Entropy martingale optimal transport and nonlinear pricing–hedging duality

Alessandro Doldi[1] · Marco Frittelli[1]

Abstract

The objective of this paper is to develop a duality between a novel entropy martingale optimal transport (EMOT) problem and an associated optimisation problem. In EMOT, we follow the approach taken in the entropy optimal transport (EOT) problem developed in Liero et al. (Invent. Math. 211:969–1117, 2018), but we add the constraint, typical of martingale optimal transport (MOT) theory, that the infimum of the cost functional is taken over martingale probability measures. In the associated problem, the objective functional, related via Fenchel conjugacy to the entropic term in EMOT, is no longer linear as in (martingale) optimal transport. This leads to a novel optimisation problem which also has a clear financial interpretation as a nonlinear subhedging problem. Our theory allows us to establish a nonlinear robust pricing–hedging duality which also covers a wide range of known robust results. We also focus on Wasserstein-induced penalisations and study how the duality is affected by variations in the penalty terms, with a special focus on the convergence of EMOT to the extreme case of MOT.

1 Introduction

As a consequence of the financial crisis in 2008, the uncertainty in the selection of a reference probability P gained increasing attention and led to the investigation of the

✉ M. Frittelli
marco.frittelli@unimi.it

A. Doldi
alessandro.doldi@unimi.it

[1] Dipartimento di Matematica, Università degli Studi di Milano, Via Saldini 50, 20133 Milano, Italy

notions of arbitrage and of the pricing–hedging duality in different settings. On the one hand, the single reference probability P was replaced with a family of – a priori non-dominated – probability measures, leading to the theory of quasi-sure stochastic analysis. On the other hand, taking an even more radical approach, a probability-free, pathwise theory of financial markets made substantial advances in the second decade of this century. In this context, it was shown in the seminal paper by Beiglböck et al. [4] that optimal transport theory is a powerful tool to prove pathwise pricing–hedging duality results. The theory we are going to present fits in this conceptual framework that we now briefly recall.

The market model is in discrete time with a finite horizon $T \in \mathbb{N}$ and zero interest rate. Let

$$\Omega := K_0 \times \cdots \times K_T$$

for closed (possibly noncompact) subsets $K_0, \ldots, K_T$ of $\mathbb{R}$ and let $X_0, \ldots, X_T$ be the canonical projections $X_t : \Omega \to K_t$ for $t = 0, 1, \ldots, T$. The process $X = (X_t)$ represents the price of some underlying asset. Later we allow a multidimensional price process, but in this introduction, we stick to the one-dimensional case for notational simplicity. We assume no reference probability measure. We write

$\text{Mart}(\Omega) := \{\text{martingale probability measures for } X \text{ under the natural filtration}\},$

and when μ is a measure defined on the Borel σ-algebra of Ω, its marginals are denoted by $\mu_0, \ldots, \mu_T$. One then considers a contingent claim $c : \Omega \to (-\infty, +\infty]$ which is allowed to depend on the whole path of the underlying asset, and one admits semistatic trading strategies for hedging. This means that in addition to dynamic trading in X via admissible integrands $\Delta \in \mathcal{H}$, one may invest in vanilla options $\varphi_t : K_t \to \mathbb{R}$. For modelling purposes, one can take vector subspaces $\mathcal{E}_t \subseteq \mathcal{C}(K_t)$ for $t = 0, \ldots, T$, where $\mathcal{C}(K_t)$ is the space of real-valued continuous functions on K_t. For each t, $\mathcal{E}_t$ is the set of static options that can be used for hedging, say affine combinations of vanilla options with different strikes and the same maturity t, and $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ is the space of all hedging instruments. The key assumption in the robust, optimal-transport-based formulation is that the marginals $(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)$ of the underlying price process X are known; see the seminal papers by Breeden and Litzenberger [13] and Hobson [27]. Such marginals can be identified if one knows a (very) large number of prices of plain vanilla options maturing at each intermediate date, for example the prices of all call options with intermediate maturities and ranging strikes. In this case, the class of arbitrage-free pricing measures that are compatible with the observed prices of the options is given by

$$\mathcal{M}(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T) := \{Q \in \text{Mart}(\Omega) : X_t \sim_Q \widehat{Q}_t \text{ for } t = 0, \ldots, T\}.$$

Let $\mathcal{H}$ consist of admissible (predictable) trading strategies, given as in Beiglböck et al. [4] via bounded continuous functions, and let

$$\mathcal{I} := \left\{ I^\Delta(x) := \sum_{t=0}^{T-1} \Delta_t(x_0, \ldots, x_t)(x_{t+1} - x_t) : \Delta \in \mathcal{H} \right\}$$

denote the corresponding set of stochastic integrals. In this framework, the subhedging duality, obtained in [4, Theorem 1.1], takes the form

$$\inf_{Q \in \mathcal{M}(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)} E_Q[c] = \sup \left\{ \sum_{t=0}^{T} E_{\widehat{Q}_t}[\varphi_t] : \varphi \in \mathcal{S}_{\mathrm{sub}}(c) \right\}, \qquad (1.1)$$

for

$$\mathcal{S}_{\mathrm{sub}}(c) := \left\{ \varphi \in \mathcal{E} : \exists \Delta \in \mathcal{H} \text{ with } \sum_{t=0}^{T} \varphi_t(x_t) + I^{\Delta}(x) \le c(x), \forall x \in \Omega \right\}, \qquad (1.2)$$

and the right-hand side of (1.1) is known as the *robust subhedging price* of c. Obviously, an analogous theory for the superhedging price can be developed as well. Several relevant papers contributed to this stream of literature, as for example Davis et al. [18], Dolinsky and Soner [20], Galichon et al. [22], Henry-Labordère et al. [26], Tan and Touzi [37]. More recent works on the topic include also Bartl et al. [3], Cheridito et al. [14], Guo and Obłój [24], Hou and Obłój [28].

1.1 The dual problem

The left-hand side of (1.1), namely $\inf_{Q \in \mathcal{M}(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)} E_Q[c]$, represents the dual problem in the financial application, but is typically the primal problem in *martingale optimal transport* (MOT). We label this case as the *sublinear case* of MOT. Inspired by the *entropy optimal transport* (EOT) introduced in Liero et al. [30], we are naturally led to the study of the *convex case* of MOT, i.e., an *entropy martingale optimal transport* (EMOT) problem, in the form

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) \right), \qquad (1.3)$$

where $\mathcal{D}_{v_t^*, \widehat{Q}_t}$ is a divergence in the usual form (see (3.5) below for an explicit expression). Notice that in the EMOT primal problem (1.3), the typical MOT constraint that Q has prescribed marginals $(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)$ is relaxed (as the infimum is taken with respect to all martingale probability measures) by penalising via $\mathcal{D}_{v_t^*, \widehat{Q}_t}$ those martingale measures Q whose marginals are far from some reference marginals $(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)$. This is a key difference with classical MOT. Nevertheless, when $\mathcal{D}_{v_t^*, \widehat{Q}_t}(\cdot) = \delta_{\{\widehat{Q}_t\}}(\cdot)$, the EMOT reduces to the classical MOT problem where only martingale probability measures with fixed marginals are allowed. Here $\delta_A := \infty 1_{A^c}$ is the characteristic function of a set A as customarily defined in convex analysis. We also stress that in (1.3), we only consider martingale probability measures, while the EOT problem of [30] is obtained by replacing in (1.3) the set $\mathrm{Mart}(\Omega)$ with $\mathrm{Meas}(\Omega)$ consisting of *all* positive finite measures μ on Ω.

In EMOT, the marginals are no longer fixed a priori as in the left-hand side of (1.1), because we may not have sufficient information to detect them with enough accuracy. This might be the case for example if there are not sufficiently many traded call and put options on the underlying assets in the market so that we cannot extract precisely

the marginals via the Breeden and Litzenberger [13] approach. Alternatively, the exact prices of the options might be unknown, e.g. by market impact effects.

1.2 The primal problem

To describe the nonlinear subhedging value, we start with the space $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ of hedging instruments consisting of some vectors of continuous functions and consider a functional $U : \mathcal{E} \to [-\infty, +\infty)$. An example is given by (a sum of) expected utility functions, as detailed below, so that U is not necessarily linear or even cash-additive.

We restore cash-additivity via the notion of the optimised certainty equivalent (OCE) studied in Ben Tal and Teboulle [5]. To this end, we introduce the generalised optimised certainty equivalent associated to U as

$$S^U(\varphi) := \sup_{\beta \in \mathbb{R}^{T+1}} \left(U(\varphi + \beta) - \sum_{t=0}^{T} \beta_t \right). \tag{1.4}$$

This is a cash-additive map (see (2.2)), yet nonlinear in general, and can be considered as a valuation of options $\varphi = (\varphi_t)$ instead of the linear cost $\sum_{t=0}^{T} E_{\widehat{Q}_t}[\varphi_t]$ in (1.1). For a possibly path-dependent contingent claim $c : \Omega \to (-\infty, +\infty]$, the nonlinear subhedging value of c when valuation is done by S^U then reads

$$\pi(c) = \sup\{S^U(\varphi) : \varphi \in \mathcal{S}_{\text{sub}}(c)\} = \sup_{\Delta \in \mathcal{H}} \sup_{\varphi \in \Phi_\Delta(c)} S^U(\varphi)$$

for $\Phi_\Delta(c) := \{\varphi \in \mathcal{E} : \sum_{t=0}^{T} \varphi_t(x_t) + I^\Delta(x) \leq c(x), \forall x \in \Omega\}$.

1.3 The duality

One of the main results of the paper in Theorem 2.4 is the duality

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \mathcal{D}_U(Q) \right) = \sup_{\Delta \in \mathcal{H}} \sup_{\varphi \in \Phi_\Delta(c)} S^U(\varphi), \tag{1.5}$$

where

$$\mathcal{D}_U(Q) = \sup_{\varphi \in \mathcal{E}} \left(U(\varphi) - \sum_{t=0}^{T} \int_{K_t} \varphi_t \mathrm{d}Q_t \right) \qquad \text{for } Q \in \text{Mart}(\Omega) \tag{1.6}$$

is the penalisation term associated to U via the Fenchel conjugate. In addition, we also prove the existence of an optimiser for the problem on the left-hand side of (1.5). We now understand that the dual problem for the latter, namely of EMOT in its general form, is the nonlinear subhedging problem appearing on the right-hand side of (1.5).

Observe that $\mathcal{D} := \mathcal{D}_U$ in (1.6) does not necessarily have an additive structure, or a divergence formulation $\mathcal{D}(Q) = \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t)$ as in (1.3), and so it does not necessarily depend on a given martingale measure $\widehat{Q}$. For example, such penalisation

terms could be induced by market prices (see Sect. 4.3) or by a Wasserstein distance (see Sect. 4.4). This additional flexibility in choosing $\mathcal{D}$ constitutes one key generalisation of the entropy optimal transport theory of Liero et al. [30]. Of course, the other difference with EOT is the presence in (1.5) of the additional supremum with respect to admissible integrands $\Delta \in \mathcal{H}$. As a consequence, on the left-hand side of (1.5), the infimum is now taken with respect to martingale probability measures instead of positive measures.

In the special case of a valuation functional U induced by utility functions, the duality (1.5) has a particularly interesting formulation (see Sects. 3.4 and 4.1 for the assumptions and more details). We provide here only two special cases. Let $\widehat{Q}_t$ be the marginals of some $\widehat{Q} \in \mathrm{Mart}(\Omega)$ and $U(\varphi) = \sum_{t=0}^{T} E_{\widehat{Q}_t}[u_t(\varphi_t)]$. If $u_t(x) = \frac{1}{\gamma_t}(1 - \exp(-\gamma_t x))$, $\gamma_0, \ldots, \gamma_t > 0$, is an exponential utility function, then (1.5) takes the form

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \frac{1}{\gamma_t} H(Q_t, \widehat{Q}_t) \right)$$

$$= \sup \left\{ \sum_{t=0}^{T} -\frac{1}{\gamma_t} \ln E_{\widehat{Q}_t}[\exp(-\gamma_t \varphi_t)] : \varphi \in \mathcal{S}_{\mathrm{sub}}(c) \right\},$$

where $H(\,\cdot\,, \cdot\,)$ denotes the relative entropy. If $u_t(x) = x$ is the linear utility function, then $\mathcal{D}_U(\,\cdot\,) = \sum_{t=0}^{T} \delta_{\widehat{Q}_t}(\,\cdot\,)$ and the EMOT reduces to the classical MOT problem.

Our framework allows us to establish and comprehend several different robust pricing–hedging duality results: new nonlinear utility-based formulations (in Corollaries 4.3, 4.4 and 5.1); the linear case (in Corollary 5.3) and the case without options (in Corollary 4.6); a new duality with penalisation functions based on market data (see Sect. 4.3) or on a Wasserstein distance (see Sect. 4.4).

One additional feature of the paper consists in replacing the set of stochastic integrals $\mathcal{I}$ with a general set $\mathcal{A}$ of suitable hedging instruments that is a general convex cone. Particular choices of such an $\mathcal{A}$, apart from the usual set of stochastic integrals, allow us to work with ε-martingale measures, supermartingales and submartingales in the duality (see Sect. 2.2.1). This extends EMOT beyond the strict martingale property.

Section 2.5 is devoted to stability and convergence issues, as we analyse how the duality is affected by variations in the penalty terms. In Examples 4.12, 4.16 and 5.4, we apply this result to the convergence of EMOT to the extreme case of MOT, and in Sect. 4.4, we focus on Wasserstein-induced penalisation terms.

2 The entropy martingale optimal transport duality

In this section, we present a precise mathematical setting, the main results and their proofs. The main result in Theorem 2.4 relies on (i) a Fenchel–Moreau argument applied to the dual system $(C_{0:T}, (C_{0:T})^*)$, where $C_{0:T}$ is a set of appropriately weighted continuous functions, (ii) the Daniell–Stone theorem that guarantees that

the elements in the dual space $(C_{0:T})^*$ that enter in the dual representation can be represented by probability measures. In order to make this possible, an order-continuity-type assumption on the valuation functional is enforced (see (2.6)).

2.1 The setting

Fix $d \in \mathbb{N}$ modelling the number of stocks in the market, and fix $d(T+1)$ closed subsets $K_0^1, \ldots, K_0^d, \ldots, K_T^1, \ldots, K_T^d$ of $\mathbb{R}$. Set, for $0 \leq s \leq t \leq T$,

$$\Omega_{s:t} := \underset{u=s}{\overset{t}{\times}} \underset{j=1}{\overset{d}{\times}} K_u^j \qquad \text{and} \qquad \Omega := \Omega_{0:T} = \underset{u=0}{\overset{T}{\times}} \underset{j=1}{\overset{d}{\times}} K_u^j.$$

Let $C(\Omega_{s:t})$ be the vector space of continuous real-valued functions on $\Omega_{s:t}$, and let

$$C_{s:t} := \left\{ \varphi \in C(\Omega_{s:t}) : \|\varphi\|_{s:t} := \sup_{x \in \Omega_{s:t}} \frac{|\varphi(x)|}{1 + \sum_{u=s}^{t} \sum_{j=1}^{d} |x_u^j|} < \infty \right\}. \qquad (2.1)$$

We introduce the space $B_{s:t}$ in a similar fashion, just substituting the requirement of continuity for φ with the request that φ be measurable with respect to the Borel σ-algebra of $\Omega_{s:t}$. Then $C_{s:t}$ and $B_{s:t}$ are Banach lattices under the norm $\|\cdot\|_{s:t}$. The topological dual of $C_{s:t}$ is denoted by $(C_{s:t})^*$.

In a discrete-time framework with finite horizon T and assuming zero interest rate, we model a market with d stocks using the canonical d-dimensional process given by $X_t^j(x) = x_t^j$, $j = 1, \ldots, d$, $t = 0, \ldots, T$, for $x \in \Omega$. We introduce the set $\mathrm{Prob}(\Omega)$ of probability measures on Ω, endowed with its Borel σ-algebra, and the set of those probability measures under which the X_t^j are integrable as

$$\mathrm{Prob}^1(\Omega) := \left\{ Q \in \mathrm{Prob}(\Omega) : E_Q\left[\sum_{t=0}^{T} \sum_{j=1}^{d} |X_t^j| \right] < \infty \right\}.$$

Fix now vector subspaces $\mathcal{E}_0, \ldots, \mathcal{E}_T$ with $\mathcal{E}_t \subseteq C_{0:t}$. The space $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ represents the class of financial instruments that can be used for static hedging. Since $\mathcal{E}_t \subseteq C_{0:t}$, we are potentially allowing to consider also Asian and path-dependent options $\varphi_t(x_0, \ldots, x_t)$ in the sets $\mathcal{E}_t$. Nonetheless, the choice $\mathcal{E}_t \subseteq C_{t:t}$ is permitted, too; see Sects. 4 and 5. Moreover, in some of the subsequent results, see Sect. 4.1, we take as $\mathcal{E}_t \subseteq C_{t:t}$ the subspace consisting of (combinations of) deterministic amounts, units of underlying stock at time t and call options with different strike prices and the same maturity t. Let $U : \mathcal{E} \to [-\infty, +\infty)$ be a proper (i.e., $\mathrm{dom}(U) := \{\varphi \in \mathcal{E} : U(\varphi) > -\infty\} \neq \emptyset$) concave functional. Recall from (1.4) the definition of $S^U : \mathcal{E} \to [-\infty, +\infty]$ which represents the valuation functional of the hedging instruments in $\mathcal{E}$. Let $\mathrm{dom}(S^U) := \{\varphi \in \mathcal{E} : S^U(\varphi) > -\infty\}$. Observe that we are considering valuation of the *process* $\varphi = (\varphi_0, \ldots, \varphi_T) \in \mathcal{E}$ rather than the valuation of the terminal payoffs only. Under the usual convention $\infty \cdot 0 = 0 \cdot \infty = 0$, one can check that the functional S^U is concave on the convex set $\mathrm{dom}(S^U)$ and

cash-additive, meaning that

$$S^U(\varphi + \alpha) = S^U(\varphi) + \sum_{t=0}^{T} \alpha_t, \qquad \forall \varphi \in \mathcal{E}, \forall \alpha \in \mathbb{R}^{T+1}. \tag{2.2}$$

Definition 2.1 Given a convex cone $\mathcal{A} \subseteq C_{0:T}$ and a Borel function c, we define

$$\pi(c) := \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi) \in [-\infty, +\infty], \tag{2.3}$$

where

$$\Phi_z(c) := \left\{ \varphi \in \mathrm{dom}(S^U) : \sum_{t=0}^{T} \varphi_t(x_0, \ldots, x_t) + z(x) \leq c(x), \forall x \in \Omega \right\}$$

and the usual convention $\sup \emptyset = -\infty$ is adopted.

We recognise that $\pi(c)$ in (2.3) is a generalised robust subhedging value for c, with a general set $-\mathcal{A}$ replacing the set of terminal values of stochastic integrals used before. Some relevant examples for choices of $\mathcal{A}$ are provided in Sect. 2.2.1.

Definition 2.2 We define the *polar* $\mathcal{A}^\circ$ of the cone $\mathcal{A} \subseteq C_{0:T}$ to be the set

$$\mathcal{A}^\circ := \{\lambda \in (C_{0:T})^* : \langle z, \lambda \rangle \leq 0, \forall z \in \mathcal{A}\}$$

where $\langle \cdot, \cdot \rangle$ is the usual pairing between $C_{0:T}$ and its topological dual $(C_{0:T})^*$, and we observe that for any λ in $(C_{0:T})^*$,

$$\sigma_{\mathcal{A}}(\lambda) := \sup_{z \in \mathcal{A}} \langle z, \lambda \rangle = \begin{cases} 0 & \text{for } \lambda \in \mathcal{A}^\circ \\ \infty & \text{otherwise} \end{cases} = \delta_{\mathcal{A}^\circ}(\lambda).$$

As will be clarified in Sect. 2.4.1, $\mathrm{Prob}^1(\Omega)$ can be identified with a subset of $(C_{0:T})^*$; so we introduce the set of probability measures

$$\mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ = \{Q \in \mathrm{Prob}^1(\Omega) : E_Q[z] \leq 0, \forall z \in \mathcal{A}\}. \tag{2.4}$$

2.2 The main results

Assumption 2.3 (i) Let $K_0^1, \ldots, K_0^d, \ldots, K_T^1, \ldots, K_T^d$ be closed subsets of $\mathbb{R}$ and denote $\Omega = \Omega_{0:T} = \bigtimes_{t=0}^{T} \bigtimes_{j=1}^{d} K_t^j$. The vector subspaces $\mathcal{E}_0, \ldots, \mathcal{E}_T$ satisfy that $\mathbb{R} \subseteq \mathcal{E}_t \subseteq C_{0:t}$, $t = 0, \ldots, T$, and we set $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$. The functional $U : \mathcal{E} \to [-\infty, +\infty)$ is concave with $U(0) \in \mathbb{R}$. Moreover, $\mathcal{A} \subseteq C_{0:T}$ is a convex cone with $0 \in \mathcal{A}$.

(ii) For every $t = 0, \ldots, T$, there exist compact sets $H_t(n) \subseteq \bigtimes_{j=1}^{d} K_t^j, n \geq 1$, and functions $0 \leq f_t^n \in \mathcal{E}_t, n \geq 1$, such that

$$1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j| \leq \sum_{t=0}^{T} f_t^n(x_0, \ldots, x_t), \qquad \forall (x_0, \ldots, x_T) \in \Omega \setminus \left(\bigtimes_{t=0}^{T} H_t(n) \right) \tag{2.5}$$

and

$$U(-af_0^n, \ldots, -af_T^n) \longrightarrow 0 \qquad \text{as } n \to \infty, \forall a \in \mathbb{R}, a > 0 . \tag{2.6}$$

Theorem 2.4 *Suppose Assumption* 2.3 *is fulfilled.*

(i) *If*

$$\pi(\widehat{c}) < \infty \text{ for some } \widehat{c} \in B_{0:T}, \tag{2.7}$$

then $\pi(c) \in \mathbb{R}$ for every $c \in B_{0:T}$ and $\pi : B_{0:T} \to \mathbb{R}$ is norm-continuous, cash-additive, concave and nondecreasing on $B_{0:T}$.

(ii) *For every lower semicontinuous $c : \Omega \to (-\infty, +\infty]$ satisfying*

$$c(x) \geq -A\left(1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j|\right), \qquad \forall x \in \Omega, \text{for some } A \in [0, \infty), \tag{2.8}$$

we have the duality

$$\inf_{Q \in \text{Prob}^1(\Omega) \cap \mathcal{A}^\circ} \left(E_Q[c] + \mathcal{D}(Q)\right) = \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi) = \pi(c), \tag{2.9}$$

where $\text{Prob}^1(\Omega) \cap \mathcal{A}^\circ$ is given in (2.4)*,*

$$\mathcal{D}(Q) = \sup_{\varphi \in \mathcal{E}} \left(U(\varphi) - \sum_{t=0}^{T} \int_{\Omega_{0:t}} \varphi_t \, \mathrm{d}Q_t\right), \tag{2.10}$$

and Q_t is the marginal of $Q \in \text{Prob}^1(\Omega)$ on $\mathcal{B}(\Omega_{0:t})$. Furthermore, if $\pi(c) < \infty$, the infimum on the left-hand side of (2.9) *is a minimum.*

Notice that the condition $\pi(\widehat{c}) < \infty$ for some $\widehat{c} \in B_{0:T}$ is not required for the validity of Theorem 2.4 (ii). In addition, we allow in (2.9) $\text{Prob}^1(\Omega) \cap \mathcal{A}^\circ = \emptyset$ with the usual convention $\inf \emptyset = +\infty$. Recall also that the existence of an optimiser in MOT implies that $\mathcal{M}(\widehat{Q}_0, \widehat{Q}_1, \ldots, \widehat{Q}_T)$ is not empty and that the marginals must be in convex order. In EMOT, the marginals are no longer assigned, and so an optimiser Q^* of the left-hand side of (2.9) belongs to $\text{Prob}^1(\Omega) \cap \mathcal{A}^\circ$ with $\mathcal{D}(Q^*) < \infty$ with no other requirement.

Corollary 2.5 *Suppose Assumption* 2.3 (i) *holds with the subsets K_t^j, $t = 0, \ldots, T$, $j = 1, \ldots, d$, of $\mathbb{R}$ being compact, $c : \Omega \to (-\infty, +\infty]$ is lower semicontinuous and $U(0) = 0$. Then* (2.9) *holds true and if $\pi(c) < \infty$, there exists an optimum for the left-hand side of* (2.9)*.*

Proof When $K_0^1, \ldots, K_0^d, \ldots, K_T^1, \ldots, K_T^d$ are compact, then $C_{0:T} = \mathcal{C}_b(\Omega)$. As $U(0) = 0$, (2.5) and (2.6) are automatically satisfied: just take $\times_{t=0}^{T} H_t(n) = \Omega$ and $f_t^n = 0, t = 0, \ldots, T, n \geq 1$. $\qquad\qquad\square$

Assumption 2.3 (ii) is inspired by Cheridito et al. [15] and is for instance satisfied if $K_t^j \subseteq [0, \infty)$, $t = 0, \ldots, T$, $j = 1, \ldots, d$, and if the valuations over a suitable sequence of call options on the underlying stocks converge to zero when the corresponding strikes diverge to infinity, as explained in the following example.

Example 2.6 Let

$$f_{j,t}^\alpha(x_{t,j}) := (|x_{t,j}| - \alpha)^+, \qquad x_{t,j} \in K_t^j, j = 1, \ldots, d, t = 0, \ldots, T, \quad (2.11)$$

and suppose that $f_{j,t}^\alpha \in \mathcal{E}_t$ for every $\alpha \geq 0$, $j = 1, \ldots, d$, $t = 0, \ldots, T$. As shown in Proposition A.1 (ii), to guarantee that (2.5) and (2.6) are satisfied, it is enough to require that U is (componentwise) nondecreasing on $\times_{t=0}^T \mathcal{E}_t$, $U(0) = 0$ and that for $\beta \in \mathbb{R}_+$ given in A.1 (i), we have

$$U(0, \ldots, 0, -af_{j,t}^{\frac{n}{\beta}}, 0, \ldots, 0) \longrightarrow 0 \qquad \text{as } n \to \infty, \forall j = 1, \ldots, d,$$

$$t = 0, \ldots, T, a \geq 0. \quad (2.12)$$

Condition (2.12) is a requirement on the valuation of single options having maturity t.

Remark 2.7 The proof of Theorem 2.4 will clarify that the use of $-\mathcal{A}$ in place of $\mathcal{A}$ in defining $\pi(c)$ is somehow a matter of taste. Now the infimum in (2.9) is in fact taken over measures in the polar $\mathcal{A}^\circ$. Instead, without the minus sign $(-\mathcal{A})$ in defining $\pi(c)$, we should work with $(-\mathcal{A})^\circ$, which is less convenient in the computations of the proof.

2.2.1 Examples for $\mathcal{A}$

We anticipate here financially relevant examples of possible choices of the convex cone $\mathcal{A}$ and the corresponding set $\mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$.

Example 2.8 To introduce martingale measures in this setup, we set

$$\mathcal{H}^d := \left\{ \Delta = (\Delta_0, \ldots, \Delta_{T-1}) : \Delta_t \in \left(\mathcal{C}_b(K_0 \times \cdots \times K_t) \right)^d \right\},$$

$$I^\Delta(x) := \sum_{t=0}^{T-1} \sum_{j=1}^d \Delta_t^j(x_0, \ldots, x_t)(x_{t+1}^j - x_t^j), \qquad \forall x \in \Omega,$$

$$\mathcal{A} = \mathcal{I} := \{I^\Delta : \Delta \in \mathcal{H}^d\} \subseteq C_{0:T}. \quad (2.13)$$

Thus the space $\mathcal{H}^d$ is the class of admissible trading strategies and $\mathcal{I}$ is the set of elementary stochastic integrals. The (possibly empty) class of martingale measures for the canonical process is denoted by $\mathrm{Mart}(\Omega)$ and consists of all probability measures on $\mathcal{B}(\Omega)$ which make each of the processes (X_t^j) a martingale under the natural filtration $\mathcal{F}_t := \sigma(X_s^j, s \leq t, j = 1, \ldots, d)$, $t = 0, \ldots, T$. Equivalently,

$$\mathrm{Mart}(\Omega) := \{Q \in \mathrm{Prob}^1(\Omega) : E_Q[I^\Delta] = 0, \forall \Delta \in \mathcal{H}^d\}.$$

It is then clear that choosing $\mathcal{A} = \mathcal{I}$, we get $\mathrm{Mart}(\Omega) = \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$. When $d = 1$, we simply write $\mathcal{H} = \mathcal{H}^1$.

Example 2.9 For every $\varepsilon \geq 0$, the set of so-called ε-*martingale measures* (see Guo and Obłój [24]) is

$$\mathrm{Mart}_\varepsilon(\Omega) := \left\{ Q \in \mathrm{Prob}^1(\Omega) : E_Q[I^\Delta] \leq \varepsilon \sum_{t=0}^{T-1} \max_{j=1,\dots,d} \|\Delta_t^j\|_\infty, \forall \Delta \in \mathcal{H}^d \right\}.$$

Thus, taking

$$\mathcal{A}^\varepsilon := \mathrm{conv}\left(\left\{ I^\Delta - \varepsilon \sum_{t=0}^{T-1} \max_{j=1,\dots,d} \|\Delta_t^j\|_\infty : \Delta \in \mathcal{H}^d \right\} \right) \tag{2.14}$$

(here $\mathrm{conv}(\cdot)$ stands for the convex hull in $C_{0:T}$, which is easily seen to be a cone since $\mathcal{H}^d$ is a vector space), one sees that

$$\mathrm{Mart}_\varepsilon(\Omega) = \mathrm{Prob}^1(\Omega) \cap (\mathcal{A}^\varepsilon)^\circ.$$

Taking in particular $\varepsilon = 0$, we have $\mathrm{Mart}_0(\Omega) = \mathrm{Mart}(\Omega)$ as in Example 2.8. It is interesting to notice that for any sequence $\varepsilon_n \downarrow 0$, we have

$$\sigma_{\mathcal{A}^{\varepsilon_n}}(Q) \uparrow \sigma_{\mathcal{I}}(Q), \qquad \forall Q \in \mathrm{Prob}^1(\Omega). \tag{2.15}$$

Example 2.10 Alternative choices for the set $\mathcal{A}$ which produce supermartingale or submartingale measures are $\mathcal{A}^\pm = \{I^\Delta : \Delta \in (\mathcal{H}^\pm)^d\}$, where we define the sets $\mathcal{H}^+ = \{\Delta \in \mathcal{H} : \Delta_t \geq 0, \forall t = 0, \dots, T\}$ and $\mathcal{H}^- = -\mathcal{H}^+$. The set $\mathcal{A}^+$ models dynamic trading with no short selling and yields

$$\mathrm{Prob}^1(\Omega) \cap (\mathcal{A}^+)^\circ = \{\text{supermartingale measures for the canonical process}\}.$$

2.2.2 Rephrasing the main results: superhedging and the martingale measures case

For a given proper concave $U : \mathcal{E} \to \mathbb{R}$, recall the definition of S^U in (1.4) and for $V(\cdot) = -U(-\cdot)$, set $S_V(\varphi) := -S^U(-\varphi)$ and

$$\mathrm{dom}(S_V) := \{\varphi \in \mathcal{E} : S_V(\varphi) < \infty\} = -\mathrm{dom}(S^U).$$

Observe that in our notation, the *superhedging* value for c is

$$\pi_+(c) := \inf_{z \in \mathcal{A}} \inf_{\varphi \in \Psi_z(c)} S_V(\varphi) \in [-\infty, +\infty],$$

where

$$\Psi_z(c) := \left\{ \varphi \in \mathrm{dom}(S_V) : \sum_{t=0}^{T} \varphi_t(x_0, \dots, x_t) + z(x) \geq c(x), \forall x \in \Omega \right\}.$$

The selection of $-\mathcal{A}$ for π and $\mathcal{A}$ for π_+ permits to recognise that the two are linked by $\pi_+(c) = -\pi(-c)$, and so the duality results for π can easily be translated into duality results for π_+. Of course, when $\mathcal{A}$ is a vector space as in the case of stochastic integrals (see (2.13) and Example 2.8), we have $\mathcal{A} = -\mathcal{A}$ and there is no need for the different choices $-\mathcal{A}$ for π and $\mathcal{A}$ for π_+.

We now rephrase our findings in Theorem 2.4, with minor additions, to get the formulations in Corollary 2.12 and Corollary 2.13 which will simplify our discussion in Sects. 4 and 5.

We associate to the functions $c : \Omega \to (-\infty, +\infty]$, $g : \Omega \to [-\infty, +\infty)$ the sets

$$\mathcal{S}_{\text{sub}}(c) := \left\{ \varphi \in \text{dom}(S^U) : \exists \Delta \in \mathcal{H}^d \text{ such that} \right.$$

$$\left. \sum_{t=0}^{T} \varphi_t(x_0, \ldots, x_t) + I^\Delta(x) \leq c(x), \forall x \in \Omega \right\}, \qquad (2.16)$$

$$\mathcal{S}_{\text{sup}}(g) := \left\{ \varphi \in \text{dom}(S_V) : \exists \Delta \in \mathcal{H}^d \text{ such that} \right.$$

$$\left. \sum_{t=0}^{T} \varphi_t(x_0, \ldots, x_t) + I^\Delta(x) \geq g(x), \forall x \in \Omega \right\}, \qquad (2.17)$$

Remark 2.11 If $\mathcal{E}_t \subseteq C_{t:t}$ for $t = 0, \ldots, T$, then $\text{dom}(S^U) \subseteq C_{0:0} \times \cdots \times C_{T:T}$ and each element φ_t in (2.16) is a function of the single variable x_t. If additionally $\text{dom}(S^U) = \mathcal{E}$ and $d = 1$, (2.16) is consistent with (1.2).

From Theorem 2.4 and the equalities $\mathcal{S}_{\text{sup}}(\cdot) = -\mathcal{S}_{\text{sub}}(-\cdot)$, $S_V(\cdot) = -S^U(-\cdot)$, one easily deduces

Corollary 2.12 *Let $\mathcal{A} = \mathcal{I}$ as in (2.13). Suppose that the assumptions in Theorem 2.4 are satisfied, $g : \Omega \to [-\infty, +\infty)$ is upper semicontinuous and also condition (2.8) holds with c replaced by $-g$. Then*

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \mathcal{D}(Q) \right) = \sup_{\varphi \in \mathcal{S}_{\text{sub}}(c)} S^U(\varphi), \qquad (2.18)$$

$$\sup_{Q \in \text{Mart}(\Omega)} \left(E_Q[g] - \mathcal{D}(Q) \right) = \inf_{\varphi \in \mathcal{S}_{\text{sup}}(g)} S_V(\varphi). \qquad (2.19)$$

If the left-hand side of (2.18) (resp. (2.19)) is finite, then an optimum exists for the left-hand side of (2.18) (resp. (2.19)).

Corollary 2.13 *If $d = 1$ and $\Omega := K_0 \times \cdots \times K_T$ for compact sets $K_0, \ldots, K_T \subseteq \mathbb{R}$, then (2.18) and (2.19) as well as existence of optima are guaranteed by the following simplified set of assumptions: $c : \Omega \to (-\infty, +\infty]$ is lower semicontinuous, $g : \Omega \to (-\infty, +\infty]$ is upper semicontinuous and $U(0) = 0$.*

Proof When $K_0, \ldots, K_T \subseteq \mathbb{R}$ are compact, we may repeat the proof of Corollary 2.12 invoking Corollary 2.5 in place of the more general Theorem 2.4. $\qquad \square$

In the subsequent sections, we only consider the subhedging price; the corresponding statements for the superhedging price can be obtained in the obvious way just described.

2.3 Literature review

We observe that the EMOT problem on the left-hand side of (1.5) was not previously considered in the literature. The associated subhedging value on the right-hand side of (1.5) is also new, even though different formulations of nonlinear subhedging prices already appeared in the literature. For example, in Föllmer and Schied [21, Sect. 4.8], the use of general risk measures in a non-robust framework allows weakening the pointwise inequality constraint in subhedging problems. In Cheridito et al. [15], the authors, now in a robust framework, consider additionally a general set of discounted trading gains that may describe different market structures, such as transaction costs or trading constraints. In the present paper, we consider instead explicitly nonlinear pricing (i.e., S^U on the right-hand side of (1.5)) of static parts of semistatic trading strategies and its impact in the duality (i.e., $\mathcal{D}_U$ on the left-hand side of (1.5)). Pennanen and Perkkiö [33] also developed a generalised optimal transport duality, which can be applied to study the pricing–hedging duality in a context similar to our additive setup of Sect. 3.

The addition of an entropic term to optimal transport problems was popularised by Cuturi [17], with several applications especially from the computational point of view (see for example the survey/monograph by Peyré and Cuturi [34, Chap. 4]). The Sinkhorn algorithm can be applied with the entropic regularisation procedure described in these works (see Benamou et al. [6] for some advantages). Convergence for this algorithm is studied e.g. in Ireland and Kullback [29] and Rüschendorf [35]. After the present paper was posted on arXiv, several relevant advances were made regarding this topic. We mention here Nutz and Wiesel [32], Bernton et al. [7], Ghosal et al. [23]. We stress that all the papers mentioned in this paragraph address a different problem: in Cuturi [17] and subsequent works, the requirement of an exact matching of the marginal distributions is maintained. In the present setting, we relax this constraint in order to model uncertainty regarding the marginals themselves.

A Sinkhorn algorithm approach was adopted in De March and Henry-Labordère [19] for building an arbitrage-free implied volatility surface from bid-ask quotes, while Henry-Labordère [25] studied a problem related to the entropic relaxation of an optimal transportation problem and Blanchet et al. [9] studied the number of operations needed for approximation of the transport cost with a given accuracy, in the case of entropic regularisation. Our framework also allows the use of a penalisation of the form $Q \mapsto \sum_{t=0}^{T} \delta_{\widehat{Q}_t}(Q_t) + \widetilde{D}(Q)$, for some entropic term $\widetilde{D}$, so that with this choice, the EMOT reduces to the MOT problem with an additional entropic regularisation term, as analysed in the abovementioned literature.

Stability issues have been studied in Backhoff-Veraguas and Pammer [2] and Neufeld and Sester [31] in what we called the sublinear case, namely with no penalty and with fixed marginals.

The works by Bernton et al. [7] and Ghosal et al. [23] study geometric properties of minimisers of the entropic OT, by means of the concept of cyclical invariance. This is

a counterpart to the characterisation, using c-cyclical monotonicity, of the geometry of optimal transport plans in the classical framework of OT. Even though a similar study of geometric properties for optimisers of EMOT would be of great interest, this topic is beyond the scope of the present paper and is left for future research.

In the framework of Liero et al. [30] (i.e., with penalisations of the marginals induced by divergence functions) and after the first version of the present work was posted on arXiv, duality results were obtained in the context of weak martingale optimal entropy transport problems by Chung and Trinh [16].

2.4 Proof of Theorem 2.4

2.4.1 The full technical setup

For a metric space $\mathbb{X}$, $\mathcal{B}(\mathbb{X})$ denotes the Borel σ-algebra and $m\mathcal{B}(\mathbb{X})$ the class of real-valued, Borel-measurable functions on $\mathbb{X}$. We define the sets

$$\text{ca}(\mathbb{X}) := \{\gamma : \mathcal{B}(\mathbb{X}) \to (-\infty, +\infty) : \gamma \text{ finite signed Borel measure on } \mathbb{X}\},$$

$$\text{Meas}(\mathbb{X}) := \{\mu : \mathcal{B}(\mathbb{X}) \to [0, \infty) : \mu \geq 0 \text{ finite Borel measure on } \mathbb{X}\},$$

$$\text{Prob}(\mathbb{X}) := \{Q : \mathcal{B}(\mathbb{X}) \to [0, 1] : Q \text{ probability Borel measure on } \mathbb{X}\},$$

$$\mathcal{C}(\mathbb{X}) := \{\varphi : \mathbb{X} \to \mathbb{R} : \varphi \text{ continuous on } \mathbb{X}\},$$

$$\mathcal{C}_b(\mathbb{X}) := \{\varphi : \mathbb{X} \to \mathbb{R} : \varphi \text{ bounded and continuous on } \mathbb{X}\}.$$

Recall that we fixed $d(T + 1)$ closed sets $K_0^1, \ldots, K_0^d, \ldots, K_T^1, \ldots, K_T^d \subseteq \mathbb{R}$, where d is the number of stocks and T the time horizon. We use the following weighted spaces of continuous functions: for an index set $I \subseteq \{1, \ldots, d\} \times \{0, \ldots, T\}$, we take

$$C_I := \left\{\phi \in \mathcal{C}\left(\underset{(j,t)\in I}{\times} K_t^j\right) : \|\phi\|_I := \sup_{x \in \times_{(j,t)\in I} K_t^j} \frac{|\phi(x)|}{1 + \sum_{(j,t)\in I} |x_t^j|} < \infty\right\}.$$

For example, we already encountered $C_{s:t}$ in (2.1), and we also consider

$$C_t := C_{t:t}.$$

The corresponding norms are denoted by $\|\cdot\|_I, \|\cdot\|_{s:t}, \|\cdot\|_t$, respectively. Some additional details on weighted spaces can be found in Appendix A.1.

Notice that if $K_0^1, \ldots, K_0^d, \ldots, K_T^1, \ldots, K_T^d$ are compact, then

$$C_{0:T} = \mathcal{C}_b(\Omega) \qquad \text{and} \qquad (C_{0:T})^* = \text{ca}(\Omega).$$

Observe that by a slight abuse of notation (regarding the domains of the functions), for index sets $I \subseteq J \subseteq \{1, \ldots, d\} \times \{0, \ldots, T\}$, there is a constant $0 < \theta \leq 1$ such that

$$C_I \subseteq C_J, \qquad \theta\|\phi\|_I \leq \|\phi\|_J \leq \|\phi\|_I, \quad \forall \phi \in C_I. \tag{2.20}$$

As already mentioned in Pennanen and Perkkiö [33] and Cheridito et al. [15], every finite signed Borel measure γ on $\mathbb{X}$ with $C_{0:T} \subseteq L^1(\mathbb{X}, \mathcal{B}(\mathbb{X}), |\gamma|)$ induces a continuous linear functional $\lambda \in (C_{0:T})^*$ via integration, namely

$$c \mapsto \langle c, \lambda \rangle = \int_{\mathbb{X}} c \, d\gamma, \qquad \forall c \in C_{0:T}.$$

The collection of such functionals, identified with the corresponding measures, is denoted by $\mathrm{ca}^1(\mathbb{X})$, that is,

$$\mathrm{ca}^1(\mathbb{X}) := \big\{ \gamma : \gamma \text{ is a finite signed Borel measure on } \mathcal{B}(\mathbb{X})$$
$$\text{with } C_{0:T} \subseteq L^1\big(\mathbb{X}, \mathcal{B}(\mathbb{X}), |\gamma|\big) \big\},$$

while the classes of nonnegative and of probability measures in $\mathrm{ca}^1(\mathbb{X})$ are denoted by $\mathrm{Meas}^1(\mathbb{X})$ and $\mathrm{Prob}^1(\mathbb{X})$, respectively. The canonical d-dimensional process is given by $X_t^j(x) = x_t^j$, $j = 1, \ldots, d$, $t = 0, \ldots, T$. Observe that every $\phi \in C_{0:T}$ satisfies $|\phi(x)| \leq \|\phi\|_{0:T}(1 + \sum_{t=0}^T \sum_{j=1}^d |x_t^j|)$, and so for any $\mu \in \mathrm{Meas}^1(\mathbb{X})$, we have $C_{0:T} \subseteq L^1(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ iff $X_t^j \in L^1(\mathbb{X}, \mathcal{B}(\mathbb{X}), \mu)$ for all j and t.

Remark 2.14 Under Assumption 2.3 (i), with the notation from there, consider the proper, convex functional $V(\varphi) := -U(-\varphi)$, with $\mathrm{dom}(V) = \{\varphi \in \mathcal{E} : V(\varphi) < \infty\}$. We define the (convex) conjugate $\mathcal{D} : \times_{t=0}^T (C_{0:t})^* \to (-\infty, +\infty]$ of U by

$$\mathcal{D}(\gamma_0, \ldots, \gamma_T) := \sup_{\varphi \in \mathcal{E}} \left(U(\varphi) - \sum_{t=0}^T \langle \varphi_t, \gamma_t \rangle \right) = \sup_{\varphi \in \mathcal{E}} \left(\sum_{t=0}^T \langle \varphi_t, \gamma_t \rangle - V(\varphi) \right). \quad (2.21)$$

Then $\mathcal{D}$ is a convex functional and $\sigma(\times_{t=0}^T (C_{0:t})^*, \mathcal{E})$-lower semicontinuous, even if we do not require that U is $\sigma(\mathcal{E}, \times_{t=0}^T (C_{0:t})^*)$-upper semicontinuous. When some $\gamma \in (C_{0:T})^*$ is given, we slightly improperly write $\mathcal{D}(\gamma) = \mathcal{D}(\gamma_0, \ldots, \gamma_T)$, where γ_t is the restriction of γ to $C_{0:t}$. We also set

$$\mathrm{dom}(\mathcal{D}) = \left\{ (\gamma_0, \ldots, \gamma_T) \in \times_{t=0}^T (C_{0:t})^* : \mathcal{D}(\gamma_0, \ldots, \gamma_T) < \infty \right\}.$$

As an immediate consequence of the definitions, the Fenchel inequality holds: if $(\varphi_0, \ldots, \varphi_T) \in \mathcal{E}$ and $(\gamma_0, \ldots, \gamma_T) \in \times_{t=0}^T (C_{0:t})^*$, then

$$\sum_{t=0}^T \langle \varphi_t, \gamma_t \rangle \leq \mathcal{D}(\gamma_0, \ldots, \gamma_T) + V(\varphi_0, \ldots, \varphi_T). \quad (2.22)$$

Remark 2.15 Another way to introduce our setting, which is used in Sects. 4.3 and 4.4, is to start with a proper convex functional $\mathcal{D} : \mathrm{ca}^1(\Omega) \to (-\infty, +\infty]$ which is $\sigma(\mathrm{ca}^1(\Omega), \mathcal{E})$-lower semicontinuous for an $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T \subseteq (C_{0:T})^{T+1}$. By the

Fenchel–Moreau theorem, we then have the representation

$$\mathcal{D}(\gamma) = \sup_{\varphi \in \mathcal{E}} \left(\sum_{t=0}^{T} \int_{\Omega} \varphi_t \, d\gamma - V(\varphi) \right),$$

where now V is the Fenchel–Moreau (convex) conjugate of $\mathcal{D}$, namely

$$V(\varphi) := \sup_{\gamma \in \mathrm{ca}^1(\Omega)} \left(\sum_{t=0}^{T} \int_{\Omega} \varphi_t \, d\gamma - \mathcal{D}(\gamma) \right). \tag{2.23}$$

Setting $U(\varphi) := -V(-\varphi)$, $\varphi \in \mathcal{E}$, we get back that $\mathcal{D}$ satisfies (2.21) and additionally that U is $\sigma(\mathcal{E}, \mathrm{ca}^1(\Omega))$-upper semicontinuous. In conclusion, a pair $(U, \mathcal{D})$ satisfying (2.21) might be defined either by providing a proper concave $U : \mathcal{E} \to [-\infty, +\infty)$ as in Sect. 2.1, or by assigning a proper convex and $\sigma(\mathcal{E}, \mathrm{ca}^1(\Omega))$-lower semicontinuous $\mathcal{D} : \mathrm{ca}^1(\Omega) \to (-\infty, +\infty]$ as explained in this remark.

2.4.2 Technical comments on Theorem 2.4

Remark 2.16 We now provide conditions ensuring that $\pi(0) < \infty$, which by Theorem 2.4 (i) implies that $\pi(c) \in \mathbb{R}$ for every $c \in B_{0:T}$.

(a) If there exists $\lambda \in \mathcal{A}^\circ \cap \partial U(0) \subseteq (C_{0:T})^*$, then $\pi(0) < \infty$. Note that here, $\partial U(0) \subseteq \times_{t=0}^{T} (C_{0:T})^*$ is the supergradient of U at $0 \in \mathcal{E}$, and we identify λ with the vector of its restrictions in writing improperly $\lambda \in \partial U(0)$). To see this, let λ satisfy $S^U(\varphi) \le \sum_{t=0}^{T} \langle \varphi_t, \lambda_t \rangle$, $\forall \varphi \in \mathcal{E}$. In particular, for all $z \in -\mathcal{A}$ and $\varphi \in \Phi_z(0)$, it then holds that $S^U(\varphi) \le \langle \sum_{t=0}^{T} \varphi_t, \lambda \rangle \le \langle \sum_{t=0}^{T} \varphi_t + z, \lambda \rangle \le 0$, as $\langle z, \lambda \rangle \ge 0$ for all $\lambda \in \mathcal{A}^\circ$, which in turns yields $\pi(0) \le 0$.

(b) We have $\pi(0) < \infty$ if and only if there exists $Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$ such that $\mathcal{D}(Q) < \infty$. Indeed, by definition, $\pi(0) \le \int_{\Omega} 0 \, dQ + \pi^*(Q) = \pi^*(Q)$. But from Lemma 2.20 (which does not rely on Lemma 2.18), we have

$$\pi^*(Q) = \mathcal{D}(Q) + \sigma_{\mathcal{A}}(Q) = \mathcal{D}(Q)$$

(the latter equality coming from $Q \in \mathcal{A}^\circ$). Hence $\pi(0) \le \mathcal{D}(Q) < \infty$. Conversely, $\pi(0) < \infty$ implies the existence of a minimum point in (2.9).

Remark 2.17 Set

$$\widetilde{\Phi}_z(c) := \left\{ \varphi \in \mathcal{E} : \sum_{t=0}^{T} \varphi_t(x_0, \dots, x_t) + z(x) \le c(x), \forall x \in \Omega \right\}$$

and observe that $\Phi_z(c) = \widetilde{\Phi}_z(c) \cap \mathrm{dom}(S^U)$. Then with $\sup \emptyset = -\infty$,

$$\pi(c) := \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi) = \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \widetilde{\Phi}_z(c)} S^U(\varphi). \tag{2.24}$$

To see this, we consider different cases for a fixed $z \in -\mathcal{A}$.

Case 1: $\Phi_z(c) = \emptyset$, which means $\sup_{\varphi \in \Phi_z(c)} S^U(\varphi) = -\infty$ by convention. If $\widetilde{\Phi}_z(c) = \emptyset$, then $\sup_{\varphi \in \widetilde{\Phi}_z(c)} S^U(\varphi) = -\infty$ by convention. If $\widetilde{\Phi}_z(c) \neq \emptyset$, then $\sup_{\varphi \in \widetilde{\Phi}_z(c)} S^U(\varphi) = -\infty$ since for every $\varphi \in \widetilde{\Phi}_z(c)$, we have $S^U(\varphi) = -\infty$ as $\varphi \notin \mathrm{dom}(S^U)$.

Case 2: $\Phi_z(c) \neq \emptyset$. Then $\widetilde{\Phi}_z(c) \neq \emptyset$, too, and since we can ignore all functions $\varphi \in \widetilde{\Phi}_z(c) \setminus \Phi_z(c)$ (which produce values $S^U(\varphi) = -\infty$), we get

$$\sup_{\varphi \in \widetilde{\Phi}_z(c)} S^U(\varphi) = \sup_{\varphi \in \Phi_z(c)} S^U(\varphi).$$

2.4.3 The proof

The proof of Theorem 2.4 is split in the following Lemmas 2.18, 2.20, 2.22–2.24 which are then combined in Lemma 2.25.

Lemma 2.18 *Under Assumption* 2.3 *and if* (2.7) *holds, Theorem* 2.4 (i) *holds. Moreover, the restriction of π to $C_{0:T}$ satisfies*

$$\pi(c) = \min_{\substack{\lambda \in (C_{0:T})^*, \\ \lambda \geq 0, \lambda(1)=1}} \left(\langle c, \lambda \rangle + \pi^*(\lambda) \right), \qquad \forall c \in C_{0:T}, \tag{2.25}$$

for

$$\pi^*(\lambda) = \sup_{c \in C_{0:T}} \left(\pi(c) - \langle c, \lambda \rangle \right), \qquad \lambda \in (C_{0:T})^*.$$

Proof Suppose that $\pi(\widehat{c}) < \infty$ for some $\widehat{c} \in B_{0:T}$. To prove that $\pi(c) > -\infty$ for every $c \in B_{0:T}$, it is (more than) enough to show that

$$\Phi_z(c) \neq \emptyset, \qquad \forall z \in -\mathcal{A}. \tag{2.26}$$

Set $H_n = H_0(n) \times \cdots \times H_T(n) \subseteq \Omega$. Observe that whenever $c \in B_{0:T}$ is given, we have for every $n \geq 1$ and $x \in H_n$ that

$$c(x) - z(x) \geq - \sup_{x \in H_n} |c(x) - z(x)|$$

$$\geq -\|c - z\|_{0:T} \sup_{x \in H_n} \left(1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j| \right) > -\infty,$$

and for every $x \in \Omega \setminus H_n$ that

$$c(x) - z(x) \geq -\|c - z\|_{0:T} \left(1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j| \right)$$

$$\geq -\|c - z\|_{0:T} - \|c - z\|_{0:T} \sum_{t=0}^{T} f_t^n(x_0, \ldots, x_t),$$

using (2.5) in the last inequality. Thus for every $x \in \Omega$,

$$c(x) - z(x) \geq -\|c - z\|_{0:T} - \sup_{x \in H_n} |c(x) - z(x)| - \|c - z\|_{0:T} \sum_{t=0}^{T} f_t^n(x_0, \dots, x_t).$$

If we now show that $(-\|c - z\|_{0:T} f_t^n)_{0 \leq t \leq T} \in \text{dom}(S^U)$ for n big enough, we then conclude that

$$\left(-\|c - z\|_{0:T} - \sup_{x \in H_n} |c(x) - z(x)| - \|c - z\|_{0:T} f_t^n \right)_{0 \leq t \leq T} \in \text{dom}(S^U)$$

by cash-additivity of S^U, and at the same time,

$$\left(-\|c - z\|_{0:T} - \sup_{x \in H_n} |c(x) - z(x)| - \|c - z\|_{0:T} f_t^n \right)_{0 \leq t \leq T} \in \Phi_z(c)$$

by definition. This in particular proves that $\pi(c) > -\infty$. Going then back to checking $(-\|c - z\|_{0:T} f_t^n)_{0 \leq t \leq T} \in \text{dom}(S^U)$, observe that

$$S^U\left((-\|c - z\|_{0:T} f_t^n)_{0 \leq t \leq T} \right)$$

$$= \sup_{\alpha \in \mathbb{R}^{T+1}} \left(U\left((-\|c - z\|_{0:T} f_t^n)_{0 \leq t \leq T} + \alpha \right) - \sum_{t=0}^{T} \alpha_t \right)$$

$$\geq U\left(-\|c - z\|_{0:T} (f_t^n)_{0 \leq t \leq T} \right)$$

$$= -V\left(\|c - z\|_{0:T} (f_t^n)_{0 \leq t \leq T} \right) \longrightarrow 0 > -\infty \qquad \text{as } n \to \infty$$

by Assumption 2.3. The fact that $\pi(c) < \infty$ will follow once we show monotonicity, cash-additivity and concavity of π. Monotonicity is trivial: if $c_1 \leq c_2$, then $\Phi_z(c_1) \subseteq \Phi_z(c_2)$ for every $z \in -\mathcal{A}$ (both sets might be empty). The cash-additivity property can be seen as follows: given $\beta \in \mathbb{R}$ and setting $1 = (1, \dots, 1) \in \mathbb{R}^T$, observe that whenever $z \in \mathcal{A}$ is given, $\varphi \in \Phi_z(c + \beta)$ is equivalent to $\varphi - \frac{\beta}{T+1} 1 \in \Phi_z(c)$ since by cash-additivity of S^U, $\text{dom}(S^U) + \mathbb{R}^{T+1} = \text{dom}(S^U)$. Consequently,

$$\pi(c + \beta) = \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c+\beta)} S^U(\varphi) = \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U\left(\varphi + \frac{\beta}{T+1} 1 \right)$$

$$= \sum_{t=0}^{T} \frac{\beta}{T+1} + \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi) = \pi(c) + \beta.$$

Coming to concavity, it is convenient to rewrite $\pi(c)$ in a slightly more convenient form as

$$\pi(c) = \sup \left\{ S^U(\varphi) : \varphi \in \text{dom}(S^U), \exists z \in -\mathcal{A} \text{ such that } \sum_{t=0}^{T} \varphi_t + z \leq c \right\} \qquad (2.27)$$

and to recall that whenever $c \in B_{0:T}$ is given, the set over which we take the supremum on the right-hand side of (2.27) is not empty by (2.26). Take then $c_i \in B_{0:T}$ and associated $z_i \in -\mathcal{A}, \varphi^i \in \operatorname{dom}(S^U)$ with $\sum_{t=0}^{T} \varphi_t^i + z_i \leq c_i$. Define $c_\alpha = \alpha c_1 + (1-\alpha)c_2$ and analogously z_α and φ^α for $\alpha \in [0,1]$. Then clearly $\sum_{t=0}^{T} \varphi_t^\alpha + z_\alpha \leq c_\alpha$. Combining this with concavity of S^U on $\operatorname{dom}(S^U)$, we obtain

$$\alpha S^U(\varphi^1) + (1-\alpha)S^U(\varphi^2)$$

$$\leq S^U(\varphi^\alpha)$$

$$\leq \sup\left\{ S^U(\varphi) : \varphi \in \operatorname{dom}(S^U), \exists z \in -\mathcal{A} \text{ such that } \sum_{t=0}^{T} \varphi_t + z \leq c_\alpha \right\}$$

$$= \pi(c_\alpha),$$

using (2.27) in the last equality. Taking now the supremum over all z_i, φ^i with $\sum_{t=0}^{T} \varphi_t^i + z_i \leq c_i$, we obtain

$$\alpha \pi(c_1) + (1-\alpha)\pi(c_2) \leq \pi\big(\alpha c_1 + (1-\alpha)c_2\big), \quad \forall \alpha \in [0,1], c_1, c_2 \in B_{0:T}. \quad (2.28)$$

Notice that up to this point, we have $\pi(c_i) \in (-\infty, +\infty]$ so that (2.28) makes sense.

Now we can combine (2.28) with the fact that $\pi(c) > -\infty$ for every $c \in B_{0:T}$ to show that $\pi(c) < \infty$ for every $c \in B_{0:T}$. Indeed, suppose that $\pi(\widetilde{c}) = \infty$ for some $\widetilde{c} \in B_{0:T}$. We know by hypothesis that $\pi(\widehat{c}) < \infty$ for some $\widehat{c} \in B_{0:T}$, and by what we have previously proved, we know that $\pi(2\widehat{c} - \widetilde{c}) > -\infty$. Observing that $\widehat{c} = \alpha(2\widehat{c} - \widetilde{c}) + (1-\alpha)\widetilde{c}$ for $\alpha = \frac{1}{2}$, we have from (2.28) that

$$\infty = \alpha\pi(2\widehat{c} - \widetilde{c}) + (1-\alpha)\pi(\widetilde{c}) \leq \pi\big(\alpha(2\widehat{c} - \widetilde{c}) + (1-\alpha)\widetilde{c}\big) = \pi(\widehat{c}) < \infty.$$

This yields a contradiction; thus there can be no $\widetilde{c} \in B_{0:T}$ with $\pi(\widetilde{c}) = \infty$. Hence $\pi : B_{0:T} \to \mathbb{R}$ is cash-additive, concave and nondecreasing on $B_{0:T}$. Then it is automatically norm-continuous on $B_{0:T}$ by the extended Namioka–Klee theorem (see Biagini and Frittelli [8]). The Fenchel–Moreau-type dual representation (2.25) holds, again by the extended Namioka–Klee theorem, this time applied to the restriction of π to $C_{0:T}$, plus standard arguments involving monotonicity and cash-additivity to prove that $\pi^*(\lambda) < \infty$ implies that $\lambda \geq 0$ and $\lambda(1) = 1$. See for example Föllmer and Schied [21, Theorem 4.16] for an exploitable technique for a similar argument. $\quad\square$

Remark 2.19 Under Assumption 2.3 and if (2.7) holds, $S^U(\varphi) < \infty$ for every $\varphi \in \operatorname{dom}(S^U)$. Indeed, choosing $c_\varphi := \sum_{t=0}^{T} \varphi_t$, we get that $\varphi \in \Phi_0(c_\varphi)$ and $S^U(\varphi) \leq \pi(c_\varphi) < \infty$ by Lemma 2.18.

Lemma 2.20 *For every* $\lambda \in (C_{0:T})^*$ *with* $\lambda \geq 0$, *we have*

$$\pi^*(\lambda) = (S^U)^*(\lambda_0, \ldots, \lambda_T) + \sigma_{\mathcal{A}}(\lambda).$$

If in addition $\lambda(1) = 1$, *then*

$$(S^U)^*(\lambda_0, \ldots, \lambda_T) := \sup_{\varphi \in \mathcal{E}}\left(S^U(\varphi) - \sum_{t=0}^{T} \langle \varphi_t, \lambda_t \rangle \right) = \mathcal{D}(\lambda_0, \ldots, \lambda_T) = \mathcal{D}(\lambda).$$

Proof Fix $\lambda \in (C_{0:T})^*$ with $\lambda \geq 0$. Then

$$
\begin{aligned}
\pi^*(\lambda) &= \sup_{c \in C_{0:T}} \left(\pi(c) - \langle c, \lambda \rangle \right) \\
&= \sup_{c \in C_{0:T}} \left(\sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi) - \langle c, \lambda \rangle \right) \\
&= \sup_{c \in C_{0:T}} \left(\sup_{z \in -\mathcal{A}} \sup_{\varphi \in \widetilde{\Phi}_z(c)} S^U(\varphi) - \langle c, \lambda \rangle \right) \\
&= \sup_{c \in C_{0:T}} \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \widetilde{\Phi}_z(c)} \left(S^U(\varphi) - \langle c, \lambda \rangle \right) \quad\quad (2.29) \\
&\leq \sup_{c \in C_{0:T}} \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \widetilde{\Phi}_z(c)} \left(S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t + z, \lambda \right\rangle \right) \\
&= \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \mathcal{E}} \left(S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t + z, \lambda \right\rangle \right) \\
&= \sup_{z \in \mathcal{A}} \langle z, \lambda \rangle + \sup_{\varphi \in \mathcal{E}} \left(S^U(\varphi) - \sum_{t=0}^{T} \langle \varphi_t, \lambda_t \rangle \right) \\
&= \sigma_{\mathcal{A}}(\lambda) + (S^U)^*(\lambda_0, \ldots, \lambda_T),
\end{aligned}
$$

where the third equality follows from (2.24). Consequently,

$$
\pi^*(\lambda) \leq \sigma_{\mathcal{A}}(\lambda) + (S^U)^*(\lambda_0, \ldots, \lambda_T). \quad\quad (2.30)
$$

At the same time, for every $\varphi \in \mathcal{E}$, $z \in -\mathcal{A}$ and for $\widehat{c} = \sum_{t=0}^{T} \varphi_t + z \in C_{0:T}$, we have that $\varphi \in \Phi_z(\widehat{c})$. Thus

$$
S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t + z, \lambda \right\rangle \leq \sup_{\varphi \in \widetilde{\Phi}_z(\widehat{c})} \left(S^U(\varphi) - \langle \widehat{c}, \lambda \rangle \right) \leq \pi^*(\lambda),
$$

using (2.29) in the second inequality, and hence

$$
\begin{aligned}
&\sup_{z \in \mathcal{A}} \langle z, \lambda \rangle + \sup_{\varphi \in \mathcal{E}} \left(S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t, \lambda \right\rangle \right) \\
&= \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \mathcal{E}} \left(S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t + z, \lambda \right\rangle \right) \leq \pi^*(\lambda). \quad\quad (2.31)
\end{aligned}
$$

Combining (2.30) and (2.31), we get $\pi^*(\lambda) = \sigma_{\mathcal{A}}(\lambda) + (S^U)^*(\lambda_0, \ldots, \lambda_T)$. If additionally $\lambda(1) = 1$, then we have

$$
\begin{aligned}
(S^U)^*(\lambda_0, \ldots, \lambda_T) &= \sup_{\varphi \in \mathcal{E}} \left(S^U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t, \lambda_t \right\rangle \right) \\
&= \sup_{\varphi \in \mathcal{E}} \left(\sup_{\alpha \in \mathbb{R}^{T+1}} \left(U(\varphi + \alpha) - \sum_{t=0}^{T} \alpha_t \right) - \left\langle \sum_{t=0}^{T} \varphi_t, \lambda \right\rangle \right) \\
&= \sup_{\varphi \in \mathcal{E}} \sup_{\alpha \in \mathbb{R}^{T+1}} \left(\left(U(\varphi + \alpha) - \sum_{t=0}^{T} \alpha_t \right) - \left\langle \sum_{t=0}^{T} \varphi_t, \lambda \right\rangle \right) \\
&= \sup_{\varphi \in \mathcal{E}} \sup_{\alpha \in \mathbb{R}^{T+1}} \left(U(\varphi + \alpha) - \sum_{t=0}^{T} \langle \varphi_t + \alpha_t, \lambda \rangle \right) \\
&= \sup_{\varphi \in \mathcal{E}} \left(U(\varphi) - \left\langle \sum_{t=0}^{T} \varphi_t, \lambda \right\rangle \right) = \mathcal{D}(\lambda_0, \ldots, \lambda_T) = \mathcal{D}(\lambda). \quad \square
\end{aligned}
$$

Remark 2.21 Under Assumption 2.3, we have $U(0) \in \mathbb{R}$ and therefore

$$
(S^U)^*(\lambda) \geq S^U(0) \geq U(0) > -\infty
$$

for every $0 \leq \lambda \in (C_{0:T})^*$.

Lemma 2.22 *Under Assumption 2.3 and if (2.7) holds, let $0 \leq \lambda \in (C_{0:T})^*$ with $\lambda(1) = 1$ be given and define $0 \leq \lambda_t = \lambda|_{C_{0:t}} \in (C_{0:t})^*$. If $(\lambda_0, \ldots, \lambda_T) \in \mathrm{dom}(\mathcal{D})$, then there exists a unique $Q \in \mathrm{Prob}^1(\Omega)$ which represents λ on $C_{0:T}$, i.e.,*

$$
\langle \varphi, \lambda \rangle = E_Q[\varphi], \qquad \forall \varphi \in C_{0:T}.
$$

Proof The proof is an adaptation of Bogachev [10, Theorem 7.10.6]. We first stress the fact that $\lambda_t = \lambda|_{C_{0:t}} \in (C_{0:t})^*$ is a consequence of (2.20). We apply Proposition A.2. To do so, we show that for a fixed $\varepsilon > 0$ and n big enough, we may define a set $H_0(n) \times \cdots \times H_T(n)$ that is compact (since so are all the factors) and satisfies the assumptions in Proposition A.2. Suppose that a given $\varphi \in C_{0:T}$ satisfies $\varphi(x) = 0$ for every $x \in \times_{t=0}^{T} H_t(n)$. We also have automatically that

$$
|\varphi(x)| \leq \|\varphi\|_{0:T} \left(1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j| \right), \qquad \forall x \in \Omega.
$$

By Assumption 2.3, we then have for every $(x_0, \ldots, x_T) \in \Omega \setminus (\times_{t=0}^{T} H_t(n))$ that

$$
|\varphi(x)| \leq \|\varphi\|_{0:T} \left(\sum_{t=0}^{T} f_t^n(x_0, \ldots, x_t) \right).
$$

Since, moreover, $\varphi = 0$ on $\times_{t=0}^{T} H_t(n)$ by assumption, we get

$$|\varphi(x)| \le \|\varphi\|_{0:T} \left(\sum_{t=0}^{T} f_t^n(x_0, \ldots, x_t) \right), \qquad \forall (x_0, \ldots, x_T) \in \Omega. \qquad (2.32)$$

Then for every $a > 0$, we have

$$|\langle \varphi, \lambda \rangle| \le \langle |\varphi|, \lambda \rangle \le \left\langle \|\varphi\|_{0:T} \sum_{t=0}^{T} f_t^n, \lambda \right\rangle = \|\varphi\|_{0:T} \sum_{t=0}^{T} \langle f_t^n, \lambda \rangle \qquad (2.33)$$

$$= \|\varphi\|_{0:T} \sum_{t=0}^{T} \langle f_t^n, \lambda_t \rangle = \|\varphi\|_{0:T} \frac{1}{a} \sum_{t=0}^{T} \langle a f_t^n, \lambda_t \rangle$$

$$\le \|\varphi\|_{0:T} \left(\frac{1}{a} \mathcal{D}(\lambda_0, \ldots, \lambda_T) + \frac{1}{a} V(a f_0^n, \ldots, a f_T^n) \right), \qquad (2.34)$$

where (2.33) follows from positivity of λ, (2.32), linearity and the fact that we have $\lambda_t := \lambda|_{C_{0:t}} \in (C_{0:t})^*$, while (2.34) follows from the Fenchel inequality (2.22).

Since $(\lambda_0, \ldots, \lambda_T) \in \mathrm{dom}(\mathcal{D})$ by hypothesis, we can select $a > 0$ such that $\frac{1}{a} \mathcal{D}(\lambda_0, \ldots, \lambda_T) \le \frac{\varepsilon}{2}$. Choose now n in such a way that $\frac{1}{a} V(a f_0^n, \ldots, a f_T^n) \le \frac{\varepsilon}{2}$ for every $s \le T$ (which is possible by Assumption 2.3). Continuing from (2.34), we get

$$|\langle \varphi, \lambda \rangle| \le \|\varphi\|_{0:T} \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \right) \le \varepsilon \|\varphi\|_{0:T}.$$

The result now follows by combining Proposition A.2 and the Daniell–Stone result in Theorem A.3. $\qquad \square$

Lemma 2.23 *Under Assumption 2.3 and if (2.7) holds, equation (2.9) is true for every $c \in C_{0:T}$, with a minimum in place of the infimum.*

Proof Combining Lemmas 2.18, 2.20 and 2.22, we have

$$\pi(c) = \min_{\substack{\lambda \in (C_{0:T})^*, \\ \lambda \ge 0, \lambda(1)=1}} \left(\langle c, \lambda \rangle + \pi^*(\lambda) \right)$$

$$= \min_{\substack{\lambda \in (C_{0:T})^*, \\ \lambda \ge 0, \lambda(1)=1}} \left(\langle c, \lambda \rangle + \mathcal{D}(\lambda) + \sigma_{\mathcal{A}}(\lambda) \right)$$

$$= \min_{\substack{\lambda \in (C_{0:T})^*, (\lambda_0, \ldots, \lambda_T) \in \mathrm{dom}(\mathcal{D}), \\ \lambda \ge 0, \lambda(1)=1}} \left(\langle c, \lambda \rangle + \mathcal{D}(\lambda) + \sigma_{\mathcal{A}}(\lambda) \right)$$

$$= \min_{\substack{Q \in \mathrm{Prob}^1(\Omega), \\ Q \in \mathrm{dom}(\mathcal{D})}} \left(E_Q[c] + \mathcal{D}(Q) + \sigma_{\mathcal{A}}(Q) \right) \qquad (2.35)$$

$$= \min_{Q \in \mathrm{Prob}^1(\Omega)} \left(E_Q[c] + \mathcal{D}(Q) + \sigma_{\mathcal{A}}(Q) \right)$$

$$= \min_{Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ} \left(E_Q[c] + \mathcal{D}(Q) \right),$$

where the first equality uses Lemma 2.18, the second Lemma 2.20, and the third and fifth equalities the fact that $\mathcal{D}$ is bounded from below by $S^U(0)$ by Remark 2.21, hence $(\lambda_0, \ldots, \lambda_T) \in \text{dom}(\mathcal{D})$ if and only if $\mathcal{D}(\lambda_0, \ldots, \lambda_T) < \infty$. Moreover, in (2.35), we use Lemma 2.22 and identify probability measures $Q \in \text{Prob}^1(\Omega)$ and their induced functionals, as well as the marginals Q_t of such measures, with the restrictions of such functionals to $C_{0:t}$. Finally, in the last equality, we just use the definition of $\sigma_{\mathcal{A}}$ and the fact that $\pi(c) < \infty$ by Lemma 2.18. $\qquad\square$

Lemma 2.24 *Under Assumption* 2.3 *and if* (2.7) *holds, the sublevel set*

$$\{Q \in \text{Prob}^1(\Omega) \cap \mathcal{A}^\circ : \mathcal{D}(Q) \le b\}$$

is $\sigma((C_{0:T})^*, C_{0:T})|_{\text{Prob}^1(\Omega)}$-*(sequentially) compact for every* $b \in \mathbb{R}$.

Proof To begin with, we show that $\{\lambda \in (C_{0:T})^* : \lambda \ge 0, \lambda(1) = 1, \pi^*(\lambda) \le b\}$ is weak* compact. First, we prove that $B := \{\lambda \in (C_{0:T})^* : \pi^*(\lambda) \le b\}$ is weak* compact. Observe that by (2.25), we have for every $r > 0$ and $\lambda \in (C_{0:T})^*$ with $\pi^*(\lambda) \le b$ that

$$
\begin{aligned}
\sup_{\substack{c \in C_{0:T}, \\ \|c\|_{0:T} \le r}} |\langle c, \lambda \rangle| &= \sup_{\substack{c \in C_{0:T}, \\ \|c\|_{0:T} \le r}} \langle c, \lambda \rangle \\
&\le \sup_{\substack{c \in C_{0:T}, \\ \|c\|_{0:T} \le r}} \left(-\pi(-c) \right) + \pi^*(\lambda) \\
&\le b + \sup_{\substack{c \in C_{0:T}, \\ \|c\|_{0:T} \le r}} \left(-\pi(-c) \right).
\end{aligned}
\tag{2.36}
$$

Now since $-\pi(\cdot)$ is real-valued, convex and continuous on $C_{0:T}$ by Lemma 2.18, it follows from Aliprantis and Border [1, Theorem 5.43] that the right-hand side in (2.36) is finite for some $r > 0$. Thus the operator norms of elements of the set B are uniformly bounded, and so B is contained in some (weak* compact, by the Banach–Alaoglu theorem) ball of $(C_{0:T})^*$. Since π^* is weak* lower semicontinuous by its very definition, its sublevel sets are weak* closed. This concludes the proof of weak* compactness of B. Next,

$$\{\lambda \in (C_{0:T})^* : \lambda \ge 0, \lambda(1) = 1, \pi^*(\lambda) \le b\}$$
$$= B \cap \{\lambda \in (C_{0:T})^* : \lambda(1) = 1\} \cap \bigcap_{\varphi \in C_{0:T}, \varphi \ge 0} \{\lambda \in (C_{0:T})^* : \lambda(\varphi) \ge 0\}$$

is the intersection of a weak* compact set and weak* closed sets; hence it is weak* compact. Combining the fact that $\sigma_{\mathcal{A}} = \delta_{\mathcal{A}^\circ}$ and Lemma 2.20,

$$\{\lambda \in (C_{0:T})^* : \lambda \ge 0, \lambda(1) = 1, \pi^*(\lambda) \le b\}$$
$$= \{\lambda \in (C_{0:T})^* : \lambda \ge 0, \lambda(1) = 1, \mathcal{D}(\lambda) \le b\} \cap \mathcal{A}^\circ.$$

By Lemma 2.22, we can identify normalised nonnegative functionals in $\mathrm{dom}(\mathcal{D})$ and measures in $\mathrm{Prob}^1(\Omega)$, so that by a slight abuse of notation, we can write

$$\{\lambda \in (C_{0:T})^* : \lambda \geq 0, \lambda(1) = 1, \mathcal{D}(\lambda) \leq b\} \cap \mathcal{A}^\circ$$
$$= \{Q \in \mathrm{Prob}_1(\Omega), \mathcal{D}(Q) \leq b\} \cap \mathcal{A}^\circ.$$

Moving to sequential compactness, the topology $\tau = \sigma((C_{0:T})^*, C_{0:T})|_{\mathrm{Prob}^1(\Omega)}$ induced on $\mathrm{Prob}^1(\Omega)$ is the topology generated by the 1-Wasserstein distance by Bolley [11, Theorem A.2] (see also the discussion in the introduction of Bolley [12], and Villani [38, Definition 6.8.(iv) and Theorem 6.9]). By the above argument, the set $\{Q \in \mathrm{Prob}_1(\Omega) : \mathcal{D}(Q) \leq b\}$ is then a compact subset of $(\mathrm{Prob}^1(\Omega), \tau)$, which is then 1-Wasserstein sequentially compact. As $\mathcal{A}^\circ$ is weak* closed by its definition, the result follows. □

Lemma 2.25 *Under Assumption* 2.3, *for every lower semicontinuous functional*

$$c : \Omega \to (-\infty, +\infty]$$

satisfying (2.8), *the duality* (2.9) *holds, and if* $\pi(c) < \infty$, *the infimum in* (2.9) *is a minimum.*

Proof Take c as in the statement. Observe that from the definition of π and the Fenchel inequality on S^U, for any $Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$, we have

$$\pi(c) = \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} S^U(\varphi)$$

$$\leq \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} \left((S^U)^*(Q_0, \ldots, Q_T) + E_Q\left[\sum_{t=0}^{T} \varphi_t \right] \right)$$

$$\leq \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} \left((S^U)^*(Q_0, \ldots, Q_T) + E_Q\left[\sum_{t=0}^{T} \varphi_t + z \right] \right)$$

$$\leq \sup_{z \in -\mathcal{A}} \sup_{\varphi \in \Phi_z(c)} \left(E_Q[c] + \mathcal{D}(Q) \right)$$

$$= E_Q[c] + \mathcal{D}(Q),$$

where the second inequality follows from $Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^0$ and the third is a consequence of Lemma 2.20. Hence

$$\pi(c) \leq \inf_{Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ} \left(E_Q[c] + \mathcal{D}(Q) \right). \tag{2.37}$$

The case $\pi(c) = \infty$ is thus trivial, and we focus on the case $\pi(c) < \infty$. Let $c^A(x) := -A(1 + \sum_{t=0}^{T} \sum_{j=1}^{d} |x_t^j|)$, for $x \in \Omega$. Then $c \geq c^A \in C_{0:T}$ and we have $\pi(c^A) \leq \pi(c) < \infty$, as can be easily verified. A standard argument produces a sequence $(c_n) \subseteq C_{0:T}$ with $c_n \uparrow c$ pointwise on Ω. We claim that given a sequence of

optima for the dual problems of $\pi(c_n)$, taking a suitable converging subsequence, the limit $\widehat{Q}$ satisfies $\widehat{Q} \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$ and $E_{\widehat{Q}}[c] + \mathcal{D}(\widehat{Q}) \leq \pi(c)$. This and (2.37) then imply (2.9).

To prove the claim, recall from Lemma 2.23 and $\infty > \pi(c) \geq \pi(c_n)$ that each dual problem for $\pi(c_n)$ admits an optimum; call it $Q^n \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ$. We proceed by observing that $\mathcal{D}(Q^n) \in \mathbb{R}$ for every n and

$$\pi(c_n) = E_{Q^n}[c_n] + \mathcal{D}(Q^n) \geq -E_{Q^n}\left[\|c_1\|_{0:T}\left(1 + \sum_{t=0}^{T} \eta_t\right)\right] + \mathcal{D}(Q^n), \quad (2.38)$$

where we set $\eta_t(x_t) = \sum_{j=1}^{d} |x_t^j|$ for $x_t \in K_t$. Now by the Fenchel inequality (2.22), mimicking the argument in (2.34),

$$E_{Q^n}\left[\|c_1\|_{0:T} \sum_{t=0}^{T} \eta_t\right] \leq \frac{1}{2}\mathcal{D}(Q^n) + \frac{1}{2}V(2\|c_1\|_{0:T}\eta_0, \ldots, 2\|c_1\|_{0:T}\eta_T).$$

Going back to (2.38), we then get

$$\pi(c_n) \geq \zeta + \frac{1}{2}\mathcal{D}(Q^n), \quad (2.39)$$

where $\zeta \in \mathbb{R}$ is a constant depending on $c_1, V, \eta_0, \ldots, \eta_T$. As $\pi(c_n) \leq \pi(c) < \infty$, we conclude that $\sup_n \mathcal{D}(Q^n) < \infty$, which in turn implies that the sequence (Q_n) lies in $\{Q \in \mathrm{Prob}^1(\Omega) \cap \mathcal{A}^\circ : \mathcal{D}(Q) \leq b\}$ for $b \in \mathbb{R}$ big enough. We know that the latter set is weak* sequentially compact by Lemma 2.24. Thus we can extract a subsequence, which we call again (Q^n), weak* converging to a $\widehat{Q} \in \mathrm{Prob}^1 \cap \mathcal{A}^\circ$. Now it is easily seen that

$$E_{\widehat{Q}}[c] + \mathcal{D}(\widehat{Q}) = \lim_n E_{\widehat{Q}}[c_n] + \mathcal{D}(\widehat{Q}) \leq \lim_n \liminf_m \left(E_{Q^m}[c_n] + \mathcal{D}(Q^m)\right)$$

$$\leq \lim_n \liminf_m \left(E_{Q^m}[c_m] + \mathcal{D}(Q^m)\right)$$

$$= \liminf_m \left(E_{Q^m}[c_m] + \mathcal{D}(Q^m)\right) = \lim_m \pi(c^m) \leq \pi(c),$$

where the first inequality uses the fact that $Q \mapsto E_Q[c_n] + \mathcal{D}(Q)$ is weak* lower semicontinuous as a sum of weak* lower semicontinuous functionals, and the second uses that $c_n \leq c_m$ if $m \geq n$.   □

2.5 Convergence of EMOT

In this section, we study some stability and convergence results for the EMOT problem. In particular, we show how under suitable convergence assumptions on the penalty terms, one can see the classical MOT as a limit case for EMOT.

We suppose that for each $n \in \mathbb{N} \cup \{\infty\}$, we are given a functional U_n and a set $\mathcal{A}_n \subseteq C_{0:T}$. We denote the corresponding value as in (2.3) by $\pi_n(c)$.

Proposition 2.26 *Suppose that for each $n \in \mathbb{N} \cup \{\infty\}$, the assumptions of Theorem 2.4 hold for $\pi_n(c)$ and that $\pi_n(c) < \infty$. Suppose that*

$$\mathcal{D}_n(Q) + \sigma_{\mathcal{A}_n}(Q) \uparrow \mathcal{D}_\infty(Q) + \sigma_{\mathcal{A}_\infty}(Q) \qquad as\ n \to \infty \qquad (2.40)$$

for every $Q \in \mathrm{Prob}^1(\Omega)$. Then $\pi_n(c) \uparrow \pi_\infty(c)$ for every $c : \Omega \to (-\infty, +\infty]$ which is lower semicontinuous and satisfies (2.8).

Proof By Lemma 2.23, the dual problem for $\pi(c_n)$ admits an optimum Q^n in the set $\mathrm{Prob}^1(\Omega) \cap \mathcal{A}_n^\circ$ for each $n \in \mathbb{N}$. Observe that $\infty > \pi_\infty(c) \geq \sup_n \pi_n(c) = \lim_n \pi_n(c)$ and that with an argument similar to the one yielding (2.39),

$$\pi_n(c) = E_{Q^n}[c] + \mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n)$$

$$\geq E_{Q^n}\left[-\|c\|_{0:T} \sum_{t=0}^T \psi_t \right] + \mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n)$$

$$\geq -\frac{1}{2}\pi_n^*(Q^n) + \frac{1}{2}\pi_n\left(-2\|c\|_{0:T} \sum_{t=0}^T \psi_t \right) + \mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n)$$

$$= \frac{1}{2}\pi_n\left(-2\|c\|_{0:T} \sum_{t=0}^T \psi_t \right) + \frac{1}{2}\mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n)$$

$$\geq \frac{1}{2}\pi_1\left(-2\|c\|_{0:T} \sum_{t=0}^T \psi_t \right) + \frac{1}{2}\mathcal{D}_1(Q^n) + \sigma_{\mathcal{A}_1}(Q^n),$$

using Lemma 2.20 to get the equality in the fourth line. As a consequence, for some constant η,

$$\infty > \pi_\infty(c) \geq \pi_n(c) \geq \eta + \frac{1}{2}\mathcal{D}_1(Q^n) + \sigma_{\mathcal{A}_1}(Q^n).$$

Hence all the measures $Q^n, n \in \mathbb{N}$, belong to a sublevel set of the form

$$\{Q \in \mathrm{Prob}^1(\Omega) \cap (\mathcal{A}_1)^\circ : \mathcal{D}_1(Q) \leq b\}$$

which is $\sigma(\mathrm{Prob}^1(\Omega), C_{0:T})$-(sequentially) compact by Lemma 2.24. Extract a subsequence, again called (Q^n), converging to a limit $Q^\infty \in \mathrm{Prob}^1(\Omega)$. Since $\mathcal{D}_n$ and $\sigma_{\mathcal{A}_n}$ are lower semicontinuous, so is $\mathcal{D}_n + \sigma_{\mathcal{A}_n}$ for every $n \in \mathbb{N} \cup \{\infty\}$. Hence

$$\mathcal{D}_\infty(Q^\infty) + \sigma_{\mathcal{A}_\infty}(Q^\infty) = \sup_K \left(\mathcal{D}_K(Q^\infty) + \sigma_{\mathcal{A}_K}(Q^\infty) \right)$$

$$\leq \sup_K \liminf_n \left(\mathcal{D}_K(Q^n) + \sigma_{\mathcal{A}_K}(Q^n) \right)$$

$$\leq \sup_K \liminf_n \left(\mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n) \right)$$

$$= \liminf_n \left(\mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n) \right),$$

using (2.40) in the first equality and the second inequality. Up to taking a further subsequence, again called (Q^n), we may assume that the $\liminf$ above is in fact a limit, so that

$$\mathcal{D}_\infty(Q^\infty) + \sigma_{\mathcal{A}_\infty}(Q^\infty) \le \lim_n \left(\mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n) \right).$$

Since $c : \Omega \to (-\infty, +\infty]$ is lower semicontinuous and satisfies (2.8) for some $A \ge 0$, there exists a sequence $(c_n) \subseteq C_{0:T}$ with $c_n \uparrow c$ pointwise on Ω, just as in the proof of Lemma 2.25. By monotone convergence, we then have $E_Q[c] = \sup_n E_Q[c_n]$. We conclude that $Q \mapsto E_Q[c]$ is the supremum of linear functionals, each being continuous with respect to the topology induced by $\sigma((C_{0:T})^*, C_{0:T})$ on $\mathrm{Prob}^1(\Omega)$. Thus $Q \mapsto E_Q[c]$ is lower semicontinuous with respect to that topology, and we obtain $E_{Q^\infty}[c] \le \liminf_n E_{Q^n}[c]$. Passing to a further subsequence, we can assume that $\liminf_n E_{Q^n}[c] = \lim_n E_{Q^n}[c]$. From the previous arguments, we then get

$$\pi_\infty(c) \le E_{Q^\infty}[c] + \mathcal{D}_\infty(Q^\infty) + \sigma_{\mathcal{A}_\infty}(Q^\infty)$$

$$\le \lim_n E_{Q^n}[c] + \mathcal{D}_\infty(Q^\infty) + \sigma_{\mathcal{A}_\infty}(Q^\infty)$$

$$\le \lim_n E_{Q^n}[c] + \lim_n \left(\mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n) \right)$$

$$= \lim_n \left(E_{Q^n}[c] + \mathcal{D}_n(Q^n) + \sigma_{\mathcal{A}_n}(Q^n) \right) = \lim_n \pi_n(c),$$

where we use that the Q^n are optima. Since we already have $\lim_n \pi_n(c) \le \pi_\infty(c)$, this concludes the proof of $\pi_n(c) \uparrow \pi_\infty(c)$. $\qquad\square$

3 Additive structure

In Sect. 2, we did not require any particular structural form of the functionals $\mathcal{D}, U$. We now assume an additive structure of U and, complementarily, an additive structure of $\mathcal{D}$. **In the entire Sect. 3**, we take for each $t = 0, \dots, T$ a vector subspace $\mathcal{E}_t \subseteq C_t = C_{t:t}$ such that $\mathcal{E}_t + \mathbb{R} = \mathcal{E}_t$ and set $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$. Observe that we automatically have $\mathcal{E} + \mathbb{R}^{T+1} = \mathcal{E}$. It is also clear that $\mathcal{E}$ is a subspace of $(C_{0:T})^{T+1}$ if we interpret $\mathcal{E}_0, \dots, \mathcal{E}_T$ as subspaces of $C_{0:T}$. We also mention here that up to now, we used for $\lambda \in (C_{0:T})^*$ (resp. for a measure $\mu \in \mathrm{ca}(\Omega)$) the notation λ_t (resp. μ_t) for the restriction to $C_{0:t}$ (resp. for the marginal on $\Omega_{0:t}$). This was motivated by the fact that we were considering general $\mathcal{E}_t \subseteq C_{0:t}$. Since we now mostly work with $\mathcal{E}_t \subseteq C_t$, we change notation slightly.

Notation 3.1 Throughout Sects. 3–5, given $\lambda \in (C_{0:T})^*$ (or a measure $\mu \in \mathrm{ca}(\Omega)$), we use the notation λ_t (resp. μ_t) for the restriction to C_t (resp. for the marginal on $\times_{j=1}^d K_t^j$).

3.1 Additive structure of U

Setup 3.2 We consider a proper concave functional $U_t : \mathcal{E}_t \to [-\infty, +\infty)$ for every $t = 0, \dots, T$. We define $\mathcal{D}_t$ on $\mathrm{ca}^1(K_t)$ similarly to (2.21) as

$$\mathcal{D}_t(\gamma_t) := \sup_{\varphi_t \in \mathcal{E}_t} \left(U_t(\varphi_t) - \int_{K_t} \varphi_t \, d\gamma_t \right), \qquad \gamma_t \in \mathrm{ca}^1(K_t),$$

and observe that $\mathcal{D}_t$ can also be viewed as defined on $\mathrm{ca}^1(\Omega)$ by using for $\gamma \in \mathrm{ca}^1(\Omega)$ the marginals $\gamma_0, \ldots, \gamma_T$ and setting $\mathcal{D}_t(\gamma) := \mathcal{D}_t(\gamma_t)$. We now define, for each $\varphi \in \mathcal{E}$, $U(\varphi) := \sum_{t=0}^T U_t(\varphi_t)$ and define $\mathcal{D}$ on $\mathrm{ca}^1(\Omega)$ using (2.21). Recall from (1.4) that

$$S^U(\varphi) := \sup_{\beta \in \mathbb{R}^{T+1}} \left(U(\varphi + \beta) - \sum_{t=0}^T \beta_t \right), \qquad \varphi \in \mathcal{E},$$

$$S^{U_t}(\varphi_t) := \sup_{\alpha \in \mathbb{R}} \left(U(\varphi_t + \alpha) - \alpha \right), \qquad \varphi_t \in \mathcal{E}_t.$$

Lemma 3.3 *In Setup* 3.2 *and under the convention* $+\infty - \infty = -\infty$, *we have*

$$\mathcal{D}(\gamma) = \sum_{t=0}^T \mathcal{D}_t(\gamma) = \sum_{t=0}^T \mathcal{D}_t(\gamma_t), \qquad \forall \gamma \in \mathrm{ca}^1(\Omega),$$

$$S^U(\varphi) = \sum_{t=0}^T S^{U_t}(\varphi_t), \qquad \forall \varphi \in \mathcal{E}, \tag{3.1}$$

and for all $\varphi \in \mathcal{E}$ *that*

$$S^U(\varphi + \beta) = S^U(\varphi) + \sum_{t=0}^T \beta_t, \qquad \forall \beta \in \mathbb{R}^{T+1},$$

$$S^{U_t}(\varphi_t + \beta) = S^{U_t}(\varphi_t) + \beta, \qquad \forall \beta \in \mathbb{R}.$$

Proof The simple proof is omitted. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ □

3.2 Duality for the general cash-additive setup

As a consequence of Theorem 2.4, we now obtain a duality in a general cash-additive setup.

Theorem 3.4 *Suppose that* $\mathcal{E}_t \subseteq C_t$ *with* $X_t \in \mathcal{E}_t$ *and that* $U_t : \mathcal{E}_t \to \mathbb{R}$ *is a concave, cash-additive functional null in* 0. *Set* $U(\varphi) := \sum_{t=0}^T U_t(\varphi_t)$ *for* $\varphi \in \mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ *and suppose that Assumption* 2.3 *is fulfilled. Recall the formulation of* $\mathcal{S}_{\mathrm{sub}}(c)$ *from* (1.2) *and consider for every* $t = 0, \ldots, T$ *the penalisations*

$$\mathcal{D}_t(Q_t) := \sup_{\varphi_t \in \mathcal{E}_t} \left(U_t(\varphi_t) - \int_{K_t} \varphi_t \, \mathrm{d}Q_t \right) \qquad \text{for } Q_t \in \mathrm{Prob}^1(K_t). \tag{3.2}$$

Let $c : \Omega \to (-\infty, +\infty]$ *be lower semicontinuous and such that* (2.8) *holds. Then*

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^T \mathcal{D}_t(Q_t) \right) = \sup \left\{ \sum_{t=0}^T U_t(\varphi_t) : \varphi \in \mathcal{S}_{\mathrm{sub}}(c) \right\} = \pi(c), \tag{3.3}$$

and the infimum in (3.3) *is a minimum provided that* $\pi(c) < \infty$.

Proof Let $\mathcal{D}$ be defined as in (2.21). Observe that we are in Setup 3.2. Lemma 3.3 tells us that $S^U(\varphi) = \sum_{t=0}^{T} S^{U_t}(\varphi_t) = \sum_{t=0}^{T} U_t(\varphi_t)$, since $U_0, \ldots, U_T$ are cash-additive, and that $\mathcal{D}$ coincides on Mart(Ω) with the penalisation term $Q \mapsto \sum_{t=0}^{T} \mathcal{D}_t(Q_t)$, where $\mathcal{D}_t(\mathcal{Q}_t)$ is given in (3.2). So $S^U = U$ and dom$(S^U) = \mathcal{E}$, all the assumptions of Theorem 2.4 are fulfilled, and so we can apply Corollary 2.12 which together with Remark 2.11 yields exactly (3.3). $\qquad\square$

3.3 Additive structure of $\mathcal{D}$

The results of this subsection will be applied in Sects. 4.3 and 4.4. In the spirit of Remark 2.15, we now reverse the procedure taken in the previous subsection: We start from some functionals $\mathcal{D}_t$ on ca$^1(K_t)$ for $t = 0, \ldots, T$ and build an additive functional $\mathcal{D}$ on ca$^1(\Omega)$. Our aim is to find the counterparts of the results in Sect. 3.1.

Setup 3.5 We consider a proper, convex, $\sigma(\text{ca}^1(K_t), \mathcal{E}_t)$-lower semicontinuous functional $\mathcal{D}_t : \text{ca}^1(K_t) \to (-\infty, +\infty]$ for every $t = 0, \ldots, T$. We extend the functionals $\mathcal{D}_t$ to ca$^1(\Omega)$ by using, for any $\gamma \in \text{ca}(\Omega)$, the marginals $\gamma_0, \ldots, \gamma_T$. If $\gamma \in \text{ca}^1(\Omega)$, we set

$$\mathcal{D}_t(\gamma) := \mathcal{D}_t(\gamma_t) \qquad \text{and} \qquad \mathcal{D}(\gamma) := \sum_{t=0}^{T} \mathcal{D}_t(\gamma) = \sum_{t=0}^{T} \mathcal{D}_t(\gamma_t).$$

We define $V(\varphi)$ for $\varphi \in \mathcal{E}$ and $V_t(\varphi_t)$ for $\varphi_t \in \mathcal{E}_t$ for $t = 0, \ldots, T$ similarly to (2.23) as

$$V(\varphi) := \sup_{\gamma \in \text{ca}^1(\Omega)} \left(\int_{\Omega} \left(\sum_{t=0}^{T} \varphi_t \right) d\gamma - \mathcal{D}(\gamma) \right),$$

$$V_t(\varphi_t) := \sup_{\gamma \in \text{ca}^1(K_t)} \left(\int_{K_t} \varphi_t \, d\gamma - \mathcal{D}_t(\gamma) \right).$$

We define on $\mathcal{E}$ the functional $U(\cdot) = -V(-\cdot)$ and similarly $U_t(\cdot) = -V_t(-\cdot)$ on $\mathcal{E}_t$ for $t = 0, \ldots, T$. Finally, $S^U(\varphi)$, $S^{U_0}(\varphi_0), \ldots, S^{U_T}(\varphi_T)$ are defined as in Setup 3.2.

Lemma 3.6 *In Setup 3.5, we have the following*:

1) *$\mathcal{D}_0, \ldots, \mathcal{D}_T$ as well as $\mathcal{D}$ are $\sigma(\text{ca}^1(\Omega), \mathcal{E})$-lower semicontinuous.*

2) *Under the additional assumption that* dom$(\mathcal{D}_t) \subseteq \text{Prob}^1(K_t)$ *for* $t = 0, \ldots, T$, *we have for any* $\varphi = (\varphi_0, \ldots, \varphi_T) \in \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ *that*

$$U(\varphi) = \sum_{t=0}^{T} U_t(\varphi_t) = \sum_{t=0}^{T} -V_t(-\varphi_t), \qquad S^U(\varphi) = \sum_{t=0}^{T} S^{U_t}(\varphi_t).$$

Proof The proof is omitted. $\qquad\square$

3.4 Divergences induced by utility functions

In this section, we provide the exact formulation of the divergences induced by utility functions $u_t : \mathbb{R} \to [-\infty, +\infty)$, distinguishing the two cases $\mathrm{dom}(u_t) = \mathbb{R}$ and $\mathrm{dom}(u_t) \supseteq [0, \infty)$.

Assumption 3.7 We consider concave, upper semicontinuous nondecreasing functions $u_0, \dots, u_T : \mathbb{R} \to [-\infty, +\infty)$ with $u_0(0) = \cdots = u_T(0) = 0$ and $u_t(x) \leq x$, $\forall x \in \mathbb{R}$ (that is, $1 \in \partial u_0(0) \cap \cdots \cap \partial u_T(0)$). For each $t = 0, \dots, T$, we define $v_t(x) := -u_t(-x)$, $x \in \mathbb{R}$, and

$$v_t^*(y) := \sup_{x \in \mathbb{R}} \left(xy - v_t(x) \right) = \sup_{x \in \mathbb{R}} \left(u_t(x) - xy \right), \qquad y \in \mathbb{R}. \qquad (3.4)$$

Remark 3.8 We observe that $v_t(y) = v_t^{**}(y) = \sup_{x \in \mathbb{R}} (xy - v_t^*(y))$ for all $y \in \mathbb{R}$ by the Fenchel–Moreau theorem and that v_t^* is convex, lower semicontinuous and bounded from below on $\mathbb{R}$. Assumption 3.7 is satisfied by a wide range of utility functions.

Fix $\widehat{\mu}_t \in \mathrm{Meas}(K_t)$. We define, for $\mu \in \mathrm{Meas}(K_t)$,

$$\mathcal{D}_{v_t^*, \widehat{\mu}_t}(\mu) := \begin{cases} \int_{K_t} v_t^*(\frac{\mathrm{d}\mu}{\mathrm{d}\widehat{\mu}_t}) \, \mathrm{d}\widehat{\mu}_t & \text{if } \mu \ll \widehat{\mu}_t, \\ \infty & \text{otherwise.} \end{cases} \qquad (3.5)$$

In the next two results, whose proofs are postponed to Appendix A.2, we provide the dual representation of the divergence terms.

Proposition 3.9 *Take $u_0, \dots, u_T$ satisfying Assumption 3.7 with*

$$\mathrm{dom}(u_0) = \cdots = \mathrm{dom}(u_T) = \mathbb{R},$$

consider closed (possibly noncompact) $K_0, \dots, K_T \subseteq \mathbb{R}$ and let $\widehat{\mu}_t \in \mathrm{Meas}(K_t)$, $t = 0, \dots, T$. Then

$$\mathcal{D}_{v_t^*, \widehat{\mu}_t}(\mu) = \sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} \varphi_t(x_t) \, \mathrm{d}\mu(x_t) - \int_{K_t} v_t \left(\varphi_t(x_t) \right) \mathrm{d}\widehat{\mu}_t(x_t) \right). \qquad (3.6)$$

Let $\widehat{Q}_t \in \mathrm{Prob}(K_t)$ and for $\mu \in \mathrm{Meas}(K_t)$, let $\mu = \mu_a + \mu_s$ be the Lebesgue decomposition of μ with respect to $\widehat{Q}_t$, where $\mu_a \ll \widehat{Q}_t$ and $\mu_s \perp \widehat{Q}_t$. Set

$$(v_t^*)'_\infty := \lim_{y \to \infty} \frac{v_t^*(y)}{y}, \qquad t = 0, \dots, T.$$

As $u_t(0) = 0$, $(v_t^*)'_\infty \in [0, \infty]$ since $v_t^*(y) \geq u_t(0) - 0y = 0$. Then we can define, for $\mu \in \mathrm{Meas}(K_t)$,

$$\mathcal{F}_t(\mu | \widehat{Q}_t) := \int_{K_t} v_t^* \left(\frac{\mathrm{d}\mu_a}{\mathrm{d}\widehat{Q}_t} \right) \mathrm{d}\widehat{Q}_t + (v_t^*)'_\infty \, \mu_s(K_t),$$

where we use the convention $\infty \cdot 0 = 0$ if $(v_t^*)'_\infty = \infty$ and $\mu_s(K_t) = 0$. Observe that the restriction of $\mathcal{F}(\cdot|\widehat{Q}_t)$ to Meas(K_t) coincides with the functional in Liero et al. [30, Equation (2.35)] with $F = v_t^*$, and that whenever dom$(u_t) = \mathbb{R}$, we have $(v_t^*)'_\infty = \lim_{y\to\infty} \frac{v_t^*(y)}{y} = \infty$ and $\mathcal{F}_t(\cdot|\widehat{Q}_t)$ coincides with $\mathcal{D}_{v_t^*,\widehat{Q}_t}(\cdot)$ on Meas(K_t); see (3.5).

Proposition 3.10 *Suppose that* $u_0, \ldots, u_T : \mathbb{R} \to [-\infty, +\infty)$ *satisfy Assumption* 3.7 *and* $K_0, \ldots, K_T \subseteq \mathbb{R}$ *are compact. If* $\widehat{Q}_t \in \mathrm{Prob}(K_t)$ *for all* $t \in \{0, \ldots, T\}$ *has full support, then*

$$\mathcal{F}_t(\mu|\widehat{Q}_t) = \sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} \varphi_t(x_t) \, d\mu(x_t) - \int_{K_t} v_t\big(\varphi_t(x_t)\big) \, d\widehat{Q}_t(x_t) \right). \quad (3.7)$$

Example 3.11 The requirement that $\widehat{Q}_0, \ldots, \widehat{Q}_T$ have full support is crucial for the proof of Proposition 3.10. We provide a simple example to show that (3.7) does not hold in general when that assumption is not fulfilled. Take $K = \{-2, 0, 2\}$, $\widehat{Q} = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_{+2}$, $\mu = \delta_0$, $u(x) = \frac{x}{x+1}$ for $x \geq -1$ and $u(x) = -\infty$ for $x < -1$. It is easy to see that v^* associated via (3.4) is given by $v^*(y) = 1 + y - 2\sqrt{y}$ for $y \geq 0$ and $v^*(y) = -\infty$ for $y < 0$, so that $(v_t^*)'_\infty = 1$. It is also easy to see that $\mu \perp \widehat{Q}$; hence $\mu_a = 0$ and $\mu_s = \mu$ in the Lebesgue decomposition with respect to $\widehat{Q}$. Hence $\mathcal{F}(\mu|\widehat{Q}) = 1 + 1\mu(K) = 2$. At the same time, we see that taking $\varphi_N \in \mathcal{C}_b(K)$ defined via $\varphi_N(-2) = \varphi_N(2) = 0, \varphi_N(0) = -N$ (observe that $u(\varphi_N) \notin \mathcal{C}_b(K)$ for N sufficiently large), we have

$$\sup_{\varphi \in \mathcal{C}_b(K)} \left(\int_K \varphi \, d\mu - \int_K v(\varphi) \, d\widehat{Q} \right) = \sup_{\varphi \in \mathcal{C}_b(K)} \left(\int_K u(\varphi) \, d\widehat{Q} - \int_K \varphi \, d\mu \right)$$

$$\geq \sup_N \left(\int_K u(\varphi_N) \, d\widehat{Q} - \int_K \varphi_N \, d\mu \right)$$

$$\geq \sup_N \left(0\frac{1}{2} + 0\frac{1}{2} - (-N) \right) = \infty.$$

4 Applications in the compact case

We suppose **in the entire Sect. 4** that the following requirements are fulfilled.

Standing Assumption 4.1 Let $d = 1$ and $\Omega := K_0 \times \cdots \times K_T$ for compact sets $K_0, \ldots, K_T \subseteq \mathbb{R}$, $K_0 = \{x_0\}$ for some $x_0 \in \mathbb{R}$, the functional $c : \Omega \to (-\infty, +\infty]$ is lower semicontinuous, $\widehat{Q} \in \mathrm{Mart}(\Omega)$ is a given martingale measure with marginals $\widehat{Q}_0, \ldots, \widehat{Q}_T$, and $c \in L^1(\widehat{Q})$.

Under this assumption, $C_{0:T} = \mathcal{C}_b(\Omega)$ and $(C_{0:T})^* = \mathrm{ca}(\Omega) = \mathrm{ca}^1(\Omega)$. We observe that the stock price (X_t) is bounded due to the compactness of $K_0, \ldots, K_T$. As a consequence, if we consider for example the call option $(X_t - \alpha)^+, \alpha \in \mathbb{R}$, then it is also bounded on Ω. The selection $\mathcal{E} \subseteq \mathcal{C}_b(K_0) \times \cdots \times \mathcal{C}_b(K_T)$ is appropriate in this context.

4.1 Subhedging with vanilla options

As in Beiglböck et al. [4], we suppose in Sect. 4.1 that the elements in $\mathcal{E}_t$ represent portfolios obtained combining call options with maturity t, units of the underlying stock at time t and deterministic amounts, that is, $\mathcal{E}_t$ consists of all the functions in $\mathcal{C}_b(K_t)$ of the form

$$\varphi_t(x_t) = a + b x_t + \sum_{n=1}^{N} c_n (x_t - \alpha_n)^+ \qquad \text{with } N \geq 1, a, b, c_n, \alpha_n \in \mathbb{R}, x_t \in K_t, \tag{4.1}$$

and take $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$. As one can see in the proofs of Corollaries 4.3 and 4.5, which are the core content of this section, one could as well take instead the space $\mathcal{E} = \mathcal{C}_b(K_0) \times \cdots \times \mathcal{C}_b(K_T)$ and preserve the validity of (4.3) and (4.5).

In all the results in Sect. 4.1, the functional U is real-valued on the whole space $\mathcal{E}$ and cash-additive, which yields $\mathrm{dom}(U) = \mathrm{dom}(S^U) = \mathcal{E}$. Thus we can use Corollaries 2.5 and 2.12, in particular (2.16) and (2.17), in the case $\mathrm{dom}(S^U) = \mathcal{E}$.

Take $U_t(\varphi_t) := \int_{K_t} u_t(\varphi_t(x_t)) \mathrm{d}\widehat{Q}_t(x_t)$. We work with the associated (one-dimensional) optimised certainty equivalent S^{U_t} that we rename, for $\varphi_t \in \mathcal{C}_b(K_t)$, as

$$S^{U_t}(\varphi_t) = \sup_{\alpha \in \mathbb{R}} \left(\int_{K_t} u_t\big(\varphi_t(x_t) + \alpha\big) \mathrm{d}\widehat{Q}_t(x_t) - \alpha \right) =: U_{\widehat{Q}_t}(\varphi_t). \tag{4.2}$$

We observe that Assumption 3.7 does *not* impose that the functions u_t are real-valued on all of $\mathbb{R}$. Nevertheless, for the functional $U_{\widehat{Q}_t}$, it can be easily shown that we have the following result whose proof is omitted.

Lemma 4.2 *Under Assumption 3.7, for each $t = 0, \ldots, T$, $U_{\widehat{Q}_t}$ is real-valued on $\mathcal{C}_b(K_t)$ and null in 0, concave, nondecreasing and cash-additive.*

Corollary 4.3 *Take $u_0, \ldots, u_T$ satisfying Assumption 3.7 and suppose that we have $\mathrm{dom}(u_0) = \cdots = \mathrm{dom}(u_T) = \mathbb{R}$. Then*

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) \right) = \sup \left\{ \sum_{t=0}^{T} U_{\widehat{Q}_t}(\varphi_t) : \varphi \in \mathcal{S}_{\mathrm{sub}}(c) \right\}. \tag{4.3}$$

Moreover, if the left-hand side of (4.3) is finite, a minimum point exists.

Proof Set $U(\varphi) = \sum_{t=0}^{T} U_{\widehat{Q}_t}(\varphi_t)$ for $\varphi \in \mathcal{E}$. By Lemma 4.2, for each $t = 0, \ldots, T$ the monotone concave functional $\varphi_t \mapsto U_{\widehat{Q}_t}(\varphi_t)$ is well defined, finite-valued, concave and nondecreasing on all of $\mathcal{C}_b(K_t)$. Hence by the extended Namioka–Klee theorem (see [8]), it is norm-continuous on $\mathcal{C}_b(K_t)$. We observe that in this case, we are in

Setup 3.2 and can apply (3.1) from Lemma 3.3. We have for every $Q \in \mathrm{Mart}(\Omega)$ that

$$
\begin{aligned}
\mathcal{D}(Q) &:= \sup_{\varphi \in \mathcal{E}} \left(U(\varphi) - \sum_{t=0}^{T} \int_{K_t} \varphi_t \, \mathrm{d}Q_t \right) \\
&= \sum_{t=0}^{T} \sup_{\varphi_t \in \mathcal{E}_t} \left(U_{\widehat{Q}_t}(\varphi_t) - \int_{K_t} \varphi_t \, \mathrm{d}Q_t \right) \\
&= \sum_{t=0}^{T} \sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(U_{\widehat{Q}_t}(\varphi_t) - \int_{K_t} \varphi_t \, \mathrm{d}Q_t \right) \qquad (4.4) \\
&= \sum_{t=0}^{T} \sup_{\varphi \in \mathcal{C}_b(K_t), \alpha_t \in \mathbb{R}} \left(\int_{K_t} u_t(\varphi_t + \alpha_t) \, \mathrm{d}\widehat{Q}_t - \int_{K_t} (\varphi_t + \alpha_t) \, \mathrm{d}Q_t \right) \\
&= \sum_{t=0}^{T} \sup_{\varphi \in \mathcal{C}_b(K_t)} \left(\int_{K_t} u_t(\varphi_t) \, \mathrm{d}\widehat{Q}_t - \int_{K_t} \varphi_t \, \mathrm{d}Q_t \right) \\
&= \sum_{t=0}^{T} \sup_{\psi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} \psi_t \, \mathrm{d}Q_t - \int_{K_t} v_t(\psi_t) \, \mathrm{d}\widehat{Q}_t \right) = \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t).
\end{aligned}
$$

Indeed, in (4.4), we combine the continuity of $U_{\widehat{Q}_t}$ on $\mathcal{C}_b(K_t)$ with the fact that $\mathcal{E}_t$ consists of all piecewise linear functions on K_t so that $\mathcal{E}_t$ is norm-dense in $\mathcal{C}_b(K_t)$; in the fourth equality, we use the fact that for $\widetilde{\varphi}_t := \varphi_t + \alpha_t$ and for every $Q \in \mathrm{Mart}(\Omega)$, $\int_{K_t} \widetilde{\varphi} \mathrm{d}Q_t = \int_{K_t} \varphi \mathrm{d}\widehat{Q}_t + \alpha_t$; in the fifth equality, we exploit the fact that $\widetilde{\varphi}_t \in \mathcal{E}_t$ for every $\varphi_t \in \mathcal{E}_t$, $\alpha_t \in \mathbb{R}$; and the last equality follows from (3.6).

Using Lemma 3.3 and the fact that $U_{\widehat{Q}_0}, \ldots, U_{\widehat{Q}_T}$ are cash-additive, we obtain $S^U(\varphi) = \sum_{t=0}^{T} S^{U_{\widehat{Q}_t}}(\varphi_t) = \sum_{t=0}^{T} U_{\widehat{Q}_t}(\varphi_t) = U(\varphi)$. By Lemma 4.2, the assumptions of Corollary 2.13 are satisfied so that we obtain

$$
\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c(X)] + \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) \right) = \sup \left\{ \sum_{t=0}^{T} U_{\widehat{Q}_t}(\varphi_t) : \varphi \in \mathcal{S}_{\mathrm{sub}}(c) \right\}.
$$

Existence of optima follows again from Corollary 2.13. □

We stress the fact that in Corollary 4.3, we assume that all the functions $u_0, \ldots, u_T$ are real-valued on all of $\mathbb{R}$. A more general result can be obtained when weakening this assumption, but it requires an additional assumption on the marginals of $\widehat{Q}$.

Corollary 4.4 *Suppose Assumption 3.7 is fulfilled. Assume that $\widehat{Q}_0, \ldots, \widehat{Q}_T$ have full support on $K_0, \ldots, K_T$, respectively. Then (4.3) holds true if we replace $\mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t)$ with $\mathcal{F}_t(Q_t | \widehat{Q}_t)$. Moreover, finiteness of the problem on the left-hand side of (4.3) implies the existence of a minimum.*

Proof The proof carries over almost literally from the proof of Corollary 4.3, except for replacing the reference to Proposition 3.9 with a reference to Proposition 3.10. $\square$

We stress that in Corollary 4.4, we impose the full support property on $K_0, \ldots, K_T$ with respect to their induced (Euclidean) topology. In particular, this means that whenever $k_t \in K_t$ is an isolated point, $\widehat{Q}_t[\{k_t\}] > 0$. This is consistent with the assumption $K_0 = \{x_0\}$, which implies that $\text{Prob}(K_0)$ reduces to the Dirac measure, i.e., $\text{Prob}(K_0) = \{\delta_{x_0}\}$.

We now take $u_t(x) = x$ for $t = 0, \ldots, T$ and get $U_{\widehat{Q}_t}(\varphi_t) = E_{\widehat{Q}_t}[\varphi_t]$. Hence an easy computation yields for all $Q \in \text{Mart}(\Omega)$ that

$$\mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) = \begin{cases} 0 & \text{if } Q_t = \widehat{Q}_t, \\ \infty & \text{otherwise.} \end{cases}$$

Recalling that $\text{Mart}(\widehat{Q}_0, \ldots, \widehat{Q}_T) = \{Q \in \text{Mart}(\Omega) : Q_t = \widehat{Q}_t, \forall t = 0, \ldots, T\}$, we recover from Corollary 4.3 the following result of Beiglböck et al. [4] (under the compactness assumption, which will be dropped in Corollary 5.3).

Corollary 4.5 *We have the equality*

$$\inf_{Q \in \text{Mart}(\widehat{Q}_0, \ldots, \widehat{Q}_T)} E_Q[c] = \sup \left\{ \sum_{t=0}^{T} E_{\widehat{Q}_t}[\varphi_t] : \varphi \in \mathcal{S}_{\text{sub}}(c) \right\}. \tag{4.5}$$

Moreover, if the left-hand side of (4.5) *is finite, a minimum point exists.*

4.2 Subhedging without options

The pricing–hedging duality without options takes the following form.

Corollary 4.6 *We have the equality*

$$\inf_{Q \in \text{Mart}(\Omega)} E_Q[c] = \sup\{m \in \mathbb{R} : \exists \Delta \in \mathcal{H} \text{ with } m + I^{\Delta} \leq c\} =: \Pi^{\text{sub}}(c). \tag{4.6}$$

Moreover, if the left-hand side of (4.6) *is finite, a minimum point exists.*

Proof We take $\mathcal{E}_0 = \cdots = \mathcal{E}_T = \mathbb{R}$ and $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T = \mathbb{R}^{T+1}$. If $u_t(x_t) = x_t$, $t = 0, \ldots, T$, and $\widehat{Q} \in \text{Mart}(\Omega)$, the functional $U_{\widehat{Q}_t}$ defined in (4.2) is given by $U_{\widehat{Q}_t}(m_t) = m_t$ and so $U(m) = \sum_{t=0}^{T} U_{\widehat{Q}_t}(m_t) = \sum_{t=0}^{T} m_t$ for all $m \in \mathcal{E}$. Hence for each $\varphi \in \mathcal{E}$ with $\varphi = (m_0, \ldots, m_T)$, $m \in \mathbb{R}^{T+1}$, we select $U(\varphi) = \sum_{t=0}^{T} m_t$. Then by the definition of $\mathcal{D}$ (see (2.10)), we get

$$\mathcal{D}(\gamma) = \begin{cases} 0 & \text{for } \gamma \in \text{ca}(\Omega) \text{ with } \gamma(\Omega) = 1, \\ \infty & \text{otherwise.} \end{cases}$$

In particular, $\mathcal{D}(Q) = 0$ for every $Q \in \text{Mart}(\Omega)$. Moreover, we observe that we have $S^U(\varphi) = U(\varphi)$ for every $\varphi \in \mathcal{E}$. Applying Corollary 2.13, we get from (2.18) that

$$\inf_{Q \in \text{Mart}(\Omega)} E_Q[c] = \sup \left\{ \sum_{t=0}^{T} m_t : \exists \, \Delta \in \mathcal{H} \text{ such that } \sum_{t=0}^{T} m_t + I^\Delta \leq c \right\}.$$

We recognise on the right-hand side above the right-hand side of (4.6). Finally, the existence of optima follows again from Corollary 2.13.   □

4.3 Penalty terms induced by market prices

In this section, we change our perspective. Instead of starting from a given U, we give a particular form of the penalisation term $\mathcal{D}$ and proceed by identifying the corresponding U in the spirit of Remark 2.15. For each $t = 0, \ldots, T$, we suppose that finite sequences $(c_{t,n})_{1 \leq n \leq N_t}$ in $\mathbb{R}$ and $(f_{t,n})_{1 \leq n \leq N_t}$ in $\mathcal{C}_b(K_t)$ are given. The functions $f_{t,n}$ represent payoffs of options whose prices $c_{t,n}$ are known from the market. We also take $\mathcal{E} = \times_{t=0}^{T} \mathcal{C}_b(K_t)$ and define

$$\text{Mart}_t(K_t) = \{\gamma_t \in \text{Prob}(K_t) : \exists \, Q \in \text{Mart}(\Omega) \text{ with } \gamma_t = Q_t\} \subseteq \text{ca}(K_t).$$

Lemma 4.7 *The set* $\text{Mart}_t(K_t)$ *is* $\sigma(\text{ca}(K_t), \mathcal{C}_b(K_t))$*-compact.*

Proof Consider the topology $\tau = \sigma(\text{ca}(\Omega), \mathcal{C}_b(\Omega))$. We see that $\text{Mart}(\Omega)$ is a τ-closed subset of the τ-compact set $\text{Prob}(\Omega)$ (which is τ-compact since Ω is a compact Polish space, see [1, Theorem 15.11]); hence it is τ-compact. Then $\text{Mart}_t(K_t)$ is the image of a τ-compact set via the marginal map $\gamma \mapsto \gamma_t$ which is τ-$\sigma(\text{ca}(K_t), \mathcal{C}_b(K_t))$-continuous; hence it is $\sigma(\text{ca}(K_t), \mathcal{C}_b(K_t))$-compact.   □

We introduce the notion of a loss function that will be useful here and also in the sequel (see Sects. 4.4 and 4.4.1) to build penalisation functions.

Definition 4.8 A function $G : \mathbb{R} \to (-\infty, +\infty]$ is called a *loss function* if it is convex, nondecreasing, lower semicontinuous and satisfies $G(0) = 0$. We set $\text{dom}(G) := \{x \in \mathbb{R} : G(x) < \infty\}$. The conjugate function $G^* : \mathbb{R} \to (-\infty, +\infty]$ defined by $G^*(y) = \sup_{x \in \mathbb{R}}(xy - G(x))$ satisfies by the monotonicity of G that $G^*(y) = \infty$ for every $y < 0$.

Our requirements allow a wide range of penalisations. For example, we might use power-like penalisations, i.e., $G(x) = \frac{x^p}{p}$ for $x > 0$ and $p \in (1, \infty)$, $G(x) = 0$ for $x \leq 0$. In that case, we have $G^*(y) = \frac{y^q}{q}$ for every $y \geq 0$ for $\frac{1}{p} + \frac{1}{q} = 1$. Alternatively, we might take

$$G(x) = \begin{cases} 0 & \text{if } x \leq \varepsilon, \\ \infty & \text{otherwise,} \end{cases} \quad \text{so that } G^*(y) = \varepsilon y \text{ for } y \geq 0. \quad (4.7)$$

For $\gamma_t \in ca(K_t)$ we set

$$\mathcal{D}_t^G(\gamma_t) := \begin{cases} \sum_{n=1}^{N_t} G_{t,n}(|\int_{K_t} f_{t,n} \, d\gamma_t - c_{t,n}|) & \text{for } \gamma_t \in \text{Mart}_t(K_t), \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 4.9 *Assume that $G_{n,t} : \mathbb{R} \to (-\infty, +\infty]$ is a loss function for all $n = 0, \ldots, N_t$ and $t = 0, \ldots, T$, and that the martingale measure $\widehat{Q} \in \text{Mart}(\Omega)$ in the Standing Assumption 4.1 also satisfies $|\int_{K_t} f_{t,n} \, d\widehat{Q}_t - c_{t,n}| \in \text{dom}(G_{t,n})$. Then*

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_t^G(Q_t) \right) = \sup \left\{ \sum_{t=0}^{T} U_t^G(\varphi_t) : \varphi \in \mathcal{S}_{\text{sub}}(c) \right\}, \qquad (4.8)$$

where

$$U_t^G(\varphi_t) := \sup_{y_t \in \mathbb{R}^{N_t}} \left(\Pi^{\text{sub}}\left(\varphi_t + \sum_{n=1}^{N_t} y_{t,n}(f_{t,n} - c_{t,n}) \right) - \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right)$$

and Π^{sub} is given in (4.6). Finally, if the left-hand side of (4.8) is finite, a minimum point exists.

Proof 1) Set $g_{t,n} := f_{t,n} - c_{t,n}$. For any $t \in \{0, \ldots, T\}$, we prove that the functional $\mathcal{D}_t^G$ is $\sigma(ca(K_t), \mathcal{C}_b(K_t))$-lower semicontinuous and that for every $\varphi_t \in \mathcal{C}_b(K_t)$, its Fenchel–Moreau (convex) conjugate satisfies

$$V_t^G(\varphi_t) := \sup_{\gamma_t \in ca(K_t)} \left(\int_{K_t} \varphi_t \, d\gamma_t - \mathcal{D}_t^G(\gamma_t) \right)$$

$$= \inf_{y_t \in \mathbb{R}^{N_t}} \left(\Pi^{\text{sup}}\left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right),$$

and thus

$$U_t^G(\varphi_t) := -V_t^G(-\varphi_t) = \sup_{y_t \in \mathbb{R}^{N_t}} \left(\Pi^{\text{sub}}\left(\varphi_t + \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) - \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right). \qquad (4.9)$$

Here we use the definition of the superhedging price as

$$\Pi^{\text{sup}}(g) := -\Pi^{\text{sub}}(-g) = \sup_{Q \in \text{Mart}(\Omega)} E_Q[g]$$

by Corollary 4.6. We observe that $\mathcal{D}_t^G$ is $\sigma(ca(K_t), \mathcal{C}_b(K_t))$-lower semicontinuous as it is a sum of functions, each being a composition of a lower semicontinuous function and a continuous function on $\text{Mart}_t(K_t)$ which is $\sigma(ca(K_t), \mathcal{C}_b(K_t))$-compact by

Lemma 4.7. We now need to compute

$$V_t^G(\varphi_t) = \sup_{\gamma_t \in \mathrm{ca}(K_t)} \left(\int_{K_t} \varphi_t \, \mathrm{d}\gamma_t - \mathcal{D}_t^G(\gamma_t) \right)$$

$$= \sup_{Q_t \in \mathrm{Mart}_t(K_t)} \left(\int_{K_t} \varphi_t \, \mathrm{d}Q_t - \mathcal{D}_t^G(Q_t) \right).$$

Recall that $G_{t,n}(x) = \sup_{y \in \mathbb{R}}(xy - G_{t,n}^*(y))$ by the Fenchel–Moreau theorem. Hence

$$V_t^G(\varphi_t) = \sup_{Q_t \in \mathrm{Mart}_t(K_t)} \left(\int_{K_t} \varphi_t \, \mathrm{d}Q_t - \sum_{n=1}^{N_t} \sup_{y_{t,n} \in \mathbb{R}} \left(y_{t,n} \int_{K_t} g_{t,n} \, \mathrm{d}Q_t - G_{t,n}^*(y_{t,n}) \right) \right)$$

$$= \sup_{Q_t \in \mathrm{Mart}_t(K_t)} \inf_{y_t \in \mathrm{dom}} \left(\int_{K_t} \left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) \mathrm{d}Q_t + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right)$$

$$=: \sup_{Q_t \in \mathrm{Mart}(K_t)} \inf_{y_t \in \mathrm{dom}} \mathcal{T}(y_t, Q_t),$$

where $\mathrm{dom} = \mathrm{dom}(G_{t,1}^*) \times \cdots \times \mathrm{dom}(G_{t,N_t}^*) \subseteq \mathbb{R}^{N_t}$. We see that $\mathcal{T}$ is real-valued on $\mathrm{dom} \times \mathrm{Mart}_t(K_t)$, convex in the first variable and concave in the second. Moreover, $\{\mathcal{T}(y_t, \cdot) \geq C\}$ is $\sigma(\mathrm{Mart}_t(K_t), \mathcal{C}_b(K_t))$-closed in $\mathrm{Mart}_t(\Omega)$ for every $y_t \in \mathrm{dom}$, and $\mathrm{Mart}_t(K_t)$ is $\sigma(\mathrm{Mart}_t(K_t), \mathcal{C}_b(K_t))$-compact by Lemma 4.7. As a consequence, $\mathcal{T}(y_t, \cdot)$ is $\sigma(\mathrm{Mart}_t(K_t), \mathcal{C}_b(K_t))$-lower semicontinuous on $\mathrm{Mart}_t(K_t)$. We can apply Simons [36, Theorem 3.1], with $A = \mathrm{dom}$ and $B = \mathrm{Mart}_t(K_t)$ endowed with the topology $\sigma(\mathrm{Mart}_t(K_t), \mathcal{C}_b(K_t))$, and interchange inf and sup. From our previous computations, we then get

$$V_t^G(\varphi_t) = \sup_{Q_t \in \mathrm{Mart}_t(K_t)} \inf_{y_t \in \mathrm{dom}} \mathcal{T}(y_t, Q_t)$$

$$= \inf_{y_t \in \mathrm{dom}} \sup_{Q_t \in \mathrm{Mart}_t(K_t)} \mathcal{T}(y_t, Q_t)$$

$$= \inf_{y_t \in \mathrm{dom}} \left(\sup_{Q_t \in \mathrm{Mart}_t(K_t)} \int_{K_t} \left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) \mathrm{d}Q_t + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right)$$

$$= \inf_{y_t \in \mathrm{dom}} \left(\sup_{Q \in \mathrm{Mart}(\Omega)} \int_{\Omega} \left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) \mathrm{d}Q + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right)$$

$$= \inf_{y_t \in \mathrm{dom}} \left(\Pi^{\mathrm{sup}} \left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right)$$

$$= \inf_{y_t \in \mathbb{R}^{N_t}} \left(\Pi^{\mathrm{sup}} \left(\varphi_t - \sum_{n=1}^{N_t} y_{t,n} g_{t,n} \right) + \sum_{n=1}^{N_t} G_{t,n}^*(y_{t,n}) \right).$$

Equation (4.9) can be obtained with minor manipulations.

2) To conclude, we are clearly in the setup of Corollary 2.13 with $\mathcal{D}$ given as in Setup 3.5 from $\mathcal{D}_0^G, \ldots, \mathcal{D}_T^G$, and by definition $\text{dom}(\mathcal{D}_t^G) \subseteq \text{Prob}(K_t)$ for each $t = 0, \ldots, T$. Using Lemma 3.6, 2) together with the computations in 1) and the fact that $S^{U_t^G} = U_t^G$ by cash-additivity of U_t^G, we get the desired equality from (2.18) in Corollary 2.13; indeed, $G_{t,n}^*$ is bounded from below and proper by our assumptions on $G_{t,n}$, and Π^{sub} is real-valued and cash-additive on bounded continuous functions. This guarantees that $V_t^G(\varphi_t)$ is null for an appropriate choice of (constant) φ_t. The existence of optima follows again from Corollary 2.13. $\qquad\square$

Remark 4.10 Our assumption of existence of a particular $\widehat{Q} \in \text{Mart}(\Omega)$ in Proposition 4.9 expresses the fact that we are assuming that our market prices $c_{t,n}$ are close enough to those given by expectations under some martingale measure.

Example 4.11 Proposition 4.9 covers a wide range of penalisations. For example, we might impose a threshold for the error in computing option prices by taking into account only those martingale measures Q such that $|\int_\Omega f_{t,n} \, dQ_t - c_{t,n}| \leq \varepsilon_{t,n}$ for some $\varepsilon_{t,n} \geq 0$. To express this, just take $G_{t,n}$ in the form (4.7) for $\varepsilon = \varepsilon_{t,n}$.

Example 4.12 We now study the convergence of the penalised problem described above to the classical MOT problem. We suppose that our information on the marginal distributions increases, by increasing the number of prices available from the market. We take $f_{t,n}(x_t) = (x_t - \alpha_n)^+$ to be call options with maturity t and strikes α_n, $n \in \mathbb{N}$, that form a dense subset of $\mathbb{R}$.

We take as loss functions $G_{t,n}(x) = 0$ for $x \leq 0$ and $G_{t,n}(x) = \infty$ for all $x > 0$, $t = 0, \ldots, T$, $n \geq 1$. This means that on the left-hand side of (4.8), the infimum is taken only over martingale measures whose theoretical prices exactly match the ones for the data, namely $c_{t,n}$. For each $t = 0, \ldots, T$, $(c_{t,n})$ is a given sequence of prices, and we suppose that they are all computed under the same martingale measure $\widehat{Q} \in \text{Mart}(\Omega)$. We consider for each $k \in \mathbb{N}$ the initial segment $c_{t,1}, \ldots, c_{t,N_t(k)}$ for sequences $N_t(k) \uparrow \infty$, $t = 0, \ldots, T$. This means that for every $Q \in \text{Mart}(\Omega)$,

$$\mathcal{D}_k(Q) := \sum_{t=0}^{T} \sum_{n=1}^{N_t(k)} G_{t,n}\left(\left|\int_{K_t} f_{t,n} \, dQ_t - c_{t,n}\right|\right)$$

$$\leq \sum_{t=0}^{T} \sum_{n=1}^{N_t(k+1)} G_{t,n}\left(\left|\int_{K_t} f_{t,n} \, dQ_t - c_{t,n}\right|\right) = \mathcal{D}_{k+1}(Q)$$

and

$$\mathcal{D}_\infty(Q) = \sup_k \mathcal{D}_{k+1}(Q) = \sum_{t=0}^{T} \sum_{n=1}^{\infty} G_{t,n}\left(\left|\int_{K_t} f_{t,n} \, dQ_t - c_{t,n}\right|\right),$$

so that

$$\mathcal{D}_\infty(Q) = \begin{cases} 0 & \text{if } \int_{K_t} f_{t,n} \, dQ_t = c_{t,n} \text{ for all } 0 \leq t \leq T, n \geq 1, \\ \infty & \text{otherwise.} \end{cases}$$

From the denseness of (α_n), we conclude that $\mathcal{D}_\infty(Q) = 0$ if $Q_t = \widehat{Q}_t$ for $0 \leq t \leq T$, and $\mathcal{D}_\infty(Q) = \infty$ otherwise. As a consequence, by Proposition 2.26, we have the convergence

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \mathcal{D}_k(Q) \right) \longrightarrow \inf_{Q \in \text{Mart}(\widehat{Q}_0,\ldots,\widehat{Q}_T)} E_Q[c] \qquad \text{as } k \to \infty.$$

4.4 Penalty terms given via Wasserstein distance

Let d_t be a metric on K_t (equivalent to the Euclidean one). The (1-)Wasserstein distance induced by d_t is called $W_t : \text{Prob}(K_t) \times \text{Prob}(K_t) \to \mathbb{R}$. Let $\text{Lip}(1, K_t)$ be the class of d_t-Lipschitz functions on K_t with Lipschitz constant at most 1. Notice that $\text{Lip}(1, K_t) \subseteq \mathcal{C}_b(K_t)$ since d_t is equivalent to the Euclidean metric. For each t, let $G_t : \mathbb{R} \to (-\infty, +\infty]$ be a loss function as in Definition 4.8. For $\text{Mart}_t(K_t)$ as in Sect. 4.3, we introduce

$$\text{Prob} \ni Q_t(K_t) \mapsto \mathcal{D}_t^W(Q_t) := \begin{cases} G_t(W_t(Q_t, \widehat{Q}_t)) & \text{for } Q_t \in \text{Mart}_t(K_t), \\ \infty & \text{otherwise.} \end{cases} \quad (4.10)$$

Then $\mathcal{D}_t^W$ is lower semicontinuous with respect to the topology of weak convergence of probability measures, since the Wasserstein metric metrises the latter for compact underlying spaces and $\text{Mart}_t(K_t)$ is compact under that topology by Lemma 4.7. We are then in Setup 3.5 and the case 2) of Lemma 3.6. As in Sect. 4.3, we take $\mathcal{E} = \times_{t=0}^T \mathcal{C}_b(K_t)$.

Proposition 4.13 *For each $t = 0, \ldots, T$, suppose that G_t is a loss function, that there exists a $Q \in \text{Mart}(\Omega)$ with $G_t(W_t(Q_t, \widehat{Q}_t)) < \infty$, where $\widehat{Q} \in \text{Mart}(\Omega)$ is the martingale measure from the Standing Assumption 4.1, and take $\mathcal{D}_t^W$ as in (4.10). Then*

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^T \mathcal{D}_t^W(Q_t) \right) = \sup \left\{ \sum_{t=0}^T U_t^W(\varphi_t) : \varphi \in \mathcal{S}_{\text{sub}}(c) \right\}, \quad (4.11)$$

where

$$U_t^W(\varphi_t) := \sup_{\substack{y \geq 0, \\ \ell_t \in \text{Lip}(1, K_t)}} \left(\Pi^{\text{sub}}(\varphi_t + y\ell_t) - \int_{K_t} y\ell_t \, d\widehat{Q}_t - G_t^*(y) \right)$$

and Π^{sub} is given in (4.6). Finally, if the left-hand side of (4.11) is finite, a minimum point exists.

Proof Starting from $\mathcal{D}_t^W$, we compute the associated V_t^W as

$$
\begin{aligned}
V_t^W(\varphi_t) &:= \sup_{\gamma \in \mathrm{ca}(K_t)} \left(\int_{K_t} \varphi_t \mathrm{d}\gamma - \mathcal{D}_t^W(\gamma) \right) \\
&= \sup_{Q \in \mathrm{Mart}_t(K_t)} \left(\int_{K_t} \varphi_t \mathrm{d}Q - G_t\big(W_t(Q, \widehat{Q}_t)\big) \right) \\
&= \sup_{Q \in \mathrm{Mart}_t(K_t)} \left(\int_{K_t} \varphi_t \mathrm{d}Q - \sup_{y \geq 0} \big(y W_t(Q, \widehat{Q}_t) - G_t^*(y)\big) \right) \\
&= \sup_{Q \in \mathrm{Mart}_t(K_t)} \inf_{y \geq 0} \left(\int_{K_t} \varphi_t \mathrm{d}Q - y W_t(Q, \widehat{Q}_t) + G_t^*(y) \right) \\
&= \sup_{Q \in \mathrm{Mart}_t(K_t)} \inf_{\substack{y \in \mathrm{dom}(G_t^*), \\ \ell_t \in \mathrm{Lip}(1, K_t)}} \left(\int_{K_t} (\varphi_t - y\ell_t) \mathrm{d}Q + \int_{K_t} y\ell_t \mathrm{d}\widehat{Q}_t + G_t^*(y) \right) \\
&= \inf_{\substack{y \in \mathrm{dom}(G_t^*), \\ \ell_t \in \mathrm{Lip}(1, K_t)}} \left(\sup_{Q \in \mathrm{Mart}_t(K_t)} \left(\int_{K_t} (\varphi_t - y\ell_t) \mathrm{d}Q \right) + \int_{K_t} y\ell_t \mathrm{d}\widehat{Q}_t + G_t^*(y) \right) \\
&= \inf_{\substack{y \in \mathrm{dom}(G_t^*), \\ \ell_t \in \mathrm{Lip}(1, K_t)}} \left(\Pi^{\sup}(\varphi_t - y\ell_t) + \int_{K_t} y\ell_t \mathrm{d}\widehat{Q}_t + G_t^*(y) \right) \\
&= \inf_{\substack{y \geq 0, \\ \ell_t \in \mathrm{Lip}(1, K_t)}} \left(\Pi^{\sup}(\varphi_t - y\ell_t) + \alpha(y, \ell_t) \right)
\end{aligned}
$$

for the penalty $\alpha(y, \ell_t) := \int_{K_t} y\ell_t \mathrm{d}\widehat{Q}_t + G_t^*(y)$. In the equality chain above, we use the following: in the third equality, the dual representation of G_t; in the fifth, the definition of $\mathrm{dom}(G_t^*)$ and the classical Kantorovich–Rubinstein duality (see Villani [38, Remark 6.5]); in the sixth, Simons [36, Theorem 3.1] (observe that $\mathrm{Mart}_t(K_t)$ is compact by Lemma 4.7); in the seventh, the definition of the superhedging price $\Pi^{\sup}(g) := -\Pi^{\mathrm{sub}}(-g) = \sup_{Q \in \mathrm{Mart}(\Omega)} E_Q[g]$ by Corollary 4.6. Once we have V_t^W, we have U_t^W, and then we can argue as in step 2) of the proof of Proposition 4.9, also regarding existence of an optimum. $\qquad\square$

Remark 4.14 If U_t^W (as well as U_t^G in Proposition 4.9) is real-valued on $\mathcal{C}_b(K_t)$, one might take $\mathcal{E}_t$ as the set of functions of the form (4.1) in place of $\mathcal{E}_t = \mathcal{C}_b(K_t)$ in both Proposition 4.9 and 4.13, using norm-density of the piecewise linear functions just as in the proof of Corollary 4.3.

Remark 4.15 The reader can check that the property $\widehat{Q} \in \mathrm{Mart}(\Omega)$ is not used in the proof, and that it would suffice to have only $\widehat{Q} \in \mathrm{Prob}(\Omega)$. This will be exploited in Sect. 4.4.1.

Example 4.16 Taking $G_t(x) = 0$ if $x \leq \varepsilon_t$ and $G_t(x) = \infty$ otherwise, we get $G_t^*(y) = \varepsilon_t y$ if $y \geq 0$ and $G_t^*(y) = \infty$ otherwise. In this case,

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_t^W(Q) \right)$$

$$= \inf\{E_Q[c] : Q \in \mathrm{Mart}(\Omega) \text{ and } W_t(Q_t, \widehat{Q}_t) \leq \varepsilon_t, t = 0, \ldots, T\}. \qquad (4.12)$$

One can verify with the same techniques as in Example 4.12 that we have convergence, as $\varepsilon_t \downarrow 0$ for every $t = 0, \ldots, T$, of the values on the right-hand side of (4.12) to the MOT value on the left-hand side of (4.5).

4.4.1 Convergence with Wasserstein-induced penalisation

As already mentioned, in the classical MOT framework, the marginals $\widehat{Q}_0, \ldots, \widehat{Q}_T$ need to be determined, potentially from the prices of many vanilla options. It is then reasonable to suppose that in a real-world situation, one proceeds by approximation, that is, one determines sequences of candidates $(\widehat{Q}_t^n) \subseteq \mathrm{Prob}(K_t)$ for $t = 0, \ldots, T$. If such an approximation scheme (whose details are beyond the scope of this paper) is working, one should have a convergence of these sequences to the true marginals. One suitable candidate for the type of convergence is the weak one, namely one might want to have $\widehat{Q}_t^n \to \widehat{Q}_t^\infty := \widehat{Q}_t$ as $n \to \infty$ for $t = 0, \ldots, T$ in the weak sense for probability measures. We suppose here that $K_0, \ldots, K_T$ are compact sets, and so weak convergence is equivalent to convergence in the Wasserstein distance. Proposition 4.17 below shows how the EMOT values treated in Proposition 4.13 and associated to the approximating measures $\widehat{Q}_t^n, t = 0, \ldots, T$, converge to the original MOT value for the true marginals $\widehat{Q}_0, \ldots, \widehat{Q}_T$, provided that the loss functions G_t^n converge appropriately.

For the next result, it is convenient to rename the martingale measure $\widehat{Q}$ from the Standing Assumption 4.1 as $\widehat{Q}^\infty$, so that $\widehat{Q}^\infty \in \mathrm{Mart}(\Omega)$ with marginals $\widehat{Q}_t^\infty$ and $c \in L^1(\widehat{Q}^\infty)$.

Proposition 4.17 *For each $n \in \mathbb{N} \cup \{\infty\}$ and $t = 0, \ldots, T$, let G_t^n be a loss functions with $G_t^n(x) \uparrow G_t^\infty(x)$ as $n \to \infty$ for every $x \in \mathbb{R}$ and $G_t^\infty(x) = \infty$ for every $x > 0$. For every $t = 0, \ldots, T$ and $n \in \mathbb{N}$, we assume that $\widehat{Q}_t^n \in \mathrm{Prob}(K_t)$, $\lim_n W_t(\widehat{Q}_t^\infty, \widehat{Q}_t^n) = 0$ and $\lim_n G_t^n(W_t(\widehat{Q}_t^\infty, \widehat{Q}_t^n)) = 0$. Then*

$$\lim_n \pi_n^W(c) = \pi_\infty^W(c) = \inf_{Q \in \mathrm{Mart}(\widehat{Q}_0^\infty, \ldots, \widehat{Q}_T^\infty)} E_Q[c], \qquad (4.13)$$

where

$$\pi_n^W(c) = \inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} G_t^n\big(W_t(Q_t, \widehat{Q}_t^n)\big) \right), \qquad n \in \mathbb{N} \cup \{\infty\}.$$

Proof The second equality in (4.13) is just a consequence of the definition of π_∞^W and of $G_t^\infty(x) = \infty$ for every $x > 0$. Since we may always pass to a subsequence

of $(\widehat{Q}^n)$, we may assume without loss of generality that $G_t^n(W_t(\widehat{Q}_t^\infty, \widehat{Q}_t^n)) < \infty$ for all $n \in \mathbb{N} \cup \{\infty\}$ and all t (the case $n = \infty$ is obvious from $G_t^\infty(0) = 0$). We first claim that $\pi_n^W(c)$ is finite for all $n \in \mathbb{N} \cup \{\infty\}$. Indeed, since Ω is compact, c is lower semicontinuous and G_t^n are nonnegative on $[0, \infty)$, we have for $n \in \mathbb{N} \cup \{\infty\}$ that

$$-\infty < \inf_{x \in \Omega} c(x) \le \inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^T G_t^n \big(W_t(Q_t, \widehat{Q}_t^n)\big) \right)$$

$$\le E_{\widehat{Q}^\infty}[c] + \sum_{t=0}^T G_t^n \big(W_t(\widehat{Q}_t^\infty, \widehat{Q}_t^n)\big) < \infty.$$

We now prove that $\pi_\infty^W(c) \ge \limsup_n \pi_n^W(c)$. Since $\pi_\infty^W(c) < \infty$, there exists an optimum $Q^\infty \in \mathrm{Mart}(\Omega)$ for $\pi_\infty^W(c)$, and its marginals satisfy $Q_t^\infty = \widehat{Q}_t^\infty$, $t = 0, \ldots, T$. Then $G_t^n(W_t(Q_t^\infty, \widehat{Q}_t^n)) = G_t^n(W_t(\widehat{Q}_t^\infty, \widehat{Q}_t^n)) \to 0$ as $n \to \infty$ and

$$\pi_\infty^W(c) = E_{Q^\infty}[c]$$

$$= \lim_n \left(E_{Q^\infty}[c] + \sum_{t=0}^T G_t^n \big(W_t(Q_t^\infty, \widehat{Q}_t^n)\big) \right)$$

$$\ge \limsup_n \inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^T G_t^n \big(W_t(Q_t, \widehat{Q}_t^n)\big) \right) = \limsup_n \pi_n^W(c).$$

It only remains to show that $\liminf_n \pi_n^W(c) \ge \pi_\infty^W(c)$. Proposition 4.13 and Remark 4.15 guarantee for each $n \in \mathbb{N}$ the existence of an optimum $Q^n \in \mathrm{Mart}(\Omega)$ for the value $\pi_n^W(c) < \infty$. From the proof of Lemma 4.7, we know that $\mathrm{Mart}(\Omega)$ is weakly compact, and so we can take a subsequence of (Q^n) such that for some $\widetilde{Q} \in \mathrm{Mart}(\Omega)$, $W_t(Q_t^{n_k}, \widetilde{Q}_t) \to 0$ as $k \to \infty$ for every t and

$$\lim_k \left(E_{Q^{n_k}}[c] + \sum_{t=0}^T G_t^{n_k} \big(W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k})\big) \right) = \liminf_n \pi_n^W(c). \tag{4.14}$$

Let $N \in \mathbb{N}$ and recall that $\sup_N G_t^N(x) = G_t^\infty(x)$. Then we compute by using the particular form of G_t^∞ that

$$\pi_\infty^W(c) = \inf_{Q \in \mathrm{Mart}(\widehat{Q}_0^\infty, \ldots, \widehat{Q}_T^\infty)} E_Q[c]$$

$$= \inf_{Q \in \mathrm{Mart}(\Omega)} \sup_{N \in \mathbb{N}} \left(E_Q[c] + \sum_{t=0}^T G_t^N \big(W_t(Q_t, \widehat{Q}_t^\infty)\big) \right)$$

$$\le \sup_{N \in \mathbb{N}} \left(E_{\widetilde{Q}}[c] + \sum_{t=0}^T G_t^N \big(W_t(\widetilde{Q}_t, \widehat{Q}_t^\infty)\big) \right)$$

$$\le \liminf_n \pi_n^W(c), \tag{4.15}$$

where the first inequality uses $\widetilde{Q} \in \mathrm{Mart}(\Omega)$ and the second is justified as follows. From the lower semicontinuity with respect to the weak convergence of $Q \mapsto \int_\Omega c \, dQ$, the lower semicontinuity of G_t^N and $W_t(\widetilde{Q}_t, \widehat{Q}_t^\infty) = \lim_k W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k})$ for all t, we obtain

$$E_{\widetilde{Q}}[c] + \sum_{t=0}^{T} G_t^N \big(W_t(\widetilde{Q}_t, \widehat{Q}_t^\infty) \big)$$

$$\leq \liminf_k \left(E_{Q^{n_k}}[c] + \sum_{t=0}^{T} G_t^N \big(W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k}) \big) \right). \tag{4.16}$$

Moreover, G_t^n is increasing in n for each t, and so for all $n_k > N$,

$$E_{Q^{n_k}}[c] + \sum_{t=0}^{T} G_t^N \big(W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k}) \big) \leq E_{Q^{n_k}}[c] + \sum_{t=0}^{T} G_t^{n_k} \big(W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k}) \big).$$

This and (4.16) imply

$$\left(E_{\widetilde{Q}}[c] + \sum_{t=0}^{T} G_t^N \big(W_t(\widetilde{Q}_t, \widehat{Q}_t^\infty) \big) \right)$$

$$\leq \liminf_k \left(E_{Q^{n_k}}[c] + \sum_{t=0}^{T} G_t^{n_k} \big(W_t(Q_t^{n_k}, \widehat{Q}_t^{n_k}) \big) \right) = \liminf_n \pi_n^W(c),$$

by (4.14). Taking the supremum over $N \in \mathbb{N}$, we obtain (4.15). $\qquad\square$

5 Applications in the noncompact case

We stress that the extension of the results in Sects. 4.3 and 4.4 to the noncompact case seems to be nontrivial. The main issues come from the verification of Assumption 2.3 (ii), when one starts the analysis from penalisation terms rather than from valuation functionals. Not excluding that such an extension is possible, we leave this topic for future research. However, the case of valuations induced by utility functions can be treated also in the noncompact case, as we describe below. In the noncompact case, Corollary 4.3 takes the following form.

Corollary 5.1 *Take $d = 1$, $K_0 = \{x_0\}$ for some $x_0 \in \mathbb{R}$ and let $K_1, \ldots, K_T \subseteq \mathbb{R}$ be closed subsets of $\mathbb{R}$. Consider utility functions $u_0, \ldots, u_T$ satisfying Assumption 3.7, and suppose $\mathrm{dom}(u_0) = \cdots = \mathrm{dom}(u_T) = \mathbb{R}$. Take for each $t = 0, \ldots, T$ the vector space $\mathcal{E}_t \subseteq C_t = C_{t:t}$ (see (2.1)) of functions of the form (4.1), let $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ and fix a $\widehat{Q} \in \mathrm{Mart}(\Omega)$ such that*

$$\int_{K_t} v_t \big(a(1 + |x_t|) \big) \, d\widehat{Q}_t(x_t) < \infty, \qquad \forall a > 0, t = 0, \ldots, T.$$

Suppose that $c : \Omega \to (-\infty, +\infty]$ is lower semicontinuous and satisfies (2.8). *Then*

$$\inf_{Q \in \text{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) \right) = \sup \left\{ \sum_{t=0}^{T} U_{\widehat{Q}_t}(\varphi_t) : \varphi \in \mathcal{S}_{\text{sub}}(c) \right\}, \quad (5.1)$$

where $U_{\widehat{Q}_t}(\varphi_t)$ is defined in (4.2) *for general $\varphi_t \in C_t$ and $\mathcal{D}_{v_t^*, \widehat{Q}_t}$ is given in* (3.5). *Moreover, the infimum in* (5.1) *is a minimum provided that the right-hand side of* (5.1) *is finite.*

Proof All the claims follow from Theorem 3.4 if we show that all its hypotheses are satisfied. To do so, we check the following properties: (i) the functional $U_t(\varphi_t) = U_{\widehat{Q}_t}(\varphi_t)$ is real-valued on C_t, concave, nondecreasing and cash-additive; (ii) $\mathcal{D}_t(Q_t) = \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t)$ for $Q \in \text{Mart}(\Omega)$; (iii) $U_t(0) = 0$ for $t = 0, \dots, T$ and the conditions (2.5) and (2.6) hold, which we do using Example 2.6.

To check (i), observe that for every $t = 0, \dots, T$ and $\varphi_t \in C_t$,

$$-\infty < -\int_{K_t} v_t \big(\|\varphi_t\|_t (1 + |x_t|) \big) \, \mathrm{d}\widehat{Q}_t(x_t)$$

$$= \int_{K_t} u_t \big(-\|\varphi_t\|_t (1 + |x_t|) \big) \, \mathrm{d}\widehat{Q}_t(x_t) \leq \int_{K_t} u_t \big(\varphi_t(x_t) \big) \, \mathrm{d}\widehat{Q}_t(x_t) \leq U_{\widehat{Q}_t}(\varphi_t)$$

$$\leq \int_{K_t} \varphi_t(x_t) \mathrm{d}\widehat{Q}_t(x_t) \leq \int_{K_t} \|\varphi_t\|_t (1 + |x_t|) \, \mathrm{d}\widehat{Q}_t(x_t) < \infty,$$

where the first inequality in the last line uses the fact that $u_t(x) \leq x$, $\forall x \in \mathbb{R}$, and the finiteness of the last term comes from $\widehat{Q} \in \text{Prob}^1(\Omega)$. Concavity, monotonicity and cash-additivity can be checked by direct computation. Coming to (ii), from Proposition 3.9, for every $Q \in \text{Prob}^1(\Omega)$ and $t = 0, \dots, T$, we have

$$\mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t) = \sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} \varphi_t(x_t) \, \mathrm{d}Q_t(x_t) - \int_{K_t} v_t \big(\varphi_t(x_t) \big) \, \mathrm{d}\widehat{Q}_t(x_t) \right)$$

$$\leq \sup_{\varphi_t \in C_t} \left(\int_{K_t} \varphi_t(x_t) \, \mathrm{d}Q_t(x_t) - \int_{K_t} v_t \big(\varphi_t(x_t) \big) \, \mathrm{d}\widehat{Q}_t(x_t) \right)$$

$$\leq \mathcal{D}_{v_t^*, \widehat{Q}_t}(Q_t),$$

where the first equality exploits (3.6) and the last inequality is from the Fenchel inequality $v_t^*(y) \geq (\varphi_t y - v_t(\varphi_t))$. To conclude the proof of (ii), we have to show that the sup in the above expression can be taken over $\mathcal{E}_t$, as in the penalty term $\mathcal{D}_t$ in Theorem 3.4. Observe that for every $\varphi_t \in C_t$, there exists a sequence $(\varphi_t^n) \subseteq C_t$, with each $\varphi_t^n \in \mathcal{E}_t$ of the form (4.1), such that $\varphi_t^n \to \varphi_t$ pointwise on K_t and $\sup_n \|\varphi_t^n\|_t < \infty$. This implies

$$\int_{K_t} \varphi_t(x_t) \, \mathrm{d}Q(x_t) - \int_{K_t} v_t \big(\varphi_t(x_t) \big) \, \mathrm{d}\widehat{Q}_t(x_t)$$

$$= \lim_n \left(\int_{K_t} \varphi_t^n(x_t) \, \mathrm{d}Q(x_t) - \int_{K_t} v_t \big(\varphi_t^n(x_t) \big) \, \mathrm{d}\widehat{Q}_t(x_t) \right)$$

for all $Q \in \text{Prob}^1(\Omega)$ by dominated convergence and using the assumption that

$$\int_{K_t} v_t\big(a(1+|x_t|)\big)\,\mathrm{d}\widehat{Q}_t(x_t) < \infty, \qquad \forall a > 0.$$

Finally, we work on (iii). Take $f_t^{\alpha_n}(x_t) = (|x_t| - \alpha_n)^+$ as in (2.11) and $\alpha_n \uparrow \infty$. Observe that $f_t^{\alpha_n}(x_t) \to 0$ as $n \to \infty$ and that the assumption $u_t(x) \le x$ for all $x \in \mathbb{R}$ implies $U_t(0) \le 0$. Then for every $a > 0$,

$$0 \ge U_t(0) \ge U_t(-af_t^{\alpha_n}) \ge \int_{K_t} u_t(-af_t^{\alpha_n})\mathrm{d}\widehat{Q}_t \longrightarrow 0 \qquad \text{as } n \to \infty, \qquad (5.2)$$

where the last limit is by dominated convergence. Indeed, $f_t^{\alpha_n}(x_t) \le 1 + |x_t|$ for every $x_t \in K_t$ so that

$$\big|u_t\big(-af_t^{\alpha_n}(x_t)\big)\big| = -u_t\big(-af_t^{\alpha_n}(x_t)\big)$$
$$\le -u_t\big(-a(1+|x_t|)\big) = v_t\big(a(1+|x_t|)\big) \in L^1(\widehat{Q}_t)$$

for every $x_t \in K_t, a > 0$. Now (5.2) yields simultaneously that $U_t(0) = 0$ and $U_t(-af_t^{\alpha_n}) \to 0$ as $n \to \infty$ for all $t = 0, \ldots, T, a > 0$. To apply Example 2.6, it is then enough to observe that taking $U(\varphi) := \sum_{t=0}^T U_t(\varphi_t)$ as in Theorem 3.4, we have $U(0) = 0$ and $U(0, \ldots, 0, -af_t^{\overset{n}{\beta}}, 0, \ldots, 0) = U_t(-af_t^{\overset{n}{\beta}}) \to 0$ as $n \to \infty$ for all $a > 0$, which is (2.12). $\qquad\qquad\square$

Remark 5.2 Observe that Corollary 5.1 remains valid for general $\widehat{Q}_t \in \text{Prob}^1(K_t)$ without requesting these are marginals of a martingale measure. Indeed, we did not use the martingale property at any point in the above proof.

Just as we obtained Corollary 4.5 from Corollary 4.3 by using the linear utility functions $u_t(x_t) = x_t$, we now deduce the following result from Corollary 5.1; see Beiglböck et al. [4, Theorem 1.1 and Corollary 1.2].

Corollary 5.3 *Take $d = 1$, $K_0 = \{x_0\}$ for some $x_0 \in \mathbb{R}$ and let $K_1, \ldots, K_T \subseteq \mathbb{R}$ be closed subsets of $\mathbb{R}$. Take for each $t = 0, \ldots, T$ the vector space $\mathcal{E}_t \subseteq C_t$ of functions of the form (4.1), let $\mathcal{E} = \mathcal{E}_0 \times \cdots \times \mathcal{E}_T$ and fix a $\widehat{Q} \in \text{Mart}(\Omega)$. Then for any $c : \Omega \to (-\infty, +\infty]$ which is lower semicontinuous and satisfies (2.8), we have*

$$\inf_{Q \in \text{Mart}(\widehat{Q}_0, \ldots, \widehat{Q}_T)} E_Q[c] = \sup\left\{\sum_{t=0}^T E_{\widehat{Q}_t}[\varphi_t] : \varphi \in \mathcal{S}_{\text{sub}}(c)\right\} = \pi(c), \qquad (5.3)$$

and if $\pi(c) < \infty$, a minimum point exists for the infimum in (5.3).

Example 5.4 We now study the convergence to the MOT problem. Take functions $u_0, \ldots, u_T : \mathbb{R} \to \mathbb{R}$ satisfying Assumption 3.7, and assume additionally that these

are all differentiable in 0 (which implies that $\{1\} = \partial u_0(0) = \cdots = \partial u_T(0)$). Observe that if we set $u_t^n(x) := nu_t(\frac{x}{n}1)$ for $x \in \mathbb{R}, t = 0, \ldots, T$, the functions $u_0^n, \ldots, u_T^n$ still satisfy Assumption 3.7. Moreover, $(v_t^n)^*(y) = \sup_{x \in \mathbb{R}}(u_t^n(x) - xy)) = nv_t^*(y)$, $y \in \mathbb{R}$. Since $u_t(0) = 0$, we have $v_t^* \geq 0$ and as a consequence $\sup_n(v_t^n)^*(y) = 0$ if $v_t^*(y) = 0$ and $\sup_n(v_t^n)^*(y) = \infty$ otherwise. Moreover, $v_t^*(y) = 0$ implies that we have $y \in \partial u_t(0) = \{1\}$. Consider the set $\mathcal{A}^\varepsilon$ of ε-martingale measures defined in (2.14), take $\widehat{Q} \in \mathrm{Mart}(\Omega)$ and a sequence $\varepsilon_n \downarrow 0$. Using (2.15) gives for every $Q \in \mathrm{Prob}^1(\Omega)$ that

$$\sum_{t=0}^{T} \mathcal{D}_{(v_t^n)^*, \widehat{Q}_t}(Q_t) + \sigma_{\mathcal{A}^{\varepsilon_n}}(Q_t) \uparrow \mathcal{D}_\infty(Q) + \sigma_{\mathcal{A}_\infty}(Q) \qquad \text{as } n \to \infty,$$

where

$$\mathcal{D}_\infty(Q) + \sigma_{\mathcal{A}_\infty}(Q) = \begin{cases} 0 & \text{if } Q \in \mathrm{Mart}(\Omega) \text{ and } Q_t = \widehat{Q}_t \text{ for } t = 0, \ldots, T, \\ \infty & \text{otherwise.} \end{cases}$$

As a consequence, by Proposition 2.26,

$$\inf_{Q \in \mathrm{Mart}(\Omega)} \left(E_Q[c] + \sum_{t=0}^{T} \mathcal{D}_{(v_t^n)^*, \widehat{Q}_t}(Q_t) \right) \longrightarrow \inf_{Q \in \mathrm{Mart}(\widehat{Q}_0, \ldots, \widehat{Q}_T)} E_Q[c] \qquad \text{as } n \to \infty.$$

Appendix

Proposition A.1 (i) *There exist constants $\gamma = \gamma(d, T) > 0, \beta = \beta(d, T) > 0$ such that for every $A > 1$, we have*

$$1 + \sum_{s=0}^{T} \sum_{j=1}^{d} |x_s^j| \leq \gamma \sum_{s=0}^{T} \sum_{j=1}^{d} f_{j,s}^{\frac{A}{\beta}}(x_s^j), \qquad \forall (x_0, \ldots, x_T) \in (\mathbb{R}^d \setminus [-A, A]^d)^{T+1},$$

where $f_{j,s}^\alpha$ is defined in (2.11).
 (ii) *The conditions in Example 2.6 are sufficient for (2.5) and (2.6) to hold.*

Proof (i) Note that $f_{j,s}^{\frac{A}{\beta}}(x_s^j) = (|x_s^j| - \frac{A}{\beta})^+$. Fix $x \in (\mathbb{R}^d \setminus [-A, A]^d)^{T+1}$. Define

$$I(x) := \{(j, s) \in \{1, \ldots, d\} \times \{0, \ldots, T\} : |x_s^j| > A\},$$

$$I^c(x) := \{(j, s) \in \{1, \ldots, d\} \times \{0, \ldots, T\} : |x_s^j| \leq A\}.$$

Then $\{1, \ldots, d\} \times \{0, \ldots, T\} = I(x) \cup I^c(x)$ and $I(x) \neq \emptyset$. Moreover, for $\beta > 1$,

$$\gamma \sum_{s=0}^{T} \sum_{j=1}^{d} f_{j,s}^{\frac{A}{\beta}}(x_s^j) - \left(1 + \sum_{s=0}^{T} \sum_{j=1}^{d} |x_s^j|\right)$$

$$= \gamma \sum_{s=0}^{T} \sum_{j=1}^{d} \left(|x_s^j| - \frac{A}{\beta}\right)^+ - \left(1 + \sum_{s=0}^{T} \sum_{j=1}^{d} |x_s^j|\right)$$

$$= \sum_{(j,s) \in I(x)} \left(\gamma\left(|x_s^j| - \frac{A}{\beta}\right) - |x_s^j|\right) + \sum_{(j,s) \in I^c(x)} \left(\gamma\left(|x_s^j| - \frac{A}{\beta}\right)^+ - |x_s^j|\right) - 1$$

$$\geq (\gamma - 1) \sum_{(j,s) \in I(x)} |x_s^j| - \frac{\gamma A}{\beta} |I(x)| - \sum_{(j,s) \in I^c(x)} |x_s^j| - 1$$

$$\geq (\gamma - 1)A - \frac{\gamma A}{\beta} d(T + 1) - d(T + 1)A - 1$$

$$= \gamma A\left(1 - d\frac{T+1}{\beta}\right) - (A + d(T+1)A + 1).$$

Choosing e.g. $\beta = 2d(T + 1)$, we can solve $\gamma A\frac{1}{2} - (A + d(T + 1)A + 1) \geq 0$ to get $\gamma \geq 2d(T + 2) + \frac{2}{A}$. This yields a possible selection $\gamma = 2d(T + 2) + 2$ as well which depends only on the dimensions d, T.

(ii) Observe first that for any concave function $F : \mathbb{R}^{T+1} \to \mathbb{R}$, we have

$$F\left(\sum_{t=0}^{T} \sum_{j=1}^{d} y_{j,t} e^{(t)}\right) \geq \frac{1}{d(T+1)} \sum_{t=0}^{T} \sum_{j=1}^{d} F\left(d(T+1)y_{j,t}e^{(t)}\right) \qquad \text{for all } y_{j,t} \in \mathbb{R},$$

where $\{e^{(t)} : t = 0, \ldots, T\}$ is the canonical basis in $\mathbb{R}^{T+1}$. Now we define the set $H_t(n) := \times_{j=1}^{d} K_t^j \cap [-n, n]^d$ and set $f_t^n := \gamma \sum_{j=1}^{d} f_{j,t}^{\frac{n}{\beta}}$ for γ, β given in (i) above. Then (i) guarantees that (2.5) holds. The concavity and monotonicity of U and (2.12) imply that for any $a \geq 0$,

$$0 = U(0) \geq U(-af_0^n, \ldots, -af_T^n)$$

$$\geq \frac{1}{d(T+1)} \sum_{t=0}^{T} \sum_{j=1}^{d} U\left(-d(T+1)a\gamma f_{j,t}^{\frac{n}{\beta}} e^{(t)}\right) \longrightarrow 0 \qquad \text{as } n \to \infty,$$

which is (2.6). □

A.1 Weighted spaces

We use the notation introduced in Sect. 2.4.1. Consider the following spaces of continuous functions. For $\psi \in \mathcal{C}(\mathbb{X})$, we set

$$C_\psi := \left\{\phi \in \mathcal{C}(\mathbb{X}) : \|\phi\|_\psi := \sup_{x \in \mathbb{X}} \frac{|\phi(x)|}{1 + |\psi(x)|} < \infty\right\}.$$

As can be easily verified by following the classical case of bounded continuous functions with the sup-norm, C_ψ is a Banach lattice under the norm $\|\cdot\|_\psi$. Notice also that $\mathcal{C}_b(\mathbb{X}) \ni \varphi \mapsto (1 + |\psi|)\varphi \in C_\psi$ defines an isomorphism between Banach spaces. The topological dual of C_ψ is denoted by $(C_\psi)^*$.

Proposition A.2 *Let $L \in (C_\psi)^*$ be continuous, linear and positive. Suppose that for every $\varepsilon > 0$, there exists a compact $K_\varepsilon \subseteq \mathbb{X}$ such that*

$$\varphi \in C_\psi \text{ with } \varphi|_{K_\varepsilon} = 0 \text{ implies that } |\langle \varphi, L \rangle| \leq \varepsilon \|\varphi\|_\psi.$$

Then for every sequence $(c_n) \subseteq C_\psi$ with $c_n \downarrow 0$ pointwise on $\mathbb{X}$, we have $L(c_n) \downarrow 0$.

Proof Fix $\varepsilon > 0$ and take the associated compact K_ε. By Dini's lemma, we have $\sup_{x \in K_\varepsilon} c_n(x) \downarrow 0$. Take n big enough such that we have $\sup_{x \in K_\varepsilon} c_n(x) < \varepsilon$. Define $0 \leq g_n^\varepsilon := \min(c_n, \varepsilon)$. Then clearly $g_n^\varepsilon(x) = |g_n^\varepsilon(x)| \leq \varepsilon \leq \varepsilon(1 + |\psi(x)|)$ for all $x \in \mathbb{X}$ implies that $\|g_n^\varepsilon\|_\psi \leq \varepsilon$ and therefore

$$|\langle g_n^\varepsilon, L \rangle| \leq \|L\| \|g_n^\varepsilon\|_\psi \leq \|L\|\varepsilon, \tag{A.1}$$

where $\|L\|$ is the operator norm (note that $\|L\| < \infty$ since L is continuous). Also, since $\sup_{x \in K_\varepsilon} c_n(x) < \varepsilon$, we get that c_n and g_n^ε coincide on K_ε so that $(c_n - g_n^\varepsilon)|_{K_\varepsilon} = 0$. By using the hypothesis on L, we then get

$$|\langle c_n - g_n^\varepsilon, L \rangle| \leq \varepsilon \|c_n - g_n^\varepsilon\|_\psi \leq \varepsilon(\|c_n\|_\psi + \|g_n^\varepsilon\|_\psi) \leq \varepsilon(\|c_1\|_\psi + \varepsilon), \tag{A.2}$$

where the last step uses the Banach lattice property of $\|\cdot\|_\psi$ and that $\|g_n^\varepsilon\|_\psi \leq \varepsilon$, as shown before. We now combine (A.1) and (A.2) to get

$$0 \leq \langle c_n, L \rangle = \langle c_n - g_n^\varepsilon, L \rangle + \langle g_n^\varepsilon, L \rangle \leq |\langle c_n - g_n^\varepsilon, L \rangle| + \langle g_n^\varepsilon, L \rangle$$
$$\leq \varepsilon(\|c_1\|_\psi + \varepsilon) + \|L\|\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $L(c_n) \downarrow 0$. $\qquad\square$

We state below the celebrated Daniell–Stone theorem.

Theorem A.3 *Let V be a vector lattice of functions (i.e., $f, g \in V$ implies that $\max(f, g) \in V$) on a set $\mathbb{X}$ such that $1 \in V$. Let L be a linear functional on V with the properties that $L(f) \geq 0$ whenever $f \geq 0$, $L(1) = 1$, and $L(f_n) \to 0$ for every sequence (f_n) of functions in V monotonically decreasing to zero. Then there exists a unique probability measure μ on the σ-algebra $\mathcal{F} = \sigma(V)$ generated by V such that $V \subseteq L^1(\mu)$ and*

$$L(f) = \int_{\mathbb{X}} f \, d\mu, \qquad \forall f \in V.$$

Proof See Bogachev [10, Theorem 7.8.1.] $\qquad\square$

A.2 Proofs

Proof of Proposition 3.9 We use Liero et al. [30, Theorem 2.7 and Remark 2.8]. To do so, let us rename $F := v_t^*$ (see (3.4) for the definition of v^*), which implies that $F^\circ(y) := -F^*(-y)$ of [30, Equation (2.45)] satisfies

$$F^\circ(y) = -F^*(-y) = -v_t^{**}(-y) = -v_t(-y) = u_t(y)$$

by the Fenchel–Moreau theorem. All the assumptions of [30, Sect. 2.3] on F are satisfied since $F(y) \geq u_t(0) - 0y = 0$ for $y \geq 0$ and $F(1) = \sup_{x \in \mathbb{R}}(u_t(x) - x) \leq 0$ (recall that $u_t(x) \leq x$ for all $x \in \mathbb{R}$). Also, we have $\lim_{y \to \infty} \frac{F(y)}{y} = F'_\infty = \infty$ since $\mathrm{dom}(u_t) = \mathbb{R}$. We can then apply [30, Theorem 2.7 and Remark 2.8] to obtain (3.6). We stress the fact that since u_t is finite-valued on all of $\mathbb{R}$, it is continuous there and for every $\varphi_t \in \mathcal{C}_b(K_t)$, we have $F^\circ(\varphi_t) = u_t(\varphi_t) \in \mathcal{C}_b(K_t)$. So the additional constraint $F^\circ(\varphi_t) \in \mathcal{C}_b(K_t)$ (below [30, Equation (2.49)]) is redundant in our setup. $\qquad\square$

Proof of Proposition 3.10 We exploit again Liero et al. [30, Theorem 2.7 and Remark 2.8] (with u_t in place of F°), as we explain now. Since u_t is nondecreasing, either its domain is of the form $[M, \infty)$ or (M, ∞), with $M \leq 0$. Given $\varphi_t \in \mathcal{C}_b(K_t)$ and $\mu \in \mathrm{Meas}(K_t)$, we have three cases:

If $\inf(\varphi_t(\mathbb{R})) > M$, then $u_t(\varphi_t) \in \mathcal{C}_b(K_t)$ since u_t is continuous on the interior of its domain.

If $\inf(\varphi_t(\mathbb{R})) < M$, then $\{\varphi_t < M\}$ is open and non-empty and hence has positive $\widehat{Q}_t$-measure as $\widehat{Q}_t$ has full support. Thus $\int_{K_t} u_t(\varphi_t) \, d\widehat{Q}_t = -\infty$.

Finally, if $\inf(\varphi_t(\mathbb{R})) = M$, then $u_t(\varphi_t) = \lim_{\varepsilon \downarrow 0} u_t(\max(\varphi_t, M + \varepsilon))$ (since u_t is nondecreasing and upper semicontinuous), $u_t(\max(\varphi_t, M + \varepsilon)) \in \mathcal{C}_b(K_t)$ as in the first case, and by the monotone convergence theorem,

$$\int_{K_t} u_t(\varphi_t) \, d\widehat{Q}_t - \int_{K_t} \varphi_t \, d\mu$$

$$= \lim_{\varepsilon \downarrow 0} \left(\int_{K_t} u_t\big(\max(\varphi_t, M + \varepsilon)\big) \, d\widehat{Q}_t - \int_{K_t} \max(\varphi_t, M + \varepsilon) \, d\mu \right).$$

Then we infer that

$$\sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} \varphi_t \, d\mu - \int_{K_t} v_t(\varphi_t) \, d\widehat{Q}_t \right)$$

$$= \sup_{\varphi_t \in \mathcal{C}_b(K_t)} \left(\int_{K_t} u_t(\varphi_t) \, d\widehat{Q}_t - \int_{K_t} \varphi_t \, d\mu \right)$$

$$= \sup \left\{ \int_{K_t} u_t(\varphi_t) \, d\widehat{Q}_t - \int_{K_t} \varphi_t \, d\mu : \varphi_t, u_t(\varphi_t) \in \mathcal{C}_b(K_t) \right\}, \qquad (A.3)$$

and from [30, Theorem 2.7 and Remark 2.8] and (A.3), we deduce the result. $\qquad\square$

Declarations

Competing Interests The authors declare no competing interests.

References

1. Aliprantis, C.D., Border, K.C.: Infinite Dimensional Analysis, 3rd edn. Springer, Berlin (2006)
2. Backhoff-Veraguas, J., Pammer, G.: Stability of martingale optimal transport and weak optimal transport. Ann. Appl. Probab. **32**, 721–752 (2022)
3. Bartl, D., Kupper, M., Prömel, D.J., Tangpi, L.: Duality for pathwise superhedging in continuous time. Finance Stoch. **23**, 697–728 (2019)
4. Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model-independent bounds for option prices—a mass transport approach. Finance Stoch. **17**, 477–501 (2013)
5. Ben-Tal, A., Teboulle, M.: An old-new concept of convex risk measures: the optimized certainty equivalent. Math. Finance **17**, 449–476 (2007)
6. Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative Bregman projections for regularized transportation problems. SIAM J. Sci. Comput. **37**, A1111–A1138 (2015)
7. Bernton, E., Ghosal, P., Nutz, M.: Entropic optimal transport: geometry and large deviations. Duke Math. J. **171**, 3363–3400 (2022)
8. Biagini, S., Frittelli, M.: On the extension of the Namioka–Klee theorem and on the Fatou property for risk measures. In: Delbaen, F., et al. (eds.) Optimality and Risk—Modern Trends in Mathematical Finance. The Kabanov Festschrift, pp. 1–28. Springer, Berlin (2009)
9. Blanchet, J., Jambulapati, A., Kent, C., Sidford, A.: Towards optimal running times for optimal transport. Preprint (2020). Available online at https://arxiv.org/abs/1810.07717
10. Bogachev, V.I.: Measure Theory. Vol. II. Springer, Berlin (2007)
11. Bolley, F.: Applications du Transport Optimal à des Problèmes de Limites de Champ Moyen. Ph.D. Thesis, Ecole Normale Supérieure de Lyon – ENS LYON (2005). Available online at https://theses.hal.science/tel-00011462
12. Bolley, F.: Separability and completeness for the Wasserstein distance. In: Donati-Martin, C., et al. (eds.) Séminaire de Probabilités XLI. Lecture Notes in Math., vol. 1934, pp. 371–377. Springer, Berlin (2008)
13. Breeden, D.T., Litzenberger, R.H.: Prices of state-contingent claims implicit in option prices. J. Bus. **51**, 621–651 (1978)
14. Cheridito, P., Kiiski, M., Prömel, D.J., Soner, H.M.: Martingale optimal transport duality. Math. Ann. **379**, 1685–1712 (2021)
15. Cheridito, P., Kupper, M., Tangpi, L.: Duality formulas for robust pricing and hedging in discrete time. SIAM J. Financ. Math. **8**, 738–765 (2017)
16. Chung, N.-P., Trinh, T.-S.: Weak optimal entropy transport problems. Preprint (2021). Available online at https://arxiv.org/abs/2101.04986
17. Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Burges, C.J.C., et al. (eds.) Advances in Neural Information Processing Systems, vol. 26, pp. 2292–2300. Curran Associates, Red Hook (2013)
18. Davis, M., Obłój, J., Raval, V.: Arbitrage bounds for prices of weighted variance swaps. Math. Finance **24**, 821–854 (2014)
19. De March, H., Henry-Labordère, P.: Building arbitrage-free implied volatility: Sinkhorn's algorithm and variants. Preprint (2020). Available online at https://arxiv.org/abs/1902.04456

20. Dolinsky, Y., Soner, H.M.: Martingale optimal transport and robust hedging in continuous time. Probab. Theory Relat. Fields **160**, 391–427 (2014)

21. Föllmer, H., Schied, A.: Stochastic Finance. An Introduction in Discrete Time, 4th revised and extended edn. de Gruyter, Berlin (2016)

22. Galichon, A., Henry-Labordère, P., Touzi, N.: A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Ann. Appl. Probab. **24**, 312–336 (2014)

23. Ghosal, P., Nutz, M., Bernton, E.: Stability of entropic optimal transport and Schrödinger bridges. J. Funct. Anal. **283**, 109622 (2022)

24. Guo, G., Obłój, J.: Computational methods for martingale optimal transport problems. Ann. Appl. Probab. **29**, 3311–3347 (2019)

25. Henry-Labordère, P.: From (martingale) Schrödinger bridges to a new class of stochastic volatility models. Preprint (2019). Available online at https://arxiv.org/abs/1904.04554

26. Henry-Labordère, P., Obłój, J., Spoida, P., Touzi, N.: The maximum maximum of a martingale with given *n* marginals. Ann. Appl. Probab. **26**, 1–44 (2016)

27. Hobson, D.G.: Robust hedging of the lookback option. Finance Stoch. **2**, 329–347 (1998)

28. Hou, Z., Obłój, J.: Robust pricing–hedging dualities in continuous time. Finance Stoch. **22**, 511–567 (2018)

29. Ireland, C.T., Kullback, S.: Contingency tables with given marginals. Biometrika **55**, 179–188 (1968)

30. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. Invent. Math. **211**, 969–1117 (2018)

31. Neufeld, A., Sester, J.: On the stability of the martingale optimal transport problem: a set-valued map approach. Stat. Probab. Lett. **176**, 109131 (2021)

32. Nutz, M., Wiesel, J.: Entropic optimal transport: convergence of potentials. Probab. Theory Relat. Fields **184**, 401–424 (2022)

33. Pennanen, T., Perkkiö, A.-P.: Convex duality in nonlinear optimal transport. J. Funct. Anal. **277**, 1029–1060 (2019)

34. Peyré, G., Cuturi, M.: Computational Optimal Transport: With Applications to Data Science. Now Publishers, Hanover (2019)

35. Rüschendorf, L.: Convergence of the iterative proportional fitting procedure. Ann. Stat. **23**, 1160–1174 (1995)

36. Simons, S.: Minimax and Monotonicity. Lecture Notes in Mathematics, vol. 1693. Springer, Berlin (1998)

37. Tan, X., Touzi, N.: Optimal transportation under controlled stochastic dynamics. Ann. Probab. **41**, 3201–3240 (2013)

38. Villani, C.: Optimal Transport. Old and New. Springer, Berlin (2009)