

# CAUSAL TRANSPORT ON PATH SPACE

BY RAMA CONT<sup>1,a</sup>, AND FANG RUI LIM<sup>2,b</sup>

<sup>1</sup>*Mathematical Institute, University of Oxford*, <sup>a</sup>[Rama.Cont@maths.ox.ac.uk](mailto:Rama.Cont@maths.ox.ac.uk)

<sup>2</sup>*Mathematical Institute, University of Oxford*, <sup>b</sup>[Fang.Lim@maths.ox.ac.uk](mailto:Fang.Lim@maths.ox.ac.uk)

We study properties of causal couplings for probability measures on the space of continuous functions. We first provide a characterization of bicausal couplings between weak solutions of stochastic differential equations. We then provide a complete description of all such bicausal Monge couplings. In particular, we show that bicausal Monge couplings of  $d$ -dimensional Wiener measures are induced by stochastic integrals of rotation-valued integrands. As an application, we give necessary and sufficient conditions for bicausal couplings to be induced by Monge maps and show that such bicausal Monge transports are dense in the set of bicausal couplings between laws of SDEs.

## CONTENTS

1	Optimal transport and its causal counterpart . . . . .	2
1.1	Contributions . . . . .	3
2	Causal couplings on path space . . . . .	4
2.1	Notations . . . . .	4
2.2	Definitions . . . . .	5
2.3	Relation with the H-hypothesis . . . . .	6
3	Causal Transport between weak solutions of stochastic differential equations . . . . .	7
3.1	Bicausal couplings of Wiener measures . . . . .	7
3.2	Bicausal couplings between solutions of SDEs . . . . .	7
3.3	Characterisation of bicausal Monge transports . . . . .	9
3.4	Proof of Theorem 3.2 and 3.3 . . . . .	12
3.4.1	Martingale Representation Theorem . . . . .	12
3.4.2	Technical proofs . . . . .	13
4	Applications . . . . .	15
4.1	Bicausal couplings induced by Monge maps . . . . .	15
4.2	Denseness of bicausal Monge maps among bicausal couplings . . . . .	17
4.3	Equivalence of bicausal Monge and Kantorovich optimal transport . . . . .	19
4.4	Examples . . . . .	20
	Acknowledgments . . . . .	22
	References . . . . .	22

---

*MSC2020 subject classifications:* Primary 49Q22, 60H10; secondary 28C20.

*Keywords and phrases:* Optimal transport, Stochastic differential equations, Wiener space.

**1. Optimal transport and its causal counterpart.** The theory of optimal transport of probability measures and the associated Wasserstein metric has been the focus of a large body of research [2, 31, 32, 37] with rich ramifications across various branches of mathematics and applications. In the context of stochastic processes, where one deals with probability measures on *filtered* spaces, it is natural to incorporate a compatibility condition with the flow of information, leading to the concept of *causal transport* [28], defined as a transport which makes use, at any time  $t$ , only the information available at  $t$ . As shown by Lassalle [28], when the underlying space is the space of trajectories of a stochastic process, the concept of causal transport has interesting connections with stochastic analysis.

A natural example of causal transport is the one induced by a *non-anticipative* map. Denote by  $\mathcal{W}^d := C([0, 1], \mathbb{R}^d)$  the space of continuous  $\mathbb{R}^d$ -valued functions on  $[0, 1]$ , with its canonical filtration  $(\mathcal{F}_t)_{t \in [0, 1]}$ . A map  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  may be naturally interpreted as a  $\mathbb{R}^d$ -valued functional  $F : [0, 1] \times \mathcal{W}^d \mapsto F(\omega, t) := T(\omega)(t) \in \mathbb{R}^d$ . Such a functional is said to be causal or non-anticipative [14, 21] if  $F(\omega, t)$  depends on the path  $\omega$  only through its values on  $[0, t]$ :

$$F(\omega, t) = F(\omega(\cdot \wedge t), t), \text{ where } s \wedge t := \min\{s, t\}.$$

The causality of  $F$  is equivalent to  $T^{-1}(\mathcal{F}_t) \subseteq \mathcal{F}_t$  for all  $t \in [0, 1]$ . Any probability measure  $\eta$  on  $\mathcal{W}^d$ , is thus transported in a causal manner, i.e. non-anticipative with respect the filtration, to an image measure  $\nu$  on  $\mathcal{W}^d$ . The map  $T$  is then an example of causal (Monge) coupling on  $\mathcal{W}^d \times \mathcal{W}^d$  between  $\eta$  and  $\nu$ .

An equivalent manner to state that  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  is causal is to require that

$$\forall t \in [0, 1], \quad \forall B \in \mathcal{F}_t, \quad \omega \mapsto \delta_{T(\omega)}(B) \text{ is } \mathcal{F}_t\text{-measurable,}$$

where  $\delta_{T(\omega)}$  is the point mass at  $T(\omega) \in \mathcal{W}^d$ . Denoting by  $X$  the canonical process on  $\mathcal{W}^d$ ,  $\delta_{T(\omega)}$  represents in fact the ‘conditional distribution’ of  $T(X)$  given  $X = \omega$ :

$$\delta_{T(\omega)} = \mathcal{L}^\eta(T(X)|X = \omega)$$

This suggests the following definition of causality for an *arbitrary* coupling: given a probability measure  $\pi$  on  $\mathcal{W}^d \times \mathcal{W}^d$  and writing  $X$  and  $Y$  as the first and second marginal processes, one replaces the above condition with the requirement that

$$(1) \quad \forall t \in [0, 1], \quad \forall B \in \mathcal{F}_t, \quad \omega \mapsto \mathcal{L}^\pi(Y|X = \omega)(B) \text{ is } \mathcal{F}_t\text{-measurable.}$$

There are some technical subtleties to take care of since the conditional distribution is only defined up to null sets, and there is a choice to be made between using the canonical filtration or its right-continuous version. We will give a precise definition below (Definition 2.1).

Since the notion of causality is not symmetric, the corresponding optimal transport problem does not yield a (symmetric) notion of distance. A remedy is to consider *bicausal* couplings, which satisfy (1) and its symmetric counterpart

$$\forall t \in [0, 1], \quad \forall B \in \mathcal{F}_t, \quad \omega \mapsto \mathcal{L}^\pi(X|Y = \omega)(B) \text{ is } \mathcal{F}_t\text{-measurable.}$$

Causal optimal transport has been studied in detail in the discrete time case [4–6, 8, 16, 17, 35, 38], where the sequential character of processes allows to study the properties over a single time period and extend them by induction. However, the continuous time case reveals much more structure, with links to stochastic analysis and functional inequalities [22, 28].

As noted by Lassalle [28], the Ito map associated with the unique strong solution of a stochastic differential equation (SDE) is a natural example of causal Monge transport from the Wiener measure to the law of the solution, while the concept of weak solution provides an example of causal coupling *à la* Kantorovich between the Wiener measure and the law of the solution. Optimal transport on the Wiener space has been investigated in [19, 20]. Causal

transport on Wiener space has been investigated in the context of filtration enlargements [1] and its relation to Talagrand's inequality [22, 28]. Backhoff et al. [5] show the relevance of causal Wasserstein distances induced by such causal transports for superhedging problems.

The present work is a detailed study of the structure of (bi-)causal couplings on path space, focusing on the case where the measures involved have a ‘differential’ structure i.e. are (weak) solutions of stochastic differential equations (SDEs). In addition to confirming that many of the results obtained in the discrete-time setting remain true for (bi-)causal transports on path space, we show that, when the measures are weak solutions of SDEs, bicausal transports between these measures inherit this differential structure and may be represented in terms of stochastic integrals or, more generally, semimartingales. In contrast to the discrete time case, we show that causal transport between laws of SDEs has a *rigid* structure: essentially, bicausal couplings of SDEs are given by bicausal coupling of the driving Brownian motions (Theorem 3.2), for which we give a complete description (Theorem 3.1).

Related questions have been studied in [7, 33], where (bi-)causal transport problems for solutions to *scalar* SDEs are investigated, as well as in [10] using a stochastic control approach.

**1.1. Contributions.** We study properties of (bi)causal couplings for probability measures on the space of continuous functions defined. We first prove a characterization of bicausal couplings between weak solutions of SDEs and then investigate the case where the bicausal couplings are induced by a Monge transport. Denoting by the  $\Pi_c(\eta, \nu)$  (resp.  $\Pi_{bc}(\eta, \nu)$ ) the set of all causal (resp. bicausal) couplings, for any map  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  which transports  $\eta$  to  $\nu$ , i.e.  $T_{\#}\eta = \nu$ , we have

$$(\text{Id}, T)_{\#}\eta \in \Pi_c(\eta, \nu) \iff T \text{ is non-anticipative.}$$

We show in Lemma 3.4 that, when  $\eta$  and  $\nu$  are unique weak solutions of SDEs,

$$(\text{Id}, T)_{\#}\eta \in \Pi_{bc}(\eta, \nu) \iff (T_t)_{t \in [0,1]} \text{ is a } \left( \eta, (\mathcal{F}_t)_{t \in [0,1]} \right)\text{-semimartingale,}$$

where  $T_t := T(\omega)(t)$ . Using a martingale representation result of Üstünel [36], we are able to completely characterise bicausal Monge transports between such measures. We show that such maps inherit the differential structure of  $\eta, \nu$  in the sense that  $T$  satisfies an SDE (9) (Theorem 3.3). In particular, we show bicausal Monge transports of  $d$ -dimensional Wiener measures are induced by stochastic integrals of adapted  $\mathcal{O}^d$ -valued integrands, where  $\mathcal{O}^d$  is the orthogonal group of  $d \times d$  matrices.

Given an SDE with unique strong solution, the Ito map provides an example of causal transport from the Wiener measure  $\mathbb{W}^d$  to the law  $\nu$  of the solution on path space. Theorem 3.3 characterizes *all* bicausal Monge maps from  $\mathbb{W}^d$  to  $\nu$  as the composition of the Ito map with integration against an adapted  $\mathcal{O}^d$ -valued process.

Using the aforementioned characterization of bicausal Monge transports, we give simple conditions for their existence (Corollaries 3.2, 3.4, 3.5). We show that any bicausal coupling between Wiener measures has a differential representation i.e. may be represented by a stochastic differential equation (Proposition 4.1). Through this SDE, we can establish necessary and sufficient conditions for bicausal couplings between weak solutions of SDEs to be induced by a Monge map (Proposition 4.1 and Corollary 4.1).

Beiglböck et al. [9] showed that in discrete time, the set  $\mathcal{T}_c(\mathbb{P}, \mathbb{P}')$  of causal Monge couplings is dense in the set of all causal couplings. We show that similar density properties hold for bicausal transports between laws of SDEs with strong, pathwise unique solutions and continuous (possibly path-dependent) drift and invertible diffusion coefficients (Proposition 4.4 and 4.5). The proofs make use of M. Émery's results on ‘almost Brownian filtrations’ [18], whose potential role in the study of causal transports was already noticed in [9, Section 5.4].

A consequence of our results is that, for many cost functions the bicausal Monge problem between weak solutions of SDEs and its Kantorovich relaxation lead to the same transport cost (Corollary 4.2).

*Outline.* Section 2.1 establish notations and defines causal couplings and maps. Section 3 contains the main results on the characterization of bicausal couplings and Monge maps between laws of solutions of SDEs, and some results regarding the existence of such transports. Section 4 considers some applications of these results.

## 2. Causal couplings on path space.

2.1. *Notations.* Let  $d \geq 1$ . Given  $v, w \in \mathbb{R}^d$ , we denote the standard Euclidean inner product as  $(v, w)_{\mathbb{R}^d}$  and norm  $\|v\|_{\mathbb{R}^d} := \sqrt{(v, v)_{\mathbb{R}^d}}$ . The space of  $d \times d$  matrices is written as  $\mathbb{R}^{d \times d}$ .  $\text{Id}_d \in \mathbb{R}^{d \times d}$  is the identity matrix. Given  $A \in \mathbb{R}^{d \times d}$ ,  $A^*$  denotes the transpose of  $A$ ,  $A^{-1}$  denotes the inverse of  $A$ ,  $\text{tr } A$  denotes its trace,  $\det A$  denotes its determinant,  $\ker A$  denotes its kernel and  $A^\dagger$  denotes its Moore-Penrose pseudoinverse [23], that is,  $A^\dagger$  is the unique matrix satisfying *i*)  $AA^\dagger A = A$ , *ii*)  $A^\dagger AA^\dagger = A^\dagger$ , *iii*)  $AA^\dagger$  and  $A^\dagger A$  are symmetric. We denote  $\mathcal{O}^d$  the set of orthogonal matrices,  $\mathcal{S}_+^d$  the set of symmetric matrices,  $\mathcal{S}_+^d$  the subset of symmetric positive semi-definite matrices and  $\mathcal{C}^d \subseteq \mathcal{S}^d$  - the subset of matrices defined as in [18]:

$$\mathcal{C}^d = \{C \in \mathcal{S}^d, \begin{pmatrix} \text{Id}_d & C \\ C^* & \text{Id}_d \end{pmatrix} \in \mathcal{S}_+^{2d}\} = \{C \in \mathcal{S}^d, \|C\|_{op} \leq 1\} = \text{conv}(\mathcal{O}^d),$$

where  $\|\cdot\|_{op}$  designates the operator norm and  $\text{conv}$  designates the convex hull. The second and third equalities are shown in [18, Proposition 1].

Denote  $\mathcal{W}^d := C([0, 1], \mathbb{R}^d)$ , the space of continuous  $\mathbb{R}^d$ -valued functions on  $[0, 1]$ . Equipped with the topology  $\tau_{\mathcal{W}^d}$  induced by the supremum norm,  $\|\cdot\|_\infty$ ,  $\mathcal{W}^d$  is a Polish space with Borel  $\sigma$ -algebra  $\mathcal{F}_1 := \sigma(\tau_{\mathcal{W}^d})$ . The canonical filtration is denoted  $(\mathcal{F}_t)_{t \in [0, 1]}$  and given by  $\mathcal{F}_t = \sigma(f_t) \subseteq \mathcal{F}_1$ , where  $f_t : \mathcal{W}^d \rightarrow (\mathcal{W}^d, \mathcal{F}_1)$ ,  $f_t(\omega) := \omega(\cdot \wedge t)$  for  $t \in [0, 1]$  and set  $(\mathcal{H}_t)_{t \in [0, 1]}$  to be the right-continuous version of the canonical filtration, i.e.

$$\mathcal{H}_t = \mathcal{F}_{t+} = \bigcap_{\epsilon > 0} \mathcal{F}_{(t+\epsilon) \wedge 1} \subseteq \mathcal{F}_1.$$

The canonical process on  $\mathcal{W}^d$  is defined as  $X : \mathcal{W}^d \times [0, 1] \rightarrow \mathbb{R}^d$ ,  $X(\omega, t) = X_t(\omega) := \omega(t)$ . When we write  $X_\cdot$ , we mean the identity map, i.e.  $X_\cdot : \mathcal{W}^d \rightarrow \mathcal{W}^d$ ,  $X_\cdot(\omega) = \omega$ . On the product  $\mathcal{W}^d \times \mathcal{W}^d$ , without risk of confusion,  $X$  will also denote the first component  $X(\omega, \omega') := \omega$  while  $Y$  denotes the second component  $Y(\omega, \omega') := \omega'$ .

Given measurable spaces  $(Z_i, \mathcal{Z}_i)$ , where  $i = 1, 2$  and  $\mathcal{Z}_i$  is some  $\sigma$ -algebra,  $\mathcal{Z}_1 \otimes \mathcal{Z}_2$  is the product  $\sigma$ -algebra on  $Z_1 \times Z_2$ .  $\mathcal{P}(Z_i)$  denotes the space of probability measures on  $Z_i$ . Given  $\eta \in \mathcal{P}(Z_1)$ , the  $\eta$ -completion of  $\mathcal{Z}_1$  is written as  $\mathcal{Z}_1^\eta$ . Given a  $\mathcal{Z}_1/\mathcal{Z}_2$ -measurable function  $f : Z_1 \rightarrow Z_2$ ,  $f_\# \eta \in \mathcal{P}(Z_2)$  is the pushforward of  $\eta$  under  $f$ , i.e.  $f_\# \eta(A) := \eta(f^{-1}(A))$  for  $A \in \mathcal{Z}_2$ . Furthermore, for  $p \geq 1$ , we write  $L^p(\eta, \mathcal{Z}_1)$  to be the set of all  $\mathcal{Z}_1$ -measurable functions  $f : Z_1 \rightarrow \mathbb{R}$  (identified up to  $\eta$ -a.s. equivalence) satisfying

$$\mathbb{E}^\eta(|f|^p) := \int_{Z_1} |f(\omega)|^p \eta(d\omega) < \infty.$$

A filtration on  $(Z_1, \mathcal{Z}_1)$  is a family of  $\sigma$ -algebras  $(\mathcal{N}_t)_{t \in [0, 1]}$ , such that  $\mathcal{N}_s \subseteq \mathcal{N}_t \subseteq \mathcal{Z}_1$  for all  $0 \leq s \leq t \leq 1$ . We call the quadruple  $(Z_1, \mathcal{Z}_1, (\mathcal{N}_t)_{t \in [0, 1]}, \eta)$  a stochastic basis if

$(\mathcal{N}_t)_{t \in [0,1]}$  is a  $\eta$ -complete, right-continuous filtration. The space of all square-integrable  $(\eta, (\mathcal{N}_t)_{t \in [0,1]})$ -martingales is denoted

$$\mathcal{M}^2\left(\eta, (\mathcal{N}_t)_{t \in [0,1]}\right) := \left\{ M = (M_t)_{t \in [0,1]} \text{ is an } (\eta, (\mathcal{N}_t)_{t \in [0,1]})\text{-martingale, } \sup_{t \in [0,1]} \mathbb{E}^\eta |M_t|^2 < \infty \right\}.$$

Finally, if  $\eta_i \in \mathcal{P}(\mathcal{Z}_i)$  for  $i = 1, 2$ ,  $\eta_1 \otimes \eta_2$  denotes the product coupling  $\eta_1 \otimes \eta_2(A \times B) := \eta_1(A)\eta_2(B)$  for  $A \in \mathcal{Z}_1, B \in \mathcal{Z}_2$ .

**2.2. Definitions.** Let  $\eta, \nu \in \mathcal{P}(\mathcal{W}^d)$  be probability measures.  $\pi \in \mathcal{P}(\mathcal{W}^d \times \mathcal{W}^d)$  is a coupling of  $(\eta, \nu)$  if

$$X_{\#}\pi = \eta \text{ and } Y_{\#}\pi = \nu.$$

The set of all such couplings is denoted  $\Pi(\eta, \nu)$ . Since  $\mathcal{W}^d \times \mathcal{W}^d$  equipped with the product topology is a Polish space, for any  $\pi \in \Pi(\eta, \nu)$  there exists an  $\eta$ -a.s. unique probability kernel  $\Theta_\pi : \mathcal{W}^d \times \mathcal{F}_1 \rightarrow [0, 1]$ ,  $(\omega, A) \mapsto \Theta_\pi^\omega(A)$  which disintegrates  $\pi$  with respect to  $\eta$  [26]:

$$\pi(A \times B) = \int_{\mathcal{W}^d} \mathbb{1}_A(\omega) \Theta_\pi^\omega(B) \eta(d\omega).$$

We call  $\Theta_\pi$  the disintegration of  $\pi$  with respect to  $\eta$ . Note that  $\Theta_\pi^\omega$  is a version of the conditional distribution of  $Y$  given  $X = \omega$ ,  $\mathcal{L}^\pi(Y|X = \omega)$ . We recall the definitions of causal transports and causal couplings [1, 28]:

**DEFINITION 2.1** (Causal couplings and causal Monge maps).

Let  $\mathcal{H}_t$  be the right-continuous version of the canonical filtration, defined as above.

1.  $\pi \in \Pi(\eta, \nu)$  is a causal coupling between  $\eta$  and  $\nu$  if

$$(2) \quad \text{for all } t \in [0, 1] \text{ and } B \in \mathcal{H}_t, \omega \mapsto \Theta_\pi^\omega(B) \text{ is } \mathcal{H}_t^\eta\text{-measurable.}$$

We denote by  $\Pi_c(\eta, \nu)$  the set of causal couplings between  $\eta$  and  $\nu$ ;

2. A measurable map  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  is a causal Monge transport between  $\eta$  and  $\nu$  if the image of  $\eta$  under  $T$  is  $\nu$  and the induced transport plan  $\pi_T := (X, T)_{\#}\eta$  is a causal transport:  $\pi_T \in \Pi_c(\eta, \nu)$ .
3.  $\pi \in \Pi(\eta, \nu)$  is a bicausal coupling between  $\eta$  and  $\nu$  if  $\pi \in \Pi(\eta, \nu)$  and  $R_{\#}\pi \in \Pi(\nu, \eta)$  where  $R$  is the exchange of coordinates:

$$R : \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathcal{W}^d \times \mathcal{W}^d$$

$$(\omega', \omega) \mapsto R(\omega, \omega') := (\omega', \omega)$$

The set of all bicausal couplings between  $\eta$  and  $\nu$  is denoted  $\Pi_{bc}(\eta, \nu)$ .

4.  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  is said to be a bicausal Monge map between  $\eta$  and  $\nu$  if  $\pi_T \in \Pi_{bc}(\eta, \nu)$ .  $\mathcal{T}_{bc}(\eta, \nu)$  denotes the set of all  $\pi \in \Pi(\eta, \nu)$  such that  $\pi = \pi_T$  for some bicausal Monge map  $T$ . Abusing notation, we will write  $T \in \mathcal{T}_{bc}(\eta, \nu)$  if  $\pi_T$  is a bicausal coupling.
5.  $\mathcal{T}_{bcb}(\eta, \nu)$  denotes the set of all couplings  $\pi \in \Pi(\eta, \nu)$  such that  $\pi = \pi_T$  for some a.s. bijective<sup>1</sup> bicausal Monge map  $T$ .

**REMARK 1.**

1. Comparing Definition 2.1 and (1), we observe two differences:

---

<sup>1</sup> $T$  is a.s. bijective if there exists  $\hat{T} : \mathcal{W}^d \rightarrow \mathcal{W}^d$  such that  $\eta$ -a.s.  $\hat{T} \circ T = X$ . and  $\nu$ -a.s.  $T \circ \hat{T} = X$ .

- we ask for measurability with respect to the completed  $\sigma$ -algebra  $\mathcal{H}_t^\eta$ , instead of just  $\mathcal{H}_t$ . This is because the disintegration  $\Theta_\pi$  is only  $\eta$ -a.s. unique;
  - Second, we have used the right-continuous filtration  $(\mathcal{H}_t)_{t \in [0,1]}$ , instead of  $(\mathcal{F}_t)_{t \in [0,1]}$ . This is motivated by the fact  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$  is a complete, right-continuous filtration for which many familiar results from stochastic analysis hold, e.g. martingales have càdlàg modifications.
2. To verify (2), it is sufficient to check measurability of  $\Theta_\pi(B)$  for all  $B \in \mathcal{F}_t$ . This observation implies that:
- if  $\mathcal{H}_t^\eta = \mathcal{F}_t^\eta$  for all  $t \in [0,1]$ , then it does not matter if we replace  $\mathcal{H}_t$  with  $\mathcal{F}_t$  in (2). Indeed, the set of causal couplings would not change. This was observed in [7, Remark 2.4] for the case when  $\eta$  is the law of a strongly Markov process. We also mention [29, Theorem 5.19] which shows that  $\mathcal{H}_t^\eta = \mathcal{F}_t^\eta$  when  $\eta$  is the law of an SDE with Lipschitz drift and uniformly elliptic diffusion coefficient; and,
  - $\Pi_c(\eta, \nu)$  and hence,  $\Pi_{bc}(\eta, \nu)$ , are compact for the topology of weak convergence by [28, Theorem 1, Remark 5].

**2.3. Relation with the H-hypothesis.** Brémaud and Yor [11] introduced the *H-hypothesis* as a property of filtrations useful in filtering theory. Two filtrations  $(\mathcal{M}_t)_{t \in [0,1]}$  and  $(\mathcal{N}_t)_{t \in [0,1]}$  on a probability space  $(\Omega, \mathcal{N}, \mathbb{P})$  such that  $\mathcal{M}_t \subseteq \mathcal{N}_t \subseteq \mathcal{N}$  are said to satisfy the H-hypothesis if every square integrable  $(\mathbb{P}, (\mathcal{M}_t))$ -martingale is also an  $(\mathbb{P}, (\mathcal{N}_t))$ -martingale:<sup>2</sup>

$$\mathcal{M}^2(\mathbb{P}, (\mathcal{M}_t)_{t \in [0,1]}) \subseteq \mathcal{M}^2(\mathbb{P}, (\mathcal{N}_t)_{t \in [0,1]}).$$

It was noticed in [1] that causality of  $\pi \in \Pi(\eta, \nu)$  is equivalent to the H-hypothesis for the filtrations  $(\mathcal{H}_t \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]}$  and  $(\mathcal{H}_t^\eta \otimes \mathcal{H}_t)_{t \in [0,1]}$  on  $(\mathcal{W}^d \times \mathcal{W}^d, \mathcal{F}_1 \otimes \mathcal{F}_1, \pi)$ . Since this is a key property that we will use repeatedly, we state it as a theorem below. Given a probability measure  $\pi$  on the product space  $(\mathcal{W}^d \times \mathcal{W}^d, \mathcal{F}_1 \otimes \mathcal{F}_1)$  we denote by  $\mathcal{P}_t^\pi$  the right-continuous  $\pi$ -completion of the canonical product filtration:

$$(3) \quad \mathcal{P}_t^\pi := \bigcap_{\epsilon > 0} (\mathcal{H}_{t+\epsilon} \otimes \mathcal{H}_{t+\epsilon})^\pi$$

**THEOREM 2.2.** *A coupling  $\pi \in \Pi(\eta, \nu)$  is a causal coupling if and only if the filtrations  $(\mathcal{H}_t^\eta \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]}$  and  $(\mathcal{P}_t^\pi)_{t \in [0,1]}$  on  $(\mathcal{W}^d \times \mathcal{W}^d, \mathcal{F}_1 \otimes \mathcal{F}_1, \pi)$  satisfy the H-hypothesis:*

$$(4) \quad \mathcal{M}^2(\pi, (\mathcal{H}_t^\eta \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]}) \subseteq \mathcal{M}^2(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]}).$$

**PROOF.** From [11, Theorem 6] or [1, Remark 2.3],  $\pi \in \Pi_c(\eta, \nu)$  if and only if the H-hypothesis holds between filtrations  $(\mathcal{H}_t \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]}$  and  $(\mathcal{H}_t^\eta \otimes \mathcal{H}_t)_{t \in [0,1]}$  on  $(\mathcal{W}^d \times \mathcal{W}^d, \mathcal{F}_1 \otimes \mathcal{F}_1, \pi)$ . Hence, necessity is clear since  $\mathcal{H}_t^\eta \otimes \mathcal{H}_t \subseteq \bigcap_{\epsilon > 0} (\mathcal{H}_{t+\epsilon} \otimes \mathcal{H}_{t+\epsilon})^\pi$ .

To prove sufficiency, let  $(M_t)_{t \in [0,1]}$  be a  $(\pi, (\mathcal{H}_t^\eta \otimes \{\emptyset, \mathcal{W}^d\}))$ -martingale, which we can assume without loss of generality to be càdlàg. It is straightforward to verify that  $(M_t)_{t \in [0,1]}$  is a  $(\pi, (\mathcal{H}_t^\eta \otimes \mathcal{H}_t))$ -martingale by causality of  $\pi$  in light of [11, Theorem 6] or [1, Remark 2.3]. Since  $(M_t)_{t \in [0,1]}$  is càdlàg, it is a  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -martingale by [34, Lemma 67.10] as required.  $\square$

**REMARK 2.** Note that the filtrations in the above theorem are complete and right-continuous. In particular, martingales have càdlàg modifications and up to standard localization techniques, it is sufficient to consider bounded martingales for checking the H-hypothesis. One can obtain other characterisations of causality using [11, Theorem 3].

<sup>2</sup>Note that this is a “property” rather than a “hypothesis” but we will stick to the historical terminology introduced in [11].



### 3. Causal Transport between weak solutions of stochastic differential equations.

3.1. *Bicausal couplings of Wiener measures.* We first examine bicausal couplings of (d-dimensional) Wiener measures.

**THEOREM 3.1** (Bicausal couplings of Wiener measures).  *$\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  if and only if there exists a stochastic basis  $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0,1]}, \mathbb{P})$  and two  $d$ -dimensional  $(\mathcal{G}_t)_{t \in [0,1]}$ -Brownian motions  $B$  and  $\bar{B}$  such that  $\pi$  is the joint law of  $(B, \bar{B})$ . In particular, there exists a  $(\mathcal{G}_t)_{t \in [0,1]}$ -progressively measurable process  $\rho : [0, 1] \times \Omega \rightarrow \mathbb{C}^d$  such that*

$$\forall t \in [0, 1] \quad [B, \bar{B}]_t = \int_0^t \rho_s ds, \quad \pi\text{-a.s.}$$

**PROOF.** Assume  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ . By Theorem 2.2, the marginal processes  $X$  and  $Y$  are  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -martingales (and, hence Brownian motions), from which the desired result follows.

To see necessity, we first show that under  $\pi := (B, \bar{B})_\# \mathbb{P}$ ,  $X$  is a  $(\pi, (\mathcal{H}_t \otimes \mathcal{H}_t)_{t \in [0,1]})$ -martingale. Fix any  $0 \leq s < t \leq 1$ ,  $A, C \in \mathcal{H}_s$  and  $(\omega, \omega') \mapsto \mathbb{1}_{A \times C}(\omega, \omega')$ . Since  $B, \bar{B}$  are  $(\mathcal{G}_t)_{t \in [0,1]}$ -Brownian motions, they are  $(\mathbb{P}, (\mathcal{G}_t))$ -martingales and  $B^{-1}(\mathcal{H}_s), \bar{B}^{-1}(\mathcal{H}_s) \subseteq \mathcal{G}_s$ . Hence, we compute

$$\mathbb{E}^\pi((X_t - X_s)\mathbb{1}_{A \times C}) = \mathbb{E}^\mathbb{P}((B_t - B_s)\mathbb{1}_{A \times C}(B, \bar{B})) = 0,$$

which verifies the desired. By continuity of sample paths and [34, Lemma 67.10], we even have that  $X$  is a  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -martingale. By the martingale representation theorem of Brownian motion, every element in  $\mathcal{M}^2(\pi, (\mathcal{H}_t^{\mathbb{W}^d} \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]})$  can be written as a stochastic integral against  $X$ , from which we conclude (4) holds and  $\pi \in \Pi_c(\mathbb{W}^d, \mathbb{W}^d)$  by Theorem 2.2. A completely symmetric argument gives  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ .

The existence of  $\rho$  follows from the Kunita-Watanabe inequality. Indeed, the aforementioned inequality gives

$$|[B, \bar{B}]_t - [B, \bar{B}]_s| \leq ([B]_t - [B]_s)^{1/2}([\bar{B}]_t - [\bar{B}]_s)^{1/2} = |t - s|,$$

which shows that  $[B, \bar{B}]$  has finite total variation, which is absolutely continuous with respect to the Lebesgue measure.  $\square$

**REMARK 3.** Note that  $(B, \bar{B})$  is not necessarily an  $\mathbb{R}^{2d}$ -valued Brownian motion. As an example, given a Brownian motion  $B$ , one can set  $\bar{B}_t = \int_0^t (\mathbb{1}_{[0,1/2)}(s) + (\mathbb{1}_{\{B_{1/2} \geq 0\}} - \mathbb{1}_{\{B_{1/2} < 0\}})\mathbb{1}_{[1/2,1]}(s))dB_s$ . By Lévy's characterization of Brownian motion,  $\bar{B} = (\bar{B}_t)_{t \in [0,1]}$  is Brownian motion with respect to the filtration generated by  $B$ , and hence,  $\text{Law}(B, \bar{B}) \in \Pi_{bc}(\mathbb{W}^1, \mathbb{W}^1)$  by Theorem 3.1. But this joint law is not a Gaussian measure since  $B_1 - \bar{B}_1 = 2\mathbb{1}_{\{B_{1/2} < 0\}}(B_1 - B_{1/2})$ .

3.2. *Bicausal couplings between solutions of SDEs.* We now provide a characterisation of bicausal couplings and Monge transports between weak solutions of SDEs. We first present the main results, Theorem 3.2 and 3.3, and their corollaries and defer technical proofs to Section 3.4. Let

$$\mathcal{A}^{d,d'} := \left\{ \alpha : [0, 1] \times \mathcal{W}^d \rightarrow \mathbb{R}^{d \times d'} : \alpha \text{ is } (\mathcal{F}_t)_{t \in [0,1]} \text{-progressively measurable} \right\}$$

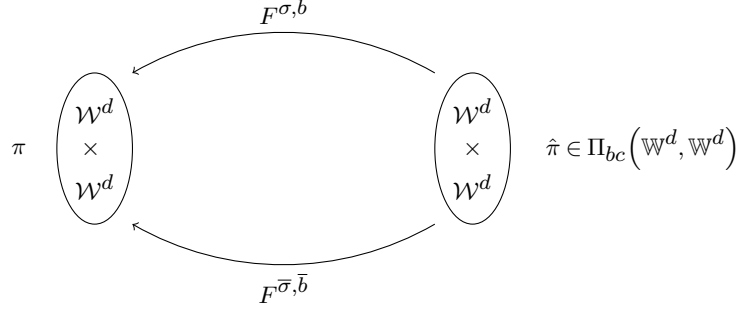


FIG 1. Bicausal couplings  $\pi \in \Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  between (strong) solutions of SDEs may be constructed by pushing forward bicausal Wiener couplings  $\hat{\pi} \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  with the Ito maps  $(F^{\sigma,b}, F^{\bar{\sigma},\bar{b}})$ .

for  $d, d' \geq 1$ . We consider the SDEs

$$(5) \quad dZ_t = b(t, Z)dt + \sigma(t, Z)dB_t, \quad Z_0 = z$$

$$(6) \quad d\bar{Z}_t = \bar{b}(t, \bar{Z})dt + \bar{\sigma}(t, \bar{Z})d\bar{B}_t, \quad \bar{Z}_0 = \bar{z}.$$

where  $\sigma, \bar{\sigma} \in \mathcal{A}^{d,d}$  and  $b, \bar{b} \in \mathcal{A}^{d,1}$  are such that

ASSUMPTION 1. There exists a unique weak solution  $\mu_z^{\sigma,b} \in \mathcal{P}(\mathcal{W}^d)$  (resp.  $\mu_{\bar{z}}^{\bar{\sigma},\bar{b}} \in \mathcal{P}(\mathcal{W}^d)$  for (5) (resp. for (6)).

The following theorem is a characterization of the set of bicausal couplings between the solutions of SDEs. It generalizes [7, Proposition 2.2] to the case where  $d > 1$  and the SDEs may not necessarily have strong solutions.

THEOREM 3.2.  $\pi \in \Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  if and only if there exists a stochastic basis  $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0,1]}, \mathbb{P})$ ,  $d$ -dimensional  $(\mathcal{G}_t)_{t \in [0,1]}$ -Brownian motions  $B$  and  $\bar{B}$  and  $(\mathcal{G}_t)_{t \in [0,1]}$ -adapted processes  $Z, \bar{Z}$  such that

- $(Z, B)$  and  $(\bar{Z}, \bar{B})$  satisfy (5) and (6) respectively, and
- $\pi$  is the joint law of  $(Z, \bar{Z})$ :  $\pi = (Z, \bar{Z})_{\#} \mathbb{P}$ .

In particular, the joint law of  $(B, \bar{B})$  under  $\mathbb{P}$  is a bicausal Wiener coupling.

In other words, all bicausal couplings may be obtained by a bicausal coupling of the Brownian motions in (5) and (6).

Consider the case of bicausal couplings between  $d$ -dimensional Wiener measures. Theorem 3.2 gives that  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  if and only if  $\pi$  is the joint law of two processes which are Brownian motions with respect to a common filtration. A consequence of this observation is that all bicausal couplings between SDEs with strong, pathwise unique solutions are obtained by taking the image of bicausal Wiener couplings in  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  under the respective Ito maps (Figure 1).

COROLLARY 3.1. Assume the SDEs (5) and (6) have strong solutions satisfying pathwise uniqueness, given by the Ito maps  $F^{\sigma,b} : (\mathcal{W}^d, \mathcal{F}_1) \rightarrow (\mathcal{W}^d, \mathcal{F}_1)$  (resp.  $F^{\bar{\sigma},\bar{b}}$ ). Then



all bicausal couplings  $\pi \in \Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  between the solutions of the SDEs may be obtained by pushing forward bicausal Wiener couplings  $\hat{\pi} \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  with the Ito maps  $(F^{\sigma,b}, F^{\bar{\sigma},\bar{b}})$ :

$$\Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}}) = \left\{ \left( F^{\sigma,b}, F^{\bar{\sigma},\bar{b}} \right)_{\#} \hat{\pi} : \hat{\pi} \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d) \right\},$$

where  $(F^{\sigma,b}, F^{\bar{\sigma},\bar{b}})_{\#} \hat{\pi}$  designates the image of the measure  $\hat{\pi}$  under the map

$$(X, Y) \in \mathcal{W}^d \times \mathcal{W}^d \mapsto (F^{\sigma,b}(X), F^{\bar{\sigma},\bar{b}}(Y)).$$

PROOF. The construction is shown in Figure 1. Note that the assumption of unique, strong solutions automatically implies weak uniqueness so that Assumption 1 holds and we can use Theorem 3.2, which yields the desired result.  $\square$

The preceding corollary allows us to deduce the properties of  $\Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  from the study of  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ , which we will use in Section 4.

**3.3. Characterisation of bicausal Monge transports.** For bicausal Monge maps, we first note that on  $(\mathcal{W}^d, \mathcal{F}_1, \mu_z^{\sigma,b})$ , the canonical process  $(X_t)_{t \in [0,1]}$  is a  $(\mu_z^{\sigma,b}, (\mathcal{H}_t))$ -semimartingale with decomposition

$$(7) \quad X_t = z + \int_0^t b(s, X) ds + M_t^{\mu_z^{\sigma,b}} \quad \mu_z^{\sigma,b}\text{-a.s. } t \in [0, 1],$$

where  $(M_t^{\mu_z^{\sigma,b}})$  is a  $(\mu_z^{\sigma,b}, (\mathcal{H}_t))$ -local martingale with quadratic variation

$$(8) \quad [M^{\mu_z^{\sigma,b}}, M^{\mu_z^{\sigma,b}}]_t = \int_0^t \sigma(s, X) \sigma^*(s, X) ds \quad \mu_z^{\sigma,b}\text{-a.s. } t \in [0, 1].$$

We then have the following characterization:<sup>3</sup>

**THEOREM 3.3 (Characterisation of bicausal Monge transports).** *Let  $\eta := \mu_z^{\sigma,b}$  and  $\nu := \mu_{\bar{z}}^{\bar{\sigma},\bar{b}}$ . For  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$ , define  $T_t(\omega) := T(\omega)(t) \in \mathbb{R}^d$  for  $t \in [0, 1]$ .  $T$  is a bicausal Monge map from  $\eta$  to  $\nu$  if and only if there exists some  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -progressively measurable process  $Q : [0, 1] \times \mathcal{W}^d \rightarrow \mathcal{O}^d$  such that  $(T_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta)_{t \in [0,1]})$ -semimartingale satisfying*

$$(9) \quad T_t = \bar{z} + \int_0^t \bar{b}(s, T) ds + \int_0^t \bar{\sigma}(s, T) Q_s \sigma^\dagger(s, X) dM_s^\eta \quad \eta\text{-a.s.}$$

and

$$(10) \quad \bar{\sigma}(t, T) Q_t = \bar{\sigma}(t, T) Q_t \sigma^\dagger(t, X) \sigma(t, X) \quad d\eta \otimes dt \quad - \text{ a.e.}$$

Since  $\sigma^\dagger(s, X) \sigma(s, X)$  is an orthogonal projection onto the orthogonal complement of  $\text{Ker } \sigma(s, X)$  (see [23, Section 5.5.4]),

$$(11) \quad (10) \iff Q_s \text{Ker}(\sigma(s, X)) \subseteq \text{Ker}(\bar{\sigma}(s, T)).$$

<sup>3</sup>We thank Y. Jiang for the sufficiency part of this theorem.

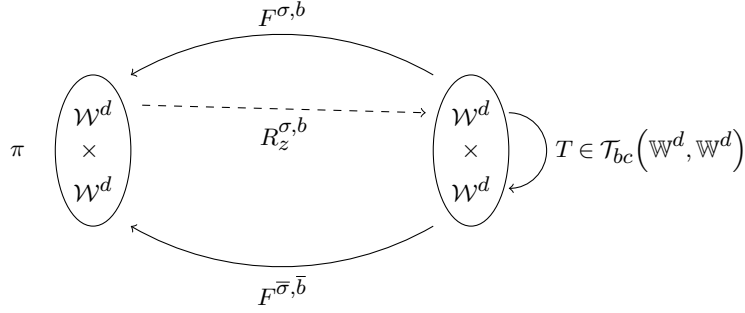


FIG 2. Any bicausal Monge transport  $\pi \in \mathcal{T}_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  can be constructed by pushing  $\mu_z^{\sigma,b}$  along  $R_z^{\sigma,b}$ , then  $T$  and finally,  $F^{\bar{\sigma},\bar{b}}$ .

PROOF OF (11). Let  $P_s := \sigma^\dagger(s, X)\sigma(s, X)$ . Then,

$$\bar{\sigma}(s, T)Q_s = \bar{\sigma}(s, T)Q_s P_s + \bar{\sigma}(s, T)Q_s(\text{Id}_d - P_s).$$

Hence, (10)  $\iff \bar{\sigma}(s, T)Q_s(\text{Id}_d - P_s) = 0$ . Since  $\text{Id}_d - P_s$  is the orthogonal projection onto  $\text{Ker } \sigma(s, X)$ , the statement follows.  $\square$

Equation (10) enables one to identify situations where bicausal Monge maps cannot exist:

COROLLARY 3.2. Under the assumptions of Theorem 3.3, if for all  $s \in [0, 1]$ ,

$$\max_{\omega \in \mathcal{W}^d} \dim(\text{Ker}(\bar{\sigma}(s, \omega))) < \min_{\omega \in \mathcal{W}^d} \dim(\text{Ker}(\sigma(s, \omega))),$$

there exists no bicausal Monge transport from  $\mu_z^{\sigma,b}$  to  $\mu_{\bar{z}}^{\bar{\sigma},\bar{b}}$ .

The bicausal Monge transports which leave the Wiener measure invariant are exactly (causal) local orthogonal transformations, which reflect the local gauge symmetry of the Wiener measure:

COROLLARY 3.3 (Bicausal Monge transports between Wiener measures). The set of all bicausal Monge transports from  $\mathbb{W}^d$  to  $\mathbb{W}^d$  is given by

$$\mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d) = \left\{ \int_0^\cdot Q_s dX_s, \quad Q : [0, 1] \times \mathcal{W}^d \mapsto \mathcal{O}^d \text{ is } (\mathcal{H}_t^{\mathbb{W}^d})_{t \in [0, 1]} \text{-progressively measurable} \right\}.$$

More generally, if  $\sigma$  is invertible, then  $\sigma^\dagger(t, X) = \sigma^{-1}(t, X)$  and (10) holds. In this case we have a complete description of the structure of bicausal Monge transports between  $\mu^{b,\sigma}$  and  $\mu^{\bar{b},\bar{\sigma}}$  (see Figure 2):

COROLLARY 3.4 (Bicausal Monge transport between strong solutions of SDEs). Assume the SDEs (5) and (6) have strong solutions satisfying pathwise uniqueness, given by the Ito maps  $F^{\sigma,b} : (\mathcal{W}^d, \mathcal{F}_1) \rightarrow (\mathcal{W}^d, \mathcal{F}_1)$  (resp.  $F^{\bar{\sigma},\bar{b}}$ ). If  $\sigma(t, \omega)$  is invertible for all  $(t, \omega)$ , then the set of all bicausal Monge transports from  $\mu_z^{\sigma,b}$  to  $\mu_{\bar{z}}^{\bar{\sigma},\bar{b}}$  is

$$\left\{ F^{\bar{\sigma},\bar{b}} \circ T \circ R_z^{\sigma,b}, \quad T \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d) \right\},$$

where  $R_z^{\sigma,b} : \mathcal{W}^d \rightarrow \mathcal{W}^d$  is defined by

$$R_z^{\sigma,b}(X) = \int_0^\cdot \sigma^{-1}(s, X) dM_s \quad \mu_z^{\sigma,b} - \text{a.s. and} \quad M_\cdot := X_\cdot - z - \int_0^\cdot b(s, X) ds.$$

PROOF. Let  $\eta := \mu_z^{\sigma,b}$  and  $(\tilde{\mathcal{F}}_t)_{t \in [0,1]}$  be the right-continuous version of the filtration generated by  $R_z^{\sigma,b} = ((R_z^{\sigma,b})_t)_{t \in [0,1]}$ , where  $(R_z^{\sigma,b})_t := \int_0^t \sigma^{-1}(s, X) dM_s$ . We claim that  $(\tilde{\mathcal{F}}_t^\eta)_{t \in [0,1]} = (\mathcal{H}_t^\eta)_{t \in [0,1]}$ . It is clear that  $\tilde{\mathcal{F}}_t^\eta \subseteq \mathcal{H}_t^\eta$  for all  $t \in [0,1]$ . To see the converse, note that  $((R_z^{\sigma,b})_t)_{t \in [0,1]}$  is a  $(\tilde{\mathcal{F}}_t^\eta)_{t \in [0,1]}$ -Brownian motion by Lévy's characterization of Brownian motion. Furthermore, by definition of  $((R_z^{\sigma,b})_t)_{t \in [0,1]}$  and (7), we have

$$X_t = z + \int_0^t b(s, X) ds + \int_0^t \sigma(s, X) d(R_z^{\sigma,b})_s, \quad t \in [0, 1] \quad \eta\text{-a.s.}$$

By pathwise uniqueness,  $X$  must be adapted to the filtration generated by Brownian motion  $R_z^{\sigma,b}$  so that  $\mathcal{H}_t^\eta \subseteq \tilde{\mathcal{F}}_t^\eta$  for all  $t \in [0, 1]$ .

Let  $\hat{T}$  be a bicausal Monge map from  $\mu_z^{\sigma,b}$  to  $\mu_z^{\bar{\sigma}, \bar{b}}$ . Then, by Theorem 3.3, there exists some  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -progressively measurable  $\mathcal{O}^d$ -valued process  $(Q_t)_{t \in [0,1]}$  such that (9) and (10) holds and by pathwise uniqueness, it is clear that  $\hat{T} = F^{\bar{\sigma}, \bar{b}}(\int_0^t Q_s d(R_z^{\sigma,b})_s)$ . But since  $(\tilde{\mathcal{F}}_t^\eta)_{t \in [0,1]} = (\mathcal{H}_t^\eta)_{t \in [0,1]}$ , then there exists a  $(\mathcal{H}_t^{\mathbb{W}^d})_{t \in [0,1]}$ -progressively measurable  $\mathcal{O}^d$ -valued process  $(\tilde{Q}_t)$  such that  $\tilde{Q}_t \circ R^{\sigma,b} = Q_t$   $\eta$ -a.s. for all  $t \in [0, 1]$ . The map  $X \mapsto T(X) := \int_0^\cdot \tilde{Q}_s dX_s$  is a bicausal Monge map from  $\mathbb{W}^d$  to  $\mathbb{W}^d$  by Corollary 3.3 and it can be readily verified that  $\eta$ -a.s.  $T(R_z^{\sigma,b}) = \int_0^\cdot Q_s d(R_z^{\sigma,b})_s$ . Hence,  $\hat{T} = F^{\bar{\sigma}, \bar{b}}(T(R_z^{\sigma,b}))$  as required.

Finally, since (10) is satisfied, for any  $T \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ , the map  $F^{\bar{\sigma}, \bar{b}}(T(R_z^{\sigma,b}))$  is a bicausal Monge map from  $\mu_z^{\sigma,b}$  to  $\mu_z^{\bar{\sigma}, \bar{b}}$  using Theorem 3.3.  $\square$

Corollary 3.4 is the case where  $\text{Ker}(\sigma(s, X)) = \{0\} \subseteq \text{Ker}(\bar{\sigma}(s, T))$  so that (10) is *always* satisfied, while Corollary 3.2 is the case where (10) is *never* satisfied. One can construct intermediate situations via conditions on the diffusion coefficient  $\bar{\sigma}$ . For example,

COROLLARY 3.5. *Let  $\sigma, \bar{\sigma} \in \mathcal{A}^{d,d}$  be such that for all  $t \in [0, 1]$ ,  $\omega, \omega' \in \mathcal{W}^d$ ,*

$$\left| \bar{\sigma}(t, \omega) - \bar{\sigma}(t, \omega') \right| \leq L \|\omega - \omega'\|_\infty \quad \text{and} \quad \int_0^1 |\bar{\sigma}(s, 0)|^2 ds < \infty.$$

*for some  $L > 0$ . If  $\text{Ker}(\bar{\sigma})$  is deterministic and for all  $s \in [0, 1]$ ,*

$$\max_{\omega \in E} \dim(\text{Ker}(\sigma(s, \omega))) \leq \dim(\text{Ker}(\bar{\sigma})(s)),$$

*there exists a bicausal Monge transport from  $\mu_z^{\sigma,0}$  to  $\mu_z^{\bar{\sigma},0}$ .*

PROOF. We construct an  $\mathcal{O}^d$ -valued process  $(Q_t)_{t \in [0,1]}$  such that for all  $s \in [0, 1], \omega \in \mathcal{W}^d$ ,

$$Q_s \text{Ker}(\sigma(s, \omega)) \subseteq \text{Ker}(\bar{\sigma}(s))$$

so that (10) holds. Let  $\bar{m}(s) := \dim(\text{Ker}(\bar{\sigma}(s)))$  and  $m(s, \omega) := \dim(\text{Ker}(\sigma(s, \omega)))$  so that  $m \leq \bar{m}$ . Letting  $\bar{\sigma} = \bar{U} \bar{\Sigma} \bar{V}^*$  be the singular value decomposition (SVD) of  $\bar{\sigma}$ , we see that for fixed  $s \in [0, 1]$ ,  $d - \bar{m} = \max \left\{ k : \bar{\Sigma}_{kk} > 0 \right\}$  (without loss of generality, the diagonal entries of  $\bar{\Sigma}$  are assumed to be decreasing). Hence, writing  $\bar{v}_i^*$  as the  $i^{\text{th}}$ -column of  $\bar{V}^*$  and  $\bar{\alpha} := d - \bar{m}$ , then  $\text{Ker}(\bar{\sigma}(s, \omega)) = \text{Span}(\{\bar{v}_i^*\}_{i=\bar{\alpha}+1}^d)$ . Since  $\text{Ker}(\bar{\sigma}(s, \omega))$  is deterministic, up to replacing the last  $\bar{m}$  columns of  $\bar{V}^*$ , we assume that  $\{\bar{v}_i^*\}_{i=\bar{\alpha}+1}^d$  are vectors depending on time, but not on  $\omega$ . Hence, let  $\bar{O}_s^*$  be any time-dependent orthogonal matrix whose last

$\overline{m}$  columns are  $\{\overline{v}_i^*\}_{i=\overline{\alpha}+1}^d$ . Letting  $\sigma = U\Sigma V^*$  be the SVD of  $\sigma$ , one easily verifies that  $Q(s, \omega) := \overline{O}_s^* V(s, \omega)$  is the desired  $Q$  noting that the kernel of  $\sigma(s, \omega)$  is given by the last  $m$  vectors of  $V_s^*$ . Picard iteration applied to (9) yields the existence of a bicausal Monge transport.  $\square$

### 3.4. Proof of Theorem 3.2 and 3.3.

**3.4.1. Martingale Representation Theorem.** A crucial ingredient in the proof of our results is a martingale representation property [36, Theorem 2] applicable to SDEs with unique weak solutions (see also [13, Ch. 7]).

**PROPOSITION 3.1.** *On  $(\mathcal{W}^d, \mathcal{F}_1, (\mathcal{H}_t^\eta)_{t \in [0,1]}, \eta)$ , where  $\eta := \mu_z^{\sigma, b}$ , we have that for any  $F \in L^2(\eta, \mathcal{F}_1)$ , there exists a  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -progressively measurable process  $\xi : [0, 1] \times \mathcal{W}^d \rightarrow \mathbb{R}^d$  such that*

$$\mathbb{E}^\eta \left( \int_0^1 \|\xi_s\|_{\mathbb{R}^d}^2 ds \right) < \infty$$

and

$$F = \int_0^1 (\xi_s, dL_s)_{\mathbb{R}^d} = \int_0^1 \sum_{i=1}^d \xi_s^i dL_s^i \quad \eta\text{-a.s.},$$

where  $L = \int_0^\cdot \sigma^\dagger(s, X) dM_s^\eta$ . In particular, any square-integrable  $(\eta, (\mathcal{H}_t^\eta)_{t \in [0,1]})$ -martingale  $(M_t)_{t \in [0,1]}$  has representation

$$M_t = \int_0^t \sum_{i=1}^d \xi_s^i dL_s^i, \quad \forall t \in [0, 1] \quad \eta\text{-a.s.}$$

**PROOF.** Consider the filtered probability space

$$(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0,1]}, \mathbb{P}) := (\mathcal{W}^d \times \mathcal{W}^d, \mathcal{F}_1 \otimes \mathcal{F}_1, (\mathcal{P}_t^{\eta \otimes \mathbb{W}^d})_{t \in [0,1]}, \eta \otimes \mathbb{W}^d),$$

using the notation in (3). Since  $X_\# \mathbb{P} = \eta$ , we can identify  $L^2(\eta, \mathcal{F}_1)$  with  $L^2(\mathbb{P}, \mathcal{F}_1 \otimes \{\emptyset, \mathcal{W}^d\})$  by the Doob-Dynkin lemma (see [26]) and we no longer distinguish these spaces. It is straightforward to verify that  $\mathbb{P} \in \Pi_{bc}(\eta, \mathbb{W}^d)$ . (7) and Theorem 2.2 implies that  $(X_t)_{t \in [0,1]}$  is a  $(\mathbb{P}, (\mathcal{G}_t))$ -semimartingale such that  $\left(X_t - z - \int_0^t b(s, X) ds\right)_{t \in [0,1]}$  is a  $(\mathbb{P}, (\mathcal{G}_t))$ -local martingale and

$$\forall t \in [0, 1], \quad [X, X]_t = \int_0^t \sigma(s, X) \sigma^*(s, X) ds \quad \mathbb{P}\text{-a.s.}$$

and  $(Y_t)_{t \in [0,1]}$  is a  $(\mathbb{P}, (\mathcal{G}_t))$ -Brownian motion independent from  $(X_t)_{t \in [0,1]}$ . Following the proof of [25, Theorem 7.1'], we define

$$(12) \quad B_t := \int_0^t \sigma^\dagger(s, X) dM_s + \int_0^t (\text{Id}_d - \sigma^\dagger(s, X) \sigma(s, X)) dY_s,$$

where  $M_t := X_t - z - \int_0^t b(s, X) ds$ . Then,  $(B_t)_{t \in [0,1]}$  is a  $(\mathbb{P}, (\mathcal{G}_t))$ -Brownian motion by Lévy's characterization of Brownian motion and  $M_t = \int_0^t \sigma(s, X) dB_s$  for all  $t \in [0, 1]$   $\mathbb{P}$ -a.s..

That is, the pair  $(X, B)$  satisfies SDE (5). Since  $\sigma^\dagger(s, X)\sigma(s, X)$  is an orthogonal projection onto the range of  $\sigma^*(s, X)$  (see [23, Section 5.5.4]), we have by [36, Theorem 2] that for any  $F \in L^2(\eta, \mathcal{F}_1)$ , there exists a  $(\mathcal{H}_t^\eta \otimes \{\emptyset, \mathcal{W}^d\})_{t \in [0,1]}$ -progressively measurable process  $(\xi_t)$  such that

$$\mathbb{E}^\mathbb{P} \left( \int_0^1 \|\xi_s\|_{\mathbb{R}^d}^2 ds \right) < \infty$$

and

$$F = \int_0^1 \left( \xi_s, \sigma^\dagger(s, X)\sigma(s, X)dB_s \right)_{\mathbb{R}^d} \mathbb{P}\text{-a.s.}$$

Using (12), one can verify  $\sigma^\dagger(s, X)\sigma(s, X)dB_s = \sigma^\dagger(s, X)dM_s$ , from which we obtain the desired result.  $\square$

### 3.4.2. Technical proofs.

PROOF OF THEOREM 3.2. Denote  $\eta := \mu_z^{\sigma, b}$  and  $\nu := \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}$ . We first prove sufficiency, that is,  $\pi = (Z, \bar{Z})_\# \mathbb{P} \in \Pi_{bc}(\eta, \nu)$ . By Assumption 1, it is clear  $\pi \in \Pi(\eta, \nu)$  and it remains to show bicausality of  $\pi$ . Define  $\mathcal{P}_t^\pi$  as in (3). We begin by verifying that  $X$  is a  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -semimartingale, for which it is enough to show that  $M_t := X_t - z - \int_0^t b(s, X)ds$  is a  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -local martingale. Up to localization, we may assume  $(M_t)_{t \in [0,1]}$  is bounded. Since  $Z^{-1}(\mathcal{H}_t) = \cap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}^Z$ , where  $(\mathcal{F}_t^Z)_{t \in [0,1]}$  is the filtration generated by  $(Z_t)_{t \in [0,1]}$ , and  $(\mathcal{G}_t)_{t \in [0,1]}$  is right-continuous by assumption, it follows that  $Z^{-1}(\mathcal{H}_t) \subseteq \mathcal{G}_t$ . Similarly,  $\bar{Z}^{-1}(\mathcal{H}_t) \subseteq \mathcal{G}_t$ . Hence, we calculate for  $0 \leq s < t \leq 1$ ,  $A, B \in \mathcal{H}_s$  and  $(\omega, \omega') \mapsto \mathbb{1}_{A \times B}(\omega, \omega')$ ,

$$\begin{aligned} \mathbb{E}^\pi(M_t \mathbb{1}_{A \times B}) &= \mathbb{E}^\mathbb{P} \left( \int_0^t \sigma(u, Z_u)dB_u \mathbb{1}_{Z^{-1}(A) \cap \bar{Z}^{-1}(B)} \right) \\ &= \mathbb{E}^\mathbb{P} \left( \int_0^s \sigma(u, Z_u)dB_u \mathbb{1}_{Z^{-1}(A) \cap \bar{Z}^{-1}(B)} \right) = \mathbb{E}^\pi(M_s \mathbb{1}_{A \times B}) \end{aligned}$$

since  $\int_0^\cdot \sigma(u, Z_u)dB_u$  is a  $(\mathbb{P}, (\mathcal{G}_t))$ -martingale. This shows that  $(M_t)_{t \in [0,1]}$  is a  $(\pi, (\mathcal{H}_t \otimes \mathcal{H}_t))$ -martingale and even a  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -martingale by continuity of sample paths and [34, Lemma 67.10]. By the stochastic integral representation of elements in  $\mathcal{M}^2(\pi, (\mathcal{H}_t^\eta \otimes \{\emptyset, \mathcal{W}^d\}))$  from Proposition 3.1 and Theorem 2.2, we conclude  $\pi \in \Pi_c(\eta, \nu)$ . A completely symmetric argument yields  $\pi \in \Pi_{bc}(\eta, \nu)$  as required.

To see necessity, let  $\pi \in \Pi_{bc}(\eta, \nu)$ . By Theorem 2.2,  $X$  and  $Y$  are  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -semimartingales. Up to enlarging the probability space, we assume that there exists a Brownian motion  $\hat{B}$  independent from  $X$  and  $Y$ . It is easy to check that  $X, Y$  and  $\hat{B}$  remain semimartingales with respect to probability measure  $\pi$  and the (complete, right-continuous) filtration generated by all three processes, denoted  $(\mathcal{G}_t)$ . As in the proof of Proposition 3.1, we construct  $(\mathcal{G}_t)$ -Brownian motions

$$\begin{aligned} B_t &:= \int_0^t \sigma^\dagger(s, X)dM_s + \int_0^t (\text{Id}_d - \sigma^\dagger(s, X)\sigma(s, X))d\hat{B}_s, \\ \bar{B}_t &:= \int_0^t \bar{\sigma}^\dagger(s, Y)dN_s + \int_0^t (\text{Id}_d - \bar{\sigma}^\dagger(s, Y)\bar{\sigma}(s, Y))d\hat{B}_s, \end{aligned}$$

where  $N_t := Y_t - \bar{z} - \int_0^t \bar{b}(s, Y) ds$ . Noting that  $M_t = \int_0^t \sigma(s, X) dB_s$  and  $N_t = \int_0^t \bar{\sigma}(s, Y) d\bar{B}_s$  yields the desired.

The final assertion regarding the correlation process  $\rho$  is exactly [18, Lemma 5].  $\square$

To prove Theorem 3.3, we require the following generalization of [28, Proposition 4].

LEMMA 3.4. *Under the assumptions of Theorem 3.3, a measurable map  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  from  $\eta$  to  $\nu$ , i.e.  $T_{\#}\eta = \nu$ , is a bicausal Monge map if and only if  $(T_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta)_{t \in [0,1]})$ -semimartingale.*

PROOF. Suppose  $T$  is a bicausal Monge map from  $\eta$  to  $\nu$ , i.e.  $\pi_T := (X, T)_{\#}\eta \in \Pi_{bc}(\eta, \nu)$ . Causality readily implies that  $(T_t)_{t \in [0,1]}$  is  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -adapted. Since  $T$  has law  $\mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}$ , then  $(T_t)_{t \in [0,1]}$  is a semimartingale with respect to probability measure  $\eta$  and its own filtration, with martingale part  $N_t := T_t - \bar{z} - \int_0^t \bar{b}(s, T) ds$ . Up to standard localization methods, we assume  $(N_t)_{t \in [0,1]}$  is bounded. Bicausality implies that  $(N_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta))$ -martingale, which implies the desired. Indeed, for  $0 \leq s \leq t \leq 1$  and  $A \in \mathcal{H}_s$ , we have

$$\begin{aligned} (13) \quad \mathbb{E}^\eta(N_t \mathbb{1}_A) &= \mathbb{E}^{\pi_T}((Y_t - \bar{z} - \int_0^t \bar{b}(u, Y) du) \mathbb{1}_{A \times \mathcal{W}^d}) \\ &= \mathbb{E}^{\pi_T}((Y_s - \bar{z} - \int_0^s \bar{b}(u, Y) du) \mathbb{1}_{A \times \mathcal{W}^d}) = \mathbb{E}^\eta(N_s \mathbb{1}_A), \end{aligned}$$

where we have used causality of  $R_{\#}\pi_T \in \Pi_c(\nu, \eta)$  and Theorem 2.2 in the second equality.

Conversely, if  $(T_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta))$ -semimartingale, then it is  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -adapted so that  $\pi_T \in \Pi_c(\eta, \nu)$ . Furthermore, it holds that the first and last term in (13) is equal, so that the second and third term are equal and  $(Y_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{P}_t^{\pi_T})_{t \in [0,1]})$ -semimartingale. By the stochastic integral representation of elements in  $\mathcal{M}^2(\pi_T, (\{\emptyset, \mathcal{W}^d\} \otimes \mathcal{H}_t^\eta))$  from Proposition 3.1 and Theorem 2.2, we conclude  $\pi_T \in \Pi_{bc}(\eta, \nu)$ .  $\square$

PROOF OF THEOREM 3.3. We start with sufficiency. Suppose  $T$  is a measurable map such that  $(T_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta))$ -semimartingale satisfying (9) and (10). By Lemma 3.4, it is sufficient to show that  $T_{\#}\eta = \nu$ . As in the proof of Proposition 3.1, up to enlarging the probability space  $(\mathcal{W}^d, \mathcal{F}_1, \eta)$ , there exists some Brownian motion  $B$  such that  $\int_0^\cdot \sigma^\dagger(s, X) dM_s^\eta = \int_0^\cdot \sigma^\dagger(s, X) \sigma(s, X) dB_s$ . Hence, (9) and (10) implies  $\eta$ -a.s.

$$\begin{aligned} T_t &= \bar{z} + \int_0^t \bar{b}(s, T) ds + \int_0^t \bar{\sigma}(s, T) Q_s \sigma^\dagger(s, X) \sigma(s, X) dB_s \\ &= \bar{z} + \int_0^t \bar{b}(s, T) ds + \int_0^t \bar{\sigma}(s, T) Q_s dB_s. \end{aligned}$$

Since  $\int_0^\cdot Q_s dB_s$  is a Brownian motion, the law of  $T$  is indeed  $\nu$  by weak uniqueness.

For necessity, if  $T$  is a bicausal Monge map from  $\eta$  to  $\nu$ , then by Lemma 3.4, we have that  $(T_t)_{t \in [0,1]}$  is a  $(\eta, (\mathcal{H}_t^\eta))$ -semimartingale with martingale part  $(N_t)_{t \in [0,1]}$  where  $N_t := T_t - \int_0^t \bar{b}(s, T) ds - \bar{z}$  and  $\eta$ -a.s.

$$[T]_t = \int_0^t \bar{\sigma}(s, T) \bar{\sigma}^*(s, T) ds \text{ for } t \in [0, 1].$$

On the other hand, standard localization procedures and Proposition 3.1 yields a  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -progressively measurable process  $H : [0, 1] \times \mathcal{W}^d \rightarrow \mathbb{R}^{d \times d}$  such that for  $t \in [0, 1]$ ,  $\eta$ -a.s.

$$(14) \quad N_t = \int_0^t H_s \sigma^\dagger(s, X) dM_s^\eta$$

with quadratic variation

$$[T]_t = [N]_t = \int_0^t \left( H_s \sigma^\dagger(s, X) \sigma(s, X) \right) \left( H_s \sigma^\dagger(s, X) \sigma(s, X) \right)^* ds \quad \eta\text{-a.s.}$$

Hence, we obtain the  $\eta$ -a.s. equality

$$\int_0^t \left( H_s \sigma^\dagger(s, X) \sigma(s, X) \right) \left( H_s \sigma^\dagger(s, X) \sigma(s, X) \right)^* ds = \int_0^t \bar{\sigma}(s, T) \bar{\sigma}^*(s, T) ds.$$

By Lebesgue Differentiation Theorem, we obtain that  $\eta$ -a.s. on a set of full Lebesgue measure

$$(H_s \sigma^\dagger(s, X) \sigma(s, X)) (H_s \sigma^\dagger(s, X) \sigma(s, X))^* = \bar{\sigma}(s, T) \bar{\sigma}^*(s, T).$$

By polar decomposition, this implies the existence of some  $(\mathcal{H}_t^\eta)_{t \in [0,1]}$ -progressively measurable<sup>4</sup>  $Q : [0, 1] \times \mathcal{W}^d \rightarrow \mathcal{O}^d$  such that  $\eta$ -a.s. on a set of full Lebesgue measure, we have

$$H_s \sigma^\dagger(s, X) \sigma(s, X) = \bar{\sigma}(s, T) Q_s \implies H_s \sigma^\dagger(s, X) = \bar{\sigma}(s, T) Q_s \sigma^\dagger(s, X)$$

$$\text{and } \bar{\sigma}(s, T) Q_s = \bar{\sigma}(s, T) Q_s \sigma^\dagger(s, X) \sigma(s, X),$$

from which we obtain the desired using (14).  $\square$

#### 4. Applications.

4.1. *Bicausal couplings induced by Monge maps.* While Theorem 3.3 gives us necessary and sufficient conditions for a measurable map  $T$  to be a bicausal Monge transport, it does not tell us *when* a bicausal transport plan is induced by a bicausal Monge map. We will now show that whether a bicausal coupling is induced by a Monge map depends on the existence of a strong solution to an associated SDE (15).

In light of Corollary 3.1, we first study the case of bicausal Monge transport between Wiener measures:

**PROPOSITION 4.1** (Bicausal Monge transport between Wiener measures). *Let  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  and  $\rho : [0, 1] \times \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathcal{C}^d$  be the  $(\mathcal{P}_t^\pi)_{t \in [0,1]}$ -progressively measurable correlation process of  $X, Y$  under  $\pi$ :*

$$\rho(t, X, Y) = \frac{d[X, Y]}{dt}, \quad \pi - \text{a.s.}$$

1. *If  $\rho$  is an  $\mathcal{O}^d$ -valued adapted process and there exists a strong, pathwise unique solution to the SDE*

$$(15) \quad d \begin{pmatrix} Z \\ \bar{Z} \end{pmatrix}_t = \begin{pmatrix} Id_d \\ \rho(t, Z, \bar{Z}) \end{pmatrix} dB_t, \quad \begin{pmatrix} Z_0 \\ \bar{Z}_0 \end{pmatrix} = 0,$$

*then  $\pi \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ .*

---

<sup>4</sup>To see progressively measurability of  $Q$ , one can show that polar decomposition can be done in a measurable way following [3].



2. If  $\pi$  is induced by a bicausal Monge map  $T$ , then  $\rho$  is  $\mathcal{O}^d$ -valued and there exists a strong solution to (15).

REMARK 4 (Measurability of the correlation process). A priori, for  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ , the marginal processes  $X$  and  $Y$  will be  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0,1]})$ -Brownian motions and hence,  $\rho$ , constructed via the Lebesgue differentiation of the quadratic covariation  $[X, Y]$ , will be  $(\mathcal{P}_t^\pi)_{t \in [0,1]}$ -progressively measurable. Yet, in (15), we consider  $\rho$  as a component of a diffusion coefficient of an SDE, which one usually assumes to be  $(\mathcal{F}_t \otimes \mathcal{F}_t)_{t \in [0,1]}$ -progressively measurable. Fortunately, one can always find some  $\tilde{\rho} : [0, 1] \times \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathcal{C}^d$   $\pi$ -indistinguishable from  $\rho$  that is  $(\mathcal{F}_t \otimes \mathcal{F}_t)_{t \in [0,1]}$ -progressively measurable by using the fact that the completion of the canonical filtration of a Brownian motion is always right-continuous and [24, Theorem 4.37.3].

PROOF. *i.* : Under  $\pi$ ,  $X$  and  $Y$  are  $(\mathcal{P}_t^\pi)_{t \in [0,1]}$ -Brownian motions such that  $[X, Y] = \int_0^\cdot \rho(s, X, Y) ds$  by Theorem 2.2. Using the fact that  $\rho$  is  $\mathcal{O}^d$ -valued, a direct calculations shows that  $[\int_0^\cdot \rho(s, X, Y) dX_s - Y] \equiv 0$  and hence,  $Y_t = \int_0^t \rho(s, X, Y) dX_s$   $\pi$ -a.s. for all  $t \in [0, 1]$ . In other words,  $(X, Y)$  is a solution to (15). By pathwise uniqueness, it follows that  $Y$  must be adapted to  $X$  and hence,  $\pi \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ .

*ii.* : Since  $\pi$  is induced by a bicausal Monge map, then Corollary 3.3 implies that  $\rho$  must be  $\mathcal{O}^d$ -valued and as before, we again have  $Y_t = \int_0^t \rho(s, X, Y) dX_s$   $\pi$ -a.s.. Hence,

$$T_t = \int_0^t \rho(s, X, T) dX_s, \quad \mathbb{W}^d\text{-a.s.}$$

and  $(T_t)_{t \in [0,1]}$  is adapted to  $(\mathcal{H}_t^{\mathbb{W}^d})_{t \in [0,1]}$  - the filtration of Brownian motion  $X$ , which implies the existence of a strong solution given by  $(X, T)$ .  $\square$

REMARK 5. The preceding proposition almost gives a complete characterization of the set  $\mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  in terms of the existence of strong solutions to (15). However, at present, the hypothesis of *ii.* seems insufficient to guarantee a *pathwise unique* strong solution to (15). If we were to assume the hypothesis of *ii.* and in addition, assume the weak uniqueness of (15), then the existence of a strong solution along with weak uniqueness implies the existence of strong, pathwise unique solution by a result of A. S. Cherny [12, Theorem 3.2]. This holds for example

1. when  $\rho$  is only a function of  $X$  and  $t$ , so that weak uniqueness of the SDE is trivial;
2. or when  $\rho$  is only a function of  $Y$  and  $t$ . Indeed, any solutions to the SDE

$$d\bar{Z}_t = \rho(t, Z, \bar{Z}) dB_t = \rho(t, \bar{Z}) dB_t$$

will always have law given by  $\mathbb{W}^d$  since  $\rho \in \mathcal{O}^d$ . Hence, the preceding SDE has a unique weak solution and, by [12, Theorem 3.1], the joint law of  $(\bar{Z}, B)$  is the unique weak solution of (15).

Finally, we note that if  $\gamma \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  is such that  $\rho(X, Y, t) = \mathbb{1}_{\{Y_t > 0\}} - \mathbb{1}_{\{Y_t \leq 0\}}$ , then (15) reduces to the well-known Tanaka's SDE from which we see that there cannot exist a strong solution. Hence,  $\gamma$  cannot be induced by a bicausal Monge transport.

For bicausal couplings between laws of SDEs with strong, pathwise unique solutions, we have the following corollary under the same assumptions as in Corollary 3.4:

COROLLARY 4.1. *Assume the SDEs (5) and (6) have strong solutions satisfying pathwise uniqueness, given by the Ito maps  $F^{\sigma,b} : (\mathcal{W}^d, \mathcal{F}_1) \rightarrow (\mathcal{W}^d, \mathcal{F}_1)$  (resp.  $F^{\bar{\sigma},\bar{b}}$ ) and fix  $\pi \in \Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$ . By Corollary 3.1, there exists  $\hat{\pi} \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  such that*

$$\pi = \left( F^{\sigma,b}, F^{\bar{\sigma},\bar{b}} \right)_{\#} \hat{\pi}.$$

*If  $\sigma(t, \omega)$  and  $\bar{\sigma}(t, \omega)$  are invertible for all  $(t, \omega) \in [0, 1] \times \mathcal{W}^d$ , then*

$$\pi \in \mathcal{T}_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}}) \iff \hat{\pi} \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d).$$

PROOF. It is easy to verify that  $\mathbb{W}^d$ -a.s.  $R_z^{\sigma,b} \circ F^{\sigma,b} = X$ . and by pathwise uniqueness,  $\mu_z^{\sigma,b}$ -a.s.  $F^{\sigma,b} \circ R_z^{\sigma,b} = X$ . In particular,  $(R_z^{\sigma,b}, R_{\bar{z}}^{\bar{\sigma},\bar{b}})_{\#} \pi = \hat{\pi}$ . Suppose now that  $\hat{\pi} \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ , i.e. there exist some measurable  $T : \mathcal{W}^d \rightarrow \mathcal{W}^d$  such that  $(X, T)_{\#} \mathbb{W}^d = \hat{\pi}$ . Then,

$$\begin{aligned} \pi &= \left( F^{\sigma,b}, F^{\bar{\sigma},\bar{b}} \right)_{\#} \hat{\pi} = \left( F^{\sigma,b}, F^{\bar{\sigma},\bar{b}} \right)_{\#} (X, T)_{\#} \mathbb{W}^d \\ &= \left( F^{\sigma,b}, F^{\bar{\sigma},\bar{b}} \circ T \right)_{\#} \left( R_z^{\sigma,b} \right)_{\#} \mu_z^{\sigma,b} = \left( X, F^{\bar{\sigma},\bar{b}} \circ T \circ R_z^{\sigma,b} \right)_{\#} \mu_z^{\sigma,b}. \end{aligned}$$

Corollary 3.4 gives that  $F_{\bar{z}}^{\bar{\sigma},\bar{b}} \circ T \circ R_z^{\sigma,b} \in \mathcal{T}_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$ . Similar relations hold for  $R_{\bar{z}}^{\bar{\sigma},\bar{b}}$  and  $F^{\bar{\sigma},\bar{b}}$  if  $\bar{\sigma}(s, \omega)$  is non-singular. Conversely, if  $\pi \in \mathcal{T}_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$ , i.e.  $\pi = (X, \hat{T})_{\#} \mu_z^{\sigma,b}$  for some measurable  $\hat{T}$ , then  $\hat{\pi} = (X, R_{\bar{z}}^{\bar{\sigma},\bar{b}} \circ \hat{T} \circ F^{\sigma,b})_{\#} \mathbb{W}^d \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ .  $\square$

4.2. *Denseness of bicausal Monge maps among bicausal couplings.* We establish conditions under which couplings induced by bicausal Monge maps are dense in the space of bicausal couplings. We first show that  $\mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  is dense in  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  for the topology of weak convergence of probability measures before moving on to the general case. We use the result of M. Émery [18, Proposition 2], which showed that the set of “mutually adapted Brownian motions” is dense in the set of “jointly immersed Brownian motions”, which is exactly equal to  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  in light of Theorem 3.2.

PROPOSITION 4.2.  *$\mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  is dense in  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  for the topology of weak convergence of probability measures.*

PROOF. By [18, Proposition 2], given any  $\pi \in \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ , there exists Brownian motions  $X^n$  and  $Y^n$  on some abstract filtered probability space  $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0,1]}, \mathbb{P})$  such that for each fixed  $n$ ,  $X^n$  and  $Y^n$  generate the same filtration and the law of  $(X^n, Y^n)$  weakly converges to  $\pi$  as  $n$  tends to infinity. Since  $X^n$  and  $Y^n$  generate the same filtration, then there exists some  $T^n : \mathcal{W}^d \rightarrow \mathcal{W}^d$  such that  $T^n(X^n) = Y^n$   $\mathbb{P}$ -a.s. and  $Y^n$  is a Brownian motion with respect to the filtration generated by  $X^n$ . Since  $Y^n$  is a Brownian motion, it is easy to see that  $T^n_{\#} \text{Law}(X^n) = T^n_{\#} \mathbb{W}^d = \mathbb{W}^d$ . Lemma 3.4 gives that  $T^n \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  and that  $\text{Law}(X^n, Y^n) = (X, T^n)_{\#} \mathbb{W}^d$  weakly converges to  $\pi$ .  $\square$

REMARK 6. A longer, but more explicit proof may be obtained along the lines of the proof of [18, Proposition 2] where  $Y^n$  is constructed explicitly as a stochastic integral against  $X^n$  of an  $\mathcal{O}^d$ -valued process, in agreement with Corollary 3.3.

In fact, the smaller set of *bijective* bicausal Monge transports is also dense subset of bicausal couplings:

PROPOSITION 4.3.  $\mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  is dense in  $\Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  for the topology of weak convergence of probability measures.

PROOF. In the proof of Proposition 4.2, since  $X^n$  and  $Y^n$  generate the same filtration (see [18, Proposition 2]), the role of  $X^n$  and  $Y^n$  are completely symmetric, i.e. there exists some  $\hat{T}^n : \mathcal{W}^d \rightarrow \mathcal{W}^d$  such that  $\hat{T}^n(Y^n) = X^n$   $\mathbb{P}$ -a.s. and  $\hat{T}^n \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ . It then holds that  $X^n = \hat{T}^n \circ T^n(X^n)$   $\mathbb{P}$ -a.s.. In other words,  $\hat{T}^n \circ T^n = X$   $\mathbb{W}^d$ -a.s.. A similar statement holds for  $T^n \circ \hat{T}^n$ .  $\square$

We now transfer this density result between couplings of Wiener measures to the density of  $\mathcal{T}_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  in  $\Pi_{bc}(\mu_z^{\sigma,b}, \mu_{\bar{z}}^{\bar{\sigma},\bar{b}})$  using the following abstract lemma:

LEMMA 4.1. Let  $\eta, \nu \in \mathcal{P}(\mathcal{W}^d)$  and sets  $\Lambda, \Upsilon \subseteq \mathcal{P}(\mathcal{W}^d \times \mathcal{W}^d)$  be such that  $\Upsilon \subseteq \Lambda \subseteq \Pi(\eta, \nu)$ ,  $\Lambda$  is closed and  $\Upsilon$  is dense in  $\Lambda$ . For any measurable  $F, \bar{F} : \mathcal{W}^d \rightarrow \mathcal{W}^d$ ,

$$(F, \bar{F})_{\#} \Lambda := \{(F, \bar{F})_{\#} \pi : \pi \in \Lambda\} \text{ is closed}$$

and

$$\{(F, \bar{F})_{\#} \pi : \pi \in \Upsilon\} \text{ is dense in } (F, \bar{F})_{\#} \Lambda,$$

where  $(F, \bar{F})_{\#} \pi$  designates the image of the measure  $\pi$  under the map

$$(X, Y) \in \mathcal{W}^d \times \mathcal{W}^d \mapsto (F(X), \bar{F}(Y)).$$

Note that the lemma is easily seen to be true if  $F, \bar{F}$  are assumed continuous since  $\Pi(\eta, \nu) \ni \pi \mapsto (F, \bar{F})_{\#} \pi$  will be a continuous function over a compact set. For  $F, \bar{F}$  measurable, one can enlarge the topology on  $\mathcal{W}^d$  without changing the Borel  $\sigma$ -algebra such that  $F$  is continuous by [27, Theorem 13.11]. By [9, Lemma 2.5], the topology of weak convergence of probability measures on  $\Pi(\eta, \nu)$  remains unchanged and we can safely conclude the lemma. We provide a more elementary and self-contained proof below relying only on Tietze extension theorem and Lusin's theorem.

PROOF. We first show that  $(F, \bar{F})_{\#} \Lambda$  is closed. Let  $(\pi_n)_{n \geq 0} \subseteq \Lambda$  be a sequence of probability measures such that  $((F, \bar{F})_{\#} \pi_n)_{n \geq 0}$  converges to some  $\xi \in \mathcal{P}(\mathcal{W}^d \times \mathcal{W}^d)$ . Since  $\Lambda$  is closed in compact  $\Pi(\eta, \nu)$ , up to a subsequence, we can assume that  $\pi_n$  converges to some  $\pi \in \Pi(\eta, \nu)$ . We now show that  $\xi = (F, \bar{F})_{\#} \pi$ . Fix a continuous, bounded  $f : \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathbb{R}$  and some  $\epsilon > 0$ . By Lusin's theorem, there exists closed sets  $K, J \subseteq \mathcal{W}^d$  such that  $\eta(K^c), \nu(J^c) < \epsilon$  and  $F|_K, \bar{F}|_J$  is continuous. By Tietze extension theorem [15, Theorem 4.1], there exist continuous  $G, \bar{G} : \mathcal{W}^d \rightarrow \mathcal{W}^d$  such that  $F|_K = G|_K$  and  $\bar{F}|_J = \bar{G}|_J$ . Note that

$$\sup_{\pi \in \Pi(\eta, \nu)} |\mathbb{E}^\pi(f(F, \bar{F}) - f(G, \bar{G}))| \leq 2C \sup_{\pi \in \Pi(\eta, \nu)} \mathbb{E}^\pi \mathbb{1}_{(K \times \bar{K})^c} \leq 4C\epsilon,$$

where  $C := \sup_{\omega, \omega'} |f(\omega, \omega')|$ . Hence, we calculate

$$\begin{aligned} |\mathbb{E}^\xi f - \mathbb{E}^\pi f(F, \bar{F})| &\leq |\mathbb{E}^\xi f - \mathbb{E}^{\pi_n} f(F, \bar{F})| + |\mathbb{E}^{\pi_n} f(F, \bar{F}) - \mathbb{E}^{\pi_n} f(G, \bar{G})| \\ &\quad + |\mathbb{E}^{\pi_n} f(G, \bar{G}) - \mathbb{E}^\pi f(G, \bar{G})| + |\mathbb{E}^\pi f(G, \bar{G}) - \mathbb{E}^\pi f(F, \bar{F})| \\ &\leq |\mathbb{E}^\xi f - \mathbb{E}^{(F, \bar{F})_{\#} \pi_n} f| + |\mathbb{E}^{\pi_n} f(G, \bar{G}) - \mathbb{E}^\pi f(G, \bar{G})| + 8C\epsilon. \end{aligned}$$

Taking  $n \rightarrow \infty$  and  $\epsilon \rightarrow 0$ , we conclude  $\mathbb{E}^\xi f = \mathbb{E}^{(F, \bar{F})_{\#} \pi} f$  for all continuous, bounded  $f$  and hence,  $\xi = (F, \bar{F})_{\#} \pi$  as desired.

To see density of  $\{(F, \bar{F})_{\#} \pi : \pi \in \Upsilon\}$  in  $(F, \bar{F})_{\#} \Lambda$ , fix  $\pi \in \Lambda$ . We wish to show there exists a sequence  $(\pi_n)_{n \geq 0} \subseteq \Upsilon$  such that  $(F, \bar{F})_{\#} \pi_n$  converges weakly to  $(F, \bar{F})_{\#} \pi$ . Since  $\Upsilon$  is dense in  $\Lambda$ , there exists a sequence  $(\pi_n)_{n \geq 0}$  weakly converging to  $\pi$ . Furthermore, since  $((F, \bar{F})_{\#} \pi_n)_{n \geq 0} \subseteq (F, \bar{F})_{\#} \Lambda \subseteq \Pi_{bc}(F_{\#} \eta, \bar{F}_{\#} \nu)$ , up to a subsequence, we can conclude that  $(F, \bar{F})_{\#} \pi_n$  converges to some  $\xi \in \mathcal{P}(\mathcal{W}^d \times \mathcal{W}^d)$  as well. By the same argument as in the previous step, we can conclude  $\xi = (F, \bar{F})_{\#} \pi$ , which implies the desired.  $\square$

**PROPOSITION 4.4.** *Assume the SDEs (5) and (6) have strong solutions satisfying path-wise uniqueness. If  $\sigma(s, \omega)$  is invertible for all  $(s, \omega)$ , then  $\mathcal{T}_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  is dense in  $\Pi_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  for the topology of weak convergence of probability measures.*

**PROOF.** Denote by  $F = F^{\sigma, b}, \bar{F} = F^{\bar{\sigma}, \bar{b}}$  the Ito maps for the SDEs (5) and (6),  $\Gamma = \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  and  $\Lambda = \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ . By applying Lemma 4.1, we conclude that  $(F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_{\#} \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  is dense in  $\Pi_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  since  $\Gamma$  is dense in  $\Lambda$  by Proposition 4.2 and  $(F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_{\#} \Pi_{bc}(\mathbb{W}^d, \mathbb{W}^d) = \Pi_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  by Corollary 3.1. Hence, it suffices to show that  $(F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_{\#} \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d) \subseteq \mathcal{T}_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$ . This follows from Corollary 3.4 since using the fact that  $F^{\sigma, b} \circ R_z^{\sigma, b} = \text{Id}$   $\mu_z^{\sigma, b}$ -a.s., we have, for any  $T \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$ ,

$$(F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_{\#} (\text{Id}, T)_{\#} \mathbb{W}^d = (F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_{\#} (\text{Id}, T)_{\#} (R_z^{\sigma, b})_{\#} \mu_z^{\sigma, b} = (\text{Id}, F^{\bar{\sigma}, \bar{b}} \circ T \circ R_z^{\sigma, b})_{\#} \mu_z^{\sigma, b}.$$

$\square$

**PROPOSITION 4.5.** *Under the assumptions of Proposition 4.4, if  $\sigma(s, \omega)$  and  $\bar{\sigma}(s, \omega)$  are invertible for all  $(s, \omega)$ , then  $\mathcal{T}_{bcb}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  is dense in  $\Pi_{bc}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$  for the topology of weak convergence of probability measures.*

**PROOF.** We repeat the proof of Proposition 4.4 with the use of Proposition 4.2 replaced by Proposition 4.3. We only need to check that if  $\hat{T} \in \mathcal{T}_{bcb}(\mathbb{W}^d, \mathbb{W}^d)$ , then  $T := F^{\bar{\sigma}, \bar{b}} \circ \hat{T} \circ R_z^{\sigma, b} \in \mathcal{T}_{bcb}(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}})$ . Let  $\hat{S} : \mathbb{W}^d \rightarrow \mathbb{W}^d$  be such that  $\hat{S} \in \mathcal{T}_{bc}(\mathbb{W}^d, \mathbb{W}^d)$  and  $\mathbb{W}^d$ -a.s.  $\hat{T} \circ \hat{S} = \hat{S} \circ \hat{T} = X$ . It is straightforward to verify that  $S := F^{\sigma, b} \circ \hat{S} \circ R_z^{\bar{\sigma}, \bar{b}}$  satisfies  $T \circ S = X$ .  $\mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}$ -a.s. and  $S \circ T = X$ .  $\mu_z^{\sigma, b}$ -a.s. as required.  $\square$

**4.3. Equivalence of bicausal Monge and Kantorovich optimal transport.** Given some cost  $c : \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$ , we define

1. the bicausal Monge transport problem:

$$(BMP) \quad M(\eta, \nu) := \inf_{\pi \in \mathcal{T}_{bc}(\eta, \nu)} \int_{\mathcal{W}^d \times \mathcal{W}^d} c(\omega, \omega') d\pi(\omega, \omega') = \inf_{\pi \in \mathcal{T}_{bc}(\eta, \nu)} \mathbb{E}^{\pi}(c).$$

2. the bicausal Kantorovich transport problem:

$$(BKP) \quad K(\eta, \nu) := \inf_{\pi \in \Pi_{bc}(\eta, \nu)} \int_{\mathcal{W}^d \times \mathcal{W}^d} c(\omega, \omega') d\pi(\omega, \omega') = \inf_{\pi \in \Pi_{bc}(\eta, \nu)} \mathbb{E}^{\pi}(c),$$

Since  $\mathcal{T}_{bc}(\eta, \nu) \subseteq \Pi_{bc}(\eta, \nu)$ , we have  $K(\eta, \nu) \leq M(\eta, \nu)$ . We can derive sufficient conditions for equality in light of Proposition 4.4:

**COROLLARY 4.2.** *Under the assumptions of Proposition 4.4, if  $c : \mathcal{W}^d \times \mathcal{W}^d \rightarrow \mathbb{R}_{\geq 0}$  is continuous and  $|c(\omega, \omega')| \leq |\phi(\omega)| + |\psi(\omega')|$  for some  $\phi \in L^1(\mu_z^{\sigma, b}, \mathcal{F}_1)$  and  $\psi \in L^1(\mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}, \mathcal{F}_1)$ , then the bicausal Monge and Kantorovich problems have the same value:*

$$K(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}) = M(\mu_z^{\sigma, b}, \mu_{\bar{z}}^{\bar{\sigma}, \bar{b}}).$$

This proof of this corollary is an adaptation of [30, Lemma 5.14] with minor changes, and is, hence, omitted.

**4.4. Examples.** A striking result, due to Lassalle [28], states that when  $c$  is the Cameron-Martin norm, the value of (BKP)-(BMP) may be computed explicitly and is given by the relative entropy.

We derive in this section other examples of explicit solutions to (BKP)-(BMP), in the case where one of the probability measures is the law of an SDE which has invertible diffusion coefficient and the other is the law of an SDE which has a deterministic drift coefficient and almost deterministic diffusion coefficient in a sense to be made precise later. In particular, we highlight how one cannot expect uniqueness of the optimal bicausal transport.

When  $\eta$  and  $\nu$  are weak solutions of SDEs, for any  $\pi \in \Pi_{bc}(\eta, \nu)$ , marginal processes  $(X_t)_{t \in [0, 1]}$  and  $(Y_t)_{t \in [0, 1]}$  are  $(\pi, (\mathcal{P}_t^\pi)_{t \in [0, 1]})$ -semimartingales by Theorem 2.2. Inspired by [5, Proposition 3.3], we will consider costs that have separate dependence on the martingale and finite variation parts of  $X$  and  $Y$ . Writing

$$X = M^\eta + V^\eta, \quad Y = M^\nu + V^\nu$$

the semimartingale decomposition of  $X$  (resp.  $Y$ ) into its martingale and finite variation parts under  $\eta$  (resp.  $\nu$ ) we consider a transport cost  $c$  given by

$$(16) \quad c(X, Y) := h(V^\eta - V^\nu) + g(\text{tr}([M^\eta - M^\nu]_1)),$$

with  $h : \mathcal{W}^d \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$  measurable and  $g : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0} \cup \{+\infty\}$  monotone increasing.

**PROPOSITION 4.6.** *Let  $\sigma \in \mathcal{A}^{d, d}$ ,  $b \in \mathcal{A}^{d, 1}$  be such that the SDE*

$$(17) \quad dX_t = b(t, X)dt + \sigma(t, X)dB_t, \quad X_0 = 0,$$

*has a unique weak solution  $\eta := \mu_0^{\sigma, b}$ , where  $\sigma(t, \omega)$  is invertible for all  $(t, \omega)$ .*

*Let  $\bar{\sigma} \in \mathcal{A}^{d, d}$  be such that  $\bar{\sigma}(t, \omega) = \kappa(t)u(t, \omega)$  with  $u(t, \omega) \in \mathcal{O}^d$  and assume the SDE*

$$(18) \quad dY_t = \bar{b}(t)dt + \bar{\sigma}(t, Y)dB_t, \quad Y_0 = 0,$$

*has a unique, strong solution with law  $\nu := \mu_0^{\bar{\sigma}, \bar{b}}$ . Denote by  $F_0^{\bar{\sigma}, \bar{b}} : \mathcal{W}^d \rightarrow \mathcal{W}^d$  the corresponding Itô map. Then an optimal transport from  $\eta$  to  $\nu$  for a cost of the form (16) is given by the bicausal Monge map*

$$\mathcal{W}^d \ni \omega \mapsto T(\omega) := F_0^{\sqrt{\bar{\sigma}\bar{\sigma}^*}, \bar{b}} \left( \int_0^\cdot Q_s \sigma^{-1}(s, X) dM^\eta(s) \right)$$

*where  $Q : [0, 1] \times \mathcal{W}^d \rightarrow \mathcal{O}^d$  is a  $(\mathcal{H}_t^\eta)_{t \in [0, 1]}$ -progressively measurable process given by  $Q_t(\omega) = Q(t, \omega) := V(t, \omega)U^*(t, \omega)$  for  $U(t, \omega), V(t, \omega) \in \mathcal{O}^d$  appearing in the SVD of matrix  $\sigma(s, \omega)\sqrt{\bar{\sigma}\bar{\sigma}^*}(s)$ . In particular,*

$$K(\eta, \nu) = M(\eta, \nu) = \mathbb{E}^\eta \left( h \left( \int_0^\cdot b(s, X) ds - \int_0^\cdot \bar{b}(s) ds \right) + \right. \\ \left. \mathbb{E}^\eta \left( g \left( \int_0^1 \text{tr}(\sigma(s, X)\sigma^*(s, X) + \bar{\sigma}\bar{\sigma}^*(s)) ds - 2 \int_0^1 \text{tr}(\sigma(s, X)\sqrt{\bar{\sigma}\bar{\sigma}^*}(s)Q(s, X)) ds \right) \right) \right).$$

REMARK 7. As the singular value decomposition is non-unique, the solution to (BMP)-(BKP) is also in general non-unique. Furthermore, the assumptions of Proposition 4.6 assumes no regularity on the drift/diffusion coefficients and still concludes the equivalence of the bicausal Kantorovich and Monge problem, which is not covered by Corollary 4.2. In the case  $d = 1$ ,  $\sigma, \bar{\sigma} \geq 0$ ,  $h(\omega) := \|\omega\|_{1\text{-var}}$  and  $g(x) := x$ , the preceding result agrees with that of [5, Proposition 3.3].

PROOF. By rotation invariance of Brownian motion and weak uniqueness, we can without loss of generality replace  $\bar{\sigma}(t, \omega)$  by the time-dependent deterministic function  $\xi : [0, 1] \rightarrow \mathbb{R}^{d \times d}$ ,  $\xi(t) := \sqrt{\bar{\sigma}\bar{\sigma}^*}(t) = \sqrt{\kappa\kappa^*}(t)$ .

Let  $\pi \in \Pi_{bc}(\eta, \nu)$ . By Theorem 3.2, there exists some stochastic basis  $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t \in [0, 1]}, \mathbb{P})$  and  $(\mathcal{G}_t)$ -adapted processes  $B, \bar{B}, Z$  and  $\bar{Z}$  satisfying Theorem 3.2. Then

$$\begin{aligned} \mathbb{E}^\pi(c) &= \mathbb{E}^\mathbb{P}(c(X - Y)) = \mathbb{E}^\mathbb{P}(h(\int_0^\cdot b(s, Z)ds - \int_0^\cdot \bar{b}(s)ds) + \\ &\quad \mathbb{E}^\mathbb{P}(g(\text{tr}(\left[\int_0^\cdot \sigma(s, Z)dB_s - \int_0^\cdot \xi(s)d\bar{B}_s\right]_1))) \end{aligned}$$

Setting  $\rho_t = d[B, \bar{B}]/dt \in \mathcal{C}^d$  as in Theorem 3.2, we have

$$\begin{aligned} &\text{tr}\left(\left[\int_0^\cdot \sigma(s, Z)dB_s - \int_0^\cdot \xi(s)d\bar{B}_s\right]_1\right) \\ &= \int_0^1 \text{tr}(\sigma(s, Z)\sigma^*(s, Z) + \bar{\sigma}\bar{\sigma}^*(s))ds - 2 \int_0^1 \text{tr}(\sigma(s, Z)\xi(s)\rho_s)ds \\ &\geq \int_0^1 \text{tr}(\sigma(s, Z)\sigma^*(s, Z) + \bar{\sigma}\bar{\sigma}^*(s))ds - 2 \int_0^1 \sup_{C \in \mathcal{C}^d} \text{tr}(\sigma(s, Z)\xi(s)C)ds \\ &= \int_0^1 \text{tr}(\sigma(s, Z)\sigma^*(s, Z) + \bar{\sigma}\bar{\sigma}^*(s))ds - 2 \int_0^1 \sup_{C \in \mathcal{O}^d} \text{tr}(\sigma(s, Z)\xi(s)C)ds, \end{aligned}$$

where we have used [10, Proposition 4.7] in the last equality. By the same proposition,

$$\sup_{C \in \mathcal{O}^d} \text{tr}(\sigma(s, Z)\xi(s)C) = \text{tr}(\sigma(s, Z)\xi(s)Q(s, Z)),$$

where  $Q$  is defined as in the statement of the proposition. Hence, by monotonicity of  $g$  and temporarily suppressing the arguments of  $\sigma$  and  $b$ ,

$$\begin{aligned} (19) \quad \mathbb{E}^\pi(c) &\geq \mathbb{E}^\eta(h(\int_0^\cdot bds - \int_0^\cdot \bar{b}(s)ds)) \\ &\quad + \mathbb{E}^\eta(g(\int_0^1 \text{tr}(\sigma\sigma^* + \bar{\sigma}\bar{\sigma}^*(s))ds - 2 \int_0^1 \text{tr}(\sigma\xi(s)Q(s, X))ds)). \end{aligned}$$

Since the RHS of (19) no longer depends on  $\pi$ , taking infimum over  $\pi \in \Pi_{bc}(\eta, \nu)$  gives that (BKP) is lower bounded by the RHS of (19). One easily checks that this lower bound is achieved by  $T$  as defined in the statement of the proposition, which is a bicausal Monge map by Theorem 3.3.  $\square$

**One dimensional case** The bicausal Kantorovich problem for laws of one-dimensional SDEs was studied in [7, 33] for a variety of costs and regularity conditions on  $\sigma, b, \bar{\sigma}$  and  $\bar{b}$ . For example, if  $b, \bar{b} : \mathbb{R} \rightarrow \mathbb{R}$  and  $\sigma, \bar{\sigma} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  are Lipschitz-continuous, then by [7, Theorem 1.3], for any  $p \geq 1$  and  $c(\omega, \omega') := \int_0^1 |\omega_s - \omega'_s|^p ds$ , an explicit solution to

$$(20) \quad \inf_{\pi \in \Pi_{bc}(\mu_0^{\sigma, b}, \mu_0^{\bar{\sigma}, \bar{b}})} \mathbb{E}^\pi(c)$$

is given by the so-called *synchronous coupling*  $\pi^* := (F^{\sigma, b}, F^{\bar{\sigma}, \bar{b}})_\# \mathbb{W}^d$ . In other words, on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with Brownian motion  $B = (B_t)_{t \in [0, 1]}$ , we construct solutions to (5) and (6) driven by the *common* noise  $B$  as  $Z = (Z_t)_{t \in [0, 1]} := F^{\sigma, b}(B)$  and  $\bar{Z} = (\bar{Z}_t)_{t \in [0, 1]} = F^{\bar{\sigma}, \bar{b}}(B)$ . Then,

$$\mathbb{E}^\mathbb{P} c(Z, \bar{Z}) = \mathbb{E}^\mathbb{P} \int_0^1 |Z_s - \bar{Z}_s|^p ds = \inf_{\pi \in \Pi_{bc}(\mu_0^{\sigma, b}, \mu_0^{\bar{\sigma}, \bar{b}})} \mathbb{E}^\pi(c).$$

If  $\sigma(x) > 0$  for all  $x \in \mathbb{R}$ , then by Corollary 3.4, we see that the synchronous coupling is induced by a Monge map  $F^{\bar{\sigma}, \bar{b}} \circ R^{\sigma, b}$ ,

$$\pi^* = (\text{Id}, F^{\bar{\sigma}, \bar{b}} \circ R^{\sigma, b})_\# \mu_0^{\sigma, b}$$

and (BKP) = (BMP) in agreement with Corollary 4.2. The filtration used to define bicausal couplings in [7] is the canonical one  $(\mathcal{F}_t)_{t \in [0, 1]}$ , instead of the right-continuous one used in this work,  $(\mathcal{H}_t)_{t \in [0, 1]}$ . However, the Lipschitz assumption in this example implies that  $\mu_0^{\sigma, b}$  and  $\mu_0^{\bar{\sigma}, \bar{b}}$  are laws of strong Markov processes and these two definitions of bicausal couplings coincide (see [7, Remark 2.3] or Remark 1 ii).

**Acknowledgments.** The authors would like to thank Yifan Jiang, Akshay Hegde, Ioannis Karatzas and Daniel Goodair for fruitful discussions and helpful comments.

## REFERENCES

- [1] ACCIAIO, B., BACKHOFF, J. and ZALASHKO, A. (2016). Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stochastic Processes and their Applications* **130** 2918–2953. <https://doi.org/10.1016/j.spa.2019.08.009>
- [2] AMBROSIO, L., BRUÉ, E. and SEMOLA, D. (2021). *Lectures on optimal transport*. Springer.
- [3] AZOFF, E. A. (1974). Borel measurability in linear algebra. *Proceedings of the American Mathematical Society* **42** 346–350.
- [4] BACKHOFF, J., BARTL, D., BEIGLBÖCK, M. and WIESEL, J. (2022). Estimating processes in adapted Wasserstein distance. *The Annals of Applied Probability* **32** 529–550.
- [5] BACKHOFF, J., BARTL, D., BEIGLBÖCK, M. and EDER, M. (2020). Adapted Wasserstein distances and stability in mathematical finance. *Finance and Stochastics* **24** 601–632. <https://doi.org/10.1007/s00780-020-00426-3>
- [6] BACKHOFF-VERAGUAS, J., BARTL, D., BEIGLBÖCK, M. and EDER, M. (2020). All adapted topologies are equal. *Probability Theory and Related Fields* **178** 1125–1172.
- [7] BACKHOFF-VERAGUAS, J., KÄLLBLAD, S. and ROBINSON, B. A. (2025). Adapted Wasserstein distance between the laws of SDEs. *Stochastic Processes and their Applications* **189** 104689. <https://doi.org/10.1016/j.spa.2025.104689>
- [8] BARTL, D., BEIGLBÖCK, M. and PAMMER, G. (2025). The Wasserstein space of stochastic processes. *Journal of the European Mathematical Society*. <https://doi.org/10.4171/JEMS/1554>
- [9] BEIGLBÖCK, M. and LACKER, D. (2018). Denseness of adapted processes among causal couplings. *arXiv preprint arXiv:1805.03185*.
- [10] BION-NADAL, J. and TALAY, D. (2019). On a Wasserstein-type distance between solutions to stochastic differential equations. *The Annals of Applied Probability* **29** 1609–1639.



- [11] BRÉMAUD, P. and YOR, M. (1978). Changes of filtrations and of probability measures. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **45** 269–295.
- [12] CHERNY, A. S. (2002). On the uniqueness in law and the pathwise uniqueness for stochastic differential equations. *Theory of Probability & Its Applications* **46** 406–419.
- [13] CONT, R. (2016). Functional Ito Calculus and functional Kolmogorov equations. In *Stochastic Integration by Parts and Functional Ito Calculus (Lecture Notes of the Barcelona Summer School in Stochastic Analysis, July 2012)* (F. Utzet and J. Vives, eds.) 115–208. Birkhauser Basel.
- [14] CONT, R. and FOURNIÉ, D.-A. (2010). Change of variable formulas for non-anticipative functionals on path space. *J. Funct. Anal.* **259** 1043–1072. <https://doi.org/10.1016/j.jfa.2010.04.017> MR2652181
- [15] DUGUNDJI, J. (1951). An extension of Tietze's theorem. *Pacific Journal of Mathematics* **1** 353 – 367.
- [16] ECKSTEIN, S. and PAMMER, G. (2024). Computational methods for adapted optimal transport. *The Annals of Applied Probability* **34** 675–713.
- [17] EDER, M. (2019). Compactness in adapted weak topologies. *arXiv preprint arXiv:1905.00856*.
- [18] ÉMERY, M. (2005). On certain almost Brownian filtrations. In *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **41** 285–305. Elsevier.
- [19] FANG, S. and NOLOT, V. (2015). Optimal transport maps on infinite dimensional spaces. *Frontiers of Mathematics in China* **10** 715–732.
- [20] FEYEL, D. and ÜSTÜNEL, A. S. (2004). Monge-Kantorovitch measure transportation and Monge-Ampere equation on Wiener space. *Probability theory and related fields* **128** 347–385.
- [21] FLIESS, M. (1981). Fonctionnelles causales non linéaires et indéterminées non commutatives. *Bull. Soc. Math. France* **109** 3–40. MR613847
- [22] FÖLLMER, H. (2022). Optimal Couplings on Wiener Space and An Extension of Talagrand's Transport Inequality. In *Stochastic Analysis, Filtering, and Stochastic Optimization: A Commemorative Volume to Honor Mark H. A. Davis's Contributions* 147–175. Springer, Cham. [https://doi.org/10.1007/978-3-030-98519-6\\_7](https://doi.org/10.1007/978-3-030-98519-6_7)
- [23] GOLUB, G. H. and VAN LOAN, C. F. (2013). *Matrix computations*. Johns Hopkins University Press.
- [24] HE, S., WANG, J. and YAN, J. A. (1992). *Semimartingale Theory and Stochastic Calculus*. Science Press.
- [25] IKEDA, N. and WATANABE, S. (1981). *Stochastic Differential Equations and Diffusion Processes*. Elsevier.
- [26] KALLENBERG, O. (1997). *Foundations of modern probability*. Springer.
- [27] KECHRIS, A. (2012). *Classical descriptive set theory*. Graduate Texts in Mathematics **156**. Springer.
- [28] LASSALLE, R. (2018). Causal transport plans and their Monge–Kantorovich problems. *Stochastic Analysis and Applications* **36** 1–33. <https://doi.org/10.1080/07362994.2017.1422747>
- [29] LIPTSER, R. S. and SHIRYAEV, A. N. (1977). *Statistics of random processes: General theory* **394**. Springer.
- [30] NUTZ, M. (2022). Introduction to entropic optimal transport. *Lecture notes, Columbia University*.
- [31] PEYRÉ, G. and CUTURI, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning* **11** 355–607.
- [32] RACHEV, S. T. and RÜSCHENDORF, L. (2006). *Mass transportation problems: Applications*. Springer.
- [33] ROBINSON, B. A. and SZÖLGYENYI, M. (2024). Bicausal optimal transport for SDEs with irregular coefficients.
- [34] ROGERS, L. C. G. and WILLIAMS, D. (2000). *Diffusions, Markov Processes, and Martingales: Volume 1, Foundations*. Cambridge University Press.
- [35] RÜSCHENDORF, L. (1985). The Wasserstein distance and approximation theorems. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **70** 117–129. <https://doi.org/10.1007/BF00532240>
- [36] ÜSTÜNEL, A. S. (2019). Martingale representation for degenerate diffusions. *Journal of Functional Analysis* **276** 3468–3483.
- [37] VILLANI, C. (2008). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg.
- [38] XU, T., WENLIANG, L. K., MUNN, M. and ACCIAIO, B. (2020). COT-GAN: Generating sequential data via causal optimal transport. *Advances in neural information processing systems* **33** 8798–8809.