

which verifies (2.5) and thus completes the proof. $\square$

Our second stability result enables us to approximate the adapted Wasserstein distance between laws of SDEs by approximating their coefficients. For this result we return to the set-up of continuous paths in Ω; we write $\|\omega\|_\infty := \sup_{s\in[0,1]} |\omega_s|$, $\omega \in \Omega$, for the sup-norm and work under the following assumption.

Assumption 2.8. Suppose that $x_0, \bar{x}_0 \in \mathbb{R}$ and $b, \bar{b} \colon [0,1] \times \Omega \to \mathbb{R}$, $\sigma, \bar{\sigma} \colon [0,1] \times \Omega \to \mathbb{R}_+$ 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 $t \in [0,1]$, $\omega \in \Omega$,

$$(2.8) \qquad |b(t,\omega)| \vee |\bar{b}(t,\omega)| \vee |\sigma(t,\omega)| \vee |\bar{\sigma}(t,\omega)| \leq K(1 + \|\omega\|_\infty);$$

(iv) there exist unique strong solutions of the SDEs (2.1).

Corollary 2.9. *Let $(x_0, \bar{x}_0, b, \bar{b}, \sigma, \bar{\sigma})$ satisfy Assumption 2.8 and write μ, ν for the laws of $X, \bar{X}$. For $n \in \mathbb{N}$, consider $(x_0^n, \bar{x}_0^n, b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)$ satisfying Assumption 2.8.(i) and (iii), with a uniform slope constant K in (2.8), and such that strong existence holds for (2.1); write $X^n, \bar{X}^n$ for a pair of solutions and μ^n, ν^n for their laws. Let $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and satisfy (2.3), and suppose that there exists some $\hat{\rho}$ as in Proposition 2.6.*

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

$$(2.9) \qquad \|\omega^n - \omega\|_\infty \to 0 \implies (b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)(t, \omega^n) \to (b, \bar{b}, \sigma, \bar{\sigma})(t, \omega), \text{ for each } t \in [0,1].$$

Then condition (2.5) is satisfied and the conclusions of Proposition 2.6 hold true.

Proof. Let ρ be an admissible correlation process. Write $\pi^{\rho,n}$ for the joint law of $(X^n, \bar{X}^n)$, for $n \in \mathbb{N}$, and π^ρ for the joint law of $(X, \bar{X})$, when the corresponding SDEs are driven by a ρ-correlated Wiener process. By Proposition B.2, we have convergence of $\pi^{\rho,n}$ to π^ρ in the p-Wasserstein distance on $\mathcal{P}(\Omega \times \Omega)$ with respect to the sup-norm. Embedding $\mathcal{P}(\Omega \times \Omega)$ into $\mathcal{P}(\hat{\Omega} \times \hat{\Omega})$, we thus have convergence of $\pi^{\rho,n}$ to π^ρ in the p-Wasserstein distance on $\mathcal{P}(\hat{\Omega} \times \hat{\Omega})$ with respect to the L^p-norm (as defined in (2.4)), which verifies (2.5) and thus completes the proof. $\qquad\square$

Remark 2.10. Whenever $(W, \bar{W})$ is a ρ-correlated Wiener process defined on some stochastic basis, under Assumption 2.8 (iv), one can uniquely construct a strong solution $(X, \bar{X})$ of the system (2.1) driven by $(W, \bar{W})$ on the same stochastic basis. We also note that Assumption 2.8 (iv) can be weakened to pathwise uniqueness only. Indeed, weak existence is guaranteed already by Assumption 2.8 (i)–(iii) and a classical result of Skorokhod (e.g. adapting the proof of [28, Theorem 21.9] to coefficients with linear growth) and so the Yamada-Watanabe criterion [28, Lemma 21.17] applies. Note finally that continuity of $b, \bar{b}, \sigma, \bar{\sigma}$ is implied by the convergence (2.9) and so we could drop Assumption 2.8 (ii) from our assumptions. We keep this assumption, however, to make it transparent that the stability result of Corollary 2.9 cannot be applied to coefficients with discontinuities.

2.3. **Approximation of SDEs in adapted Wasserstein distance.** We next make use of Proposition 2.2 to prove a more general version of Theorem 1.4 on the approximation of the laws of SDEs in adapted Wasserstein distance. In Proposition 3.21, we will apply this result to the convergence of the Knothe–Rosenblatt rearrangement to the synchronous coupling. For $p \geq 1$, and $\pi, \pi' \in \mathcal{P}(\Omega \times \Omega)$ with finite p^{th} moment, we define the adapted p-Wasserstein distance between π and π' analogously to (1.2); specifically, defining the set of bicausal couplings, $\mathrm{Cpl}_{\mathrm{bc}}(\pi, \pi')$, analogously to Definition 1.1 when $\Omega \times \Omega$ is equipped with the product filtration, $\mathcal{AW}_p(\pi, \pi')$ is given by (2.4) when replacing the set $\mathrm{Cpl}(\pi, \pi')$ by $\mathrm{Cpl}_{\mathrm{bc}}(\pi, \pi')$.

Theorem 2.11. *Suppose that for some $p \geq 1$, $b, \bar{b} \colon C([0,1], \mathbb{R}) \to \mathbb{R}$, $\sigma, \bar{\sigma} \colon C([0,1], \mathbb{R}) \to \mathbb{R}_+$ there exist p-integrable unique strong solutions $(X, \bar{X})$ of (2.1). Suppose also that, for all $h > 0$, and $b^h, \bar{b}^h \colon C([0,1], \mathbb{R}) \to \mathbb{R}$, $\sigma^h, \bar{\sigma}^h \colon C([0,1], \mathbb{R}) \to \mathbb{R}$ there exist p-integrable unique strong solutions $(X^h, \bar{X}^h)$ of (2.1). Moreover, suppose that $X^h \to X$ and $\bar{X}^h \to \bar{X}$ in L^p.*

Then, for any bicausal coupling $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mathrm{Law}(X), \mathrm{Law}(\bar{X}))$, there exists a probability space supporting a correlated Wiener process $(W, \bar{W})$ such that π is equal to the joint law of $(X, \bar{X})$ driven by $(W, \bar{W})$ and, for π^h equal to the joint law of $(X^h, \bar{X}^h)$ driven by $(W, \bar{W})$, we have $\lim_{h \to 0} \mathcal{AW}_p(\pi^h, \pi) = 0$.

Remark 2.12. Under Lipschitz conditions on the coefficients of the SDE (1.3), the Euler–Maruyama scheme (1.4) converges to the unique solution of (1.3) in L^p, for all $p \geq 1$ (see, e.g. [30]), and so Theorem 2.11 implies the result of Theorem 1.4. Under the same conditions, the monotone Euler–Maruyama scheme that we develop below in Definition 3.13 also converges to the solution of (1.3) in L^p, for all $p \geq 1$; see Propositions 3.16 and 3.21.

Proof of Theorem 2.11. Take $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mathrm{Law}(X), \mathrm{Law}(\bar{X}))$. By Proposition 2.2, there exists a correlated Wiener process $(W, \bar{W})$ such that π is the joint law of $(X, \bar{X})$ driven by $(W, \bar{W})$. Take $(X^h, \bar{X}^h)$ driven by $(W, \bar{W})$. Then $\mathrm{Law}((X^h, \bar{X}^h), (X, \bar{X})) \in \mathrm{Cpl}_{\mathrm{bc}}(\pi^h, \pi)$. Indeed, for $t \in [0,1]$, and any bounded measurable $f \colon C([0,t], \mathbb{R}) \times C([0,t], \mathbb{R}) \to \mathbb{R}$,

$$\mathbb{E}\Big[f((X^h_s, \bar{X}^h_s)_{s \in [0,t]}) \mid \mathcal{F}^X_1 \otimes \mathcal{F}^{\bar{X}}_1 \Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[f((X^h_s, \bar{X}^h_s)_{s \in [0,t]}) \mid (\mathcal{F}^X_t \otimes \mathcal{F}^{\bar{X}}_t) \vee \sigma\{(W_u - W_t, \bar{W}_u - \bar{W}_t) \colon u \in (t, 1]\} \Big] \mid \mathcal{F}^X_1 \otimes \mathcal{F}^{\bar{X}}_1 \Big]$$

$$= \mathbb{E}\Big[\mathbb{E}\Big[f((X^h_s, \bar{X}^h_s)_{s \in [0,t]}) \mid \mathcal{F}^X_t \otimes \mathcal{F}^{\bar{X}}_t \Big] \mid \mathcal{F}^X_1 \otimes \mathcal{F}^{\bar{X}}_1 \Big]$$

$$= \mathbb{E}\Big[f((X^h_s, \bar{X}^h_s)_{s \in [0,t]}) \mid \mathcal{F}^X_t \otimes \mathcal{F}^{\bar{X}}_t \Big],$$

where the second equality follows from the fact that $\sigma\{(W_u - W_t, \bar{W}_u - \bar{W}_t) \colon u \in (t, 1]\}$ is independent of $(X^h_s, \bar{X}^h_s)_{s \in [0,t]}$ and $\mathcal{F}^X_t \otimes \mathcal{F}^{\bar{X}}_t$. 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 bicausality.

Finally, using the convergence in L^p combined with the same arguments as used in Corollary 2.7, we conclude that $\mathcal{AW}_p(\mathrm{Law}(X^h, \bar{X}^h), \pi)$ converges to zero, as $h \to 0$. $\qquad \square$

3. THE SYNCHRONOUS COUPLING: PROPERTIES AND OPTIMALITY

We now return to the setting of SDEs with time-homogeneous Markovian coefficients. Specifically, we consider functions $b \colon \mathbb{R} \to \mathbb{R}$ and $\sigma \colon \mathbb{R} \to \mathbb{R}_+$ such that there exists a unique strong solution of the SDE (1.3); 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.

In order to define the synchronous coupling between any two such laws, say $\mu = \mu^{b,\sigma}$ and $\nu = \mu^{\bar{b}, \bar{\sigma}}$, consider the system of SDEs, defined for $t \in [0,1]$,

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

We then 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 strong solution of (3.1) driven by (W, W). Then the *synchronous coupling* is defined as the joint law

$$\pi^{\mathrm{sync}}_{\mu,\nu} := \mathrm{Law}(X^{\mathrm{sync}}, \bar{X}^{\mathrm{sync}}).$$

---

[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 $\mathrm{sign}(\sigma \bar{\sigma})$, where $\mathrm{sign}(x) = 1$ if $x \geq 0$ and -1 otherwise. This determines the synchronous coupling uniquely, even in situations where $\sigma \bar{\sigma}$ can become zero.

We will also refer to the coupling obtained via the above procedure with $\bar{W} = -W$ as the *anti-synchronous coupling*, $\pi_{\mu,\nu}^{\text{async}}$.

The main aim of this section is to establish the 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 also establish this result for a particular class of coefficients which allows for discontinuities in the drift coefficient. As part of our analysis, we also provide an optimality result for the Knothe–Rosenblatt rearrangement and establish the link to the synchronous coupling.

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 [42] and Knothe [31]. This coupling can be seen as an extension of the monotone rearrangement between marginal laws on $\mathbb{R}$ to the case of coupling laws of $\mathbb{R}$-valued discrete-time processes. We illustrate the Knothe–Rosenblatt rearrangement in Figure 1. The aim here is to investigate its optimality properties. To this end, we first introduce some notation.

For a probability measure μ on $\mathbb{R}$, the *cumulative distribution function* $F_\mu \colon \mathbb{R} \to [0,1]$, is given by

$$F_\mu(u) = \mu(-\infty, u], \quad u \in \mathbb{R}.$$

The *quantile function* $F_\mu^{-1} \colon [0,1] \to \mathbb{R}$ is defined as its left-continuous inverse; that is, $F_\mu^{-1}(y) = \inf\{u \in \mathbb{R} \colon F_\mu(u) \geq y\}$. Given two probability measures μ, ν on $\mathbb{R}$, recall that μ is said to dominate ν in *first order stochastic dominance* if $F_\mu(u) \leq F_\nu(u)$, for all $u \in \mathbb{R}$.

Given a probability measure μ on $\mathbb{R}^n$, $n \in \mathbb{N}$, let μ_1 be its marginal distribution onto the first component and write $\mu_{x_1,\ldots,x_k}$, $x_1,\ldots,x_k \in \mathbb{R}$, $k = 1,\ldots,n-1$, for the one-dimensional conditional distribution in the $(k+1)^{\text{th}}$ coordinate given the first k coordinates; that is

$$(3.2) \qquad \mu(\mathrm{d}x_1,\ldots,\mathrm{d}x_n) = \mu_1(\mathrm{d}x_1)\mu_{x_1}(\mathrm{d}x_2)\ldots\mu_{x_1,\ldots,x_{n-1}}(\mathrm{d}x_n).$$

If μ defines a Markov process, then the transition kernels $\mu_{x_1,\ldots,x_k}$, $k = 1,\ldots,n-1$, will only depend on x_k but we keep writing out the full tuple for notational reasons.

Two functions $f, g \colon \mathbb{R} \to \mathbb{R}$ 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. Here, and throughout the paper, we use the terms increasing and decreasing in a weak sense; we do not require strict monotonicity.

We are now ready to define the Knothe–Rosenblatt rearrangement:

Definition 3.1 (Knothe–Rosenblatt rearrangement)**.** Given probability measures μ, ν on $\mathbb{R}^n$, let $U_1,\ldots,U_n$ be independent uniform random variables on $[0,1]$, define $X_1 = F_{\mu_1}^{-1}(U_1)$, $Y_1 = F_{\nu_1}^{-1}(U_1)$, and, for $k = 2,\ldots,n$, define inductively the random variables

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

The *Knothe–Rosenblatt rearrangement* between the marginals μ and ν is then given by

$$\pi_{\mu,\nu}^{\text{KR}} := \text{Law}(X_1,\ldots,X_n,Y_1,\ldots,Y_n).$$

For $n = 1$, $\pi_{\mu,\nu}^{\text{KR}} = \text{Law}(F_{\mu_1}^{-1}(U_1), F_{\nu_1}^{-1}(U_1))$ is known as the *monotone rearrangement*.

Remark 3.2. If μ is absolutely continuous with respect to the Lebesgue measure, then $\pi_{\mu,\nu}^{\text{KR}}$ is induced by the Monge map $(x_1,\ldots,x_n) \mapsto T(x_1,\ldots,x_n) = (T^1(x_1), T^2(x_2;x_1),\ldots,T^n(x_n;x_1\ldots,x_{n-1}))$ given by $T^1(x_1) = F_{\nu_1}^{-1} \circ F_{\mu_1}(x_1)$ and, for $k = 2,\ldots,n$,

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

Remark 3.3. The following condition is equivalent to the condition in Definition 3.1: For independent uniform random variables $U_1, \ldots, U_n$ on $[0, 1]$, we have the representation

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

where the functions T_i, S_i are co-monotone in their first argument, for all $i = 1, \ldots, n$. By composing each T_i and S_i in their first argument with suitable increasing functions, we could replace each U_i by an independent normal random variable $Z_i \sim \mathcal{N}(0, 1/n)$, for $i = 1, \ldots, n$. This shows more clearly the resemblance between the synchronous coupling in continuous time and the Knothe–Rosenblatt rearrangement in discrete time. Much of what we do in this paper can be interpreted as justifying a convergence of sorts of the latter to the former when $n \to \infty$.

We now consider the optimality properties of the Knothe–Rosenblatt rearrangement. Suppose that the cost function $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is *submodular*[4]; i.e.

$$(3.3) \qquad c(x, y) + c(\bar{x}, \bar{y}) - c(x, \bar{y}) - c(\bar{x}, y) \leq 0, \quad \text{for all } x \leq \bar{x}, \ y \leq \bar{y}.$$

For probability measures μ, ν on $\mathbb{R}^n$, $n \in \mathbb{N}$, we define the set $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ of bicausal couplings analogously to Definition 1.1 and, for $p \geq 1$, we write $\mu \in \mathcal{P}_p(\mathbb{R}^n)$ if $\int_{\mathbb{R}^n} \sum_{k=1}^n |x_k|^p \mu(\mathrm{d}x) < \infty$.

We relate the optimality of the Knothe–Rosenblatt rearrangement for a class of discrete-time bicausal transport problems to the concept of stochastic monotonicity. For Markov processes, the notion of stochastic monotonicity goes back to Daley [19]. The following definition generalises this concept to arbitrary processes.

Definition 3.4 (stochastic co-monotonicity). A probability measure μ on $\mathbb{R}^n$ (or a stochastic process with law μ) is *stochastically increasing* (resp. decreasing) if, for $k = 1, \ldots, n-1$, the map $(x_1, \ldots, x_k) \mapsto \mu_{x_1, \ldots, x_k}$ is increasing (resp. decreasing) in first order stochastic dominance on $\mathcal{P}(\mathbb{R})$ with respect to the product order on $\mathbb{R}^k$.[5]

We say that two probability measures on $\mathbb{R}^n$ are *stochastically co-monotone* if they are both stochastically increasing, both stochastically decreasing, or one is both stochastically increasing and decreasing and the other is arbitrary.

The following optimality result forms the basis for our proof of optimality of the synchronous coupling in continuous time. For the particular case of a two-period problem and a cost function of the form $c(x, y) = f(x - y)$, $x, y \in \mathbb{R}$, for some convex function f, the result was established in [44, Corollary 2]; for the same type of cost functions and multi-period Markov processes, the result was established in [5, Proposition 5.3]; see, however, Remark 3.8 below.

Proposition 3.5 (optimality of Knothe–Rosenblatt). *For $p \geq 1$ and $n \in \mathbb{N}$, let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$ be stochastically co-monotone. Suppose that, for some $K > 0$, $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies (2.3) and (3.3). Then the Knothe–Rosenblatt rearrangement between μ and ν is optimal for the bicausal transport problem*

$$(3.4) \qquad \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)} \int \sum_{k=1}^n c(x_k, y_k) \pi(\mathrm{d}x, \mathrm{d}y).$$

Suppose additionally that there exists a strictly convex function $f \colon \mathbb{R} \to \mathbb{R}$ and a function $\tilde{c} \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfying (3.3) such that $c(x, y) = f(x - y) + \tilde{c}(x, y)$ for all $x, y \in \mathbb{R}$. Then the Knothe–Rosenblatt rearrangement is the unique optimiser of (3.4).

---

[4]Submodular functions are also referred to as quasi-monotone or L-subadditive.
[5]I.e. $\mu_{\bar{x}_1, \ldots, \bar{x}_k}$ dominates $\mu_{x_1, \ldots, x_k}$ in first order stochastic dominance for any $(x_1, \ldots, x_k), (\bar{x}_1, \ldots, \bar{x}_k) \in \mathbb{R}^k$ with $x_i \leq \bar{x}_i$, for all $i = 1, \ldots, k$.

Before proceeding to the proof, a few remarks concerning classical optimal transport on the line are necessary.

Remark 3.6. Let $p \geq 1$, $\alpha, \beta \in \mathcal{P}_p(\mathbb{R})$, and suppose that, for some $K > 0$, $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ satisfies (2.3) and (3.3). Then, by Rüschendorf [43, Corollary 3], the monotone rearrangement defined in Definition 3.1 attains the infimum in the optimal transport problem

$$\inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int c(x, y) \pi(\mathrm{d}x, \mathrm{d}y).$$

This characterisation of optimality in terms of monotonicity is inherently one-dimensional. To prove optimality of the Knothe–Rosenblatt rearrangement, we iterate this result over multiple time steps. Since we deduce the optimality of the synchronous coupling from the discrete-time result by a limiting argument, the proof of our continuous-time result is also specific to one-dimensional processes; see Proposition 3.19.

Remark 3.7. Let $f \colon \mathbb{R} \to \mathbb{R}$ be strictly convex and $c_f(x, y) := f(x - y)$, for any $x, y \in \mathbb{R}$. Let $p \geq 1$ and $\alpha, \beta \in \mathcal{P}_p(\mathbb{R})$. Then, by Remark 3.6, the monotone rearrangement defined in Definition 3.1 attains the infimum

$$\inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int f(x - y) \pi(\mathrm{d}x, \mathrm{d}y).$$

By [11, Theorem 1], any coupling that attains this infimum is c_f-cyclically monotone. One can verify that, for f strictly convex on $\mathbb{R}$, the only c_f-cyclically monotone coupling is the monotone rearrangement, and thus this is the *unique* optimiser.

Now let $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ take the form prescribed in the uniqueness part of Proposition 3.5; i.e. $c = c_f + \tilde{c}$ with c_f as defined above and $\tilde{c}$ submodular. By Remark 3.6, the monotone rearrangement also attains the infima

$$\inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int c(x, y) \pi(\mathrm{d}x, \mathrm{d}y) \quad \text{and} \quad \inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int \tilde{c}(x, y) \pi(\mathrm{d}x, \mathrm{d}y).$$

Therefore

$$\inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int c(x, y) \pi(\mathrm{d}x, \mathrm{d}y) = \inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int f(x - y) \pi(\mathrm{d}x, \mathrm{d}y) + \inf_{\pi \in \mathrm{Cpl}(\alpha, \beta)} \int \tilde{c}(x, y) \pi(\mathrm{d}x, \mathrm{d}y),$$

and uniqueness of the optimiser for the problem on the left-hand side follows.

Proof of Proposition 3.5. Similarly to (3.2), we identify each measure π on $\mathbb{R}^n \times \mathbb{R}^n$ with its associated sequence of consecutive disintegration kernels

$$(3.5) \quad \pi(\mathrm{d}x_1, \ldots, \mathrm{d}x_n, \mathrm{d}y_1, \ldots, \mathrm{d}y_n) = \pi_1(\mathrm{d}x_1, \mathrm{d}y_1) \pi_{x_1, y_1}(\mathrm{d}x_2, \mathrm{d}y_2) \ldots \pi_{x_{1:n-1}, y_{1:n-1}}(\mathrm{d}x_n, \mathrm{d}y_n),$$

where we use the shorthand notation $x_{1:k}$ for $x_1, \ldots, x_k$, and analogously for the y-component. By [5, Proposition 5.1], $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu)$ if and only if $\pi_1 \in \mathrm{Cpl}(\mu_1, \nu_1)$ and $\pi_{x_{1:k}, y_{1:k}} \in \mathrm{Cpl}(\mu_{x_{1:k}}, \nu_{y_{1:k}})$, for $k = 1, \ldots, n - 1$. This allows us to solve the optimisation problem (3.4) by backward induction. Define $v_n(x_{1:n}, y_{1:n}) \equiv 0$ and, for $k = 2, \ldots, n$, define v_{k-1} inductively by

$$(3.6) \quad v_{k-1}(x_{1:k-1}, y_{1:k-1}) = \inf_{\pi \in \mathrm{Cpl}\left(\mu_{x_{1:k-1}}, \nu_{y_{1:k-1}}\right)} \int c(x_k, y_k) + v_k(x_{1:k}, y_{1:k}) \, \pi(\mathrm{d}x_k, \mathrm{d}y_k).$$

Then, by [5, Proposition 5.2], the value of problem (3.4) is given by $v_0 = \inf_{\pi \in \mathrm{Cpl}(\mu_1, \nu_1)} \int c(x_1, y_1) + v_1(x_1, y_1) \, \pi(\mathrm{d}x_1, \mathrm{d}y_1)$, and (3.4) admits a minimiser π, which is obtained from the sequence of locally optimal disintegration kernels via (3.5).

For $k = n$, the infimum in (3.6) is attained by the monotone rearrangement between $\mu_{x_1:x_{n-1}}$ and $\nu_{y_1:y_{n-1}}$, since c satisfies (3.3); see Remark 3.6. Therefore

$$v_{n-1}(x_{1:n-1}, y_{1:n-1}) = \int_0^1 c\left(F^{-1}_{\mu_{x_1:x_{n-1}}}(u), F^{-1}_{\nu_{y_1:y_{n-1}}}(u)\right) \mathrm{d}u.$$

For $k = n - 1$, the assumption of stochastic monotonicity gives that $x_{n-1} \mapsto F^{-1}_{\mu_{x_1:x_{n-1}}}(u)$ and $y_{n-1} \mapsto F^{-1}_{\nu_{y_1:y_{n-1}}}(u)$ are co-monotone for any $x_{1:n-2}$, $y_{1:n-2}$ and $u \in [0,1]$. We also observe that (3.3) is a linear constraint and that, if c satisfies (3.3) and $f, g \colon \mathbb{R} \to \mathbb{R}$ are co-monotone, then $(x, y) \mapsto c(f(x), g(y))$ also satisfies (3.3). Thus $(x_{n-1}, y_{n-1}) \mapsto c(x_{n-1}, y_{n-1}) + v_{n-1}(x_{1:n-1}, y_{1:n-1})$ satisfies (3.3) for any $x_{1:n-2}$, $y_{1:n-2}$. As a consequence, the monotone rearrangement again attains the infimum on the right-hand side of (3.6), and so

$$
v_{n-2}(x_{1:n-2}, y_{1:n-2}) = \int_0^1 \left\{ c\left(F^{-1}_{\mu_{x_1:x_{n-2}}}(s), F^{-1}_{\nu_{y_1:y_{n-2}}}(s) \right) \right.
$$
$$
\left. + \int_0^1 c\left(F^{-1}_{\mu_{x_1:x_{n-2},x_{n-1}}}(u), F^{-1}_{\nu_{y_1:y_{n-2},y_{n-1}}}(u) \right) \mathrm{d}u \Big|_{(x_{n-1}, y_{n-1}) = \left(F^{-1}_{\mu_{x_1:x_{n-2}}}(s), F^{-1}_{\nu_{y_1:y_{n-2}}}(s) \right)} \right\} \mathrm{d}s.
$$

For $k = n - 2$, the assumption of stochastic monotonicity gives that $x_{n-2} \mapsto \mu_{x_1:x_{n-2}}$ and $y_{n-2} \mapsto \nu_{y_1:y_{n-2}}$ as well as $(x_{n-2}, x_{n-1}) \mapsto \mu_{x_1:x_{n-2},x_{n-1}}$ and $(y_{n-2}, y_{n-1}) \mapsto \nu_{y_1:y_{n-2},y_{n-1}}$ are co-monotone for any $x_{1:n-3}$, $y_{1:n-3}$. Using once again the fact that (3.3) is a linear constraint which is preserved under compositions with co-monotone functions, we have that $(x_{n-2}, y_{n-2}) \mapsto c(x_{n-2}, y_{n-2}) + v_{n-2}(x_{1:n-2}, y_{1:n-2})$ satisfies (3.3), and so the monotone rearrangement attains the infimum on the right-hand side of (3.6).

By induction, the analogous conclusion holds for $k = 1, \ldots, n - 3$, and we conclude that the Knothe–Rosenblatt rearrangement attains the infimum in (3.4).

To show uniqueness of the optimal coupling, first note that, for any coupling π that attains the infimum in (3.4), the disintegration kernel $\pi_{x_{1:k},y_{1:k}}$ attains the infimum v_{k-1} in (3.6), for each $k = 2, \ldots, n$, and π_1 attains the infimum in v_0. Thus it suffices to prove uniqueness of the optimiser for each v_{k-1}, $k = 1, \ldots, n$.

Suppose that $c(x, y) = f(x - y) + \tilde{c}(x, y)$, for all $x, y \in \mathbb{R}$, where f is strictly convex and $\tilde{c}$ satisfies (3.3). Then Remark 3.7 directly implies uniqueness of the optimiser for the infimum in v_{n-1}. Iterating as above, we see that, for each $k = 1, \ldots, n$, v_{k-1} takes the form required to apply Remark 3.7 and thus we conclude that the Knothe–Rosenblatt rearrangement is the unique optimiser for (3.4). $\qquad\square$

Remark 3.8. For Markov processes the assumption of stochastic co-monotonicity reduces to requiring $x_k \mapsto \mu_{x_1,\ldots,x_k}$ and $x_k \mapsto \nu_{x_1,\ldots,x_k}$ to be co-monotone, $k = 1, \ldots, n - 1$, (since $\mu_{x_1,\ldots,x_k}$ only depends on x_k). In general, however, this property is not sufficient to ensure optimality of the Knothe–Rosenblatt rearrangement. Indeed, consider the following marginals which satisfy this property but are not stochastically co-monotone in the sense of Definition 3.4: $\mu = \frac{1}{4}(\delta_{(1,2,7)} + \delta_{(1,0,5)} + \delta_{(-1,0,-5)} + \delta_{(-1,-2,-7)})$ and $\nu = \frac{1}{4}(\delta_{(1,2,-5)} + \delta_{(1,0,-7)} + \delta_{(-1,0,7)} + \delta_{(-1,-2,5)})$. For the cost $c(x, y) = |x - y|$, the Knothe–Rosenblatt rearrangement induces a cost of 12, while one can find another bicausal coupling that induces a cost of 4. This example reveals that the conclusion of [5, Proposition 5.3] does not hold under the stated assumptions. In Proposition 3.5 above, we correct and extend this result.

Remark 3.9. A well-studied example of a cost function satisfying the above conditions is given by $c(x, y) = |x - y|^p$, $x, y \in \mathbb{R}$, for some $p \geq 1$. In this case, for $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$, the infimum in Proposition 3.5 is the *discrete-time adapted Wasserstein distance*

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

cf. [7, Equation (6)].

Remark 3.10. The proof of Proposition 3.5 can be seen as a discrete-time analogue of the proof of optimality of the synchronous coupling between one-dimensional diffusions in [16]. Indeed, both proofs are based on dynamic programming arguments. More precisely, under sufficient regularity

assumptions, the crucial observation in the verification argument of [16] is that the second order cross-derivative of the value function is negative, while the algebraic analogue (3.3) of this condition appears in the induction step in the proof of Proposition 3.5.

Remark 3.11. Note that the finite-dimensional distributions of any one-dimensional continuous strong Markov process are stochastically increasing, as was shown in Beiglböck, Pammer, and Schachermayer [12, Proposition 5.2] by use of a coupling argument originating from Hobson [24]. When such a strong Markov process is the solution of an SDE, our methods give an alternative proof of this fact; see Remark 3.18.

3.2. **A monotone numerical scheme.** In this section, we seek to discretise the SDEs (3.1) via a numerical scheme which is stochastically monotone; in light of Corollary 2.7 and Proposition 3.5, this will lead to a discrete-time bicausal transport problem whose value converges to that of our continuous-time problem. We make the following observation.

Remark 3.12. Recall the Euler–Maruyama scheme for the SDE (1.3) with coefficients $b\colon \mathbb{R} \to \mathbb{R}$, $\sigma\colon \mathbb{R} \to \mathbb{R}_+$ driven by a Wiener process W, as defined in (1.4):

For $N \in \mathbb{N}$, let $h = 1/N$ and $X_0^h = x_0$. Then, for each $k = 0, \dots, N-1$ and $t \in (kh, (k+1)h]$, define

$$(3.7) \qquad X_t^h = X_{kh}^h + (t - kh)b(X_{kh}^h) + \sigma(X_{kh}^h)(W_t - W_{kh}).$$

We call the process $(X_t^h)_{t\in[0,1]}$ the *Euler–Maruyama scheme* and refer to $(X_{kh}^h)_{k\in\{0,\dots,N-1\}}$ as the *discrete-time Euler–Maruyama scheme*.

Now suppose that b is Lipschitz, with Lipschitz constant C, and that σ is *constant*. Then, for $h < C^{-1}$, the process $(X_{kh}^h)_{k\in\{0,\dots,N-1\}}$ is stochastically increasing. Indeed the function $x \mapsto x + hb(x)$ is increasing and so, for $a \in \mathbb{R}$ and $x < x'$,

$$\mathbb{P}\big[x + hb(x) + \sigma(W_{(k+1)h} - W_{kh}) \le a\big] \ge \mathbb{P}\big[x' + hb(x') + \sigma(W_{(k+1)h} - W_{kh}) \le a\big].$$

When we take a non-constant diffusion coefficient σ, the above discrete-time Euler–Maruyama scheme may no longer be stochastically monotone. We therefore define the following variant of the Euler–Maruyama scheme, in which the Brownian increments are truncated, such that we recover the desired stochastic monotonicity.

Let W be a standard one-dimensional Wiener process, let $h > 0$, and define the truncation level $A_h := 4\sqrt{-h \log h}$. Let $W_0^h = 0$. For each $k = 0, 1, \dots$, define the stopping time τ_k^h to be the first time after kh that the Brownian increment $(W_\cdot - W_{kh})$ leaves the interval $(-A_h, A_h)$, and define

$$W_t^h := W_{kh}^h + (W_\cdot - W_{kh})_{t \wedge \tau_k^h}, \quad \text{for} \quad t \in (kh, (k+1)h].$$

Definition 3.13 (monotone Euler–Maruyama scheme)**.** Consider the SDE (1.3) for some coefficients $b\colon \mathbb{R} \to \mathbb{R}$, $\sigma\colon \mathbb{R} \to \mathbb{R}_+$ and a Wiener process W. Fix $N \in \mathbb{N}$, $h = \frac{1}{N}$, and define W^h as above. Let $X_0^h = x_0$. Then, for each $k = 0, \dots, N-1$ 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* and refer to $(X_{kh}^h)_{k\in\{0,\dots,N-1\}}$ as the *discrete-time monotone Euler–Maruyama scheme*.

We now verify that, for Lipschitz coefficients and sufficiently small $h > 0$, the discrete-time monotone Euler–Maruyama scheme is stochastically monotone.

Lemma 3.14. *Suppose that the coefficients b and σ in (1.3) are Lipschitz. Then, for sufficiently small $h > 0$, the discrete-time monotone Euler–Maruyama scheme $(X_{kh}^h)_{k\in\{0,\dots,N-1\}}$ for (1.3) is stochastically increasing.*

Proof. Let $x, x' \in \mathbb{R}$ such that $x < x'$ and let $k \in \{0, \ldots, N - 2\}$. Define the random variables $Y = x + hb(x) + \sigma(x)(W^h_{(k+1)h} - W^h_{kh})$ and $Y' = x' + hb(x') + \sigma(x')(W^h_{(k+1)h} - W^h_{kh})$. 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, of Y and Y'. $\quad\square$

Combined with Proposition 3.5, Lemma 3.14 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. Next 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. *Fix a Wiener process W and consider 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 their respective laws. 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 $k \in \{0, \ldots, N-2\}$ 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 $\Delta_k W^h$, $k = 0, \ldots, N - 2$, for both X^h and $\bar{X}^h$, the characterisation of the Knothe–Rosenblatt rearrangement given in Remark 3.3 implies that the joint law of X^h and $\bar{X}^h$ coincides with the Knothe–Rosenblatt rearrangement between μ^h and ν^h. $\quad\square$

We finally establish that the monotone scheme converges in the L^p-norm to the solution of the SDE; the proof is deferred to Appendix A.

Proposition 3.16. *Let $p \geq 1$. Suppose that the coefficients b and σ in* (1.3) *are Lipschitz and denote by X its unique strong solution. Consider the associated monotone Euler–Maruyama scheme X^h given in Definition 3.13. Then we have the L^p-convergence*

$$\lim_{h \to 0} \mathbb{E}\left[\sup_{0 \leq t \leq 1} |X^h_t - X_t|^p\right] = 0,$$

and the associated discrete-time monotone Euler–Maruyama scheme converges to X in L^p, in the sense that

$$\lim_{h \to 0} \mathbb{E}\left[\sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X^h_{kh} - X_t|^p \mathrm{d}t\right] = 0.$$

Remark 3.17. Milstein et al. [36] proved L^p-convergence of a Milstein scheme with a truncated Gaussian driving noise, for SDEs with sufficiently regular coefficients. A fully implicit Euler–Maruyama scheme with the same truncated noise was also introduced in [36], but no general convergence result was given. Seemingly independently, Liu and Pagès [35] proved L^p-convergence of an Euler–Maruyama scheme with truncated Brownian increments for McKean–Vlasov equations. Since the first version of the present article appeared online, Jourdain and Pagès [26] also proved the L^p-convergence of a similar Euler–Maruyama scheme with truncated Brownian increments, allowing also for time-dependent coefficients. In [26, 35], the authors exploit a monotonicity property similar to Lemma 3.14 in order to study the (monotone) convex ordering of continuous-time processes.

We note that, as in [36], the truncation level A_h is chosen in such a way that the moments of the error introduced by the truncation decay sufficiently fast as $h \to 0$.

Remark 3.18. Supposing that the coefficients of the SDE (1.3) are Lipschitz, the unique strong solution X is a Feller process and so, as noted in Remark 3.11, the finite-dimensional distributions of X are stochastically increasing. We sketch an alternative proof of this fact based on approximation by the monotone Euler–Maruyama scheme. For h sufficiently small, the discrete-time monotone Euler–Maruyama scheme X^h is stochastically increasing by Lemma 3.14. For $t_1, t_2 \in [0,1] \cap \mathbb{Q}$ with $t_1 < t_2$, choose h such that $t_1 = k_1 h$, $t_2 = k_2 h$, for some $k_1, k_2 \in \mathbb{N}_0$. Then the transition kernels for X^h from t_1 to t_2 are also stochastically increasing. The uniform-in-time L^p-convergence of X^h to X shown in Proposition 3.16 implies convergence of finite-dimensional distributions and therefore also of transition kernels. Refining the discretisation grid of X^h appropriately, we conclude that the transition kernels for X from t_1 to t_2 are stochastically increasing, as required.

We are now ready to prove the following result, which implies the conclusion of Theorem 1.3 when the coefficients in (3.1) are Lipschitz.

Proposition 3.19. *Let* $b, \bar{b} \colon \mathbb{R} \to \mathbb{R}$ *and* $\sigma, \bar{\sigma} \colon \mathbb{R} \to \mathbb{R}_+$ *be Lipschitz continuous. For* $N \in \mathbb{N}$, *set* $h = \frac{1}{N}$ *and let* μ^h, ν^h *be the laws of the discrete-time monotone Euler–Maruyama schemes for* (3.1). *Then, for* $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *continuous and satisfying* (2.3) *and* (3.3), *for some* $p \geq 1$ *and* $K > 0$,

$$\lim_{h \to 0} \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)} h \int \sum_{k=0}^{N-1} c(x_k, \bar{x}_k) \mathrm{d}\pi = \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \, \mathrm{d}\pi;$$

moreover, the infimum on the right hand side is attained by the synchronous coupling.

Remark 3.20. 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 A.1. Also, according to Lemma A.8, the associated monotone Euler–Maruyama schemes are bounded in L^p, $p \geq 1$. The bicausal optimal transport problems in the statement of Proposition 3.19 are therefore well-defined.

Proof of Proposition 3.19. Let $X^h, \bar{X}^h \colon \Omega \to \hat{\Omega}$ be measurable maps defining constant interpolations of the monotone Euler–Maruyama schemes for (3.1). That is, for any $k \in \{0, \ldots, N-1\}$ and $t \in [kh, (k+1)h)$, and for any Wiener process W, $(X^h \circ W)_t$ and $(\bar{X}^h \circ W)_t$ coincide with the respective monotone Euler–Mauryama schemes driven by W at time kh. Then, for any correlated Wiener process $(W, \bar{W})$, defining $\pi := \mathrm{Law}(((X^h \circ W)_{kh})_{k \in \{0, \ldots, N-1\}}, ((\bar{X}^h \circ \bar{W})_{kh})_{k \in \{0, \ldots, N-1\}})$,

$$(3.8) \qquad \int h \sum_{k=0}^{N-1} c(x_k, y_k) \pi(\mathrm{d}x, \mathrm{d}y) = \mathbb{E}\left[\int_0^1 c((X^h \circ W)_t, (\bar{X}^h \circ \bar{W})_t) \mathrm{d}t \right],$$

and $\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)$. By Lemma 3.14 and Proposition 3.5, for $h > 0$ sufficiently small, the Knothe–Rosenblatt rearrangement $\pi^{\mathrm{KR}}_{\mu^h, \nu^h}$ attains the infimum

$$(3.9) \qquad \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^h, \nu^h)} \int h \sum_{k=0}^{N-1} c(x_k, y_k) \pi(\mathrm{d}x, \mathrm{d}y).$$

Moreover, by Lemma 3.15, $\pi^{\mathrm{KR}}_{\mu^h, \nu^h} = \mathrm{Law}(((X^h \circ W)_{kh})_{k \in \{0, \ldots, N-1\}}, ((\bar{X}^h \circ W)_{kh})_{k \in \{0, \ldots, N-1\}})$. Hence the perfectly correlated Wiener process (W, W) attains the infimum taken over all correlated Wiener processes on the right-hand side of (3.8), and this infimum coincides with (3.9).

By Proposition 3.16, for any Wiener process W, $X^h \circ W$ converges in L^p to $X^{b,\sigma}$ driven by W, and analogously for $\bar{X}^h$. Hence we can conclude by Corollary 2.7. $\qquad\square$

The above result shows that the continuous-time adapted Wasserstein distance between laws of solutions of (3.1) with Lipschitz coefficients is the limit of discrete-time adapted Wasserstein

distances, each of which are attained by the Knothe–Rosenblatt rearrangement; meanwhile, the synchronous coupling attains the continuous-time adapted Wasserstein distance. In fact, we also have convergence of the optimisers. In view of Theorem 2.11 and the L^p-convergence that we prove in Proposition 3.16, our particular choice of continuous-time scheme leads to the following convergence of the Knothe–Rosenblatt rearrangement to the synchronous coupling in the adapted Wasserstein distance. For this reason, we argue that the synchronous coupling can be viewed as a continuous-time analogue of the Knothe–Rosenblatt rearrangement.

Proposition 3.21. *Let $b, \bar{b} \colon \mathbb{R} \to \mathbb{R}$ and $\sigma, \bar{\sigma} \colon \mathbb{R} \to \mathbb{R}_+$ be Lipschitz continuous. Consider the SDEs (3.1) driven by a common Wiener process W, and write $X, \bar{X}$ for the solutions of (3.1) and μ, ν for their laws. For any $N \in \mathbb{N}$ and $h = 1/N$, let $X^h, \bar{X}^h$ be the monotone Euler–Maruyama schemes for (3.1) driven by the common Wiener process W. Further, let $\mu^h = \mathrm{Law}((X^h_{kh})_{k \in \{0,\ldots,N-1\}})$, $\nu^h = \mathrm{Law}((\bar{X}^h_{kh})_{k \in \{0,\ldots,N-1\}})$ denote the laws of the discrete-time monotone Euler–Maruyama schemes.*

Then, for any $N \in \mathbb{N}$ and $h = 1/N$, $\mathrm{Law}(X^h, \bar{X}^h)$ is an interpolation of the Knothe–Rosenblatt rearrangement, in the sense that

$$\text{(3.10)} \qquad \mathrm{Law}\big((X^h_{kh}, \bar{X}^h_{kh})_{k \in \{0,\ldots,N-1\}}\big) = \pi^{\mathrm{KR}}_{\mu^h, \nu^h}.$$

Moreover, for any $p \geq 1$,

$$\text{(3.11)} \qquad \lim_{h \to 0} \mathcal{AW}_p\big(\mathrm{Law}(X^h, \bar{X}^h), \pi^{\mathrm{sync}}_{\mu,\nu}\big) = 0.$$

Proof. By definition of the synchronous coupling, $\mathrm{Law}(X, \bar{X}) = \pi^{\mathrm{sync}}_{\mu,\nu}$. For fixed $N \in \mathbb{N}$ and $h = 1/N$, Lemma 3.15 implies that $\mathrm{Law}(X^h, \bar{X}^h)$ interpolates the Knothe–Rosenblatt rearrangement in the sense of (3.10).

For any $p \geq 1$, we have the L^p convergence $X^h \to X$ and $\bar{X}^h \to \bar{X}$ by Proposition 3.16. Since the same correlated Wiener process (W, W) is chosen to drive both the SDEs (3.1) and the monotone Euler–Maruyama schemes, the desired $\mathcal{AW}_p$ convergence (3.11) then follows from Theorem 2.11. $\square$

3.3. $\mathcal{AW}_p$ 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 start by making some remarks on Assumption 1.2 and providing examples of coefficients that satisfy this assumption.

Remark 3.22. Under Assumption 1.2, strong existence is guaranteed for the SDEs (3.1). Indeed, by a result of Skorokhod [45], there exist weak solutions of the SDEs under the given continuity and linear growth assumptions on the coefficients. Then, by the Yamada-Watanabe criterion [29, Ch. 5, Corollary 3.23], the combination of pathwise uniqueness and weak existence implies the existence of strong solutions. We refer to [18] for an example of a Markovian SDE for which strong existence does not hold.

Remark 3.23. Assumption 1.2 is satisfied, for example, in the following cases:

(i) $b, \bar{b}, \sigma, \bar{\sigma}$ are Lipschitz [17, Proposition 1.9 (Itô)] — see Proposition 3.19;
(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 [17, Proposition 1.10 (Zvonkin)] — see Proposition 3.30;

(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 [17, 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 $h \colon \mathbb{R}_+ \to \mathbb{R}$ satisfying $\int_0^{0+} h^{-2}(x) \mathrm{d}x = +\infty$ [17, Proposition 1.12 (Yamada–Watanabe)].

We are now ready to provide the following theorem, which, in particular, implies our main result, Theorem 1.3.

Theorem 3.24. *Suppose that (b, σ) and $(\bar{b}, \bar{\sigma})$ satisfy Assumption 1.2. Let $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and satisfy (2.3) and (3.3), for some $p \geq 1$ and $K > 0$. Then the synchronous coupling attains the following infimum:*

$$\inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \, \mathrm{d}\pi.$$

If the inequality in (3.3) is reversed, then this infimum is attained by the anti-synchronous coupling.

Proof. Under Lipschitz conditions on the coefficients of the SDEs (3.1), we have already proved the conclusion of the theorem in Proposition 3.19. Now suppose that the more general condition of Assumption 1.2 is satisfied, namely that the coefficients are continuous with linear growth, and that pathwise uniqueness holds. Then, according to Remark 2.10, the SDEs (3.1) satisfy Assumption 2.8, where the coefficients $b, \bar{b}, \sigma, \bar{\sigma}$ now are Markovian and time-homogeneous and we identify, for example, $b(t, \omega) = b(\omega_t)$, $t \in [0, 1]$, $\omega \in \Omega$. 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 (2.9). Defining μ^n, ν^n as the laws of X^{b^n, σ^n}, $X^{\bar{b}^n, \bar{\sigma}^n}$, we have by Proposition 3.19 that the synchronous coupling $\pi^{\mathrm{sync}}_{\mu^n, \nu^n}$ attains the infimum for the corresponding bicausal transport problem. The first part thus follows by Corollary 2.9. Finally, if c satisfies (3.3) with the inequality reversed, then Proposition 3.5 and Lemma 3.15, and thus Proposition 3.19, hold with obvious modifications and the last part then follows by use of the same arguments as used above. $\square$

Remark 3.25. For the case of $\mathcal{AW}_2$ with sufficiently regular coefficients, uniqueness of the optimal coupling follows from [16], in view of Remark 2.3. Namely, when $b, \bar{b}, \sigma, \bar{\sigma}$ are continuously differentiable with α-Hölder first derivative for some $\alpha \in (0, 1)$, and $\sigma, \bar{\sigma}$ are bounded away from zero, [16, Section 2.1] gives a PDE formulation of $\mathcal{AW}_2$, from which we can deduce uniqueness of the optimiser. Recovering and extending this uniqueness result via probabilistic arguments is left open for future research.

Remark 3.26. 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 3.28 (or vice-versa), then the conclusion of Theorem 1.3 still holds, by the following reasoning.

Suppose that $(\bar{b}, \bar{\sigma})$ satisfy Assumption 3.28. If (b, σ) are Lipschitz, then by examining the proofs of Proposition 3.19 and Proposition 3.30, 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 B.2 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 (2.9). 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 3.27 (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_t^h in Definition 3.13 is replaced by $X_t^h = X_{kh}^h + (t - kh)b(kh, X_{kh}^h) + \sigma(kh, X_{kh}^h)(W_t^h - W_{kh}^h)$. Under the given conditions on the coefficients, the corresponding standard Euler–Maruyama scheme converges in L^2 by [30, 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.

3.4. **Extension to discontinuous drifts.** In this section, we establish the conclusion of Theorem 1.3 under a different set of assumptions, which allows for discontinuities in the drift. For further results in this direction, see the follow-up work [41]. For the coefficients under consideration, we are able to employ a similar approach as used for the case of Lipschitz coefficients above. Specifically, we first apply a Zvonkin-type transformation to remove the drift and then use the monotone Euler–Maruyama scheme for the resulting martingales which feature Lipschitz diffusion coefficients. To this end, we work under the following assumption.

Assumption 3.28. 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 3.29. By Zvonkin's theorem [49], there exist unique strong solutions $(X_t)_{t \geq 0}, (\bar{X}_t)_{t \geq 0}$ of the SDEs (3.1) under Assumption 3.28 (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 3.30. *Suppose that (b, σ) and $(\bar{b}, \bar{\sigma})$ satisfy Assumption 3.28. Let $c \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and satisfy (2.3) and (3.3), for some $p \geq 1$ and $K > 0$. Then, the synchronous coupling attains the following infimum:*

$$\inf_{\pi \in \mathrm{Cpl_{bc}}(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})} \iint_0^1 c(\omega_t, \bar{\omega}_t) \mathrm{d}t \, \mathrm{d}\pi.$$

In particular, for any $p \geq 1$, the synchronous coupling attains the infimum in $\mathcal{AW}_p(\mu^{b,\sigma}, \mu^{\bar{b},\bar{\sigma}})$.

To prove this result, we use exactly the drift-removing transformation introduced by Zvonkin in [49] to prove existence and uniqueness of strong solutions. Define the increasing map $T \colon \mathbb{R} \to \mathbb{R}_+$ 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 (1.3) with coefficients b, σ. Then, by Itô's formula, Y solves the SDE

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

Lemma 3.31. *Suppose that $b\colon \mathbb{R} \to \mathbb{R}$ is bounded and measurable, and that $\sigma\colon \mathbb{R} \to \mathbb{R}_+$ is bounded, uniformly positive, and Lipschitz. Then the map $(\sigma T') \circ T^{-1}\colon \mathbb{R} \to \mathbb{R}_+$ 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. $\qquad\square$

In light of Lemma 3.31, we can apply the monotone Euler–Maruyama scheme defined in Definition 3.13 to the transformed SDE (3.12). Fix $N \in \mathbb{N}$ and $h = \frac{1}{N}$. The full scheme for the SDE (1.3) in this case is then as follows. Let $X_0^h = x_0$ and, for $k = 0, \ldots, N-1$ and $t \in (kh, (k+1)h]$, define

$$(3.13) \qquad X_t^h := T^{-1}\big[T(X_{kh}^h) + \sigma(X_{kh}^h)T'(X_{kh}^h)(W_t^h - W_{kh}^h)\big].$$

The map T is increasing and invertible with increasing inverse. Therefore, for $h > 0$ sufficiently small and $k = 0, \ldots, N-1$, we can use the same arguments as in the proof of Lemma 3.14, along with the fact that $\sigma T' \circ T^{-1}$ is Lipschitz, to see that $X_{kh}^h \mapsto X_{(k+1)h}^h$ is a concatenation of three increasing maps. Hence the process $(X_{kh}^h)_{k=0,\ldots,N}$ is stochastically increasing.

Remark 3.32. Assumption 3.28.(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 (A.1) to the solution Y of (3.12) thus gives us that the solution X of (1.3) is also p-integrable, for $p \geq 1$. Similarly, applying Lemma A.8 to the monotone Euler–Maruyama scheme Y^h for (3.12) yields an L^p-bound for X^h defined by (3.13), for $h > 0$ and $p \geq 1$. The bicausal optimal transport problems appearing in Proposition 3.30 are therefore well-defined.

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

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

Hence, from the L^p-convergence of Y^h to Y given by Proposition 3.16, we can deduce L^p-convergence of X^h to X.

Proof of Proposition 3.30. Thanks to Lemma 3.31 and Remark 3.32, the result follows by use of the same arguments as used to prove Proposition 3.19. $\qquad\square$

Remark 3.33. Since our stability result in Corollary 2.9 requires the coefficients to be continuous, with the methods employed in this paper, Proposition 3.30 is the most general result that we are able to obtain for SDEs with discontinuous coefficients.

4. On the topology induced by the adapted Wasserstein distance

In this section, we apply the stability result of Proposition B.2 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 (1.3) 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 $\mathcal{P}^\Lambda := \{\mu^{b,\sigma} : (b, \sigma) \in \mathcal{A}^\Lambda \times \mathcal{A}^\Lambda\} \subset \mathcal{P}^*$. 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 A.1. 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

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

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 (1.3) 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 B.2, 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 B.1 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 $\pi_{\mu_n, \mu}^{\mathrm{sync}} \in \mathrm{Cpl}_{\mathrm{bc}}(\mu_n, \mu)$, 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 $\mathrm{Cpl}_{\mathrm{bc}}(\mu, \nu) \subseteq \mathrm{Cpl}_{\mathrm{c}}(\mu, \nu) \subseteq \mathrm{Cpl}(\mu, \nu)$, 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. $\qquad\square$

Remark 4.1. The result of Theorem 1.5 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

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

applying the same Arzelà-Ascoli argument. In this case, existence and uniqueness of strong solutions is guaranteed by a result of Engelbert and Schmidt [17, Proposition 1.11] (c.f. Remark 3.23), as the continuity of the coefficients and the strict positivity of σ implies that the local integrability condition required for this result is satisfied.

Remark 4.2. In contrast to the optimality results, Theorem 1.5 admits a multidimensional extension. We first observe that the stability result Proposition B.2 holds in arbitrary dimensions, following the same proof as in dimension one. For two $\mathbb{R}^d$-valued SDEs, we define the synchronous coupling as the joint law of their solutions when they are driven by a common d-dimensional Brownian motion, and we define the subset $\mathcal{P}^\Lambda$ of probability measures on $C([0, T], \mathbb{R}^d)$ analogously to the one-dimensional setting. Then one can adapt the proof of Theorem 1.5 to prove the same topological equivalence in general dimensions. We note that this multidimensional result extends [16, Propositions 1.8 and 1.9].

5. 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 [6].

Example 5.1 (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

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

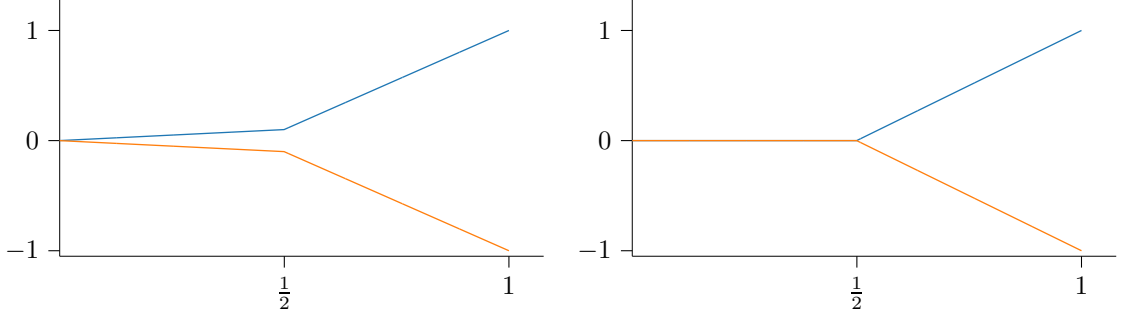$$X_{\frac{1}{2}}^n = -\frac{1}{n} \quad \text{and} \quad X_1^n = -1, \quad \text{with probability } \frac{1}{2},$$

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.

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 bicausal, however. In fact the only bicausal 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 $V^n := \inf_{Z \ \mathcal{F}_{\frac{1}{2}}^{X^n}\text{-measurable}} \mathbb{E}|X_1^n - Z|^2$, for each $n \in \mathbb{N} \cup \{\infty\}$, with $\mathcal{F}^{X^n}$ denoting the raw natural filtration of X^n. Since X_1^n is $\mathcal{F}_{\frac{1}{2}}^{X^n}$-measurable, for each $n \in \mathbb{N}$, we have $V^n = 0$. On the other hand, $V^\infty = 1$, since the sigma-algebra $\mathcal{F}_{\frac{1}{2}}^{X^\infty}$ 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.

5.1. **Non-Markovianity.** 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).

Example 5.2 (non-Markovian counterexample)**.** This example already appears in [6], as a counterexample in a different setting.

Let $C > 0$, $h \in (0,1)$, and define $b(t, \omega) := C \operatorname{sign}(\omega_h) \mathbb{1}_{\{t > h\}}$, for $t \in [0, 1]$, $\omega \in \Omega$. Then let $\mu := \mathrm{Law}(X)$, where X is the unique strong solution of

$$\mathrm{d}X_t = b(t, X)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0;$$

note that strong existence and pathwise uniqueness are guaranteed by Zvonkin [49]. Similarly, let $\nu := \mathrm{Law}(\bar{X})$, where $\bar{X}$ is the unique strong solution of

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

Now consider the couplings

$$\pi^{\text{sync}} := \text{Law}(W_\cdot + C\,\text{sign}(W_h)[\cdot - h]_+ \ , \ W_\cdot - C\,\text{sign}(W_h)[\cdot - h]_+);$$

$$\pi^{\text{async}} := \text{Law}(W_\cdot + C\,\text{sign}(W_h)[\cdot - h]_+ \ , \ -W_\cdot + C\,\text{sign}(W_h)[\cdot - h]_+).$$

Note that $\pi^{\text{sync}}, \pi^{\text{async}} \in \text{Cpl}_{\text{bc}}(\mu, \nu)$. 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

$$\mathbb{E}^{\pi^{\text{sync}}}\left[\int_0^1 |\omega_t - \bar\omega_t|^2 \mathrm{d}t\right] = \mathbb{E}\left[\int_h^1 (2C\,\text{sign}(W_h)[t-h])^2 \mathrm{d}t\right] = \frac{4}{3}C(1-h)^3, \quad \text{while}$$

$$\mathbb{E}^{\pi^{\text{async}}}\left[\int_0^1 |\omega_t - \bar\omega_t|^2 \mathrm{d}t\right] = \mathbb{E}\left[\int_0^1 (2W_t)^2 \mathrm{d}t\right] = 2.$$

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 5.3 (kinetic SDEs). Let $b, \bar b \colon \mathbb{R} \to \mathbb{R}$ be increasing and Lipschitz continuous, and consider, for $t \in [0, 1]$, the kinetic equations

(5.1)
$$dX_t = b\left(\int_0^t X_s \mathrm{d}s\right)\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = 0;$$
$$d\bar X_t = \bar b\left(\int_0^t \bar X_s \mathrm{d}s\right)\mathrm{d}t + \mathrm{d}\bar W_t, \quad \bar X_0 = 0.$$

A reasonable time-discretisation of such SDEs would be to define, for $N \in \mathbb{N}$ and $h = \frac{1}{N}$, the process $(X_k^h)_{k=0,\dots,N}$ by $X_0^h = 0$ and for $k = 1, \dots, N$,

$$X_k^h = X_{k-1}^h + b\left(h \sum_{i < k} X_i^h\right)h + W_{kh} - W_{(k-1)h}.$$

This scheme is stochastically monotone in the sense of Definition 3.4. Defining $\bar X^h$ in the same way, Proposition 3.5 then gives that the Knothe–Rosenblatt rearrangement is optimal among bicausal couplings between the laws of X^h and $\bar X^h$ (for the same class of cost functions). Similarly to Lemma 3.15, we also have that the Knothe–Rosenblatt rearrangement is given by $\text{Law}(X^h, \bar X^h)$ when the discretisation schemes X^h and $\bar X^h$ are defined with respect to a common Wiener process $W = \bar W$.

By use of arguments similar to those employed for the Markovian case in Section 3, we expect to obtain appropriate convergence of the scheme to the true solution. Optimality of the synchronous coupling for the continuous-time bicausal transport problem between the laws of solutions to (5.1) would then follow from Proposition 2.6. However, we do not carry out this analysis rigorously.

5.2. **Multidimensional examples.** As stated in Remark 3.6, our proof of optimality of the synchronous coupling does not extend to higher dimensions. In this section we present multidimensional examples for which the synchronous coupling is not optimal. In particular, these examples illustrate that, in higher dimensions, Markovianity is no longer an indicator of optimality of the synchronous coupling. The kinetic system (5.1) in Example 5.3 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 a similar fashion, Example 5.2 can be made Markovian by introducing an additional state variable.

Example 5.4 (two-dimensional counterexample I). Let $A > 0$ be small and $C > 0$ large, let B, W be independent Wiener processes and consider the SDEs

$$\mathrm{d}X_t = C\big(\mathbb{1}_{\{Y_t > -A\}} - \mathbb{1}_{\{Y_t < A\}}\big)\mathrm{d}t + \mathrm{d}B_t, \quad X_0 = 0,$$

$$\mathrm{d}Y_t = \mathbb{1}_{\{|Y_t| < A\}}\mathrm{d}W_t, \quad Y_0 = 0.$$

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

$$\mathrm{d}\bar{X}_t = -C\big(\mathbb{1}_{\{\bar{Y}_t > -A\}} - \mathbb{1}_{\{\bar{Y}_t < A\}}\big)\mathrm{d}t + \mathrm{d}\bar{B}_t, \quad \bar{X}_0 = 0,$$

$$\mathrm{d}\bar{Y}_t = \mathbb{1}_{\{|\bar{Y}_t| < A\}}\mathrm{d}\bar{W}_t, \quad \bar{Y}_0 = 0.$$

The process Y (resp. $\bar{Y}$) is a Wiener process until hitting $-A$ or A, where it freezes. The process X (resp. $\bar{X}$) is also a Wiener process 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.

In the previous example, we considered two-dimensional systems driven by two-dimensional Wiener processes. A natural question is whether a counterexample to the optimality of the synchronous coupling can be constructed when the driving Wiener processes are one dimensional. The next example addresses this question.

Example 5.5 (two-dimensional counterexample II). Let $W, \bar{W}$ be one-dimensional standard Wiener processes and let $h \in (0, 1)$, $C \in (0, \infty)$. Let (X, Y) be strong solutions of the SDEs

$$\mathrm{d}X_t = C\mathrm{sign}(Y_t)\mathbb{1}_{\{t > h\}}\mathrm{d}t + \mathrm{d}W_t,$$

$$\mathrm{d}Y_t = \mathbb{1}_{\{t \leq h\}}\mathrm{d}W_t,$$

with $X_0 = Y_0 = 0$, and write $\mu = \mathrm{Law}(X, Y)$. Also let $(\bar{X}, \bar{Y})$ be strong solutions of the SDEs

$$\mathrm{d}\bar{X}_t = -C\mathrm{sign}(\bar{Y}_t)\mathbb{1}_{\{t > h\}}\mathrm{d}t + \mathrm{d}\bar{W}_t,$$

$$\mathrm{d}\bar{Y}_t = \mathbb{1}_{\{t \leq h\}}\mathrm{d}\bar{W}_t,$$

with $\bar{X}_0 = \bar{Y}_0$, and write $\nu = \mathrm{Law}(\bar{X}, \bar{Y})$. Then we can again define the synchronous coupling $\pi_{\mu,\nu}^{\mathrm{sync}} := \mathrm{Law}(X, Y, \bar{X}, \bar{Y})$ when $W = \bar{W}$, and the antisynchronous coupling $\pi_{\mu,\nu}^{\mathrm{async}} := \mathrm{Law}(X, Y, \bar{X}, \bar{Y})$ when $W = -\bar{W}$.

Observe that, for any $t \in [0, 1]$, we have $Y_t = W_{t \wedge h}$ and $\bar{Y}_t = \bar{W}_{t \wedge h}$, and so

$$\mathrm{d}X_t = C\mathrm{sign}(W_h)\mathbb{1}_{\{t > h\}}\mathrm{d}t + \mathrm{d}W_t,$$

$$\mathrm{d}\bar{X}_t = -C\mathrm{sign}(\bar{W}_h)\mathbb{1}_{\{t > h\}}\mathrm{d}t + \mathrm{d}\bar{W}_t,$$

as in Example 5.2. The squared L^2-cost under the synchronous coupling is now calculated to be

$$\mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{sync}}}\left[\int_0^1 |X_t - \bar{X}_t|^2 \mathrm{d}t + \int_0^1 |Y_t - \bar{Y}_t|^2 \mathrm{d}t\right] = \mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{sync}}}\left[\int_0^1 |X_t - \bar{X}_t|^2 \mathrm{d}t\right] = \frac{4}{3}C(1 - h)^3.$$

And for the antisynchronous coupling, the cost is given by

$$\mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{async}}}\left[\int_0^1 |X_t - \bar{X}_t|^2 \mathrm{d}t + \int_0^1 |Y_t - \bar{Y}_t|^2 \mathrm{d}t\right] = \mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{async}}}\left[\int_0^1 |X_t - \bar{X}_t|^2 \mathrm{d}t\right] + 4\int_0^1 \mathbb{E}\big[|W_{t \wedge h}|^2\big]\mathrm{d}t$$

$$= 2 + 4\int_0^1 (t \wedge h)\mathrm{d}t \leq 2 + 4h.$$

Taking C sufficiently large and h sufficiently small, we conclude that the synchronous coupling may fail to be optimal.

5.3. Adapted Wasserstein distance with L^∞ norm.

Suppose that we wish to replace the L^p norm on $\Omega = C([0,1], \mathbb{R})$ with the L^∞ norm and find the associated adapted 1-Wasserstein distance between measures μ, ν on Ω:

$$\inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \mathbb{E}^\pi \left[\sup_{t \in [0,1]} |\omega_t - \bar{\omega}_t| \right].$$

Discretising the problem as before, we arrive at

$$(5.2) \qquad \inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu^N, \nu^N)} \mathbb{E}^\pi \left[\max_{k \in \{1,\dots,N\}} |x_k - \bar{x}_k| \right],$$

for some $N \in \mathbb{N}$ and measures μ_N, ν_N on $\mathbb{R}^N$. This discrete-time bicausal optimal transport problem does not satisfy the assumptions of Proposition 3.5 for optimality of the Knothe–Rosenblatt rearrangement, since the cost function is not of a separable form.

We now give a counterexample to optimality of the Knothe–Rosenblatt rearrangement for the problem (5.2), when $N = 2$ and the marginals are atomic. We leave open the question of finding the optimiser when the marginals are the laws of some numerical scheme for an SDE.

Example 5.6. Define the measures $\mu := 1/2 \cdot (\delta_{(-3,-7)} + \delta_{(1,4)})$ and $\nu := 1/2 \cdot (\delta_{(2,4)} + \delta_{(5,6)})$ on $\mathbb{R}^2$. We will show that the Knothe–Rosenblatt rearrangement is suboptimal for

$$\inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \mathbb{E}^\pi [|x_1 - \bar{x}_1| \vee |x_2 - \bar{x}_2|].$$

We illustrate processes with law μ and ν, respectively, along with two bicausal couplings in Figure 3.
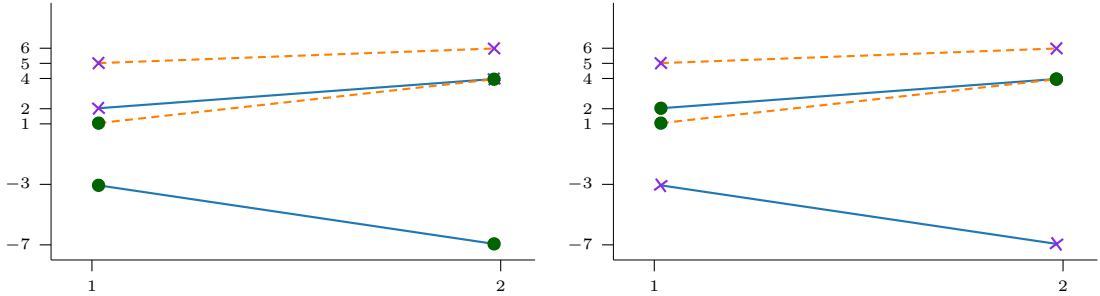

FIGURE 3. The Knothe–Rosenblatt rearrangement $\pi_{\mu,\nu}^{\mathrm{KR}}$ is shown on the left, and the coupling $\pi_{\mu,\nu}^{\mathrm{AT}}$ on the right. The solid blue lines and dashed orange lines represent processes with law μ and ν, respectively. At each time, the points with the same colour and style are coupled with each other.

For the Knothe–Rosenblatt rearrangement $\pi_{\mu,\nu}^{\mathrm{KR}}$ we have

$$\mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{KR}}}[|x_1 - \bar{x}_1| \vee |x_2 - \bar{x}_2|] = 1/2 \cdot (|-3-2| \vee |-7-4|) + 1/2 \cdot (|1-5| \vee |4-6|)$$

$$= 1/2 \cdot (11 + 4) = 15/2.$$

On the other hand, consider the coupling $\pi_{\mu,\nu}^{\mathrm{AT}}$ that is defined similarly to $\pi_{\mu,\nu}^{\mathrm{KR}}$ but with the first marginals coupled via the *antitone* rearrangement rather than the monotone rearrangement. Then

$$\mathbb{E}^{\pi_{\mu,\nu}^{\mathrm{AT}}}[|x_1 - \bar{x}_1| \vee |x_2 - \bar{x}_2|] = 1/2 \cdot (|-3-5| \vee |-7-6|) + 1/2 \cdot (|1-2| \vee |4-4|)$$

$$= 1/2 \cdot (13 + 1) = 14/2 < 15/2.$$

Thus the Knothe–Rosenblatt rearrangement is not an optimiser.

Note that both μ and ν are stochastically increasing. Therefore Proposition 3.5 implies that the Knothe–Rosenblatt rearrangement $\pi_{\mu,\nu}^{\mathrm{KR}}$ is an optimiser of

$$\inf_{\pi \in \mathrm{Cpl}_{\mathrm{bc}}(\mu,\nu)} \mathbb{E}^{\pi}[|x_1 - \bar{x}_1| + |x_2 - \bar{x}_2|].$$

In fact one can compute that both $\pi_{\mu,\nu}^{\mathrm{KR}}$ and $\pi_{\mu,\nu}^{\mathrm{AT}}$ attain the same value for this problem and are therefore both optimisers.

References

[1] B. Acciaio, J. Backhoff-Veraguas, and R. Carmona. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM J. Control Optim.*, 57(6): 3666–3693, 2019. doi: 10.1137/18M1196479.

[2] B. Acciaio, J. Backhoff-Veraguas, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stoch. Proc. Appl.*, 130(5):2918–2953, 2020. ISSN 0304-4149. doi: 10.1016/j.spa.2019.08.009.

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

[4] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008. ISBN 978-3-7643-8721-1.

[5] J. Backhoff-Veraguas, M. Beiglböck, Y. Lin, and A. Zalashko. Causal transport in discrete time and applications. *SIAM J. Optim.*, 27(4):2528–2562, 2017. ISSN 1052-6234. doi: 10. 1137/16M1080197.

[6] 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. ISSN 0949-2984. doi: 10.1007/s00780-020-00426-3.

[7] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probab. Theory Relat. Fields*, 178(3-4):1125–1172, 2020. ISSN 0178-8051. doi: 10.1007/ s00440-020-00993-8.

[8] J. Backhoff-Veraguas, M. Beiglböck, M. Eder, and A. Pichler. Fundamental properties of process distances. *Stoch. Proc. Appl.*, 130(9):5575–5591, 2020. ISSN 0304-4149. doi: 10.1016/ j.spa.2020.03.017.

[9] D. Bartl, M. Beiglböck, and G. Pammer. The Wasserstein space of stochastic processes. *J. Eur. Math. Soc.*, 2024. To appear. arXiv:2104.14245.

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

[11] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *J. Funct. Anal.*, 256(6):1907–1927, 2009. ISSN 0022-1236.

[12] M. Beiglböck, G. Pammer, and W. Schachermayer. From Bachelier to Dupire via optimal transport. *Finance Stoch.*, 26(1):59–84, 2022. ISSN 0949-2984. doi: 10.1007/ s00780-021-00466-3.

[13] M. Beiglböck, G. Pammer, and A. Posch. The Knothe–Rosenblatt distance and its induced topology. *arXiv:2312.16515*, 2023.

[14] M. Beiglböck, G. Pammer, and S. Schrott. Denseness of biadapted Monge mappings. *Ann. Inst. H. Poincaré Probab. Statist.*, Oct. 2024. doi: 10.48550/arxiv.2210.15554.

[15] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. A Wiley-Interscience Publication.

[16] 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. ISSN 1050-5164. doi: 10.1214/18-AAP1423.

[17] A. S. Cherny and H.-J. Engelbert. *Singular Stochastic Differential Equations*, volume 1858 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-24007-5 978-3-540-31560-5. doi: 10.1007/b104187.

[18] A. M. G. Cox and B. A. Robinson. SDEs with no strong solution arising from a problem of stochastic control. *Electron. J. Probab.*, 28:1–24, 2023. ISSN 1083-6489. doi: 10.1214/23-EJP995.

[19] D. J. Daley. Stochastically monotone Markov chains. *Z. Wahrsch. Verw. Gebiete*, 10(4):305–317, 1968.

[20] S. Eckstein and G. Pammer. Computational methods for adapted optimal transport. *Ann. Appl. Probab.*, 34(1A):675 – 713, 2024. doi: 10.1214/23-AAP1975.

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

[22] I. I. Gihman and A. V. Skorohod. *Controlled Stochastic Processes*. Springer New York, New York, NY, 1979. ISBN 978-1-4612-6204-6 978-1-4612-6202-2. doi: 10.1007/978-1-4612-6202-2.

[23] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *J. Math. Econom.*, 25(4):443–464, 1996. ISSN 0304-4068. doi: 10.1016/0304-4068(95)00739-3.

[24] D. G. Hobson. Volatility misspecification, option pricing and superreplication via coupling. *Ann. Appl. Probab.*, 8(1):193–205, 1998. ISSN 1050-5164. doi: 10.1214/aoap/1027961040.

[25] J. Jacod and J. Mémin. Weak and strong solutions of stochastic differential equations: existence and stability. In *Stochastic Integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pages 169–212. Springer, Berlin, 1981.

[26] B. Jourdain and G. Pagès. Convex ordering of solutions to one-dimensional SDEs. *arXiv:2312.09779*, 2023.

[27] S. Källblad. A dynamic programming approach to distribution-constrained optimal stopping. *Ann. Appl. Probab.*, 32(3):1902–1928, 2022.

[28] O. Kallenberg. *Foundations of Modern Probability*. Probability and its Applications. Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

[29] 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. ISBN 0-387-97655-8. doi: 10.1007/978-1-4612-0949-2.

[30] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992. ISBN 978-3-642-08107-1 978-3-662-12616-5. doi: 10.1007/978-3-662-12616-5.

[31] H. Knothe. Contributions to the theory of convex bodies. *Mich. Math. J.*, 4:39–52, 1957. ISSN 0026-2285.

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

[33] R. Lassalle. Causal transference plans and their Monge–Kantorovich problems. *Stoch. Anal. Appl.*, 36(3):452–484, 2018.

[34] J.-F. Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer International Publishing, Cham, 2016. ISBN 978-3-319-31088-6 978-3-319-31089-3. doi: 10.1007/978-3-319-31089-3.

[35] Y. Liu and G. Pagès. Monotone convex order for the McKean–Vlasov processes. *Stoch. Proc. Appl.*, 152:312–338, 2022.

[36] G. N. Milstein, Y. M. Repin, and M. V. Tretyakov. Numerical methods for stochastic systems preserving symplectic structure. *SIAM J. Numer. Anal.*, 40(4):1583–1604, 2002. ISSN 0036-1429, 1095-7170. doi: 10.1137/S0036142901395588.

[37] G. Pammer. A note on the adapted weak topology in discrete time. *Electron. Commun. Probab.*, 29:1 – 13, 2024. doi: 10.1214/23-ECP572.

[38] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012. ISSN 1052-6234. doi: 10.1137/110825054.

[39] G. C. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014. ISBN 978-3-319-08842-6; 978-3-319-08843-3. doi: 10.1007/978-3-319-08843-3.

[40] A. Pichler and M. Weinhardt. The nested Sinkhorn divergence to learn the nested distance. *Comput. Manag. Sci.*, pages 1–25, 2021.

[41] B. A. Robinson and M. Szölgyenyi. Bicausal optimal transport for SDEs with irregular coefficients. *arXiv:2403.09941*, 2024.

[42] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Stat.*, 23(3):470–472, 1952.

[43] L. Rüschendorf. Solution of a statistical optimization problem by rearrangement methods. *Metrika*, 30(1):55–61, 1983. ISSN 0026-1335, 1435-926X. doi: 10.1007/bf02056901.

[44] L. Rüschendorf. The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete*, 70(1):117–129, 1985. ISSN 0044-3719. doi: 10.1007/BF00532240.

[45] 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.

[46] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. ISBN 0-8218-3312-X.

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

[48] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.

[49] A. K. Zvonkin. A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR Sbornik*, 22(1):129–149, Feb. 1974. ISSN 0025-5734. doi: 10.1070/SM1974v022n01ABEH001689.

Appendix A. Convergence of the monotone scheme

The aim here is to prove Proposition 3.16 which states the L^p-convergence of the monotone Euler–Maruyama scheme (given in Definition 3.13) to the unique strong solution of (1.3) when the coefficients are Lipschitz. 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; making use of a bound on the pth moments of the monotone scheme, we then deduce L^p-convergence. Throughout the following proofs, we make use of generic constants, which may change from one line to the next.

Remark A.1. Under the assumption that the coefficients b, σ in (1.3) are Lipschitz, there exists a unique strong solution X to (1.3) according to a classical result of Itô; see, e.g. [17, Proposition 1.9]. Moreover, for $p > 0$, the process X satisfies the following moment bounds. There exist constants $C_p, \tilde{C}_p > 0$ such that

$$(A.1) \qquad \mathbb{E}\left[\sup_{0 \leq t \leq 1} |X_t|^p\right] \leq C_p,$$

and, for any $s, t \in [0, 1]$ with $s < t$,

$$(A.2) \qquad \mathbb{E}\left[\sup_{s \le u \le t} |X_u - X_s|^p\right] \le \tilde{C}_p (t-s)^{\frac{p}{2}}.$$

These bounds follow from a standard application of Doob's martingale inequality, the Burkholder–Davis–Gundy inequality, and Grönwall's lemma; see, e.g. [22, Lemma 3.8 and Equation (3.48)].

Remark A.2. 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 augmented to satisfy the usual conditions.

We start with a lemma, which we adapt from [36, Lemma 2.1], that gives a bound on the fourth moment of the error created by the truncation. This bound is used in the proof of Proposition A.5 and thus justifies the choice of the truncation level A_h.

Lemma A.3. *For $N \in \mathbb{N}$, $h = 1/N$, and fixed $k \in \{0, \ldots, N-1\}$, we have the fourth moment bound*

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

Proof. First note that

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

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

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

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

Remark A.4. 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 A.3.

In order to prove the L^2-convergence of the monotone Euler–Maruyama scheme X^h to the unique strong solution X of (1.3), we first recall the following estimates for the standard Euler–Maruyama scheme $\tilde{X}^h$ (defined in (3.7)) when the coefficients are Lipschitz. From the proof of [30, Theorem 10.2.2], for example, there exists a constant $\tilde{C}_0$ such that, for any $h > 0$, we have the L^2 estimate

$$(A.3) \qquad \mathbb{E}\left[\sup_{0 \le s \le 1} |\tilde{X}_s^h - X_s|^2\right] \le \tilde{C}_0 h.$$

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

$$(A.4) \qquad \mathbb{E}\left[\sup_{0 \le s \le 1} |\tilde{X}_s^h|^p\right] \le \tilde{C}_p.$$

Proposition A.5. *Suppose that the coefficients b and σ in (1.3) are Lipschitz. Then there exists a constant $C > 0$ such that, for any $N \in \mathbb{N}$ and $h = \frac{1}{N}$,*

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

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

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

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

so that

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

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

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

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

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

As noted in Remark A.2, W^h is an $\mathcal{F}^W$-martingale, and we see that $\mathrm{d}\langle W^h\rangle_t \leq \mathrm{d}t$. Thus, by Doob's martingale inequality,

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

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

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

$$U_t \leq 4\mathbb{E}\left[\left|\sum_{k=0}^{n_t-1}\sigma(\tilde{X}_{kh}^h)\left[\Delta_k W - \Delta_k W^h\right] + \sigma(\tilde{X}_{t_{n_t}}^h)\left[W_t - W_{t_{n_t}} - (W_t^h - W_{t_{n_t}}^h)\right]\right|^2\right]$$

$$\leq 4N\sum_{k=0}^{N-1}\mathbb{E}\left[\sigma(\tilde{X}_{kh}^h)^2(\Delta_k W - \Delta_k W^h)^2\right],$$

and, by the Cauchy–Schwarz inequality,

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

Applying Lemma A.3, we can bound each term $\mathbb{E}\left[(\Delta_k W - \Delta_k W^h)^4\right] \leq Ch^{10}$. Using the Lipschitz property of σ and the L^p bound (A.4) for the Euler–Maruyama scheme, we can also bound $\mathbb{E}[\sigma(\tilde{X}_{kh}^h)^4] \leq C$, for any $k = 0, \ldots, N$. Therefore we have

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

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

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

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

Remark A.6. Similarly to Remark A.4, the power in the bound in Proposition A.5 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 (1.3).

Corollary A.7. *Suppose that the coefficients b and σ in (1.3) are Lipschitz. Then there exists a constant $C > 0$ such that, for any $h > 0$ sufficiently small,*

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

Proof. Combining the rate of L^2 convergence of the Euler–Maruyama scheme given in (A.3) with the estimate of the L^2-error between the Euler–Maruyama scheme and the monotone Euler–Maruyama scheme given in Proposition A.5, we can conclude via a simple application of the triangle inequality that

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

$\square$

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

Lemma A.8. *Suppose that the coefficients b and σ in (1.3) are Lipschitz. Then, for $p \geq 1$, 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 $h > 0$,*

$$\mathbb{E}\left[\sup_{0 \leq t \leq 1} |X_t^h|^p\right] \leq C_p.$$

Proof. Follows from a standard application of martingale inequalities and Grönwall's lemma, similarly to Remark A.1. $\square$

Convergence in L^p now follows immediately.

Proof of Proposition 3.16. By Corollary A.7, 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 A.8 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 (A.1). Combining the L^2-convergence with the bounds in L^{p+1} implies L^p-convergence, as required.

To prove L^p-convergence of the discrete-time scheme, note that

$$\mathbb{E}\left[\sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_{kh} - X_{kh}^h|^p \mathrm{d}t\right] \leq T\mathbb{E}\left[\sup_{0 \leq t \leq 1} |X_t - X_t^h|^q\right] \xrightarrow{h \to 0} 0,$$

and (A.2) provides the estimate

$$\mathbb{E}\left[\sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}|^p \mathrm{d}t\right] \leq h \sum_{k=0}^{N-1} \mathbb{E}\left[\sup_{kh \leq t \leq (k+1)h} |X_t - X_{kh}|^p\right] \leq \tilde{C}_p T h^{\frac{p}{2}}.$$

We conclude by the triangle inequality. $\square$

Appendix B. A stability result for SDEs

We here establish a stability result for the parameter dependence of path-dependent SDEs driven by correlated Wiener processes. The result can also be obtained from [25, Theorem 3.24], for example, but for completeness we provide a direct proof.

In order to formulate the result, recall the notation $\Omega = C([0,1], \mathbb{R})$ and $\|\omega\|_\infty := \sup_{s \in [0,1]} |\omega_s|$, $\omega \in \Omega$, for the sup-norm. 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 $p \geq 1$, and $\pi, \pi' \in \mathcal{P}(\Omega \times \Omega)$ with finite p^{th} moment, we here define the p-Wasserstein distance (with respect to the sup-norm) between π and π' to be

$$(B.1) \qquad \inf_{\alpha \in \text{Cpl}(\pi, \pi')} \mathbb{E}^{\alpha}[\|\omega - \omega'\|_{\infty}^p + \|\bar{\omega} - \bar{\omega}'\|_{\infty}^p],$$

where $\text{Cpl}(\pi, \pi')$ denotes the set of probability measures on $(\Omega \times \Omega) \times (\Omega \times \Omega)$ with marginal distribution onto the first (resp. last) two coordinates given by π (resp. π') and $((\omega, \bar{\omega}), (\omega', \bar{\omega}'))$ denotes the canonical process.

Remark B.1. For any $p \geq 1$, π_n converges to π with respect to the p-Wasserstein distance on $\mathcal{P}(\Omega \times \Omega)$, as defined in (B.1), if and only if, for any continuous function $\phi \colon \Omega \times \Omega \to \mathbb{R}$ with at most polynomial growth of order p — i.e. $|\phi(\omega, \bar{\omega})| \leq C(1 + \|\omega\|_{\infty}^p + \|\bar{\omega}\|_{\infty}^p)$, $(\omega, \bar{\omega}) \in \Omega \times \Omega$ — it holds that $\mathbb{E}^{\pi_n}[\phi(\omega, \bar{\omega})] \to \mathbb{E}^{\pi}[\phi(\omega, \bar{\omega})]$ (see, e.g. [46, Theorem 7.12]).

Proposition B.2. *Let $(W, \bar{W})$ be a ρ-correlated Wiener process, for some progressively measurable process ρ, as defined in Definition 2.1. Suppose that $(x_0, \bar{x}_0, b, \bar{b}, \sigma, \bar{\sigma})$ satisfies Assumption 2.8, and write $(X, \bar{X})$ for the unique strong solution of (2.1) 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 2.8.(i) and (iii), with a uniform slope constant K in (2.8), and such that strong existence holds for (2.1); let $(X^{b^n, \sigma^n}, \bar{X}^{\bar{b}^n, \bar{\sigma}^n})$ be one such strong solution, when (2.1) 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:

$$(B.2) \qquad \|\omega^n - \omega\|_{\infty} \to 0 \implies (b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)(t, \omega^n) \to (b, \bar{b}, \sigma, \bar{\sigma})(t, \omega), \text{ for each } t \in [0, 1].$$

Then, for any $p \geq 1$,

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

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

Proof. Similarly to Remark A.1, 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 pth moments of $\|X^{b^n, \sigma^n}\|_{\infty}$ and $\|\bar{X}^{\bar{b}^n, \bar{\sigma}^n}\|_{\infty}$ are bounded uniformly 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 2.8.(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 $(X, \bar{X})$, since the latter is then well posed.

Let $\eta \in \mathcal{P}(\Omega \times \Omega)$ be one such weak accumulation point. Then, after possibly passing to a subsequence, Skorokhod's representation theorem ensures the existence of stochastic processes $(X^n, \bar{X}^n, W^n, \bar{W}^n)_{n \in \mathbb{N}}$ and $(Y, \bar{Y}, W^{\infty}, \bar{W}^{\infty})$ defined on a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

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

$(X^n, \bar{X}^n, W^n, \bar{W}^n) \to (Y, \bar{Y}, W^{\infty}, \bar{W}^{\infty})$ pointwise, and $\text{Law}(Y, \bar{Y}) = \eta$. 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 (2.1), 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 $m_{\varepsilon} := (\mathrm{d}t \times \mathbb{P})(\{(t, \omega) : (t, W^n(\omega), \bar{W}^n(\omega)) \notin E\}) \leq \varepsilon$ 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 $\rho^{\varepsilon} \colon [0, 1] \times \Omega \times \Omega \to [-1, 1]$, 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 $f \colon [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ which is differentiable in time, twice differentiable in space, and with corresponding bounded and continuous derivatives, it holds that $R_f^n = 0$, where

$$R_f^n := \mathbb{E}\Big[f(T, X_1^n, \bar{X}_1^n) - f(0, x_0^n, \bar{x}_0^n) -$$

$$\int_0^1 \big\{ \partial_t f + b^n \partial_x f + \bar{b}^n \partial_{\bar{x}} f + \frac{1}{2}(\sigma^n)^2 \partial_{xx} f + \frac{1}{2}(\bar{\sigma}^n)^2 \partial_{\bar{x}\bar{x}} f + \rho(t, W^n, \bar{W}^n) \sigma^n \bar{\sigma}^n \partial_{x\bar{x}} f \big\}(t, X^n, \bar{X}^n) dt \Big],$$

and we identify $f(t, Z, \bar{Z}) \equiv f(t, Z_t, \bar{Z}_t)$, for any processes $Z, \bar{Z}$. On the other hand, we may also define

$$R_f^\infty := \mathbb{E}\Big[f(T, Y_1, \bar{Y}_1) - f(0, x_0, \bar{x}_0) -$$

$$\int_0^1 \big\{ \partial_t f + b \partial_x f + \bar{b} \partial_{\bar{x}} f + \frac{1}{2}\sigma^2 \partial_{xx} f + \frac{1}{2}\bar{\sigma}^2 \partial_{\bar{x}\bar{x}} f + \rho(t, W^\infty, \bar{W}^\infty) \sigma \bar{\sigma} \partial_{x\bar{x}} f \big\}(t, Y, \bar{Y}) dt \Big],$$

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[6] ρ^ε, but otherwise unchanged. 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 (B.2), the integrand converges $dt \times \mathbb{P}$-almost surely. 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 $|R_f^n - R_f^{n,\varepsilon}| \leq C\varepsilon$, 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 $\bar{Y}$. Thus

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

and we can conclude by uniqueness of solutions to the martingale problem associated to the system for $(X, \bar{X})$. $\qquad \square$

---

[6]Crucially, we do not claim that the martingale problem with initial condition $(x_0^n, \bar{x}_0^n)$ and coefficients $(b^n, \bar{b}^n, \sigma^n, \bar{\sigma}^n)$, driven by a ρ^ε-correlated Brownian motion, is well-posed. Instead we view the introduction of $\rho^\varepsilon, R_f^{n,\varepsilon}, R_f^{\infty,\varepsilon}$, as a technical means toward the goal of showing that $R_f^\infty = 0$.