



Adapted Wasserstein distances and stability in mathematical finance

Julio Backhoff-Veraguas^{1,2} · Daniel Bartl¹ ·
Mathias Beiglböck¹ · Manu Eder¹

Received: 6 March 2019 / Accepted: 8 January 2020 / Published online: 4 June 2020
© The Author(s) 2020

Abstract Assume that an agent models a financial asset through a measure \mathbb{Q} with the goal to price/hedge some derivative or optimise some expected utility. Even if the model \mathbb{Q} is chosen in the most skilful and sophisticated way, the agent is left with the possibility that \mathbb{Q} does not provide an *exact* description of reality. This leads us to the following question: will the hedge still be somewhat meaningful for models in the proximity of \mathbb{Q} ?

If we measure proximity with the usual Wasserstein distance (say), the answer is No. Models which are similar with respect to the Wasserstein distance may provide dramatically different information on which to base a hedging strategy.

Remarkably, this can be overcome by considering a suitable *adapted* version of the Wasserstein distance which takes the temporal structure of pricing models into account. This adapted Wasserstein distance is most closely related to the nested distance as pioneered by Pflug and Pichler (SIAM J. Optim. 20:1406–1420, 2009, SIAM J. Optim. 22:1–23, 2012, Multistage Stochastic Optimization, 2014). It allows us to establish Lipschitz properties of hedging strategies for semimartingale models in

J. Backhoff gratefully acknowledges financial support by the FWF through grant P30750 and by the Vienna University of Technology. D. Bartl has been funded by the Austrian Science Fund (FWF) under Project P28661. M. Beiglböck and M. Eder gratefully acknowledge financial support by the FWF through grant Y782.

✉ J. Backhoff-Veraguas
julio.backhoff@univie.ac.at

D. Bartl
daniel.bartl@univie.ac.at

M. Beiglböck
mathias.beiglboeck@univie.ac.at

M. Eder
manuel.eder@univie.ac.at

¹ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, Vienna 1090, Austria

² Present address: University of Twente, Drienerlolaan 5, 7522 NB Enschede, Netherlands

discrete and continuous time. Notably, these abstract results are sharp already for Brownian motion and European call options.

Keywords Hedging · Utility maximisation · Optimal transport · Causal optimal transport · Wasserstein distance · Sensitivity · Stability

Mathematics Subject Classification (2010) 91G80 · 60G42 · 60G44 · 90C15

JEL Classification G11 · C32 · C62

1 Introduction

1.1 Outline

Assume that a reference measure \mathbb{P} is used to model the evolution of a financial asset X with the purpose to hedge a financial claim or to maximise some expected utility. We do not expect that the model \mathbb{P} captures reality in an absolutely accurate way. However, supposing that \mathbb{P} is close enough to reality (described by a probability \mathbb{Q}), we still hope that a strategy which is developed for \mathbb{P} leads to reasonable results.

A main goal of this paper is to establish this intuitive idea rigorously based on a new notion of *adapted Wasserstein distance* \mathcal{AW}_p between semimartingale measures. To fix ideas, we provide a first example of the results we are after.

Theorem 1.1 *Let \mathbb{P}, \mathbb{Q} be continuous semimartingale models for the asset price process X , and assume that $C(X)$ denotes an L -Lipschitz payoff of a (path-dependent) derivative C . Assume that a predictable trading strategy $H = (H_t)$, $|H| \leq k$, and an initial endowment $m \in \mathbb{R}$ constitute a \mathbb{P} -superhedge of $C(X)$, i.e.,*

$$C(X) \leq m + (H \cdot X)_T, \quad \mathbb{P}\text{-almost surely.}$$

Then there is a predictable G such that m, G constitute an “almost” \mathbb{Q} -superhedge in the sense that

$$\mathbb{E}_{\mathbb{Q}}[(C(X) - m - (G \cdot X)_T)^+] \leq 6(k + L)\mathcal{AW}_1(\mathbb{P}, \mathbb{Q}). \quad (1.1)$$

While the adapted Wasserstein distance will be defined in abstract terms (see (1.3)), it relates directly to the model parameters for “simple” models. In particular, if \mathbb{P}, \mathbb{Q} are Brownian models with different volatilities, then the distance between these models is just the difference of the volatilities. Moreover, the bound in (1.1) (as well as further Lipschitz bounds given below) are already sharp in such a simple setting and for C a European call option.

Below we provide a number of results with a similar flavour as Theorem 1.1. For example, we provide versions where the hedging error is controlled in terms of risk measures, and we show that a Lipschitz bound of the type (1.1) applies (with bigger constants) if the same trading strategy H is applied in the model \mathbb{P} as well as in the

model \mathbb{Q} . Importantly, we establish that comparable results of Lipschitz-continuity apply to utility maximisation and utility indifference pricing.

We emphasise that familiar concepts such as the Lévy–Prokhorov metric or the usual Wasserstein distance do not appear suitable to derive results comparable to Theorem 1.1. For example, in the vicinity of financially meaningful models, there are models with *arbitrarily high arbitrage* even for bounded strategies; similar phenomena appear with respect to completeness/incompleteness. Instead, we introduce an adapted Wasserstein distance \mathcal{AW}_p which takes the temporal structure of semi-martingale models into account. These distances are conceptually closely related to the nested distance as pioneered by Pflug and Pichler [47, 48, 49]; see Acciaio et al. [1], Glanzer et al. [26], Bion-Nadal and Talay [18] for first articles which link such a type of distance to finance. We describe these contributions more closely in Sect. 2 below.

1.2 Notation and adapted Wasserstein distances

Throughout, we let

$$\Omega := \mathbb{R}^T \quad \text{or} \quad \Omega := C([0, T]).$$

The first setting is referred to as the discrete-time case, and the second as the continuous-time case.¹ In the first case, we denote by $I = \{1, \dots, T\}$ the time-index set, and in the second $I = [0, T]$. Throughout the article, we provide definitions and results without specifying which of the two cases we are referring to; this means that the definitions/results apply in both cases. Only occasionally we consider one case specifically, and in such a situation, we state this explicitly.

We interpret Ω as the set of all possible evolutions (in time) of the one-dimensional asset price. Importantly, *mutatis mutandis*, all our results (except Propositions 3.3, 3.6 and Example 3.4) remain true for multidimensional asset price processes (corresponding to $\Omega = (\mathbb{R}^d)^T$ resp. $\Omega = C([0, T]; \mathbb{R}^d)$). We chose to go for the one-dimensional version to simplify notation.

The mappings $X, Y: \Omega \rightarrow \Omega$ denote the canonical processes (i.e., the identity map), and we make the convention that on $\Omega \times \Omega$, the process X denotes the first coordinate and Y the second one. The spaces Ω and $\Omega \times \Omega$ are endowed with the maximum norm and the corresponding Borel σ -field. In continuous time, the space Ω is endowed with the right-continuous filtration generated by X ; in discrete time, we use the plain filtration generated by X . In any case, we denote this filtration by $\mathbb{F} = (\mathcal{F}_t)$ and endow $\Omega \times \Omega$ with the product filtration $\mathbb{F} \otimes \mathbb{F}$. Given a σ -field \mathcal{G} and a probability \mathbb{P} on \mathcal{G} , we write $\mathcal{G}^{\mathbb{P}}$ for the \mathbb{P} -completion of \mathcal{G} . The set $\text{Cpl}(\mathbb{P}, \mathbb{Q})$ of *couplings* between probability measures \mathbb{P}, \mathbb{Q} consists of all probability measures π on $\Omega \times \Omega$ such that $X(\pi) = \mathbb{P}$ and $Y(\pi) = \mathbb{Q}$. A *Monge coupling* is a coupling that is of the form $\pi = (\text{Id}, T)(\mathbb{P})$ for some Borel mapping $T: \Omega \rightarrow \Omega$ that transports \mathbb{P}

¹Indeed, the arguments in the discrete and the continuous case use the same set of ideas, but the presentation is significantly less technical in the discrete case, which was an important reason to include the discrete case in the paper.

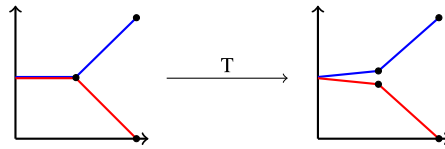


Fig. 1 Map T sends the blue path on the left to the blue path on the right, and similarly for the red paths. The stochastic processes depicted are close in Wasserstein sense, but very different for utility maximisation

to \mathbb{Q} , i.e., satisfies $T(\mathbb{P}) = \mathbb{Q}$. Given a metric d on Ω and $p \geq 1$, the p -Wasserstein distance of \mathbb{P}, \mathbb{Q} is

$$\mathcal{W}_p(\mathbb{P}, \mathbb{Q}) = \inf\{\mathbb{E}_\pi[d(X, Y)^p]^{1/p} : \pi \in \text{Cpl}(\mathbb{P}, \mathbb{Q})\}. \tag{1.2}$$

In many cases of practical interest, the infimum in (1.2) remains unchanged if one minimises only over Monge couplings; cf. [50].

Before defining the adapted Wasserstein distance between measures \mathbb{P} and \mathbb{Q} on Ω , let us hint why distances related to weak convergence are not suitable for the results we have in mind. Assume for example that we are interested in a utility maximisation problem in two periods and that Fig. 1 describes the laws \mathbb{P}, \mathbb{Q} of two traded assets. Clearly, they are very close in the Wasserstein distance, as follows from considering the obvious Monge coupling induced by $T : \Omega \rightarrow \Omega, T(\mathbb{P}) = \mathbb{Q}$ depicted in Fig. 1. At the same time, the outcome of utility maximisation is certainly very different. Similarly, \mathbb{P} is a martingale measure while \mathbb{Q} allows arbitrage. The clear reason for that is the different structure of information available at time 1.

To exhibit why the Wasserstein distance does not reflect this different structure of information, let us review the transport condition $T(\mathbb{P}) = \mathbb{Q}$. We rephrase it as

$$(T_1(X_1, X_2), T_2(X_1, X_2)) \stackrel{(d)}{=} (Y_1, Y_2), \tag{1.3}$$

where $\stackrel{(d)}{=}$ stands for equality in law. While this condition is of course perfectly natural in mass transport, (1.3) almost seems like cheating when viewed from a probabilistic perspective: the map T_1 should not be allowed to consider the future value X_2 in order to determine Y_1 . To define an adapted version of the Wasserstein distance, the “process” $(T_i)_{i=1,2}$ should be taken to be *adapted* in order to account for the different information structures of \mathbb{P} and \mathbb{Q} .

Naturally, our formal definition of adapted Wasserstein distances will not refer to adapted Monge transports, but rather to couplings which are “adapted” in an appropriate sense. Following Lassalle [41], we call such couplings (bi-)causal. Since the definition below may appear a bit technical at first glance, the following may be reassuring: In the discrete-time setting and for measures \mathbb{P} absolutely continuous with respect to Lebesgue measure, the weak closure (in the sense of weak convergence of measures) of the set of adapted Monge couplings, i.e., $\pi = (\text{Id}, T)(\mathbb{P})$ for T adapted, is precisely the set of all causal couplings; see Lacker [38].

Definition 1.2 For a coupling π of $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\Omega)$, let $\pi(d\omega, d\eta) = \mathbb{P}(d\omega)\pi_\omega(d\eta)$ denote a regular disintegration with respect to \mathbb{P} . The set $\text{Cpl}_C(\mathbb{P}, \mathbb{Q})$ of *causal couplings*

plings consists of all $\pi \in \text{Cpl}(\mathbb{P}, \mathbb{Q})$ such that for all $t \in I$ and $A \in \mathcal{F}_t$,

$$\omega \mapsto \pi_\omega(A) \text{ is } \mathcal{F}_t^{\mathbb{P}}\text{-measurable.}$$

The set of all *bi-causal* couplings $\text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q})$ consists of all $\pi \in \text{Cpl}_{\text{C}}(\mathbb{P}, \mathbb{Q})$ such that also $S(\pi) \in \text{Cpl}_{\text{C}}(\mathbb{Q}, \mathbb{P})$, where $S : \Omega \times \Omega \rightarrow \Omega \times \Omega$, $S(\omega, \eta) := (\eta, \omega)$.

In discrete time, a coupling π is causal if and only if

$$\pi((Y_1, \dots, Y_t) \in A | X) = \pi((Y_1, \dots, Y_t) \in A | X_1, \dots, X_t)$$

\mathbb{P} -a.s. for every t and Borel set $A \subseteq \mathbb{R}^t$, that is, at time t , given the past (X_1, \dots, X_t) of X , the distribution of Y_t does not depend on the future (X_{t+1}, \dots, X_N) of X .

Replacing couplings by bi-causal couplings in (1.2), one arrives at the nested distance as introduced by Pflug and Pichler [46, 47]. Since our goal is to compare also semimartingale models in continuous time, we work with an adapted Wasserstein distance that is defined slightly differently. (Notably, it is straightforward that the two distances are equivalent for probabilities on \mathbb{R}^N . We elaborate in Sect. 3.3 below why the definition in (1.4) is more appropriate for our purposes even in discrete time.)

In continuous time, we denote by $\mathcal{SM}(\Omega)$ the set of all probabilities \mathbb{P} on (the Borel σ -field of) Ω under which the canonical process X is a (continuous) semimartingale. In discrete time, $\mathcal{SM}(\Omega)$ denotes the set of all Borel probabilities \mathbb{P} on Ω under which X is integrable. In both cases, we can uniquely decompose $X = M + A$, with A a finite variation predictable process starting at zero and M a local martingale. Indeed, in the first case, X is a special semimartingale and M and A can be chosen continuous as well, and in the second case, this is the Doob decomposition of an integrable adapted discrete-time process. For $p \in [1, \infty)$, we denote by $\mathcal{SM}_p(\Omega)$ the subset of $\mathcal{SM}(\Omega)$ for which

$$\mathbb{E}_{\mathbb{P}}[|M|_T^{p/2} + |A|_{1\text{-var}}^p] < \infty,$$

where $[\cdot]$ is the quadratic variation and $|\cdot|_{1\text{-var}}$ the first variation. Note also that by the BDG inequality, $\mathbb{E}_{\mathbb{P}}[\sup_{s \leq T} |M_s|] < \infty$ for $\mathbb{P} \in \mathcal{SM}_p(\Omega)$; hence M is then a true martingale.

Definition 1.3 For $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$, $p \geq 1$, we define the *adapted Wasserstein distance* as

$$\begin{aligned} \mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) \\ := \inf \left\{ \mathbb{E}_{\pi} \left[|M^X - M^Y|_T^{p/2} + |A^X - A^Y|_{1\text{-var}}^p \right]^{1/p} : \pi \in \text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q}) \right\}, \end{aligned} \quad (1.4)$$

where $X = M^X + A^X$, $Y = M^Y + A^Y$ denote the semimartingale decompositions of X and Y , respectively.

It is shown in Lemma 3.1 that \mathcal{AW}_p is well defined (i.e., that $X - Y$ is a semimartingale under every bi-causal coupling) and in Lemma 3.2 that \mathcal{AW}_p in fact defines a metric.

Remark 1.4 In the continuous-time setup, the adapted Wasserstein distance can also be computed through

$$\mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) = \inf \left\{ \mathbb{E}_\pi \left[|X - Y|_T^{p/2} + \text{MV}_T[|X - Y|^p] \right]^{1/p} : \pi \in \text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q}) \right\}.$$

Here MV denotes the mean variation, i.e.,

$$\text{MV}_T[Z] = \sup_{\Delta} \sum_{t_j \in \Delta} |\mathbb{E}_\pi[Z_{t_{j+1}} - Z_{t_j} | \mathcal{F}_{t_j}]|,$$

where the supremum is taken over all finite partitions Δ of $[0, T]$.

In Sect. 3.2 below, we give explicit formulae for the adapted Wasserstein distance in the case of semimartingale measures described by simple SDEs.

1.3 Stability of superhedging

For the rest of this article, fix some $k \in \mathbb{R}_+$ and let \mathcal{H}_k be the set of all predictable processes

$$H : \Omega \times I \rightarrow [-k, k].$$

For every $p \geq 1$, write b_p for the “upper” Burkholder–Davis–Gundy (BDG) constant. In particular, it is known that $b_1 \leq 6$ and that $b_2 = 2$.

Our first main result concerns the stability of superhedging and constitutes a stronger version of Theorem 1.1 stated above.

Theorem 1.5 *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_1(\Omega)$, $H \in \mathcal{H}_k$ and let $C : \Omega \rightarrow \mathbb{R}$ be Lipschitz with constant L . Then the hedging error under \mathbb{Q} is bounded by the distance of \mathbb{P} and \mathbb{Q} plus the hedging error under \mathbb{P} in the following sense: There exists $G \in \mathcal{H}_k$ such that*

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[(C - m - (G \cdot X)_T)^+] &\leq \mathbb{E}_{\mathbb{P}}[(C - m - (H \cdot X)_T)^+] \\ &\quad + b_1(k + L)\mathcal{AW}_1(\mathbb{P}, \mathbb{Q}). \end{aligned} \tag{WHI}$$

Assume in addition that $H_t : \Omega \rightarrow \mathbb{R}$ is Lipschitz with constant \tilde{L} for every $t \in I$. Then we can take $G = H$ and obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}[(C - m - (H \cdot X)_T)^+] &\leq \mathbb{E}_{\mathbb{P}}[(C - m - (H \cdot X)_T)^+] \\ &\quad + b_1(k + L)\mathcal{AW}_1(\mathbb{P}, \mathbb{Q}) + \beta\mathcal{AW}_2(\mathbb{P}, \mathbb{Q}), \end{aligned} \tag{SHI}$$

where $\beta := 2\sqrt{2}b_1\tilde{L} \min\{\mathcal{AW}_2(\mathbb{P}, \delta_0), \mathcal{AW}_2(\mathbb{Q}, \delta_0)\}$.

Importantly, it is impossible to transfer a superhedge under \mathbb{P} into a superhedge under \mathbb{Q} . This occurs already in a one-period framework and is not a by-product of our definition of the adapted Wasserstein distance; see Remark 5.2. A similar reasoning requires to consider only trading strategies bounded by k ; see Remark 5.3.

It is worthwhile to compare the inequalities (WHI) and (SHI):

(S) In a certain sense, the “strong hedging inequality” (SHI) seems to be the more relevant assertion; after all, a trader does not know that the model \mathbb{Q} (rather than the model \mathbb{P}) describes reality and hence she might (somewhat stubbornly) stick to the initial plan of hedging her risk according to the strategy H . The inequality (SHI) then allows quantifying the losses due to this model error.

(W) However, the “weak hedging inequality” (WHI) also has a particular merit. Suppose that a trader W starts with the prior belief that the asset price evolves according to a Black–Scholes model with volatility σ_1 , but soon after time 0 realises that a volatility $\sigma_2 \neq \sigma_1$ yields a more adequate description of reality. If the witty trader W makes an accurate guess about the correct model and updates her trading strategy accordingly, her losses can be controlled through the tighter bound in (WHI).

In Theorem 4.2, we provide a version of Theorem 1.5 where $(\cdot)^+$ is replaced by a convex, strictly increasing loss function $\ell: \mathbb{R} \rightarrow \mathbb{R}_+$.

Another way to gauge the effectiveness of an almost superhedge is by means of risk measures. We postpone the general formulation to Theorem 4.3 and first present a version that appeals to the average value of risk $\text{AVaR}_\alpha^\mathbb{P}$. Recall that for a random variable $Z: \Omega \rightarrow \mathbb{R}$,

$$\text{AVaR}_\alpha^\mathbb{P}(Z) := \inf_{m \in \mathbb{R}} \mathbb{E}_\mathbb{P}[m + (Z - m)^+ / \alpha]$$

is the average value at risk at level $\alpha \in (0, 1)$ under model \mathbb{P} . We then have

Theorem 1.6 *Assume that $C: \Omega \rightarrow \mathbb{R}$ is Lipschitz with constant L . Then*

$$\left| \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{P}(C - (H \cdot X)_T) - \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{Q}(C - (H \cdot X)_T) \right| \leq r \mathcal{AW}_1(\mathbb{P}, \mathbb{Q})$$

for $r := b_1(L + k)/\alpha$. If $H \in \mathcal{H}_k$ is such that $H_t: \Omega \rightarrow [-k, k]$ is Lipschitz with constant \tilde{L} for every $t \in I$ and β is the constant defined in Theorem 1.5, then

$$\begin{aligned} \left| \text{AVaR}_\alpha^\mathbb{P}(C - (H \cdot X)_T) - \text{AVaR}_\alpha^\mathbb{Q}(C - (H \cdot X)_T) \right| &\leq r \mathcal{AW}_1(\mathbb{P}, \mathbb{Q}) \\ &\quad + \frac{\beta}{\alpha} \mathcal{AW}_2(\mathbb{P}, \mathbb{Q}). \end{aligned}$$

The interpretation of this result is similar to that of Theorem 1.5: As $\text{AVaR}_\alpha^\mathbb{P}(\cdot)$ is translation invariant, one has

$$\inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{P}(C - (H \cdot X)_T) = \inf \left\{ m \in \mathbb{R} : \begin{array}{l} \text{there is } H \in \mathcal{H}_k \text{ such that} \\ \text{AVaR}_\alpha^\mathbb{P}(C - m - (H \cdot X)_T) \leq 0 \end{array} \right\},$$

and the right-hand side constitutes a relaxed version of the superhedging price. Notably, the explicit calculations of the adapted Wasserstein distance given in Sect. 3.2 imply that Theorem 1.6 (and similarly Theorem 1.5) are sharp.

Example 1.7 For hedging in a Brownian framework, consider a European call option $C(X) = (X_T - K)^+$, where for simplicity $K = 0$. Moreover, let \mathbb{P}^σ be Wiener measure with constant volatility $\sigma \geq 0$. Then for every $\sigma, \hat{\sigma} \geq 0, k \geq 1$ and $\alpha \in (0, 1)$, it

holds that (we defer the proof of this fact to Sect. 4)

$$\begin{aligned} & \left| \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^{\mathbb{P}^\sigma} (C - (H \cdot X)_T) - \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^{\mathbb{P}^{\hat{\sigma}}} (C - (H \cdot X)_T) \right| \\ &= |\mathbb{E}_{\mathbb{P}^\sigma} [C] - \mathbb{E}_{\mathbb{P}^{\hat{\sigma}}} [C]| = \frac{1}{\sqrt{2\pi}} T |\sigma - \hat{\sigma}| = \frac{1}{\sqrt{2\pi}} \mathcal{AW}_1(\mathbb{P}^\sigma, \mathbb{P}^{\hat{\sigma}}). \end{aligned}$$

This shows that the estimate in Theorem 1.6 is tight (up to constants) in the sense that it is essentially impossible to improve on the probability metric \mathcal{AW}_1 .

We make the important remark that Glanzer et al. [26] use the nested distance to control acceptability prices in discrete-time models in a Lipschitz fashion through the nested distance of these models. Specifically, in a discrete-time one-period framework, [26, Proposition 3] and Theorem 1.6 yield almost the same assertion; in that setup, the only difference is that [26, Proposition 3] does not specify a Lipschitz constant and does not assume uniform boundedness of the admissible hedging strategy. (However, the latter seems to be in conflict with our Remark 5.3 below.)

1.4 Stability of utility maximisation and utility indifference pricing

We move on to consider the continuity of utility maximisation. Let $U : \mathbb{R} \rightarrow \mathbb{R}$ be a utility function which is concave and increasing, and denote by U' the left-continuous version of the derivative. We have

Theorem 1.8 *Let $C : \Omega \rightarrow \mathbb{R}$ be Lipschitz-continuous and assume that there exists $c \geq 0$ such that $U'(x) \leq c(1 + |x|^{p-1})$ for all x . Then for every $R \geq 0$, there exists a constant K such that*

$$\left| \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} [U(C + (H \cdot X)_T)] - \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}} [U(C + (H \cdot X)_T)] \right| \leq K \mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$$

for all $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$ with $\mathcal{AW}_p(\mathbb{P}, \delta_0), \mathcal{AW}_p(\mathbb{Q}, \delta_0) \leq R$.

The failure of usual Wasserstein distances to guarantee stability of utility maximisation is illustrated in Remark 5.1.

A common way of quantifying the value of a claim is via utility indifference pricing:² Given a claim C , the utility indifference (bid) price v is defined as a solution of the equation

$$\sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} [U(C - v + (H \cdot X)_T)] = \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} [U((H \cdot X)_T)].$$

Continuing in the spirit of the present paper, we are interested in the stability of $\mathbb{P} \mapsto v(\mathbb{P})$, where the latter denotes a utility indifference price associated to the model \mathbb{P} . If U is strictly increasing, then v is unique.

²We are grateful to an anonymous referee for pointing out that we could include the stability of utility indifference pricing with respect to the adapted Wasserstein distance.

Theorem 1.9 *Let $C: \Omega \rightarrow \mathbb{R}$ be Lipschitz-continuous and assume that there exists $c \geq 0$ such that $0 < U'(x) \leq c(1 + |x|^{p-1})$ for all x . Then for every $R \geq 0$, there exists a constant K such that*

$$|v(\mathbb{P}) - v(\mathbb{Q})| \leq K \mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$$

for all $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$ with $\mathcal{AW}_p(\mathbb{P}, \delta_0), \mathcal{AW}_p(\mathbb{Q}, \delta_0) \leq R$.

1.5 Structure of the paper

In Sect. 2, we briefly review the literature related to this paper. In Sect. 3, we establish some basic properties of the adapted Wasserstein distance, discuss the choice of cost function and give some examples. Moreover, we derive a contraction principle (Theorem 3.10) which relates the adapted Wasserstein distance with a “weak” (in the sense of Gozlan et al. [28]) transport distance. This result forms the basis for the proofs of the results mentioned in the introduction, as well as certain extensions of these results; see Sect. 4. Finally, we conclude with some remarks in Sect. 5.

2 Literature

The articles closest in spirit to ours are [1, 18, 26]. Acciaio et al. [1] consider an object related to the adapted Wasserstein distance in continuous time in connection with utility maximisation, enlargement of filtrations and optimal stopping. Glanzer et al. [26] prove a deviation inequality for the so-called nested distance in a discrete-time framework,³ and consider acceptability pricing over an ambiguity set described through the nested distance. Bion-Nadal and Talay [18] study via PDE arguments a continuous-time optimisation problem which is related to the adapted Wasserstein distance.

The concept of causal couplings, and optimal transport over causal couplings, has been recently popularised by Lassalle [41], although precursors can be found in the works by Yamada and Watanabe [55] and Rüschenendorf [52]. This notion is central to the recent articles by Acciaio et al. [1] and Backhoff-Veraguas et al. [10, 8, 9].

The idea of strengthening weak convergence of measures in order to account for a temporal evolution has some history. Indeed, several authors have independently introduced different approaches to address this challenge. The seminal unpublished work by Aldous [2] introduces the notion of extended weak convergence for the study of stability of optimal stopping problems. The principal idea is not to compare the laws of processes directly, but rather the laws of the corresponding prediction processes. Independently, Hellwig [29] introduces the information topology for the stability of equilibrium problems in economics. Roughly, two probability measures on a product $X_1 \times \cdots \times X_N$ of finitely many spaces are considered to be close if for each

³Note added in revision: improved convergence rates have been recently obtained in Backhoff-Veraguas et al. [7] for a related sample-based estimator. Together with the results of the present article, this gives statistical consistency for an empirical version of the financial problems considered.

$t \leq N$, the projections onto the first t coordinates as well as the corresponding conditional (regular) disintegrations are close. Unrelated to these developments, Pflug and Pichler [46, 47, 48] have introduced nested distances for the stability of stochastic programming in discrete time. The nested distance is the obvious role model for the adapted Wasserstein distances considered in this article, and (as mentioned above) for a fixed number of time steps and $p \geq 1$, they are obviously equivalent. Yet another idea to account for the temporal evolution of processes would be to symmetrise the causal transport costs $\mathcal{W}_c(\mathbb{P}, \mathbb{Q})$ defined by Lassalle [41] by taking the maximum or sum of $\mathcal{W}_c^2(\mathbb{P}, \mathbb{Q})$ and $\mathcal{W}_c^2(\mathbb{Q}, \mathbb{P})$; this was pointed out by Soumik Pal.

In parallel work [6], the four authors of the present article investigate the relations between these concepts in detail. Remarkably, in (finite) discrete time, *all* of the concepts mentioned above (adapted Wasserstein distances, extended weak convergence, information topology, nested distances, symmetrised causal transport costs) define the *same topology*. As noted above, this “weak adapted topology” refines the usual weak topology (properly for $T \geq 2$; see also Remark 5.2). The articles [8, 6, 23] investigate basic properties of this topology; e.g., the weak adapted topology is Polish [8, Sect. 5], and sets are totally bounded with respect to the adapted Wasserstein distance/nested distance if and only if they are totally bounded with respect to the usual Wasserstein distance [6, Lemma 1.6]. For recent applications of these concepts to optimal transport and probabilistic variants thereof, we refer to Backhoff-Veraguas et al. [11, 12] and Wiesel [54].

In contrast, fundamental topological properties of the above-mentioned concepts in the continuous-time case seem to be much less understood and, as far as the authors are concerned, pose an interesting challenge for future research. Specifically, it is not clear to us whether the topology associated to the adapted Wasserstein distance is Polish in the continuous-time case. In a similar vein, we expect that results analogous to those of the present article should apply in the case of càdlàg paths, but such an extension is beyond the scope of our current understanding of adapted Wasserstein distances.

The question of stability in mathematical finance has been studied from different perspectives over the years. Notably, starting with the articles of Lyons [42] and Avellaneda et al. [5], the area of robust finance has mainly focused on extremal models and hedging strategies which dominate the payoff for every model in a specified class. Following the publication of Hobson’s seminal article [32], connections with the Skorokhod embedding problem have been a driving force of the field; see the surveys of Hobson [34] and Obłój [44]. Recently, this has been complemented by techniques coming from (martingale) optimal transport; early papers which advance this viewpoint include Hobson [35], Beiglböck et al. [15, 16], Galichon et al. [25], Bouchard and Nutz [19], Dolinsky and Soner [22], Campi et al. [20], and Beiglböck and Siorpaes [14]. The literature on “local” misspecification of volatility in a sense more closely related to the present article appears more sparse. El Karoui et al. [24] establish in a stochastic volatility framework that if the misspecified volatility dominates the true volatility, then the misspecified price of call options dominates the real price; see also the elegant account of Hobson [33]. More recently, the question of pricing and hedging under uncertainty about the volatility of a reference local volatility model is studied by Herrmann et al. [31] (see also Herrmann and Muhle-Karbe

[30]). Less plausible models are penalised through a mean-square distance to the volatility of the reference model, and the authors obtain explicit formulas for prices and hedging strategies in a limit for small uncertainty aversion. Becherer and Kentia [13] derive worst-case good-deal bounds under model ambiguity which concerns drift as well as volatility. Indeed, discussions with Dirk Becherer motivated us to consider also models with drift in our results on stability of superhedging. The behaviour of the superhedging price in a ball (with respect to various notions of distance) around a reference model is studied in depth by Obłój and Wiesel [45] for a d -dimensional asset and one time period.

A notable implication of our work is that it yields a coherent way to measure model uncertainty (in the sense of Cont's influential article [21]): Fix a subset M_0 of the set M of all consistent models, i.e., martingale measures which are consistent with benchmark instruments whose price can be observed on the market. Given M_0 , the model uncertainty associated to a derivative f can be gauged through

$$\rho_{M_0}(f) := \sup\{\mathbb{E}_{\mathbb{Q}} f : \mathbb{Q} \in M_0\} - \inf\{\mathbb{E}_{\mathbb{Q}} f : \mathbb{Q} \in M_0\}.$$

The worst-case approach typically pursued in robust finance then yields $\rho_{M_0}(f)$ for $M_0 = M$, but it appears equally natural to take M_0 to be an infinitesimal ball around a reference model. This approach is being pursued by Bartl, Drapeau, Obłój, Wiesel and one of the present authors in a one-period framework. Our results indicate that an adapted Wasserstein distance provides a way to extend this to a multi-period setup, and we intend to pursue this further in future work.

On a different note, much work has been done regarding the convergence of discrete-time models to their continuous-time analogues. Due to the vastness of this literature, we refer the reader to the book by Prigent [51] for references. Finally, in more recent times and starting from the works of Kardaras and Žitković, the stability of utility maximisation has been studied in Kardaras and Žitković [37], Larsen [39], Larsen and Žitković [40], Mocha and Westray [43] and Weston [53], among others.

3 The adapted Wasserstein distance

3.1 Basic properties of \mathcal{AW}_p

The following lemma shows that \mathcal{AW}_p is well defined.

Lemma 3.1 *Let \mathbb{P}, \mathbb{Q} be integrable (semi-)martingale measures for $X, Y: \Omega \rightarrow \Omega$ respectively, and let π be a bi-causal coupling between \mathbb{P} and \mathbb{Q} . Then the process $X, Y, X - Y: \Omega \times \Omega \rightarrow \Omega$ are (semi-)martingales with respect to π . Further, if $X = M + A$ denotes the semimartingale decomposition under \mathbb{P} , then up to evanescence, $M + A$ is the semimartingale decomposition of X under π .*

Proof Let $X = M + A$ be the semimartingale decomposition under \mathbb{P} and consider M and A as processes on $\Omega \times \Omega$ via $M(\omega, \eta) := M(\omega)$ and $A(\omega, \eta) := A(\omega)$. Further let $\pi = \mathbb{P}(d\omega)\pi_\omega(d\eta)$ be a bi-causal coupling between \mathbb{P} and \mathbb{Q} . To show that $X = M + A$ remains the semimartingale decomposition under π , it is enough to show

that M is a martingale under π . To that end, let $0 \leq s \leq t$ and let $Z : \Omega \times \Omega \rightarrow \mathbb{R}$ be $\mathcal{F}_s \otimes \mathcal{F}_s$ -measurable and bounded. (Recall that $\mathbb{F} = (\mathcal{F}_t)$ denotes the right-continuous filtration generated by X and that we endow $\Omega \times \Omega$ with the filtration $\mathbb{F} \otimes \mathbb{F}$.) Then the random variable $Z' : \Omega \rightarrow \mathbb{R}$ defined by

$$Z'(\omega) := \int Z(\omega, \eta) \pi_\omega(d\eta) \quad \text{is } \mathcal{F}_s^{\mathbb{P}}\text{-measurable,}$$

and clearly bounded. Indeed, if $Z(\omega, \eta) = Z^1(\omega)Z^2(\eta)$ for \mathcal{F}_s -measurable bounded functions Z^1 and Z^2 , it follows from the definition of bi-causality that Z' is $\mathcal{F}_s^{\mathbb{P}}$ -measurable; the general statement then follows from a monotone class argument. Therefore

$$\begin{aligned} \mathbb{E}_\pi[(M_t - M_s)Z] &= \int (M_t(\omega) - M_s(\omega)) \int Z(\omega, \eta) \pi_\omega(d\eta) \mathbb{P}(d\omega) \\ &= \mathbb{E}_{\mathbb{P}}[(M_t - M_s)Z'] \\ &= 0 \end{aligned}$$

by the martingale property of M under \mathbb{P} . This shows that M is a martingale under π and therefore that $X = M + A$ is the semimartingale decomposition under π . \square

Lemma 3.2 \mathcal{AW}_p defines a metric on the set $\mathcal{SM}_p(\Omega)$.

We note that very similar arguments could be used to show that \mathcal{AW}_p defines a metric for semimartingales with infinite time horizon on \mathbb{N} or $[0, \infty)$.

Proof of Lemma 3.2 It is clear that $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) = \mathcal{AW}_p(\mathbb{Q}, \mathbb{P}) \geq 0$ for all \mathbb{P}, \mathbb{Q} in $\mathcal{SM}_p(\Omega)$. Suppose that $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) = 0$. As $\|\cdot\|_\infty \leq |\cdot|_{1\text{-var}}$, it is immediate that if π participates in the infimum defining $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$ and $X - Y = M + A$, then

$$\mathbb{E}_\pi[\|X - Y\|_\infty^p] \leq 2^{p-1} \mathbb{E}_\pi[\|M\|_\infty^p + |A|_{1\text{-var}}^p] \leq 2^{p-1} b_p \mathbb{E}_\pi[|M|_T^{p/2} + |A|_{1\text{-var}}^p],$$

where b_p denotes the BDG constant and we used the BDG inequality for the martingale M . Hence the usual Wasserstein distance between \mathbb{P} and \mathbb{Q} (defined with respect to the $\|\cdot\|_\infty$ -norm) is dominated from above by $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$, and so $\mathbb{P} = \mathbb{Q}$.

We now prove the triangle inequality. Let $\mathbb{P}, \mathbb{Q}, \mathbb{R}$ given. We fix $\varepsilon > 0$ and assume π is bi-causal ε -optimal for $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$ and $\tilde{\pi}$ is bi-causal ε -optimal for $\mathcal{AW}_p(\mathbb{Q}, \mathbb{R})$. In the next couple of lines, ω always denotes the first coordinate of a vector in Ω^3 , η the second and γ the last. Let

$$\pi(d\omega, d\eta) = \pi_\eta(d\omega) \mathbb{Q}(d\eta) \quad \text{and} \quad \tilde{\pi}(d\eta, d\gamma) = \tilde{\pi}_\eta(d\gamma) \mathbb{Q}(d\eta)$$

be disintegrations, and define $\Pi \in \mathcal{P}(\Omega^3)$ by

$$\Pi(d\omega, d\eta, d\gamma) = \pi_\eta(d\omega) \tilde{\pi}_\eta(d\gamma) \mathbb{Q}(d\eta).$$

If $\bar{\pi}(d\omega, d\gamma) := \int_{\Omega} \Pi(d\omega, d\eta, d\gamma)$ is the projection of Π onto the first and third components, then it is clear that the first and second marginals of $\bar{\pi}$ are \mathbb{P} and \mathbb{R} , respectively. Moreover, a disintegration of $\bar{\pi} = \bar{\pi}_{\omega}(d\gamma) \mathbb{P}(d\omega)$ is given by

$$\bar{\pi}_{\omega}(d\gamma) = \int_{\Omega} \tilde{\pi}_{\eta}(d\gamma) \pi_{\omega}(d\eta),$$

where as indicated above, π_{ω} now denotes the disintegration of π with respect to the first coordinate, that is, $\pi(d\omega, d\eta) = \pi_{\omega}(d\eta) \mathbb{P}(d\omega)$. We claim that for every $A \in \mathcal{F}_t$, the mapping $\omega \mapsto \bar{\pi}_{\omega}(A)$ is $\mathcal{F}_t^{\mathbb{P}}$ -measurable. Indeed, by bi-causality of $\tilde{\pi}$, one has that $\eta \mapsto \tilde{\pi}_{\eta}(A)$ is $\mathcal{F}_t^{\mathbb{Q}}$ -measurable. Thus there are an \mathcal{F}_t -measurable function X and a \mathbb{Q} -almost surely zero function N such that $\tilde{\pi}_{\eta}(A) = X(\eta) + N(\eta)$ for all $\eta \in \Omega$. Then $\bar{\pi}_{\omega}(A) = \int_{\Omega} X(\eta) \pi_{\omega}(d\eta) + \int_{\Omega} N(\eta) \pi_{\omega}(d\eta)$ for all $\eta \in \Omega$. The first term is $\mathcal{F}_t^{\mathbb{P}}$ -measurable (by bi-causality of π), and as π is a coupling between \mathbb{P} and \mathbb{Q} , one has that $\int_{\Omega} N(\eta) \pi_{\omega}(d\eta) = 0$ for \mathbb{P} -almost all $\omega \in \Omega$.

The argument for $\bar{\pi} = \bar{\pi}_{\gamma}(d\omega) \mathbb{R}(d\gamma)$ is similar and therefore $\bar{\pi}$ is a bi-causal coupling between \mathbb{P} and \mathbb{R} . Finally, it follows as in the proof of Lemma 3.1 that if $X = M^X + A^X$, $Y = M^Y + A^Y$ and $Z = M^Z + A^Z$ are the semimartingale decompositions under \mathbb{P} , \mathbb{Q} , and \mathbb{R} , then they remain the semimartingale decompositions under Π on Ω^3 endowed with the product filtration.

To finish the proof of the triangle inequality, we observe that

$$\begin{aligned} \mathcal{AW}_p(\mathbb{P}, \mathbb{R}) &\leq \mathbb{E}_{\bar{\pi}}[|M^X - M^Z|_T^{p/2} + |A^X - A^Z|_{1\text{-var}}^p]^{1/p} \\ &= \mathbb{E}_{\Pi}[(M^X - M^Y) + (M^Y - M^Z)]_T^{p/2} \\ &\quad + |(A^X - A^Y) + (A^Y - A^Z)|_{1\text{-var}}^p]^{1/p}. \end{aligned}$$

The function $M \mapsto \mathbb{E}_{\Pi}[|M|_T^{p/2}]^{1/p}$ is known to be a norm on the space $\mathcal{M}_p(\Pi)$ of Π -martingales starting at zero whose supremum is p -integrable. Likewise, the function $A \mapsto \mathbb{E}_{\Pi}[|A|_{1\text{-var}}^p]^{1/p}$ is a norm on the space of finite variation processes with p -integrable variation. Hence

$$(M, A) \mapsto \|(M, A)\| := \mathbb{E}_{\Pi}[|M|_T^{p/2} + |A|_{1\text{-var}}^p]^{1/p}$$

is a norm on the product of these spaces. We conclude the proof for the triangle inequality with

$$\begin{aligned} \mathcal{AW}_p(\mathbb{P}, \mathbb{R}) &\leq \|(M^X - M^Y, A^X - A^Y) + (M^Y - M^Z, A^Y - A^Z)\| \\ &\leq \|(M^X - M^Y, A^X - A^Y)\| + \|(M^Y - M^Z, A^Y - A^Z)\| \\ &= \mathbb{E}_{\pi}[|M^X - M^Y|_T^{p/2} + |A^X - A^Y|_{1\text{-var}}^p]^{1/p} \\ &\quad + \mathbb{E}_{\tilde{\pi}}[|M^Y - M^Z|_T^{p/2} + |A^Y - A^Z|_{1\text{-var}}^p]^{1/p} \\ &\leq 2\varepsilon + \mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) + \mathcal{AW}_p(\mathbb{Q}, \mathbb{R}), \end{aligned}$$

as the semimartingale decomposition of $X - Y$ under π is $(M^X - M^Y) + (A^X - A^Y)$, with an analogous expression for $Y - Z$ under $\tilde{\pi}$.

To conclude the proof, it remains to show that we have $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) < \infty$ for all $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$. But Lemma 3.1 gives $\mathcal{AW}_p(\mathbb{P}, \delta_0) = \mathbb{E}_{\mathbb{P}}[[M]_T^{p/2} + |A|_{1\text{-var}}^p]^{1/p}$, where $X = M + A$ is the semimartingale decomposition under \mathbb{P} . Therefore the triangle inequality implies that \mathcal{AW}_p is real-valued on $\mathcal{SM}_p(\Omega)$. \square

3.2 Examples and explicit calculations

We start by a simple result which permits to give a closed-form expression of the adapted Wasserstein distance in given continuous-time situations.

Proposition 3.3 *For $i \in \{1, 2\}$, consider the SDEs with bounded progressive coefficients*

$$dX_t^i = \mu_i(t, (X_s^i)_{s \leq t})dt + \sigma_i(t, (X_s^i)_{s \leq t})dB_t^i. \tag{3.1}$$

Assume that each SDE admits a unique strong solution and denote by $\mathbb{P}^{\mu_i, \sigma_i}$ the respective laws. Further assume that

- μ_1 is a function of time only (namely $\mu_1 : [0, T] \rightarrow \mathbb{R}$);
- $\sigma_1, \sigma_2 \geq 0$ and at least one of them is a function of time only.

Then the synchronous coupling (namely $\pi^ =$ joint law of (X^1, X^2) , where $B^1 = B^2$ in (3.1)) is optimal in the definition of $\mathcal{AW}_p(\mathbb{P}^{\mu_1, \sigma_1}, \mathbb{P}^{\mu_2, \sigma_2})$.*

The discrete-time version of the above synchronous coupling is given by the Knothe–Rosenblatt rearrangement [10], and a variant of the previous result can also be obtained in the discrete-time framework.

Proof Proof of Proposition 3.3 Let π be a feasible coupling for $\mathcal{AW}_p(\mathbb{P}^{\mu_1, \sigma_1}, \mathbb{P}^{\mu_2, \sigma_2})$, leading to a finite cost. For this proof, we denote the coordinate process on $\Omega \times \Omega$ by (X^1, X^2) . As before, let $X^i = A^i + M^i$ be the unique continuous semimartingale decomposition of X^i under the $\mathbb{P}^{\mu_i, \sigma_i}$ -completion of its right-continuous filtration. Observe that $t \mapsto \frac{d}{dt}A_t^1$ is a.s. deterministic, by the assumption on μ_1 , and that the law of $t \mapsto \frac{d}{dt}A_t^2$ is independent of the coupling π . Both facts can be easily derived from the identity

$$\frac{d}{dt}A_t^i = \lim_{\varepsilon \searrow 0} \frac{\mathbb{E}_{\pi}[X_{t+\varepsilon}^i | \mathcal{F}_t^{X^i}] - X_t^i}{\varepsilon},$$

which by the Lebesgue differentiation theorem holds $dt \otimes d\pi$ -a.e. As a consequence, the term $\mathbb{E}_{\pi}[|A^1 - A^2|_{1\text{-var}}^p]$ is independent of the coupling π , and so we may ignore it and only focus on the term $\mathbb{E}_{\pi}[|M^1 - M^2|_T^{p/2}]$.

By Doob’s martingale representation [36, Theorem 3.4.2], on a possibly enlarged filtered probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\pi})$, we may represent the martingale (M^1, M^2) by

$$M^i = \int \sigma_{i1}dW + \int \sigma_{i2}d\hat{W},$$

where W, \hat{W} are independent standard one-dimensional Brownian motions and $\sigma_{ik}, i, k \in \{1, 2\}$, real-valued processes, both of them adapted in the enlarged filtered space. In the following, we omit the argument $(X_s^i)_{s \leq t}$ from σ_i . Necessarily, we have

$$\sigma_{i,t}^2 = \frac{d}{dt}[M^i]_t = \sigma_{i1,t}^2 + \sigma_{i2,t}^2 \quad dt \otimes d\tilde{\pi}\text{-a.e.}$$

By the Cauchy–Schwarz inequality, we deduce that almost surely,

$$[M^1, M^2]_T = \int_0^T (\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22})dt \leq \int_0^T \sigma_1\sigma_2dt,$$

and accordingly we get the lower bound

$$\mathbb{E}_\pi[[M^1 - M^2]_T^{p/2}] \geq \mathbb{E}_\pi\left[\left(\int_0^T (\sigma_1 - \sigma_2)^2 dt\right)^{p/2}\right].$$

As in the beginning of the proof, the right-hand side does not depend on the coupling π thanks to one of the σ^i being a function of time only. To conclude, observe that for the synchronous coupling π^* , we have equality in the above equation. \square

As an easy consequence, we have

Example 3.4 For bounded Lipschitz functions $\mu_1, \mu_2, \sigma_1, \sigma_2$, we denote by $\mathbb{P}^{\mu_i, \sigma_i}$ the law of the diffusion

$$dX_t^i = \mu_i(t, X_t^i)dt + \sigma_i(t, X_t^i)dB_t.$$

Assume that

- μ_i is independent of the x -variable for some $i \in \{1, 2\}$, and
- σ_k is independent of the x -variable for some $k \in \{1, 2\}$.

For $j \in \{1, 2\} \setminus \{i\}$ and $\ell \in \{1, 2\} \setminus \{k\}$, we have

$$\begin{aligned} \mathcal{AW}_p(P^{\mu_1, \sigma_1}, P^{\mu_2, \sigma_2})^p &= \mathbb{E}\left[\left(\int_0^T (\sigma_\ell(t, X_t^\ell) - \sigma_k(t))^2 dt\right)^{p/2}\right] \\ &\quad + \mathbb{E}\left[\left(\int_0^T |\mu_j(t, X_t^j) - \mu_i(t)| dt\right)^p\right]. \end{aligned}$$

We now illustrate that in general, it is not true that the straightforward synchronous coupling of Proposition 3.3 is optimal. As a consequence, we do not expect a closed-form expression for the adapted Wasserstein distance. A discrete-time version of this observation is discussed in [8, Sect. 7].

Example 3.5 Consider $d = 1, T = 2$ and for each $c \in \mathbb{R}$ introduce

$$\mu_t^c(\omega) := c1_{[1,2]}(t) \text{sign}(\omega_1), \quad \hat{\mu}_t^c(\omega) := -\mu_t^c(\omega).$$

Assuming that B is a Brownian motion and for $\sigma \in \mathbb{R}_+$, we introduce the couplings

$$\begin{aligned} \pi_1 &:= \text{Law} \left(\sigma B + \int \mu_t^c(B)dt, \sigma B + \int \hat{\mu}_t^c(B)dt \right), \\ \pi_2 &:= \text{Law} \left(\sigma B + \int \mu_t^c(B)dt, -\sigma B + \int \hat{\mu}_t^c(-B)dt \right). \end{aligned}$$

These couplings share the same marginals and each of them is bi-causal. Writing as before $X - Y = M + A$, it is easy to compute

$$\begin{aligned} \mathbb{E}_{\pi_1} \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right] &= (2c)^p, \\ \mathbb{E}_{\pi_2} \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right] &= (8\sigma^2)^{p/2}. \end{aligned}$$

We conclude that for each p , there are plenty of pairs (c, σ) such that the ‘‘synchronous’’ coupling π_1 is not optimal between its marginals for the metric \mathcal{AW}_p .

To close this section, we estimate the distance between two geometric Brownian motions with different volatilities.

Proposition 3.6 *For $i = 1, 2$, let \mathbb{P}^{σ_i} denote the law of the solution to the SDE $dZ_t^i = \sigma_i Z_t^i dB_t^i$ with $Z_0^i = 1$, where B^i denotes Brownian motion and $\sigma_i \in \mathbb{R}_+$. Letting $R \sim \mathcal{N}(0, T)$, we then have*

$$\mathcal{AW}_2(\mathbb{P}^{\sigma_1}, \mathbb{P}^{\sigma_2})^2 = \mathbb{E} \left[\left(e^{\sigma_1 R - \frac{\sigma_1^2 T}{2}} - e^{\sigma_2 R - \frac{\sigma_2^2 T}{2}} \right)^2 \right] = e^{\sigma_1^2 T} - 2e^{\sigma_1 \sigma_2 T} + e^{\sigma_2^2 T},$$

and for $p > 1$,

$$\mathcal{AW}_p(\mathbb{P}^{\sigma_1}, \mathbb{P}^{\sigma_2})^p \leq c_p \mathbb{E} \left[\left(e^{\sigma_1 R - \frac{\sigma_1^2 T}{2}} - e^{\sigma_2 R - \frac{\sigma_2^2 T}{2}} \right)^p \right],$$

where c_p is the constant in the BDG inequality which allows controlling the quadratic variation by the terminal value.

Proof We have

$$\begin{aligned} &\mathcal{AW}_p(\mathbb{P}^{\sigma_1}, \mathbb{P}^{\sigma_2})^p \\ &= \inf \left\{ \mathbb{E}_{\pi} \left[[Z^1 - Z^2]_T^{p/2} \right] : \pi \in \text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q}) \right\} \\ &\leq c_p \inf \left\{ \mathbb{E}_{\pi} \left[(Z_T^1 - Z_T^2)^p \right] : \pi \in \text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q}) \right\} \\ &= c_p \inf \left\{ \int \left(e^{\sigma_1 r_1 - \frac{\sigma_1^2 T}{2}} - e^{\sigma_2 r_2 - \frac{\sigma_2^2 T}{2}} \right)^p d\pi(r_1, r_2) : \pi \in \text{Cpl}(\gamma_T, \gamma_T) \right\} \\ &= c_p \mathbb{E} \left[\left(e^{\sigma_1 R - \frac{\sigma_1^2 T}{2}} - e^{\sigma_2 R - \frac{\sigma_2^2 T}{2}} \right)^p \right], \end{aligned}$$

where γ_T denotes a centered Gaussian with variance T . For $p = 2$ and $c_2 = 1$, we obtain equality. □

3.3 Choice of the “cost functional”

Recall from Definition 1.3 that the adapted Wasserstein distance is given through

$$\mathcal{AW}_p(\mathbb{P}, \mathbb{Q}) := \inf\{\Phi : \pi \in \text{Cpl}_{\text{BC}}(\mathbb{P}, \mathbb{Q})\},$$

where the “cost functional”

$$\Phi = \mathbb{E}_\pi \left[|M^X - M^Y|_T^{p/2} + |A^X - A^Y|_{1\text{-var}}^p \right]^{1/p} \quad (3.2)$$

is defined using the semimartingale decompositions $X = M^X + A^X$, $Y = M^Y + A^Y$. The distinctive property of this “quadratic plus first variation” functional is that it exhibits the proper scaling to interpret the discrete-time case as an approximation to the continuous-time counterpart. To see this, consider $\Omega = C([0, 1])$ and let \mathbb{P}^σ be the law of X , where $X_t = \int_0^t \sigma_s dB_s$, B is a Brownian motion and $\sigma \in C([0, 1])$, $\sigma \geq 0$. For each N , denote by \mathbb{P}_N^σ the law of a random walk on $\{0, 1/N, 2/N, \dots, 1\}$ with independent increments from n/N to $(n+1)/N$ distributed according to $\mathcal{N}(0, \sigma_{n/N}^2/N)$. Then one can compute that for $0 \leq \sigma, \sigma' \in C([0, 1])$,

$$\begin{aligned} \mathcal{AW}_2(\mathbb{P}_N^\sigma, \mathbb{P}_N^{\sigma'}) &= \left(\sum_{n=0}^{N-1} \frac{1}{N} |\sigma_{n/N} - \sigma'_{n/N}|^2 \right)^{1/2} \\ &\longrightarrow \left(\int_0^1 |\sigma_t - \sigma'_t|^2 dt \right)^{1/2} = \mathcal{AW}_2(\mathbb{P}^\sigma, \mathbb{P}^{\sigma'}). \end{aligned}$$

For comparison, consider the consequences of replacing the “cost functional” Φ in (3.2) with $\Phi = \mathbb{E}_\pi \left[\sum_{i=0}^N (X_i - Y_i)_i^2 \right]^{1/2}$ corresponding to a quadratic nested distance (in terms of Pflug and Pichler [47]). While $\widetilde{\mathcal{AW}}_2$ and \mathcal{AW}_2 are equivalent metrics for each fixed N , $\widetilde{\mathcal{AW}}_2$ does not exhibit the appropriate scaling for large N . A straightforward computation shows that $\widetilde{\mathcal{AW}}_2(\mathbb{P}_N^\sigma, \mathbb{P}_N^{\sigma'}) \rightarrow \infty$ as $N \rightarrow \infty$ whenever $\sigma \neq \sigma'$. In consequence, bounds on the hedging error in terms of $\widetilde{\mathcal{AW}}_2(\mathbb{P}_N^\sigma, \mathbb{P}_N^{\sigma'})$ become progressively weaker as $N \rightarrow \infty$. In particular, they do not allow a meaningful continuous-time limit.

When restricting solely to martingale measures \mathbb{P}, \mathbb{Q} , a sensible alternative to (3.2) would be to consider the maximum norm, i.e., $\Phi' = \mathbb{E}_\pi [\sup_t |X_t - Y_t|^p]^{1/p}$. In fact, by the BDG inequalities, this is essentially equivalent to our choice in (3.2). However, when considering semimartingales, this cost is too coarse. For example, let (ω_n) be a sequence in Ω which converges to zero in the maximum norm, but for which the first variation tends to infinity. Then $\mathbb{P}_n := \delta_{\omega_n}$ converges to $\mathbb{P} := \delta_0$ (when the adapted distance is defined only with the maximum norm as cost), but none of our optimisation problems converge (take a strategy $H \in \mathcal{H}_k$ for which $(H(X) \bullet X)_T \approx k|\omega_n|_{1\text{-var}}$ almost surely).

3.4 Stochastic integrals and a contraction principle

We present here the two technical results which underlie the proofs of the main theorems in the article. The first one is

Lemma 3.7 *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_1(\Omega)$, $H \in \mathcal{H}_k$ and π be a bi-causal coupling between \mathbb{P} and \mathbb{Q} . Then there exists a process $G \in \mathcal{H}_k$ such that $G_t(Y) = \mathbb{E}_\pi[H_t(X)|Y]$ for every t , π -almost surely. Moreover, we have $(G(Y) \cdot Y)_T = \mathbb{E}_\pi[(H(X) \cdot Y)_T|Y]$ π -almost surely.*

Proof In discrete time, we can always write $H = \sum_{t=1}^N H_t 1_{\{t\}}$ for Borel functions $H_t: \mathbb{R}^{t-1} \rightarrow [-k, k]$. Let $\pi = \pi_\eta(d\omega) \mathbb{P}(d\omega)$ be a disintegration and define

$$G'_t(\eta) := \int H_t(\omega)\pi_\eta(d\omega)$$

for every t and $\eta \in \Omega$. By the definition of a bi-causal coupling, G'_t is $\mathcal{F}_{t-1}^{\mathbb{Q}}$ -measurable. It remains to pick functions G_t which are \mathcal{F}_{t-1} -measurable such that $G_t = G'_t$ \mathbb{Q} -almost surely. Since $\mathbb{E}_\pi[H_t(X)|Y] = G_t(Y)$ π -almost surely, it is clear that $(G(Y) \cdot Y)_T = \mathbb{E}_\pi[(H(X) \cdot Y)_T|Y]$ π -almost surely.

In continuous time, we take G to be the predictable projection of H under the reference measure π , with respect to the π -completion of the filtration $\{\emptyset, \Omega\} \otimes \mathbb{F}^Y$. By [1, Lemma C.1], the result is π -indistinguishable from a predictable process under the \mathbb{Q} -completion of the filtration \mathbb{F}^Y . The t -by- t , π -almost sure equality $G_t(Y) = \mathbb{E}_\pi[H_t(X)|Y]$ is then a consequence of the definition of the predictable projection. The π -almost sure equality $(G(Y) \cdot Y)_T = \mathbb{E}_\pi[(H \cdot Y)_T|Y]$ is established in Lemma 3.8 below, assuming that $\mathbb{E}_{\mathbb{Q}}[[Y]_T] < \infty$. The general case follows by localisation. □

Lemma 3.8 *In the continuous-time context of Lemma 3.7, assume further that we have $\mathbb{E}_{\mathbb{Q}}[[Y]_T] < \infty$. Then $(G(Y) \cdot Y)_T = \mathbb{E}_\pi[(H(X) \cdot Y)_T|Y]$ π -almost surely.*

Proof The statement is true if instead of the stochastic integrals, we consider the integrals with respect to the finite variation part of Y (either by properties of Riemann–Stieltjes integrals, or directly from the definition of the predictable projection). For this reason, we may now assume that Y is itself a martingale.

We first take for granted the following result: If h is bounded and predictable in the filtration of (X, Y) and if g denotes its predictable projection in the filtration of Y under the measure π , then

$$\mathbb{E}_\pi \left[\int_0^T |g_t|^2 d[Y]_t \right] \leq \mathbb{E}_\pi \left[\int_0^T |h_t|^2 d[Y]_t \right]. \tag{3.3}$$

We know that there exists a sequence (H^n) of predictable simple processes such that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\pi \left[\int_0^T |H_t - H_t^n|^2 d[Y]_t \right] = 0.$$

By the Itô isometry, the stochastic integrals $(H^n \cdot Y)_T$ converge in $L^2(\pi)$ to $(H \cdot Y)_T$. Denoting by G^n the predictable projection of H^n with respect to the Y -filtration, we deduce from (3.3) that

$$\lim_{n \rightarrow \infty} \mathbb{E}_\pi \left[\int_0^T |G_t - G_t^n|^2 d[Y]_t \right] = 0;$$

so again by the Itô isometry, $(G^n \cdot Y)_T$ converges in $L^2(\pi)$ to $(G \cdot Y)$. The π -almost sure equality $(G^n \cdot Y)_T = \mathbb{E}_\pi[(H^n \cdot Y)_T | Y]$ follows easily by the bi-causality of the coupling π , and by taking L^2 -limits, the desired conclusion is obtained.

To finish the proof, we must establish (3.3). First we observe that

$$\begin{aligned} \mathbb{E}_\pi \left[\int_0^T |g_t|^2 d[Y]_t \right]^{1/2} &= \sup_{\substack{f \text{ is } Y\text{-predictable,} \\ \|f\| \leq 1}} \mathbb{E}_\pi \left[\int_0^T f_t g_t d[Y]_t \right] \\ &= \sup_{\substack{f \text{ is } Y\text{-predictable,} \\ \|f\| \leq 1}} \mathbb{E}_\pi \left[\int_0^T f_t h_t d[Y]_t \right], \end{aligned}$$

as follows from predictable projection and with $\|f\|^2 := \mathbb{E}_\pi[\int_0^1 |f_t|^2 d[Y]_t]$. The result (3.3) is then a consequence of the equality

$$\mathbb{E}_\pi \left[\int_0^T |h_t|^2 d[Y]_t \right]^{1/2} = \sup_{\substack{f \text{ is } (X, Y)\text{-predictable,} \\ \|f\| \leq 1}} \mathbb{E}_\pi \left[\int_0^T f_t h_t d[Y]_t \right]. \quad \square$$

Our next crucial technical result is given in Theorem 3.10 below. But first we need some preparation.

Lemma 3.9 *Let $p \geq 1$. Let $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$, let π be a bi-causal coupling between \mathbb{P} and \mathbb{Q} , let $H \in \mathcal{H}_k$ and write $X - Y = M + A$ for the semimartingale decomposition under π . Then we have*

$$\begin{aligned} \mathbb{E}_\pi [\|X - Y\|_\infty^p] &\leq 2^{p-1} b_p \mathbb{E}_\pi [[M]_T^{p/2} + |A|_{1\text{-var}}^p], \\ \mathbb{E}_\pi [|(H(X) \cdot X)_T - (H(X) \cdot Y)_T|^p] &\leq 2^{p-1} b_p k^p \mathbb{E}_\pi [[M]_T^{p/2} + |A|_{1\text{-var}}^p], \end{aligned}$$

where b_p is the upper constant in the BDG inequality. If further $H_t: \Omega \rightarrow \mathbb{R}$ is \tilde{L} -Lipschitz-continuous for every t , then we have

$$\begin{aligned} \mathbb{E}_\pi [|(H(X) \cdot X)_T - (H(Y) \cdot Y)_T|^p] &\leq 2^{2p-2} b_p k^p \mathbb{E}_\pi [[M]_T^{p/2} + |A|_{1\text{-var}}^p] \\ &\quad + \alpha \mathbb{E}_\pi [[M]_T^p + |A|_{1\text{-var}}^{2p}]^{1/2} \end{aligned}$$

where $\alpha = 2^{3p-2} \tilde{L}^p b_p b_{2p}^{1/2} \min\{\mathcal{AW}_{2p}(\mathbb{P}, \delta_0)^p, \mathcal{AW}_{2p}(\mathbb{Q}, \delta_0)^p\}$.

Proof The elementary inequality $(x + y)^p \leq 2^{p-1} x^p + 2^{p-1} y^p$ for $x, y \geq 0$ together with the BDG inequality and the fact that $\|\cdot\|_\infty \leq |\cdot|_{1\text{-var}}$ implies

$$\begin{aligned} \mathbb{E}_\pi [\|X - Y\|_\infty^p] &\leq 2^{p-1} \mathbb{E}_\pi [\|M\|_\infty^p] + 2^{p-1} \mathbb{E}_\pi [|A|_{1\text{-var}}^p] \\ &\leq 2^{p-1} b_p \mathbb{E}_\pi [[M]_T^{p/2} + |A|_{1\text{-var}}^p]. \end{aligned}$$

This proves the first part. The same arguments imply

$$\begin{aligned} & \mathbb{E}_\pi \left[\left| (H(X) \cdot X)_T - (H(X) \cdot Y)_T \right|^p \right] \\ & \leq 2^{p-1} \mathbb{E}_\pi \left[\left| (H(X) \cdot M)_T \right|^p \right] + 2^{p-1} \mathbb{E}_\pi \left[\left| (H(X) \cdot A)_T \right|^p \right] \\ & \leq 2^{p-1} k^p b_p \mathbb{E}_\pi \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right], \end{aligned}$$

from which the second part follows. To prove the third claim, write

$$\begin{aligned} & \mathbb{E}_\pi \left[\left| (H(X) \cdot X)_T - (H(Y) \cdot Y)_T \right|^p \right] \\ & \leq 2^{p-1} \mathbb{E}_\pi \left[\left| \left((H(X) - H(Y)) \cdot X \right)_T \right|^p \right] \\ & \quad + 2^{p-1} \mathbb{E}_\pi \left[\left| (H(Y) \cdot X)_T - (H(Y) \cdot Y)_T \right|^p \right]. \end{aligned}$$

The second term is at most $2^{p-1} 2^{p-1} k^p b_p \mathbb{E}_\pi \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right]$ by the second part. It remains to estimate $\mathbb{E}_\pi \left[\left| \left((H(X) - H(Y)) \cdot X \right)_T \right|^p \right]$. Write $X = N + B$ for the semimartingale decomposition of X under \mathbb{P} . By Lemma 3.1, the semimartingale decomposition under π is still $X = N + B$. Moreover, the BDG inequality, the Lipschitz-continuity of H and Hölder’s inequality imply that

$$\begin{aligned} & \mathbb{E}_\pi \left[\left| \left((H(X) - H(Y)) \cdot X \right)_T \right|^p \right] \\ & \leq 2^{p-1} \mathbb{E}_\pi \left[\left| \left((H(X) - H(Y)) \cdot N \right)_T \right|^p + \left| \left((H(X) - H(Y)) \cdot B \right)_T \right|^p \right] \\ & \leq 2^{p-1} \mathbb{E}_\pi \left[\|H(X) - H(Y)\|_\infty^p (b_p [N]_T^{p/2} + |B|_{1\text{-var}}^p) \right] \\ & \leq 2^{p-1} b_p \tilde{L}^p \mathbb{E}_\pi \left[\|X - Y\|_\infty^{2p} \right]^{1/2} \mathbb{E}_\pi \left[([N]_T^p + |B|_{1\text{-var}}^p)^2 \right]^{1/2}. \end{aligned}$$

It now follows from the first part that

$$\mathbb{E}_\pi \left[\|X - Y\|_\infty^{2p} \right]^{1/2} \leq (2^{2p-1} b_{2p})^{1/2} \mathbb{E}_\pi \left[[M]_T^p + |A|_{1\text{-var}}^{2p} \right]^{1/2},$$

and by Lemma 3.1, we have

$$\mathbb{E}_\pi \left[([N]_T^{p/2} + |B|_{1\text{-var}}^p)^2 \right]^{1/2} \leq 2^{1/2} \mathcal{AW}_{2p}(\mathbb{P}, \delta_0)^p.$$

Putting all estimates together and replacing X and Y yields the claim. □

Denote by $\mathcal{P}_p(\mathbb{R})$ the set of all Borel probability measures μ on \mathbb{R} such that $\int |x|^p \mu(dx) < \infty$. Moreover, let $d_p(\mu, \nu)$ be the usual p -Wasserstein distance, and let d_p^w be the weak p -Wasserstein cost, that is,

$$\begin{aligned} d_p(\mu, \nu) & := \inf \left\{ \left(\int |x - y|^p \gamma(dx, dy) \right)^{1/p} : \gamma \text{ is a coupling of } \mu \text{ and } \nu \right\}, \\ d_p^w(\mu, \nu) & := \inf \left\{ \left(\int \left| x - \int y \gamma^x(dy) \right|^p \mu(dx) \right)^{1/p} : \gamma \text{ is a coupling of } \mu \text{ and } \nu \right\}. \end{aligned}$$

Here $\gamma = \mu(dx)\gamma^x(dy)$ denotes the disintegration. Note that d_p^w is not symmetric and that as a consequence of Jensen’s inequality, we always have $d_p^w \leq d_p$. Problems akin to $d_p^w(\mu, \nu)$ go under the name of “weak optimal transport” and have been recently introduced by Gozlan et al. in [28], but see also Alfonsi et al. [3], Alibert et al. [4], Backhoff-Veraguas et al. [11, 9] and Gozlan et al. [27]. We have

Theorem 3.10 *Let $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$, let π be a bi-causal coupling between \mathbb{P} and \mathbb{Q} , let $C : \Omega \rightarrow \mathbb{R}$ be Lipschitz with constant L , and let $H \in \mathcal{H}_k$. Further denote by $X - Y = M + A$ the semimartingale decomposition under π and let $G \in \mathcal{H}_k$ be such that $(G(Y) \cdot Y)_T = \mathbb{E}_\pi[(H(X) \cdot Y)_T | Y]$ π -almost surely. Then*

$$\begin{aligned} & d_p^w \left(\left(C(Y) + (G(Y) \cdot Y)_T \right) (\mathbb{Q}), \left(C(X) + (H(X) \cdot X)_T \right) (\mathbb{P}) \right) \\ & \leq 2^{(p-1)/p} b_p^{1/p} (k + L) \mathbb{E}_\pi \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right]^{1/p}. \end{aligned} \tag{3.4}$$

Now assume in addition that $H_t : \Omega \rightarrow \mathbb{R}$ is \tilde{L} -Lipschitz-continuous for every t . Then

$$\begin{aligned} & d_p \left(\left(C(Y) + (H(Y) \cdot Y)_T \right) (\mathbb{Q}), \left(C(X) + (H(X) \cdot X)_T \right) (\mathbb{P}) \right) \\ & \leq 2^{(3p-3)/p} b_p^{1/p} (k + L) \mathbb{E}_\pi \left[[M]_T^{p/2} + |A|_{1\text{-var}}^p \right]^{1/p} \\ & \quad + \alpha^{1/p} \mathbb{E}_\pi \left[[M]_T^p + |A|_{1\text{-var}}^{2p} \right]^{1/2p}, \end{aligned}$$

where α is the constant from Lemma 3.9.

Proof We start by proving the first claim. Let π be as stated, and define

$$a(X) := C(X) + (H(X) \cdot X)_T, \quad b(Y) := C(Y) + (G(Y) \cdot Y)_T.$$

Now let $\gamma := (b(Y), a(X))(\pi)$ so that γ is trivially a coupling between $b(Y)(\mathbb{Q})$ and $a(X)(\mathbb{P})$. Therefore

$$d_p^w(b(Y)(\mathbb{Q}), a(X)(\mathbb{P})) \leq \mathbb{E}_\pi \left[|b(Y) - \mathbb{E}_\pi[a(X)|b(Y)]|^p \right]^{1/p}.$$

By assumption, it holds that

$$\mathbb{E}_\pi \left[(G(Y) \cdot Y)_T - (H(X) \cdot X)_T | Y \right] = \mathbb{E}_\pi \left[(H(X) \cdot Y)_T - (H(X) \cdot X)_T | Y \right].$$

Thus by using the tower property and Jensen’s inequality, it follows that

$$\begin{aligned} & \mathbb{E}_\pi \left[|b(Y) - \mathbb{E}_\pi[a(X)|b(Y)]|^p \right]^{1/p} \\ & \leq \mathbb{E}_\pi \left[\left| \mathbb{E}_\pi[C(Y) - C(X)|Y] + \mathbb{E}_\pi \left[(G(Y) \cdot Y)_T - (H(X) \cdot X)_T | Y \right] \right|^p \right]^{1/p} \\ & \leq \mathbb{E}_\pi \left[|C(Y) - C(X)|^p \right]^{1/p} + \mathbb{E}_\pi \left[\left| (H(X) \cdot Y)_T - (H(X) \cdot X)_T \right|^p \right]^{1/p}. \end{aligned}$$

The claim now follows from the first and second estimates in Lemma 3.9.

If H is additionally Lipschitz, let

$$d(X) := C(X) + (H(X) \bullet X)_T, \quad e(Y) := C(Y) + (H(Y) \bullet Y)_T$$

as well as $\gamma := (e(Y), d(Y))(\pi)$. Then similarly as before,

$$\begin{aligned} & d_p(e(Y)(\mathbb{Q}), d(X)(\mathbb{P})) \\ & \leq \mathbb{E}_\pi[|e(Y) - d(Y)|^p]^{1/p} \\ & \leq \mathbb{E}_\pi[|C(Y) - C(X)|^p]^{1/p} + \mathbb{E}_\pi[|(H(Y) \bullet Y)_T - (H(X) \bullet X)_T|^p]^{1/p}, \end{aligned}$$

and the claim follows from the first and third estimates in Lemma 3.9. □

Remark 3.11 An evident question is whether an estimate for the usual Wasserstein distance holds true without the (Lipschitz-)continuity assumption on H , i.e., whether (3.4) holds for d_p instead of d_p^w . The following example shows that this is not true. In a two-period discrete-time model ($T = 2$), let

$$\mathbb{P} := \delta_0 \otimes ((\delta_1 + \delta_{-1})/2), \quad \mathbb{P}_\varepsilon := ((\delta_\varepsilon + \delta_{-\varepsilon})/2) \otimes ((\delta_1 + \delta_{-1})/2)$$

so that $\mathcal{AW}_p(\mathbb{P}_\varepsilon, \mathbb{P}) \rightarrow 0$ as $\varepsilon \rightarrow 0$ for every p . Then we define $H_1 := 0$ and $H_2 := 1_{(0,\infty)} - 1_{(-\infty,0)}$. For the projection under any bi-causal coupling between \mathbb{P}_ε and \mathbb{P} of H onto Y , one computes $G_1 = 0$ and $G_2 = 0$. In particular, $(G(Y) \bullet Y)_T = 0$ \mathbb{P} -almost surely. However, for every $\varepsilon > 0$, one has $\mathbb{P}_\varepsilon[(H(X) \bullet X)_T \geq 1 - \varepsilon] \geq 1/4$ which implies that the respective laws cannot converge.

4 Proofs of the results stated in the introduction and extensions

Thanks to the work done in the previous section, the strategy for the proofs boils down to two parts. In the first step, one forgets about the space Ω and only focuses on continuity of the problem at hand with respect to d_p or d_p^w when image measures on \mathbb{R} are plugged in; e.g. in utility maximisation, this means to study continuity of $\mu \mapsto \int U(x) \mu(dx)$. In the second step, one uses the obtained continuity and the contraction result in Theorem 3.10.

4.1 Proof of Theorem 1.5

We need the following elementary estimate.

Lemma 4.1 *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ and let $f: \mathbb{R} \rightarrow \mathbb{R}$ be convex and Lipschitz. Then*

$$\int f(x) \mu(dx) - \int f(y) \nu(dy) \leq L d_1^w(\mu, \nu), \tag{4.1}$$

where L is the Lipschitz constant of f .

Proof Let γ be a coupling of μ and ν . Applying Jensen's inequality, we obtain

$$\begin{aligned} \int f(x) \mu(dx) - \int f(y) \nu(dy) &= \int (f(x) - f(y)) \gamma(dx, dy) \\ &= \int \left(f(x) - \int f(y) \gamma^x(dy) \right) \mu(dx) \\ &\leq \int \left(f(x) - f\left(\int y \gamma^x(dy)\right) \right) \mu(dx) \\ &\leq L \int \left| x - \int y \gamma^x(dy) \right| \mu(dx). \end{aligned}$$

As γ was arbitrary, this implies the claim. \square

In fact, there is equality in the previous lemma if one takes the supremum on the left-hand side of (4.1) over all L -Lipschitz convex functions; this is shown in Gozlan et al. [28, Proposition 3.2].

We now turn to the

Proof of Theorem 1.5 For $n > 0$, let π be a bi-causal coupling which attains the infimum in the definition of $\mathcal{AW}_1(\mathbb{P}, \mathbb{Q})$ up to $1/n$. By Lemma 3.7, there is $G^n \in \mathcal{H}_k$ such that $(G^n(Y) \cdot Y)_T = \mathbb{E}_\pi[(H(X) \cdot Y)_T | Y]$ π -almost surely. Define

$$\mu^n := \left(C(Y) + (G^n(Y) \cdot Y)_T \right) (\mathbb{Q}), \quad \nu := \left(C(X) + (H(X) \cdot X)_T \right) (\mathbb{P}).$$

(Note that $\mu^n, \nu \in \mathcal{P}_1(\mathbb{R})$ as $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_1(\Omega)$.) By Lemma 4.1, we have

$$\begin{aligned} \mathbb{E}_\mathbb{Q} \left[\left(C(Y) - m - (G^n(Y) \cdot Y)_T \right)^+ \right] - \mathbb{E}_\mathbb{P} \left[\left(C(X) - m - (H(X) \cdot X)_T \right)^+ \right] \\ \leq d_1^w(\mu^n, \nu). \end{aligned}$$

From Theorem 3.10, we obtain

$$\begin{aligned} \mathbb{E}_\mathbb{Q} \left[\left(C(Y) - m - (G^n(Y) \cdot Y)_T \right)^+ \right] \leq \mathbb{E}_\mathbb{P} \left[\left(C(X) - m - (H(X) \cdot X)_T \right)^+ \right] \\ + b_1(k+L)(\mathcal{AW}_1(\mathbb{P}, \mathbb{Q}) + 1/n). \quad (4.2) \end{aligned}$$

Assume first that $\mathbb{E}_\mathbb{Q}[[Y]_T] < \infty$ and denote by A the finite variation process associated to Y . Then, as (G^n) is uniformly in n bounded by k , there exist a predictable G and a sequence of forward-convex combinations of (G^n) which converge in $L^2(d\mathbb{Q} \otimes d([Y] + A))$ to G . This, (4.2) and the convexity of $(\cdot)^+$ lead to the desired conclusion. The general case follows by a simple but notationally heavy localisation argument.

The proof in case that $G = H$ and H is Lipschitz follows analogously from the second part of Theorem 3.10. \square

4.2 Proof of Theorem 1.6

In a first step notice that for all \mathbb{P}, \mathbb{P}' and random variables Z, Z' , it follows as in Lemma 4.1 that

$$\text{AVaR}_\alpha^\mathbb{P}(Z) - \text{AVaR}_\alpha^{\mathbb{P}'}(Z') \leq d_1^w(Z(\mathbb{P}), Z'(\mathbb{P}'))/\alpha.$$

Indeed, if γ is a coupling from $\mu := Z(\mathbb{P})$ to $\nu := Z'(\mathbb{P}')$, then

$$\begin{aligned} & \text{AVaR}_\alpha^\mathbb{P}(Z) - \text{AVaR}_\alpha^{\mathbb{P}'}(Z') \\ &= \inf_m \int \left(\frac{1}{\alpha}(x - m)^+ - m \right) \mu(dx) - \inf_m \frac{1}{\alpha} \int \left(\int (y - m)^+ \gamma^x(dy) - m \right) \nu(dy) \\ &\leq \sup_m \frac{1}{\alpha} \int ((x - m)^+ - (y - m)^+) \gamma(dx, dy) \\ &\leq \sup_m \frac{1}{\alpha} \int \left((x - m)^+ - \left(\int y \gamma^x(dy) - m \right)^+ \right) \mu(dx) \\ &\leq \frac{1}{\alpha} \int \left| x - \int y \gamma^x(dy) \right| \mu(dx), \end{aligned}$$

so that minimising over γ yields the claim.

The rest of the proof now follows the same line of argumentation as in the proof of Theorem 1.5. Fix $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_1(\Omega)$. Assume (only for notational simplicity) that there exists a bi-causal coupling π which attains the infimum in the definition of $\mathcal{AW}_1(\mathbb{P}, \mathbb{Q})$, and that there exist $H^* \in \mathcal{H}_k$ such that

$$\text{AVaR}_\alpha^\mathbb{P}(C(X) - (H^*(X) \cdot X)_T) = \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H(X) \cdot X)_T).$$

By Lemma 3.7, there is $G^* \in \mathcal{H}_k$ such that $(G^*(Y) \cdot Y)_T = \mathbb{E}_\pi[(H^*(X) \cdot X)_T | Y]$ π -almost surely. Therefore

$$\begin{aligned} & \inf_{G \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{Q}(C(Y) - (G(Y) \cdot Y)_T) - \inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H(X) \cdot X)_T) \\ &\leq \text{AVaR}_\alpha^\mathbb{Q}(C(Y) - (G^*(Y) \cdot Y)_T) - \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H^*(X) \cdot X)_T) \\ &\leq \frac{1}{\alpha} d_1^w \left((C(Y) - (G^*(Y) \cdot Y)_T)(\mathbb{Q}), (C(X) - (H^*(X) \cdot X)_T)(\mathbb{P}) \right) \\ &\leq \frac{b_1(k + L)}{\alpha} \mathcal{AW}_1(\mathbb{P}, \mathbb{Q}), \end{aligned}$$

where the last inequality is due to Theorem 3.10. Interchanging the role of \mathbb{P} and \mathbb{Q} yields the desired conclusion. The proof for the second estimate follows analogously. □

4.3 Proof of Example 1.7

First note that $\text{AVaR}_\alpha^\mathbb{P}(Z) \geq \mathbb{E}_\mathbb{P}[Z]$ for every integrable random variable Z . Indeed, this follows from integrating the pointwise inequality

$$x = x + m - m \leq (x + m)^+ / \alpha - m.$$

Therefore, as the Brownian stochastic integral has expectation zero, we conclude that $\inf_{H \in \mathcal{H}_k} \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H(X) \cdot X)_T) \geq \mathbb{E}_\mathbb{P}[C(X)]$. On the other hand, define

$$f(t, x) := \int C(x + y) \mathcal{N}(0, \sigma^2(T - t))(dy) \quad \text{for } (t, x) \in [0, T] \times \mathbb{R},$$

where $\mathcal{N}(0, \sigma^2(T - t))$ stands for the normal distribution with mean 0 and variance $\sigma^2(T - t)$. Then $C(X) = f(T, X_T)$ and $\mathbb{E}_\mathbb{P}[f(t, X_t) | \mathcal{F}_s] = f(s, X_s)$ for every $0 \leq s \leq t \leq T$. Thus Itô's formula and the fact that the martingale property implies that the finite variation part vanishes imply $f(T, X_T) = f(0, 0) + (H^*(X) \cdot X)_T$ for the predictable trading strategy $H_t^* := \partial_x f(t, X_t)$. As further $|H_t^*| \leq 1$ for every t and $f(0, 0) = \sigma / \sqrt{2\pi}$, one has

$$\inf_{H \in \mathcal{H}_1} \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H(X) \cdot X)_T) \leq \text{AVaR}_\alpha^\mathbb{P}(C(X) - (H^*(X) \cdot X)_T) = \frac{\sigma}{\sqrt{2\pi}}.$$

The proof now follows from the explicit formula for the adapted Wasserstein distance derived in Example 3.4 and the fact that $\mathbb{E}_\mathbb{P}[C(X)] = \sigma / \sqrt{2\pi}$. \square

4.4 Proof of Theorem 1.8

Recall that we have $U'(x) \leq c(1 + |x|^{p-1})$ for all $x \in \mathbb{R}$ and some constant c . Let $\mathbb{P}, \mathbb{Q} \in \mathcal{SM}_p(\Omega)$ be arbitrary and assume (only for notational simplicity) that there is $H^* \in \mathcal{H}_k$ such that

$$\mathbb{E}_\mathbb{P}\left[U\left(C(X) + (H^*(X) \cdot X)_T\right)\right] = \sup_{H \in \mathcal{H}_k} \mathbb{E}_\mathbb{P}\left[U\left(C(X) + (H(X) \cdot X)_T\right)\right]$$

and that there is a bi-causal coupling π between \mathbb{P} and \mathbb{Q} which is optimal for $\mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$. Due to Lemma 3.7, there exists $G^* \in \mathcal{H}_k$ with the property that $(G^*(Y) \cdot Y)_T = \mathbb{E}_\pi[(H^*(X) \cdot Y)_T | Y]$ π -almost surely. Let

$$\mu := \left(C(Y) + (G^*(Y) \cdot Y)_T\right)(\mathbb{Q}), \quad \nu := \left(C(X) + (H^*(X) \cdot X)_T\right)(\mathbb{P}),$$

and let γ be an (almost) optimal coupling for $d_p^w(\mu, \nu)$. As U is concave and increasing, we have $U(y) - U(x) \leq U'(\min\{x, y\})|x - y|$. Using Jensen's inequality for the concave function U gives

$$\begin{aligned} & \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} \left[U \left(C(X) + (H(X) \cdot X)_T \right) \right] - \sup_{G \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}} \left[U \left(C(Y) + (G(Y) \cdot Y)_T \right) \right] \\ & \leq \mathbb{E}_{\mathbb{P}} \left[U \left(C(X) + (H^*(X) \cdot X)_T \right) \right] - \mathbb{E}_{\mathbb{Q}} \left[U \left(C(Y) + (G^*(Y) \cdot Y)_T \right) \right] \\ & = \int (U(y) - U(x)) \gamma(dx, dy) \leq \int \left(U \left(\int y \gamma^x(dy) \right) - U(x) \right) \mu(dx) \\ & \leq \left(\int \left| U' \left(\min \left\{ x, \int y \gamma^x(dy) \right\} \right) \right|^q \mu(dx) \right)^{1/q} d_p^w(\mu, \nu), \end{aligned}$$

where we used Hölder’s inequality in the last line and q denotes the conjugate exponent of p , i.e., $1/p + 1/q = 1$. As $q(p - 1) = p$, the growth assumption on U' implies that $|U'(\min\{x, y\})|^q \leq c(1 + |x|^p + |y|^p)$ for some (new) constant c . Then by Lemma 3.9, we have

$$\begin{aligned} & \int \left| U' \left(\min \left\{ x, \int y \gamma^x(dy) \right\} \right) \right|^q \mu(dx) \\ & \leq c \left(1 + \int |x|^p \mu(dx) + \int \left| \int y \gamma^x(dy) \right|^p \mu(dx) \right) \\ & \leq c \left(1 + \int |x|^p \mu(dx) + \int |y|^p \nu(dy) \right) \\ & \leq \tilde{c} (1 + \mathcal{AW}_p(\mathbb{Q}, \delta_0)^p + \mathcal{AW}_p(\mathbb{P}, \delta_0)^p) =: e \end{aligned}$$

for $e := \tilde{c}(1 + R^p + R^p)$. Exchanging the roles of \mathbb{P} and \mathbb{Q} and using Theorem 3.10 completes the proof. □

4.5 Proof of Theorem 1.9

In a first step, we claim that $v(\mathbb{P})$ is uniformly bounded over all \mathbb{P} which satisfy $\mathcal{AW}_p(\mathbb{P}, \delta_0) \leq R$. Indeed, using the growth assumption on U , the fact that U is strictly increasing and the BDG inequality to control the p th moment of $(H \cdot X)_T$, it follows that there exist $a, A \in \mathbb{R}$ such that

$$\inf U < a \leq \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} \left[U \left((H \cdot X)_T \right) \right] \leq A < \sup U \tag{4.3}$$

for all \mathbb{P} with $\mathcal{AW}_p(\mathbb{P}, \delta_0) \leq R$. Now assume that there exists a sequence (\mathbb{P}_n) with $\mathcal{AW}_p(\mathbb{P}_n, \delta_0) \leq R$, but $v(\mathbb{P}_n) \rightarrow \infty$. Then using the BDG inequality once more, it follows that

$$\sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}_n} \left[U \left(C - v(\mathbb{P}_n) + (H \cdot X)_T \right) \right] \longrightarrow \inf U$$

which is a contradiction to (4.3). The case $v(\mathbb{P}_n) \rightarrow -\infty$ is excluded analogously.

At this point, using the definition of $v(\mathbb{P})$, a twofold application of Theorem 1.8 yields

$$\left| \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}}[U(C - v(\mathbb{P}) + (H \cdot X)_T)] - \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}}[U((H \cdot X)_T)] \right| \leq 2K \mathcal{AW}_p(\mathbb{P}, \mathbb{Q}).$$

Indeed, while a direct application of the theorem would give a constant K which depends on $v(\mathbb{P})$, an inspection of its proof shows that the constant K depends only on the size of $v(\mathbb{P})$. By the first step, this is bounded uniformly over \mathbb{P} with $\mathcal{AW}_p(\mathbb{P}, \delta_0) \leq R$.

Now let $\varepsilon > 0$ and $H \in \mathcal{H}_k$ be arbitrary and set $Y := Y^H := C - v(\mathbb{P}) + (H \cdot X)_T$. Then it follows that there is some constant $c > 0$ (depending on $R \geq \mathcal{AW}_p(\mathbb{Q}, \delta_0)$ and U) such that

$$\mathbb{E}_{\mathbb{Q}}[U(Y + \varepsilon)] = \mathbb{E}_{\mathbb{Q}}[U(Y)] + \mathbb{E}_{\mathbb{Q}} \left[\int_Y^{Y+\varepsilon} U'(z) dz \right] \geq \mathbb{E}_{\mathbb{Q}}[U(Y)] + \varepsilon c.$$

Indeed, this would follow directly if Y were bounded by a fixed constant, but readily extends to the present setting as $\mathbb{E}_{\mathbb{Q}}[|Y|^p] \leq C$ for some constant $C > 0$, independent of H and \mathbb{Q} , as long as $\mathcal{AW}_p(\mathbb{Q}, \delta_0) \leq R$. In a similar manner, we also obtain $\mathbb{E}_{\mathbb{Q}}[U(Y - \varepsilon)] \leq \mathbb{E}_{\mathbb{Q}}[U(Y)] - \varepsilon c$. Putting everything together yields

$$\begin{aligned} & \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}} \left[U \left(C - (v(\mathbb{P}) + \varepsilon) + (H \cdot X)_T \right) \right] \\ & < \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}} [U((H \cdot X)_T)] < \sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{Q}} \left[U \left(C - (v(\mathbb{P}) - \varepsilon) + (H \cdot X)_T \right) \right] \end{aligned}$$

for some $\varepsilon < \tilde{C} \mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$, where \tilde{C} is a new constant emerging from K and c . Thus $|v(\mathbb{Q}) - v(\mathbb{P})| \leq \varepsilon \leq \tilde{C} \mathcal{AW}_p(\mathbb{P}, \mathbb{Q})$ which completes the proof. \square

4.6 Two generalisations

The following two results can be proved using almost the same arguments as in the proofs of Theorems 1.6 and 1.8. In particular, the proofs boil down to establishing convergence for image measures with respect to d_p and give no new insight on adapted Wasserstein distances, so we skip them.

Proposition 4.2 *Let $\ell: \mathbb{R} \rightarrow \mathbb{R}_+$ be a convex and strictly increasing function and let $\delta > 0$. Assume that $p \geq 1$ is such that $\ell'(x) \leq c(1 + |x|^{p-1})$ for some constant c . Then for every Lipschitz-continuous function $C: \Omega \rightarrow \mathbb{R}$, the function*

$$\mathbb{P} \mapsto \inf \left\{ m \in \mathbb{R} : \exists H \in \mathcal{H}_k \text{ such that } \mathbb{E}_{\mathbb{P}} \left[\ell \left(C(X) - (H(X) \cdot X)_T - m \right) \right] \leq \delta \right\}$$

is continuous on $(\mathcal{SM}_p(\Omega), \mathcal{AW}_p)$.

Let ρ be a law-invariant risk measure which we directly view as a functional from $\mathcal{P}_p(\mathbb{R})$ to \mathbb{R} . For $\mathbb{P} \in \mathcal{SM}_p(\Omega)$ and a random variable $Z: \Omega \rightarrow \mathbb{R}$ such that

$Z(\mathbb{P}) \in \mathcal{P}_p(\mathbb{R})$, we write $\rho^{\mathbb{P}}(Z) = \rho(Z(\mathbb{P}))$. A typical example of a law-invariant risk measure which satisfies $\rho(\mu) - \rho(\nu) \leq Ld^w(\mu, \nu)$ for some constant L depending on the p th moment of μ and ν is the optimised certainty equivalent, introduced to the mathematical finance community in Ben-Tal and Teboulle [17]. For a convex, increasing function $\ell: \mathbb{R} \rightarrow \mathbb{R}$ which is bounded from below and satisfies $\ell(x)/x \rightarrow \infty$ as $x \rightarrow \infty$, the optimised certainty equivalent is defined via

$$\rho^{\mathbb{P}}(Z) := \inf_{m \in \mathbb{R}} (\mathbb{E}_{\mathbb{P}}[\ell(Z - m)] + m) = \inf_{m \in \mathbb{R}} \left(\int \ell(x - m)(Z(\mathbb{P}))(dx) + m \right).$$

If $\ell'(x) \leq c(1 + |x|^{p-1})$, then it follows that the infimum over m can be taken in some compact set depending on the p th moments. Due to cash additivity of ρ , the following proposition has the same interpretation as Theorem 1.6.

Proposition 4.3 *Assume that $\rho: \mathcal{P}_p(\mathbb{R}) \rightarrow \mathbb{R}$ satisfies $\rho(\mu) - \rho(\nu) \leq Ld^w(\mu, \nu)$ for some constant L depending on the p -th moment of μ and ν . Then for every Lipschitz function $C: \Omega \rightarrow \mathbb{R}$, the mapping*

$$\mathbb{P} \mapsto \inf_{H \in \mathcal{H}_k} \rho^{\mathbb{P}} \left(C(X) - (H(X) \cdot X)_T \right)$$

is locally Lipschitz-continuous on $(\mathcal{SM}_p(\Omega), \mathcal{AW}_p)$.

Finally, let us point out that although it is not a convex risk measure, Value-at-Risk (VaR) would be another natural candidate to study continuity. However, as VaR is not continuous with respect to weak convergence, continuity of the mapping $\mathbb{P} \mapsto \inf\{m \in \mathbb{R} : \text{there is } H \in \mathcal{H}_k \text{ with } \text{VaR}^{\mathbb{P}}(C(X) - m - (H(X) \cdot X)_T) \leq 0\}$ fails already in a one-period model.

5 Final remarks

The first two remarks illustrate situations where using the usual Wasserstein distance does not work.

Remark 5.1 We note that convergence in the usual Wasserstein distance is not sufficient to obtain continuity in any of the problems we study in this paper. Consider a two-period market with

$$\begin{aligned} \mathbb{P}_n &= \frac{1}{4}(\delta_{(1/n,1)} + \delta_{(1/n,0)} + \delta_{(-1/n,0)} + \delta_{(-1/n,-1)}), \\ \mathbb{P} &= \frac{1}{4}(\delta_{(0,1)} + 2\delta_{(0,0)} + \delta_{(0,-1)}). \end{aligned}$$

Then \mathbb{P} and each \mathbb{P}_n satisfy the classical no-arbitrage condition, unlike the situation described in Fig. 1. While (\mathbb{P}_n) converges to \mathbb{P} in the usual Wasserstein distance,

one can verify that convergence in the nested distance does not hold. For example, in utility maximisation of the trivial claim $C \equiv 0$, we have

$$\sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}} \left[U \left(C(X) + (H(X) \cdot X)_T \right) \right] = U(0)$$

by Jensen's inequality (as X is a martingale under \mathbb{P}). For \mathbb{P}_n , taking the strategy H^* consisting of $H_0^* = 0$ and $H_1^*(x) = k \operatorname{sign}(x)$, one gets

$$\sup_{H \in \mathcal{H}_k} \mathbb{E}_{\mathbb{P}_n} \left[U \left(C(X) + (H(X) \cdot X)_T \right) \right] \geq \mathbb{E}_{\mathbb{P}_n} \left[U \left(C(X) + (H^*(X) \cdot X)_T \right) \right] \rightarrow U(k),$$

showing the lack of continuity.

Remark 5.2 As explained in the introduction, the objective in Theorem 1.5 can be seen as a relaxed version of the superhedging problem. The reason to consider this relaxation is not a technical simplification, but necessary to obtain continuity without further assumptions. Indeed, the problem of superhedging,

$$\inf \{ m \in \mathbb{R} : \text{there is } H \in \mathcal{H}_k \text{ such that } m + (H \cdot X)_T \geq C(X) \text{ } \mathbb{P}\text{-almost surely} \}$$

is not continuous in \mathbb{P} with respect to the adapted distance for any $k \in [0, \infty]$. In fact, this already happens in one period, where the adapted and the usual Wasserstein distances coincide. Consider a sequence (\mathbb{P}_n) of measures with full support which converge weakly to a measure \mathbb{P} . Then the superhedging price with respect to \mathbb{P}_n equals the concave envelope of C , while the superhedging price with respect to \mathbb{P} equals the concave envelope of C restricted to the support of \mathbb{P} . For a recent paper on this problem in one period, see the work of Obłój and Wiesel [45].

The final remark illustrates that uniformly bounded strategies are necessary.

Remark 5.3 Similarly as in Remark 5.2, the restriction to trading strategies in \mathcal{H}_k (i.e., uniformly bounded strategies) is also not merely a technical simplification. For example, in a one-period framework, the measures $\mathbb{P}^\varepsilon := (1 - \varepsilon)\delta_{(0,\varepsilon)} + \varepsilon\delta_{(0,-\varepsilon)}$ converge to $\mathbb{P} := \delta_{(0,0)}$ in every (adapted) Wasserstein distance. However, we have for small $\varepsilon > 0$ that

$$\inf_{H \in \mathcal{H}_\infty} \operatorname{AVaR}_\alpha^{\mathbb{P}^\varepsilon}((H \cdot X)_T) = -\infty, \quad \inf_{H \in \mathcal{H}_\infty} \operatorname{AVaR}_\alpha^{\mathbb{P}}((H \cdot X)_T) = 0,$$

where $\mathcal{H}_\infty := \bigcup_{k \in \mathbb{N}} \mathcal{H}_k$ is the set of all bounded trading strategies.

Acknowledgements All authors are grateful to the anonymous referees whose insightful comments had a significant impact on this article.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Acciaio, B., Backhoff-Veraguas, J., Zalashko, A.: Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization. *Stoch. Process. Appl.* **130**, 2918–2953 (2020).
2. Aldous, D.J.: Weak Convergence and General Theory of Processes. Unpublished monograph; Department of Statistics, University of California, Berkeley (1981). Available online at <https://www.stat.berkeley.edu/~aldous/Papers/weak-gtp.pdf>
3. Alfonsi, A., Corbetta, J., Jourdain, B.: Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *Int. J. Theor. Appl. Finance* **22**, 1950002-1–1950002-41 (2019)
4. Alibert, J.J., Bouchitte, G., Champion, T.: A new class of cost for optimal transport planning. *Eur. J. Appl. Math.* **30**, 1229–1263 (2019)
5. Avellaneda, M., Levy, A., Parás, A.: Pricing and hedging derivative securities in markets with uncertain volatilities. *Appl. Math. Finance* **2**, 73–88 (1995)
6. Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., Eder, M.: All adapted topologies are equal. Preprint (2019). Available online at <https://arxiv.org/abs/1905.00368>
7. Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., Wiesel, J.: Estimating processes in adapted Wasserstein distance. Preprint (2020). Available online at <https://arxiv.org/abs/2002.07261>
8. Backhoff-Veraguas, J., Beiglböck, M., Eder, M., Pichler, A.: Fundamental properties of process distances. *Stoch. Process. Appl.* (2020, forthcoming). <https://doi.org/10.1016/j.spa.2020.03.017>
9. Backhoff-Veraguas, J., Beiglböck, M., Huesmann, M., Källblad, S.: Martingale Benamou–Brenier: a probabilistic perspective. *Ann. Probab.* (2018, forthcoming). Available online at <https://www.e-publications.org/ims/submission/AOP/user/submissionFile/39384?confirm=ec3bf61d>
10. Backhoff-Veraguas, J., Beiglböck, M., Lin, Y., Zalashko, A.: Causal transport in discrete time and applications. *SIAM J. Optim.* **27**, 2528–2562 (2017)
11. Backhoff-Veraguas, J., Beiglböck, M., Pammer, G.: Existence, duality, and cyclical monotonicity for weak transport costs. *Calc. Var. Partial Differ. Equ.* **58**, 203-1–203-28 (2019)
12. Backhoff-Veraguas, J., Pammer, G.: Stability of martingale optimal transport and weak optimal transport. Preprint (2019). Available online at <https://arxiv.org/abs/1904.04171>
13. Becherer, D., Kentia, K.: Good deal hedging and valuation under combined uncertainty about drift and volatility. *Probab. Uncertain. Quant. Risk* **2**, 13-1–13-40 (2017)
14. Beiglböck, M., Cox, A., Huesmann, M.: The geometry of multi-marginal Skorokhod embedding. *Probab. Theory Relat. Fields* **176**, 1045–1096 (2020)
15. Beiglböck, M., Henry-Labordère, P., Penkner, F.: Model-independent bounds for option prices: a mass transport approach. *Finance Stoch.* **17**, 477–501 (2013)
16. Beiglböck, M., Juillet, N.: On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.* **44**, 42–106 (2016)
17. Ben-Tal, A., Teboulle, M.: An old-new concept of convex risk measures: the optimized certainty equivalent. *Math. Finance* **17**, 449–476 (2007)
18. Bion-Nadal, J., Talay, D.: On a Wasserstein-type distance between solutions to stochastic differential equations. *Ann. Appl. Probab.* **29**, 1609–1639 (2019)
19. Bouchard, B., Nutz, M.: Arbitrage and duality in nondominated discrete-time models. *Ann. Appl. Probab.* **25**, 823–859 (2015)
20. Campi, L., Laachir, I., Martini, C.: Change of numeraire in the two-marginals martingale transport problem. *Finance Stoch.* **21**, 471–486 (2017)
21. Cont, R.: Model uncertainty and its impact on the pricing of derivative instruments. *Math. Finance* **16**, 519–547 (2006)

22. Dolinsky, Y., Soner, H.M.: Martingale optimal transport and robust hedging in continuous time. *Probab. Theory Relat. Fields* **160**, 391–427 (2014)
23. Eder, M.: Compactness in adapted weak topologies. Preprint (2019). Available online at <https://arxiv.org/abs/1905.00856>
24. El Karoui, N., Jeanblanc-Picqué, M., Shreve, S.: Robustness of the Black and Scholes formula. *Math. Finance* **8**, 93–126 (1998)
25. 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)
26. Glanzer, M., Pflug, G.C., Pichler, A.: Incorporating statistical model error into the calculation of acceptability prices of contingent claims. *Math. Program.* **174**, 499–524 (2019)
27. Gozlan, N., Roberto, C., Samson, P.M., Shu, Y., Tetali, P.: Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. Henri Poincaré Probab. Stat.* **54**, 1667–1693 (2018)
28. Gozlan, N., Roberto, C., Samson, P.M., Tetali, P.: Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **273**, 3327–3405 (2017)
29. Hellwig, M.F.: Sequential decisions under uncertainty and the maximum theorem. *J. Math. Econ.* **25**, 443–464 (1996)
30. Herrmann, S., Muhle-Karbe, J.: Model uncertainty, recalibration, and the emergence of delta-vega hedging. *Finance Stoch.* **21**, 873–930 (2017)
31. Herrmann, S., Muhle-Karbe, J., Seifried, F.T.: Hedging with small uncertainty aversion. *Finance Stoch.* **21**, 1–64 (2017)
32. Hobson, D.: Robust hedging of the lookback option. *Finance Stoch.* **2**, 329–347 (1998)
33. Hobson, D.: Volatility misspecification, option pricing and superreplication via coupling. *Ann. Appl. Probab.* **8**, 193–205 (1998)
34. Hobson, D.: The Skorokhod embedding problem and model-independent bounds for option prices. In: Carmona, R., et al. (eds.) *Paris-Princeton Lectures on Mathematical Finance 2010*. Lecture Notes in Math., vol. 2003, pp. 267–318. Springer, Berlin (2011)
35. Hobson, D., Neuberger, A.: Robust bounds for forward start options. *Math. Finance* **22**, 31–56 (2012)
36. Karatzas, I., Shreve, S.: *Brownian Motion and Stochastic Calculus*. Springer, Berlin (1988)
37. Kardaras, C., Žitković, G.: Stability of the utility maximization problem with random endowment in incomplete markets. *Math. Finance* **21**, 313–333 (2011)
38. Lacker, D.: Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. Preprint (2018). Available online at <https://arxiv.org/abs/1805.03185>
39. Larsen, K.: Continuity of utility-maximization with respect to preferences. *Math. Finance* **19**, 237–250 (2009)
40. Larsen, K., Žitković, G.: Stability of utility-maximization in incomplete markets. *Stoch. Process. Appl.* **117**, 1642–1662 (2007)
41. Lassalle, R.: Causal transference plans and their Monge–Kantorovich problems. *Stoch. Anal. Appl.* **36**, 452–484 (2018)
42. Lyons, T.J.: Uncertain volatility and the risk-free synthesis of derivatives. *Appl. Math. Finance* **2**, 117–133 (1995)
43. Mocha, M., Westray, N.: The stability of the constrained utility maximization problem: a BSDE approach. *SIAM J. Financ. Math.* **4**, 117–150 (2013)
44. Oblój, J.: The Skorokhod embedding problem and its offspring. *Probab. Surv.* **1**, 321–390 (2004)
45. Oblój, J., Wiesel, J.: Statistical estimation of superhedging prices. Preprint (2018). Available online at <https://arxiv.org/abs/1807.04211>
46. Pflug, G.C.: Version-independence and nested distributions in multistage stochastic optimization. *SIAM J. Optim.* **20**, 1406–1420 (2009)
47. Pflug, G.C., Pichler, A.: A distance for multistage stochastic optimization models. *SIAM J. Optim.* **22**, 1–23 (2012)
48. Pflug, G.C., Pichler, A.: *Multistage Stochastic Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, Berlin (2014)
49. Pflug, G.C., Pichler, A.: From empirical observations to tree models for stochastic optimization: convergence properties. *SIAM J. Optim.* **26**, 1715–1740 (2016)
50. Pratelli, A.: On the equality between Monge’s infimum and Kantorovich’s minimum in optimal mass transportation. *Ann. Inst. Henri Poincaré Probab. Stat.* **43**, 1–13 (2007)
51. Prigent, J.L.: *Weak Convergence of Financial Markets*. Springer, Berlin (2003)
52. Rüschendorf, L.: The Wasserstein distance and approximation theorems. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **70**, 117–129 (1985)

53. Weston, K.: Stability of utility maximisation in nonequivalent markets. *Finance Stoch.* **20**, 511–541 (2016)
54. Wiesel, J.: Continuity of the martingale optimal transport problem on the real line. Preprint (2019). Available online at <https://arxiv.org/abs/1905.04574>
55. Yamada, T., Watanabe, S.: On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11**, 155–167 (1971)