Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 130 (2020) 1568–1615

stochastic processes and their applications

www.elsevier.com/locate/spa

Multiperiod martingale transport

Marcel Nutz[a,b,*,1], Florian Stebegg[a], Xiaowei Tan[b]

[a] *Department of Statistics, Columbia University, United States*
[b] *Department of Mathematics, Columbia University, United States*

Received 25 June 2017; received in revised form 26 January 2019; accepted 16 May 2019
Available online 1 June 2019

Abstract

Consider a multiperiod optimal transport problem where distributions $\mu_0, \dots, \mu_n$ are prescribed and a transport corresponds to a scalar martingale X with marginals $X_t \sim \mu_t$. We introduce particular couplings called left-monotone transports; they are characterized equivalently by a no-crossing property of their support, as simultaneous optimizers for a class of bivariate transport cost functions with a Spence–Mirrlees property, and by an order-theoretic minimality property. Left-monotone transports are unique if μ_0 is atomless, but not in general. In the one-period case $n = 1$, these transports reduce to the Left-Curtain coupling of Beiglböck and Juillet. In the multiperiod case, the bivariate marginals for dates $(0, t)$ are of Left-Curtain type, if and only if $\mu_0, \dots, \mu_n$ have a specific order property. The general analysis of the transport problem also gives rise to a strong duality result and a description of its polar sets. Finally, we study a variant where the intermediate marginals $\mu_1, \dots, \mu_{n-1}$ are not prescribed.

MSC: 60G42; 49N05

Keywords: Optimal transport; Martingale coupling; Duality

1. Introduction

Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be a vector of probability measures μ_t on the real line. A measure P on $\mathbb{R}^{n+1}$ whose marginals are given by $\boldsymbol{\mu}$ is called a coupling (or transport) of $\boldsymbol{\mu}$, and the set of all such measures is denoted by $\Pi(\boldsymbol{\mu})$. We shall be interested in couplings P that are martingales; that is, the identity $X = (X_0, \dots, X_n)$ on $\mathbb{R}^{n+1}$ is a martingale under P. Hence,

* Correspondence to: 1255 Amsterdam Avenue, New York, NY 10027, USA.
E-mail addresses: mnutz@columbia.edu (M. Nutz), florian.stebegg@columbia.edu (F. Stebegg), xt2161@columbia.edu (X. Tan).

[1] Research supported by an Alfred P. Sloan Fellowship and NSF Grants DMS-1512900 and DMS-1812661.

https://doi.org/10.1016/j.spa.2019.05.010

we will assume that all marginals have a finite first moment and denote by $\mathcal{M}(\boldsymbol{\mu})$ the set of martingale couplings. A classical result of Strassen [41] shows that $\mathcal{M}(\boldsymbol{\mu})$ is nonempty if and only if the marginals are in convex order, denoted by $\mu_{t-1} \leq_c \mu_t$ and defined by the requirement that $\mu_{t-1}(\phi) \leq \mu_t(\phi)$ for any convex function ϕ, where $\mu(\phi) := \int \phi \, d\mu$.

The first goal of this paper is to introduce and study a family of "canonical" couplings $P \in \mathcal{M}(\mu)$ that we call left-monotone. These couplings specialize to the Left-Curtain coupling of [8] in the one-step case $n = 1$ and share, broadly speaking, several properties reminiscent of the Hoeffding–Fréchet coupling of classical optimal transport. Indeed, left-monotone couplings will be characterized by order-theoretic minimality properties, as simultaneous optimal transports for certain classes of reward (or cost) functions, and through no-crossing conditions on their supports.

The second goal is to develop a strong duality theory for multiperiod martingale optimal transport, along the lines of [10] for the one-period martingale case and [35] for the classical optimal transport problem. That is, we introduce a suitable dual optimization problem and show the absence of a duality gap as well as the existence of dual optimizers for general transport reward (or cost) functions. The duality result is a crucial tool for the study of the left-monotone couplings.

We also develop similar results for a variant of our problem where the intermediate marginals $\mu_1, \dots, \mu_{n-1}$ are not prescribed (Section 9), but we shall focus on the full marginal case for the purpose of the Introduction.

1.1. Left-monotone transports

For the sake of orientation, let us first state the main result and then explain the terminology contained therein. The following is a streamlined version—the results in the body of the paper are stronger in some technical aspects.

Theorem 1.1. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order and $P \in \mathcal{M}(\boldsymbol{\mu})$ a martingale transport between these marginals. The following are equivalent:*

(i) P is a simultaneous optimal transport for $f(X_0, X_t)$, $1 \leq t \leq n$ whenever $f : \mathbb{R}^2 \to \mathbb{R}$ is a smooth second-order Spence–Mirrlees function.
(ii) P is concentrated on a left-monotone set $\Gamma \subseteq \mathbb{R}^{n+1}$.
(iii) P transports $\mu_0|_{(-\infty,a]}$ to the obstructed shadow $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{(-\infty,a]})$ in step t, for all $1 \leq t \leq n$ and $a \in \mathbb{R}$.

There exists $P \in \mathcal{M}(\boldsymbol{\mu})$ satisfying (i)–(iii), and any such P is called a left-monotone transport. If μ_0 is atomless, then P is unique.

Let us now discuss the items in the theorem.

(i) Optimal transport. This property characterizes P as a simultaneous optimal transport. Given a function $f : \mathbb{R}^{n+1} \to \mathbb{R}$, we may consider the martingale optimal transport problem with reward f (or cost $-f$),

$$\mathbf{S}_{\boldsymbol{\mu}}(f) = \sup_{P \in \mathcal{M}(\boldsymbol{\mu})} P(f); \tag{1.1}$$

recall that $P(f) = \mathbb{E}^P[f(X_0, \dots, X_n)]$. A Lipschitz function $f \in C^{1,2}(\mathbb{R}^2; \mathbb{R})$ is called a *smooth second-order Spence–Mirrlees* function if it satisfies the cross-derivative condition

$f_{xyy} > 0$; this has also been called the martingale Spence–Mirrlees condition in analogy to the classical Spence–Mirrlees condition $f_{xy} > 0$. Given such a function of two variables and $1 \le t \le n$, we may consider the n-step martingale optimal transport problem with reward $f(X_0, X_t)$. Characterization (i) states that a left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ is an optimizer simultaneously for the n transport problems $f(X_0, X_t)$, $1 \le t \le n$, for some (and then all) smooth second-order Spence–Mirrlees functions f.

In the one-step case, a corresponding result holds for the Left-Curtain coupling [8]; here the simultaneous optimization becomes a single one. In view of the characterization in (i), an immediate consequence is that if there exists $P \in \mathcal{M}(\boldsymbol{\mu})$ such that all bivariate projections $P_{0t} = P \circ (X_0, X_t)^{-1} \in \mathcal{M}(\mu_0, \mu_t)$ are of Left-Curtain type, then P is left-monotone. However, such a transport does not exist unless the marginals satisfy a very specific condition (see Proposition 6.9), and in general the bivariate projections of a left-monotone transport are *not* of Left-Curtain type.

(ii) Geometry. The second item characterizes P through a geometric property of its support. A set $\Gamma \subseteq \mathbb{R}^{n+1}$ will be called *left-monotone* if it has the following no-crossing property for all $1 \le t \le n$: Let $\boldsymbol{x} = (x_0, \dots, x_{t-1})$, $\boldsymbol{x}' = (x'_0, \dots, x'_{t-1}) \in \mathbb{R}^t$ and

$$y^-, y^+, y' \in \mathbb{R} \quad \text{with} \quad y^- < y^+$$

be such that $(\boldsymbol{x}, y^+), (\boldsymbol{x}, y^-), (\boldsymbol{x}', y')$ are in the projection of Γ to the first $t+1$ coordinates. Then,

$$y' \notin (y^-, y^+) \quad \text{whenever} \quad x_0 < x'_0.$$

That is, if we consider two paths in Γ starting at x_0 and coinciding up to $t-1$, and a third path starting at x'_0 to the right of x_0, then at time t the third path cannot step in-between the first two—this is illustrated in Fig. 1. Item (ii) states that a left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ can be characterized by the fact that it is concentrated on a left-monotone set Γ. (In Theorem 7.16 we shall state a stronger result: we can find a left-monotone set that carries all left-monotone transports at once.)

In the one-step case $n = 1$, left-monotonicity coincides with the Left-Curtain property of [8]. However, we emphasize that for $t > 1$, our no-crossing condition differs from the Left-Curtain property of the bivariate projection $(X_0, X_t)(\Gamma)$ as the latter would not contain the restriction that the first two paths have to coincide up to $t-1$ (see also Example 6.10). This corresponds to the mentioned fact that the bivariate marginal P_{0t} need not be of Left-Curtain type. On the other hand, the geometry of the projection $(X_{t-1}, X_t)(\Gamma)$ is also quite different from the Left-Curtain one, as our condition may rule out third paths crossing from the right *and* left at $t-1$, depending on the starting point x'_0 rather than the location of x'_{t-1}.

(iii) Convex ordering. This property characterizes left-monotone transports in an order-theoretic way and will be used in the existence proof. To explain the idea, suppose that μ_0 consists of finitely many atoms at $x_1, \dots, x_N \in \mathbb{R}$. Then, for any fixed t, a coupling of μ_0 and μ_t can be defined by specifying a "destination" measure for each atom. We consider all chains[2] $\mu_0|_{x_i} \le_c \theta_1 \le_c \cdots \le_c \theta_t$ of measures θ_s in convex order that satisfy the marginals constraints $\theta_s \le \mu_s$ for $s \le t$. Of these chains, keep only the terminal measures θ_t and compare them according to the convex order. The *obstructed shadow of* $\mu_0|_{x_1}$ *in* μ_t *through* $\mu_1, \dots, \mu_{t-1}$, denoted $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{x_i})$, is defined as the unique least element[3] among the θ_t. A particular

[2] Here $\mu_0|_{x_i}$ denotes a Dirac measure of mass $\mu_0(\{x_i\})$ at x_i.

[3] See Definition 6.6 and Lemma 6.7 for details on this construction.

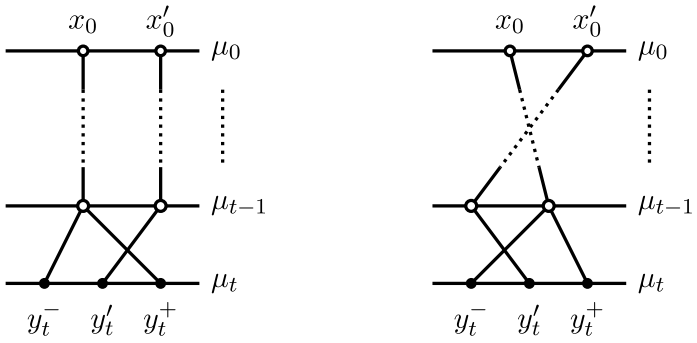

Fig. 1. Two examples of forbidden configurations in left-monotone sets.

coupling of μ_0 and μ_t is the one that successively maps the atoms $\mu_0|_{x_i}$ to their obstructed shadows in the remainder of μ_t, starting with the left-most atom x_i and continuing from left to right. In the case of general measures, we consider the restrictions $\mu_0|_{(-\infty,a]}$ instead of successively mapping the atoms. Characterization (iii) then states that a left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ maps $\mu_0|_{(-\infty,a]}$ to its obstructed shadow at date t for all $1 \le t \le n$ and $a \in \mathbb{R}$. This shows in particular that the bivariate projections $P_{0t} = P \circ (X_0, X_t)^{-1}$ of a left-monotone coupling are uniquely determined. In the body of the text, we shall also give an alternative definition of the obstructed shadow by iterating unobstructed shadows through the marginals up to date t; see Section 6.

The above specializes to the construction of [8] for the one-step case, which corresponds to the situation of $t = 1$ where there are no intermediate marginals obstructing the shadow. When $t > 1$, the obstruction by the intermediate marginals once again entails that P_{0t} need not be of Left-Curtain type. More precisely, Characterization (iii) gives rise to a sharp criterion (Proposition 6.9) on the marginals $\boldsymbol{\mu}$, describing exactly when this coincidence arises.

(Non-)uniqueness. We have seen above that for a left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ the bivariate projections P_{0t}, $1 \le t \le n$ are uniquely determined. In particular, for $n = 1$, we recover the result of [8] that the left-monotone coupling is unique. For $n > 1$, the situation turns out to be quite different depending on the nature of the first marginal. On the one extreme, we shall see that when μ_0 is atomless, there is a unique left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$. Moreover, P has a degenerate structure reminiscent of Brenier's theorem: it can be disintegrated as $P = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$ where each one-step transport kernel κ_t is concentrated on the graphs of two functions. On the other extreme, if μ_0 is a Dirac mass, the typical case is that there are infinitely many left-monotone couplings—see Section 8 for a detailed discussion. We shall also show that left-monotone transports are not Markovian in general, even if uniqueness holds (Example 7.17).

1.2. Duality

The analysis of left-monotone transports is based on a duality result that we develop for general reward functions $f : \mathbb{R}^{n+1} \to (-\infty, \infty]$ with an integrable lower bound. Formally, the dual problem (in the sense of linear programming) for the transport problem $\mathbf{S}_{\boldsymbol{\mu}}(f) = \sup_{P \in \mathcal{M}(\boldsymbol{\mu})} P(f)$ is the minimization

$$\mathbf{I}_{\boldsymbol{\mu}}(f) := \inf_{(\boldsymbol{\phi}, H)} \sum_{t=0}^{n} \mu_t(\phi_t)$$

where the infimum is taken over vectors $\boldsymbol{\phi} = (\phi_0, \dots, \phi_n)$ of real functions and predictable processes $H = (H_1, \dots, H_n)$ such that

$$\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n \geq f; \tag{1.2}$$

here $(H \cdot X)_n := \sum_{t=1}^{n} H_t (X_t - X_{t-1})$ is the discrete-time integral. The desired result (Theorem 5.2) states that there is no duality gap, i.e. $\mathbf{I}_{\boldsymbol{\mu}}(f) = \mathbf{S}_{\boldsymbol{\mu}}(f)$, and that the dual problem is attained whenever it is finite. From the analysis for the one-step case in [10] we know that this assertion fails for the above naive formulation of the dual, and requires several relaxations regarding the integrability of the functions ϕ_t and the domain $\mathcal{V} \subseteq \mathbb{R}^{n+1}$ where the inequality (1.2) is required. Specifically, the inequality needs to be relaxed on sets that are $\mathcal{M}(\boldsymbol{\mu})$-polar; i.e. not charged by any transport $P \in \mathcal{M}(\boldsymbol{\mu})$. These sets are characterized in Theorem 3.1 where we show that the $\mathcal{M}(\boldsymbol{\mu})$-polar sets are precisely the (unions of) sets which project to a two-dimensional polar set of $\mathcal{M}(\mu_{t-1}, \mu_t)$ for some $1 \leq t \leq n$.

The duality theorem gives rise to a monotonicity principle (Theorem 5.4) that underpins the analysis of the left-monotone couplings. Similarly to the cyclical monotonicity condition in classical transport, it allows one to study the geometry of the support of optimal transports for a given function f.

1.3. Background and related literature

The martingale optimal transport problem (1.1) was introduced in [5] with the dual problem as a motivation. Indeed, in financial mathematics the function f is understood as the payoff of a derivative written on the underlying X and (1.2) corresponds to superhedging f by statically trading in European options $\phi_t(X_t)$ and dynamically trading in the underlying according to the strategy H. The value $\mathbf{I}_{\boldsymbol{\mu}}(f)$ then corresponds to the lowest price of f for which the seller can enter a model-free hedge $(\boldsymbol{\phi}, H)$ if the marginals $X_t \sim \mu_t$ are known from option market data. In [5], it was shown (with the above, "naive" formulation of the dual problem) that there is no duality gap if f is sufficiently regular, whereas dual existence was shown to fail even in regular cases. The idea of model-free hedging as well as the connection to Skorokhod embeddings goes back to [27]; we refer to [12–14,28,38,43] for further references. A specific multiperiod martingale optimal transport problem also arises in the study of the maximal maximum of a martingale given n marginals [23].

The one-step case $n = 1$ has been studied in great detail. In particular, [8] introduced the Left-Curtain coupling and pioneered numerous ideas underlying Theorem 1.1, [25] provided an explicit construction of that coupling, and [32] established the stability with respect to the marginals. Our duality results specialize to the ones of [10] when $n = 1$. Unsurprisingly, we shall exploit many arguments and results from these papers wherever possible. As indicated above, and as will be seen in the proofs below, the multistep case allows for a richer structure and necessitates novel ideas; for instance, the analysis of the polar sets (Theorem 3.1) is surprisingly involved. Other works in the one-step martingale case have studied reward functions f such as forward start straddles [29,30] or Asian payoffs [40]. We also refer to [15,21] for recent developments with multidimensional marginals.

One-step martingale optimal transport problems can alternately be studied as optimal Skorokhod embedding problems with marginal constraints; cf. [2,3,6,7]. A multi-marginal extension [1] of [2] is in preparation at the time of writing and the authors have brought to our attention that it will offer a version of Theorem 1.1 in the Skorokhod picture, at least in the

case where μ_0 is atomless and some further conditions are satisfied. The Skorokhod embedding problem with multi-marginal constraint was also studied in [22].

A multi-step coupling quite different from ours can be obtained by composing in a Markovian fashion the Left-Curtain transport kernels from μ_{t-1} to μ_t, $1 \le t \le n$, as discussed in [25]. In [33] the continuous-time limits of such couplings for $n \to \infty$ are studied to find solutions of the so-called Peacock problem [26] where the marginals for a continuous-time martingale are prescribed; see also [24] and [34] for other continuous-time results with full marginal constraint. Early contributions related to the continuous-time martingale transport problem include [17,18,20,36,39,42].

The remainder of the paper is organized as follows. Section 2 fixes basic terminology and recalls the necessary results from the one-step case. In Section 3, we characterize the polar structure of $\mathcal{M}(\boldsymbol{\mu})$. Section 4 introduces and analyzes the space that is the domain of the dual problem in Section 5, where we state the duality theorem and the monotonicity principle. Section 6 introduces left-monotone transports by the shadow construction and Section 7 develops the equivalent characterizations in terms of support and optimality properties. The (non-)uniqueness of left-monotone transports is discussed in Section 8. We conclude with the analysis of the problem with unconstrained intermediate marginals in Section 9.

2. Preliminaries

Throughout this paper, μ_t, μ, ν denote finite measures on $\mathbb{R}$ with finite first moment, the total mass not necessarily being normalized. Generalizing the notation from the Introduction to a vector $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ of such measures, we will write $\Pi(\boldsymbol{\mu})$ for the set of couplings; that is, measures P on $\mathbb{R}^{n+1}$ such that $P \circ X_t^{-1} = \mu_t$ for $0 \le t \le n$ where $X = (X_0, \dots, X_n) : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is the identity. Moreover, $\mathcal{M}(\boldsymbol{\mu})$ is the subset of all $P \in \Pi(\boldsymbol{\mu})$ that are martingales, meaning that

$$\int X_s \mathbf{1}_A(X_0, \dots, X_s) dP = \int X_t \mathbf{1}_A(X_0, \dots, X_s) dP$$

for all $s \le t$ and Borel sets $A \in \mathfrak{B}(\mathbb{R}^{s+1})$.

We denote by $\mathbb{F} = \{\mathfrak{F}_t\}_{0 \le t \le n}$ the canonical filtration $\mathfrak{F}_t := \sigma(X_0, \dots, X_t)$. As usual, an $\mathbb{F}$-predictable process $H = \{H_t\}_{1 \le t \le n}$ is a sequence of real functions on $\mathbb{R}^{n+1}$ such that H_t is $\mathfrak{F}_{t-1}$-measurable; i.e. $H_t = h_t(X_0, \dots, X_{t-1})$ for some Borel-measurable $h_t : \mathbb{R}^t \to \mathbb{R}$. Given an $\mathbb{F}$-predictable process H, the discrete stochastic integral $\{(H \cdot X)_t\}_{0 \le t \le n}$ is defined by

$$(H \cdot X)_t := \sum_{s=1}^{t} H_s \cdot (X_s - X_{s-1}).$$

If X is a martingale under some measure P, then $H \cdot X$ is a generalized (not necessarily integrable) martingale in the sense of generalized conditional expectations; cf. [31, Proposition 1.64].

We say that $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ is in convex order if $\mu_{t-1} \le_c \mu_t$ for all $1 \le t \le n$; that is, $\mu_{t-1}(\phi) \le \mu_t(\phi)$ for any convex function $\phi : \mathbb{R} \to \mathbb{R}$. This implies that μ_{t-1} and μ_t have the same total mass. The order can also be characterized by the potential functions

$$u_{\mu_t} : \mathbb{R} \to \mathbb{R}, \quad u_{\mu_t}(x) := \int |x - y| \, \mu_t(dy).$$

The following properties are elementary:

(i) u_{μ_t} is nonnegative and convex,

(ii) $\partial^+ u_{\mu_t}(x) - \partial^- u_{\mu_t}(x) = 2\mu_t(\{x\})$,
(iii) $\lim_{|x|\to\infty} u_{\mu_t}(x) = \infty \mathbf{1}_{\mu_t \neq 0}$,
(iv) $\lim_{|x|\to\infty} u_{\mu_t}(x) - \mu_t(\mathbb{R})|x - \text{bary}(\mu_t)| = 0$,

where ∂^+ and ∂^- denote the right and left derivatives, respectively, and $\text{bary}(\mu_t) = (\int x d\mu_t)/\mu_t(\mathbb{R})$ is the barycenter. We can therefore extend u_{μ_t} continuously to $\bar{\mathbb{R}} = [-\infty, \infty]$. The following result of Strassen is classical (cf. [41]; the last statement is obtained as e.g. in [19, Corollary 2.95]).

Proposition 2.1. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be finite measures on $\mathbb{R}$ with finite first moments and equal total mass. The following are equivalent:*

(i) $\mu_0 \leq_c \cdots \leq_c \mu_n$,
(ii) $u_{\mu_0} \leq \cdots \leq u_{\mu_n}$,
(iii) $\mathcal{M}(\boldsymbol{\mu}) \neq \emptyset$,
(iv) there exist stochastic kernels $\kappa_t(x_0, \dots, x_{t-1}, dx_t)$ such that

$$\int |x_t|\, \kappa_t(x_0, \dots, x_{t-1}, dx_t) < \infty \text{ and } \int x_t\, \kappa_t(x_0, \dots, x_{t-1}, dx_t) = x_{t-1}$$

for all $(x_0, \dots, x_t) \in \mathbb{R}^t$ and $1 \leq t \leq n$, and

$$\mu_t = (\mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n) \circ (X_t)^{-1} \quad \text{for all } 0 \leq t \leq n.$$

All kernels will be stochastic (i.e. normalized) in what follows. A kernel κ_t with the first property in (iv) is called *martingale kernel*.

2.1. The one-step case

For the convenience of the reader, we summarize some results from [8] and [10] for the one-step problem ($n = 1$) which will be used later on. In this section we write (μ, ν) instead of (μ_0, μ_1) for the given marginals in convex order.

Definition 2.2. The pair $\mu \leq_c \nu$ is *irreducible* if the set $I = \{u_\mu < u_\nu\}$ is connected and $\mu(I) = \mu(\mathbb{R})$. In this situation, let J be the union of I and any endpoints of I that are atoms of ν; then (I, J) is the *domain* of $\mathcal{M}(\mu, \nu)$.

The first result is a decomposition of the transport problem into irreducible parts; cf. [8, Theorem 8.4].

Proposition 2.3. *Let $\mu \leq_c \nu$ and let $(I_k)_{1\leq k\leq N}$ be the (open) components of $\{u_\mu < u_\nu\}$, where $N \in \{0, 1, \dots, \infty\}$. Set $I_0 = \mathbb{R}\backslash \cup_{k\geq 1} I_k$ and $\mu_k = \mu|_{I_k}$ for $k \geq 0$, so that $\mu = \sum_{k\geq 0} \mu_k$. Then, there exists a unique decomposition $\nu = \sum_{k\geq 0} \nu_k$ such that*

$$\mu_0 = \nu_0 \quad \text{and} \quad \mu_k \leq_c \nu_k \quad \text{for all} \quad k \geq 1,$$

and this decomposition satisfies $I_k = \{u_{\mu_k} < u_{\nu_k}\}$ for all $k \geq 1$. Moreover, any $P \in \mathcal{M}(\mu, \nu)$ admits a unique decomposition $P = \sum_{k\geq 0} P_k$ such that $P_k \in \mathcal{M}(\mu_k, \nu_k)$ for all $k \geq 0$.

We observe that the measure P_0 in Proposition 2.3 transports μ_0 to itself and is concentrated on $\Delta_0 := \Delta \cap I_0^2$ where $\Delta = \{(x, x) : x \in \mathbb{R}\}$ is the diagonal. Thus, the transport problem with index $k = 0$ is not actually an irreducible one, but we shall nevertheless refer to (I_0, I_0)

as the domain of this problem. When we want to emphasize the distinction, we call (I_0, I_0) the *diagonal domain* and $(I_k, J_k)_{k\geq 1}$ the *irreducible domains* of $\mathcal{M}(\mu,\nu)$. Similarly, the sets $V_k := I_k \times J_k$, $k \geq 1$ will be called the *irreducible components* and $V_0 := \Delta_0$ will be called the *diagonal component* of $\mathcal{M}(\mu,\nu)$. This terminology refers to the following result of [10, Theorem 3.2] which essentially states that the components are the only sets that can be charged by a martingale transport. We call a set $B \subseteq \mathbb{R}^2$ $\mathcal{M}(\mu,\nu)$-polar if it is P-null for all $P \in \mathcal{M}(\mu,\nu)$, where a nullset is, as usual, any set contained in a Borel set of zero measure.

Proposition 2.4. *Let $\mu \leq_c \nu$ and let $B \subseteq \mathbb{R}^2$ be a Borel set. Then B is $\mathcal{M}(\mu,\nu)$-polar if and only if there exist a μ-nullset N_μ and a ν-nullset N_ν such that*

$$B \subseteq (N_\mu \times \mathbb{R}) \cup (\mathbb{R} \times N_\nu) \cup \left(\bigcup_{k\geq 0} V_k\right)^c.$$

The following result of [10, Lemma 3.3] will also be useful; it is the main ingredient in the proof of the preceding proposition.

Lemma 2.5. *Let $\mu \leq_c \nu$ be irreducible and let π be a finite measure on $\mathbb{R}^2$ whose marginals π_1, π_2 satisfy[4] $\pi_1 \leq \mu$ and $\pi_2 \leq \nu$. Then, there exists $P \in \mathcal{M}(\mu,\nu)$ such that P dominates π in the sense of absolute continuity.*

3. The polar structure

The goal of this section is to identify all obstructions to martingale transports imposed by the marginals $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$, and thus, conversely, the sets that can indeed be charged. We recall that a subset B of $\mathbb{R}^{n+1}$ is called $\mathcal{M}(\boldsymbol{\mu})$-polar if it is a P-nullset for all $P \in \mathcal{M}(\boldsymbol{\mu})$. The result for the one-step case in Proposition 2.4 already exhibits an obvious type of polar set $B \subseteq \mathbb{R}^{n+1}$: if for some t there is an $\mathcal{M}(\mu_{t-1}, \mu_t)$-polar set $B' \subseteq \mathbb{R}^2$ such that $B \subseteq \mathbb{R}^{t-1} \times B' \times \mathbb{R}^{n-t}$, then B must be $\mathcal{M}(\boldsymbol{\mu})$-polar. The following shows that unions of such sets are in fact the only polar sets of $\mathcal{M}(\boldsymbol{\mu})$.

Theorem 3.1 (*Polar Structure*)**.** *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order. Then a Borel set $B \subseteq \mathbb{R}^{n+1}$ is $\mathcal{M}(\boldsymbol{\mu})$-polar if and only if there exist μ_t-nullsets N_t such that*

$$B \subseteq \bigcup_{t=0}^{n} (X_t)^{-1}(N_t) \cup \bigcup_{t=1}^{n} (X_{t-1}, X_t)^{-1} \left(\bigcup_{k\geq 0} V_k^t\right)^c \tag{3.1}$$

where $(V_k^t)_{k\geq 1}$ are the irreducible components of $\mathcal{M}(\mu_{t-1}, \mu_t)$ and V_0^t is the corresponding diagonal component.

Before stating the proof, we introduce some additional terminology. The second part of (3.1) can be expressed as

$$\begin{aligned} \bigcup_{t=1}^{n} (X_{t-1}, X_t)^{-1} \left(\bigcup_{k\geq 0} V_k^t\right)^c &= \left(\bigcap_{t=1}^{n} \bigcup_{k\geq 0} (X_{t-1}, X_t)^{-1}(V_k^t)\right)^c \\ &= \left(\bigcup_{k_1,\dots,k_n \geq 0} \bigcap_{t=1}^{n} (X_{t-1}, X_t)^{-1}(V_{k_t}^t)\right)^c. \end{aligned} \tag{3.2}$$

[4] By $\pi_1 \leq \mu$ we mean that $\pi_1(A) \leq \mu(A)$ for every Borel set $A \subseteq \mathbb{R}$.

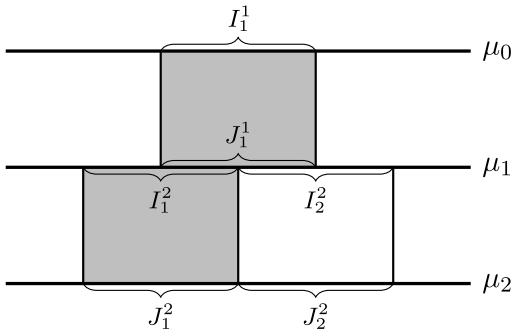

Fig. 2. The shaded area represents $V_{\boldsymbol{k}}$ for $\boldsymbol{k} = (1, 1)$.

For every $\boldsymbol{k} = (k_1, \dots, k_n)$, the set

$$V_{\boldsymbol{k}} = \bigcap_{t=1}^{n} (X_{t-1}, X_t)^{-1}(V_{k_t}^t) \subseteq \mathbb{R}^{n+1}$$

as occurring in the last expression of (3.2) will be referred to as an *irreducible component* of $\mathcal{M}(\boldsymbol{\mu})$; these sets are disjoint since $V_k^t \cap V_{k'}^t = \emptyset$ for $k \neq k'$. Moreover, we call their union

$$\mathcal{V} = \cup_{\boldsymbol{k}} V_{\boldsymbol{k}}$$

the *effective domain* of $\mathcal{M}(\boldsymbol{\mu})$.

Roughly speaking, an irreducible component $V_{\boldsymbol{k}}$ is a chain of irreducible components from the individual steps $(t-1, t)$. In the one-step case considered in [8,10], it was possible and useful to decompose the transport problem into its irreducible components and study those separately to a large extent; cf. Proposition 2.3. This is impossible in the multistep case, as illustrated by the following example.

Example 3.2. Consider the two-step martingale transport problem with marginals $\mu_0 = \delta_0$, $\mu_1 = \frac{1}{2}(\delta_{-1} + \delta_1)$ and $\mu_2 = \frac{1}{4}(\delta_{-2} + 2\delta_0 + \delta_2)$. Then the irreducible components are given by

$$\begin{aligned}
V_{00} &= \{(x, x, x) : x \notin (-2, 2)\} \\
V_{01} &= \{(x, x) : x \in (-2, -1]\} \times [-2, 0] \\
V_{02} &= \{(x, x) : x \in [1, 2)\} \times [0, 2] \\
V_{10} &= (-1, 1) \times \{0\} \times \{0\} \\
V_{11} &= (-1, 1) \times [-1, 0) \times [-2, 0] \\
V_{12} &= (-1, 1) \times (0, 1] \times [0, 2].
\end{aligned}$$

There is only one martingale transport $P \in \mathcal{M}(\boldsymbol{\mu})$, given by

$$P = \frac{1}{4}(\delta_{(0,-1,-2)} + \delta_{(0,-1,0)} + \delta_{(0,1,0)} + \delta_{(0,1,2)}).$$

While P is supported on $V_{11} \cup V_{12}$, it cannot be decomposed into two martingale parts that are supported on V_{11} and V_{12}, respectively: V_{11} and V_{12} are disjoint, but $P|_{V_{11}} = \frac{1}{4}(\delta_{(0,-1,-2)} + \delta_{(0,-1,0)})$ is not a martingale.

The main step in the proof of Theorem 3.1 will be the following lemma.

Lemma 3.3. *Let $V_{\boldsymbol{k}}$ be an irreducible component of $\mathcal{M}(\boldsymbol{\mu})$ and consider a measure π concentrated on $V_{\boldsymbol{k}}$ such that $\pi_t \leq \mu_t$ for $t = 0, \dots, n$. Then there exists a transport $P \in \mathcal{M}(\boldsymbol{\mu})$ which dominates π in the sense of absolute continuity.*

Deferring the proof, we first show how this implies the theorem.

Proof of Theorem 3.1. Clearly $(X_t)^{-1}(N_t)$ is $\mathcal{M}(\boldsymbol{\mu})$-polar for $t = 0, \dots, n$ and $(X_{t-1}, X_t)^{-1}\left(\bigcup_{k\geq 0} V_k^t\right)^c$ is $\mathcal{M}(\boldsymbol{\mu})$-polar for $t = 1, \dots, n$. This shows that (3.1) is sufficient for $B \subseteq \mathbb{R}^{n+1}$ to be $\mathcal{M}(\boldsymbol{\mu})$-polar.

Conversely, suppose that (3.1) does not hold; we show that B is not $\mathcal{M}(\boldsymbol{\mu})$-polar. In view of (3.2), by passing to a subset of B if necessary, we may assume that

$$B \subseteq \mathcal{V} = \bigcup_{\boldsymbol{k}} V_{\boldsymbol{k}} = \bigcup_{\boldsymbol{k}} \bigcap_{t=1}^{n} (X_{t-1}, X_t)^{-1}(V_{k_t}^t).$$

We may also assume that there are no μ_t-nullsets N_t such that $B \subseteq \cup_{t=0}^n (X_t)^{-1}(N_t)$. By a result of classical optimal transport [4, Proposition 2.1], this entails that B is not $\Pi(\boldsymbol{\mu})$-polar; i.e. we can find a measure $\rho \in \Pi(\boldsymbol{\mu})$ such that $\rho(B) > 0$.

We now write $B = \bigcup_{\boldsymbol{k}} B \cap V_{\boldsymbol{k}}$. As $\rho(B) = \sum_{\boldsymbol{k}} \rho(B \cap V_{\boldsymbol{k}}) > 0$, we can find some $\boldsymbol{k}$ such that $\rho(B \cap V_{\boldsymbol{k}}) > 0$. But then $\pi := \rho|_{V_{\boldsymbol{k}}}$ satisfies the assumptions of Lemma 3.3 which yields $P \in \mathcal{M}(\boldsymbol{\mu})$ such that $P \gg \pi$. In particular, $P(B) > 0$ and B is not $\mathcal{M}(\boldsymbol{\mu})$-polar. □

3.1. Proof of Lemma 3.3

The reasoning for Lemma 3.3 follows an induction on the number n of time steps; its rigorous formulation requires a certain amount of control over subsequent steps of the transport problem. Thus, we first state a more quantitative version of (the core part of) the lemma that is tailored to the inductive argument.

Definition 3.4. Let $\boldsymbol{\mu}$ be in convex order and $\mathcal{V}$ the effective domain of $\mathcal{M}(\boldsymbol{\mu})$. We say that a finite measure π has a *compact support family* if there are disjoint compact product sets[5] $K_1, \dots, K_m \subseteq \mathcal{V}$ with $\pi(\cup_i K_i) = \pi(\mathbb{R}^{n+1})$ such that $K_i \subseteq V_{\boldsymbol{k}_i}$ for some irreducible component $V_{\boldsymbol{k}_i}$ for all $i = 1, \dots, m$.

Definition 3.5. Let $\boldsymbol{\mu}$ be in convex order, $t \leq n$ and $\sigma \leq \mu_t$ a finite measure on $\mathbb{R}$. If $t = n$, we say that σ is *diagonally compatible* (with $\boldsymbol{\mu}$) if there is a finite family of compact sets $L_1, \dots, L_m \subseteq \mathbb{R}$ with $\sigma(\cup_i L_i) = \sigma(\mathbb{R})$. Whereas if $t < n$, we require in addition that for every i, either (a) $L_i \subseteq I_k$ for some irreducible component (I_k, J_k) of $\mathcal{M}(\mu_t, \mu_{t+1})$ or (b) $L_i \subseteq I_0$ and there is $t + 1 \leq t' \leq n$ such that $L_i \subseteq I_0^s$ for the diagonal components of $\mathcal{M}(\mu_s, \mu_{s+1})$ for all $t \leq s < t'$ and $L_i \subseteq I_k^{t'}$ for some (non-diagonal) irreducible component $(I_k^{t'}, J_k^{t'})$ of $\mathcal{M}(\mu_{t'}, \mu_{t'+1})$, where we set $I_k^n = J_k^n = \mathbb{R}$ for notational convenience.

Lemma 3.6. *Let $t < n$ and let $L \subseteq I_0$ be a compact interval contained in the diagonal component of $\mathcal{M}(\mu_t, \mu_{t+1})$ such that $\mu_t(L) > 0$. There exist a compact interval $L' \subseteq L$ with $\mu_t(L') > 0$ and $t + 1 \leq t' \leq n$ such that $L' \subseteq I_0^s$ for the diagonal component of $\mathcal{M}(\mu_s, \mu_{s+1})$ for all $t \leq s < t'$ and $L' \subseteq I_k^{t'}$ for some (non-diagonal) irreducible component $(I_k^{t'}, J_k^{t'})$ of $\mathcal{M}(\mu_{t'}, \mu_{t'+1})$, where we again set $I_k^n = J_k^n = \mathbb{R}$ for notational convenience.*

[5] By a compact product set we mean a set $K = A_0 \times \cdots \times A_n$ where each $A_t \subseteq \mathbb{R}$ is compact.

Proof. The statement is trivially satisfied for $t = n - 1$ as we can just take $L' = L$. For $t < n - 1$, consider the family of irreducible components (I_k^{t+1}, J_k^{t+1}) of $\mathcal{M}(\mu_{t+1}, \mu_{t+2})$. We distinguish three cases.

(i) First, consider the case where $L \cap I_k^{t+1} = \emptyset$ for all $k \geq 1$, then L is contained in the diagonal component of $\mathcal{M}(\mu_{t+1}, \mu_{t+2})$.

(ia) If $L = \{x\}$ consists of a single point with positive mass, then we can conclude by induction from the result for $t + 1$.

(ib) If no endpoint of L is on the boundary of some component I_k^t, then observe that $\mu_t|_L = \mu_{t+1}|_L$. We can find $L' \subseteq L$ from the statement of the lemma for $t + 1$. Then L' gives the result as $\mu_t(L') = \mu_{t+1}(L') > 0$.

(ic) If L contains more than one point, and also the endpoint of some component I_k^t. When this endpoint x has positive point mass, we can set $L' = \{x\}$ and conclude as in (ia). If the endpoint has zero mass, we can find $\bar{L} \subseteq L$ compact with $\mu_t(\bar{L}) > 0$ that does not contain this endpoint and argue as in (ib). (Observe that there might be at most two endpoints.)

(ii) Next, let $k \geq 1$ be such that $\mu_{t+1}(L \cap I_k^{t+1}) > 0$ (and in particular $L \cap I_k^{t+1} \neq \emptyset$). Then we can find a compact interval $L' \subseteq L \cap I_k^{t+1}$ such that $\mu_t(L') > 0$ and we directly see that L' satisfies the statement of the lemma.

(iii) Finally, suppose that there is $k \geq 1$ with $L \cap I_k^{t+1} \neq \emptyset$ but $\mu_t(L \cap I_k^{t+1}) = 0$. In particular this means that $L \not\subseteq I_k^{t+1}$. It furthermore means that $I_k^{t+1} \not\subseteq L$, as otherwise $\mu_{t+1}(I_k^{t+1}) = \mu_t(I_k^{t+1}) = \mu_t(L \cap I_k^{t+1}) = 0$ which contradicts the definition of I_k^{t+1}. As L is a compact interval and I_k^{t+1} is an open interval, we have that $L \backslash I_k^{t+1}$ is a compact interval and $\mu_t(L \backslash I_k^{t+1}) = \mu_t(L) > 0$. Notice that there can be at most two such components I_k^{t+1} for fixed L and we will be in case (i) after removing both of them if necessary. □

Lemma 3.7. *Let $t \leq n$ and let $J \subseteq \mathbb{R}$ be an interval such that $\mu_t(J) > 0$. Then we can find a compact interval $K \subseteq J$ with $\mu_t(K) > 0$ such that $\mu_t|_K$ is diagonally compatible.*

Proof. The case $t = n$ is trivial. Thus, let $t < n$. We consider the family $\{I_k\}_{k \geq 1}$ of open sets corresponding to the irreducible components of $\mathcal{M}(\mu_t, \mu_{t+1})$ and distinguish two cases.

(i) There is some $k \geq 1$ such that $\mu_t(I_k \cap J) > 0$. In this case, we can choose a compact interval $K \subseteq I_k \cap J$ such that $\mu_t(K) > 0$.

(ii) Now suppose that $\mu_t(I_k \cap J) = 0$ for all $k \geq 1$. Then we first notice that there are at most two components I_{k_1}, I_{k_2} so that $I_{k_i} \cap J \neq \emptyset$ and $J \backslash (I_{k_1} \cup I_{k_2})$ is still a nonempty interval with positive μ_t-mass, since I_k cannot be contained in J. We can therefore assume without loss of generality that $J \subseteq I_0$ and is compact. Now we can apply Lemma 3.6 to find a subinterval $K \subseteq J$ such that $\mu_t|_K$ is diagonally compatible. □

Lemma 3.8. *Let $t \leq n$ and let π be a measure on $\mathbb{R}^{t+1}$ that has a compact support family with respect to $\mu_0, \ldots, \mu_t$ and satisfies $\pi_s \leq \mu_s$ for $s \leq t$. In addition, suppose that π_t is diagonally compatible.*

Then there is a martingale measure Q on $\mathbb{R}^{t+1}$ that dominates π in the sense of absolute continuity and has a compact support family with respect to $\mu_0, \ldots, \mu_t$ and satisfies $Q_s \leq \mu_s$ for $s \leq t$. In addition, Q_t can be chosen to be diagonally compatible. Finally, Q can be chosen such that $dQ = g d\pi + d\sigma$ where the density g is bounded and the measure σ is singular with respect to π.

Proof. We proceed by induction on t. For $t = 0$ there is nothing to prove; we can set $Q = \pi$.

Consider $t \geq 1$ and assume that the lemma has already been shown for $(t-1)$-step measures. We disintegrate

$$\pi = \pi' \otimes \kappa(x_0, \dots, x_{t-1}, dx_t) \tag{3.3}$$

and observe that π' satisfies the conditions of the lemma. In particular, π'_{t-1} must be diagonally compatible: the compact sets that it is supported on are either contained in irreducible components of $\mathcal{M}(\mu_{t-1}, \mu_t)$ or in the diagonal component. Any such compact subset of the diagonal component of $\mathcal{M}(\mu_{t-1}, \mu_t)$ must correspond to one of the finitely many compact sets in the support of π_t so that they inherit the compatibility property from these sets.

By the induction assumption, we then find a martingale measure $Q' \gg \pi'$ on $\mathbb{R}^t$ with the stated properties. In particular, the marginal Q'_{t-1} is diagonally compatible with $\boldsymbol{\mu}$.

Again, let $\{I_k\}_{k\geq 1}$ be the open intervals from the irreducible domains (I_k, J_k) of $\mathcal{M}(\mu_{t-1}, \mu_t)$ and let I_0 denote the corresponding diagonal domain. We shall construct a martingale kernel $\hat{\kappa}$ by suitably manipulating κ. Let us observe that since π is concentrated on $\mathcal{V}$ and has a compact support family with respect to $\mu_0, \dots, \mu_t$, the following hold for π'-a.e. $\boldsymbol{x} = (x_0, \dots, x_{t-1}) \in \mathbb{R}^t$ and a finite family of compact sets L_i with properties (a) or (b) from Definition 3.5:

- $\kappa(\boldsymbol{x}, \cdot) = \delta_{x_{t-1}}$ whenever $x_{t-1} \in I_0$,
- $\kappa(\boldsymbol{x}, \cdot)$ is concentrated on some L_i with $L_i \subseteq J_k$ for $x_{t-1} \in I_k$ with $k \geq 1$ and $Q'_{t-1}(I_k) > 0$.

By changing κ on a π'-nullset, we may assume that these two properties hold for all $\boldsymbol{x} \in \mathbb{R}^t$.

Step 1. Next, we argue that we may change Q' and κ such that the marginal $(Q' \otimes \kappa)_t = (Q' \otimes \kappa) \circ X_t^{-1}$ satisfies

$$(Q' \otimes \kappa)_t \leq \mu_t. \tag{3.4}$$

Indeed, recall that $dQ' = dQ'_{abs} + d\sigma' = g' d\pi' + d\sigma'$ where the density g' is bounded and σ' is singular with respect to π'. Using the Lebesgue decomposition theorem, we find a Borel set $A \subseteq \mathbb{R}^t$ such that $\sigma'(A) = \sigma'(\mathbb{R}^t)$ and $\pi'(A) = 0$. By scaling Q' with a constant we may assume that $g' \leq 1/2$. As $\pi_t \leq \mu_t$, the marginal $(Q'_{abs} \otimes \kappa)_t$ is then bounded by $\frac{1}{2}\mu_t$, and it remains to bound $(\sigma' \otimes \kappa)_t$ in the same way.

Note that $Q'_{t-1} \leq \mu_{t-1}$ implies $\sigma'_{t-1} \leq \mu_{t-1}$. We may change κ arbitrarily on the set A without invalidating (3.3). Indeed, for each irreducible component (I_k, J_k) of $\mathcal{M}(\mu_{t-1}, \mu_t)$ we choose and fix a compact interval $K_k \subseteq J_k$ with $\mu_t(K_k) > 0$ such that $\mu_t|_{K_k}$ is diagonally compatible; this is possible by Lemma 3.7. For $\boldsymbol{x} = (x_0, \dots, x_{t-1}) \in A$ such that $x_{t-1} \in I_k$ we then define

$$\kappa(\boldsymbol{x}, \cdot) := \frac{1}{\mu_t(K_k)} \mu_t|_{K_k}.$$

Set $\epsilon_k = \mu_t(K_k)/\mu_{t-1}(I_k)$. Then

$$\epsilon := \inf_{k:\, Q'_{t-1}(I_k)>0} \epsilon_k \wedge 1$$

is strictly positive because there are only finitely many k with $Q'_{t-1}(I_k) > 0$ (this is the purpose of the induction assumption that Q'_{t-1} is diagonally compatible). As $\sigma'_{t-1} \leq \mu_{t-1}$, we may scale Q' once again to obtain $\sigma'_{t-1} \leq \frac{\epsilon}{6}\mu_{t-1}$. We now have

$$(\sigma'|_{\mathbb{R}^{t-1}\times I_k} \otimes \kappa)_t = \sigma'_{t-1}(I_k) \frac{1}{\mu_t(K_k)} \mu_t|_{K_k} \leq \frac{\epsilon}{6} \frac{\mu_{t-1}(I_k)}{\mu_t(K_k)} \mu_t|_{K_k} \leq \frac{1}{6}\mu_t|_{K_k}.$$

For the diagonal domain I_0 the corresponding inequality holds because we have $\kappa(\boldsymbol{x},\cdot)=\delta_{x_{t-1}}$ for $x_{t-1}\in I_0$ and $\sigma'_{t-1}|_{I_0}\leq\frac{1}{6}\mu_{t-1}|_{I_0}\leq\frac{1}{6}\mu_t|_{I_0}$. As a consequence, we have $(\sigma'\otimes\kappa)_t\leq\frac{1}{2}\mu_t$ as desired, so that we may assume (3.4) in what follows.

Step 2. We now construct a martingale kernel $\hat{\kappa}$ such that $Q=Q'\otimes\hat{\kappa}$ has the required properties. For a fixed irreducible component (I_k,J_k) we have that $Q'_{t-1}|_{I_k}=Q'_{t-1}|_K$ for some compact $K\subseteq I_k$. We can find compact intervals $B^-,B^+\subseteq J_k$ with $\mu_t(B^-)>0$ and $\mu_t(B^+)>0$ such that B^- is to the left of K and B^+ is to the right of K, in the sense that $x<y<z$ for $x\in B^-$, $y\in K$ and $z\in B^+$. By Lemma 3.7, we can further assume that we have $B^+\subseteq I^t_k$ and $B^-\subseteq I^t_{k'}$ for some $k,k'\geq 0$, where $(I^t_l)_{l\geq 0}$ belong to the components of $\mathcal{M}(\mu_t,\mu_{t+1})$, and that $\mu_t|_{B^\pm}$ is diagonally compatible.

Next, we define two nonnegative functions $\boldsymbol{x}\mapsto\varepsilon^-(\boldsymbol{x}),\varepsilon^+(\boldsymbol{x})$ for $\boldsymbol{x}=(x_0,\dots,x_{t-1})\in\mathbb{R}^{t-1}\times K$ as follows:

- for $\boldsymbol{x}$ such that $\mathrm{bary}(\kappa(\boldsymbol{x},\cdot))<x_{t-1}$, let ε^+ be the unique number such that $\kappa(\boldsymbol{x},\cdot)+\varepsilon^+(\boldsymbol{x})\cdot\mu_t|_{B^+}$ has barycenter x_{t-1},
- for $\boldsymbol{x}$ such that $\mathrm{bary}(\kappa(\boldsymbol{x},\cdot))>x_{t-1}$, let ε^- be the unique number such that $\kappa(\boldsymbol{x},\cdot)+\varepsilon^-(\boldsymbol{x})\cdot\mu_t|_{B^-}$ has barycenter x_{t-1},
- $\varepsilon^\pm(\boldsymbol{x})=0$ otherwise.

Observe that these numbers always exist because B^- and B^+ have positive mass and positive distance from the points $x_{t-1}\in K$. We now define the martingale kernel $\hat{\kappa}$ by

$$\hat{\kappa}(\boldsymbol{x}):=c(\varepsilon^-\cdot\mu_t|_{B^-}+\kappa+\varepsilon^+\cdot\mu_t|_{B^+})$$

where $0<c\leq 1$ is a normalizing constant such that $\hat{\kappa}$ is again a stochastic kernel. We also define $\hat{\kappa}(\boldsymbol{x})=\kappa(\boldsymbol{x})$ for $\boldsymbol{x}$ on the diagonal domain.

For each $k\geq 1$, let $B^\pm_k$ denote the sets associated with I_k as above. Once again, the number

$$C:=\frac{1}{3}\inf_{k:\,Q'_{t-1}(I_k)>0}[\mu_t(B^-_k)\wedge\mu_t(B^+_k)]$$

is strictly positive because there are only finitely many k with $Q'_{t-1}(I_k)>0$. We can now define

$$Q:=C\cdot(Q'\otimes\hat{\kappa}).$$

Then Q is a martingale transport whose marginals satisfy $Q_s\leq Q'_s\leq\mu_s$ for $0\leq s\leq t-1$ whereas $Q_t\leq\mu_t$ by (3.4), the construction of $\hat{\kappa}$ and the choice of C; indeed, for every $x_{t-1}\in I^t_k$ we have

$$\begin{aligned}3C\hat{\kappa}(\boldsymbol{x})&\leq 3C\varepsilon^-\cdot\mu_t|_{B^-}+3C\kappa+3C\varepsilon^+\cdot\mu_t|_{B^+}\\&\leq\mu_t|_{B^-}+\kappa+\mu_t|_{B^+}\leq 2\mu_t+\kappa.\end{aligned}$$

To see that Q_t is diagonally compatible, observe that Q_t is supported by a finite family of compact sets consisting of the following:

- a finite family of compact sets $\tilde{L}_i\subseteq I_0$ such that $Q'_{t-1}|_{\tilde{L}_i}$ is diagonally compatible (from the induction hypothesis that Q'_{t-1} is diagonally compatible),
- a finite family of compact sets $L_i\subseteq J_k$ for some $k\geq 1$ with $Q'_{t-1}(I_k)>0$ such that $Q_t|_{L_i}\leq\mu_t|_{L_i}$ is diagonally compatible, and
- the sets $B^\pm_k$ for the finitely many k such that $Q'_{t-1}(I_k)>0$, where $Q_t|_{B^\pm_k}\leq\mu_t|_{B^\pm_k}$ is diagonally compatible.

It remains to check that Q has the required decomposition with respect to π. Indeed, $\hat{\kappa}$ can be decomposed as

$$\hat{\kappa} = c\kappa + (1-c)\kappa^{\perp}$$

where $\kappa^{\perp}$ is singular to κ. Recalling the decomposition $Q' = Q'_{abs} + \sigma'$, we then have

$$Q' \otimes \hat{\kappa} = c\,Q'_{abs} \otimes \kappa \;+\; (1-c)Q'_{abs} \otimes \kappa^{\perp} \;+\; \sigma' \otimes \hat{\kappa}.$$

The last two terms are singular with respect to $\pi = \pi' \otimes \kappa$, and the first term is absolutely continuous with bounded density. □

Proof of Lemma 3.3. Let π be a measure with marginals $\pi_t \leq \mu_t$ for all t which is concentrated on some irreducible component $V = V_{\boldsymbol{k}}$ and thus, in particular, on the effective domain $\mathcal{V}$.

Step 1. We first decompose $\pi = \sum_{m=1}^{\infty} \pi^m$ such that each π^m satisfies the requirements of Lemma 3.8 with $t = n$.

Indeed, let $V = \cap_{t=1}^{n}(X_{t-1}, X_t)^{-1}(V_{k_t}^t)$ and suppose first that $k_t \neq 0$ for $1 \leq t \leq n$. Then, we can write V as a product of nonempty intervals: $V = A_0 \times \cdots \times A_n$ where $A_0 = I_{k_1}^1$, $A_n = J_{k_n}^n$ and $A_t = J_{k_t}^t \cap I_{k_{t+1}}^{t+1}$ for $1 < t < n$. Thus, we can choose increasing families of compact intervals K_t^m such that $A_t = \cup_{m\geq 1} K_t^m$ for all t. Setting $\pi^1 := \pi|_{\prod_{t=0}^n K_t^1}$ and $\pi^m := \pi|_{\prod_{t=0}^n K_t^m \setminus \prod_{t=0}^n K_t^{m-1}}$ for $m > 1$ yields the required decomposition.

If $k_t = 0$ for one or more $1 \leq t \leq n$, we have $V \subseteq A_0 \times \cdots \times A_n$, where A_t is defined as above when $k_t \neq 0 \neq k_{t+1}$ but we use $\mathbb{R}$ instead of $J_{k_t}^t$ when $k_t = 0$ and $\mathbb{R}$ instead of $I_{k_{t+1}}^{t+1}$ when $k_{t+1} = 0$. After these modifications, π^m can be defined as above; recall that diagonal components are always closed.

Step 2. For each of the measures π^m, Lemma 3.8 yields a martingale measure $Q^m \gg \pi^m$ with the properties stated in the lemma. In particular, each Q^m has a compact support family. We show below that there exist $P^m \in \mathcal{M}(\boldsymbol{\mu})$ such that $P^m \gg Q^m$, and then $P := \sum 2^{-m} P^m$ satisfies $P \in \mathcal{M}(\boldsymbol{\mu})$ and $P \gg \pi$ as desired.

To complete the proof, it remains to show that for fixed $m \geq 1$ there exist $0 < \epsilon < 1$ and $\bar{Q}^m \in \mathcal{M}(\boldsymbol{\mu} - \epsilon(Q_0^m, \ldots, Q_n^m))$, as we may then conclude by setting $P^m := \epsilon Q^m + \bar{Q}^m \in \mathcal{M}(\boldsymbol{\mu})$. By Proposition 2.1, the set $\mathcal{M}(\boldsymbol{\mu} - \epsilon(Q_0^m, \ldots, Q_n^m))$ is nonempty if the marginals are in convex order, or equivalently if the potential functions satisfy

$$u_{\mu_{t-1}} - \epsilon u_{Q_{t-1}^m} \leq u_{\mu_t} - \epsilon u_{Q_t^m} \tag{3.5}$$

for $t = 1, \ldots, n$. Thus, it suffices to find $\epsilon > 0$ with this property for fixed t, and we have reduced to a question about a one-step martingale transport problem. Indeed, we have $u_{\mu_{t-1}} \leq u_{\mu_t}$ on $\mathbb{R}$. Since Q^m has a compact support family and in particular is supported by $\mathcal{V}$, there is a finite collection of compact sets $K_j \subseteq \mathbb{R}$ such that each K_j is contained in one of the intervals $I_{k_j}^{t-1}$ from the decomposition of (μ_{t-1}, μ_t) into irreducible components, Q^m transports mass from K_j to itself for each j, and Q^m is the identical Monge transport on the complement $(\cup_j K_j)^c$. On each K_j, Steps (a) and (b) in the proof of [10, Lemma 3.3] yield $\epsilon > 0$ such that (3.5) holds on K_j, and we can choose $\epsilon > 0$ independently of j since there are finitely many j. On the other hand, (3.5) trivially holds on $(\cup_j K_j)^c$ since $u_{Q_{t-1}^m} = u_{Q_t^m}$ on that set. This completes the proof. □

4. The dual space

In this section we introduce the domain of the dual optimization problem and show that it has a certain closedness property. The latter will be crucial for the duality theorem in the subsequent section.

We shall need a generalized notion of integrability for the elements of the dual space. To this end, we first recall the integral for concave functions as detailed in [10, Section 4.1].

Definition 4.1. Let $\mu \leq_c \nu$ be irreducible with domain (I, J) and let $\chi : J \to \mathbb{R}$ be a concave function. We define

$$(\mu - \nu)(\chi) := \frac{1}{2}\int_I (u_\mu - u_\nu) d\chi'' + \int_{J\setminus I} |\Delta\chi| d\nu \in [0, \infty]$$

where $-\chi''$ is the (locally finite) second derivative measure of $-\chi$ on I and $|\Delta\chi|$ is the absolute magnitude of the jumps of χ at the boundary points $J\setminus I$.

Remark 4.2. As shown in [10, Lemma 4.1], this integral is well-defined and satisfies

$$(\mu - \nu)(\chi) = \int_I \left[\chi(x) - \int_J \chi(y)\,\kappa(x, dy)\right]\mu(dx)$$

for any $P = \mu \otimes \kappa \in \mathcal{M}(\mu, \nu)$. Moreover, it coincides with the difference $\mu(\chi) - \nu(\chi)$ of the usual integrals when $\chi \in L^1(\mu) \cap L^1(\nu)$.

For later reference, we record two more properties of the integral.

Lemma 4.3. *Let $\mu \leq_c \nu$ be irreducible with domain (I, J) and let $\chi : J \to \mathbb{R}$ be concave.*

(i) Assume that I has a finite right endpoint r and $\chi(a) = \chi'(a) = 0$ for some $a \in I$. Then $\chi \leq 0$ and $\chi \mathbf{1}_{[a,\infty)}$ is concave. If ν has an atom at r, then

$$\chi(r) \geq -\frac{C}{\nu(\{r\})}(\mu - \nu)(\chi \mathbf{1}_{[a,\infty)})$$

for a constant $C \geq 0$ depending only on μ, ν.

(ii) For $a, b \in \mathbb{R}$, the concave function $\bar{\chi}(x) := \chi(x) + ax + b$ satisfies

$$(\mu - \nu)(\bar{\chi}) = (\mu - \nu)(\chi).$$

Proof. The first part is [10, Remark 4.6] and the second part follows directly from $\bar{\chi}'' = \chi''$ and $\Delta\bar{\chi} = \Delta\chi$. □

Let us now return to the multistep case with a vector $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_n)$ of measures in convex order and introduce $\boldsymbol{\mu}(\boldsymbol{\phi}) := \sum_{t=0}^n \mu_t(\phi_t)$ in cases where we do not necessarily have $\phi_t \in L^1(\mu_t)$. As mentioned previously, in contrast to [10], the multistep transport problem does not decompose into irreducible components, forcing us to directly give a global definition of the integral.

Definition 4.4. Let $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_n)$ be a vector of Borel functions $\phi_t : \mathbb{R} \to \bar{\mathbb{R}}$. A vector $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_n)$ of Borel functions $\chi_t : \mathbb{R} \to \mathbb{R}$ is called a *concave moderator* for $\boldsymbol{\phi}$ if for $1 \leq t \leq n$,

(i) $\chi_t|_J$ is concave for every domain (I, J) of an irreducible component of $\mathcal{M}(\mu_{t-1}, \mu_t)$,

(ii) $\chi_t|_{I_0} \equiv 0$ for the diagonal domain I_0 of $\mathcal{M}(\mu_{t-1}, \mu_t)$,
(iii) $\phi_t - \chi_{t+1} + \chi_t \in L^1(\mu_t)$,

where $\chi_{n+1} \equiv 0$. We also convene that $\chi_0 \equiv 0$. The *moderated integral* of $\boldsymbol{\phi}$ is then defined by

$$\boldsymbol{\mu}(\boldsymbol{\phi}) := \sum_{t=0}^{n} \mu_t(\phi_t - \chi_{t+1} + \chi_t) + \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t) \in (-\infty, \infty], \tag{4.1}$$

where $(\mu_{t-1} - \mu_t)^k(\chi_t)$ denotes the integral of Definition 4.1 on the kth irreducible component of $\mathcal{M}(\mu_{t-1}, \mu_t)$.

Remark 4.5. The moderated integral is independent of the choice of the moderator $\boldsymbol{\chi}$. To see this, consider a second moderator $\tilde{\boldsymbol{\chi}}$ for $\boldsymbol{\phi}$; then we have $(\tilde{\chi}_{t+1} - \chi_{t+1}) - (\tilde{\chi}_t - \chi_t) \in L^1(\mu_t)$. We may assume that (4.1) is finite for at least one of the moderators. Using Remark 4.2 with arbitrary κ_t such that $\mu_{t-1} \otimes \kappa_t \in \mathcal{M}(\mu_{t-1}, \mu_t)$ for $1 \leq t \leq n$, as well as Fubini's theorem for kernels,

$$\begin{aligned}
&\sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t) - (\mu_{t-1} - \mu_t)^k(\tilde{\chi}_t) \\
&\quad = \int \cdots \int \sum_{t=1}^{n} \chi_t(x_{t-1}) - \chi_t(x_t)\, \kappa_n(x_{n-1}, dx_n) \cdots \kappa_1(x_0, dx_1)\mu_0(dx_0) \\
&\qquad - \int \cdots \int \sum_{t=1}^{n} \tilde{\chi}_t(x_{t-1}) - \tilde{\chi}_t(x_t)\, \kappa_n(x_{n-1}, dx_n) \cdots \kappa_1(x_0, dx_1)\mu_0(dx_0) \\
&\quad = \sum_{t=0}^{n} \mu_t((\chi_{t+1} - \tilde{\chi}_{t+1}) - (\chi_t - \tilde{\chi}_t)).
\end{aligned}$$

It now follows that (4.1) yields the same value for both moderators.

For later reference, we also record the following property.

Remark 4.6. If $\boldsymbol{\chi}$ is a concave moderator, Definition 4.4(ii) implies that

$$\chi_t = \sum_{k\geq 1} \chi_t|_{I_k^t} = \sum_{k\geq 1} \chi_t|_{J_k^t}$$

where (I_k^t, J_k^t) is the kth irreducible domain of $\mathcal{M}(\mu_{t-1}, \mu_t)$.

Next, we introduce the space of functions which have a finite integral in the moderated sense.

Definition 4.7. We denote by $L^c(\boldsymbol{\mu})$ the space of all vectors $\boldsymbol{\phi}$ admitting a concave moderator $\boldsymbol{\chi}$ with $\sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t) < \infty$.

It follows that $\boldsymbol{\mu}(\boldsymbol{\phi})$ is finite for $\boldsymbol{\phi} \in L^c(\boldsymbol{\mu})$, and we have $\boldsymbol{\mu}(\phi) = \sum_t \mu_t(\phi_t)$ for $\boldsymbol{\phi} \in \varPi_{t=0}^{n} L^1(\mu_t)$. The definition is also consistent with the expectation under martingale transports, in the following sense.

Lemma 4.8. *Let $\boldsymbol{\phi} \in L^c(\boldsymbol{\mu})$ and let $H = (H_1, \dots, H_n)$ be $\mathbb{F}$-predictable. If*

$$\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n$$

is bounded from below on the effective domain $\mathcal{V}$ of $\mathcal{M}(\boldsymbol{\mu})$, then

$$\boldsymbol{\mu}(\boldsymbol{\phi}) = P\left[\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n\right], \quad P \in \mathcal{M}(\boldsymbol{\mu}).$$

Proof. Let $P \in \mathcal{M}(\boldsymbol{\mu})$, let $\boldsymbol{\chi}$ be a concave moderator for $\boldsymbol{\phi}$, and assume without loss of generality that 0 is the lower bound. Using Remark 4.6, we have that $\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n$ equals

$$\sum_{t=0}^{n} (\phi_t - \chi_{t+1} + \chi_t)(X_t) + \sum_{t=1}^{n} \sum_{k\geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n \geq 0.$$

By assumption, the functions $(\phi_t - \chi_{t+1} + \chi_t)(X_t)$ are P-integrable. Therefore, the negative part of the remaining expression must also be P-integrable. Writing $P_t := P \circ (X_0, \dots, X_t)^{-1}$ and using that $(\chi_t|_{J_k^t})^+$ has linear growth, we see that for any disintegration $P = P_{n-1} \otimes \kappa_n$,

$$\begin{aligned}
&\int \left[\sum_{t=1}^{n} \sum_{k\geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n\right] \kappa_n(X_0, \dots, X_{n-1}, dX_n) \\
&\quad = \sum_{t=1}^{n-1} \sum_{k\geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_{n-1} \\
&\qquad + \sum_{k\geq 1} \int \left[\chi_n|_{I_k^n}(X_{n-1}) - \chi_n|_{J_k^n}(X_n)\right] \kappa_n(X_0, \dots, X_{n-1}, dX_n).
\end{aligned}$$

Iteratively integrating with kernels such that $P_t = P_{t-1} \otimes \kappa_t$ and observing that we can apply Fubini's theorem to $\sum_{t=1}^{n} \sum_{k\geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n$ as its negative part is P-integrable, we obtain

$$P\left[\sum_{t=1}^{n} \sum_{k\geq 1} (\chi_t|_{I_k^t}(X_{t-1}) - \chi_t|_{J_k^t}(X_t)) + (H \cdot X)_n\right] = \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t)$$

and the result follows. □

We can now define our dual space. It will be convenient to work with nonnegative reward functions f for the moment—we shall relax this constraint later on; cf. Remark 5.3.

Definition 4.9. Let $f : \mathbb{R}^{n+1} \to [0, \infty]$. We denote by $\mathcal{D}_{\boldsymbol{\mu}}(f)$ the set of all pairs $(\boldsymbol{\phi}, H)$ where $\boldsymbol{\phi} \in L^c(\boldsymbol{\mu})$ and $H = (H_1, \dots, H_n)$ is an $\mathbb{F}$-predictable process such that

$$\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n \geq f \quad \text{on} \quad \mathcal{V}.$$

By Lemma 4.8, the expectation of the left hand side under any $P \in \mathcal{M}(\boldsymbol{\mu})$ is given by the moderated integral $\boldsymbol{\mu}(\boldsymbol{\phi})$; this will be seen as the dual cost of $(\boldsymbol{\phi}, H)$ when we consider the dual problem $\inf_{(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)} \boldsymbol{\mu}(\boldsymbol{\phi})$ in Section 5.

The following closedness property is the key result about the dual space.

Proposition 4.10. *Let $f^m : \mathbb{R}^{n+1} \to [0,\infty]$, $m \geq 1$ be a sequence of functions such that*

$$f^m \to f \qquad \text{pointwise}$$

and let $(\boldsymbol{\phi}^m, H^m) \in \mathcal{D}_{\boldsymbol{\mu}}(f^m)$ be such that $\sup_m \boldsymbol{\mu}(\boldsymbol{\phi}^m) < \infty$. Then there exist $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$ with

$$\boldsymbol{\mu}(\boldsymbol{\phi}) \leq \liminf_{m\to\infty} \boldsymbol{\mu}(\boldsymbol{\phi}^m).$$

4.1. Proof of Proposition 4.10

An attempt to prove Proposition 4.10 directly along the lines of [10] runs into a technical issue in controlling the concave moderators. Roughly speaking, they do not allow sufficiently many normalizations; this is related to the aforementioned fact that the multistep problem cannot be decomposed into its components. We shall introduce a generalized dual space with families of functions indexed by the components, and prove a "lifted" version of Proposition 4.10 in this larger space. Once that is achieved, we can infer the closedness result in the original space as well. (The reader willing to admit Proposition 4.10 may skip this subsection without much loss of continuity.)

Definition 4.11. Let $\boldsymbol{\phi} = \{\phi_t^k : 0 \leq t \leq n,\ k \geq 0\}$ be a family of Borel functions, consisting of one function $\phi_t^k : J_k^t \to \bar{\mathbb{R}}$ for each irreducible component (I_k^t, J_k^t) of $\mathcal{M}(\mu_{t-1}, \mu_t)$ as indexed by $k \geq 1$ and $1 \leq t \leq n$, functions $\phi_t^0 : I_0^t \to \bar{\mathbb{R}}$ for the diagonal components I_0^t indexed by $1 \leq t \leq n$, and a single function $\phi_0^0 : \mathbb{R} \to \bar{\mathbb{R}}$ for $t = 0$. Similarly, let $\boldsymbol{\chi} = \{\chi_t^k : 1 \leq t \leq n,\ k \geq 0\}$ be a family of functions, consisting of one concave function $\chi_t^k : J_k^t \to \mathbb{R}$ for each irreducible component (I_k^t, J_k^t) and Borel functions $\chi_t^0 : I_0^t \to \mathbb{R}$ for the diagonal components. We also convene that $\chi_0^0 \equiv 0$ and define the functions[6] $\chi_t := \sum_{k\geq 0} \chi_t^k|_{I_k^t}$ for $t = 1, \dots, n$, as well as $\chi_{n+1} \equiv 0$.

We call $\boldsymbol{\chi}$ a *concave moderator* for $\boldsymbol{\phi}$ if for all $t = 0, \dots, n$ and $k \geq 0$,

$$\phi_t^k + \chi_t^k - \chi_{t+1} \in L^1(\mu_t^k)$$

and the sum $\sum_{k\geq 0} \mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1})$ converges in $(-\infty, \infty]$, where μ_t^k is the second marginal of the kth irreducible component in the decomposition of $\mathcal{M}(\mu_{t-1}, \mu_t)$ as in Proposition 2.3 and $\mu_0^0 \equiv \mu_0$. The generalized[7] moderated integral is then defined by

$$\boldsymbol{\mu}(\boldsymbol{\phi}) := \sum_{t=0}^{n} \sum_{k\geq 0} \mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1}) + \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k).$$

We denote by $L^{c,g}(\boldsymbol{\mu})$ the set of all families $\boldsymbol{\phi}$ which admit a concave moderator $\boldsymbol{\chi}$ such that

$$\sum_{t=0}^{n} \sum_{k\geq 0} |\mu_t^k(\phi_t^k + \chi_t^k - \chi_{t+1})| + \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k) < \infty.$$

For $\boldsymbol{\phi} \in L^{c,g}(\boldsymbol{\mu})$, the value of $\boldsymbol{\mu}(\boldsymbol{\phi})$ is independent of the choice of the moderator $\boldsymbol{\chi}$. This is shown similarly as in Remark 4.5. We can now introduce the generalized dual space.

[6] The restriction to I_k^t is important to avoid "double counting" in the sums. Note that the intervals J may overlap at their endpoints.

[7] This integral is not related to the notion of a generalized martingale.

Definition 4.12. Let $f : \mathbb{R}^{n+1} \to [0, \infty]$. We denote by $\mathcal{D}^g_{\boldsymbol{\mu}}(f)$ the set of all pairs $(\boldsymbol{\phi}, H)$ where $\boldsymbol{\phi} \in L^{c,g}(\boldsymbol{\mu})$, $H = (H_1, \dots, H_n)$ is $\mathbb{F}$-predictable, and

$$\sum_{t=0}^{n} \phi_t^{k_t}(x_t) + (H \cdot \boldsymbol{x})_n \geq f(\boldsymbol{x})$$

for all $\boldsymbol{x} = (x_0, \dots, x_n)$ and $\boldsymbol{k} = (k_0, \dots, k_n)$ such that $(x_{t-1}, x_t) \in (I^t_{k_t}, J^t_{k_t})$ for some (irreducible or diagonal) component[8] and $t = 1, \dots, n$.

We observe that for any $\boldsymbol{x} \in \mathcal{V}$ the corresponding $\boldsymbol{k} = (k_0, \dots, k_n)$ is uniquely defined, where the index $k_0 \equiv 0$ exists purely for notational convenience.

For later reference, the following lemma elaborates on certain degrees of freedom in choosing elements of $\mathcal{D}^g_{\boldsymbol{\mu}}(f)$.

Lemma 4.13. *Let $(\boldsymbol{\phi}, H) \in \mathcal{D}^g_{\boldsymbol{\mu}}(f)$ and let $\boldsymbol{\chi}$ be a corresponding concave moderator. Let $1 \leq t \leq n$, let (I^t_k, J^t_k) be the domain of an irreducible component of $\mathcal{M}(\mu_{t-1}, \mu_t)$ and $c_1, c_2 \in \mathbb{R}$. Introduce new families $(\tilde{\boldsymbol{\phi}}, \tilde{H})$ and $\tilde{\boldsymbol{\chi}}$ by either (i) or (ii):*

(1) Define

$$\begin{aligned}
\tilde\phi^k_t(y) &= \phi^k_t(y) - (c_1 y - c_2), \qquad \tilde\chi^k_t(y) = \chi^k_t(y) + (c_1 y - c_2),\\
\tilde\phi^{k'}_{t-1}(x) &= \phi^{k'}_{t-1}(x) + (c_1 x - c_2)|_{I^t_k}, \qquad \tilde\chi^{k'}_{t-1} = \chi^{k'}_{t-1},\\
\tilde\phi^{k'}_s &= \phi^{k'}_s, \qquad \tilde\chi^{k'}_s = \chi^{k'}_s \quad \text{for } s \notin \{t-1, t\},\\
\tilde H_t &= H_t + c_1|_{X^{-1}_{t-1}(I^t_k)}, \qquad \tilde H_s = H_s \quad \text{for } s \neq t
\end{aligned}$$

where k' runs over all components of the corresponding step in the subscript.

(2) Define

$$\begin{aligned}
\tilde\phi^0_t &= \phi^0_t + \chi^0_t - \chi^0_{t+1}|_{I^t_0}, \qquad \tilde\chi^0_t = 0, \quad \text{and}\\
\tilde\phi^k_t &= \phi^k_t - \chi^0_{t+1}, \qquad \tilde\chi^k_t = \chi^k_t \quad \text{for } k \geq 1,\ t = 0, \dots, n.
\end{aligned}$$

Then $(\tilde{\boldsymbol{\phi}}, \tilde H) \in \mathcal{D}^g_{\boldsymbol{\mu}}(f)$ and $\tilde{\boldsymbol{\chi}}$ is a corresponding concave moderator. Moreover, we have

$$\begin{aligned}
&\sum_{t=0}^{n} \phi^{k_t}_t(x_t) + (H \cdot \boldsymbol{x})_n = \sum_{t=0}^{n} \tilde\phi^{k_t}_t(x_t) + (\tilde H \cdot \boldsymbol{x})_n \quad \text{and}\\
&\phi^k_t + \chi^k_t - \chi_{t+1} = \tilde\phi^k_t + \tilde\chi^k_t - \tilde\chi_{t+1} \quad \text{for all } k \geq 1,\ t = 0, \dots, n,
\end{aligned}$$

as well as $\boldsymbol{\mu}(\boldsymbol{\phi}) = \boldsymbol{\mu}(\tilde{\boldsymbol{\phi}})$.

Proof. (i) If $\boldsymbol{x}$ is such that $(x_{t-1}, x_t) \notin I^t_k \times J^t_k$, then $\tilde\phi^{k_t}_t(x_t) = \phi^{k_t}_t(x_t)$ for $t = 0, \dots, n$ and $\tilde H(\boldsymbol{x}) = H(\boldsymbol{x})$. Otherwise,

$$\begin{aligned}
\tilde\phi^{k_t}_t(x_t) + \tilde\phi^{k_{t-1}}_{t-1}(x_{t-1}) + \tilde H_t(x_t - x_{t-1}) &= \phi^{k_t}_t(x_t) + \phi^{k_{t-1}}_{t-1}(x_{t-1})\\
&\quad + H_t(x_t - x_{t-1}),\\
\tilde\phi^k_t + \tilde\chi^k_t - \tilde\chi_{t+1} &= \phi^k_t + \chi^k_t - \chi_{t+1}, \text{ and}\\
\tilde\phi^{k'}_{t-1} + \tilde\chi^{k'}_{t-1} - \tilde\chi_t &= \phi^{k'}_{t-1} + \chi^{k'}_{t-1} - \chi_t.
\end{aligned}$$

[8] Given an irreducible component (I, J), the notation $(x, y) \in (I, J)$ means that $x \in I, y \in J$, whereas for a diagonal component (I_0, I_0) it is to be understood as $x = y \in I_0$.

Along with the fact that $(\mu_t - \mu_{t-1})^k(\chi_t^k) = (\mu_t - \mu_{t-1})^k(\tilde{\chi}_t^k)$, these identities imply the assertions.

(ii) Similarly as in (i), the terms in question coincide by construction. □

Remark 4.14. The modification of Lemma 4.13(i) can be applied simultaneously for infinitely many k's without difficulties. In this case we set

$$\tilde{\phi}_{t-1}^{k'}(x) := \phi_{t-1}^{k'}(x) + \sum_{k\geq 1}(c_1^k x - c_2^k)|_{I_k^t},$$

as well as $\tilde{\phi}_t^k(y) = \phi_t^k(y) - (c_1^k y - c_2^k)$ and $\tilde{\chi}_t^k(y) = \chi_t^k(y) + (c_1^k y - c_2^k)$ for the components $k \geq 1$ in step t. The pointwise equalities still hold as above and in particular, the moderated integral does not change.

Remark 4.15. Any $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$ induces an element $(\boldsymbol{\phi}^g, H) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f)$ with $\boldsymbol{\mu}(\boldsymbol{\phi}^g) = \boldsymbol{\mu}(\boldsymbol{\phi})$ by choosing some concave moderator $\boldsymbol{\chi}$ for $\boldsymbol{\phi}$ and setting

$$\phi_t^k := \phi_t|_{J_k^t}, \qquad \chi_t^k := \chi_t|_{J_k^t}.$$

We now show the analogue to Lemma 4.8 for the generalized dual space.

Lemma 4.16. *Let $\boldsymbol{\phi} \in L^{c,g}(\boldsymbol{\mu})$ and let $H = (H_1, \dots, H_n)$ be $\mathbb{F}$-predictable. If*

$$\sum_{t=0}^n \phi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t) + (H \cdot \boldsymbol{x})_n$$

is bounded from below on the effective domain $\mathcal{V}$ of $\mathcal{M}(\boldsymbol{\mu})$, then

$$\boldsymbol{\mu}(\phi) = P\left[\sum_{t=0}^n \phi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t) + (H \cdot \boldsymbol{x})_n\right], \quad P \in \mathcal{M}(\boldsymbol{\mu}).$$

Proof. Let $P \in \mathcal{M}(\boldsymbol{\mu})$, let $\boldsymbol{\chi}$ be a concave moderator for $\boldsymbol{\phi}$ such that $\chi_t^0 \equiv 0$ and assume that 0 is the lower bound. It is easy to see that $\sum_{t=0}^n \phi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t) + (H \cdot \boldsymbol{x})_n$ equals

$$\sum_{t=0}^n (\phi_t^{\boldsymbol{k}_t(\boldsymbol{x})} - \chi_{t+1} + \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})})(x_t) + \sum_{t=1}^n (\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_{t-1}) - \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t)) + (H \cdot \boldsymbol{x})_n \geq 0.$$

By assumption $\sum_{t=0}^n (\phi_t^{\boldsymbol{k}_t(\boldsymbol{x})} - \chi_{t+1} + \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})})(x_t)$ is P-integrable. Therefore, the negative part of the remaining expression must also be P-integrable. Writing $P_t := P \circ (X_0, \dots, X_t)^{-1}$ and using that $(\chi_t^k)^+$ has linear growth, we see that for any disintegration $P = P_{n-1} \otimes \kappa_n$,

$$\begin{aligned}
&\int \left[\sum_{t=1}^n (\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_{t-1}) - \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t)) + (H \cdot \boldsymbol{x})_n\right] \kappa_n(x_0, \dots, x_{n-1}, dx_n) \\
&\quad = \sum_{t=1}^{n-1} (\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_{t-1}) - \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t)) + (H \cdot \boldsymbol{x})_{n-1} \\
&\qquad + \int \left[(\chi_n^{\boldsymbol{k}_n(\boldsymbol{x})}(x_{n-1}) - \chi_n^{\boldsymbol{k}_n(\boldsymbol{x})}(x_n))\right] \kappa_n(x_0, \dots, x_{n-1}, dx_n).
\end{aligned}$$

Iteratively integrating with kernels such that $P_t = P_{t-1} \otimes \kappa_t$ and observing that we can apply Fubini's theorem to $\sum_{t=1}^n (\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_{t-1}) - \chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t)) + (H \cdot \boldsymbol{x})_n$ as its negative part is P-integrable,

we obtain

$$P\left[\sum_{t=1}^{n}(\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_{t-1})-\chi_t^{\boldsymbol{k}_t(\boldsymbol{x})}(x_t))+(H\cdot\boldsymbol{x})_n\right]=\sum_{t=1}^{n}\sum_{k\geq 1}(\mu_{t-1}-\mu_t)^k(\chi_t^k)$$

and the result follows. □

Next, we establish that lifting from $\mathcal{D}_{\boldsymbol{\mu}}(f)$ to $\mathcal{D}^g_{\boldsymbol{\mu}}(f)$ does not change the range of dual costs.

Proposition 4.17. *Let $f:\mathbb{R}^{n+1}\to[0,\infty]$. We have*

$$\{\boldsymbol{\mu}(\boldsymbol{\phi}^g):(\boldsymbol{\phi}^g,H)\in\mathcal{D}^g_{\boldsymbol{\mu}}(f)\}=\{\boldsymbol{\mu}(\boldsymbol{\phi}):(\boldsymbol{\phi},H)\in\mathcal{D}_{\boldsymbol{\mu}}(f)\}.$$

Proof. Remark 4.15 shows the inclusion "$\supseteq$". To show the reverse, we may apply Lemma 4.13 (i) together with Remark 4.14 to modify a given pair $(\boldsymbol{\phi}^g,H)\in\mathcal{D}^g_{\boldsymbol{\mu}}(f)$ such that $\phi_t^k(x)=0$ for $x\in J_k^t\backslash I_k^t$, for all irreducible domains (I_k^t,J_k^t) of $\mathcal{M}(\mu_{t-1},\mu_t)$ and $1\leq t\leq n$. Here we have used that $x\in J_k^t\backslash I_k^t$ implies $\mu_k^t(\{x\})>0$, cf. Definition 2.2, and therefore $\phi^g\in L^{c,g}(\boldsymbol{\mu})$ implies $\phi_t^k(x)\in\mathbb{R}$; that is, such endpoints can indeed be shifted to 0 by adding affine functions to ϕ_t^k.

Let $\boldsymbol{\chi}^g$ be a concave moderator for $\boldsymbol{\phi}^g$. Using Lemma 4.3(ii) and again Lemma 4.13 as above, we can modify χ_t^k to satisfy $\chi_t^k(x)=0$ for $x\in J_k^t\backslash I_k^t$, for all irreducible domains (I_k^t,J_k^t) of $\mathcal{M}(\mu_{t-1},\mu_t)$ and $1\leq t\leq n$. Here, the finiteness of χ_t^k at the endpoints follows from Lemma 4.3(i) and $(\mu_{t-1}-\mu_t)^k(\chi_t^k)<\infty$.

Still denoting the modified dual element by $(\boldsymbol{\phi}^g,H)$, we define $\boldsymbol{\phi}\in L^c(\boldsymbol{\mu})$ and a corresponding concave moderator $\boldsymbol{\chi}$ by

$$\phi_t(x):=\phi_t^k(x),\quad \chi_t(x):=\chi_t^k(x),\quad \text{for } x\in J_k^t;$$

they are well-defined since ϕ_t^k and χ_t^k vanish at points that belong to more than one set J_k^t. We have $\boldsymbol{\mu}(\boldsymbol{\phi})=\boldsymbol{\mu}(\boldsymbol{\phi}^g)$ by construction and the result follows. □

Definition 4.18. Let $1\leq t\leq n$ and $x_t\in\mathbb{R}$. A sequence $\boldsymbol{x}=(x_0,\dots,x_t)$ is a *predecessor path* of x_t if there are indices $(k_0,\dots,k_t)$ such that $(x_{s-1},x_s)\in(I_{k_s}^s,J_{k_s}^s)$ for some component (irreducible or diagonal) of $\mathcal{M}(\mu_{s-1},\mu_s)$, for all $1\leq s\leq t$. We write $\Bbbk(\boldsymbol{x})$ for the (unique) associated sequence $(k_0,\dots,k_t)$ followed by the path $\boldsymbol{x}$ in the above sense, and $\Psi_t^k(x_t)$ for the set of all predecessor paths with $k_t=k$.

These notions will be useful in the next step towards the closedness result, which is to "regularize" the concave moderators. For concreteness in some of the expressions below, we convene that $\infty-\infty:=\infty$.

Lemma 4.19. *Let $(\boldsymbol{\phi},H)\in\mathcal{D}^g_{\boldsymbol{\mu}}(0)$. There is a concave moderator $\boldsymbol{\chi}$ of $\boldsymbol{\phi}$ such that*

$$\phi_t^k+\chi_t^k-\chi_{t+1}\geq 0\quad \text{on } J_k^t\quad \text{for all}\quad t=0,\dots,n,\quad k\geq 1,\ \text{and} \tag{4.2}$$

$$\phi_t^0+\chi_t^0-\chi_{t+1}\geq 0\quad \mu_t\text{-a.s. on } I_0^t\quad \text{for all}\quad t=1,\dots,n. \tag{4.3}$$

As a consequence,

$$\sum_{t=1}^{n}\sum_{k\geq 1}(\mu_{t-1}-\mu_t)^k(\chi_t^k)\leq\boldsymbol{\mu}(\boldsymbol{\phi}).$$

Proof. Fix $1 \le t \le n$ and let (I_k^t, J_k^t) be the domain of some component of $\mathcal{M}(\mu_{t-1}, \mu_t)$. We define $\boldsymbol{\chi} = (\chi_t^k)$ by $\chi_0^0 = 0$ and

$$\chi_t^k(x_t) = \inf_{\boldsymbol{x} \in \Psi_t^k(x_t)} \left\{ \sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\boldsymbol{x})}(x_s) + (H \cdot \boldsymbol{x})_t \right\};$$

then χ_t^k is concave on J_k^t for $k \ge 1$ as an infimum of affine functions.

We first show that

$$\{\chi_t^k = +\infty\} \subseteq \left\{\phi_{t-1}^{k'} = +\infty\right\} \cup \{\chi_{t-1}^{k'} = +\infty\}.$$

In particular, such points only exist after a chain of diagonal components from a point where $\phi_t^k(x_t) = \infty$. Suppose $\chi_t^k(x_t) = +\infty$ and $k \ge 1$, then the predecessor paths of x_t agree with the predecessor paths of all of J_t^k up to $t-1$, but $\{\sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\boldsymbol{x})}(x_s) < \infty\}$ must hold $\mathcal{M}(\boldsymbol{\mu})$-q.s. as $\boldsymbol{\phi} \in L^{c,g}(\boldsymbol{\mu})$. We must therefore have $x_t \in I_t^0$. Then, by definition, $\chi_t^0(x_t) = \chi_{t-1}^k(x_t) + \phi_{t-1}^k(x_t)$ and the claim follows.

Next, we verify that $\boldsymbol{\chi}$ satisfies (4.2) and (4.3). For notational convenience we for now set $\chi_{n+1} \equiv \inf_{\boldsymbol{x} \in \mathcal{V}} \left\{ \sum_{s=0}^{n} \phi_s^{\mathbb{k}_s(\boldsymbol{x})}(x_s) + (H \cdot \boldsymbol{x})_n \right\} \ge 0$. Restricting the infimum in the definition of $\boldsymbol{\chi}$ to the set of paths $\boldsymbol{x}$ with $x_{t+1} = x_t \in I_{k'}^{t+1} \cap J_k^t$ yields

$$\begin{aligned}
\chi_{t+1}(x_t) = \chi_{t+1}^{k'}(x_t) &= \inf_{\boldsymbol{x} \in \Psi_{t+1}^{k'}(x_t)} \left\{ \sum_{s=0}^{t} \phi_s^{\mathbb{k}_s(\boldsymbol{x})}(x_s) + (H \cdot \boldsymbol{x})_{t+1} \right\} \\
&\le \inf_{\boldsymbol{x} \in \Psi_t^k(x_t)} \left\{ \sum_{s=0}^{t-1} \phi_s^{\mathbb{k}_s(\boldsymbol{x})}(x_s) + (H \cdot \boldsymbol{x})_t \right\} + \phi_t^k(x_t) \\
&= \chi_t^k(x_t) + \phi_t^k(x_t).
\end{aligned}$$

Since $\cup_{k' \ge 0} I_{k'}^{t+1} = \mathbb{R}$, this will imply (4.2) after we check that $\chi_t^k > -\infty$ for $k \ge 1$ and $\chi_t^0 > -\infty$ holds μ_t^0-a.s., which also implies that $\chi_t > -\infty$ holds μ_{t-1}-almost surely. We show this inductively for $t \ge 1$.

Clearly $\chi_{n+1} \ge 0 > -\infty$. Now, for $t \le n$ the induction hypothesis is that $\chi_{t+1} > -\infty$ holds almost surely μ_t.

From $\boldsymbol{\phi} \in L^{c,g}$ and $\chi_{t+1} > -\infty$ μ_t-a.s. we have that

$$\phi_t^k < \infty, \quad \chi_{t+1} > -\infty \quad \text{hold } \mu_t^k\text{-a.s.}$$

As χ_t^k is concave and J_t^k is the convex hull of the topological support of μ_t^k we then get $\chi_t^k > -\infty$ on all of J_t^k from the previous inequality.

For $k = 0$, the inequality yields $\{\chi_t^0 = -\infty\} \subseteq \{\chi_{t+1} = -\infty\} \cup \{\phi_t^0(x_t) = \infty\}$ and both of these sets are μ_t nullsets. Finally $\mu_{t-1}(\{\chi_t = -\infty\}) = 0$ as this is a subset of the diagonal component where μ_{t-1} is dominated by μ_t.

Set $\bar{\phi}_t^k := \phi_t^k + \chi_t^k - \chi_{t+1}|_{J_k^t}$ for $0 \le t \le n$; then $\bar{\phi}_t^k \ge 0$. Moreover, choose an arbitrary $P \in \mathcal{M}(\boldsymbol{\mu})$ with disintegration $P = \mu_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n$ for some stochastic kernels $\kappa_t(x_0, \ldots, x_{t-1}, dx_t)$. From Lemma 4.16 we know that

$$\boldsymbol{\mu}(\boldsymbol{\phi}) = P\left[\sum_{t=0}^{n} \phi_t^{\mathbb{k}_t(X)}(X_t) + (H \cdot X)_n \right] < \infty.$$

We can therefore apply Fubini's theorem for kernels as in the proof of Lemma 4.16 to the expression

$$\begin{aligned} 0 &\leq \sum_{t=0}^{n} \phi_t^{\mathbb{k}_t(\boldsymbol{x})}(x_t) + (H \cdot \boldsymbol{x})_n \\ &= \sum_{t=0}^{n} \bar{\phi}_t^{\mathbb{k}_t(\boldsymbol{x})}(x_t) + \sum_{t=1}^{n} \left(\chi_t(x_{t-1}) - \chi_t^{\mathbb{k}_t(\boldsymbol{x})}(x_t) \right) + (H \cdot \boldsymbol{x})_n \end{aligned}$$

and obtain

$$P\left[\sum_{t=0}^{n} \phi_t^{\mathbb{k}_t(X)}(X_t) + (H \cdot X)_n \right] = \sum_{t=0}^{n} \sum_{k \geq 0} \mu_t^k(\bar{\phi}_t^k) + \sum_{t=1}^{n} \sum_{k \geq 1} (\mu_{t-1} - \mu_t)^k(\chi_t^k)$$

which shows that the right hand side is finite, and therefore $\boldsymbol{\chi}$ is a concave moderator for $\boldsymbol{\phi}$. Finally, the second claim follows from $\mu_t^k(\bar{\phi}_t^k) \geq 0$. □

The last tool for our closedness result is a compactness property for concave functions in the one-step case; cf. [10, Proposition 5.5].

Proposition 4.20. *Let $\mu \leq_c \nu$ be irreducible with domain (I, J) and let $a \in I$ be the common barycenter of μ and ν. Let $\chi_m : J \to \mathbb{R}$ be concave functions such that*[9]

$$\chi_m(a) = \chi_m'(a) = 0 \quad \text{and} \quad \sup_{m \geq 1} (\mu - \nu)(\chi_m) < \infty.$$

There exists a subsequence χ_{m_k} which converges pointwise on J to a concave function $\chi : J \to \mathbb{R}$, and $(\mu - \nu)(\chi) \leq \liminf_k (\mu - \nu)(\chi_{m_k})$.

We are now ready to state and prove the analogue of Proposition 4.10 in the generalized dual.

Proposition 4.21. *Let $f^m : \mathbb{R}^{n+1} \to [0, \infty]$, $m \geq 1$ be a sequence of functions such that*

$$f^m \to f \qquad \text{pointwise}$$

and let $(\boldsymbol{\phi}^m, H^m) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f^m)$ be such that $\sup_m \boldsymbol{\mu}(\boldsymbol{\phi}^m) < \infty$. Then there exist $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f)$ with

$$\boldsymbol{\mu}(\boldsymbol{\phi}) \leq \liminf_{m \to \infty} \boldsymbol{\mu}(\boldsymbol{\phi}^m).$$

Proof. Since $(\boldsymbol{\phi}^m, H^m) \in \mathcal{D}_{\boldsymbol{\mu}}^g(f^m)$ and $f^m \geq 0$, we can introduce a sequence of concave moderators $\boldsymbol{\chi}_m$ as in Lemma 4.19. A normalization of $(\boldsymbol{\phi}^m, H^m)$ as in Lemma 4.13(i) and (ii), in the general form of Remark 4.14, allows us to assume without loss of generality that $\chi_{t,m}^0 \equiv 0$ and $\chi_{t,m}^k(a_t^k) = (\chi_{t,m}^k)'(a_t^k) = 0$, where a_t^k is the barycenter of μ_t^k—this modification is the main merit of lifting to the generalized dual space. While the generalized dual gives enough degrees of freedom to choose this normalization, the dual without the generalization does not. This is related to the possible overlap of the intervals I, J at the different times t; see also Fig. 2 and the paragraph preceding Example 3.2.

By passing to a subsequence as in Proposition 4.20 for each component and using a diagonal argument, we obtain pointwise limits $\chi_t^k : J_k^t \to \mathbb{R}$ for $\chi_{t,m}^k$ after passing to another subsequence.

[9] To be specific, let us convene that χ_m' is the left derivative—this is not important here.

Since $\phi^k_{t,m} + \chi^k_{t,m} - \chi_{t+1,m} \geq 0$ on J^t_k[10] and $\chi^k_{t,m} \to \chi^k_t$ as well as $\chi_{t+1,m} \to \chi_{t+1}$, we can apply Komlos' lemma (in the form of [16, Lemma A1.1] and its remark) to find convex combinations $\tilde{\phi}^k_{t,m} \in \operatorname{conv}\{\phi^k_{t,m}, \phi^k_{t,m+1}, \dots\}$ which converge μ^k_t-a.s. for $0 \leq t \leq n$. We may assume without loss of generality that $\tilde{\phi}^k_{t,m} = \phi^k_{t,m}$. Thus, we can set

$$\begin{aligned}\phi^k_t &:= \limsup \phi^k_{t,m} \quad \text{on } J^t_k \quad \text{for } t = 1, \dots, n,\\ \phi_0 &:= \liminf \phi_{0,m}\end{aligned}$$

to obtain

$$\phi^k_{t,m} \to \phi^k_t \quad \mu^k_t\text{-a.s.} \quad \text{and} \quad \phi^k_t + \chi^k_t - \chi_{t+1} \geq 0 \text{ on } J^t_k.$$

We can now apply Fatou's lemma and Proposition 4.20 to deduce that

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{\phi}) &= \sum_{t=0}^{n} \sum_{k\geq 0} \mu^k_t(\phi^k_t + \chi^k_t - \chi_{t+1}) + \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi^k_t)\\ &\leq \sum_{t=0}^{n} \sum_{k\geq 0} \liminf \mu^k_t(\phi^k_{t,m} + \chi^k_{t,m} - \chi_{t+1,m})\\ &\quad + \sum_{t=1}^{n} \sum_{k\geq 1} \liminf (\mu_{t-1} - \mu_t)^k(\chi^k_{t,m})\\ &\leq \liminf \left[\sum_{t=0}^{n} \sum_{k\geq 0} \mu^k_t(\phi^k_{t,m} + \chi^k_{t,m} - \chi_{t+1,m}) + \sum_{t=1}^{n} \sum_{k\geq 1} (\mu_{t-1} - \mu_t)^k(\chi^k_{t,m}) \right]\\ &= \liminf \boldsymbol{\mu}(\boldsymbol{\phi}^m) < \infty.\end{aligned}$$

In particular, we see that $\boldsymbol{\phi} \in L^{c,g}(\boldsymbol{\mu})$ with concave moderator $\boldsymbol{\chi}$.

It remains to construct the predictable process $H = (H_1, \dots, H_n)$. With a mild abuse of notation, we shall identify $H_t(x_0, \dots, x_n)$ with the corresponding function of $(x_0, \dots, x_{t-1})$ in this proof.

We first define for each $\boldsymbol{k} = (k_0, \dots, k_t)$ and $\boldsymbol{x} = (x_0, \dots, x_t)$ such that $\boldsymbol{k} = \Bbbk(\boldsymbol{x})$, the functions $G^{\boldsymbol{k}}_{t,m}$ and $G^{\boldsymbol{k}}_t$ by

$$\begin{aligned}G^{\boldsymbol{k}}_{t,m}(\boldsymbol{x}) &:= \sum_{s=0}^{t} \phi^{k_s}_{s,m}(x_s) + \sum_{s=1}^{t} H_{s,m}(x_0, \dots, x_{s-1}) \cdot (x_s - x_{s-1}),\\ G^{\boldsymbol{k}}_t(\boldsymbol{x}) &:= \liminf G^{\boldsymbol{k}}_{t,m}(\boldsymbol{x}).\end{aligned}$$

Given $\boldsymbol{k} = (k_0, \dots, k_t)$, we write $\boldsymbol{k}' = (k_0, \dots, k_{t-1})$. We claim that there exists an $\mathbb{F}$-predictable process H such that for all $1 \leq t \leq n$,

$$G^{\boldsymbol{k}'}_{t-1}(x_0, \dots, x_{t-1}) + \phi^{k_t}_t(x_t) + H_t(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1}) \geq G^{\boldsymbol{k}}_t(x_0, \dots, x_t). \tag{4.4}$$

Once this is established, the proposition follows by induction since $G^{(0)}_0(x_0) = \phi_0(x_0)$ and $G^{\boldsymbol{k}}_n(x_0, \dots, x_n) \geq f(x_0, \dots, x_n)$.

To prove the claim, write g^{conc} for the concave hull of a function g and observe that

$$\begin{aligned}&\liminf [G^{\boldsymbol{k}'}_{t-1,m}(x_0, \dots, x_{t-1}) + H_{t,m}(x_0, \dots, x_{t-1}) \cdot (x_t - x_{t-1})]\\ &\qquad \geq \liminf [(G^{\boldsymbol{k}}_{t,m}(x_0, \dots, x_{t-1}, \cdot) - \phi^{k_t}_{t,m}(\cdot))^{\mathrm{conc}}(x_t)]\end{aligned}$$

[10] Observe that this inequality will still hold after modifying $\boldsymbol{\phi}$ and $\boldsymbol{\chi}$ as in Lemma 4.13.

$$\begin{aligned}
&\geq [\liminf(G^{\boldsymbol{k}}_{t,m}(x_0,\dots,x_{t-1},\cdot) - \phi^{k_t}_{t,m}(\cdot))]^{\mathrm{conc}}(x_t)\\
&\geq [G^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},\cdot) - \phi^{k_t}_t(\cdot)]^{\mathrm{conc}}(x_t)\\
&=: \hat\phi^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},x_t).
\end{aligned}$$

By construction, $\hat\phi^{\boldsymbol{k}}_t$ is concave in the last variable and satisfies

$$G^{\boldsymbol{k}'}_{t-1}(x_0,\dots,x_{t-1}) \geq \hat\phi^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},x_{t-1}).$$

Let $\partial_t\hat\phi^{\boldsymbol{k}}_t$ denote the left partial derivative in the last variable and set

$$H^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1}) := \partial_t\hat\phi^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},x_{t-1})$$

for $\boldsymbol{k}_t \geq 1$ and $H^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1}) = 0$ for $\boldsymbol{k}_t = 0$; then we have

$$\begin{aligned}
&G^{\boldsymbol{k}'}_{t-1}(x_0,\dots,x_{t-1}) + H^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1})\cdot(x_t - x_{t-1})\\
&\quad\geq \hat\phi^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},x_{t-1}) + H^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1})\cdot(x_t - x_{t-1})\\
&\quad\geq \hat\phi^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1},x_t)\\
&\quad\geq G^{\boldsymbol{k}}_t(x_0,\dots,x_t) - \phi^{k_t}_t(x_t).
\end{aligned}$$

Finally, for any $(x_0,\dots,x_{t-1}) \in \mathbb{R}^t$, we define $H_t(x_0,\dots,x_{t-1})$ as

$$\begin{cases} H^{\boldsymbol{k}}_t(x_0,\dots,x_{t-1}), & \text{if } \boldsymbol{k} = \mathbb{k}(x_0,\dots,x_{t-1},x_t) \text{ for some } x_t \in \mathbb{R}\\ 0, & \text{otherwise;}\end{cases}$$

this is well-defined since $\mathbb{k}(x_0,\dots,x_t)$ depends only on $(x_0,\dots,x_{t-1})$. The predictable process H satisfies (4.4) and thus the proof is complete. □

Proof of Proposition 4.10. In view of Remark 4.15 and Proposition 4.17, the result follows from Proposition 4.21. □

5. Duality theorem and monotonicity principle

The first goal of this section is a duality result for the multistep martingale transport problem; it establishes the absence of a duality gap and the existence of optimizers in the dual problem. (As is well known, an optimizer for the primal problem only exists under additional conditions, such as continuity of f.) The second goal is a monotonicity principle describing the geometry of optimal transports; it will be a consequence of the duality result.

As above, we consider a fixed vector $\boldsymbol{\mu} = (\mu_0,\dots,\mu_n)$ of marginals in convex order. The primal and dual problems as defined follows.

Definition 5.1. Let $f : \mathbb{R}^{n+1} \to [0,\infty]$. The *primal problem* is

$$\mathbf{S}_{\boldsymbol{\mu}}(f) := \sup_{P\in\mathcal{M}(\boldsymbol{\mu})} P(f) \in [0,\infty],$$

where $P(f)$ refers to the outer integral if f is not measurable. The *dual problem* is

$$\mathbf{I}_{\boldsymbol{\mu}}(f) := \inf_{(\boldsymbol{\phi},H)\in\mathcal{D}_{\boldsymbol{\mu}}(f)} \boldsymbol{\mu}(\boldsymbol{\phi}) \in [0,\infty].$$

We recall that a function $f : \mathbb{R}^{n+1} \to [0,\infty]$ is called *upper semianalytic* if the sets $\{f \geq c\}$ are analytic for all $c \in \mathbb{R}$, where a subset of $\mathbb{R}^{n+1}$ is called analytic if it is the image of a Borel subset of a Polish space under a Borel mapping. Any Borel function is upper semianalytic

and any upper semianalytic function is universally measurable; we refer to [11, Chapter 7] for background. The following is the announced duality result.

Theorem 5.2 (*Duality*). *Let* $f : \mathbb{R}^{n+1} \to [0, \infty]$.

(i) If f *is upper semianalytic, then* $\mathbf{S}_{\boldsymbol{\mu}}(f) = \mathbf{I}_{\boldsymbol{\mu}}(f) \in [0, \infty]$.
(ii) If $\mathbf{I}_{\boldsymbol{\mu}}(f) < \infty$, *there exists a dual optimizer* $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$.

Proof. Given our preceding results, much of the proof follows the lines of the corresponding result for the one-step case in [10, Theorem 6.2]; therefore, we shall be brief. We mention that the present theorem is slightly more general than the cited one in terms of the measurability condition (f is upper semianalytic instead of Borel); this is due to the global proof given here.

Step 1. Using Lemma 4.8 we see that $\mathbf{S}_{\boldsymbol{\mu}}(f) \leq \mathbf{I}_{\boldsymbol{\mu}}(f)$ holds for all upper semicontinuous $f : \mathbb{R}^{n+1} \to [0, \infty]$.

Step 2. Using the de la Vallée–Poussin theorem and our assumption that the marginals have a finite first moment, there exist increasing, superlinearly growing functions $\zeta_{\mu_t} : \mathbb{R}^+ \to \mathbb{R}^+$ such that $x \mapsto \zeta_{\mu_t}(|x|)$ is μ_t-integrable for all $0 \leq t \leq n$. Define

$$\zeta(x_0, \dots, x_n) := 1 + \sum_{t=0}^{n} \zeta_{\mu_t}(|x_t|)$$

and let C_ζ be the vector space of all continuous functions f such that f/ζ vanishes at infinity. Then, a Hahn–Banach separation argument can be used to show that $\mathbf{S}_{\boldsymbol{\mu}}(f) \geq \mathbf{I}_{\boldsymbol{\mu}}(f)$ holds for all $f \in C_\zeta$; the details of the argument are the same as in the proof of [10, Lemma 6.4].

Step 3. Let f be bounded and upper semicontinuous; then there exists a sequence of bounded continuous functions $f^m \in C_b(\mathbb{R}^{n+1})$ which decrease to f pointwise. As $C_b(\mathbb{R}^{n+1}) \subseteq C_\zeta$, we have $\mathbf{S}_{\boldsymbol{\mu}}(f^m) = \mathbf{I}_{\boldsymbol{\mu}}(f^m)$ for all m by the first two steps.

Let $\mathcal{U}$ be the set of all bounded, nonnegative, upper semicontinuous functions on $\mathbb{R}^{n+1}$. We recall that a map $\mathbf{C} : [0, \infty]^{\mathbb{R}^{n+1}} \to [0, \infty]$ is called a $\mathcal{U}$-capacity if it is monotone, sequentially continuous upwards on $[0, \infty]^{\mathbb{R}^{n+1}}$ and sequentially continuous downwards on $\mathcal{U}$. The functional $f \mapsto \mathbf{S}_{\boldsymbol{\mu}}(f)$ is a $\mathcal{U}$-capacity; this follows from the weak compactness of $\mathcal{M}(\boldsymbol{\mu})$ and the arguments in [35, Propositions 1.21, 1.26].

It follows that $\mathbf{S}_{\boldsymbol{\mu}}(f^m) \to \mathbf{S}_{\boldsymbol{\mu}}(f)$. By the monotonicity of $f \mapsto \mathbf{I}_{\boldsymbol{\mu}}(f)$ and Step 1 we obtain

$$\mathbf{I}_{\boldsymbol{\mu}}(f) \leq \lim \mathbf{I}_{\boldsymbol{\mu}}(f^m) = \lim \mathbf{S}_{\boldsymbol{\mu}}(f^m) = \mathbf{S}_{\boldsymbol{\mu}}(f) \leq \mathbf{I}_{\boldsymbol{\mu}}(f).$$

Step 4. Since $\mathbf{S}_{\boldsymbol{\mu}} = \mathbf{I}_{\boldsymbol{\mu}}$ on $\mathcal{U}$ by Step 3, $\mathbf{I}_{\boldsymbol{\mu}}$ is sequentially downward continuous on $\mathcal{U}$ like $\mathbf{S}_{\boldsymbol{\mu}}$. On the other hand, Proposition 4.10 implies that it is sequentially upwards continuous on $[0, \infty]^{\mathbb{R}^{n+1}}$. As a result, $\mathbf{I}_{\boldsymbol{\mu}}$ is a $\mathcal{U}$-capacity.

Step 5. Let $f : \mathbb{R}^{n+1} \to [0, \infty]$ be upper semianalytic. For any $\mathcal{U}$-capacity $\mathbf{C}$, Choquet's capacitability theorem shows that

$$\mathbf{C}(f) = \sup\{\mathbf{C}(g) : g \in \mathcal{U}, g \leq f\}.$$

As $\mathbf{S}_{\boldsymbol{\mu}}$ and $\mathbf{I}_{\boldsymbol{\mu}}$ are $\mathcal{U}$-capacities that coincide on $\mathcal{U}$, it follows that $\mathbf{S}_{\boldsymbol{\mu}}(f) = \mathbf{I}_{\boldsymbol{\mu}}(f)$. This completes the proof of (i).

Step 6. To see that the infimum $\mathbf{I}_{\boldsymbol{\mu}}(f)$ is attained if it is finite, we merely need to apply Proposition 4.10 with the constant sequence $f^m = f$. □

We can easily relax the lower bound on f.

Remark 5.3. Let $f : \mathbb{R}^{n+1} \to (-\infty, \infty]$ and suppose there exist $\boldsymbol{\phi} \in \prod_{t=0}^{n} L^1(\mu_t)$ and a predictable process H such that

$$f \geq \sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n \quad \text{on} \quad \mathcal{V}.$$

Then we can apply Theorem 5.2 to $\big[f - \sum_{t=0}^{n} \phi_t(X_t) - (H \cdot X)_n\big]^+$ and obtain the analogue of its assertion for f.

The duality result gives rise to a monotonicity principle describing the support of optimal martingale transports, in the spirit of the cyclical monotonicity condition from classical transport theory. The following generalizes the results of [8, Lemma 1.11] and [10, Corollary 7.8] for the one-step martingale transport problem.

Theorem 5.4 (*Monotonicity Principle*)**.** *Let $f : \mathbb{R}^{n+1} \to [0, \infty]$ be Borel and suppose that $\mathbf{S}_{\boldsymbol{\mu}}(f) < \infty$. There exists a Borel set $\Gamma \subseteq \mathbb{R}^{n+1}$ with the following properties.*

(i) A measure $P \in \mathcal{M}(\boldsymbol{\mu})$ is concentrated on Γ if and only if it is optimal for $\mathbf{S}_{\boldsymbol{\mu}}(f)$.
(ii) Let $\bar{\boldsymbol{\mu}} = (\bar{\mu}_0, \dots, \bar{\mu}_n)$ be another vector of marginals in convex order. If $\bar{P} \in \mathcal{M}(\bar{\boldsymbol{\mu}})$ is concentrated on Γ, then $\bar{P}$ is optimal for $\mathbf{S}_{\bar{\boldsymbol{\mu}}}(f)$.

Indeed, if $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$ is an optimizer for $\mathbf{I}_{\boldsymbol{\mu}}(f)$, then we can take

$$\Gamma := \left\{ \sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n = f \right\} \cap \mathcal{V}.$$

Proof. As $\mathbf{S}_{\boldsymbol{\mu}}(f) < \infty$, Theorem 5.2 shows that $\mathbf{I}_{\boldsymbol{\mu}}(f) = \mathbf{S}_{\boldsymbol{\mu}}(f) < \infty$ and that there exists a dual optimizer $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$. In particular, we can define Γ as above.

(i) As $0 \leq f$ and $P(f) \leq \mathbf{S}_{\boldsymbol{\mu}}(f) < \infty$ for all $P \in \mathcal{M}(\boldsymbol{\mu})$, we see that f is P-integrable for all $P \in \mathcal{M}(\boldsymbol{\mu})$. Since $\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n \geq 0$ on the effective domain $\mathcal{V}$, and $P\big[\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n\big] = \boldsymbol{\mu}(\boldsymbol{\phi}) = \mathbf{I}_{\boldsymbol{\mu}}(f) < \infty$ by Lemma 4.8, we also obtain the P-integrability of $\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n$. In particular,

$$0 \leq P\left[\sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n - f \right] = \boldsymbol{\mu}(\boldsymbol{\phi}) - P(f) = \mathbf{S}_{\boldsymbol{\mu}}(f) - P(f)$$

and equality holds if and only if P is concentrated on Γ.

(ii) We may assume that $\bar{P}$ is a probability measure with $\bar{P}(f) < \infty$. As a first step, we show that the effective domain $\bar{\mathcal{V}}$ of $\mathcal{M}(\bar{\boldsymbol{\mu}})$ is a subset of the effective domain $\mathcal{V}$ of $\mathcal{M}(\boldsymbol{\mu})$. To that end, it is sufficient to show that if $1 \leq t \leq n$ and $x \in \mathbb{R}$ are such that $u_{\mu_{t-1}}(x) = u_{\mu_t}(x)$, then $u_{\bar{\mu}_{t-1}}(x) = u_{\bar{\mu}_t}(x)$, and if moreover $\partial^+ u_{\mu_{t-1}}(x) = \partial^+ u_{\mu_t}(x)$, then $\partial^+ u_{\bar{\mu}_{t-1}}(x) = \partial^+ u_{\bar{\mu}_t}(x)$, and similarly for the left derivative ∂^- (cf. Proposition 2.3). Indeed, for t and x such that $u_{\mu_{t-1}}(x) = u_{\mu_t}(x)$, our assumption that $\Gamma \subseteq \mathcal{V}$ implies

$$\Gamma \subseteq (X_{t-1}, X_t)^{-1}\big((-\infty, x]^2 \cup [x, \infty)^2\big).$$

Using also that $\mathbb{E}^{\bar{P}}[X_t | \mathfrak{F}_{t-1}] = X_{t-1}$ and that $\bar{P}$ is concentrated on Γ,

$$\begin{aligned} u_{\bar{\mu}_{t-1}}(x) &= \mathbb{E}^{\bar{P}}[|X_{t-1} - x|] \\ &= \mathbb{E}^{\bar{P}}[(X_{t-1} - x)\mathbf{1}_{X_{t-1} \geq x}] + \mathbb{E}^{\bar{P}}[(x - X_{t-1})\mathbf{1}_{X_{t-1} \leq x}] \end{aligned}$$

$$= \mathbb{E}^{\bar{P}}[(X_t - x)\mathbf{1}_{X_{t-1}\geq x}] + \mathbb{E}^{\bar{P}}[(x - X_t)\mathbf{1}_{X_{t-1}\leq x}]$$
$$= \mathbb{E}^{\bar{P}}[|X_t - x|] = u_{\bar{\mu}_t}(x)$$

as desired. If in addition $\partial^+ u_{\mu_{t-1}}(x) = \partial^+ u_{\mu_t}(x)$, then $\Gamma \subseteq \mathcal{V}$ implies

$$\Gamma \subseteq (X_{t-1}, X_t)^{-1}\big((-\infty, x]^2 \cup (x, \infty)^2\big).$$

As $\bar{P}$ is concentrated on Γ, it follows that

$$\partial^+ u_{\bar{\mu}_{t-1}}(x) = \bar{P}[X_{t-1} \leq x] - \bar{P}[X_{t-1} > x]$$
$$= \bar{P}[X_t \leq x] - \bar{P}[X_t > x] = \partial^+ u_{\bar{\mu}_t}(x)$$

as desired. The same argument can be used for the left derivative and we have shown that $\bar{\mathcal{V}} \subseteq \mathcal{V}$.

In view of that inclusion, the inequality $\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n \geq f$ holds on $\bar{\mathcal{V}}$. Since $\bar{P}$ is concentrated on Γ,

$$\bar{P}\left[\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n\right] = \bar{P}(f) < \infty.$$

We may follow the arguments in the proof of Lemma 4.19 to construct a moderator $\boldsymbol{\chi}$ and establish that $(\boldsymbol{\phi}, H) \in \mathcal{D}^g_{\bar{\boldsymbol{\mu}}}(f)$, where we are implicitly using the embedding detailed in Remark 4.15. (Note that the proof of Lemma 4.19 uses the condition $(\boldsymbol{\phi}, H) \in \mathcal{D}^g_{\bar{\boldsymbol{\mu}}}(0)$ only to establish $\bar{P}\left[\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n\right] < \infty$. In the present situation the latter is known a priori and the condition is not needed.) Then, we can modify $\boldsymbol{\chi}$ as in the proof of Proposition 4.17 to see that $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\bar{\boldsymbol{\mu}}}(f)$. As a result, we may apply Lemma 4.8 to obtain that

$$\bar{P}(f) = \bar{P}\left[\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n\right] = \bar{\boldsymbol{\mu}}(\boldsymbol{\phi}),$$

whereas for any other $P' \in \mathcal{M}(\bar{\boldsymbol{\mu}})$ we have

$$P'(f) \leq P'\left[\sum_{t=0}^n \phi_t(X_t) + (H \cdot X)_n\right] = \bar{\boldsymbol{\mu}}(\boldsymbol{\phi}) = \bar{P}(f).$$

This shows that $\bar{P} \in \mathcal{M}(\bar{\boldsymbol{\mu}})$ is optimal. □

6. Left-monotone transports

In this section we define left-monotone transports through a shadow property and prove their existence.

6.1. Preliminaries

Before moving on to the n-step case, we recall the essential definitions and results regarding the one-step version of the left-monotone transport (also called the Left-Curtain coupling). The first notion is the so-called shadow, and it will be useful to define it for measures $\mu \leq_{pc} \nu$ in *positive convex order*, meaning that $\mu(\phi) \leq \nu(\phi)$ for any nonnegative convex function ϕ. Clearly, this order is weaker than the convex order $\mu \leq_c \nu$, and it is worth noting that μ may have a smaller mass than ν. The following is the result of [8, Lemma 4.6].

Lemma 6.1. *Let $\mu \leq_{pc} \nu$. Then the set*

$$[\![\mu; \nu]\!] := \{\theta : \mu \leq_c \theta \leq \nu\}$$

is non-empty and contains a unique least element $\mathcal{S}^\nu(\mu)$ for the convex order:

$$\mathcal{S}^\nu(\mu) \leq_c \theta \text{ for all } \theta \in [\![\mu; \nu]\!].$$

The measure $\mathcal{S}^\nu(\mu)$ is called the shadow *of μ in ν.*

It will be useful to have the following picture in mind: if μ is a Dirac measure, its shadow in ν is a measure θ of equal mass and barycenter, chosen such as to have minimal variance subject to the constraint $\theta \leq \nu$.

The second notion is a class of reward functions.

Definition 6.2. A Borel function $f : \mathbb{R}^2 \to \mathbb{R}$ is called *second-order Spence–Mirrlees* if $y \mapsto f(x', y) - f(x, y)$ is strictly convex for any $x < x'$.

We note that if f is sufficiently differentiable, this can be expressed as the cross-derivative condition $f_{xyy} > 0$ which has also been called the martingale Spence–Mirrlees condition, in analogy to the classical Spence–Mirrlees condition $f_{xy} > 0$.

In the one-step case, the left-monotone transport is unique and can be characterized as follows; cf. [8, Theorems 4.18, 4.21, 6.1] where this transport is called the Left-Curtain coupling, as well as [37, Theorem 1.2] for the third equivalence in the stated generality.

Proposition 6.3. *Let $\mu \leq_c \nu$ and $P \in \mathcal{M}(\mu, \nu)$. The following are equivalent:*

(i) For all $x \in \mathbb{R}$ and $A \in \mathfrak{B}(\mathbb{R})$,

$$P[(-\infty, x] \times A] = \mathcal{S}^\nu(\mu|_{(-\infty,x]})(A).$$

(ii) P is concentrated on a Borel set $\Gamma \subseteq \mathbb{R}^2$ satisfying

$$(x, y^-), (x, y^+), (x', y') \in \Gamma, \quad x < x' \quad \Rightarrow \quad y' \notin (y^-, y^+).$$

(iii) P is an optimizer of $\mathbf{S}_{\mu,\nu}(f)$ for some (and then all) $f : \mathbb{R}^2 \to \mathbb{R}$ second-order Spence–Mirrlees such that there exist functions $a \in L^1(\mu)$, $b \in L^1(\nu)$ with $|f(x, y)| \leq a(x) + b(y)$.

There exists a unique measure $\bar{P} \in \mathcal{M}(\mu, \nu)$ satisfying (i)–(iii), and $\bar{P}$ is called the (one-step) left-monotone transport.

If μ is a discrete measure, the characterization in (i) can be understood as follows: the left-monotone transport $\bar{P}$ processes the atoms of μ from left to right, mapping each one of them to its shadow in the remaining target measure.

Next, we record two more results about shadows that will be used below. The first one, cited from [9, Theorem 3.1], generalizes the above idea in the sense that the atoms are still mapped to their shadows but can be processed in any given order; in the general (non-discrete) case, such an order is defined by a coupling π from the uniform measure to μ.

Proposition 6.4. *Let $\mu \leq_c \nu$ and $\pi \in \Pi(\lambda, \mu)$ where λ denotes the Lebesgue measure on $[0, 1]$. Then there exists a unique measure $Q \in \Pi(\lambda, \mu, \nu)$ on $\mathbb{R}^3$ such that $Q \circ (X_0, X_1)^{-1} = \pi$ and*

$$Q|_{[0,s]\times\mathbb{R}\times\mathbb{R}} \circ (X_1, X_2)^{-1} \in \mathcal{M}(\pi_s, \mathcal{S}^\nu(\pi_s)), \quad s \in \mathbb{R},$$

where $\pi_s := \pi|_{[0,s]\times\mathbb{R}} \circ (X_1)^{-1}$.

We shall also need the following facts about shadows.

Lemma 6.5.

(i) Let μ_1, μ_2, ν be finite measures satisfying $\mu_1 + \mu_2 \leq_{pc} \nu$. Then $\mu_2 \leq_{pc} \nu - \mathcal{S}^{\nu}(\mu_1)$ and $\mathcal{S}^{\nu}(\mu_1 + \mu_2) = \mathcal{S}^{\nu}(\mu_1) + \mathcal{S}^{\nu - \mathcal{S}^{\nu}(\mu_1)}(\mu_2)$.

(ii) Let μ, ν_1, ν_2 be finite measures such that $\mu \leq_{pc} \nu_1 \leq_c \nu_2$. Then, it follows that $\mathcal{S}^{\nu_1}(\mu) \leq_{pc} \nu_2$. Moreover, $\mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) = \mathcal{S}^{\nu_2}(\mu)$ if and only if $\mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu)$.

Proof. Part (i) is [8, Theorem 4.8]. To obtain the first statement in (ii), we observe that $\mathcal{S}^{\nu_1}(\mu) \leq \nu_1 \leq_c \nu_2$ and hence

$$\mathcal{S}^{\nu_1}(\mu)(\phi) \leq \nu_1(\phi) \leq \nu_2(\phi)$$

for any nonnegative convex function ϕ. Turning to the second statement, the "only if" implication follows directly from the definition of the shadow in Lemma 6.1. To show the reverse implication, suppose that $\mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu)$. Then, we have

$$\mu \leq_c \mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \leq \nu_2 \quad \text{and} \quad \mathcal{S}^{\nu_1}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mu) \leq \nu_2.$$

These inequalities imply that

$$\mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \in [\![\mu; \nu_2]\!] \quad \text{and} \quad \mathcal{S}^{\nu_2}(\mu) \in \left[\!\left[\mathcal{S}^{\nu_1}(\mu); \nu_2\right]\!\right],$$

and now the minimality property of the shadow shows that

$$\mathcal{S}^{\nu_2}(\mu) \leq_c \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \quad \text{and} \quad \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_1}(\mu)) \leq_c \mathcal{S}^{\nu_2}(\mathcal{S}^{\nu_2}(\mu)) = \mathcal{S}^{\nu_2}(\mu)$$

as desired. □

6.2. Construction of a multistep left-monotone transport

Our next goal is to define and construct a multistep left-monotone transport. The following concept will be crucial.

Definition 6.6. Let $\mu_0 \leq_{pc} \mu_1 \leq_c \cdots \leq_c \mu_n$. For $1 \leq t \leq n$, the *obstructed shadow of μ_0 in μ_t through $\mu_1, \ldots, \mu_{t-1}$* is iteratively defined by

$$\mathcal{S}^{\mu_1,\ldots,\mu_t}(\mu_0) := \mathcal{S}^{\mu_t}(\mathcal{S}^{\mu_1,\ldots,\mu_{t-1}}(\mu_0)).$$

The obstructed shadow is well-defined due to Lemma 6.5(ii). An alternative definition is provided by the following characterization.

Lemma 6.7. *Let $\mu_0 \leq_{pc} \mu_1 \leq_c \cdots \leq_c \mu_n$ and $1 \leq t \leq n$. Then $\mathcal{S}^{\mu_1,\ldots,\mu_t}(\mu_0)$ is the unique least element of the set*

$$[\![\mu_0; \mu_t]\!]^{\mu_1,\ldots,\mu_{t-1}} := \{\theta_t \leq \mu_t : \exists \theta_s \leq \mu_s,\ 1 \leq s \leq t-1,\ \mu_0 \leq_c \theta_1 \leq_c \cdots \leq_c \theta_t\}$$

for the convex order; that is, $\mathcal{S}^{\mu_1,\ldots,\mu_t}(\mu_0) \leq_c \theta$ for all elements θ.

Proof. For $t = 1$ this holds by the definition of the shadow in Lemma 6.1. For $t > 1$, we inductively assume that $\mathcal{S}^{\mu_1,\ldots,\mu_{t-1}}(\mu_0)$ is the least element of $[\![\mu_0; \mu_{t-1}]\!]^{\mu_1,\ldots,\mu_{t-2}}$. Consider an arbitrary element $\theta_t \in [\![\mu_0; \mu_t]\!]^{\mu_1,\ldots,\mu_{t-1}}$ and fix some

$$\mu_0 \leq_c \theta_1 \leq_c \cdots \leq_c \theta_{t-1} \leq_c \theta_t \quad \text{with} \quad \theta_s \leq \mu_s, \quad 1 \leq s \leq t-1.$$

Then, $\theta_{t-1} \in [\![\mu_0; \mu_{t-1}]\!]^{\mu_1,\dots,\mu_{t-2}}$ and in particular $\mathcal{S}^{\mu_1,\dots,\mu_{t-1}}(\mu_0) \leq_c \theta_{t-1}$. Recall that $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0)$ is defined as the least element for $\leq_c$ of

$$\begin{aligned}[\![\mathcal{S}^{\mu_1,\dots,\mu_{t-1}}(\mu_0); \mu_t]\!] &= \{\theta \leq \mu_t : \mathcal{S}^{\mu_1,\dots,\mu_{t-1}}(\mu_0) \leq_c \theta\} \\ &\supseteq \{\theta \leq \mu_t : \theta_{t-1} \leq_c \theta\} \ni \theta_t.\end{aligned}$$

Hence, $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0) \leq_c \theta_t$, and as $\theta_t \in [\![\mu_0; \mu_t]\!]^{\mu_1,\dots,\mu_{t-1}}$ was arbitrary, this shows that $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0)$ is a least element of $[\![\mu_0; \mu_t]\!]^{\mu_1,\dots,\mu_{t-1}}$. The uniqueness of the least element follows from the general fact that $\theta_t^1 \leq_c \theta_t^2$ and $\theta_t^2 \leq_c \theta_t^1$ imply $\theta_t^1 = \theta_t^2$. □

We can now state the main result of this section.

Theorem 6.8. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order. Then there exists $P \in \mathcal{M}(\boldsymbol{\mu})$ such that the bivariate projections $P_{0t} := P \circ (X_0, X_t)^{-1}$ satisfy*

$$P_{0t}[(-\infty, x] \times A] = \mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{(-\infty,x]})(A) \quad \text{for} \quad x \in \mathbb{R},\ A \in \mathfrak{B}(\mathbb{R}),$$

for all $1 \leq t \leq n$. Any such $P \in \mathcal{M}(\boldsymbol{\mu})$ is called a left-monotone transport.

We observe that an n-step left-monotone transport is defined purely in terms of its bivariate projections $P \circ (X_0, X_t)^{-1}$. In the one-step case, this completely determines the transport. For $n > 1$, we shall see that there can be multiple (and then infinitely many) left-monotone transports; in fact, they form a convex compact set. This will be discussed in more detail in Section 8, where it will also be shown that uniqueness does hold if μ_0 is atomless.

Proof of Theorem 6.8. *Step 1.* We first construct measures $\pi_t \in \Pi(\lambda, \mu_t)$, $0 \leq t \leq n$ such that

$$\pi_t|_{[0,\mu_0((-\infty,x])]\times\mathbb{R}} \circ X_1^{-1} = \mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{(-\infty,x]})$$

for all $x \in \mathbb{R}$, as well as measures $Q_t \in \Pi(\lambda, \mu_{t-1}, \mu_t)$, $1 \leq t \leq n$ such that

$$\begin{aligned}&Q_t|_{[0,\mu_0((-\infty,x])]\times\mathbb{R}\times\mathbb{R}} \circ (X_1, X_2)^{-1} \in \\ &\qquad \mathcal{M}\big(\mathcal{S}^{\mu_1,\dots,\mu_{t-1}}(\mu_0|_{(-\infty,x]}), \mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{(-\infty,x]})\big)\end{aligned} \tag{6.1}$$

for all $x \in \mathbb{R}$. Indeed, for $t = 0$, we take $\pi_0 \in \Pi(\lambda, \mu_0)$ to be the quantile[11] coupling. Then, applying Proposition 6.4 to π_0 yields the measure Q_1, and we can define $\pi_1 := Q_1 \circ (X_0, X_2)^{-1}$. Proceeding inductively, applying Proposition 6.4 to π_{t-1} yields Q_t which in turn allows us to define $\pi_t := Q_t \circ (X_0, X_2)^{-1}$.

Step 2. For $1 \leq t \leq n$, consider a disintegration $Q_t = \pi_{t-1} \otimes \kappa_t$ of Q_t. By (6.1), we may choose $\kappa_t(s, x_{t-1}, dx_t)$ to be a martingale kernel; that is,

$$\int x_t \, \kappa_t(s, x_{t-1}, dx_t) = x_{t-1}$$

holds for all $(s, x_{t-1}) \in \mathbb{R}^2$. We now define a measure $\pi \in \Pi(\lambda, \mu_0, \dots, \mu_n)$ on $\mathbb{R}^{n+2}$ via

$$\pi = \pi_0 \otimes \kappa_1 \otimes \cdots \otimes \kappa_n.$$

Then, π satisfies

$$\pi \circ (X_0, X_t)^{-1} = \pi_{t-1} \quad \text{and} \quad \pi \circ (X_0, X_t, X_{t+1})^{-1} = Q_t$$

for $1 \leq t \leq n$, and setting $P = \pi \circ (X_1, \dots, X_{n+1})^{-1}$ yields the theorem. □

[11] The quantile coupling (or Fréchet–Hoeffding coupling) is given by the law of $(F_\lambda^{-1}, F_{\mu_0}^{-1})$ under λ, where $F_{\mu_0}^{-1}$ is the inverse c.d.f. of μ_0.

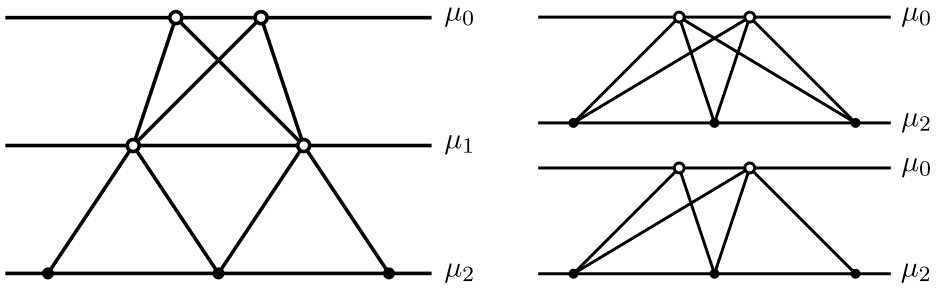

Fig. 3. The left panel shows the support of the left-monotone transport P from Example 6.10. The right panel shows the support of P_{02} (top) and the support of the left-monotone transport in $\mathcal{M}(\mu_0, \mu_2)$ (bottom). The elements of the support are represented by the diagonal lines.

The following result studies the bivariate projections P_{0t} of a left-monotone transport and shows in particular that P_{0t} may differ from the Left-Curtain coupling [8] in $\mathcal{M}(\mu_0, \mu_t)$.

Proposition 6.9. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order and let $P \in \mathcal{M}(\boldsymbol{\mu})$ be a left-monotone transport. The following are equivalent:*

(i) The bivariate projection $P_{0t} = P \circ (X_0, X_t)^{-1} \in \mathcal{M}(\mu_0, \mu_t)$ is left-monotone for all $1 \le t \le n$.
(ii) The marginals $\boldsymbol{\mu}$ satisfy

$$\mathcal{S}^{\mu_1}(\mu_0|_{(-\infty,x]}) \le_c \cdots \le_c \mathcal{S}^{\mu_n}(\mu_0|_{(-\infty,x]}) \quad \text{for all} \quad x \in \mathbb{R}. \tag{6.2}$$

Proof. Given $\mu \le \mu_0$, an iterative application of Lemma 6.5(ii) shows that the obstructed shadows coincide with the ordinary shadows, i.e. $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu) = \mathcal{S}^{\mu_t}(\mu)$ for $1 \le t \le n$, if and only if $\mathcal{S}^{\mu_1}(\mu) \le_c \cdots \le_c \mathcal{S}^{\mu_n}(\mu)$. The proposition follows by applying this observation to $\mu = \mu_0|_{(-\infty,x]}$. □

The following example illustrates the proposition and shows that (6.2) may indeed fail.

Example 6.10. Consider the marginals

$$\mu_0 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1, \quad \mu_1 = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_2, \quad \mu_2 = \frac{1}{4}\delta_{-4} + \frac{1}{2}\delta_0 + \frac{1}{4}\delta_4.$$

Then the set $\mathcal{M}(\boldsymbol{\mu})$ consists of a single transport P; cf. the left panel of Fig. 3. Thus, P is necessarily left-monotone. Similarly, $P_{01} = P \circ (X_0, X_1)^{-1}$ is the unique element of $\mathcal{M}(\mu_0, \mu_1)$. However, $P_{02} = P \circ (X_0, X_2)^{-1}$ is given by

$$\frac{3}{16}\delta_{(-1,-4)} + \frac{1}{4}\delta_{(-1,0)} + \frac{1}{16}\delta_{(-1,4)} + \frac{1}{16}\delta_{(1,-4)} + \frac{1}{4}\delta_{(1,0)} + \frac{3}{16}\delta_{(1,4)}$$

whereas the unique left-monotone transport in $\mathcal{M}(\mu_0, \mu_2)$ can be found to be

$$\frac{1}{8}\delta_{(-1,-4)} + \frac{3}{8}\delta_{(-1,0)} + \frac{1}{8}\delta_{(1,-4)} + \frac{1}{8}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,4)}.$$

Therefore, there exists no transport $P \in \mathcal{M}(\boldsymbol{\mu})$ such that both P_{01} and P_{02} are left-monotone, and Proposition 6.9 shows that (6.2) fails.

Remark 6.11. Of course, all our results on left-monotone transports have "right-monotone" analogues, obtained by reversing the orientation on the real line (i.e. replacing $x \mapsto -x$ everywhere).

7. Geometry and optimality properties

In this section we introduce the optimality properties for transports and the geometric properties of their supports that were announced in the Introduction, and prove that they equivalently characterize left-monotone transports.

7.1. Geometry of optimal transports for reward functions of Spence–Mirrlees type

The first goal is to show that optimal transports for specific reward functions are concentrated on sets $\Gamma \subseteq \mathbb{R}^{n+1}$ satisfying certain no-crossing conditions that we introduce next. Given $1 \leq t \leq n$, we write

$$\Gamma^t = \{(x_0, \dots, x_t) \in \mathbb{R}^{t+1} : (x_0, \dots, x_n) \in \Gamma \text{ for some } (x_{t+1}, \dots, x_n) \in \mathbb{R}^{n-t}\}$$

for the projection of Γ onto the first $t+1$ coordinates.

Definition 7.1. Let $\Gamma \subseteq \mathbb{R}^{n+1}$ and $1 \leq t \leq n$. Consider $\boldsymbol{x} = (x_0, \dots, x_{t-1})$, $\boldsymbol{x}' = (x'_0, \dots, x'_{t-1}) \in \mathbb{R}^t$ and $y^+, y^-, y' \in \mathbb{R}$ with $y^- < y^+$ such that $(\boldsymbol{x}, y^+), (\boldsymbol{x}, y^-), (\boldsymbol{x}', y') \in \Gamma^t$. Then, the projection

$$\Gamma^t \text{ is } \textit{left-monotone} \quad \text{if} \quad y' \notin (y^-, y^+) \text{ whenever } x_0 < x'_0.$$

The set Γ is left-monotone[12] if Γ^t is left-monotone for all $1 \leq t \leq n$.

We also need the following notion.

Definition 7.2. Let $\Gamma \subseteq \mathbb{R}^{n+1}$ and $1 \leq t \leq n$. The projection Γ^t is *nondegenerate* if for all $\boldsymbol{x} = (x_0, \dots, x_{t-1}) \in \mathbb{R}^t$ and $y \in \mathbb{R}$ such that $(\boldsymbol{x}, y) \in \Gamma^t$, the following hold:

(i) if $y > x_{t-1}$, there exists $y' < x_{t-1}$ such that $(\boldsymbol{x}, y') \in \Gamma^t$;
(ii) if $y < x_{t-1}$, there exists $y' > x_{t-1}$ such that $(\boldsymbol{x}, y') \in \Gamma^t$.

The set Γ is called nondegenerate[13] if Γ^t is nondegenerate for all $1 \leq t \leq n$.

Broadly speaking, this definition says that for any path to the right in Γ there exists a path to the left, and vice versa. For a set supporting a martingale, nondegeneracy is not a restriction, in the following sense.

Remark 7.3. Let $\boldsymbol{\mu}$ be in convex order, $\mathcal{V}$ its effective domain and $\Gamma \subseteq \mathcal{V}$.

(i) There exists a nondegenerate, universally measurable set $\Gamma' \subseteq \Gamma$ such that $P(\Gamma') = 1$ for all $P \in \mathcal{M}(\boldsymbol{\mu})$ with $P(\Gamma) = 1$.

(ii) Fix $P \in \mathcal{M}(\boldsymbol{\mu})$ with $P(\Gamma) = 1$. There exists a nondegenerate, Borel-measurable set $\Gamma'_P \subseteq \Gamma$ such that $P(\Gamma'_P) = 1$.

[12] This terminology for Γ is abusive since $\Gamma = \Gamma^n$ is in fact a projection itself—it will be clear from the context what is meant.

[13] Footnote 12 applies here as well.

Proof. Let N_t be the set of all $\boldsymbol{x} \in \Gamma^t$ such that (i) or (ii) of Definition 7.2 fail. If P is a martingale with $P(\Gamma) = 1$, we see that $N_t \times \mathbb{R}^{n-t+1}$ is P-null. Moreover, N_t is universally measurable (as the projection of a Borel set) and we can set

$$\Gamma' := \Gamma \setminus \bigcup_{t=1}^{n} (N_t \times \mathbb{R}^{n-t+1})$$

to prove (i). Turning to (ii), universal measurability implies that there exists a Borel set $N_t' \supseteq N_t$ such that $N_t' \setminus N_t$ is P_{t-1}-null, where $P_{t-1} = P \circ (X_0, \dots, X_{t-1})^{-1}$. We can then set $\Gamma_P' := \Gamma \setminus \cup_{t=1}^{n} (N_t' \times \mathbb{R}^{n-t+1})$. □

Next, we introduce a notion of competitors along the lines of [8, Definition 1.10].

Definition 7.4. Let π be a finite measure on $\mathbb{R}^{t+1}$ whose marginals have finite first moments and consider a disintegration $\pi = \pi_t \otimes \kappa$, where π_t is the projection of π onto the first t coordinates. A measure $\pi' = \pi_t \otimes \kappa'$ is a *t-competitor* of π if it has the same last marginal and

$$\text{bary}(\kappa(\boldsymbol{x}, \cdot)) = \text{bary}(\kappa'(\boldsymbol{x}, \cdot)) \quad \text{for } \pi_t\text{-a.e.} \quad \boldsymbol{x} = (x_0, \dots, x_{t-1}).$$

Using these definitions, we now formulate a variant of the monotonicity principle stated in Theorem 5.4 (i) that will be convenient to infer the geometry of Γ.

Lemma 7.5. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order, $1 \le t \le n$ and let $\bar{f} : \mathbb{R}^{t+1} \to [0, \infty)$ be Borel. Consider $f(X_0, \dots, X_n) := \bar{f}(X_0, \dots, X_t)$ and suppose that $\mathbf{I}_{\boldsymbol{\mu}}(f) < \infty$. Let $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f)$ be an optimizer for $\mathbf{I}_{\boldsymbol{\mu}}(f)$ with the property that $\phi_s \equiv H_s \equiv 0$ for $s = t+1, \dots, n$ and define the set*

$$\Gamma := \left\{ \sum_{t=0}^{n} \phi_t(X_t) + (H \cdot X)_n = f \right\} \cap \mathcal{V}.$$

Let π be a finitely supported probability on $\mathbb{R}^{t+1}$ which is concentrated on Γ^t. Then $\pi(\bar{f}) \ge \pi'(\bar{f})$ for any t-competitor π' of π that is concentrated on $\mathcal{V}^t$.

Proof. Recall that the projections π_t and π_t' onto the first t coordinates coincide. Thus,

$$\begin{aligned}
\pi[H_t \cdot (X_t - X_{t-1})] &= \int H_t \cdot (\text{bary}(\kappa(X_0, \dots, X_{t-1}, \cdot)) - X_{t-1}) d\pi_t \\
&= \int H_t \cdot (\text{bary}(\kappa'(X_0, \dots, X_{t-1}, \cdot)) - X_{t-1}) d\pi_t' \\
&= \pi'[H_t \cdot (X_t - X_{t-1})].
\end{aligned}$$

Using also that the last marginals coincide, we deduce that

$$\pi[\bar{f}] = \pi\left[\sum_{s=0}^{t} \phi_s(X_s) + (H \cdot X)_t\right] = \pi'\left[\sum_{s=0}^{t} \phi_s(X_s) + (H \cdot X)_t\right] \ge \pi'[\bar{f}]. \quad \square$$

Next, we formulate an intermediate result relating optimality for Spence–Mirrlees reward functions to left-monotonicity of the support.

Lemma 7.6. *Let $1 \le t \le n$ and let $\Gamma \subseteq \mathcal{V}$ be a subset such that Γ^t is nondegenerate. Moreover, let $f : \mathbb{R}^{t+1} \to \mathbb{R}$ be of the form $f(X_0, \dots, X_t) = \bar{f}(X_0, X_t)$ for a second-order Spence–Mirrlees function $\bar{f}$. Assume that for any finitely supported probability π that*

is concentrated on Γ^t and any t-competitor π' of π that is concentrated on $\mathcal{V}^t$, we have $\pi(f) \geq \pi'(f)$. Then, the projection Γ^t is left-monotone.

Proof. Consider $(\boldsymbol{x}, y_1), (\boldsymbol{x}, y_2), (\boldsymbol{x}', y') \in \Gamma^t$ satisfying $x_0 < x_0'$ and suppose for contradiction that $y_1 < y' < y_2$. We define $\lambda = \frac{y_2 - y'}{y_2 - y_1}$ and

$$\pi = \frac{\lambda}{2}\delta_{(\boldsymbol{x}, y_1)} + \frac{1-\lambda}{2}\delta_{(\boldsymbol{x}, y_2)} + \frac{1}{2}\delta_{(\boldsymbol{x}', y')}$$
$$\pi' = \frac{\lambda}{2}\delta_{(\boldsymbol{x}', y_1)} + \frac{1-\lambda}{2}\delta_{(\boldsymbol{x}', y_2)} + \frac{1}{2}\delta_{(\boldsymbol{x}, y')}.$$

Then π and π' have the same projection $\pi_t = \pi'_t$ on the first t marginals and their last marginals also coincide. Moreover, disintegrating $\pi = \pi_t \otimes \kappa$ and $\pi' = \pi_t \otimes \kappa'$, the measures $\kappa(\boldsymbol{x}), \kappa(\boldsymbol{x}'), \kappa'(\boldsymbol{x}), \kappa(\boldsymbol{x}')$ all have barycenter y'. Therefore, π and π' are t-competitors. We must also have that π' is concentrated on $\mathcal{V}^t$, by the shape of $\mathcal{V}$. Now our assumption implies that $\pi(f) \geq \pi'(f)$, but the second-order Spence–Mirrlees property of $\bar{f}$ implies that $\pi(f) < \pi'(f)$. □

7.2. Geometry of left-monotone transports

Next, we establish that transports with left-monotone support are indeed left-monotone in the sense of Theorem 6.8.

Theorem 7.7. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order and let $P \in \mathcal{M}(\boldsymbol{\mu})$ be concentrated on a nondegenerate, left-monotone set $\Gamma \subseteq \mathbb{R}^{n+1}$. Then P is left-monotone.*

Before stating the proof of the theorem, we record two auxiliary results about measures on the real line. The first one is a direct consequence of Proposition 2.1.

Lemma 7.8. *Let $a < b$ and $\mu \leq_c \nu$. If ν is concentrated on $(-\infty, a]$, then so is μ, and moreover $\nu(\{a\}) \geq \mu(\{a\})$. The analogue holds for $[b, \infty)$.*

The second result is [8, Lemma 5.2].

Lemma 7.9. *Let σ be a nontrivial signed measure on $\mathbb{R}$ with $\sigma(\mathbb{R}) = 0$ and let $\sigma = \sigma^+ - \sigma^-$ be its Hahn decomposition. There exist $a \in \operatorname{supp}(\sigma^+)$ and $b > a$ such that $\int (b-y)^+ \mathbf{1}_{[a,\infty)}\, d\sigma(y) > 0$.*

We can now give the proof of the theorem; it is inspired by [8, Theorem 5.3] which corresponds to the case $n = 1$.

Proof of Theorem 7.7. Since the case $n = 1$ is covered by Proposition 6.3, we may assume that the theorem has been proved for transports with $n - 1$ steps and focus on the induction argument.

For every $x \in \mathbb{R}$ we denote by μ_x^t the marginal $(P|_{(-\infty,x]\times\mathbb{R}^n}) \circ X_t^{-1}$. In particular, we then have $\mu_x^0 = \mu_0|_{(-\infty,x]}$ and μ_x^t is the image of μ_x^0 under P after t steps. For the sake of brevity, we also set $\nu_x^t := \mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_x^0)$. By definition, P is left-monotone if $\mu_x^t = \nu_x^t$ for all $x \in \mathbb{R}$ and $t \leq n$, and by the induction hypothesis, we may assume that this holds for $t \leq n-1$.

We argue by contradiction and assume that there exists $x \in \mathbb{R}$ such that $\mu_x^n \neq \nu_x^n$. Then, the signed measure

$$\sigma := \nu_x^n - \mu_x^n$$

is nontrivial and we can find $a < b$ with $a \in \operatorname{supp}(\sigma^+)$ as in Lemma 7.9. Observe that $\sigma^+ \le \mu_n - \mu_x^n$ where $\mu_n - \mu_x^n$ is the image of $\mu_n|_{(x,\infty)}$ under P. Hence, $a \in \operatorname{supp}(\mu_n - \mu_x^n)$ and as P is concentrated on Γ, we conclude that there exists a sequence of points

$$\boldsymbol{x}^m = (x_0^m, \dots, x_n^m) \in \Gamma \quad \text{with } x < x_0^m \text{ and } x_n^m \to a. \tag{7.1}$$

Moreover, by the characterization of the obstructed shadow in Lemma 6.7, we must have

$$\nu_x^n \le_c \mu_x^n$$

as $\mu_x^n \in [\![\mu_x^0; \mu_n]\!]^{\mu_1,\dots,\mu_{n-1}}$ due to the fact that μ_x^n is the image of μ_x^0 under a martingale transport.

Step 1. We claim that for all $\boldsymbol{x} = (x_0, \dots, x_{n-1})$ with $x_0 \le x$ and $x_{n-1} \le a$, it holds that

$$\Gamma_{\boldsymbol{x}} \cap (a, \infty) = \emptyset,$$

where $\Gamma_{\boldsymbol{x}} = \{y \in \mathbb{R} : (\boldsymbol{x}, y) \in \Gamma\}$ is the section of Γ at $\boldsymbol{x}$. By way of contradiction, assume that for some $\boldsymbol{x}$ with $x_0 \le x$ and $x_{n-1} \le a$ we have $\Gamma_{\boldsymbol{x}} \cap (a, \infty) \neq \emptyset$, then in particular $\Gamma_{\boldsymbol{x}} \cap (x_{n-1}, \infty) \neq \emptyset$. In view of the nondegeneracy of Γ, we conclude that $\Gamma_{\boldsymbol{x}} \cap (-\infty, x_{n-1}) \neq \emptyset$ and hence that $\Gamma_{\boldsymbol{x}} \cap (-\infty, a) \neq \emptyset$. This yields a contradiction to the left-monotonicity of Γ by using $\boldsymbol{x}^m$ from (7.1) for $\boldsymbol{x}'$ in Definition 7.1 for large enough m, and the proof of the claim is complete.

Step 2. Similarly, we can show that for all $\boldsymbol{x} = (x_0, \dots, x_{n-1})$ with $x_0 \le x$ and $x_{n-1} \ge a$,

$$\Gamma_{\boldsymbol{x}} \cap (-\infty, a) = \emptyset.$$

Step 3. Next, we consider the marginals

$$\mu_{x,a}^t := \left(P|_{(-\infty,x]\times\mathbb{R}^{n-2}\times(-\infty,a]\times\mathbb{R}}\right) \circ X_t^{-1}.$$

Then, in particular, $\mu_{x,a}^{n-1} = \mu_x^{n-1}|_{(-\infty,a]}$ and $\mu_{x,a}^n$ is the image of $\mu_{x,a}^{n-1}$ under the last step of P. Step 1 of the proof thus implies that $\mu_{x,a}^n$ is concentrated on $(-\infty, a]$. We also write

$$\nu_{x,a}^n := \mathcal{S}^{\mu_n}(\mu_x^{n-1}|_{(-\infty,a]}).$$

We have $\mu_{x,a}^{n-1} \le_c \mu_{x,a}^n$ as $\mathcal{M}(\mu_{x,a}^{n-1}, \mu_{x,a}^n) \neq \emptyset$, and $\mu_{x,a}^n \le \mu_x^n \le \mu_n$. Therefore,

$$\nu_{x,a}^n \le_c \mu_{x,a}^n \tag{7.2}$$

by the minimality of the shadow. Next, we show that

$$\nu_x^n - \nu_{x,a}^n \le_c \mu_x^n - \mu_{x,a}^n. \tag{7.3}$$

Observe that $\mu_x^n - \mu_{x,a}^n$ is the image of $\mu_x^{n-1}|_{(a,\infty)}$ under P and therefore concentrated on $[a, \infty)$ by Step 2. Using this observation, that $\mu_{x,a}^n$ is concentrated on $(-\infty, a]$ as mentioned above, and the fact that $\nu_{x,a}^n(\{a\}) \le \mu_{x,a}^n(\{a\})$ as a consequence of (7.2) and Lemma 7.8, we have

$$\mu_x^n - \mu_{x,a}^n = (\mu_x^n - \mu_{x,a}^n)|_{[a,\infty)} \le (\mu_n - \mu_{x,a}^n)|_{[a,\infty)} \le (\mu_n - \nu_{x,a}^n)|_{[a,\infty)} \le \mu_n - \nu_{x,a}^n.$$

We also have $\mu_x^{n-1}|_{(a,\infty)} \le_c \mu_x^n - \mu_{x,a}^n$ since the latter measure is the image of the former under P. Together with the preceding display, we have established that

$$\mu_x^n - \mu_{x,a}^n \in [\![\mu_x^{n-1}|_{(a,\infty)}; \mu_n - \nu_{x,a}^n]\!].$$

On the other hand,

$$\nu_x^n - \nu_{x,a}^n = \mathcal{S}^{\mu_n - \nu_{x,a}^n}(\mu_x^{n-1}|_{(a,\infty)})$$

from the additivity property of the shadow in Lemma 6.5(i), and therefore (7.3) follows by the minimality of the shadow.

Step 4. Recall from Step 3 that $\mu^n_{x,a}$ is concentrated on $(-\infty, a]$ and that $\mu^n_x - \mu^n_{x,a}$ is concentrated on $[a, \infty)$. Therefore, $\nu^n_{x,a}$ is concentrated on $(-\infty, a]$ and $\nu^n_x - \nu^n_{x,a}$ is concentrated on $[a, \infty)$, by Lemma 7.8. Moreover, we have $\nu^n_{x,a}(\{a\}) \le \mu^n_{x,a}(\{a\})$ by the same lemma, and finally, the function $y \mapsto (b-y)^+ \mathbf{1}_{[a,\infty)}(y)$ is convex on $[a, \infty)$ as $a < b$. Using these facts and (7.3),

$$\begin{aligned}
&\int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) \nu^n_x(dy) \\
&\qquad = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)(\nu^n_x - \nu^n_{x,a})(dy) + (b-a)\nu^n_{x,a}(\{a\}) \\
&\qquad \le \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y)(\mu^n_x - \mu^n_{x,a})(dy) + (b-a)\mu^n_{x,a}(\{a\}) \\
&\qquad = \int (b-y)^+ \mathbf{1}_{[a,\infty)}(y) \mu^n_x(dy).
\end{aligned}$$

This contradicts the choice of a and b, cf. Lemma 7.9, and thus completes the proof. □

7.3. Optimality properties

In this section we relate left-monotone transports and left-monotone sets to the optimal transport problem for Spence–Mirrlees functions.

Theorem 7.10. *For $1 \le t \le n$, let $f_t : \mathbb{R}^2 \to \mathbb{R}$ be second-order Spence–Mirrlees functions such that $|f_t(x,y)| \le a_0(x) + a_t(y)$ for some $a_0 \in L^1(\mu_0)$ and $a_t \in L^1(\mu_t)$. There exists a universally measurable, nondegenerate, left-monotone set $\Gamma' \subseteq \mathbb{R}^{n+1}$ such that any simultaneous optimizer $P \in \mathcal{M}(\boldsymbol{\mu})$ for $\mathbf{S}_{\boldsymbol{\mu}}(f_t(X_0, X_t))$, $1 \le t \le n$ is concentrated on Γ'. In particular, any such P is left-monotone.*

Proof. The last assertion follows by an application of Theorem 7.7, so we may focus on finding Γ'. For each $1 \le t \le n$, we use Theorem 5.2 and Remark 5.3 to find a dual optimizer $(\boldsymbol{\phi}, H) \in \mathcal{D}_{\boldsymbol{\mu}}(f_t)$ for $\mathbf{I}_{\boldsymbol{\mu}}(f_t(X_0, X_t))$ and define the Borel set

$$\Gamma(t) := \left\{ \sum_{s=0}^{n} \phi_s(X_s) + (H \cdot X)_n = f_t \right\} \cap \mathcal{V}.$$

Here, we may choose a dual optimizer such that $\phi_s \equiv H_s \equiv 0$ for $s = t+1, \dots, n$. (This can be seen by applying Theorem 5.2 to the transport problem involving only the marginals $(\mu_0, \dots, \mu_t)$ and taking the corresponding dual optimizer.) Theorem 5.4 shows that any simultaneous optimizer $P \in \mathcal{M}(\boldsymbol{\mu})$ is concentrated on $\Gamma(t)$ for all t, and hence also on the Borel set

$$\Gamma := \bigcap_{t=1}^{n} \Gamma(t).$$

Using Remark 7.3(i), we find a universally measurable, nondegenerate subset $\Gamma' \subseteq \Gamma$ with the same property. Since the projection $(\Gamma')^t$ is contained in the projection $(\Gamma(t))^t$, Lemma 7.5 and Lemma 7.6 yield that $(\Gamma')^t$ is left-monotone for all t; that is, Γ' is left-monotone. □

Remark 7.11. In Theorem 7.10, if we only wish to find a nondegenerate, left-monotone set $\Gamma'_P \subseteq \mathbb{R}^{n+1}$ such that a given simultaneous optimizer $P \in \mathcal{M}(\boldsymbol{\mu})$ is concentrated on Γ'_P, then we may choose Γ'_P to be Borel instead of universally measurable. This follows by replacing the application of Remark 7.3(i) by Remark 7.3(ii) in the proof.

The following is a converse to Theorem 7.10.

Theorem 7.12. *Given $1 \le t \le n$, let $f \in C^{1,2}(\mathbb{R}^2)$ be such that $f_{xyy} \ge 0$ and suppose that the following integrability condition holds:*

$$\begin{cases} f(X_0, X_t), \quad f(0, X_t), \quad f(X_0, 0), \quad \bar{h}(X_0)X_0, \quad \bar{h}(X_0)X_t \\ \text{are } P\text{-integrable for all } \ P \in \mathcal{M}(\boldsymbol{\mu}), \end{cases} \tag{7.4}$$

where $\bar{h}(x) := \partial_y|_{y=0}[f(x, y) - f(0, y)]$. Then every left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ is an optimizer for $\mathbf{S}_{\boldsymbol{\mu}}(f)$.

The integrability condition clearly holds when f is Lipschitz continuous; in particular, a smooth second-order Spence–Mirrlees function (as defined in the Introduction) satisfies the assumptions of the theorem for any $\boldsymbol{\mu}$.

The proof will be given by an approximation based on the following building blocks for Spence–Mirrlees functions; the construction is novel and may be of independent interest.

Lemma 7.13. *Let $1 \le t \le n$ and let $f(X_0, \dots, X_n) := \mathbf{1}_{(-\infty,a]}(X_0)\varphi(X_t)$ for a concave function φ and $a \in \mathbb{R}$. Then every left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ is an optimizer for $\mathbf{S}_{\boldsymbol{\mu}}(f)$.*

Proof. In view of Lemma 6.7, this follows directly by applying the defining shadow property from Theorem 6.8 with $x = a$. □

The integrability condition (7.4) implies that setting

$$g(x, y) := f(x, 0) + f(0, y) - f(0, 0) + \bar{h}(x)y,$$

the three terms constituting

$$g(X_0, X_t) = [f(X_0, 0) + \bar{h}(X_0)X_0] + [f(0, X_t) - f(0, 0)] + [\bar{h}(X_0)(X_t - X_0)]$$

are P-integrable and $P[g(X_0, X_t)]$ is constant over $P \in \mathcal{M}(\boldsymbol{\mu})$. By replacing f with $f - g$, we may thus assume without loss of generality that

$$f(x, 0) = f(0, y) = f_y(x, 0) = 0 \quad \text{for all} \quad (x, y) \in \mathbb{R}^2. \tag{7.5}$$

After this normalization, integration by parts yields the representation

$$f(x, y) = \int_0^y \int_0^x (y - t) f_{xyy}(s, t)\, ds\, dt. \tag{7.6}$$

Lemma 7.14. *Theorem 7.12 holds under the following additional condition: there exists a constant $c > 0$ such that*

$$x \mapsto f(x, y) \text{ is constant on } \{x > c\} \text{ and on } \{x < -c\},$$
$$y \mapsto f(x, y) \text{ is affine on } \{y > c\} \text{ and on } \{y < -c\}.$$

Proof. Integration by parts implies that for all $(x, y) \in \mathbb{R}^2$, we have the representation

$$\begin{aligned} f(x,y) = &- \int_{-c}^{c}\int_{-c}^{c} \mathbf{1}_{(-\infty,s]}(x)(y-t)^+ f_{xyy}(s,t)\,ds\,dt \\ &+ [f(x,-c) - (-c)f_y(x,-c)] \\ &+ [f(c,y) - f(c,-c) - f_y(c,-c)(y-(-c))] \\ &+ f_y(x,-c)y. \end{aligned}$$

The last three terms are of the form $g(x, y) = \tilde{\phi}(x) + \tilde{\psi}(y) + \tilde{h}(x)y$ and of linear growth due to the additional condition. Hence, as above, $P'[g(X_0, X_t)] = C$ is constant for $P' \in \mathcal{M}(\boldsymbol{\mu})$. If $P \in \mathcal{M}(\boldsymbol{\mu})$ is left-monotone and $P' \in \mathcal{M}(\boldsymbol{\mu})$ is arbitrary, Fubini's theorem and Lemma 7.13 yield that

$$\begin{aligned} P[f] &= -\int_{-c}^{c}\int_{-c}^{c} P[\mathbf{1}_{(-\infty,s]}(x)(y-t)^+] f_{xyy}(s,t)\,ds\,dt + C \\ &\geq -\int_{-c}^{c}\int_{-c}^{c} P'[\mathbf{1}_{(-\infty,s]}(x)(y-t)^+] f_{xyy}(s,t)\,ds\,dt + C \\ &= P'[f], \end{aligned}$$

where P, P' are understood to integrate with respect to (x, y) and the application of Fubini's theorem is justified by the nonnegativity of the integrand. □

Proof of Theorem 7.12. Let f be as in the theorem. We shall construct functions f^m, $m \geq 1$ satisfying the assumption of Lemma 7.14 as well as $P[f^m] \to P[f]$ for all $P \in \mathcal{M}(\boldsymbol{\mu})$. Once this is achieved, the theorem follows from the lemma.

Indeed, we may assume that f is normalized as in (7.5). Let $m \geq 1$ and let $\rho_m : \mathbb{R} \to [0, 1]$ be a smooth function such that $\rho_m = 1$ on $[-m, m]$ and $\rho_m = 0$ on $[-m-1, m+1]^c$. In view of (7.6), we define f^m by

$$f^m(x,y) = \int_0^y \int_0^x (y-t) f_{xyy}(s,t)\rho_m(s)\rho_m(t)\,ds\,dt.$$

It then follows that f^m satisfies the assumptions of Lemma 7.14 with the constant $c = m + 1$. Moreover, we have

$$0 \leq f^m(x,y) \leq f^{m+1}(x,y) \leq f(x,y) \quad \text{for} \quad x \geq 0$$

and the opposite inequalities for $x \leq 0$, as well as $f^m(x, y) \to f(x, y)$ for all (x, y).

Let $P \in \mathcal{M}(\boldsymbol{\mu})$. Since f is P-integrable, applying monotone convergence separately on $\{x \geq 0\}$ and $\{x \leq 0\}$ yields that $P[f^m] \to P[f]$, and the proof is complete. □

Remark 7.15. The function

$$\bar{f}(x,y) := \tanh(x)\sqrt{1+y^2}$$

satisfies the conditions of Theorem 7.12 for all marginals $\boldsymbol{\mu}$ in convex order, since the latter are assumed to have a finite first moment.

We can now collect the preceding results to obtain, in particular, the equivalences stated in Theorem 1.1.

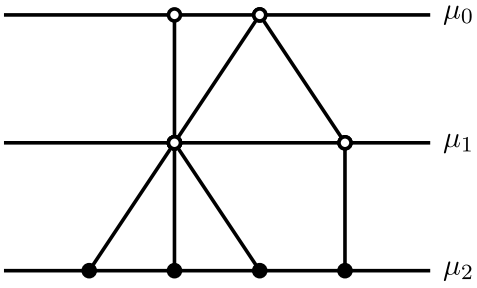

Fig. 4. Support of the non-Markovian transport in Example 7.17.

Theorem 7.16. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order. There exists a left-monotone, nondegenerate, universally measurable set $\Gamma \subseteq \mathbb{R}^{n+1}$ such that for any $P \in \mathcal{M}(\boldsymbol{\mu})$, the following are equivalent:*

- *(i) P is an optimizer for $\mathbf{S}_{\boldsymbol{\mu}}(f(X_0, X_t))$ whenever f is a smooth second-order Spence–Mirrlees function and $1 \le t \le n$,*
- *(ii) P is concentrated on Γ,*
- *(ii') P is concentrated on a left-monotone set,*
- *(iii) P is left-monotone; i.e. P_{0t} transports $\mu_0|_{(-\infty,a]}$ to $\mathcal{S}^{\mu_1,\dots,\mu_t}(\mu_0|_{(-\infty,a]})$ for all $1 \le t \le n$ and $a \in \mathbb{R}$.*

Moreover, there exists $P \in \mathcal{M}(\boldsymbol{\mu})$ satisfying (i)–(iii).

Proof. Let Γ be the set provided by Theorem 7.10 for the function $f_t = \bar{f}$ of Remark 7.15. Given $P \in \mathcal{M}(\boldsymbol{\mu})$, Theorem 7.10 shows that (i) implies (ii) which trivially implies (ii'). Theorem 7.7 and Remark 7.3 show that (ii') implies (iii), and Theorem 7.12 shows that (iii) implies (i). Finally, the existence of a left-monotone transport was stated in Theorem 6.8. □

We conclude this section with an example showing that left-monotone transports are not Markovian in general, even if they are unique and (6.2) holds for $\boldsymbol{\mu}$.

Example 7.17. Consider the marginals

$$\mu_0 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \mu_1 = \frac{3}{4}\delta_0 + \frac{1}{4}\delta_2, \quad \mu_2 = \frac{1}{8}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{8}\delta_1 + \frac{1}{4}\delta_2.$$

The transport $P \in \mathcal{M}(\boldsymbol{\mu})$ given by

$$P = \frac{1}{2}\delta_{(0,0,0)} + \frac{1}{8}\delta_{(1,0,-1)} + \frac{1}{8}\delta_{(1,0,1)} + \frac{1}{4}\delta_{(1,2,2)}$$

is left-monotone because its support is left-monotone (Fig. 4), and it is clearly not Markovian. On the other hand, it is not hard to see that this is the only way to build a left-monotone transport in $\mathcal{M}(\boldsymbol{\mu})$.

8. Uniqueness of left-monotone transports

In this section we consider the (non-)uniqueness of left-monotone transports. It turns out the presence of atoms in μ_0 is important in this respect—let us start with the following simple observation.

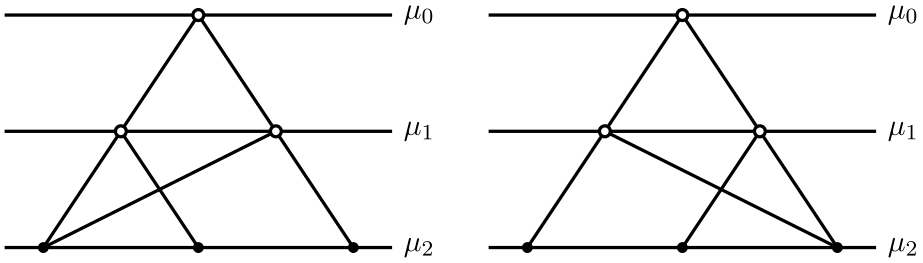

Fig. 5. Supports of two left-monotone transports for the same marginals.

Remark 8.1. Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order. If μ_0 is a Dirac mass, then every $P \in \mathcal{M}(\boldsymbol{\mu})$ is left-monotone. Indeed, $\mathcal{M}(\mu_0, \mu_t)$ is a singleton for every $1 \le t \le n$, hence P_{0t} must be the (one-step) left-monotone transport.

Exploiting this observation, the following shows that left-monotone transports need not be unique when $n \ge 2$.

Example 8.2. Let $\mu_0 = \delta_0$, $\mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, $\mu_2 = \frac{3}{8}\delta_{-2} + \frac{1}{4}\delta_0 + \frac{3}{8}\delta_2$. By the remark, any element in $\mathcal{M}(\boldsymbol{\mu})$ is left-monotone. Moreover, $\mathcal{M}(\boldsymbol{\mu})$ is a continuum since $\mathcal{M}(\mu_1, \mu_2)$ contains the convex hull of the two measures

$$P_l = \frac{1}{4}\delta_{(-1,-2)} + \frac{1}{4}\delta_{(-1,0)} + \frac{1}{8}\delta_{(1,-2)} + \frac{3}{8}\delta_{(1,2)},$$
$$P_r = \frac{3}{8}\delta_{(-1,-2)} + \frac{1}{8}\delta_{(-1,2)} + \frac{1}{4}\delta_{(1,0)} + \frac{1}{4}\delta_{(1,2)}.$$

The corresponding supports are depicted in Fig. 5.

The example illustrates that non-uniqueness can typically be expected when μ_0 has atoms. On the other hand, we have the following uniqueness result.

Theorem 8.3. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order. If μ_0 is atomless, there exists a unique left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$.*

The remainder of this section is devoted to the proof. Let us call a kernel $\kappa(x, dy)$ *binomial* if for all $x \in \mathbb{R}$, the measure $\kappa(x, dy)$ consists of (at most) two point masses. A martingale transport will be called binomial if it can be disintegrated using only binomial kernels. We shall show that when μ_0 is atomless, any left-monotone transport is a binomial martingale, and then conclude the uniqueness via a convexity argument.

The first step is the following set-theoretic result.

Lemma 8.4. *Let $k \ge 1$ be an integer and $\Gamma \subseteq \mathbb{R}^{t+1}$. For $\boldsymbol{x} \in \mathbb{R}^t$, we denote by $\Gamma_{\boldsymbol{x}} := \{y \in \mathbb{R} : (\boldsymbol{x}, y) \in \Gamma\}$ the section at $\boldsymbol{x}$. If the set*

$$\{\boldsymbol{x} \in \mathbb{R}^t : |\Gamma_{\boldsymbol{x}}| \ge k\}$$

is uncountable, then it has an accumulation point. More precisely, there are $\boldsymbol{x} = (x_0, \dots, x_t) \in \mathbb{R}^t$ and $y_1 < \cdots < y_k$ in $\Gamma_{\boldsymbol{x}}$ such that for all $\epsilon > 0$ there exist $\boldsymbol{x}' = (x'_0, \dots, x'_t) \in \mathbb{R}^t$ and $y'_1 < \cdots < y'_k$ in $\Gamma_{\boldsymbol{x}'}$ satisfying

(i) $\|\boldsymbol{x} - \boldsymbol{x}'\| < \epsilon$,

(ii) $x_0 < x_0'$,
(iii) $\max_{i=1,\dots,k} |y_i - y_i'| < \epsilon$.

Proof. The proof is similar to the one of [8, Lemma 3.2] and therefore omitted. □

The following statement on the binomial structure generalizes a result of [8] for the one-step case and is of independent interest.

Proposition 8.5. *Let $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ be in convex order and let μ_0 be atomless. There exists a universally measurable set $\Gamma \subseteq \mathbb{R}^{n+1}$ such that every left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ is concentrated on Γ and such that for all $1 \le t \le n$ and $\boldsymbol{x} \in \mathbb{R}^t$,*

$$|\{y \in \mathbb{R} : (X_0, \dots, X_t)^{-1}(\boldsymbol{x}, y) \cap \Gamma \neq \emptyset\}| \le 2. \tag{8.1}$$

In particular, every left-monotone transport $P \in \mathcal{M}(\boldsymbol{\mu})$ is a binomial martingale.

Proof. Let Γ be as in Theorem 7.16; then every left-monotone $P \in \mathcal{M}(\boldsymbol{\mu})$ is concentrated on Γ. Let A_t be the set of all $\boldsymbol{x} \in \mathbb{R}^t$ such that (8.1) fails. Suppose that A_t is uncountable; then Lemma 8.4 yields points $\boldsymbol{x}, \boldsymbol{x}'$ such that for some $y_1, y_2 \in \Gamma^t_{\boldsymbol{x}}$ and $y \in \Gamma^t_{\boldsymbol{x}'}$ we have $y_1 < y < y_2$. This contradicts the left-monotonicity of Γ (Definition 7.1), thus A_t must be countable. Hence, $(X_0, \dots, X_{t-1})^{-1}(A_t)$ is Borel and P-null for all $P \in \mathcal{M}(\boldsymbol{\mu})$, as μ_0 is atomless. The set $\Gamma' = \Gamma \setminus \cup_{t=1}^n (X_0, \dots, X_{t-1})^{-1}(A_t)$ then has the required properties. □

Proof of Theorem 8.3. We will prove this result using induction on n. For $n = 1$ the result holds by Proposition 6.3, with or without atoms. To show the induction step, let P' be the unique left-monotone transport in $\mathcal{M}(\mu_0, \dots, \mu_{n-1})$ and let $P_1 = P' \otimes \kappa_1$ and $P_2 = P' \otimes \kappa_2$ be disintegrations of two n-step left-monotone transports. Then,

$$\frac{P_1 + P_2}{2} = P' \otimes \frac{\kappa_1 + \kappa_2}{2}$$

is again left-monotone, and Proposition 8.5 yields that $(\kappa_1 + \kappa_2)/2$ must be a binomial kernel P'-a.s. Using also the martingale property of κ_1 and κ_2, this can only be true if $\kappa_1 = \kappa_2$ holds P'-a.s., and therefore $P_1 = P_2$. □

9. Free intermediate marginals

In this section we discuss a variant of our transport problem where the intermediate marginal constraints $\mu_1, \dots, \mu_{n-1}$ are omitted; that is, only the first and last marginals μ_0, μ_n are prescribed. (One could similarly adapt the results to a case where some, but not all of the intermediate marginals are given.)

The primal space will be denoted by $\mathcal{M}^n(\mu_0, \mu_n)$ and consists of all martingale measures P on $\mathbb{R}^{n+1}$ such that $\mu_0 = P \circ (X_0)^{-1}$ and $\mu_n = P \circ (X_n)^{-1}$. To make the connection with the previous sections, we note that

$$\mathcal{M}^n(\mu_0, \mu_n) = \bigcup \mathcal{M}(\boldsymbol{\mu})$$

where the union is taken over all vectors $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_{n-1}, \mu_n)$ in convex order.

9.1. Polar structure

We first characterize the polar sets of $\mathcal{M}^n(\mu_0, \mu_n)$. To that end, we introduce an analogue of the irreducible components.

Definition 9.1. Let $\mu_0 \leq_c \mu_n$ and let $(I_k, J_k) \subseteq \mathbb{R}^2$ be the corresponding irreducible domains in the sense of Proposition 2.3. The *n-step components* of $\mathcal{M}^n(\mu_0, \mu_n)$ are the sets[14]

(i) $I_k^n \times J_k$, where $k \geq 1$,
(ii) $I_0^{n+1} \cap \Delta_n$,
(iii) $I_k^t \times \{p\}^{n-t+1}$, where $p \in J_k \setminus I_k$ and $1 \leq t \leq n$, $k \geq 1$.

The characterization then takes the following form.

Theorem 9.2 (*Polar Structure*). *Let $\mu_0 \leq_c \mu_n$. A Borel set $B \subseteq \mathbb{R}^{n+1}$ is $\mathcal{M}^n(\mu_0, \mu_n)$-polar if and only if there exist a μ_0-nullset N_0 and a μ_n-nullset N_n such that*

$$B \subseteq (N_0 \times \mathbb{R}^n) \cup (\mathbb{R}^n \times N_n) \cup \Big(\bigcup V_j\Big)^c$$

where the union runs over all n-step components V_j of $\mathcal{M}^n(\mu_0, \mu_n)$.

It turns out that our previous results can be put to work to prove the theorem, by means of the following lemma which may be of independent interest.

Lemma 9.3. *Let $\mu \leq_c \nu$ be irreducible with domain (I, J) and let ρ be a probability concentrated on J. Then, there exists a probability $\mu \leq_c \theta \leq_c \nu$ satisfying $\theta \gg \rho$ such that $\mu \leq_c \theta$ and $\theta|_I \leq_c (\nu - \theta|_{J\setminus I})$ are both irreducible.*

Proof. Step 1. We first assume that $\rho = \delta_x$ for some $x \in J$ and show that there exists θ satisfying

$$\mu \leq_c \theta \leq_c \nu \quad \text{and} \quad \theta \gg \delta_x.$$

If ν has an atom at x, we can choose $\theta = \nu$. Thus, we may assume that $\nu(\{x\}) = 0$ and in particular that $x \in I$. Let a be the common barycenter of μ and ν and suppose that $x < a$. For all $b \in \mathbb{R}$ and $0 \leq c \leq \nu(\{b\})$, the measure

$$\nu_{b,c} := \nu|_{(-\infty,b)} + c\delta_b$$

satisfies $\nu_{b,c} \leq \nu$, and as $x < a$ there are unique b, c such that $\mathrm{bary}(\nu_{b,c}) = x$. Setting $\alpha = \nu_{b,c}$ and $\epsilon_0 = \alpha(\mathbb{R})$, we then have $\epsilon_0\delta_x \leq_c \alpha \leq \nu$, and a similar construction yields this result for $x \geq a$. The existence of such α implies that

$$\epsilon\delta_x \leq_{pc} \nu, \quad 0 \leq \epsilon \leq \epsilon_0$$

and thus the shadow $\mathcal{S}^\nu(\epsilon\delta_x)$ is well-defined. This measure is given by the restriction of ν to an interval (possibly including fractions of atoms at the endpoints); cf. [8, Example 4.7]. Moreover, the interval is bounded after possibly reducing the mass ϵ_0. Thus, for all $\epsilon < \epsilon_0$, the difference of potential functions

$$u_{\mathcal{S}^\nu(\epsilon\delta_x)} - u_{\epsilon\delta_x} \geq 0$$

vanishes outside a compact interval, and it converges uniformly to zero as $\epsilon \to 0$.

[14] A superscript m indicates the m-fold Cartesian product; Δ_n is the diagonal in $\mathbb{R}^{n+1}$.

On the other hand, as $\mu \leq_c \nu$ is irreducible, the difference $u_\nu - u_\mu \geq 0$ is uniformly bounded away from zero on compact subsets of I and has nonzero derivative on $J \setminus I$. Together, it follows that

$$u_\nu - u_{\mathcal{S}^\nu(\epsilon\delta_x)} + u_{\epsilon\delta_x} \geq u_\mu \tag{9.1}$$

for small enough $\epsilon > 0$, so that

$$\theta := \nu - \mathcal{S}^\nu(\epsilon\delta_x) + \epsilon\delta_x$$

satisfies $\mu \leq_c \theta \leq_c \nu$; moreover, $\theta \gg \delta_x$ as $\nu(\{x\}) = 0$.

Step 2. We turn to the case of a general probability measure ρ on J. By Step 1, we can find a measure θ_x for each $x \in J$ such that

$$\mu \leq_c \theta_x \leq_c \nu \quad \text{and} \quad \theta_x \gg \delta_x.$$

The map $x \mapsto \theta_x$ can easily be chosen to be measurable (by choosing the ϵ for (9.1) in a measurable way). We can then define the probability measure

$$\theta'(A) := \int_J \theta_x(A)\rho(dx), \quad A \in \mathfrak{B}(\mathbb{R})$$

which satisfies $\mu \leq_c \theta' \leq_c \nu$. Moreover, we have $\theta' \gg \rho$; indeed, if $A \in \mathfrak{B}(\mathbb{R})$ is a θ'-nullset, then $\theta_x(A) = 0$ for ρ-a.e. x and thus $\rho(A) = 0$ as $\theta_x \gg \delta_x$.

Finally, $\theta := (\mu + \theta' + \nu)/3$ shares these properties. As $u_\mu < u_\nu$ on I due to irreducibility, we have $u_\mu < u_\theta < u_\nu$ on I and it follows that $\mu \leq_c \theta$ and $\theta|_I \leq_c (\nu - \theta|_{J\setminus I})$ are irreducible. □

Lemma 9.4. *Let $\mu_0 \leq_c \mu_n$ and let π be a measure on $\mathbb{R}^{n+1}$ which is concentrated on an n-step component V of $\mathcal{M}^n(\mu_0, \mu_n)$ and whose first and last marginals satisfy*

$$\pi_0 \leq \mu_0, \quad \pi_n \leq \mu_n.$$

Then there exists $P \in \mathcal{M}^n(\mu_0, \mu_n)$ such that $P \gg \pi$.

Proof. If $V = I_0^{n+1} \cap \Delta_n$, then π must be an identical transport and we can take P to be any element of $\mathcal{M}(\mu_0, \mu_0, \ldots, \mu_0, \mu_n)$. Thus, we may assume that V is of type (i) or (iii) in Definition 9.1, and then, by fixing $k \geq 1$, that $\mu_0 \leq_c \mu_n$ is irreducible with domain (I, J).

Using Lemma 9.3, we can find intermediate marginals μ_t with

$$\mu_0 \leq_c \mu_1 \leq_c \cdots \leq_c \mu_{n-1} \leq_c \mu_n$$

such that $\mu_t \gg \pi_t$ for all $1 \leq t \leq n-1$, and each of the steps $\mu_{t-1} \leq_c \mu_t$, $1 \leq t \leq n$ has a single irreducible domain given by (I, J) as well as (possibly) a diagonal component on $J \setminus I$. We note that V is an irreducible component of $\mathcal{M}(\mu_0, \mu_1, \ldots, \mu_n)$ as introduced after Theorem 3.1.

Let $f_t = d\pi_t/d\mu_t$ be the Radon–Nikodym derivative of the marginal at date t. For $m \geq 1$, we define the measure $\pi^m \ll \pi$ by

$$\pi^m(dx_0, \ldots, dx_n) = 2^{-m} \left(\prod_{t=1}^{n-1} \mathbf{1}_{f_t(x_t) \leq 2^m} \right) \pi(dx_0, \ldots, dx_n).$$

Then, the marginals π_t^m satisfy the stronger condition $\pi_t^m \leq \mu_t$ for $0 \leq t \leq n$. Thus, we can apply Lemma 3.3 to $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_n)$ and the irreducible component V, to find $P^m \in \mathcal{M}(\boldsymbol{\mu}) \subseteq \mathcal{M}^n(\mu_0, \mu_n)$ such that $P^m \gg \pi^m$. Noting that $\sum_{m\geq 1} 2^{-m}\pi^m \gg \pi$, we see that $P := \sum_{m\geq 1} 2^{-m} P^m \gg \pi$ satisfies the requirements of the lemma. □

Proof of Theorem 9.2. The result is deduced from Lemma 9.4 by following the argument in the proof of Theorem 3.1. □

9.2. Duality

In this section we formulate a duality theorem for the transport problem with free intermediate marginals.

Definition 9.5. Let $f : \mathbb{R}^{n+1} \to [0, \infty]$. The *primal problem* is

$$\mathbf{S}^n_{\mu_0,\mu_n}(f) := \sup_{P \in \mathcal{M}^n(\mu_0,\mu_n)} P(f) \in [0, \infty]$$

and the dual problem is

$$\mathbf{I}^n_{\mu_0,\mu_n}(f) := \inf_{(\phi,\psi,H)\in\mathcal{D}^n_{\mu_0,\mu_n}(f)} \mu_0(\phi) + \mu_n(\psi) \in [0, \infty],$$

where $\mathcal{D}^n_{\mu_0,\mu_n}(f)$ consists of all triplets (ϕ, ψ, H) such that $(\phi, \psi) \in L^c(\mu_0, \mu_n)$ and $H = (H_1, \dots, H_n)$ is $\mathbb{F}$-predictable with

$$\phi(X_0) + \psi(X_n) + (H \cdot X)_n \geq f \quad \mathcal{M}^n(\mu_0, \mu_n)\text{-q.s.}$$

i.e. the inequality holds P-a.s. for all $P \in \mathcal{M}^n(\mu_0, \mu_n)$.

The analogue of Theorem 5.2 reads as follows.

Theorem 9.6 (*Duality*). *Let $f : \mathbb{R}^{n+1} \to [0, \infty]$.*

(i) If f is upper semianalytic, then $\mathbf{S}^n_{\mu_0,\mu_n}(f) = \mathbf{I}^n_{\mu_0,\mu_n}(f) \in [0, \infty]$.
(ii) If $\mathbf{I}^n_{\mu_0,\mu_n}(f) < \infty$, there exists a dual optimizer $(\phi, \psi, H) \in \mathcal{D}^n_{\mu_0,\mu_n}(f)$.

The main step for the proof is again a closedness result. We shall only discuss the case where $\mu_0 \leq_c \mu_n$ is irreducible; the extension to the general case can be obtained along the lines of Section 4.

Proposition 9.7. *Let $\mu_0 \leq_c \mu_n$ be irreducible and let $f^m : \mathbb{R}^{n+1} \to [0, \infty]$ be a sequence of functions such that $f^m \to f$ pointwise. Moreover, let $(\phi^m, \psi^m, H^m) \in \mathcal{D}^n_{\mu_0,\mu_n}(f^m)$ be such that $\sup_m \mu_0(\phi^m) + \mu_n(\psi^m) < \infty$. Then there exist $(\phi, \psi, H) \in \mathcal{D}^n_{\mu_0,\mu_n}(f)$ such that*

$$\mu_0(\phi) + \mu_n(\psi) \leq \liminf_{m\to\infty} \mu_0(\phi^m) + \mu_n(\psi^m).$$

Proof. Let μ_t, $1 \leq t \leq n-1$ be such that $\boldsymbol{\mu} = (\mu_0, \dots, \mu_n)$ is in convex order and $\mu_{t-1} \leq_c \mu_t$ is irreducible for all $1 \leq t \leq n$; such μ_t are easily constructed by prescribing their potential functions. Setting $\boldsymbol{\phi}^m = (\phi^m, 0, \dots, 0, \psi^m)$ we have $(\boldsymbol{\phi}^m, H^m) \in \mathcal{D}^g_{\boldsymbol{\mu}}(f^m)$ and can thus apply Proposition 4.21 to obtain $(\boldsymbol{\phi}, H) \in \mathcal{D}^g_{\boldsymbol{\mu}}(f)$. The construction in the proof of that proposition yields $\phi_t \equiv 0$ for $1 \leq t \leq n-1$. Therefore, $(\phi_0, \phi_n, H) \in \mathcal{D}^n_{\mu_0,\mu_n}(f)$ and

$$\mu_0(\phi_0) + \mu_n(\phi_n) = \boldsymbol{\mu}(\boldsymbol{\phi}) \leq \liminf_{m\to\infty} \boldsymbol{\mu}(\boldsymbol{\phi}^m) = \liminf_{m\to\infty} \mu_0(\phi^m) + \mu_n(\psi^m). \quad \square$$

Proof of Theorem 9.6. On the strength of Proposition 9.7, the proof is analogous to the one of Theorem 5.2. □

9.3. Monotone transport

The analogue of our result on left-monotone transports is somewhat degenerate: with unconstrained intermediate marginals, the corresponding coupling is the identical transport in the first $n-1$ steps and the (one-step) left-monotone transport in the last step. The full result runs as follows.

Theorem 9.8. *Let $P \in \mathcal{M}^n(\mu_0, \mu_n)$. The following are equivalent:*

(i) P is a simultaneous optimizer for $\mathbf{S}^n_{\mu_0,\mu_n}(f(X_0, X_t))$ for all smooth second-order Spence–Mirrlees functions f and $1 \le t \le n$.

(ii) P is concentrated on a left-monotone set $\Gamma \subset \mathbb{R}^{n+1}$ such that

$$\Gamma^{n-1} = \{(x, \dots, x) : x \in \Gamma^0\}.$$

(iii) For $0 \le t \le n-1$, we have $P \circ (X_t)^{-1} = \mu_0$ and $P \circ (X_t, X_n)^{-1}$ is the (one-step) left-monotone transport in $\mathcal{M}(\mu_0, \mu_n)$.

There exists a unique $P \in \mathcal{M}^n(\mu_0, \mu_n)$ satisfying (i)–(iii).

Proof. A transport P as in (iii) exists and is unique, because the identical transport between equal marginals and the left-monotone transport in $\mathcal{M}(\mu_0, \mu_n)$ exist and are unique; cf. Proposition 6.3. The equivalence of (ii) and (iii) follows from the same proposition and the fact that the only martingale transport from μ_0 to μ_0 is the identity.

Let $P \in \mathcal{M}^n(\mu_0, \mu_n)$ satisfy (i). In particular, P is then an optimizer for $\mathbf{S}^n_{\mu_0,\mu_n}(f(X_0, X_n))$, which by Proposition 6.3 implies that $P_{0n} = P \circ (X_0, X_n)^{-1}$ is the (one-step) left-monotone transport in $\mathcal{M}(\mu_0, \mu_n)$. For $t = 1, \dots, n-1$, P is an optimizer for $\mathbf{S}^n_{\mu_0,\mu_n}(-\mathbf{1}_{\{X_0 \le a\}}|X_t - b|)$, for all $a, b \in \mathbb{R}$. This implies that P_{0t} transports $\mu_0|_{(-\infty,a]}$ to the minimal element of $\{\theta : \mu_0|_{(-\infty,a]} \le_c \theta \le_{pc} \mu_n\}$ in the sense of the convex order, which is $\theta = \mu_0|_{(-\infty,a]}$. Therefore, P_{0t} must be the identical transport for $t = 1, \dots, n-1$ and all but the last marginal are equal to μ_0.

Conversely, let $P \in \mathcal{M}^n(\mu_0, \mu_n)$ have the properties from (iii). Then, P is optimal for $\mathbf{S}^n_{\mu_0,\mu_n}(-\mathbf{1}_{\{X_0 \le a\}}(X_t - b)^+)$ for all $1 \le t \le n$ and this can be extended to the optimality (i) for smooth second-order Spence–Mirrlees functions as in the proof of Theorem 7.12. □

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Beiglböck, A.M.G. Cox, M. Huesmann, The geometry of multi-marginal Skorokhod embedding, 2017, Preprint arXiv:1705.09505v1.

[2] M. Beiglböck, A.M.G. Cox, M. Huesmann, Optimal transport and Skorokhod embedding, Invent. Math. 208 (2) (2017) 327–400.

[3] M. Beiglböck, A.M.G. Cox, M. Huesmann, N. Perkowski, D.J. Prömel, Pathwise super-replication via Vovk's outer measure, Finance Stoch. 21 (4) (2017) 1141–1166.

[4] M. Beiglböck, M. Goldstern, G. Maresch, W. Schachermayer, Optimal and better transport plans, J. Funct. Anal. 256 (6) (2009) 1907–1927.

[5] M. Beiglböck, P. Henry-Labordère, F. Penkner, Model-independent bounds for option prices: A mass transport approach, Finance Stoch. 17 (3) (2013) 477–501.

[6] M. Beiglböck, P. Henry-Labordère, N. Touzi, Monotone martingale transport plans and Skorokhod embedding, Stochastic Process. Appl. 127 (9) (2017) 3005–3013.
[7] M. Beiglböck, M. Huesmann, F. Stebegg, Root to Kellerer, in: Séminaire de probabilités XLVIII, in: Lecture Notes in Math., vol. 2168, Springer, Berlin, 2016, pp. 1–12.
[8] M. Beiglböck, N. Juillet, On a problem of optimal transport under marginal martingale constraints, Ann. Probab. 44 (1) (2016) 42–106.
[9] M. Beiglböck, N. Juillet, Shadow couplings, 2016, Preprint arXiv:1609.03340v1.
[10] M. Beiglböck, M. Nutz, N. Touzi, Complete duality for martingale optimal transport on the line, Ann. Probab. 45 (5) (2017) 3038–3074.
[11] D.P. Bertsekas, S.E. Shreve, Stochastic Optimal Control. The Discrete-Time Case, Academic Press, New York, 1978.
[12] S. Biagini, B. Bouchard, C. Kardaras, M. Nutz, Robust fundamental theorem for continuous processes, Math. Finance 27 (4) (2017) 963–987.
[13] B. Bouchard, M. Nutz, Arbitrage and duality in nondominated discrete-time models, Ann. Appl. Probab. 25 (2) (2015) 823–859.
[14] A.M.G. Cox, J. Obłój, Robust pricing and hedging of double no-touch options, Finance Stoch. 15 (3) (2011) 573–605.
[15] H. De March, N. Touzi, Irreducible convex paving for decomposition of multi-dimensional martingale transport plans, Ann. Probab. 47 (3) (2019) 1726–17747.
[16] F. Delbaen, W. Schachermayer, A general version of the fundamental theorem of asset pricing, Math. Ann. 300 (1994) 463–520.
[17] Y. Dolinsky, H.M. Soner, Martingale optimal transport and robust hedging in continuous time, Probab. Theory Related Fields 160 (1–2) (2014) 391–427.
[18] Y. Dolinsky, H.M. Soner, Martingale optimal transport in the Skorokhod space, Stochastic Process. Appl. 125 (10) (2015) 3893–3931.
[19] H. Föllmer, A. Schied, Stochastic Finance: An Introduction in Discrete Time, third edition, W. de Gruyter, Berlin, 2011.
[20] A. Galichon, P. Henry-Labordère, N. Touzi, A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options, Ann. Appl. Probab. 24 (1) (2014) 312–336.
[21] N. Ghoussoub, Y.-H. Kim, T. Lim, Structure of optimal martingale transport in general dimensions, Ann. Probab. 47 (1) (2019) 109–164.
[22] G. Guo, X. Tan, N. Touzi, Optimal Skorokhod embedding under finitely many marginal constraints, SIAM J. Control Optim. 54 (4) (2016) 2174–2201.
[23] P. Henry-Labordère, J. Obłój, P. Spoida, N. Touzi, Maximum maximum of martingales given marginals, Ann. Appl. Probab. 26 (1) (2016) 1–44.
[24] P. Henry-Labordère, X. Tan, N. Touzi, An explicit version of the one-dimensional Brenier's theorem with full marginals constraint, Stochastic Process. Appl. 126 (9) (2016) 2800–2834.
[25] P. Henry-Labordère, N. Touzi, An explicit martingale version of the one-dimensional Brenier theorem, Finance Stoch. 20 (3) (2016) 635–668.
[26] F. Hirsch, C. Profeta, B. Roynette, M. Yor, Peacocks and Associated Martingales, with Explicit Constructions, Springer, Milan, 2011.
[27] D. Hobson, Robust hedging of the lookback option, Finance Stoch. 2 (4) (1998) 329–347.
[28] D. Hobson, The Skorokhod embedding problem and model-independent bounds for option prices, in: Paris-Princeton Lectures on Mathematical Finance 2010, in: Lecture Notes in Math., vol. 2003, Springer, Berlin, 2011, pp. 267–318.
[29] D. Hobson, M. Klimmek, Robust price bounds for the forward starting straddle, Finance Stoch. 19 (1) (2015) 189–214.
[30] D. Hobson, A. Neuberger, Robust bounds for forward start options, Math. Finance 22 (1) (2012) 31–56.
[31] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, second ed., Springer, Berlin, 2003.
[32] N. Juillet, Stability of the shadow projection and the left-curtain coupling, Ann. Inst. Henri Poincaré Probab. Stat. 52 (4) (2016) 1823–1843.
[33] N. Juillet, Martingales associated to Peacocks using the curtain coupling, Electron. J. Probab. 23 (8) (2018) 1–29.
[34] S. Källblad, X. Tan, N. Touzi, Optimal Skorokhod embedding given full marginals and Azéma–Yor peacocks, Ann. Appl. Probab. 27 (2) (2017) 686–719.

[35] H.G. Kellerer, Duality theorems for marginal problems, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 67 (4) (1984) 399–432.
[36] A. Neufeld, M. Nutz, Superreplication under volatility uncertainty for measurable claims, Electron. J. Probab. 18 (48) (2013) 1–14.
[37] M. Nutz, F. Stebegg, Canonical supermartingale couplings, Ann. Probab 46 (6) (2018) 3351–3398.
[38] J. Obłój, The Skorokhod embedding problem and its offspring, Probab. Surv. 1 (2004) 321–390.
[39] H.M. Soner, N. Touzi, J. Zhang, Dual formulation of second order target problems, Ann. Appl. Probab. 23 (1) (2013) 308–347.
[40] F. Stebegg, Model-independent pricing of Asian options via optimal martingale transport, 2014, Preprint arXiv:1412.1429v1.
[41] V. Strassen, The existence of probability measures with given marginals, Ann. Math. Stat. 36 (1965) 423–439.
[42] X. Tan, N. Touzi, Optimal transportation under controlled stochastic dynamics, Ann. Probab 41 (5) (2013) 3201–3240.
[43] N. Touzi, Martingale inequalities, optimal martingale transport, and robust superhedging, in: Congrès SMAI 2013, in: ESAIM Proc. Surveys, vol. 45, EDP Sci., Les Ulis, 2014, pp. 32–47.