Monotone Sharpe ratios and related measures of investment performance

Mikhail Zhitlukhin

Abstract We introduce a new measure of performance of investment strategies, the monotone Sharpe ratio. We study its properties, establish a connection with coherent risk measures, and obtain an efficient representation for using in applications.

1 Introduction

This paper concerns the problem of evaluation of performance of investment strategies. By performance, in a broad sense, we mean a numerical quantity which characterizes how good the return rate of a strategy is, so that an investor typically wants to find a strategy with high performance.

Apparently, the most well-known performance measure is the Sharpe ratio, the ratio of the expectation of a future return, adjusted by a risk-free rate or another benchmark, to its standard deviation. It was introduced by William F. Sharpe in the 1966 paper [23], a more modern look can be also found in [24]. The Sharpe ratio is based on the Markowitz mean-variance paradigm [14], which assumes that investors need to care only about the mean rate of return of assets and the variance of the rate of return: then in order to find an investment strategy with the smallest risk (identified with the variance of return) for a given desired expected return, one just needs to find a strategy with the best Sharpe ratio and diversify appropriately between this strategy and the risk-free asset (see a brief review in Section 2 below). Despite its simplicity, as viewed from today's economic science, the Markowitz portfolio theory was a major breakthrough in mathematical finance. Even today, more than 65 years later, analysts still routinely compute Sharpe ratios of investment portfolios and use it, among other tools, to evaluate performance.

In the present paper we look at this theory in a new way, and establish connections with much more recent developments. The main part of the material of

Steklov Mathematical Institute of Russian Academy of Sciences, 8 Gubkina st., Moscow 119991, Russia. e-mail: E-mail: mikhailzh@mi.ras.ru.

the paper developed from a well-known observation that variance is not identical to risk: roughly speaking, one has to distinguish between “variance above mean” (which is good) and “variance below mean” (which is bad). In particular, the Sharpe ratio lacks the property of monotonicity, i.e. there might exist an investment strategy which always yields a return higher than another strategy, but has a smaller Sharpe ratio. The original goal of this work was to study a modification of the Sharpe ratio, which makes it monotone. Some preliminary results were presented in [27, 28]. It turned out, that the modified Sharpe ratio posses interesting properties and is tightly connected to the theory of risk measures. The study of them is the subject of this paper.

The modification of the Sharpe ratio we consider, which we call the monotone Sharpe ratio, is defined as the maximum of the Sharpe ratios of all probability distributions that are dominated by the distribution of the return of some given investment strategy. In this paper we work only with *ex ante* performance measure, i.e. assume that probability distributions of returns are known or can be modeled, and one needs to evaluate their performance; we leave aside the question how to construct appropriate models and calibrate them from data.

The theory we develop focuses on two aspects: on one hand, to place the new performance measure on a modern theoretical foundation, and, on the other hand, take into account issues arising in applications, like a possibility of fast computation and good properties of numerical results. Regarding the former aspect, we can mention the paper of Cherny and Madan [3], who studied performance measures by an axiomatic approach. The abstract theory of performance measures they proposed is tightly related to the theory of convex and coherent risk measures, which has been a major breakthrough in the mathematical finance in the past two decades. We show that the monotone Sharpe ratio satisfies those axioms, which allows to apply results from the risk measures theory to it through the framework of Cherny and Madan. Also we establish a connection with more recently developed objects, the so-called buffered probabilities, first introduced by Rockafellar and Royset in [21] and now gaining popularity in applications involving optimization under uncertainty. Roughly speaking, they are “nice” alternatives to optimization criteria involving probabilities of adverse events, and lead to solutions of optimization problems which have better mathematical properties compared to those when standard probabilities are used. One of main implications of our results is that the portfolio selection problem with the monotone Sharpe ratio is equivalent to minimization of the buffered probability of loss.

Addressing the second aspect mentioned above, our main result here is a representation of the monotone Sharpe ratio as a solution of some convex optimization problem, which gives a computationally efficient way to evaluate it. Representations of various functionals in such a way are well-known in the literature on convex optimization. For example, in the context of finance, we can mention the famous result of Rockafellar and Uryasev [22] about the representation of the conditional value at risk. That paper also provides a good explanation why such a representation is useful in applications (we also give a brief account on that below).

Our representation also turns out to be useful in stochastic control problems related to maximization of the Sharpe ratio in dynamic trading. Those problems are known in the literature as examples of stochastic control problems where the Bellman optimality principle cannot be directly applied. With our theory, we are able to find the optimal strategies in a shorter and simpler way, compared to the results previously known in the literature.

Finally, we would like to mention, that in the literature a large number of performance measures have been studied. See for example papers [10, 5, 4] providing more than a hundred examples of them addressing various aspects of evaluation of quality of investment strategies. We believe that due to both the theoretical foundation and the convenience for applications, the monotone Sharpe ratio is a valuable contribution to the field.

The paper is organized as follows. In Section 2 we introduce the monotone Sharpe ratio and study its basic properties which make it a reasonable performance measure. There we also prove one of the central results, the representation as a solution of a convex optimization problem. In Section 3, we generalize the concept of the buffered probability and establish a connection with the monotone Sharpe ratio, as well as show how it can be used in portfolio selection problems. Section 4 contains applications to dynamic problems.

2 The monotone Sharpe ratio

2.1 Introduction: Markowitz portfolio optimization and the Sharpe ratio

Consider a one-period market model, where an investor wants to distribute her initial capital between $n+1$ assets: one riskless asset and n risky assets. Assume that the risky assets yield return R_i, $i = 1,\dots,n$, so that \$1 invested "today" in asset i turns into $\$(1+R_i)$ "tomorrow"; the rates of return R_i are random variables with known distributions, such that $R_i > -1$ with probability 1 (no bankrupts happen). The rate of return of the riskless asset is constant, $R_0 = r > -1$. We always assume that the probability distributions of R_i are known and given, and, for example, do not consider the question how to estimate them from past data. In other words, we always work with *ex ante* performance measures (see [24]).

An investment portfolio of the investor is identified with a vector $x \in \mathbb{R}^{n+1}$, where x_i is the proportion of the initial capital invested in asset i. In particular, $\sum_i x_i = 1$. Some coordinates x_i may be negative, which is interpreted as short sales ($i = 1,\dots,n$) or loans ($i = 0$). It is easy to see that the total return of the portfolio is $R_x = \langle x, R\rangle := \sum_i x_i R_i$.

The Markowitz model prescribes the investor to choose the optimal investment portfolio in the following way: she should decide what expected return $\mathrm{E}R_x$ she wants to achieve, and then find the portfolio x which minimizes the variance of the return $\operatorname{Var} R_x$. This leads to the quadratic optimization problem:

$$\begin{aligned} \text{minimize} \quad & \operatorname{Var} R_x \text{ over } x \in \mathbb{R}^{n+1} \\ \text{subject to} \quad & \mathrm{E} R_x = \mu \\ & \textstyle\sum_i x_i = 1. \end{aligned} \tag{1}$$

Under mild conditions on the joint distribution of R_i, there exists a unique solution x^*, which can be easily written explicitly in terms of the covariance matrix and the vector of expected returns of R_i (the formula can be found in any textbook on the subject, see, for example, Chapter 2.4 in [18]).

It turns out that points $(\sigma_{x^*}, \mu_{x^*})$, where $\sigma_{x^*} = \sqrt{\operatorname{Var} R_{x^*}}$, $\mu_{x_*} = \mathrm{E} R_{x^*}$ correspond to the optimal portfolios for all possible expected returns $\mu \in [r, \infty)$, lie on the straight line in the plane (σ, μ), called the efficient frontier. This is the set of portfolios the investor should choose from – any portfolio below this line is inferior to some efficient portfolio (i.e. has the same expected return but larger variance), and there are no portfolios above the efficient frontier.

The slope of the efficient frontier is equal to the Sharpe ratio of any efficient portfolio containing a non-zero amount of risky assets (those portfolios have the same Sharpe ratio). Recall that the Sharpe ratio of return R is defined as the ratio of the expected return adjusted by the risk-free rate to its standard deviation

$$S(R) = \frac{\mathrm{E}(R - r)}{\sqrt{\operatorname{Var} R}}.$$

In particular, to solve problem (1), it is enough to find some efficient portfolio $\hat{x}$, and then any other efficient portfolio can be constructed by a combination of the riskless portfolio $x_0 = (1, 0, \dots, 0)$ and $\hat{x}$, i.e $x^* = (1 - \lambda) x_0 + \lambda \hat{x}$, where $\lambda \in [0, +\infty)$ is chosen to satisfy $\mathrm{E} R_{x^*} = \mu \geq r$. This is basically the statement of the Mutual Fund Theorem. Thus, the Sharpe ratio can be considered as a measure of performance of an investment portfolio and an investor is interested in finding a portfolio with the highest performance. In practice, broad market indices can be considered as quite close to efficient portfolios.

The main part of the material in this paper grew from the observation that the Sharpe ratio is not monotone: for two random variables X, Y the inequality $X \leq Y$ a.s. does not imply the same inequality between their Sharpe ratios, i.e. that $S(X) \leq S(Y)$. Here is an example: let X have the normal distribution with mean 1 and variance 1 and $Y = X \wedge 1$; obviously, $S(X) = 1$ but one can compute that $S(Y) > 1$. From the point of view of the portfolio selection problem, this fact means that it is possible to increase the Sharpe ratio by disposing part of the return (or consuming it). This doesn't agree well with the common sense interpretation of efficiency. Therefore, one may want to look for a replacement of the Sharpe ratio, which will not have such a non-natural property.

In this paper we'll use the following simple idea: if it is possible to increase the Sharpe ratio by disposing a part of the return, let's define the new performance measure as the maximum Sharpe ratio that can be achieve by such a disposal. Namely, define the new functional by

$$\mathbb{S}(X) = \sup_{C\geq 0} S(X-C),$$

where the supremum is over all non-negative random variables C (defined on the same probability space as X), which represent the disposed return. In the rest of this section, we'll study such functionals and how they can be used in portfolio selection problems. We'll work in a more general setting and consider not only the ratio of expected return to standard deviation of return but also ratios of expected return to deviations in L^p. The corresponding definitions will be given below.

2.2 *The definition of the monotone Sharpe ratio and its representation*

In this section we'll treat random variables as returns of some investment strategies, unless other is stated. That is, large values are good, small values are bad. Without loss of generality, we'll assume that the risk-free rate is zero, otherwise one can replace a return X with $X - r$, and all the results will remain valid.

First we give the definition of a deviation measure in L^p, $p \in [1,\infty)$, which will be used in the denominator of the Sharpe ratio instead of the standard deviation (the latter one is a particular case for $p=2$). Everywhere below, $\|\cdot\|_p$ denotes the norm in L^p, i.e. $\|X\|_p = (\mathrm{E}|X|^p)^{\frac{1}{p}}$.

Definition 1. We define the L^p-deviation of a random variable $X \in \mathrm{L}^p$ as

$$\sigma_p(X) = \min_{c\in\mathbb{R}} \|X-c\|_p.$$

In the particular case $p=2$, as is well-known, $\sigma_2(X)$ is the standard deviation, and the minimizer is $c^* = \mathrm{E}X$. For $p=1$, the minimizer $c^* = \mathrm{med}(X)$, the median of the distribution of X, so that $\sigma_1(X)$ is the absolute deviation from the median. It is possible to use other deviation measures to define the monotone Sharpe ratio, for example $\|X-\mathrm{E}X\|_p$, but the definition given above seems to be the most convenient for our purposes.

Observe that σ_p obviously satisfies the following properties, which will be used later: (a) it is sublinear; (b) it is uniformly continuous on L^p; (c) $\sigma_p(X)=0$ if and only if X is a constant a.s.; (d) for any σ-algebra $\mathscr{G} \subset \mathscr{F}$, where $\mathscr{F}$ is the original σ-algebra on the underlying probability space for X, we have $\sigma_p(\mathrm{E}(X \mid \mathscr{G})) \leq \sigma_p(X)$; (e) if X and Y have the same distributions, then $\sigma_p(X) = \sigma_p(Y)$.

Definition 2. The *monotone Sharpe ratio in* L^p of a random variable $X \in \mathrm{L}^p$ is defined by

$$\mathbb{S}_{\mathrm{p}}(X) = \sup_{Y\leq X} \frac{\mathrm{E}Y}{\sigma_p(Y)}, \tag{2}$$

where the supremum is over all $Y \in \mathrm{L}^p$ such that $Y \le X$ a.s. For $X = 0$ a.s. we set by definition $\mathbb{S}_\mathrm{p}(0) = 0$.

One can easily see that if $p > 1$, then $\mathbb{S}_\mathrm{p}(X)$ assumes value in $[0,\infty]$. Indeed, if $\mathrm{E}X \le 0$, then $\mathbb{S}_\mathrm{p}(X) = 0$ as it is possible to take $Y \le X$ with arbitrarily large L^p-deviation keeping $\mathrm{E}Y$ bounded. On the other hand, if $X \ge 0$ a.s. and $\mathrm{P}(X > 0) \ne 0$, then $\mathbb{S}_\mathrm{p}(X) = +\infty$ as one can consider $Y_\varepsilon = \varepsilon\,\mathrm{I}(X \ge \varepsilon)$ with $\varepsilon \to 0$ for which $\mathrm{E}Y_\varepsilon/\sigma_p(Y_\varepsilon) \to \infty$.

Thus, the main case of interest will be when $\mathrm{E}X > 0$ and $\mathrm{P}(X < 0) \ne 0$; then $0 < \mathbb{S}_\mathrm{p}(X) < \infty$. For this case, the following theorem provides the representation of $\mathbb{S}_\mathrm{p}$ as a solution of some convex optimization problem.

Theorem 1. *Suppose* $X \in \mathrm{L}^p$ *and* $\mathrm{E}(X) > 0$, $\mathrm{P}(X < 0) \ne 0$. *Then the following representations of the monotone Sharpe ratio are valid.*
1) For $p \in (1,\infty)$ *with* q *such that* $\frac{1}{p} + \frac{1}{q} = 1$*:*

$$(\mathbb{S}_p(X))^q = \max_{a,b\in\mathbb{R}} \Big\{ b - \mathrm{E}\Big(\tfrac{q-1}{q^p} \big|(aX+b)_+ - q\big|^p + (aX+b)_+ \Big) \Big\}. \tag{3}$$

2) For $p = 1,2$*:*

$$\frac{1}{1 + (\mathbb{S}_\mathrm{p}(X,r))^p} = \min_{c\in\mathbb{R}} \mathrm{E}(1 - cX)_+^p. \tag{4}$$

The main point about this theorem is that it allows to reduce the problem of computing $\mathbb{S}_\mathrm{p}$ as the supremum over the set of random variables to the optimization problem with one or two real parameters and the convex objective function. The latter problem is much easier than the former one, since there exist efficient algorithms of numerical convex optimization. This gives a convenient way to compute $\mathbb{S}_\mathrm{p}(X)$ (though only numerically, unlike the standard Sharpe ratio). We'll also see that the representation is useful for establishing some theoretical results about $\mathbb{S}_\mathrm{p}$.

For the proof, we need the following auxiliary lemma.

Lemma 1. *Suppose* $X \in \mathrm{L}^p$, $p \in [1,\infty)$*, and* q *is such that* $\frac{1}{p} + \frac{1}{q} = 1$. *Then*

$$\sigma_p(X) = \max\{\mathrm{E}(RX) \mid R \in \mathrm{L}^q,\ \mathrm{E}R = 0, \|R\|_q \le 1\}.$$

Proof. Suppose $\sigma_p(X) = \|X - c^*\|_p$. By Hölder's inequality, for any $R \in \mathrm{L}^q$ with $\mathrm{E}R = 0$ and $\|R\|_q \le 1$ we have

$$\mathrm{E}(RX) = \mathrm{E}(R(X - c^*)) \le \|R\|_q \cdot \|X - c^*\|_p \le \|X - c^*\|_p.$$

On the other hand, the two inequalities turn into equalities for

$$R^* = \frac{\mathrm{sgn}(X - c^*) \cdot |X - c^*|^{p-1}}{\|X - c^*\|_p^{p-1}}$$

and R^* satisfies the above constraints.

Proof (Proof of Theorem 1). Without loss of generality, assume $\mathrm{E}X = 1$. First we're going to show that $\mathbb{S}_\mathrm{p}$ can be represented through the following optimization problem:

$$\mathbb{S}_\mathrm{p}(X) = \inf_{R\in \mathrm{L}^q}\{\|R\|_q \mid R \le 1 \text{ a.s.},\ \mathrm{E}R = 0,\ \mathrm{E}(RX) = 1\}. \tag{5}$$

In (2), introduce the new variables: $c = (\mathrm{E}Y)^{-1} \in \mathbb{R}$ and $Z = cY \in \mathrm{L}^p$. Then

$$\frac{1}{\mathbb{S}_p(X)} = \inf_{\substack{Z\in \mathrm{L}^p\\ c\in\mathbb{R}}}\{\sigma_p(Z) \ \mid Z \le cX,\ \mathrm{E}Z = 1\}.$$

Consider the dual of the optimization problem in the RHS (see the Appendix for a brief overview of duality methods in optimization). Define the dual objective function $g\colon \mathrm{L}^q_+ \times \mathbb{R} \to \mathbb{R}$ by

$$g(u,v) = \inf_{\substack{Z\in \mathrm{L}^p\\ c\in\mathbb{R}}}\{\sigma_p(Z) + \mathrm{E}(u(Z - cX)) - v(\mathrm{E}Z - 1)\}.$$

The dual problem consists in maximizing $g(u,v)$ over all $u \in \mathrm{L}^q_+$, $v \in \mathbb{R}$. We want to show that the strong duality takes place, i.e. that the values of the primal and the dual problems are equal:

$$\frac{1}{\mathbb{S}_p(X)} = \sup_{\substack{u\in \mathrm{L}^q_+\\ v\in\mathbb{R}}} g(u,v).$$

To verify the sufficient condition for the strong duality from Theorem 7, introduce the optimal value function $\phi\colon \mathrm{L}^p \times \mathbb{R} \to [-\infty,\infty)$

$$\phi(a,b) = \inf_{\substack{Z\in \mathrm{L}^p\\ c\in\mathbb{R}}}\{\sigma_p(Z) \ \mid Z - cX \le a,\ \mathrm{E}Z - 1 = b\}$$

(obviously, $(\mathbb{S}_\mathrm{p}(X))^{-1} = \phi(0,0)$). Observe that if a pair (Z_1,c_1) satisfies the constraints in $\phi(a_1,b_1)$ then the pair (Z_2,c_2) with

$$c_2 = c_1 + b_2 - b_1 + \mathrm{E}(a_1 - a_2), \quad Z_2 = Z_1 + a_2 - a_1 + (c_2 - c_1)X,$$

satisfies the constraints in $\phi(a_2,b_2)$. Clearly, $\|Z_1 - Z_2\|_p + |c_1 - c_2| = O(\|a_1 - a_2\|_p + |b_1 - b_2|)$, which implies that $\phi(a,b)$ is continuous, so the strong duality holds.

Let us now transform the dual problem. It is obvious that if $\mathrm{E}(uX) \neq 0$, then $g(u,v) = -\infty$ (minimize over c). For u such that $\mathrm{E}(uX) = 0$, using the dual representation of $\sigma_p(X)$, we can write

$$g(u,v) = \inf_{Z\in \mathrm{L}^p}\sup_{R\in\mathscr{R}} \mathrm{E}(Z(R + u - v) + v) \qquad \text{if } \mathrm{E}(uX) = 0,$$

where $\mathscr{R} = \{R \in \mathrm{L}^q : \mathrm{E}R = 0,\ \|R\|_q \le 1\}$ is the dual set for σ_p from Lemma 1. Observe that the set $\mathscr{R}$ is compact in the weak-$*$ topology by the Banach-Alaoglu theorem. Consequently, by the minimax theorem (see Theorem 8), the supremum and infimum can be swapped. Then it is easy to see that $g(u,v) > -\infty$ only if there exists $R \in \mathscr{R}$ such that $R + u - v = 0$ a.s., and in this case $g(u,v) = v$. Therefore, the dual problem can be written as follows:

$$\begin{aligned}\frac{1}{\mathbb{S}_{\mathrm{p}}(X)} &= \sup_{\substack{u \in \mathrm{L}^q \\ v \in \mathbb{R}}} \{v \mid u \geq 0 \text{ a.s., } \mathrm{E}(uX) = 0,\ v - u \in \mathscr{R}\} \\ &= \sup_{R \in \mathscr{R}} \{\mathrm{E}(RX) \mid R \leq \mathrm{E}(RX) \text{ a.s.}\} \\ &= \sup_{R \in \mathrm{L}^q} \{\mathrm{E}(RX) \mid R \leq \mathrm{E}(RX) \text{ a.s., } \mathrm{E}R = 0,\ \|R\|_q \leq 1\},\end{aligned}$$

where in the second equality we used that if $v - u = R \in \mathscr{R}$, then the second constraint imply that $v = \mathrm{E}(RX)$ since it is assumed that $\mathrm{E}X = 1$. Now by changing the variable R to $R/\mathrm{E}(RX)$ in the right-hand side, we obtain representation (5).

From (5), it is obvious that for $p > 1$

$$(\mathbb{S}_{\mathrm{p}}(X))^q = \inf_{R \in \mathrm{L}^q} \{\mathrm{E}|R|^q \mid R \leq 1 \text{ a.s., } \mathrm{E}R = 0,\ \mathrm{E}(RX) = 1\}. \tag{6}$$

We'll now consider the optimization problem dual to this one. Denote its optimal value function by $\phi\colon \mathrm{L}^q \times \mathbb{R} \times \mathbb{R} \to \overline{\mathbb{R}}$. It will be more convenient to change the optimization variable R here by $1 - R$ (which clearly doesn't change the value of ϕ), so that

$$\phi(a,b,c) = \inf_{R \in \mathrm{L}^q} \{\mathrm{E}|R-1|^q \mid R \geq a \text{ a.s., } \mathrm{E}R = 1 + b,\ \mathrm{E}(RX) = c\}.$$

Let us show that ϕ is continuous at zero. Denote by $C(a,b,c) \subset \mathrm{L}^q$ the set of $R \in \mathrm{L}^q$ satisfying the constraints of the problem. It will be enough to show that if $\|a\|_q, |b|, |c|$ are sufficiently small then for any $R \in C(0,0,0)$ there exists $\widetilde{R} \in C(a,b,c)$ such that $\|R - \widetilde{R}\|_q \leq (\|R\|_q + K)(\|a\|_q + |b| + |c|)$ and vice versa. Here K is some fixed constant.

Since $\mathrm{P}(X < 0) \neq 0$, there exists $\xi \in \mathrm{L}^\infty$ such that $\xi \geq 0$ a.s. and $\mathrm{E}(\xi X) = -1$. If $R \in C(0,0,0)$, then one can take the required $\widetilde{R} \in C(a,b,c)$ in the form

$$\widetilde{R} = \begin{cases} a + \lambda_1 R + \lambda_2 \xi, & \text{if } \mathrm{E}(aX) \geq 0, \\ a + \mu_1 R + \mu_2, & \text{if } \mathrm{E}(aX) < 0, \end{cases}$$

where the non-negative constants $\lambda_1, \lambda_2, \mu_1, \mu_2$ can be easily found from the constraint $\widetilde{R} \in C(a,b,c)$, and it turns out that $\lambda_1, \mu_1 = 1 + O(\|a\|_q + |b| + |c|)$ and $\lambda_2, \mu_2 = O(\|a\|_q + |b| + |c|)$. If $R \in C(a,b,c)$, then take

$$\widetilde{R} = \begin{cases} \lambda_1 (R - a + \lambda_2 \xi), & \text{if } c \geq \mathrm{E}(aX), \\ \mu_1 (R - a + \mu_2), & \text{if } c < \mathrm{E}(aX), \end{cases}$$

with λ_i, μ_i making $\widetilde{R} \in C(0,0,0)$.

Thus, the strong duality holds in (6) and we have

$$\mathbb{S}_{\mathrm{p}}(X) = \sup_{\substack{u\in \mathrm{L}^q_+\\ v,w\in\mathbb{R}}} g(u,v,w) \tag{7}$$

with the dual objective function $g\colon \mathrm{L}^q_+ \times \mathbb{R}\times\mathbb{R} \to \overline{\mathbb{R}}$

$$\begin{aligned} g(u,v,w) &= \inf_{R\in \mathrm{L}^q} \mathrm{E}(|R|^q + R(u+v+wX) - u - w)\\ &= -\mathrm{E}\Big(\tfrac{q-1}{q^p}|u+v+wX|^p + u + w\Big), \end{aligned}$$

where the second inequality is obtained by choosing R which minimizes the expression under the expectation for every random outcome.

Observe that for any fixed $v,w\in\mathbb{R}$ the optimal $u^* = u^*(v,w)$ in (7) can be found explicitly: $u^* = (v+wX+q)_-$. Then by straightforward algebraic transformation we obtain (3).

For $p=2$, from (3) we get

$$(\mathbb{S}_2(X))^2 = \max_{a,b\in\mathbb{R}} \Big\{ b - \frac{1}{4}\mathrm{E}(aX+b)_+^2 - 1\Big\}$$

It is easy to see that it is enough to maximize only over $b\geq 0$. Maximizing over b and introducing the variable $c = -\frac{a}{b}$, we obtain representation (4) for $p=2$.

To obtain representation (4) for $p=1$, let's again consider problem (5). Similarly to (7) (the only change will be to use $\|R\|_q$ instead of $\mathrm{E}|R|^q$), we can obtain that

$$\mathbb{S}_1(X) = \sup_{\substack{u\in \mathrm{L}^\infty_+\\ v,w\in\mathbb{R}}} g(u,v,w),$$

where now we denote

$$g(u,v,w) = \inf_{R\in\mathrm{L}^\infty} \{\|R\|_\infty + \mathrm{E}(R(u+v+wX)-u) - w\}.$$

Observe that a necessary condition for $g(u,v,w) > -\infty$ is that $\mathrm{E}|u+v+wX| \leq 1$: otherwise take $\widetilde{R} = c(\mathrm{I}(u+v+wX\leq 0) - \mathrm{I}(u+v+wX>0))$ and let $c\to\infty$. Under this condition we have $g(u,v,w) = -\mathrm{E}u - w$ since from Hölder's inequality $|\mathrm{E}((\alpha+v+wX)R)| \leq \|R\|_\infty$ and therefore the infimum in g is attained at $R=0$ a.s. Consequently, the dual problem becomes

$$\mathbb{S}_1(X) = -\inf_{\substack{u\in \mathrm{L}^\infty\\ v,w\in\mathbb{R}}} \{\mathrm{E}u + w \mid u\geq 0 \text{ a.s., } \mathrm{E}|u+v+wX|\leq 1\}. \tag{8}$$

Observe that the value of the infimum is non-positive, and so it is enough to restrict the values of w to $\mathbb{R}_-$ only. Let's fix $v\in\mathbb{R}$, $w\in\mathbb{R}_-$ and find the optimal $u^* = u^*(v,w)$. Clearly, whenever $v+wX(\omega)\geq 0$, it's optimal to take $u^*(\omega)=0$. Whenever $v+wX(\omega)<0$, we should have $u^*(\omega)\leq |v+wX(\omega)|$, so that $u(\omega)+$

$v+wX(\omega)\le 0$ (otherwise, the choice $u^*(\omega)=|v+wX(\omega)|$ will be better). Thus for the optimal u^*

$$\mathrm{E}\,|u^*+v+wX| = \mathrm{E}\,|v+wX| - \mathrm{E}\,u^*.$$

In particular, for the optimal u^* the inequality in the second constraint in (8) should be satisfied as the equality, since otherwise it would be possible to find a smaller u^*. Observe that if $\mathrm{E}(v+wX)_+>1$, then no $u\in \mathrm{L}^\infty$ exists which satisfies the constraint of the problem. On the other hand, if $\mathrm{E}(v+wX)_+\le 1$ then at least one such u exists. Consequently, problem (8) can be rewritten as follows:

$$-\mathbb{S}_1(X) = \inf_{v\in\mathbb{R}, w\in\mathbb{R}_-} \{\mathrm{E}\,|v+wX| + w - 1 \mid \mathrm{E}(v+wX)_+ \le 1\}.$$

Clearly, $\mathrm{E}\,|v^*+w^*X|\le 0$ for the optimal pair (v^*,w^*), so the constraint should be satisfied as the equality (otherwise multiply both v,w by $1/\mathrm{E}\,|v+wX)_+$, which will decrease the value of the objective function). By a straightforward transformation, we get

$$1+\mathbb{S}_1(X) = \sup_{v\in\mathbb{R}, w\in\mathbb{R}_-} \{v \mid \mathrm{E}(v+wX)_+ = 1\}$$

and introducing the new variable $c=w/v$, we obtain representation (3).

2.3 Basic properties

Theorem 2. *For any $p\in[1,\infty)$, the monotone Shape ratio in L^p satisfies the following properties.*

(a) (Quasi-concavity) For any $c\in\mathbb{R}$, the set $\{X\in \mathrm{L}^p : \mathbb{S}_\mathrm{p}(X)\ge c\}$ is convex.
(b) (Scaling invariance) $\mathbb{S}_\mathrm{p}(\lambda X)=\mathbb{S}_\mathrm{p}(X)$ for any real $\lambda>0$.
(c) (Law invariance) If X and Y have the same distribution, then $\mathbb{S}_\mathrm{p}(X)=\mathbb{S}_\mathrm{p}(Y)$.
(d) (2nd order monotonicity) If X dominates Y in the second stochastic order, then $\mathbb{S}_\mathrm{p}(X)\ge\mathbb{S}_\mathrm{p}(Y)$.
(e) (Continuity) $\mathbb{S}_\mathrm{p}(X)$ is continuous with respect to L^p-norm at any X such that $\mathrm{E}X>0$ and $\mathrm{P}(X<0)\ne 0$.

Before proving this theorem, let us briefly discuss the properties in the context of the portfolio selection problem.

The quasi-concavity implies that the monotone Sharpe ratio favors portfolio diversification: if $\mathbb{S}_\mathrm{p}(X)\ge c$ and $\mathbb{S}_\mathrm{p}(Y)\ge c$, then $\mathbb{S}_\mathrm{p}(\lambda X+(1-\lambda)Y)\ge c$ for any $\lambda\in[0,1]$, where $\lambda X+(1-\lambda)Y$ can be thought of as diversification between portfolios with returns X and Y. Note that the property of quasi-concavity is weaker than concavity; it's not difficult to provide an example showing that the monotone Sharpe ratio is not concave.

The scaling invariance can be interpreted as that the monotone Sharpe ratio cannot be changed by leveraging a portfolio (in the same way as the standard Sharpe ratio). Namely, suppose $X=R_x$, where $R_x=\langle x,R\rangle$ is the return of portfolio $x\in\mathbb{R}^{n+1}$ (as in Section 2.1), $\sum_i x_i=1$. Consider a leveraged portfolio $\widetilde{x}$ with $\widetilde{x}_i=\lambda x_i$, $i\ge 1$

and $\widetilde{x}_0 = 1-\sum \widetilde{x}_i$, i.e. a portfolio which is obtained from x by proportionally scaling all the risky positions. Then it's easy to see that $R_{\widetilde{x}} = \lambda R_x$, and so $\mathbb{S}_{\rm p}(R_x) = \mathbb{S}_{\rm p}(R_{\widetilde{x}})$.

Law invariance, obviously, states that we are able to evaluate the performance knowing only the distribution of the return. The interpretation of the continuity property is also clear.

The 2nd order monotonicity means that $\mathbb{S}_{\rm p}$ is consistent with preferences of risk-averse investors. Recall that it is said that the distribution of a random variable X dominates the distribution of Y in the 2nd stochastic order, which we denote by $X \succcurlyeq Y$, if $\mathrm{E}U(X) \ge \mathrm{E}U(Y)$ for any increasing concave function U such that $\mathrm{E}U(X)$ and $\mathrm{E}U(Y)$ exist. Such a function U can be interpreted as a utility function, and then the 2nd order stochastic dominance means that X is preferred to Y by any risk averse investor.

Regarding the properties from Theorem 2, let us also mention the paper [3], which studies performance measures by an axiomatic approach in a fashion similar to the axiomatics of coherent and convex risk measures. The authors define a performance measure (also called an acceptability index) as a functional satisfying certain properties, then investigate implications of those axioms, and show a deep connection with coherent risk measures, as well as provide examples of performance measures. The minimal set of four axioms a performance measure should satisfy consists of the quasi-concavity, monotonicity, scaling invariance and semicontinuity (in the form of the so-called Fatou property in L^∞, as the paper [3] considers only functionals on L^∞). In particular, the monotone Sharpe ratio satisfies those axioms and thus provides a new example of a performance measure in the sense of this system of axioms. It also satisfies all the additional natural properties discussed in that paper: the law invariance, the arbitrage consistency ($\mathbb{S}_{\rm p}(X) = +\infty$ iff $X \ge 0$ a.s. and $\mathrm{P}(X>0) \ne 0$) and the expectation consistency (if $\mathrm{E}X < 0$ then $\mathbb{S}_{\rm p}(X) = 0$, and if $\mathrm{E}X > 0$ then $\mathbb{S}_{\rm p}(X) > 0$; this property is satisfied for $p>1$).

Proof (Proof of Theorem 2). Quasi-concavity follows from that the L^p–Sharpe ratio $S_p(X) = \frac{\mathrm{E}X}{\sigma_p(X)}$ is quasi-concave. Indeed, if $S_p(X) \ge c$ and $S_p(Y) \ge c$, then

$$S_p(\lambda X + (1-\lambda)Y) \ge \frac{\lambda\,\mathrm{E}X + (1-\lambda)\,\mathrm{E}Y}{\lambda\sigma_p(X) + (1-\lambda)\sigma_p(Y)} \ge c$$

for any $\lambda \in [0,1]$. Since $\mathbb{S}_p$ is the maximum of $f_Z(X) = S_p(X-Z)$ over $Z \in \mathrm{L}^p_+$, the quasi-concavity is preserved.

The scaling invariance is obvious. Since the expectation and the L^p-deviation are law invariant, in order to prove the law invariance of $\mathbb{S}_{\rm p}$, it is enough to show that the supremum in the definition of $\mathbb{S}_{\rm p}(X)$ can be taken over only $Y \le X$ which are measurable with respect to the σ-algebra generated by X, or, in other words, $Y = f(X)$ for some measurable function f on $\mathbb{R}$. But this follows from the fact that if for any $Y \le X$ one considers $\widetilde{Y} = \mathrm{E}(Y \mid X)$, then $\widetilde{Y} \le X$, $\mathrm{E}(\widetilde{Y}) = \mathrm{E}Y$ and $\sigma_p(Y) \le \sigma_p(Y)$, hence $S_p(\widetilde{Y}) \ge S_p(Y)$.

To prove the 2nd order monotonicity, recall that another characterization of the 2nd order stochastic dominance is as follows: $X_1 \preccurlyeq X_2$ if and only if there exist random variables X_2' and Z (which may be defined on a another probability space) such that $X_2 \stackrel{d}{=} X_2'$, $X_1 \stackrel{d}{=} X_2' + Z$ and $\mathrm{E}(Z \mid X_2') \le 0$. Suppose $X_1 \preccurlyeq X_2$. From the law invariance, without loss of generality, we may assume that X_1, X_2, Z are defined on the same probability space. Then for any $Y_1 \le X_1$ take $Y_2 = \mathrm{E}(Y_1 \mid X_2)$. Clearly, $Y_2 \le X_2$, $\mathrm{E}Y_2 = \mathrm{E}Y_1$ and $\sigma_p(Y_2) \le \sigma_p(Y_1)$. Hence $\sigma_p(X_1) \le \sigma_p(X_2)$.

Finally, the continuity of $\mathbb{S}_\mathrm{p}(X)$ follows from that the expectation and the L^p-deviation are uniformly continuous.

3 Buffered probabilities

In the paper [21] was introduced the so-called buffered probability, which is defined as the inverse function of the conditional value at risk (with respect to the risk level). The authors of that and other papers (for example, [7]) argue that in stochastic optimization problems related to minimization of probability of adverse events, the buffered probability can serve as a better optimality criterion compared to the usual probability.

In this section we show that the monotone Sharpe ratio is tightly related to the buffered probability, especially in the cases $p = 1, 2$. In particular, this will provide a connection of the monotone Share ratio with the conditional value at risk. We begin with a review of the conditional value at risk and its generalization to the spaces L^p. Then we give a definition of the buffered probability, which will generalize the one in [21, 13] from L^1 to arbitrary L^p.

3.1 A review of the conditional value at risk

Let X be a random variable, which describes loss. As opposed to the previous section, now large values are bad, small values are good (negative values are profits). For a moment, to avoid technical difficulties, assume that X has a continuous distribution.

Denote by $Q(X, \lambda)$ the λ-th quantile of the distribution of X, $\lambda \in [0,1]$, i.e. $Q(X, \lambda)$ is a number $x \in \overline{\mathbb{R}}$, not necessarily uniquely defined, such that $\mathrm{P}(X \le x) = \lambda$. The quantile $Q(X, \lambda)$ is also called the value at risk[1] (VAR) of X at level λ, and it shows that in the worst case of probability $1 - \lambda$, the loss will be at least $Q(X, \lambda)$. This interpretation makes VAR a sort of a measure of risk (in a broad meaning of this term), and it is widely used by practitioners.

However, it is well-known that VAR lacks certain properties that one expects from a measure of risk. One of the most important drawbacks is that it doesn't show what happens with probability less than $1 - \lambda$. For example, an investment strategy which loses \$1 mln with probability 1% and \$2 mln with probability 0.5% is quite

[1] Some authors use definitions of VAR and CVAR which are slightly different from the ones used here: for example, take $(-X)$ instead of X, or $1 - \lambda$ instead of λ, etc.

different from a strategy which loses \$1 mln and \$10 mln with the same probabilities, however they will have the same VAR at the 99% level. Another drawback of VAR is that it's not convex – as a consequence, it may not favor diversification of risk, which leads to concentration of risk (above $1-\lambda$ level).

The conditional value at risk (CVAR; which is also called the average value at risk, or the expected shortfall, or the superquantile) is considered as an improvement of VAR. Recall that if $X \in \mathrm{L}^1$ and has a continuous distribution, then CVAR of X at risk level $\lambda \in [0,1]$ can be defined as the conditional expectation in its right tail of probability $1-\lambda$, i.e.

$$\mathrm{CVAR}(X,\lambda) = \mathrm{E}(X \mid X > Q(X,\lambda)) \tag{9}$$

We will also use the notation $\mathbb{Q}(X,\lambda) = \mathrm{CVAR}(X,\lambda)$ to emphasize the connection with quantiles.

CVAR provides a basic (and the most used) example of a coherent risk measure. The theory of risk measures, originally introduced in the seminal paper [1], plays now a prominent role in applications in finance. We are not going to discuss all the benefits of using coherent (and convex) risk measures in optimization problems; a modern review of the main results in this theory can be found, for example, in the monograph [8].

Rockafellar and Uryasev [22] proved that CVAR admits the following representation though the optimization problem

$$\mathbb{Q}(X,\lambda) = \min_{c\in\mathbb{R}} \left(\frac{1}{1-\lambda} \mathrm{E}(X-c)_+ + c \right). \tag{10}$$

Actually, this formula can be used as a general definition for CVAR, which works in the case of any distribution of X, not necessarily continuous. The importance of this representation is that it provides an efficient method to compute CVAR, which in practical applications often becomes much faster than e.g. using formula (9). It also behaves "nicely" when CVAR is used as a constraint or an optimality criterion in convex optimization problems, for example portfolio selection. Details can be found in [22].

Representation (10) readily suggests how CVAR can be generalized to "put more weight" on the right tail of the distribution of X, which provides a coherent risk measure for the space L^p.

Definition 3. For $X \in \mathrm{L}^p$, define the L^p-CVAR at level $\lambda \in [0,1)$ by

$$\mathbb{Q}_p(X,\lambda) = \min_{c\in\mathbb{R}} \left(\frac{1}{1-\lambda} \|(X-c)_+\|_p + c \right).$$

The L^p-CVAR was studied, for example, in the papers [9, 2]. In particular, in [9], it was argued that higher values of p may provide better results than the standard CVAR ($p=1$) in certain portfolio selection problems. For us, the cases $p=1,2$ will be the most interesting due the direct connection with the monotone Sharpe ratio, as will be shown in the next section.

It is known that the following dual representation holds for L^p-CVAR, which we will use below: for any $X \in \mathrm{L}^p$ and $\lambda \in [0,1)$

$$\mathbb{Q}_p(X,\lambda) = \sup\bigl\{\mathrm{E}(RX) \mid R \in \mathrm{L}^q_+,\ \|R\|_q \le (1-\lambda)^{-1},\ \mathrm{E}R = 1\bigr\}, \tag{11}$$

where, as usual, $\frac{1}{p} + \frac{1}{q} = 1$. This result is proved in [2].

3.2 The definition of buffered probability and its representation

Consider the function inverse to CVAR in λ, that is for a random variable X and $x \in \mathbb{R}$ define $\mathbb{P}(X,x) = \lambda$ where λ is such that $\mathbb{Q}(X,\lambda) = x$ (some care should be taken at points of discontinuity, a formal definition is given below). In the papers [21, 12, 13], $\mathbb{P}(X,x)$ was called the "buffered" probability that $X > x$; we explain the rationale behind this name below. At this moment, it may seem that from a purely mathematical point of view such a simple operation as function inversion probably shouldn't deserve much attention. But that's not the case if we take applications into account. For this reason, before we give any definitions, let us provide some argumentation why studying $\mathbb{P}(X,x)$ may be useful for applications.

In many practical optimization problems one may want to consider constraints defined in terms of probabilities of adverse event, or to use those probabilities as optimization criteria. For example, an investment fund manager may want to maximize the expected return of her portfolio under the constraint that the probability of a loss more than \$1 mln should be less than 1%; or an engineer wants to minimize the construction cost of a structure provided that the tension in its core part can exceed a critical threshold with only a very small probability during its lifetime.

Unfortunately, the probability has all the same drawbacks as the value at risk, which were mentioned above: it's not necessarily convex, continuous and doesn't provide information about how wrong things can go if an adverse event indeed happens. For those reasons, CVAR may be a better risk measure, which allows to avoid some of the problems. For example, if using CVAR, the above investor can reformulate her problem as maximization of the expected return given that the average loss in the worst 1% of cases doesn't exceed \$1 mln. However, such a setting of the problem may be inconvenient, as CVAR "speaks" in terms of quantiles, but one may need the answer in terms of probabilities. For example, \$1 mln may be value of liquid assets of the fund which can be quickly and easily sold to cover a loss; so the manager must ensure that the loss doesn't exceed this amount. But it is not clear how she can use the information about the average loss which CVAR provides. A similar problem arises in the example with an engineer.

In [21], Rockafellar and Royset proposed the idea that the inverse of CVAR may be appropriate for such cases: since quantiles and probabilities are mutually inverse, and CVAR is a better alternative to quantiles, then one can expect that the inverse of CVAR, the buffered probability, could be a better alternative to probability. Here, we follow this idea.

Note that, in theory, it is possible to invert CVAR as a function in λ, but, in practice, computational difficulty may be a serious problem for doing that: it may take too much time to compute CVAR for a complex system even for one fixed level of risk λ, so inversion, which requires such a computation for several λ, may be not feasible (and this is often the case in complex engineering or financial models). Therefore, we would like to be able to work directly with buffered probabilities, and have an efficient method to compute them. We'll see that the representation given below turns out to give more than just an efficient method of computation. In particular, in view of Section 2, it will show a connection with the monotone Sharpe ratio, a result which is by no means obvious.

The following simple lemma will be needed to show that it is possible to invert CVAR.

Lemma 2. *For $X \in \mathrm{L}^p$, $p \in [1,\infty)$, the function $f(\lambda) = \mathbb{Q}_p(X,\lambda)$ defined for $\lambda \in [0,1)$ has the following properties:*

1. $f(0) = \mathrm{E}X$;
2. $f(\lambda)$ is continuous and non-decreasing;
3. $f(\lambda)$ is strictly increasing on the set $\{\lambda : f(\lambda) < \operatorname{ess\,sup} X\}$;
4. if $P := \mathrm{P}(X = \operatorname{ess\,sup} X) > 0$, then $f(\lambda) = \operatorname{ess\,sup} X$ for $\lambda \in [1-P^{1/p},1)$.

Proof. The first property obviously follows from the dual representation, and the second one can be easily obtained from the definition. To prove the third property, observe that if $\mathbb{Q}_p(X,\lambda) < \operatorname{ess\,sup} X$, then the minimum in the definition is attained at some $c^* < \operatorname{ess\,sup} X$. So, for any $\lambda' < \lambda$ we have $\mathbb{Q}_p(X,\lambda') \le \frac{1}{1-\lambda'}\|(X-c^*)_+\|_p + c^* < \mathbb{Q}_p(X,\lambda)$ using that $\|(X-c^*)_+\|_p > 0$.

Finally, the fourth property follows from that if $P > 0$, and, in particular, $\operatorname{ess\,sup} X < \infty$, then $\mathbb{Q}_p(X,\lambda) \le \operatorname{ess\,sup} X$ for any $\lambda \in [0,1)$, as one can take $c = \operatorname{ess\,sup} X$ in the definition. On the other hand, for $\lambda_0 = 1 - P^{\frac{1}{p}}$ we have that $R = P^{-1}\,\mathrm{I}(X = \operatorname{ess\,sup} X)$ satisfies the constraint in the dual representation and $\mathrm{E}(RX) = \operatorname{ess\,sup} X$. Hence $\mathbb{Q}(X,\lambda_0) = \operatorname{ess\,sup} X$, and then $\mathbb{Q}_p(X,\lambda) = \operatorname{ess\,sup} X$ for any $\lambda \ge \lambda_0$ by the monotonicity.

Definition 4. For $X \in \mathrm{L}^p$, $p \in [1,\infty)$, and $x \in \mathbb{R}$, set

$$\mathbb{P}_p(X,x) = \begin{cases} 0, & \text{if } x > \operatorname{ess\,sup} X, \\ (\mathrm{P}(X = \sup X))^{\frac{1}{p}}, & \text{if } x = \operatorname{ess\,sup} X, \\ 1 - \mathbb{Q}_p^{-1}(X,x), & \text{if } \mathrm{E}X < x < \operatorname{ess\,sup} X, \\ 1, & \text{if } x \le \mathrm{E}X. \end{cases}$$

The "main" case in this definition is the third one. In particular, one can see that for a random variable X which has a distribution with a support unbounded from above, the first and the second cases do not realize. Figure 1 schematically shows the relation between the quantile function, the CVAR, the probability distribution function, and the buffered probability. In particular, it is easy to see that always $\mathbb{P}_p(X,x) \ge \mathrm{P}(X > x)$. According to the terminology of [21], the difference between these two quantities is a "safety buffer", hence the name buffered probability.

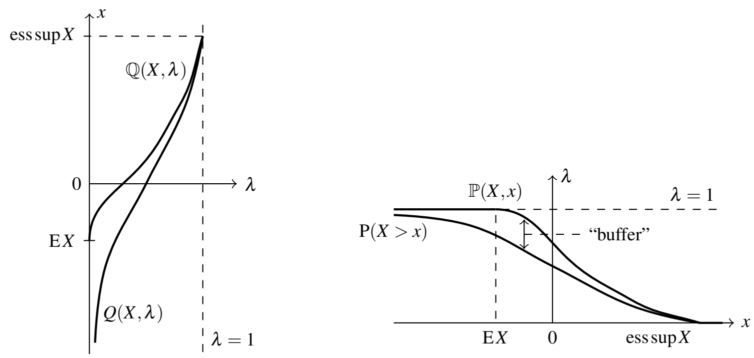

Fig. 1 Left: quantile and distribution functions. Right: complementary probability distribution function and buffered probability $\mathbb{P}(X,x)$. In this example, $\operatorname{ess\,sup} X < \infty$, but $\mathrm{P}(X = \operatorname{ess\,sup} X) = 0$, so $\mathbb{P}(X,x)$ is continuous everywhere.

Theorem 3. *For any* $X \in \mathrm{L}^p$

$$\mathbb{P}_p(X,x) = \min_{c\ge 0} \|(c(X-x)+1)_+\|_p. \tag{12}$$

Proof. For the case $p = 1$ this result was proved in [13]. Here, we follow the same idea, but for general $p \in [1,\infty)$. Without loss of generality, we can assume $x = 0$, otherwise consider $X - x$ instead of X.

Case 1: $\mathrm{E}X < 0$, $\operatorname{ess\,sup} X > 0$. By Lemma 2 and the definition of $\mathbb{Q}_p$ we have

$$\begin{aligned}\mathbb{P}_p(X,0) &= \min\{\lambda \in (0,1) \mid \mathbb{Q}_p(X,1-\lambda) = 0\}\\ &= \min_{\lambda\in(0,1)} \{\lambda \mid \min_{c\in\mathbb{R}} (\tfrac{1}{\lambda}\|((X+c)_+\|_p - c) = 0\}\\ &= \min_{\substack{\lambda\in(0,1)\\ c\in\mathbb{R}}} \{\lambda \mid \|(X+c)_+\|_p = \lambda c\}.\end{aligned}$$

Observe that the minimum here can be computed only over $c > 0$ (since for $c \le 0$ the constraint is obviously not satisfied). Then dividing the both parts of the equality in the constraint by c we get

$$\mathbb{P}_p(X,0) = \min_{c>0} \|(c^{-1}X+1)_+\|_p,$$

which is obviously equivalent to (12).

Case 2: $\mathrm{E}X \ge 0$. We need to show that $\min_{c\ge 0} \|(cX+1)_+\|_p = 1$. This clearly follows from that for any $c \ge 0$ we have $\min_{c\ge 0} \|(cX+1)_+\|_p \ge \min_{c\ge 0} \mathrm{E}(cX+1) = 1$.

Case 3: $\operatorname{ess\,sup} X = 0$. Now $\|(cX+1)_+\|_p \ge \mathrm{P}(X=0)^{1/p}$ for any $c \ge 0$, while $\|(cX+1)_+\|_p \to \mathrm{P}(X=0)^{1/p}$ as $c \to +\infty$. Hence $\min_{c\ge 0} \|(cX+1)_+\|_p = \mathrm{P}(X=0)^{1/p}$ as claimed.

Case 4: $\operatorname{ess\,sup} X < 0$. Similarly, $\|(cX+1)_+\|_p \to 0$ as $c \to +\infty$.

From formula (12), one can easily see the connection between the monotone Sharpe ratio and the buffered probability for $p = 1,2$: for any $X \in \mathrm{L}^p$

$$\frac{1}{1+(\mathbb{S}_\mathrm{p}(X))^p} = (\mathbb{P}_p(-X,0))^p.$$

In particular, if X is as the return of a portfolio, then a portfolio selection problem where one wants to maximize the monotone Sharpe ratio of the portfolio return becomes equivalent to the minimization of the buffered probability that $(-X)$ exceeds 0, i.e. the buffered probability of loss. This is a nice (and somewhat unexpected) connection between the classical portfolio theory and modern developments in risk evaluation!

One can ask a question whether a similar relation between $\mathbb{P}_p$ and $\mathbb{S}_\mathrm{p}$ holds for other values of p. Unfortunately, in general, there seems to be no simple formula connecting them. It can be shown that they can be represented as the following optimization problems:

$$\mathbb{S}_\mathrm{p}(X) = \min_{R \in \mathrm{L}^q_+} \{\|R-1\|_q \mid \mathrm{E}R = 1,\ \mathrm{E}(RX) = 1\},$$
$$\mathbb{P}_p(X,0) = \min_{R \in \mathrm{L}^q_+} \{\|R\|_q \mid \mathrm{E}R = 1,\ \mathrm{E}(RX) = 1\},$$

which have the same constraint sets but different objective functions. The first formula here easily follows from (5), the second one can be obtained using the dual representation of CVAR (11).

3.3 Properties

In this section we investigate some basic properties of $\mathbb{P}_p(X,x)$ both in X and x, and discuss its usage in portfolio selection problem. One of the main points of this section is that buffered probabilities (of loss) can be used as optimality criteria, similarly to monotone Sharpe ratios (and in the cases $p \neq 1,2$ they are more convenient due to s simpler representation).

Theorem 4. *Suppose $X \in \mathrm{L}^p$, $x \in \mathbb{R}$ and $p \in [1,\infty)$. Then $\mathbb{P}_p(X,x)$ has the following properties.*

1. *The function $x \mapsto \mathbb{P}_p(X,x)$ is continuous and strictly decreasing on $[\mathrm{E}X, \operatorname{ess\,sup}(X))$, and non-increasing on the whole $\mathbb{R}$.*
2. *The function $X \mapsto \mathbb{P}_p(X,x)$ is quasi-convex, law invariant, 2nd order monotone, continuous with respect to the L^p-norm, and concave with respect to mixtures of distributions.*
3. *The function $p \mapsto \mathbb{P}_p(X,x)$ is non-decreasing in p.*

For $p = 1$, similar results can be found in [13]; the proofs are similar as well (except property 3, but it obviously follows from the Lyapunov inequality), so we do not provide them here.

Regarding the second property note that despite $\mathbb{P}_p(X,x)$ is quasi-convex in X, it's not convex in X as the following simple example shows: consider $X \equiv 2$ and $Y \equiv -1$; then $\mathbb{P}((X+Y)/2,\ 0) = 1 \not\leq \frac{1}{2} = \frac{1}{2}\mathbb{P}(X,0) + \frac{1}{2}\mathbb{P}(Y,0)$.

Also recall that the mixture of two distributions on $\mathbb{R}$ specified by their distribution functions $F_1(x)$ and $F_2(x)$ is defined as the distribution $F(x) = \lambda F_1(x) + (1-\lambda)F_2(x)$ for any fixed $\lambda \in [0,1]$. We write $X \overset{d}{=} \lambda X_1 \oplus (1-\lambda)X_2$ if the distribution of a random variable X is the mixture of the distributions of X_1 and X_2. If ξ is a random variable taking values $1,2$ with probabilities $\lambda, 1-\lambda$ and independent of X_1, X_2, then clearly $X \overset{d}{=} X_\xi$. Concavity of $\mathrm{P}_p(x,x)$ with respect to mixtures of distributions means that $\mathbb{P}_p(X,x) \geq \lambda \mathbb{P}_p(X_1,x) + (1-\lambda)\mathbb{P}_p(X_2,x)$.

Now let's look in more details on how a simple portfolio selection problem can be formulated with $\mathbb{P}_p$. Assume the same setting as in Section 2.1: R is a $(n+1)$-dimensional vector of asset returns, the first asset is riskless with the rate of return r, and the other n assets are risky with random return in L^p. Let $R_x = \langle x, R\rangle$ denote the return of a portfolio $x \in \mathbb{R}^{n+1}$, and $\delta > 0$ be a fixed number, a required expected return premium. Consider the following optimization problem:

$$\begin{aligned} \text{minimize} \quad & \mathbb{P}_p(r - R_x,\ 0) \text{ over } x \in \mathbb{R}^{n+1} \\ \text{subject to} \quad & \mathrm{E}(R_x - r) = \delta, \\ & \textstyle\sum_i x_i = 1. \end{aligned} \tag{13}$$

In other words, an investor wants to minimize the buffered probability that the return of her portfolio will be less than the riskless return subject to the constraint on the expected return. Denote the vector of adjusted risky returns $\overline{R} = (R_1 - r, \dots, R_n - r)$, and the risky part of the portfolio $\overline{x} = (x_1, \dots, x_n)$. Using the representation of $\mathbb{P}_p$, the problem becomes

$$\begin{aligned} \text{minimize} \quad & \mathrm{E}(1 - \langle \overline{x}, \overline{R}\rangle)_+^p \text{ over } \overline{x} \in \mathbb{R}^n \\ \text{subject to} \quad & \mathrm{E}\,\langle \overline{x}, \overline{R}\rangle \geq 0. \end{aligned} \tag{14}$$

If we find a solution $\overline{x}^*$ of this problem, then the optimal portfolio in problem (13) can be found as follows:

$$x_i^* = \frac{\delta \overline{x}_i^*}{\mathrm{E}\,\langle \overline{x}^*, \overline{R}\rangle},\ i = 1, \dots, n, \qquad x_0^* = 1 - \sum_{i=1}^n x_i^*.$$

Moreover, observe that the constraint $\mathrm{E}\,\langle \overline{x}, \overline{R}\rangle \geq 0$ can be removed in (14) since the value of the objective function is not less than 1 in the case if $\mathrm{E}\,\langle \overline{x}, \overline{R}\rangle < 0$, which is not optimal. Thus, (14) becomes an unconstrained problem.

4 Dynamic problems

This section illustrates how the developed theory can be used to give new elegant solutions of dynamic portfolio selection problems when an investor can continuously trade in the market. The results we obtain are not entirely new, but their proofs are considerably shorter and simpler than in the literature.

4.1 A continuous-time market model and two investment problems

Suppose there are two assets traded in the market: a riskless asset with price B_t and a risky asset with price S_t at time $t \in [0,\infty)$. The time runs continuously. Without loss of generality, we assume $B_t \equiv 1$. The price of the risky asset is modeled by a geometric Brownian motion with constant drift μ and volatility σ, i.e.

$$S_t = S_0 \exp\biggl(\sigma W_t + \Bigl(\mu - \frac{\sigma^2}{2} t\Bigr)\biggr), \qquad t \geq 0,$$

where W_t is a Brownian motion (Wiener process). Without loss of generality, $S_0 = 1$. It is well-known that the process S_t is the unique strong solution of the stochastic differential equation (SDE)

$$dS_t = S_t(\mu dt + \sigma dW_t).$$

We consider the following two problems of choosing an optimal investment strategy in this market model.

Problem 1. Suppose a trader can manage her portfolio dynamically on a time horizon $[0,T]$. A trading strategy is identified with a scalar control process u_t, which is equal to the amount of money invested in the risky asset at time t. The amount of money v_t is invested in the riskless asset. The value $X_t^u = u_t + v_t$ of the portfolio with the starting value $X_0^u = x_0$ satisfies the controlled SDE

$$dX_t^u = u_t(\mu dt + \sigma dW_t), \qquad X_0^u = x_0. \tag{15}$$

This equation is well-known and it expresses the assumption that the trading strategy is self-financing, i.e. it has no external inflows or outflows of capital. Note that v_t doesn't appear in the equation since it can be uniquely recovered as $v_t = X_t^u - u_t$.

To have X^u correctly defined, we'll assume that u_t is predictable with respect to the filtration generated by W_t and $\mathrm{E}\int_0^T u_t^2 dt < \infty$. We'll also need to impose the following mild technical assumption:

$$\mathrm{E}\exp\biggl(\frac{\sigma^2 p^2}{2}\int_0^T \frac{u_t^2}{(1-X_t^u)^2} dt\biggr) < \infty. \tag{16}$$

The class of all the processes u_t satisfying these assumptions will be denoted by $\mathscr{U}$. Actually, it can be shown that (16) can be removed without changing the class of optimal strategies in the problem formulated below, but to keep the presentation simple, we will require it to hold.

The problem consists in minimizing the buffered probability of loss by time T. So the goal of the trader is to solve the following control problem with some fixed $p \in (1,\infty)$:

$$V_1 = \inf_{u\in\mathscr{U}} \mathbb{P}_p(x_0 - X_T^u, 0). \tag{17}$$

For $p = 2$ this problem is equivalent to the problem of maximization of the monotone Sharpe ratio $\mathbb{S}_\mathrm{p}(X_T^u - x_0)$. Moreover, we'll also show that the same solution is obtained in the problem of maximization of the standard Sharpe ratio, $S(X_T^u - x_0)$. Note that we don't consider the case $p = 1$.

From (15) and (17), it is clear that without loss of generality we can (and will) assume $x_0 = 0$. It is also clear that there is no unique solution of problem (17): if some u^* minimizes $\mathbb{P}_p(x_0 - X_T^u, 0)$ then so does any $u_t = cu_t^*$ with a constant $c > 0$. Hence, additional constraints have to be imposed if one wants to have a unique solution, for example a constraint on the expected return like $\mathrm{E}X_T^u = x_0 + \delta$. This is similar to the standard Markowitz portfolio selection problem, as discussed in Section 2.1.

Problem 2. Suppose at time $t = 0$ a trader holds one unit of the risky asset with the starting price $S_0 = 1$ and wants to sell it better than some goal price $x \geq 1$. The asset is indivisible (e.g. a house) and can be sold only at once.

A selling strategy is identified with a Markov time of the process S_t. Recall that a random variable τ with values in $[0,\infty]$ is called a Markov time if the random event $\{\tau \leq t\}$ is in the σ-algebra $\sigma(S_r;\ r \leq t)$ for any $t \geq 0$. The notion of a stopping time reflects the idea that no information about the prices in the future can be used at the moment when the trader decides to sell the asset. The random event $\{\tau = \infty\}$ is interpreted as the situation when the asset is never sold, and we set $S_\infty := 0$. We'll see that the optimal strategy in the problem we formulate below will not sell the asset with positive probability.

Let $\mathscr{M}$ we denote the class of all Markov times of the process S_t. We consider the following optimal stopping problem for $p \in (1,\infty)$:

$$V_2 = \inf_{\tau\in\mathscr{M}} \mathbb{P}_p(x - S_\tau, 0), \tag{18}$$

i.e. minimization of the buffered probability to sell worse then for the goal price x. Similarly to Problem 1, in the case $p = 2$, it'll be shown that this problem is equivalent to maximization of the monotone Sharpe ratio $\mathbb{S}_2(S_\tau - x)$, and the optimal strategy also maximizes the standard Sharpe ratio.

4.2 A brief literature review

Perhaps, the most interesting thing to notice about the two problems is that they are “not standard” from the point of view of the stochastic control theory for diffusion processes and Brownian motion. Namely, they don’t directly reduce to solutions of some PDEs, which for “standard” problems is possible via the Hamilton–Jacobi–Bellman equation. In Problems 1 and 2 (and also in the related problems of maximization of the standard Sharpe ratio), the HJB equation cannot be written because the objective function is not in the form of the expectation of some functional of the controlled process, i.e. not $\mathrm{E}F(X_r^u;\ r \le T)$ or $\mathrm{E}F(S_r;\ r \le \tau)$. Hence, another approach is needed to solve them.

Dynamic problems of maximization of the standard Sharpe ratio and related problems with mean-variance optimality criteria have been studied in the literature by several authors. We just briefly mention several of them.

Richardson [19] was, probably, the first who solved a dynamic portfolio selection problem under a mean-variance criterion (the earlier paper of White [26] can also be mentioned); he used “martingale” methods. Li and Ng [11] studied a multi-asset mean-variance selection problem, which they solved through auxiliary control problems in the standard setting. A similar approach was also used in the recent papers by Pedersen and Peskir [15, 16]. The first of them provides a solution for the optimal selling problem (an analogue of our Problem 2), the second paper solves the portfolio selection problem (our Problem 1). There are other results in the literature, a comprehensive overview can be found in the above-mentioned papers by Pedersen and Peskir, and also in the paper [6].

It should be also mentioned that a large number of papers study the so-called problem of time inconsistency of the mean-variance and similar optimality criteria, which roughly means that at a time $t > t_0$ it turns out to be not optimal to follow the strategy, which was optimal at time t_0. Such a contradiction doesn’t happen in standard control problems for Markov processes, where the Bellman principle can be applied, but it is quite typical for non-standard problems. Several approaches to redefine the notion of an optimal strategy that would take into account time inconsistency are known: see, for example, the already mentioned papers [15, 16, 6] and the references therein. We will not deal with the issue of time inconsistency (our solutions are time inconsistent).

Compared to the results in the literature, the solutions of Problems 1 and 2 in the case $p = 2$ readily follows from earlier results (e.g. from [19, 15, 16]); the other cases can be also studied by previously known methods. Nevertheless, the value of this paper is in the new approach to solve them through the monotone Sharpe ratio and buffered probabilities. This approach seems to be simpler than previous ones (the reader can observe how short the solutions presented below compared to [15, 16]) and promising for more general settings.

4.3 Solution of Problem 1

Theorem 5. *The class of optimal control strategies in Problem 1 is given by*

$$u_t^c = \frac{\mu}{\sigma^2(p-1)}(c - X_t^{u^c}),$$

where $c > 0$ can an arbitrary constant. The process $Y_t^{u^c} = c - X_t^{u^c}$ is a geometric Brownian motion satisfying the SDE

$$\frac{dY_t^{u^c}}{Y_t^{u^c}} = -\frac{\mu^2}{\sigma^2(p-1)}dt - \frac{\mu}{\sigma(p-1)}dW_t, \qquad Y_0^{u^c} = c.$$

Proof. Assuming $x_0 = 0$, from the representation of $\mathbb{P}_p(X,x)$ we have

$$V_1 = \min_{c\geq 0}\min_{u\in\mathscr{U}} \|(1-cX_T^u)_+\|_p = \min_{u\in\mathscr{U}}\|(1-X_T^u)_+\|_p, \tag{19}$$

where in the second equality we used that the constant c can be included in the control, since $cX^u = X^{cu}$. Denote $\widetilde{X}_t^u = 1 - X_t^u$, so that the controlled process $\widetilde{X}^u$ satisfies the equation

$$d\widetilde{X}_t^u = -\mu u_t dt - \sigma u_t dW_t, \qquad \widetilde{X}_0^u = 1.$$

Then

$$V_1^p = \min_{u\in\mathscr{U}} \mathrm{E}\,|\widetilde{X}_T^u|^p, \tag{20}$$

where $(\cdot)_+$ from (19) was removed since it is obvious that as soon as $\widetilde{X}_t^u$ reaches zero, it is optimal to choose $u \equiv 0$ afterwards, so the process stays at zero.

Let $v_t = v_t(u) = -u_t/\widetilde{X}_t^u$. Then for any $u \in \mathscr{U}$ we have

$$\mathrm{E}\,|\widetilde{X}_T^u|^p = \mathrm{E}\Big\{Z_T \exp\Big(\textstyle\int_0^T \big(\mu p v_s + \frac{1}{2}\sigma^2(p^2-p)v_s^2\big)ds\Big)\Big\},$$

where Z is the stochastic exponent process $Z = \mathscr{E}(\sigma p v)$. From Novikov's condition, which holds due to (16), Z_t is a martingale and $EZ_T = 1$. By introducing the new measure Q on the σ-algebra $\mathscr{F}_T = \sigma(W_t,\ t \leq T)$ with the density $dQ = Z_T dP$ we obtain

$$\mathrm{E}\,|\widetilde{X}_T^u|^p = \mathrm{E}^Q\Big\{\exp\Big(\textstyle\int_0^T \big(\mu p v_s + \frac{1}{2}\sigma^2(p^2-p)v_s^2\big)ds\Big)\Big\}.$$

Clearly, this expression can be minimized by minimizing the integrand for each t, i.e. by

$$v_t^* = -\frac{\mu}{\sigma^2(p-1)} \text{ for all } t \in [0,T].$$

Obviously, it satisfies condition (16), so the corresponding control process

$$u_t^* = \frac{\mu}{\sigma^2(p-1)}\widetilde{X}_t^u = \frac{\mu}{\sigma^2(p-1)}(1 - X_t^u)$$

is optimal in problem (20). Consequently, any control process $u_t^c = cu_t^*$, $c > 0$, will be optimal in (17). Since $X_t^{u^c} = cX_t^{u^*}$, we obtain the first claim of the theorem. The representation for $Y_t^{u^c}$ follows from straightforward computations.

Corollary 1. *Let $u^* = \frac{\mu}{\sigma^2}(c - X_t^u)$, with some $c > 0$, be an optimal control strategy in problem* (17) *for $p = 2$. Then the standard Sharpe ratio of $X_T^{u^*}$ is equal to its monotone Sharpe ratio, $S(X_T^{u^*}) = \mathbb{S}_2(S_T^{u^*})$.*

In particular, u^* also maximizes the standard Sharpe ratio of the return X_T^u, i.e. $S(X_T^u) \le S(X_T^{u^*})$ for any $u \in \mathscr{U}$.

Proof. Suppose there is $Y \in \mathrm{L}^2$ such that $Y \le X_T^{u^*}$ and $S(Y) > S(X_T^{u^*})$. It is well-known that the market model we consider is no-arbitrage and complete. This implies that there exists $y_0 < 0$ and a control u_t such that $X_0^u = y_0$ and $X_T^u = Y$. The initial capital y_0 is negative, because otherwise an arbitrage opportunity can be constructed. But then the capital process $\widetilde{X}_t = X_t^u - y_0$ would have a higher Sharpe ratio than Y and hence a higher monotone Sharpe ratio than X_T^{u*}. A contradiction. This proves the first claim of the corollary, and the second one obviously follows from it.

4.4 Solution of Problem 2

We'll assume that $x \ge 1$, $\mu \in \mathbb{R}$, $\sigma > 0$, $p > 1$ are fixed throughout and use the following auxiliary notation:

$$\gamma = \frac{2\mu}{\sigma^2}, \qquad C(b) = \left(\frac{b}{1 + \frac{x}{b-x}(1 - b^{1-\gamma})^{\frac{1}{p-1}}} - x\right)^{-1} \text{ for } b \in [x, \infty).$$

Theorem 6. *The optimal selling time τ^* in problem* (18) *is as follows.*

1. *If $\mu \le 0$, then τ^* can be any Markov time: $\mathbb{P}_p(x - S_\tau, 0) = 1$ for any $\tau \in \mathscr{M}$.*
2. *If $\mu \ge \frac{\sigma^2}{2}$, then S_t reaches any level $x' > x$ with probability* 1 *and any stopping time of the form $\tau^* = \inf\{t \ge 0 : S_t = x'\}$ is optimal.*
3. *If $0 < \mu < \frac{\sigma^2}{2}$, then the optimal stopping time is*

$$\tau^* = \inf\{t \ge 0 : S_t \ge b^*\},$$

where $b^* \in [x, \infty)$ is the point of minimum of the function

$$f(b) = ((1 + C(b)(x-b))^p b^{\gamma-1} + (1 + C(b)x)^p (1 - b^{\gamma-1}), \qquad b \in [x, \infty),$$

and we set $\tau^* = +\infty$ on the random event $\{S_t < b \text{ for all } t \ge 0\}$.

Observe that if $0 < \mu < \frac{\sigma^2}{2}$, i.e. $\gamma \in (0,1)$, then the function $f(b)$ attains its minimum on $[x, \infty)$, since it is continuous with the limit values $f(x) = f(\infty) = 1$.

Proof. From the representation for $\mathbb{P}_p$ we have

$$V_2^p = \inf_{c\geq 0} \inf_{\tau\in\mathscr{M}} \mathrm{E}\,|(1+c(x-S_\tau))_+|^p.$$

Let $Y_t^c = 1 + c(x - S_\tau)$. Observe that if $\mu \leq 0$, then Y_t^c is a submartingale for any $c \geq 0$, and so by Jensen's inequality $|(Y_t)_+)|^p$ is a submartingale as well. Hence for any $\tau \in \mathscr{M}$ we have $\mathrm{E}\,|(1+c(x-S_\tau))_+|^p \geq 1$, and then $V_2 = 1$.

If $\mu \geq \frac{\sigma^2}{2}$, then from the explicit representation $S_t = \exp(\sigma W_t + (\mu - \frac{\sigma^2}{2})t)$ one can see that S_t reaches any level $x' \geq 1$ with probability 1 (as the Brownian motion W_t with non-negative drift does so). Then for any $x' > x$ we have $\mathbb{P}_p(x - S_{\tau_{x'}}, 0) = 0$, where $\tau_{x'}$ is the first moment of reaching x'.

In the case $\mu \in (0, \frac{\sigma^2}{2})$, for any $c \geq 0$, consider the optimal stopping problem

$$V_{2,c} = \inf_{\tau\in\mathscr{M}} \mathrm{E}\,|(1+c(x-S_\tau))_+|^p.$$

This is an optimal stopping problem for a Markov process S_t. From the general theory (see e.g. [17]) it is well known that the optimal stopping time here is of the threshold type:

$$\tau_c^* = \inf\{t \geq 0 : S_t \geq b_c\},$$

where $b_c \in [x, x + \frac{1}{c}]$ is some optimal level, which has to be found. Then the distribution of $S_{\tau_c^*}$ is binomial: it assumes only two values b_c and 0 with probabilities p_c and $1 - p_c$, where $p_c = b_c^{\gamma-1}$ as can be easily found from the general formulas for boundary crossing probabilities for a Brownian motion with drift. Consequently,

$$V_2^p = \inf_{b\geq x} \inf_{c\leq \frac{1}{(b-x)}} \Big((1+c(x-b))^p b^{\gamma-1} + (1+cx)^p(1-b^{\gamma-1})\Big).$$

It is straightforward to find that for any $b \geq x$ the optimal $c^*(b)$ is given by $c^*(b) = C(b)$, which proves the claim of the theorem.

Corollary 2. *Assume $\mu \in (0, \frac{\sigma^2}{2})$ and $p = 2$. Let τ^* denote the optimal stopping time from Theorem 6. Then the standard Sharpe ratio of $S_{\tau^*} - x$ is equal to its monotone Sharpe ratio, $S(S_{\tau^*} - x) = \mathbb{S}_2(S_{\tau^*} - x)$. In particular, τ^* also maximizes the standard Sharpe ratio of $S_\tau - x$, i.e. $S(S_\tau - x) \leq S(S_{\tau^*} - x)$ for any $\tau \in \mathscr{M}$.*

Moreover, in this case the optimal threshold b^* can be found as the point of maximum of the function

$$g(b) = \frac{b^\gamma - x}{b^{\frac{\gamma+1}{2}}(1-b^{\gamma-1})^{\frac{1}{2}}}.$$

Proof. Suppose $Y \leq S_{\tau^*} - x$. As shown above, it is enough to consider only Y which are measurable with respect to the σ-algebra generated by the random variable S_{τ^*}. Since S_{τ^*} has a binomial distribution, then Y should also have a binomial distribution, assuming values $y_1 \leq b^* - x$ and $y_2 \leq -x$ with the same probabilities $(b^*)^{\gamma-1}$ and $1 - (b^*)^{\gamma-1}$ as S_{τ^*} assumes the values b^* and 0. Using this, it is now not difficult to see that $S(Y) \leq S(S_{\tau^*} - x)$, which proves the first claim.

The second claim follows from that for any stopping time of the form $\tau_b = \{t \geq 0 : S_t = b\}$, $b \in [x,\infty)$ we have $S(S_{\tau_b} - x) = g(b)$.

5 Appendix

This appendix just reminds some facts from convex optimization and related results which were used in the paper.

5.1 Duality in optimization

Let $\mathscr{Z}$ be a topological vector space and $f(z)$ a real-valued function on $\mathscr{Z}$. Consider the optimization problem

$$\text{minimize } f(z) \text{ over } z \in \mathscr{Z}. \tag{21}$$

A powerful method to analyze such an optimization problem consists in considering its dual problem. To formulate it, suppose that $f(z)$ can be represented in the form $f(z) = F(z,0)$ for all $z \in \mathscr{Z}$, where $F(z,a)\colon \mathscr{Z} \times \mathscr{A} \to \mathbb{R}$ is some function, and $\mathscr{A}$ is another topological vector space (a convenient choice of F and $\mathscr{A}$ plays an important role).

Let $\mathscr{A}^*$ denote the topological dual of $\mathscr{A}$. Define the Lagrangian $L\colon \mathscr{Z} \times \mathscr{A}^* \to \overline{\mathbb{R}}$ and the dual objective function $g\colon \mathscr{A}^* \to \overline{\mathbb{R}}$ by

$$L(z,u) = \inf_{a\in\mathscr{A}} \{F(z,a) + \langle a,u\rangle\}, \qquad g(u) = \inf_{z\in\mathscr{Z}} L(z,u).$$

Then the dual problem is formulated as the optimization problem

$$\text{maximize } g(u) \text{ over } u \in \mathscr{A}^*.$$

If we denote by V_P and V_D the optimal values of the primal and dual problems respectively (i.e. the infimum of $f(z)$ and the supremum of $g(u)$ respectively), then it is easy too see that $V_P \geq V_D$ always.

We are generally interested in the case when the strong duality takes place, i.e. $V_P = V_D$, or, explicitly,

$$\min_{z\in\mathscr{Z}} f(z) = \max_{u\in\mathscr{A}^*} g(u). \tag{22}$$

Introduce the optimal value function $\phi(a) = \inf_{z\in\mathscr{Z}} F(z,a)$. The following theorem provides a sufficient condition for the strong duality (22) (see Theorem 7 in [20]).

Theorem 7. *Suppose F is convex in (z,a) and $\phi(0) = \liminf_{a\to 0} \phi(a)$. Then (22) holds.*

Let us consider a particular case of problem (21) which includes constraints in the form of equalities and inequalities. Assume that $\mathscr{Z} = \mathrm{L}^p$ for some $p \in [1,\infty)$

and two functions $h_i\colon \mathrm{L}^p \to \mathrm{L}^{r_i}(\mathbb{R}^{n_i})$, $i = 1,2$ are given (the spaces L^p and L^{r_i} are not necessarily defined on the same probability space). Consider the problem

$$\begin{aligned} &\text{minimize} && f(z) \text{ over } z \in \mathrm{L}^p \\ &\text{subject to} && g(z) \le 0 \text{ a.s.} \\ &&& h(z) = 0 \text{ a.s.} \end{aligned}$$

This problem can be formulated as a particular case of the above abstract setting by defining

$$F(z,a_1,a_2) = \begin{cases} f(z), & \text{if } g(z) \le a_1 \text{ and } h(z) = a_2 \text{ a.s.}, \\ +\infty, & \text{otherwise.} \end{cases}$$

The Lagrangian of this problem is

$$\begin{aligned} L(z,u_1,u_2) &= \inf_{a_1,a_2} \big\{ F(z,a_1,a_2) + \langle a_1,u_1\rangle + \langle a_2,u_2\rangle \big\} \\ &= \begin{cases} f(z) + \langle g(z),u_1\rangle + \langle h(z),u_2\rangle, & \text{if } u_1 \ge 0 \text{ a.s.}, \\ -\infty, & \text{otherwise,} \end{cases} \end{aligned}$$

where we denote $\langle a,u\rangle = \mathrm{E}(\sum_i a_i u_i)$.

So the dual objective function

$$g(u,v) = \inf_{z\in L^p} \{ f(z) + \langle g(z),u\rangle + \langle h(z),v\rangle \} \qquad \text{for } u \ge 0 \text{ a.s.},$$

and the dual optimization problem

$$\begin{aligned} &\text{maximize} && g(u,v) \text{ over } u \in L^{r'},\ v \in \mathrm{L}^{w'} \\ &\text{subject to} && u \ge 0. \end{aligned}$$

The strong duality equality:

$$\min_z \{ f(z) \mid g(z) \le 0,\ h(z) = 0 \} = \max_{u,v} \{ g(u,v) \mid u \ge 0 \}$$

5.2 *The minimax theorem*

Theorem 8 (Sion's minimax theorem, Corollary 3.3 in [25]). *Suppose X,Y are convex spaces such that one of them is compact, and $f(x,y)$ is a function on $X \times Y$, such that $x \mapsto f(x,y)$ is quasi-concave and u.s.c. for each fixed y and $y \mapsto f(x,y)$ is quasi-convex and l.s.c. for each fixed x. Then*

$$\sup_{x\in X} \inf_{y\in Y} f(x,y) = \inf_{y\in Y} \sup_{x\in X} f(x,y).$$

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Mathematical Finance **9**(3), 203–228 (1999)
2. Cheridito, P., Li, T.: Risk measures on Orlicz hearts. Mathematical Finance **19**(2), 189–214 (2009)
3. Cherny, A., Madan, D.: New measures for performance evaluation. The Review of Financial Studies **22**(7), 2571–2606 (2008)
4. Cogneau, P., Hübner, G.: The (more than) 100 ways to measure portfolio performance. part 2: special measures and comparison. Journal of Performance Measurement **14**(Fall), 56–69 (2009)
5. Cogneau, P., Hübner, G.: the (more than) 100 ways to measure portfolio performance. part 1: standardized risk-adjusted measures. Journal of Performance Measurement **13**(Summer), 56–71 (2009)
6. Cui, X., Li, D., Wang, S., Zhu, S.: Better than dynamic mean-variance: time inconsistency and free cash flow stream. Mathematical Finance **22**(2), 346–378 (2012)
7. Davis, J.R., Uryasev, S.: Analysis of tropical storm damage using buffered probability of exceedance. Natural Hazards **83**(1), 465–483 (2016)
8. Föllmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time. Walter de Gruyter (2011)
9. Krokhmal, P.A.: Higher moment coherent risk measures. Quantitative Finance **7**, 373–387 (2007)
10. Le Sourd, V.: Performance measurement for traditional investment. Financial Analysts Journal **58**(4), 36–52 (2007)
11. Li, D., Ng, W.L.: Optimal dynamic portfolio selection: multiperiod mean-variance formulation. Mathematical Finance **10**(3), 387–406 (2000)
12. Mafusalov, A., Shapiro, A., Uryasev, S.: Estimation and asymptotics for buffered probability of exceedance. European Journal of Operational Research (2018)
13. Mafusalov, A., Uryasev, S.: Buffered probability of exceedance: mathematical properties and optimization algorithms. Risk Management and Financial Engineering Lab, Department of Industrial and Systems Engineering, University of Florida, Research Report **1**, 2015–5 (2014)
14. Markowitz, H.: Portfolio selection. The Journal of Finance **7**(1), 77–91 (1952)
15. Pedersen, J.L., Peskir, G.: Optimal mean–variance selling strategies. Mathematics and Financial Economics **10**(2), 203–220 (2016)
16. Pedersen, J.L., Peskir, G.: Optimal mean-variance portfolio selection. Mathematics and Financial Economics **11**(2), 137–160 (2017)
17. Peskir, G., Shiryaev, A.: Optimal Stopping and Free-Boundary Problems. Springer (2006)
18. Pliska, S.: Introduction to Mathematical Finance. Blackwell Publishers Oxford (1997)
19. Richardson, H.R.: A minimum variance result in continuous trading portfolio optimization. Management Science **35**(9), 1045–1055 (1989)
20. Rockafellar, R.T.: Conjugate Duality and Optimization, vol. 16. SIAM (1974)
21. Rockafellar, R.T., Royset, J.O.: On buffered failure probability in design and optimization of structures. Reliability Engineering & System Safety **95**(5), 499–510 (2010)
22. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. Journal of Risk **2**, 21–42 (2000)
23. Sharpe, W.F.: Mutual fund performance. The Journal of Business **39**(1), 119–138 (1966)
24. Sharpe, W.F.: The sharpe ratio. Journal of Portfolio Management **21**(1), 49–58 (1994)
25. Sion, M.: On general minimax theorems. Pacific Journal of Mathematics **8**(1), 171–176 (1958)
26. White, D.: Dynamic programming and probabilistic constraints. Operations Research **22**(3), 654–664 (1974)
27. Zhitlukhin, M.: A second-order monotone modification of the sharpe ratio. In: Recent Advances in Financial Engineering 2014: Proceedings of the TMU Finance Workshop 2014, pp. 217–226. World Scientific (2016)
28. Zhitlukhin, M.: On maximization of the expectation-to-deviation ratio of a random variable. Russian Mathematical Surveys **72**(4), 765 (2017)