Convex duality in nonlinear optimal transport

Teemu Pennanen[*] Ari-Pekka Perkkiö[†]

October 30, 2017

Abstract

This article studies problems of optimal transport, by embedding them in a general functional analytic framework of convex optimization. This provides a unified treatment of a large class of related problems in probability theory and allows for generalizations of the classical problem formulations. General results on convex duality yield dual problems and optimality conditions for these problems. When the objective takes the form of a convex integral functional, we obtain more explicit optimality conditions and establish the existence of solutions for a relaxed formulation of the problem. This covers, in particular, the mass transportation problem and its nonlinear generalizations.

Keywords. mass transport; martingale transport; Schrödinger problem; convex duality; integral functionals

AMS subject classification codes. 46N10, 46N30

1 Introduction

Let S_t, $t = 0, \dots, T$ be Polish spaces and $S = S_0 \times \cdots \times S_T$. Let M_t and M be spaces of $\mathbb{R}^d$-valued Borel measures on S_t and S, respectively, and consider the optimization problem

$$\text{minimize} \quad \sum_{t=0}^{T} G_t^*(\lambda_t) + H^*(\lambda) \quad \text{over } \lambda \in M, \tag{D}$$

where G_t^* and H^* are convex functions on M_t and M, respectively and λ_t is the marginal of λ on S_t.

The above covers a wide range of optimization problems encountered in probability theory and finance. In particular, when $T = d = 1$, $G_t^* = \delta_{\{\mu_t\}}$ and[1]

[*]Department of Mathematics, King's College London, Strand, London, WC2R 2LS, United Kingdom

[†]Mathematics Institute, Ludwig-Maximilian University of Munich, Theresienstr. 39, 80333 Munich, Germany

[1]Given a set C, its *indicator function* δ_C takes the value 0 on C and $+\infty$ outside of C.

$H^*(\lambda) = \int_S c d\lambda + \delta_{M_+}(\lambda)$ for given $\mu_t \in M_t$ and a lower semicontinuous nonnegative function c, we cover the classical Monge–Kantorovich mass transportation problem. Choosing $H^* = \delta_\Lambda$ for a closed convex set $\Lambda \subset M$ of probability measures, we obtain the problem from Strassen [21] of finding probability measures with given marginals. When $H^*(\lambda)$ is the entropy relative to a given reference measure, we recover the classical Schrödinger problem; see e.g. [5, 7] and the references therein. Problems where the effective domain of H^* is contained in the set of martingale measures have been recently proposed in mathematical finance e.g. in [2].

Allowing for more general choices of G_t^* is relevant e.g. in economic applications where λ_t is not necessarily fixed but can react to demand with an increasing marginal costs of production. In the case of finite S, such problems have been extensively studied in [17]. In the financial context of [2], more general convex functionals G_t^* arise naturally when price quotes for derivatives come with bid-ask spreads and finite quantities.

This paper develops a duality theory for (D) by embedding it in the general conjugate duality framework of Rockafellar [16]. This provides a unified treatment of a wide range of problems in deriving optimality conditions and criteria for the existence of optimal solutions. The duality approach yields simplified proofs and generalizations of many classical results in applied probability.

As examples, we extend some well-known results on the existence of probability measure with given marginals, on the Schrödinger problem and on model-free superhedging of financial derivatives. Our main theorem on problem (D) yields extensions of the main results of [21], [5] and [2] to models with general marginal functionals G_t^*.

When the functions G_t^* and H^* have the additional structure of integral functionals, the optimality conditions allow for pointwise characterizations and the problem dual to (D) allows for a relaxation where the optimum is attained under fairly general conditions. Our existence results extend the existing results on the dual of the Monge–Kantorovich problem to a wider class of problems. In particular, we obtain a necessary and sufficient conditions for optimal transportation plans in mass transportation with capacity constraints. We obtain a similar result for the Schrödinger problem which also seems new.

This paper combines techniques from convex analysis, measure theory and the theory of integral functionals of continuous functions. The general duality results are derived from the functional analytic framework of [16] while the theory of integral functionals allows for a more explicit form of optimality conditions and for a relaxation of the problem dual to (D). The generality of our setting requires an extended conjugacy theorem for integral functionals proved in the appendix. The attainment of the minimum in the dual of (D) is established by borrowing techniques from convex stochastic optimization [10].

2 Conjugate duality

This section derives (D) as a dual problem of a convex optimization problem on a Banach space of continuous functions. In some applications it is convenient to allow for unbounded continuous functions so we will follow [21] and allow for continuous functions that become bounded when scaled by a possibly unbounded continuous function.

Given a continuous $\psi_t : S_t \to [1, \infty)$,

$$C_t := \{x_t \in C(S_t; \mathbb{R}^d) \mid x_t/\psi_t \in C_b(S_t; \mathbb{R}^d)\}$$

is a Banach space under the norm $\|x_t\|_{C_t} := \|x_t/\psi_t\|_{C_b(S_t;\mathbb{R}^d)}$, where $C_b(S_t; \mathbb{R}^d)$ is the space of bounded continuous functions with the supremum norm. The space M_t of $\mathbb{R}^d$-valued finite Borel measures under which ψ_t is integrable may be identified with a linear subspace of the norm dual C_t^* of C_t. Indeed, for every $\lambda_t \in M_t$,

$$x_t \mapsto \int_{S_t} x_t d\lambda_t := \sum_{i=1}^d \int_{S_t} x_t^i d\lambda_t^i = \sum_{i=1}^d \int_{S_t} x_t^i/\psi_t d(\psi_t \lambda_t^i)$$

is a continuous linear functional on C_t. If S_t is compact, then Riesz representation (see e.g. [3, Theorem 7.10.4]) implies that $C_t^* = M_t$ but, in general, the inclusion $M_t \subseteq C_t^*$ may be strict. Similarly, defining

$$\psi(s) := \sum_{t=0}^T \psi_t(s_t),$$

the space M of finite $\mathbb{R}^d$-valued Borel measures on S under which ψ is integrable is a linear subspace of the Banach dual of

$$C := \{u \in C(S; \mathbb{R}^d) \mid u/\psi \in C_b(S; \mathbb{R}^d)\}.$$

When ψ_t are bounded, we have $C_t = C_b(S_t; \mathbb{R}^d)$ and $C = C_b(S; \mathbb{R}^d)$ the duals of which contain all finite $\mathbb{R}^d$-valued Borel measures on S_t and S, respectively.

Let G_t be a proper convex function on C_t, $t = 0, \ldots, T$, let H be a proper convex function on C, and consider the problem

$$\text{minimize} \quad \sum_{t=0}^T G_t(x_t) + H\left(-\sum_{t=0}^T x_t \circ \pi_t\right) \quad \text{over } x \in \prod_{t=0}^T C_t, \tag{P}$$

where $x = (x_t)_{t=0}^T$ and $\pi_t(s) := s_t$. The general duality results below depend on the properties of the *optimum value function*.

$$\varphi(u) := \inf_x \left\{\sum_{t=0}^T G_t(x_t) + H\left(u - \sum_{t=0}^T x_t \circ \pi_t\right)\right\}$$

defined on C.

Throughout, we will endow the dual space C^* of C by the weak*-topology. The spaces C and C^* are then in separating duality under the natural bilinear form

$$\langle u, \lambda \rangle := \lambda(u).$$

Similarly for C_t^*. It turns out that the conjugate

$$\varphi^*(\lambda) := \sup_{u \in C} \{\langle u, \lambda \rangle - \varphi(u)\}$$

of φ can be expressed as

$$\varphi^*(\lambda) = \sum_{t=0}^{T} G_t^*(\lambda_t) + H^*(\lambda),$$

where G_t^* is the conjugate of G_t, H^* is the conjugate of H and $\lambda_t \in C_t^*$ denotes the continuous linear functional $x_t \mapsto \langle x_t \circ \pi_t, \lambda \rangle$ on C_t, the t-th *marginal* of λ.

The infimum of φ^* over C^* equals $-\varphi^{**}(0)$ so if φ is lower semicontinuous and the optimum value inf (P) of (P) is finite, then the biconjugate theorem implies that $-\inf$ (P) equals the optimum value of

$$\text{minimize} \quad \sum_{t=0}^{T} G_t^*(\lambda_t) + H^*(\lambda) \quad \text{over } \lambda \in C^*. \tag{DR}$$

This may be viewed as a "relaxation" of (D) from the space M of Borel measures to all of C^*. Clearly, if $\operatorname{dom} \varphi^* \subseteq M$, then (DR) coincides with (D). The following lemma gives a sufficient condition for this. It is a simple extension of [6, Lemma 4.10] that was formulated for $T = 1$ and $\psi_t \equiv 1$.

Lemma 1. *If* $\operatorname{dom} \varphi^* \subset C_+^*$ *and* $\operatorname{dom} G_t^* \subseteq M_t$ *for each* $t = 0, \dots, T$*, then* $\operatorname{dom} \varphi^* \subseteq M$*.*

Proof. By [3, Theorem 7.10.6], $\lambda \in \operatorname{dom} \varphi^*$ is a Radon measure (since S is Polish, this is equivalent to being a Borel measure [3, Theorem 7.1.7]) if and only if, for every $\epsilon > 0$, there exists a compact $K \subset S$ such that if $u \in C_b$ is zero on K, then $|\langle u, \lambda \rangle| \leq \epsilon \|u\|$.

Let $\epsilon > 0$. By assumption, $\lambda_t \in M_t$ and they are nonnegative since $\operatorname{dom} \varphi^* \subset C_+^*$. By [3, Theorem 7.1.7], there exist compact sets K_t such that $\lambda_t(K_t^C) < \epsilon/(T+1)$. Let $u \in C_b$ be zero on $\prod K_t$. Since λ is an additive set function, and $|u| \leq 1_{(\prod K_t)^C} \|u\|_{C_b}$,

$$\begin{aligned}
|\langle u, \lambda \rangle| &\leq \int 1_{(\prod K_t)^C} \|u\| d\lambda \\
&= \lambda \left(\bigcup \pi_t^{-1}(K_t)^C \right) \|u\| \\
&\leq \sum_t \lambda \left(\pi_t^{-1}(K_t)^C \right) \|u\| \\
&\leq \sum_t \lambda_t \left(K_t^C \right) \|u\| \\
&= \epsilon \|u\|
\end{aligned}$$

which completes the proof. □

The set of relaxed dual solutions coincides with the *subdifferential* $\partial\varphi(0)$ of φ at the origin. If $\partial\varphi(0)$ is nonempty, then φ is closed at the origin and there is no duality gap. The following result gives a sufficient condition for the existence in (DR). It involves the domain

$$\operatorname{dom}\varphi = \operatorname{dom} H + \{\sum_{t=0}^T x_t \circ \pi_t \mid x_t \in \operatorname{dom} G_t\}$$

of the optimum value function of (P)

Theorem 2. *If G_t and H be proper lsc functions such that the set*

$$\bigcup_{\alpha>0} \alpha \operatorname{dom}\varphi$$

is a nonempty closed linear subspace of C, then the optimum in (DR) *is attained, there is no duality gap and an x solves* (P) *if and only if there is a $\lambda \in C^*$ such that*

$$\begin{aligned} \partial G_t(x_t) &\ni \lambda_t, \quad t = 0, \dots, T, \\ \partial H(-\sum_{t=0}^T x_t \circ \pi_t) &\ni \lambda, \end{aligned}$$

and then λ solves (DR).

Proof. Problem (P) fits the conjugate duality framework of [16] with $X = \prod_{t=0}^T C_t$, $U = C$ and

$$F(x,u) := \sum_{t=0}^T G_t(x_t) + H\left(u - \sum_{t=0}^T x_t \circ \pi_t\right).$$

The associated *Lagrangian* L is the convex-concave function defined for each $x \in \prod_{t=0}^T C_t$ and $\lambda \in M$ by

$$\begin{aligned} L(x,\lambda) &:= \inf_u \{F(x,u) - \langle u, \lambda\rangle\} \\ &= \sum_{t=0}^T G_t(x_t) - \sum_{t=0}^T \langle x_t \circ \pi_t, \lambda\rangle - H^*(\lambda) \\ &= \sum_{t=0}^T G_t(x_t) - \sum_{t=0}^T \langle x_t, \lambda_t\rangle - H^*(\lambda). \end{aligned}$$

The conjugate of F can thus be expressed for each $\theta \in \prod_t M_t$ and $\lambda \in M$ as

$$\begin{aligned}F^*(\theta,\lambda) &= \sup_x\{\langle x,\theta\rangle - L(x,\lambda)\}\\ &= \sup_x\{\sum_{t=0}^T\langle x_t,\theta_t\rangle - \sum_{t=0}^T G_t(x_t) + \sum_{t=0}^T\langle x_t,\lambda_t\rangle + H^*(\lambda)\}\\ &= \sum_{t=0}^T G_t^*(\lambda_t+\theta_t) + H^*(\lambda).\end{aligned}$$

Thus

$$\varphi^*(\lambda) = F^*(0,\lambda) = \sum_{t=0}^T G_t^*(\lambda_t) + H^*(\lambda).$$

By [23, Theorem 2.7.1(vii)], φ is continuous at the origin relative to aff dom ϕ, so $\partial\varphi(0) \neq \emptyset$ by [23, Theorem 2.4.12]. The claims now follow from Theorems 15 and 16 of [16]. □

Remark 3. The second condition in Theorem 2 holds, in particular, if $0 \in \operatorname{int}\operatorname{dom}\varphi$, which holds, in particular, if

$$0 \in \operatorname{int}\operatorname{dom} H + \{\sum_{t=0}^T x_t \circ \pi_t \,|\, x_t \in \operatorname{dom} G_t\}.$$

In the scalar case $d = 1$ this last condition holds, in particular, if H is nondecreasing with $H(0) < \infty$ and there exist $x_t \in \operatorname{dom} G_t$ such that $\sum_{t=0}^T x_t \circ \pi_t \geq \epsilon\psi$ for some $\epsilon > 0$. This is satisfied e.g. in the applications of Section 7 below where $\operatorname{dom} G_t = C_t$ for all t.

The general results in conjugate duality would also give sufficient conditions for the existence of primal solutions but in many applications, the primal optimum is not attained in $\prod_{t=0}^T C_t$. In Sections 5 and 6 below, we will extend the domain of definition of the primal objective and give sufficient conditions for the attainment of the primal optimum in a larger space of measurable functions.

3 Examples

This section illustrates the general results of Section 2 by extending three well-known results in measure theory and mathematical finance. From now on, we will use the simplified notation

$$\sum_{t=0}^T x_t := \sum_{t=0}^T x_t \circ \pi_t.$$

3.1 Probability measures with given marginals

The first application deals with the classical problem on the existence of probability measures with given marginals. The following extends the existence result of [21] by allowing for more general conditions on the marginals. As usual, the *support function* of a set D in a locally convex space X is the lower semicontinuous convex function σ_D on the dual space V of X given by

$$\sigma_D(v) := \sup_{x\in D}\langle x, v\rangle.$$

Theorem 4. *Let $\Lambda \subset M$ and $\Lambda_t \subset M_t$ be weakly compact and convex. There exists $\lambda \in \Lambda$ with $\lambda_t \in \Lambda_t$ if and only if*

$$\sum_{t=0}^{T} \sigma_{\Lambda_t}(x_t) + \sigma_\Lambda\left(-\sum_{t=0}^{T} x_t\right) \geq 0 \qquad \forall x \in \prod_{t=0}^{T} C_t.$$

Proof. This fits Theorem 2 with $H = \sigma_\Lambda$ and $G_t = \sigma_{\Lambda_t}$. Indeed, by the biconjugate theorem (see e.g. [16, Theorem 5]), we then have $H^* = \delta_\Lambda$ and $G_t^* = \delta_{\Lambda_t}$, so the objective of (D) is simply the indicator of the set

$$\{\lambda \in \Lambda \,|\, \lambda_t \in \Lambda_t\}.$$

The existence is thus equivalent to the optimum value of (D) being equal to zero. Since Λ is bounded, $\operatorname{dom}\varphi = C$, so the domain condition of the Theorem 2 is satisfied. Thus, there is no duality gap so $\inf(D) = 0$ if and only if $\inf(P) = 0$, which holds exactly when the condition in the statement holds. □

When $T = d = 1$, Λ is a subset of probability measures and $\Lambda_t = \{\mu_t\}$ for given probability measures μ_t on S_t, Theorem 4 reduces to Theorem 7 of [21].

3.2 Schrödinger problem

Let $d = 1$ and let $R \in M$ and $\mu_t \in M_t$ be probability measures. The associated *Schrödinger problem* is the convex minimization problem

$$\begin{aligned}&\text{minimize} \quad \int_S \ln(d\lambda/dR)d\lambda \quad \text{over } \lambda \in M_+(S)\\ &\text{subject to} \quad \lambda \ll R, \quad \lambda_t = \mu_t \quad t = 0, \ldots, T.\end{aligned}$$

Such problems have been extensively studied in the literature; see e.g. [4] and the references there.

This fits the format of (P) with $G_t(x_t) = \int_{S_t} x_t d\mu_t$ and

$$H(u) = \ln \int_S e^u dR.$$

Indeed, H is proper convex lsc function with the conjugate

$$H^*(\lambda) = \begin{cases}\int_S \ln(d\lambda/dR)d\lambda & \text{if } \lambda \in \mathcal{P},\\ +\infty & \text{otherwise,}\end{cases}$$

where $\mathcal{P} \subset M$ is the set of probability measures. The expression of the conjugate is derived e.g. in [15, Section 3] under the assumption that S is a compact Hausdorff space and $\psi = 1$. Combined with with Theorem 9 below, the same argument works in the case of Polish S and general ψ.

Allowing for general proper lsc convex G_t, gives rise to the following generalized formulation of the Schrödinger problem

$$\begin{aligned} &\text{minimize} \quad \sum_{t=0}^{T} G_t^*(\lambda_t) + \int_S \ln(d\lambda/dR) d\lambda \quad \text{over } \lambda \in M_+(S) \\ &\text{subject to} \quad \lambda \ll R. \end{aligned} \tag{1}$$

This allows for situations where the marginals are not known exactly. Theorem 2 combined with Remark 3 gives the following.

Theorem 5. *Assume that there exist $x_t \in \operatorname{dom} G_t$ such that $\sum_{t=0}^{T} x_t \geq \epsilon\psi$ for some $\epsilon > 0$. Then the optimum in (1) is attained and the optimum value coincides with the negative of the optimum value of*

$$\text{minimize} \quad \sum_{t=0}^{T} G_t(x_t) + \ln \int_S \exp\left(-\sum_{t=0}^{T} x_t\right) dR \quad \text{over } x \in \prod_{t=0}^{T} C_t.$$

When $T = 1$ and $G_t(x_t) = \int_{S_t} x_t d\mu_t$, we recover the dual of the Schrödinger problem studied in [7]. In Section 7.3 below, we will associate (7.3) with another dual problem for which the optimum is attained. This yields necessary and sufficient conditions for the minimizers of the Schrödinger problem. This provides a duality proof of the optimality conditions given in [5, Theorem 3.43].

3.3 Model-independent superhedging

Let $d = 1$, $S_t = \mathbb{R}^n$ and $\psi_t(s_t) = 1 + |s_t|$ for all t and $H = \delta_{\mathcal{C}_{\hat{u}}}$, where

$$\mathcal{C}_{\hat{u}} := \{u \in C \,|\, \exists z \in \mathcal{N} : \; \hat{u}(s) + u(s) \leq \sum_{t=0}^{T-1} z_t(s^t) \cdot \Delta s_{t+1}\}$$

for an upper semicontinuous function $\hat{u}$ and

$$\mathcal{N} := \{(z_t)_{t=0}^{T-1} \,|\, z_t \in \mathcal{L}^\infty(S^t; \mathbb{R}^n) \quad t = 0, \dots, T\},$$

where $S^t := S_0 \times \cdots \times S_t$. Problem (P) becomes

$$\begin{aligned} &\text{minimize} \quad \sum_{t=0}^{T} G_t(x_t) \quad \text{over} \quad x \in \prod_{t=0}^{T} C_t, \; z \in \mathcal{N} \\ &\text{subject to} \quad \hat{u}(s) \leq \sum_{t=0}^{T-1} z_t(s^t) \cdot \Delta s_{t+1} + \sum_{t=0}^{T} x_t(s_t) \quad \forall s. \end{aligned} \tag{2}$$

This can be interpreted as a problem of optimal superhedging $\hat{u}$ in a financial market where G_t gives the cost of buying an s_t-dependent cash-flow x_t paid out at time t and the sum involving z represents the gains from a self-financing trading strategy described by z. When

$$G_t(x_t) = \int_{S_t} x_t d\mu_t$$

for given probability measures μ_t, we recover the superhedging problem studied in [2]. Nonlinear functions G_t arise naturally in practice where one faces bid-ask spreads and price quotes are available only for finite quantities.

We will denote the set of nonnegative *martingale measures* by

$$\mathcal{M} := \{\lambda \in M_+ \mid \int_S \sum_{t=0}^{T} z_t(s^t) \cdot \Delta s_{t+1} d\lambda = 0 \quad \forall z \in \mathcal{N}\}.$$

Lemma 6. *Assume that $\hat{u} \leq K\psi$ for some $K \in \mathbb{R}$. Then for $\lambda \in M$, the conjugate of H can be expressed as*

$$H^*(\lambda) = \begin{cases} -\int_S \hat{u} d\lambda & \textit{if } \lambda \in \mathcal{M}, \\ +\infty & \textit{otherwise.} \end{cases}$$

Proof. It is clear that $H^*(\lambda) = +\infty$ unless $\lambda \geq 0$. For $\lambda \geq 0$,

$$H^*(\lambda) \leq \sup_{z \in \mathcal{N}} \int_S \sum_{t=0}^{T} z_t(s^t) \cdot \Delta s_{t+1} d\lambda - \int_S \hat{u} d\lambda = \begin{cases} -\int_S \hat{u} d\lambda & \text{if } \lambda \in \mathcal{M}, \\ +\infty & \text{otherwise.} \end{cases}$$

On the other hand,

$$H^* \geq \sigma_\Gamma + \sigma_{\mathcal{C}_0^c},$$

where $\Gamma = \{u \in C \mid u \leq -\hat{u}\}$ and

$$\mathcal{C}_0^c := \{u \in C \mid \exists z \in \tilde{\mathcal{N}} : \; u(s) \leq \sum_{t=0}^{T-1} z_t(s^t) \cdot \Delta s_{t+1}\}$$

with $\tilde{\mathcal{N}} \subset \mathcal{N}$ denoting the continuous bounded strategies. When $\hat{u} \leq K\psi$ for some $K \in \mathbb{R}$, then for $\lambda \geq 0$,

$$\sigma_\Gamma(\lambda) = -\int_S \hat{u} d\lambda,$$

by Theorem 9 below. By standard approximation arguments, $\sigma_{\mathcal{C}_0^c} = \delta_{\mathcal{M}}$ (see e.g. [21, page 435] or [2, Lemma 2.3]). □

When $\operatorname{dom} G_t^* \subset M_t$ for all $t = 0, \dots, T$, the feasible dual solutions are in M, by Lemma 1, so problem (D) can be written as

$$\text{minimize} \quad \sum_{t=0}^{T} G_t^*(\lambda_t) - \int_S \hat{u} d\lambda \quad \text{over} \quad \lambda \in \mathcal{M}. \tag{3}$$

Combining Theorem 2 with Remark 3 gives the following.

Theorem 7. *Assume that* $\operatorname{dom} G_t^* \subset M_t$ *for all* $t = 0, \dots, T$, *that* $\hat{u} \le K\psi$ *for some* $K \in \mathbb{R}$ *and that* (2) *remains feasible when* $\hat{u}$ *is increased by* $\epsilon\psi$ *for some* $\epsilon > 0$. *Then the optimum in* (3) *is attained and the optimum value coincides with the negative of the optimum value of* (2).

When $G_t(x_t) = \int x_t d\mu_t$ for given $\mu_t \in M_t$, the feasibility condition is trivially satisfied and we recover Theorem 1.1 of [2]. In fact, Theorem 7 is slightly sharper than [2, Theorem 1.1] since we obtain the absence of a duality gap for continuous functions x_t.

We denote by C^c the subset of convex functions in C. Allowing for unbounded continuous functions is essential here as the only bounded convex functions are the constant functions. The following corollary of Theorem 7 extends [21, Theorem 8] on the existence of martingale measures with given marginals.

Corollary 8. *Let* $\Lambda_t \subset M_t$ *be weakly closed convex sets of probability measures. There exists* $\lambda \in \mathcal{M}$ *with* $\lambda_t \in \Lambda_t$ *if and only if*

$$\sum_{t=0}^{T} \sigma_{\Lambda_t}(w_t - w_{t+1}) \ge 0$$

for all $w \in \prod_{t=0}^{T+1} C_t^c$ *with* $w_0 \ge 0$ *and* $w_{T+1} = 0$.

Proof. Let $G_t = \sigma_{\Lambda_t}$ and $\hat{u} = 0$ in Theorem 7. Given an $x \in \prod_{t=0}^{T} C_t$, define $w_r \in C_r$ for $r = 0, \dots, T$ by

$$w_r(s_r) := \sum_{t=r}^{T} x_t(s_r).$$

If $w_0 \ge 0$ and w_r is convex for each r, then x is feasible. Indeed, if for some r,

$$0 \le \sum_{t=0}^{r-1} z_t(s^t) \cdot \Delta s_{t+1} + \sum_{t=0}^{r-1} x_t(s_t) + w_r(s_r) \quad \forall s \in S \tag{H_r}$$

and we choose $-z_r(s^r) \in \partial w_{r+1}(s_r)$, then

$$0 \le z_r(s_r) \cdot \Delta s_{r+1} + w_{r+1}(s_{r+1}) - w_{r+1}(s_r),$$

which combined with (H_r) gives (H_{r+1}). For $r = 0$, (H_r) simply means $w_0 \ge 0$.

On the other hand, since σ_{Λ_t} are nondecreasing, it is optimal to choose x_t so that w_r are convex. Indeed, for $r = T$ this is clear as (H_r) implies that the optimal x_T is given as a pointwise supremum of affine functions of s_T whose gradients are in $\mathcal{L}^\infty$. If w_t is convex for $t > r$, then (H_r) is necessary and sufficient for feasibility so it is optimal to choose w_r as small as possible subject to (H_r), which again means that w_r is convex. Moreover, since z_t are bounded, $w_r \in C_r^c$. The optimum value thus equals that of

$$\begin{aligned} &\text{minimize} \quad \sum_{t=0}^{T} G_t(w_t - w_{t+1}) \quad \text{over} \quad w \in \prod_{t=0}^{T} C_t^c, \\ &\text{subject to} \quad w_0 \ge 0, \end{aligned}$$

where $w_{T+1} := 0$. □

Note that if $\Lambda_t = \{\mu_t\}$ for each t, then

$$\sum_{t=0}^{T} \sigma_{\Lambda_t}(w_t - w_{t+1}) = \sum_{t=0}^{T} \int_{\mathbb{R}^n} w_t d(\mu_t - \mu_{t-1}),$$

and Corollary 8 reduces to [21, Theorem 8], which says that there exists a martingale measure with marginals μ_t if and only if μ_t are in convex order.

4 Integral functionals

From now on, we assume extra structure on G_t and H that will

1. allow us to write the optimality conditions in a more explicit pointwise form,
2. suggests a natural relaxation of problem (P) to a larger space of measurable functions where the infimum is more likely to be attained.

More precisely, we assume that each G_t is an integral functional of the form

$$G_t(x_t) = \int_{S_t} g_t(x_t(s_t), s_t) d\mu_t(s_t) + \delta_{C(D_t)}(x_t),$$

where μ_t is a probability measure on S_t, g_t is *a convex $\mathcal{B}(S_t)$-normal integrand*[2] *on $\mathbb{R}^d$*, $D_t(s_t) := \operatorname{cl} \operatorname{dom} g_t(\cdot, s_t)$ and

$$C(D_t) := \{u \in C_t \mid u(s_t) \in D_t(s_t) \ \forall s_t \in S_t\}$$

is the set of selections of D_t. Similarly, we assume that

$$H(u) = \int_S h(u(s), s) d\mu(s) + \delta_{C(D)}(u)$$

where μ is a probability measure on S, h is a convex $\mathcal{B}(S)$-normal integrand on $\mathbb{R}^d$ and $D(s) =: \operatorname{cl} \operatorname{dom} h(\cdot, s)$.

We define a function h^∞ on $\mathbb{R}^d \times S$ by setting $h^\infty(\cdot, s)$ equal to the recession function of $h(\cdot, s)$. Recall that the recession function of a lsc convex function k is given by

$$k^\infty(x) := \sup_{\alpha > 0} \frac{k(\bar{x} + \alpha x) - k(\bar{x})}{\alpha},$$

which is independent of the choice $\bar{x} \in \operatorname{dom} k$; see [14, Theorem 8.5]. By [18, Exercise 14.54(a)], h^∞ is a convex $\mathcal{B}(S)$-normal integrand on $\mathbb{R}^d$. Recall that a set-valued mapping $D : S \rightrightarrows \mathbb{R}^d$ is *inner semicontinuous* (isc) if the inverse

[2]This means that the set-valued mapping $s_t \mapsto \{(x_t, \alpha) \in \mathbb{R}^d \times \mathbb{R} \mid g_t(x_t, s_t) \leq \alpha\}$ is $\mathcal{B}(S_t)$-measurable and closed convex-valued; see e.g. [18, Chapter 14].

image under D of every open set in $O \subset \mathbb{R}^d$ is open in S, the inverse image being defined by

$$D^{-1}(O) := \{s \in S \,|\, D(s) \cap O \neq \emptyset\}.$$

The following result characterizes the conjugate and the subdifferential of H. Its proof can be found in the appendix.

Given a $\lambda \in M$, we denote its absolutely continuous and singular parts, respectively, with respect to μ by λ^a and λ^s. The *normal cone* of $D(s)$ at a point u is defined as the subdifferential of $\delta_{D(s)}$ at u. More explicitly, it is the closed convex cone $N_{D(s)}(u)$ given by

$$N_{D(s)}(u) = \{y \in \mathbb{R}^d \,|\, (u' - u) \cdot y \leq 0 \quad \forall u' \in D(s)\}$$

for $u \in D(s)$ and $N_{D(s)}(u) = \emptyset$ for $u \notin D(s)$.

Theorem 9. *Assume that* $D(s) := \operatorname{dom} h(\cdot, s)$ *is isc,* $\operatorname{cl}\operatorname{dom} H = C(D)$ *and that* H *is finite and continuous at some* $u \in C$*. Then* H *is a proper convex lsc function and the restriction to* M *of its conjugate is given by*

$$H^*(\lambda) = \int_S h^*(d\lambda^a/d\mu)d\mu + \int_S (h^*)^\infty(d\lambda^s/d|\lambda^s|)d|\lambda^s|.$$

Moreover, $\lambda \in \partial H(u) \cap M$ *if and only if*

$$\begin{aligned} d\lambda^a/d\mu &\in \partial h(u) \quad \mu\text{-a.e.} \\ d\lambda^s/d|\lambda^s| &\in N_D(u) \quad |\lambda^s|\text{-a.e.} \end{aligned}$$

If $\operatorname{dom} H = C$*, then* $\operatorname{dom} H^* \subseteq \{\lambda \in M \mid \lambda \ll \mu\}$*.*

Combining Theorem 9 with Lemma 1 gives the following.

Corollary 10. *Assume that* H *and* G_t *all satisfy the assumptions of Theorem 9 and that* $\operatorname{dom} H = C$ *or* $\operatorname{dom} G_t = C_t$ *for all* $t = 0, \ldots, T$*. Then* $\operatorname{dom} \varphi^* \subset M$*,* $\lambda_t \ll \mu_t$ *for all* $\lambda \in \operatorname{dom} \varphi^*$ *and the optimality conditions in Theorem 2 can be written as*

$$\begin{aligned} d\lambda_t/d\mu_t &\in \partial g_t(x_t) \quad \mu_t\text{-a.e.} \quad t = 0, \ldots, T \\ d\lambda^a/d\mu &\in \partial h(-\sum_{t=0}^T x_t) \quad \mu\text{-a.e.} \\ d\lambda^s/d|\lambda^s| &\in N_D(-\sum_{t=0}^T x_t) \quad |\lambda^s|\text{-a.e..} \end{aligned}$$

The optimality conditions characterize the optimal primal-dual pairs of solutions but in many applications, the primal optimum is not attained in the space of continuous functions. This motivates a relaxation of the primal problem to a larger space where the optimal solutions are more likely to exist.

5 Relaxation of the primal problem

In general, primal solutions do not exist in the space of continuous functions but we will establish the existence of solutions in a larger space of measurable functions when the functionals G_t and H are integral functionals as in Section 4 above and μ_t is the t-th marginal of μ.

More precisely, we study the problem

$$\text{minimize} \quad \int_S \left[\sum_{t=0}^T g_t(x_t) + h(-\sum_{t=0}^T x_t)\right] d\mu \quad \text{over} \quad x \in \Phi, \tag{PR}$$

where

$$\Phi := \{x \in \prod_{t=0}^T \mathcal{L}_t^0 \,|\, x_t \in D_t,\ -\sum_{t=0}^T x_t \in D \quad (\mu_t)_{t=0}^T\text{-a.e.}\},$$

where $\mathcal{L}_t^0 := \mathcal{L}^0(S_t, \mathcal{B}(S_t); \mathbb{R}^d)$ and $(\mu_t)_{t=0}^T$-almost everywhere means that the property holds on a Cartesian product of sets of full measure on S_t.

Lemma 11. *If $A \in \mathcal{B}(S)$ occurs (μ_t)-almost everywhere then $\mu(A) = 1$.*

Proof. By definition, A occurs (μ_t)-almost everywhere if there exist $A_t \in S_t$ with $\mu_t(A_t) = 1$ such that $\prod_{t=0}^T A_t \subset A$. Noting that $\prod_{t=0}^T A_t = \bigcap_{t=0}^T \pi_t^{-1}(A_t)$, where π_t is the projection $s \mapsto s_t$, we get

$$\mu((\prod_{t=0}^T A_t)^c) = \mu\left(\bigcup_{t=0}^T (\pi_t^{-1}(A_t))^c\right) \leq \sum_{t=0}^T \mu\left(\pi_t^{-1}(A_t^c)\right) = \sum_{t=0}^T \mu_t(A_t^c),$$

where $\mu_t(A_t^c) = 0$. □

Sufficient conditions for attainment of the minimum in (PR) will be given in Theorem 15 below. Clearly, the optimum value of (PR) minorizes that of (P). To guarantee that the optimum value of (PR) is still greater than $-\inf$ (D) we will assume the following.

Assumption 1. *Feasible x in* (PR) *and λ in* (D) *satisfy*

$$\int_S \left[\sum_{t=0}^T g_t(x_t)\right] d\mu < \infty \quad \textit{and} \quad \int_S h(-\sum_{t=0}^T x_t) d\mu < \infty$$

and

$$\int_S \left[\sum_{t=0}^T x_t \cdot \frac{d\lambda_t}{d\mu_t}\right] d\mu = \int_S \left[\sum_{t=0}^T x_t \cdot \frac{d\lambda}{d|\lambda|}\right] d|\lambda|.$$

Sufficient conditions for Assumption 1 will be given at the end of this section. The following statement shows that (PR) can indeed be considered as a valid dual to (D).

Theorem 12. *Assume that the normal integrands g_t and h satisfy the conditions of Corollary 10 and that Assumption 1 holds. Then*

$$\inf(\mathrm{D}) \leq \inf(\mathrm{PR}) \leq \inf(\mathrm{P})$$

and feasible solutions x in (PR) *and λ in* (D) *are optimal with* $\inf(\mathrm{PR}) = -\inf(\mathrm{D})$ *if and only if*

$$d\lambda_t/d\mu_t \in \partial g_t(x_t) \quad \mu_t\text{-a.e.} \quad t = 0, \dots, T$$

$$d\lambda^a/d\mu \in \partial h(-\sum_{t=0}^{T} x_t) \quad \mu\text{-a.e.}$$

$$d\lambda^s/d|\lambda^s| \in N_D(-\sum_{t=0}^{T} x_t) \quad |\lambda^s|\text{-a.e.}$$

Proof. Let x and λ be feasible in (PR) and (D), respectively. By Lemma 11, the condition $\lambda_t \ll \mu_t$ implies

$$x_t \in D_t, \ -\sum_{t=0}^{T} x_t \in D \quad \lambda\text{-a.e.}$$

Thus, by Fenchel's inequality,

$$g_t(x_t) + g_t^*(d\lambda_t/d\mu_t) \geq x_t \cdot (d\lambda_t/d\mu_t) \quad \mu_t\text{-a.e.} \tag{4}$$

$$h(-\sum_{t=0}^{T} x_t) + h^*(d\lambda^a/d\mu) \geq (-\sum_{t=0}^{T} x_t) \cdot (d\lambda^a/d\mu) \quad \mu\text{-a.e.} \tag{5}$$

$$(h^*)^\infty(d\lambda^s/d|\lambda^s|) \geq (-\sum_{t=0}^{T} x_t) \cdot (d\lambda^s/d|\lambda^s|) \quad |\lambda^s|\text{-a.e..} \tag{6}$$

Summing up, (4) gives

$$\sum_{t=0}^{T} g_t(x_t) + \sum_{t=0}^{T} g_t^*(d\lambda_t/d\mu_t) \geq \sum_{t=0}^{T} x_t \cdot (d\lambda_t/d\mu_t) \quad (\mu_t)\text{-a.e.,}$$

where, by Lemma 11, the inequality holds μ-almost everywhere as well. Integrating, we get

$$\int_S \left[\sum_{t=0}^{T} g_t(x_t)\right] d\mu + \int_S \left[\sum_{t=0}^{T} g_t^*(d\lambda_t/d\mu_t)\right] d\mu \geq \int_S \left[\sum_{t=0}^{T} x_t \cdot (d\lambda_t/d\mu_t)\right] d\mu$$

On the other hand, (5) and (6) give

$$\int_S h(-\sum_{t=0}^{T} x_t) d\mu + H^*(\lambda) \geq \int_S (-\sum_{t=0}^{T} x_t) \cdot \frac{d\lambda}{d|\lambda|} d|\lambda|.$$

By the first part of Assumption 1, the left hand sides of the above two inequalities are finite so

$$\int_S \left[\sum_{t=0}^T g_t(x_t) + h(-\sum_{t=0}^T x_t)\right] d\mu + \sum_{t=0}^T G_t^*(\lambda_t) + H^*(\lambda)$$
$$\geq \int_S \left[\sum_{t=0}^T x_t \cdot (d\lambda_t/d\mu_t)\right] d\mu + \int_S (-\sum_{t=0}^T x_t) \cdot \frac{d\lambda}{d|\lambda|} d|\lambda|,$$

where the right hand side vanishes by the second part of Assumption 1. Thus, $-\inf(\mathrm{D}) \leq \inf(\mathrm{PR})$. The above also shows that this holds as an equality if and only if (4)–(6) hold as equalities almost everywhere, which in turn is equivalent to the subdifferential conditions in the statement; see e.g. [14, Theorem 23.5]. □

The following lemma gives sufficient conditions for the first part of Assumption 1.

Lemma 13. *Assume that there exist $\bar{v} \in \mathcal{L}^\infty$, $\beta \in \mathcal{L}^1$ and $\delta > 0$ such that $g_t^*(\bar{v}(s), s) \leq \beta(s)$ and $h^*(\bar{v}(s) + v, s) \leq \beta(s)$ for $v \in \mathbb{R}^d$ with $|v| \leq \delta$. Then, the first part of Assumption 1 holds. If in addition, $\mu = \prod_{t=0}^T \mu_t$, then $x_t \in \mathcal{L}_t^1$ for every feasible x in* (PR).

Proof. By Fenchel's inequality,

$$\begin{aligned}\sum_{t=0}^T g_t(x_t, s_t) + h(-\sum_{t=0}^T x_t, s) &\geq \sum_{t=0}^T [\bar{v}(s) \cdot x_t - g_t^*(\bar{v}(s), s)] \\ &\quad - (\bar{v}(s) + v) \cdot \sum_{t=0}^T x_t - h^*(\bar{v}(s) + v, s) \\ &\geq -v \cdot \sum_{t=0}^T x_t - (T+2)\beta(s).\end{aligned}$$

Since this holds for any $v \in \mathbb{R}^d$ with $|v| \leq \delta$, the sum $\sum_{t=0}^T x_t$ is μ-integrable if x is feasible in (PR). By Fenchel's inequality again,

$$\sum_{t=0}^T g_t(x_t) \geq \bar{v}(s) \cdot \sum_{t=0}^T x_t - (1+T)\beta(s) \quad \text{and} \quad h(-\sum_{t=0}^T x_t) \geq \bar{v}(s) \cdot \sum_{t=0}^T x_t - \beta(s)$$

so the first part of Assumption 1 is satisfied. If $\mu = \prod_{t=0}^T \mu_t$, then by Fubini's theorem, μ-integrability of $\sum_{t=0}^T x_t$ implies that each x_t is μ_t-integrable. □

The second part of Assumption 1 clearly holds when feasible solutions x of (PR) have x_t bounded. More generally, it holds if each x_t is λ_t-integrable. This holds under all the assumptions of Lemma 13, when $d = 1$ and feasible λ

satisfy $\lambda_t = \mu_t$. This last condition holds in problems with given marginals; see Section 7 below.

In some problems it is essential not to require the integrability of x_t; see Section 7.3 below. The following lemma addresses such situations but, interestingly, the argument only works when $T = d = 1$. The idea for the proof is taken from that of [5, Corollary 3.15].

Lemma 14. *If $T = d = 1$ and feasible λ satisfies $\lambda \in M_+$ and $\lambda_t = \mu_t$, then the first part of Assumption 1 implies the second part.*

Proof. The proof of Theorem 12 shows that when $T = d = 1$, $\mu, \lambda \in M_+$ and $\lambda_t = \mu_t$, feasible solutions x and λ satisfy

$$\int_S \left[\sum_{t=0}^T x_t\right]^+ d\mu < \infty \quad \text{and} \quad \int_S \left[\sum_{t=0}^T x_t\right]^- d\lambda < \infty$$

Let x_t^ν be the pointwise projection of x_t to the unit ball of radius ν. We have

$$\int_S \left(\sum_{t=0}^T x_t^\nu\right) d\mu = \sum_{t=0}^T \int_{S_t} x_t^\nu d\mu_t = \sum_{t=0}^T \int_{S_t} x_t^\nu d\lambda_t = \int_S \left(\sum_{t=0}^T x_t^\nu\right) d\lambda$$

and when $T = 1$,

$$[\sum_{t=0}^T x_t^\nu]^+ \le [\sum_{t=0}^T x_t]^+ \quad \text{and} \quad [\sum_{t=0}^T x_t^\nu]^- \le [\sum_{t=0}^T x_t]^-$$

so, by Fatou's lemma,

$$\int_S \left(\sum_{t=0}^T x_t\right) d\mu \ge \limsup \int_S \left(\sum_{t=0}^T x_t^\nu\right) d\mu \ge \liminf \int_S \left(\sum_{t=0}^T x_t^\nu\right) d\lambda \ge \int_S \left(\sum_{t=0}^T x_t\right) d\lambda.$$

This implies

$$\int_S \left[\sum_{t=0}^T x_t\right]^- d\mu < \infty \quad \text{and} \quad \int_S \left[\sum_{t=0}^T x_t\right]^+ d\lambda < \infty,$$

so the same argument gives the reverse inequality. □

6 Existence of relaxed primal solutions

We now turn to the existence of solutions in the relaxed problem (PR). We start more abstractly by considering problems of the form

$$\text{minimize} \quad \int_S f(x) d\mu \quad \text{over} \quad x \in \Phi, \tag{$\bar{P}$}$$

where f is a convex normal $\mathcal{B}(S)$-integrand on $\mathbb{R}^{(1+T)d}$ and

$$\Phi := \left\{ x \in \prod_{t=0}^{T} \mathcal{L}_t^0 \;\middle|\; x \in \operatorname{cl}\operatorname{dom} f \quad (\mu_t)_{t=0}^T\text{-a.e.} \right\}.$$

Problem (PR) fits $(\bar{P})$ with

$$f(x,s) = \sum_{t=0}^{T} g_t(x_t, s_t) + h(-\sum_{t=0}^{T} x_t, s) \tag{7}$$

under the following.

Assumption 2. *The set*

$$\{x \in \mathbb{R}^{(1+T)d} \,|\, x_t \in \operatorname{rint} D_t(s_t),\; -\sum_{t=0}^{T} x_t \in \operatorname{rint} D(s)\}$$

is nonempty for every $s \in S$.

Indeed, by [18, Propositions 14.44(d) and 14.45(a)], f defined by (7) is a normal integrand, and, by [14, Theorem 9.3], Assumption 2 implies

$$\operatorname{cl}\operatorname{dom} f(\cdot, s) = \{x \,|\, x_t \in D_t(s),\; -\sum_{t=0}^{T} x_t \in D(s)\}.$$

Assumption 2 is automatically satisfied if $D_t(s_t) = \mathbb{R}^d$ for all t since $\operatorname{rint} D(s) \neq \emptyset$, by [14, Theorem 6.2].

Except for the filtration property, problem $(\bar{P})$ is similar to the general stochastic optimization problem studied in [10]. The following variant of [10, Theorem 2] gives sufficient conditions for the existence of solutions in $(\bar{P})$. Its proof uses [14, Corollary 8.6.1] which says that if $x \in \mathbb{R}^{(1+T)d}$ is such that $f^\infty(x,s) \leq 0$ and $f^\infty(-x,s) \leq 0$, then $f(\bar{x}+x,s) = f(\bar{x},s)$ for every $\bar{x} \in \operatorname{dom} f(\cdot,s)$. The Borel sigma-algebra generated on S by the projection of s to s_t will be denoted by $\mathcal{F}_t$.

The statements below involve the set

$$N := \{x \in \mathbb{R}^{(1+T)d} \mid \sum x_t = 0\}.$$

Theorem 15. *Assume that $\prod_{t=0}^T \mu_t \ll \mu$, there exists $m \in \mathcal{L}^1(S,\mathcal{F},\mu)$ such that*

$$f(x,s) \geq m(s) \quad \forall x \in \mathbb{R}^{(1+T)d} \;\; \mu\text{-a.e.},$$

and that for every $s \in S$

$$\{x \in \mathbb{R}^{(1+T)d} \mid f^\infty(x,s) \leq 0\} = N. \tag{8}$$

Then $(\bar{P})$ has a solution.

Proof. Let x be feasible in $(\bar{P})$. Since the set $N := \{x \in \mathbb{R}^{(1+T)d} \mid \sum x_t = 0\}$ is linear, condition (8) implies, by [14, Corollary 8.6.1], that $f(\cdot, s)$ is constant in the directions of N. Let

$$N_t := \{x_t \in \mathbb{R}^d \mid \exists z \in \mathbb{R}^{d\prod_{t=0}^T(1+T)} : (0,\dots,0,x_t,z_{t+1},\dots,z_T) \in N\}.$$

The s-wise orthogonal projection $\tilde{x}_0$ of x_0 to N_0 has an extension $\tilde{x} \in \mathcal{L}^0(\mathcal{F}_0; N)$ such that $x_0 - \tilde{x}_0 \in N_0^\perp$. Defining $\bar{x}^0 := x - \tilde{x}$, we have $f(\bar{x}^0) = f(x)$ everywhere in S and $\bar{x}^0 \in \operatorname{cl}\operatorname{dom} f$ $(\mu_t)_{t=0}^T$-almost everywhere. Repeating the argument for $t = 1, \dots, T$, we arrive at an

$$\bar{x}^T \in \prod \mathcal{L}_t^0 + \sum_{t=0}^T \mathcal{L}^0(\mathcal{F}_t; N)$$

with $f(\bar{x}^T) = f(x)$ and $\bar{x}_t^T \in N_t^\perp$ everywhere in S for all t and $\bar{x}^T \in \operatorname{cl}\operatorname{dom} f$ $(\mu_t)_{t=0}^T$-almost everywhere.

Let $(x^\nu)_{\nu=1}^\infty \subset \Phi$ such that $Ef(x^\nu) \le \inf(\bar{P}) + 2^{-\nu}$. Since $f(x^\nu)$ is bounded in $\mathcal{L}^1$, Komlos' theorem gives the existence of a subsequence of convex combinations (still denoted by (x^ν)) and $\beta \in \mathcal{L}^0$ such that $f(x^\nu) \le \beta$ almost surely.

By the first paragraph, there exists $\bar{x}^\nu$ with $f(\bar{x}^\nu) = f(x^\nu)$ and $\bar{x}_t^\nu \in N_t^\perp$ everywhere in S for all t, $\bar{x}^\nu \in \operatorname{cl}\operatorname{dom} f$ $(\mu_t)_{t=0}^T$-almost everywhere, and $\bar{x}^\nu \in \prod \mathcal{L}_t^0 + \sum_{t=0}^T \mathcal{L}^0(\mathcal{F}_t; N)$. Thus $\bar{x}^\nu \in \{x \in \mathcal{L}^0 \mid x \in \Gamma \text{ a.e.}\}$, where

$$\Gamma(s) := \{x \in \mathbb{R}^{d(T+1)} \mid x_t \in N_t^\perp,\ f(x,s) \le \beta(s)\}.$$

By Corollary 8.3.3 and Theorem 8.7 of [14], the recession cone of $\Gamma(s)$ is given by $\Gamma^\infty(s) = \{x \mid x_t \in N_t^\perp, x \in N\}$. For $x \in \Gamma^\infty(s)$, we have $x_0 \in N_t^\perp \cap N_t$, so $x_0 = 0$. Repeating the argument for $t = 1, \dots, T$, we get that $x = 0$ and so $\Gamma^\infty = \{0\}$ μ-almost everywhere. By [14, Theorem 8.4], the sequence $(\bar{x}^\nu)$ is thus almost surely bounded. By Komlos' theorem, there exists a subsequence of convex combinations and $\bar{x} \in \mathcal{L}^0$ such that $(\bar{x}^\nu) \to \bar{x}$ μ-almost everywhere. By Fatou's lemma, $\int f(\bar{x})d\mu \le \liminf \int f(\bar{x}^\nu)d\mu \le \inf(\bar{P})$.

Since $\prod_{t=0}^T \mu_t \ll \mu$, the sequence $(\bar{x}^\nu)$ converges $\prod_{t=0}^T \mu_t$-almost everywhere. By Lemma 18 below, $\bar{x}^\nu = \sum_{t=0}^T (\tilde{x}^\nu)^t$ for some μ_t-almost everywhere converging $(\tilde{x}^\nu)^t \in \mathcal{L}^0(\mathcal{B}(S_t), \mu_t)$, so $\bar{x} \in \operatorname{cl}\operatorname{dom} f$ $(\mu_t)_{t=0}^T$-almost everywhere. Let $\bar{x}^t$ be the limit of $(\tilde{x}^\nu)^t$. For every t',

$$\hat{x}^{t'} := (-\bar{x}_0^{t'}, \dots, -\bar{x}_{t'-1}^{t'}, \sum_{t \ne t'} \bar{x}_t^{t'}, -\bar{x}_{t'+1}^{t'}, \dots, -\bar{x}_T^{t'})$$

belongs to $\mathcal{L}^0(N)$, so $x := \bar{x} + \sum_{t'=0}^T \hat{x}^{t'}$ satisfies $f(x) = f(\bar{x})$. We also have that $x \in \Phi$ so x is optimal. □

Remark 16. The conclusion of Theorem 15 still holds if f is coercive in the sense that $\{x \mid f^\infty(x,s) \le 0\} = \{0\}$. In fact, the proof then simplifies considerably. A more general condition that covers both the condition of Theorem 15 as well as the coercivity condition is that there is a subset J of the indices $\{0,\dots,T\}$ such that

$$\{x \mid f^\infty(x,s) \le 0\} = \{x \mid \sum_{t\in J} x_t = 0,\ x_t = 0 \quad \forall t \notin J\}$$

for all $s \in S$.

Remark 17. When f is given by (7), we have

$$f^\infty(x,s) = \sum_{t=0}^{T} g_t^\infty(x_t,s_t) + h^\infty(-\sum_{t=0}^{T} x_t, s),$$

by [14, Theorems 9.3 and 9.5] as soon as f is proper.

The following lemma was used in the proof of Theorem 15. For $T = 1$, more general results can be found e.g. in [5]; see also [20].

Lemma 18. *If $(x^\nu) \subset \sum_{t=0}^T \mathcal{L}^0(\mathcal{F}_t, \prod_{t=0}^T \mu_t)$ converges $\prod_{t=0}^T \mu_t$-almost everywhere, then there exists μ_t-almost everywhere converging sequences $((\tilde{x}^\nu)^t) \subset \mathcal{L}^0(\mathcal{B}(S_t), \mu_t)$ such that $\sum (\tilde{x}^\nu)^t = x^\nu$.*

Proof. The statement is clearly valid for $T = 0$. We proceed by induction on T. Let $(\sum_{t=0}^T x_t^\nu)$ be a converging sequence in $\sum_{t=0}^T \mathcal{L}^0(\mathcal{F}_t, \prod_{t=0}^T \mu_t)$ and let $A \subseteq S$ be the set where the convergence holds.

Let $A_T(s_T) := \{s^{T-1} \mid (s^{T-1}, s_T) \in A\}$. Since $(\prod_{t=0}^T \mu_t)(A) = 1$, we have $\mu_T(\bar{A}_T) = 1$, where $\bar{A}_T := \{s_T \mid \prod_{t=0}^{T-1} \mu_t(A_T(s_T)) = 1\}$. Let $\bar{s}_T \in \bar{A}_T$,

$$\begin{aligned}
(\tilde{x}^\nu)^T(s_T) &= (x^\nu)^T(s_T) - (x^\nu)^T(\bar{s}_T),\\
(\tilde{x}^\nu)^{T-1}(s_{T-1}) &= (x^\nu)^{T-1}(s_{T-1}) + (x^\nu)^T(\bar{s}_T),\\
(\tilde{x}^\nu)^t &= x_t^\nu \qquad t = 0,\dots T-2,
\end{aligned}$$

so that $\sum_t (x^\nu)^t = \sum_t (\tilde{x}^\nu)^t$. We have that $\sum_{t=0}^{T-1} (\tilde{x}^\nu)^t$ converges $\prod_{t=0}^{T-1} \mu_t$-almost everywhere and, by the induction hypothesis, there exist μ_t-almost everywhere converging sequences $((\hat{x}^\nu)^t) \subset \mathcal{L}^0(\mathcal{B}(S_t), \mu_t)$ such that $\sum_{t=0}^{T-1} (\hat{x}^\nu)^t = \sum_{t=0}^{T-1} (\tilde{x}^\nu)^t$. We also get that

$$\sum_{t=0}^{T} (x^\nu)^t - \sum_{t=0}^{T-1} (\tilde{x}^\nu)^t = (\tilde{x}^\nu)^T$$

converges μ_T-almost everywhere. This completes the induction argument. □

7 Applications to problems with fixed marginals

This section illustrates the results of the previous sections in the case of fixed marginals. More precisely, will assume throughout that $d = 1$ (the measures are scalar-valued), and that G_t and H are given in terms of integral functionals with $g_t(x_t, s_t) = x_t$ for each t and $h(\cdot, s)$ nondecreasing. In this case, $g_t^*(\cdot, s_t) = \delta_{\{1\}}$ and $\operatorname{dom} H^* \subset C_+^*$ so the assumptions of Lemma 1 are satisfied and problem (D) can be written as

$$\begin{array}{lll} \text{minimize} & H^*(\lambda) & \text{over } \lambda \in M \\ \text{subject to} & \lambda_t = \mu_t & t = 0, \dots, T, \end{array} \tag{9}$$

while the relaxed primal (PR) problem becomes

$$\text{minimize} \quad \int_S \left[\sum_{t=0}^T x_t + h(-\sum_{t=0}^T x_t) \right] d\mu \quad \text{over} \quad x \in \Phi, \tag{10}$$

where

$$\Phi = \{x \in \prod_{t=0}^T \mathcal{L}_t^0 \mid -\sum_{t=0}^T x_t \in D \quad (\mu_t)_{t=0}^T\text{-a.e.}\}.$$

Combining Theorems 2, 12 and 15 gives the following.

Theorem 19. *Assume that h satisfies the assumptions of Theorem 9, that $h^*(v, \cdot)$ is μ-integrable for $v \in \mathbb{R}$ in a neighborhood of 1 and that either $\mu = \prod_{t=0}^T \mu_t$ or $T = 1$ and $\prod_{t=0}^T \mu_t \ll \mu$. Then the optima in (9) and (10) are attained, there is no duality gap and feasible solutions x and λ are optimal if and only if*

$$d\lambda^a / d\mu \in \partial h(-\sum_{t=0}^T x_t) \quad \mu\text{-a.e.},$$

$$d\lambda^s / d|\lambda^s| \in N_D(-\sum_{t=0}^T x_t) \quad |\lambda^s|\text{-a.e.}$$

If, in addition, $\mu = \prod_{t=0}^T \mu_t$, then $x_t \in \mathcal{L}_t^1$ for each feasible x in (10).

Proof. Since $\operatorname{dom} G_t = C_t$ for all t and H is nondecreasing, Theorem 2 implies that the optimum in (9) is attained and that there is no duality gap. To prove the attainment in (10), we apply Theorem 15 with

$$f(x, s) = \sum_{t=0}^T x_t + h(-\sum_{t=0}^T x_t, s).$$

Assumption 2 holds trivially since $\operatorname{dom} g_t = \mathbb{R}^d$ for each t, so (10) coincides with $(\bar{P})$. By the Fenchel inequality,

$$f(x, s) \geq (1 - v) \sum_{t=0}^T x_t - h^*(v, s), \tag{11}$$

so the integrability condition implies that the lower bound in Theorem 15 holds with $m(s) = h^*(1, s)$. This also gives

$$f^\infty(x, s) \geq (1 - v) \sum_{t=0}^T x_t$$

for v in a neighborhood of 1, so $f^\infty(x, s) \geq \epsilon | \sum_{t=0}^T x_t|$ for some $\epsilon > 0$. It follows that f satisfies (8). Thus, by Theorem 15, the optimum in (10) is attained.

By Lemma 13, the integrability condition implies that the first part of Assumption 1 holds. If $T = 1$, Lemma 14 implies that the second part of Assumption 1 is satisfied as well. If, on the other hand, $\mu = \prod_{t=0}^T \mu_t$, then, by Lemma 13, $x_t \in \mathcal{L}_t^1$ and the second part of Assumption 1 is again holds. The rest now follows from Theorem 12 by observing that, when $g_t(x_t, s_t) = x_t$, the condition $d\lambda_t/d\mu_t \in \partial g_t(x_t)$ simply means that $\lambda_t = \mu_t$. □

7.1 Monge–Kantorovich problem

Let c be a measurable function on S and let $h(u, s) = \delta_{(-\infty, c(s)]}(u)$. In this case,

$$h^*(v, s) = \begin{cases} c(s)v & \text{if } v \geq 0, \\ +\infty & \text{otherwise} \end{cases}$$

and problem (9) can be written as

$$\begin{aligned} &\text{minimize} \quad \int_S c d\lambda \quad \text{over } \lambda \in M_+ \\ &\text{subject to} \quad \lambda_t = \mu_t \quad t = 0, \dots, T. \end{aligned} \tag{12}$$

When $T = 1$, we recover the classical Monge–Kantorovich mass transportation problem; see e.g. [1], [22], [6], [13] and their references. On the other hand, if S_t coincide for all t, problem (12) can be interpreted as the problem of finding a stochastic process $X = (X_t)_{t=0}^T$ such that X_t has distribution μ_t and the expectation of $c(X)$ is minimized. It should be noted that (12) depends on μ only through its marginals μ_t. Thus, we choose

$$\mu = \prod_{t=0}^T \mu_t.$$

Problem (10) becomes

$$\text{minimize} \quad \int_S \sum_{t=0}^T x_t d\mu \quad \text{over} \quad x \in \Phi, \tag{13}$$

where

$$\Phi = \{x \in \prod_{t=0}^T \mathcal{L}_t^0 \mid -\sum_{t=0}^T x_t \leq c \quad (\mu_t)_{t=0}^T\text{-a.e.}\}.$$

Indeed, by Lemma 11, $x \in \Phi$ implies $-\sum_{t=0}^{T} x_t \in D$ μ-almost everywhere so

$$\int_S \left[\sum_{t=0}^{T} x_t + h(-\sum_{t=0}^{T} x_t)\right] d\mu = \int_S \sum_{t=0}^{T} x_t d\mu.$$

Theorem 20. *Assume that c is lower semicontinuous and μ-integrable with $c \geq K\psi$ for some $K \in \mathbb{R}$. Then the optima in* (12) *and* (13) *are attained, there is no duality gap and feasible solutions λ and x are optimal if and only if*

$$\int_S \left(c + \sum_{t=0}^{T} x_t\right) d\lambda = 0.$$

Moreover, if x is feasible in (13)*, then $x_t \in \mathcal{L}_t^1$ so the objective of* (13) *can be written as*

$$\int_S \sum_{t=0}^{T} x_t d\mu = \sum_{t=0}^{T} \int_{S_t} x_t d\mu_t.$$

Proof. We now have $D(s) = \{u \in \mathbb{R} \,|\, u \leq c(s)\}$ which is inner semicontinuous if and only if c is lower semicontinuous; see [8, Example 1.2*]. The lower bound on c implies that h satisfies the assumptions of Theorem 9. Since c is μ-integrable, all the conditions of Theorem 19 are satisfied. The form of the optimality conditions follows simply by observing that now, $\partial h = N_D$. □

Instead of the lower bound $c \geq K\psi$, [22, Theorem 5.10] assumes the existence of $c_t \in \mathcal{L}_t^1$ such that $c \geq \sum_t c_t$. However, if there is no $K \in \mathbb{R}$ such that $c \geq K\psi$, then problem (P) is infeasible so the duality argument fails and, in particular, the first conclusion of [22, Theorem 5.10] does not hold. The function c is integrable, in particular, if there exist $c_t \in \mathcal{L}_t^1$ such that $c \leq \sum_t c_t$. This latter condition is assumed e.g. in [22, Theorem 5.10] in establishing the existence of solutions.

Remark 21. Feasibility of an x means that the inequality constraint holds on a product set $A^x = A_0^x \times \cdots \times A_T^x$, where $\mu_t(A_t^x) = 1$. Thus, every dual feasible solution λ satisfies

$$\lambda((A^x)^c) \leq \sum_{t=0}^{T} \lambda_t((A_t^x)^c) = \sum_{t=0}^{T} \mu_t((A_t^x)^c) = 0.$$

The optimality conditions thus imply that the optimal dual solutions λ are supported by the sets

$$\Gamma_x := \{s \in A^x \,|\, c(s) + \sum_{t=0}^{T} x_t(s_t) \leq 0\},$$

where x runs through optimal primal solutions. The sets Γ_x are *c-monotone* in the sense that

$$\sum_{i=1}^{n} c(s_0^i, \ldots, s_T^i) \leq \sum_{i=1}^{n} c(s_0^{P_0(i)}, \ldots, s_T^{P_T(i)})$$

for any $(s_0^i, \dots, s_T^i) \in \Gamma_x$, $i = 1, \dots n$ and any permutations P_t of the indices i. Indeed,

$$\sum_{i=1}^n c(s_0^i, \dots, s_T^i) \le -\sum_{i=1}^n \sum_{t=0}^T x_t(s_t^i) = -\sum_{i=1}^n \sum_{t=0}^T x_t(s_t^{P_t(i)}) \le \sum_{i=1}^n c(s_0^{P_0(i)}, \dots, s_T^{P_T(i)}),$$

where the last inequality follows from the feasibility of x on A^x. This is a multivariate generalization of the c-cyclical monotonicity property studied e.g. in [13] and [22]. When $T = 1$, it is known that a feasible λ is optimal if it is concentrated on a c-monotone set. It would be natural to conjecture that this holds also for $T > 1$.

7.2 Capacity constraints

Let c and ϕ be nonnegative measurable functions on S and let

$$h(u, s) = \phi(s)[u - c(s)]^+.$$

We get

$$h^*(v, s) = c(s)v + \delta_{[0,\phi(s)]}(v)$$

so problem (9) can be written as

$$\begin{aligned} &\text{minimize} && \int_S c d\lambda \quad \text{over } \lambda \in M_+ \\ &\text{subject to} && \lambda \ll \mu, \quad \frac{d\lambda}{d\mu} \le \phi, \quad \lambda_t = \mu_t \quad t = 0, \dots, T. \end{aligned} \tag{14}$$

This models *capacity constraints* on the transport plan requiring $\lambda \le \phi\mu$, where the inequality is taken with respect to the natural order on M. Constrained variations of the Monge–Kantorovich problem are considered also in [13, Chapter 7]. What is called "capacity constraints" in [13, Section 7.3], however, is different from the constraints of (14). In the case of finite S, problem (14) reduces to a network flow problem where the flow on each arc of the network is bounded from above by the value of ϕ; see [17] for a comprehensive study of linear and nonlinear network flow problems.

Problem (10) becomes

$$\text{minimize} \quad \int_S \left[\sum_{t=0}^T x_t + \phi[\sum_{t=0}^T x_t + c]^- \right] d\mu \quad \text{over} \quad x \in \Phi, \tag{15}$$

where

$$\Phi = \prod_{t=0}^T \mathcal{L}_t^0.$$

Theorem 19 gives the following.

Theorem 22. *Assume that* $\mu = \prod_{t=0}^T \mu_t$ *and that* c *and* ϕ *are* μ*-integrable with* $c \geq K\psi$ *and* $\phi \geq v$ *for some* $K \in \mathbb{R}$ *and* $v > 1$*. Then the optima in* (14) *and* (15) *are attained, there is no duality gap and feasible solutions* λ *and* x *are optimal if and only if*

$$
\begin{aligned}
d\lambda/d\mu = 0 \quad &\textit{if} \quad -\sum_{t=0}^T x_t < c,\\
d\lambda/d\mu \in [0,\phi] \quad &\textit{if} \quad -\sum_{t=0}^T x_t = c,\\
d\lambda/d\mu = \phi \quad &\textit{if} \quad -\sum_{t=0}^T x_t > c.
\end{aligned}
$$

Moreover, if x *is feasible in* (15)*, then* $x_t \in \mathcal{L}^1_t$*.*

In the case of finite S, the optimality conditions in Theorem 22 correspond to the classical complementary slackness conditions in constrained network optimization problems; see [17].

7.3 Schrödinger problem

We now return to the Schrödinger problem

$$
\begin{aligned}
&\text{minimize} \quad \int_S \ln(d\lambda/dR)d\lambda \quad \text{over } \lambda \in M_+\\
&\text{subject to} \quad \lambda \ll R, \quad \lambda_t = \mu_t \quad t = 0,\dots,T
\end{aligned}
$$

studied in Section 3.2. We will derive optimality conditions and a dual problem under the assumption that there exists a feasible λ equivalent to R. Denoting the feasible point by μ and $\phi := d\mu/dR$, the problem can then be written as

$$
\begin{aligned}
&\text{minimize} \quad \int_S \frac{d\lambda}{d\mu}\ln(\phi\frac{d\lambda}{d\mu})d\mu \quad \text{over } \lambda \in M_+\\
&\text{subject to} \quad \lambda \ll \mu, \quad \lambda_t = \mu_t \quad t = 0,\dots,T.
\end{aligned}
$$

This fits the format of (9) with $h(u,s) = \frac{e^u - 1}{\phi(s)}$. Indeed, we have

$$
h^*(v,s) = \begin{cases} v\ln(\phi(s)v) - v + 1/\phi(s) & \text{if } v > 0,\\ 0 & \text{if } v = 0,\\ +\infty & \text{otherwise,} \end{cases}
$$

so that $(h^*)^\infty(\cdot,s) = \delta_{\{0\}}$ for all $s \in S$ and

$$
H^*(\lambda) = \begin{cases} \int_S \frac{d\lambda}{d\mu}\ln(\phi\frac{d\lambda}{d\mu})d\mu & \text{if } \lambda \ll \mu,\\ +\infty & \text{otherwise.} \end{cases}
$$

The relaxed primal problem becomes

$$\text{minimize} \quad \int_S \left[\sum_{t=0}^T x_t + \frac{\exp(-\sum_{t=0}^T x_t) - 1}{\phi} \right] d\mu \quad \text{over} \quad x \in \prod_{t=0}^T \mathcal{L}^0_t.$$

Note that even when restricted to $x \in \prod C_t$, the objective is different from that in Theorem 5.

Theorem 19 gives the following.

Theorem 23. *Assume that $T = 1$, $\prod_{t=0}^T \mu_t \ll R$ and that (7.3) admits a feasible solution equivalent to R. Then the optimum in (7.3) is attained and the optimal solutions λ are characterized by the existence of an $x \in \prod_{t=0}^T \mathcal{L}^0_t$ such that*

$$d\lambda/dR = \exp(-\sum_{t=0}^T x_t) \quad R\text{-a.e.}.$$

If $\prod_{t=0}^T \mu_t$ is feasible and equivalent to R, then the same conclusion holds for any T and, moreover, $x_t \in \mathcal{L}^1_t$ for feasible x in (7.3).

Proof. Since $\mu \approx R$, the condition $\prod_{t=0}^T \mu_t \ll R$ means that $\prod_{t=0}^T \mu_t \ll \mu$. The feasibility of μ in (7.3) (and the definition of ϕ) implies that the integrability condition in Theorem 19 is satisfied. It is clear that h satisfies the other conditions as well. The optimality conditions mean that $\lambda \approx \mu$ and

$$\frac{d\lambda}{d\mu} = \frac{\exp(-\sum_{t=0}^T x_t)}{\phi} \quad \mu\text{-a.e.}$$

which reduces to the one in the statement since $\mu \approx R$ and $\phi = d\mu/dR$. $\square$

The necessity and sufficiency of the optimality condition in Theorem 23 was established for feasible solutions equivalent to R in [5, Theorem 3.43] under the assumption that $R \ll \prod_{t=0}^T \mu_t$. Theorem 23 above gives the equivalence when $\prod_{t=0}^T \mu_t \ll R$ without assuming apriori the equivalence with R.

The last statement of Theorem 23 seems new. An alternative condition for the integrability of x_t is given in [19, Proposition 1]. Example 1 of [19] shows that the integrability may fail without additional conditions.

8 Appendix

In this appendix we prove Theorem 9. The proof follows the arguments in [12] which in turn are based on those in [15] and [11]. We reproduce the proofs here since we allow for unbounded scaling functions ψ_t and we do not assume that S is locally compact.

Let h be a convex normal $\mathcal{B}(S)$-integrand on $\mathbb{R}^d$, μ a nonnegative Radon measure on S and let

$$I_h(u) = \int h(u) d\mu.$$

Theorem 24. *If I_h is finite and continuous at some point on C, then I_h is lsc and I_h^* is proper and given by*

$$I_h^*(\lambda) = \min_{\lambda' \in M} \{I_{h^*}(d\lambda'/d\mu) + \sigma_{\operatorname{dom} I_h}(\lambda - \lambda') \mid \lambda' \ll \mu\}.$$

Proof. Defining the convex function $\bar{I}_h$ to L^∞ by

$$\bar{I}_h(u) = \int h(u) d\mu,$$

we have $I_h = \bar{I}_h \circ A$, where $A : C \to L^\infty(\mu)$ is the natural embedding. We equip L^∞ with the essential supremum-norm. By [15, Theorem 2], the continuity of I_h at a point $\bar{u}$ implies that $\bar{I}_h$ is proper and continuous at $A\bar{u}$. Thus, by [16, Theorem 19],

$$I_h^*(\lambda) = \inf_{\theta \in (L^\infty)^*} \{\bar{I}_h^*(\theta) \mid A^*\theta = \lambda\}.$$

By [15, Theorem 1], the conjugate of $\bar{I}_h$ on $(L^\infty)^*$ can be expressed in terms of the Yosida-Hewitt decomposition $\theta = \theta^a + \theta^s$ as

$$\bar{I}_h^*(\theta) = I_{h^*}(d\theta^a/d\mu) + \sigma_{\operatorname{dom} \bar{I}_h}(\theta^s).$$

We thus get

$$I_h^*(\lambda) = \inf_{\theta \in (L^\infty)^*} \{I_{h^*}(d\theta^a/d\mu) + \sigma_{\operatorname{dom} \bar{I}_h}(\theta^s) \mid A^*(\theta^a + \theta^s) = \lambda\}. \tag{16}$$

It suffices to show that

$$I_h^*(\lambda) = \inf_{\tilde{\theta} \in (L^\infty)^*, \theta^a \ll \mu} \{I_{h^*}(d\theta^a/d\mu) + \sigma_{\operatorname{dom} \bar{I}_h}(\tilde{\theta}) \mid A^*(\theta^a + \tilde{\theta}) = \lambda\}. \tag{17}$$

Indeed, the formula in the statement follows by writing this as

$$I_h^*(\lambda) = \inf_{\theta^a \ll \mu} \left\{ I_{h^*}(d\theta^a/d\mu) + \inf_{\tilde{\theta} \in (L^\infty)^*} \{\sigma_{\operatorname{dom} \bar{I}_h}(\tilde{\theta}) \mid A^*\tilde{\theta} = \lambda - A^*\theta^a\} \right\},$$

and using the expression

$$\sigma_{\operatorname{dom} I_h}(\lambda - A^*\theta^a) = \inf_{\tilde{\theta} \in (L^\infty)^*} \{\sigma_{\operatorname{dom} \bar{I}_h}(\tilde{\theta}) \mid A^*\tilde{\theta} = \lambda - A^*\theta^a\},$$

which is obtained by applying [16, Theorem 19] to the function $\delta_{\operatorname{dom} I_h} = \delta_{\operatorname{dom} \bar{I}_h} \circ A$.

To prove (17), let $\tilde{\theta} \in (L^\infty)^*$ such that $A^*(\theta^a + \tilde{\theta}) = \lambda$. For any $u \in C$,

$$\langle u, \lambda \rangle - I_h(u) = \langle Au, \theta^a \rangle - \bar{I}_h(Au) + \langle u, A^*\tilde{\theta} \rangle,$$

so taking supremum over $u \in \operatorname{dom} I_h$ gives

$$I_h^*(\lambda) \leq I_{h^*}(d\theta^a/d\mu) + \sigma_{\operatorname{dom} \bar{I}_h}(\tilde{\theta}).$$

Minimizing over $\tilde{\theta} \in L^\infty(S)^*$ and $\theta^a \ll \mu$ such that $A^*(\theta^a + \tilde{\theta}) = \lambda$ gives

$$I_h^*(\lambda) \le \inf_{\tilde{\theta}\in(L^\infty)^*,\theta^a\ll\mu} \{I_{h^*}(d\theta^a/d\mu) + \sigma_{\mathrm{dom}\,\bar{I}_h}(\tilde{\theta}) \mid A^*(\theta^a + \tilde{\theta}) = \lambda\}.$$

The reverse inequality follows by noting that if we restrict $\tilde{\theta}$ to be purely singular with respect to μ, we obtain the right hand side of (16). □

Theorem 25. *If D is isc and $C(D) \neq \emptyset$, then for each $\lambda \in M$,*

$$\sigma_{C_b(D)}(\lambda) = \int (h^*)^\infty(d\lambda/d|\lambda|)d|\lambda|.$$

Proof. By Fenchel's inequality,

$$\langle y, \lambda\rangle \le \int \sigma_D(d\lambda/d|\lambda|)d|\lambda| \tag{18}$$

for every $y \in C_b(D)$, so it suffices to show

$$\sup_{y\in C_b(D)} \langle y, \lambda\rangle \ge \int \sigma_S(d\lambda/d|\lambda|)d|\lambda|.$$

We have, by [18, Theorem 14.60],

$$\sup_{w\in L^\infty(\lambda;D)} \int wd\lambda = \int \sigma_D(d\lambda/d|\lambda|)d|\lambda|.$$

Let $\tilde{y} \in C_b(D)$,

$$\alpha < \int \sigma_D(d\lambda/d|\lambda|)d|\lambda|$$

and $w \in L^\infty(\lambda; S)$ be such that $\int wd\lambda > \alpha$. By Lusin's theorem [3, Theorem 7.1.13], there is an open $\tilde{O} \subset S$ such that $\int_{\tilde{O}}(|\tilde{y}|+|w_t|)d|\lambda| < \epsilon/2$, $\tilde{O}^C$ is compact and w is continuous relative to $\tilde{O}^C$. The mapping

$$\Gamma(s) = \begin{cases} w(s) & \text{if } s \in \tilde{O}^C \\ D(s) & \text{if } s \in \tilde{O} \end{cases}$$

is isc convex closed nonempty-valued so that, by [8, Theorem 3.1"'], there is a continuous $\hat{y}$ on S with $\hat{y} = w$ on $\tilde{O}^C$ and $\hat{y} \in D$ everywhere. Since $\hat{y}$ is continuous and bounded on $\tilde{O}^C$ which is compact, there is an open $\hat{O}$ such that $\hat{y}$ is bounded on $\hat{O}$. Since $\hat{O}^C$ is a countable intersection of open sets, we may choose $\hat{O}$ in a way that $\int_{\hat{O}\setminus\tilde{O}^C} |\hat{y}_t|d|\lambda| < \epsilon/2$.

Since $\hat{O}$ and $\tilde{O}$ form an open cover of T and since T is normal, there is, by [9, Theorem 36.1], a continuous partition of unity $(\hat{\alpha}, \tilde{\alpha})$ subordinate to $(\hat{O}, \tilde{O})$. Defining $y := \hat{\alpha}\hat{y} + \tilde{\alpha}\tilde{y}$, we have $y \in C_b(D)$ and

$$\int yd\lambda \ge \int_{\tilde{O}^C} wd\lambda - \int_{\hat{O}\setminus\tilde{O}^C} \hat{\alpha}|\hat{y}|d|\lambda| - \int_{\tilde{O}} \tilde{\alpha}|\tilde{y}|d|\lambda| \ge \int \alpha - \epsilon,$$

which finishes the proof of necessity, since $\alpha < \int \sigma_S(d\lambda/d|\lambda|)d|\lambda|$ was arbitrary. □

Theorem 26. *Assume that $D(s) := \operatorname{dom} h(\cdot, s)$ is isc, $\operatorname{cl}\operatorname{dom} H = C_b(D)$ and that H is finite and continuous at some $u \in C_b$. Then H is a proper convex lsc function and the restriction to M of its conjugate is given by*

$$H^*(\lambda) = \int_S h^*(d\lambda^a/d\mu)d\mu + \int_S (h^*)^\infty(d\lambda^s/d|\lambda^s|)d|\lambda^s|,$$

where λ^s is the singular part of $\lambda \in M$ in its Lebesgue decomposition with respect to μ. If $\operatorname{dom} H = C_b$, then $\operatorname{dom} H^*$ is contained in the set of Borel-measures absolutely continuous w.r.t. μ.

Proof. Since $\operatorname{int}\operatorname{dom} I_h \cap C_b(D) \neq \emptyset$, [16, Theorem 20] gives

$$H^*(\lambda) = (I_h + \delta_{C_b(D)})^*(\lambda) = \min_{\lambda''}\{I_h^*(\lambda - \lambda'') + \sigma_{C_b(D)}(\lambda'')\}$$

Thus, by Theorem 24,

$$\begin{aligned} &H^*(\lambda) \\ &= \min_{\lambda''}\{\min_{\lambda'}\{I_{h^*}(d\lambda'/d\mu) + \sigma_{\operatorname{dom} I_h}(\lambda - \lambda' - \lambda'') \mid \lambda' \ll \mu\} + \sigma_{C_b(D)}(\lambda'')\} \\ &= \min_{\lambda'}\left\{ I_{h^*}(d\lambda'/d\mu) + \min_{\lambda''}\{\sigma_{\operatorname{dom} I_h}(\lambda - \lambda' - \lambda'') + \sigma_{C_b(D)}(\lambda'')\} \;\middle|\; \lambda' \ll \mu \right\}. \end{aligned}$$

Since $\operatorname{int}\operatorname{dom} I_h \cap C_b(D) \neq \emptyset$, [16, Theorem 20] again gives

$$H^*(\lambda) = \min_{\lambda''}\{\sigma_{\operatorname{dom} I_h}(\lambda - \lambda'') + \sigma_{C(D)}(\lambda'')\}.$$

Since, by assumption, $C_b(D) = \operatorname{cl}\operatorname{dom} H = \operatorname{cl}(\operatorname{dom} I_h \cap C_b(D))$, the left side equals $\sigma_{C_b(D)}(\lambda)$. Thus

$$H^*(\lambda) = \min_{\lambda'}\{I_{h^*}(d\lambda'/d\mu) + \sigma_{C_b(D)}(\lambda - \lambda') \mid \lambda' \ll \mu\}. \tag{19}$$

For $\lambda \in M$, Theorem 25 now gives

$$\begin{aligned} H^*(\lambda) = \min_{\lambda'}\{&\int h^*(d\lambda'/d\mu)d\mu + \int (h^*)^\infty(d(\lambda - \lambda')/d\mu)d\mu\} \\ &+ \int (h^*)^\infty(d(\lambda^s)/d|\lambda^s|)d|\lambda^s|, \end{aligned}$$

By [14, Corollary 8.5.1], the last minimum is attained at $d\lambda'/d\mu = d\lambda/d\mu$, so the last expression equals $J_{h^*}(\lambda)$.

If $\operatorname{dom} H = C_b$, (19) implies the claim. □

Proof of Theorem 9. Defining $\tilde{h}(u,s) = h(\psi(s)u, s)$, $\tilde{D}(s) = \operatorname{cl}\operatorname{dom}\tilde{h}(s)$ and

$$\tilde{H}(u) := I_{\tilde{h}}(u) + \delta_{C_b(\tilde{D})},$$

on C_b, we get

$$\begin{aligned}H^*(\lambda) &= \sup_{u\in C}\{\langle u,\lambda\rangle - H(u)\}\\ &= \sup_{u\in C_b}\{\langle u,\psi\lambda\rangle - H(\psi u)\}\\ &= \tilde{H}^*(\psi\lambda).\end{aligned}$$

Clearly, $\tilde{D}(s) = \{u \mid \psi(s)u \in \operatorname{dom} h(s)\}$. By [8, Proposition 2.2], D is isc if and only if $\tilde{D}$ is isc. It is thus clear that H satisfies the assumptions of Theorem 9 if and only if $\tilde{H}$ satisfies those of Theorem 26. Since $\tilde{h}^*(y,s) = h^*(y/\psi(s),s)$, an application of Theorem 26 to $\tilde{H}^*(\psi\lambda)$ gives the expression for H^* in the statement.

As to the subdifferential formulas, we have $\lambda \in \partial H(u)\cap M$ if and only if $H(u) + J_{h^*}(\lambda) = \langle u,\lambda\rangle$. For any $u\in\operatorname{dom} H$ and $\lambda\in M$, we have the Fenchel's inequalities

$$\begin{aligned}h(u) + h^*(d\lambda^a/d\mu) &\geq u\cdot(d\lambda^a/d\mu) \quad \mu\text{-a.e.},\\ (h^*)^\infty(d\lambda^s/d|\lambda^s|) &\geq y\cdot(d\lambda^s/d|\lambda^s|) \quad |\lambda^s|\text{-a.e.},\end{aligned}$$

which hold as equalities if and only if $H(u) + J_{h^*}(\lambda) = \langle u,\lambda\rangle$. These equalities are equivalent to the given pointwise subdifferential conditions.

Since $\operatorname{dom} H^* = \{\lambda\in C^* \mid \psi\lambda\in\operatorname{dom}\tilde{H}^*\}$, the last claim follows from that of Theorem 26. □

References

[1] L. Ambrosio, L. A. Caffarelli, Y. Brenier, G. Buttazzo, and C. Villani. *Optimal transportation and applications*, volume 1813 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin; Centro Internazionale Matematico Estivo (C.I.M.E.), Florence, 2003. Lectures from the C.I.M.E. Summer School held in Martina Franca, September 2–8, 2001, Edited by Caffarelli and S. Salsa.

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

[3] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[4] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probability*, 3:146–158, 1975.

[5] H. Föllmer and N. Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Probab.*, 25(2):901–926, 1997.

[6] C. Léonard. A set of lecture notes on convex optimization with some applications to probability theory incomplete draft. *Lecture notes*, 2006.

[7] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.

[8] E. Michael. Continuous selections. I. *Ann. of Math. (2)*, 63:361–382, 1956.

[9] J. Munkres. *Topology.* Prentice Hall, 2 edition, 2000.

[10] T. Pennanen and A.-P. Perkkiö. Stochastic programs without duality gaps. *Mathematical Programming*, 136(1):91–110, 2012.

[11] A.-P. Perkkiö. Continuous essential selections and integral functionals. *Set-Valued Var. Anal.*, 22(1):45–58, 2014.

[12] A.-P. Perkkiö. Conjugates of integral functionals of continuous functions. *Journal of Mathematical Analysis and Applications*, to appear.

[13] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory.

[14] R. T. Rockafellar. *Convex analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[15] R. T. Rockafellar. Integrals which are convex functionals. II. *Pacific J. Math.*, 39:439–469, 1971.

[16] R. T. Rockafellar. *Conjugate duality and optimization.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974.

[17] R. T. Rockafellar. *Network flows and monotropic optimization.* Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, 1984. A Wiley-Interscience Publication.

[18] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[19] L. Rüschendorf and W. Thomsen. Note on the Schrödinger equation and I-projections. *Statist. Probab. Lett.*, 17(5):369–375, 1993.

[20] L. Rüschendorf and W. Thomsen. Closedness of sum spaces and the generalized Schrödinger problem. *Teor. Veroyatnost. i Primenen.*, 42(3):576–590, 1997.

[21] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[22] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[23] C. Zălinescu. *Convex analysis in general vector spaces.* World Scientific Publishing Co., Inc., River Edge, NJ, 2002.