

# BICAUSAL OPTIMAL TRANSPORT FOR SDES WITH IRREGULAR COEFFICIENTS

MICHAELA HITZ AND BENJAMIN A. ROBINSON

*Department of Statistics, University of Klagenfurt, Austria*

**ABSTRACT.** We solve constrained optimal transport problems in which the marginal laws are given by the laws of solutions of stochastic differential equations (SDEs). We consider SDEs with irregular coefficients, making only minimal regularity assumptions. We show that the so-called synchronous coupling is optimal among bicausal couplings, that is couplings that respect the flow of information encoded in the stochastic processes. Our results provide a method to numerically compute the adapted Wasserstein distance between laws of SDEs with irregular coefficients. We show that this can be applied to quantifying model uncertainty in stochastic optimisation problems. Moreover, we introduce a transformation-based semi-implicit numerical scheme and establish the first strong convergence result for SDEs with exponentially growing and discontinuous drift.

## 1. INTRODUCTION

A fundamental question in stochastic modelling is that of quantifying the effects of model uncertainty. In this context it is of interest to compute an appropriate distance between different stochastic models. In particular, we focus on models that are described by stochastic differential equations (SDEs), potentially having irregular coefficients. The main contribution of this paper is to provide a methodology to efficiently compute such a distance between the laws of SDEs, by solving an optimal transport problem. Our approach successfully brings together optimal transport and numerical analysis of SDEs. Moreover, for SDEs with exponentially growing and discontinuous drift, we prove strong well-posedness and strong convergence of a novel numerical scheme.

In order to compare stochastic models, a reasonable choice of distance is a modification of the Wasserstein distance on the space of probability measures. The Wasserstein distance arises from an optimal transport problem and is defined as the infimum of an expected cost over the set of couplings  $\text{Cpl}(\mu, \nu)$  between probability measures  $\mu$  and  $\nu$ , that is the set of probability measures on the product space having marginals  $\mu$  and  $\nu$ . This distance metrises the weak topology. However, in many situations the weak topology is unsuitable for comparing stochastic processes, since it does not take into account the information structure that is encoded in the filtrations generated by the processes. As a remedy, the adapted Wasserstein distance and associated bicausal optimal transport problem has been proposed. This constitutes the main object of study in this paper.

We consider stochastic models described by scalar SDEs of the form

$$(1.1) \quad dX_t = b_t(X_t) dt + \sigma_t(X_t) dW_t, \quad t \in [0, T],$$

---

*E-mail address:* michaela.hitz@aau.at, benjamin.robinson@aau.at.

*Date:* February 6, 2026.

*2020 Mathematics Subject Classification.* 60H10, 49Q22, 65C30, 60H35.

*Key words and phrases.* Adapted Wasserstein distance, bicausal couplings, numerical methods for stochastic differential equations, optimal transport, uncertainty quantification.

This research was funded in part by the Austrian Science Fund (FWF) [10.55776/Y782], [10.55776/P35519], [10.55776/P34743]. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

with  $X_0 = x_0 \in \mathbb{R}$ , where  $T \in (0, \infty)$ ,  $b: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma: [0, T] \times \mathbb{R} \rightarrow [0, \infty)$  are measurable functions, and  $W = (W_t)_{t \in [0, T]}$  is a standard scalar Brownian motion. Now consider two different models with data  $(b, \sigma, x_0, W)$ , respectively  $(\bar{b}, \bar{\sigma}, x_0, \bar{W})$ . For  $\Omega = C([0, T], \mathbb{R})$ ,  $p \geq 1$ , and  $\mathcal{P}_p(\Omega)$  being the set of all probability measures on  $\Omega$  with finite  $p^{\text{th}}$  moment, denote the law of the solutions of these SDEs by  $\mu \in \mathcal{P}_p(\Omega)$ , respectively  $\nu \in \mathcal{P}_p(\Omega)$ . Let  $(\omega, \bar{\omega})$  denote the canonical process on  $\Omega \times \Omega$ . Then the adapted Wasserstein distance between  $\mu$  and  $\nu$  is defined by

$$\mathcal{AW}_p^p(\mu, \nu) := \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \mathbb{E}^\pi \left[ \int_0^T |\omega_t - \bar{\omega}_t|^p dt \right],$$

where we optimise over the set  $\text{Cpl}_{\text{bc}}(\mu, \nu)$  of bicausal couplings between  $\mu$  and  $\nu$ , that is the set of couplings that satisfy an additional constraint that encodes the filtrations of the processes. Informally, this bicausality condition imposes that for any  $t \in [0, T]$ , the value  $X_t$  is conditionally independent of the process  $\bar{X}$  given  $(\bar{X}_s)_{s \in [0, t]}$ , with the symmetric condition when exchanging the roles of  $X$  and  $\bar{X}$ . In the context of optimisation problems for stochastic processes, such as those that arise in mathematical finance and mathematical biology, the causality condition appears natural. One is typically only able to make decisions based on information contained in the filtration of the process at the present time.

We aim to solve such a bicausal transport problem under minimal regularity assumptions on the coefficients of the SDE (1.1). For one-dimensional SDEs (1.1) with smooth time-homogeneous coefficients, Bion-Nadal and Talay [12, §2.1] exploit PDE methods to show that the so-called synchronous coupling attains the adapted Wasserstein distance  $\mathcal{AW}_2$ . This coupling is the joint law of the SDEs when they are driven by a common Brownian motion. Following this work, Backhoff-Veraguas, Källblad, and Robinson [8, Theorem 1.3] show that the synchronous coupling attains  $\mathcal{AW}_p$ ,  $p \geq 1$ , between the laws of SDEs with globally Lipschitz coefficients. The result of [8] holds also for continuous linearly growing coefficients, provided that pathwise uniqueness holds. However, no explicit conditions for this are given in [8].

We show that optimality of the synchronous coupling holds in much greater generality, considering also SDEs with irregular coefficients. We consider two classes of irregularity:

- (1) discontinuous drift that grows up to exponentially, degenerate diffusion;
- (2) bounded measurable drift, bounded diffusion that is  $\alpha$ -Hölder, with  $\alpha \in [1/2, 1]$ .

Our main result can be stated informally as follows:

**Main result.** *For models with data  $(b, \sigma, x_0, W)$ , respectively  $(\bar{b}, \bar{\sigma}, x_0, \bar{W})$ , each potentially having any one of the above irregularities, the synchronous coupling is optimal. Furthermore, one can compute the adapted Wasserstein distance by approximating each scalar SDE with any numerical scheme for which strong convergent rates are known.*

Given that the synchronous coupling attains the adapted Wasserstein distance, the problem of computing this distance reduces to that of simulating two one-dimensional SDEs, using the same noise for each, and applying (multilevel-)Monte Carlo methods. Such computational efficiency is not usually expected, even in the case of classical optimal transport, without some regularisation. Moreover, this distance provides a computationally efficient measure of model uncertainty for optimal stopping problems.

The methods of proof in our paper are based on transformation and discretisation of SDEs. In case (1), for SDEs with discontinuous and exponentially growing drift, additionally satisfying a piecewise one-sided Lipschitz condition, we introduce a transformation-based semi-implicit Euler–Maruyama scheme for which we prove strong convergence. We use the transformation method developed in [47] to handle the discontinuities, and combine it with a semi-implicit scheme to

handle the super-linear growth. We design this numerical scheme such that it is stochastically monotone. For measures on  $\mathbb{R}^n$  satisfying such a monotonicity condition, it is known that the well-studied Knothe–Rosenblatt rearrangement is an optimiser for the bicausal transport problem between them. Applying the strong convergence of the scheme and a stability result for bicausal optimal transport, we prove the optimality of the synchronous coupling. In case (2), with bounded and measurable drift, we apply the Zvonkin transformation to obtain SDEs with sufficiently regular coefficients, for which optimality results are already known. We show that such a transformation preserves the optimality property of the synchronous coupling. The novel strong convergence result in case (1) can be stated informally as follows:

**Auxiliary result.** *For the SDE (1.1) with drift  $b$  that has exponential growth and satisfies mild piecewise continuity conditions, and with diffusion  $\sigma$  that may be degenerate except at the points of discontinuity of  $b$ , there exists a unique strong solution with bounded moments of all orders, and a transformation-based drift-implicit Euler–Maruyama scheme converges strongly with known rates.*

We conclude this section by commenting on the classes of SDEs covered by our assumptions. To the best of our knowledge, the only existing optimality result for SDEs with discontinuous drift is [8, Proposition 3.30], which considers bounded coefficients with additional regularity and integrability conditions. Our case (2) extends the bounded coefficients case to the full generality of the existence and uniqueness result of Zvonkin [88]. For case (1), we are able to treat unbounded coefficients with discontinuous drift and even possibly degenerate diffusion coefficient. While a boundedness assumption is artificial for many applications, allowing for discontinuous drift coefficients that may grow super-linearly permits us to treat important examples, such as the value of a dividend-paying firm in a Black–Scholes market; see Example 4.3. For our optimality proof in this setting, we require strong existence and uniqueness, as well as a strongly converging numerical scheme. As described above, we also go beyond existing results in this direction. One may naturally ask whether our results can be further extended to path-dependent coefficients or multi-dimensional equations. In both cases, however, counterexamples to optimality of the synchronous coupling are given in [8, Section 5].

**1.1. Structure of the article.** Since this work combines optimal transport and numerical analysis of SDEs, we keep it self-contained such that it is accessible to readers from both fields. We provide the necessary background on optimal transport theory in Section 2.1 and its applications in the theory of stochastic processes in Section 2.2. In Section 2.3, we discuss numerical approximation schemes for SDEs and their application to bicausal optimal transport. We present some preliminary stability and optimality results in Section 3.

We present our first theorem in Section 4, where we consider the first class of irregularity of the coefficients stated above, that is discontinuous drift with exponential growth and degenerate diffusion. We define a novel numerical scheme, prove its strong convergence, and provide convergence rates in Section 4.2. We prove optimality of the synchronous coupling for this class of coefficients in Section 4.3.

In Section 5 we consider the second class of irregularity of the coefficients stated above, that is bounded measurable drift and Hölder continuous diffusion. Applying the Zvonkin transformation, we also prove optimality of the synchronous coupling for this class of coefficients.

We combine the results of the preceding sections in Section 6 to prove the main theorem of this paper. In Section 7 we apply our results to compute adapted Wasserstein distances, and finally we present an application to the quantification of model uncertainty for stochastic optimisation problems in Section 8.

The proofs of several results from Section 3 and Section 4 are collected in Appendix A and Appendix B, respectively.

The relevant literature can be found in the respective sections.

## 2. OPTIMAL TRANSPORT AND APPLICATIONS TO STOCHASTIC PROCESSES

Here we introduce the concepts from optimal transport that will be needed for our applications. The field has a long history, originating with Monge [55] in 1781 and taken up again by Kantorovich [40] in 1942. For the interested reader, modern accounts can be found, for example, in Villani [83], Ambrosio, Gigli, and Savaré [3], and Santambrogio [77], with the latter taking a more applied point of view.

**2.1. Introduction to optimal transport.** Given separable metric spaces  $(\mathcal{X}, d)$  and  $(\mathcal{Y}, \rho)$ , let  $\mathcal{B}(\mathcal{X})$ ,  $\mathcal{B}(\mathcal{Y})$  denote the respective Borel sigma-algebras, and  $\mathcal{P}(\mathcal{X})$ ,  $\mathcal{P}(\mathcal{Y})$  the spaces of probability measures on  $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ ,  $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ , respectively. Further, let  $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$  denote the set of probability measures on the product space equipped with the product Borel sigma-algebra. For  $\mu \in \mathcal{P}(\mathcal{X})$ ,  $\nu \in \mathcal{P}(\mathcal{Y})$ , define the set of *couplings*  $\text{Cpl}(\mu, \nu)$  by

$$\text{Cpl}(\mu, \nu) := \{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \pi(A, \mathcal{Y}) = \mu(A), \pi(\mathcal{X}, B) = \nu(B), \text{ for all } A \in \mathcal{B}(\mathcal{X}), B \in \mathcal{B}(\mathcal{Y}) \}.$$

Consider a lower semicontinuous *cost function*  $C: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{\infty\}$ .

The *optimal transport problem* [40] is to find

$$(\text{OT}) \quad \inf_{\substack{\pi \in \text{Cpl}(\mu, \nu) \\ \text{Law}(X, Y) = \pi}} \mathbb{E}[C(X, Y)],$$

supposing sufficient conditions on  $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$ ,  $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \nu)$  and  $C$  such that the expectation in (OT) is well defined for all  $\pi \in \text{Cpl}(\mu, \nu)$ .

**2.1.1. Brenier's theorem.** Central to the theory of optimal transport is the following key result of Brenier [15]: *an optimiser for the quadratic cost is given by the gradient of a convex function*. More precisely, let  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$  be equipped with the Euclidean metric  $|\cdot|$  and set  $c(x, y) := |x - y|^2$ . Suppose that  $\mu$  is absolutely continuous with respect to the Lebesgue measure. Then (OT) admits a unique optimiser  $\pi^{\text{mono}}$  and there exists a convex function  $v: \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\pi^{\text{mono}} = (\text{id}, \nabla v)_{\#} \mu$ , the pushforward of the measure  $\mu$  by the map  $(\text{id}, \nabla v): \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ . In the one-dimensional case,  $\nabla v = F_{\nu}^{-1} \circ F_{\mu}$ , where  $F_{\eta}, F_{\eta}^{-1}$  denote the cumulative distribution function and quantile function of a measure  $\eta$ , and the coupling  $\pi^{\text{mono}}$  is known as the *monotone rearrangement* (or the *Hoeffding–Fréchet coupling*).

**2.1.2. Wasserstein distance.** Let  $\mathcal{X} = \mathcal{Y}$  be equipped with some metric  $d$  such that  $(\mathcal{X}, d)$  is a Radon space. Then, following [3, Section 7.1], a metric is induced on the space of probability measures  $\mathcal{P}_p(\mathcal{X})$ , for any  $p \geq 1$ , by taking the cost function  $C = d^p$  in (OT). For measures  $\mu, \nu \in \mathcal{P}_p(\mathcal{X})$  the *p-Wasserstein distance*  $\mathcal{W}_p(\mu, \nu)$  is defined by

$$(2.1) \quad \mathcal{W}_p^p(\mu, \nu) := \inf_{\substack{\pi \in \text{Cpl}(\mu, \nu) \\ \text{Law}(X, Y) = \pi}} \mathbb{E}[d(X, Y)^p].$$

The *p-Wasserstein distance* metrises the weak topology on  $\mathcal{P}_p(\mathcal{X})$  (see [3, Proposition 7.1.5]) and is thus a natural candidate to compare probability measures on  $\mathcal{X}$ .

**2.2. Optimal transport applied to stochastic processes.** We now consider optimal transport between the laws of real-valued stochastic processes. First we treat discrete-time processes.

Set  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$ , for some  $n \in \mathbb{N}$ . Equip this space with the  $L^p$  norm, defined by  $\|x\|_p^p := \sum_{i=1}^n |x_i|^p$  for  $x \in \mathbb{R}^n$ , for some  $p \geq 1$ , and with the corresponding Borel sigma-algebra  $\mathcal{B}(\mathbb{R}^n)$ .

Let  $x = (x_1, \dots, x_n)$  denote the canonical process on  $\mathbb{R}^n$ , and  $(x, \bar{x})$  the canonical process on the product space  $\mathbb{R}^n \times \mathbb{R}^n$ . Let  $\mathcal{F} = (\mathcal{F}_k)_{k \in \{1, \dots, n\}}$  denote the canonical filtration on  $\mathbb{R}^n$ , and note that  $\mathcal{F}_n = \mathcal{B}(\mathbb{R}^n)$ . Given a random variable  $X$  on  $(\mathbb{R}^n, \mathcal{F}_n)$  with law  $\eta$ , let  $\mathcal{F}^X = (\mathcal{F}_k^X)_{k \in \{1, \dots, n\}}$  denote the completion of its natural filtration with respect to  $\eta$ .

As stated in Section 2.1.2, the Wasserstein distance is a natural candidate to measure the distance between two probability measures on  $(\mathbb{R}^n, \mathcal{F})$ . However, in the case that such probability measures are the laws of stochastic processes, the usual Wasserstein distance (and associated weak topology) lacks some desirable properties. For example, Aldous [2, Section 11] already noted that deterministic processes can converge weakly to martingales. We illustrate this with the following example from Backhoff-Veraguas, Bartl, Beiglböck, and Eder [6]; cf. Aldous [2, Example 11.4]. A continuous-time analogue of this example also appears in Backhoff-Veraguas et al. [8, Example 5.1].

**Example 2.1.** Consider the two-step real-valued processes  $X, X^\varepsilon$ , with laws  $\mu, \mu^\varepsilon$ , for each  $\varepsilon > 0$ , defined by

$$\begin{aligned} \mathbb{P}[(X_1^\varepsilon, X_2^\varepsilon) = (\varepsilon, 1)] &= \mathbb{P}[(X_1^\varepsilon, X_2^\varepsilon) = (-\varepsilon, -1)] = \frac{1}{2}, \\ \mathbb{P}[(X_1, X_2) = (0, 1)] &= \mathbb{P}[(X_1, X_2) = (0, -1)] = \frac{1}{2}. \end{aligned}$$

The paths of the processes are shown in Figure 1. Define  $\pi^* \in \text{Cpl}(\mu^\varepsilon, \mu)$  such that positive

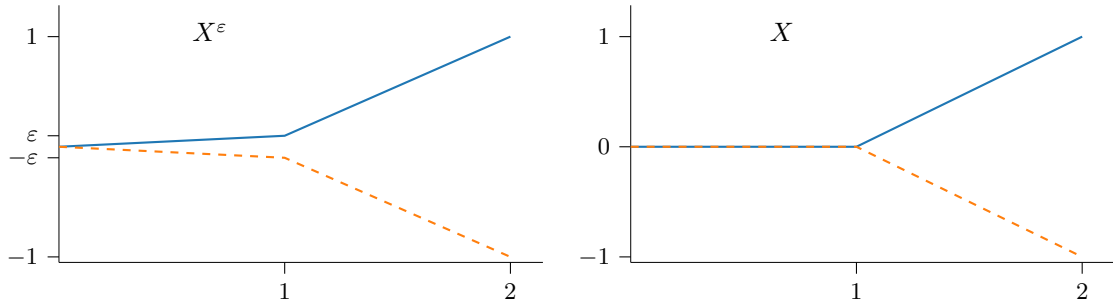


FIGURE 1. A sequence of two-step deterministic processes  $X^\varepsilon$  converges weakly as  $\varepsilon \rightarrow 0$  to a martingale  $X$  [6, Figure 1].

(resp. negative) values of  $X^\varepsilon$  are coupled with positive (resp. negative) values of  $X$ ; that is the solid blue paths in Figure 1 are coupled with each other, and the dashed orange paths are coupled with each other. Then we find that, for  $\text{Law}(X^\varepsilon, X) = \pi^*$ ,

$$\mathcal{W}_p^p(\mu^\varepsilon, \mu) \leq \sum_{i=1}^2 \mathbb{E}[|X_i^\varepsilon - X_i|^p] = \varepsilon^p,$$

implying the weak convergence  $\mu^\varepsilon \rightharpoonup \mu$  as  $\varepsilon \rightarrow 0$ .

On the other hand, consider the values of optimal stopping problems  $V^\varepsilon := \sup_\tau \mathbb{E}[X_\tau^\varepsilon]$  and  $V := \sup_\tau \mathbb{E}[X_\tau]$ , where the suprema are taken over stopping times taking values in  $\{1, 2\}$ . Since  $X$  is a martingale centred at 0, we immediately see that  $V = 0$ . However, in the case of  $X^\varepsilon$ , the optimal stopping time is 2 in the case that  $X_1^\varepsilon > 0$ , and 1 otherwise, and so we obtain

$$V^\varepsilon = \frac{1}{2}(1 - \varepsilon) \rightarrow \frac{1}{2}.$$

Thus we see that the processes  $X^\varepsilon$  and  $X$  have a very different information structure, but nevertheless their laws are close in the usual Wasserstein distance. This example motivates the introduction of the *adapted Wasserstein distance* and the associated *bicausal optimal transport* problem.

**2.2.1. Bicausal optimal transport.** We have seen that the drawback of the Wasserstein distance for comparing stochastic processes is that this distance is not able to distinguish different information structures. To rectify this shortcoming, we consider Wasserstein distances with the additional constraint that the couplings must respect the flow of information. This constraint has been formalised in various ways, including the *Markov constructions* of Rüschendorf [75], the *nested distances* of Pflug and Pichler [66, 67], a stochastic control problem of Bion-Nadal and Talay [12], the *causal transport plans* of Lassalle [45], and the *adapted Wasserstein distance* introduced by Backhoff-Veraguas et al. [6]. Here we will use the terminology *adapted Wasserstein distance*, which we define below. Although there are other candidates for distances between stochastic processes, many such distances have been shown to be topologically equivalent to the adapted Wasserstein distance; see, for example, [7, 14].

Analogously to the classical Wasserstein distance, the adapted Wasserstein distance is induced by a *bicausal optimal transport* problem.

**Definition 2.2.** Let  $\mu, \nu$  be probability measures on  $(\mathbb{R}^n, \mathcal{F})$  and let  $\pi \in \text{Cpl}(\mu, \nu)$ . The coupling  $\pi$  is *causal* if, for any random variables  $X, Y$  with  $\text{Law}(X, Y) = \pi$ , we have the conditional independence

$$(2.2) \quad \mathcal{F}_k^Y \text{ is independent of } \mathcal{F}_n^X \text{ under } \pi \text{ conditional on } \mathcal{F}_k^X,$$

for all  $k \in \{1, \dots, n\}$ . If  $\pi \in \text{Cpl}(\mu, \nu)$  is causal, and the coupling  $\theta_{\#}\pi \in \text{Cpl}(\nu, \mu)$  is causal, where  $\theta(x, y) = (y, x)$ , then we say that  $\pi$  is *bicausal*. We write  $\text{Cpl}_{\text{bc}}(\mu, \nu)$  for the set of bicausal couplings.

In words,  $\pi = \text{Law}(X, Y)$  is causal if the present value of  $Y$  is independent of the future of the process  $X$ , conditional on the past of  $X$ .

We assume that the cost function  $C: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  in (OT) takes the following form. Suppose that there exist continuous functions  $c_k: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ , for  $k \in \{1, \dots, n\}$  such that  $C(x, y) = \sum_{k=1}^n c_k(x_k, y_k)$ , for any  $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \mathbb{R}^n$ . For each  $k \in \{1, \dots, n\}$ , suppose that the function  $c_k$  has *polynomial growth*; i.e. for some  $p \geq 1$ , there exists  $K \geq 0$  such that, for all  $x, y \in \mathbb{R}$ ,  $k \in \{1, \dots, n\}$ ,

$$(2.3) \quad |c_k(x, y)| \leq K[1 + |x|^p + |y|^p],$$

and that  $c_k$  is *quasi-monotone*<sup>1</sup> in the sense that

$$(2.4) \quad c_k(x, y) + c_k(x', y') - c_k(x, y') - c_k(x', y) \geq 0, \quad \text{for all } x \leq x', y \leq y'.$$

For  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$ , the *bicausal optimal transport problem* is then to find (OT) with the usual set of couplings  $\text{Cpl}(\mu, \nu)$  replaced by the set of bicausal couplings  $\text{Cpl}_{\text{bc}}(\mu, \nu)$ , i.e.

$$(2.5) \quad \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \mathbb{E}^{\pi} \left[ \sum_{k=1}^n c_k(x_k, \bar{x}_k) \right].$$

Taking each function  $c_k$  in (2.5) to be the  $p^{\text{th}}$  power of the Euclidean distance, we define an adapted analogue of the Wasserstein distance (2.1) as follows.

**Definition 2.3** (adapted Wasserstein distance – discrete time). Let  $p \geq 1$  and  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$ . Then the *adapted Wasserstein distance*  $\mathcal{AW}_p(\mu, \nu)$  is defined by

$$\mathcal{AW}_p^p(\mu, \nu) := \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \mathbb{E}^{\pi} \left[ \sum_{k=1}^n |x_k - \bar{x}_k|^p \right].$$

<sup>1</sup>also called *L-superadditive* or *supermodular*; see Block, Griffith, and Savits [13, Remark 1.1]

**Remark 2.4.** The term *adapted* appears natural here, since Definition 2.2 constrains the usual set of couplings to those that satisfy a condition analogous to adaptedness of stochastic processes. The term *causal*, introduced in [45], does not necessarily correspond to the notion of causality in statistics; for example, the product coupling is bicausal. On the other hand, Cheridito and Eckstein [16] relate causal optimal transport to the causal structure of graphical models.

**Example 2.5** (Example 2.1 revisited). For the laws given in Example 2.1, the adapted Wasserstein distance  $\mathcal{AW}_p(\mu^\varepsilon, \mu)$  is bounded away from 0, uniformly in  $\varepsilon > 0$ , for any  $p \geq 1$ . Indeed, under any  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \mu^\varepsilon)$ ,  $\mathcal{F}_1^{X^\varepsilon}$  is independent of  $\mathcal{F}_2^X$  conditional on  $\mathcal{F}_1^X$ . Since  $\mathcal{F}_1^X$  is trivial,  $\mathcal{F}_1^{X^\varepsilon}$  and  $\mathcal{F}_2^X$  are independent under  $\pi$ , without any conditioning. Moreover,  $X_2^\varepsilon$  is fully determined by  $X_1^\varepsilon$ , so  $\mathcal{F}_2^{X^\varepsilon} = \mathcal{F}_1^{X^\varepsilon}$ . Thus  $\{\mathcal{F}_k^{X^\varepsilon}\}_{k \in \{1,2\}}$  is independent of  $\mathcal{F}_2^X$  under  $\pi$  and, using again that  $\mathcal{F}_1^X$  is trivial, is independent of  $\{\mathcal{F}_k^X\}_{k \in \{1,2\}}$ . Hence the product coupling is the only element of  $\text{Cpl}_{\text{bc}}(\mu, \mu^\varepsilon)$ , and so  $\mathcal{AW}_p(\mu^\varepsilon, \mu) = \varepsilon^p + 2^{p-1} > 2^{p-1}$ . Recalling that, for the optimal stopping problems in Example 2.1,  $V^\varepsilon \not\rightarrow V$  as  $\varepsilon \rightarrow 0$ , the adapted Wasserstein distance appears to be better suited to comparing stochastic processes than the usual Wasserstein distance.

Stability of various optimisation problems with respect to the adapted Wasserstein distance is shown in Backhoff-Veraguas et al. [6]. Moreover, optimal stopping problems are in fact continuous with respect to the adapted Wasserstein distance by [7, Lemma 7.1]; cf. [2, Theorem 17.2]. This confirms our intuition from Examples 2.1 and 2.5.

**2.2.2. Optimality of bicausal couplings.** Under the following condition of stochastic monotonicity, Rüschendorf [75] identified an optimiser for a bicausal transport problem between the laws of discrete-time Markov processes. For a measure  $\eta$  on  $\mathbb{R}^n$ , denote by  $\eta_1$  the first one-dimensional marginal of  $\eta$  and, for each  $k \in \{2, \dots, n\}$ , denote by  $\eta_{x_1, \dots, x_{k-1}}$  the  $k^{\text{th}}$  conditional marginal of  $\eta$ , given that the first  $k-1$  coordinates are equal to  $(x_1, \dots, x_{k-1})$ .

**Definition 2.6** (stochastic co-monotonicity [18]). Let  $(X_n)_{n \in \{1, \dots, N\}}$  be a real-valued Markov process with law  $\mu \in \mathcal{P}(\mathbb{R}^n)$ . The process  $X$ , or the measure  $\mu$ , is said to be *stochastically increasing* (resp. *decreasing*) if  $x_k \mapsto \mu_{x_1, \dots, x_{k-1}, x_k}$  is increasing (resp. decreasing) in first order stochastic dominance, for each  $k \in \{1, \dots, N-1\}$ . Two such processes or measures are *stochastically co-monotone* if they are both stochastically increasing, both stochastically decreasing, or one is stochastically increasing and decreasing while the other is arbitrary.

As we shall see below, an optimal bicausal coupling under this monotonicity condition is given by the Knothe–Rosenblatt rearrangement, introduced independently by Rosenblatt [74] and Knothe [43]. This coupling is a multi-dimensional extension of the monotone rearrangement described in Section 2.1.1. Under the Knothe–Rosenblatt rearrangement, the first marginals are coupled monotonically and then, inductively, the conditional marginals are coupled monotonically; see [8, Figure 1] for an illustration.

**Definition 2.7** (Knothe–Rosenblatt rearrangement). Given probability measures  $\mu, \nu$  on  $\mathbb{R}^n$ , let  $U_1, \dots, U_n$  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  $k \in \{2, \dots, n\}$ , define inductively the random variables

$$X_k = F_{\mu_{x_1, \dots, x_{k-1}}}^{-1}(U_k), \quad Y_k = F_{\nu_{y_1, \dots, y_{k-1}}}^{-1}(U_k).$$

The Knothe–Rosenblatt rearrangement between the marginals  $\mu$  and  $\nu$  is then given by

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



We will make use of the following equivalent definition, which is proved in Backhoff-Veraguas, Beiglböck, Lin, and Zalashko [5, Proposition 5.8]. Similarly to Definition 2.6 of stochastic comonotonicity, we say that two functions on  $\mathbb{R}$  are *co-monotone* if they are both increasing or both decreasing.

**Proposition 2.8.** *A coupling  $\pi \in \text{Cpl}(\mu, \nu)$  is equal to  $\pi_{\mu, \nu}^{\text{KR}}$  if and only if  $\pi = \text{Law}(X, Y)$ , where*

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

for  $U_1, \dots, U_n$  independent uniform random variables on  $[0, 1]$ , and functions  $T_i, S_i: \mathbb{R}^i \rightarrow \mathbb{R}$  that are co-monotone in  $U_i$ , for each  $i \in \{1, \dots, n\}$ .

The following optimality result is a straightforward modification of [8, Proposition 3.5], which generalises earlier results from Backhoff-Veraguas et al. [5], Rüschendorf [75]. We note that, although [8, Proposition 3.5] considers a single function  $c \equiv c_k$  for all  $k \in \{1, \dots, n\}$ , only minor changes to the proof are required for our setting and we do not repeat the proof here.

**Proposition 2.9.** *For  $p \geq 1$ ,  $n \in \mathbb{N}$ , let  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^n)$  be the laws of Markov processes. Consider the bicausal optimal transport problem (2.5) with continuous functions  $c_k: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ ,  $k \in \{1, \dots, n\}$ , satisfying (2.3) and (2.4). If  $\mu$  and  $\nu$  are stochastically co-monotone, then the Knothe–Rosenblatt rearrangement attains the infimum in (2.5).*

**Remark 2.10.** We show in Example 5.11 that, when first order stochastic dominance is replaced by second order stochastic dominance for one of the marginals in the definition of stochastic comonotonicity, optimality of the Knothe–Rosenblatt rearrangement may fail.

**2.2.3. Continuous-time bicausal optimal transport.** The notions of bicausality and adapted Wasserstein distance extend to continuous-time processes.

Fix a finite time horizon  $T > 0$  and let  $\mathcal{X} = \mathcal{Y} = \Omega := C([0, T], \mathbb{R})$  be the space of real-valued continuous paths on the time interval  $[0, T]$ . Analogously to the discrete-time setting, equip the space  $\Omega$  with the  $L^p$  norm, defined by  $\|\gamma\|_p^p := \int_0^T |\gamma_t|^p dt$  for  $\gamma \in \Omega$ , and with the corresponding Borel sigma-algebra  $\mathcal{B}(\Omega)$ . Let  $\omega = (\omega_t)_{t \in [0, T]}$  denote the canonical process on  $\Omega$  and  $\mathcal{F} = (\mathcal{F}_t)_{t \in [0, T]}$  the canonical filtration, noting that  $\mathcal{F}_T = \mathcal{B}(\Omega)$ . Let  $(\omega, \bar{\omega}) = ((\omega_t, \bar{\omega}_t))_{t \in [0, T]}$  denote the canonical process on the product space  $\Omega \times \Omega$ , and equip this space with the product filtration. As in the discrete-time case, for a random variable  $X$  on  $(\Omega, \mathcal{F}_T)$  with law  $\eta$ , the completion of its natural filtration under  $\eta$  is denoted by  $\mathcal{F}^X = (\mathcal{F}_t^X)_{t \in [0, T]}$ .

Let  $\mu, \nu \in \mathcal{P}_p(\Omega)$ , for some  $p \geq 1$ . We say that a coupling  $\pi \in \text{Cpl}(\mu, \nu)$  is bicausal if it satisfies the continuous-time analogue of Definition 2.2; i.e.  $\pi$  is causal if (2.2) holds for all  $t \in [0, T]$ , for any random variables  $X, Y$  with  $\pi = \text{Law}(X, Y)$ , and  $\pi$  is bicausal if (2.2) also holds with the roles of  $X$  and  $Y$  reversed. Again we write  $\text{Cpl}_{\text{bc}}(\mu, \nu)$  for the set of bicausal couplings. Note that, by [45, Example 1(v), Lemma 4, and Proposition 2], our definition is equivalent to Backhoff-Veraguas et al. [6, Definition 1.2] and Lassalle [45, Definition 1].

In the continuous-time setting, we consider cost functions  $C: \Omega \times \Omega \rightarrow \mathbb{R}$  in (OT) that take the form  $C(\omega, \bar{\omega}) = \int_0^T c_t(\omega_t, \bar{\omega}_t) dt$ , for any  $\omega, \bar{\omega} \in \Omega$ , where  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is a measurable function. Additionally, we make the following assumption.

**Assumption 2.11.** Suppose that  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is continuous on  $[0, T] \times M$  for any compact set  $M \subseteq \mathbb{R} \times \mathbb{R}$ , and suppose that, for some  $p \geq 1$ , there exists  $K \geq 0$  such that, for each fixed  $t \in [0, T]$ ,  $c_t(\cdot, \cdot): \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfies (2.3) and (2.4).



For  $\mu, \nu \in \mathcal{P}_p(\Omega)$ , the continuous-time *bicausal optimal transport problem* between  $\mu$  and  $\nu$  is to find the value

$$(BCOT) \quad V_c(\mu, \nu) := \inf_{\pi \in \text{Cpl}_{bc}(\mu, \nu)} \mathbb{E}^\pi \left[ \int_0^T c_t(\omega_t, \bar{\omega}_t) dt \right].$$

We define the adapted Wasserstein distance in the continuous-time setting as follows.

**Definition 2.12** (adapted Wasserstein distance – continuous time). For  $p \geq 1$  and  $\mu, \nu \in \mathcal{P}_p(\Omega)$ , the *adapted Wasserstein distance*  $\mathcal{AW}_p(\mu, \nu)$  is defined by

$$\mathcal{AW}_p^p(\mu, \nu) := \inf_{\pi \in \text{Cpl}_{bc}(\mu, \nu)} \mathbb{E}^\pi \left[ \int_0^T |\omega_t - \bar{\omega}_t|^p dt \right].$$

The adapted Wasserstein distance  $\mathcal{AW}_2$ , in essentially the form of Definition 2.12, was first studied by Bion-Nadal and Talay [12] via a stochastic control formulation and associated Hamilton–Jacobi–Bellman equation. Backhoff-Veraguas et al. [8] studied  $\mathcal{AW}_p$  as defined in Definition 2.12, for general  $p \geq 1$ . We will see in Section 7 that this distance is amenable to efficient computation, and in Section 8 we show that this distance can be applied to robust stochastic optimisation. For semimartingales, Backhoff-Veraguas et al. [6] consider an alternative definition of adapted Wasserstein distance based on the Doob decomposition. While this definition also yields stability of various stochastic optimisation problems, we are not aware of efficient methods for the numerical computation of such a distance.

We consider the problem (BCOT) between the laws of stochastic processes that are described by SDEs. To the best of our knowledge, for such continuous-time problems, Bion-Nadal and Talay [12] and Backhoff-Veraguas et al. [6] were the first to find an explicit optimiser, the so-called *synchronous coupling*, which we define after the following preparations.

**Definition 2.13.** For an adapted process  $\rho = (\rho_t)_{t \in [0, T]}$  taking values in  $[-1, 1]$ , we call  $(W, \bar{W})$  a  $\rho$ -correlated Brownian motion if  $W, \bar{W}$  are each one-dimensional Brownian motions and, for each  $t \in [0, T]$ , their correlation is given by  $\text{corr}(W_t, \bar{W}_t) = \rho_t$ .

Let  $b, \bar{b}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma, \bar{\sigma}: [0, T] \times \mathbb{R} \rightarrow [0, \infty)$  be measurable functions, and let  $x_0 \in \mathbb{R}$ . Suppose that there exists a unique strong solution  $(X, \bar{X})$  of the system of SDEs

$$(2.6) \quad \begin{cases} dX_t = b_t(X_t) dt + \sigma_t(X_t) dW_t, & X_0 = x_0, \\ d\bar{X}_t = \bar{b}_t(\bar{X}_t) dt + \bar{\sigma}_t(\bar{X}_t) d\bar{W}_t, & \bar{X}_0 = x_0, \end{cases}$$

on the time interval  $[0, T]$ , where  $(W, \bar{W})$  is a correlated Brownian motion, and write  $\mu, \nu$  for the laws of  $X, \bar{X}$ , respectively. It is shown in [8, Proposition 2.2] that the set of bicausal couplings  $\text{Cpl}_{bc}(\mu, \nu)$  is equal to the set of all couplings of the form  $\pi = \text{Law}(X, \bar{X})$ , where (2.6) is driven by some correlated Brownian motion. In particular, the synchronous coupling is defined as follows.

**Definition 2.14.** Let  $(X, \bar{X})$  be the unique strong solution of (2.6) with  $W = \bar{W}$ . Then the *synchronous coupling* is defined as  $\pi_{\mu, \nu}^{\text{sync}} := \text{Law}(X, \bar{X})$ , the joint law of the solutions when the SDEs are driven by a common Brownian motion.

**2.3. Numerical methods for SDEs.** In order to solve SDEs of the form (1.1), one usually has to resort to numerical methods. The simplest of these is the *Euler–Maruyama scheme*. Let  $N \in \mathbb{N}$  and  $h = T/N$ . We suppose here and throughout the paper that  $N > T$  so that  $h < 1$ . The Euler–Maruyama scheme  $(X_{kh}^{h, \text{EM}})_{k \in \{0, \dots, N\}}$  is defined by  $X_0^{h, \text{EM}} = x_0$  and, for all  $k \in \{0, \dots, N-1\}$

$$(2.7) \quad X_{(k+1)h}^{h, \text{EM}} = X_{kh}^{h, \text{EM}} + b_{kh}(X_{kh}^{h, \text{EM}})h + \sigma_{kh}(X_{kh}^{h, \text{EM}})\Delta W_{k+1},$$

where  $\Delta W_{k+1} := W_{(k+1)h} - W_{kh} \sim \mathcal{N}(0, h)$ . One can also define a *semi-implicit* Euler–Maruyama scheme  $(X_{kh}^{h, \text{IEM}})_{k \in \{0, \dots, N\}}$  by  $X_0^{h, \text{IEM}} = x_0$  and, for all  $k \in \{0, \dots, N-1\}$ ,

$$(2.8) \quad X_{(k+1)h}^{h, \text{IEM}} = X_{kh}^{h, \text{IEM}} + b_{kh}(X_{(k+1)h}^{h, \text{IEM}})h + \sigma_{kh}(X_{kh}^{h, \text{IEM}})\Delta W_{k+1}.$$

Under global Lipschitz conditions on the coefficients, both of the above schemes converge in the strong  $L^p$ -sense to the true solution of the SDE; see, for example, Kloeden and Platen [42]. In recent years, numerical methods for SDEs with irregular (non-globally Lipschitz continuous) coefficients have been studied intensively. We refer to Remarks 4.4 and 5.2 and Examples 7.3 and 7.4 for references to relevant results.

The *monotone Euler–Maruyama scheme* is introduced in [8, Definition 3.13]. This scheme is defined analogously to (2.7) with the Brownian increments  $\Delta W_{k+1}$  replaced by *truncated Brownian increments*  $\Delta W_{k+1}^h$ , defined in Definition 2.15 below. Similarly, we define the *monotone semi-implicit Euler–Maruyama scheme* by replacing the Brownian increments in (2.8) by truncated Brownian increments.

**Definition 2.15** (truncated Brownian motion; cf. [54, §2.1]). Let  $(W_t)_{t \in [0, T]}$  be a standard Brownian motion, let  $N \in \mathbb{N}$  and set  $h = T/N$ . Define  $A_h = 2\sqrt{-h \log h}$  and, for each  $k \in \{0, \dots, N-1\}$ , define the stopping time  $\tau_k^h := \inf\{t \geq kh : |W_t - W_{kh}| > A_h\}$ . The *truncated Brownian motion*  $(W_t^h)_{t \in [0, T]}$  is defined by  $W_0^h = 0$  and  $W_t^h := W_{kh}^h + W_{t \wedge \tau_k^h} - W_{kh}$ , for  $t \in (kh, (k+1)h]$ ,  $k \in \{0, \dots, N-1\}$ . The *truncated Brownian increments* are defined by  $\Delta W_{k+1}^h := W_{(k+1)h}^h - W_{kh}^h = W_{(k+1)h \wedge \tau_k^h} - W_{kh}$ ,  $k \in \{0, \dots, N-1\}$ .

If the coefficients  $b$  and  $\sigma$  of the SDE (1.1) are time-homogeneous and globally Lipschitz continuous, [8, Lemma 3.14] shows that, in contrast to (2.7), the monotone Euler–Maruyama scheme is stochastically increasing for sufficiently small step size. The scheme also converges strongly to the solution of (1.1); in particular, [8, Proposition 3.16] shows that the linear interpolation of the monotone Euler–Maruyama scheme converges in  $L^p$  uniformly in time. Combined with a stability argument, these properties are key to the proof of optimality of the synchronous coupling in [8, Proposition 3.19]. Similar numerical schemes based on truncated Brownian increments are also studied in Liu and Pagès [50] and Jourdain and Pagès [38], with applications to (monotone) convex ordering of SDEs.

A different perspective on the interplay between optimal transport and numerical methods for SDEs is given by Davie [21].

### 3. PRELIMINARY RESULTS

We present two general stability results for bicausal optimal transport, which we will use in our optimality proofs. The first allows us to pass from discrete to continuous time, and the second allows us to approximate the coefficients of SDEs. As an immediate consequence, we obtain a first generalisation of [8, Theorem 1.3], which we will later apply in combination with a transformation of SDEs in order to prove Theorem 5.8. The proofs of the results in this preliminary section are presented in Appendix A.

**Assumption 3.1.** Suppose that  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  satisfy the following assumptions:

- (i)  $(X, \bar{X})$  are the unique strong solutions of the SDEs (2.6) driven by some correlated Brownian motion  $(W, \bar{W})$ , where  $X = (X_t)_{t \in [0, T]}$ ,  $\bar{X} = (\bar{X}_t)_{t \in [0, T]}$ , with  $X_0 = \bar{X}_0 = x_0$ ;
- (ii) for any  $N \in \mathbb{N}$  with  $N > T$  and  $h = T/N$ ,  $X^h = (X_{kh})_{k \in \{0, \dots, N\}}$  and  $\bar{X}^h = (\bar{X}_{kh})_{k \in \{0, \dots, N\}}$ , with  $X_0^h = \bar{X}_0^h = x_0$ ;

(iii) there exist measurable maps

$$\phi, \bar{\phi}: (0, 1) \times \mathbb{N} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}, \quad \text{and} \quad f: (0, 1) \times \mathbb{N} \times C([0, T], \mathbb{R}) \rightarrow \mathbb{R},$$

such that  $\phi$  and  $\bar{\phi}$  are co-monotone in their final argument and, for any correlated Brownian motion  $(W, \bar{W})$ , any  $N \in \mathbb{N}$  with  $N > T$  and  $h = T/N$ , and any  $k \in \{0, \dots, N-1\}$ ,

$$X_{(k+1)h}^h = \phi(h, k, X_{kh}^h, f(h, k, W)), \quad \bar{X}_{(k+1)h}^h = \bar{\phi}(h, k, \bar{X}_{kh}^h, f(h, k, \bar{W}));$$

(iv) there exists  $N_0 \in \mathbb{N}$  with  $N_0 > T$  such that, for any  $N \in \mathbb{N}$  with  $N \geq N_0$  and  $h = T/N$ ,  $X^h$  and  $\bar{X}^h$  are stochastically co-monotone (see Definition 2.6);

(v) in the case that  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  are driven by the same correlated Brownian motion  $(W, \bar{W})$ , for some  $p \geq 1$ , the linear interpolation of  $X^h$  (resp.  $\bar{X}^h$ ) converges in  $L^p$  to  $X$  (resp.  $\bar{X}$ ); i.e.

$$\lim_{h \rightarrow 0} \mathbb{E} \left[ \sum_{k=1}^N \int_{(k-1)h}^{kh} |X_{kh}^h - X_s|^p ds \right] = \lim_{h \rightarrow 0} \mathbb{E} \left[ \sum_{k=1}^N \int_{(k-1)h}^{kh} |\bar{X}_{kh}^h - \bar{X}_s|^p ds \right] = 0.$$

If  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  satisfy Assumption 3.1, then we say that  $X^h$  (resp.  $\bar{X}^h$ ) is a numerical scheme for  $X$  (resp.  $\bar{X}$ ).

**Example 3.2.** Suppose that  $(b, \sigma)$  are time-homogeneous and Lipschitz.

(1) The monotone Euler–Maruyama scheme  $X^h$  for  $X$  corresponds to taking  $\phi, f$  in Assumption 3.1 such that, for any Brownian motion  $B$ , any  $N \in \mathbb{N}$  sufficiently large,  $h = T/N$  and  $k \in \{0, \dots, N-1\}$ ,  $f(h, k, B) = \Delta B_{k+1}^h$ , the truncated Brownian increment defined in Definition 2.15, and

$$\phi(h, k, x, f(h, k, B)) = x + b(x)h + \sigma(x)f(h, k, B).$$

(2) The semi-implicit monotone Euler–Maruyama scheme  $X^h$  for  $X$  corresponds to taking  $\phi, f$  in Assumption 3.1 such that, for any Brownian motion  $B$ , any  $N \in \mathbb{N}$  sufficiently large,  $h = T/N$  and  $k \in \{0, \dots, N-1\}$ ,  $f(h, k, B) = \Delta B_{k+1}^h$  and

$$\phi(h, k, x, f(h, k, B)) = (\text{id} - bh)^{-1}(x + \sigma(x)f(h, k, B)).$$

**Proposition 3.3.** Suppose that  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  satisfy Assumption 3.1 for some  $N_0 \in \mathbb{N}$  and  $p \geq 1$ , and write  $\mu = \text{Law}(X)$ ,  $\nu = \text{Law}(\bar{X})$ ,  $\mu^h = \text{Law}(X^h)$ ,  $\nu^h = \text{Law}(\bar{X}^h)$ . Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11 with the same value of  $p$ . Then

$$(3.1) \quad \lim_{h \rightarrow 0} \inf_{\pi^h \in \text{Cpl}_{\text{bc}}(\mu^h, \nu^h)} \mathbb{E}^{\pi^h} \left[ h \sum_{k=1}^N c_{kh}(x_k, y_k) \right] = \inf_{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)} \mathbb{E}^{\pi} \left[ \int_0^T c_t(\omega_t, \bar{\omega}_t) dt \right],$$

and the synchronous coupling  $\pi_{\mu, \nu}^{\text{sync}}$  attains the infimum on the right hand side.

**Lemma 3.4.** Let  $b, \bar{b}, b^N, \bar{b}^N: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma, \bar{\sigma}, \sigma^N, \bar{\sigma}^N: [0, T] \times \mathbb{R} \rightarrow [0, \infty)$  be measurable functions with uniform linear growth in space. For each  $t \in [0, T]$ , suppose that  $b_t, \bar{b}_t, \sigma_t, \bar{\sigma}_t$  are continuous in space and that the following convergence holds uniformly on compact sets as  $N \rightarrow \infty$ :

$$b_t^N \rightarrow b_t, \quad \bar{b}_t^N \rightarrow \bar{b}_t, \quad \sigma_t^N \rightarrow \sigma_t, \quad \bar{\sigma}_t^N \rightarrow \bar{\sigma}_t.$$

Suppose moreover that, for coefficients  $(b, \sigma)$ ,  $(\bar{b}, \bar{\sigma})$ , there exist unique strong solutions  $(X, \bar{X})$  of the SDEs (2.6), and that for coefficients  $(b^N, \sigma^N)$ ,  $(\bar{b}^N, \bar{\sigma}^N)$ ,  $N \in \mathbb{N}$ , there exist unique strong solutions  $(X^N, \bar{X}^N)$  of the SDEs (2.6), for each  $N \in \mathbb{N}$ . Write  $\mu = \text{Law}(X)$ ,  $\nu = \text{Law}(\bar{X})$  and  $\mu^N = \text{Law}(X^N)$ ,  $\nu^N = \text{Law}(\bar{X}^N)$ .

Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11 and suppose that, for each  $N \in \mathbb{N}$ ,  $\pi_{\mu^N, \nu^N}^{\text{sync}}$  attains the infimum  $V_c(\mu^N, \nu^N)$  in (BCOT). Then

$$\lim_{N \rightarrow \infty} V_c(\mu^N, \nu^N) = V_c(\mu, \nu),$$

and  $\pi_{\mu, \nu}^{\text{sync}}$  attains the infimum  $V_c(\mu, \nu)$ .

Under the following assumption on both  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$ , the unique strong solutions of (2.6) and their corresponding monotone Euler–Maruyama schemes satisfy Assumption 3.1 (cf. Example 3.2) and we can apply Proposition 3.3.

**Assumption 3.5.** Suppose that  $(b, \sigma)$  satisfy the following:

- (1)  $t \mapsto b_t(x), t \mapsto \sigma_t(x)$  are Lipschitz uniformly in  $x \in \mathbb{R}$ ;
- (2)  $x \mapsto b_t(x), x \mapsto \sigma_t(x)$  are Lipschitz uniformly in  $t \in [0, T]$ .

**Proposition 3.6.** Suppose that the coefficients  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  of (2.6) satisfy Assumption 3.5 and write  $\mu, \nu$  for the laws of the respective strong solutions  $X, \bar{X}$ . Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Then the synchronous coupling attains the value  $V_c(\mu, \nu)$  of the bicausal optimal transport problem (BCOT).

Combining Proposition 3.6 with the stability result Lemma 3.4 leads to the following generalisation of [8, Theorem 1.3, Remark 3.27] to the case of both time-dependent coefficients and cost functional.

**Assumption 3.7.** Suppose that  $(b, \sigma)$  satisfy the following:

- (1)  $b$  and  $\sigma$  are continuous;
- (2)  $b$  and  $\sigma$  have linear growth in  $x$ ;
- (3) pathwise uniqueness holds for the SDE (1.1) with coefficients  $(b, \sigma)$ .

**Corollary 3.8.** Suppose that the coefficients  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  of (2.6) satisfy Assumption 3.7. Then there exist unique strong solutions  $X, \bar{X}$  of the SDEs (2.6), and we let  $\mu, \nu$  denote their laws. Further, let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Then the synchronous coupling attains the value  $V_c(\mu, \nu)$  of the bicausal optimal transport problem (BCOT).

#### 4. DISCONTINUOUS DRIFT WITH EXPONENTIAL GROWTH

Consider the SDE (1.1) with time-homogeneous coefficients. We allow the drift coefficient  $b$  to exhibit both *discontinuities* and *exponential growth*. We will impose only mild piecewise continuity assumptions on  $b$ , and we will assume that the diffusion coefficient  $\sigma$  is globally Lipschitz continuous and that  $\sigma$  is non-zero only at the points of discontinuity of  $b$ . In this setting we will prove:

- existence and uniqueness of strong solutions of (1.1);
- strong  $L^p$ -convergence of a transformed semi-implicit Euler–Maruyama scheme;
- optimality of the synchronous coupling between the laws of two such SDEs for the problem (BCOT).

To the best of our knowledge, this is the first work to treat SDEs with both discontinuous and exponentially growing drift. Precisely, we make the following assumptions on the coefficients  $(b, \sigma)$ .

**Assumption 4.1.** Let  $(b, \sigma)$  be time-homogeneous; i.e.  $b_t \equiv b: \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma_t \equiv \sigma: \mathbb{R} \rightarrow [0, \infty)$ , for all  $t \in [0, T]$ . Suppose that there exists  $m \in \mathbb{N} \cup \{0\}$  and  $\xi_1, \dots, \xi_m \in \mathbb{R}$ , with  $-\infty = \xi_0 < \xi_1 < \xi_1 < \dots, \xi_m < \xi_{m+1} = \infty$ , and there exist constants  $L_b, K_b, L_\sigma, \gamma, \eta \in (0, \infty)$  such that  $(b, \sigma)$  satisfy the following:

- (A1) on each interval  $(\xi_k, \xi_{k+1})$ ,  $k \in \{0, \dots, m\}$ ,

- (i)  $b$  is one-sided Lipschitz; i.e.  $(x - y)(b(x) - b(y)) \leq L_b|x - y|^2$ , for all  $x, y \in (\xi_k, \xi_{k+1})$ ,
- (ii)  $b$  satisfies the following exponential growth and local Lipschitz condition: for all  $x, y \in (\xi_k, \xi_{k+1})$ ,

$$|b(x) - b(y)| \leq K_b(\exp\{\gamma|x|^\eta\} + \exp\{\gamma|y|^\eta\})|x - y|;$$

(A2)  $\sigma$  is globally Lipschitz; i.e.  $|\sigma(x) - \sigma(y)| \leq L_\sigma|x - y|$  for all  $x, y \in \mathbb{R}$ ;

(A3)  $\sigma(\xi_k) \neq 0$ , for all  $k \in \{1, \dots, m\}$ .

**Remark 4.2.** We note that we do not impose any uniform non-degeneracy condition on the diffusion coefficient  $\sigma$ . This is in contrast to the setting of Dareiotis, Gerencsér, and Lê [20], for example.

**Example 4.3.** Consider the introductory example from [35]:

$$(4.1) \quad dX_t = -X_t^3 dt + dW_t, \quad X_0 = x_0 \in \mathbb{R},$$

and a modification of the example in [81, Section 2]:

$$(4.2) \quad dX_t = \left( \frac{1}{2} - 2\text{sign}(X_t - 1) \right) dt + |X_t| dW_t, \quad X_0 = x_0 \in \mathbb{R}.$$

A more practical example is the value of a company in a Black–Scholes market, with parameters  $\mu, \rho \in (0, \infty)$ , that pays out dividends following a threshold strategy; i.e. dividends are paid out at a rate  $u \in (0, \infty)$  whenever the value of the company exceeds a certain threshold level  $\ell \in (0, \infty)$ :

$$(4.3) \quad dX_t = (\mu - u\mathbf{1}_{\{X_t \geq \ell\}})X_t dt + \rho X_t dW_t, \quad X_0 = x_0 \in (0, \infty).$$

The coefficients of the SDEs (4.1), (4.2), and (4.3) each satisfy Assumption 4.1.

**Remark 4.4.**

- (i) In the case that  $m = 0$ , the drift coefficient  $b$  is continuous, and there is no non-degeneracy condition on  $\sigma$ . In this setting, convergence in probability and almost sure convergence of the explicit Euler–Maruyama scheme were shown by Gyöngy and Krylov [28] and Gyöngy [27], respectively. However, Hutzenthaler et al. [35, Theorem 1] shows that, for the SDE (4.1), this scheme has unbounded moments and diverges in the strong  $L^p$  sense. Two approaches to recover moment bounds and strong convergence have been taken in the literature.
  - (a) Tamed numerical schemes, in which polynomially growing coefficients are appropriately rescaled, have been studied, for example, by Hutzenthaler, Jentzen, and Kloeden [36], Sabanis [76]. These are explicit schemes, which are computationally efficient. However, the technique of taming does not readily yield a scheme satisfying the monotonicity property Assumption 3.1 (iv) and to our knowledge has not been applied to exponentially growing coefficients.
  - (b) For a semi-implicit Euler–Maruyama scheme, Hu [33] proves  $L^2$ -convergence of order  $1/2$  at the terminal time. When  $b$  has at most polynomial growth, Higham, Mao, and Stuart [31] prove uniform-in-time  $L^2$ -convergence at the same rate. Under additional assumptions, Hutzenthaler and Jentzen [34], Mao and Szpruch [51, 52] also study strong convergence of this scheme. However, under Assumption 4.1 with  $m = 0$ , we are not aware of any existing results on uniform-in-time  $L^p$ -convergence.
- (ii) Strengthening Assumption 4.1 (A1) to piecewise Lipschitz continuity, we are in the setting of Leobacher and Szölgényi [46, 47], who were the first to consider transformation-based schemes for SDEs with piecewise-Lipschitz drift. Under these assumptions, the authors

proved strong well-posedness of the SDEs and strong  $L^p$ -convergence of a transformation-based Euler–Maruyama scheme. Leobacher and Szölgyenyi [49], Müller-Gronbach and Yaroslavl'tseva [56] also prove strong convergence rates for the classical Euler–Maruyama scheme. Higher-order transformation-based schemes and adaptive schemes have been introduced in Müller-Gronbach and Yaroslavl'tseva [57], Neuenkirch, Szölgyenyi, and Szpruch [61], Przybyłowicz, Schwarz, and Szölgyenyi [71], Yaroslavl'tseva [87]. Przybyłowicz and Szölgyenyi [69], Przybyłowicz, Szölgyenyi, and Xu [70] additionally allow for processes with jumps. Overviews of this line of research can be found in Leobacher and Szölgyenyi [48], Szölgyenyi [81].

- (iii) Recent results of Müller-Gronbach, Sabanis, and Yaroslavl'tseva [59], Spendier and Szölgyenyi [79], as well as Hu and Gan [32], prove existence and uniqueness of strong solutions and convergence of tamed Euler–Maruyama schemes for SDEs with drift coefficients that are both discontinuous and superlinearly growing. However, these results impose a polynomial growth condition on the drift coefficient, whereas we allow for exponential growth. We note, however, that [59] allows for both the drift and diffusion coefficient to have polynomial growth.

**4.1. A transformed SDE.** In order to remove the discontinuities from the drift coefficient of the SDE (1.1), we define a transformation of space  $G: \mathbb{R} \rightarrow \mathbb{R}$  as in Leobacher and Szölgyenyi [47].

First we define  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  by

$$\phi(u) = \begin{cases} (1+u)^3(1-u)^3 & |u| \leq 1, \\ 0 & |u| > 1. \end{cases}$$

Then, for constants  $c_0 \in (0, \infty)$  and  $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R} \setminus \{0\}$ , define  $\bar{\phi}_k: \mathbb{R} \rightarrow \mathbb{R}$ , for each  $k \in \{1, \dots, m\}$ , by

$$(4.4) \quad \bar{\phi}_k(x) := \phi\left(\frac{x - \xi_k}{c_0}\right)(x - \xi_k)|x - \xi_k|, \quad \text{for } x \in \mathbb{R},$$

and define  $G: \mathbb{R} \rightarrow \mathbb{R}$  by

$$(4.5) \quad G(x) := x + \sum_{k=1}^m \alpha_k \bar{\phi}_k(x), \quad \text{for } x \in \mathbb{R}.$$

Before fixing the constants in the definition of  $G$ , we make the following remark.

**Remark 4.5.** On the finite intervals  $(\xi_k, \xi_{k+1})$ , for  $k \in \{1, \dots, m-1\}$ , the local Lipschitz constant in Assumption 4.1 (A1) (i) is bounded by  $2K_b \exp\{\gamma(|\xi_k| \vee |\xi_{k+1}|)^\eta\} < \infty$ , and thus  $b$  extends to a Lipschitz function on the compact interval  $[\xi_k, \xi_{k+1}]$ . In particular, there exist finite one-sided limits  $b(\xi_k+) := \lim_{x \searrow \xi_k} b(x)$  and  $b(\xi_{k+1}-) := \lim_{x \nearrow \xi_{k+1}} b(x)$ . Similarly,  $b$  extends to a Lipschitz function on  $[\xi_1 - c, \xi_1]$  and  $[\xi_m, \xi_m + c]$ , for any  $c \in \mathbb{R}$ , and the one-sided limits  $b(\xi_1-), b(\xi_m+)$  exist and are finite.

By Remark 4.5 and Assumption 4.1 (A3), for each  $k \in \{1, \dots, m\}$  we can now define

$$(4.6) \quad \alpha_k := \frac{1}{2} \sigma(\xi_k)^{-2} (b(\xi_k-) - b(\xi_k+)),$$

and choose

$$(4.7) \quad c_0 < \min_{k \in \{1, \dots, m\}} \left\{ \frac{1}{6|\alpha_k|} \right\} \wedge \min_{k \in \{1, \dots, m-1\}} \left\{ \frac{1}{2} (\xi_{k+1} - \xi_k) \right\}.$$

We will see that this choice of constants ensures that the transformation  $G$  acts only locally at each discontinuity, is strictly increasing, and the drift of the transformed SDE is continuous. We prove the following properties of the transformation  $G$  in Appendix B.

**Lemma 4.6.** *For each  $k \in \{1, \dots, m\}$ ,  $\bar{\phi}_k$  defined in (4.4) with  $c_0$  chosen according to (4.7) satisfies the following:*

- $\bar{\phi}_k(x) = \bar{\phi}'_k(x) = \bar{\phi}''_k(x) = 0$  for  $|x - \xi_k| \geq c_0$ ,
- $\bar{\phi}_k(\xi_k) = \bar{\phi}'_k(\xi_k) = 0$ ,  $\bar{\phi}''_k(\xi_k-) = -2$ ,  $\bar{\phi}''_k(\xi_k+) = 2$ ,
- $\bar{\phi}'_k(x) \in [-6c_0, 6c_0]$ , for all  $x \in \mathbb{R}$ ,
- $\bar{\phi}_k: \mathbb{R} \rightarrow \mathbb{R}$  is Lipschitz, has Lipschitz first derivative, and has piecewise Lipschitz almost-everywhere second derivative with discontinuity point  $\xi_k$ .

**Lemma 4.7** (Cf. [47, Lemma 2.2]). *The transformation  $G$  defined by (4.5), (4.6), and (4.7) satisfies the following:*

- $G = \text{id}$  on the set  $\mathbb{R} \setminus \bigcup_{k \in \{1, \dots, m\}} (\xi_k - c_0, \xi_k + c_0)$ ,
- $G$  is strictly increasing with strictly increasing inverse  $G^{-1}: \mathbb{R} \rightarrow \mathbb{R}$ ,
- $G, G^{-1}$  are bounded and Lipschitz with Lipschitz constants  $L_G, L_{G^{-1}} \in (0, \infty)$ ,
- $G, G^{-1}$  have bounded Lipschitz first derivatives  $G', (G^{-1})'$ , and have bounded piecewise Lipschitz almost-everywhere second derivatives  $G'', (G^{-1})''$  with discontinuity points  $\xi_1, \dots, \xi_m$ .

**Example 4.8.** Suppose that  $m = 1$ ; i.e. there is only a single exceptional point  $\xi \in \mathbb{R}$  at which  $b$  is not Lipschitz. Figure 2 illustrates that  $G$  is close to linear, increasing, and has Lipschitz first derivative and Lipschitz-almost-everywhere second derivative.

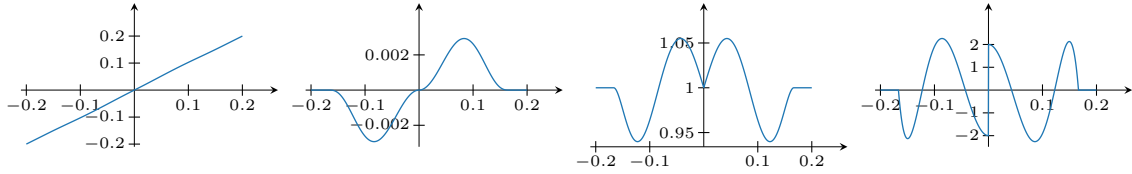


FIGURE 2. From left to right: the functions  $G$ ,  $G - \text{id}$ ,  $G'$ , and  $G''$ , for  $\xi = 0$ ,  $\alpha = 1$ ,  $c = 1/6$ .

Supposing that there exists a strong solution  $X$  of (1.1), we define  $Z_t := G(X_t)$ , for all  $t \in [0, T]$ . Formally, applying Itô's formula, we find that  $Z$  solves the SDE

$$(4.8) \quad dZ_t = \tilde{b}(Z_t) dt + \tilde{\sigma}(Z_t) dW_t, \quad Z_0 = G(x_0),$$

where  $\tilde{b}: \mathbb{R} \rightarrow \mathbb{R}$ ,  $\tilde{\sigma}: \mathbb{R} \rightarrow [0, \infty)$  are defined by

$$(4.9) \quad \begin{aligned} \tilde{b} &:= (bG') \circ G^{-1} + \frac{1}{2}(\sigma^2 G'') \circ G^{-1}, \\ \tilde{\sigma} &:= (\sigma G') \circ G^{-1}. \end{aligned}$$

Note that, by Lemma 4.7,  $\tilde{b} = b$  and  $\tilde{\sigma} = \sigma$  on the set  $\mathbb{R} \setminus \bigcup_{k \in \{1, \dots, m\}} (\xi_k - c_0, \xi_k + c_0)$ . That is, the transformation only affects the coefficients on a neighbourhood of each discontinuity point.

To prove regularity of the transformed coefficients, we will make use of the following lemma, which is analogous to [46, Lemma 2.6]. Again we postpone the proof to Appendix B.

**Lemma 4.9.** *Suppose that  $f: \mathbb{R} \rightarrow \mathbb{R}$  satisfies Assumption 4.1 (A1) and is continuous. Then  $f$  satisfies Assumption 4.1 (A1) (i) and (ii) globally on  $\mathbb{R}$ .*

We say that coefficients  $(b, \sigma)$  satisfy the *monotonicity condition* if there exists  $\alpha, \beta \in (0, \infty)$  such that, for all  $x \in \mathbb{R}$ ,

$$(4.10) \quad xb(x) \vee |\sigma(x)|^2 \leq \alpha + \beta|x|^2.$$



We let  $L^1_{\text{loc}}(\mathbb{R})$  denote the space of locally integrable functions on  $\mathbb{R}$ . The proof of the following properties of the transformed coefficients is once again found in Appendix B.

**Lemma 4.10.** *Suppose that  $(b, \sigma)$  satisfy Assumption 4.1. Then there exist constants  $L_{\tilde{b}}, K_{\tilde{b}}, L_{\tilde{\sigma}}, \tilde{\gamma}, \tilde{\eta}, \alpha, \beta \in (0, \infty)$  such that the transformed coefficients  $(\tilde{b}, \tilde{\sigma})$  defined in (4.9) satisfy the following:*

- (i)  $\tilde{b}$  is one-sided Lipschitz; i.e.  $(x - y)(\tilde{b}(x) - \tilde{b}(y)) \leq L_{\tilde{b}}|x - y|^2$ , for all  $x, y \in \mathbb{R}$ ;
  - (ii)  $\tilde{b}$  satisfies the following exponential growth and local Lipschitz condition: for all  $x, y \in \mathbb{R}$ ,
- $$(4.11) \quad |\tilde{b}(x) - \tilde{b}(y)| \leq K_{\tilde{b}}(\exp\{\tilde{\gamma}|x|^{\tilde{\eta}}\} + \exp\{\tilde{\gamma}|y|^{\tilde{\eta}}\})|x - y|;$$
- (iii)  $\tilde{b}$  is locally absolutely continuous; i.e. there exists  $\tilde{b}' \in L^1_{\text{loc}}(\mathbb{R})$  such that, for any  $x, y \in \mathbb{R}$ ,
- $$\tilde{b}(y) - \tilde{b}(x) = \int_x^y \tilde{b}'(r) dr;$$
- (iv)  $\tilde{\sigma}$  is globally Lipschitz; i.e.  $|\tilde{\sigma}(x) - \tilde{\sigma}(y)| \leq L_{\tilde{\sigma}}|x - y|$ , for all  $x, y \in \mathbb{R}$ ;
  - (v)  $(\tilde{b}, \tilde{\sigma})$  satisfy the monotonicity condition (4.10) with constants  $\alpha, \beta$ .

The following result gives well-posedness of the transformed SDE, as well as moment bounds and one-step error bounds that will be used in the convergence analysis of the numerical scheme.

**Proposition 4.11.** *Suppose that  $(b, \sigma)$  satisfy Assumption 4.1. Then the SDE (4.8) admits a unique strong solution  $(Z_t)_{t \in [0, T]}$ . Moreover, for any  $p \geq 1$ , there exist constants  $C_0(p, T), C_1(p, T), C_2(p, T) > 0$  such that*

$$(4.12) \quad \mathbb{E} \left[ \sup_{t \in [0, T]} |Z_t|^p \right] \leq C_0(p, T)$$

and, for all  $s, t \in [0, T]$  with  $s < t$ ,

$$(4.13) \quad \mathbb{E} \left[ \sup_{r \in [s, t]} |Z_r - Z_s|^p \right] \leq C_1(p, T)(t - s)^{\frac{p}{2}}$$

and

$$(4.14) \quad \mathbb{E} \left[ \sup_{r \in [s, t]} |\tilde{b}(Z_r) - \tilde{b}(Z_s)|^p \right] \leq C_2(p, T)(t - s)^{\frac{p}{2}}.$$

*Proof.* By Lemma 4.10,  $(\tilde{b}, \tilde{\sigma})$  satisfy the monotonicity condition (4.10),  $\tilde{b}$  is one-sided Lipschitz, and  $\tilde{\sigma}$  is Lipschitz. Under these conditions, the classical result of Krylov [44, Theorem 1] shows existence and uniqueness of a strong solution  $Z$  of (4.8). The monotonicity condition (4.10) also implies moment bounds (4.12) on the solution, as shown, for example, in [31, Lemma 3.2].

Given that  $\tilde{b}$  also satisfies the exponential growth condition (4.11), we can follow the proofs of Hu [33, Lemma 3.4, Theorem 3.5] to prove the bounds (4.13) and (4.14). We note that Hu [33] assumes that the coefficients  $(\tilde{b}, \tilde{\sigma})$  are continuously differentiable, but an inspection of the proofs reveals that no such assumptions are necessary; in particular, Lemma 4.10 shows that the fundamental theorem of calculus holds.  $\square$

We now prove our existence and uniqueness result, following the proof of [47, Theorem 2.6].

**Theorem 4.12.** *Suppose that the coefficients  $(b, \sigma)$  satisfy Assumption 4.1. Then there exists a unique strong solution  $X$  of (1.1) with coefficients  $(b, \sigma)$  and initial condition  $X_0 = x_0 \in \mathbb{R}$ . Moreover, for any  $p \geq 1$ , there exists a constant  $C(p, T) > 0$  such that*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t|^p \right] \leq C(p, T).$$

*Proof.* By Proposition 4.11, there exists a unique strong solution  $Z$  of the SDE (4.8). Now define  $X_t = G^{-1}(Z_t)$ , for all  $t \in [0, T]$ . By Lemma 4.7,  $G^{-1}$  is Lipschitz with Lipschitz constant  $L_{G^{-1}}$ , and by Proposition 4.11,  $Z$  satisfies the one-step error bound (4.13). Hence, for  $p \geq 1$ ,

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t|^p \right] \leq 2^{p-1} L_{G^{-1}}^p \mathbb{E} \left[ \sup_{t \in [0, T]} |Z_t - Z_0|^p \right] + 2^{p-1} |x_0|^p \leq C(p, T),$$

where  $C(p, T) := 2^{p-1} L_{G^{-1}}^p C_1(p, T) T^{\frac{p}{2}} + 2^{p-1} |x_0|^p$ .

By Lemma 4.7,  $G^{-1}: \mathbb{R} \rightarrow \mathbb{R}$  is continuously differentiable and has almost everywhere second derivative that is locally integrable. Thus we may apply Itô's formula (see, e.g. [73, §VI Theorem 1.5 and Remarks]) to see that  $X$  satisfies the SDE (1.1) with coefficients  $(b, \sigma)$ , and  $X_0 = G^{-1}(Z_0) = x_0$ . Since  $Z$  is a strong solution, and  $X$  is a deterministic transformation of  $Z$  at each time,  $X$  is also a strong solution. Conversely, for any strong solution  $X'$  of (1.1) with coefficients  $(b, \sigma)$  and with  $X'_0 = x_0$ , the process  $Z'$  defined by taking  $Z'_t = G(X'_t)$  for  $t \in [0, T]$  is a strong solution of (4.8). By pathwise uniqueness,  $Z'_t = Z_t$  for all  $t \in [0, T]$  almost surely. Thus,  $X'_t = G^{-1}(Z'_t) = G^{-1}(Z_t) = X_t$  for all  $t \in [0, T]$  almost surely. We conclude that strong existence and pathwise uniqueness hold for (1.1) with coefficients  $(b, \sigma)$ .  $\square$

**4.2. Transformed semi-implicit Euler–Maruyama scheme.** We now define a numerical scheme via the transformation introduced in the previous section.

**Definition 4.13** (transformed (monotone) semi-implicit Euler–Maruyama scheme). Fix  $N \in \mathbb{N}$ , with  $N > T$ , and set  $h = T/N$ . For the SDE (1.1) with coefficients  $(b, \sigma)$  satisfying Assumption 4.1, let  $G$  be defined as in (4.5) and  $(\tilde{b}, \tilde{\sigma})$  as in (4.9).

Define the *transformed monotone semi-implicit Euler–Maruyama scheme*  $X^h$  by  $X_0^h = x_0$  and, for  $k \in \{0, \dots, N-1\}$ ,

$$(4.15) \quad X_{(k+1)h}^h := G^{-1} \left( G(X_{kh}^h) + (\tilde{b} \circ G)(X_{(k+1)h}^h) \cdot h + (\tilde{\sigma} \circ G)(X_{kh}^h) \Delta W_{k+1}^h \right),$$

where  $\Delta W_{k+1}^h$  is a truncated Brownian increment, as in Definition 2.15.

We may define the *transformed semi-implicit Euler–Maruyama scheme*  $\tilde{X}^h$  analogously, replacing the truncated increments with the usual Brownian increments  $\Delta W_k$ .

**Remark 4.14.**

- (i) Similarly to the monotone Euler–Maruyama scheme discussed in Section 2.3, the term *monotone* in Definition 4.13 is motivated by the fact that the scheme will be shown to be stochastically monotone when  $\tilde{b}$  is one-sided Lipschitz and  $\tilde{\sigma}$  is Lipschitz; see Lemma 4.21.
- (ii) If  $b$  is continuous, then the transformation  $G$  is the identity and  $\tilde{X}^h$  is the classical semi-implicit (or backward) Euler–Maruyama scheme, as studied for example in [31, 33, 51, 52].

For the scheme defined in Definition 4.13, we will prove moment bounds and strong  $L^p$ -convergence for  $p \geq 1$ . In particular, for the semi-implicit Euler–Maruyama scheme, we generalise the moment bounds from Higham et al. [31] and the  $L^2$  convergence from Hu [33]. We note that [31, 33] also assume continuous differentiability of  $b$  and  $\sigma$ , but [33] uses this condition only for existence and uniqueness of solutions, which we have established under Assumption 4.1 in Theorem 4.12. Similarly, the result of [31] can be established without the continuous differentiability assumption in our setting.

We will repeatedly make use of the following lemma.

**Lemma 4.15.** *Suppose that  $f: \mathbb{R} \rightarrow \mathbb{R}$  is one-sided Lipschitz with constant  $L_f$  and let  $h \in (0, 1/L_f)$ . Then the map  $(\text{id} - hf)$  is strictly increasing and invertible with strictly increasing inverse.*

*Proof.* Suppose that  $x, y \in \mathbb{R}$ ,  $x < y$ . By the one-sided Lipschitz condition,

$$h(x - y)(f(x) - f(y)) \leq hL_f|x - y|^2 < |x - y|^2.$$

Then, since  $x - y < 0$ ,

$$h(f(x) - f(y)) > (x - y)^{-1}|x - y|^2 = x - y,$$

or equivalently

$$(\text{id} - hf)(x) < (\text{id} - hf)(y).$$

Hence  $(\text{id} - hf)$  is strictly increasing, and thus invertible with strictly increasing inverse.  $\square$

**Lemma 4.16.** *Suppose that the coefficients of (1.1) satisfy Assumption 4.1, let  $\tilde{b}$  be the transformed drift in (4.9), and let  $L_{\tilde{b}} \in (0, \infty)$  be as in Lemma 4.10. Then, for all  $h < 1/L_{\tilde{b}}$ , the transformed (monotone) semi-implicit Euler–Maruyama scheme (4.15) is well defined.*

*Proof.* By Lemma 4.10,  $\tilde{b}$  is one-sided Lipschitz with constant  $L_{\tilde{b}}$ . Then, for  $h < 1/L_{\tilde{b}}$ , the map  $(\text{id} - h\tilde{b})$  is invertible by Lemma 4.15. We can rewrite (4.15) as

$$(\text{id} - h\tilde{b})(G(X_{(k+1)h}^h)) = G(X_{kh}^h) + (\tilde{\sigma} \circ G)(X_{kh}^h)\Delta W_{k+1}^h,$$

and similarly in the case of non-truncated Brownian increments. Since  $G$  is also invertible by Lemma 4.7, we may conclude.  $\square$

**Proposition 4.17.** *Suppose that the coefficients of (1.1) satisfy Assumption 4.1, let  $(\tilde{b}, \tilde{\sigma})$  be the transformed coefficients in (4.9), and let  $\beta, L_{\tilde{b}} \in (0, \infty)$  be as in Lemma 4.10. Let  $N \in \mathbb{N}$  with  $N > (1 \vee 4\beta \vee 2L_{\tilde{b}})T$ , let  $h = T/N$ , and let  $X^h$  denote the transformed monotone semi-implicit Euler–Maruyama scheme defined by (4.15).*

*Then, for any  $p \geq 1$ , there exists a constant  $C(p, T) > 0$ , not depending on  $h$ , such that*

$$(4.16) \quad \mathbb{E} \left[ \sup_{k \in \{0, \dots, N\}} |X_{kh}^h|^p \right] \leq C(p, T).$$

*Analogous bounds hold for the scheme  $\tilde{X}^h$  defined with respect to non-truncated Brownian increments.*

*Proof.* Since  $h < 1/L_{\tilde{b}}$ , Lemma 4.16 implies that  $X^h$  and  $\tilde{X}^h$  are well defined.

First suppose that  $p > 4$ . For  $k \in \{0, \dots, N\}$ , define  $Z_{kh}^h := G(X_{kh}^h)$ . By Lemma 4.7,  $G^{-1}$  is Lipschitz with constant  $L_{G^{-1}}$ , and so

$$(4.17) \quad \mathbb{E} \left[ \sup_{k \in \{0, \dots, N\}} |X_{kh}^h|^p \right] \leq 3^{p-1} L_{G^{-1}}^p \mathbb{E} \left[ \sup_{k \in \{0, \dots, N\}} |Z_{kh}^h|^p \right] + 3^{p-1} L_{G^{-1}}^p |G(x_0)|^p + 3^{p-1} |x_0|^p.$$

We aim to bound the moments of  $\sup_{k \in \{0, \dots, N\}} |Z_{kh}^h|$ .

Define  $K_h := (1 - 2\beta h)^{-1} < 2$  and, for  $k \in \{0, \dots, N\}$ , define  $Y_{kh}^h := (\text{id} - h\tilde{b})(Z_{kh}^h)$ . By Lemma 4.10,  $(\tilde{b}, \tilde{\sigma})$  satisfy the monotonicity condition (4.10). As in [52, Lemma 3.2], this implies that, for any  $k \in \{0, \dots, N\}$ ,

$$(4.18) \quad |Z_{kh}^h|^2 \leq K_h (|Y_{kh}^h|^2 + 2\alpha h).$$

Therefore it suffices to bound the moments of  $\sup_{k \in \{0, \dots, N\}} |Y_{kh}^h|$ . Observe that, for any  $k \in \{0, \dots, N-1\}$ ,

$$(4.19) \quad Y_{(k+1)h}^h = Z_{kh}^h + \tilde{\sigma}(Z_{kh}^h)\Delta W_{k+1}^h,$$

and

$$Z_{kh}^h = Y_{kh}^h + h\tilde{b}(Z_{kh}^h).$$

Thus  $(Y_{kh}^h)_{k \in \{0, \dots, N\}}$  coincides with the so-called *split-step backward Euler method* described in [31, Section 3.3], started from  $Y_0^h = x_0 - h\tilde{b}(x_0)$  and with the increments  $(\Delta W_{k+1})_{k \in \{0, \dots, N-1\}}$  replaced by  $(\Delta W_{k+1}^h)_{k \in \{0, \dots, N-1\}}$ . We now closely follow the proof of [31, Lemma 3.7] to obtain moment bounds.

To ensure finite moments, for  $R > 0$ , define the stopping time

$$\theta_R := \inf\{k \in \{0, \dots, N\} : |Y_{kh}^h| > R\},$$

with the convention that  $\inf \emptyset := \infty$ . For  $k \in \{0, \dots, \theta_R - 1\}$ , we have the bound  $|Y_{kh}^h| \leq R$ . To bound the moments of  $|Y_{\theta_R h}^h|$  we first note that, by (4.18),

$$|Z_{(\theta_R-1)h}^h|^p \leq 2^{\frac{p}{2}-1} K_h^{\frac{p}{2}} (|Y_{(\theta_R-1)h}^h|^p + (2\alpha h)^{\frac{p}{2}}) \leq 2^{p-1} (R^p + (2\alpha)^{\frac{p}{2}}) =: C(p, R).$$

Note that, in Definition 2.15 of the truncation Brownian motion, we have  $A_h < 2$ . Then, combining (4.19) with the monotonicity condition (4.10) proved in Lemma 4.10, we have

$$\begin{aligned} \mathbb{E}[|Y_{\theta_R h}^h|^p] &\leq 2^{p-1} \mathbb{E}[|Z_{(\theta_R-1)h}^h|^p] + 2^{p-1} A_h^p \mathbb{E}[|\tilde{\sigma}(Z_{(\theta_R-1)h}^h)|^p] \\ &\leq 2^{p-1} \mathbb{E}[|Z_{(\theta_R-1)h}^h|^p] + 2^{p-1} 2^{\frac{p}{2}-1} A_h^p (\alpha^{\frac{p}{2}} + \beta^{\frac{p}{2}} \mathbb{E}[|Z_{(\theta_R-1)h}^h|^p]) \\ &\leq 2^{p-1} C(p, R)^p (1 + 2^{\frac{3p}{2}-1} \beta^{\frac{p}{2}}) + 2^{\frac{5p}{2}-2} \alpha^{\frac{p}{2}} =: \tilde{C}(p, R). \end{aligned}$$

Hence

$$(4.20) \quad \sup_{k \in \{0, \dots, N\}} \mathbb{E}[|Y_{(k \wedge \theta_R)h}^h|^p] \leq \max\{R^p, \tilde{C}(p, R)\} < \infty.$$

Combining (4.18) with (4.19) gives

$$\begin{aligned} |Y_{(k+1)h}^h|^2 &= |Z_{kh}^h|^2 + 2Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h + |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2 \\ &\leq K_h |Y_{kh}^h|^2 + 2\alpha h K_h + 2Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h + |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2, \end{aligned}$$

for any  $k \in \{0, \dots, N-1\}$ . For all  $\ell \in \{1, \dots, N\}$ , noting that  $K_h = 1 + 2\beta h K_h$ , and iterating the above inequality, we have

$$\begin{aligned} |Y_{(\ell \wedge \theta_R)h}^h|^2 &\leq |Y_0^h|^2 + 2\beta h K_h \sum_{k=0}^{(\ell \wedge \theta_R)-1} |Y_{kh}^h|^2 + 2\alpha T K_h + 2 \sum_{k=0}^{(\ell \wedge \theta_R)-1} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h \\ &\quad + \sum_{k=0}^{(\ell \wedge \theta_R)-1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2. \end{aligned}$$

We now raise each side to the power  $p/2$  and use the inequality  $(a_1 + a_2 + a_3 + a_4)^{\frac{p}{2}} \leq 4^{\frac{p}{2}-1} (|a_1|^{\frac{p}{2}} + |a_2|^{\frac{p}{2}} + |a_3|^{\frac{p}{2}} + |a_4|^{\frac{p}{2}})$ , for any  $a_1, a_2, a_3, a_4 \in \mathbb{R}$ . Setting  $c_0(p, T) := 2^{\frac{3p}{2}-2} (|G(x_0)|^2 + |\tilde{b}(G(x_0))|^2 + 2\alpha T)^{\frac{p}{2}}$ , we have

$$(4.21) \quad \begin{aligned} |Y_{(\ell \wedge \theta_R)h}^h|^p &\leq c_0(p, T) + 2^{\frac{3p}{2}-2} \left( \beta h K_h \sum_{k=0}^{(\ell \wedge \theta_R)-1} |Y_{kh}^h|^2 \right)^{\frac{p}{2}} + 2^{\frac{3p}{2}-2} \left| \sum_{k=0}^{(\ell \wedge \theta_R)-1} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h \right|^{\frac{p}{2}} \\ &\quad + 2^{p-2} \left( \sum_{k=0}^{(\ell \wedge \theta_R)-1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2 \right)^{\frac{p}{2}}. \end{aligned}$$

For all  $M \in \{1, \dots, N\}$ , we seek a bound for  $\mathbb{E}[\sup_{\ell \in \{0, \dots, M\}} |Y_{(\ell \wedge \theta_R)h}^h|^p]$ . We will treat each term in the sum (4.21) in turn. First, since  $p > 2$ , an application of Hölder's inequality gives

$$(4.22) \quad 2^{\frac{3p}{2}-2} \left( \beta h K_h \sum_{k=0}^{(\ell \wedge \theta_R)-1} |Y_{kh}^h|^2 \right)^{\frac{p}{2}} \leq c_1(p, T) h \sum_{k=0}^{\ell-1} |Y_{(k \wedge \theta_R)h}^h|^p,$$

where  $c_1(p, T) := 2^{2p-2} \beta^{\frac{p}{2}} T^{\frac{p}{2}-1}$ .

To bound the second term in (4.21), we apply the Burkholder–Davis–Gundy inequality, noting that the truncated Brownian motion  $W^h$  is an  $\mathcal{F}^W$ -martingale with quadratic variation  $\langle W^h \rangle$  satisfying the bound  $\langle W^h \rangle_t \leq \langle W \rangle_t = t$ , for all  $t \geq 0$ . Thus there exists a constant  $c_2(p) > 0$  such that

$$\begin{aligned} \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left| \sum_{k=0}^{(\ell \wedge \theta_R) - 1} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h \right|^2 \right] &= \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left| \sum_{k=0}^{(\ell \wedge \theta_R) - 1} \int_{kh}^{(k+1)h} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) dW_s^h \right|^2 \right] \\ &\leq c_2(p) \mathbb{E} \left[ \left( \sum_{k=0}^{(M \wedge \theta_R) - 1} h |Z_{kh}^h \tilde{\sigma}(Z_{kh}^h)|^2 \right)^{\frac{p}{4}} \right]. \end{aligned}$$

Now, noting that  $p > 4$ , Hölder's inequality gives

$$\mathbb{E} \left[ \left( \sum_{k=0}^{(M \wedge \theta_R) - 1} h |Z_{kh}^h \tilde{\sigma}(Z_{kh}^h)|^2 \right)^{\frac{p}{4}} \right] \leq T^{\frac{p}{4} - 1} h \sum_{k=0}^{M-1} \mathbb{E} \left[ |Z_{(k \wedge \theta_R)h}^h \tilde{\sigma}(Z_{(k \wedge \theta_R)h}^h)|^{\frac{p}{2}} \right].$$

Using the monotonicity condition (4.10) and the inequality  $a \leq \frac{1}{2}(1 + a^2)$  for any  $a \in \mathbb{R}$ , we find a constant  $c_3(p) > 0$  such that

$$\begin{aligned} |Z_{(k \wedge \theta_R)h}^h \tilde{\sigma}(Z_{(k \wedge \theta_R)h}^h)|^{\frac{p}{2}} &\leq |Z_{(k \wedge \theta_R)h}^h|^{\frac{p}{2}} (\alpha + \beta |Z_{(k \wedge \theta_R)h}^h|^2)^{\frac{p}{4}} \\ &\leq c_3(p) |Z_{(k \wedge \theta_R)h}^h|^{\frac{p}{2}} (\alpha^{\frac{p}{4}} + \beta^{\frac{p}{4}} |Z_{(k \wedge \theta_R)h}^h|^{\frac{p}{2}}) \\ &\leq c_3(p) \left[ \frac{1}{2} + \left( \frac{\alpha^{\frac{p}{2}}}{2} + \beta^{\frac{p}{4}} \right) |Z_{(k \wedge \theta_R)h}^h|^p \right], \end{aligned}$$

for all  $k \in \{0, \dots, N\}$ . Combining the previous estimates and setting  $c_4(p, T) := 2^{-1} T^{\frac{p}{4}} c_2(p) c_3(p)$  and  $c_5(p, T) := (2^{-1} \alpha^{\frac{p}{2}} + \beta^{\frac{p}{4}}) T^{\frac{p}{4} - 1} c_2(p) c_3(p)$ , we arrive at

$$(4.23) \quad \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left| \sum_{k=0}^{(\ell \wedge \theta_R) - 1} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h \right|^2 \right] \leq c_4(p, T) + c_5(p, T) h \sum_{k=0}^{M-1} \mathbb{E} [|Z_{(k \wedge \theta_R)h}^h|^p].$$

For the final term in (4.21), note that

$$\begin{aligned} (4.24) \quad \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left( \sum_{k=0}^{(\ell \wedge \theta_R) - 1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2 \right)^{\frac{p}{2}} \right] &\leq N^{\frac{p}{2} - 1} \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \sum_{k=0}^{(\ell \wedge \theta_R) - 1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^p \right] \\ &\leq N^{\frac{p}{2} - 1} \sum_{k=0}^{M-1} \mathbb{E} [|\tilde{\sigma}(Z_{kh}^h)|^p |\Delta W_{k+1}^h|^p]. \end{aligned}$$

For each  $k \in \{0, \dots, N-1\}$ , we have  $\mathbb{E} [|\tilde{\sigma}(Z_{kh}^h)|^p |\Delta W_{k+1}^h|^p] = \mathbb{E} [|\tilde{\sigma}(Z_{kh}^h)|^p] \mathbb{E} [|\Delta W_{k+1}^h|^p]$ , by independence of increments. The monotonicity condition (4.10) implies that  $\mathbb{E} [|\tilde{\sigma}(Z_{kh}^h)|^p] \leq c_6(p) (\alpha^{\frac{p}{2}} + \beta^{\frac{p}{4}} \mathbb{E} [|Z_{kh}^h|^p])$ , for some constant  $c_6(p) > 0$ . For the truncated Brownian motion  $W^h$ , there exists a constant  $c_7(p) > 0$  such that  $\mathbb{E} [|\Delta W_{k+1}^h|^p] \leq \mathbb{E} [|\Delta W_{k+1}|^p] \leq c_7(p) h^{\frac{p}{2}}$ , where the first inequality can be shown by conditioning the Brownian increment and applying Jensen's inequality.

Hence, by (4.24), we get

$$(4.25) \quad \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left( \sum_{k=0}^{(\ell \wedge \theta_R) - 1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2 \right)^{\frac{p}{2}} \right] \leq c_8(p, T) + c_9(p, T) h \sum_{k=0}^{M-1} \mathbb{E} [|Z_{(k \wedge \theta_R)h}^h|^p],$$

where  $c_8(p, T) := c_6(p) c_7(p) \alpha^{\frac{p}{2}} T^{\frac{p}{2}}$  and  $c_9(p, T) := c_6(p) c_7(p) \beta^{\frac{p}{4}} T^{\frac{p}{2} - 1}$ .

Combining (4.18) with (4.23) and (4.25) gives

$$\begin{aligned} \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left| \sum_{k=0}^{(\ell \wedge \theta_R)-1} Z_{kh}^h \tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h \right|^{\frac{p}{2}} \right] &\leq \tilde{c}_4(p, T) + \tilde{c}_5(p, T) h \sum_{k=0}^{M-1} \mathbb{E}[|Y_{(k \wedge \theta_R)h}^h|^p], \\ \mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} \left( \sum_{k=0}^{(\ell \wedge \theta_R)-1} |\tilde{\sigma}(Z_{kh}^h) \Delta W_{k+1}^h|^2 \right)^{\frac{p}{2}} \right] &\leq \tilde{c}_8(p, T) + \tilde{c}_9(p, T) h \sum_{k=0}^{M-1} \mathbb{E}[|Y_{(k \wedge \theta_R)h}^h|^p], \end{aligned}$$

for constants  $\tilde{c}_4(p, T) = c_4(p, T) + 2^{\frac{3p}{2}-1} \alpha^{\frac{p}{2}} T c_5(p, T)$ ,  $\tilde{c}_5(p, T) = 2^{p-1} c_5(p, T)$ ,  $\tilde{c}_8(p, T) = c_8(p, T) + 2^{\frac{3p}{2}-1} \alpha^{\frac{p}{2}} T c_9(p, T)$ ,  $\tilde{c}_9(p, T) = 2^{p-1} c_9(p, T)$ . Further combining this with (4.21) and (4.22), we find that

$$\mathbb{E} \left[ \sup_{\ell \in \{0, \dots, M\}} |Y_{(\ell \wedge \theta_R)h}^h|^p \right] \leq \tilde{c}_0(p, T) + \tilde{c}_1(p, T) h \sum_{k=0}^{M-1} \mathbb{E} \left[ \sup_{j \in \{0, \dots, k\}} |Y_{(j \wedge \theta_R)h}^h|^p \right],$$

where  $\tilde{c}_0(p, T) = c_0(p, T) + 2^{\frac{3p}{2}-2} \tilde{c}_4(p, T) + 2^{p-2} \tilde{c}_8(p, T)$  and  $\tilde{c}_1(p, T) = c_1(p, T) + 2^{\frac{3p}{2}-2} \tilde{c}_5(p, T) + 2^{p-2} \tilde{c}_9(p, T)$ .

In light of (4.20), we can apply the discrete Gronwall inequality (see, e.g. [51, Lemma 3.4]) to conclude that

$$\mathbb{E} \left[ \sup_{\ell \in \{0, \dots, N\}} |Y_{(\ell \wedge \theta_R)h}^h|^p \right] \leq \tilde{c}_0(p, T) e^{\tilde{c}_1(p, T)T}.$$

We can now take the limit  $R \rightarrow \infty$  and apply Fatou's lemma to find

$$\mathbb{E} \left[ \sup_{\ell \in \{0, \dots, N\}} |Y_{\ell h}^h|^p \right] \leq \tilde{c}_0(p, T) e^{\tilde{c}_1(p, T)T}.$$

A further application of (4.18) yields

$$\mathbb{E} \left[ \sup_{\ell \in \{0, \dots, N\}} |Z_{\ell h}^h|^p \right] \leq 2^{\frac{p}{2}-1} K_h^{\frac{p}{2}} \left( \tilde{c}_0(p, T) e^{\tilde{c}_1(p, T)T} + (2\alpha h)^{\frac{p}{2}} \right).$$

Set  $C_0(p, T) := 2^{p-1}(\tilde{c}_0(p, T) e^{\tilde{c}_1(p, T)T} + (\frac{\alpha}{\beta})^{\frac{p}{2}})$  and  $C(p, T) := 2^{2(p-1)} L_{G^{-1}}^p(C_0(p, T) + |G(x_0)|^p) + 2^{p-1}|x_0|^p$ . Applying the bound (4.17), we conclude that (4.16) holds.

Given the bound (4.16) for some  $p = \bar{p} > 4$ , we deduce that (4.16) also holds for any  $p \in [1, 4]$  with  $C(p, T) = C(\bar{p}, T)^{\frac{p}{\bar{p}}}$ , by applying Jensen's inequality.

The result for  $\tilde{X}^h$  is proved analogously.  $\square$

We prove the following strong convergence rates.

**Theorem 4.18.** *Suppose that the coefficients of (1.1) satisfy Assumption 4.1, let  $X$  be the unique strong solution of (1.1), let  $(\tilde{b}, \tilde{\sigma})$  be the transformed coefficients in (4.9), and let  $\beta, L_{\tilde{b}} \in (0, \infty)$  be as in Lemma 4.10. Let  $N \in \mathbb{N}$  with  $N > (1 \vee 4\beta \vee 2L_{\tilde{b}})T$ , let  $h = T/N$ , and let  $X^h$  denote the transformed monotone semi-implicit Euler–Maruyama scheme defined by (4.15).*

*Then, for any  $p \geq 1$ , there exists  $C \in (0, \infty)$  such that*

$$\sup_{k \in \{1, \dots, N\}} \mathbb{E}[|X_{kh} - X_{kh}^h|^p]^{\frac{1}{p}} \vee \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^p dt \right]^{\frac{1}{p}} \leq \begin{cases} Ch^{\frac{1}{2}}, & p \in [1, 2], \\ Ch^{\frac{1}{p(p-1)}}, & p \in [2, \infty). \end{cases}$$

*Analogous estimates hold for the scheme  $\tilde{X}^h$  defined with respect to non-truncated Brownian increments.*

**Remark 4.19.**

- (i) To our knowledge, this is the first result on strong convergence of an implicit scheme in the presence of *both* discontinuities *and* exponential growth in the drift coefficient.

- (ii) For  $p \in [1, 2]$ , we obtain the strong  $L^p$  convergence rate of  $1/2$  that we expect. For  $p > 2$ , we leave the question of improving the strong  $L^p$  convergence rate for future research.
- (iii) In the case that  $b$  is continuous, our result still strengthens Hu [33, Theorem 2.4], since we relax the continuous differentiability assumption on the coefficients and consider general  $p \in [1, \infty)$ , not only  $p = 2$ .

*Proof of Theorem 4.18.* By assumption,  $h$  is chosen sufficiently small that  $X^h$  and  $\tilde{X}^h$  are well defined by Lemma 4.16, and that the moment bounds shown in Proposition 4.17 hold.

We first consider  $p = 2$ . As in the proof of Theorem 4.12,  $X_t = G^{-1}(Z_t)$ , for all  $t \in [0, T]$ , where  $Z$  is the unique strong solution of (1.1) with coefficients  $(\tilde{b}, \tilde{\sigma})$ , as defined in (4.9), and  $Z_0 = G(x_0)$ . Now, for  $k \in \{0, \dots, N\}$ , define  $Z_{kh}^h := G(X_{kh}^h)$ , so that  $Z_0^h = Z_0$  and, for  $k \in \{0, \dots, N-1\}$ ,

$$(4.26) \quad Z_{(k+1)h}^h = Z_{kh}^h + h\tilde{b}(Z_{(k+1)h}^h) + \tilde{\sigma}(Z_{kh}^h)\Delta W_{k+1}^h.$$

Then, using the Lipschitz property of  $G^{-1}$  proved in Lemma 4.7, we have

$$(4.27) \quad \begin{aligned} \mathbb{E}[|X_{kh} - X_{kh}^h|^2] &\leq L_{G^{-1}}^2 \mathbb{E}[|Z_{kh} - Z_{kh}^h|^2], \\ \mathbb{E}\left[\sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^2 dt\right] &\leq L_{G^{-1}}^2 \mathbb{E}\left[\sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |Z_t - Z_{kh}^h|^2 dt\right]. \end{aligned}$$

Therefore it suffices to analyse the  $L^2$  error between  $Z^h$  and  $Z$ .

For any  $k \in \{0, \dots, N-1\}$ , we can write

$$Z_{(k+1)h} = Z_{kh} + \tilde{b}(Z_{(k+1)h})h + \tilde{\sigma}(Z_{kh}^h)\Delta W_{k+1}^h + B_{k+1} + \Sigma_{k+1} + \Sigma_{k+1}^h + \mathcal{E}_{k+1},$$

where we define the error terms

$$\begin{aligned} B_{k+1} &:= \int_{kh}^{(k+1)h} (\tilde{b}(Z_s) - \tilde{b}(Z_{(k+1)h})) ds, & \Sigma_{k+1} &:= \int_{kh}^{(k+1)h} (\tilde{\sigma}(Z_s) - \tilde{\sigma}(Z_{kh})) dW_s, \\ \Sigma_{k+1}^h &:= (\tilde{\sigma}(Z_{kh}) - \tilde{\sigma}(Z_{kh}^h))\Delta W_{k+1}^h, & \mathcal{E}_{k+1} &:= \tilde{\sigma}(Z_{kh})[\Delta W_{k+1} - \Delta W_{k+1}^h]. \end{aligned}$$

Note that, if we define  $\tilde{Z}_{kh}^h := G(\tilde{X}_{kh}^h)$ , for  $k \in \{0, \dots, N\}$ , then  $\tilde{Z}^h$  satisfies analogous estimates to (4.27) and also satisfies (4.26) with the truncated increment  $\Delta W_{k+1}^h$  replaced by  $\Delta W_{k+1}$ . Thus we can write

$$Z_{(k+1)h} = Z_{kh} + \tilde{b}(Z_{(k+1)h})h + \tilde{\sigma}(\tilde{Z}_{kh}^h)\Delta W_{k+1} + B_{k+1} + \Sigma_{k+1} + \tilde{\Sigma}_{k+1},$$

with  $\tilde{\Sigma}_{k+1} := (\tilde{\sigma}(Z_{kh}) - \tilde{\sigma}(\tilde{Z}_{kh}^h))\Delta W_{k+1}$ . We can bound the second moment of  $\tilde{\Sigma}_{k+1}$  in exactly the same way as we do for  $\Sigma_{k+1}^h$  below. Thus the estimates that we prove for  $Z^h$  also hold for  $\tilde{Z}^h$  with only a minor modification in the proof.

Now we consider

$$Z_{(k+1)h} - Z_{(k+1)h}^h = Z_{kh} - Z_{kh}^h + (\tilde{b}(Z_{(k+1)h}) - \tilde{b}(Z_{(k+1)h}^h))h + B_{k+1} + \Sigma_{k+1} + \Sigma_{k+1}^h + \mathcal{E}_{k+1}.$$

By Lemma 4.10, there exists  $\tilde{b}' \in L_{\text{loc}}^1(\mathbb{R})$  such that, for all  $x, y \in \mathbb{R}$ ,

$$\tilde{b}(y) - \tilde{b}(x) = \int_x^y \tilde{b}'(s) ds = \left( \int_0^1 \tilde{b}'(uy + (1-u)x) du \right) (y - x).$$

Hence

$$(4.28) \quad \begin{aligned} &\left(1 - h \int_0^1 \tilde{b}'(uZ_{(k+1)h} - (1-u)Z_{(k+1)h}^h) du\right) (Z_{(k+1)h} - Z_{(k+1)h}^h) \\ &= (Z_{kh} - Z_{kh}^h) + B_{k+1} + \Sigma_{k+1} + \Sigma_{k+1}^h + \mathcal{E}_{k+1}. \end{aligned}$$



By the one-sided Lipschitz property shown in Lemma 4.10, we have that  $\tilde{b}' \leq L_{\tilde{b}}$  Lebesgue-almost everywhere. Then, similarly to [33, Proof of Theorem 2.4], since  $h < 1/L_{\tilde{b}}$ , the inverse

$$\mathcal{I}_{k+1} := \left( 1 - h \int_0^1 \tilde{b}'(uZ_{(k+1)h} - (1-u)Z_{(k+1)h}^h) du \right)^{-1}$$

exists and  $|\mathcal{I}_{k+1}| \leq (1 - L_{\tilde{b}}h)^{-1}$ . Hence, by (4.28),

$$Z_{(k+1)h} - Z_{(k+1)h}^h = \mathcal{I}_{k+1}(Z_{kh} - Z_{kh}^h) + \mathcal{I}_{k+1}B_{k+1} + \mathcal{I}_{k+1}\Sigma_{k+1} + \mathcal{I}_{k+1}\Sigma_{k+1}^h + \mathcal{I}_{k+1}\mathcal{E}_{k+1},$$

and iterating gives

$$(4.29) \quad \begin{aligned} Z_{(k+1)h} - Z_{(k+1)h}^h &= \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j B_j + \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j \Sigma_j \\ &\quad + \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j \Sigma_j^h + \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j \mathcal{E}_j. \end{aligned}$$

We now bound the second moment of the first term. By the Cauchy–Schwarz inequality,

$$\left| \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j B_j \right|^2 \leq \sum_{j=1}^{k+1} |\mathcal{I}_{k+1} \dots \mathcal{I}_j|^2 \cdot \sum_{j=1}^{k+1} |B_j|^2 \leq (k+1)(1 - L_{\tilde{b}}h)^{-2(k+1)} \sum_{j=1}^{k+1} |B_j|^2.$$

Since  $L_{\tilde{b}}h \leq \frac{1}{2}$ , setting  $c_1 := 4L_{\tilde{b}}$  gives

$$(1 - L_{\tilde{b}}h)^{-2(k+1)} \leq \exp \left\{ 2(k+1) \frac{L_{\tilde{b}}h}{1 - L_{\tilde{b}}h} \right\} \leq \exp \{c_1 T\}.$$

For each  $j \in \{1, \dots, N\}$ , applying the Cauchy–Schwarz inequality and the estimate (4.14) from Proposition 4.11 gives us

$$\mathbb{E}[|B_j|^2] \leq h \int_{(j-1)h}^{jh} \mathbb{E} \left[ |\tilde{b}(Z_s) - \tilde{b}(Z_{jh})|^2 \right] ds \leq C_2(2, T)h^3.$$

Therefore, setting  $c_2 = C_2(2, T)$ , we have

$$(4.30) \quad \mathbb{E} \left[ \left| \sum_{j=1}^{k+1} \mathcal{I}_{k+1} \dots \mathcal{I}_j B_j \right|^2 \right] \leq c_2 T^2 \exp(c_1 T) h.$$

To estimate the remaining terms in (4.29), we set  $\mathcal{S}_0 = 0$  and, for all  $k \in \{1, \dots, N\}$ , denote

$$\mathcal{S}_k := \sum_{j=1}^k \mathcal{I}_k \dots \mathcal{I}_j \Sigma_j + \sum_{j=1}^k \mathcal{I}_k \dots \mathcal{I}_j \Sigma_j^h + \sum_{j=1}^k \mathcal{I}_k \dots \mathcal{I}_j \mathcal{E}_j.$$

Observe that, for all  $k \in \{0, \dots, N-1\}$ ,

$$\mathcal{S}_{k+1} = \mathcal{I}_{k+1}(\mathcal{S}_k + \Sigma_{k+1} + \Sigma_{k+1}^h + \mathcal{E}_{k+1}).$$

By independence of Brownian increments, we have that  $\mathcal{S}_k$  is independent of  $\Sigma_{k+1}$ ,  $\Sigma_{k+1}^h$ , and  $\mathcal{E}_{k+1}$ . Using Pythagoras' theorem and then the Cauchy–Schwarz inequality, we see that the second moment is bounded by

$$(4.31) \quad \mathbb{E}[|\mathcal{S}_{k+1}|^2] \leq (1 - L_{\tilde{b}}h)^{-2} (\mathbb{E}[|\mathcal{S}_k|^2] + 3\mathbb{E}[|\Sigma_{k+1}|^2] + 3\mathbb{E}[|\Sigma_{k+1}^h|^2] + 3\mathbb{E}[|\mathcal{E}_{k+1}|^2]).$$

We note that, in order to iterate this bound, it is essential that no constant appears in front of the term  $\mathbb{E}[|\mathcal{S}_k|^2]$ . Thus the above argument using independence does not directly generalise to  $p^{\text{th}}$  moments for  $p \neq 2$ .

We now bound the right-hand side of (4.31) term by term. By Lemma 4.10,  $\tilde{\sigma}$  is Lipschitz with constant  $L_{\tilde{\sigma}}$ . Therefore, by the Itô isometry,

$$\mathbb{E}[|\Sigma_{k+1}|^2] = \mathbb{E}\left[\int_{kh}^{(k+1)h} |\tilde{\sigma}(Z_s) - \tilde{\sigma}(Z_{kh})|^2 ds\right] \leq L_{\tilde{\sigma}}^2 \mathbb{E}\left[\int_{kh}^{(k+1)h} |Z_s - Z_{kh}|^2 ds\right].$$

Then the bound (4.13) gives a constant  $c_3 \in (0, \infty)$  such that

$$(4.32) \quad \mathbb{E}[|\Sigma_{k+1}|^2] \leq c_3 L_{\tilde{\sigma}}^2 \int_{kh}^{(k+1)h} |s - kh| ds \leq c_3 L_{\tilde{\sigma}}^2 h^2.$$

By the independence of  $Z_{kh}$  from the increments  $\Delta W_{k+1}$  and  $\Delta W_{k+1}^h$ , together with the linear growth of  $\tilde{\sigma}$  and the second moment bound from Proposition 4.17, there exists  $c_4 \in (0, \infty)$  such that

$$\mathbb{E}[|\mathcal{E}_{k+1}|^2] = \mathbb{E}[|\tilde{\sigma}(Z_{kh})|^2] \cdot \mathbb{E}[|\Delta W_{k+1} - \Delta W_{k+1}^h|^2] \leq c_4 \mathbb{E}[|\Delta W_{k+1} - \Delta W_{k+1}^h|^2].$$

By [8, Lemma A.3], we have the bound  $\mathbb{E}[|\Delta W_{k+1} - \Delta W_{k+1}^h|^2] \leq 2h^3$ , and so there exists  $c_5 \in (0, \infty)$  such that

$$(4.33) \quad \mathbb{E}[|\mathcal{E}_{k+1}|^2] \leq c_5 h^3.$$

An application of the triangle inequality then yields the existence of a constant  $c_6 \in (0, \infty)$  such that  $\mathbb{E}[|\Delta W_{k+1}^h|^2] \leq 2(\mathbb{E}[|\mathcal{E}_{k+1}|^2] + \mathbb{E}[|\Delta W_{k+1}|^2]) \leq c_6 h$ . This, the independence of  $Z_{kh}$  and  $Z_{kh}^h$  from the increment  $\Delta W_{k+1}^h$ , and the Lipschitz continuity of  $\tilde{\sigma}$  yield the bound

$$(4.34) \quad \begin{aligned} \mathbb{E}[|\Sigma_{k+1}^h|^2] &= \mathbb{E}[|\tilde{\sigma}(Z_{kh}) - \tilde{\sigma}(Z_{kh}^h)|^2] \cdot \mathbb{E}[|\Delta W_{k+1}^h|^2] \\ &\leq c_6 L_{\tilde{\sigma}}^2 \mathbb{E}[|Z_{kh} - Z_{kh}^h|^2] h. \end{aligned}$$

Combining (4.29) with the definition of  $\mathcal{S}_k$  and the bound (4.30) gives

$$(4.35) \quad \mathbb{E}[|Z_{kh} - Z_{kh}^h|^2] \leq 2c_2 T^2 \exp(c_1 T) h + 2\mathbb{E}[|\mathcal{S}_k|^2].$$

We now substitute the bounds (4.32), (4.33), and (4.34) into (4.31), to get

$$\mathbb{E}[|\mathcal{S}_{k+1}|^2] \leq (1 - L_{\tilde{\sigma}} h)^{-2} (\mathbb{E}[|\mathcal{S}_k|^2] + 3c_3 L_{\tilde{\sigma}}^2 h^2 + 3c_6 L_{\tilde{\sigma}}^2 \mathbb{E}[|Z_{kh} - Z_{kh}^h|^2] h + 3c_5 h^3).$$

Then (4.35) gives

$$\mathbb{E}[|\mathcal{S}_{k+1}|^2] \leq (1 - L_{\tilde{\sigma}} h)^{-2} (1 + 6c_6 L_{\tilde{\sigma}}^2 h) \mathbb{E}[|\mathcal{S}_k|^2] + c_7 h^2 (1 - L_{\tilde{\sigma}} h)^{-2},$$

where  $c_7 := 3(c_3 L_{\tilde{\sigma}}^2 + 2c_2 c_6 L_{\tilde{\sigma}}^2 T^2 \exp(c_1 T) + c_5)$ . Since  $L_{\tilde{\sigma}} h \leq \frac{1}{2}$ , we can check that  $1 \leq (1 - L_{\tilde{\sigma}} h)^{-2} (1 + 6c_6 L_{\tilde{\sigma}}^2 h) \leq 1 + c_8 h$ , where  $c_8 := 24c_6 L_{\tilde{\sigma}}^2$ , and we have

$$\mathbb{E}[|\mathcal{S}_{k+1}|^2] \leq (1 + c_8 h) \mathbb{E}[|\mathcal{S}_k|^2] + 4c_7 h^2.$$

Iterating, we obtain

$$\mathbb{E}[|\mathcal{S}_{k+1}|^2] \leq 4c_7 h^2 \sum_{j=0}^k (1 + c_8 h)^j \leq 4c_7 h^2 \sum_{j=0}^k \exp(c_8 h N) \leq 4c_7 T \exp(c_8 T) h.$$

Combining this with (4.35) gives a constant  $c_9 \in (0, \infty)$  such that

$$(4.36) \quad \mathbb{E}[|Z_{kh} - Z_{kh}^h|^2] \leq c_9 h.$$

Now by (4.13), for  $k \in \{0, \dots, N-1\}$  and  $t \in [kh, (k+1)h)$ , there exists a constant  $c_{10} \in (0, \infty)$  such that  $\mathbb{E}[|Z_t - Z_{kh}|^2] \leq c_{10}h$ . Combining this with (4.36), we have

$$(4.37) \quad \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |Z_t - Z_{kh}^h|^2 dt \right] \leq 2 \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} (\mathbb{E}[|Z_t - Z_{kh}|^2] + \mathbb{E}[|Z_{kh}^h - Z_{kh}|^2]) dt \\ \leq 2Nh(c_9h + c_{10}h) = \tilde{C}h,$$

for  $\tilde{C} := 2T(c_9 + c_{10})$ .

The estimates (4.27) applied to (4.36) and (4.37) yield

$$\sup_{k \in \{1, \dots, N\}} \mathbb{E}[|X_{kh} - X_{kh}^h|^2]^{1/2} \leq L_{G^{-1}} c_9^{1/2} h^{1/2},$$

and

$$(4.38) \quad \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^2 dt \right]^{1/2} \leq L_{G^{-1}} \tilde{C}^{1/2} h^{1/2}.$$

For  $p \in [1, 2)$ , we obtain the same rate of convergence by applying Jensen's inequality. For  $p \in (2, \infty)$ , we apply Hölder's inequality, followed by the  $L^2$  bound (4.38) and the  $(p+1)$ -moment bounds for the scheme and the solution of the SDE that are shown in Proposition 4.17 and Proposition 4.11, respectively. Hence, there exists a constant  $C \in (0, \infty)$  such that

$$\mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^p dt \right]^{\frac{1}{p}} = \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^{\frac{2}{p-1}} \cdot |X_t - X_{kh}^h|^{\frac{(p-2)(p+1)}{p-1}} dt \right]^{\frac{1}{p}} \\ \leq \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^2 dt \right]^{\frac{1}{p(p-1)}} \mathbb{E} \left[ \sum_{k=0}^{N-1} \int_{kh}^{(k+1)h} |X_t - X_{kh}^h|^{p+1} dt \right]^{\frac{p-2}{p(p-1)}} \\ \leq Ch^{\frac{1}{p(p-1)}}.$$

In the same way, we show that

$$\sup_{k \in \{1, \dots, N\}} \mathbb{E}[|X_{kh} - X_{kh}^h|^p]^{\frac{1}{p}} \leq Ch^{\frac{1}{p(p-1)}},$$

which concludes the proof.  $\square$

**Remark 4.20** (additive noise). Under the additional assumption that  $\sigma \equiv 1$ , we only require estimates for the error terms  $B_k$ , but not for  $\Sigma_k, \tilde{\Sigma}_k^h$ , in the proof of convergence of the scheme  $\tilde{X}^h$ . Repeated applications of Hölder's inequality yield  $p^{\text{th}}$  moment estimates analogous to (4.30) with order  $h^{\frac{p}{2}}$ . Proceeding as in the proof of Theorem 4.18, we find the strong  $L^p$  convergence rate  $1/2$ , for all  $p \geq 1$ , for the scheme  $\tilde{X}^h$  in the case of additive noise.

**4.3. Application to optimal transport.** We next apply the transformed monotone semi-implicit Euler–Maruyama scheme from Definition 4.13 to finding the optimiser of (BCOT) between the laws of SDEs satisfying Assumption 4.1. To this end, we will verify the following properties of the scheme.

**Lemma 4.21.** *Suppose that the coefficients of (1.1) satisfy Assumption 4.1. Then, for all  $h$  sufficiently small, the transformed monotone semi-implicit Euler–Maruyama scheme (4.15) is stochastically increasing and, for  $k \in \{0, \dots, N-1\}$ , the map  $\Delta W_{k+1}^h \mapsto X_{(k+1)h}^h$  is increasing.*

*Proof.* By Lemma 4.10,  $\tilde{\sigma}$  is Lipschitz with constant  $L_{\tilde{\sigma}} \in (0, \infty)$ . Fix  $k \in \{0, \dots, N-1\}$  and  $x, y \in \mathbb{R}$  with  $x < y$ . Then, using Definition 2.15 of  $\Delta W^h$ ,

$$\begin{aligned} x + \tilde{\sigma}(x)\Delta W_{k+1}^h - (y + \tilde{\sigma}(y)\Delta W_{k+1}^h) &= x - y + (\tilde{\sigma}(x) - \tilde{\sigma}(y))\Delta W_{k+1}^h \\ &\leq (1 + L_{\tilde{\sigma}}\Delta W_{k+1}^h)(x - y) \\ &\leq (1 - L_{\tilde{\sigma}}A_h)(x - y). \end{aligned}$$

For  $h > 0$  sufficiently small that  $1 - L_{\tilde{\sigma}}A_h > 0$ , we have that the map  $(\text{id} + \Delta W_{k+1}^h\tilde{\sigma})$  is strictly increasing with strictly increasing inverse. Let us also take  $h$  sufficiently small that  $(\text{id} - h\tilde{b})^{-1}$  is well defined and strictly increasing by Lemma 4.15. Then we can rewrite (4.15) as

$$(4.39) \quad X_{(k+1)h}^h = G^{-1} \circ (\text{id} - h\tilde{b})^{-1} \circ (\text{id} + \Delta W_{k+1}^h\tilde{\sigma}) \circ G(X_{kh}^h).$$

Recall that  $G$  and  $G^{-1}$  are strictly increasing by Lemma 4.7. Thus  $X_{(k+1)h}^h$  is obtained from  $X_{kh}^h$  by the concatenation of four strictly increasing maps. This shows that  $(X_{kh}^h)_{k \in \{0, \dots, N\}}$  is stochastically increasing. Noting that  $\tilde{\sigma}$  is non-negative, the second claim also follows from (4.39).  $\square$

We can now state and prove our optimality result.

**Theorem 4.22.** *Suppose that the coefficients  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  of (2.6) satisfy Assumption 4.1 and write  $\mu, \nu$  for the laws of the respective strong solutions  $X, \bar{X}$ . Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Then the synchronous coupling attains the value  $V_c(\mu, \nu)$  of the bicausal optimal transport problem (BCOT).*

*Proof.* We proved strong existence and pathwise uniqueness in Theorem 4.12. For  $N \in \mathbb{N}$ ,  $h = T/N$ , denote by  $(X_{kh}^h)_{k \in \{1, \dots, N\}}$  and  $(\bar{X}_{kh}^h)_{k \in \{1, \dots, N\}}$  the transformed monotone semi-implicit Euler–Maruyama scheme for the solutions  $X$  and  $\bar{X}$  of (1.1), respectively, as defined in Definition 4.13. By Lemma 4.21,  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  satisfy Assumptions 3.1 (iii) and (iv). By Theorem 4.18, Assumption 3.1 (v) is also satisfied. We conclude by applying Proposition 3.3.  $\square$

## 5. SDES WITH BOUNDED MEASURABLE DRIFT

We now consider a second class of irregular coefficients. We only require the drift coefficient to be bounded and measurable, and we allow the diffusion coefficient to be Hölder continuous in space rather than Lipschitz, but we do require a uniform non-degeneracy condition and boundedness of the diffusion, which was not needed in Section 4. The coefficients may be time-dependent. In this setting, we again obtain optimality of the synchronous coupling for (BCOT) between the laws of SDEs. Similarly to Section 4, the proof relies on a transformation of the SDEs. Here we use the drift-removing transformation that is defined by Zvonkin [88, Theorem 1]. We solve a transformed bicausal optimal transport problem, which we show to be equivalent to (BCOT). This is a different approach from Section 4; in Section 5.1 we show that the key monotonicity result required for the former approach is not satisfied in the present setting, and we provide a counterexample to show that a weaker monotonicity condition is not sufficient.

We make the following assumptions on the coefficients  $(b, \sigma)$  of the SDE (1.1), following Zvonkin [88].

**Assumption 5.1.** Suppose that the coefficients  $(b, \sigma)$  satisfy the following:

- (i)  $b$  is bounded and measurable;
- (ii)  $\sigma$  is bounded and continuous and, for some  $\alpha \in [1/2, 1]$ ,  $x \mapsto \sigma_t(x)$  is  $\alpha$ -Hölder continuous, uniformly in  $t \in [0, T]$ ;
- (iii)  $\sigma$  is uniformly non-degenerate; i.e. there exists  $C > 0$  such that  $\sigma_t^2(x) \geq C$ , for all  $(t, x) \in [0, T] \times \mathbb{R}$ .

**Remark 5.2.**

- (1) Under Assumption 5.1, Zvonkin [88, Theorem 4] proves existence and uniqueness of strong solutions of the SDE (1.1) via a transformation that removes the drift.
- (2) A similar transformation to that of [88] was used by Talay [82] to analyse the numerical approximation of SDEs. More recently, Backhoff-Veraguas et al. [8], Gerencsér, Lampl, and Ling [24], Neuenkirch and Szölgyenyi [60], Ngo and Taguchi [62, 63, 64] use a similar drift-removing transformation in the numerical approximation of SDEs with irregular coefficients, under stronger conditions than Assumption 5.1.
- (3) For the SDE (1.1) with coefficients satisfying Assumption 5.1, strong convergence rates for the Euler–Maruyama scheme have been obtained by Gyöngy and Rásonyi [29] under additional conditions on the drift, and by Dareiotis and Gerencsér [19], Dareiotis et al. [20] for time-homogeneous coefficients under additional conditions on the diffusion.

**Example 5.3.** Consider the SDE

$$dX_t = \text{sign}(X_t) dt + (1 + \sqrt{|X_t|} \mathbf{1}_{\{|X_t| \leq 4\}} + 2 \mathbf{1}_{\{|X_t| > 4\}}) dW_t; \quad X_0 = x_0 \in \mathbb{R}.$$

The coefficients of this SDE satisfy Assumption 5.1.

Suppose that  $(b, \sigma)$  satisfy Assumption 5.1, and let  $X$  be the unique strong solution of (1.1) with coefficients  $(b, \sigma)$ . Zvonkin [88, Theorem 1] gives the existence of a function  $u: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  that is in the Sobolev space  $W_p^{1,2}([0, T] \times D)$ , for any bounded domain  $D \subset \mathbb{R}$  and  $p > \frac{3}{2}$ , and solves the PDE problem

$$\begin{aligned} \partial_t u_t(x) + \partial_x u_t(x) b_t(x) + \frac{1}{2} \partial_{xx} u_t(x) \sigma_t^2(x) &= 0, \quad (t, x) \in [0, T] \times \mathbb{R}, \\ u_T(x) &= x, \quad x \in \mathbb{R}. \end{aligned}$$

By [88, Theorem 2], there exists an inverse function  $v: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $(v_t \circ u_t)(x) = x$  for all  $t \in [0, T]$ ,  $x \in \mathbb{R}$ .

Define  $Z_t := u_t(X_t)$  for all  $t \in [0, T]$ . Then [88, Theorem 3] shows that we can apply Itô's formula to obtain

$$(5.1) \quad dZ_t = \tilde{\sigma}_t(Z_t) dW_t, \quad Z_0 = u_0(x_0),$$

where  $\tilde{\sigma}_t := (\partial_x u_t \cdot \sigma_t) \circ v_t$ , for all  $t \in [0, T]$ .

We recall the following properties of  $u$  and  $v$  from [88, Theorem 2] without proof.

**Lemma 5.4.** *Suppose that  $(b, \sigma)$  satisfy Assumption 5.1. Then, uniformly in  $t \in [0, T]$ ,*

- (i)  $x \mapsto u_t(x)$  and  $x \mapsto v_t(x)$  are strictly increasing and Lipschitz continuous;
- (ii)  $x \mapsto \partial_x u_t(x)$  is bounded from above, bounded away from zero, and  $\beta$ -Hölder continuous, for any  $\beta \in (0, 1)$ ;

and, uniformly in  $x \in D$ , for any bounded domain  $D \subset \mathbb{R}$ ,

- (iii)  $t \mapsto u_t(x)$ ,  $t \mapsto v_t(x)$ , and  $t \mapsto \partial_x u_t(x)$  are  $\beta$ -Hölder continuous for any  $\beta \in (0, 1)$ .

The following is an immediate consequence of Lemma 5.4.

**Corollary 5.5.** *Suppose that  $(b, \sigma)$  satisfies Assumption 5.1. Then the diffusion coefficient  $\tilde{\sigma}$  of the transformed SDE (5.1) is continuous, bounded from above, and bounded away from zero. Moreover,  $x \mapsto \tilde{\sigma}_t(x)$  is  $\alpha$ -Hölder continuous, uniformly in  $t \in [0, T]$ .*

Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Suppose that  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  satisfy Assumption 5.1, and write  $\mu, \nu$  for the laws of the solutions  $X, \bar{X}$  of (2.6). For  $(\bar{b}, \bar{\sigma})$ , define functions  $\bar{u}, \bar{v}$  analogously to  $u, v$  and define  $\bar{Z}_t = \bar{u}_t(\bar{X}_t)$  for all  $t \in [0, T]$ .

In order to find the value  $V_c(\mu, \nu)$  of (BCOT), we consider the following auxiliary problem. Write  $\tilde{\mu}, \tilde{\nu}$  for the laws of  $Z, \bar{Z}$ , respectively, and define a cost function  $\tilde{c}: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  by

$$(5.2) \quad \tilde{c}_t(z, \bar{z}) := c_t(v_t(z), \bar{v}_t(\bar{z})).$$

Consider the problem

$$(5.3) \quad V_{\tilde{c}}(\tilde{\mu}, \tilde{\nu}) := \inf_{\pi \in \text{Cpl}_{\text{bc}}(\tilde{\mu}, \tilde{\nu})} \mathbb{E}^\pi \left[ \int_0^T \tilde{c}_t(\omega_t, \bar{\omega}_t) dt \right].$$

We now verify that  $\tilde{c}$  inherits the continuity, polynomial growth and quasi-monotonicity properties from  $c$ .

**Lemma 5.6.** *Suppose that  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  satisfy Assumption 5.1 and that  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfies Assumption 2.11 for some  $p \geq 1$ . Then the function  $\tilde{c}_t: \mathbb{R} \times \mathbb{R}$  defined by (5.2) also satisfies Assumption 2.11 for the same value of  $p$ .*

*Proof.* By Lemma 5.4,  $v, \bar{v}: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  are continuous. Thus, since  $c$  is continuous on  $[0, T] \times M$  for any compact  $M \subseteq \mathbb{R} \times \mathbb{R}$ , the same property holds for  $\tilde{c}$ . By assumption, there exist  $p \geq 1$  and  $K \geq 0$  such that, for all  $t \in [0, T]$ ,  $c_t$  satisfies the growth bound (2.3). Now fix  $t \in [0, T]$ . Then

$$|\tilde{c}_t(z, \bar{z})| = |c_t(v_t(z), \bar{v}_t(\bar{z}))| \leq K(1 + |v_t(z)|^p + |\bar{v}_t(\bar{z})|^p).$$

Writing  $L$  for the Lipschitz constant of  $v_t$ , we have  $|v_t(z)| \leq L|z| + L|u_t(x_0)| + |x_0|$ , and similarly for  $\bar{v}_t(\bar{z})$ . Thus the growth bound (2.3) holds for  $\tilde{c}_t$  with a constant  $\tilde{K}$  independent of  $t$ , and with power  $p$ . It remains to prove that  $\tilde{c}_t$  satisfies (2.4). Since  $v_t, \bar{v}_t$  are increasing by Lemma 5.4, this follows immediately from the fact that  $c_t$  satisfies (2.4).  $\square$

Now we show that (5.3) is equivalent to (BCOT).

**Lemma 5.7.** *Suppose that  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  satisfy Assumption 5.1. Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be a measurable function. Then  $V_c(\mu, \nu) = V_{\tilde{c}}(\tilde{\mu}, \tilde{\nu})$ .*

*Proof.* Let  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$ . By [8, Proposition 2.2],  $\pi = \text{Law}(X, \bar{X})$ , where  $(X, \bar{X})$  is the unique strong solution of the system of SDEs (2.6) driven by some correlated Brownian motion  $(W, \bar{W})$ . Now define  $(Z, \bar{Z})$  by  $Z_t = u_t(X_t)$ ,  $\bar{Z}_t = \bar{u}_t(\bar{X}_t)$ , for all  $t \in [0, T]$ . Then  $Z$  is the unique strong solution of the SDE (5.1) driven by  $W$ , and  $\bar{Z}$  of the analogous transformed SDE for  $(\bar{b}, \bar{\sigma})$  driven by  $\bar{W}$ . Hence, by [8, Proposition 2.2] again, we can define a bicausal coupling by  $\tilde{\pi} := \text{Law}(Z, \bar{Z}) \in \text{Cpl}_{\text{bc}}(\tilde{\mu}, \tilde{\nu})$ . By the definition of  $\tilde{c}$ , we can write

$$(5.4) \quad \mathbb{E}^{\tilde{\pi}} \left[ \int_0^T \tilde{c}_t(\omega_t, \bar{\omega}_t) dt \right] = \mathbb{E} \left[ \int_0^T c_t(v_t(Z_t), \bar{v}_t(\bar{Z}_t)) dt \right] = \mathbb{E}^\pi \left[ \int_0^T c_t(\omega_t, \bar{\omega}_t) dt \right].$$

Starting from an arbitrary  $\tilde{\pi} \in \text{Cpl}_{\text{bc}}(\tilde{\mu}, \tilde{\nu})$ , we can construct  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$  in a similar manner, via the inverse transform  $v$ . Thus, taking the infimum on both sides of (5.4), we obtain  $V_c(\mu, \nu) = V_{\tilde{c}}(\tilde{\mu}, \tilde{\nu})$ .  $\square$

Combining the preceding results, we obtain optimality of the synchronous coupling for (BCOT).

**Theorem 5.8.** *Suppose that  $(b, \sigma)$ ,  $(\bar{b}, \bar{\sigma})$  satisfy Assumption 5.1, and write  $\mu, \nu$  for the laws of  $X, \bar{X}$ , respectively. Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Then the synchronous coupling attains the value  $V_c(\mu, \nu)$  of the bicausal optimal transport problem (BCOT).*

*Proof.* Thanks to Lemma 5.7, it suffices to prove that the synchronous coupling  $\pi_{\tilde{\mu}, \tilde{\nu}}^{\text{sync}}$  attains the value  $V_{\tilde{c}}(\tilde{\mu}, \tilde{\nu})$ , where  $\tilde{\mu}, \tilde{\nu}$  are the laws of solutions of transformed SDEs of the form (5.1). By Corollary 5.5, the coefficients of these SDEs are bounded and continuous, and pathwise uniqueness

holds by Watanabe and Yamada [85, Theorem 2]. Thus the coefficients of each of the transformed SDEs satisfy Assumption 3.7. By Lemma 5.6,  $\tilde{c}$  satisfies the conditions of Corollary 3.8, and applying this result completes the proof.  $\square$

**5.1. Lack of monotonicity and counterexample.** In the proof of Theorem 5.8, we consider transformed SDEs with Hölder continuous diffusion coefficients that may not be Lipschitz. Recall the monotone Euler–Maruyama scheme from Section 2.3 that was introduced in [8, Definition 3.13]. We show that this scheme is not stochastically monotone for SDEs whose diffusion coefficient is not Lipschitz. Thus we could not apply Proposition 3.3 with the monotone Euler–Maruyama scheme to prove optimality in this case.

**Proposition 5.9.** *Let  $\sigma: \mathbb{R} \rightarrow (0, \infty)$  be  $\alpha$ -Hölder continuous for some  $\alpha \in [1/2, 1)$ , bounded from above, and bounded away from zero. Consider the SDE  $dZ_t = \sigma(Z_t) dW_t$ , with  $Z_0 = z_0 \in \mathbb{R}$ . If  $\sigma$  is not Lipschitz continuous, then the monotone Euler–Maruyama scheme for  $Z$  is not stochastically monotone for any step size.*

**Remark 5.10.** In the setting of Proposition 5.9, the monotone Euler–Maruyama scheme coincides with the transformed monotone semi-implicit Euler–Maruyama scheme, since there is no drift term.

*Proof of Proposition 5.9.* The conditions of the proposition are sufficient to guarantee the existence of a unique strong solution  $Z$  of the SDE by, e.g. [88, Theorem 4]. For any  $N \in \mathbb{N}$  and  $h = \frac{T}{N}$ , let  $Z^h$  denote the monotone Euler–Maruyama scheme for  $Z$  with step size  $h$ , and let  $A_h$  be the truncation level given in Definition 2.15. Note that, since  $\sigma$  is not Lipschitz continuous, there exist  $z, \bar{z} \in \mathbb{R}$  with  $z < \bar{z}$  such that

$$(5.5) \quad |\sigma(\bar{z}) - \sigma(z)| > A_h^{-1}(\bar{z} - z) > 0.$$

Now consider the random variables  $Y := z + \sigma(z)\xi^h$ ,  $\bar{Y} := \bar{z} + \sigma(\bar{z})\xi^h$ , where  $\xi^h$  is a random variable with the same law as an increment of the truncated Brownian motion  $W^h$ . For all  $a \in \mathbb{R}$ , we have

$$\mathbb{P}[Y \leq a] = \mathbb{P}\left[\xi^h \leq \frac{a - z}{\sigma(z)}\right] = \begin{cases} 0, & a \leq z - A_h\sigma(z), \\ 2\Phi\left(h^{-\frac{1}{2}} \frac{a - z}{\sigma(z)}\right), & a \in (z - A_h\sigma(z), z + A_h\sigma(z)), \\ 1, & a \geq z + A_h\sigma(z), \end{cases}$$

where  $\Phi: \mathbb{R} \rightarrow [0, 1]$  is the distribution function of a standard Gaussian, which is strictly increasing. The same holds for  $\bar{Y}$  with  $z$  replaced by  $\bar{z}$ . We aim to show that there exist  $a_*, a^* \in \mathbb{R}$  such that  $\mathbb{P}[\bar{Y} \leq a_*] < \mathbb{P}[Y \leq a_*]$  and  $\mathbb{P}[Y \leq a^*] < \mathbb{P}[\bar{Y} \leq a^*]$ . This would imply that  $Z^h$  is neither stochastically decreasing nor increasing. Suppose first that  $\sigma(\bar{z}) \geq \sigma(z)$ . Then rearranging (5.5) gives

$$\bar{z} - A_h\sigma(\bar{z}) < z - A_h\sigma(z).$$

So there exists  $a^* \in (\bar{z} - A_h\sigma(\bar{z}), z - A_h\sigma(z))$ , and we see that

$$\mathbb{P}[Y \leq a^*] = 0, \text{ while } \mathbb{P}[\bar{Y} \leq a^*] > 0.$$

Set  $a_* = z + A_h\sigma(z)$  and note that  $a_* < \bar{z} + A_h\sigma(\bar{z})$ . Therefore  $\mathbb{P}[Y \leq a_*] = 1$  and  $\mathbb{P}[\bar{Y} \leq a_*] < 1$ .

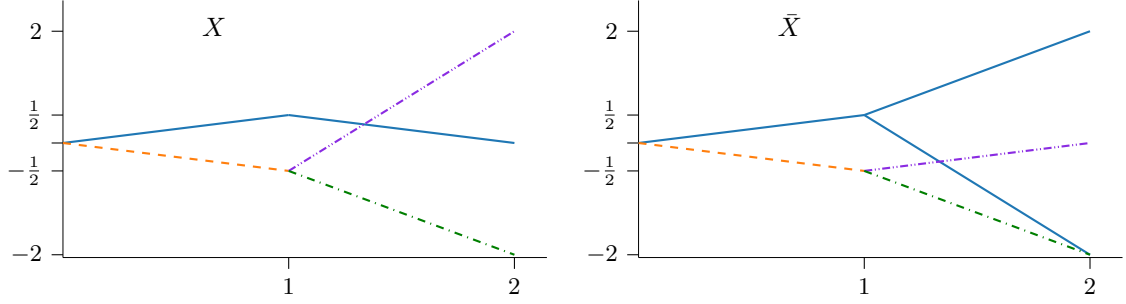
On the other hand, if  $\sigma(\bar{z}) < \sigma(z)$ , then  $\bar{z} + A_h\sigma(\bar{z}) < z + A_h\sigma(z)$ , and so there exists  $a^*$  such that  $\mathbb{P}[\bar{Y} \leq a^*] = 1$  and  $\mathbb{P}[Y \leq a^*] < 1$ . Now setting  $a_* = \bar{z} - A_h\sigma(\bar{z}) > z - A_h\sigma(z)$  gives  $\mathbb{P}[\bar{Y} \leq a_*] = 0$  and  $\mathbb{P}[Y \leq a_*] > 0$ . Hence  $Z^h$  is not stochastically monotone.  $\square$

A natural generalisation of stochastic monotonicity is monotonicity with respect to second order stochastic dominance. We now provide an example to show that, for two marginals that are increasing in second order stochastic dominance, the Knothe–Rosenblatt rearrangement may fail

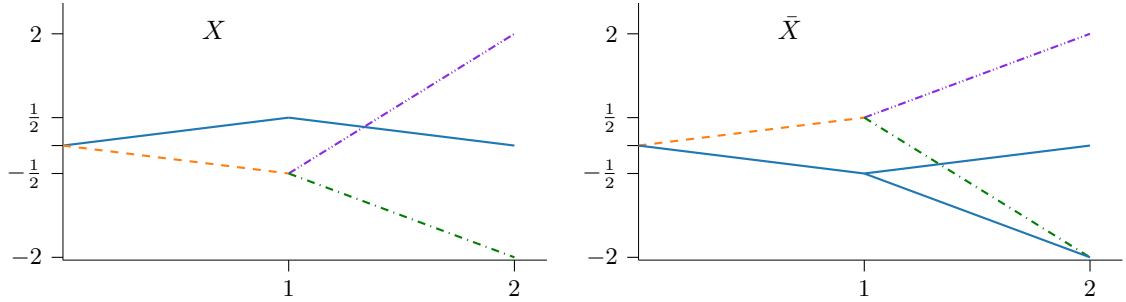


to be optimal for the bicausal transport problem, if one of the marginals is not stochastically increasing.

**Example 5.11** (second order stochastic dominance is insufficient for optimality). Consider the discrete-time Markov processes  $X = (X_1, X_2)$ ,  $\bar{X} = (\bar{X}_1, \bar{X}_2)$ , with first marginals given by  $\mathbb{P}[X_1 = -1/2] = \mathbb{P}[X_1 = 1/2] = \mathbb{P}[\bar{X}_1 = -1/2] = \mathbb{P}[\bar{X}_1 = 1/2] = 1/2$ , and conditional second marginals given by  $\mathbb{P}[X_2 = 2 \mid X_1 = -1/2] = \mathbb{P}[X_2 = -2 \mid X_1 = -1/2] = 1/2$ ,  $\mathbb{P}[X_2 = 0 \mid X_1 = 1/2] = 1$ , and  $\mathbb{P}[\bar{X}_2 = -2 \mid \bar{X}_1 = -1/2] = \mathbb{P}[\bar{X}_2 = 0 \mid \bar{X}_1 = -1/2] = \mathbb{P}[\bar{X}_2 = -2 \mid \bar{X}_1 = 1/2] = \mathbb{P}[\bar{X}_2 = 2 \mid \bar{X}_1 = 1/2] = 1/2$ .



(A) The laws of  $X$  and  $\bar{X}$  coupled with the Knothe–Rosenblatt rearrangement  $\pi_{\mu,\nu}^{\text{KR}}$ .



(B) The laws of  $X$  and  $\bar{X}$  coupled with an alternative bicausal coupling  $\pi_{\mu,\nu}^{\text{AT}}$ .

FIGURE 3. Two couplings of the laws of  $X$  and  $\bar{X}$ . In each case, paths with the same colour and line style are coupled with each other.

We claim that  $X$  is increasing with respect to second order stochastic dominance, but not with respect to first order stochastic dominance. Denote by  $F_-, F_+ : \mathbb{R} \rightarrow [0, 1]$  the distribution functions of the second marginal of  $X$ , conditional on  $X_1 = -\frac{1}{2}$  and  $X_1 = \frac{1}{2}$ , respectively; that is

$$F_-(x) = \begin{cases} 0, & x < -2, \\ \frac{1}{2}, & x \in [-2, 2), \\ 1, & x \geq 2, \end{cases} \quad F_+(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

We see that, for any  $x \in [0, 2)$ ,  $F_-(x) < F_+(x)$ , and so  $X$  is not stochastically increasing. However, integrating, we find that

$$\int_{-\infty}^x (F_-(u) - F_+(u)) \, du = \begin{cases} 0, & x < -2, \\ \frac{x+2}{2} \geq 0, & x \in [-2, 0), \\ \frac{2-x}{2} \geq 0, & x \in [0, 2), \\ 0, & x \geq 2. \end{cases}$$

Hence  $X$  is increasing with respect to second order stochastic dominance. By inspection of the conditional distribution functions for  $\bar{X}$ , we see that  $\bar{X}$  is stochastically monotone, and therefore

also increasing with respect to second order stochastic dominance. Thus  $X$  and  $\bar{X}$  are co-monotone with respect to *second order* stochastic dominance, but they are *not* stochastically co-monotone (i.e. co-monotone with respect to first order stochastic dominance).

Let  $\mu = \text{Law}(X)$  and  $\nu = \text{Law}(\bar{X})$ . We now show that second order monotonicity is not sufficient for optimality of the Knothe–Rosenblatt rearrangement. We aim to solve (2.5) for the quadratic cost; that is we want to find

$$\mathcal{AW}_2^2(\mu, \nu) = \inf_{\substack{\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu) \\ \text{Law}(X, \bar{X}) = \pi}} \mathbb{E}[|X_1 - \bar{X}_1|^2 + |X_2 - \bar{X}_2|^2].$$

We first compute the cost induced by the Knothe–Rosenblatt rearrangement  $\pi_{\mu, \nu}^{\text{KR}}$ , shown in Figure 3a. Under this coupling, the first marginals are coupled monotonically, and so we have  $\mathbb{E}[|X_1 - \bar{X}_1|^2] = 0$ . Now, conditional on the event that  $(X_1, \bar{X}_1) = (\frac{1}{2}, \frac{1}{2})$ , we couple the second marginals monotonically and compute  $\mathbb{E}[|X_2 - \bar{X}_2|^2] = \frac{1}{2}(|0 - 2|^2 + |0 + 2|^2) = 4$ . Conditional on the complementary event  $(X_1, \bar{X}_1) = (-\frac{1}{2}, -\frac{1}{2})$ , we find  $\mathbb{E}[|X_2 - \bar{X}_2|^2] = \frac{1}{2}(|2 - 0|^2 + |-2 + 2|^2) = 2$ . We conclude that, for  $\text{Law}(X, \bar{X}) = \pi_{\mu, \nu}^{\text{KR}}$ ,

$$\mathbb{E}[|X_1 - \bar{X}_1|^2 + |X_2 - \bar{X}_2|^2] = 0 + \frac{1}{2}(4 + 2) = 3.$$

Now consider an alternative bicausal coupling  $\pi_{\mu, \nu}^{\text{AT}}$ , defined by taking the *antitone* coupling between the first marginals (i.e. the opposite of the monotone coupling) and then proceeding as in the Knothe–Rosenblatt rearrangement by coupling the conditional second marginals monotonically, as shown in Figure 3b. Under this coupling there is a non-zero contribution to the cost from the first marginals:  $\mathbb{E}[|X_1 - \bar{X}_1|^2] = 1$ . Conditional on  $(X_1, \bar{X}_1) = (\frac{1}{2}, -\frac{1}{2})$ , we have  $\mathbb{E}[|X_2 - \bar{X}_2|^2] = \frac{1}{2}(|0 - 0|^2 + |0 + 2|^2) = 2$ , and conditional on the complement,  $\mathbb{E}[|X_2 - \bar{X}_2|^2] = \frac{1}{2}(|2 - 2|^2 + |-2 + 2|^2) = 0$ . Thus taking  $\text{Law}(X, \bar{X}) = \pi_{\mu, \nu}^{\text{AT}}$  gives a total cost of

$$\mathbb{E}[|X_1 - \bar{X}_1|^2 + |X_2 - \bar{X}_2|^2] = 1 + \frac{1}{2}(2 + 0) = 2 < 3.$$

Hence the Knothe–Rosenblatt rearrangement is not optimal.

## 6. GENERAL OPTIMALITY OF THE SYNCHRONOUS COUPLING

In the previous sections we have identified different classes of coefficients of the SDEs (2.6) for which the synchronous coupling is optimal for (BCOT). In fact, the same optimality result holds for SDEs whose coefficients belong to two different classes; that is they satisfy one of Assumptions 3.7, 4.1, and 5.1.

**Theorem 6.1.** *Suppose that  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  each satisfy any one of Assumptions 3.7, 4.1, and 5.1, where the two pairs of coefficients may each satisfy a different assumption. Write  $\mu, \nu$  for the laws of  $X, \bar{X}$ , respectively. Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Then the synchronous coupling attains the value  $V_c(\mu, \nu)$  of the bicausal optimal transport problem (BCOT).*

*Proof.* First suppose that  $(b, \sigma)$  satisfies either the global Lipschitz condition Assumption 3.5 (this implies Assumption 3.7) or Assumption 4.1. Then take  $X^h$  to be the monotone Euler–Maruyama scheme defined in [8, Definition 3.13], or the transformed monotone semi-implicit Euler–Maruyama scheme defined in Definition 4.13, respectively. If  $(\bar{b}, \bar{\sigma})$  also satisfies one of the above conditions, then taking  $\bar{X}^h$  to be the respective numerical scheme,  $(X, \bar{X})$  and  $(X^h, \bar{X}^h)$  satisfy Assumption 3.1. Then we can conclude by Proposition 3.3. If  $(\bar{b}, \bar{\sigma})$  satisfies Assumption 3.7, but not necessarily a global Lipschitz condition, then we can approximate  $(\bar{b}, \bar{\sigma})$  locally uniformly in space by a sequence of Lipschitz functions that have uniform linear growth. Then we conclude by applying the stability result Lemma 3.4 to the previous case.

Now suppose that  $(b, \sigma)$  satisfies either Assumption 3.7 or Assumption 4.1, and  $(\bar{b}, \bar{\sigma})$  satisfies Assumption 5.1. Then we adapt the proof of Theorem 5.8 as follows. Take the transformation  $\bar{u}$  to be as in the proof of Theorem 5.8, and replace the transformation  $u$  with the identity. Then, following the proof of Theorem 5.8, we show that (BCOT) is equivalent to a bicausal transport problem between the laws of transformed SDEs, whose coefficients satisfy either Assumption 3.7 or Assumption 4.1. From the cases considered above, we know that the synchronous coupling is optimal for the transformed problem, and by equivalence of the problems, the synchronous coupling is also optimal for (BCOT).

Noting that the roles of  $\mu$  and  $\nu$  are interchangeable, all remaining cases are covered by Corollary 3.8 and Theorems 4.22 and 5.8.  $\square$

## 7. COMPUTATION OF THE ADAPTED WASSERSTEIN DISTANCE

Our results immediately suggest an efficient method for computing the value of a bicausal optimal transport problem. This is in sharp contrast to the general situation in classical and constrained optimal transport problems, when the optimiser is not known. Therefore our results contribute to the efficiency of computations in applications, such as in the calculation of robust bounds for stochastic optimisation problems, as discussed in Section 8. In Section 7.2 we present numerical examples, in which we explicitly compute adapted Wasserstein distances.

In classical optimal transport, entropic regularisation is used in computation. Sinkhorn's algorithm can be used to efficiently compute the value of the regularised problem, but not of the original unregularised problem; see Cuturi [17], Peyré and Cuturi [65]. For bicausal optimal transport in discrete time, Eckstein and Pammer [22], Pichler and Weinhardt [68] introduce two different adapted analogues of Sinkhorn's algorithm. However, each of these algorithms either include a discretisation of the marginal measures or work only on a discrete state space, and generalising them to both continuous time and space would be prohibitively costly. Under rather strong assumptions on the coefficients, Bion-Nadal and Talay [12] consider a stochastic control problem for SDEs, which is shown to be equivalent to finding the adapted 2-Wasserstein distance in [8, Proposition 2.2]. This control formulation gives rise to a Hamilton–Jacobi–Bellman equation, which one could solve numerically.

**7.1. Computation via the synchronous coupling.** Let  $c: [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  satisfy Assumption 2.11. Suppose that  $(b, \sigma)$  satisfies one of Assumptions 3.7, 4.1, and 5.1, and that  $(\bar{b}, \bar{\sigma})$  also satisfies one of these assumptions. Then, by Theorem 6.1, the synchronous coupling is optimal for  $V_c(\mu, \nu)$ . Writing  $(X, \bar{X})$  for the solution of the SDEs (2.6) driven by a common Brownian motion  $W$ , we thus have

$$V_c(\mu, \nu) = \mathbb{E} \left[ \int_0^T c_t(X_t, \bar{X}_t) dt \right].$$

In particular, taking  $c_t(x, y) = |x - y|^p$ , for all  $t \in [0, T]$ ,  $x, y \in \mathbb{R}$ , the adapted Wasserstein distance is given by

$$\mathcal{AW}_p^p(\mu, \nu) = \mathbb{E} \left[ \int_0^T |X_t - \bar{X}_t|^p dt \right].$$

For  $N \in \mathbb{N}$ , set  $h = T/N$  and take  $X^h, \bar{X}^h$  to be càdlàg processes adapted to the natural filtration of  $W$  such that  $X^h$  (resp.  $\bar{X}^h$ ) approximates  $X$  (resp.  $\bar{X}$ ) strongly in  $L^p$ . Define  $S^h := \mathbb{E}[\int_0^T |X_t^h - \bar{X}_t^h|^p dt]^{\frac{1}{p}}$ . Then

$$(7.1) \quad |\mathcal{AW}_p(\mu, \nu) - S^h| \leq \mathbb{E} \left[ \int_0^T |X_t - X_t^h|^p dt \right]^{\frac{1}{p}} + \mathbb{E} \left[ \int_0^T |\bar{X}_t - \bar{X}_t^h|^p dt \right]^{\frac{1}{p}} \xrightarrow{h \rightarrow 0} 0.$$

Note that each term in the sum on the right hand side of (7.1) depends on only one of the processes  $X, \bar{X}$  and so these terms can be estimated separately. Thus the problem of computing the adapted Wasserstein distance is reduced to the well-studied problem of numerically solving two one-dimensional SDEs. Monte Carlo methods can then be applied to compute these approximations.

There are now two error sources to consider in the numerical computation of  $\mathcal{AW}_p(\mu, \nu)$ : the convergence rate of the schemes  $X^h, \bar{X}^h$ , and the computational cost of Monte Carlo estimation for  $S^h$ . Given numerical schemes for which we have strong convergence rates, one can efficiently implement a multi-level Monte Carlo (MLMC) method to reduce the overall computational cost; see, for example, Giles and Szpruch [26].

**Remark 7.1.** The bound (7.1) shows that, for each SDE, we may take any numerical scheme that converges strongly in  $L^p$ . For example, for any SDE with Lipschitz coefficients, we can apply the Euler–Maruyama scheme defined in Section 2.3. In the case of discontinuous and exponentially growing drift, we proved strong  $L^p$  convergence of the transformed (monotone) semi-implicit Euler–Maruyama scheme in Theorem 4.18, which allows us to compute the adapted Wasserstein distance when one of the SDEs has coefficients satisfying Assumption 4.1. When additional conditions hold, however, we may exploit the existing convergence results that are discussed in Remark 4.4 to improve computational efficiency. While stochastic monotonicity of the scheme is essential to the proof of optimality in Theorem 4.22, this condition is not needed for the computation of the adapted Wasserstein distance. Under Assumption 5.1, suitable numerical schemes are discussed in Remark 5.2.

**Example 7.2.** As a toy example, consider the situation that  $b, \bar{b}$  are Lipschitz and  $\sigma, \bar{\sigma}$  are strictly positive constants. In this case, taking  $X^h, \bar{X}^h$  as the standard Euler–Maruyama schemes for  $X, \bar{X}$ , respectively, (7.1) implies that  $|\mathcal{AW}_p(\mu, \nu) - S^h| \leq C(p, T)N^{-1}$ . This follows from the order 1 strong convergence of the Euler–Maruyama scheme in the case of Lipschitz drift and additive noise, where it is equivalent to the Milstein scheme [53].

**7.2. Numerical examples.** In each of the examples below, we consider an SDE with solution  $X$  and, for  $k \in \mathbb{N}$ , the solution  $\bar{X}^{(k)}$  of an SDE obtained by some perturbation of the coefficients of the original SDE, where the magnitude of the perturbation increases with the parameter  $k$ . We fix the terminal time  $T = 1$  and consider the adapted Wasserstein distance  $\mathcal{AW}_2(\mu, \nu^{(k)})$ , where  $\mu$  and  $\nu^{(k)}$  denote the laws of  $X$  and  $\bar{X}^{(k)}$ , respectively. Supposing that the coefficients of the SDEs for  $X$  and  $\bar{X}^{(k)}$  satisfy the assumptions of Theorem 6.1, this distance is attained by the synchronous coupling. For  $h > 0$ , let  $X^h$  (resp.  $\bar{X}^{(k),h}$ ) denote numerical approximations of  $X$  (resp.  $\bar{X}^{(k)}$ ) that converge strongly in  $L^2$  with known rates and are each driven by a common one-dimensional Brownian motion. We fix the step size  $h = 2^{-12}$  for the numerical approximations and average over  $2^{12}$  sample paths. In light of the bound (7.1), we can then approximate the squared adapted Wasserstein distance  $\mathcal{AW}_2^2(\mu, \nu^{(k)})$  by the Monte Carlo estimate

$$(7.2) \quad \widehat{\mathcal{AW}}_2^2(\mu, \nu^{(k)}) = \frac{1}{2^{12}(2^{12} + 1)} \sum_{i=1}^{2^{12}} \sum_{j=0}^{2^{12}} |X_{jh}^h(\omega_i) - \bar{X}_{jh}^{(k),h}(\omega_i)|^2.$$

In each example below, we plot  $\widehat{\mathcal{AW}}_2^2(\mu, \nu^{(k)})$  against the size of the perturbation.

**Example 7.3** (An SDE with discontinuous drift). Consider SDEs

$$(7.3) \quad \begin{aligned} dX_t &= dW_t, \quad X_0 = x_0, \\ d\bar{X}_t^{(k)} &= \frac{k}{10} \text{sign}(\bar{X}_t^{(k)}) dt + dW_t, \quad \bar{X}_0^{(k)} = x_0, \quad k \in \{1, \dots, 10\}. \end{aligned}$$

The solution  $X$  of the first SDE in (7.3) can be approximated by the Euler–Maruyama scheme with strong convergence order 1. For the SDEs in the second line of (7.3), however, Ellinger, Müller-Gronbach, and Yaroslavl'tseva [23], Müller-Gronbach and Yaroslavl'tseva [58], Przybyłowicz, Schwarz, and Szölgyenyi [72] show an upper bound of  $3/4$  for the strong convergence rate of any numerical scheme that uses finitely many approximations of the driving Brownian motion (such as the Euler–Maruyama scheme). This is one of the structural properties that distinguishes the class of SDEs with discontinuous drift. For the adapted Wasserstein distance between the induced measures, however, we observe no gap comparable to the gap between the strong convergence rates. Each of the SDEs in (7.3) satisfies Assumption 4.1, and so Theorem 6.1 implies that the synchronous coupling attains the adapted Wasserstein distance between the laws of the solutions. Thus, we can estimate this distance by (7.2), taking  $X^h, \bar{X}^{(k),h}$  to be the respective Euler–Maruyama schemes, each driven by a common Brownian motion. The relationship between this approximation of the adapted Wasserstein distance and the size of the jump in the drift is shown in Figure 4.

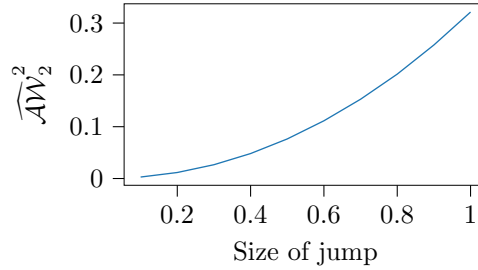


FIGURE 4. Approximation of the adapted Wasserstein distance for the SDEs (7.3).

**Example 7.4** (CIR model). Consider a CIR process  $(X_t)_{t \in [0,1]}$ , solving the SDE

$$(7.4) \quad dX_t = \kappa(\eta - X_t) dt + \gamma\sqrt{X_t} dW_t, \quad X_0 = x_0,$$

for some  $\kappa \in \mathbb{R}$ ,  $\eta \in [0, \infty)$ ,  $\gamma, x_0 \in (0, \infty)$ . We suppose that  $2\kappa\eta \geq \gamma^2$ , so that  $X_t > 0$  for all  $t \in [0, 1]$  by Feller's test; see, for example, [41, Theorem 5.29]. We call the parameter  $\eta$  the mean reversion level,  $\kappa$  the mean reversion speed, and  $\gamma$  the diffusion parameter. Such a process is commonly used to model interest rates.

Here we have a Lipschitz-continuous drift coefficient and a  $1/2$ -Hölder continuous diffusion coefficient. The classical result of Yamada and Watanabe [86, Theorem 1] gives pathwise uniqueness for (7.4). Thus the coefficients of (7.4) satisfy Assumption 3.7. Consider also the process  $\bar{X}^{(k)}$  solving the same SDE (7.4) with perturbed parameters  $\bar{\kappa}, \bar{\eta}, \bar{\gamma}$  and  $\bar{X}_0^{(k)} = x_0$ . Since the coefficients of this SDE also satisfy Assumption 3.7, we have by Theorem 6.1 that the synchronous coupling attains the adapted Wasserstein distance between their laws.

Since the CIR process remains positive, its approximation requires a modification to the standard Euler–Maruyama scheme. For example, we can take the symmetrised Euler–Maruyama scheme that is defined in Berkaoui, Bossy, and Diop [11] by taking the absolute value at each step of the Euler–Maruyama scheme. Under the condition that  $2\kappa\eta \geq \gamma^2$ , [11, Theorem 2.2] shows that the symmetrised Euler–Maruyama scheme converges strongly in  $L^2$  at rate  $1/2$ . Note that [30, Theorem 1.1] shows that this convergence rate does not hold when  $2\kappa\eta < \gamma^2$ . In this example we fix the parameters  $\kappa = \eta = \gamma = 1$  for the process  $X$ , and for the process  $\bar{X}^{(k)}$  we perturb each of these parameters in turn, fixing the other two equal to those of the process  $X$ . For each perturbation, we approximate  $X, \bar{X}^{(k)}$  by the numerical scheme described above, taking the same Brownian increments in both  $X^h$  and  $\bar{X}^{(k),h}$ . We then estimate the adapted Wasserstein distance between the laws of  $X$  and  $\bar{X}^{(k)}$  by (7.2).

Figure 5 shows an approximation of the adapted Wasserstein distance plotted against the absolute size of the perturbation for each parameter on the left-hand side, and the same on a log-scale on the right-hand side. As expected, we observe clearly in Figure 5 that increasing the perturbation in each of the parameters increases the adapted Wasserstein distance between the reference measure and the measure induced by the process with perturbed parameters. We observe that a perturbation of the diffusion parameter has a bigger effect than that of the mean reversion level, which has a much bigger effect than a perturbation of the mean reversion speed. This behaviour depends, however, on the choice of the parameters for the process  $X$ . As the log-plot confirms, the adapted Wasserstein distance grows exponentially in the size of the perturbation. This observation confirms that the adapted Wasserstein distance successfully captures the difference between CIR models with different parameters.

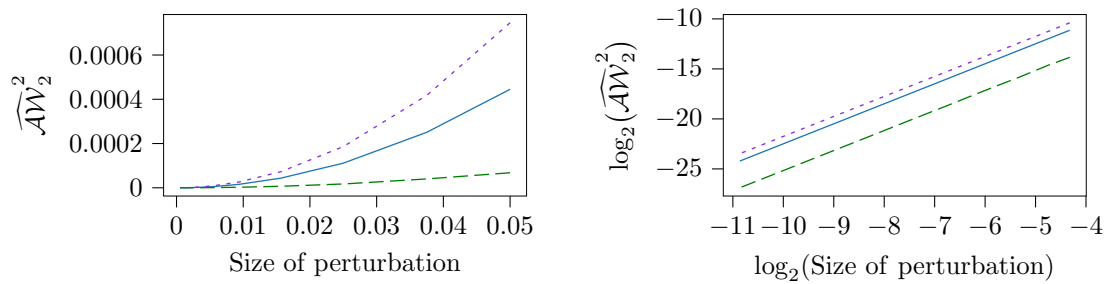


FIGURE 5. Approximation of the adapted Wasserstein distance for the CIR process; the dotted purple line shows the effect of a perturbed diffusion parameter  $\gamma$ , the solid blue line shows the effect of a perturbed mean reversion level  $\eta$ , and the dashed green line shows the effect of a perturbed mean reversion speed  $\kappa$ .

## 8. APPLICATION TO ROBUST OPTIMISATION

We finally turn to applications in the robust approach to stochastic optimisation. Several results have been proven on stability of various stochastic optimisation problems [1, 6] and (Lipschitz) continuity of optimal stopping problems [1, 2, 7], in both discrete and continuous time, with respect to the adapted Wasserstein distance. In a discrete-time setup, the adapted Wasserstein distance has also been applied to quantify model-sensitivity of multi-period optimisation problems in [4], [9], and [37], with the latter also treating model-sensitivity of continuous-time problems.

As a concrete application of the adapted Wasserstein distance to robust optimisation, we consider optimal stopping problems for continuous-time stochastic processes. Such problems are often analytically intractable and the numerical approximation of their solution may be computationally intensive; see, e.g. [10] and the references therein. In Theorem 8.1 we prove Lipschitz continuity of optimal stopping problems with respect to the adapted Wasserstein distance, which admits an efficient numerical approximation, following the approach of Section 7.1. Hence, our methods provide computationally feasible bounds on the discrepancy between the values of optimal stopping problems under different probability measures, making such problems tractable for applications.

**8.1. Optimal stopping.** We consider the following class of optimal stopping problems. For an  $\mathcal{F}$ -adapted process  $L: [0, T] \times \Omega \rightarrow \mathbb{R}$  and  $p \geq 1$ , define the value function  $v^L: \mathcal{P}_p(\Omega) \rightarrow \mathbb{R}$  of the associated optimal stopping problem for any  $\mu \in \mathcal{P}_p(\Omega)$  by

$$v^L(\mu) := \inf_{\tau} \mathbb{E}^{\mu}[L(\tau(\omega), \omega)],$$

where the infimum is taken over all stopping times with respect to  $\mathcal{F}^\mu$ , the completion of the canonical filtration under  $\mu$ . We prove the following Lipschitz estimate, which quantifies the model uncertainty in the choice of the measure  $\mu \in \mathcal{P}_p(\Omega)$ .

**Theorem 8.1.** *Let  $p \geq 1$ , let  $L: [0, T] \times \Omega \rightarrow \mathbb{R}$  be  $\mathcal{F}$ -adapted and suppose that there exists a Lipschitz constant  $C_L > 0$  such that, for all  $t \in [0, T]$ ,  $\gamma, \bar{\gamma} \in \Omega$ ,*

$$|L(t, \gamma) - L(t, \bar{\gamma})| \leq C_L \|\gamma - \bar{\gamma}\|_p.$$

*Then, for any  $\mu, \nu \in \mathcal{P}_p(\Omega)$ , we have the bound*

$$|v^L(\mu) - v^L(\nu)| \leq C_L \mathcal{AW}_p(\mu, \nu).$$

**Example 8.2.** Let  $f: \mathbb{R} \rightarrow \mathbb{R}$  be Lipschitz and define  $L: [0, T] \times \Omega \rightarrow \mathbb{R}$ , for any  $t \in [0, T]$ ,  $\gamma \in \Omega$ , by

$$L(t, \gamma) := f\left(\int_0^t \gamma_s \, ds\right).$$

Then  $L$  satisfies the conditions of Theorem 8.1 for any  $p \geq 1$ . In this example, the value  $v^L$  of the corresponding optimal stopping problem has a financial interpretation as an American-style Asian option.

This result could also be derived from [1, Proposition 4.4], but we rather provide a direct proof, following a similar approach to the proof of the analogous discrete-time result [7, Lemma 7.1].

*Proof of Theorem 8.1.* Let  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$  and let  $(X, Y)$  be a pair of random variables on  $\Omega \times \Omega$  with joint law  $\pi$ . As before, let  $\mathcal{F}^X$  and  $\mathcal{F}^Y$  denote the completions of the natural filtrations of  $X$  and  $Y$  under the measures  $\mu$  and  $\nu$ , respectively. Fix  $\varepsilon > 0$ , and take an  $\varepsilon$ -optimal  $\mathcal{F}^Y$ -stopping time  $\tau^Y$ ; i.e.

$$(8.1) \quad \mathbb{E}[L(\tau^Y, Y)] \leq v^L(\nu) + \varepsilon.$$

For each  $u \in [0, 1]$ , define  $\theta_u := \inf\{t \geq 0 : \pi(\tau^Y \leq t \mid \mathcal{F}_t^X) \geq u\}$ . We claim that  $\theta_u$  is an  $\mathcal{F}^X$ -stopping time. By continuity of the paths of processes  $X$  and  $Y$ , we know that, for any  $u \in [0, 1]$  and  $t \in [0, T]$ ,

$$\{\theta_u \leq t\} = \{\pi(\tau^Y \leq t \mid \mathcal{F}_t^X) \geq u\}.$$

So it suffices to show that, for each  $t \in [0, T]$ , the conditional probability  $\pi(\tau^Y \leq t \mid \mathcal{F}_t^X)$  is  $\mathcal{F}_t^X$ -measurable. Since  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$ , the causality condition (2.2) states that, under  $\pi$ ,  $\mathcal{F}_t^Y$  is conditionally independent of  $\mathcal{F}_t^X$ , given  $\mathcal{F}_t^X$ . An equivalent formulation of this conditional independence condition (see, e.g. Kallenberg [39, Theorem 8.9]) is that for any  $F \in \mathcal{F}_t^Y$ ,

$$(8.2) \quad \pi(F \mid \mathcal{F}_t^X \vee \mathcal{F}_t^X) = \pi(F \mid \mathcal{F}_t^X).$$

In fact,  $\mathcal{F}_t^X \subset \mathcal{F}_t^X$ , and so  $\mathcal{F}_t^X \vee \mathcal{F}_t^X = \mathcal{F}_t^X$ . By the definition of a stopping time,  $\{\tau^Y \leq t\} \in \mathcal{F}_t^Y$ . Therefore (8.2) gives us

$$\pi(\tau^Y \leq t \mid \mathcal{F}_t^X) = \pi(\tau^Y \leq t \mid \mathcal{F}_t^X),$$

which is then  $\mathcal{F}_t^X$ -measurable. This proves our claim that  $\theta_u$  is an  $\mathcal{F}^X$ -stopping time for each  $u \in [0, 1]$ . Thus

$$v^L(\mu) \leq \inf_{u \in [0, 1]} \mathbb{E}[L(\theta_u, X)] \leq \int_0^1 \mathbb{E}[L(\theta_u, X)] \, du.$$

Applying the tower property of conditional expectation and Fubini's theorem to exchange the order of integration, we arrive at

$$(8.3) \quad v^L(\mu) \leq \mathbb{E}\left[\int_0^1 \mathbb{E}^\pi[L(\theta_u, X) \mid \mathcal{F}_t^X] \, du\right].$$



Now fix a path  $\gamma \in \Omega$  of the process  $X$ . We can define a probability measure  $\mathbb{P}_\gamma$  on  $[0, T]$  by  $\mathbb{P}_\gamma([0, t)) = \pi(\tau^Y \leq t \mid \mathcal{F}_T^X)(\gamma)$ , for all  $t \in [0, T]$ . Then  $u \mapsto \theta_u(\gamma)$  is the quantile function of  $\mathbb{P}_\gamma$ , and we see that

$$\mathbb{E}[L(\tau^Y, X) \mid \mathcal{F}_T^X](\gamma) = \int_0^T L(t, \gamma) d\mathbb{P}_\gamma(t) = \int_0^1 L(\theta_u(\gamma), \gamma) du = \int_0^1 \mathbb{E}[L(\theta_u, X) \mid \mathcal{F}_T^X](\gamma) du.$$

Taking the expectation with respect to  $\pi$  and substituting into (8.3), we find

$$v^L(\mu) \leq \mathbb{E}^\pi \left[ \mathbb{E}[L(\tau^Y, X) \mid \mathcal{F}_T^X](\omega) \right] = \mathbb{E}[L(\tau^Y, X)].$$

By (8.1), the Lipschitz continuity of  $L$  in the second argument, and Jensen's inequality, we have

$$v^L(\mu) - v^L(\nu) \leq C_L \cdot \mathbb{E}[\|X - Y\|_p] + \varepsilon \leq C_L \cdot \mathbb{E}[\|X - Y\|_p^p]^{\frac{1}{p}} + \varepsilon.$$

Exchanging the roles of  $\mu$  and  $\nu$  and noting that  $\varepsilon > 0$  and  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$  were chosen arbitrarily completes the proof.  $\square$

## APPENDIX A. PROOFS FROM SECTION 3

*Proof of Proposition 3.3.* For  $h > 0$ , Assumption 3.1 (iv) implies that the Knothe–Rosenblatt rearrangement is optimal for (2.5) between  $\mu^h$  and  $\nu^h$ , by Proposition 2.9. Taking a common Brownian motion  $W = \bar{W}$  in Assumption 3.1 (iii), the Knothe–Rosenblatt rearrangement is equal to  $\text{Law}(X^h, \bar{X}^h)$ , by Proposition 2.8. This together with Assumption 3.1 (v) allows us to conclude that the limit (3.1) holds and the synchronous coupling attains the value (BCOT) of the continuous-time bicausal transport problem. This final step is a straightforward modification of the proof of stability of bicausal optimal transport in [8, Proposition 2.6, Corollary 2.7], and we omit the details here.  $\square$

*Proof of Lemma 3.4.* Let  $\pi \in \text{Cpl}_{\text{bc}}(\mu, \nu)$ . Then by [8, Proposition 2.2] on the relation between bicausal couplings and correlated Brownian motions, there exists a correlated Brownian motion  $(W, \bar{W})$  such that  $\pi = \text{Law}(X, \bar{X})$  when (2.6) is driven by  $(W, \bar{W})$ . Taking the same driving Brownian motion, we write  $\pi^N := \text{Law}(X^N, \bar{X}^N) \in \text{Cpl}_{\text{bc}}(\mu^N, \nu^N)$ , for each  $N \in \mathbb{N}$ . Interpreting each such system of SDEs as a two-dimensional SDE, the stability result of Stroock and Varadhan [80, Theorem 11.1.4] implies the weak convergence  $\pi^N \rightharpoonup \pi$  as  $N \rightarrow \infty$ . Moreover, the uniform linear growth bound implies that, for any  $p \geq 1$ , there exists a constant  $C > 0$  such that  $\mathbb{E}^{\pi^N}[\sup_{t \in [0, T]} |(\omega_t, \bar{\omega}_t)|^p] \leq C$  for all  $N \in \mathbb{N}$ ; see, e.g. [25, Lemma 3.8]. Thus  $\mathbb{E}^{\pi^N}[\phi(\omega, \bar{\omega})] \rightarrow \mathbb{E}^\pi[\phi(\omega, \bar{\omega})]$ , for any functional  $\phi: \Omega \times \Omega \rightarrow \mathbb{R}$  that is continuous with order  $p$  polynomial growth with respect to the uniform norm on  $\Omega \times \Omega$ ; see, e.g. [84, Definition 6.8]. We claim that  $(\omega, \bar{\omega}) \mapsto \int_0^T c_t(\omega_t, \bar{\omega}_t) dt$  satisfies these conditions. Indeed the polynomial growth of order  $p$  follows immediately from the bound (2.3), which holds uniformly in  $t$ . Continuity follows from (2.3) and the continuity of the maps  $(x, y) \mapsto c_t(x, y)$  for each  $t \in [0, T]$  via the Lebesgue–Vitali dominated convergence theorem. Hence

$$(A.1) \quad \mathbb{E}^{\pi^N} \left[ \int_0^T c_t(\omega_t, \bar{\omega}_t) dt \right] \xrightarrow{N \rightarrow \infty} \mathbb{E}^\pi \left[ \int_0^T c_t(\omega_t, \bar{\omega}_t) dt \right].$$

Now take a common one-dimensional Brownian motion driving (2.6) for each set of coefficients. Then we have

$$\pi_{\mu^N, \nu^N}^{\text{sync}} = \text{Law}(X^N, \bar{X}^N) \xrightarrow{N \rightarrow \infty} \text{Law}(X, \bar{X}) = \pi_{\mu, \nu}^{\text{sync}}.$$

Given that  $\pi_{\mu^N, \nu^N}^{\text{sync}}$  minimises the left-hand side of (A.1) over  $\text{Cpl}_{\text{bc}}(\mu^N, \nu^N)$  for each  $N \in \mathbb{N}$ , we have that  $\pi_{\mu, \nu}^{\text{sync}}$  minimises the right-hand side of (A.1) over  $\text{Cpl}_{\text{bc}}(\mu, \nu)$  and the required convergence holds.  $\square$

*Proof of Proposition 3.6.* Under Assumption 3.5, we can show that the solutions of (2.6) and the corresponding monotone Euler–Maruyama schemes satisfy Assumption 3.1, following the proofs of [8, Lemma 3.14, Lemma 3.15, Proposition 3.16]. The result then follows from Proposition 3.3.  $\square$

*Proof of Corollary 3.8.* For each SDE, existence of a weak solution is guaranteed by Skorokhod [78, Chapter 3, Section 3], and given that we assumed pathwise uniqueness, Yamada and Watanabe [86, Corollary 1] implies that there exists a unique strong solution. We can approximate the coefficients  $(b, \sigma)$  and  $(\bar{b}, \bar{\sigma})$  locally uniformly in space by coefficients satisfying Assumption 3.5 and the conditions of Lemma 3.4. We conclude by combining Proposition 3.6 and Lemma 3.4.  $\square$

## APPENDIX B. PROOFS FROM SECTION 4.1

*Proof of Lemma 4.6.* Direct computation of the first and second derivatives of  $\bar{\phi}_k$  yield the first three points. We also observe that the first derivative is continuous and bounded, and hence  $\bar{\phi}_k$  is Lipschitz. The second derivative is defined almost everywhere and is continuous and bounded on  $(-\infty, \xi_k) \cup (\xi_k, \infty)$ . Combined with the fact that the one-sided second derivatives at  $\xi_k$  exist and are finite, this shows that  $\bar{\phi}'_k$  is Lipschitz. Finally, differentiating  $\bar{\phi}'_k$  once again on  $(-\infty, \xi_k) \cup (\xi_k, \infty)$  gives a bounded third derivative, and hence  $\bar{\phi}''_k$  is piecewise Lipschitz with discontinuity point  $\xi_k$ .  $\square$

*Proof of Lemma 4.7.* By Lemma 4.6,  $\bar{\phi}_k(x) = 0$  for  $|x - \xi_k| \geq c_0$  for each  $k \in \{1, \dots, m\}$ , and so we have  $G(x) = x$  for  $x \in \mathbb{R} \setminus \bigcup_{k \in \{1, \dots, m\}} (\xi_k - c_0, \xi_k + c_0)$ . Moreover, for any  $k \in \{1, \dots, m\}$  and  $x \in [\xi_k - c_0, \xi_k + c_0]$ ,  $G(x) = x + \alpha_k \bar{\phi}_k(x)$  and  $G'(x) = 1 + \alpha_k \bar{\phi}'_k(x)$ , implying that  $G'(x) \in [1 - 6c_0|\alpha_k|, 1 + 6c_0|\alpha_k|]$  by Lemma 4.6. Thus, as in [47, Lemma 2.2], we observe that our choice of  $c_0$  in (4.7) guarantees that  $G$  is strictly increasing on  $\mathbb{R}$  and therefore admits a strictly increasing inverse  $G^{-1}: \mathbb{R} \rightarrow \mathbb{R}$ . From the definition of  $G$  and the properties proved in Lemma 4.6, we immediately deduce that  $G$  is bounded and Lipschitz with bounded Lipschitz first derivative  $G'$  and bounded piecewise Lipschitz almost-everywhere second derivative  $G''$ , with discontinuity points  $\xi_1, \dots, \xi_m$ . By the inverse function theorem,  $G^{-1}$  has the same properties.  $\square$

*Proof of Lemma 4.9.* Fix  $c > 0$  and let  $L_f, K_f, \gamma, \eta$  denote the constants from Assumption 4.1 (A1) that appear in the one-sided Lipschitz and local Lipschitz conditions for  $f$ . By Remark 4.5 and the continuity of  $f$ , we have that  $f$  is Lipschitz on the compact interval  $[\xi_1 - c, \xi_m + c]$  and hence satisfies Assumption 4.1 (A1) (i) and (ii) on  $[\xi_1 - c, \xi_m + c]$ . Suppose that  $x \in [\xi_1 - c, \xi_m + c)$ ,  $y \in (-\infty, \xi_1 - c]$ . Then, since  $y < \xi_1 - c < x$ , we have

$$\begin{aligned} (x - y)(f(x) - f(y)) &= \frac{x - y}{x - (\xi_1 - c)} (x - (\xi_1 - c))(f(x) - f(\xi_1 - c)) \\ &\quad + \frac{x - y}{(\xi_1 - c) - y} ((\xi_1 - c) - y)(f(\xi_1 - c) - f(y)) \\ &\leq \frac{x - y}{x - (\xi_1 - c)} L_f |x - (\xi_1 - c)|^2 + \frac{x - y}{(\xi_1 - c) - y} L_f |(\xi_1 - c) - y|^2 \\ &= L_f (x - y) (x - (\xi_1 - c) + (\xi_1 - c) - y) = L_f |x - y|^2. \end{aligned}$$

Similarly,

$$\begin{aligned} |f(x) - f(y)| &\leq |f(x) - f(\xi_1 - c)| + |f(\xi_1 - c) - f(y)| \\ &\leq K_f (e^{\gamma|x|^\eta} + e^{\gamma|\xi_1 - c|^\eta}) (x - (\xi_1 - c)) + K_f (e^{\gamma|\xi_1 - c|^\eta} + e^{\gamma|y|^\eta}) ((\xi_1 - c) - y) \\ &\leq K_f (2e^{\gamma|x|^\eta} + e^{\gamma|y|^\eta}) (x - y) + K_f (e^{\gamma|x|^\eta} + 2e^{\gamma|y|^\eta}) (x - y) \\ &= 3K_f (e^{\gamma|x|^\eta} + e^{\gamma|y|^\eta}) |x - y|. \end{aligned}$$

The remaining cases are treated analogously, and we conclude that  $f$  satisfies Assumption 4.1 (A1) (i) and (ii) globally on  $\mathbb{R}$ .  $\square$

*Proof of Lemma 4.10.* First, as in [47], we note that the constants  $\alpha_k$ ,  $k \in \{1, \dots, m\}$  are chosen in (4.6) such that  $\tilde{b}$  is continuous on  $\mathbb{R}$ . Indeed by Lemma 4.6, for each  $k \in \{1, \dots, m\}$ ,  $G'(\xi_k) = 1 + \alpha_k \bar{\phi}'_k(\xi_k) = 1$ ,  $G''(\xi_k+) = \alpha_k \bar{\phi}''_k(\xi_k+) = 2\alpha_k$ ,  $G''(\xi_k-) = -2\alpha_k$ , and so

$$(bG' + \frac{1}{2}\sigma^2 G'')(\xi_k+) - (bG' + \frac{1}{2}\sigma^2 G'')(\xi_k-) = b(\xi_k+) - b(\xi_k-) + 2\alpha_k \sigma^2(\xi_k) = 0.$$

Then, by Lemma 4.7,  $bG' + \frac{1}{2}\sigma^2 G''$  is continuous on  $\mathbb{R}$ . Also by Lemma 4.7,  $G^{-1}$  is Lipschitz, and so  $\tilde{b}$  is continuous on  $\mathbb{R}$ .

Let  $c > 0$ . Then, on each finite interval  $(\xi_1 - c, \xi_1)$ ,  $(\xi_m, \xi_m + c)$ ,  $(\xi_k, \xi_{k+1})$ , for  $k \in \{1, \dots, m-1\}$ , we have that  $b$ ,  $\sigma$ , and  $G'$  and  $G''$  are bounded and Lipschitz by Remark 4.5, Assumption 4.1 (A2), and Lemma 4.7, respectively. Thus on each such interval,  $bG' + \frac{1}{2}\sigma^2 G''$  is Lipschitz as the sum of products of bounded Lipschitz functions. Taking the composition with the Lipschitz function  $G^{-1}$ , we see that  $\tilde{b}$  is Lipschitz, and thus satisfies Assumption 4.1 (A1) (i) and (ii), on each interval  $(\xi_1 - c, \xi_1)$ ,  $(\xi_m, \xi_m + c)$ ,  $(\xi_k, \xi_{k+1})$ , for  $k \in \{1, \dots, m-1\}$ . On the intervals  $(-\infty, \xi_1 - c)$  and  $(\xi_m + c, +\infty)$ , we have that  $\tilde{b} = b$  and thus satisfies Assumption 4.1 (A1) (i) and (ii). Since  $\tilde{b}$  is continuous, Lemma 4.9 implies statements (i) and (ii) of the lemma. Statement (iii) follows from statement (ii).

Now,  $\sigma G'$  is globally Lipschitz, since it is the product of bounded Lipschitz functions on  $[\xi_1 - c_0, \xi_m + c_0]$  and is equal to  $\sigma$  on  $(-\infty, \xi_1 - c_0] \cup [\xi_m + c_0, +\infty)$ ; cf. [47, Lemma 2.5]. Taking the composition with the Lipschitz function  $G^{-1}$  preserves the Lipschitz property, and this concludes the proof of statement (iv).

Finally, statement (v) of the lemma follows from statements (i) and (iv).  $\square$

## REFERENCES

- [1] B. Acciaio, J. Backhoff-Veraguas, and A. Zalashko. Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Process. Appl.*, 130(5):2918–2953, 2020. doi: 10.1016/j.spa.2019.08.009.
- [2] D. J. Aldous. Weak convergence and general theory of processes. Unpublished monograph: Department of Statistics, University of California, Berkeley, 1981.
- [3] 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. doi: 10.1007/978-3-7643-8722-8.
- [4] B. Analui and G. C. Pflug. On distributionally robust multiperiod stochastic optimization. *Comput. Manag. Sci.*, 11:197–220, 2014. doi: 10.1007/s10287-014-0213-y.
- [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. 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. 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 Related Fields*, 178(3-4):1125–1172, 2020. doi: 10.1007/s00440-020-00993-8.
- [8] J. Backhoff-Veraguas, S. Källblad, and B. A. Robinson. Adapted Wasserstein distance between the laws of SDEs. *Stochastic Process. Appl.*, 189:104689, 2025. doi: 10.1016/j.spa.2025.104689.

- [9] D. Bartl and J. Wiesel. Sensitivity of multiperiod optimization problems with respect to the adapted Wasserstein distance. *SIAM J. Financial Math.*, 14(2):704–720, 2023. doi: 10.1137/22M1537746.
- [10] S. Becker, P. Cheridito, A. Jentzen, and T. Welti. Solving high-dimensional optimal stopping problems using deep learning. *Eur. J. Appl. Math.*, 32(3):470–514, 2021. doi: 10.1017/S0956792521000073.
- [11] A. Berkouki, M. Bossy, and A. Diop. Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence. *ESAIM Probab. Stat.*, 12:1–11, 2008. doi: 10.1051/ps:2007030.
- [12] 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. doi: 10.1214/18-AAP1423.
- [13] H. W. Block, W. S. Griffith, and T. H. Savits.  $L$ -superadditive structure functions. *Adv. in Appl. Probab.*, 21(4):919–929, 1989. doi: 10.2307/1427774.
- [14] P. Bonnier, C. Liu, and H. Oberhauser. Adapted topologies and higher rank signatures. *Ann. Appl. Probab.*, 33(3):2136–2175, 2023. doi: 10.1214/22-aap1862.
- [15] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991. doi: 10.1002/cpa.3160440402.
- [16] P. Cheridito and S. Eckstein. Optimal transport and Wasserstein distances for causal models. *Bernoulli*, 31(2):1351–1376, 2025. doi: 10.3150/24-bej1773.
- [17] M. Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26, pages 2292–2300, 2013.
- [18] D. J. Daley. Stochastically monotone Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 10:305–317, 1968. doi: 10.1007/BF00531852.
- [19] K. Dareiotis and M. Gerencsér. On the regularisation of the noise for the Euler–Maruyama scheme with irregular drift. *Electron. J. Probab.*, 25:1–18, 2020. doi: 10.1214/20-ejp479.
- [20] K. Dareiotis, M. Gerencsér, and K. Lê. Quantifying a convergence theorem of Gyöngy and Krylov. *Ann. Appl. Probab.*, 33(3):2291–2323, 2023. doi: 10.1214/22-aap1867.
- [21] A. M. Davie. Approximation of SDE solutions using local asymptotic expansions. *Stoch. Partial Differ. Equ. Anal. Comput.*, 10(3):720–734, 2022. doi: 10.1007/s40072-021-00232-8.
- [22] S. Eckstein and G. Pammer. Computational methods for adapted optimal transport. *Ann. Appl. Probab.*, 34(1A):675–713, 2024. doi: 10.1214/23-aap1975.
- [23] S. Ellinger, T. Müller-Gronbach, and L. Yaroslavtseva. On optimal error rates for strong approximation of SDEs with a drift coefficient of fractional Sobolev regularity, 2024. arXiv:2402.13732.
- [24] M. Gerencsér, G. Lampl, and C. Ling. The Milstein scheme for singular SDEs with Hölder continuous drift. *IMA J. Numer. Anal.*, 45(5):3077–3108, 2025. doi: 10.1093/imanum/drae083.
- [25] I. I. Gihman and A. V. Skorohod. *Controlled Stochastic Processes*. Springer-Verlag, New York-Heidelberg, 1979. ISBN 0-387-90410-7. doi: 10.1007/978-1-4612-6202-2. Translated from the Russian by Samuel Kotz.
- [26] M. Giles and L. Szpruch. Multilevel Monte Carlo methods for applications in finance. In *Recent Developments in Computational Finance*, volume 14 of *Interdiscip. Math. Sci.*, pages 3–47. World Sci. Publ., Hackensack, NJ, 2013. doi: 10.1142/9789814436434\_0001.
- [27] I. Gyöngy. A note on Euler’s approximations. *Potential Anal.*, 8(3):205–216, 1998. doi: 10.1023/A:1008605221617.
- [28] I. Gyöngy and N. Krylov. Existence of strong solutions for Itô’s stochastic equations via approximations. *Probab. Theory Related Fields*, 105(2):143–158, 1996. doi: 10.1007/BF01203833.

- [29] I. Gyöngy and M. Rásonyi. A note on Euler approximations for SDEs with hölder continuous diffusion coefficients. *Stochastic Process. Appl.*, 121(10):2189–2200, 2011. doi: 10.1016/j.spa.2011.06.008.
- [30] M. Hefter and A. Jentzen. On arbitrarily slow convergence rates for strong numerical approximations of Cox–Ingersoll–Ross processes and squared Bessel processes. *Finance Stoch.*, 23(1):139–172, 2019. doi: 10.1007/s00780-018-0375-5.
- [31] D. J. Higham, X. Mao, and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, 40(3):1041–1063, 2002. doi: 10.1137/S0036142901389530.
- [32] H. Hu and S. Gan. Strong convergence of the tamed Euler scheme for scalar SDEs with superlinearly growing and discontinuous drift coefficient, 2022. arXiv:2206.00088.
- [33] Y. Hu. Semi-implicit Euler-Maruyama scheme for stiff stochastic equations. In H. Körezlioglu, B. Øksendal, and A. S. Üstünel, editors, *Stochastic Analysis and Related Topics, V (Silivri, 1994)*, volume 38 of *Progr. Probab.*, pages 183–202. Birkhäuser Boston, Boston, MA, 1996. doi: 10.1007/978-1-4612-2450-1\_9.
- [34] M. Hutzenthaler and A. Jentzen. Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients. *Mem. Amer. Math. Soc.*, 236(1112), 2015. doi: 10.1090/memo/1112.
- [35] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467(2130):1563–1576, 2011. doi: 10.1098/rspa.2010.0348.
- [36] M. Hutzenthaler, A. Jentzen, and P. E. Kloeden. Strong convergence of an explicit numerical method for SDEs with nonglobally Lipschitz continuous coefficients. *Ann. Appl. Probab.*, 22(4):1611–1641, 2012. doi: 10.1214/11-AAP803.
- [37] Y. Jiang and J. Oblój. Sensitivity of causal distributionally robust optimization, 2024. arXiv:2408.17109.
- [38] B. Jourdain and G. Pagès. Convex ordering of solutions to one-dimensional SDEs, 2023. arXiv:2312.09779.
- [39] O. Kallenberg. *Foundations of Modern Probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer, Cham, third edition, 2021. ISBN 978-3-030-61871-1; 978-3-030-61870-4. doi: 10.1007/978-3-030-61871-1.
- [40] L. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.
- [41] 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.
- [42] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*, volume 23 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1992. ISBN 3-540-54062-8. doi: 10.1007/978-3-662-12616-5.
- [43] H. Knothe. Contributions to the theory of convex bodies. *Michigan Math. J.*, 4:39–52, 1957. doi: 10.1307/mmj/1028990175.
- [44] N. V. Krylov. A simple proof of the existence of a solution of Itô’s equation with monotone coefficients. *Theory Probab. Appl.*, 35(3):583–587, 1991. doi: 10.1137/1135082.
- [45] R. Lassalle. Causal transference plans and their Monge–Kantorovich problems. *Stoch. Anal. Appl.*, 36(3):452–484, 2018. doi: 10.1080/07362994.2017.1422747.

- [46] G. Leobacher and M. Szölgényi. A numerical method for SDEs with discontinuous drift. *BIT*, 56(1):151–162, 2016. doi: 10.1007/s10543-015-0549-x.
- [47] G. Leobacher and M. Szölgényi. A strong order 1/2 method for multidimensional SDEs with discontinuous drift. *Ann. Appl. Probab.*, 27(4):2383–2418, 2017. doi: 10.1214/16-AAP1262.
- [48] G. Leobacher and M. Szölgényi. Numerical methods for SDEs with drift discontinuous on a set of positive reach. *Internationale Mathematische Nachrichten*, 235(1):1–16, 2017. doi: 10.48550/arXiv.1708.06188.
- [49] G. Leobacher and M. Szölgényi. Convergence of the Euler–Maruyama method for multidimensional SDEs with discontinuous drift and degenerate diffusion coefficient. *Numer. Math.*, 138(1):219–239, 2018. doi: 10.1007/s00211-017-0903-9.
- [50] Y. Liu and G. Pagès. Monotone convex order for the McKean–Vlasov processes. *Stochastic Process. Appl.*, 152:312–338, 2022. doi: 10.1016/j.spa.2022.06.003.
- [51] X. Mao and L. Szpruch. Strong convergence and stability of implicit numerical methods for stochastic differential equations with non-globally Lipschitz continuous coefficients. *J. Comput. Appl. Math.*, 238:14–28, 2013. doi: 10.1016/j.cam.2012.08.015.
- [52] X. Mao and L. Szpruch. Strong convergence rates for backward Euler–Maruyama method for non-linear dissipative-type stochastic differential equations with super-linear diffusion coefficients. *Stochastics*, 85(1):144–171, 2013. doi: 10.1080/17442508.2011.651213.
- [53] G. N. Milstein. Approximate integration of stochastic differential equations. *Theory Probab. Appl.*, 19(3):557–562, 1975. doi: 10.1137/1119062.
- [54] 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. doi: 10.1137/S0036142901395588.
- [55] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l’Académie Royale des Sciences de Paris*, pages 666–704, 1781.
- [56] T. Müller-Gronbach and L. Yaroslavl’tseva. On the performance of the Euler–Maruyama scheme for SDEs with discontinuous drift coefficient. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(2):1162–1178, 2020. doi: 10.1214/19-AIHP997.
- [57] T. Müller-Gronbach and L. Yaroslavl’tseva. A strong order 3/4 method for SDEs with discontinuous drift coefficient. *IMA J. Numer. Anal.*, 42(1):229–259, 2022. doi: 10.1093/imanum/draa078.
- [58] T. Müller-Gronbach and L. Yaroslavl’tseva. Sharp lower error bounds for strong approximation of SDEs with discontinuous drift coefficient by coupling of noise. *Ann. Appl. Probab.*, 33(2):1102–1135, 2023. doi: 10.1214/22-aap1837.
- [59] T. Müller-Gronbach, S. Sabanis, and L. Yaroslavl’tseva. Existence, uniqueness and approximation of solutions of SDEs with superlinear coefficients in the presence of discontinuities of the drift coefficient. *Commun. Math. Sci.*, 23(2):389–429, 2025. doi: 10.4310/cms.241217213438.
- [60] A. Neuenkirch and M. Szölgényi. The Euler–Maruyama scheme for SDEs with irregular drift: convergence rates via reduction to a quadrature problem. *IMA J. Numer. Anal.*, 41(2):1164–1196, 2021. doi: 10.1093/imanum/draa007.
- [61] A. Neuenkirch, M. Szölgényi, and L. Szpruch. An adaptive Euler–Maruyama scheme for stochastic differential equations with discontinuous drift and its convergence analysis. *SIAM J. Numer. Anal.*, 57(1):378–403, 2019. doi: 10.1137/18M1170017.
- [62] H.-L. Ngo and D. Taguchi. On the Euler–Maruyama approximation for one-dimensional stochastic differential equations with irregular coefficients. *IMA J. Numer. Anal.*, 37(4):1864–1883, 2017. doi: 10.1093/imanum/drw058.

- [63] H.-L. Ngo and D. Taguchi. Strong convergence for the Euler–Maruyama approximation of stochastic differential equations with discontinuous coefficients. *Statist. Probab. Lett.*, 125: 55–63, 2017. doi: 10.1016/j.spl.2017.01.027.
- [64] H.-L. Ngo and D. Taguchi. On the Euler–Maruyama scheme for SDEs with bounded variation and Hölder continuous coefficients. *Math. Comput. Simulation*, 161:102–112, 2019. doi: 10.1016/j.matcom.2019.01.012.
- [65] G. Peyré and M. Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/22000000073.
- [66] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM J. Optim.*, 22(1):1–23, 2012. doi: 10.1137/110825054.
- [67] 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.
- [68] A. Pichler and M. Weinhardt. The nested Sinkhorn divergence to learn the nested distance. *Comput. Manag. Sci.*, 19:269–293, 2021. doi: 10.1007/s10287-021-00415-7.
- [69] P. Przybyłowicz and M. Szölgényi. Existence, uniqueness, and approximation of solutions of jump-diffusion SDEs with discontinuous drift. *Appl. Math. Comput.*, 403:126191, 2021. doi: 10.1016/j.amc.2021.126191.
- [70] P. Przybyłowicz, M. Szölgényi, and F. Xu. Existence and uniqueness of solutions of SDEs with discontinuous drift and finite activity jumps. *Statist. Probab. Lett.*, 174(109072):109072, 2021. doi: 10.1016/j.spl.2021.109072.
- [71] P. Przybyłowicz, V. Schwarz, and M. Szölgényi. A higher order approximation method for jump-diffusion SDEs with discontinuous drift coefficient. *J. Math. Anal. Appl.*, 538(1):128319, 2024. doi: 10.1016/j.jmaa.2024.128319.
- [72] P. Przybyłowicz, V. Schwarz, and M. Szölgényi. Lower error bounds and optimality of approximation for jump-diffusion SDEs with discontinuous drift. *BIT*, 64(4):35, 2024. doi: 10.1007/s10543-024-01036-7.
- [73] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, third edition, 1999. ISBN 3-540-64325-7. doi: 10.1007/978-3-662-06400-9.
- [74] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Statistiscs*, 23(3): 470–472, 1952. doi: 10.1214/aoms/1177729394.
- [75] L. Rüschendorf. The Wasserstein distance and approximation theorems. *Z. Wahrsch. Verw. Gebiete*, 70(1):117–129, 1985. doi: 10.1007/BF00532240.
- [76] S. Sabanis. A note on tamed Euler approximations. *Electron. Commun. Probab.*, 18:1–10, 2013. doi: 10.1214/ECP.v18-2824.
- [77] F. Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser, Cham, 2015. ISBN 978-3-319-20827-5; 978-3-319-20828-2. doi: 10.1007/978-3-319-20828-2.
- [78] A. V. Skorokhod. *Studies in the Theory of Random Processes*. Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, MA, 1965.
- [79] K. Spendier and M. Szölgényi. Convergence of the tamed-Euler–Maruyama method for SDEs with discontinuous and polynomially growing drift. In A. Hinrichs, P. Kritzer, and F. Pillichshammer, editors, *Monte Carlo and Quasi-Monte Carlo Methods. MCQMC 2022*, volume 460 of *Springer Proc. Math. Stat.*, pages 601–623. Springer International Publishing, 2024. doi: 10.1007/978-3-031-59762-6\_30.



- [80] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-28998-2; 3-540-28998-4. doi: 10.1007/3-540-28999-2. Reprint of the 1997 edition (Grundlehren der mathematischen Wissenschaft, Vol. 233).
- [81] M. Szölgyenyi. Stochastic differential equations with irregular coefficients: mind the gap! *Internationale Mathematische Nachrichten*, 246:43–56, 2021. doi: 10.48550/arXiv.2104.11505.
- [82] D. Talay. Résolution trajectorielle et analyse numérique des équations différentielles stochastiques. *Stochastics*, 9(4):275–306, 1983. doi: 10.1080/17442508308833257.
- [83] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. ISBN 0-8218-3312-X. doi: 10.1090/gsm/058.
- [84] C. Villani. *Optimal Transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009. doi: 10.1007/978-3-540-71050-9.
- [85] S. Watanabe and T. Yamada. On the uniqueness of solutions of stochastic differential equations II. *J. Math. Kyoto Univ.*, 11(3):553–563, 1971. doi: 10.1215/kjm/1250523620.
- [86] T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971. doi: 10.1215/kjm/1250523691.
- [87] L. Yaroslavtseva. An adaptive strong order 1 method for SDEs with discontinuous drift coefficient. *J. Math. Anal. Appl.*, 513(2):126180, 2022. doi: 10.1016/j.jmaa.2022.126180.
- [88] A. K. Zvonkin. A transformation of the phase space of a diffusion process that removes the drift. *Math. USSR-Sb.*, 22(1):129–149, 1974. doi: 10.1070/SM1974v022n01ABEH001689.