universität wien

DISSERTATION/DOCTORAL THESIS

Titel der Dissertation / Title of the Doctoral Thesis

"Causal optimal transport: theory and applications"

verfasst von / submitted by

Anastasiia Zalashko, MSc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 796 605 405
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Mathematik
Betreut von / Supervisor:	o. Univ.-Prof. Dr. Walter Schachermayer
Mitbetreut von / Co-Supervisors:	Dr. Beatrice Acciaio Dr. Julio Daniel Backhoff Veraguas

Abstract

This thesis studies the causal optimal transport problem between laws of stochastic processes. Causality here means to add a physically relevant time constraint into the classical optimal transport problem restricting the admissible set of transport plans to those that perform the allocation of mass in an adapted (nonanticipative) way. Intuitively, one can say that the relationship between causal plans and adapted processes is the same as that between transport plans (Kantorovich) and classical transport maps (Monge).

The main aim of the thesis is to show the full theoretical potential and the modelling power of causal optimal transport problems by analysing their structure and by highlighting their significant features. After giving an outlook of classical optimal transport, the study starts with the finite-dimensional/discrete-time case, tackling the questions of duality and characterization of optimizers. Under appropriate assumptions, the causal analogue to the Brenier map of the classical case is identified as the so-called Knothe-Rosenblatt rearrangement. Furthermore, a connection between transport-information inequalities and stochastic optimization is presented. As the next step, suitable extensions of the discrete-time results to the infinite-dimensional/continuous-time setup are obtained. Most importantly, the projections/limits arguments enable to recover the equality between causal transports with Cameron-Martin cost and the relative entropy for a specific class of processes.

Varying the underlying filtrations in the definition of causality allowed to provide a novel point of view on various problems related to enlargement of filtrations and Girsanov theory in stochastic analysis. In particular, necessary and sufficient conditions for a Brownian motion to remain a semimartingale in an enlarged filtration, in terms of certain minimization problems over sets of causal transport plans, are given. As an application of these ideas, those are further used for estimation of the value of having additional information, for some classical stochastic optimization problems.

Zusammenfassung

Diese Dissertation beschäftigt sich mit kausalem optimalen Transport zwischen Verteilungen von stochastischen Prozessen. Die Kausalität bedeutet hier, eine physische relevante Zeiteinschränkung in das klassische optimale Transportproblem einzufügen, das die zulässige Menge von Transportplänen auf diejenigen beschränkt, die die Zuteilung von Masse in einer adaptierten (nicht antizipativen) Weise durchführen. Intuitiv kann man sagen, dass die Beziehung zwischen Kausalplänen und adaptierten Prozessen die gleiche ist wie zwischen Transportplänen (Kantorowitsch) und klassischen Transportabbildungen (Monge).

Das Hauptziel der Dissertation ist, das volle Potential in Theorie sowie Modellierung kausaler optimaler Transportprobleme zu zeigen, indem ihre Struktur analysiert und ihre wesentlichen Merkmale hervorgehoben werden. Nach einem Abriss des klassischen optimalen Transport(problems) beginnt die Analyse mit dem endlich-dimensionalen/diskreten Zeitfall, um die Fragen der Dualität und Charakterisierung von Optimierern zu lösen. Unter angemessenen Voraussetzungen wird das kausale Analogon zu Brenier-Abbildung des klassischen Falles als sogenannte Knothe-Rosenblatt-Umordnung identifiziert. Darüber hinaus wird eine Verbindung zwischen Transport-Informations-Ungleichungen und stochastischer Optimierung präsentiert. Als nächster Schritt werden angemessene Fortsetzungen der diskreten Zeit-Ergebnisse zur unendlich-dimensionalen/kontinuierlichen Zeitaufstellung erhalten. Hierbei ist zentral, dass die Projektionen/Grenzwert-Argumente die Wiederherstellung der Gleichheit zwischen kausalen Transporten mit Cameron-Martin-Kosten und der relativen Entropie für eine bestimmte Klasse von Prozessen ermöglichen.

Variation der zugrundeliegenden Filtrationen in der Definition der Kausalität ermöglicht es, eine neue Sichtweise über verschiedene Probleme im Zusammenhang mit der Vergrößerung der Filtrationen und der Girsanov-Theorie in der stochastischen Analyse zu bilden. Insbesondere sind notwendige und hinreichende Bedingungen für eine Brownsche Bewegung als ein Semimartingal in einer vergrößerten Filtration erhalten zu bleiben, in Bezug auf bestimmte Minimierungsprobleme über Mengen von kausalen Transportplänen gegeben. Als Anwendung der oben genannten Ideen dient die Veranlagung des Wertes zusätzlicher Informationen in einigen klassischen stochastische Optimierungsprobleme.

Acknowledgements

I was so lucky to meet all the wonderful people during my PhD journey, who helped me to achieve this goal.

First and foremost, I would like to express my special gratitude and thanks to my advisor Walter Schachermayer for giving me this chance of being his PhD student, for the freedom in doing research, for his support and patience. I also gratefully acknowledge the financial support by Doktoratskolleg W1245 "Dissipation and Dispersion in Nonlinear PDEs" of the Austrian Science Fund (FWF).

I would like to sincerely thank Mathias Beiglböck, who initiated the project on the causal optimal transport.

I am grateful to my co-advisor Beatrice Acciaio for encouraging my research, for the constant support and for making my stay at LSE enjoyable and productive.

I am very thankful to Martin Huesmann for all the insightful discussions during my research visit to the University of Bonn.

Most of all, I owe special thanks to Julio Backhoff for all his constant help and support, for sharing his knowledge with me, for guiding me and dedicating me so much of his time.

Thanks a lot to my colleagues for creating such an enjoyable and inspiring atmosphere: Jiří Černý, Christa Cuchiero, Lingqi Gu, Astrid Kollros, Yiqing Lin, Radka Picková, Mathias Pohl, Alexander Ristig, Cong Shi, Oleg Szehr, Ludovic Tangpi and Junjian Yang. Also I would like to thank Stefanie Hirsch for giving me the most valuable piece of advice at the beginning of my PhD.

Finally, I would like to thank Alexander for all his never-ending love, support and encouragement! It would not have been possible without you!

Всему хорошему, что у меня есть, я обязана свои родителям! С бесконечной любовью и благодарностью, я посвящаю свою диссертацию вам, мои самые дорогие и любимые, Мамочка и Папочка!

Contents

Introduction

From the seminal works of Monge [Mon84] and Kantorovich [Kan42], the theory of optimal transport has widely developed and established itself as an active research area, with growing applications in the most various areas of sciences and engineering. The central problem of optimal transport consists in finding transport plans that perform the allocation of a given source measure to a target one in an optimal way, where optimality is measured against a given cost function on the product space. For a detailed introduction to the topic we refer the reader to the following books [RR98, Vil03, Vil08, San15] and the references therein.

In this thesis we investigate causal optimal transport problems between two given probability measures interpreted as laws of stochastic processes. The essential difference between classical and causal optimal transport lies in the fact that time is introduced in the latter. As a consequence, the set of admissible transport plans is restricted to those that satisfy a particular non-anticipativity constraint. A transport plan between two processes is said to be *causal* if, from an observed trajectory of the first process, the "mass" can be split at each moment of time into the second process only based on the information available up to that time. It is illustrative to think of the deterministic case (i.e. when there is no splitting of mass); such a causal plan is then an actual mapping which is further adapted, and so the relationship between causal plans and adapted processes is the same as between classical transport plans (Kantorovich) and transport maps (Monge).

Mathematically, in order to define a causal transport problem, one needs to be given two *filtered* probability spaces $(\mathcal{X}, \{\mathcal{F}_t\}_{t=0}^T, \mu)$ and $(\mathcal{Y}, \{\mathcal{G}_t\}_{t=0}^T, \nu)$. A transport plan π (or a coupling), which is a probability measure on $\mathcal{X} \times \mathcal{Y}$ having the prescribed marginals μ, ν, is further called *causal* if the regular conditional kernel of π w.r.t. the first coordinate is measurable w.r.t. the filtration $\{\mathcal{F}_t\}$. In other words, the amount of "mass" transported by π to a subset of the target space $\mathcal{Y}$ belonging to $\mathcal{G}_t$, is solely determined by the information contained in $\mathcal{F}_t$. From this we clearly see the decisive role of filtrations in the definition of causality.

The idea of imposing a "causality" constraint on a transport plan between laws of stochastic processes is actually not new. From the earliest references on the topic, this definition can be found in the lines of the proof of the Yamada-Watanabe criterion for stochastic differential equations [YW71]. Later in [Jac80], distinguishing causality as a special property of joint solution measures (joint laws of solutions of stochastic differential equations), the author explored the connection between the questions of existence/uniqueness of weak/strong solutions of these equations and properties of the set of those measures. Kurtz in [Kur07, Kur14] continued this line of research, using the name "compatibility" for this special type of constraint.

A systematic treatment and use of causality as an interesting property of abstract transport plans and their associated optimal transport problems was first made by Lassalle in [Las13] in the general context of Polish spaces (then updated in [Las15]). As an application, the author considered the Wiener space setting of the problem and established that weak solutions to Brownian motion-driven stochastic differential equations can be conceived as causal transport plans between the Wiener measure and a target measure, and found that such plans are automatically *bicausal*, and optimal for a Cameron-Martin-type cost. The notion of bicausality is being the symmetrized analogue of causality. We stress that the main motivation for our work comes from this connection with stochastic analysis, which we considerably enhance. Our larger goal is to deepen the theoretical understanding of causal transport problems, as well as to explore their modelling potential by looking first at the discrete-time setup, and then extending the obtained results to the infinite-dimensional setting.

After giving an outlook of classical optimal transport, we start the thesis tackling the questions of duality and characterization of optimizers. The causality property can be characterized as a linear constraint on transport plans, which naturally leads to the formulation of a dual problem to causal optimal transport. In a finite-dimensional/discrete-time case, when working with canonical filtrations, we can embed this problem in the class of optimal transport problems under (infinitely many) linear constraints, as considered by Beiglböck, Griessler [BG14] and Zaev [Zae15]. However, in the general case of different filtrations, we cannot invoke the existing results in order to prove that the values of the primal and dual problems coincide. For the conditions on the existence of an optimal causal transport, a dual formulation, and the proof of no-duality-gap, we refer to Theorems 2.1.10 and 4.1.6 in discrete and continuous-time setting respectively.

By studying the conditional distributions of causal transports in a discrete-time set-up, we are able to tackle many instances of the causal optimal transport problem by means of a recursion, which we call the Dynamic Programming Principle; see Theorem 2.2.2. The appeal of these recursions is that instead of one "multi-dimensional" transport problem over causal plans, we obtain recursively several "one-dimensional" problems, each one of them a "general" (i.e. non-linear) transport problem as introduced recently in [GRST15]. In this way, we reduce the dimensionality of the problem at the expense of introducing non-linearities. This is a distinguishing feature which is absent in classical optimal transport.

Causal optimal transport is related to the notion of *nested distance*, whose systematic investigation was initiated by Pflug [Pfl09] and Pflug, Pichler [PP12, PP14, PP15], and had a precursor in the "Markov-constructions" studied by Rüschendorf [Rüs85]. Roughly speaking, the nested distance is defined through a problem of optimal transport over plans which are *bicausal*. Moreover, in [PP12] the authors applied these considerations to the practical problem of reducing the complexity of multistage stochastic programs, by showing that the difference between the optimal value of a program w.r.t. two different noise distributions is dominated by the nested distance between them. These contributions were the first indicators of the potential of causal transport for mathematical modelling and applications.

We deepen the understanding of the relation between nested distance and multistage stochastic programming by showing that the discrepancy between such stochastic programs can be simply gauged by an entropy, which is easier to compute in practice than the nested distance itself. We also highlight the connection between causality and functional inequalities when, in Section 2.5, we establish Talagrand's celebrated inequality (see [Tal96]) for the standard Gaussian measure by interpreting the author's "tensorization/inductive trick" as an instance of our aforementioned recursions; we refer the reader to [Led01, GL10] for an account on functional/geometric inequalities and the related concept of concentration of measure.

Optimal transport has gained much momentum since Brenier's contribution [Bre91] on uniqueness and identification of optimal transport. Along these lines, we are able to identify a causal discrete-time optimizer in a setting that is relevant to stochastic analysis. In Theorem 2.2.13 we establish that the Knothe-Rosenblatt rearrangement [Kno57] (also known as multidimensional quantile transform/ increasing triangular transformation) is causal optimal for the squared euclidean distance (or more generally convex and separable in the sense of (2.2.21) below) if the source measure is of product type. If further the source measure has absolutely continuous marginals, then the Knothe-Rosenblatt rearrangement is of Monge-type. Hence, it can be viewed as a causal version of the classical Brenier map. In our opinion this result adds to the appeal of this rearrangement, which is in any case widely used in analysis, statistics, and operations research in the context of scenario generation.

Coming from our motivation, a large part of the thesis is devoted to the study of continuous-time causal transport problems on Wiener space. A peculiar feature of these infinite-dimensional transport problems is that their time-discretized versions are very explicit, with the source measure having independent increments, and the cost function being of a separable structure in terms of those increments; for the related discussion see Section 3.5. In view of this fact, the Knothe-Rosenblatt rearrangement appears to be an optimizer for the corresponding finite-dimensional projection of a causal transport problem on the Wiener space with the Cameron-Martin cost function. Exploiting further the time-discretization arguments allows us to provide a proof of Talagrand's inequality for the Wiener measure, first obtained by Feyel and Üstünel in [FÜ04] (see Lemma 3.7.3). Moreover, this approach enables us to recover the equality between the causal transport problems with Cameron-Martin cost, and the relative entropy, for a specific class of processes given in Proposition 3.7.4.

Due to the connection between solutions of Brownian motion-driven stochastic differential equations and optimizers of causal transport problems on Wiener spaces with the Cameron-Martin cost, it is a question of special interest to find the conditions under which optimal transport plans are generated by maps, i.e. of Monge form. In particular, this would yield the existence of strong solutions of those equations. However, recognizing the discrete-time causal Monge optimizer is not enough to solve the problem, since there is no guarantee that it will remain a transport map in the limit. The advantage of the Knothe-Rosenblatt rearrangement, as we show in Section 3.8, is that it might be used to construct a feasible transport plan on a path space that mimics the value of the time-discretized transport problem. We think that the building blocks we provide in Chapter 3 of this thesis should lay down the foundations for an optimal transport proof of the existence of strong solutions to Brownian-driven stochastic differential equations.

Varying the filtrations in the definition of causality allows us to provide a novel point of view on various problems related to enlargement of filtrations and Girsanov theory in stochastic analysis, which is the content of Chapter 4. We recall that the central question of enlargement of filtrations is whether the semimartingale property is preserved when passing from a given filtration to a larger one; see [BY78, JY78, Jeu80, Jac85] for some of the earliest works on the subject.

We start from the observation that if a process B, which is a Brownian motion on some probability space, is a semimartingale with respect to a filtration $\mathcal{G}$ larger than its natural filtration, then its unique continuous semimartingale decomposition takes the form

$$dB_t = d\tilde{B}_t + dA_t. \tag{1}$$

Here $\tilde{B}$ denotes a $\mathcal{G}-$Brownian motion and A a continuous $\mathcal{G}$-adapted finite variation process. As it turns out to be, the joint law of $(\tilde{B}, B)$ is a causal transport plan on path space, when considering the canonical and an appropriate enlarged filtration (see Section 4.1 for the precise framework). Since $\tilde{B}$ is an anticipative but deterministic mapping of B, much as a Monge map in classical transport, one can say that causal transport plans correspond to the Kantorovich-type generalization of such anticipative mappings.

The main result of the thesis in this regard, is the characterization of the semimartingale preservation property in an enlarged filtration, for a process which is a Brownian motion in the original filtration. Necessary and sufficient conditions for this preservation property are given in terms of causal transport problems on continuous path space; see Theorem 4.4.2. In addition, we can give necessary and sufficient conditions not only for the semimartingale preservation property to hold, but also to ensure that the finite variation process in the semimartingale decomposition (1) is absolutely continuous; see Theorem 4.4.5.

In the absolutely continuous case described above, we further identify a non-linear dual problem which we can fully solve and relate to the original causal transport problem. This is novel even in the absence of anticipation/enlargements. Interestingly, this gives a different proof of the semimartingale preservation property, and is achieved through optimal transport and convex analysis techniques, without resorting to stochastic analysis arguments; see Theorem 4.5.2.

The final contribution of this thesis consists in showing how causal transport problems can be used in order to estimate the value of additional information (i.e. the value of having access to an information flow described by a finer filtration). The latter is defined as the difference between the values of a given stochastic optimization problem when the optimization is run over a smaller “original” filtration, and when it is done over a finer “enlarged” one. It turns out that, for stochastic control of linear systems (concretely, utility maximization) and optimal stopping, this difference can be estimated by means of a causal transport problem in a robust way. We refer to [PK96] for the original motivation and [Pfl09, PP12] for a discrete-time approach related to ours.

The thesis is organized as follows.

Chapter 1 presents an outlook of classical optimal transport. In Section 1.1 we give a short introduction to the field recalling the basic facts needed throughout the thesis. In Section 1.2 we discuss transport problems with linear constraints. In Section 1.3 we provide an extension of classical duality/attainability results for a specific class of cost functions, that is further applied for linearly constrained transport problems.

Chapter 2 is based on the joint work with J. Backhoff, M. Beiglböck and Y. Lin [BBLZ16]. In Section 2.1 we introduce the discrete-time setting of the causal transport problem, give an equivalent characterization to test causality, and prove attainability and duality for the causal transport problem. In Section 2.2 we present the recursive formulation of the problem and identify the Knothe-Rosenblatt rearrangement as a causal optimizer. In Section 2.3 we examine further properties of the aforementioned rearrangement. Sections 2.4 and 2.5 are devoted to the discussion of some aspects of the bicausal transport together with connections to stochastic programming, providing the bicausal transport information inequality. And finally, in Section 2.6, we shortly discuss the anticipation of information in discrete time.

Chapter 3 starts with the introduction of causal transport problems in the infinite-dimensional setup (Section 3.1). After providing some relevant examples in Section 3.2, we review the existing literature related to transport on Wiener spaces (Section 3.3). In Section 3.4 we discuss the existing connection between optimal transport and stochastic differential equations. In Section 3.5 we show how causal transport problems on Wiener space with Cameron-Martin-type costs can be seen as the limiting objects of their finite-dimensional counterparts, via time-discretization. In Section 3.6 we present an approximation scheme for a class of stochastic control problems, which makes use of the knowledge hitherto obtained. The projections/limits arguments enable us to recover famous Talagrand's inequality and the equality between causal transports with Cameron-Martin cost and the relative entropy for a specific class of processes; see Section 3.7. At the end of this chapter, Section 3.8, we give one procedure for embedding a discrete-time optimizer into a feasible transport plan on path space. We close the chapter with a brief discussion on possible extensions and lines of research opened by the present work, see Section 3.9. Parts of this chapter benefited from discussions with J. Backhoff, M. Beiglböck and M. Huesmann.

Chapter 4 is based on the joint work with B. Acciaio and J. Backhoff [ABZ16]. We start with the introduction of the main concepts and ingredients of causal transport problems in the most general case with different filtrations and in infinite dimension, see Section 4.1. Here we also explain the connection with enlargements of filtrations, giving some examples in Section 4.2. In Section 4.3 we discuss the link between causality, optional and dual projections in stochastic analysis. Section 4.4 contains our main results on the semimartingale preservation property, along with some comments on the literature. Section 4.5 is devoted to a study of duality. Finally, in Section 4.6 we show how certain causal optimal transport problems might be used for the estimation of the value of having additional information in several continuous-time stochastic optimization problems.

Notation and conventions

For a Polish space $\mathcal{Z}$, we use $\mathcal{P}(\mathcal{Z})$ to denote the Borel probability measures on $\mathcal{Z}$, and endow it with the topology of weak convergence of measures. By $\mathcal{B}(\mathcal{Z})$ we denote the Borel σ-field on $\mathcal{Z}$, and for any σ-field $\mathcal{J} \subseteq \mathcal{B}(\mathcal{Z})$ we write $B(\mathcal{Z}, \mathcal{J})$ (resp. $B_b(\mathcal{Z}, \mathcal{J})$) for the set of all real-valued functions on $\mathcal{Z}$ that are measurable (resp. bounded measurable) w.r.t. $\mathcal{J}$. With $\mathcal{C}(\mathcal{Z})$ (resp. $\mathcal{C}_b(\mathcal{Z})$) we mean the set of all continuous (resp. bounded continuous) real-valued functions defined on $\mathcal{Z}$. Measurability with respect to a sigma algebra is often understood as equality to a correspondingly measurable function modulo a null set w.r.t. the measure unequivocally relevant to the given context.

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a measurable map $f : \Omega \to \mathcal{Z}$, $f_*\mathbb{P} \in \mathcal{P}(\mathcal{Z})$ denotes the push forward of $\mathbb{P}$ by f, which is by definition $f_*\mathbb{P}(A) = \mathbb{P}[f^{-1}(A)]$ for any set $A \in \mathcal{B}(\mathcal{Z})$.

For a product of sets $\mathcal{X} \times \mathcal{Y}$ we denote by p^1, p^2 the projection onto the first resp. second coordinate. We denote by π^x, π^y the regular kernels of a measure π on $\mathcal{X} \times \mathcal{Y}$ w.r.t. its first and second coordinate respectively. Analogous notation extends to products of more than two spaces. Therefore, $\int f(y)\pi^x(dy)$ gives a version of the conditional expectation of $f(y)$ given x under measure π, sometimes also denoted $E^\pi[f(Y)|X = x]$ in the literature, and so forth. However, for the conditioning w.r.t. a given σ-field $\mathcal{J} \subseteq \mathcal{B}(\mathcal{Z})$ we shall follow the accepted practice of denoting it by $\mathbb{E}^\pi[f|\mathcal{J}]$.

On $\mathbb{R}^N \times \mathbb{R}^N$ we denote by $(x_1, \dots, x_N)$ the first half and $(y_1, \dots, y_N)$ the second half of the coordinates, and we convene that for a probability π on $\mathbb{R}^N \times \mathbb{R}^N$ (respect. η on $\mathbb{R}^N$), $\pi^{x_1,\dots,x_t,y_1,\dots,y_t}$ (respect. $\eta^{x_1,\dots,x_t}$) denotes the two-dimensional measure on (x_{t+1}, y_{t+1}) (respect. one-dimensional measure on x_{t+1}) given by regular disintegration of π w.r.t. $(x_1, \dots, x_t, y_1, \dots, y_t)$ (respect. η w.r.t. $(x_1, \dots, x_t)$). Also, a statement like "for π-a.e. $x_1, \dots, x_t, y_1, \dots, y_t$" or "for η-a.e. $x_1, \dots, x_t$" is meant to denote respectively "almost-everywhere" with respect to the projections of π onto $x_1, \dots, x_t, y_1, \dots, y_t$ or η onto $x_1, \dots, x_t$.

If f and g are real-valued functions on $\mathcal{X}$ resp. $\mathcal{Y}$, we denote $f \oplus g(x, y) := f(x) + g(y)$. If a probability measure ν is absolutely continuous w.r.t. a measure μ, we denote it by $\nu \ll \mu$. The notation γ is reserved for the Wiener measure on the space of continuous functions on the interval $[0, T]$, denoted by $\mathcal{C} := \mathcal{C}[0, T]$. We write δ_x for the Dirac mass at point x. See p.121 for an exhaustive index of more specific notation.

Chapter 1

Optimal transport: an outlook

The aim of this chapter is to recall the most important results from optimal transport needed throughout the thesis, and to set up the notations. We start with Section 1.1 on classical transport, discussing existence of optimizers and duality in this setting. In Section 1.2 we proceed with the short review of the rather new topic of optimal transport under additional linear constraints. We finish this chapter with Section 1.3, where we give an extension of classical duality results to a specific class of cost functions. As an application, we further provide a duality result for linearly constrained transport problems for the same class of costs.

1.1 Classical optimal transport

Let $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mu)$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \nu)$ be two Polish probability spaces. We consider the *Monge-Kantorovich transport problem* for the given probability measures μ and ν, known from the pioneering works of Monge [Mon84] and Kantorovich [Kan42]. We denote by $\Pi(\mu, \nu) \subset \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ the set of those probability measures on $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{X}$-marginal μ and $\mathcal{Y}$-marginal ν, namely

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}), \text{ s.t. } p^1_* \pi = \mu \text{ and } p^2_* \pi = \nu\} .$$

The elements $\pi \in \Pi(\mu, \nu)$ are referred to as transport plans between the measures μ and ν. This set is never empty since the product measure $\mu \otimes \nu$ is an admissible transport plan. For any set $A \times B \in \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{Y})$, the value $\pi(A \times B)$ prescribes intuitively how much mass is moved from the set A to B. Sometimes this movement of mass can be described through a map $T : \mathcal{X} \to \mathcal{Y}$, meaning that from a single point $x \in \mathcal{X}$, the mass is transported to a prescribed destination $y := T(x)$. Those maps T that preserve the marginals μ and ν, i.e. $T_* \mu = \nu$, are called transport maps.

We will often consider pairs of random variables (X, Y) defined on some probability space $(\Omega, \mathbb{P})$ and taking values in resp. $\mathcal{X}$ and $\mathcal{Y}$, and refer to them as transport plans as well. Any property on (X, Y) should then be understood as a property on $(X, Y)_* \mathbb{P}$.

The classical optimal transport problem consists in minimizing the cost of transporting the (source) measure μ to the (target) measure ν, with respect to a given cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$. An application where the infinite value of the cost function might naturally occur is the transport on Wiener spaces with the Cameron-Martin cost function; see Chapter 3 for more on this.

The (primal) optimal transport problem is formulated then as follows

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y). \tag{OT}$$

The minimizers of this problem are called *optimal transport plans* between μ and ν, should they exist. If π is of the form $(\mathrm{id},T)_*\mu$ for some measurable map $T:\mathcal{X}\to\mathcal{Y}$ (i.e. no splitting of mass occurs), then the map T is called an *optimal transport map* from μ to ν. The transport problem (OT) is well defined for Borel measurable cost functions c. In most applications of optimal transport theory, the cost function c is considered to be lower semicontinuous.

The proof of the existence of optimal transport plans is usually based on two results, which we recall below. These are the lower semicontinuity of the cost functional (Lemma 1.1.3) and compactness of the set of all transport plans $\Pi(\mu,\nu)$, which is a consequence of tightness of the latter set (Lemma 1.1.1) combined together with Theorem 1.1.2. For the well known proofs of the results below we refer the reader to [Vil08] or [San15].

Lemma 1.1.1 (Tightness of transport plans)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ and $\mathcal{B}\subset\mathcal{P}(\mathcal{Y})$ be tight subsets of $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ respectively. Then the set $\Pi(\mathcal{A},\mathcal{B})$ of all transport plans whose marginals lie in $\mathcal{A}$ and $\mathcal{B}$ respectively, is itself tight in $\Pi(\mathcal{X}\times\mathcal{Y})$. In particular, $\Pi(\mu,\nu)$ is tight.*

Theorem 1.1.2 (Prokhorov's theorem)**.** *If $\mathcal{X}$ is a Polish space, then a set $\mathcal{A}\subset\mathcal{P}(\mathcal{X})$ is precompact for the weak topology if and only if it is tight, i.e. for any $\varepsilon>0$ there is a compact set K_ε such that $\mu(\mathcal{X}\setminus K_\varepsilon)\le\varepsilon$ for all $\mu\in\mathcal{A}$.*

Lemma 1.1.3 (Lower semicontinuity of the cost functional)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, and $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}\cup\{+\infty\}$ a lower semicontinuous cost function bounded from below. Let $(\pi_k)_{k\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{X}\times\mathcal{Y}$, converging weakly to some $\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$. Then*

$$\int_{\mathcal{X}\times\mathcal{Y}} c d\pi \le \liminf_{k\to\infty}\int_{\mathcal{X}\times\mathcal{Y}} c d\pi_k.$$

In particular, if c is nonnegative, then $\pi\mapsto\int c d\pi$ is lower semicontinuous on $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, equipped with the topology of weak convergence.

Theorem 1.1.4. *Let $(\mathcal{X},\mu)$ and $(\mathcal{Y},\nu)$ be two Polish probability spaces, let $a:\mathcal{X}\to\mathbb{R}\cup\{-\infty\}$ and $b:\mathcal{X}\to\mathbb{R}\cup\{-\infty\}$ be two upper semicontinuous functions such that $a\in L^1(\mu), b\in L^1(\nu)$. Let $c:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous cost function, such that $c(x,y)\ge a(x)+b(y)$ for all x,y. Then there exists an optimal transport plan for* (OT)*.*

The lower bound assumption on c guarantees that the expected cost $\int c(x,y)d\pi(x,y)$ is well-defined on $\mathbb{R}\cup\{+\infty\}$. In most applications in this thesis we may choose $a=b=0$.

We would like to remark that an optimal transport plan might fail to exist if the cost function does not satisfy proper continuity assumptions.

Example 1.1.5. [BGMS09, Example 1.3] Consider $\mathcal{X} = [0,1)$, $\mathcal{Y} = [1,2)$, with and μ, ν be corresponding restrictions of Lebesgue measure, and where the cost is $c(x,y) = (x-y)^2$. The cheapest way to do the transport is to shift every point by 1, i.e. the optimal transport map is $T(x) = x+1$. This results in an optimal transport cost of 1. If we now alter the cost function as follows

$$\tilde{c}(x,y) = \begin{cases} 2, & \text{if } y = x+1, \\ c(x,y), & \text{otherwise,} \end{cases}$$

it becomes impossible to find a transport plan $\pi \in \Pi(\mu,\nu)$ with total transport cost equal to 1, but it is still possible to achieve transport costs arbitrarily close to 1. (For instance, shift $[0, 1-\varepsilon)$ to $[1+\varepsilon, 2)$ and $[1-\varepsilon, 1)$ to $[1, 1+\varepsilon)$ for small $\varepsilon > 0$.)

The following result tells us about the convexity of the optimal cost; for the proof see [Vil08, Theorem 4.8]. We will use the below estimates in Sections 2.2.1 and 2.2.2.

Theorem 1.1.6. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function, and let $C(\mu,\nu)$ be the optimal transport cost functional on $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ associated with the given measures μ, ν, i.e. the value of* (OT). *Let (Θ, η) be a probability space, and let μ_θ, ν_θ, be two measurable functions defined on Θ, with values in $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ respectively. Assume that $c(x,y) \geq a(x) + b(y)$, where $a \in L^1(d\mu_\theta d\eta(\theta))$, $b \in L^1(d\nu_\theta d\eta(\theta))$. Then*

$$C\left(\int_\Theta \mu_\theta d\eta(\theta), \int_\Theta \nu_\theta d\eta(\theta)\right) \leq \int_\Theta C\left(\mu_\theta, \nu_\theta\right) d\eta(\theta).$$

Optimal transport problems have a rich duality theory. To formulate a dual problem, we define

$$\Psi := \left\{ (\phi,\psi) : \begin{array}{c} \phi : \mathcal{X} \to [-\infty,\infty) \ \mu-\text{integrable}, \ \psi : \mathcal{Y} \to [-\infty,\infty) \ \nu-\text{integrable}, \\ \phi(x) + \psi(y) \leq c(x,y) \text{ for all } (x,y) \in \mathcal{X} \times \mathcal{Y} \end{array} \right\}. \tag{1.1.1}$$

Then duality means the equivalence between (OT) and the following maximization problem

$$\sup_{(\phi,\psi)\in\Psi} \left\{ \int \phi d\mu + \int \psi d\nu \right\}. \tag{1.1.2}$$

The validity of the above duality was established by Kellerer [Kel84] for lower semicontinuous or just Borel measurable cost functions, bounded by the sum of two integrable functions. In [BLS12], duality results were extended to the class of Borel-measurable uniformly bounded cost functions, or more generally, $\mu \otimes \nu$-a.s. finitely valued. Moreover, the latter condition is sufficient for the existence of dual optimizers in (1.1.2), see [BS11, Theorem 2]. However, as Example 4.1 in [BS11] shows, in general, Monge-Kantorovich duality might fail for a measurable cost function. We recall this example below.

Example 1.1.7. Let $\mathcal{X} = \mathcal{Y} = [0,1]$, and μ, ν Lebesgue measure on $[0,1]$. We take the cost function c defined as

$$c(x,y) = \begin{cases} \infty, & x < y, \\ 1, & x = y, \\ 0, & x > y. \end{cases}$$

The optimal transport plan π is concentrated on the diagonal and gives an optimal cost equal to one. Assume that $\phi, \psi : [0,1] \to [-\infty, \infty)$ are integrable functions satisfying $\phi(x) + \psi(y) \leq c(x,y)$ for all $x, y \in [0,1]$. Then

$$\int \phi d\mu + \int \psi d\nu = \lim_{\alpha \downarrow 0} \int_0^{1-\alpha} (\phi(x+\alpha) + \psi(x))dx \leq \lim_{\alpha \downarrow 0} \int_0^{1-\alpha} c(x+\alpha, x)dx = 0.$$

Therefore, we see that there is a "duality gap", since the value of the primal problem is one, and that of the dual is zero.

1.2 Optimal transport with linear constraints

In this section we consider optimal transport with additional linear constraints, which is a rather young area of research in comparison with classical transport. Linearly constrained transport means that in the setting of classical transport one fixes some functional subspace $\mathbb{F}$, where admissible transport plans are assumed to vanish. Let $\mathbb{F} \subset \mathcal{B}_b(\mathcal{X} \times \mathcal{Y})$, then π is admissible if for any $F \in \mathbb{F}$ it holds that

$$\int F(x,y) d\pi(x,y) = 0.$$

Further, the set of constrained transport plans with given marginals will be denoted by

$$\Pi_{\mathbb{F}}(\mu, \nu) \subset \Pi(\mu, \nu).$$

The constrained transport problem then looks as follows. For a given cost function $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ we are minimizing the total expected cost among the restricted set of transport plans

$$\inf_{\pi \in \Pi_{\mathbb{F}}(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y) d\pi(x,y). \tag{COT}$$

For an introduction to the general problem and the existing results we refer the reader to the works of Beiglböck and Griessler [BG14] and Zaev [Zae15].

Different choices of the subspace $\mathbb{F}$ lead to the various properties that transport plans will satisfy. For example, if one takes $\mathcal{X} = \mathcal{Y} = \mathbb{R}$ and

$$\mathbb{F} = \{F(x)(y-x), \quad F \in \mathcal{C}_b(\mathbb{R})\},$$

it is easy to see that the set $\Pi_{\mathbb{F}}(\mu, \nu)$ corresponds to all transport plans between the given marginals that are two-step martingales. Martingale optimal transport is a very active area of research, and for the existing results see [BJ12, BHLP13, BNT15] to name a few. We refer the reader to the references given in these articles for a comprehensive list of contributions.

If we assume that all the functions F from the set $\mathbb{F}$ are further continuous, then it is easy to see that the set of transport plans $\Pi_{\mathbb{F}}(\mu, \nu)$ is closed, and as a consequence of Prokhorov's theorem, it is compact. Moreover, in this case the functional

$$\pi \mapsto \int F d\pi,$$

is continuous for $F \in \mathbb{F}$. Compactness and continuity imply that the transport problem (COT) is attained if and only if the set $\Pi_{\mathbb{F}}(\mu, \nu)$ is non-empty. However, if we allow for a more general set of test functions $\mathbb{F}$, meaning that we relax the continuity assumption on the functions $F \in \mathbb{F}$, then the compactness of $\Pi_{\mathbb{F}}(\mu, \nu)$ is not straightforward anymore, as we will see in the next chapter (see Lemma 2.1.8).

As in classical transport, there is also a dual formulation for (COT):

$$\sup_{\substack{(\phi,\psi)\in\Psi, F\in\mathbb{F} \text{ s.t.}\\ \phi\oplus\psi\le c+F}} \left\{\int \phi d\mu + \int \psi d\nu\right\}, \tag{1.2.1}$$

where Ψ is defined in (1.1.1). The equivalence between (COT) and the maximization problem (1.2.1) for a class $\mathbb{F} \subset \mathcal{C}_b(\mathcal{X}\times\mathcal{Y})$ and continuous cost function c was shown in [Zae15]. His proof can be easily extended to the case of lower semicontinuous cost functions using the same arguments as in the classical case (see e.g. [Vil08]).

1.3 Extensions of duality results

In this section we generalize classical results on attainability and duality for transport problems of the form (OT) for the class of cost functions of the form

$$c(x,y) := \tilde{c}(x,y) + f(x)g(y), \tag{1.3.1}$$

with $\tilde{c}$ to be any lower semicontinuous function, and either f or g continuous. Even though there is no systematic duality theory for Borel measurable cost functions c with values in the extended real line $(-\infty,+\infty]$, the structure of the cost functions we consider allows us to prove attainability/duality results for (OT) in a simple and self-contained way; see Proposition 1.3.1. Moreover, we give a proof of duality for linearly constrained transport problems, for the same type of costs, see Corollary 1.3.3. The latter extension will be important for us when proving duality for causal optimal transport; see Theorem 2.1.10 for the result in discrete time, and Theorem 4.1.6 for a continuous time analogue.

Proposition 1.3.1. *Let $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ be bounded Borel functions and $\tilde{c} : \mathcal{X}\times\mathcal{Y} \to (-\infty,\infty]$ lower semicontinuous and bounded from below. Suppose that either f or g is further continuous, and define*

$$c(x,y) := \tilde{c}(x,y) + f(x)g(y).$$

Then the optimal transport problem (OT) *corresponding to the cost c is attained. Furthermore, there is no duality gap:*

$$\inf_{\pi\in\Pi(\mu,\nu)} \int c d\pi = \sup_{\substack{\phi\in\mathcal{C}_b(\mathcal{X}),\, \psi\in\mathcal{C}_b(\mathcal{Y})\\ \phi\oplus\psi\le c}} \left\{\int \phi(x)\mu(dx) + \int \psi(y)\nu(dy)\right\}.$$

From the proof it will be clear that one can take c above to contain a finite sum of terms of the form $f(x)g(y)$ as described. Before we give the proof of the above proposition, we prove the following lemma.

Lemma 1.3.2. *Let $(\mathcal{X},\mu)$ and $(\mathcal{Y},\nu)$ be Polish probability spaces. Let $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ be bounded Borel functions, at least one of which is continuous. Then the function*

$$\pi \mapsto \int f(x)g(y)d\pi(x,y)$$

is continuous on $\Pi(\mu,\nu)$ with respect to the weak topology.

Proof. We assume w.l.o.g. that f is continuous and let $M \in \mathbb{R}$ be such that $|f| \le M$, and $g_k : \mathcal{Y} \to \mathbb{R}, k \in \mathbb{N}$, be a sequence of bounded continuous functions converging to g in $L^1(\nu)$. Consider a sequence of measures $(\pi_n)_n \subseteq \Pi(\mu,\nu)$ such that π_n converges weakly to

π for some $\pi \in \Pi(\mu,\nu)$. Recall that $\Pi(\mu,\nu)$ is a compact subset of $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ with respect to the weak topology. For any $\varepsilon > 0$, take $k(\varepsilon)$ such that $\|g - g_k\|_{L^1(\nu)} \leq \varepsilon/(2M+1)$ for all $k \geq k(\varepsilon)$, and take $n(\varepsilon)$ such that for all $n \geq n(\varepsilon)$

$$\left| \int f(x)g_{k(\varepsilon)}(y)d\pi(x,y) - \int f(x)g_{k(\varepsilon)}(y)d\pi_n(x,y)\right| \leq \varepsilon/(2M+1).$$

Then, for all $n \geq n(\varepsilon)$ we have

$$\begin{aligned} &\left| \int f(x)g(y)d\pi(x,y) - \int f(x)g(y)d\pi_n(x,y)\right| \leq \int |f(x)||g(y)-g_{k(\varepsilon)}(y)|d\pi(x,y)+ \\ &\left|\int f(x)g_{k(\varepsilon)}(y)d\pi(x,y) - \int f(x)g_{k(\varepsilon)}(y)d\pi_n(x,y)\right| + \int |f(x)||g(y)-g_{k(\varepsilon)}(y)|d\pi_n(x,y) \\ &\qquad \leq \tfrac{M\varepsilon}{2M+1} + \tfrac{\varepsilon}{2M+1} + \tfrac{M\varepsilon}{2M+1} \leq \varepsilon, \end{aligned}$$

which proves the desired statement. □

Proof of Proposition 1.3.1. For the proof of this result we employ classical arguments, as in [Vil03] or [BS11, Section 1.3]. Since the cost function $\tilde{c}$ is lower semicontinuous, there exists a sequence $(\tilde{c}_n)_n$ of bounded continuous functions on $\mathcal{X}\times\mathcal{Y}$ such that $\tilde{c}_n$ is increasing to $\tilde{c}$. We are going to show that the sequence

$$P(c_n) := \inf_{\pi\in\Pi(\mu,\nu)} \int c_n d\pi$$

converges to $P(c) := \inf_{\pi\in\Pi(\mu,\nu)} \int c d\pi$, where we denote $c_n(x,y) = \tilde{c}_n(x,y) + f(x)g(y)$. In order to do so, for each $n \in \mathbb{N}$, we pick $\pi_n \in \Pi(\mu,\nu)$ such that $\int c_n d\pi_n \leq P(c_n) + 1/n$. Since $\Pi(\mu,\nu)$ is weakly compact, we may assume that $(\pi_n)_n$ converges weakly to some transport plan $\pi \in \Pi(\mu,\nu)$. Then, it holds

$$\begin{aligned} P(c) &\leq \int c d\pi = \lim_m \int c_m d\pi = \lim_m(\lim_n \int c_m d\pi_n) \leq \lim_m(\lim_n \int c_n d\pi_n) \\ &= \lim_n \int c_n d\pi_n = \lim_n P(c_n) \leq P(c), \end{aligned}$$

where we used monotone convergence in the first equality, Lemma 1.3.2 to ensure that $\int c_m d\pi = \lim_n \int c_m d\pi_n$, and the fact that c_n is an increasing sequence with $P(c_n) \leq \int c_n d\pi_n \leq P(c_n) + 1/n$. This concludes the proof of our claim. Moreover, by compactness and Lemma 1.3.2, it actually holds that π is an optimizer for the cost c. The function c_n is Borel bounded, so by [Kel84, Theorem 2.14] we have that duality holds for it. Thus we can pick (ψ_n, ϕ_n) such that

$$\int \psi_n d\mu + \int \phi_n d\nu \geq P(c_n) - 1/n,$$

and hence

$$\sup_n(\int \psi_n d\mu + \int \phi_n d\nu) \geq \lim_n P(c_n) - 1/n = P(c).$$

Since $\psi_n(x) + \phi_n(y) \leq c_n(x,y) \leq c(x,y)$, duality is established. □

Corollary 1.3.3. *Let $\mathfrak{F}$ (resp. $\mathfrak{G}$) be a non-empty collection of real-valued bounded Borel functions on $\mathcal{X}$ (resp. $\mathcal{Y}$), and define*[1] $\mathbb{F} := span\{fg : f \in \mathfrak{F}, g \in \mathfrak{G}\}$. *We define the optimal transport problem with linear constraints determined by $\mathbb{F}$ as before:*

$$\inf_{\pi\in\Pi_{\mathbb{F}}(\mu,\nu)} \int c d\pi. \tag{1.3.2}$$

Assume that $c : \mathcal{X}\times\mathcal{Y} \to (-\infty,\infty]$ is lower semicontinuous and bounded from below, and that either all elements of $\mathfrak{F}$, or all elements of $\mathfrak{G}$, are continuous. Then (1.3.2) *is*

[1] Here and thereafter, *span* denotes the linear space of functions obtained by finite linear combinations of those functions in the generating class.

attained, and there is no duality gap:

$$\inf_{\pi\in\Pi_{\mathbb{F}}(\mu,\nu)} \int c d\pi = \sup_{\substack{\phi\in\mathcal{C}_b(\mathcal{X}),\, \psi\in\mathcal{C}_b(\mathcal{Y}),\, F\in\mathbb{F} \\ \phi\oplus\psi\leq c+F}} \left\{\int \phi(x)\mu(dx) + \int \psi(y)\nu(dy)\right\}.$$

Note that this result is not a consequence of [Zae15], who only deals with continuous test functions in the class $\mathbb{F}$.

Proof. The set $\Pi_{\mathbb{F}}(\mu,\nu)$ is a weakly closed subset of the compact set $\Pi(\mu,\nu)$ (follows from Lemma 1.3.2). This is enough to ensure the attainability of (1.3.2).

It is immediate that

$$\inf_{\pi\in\Pi_{\mathbb{F}}(\mu,\nu)} \int c d\pi = \inf_{\pi\in\Pi(\mu,\nu)} \sup_{F\in\mathbb{F}} \int (c+F) d\pi = \sup_{F\in\mathbb{F}} \inf_{\pi\in\Pi(\mu,\nu)} \int (c+F) d\pi,$$

by the usual minimax arguments (see e.g. [Sio58]), and by observing that the affine bilinear objective function is lower semicontinuous in π by Lemma 1.3.2, as π varies over a compact set. By Proposition 1.3.1 we find

$$\inf_{\pi\in\Pi_{\mathbb{F}}(\mu,\nu)} \int c d\pi = \sup_{F\in\mathbb{F}} \sup_{\substack{\phi\in\mathcal{C}_b(\mathcal{X}), \psi\in\mathcal{C}_b(\mathcal{Y}) \\ \phi\oplus\psi\leq c+F}} \left\{\int \phi(x)\mu(dx) + \int \psi(y)\nu(dy)\right\},$$

yielding the desired result. $\square$

Chapter 2

Causal transport in discrete time

This chapter is devoted to the analysis of causal transport problems in discrete time. First, we show that causality is a linear constraint on measures and provide an equivalent characterization to test this property (see Proposition 2.1.5). Next, in Theorem 2.1.10 we present a dual formulation, proving that there is no duality gap. One of the main results of this chapter is the dynamic programming principle (DPP) for causal transport problems given in Theorem 2.2.2. We establish it with the help of the so-called quasi-Markov transport plans, that do posses a DPP. We show that the value of transport problems induced by the latter coincides with causal optimal transport under a markovianity assumption on the source measure and for separable cost functions (see Theorem 2.2.5 and Proposition 2.2.6). Our next main result, Theorem 2.2.13, establishes the equivalence of causal and bicausal transport problems, and furthermore, it identifies the Knothe-Rosenblatt rearrangement as a causal optimizer for convex separable costs. The key here is first to identify the mentioned rearrangement as a bicausal optimizer, which we do in Proposition 2.2.14 generalizing a corresponding result in [Rüs85], and then to prove that under the given assumptions the values of the causal and bicausal problems coincide. As a consequence, we provide an extension of the previous result to the case when the source measure has independent increments and the cost is suitably modified (see Corollary 2.2.17). In Section 2.3 we examine to which extent the aforementioned rearrangement is canonical, characterizing it as the unique bicausal transport plan which is an increasing triangular transport. Sections 2.4 and 2.5 are devoted to the discussion of the connection between bicausal transport problems and stochastic programming. We show that many stochastic programs are concave/convex along what we define as lexicographical displacement interpolations, in analogy to the concepts of displacement interpolation and displacement convex functionals in classical optimal transport theory (see [Vil03, chapter 5] or McCann [McC97]), where the role of Brenier's map is taken by the Knothe-Rosenblatt rearrangement. Further, in Theorem 2.5.1 we give conditions under which the nested distance of "order one", which gauges the discrepancy between stochastic programs driven by different noise distributions and a common Lipschitz-cost criterion, can itself be assessed by the square root of the relative entropy between such processes. In other words, we establish a transport-information inequality for this nested distance of "order one". We finish this chapter with a short discussion on enlargements of filtrations in a discrete-time setting; this is a warm-up for Chapter 4.

2.1 Causality as a linear constraint

Let $\mathcal{X}$ and $\mathcal{Y}$ be closed subsets of $\mathbb{R}^N$ that will be the ambient spaces until the end of the chapter. We endow the spaces $\mathcal{X}$ and $\mathcal{Y}$ with filtrations $\mathcal{F} = (\mathcal{F}_t)_{t=1}^N$ and $\mathcal{G} = (\mathcal{G}_t)_{t=1}^N$ respectively. We assume that

$$\mathcal{F} \text{ is generated by the coordinate process,}$$

i.e. $\mathcal{F}_t$ is the smallest σ-algebra s.t. $x \in \mathcal{X} \mapsto (x_1, \ldots, x_t) \in \mathbb{R}^t$ is measurable, and so forth. Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be given two probability measures. In the following we assume w.l.o.g. that $\mathcal{X} = supp(\mu)$ and $\mathcal{Y} = supp(\nu)$, whenever dealing with transport problems between μ and ν.

Most of the results would still hold for $\mathcal{X} = \mathcal{Y} = S^N$, with S a Polish space, i.e. we could work with S-valued discrete-time stochastic processes in N-steps as well. It is mostly for the sake of familiarity that we shall take $S = \mathbb{R}$ throughout the whole chapter (moreover, we could also take products of different Polish spaces). We shall commonly denote by $(x, y) = (x_1, \ldots, x_N, y_1, \ldots, y_N)$ a generic element in $\mathbb{R}^N \times \mathbb{R}^N$.

For a process Z on a given probability space $(\Omega, \mathbb{P})$, we denote by $\mathcal{F}^Z := Z^{-1}(\mathcal{F})$ the filtration generated by Z on Ω. For the rest of this chapter, let $(\Omega, \mathbb{P})$ be a fixed probability space.

Definition 2.1.1 (Causal transport plan)**.** *A transport plan $\pi \in \Pi(\mu, \nu)$ is called causal between $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ if, for any $t \in \{1, \ldots, N\}$ and any set $B \in \mathcal{G}_t$, the map*

$$\mathcal{X} \ni x \to \pi^x(B)$$

is measurable with respect to $\mathcal{F}_t$. The set of all such plans will be denoted

$$\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu).$$

Analogously, we will be interested in transport plans that are "causal in both directions", or *bicausal* in our terminology. The set of all such plans is explicitly given by

$$\Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu, \nu) = \left\{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu) \text{ s.t. } e_*\pi \in \Pi^{\mathcal{G},\mathcal{F}}(\nu, \mu)\right\},$$

where $e(x, y) = (y, x)$. As in the usual optimal transport problem, the set of all causal plans $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$, as well as $\Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu, \nu)$, are always non-empty because the product measure $\mu \otimes \nu \in \Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu, \nu)$. Further, as in the classical setting, we shall consider the case of Borel measurable transformations $T : \mathcal{X} \to \mathcal{Y}$ satisfying $T_*\mu = \nu$. (Monge) transport maps T are termed (bi)causal if the associated $\pi^T := (\mathrm{id} \times T)_*\mu$ belongs to $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$ (resp. $\Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu, \nu)$).

Definition 2.1.2 (Causal coupling)**.** *A pair (X, Y) of continuous processes on $(\Omega, \mathbb{P})$, is called a* causal coupling *if $(X, Y)_*\mathbb{P}$ is a causal transport plan between $(\mathcal{X}, \mathcal{F}, X_*\mathbb{P})$ and $(\mathcal{Y}, \mathcal{G}, Y_*\mathbb{P})$.*

The condition for a transport plan to be causal, written in terms of stochastic processes, looks as follows: for all $t = 1, \ldots, N$ and any set $B_t \in \mathcal{G}_t$,

$$\mathbb{P}\big((Y_1, \ldots, Y_N) \in B_t \mid X_1, \ldots, X_N\big) = \mathbb{P}\big((Y_1, \ldots, Y_N) \in B_t \mid X_1, \ldots X_t\big).$$

In the case $\mathcal{F} = \mathcal{G}$, which will occupy us for most of the chapter, the above equality reads as "given the past of X, the past of Y and the future of X are independent". This is perhaps best interpreted by the following equivalent formulation (see e.g. [Kal02, Proposition 6.13]):

$$Y_t = F_t(X_1, \dots, X_t, U_t) \quad \forall t \in \{1, \dots, N\},$$

for some measurable functions F_t and where each U_t is a uniform random variable independent of $X_1, \dots, X_N$; however, the U_t's need not be independent of each other.

Remark 2.1.3. Clearly, if $\mathcal{F} = \mathcal{G}$, a transport map T is causal if and only if it is adapted, in the sense that there exist a Borel measurable $T^t : \mathbb{R}^t \to \mathbb{R}$ such that for μ-a.e. $(x_1, \dots, x_N)$:

$$T(x_1, \dots, x_N) \;=\; (T^1(x_1), T^2(x_1, x_2), \dots, T^N(x_1, \dots, x_N)).$$

Bicausality is more subtle, but it holds if for instance T admits a measurable μ-a.s. left inverse which is also adapted.

Example 2.1.4. This example shows that there may be no causal Monge maps even if classical Monge maps do exist. One can easily observe that for the measures

$$\mu = a_1\, \delta_{(1,2)} + a_2\, \delta_{(1,0)} + a_3\, \delta_{(-1,0)} + a_4\, \delta_{(-1,-2)},$$

and

$$\nu = a_1\, \delta_{(1,2)} + a_3\, \delta_{(1,0)} + a_2\, \delta_{(-1,0)} + a_4\, \delta_{(-1,-2)},$$

where $a_i, i = 1, \dots, 4$ are positive numbers that sum up to one and $a_i \neq a_j$, the only Monge map is given by:

$$T(1,2) = (1,2)\ ,\ T(-1,-2) = (-1,-2)\ ,\ T(1,0) = (-1,0)\ ,\ T(-1,0) = (1,0)$$

and the corresponding transport plan is

$$\begin{aligned}\pi^T = a_1\, \pi((1,2);(1,2)) \quad &+ a_2\, \pi((1,0);(-1,0)) \\ + a_3\, \pi((-1,0);(1,0)) &+ a_4\, \pi((-1,-2);(-1,-2)).\end{aligned}$$

This transport plan is clearly non-causal since for example in coupling notation

$$\mathbb{P}\left[Y_1 = 1 | X_1 = -1, X_2 = -2\right] = 0,$$

whereas

$$\mathbb{P}\left[Y_1 = 1 | X_1 = -1\right] = \frac{a_3}{a_3 + a_4}.$$

Hence, there is no causal Monge map pushing forward μ to ν; i.e. for any causal transport plan the mass must split.

Now we introduce our main optimization problem, the *causal optimal transport problem*: given some Borel cost function c defined on $\mathbb{R}^N \times \mathbb{R}^N$ and the probability measures μ, ν, find the minimal cost at which they can be coupled in a causal way, i.e. consider

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \int c\, d\pi. \tag{2.1.1}$$

Minimizing over the set $\Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ defines the bicausal optimal transport problem

$$\inf_{\pi\in\Pi_{bc}^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \int c\, d\pi. \tag{2.1.2}$$

These should be compared to the classical problem of optimal transport in which minimization is done over $\Pi(\mu,\nu)$. In the following, we shall usually assume c to be bounded from below; in principle it would also suffice that $c(x,y)\geq a(x)+b(y)$ for $a\in L^1(\mu)$, $b\in L^1(\nu)$.

In what follows, given two measures μ,ν on $\mathbb{R}^N$, we will study (bi)causal transport problems between $(\mathbb{R}^N,\mathcal{F},\mu)$ and $(\mathbb{R}^N,\mathcal{G},\nu)$. It is clear that in general the value of of the classical transport problem (OT) is smaller than or equal to the value of the causal transport problem (2.1.1), which in turn is smaller than or equal to the value of the bicausal transport problem (2.1.2), see Section 1.1.

The following proposition shows that causality can be seen as a linear constraint on measures on the product space, stated in terms of a special class of test functions or via discrete stochastic integrals.

Proposition 2.1.5. *For a probability measure $\pi\in\Pi(\mu,\nu)$, the following are equivalent:*

1. *π is a causal transport plan w.r.t. $\mathcal{F}$ and $\mathcal{G}$;*
2. *For all $t\in\{1,\dots,N\}$, $g_t\in C_b(\mathbb{R}^N)$ and $\mathcal{G}_t$-measurable functions $h_t\in B_b(\mathbb{R}^N)$, we have*

$$\int h_t(y_1,\dots,y_N)\left\{g_t(x_1,\dots,x_N)-\int g_t(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right\}d\pi=0; \tag{2.1.3}$$

3. *For every bounded $\mathcal{G}$-adapted process H and each continuous bounded $(\mu,\mathcal{F})$-martingale M we have*

$$\int\sum_{t<N} H_t(y_1,\dots,y_N)\left[M_{t+1}(x_1,\dots,x_{t+1})-M_t(x_1,\dots,x_t)\right]d\pi=0;$$

4. *Every square integrable bounded $(\mu,\mathcal{F})$-martingale is a square integrable $(\pi,\mathcal{F}\otimes\mathcal{G})$-martingale;*
5. *For all $t\in\{1,\dots,N\}$ the σ-fields $\mathcal{G}_t$ and $\mathcal{F}_N$ are conditionally independent with respect to π given $\mathcal{F}_t$.*

Proof. STEP 1. Equivalence between Points 1 and 2:

For $\mathcal{G}_t$-measurable $h_t\in B_b(\mathbb{R}^N)$ denote

$$f^h(x_1,\dots,x_N):=\int h_t(y_1,\dots,y_N)\pi^{x_1,\dots,x_N}(dy_1,\dots,dy_N).$$

By definition $\pi\in\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ if and only if for all $t\leq N$ the regular conditional kernel of π is $\mathcal{F}_t-$measurable, i.e. for all such f^h we have

$$f^h(x_1,\dots,x_N)=\int f^h(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N),$$

which is equivalent to the following:

$$\int g(x_1,\dots,x_N)\left[f^h(x_1,\dots,x_N)-\int f^h(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\mu=0, \tag{2.1.4}$$

for every function $g\in C_b(\mathbb{R}^N)$ and for all $t\leq N$. We can take the functions g's continuous and not merely Borel bounded, since under the given assumptions any bounded Borel function can be approximated by a sequence of continuous bounded ones in $L^p(\mu)$-norm for every $1\leq p<\infty$, where μ is a Borel finite measure on a Polish space. Having observed that

$$\begin{aligned}
&\int f^h(x_1,\dots,x_N)\left[\int g(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\mu=\\
&\int\left[\int f^h(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]\left[\int g(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\mu=\\
&\int g(x_1,\dots,x_N)\left[\int f^h(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\mu,
\end{aligned}$$

equation (2.1.4) is equivalent to the following

$$\int f^h(x_1,\dots,x_N)\left[g(x_1,\dots,x_N)-\int g(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\mu=0.$$

By the tower property of conditional expectation the above equation is equivalent to:

$$\int h_t(y_1,\dots,y_N)\left[g_t(x_1,\dots,x_N)-\int g_t(x_1,\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\right]d\pi \;=\; 0.$$

STEP 2. Equivalence between Points 2 and 3: Choosing $g_t=M_{t+1}$ and $h_t=H_t$ for each $t<N$, and summing up, proves Point 3 from Point 2. Conversely, given t, h_t and g_t we build $H_s=h_t\mathbf{1}_{s\geq t}$ and

$$M_s=\int g_t(x_1,\dots,x_s,x_{s+1},\dots,x_N)\mu^{x_1,\dots,x_s}(dx_{s+1},\dots,dx_N),$$

and conclude by telescopic sum over s.
STEP 3. Equivalence between Points 1 and 4: Let M be a $(\mu,\mathcal{F})$-martingale. For $0\leq s<t\leq N$ and $\mathcal{F}_s\otimes\mathcal{G}_s-$ adapted functions f_s we have

$$\begin{aligned}
&\int M_t(x_1,\dots,x_t)f_s(x_1,\dots,x_s,y_1,\dots,y_s)d\pi(x_1,\dots,x_t,y_1,\dots,y_s)=\\
&\int M_t(x_1,\dots,x_t)\left(\int f_s(x_1,\dots,x_s,y_1,\dots,y_s)d\pi^{x_1,\dots,x_t}(y_1,\dots,y_s)\right)d\mu(x_1,\dots,x_t)=\\
&\int M_t(x_1,\dots,x_t)\left(\int f_s(x_1,\dots,x_s,y_1,\dots,y_s)d\pi^{x_1,\dots,x_s}(y_1,\dots,y_s)\right)d\mu(x_1,\dots,x_t)=\\
&\int M_s(x_1,\dots,x_s)\left(\int f_s(x_1,\dots,x_s,y_1,\dots,y_s)d\pi^{x_1,\dots,x_s}(y_1,\dots,y_s)\right)d\mu(x_1,\dots,x_s)=\\
&\int M_s(x_1,\dots,x_s)f_s(x_1,\dots,x_s,y_1,\dots,y_s)d\pi(x_1,\dots,x_s,y_1,\dots,y_s),
\end{aligned}$$

where causality is used in the second equality. Conversely, let every bounded $(\mu,\mathcal{F})$-martingale be $(\pi,\mathcal{F}\otimes\mathcal{G})$-martingale. Then for any $\mathcal{G}$-adapted process H we have

$$\begin{aligned}
&\int\sum_{t<N}H_t(y_1,\dots,y_N)M_{t+1}(x_1,\dots,x_{t+1})d\pi(x_1,\dots,x_N,y_1,\dots,y_N)=\\
&\int\sum_{t<N}H_t(y_1,\dots,y_N)\int M_{t+1}(x_1,\dots,x_{t+1})d\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(x_{t+1},\dots,x_N,y_{t+1},\dots,y_N)d\pi=
\end{aligned}$$

$$\int \sum_{t<N} H_t(y_1,\ldots,y_N)M_t(x_1,\ldots,x_t)d\pi(x_1,\ldots,x_N,y_1,\ldots,y_N),$$

and causality follows from Point 3.

STEP 4. Equivalence between Points 4 and 5: This equivalence can be shown as in [BY78, Theorem 3]. Point 4 can equivalently be stated as: for all $t \in \{1,\ldots,N\}$ and for every μ-a.s. integrable $\mathcal{F}_N$-adapted process X the following holds

$$\begin{aligned}
&\int X(x_1,\ldots,x_N)d\mu^{x_1,\ldots,x_t}(x_{t+1},\ldots,x_N) = \\
&\int X(x_1,\ldots,x_N)d\pi^{x_1,\ldots,x_t}(x_{t+1},\ldots,x_N,y_1,\ldots,y_N) = \\
&\int X(x_1,\ldots,x_N)d\pi^{x_1,\ldots,x_t,y_1,\ldots,y_t}(x_{t+1},\ldots,x_N,y_{t+1},\ldots,y_N),
\end{aligned}$$

which is one of the possible ways to define the conditional independence of filtrations $\mathcal{G}_t$ and $\mathcal{F}_N$ with respect to π given $\mathcal{F}_t$. $\square$

2.1.1 Duality

As any linear optimization problem, causal transport problem admits a dual formulation. The main result of this section is Theorem 2.1.10 which gives the attainability and no-duality-gap result. As in the classical setting, there exists a dual formulation to the (bi)causal transport problem. We have already observed in Proposition 2.1.5 that these problems form a subclass of optimal transport problems under linear constraints discussed in Section 1.2. Let us make this precise in the present setting, by defining the following sets of test functions that will be instrumental for the dual formulation:

$$\mathbb{F} := \left\{ \begin{array}{c} F:\mathbb{R}^N\times\mathbb{R}^N\to\mathbb{R} \text{ s.t. } F(x_1,\ldots,x_N,y_1,\ldots,y_N) = \\ \sum_{t<N} h_t(y_1,\ldots,y_N)\left[g_t(x_1,\ldots,x_N) - \int g_t(x_1,\ldots,x_N)\mu^{x_1,\ldots,x_t}(dx_{t+1},\ldots,dx_N)\right], \\ \text{with } h_t\in B_b(\mathbb{R}^N,\mathcal{G}_t),\ g_t\in C_b(\mathbb{R}^N) \text{ for all } t<N \end{array} \right\}, \tag{2.1.5}$$

$$\mathbb{S} := \left\{ \begin{array}{c} S:\mathbb{R}^N\times\mathbb{R}^N\to\mathbb{R} \text{ s.t. } S(x_1,\ldots,x_N,y_1,\ldots,y_N) = \\ \sum_{t<N} H_t(y_1,\ldots,y_N)\left[M_{t+1}(x_1,\ldots,x_{t+1}) - M_t(x_1,\ldots,x_t)\right], \\ \text{with } H_t\in B_b(\mathbb{R}^N,\mathcal{G}_t),\ M_t\in C_b(\mathbb{R}^t) \text{ for all } t<N, \text{ and with } M \text{ a martingale} \end{array} \right\}. \tag{2.1.6}$$

Let us also notice that $\mu^{x_1,\ldots,x_{N-1}}$ is a univariate measure, and more generally we will commonly write $\mu^{x_1,\ldots,x_t}(dx_{t+1},\ldots,dx_{t+k})$ for the measure $\mu^{x_1,\ldots,x_t}(dx_{t+1},\ldots,dx_{t+k},\mathbb{R}^{N-t-k})$ and similarly $\mu(dx_1,\ldots,dx_t)$ for the projection of μ into the first t-marginals.

The following assumption eases the proof of Theorem 2.1.10:

Assumption 2.1.6. *The measure μ is successively weakly continuous in the sense that for each $t<N$, there is a version of the regular conditional kernel of μ w.r.t. its first t variables s.t.*

$$(x_1,\ldots,x_t)\in supp(\mu(dx_1,\ldots,dx_t)) \mapsto \mu^{x_1,\ldots,x_t}(dx_{t+1},\ldots,dx_N)\in\mathcal{P}(\mathbb{R}^{N-t}),$$

is continuous w.r.t. the weak topology in the range and the relative topology in the domain.

If $supp(\mu)$ contains no accumulation points, as in the random walk/event tree setting, Assumption 2.1.6 is vacuously fulfilled. On the other extreme, there are many discrete-time processes with full support satisfying it, e.g. the Gaussian case.

We now proceed to the duality and attainment questions. First:

Lemma 2.1.7. *Suppose Assumption 2.1.6 holds. For $g \in C_b(\mathbb{R}^N)$, the functions*

$$(x_1, \dots, x_N) \mapsto g(x_1, \dots, x_N) - \int g(x_1, \dots, x_N)\mu^{x_1,\dots,x_t}(dx_{t+1}, \dots, dx_N),$$

belong themselves to $C_b(\mathbb{R}^N)$.

Proof. It is suffices to check the continuity of

$$(x_1, \dots, x_t) \mapsto \int g(x_1, \dots, x_N)\mu^{x_1,\dots,x_t}(dx_{t+1}, \dots, dx_N).$$

Let $x^n := (x_1^n, \dots, x_t^n)$ converge to $y := (y_1, \dots, y_t)$, thus we may assume all x^n belongs to a fixed ball B around y^n. Let us fix an arbitrary $\varepsilon > 0$. By assumption μ^{x^n} converges weakly to μ^y and so in particular $\{\mu^{x^n}\}_{n\in\mathbb{N}}$ is tight. Thus we may find a compact in $\mathbb{R}^{N-t}$ such that $\sup_n \mu^{x^n}(K^c) \le \varepsilon/3$. For large enough n_0 we also have that

$$\sup_{n>n_0, z\in K} |g(x^n, z) - g(y, z)| \le \varepsilon/3,$$

since g restricted to $B \times K$ must be uniformly continuous. We then write:

$$\begin{aligned}&\left|\int g(x^n, z)\mu^{x^n}(dz) - \int g(y, z)\mu^y(dz)\right| \le \left|\int_K [g(x^n, z) - g(y, z)]\mu^{x^n}(dz)\right| + \\ &\left|\int_{K^c} [g(x^n, z) - g(y, z)]\mu^{x^n}(dz)\right| + \left|\int g(y, z)[\mu^{x^n}(dz) - \mu^y(dz)]\right|\end{aligned}$$

We easily see that each term is smaller than $\varepsilon/3$ for n large enough and the continuity follows. □

This lemma also implies that under Assumption 2.1.6 we can test causality against continuous bounded functions g in the set $\mathbb{F}$ (resp. M in $\mathbb{S}$), instead of just bounded Borel.

Lemma 2.1.8. *Under Assumption 2.1.6 the set $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$ of causal transport plans is compact for weak convergence.*

Proof. Since $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu) \subseteq \Pi(\mu, \nu)$, and the latter is weakly compact, we only need to show that $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$ is weakly closed. From Lemma 2.1.7 and Lemma 1.3.2 we know that the functions $\pi \mapsto \int F d\pi, F \in \mathbb{F}$, are continuous with respect to the weak topology. The result follows from Proposition 2.1.5, since the condition $\int F d\pi = 0$ is closed for all $F \in \mathbb{F}$. □

Remark 2.1.9. Observe that the set of causal transport plans with the fixed first marginal, denoted by $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \cdot)$, is closed too. This result will be needed in Section 2.6.

Now we are ready to present the basic primal attainability/no-duality-gap result:

Theorem 2.1.10 (Causal transport duality)**.** *Suppose that $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous and bounded from below, and that Assumption 2.1.6 holds. Then there is no duality gap*

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \int c d\pi \quad = \sup_{\substack{\phi,\psi \in C_b(\mathbb{R}^N), F \in \mathbb{F} \\ \phi \oplus \psi \leq c + F}} \left[\int \phi d\mu + \int \psi d\nu\right] \quad = \sup_{\substack{\phi,\psi \in C_b(\mathbb{R}^N), S \in \mathbb{S} \\ \phi \oplus \psi \leq c + S}} \left[\int \phi d\mu + \int \psi d\nu\right],$$

and the infimum on the l.h.s. (i.e. (2.1.1)*) is attained.*

It is also easy to see that the dual problem can be reduced to both:

$$\sup_{\substack{\psi \in C_b(\mathbb{R}^N), F \in \mathbb{F}, \\ \psi \leq c + F}} \int \psi d\nu \quad \text{and} \quad \sup_{\substack{\psi \in C_b(\mathbb{R}^N), S \in \mathbb{S}, \\ \psi \leq c + S}} \int \psi d\nu.$$

Proof. From Lemma 1.3.2 we know that the set $\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ is weakly compact and attainability follows by classical arguments (every lower semicontinuous functions attains its minimum on the compact set). Next, we observe that under the weak continuity assumption on μ, all the x-dependent factors generating $\mathbb{F}$ and $\mathbb{S}$ are continuous and bounded. Then duality is a direct consequence of Proposition 1.3.3, upon observing that the sets $\mathbb{F}$ and $\mathbb{S}$ have the correct structure. The fact that ϕ disappears from the dual problem follows from the fact that $\phi - \int \phi d\mu$ belongs to $\mathbb{F}$ (resp. $\mathbb{S}$).

□

Our duality result cannot be seen as a consequence of duality given in [Zae15] or [Las15]. In the first paper the author gives the proof for the set of linear constraints with continuous and bounded test functions, whereas in the second paper the author makes restrictive assumptions on filtrations requiring them to be of a particular structure. Neither of these is suitable for us, taking into account that we would like to work in the most general setting that allows to consider the enlargements of filtrations, see Section 2.6 for a discrete time setting and Chapter 4 for a discussion in a continuous time.

2.2 Main results

In this section we discuss in details the causal and bicausal transport problems in an adapted setting, i.e. we assume that filtrations

$$\mathcal{F} = \mathcal{G} \text{ are generated by the coordinate processes.}$$

The main results are the recursive formulations for both problems and the characterization of bicausal optimizers, given by the Knothe-Rosenblatt rearrangement. Moreover, we give the conditions under which the latter can be seen as causal optimizer as well.

It is known that is possible to decompose a measure $\pi \in \Pi(\mathbb{R}^N, \mathbb{R}^N)$ uniquely in a suitable way, by successive disintegration first w.r.t. $(x_1, \dots, x_{N-1}, y_1, \dots, y_{N-1})$, then w.r.t. $(x_1, \dots, x_{N-2}, y_1, \dots, y_{N-2})$ and so forth in a recursive way, obtaining

$$\pi(dx_1, \dots, dx_N, dy_1, \dots, dy_N) = \bar{\pi}(dx_1, dy_1)\pi^{x_1,y_1}(dx_2, dy_2) \dots \pi^{x_1,\dots,x_{N-1},y_1,\dots,y_{N-1}}(dx_N, dy_N), \quad (2.2.1)$$

so that each $\pi_t(dx_t, dy_t) := \pi^{x_1,\ldots,x_{t-1},y_1,\ldots,y_{t-1}}(dx_t, dy_t)$ is a regular conditional probability or kernel (see [Bog07, Theorem 10.4.14, Corollary 10.4.17] and [Kal02, Lemma 1.41]). We will freely perform concatenation of these objects or further decompose them into smaller kernels, as justified in [BS78, Chapter 7.4.3].

We start with a lemma that allows us to characterize the causal transport plans using the successive disintegrations of measures on a product space.

Lemma 2.2.1. *For $\pi \in \Pi(\mathbb{R}^N, \mathbb{R}^N)$, the following statements are equivalent:*

1. *π is a causal transport between $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{F}, \nu)$;*

2. *Decomposing π in terms of successive regular kernels*

$$\pi(dx_1,\ldots,dx_N,dy_1,\ldots,dy_N) = \bar{\pi}(dx_1,dy_1)\pi^{x_1,y_1}(dx_2,dy_2)\ldots\pi^{x_1,\ldots,x_{N-1},y_1,\ldots,y_{N-1}}(dx_N,dy_N), \quad (2.2.2)$$

then $\bar{\pi} \in \Pi(p^1_*\mu, p^1_*\nu)$ and for $t < N$ and π-almost all $x_1, \ldots, x_t, y_1, \ldots, y_t$

$$p^1_*\pi^{x_1,\ldots,x_t,y_1,\ldots,y_t}(dx_{t+1}) \quad = \quad \mu^{x_1,\ldots,x_t}(dx_{t+1}), \quad (2.2.3)$$

and for ν-almost all $y_1, \ldots, y_t$

$$\pi^{y_1,\ldots,y_t}(dy_{t+1}) = \nu^{y_1,\ldots,y_t}(dy_{t+1}). \quad (2.2.4)$$

Notice that (2.2.4) is equivalent to

$$\int \pi^{x_1,\ldots,x_t,y_1,\ldots,y_t}(\mathbb{R}, dy_{t+1})\pi^{y_1,\ldots,y_t}(dx_1,\ldots,dx_t) \quad = \quad \nu^{y_1,\ldots,y_t}(dy_{t+1}),$$

which is more convenient for the derivation of the dynamic programming principle to come.

Proof. Let $\pi \in \Pi(\mu, \nu)$ be decomposed as in (2.2.2). It is causal if and only of for any time $t \le n \le N$,

$$\pi^{x_1,\ldots,x_t,y_1,\ldots,y_t}(dx_1,\ldots,dx_n) = \pi^{x_1,\ldots,x_t}(dx_1,\ldots,dx_n),$$

see [Kal02, Proposition 6.6]. Since the x-marginal of π is μ, (2.2.3) follows. For the other condition, observe that for any $t < N$

$$\begin{aligned}
&\textstyle\int h(y_1,\ldots,y_t)\pi(dx_1,\ldots,dx_t,dy_1,\ldots,dy_t) = \qquad (2.2.5)\\
&\textstyle\int h(y_1,\ldots,y_t)\pi^{x_1,\ldots,x_{t-1},y_1,\ldots,y_{t-1}}(dx_t,dy_t)\pi(dx_1,\ldots,dx_{t-1},dy_1,\ldots,dy_{t-1}) =\\
&\textstyle\int h(y_1,\ldots,y_t)\pi^{x_1,\ldots,x_{t-1},y_1,\ldots,y_{t-1}}(dx_t,dy_t)\pi^{y_1,\ldots,y_{t-1}}(dx_1,\ldots,dx_{t-1})\pi(dy_1,\ldots,dy_{t-1}),
\end{aligned}$$

where the dependence on the x's is integrated out, so what is left

$$\begin{aligned}
&\textstyle\int h(y_1,\ldots,y_t)\pi^{y_1,\ldots,y_{t-1}}(dy_t)\pi(dy_1,\ldots,dy_{t-1}) =\\
&\qquad\qquad\textstyle\int h(y_1,\ldots,y_t)\pi^{y_1,\ldots,y_{t-1}}(dy_t)\nu(dy_1,\ldots,dy_{t-1}).
\end{aligned}$$

Hence, $\pi^{y_1,\dots,y_{t-1}}(dy_t)$ must coincide with a kernel of ν from the uniqueness of disintegration, and (2.2.4) follows. For the converse statement, the fact that the x-marginal of π is μ follows from (2.2.3) and $\bar{\pi} \in \Pi(p^1_*\mu, p^1_*\nu)$, whereas the fact that the y-marginal of π is ν follows from $\bar{\pi} \in \Pi(p^1_*\mu, p^1_*\nu)$ and (2.2.4) after the inductive iteration of arguments in (2.2.5) for any $t < N$. To see that these conditions imply causality, it is enough to verify (2.1.3) for any $t = 1, \dots, N-1$. Having observed that functions h_t in (2.1.3) depend only on $y_1, \dots, y_t$, the latter can be written then as

$$\int h_t(y_1,\dots,y_t)\big[g_t(x_1,\dots,x_N) - \int g_t(x_1,\dots,x_t,x_{t+1},\dots,x_N)\mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N)\big]$$
$$\mu^{x_1,\dots,x_{N-1}}(dx_N)\dots\mu^{x_1,\dots,x_t}(dx_{t+1})\pi^{x_1,\dots,x_{t-1},y_1,\dots,y_{t-1}}(dx_t,dy_t)\dots\pi^{x_1,y_1}(dx_2,dy_2)\bar{\pi}(dx_1,dy_1),$$

which is zero because from disintegration property

$$\mu^{x_1,\dots,x_{N-1}}(dx_N)\dots\mu^{x_1,\dots,x_t}(dx_{t+1}) = \mu^{x_1,\dots,x_t}(dx_{t+1},\dots,dx_N).$$

Hence, we conclude.

□

We would like to remark that in the case when both filtrations on $\mathcal{X} = \mathcal{Y} = \mathbb{R}^N$ are coordinate, we can test causality against continuous and bounded functions, i.e. functions h_t in (2.1.3) can be taken from $C_b(\mathbb{R}^t)$. This further implies that causal transport duality for continuous cost functions can be seen as a particular case of the Monge-Kantorovich problem with additional linear constraints considered in [Zae15, Theorem 2.5]. Finally duality for lower semicontinuous cost functions is achieved by the usual approximation arguments in [Vil03, Chapter 1], owing to the compactness of the set of all causal plans.

2.2.1 Dynamic programming principle

The aim of this section is to introduce the dynamic programming principle (DPP) for a class of causal transport problems. We believe that generally a causal transport problem

$$\inf_{\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int c d\pi. \tag{Pc}$$

does not allow for a meaningful and useful "factorized" or "recursive" formulation, along the lines of what a DPP ought to be. However, if we impose more structure in the transport problem, we can get the recursive formulation. The importance of decomposition (2.2.1) together with Lemma 2.2.1 lies in the fact that it suggests that the causal optimal transport problem can be solved recursively if the starting measure μ is Markovian and the cost function has a "semiseparable" structure.

Theorem 2.2.2 (DPP for causal plans)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$, suppose that μ is a Markov measure and that the cost is semiseparable in the sense that for non-negative l.s.c. functions c_t we have*

$$c = \sum_t c_t(x_t, y_1, \dots, y_t).$$

Then, starting from $V^c_N := 0$ and defining recursively for $t = N, \dots, 2$:

$$V^c_{t-1}(y_1,\dots,y_{t-1};m(dx_{t-1})) \;:=\; \inf_{\pi_t\in\Pi\left(\int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)\,,\,\nu^{y_1,\dots,y_{t-1}}(dy_t)\right)} \int \pi_t(dx_t,dy_t)\Big\{c_t(x_t,y_1,\dots,y_t) + V^c_t(y_1,\dots,y_t;\pi^{y_t}_t(dx_t))\Big\}, \quad (2.2.6)$$

and

$$V^c_0 := \inf_{\pi_1\in\Pi(p^1_*\mu,p^1_*\nu)} \int \pi_1(dx_1,dy_1)\{c_1(x_1,y_1)+V^c_1(y_1;\pi^{y_1}_1(dx_1))\},$$

the integrals above are well-defined and $V^c_0 = value(\text{Pc})$*, i.e. the recursion determines* (Pc)*. Furthermore, if Assumption 2.1.6 holds, there exists a causal optimizer* $\tilde\pi$ *with the additional property that for all* t*:*

$$\tilde\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1}) = \tilde\pi^{x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1}),$$

and each of the one-step optimization problems in (2.2.6) *corresponds to a general transport problem in [GRST15] which is convex and l.s.c. in the kernel (i.e. in* $\pi^{y_t}_t$*, for fixed* $y_1,\dots,y_{t-1}$*).*

The very last line of the previous result means concretely that

$$V^c_{t-1}(y_1,\dots,y_{t-1};m(dx_{t-1})) \;=\; \inf_{\substack{y_t\mapsto\pi^{y_t}_t \text{ s.t.}\\ \int_{y_t}\nu^{y_1,\dots,y_{t-1}}(dy_t)\pi^{y_t}_t(dx_t)=\int_{x_{t-1}}m(dx_{t-1})\mu^{x_{t-1}}(dx_t)}} \int_{y_t}\nu^{y_1,\dots,y_{t-1}}(dy_t)\Big\{\int_{x_t}\pi^{y_t}_t(dx_t)c_t(x_t,y_1,\dots,y_t) + V^c_t(y_1,\dots,y_{t-1},y_t;\pi^{y_t}_t(dx_t))\Big\}, \quad (2.2.7)$$

and that the function in curly brackets in the r.h.s. is convex and l.s.c. in $\pi^{y_t}_t$ (ceteris paribus). The same function in brackets is only jointly universally measurable (actually lower semianalytic, l.s.a. in short) as far as we can say, regardless of Assumption 2.1.6. If on the other hand Assumption 2.2.8 holds, then this function is jointly l.s.c. Problems of this kind have been analysed in [GRST15] .

To fully grasp the nature of the DPP one should better look at an example; this requires the use of some extra notation which will be also useful for Theorem 2.2.13 below. We denote by

$$F_\eta(\cdot) := \eta((-\infty,\cdot])$$

the usual cumulative distribution function of a probability measure η on the line, and by $F^{-1}_\eta(u)$ its left-continuous generalized inverse, i.e.

$$F^{-1}_\eta(u) = \inf\{y : F_\eta(y)\ge u\}.$$

Example 2.2.3. Take $N = 2$ and $c = (x_1-y_1)^2+(x_2-y_2)^2$. Using the well-known optimality of the monotone coupling on the line (see e.g. [Vil03, Theorem 2.18]) we get:

$$V_1(y_1;\pi_1^{y_1}(dx_1)) = \inf_{\pi_2\in\Pi(\int_{x_1}\pi_1^{y_1}(dx_1)\mu^{x_1},\nu^{y_1})} \int \pi_2(dx_2,dy_2)[x_2-y_2]^2$$
$$= \int_0^1 \left[F^{-1}_{\nu^{y_1}}(u) - F^{-1}_{\int_{x_1}\pi_1^{y_1}(dx_1)\mu^{x_1}}(u)\right]^2 du,$$

$$V_0 = \inf_{\pi_1\in\Pi(p^1_*\mu,p^1_*\nu)}\Big\{\int_{x_1,y_1}\pi_1(dx_1,dy_1)[x_1-y_1]^2+\int_{y_1}\nu(dy_1)\int_0^1\left[F^{-1}_{\nu^{y_1}}(u)-F^{-1}_{\int_{x_1}\pi_1^{y_1}(dx_1)\mu^{x_1}}(u)\right]^2du\Big\}.$$

From this we see that the cost function in V_0 has indeed a non-linear behaviour in its last term, here in terms of π_1.

In order to prove Theorem 2.2.2, first, we introduce causal quasi-Markov transport plans and show that their associated optimal transport problem do possess a DPP. This fact (Theorem 2.2.5) together with Proposition 2.2.6 will then allow us to prove the desired result.

Definition 2.2.4. *A causal transport plan $\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ is called causal quasi-Markov, which we denote by $\pi \in \Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)$, if for every t*

$$\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1}) = \pi^{x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1}).$$

Of course $\Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)\neq\emptyset$ if and only if μ is a Markov measure. To wit, if $\pi\in\Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)$ then

$$\mu^{x_1,\dots,x_t}(dx_{t+1}) = \pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1}) = \pi^{x_t,y_1,\dots,y_t}(dx_{t+1}) = \mu^{x_1}(dx_{t+1}),$$

which implies the Markov property of μ. Conversely, if μ is a Markov measure, then the independent coupling of μ and ν belongs to $\Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)$. In either case, $\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1}) = \mu^{x_t}(dx_{t+1})$ must be satisfied.
Let us stress that the causal quasi-Markov constraint is not a linear constraint, and the set $\Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)$ of all these measures is not convex.

We introduce now the causal quasi-Markov transport problem:

$$\inf\left\{\int c\,d\pi,\ \pi\in\Pi^{\mathcal{F},\mathcal{F}}_{cqm}(\mu,\nu)\right\}. \tag{2.2.8}$$

The following theorem shows the DPP for (2.2.8):

Theorem 2.2.5 (DPP for (2.2.8)**).** *Let $\mu,\nu\in\mathcal{P}(\mathbb{R}^N)$, suppose that μ is Markov and that the cost is semiseparable in the sense that $c=\sum_t c_t(x_t,y_1,\dots,y_t)$ for non-negative Borel functions c_t. Then starting from $V^c_N := 0$ and defining recursively for $t=N,\dots,2$:*

$$\begin{aligned} V^c_{t-1}(y_1,\dots,y_{t-1};m(dx_{t-1})) \;:=&\\ \inf_{\pi_t\in\Pi\left(\int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)\,,\,\nu^{y_1,\dots,y_{t-1}}(dy_t)\right)} &\int \pi_t(dx_t,dy_t)\big\{c_t(x_t,y_1,\dots,y_t)\\ &+V^c_t(y_1,\dots,y_t;\pi_t^{y_t})\big\}, \end{aligned} \tag{2.2.9}$$

we have that

$$V_0^c := \inf_{\pi_1 \in \Pi(p^1_* \mu, p^1_* \nu)} \int \pi_1(dx_1, dy_1)\{c_1(x_1, y_1) + V_1^c(y_1; \pi_1^{y_1})\} \;=\; value(2.2.8).$$

Furthermore, each function V_{t-1}^c is jointly universally measurable and convex in its last component. If moreover, Assumption 2.1.6 holds and each c_t is l.s.c, then $V_{t-1}^c(y_1, \dots, y_{t-1}; \cdot)$ is l.s.c. and the infimum in (2.2.8) is attained.

To be precise, we shall prove that V_{t-1}^c is lower semianalytic (i.e. $\{V_{t-1}^c < r\}$ is analytic for each r; see [BS78, Definition 7.21]), which implies universal measurability. We recall that a set is analytic if it is a continuous image of a Borel set in a Polish space.

Proof. We split the proof in several steps. From the definition of the value function V_t^c we see that it is bounded from below and convex in its last component for any t. To prove convexity, we denote by π_1, π_2 optimizers for $V_t^c(y_1, \dots, y_t; m)$ and $V_t^c(y_1, \dots, y_t; \bar{m})$ respectively, and look at the linear combination of two value functions

$$\begin{aligned}
&\lambda V_t^c(y_1, \dots, y_t; m) + (1-\lambda) V_t^c(y_1, \dots, y_t; \bar{m}) = \\
&\lambda \mathbb{E}^{\pi_1}\left[\int \pi_{t+1}(dx_{t+1}, dy_{t+1})\big\{c_{t+1}(x_{t+1}, y_1, \dots, y_{t+1}) + V_{t+1}^c(y_1, \dots, y_{t+1}; \pi_{t+1}^{y_{t+1}})\big\}\right] + \\
&(1-\lambda) \mathbb{E}^{\pi_2}\left[\int \pi_{t+1}(dx_{t+1}, dy_{t+1})\big\{c_{t+1}(x_{t+1}, y_1, \dots, y_{t+1}) + V_{t+1}^c(y_1, \dots, y_{t+1}; \pi_{t+1}^{y_{t+1}})\big\}\right] = \\
&\mathbb{E}^{\lambda\pi_1 + (1-\lambda)\pi_2}\left[\int \pi(dx_{t+1}, dy_{t+1})\big\{c_{t+1}(x_{t+1}, y_1, \dots, y_{t+1}) + V_{t+1}^c(y_1, \dots, y_{t+1}; \pi_{t+1}^{y_{t+1}})\big\}\right] \geq \\
&V_t^c(y_1, \dots, y_t; \lambda m + (1-\lambda)\bar{m}),
\end{aligned}$$

where in the last inequality we used that the measure $\lambda\pi_1 + (1-\lambda)\pi_2$ is an element of the set $\Pi\Big(\int_{x_t} (\lambda m + (1-\lambda)\bar{m})\,(dx_t)\mu^{x_t}(dx_{t+1})\,,\, \nu^{y_1, \dots, y_t}(dy_{t+1})\Big)$. Hence, the convexity of the value function V_t^c follows (see e.g. [Vil08, Theorem 4.8]).

STEP 1: We first show that the sets

$$D_{t-1} := \big\{(y_1, \dots, y_{t-1}, m, \pi) \text{ s.t. } \pi \in \Pi\big(\int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)\,,\, \nu^{y_1, \dots, y_{t-1}}(dy_t)\big)\big\}$$

are Borel, so in particular analytic. Indeed, observe first that

$$\bar{D}_{t-1} := \{(p, q, \pi) \text{ s.t. } \pi \in \Pi\left(q(dx_t)\,,\, p(dy_t)\right)\}$$

is closed w.r.t. weak convergence of probability measures (this follows from compactness of the set $\Pi\left(q(dx_t)\,,\, p(dy_t)\right)$ and from Prokhorov's Theorem 1.1.2). So, if we denote by Ψ the map

$$(y_1, \dots, y_{t-1}, m, \pi) \mapsto \big(\nu^{y_1, \dots, y_{t-1}}(dy_t), \int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t), \pi\big),$$

we have $D_{t-1} = \Psi^{-1}(\bar{D}_{t-1})$. We know that the application $(y_1, \dots, y_{t-1}) \mapsto \nu^{y_1, \dots, y_{t-1}}(dy_t)$ is Borel measurable (see [Kal02, Lemma 1.40]), so it suffices to show that the map

$$m(dx_{t-1}) \mapsto \int\limits_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)$$

is Borel. For this, it suffices to show that for Borel sets $\{A_i, B_i\}_{i=1}^r$ the sets of the form

$$\{m : \int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(A_i) \in B_i \text{ for } i = 1, \ldots, r\}$$

are again Borel. Since for any i the functions $x_{t-1} \mapsto \tilde{g}_i(x_{t-1}) := \mu^{x_{t-1}}(A_i)$ are bounded Borel, everything ultimately boils down to proving that the sets

$$\left\{m : \int_{x_{t-1}} m(dx_{t-1})\tilde{g}_i(x_{t-1}) \in B_i\right\} \tag{2.2.10}$$

are Borel for any bounded function $\tilde{g}_i$, $i = 1, \ldots, r$, or equivalently that the functions

$$m(dx_{t-1}) \mapsto \int\limits_{x_{t-1}} m(dx_{t-1})g(x_{t-1}) \tag{2.2.11}$$

are Borel measurable. But this follows now from an application of the functional monotone class theorem (see e.g. [AB06, Theorem 15.13]).

For future use in Step 5, we observe that as soon as Assumption 2.1.6 holds, the sets

$$\mathcal{K} := \left\{(m, \pi) \text{ s.t. } \pi \in \Pi\Big(\int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t) \,,\, \nu^{y_1,\ldots,y_{t-1}}(dy_t)\Big)\right\},$$

which are generally Borel only (as fibers of the set D_{t-1}), are further closed. To see this, we take a sequence of elements $(m_n, \pi_n)_n \in \mathcal{K}$ that converges weakly to (m, π), and since for any $f \in C_b(\mathbb{R})$ the function $\int\limits_{x_t} \mu^{x_{t-1}}(dx_t)f(x_t) \in C_b(\mathbb{R})$, it follows that

$$\int\limits_{x_{t-1}} m_n(dx_{t-1}) \int\limits_{x_t} \mu^{x_{t-1}}(dx_t)f(x_t) \rightharpoonup \int\limits_{x_{t-1}} m(dx_{t-1}) \int\limits_{x_t} \mu^{x_{t-1}}(dx_t)f(x_t),\ n \to \infty.$$

STEP 2: Let us now prove that the recursion is well-defined; namely, that the involved integrated functions are defined at most except for a null-set w.r.t. the integral. To be precise, we will show that these functions are lower semianalytic, and so universally measurable, thus in particular their integrals are well defined w.r.t. any Borel measure. For that matter, we shall first prove that

$$y_1, \ldots, y_t, m(dx_t) \mapsto V_t^c(y_1, \ldots, y_t; m) \tag{2.2.12}$$

is lower semianalytic. By [Kal02, Lemma 1.40], we know that for the regular conditional kernel of a given measure $\pi(dx_t, dy_t)$, the application $y_t \mapsto \pi^{y_t}(dx_t)$ is Borel measurable, then so is the map $(y_1, \ldots, y_t, \pi) \mapsto (y_1, \ldots, y_t, \pi^{y_t}(dx_t))$. Therefore, by [BS78, Lemma 7.30(3)] (on the composition of l.s.a. and Borel maps) we deduce that $(y_1, \ldots, y_t, \pi) \mapsto V_t^c(y_1, \ldots, y_t; \pi^{y_t})$ is l.s.a for π participating in the infimum in (2.2.9), and in particular the integral is indeed well-defined.

To show that functions (2.2.12) are l.s.a. for any $t \le N$, we will proceed by induction and we start with V_{N-1}^c. The cost

$$(y_1, \ldots, y_{N-1}, m, \pi) \mapsto \int \pi(dx_N, dy_N)c_N(x_N, y_1, \ldots, y_N)$$

is Borel measurable and so in particular l.s.a. The proof is the same as for functions in

(2.2.11) with $\tilde{g}_1 = -[c_N \wedge s]$, and take $s \to +\infty$. We can write $V^c_{N-1}(y_1, \dots, y_{N-1}; m)$ as the infimum of this cost over the fiber of the set D_{N-1} at $(y_1, \dots, y_{N-1}, m)$. By Step 1 we know that the set D_{N-1} is analytic, and by [BS78, Proposition 7.47], the function V^c_{N-1} is l.s.a. on the projection of D_{N-1} onto its $(y_1, \dots, y_{N-1}, m)$ components. By reverse induction, let us suppose that V^c_t is l.s.a. and prove that so is V^c_{t-1}. The cost to consider is now

$$(y_1, \dots, y_{t-1}, m, \pi) \mapsto \int \pi(dx_t, dy_t) c_t(x_t, y_1, \dots, y_t) + \int \nu^{y_1,\dots,y_{t-1}}(dy_t) V^c_t(y_1, \dots, y_t; \pi^{y_t}_t),$$

the first term of which is Borel as before. The second term is l.s.a. by virtue of the inductive step, the discussion about composition of l.s.a. and Borel maps above, and by [BS78, Proposition 7.48] on the integration of l.s.a. functions with respect to Borel kernels. By [BS78, Lemma 7.30(4)] we get that their sum is l.s.a. and so by [BS78, Proposition 7.47], we conclude that V^c_{t-1} is l.s.a. as the infimum of a l.s.a. function over the fiber of an analytic set.

STEP 3: By the previous point, and the way we wrote (2.2.9) as an infimum of a l.s.a. cost over a fiber of an analytic set in Step 2, we can perform for any $\varepsilon > 0$ by [BS78, Proposition 7.50(b)] a selection

$$(y_1, \dots, y_{t-1}, m) \mapsto L^{y_1,\dots,y_{t-1};m}_{t-1,\varepsilon}(dx_t, dy_t) \in \Pi\left(\int_{x_{t-1}} m(dx_{t-1}) \mu^{x_{t-1}}(dx_t)\,,\, \nu^{y_1,\dots,y_{t-1}}(dy_t) \right),$$

so that the above mapping is universally measurable (i.e. measurable w.r.t. any completion of the corresponding Borel sets of its domain) and

$$\begin{aligned} V^c_{t-1}(y_1, \dots, y_{t-1}, m) + \varepsilon \geq \int L^{y_1,\dots,y_{t-1};m}_{t-1,\varepsilon}(dx_t, dy_t) \big[& c_t(x_t, y_1, \dots, y_t) \\ & + V^c_t(y_1, \dots, y_t; (L^{y_1,\dots,y_{t-1};m}_{t-1,\varepsilon})^{y_t}) \big], \end{aligned}$$

so each L is an ε-optimizer for the corresponding problem. We will now build a measure, whose successive kernels will solve the recursions (2.2.9) at each step, modulo an ε margin. Start with any ε-optimizer $\pi_{0,\varepsilon}(dx_1, dy_1)$ of V^c_0. Then take

$$y_1 \mapsto \pi^{y_1}_{1,\varepsilon}(dx_2, dy_2) := L^{y_1;\pi^{y_1}_{0,\varepsilon}}_{1,\varepsilon}(dx_2, dy_2),$$

which is universally measurable by the composition result [BS78, Proposition 7.44]. Inductively, if $(y_1, \dots, y_{t-1}) \mapsto \pi^{y_1,\dots,y_{t-1}}_{t-1,\varepsilon}(dx_t, dy_t)$ has been constructed in a universally measurable way, we build

$$(y_1, \dots, y_t) \mapsto \pi^{y_1,\dots,y_t}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}) := L^{y_1,\dots,y_t;(\pi^{y_1,\dots,y_{t-1}}_{t-1,\varepsilon})^{y_t}}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}),$$

which is universally measurable by [BS78, Proposition 7.44], the inductive step, and the fact that the function that associates a regular kernel to a product measure is Borel. This finishes the induction, and by construction

$$\pi^{y_1,\dots,y_t}_{t,\varepsilon}(dx_{t+1}, dy_{t+1}) \in \Pi\left(\int_{x_t} (\pi^{y_1,\dots,y_{t-1}}_{t-1,\varepsilon})^{y_t}(dx_t) \mu^{x_t}(dx_{t+1}), \nu^{y_1,\dots,y_t}(dy_{t+1}) \right), \quad (2.2.13)$$

and $\pi_{t,\varepsilon}^{y_1,\dots,y_t}$ attains $V_t^c(y_1,\dots,y_t;(\pi_{t-1,\varepsilon}^{y_1,\dots,y_{t-1}})^{y_t})$ except for an ε margin. We now define

$$(x_t,y_1,\dots,y_t)\mapsto \Gamma_{t,\varepsilon}^{x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1}) := \mu^{x_t}(dx_{t+1})(\pi_{t,\varepsilon}^{y_1,\dots,y_t})^{x_{t+1}}(dy_{t+1}),$$

which is universally measurable by [BS78, Proposition 7.44] again. By [BS78, Proposition 7.45] the successive concatenation of these kernel induces a unique Borel measure Γ_ε such that

$$\begin{aligned}\Gamma_\varepsilon(dx_1,\dots,dx_N,dy_1,\dots,dy_N) \;=\; & \pi_{0,\varepsilon}(dx_1,dy_1)\Gamma_{1,\varepsilon}^{x_1,y_1}(dx_2,dy_2)\dots\\ & \Gamma_{t,\varepsilon}^{x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1})\dots\Gamma_{N-1,\varepsilon}^{x_{N-1},y_1,\dots,y_{N-1}}(dx_N,dy_N).\end{aligned}$$

By construction, this measure is causal quasi-Markov and its x-marginal is exactly μ. As for the y-marginal, it is easy to see that $\Gamma_\varepsilon(dy_1)=\nu(y_1)$ and that

$$\begin{aligned}\Gamma_\varepsilon(dy_1,dy_2) &= \int_{x_1,x_2} \pi_{0,\varepsilon}(dx_1,dy_1)\mu^{x_1}(dx_2)(\pi_{1,\varepsilon}^{y_1})^{x_2}(dy_2)\\ &= \nu(dy_1)\int_{x_2}\Big(\int_{x_1}\pi_{0,\varepsilon}^{y_1}(dx_1)\mu^{x_1}(dx_2)\Big)(\pi_{1,\varepsilon}^{y_1})^{x_2}(dy_2)\\ &= \nu(dy_1)\int_{x_2}\pi_{1,\varepsilon}^{y_1}(dx_2)(\pi_{1,\varepsilon}^{y_1})^{x_2}(dy_2) = \nu(dy_1)\int_{x_2}\pi_{1,\varepsilon}^{y_1}(dx_2,dy_2)\\ &= \nu(dy_1)\nu^{y_1}(dy_2).\end{aligned}$$

Observe that inside the brackets in the second line is exactly the first marginal of $\pi_{1,\varepsilon}^{y_1}$, which follows from (2.2.13). Inductively, one can verify that Γ_ε has ν as its y-marginal.
STEP 4: We now prove the equality $V_0^c = value(2.2.8)$. By the previous step, $V_0^c + N\varepsilon \geq value(2.2.8)$, since Γ_ε is feasible for (2.2.8) and because by construction Γ_ε was designed to be ε-optimal at each step, so that overall $V_0^c+N\varepsilon\geq\int cd\Gamma_\varepsilon$. By letting $\varepsilon\to 0$, we get $V_0^c\geq value(2.2.8)$. On the other hand, given any $\pi\in\Pi_{cqm}^{\mathcal{F},\mathcal{F}}(\mu,\nu)$, we clearly have that

$$\pi^{y_1,\dots,y_t}(dx_{t+1},dy_{t+1})\in\Pi\Big(\int_{x_t}(\pi^{y_1,\dots,y_{t-1}})^{y_t}(dx_t)\mu^{x_t}(dx_{t+1})\,,\,\nu^{y_1,\dots,y_t}(dy_{t+1})\Big), \qquad (2.2.14)$$

and further we can write

$$\begin{aligned}\int c_t d\pi = \int \pi(dy_1,\dots,dy_{t-1})\pi^{y_1,\dots,y_{t-1}}(dx_t,dy_t)c_t(x_t,y_1,\dots,y_t) =\\ \int\pi(dx_1,dy_1)\pi^{y_1}(dx_2,dy_2)\pi^{y_1,y_2}(dx_3,dy_3)\dots\pi^{y_1,\dots,y_{t-1}}(dx_t,dy_t)c_t(x_t,y_1,\dots,y_t),\end{aligned}$$

so

$$\int cd\pi = \int\pi(dx_1,dy_1)\left\{c_1+\int\pi^{y_1}(dx_2,dy_2)\left\{c_2+\int\pi^{y_1,y_2}(dx_3,dy_3)\left\{c_3+\dots\right\}\right\}\right\}. \qquad (2.2.15)$$

Combining this with (2.2.14) we get $value(2.2.8)\geq V_0^c$ and conclude the desired equality.

From now on we take Assumption 2.1.6 for granted and assume l.s.c. costs.

STEP 5: We establish now that the function V_{t-1}^c is lower semicontinuous in its last argument (i.e. in m), the other ones being fixed. The desired lower semicontinuity could be proved by hand using convex combination of kernels as in Step 6 below, but a simpler argument is to invoke a "maximum theorem" as in [BS78, Proposition 7.33], which establishes the lower semicontinuity of a value function as long as the cost is jointly lower semicontinuous (in our case w.r.t. (m,π), since the y's are fixed, and this clearly holds) and as soon as the constraint set is a fiber of a closed set (which is true by the last observation in Step 1). The lower semicontinuity in (m,π) is proved by reverse induction, much as in Step 2, and for [BS78, Proposition 7.33] one can assume that the π's live a priori in a compact space, since varying the m's in a tight set only will only move the constraint set inside a larger tight set in the product.

STEP 6: We now prove that each of the problems in (2.2.9) is attained. The $\int c_t d\pi$ part of the cost is l.s.c. and linear, so it causes no trouble and we may assume $c_t \equiv 0$. We take a minimizing sequence π_n for (2.2.9):

$$V^c_{t-1}(y_1,\dots,y_{t-1};m) = \lim_n \int \nu^{y_1,\dots,y_{t-1}}(dy_t)V^c_t(y_1,\dots,y_t;\pi^{y_t}_n).$$

Trivially the sequence $\int_{y_t} \nu^{y_1,\dots,y_{t-1}}(dy_t)\pi^{y_t}_n(dx_t)$ is tight, since it is identically equal to the measure $\int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)$. By [Bal, Theorem 3.15] we conclude that there is a subsequence of kernels $\{y_t \mapsto \pi^{y_t}_n\}_n$, which is again denoted by $\pi^{y_t}_n$, and a kernel $\{y_t \mapsto \pi^{y_t}_t\}$ such that the sequence of Césaro averages

$$y_t \mapsto \tilde{\pi}^{y_t}_n := \frac{1}{n}\sum_{i\leq n} \pi^{y_t}_i$$

converges weakly, for $\nu^{y_1,\dots,y_{t-1}}(dy_t)$-a.e. y_t, to the kernel $y_t \mapsto \pi^{y_t}$ (i.e. as measures on x_t, see [Bal, Definition 3.10]), and further, by [Bal, Corollary 3.14], it holds that

$$\int \nu^{y_1,\dots,y_{t-1}}(dy_t)\pi^{y_t}_t(dx_t) = \int_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t),$$

so the measure $\pi := \nu^{y_1,\dots,y_{t-1}}(dy_t)\pi^{y_t}_t(dx_t)$ is feasible for (2.2.9). All in all it follows

$$\begin{aligned} V^c_{t-1}(y_1,\dots,y_{t-1};m) &= \liminf_n \frac{1}{n}\sum_{i\leq n} \int \nu^{y_1,\dots,y_{t-1}}(dy_t)V^c(y_1,\dots,y_t;\pi^{y_t}_i) \\ &\geq \liminf_n \int \nu^{y_1,\dots,y_{t-1}}(dy_t)V^c(y_1,\dots,y_t;\tilde{\pi}^{y_t}_n) \\ &\geq \int \nu^{y_1,\dots,y_{t-1}}(dy_t)\liminf_n V^c(y_1,\dots,y_t;\tilde{\pi}^{y_t}_n) \\ &= \int \nu^{y_1,\dots,y_{t-1}}(dy_t)V^c(y_1,\dots,y_t;\pi^{y_t}_t), \end{aligned}$$

by convexity, Fatou's Lemma (remember the V^c's are bounded below), and the lower semicontinuity established in the previous step. All in all, we conclude that π is feasible and optimal.

STEP 7: By the previous step, we may now go back to Step 3 and find by [BS78, Proposition 7.50(b)] a selection of optimizers at each time (as opposed to only ε-optimizers) $L^{y_1,\dots,y_t;m}_t$. Then we can redo Step 3 with $\varepsilon = 0$, building a global measure Γ which exactly solves the recursion and is feasible for (2.2.8), and so is optimal for it. □

We would like to point out that even if the cost function c does not have the stated semiseparable structure, a look at the proof shows that the recursion is well-posed and that its value gives an upper bound for $value$(2.2.8) (see Step 4 of the proof). The semiseparable structure is crucial only for the lower bound (see (2.2.15)).

As we have mentioned in Theorem 2.2.2, each optimization problem in DPP is of the form of general (non-linear) transports on the line. Under Assumption 2.2.8 and for l.s.c. cost functions each of the value functions in (2.2.9) is convex in the kernel and jointly l.s.c. Assuming additionally that all the regular kernels of the measures μ, ν are compactly supported, we can use the duality result of Gozlan et al. [GRST15, Theorem 9.5]. Denote $\eta(dx_t) = \int\limits_{x_{t-1}} m(dx_{t-1})\mu^{x_{t-1}}(dx_t)$, and

$$\bar{c}_t^{y_1,\dots,y_{t-1}}(y_t;p) = \int\limits_{x_t} c_t(x_t, y_1, \dots, y_t)p(dx_t) + V_t^c(y_1, \dots, y_{t-1}, y_t; p(dx_t)).$$

Then the following dual formulation holds:

$$V_{t-1}^c(y_1, \dots, y_{t-1}; m(dx_{t-1})) = \sup_{\phi}\left\{\int R_{\bar{c}_t^{y_1,\dots,y_{t-1}}}\phi(y_t)\nu^{y_1,\dots,y_{t-1}}(dy_t) - \int \phi(x_t)\eta(dx_t)\right\}, \tag{2.2.16}$$

where

$$R_{\bar{c}_t^{y_1,\dots,y_{t-1}}}\phi(y_t) = \inf_{p\in\mathcal{P}(\mathbb{R})} \left\{\int \phi(x_t)p(dx_t) + \bar{c}_t^{y_1,\dots,y_{t-1}}(y_t;p)\right\}.$$

As a by-product of the previous proof, using the arguments from Step 6 therein, one can prove attainability of general transport problems as in [GRST15] in the presence of lower semicontinuity and convexity of the cost w.r.t. the kernel, without the assumption that the state space be compact or discrete-like.

We finally establish a sufficient condition for (2.2.8) and (Pc) to be equivalent:

Proposition 2.2.6. *Let μ be Markov and the cost be semiseparable. Then $value$(2.2.8) = $value$(Pc). Moreover, if (Pc) is attained, then there is some optimal causal transport which is further causal quasi-Markov.*

Proof. We prove that under the given assumptions any causal plan has a related causal quasi-Markov plan which incurs in the same cost. Let $\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu)$, and build

$$d\hat{\pi} = \pi(dx_1, dy_1)\pi^{x_1,y_1}(dx_2, dy_2)\pi^{x_2,y_1,y_2}(dx_3, dy_3)\dots\pi^{x_{N-1},y_1,\dots,y_{N-1}}(dx_N, dy_N).$$

As $\hat{\pi}^{x_1,\dots,x_t,y_1,\dots,y_t} = \pi^{x_t,y_1,\dots,y_t} = \hat{\pi}^{x_t,y_1,\dots,y_t}$, we have that $\hat{\pi}$ is quasi-Markov and its x-marginal is μ. In fact,

$$\begin{aligned}
\hat{\pi}^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1}) &= \pi^{x_t,y_1,\dots,y_t}(dx_{t+1}) \\
&= \int_{x_1,\dots,x_{t-1}} \pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1})\pi^{x_t,y_1,\dots,y_t}(dx_1, \dots, dx_{t-1}) \\
&= \int_{x_1,\dots,x_{t-1}} \mu^{x_t}(dx_{t+1})\pi^{x_t,y_1,\dots,y_t}(dx_1, \dots, dx_{t-1}) = \mu^{x_t}(dx_{t+1}),
\end{aligned}$$

so it is also causal.
We finally prove that for every function of the form $H := H(x_t, y_1, \dots, y_t)$, the expected

value w.r.t. π and $\hat{\pi}$ coincides.

$$
\begin{aligned}
\int H d\hat{\pi} &= \int H \pi(dx_1, dx_2, dy_1, dy_2) \pi^{x_2,y_1,y_2}(dx_3, dy_3) \hat{\pi}^{x_3,y_1,y_2,y_3}(dx_4, \dots, dy_4, \dots) \\
&= \int H \pi(dx_2, dy_1, dy_2) \pi^{x_2,y_1,y_2}(dx_3, dy_3) \hat{\pi}^{x_3,y_1,y_2,y_3}(dx_4, \dots, dy_4, \dots) \\
&= \int H \pi(dx_2, dx_3, dy_1, dy_2, dy_3) \pi^{x_3,y_1,y_2,y_3}(dx_4, dy_4) \hat{\pi}^{x_4,y_1,y_2,y_3,y_4}(dx_5, \dots, dy_5, \dots) \\
&= \int H \pi(dx_3, dy_1, dy_2, dy_3) \pi^{x_3,y_1,y_2,y_3}(dx_4, dy_4) \hat{\pi}^{x_4,y_1,y_2,y_3,y_4}(dx_5, \dots, dy_5, \dots) \\
\dots &= \int H \pi(dx_{t-1}, dy_1, \dots, dy_{t-1}) \pi^{x_{t-1},y_1,\dots,y_{t-1}}(dx_t, dy_t) \;=\; \int H d\pi.
\end{aligned}
$$

This shows first that π and $\hat{\pi}$ incur in the same cost under the semiseparability assumption, and second that the y-marginal of $\hat{\pi}$ is ν, which finally establishes $\hat{\pi} \in \Pi_{cqm}^{\mathcal{F},\mathcal{F}}(\mu, \nu)$. □

All in all, the proof of Theorem 2.2.2 is now trivial:

Proof of Theorem 2.2.2. By Proposition 2.2.6, $value$(2.2.8) $=$ $value$(Pc), and by Theorem 2.2.5 we have the recursion and all the other properties. □

We would like to remark that with Proposition 2.2.6 we can conclude that, whenever the measure μ is Markov and the cost is l.s.c. and semiseparable, the problem (2.2.8) has a solution. Indeed, since (Pc)=(2.2.8) and by [Las15, Corollary 1] the former is attained, one constructs also an optimizer for the latter. Our Theorem 2.2.5 gives the same result with a stronger assumption in a self-contained way, through DPP.

2.2.2 Bicausal transport problem

The focus of this section is on bicausal transport problems. Our next main result, given in Theorem 2.2.13, establishes the equivalence of (Pc) and

$$
\inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int c d\pi. \tag{Pbc}
$$

under Condition 2.2.22 below. Furthermore, it identifies the causal optimizer of (Pc) for convex separable costs (as well as clarifies when these are Monge maps), in a setting relevant for future applications. We finish this section with the dynamic programming principle for (Pbc) which is much more tractable in comparison with the causal DPP.

First, we start with an equivalent characterization of bicausal transport plans and duality in this setting. Observe that in this case the condition on the second marginal and its kernels is simplified comparing to the case of causal transport plans. The proof is the same as in Lemma 2.2.1, so we omit it.

Proposition 2.2.7. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^N)$. If $\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu, \nu)$ is decomposed as in* (2.2.2)*, then the following conditions on the kernels hold:*

(i) $\bar{\pi} \in \Pi(p_*^1\mu, p_*^1\nu)$, and

(ii) successively for $t < N$ and for π-almost every $x_1, \dots, x_t, y_1, \dots, y_t$ it holds that

$$
\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1}, dy_{t+1}) \in \Pi(\mu^{x_1,\dots,x_t}(dx_{t+1}), \nu^{y_1,\dots,y_t}(dy_{t+1})).
$$

Conversely, given regular kernels

$$\bar{\pi}(dx_1, dy_1), \pi^{x_1,y_1}(dx_2, dy_2), \ldots, \pi^{x_1,\ldots,x_{N-1},y_1,\ldots,y_{N-1}}(dx_N, dy_N)$$

satisfying the properties (i) − (ii), the measure π constructed as in (2.2.2) *belongs to* $\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)$.

The analogue of Theorem 2.1.10 is obtained for the bicausal setting. As in the causal case, we define the sets

$$\mathbb{F}' := \left\{ \begin{array}{c} F' : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \text{ s.t. } F'(x_1,\ldots,x_N,y_1,\ldots,y_N) = \\ \sum\limits_{t<N} h_t(y_1,\ldots,y_t)\left[g_t(x_1,\ldots,x_N) - \int g_t(x_1,\ldots,x_t,x_{t+1},\ldots,x_N)\mu^{x_1,\ldots,x_t}(dx_{t+1},\ldots,dx_N)\right] + \\ \sum\limits_{t<N} h'_t(x_1,\ldots,x_t)\left[g'_t(y_1,\ldots,y_N) - \int g'_t(y_1,\ldots,y_t,y_{t+1},\ldots,y_N)\nu^{y_1,\ldots,y_t}(dy_{t+1},\ldots,dy_N)\right], \\ \text{with } h_t, h'_t \in B_b(\mathbb{R}^t),\ g_t, g'_t \in C_b(\mathbb{R}^N) \text{ for all } t<N \end{array} \right\}, \tag{2.2.17}$$

$$\mathbb{S}' := \left\{ \begin{array}{c} S' : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \text{ s.t. } S'(x_1,\ldots,x_N,y_1,\ldots,y_N) = \\ \sum\limits_{t<N} H_t(y_1,\ldots,y_t)\left[M_{t+1}(x_1,\ldots,x_{t+1}) - M_t(x_1,\ldots,x_t)\right] + \\ \sum\limits_{t<N} H'_t(x_1,\ldots,x_t)\left[M'_{t+1}(y_1,\ldots,y_{t+1}) - M'_t(y_1,\ldots,y_t)\right], \\ \text{with } H_t, H'_t \in B_b(\mathbb{R}^t), M_t, M'_t \in C_b(\mathbb{R}^t) \text{ for all } t<N, \text{ and with } M \text{ a martingale} \end{array} \right\}. \tag{2.2.18}$$

The following strengthened version of Assumption 2.1.6 will be needed:

Assumption 2.2.8. *Both μ and ν are successively weakly continuous.*

Under the above assumption, mimicking the proofs of Lemma 2.1.7 and Lemma 1.3.2, one gets the same continuity results for the test functions in the sets $\mathbb{F}'$ and $\mathbb{S}'$. This further implies that the set of bicausal transport plans $\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ is weakly closed, hence compact. Bicausal duality can be seen as a corollary of Theorem 2.1.10. The duality theorem for bicausal transport plans was first obtained in [PP12, Theorem 7.2].

Corollary 2.2.9. *Suppose that $c : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is lower semicontinuous and bounded from below, and that Assumption 2.2.8 holds. Then there is no duality gap:*

$$\inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int c d\pi = \sup_{\substack{\phi,\psi \in C_b(\mathbb{R}^N), F' \in \mathbb{F}' \\ \phi \oplus \psi \leq c + F'}} \left[\int \phi d\mu + \int \psi d\nu\right] = \sup_{\substack{\phi,\psi \in C_b(\mathbb{R}^N), S' \in \mathbb{S}' \\ \phi \oplus \psi \leq c + S'}} \left[\int \phi d\mu + \int \psi d\nu\right]. \tag{Dbc}$$

Moreover, the infimum in the l.h.s. is attained.

The Knothe-Rosenblatt rearrangement defined below plays a central role in this work. A priori this transformation of measures has nothing to do with optimal transport, and was independently proposed by Rosenblatt [Ros52] for statistical purposes and by Knothe [Kno57] for applications to geometric inequalities. Section 2.3 is devoted to its closer analysis. Further we denote F_{η_1} the distribution of $p^1_*\eta$ whenever η is a measure in multiple dimensions.

Definition 2.2.10. *The increasing N-dimensional **Knothe-Rosenblatt** rearrangement*[1] *of μ and ν is defined as the law of the random vector $(X_1^*, \dots, X_N^*, Y_1^*, \dots, Y_N^*)$ where*

$$\begin{aligned} X_1^* &= F_{\mu_1}^{-1}(U_1), \qquad & Y_1^* &= F_{\nu_1}^{-1}(U_1), \qquad && \text{and inductively} \\ X_t^* &= F_{\mu^{X_1^*,\dots,X_{t-1}^*}}^{-1}(U_t), \qquad & Y_t^* &= F_{\nu^{Y_1^*,\dots,Y_{t-1}^*}}^{-1}(U_t), \qquad && \text{for } t = 2, \dots, N, \end{aligned} \tag{2.2.19}$$

for $U_1, \dots, U_N$ independent and uniformly distributed random variables on $[0,1]$. Additionally, if μ-a.s. all the conditional distributions of μ are atomless (e.g. if μ has a density), then this rearrangement is induced by the (Monge) map

$$(x_1, \dots, x_N) \mapsto T(x_1, \dots, x_N) := (T^1(x_1), T^2(x_2; x_1), \dots, T^N(x_N; x_1, \dots .x_{N-1})),$$

where

$$\begin{aligned} T^1(x_1) :=&\ F_{\nu_1}^{-1} \circ F_{\mu_1}(x_1), \\ T^t(x_t; x_1, \dots, x_{t-1}) :=&\ F_{\nu^{T^1(x_1),\dots,T^{t-1}(x_{t-1};x_1,\dots,x_{t-2})}}^{-1} \circ F_{\mu^{x_1,\dots,x_{t-1}}}(x_t), \quad t \geq 2. \end{aligned} \tag{2.2.20}$$

The following example shows that the assumption on the measure μ having no atoms is necessary in order to guarantee that the Knothe-Rosenblatt map indeed matches the marginals.

Example 2.2.11. Let $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$. Then the distribution function that corresponds to the measure μ is

$$F_\mu(x) = \begin{cases} 0, & x \in (-\infty, -1)\,, \\ \frac{1}{2}, & x \in [\,-1, 1\,)\,, \\ 1, & x \in [\,1, \infty\,)\,. \end{cases}$$

Its inverse is

$$F_\mu^{-1}(x) = \begin{cases} -1, & t \in \left[0.\frac{1}{2}\right], \\ 1, & x \in \left(\frac{1}{2}, 1\right]. \end{cases}$$

Let $\nu = \frac{1}{3}(\delta_{-1} + \delta_0 + \delta_1)$. Then

$$F_\nu(y) = \begin{cases} 0, & y \in (-\infty, -1)\,, \\ \frac{1}{3}, & y \in [\,-1, 0\,)\,, \\ \frac{2}{3}, & y \in [\,0, 1\,)\,, \\ 1, & y \in [\,1, \infty\,)\,, \end{cases}$$

$$F_\nu^{-1}(y) = \begin{cases} -1, & t \in \left[0.\frac{1}{3}\right], \\ 0, & t \in \left(\frac{1}{3}, \frac{2}{3}\right], \\ 1, & t \in \left(\frac{2}{3}, 1\right]. \end{cases}$$

The increasing rearrangement $T(x) = F_\nu^{-1}(F_\mu(x))$ does not push forward μ to ν, since

[1] The reader might know it by the name *quantile transform* or *Knothe-Rosenblatt coupling.*

for example $\nu(0) = \frac{1}{3}$, which is not the same as

$$\mu(T^{-1}(0)) = \mu\left(F_\mu^{-1}\left(F_\nu(0)\right)\right) = \mu\left(F_\mu^{-1}\left(\tfrac{2}{3}\right)\right) = \mu(1) = \tfrac{1}{2}.$$

Remark 2.2.12. We reassure the reader that the Knothe-Rosenblatt rearrangements in the form (2.2.19) are always bicausal, and in the Monge form (2.2.20) this is also the case as soon as all conditional distributions of the source measure are atomless. We illustrate the argument for (2.2.19) as follows. For each bounded Borel $g(\cdot,\cdot)$ we define

$$y_1 \mapsto G(y_1) := \textstyle\int_0^1 g(y_1, F^{-1}_{\nu^{y_1}}(v))dv,$$

and denote $X_1^* := F^{-1}_{\mu_1}(U_1)$, $Y_1^* := F^{-1}_{\nu_1}(U_1)$ and $Y_2^* := F^{-1}_{\nu^{F^{-1}_{\nu_1}(U_1)}}(U_2) = F^{-1}_{\nu^{Y_1^*}}(U_2)$. Thus

$$G(Y_1^*) = \mathbb{E}[g(Y_1^*, Y_2^*)\,|\,Y_1^*],$$

since Y_1^* is U_1-measurable and U_1 is independent of U_2. We hence get

$$\begin{aligned}
E[f(X_1^*)g(Y_1^*, Y_2^*)] &= \textstyle\int_0^1\int_0^1 f(F^{-1}_{\mu_1}(u_1))g\Big(F^{-1}_{\nu_1}(u_1), F^{-1}_{\nu^{F^{-1}_{\nu_1}(u_1)}}(u_2)\Big)du_2du_1\\
&= \textstyle\int_0^1 f(F^{-1}_{\mu_1}(u_1))\left[\int_0^1 g\Big(F^{-1}_{\nu_1}(u_1), F^{-1}_{\nu^{F^{-1}_{\nu_1}(u_1)}}(u_2)\Big)du_2\right]du_1\\
&= \textstyle\int_0^1 f(F^{-1}_{\mu_1}(u_1))G\left(F^{-1}_{\nu_1}(u_1)\right)du_1\\
&= \mathbb{E}[\,f(X_1^*)\,G(Y_1^*)]\\
&= \mathbb{E}[\,\mathbb{E}[f(X_1^*)|Y_1^*]\,G(Y_1^*)]\\
&= \mathbb{E}[\,\mathbb{E}[f(X_1^*)|Y_1^*]\,\mathbb{E}[g(Y_1^*, Y_2^*)\,|\,Y_1^*]\,]\\
&= \mathbb{E}[\,\mathbb{E}[f(X_1^*)|Y_1^*]\,g(Y_1^*, Y_2^*)].
\end{aligned}$$

We have just established that for all Borel f : $\mathbb{E}[f(X_1^*)|Y_1^*, Y_2^*] = \mathbb{E}[f(X_1^*)|Y_1^*]$. Thus, the law of X_1^* given (Y_1^*, Y_2^*) is equal to the law of X_1^* given Y_1^*. The same holds inverting the roles of μ and ν, by symmetry, and also going to greater time indices. As for (2.10), the argument actually follows directly upon noticing for example that for $\mu-$a.e. x_1 the measure $[F_{\mu^{x_1}}]_*\mu^{x_1}$ is equal to the Lebesgue measure on $[0,1]$, under the given assumptions. We can then apply the previous arguments with uniform random variables etc. Even though these considerations are not explicit in [Rüs85], they underpin some of the results therein.

Now we are ready to present our next main result on the equality between causal and bicausal transport problem, and as a consequence of this, the characterization of optimal causal transport plans.

Theorem 2.2.13. *Assume that c is l.s.c. bounded from below and has a separable structure*

$$c(x_1, \dots, x_N, y_1, \dots, y_N) = \sum_{t\le N} c_t(x_t, y_t). \tag{2.2.21}$$

Further suppose that the starting measure μ is the product of its marginals, i.e.

$$\mu(dx_1, \ldots, dx_N) = \mu_1(dx_1) \ldots \mu_N(dx_N). \tag{2.2.22}$$

Then the values of (Pc) *and* (Pbc) *coincide. If moreover, $c_t(x_t, y_t) = c_t(x_t - y_t)$ and c_t is convex* [2]*, then a solution to* (Pc) *is given by the Knothe-Rosenblatt rearrangement. Additionally, if each μ_i is atomless (e.g. if it has a density), then this rearrangement is induced by the Monge map* (2.2.20).

In [Las13, Lemma 5] it was shown that the optimal solution to the causal transport problem in Wiener Space, with Cameron-Martin cost and Wiener measure as source, is automatically bicausal. The results above give the causal/bicausal equality and existence as well as characterization of the Monge solution to what can be thought of as "finite-dimensional projections" of that problem. As we will show further in Chapter 3, causal/bicausal equality in continuous-time is a result of the Cameron-Martin cost being written in terms of "speed" and the source measure having independent increments.

Throughout this work we mostly focus on convex costs of the difference, and as a result we only deal with the increasing Knothe-Rosenblatt rearrangement. Evidently, for concave costs the correct notion would be to take the decreasing rearrangement defined in the obvious way.

The proof of Theorem 2.2.13 rests on the result, which shows that under Conditions (2.2.23) the Knothe-Rosenblatt coupling is an optimal bicausal transport plan. In [Rüs85, Corollary 2] this result was shown under the assumption of "monotone regression dependence". For $N = 2$ it means that, for each t, $x \mapsto F_{\mu^x}(t)$ and $y \mapsto F_{\nu^y}(t)$ are simultaneously either decreasing or increasing. Conditions (2.2.23), stated below, is slightly more general, encompassing at the same time the independence condition (2.2.22) and Rüschendorf's "monotone regression dependence".

Proposition 2.2.14. *For each $t = 1, \ldots, N-1$, all $(x_1, \ldots, x_{t-1}), (y_1, \ldots, y_{t-1})$, and $u \in \mathbb{R}$, suppose one of the following holds:*

(a) Either $F_{\mu^{x_1,\ldots,x_{t-1},\cdot}}(u)$ or $F_{\nu^{y_1,\ldots,y_{t-1},\cdot}}(u)$ is constant; or (2.2.23)

(b) The functions $F_{\mu^{x_1,\ldots,x_{t-1},\cdot}}(u)$ and $F_{\nu^{y_1,\ldots,y_{t-1},\cdot}}(u)$ are monotone and have the same type of monotonicity.

Assume further that

$$c(x_1, \ldots, x_N, y_1, \ldots, y_N) := \sum_{t \le N} c_t(x_t - y_t),$$

where each c_t is convex. Then a solution to (Pbc) *is given by the Knothe-Rosenblatt rearrangement* (2.2.19)*. Additionally, if μ-a.s. all the conditional distributions of μ are atomless (e.g. if μ has a density), then this rearrangement is induced by the Monge map determined by* (2.2.20).

[2] Here we mean $c_t(x_t, y_t) = h_t(x_t - y_t)$ and redefine c_t as h_t. This "univariate" structure is not indispensable for the result: it can be replaced by a more general Spence-Mirrlees condition.

First, we would like to give the illustration when Conditions (2.2.23) hold.

Example 2.2.15. Let X_1, X_2 be independent and Y_1, Y_2 be arbitrary random variables. Let E be independent of the previous 4 random variables. Define $X_3 := X_2 + E$ and $Y_3 := Y_2 + E$. Denote $\mu = Law(X_1, X_2, X_3)$ and $\nu = Law(Y_1, Y_2, Y_3)$. Then already at time $t = 1$ Condition (b) fails since $F_{\mu^{\cdot}}(u)$ is constant by independence of X_1 and X_2, and on the other hand $F_{\mu^{\cdot}}(u)$ is arbitrary, however (a) holds. But at time $t = 2$ Rüschsendorf's condition of monotone regression dependence (b) does hold, whereas (a) does not. All in all, Conditions 2.2.23 hold here.

Before providing the proof of Proposition 2.2.14, first we prove the following lemma.

Lemma 2.2.16. *For any convex function $c : \mathbb{R} \to \mathbb{R}$ and any functions $f, g : \mathbb{R} \to \mathbb{R}$ that have the same type of monotonicity, and for any pairs $x \le \bar{x}$ and $y \le \bar{y}$, the following holds*

$$c(f(x) - g(y)) + c(f(\bar{x}) - g(\bar{y})) \le c(f(\bar{x}) - g(y)) + c(f(x) - g(\bar{y})) \tag{2.2.24}$$

Proof. It is known that any convex function has a monotone slope, meaning that for any $h \ge 0$ the function

$$\frac{c(z) - c(z-h)}{h},$$

is increasing in z. This implies that for any $z \le \bar{z}$

$$c(z) - c(z-h) \le c(\bar{z}) - c(\bar{z} - h).$$

If both functions f and g are non-decreasing, we take $z := f(x) - g(y)$ and $\bar{z} := f(\bar{x}) - g(y)$, which gives us $h = g(\bar{y}) - g(y) \ge 0$ and $z \le \bar{z}$. For non increasing functions f and g, we write

$$c(f(x) - g(y)) = c(-[-f(x) - (-g(y))]),$$

and conclude as before. □

Proof of Proposition 2.2.14. We start with $N = 2$. From classical optimal transport (see [Vil03]) it is known that

$$\inf_{\pi_2 \in \Pi(\mu^{x_1}, \nu^{y_1})} \int c_2(x_2 - y_2)\pi_2(dx_2, dy_2) = \int c_2(F^{-1}_{\mu^{x_1}}(u_2) - F^{-1}_{\nu^{y_1}}(u_2))du_2.$$

As we will see in Proposition 2.2.20, the bicausal transport problem is equivalent to

$$\inf_{\pi_1 \in \Pi(p^1_*\mu, p^1_*\nu)} \int \pi_1(dx_1, dy_1) \left[c_1(x_1 - y_1) + \inf_{\pi_2 \in \Pi(\mu^{x_1}, \nu^{y_1})} \int c_2(x_2 - y_2)\pi_2(dx_2, dy_2) \right].$$

Therefore, the function in square brackets above equals

$$V(x_1, y_1) := c_1(x_1 - y_1) + \int c_2(F^{-1}_{\mu^{x_1}}(u_2) - F^{-1}_{\nu^{y_1}}(u_2))du_2. \tag{2.2.25}$$

For every fixed value u_2, and for any $x_1 \le \bar{x}_1$ and $y_1 \le \bar{y}_1$, we have under Conditions (2.2.23) (a) or (b) and from Lemma 2.2.16 the following

$$\begin{aligned}
&c_1(x_1-y_1)+c_1(\bar{x}_1-\bar{y}_1)-c_1(\bar{x}_1-y_1)-c_1(x_1-\bar{y}_1)+\\
&c_2(F^{-1}_{\mu^{x_1}}(u_2)-F^{-1}_{\nu^{y_1}}(u_2))+c_2(F^{-1}_{\mu^{\bar{x}_1}}(u_2)-F^{-1}_{\nu^{\bar{y}_1}}(u_2))-\\
&c_2(F^{-1}_{\mu^{\bar{x}_1}}(u_2)-F^{-1}_{\nu^{y_1}}(u_2))-c_2(F^{-1}_{\mu^{x_1}}(u_2)-F^{-1}_{\nu^{\bar{y}_1}}(u_2))\leq 0.
\end{aligned}\tag{2.2.26}$$

Integrating against u_2 yields

$$V(x_1,y_1)+V(\bar{x}_1,\bar{y}_1)\leq V(\bar{x}_1,y_1)+V(x_1,\bar{y}_1).$$

By classical transport, the optimal coupling between $p^1_*\mu$ and $p^1_*\nu$ is then the increasing monotone rearrangement. From the beginning of the proof, given (x_1,y_1), the optimal coupling between μ^{x_1} and ν^{y_1} is likewise the increasing monotone rearrangement. Iterating our arguments for general N the result holds. $\square$

Proof of Theorem 2.2.13. The result follows from Proposition 2.2.14, if the causal/ bicausal equality is proved. We stress that Condition (2.2.23)(a), needed for Proposition 2.2.14, is satisfied in the present case because the conditional distribution functions associated to μ do not depend on the conditioning argument. We now prove the causal/bicausal equality. Start with $\pi\in\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ and decompose it as in (2.2.2). From Proposition 2.1.5, we know that the following conditions ($t<N$) are satisfied by the kernels $\bar{\pi}(dx_1,dy_1),\pi^{x_1,y_1}(dx_2,dy_2)$, up to $\pi^{x_1,\dots,x_{N-1},y_1,\dots,y_{N-1}}(dx_N,dy_N)$:

$$\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1},\mathbb{R})=\mu_{t+1}(dx_{t+1})\tag{2.2.27}$$

and

$$\int\limits_{x_1,\dots,x_t}\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(\mathbb{R},dy_{t+1})\pi^{y_1,\dots,y_t}(dx_1,\dots,dx_t)\;=\;\nu^{y_1,\dots,y_t}(dy_{t+1})=\pi^{y_1,\dots,y_t}(\mathbb{R},dy_{t+1}).\tag{2.2.28}$$

We can rewrite (2.2.27) in the following way:

$$\int\limits_{x_1,\dots,x_t}\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1},\mathbb{R})\pi^{y_1,\dots,y_t}(dx_1,\dots,dx_t)\;=\;\mu_{t+1}(dx_{t+1})=\pi^{y_1,\dots,y_t}(dx_{t+1},\mathbb{R}).\tag{2.2.29}$$

Therefore, we can construct a new plan $\tilde{\pi}$ as follows

$$\tilde{\pi}(dx_1,\dots,dx_N,dy_1,\dots,dy_N)=\bar{\pi}(dx_1,dy_1)\pi^{y_1}(dx_2,dy_2)\dots\pi^{y_1,\dots,y_{N-1}}(dx_N,dy_N),$$

with

$$\pi^{y_1,\dots,y_t}(dx_{t+1},dy_{t+1})=\int\limits_{x_1,\dots,x_t}\pi^{x_1,\dots,x_t,y_1,\dots,y_t}(dx_{t+1},dy_{t+1})\pi^{y_1,\dots,y_t}(dx_1,\dots,dx_t).\tag{2.2.30}$$

Due to (2.2.28) and (2.2.29), one can see that for any $t < N$ each kernel

$$\pi^{y_1,\dots,y_t} \in \Pi(\mu_{t+1}(dx_{t+1}), \nu^{y_1,\dots,y_t}(dy_{t+1})).$$

Then, from Proposition (2.2.7), we know that $\tilde{\pi} \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)$. Writing down the minimization over the kernels (2.2.30), we have exactly (Pbc). Since the kernels of $\tilde{\pi}$ and π are connected via (2.2.28) and (2.2.29), one can observe that the value of the bicausal transport problem is less than or equal to the causal one, giving us the desired result. $\square$

The explicit form of the optimizers in Theorem 2.2.13 is obviously only true for $\mathbb{R}$-valued processes. Interestingly, if the transport problem consisted in mapping (in a bicausal way) $\mathbb{R}^d$-valued process, the recursive structure is just as in the case $d = 1$ and so if μ were the product of its marginals (each of them in $\mathcal{P}(\mathbb{R}^d)$ now) we would get a similar conclusion with the role of the monotone rearrangements taken by general Brenier maps, under suitable conditions.

Corollary 2.2.17. *Assume that c is l.s.c. bounded from below and is of the form*

$$c(x_1,\dots,x_N,y_1,\dots,y_N) = c_1(x_1,y_1) + \sum_{1<t\leq N} c_t(x_t - x_{t-1}, y_t - y_{t-1}), \tag{2.2.31}$$

and that the source measure μ has independent increments. Then the values of (Pc) *and* (Pbc) *coincide. If moreover, $c_t(a,b) = c_t(a-b)$ and c_t is convex, then a solution to* (Pc) *is given by the Knothe-Rosenblatt rearrangement* (2.2.19)*, and if additionally each μ_i is atomless, then this rearrangement is induced by the Monge map* (2.2.20)*.*

Proof. First observe that for the map

$$\Theta : (x_1,\dots,x_N) \mapsto (x_1, x_2 - x_1, \dots, x_N - x_{N-1}), \tag{2.2.32}$$

the following equality holds

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \Big(c_1(x_1,y_1) + \sum_{1<t\leq N} c_t(x_t - x_{t-1}, y_t - y_{t-1})\Big) d\pi(x_1,\dots,x_N,y_1,\dots,y_N) = \tag{2.2.33}$$

$$\inf_{\tilde{\pi}\in\Pi^{\mathcal{F},\mathcal{F}}(\Theta_*\mu,\Theta_*\nu)} \int \sum_{t=1}^{N} c_t(x_t,y_t) d\tilde{\pi}(x_1,\dots,x_N,y_1,\dots,y_N). \tag{2.2.34}$$

Having recognized that the measure $\Theta_*\mu$ has independent components, the causal/bicausal equality follows from Theorem 2.2.13. Moreover, it also implies that under the given assumptions on c_t, a solution to (2.2.34) is given by the Knothe-Rosenblatt rearrangement between the measures $\Theta_*\mu$ and $\Theta_*\nu$, denoted by $\tilde{\pi}^{KR}$. Then we claim, that $\pi^{KR} := \Theta_*^{-1}\tilde{\pi}^{KR}$ is a solution to (2.2.33), with

$$\Theta^{-1} : (x_1,\dots,x_N) \mapsto (x_1, x_1 + x_2, \dots, x_1 + \dots + x_N).$$

Indeed, we easily see that $\pi^{KR} \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$. Moreover, we observe that π^{KR} is again of the Knothe-Rosenblatt form. By the equality (2.2.33)/(2.2.34) we conclude.

$\square$

Previously we have discussed that the values of the causal and bicausal problems coincide in the case when the starting measure μ is the product of its marginals and the cost function has a separable structure. The following two examples, solved numerically using linear programming solver, show that dropping either assumption causes this equality to break down.

Example 2.2.18. Take

$$\mu = 0.16\,\delta_{(1,1)} + 0.24\,\delta_{(1,-1)} + 0.24\,\delta_{(-1,1)} + 0.36\,\delta_{(-1,-1)},$$

$$\nu = 0.25\,\delta_{(1,1)} + 0.25\,\delta_{(1,-1)} + 0.25\,\delta_{(-1,1)} + 0.25\,\delta_{(-1,-1)}.$$

Here μ is the product of its marginals, but the cost function $c(x_1, x_2, y_1, y_2) = \mathbb{I}_{(x_1,x_2)\neq(y_1,y_2)}$ is non-separable. Then an optimal causal transport plan is

$$\begin{aligned}\pi_c = {} & 0.16\ \pi((1,1);(1,1)) && + 0.24\,\pi((1,-1);(1,-1)) && + 0.03\,\pi((-1,1);(1,1)) \\ & + 0.01\ \pi((-1,1);(1,-1)) && + 0.2\ \ \pi((-1,1);(-1,1)) && + 0.06\,\pi((-1,-1);(1,1)) \\ & + 0.05\ \pi((-1,-1);(-1,1)) && + 0.25\,\pi((-1,-1);(-1,-1)),\end{aligned}$$

giving the optimal causal value 0.15, and an optimal bicausal one

$$\begin{aligned}\pi_{bc} = {} & 0.16\,\pi((1,1);(1,1)) && + 0.04\,\pi((1,-1);(1,1)) && + 0.2\,\pi((1,-1);(1,-1)) \\ & + 0.04\,\pi((-1,1);(1,1)) && + 0.2\ \ \pi((-1,1);(-1,1)) && + 0.01\,\pi((-1,-1);(1,1)) \\ & + 0.05\,\pi((-1,-1);(1,-1)) && + 0.05\,\pi((-1,-1);(-1,1)) && + 0.25\,\pi((-1,-1);(-1,-1)),\end{aligned}$$

giving the value 0.19.

We stress that in case the independent marginals Condition (2.2.23)(a) does not hold, and even if the assumption of "monotone regression dependence" (b) is true, as the following example shows, the values of causal and bicausal transport problems may differ.

Example 2.2.19. Consider a quadratic-separable cost function and define μ, failing to be the product of its marginals, and ν as:

$$\begin{aligned}\mu &= 0.18\,\delta_{(1,2)} + 0.24\,\delta_{(1,0)} + 0.18\,\delta_{(1,-2)} + 0.08\,\delta_{(-1,2)} + 0.12\,\delta_{(-1,0)} + 0.2\,\delta_{(-1,-2)}, \\ \nu &= 0.1\,\delta_{(1,2)} + 0.26\,\delta_{(1,-2)} + 0.16\,\delta_{(-1,2)} + 0.48\,\delta_{(-1,-2)}.\end{aligned}$$

An optimal causal plan is

$$\begin{aligned}\pi_c = {} & 0.144\,\pi((1,0);(1,-2)) && + 0.008\,\pi((1,0);(-1,2)) && + 0.088\,\pi((1,0);(-1,-2)) \\ & + 0.1\ \ \ \pi((1,2);(1,2)) && + 0.008\,\pi((1,2);(1,-2)) && + 0.072\,\pi((1,2);(-1,2)) \\ & + 0.108\,\pi((1,-2);(1,-2)) && + 0.072\,\pi((1,-2);(-1,-2)) && + 0.12\ \ \pi((-1,0);(-1,-2)) \\ & + 0.08\ \ \pi((-1,2);(-1,2)) && + 0.2\ \ \ \ \pi((-1,-2);(-1,-2)),\end{aligned}$$

giving the value 2.528, and a bicausal optimal one

$$\begin{aligned}\pi_{bc} = {} & 0.144\,\pi((1,0);(1,-2)) && + 0.096\,\pi((1,0);(-1,-2)) && + 0.1\ \ \ \pi((1,2);(1,2)) \\ & + 0.008\,\pi((1,2);(1,-2)) && + 0.06\ \ \pi((1,2);(-1,2)) && + 0.012\,\pi((1,2);(-1,-2))\end{aligned}$$

$$+ 0.108\,\pi((1,-2);(1,-2)) \quad + 0.072\,\pi((1,-2);(-1,-2)) + 0.02\ \ \pi((-1,0);(-1,2))$$
$$+ 0.1\quad \pi((-1,0);(-1,-2)) + 0.08\ \ \pi((-1,2);(-1,2)) \quad + 0.2\ \ \pi((-1,-2);(-1,-2)),$$

giving the value 2.72.

The previous example also shows us that Rüschendorf's monotone regression condition, which holds here, is insufficient to guarantee the equality between (Pc) and (Pbc) even for separable costs.

Dynamic programming principle: Bicausality

We now give the DPP for the bicausal case, which appeared in [Rüs85, Theorem 3] and [PP14]. We will prove it with different methods, obtaining further attainability and some regularity of the value function. Observe that no Markovianity of μ or separability of the costs is needed, and that the value functions in DPP (and recursion itself) are more tractable than in the causal quasi-Markov case (2.2.9).

Proposition 2.2.20. *Given a Borel bounded from below cost function c, we have that the recursive optimization problem*

$$\inf_{\pi_1\in\Pi(p^1_*\mu,p^1_*\nu)} \int \pi_1(dx_1,dy_1) \inf_{\pi_2\in\Pi(\mu^{x_1},\nu^{y_1})} \int \pi_2(dx_2,dy_2)\dots$$
$$\dots \inf_{\pi_N\in\Pi(\mu^{x_1,\dots,x_{N-1}},\nu^{y_1,\dots,y_{N-1}})} \int \pi_N(dx_N,dy_N)c(x_1,\dots,x_N,y_1,\dots,y_N) \quad \text{(Dyn-Pbc)}$$

is well-defined, namely the successive integrals

$$V^c_t(x_1,\dots,x_t,y_1,\dots,y_t) = \inf_{\pi_{t+1}\in\Pi(\mu^{x_1,\dots,x_t},\nu^{y_1,\dots,y_t})} \int \pi_{t+1}(dx_{t+1},dy_{t+1})V^c_{t+1}(x_1,\dots,x_{t+1},y_1,\dots,y_{t+1}) \quad (2.2.35)$$

are well-defined, and the values of (Dyn-Pbc), (Pbc) *and*

$$V^c_0 := \inf_{\pi_1\in\Pi(p^1_*\mu,p^1_*\nu)} \int V^c_1(x_1,y_1)\pi_1(dx_1,dy_1)$$

coincide. If further c is l.s.c. and Assumption 2.2.8 holds, then there is a bicausal optimizer and the value functions V^c_t are all l.s.c.

Proof. Since $(x_1,\dots,x_t,y_1,\dots,y_t)\mapsto(\mu^{x_1,\dots,x_t},\nu^{y_1,\dots,y_t})$ is Borel and the set $\{(p,q,\pi) : \pi\in\Pi(p,q)\}$ is closed, we get that

$$D_t := \{(x_1,\dots,x_t,y_1,\dots,y_t,\pi) : \pi\in\Pi(\mu^{x_1,\dots,x_t},\nu^{y_1,\dots,y_t})\}$$

is Borel. As in the proof of Theorem (2.2.5), we start by proving recursively that the functions V^c_t are lower semianalytic (l.s.a). As in Step 2 therein, one observes first that

$$(x_1,\dots,x_{N-1},y_1,\dots,y_{N-1},\pi)\mapsto \int \pi(dx_N,dy_N)c(x_1,\dots,x_N,y_1,\dots,y_N)$$

is Borel, and so l.s.a. Since $V^c_{N-1}(x_1,\dots,x_{N-1},y_1,\dots,y_{N-1})$ is the infimum of this function over the fiber of D_{N-1} at $(x_1,\dots,x_{N-1},y_1,\dots,y_{N-1})$, we get by [BS78, Proposition 7.47] that V^c_{N-1} is l.s.a. and in particular universally measurable. The recursive step is obvious, and we get this result for each V^c_t, therefore the integrals in (2.2.35) are well-defined. Now take, as in Step 3 of the proof of Theorem (2.2.5), a universally measurable selection of ε-optimizers for V^c_t; call these $\pi^{x_1,\dots,x_t,y_1,\dots,y_t}_{t,\varepsilon}$. Build then Γ_ε as the unique Borel measure that results from concatenating these ([BS78, Proposition 7.45]). By construction, each Γ_ε is in $\Pi(\mu,\nu)$ (see Proposition 2.2.7) and $V^c_0 + N\varepsilon \geq \int c d\Gamma_\varepsilon \geq value(\text{Pbc})$, and so $V^c_0 \geq value(\text{Pbc})$. The reverse inequality follows directly from Proposition 2.2.7.

If Assumption 2.2.8 holds, then D_t is closed, and so if further c is l.s.c. we argue as in Step 5 of the proof of Theorem (2.2.5), proving that the value functions are l.s.c. as well. In turn, applying now [BS78, Proposition 7.50(b)] we can obtain a universally measurable selection of optimizers $\pi^{x_1,\dots,x_t,y_1,\dots,y_t}_t$ and with them we build an optimal bicausal transport Γ (just like the previous paragraph with $\varepsilon = 0$). □

As we mentioned before, in [Rüs85, Theorem 3] the validity of a dynamic programming principle is already obtained, as well as the measurability of the value function (the proof does not provide full arguments and it should be complemented with [LV14, Proposition 3.6]). In addition, our proof of Proposition 2.2.20 delivers further semicontinuity of the value function at every intermediate step, via self-contained arguments.

Since at each step of the dynamic programming principle (Dyn-Pbc), or equivalently in (2.2.35), we have a usual optimal transport problem, we can write these in their dual formulation; for simplicity we assume that c is l.s.c. and non-negative (see however [BS11] for an extension). In effect, the value function at time t for each $t = N-1, N-2, \dots, 1$ is obtained recursively as:

$$\begin{aligned}
&V^c(t; x_1,\dots,x_t,y_1\dots,y_t) := \\
&\sup_{\substack{\phi_{t+1},\psi_{t+1}\in C_b(\mathbb{R}),\\ \phi_{t+1}(x_{t+1})+\psi_{t+1}(y_{t+1}) \leq V^c(t+1;x_1,\dots,x_{t+1},y_1,\dots,y_{t+1})}} \Big\{ \textstyle\int \phi_{t+1}(x_{t+1})\mu^{x_1,\dots,x_t}(dx_{t+1}) \\
&\qquad\qquad + \textstyle\int \psi_{t+1}(y_{t+1})\nu^{y_1,\dots,y_t}(dy_{t+1}) \Big\},
\end{aligned}$$

so the value of (Dyn-Pbc) is also given by

$$V^c(0) := \sup_{\substack{\phi_1,\psi_1\in C_b(\mathbb{R})\\ \phi_1(x_1)+\psi_1(y_1)\leq V^c(1;x_1,y_1)}} \int \phi_1(x_1)p^1_*\mu(dx_1) + \int \psi_1(y_1)p^1_*\nu(dy_1).$$

Observe that this recursive structure of the dual problem is not obvious from (Dbc), but can be guessed a posteriori thanks to the apparent "primal" recursive structure. A simpler picture of (Dyn-Pbc) arises when c has a separable structure as in (2.2.21); see [PP12, Theorem 14].

We close the present subsection with a preliminary illustration of how the causal DPP can be used in practice; here we show a condition giving the equivalence between (Pc) and (Pbc):

Corollary 2.2.21. *Consider $N = 2$ and a separable cost of the form*

$$c = c_1(x_1, y_1) + |x_2 - y_2|,$$

with c_1 l.s.c. and bounded from below. Suppose that for every z, y_1 one of the following two cases holds

$$\text{either } [\forall x_1 : F_{\mu^{x_1}}(z) \geq F_{\nu^{y_1}}(z)] \quad \text{or} \quad [\forall x_1 : F_{\mu^{x_1}}(z) \leq F_{\nu^{y_1}}(z)] .$$

Then (Pc) *is equivalent to* (Pbc).

Proof. Borrowing from Theorem 2.2.2, we call

$$V_1(y_1, \pi^{y_1}) = \inf_{m \in \Pi(\int \pi^{y_1}(dx_1)\mu^{x_1}, \nu^{y_1})} \int |x_2 - y_2| m(dx_2, dy_2),$$

which by e.g. [Vil03, eq.(2.48), p.75] equals $\int_z \left| F_{\int \pi^{y_1}(dx_1)\mu^{x_1}}(z) - F_{\nu^{y_1}}(z) \right| dz$. Observing that $F_{\int\limits_{x_1} \pi^{y_1}(dx_1)\mu^{x_1}}(z) = \int\limits_{x_1} \pi^{y_1}(dx_1) F_{\mu^{x_1}}(z)$, by our technical assumption

$$\begin{aligned}
\int\limits_z \left| F_{\int\limits_{x_1} \pi^{y_1}(dx_1)\mu^{x_1}}(z) - F_{\nu^{y_1}}(z) \right| dz &= \int\limits_z \left| \int\limits_{x_1} \pi^{y_1}(dx_1) F_{\mu^{x_1}}(z) - F_{\nu^{y_1}}(z) \right| dz \\
&= \int\limits_z \int\limits_{x_1} \pi^{y_1}(dx_1) \left| F_{\mu^{x_1}}(z) - F_{\nu^{y_1}}(z) \right| dz. \qquad (2.2.36)
\end{aligned}$$

Together with Fubini's Theorem, we get

$$V_1(y_1, \pi^{y_1}) = \int_{x_1} \pi^{y_1}(dx_1) \inf_{m \in \Pi(\mu^{x_1}, \nu^{y_1})} \int |x_2 - y_2| dm.$$

Then, from Theorem 2.2.2 we obtain that

$$Value(\text{Pc}) = \inf_{\pi \in \Pi(p^1_* \mu, p^1_* \nu)} \int \pi(dx_1, dy_1) \left\{ c_1(x_1, y_1) + \inf_{m \in \Pi(\mu^{x_1}, \nu^{y_1})} \int |x_2 - y_2| dm \right\},$$

which as we saw in Proposition 2.2.20, equals the bicausal DPP, hence, we conclude. □

In classical optimal transport it is known that the quadratic cost optimizer between two Gaussian measures is always Gaussian (see [DL82]). The following example shows us that in case of causal optimal transport we can guarantee the same result only under the assumption that the starting measure has independent marginals. W.l.o.g. we can assume that the measures are centered.

Example 2.2.22. Let $X = (X_1, X_2) \sim \mathcal{N}(0, diag(d_1, d_2)) = \mu$, $Y = (Y_1, Y_2) \sim \mathcal{N}(0, \Sigma_Y) = \nu$, where $\Sigma_Y = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. We want to transport μ to ν in a causal way. Let π be any causal transference plan that does this transportation (we likewise denote π any causal coupling (X, Y) of measures μ and ν). We observe that

$$Cov(X_2, Y_1) = \mathbb{E}[X_2 Y_1] = \mathbb{E}\left(X_2 \mathbb{E}[Y_1 \mid X_1, X_2]\right) = \mathbb{E}\left(X_2 \mathbb{E}[Y_1 \mid X_1]\right) = 0,$$

using the causality property and assumption of independence of marginals. So we conclude that the covariance matrix of all causal plans π has the following structure

$$\Sigma = \begin{pmatrix} d_1 & 0 & a_{11} & a_{12} \\ 0 & d_2 & 0 & a_{22} \\ a_{11} & 0 & \sigma_{11} & \sigma_{12} \\ a_{12} & a_{22} & \sigma_{12} & \sigma_{22} \end{pmatrix}, \tag{2.2.37}$$

where $a_{11} = cov(X_1, Y_1), a_{12} = cov(X_1, Y_2), a_{22} = cov(X_2, Y_2)$. Having this covariance structure is hence necessary, but possibly not sufficient for causality. However, if we assume that the vector (X_1, X_2, Y_1, Y_2) has multivariate Gaussian distribution, we can actually find the conditional laws $Law\Big(Y_1 \mid X_1, X_2\Big)$ and $Law\Big(Y_1 \mid X_1\Big)$, which are Gaussian as well. Namely,

$$\Big(Y_1 \mid X_1 = x_1, X_2 = x_2\Big) \sim \mathcal{N}\Big(\bar{\mu}, \bar{\Sigma}\Big),$$

where $\bar{\mu} = \Big(x_1 a_{11} d_2 + x_2 a_{12} d_1\Big)/d_1 d_2$ and $\bar{\Sigma} = \sigma_{11} - \Big(a_{11}^2 d_2 + a_{12}^2 d_1)\Big)/d_1 d_2$. From the other hand,

$$\Big(Y_1 \mid X_1 = x_1\Big) \sim \mathcal{N}\Big(\mu', \Sigma'\Big),$$

with $\mu' = a_{11} d_1^{-1} x_1$ and $\Sigma' = \sigma_{11} - a_{11}^2 d_1^{-1}$. Since two Gaussian distributions coincide whenever they have the same means and variances, we get that the Gaussian coupling (X_1, X_2, Y_1, Y_2) is causal if and only if $a_{12} = 0$, and so the covariance structure Σ is necessary and sufficient in this case, showing us at the same time that the causal optimizer is jointly Gaussian.

Minimizing the quadratic cost $\mathbb{E}\left[|X_1 - Y_1|^2 + |X_2 - Y_2|^2\right]$ over all causal couplings of μ and ν, and observing that the cost can clearly be written in terms of the covariance matrix of the coupling, we conclude from the previous paragraph that we are entitled to consider joint Gaussian couplings only.

2.3 Discussion on the Knothe-Rosenblatt rearrangement

In this section we answer the question: to what extent the Knothe-Rosenblatt rearrangement is canonical? Here we show how it is characterized as the only *increasing* transport (in a precise sense) which is also bicausal. In Section 2.4 we will further highlight the importance of this rearrangement by viewing it in light of a modern tenant of optimal transport theory.

As we have seen, the Knothe-Rosenblatt rearrangement (2.2.19)-(2.2.20) appears quite naturally in our setting. In light of Remark 2.2.12, we would like to characterize it as the unique bicausal transport plan with a desirable "increasingness" property. The correct concept turns out to be that of increasing triangular transformations, found in [BKM05], which we recall here:

Definition 2.3.1. *A map $T : \mathbb{R}^N \to \mathbb{R}^N$ is an increasing triangular transformation (in short ITT) with source μ if for each $t \in \{1, \dots, N\}$ there is a Borel function*

$$\mathbb{R}^t \ni (x_1, \dots, x_t) \mapsto T^t(x_t; x_1, \dots, x_{t-1}) \in \mathbb{R},$$

such that $T^t(\cdot\,; x_1, \dots, x_{t-1})$ is non-decreasing[3] *for μ-a.e. $(x_1, \dots, x_{t-1})$, and further*

$$T(x_1, \dots, x_N) \;=\; (T^1(x_1), T^2(x_2; x_1), \dots, T^N(x_N; x_1, \dots, x_{N-1})), \;\; \mu - a.s.$$

The terminology comes from the fact that a differentiable (for example, linear) triangular transformation has a Jacobian matrix in triangular form. Clearly the Knothe-Rosenblatt map (2.2.20) is an ITT. As Lemma 2.3.2 shows, given that all conditional distributions of μ and ν are atomless, there is a canonical ITT that pushes μ into ν which is unique up to μ-negligible sets: the Knothe-Rosenblatt map.

Lemma 2.3.2 ([BKM05])**.** *Let μ be a Borel probability measure on $\mathbb{R}^N$. Let $T = (T_1, \dots, T_N)$ and $S = (S_1, \dots, S_N)$ be increasing triangular Borel mappings such that $T_*\mu = S_*\mu$ and for each n the mapping $(T_1, \dots, T_n)$ is injective on a Borel subset of full measure with respect to the projection of μ onto $\mathbb{R}^n$. Then $T(x) = S(x)$ $\mu-a.e.$*
In particular, if the projections of the measures μ and ν onto $\mathbb{R}^n$ are absolutely continuous, then there exists a canonical triangular mapping - the Knothe-Rosenblatt map, which is unique up to $\mu-$equivalence in the class of increasing Borel triangular mappings transforming μ into ν.

Under the same conditions on μ and ν (absolute continuity of kernels), this characterization has a geometric counterpart (see e.g. [San15, Chapter 2.3, Remark 2.20]): the Knothe-Rosenblatt map is the unique map which is increasing in the lexicographical order of $\mathbb{R}^N$ and pushes μ into ν. We remind, $x < y$ in lexicographical order if there exists $k \in \{1, \dots, N\}$ s.t. $x_j = y_j$ for $j < k$, and $x_k < y_k$. Under the same assumptions [CGS10] obtains the Knothe-Rosenblatt map as a natural limit of Brenier maps[4].

The assumption that the target measure ν possesses atomless conditional disintegrations is essential for uniqueness.

Example 2.3.3. [BKM05] Let μ be Lebesgue measure on the square $[0,1]^2$ and let ν be the Lebesgue measure on the unit interval of the second coordinate line. One can find at least two maps that transport μ to ν. Consider $T(x_1, x_2) = (0, x_2)$ and $S(x_1, x_2) = (0, S_2)$, where

$$S_2(x_1, x_2) = \begin{cases} \frac{(x_2+1)}{2}, & 0 \le x_1 \le \frac{1}{2}, \\ \frac{x_2}{2}, & \frac{1}{2} < x_1 \le 1. \end{cases}$$

Both of them are ITT, however, only T is the Knothe-Rosenblatt rearrangement - unique bicausal increasing transformation between μ and ν. Observing that

$$\mathbb{P}\left[X_1 \le x_1 | Y_1 = 0, Y_2 = \tfrac{x_2}{2}\right] = 2\min(1, x_1) - 1 \ne \mathbb{P}\left[X_1 \le x_1 | Y_1 = 0\right] = \mathbb{P}\left[X_1 \le x_1\right] = x_1,$$

[3] For $t = 1$ one should understand T^1 as a non-decreasing function of x_1 only.

[4] Beware that [San15, CGS10] read coordinates from N to 1 rather than from 1 to N as we do.

the reader can convince himself/herself that the map S is indeed not bicausal, hence is not Knothe-Rosenblatt.

The same idea can be used to prove existence of infinitely many increasing triangular transformations for the given marginals (defining S_2 through a different rule on x_1 and adjusting its value correspondingly). Notice that neither T nor S are lexicographically increasing transformations. An example of the latter in this case can be $L(x_1, x_2) = (0, L_2)$, where

$$L_2(x_1, x_2) = \begin{cases} \frac{x_2}{2}, & 0 \leq x_1 \leq \frac{1}{2}, \\ \frac{(x_2+1)}{2}, & \frac{1}{2} < x_1 \leq 1. \end{cases}$$

The lexicographically increasing map, when it exists, is an increasing triangular transformation. If the measures μ, ν and their conditional distributions are atomless, then there exists a unique canonical triangular transformation of μ into ν ([BKM05]), and the three objects, Knothe-Rosenblatt rearrangement, lexicographically increasing map and increasing triangular transformation, coincide.

We now ask ourselves: what distinguishes/characterizes the Knothe-Rosenblatt map without any assumption on ν? Here is the first result:

Proposition 2.3.4. *Suppose all conditional distributions of μ are atomless. Then the Knothe-Rosenblatt map from μ into ν, defined in* (2.2.20)*, is the unique bicausal transport plan between μ and ν which is induced by an increasing triangular transformation.*

Proof. Let π be a bicausal transport plan between μ and ν. By Proposition 2.2.7.(ii) we know that $\pi^{x_1,\dots,x_{t-1},y_1,\dots,y_{t-1}} \in \Pi(\mu^{x_1,\dots,x_{t-1}}, \nu^{y_1,\dots,y_{t-1}})$. If π is further induced by an ITT (say T), this means that $T^t(\cdot; x_1, \dots, x_{t-1})$ is pushing $\mu^{x_1,\dots,x_{t-1}}$ into $\nu^{y_1,\dots,y_{t-1}}$ in an increasing way (of course, under the understanding that $(y_1, \dots, y_{t-1}) = (T^1, \dots, T^{t-1})(x_1, \dots, x_{t-1})$). This immediately shows that $T^1 = F_{\nu_1}^{-1} \circ F_{\mu_1}$ except at most in a μ-null set. A straightforward induction argument proves that μ-a.s. the Knothe-Rosenblatt map and T are equal. $\square$

The relevance of this result is that, as Example 2.3.3 reveals, when ν has atoms in its conditional distributions then there may be many ITT's from μ into ν. The same counterexample proves that the Knothe-Rosenblatt map need not be increasing in lexicographical order in this case, so it cannot be characterized in terms of lexicographical order in this generality. Furthermore, under this pathology of ν, the Knothe-rosenblatt map need not be the natural limit of Brenier maps (see [San15, Example 2.26]). So, as far as we know, Proposition 2.3.4 is the only robust characterization of the Knothe-Rosenblatt map. We finally stress that this result can be extended to the case where μ and ν are arbitrary (so the Knothe-Rosenblatt rearrangement (2.2.19) need not be induced by a map) provided one generalizes the definition of ITT to transport plans:

Definition 2.3.5. *$\pi \in \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N)$ is an increasing triangular transport if $\pi(dx_1, dy_1)$, as well as $\pi^{x_1,\dots,x_{t-1},y_1,\dots,y_{t-1}}(dx_t, dy_t)$ for π-almost every $(x_1, \dots, x_{t-1}, y_1, \dots, y_{t-1})$, all have monotone support as bivariate measures.*[5]

[5] A set $\Gamma \subset \mathbb{R}^2$ is called monotone if $(x, y), (\bar{x}, \bar{y}) \in \Gamma$ and $x < \bar{x}$ implies $y \leq \bar{y}$.

Proposition 2.3.6. *The Knothe-Rosenblatt rearrangement from μ into ν, defined in* (2.2.19)*, is the unique bicausal transport plan between μ and ν which is an increasing triangular transport.*

The proof is essentially the same as for Proposition 2.3.4, so we omit it.

2.3.1 Explicit examples of Knothe-Rosenblatt rearrangement

The following example aims to show that the Knothe-Rosenblatt rearrangement between two Gaussians always defines a jointly Gaussian transport plan.

Example 2.3.7. Consider two vectors $X = (X_1, X_2) \sim \mathcal{N}(0, diag(d_1, d_2))$ and $Y = (Y_1, Y_2) \sim \mathcal{N}(0, \Sigma)$, with $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$. Note, that X_1 and X_2 are independent, and the Knothe-Rosenblatt coupling between the vectors X and Y in this case is defined as

$$\begin{aligned} X_1^* &= F_{\mu_1}^{-1}(U_1), \quad Y_1^* = F_{\nu_1}^{-1}(U_1), \\ X_2^* &= F_{\mu_2}^{-1}(U_2), \quad Y_2^* = F_{\nu^{Y_1^*}}^{-1}(U_2), \end{aligned} \tag{2.3.1}$$

with $\mu_1 = \mathcal{N}(0, d_1)$, $\mu_2 = \mathcal{N}(0, d_2)$ and $\nu_1 = \mathcal{N}(0, \sigma_{11})$. Now, replace the independent uniform U_i's by independent standard normals $V_i's$, i.e. $V_1 = \Phi^{-1}(U_1), V_2 = \Phi^{-1}(U_2)$, where Φ is the cdf of a standard normal random variable. Then we have the following representations

$$X_1 = d_1 V_1, \; X_2 = d_2 V_2, \; Y_1 = \sigma_{11} V_1.$$

The conditional distribution of Y_2 given $Y_1 = y_1$ is $\mathcal{N}\left(\frac{\sigma_{12}}{\sigma_{22}} y_1, \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}}\right)$. And so,

$$Y_2 | Y_1 = \tfrac{\sigma_{12}\sigma_{11}}{\sigma_{22}} V_1 + \tfrac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} V_2 = c_1 V_1 + c_2 V_2.$$

Note that c_2 does not depend on V_1, because the conditional distribution of Y_2 given Y_1 is normal with mean depending on Y_1, but variance is independent of Y_1. Having this representations, the joint Gaussianity of the vector (X_1, X_2, Y_1, Y_2) is trivial.

Now we give a detailed example on the construction of the Knothe-Rosenblatt rearrangement between two given measures.

Example 2.3.8. Let $(G_1, G_2) \sim \mathcal{N}(0, I_2)$ be a bivariate normally distributed random vector. Define $X_1 = G_1, X_2 = G_1 + G_2$ and $Y_1 = \text{sign}(X_1), Y_2 = \text{sign}(X_2)$. Let μ and ν be the joint laws of the vectors (X_1, X_2) and (Y_1, Y_2), i.e.

$$\mu = \mathcal{L}(X_1, X_2) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \right),$$

and

$$\nu = \mathcal{L}(Y_1, Y_2) = \tfrac{3}{8}\delta_{(1,1)} + \tfrac{1}{8}\delta_{(1,-1)} + \tfrac{1}{8}\delta_{(-1,1)} + \tfrac{3}{8}\delta_{(-1,-1)}.$$

Since μ-a.s. all the conditional distributions of μ are atomless, then the Knothe-Rosenblatt rearrangement between μ and ν is induced by the Monge map and defined as

$$\begin{aligned} Y_1^* = T^1(x_1) &:= F_{\nu_1}^{-1} \circ F_{\mu_1}(x_1), \\ Y_2^* = T^2(x_2; x_1) &:= F_{\nu^{Y_1^*}}^{-1} \circ F_{\mu^{x_1}}(x_2). \end{aligned} \tag{2.3.2}$$

F_{μ_1} is standard normal distribution. The first marginal of ν is

$$\nu_1 = \tfrac{1}{2}\delta_1 + \tfrac{1}{2}\delta_{-1},$$

and its distribution function is

$$F_{\nu_1}(y_1) = \begin{cases} 0, & y_1 \in (-\infty, -1\,], \\ \frac{1}{2}, & y_1 \in [\,-1, 1\,), \\ 1, & y_1 \in [\,1, \infty\,). \end{cases}$$

which has generalized inverse

$$F_{\nu_1}^{-1}(t) = \begin{cases} -1, & t \in (\,0, \frac{1}{2}\,], \\ 1, & t \in (\,\frac{1}{2}, 1\,]. \end{cases}$$

Therefore,

$$T_1(X_1^*) := Y_1^* = \begin{cases} -1, & F_{\mu_1}(x_1) \in (\,0, \frac{1}{2}\,] \\ 1, & F_{\mu_1}(x_1) \in (\,\frac{1}{2}, 1\,] \end{cases} = \begin{cases} -1, & X_1^* \leq 0 \\ 1, & X_1^* \geq 0 \end{cases} = \begin{cases} -1, & G_1 \leq 0 \\ 1, & G_1 \geq 0. \end{cases}$$

The conditional distribution of μ given X_1^* is

$$\mu^{x_1} = \mathcal{L}(X_2^* | X_1^* = x_1) = \mathcal{L}(G_1 + G_2 | X_1^* = x_1) = \mathcal{N}(x_1, 1).$$

Let us find the conditional distribution function $F_{\nu^{Y_1^*}}$.
If $Y_1^* = 1$, then the conditional cdf is

$$F_{\nu^{Y_1^*=1}}(y_2) = \begin{cases} 0, & y_2 \in (-\infty, -1\,] \\ \frac{1}{4}, & y_2 \in [\,-1, 1\,) \\ 1, & y_2 \in [\,1, \infty\,), \end{cases}$$

and its inverse is

$$F_{\nu^{Y_1^*=1}}^{-1}(t) = \begin{cases} -1, & t \in (\,0, \frac{1}{4}\,], \\ 1, & t \in (\,\frac{1}{4}, 1\,]. \end{cases}$$

If $Y_1^* = -1$, then

$$F_{\nu^{Y_1^*=-1}}(y_2) = \begin{cases} 0, & y_2 \in (-\infty, -1\,] \\ \frac{3}{4}, & y_2 \in [\,-1, 1\,) \\ 1, & y_2 \in [\,1, \infty\,), \end{cases}$$

and its inverse is

$$F^{-1}_{\nu^{Y_1^*=-1}}(t) = \begin{cases} -1, & t \in (\,0, \frac{3}{4}\,] \\ 1, & t \in (\,\frac{3}{4}, 1\,] \end{cases} .$$

Combining these together, we get

$$\begin{aligned} Y_2^* &= F^{-1}_{\nu^{Y_1^*}}(F_{\mu^{x_1}}(x_2)) \\ &= \begin{cases} -1, & Y_1^* = -1,\, F_{\mu^{x_1}}(x_2) \in (\,0, \frac{3}{4}\,] \text{ or } Y_1^* = 1,\, F_{\mu^{x_1}}(x_2) \in (\,0, \frac{1}{4}\,], \\ \;\;1, & Y_1^* = -1,\, F_{\mu^{x_1}}(x_2) \in (\,\frac{3}{4}, 1\,] \text{ or } Y_1^* = 1,\, F_{\mu^{x_1}}(x_2) \in (\,\frac{1}{4}, 1\,]. \end{cases} \end{aligned}$$

Observe that, e.g. $F_{\mu^{x_1}}(x_2) \in (\,0, \frac{3}{4}\,]$ is equivalent to $\Phi(X_2^* - X_1^*) < \frac{3}{4}$, where Φ is the standard normal cdf. Hence, the expression above can be rewritten as

$$\begin{aligned} &T_2(X_1^*, X_2^*) := Y_2^* = \\ &F^{-1}_{\nu^{Y_1^*}}(F_{\mu^{x_1}}(x_2)) = \begin{cases} -1, & X_1^* \le 0,\, \Phi(X_2^* - X_1^*) \le \frac{3}{4} \text{ or } X_1^* \ge 0,\, \Phi(X_2^* - X_1^*) \le \frac{1}{4}, \\ \;\;1, & X_1^* \le 0,\, \Phi(X_2^* - X_1^*) > \frac{3}{4} \text{ or } X_1^* \ge 0,\, \Phi(X_2^* - X_1^*) > \frac{1}{4}. \end{cases} \end{aligned}$$

It is trivial to check that the constructed coupling $(X_1^*, X_2^*, Y_1^*, Y_2^*)$ is causal. To demonstrate that it is indeed bicausal, we need to verify the following

$$\mathbb{P}[X_1^* \le x_1 | Y_1^* = 1, Y_2^* = 1] = \mathbb{P}[X_1^* \le x_1 | Y_1^* = 1, Y_2^* = -1] = \mathbb{P}[X_1^* \le x_1 | Y_1^* = 1]$$

and

$$\mathbb{P}[X_1^* \le x_1 | Y_1^* = -1, Y_2^* = 1] = \mathbb{P}[X_1^* \le x_1 | Y_1^* = -1, Y_2^* = -1] = \mathbb{P}[X_1^* \le x_1 | Y_1^* = -1].$$

Let us find the conditional probabilities

$$\mathbb{P}[X_1^* \le x_1 | Y_1^* = 1] = \tfrac{\mathbb{P}[X_1^* \le x_1, Y_1^* = 1]}{\mathbb{P}[Y_1^* = 1]} = 2\mathbb{P}[X_1^* \le x_1, X_1^* \ge 0] = 2\Phi(x_1) - 1;$$

$$\begin{aligned} \mathbb{P}[X_1^* \le x_1 | Y_1^* = 1, Y_2^* = 1] &= \tfrac{8}{3}\mathbb{P}\left[X_1^* \le x_1, X_1^* \ge 0, \Phi(X_2^* - X_1^*) > \tfrac{1}{4}\right] \\ &= \tfrac{8}{3}\mathbb{P}[0 \le G \le x_1]\mathbb{P}\left[\Phi(G') > \tfrac{1}{4}\right] \\ &= \tfrac{8}{3}\left(\Phi(x_1) - \tfrac{1}{2}\right)\tfrac{3}{4} = 2\Phi(x_1) - 1; \end{aligned}$$

$$\begin{aligned} \mathbb{P}[X_1^* \le x_1 | Y_1^* = 1, Y_2^* = -1] &= 8\mathbb{P}\left[X_1^* \le x_1, X_1^* \ge 0, \Phi(X_2^* - X_1^*) \le \tfrac{1}{4}\right] \\ &= 8\mathbb{P}[0 \le G \le x_1]\mathbb{P}\left[\Phi(G') \le \tfrac{1}{4}\right] \\ &= 8\left(\Phi(x_1) - \tfrac{1}{2}\right)\tfrac{1}{4} = 2\Phi(x_1) - 1. \end{aligned}$$

In the same way one can convince himself that

$$\begin{aligned} &\mathbb{P}[X_1^* \le x_1 | Y_1^* = -1, Y_2^* = 1] \\ &= \mathbb{P}[X_1^* \le x_1 | Y_1^* = -1, Y_2^* = -1] \\ &= \mathbb{P}[X_1^* \le x_1 | Y_1^* = -1] \end{aligned}$$

$$= 2\Phi(\min(0, x_1)).$$

Therefore, $T := (T_1(X_1^*), T_2(X_1^*, X_2^*))$ is indeed a bicausal transport between the given measures μ and ν. One can assure herself/himself that it is actually an increasing triangular transformation which is not lexicographically increasing. The last property fails because one can find such pairs of vectors $x = (x_1, x_2)$ and $\bar{x} = (\bar{x}_1, \bar{x}_2)$, that $T_1(x_1) = T_1(\bar{x}_1)$, and tweak x_2 and $\bar{x}_2$ in such a way that for $x_2 > \bar{x}_2$ it will happen that $T_2(x_1, x_2) < T_2(\bar{x}_1, \bar{x}_2)$.

2.4 Applications of the bicausal case

Inspired by the concepts of displacement interpolation/convexity (as in [Vil03, Chapter 5]), we will define a related notion where the Knothe-Rosenblatt rearrangement replaces the role of the Brenier's map, and dub it lexicographical displacement interpolation/convexity (see Section 2.4.1). We expect that the lexicographical displacement interpolations have a geometric meaning as geodesic curves, however, the realization of this idea is left as an open problem, which we consider as a relevant step towards an understanding of the geometrical side of the bicausal/nested transport problem. We only present indirect evidences in this respect. Still, this also provides an interesting link to stochastic programming and nested distances (see Section 2.4.2). Concretely, we show that under convexity and Lipschitz conditions on the cost, stochastic programs are lexicographical displacement concave, in analogy to the potential energy in optimal transport ([Vil03, Chapter 5.2]).

2.4.1 Geometric properties

We start with a discussion about geometric properties of the space of probability measures endowed with a bicausal Wasserstein distance (equiv. nested distance). We notice first that the Knothe-Rosenblatt rearrangement offers a way to interpolate in a meaningful and non-linear way between stochastic programs. From Remark 2.2.12 we know that this rearrangement is bicausal, and, as discussed in the previous section, if all conditional probabilities $\mu^{x_1,\dots,x_n}, \nu^{y_1,\dots,y_n}$ are atomless, then it is induced by a bicausal map which is characterized as the μ-a.s. unique transformation, which is increasing w.r.t. lexicographical order and which pushes forward μ to ν. Let us first define what we call the *lexicographical displacement interpolation* between μ and ν.

Definition 2.4.1. *Let $\pi = \pi(\mu, \nu)$ be the Knothe-Rosenblatt rearrangement as in* (2.2.19). *The lexicographical displacement interpolation between μ and ν is then defined as the function*

$$t \in [0,1] \mapsto [\mu, \nu]_t := [(x,y) \mapsto (1-t)x + ty]_* \pi \in \mathcal{P}(\mathbb{R}^N). \tag{2.4.1}$$

If all conditional probabilities $\mu^{x_1,\dots,x_n}$ are atomless, we also have

$$[\mu, \nu]_t = ([1-t]id + tT)_* \mu,$$

where $T = T(\mu,\nu)$ is the Knothe-Rosenblatt map defined in (2.2.20)*, and* id *denotes the identity map in $\mathbb{R}^N$.*

In particular we have that $[\mu,\nu]_0 = \mu$ and $[\mu,\nu]_1 = \nu$; therefore, the name.

Let us now discuss the geometric interpretation that lexicographical displacement interpolation should have. Assuming that μ has independent marginals, we start from the observation that for costs of the type

$$c_p(x_1,\dots,x_N,y_1,\dots,y_N) := \sum_{i=1}^N |x_i - y_i|^p,$$

lexicographical displacement interpolation has "constant speed" w.r.t. the associated bicausal / nested distances[6], namely

$$\left(\inf_{\pi\in\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,[\mu,\nu]_t)} \int c_p d\pi\right)^{1/p} = t\left(\inf_{\pi\in\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int c_p d\pi\right)^{1/p}, \tag{2.4.2}$$

whenever μ has atomless conditional distributions. Indeed, by assumption the map $x \mapsto (1-t)x + tT(x)$ pushes forward μ to $[\mu,\nu]_t$. Further, it is bicausal and an increasing triangular transformation (see Section 2.3), so it is the Knothe-Rosenblatt rearrangement of μ into $[\mu,\nu]_t$, and so optimal for the l.h.s. above. Therefore,

$$\begin{aligned}\left(\inf_{\pi\in\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,[\mu,\nu]_t)} \int c_p d\pi\right)^{1/p} &= \int \sum_{i=1}^N |x_i - (1-t)x_i - tT^i(x_i;x_1,\dots,x_{i-1})|^p d\mu(x_1,\dots,x_N)\\ &= t^p \int \sum_{i=1}^N |x_i - T^i(x_i;x_1,\dots,x_{i-1})|^p d\mu(x_1,\dots,x_N),\end{aligned}$$

which is the r.h.s. of (2.4.2). The argument is of course reminiscent to the Brenier case. This suggests that for the case $p = 2$ one would want to interpret the corresponding bicausal distance as a geodesic length over the space of probability measures (say absolutely continuous ones) when given a differentiable structure and corresponding metric. This is discussed in [Vil03, Chapter 8] for the classical transport case, and no such thing has yet been accomplished for the present bicausal setting. In a way, a first step in this direction was done by Mikami [Mik12], where a Hamilton-Jacobi formulation for the dual of the bicausal problem was derived.

2.4.2 Stochastic programming

The goal here is to explore and to further the connection between (non-anticipative) multistage stochastic programming and bicausal optimal transport, discovered by Pflug and Pichler [PP12]. We start by defining the value function of a stochastic program

$$v(\eta) := \inf_{u_1(),\dots,u_N()} \int H(x_1,\dots,x_N,u_1(x_1),u_2(x_1,x_2),\dots,u_N(x_1,\dots,x_N))\eta(dx_1,\dots,dx_N), \tag{2.4.3}$$

[6] That this constitutes an actual distance was proved in the appendix of [PP12].

as a function of the noise distribution η. Here $H : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is the objective function and the minimization is taken over Borel adapted controls.

We prove now that many stochastic optimization problems are concave along the curves given by (2.4.1). This result should be seen as a complement to the results of Pflug and Pichler [Pfl09, PP12], which give conditions under which $|v(\nu) - v(\mu)|$ can be gauged by the value of a bicausal problem between μ and ν (see Section 2.5 for a detailed discussion). Indeed, Theorem 2.4.2 highlights the connection between the most eminent of bicausal maps, i.e. the Knothe-Rosenblatt rearrangement, and multistage stochastic programming. On the other hand, the next result can be related to [Vil03, Open Problem 5.17], with the caveat that we replaced the role of the Brenier's map by the Knothe-Rosenblatt rearrangement when defining the interpolations.

Theorem 2.4.2. *Let $H : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ be bounded from below, concave in the first variable while convex in the second one. Then the functions*

$$t \in [0,1] \mapsto v([\mu,\nu]_t)$$

are concave for each ν and μ such that all conditional probabilities $\mu^{x_1,\dots,x_n}$ are atomless, so we may say that $v(\cdot)$ is lexicographical displacement concave. If further $v(\mu)$ is attained, then

$$v(\nu) \leq v(\mu) + \inf_{u^* \in \arg\min(v(\mu))} \int \inf_{\xi \in \partial_x H(x,u^*(x))} \xi \cdot [T(x) - x] d\mu(x).$$

Proof. Take μ, ν, H as stated, and $a, b, s, t \in [0,1]$ with $a + b = 1$. We will show that $v([\mu,\nu]_{at+bs}) \geq av([\mu,\nu]_t) + bv([\mu,\nu]_s)$. We write

$$\begin{aligned}
A_n(x) &= (1-t)x_n + tT^n(x_n; x_1, \dots, x_{n-1}) \\
B_n(x) &= (1-s)x_n + sT^n(x_n; x_1, \dots, x_{n-1}) \\
C_n(x) &= (1-at-bs)x_n + [at+bs]T^n(x_n; x_1, \dots, x_{n-1}),
\end{aligned}$$

with $x = (x_1, \dots, x_n)$, so we have

$$v([\mu,\nu]_{at+bs}) = \inf_{u_1(),\dots,u_N()} \int H\Big(C_1(x), \dots, C_N(x), u_1\big(C_1(x)\big), \dots, u_N\big(C_1(x), \dots, C_N(x)\big)\Big) d\mu(x),$$

and, by the concavity assumption,

$$\begin{aligned}
&v([\mu,\nu]_{at+bs}) \geq \qquad (2.4.4) \\
&a \left\{ \inf_{u_1(),\dots,u_N(\cdot)} \int H\Big(A_1(x), \dots, A_N(x), u_1\big(C_1(x)\big), \dots, u_N\big(C_1(x), \dots, C_N(x)\big)\Big) d\mu(x) \right\} + \\
&b \left\{ \inf_{u_1(),\dots,u_N(\cdot)} \int H\Big(B_1(x), \dots, B_N(x), u_1\big(C_1(x)\big), \dots, u_N\big(C_1(x), \dots, C_N(x)\big)\Big) d\mu(x) \right\}.
\end{aligned}$$

If say $t < 1$, we have that $A_1(x) = A_1(x_1)$ is injective (as it is strictly increasing) so defining $\tilde{u}_1(\cdot) = u_1 \circ C_1 \circ A_1^{-1}(\cdot)$ we have $\tilde{u}_1(A_1(x)) = u_1(C_1(x))$. For $n = 2$ we introduce

$$\tilde{u}_2(z_1, z_2) = u_2\big(\ C_1 \circ A_1^{-1}(z_1)\ ,\ C_2\big(A_1^{-1}(z_1), A_2^{-1}(z_2|A_1^{-1}(z_1))\big)\ \big),$$

which is well-defined due to the function A_2 being strictly increasing in its second variable, and verify that $\tilde{u}_2(A_1(x), A_2(x)) = u_2(C_1(x), C_2(x))$. Inductively, we obtain easily for each $n \le N$ a $\tilde{u}_n$ s.t. $\tilde{u}_n(A_1(x), \ldots, A_n(x)) = u_n(C_1(x), \ldots, C_n(x))$. Hence the first term on the r.h.s. of (2.4.4) is bounded from below by $v([\mu, \nu]_t)$. Using similar arguments for the second term, we see that $v([\mu, \nu]_{at+bs}) \ge av([\mu, \nu]_t) + bv([\mu, \nu]_s)$ holds if $s, t < 1$. The case $s = 1$ or $t = 1$ can be obtained by a limiting and a Komlos-Mazur type argument (with convex combinations; here the convexity assumption is used). A direct argument, inspired by [PP12], is as follows:

$$\begin{aligned}&\int H\big(A_1(x), \ldots, A_N(x), u_1\big(C_1(x)\big), \ldots, u_N\big(C_1(x), \ldots, C_N(x)\big)\big)\mu(dx) \ge \\ &\quad \mathbb{E}^\mu\left[H\Big(A_1, \ldots, A_N, \mathbb{E}^\mu[u_1(C_1)|A_1, \ldots, A_N], \ldots, \mathbb{E}^\mu[u_N(C_1, \ldots, C_N)|A_1, \ldots, A_N]\Big)\right],\end{aligned}$$

by the convexity assumption and Jensen's inequality. So now observing that if e.g. $t = 1$ then $A = T$, which is bicausal from μ to ν, we get that $\mathbb{E}^\mu[u_n(C_1, \ldots, C_n)|A_1, \ldots, A_N] = \tilde{u}_n(A_1, \ldots, A_n)$ for each n and we conclude as before. The case $s = 1$ is analogous.

For the last statement we obtain by the concavity assumption that

$$H([1-t]x + tT(x), u^*(x)) \le H(x, u^*(x)) + t\xi(x) \cdot [T(x) - x],$$

where u^* is any optimizer for $v(\mu)$, and $\xi(x)$ is any measurable selection of

$$x \mapsto \partial_x H(x, u^*(x)),$$

the partial superdifferential w.r.t. the first variable. Such a selection exists by [RW98, Theorem 14.56], for which H must be a normal integrand, but this is true by [RW98, Proposition 14.39]. In particular, for $t = 1$ and integrating, we get

$$\int H(T(x), u^*(x))\mu(dx) \le v(\mu) + \int \xi(x) \cdot [T(x) - x]d\mu(x).$$

By the same arguments as before (the convexity assumption, the bicausality of T and $T_*\mu = \nu$), the l.h.s. is an upper bound for $v(\nu)$. Therefore, the proof is concluded if we can find a measurable selector $\xi(\cdot)$ such that

$$\xi(x) \in argmin_{\xi \in \partial_x H(x, u^*(x))} \left\{\xi \cdot [T(x) - x]\right\},$$

The existence of such a selector follows from the measurable maximum theorem [AB06, Theorem 18.19] after observing that $(x, \xi) \mapsto \xi \cdot [T(x) - x]$ is a Carathéodory function and that the correspondence $x \mapsto \partial_x H(x, u^*(x))$ is nonempty compact-valued and weakly measurable (i.e. measurable in the sense of [RW98, Theorem 14.56]). □

2.5 Digression into classical functional inequalities

The aim of this section is to explore the connection between (non-anticipative) multistage stochastic programming and bicausal optimal transport from the point of view of geometric/functional inequalities. We finish this section with an application of DPP discussed in Section 2.2.2 giving the strengthened version of Talagrand inequality.

As established in [PP12, Theorem 6.1], as soon as the cost function H is 1-Lipschitz in its first argument and convex in the second one, the discrepancy between the values of a stochastic programs w.r.t. the noise distribution μ and ν (see (2.4.3) for the definition of v) is less than or equal to the bicausal distance between measures μ and ν w.r.t. the cost

$$c(x,y) := \|x-y\|_1 = \sum_{i=1}^{N} |x_i - y_i|,$$

namely

$$|v(\mu) - v(\nu)| \leq \inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \|x-y\|_1 d\pi(x,y).$$

This result means that if say μ is a "complicated law" (for instance given by a nightmarish tree), then finding a "simpler law" ν close enough in the causal distance sense guarantees that uniformly in the given class of stochastic programs the discrepancy is controlled. Notice that, in our interpretation, the complicated measure μ is expected to dominate (in the sense of absolute continuity) the simpler measure ν. Therefore, the practical question is how to efficiently minimize this bicausal distance over a given set of measures. We argue that in many situations a transport-information inequality for such bicausal distances permits to gauge the discrepancy of stochastic programs replacing the computation of such distance by the evaluation of a relative entropy.

We recall, that the relative entropy of the measure ν w.r.t. the given measure μ is denoted by $\mathrm{Ent}(\nu|\mu)$ and it equals

$$\mathrm{Ent}(\nu|\mu) = \int \rho \log \rho, \quad \nu = \rho\mu,$$

with the understanding that $\mathrm{Ent}(\nu|\mu) = +\infty$ if ν is not absolutely continuous w.r.t. μ. By transport-information inequalities we mean functional inequalities of the form

$$C(\mu,\nu) \leq \mathcal{E}_\mu(\nu), \ \forall \nu \in \mathcal{P}(\mathcal{X}),$$

where $C(\mu,\nu)$ is the optimal transport cost between μ and ν, w.r.t. some specific cost function, and $\mathcal{E}_\mu(\nu)$ is some nonlinear functional of μ ("energy"). The most important class of functional inequalities of this type occurs when the cost function is a distance function $c(x,y) = d(x,y)^p$, and the "energy" functional is the relative entropy.

We say that a probability measure μ satisfies the transportation cost-information inequality T_p if there is some constant $C > 0$ such that, for any probability measure ν,

$$\mathcal{W}_p(\mu,\nu) \leq \sqrt{2\, C\, \mathrm{Ent}(\nu|\mu)}, \tag{T_p}$$

where $\mathcal{W}_p^p(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \left(\int\int d(x,y)^p d\pi(x,y) \right)$. The cases $p = 1$ and $p = 2$ are particularly interesting. The $T_1(C)$ inequality is related to the concept of concentration of measures, and this fact first appeared in [Mar86, Mar96]. Gaussian concentration is a terminology meaning that some probability measure enjoys the properties of concentration which are similar to those of a Gaussian measure. A certain form of Gaussian concentration is actually equivalent to $T_1(C)$ (see [Vil08, Theorem 22.10]). As for the $T_2(C)$

inequality, it was first established by Talagrand (see [Tal96]) for the Gaussian measure and it gives the upper bound for the squared Wasserstein distance

$$\mathcal{W}_2^2(G^N, \nu) \leq 2\mathrm{Ent}(\nu|G^N), \tag{T_2}$$

where G^N is a unit standard Gaussian measure. Equality holds if and only if ν is an affine translate of G^N.

Theorem 2.5.1. *Let $\mu \in \mathcal{P}(\mathbb{R}^N)$ satisfy:*

(EXP) for each t there exist $a_t > 0, \lambda_t \in \mathbb{R}$ such that $\int e^{a_t x_t^2} \mu^{x_1,\ldots,x_{t-1}}(dx_t) \leq e^{\lambda_t}$ μ-a.s.;

(LIP) there is $C > 0$ such that, for all $t \in \{1, \ldots, N\}$ and every 1-Lipschitz function $f : \mathbb{R} \to \mathbb{R}$, the function

$$(x_1, \ldots, x_{t-1}) \mapsto \int f(x_t) \mu^{x_1,\ldots,x_{t-1}}(dx_t)$$

is μ-a.s. equal to a C-Lipschitz function.

Then we have the following bicausal transport-information inequality:

$$\text{for all } \nu \in \mathcal{P}(\mathbb{R}^N): \quad \mathcal{W}_{bc,1}(\mu, \nu) := \inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \|x - y\|_1 d\pi(x, y) \leq K\sqrt{\mathrm{Ent}(\nu|\mu)}, \tag{2.5.1}$$

where the constant K is defined as

$$K := \sqrt{2 \sum_{j=1}^{N} (1 + C)^{2j} \frac{(1+\lambda_{N-j})}{a_{N-j}^2}}.$$

In particular, for every "cost criterion" $H : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ bounded from below, r-Lipschitz in its first argument and convex in its second, we have for the corresponding stochastic programs

$$|v(\mu) - v(\nu)| \leq rK\sqrt{\mathrm{Ent}(\nu|\mu)}, \tag{2.5.2}$$

with $v(\cdot)$ defined as in (2.4.3).

Remark 2.5.2. A few observations are in order:

1. Crucially, the presence of the entropy in Theorem 2.5.1 stems from our assumption (EXP) which is equivalent to a Gaussian concentration and allows us to prove similar result in the presence of causality constraint.

2. The entropic upper bounds in (2.5.1)-(2.5.2) of course trivialize, becoming $+\infty$, if $\nu \not\ll \mu$. On the other hand, one could expect the r.h.s. of (2.5.2) to be easier to compute (when finite) than a transport-type quantity.

3. W.l.o.g. the Lipschitz property above is meant with respect to the sum-of-absolute-values norms in the respective spaces. The tower property (EXP) implies

$$\int e^{a_1 x_1^2 + \cdots + a_N x_N^2} \mu(dx_1, \ldots, dx_N) \leq e^{\lambda_1 + \cdots + \lambda_N} < \infty,$$

and in particular every Lipschitz function is μ-integrable.

4. We stress that Assumption (EXP) is reasonable in practice; it is automatically satisfied in the finite discrete case, for empirical measures, or when μ has bounded support. The corresponding non-causal version of (2.5.1), which can be found in [Vil08, Theorem 22.10] or [DGW04, Theorem 2.3] and crucially does not implies our result, actually also needs a condition equivalent to (EXP).

5. Assumption (LIP) implies by the Kantorovich-Rubinstein Theorem (e.g. [Vil03, Theorem 1.14]), that

$$\mathcal{W}_1(\mu^{x_1,\dots,x_{t-1}},\mu^{z_1,\dots,z_{t-1}}) \leq C\sum_{i<t}|x_i-z_i|,$$

where $\mathcal{W}_1$ is the usual 1-Wasserstein distance on $\mathcal{P}(R)$. If μ is a Markov law, then the r.h.s. here can be taken as $C|x_{t-1}-z_{t-1}|$. On the other hand, if μ is a martingale law, the l.h.s. is in turn bounded from below by $|x_{t-1}-z_{t-1}|$.

6. All in all, Assumption (LIP) is the most fundamental for Theorem 2.5.1. It is clearly satisfied in the discrete case or for empirical measures. In general it is also implied if μ has independent increments (with $C=1$ then), or if e.g. μ is the law of the solution $X_1,\dots,X_N$ of a uniformly Lipschitz random dynamical system of the form $X_{t+1}=R(Z_t,(X_1,\dots,X_t),t)$, where R is Borel in the first argument, Lipschitz in the second one (uniformly w.r.t. the first one), and Z_t is independent of $(X_1,\dots,X_t)$.

7. We observe that the constant in (2.5.1) can be very plausibly improved (making it of order $\sqrt{N}$ is probably the best one can hope for). This is done for example in [DGW04, Theorem 2.5] for the non-causal but Markov case, and necessitates more care and further assumptions.

Proof. By (EXP) and [Vil08, Theorem 22.10 + (22.16)] we have that $\mu^{x_1,\dots,x_{t-1}}$ satisfies the next T_1 functional inequality (for every Borel prob. measure m):

$$\begin{aligned}\mathcal{W}_1(\mu^{x_1,\dots,x_{t-1}},m) &\leq \tfrac{\sqrt{2}}{a_t}\left(1+\log\textstyle\int e^{a_t x_t^2}\mu^{x_1,\dots,x_{t-1}}(dx_t)\right)^{1/2}\sqrt{\mathrm{Ent}(m|\mu^{x_1,\dots,x_{t-1}})} \qquad (2.5.3)\\ &\leq \tfrac{\sqrt{2(1+\lambda_t)}}{a_t}\sqrt{\mathrm{Ent}(m|\mu^{x_1,\dots,x_{t-1}})}.\end{aligned}$$

Let us denote by K_{t-1} the constant in the r.h.s. above. By the triangle inequality and (LIP):

$$\begin{aligned}\mathcal{W}_1(\mu^{x_1,\dots,x_{t-1}},\nu^{y_1,\dots,y_{t-1}}) &\leq \mathcal{W}_1(\mu^{x_1,\dots,x_{t-1}},\mu^{y_1,\dots,y_{t-1}})+\mathcal{W}_1(\mu^{y_1,\dots,y_{t-1}},\nu^{y_1,\dots,y_{t-1}})\\ &\leq C\sum_{i<t}|x_i-y_i|+K_{t-1}\sqrt{\mathrm{Ent}(\nu^{y_1,\dots,y_{t-1}}|\mu^{y_1,\dots,y_{t-1}})},\end{aligned}$$

for $\mu\otimes\nu$-a.e. $(x_1,\dots,x_{t-1},y_1,\dots,y_{t-1})$; we are entitled to do this since we may assume w.l.o.g. that $\nu\ll\mu$. Using (Dyn-Pbc) and applying the above computation recursively, one arrives at:

$$\mathcal{W}_{bc,1}(\mu,\nu) \le \int \nu(dy_1,\dots,dy_N) \sum_{j<N} K_{N-j-1}(1+C)^j \sqrt{\mathrm{Ent}(\nu^{y_1,\dots,y_{N-j-1}}|\mu^{y_1,\dots,y_{N-j-1}})},$$

so using Cauchy-Schwartz for the sum and for the integral we get

$$\mathcal{W}_{bc,1}(\mu,\nu) \le \sqrt{A}\sqrt{B},$$

where

$$B = \int \nu(dy_1,\dots,dy_N) \sum_{j<N} \mathrm{Ent}(\nu^{y_1,\dots,y_{N-j-1}}|\mu^{y_1,\dots,y_{N-j-1}}),$$

which by the additive property of the entropy (e.g. [Var84, Lemma 10.3]) equals $\mathrm{Ent}(\nu|\mu)$, and where

$$A = 2\sum_{j<N}(1+C)^{2j}\frac{(1+\lambda_{N-j})}{a_{N-j}^2}.$$

□

We finish the section with the following proposition, where we give another instance where Theorem 2.2.13 (on the equality between causal and bicausal transport problems and their optimizer) and Condition 2.2.22 (on the independence of marginals of the source measure) come in handy. We establish here the slightly stronger related bicausal inequality.

Proposition 2.5.3. *Let G^N be the unit standard Gaussian measure on $\mathbb{R}^N$. Then for all $\nu \in \mathcal{P}(\mathbb{R}^N)$ the following bicausal transport-entropy inequality holds*

$$\mathcal{W}_{bc,2}^2(G^N,\nu) := \inf_{\pi\in\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \sum_{i=1}^N (x_i-y_i)^2 d\pi(x_1,\dots,x_N,y_1,\dots,y_N) \le 2\mathrm{Ent}(\nu|G^N). \tag{CT_2}$$

The inequality (CT_2) clearly implies (T_2) for G^N. It is the DPP for the bicausal transport problem that replaces the role of the usual tensorization trick; strictly speaking, the DPP is just giving a name to an intermediate step in Talagrand's well-known proof. The proof given here applies of course to other product measures, and in that setting is reminiscent of [Vil08, Proposition 22.5]; we stick to the gaussian case only for the sake of concreteness and because this relates to the Wiener case in continuous time, which we discuss in Chapter 3.

Proof. We start assuming (CT_2) for $N=1$, which holds by [Tal96], since in dimension 1 the values of classical, causal and bicausal transport problems coincide, and prove it for $N=2$. The general inductive argument is then obvious and therefore skipped. From e.g. [Var84, Lemma 10.3] it is known that for any $\nu \in \mathcal{P}(\mathbb{R}^2)$ the following tensorization of relative entropy holds:

$$\mathrm{Ent}(\nu|G^2) = \mathrm{Ent}(\nu|G^1\otimes G^1) = \mathrm{Ent}(p_*^1\nu|G^1) + \int \mathrm{Ent}(\nu^{y_1}|G^1)p_*^1\nu(dy_1)$$

$$\geq \tfrac{\mathcal{W}_2^2(G^1,p^1_*\nu)}{2} + \int \tfrac{\mathcal{W}_2^2(G^1,\nu^{y_1})}{2} p^1_*\nu(dy_1).$$

On the other hand, by the bicausal dynamic programming principle (Proposition 2.2.20) we get

$$\begin{aligned}\mathcal{W}_{bc,2}^2(G^2,\nu) &= \inf_{\pi\in\Pi(G^1,p^1_*\nu)} \int \{|x_1-y_1|^2 + \mathcal{W}_2(G^1,\nu^{y_1})\}\, \pi(dx_1,dy_1) \\ &= \mathcal{W}_2^2(G^1,p^1_*\nu) + \int \mathcal{W}_2^2(G^1,\nu^{y_1}) p^1_*\nu(dy_1).\end{aligned}$$

□

Remark 2.5.4. It is clear that equality in (CT_2) cannot generally hold, since for $N=1$ we have $\mathcal{W}_2^2(G,\nu) = \mathcal{W}_{bc,2}^2(G,\nu)$ and equality in (T_2) is not usually satisfied. On the other hand, if ν is an affine translate of G^N, then of course $2\mathrm{Ent}(\nu|G^N) = \mathcal{W}_2^2(G^N,\nu) = \mathcal{W}_{bc,2}^2(G^N,\nu)$. More generally, there is equality in (CT_2) provided the kernels $\nu^{y_1,\dots,y_t}(dy_{t+1})$ are affine translates of G^1, for each t and ν-a.e. y.

2.6 Anticipation of information: a prelude

We finish this chapter with a short discussion on enlargements of filtrations. This setting allows to encompass possible *anticipation* of some information regarding the evolution of the coordinate process. The aim of this section is to ease the understanding of this framework, which is discussed in details in Chapter 4, and provide the reader with a smooth transition from discrete to continuous time.

Let $w=(w_1,\dots,w_N)$ be the coordinate process on $\mathbb{R}^N$, and let $\mathcal{F}$ be the filtration generated by w:

$$\mathcal{F}_t := \sigma(w_s, s\leq t).$$

We fix arbitrary $\mathcal{F}_N$-measurable functions $\mathfrak{g}_t:\mathbb{R}^N\to\mathbb{R}$, and consider the enlargement of filtration $\mathcal{F}$ with the process $\mathfrak{g}(W)=(\mathfrak{g}_1(w),\dots,\mathfrak{g}_N(w))$. This leads to the *(progressively) enlarged filtration* $\mathcal{G}=(\mathcal{G}_t)_{t=1}^N$, defined via

$$\mathcal{G}_t := \mathcal{F}_t \vee \sigma(\{\mathfrak{g}_s, s\leq t\}). \tag{2.6.1}$$

We stress that in case $\mathfrak{g}_t(w)=L(w)$ for any $t=1,\dots,N$, this leads to the so-called *initial* enlargement of the filtration with L.

We remind that for a process Z on a given probability space $(\Omega,\mathbb{P})$, we denote by $\mathcal{F}^Z := Z^{-1}(\mathcal{F})$ the filtration generated by Z on Ω, and by $\mathcal{F}^{Z,\mathfrak{g}} := Z^{-1}(\mathcal{G})$ the filtration obtained by enlarging $\mathcal{F}^Z$ with the process $(\mathfrak{g}_t(Z))_{t=1}^N$, in analogy to (2.6.1) (note that $\mathfrak{g}_t(Z)$ are $\mathcal{F}^Z_N$-measurable for all $t\in\{1,\dots,N\}$).

The equivalent characterization of causality in the case of arbitrary filtrations was given in Proposition 2.1.5, and of course it includes enlargements as a special case. The condition for a transport plan to be causal, written in terms of stochastic processes, in this particular case looks as follows: for all $t=1,\dots,N$ and any set $B_t\in\mathcal{G}_t$,

$$\mathbb{P}\big((Y_1,\dots,Y_N)\in B_t \mid X_1,\dots,X_N\big) = \mathbb{P}\big((Y_1,\dots,Y_N)\in B_t \mid X_1,\dots X_t\big).$$

Heuristically this reads as "given the past of X, the past of $(Y, \mathfrak{g}(Y))$ and the future of X are independent".

The following proposition gives the connection between two causal transport problems: with and without enlargement of filtrations.

Proposition 2.6.1. *In case of initial enlargement of the coordinate filtration $\mathcal{F}$, i.e. $\mathcal{G} = \mathcal{F} \vee \sigma(L)$ fore some random variable L, for any given lower semicontinuous bounded from below cost function c and any probability measures μ and ν, we have*

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \int c(x_1, \dots, x_N, y_1, \dots, y_N) d\pi(x_1, \dots, x_N, y_1, \dots, y_N) = \tag{2.6.2}$$
$$\int \Big\{ \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu^{L=\tilde{y}})} \int c(x_1, \dots, x_N, y_1, \dots, y_N) d\pi(x_1, \dots, x_N, y_1, \dots, y_N) \Big\} d\ell(\tilde{y}),$$

where ℓ is the distribution of the random variable L.

Note that $\nu^{L=\tilde{y}}$ is the regular conditional kernel of ν, when the value of the random variable L is fixed and equals $\tilde{y}$. For the proof of this proposition we need the following lemma.

Lemma 2.6.2. *In case of initial enlargement of the coordinate filtration $\mathcal{F}$,*

$$\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu) \text{ if and only if } \pi^{L=\tilde{y}} \in \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu^{L=\tilde{y}}), \text{ for } L_*\nu - a.e.\ \tilde{y},$$

where $\pi^{L=\tilde{y}}$ is the $\pi-$conditional distribution of $(x_1, \dots, x_N, y_1, \dots, y_N)$ given $L(y_1, \dots, y_N) = \tilde{y}$.

The set $\Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu^{L=\tilde{y}})$ is the set of causal transport plans between $(\mathbb{R}^N, \mathcal{F}, \mu)$ and $(\mathbb{R}^N, \mathcal{F}, \nu^{L=\tilde{y}})$, i.e. the set of those transport plans that are adapted to the coordinate filtration $\mathcal{F}$ without enlargement.

Proof. Let $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$. First, observe that in the case of enlargement of filtration the functions h_t, used in the equivalent characterization of causality (2.1.3), might possibly depend on $L(y_1, \dots, y_N)$. Then from Proposition 2.1.5 we have that, for any $\mathcal{F}$-measurable function g_t and $\mathcal{G}_t$-measurable function h_t of the form $h_t(y_1, \dots, y_t, L(y_1, \dots, y_N)) = f_t(y_1, \dots, y_t) R(L(y_1, \dots, y_N))$, with $f \in C_b(\mathbb{R}^t), R \in C_b(\mathbb{R}^N)$,

$$\begin{aligned}
0 &= \int h_t(y_1, \dots, y_t, L(y_1, \dots, y_N)) \big[g_t(x_1, \dots, x_N) - \\
&\int g_t(x_1, \dots, x_N) \mu^{x_1, \dots, x_t}(dx_{t+1}, \dots, dx_N) \big] d\pi(x_1, \dots, x_N, y_1, \dots, y_N) \\
&= \int R(\tilde{y}) \int f_t(y_1, \dots, y_t) \big[g_t(x_1, \dots, x_N) - \\
&\int g_t(x_1, \dots, x_N) \mu^{x_1, \dots, x_t}(dx_{t+1}, \dots, dx_N) \big] d\pi^{L=\tilde{y}}(x_1, \dots, x_N, y_1, \dots, y_N) d\ell(\tilde{y}).
\end{aligned}$$

Since the above equality holds for any continuous bounded function R, we can conclude that

$$\int f_t(y_1, \dots, y_t) \big[g_t(x_1, \dots, x_N) - \\ \int g_t(x_1, \dots, x_N) \mu^{x_1, \dots, x_t}(dx_{t+1}, \dots, dx_N) \big] d\pi^{L=\tilde{y}}(x_1, \dots, x_N, y_1, \dots, y_N) = 0,$$

proving that $\pi^{L=\tilde{y}} \in \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu^{L=\tilde{y}})$. The reverse implication is trivial. Causality implies that $L(y_1, \ldots, y_N)$ is independent of $(x_1, \ldots, x_N)$, hence the first marginal of $\pi^{L=\tilde{y}}$ is the same for all $\tilde{y}$, namely μ. As for the second marginal, it is clear that

$$\pi^{L=\tilde{y}}(dy_1, \ldots, dy_N) = \nu^{L=\tilde{y}}(dy_1, \ldots, dy_N).$$

Hence, we conclude. □

Proof of Proposition 2.6.1. Let π be a feasible transport plan for the transport problem in the l.h.s. of (2.6.2). Then

$$\begin{aligned}&\int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N) = \\ &\int \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi^{L=\tilde{y}}(x_1, \ldots, x_N, y_1, \ldots, y_N) d\ell(\tilde{y}).\end{aligned}$$

From Lemma 2.6.2 we know that the transport plan $\pi^{L=\tilde{y}} \in \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu^{L=\tilde{y}})$, hence optimizing over all measures in $\Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu^{L=\tilde{y}})$ we get the inequality

$$\begin{aligned}&\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N) \geq \\ &\int \Big\{ \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu^{L=\tilde{y}})} \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N) \Big\} d\ell(\tilde{y}).\end{aligned}$$

To get the reverse inequality, first we need to show the variant of measurable selection result. Let $\Phi : \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N) \to \mathcal{P}(\mathbb{R}^N) \times \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N)$ be the continuous set-valued function that for each measure $\pi \in \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N)$ assigns the pairs of probability measures $(p^2_* \pi, \pi)$. We introduce $\Pi^{\mathcal{F},\mathcal{F}}(\mu, \cdot)$ which is the set of causal transport plans with the first marginal μ (see Remark 2.1.9). Then $D := \Phi(\Pi^{\mathcal{F},\mathcal{F}}(\mu, \cdot))$ is analytic as a continuous image of a closed set. We denote $D^\nu := \big\{\pi \in \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N) \text{ s.t. } (\nu, \pi) \in D\big\} = \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu)$. Consider the following optimization problem

$$\begin{aligned}f(\nu) :&= \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N) \\ &= \inf_{\pi \in D^\nu} \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N).\end{aligned} \tag{2.6.3}$$

Then by [BS78, Proposition 7.50(a)] there exist for every $\varepsilon > 0$ a universally measurable function $\phi : \mathcal{P}(\mathbb{R}^N) \to \mathcal{P}(\mathbb{R}^N \times \mathbb{R}^N)$ s.t. $Gr(\phi) \subseteq D$, i.e. $\phi(\nu) \in \Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu)$, and $f(\nu) \geq \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\phi(\nu) - \varepsilon$. Hence, $\phi(\nu)$ is a measurable selection of $\varepsilon-$optimizers for the transport problem (2.6.3). Moreover, since the map $\tilde{y} \mapsto \nu^{L=\tilde{y}}$ is universally measurable, $\tilde{y} \mapsto \phi(\nu^{L=\tilde{y}})$ is likewise measurable, as the composition of two universally measurable maps (see [BS78, Proposition 7.44]). Thus, we have that for any $\tilde{y}$ an $\varepsilon-$optimizer of

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu^{L=\tilde{y}})} \int c(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi(x_1, \ldots, x_N, y_1, \ldots, y_N),$$

that can be chosen in a measurable way. Denote $\pi^{\tilde{y}}(dx, dy) := \phi(\nu^{L=\tilde{y}})$. We have that $\pi = \int \pi^{\tilde{y}} d\ell(\tilde{y})$ is well defined (see [BS78, Proposition 7.45]), and from Lemma 2.6.2 it is a feasible element in $\Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$. Hence, the converse inequality follows.

□

In continuous time it is a difficult problem to give conditions under which any $\mathcal{F}-$martingale remains a semimartingale in a bigger filtration $\mathcal{G} \supset \mathcal{F}$. However, in discrete time this is always true and it is an easy consequence of Doob's decomposition.

Proposition 2.6.3. *[BSJR16] If $\mathcal{F} \subset \mathcal{G}$ and X is any $\mathcal{F}-$martingale, it is also a $\mathcal{G}-$semimartingale with decomposition*

$$X = X^{\mathcal{G}} + A^{\mathcal{G}},$$

where $X^{\mathcal{G}}$ is a $\mathcal{G}-$martingale, $A^{\mathcal{G}}$ is $\mathcal{G}-$predictable, $A_0^{\mathcal{G}} = 0$. Moreover,

$$\Delta A_n^{\mathcal{G}} = \mathbb{E}\left[\Delta X_n \mid \mathcal{G}_{n-1}\right], \, n \geq 1.$$

Let us study the following particular example of "Brownian bridge" in discrete time.

Example 2.6.4 (Brownian bridge). Let $(Z_i, i \geq 1)$ be a sequence of i.i.d. random variables with zero mean and the process X be of the form $X_n := \sum_{i=1}^n Z_i, n = 1, \dots, N$. Denote by μ the law of X and by $\mathcal{F}^X$ the filtration generated by X. Of course, X is a discrete-time $\mathcal{F}^X-$martingale. Consider the initial enlargement of $\mathcal{F}^X$ by the value Z_N, i.e. $\mathcal{G} = \mathcal{F}^X \vee \sigma(Z_N)$. Then the $\mathcal{G}-$ martingale $X^{\mathcal{G}}$ has the following decomposition

$$X_n^{\mathcal{G}} = X_n - \textstyle\sum_{k=1}^n \frac{X_N - X_{k-1}}{N-(k-1)}. \tag{2.6.4}$$

In general, the coupling $(X^{\mathcal{G}}, X)$ is not causal from $\mathcal{F}^X$ to $\mathcal{G}$. To see this consider the case of $N = 3$. Then

$$\begin{aligned} X^{\mathcal{G}} &= \left(\tfrac{X_1 - X_3}{3}, \tfrac{3X_1 + 6X_2 - 5X_3}{6}, \tfrac{3X_1 + 6X_2 - 5X_3}{6}\right) \\ &= \left(\tfrac{2Z_1 - Z_2 - Z_3}{3}, \tfrac{4Z_1 + Z_2 - 5Z_3}{6}, \tfrac{4Z_1 + Z_2 - 5Z_3}{6}\right). \end{aligned}$$

Having this decomposition we see that causality fails upon noticing that $X_2^{\mathcal{G}}$ is not independent on X_1 and $X_1^{\mathcal{G}}$ jointly.

It is not surprising that the causality property fails in this particular example. As we will see in Example 4.1.5, in continuous time the independence of increments is neither enough to guarantee causality. However, in the Brownian case, the coupling between Brownian motion in enlarged and original filtrations is always causal. This happens due to the joint independence of increments on the past of the original process and on the enlarged filtration up to each moment of time t. In order to encompass a broader class of martingales, we introduced another condition to test causality in continuous time, see Remark 4.4.7.

Chapter 3

From discrete to continuous time

This chapter focuses on the understanding of the causal transport problem on Wiener space through time discretizations, using the results obtained for discrete-time causal transport problems. In Section 3.1 we introduce causal optimal transport problems on Wiener spaces and fix the notations for the rest of the chapter. In Section 3.2 we provide two examples, that illustrate two extreme cases: when optimizers of classical/causal/bicausal transport problems with the Cameron-Martin cost coincide, and when they are different, which is the case in general. In the following two sections we review the existing literature and the relevant results on transport on Wiener spaces (Section 3.3) and the existing connection between optimal transport and stochastic differential equations (Section 3.4). Section 3.5 starts with the discussion on the projections procedure showing in Lemma 3.5.1 that this construction preserves causality and that the sequence of the time-discretized Cameron-Martin costs is increasing, see Lemma 3.5.5. These building blocks allow us to prove in Proposition 3.5.7 that causal optimal transport problems on Wiener spaces are the limiting objects of their finite-dimensional counterparts. Section 3.6 contains further possible applications of the discretization procedure. Namely, we give there an approximation scheme for a particular class of stochastic control problems. In Section 3.7 we recover the famous Talagrand inequality on Wiener space using the projections arguments. Moreover, we provide the proof for the equality between the causal transport problems with Cameron-Martin cost and relative entropy for a class of Markovian SDEs (see Theorem 3.7.4). This result in its general form first appeared in [Las13, Theorem 4], with the proof based on the well-known Föllmer's formula and the related Föllmer drift (also employed in [Leh13] for similar purposes). To the contrary, our approach makes use of the duality theory developed for discrete-time causal transport problems. Next in Section 3.8 we give an explicit procedure for embedding discrete-time optimizers into admissible transport plans on Wiener space. We finish this chapter with Section 3.9, where we give conjectures on how to recover the above mentioned equality with the entropy for a general case, for instance by showing that the discrepancies between time-discretized causal transport problems and the relative entropy disappear in the limit.

3.1 Notation specific for continuous time

We start with setting up the notation for the rest of the chapter. We denote by $\mathcal{C} := \mathcal{C}[0,1]$ the space of continuous functions $f : [0,1] \to \mathbb{R}$ such that $f(0) = 0$. Let $W = (W_t)_{t\in[0,1]}$ be the coordinate process on $\mathcal{C}$, $W_t(\omega) = \omega_t$ for $\omega \in \mathcal{C}$, and let $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,1]}$ be the right-continuous version of the filtration on $\mathcal{C}$ generated by W:

$$\mathcal{F}_t := \bigcap_{u>t} \sigma(W_s, s \le u).$$

In the following, we denote by $\mathcal{F}^\mu$ the usual μ-completed filtration containing $\mathcal{F}$, with analogous notation throughout the rest of the chapter. Further, γ will stand for the Wiener measure on $\mathcal{C}$ (see e.g. [KS91, Definition 2.4.21]). The generic element on $\mathcal{C} \times \mathcal{C}$ is denoted by $(\omega, \bar{\omega})$.

Definition 3.1.1 (Causal transport plan). *A transport plan $\pi \in \Pi(\mu,\nu)$ is called* causal *between $(\mathcal{C},\mathcal{F},\mu)$ and $(\mathcal{C},\mathcal{F},\nu)$ if, for any $t \in [0,1]$ and any set $A \in \mathcal{F}_t$, the map*

$$\mathcal{C} \ni \omega \mapsto \pi^\omega(A) = \mathbb{E}^\pi[\mathbb{I}_A | \mathcal{F}_1 \otimes \{\emptyset, \mathcal{C}\}](\omega)$$

is measurable with respect to $\mathcal{F}^\mu_t$, where $\pi^\omega(d\bar{\omega}) := \pi^\omega(\{\omega\} \times d\bar{\omega})$ is the regular conditional kernel of π w.r.t. the first coordinate.

For the definition of causality in a discrete setting see Definition 2.1.1, and for the continuous-time case with arbitrary filtrations see Definition 4.1.1. In what follows, we will consider (bi)causal transport problems between $(\mathcal{C},\mathcal{F},\mu)$ and $(\mathcal{C},\mathcal{F},\nu)$, where most often we will take $\mu = \gamma$ and $\nu \ll \gamma$. As before, the sets of causal and bicausal transport plans are denoted by $\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$, resp. $\Pi^{\mathcal{F},\mathcal{F}}_{bc}(\mu,\nu)$.

In the following remark we collect some useful equivalent characterizations of causality. For the discrete-time analogues and the proofs of the below equivalences we refer the reader to Proposition 2.1.5.

Remark 3.1.2. For a probability measure $\pi \in \Pi(\mu,\nu)$, the following are equivalent:

1. π is a causal transport plan between $(\mathcal{C},\mathcal{F},\mu)$ and $(\mathcal{C},\mathcal{F},\nu)$;
2. $\pi\left(\mathcal{C} \times D_t | \mathcal{F}_t \otimes \{\emptyset,\mathcal{C}\}\right) = \pi\left(\mathcal{C} \times D_t | \mathcal{F}_1 \otimes \{\emptyset,\mathcal{C}\}\right)$ π-a.s., for all $t \in [0,1]$, $D_t \in \mathcal{F}_t$;
3. the σ-fields $\{\emptyset,\mathcal{C}\} \otimes \mathcal{F}_t$ and $\mathcal{F}_1 \otimes \{\emptyset,\mathcal{C}\}$ are conditionally independent with respect to π given $\mathcal{F}_t \otimes \{\emptyset,\mathcal{C}\}$, for all $t \in [0,1]$.

Remark 3.1.3. Under the weak-continuity assumption on the first marginal μ, i.e. $\forall f \in C_b(\mathcal{C})$, $t \in [0,T]$ the map

$$x \mapsto \mathbb{E}^\mu[f|\mathcal{F}_t](x) \text{ is continuous,}$$

(see Assumption 2.1.6 written for discrete time), we can prove that the set $\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ is compact and $\Pi^{\mathcal{F},\mathcal{F}}(\mu,\cdot)$ is closed. The proof is essentially the same as in the discrete-time case (see Lemma 2.1.8 and Remark 2.1.9). Note that the above assumption is satisfied by the Wiener measure.

For any function $g \in \mathcal{C}$ we introduce the convenient notation

$$\langle g_t \rangle := \begin{cases} \dot{g}_t & \text{, if } g \text{ is absolutely continuous} \\ +\infty & \text{, else.} \end{cases}$$

A prominent cost on $\mathcal{C} \times \mathcal{C}$, which is an analogue of the squared Euclidean distance in the finite dimensional case, is *the Cameron-Martin (semi)distance*

$$(\omega, \bar{\omega}) \in \mathcal{C} \times \mathcal{C} \to c(\omega, \bar{\omega}) := \int_0^1 \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 dt \in [0, \infty].$$

This cost induces the following transport problem introduced by Feyel and Üstunell [FÜ02]

$$\mathcal{W}_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int \int_0^1 \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 dt \, d\pi(\omega, \bar{\omega}), \tag{3.1.1}$$

as well as its causal counterpart

$$\mathcal{W}_{c,2}^2(\mu, \nu) := \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \int_0^1 \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 dt \, d\pi(\omega, \bar{\omega}), \tag{3.1.2}$$

and bicausal one

$$\mathcal{W}_{bc,2}^2(\mu, \nu) := \inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \int_0^1 \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 dt \, d\pi(\omega, \bar{\omega}). \tag{3.1.3}$$

Naturally, one can consider more general costs of the form

$$(\omega, \bar{\omega}) \in \mathcal{C} \times \mathcal{C} \to c^h(\omega, \bar{\omega}) := \int_0^1 h(\langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle) dt, \tag{3.1.4}$$

with h e.g. a convex function. We will refer to this kind of costs as *the Cameron-Martin type costs*. From now on the main object of our investigation is the causal transport problem on Wiener space with the Cameron-Martin cost.

3.2 Examples

The following example aims to illustrate the difference between the optimizers of the transport problems (3.1.1) and (3.1.2).

Example 3.2.1. Let $\nu \in \mathcal{P}(\mathcal{C}[0, 1])$ be absolutely continuous w.r.t. the Wiener measure $\mu = \gamma$ with density

$$\frac{d\nu}{d\gamma} = f(W_1),$$

where f is an arbitrary continuous bounded function function which acts on the terminal value of the coordinate process W. Let $S : \mathbb{R} \to \mathbb{R}$ be the monotone increasing rearrangement of $\gamma_1 := p_*\gamma$ to $\nu_1 := p_*\nu$, with $p(W) = W_1$, meaning that $\mathbb{E}^\gamma[h(S(W_1))] = \mathbb{E}^\nu[h(W_1)], \forall h \in \mathcal{C}_b(\mathbb{R})$. Define the process

$$X_t(\omega) := \omega_t + t(S(\omega_1) - \omega_1), \; t \in [0, 1].$$

We claim that the transport plan $\bar{\pi} := (\mathrm{id}, X)_*\gamma$ is optimal for the Feyel-Üstunel transport problem (3.1.1), where $X := (X_t)_{t\in[0,1]}$. To show this, first we need to verify that the pushforward of γ by X is exactly ν, meaning that for any function $g \in \mathcal{C}_b(\mathcal{C})$ the following holds

$$\mathbb{E}^\nu[g(\omega)] = \mathbb{E}^\gamma[g(X(\omega))],$$

or equivalently,

$$\mathbb{E}^\gamma[g(\omega)f(\omega_1)] = \mathbb{E}^\gamma[g(X(\omega))]. \tag{3.2.1}$$

To see this, we proceed as follows:

$$\mathbb{E}^\gamma[g(X(\omega))] \;\; = \;\; \mathbb{E}^\gamma[g\{\omega_t + t(S(\omega_1) - \omega_1)\}_{t\in[0,1]}]$$

$$
\begin{aligned}
&= \int \mathbb{E}^\gamma[g\{\omega_t + tx - t\omega_1\}_{t\in[0,1]}] S_*\gamma_1(dx) \\
&= \int \mathbb{E}^\gamma[g(\omega)|\omega_1 = x]\nu(dx) \\
&= \mathbb{E}^\gamma[\mathbb{E}^\gamma[g(\omega)|\omega_1]f(\omega_1)] = \mathbb{E}^\gamma[g(\omega)f(\omega_1)].
\end{aligned}
$$

Indeed, in the second equality we used that $\{\omega_t - t\omega_1\}$ is independent of ω_1 under γ, and in the third equality that the law of $\{\omega_t + tx - t\omega_1\}_t$ under γ equals to $\gamma^{\omega_1=x}$. As for the optimality of the transport plan $\bar{\pi} := (\mathrm{id}, X)_*\gamma$, it follows from

$$
\begin{aligned}
\mathbb{E}^{\bar{\pi}}\left[\int_0^1 \langle\cdot\bar{\omega}_t - \omega_t\cdot\rangle^2 dt\right] = \mathbb{E}^{\bar{\pi}}\left[(S(\omega_1) - \omega_1)^2\right] &= \inf_{p\in\Pi(\gamma_1,\nu_1)} \int |x-y|^2 p(dx,dy) \\
&\leq \inf_{\pi\in\Pi(\gamma,\nu)} \mathbb{E}^{\pi}\left[\int_0^1 \langle\cdot\bar{\omega}_t - \omega_t\cdot\rangle^2 dt\right],
\end{aligned}
$$

by optimality of the rearrangement S (e.g. see [Vil03, Theorem 2.18]) and then Jensen's inequality.

On the other hand, to find a solution to the causal/bicausal variant we use Girsanov Theorem as follows. Define $Z_t := \mathbb{E}^\gamma[f(W_1)|\mathcal{F}_t]$, $t \in [0,1]$, which is a $\gamma-$martingale. By the Markov property and independence of increments, we have that $Z_t = u^f(t, W_t)$, where $u^f(t,x)$ is smooth for $t < 1$, and non-negative. Denoting $\partial_x u^f(t,x)$ the partial derivative w.r.t. the variable x, by Itô formula we get

$$
dZ_t = \partial_x u^f(t,W_t)dW_t = u^f(t,W_t)\left(\frac{\partial_x u^f}{u^f}\right)(t,W_t)dW_t = Z_t\,\partial_x \ln\left(u^f(t,W_t)\right)dW_t.
$$

Then from Girsanov Theorem we find that

$$
X_t = W_t - \int_0^t \partial_x \ln(u^f(s,W_s))ds \text{ is a } \nu - \text{Brownian motion.}
$$

Hence, the pair of processes (X, W) is under ν a weak solution to the SDE

$$
dY_t = dB_t - \partial_x \ln\left(u^f(t,Y_t)\right)dt,
$$

see [KS91, Proposition 5.3.6]. By [Las15, Theorem 4], we then know that the joint law of (X, W) under ν is a bicausal transport plan with the first marginal γ and the second marginal ν, which is optimal for (3.1.2) and (3.1.3); see Section 3.4 for more details.

We remark, that Feyel and Üstunel proved a general result on existence and structure of optimizers for (3.1.1) (see Section 3.3 for more details on this), without giving explicit formulae. Therefore, the above example might be of its own interest.

To the contrary, in the simple case of a deterministic shift of the Wiener measure, the optimizers of the classical, causal and bicausal transport problems with the Cameron-Martin cost coincide, as the following example shows.

Example 3.2.2. Consider the transport between the measures γ and $\nu = A_*\gamma$, where $A_t := \omega_t + \int_0^t \alpha_s ds$, $t \in [0,1]$, is a shift of the Wiener measure, with α a deterministic and square integrable process. We claim that $\bar{\pi} := (\mathrm{id}, A)_*\gamma$ is a bicausal transport plan, and that is an optimizer of the transport problems (3.1.1), (3.1.2) and (3.1.3). By classical duality results (e.g. see [Vil03, Theorem 1.3]) we know that

$$
\frac{1}{2}\mathcal{W}_2^2(\gamma,\nu) = \sup_{\phi(\omega)+\psi(\bar{\omega})\leq\frac{1}{2}\int_0^T\langle\cdot\omega_t-\bar{\omega}_t\cdot\rangle^2 dt} \int \phi(\omega)d\gamma + \int \psi(\bar{\omega})d\nu.
$$

Taking $\phi(\omega) := -\int_0^T \alpha_s d\omega_s$ and $\psi(\bar{\omega}) := \int_0^T \alpha_s d\bar{\omega}_s - \frac{1}{2}\int_0^T \alpha_s^2 ds$, and considering $\bar{\omega}_t = \omega_t +$

$\int_0^t h_s ds$ with $h = h(\omega, \bar{\omega})$, we get

$$\phi(\omega) + \psi(\bar{\omega}) = -\int_0^T \alpha_s d\omega_s + \int_0^T \alpha_s d\omega_s + \int_0^T \alpha_s h_s ds - \frac{1}{2}\int_0^T \alpha_s^2 ds =$$
$$= -\int_0^T \frac{(\alpha_s - h_s)^2}{2} ds + \int_0^T \frac{h_s^2}{2} ds \leq \frac{1}{2}\int_0^T \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 ds.$$

If such h does not exist for given $(\omega, \bar{\omega})$, the inequality is trivial. Hence, (ϕ, ψ) are dual feasible elements. On the other hand, by definition of ν, we have

$$\begin{aligned}\int \phi(\omega) d\gamma + \int \psi(\bar{\omega}) d\nu &= \mathbb{E}^\gamma\left[-\int_0^T \alpha_s d\omega_s\right] + \mathbb{E}^\nu\left[\int_0^T \alpha_s d\bar{\omega}_s - \frac{1}{2}\int_0^T \alpha_s^2 ds\right] \\ &= \mathbb{E}^\gamma\left[\int_0^T \alpha_s \left(d\omega_s + \alpha_s ds\right) - \frac{1}{2}\int_0^T \alpha_s^2 ds\right] \\ &= \mathbb{E}^\gamma\left[\frac{1}{2}\int_0^T \alpha_s^2 ds\right] = \frac{1}{2}\int_0^T \alpha_s^2 ds.\end{aligned}$$

Indeed, note that $\mathbb{E}^\gamma\left[\int_0^T \alpha_s d\omega_s\right] = 0$, since ω is Brownian motion under γ. Thus,

$$\frac{1}{2}\mathcal{W}_2^2(\gamma, \nu) = \sup_{\phi(\omega)+\psi(\bar{\omega}) \leq \frac{1}{2}\int_0^T \langle\!\cdot \omega_t - \bar{\omega}_t \cdot\!\rangle^2 dt} \int \phi(\omega) d\gamma + \int \psi(\bar{\omega}) d\nu \geq \frac{1}{2}\int_0^T \alpha_s^2 ds.$$

We conclude that $\bar{\pi}$ is optimizer of (3.1.1), since for it this lower bound is attained. It is easy to see that $\bar{\pi}$ is a bicausal transport plan. Together with the previous result, we conclude that it is optimal causal as well as bicausal transport plan for (3.1.2), resp. (3.1.3).

3.3 Transport on Wiener spaces: literature overview

From Brenier [Bre91] we know that any absolutely continuous probability measure $\mu \in \mathcal{P}(\mathbb{R}^N)$ can be transformed into an arbitrary probability measure $\nu \in \mathcal{P}(\mathbb{R}^N)$ by means of a transformations of the form $T(x) = \nabla\phi(x)$, where $\phi(x)$ is a convex function, under the assumption that finite second moments of those measures exist. Such transformations are natural analogues of increasing functions on the real line. In the Wiener space setting, i.e. for optimal transport on $\mathcal{C}$ with the Cameron-Martin cost, an analogous result was established by Feyel and Üstunel in [FÜ02]. In particular, they show that if the given measure $\nu \in \mathcal{P}(\mathcal{C})$ is of finite Wasserstein distance from the Wiener measure γ, then there exists a function ϕ, which is roughly a counterpart of a convex function, such that the transformation defined by

$$T(\omega) = \omega + \nabla\phi(\omega) \tag{3.3.1}$$

transports γ to ν. One difficulty in the infinite dimensional setting comes from the fact that the Cameron-Martin cost is not continuous with respect to the Fréchet topology of $\mathcal{C}$, and hence, for instance, the weak convergence of the probability measures does not imply the convergence of the integrals of the cost function. Moreover, in a given sense, the Cameron-Martin cost function takes the value plus infinity “very often”([FÜ04]).

In general, the transformations of the form (3.3.1) are not adapted to the canonical filtration, and might be anticipative. To the contrary, abstract Girsanov transformations of the classical Wiener space, namely

$$\tilde{T}(\omega) = \omega + \int_0^t v_s ds, \text{ with } v \text{ adapted}, \tag{3.3.2}$$

are of course non-anticipative by construction. Lassalle ([Las13]) was the first one to introduce a class of transport problems incorporating the nonanticipativity condition as a special property of transport plans, and called it causality. In fact, the maps of the form (3.3.2) are feasible elements to the causal optimal transport problems (3.1.2) via

$$\tilde{\pi}(d\omega, d\bar{\omega}) = \mu(d\omega)\delta_{\tilde{T}(\omega)}(d\bar{\omega}).$$

On the other hand, the object defined in (3.3.1), which is a solution of the transport problem (3.1.1), is in general not even a feasible element to causal transport problems, as we have seen in Example 3.2.1.

3.4 Causal transport and SDEs: literature overview

In this section we explain in details the existing connection between (bi)causal optimal transport plans and solutions of Brownian-driven stochastic differential equations (SDEs) of the form

$$dX_t = dB_t + v_t(X)dt, \quad X_0 = 0, \tag{3.4.1}$$

where B is a Brownian motion and v is an adapted function of the path of the process X.

First, we recall the terminology about weak and strong solutions of SDEs. A pair of processes (B, X) defined on some filtered probability space $(\Omega, \mathbb{F}, (\mathcal{F}_t), \mathbb{P})$ is called a weak solution of the above SDE if it satisfies (3.4.1), if the process B is an $(\mathcal{F}_t)-$Brownian motion and X is $(\mathcal{F}_t)-$adapted. The pair of processes is called a strong solution if it is defined on a given probability space, satisfies the SDE (3.4.1) and, importantly, the process X is adapted to the filtration of Brownian motion. So, the existence of a weak solution ensures that a solution exists on some probability space for some stochastic inputs, whereas a strong solution is defined on a given probability space and for given stochastic inputs. For a more detailed overview and the existing results on existence/uniqueness of solutions we refer the reader to [Itô50, YW71, Ok03, Pro05].

Finding the conditions that guarantee existence and uniqueness of weak/strong solutions is the central topic in SDE-related literature. One of the first criterion for the existence of strong solutions was given by Yamada and Watanabe in [YW71]. Among other results, the authors proved that for Itô equations of the form (3.4.1), the conditional probability kernels of the joint laws of weak solutions are adapted to the Brownian filtration. Later, in [Jac80] the author started the analysis of the existence of solutions from the perspective of properties of measures on the product path space. In particular, he observed that a measure $\pi = (B, X)_*\mathbb{P}$, which is a joint law of a weak solution (B, X) on some probability space $(\Omega, \mathbb{P})$, corresponds to a strong solution if the conditional probability kernel of π, disintegrated w.r.t. the first coordinate, is a Dirac delta at a given point, i.e. $|\text{supp}(\pi^\omega)| \leq 1$. Moreover, he showed that every strong solution measure (i.e. those measures that correspond to joint laws of strong solutions) is extremal in the set of all solution measures. The converse is not necessarily true, since it might happen that there exists only one joint solution measure, which is not strong. Kurtz has continued this line of research in [Kur07, Kur14], analysing the properties of these measures and giving the existence/uniqueness proofs using different methods. He referred to the adaptedness property of kernels as a compatibility constraint.

Even though the analysis of the properties of solution measures was started before Lassalle [Las13], he was the first one who made the connection between SDEs and optimal transport explicit. He showed that for a given measure $\nu \in \mathcal{P}(\mathcal{C})$, which is absolutely continuous w.r.t. the Wiener measure γ, there is a drift v_t s.t. the joint law of (B, X), which is a weak solution of (3.4.1), is optimal for $\mathcal{W}^2_{c,2}(\gamma, \nu)$. The proof relies on an optimal transport formulation of a well-known representation of the relative entropy w.r.t. γ, inherited from Föllmer's formula [Föl86, Proposition 2.11]. Also in [Las15, Theorem 4] the author recognized that (3.4.1) has a strong solution if and only if the causal transport problem (3.1.2) has a Monge optimizer, i.e. the optimal transport plan is generated by an adapted map of the form (3.3.2). And so, the existence of causal Monge optimizers of (3.1.2) guarantees the existence of strong solutions to (3.4.1). We would like to remark that causal transport problems between two stochastic processes allow for a broader class of applications, not necessarily restricted to SDEs, as discussed in Section 2.4.

To construct a weak solution, one does not fix a priori any probability space. However, once one knows that a weak solution exists on some filtered probability space, one may ask a further question, if it possible to build a strong solution on this space. This is not always possible, as it is known from the celebrated Tsilrelson's SDE (see [Tsi75]), that has a weak solution, but no strong one. A similar situation occurs with solutions to causal transport problems: Monge solutions do not always exist. Therefore, it is an interesting question to find conditions under which an optimizer to (3.1.2) is given by a Monge map. Using our results from the discrete-time setting given in Chapter 2, in Section 3.9 we suggest some steps in this direction.

3.5 Projections and limits

The aim of this section is to understand the causal/bicausal equality of [Las13] in continuous time. As we show, this equality is simply the result of the Cameron-Martin cost being written in terms of "speed" and the source measure having independent increments.

To start with, we examine how a (bi)causal transport problem from $(\mathcal{C}, \mathcal{F}, \mu)$ into $(\mathcal{C}, \mathcal{F}, \nu)$ on a time-index set $\mathbb{T}$ can be projected down (i.e. pushed-forward) to a similar one from $(\mathbb{R}^N, \mathcal{F}^N, \mu^N)$ to $(\mathbb{R}^N, \mathcal{F}^N, \nu^N)$ on a smaller index set $\mathbb{T}^N = \{t_1, \dots, t_N\} \subseteq \mathbb{T}$, where $\mathcal{F}^N = (\mathcal{F}^N_i)^N_{i=1}$ is the coordinate filtration on $\mathbb{R}^N$. This discretization procedure will be then applied in order to project (bi)causal transport problems on Wiener space down to $\mathbb{R}^N$. We will show that the projections constructed below preserve (bi)causality under an additional assumptions on the marginals. We finish this section by analysing the continuous-time limit of the projected transport problem to the original one on Wiener space.

We introduce the process $b : \mathcal{C} \mapsto \mathbb{R}^N$ by defining

$$(\omega_t)_{t \in \mathbb{T}} \mapsto b(\omega) = (\omega_{t_1}, \dots, \omega_{t_N}). \tag{3.5.1}$$

It can be seen as an N-step process on the measurable space $\mathcal{C}$. We will use b to push forward the transport problem on $\mathcal{C}$ down to $\mathbb{R}^N$. The filtration generated by the process

b on $\mathcal{C}$, namely $\{\sigma(\{\omega_{t_1},\dots,\omega_{t_i}\})\}_{i=1}^N$, is contained in the canonical filtration $\mathcal{F}$ in the obvious way.

For the given measures $\mu \in \mathcal{P}(\mathcal{C})$, $\nu \in \mathcal{P}(\mathcal{C})$ let

$$\mu^N \in \mathcal{P}(\mathbb{R}^N),\ \nu^N \in \mathcal{P}(\mathbb{R}^N) \text{ be defined as } \mu^N := b_*\mu,\ \nu^N := b_*\nu,$$

the push-forwards of μ and ν by b. Further, if $c : \mathcal{C} \times \mathcal{C} \to \mathbb{R} \cup \{+\infty\}$ is a bounded from below cost criterion, we define a new cost by setting

$$c^N : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}, \quad c^N(x,y) := \inf_{(\omega,\bar{\omega}) \in b^{-1}(x_1,\dots,x_N) \times b^{-1}(y_1,\dots,y_N)} c(\omega,\bar{\omega}).$$

Given the natural construction above, we would like to show that (bi)causality is preserved.

Lemma 3.5.1. *Let π be a causal (respect. bicausal) transport plan between μ and ν on $\mathcal{C} \times \mathcal{C}$. Define*

$$\pi^N := (b \times b)_*\pi,$$

which is a probability measure on $\mathbb{R}^N \times \mathbb{R}^N$. If μ is the law of a Markov process (respect. additionally ν is the law of a Markov process) then π^N is a causal (respect. bicausal) transport plan between μ^N and ν^N on $\mathbb{R}^N \times \mathbb{R}^N$.

Proof. Clearly π^N has the prescribed marginals. We remind that π^ω and $[\pi^N]^x$ denote respectively the kernels of π w.r.t. μ and of π^N w.r.t. μ^N. From the disintegration of the measure π^N we know

$$\mathbb{E}^{\pi^N}[\mathbb{I}_A(x)\mathbb{I}_B(y)] = \mathbb{E}^{\mu^N}[\mathbb{I}_A(x)[\pi^N]^x(B)], \tag{3.5.2}$$

for arbitrary sets $A, B \in \mathcal{F}^N$. On the other hand, using the change of variables formula, we have

$$\begin{aligned}&\mathbb{E}^{\pi^N}[\mathbb{I}_A(x)\mathbb{I}_B(y)] = \mathbb{E}^{\pi}[\mathbb{I}_A(b(\omega))\mathbb{I}_B(b(\bar{\omega}))] = \mathbb{E}^{\mu}[\mathbb{I}_A(b(\omega))\pi^\omega(b^{-1}(B))] = \\ &\mathbb{E}^{\mu}[\mathbb{I}_A(b(\omega))\mathbb{E}^{\mu}[\pi^\omega(b^{-1}(B))|\sigma(b)]] = \mathbb{E}^{\mu^N}[\mathbb{I}_A(x)\mathbb{E}^{\mu}[\pi^\omega(b^{-1}(B))|b = x]],\end{aligned}$$

where $\sigma(b)$ denotes the sigma-algebra generated by the function b. Thus we see that the kernel of π^N w.r.t. μ^N is given by

$$[\pi^N]^x(B) := \mathbb{E}^{\mu}[\pi^\omega(b^{-1}(B))|b = x].$$

For the sets $B \in \mathcal{F}_i^N$, the inverse image $b^{-1}(B) \in \mathcal{F}_{t_i}$, and by causality $\pi^\omega(b^{-1}(B))$ is $\mathcal{F}_{t_i}$-measurable. It remains to show that

$$\mathbb{E}^{\mu}[\pi^\omega(b^{-1}(B))|x_1,\dots x_N] = \mathbb{E}^{\mu}[\pi^\omega(b^{-1}(B))|x_1,\dots x_i]. \tag{3.5.3}$$

Indeed, this would imply that the kernel $[\pi^N]^x(B)$ is $\mathcal{F}_i^N-$measurable. Note that (3.5.3) is true as soon as μ is Markov, since then for any $X \in \mathcal{F}_{t_i}$, $A \in \mathcal{F}_i^N$ and $C \in \sigma(x_i,\dots x_N)$ it holds

$$\begin{aligned}&\mathbb{E}^{\mu}[\mathbb{I}_A\mathbb{I}_C\mathbb{E}^{\mu}[\mathbb{I}_X|x_1,\dots x_N]] = \mathbb{E}^{\mu}[\mathbb{I}_X\mathbb{I}_A\mathbb{I}_C] = \mathbb{E}^{\mu}[\mathbb{I}_X\mathbb{I}_A\mathbb{E}^{\mu}[\mathbb{I}_C|\mathcal{F}_{t_i}]] = \mathbb{E}^{\mu}[\mathbb{I}_X\mathbb{I}_A\mathbb{E}^{\mu}[\mathbb{I}_C|x_i]] = \\ &\mathbb{E}^{\mu}[\mathbb{I}_X\mathbb{I}_A\mathbb{E}^{\mu}[\mathbb{I}_C|x_1,\dots,x_i]] = \mathbb{E}^{\mu}[\mathbb{E}^{\mu}[\mathbb{I}_X|x_1,\dots,x_i]\mathbb{I}_A\mathbb{E}^{\mu}[\mathbb{I}_C|x_1,\dots,x_i]] = \mathbb{E}^{\mu}[\mathbb{I}_A\mathbb{I}_C\mathbb{E}^{\mu}[\mathbb{I}_X|x_1,\dots,x_i]].\end{aligned} \tag{3.5.4}$$

In the bicausal case we proceed analogously. □

We stress that if the source measure μ is not Markov, it is no longer true that the projection constructed above preserves causality. To see this consider the following simple example, suggested to us by Mathias Beiglböck.

Example 3.5.2. Let X_1, X_2 be two random variables on some probability space $(\Omega, \mathbb{P})$. Consider the process $\bar{X} = (\bar{X}_1, \bar{X}_2, \bar{X}_3) := (X_1, X_2, X_1)$ in 3-steps, which is not Markov.

It is transported to $\bar{Y}$ by an adapted map T, i.e.

$$\bar{Y} = (\bar{Y}_1, \bar{Y}_2, \bar{Y}_3) := (T^1(X_1), T^2(X_1, X_2), T^3(X_1, X_2, X_1)).$$

From this construction it is obvious that the coupling $(\bar{X}, \bar{Y})$ is causal. However, the coupling $(\bar{X}_2, \bar{X}_3, \bar{Y}_2, \bar{Y}_3)$, obtained by projecting $(\bar{X}, \bar{Y})$ on the smaller time index set (by deleting the first time), is not causal anymore. In order to see that causality fails, we verify that

$$Law(\bar{Y}_2 | \bar{X}_2, \bar{X}_3) = Law(T^2(X_1, X_2)|X_2, X_1) = \delta_{T^2(X_1, X_2)}$$

does not coincide with

$$Law(\bar{Y}_2 | \bar{X}_2) = Law(T^2(X_1, X_2)|X_2).$$

The following corollary shows that the finite sampling of the transport problems on $\mathcal{C} \times \mathcal{C}$ decreases the optimal expected cost.

Corollary 3.5.3. *Let* Π *denote* $\Pi^{\mathcal{F},\mathcal{F}}(\mu, \nu)$ *(respect.* $\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu, \nu)$*) and* Π^N *denote* $\Pi^{\mathcal{F}^N,\mathcal{F}^N}(\mu^N, \nu^N)$ *(respect.* $\Pi_{bc}^{\mathcal{F}^N,\mathcal{F}^N}(\mu^N, \nu^N)$*). If* μ *is the law of a Markov process w.r.t. the natural filtration (respect. additionally* ν *is the law of a Markov process w.r.t. this filtration), then we get the following inequality*

$$\inf_{\pi^N \in \Pi^N} \int c^N(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi^N(x_1, \ldots, x_N, y_1, \ldots, y_N) \quad \leq \quad \inf_{\pi \in \Pi} \int c(\omega, \bar{\omega}) d\pi(\omega, \bar{\omega}).$$

Proof. We take a transport plan $\pi \in \Pi$ and build a measure $\pi^N \in \Pi^N$ with the help of the projection argument in Lemma 3.5.1. From the definition of the cost function c^N we see that

$$c^N(b(\omega), b(\bar{\omega})) \leq c(\omega, \bar{\omega}),$$

for any arbitrary pair $(\omega, \bar{\omega})$. Then, by the definition of pushforward measure, we get

$$\int c^N(x_1, \ldots, x_N, y_1, \ldots, y_N) d\pi^N(x_1, \ldots, x_N, y_1, \ldots, y_N) = \int c^N(b(\omega), b(\bar{\omega})) d\pi(\omega, \bar{\omega}) \leq \int c(\omega, \bar{\omega}) d\pi(\omega, \bar{\omega}).$$

The result easily follows from this inequality. $\square$

The nice feature of the Cameron-Martin type costs is that their projection into finite-time indices is very explicit.

Lemma 3.5.4. *Let* $c = c^h$ *be the Cameron-Martin type cost function on* $\mathcal{C} \times \mathcal{C}$ *introduced in* (3.1.4)*. Then its projected cost* $c^N(x_1, \ldots, x_N, y_1, \ldots, y_N)$ *on the time grid* $\mathbb{T}^N = \{t_1, \ldots, t_N\}$ *is given by:*

$$c^{N,h}(x_1, \ldots, x_N, y_1, \ldots, y_N) = \sum_{i=1}^{N-1} h\left(\frac{\Delta x_{i+1} - \Delta y_{i+1}}{\Delta t_{i+1}}\right) \Delta t_{i+1}, \tag{3.5.5}$$

where Δ *represents the difference operator* $\Delta f_{t+1} := f_{t+1} - f_t$*.*

Proof. We remind that the cost function $c^{N,h}$ was defined as follows

$$c^{N,h}(x_1, \ldots, x_N, y_1, \ldots, y_N) := \inf_{(\omega, \bar{\omega}) \in b^{-1}(x_1, \ldots, x_N) \times b^{-1}(y_1, \ldots, y_N)} c^h(\omega, \bar{\omega}),$$

meaning that we optimize over all possible paths $(\omega, \bar{\omega})$ that at the fixed moments of time $t_1, \ldots, t_N$ take the prescribed values (x_i, y_i) for $i = 1, \ldots, N$. It is of course clear that $c^{N,h}$

is bounded above by the value attained on the pair of piecewise linear functions that take given values at times $t_1, \ldots, t_N$, i.e.

$$c^{N,h}(x_1, \ldots, x_N, y_1, \ldots, y_N) \leq \sum_{i=1}^{N-1} \int_{t_i}^{t_{i+1}} h(\langle\cdot\omega_t - \bar{\omega}_t\cdot\rangle)dt = \sum_{i=1}^{N-1} h\left(\tfrac{\Delta x_{i+1} - \Delta y_{i+1}}{\Delta t_{i+1}}\right)\Delta t_{i+1}.$$

The reverse inequality is an application of Jensen's inequality

$$\begin{aligned}
\int_0^1 h(\langle\cdot\omega_t - \bar{\omega}_t\cdot\rangle)dt &= \sum_{i=1}^{N-1} \Delta t_{i+1}\Big(\int_{t_i}^{t_{i+1}} h(\langle\cdot\omega_t - \bar{\omega}_t\cdot\rangle)\tfrac{dt}{\Delta t_{i+1}}\Big) \\
&\geq \sum_{i=1}^{N-1} h\Big(\int_{t_i}^{t_{i+1}} \langle\cdot\omega_t - \bar{\omega}_t\cdot\rangle\tfrac{dt}{\Delta t_{i+1}}\Big)\Delta t_{i+1} \\
&= \sum_{i=1}^{N-1} h\left(\tfrac{\Delta x_{t_{i+1}} - \Delta y_{t_{i+1}}}{\Delta t_{i+1}}\right)\Delta t_{i+1}.
\end{aligned}$$

$\square$

It is interesting to note that the sequence of the projected costs $c^{N,h}$ defined in (3.5.5) is increasing to the Cameron-Martin type cost, as we now prove.

Lemma 3.5.5. *For $N \in \mathbb{N}$, let $\mathbb{T}^N$ be a refining partition of the interval $[0,1]$, i.e. $\mathbb{T}^{N+1} \subseteq \mathbb{T}^N$, whose mesh size goes to zero. For any convex function h, the sequence of costs*

$$c^{N,h}(\omega_{t_1}, \ldots, \omega_{t_N}, \bar{\omega}_{t_1}, \ldots, \bar{\omega}_{t_N}) := \sum_{i=1}^{N-1} h\left(\tfrac{\Delta\omega_{t_{i+1}} - \Delta\bar{\omega}_{t_{i+1}}}{\Delta t_{i+1}}\right)\Delta t_{i+1}$$

is increasing to $c^h := \int_0^1 h(\langle\cdot\omega_t - \bar{\omega}\cdot\rangle)dt$.

Proof. First, we show that the sequence $\{c^{N,h}\}$ is increasing. Consider the time interval $[t_i^N, t_{i+1}^N]$ from the partition $\mathbb{T}^N$. Since the partition is refining, we can assume w.l.o.g. that for some time index s, the point from the partition $\mathbb{T}^{N+1}$ is in the interval $t_{s-1}^{N+1} := t_i^N \leq t_s^{N+1} \leq t_{i+1}^N =: t_{s+1}^{N+1}$. Then Jensen's inequality implies that

$$\begin{aligned}
(t_{i+1}^N - t_i^N)\Big[\frac{t_s^{N+1} - t_i^N}{t_{i+1}^N - t_i^N}\, h\Big(\frac{\omega_{t_s^{N+1}} - \omega_{t_i^N} - \bar{\omega}_{t_s^{N+1}} + \bar{\omega}_{t_i^N}}{t_s^{N+1} - t_i^N}\Big) + \\
\frac{t_{i+1}^N - t_s^{N+1}}{t_{i+1}^N - t_i^N}\, h\Big(\frac{\omega_{t_{i+1}^N} - \omega_{t_s^{N+1}} - \bar{\omega}_{t_{i+1}^N} + \bar{\omega}_{t_s^{N+1}}}{t_{i+1}^N - t_s^{N+1}}\Big)\Big] \geq \\
(t_{i+1}^N - t_i^N)\, h\Big(\frac{\omega_{t_s^{N+1}} - \omega_{t_i^N} - \bar{\omega}_{t_s^{N+1}} + \bar{\omega}_{t_i^N} + \omega_{t_{i+1}^N} - \omega_{t_s^{N+1}} - \bar{\omega}_{t_{i+1}^N} + \bar{\omega}_{t_s^{N+1}}}{t_{i+1}^N - t_i^N}\Big) = \\
h\Big(\frac{\Delta\omega_{t_{i+1}^N} - \Delta\bar{\omega}_{t_{i+1}^N}}{\Delta t_{i+1}^N}\Big)\Delta t_{i+1}^N.
\end{aligned}$$

And so, we see that the l.h.s., representing two elements in the sum $c^{N+1,h}$, is bigger than or equal to the corresponding one from the sum $c^{N,h}$. Iterating these arguments for any time index i, one can show that $c^{N,h} \leq c^{N+1,h}$, proving that the sequence of costs is indeed increasing. By the convergence of Riemann sums we see that indeed $\lim_{N\to\infty} c^{N,h} = c^h$. $\square$

As a result of the previous lemma, we can show the convergence of the corresponding transport problems.

Corollary 3.5.6. *If μ is the law of a Markov process, then the following sequence of values of transport problems*

$$P(c^{N,h}) := \inf_{\pi^N \in \Pi^{\mathcal{F}^N,\mathcal{F}^N}(\mu^N,\nu^N)} \int \sum_{i=0}^{N} h\left(\tfrac{\Delta x_{i+1}-\Delta y_{i+1}}{\Delta t_{i+1}}\right) \Delta t_{i+1} d\pi^N(x,y)$$

is increasing to

$$P(c^h) := \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \int \int_0^1 h(\dot{\omega}_t - \dot{\bar{\omega}})dtd\pi(\omega,\bar{\omega}).$$

Proof. First, note that transport problem $P(c^h)$ is defined on $\mathcal{C} \times \mathcal{C}$, while $P(c^{N,h})$ is defined on $\mathbb{R}^N \times \mathbb{R}^N$. However, from the way we defined the finite dimensional projections of the measures (see Lemma 3.5.1) and from the change of variables formula, we know

$$\int c^{N,h}(b(\omega_{t_1},\ldots,\omega_{t_N}), b(\bar{\omega}_{t_1},\ldots,\bar{\omega}_{t_N}))d\pi(\omega,\bar{\omega}) = \\ \int c^{N,h}(x_1,\ldots,x_N,y_1,\ldots,y_N)d\pi^N(x_1,\ldots,x_N,y_1,\ldots,y_N), \quad (3.5.6)$$

where $\omega_{t_1},\ldots,\omega_{t_N}$ come from the continuous path ω sampled at the time points $t_1,\ldots,t_N$. This implies that $P(c^{N,h})$ coincides with the values of the transport problem on $\mathcal{C} \times \mathcal{C}$ with cost $c^{N,h}(b(\cdot),b(\cdot))$. Knowing that the sequence of costs $c^{N,h}$ is increasing to c^h (Lemma 3.5.5) and that the set $\Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ is compact (see Remark 3.1.3), it is straightforward to show the convergence of $P(c^{N,h})$ to $P(c^h)$ (for the arguments see the proof of Proposition 1.3.1). □

Now we are ready to present the main result of this section. We consider the following discretization of the transport problem on Wiener space. For $N \in \mathbb{N}$ we denote by $\mathbb{T}^N$ the N-generation dyadic grid on $[0,1]$, given by $\{k2^{-N} : k = 0,\ldots,2^N\}$. We write

$$\mathcal{W}^{2,N}_{c,2}(\gamma^N,\nu^N) = \inf_{\pi \in \Pi^{\mathcal{F}^N,\mathcal{F}^N}(\gamma^N,\nu^N)} \int \sum_{i=1}^{2^N-1} 2^{-N} \left\{\tfrac{\Delta x_{i+1}-\Delta y_{i+1}}{2^{-N}}\right\}^2 d\pi(x_0,\ldots,x_{2^N},y_0,\ldots,y_{2^N}) \quad (3.5.7)$$

for the projection on $\mathbb{R}^{2^N}$ of the transport problem on Wiener space, where the superscript N next to $\mathcal{W}$ means that we consider the transport problem with the discretized cost function c^N. From Corollary 3.5.3 we know that

$$\mathcal{W}^{2,N}_{c,2}(\gamma^N,\nu^N) \le \mathcal{W}^2_{c,2}(\gamma,\nu).$$

In fact, as expected, we get that the problem on the right-hand side is the limiting object of the one on the left-hand side.

Proposition 3.5.7. *For the Wiener measure γ, an arbitrary probability measure $\nu \in \mathcal{P}(\mathcal{C})$, and their finite dimensional projections γ^N, $\nu^N \in \mathcal{P}(\mathbb{R}^N)$, we have*

$$\mathcal{W}^{2,N}_{c,2}(\gamma^N,\nu^N) \nearrow \mathcal{W}^2_{c,2}(\gamma,\nu).$$

The proof of this result easily follows from Lemma 3.5.5 and Corollary 3.5.6.

3.6 Stochastic control approximations

In this section we would like to give another possible application of the discretization procedure, namely for the approximation of stochastic control problems. Consider the following problem

$$\inf_{v_t \text{ s.t. } dX_t = v_t dt + dB_t} \mathbb{E}\left[\Big(\int_0^1 h(v_t)dt + F(X_1)\Big)\right], \tag{3.6.1}$$

where we optimize over all drifts v_t adapted to the filtration generated by Brownian motion B s.t. the stochastic differential equation

$$dX_t = v_t dt + dB_t,$$

is satisfied. This problem can be equivalently written as follows

$$\begin{aligned}\inf_{\nu \in \mathcal{P}(\mathcal{C})} \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma,\nu)} \int \Big(\int_0^1 h(\langle\!\cdot\,\omega_t - \bar{\omega}_t\cdot\!\rangle)dt + F(\bar{\omega}_1)\Big)d\pi(\omega,\bar{\omega}) = \\ \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma,\cdot)} \int \Big(\int_0^1 h(\langle\!\cdot\,\omega_t - \bar{\omega}_t\cdot\!\rangle)dt + F(\bar{\omega}_1)\Big)d\pi(\omega,\bar{\omega}).\end{aligned} \tag{3.6.2}$$

The idea is to show that the above stochastic control problem (3.6.1) can be approximated by the following

$$\inf_{\pi \in \Pi^{\mathcal{F}^N,\mathcal{F}^N}(\gamma^N,\cdot)} \int \big(c^{N,h}(x_1,\ldots,x_N,y_1,\ldots,y_N)dt + F^N(y_N)\big)d\pi^N(x_1,\ldots,x_N,y_1,\ldots,y_N). \tag{3.6.3}$$

To do this, we first need to prove a result similar to the one in Proposition 3.5.7, for causal transport problems where only the first marginal is being fixed. We remind that this set of measures, denoted by $\Pi^{\mathcal{F},\mathcal{F}}(\gamma,\cdot)$, is closed (see Remark 3.1.3), but not compact.

Lemma 3.6.1. *Let $h(x) \geq C\,|x|^p$, for some $p > 1$, and all x large enough . We consider*

$$\begin{aligned}\tilde{P}(c^{N,h}) &:= \inf_{\pi \in \Pi^{\mathcal{F}^N,\mathcal{F}^N}(\gamma^N,\cdot)} \int \sum_{i=0}^{N} h\left(\tfrac{\Delta x_{i+1} - \Delta y_{i+1}}{\Delta t_{i+1}}\right)\Delta t_{i+1} d\pi^N(x,y),\\ \tilde{P}(c^h) &:= \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma,\cdot)} \int \int_0^1 h(\langle\!\cdot\,\omega_t - \bar{\omega}\cdot\!\rangle)dt\, d\pi < \infty.\end{aligned}$$

Then $\tilde{P}(c^{N,h})$ is increasing to $\tilde{P}(c^h)$.

Before we give the proof of the above result, we prove the following tightness lemma.

Lemma 3.6.2. *Let $\{\pi_n\}_n \in \Pi(\gamma,\cdot)$ such that for some $p > 1$ there exists a constant C s.t.*

$$\int \int_0^1 \langle\!\cdot\,\omega_t - \bar{\omega}_t\cdot\!\rangle^p dt d\pi_n \leq C, \quad \forall n.$$

Then the sequence $\{\pi_n\}_n$ is tight.

Proof. By Lemma 1.1.1, we only need to prove that the sequence of second marginals $\nu_n := p^2_*\pi_n$ is tight, since the first marginal $p^1_*\pi_n = \gamma$ is fixed and not moving with n. By [Bil99, Theorem 7.3], tightness of $\{\nu_n\}$ is equivalent to the statement

$$\text{for all } \varepsilon > 0 : \lim_{\delta\to 0} \limsup_{n\to\infty} \nu_n \left\{ \sup_{s,t\in[0,1],|s-t|\leq\delta} |\bar{\omega}_s - \bar{\omega}_t| > \varepsilon \right\} = 0.$$

From Chebyshev's inequality we know that

$$\nu_n \left\{ \sup_{s,t\in[0,1],|s-t|\leq\delta} |\bar{\omega}_s - \bar{\omega}_t| > \varepsilon \right\} \leq \tfrac{1}{\varepsilon} \int \left\{ \sup_{s,t\in[0,1],|s-t|\leq\delta} |\bar{\omega}_s - \bar{\omega}_t| \right\} d\nu_n.$$

Denoting $g_t := \omega_t - \bar{\omega}_t$, we see that

$$\int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |g_s - g_t| \right\} d\pi_n \le \int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |s-t| \left| \int_s^t \langle\!\cdot g_r \cdot\!\rangle \tfrac{dr}{|s-t|} \right| \right\} d\pi_n \le$$

$$\int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |s-t|^{\frac{p-1}{p}} (\int_s^t \langle\!\cdot g_r \cdot\!\rangle^p dr)^{1/p} \right\} d\pi_n \le \delta^{\frac{p-1}{p}} \int \left(\int_0^1 \langle\!\cdot g_r \cdot\!\rangle^p dr \right)^{1/p} d\pi_n \le$$

$$\delta^{\frac{p-1}{p}} \left(\int \int_0^1 \langle\!\cdot g_r \cdot\!\rangle^p dr d\pi_n \right)^{1/p} \le \delta^{\frac{p-1}{p}} C^{1/p},$$

where we used Hölder's inequality in the second step and Jensen's inequality in the fourth. Since by Gaussian estimates $\int |\omega_t - \omega_s|^4 d\gamma \le C_1|t-s|^2$, it follows by [RY99, Theorem I.2.1] that

$$\int \left[\sup_{s\neq t} \left(\tfrac{|\omega_t-\omega_s|}{|t-s|^{\frac{1}{8}}}\right)^4 \right] d\gamma < \infty.$$

Then

$$\int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |\omega_t - \omega_s| \right\} d\gamma \le \delta^{\frac{1}{8}} \int \left\{ \sup_{s\neq t} \tfrac{|\omega_t-\omega_s|}{|t-s|^{\frac{1}{8}}} \right\} d\gamma \le \delta^{\frac{1}{8}} \sqrt[4]{\int \left[\left(\sup_{s\neq t} \tfrac{|\omega_t-\omega_s|}{|t-s|^{\frac{1}{8}}} \right)^4 \right] d\gamma} = C_2 \delta^{\frac{1}{8}}.$$

Then collecting the bounds together, we get

$$\int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |\bar{\omega}_s - \bar{\omega}_t| \right\} d\nu_n \le \int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |\omega_s - \omega_t - (\bar{\omega}_s - \bar{\omega}_t)| \right\} d\pi_n +$$

$$\int \left\{ \sup_{s,t\in[0,1],|s-t|\le\delta} |\omega_s - \omega_t| \right\} d\gamma \le \delta^{\frac{p-1}{p}} C^{1/p} + C_2 \delta^{\frac{1}{8}}.$$

This bound tends to 0, as $\delta \to 0$, and so we conclude that $\{\pi_n\}$ is tight. □

Proof of Lemma 3.6.1. Remembering that $\tilde{P}(c^{N,h})$ is equivalent to a problem on $\mathcal{C} \times \mathcal{C}$ with cost $c^{N,h}(b(\cdot), b(\cdot))$, for each n we pick a sequence $\{\pi_n\} \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma, \cdot)$ s.t.

$$\int c^{N,h} d\pi_n \le \tilde{P}(c^{N,h}) + \tfrac{1}{n} \le \tilde{P}(c^h) + \tfrac{1}{n}.$$

Lemma 3.6.2 implies that $\{\pi_n\}$ is tight and up to a subsequence converges to some $\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma, \cdot)$. The proof then follows as in Corollary 3.5.6. □

Proposition 3.6.3. *If* (3.6.1) *is finite, then for any lower semicontinuous function F : $\mathbb{R} \to \mathbb{R}$ bounded from below, and $h(x) \ge C\,|x|^p$, for some $p > 1$, and all x large enough, the sequence* (3.6.3) *converges to* (3.6.1).

Proof. Since the function F is lower semicontinuous, there exists a sequence of bounded continuous functions $\{F^N\}$ increasing to F. Following the arguments of Proposition 1.3.1 applied for a sequence of costs $\tilde{c}^{N,h} := c^{N,h} + F^N$ together with Lemma 3.6.1, we can show the convergence of $P(\tilde{c}^{N,h})$ to $P(\tilde{c}^h)$, where $\tilde{c}^h := c^h + F$. □

3.7 Entropy

In [Las13] the author showed that the relative entropy appears to be a causal counterpart to the squared Wasserstein distance

$$\mathcal{W}^2_{c,2}(\gamma, \nu) = 2\mathrm{Ent}(\nu|\gamma). \tag{3.7.1}$$

This trivially entails Talagrand's inequality [Tal96] for the Wiener measure. The extension of the finite dimensional version of this famous inequality to the setting of Wiener spaces was first done by [FÜ02, FÜ04]. As an immediate consequence of the finite-dimensional time-discretized results from the previous section we can recover this inequality as well. Interestingly, our method is different from [Las13] and [Leh13]. We give the proof in Lemma 3.7.3. But before, for completeness, we recall several relevant facts about relative entropy. We start with a well-known lemma, stating that the relative entropy decreases when measures are pushed forward by measurable maps. For the sake of completeness, we provide a proof below.

Lemma 3.7.1. *Let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ be two probability measures s.t. ν is absolutely continuous w.r.t. μ. The for any measurable map T we have*

$$\mathrm{Ent}(T_*\nu|T_*\mu) \leq \mathrm{Ent}(\nu|\mu),$$

with equality if and only if the density $d\nu/d\mu$ is a function of T.

Proof. Denote the pushforwards $\tilde{\mu} := T_*\mu$ and $\tilde{\nu} := T_*\nu$. First observe that from the change of variables formula we have

$$\mathbb{E}^{\tilde{\nu}}[f] = \mathbb{E}^{\nu}[f \circ T] = \mathbb{E}^{\mu}\left[f \circ T \,\tfrac{d\nu}{d\mu}\right] = \mathbb{E}^{\mu}\left[f \circ T \,\mathbb{E}^{\mu}\left[\tfrac{d\nu}{d\mu}\Big|\sigma(T)\right]\right] = \mathbb{E}^{\tilde{\mu}}\left[f \,\mathbb{E}^{\mu}\left[\tfrac{d\nu}{d\mu}\Big|T = \cdot\right]\right],$$

for any $f \in \mathcal{C}_b(\mathcal{X})$, and where $\sigma(T)$ denotes the $\sigma-$algebra generated by T. From the derivations above we see that $\tilde{\nu} \ll \tilde{\mu}$ with

$$\tilde{\mu} - \text{a.e. } \tilde{x}, \quad \frac{d\tilde{\nu}}{d\tilde{\mu}}(\tilde{x}) = \mathbb{E}^{\mu}\left[\frac{d\nu}{d\mu}\Big|T = \tilde{x}\right]$$

and

$$\mu - \text{ a.e. } x, \quad \frac{d\tilde{\nu}}{d\tilde{\mu}} \circ T(x) = \mathbb{E}^{\mu}\left[\frac{d\nu}{d\mu}\Big|T = T(x)\right].$$

Then, substituting this in the formula for the relative entropy, we have

$$\begin{aligned}
\mathrm{Ent}(\tilde{\nu}|\tilde{\mu}) &= \mathbb{E}^{\tilde{\mu}}\left[\tfrac{d\tilde{\nu}}{d\tilde{\mu}} \ln\left(\tfrac{d\tilde{\nu}}{d\tilde{\mu}}\right)\right] = \mathbb{E}^{\mu}\left[\tfrac{d\tilde{\nu}}{d\tilde{\mu}} \circ T \,\ln\left(\tfrac{d\tilde{\nu}}{d\tilde{\mu}} \circ T\right)\right] \\
&= \mathbb{E}^{\mu}\left[\mathbb{E}^{\mu}\left[\tfrac{d\nu}{d\mu}\Big|\sigma(T)\right] \ln\left(\mathbb{E}^{\mu}\left[\tfrac{d\nu}{d\mu}\Big|\sigma(T)\right]\right)\right] \\
&\leq \mathbb{E}^{\mu}\left[\tfrac{d\nu}{d\mu} \ln\left(\tfrac{d\nu}{d\mu}\right)\right] = \mathrm{Ent}(\nu|\mu),
\end{aligned}$$

where in the last step we used conditional Jensen's inequality for the strictly convex function $x\ln(x)$. Of course, it is clear that if, and only if, $\frac{d\nu}{d\mu}$ is a function of T, we get equality in the last step. □

Remark 3.7.2. We recall that if $G^1 = \mathcal{N}(m_1, \sigma^2)$ is the $1-$dimensional Gaussian measure, and ν is a translate of it, i.e. $\nu = \mathcal{N}(m_2, \sigma^2)$, then the value of the relative entropy is

$$\mathrm{Ent}(\nu|G^1) = \frac{(m_2 - m_1)^2}{2\sigma^2}. \tag{3.7.2}$$

Moreover, this value also coincides with the Wasserstein transport cost between these two measures when the cost functions is quadratic (see [DL82]).

Now we prove that the value of the causal transport problem with the Cameron-Martin cost is bounded above by the relative entropy.

Lemma 3.7.3. *Let γ be the Wiener measure, and $\nu \in \mathcal{P}(\mathcal{C})$. Then we have*

$$\mathcal{W}_2^2(\gamma, \nu) \leq \mathcal{W}_{c,2}^2(\gamma, \nu) \leq 2\,\mathrm{Ent}(\nu|\gamma). \tag{3.7.3}$$

Proof. The leftmost inequality is trivial. To prove the remaining one, for the given measures γ and ν we first construct their finite dimensional projections γ^N and ν^N as discussed in Section 3.5. Then we push forward the marginals by the map

$$\Theta : (x_0, \dots, x_{2^N}) \mapsto 2^{\frac{N}{2}}(x_1 - x_0, \dots, x_{2^N} - x_{2^{N-1}}),$$

in order to guarantee that the starting measure becomes the standard $N-$dimensional Gaussian, denoted by G^N. We denote $\tilde{\nu}^N := \Theta_* \nu^N$. Then from Corollary 2.2.17 we know that

$$\mathcal{W}_{c,2}(\gamma^N, \nu^N) = \inf_{\pi \in \Pi^{\mathcal{F}^N, \mathcal{F}^N}(G^N, \tilde{\nu}^N)} \int \Big[\sum_{k=0}^{2^N} (x_k - y_k)^2 \Big] d\pi(x_0, \dots, x_{2^N}, y_0, \dots, y_{2^N}).$$

By the causal Talagrand inequality (see Proposition 2.5.3) we get that

$$\mathcal{W}_{c,2}^{2,N}(\gamma^N, \nu^N) = \mathcal{W}_{bc,2}^{2,N}(\gamma^N, \nu^N) \leq 2\mathrm{Ent}(\tilde{\nu}^N | G^N) = 2\mathrm{Ent}(\nu^N | \gamma^N),$$

because there is a bijection pushing forward (γ^N, ν^N) to $(G^N, \tilde{\nu}^N)$. From Lemma 3.7.1 we know that $\mathrm{Ent}(\nu^N|\gamma^N) \leq \mathrm{Ent}(\nu|\gamma)$. The result follows from the continuous-time limit of Proposition 3.5.7. □

As a final point of this section, we recover the equality (3.7.1) between causal transport problems and the relative entropy in a special case using our projection arguments. We restrict ourselves to transport problems $\mathcal{W}_{c,2}^2(\gamma, \nu)$ between the Wiener measure γ and the measure $\nu \in \mathcal{P}(\mathcal{C})$, where the latter is the law of a process X that satisfies the following Markovian SDE

$$dX_t = v_t(X_t)dt + dB_t, \quad X_0 = 0, \tag{3.7.4}$$

with continuous and bounded drift v_t and B a Brownian motion. As we have already seen before, the finite-dimensional projection on the dyadic grid of the transport problem $\mathcal{W}_{c,2}^2(\gamma, \nu)$ is given by

$$\mathcal{W}_{c,2}^{2,N}(\gamma^N, \nu^N) = \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\gamma^N, \nu^N)} \int \sum_{i=1}^{2^N - 1} 2^{-N} \left\{ \frac{\Delta x_{i+1} - \Delta y_{i+1}}{2^{-N}} \right\}^2 d\pi(x_0, \dots, x_{2^N}, y_0, \dots, y_{2^N}). \tag{3.7.5}$$

Recall from the duality theory developed in Section 2.1.1 that the dual of the transport problem above is

$$\sup_{\substack{\psi \in C_b(\mathbb{R}^{2^N}), F \in \mathbb{F}, \\ \psi \leq c^N + F}} \int \psi(y_0, \dots, y_{2^N}) d\nu^N(y_0, \dots, y_{2^N}), \tag{3.7.6}$$

where the class of test functions $\mathbb{F}$ was defined in (2.1.5).

Proposition 3.7.4. *Let $\nu \in \mathcal{P}(\mathcal{C})$ be the law of the solution of the SDE* (3.7.4)*. Then we have*

$$\mathcal{W}_{c,2}^2(\gamma, \nu) = 2\mathrm{Ent}(\nu|\gamma).$$

Proof. Note that having proved (3.7.3), we actually only need to show that $\mathcal{W}_{c,2}^2(\gamma, \nu) \geq 2\mathrm{Ent}(\nu|\gamma)$. To do this we are going to use duality and the limiting arguments. We introduce feasible elements for (3.7.6) as follows:

$$\bar{\psi}(y_0, \dots, y_{2^N}) = 2 \sum_{i=1}^{2^N - 1} v_{t_i}(y_i) \Delta y_{i+1} - \sum_{i=1}^{2^N - 1} 2^{-N} v_{t_i}^2(y_i)$$

and

$$\bar{F}(x_0,\dots,x_{2^N},y_0,\dots,y_{2^N}) = 2\sum_{i=1}^{2^N-1} v_{t_i}(y_i)\Delta x_{i+1}.$$

Indeed, it is easy to see that $\bar{F}\in\mathbb{F}$. To show that $\bar{\psi}$ and $\bar{F}$ are admissible dual elements, we need to verify that

$$2\sum_{i=1}^{2^N-1} v_{t_i}(y_i)\Delta y_{i+1} - \sum_{i=1}^{2^N-1} 2^{-N}v_{t_i}^2(y_i) \le \sum_{i=1}^{2^N-1} 2^{-N}\left\{\tfrac{\Delta x_{i+1}-\Delta y_{i+1}}{2^{-N}}\right\}^2 + 2\sum_{i=1}^{2^N-1} v_{t_i}(y_i)\Delta x_{i+1},$$

or equivalently that

$$\sum_{i=1}^{2^N-1} 2^{-N}\left\{\tfrac{\Delta x_{t_{i+1}}-\Delta y_{t_{i+1}}}{2^{-N}}\right\}^2 + 2\sum_{i=1}^{2^N-1} v_{t_i}(y_{t_i})(\Delta x_{t_{i+1}}-\Delta y_{t_{i+1}}) + \sum_{i=1}^{2^N-1} 2^{-N}v_{t_i}^2(y_{t_i}) \ge 0.$$

But it is obviously true, since this inequality is equivalent to

$$\sum_{i=1}^{2^N-1}\left(v_{t_i}(y_i)+\frac{\Delta x_{i+1}-\Delta y_{i+1}}{2^{-N}}\right)^2 \ge 0.$$

And so, we get by weak duality

$$\begin{aligned}\mathcal{W}_{c,2}^{2,N}(\gamma^N,\nu^N) &\ge 2\int\sum_{i=1}^{2^N-1}(v_{t_i}(y_i)\Delta y_{i+1}-2^{-N}v_{t_i}^2(y_i))d\nu^N(y_0,\dots,y_{2^N}) \qquad (3.7.7)\\ &= 2\mathbb{E}\left[\sum_{i=1}^{2^N-1}(v_{t_i}(X_i)\Delta X_{i+1}-2^{-N}v_{t_i}^2(X_i))\right],\end{aligned}$$

where X as in (3.7.4). Taking limits on both sides we get

$$\mathcal{W}_{c,2}^2(\gamma,\nu) \ge 2\int\int v_t(\bar{\omega}_t)d\bar{\omega}_t d\nu(\bar{\omega}) - \int\int v_t^2(\bar{\omega}_t)dtd\nu(\bar{\omega}),$$

by stochastic integration theory and the convergence of Riemann sums. The r.h.s. above equals

$$2\int\int v_t(\bar{\omega}_t)[d\bar{\omega}_t - v_t(\bar{\omega}_t)dt]d\nu(\bar{\omega}) + \int\int v_t^2(\bar{\omega}_t)dtd\nu(\bar{\omega}) = \int\int v_t^2(\bar{\omega}_t)dtd\nu(\bar{\omega}),$$

because the term inside the squared brackets is a $\nu-$ Brownian motion. Together with Proposition 3.5.7, we obtain that

$$\mathcal{W}_{c,2}^2(\gamma,\nu) \ge \int\int_0^1 v_t^2(\bar{\omega}_t)dtd\nu(\bar{\omega}) \ge 2\mathrm{Ent}(\nu|\gamma),$$

where the last inequality follows from [Leh13, Proposition 1], namely by Itô formula. □

Remark 3.7.5. We expect that similar, if more involved arguments, are also applicable if ν is not Markov.

3.8 An explicit construction

So far, we have projected down transport problems from $\mathcal{C}\times\mathcal{C}$ to $\mathbb{R}^N\times\mathbb{R}^N$. In this section we would like to do the opposite and to explain one procedure for embedding a discrete-time optimizer into a feasible transport plan on path space, and discuss further subtleties. For the rest of the section we assume that $\mathcal{W}_{c,2}^2(\gamma,\nu)<\infty$.

We remind that Theorem 2.2.13 implies that $\mathcal{W}_{c,2}(\gamma^N,\nu^N)$ (finite-dimensional projection on the dyadic grid, see (3.5.7)) is attained by a bicausal Monge (Knothe-Rosenblatt) map denoted by $T^N=(T_0^N,\dots,T_{2^N}^N)$, for which in particular $T_*^N\gamma^N=\nu^N$. Therefore, the optimal coupling for the transport problem $\mathcal{W}_{c,2}(\gamma^N,\nu^N)$ is given by $(\mathrm{Id}\times T^N)_*\gamma^N$, or equivalently by $\gamma^N(dx_0,\dots,dx_{2^N})\delta_{T^N(dx_0,\dots,dx_{2^N})}(dy_0,\dots,dy_{2^N})$. In order to embed it in a measure on $\mathcal{C}\times\mathcal{C}$, with some abuse of notation, for a continuous path $\omega\in\mathcal{C}$ we define

$$T_k^N(\omega)=T_k^N(\omega_{t_0},\omega_{t_{2^{-N}}},\dots,\omega_{t_{k2^{-N}}}),\quad \text{for } k=0,\dots,2^N,$$

which is the value attained by the Knothe-Rosenblatt rearrangement on the path ω that is finitely sampled at the points of the dyadic grid $\mathbb{T}^N:=\{k2^N, k=0,\dots,2^{-N}\}$. Further we introduce a process $M_t^N(\omega), t\in[0,1]$ on $\mathcal{C}$ as follows. For each time t there exists a unique index $\bar{k}$ s.t. t belongs to the respective subinterval of the partition of $[0,1]$, i.e. $\bar{k}2^{-N}\le t<(\bar{k}+1)2^{-N}$. We define

$$M_t^N(\omega):=\omega_t+T_{\bar{k}}^N(\omega)-\omega_{t_{\bar{k}2^{-N}}}+(t2^N-\bar{k})\left[T_{\bar{k}+1}^N(\omega)-T_{\bar{k}}^N(\omega)-\left(\omega_{t_{(\bar{k}+1)2^{-N}}}-\omega_{t_{\bar{k}2^{-N}}}\right)\right]. \tag{3.8.1}$$

This is an anticipating continuous process on $[0,1]$, which is the sum of the identity and a piecewise linear process that looks (at most) 2^{-N}-amounts-of-time ahead in the future. Observe that $M_t^N(\omega)-\omega_t$ is nothing more than the piecewise linear interpolation between the values of the vector $T_k^N(\omega)$ on the points of time grid $\mathbb{T}^N$.

By telescopic sum, we get that at the fixed moments of the time grid the values of the process $M_t^N(\omega)$ coincide with the values of the Knothe-Rosenblatt rearrangement, i.e.

$$M_{k2^{-N}}^N(\omega)\;=\;\omega_{t_{k2^{-N}}}+\sum_{s=0}^{k-1}\left\{T_{s+1}^N(\omega)-T_s^N(\omega)-\left(\omega_{t_{(s+1)2^{-N}}}-\omega_{t_{s2^{-N}}}\right)\right\}=T_k^N(\omega), \tag{3.8.2}$$

where $k=0,\dots,2^N$. This implies that the law of the sampled process $(M_0^N,\dots,M_{k2^{-N}}^N,\dots,M_1^N)$ under the Wiener measure γ is exactly ν^N, although the full law of M^N is not ν. Further, we find

$$\frac{d}{dt}(\omega_\cdot-M_\cdot^N(\omega))\;=\;2^N\left[\omega_{t_{(\bar{k}+1)2^{-N}}}-\omega_{t_{\bar{k}2^{-N}}}-\left(T_{\bar{k}+1}^N(\omega)-T_{\bar{k}}^N(\omega)\right)\right]. \tag{3.8.3}$$

Remark 3.8.1. The process M^N is related to the Brownian bridge; to be precise, it is a mixture of Brownian bridges (therefore not necessarily Gaussian nor Markovian), exactly constructed in a strong (pathwise) way to match the second marginal ν^N on the grid $\mathbb{T}^N$. To be more specific, the process M^N is defined in such a way that at the beginning of each interval of the partition of $[0,1]$, we fix the value that the process M^N will take at the end of each time interval.

Having defined the process M^N, we construct a measure Γ on $\mathcal{C}\times\mathcal{C}$ as follows

$$\Gamma(d\omega,d\bar{\omega}):=\gamma(d\omega)\delta_{M^N(\omega)}(d\bar{\omega}).$$

Observe that Γ is not a causal transport plan since the process M^N is anticipative. The cost corresponding to the transport plan Γ is

$$\int\int_0^1\langle\cdot\omega_t-\bar{\omega}_t\cdot\rangle^2dt\,d\Gamma(\omega,\bar{\omega})=\int\int_0^1\left[\tfrac{d}{dt}(\omega-M^N(\omega))\right]^2dt\,d\gamma(\omega)= \tag{3.8.4}$$

$$\int\sum_{k=0}^{2^N-1}2^{-N}\left[\tfrac{\Delta\omega_{t_{k+1}}-\Delta T_{k+1}^N(\omega_{t_0},\omega_{t_1},\dots,\omega_{t_{k+1}})}{2^{-N}}\right]^2d\gamma^N(\omega_{t_0},\dots,\omega_{t_{2^N}})=\mathcal{W}_{c,2}^2(\gamma^N,\nu^N). \tag{3.8.5}$$

And so the non-causal measure Γ on the path space mimics the problem on the $N-$th dyadic grid. This, and the fact that Γ is Monge, are the main advantages of this coupling.

Now we prove that the limits of Γ are causal transport plans in $\Pi^{\mathcal{F},\mathcal{F}}(\gamma,\nu)$. Lemma 3.6.2 implies that the sequence $\{\Gamma_N\}$ is tight. Indeed, the r.h.s. of (3.8.4) is bounded by $\mathcal{W}_{c,2}^2(\gamma,\nu)$, which we assumed to be finite. Let us call $\bar{\Gamma}$ any weak accumulation point of the sequence $\{\Gamma_N\}$. Since

$$(M_0^N, M_{2^{-N}}^N, \ldots, M_1^N)_*\gamma = (T_0^N, \ldots, T_{2^N}^N)_*\gamma^N = \nu^N,$$

$M_*^N\gamma$ converges to ν in the sense of finite dimensional distribution on the dyadics $\cup_N \mathbb{T}^N$, and so by continuity also in the sense of finite dimensional distribution on $[0,1]$. [Bil99, Theorem 7.1] implies that $M_*^N\gamma$ converges weakly to ν, and so the second marginal of $\bar{\Gamma}$ must be ν, which implies that $\bar{\Gamma} \in \Pi(\gamma,\nu)$. Furthermore, since $\omega \mapsto M^N(\omega)$ is 2^{-N}-anticipating, we get by [Las15, Remark 1(ii)] that $\bar{\Gamma}$ is causal, even if each Γ_N was anticipative.

3.9 Conjectures

It is an interesting question to recover the desired equality $\mathcal{W}_{c,2}^2(\gamma,\nu) = 2\mathrm{Ent}(\nu|\gamma)$ by means of our projections arguments in a full generality. Below we suggest some steps in this direction, which are notably different form the duality arguments in Proposition 3.7.4.

Let ν be any measure on the path-space $\mathcal{C}$ with finite entropy w.r.t. the Wiener measure γ. As before, we consider γ^N, ν^N to be the finite-dimensional projections of γ and ν on the dyadic grid $\mathbb{T}^N := \{k2^N, k = 0, \ldots, 2^{-N}\}$. Further, we consider $G^N := \Theta_*\gamma^N$ and $\tilde{\nu}^N := \Theta_*\nu^N$ to be the push-forwards by the process $\Theta(\omega_{t_1}, \ldots, \omega_{t_N}) = 2^{\frac{N}{2}}(\omega_{t_1}, \omega_{t_2} - \omega_{t_1}, \ldots, \omega_{t_N} - \omega_{t_{N-1}})$, as in the proof of Lemma 3.7.1 (see also Corollary 2.2.17 for the related discussion). From this construction we get that G^N is a joint N-dimensional standard Gaussian. Recall that in our particular case the value of the causal transport problem $\mathcal{W}_{c,2}(G^N, \tilde{\nu}^N)$ coincides with the corresponding bicausal one (see Theorem 2.2.13). From the dynamic programming principle obtained in Proposition 2.2.20, we conclude that

$$\mathcal{W}_{c,2}(\gamma^N, \nu^N) = \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(G^N, \tilde{\nu}^N)} \int \sum_{i=0}^{2^N} (x_{t_i} - y_{t_i})^2 \, d\pi(x_{t_0}, \ldots, x_{t_{2^N}}, y_{t_0}, \ldots, y_{t_{2^N}}) \tag{3.9.1}$$

$$= \int \Big\{ \sum_{i=0}^{2^N} \mathcal{W}_2\left(G^1, [\nu^N]^{y_{t_1}, \ldots, y_{t_i}}\right) \Big\} d\nu^N(y_{t_0}, \ldots, y_{t_{2^N}}), \tag{3.9.2}$$

where $[\nu^N]^{y_{t_1}, \ldots, y_{t_i}}$ denotes the kernel of ν^N on $y_{t_{i+1}}$ given $y_{t_1}, \ldots, y_{t_i}$; in the special case when $i = 0$ the notation $[\nu^N]^{y_{t_1}, \ldots, y_{t_0}}$ simply means the (unconditional) first marginal of the measure ν. From Lemma 3.5.5 we see that the l.h.s. of (3.9.1) is increasing in N. Furthermore, it is apparent that

$$\mathcal{W}_{c,2}(\gamma^N, \nu^N) = \int \Big\{ \sum_{i=0}^{2^N} \mathcal{W}_2\left(G^1, L^{i,N}\right) \Big\} d\nu(\bar{\omega}), \tag{3.9.3}$$

where we define the measure $L^{i,N}$ to be the pushforward by the random variable $\omega \mapsto \omega_{t_i}$ of the measure ν conditioned on the sigma-algebra $\mathcal{F}_i^N := \sigma\{\omega_{t_0}, \ldots, \omega_{t_k}, k = 0, \ldots, i2^{-N}\}$, and so $L^{i,N} = L^{i,N}(\omega_{t_0}, \ldots, \omega_{t_i})$ is a measure on $\mathbb{R}$.

On the other hand, from the Girsanov Theorem we know that there exists a so-called drift process v_t, $t \in [0,1]$ such that $d\bar{\omega}_t = v_t dt + d\omega_t$ holds ν-a.s., where ω and $\bar{\omega}$ are Brownian motions under γ and ν respectively. In particular, we have

$$\Delta\bar{\omega}_{t_{i+1}} = v_{t_{i+1}} 2^{-N} + \Delta\omega_{t_{i+1}}.$$

Define now by $\ell^{i,N}$ the ν-law of $v_{t_i}2^{-N} + \Delta\omega_{t_{i+1}}$ given $\mathcal{F}_i^N$, which is Gaussian with mean

$v_{t_i}2^{-N}$ and variance 2^{-N}. From Remark 3.7.2 we know that $\text{Ent}(\ell^{i,N}|\mathcal{N}) = 2^{-1-N}v_{t_i}^2$. For this reason, we get

$$\int \Big\{ \sum_{i=0}^{2^N} \mathcal{W}_2\left(G^1, \ell^{i,N}\right) \Big\} d\nu(\bar{\omega}) = 2 \int \left\{ \sum_{i=0}^{2^N} \text{Ent}(\ell^{i,N}|\mathcal{N}) \right\} d\nu(\bar{\omega}) = \int \Big\{ \sum_{i=0}^{2^N} 2^{-N} v_{t_i}^2 \Big\} d\nu(\bar{\omega})$$

$$\to \int \Big[\int_0^1 v_s^2(\bar{\omega}) ds \Big] d\nu(\bar{\omega}), \tag{3.9.4}$$

where the limit comes from the convergence of Riemann sums plus positivity of the integrand. Further, from Föllmer's formula [Föl86, Proposition 2.11], it is known that the expression (3.9.4) can be bounded below by the relative entropy between γ and ν, and so we get

$$2\text{Ent}(\nu|\gamma) - \lim_{N\to\infty} \int \Big\{ \sum_{i=0}^{2^N} \mathcal{W}_2\left(G^1, \ell^{i,N}\right) \Big\} d\nu(\bar{\omega}) \leq 0.$$

Now taking limit in N together with the Proposition 3.5.7 and (3.9.3), we have shown that

$$\mathcal{W}_{c,2}(\gamma, \nu) \geq 2\text{Ent}(\nu|\gamma) + \lim_{N\to\infty} \int \Big\{ \sum_{i=0}^{2^N} \left[\mathcal{W}_2\left(G^1, L^{i,N}\right) - \mathcal{W}_2\left(G^1, \ell^{i,N}\right) \right] \Big\} d\nu(\bar{\omega}). \tag{3.9.5}$$

We believe that the r.h.s. above, being always smaller than or equal to $\mathcal{W}_{c,2}(\gamma, \nu)$, should actually be equal to this quantity. We also believe that the limit above should be greater than or equal to zero (it suffices to verify this limit omega by omega). All in all one can expect to conclude that $\mathcal{W}_{c,2}(\gamma, \nu) \geq 2\text{Ent}(\nu|\gamma)$, and thus the desired equality $\mathcal{W}_{c,2}(\gamma, \nu) = 2\text{Ent}(\nu|\gamma)$ should follow.

Remark 3.9.1. Note that the conjectures we made above open the doors towards the optimal transport proof of the existence of strong solutions of the Brownian-driven SDEs of the form (3.4.1). Knowing that the optimizer of the projected transport problem is generated my a map, the limiting arguments and the entropy equality should be the building blocks for this proof. However, it is for example unclear, how to guarantee that in the limit a discrete-time optimizer remains of Monge form.

Chapter 4

Enlargements of filtrations and applications in continuous time

In this chapter we consider causal transport problems in an infinite dimensional setting that allows for anticipation of information. Throughout the chapter we talk of *(continuous) semimartingale decomposition*, referring to the unique decomposition of a continuous semimartingale into a continuous local martingale and a continuous finite variation process. The notion of causality can be used to study semimartingale decompositions in the setting of enlargement of filtrations, and this is the object of analysis of this chapter.

After setting up the notations and giving the preliminary results in Section 4.1, we illustrate a first connection between decomposition of semimartingales and causality in Section 4.2. Next, in Section 4.3 we discuss the connection that causality imposes between optional and dual projections of stochastic processes. In Section 4.4 we show how the causal transport problem for specific cost functions allows to characterize semimartingale preservation under filtration enlargement. Necessary and sufficient conditions for a Brownian motion to remain a semimartingale in the enlarged filtration, which is the main result of this chapter, are given in Theorems 4.4.2 and 4.4.4. When the cost function is of Cameron-Martin type, and the filtration enlargement is done entirely at time zero, the causal transport problem can be interpreted in terms of entropy and mutual information; see Section 4.4.3. Thus we are inclined to say that, irrespective of the cost function and the kind of enlargement, the value of our causal transport problems can be seen as a mutual information in a wider sense. A generalization of the definition of causality allows us to determine necessary and sufficient conditions for a general continuous semimartingale to remain a semimartingale with respect to an enlarged filtration. In Section 4.5 we resume the discussion on the duality results initiated in Chapter 2. In particular, we introduce the so called refined dual problem, for which we can give an explicit characterization of optimizers. Finally in Section 4.6 we show how the causal transport problem can be used to estimate the value of additional information with respect to stochastic optimization problems.

Convention for this chapter: Integration w.r.t. a measure π is denoted here as $\mathbb{E}^\pi$, since this is the most efficient and consistent notation in the present setting.

4.1 Path space and filtration enlargement

The setting of this chapter is the same as in the previous one, with the most important difference that we allow for different filtrations on the target space without assuming them to be canonical; for the reminder of notations see Section 3.1. As before, $W = (W_t)_{t\in[0,T]}$ denotes the coordinate process on $\mathcal{C}$, and $(\mathcal{F}_t)_{t\in[0,T]}$ is the right-continuous version of the filtration generated by W. The usual μ-completed filtration containing $\mathcal{F}$ is denoted by $\mathcal{F}^\mu$.

In order to consider *anticipation* of some information regarding the evolution of the coordinate process, we fix arbitrary $\mathcal{F}_T$-measurable functions $\mathfrak{g}_t : \mathcal{C} \to \mathcal{S}$, where $\mathcal{S}$ is a (fixed) Polish space, and enlarge the filtration $\mathcal{F}$ with that of the $\mathcal{S}$-valued process $(\mathfrak{g}_t(W))_{t\in[0,T]}$. This leads to the *(progressively) enlarged filtration* $\mathcal{G} = (\mathcal{G}_t)_{t\in[0,T]}$, defined via

$$\mathcal{G}_t := \bigcap_{0<\epsilon\leq T-t} \mathcal{G}^0_{t+\epsilon},\ t < T, \text{ and } \mathcal{G}_T := \mathcal{G}^0_T, \quad \text{where } \ \mathcal{G}^0_t := \mathcal{F}_t \vee \sigma(\{\mathfrak{g}_s, s \leq t\}). \quad (4.1.1)$$

In the discrete time setting we have already had the first glance of the questions of anticipation of information in a simplified setting, and stated the first results (see Section 2.6). In this chapter we continue the discussion giving a more detailed analysis of the connection between causal transport problems and enlargements of filtrations.

Below we provide the most general definition of causality allowing for different state spaces and filtrations on them. This is to be compared with the one in the discrete-time setting (see Definition 2.1.1).

Definition 4.1.1 (Causal transport plan)**.** *A transport plan $\pi \in \Pi(\mu, \nu)$ is called* causal *between $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ if, for any $t \in [0, T]$ and any set $A \in \mathcal{G}_t$, the map*

$$\mathcal{X} \ni x \mapsto \pi^x(A) = \mathbb{E}^\pi[\mathbb{I}_A|\mathcal{F}_T](x)$$

is measurable with respect to $\mathcal{F}^\mu_t$, where $\pi^x(dy) := \pi(\{x\} \times dy)$ is the regular conditional kernel of π w.r.to the first coordinate.

We stress that in [Las15] the author actually uses a weaker definition, but establishes sufficient conditions for its equivalence with the one we give. We shall need and use the concept of causality as introduced in Definition 4.1.1 throughout this chapter.

For a continuous process Z on a given probability space $(\Omega, \mathbb{P})$, we denote by $\mathcal{F}^Z := Z^{-1}(\mathcal{F})$ the right-continuous version of the filtration generated by Z on Ω, and by $\mathcal{F}^{Z,\mathfrak{g}} := Z^{-1}(\mathcal{G})$ the filtration obtained by enlarging $\mathcal{F}^Z$ with the process $(\mathfrak{g}_t(Z))_{t\in[0,T]}$, in analogy to (4.1.1) (note that $\mathfrak{g}_t(Z)$ are $\mathcal{F}^Z_T$-measurable for all $t \in [0, T]$). For the rest of this section we assume $(\Omega, \mathbb{P})$ to be a fixed probability space. We remind that a pair (X, Y) of continuous processes on $(\Omega, \mathbb{P})$, is called a *causal coupling* w.r.t. the filtrations $\mathcal{F}^X$ and $\mathcal{F}^{Y,\mathfrak{g}}$ if $(X, Y)_*\mathbb{P}$ is a causal transport plan between $(\mathcal{C}, \mathcal{F}, X_*\mathbb{P})$ and $(\mathcal{C}, \mathcal{G}, Y_*\mathbb{P})$.

In what follows, given two measures μ, ν on the space of continuous functions $\mathcal{C} := \mathcal{C}[0, T]$, we will study causal transport plans between $(\mathcal{C}, \mathcal{F}, \mu)$ and $(\mathcal{C}, \mathcal{G}, \nu)$

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \mathbb{E}^\pi[c]. \quad (4.1.2)$$

Working with the path space $\mathcal{C}$ eases the exposition of our analysis. We point out, however, that our results have a natural extension in the multidimensional setting $\mathcal{C}^d$, i.e. for multidimensional continuous processes, see Remarks 4.4.3 and 4.5.4.

As we have already seen in Section 2.1.1, causal transport problems form a subclass of optimal transport problems under linear constraints. The class $\mathbb{F}$ of functions to test causality, that was defined in a discrete time set-up in (2.1.5), in the present setting takes the form

$$\mathbb{F} \;:=\; span\left(\{g\left[f - \mathbb{E}^\mu[f|\mathcal{F}_t]\right] \,:\, f \in C_b(\mathcal{C}), g \in B_b(\mathcal{C}, \mathcal{G}_t), t \in [0,T]\}\right).$$

Then, a transport plan $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ if and only if $\mathbb{E}^\pi[F] = 0$ for all $F \in \mathbb{F}$.

In the following remark we state without proof some useful equivalent characterizations of causality, which were given in Proposition 2.1.5 in discrete time setting and recalled in Remark 3.1.2 in the setting of Wiener spaces and canonical filtrations.

Remark 4.1.2. For a probability measure $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, the following are equivalent:

1. π is a causal transport plan w.r.to $\mathcal{F}$ and $\mathcal{G}$;
2. $\pi\left(\mathcal{X} \times D_t|\mathcal{F}_t \otimes \{\emptyset, \mathcal{Y}\}\right) = \pi\left(\mathcal{X} \times D_t|\mathcal{F}_T \otimes \{\emptyset, \mathcal{Y}\}\right)$ π-a.s., for all $t \in [0,T]$,$D_t \in \mathcal{G}_t$;
3. the σ-fields $\{\emptyset, \mathcal{X}\} \otimes \mathcal{G}_t$ and $\mathcal{F}_T \otimes \{\emptyset, \mathcal{Y}\}$ are conditionally independent with respect to π given $\mathcal{F}_t \otimes \{\emptyset, \mathcal{Y}\}$, for all $t \in [0,T]$;
4. *The H-hypothesis* holds between $\mathcal{F}^\mu \otimes \{\emptyset, \mathcal{Y}\}$ and $\mathcal{F}^\mu \otimes \mathcal{G}$ with respect to π, meaning that every square integrable $(\pi, \mathcal{F}^\mu \otimes \{\emptyset, \mathcal{Y}\})$-martingale is square integrable $(\pi, \mathcal{F}^\mu \otimes \mathcal{G})$-martingale.

We just stress, for the sake of the reader, the reason why causality/conditional independence implies the H-hypothesis: if M is a $(\pi, \mathcal{F}^\mu \otimes \{\emptyset, \mathcal{Y}\})$-martingale, then

$$\mathbb{E}^\pi[M_{t+s}|\mathcal{F}^\mu_t \otimes \mathcal{G}_t] = \mathbb{E}^\pi[M_{t+s}|\mathcal{F}^\mu_t \times \{\emptyset, \mathcal{Y}\}] = M_t,$$

hence M is a $(\pi, \mathcal{F}^\mu \otimes \mathcal{G})$-martingale.

Remark 4.1.3. Definition 4.1.1 and Remark 4.1.2 imply that (X,Y) is a causal coupling if and only if for all t, $\mathcal{F}^{Y,\mathfrak{g}}_t$ and $\mathcal{F}^X_T$ are conditionally independent given $\mathcal{F}^X_t$, that is, the H-hypothesis holds between $\mathcal{F}^X$ and $\mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}}$, with respect to $\mathbb{P}$.

In a Brownian setting there is an easier way to check causality, as the following lemma shows.

Lemma 4.1.4. *Let X be a Brownian motion in its right-continuous natural filtration, and Y be a continuous process. Then the following are equivalent:*

1. *(X,Y) is a causal coupling w.r.t. $\mathcal{F}^X$ and $\mathcal{F}^{Y,\mathfrak{g}}$;*
2. *X is a Brownian motion in $\mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}}$;*
3. *there is a filtration $\mathcal{A}$ on Ω s.t. X is a $\mathcal{A}$-Brownian motion and $\mathcal{F}^{Y,\mathfrak{g}} \subseteq \mathcal{A}$.*

Proof. 1⇒2: We need to show that X is a $\mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}}$-martingale. For $0 \le s < t \le T$ and $f_s \in L^\infty(\mathcal{F}^X_s \vee \mathcal{F}^{Y,\mathfrak{g}}_s)$, we have

$$\begin{aligned}\mathbb{E}\left[X_t f_s\right] &= \mathbb{E}\left[\mathbb{E}\left[X_t f_s \mid (X_u)_{u\le t}\right]\right] = \mathbb{E}\left[X_t \mathbb{E}\left[f_s \mid (X_u)_{u\le t}\right]\right] = \mathbb{E}\left[X_t \mathbb{E}\left[f_s \mid (X_u)_{u\le s}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[X_t \mid (X_u)_{u\le s}\right]\mathbb{E}\left[f_s \mid (X_u)_{u\le s}\right]\right] = \mathbb{E}\left[X_s \mathbb{E}\left[f_s \mid (X_u)_{u\le s}\right]\right] = \mathbb{E}\left[X_s f_s\right],\end{aligned}$$

where causality is used in the third equality.
2⇒3 follows by taking $\mathcal{A} := \mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}}$.
3⇒1: For $t \in [0,T]$ and $D_t \in \mathcal{F}_t^{Y,\mathfrak{g}}$,

$$\mathbb{P}(D_t \mid (X_s)_{s\le T}) = \mathbb{P}(D_t \mid (X_s)_{s\le t}, \{X_{t+u} - X_t, u \in [0, T-t]\}) = \mathbb{P}(D_t \mid (X_s)_{s\le t}),$$

since X is a $\mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}}$-Brownian motion, hence its increments $X_{t+s} - X_t$ are jointly independent of $(X_s)_{s\le t}$ and of the event $D_t \in \mathcal{F}_t^{Y,\mathfrak{g}}$. This shows that (X,Y) is causal, by Definition 2.1.2 and Remark 4.1.2-2. □

If $\mathcal{F}^{Y,\mathfrak{g}} \subseteq \mathcal{F}^X$, then (X,Y) is causal coupling for any martingale X in its own filtration, since in this case $\mathcal{F}^X \vee \mathcal{F}^{Y,\mathfrak{g}} = \mathcal{F}^X$, and $H-$hypothesis trivially holds.

Example 4.1.5 (Poisson bridge). Let M be a Poisson process with constant intensity $\lambda = 1$, $\mathcal{F}_t^M = \sigma(M_s, s \le t)$ its natural filtration and $T > 0$ a fixed time. Let $\mathcal{F}^{M,\mathfrak{g}} = \mathcal{F}_t^M \vee M_T$ be the natural filtration of M enlarged with the terminal value M_T of the process M. Then the process

$$\tilde{M}_t = M_t - \int_0^t \tfrac{M_T - M_s}{T-s} ds,\ t \in [0,T], \tag{4.1.3}$$

is a $\mathcal{F}^{M,\mathfrak{g}}-$martingale (for the proof see [JYC09, Proposition 8.5.1.1]). However, the coupling $(\tilde{M}, M)$ is not causal. To see this, we consider the set $A := \{M_T = 0\} \subset \mathcal{F}_0^{M,\mathfrak{g}}$. Since $M_t = 0$, $\forall t \in [0,T]$ on A, from (4.1.3) we get that $\tilde{M}_t = 0$, $\forall t \in [0,T]$. And so, $\tilde{M}$ is not independent of M_T.

In [Las15, Section 3] the author proves, in the general setting of Polish spaces, that the set $\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ is closed for weak convergence. Thus, an attainability and duality theory for the problem of optimal transport under the causality constraint follows. This is done there at the expense of a regularity assumptions of sorts on the filtration $\mathcal{G}$ (see [Las15, Definition 3]). Such an assumption would, in our context, drastically limit the applicability of the transport approach. However, we can still obtain an attainability and duality theory without assumptions on $\mathcal{G}$. We do this at the price of requiring the first marginal μ to be "weakly continuous", i.e. $\forall f \in C_b(\mathcal{C}),\ t \in [0,T]$ the map

$$x \mapsto \mathbb{E}^\mu[f|\mathcal{F}_t](x) \text{ is continuous.} \tag{4.1.4}$$

(see Assumption 2.1.6 written for discrete time). We state without a proof the duality result written in our particular setting we are working in.

Theorem 4.1.6 (Causal transport duality)**.** *Let $c : \mathcal{C} \times \mathcal{C} \to (-\infty, \infty]$ be bounded from below and lower semicontinuous. Further assume that μ satisfies the weak continuity property* (4.1.4)*. Then* (4.1.2) *is attained and there is no duality gap:*

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \mathbb{E}^\pi[c] = \sup_{\substack{\phi \in C_b(\mathcal{C}),\, \psi \in C_b(\mathcal{C}),\, F \in \mathbb{F} \\ \phi \oplus \psi \le c + F}} \mathbb{E}^\mu[\phi] + \mathbb{E}^\nu[\psi] = \sup_{\substack{\psi \in C_b(\mathcal{C}),\, F \in \mathbb{F} \\ \psi \le c + F}} \mathbb{E}^\nu[\psi],$$

where the class $\mathbb{F}$ is defined as follows

$$\mathbb{F} \ := \ span\left(\{g(\bar{\omega})\left[f(\omega) - \mathbb{E}^\mu[f|\mathcal{F}_t](\omega)\right] \,:\, f \in C_b(\mathcal{C}), g \in B_b(\mathcal{C}, \mathcal{G}_t), t \in [0,T]\}\right). \tag{4.1.5}$$

The proof of this result is essentially the same as in Theorem 2.1.10. The weak continuity property of μ, as a measure on a path space $\mathcal{C}$, is fulfilled if it is the law of a Feller process. So the case that interests us, Wiener measure on continuous path space, is fully covered. This assumption guarantees that the set $\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$ of causal couplings is compact for weak convergence; for the proof see Lemma 2.1.8.

4.2 On enlargements of Brownian filtration

The central question of the theory of enlargements of filtrations is to find the conditions under which the preservation of the semimartingale property holds, i.e. any (semi)martingale in its own filtration can be represented as a sum of a martingale in an enlarged filtration and an adapted finite variation process. The two most studied kinds of filtration enlargements, which are particular cases of (4.1.1) are the following ones:

- *initial enlargement* with a ($\mathcal{F}_T$-measurable) random variable, say L (so $\mathfrak{g}_t = L$, $t \in [0, T]$);
- *progressive enlargement with a random time* (non-negative $\mathcal{F}_T$-measurable random variable), say τ (so $\mathfrak{g}_t = \tau \wedge t$, $t \in [0, T]$), in which case $\mathcal{G}$ turns τ into a stopping time.

Our setting allows us to consider the most general types of enlargements, covering the typical cases mentioned above. In fact, not many works are devoted to the study of general enlargements of filtration beyond initial and progressive enlargements, see e.g. [ADI07] and [KP15]. We refer the reader to the monographs [Jeu80], [MY06], [JYC09, Sec. 5.9], [Pro05, Ch. VI] for an account of the main results and the literature on filtration enlargements.

Below we collect some well-known examples of filtration enlargements in a Brownian setting, which will be useful for future reference (see e.g. [MY06] for these and many other examples). Let B be a Brownian motion on a probability space $(\Omega, \mathbb{P})$. If B remains a semimartingale with respect to the enlarged filtration $\mathcal{F}^{B,\mathfrak{g}}$, then its unique continuous semimartingale decomposition takes the form

$$dB_t = d\tilde{B}_t + dA_t,$$

where $\tilde{B}$ is a $\mathcal{F}^{B,\mathfrak{g}}$-Brownian motion and A is a continuous $\mathcal{F}^{B,\mathfrak{g}}$-adapted finite variation process. In particular, for every finite horizon $T > 0$, by Lemma 4.1.4, we have that $(\tilde{B}, B)$ is a causal transport plan between $\mathcal{F}^{\tilde{B}}$ and $\mathcal{F}^{B,\mathfrak{g}}$, that is, $(\tilde{B}, B)_*\mathbb{P} \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma, \gamma)$.

In what follows we recall some specific enlargement of the filtration $\mathcal{F}^B$, which will be referred to later on in the article.

(1) *Initial enlargement with finitely many atoms*: initial enlargement with a discrete random variable, say L (namely $\mathfrak{g}_t(B) = L$ for all $t \in [0, T]$), that takes values $l_1, \ldots, l_n$, for some $n \in \mathbb{N}$. This corresponds to enlarging the filtration at time zero with the sets $C_i = \{L = l_i\}, i = 1, \ldots, n$. In this case the decomposition of B in the enlarged filtration takes the form

$$dB_t = d\tilde{B}_t + \textstyle\sum_{i=1}^n \mathbb{I}_{C_i} \frac{\eta^i_t}{M^i_t} dt, \tag{4.2.1}$$

where $M^i_t = \mathbb{P}(C_i | \mathcal{F}^B_t)$, which by martingale representation can be written as $M^i_t = M^i_0 + \int\limits_0^t \eta^i_s dB_s$, for some predictable process η^i.

(2) *Brownian bridge*: initial enlargement by the value of the Brownian motion at the terminal time T, namely $\mathfrak{g}_t(B) = B_T$ for all $t \in [0, T]$. In this case it is well-known that the decomposition of B in the enlarged filtration $\mathcal{F}^{B,\mathfrak{g}} = \mathcal{F}^B \vee \sigma(B_T)$ is

$$dB_t = d\tilde{B}_t + \tfrac{B_T - B_t}{T-t} dt. \tag{4.2.2}$$

(3) *Progressive enlargement with last hitting time*: progressive enlargement with the random time $\tau = \sup\{u \le T, B_u = 0\}$. In this case, using the notation $\Phi(x) = \sqrt{2/\pi}\int\limits_x^\infty e^{-u^2/2}du$, the decomposition of B in the enlarged filtration is the following:

$$dB_t = d\tilde{B}_t - \sqrt{\tfrac{2}{\pi}}\frac{\exp\left(-B_t^2/2(T-t)\right)}{\sqrt{T-t}}\left(\frac{\mathrm{sgn}(B_t)}{\Phi(|B_t|/\sqrt{T-t})}\mathbb{I}_{\{t\le\tau\}} - \frac{\mathrm{sgn}(B_T)}{1-\Phi(|B_t|/\sqrt{T-t})}\mathbb{I}_{\{\tau\le t\le T\}}\right)dt. \quad (4.2.3)$$

(4) *Bessel process*: define a 3-dimensional Bessel process by $dR_t = \frac{1}{R_t}dt + dB_t$, and denote $J_t := \inf_{s\ge t} R_s$. Then the process $\tilde{B}_t := R_t - 2J_t$ is a Brownian motion in the filtration obtained by enlarging $\mathcal{F}^B$ with the process J, and

$$dB_t = d\tilde{B}_t + (2dJ_t - \tfrac{1}{R_t}dt).$$

Note that, contrary to the previous cases, here the finite variation process in the semimartingale decomposition of B in the enlarged filtration is not absolutely continuous with respect to Lebesgue measure.

4.3 Optional projections in causal transport

In Sections 4.4 and 4.6 we will intensively use projections and dual projections, which we briefly recall below. We refer to [DM80, Ch. VI] for an accurate study of the subject.

Let X be a positive or bounded measurable process on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$. The *optional projection* of X is the unique (up to indistinguishability) optional process Y such that

$$\mathbb{E}\left[X_\tau \mathbb{I}_{\{\tau<\infty\}}|\mathcal{H}_\tau\right] = Y_\tau \mathbb{I}_{\{\tau<\infty\}} \text{ a.s.}$$

for every stopping time τ. The *predictable projection* of X is the unique (up to indistinguishability) predictable process N such that

$$\mathbb{E}\left[X_\tau \mathbb{I}_{\{\tau<\infty\}}|\mathcal{H}_{\tau-}\right] = N_\tau \mathbb{I}_{\{\tau<\infty\}} \text{ a.s.}$$

for every predictable stopping time τ. These notions of projection can be given for a broader class of processes, including those of integrable variation; see [DM80, Remark VI.44-(f)].

Now, let H be a raw process of integrable variation on $(\Omega, \mathcal{H}, \mathbb{P})$, i.e. a stochastic process which is nonnegative, and whose paths are càdlàg, but which is not $(\mathcal{H}_t)-$ adapted. The *dual optional (resp. predictable) projection* of H is the optional (resp. predictable) integrable variation process U defined by

$$\mathbb{E}\left[\int\limits_0^\infty X_t dU_t\right] = \mathbb{E}\left[\int\limits_0^\infty X_t dH_t\right]$$

for any bounded optional (resp. predictable) process X. W.l.o.g. we assume $U_0 = 0$.

In this chapter, we denote by ${}^o\Lambda^{\mathcal{F}}$ (resp. ${}^{\mathcal{F}}\Lambda^o$) the optional projection (resp. dual optional projection) of a process Λ with respect to $(\pi, \{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi)$. The following lemma is fundamental for the proof of Theorem 4.4.1 and Proposition 4.6.4. It follows directly from [DM80, Lemma 7, Appendix I].

Lemma 4.3.1. *Let $\pi \in \Pi(\mu, \nu)$ be a (non-necessarily causal) transport plan, and let Λ be a $(\mathcal{B}([0,T]) \otimes \mathcal{F}_T \otimes \mathcal{G}_T)$-measurable process on $\mathcal{C} \times \mathcal{C}$ of integrable variation. Then:*

1. *The optional projection of Λ with respect to $(\pi, \{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi)$ (resp. $(\pi, \{\{\emptyset, \mathcal{C}\} \otimes \mathcal{G}\}^\pi)$), which we denote by ${}^o\Lambda^{\mathcal{F}}$ (resp. ${}^o\Lambda^{\mathcal{G}}$) is π-indistinguishable from an optional process with respect to $(\mu, \mathcal{F}^\mu)$ (resp. $(\nu, \mathcal{G}^\nu)$), so w.l.o.g. one may assume*
$$ {}^o\Lambda^{\mathcal{F}}(\omega, \bar{\omega}) = {}^o\Lambda^{\mathcal{F}}(\omega), \qquad {}^o\Lambda^{\mathcal{G}}(\omega, \bar{\omega}) = {}^o\Lambda^{\mathcal{G}}(\bar{\omega}). $$
The analogous statement holds for the predictable projections.

2. *The dual optional projection of Λ with respect to $(\pi, \{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi)$ (resp. $(\pi, \{\{\emptyset, \mathcal{C}\} \otimes \mathcal{G}\}^\pi)$), which we denote by ${}^{\mathcal{F}}\Lambda^o$ (resp. ${}^{\mathcal{G}}\Lambda^o$), is π-indistinguishable from an optional process with respect to $(\mu, \mathcal{F}^\mu)$ (resp. $(\nu, \mathcal{G}^\nu)$), so w.l.o.g. one may assume*
$$ {}^{\mathcal{F}}\Lambda^o(\omega, \bar{\omega}) = {}^{\mathcal{F}}\Lambda^o(\omega), \qquad {}^{\mathcal{G}}\Lambda^o(\omega, \bar{\omega}) = {}^{\mathcal{G}}\Lambda^o(\bar{\omega}). $$
The analogous statement holds for the dual predictable projections.

The next proposition shows the strong relation that causality imposes between projections. This will be crucial in our applications to optimal stopping problems in Section 4.6.1. We comment on the noteworthiness of this result after its proof.

Proposition 4.3.2. *Let $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$ be a causal transport plan, and let Λ be an $(\mathcal{F} \otimes \mathcal{G})$-adapted, integrable variation càdlàg process with $\Lambda_0 = 0$. Then*
$$ {}^{\mathcal{F}}\Lambda^o \quad \text{is } \pi\text{-indistinguishable from} \quad {}^o\Lambda^{\mathcal{F}}. $$

Proof. We drop the superscript $\mathcal{F}$ to simplify the notation. Fix $t \in [0, T]$, and consider the raw process $X = (X_s)_{s \in [0,t]}$, with constant paths, given by
$$ X_s = \mathbb{I}_{\{{}^o\Lambda_t > \Lambda^o_t\}}, \qquad s \in [0, t]. $$
Its optional projection with respect to $(\pi, \{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi)$ satisfies
$$ {}^oX_s = \pi[{}^o\Lambda_t > \Lambda^o_t | \{\mathcal{F}_s \otimes \{\emptyset, \mathcal{C}\}\}^\pi], \qquad s \in [0, t]. $$
Note that $({}^oX_s)_{s \in [0,t]}$ is an $\{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi$-martingale, hence π-indistinguishable from an $(\mathcal{F} \otimes \{\emptyset, \mathcal{C}\})$-martingale, by [DM80, Lemma 7, Appendix I], which is then also an $(\mathcal{F} \otimes \mathcal{G})$-martingale by causality; see Remark 4.1.2. This means that $({}^oX_s)_{s \in [0,t]}$, optional projection of X w.r.t. $\{\mathcal{F} \otimes \{\emptyset, \mathcal{C}\}\}^\pi$, is the càdlàg version of the $(\mathcal{F} \otimes \mathcal{G})$-martingale $s \mapsto \pi[{}^o\Lambda_t > \Lambda^o_t | \mathcal{F}_s \otimes \mathcal{G}_s]$, $0 \le s \le t$, thus π-indistinguishable from the optional projection of X w.r.t. $(\pi, \{\mathcal{F} \otimes \mathcal{G}\}^\pi)$.

Now, by definition of optional projection,
$$ \mathbb{E}^\pi[\Lambda_t X_t] = \mathbb{E}^\pi[{}^o\Lambda_t X_t]. $$
On the other hand, since X is constant and $\Lambda_0 = 0$,
$$ \mathbb{E}^\pi[\Lambda_t X_t] = \mathbb{E}^\pi[\Lambda_t X_0] = \mathbb{E}^\pi\left[\int_0^t X_s d\Lambda_s\right]. $$
Now, using [DM80, Remark VI.58-(d)] and the fact that Λ is an $(\mathcal{F}\otimes\mathcal{G})$-adapted integrable variation process, we have that $\mathbb{E}^\pi\left[\int_0^t X_s d\Lambda_s\right] = \mathbb{E}^\pi\left[\int_0^t Y_s d\Lambda_s\right]$, where $(Y_s)_{s \in [0,t]}$ is the optional projection of X w.r.t. $\mathcal{F} \otimes \mathcal{G}$. Therefore we have
$$ \begin{aligned} \mathbb{E}^\pi[\Lambda_t X_t] &= \mathbb{E}^\pi\left[\int_0^t Y_s d\Lambda_s\right] = \mathbb{E}^\pi\left[\int_0^t {}^oX_s d\Lambda_s\right] = \mathbb{E}^\pi\left[\int_0^t X_s d\Lambda^o_s\right] \\ &= \mathbb{E}^\pi\left[\Lambda^o_t X_t\right] - \mathbb{E}^\pi\left[\Lambda^o_0 X_t\right] = \mathbb{E}^\pi\left[\Lambda^o_t X_t\right], \end{aligned} $$

where in the second equality we use that ${}^oX$ and Y are π-indistinguishable, and in the third one the definition of dual optional projection (recall $\Lambda^o_0 = 0$). Altogether we have

$$\mathbb{E}^\pi[{}^o\Lambda_t \mathbb{I}_{\{{}^o\Lambda_t > \Lambda^o_t\}}] = \mathbb{E}^\pi[\Lambda^o_t \mathbb{I}_{\{{}^o\Lambda_t > \Lambda^o_t\}}],$$

hence $\{{}^o\Lambda_t > \Lambda^o_t\}$ is π-negligible. Arguing similarly, we get that $\{{}^o\Lambda_t \neq \Lambda^o_t\}$ is π-negligible. Since this is true for all $t \in [0, T]$, we have that ${}^o\Lambda$ and Λ^o are versions of each other and hence π-indistinguishable since both càdlàg (see [DM80, Theorem VI.47]). $\square$

The essential difference between ${}^o\Lambda^{\mathcal{F}}$ and ${}^{\mathcal{F}}\Lambda^o$, as explained in [DM80, Remark VI.74-(c)], is that while the first one formalizes $\mathbb{E}[\Lambda_t|\mathcal{F}_t]$ the second one does so to $\int_0^t \mathbb{E}[d\Lambda_s|\mathcal{F}_s]$. What we justify in Proposition 4.3.2 is the fact that under causality we have:

$$\begin{aligned} {}^{\mathcal{F}}\Lambda^o_t - {}^{\mathcal{F}}\Lambda^o_0 &= \int_0^t \mathbb{E}^\pi[d\Lambda_s|\mathcal{F}_s] = \int_0^t \mathbb{E}^\pi[d\Lambda_s|\mathcal{F}_t] = \mathbb{E}^\pi[\int_0^t d\Lambda_s|\mathcal{F}_t] \\ &= \mathbb{E}^\pi[\Lambda_t - \Lambda_0|\mathcal{F}_t] = {}^o\Lambda^{\mathcal{F}}_t - {}^o\Lambda^{\mathcal{F}}_0. \end{aligned}$$

This phenomenon is not symmetric, just as causality, i.e. one does not expect Proposition 4.3.2 to hold for projections w.r.t. $\mathcal{G}$.

4.4 Causal optimal transport and semimartingale decomposition

In this section we give one of our main results. We show the connection between the semimartingale decompositions arising in enlargement of filtrations, and certain causal optimal transport problems. We consider both the case when the finite variation part in the semimartingale decomposition is absolutely continuous and when it is not; we do it separately for the sake of applications and connection with the literature. For simplicity of exposition all the results are shown in a one-dimensional Brownian setting. For the extension of the next results and arguments to the case of multidimensional continuous paths see Remark 4.4.3 and4.5.4.

Working in the continuous path space framework of Section 4.1, we consider the filtrations $\mathcal{F}$ and $\mathcal{G}$ as defined there. Recall that γ is the Wiener measure on the path space $\mathcal{C}$, and that $(\omega, \bar\omega)$ denotes a generic element in $\mathcal{C} \times \mathcal{C}$. By id we mean the identity mapping on $\mathcal{C}$. The following notation will be frequently used: for a process/path Z we denote by

$$V_t(Z) \ = \sup_{0\le t_1\le\cdots\le t_k\le t} \ \sum_{i<k} |Z_{t_i} - Z_{t_{i+1}}|$$

the total variation of Z up to time t.

4.4.1 The general case

We first need to obtain the following result, reminiscent of [Las15]:

Theorem 4.4.1. *Let ν be a measure on $\mathcal{C}$ such that $\nu \ll \gamma$. Then the following are equivalent:*

(i) for some continuous, $\mathcal{G}^\nu$-adapted, integrable variation process $A = A(\bar\omega)$, i.e. such that

$$\mathbb{E}^\nu[V_T(A)] < \infty, \tag{4.4.1}$$

the process $\xi_t(\bar\omega) := \bar\omega_t - A_t(\bar\omega)$ is a $(\nu, \mathcal{G}^\nu)$-Brownian motion;

(ii) the following causal optimal transport problem is finite:

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)} \mathbb{E}^\pi[V_T(\bar\omega-\omega)]. \tag{4.4.2}$$

Moreover, whenever (i)-(ii) hold, we have that:
1. the transport plan $\hat\pi := (\xi, \mathrm{id})_*\nu$ belongs to $\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$, and is optimal for (4.4.2);
2. for every transport plan $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ with finite cost in (4.4.2), the process $\tilde A(\omega,\bar\omega) := A(\bar\omega)$ in (i) is the $(\pi, \{\{\emptyset, \mathcal{C}\}\otimes\mathcal{G}\}^\pi)$-dual predictable projection of the process $(\omega,\bar\omega) \mapsto \Lambda(\omega,\bar\omega) := \bar\omega - \omega$.

As we have commented before, since the Wiener measure γ satisfies the weak continuity property (4.1.4), the set $\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ is weakly compact. Moreover, the total variation of the difference of the coordinate processes is bounded from below and lower semicontinuous w.r.t. supremum norm, hence Theorem 4.1.6 apply, giving us the primal attainability in (4.4.2) and duality. In the ensuing proof, attainability is also established but via stochastic analysis arguments.

Proof. $(i) \Rightarrow (ii)$: From the process ξ in (i), define the coupling $\hat\pi := (\xi, \mathrm{id})_*\nu$. The fact that $\hat\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ follows as in the proof of Lemma 4.1.4, and $\mathbb{E}^{\hat\pi}[V_T(\bar\omega-\omega)] = \mathbb{E}^\nu[V_T(A)(\bar\omega)] < \infty$.

$(ii) \Rightarrow (i)$: Fix $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ with $\mathbb{E}^\pi[V_T(\bar\omega-\omega)] < \infty$. The continuous process $\Lambda_t(\omega,\bar\omega) := \bar\omega_t - \omega_t$ is of integrable variation with respect to π, hence we can define its dual predictable projection $\tilde A(\omega,\bar\omega)$ with respect to $\{\{\emptyset,\mathcal{C}\}\otimes\mathcal{G}\}^\pi$, the π-completion of $\{\emptyset,\mathcal{C}\}\otimes\mathcal{G}$. In particular, $\tilde A$ is an integrable variation process on $\mathcal{C}\times\mathcal{C}$. By Lemma 4.3.1, we may assume that $\tilde A$ does not depend on the first coordinate, thus it corresponds to a $\mathcal{G}^\nu$-predictable integrable variation process A on $\mathcal{C}$, in the sense that $\tilde A(\omega,\bar\omega) = A(\bar\omega)$, which gives (4.4.1). Altogether we have

$$\mathbb{E}^\pi\left[\int_0^T X_t(\bar\omega)d\Lambda_t(\omega,\bar\omega)\right] = \mathbb{E}^\nu\left[\int_0^T X_t(\bar\omega)dA_t(\bar\omega)\right], \tag{4.4.3}$$

for every $\mathcal{G}^\nu$-predictable bounded process X. Moreover, we have

$$\mathbb{E}^\nu[V_T(A)] = \mathbb{E}^\pi[V_T(\tilde A)] \le \mathbb{E}^\pi[V_T(\Lambda)] < \infty. \tag{4.4.4}$$

Now, note that the jump times of A are $\mathcal{G}^\nu$-predictable, by [DM80, Theorem B, page xiii], and that for each jump time τ, $\Delta A_\tau = \mathbb{E}^\pi[\Delta\Lambda_\tau|\{\emptyset,\mathcal{C}\}\otimes\mathcal{G}^\nu_{\tau-}] = 0$ a.s., from the continuity of Λ; see [DM80, Theorem VI.76]. Therefore, A is continuous.

We now define the process ξ as in (i), and need to show that it is a $(\nu,\mathcal{G}^\nu)$-Brownian motion. For $0 \le s < t \le T$ and $f_s \in L^\infty(\mathcal{G}^\nu_s)$, we have

$$\begin{aligned}
\mathbb{E}^\nu[(\xi_t(\bar\omega)-\xi_s(\bar\omega))f_s(\bar\omega)] &= \mathbb{E}^\nu[(\bar\omega_t-\bar\omega_s-\textstyle\int_s^t dA_u(\bar\omega))f_s(\bar\omega)]\\
&= \mathbb{E}^\pi[(\omega_t-\omega_s)f_s(\bar\omega)] + \mathbb{E}^\pi[(\textstyle\int_s^t d\Lambda_u - \int_s^t dA_u(\bar\omega))f_s(\bar\omega)]\\
&= \mathbb{E}^\pi[(\textstyle\int_s^t d\Lambda_u - \int_s^t dA_u(\bar\omega))f_s(\bar\omega)] = 0,
\end{aligned}$$

where the third equality follows since ω, which is a $(\gamma,\mathcal{F}^\gamma)$-martingale, is consequently by causality a $(\pi,\mathcal{F}^\gamma\otimes\mathcal{G})$-martingale, thus also a $(\pi,(\mathcal{F}^\gamma\otimes\mathcal{G})^\pi)$-martingale and in particular a $(\pi,\mathcal{F}^\gamma\otimes\mathcal{G}^\nu)$-martingale. The last equality follows from (4.4.3) with $X_t := f_s 1_{]s,T]}(t)$. This shows that ξ is a $(\nu,\mathcal{G}^\nu)$-martingale, and we conclude by an application of Lévy theorem together with Girsanov theorem; indeed, the quadratic variation of ξ at t must be t, by the assumption $\nu \ll \gamma$.

Finally, with the aid of (4.4.4), we see that the transport plan $\hat{\pi} := (\xi, \mathrm{id})_*\nu$ is optimal for (4.4.2). Moreover, since $\bar{\omega}$ is continuous, the uniqueness of the semimartingale decomposition entails that the process A found in the proof of $(ii) \Rightarrow (i)$ must have not depended on which π (with finite cost in (4.4.2)) we started with. $\square$

We now state our main result, that provides a necessary and sufficient condition for a Brownian motion to remain a semimartingale in an enlarged filtration. We use the notations introduced in Section 4.1.

Theorem 4.4.2 (Semimartingale preservation property)**.** *The following are equivalent:*

(i) any process B which is a Brownian motion in its natural filtration $\mathcal{F}^B$ on some probability space $(\Omega, \mathbb{P})$, remains a semimartingale in the enlarged filtration $\mathcal{F}^{B,\mathfrak{g}(B)}$;

(ii) the causal transport problem (4.4.2) *is finite for some measure $\nu \sim \gamma$.*

Moreover, when (i)-(ii) hold, and denoting by $B = \widetilde{B} + N$ the semimartingale decomposition of B in $\mathcal{F}^{B,\mathfrak{g}}$, we have that $(\widetilde{B}, B)$ is a causal coupling with respect to $\mathcal{F}^{\tilde{B}}$ and $\mathcal{F}^{B,\mathfrak{g}}$.

Proof. $(ii) \Rightarrow (i)$: By Theorem 4.4.1, there exists a continuous, $\mathcal{G}^\nu$-adapted, integrable variation process A such that the process $\xi_t(\bar{\omega}) := \bar{\omega}_t - A_t(\bar{\omega})$ is a $(\nu, \mathcal{G})$-Brownian motion. Since $\nu \sim \gamma$, Girsanov theorem implies that $\bar{\omega}$ is a $(\gamma, \mathcal{G})$-semimartingale. Moreover, since $\bar{\omega}$ is the coordinate process on $\mathcal{C}$, and from $B_*\mathbb{P} = \gamma$ and $\mathcal{F}^{B,\mathfrak{g}} = B^{-1}(\mathcal{G})$, we have that B is a $(\mathbb{P}, \mathcal{F}^{B,\mathfrak{g}})$-semimartingale.

$(i) \Rightarrow (ii)$: Let $B = M + U$ be the semimartingale decomposition of B in $\mathcal{F}^{B,\mathfrak{g}}$, with M a $(\mathbb{P}, \mathcal{F}^{B,\mathfrak{g}})$-Brownian motion and U a finite variation process, so that in particular $V_T(U) < \infty$. Since $0 < (1 + V_T(U))^{-1} \leq 1$, we have

$$c^{-1} := \mathbb{E}^{\mathbb{P}}\left[(1 + V_T(U))^{-1}\right] \in (0, 1],$$

and $Z_T := c\,(1 + V_T(U))^{-1} > 0$ $\mathbb{P}$-a.s., with $\mathbb{E}^{\mathbb{P}}[Z_T] = 1$. We can then define a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}^B_T)$ via $\frac{d\mathbb{Q}}{d\mathbb{P}} := Z_T$, so that $\mathbb{Q} \sim \mathbb{P}$ and

$$\mathbb{E}^{\mathbb{Q}}[V_T(U)] = c\mathbb{E}^{\mathbb{P}}\left[\tfrac{V_T(U)}{1+V_T(U)}\right] \leq c < \infty. \tag{4.4.5}$$

Let Z be the $(\mathbb{P}, \mathcal{F}^{B,\mathfrak{g}})$-martingale defined from Z_T. Then, by Girsanov theorem, B has decomposition

$$B_t = \widetilde{M}_t + \left(\int_0^t \tfrac{d\langle Z,M\rangle_s}{Z_s} + U_t\right),$$

where $\widetilde{M} := M - \int_0^\cdot \frac{d\langle Z,M\rangle_s}{Z_s}$ is a $(\mathbb{Q}, \mathcal{F}^{B,\mathfrak{g}})$-Brownian motion. Moreover,

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[\int_0^T \left|\tfrac{d\langle Z,M\rangle_s}{Z_s}\right|\right] &= \mathbb{E}^{\mathbb{P}}\left[\int_0^T |d\langle Z, M\rangle_s|\right] \leq \mathbb{E}^{\mathbb{P}}\left[\sqrt{\langle Z, Z\rangle_T}\sqrt{\langle M, M\rangle_T}\right] \\
&= \sqrt{T}\mathbb{E}^{\mathbb{P}}\left[\sqrt{\langle Z, Z\rangle_T}\right] \leq K\sqrt{T}\mathbb{E}^{\mathbb{P}}\left[\sup_{t\in[0,T]} Z_t\right] \leq cK\sqrt{T},
\end{aligned}$$

by the Kunita-Watanabe inequality and the Burkholder-Davis-Gundy inequality (with constant K). Together with (4.4.5), this implies that the process $N := \int_0^\cdot \frac{d\langle Z,M\rangle_s}{Z_s} + U$ is of $\mathbb{Q}$-integrable variation. This shows that (ii) holds with $\nu := B_*\mathbb{Q}$ (as $\gamma = B_*\mathbb{P}$ and $\mathbb{Q} \sim \mathbb{P}$ imply $\nu \sim \gamma$).

Finally, by Lemma 4.1.4, $(\widetilde{B}, B)$ is a causal coupling with respect to $\mathcal{F}^{\tilde{B}}$ and $\mathcal{F}^{B,\mathfrak{g}}$. $\square$

From the above results it is clear that if there is one causal transport in $\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$, for some measure $\nu \sim \gamma$, for which the difference of the coordinates is of integrable variation and a.s. absolutely continuous, then the finite variation part in the semimartingale decomposition of the Brownian motion in the enlarged filtration is also absolutely continuous, see Section 4.4.2.

Remark 4.4.3 (Multidimensional processes). We want to point out that the previous theorems can be easily extended to a multidimensional setting. Indeed, instead of the path space $\mathcal{C}$, that accommodates 1-dimensional continuous processes, we can consider $\mathcal{C}^d$, path space for d-dimensional continuous processes. In this case, we write

$$\{(a^i_t)_{t\in[0,T]}\}_{i=1}^d \mapsto V_T(a) := \sum_{i=1}^d V_T(a^i)$$

in (4.4.1)-(4.4.2) for the variation of a multidimensional process. Then the proof of Theorem 4.4.1 follows exactly the same arguments, where the dual projections are now taken componentwise. As for Theorem 4.4.2, one should define Z_T as $c/(1+\sum_{i=1}^d V_T(U_i))$ instead.

4.4.2 The absolutely continuous case

In many well-known filtration enlargements, the finite variation part in the semimartingale decomposition of the Brownian motion in the enlarged filtration is absolutely continuous with respect to Lebesgue (as in the examples (4.2.1), (4.2.2) and (4.2.3) above), i.e. is of the form

$$dB_t = d\widetilde{B}_t + b_t(B)dt.$$

This is true, for example, in the case of initial enlargement of filtrations under Jacod's assumption (see [Jac85]) and under Yor's method (see [Yor12, Section 12.1]), as well as in the case of progressive enlargement with a random time (see [JY78] and [Jeu80]); for general enlargements see [ADI06]. That is why this is a framework of major interest which deserves a deeper analysis. In analogy to Theorems 4.4.1 and 4.4.2, we can give necessary and sufficient conditions for such a decomposition to hold, together with a characterization of the process b in terms of causal transport.

Theorem 4.4.4. *Let ν be some measure on $\mathcal{C}$ such that $\nu \ll \gamma$, and let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a convex even function such that $\rho(+\infty) = +\infty$ and $\rho(0) = 0$. Then the following are equivalent:*

(i) for some $\mathcal{G}$-predictable process $\alpha = \alpha(\bar\omega)$ such that

$$\mathbb{E}^\nu\left[\int_0^T \rho(\alpha_s)ds\right] < \infty,$$

the process $\xi_t(\bar\omega) := \bar\omega_t - \int_0^t \alpha_s(\bar\omega)ds$ is a $(\nu, \mathcal{G}^\nu)$-Brownian motion;

(ii) the following causal optimal transport problem is finite:

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)} \mathbb{E}^\pi\left[\int_0^T \rho(\langle\!\cdot\,\bar\omega_t - \omega_t\!\cdot\rangle)dt\right]. \tag{4.4.6}$$

Moreover, whenever (i)-(ii) hold, then $\hat\pi := (\xi, \mathrm{id})_*\nu$ belongs to $\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$, it is optimal for (4.4.6), and for every $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ with finite cost in (4.4.6), it holds that the *process* $\tilde{\alpha}_t(\omega, \bar{\omega}) := \alpha(\bar{\omega})$ *equals the predictable projection of* $\langle \cdot \bar{\omega}_t - \omega_t \cdot \rangle$ *with respect to* $(\pi, \{\emptyset, \mathcal{C}\} \otimes \mathcal{G}^\nu)$.

Theorem 4.4.5 (Semimartingale preservation property). *The following are equivalent:*

(i) any process B *which is a Brownian motion in its natural filtration* $\mathcal{F}^B$ *on some probability space* $(\Omega, \mathbb{P})$*, remains a semimartingale in the enlarged filtration* $\mathcal{F}^{B,\mathfrak{g}}$*, with decomposition*

$$dB_t = d\widetilde{B}_t + b_t(B)dt; \tag{4.4.7}$$

(ii) the causal transport problem (4.4.6) *is finite for some measure* $\nu \sim \gamma$ *and some function* ρ *as in Theorem 4.4.4.*

Moreover, if (ii) *holds for* $\nu = \gamma$*, then the value of the causal transport problem* (4.4.6) *equals* $\mathbb{E}^\gamma[\int_0^T \rho(b_t(B))dt] < \infty$*, hence the information drift in* (4.4.7) *is* ρ*-integrable.*

The proofs of the above theorems follow the same steps of the proofs of Theorems 4.4.1 and 4.4.2, so we omit them. One simply observes that $\rho(x) \geq m|x| + n$ for some $m > 0$, so the relevant processes in the proofs are of integrable and a.s. absolutely continuous variation. Then one recalls that for an integrable variation process X which is absolutely continuous, say $X = \int_0^\cdot a_t dt$, the dual predictable projection of X w.r.t. some filtration $\mathcal{H}$ is also absolutely continuous, and is indistinguishable from both $\int_0^\cdot {}^p a_t dt$ and $\int_0^\cdot {}^o a_t dt$, where ${}^p a$ and ${}^o a$ are the predictable and optional projections of a w.r.t. $\mathcal{H}$, respectively.

Remark 4.4.6. For $\rho(x) = x^2/2$, the cost in (4.4.6) is Cameron-Martin cost, which was discussed in details in Chapter 3. In this case, by Theorem 4.4.5 finiteness of problem (4.4.6) for $\nu = \gamma$ is equivalent to square integrability of the drift in (4.4.7). When this holds, one can apply Girsanov theorem, which ensures that B is a Brownian motion in $\mathcal{G}$ under a change of measure. Therefore, by martingale representation, all $\mathcal{F}^B$-semimartingales remain semimartingales w.r.t. $\mathcal{F}^{B,\mathfrak{g}}$. Square integrability of the drift holds for example when initially enlarging with a discrete random variable as in case (1) of Section 4.2, while it fails in the Brownian bridge case and for progressive enlargements with last hitting times, as in (2) and (3) of Section 4.2.

Remark 4.4.7 (General continuous semimartingales). The theorems above, as well as those in the non-absolutely continuous setting, have an analogue outside the Brownian framework. In order to establish this, we need a condition for transport plans that generalizes the concept of causality introduced in Definition 4.1.1, namely:

$$\mathbb{E}^\pi[(\omega_t - \omega_s) f_s(\bar{\omega})] = 0 \qquad \forall\, 0 \leq s < t \leq T,\ f_s \in L^\infty(\mathcal{C}, \mathcal{G}_s^\nu, \nu). \tag{4.4.8}$$

In particular, if X is a continuous semimartingale on a probability space $(\Omega, \mathbb{P})$, which remains a semimartingale in the enlarged filtration $\mathcal{F}^{X,\mathfrak{g}}$ with canonical decomposition $X = \widetilde{X} + N$, then the transport plan $(\widetilde{X}, X)_* \mathbb{P}$ satisfies (4.4.8).

In this framework, an analogue of Theorem 4.4.1 can be established, where now the process ξ in (i) is only required to be a $(\nu, \mathcal{G}^\nu)$-martingale, and where the optimal transport problem in (ii) is formulated over transport plans in $\Pi(\mu, \nu)$, for some martingale law μ, which satisfy (4.4.8). This result then leads to an analogue of Theorem 4.4.2, giving a necessary and sufficient condition for any continuous semimartingale X to remain a semimartingale in the enlarged filtration $\mathcal{F}^{X,\mathfrak{g}}$. In the same way one has the analogues of Theorems 4.4.4 and 4.4.5 for general continuous semimartingales.

4.4.3 Initial enlargement: Jacod's condition, entropy, and mutual information

The focus of this section is on the initial enlargement of canonical filtration with some random variable. Our aim here is to connect the classical results from enlargement theory, the existing results in the literature on utility maximization of informed agents and our transport approach.

The well-known Jacod's condition (see [Jac85]) guarantees that semimartingale preservation property in the case of initial enlargement holds if and only if for each $t < T$

$$\eta_t^\omega(dx) \ll Law(L),$$

where $\eta_t^\omega(dx)$ is a family of regular conditional distributions such that $\eta_t^{\cdot}(A)$ is a version of $\mathbb{E}[\mathbb{I}_{L\in A}|\mathcal{F}_t]$.

As explained in the "Comparison with Jacod's condition" Section in [ADI07], Jacod's method for initial enlargements [Jac85] (see our Sections 4.1 and 4.2) can be interpreted in the following way: starting with a Brownian motion B on some probability space $(\Omega, \mathbb{P})$, considering an initial enlargement of the Brownian filtration $\mathcal{F}^B$ with a random variable $L(B)$ and assuming that for almost all x, $\mathbb{P}^{L=x} := \mathbb{P}(\cdot|L(B) = x) \ll \mathbb{P}$, one applies Girsanov theorem and finds $B - A^x$ to be a (local) martingale w.r.t. $\mathbb{P}^{L=x}$. Then combining these, one obtains that $B - A^{L(B)}$ is a (local) martingale w.r.t. $\mathbb{P}$ and the enlarged filtration $\mathcal{F}^{B,L}$. Remember that under Jacod's condition the finite variation process $A^{L(B)}$ is absolutely continuous with respect to Lebesgue, that is, $A_t^{L(B)} = \int_0^t \alpha_s ds$. Notably, there is a causal optimal transport counterpart to the above mentioned method. Denoting γ to be the law of B,

$$\ell(dx) := L(B)_*\mathbb{P}, \quad \text{and } \gamma^{L=x} = \text{“conditional law of } B \text{ given } L(B) = x\text{”},$$

we have

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}^L}(\gamma,\gamma)} \mathbb{E}^\pi\Big[\int_0^T \rho(\langle\!\cdot\bar{\omega}_t - \omega_t\cdot\!\rangle)dt\Big] = \int \inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}}(\gamma,\gamma^{L=x})} \mathbb{E}^\pi\Big[\int_0^T \rho(\langle\!\cdot\bar{\omega}_t - \omega_t\cdot\!\rangle)dt\Big]\ell(dx). \quad (4.4.9)$$

The proof of the above result was essentially given in Proposition 2.6.1 in discrete time.

From [Las15], we know that in the Cameron-Martin case of $\rho(x) = x^2/2$ the integrand in the r.h.s. of (4.4.9) equals the relative entropy of $\gamma^{L=x}$ w.r.t. γ, whenever this is finite. For us this means that

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}^L}(\gamma,\gamma)} \tfrac{1}{2}\mathbb{E}^\pi\Big[\int_0^T (\langle\!\cdot\bar{\omega}_t - \omega_t\cdot\!\rangle)^2 dt\Big] = \int \mathrm{Ent}(\gamma^{L=x}|\gamma)\ell(dx). \quad (4.4.10)$$

Since the relative entropy $\mathrm{Ent}(\gamma^{L=x}|\gamma)$ is further integrated w.r.t. the law of $L(B)$, we get that the r.h.s. in (4.4.10) corresponds to the so-called *Mutual Information* between B and $L(B)$, denoted by $\mathrm{I}(B, L(B))$. It is defined as the relative entropy of the joint law $P_{B,L(B)}$ w.r.t. the decoupling measure $P_B \otimes P_{L(B)}$, namely:

$$\mathrm{I}(B, L(B)) := \mathrm{Ent}(P_{B,L(B)}|P_B \otimes P_{L(B)}).$$

This value is infinite if $\gamma^{L=x} \not\ll \gamma$ $\forall x$ ℓ-a.s. Otherwise, we prove the above claimed equality.

Lemma 4.4.8. *Assuming that $\gamma^{L=x} \ll \gamma$ $\forall x$ ℓ-a.s., then*

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}^L}(\gamma,\gamma)} \mathbb{E}^\pi\Big[\tfrac{1}{2}\int_0^T (\langle\!\cdot\bar{\omega}_t - \omega_t\cdot\!\rangle)^2 dt\Big] = \int Ent(\gamma^{L=x}|\gamma)l(dx) = I(B, L(B)).$$

Proof. By definition of mutual information we have

$$\begin{aligned} I(B,L(B)) &= \int\int \frac{dP_{B,L(B)}}{d(P_B\otimes P_{L(B)})}\ln\left(\frac{dP_{B,L(B)}}{d(P_B\otimes P_{L(B)})}\right)dP_B dP_{L(B)} \\ &= \int\left(\int \frac{d\gamma^{L=x}}{d\gamma}\ln\left(\frac{d\gamma^{L=x}}{d\gamma}\right)d\gamma\right)l(dx) \\ &= \int \mathrm{Ent}(\gamma^{L=x}|\gamma)l(dx). \end{aligned}$$

□

Note that the lemma above recovers the result of [ADI06, Theorem 5.13] using our methods. If the initial enlargement is done by a discrete random variable as in Section 4.2-(1), then

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}^L}(\gamma,\gamma)} \mathbb{E}^\pi\left[\frac{1}{2}\int_0^T (\langle\bar{\omega}_t-\omega_t\rangle)^2dt\right] = -\sum_{i=1}^n p_i\ln(p_i),$$

where $p_i=\mathbb{P}(C_i)$, and the term on the r.h.s. is referred to as the entropy of the partition $\{C_i\}_{i=1}^n$; see [Yor85] and [AIS98].

On the other hand, by Theorem 4.4.5, the l.h.s. in (4.4.10) is finite if and only if the information drift α in the semimartingale decomposition of B w.r.t. $\mathcal{F}^{B,L}$ is square integrable, in which case the value of the causal transport problem equals $\frac{1}{2}\mathbb{E}^\gamma[\int_0^T\alpha_t^2dt]$. In [PK96] (see also [ADI06], [AIS98]) it is proved that this value corresponds to the additional utility obtained by an investors who maximizes the expected log-utility of terminal wealth in a certain complete market model w.r.t. $\mathcal{F}^{B,L}$, compared to an investor w.r.t. $\mathcal{F}^B$. Further, it is known that this value also equals the relative entropy $\mathrm{Ent}(\mathbb{P}|\mathbb{Q})$, where $\mathbb{Q}$ is a probability measure under which B is a $(\mathbb{Q},\mathcal{F}^{B,L})$-Brownian motion; see Remark 4.4.6. Putting things together, we have

$$\inf_{\pi\in\Pi^{\mathcal{F},\mathcal{F}^L}(\gamma,\gamma)} \frac{1}{2}\mathbb{E}^\pi\left[\int_0^T (\langle\bar{\omega}_t-\omega_t\rangle)^2dt\right] = \mathrm{Ent}(\mathbb{P}|\mathbb{Q}) = \int \mathrm{Ent}(\gamma^{L=x}|\gamma)l(dx) = \mathrm{I}(B,L(B)),$$

under the assumption that $\gamma^{L=x}\ll\gamma$ $\forall x$ ℓ-a.s.

4.5 Refined Duality

As we have already seen, the proofs of Theorems 4.4.1 and 4.4.2 as well as the sketches of the proofs of Theorems 4.4.4 and 4.4.5 are of stochastic analysis flavour. In this section we discuss what the optimal transport perspective has to say in the absolutely continuous case. Throughout this section we consider ν to be a fixed measure on $\mathcal{C}$.

An interesting observation is that in this setting we can actually say more about the problem dual to (4.4.6), which corresponds to

$$\sup_{\substack{\psi\in C_b(\mathcal{C}),\,F\in\mathbb{F}\\ \psi(\bar{\omega})\le\int_0^T\rho(\langle\bar{\omega}_t-\omega_t\rangle)dt+F(\omega,\bar{\omega})}} \left\{\int\psi(\bar{\omega})\nu(d\bar{\omega})\right\}, \tag{4.5.1}$$

where the class $\mathbb{F}$ is defined in (4.1.5). Let us define the *refined dual* problem as

$$\sup_{\zeta\in S_a(\mathcal{G})} \left\{\int\left[\int_0^T\zeta_t(\bar{\omega})d\bar{\omega}_t-\int_0^T\rho^*(\zeta_t(\bar{\omega}))dt\right]\nu(d\bar{\omega})\right\}, \tag{4.5.2}$$

where ρ^* denotes the convex conjugate of ρ, and $S_a(\mathcal{G})$ denotes the set of simple previsible processes w.r.t. $\mathcal{G}$. Namely, for $\mathcal{G}$-stopping times $0 \le \tau_1 \le \cdots \le \tau_m \le T$ we define

$$S_a(\mathcal{G}) := \left\{ \sum_{i=1}^{m} \zeta^i(\bar\omega)\mathbb{I}_{(\tau_i,\tau_{i+1}]}(t) : m \in \mathbb{N}, \zeta^i \in B_b(\mathcal{G}_{\tau_i}) \right\},$$

so the first integral in (4.5.2) is defined as a finite sum as customary.

The appeal of the refined dual (4.5.2) comes from the next argument:

Lemma 4.5.1. *Let $\rho : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex. Weak-refined duality holds, in the sense that the value of the primal problem* (4.4.6) *can only be larger than that of the refined dual* (4.5.2)*.*

Proof. W.l.o.g. we assume that (4.4.6) is finite. Let $\psi(\bar\omega) = \int_0^T \zeta_t(\bar\omega)d\bar\omega_t - \int_0^T \rho^*(\zeta_t(\bar\omega))dt$ and $F(\omega,\bar\omega) = \int_0^T \zeta_t(\bar\omega)d\omega_t \in \mathbb{F}$, with $\zeta \in S_a(\mathcal{G})$. Taking any $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\nu)$ with finite cost, and denoting $\langle\!\cdot\bar\omega_t - \omega_t\cdot\!\rangle = \alpha_t$, by Fenchel-Young inequality we have

$$\int_0^T \zeta_t(\bar\omega)d(\bar\omega_t - \omega_t) \le \int_0^T \rho(\alpha_t)dt + \int_0^T \rho^*(\zeta_t(\bar\omega))dt.$$

Hence,

$$\psi(\bar\omega) \le \int_0^T \rho(\alpha_t)dt + F(\omega,\bar\omega) = \int_0^T \rho(\langle\!\cdot\bar\omega_t - \omega_t\cdot\!\rangle)dt + F(\omega,\bar\omega) \quad \pi\text{-a.s.}$$

Since by causality $\mathbb{E}^\pi[F] = 0$, we conclude that the value of the primal problem is not any less than (4.5.2). □

Actually, we can show that strong duality holds in this setting. We point out that even without anticipation of information (i.e. $\mathcal{F} = \mathcal{G}$) the following theorem is a novel result.

Theorem 4.5.2. *Let ν be some measure on $\mathcal{C}$ such that $\nu \ll \gamma$, and let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a strictly convex even function such that $\rho(+\infty) = +\infty$ and $\rho(0) = 0$. Suppose further that ρ satisfies for some $C, x_0 > 0$ and $\ell > 1$:*
$\rho(x) = 0 \Leftrightarrow x = 0$, $\frac{\rho(0)}{0} = 0$, $\frac{\rho(+\infty)}{+\infty} = +\infty$, $\rho(2x) \le C\rho(x)$ *and* $\rho(x) \le \frac{\rho(\ell x)}{2\ell}$ *if* $x \ge x_0$.
Then:
(i) The primal (4.4.6)*, the dual* (4.5.1) *and the refined dual* (4.5.2)*, have the same value.*

From now on we assume that this common value is finite.

(ii) The refined dual (4.5.2) *can be computed (without changing its value) over the set M^{ρ^*}, which is the closure of $S_a(\mathcal{G})$ w.r.t. the so-called gauge norm*

$$\zeta \mapsto \|\zeta\|_{\rho^*} := \inf\left\{ \beta > 0 : \int \int_0^T \rho^*(\zeta_t(\bar\omega)/\beta)dt\, \nu(d\bar\omega) \le 1 \right\}, \tag{4.5.3}$$

and it is attained there by a unique optimizer $\hat\zeta = \hat\zeta(\bar\omega)$.
(iii) The optimal drift α in Theorem 4.4.4(i) is related to $\hat\zeta$ through

$$\rho^*(\hat\zeta_t(\bar\omega)) + \rho(\alpha_t(\bar\omega)) = \alpha_t(\bar\omega)\hat\zeta_t(\bar\omega) \quad d\nu \otimes dt\text{-a.s,} \tag{4.5.4}$$

namely $\alpha_t(\bar\omega) = (\rho^*)'(\hat\zeta_t(\bar\omega))$, or equivalently, $\hat\zeta_t(\bar\omega) \in \partial\rho(\alpha_t(\bar\omega))$, with ∂ denoting sub-differential.
(iv) $\xi_t(\bar\omega) := \bar\omega_t - \int_0^t \alpha_s(\bar\omega)ds$ is a $(\nu, \mathcal{G}^\nu)$-Brownian motion, so if further $\gamma \ll \nu$ we have that the canonical process $\bar\omega$ is a $(\gamma, \mathcal{G}^\nu)$-semimartingale.

The typical examples for which the given conditions on ρ are satisfied, are power functions $\rho(x) \sim x^p$ with exponent $1 < p < +\infty$, which covers the Cameron-Martin case, as well as $\rho(x) = |x|^a(1 + |\log|x||)$ for $a > 1$; see the comments after [RR91, Ch. II.2.3, Corollary 4]. Note that the formula (4.5.4) gives the recipe to find a primal optimizer if a dual one is known.

Remark 4.5.3. We stress that the dual (4.5.1) is most often not attained on continuous functions. Still, the refined dual (4.5.2) admits an optimizer which induces a formal optimal element for (4.5.1) by setting $F(\omega, \bar\omega) := \int_0^T \hat\zeta_t(\bar\omega)d\omega_t$ and

$$\psi(\bar\omega) := \int_0^T \hat\zeta_t(\bar\omega)d\bar\omega_t - \int_0^T \rho^*(\hat\zeta_t(\bar\omega))dt = \int_0^T \hat\zeta_t(\bar\omega)d\xi_t(\bar\omega) + \int_0^T \rho(\alpha_t(\bar\omega))dt.$$

Notice that, even though the optimal $\hat\zeta$ depends on the cost function ρ, the optimal drift α does not. In the Cameron-Martin case $\rho(x) = x^2/2$, we actually get from (4.5.4) that $\hat\zeta = \alpha$ and so

$$\psi(\bar\omega) = \int_0^T \alpha_t(\bar\omega)d\bar\omega_t - \int_0^T (\alpha_t(\bar\omega))^2/2\, dt.$$

Furthermore, in the absence of enlargement (i.e. $\mathcal{F} = \mathcal{G}$) we find by Girsanov theorem the identity

$$\tfrac{d\nu}{d\gamma}(\bar\omega) = \exp(\psi(\bar\omega)).$$

In words: the causal Kantorovich potential ψ between Wiener measure and ν is the logarithm of their relative density.

Remark 4.5.4 (Multidimensional processes). As seen in Remark 4.4.3 for the general case, also Theorem 4.5.2 has an analogue in the multidimensional setting. It suffices to define the gauge norm (4.5.3) as acting on the euclidean norm of $\zeta_t(\bar\omega)$, interpret the r.h.s. of (4.5.4) as inner product, etc. This is straightforward, so we do not give the details.

4.5.1 Extensions of stochastic integrals and elements of Orlicz space theory

In order to prove Theorem 4.5.2 we need to extend the stochastic integral $\int_0^T \zeta_t(\bar\omega)d\bar\omega_t$ beyond simple $\mathcal{G}$-previsible integrands via functional analytic arguments, much inspired by [Léo12]. Of course, this could have been done via Theorem 4.4.4, using that a fortiori the coordinate process $\bar\omega$ is a $(\nu, \mathcal{G})$-semimartingale. We avoided this to show that there is a true transport/functional method for this. Likewise, Point (iv) is obtained without using previous results. But first, we recall the necessary elementary facts on Orlicz spaces needed throughout this section.

As presented in [RR91], a convex even function $\Phi : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ satisfying $\Phi(0) = 0$ and $\Phi(\infty) = \infty$, is called a *Young function.* If such a function is finite-valued, it is zero only at the origin, and satisfies both $\Phi(0)/0 = 0$ and $\Phi(\infty)/\infty = \infty$, then it is called an *N-function.* We remark that Φ is a Young function (resp. N-function) if and only if its conjugate Φ^* is so.

From now on we identify processes which are $d\nu \times dt$-a.e. equal. Assuming that Φ is a Young function, we define

$$M^\Phi := \left\{\zeta : \mathcal{C} \times [0,T] \to \mathbb{R},\ \text{s.t. } \zeta \text{ is } \mathcal{G}\text{-previsible and } \forall k > 0 : \int\int_0^T \Phi(k\zeta_t(\bar\omega))dt\nu(d\bar\omega) < \infty\right\},$$

which is a closed subspace (sometimes called Orlicz heart or Morse-Transue space) of the so-called Orlicz space

$$L^{\Phi} := \left\{\zeta : \mathcal{C} \times [0,T] \to \mathbb{R}, \text{ s.t. } \zeta \text{ is } \mathcal{G}\text{-previsible and } \exists k > 0 : \int\int_0^T \Phi(k\zeta_t(\bar\omega))dt\nu(d\bar\omega) < \infty\right\},$$

when endowed with the *gauge norm*

$$\|\zeta\|_{\Phi} := \inf\left\{\beta > 0 : \int\int_0^T \Phi(\zeta_t(\bar\omega)/\beta)dt\,\nu(d\bar\omega) \le 1\right\}.$$

The gauge norm actually turns L^{ρ^*} into a Banach space, and if e.g. Φ is an N-function then the norm-dual of M^{Φ} is L^{Φ^*} ([RR91, Ch. III.3.3, Theorem 10] and [RR91, Ch. IV.4.1, Theorem 6]).

We now introduce growth conditions on Φ and Φ^*. We say that a Young function Φ is in Δ_2 if there are some $C, x_0 > 0$ s.t. whenever $x \ge x_0$ we have $\Phi(2x) \le C\Phi(x)$. This is equivalent to the following condition on Φ^*: there exist $\ell > 1, x_0 > 0$ s.t. whenever $x \ge x_0$ we have $\Phi^*(x) \le \frac{\Phi^*(\ell x)}{2\ell}$. When Φ is an N-function, then by [RR91, Ch. II.2.3, Theorem 3]:

$$\Phi \text{ in } \Delta_2 \iff \exists x_0 > 0, \epsilon > 1 \text{ s.t. } \left[x \ge x_0 \Rightarrow \frac{x\Phi'(x)}{\Phi(x)} \le \epsilon\right]. \tag{4.5.5}$$

Clearly, when Φ is in Δ_2 we have $M^{\Phi} = L^{\Phi}$. The reflexivity of L^{Φ} is essentially equivalent to Φ and Φ^* being in Δ_2.

Below we provide a technical lemma needed in the proof of Theorem 4.5.2:

Lemma 4.5.5. *If ρ^* is an N-function in Δ_2, then*

$$(\omega, \bar\omega) \in \mathcal{C} \times \mathcal{C} \mapsto \int_0^T \rho(\langle\!\cdot\bar\omega_t - \omega_t\!\cdot\rangle)dt$$

is lower semicontinuous when $\mathcal{C} \times \mathcal{C}$ is equipped with the "sum of the uniform norms" norm.

Proof. It suffices to show the lower semicontinuity of $\bar\omega \mapsto r(\bar\omega) := \int_0^T \rho(\langle\!\cdot\bar\omega_t\!\cdot\rangle)dt$. Let $\mathcal{C}_K := \{\bar\omega \in \mathcal{C} : r(\bar\omega) \le K\}$. Since ρ^* is in Δ_2, we have by [RR91, Ch. II.2.3, Corollary 5] that ρ grows at least as fast as some power function with exponent $p > 1$. This shows that $\mathcal{C}_K$ is bounded in $W^{1,p}([0,T])$, the Sobolev space of absolutely continuous functions with p-integrable first derivative. By classical arguments we have that if $\bar\omega^n \to \bar\omega$ uniformly, with $\bar\omega^n \in \mathcal{C}_K$, then $\bar\omega \in W^{1,p}([0,T])$ and in particular $\bar\omega$ is absolutely continuous too. By Fatou's Lemma we further get $\bar\omega \in \mathcal{C}_K$, yielding the desired result. □

Now we are ready to define a stochastic integral $\int_0^T \zeta_t(\bar\omega)d\bar\omega_t$ for integrands from the set M^{ρ^*}.

Lemma 4.5.6. *Suppose that ρ^* is a Young function having a global minimum at the origin. Then*

$$\overline{S_a(\mathcal{G})}^{\|\cdot\|_{\rho^*}} = M^{\rho^*}. \tag{4.5.6}$$

If further the refined dual (4.5.2) *is finite, then the functional*

$$\zeta = \sum \zeta^i \mathbb{I}_{(\tau_i, \tau_{i+1}]} \in S_a(\mathcal{G}) \mapsto \int\int_0^T \zeta_t d\bar\omega_t\,\nu(d\bar\omega) := \int \sum \zeta^i(\bar\omega)[\bar\omega_{\tau_{i+1}} - \bar\omega_{\tau_i}]\nu(d\bar\omega)$$

can be uniquely extended to M^{ρ^*} by continuity. Using the same notation for its extension, we can replace the optimization variables in (4.5.2) *by taking "$\zeta \in M^{\rho^*}$" without changing the value of the optimization problem.*

Proof. By [RR91, Ch. III.3.4, Proposition 3], we have (4.5.6) under the given hypotheses; the measure considered being $d\nu \times dt$, the sigma-algebra being the $\mathcal{G}$-previsible one, and the Young function being ρ^*. One need only observe that the previsible sigma-algebra is generated by the algebra of sets, whose elements are finite disjoint unions of "base" sets of the form $D \times (\underline{\tau}, \overline{\tau}]$, with $D \in \mathcal{G}_{\underline{\tau}}$ and $\mathcal{G}$-stopping times $\underline{\tau} \leq \overline{\tau}$. If we denote by v the value of problem (4.5.2), from now on assumed finite, we then have for all $F \in S_a(\mathcal{G}), \beta > 0$:

$$\int \int_0^T (\zeta_t/\beta) d\bar{\omega}_t \, \nu(d\bar{\omega}) \leq v + \int \int_0^T \rho^*(\zeta_t(\bar{\omega})/\beta) dt \, \nu(d\bar{\omega}),$$

so by definition

$$\int \int_0^T \zeta_t \, d\bar{\omega}_t \, \nu(d\bar{\omega}) \leq (v+1)\|\zeta\|_{\rho^*}. \tag{4.5.7}$$

This shows that the discrete integral, seen as a continuous linear functional on $S_a(\mathcal{G})$, can be uniquely and continuously extended to the norm closure of this space, which we know to coincide with M^{ρ^*}. Because the convex functional $\zeta \mapsto \int \int_0^T g^*(\zeta_t) \, dt \, \nu(d\bar{\omega})$ is finite throughout M^{ρ^*}, for all $\zeta, \bar{\zeta} \in M^{\rho^*}$ with $\|\zeta - \bar{\zeta}\|_{\rho^*} \leq 1/2$ we have that

$$\int \int_0^T \rho^*(\bar{\zeta}_t) \, dt \, \nu(d\bar{\omega}) \leq 1/2 \int \int_0^T \rho^*(2\zeta_t) \, dt \, \nu(d\bar{\omega}) + 1/2 \int \int_0^T \rho^*\left(\frac{\zeta_t - \bar{\zeta}_t}{1/2}\right) \, dt \, \nu(d\bar{\omega}) \leq C_\zeta + 1,$$

where C_ζ is a constant only depending on ζ. This shows that the convex functional is locally bounded and thus continuous by classical results; the last statement then follows. □

The lemma below shows that M^{ρ^*} is the smallest set, where the refined dual problem (4.5.2) is attained.

Lemma 4.5.7. *Assume that ρ^* (equiv. ρ) is an N-function, that ρ is in Δ_2, and that the refined dual* (4.5.2) *is finite. Then* (4.5.2) *is attained in M^{ρ^*}.*

Proof. By Lemma 4.5.6, we can consider (4.5.2) as defined over M^{ρ^*}. By [RR91, Ch. IV.4.1, Theorem 6] we see that L^{ρ^*} is the norm dual space of M^ρ, so the classical Banach-Alaoglu's theorem implies that closed balls in M^{ρ^*} are $\sigma(M^{\rho^*}, M^\rho)$-compact, since M^{ρ^*} is a norm-closed and convex subset of L^{ρ^*}. By [RR91, Ch. V.5.3, Theorem 3] and the comment following its proof, we see that ρ in Δ_2 implies that

$$\lim_{\|\zeta\|_{\rho^*} \to \infty} \frac{\int \int_0^T \rho^*(\zeta_t) dt \, \nu(d\bar{\omega})}{\|\zeta\|_{\rho^*}} = +\infty. \tag{4.5.8}$$

As a consequence of this and (4.5.7) (which holds also in M^{ρ^*}), we get

$$\int [\int_0^T \zeta_t d\bar{\omega}_t - \int_0^T \rho^*(\zeta_t) dt] \nu(d\bar{\omega}) \leq \|\zeta\|_{\rho^*} \{ v + 1 - \frac{\int \int_0^T \rho^*(\zeta_t) dt \, \nu(d\bar{\omega})}{\|\zeta\|_{\rho^*}} \},$$

where v is the value of (4.5.2), and by (4.5.8) the r.h.s. above goes to $-\infty$ as $\|\zeta\|_{\rho^*} \to \infty$. This shows that in computing (4.5.2) one may restrict the problem to a big enough fixed ball in M^{ρ^*}, which is $\sigma(M^{\rho^*}, M^\rho)$-compact. As in the end of the proof of Lemma 4.5.6, we observe that the objective function is norm-continuous, and because it is concave it is also $\sigma(M^{\rho^*}, M^\rho)$-upper semicontinuous. The existence of an optimizer in M^{ρ^*} follows. □

Lemma 4.5.8. *Suppose* (4.5.2) *is finite and attained by some $\hat{\zeta} \in M^{\rho^*}$, and that ρ^* is a differentiable N-function (equiv. ρ is a strictly convex N-function) which is in Δ_2. Setting*

$$\alpha_t(\bar{\omega}) := (\rho^*)'(\hat{\zeta}_t(\bar{\omega})) \quad \text{and} \quad \xi_t(\bar{\omega}) := \bar{\omega}_t - \int_0^t \alpha_t(\bar{\omega}) dt,$$

we have that $\int \int_0^T \rho(\alpha_t(\bar{\omega})) dt \, \nu(d\bar{\omega}) < +\infty$, and so $\alpha \in L^\rho$ and ξ is a $(\nu, \mathcal{G})$-martingale.

If further $\nu \ll \gamma$, then ξ is a $(\nu, \mathcal{G})$-Brownian motion, $(\xi, \mathrm{id})_* \nu \in \Pi^{\mathcal{F}, \mathcal{G}}(\gamma, \nu)$, the primal problem (4.4.6) *is finite and its value is equal to* (4.5.1) *and* (4.5.2).

Proof. We first observe that by the identity of sub-differentials

$$\int\int_0^T \rho\circ(\rho^*)'(\hat{\zeta}_t(\bar{\omega}))dt\,\nu(d\bar{\omega}) = -\int\int_0^T \rho^*(\hat{\zeta}_t(\bar{\omega}))dt\,\nu(d\bar{\omega}) + \int\int_0^T \hat{\zeta}_t(\bar{\omega})(\rho^*)'(\hat{\zeta}_t(\bar{\omega}))dt\,\nu(d\bar{\omega}), \tag{4.5.9}$$

so the finiteness of the l.h.s. is equivalent to the finiteness of the second term in the r.h.s. Since ρ^* is an N-function in Δ_2 we have by (4.5.5) that when $\hat{\zeta}_t(\bar{\omega})$ is large the integrand $\hat{\zeta}_t(\bar{\omega})(\rho^*)'(\hat{\zeta}_t(\bar{\omega}))$ is dominated by a (fixed) constant times $\rho^*(\hat{\zeta}_t(\bar{\omega}))$, and so we conclude that the left- and right-hand sides above are indeed finite. In particular $\alpha \in L^\rho$ holds, and the map

$$h \in M^{\rho^*} \mapsto \int\int_0^T \alpha_t(\bar{\omega})h_t(\bar{\omega})dt\nu(d\bar{\omega})$$

is finite-valued and continuous by [RR91, Ch. III.3.3, Proposition 1].

Since (4.5.2) is a concave problem, we have that if $\hat{\zeta}$ is optimizer of $\sup_{\zeta\in M^{\rho^*}} H(\zeta)$, then $\forall h \in M^{\rho^*}$ it holds that $\frac{\partial}{\partial\varepsilon}H(\hat{\zeta}+\varepsilon h)|_{\varepsilon=0} = 0$, where

$$H(\zeta) := [\int\int_0^T \zeta_t(\bar{\omega})\ d\bar{\omega}_t\nu(d\bar{\omega}) - \int\int_0^T \rho^*(\zeta_t(\bar{\omega}))dt\nu(d\bar{\omega})].$$

Thus we get that

$$\int\int_0^T h_t(\bar{\omega})d\bar{\omega}_t\nu(d\bar{\omega}) - \int\int_0^T \alpha_t(\bar{\omega})h_t(\bar{\omega})dt\nu(d\bar{\omega}) = 0 \quad \forall h \in M^{\rho^*}.$$

This means that $\int\int_0^T h_t(\bar{\omega})d\xi_t\nu(d\bar{\omega}) = 0$ for all such h, which implies that ξ is indeed a $(\nu,\mathcal{G})$-martingale. Since the bracket of the canonical process is the identity under γ, this is inherited by ξ by Girsanov theorem under the assumption $\nu \ll \gamma$, so by Levy's theorem ξ is then a $(\nu,\mathcal{G})$-Brownian motion. By Lemma 4.1.4, $(\xi, id)_*\nu$ is causal. In light of the finiteness in (4.5.9), this proves that the primal problem (4.4.6) is finite. The lower semicontinuity of $(\omega,\bar{\omega}) \mapsto \int_0^T \rho(\langle\bar{\omega}_t - \omega_t\rangle)dt$ was established in Lemma 4.5.5, and so by Theorem 4.1.6 we get that there is no duality gap. To conclude the proof, we only need to check the equality between (4.4.6) and the refined dual. For this, we rewrite (4.5.9) as

$$\int\int_0^T \rho(\alpha_t(\bar{\omega}))dt\,\nu(d\bar{\omega}) = -\int\int_0^T \rho^*(\hat{\zeta}_t(\bar{\omega}))dt\,\nu(d\bar{\omega}) + \int\int_0^T \hat{\zeta}_t(\bar{\omega})d\bar{\omega}_t\,\nu(d\bar{\omega}),$$

where we used that $\int\int_0^T \hat{\zeta}_t(\bar{\omega})d\xi_t\nu(d\bar{\omega}) = 0$. This proves that the refined dual has a greater value than the primal, and we conclude by Lemma 4.5.1. □

We can finally give the proof of Theorem 4.5.2:

Proof of Theorem 4.5.2. Under the assumptions made, both ρ and ρ^* are N-functions in Δ_2. For Point (i), and thanks to Lemma 4.5.1, we only need to prove that when the refined dual is finite, its value coincides with the primal one. This follows by Lemmas 4.5.7 and 4.5.8. Point (ii) is contained in Lemmas 4.5.6 and 4.5.7. Point (iii) follows from Lemma 4.5.8, and because the strict convexity of ρ implies the differentiability of ρ^*. Point (iv) is given by the last lemma as well together with an application of Girsanov theorem. □

We stress that Point (iv) of Theorem 4.5.2 can also be obtained via more sophisticated stochastic analysis arguments: if the primal problem (equiv. the refined dual) is finite, then as in the proof of Lemma 4.5.6 one shows that for all $\zeta \in S_a(\mathcal{G}) : \int\int_0^T \zeta_t\,d\bar{\omega}_t\,\nu(d\bar{\omega}) \leq (v+1)\|\zeta\|_{\rho^*}$, where v is the common optimal value. If $|\zeta| \leq C$ then $\|\zeta\|_{\rho^*} \leq C\|1\|_{\rho^*} < \infty$. This suggests that $\{\int_0^T \zeta_t\,d\bar{\omega}_t : \zeta \in S_a(\mathcal{G}),\ \|\zeta\|_\infty \leq 1\}$ is bounded in ν-probability, so by the Bichteler-Dellacherie theorem $\bar{\omega}$ is a $(\nu, \mathcal{G})$-semimartingale, and we conclude by Girsanov theorem. Such a proof is reminiscent of original arguments in [Jac85].

4.6 Applications to stochastic optimization problems

In this section we show how the causal transport framework allows us to give an estimate of the value of additional information, for some classical stochastic optimization problems in continuous-time. By value of information we mean the difference between the optimal value of these problems with and without additional information (i.e. w.r.t. the enlarged and the original filtration, respectively). The main idea is to take "causal projections" of candidate optimizers in the problem with the larger filtration, so building a feasible element in the problem with the smaller filtration, making a comparison possible. In discrete-time this idea goes back to [Pfl09, PP12], where the application is not about enlarging filtrations, but about gauging the dependence of multistage stochastic programming problems w.r.t. the reference probability measures (we recalled this in Section 2.4.2). We start with Section 4.6.1, on optimal stopping problems, for which the outlined projection approach is more delicate and fully novel to the best of our knowledge (even in discrete-time). Then in Section 4.6.2 we deal with utility maximization with portfolio constraints; this is a prominent example of a controlled linear system, and indeed the same arguments would be applicable to such systems in general. In both optimization problems considered below, we will obtain very robust estimates in terms of causal minimization problem.

4.6.1 Optimal stopping

Here we consider the framework of Section 4.1, with canonical space $\mathcal{C} := \mathcal{C}[0, T]$ and filtrations $\mathcal{F} \subset \mathcal{G}$ on it. We begin with the definition of a randomized stopping time:

Definition 4.6.1. *A randomized stopping time Σ with respect to a filtration $\mathcal{H}$ and a probability measure $\mu \in \mathcal{P}(\mathcal{C})$, written $\Sigma \in RST(\mathcal{H}, \mu)$, is an increasing right-continuous $\mathcal{H}$-adapted process on $[0, T]$, with $\Sigma_0 = 0$ and $\Sigma_T = 1$, $\mu - a.s.$*

This notion generalizes the concept of stopping time, say τ, according to which a path ω is stopped at a unique point in time $\tau(\omega)$. Stopping according to a randomized stopping time Σ means that a path ω is stopped in $[0, t]$ with probability $\Sigma_t(\omega)$. We recommend [BCH16, Section 3.2] for a modern view on this matter, and refer to [BC77] for the original motivation/definition.

The next lemma is of fundamental importance for our applications. It identifies what causal dual optional projections do to randomized stopping times:

Lemma 4.6.2. *Let $\Sigma \in RST(\mathcal{G}, \nu)$. Then, for any $\mu \in \mathcal{P}(\mathcal{C})$ and any causal transport plan $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu, \nu)$, there is a randomized stopping time $\tilde{\Sigma} \in RST(\mathcal{F}, \mu)$ such that*

$$\mathbb{E}^{\pi}\left[\int_0^T \ell(\omega, t) d\Sigma_t(\bar{\omega})\right] = \mathbb{E}^{\mu}\left[\int_0^T \ell(\omega, t) d\tilde{\Sigma}_t(\omega)\right],$$

for all $\mathcal{F}$-optional processes $(\omega, t) \mapsto \ell(\omega, t)$ which are bounded or positive.

Proof. Let $\tilde{\Sigma}(\omega,\bar{\omega})$ be the dual optional projection of $\Sigma = \Sigma(\bar{\omega})$ with respect to $(\pi, (\mathcal{F}\otimes\{\emptyset,\mathcal{C}\})^\pi)$. From Lemma 4.3.1, we may assume that $\tilde{\Sigma}(\omega,\bar{\omega}) = \tilde{\Sigma}(\omega)$, and from Proposition 4.3.2 we have that $\tilde{\Sigma}$ equals the optional projection of Σ with respect to $(\pi, (\mathcal{F}\otimes\{\emptyset,\mathcal{C}\})^\pi)$. Moreover, by [DM80, Lemma 7, Appendix I], we can assume $\tilde{\Sigma}$ to be $(\mathcal{F}\otimes\{\emptyset,\mathcal{C}\})$-optional. This implies that $\tilde{\Sigma}$ lies in the interval $[0,1]$ too, and hence belongs to $RST(\mathcal{F},\mu)$. □

Our purpose is to quantitatively gauge, via causal transport arguments, the dependence of optimal stopping problems on $[0,T]$ with respect to the filtration or the reference probability measure. See [LP90, CT07] or the seminal but unpublished work [Ald81], for the related issue of (qualitative) stability of these problems. Lemma 4.6.2 above suggests that we should rather define optimal stopping over randomized stopping times. It is well-known that, in the non-anticipative case, one can move between formulations over stopping times and over randomized stopping times. That this is also true in the anticipative case is somewhat hidden in the aforementioned articles, so we sketch the arguments for convenience of the reader:

Lemma 4.6.3. *Let $\nu\in\mathcal{P}(\mathcal{C})$, and let $(\omega,t)\mapsto\ell(\omega,t)$ be measurable, $\mathcal{F}$-optional, bounded or positive. Then*

$$\inf\left\{\mathbb{E}^\nu\left[\ell_\tau\right]:\ \tau \text{ a } \mathcal{G}\text{-stopping time on } \mathcal{C},\ \tau\le T\right\} \;=\; \inf_{\Sigma\in RST(\mathcal{G},\nu)} \mathbb{E}^\nu\left[\textstyle\int \ell_t d\Sigma_t\right]. \tag{4.6.1}$$

Furthermore, let $(\Omega,\mathcal{H},\mathbb{P})$ be a complete filtered probability space, $X:\Omega\to\mathcal{C}$ measurable and $\mathcal{H}$-adapted with $X_*\mathbb{P}=\nu$. Assuming that

$$\forall t\le T,\ X^{-1}(\mathcal{G}_T) \text{ is conditionally independent of } \mathcal{H}_t \text{ given } X^{-1}(\mathcal{G}_t), \tag{4.6.2}$$

we further have that the common value in (4.6.1) equals

$$\inf\left\{\mathbb{E}^{\mathbb{P}}\left[\ell(X,\tau)\right]:\ \tau \text{ a } \mathcal{H}\vee X^{-1}(\mathcal{G})\text{-stopping time on } \Omega,\ \tau\le T\right\}. \tag{4.6.3}$$

Proof. We first prove (4.6.1), following [CT07, Proof of Lemma 9]. Evidently the r.h.s. in (4.6.1) is the lesser one. For the converse inequality, take $\Sigma\in RST(\mathcal{G},\nu)$ and define

$$(\omega,x)\in\mathcal{C}\times[0,1]\mapsto\tau(\omega,x):=\inf\{t\in[0,T]:\Sigma_t>x\},$$

so by [DM80, Ch. VI.55] we have

$$\mathbb{E}^\nu\left[\textstyle\int\ell_t d\Sigma_t\right]=\int\int_0^T\ell(\omega,t)d\Sigma_t(\omega)\nu(d\omega)=\int\int_0^1\ell(\omega,\tau(\omega,x))\,dx\,\nu(d\omega).$$

Observe that for x fixed and each t we have $\{\omega\in\mathcal{C}:\ \tau(\omega,x)>t\}=\{\omega\in\mathcal{C}:\ \Sigma_t\le x\}\in\mathcal{G}_t$, hence $\omega\mapsto\tau(\omega,x)$ is a $\mathcal{G}$-stopping time on $\mathcal{C}$. Applying Fubini-Tonelli theorem, we find

$$\mathbb{E}^\nu\left[\textstyle\int\ell_t d\Sigma_t\right]=\int_0^1\int\ell(\omega,\tau(\omega,x))\,\nu(d\omega)\,dx,$$

and since the integrand in the r.h.s. here is for each x larger than the l.h.s. of (4.6.1), this establishes the equality. As for (4.6.3), one follows the arguments in [LP90, Proposition 3.5], or more precisely their extension in [CT07, Lemma 17]. □

If ν above is Markov (resp. Wiener) and $\mathcal{F}=\mathcal{G}$, then (4.6.2) is equivalent to X being Markov (resp. Brownian motion) w.r.t. $\mathcal{H}$. This should convey the message that both Condition (4.6.2) and Problem (4.6.3) are natural in our more general context.

We now look at optimal stopping under $(\mathcal{F},\mu)$, which by the previous lemma equals

$$v^{\mathcal{F},\mu}:=\inf_{\Sigma\in RST(\mathcal{F},\mu)}\mathbb{E}^\mu\left[\textstyle\int\ell_t d\Sigma_t\right]. \tag{$OS(\mathcal{F},\mu)$}$$

We want to compare this problem with the one where extra information (anticipation) is available and/or the law of the process to be stopped is different, namely (again by Lemma 4.6.3)

$$v^{\mathcal{G},\nu} := \inf_{\Sigma \in RST(\mathcal{G},\nu)} \mathbb{E}^{\nu}\left[\int \ell_t d\Sigma_t\right]. \tag{$OS(\mathcal{G},\nu)$}$$

The comparison of $v^{\mathcal{F},\mu}$ with $v^{\mathcal{G},\mu}$ corresponds to assessing the cost of information/anticipation. On the other hand, the comparison of $v^{\mathcal{F},\mu}$ with $v^{\mathcal{F},\nu}$ corresponds to the study of the dependence of non-anticipating optimal stopping with respect to different reference measures, or equivalently, with respect to different processes. We have:

Proposition 4.6.4. *Assume that $v^{\mathcal{F},\mu}$ and $v^{\mathcal{G},\nu}$ are both finite, and that the cost function $\ell : \mathcal{C} \times \mathbb{R}_+$ is optional and K-Lipschitz in its first argument with respect to a metric d on $\mathcal{C}$, uniformly in time (i.e. in the second argument). Then we have*

$$v^{\mathcal{F},\mu} - v^{\mathcal{G},\nu} \leq K \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})]. \tag{4.6.4}$$

In particular, in the two special cases of interest we have:

(i) If $\mu = \nu$, then

$$0 \leq v^{\mathcal{F},\mu} - v^{\mathcal{G},\mu} \leq K \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\mu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})]; \tag{4.6.5}$$

(ii) If $\mathcal{F} = \mathcal{G}$, then

$$\left|v^{\mathcal{F},\mu} - v^{\mathcal{F},\nu}\right| \leq K \inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})], \tag{4.6.6}$$

where the constraint in the transport problem in the right-hand side of (4.6.6) *means that both $\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ and $\tilde{\pi} \in \Pi^{\mathcal{F},\mathcal{F}}(\nu,\mu)$, where $\tilde{\pi} = ((\omega,\bar{\omega}) \mapsto (\bar{\omega},\omega))_*\pi$.*

Proof. Take an optimizer Σ for $(OS(\mathcal{G},\nu))$ (same argument holds for an optimizing sequence). We write this in the $\bar{\omega}$-variable and consider any causal transport π between μ and ν. Let $\tilde{\Sigma} \in RST(\mathcal{F}\,\mu)$ be the randomized stopping time associated to Σ, as in Lemma 4.6.2. We have

$$v^{\mathcal{F},\mu} \leq \mathbb{E}^{\mu}\left[\int_0^T l(\omega,t) d\tilde{\Sigma}_t(\omega)\right] = \mathbb{E}^{\pi}\left[\int_0^T l(\omega,t) d\Sigma_t(\bar{\omega})\right].$$

Hence the difference can be bounded above as follows

$$v^{\mathcal{F}\mu} - v^{\mathcal{G},\nu} \leq \mathbb{E}^{\pi}\left[\int_0^T [l(\omega,t) - l(\bar{\omega},t)] d\Sigma_t(\bar{\omega})\right] \leq KT\mathbb{E}^{\pi}[d(\omega,\bar{\omega})].$$

Being π a generic causal transport between the measures μ and ν, we get the bound in (4.6.4).

In the case (i), obviously $v^{\mathcal{F},\mu} \geq v^{\mathcal{G},\mu}$, since only the set of feasible optimization variables changes. As for the case (ii), exchanging the roles of μ and ν we get

$$\left|v^{\mathcal{F},\mu} - v^{\mathcal{F},\nu}\right| \leq K \max\left\{\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})], \inf_{\pi \in \Pi^{\mathcal{F},\mathcal{F}}(\nu,\mu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})]\right\} \leq K \inf_{\pi \in \Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)} \mathbb{E}^{\pi}[d(\omega,\bar{\omega})].$$

The last inequality follows since the cost d is symmetric (as a metric), implying that the r.h.s. can be computed on $\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\mu,\nu)$ or $\Pi_{bc}^{\mathcal{F},\mathcal{F}}(\nu,\mu)$ equivalently. □

Replacing the Lipschitz condition in Proposition 4.6.4 by uniform continuity, one obtains analogue results involving a modulus of continuity. Also observe that $\ell(x,t) = f(x_t)$ and $\ell(x,t) = f(\sup_{s \leq t} x_s)$ satisfy the assumptions of Proposition 4.6.4, with $d(\omega,\bar{\omega}) =$

$\|\omega - \bar{\omega}\|_\infty$, if f is Lipschitz. In this case the bound in (4.6.4) can be further majorized up to a multiplicative constant by

$$\inf_{\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)} \mathbb{E}^\pi[V_T(\bar{\omega} - \omega)].$$

4.6.2 Utility maximization

This part follows in spirit the previous section. We want to compare the optimal value of expected utility from terminal wealth, over a fixed finite time horizon $[0, T]$, when the reference filtration is enlarged by anticipation of information in the sense of Section 4.1. A wide literature is devoted to the utility maximization problem, in complete or incomplete markets, and with or without additional constraints; see [XS92, KLSX91, KLS87, CK92] among the earliest articles on the subject. Pikovsky and Karatzas [PK96] were the first ones to include anticipation of information. In a complete market, and for initial filtration enlargements, they provide the explicit value of this anticipation of information, in terms of log-utility maximization, with or without short-selling constraints; see also [AIS98, ADI06].

In this section we consider possibly incomplete markets, and any kind of anticipation of information (not just initial), and give an estimate of the value of information in terms of utility maximization under short-selling constraints, for a class of utility functions which includes the logarithm among other well-known ones; see Assumption 4.6.6. In order to do this, we set $\mathcal{X} = \mathcal{Y} = \mathcal{C}^d = \mathcal{C}([0, T], \mathbb{R}^d)$, which is the space of continuous $\mathbb{R}^d$-valued functions on the interval $[0, T]$, and keep the notation $\omega, \bar{\omega}$ for the coordinate processes. Let $(\Omega, \mathbb{P})$ be a probability space, equipped with $\mathcal{F}^B$, the natural filtration of a d-dimensional Brownian motion $B = (B_1, \ldots, B_d)^*$, i.e. $\mathcal{F}^B = \sigma(B_1, \ldots, B_d)$, augmented so as to satisfy the usual conditions. We use notation analogous to that of Section 4.1, and denote by $\mathcal{F}^{B,\mathfrak{g}}$ the enlargement of the filtration $\mathcal{F}^B$ with the process $(\mathfrak{g}_t(B))_t$, where $\mathfrak{g}_t : \mathcal{C}^d \to \mathcal{S}$. Hence $\mathcal{F}^{B,\mathfrak{g}}$ represents the information available to the informed agent. Throughout we assume:

Assumption 4.6.5. *The process B remains a semimartingale with respect to $\mathcal{F}^{B,\mathfrak{g}}$, say with semimartingale decomposition $B = \widetilde{B} + A$.*

We consider a financial market consisting of a riskless asset (bond), which we normalize to 1, and $m \le d$ risky assets whose price dynamics are described by the stochastic equations

$$dS_t^i = S_t^i\big(b_t^i\, dt + \sum_{j=1}^d \sigma_t^{i,j} dB_t^j\big), \quad i = 1, \ldots, m,$$

with initial condition $S_0^i = s_0^i > 0$. The vector process $b = (b_t^1, \ldots, b_t^m)^*$ of mean rates of return is assumed to be $\mathcal{F}^B$-progressively measurable and L-Lipschitz uniformly in time, i.e.

$$|b_t^i(\omega^1) - b_t^i(\omega^2)| \le L \sum_{k=1}^d \sup_{0\le s\le t} |\omega_s^{1,k} - \omega_s^{2,k}| \quad \forall\, t, i \text{ and } \forall\, \omega^1, \omega^2 \in \mathcal{C}^d. \qquad (4.6.7)$$

The $m \times d$ volatility matrix $\sigma_t = (\sigma_t^{i,j})_{1\le i\le m, 1\le j\le d}$ has full rank, it is $\mathcal{F}^B$-progressively measurable and M-Lipschitz uniformly in time, i.e.

$$|\sigma_t^{i,j}(\omega^1) - \sigma_t^{i,j}(\omega^2)| \leq M \sum_{k=1}^{d} \sup_{0\leq s\leq t} |\omega_s^{1,k} - \omega_s^{2,k}| \quad \forall\, t, i, j \text{ and } \forall\, \omega^1, \omega^2 \in \mathcal{C}^d, \tag{4.6.8}$$

and there exists some constant C s.t. $|\sigma_t^{i,j}| \leq C$ for each time t and for any i, j . We denote by λ_t^i the proportion of an agent's wealth invested in the ith stock at time t ($1 \leq i \leq m$), the remaining proportion $1 - \sum_{i=1}^m \lambda_t^i$ being invested in the bond. We shall forbid short-selling of stocks and bond, which corresponds to the constraint $\lambda_i \in [0,1]$ for all i, and $\sum_{i=1}^m \lambda_t^i \leq 1$. We write $\lambda \in \mathbb{A}$ for this constraint; the case of arbitrary compact-convex constraints can be treated in the same way. Let $\mathcal{A}(\mathcal{F}^{B,\mathfrak{g}})$ and $\mathcal{A}(\mathcal{F}^B)$ be the sets of *admissible portfolios* for the agent with and without anticipative information, i.e. the sets of $\mathcal{F}^{B,\mathfrak{g}}$-, respectively $\mathcal{F}^B$-progressively measurable $\mathbb{A}$-valued processes $(\lambda_t)_{t\in[0,T]}$. Denoting by X^λ the wealth process corresponding to a portfolio λ and starting from a unit of capital, we have $dX_t^\lambda = X_t^\lambda \lambda_t^*[b_t dt + \sigma_t dB_t]$, that is,

$$X_t^\lambda = \exp\left(\int_0^t (\lambda_s^* b_s - \tfrac{1}{2}||\sigma_s^* \lambda_s||^2) ds + \int_0^t \lambda_s^* \sigma_s dB_s\right).$$

The above expression makes sense for portfolios in $\mathcal{A}(\mathcal{F}^{B,\mathfrak{g}})$ by Assumption 4.6.5. We also need:

Assumption 4.6.6. *The utility function $U : \mathbb{R}_+ \to \mathbb{R}$ is concave, increasing, and such that, for some $K \in \mathbb{R}_+$, we have $F := U \circ exp$ is K-Lipschitz, concave and increasing.*

We remark that this assumption is fulfilled e.g. by U negative power utility $U(x) = \frac{x^a}{a}$ for $a \leq 0$, or logarithmic utility $U(x) = ln(x)$, or exponential utility $U(x) = -\frac{1}{a}e^{-ax}$ for $a \geq 1$. The function F is 1-Lipschitz for the first two examples, and e^{-a}-Lipschitz for the last one.

The utility maximization problem without anticipation of information is then given by

$$v^B = \sup_{\lambda\in\mathcal{A}(\mathcal{F}^B)} \mathbb{E}[U(X_T^\lambda)]. \tag{$U(B)$}$$

We proceed to compare this value with the following problem under anticipation of information:

$$v^{B,\mathfrak{g}} = \sup_{\lambda\in\mathcal{A}(\mathcal{F}^{B,\mathfrak{g}})} \mathbb{E}[U(X_T^\lambda)], \tag{$U(B,\mathfrak{g})$}$$

hence obtaining a bound on the price of information relative to the risk-attitude encoded by U.

Proposition 4.6.7. *The difference between the value functions of informed and uninformed agents can be bounded as follows (with the convention $+\infty - \infty = 0$)*

$$0 \leq v^{B,\mathfrak{g}} - v^B \leq \tilde{K} \inf_{\pi\in\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\gamma)} \mathbb{E}^\pi[V_T(\bar{\omega} - \omega)], \tag{4.6.9}$$

for some explicit constant $\tilde{K}$, see (4.6.13).

We recall that by the total variation of an $\mathbb{R}^d$-valued process X we mean the sum of total variations of its components, i.e. $V_T(X) = \sum_{i=1}^d V_T(X_i)$. Thanks to the multidimensional version of Theorem 4.4.2 (see Remark 4.4.3), we have that if the causal problem in (4.6.9) is finite, then its value equals $\mathbb{E}^\gamma\, [V_T(A)]$. The appeal of (4.6.9) is that in principle one need not know the specific form of the process A.

Proof. In the case $v^B = \infty$ we do not have anything to prove, hence we assume v^B to be finite. In the path space $\mathcal{C}^d$, the expected utility from terminal wealth for the agents with and without anticipative information is given by

$$\mathbb{E}^\gamma\left[F\left(\int_0^T(\lambda_t^* b_t - \tfrac{1}{2}||\sigma_t^*\lambda_t||^2)dt + \int_0^T \lambda_t^*\sigma_t d\omega_t,\right)\right]$$

where λ is $\mathcal{G}$- and $\mathcal{F}$-progressively measurable, respectively.

We now fix a causal transport $\pi \in \Pi^{\mathcal{F},\mathcal{G}}(\gamma,\gamma) \subset \mathcal{P}(C^d \times \mathcal{C}^d)$ for which the total variation $V_T(\bar\omega - \omega)$ is π-a.s. finite, and consider $(U(B))$ to be solved in the ω variable and $(U(B,\mathfrak{g}))$ in the $\bar\omega$ variable. Assume $v^{B,\mathfrak{g}} < \infty$ and that there is an optimizer $\hat\lambda(\bar\omega) = (\hat\lambda^1(\bar\omega),\dots,\hat\lambda^m(\bar\omega))$ for problem $(U(B,\mathfrak{g}))$ (else one can argue in the same way for every element λ^n of a sequence such that $\mathbb{E}[U(X_T^{\lambda^n})] \to v^{B,\mathfrak{g}}$ for $n \to \infty$). We denote by $\widetilde\lambda = (\widetilde\lambda^1,\dots,\widetilde\lambda^m)$ its optional projection with respect to $(\pi, \mathcal{F}\otimes\{\emptyset,\mathcal{C}\})$, so that in particular

$$\widetilde\lambda_t^i(\omega) = \widetilde\lambda_t^i(\omega,\bar\omega) = \mathbb{E}^\pi[\hat\lambda_t^i(\bar\omega)|\mathcal{F}_t] = \mathbb{E}^\pi[\hat\lambda_t^i(\bar\omega)|\mathcal{F}_T],\ i = 1\dots,m,$$

(for simplicity, here and in what follows, we use the notation $\mathbb{E}^\pi[.|\mathcal{F}_t]$ for $\mathbb{E}^\pi[.|\mathcal{F}_t\otimes\{\emptyset,\mathcal{C}\}]$). The last equality follows by causality and is crucial for the next argument. Moreover, note that $\widetilde\lambda \in \mathcal{A}(\mathcal{F}^B)$, which yields

$$\begin{aligned}
v^B &\geq \mathbb{E}^\gamma\Big[F\Big(\int_0^T(\widetilde\lambda_t^*(\omega)b_t(\omega) - \tfrac{1}{2}||\sigma_t^*(\omega)\widetilde\lambda_t(\omega)||^2)dt + \int_0^T \widetilde\lambda_t^*(\omega)\sigma_t(\omega)d\omega_t\Big)\Big]\\
&\geq \mathbb{E}^\pi\Big[F\Big(\mathbb{E}^\pi\Big[\int_0^T(\hat\lambda_t^*(\bar\omega)b_t(\omega) - \tfrac{1}{2}||\sigma_t^*(\omega)\hat\lambda_t(\bar\omega)||^2)dt + \int_0^T \hat\lambda_t^*(\bar\omega)\sigma_t(\omega)d\omega_t\Big|\mathcal{F}_T\Big]\Big)\Big]\\
&\geq \mathbb{E}^\pi\Big[F\Big(\int_0^T(\hat\lambda_t^*(\bar\omega)b_t(\omega) - \tfrac{1}{2}||\sigma_t^*(\omega)\hat\lambda_t(\bar\omega)||^2)dt + \int_0^T \hat\lambda_t^*(\bar\omega)\sigma_t(\omega)d\omega_t\Big)\Big],
\end{aligned}$$

by Jensen's inequality, being F concave and increasing, and by causality. Therefore, by Lipschitz continuity of the function F, we have

$$\begin{aligned}
&0 \leq v^{B,\mathfrak{g}} - v^B\\
&\leq K\mathbb{E}^\pi\Big[\Big|\int_0^T \hat\lambda_t^*(\bar\omega)\,(b_t(\bar\omega) - b_t(\omega))\,dt - \tfrac{1}{2}\int_0^T\big(||\sigma_t^*(\bar\omega)\hat\lambda_t(\bar\omega)||^2 - ||\sigma_t^*(\omega)\hat\lambda_t(\bar\omega)||^2\big)dt\\
&+ \int_0^T \hat\lambda_t^*(\bar\omega)\,(\sigma_t(\bar\omega)d\bar\omega_t - \sigma_t(\omega)d\omega_t)\Big|\Big]\\
&\leq K\mathbb{E}^\pi\Big[\int_0^T|\hat\lambda_t^*(\bar\omega)(b_t(\bar\omega) - b_t(\omega))|dt + \Big|\int_0^T \hat\lambda_t^*(\bar\omega)\,(\sigma_t(\bar\omega)d\bar\omega_t - \sigma_t(\omega)d\omega_t)\Big|\\
&+ \tfrac{1}{2}\int_0^T\sum_{j=1}^d\Big[\sum_{i=1}^m|\hat\lambda_t^i(\bar\omega)|\,|\sigma_t^{i,j}(\bar\omega) - \sigma_t^{i,j}(\omega)|\Big]\Big[\sum_{i=1}^m|\hat\lambda_t^i(\bar\omega)|\,|\sigma_t^{i,j}(\bar\omega) + \sigma_t^{i,j}(\omega)|\Big]dt\Big]\\
&\leq K\mathbb{E}^\pi\Big[\int_0^T\sum_{i=1}^m|b_t^i(\bar\omega) - b_t^i(\omega)|\,dt + \Big|\int_0^T \hat\lambda_t^*(\bar\omega)\sigma_t(\bar\omega)(d\bar\omega_t - d\omega_t)\Big| \qquad (4.6.10)\\
&+ \Big|\int_0^T \hat\lambda_t^*(\bar\omega)(\sigma_t(\bar\omega) - \sigma_t(\omega))d\omega_t\Big| \qquad (4.6.11)\\
&+ \tfrac{1}{2}\int_0^T\sum_{j=1}^d\Big[\sum_{i=1}^m|\sigma_t^{i,j}(\bar\omega) - \sigma_t^{i,j}(\omega)|\Big]\Big[\sum_{i=1}^m|\sigma_t^{i,j}(\bar\omega) + \sigma_t^{i,j}(\omega)|\Big]\,dt\Big]. \qquad (4.6.12)
\end{aligned}$$

Now, by (4.6.7) and since $\bar\omega_0 - \omega_0 = 0$ on $\mathcal{C}^d$, we have that π-a.s.

$$\int_0^T\sum_{i=1}^m|b_t^i(\bar\omega) - b_t^i(\omega)|\,dt \leq LTm\sum_{k=1}^d\sup_{0\leq t\leq T}\left|\bar\omega_t^k - \omega_t^k\right| \leq LTm\sum_{k=1}^d V_T(\bar\omega^k - \omega^k) = LTmV_T(\bar\omega - \omega).$$

The second term in (4.6.10) is easily bounded π-a.s. as follows

$$\Big|\int_0^T \hat\lambda_t^*(\bar\omega)\sigma_t(\bar\omega)(d\bar\omega_t - d\omega_t)\Big| = \sum_{j=1}^d\sum_{i=1}^m\Big|\int_0^T \hat\lambda_t^i(\bar\omega)\sigma_t^{i,j}(\bar\omega)(d\bar\omega_t^j - d\omega_t^j)\Big| \leq CmV_T(\bar\omega - \omega).$$

To get the upper bound for (4.6.11), we proceed as follows. We denote $X_t = \int_0^t\hat\lambda_s^*(\bar\omega)\,(\sigma_s(\bar\omega) - \sigma_s(\omega))\,d\omega_s$. This is a martingale under π, because by causality the coordinate process ω is an $\mathcal{F}\otimes\mathcal{G}$-martingale. Hence, we can apply the Burkholder-

Davis-Gundy inequality, obtaining

$$\mathbb{E}^\pi[|X_T|] \le C_1\mathbb{E}^\pi\Big[\sqrt{\langle X,X\rangle_T}\Big] = C_1\mathbb{E}^\pi\Big[\sqrt{\int_0^T \sum_{j=1}^d \Big|\sum_{i=1}^m \hat\lambda_t^i(\bar\omega)(\sigma_t^{i,j}(\bar\omega)-\sigma_t^{i,j}(\omega))\Big|^2 dt}\Big]$$

$$\le C_1\sqrt{m}\,\mathbb{E}^\pi\Big[\sqrt{\int_0^T\sum_{j=1}^d\sum_{i=1}^m|\sigma_t^{i,j}(\bar\omega)-\sigma_t^{i,j}(\omega)|^2dt}\Big] \le C_1\,M\sqrt{m\,d}\;\mathbb{E}^\pi\Big[\sqrt{\int_0^T\Big\{\sum_{k=1}^d\sup_{0\le s\le t}|\bar\omega_s^k-\omega_s^k|\Big\}^2}\Big]$$

$$\le C_1\,M\sqrt{T\,m\,d}\;\mathbb{E}^\pi\Big[\sqrt{\Big\{\sum_{k=1}^d\sup_{0\le s\le T}|\bar\omega_s^k-\omega_s^k|\Big\}^2dt}\Big] \le\; C_1\,M\sqrt{T\;m\,d}\,\mathbb{E}^\pi\left[V_T(\bar\omega-\omega)\right].$$

Hence, the difference $v^{B,\mathfrak{g}}-v^B$ is bounded above by $\tilde K\mathbb{E}^\pi[V_T(\bar\omega-\omega)]$, with

$$\tilde K = K(LTm+m^2dCMT+Cm+C_1M\sqrt{Tdm}). \tag{4.6.13}$$

Since π was a generic transport in $\Pi^{\mathcal{F},\mathcal{G}}(\gamma,\gamma)$, this concludes the proof for the case $v^{B,\mathfrak{g}}<\infty$. The case $v^{B,\mathfrak{g}}=\infty$ follows similarly, working along a sequence λ^n s.t. $\mathbb{E}[U(X_T^{\lambda^n})]\to\infty$. $\square$

List of frequently used notation

(OT)	p.20	Optimal transport problem
(COT)	p.22	Optimal transport problem with linear constraints
(Pc)	p.29,36	Causal optimal transport problem
(Pbc)	p.30,45	Bicausal optimal transport problem
(Pcqm)	p.38	Causal quasi-Markov optimal transport problem
$\Pi(\mu,\nu)$	p.19	Set of admissible transport plans between μ and ν
$\Pi_{\mathbb{F}}(\mu,\nu)$	p.22	Set of constrained transport plans between μ and ν
$\Pi^{\mathcal{F},\mathcal{G}}(\mu,\nu)$	p.28	Set of causal transport plans between μ and ν
$\Pi^{\mathcal{F},\mathcal{G}}(\mu,\cdot)$	p.33	Set of causal transport plans with the first marginal μ
$\Pi^{\mathcal{F},\mathcal{G}}_{bc}(\mu,\nu)$	p.28	Set of bicausal transport plans
$\Pi^{\mathcal{F},\mathcal{G}}_{cqm}(\mu,\nu)$	p.38	Set of causal quasi-Markov transport plans
$\mathcal{F}^Z$	p.28	Filtration generated by a process Z
$\mathcal{F}^\mu$	p.96	μ-completion of filtration $\mathcal{F}$
$F_\eta(\cdot)$	p.37	Cumulative distribution function of a probability measure η
$F_\eta^{-1}(u)$	p.37	Left-continuous generalized inverse of cumulative distribution function F_η
$p^1_*\mu$	p.35	First marginal of the measure μ
$\mathrm{Ent}(\nu\vert\mu)$	p.67	Relative entropy of ν w.r.t. μ
$\mathcal{W}_p^p(\mu,\nu)$	p.67	$p-$Wasserstein distance
$\mathcal{W}_{c,2}^{2,N}(\gamma^N,\nu^N)$	p.85	Time-discretized transport problem on Wiener space with Cameron-Martin cost
G^N	p.68	$N-$dimensional standard Gaussian measure
T_2	p.68	Talagrand's inequality for the Gaussian measure
CT_2	p.70	Bicausal transport-entropy inequality
$\langle\!\cdot\omega_t\cdot\!\rangle$	p.76	Time derivative of the path ω
c^h	p.77	Cameron-Martin type cost function
$c^{N,h}$	p.83	Time discretization of the Cameron-Martin type cost function
c^N	p.82	Time discretization of the cost function c
$\mathcal{C}[0,T]$	p.76	Space of continuous functions on the interval $[0,T]$
Δ	p.83	Difference operator
${}^o\Lambda^{\mathcal{F}}$	p.100	Optional projection of Λ
${}^{\mathcal{F}}\Lambda^o$	p.100	Dual optional projection of Λ
$V_t(Z)$	p.102	Total variation of a process Z up to time t

Σ	p.114	Randomized stopping time
ρ^*	p.108	Convex conjugate of function ρ
∂	p.109	Sub-differential
$\langle Z, M\rangle$	p.104	Covariation of processes Z and M
$\langle X, X\rangle$	p.104	Quadratic variation of a process X

Index

Bibliography

[AB06] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis*, third ed., Springer, Berlin, 2006, A hitchhiker's guide. MR 2378491 (2008m:46001)

[ABZ16] B. Acciaio, J. Backhoff, and A. Zalashko, *Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization*, Submitted, arXiv:1611.02610, 2016.

[ADI06] S. Ankirchner, S. Dereich, and P. Imkeller, *The Shannon information of filtrations and the additional logarithmic utility of insiders*, Ann. Probab. **34** (2006), no. 2, 743–778. MR 2223957

[ADI07] ———, *Enlargement of filtrations and continuous Girsanov-type embeddings*, Séminaire de Probabilités XL, Lecture Notes in Math., vol. 1899, Springer, Berlin, 2007, pp. 389–410. MR 2409018

[AIS98] J. Amendinger, P. Imkeller, and M. Schweizer, *Additional logarithmic utility of an insider*, Stochastic processes and their applications **75** (1998), no. 2, 263–286.

[Ald81] D. Aldous, *Weak convergence and the general theory of processes (weak convergence of stochastic processes for processes viewed in the strasbourg manner)*, unpublished, 1981.

[Bal] E.J. Balder, *Lectures on Young measures*, cahiers de mathematiques de la décision 9517 ed., CEREMADE-Université Paris-Dauphine.

[BBLZ16] J. Backhoff, M. Beiglböck, Y. Lin, and A. Zalashko, *Causal transport in discrete time and applications*, to appear in SIAM Journal on Optimization, 2016.

[BC77] J. R. Baxter and R. V. Chacon, *Compactness of stopping times*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **40** (1977), no. 3, 169–181. MR 0517871

[BCH16] M. Beiglböck, A. Cox, and M. Huesmann, *Optimal transport and skorokhod embedding*, To appear Inventiones Mathematicae, arXiv:1307.3656v4, 2016.

[BG14] M. Beiglböck and C. Griessler, *An optimality principle with applications in optimal transport and its offsprings*, Submitted, arXiv:1404.7054, 2014.

[BGMS09] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer, *Optimal and better transport plans*, J. Funct. Anal. **256** (2009), no. 6, 1907–1927. MR 2498564

[BHLP13] M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance Stoch. **17** (2013), no. 3, 477–501. MR 3066985

[Bil99] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication. MR 1700749 (2000e:60008)

[BJ12] M. Beiglböck and N. Juillet, *On a problem of optimal transport under marginal martingale constraints*, To appear in Ann. Probab., 2012.

[BKM05] V. I. Bogachev, A. V. Kolesnikov, and K. V. Medvedev, *Triangular transformations of measures*, Mat. Sb. **196** (2005), no. 3, 3–30. MR 2144274 (2006c:28012)

[BLS12] M. Beiglböck, C. Léonard, and W. Schachermayer, *A general duality theorem for the Monge-Kantorovich transport problem*, Studia Math. **209** (2012), no. 2, 151–167. MR 2943840

[BNT15] M. Beiglböck, M. Nutz, and N. Touzi, *Complete duality for martingale optimal transport on the line*, To appear Ann. Probab., arXiv:1507.00671, 2015.

[Bog07] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007. MR 2267655 (2008g:28002)

[Bre91] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417. MR 1100809

[BS78] D. P. Bertsekas and S. E. Shreve, *Stochastic optimal control*, Mathematics in Science and Engineering, vol. 139, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978, The discrete time case. MR 511544 (80d:93081)

[BS11] M. Beiglböck and W. Schachermayer, *Duality for Borel measurable cost functions*, Trans. Amer. Math. Soc. **363** (2011), no. 8, 4203–4224. MR 2792985 (2012k:49108)

[BSJR16] C. Blanchet-Scalliet, M. Jeanblanc, and R. R. Romero, *Enlargement of filtration in discrete time*, hal.archives-ouvertes.fr/hal-01253214, 2016.

[BY78] P. Brémaud and M. Yor, *Changes of filtrations and of probability measures*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **45** (1978), no. 4, 269–295.

[CGS10] G. Carlier, A. Galichon, and F. Santambrogio, *From Knothe's transport to Brenier's map and a continuation method for optimal transport*, SIAM J. Math. Anal. **41** (2009/10), no. 6, 2554–2576. MR 2607321

[CK92] J. Cvitanić and I. Karatzas, *Convex duality in constrained portfolio optimization*, Ann. Appl. Probab. **2** (1992), no. 4, 767–818. MR 1189418

[CT07] F. Coquet and S. Toldo, *Convergence of values in optimal stopping and convergence of optimal stopping times*, Electron. J. Probab. **12** (2007), no. 8, 207–228. MR 2299917

[DGW04] H. Djellout, A. Guillin, and L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, Ann. Probab. **32** (2004), no. 3B, 2702–2732. MR 2078555 (2005i:60031)

[DL82] D. C. Dowson and B. V. Landau, *The Fréchet distance between multivariate normal distributions*, J. Multivariate Anal. **12** (1982), no. 3, 450–455. MR 666017 (84a:62078)

[DM80] C. Dellacherie and P.-A. Meyer, *Probabilités et potentiel. Chapitres V à VIII*, revised ed., Actualités Scientifiques et Industrielles, vol. 1385, Hermann, Paris, 1980, Théorie des martingales. [Martingale theory]. MR 566768

[Föl86] H. Föllmer, *Time reversal on Wiener space*, Stochastic processes—mathematics and physics (Bielefeld, 1984), Lecture Notes in Math., vol. 1158, Springer, Berlin, 1986, pp. 119–129. MR 838561

[FÜ02] D. Feyel and A. S. Üstünel, *Measure transport on Wiener space and the Girsanov theorem*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 11, 1025–1028. MR 1913729

[FÜ04] D. Feyel and A. S. Üstünel, *Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space*, Probab. Theory Related Fields **128** (2004), no. 3, 347–385. MR 2036490 (2004m:60121)

[GL10] N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736. MR 2895086

[GRST15] N. Gozlan, C. Roberto, P-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, Preprint arXiv:1412.7480v4, 2015.

[Itô50] K. Itô, *Stochastic differential equations in a differentiable manifold*, Nagoya Math. J. **1** (1950), 35–47. MR 0038596

[Jac80] J. Jacod, *Weak and strong solutions of stochastic differential equations*, Stochastics **3** (1980), 171–191.

[Jac85] ———, *Grossissement initial, hypothèse (H'), et théorème de Girsanov*, Grossissements de Filtrations: Exemples et Applications (T. Jeulin and M. Yor, eds.), Lecture Notes in Mathematics, vol. 1118, Springer, Berlin - Heidelberg, 1985, pp. 15–35.

[Jeu80] T. Jeulin, *Semi-martingales et grossissement d'une filtration*, Lecture Notes in Mathematics, vol. 833, Springer, Berlin, 1980. MR MR604176 (82h:60106)

[JY78] T. Jeulin and M. Yor, *Grossissement d'une filtration et semi-martingales: formules explicites*, Séminaire de Probabilités, XII (Univ. Strasbourg, Strasbourg, 1976/1977), Lecture Notes in Math., vol. 649, Springer, Berlin, 1978, pp. 78–97. MR MR519998

[JYC09] M. Jeanblanc, M. Yor, and M. Chesney, *Mathematical methods for financial markets*, Springer Finance, Springer-Verlag London Ltd., London, 2009. MR 2568861

[Kal02] O. Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR 1876169 (2002m:60002)

[Kan42] L. V. Kantorovich, *On the transfer of masses*, Dokl. Akad. Nauk. SSSR, vol. 37, 1942, pp. 227–229.

[Kel84] H. G. Kellerer, *Duality theorems for marginal problems*, Z. Wahrsch. Verw. Gebiete **67** (1984), no. 4, 399–432. MR 761565

[KLS87] I. Karatzas, J. P. Lehoczky, and S. E. Shreve, *Optimal portfolio and consumption decisions for a "small investor" on a finite horizon*, SIAM J. Control Optim. **25** (1987), no. 6, 1557–1586. MR 912456

[KLSX91] I. Karatzas, J. P. Lehoczky, S. E. Shreve, and G.-L. Xu, *Martingale and duality methods for utility maximization in an incomplete market*, SIAM J. Control Optim. **29** (1991), no. 3, 702–730. MR 1089152

[Kno57] H. Knothe, *Contributions to the theory of convex bodies*, Michigan Math. J. **4** (1957), 39–52. MR 0083759

[KP15] Y. Kchia and P. Protter, *Progressive filtration expansions via a process, with applications to insider trading*, Int. J. Theor. Appl. Finance **18** (2015), no. 4, 1550027, 48. MR 3358108

[KS91] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, second ed., Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991. MR 1121940

[Kur07] T. G. Kurtz, *The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities*, Electron. J. Probab **12** (2007), 951–965.

[Kur14] ———, *Weak and strong solutions of general stochastic models*, Electron. Commun. Probab. **19** (2014), no. 58, 16. MR 3254737

[Las13] R. Lassalle, *Causal transference plans and their Monge-Kantorovich problems*, arXiv:1303.6925, 2013.

[Las15] ———, *Causal transference plans and their Monge-Kantorovich problems*, Submitted, arXiv:1303.6925.v2, 2015.

[Led01] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR 1849347

[Leh13] J. Lehec, *Representation formula for the entropy and functional inequalities*, Ann. Inst. Henri Poincaré Probab. Stat. **49** (2013), no. 3, 885–899. MR 3112438

[Léo12] C. Léonard, *Girsanov theory under a finite entropy condition*, Séminaire de Probabilités XLIV, Lecture Notes in Math., vol. 2046, Springer, Heidelberg, 2012, pp. 429–465. MR 2953359

[LP90] D. Lamberton and G. Pagès, *Sur l'approximation des réduites*, Ann. Inst. H. Poincaré Probab. Statist. **26** (1990), no. 2, 331–355. MR 1063754

[LV14] L. Leskella and M. Vihola, *Conditional convex orders and measurable martingale couplings*, arXiv:1404.0999v2, 2014.

[Mar86] K. Marton, *A simple proof of the blowing-up lemma*, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446. MR 838213

[Mar96] ———, *Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866. MR 1404531

[McC97] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), no. 1, 153–179. MR 1451422

[Mik12] T. Mikami, *A characterization of the Knothe-Rosenblatt processes by a convergence result*, SIAM J. Control Optim. **50** (2012), no. 4, 1903–1920. MR 2974723

[Mon84] G Monge, *Mémoire sur la théorie des déblais et des remblais, histoire de l'académie royale des sciences année 1781*, Avec les Memoires de Mathematique & de Physique, pour la même Annee](2e partie)(1784) Histoire (1784), 34–38.

[MY06] R. Mansuy and M. Yor, *Random times and enlargements of filtrations in a brownian setting*, Springer, 2006.

[Ok03] B. Oksendal, *Stochastic differential equations*, sixth ed., Universitext, Springer-Verlag, Berlin, 2003, An introduction with applications. MR 2001996

[Pfl09] G. Ch. Pflug, *Version-independence and nested distributions in multistage stochastic optimization*, SIAM Journal on Optimization **20** (2009), no. 3, 1406–1420.

[PK96] I. Pikovsky and I. Karatzas, *Anticipative portfolio optimization*, Adv. in Appl. Probab. **28** (1996), no. 4, 1095–1122. MR 1418248

[PP12] G. Ch. Pflug and A. Pichler, *A distance for multistage stochastic optimization models*, SIAM J. Optim. **22** (2012), no. 1, 1–23. MR 2902682

[PP14] ______, *Multistage stochastic optimization*, Springer Series in Operations Research and Financial Engineering, Springer, Cham, 2014. MR 3288310

[PP15] ______, *Dynamic generation of scenario trees*, Comput. Optim. Appl. **62** (2015), no. 3, 641–668. MR 3426120

[Pro05] P. E. Protter, *Stochastic integration and differential equations*, Stochastic Modelling and Applied Probability, vol. 21, Springer-Verlag, Berlin, 2005, Second edition. Version 2.1, Corrected third printing. MR 2273672

[Ros52] M. Rosenblatt, *Remarks on a multivariate transformation*, Ann. Math. Statistics **23** (1952), 470–472. MR 0049525

[RR91] M. Rao and Z. Ren, *Theory of Orlicz spaces*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 146, Marcel Dekker, Inc., New York, 1991. MR 1113700

[RR98] S. T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. II*, Probability and its Applications (New York), Springer-Verlag, New York, 1998, Applications. MR 1619171

[Rüs85] L. Rüschendorf, *The Wasserstein distance and approximation theorems*, Z. Wahrsch. Verw. Gebiete **70** (1985), no. 1, 117–129. MR 795791 (86m:60004)

[RW98] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317, Springer-Verlag, Berlin, 1998. MR 1491362

[RY99] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 293, Springer-Verlag, Berlin, 1999. MR 1725357

[San15] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015, Calculus of variations, PDEs, and modeling. MR 3409718

[Sio58] M. Sion, *On general minimax theorems*, Pacific J. Math. **8** (1958), 171–176. MR 0097026

[Tal96] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. MR 1392331 (97d:60029)

[Tsi75] B. S. Tsirelson, *An example of a stochastic differential equation that has no strong solution*, Teor. Verojatnost. i Primenen. **20** (1975), no. 2, 427–430. MR 0375461

[Var84] S. R. S. Varadhan, *Large deviations and applications*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984. MR 758258 (86h:60067b)

[Vil03] C. Villani, *Topics in optimal transportation*, no. 58, American Mathematical Soc., 2003.

[Vil08] ______, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.

[XS92] G.-L. Xu and S. E. Shreve, *A duality method for optimal consumption and investment under short-selling prohibition. I. General market coefficients*, Ann. Appl. Probab. **2** (1992), no. 1, 87–112. MR 1143394

[Yor85] M. Yor, *Entropie d'une partition, et grossissement initial d'une filtration*, Grossissements de filtrations: exemples et applications, Springer, 1985, pp. 45–58.

[Yor12] ______, *Some aspects of brownian motion: Part ii: Some recent martingale problems*, Birkhäuser, 2012.

[YW71] T. Yamada and S. Watanabe, *On the uniqueness of solutions of stochastic differential equations*, Journal of Mathematics of Kyoto University **11** (1971), no. 1, 155–167.

[Zae15] D. Zaev, *On the Monge–Kantorovich Problem with Additional Linear Constraints*, Mat. Zametki **98** (2015), no. 5, 664–683. MR 3438523