NIKOLAY LYSENKO

MAXIMIZATION OF FUNCTIONALS DEPENDING ON THE TERMINAL VALUE AND THE RUNNING MAXIMUM OF A MARTINGALE: A MASS TRANSPORT APPROACH

It is known that the Azéma-Yor solution to the Skorokhod embedding problem maximizes the law of the running maximum of a uniformly integrable martingale with a given terminal value distribution. Recently this optimality property has been generalized to expectations of certain bivariate cost functions depending on the terminal value and the running maximum.

In this paper we give an extension of this result to another class of functions. In particular, we study a class of cost functions for which the corresponding optimal embeddings are not Azéma-Yor. The suggested approach is quite straightforward modulo basic facts of the Monge-Kantorovich mass transportation theory. Loosely speaking, the joint distribution of the running maximum and the terminal value in the Azéma-Yor embedding is concentrated on the graph of a monotone function, and we show that this fact follows from the cyclical monotonicity criterion for solutions to the Monge-Kantorovich problem.

1. INTRODUCTION

Let W_t be the canonical Wiener process and let μ be a centered probability measure on $\mathbb{R}$, i.e., a probability measure such that $\int_{\mathbb{R}} x \, \mu(dx) = 0$ and $\int_{\mathbb{R}_{\geq 0}} x \, \mu(dx) < +\infty$. Throughout μ will be treated as the law of the terminal value of a uniformly integrable martingale started at 0.

One of the key concepts considered in this paper is the Skorokhod embedding (see [8]). The following formulation can be found in [4].

Problem 1 (Skorokhod). For a given measure μ, find a uniformly integrable stopping time τ (with respect to the filtration generated by the canonical Wiener process) such that $W_\tau \sim \mu$.

Several constructions for τ are known, and some of them possess different optimality properties. Detailed surveys are presented in works of Obłój [6] and Hobson [4]. It is sufficient for our purposes to mention only the Azéma-Yor embedding [1], because it plays a significant role in the setup of the paper.

Hobson and Klimmek [5] have considered maximization of expected values of cost functions depending on the terminal value and the running maximum. They have shown that the Azéma-Yor embedding solves this problem within the class of all uniformly integrable martingales for a broad class of functions. Such functions $F(w, s)$ are assumed to be continuous and differentiable with respect to the running maximum F_s. In addition, $F_s/(s - w)$ is assumed to be nondecreasing in w. Actually, paper [5] contains a lot of

other interesting results, but the aforementioned fact motivates our study of optimality of the Azéma-Yor solution for bivariate cost functions.

For additional information about the Azéma-Yor embedding, one can consult Section 3.1 of [5] and the cited comprehensive surveys [6] and [4]. However, the only property of the Azéma-Yor embedding used in this paper is as follows.

Theorem 1 (Azéma and Yor; [1])**.** *Consider the class of uniformly integrable martingales with a given distribution μ at the terminal moment T. For each positive l, the probability that the running maximum of a process from this class is greater than or equal to l does not exceed $\mu([\beta_\mu^{-1}(l); +\infty))$, where β_μ^{-1} is the inverse barycenter function of μ, i.e., the function inverse to the function $\beta_\mu(k) = \mathbb{E}_\mu[x | x \geq k]$. Moreover, these values are attained for the Azéma-Yor embedding for all $l \geq 0$ simultaneously.*

The paper of Hobson and Klimmek uses the results obtained by Rogers [7]. These results are applied in our paper as well. The following necessary and sufficient condition for a measure on $\mathbb{R} \times \mathbb{R}$ to be the joint distribution of the terminal value and the running maximum of a uniformly integrable martingale is of particular importance.

Theorem 2 (Rogers; based on Corollary 2.4 of [7])**.** *A measure π on $\mathbb{R} \times \mathbb{R}$ is the joint distribution of the final value and the running maximum of a uniformly integrable martingale starting at 0 if and only if the following assumptions (1.1)-(1.4) are satisfied:*

$$(1.1) \qquad \int_{\mathbb{R} \times \mathbb{R}} |x| \, \pi(dx, dy) < +\infty,$$

$$(1.2) \qquad \int_{\mathbb{R} \times \mathbb{R}} x \, \pi(dx, dy) = 0,$$

$$(1.3) \qquad \operatorname{supp} \pi \subset M := (\mathbb{R} \times \mathbb{R}_{\geq 0}) \cap \{(x, y) : y - x \geq 0\},$$

$$(1.4) \qquad \mathbb{E}_\pi[x \,|\, y \geq s] \geq s, \quad \forall s \geq 0.$$

The original formulation of Rogers contains one more assumption, namely, that

$$\mathbb{E}_\pi[x \,|\, y \geq s]$$

is nondecreasing with respect to s. This assumption, however, can be omitted, because it follows immediately from (1.3) and (1.4). Indeed, let us assume the existence of s_1 and s_2 such that $s_1 < s_2$, but $\mathbb{E}_\pi[x \,|\, y \geq s_1] > \mathbb{E}_\pi[x \,|\, y \geq s_2]$. Note that

$$\mathbb{E}_\pi[x \,|\, y \geq s_1] = \lambda \mathbb{E}_\pi[x \,|\, s_1 \leq y < s_2] + (1 - \lambda)\mathbb{E}_\pi[x \,|\, y \geq s_2]$$

for some $0 < \lambda \leq 1$. Hence $\mathbb{E}_\pi[x \,|\, s_1 \leq y < s_2] \geq \mathbb{E}_\pi[x \,|\, y \geq s_1] > \mathbb{E}_\pi[x \,|\, y \geq s_2] \geq s_2$, which is prohibited by (1.3).

An application of the Monge-Kantorovich theory can be considered as the third key component of our approach. From the rigorous point of view, the considered problem is not covered by the optimal transport theory, but, fortunately, a reduction to the mass transportation problem is possible after some preparations. The idea to treat the maximization problem for the expectation of a bivariate function as a transportation problem with some restrictions has appeared in discussions with Alexander Kolesnikov. More information about the optimal transportation theory can be found in [9] and [2], and constrained transportation problems are studied, for instance, in [10].

In order to formulate the main problem of this article, it is convenient to introduce some auxiliary notions.

Definition 1. Given a centered probability measure μ and a positive number T, the set of μ-*admissible* (or simply admissible) processes is defined to be the set of all uniformly integrable martingales starting at 0 at $t = 0$ that are distributed according to μ at the terminal time $t = T$.

Definition 2. A bivariate function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is called *serrated* if, for every $w_0 \in \mathbb{R}$, the univariate function $F(w_0, s)$ is increasing for $s \in (-\infty; a_{w_0})$ and decreasing for $s \in (a_{w_0}; +\infty)$, where $a_{w_0} \in \mathbb{R} \cup \{-\infty, +\infty\}$ (the value plus or minus infinity means that $F(w_0, s)$ is decreasing or increasing, respectively). The set $R_F = \{(w_0, a_{w_0})\colon w_0 \in \mathbb{R}\} \subset \mathbb{R} \times (\mathbb{R} \cup \{-\infty, +\infty\})$ is called the *ridge* of F.

Recall another useful definition that comes from applications of the optimal transportation theory.

Definition 3. A bivariate function $F\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is called *supermodular* if it has the following property:

$$(1.5) \quad \forall w_1, w_2, s_1, s_2 \in \mathbb{R} \quad \begin{cases} w_1 < w_2 \\ s_1 < s_2 \end{cases} \Rightarrow F(w_1, s_1) + F(w_2, s_2) \geq F(w_1, s_2) + F(w_2, s_1).$$

If the rightmost inequality in (1.5) is strict, the function is called *strictly supermodular*.

Remark 1. A strictly supermodular serrated function has a nondecreasing ridge.

Assume that we are given a continuous strictly supermodular serrated function $F(w, s)$, a centered probability measure μ, and a positive number T. The problem studied in Section 2 is this: find a μ-admissible process M_t maximizing the expected value of $F(M_T, S_T)$, where $S_T = \max_{0 \leq t \leq T} M_t$, among of all μ-admissible processes. Although this case is narrow, it enables one to see the core idea clearly. However, the suggested proof is not the simplest one.

If $F(w, s)$ is a strictly supermodular function and $G(w, s)$ is a (not necessary strictly) supermodular function, then their sum $H(w, s) = F(w, s) + G(w, s)$ is again strictly supermodular. This obvious fact leads to the idea how a conclusion might be drawn in some cases of discontinuous functions. Such cases are considered in Subsection 3.1, and a construction analogous to the aforementioned result of [5] is represented in Subsection 3.2.

We emphasize that in Section 3 neither differentiability nor even continuity of F is required. Furthermore, examples of proper discontinuous strictly supermodular serrated functions are discussed in Subsection 3.1. Thus, the results obtained in our article are not covered by [5].

Finally, let us note that the problem under consideration has a natural interpretation in terms of model-independent finance. Suppose that an exotic derivative with payout function $F(M_T, S_T)$ is going to be underwritten at the moment $t = 0$, where M_t is the price process of the underlying asset, and again $S_T = \max_{0 \leq t \leq T} M_t$. The problem is to determine the no-arbitrage price of the derivative.

What data is available? The idea to retrieve a market prognosis from the current quotations of liquid European call options goes back to Breeden and Litzenberger [3]. Their construction enables one to reconstruct the measure μ under the assumption of the presence of the continuum of liquid calls with the same maturity T. Another piece of information is that the underlying asset price process must be a uniformly integrable martingale started at some fixed price. If someone has a model for the underlying asset price, it must be calibrated to this μ in order to get the exact value. Nevertheless, in model-independent finance all the models that are consistent with available information are treated as realistic, regardless their peculiarities. Thereby the following question arises: what are the upper and the lower bounds for the derivative price? In other

words, the seller is interested not in a unique price, but in a range of no-arbitrage prices. It is easy to see that this financial problem agrees with the former probabilistic problem.

2. The basic approach: the result for continuous strictly supermodular functions

The lemma below can be considered as an analogue of variational theorems on c-cyclical monotonicity of solutions to the Monge-Kantorovich problem. This is the reason why the approach of this paper is called a mass transport approach.

Lemma 1. *Let $F(w, s)$ be a strictly supermodular continuous function. Assume we are given a centered probability measure μ and a positive T. Then every μ-admissible process maximizing expectation of $F(M_T, S_T)$ has a nondecreasing joint distribution of the terminal value and the running maximum π. The latter means that there are no intervals $I_1, I_2 \subset \mathbb{R}_1$ and $J_1, J_2 \subset \mathbb{R}_2$ (subscripts below $\mathbb{R}$ indicate coordinate axis) with the properties*

$$(2.1) \qquad\qquad \sup I_1 \leq \inf I_2, \ \inf J_1 \geq \sup J_2,$$

$$(2.2) \qquad\qquad \pi(I_1 \times J_1) > 0, \ \pi(I_2 \times J_2) > 0.$$

Proof. Assume that the joint distribution of the terminal value and the running maximum of an admissible process π is not nondecreasing. Let us construct a competitor $\bar{\pi}$ on $\mathbb{R} \times \mathbb{R}$, which is the joint distribution generated by an admissible process giving a better value to the cost functional.

Since π is not nondecreasing, there are intervals I_1, I_2, J_1 and J_2 for which (2.1) and (2.2) are true.

For the sake of technical purposes, it is appropriate to trim them. The set $I_1 \times J_1$ lies inside a compact set, so there is a point $(x_1, y_1) \in I_1 \times J_1$ such that every ε-neighborhood ($\varepsilon > 0$) of this point has a non-zero π-mass. There exists a point $(x_2, y_2) \in I_2 \times J_2$ with the same property. Consider a compact set C containing both $I_1 \times J_1$ and $I_2 \times J_2$ and then choose δ small enough to ensure that for all $(a_1, b_1), (a_2, b_2) \in C$ we have

$$(2.3) \qquad (|(a_1, b_1) - (a_2, b_2)| < \delta) \Rightarrow (|F(a_1, b_1) - F(a_2, b_2)| < Q),$$

where $Q = \frac{1}{4}(F(x_1, y_2) + F(x_2, y_1) - F(x_1, y_1) - F(x_2, y_2))$. This can be done because of the uniform continuity property. Now let $U_1 = (x_1 - \frac{\delta}{2}; x_1 + \frac{\delta}{2})$ and $V_1 = (y_1 - \frac{\delta}{2}; y_1 + \frac{\delta}{2})$. If $y_2 \neq 0$, let U_2 and V_2 be intervals such that $\operatorname{diam} U_2 = \operatorname{diam} V_2 < \min(\delta, 2y_2)$ and (x_2, y_2) is the center of the square $U_2 \times V_2$. Else (i.e. if $y_2 = 0$) let $U_2 = (x_2 - \frac{\delta}{2}; x_2 + \frac{\delta}{2})$ and $V_2 = [0; \frac{\delta}{2})$ (it does not matter that V_2 is not open). The last conditions are helpful when proving that (1.3) remains valid. We recall that the aim is to trim initial intervals, so if $U_1 \times V_1 \subset I_1 \times J_1$ and $U_2 \times V_2 \subset I_2 \times J_2$ are not satisfied, δ must be chosen little enough to satisfy these conditions.

Now all the preparations are done, and it is possible to construct $\bar{\pi}$ iteratively.

Denote $\min(\pi(U_1 \times V_1), \pi(U_2 \times V_2))$ as m. Then we construct π_1 as a measure on $\mathbb{R} \times \mathbb{R}$ defined in the following way:

(1) it coincides with π on subsets of $(\mathbb{R} \times \mathbb{R}) \setminus (U_1 \times V_1)$,
(2) for each $\mathcal{S} \subset U_1 \times V_1$ one has $\pi_1(\mathcal{S}) = \frac{\pi(U_1 \times V_1) - m}{\pi(U_1 \times V_1)} \pi(\mathcal{S})$,
(3) it is extended additively to other measurable subsets.

Note that π_1 is not a probability measure, because its mass equals $1 - m$. One can treat π_1 as the result of subtraction of the measure σ_1 from π.

Let σ_2 be a measure of mass m with support contained inside of $U_2 \times V_1$ such that, for each neighborhood $U \subset U_2$, $\sigma_2(U \times V_1)$ is proportional to $\pi(U \times V_2)$ and, for each neighborhood $V \subset V_1$, $\sigma_2(U_2 \times V)$ is proportional to $\pi(U_1 \times V)$ (note that σ_2 is not uniquely defined). Let us define π_2 as follows:

(1) π_2 coincides with π_1 on subsets of $(\mathbb{R} \times \mathbb{R}) \setminus (U_2 \times V_1)$,
(2) for each $\mathcal{S} \subset U_2 \times V_1$ $\pi_2(\mathcal{S}) = \pi(\mathcal{S}) + \sigma_2(\mathcal{S})$,
(3) it is extended additively to other measurable subsets.

Here the first step is done. The second (and the final) step is to reassign the mass m from $U_2 \times V_2$ to $U_1 \times V_2$ such that the projection on the first coordinate equals μ. In order to do this we define π_3 in the following way:

(1) π_3 coincides with π_2 on subsets of $(\mathbb{R} \times \mathbb{R}) \setminus (U_2 \times V_2)$,
(2) for each $\mathcal{S} \subset U_2 \times V_2$ one has $\pi_3(\mathcal{S}) = \frac{\pi(U_1 \times V_1) - m}{\pi(U_1 \times V_1)} \pi(\mathcal{S})$,
(3) it is extended additively to other measurable subsets.

The measure π_3 can be interpreted as the result of subtraction of the measure σ_3.

Let σ_4 be a measure of mass m supported inside of $U_1 \times V_2$ such that for each neighborhood $U \subset U_1$, $\sigma_4(U \times V_2)$ is proportional to $\pi(U \times V_1)$ and for each neighborhood $V \subset V_2$, $\sigma_4(U_1 \times V)$ is proportional to $\pi(U_2 \times V)$ (again σ_4 is not uniquely defined). Finally, define π_4 as a measure with the properties

(1) π_4 coincides with π_3 on subsets of $(\mathbb{R} \times \mathbb{R}) \setminus (U_1 \times V_2)$,
(2) for each $\mathcal{S} \subset U_1 \times V_2$ one has $\pi_4(\mathcal{S}) = \pi(\mathcal{S}) + \sigma_4(\mathcal{S})$,
(3) π_4 is extended additively to other measurable subsets

It remains to prove that $\bar{\pi} = \pi_4$ is the desired competitor.

First we observe that $\mathrm{Pr}_1 \pi_4 = \mu$, because $\mathrm{Pr}_1 \sigma_1 = \mathrm{Pr}_1 \sigma_4$ and $\mathrm{Pr}_1 \sigma_2 = \mathrm{Pr}_1 \sigma_3$. This fact implies that (1.1) and (1.2) are satisfied.

Assumption (1.3) is satisfied too, because $(x_2, y_2) \in M = (\mathbb{R} \times \mathbb{R}_{\geq 0}) \cap \{(x, y) : y - x \geq 0\}$, and this provides that $U_1 \times V_2$ and $U_2 \times V_1$ are subsets of the required set M; it can be clearly seen from the construction.

Further, (1.4) holds true, because $\forall s \in \mathbb{R}_{\geq 0}$ the three statements are true:

$$(2.4) \qquad \mathbb{P}_\pi[x \in U_1 \,|\, y \geq s] \geq \mathbb{P}_{\pi_4}[x \in U_1 \,|\, y \geq s],$$

$$(2.5) \qquad \mathbb{P}_\pi[x \in U_2 \,|\, y \geq s] \leq \mathbb{P}_{\pi_4}[x \in U_2 \,|\, y \geq s],$$

$$(2.6) \qquad \mathbb{P}_\pi[x \notin U_1 \text{ and } x \notin U_2 \,|\, y \geq s] = \mathbb{P}_{\pi_4}[x \notin U_1 \text{ and } x \notin U_2 \,|\, y \geq s].$$

Putting (2.4)-(2.6) together yields the desired result that $\forall s \in \mathbb{R}_{\geq 0}$

$$(2.7) \qquad \mathbb{E}_{\pi_4}[x \,|\, y \geq s] \geq \mathbb{E}_\pi[x \,|\, y \geq s] \geq s.$$

Thus, π_4 is the joint distribution of the terminal value and the running maximum of a uniformly integrable martingale starting from 0 and, moreover, this process is μ-admissible, because the law of the terminal value is μ.

Finally, we have to prove that the process corresponding to π_4 is better. To this end we prove the following inequality:

$$(2.8) \qquad \int F(w, s)\, (\sigma_2 + \sigma_4)(dw, ds) - \int F(w, s)\, (\sigma_1 + \sigma_3)(dw, ds) > 0.$$

The desired result follows from the following line of computations:

$$\int F(w, s)(\sigma_2 + \sigma_4)(dw, ds) - \int F(w, s)(\sigma_1 + \sigma_3)(dw, ds)$$
$$= \int \big(F(x_1, y_2) + (F(w, s) - F(x_1, y_2))\big) \sigma_2(dw, ds)$$

$$+ \int \big(F(x_2, y_1) + (F(w, s) - F(x_2, y_1)) \big) \sigma_4(dw, ds)$$

$$- \int \big(F(x_2, y_2) + (F(w, s) - F(x_2, y_2)) \big) \sigma_3(dw, ds)$$

$$- \int \big(F(x_1, y_1) + (F(w, s) - F(x_1, y_1)) \big) \sigma_1(dw, ds)$$

$$= \int F(x_1, y_2) \sigma_2(dw, ds) + \int (F(w, s) - F(x_1, y_2)) \sigma_2(dw, ds)$$

$$+ \int F(x_2, y_1) \sigma_4(dw, ds) + \int (F(w, s) - F(x_2, y_1)) \sigma_4(dw, ds)$$

$$- \int F(x_2, y_2) \sigma_3(dw, ds) - \int (F(w, s) - F(x_2, y_2)) \sigma_3(dw, ds)$$

$$- \int F(x_1, y_1) \sigma_1(dw, ds) - \int (F(w, s) - F(x_1, y_1)) \sigma_1(dw, ds)$$

$$= (F(x_1, y_2) + F(x_2, y_1) - F(x_2, y_2) - F(x_1, y_1)) m$$

$$+ \int \big(F(w, s) - F(x_1, y_2) \big) \sigma_2(dw, ds) + \int \big(F(w, s) - F(x_2, y_1) \big) \sigma_4(dw, ds)$$

$$- \int \big(F(w, s) - F(x_2, y_2) \big) \sigma_3(dw, ds) - \int \big(F(w, s) - F(x_1, y_1) \big) \sigma_1(dw, ds)$$

$$> (F(x_1, y_2) + F(x_2, y_1) - F(x_2, y_2) - F(x_1, y_1)) m$$

$$- \Big(\int Q \sigma_2(dw, ds) + \int Q \sigma_3(dw, ds) + \int Q \sigma_4(dw, ds) + \int Q \sigma_1(dw, ds) \Big) = 0.$$

This completes the proof. □

With the help of Lemma 1 we establish several results on optimal admissible processes. We start with two extremal types of strictly supermodular serrated functions and get a generalizing statement at the end of this section.

Theorem 3. *Let $F(w, s)$ be a continuous strictly supermodular serrated function and let its ridge R_F be the set $\{(w, +\infty) : w \in \mathbb{R}\}$. Then the functional $\mathbb{E}[F(w, s)]$ considered on the set of all μ-admissible processes is maximized by the Azéma-Yor embedding.*

Proof. By Lemma 1, the joint distribution generated by an optimal process must have a nondecreasing support. A μ-admissible process can not generate a distribution with non-decreasing support which has a non-zero mass higher than the graph of the barycenter function β_μ. It follows from the Azéma-Yor embedding optimality property for nondecreasing functions depending only on the running maximum (see Theorem 1). Thereby all the mass is placed not higher than the barycenter function graph. Since for every w_0 the univariate functions $F(w_0, s)$ is assumed to be increasing, the Azéma-Yor embedding is the optimal admissible process.

In case of non-atomic μ this can be seen from the following line of computations, where U stands for the joint distribution of an arbitrary μ-admissible process, U_w stands for the conditional distribution given fixed w, and AY stands for the joint distribution of the Azéma-Yor embedding:

$$\mathbb{E}_U[F(M_T, S_T)] = \int \Big(\int F(w, s) U_w(ds) \Big) \mu(dw)$$

$$\leq \int F(w, \beta_\mu(w)) \mu(dw) = \mathbb{E}_{AY}[F(M_T, S_T)].$$

The inequality is an equality if U is equivalent to AY.

If μ has atoms, $\int F(w, \beta_\mu(w))\mu(dw)$ must be replaced by its analogue, where instead of discontinuous $\beta_\mu(w)$ there is a continuous line with vertical intervals and masses of atoms are properly distributed along these intervals. $\qquad\square$

Remark 2. An example of a function satisfying assumptions of Theorem 3 is $F(w, s) = (\arctan w + 2)s$.

Proposition 1. *Let $F(w, s)$ be a measurable serrated function such that its ridge R_F is the set $\{(w, -\infty) : w \in \mathbb{R}\}$. Then the functional $\mathbb{E}[F(w, s)]$ considered on the set of all μ-admissible processes is maximized by a the pure jump process, i.e., a process which is constant on the time interval $[0; T/2)$, jumps to a value of a random variable with law μ at the $t = T/2$, and equals another constant on the time interval $[T/2; T]$.*

Proof. The described process generates the joint distribution with support contained in the boundary of M (see (1.3) for the definition of M). Since (1.3) prohibits placing a non-zero mass lower than this boundary and since for every w_0 the univariate functions $F(w_0, s)$ is assumed to be decreasing, this is an optimal admissible process.

Again, this can be seen from the following line of computations, where U stands for the joint distribution of an arbitrary μ-admissible process, U_w stands for the conditional distribution given fixed w, and PJ stands for the joint distribution of the pure jump process:

$$\mathbb{E}_U[F(M_T, S_T)] = \int \left(\int F(w, s) U_w(ds) \right) \mu(dw)$$
$$\leq \int F(w, \max(0, w)) \mu(dw) = \mathbb{E}_{PJ}[F(M_T, S_T)].$$

The inequality is an equality if U is equivalent to PJ. $\qquad\square$

Remark 3. An example of a function satisfying assumptions of Proposition 1 is $F(w, s) = -|w|s$.

Remark 4. In Proposition 1, the μ-admissibility can be replaced by a less restrictive condition, since the martingale property is not used. However, the optimal process (to be precise, at least one of optimal processes) is still a martingale.

We will see that, roughly speaking, a general optimal process has the joint distribution of the final value and the running maximum in an "intermediate position" between the two discussed extremal distributions. For the sake of clarity below it is assumed that μ has no atoms (the proof of Theorem 3 contains a precept how to generalize a further reasoning to the atomic case).

Lemma 2. *Every nondecreasing function f taking values in the intersection of M with the closed subgraph of the barycenter function of μ induces a measure that is the joint distribution of the terminal value and the running maximum of a μ-admissible process.*

Proof. Consider the measure $f_\#\mu$. The properties (1.1) and (1.2) are immediate. The hypothesis of the lemma ensures that (1.3) is satisfied. Finally, (1.4) follows from the inequality

$$(2.9) \qquad\qquad \mathbb{E}_{f_\#\mu}[x|y \geq s] \geq \mathbb{E}_{AY}[x|y \geq s] \geq s,$$

where the subscript AY refers to the distribution generated by the Azéma-Yor embedding (note that $\mathbb{E}_{AY}[x|y \geq s] = s$ in any standard situation). $\qquad\square$

Theorem 4. *Let $F(w, s)$ be a continuous strictly supermodular serrated function. Then among all μ-admissible processes the functional $\mathbb{E}[F(w, s)]$ is maximized by the process*

with the property that the corresponding joint distribution of the final value and the running maximum is $g_\#\mu$, where g is defined by

$$(2.10) \qquad\qquad g(w) = \min(\beta_\mu(w), \max(a_w, 0, w)),$$

β_μ is the same as in Theorem 1, and a_w is the same as in Definition 2.

Proof. The described process exists and is μ-admissible, because the maximum of two monotone functions is a monotone function and the minimum of two monotone functions is again a monotone function, so g is covered by Lemma 2. Explicit description of this process can be found in the proof of Theorem 2.2 of [7].

The process is optimal due to the following reasoning. Lemma 1 implies that the optimal joint distribution of the terminal value and the running maximum of an admissible process must have monotone support. A joint distribution with a monotone support cannot place non-zero mass beyond the closed subgraph of the barycenter function of μ, since this is prohibited by Theorem 1. Also condition (1.3) shows that every joint distribution of the terminal value and the running maximum of an admissible process cannot place non-zero mass beyond M. Combining this together, we obtain that for each fixed w within the region where all mass must be placed the best possible point is the point that belongs to the graph of g. Thus, the process described in the formulation is optimal.

As before, the last statement can be seen from the following line of computations, where U stands for the joint distribution of an arbitrary μ-admissible process and U_w stands for the conditional distribution given fixed w:

$$\mathbb{E}_U[F(M_T, S_T)] = \int \left(\int F(w, s) U_w(ds) \right) \mu(dw)$$

$$\leq \int F(w, g(w)) \mu(dw) = \mathbb{E}_{g_\#\mu}[F(M_T, S_T)].$$

The inequality is an equality if U is equivalent to $g_\#\mu$. $\qquad\square$

Remark 5. The following function satisfies the assumptions of Theorem 4:

$$F(w, s) = (\arctan w + 2)s - 4 \operatorname{Ind}_{\{(w,s) \,:\, s \geq R(w)\}}(s - R(w)),$$

where R is an increasing function. The ridge of $F(w, s)$ is the graph of $s = R(w)$.

Remark 6. In the above theorems, when speaking about the maximization of the functional $\mathbb{E}[F(M_T, S_T)]$, we mean that there are no processes that allow achieving higher values. If a μ-admissible process U is such that one of the theorems asserts that it is optimal for the corresponding F, but $\mathbb{E}_U[F(M_T, S_T)] = -\infty$, then, of course, all other μ-admissible processes are also optimal for the same problem.

3. Generalizations

3.1. **Introducing discontinuity.** We have already mentioned that the above approach to the proofs of Theorems 3 and 4 is not the simplest one. Instead of it one can apply directly the optimal transportation theory. The standard assumption ensuring the existence of a solution to the Monge-Kantorovich problem is the lower semicontinuity of the cost function. In addition, the solutions to the Monge-Kantorovich problem admit the so-called cyclical monotonicity property, which can be established under assumptions that are at least not stronger than the assumption of lower semicontinuity (see [2] for references to recent results in this direction).

Let us give some basic definitions.

Problem 2. Suppose that $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a measurable function (often called a cost function). Suppose also that μ and ν are Borel measures on $\mathbb{R}$ and $\Pi(\mu, \nu)$ is the set of all Borel measures π on $\mathbb{R} \times \mathbb{R}$ such that $\Pr_1 \pi = \mu$ and $\Pr_2 \pi = \nu$. *The Monge-Kantorovich problem is this:*

$$(3.1) \qquad \int_{\mathbb{R} \times \mathbb{R}} c(x, y)\pi(dx, dy) \to \min_{\pi \in \Pi(\mu, \nu)}.$$

Actually, above definition is not general. It is possible to consider $X \times Y$ for measure spaces X and Y instead of $\mathbb{R} \times \mathbb{R}$ or to introduce more than two axes (also known as marginals).

Definition 4. Suppose that $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a measurable function. A subset $\Gamma \subset \mathbb{R} \times \mathbb{R}$ is called *c-cyclically monotone* if for every non-empty sequence of its elements $(x_1, y_1), ..., (x_n, y_n)$ it is true that

$$(3.2) \quad c(x_1, y_1) + c(x_2, y_2) + ... + c(x_n, y_n) \leq c(x_1, y_n) + c(x_2, y_1) + ... + c(x_n, y_{n-1}).$$

Below we present the approach mentioned in the Introduction.

Theorem 5. *In the formulations of Theorems 3 and 4, the continuity can be replaced by the upper semicontinuity.*

Proof. Suppose that the joint distribution π of the final value and the running maximum of a μ-admissible process is given. Set $\nu = \Pr_2 \pi$. Let $\overline{\pi}$ be a solution to the Monge-Kantorovich problem with marginals μ and ν and the cost function $-F(w, s)$. Here the sign is reversed, because the initial problem is a maximization problem, but the Monge-Kantorovich problem is a minimization problem. A solution to the described Monge-Kantorovich problem exists, because the cost function $-F$ is lower semicontinuous.

It can be easily verified that the *c*-cyclical monotonicity (i.e. $-F(w, s)$-cyclical monotonicity) implies that $\overline{\pi}$ is concentrated on the graph of a monotone function T (this is a standard observation coming from the optimal transportation theory):

$$(3.3) \qquad \overline{\pi}\Big(\{(x, T(x)) : \ x \in \mathbb{R}\}\Big) = 1.$$

Let us show that $\overline{\pi}$ is the joint distribution of an admissible process. Conditions (1.1) and (1.2) are satisfied automatically. Further, $\operatorname{supp} \overline{\pi} \subset \mathbb{R} \times \mathbb{R}_{\geq 0}$, because $\operatorname{supp} \nu \subset \mathbb{R}_{\geq 0}$, meanwhile $\operatorname{supp} \overline{\pi} \subset \{(x, y) : y - x \geq 0\}$ due to the following reasoning. For the initial joint distribution π it is true that

$$(3.4) \qquad \forall k > 0 \ \nu([0; k]) = \pi(\mathbb{R} \times [0; k]) = \pi((-\infty; k] \times [0; k]) \leq \mu((-\infty; k]).$$

Combining (3.4) with (3.3) yields that (1.3) is checked. Finally, (1.4) holds true, because $\forall k > 0 \ \nu([k; +\infty)) \leq \nu_{AY}([k; +\infty))$, where subscript AY refers to the joint distribution generated by the Azéma-Yor embedding, so $\mathbb{E}_{\overline{\pi}}[x|y \geq s] \geq \mathbb{E}_{AY}[x|y \geq s] \geq s$.

The construction ensures that the process that generates $\overline{\pi}$ is not worse than the initial process. Since its joint distribution is supported on the graph of a monotone function, the further proof can follow the arguments of Theorems 3 and 4. $\qquad\square$

However, as far as the author knows, in the Monge-Kantorovich theory there is no analogous result for upper semicontinuity instead of lower semicontinuity. Fortunately, the approach of Section 2 is applicable to some lower semicontinuous functions.

Proposition 2. *Assume that $F(w, s)$ is a continuous supermodular serrated function and $G_{a,b}(w, s) = \operatorname{Ind}_{\{(w,s) \,:\, w \geq a,\, s \geq b\}}$, where (a, b) is in the ridge of $F(w, s)$. Then Theorems 3 and 4 hold true for $(F + G_{a,b})(w, s)$, which is a lower semicontinuous supermodular serrated function with the same ridge.*

Proof. Note that $(F+G_{a,b})(w,s)$ possesses declared properties due to its construction. It is sufficient to prove only a generalization of Lemma 1, because the remaining reasoning from the theorems is still applicable.

Consider four measures σ_1, σ_2, σ_3 and σ_4 which are employed in the proof of Lemma 1 and have equal masses m. The desired result follows from the inequality

$$(3.5) \qquad \int G_{a,b}(w,s)(\sigma_2 + \sigma_4 - \sigma_1 - \sigma_3)(dw, ds) \geq 0.$$

The above inequality can be proved by a simple analysis of possible configurations of supports of this four measures relatively to $\{(w,s) : w \geq a, s \geq b\}$. This analysis is based on applications of the equalities $\mathrm{Pr}_1 \sigma_1 = \mathrm{Pr}_1 \sigma_4$, $\mathrm{Pr}_2 \sigma_1 = \mathrm{Pr}_2 \sigma_4$, $\mathrm{Pr}_1 \sigma_2 = \mathrm{Pr}_1 \sigma_3$, and $\mathrm{Pr}_2 \sigma_3 = \mathrm{Pr}_2 \sigma_4$. It is just a plain geometry and arithmetic. $\qquad \square$

Remark 7. Of course, the previous proposition is true not only for $\mathrm{Ind}_{\{(w,s)\,:\,w \geq a,\, s \geq b\}}$, but also remains valid for $\mathrm{Ind}_{\{(w,s)\,:\,w \leq a,\, s \leq b\}}$, $-\mathrm{Ind}_{\{(w,s)\,:\,w > a,\, s < b\}}$, $-\mathrm{Ind}_{\{(w,s)\,:\,w < a,\, s > b\}}$, and positive linear combinations of such functions.

3.2. **Towards the result of Hobson and Klimmek.** Everywhere above the second marginal of the joint distribution was fixed. This is consistent with the spirit of the Monge-Kantorovich theory, but the price for this is a quite restrictive requirement of supermodularity. If $F(w,s)$ is continuously differentiable with respect to s, its supermodularity is equivalent to the assumption that all the univariate functions $F_s(w, s_0)$ with arbitrarily fixed s_0 are increasing. In [5] a weaker assumption was suggested: why not suppose that only $F_s(w, s_0)/(s_0 - w)$ is monotonic? Actually, the reader can observe that in the proof of Lemma 1 in (2.7) condition (1.4) is satisfied in a non-optimal manner and for some s the inequality $\mathbb{E}_{\bar{\pi}}[x \,|\, y \geq s] \geq s$ is strict, which means that an additional mass may be lifted up. This is refined in the current subsection.

To make the idea clear, do the following. Consider the function $F(w,s)$ with a continuous partial derivative $F_s(w,s)$. Moreover, we assume that for each $s_0 > 0$ the function $F_s(w, s_0)/(s_0 - w)$ is increasing for $w < s_0$. Suppose that π is the joint distribution of a martingale with terminal value law μ and that there are points (w_1, s_1) and (w_2, s_2), with $w_1 < w_2$ and $s_1 > s_2$, such that π has atoms with masses at least m at these points. Remove the mass $(s_2 - w_2)/(s_2 - w_1)m$ from (w_1, s_1) to (w_1, s_2) and remove the mass m from (w_2, s_2) to the point (w_2, s_1) and the vertical open interval between (w_2, s_2) and (w_2, s_1) in such a way that for every q, $s_2 < q \leq s_1$, the mass lifted not lower than q is $((q-w_1)(s_2-w_2))/((q-w_2)(s_2-w_1))m$. It can be seen clearly that the coefficients were chosen in order to provide that (1.4) remains true. Because the properties (1.1)-(1.3) here are trivial, the new measure is a joint distribution as well. Moreover, this measure is a better competitor, since the gain from the reassignment is

(3.6)
$$\Delta := \int_{(w_2,s_2)}^{(w_2,s_1)} F_s(w_2, \xi) \frac{(\xi - w_1)(s_2 - w_2)}{(\xi - w_2)(s_2 - w_1)} m \, d\xi - \int_{(w_1,s_2)}^{(w_1,s_1)} F_s(w_1, \xi) \frac{(s_2 - w_2)}{(s_2 - w_1)} m \, d\xi.$$

The assumptions about $F(w,s)$ imply that the latter is positive.

The proof of the following statement is omitted because up to some inessential technicalities it is reduced to the proof of an appropriate analogue of Lemma 1 and follows the same line.

Theorem 6. *The results of Theorems 3 and 4 hold true under the following assumptions imposed on F instead of strict supermodularity: $F(w,s)$ is continuous, there exists a continuous partial derivative $F_s(w,s)$, and, in addition, for each $s_0 > 0$ the function $F_s(w, s_0)/(s_0 - w)$ is increasing for $w < s_0$.*

Remark 8. The case of nondecreasing $F_s(w, s_0)/(s_0 - w)$ for every $s_0 > 0$ and $w < s_0$ is considered in [5], but it is not covered here.

References

1. J. Azéma and M. Yor, *Le probléme de Skorokhod: Compléments á "Une solution simple au probléme de Skorokhod"*, In Séminaire de Probabilités, XIII (Univ. Strasbourg, Strasbourg), 1979.

2. V. I. Bogachev and A. V. Kolesnikov, *The Monge–Kantorovich problem: achievements, connections, and perspectives*, Uspekhi Mat. Nauk **67** (2012), no. 5(407), 3–110.

3. D. T. Breeden and R. H. Litzenberger, *Prices of state-contingent claims implicit in options prices*, J. Business **51** (1978), 621–651.

4. D. G. Hobson, *The Skorokhod embedding problem and model-independent bounds for option prices*, Volume 2003 of Lecture Notes in Math., 267–318, Springer, Berlin, 2011.

5. D. G. Hobson and M. Klimmek, *Maximizing functionals of the maximum in the Skorokhod embedding problem and an application to variance swaps*, The Annals of Applied Probability **23** (2013), no. 5, 2020–2052.

6. J. Obłój, *The Skorokhod embedding problem and its offspring* Probab. Surv. **1** (2004), 321–390 (electronic).

7. L. C. G. Rogers, *The joint law of the maximum and terminal value of a martingale*, Probab. Theory Related Fields **95** (1993), 451–466.

8. A. V. Skorokhod, *Studies in the Theory of Random Processes*, Addison-Wesley, Reading, MA, 1965.

9. C. Villani, *Topics in optimal transportation*, Grad. Stud. Math. **58**, Amer. Math. Soc., Providence, RI, 2003, xvi + 370 pp.

10. D. A. Zaev, *On the Monge-Kantorovich problem with additional linear constraints*, Mat. Zametki **98(5)** (2015), 664–683.

National Research University Higher School of Economics, Faculty of Mathematics, 119048 Moscow, Usacheva str., 6