Mean-Reverting Portfolio Design via Majorization-Minimization Method

Ziping Zhao and Daniel P. Palomar
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology, Hong Kong
Email: {ziping.zhao, palomar}@ust.hk

Abstract—**This paper considers the mean-reverting portfolio design problem arising from statistical arbitrage in the financial markets. The problem is formulated by optimizing a criterion characterizing the mean-reversion strength of the portfolio and taking into consideration the variance of the portfolio and an investment budget constraint at the same time. An efficient algorithm based on the majorization-minimization (MM) method is proposed to solve the problem. Numerical results show that our proposed mean-reverting portfolio design method can significantly outperform every underlying single spread and the benchmark method in the literature.**

I. INTRODUCTION

Pairs trading [1]–[3] is a well known trading strategy which was pioneered by scientists Gerry Bamberger and David Shaw, and the quantitative trading group led by Nunzio Tartaglia at Morgan Stanley in the mid 1980s. As indicated by the name, it is an investment strategy that focuses on a pair of assets at the same time. Investors embracing this strategy do not need to focus on the absolute future price of every single asset, which by nature is hard to assess, but only the relative price of this pair. To arbitrage from the market, investors need to buy the under-priced asset, and short-sell the over-priced one and then profits are locked in after trading positions are unwound when the relative mispricing corrects itself in the future. In general, pairs trading strategy with only two trading assets falls into the umbrella of statistical arbitrage [4]–[6] where the underlying trading basket consists of three or more assets.

In [7], the authors first came up with the concept of cointegration to describe the linear stationary relationship of nonstationary time series which are said to be cointegrated. Later, the cointegrated vector autoregressive model [8] was proposed to discover cointegration relations which could be used to get statistical arbitrage opportunities. Empirical and technical analyses show that cointegration relations really exist in financial markets, for example, the stock prices of the two American famous consumer staple companies Coca-Cola and PepsiCo and those of the two energy companies Ensco and Noble Corporation.

Assets in a cointegration relation can be used to form a portfolio and traded based on its stationary mean-reversion property. Such a portfolio designed on a basket of assets is named as a mean-reverting portfolio (MRP) which is also called a spread. In practice, spreads are usually constructed

using heuristic strategies and inherent correlations may exist among them. However, they are commonly traded separately with their possible connections neglected. So a natural question is whether we can find an optimized portfolio of these spreads which could outperform every single one. In this paper, this issue is precisely addressed by designing a portfolio that combines multiple underlying spreads.

In order to design a mean-reverting portfolio, there are two main factors to consider: i) the designed MRP should exhibit a strong mean-reversion indicating it can have frequent mean-crossing points and hence bring in trading opportunities, and ii) the designed MRP should exhibit enough variance meaning that it could lead to sufficient profits in every trading and at the same time reduce the possibility that the believed mean-reversion equilibrium breaks down. In [9], the author first gave insight on the MRP design problem without considering the variance control. Later, authors in [10] realized that directly solving the problem in [9] could result in a portfolio with very low variance, then the variance control was taken into consideration and new criteria to characterize the mean-reversion property were also proposed for the MRP design problem. In [10], semidefinite relaxation methods were used to solve the nonconvex problem formulations, but these methods are very computationally costly in general. Besides that, the existing optimization problems in [9], [10] were all carried out by imposing an ℓ_2-norm constraint on the portfolio weight which is mathematically convenient but carries no physical meaning in finance.

In this paper, we deal with the MRP design problem by optimizing a mean-reversion criterion called portmanteau statistics, together with the variance control and an investment budget constraint. An efficient solving algorithm based on the MM method is proposed; this allows the efficient resolution of the problem by a sequence of generalized trust region subproblems. Numerical results show that our proposed method significantly outperforms every single spread and the benchmark method in [10] by comparing some performance metrics through an designed trading experiment.

II. MEAN-REVERTING PORTFOLIO DESIGN

A. Mean-Reverting Portfolio

For a financial asset like a common stock, its log-price at time t is computed as $y_t = \log(p_t)$, where $\log(\cdot)$ is the natural logarithm, and p_t denotes its price at time t. For M assets, their

log-prices can be denoted as $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{M,t}]^T$. An MRP is defined by the hedge ratio $\mathbf{w}_s$, and its corresponding log-price spread s_t is defined as

$$s_t = \mathbf{w}_s^T \mathbf{y}_t = \sum_{m=1}^{M} w_{s,m} y_{m,t}, \tag{1}$$

where $\mathbf{w}_s = [w_{s,1}, w_{s,2}, \ldots, w_{s,M}]^T$. Vector $\mathbf{w}_s$ indicates the market value proportions invested on the assets. For $m = 1, 2, \ldots, M$, $w_{s,m} > 0$, $w_{s,m} < 0$ and $w_{s,m} = 0$ mean a long position (i.e., the asset is bought), a short position (i.e., the asset is short-sold), and no position, respectively.

If we consider N spreads, the log-prices for them can be denoted as $\mathbf{s}_t = [s_{1,t}, s_{2,t}, \ldots, s_{N,t}]^T$. Then an MRP can be designed on these spreads using weight $\mathbf{w}$ with its spread written as

$$z_t = \mathbf{w}^T \mathbf{s}_t = \sum_{n=1}^{N} w_n s_{n,t}, \tag{2}$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$ indicates the market values on the spreads. Since we can get the centered counterpart $\tilde{\mathbf{s}}_t = \mathbf{s}_t - \mathsf{E}[\mathbf{s}_t]$ for $\mathbf{s}_t$, we will use $\mathbf{s}_t$ to denote $\tilde{\mathbf{s}}_t$ in the following.

B. Portmanteau Statistics $\mathrm{por}_z(p, \mathbf{w})$

The mean-reversion property for a time series z_t is characterized by the portmanteau statistics of order p [11]. For a centered univariate stationary process z_t, it is defined as

$$\mathrm{por}_z(p) \triangleq T \sum_{i=1}^{p} \rho_i^2, \tag{3}$$

where ρ_i is the i-th order autocorrelation of z_t defined by $\rho_i = \frac{\mathsf{E}[z_t z_{t-i}]}{\mathsf{E}[z_t^2]}$. The portmanteau statistics is used to test whether a random process is close to a white noise. From the definition, we have $\mathrm{por}_z(p) \geq 0$ and the minimum of $\mathrm{por}_z(p)$ is attained by a white noise process, i.e., the portmanteau statistics for a white noise process is 0 for any p. Under this criterion, in order to get a spread z_t as close as possible to a white noise process, we need to minimize $\mathrm{por}_z(p)$ for a prespecified p. Given a spread $z_t = \mathbf{w}^T \mathbf{s}_t$, since $\rho_i = \frac{\mathbf{w}^T \mathsf{E}[\mathbf{s}_t \mathbf{s}_{t+i}^T] \mathbf{w}}{\mathbf{w}^T \mathsf{E}[\mathbf{s}_t \mathbf{s}_t^T] \mathbf{w}} = \frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}$ in which $\mathbf{M}_i$ is the i-th order autocovariance matrix of $\mathbf{s}_t$, and specifically, when $i = 0$, $\mathbf{M}_0$ stands for the (positive definite) covariance matrix, the $\mathrm{por}_z(p, \mathbf{w})$ is written as

$$\mathrm{por}_z(p, \mathbf{w}) = T \sum_{i=1}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2. \tag{4}$$

The matrices $\mathbf{M}_i$s in (4) are assumed symmetric since they can always be symmetrized.

C. MRP Design by Minimizing $\mathrm{por}_z(p, \mathbf{w})$

The MRP design problem can be formulated by minimizing the portmanteau statistics $\mathrm{por}_z(p, \mathbf{w})$ and at the same time setting the variance of z_t ($\mathrm{Var}[z_t] = \mathbf{w}^T \mathsf{E}[\mathbf{s}_t \mathbf{s}_t^T] \mathbf{w} = \mathbf{w}^T \mathbf{M}_0 \mathbf{w}$) to a certain level ν which could be written as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{i=1}^{p} \left(\mathbf{w}^T \mathbf{M}_i \mathbf{w} \right)^2 \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu \\ & \mathbf{1}^T \mathbf{w} = 1, \end{aligned} \tag{5}$$

where $\mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu$ is the spread variance control with ν being a prespecified variance level and $\mathbf{1}^T \mathbf{w} = 1$ is the budget constraint which means that all the money we invest on the long and short positions should sum up to the investment budget. Here, the denominator in $\mathrm{por}_z(p, \mathbf{w})$ is reduced because of the existence the variance constraint. In this formulation, we allow short positions in the portfolio construction which can enable the portfolio managers to achieve leverage by financing portfolio not only with the budget but with the proceeds from the assets held short positions. It is easy to see that the problem in (5) is nonconvex due to the nonconvexity of the objective function and the constraint set.

III. PROBLEM SOLVING METHOD VIA MAJORIZATION-MINIMIZATION

In this section, we first briefly discuss the majorization-minimization or minorization-maximization (MM) method, and then the solving approach for problem (5) is given based on the MM method.

A. The MM Method

The MM method [12]–[14] refers to the majorization-minimization or minorization-maximization method which is a generalization of the well-known expectation-maximization (EM) algorithm. The idea behind MM is that instead of dealing with the original optimization problem which could be difficult to tackle directly, it solves a series of simple surrogate subproblems.

Suppose the optimization problem is as follows:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{6}$$

where the constraint set $\mathcal{X} \subseteq \mathbb{R}^N$. In general, there is no assumption on the convexity and differentiability of $f(\mathbf{x})$. The MM method aims to solve this problem by optimizing a sequence of surrogate functions that majorize the objective function $f(\mathbf{x})$ over the set $\mathcal{X}$. More specifically, starting from an initial feasible point $\mathbf{x}^{(0)}$, the algorithm produces a sequence $\{\mathbf{x}^{(k)}\}$ according to the following update rule:

$$\mathbf{x}^{(k+1)} \in \underset{\mathbf{x} \in \mathcal{X}}{\arg\min}\, u\left(\mathbf{x}, \mathbf{x}^{(k)}\right), \tag{7}$$

where $\mathbf{x}^{(k)}$ is the point generated by the update rule at the k-th iteration and the surrogate function $u\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ is the corresponding majorization function of $f(\mathbf{x})$ at point $\mathbf{x}^{(k)}$. A surrogate function is called a majorization function of $f(\mathbf{x})$ at point $\mathbf{x}^{(k)}$ if it satisfies the following properties:

$$\begin{aligned} u\left(\mathbf{x}, \mathbf{x}^{(k)}\right) &\geq f(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathcal{X}, \\ u\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}\right) &= f\left(\mathbf{x}^{(k)}\right). \end{aligned} \tag{8}$$

That is to say, the surrogate function $u\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ should be an upper bound of the original function $f(\mathbf{x})$ over $\mathcal{X}$ and coincide with $f(\mathbf{x})$ at point $\mathbf{x}^{(k)}$. Although the definition of $u\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ gives us a lot of flexibility for choosing it, in practice, the surrogate function $u\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ should be properly chosen so as to make the iterative update in (7) easy to compute. The MM method iteratively runs until some convergence criterion is met.

In the following, we will apply the MM method to the problem in (5).

B. MRP Design Algorithm via MM

To solve problem (5) via majorization-minimization, the key step is to find a majorization function of the objective function such that the majorized subproblem is easy to solve. Observe that the objective function is quartic in $\mathbf{w}$. The following mathematical manipulations are necessary. We first compute the Cholesky decomposition of $\mathbf{M}_0$ which is $\mathbf{M}_0 = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower triangular and nonsingular matrix. Let us define $\bar{\mathbf{w}} = \mathbf{L}^T\mathbf{w}$, $\bar{\mathbf{M}}_i = \mathbf{L}^{-1}\mathbf{M}_i\left(\mathbf{L}^T\right)^{-1}$, $\mathbf{c} = \mathbf{L}^{-1}\mathbf{1}$, and $\bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T$. Then problem (5) can be written as

$$
\begin{aligned}
&\underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\operatorname{minimize}} && \sum_{i=1}^{p}\left(\operatorname{Tr}\left(\bar{\mathbf{M}}_i\bar{\mathbf{W}}\right)\right)^2 \\
&\text{subject to} && \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\
& && \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\
& && \mathbf{c}^T\bar{\mathbf{w}} = 1.
\end{aligned} \tag{9}
$$

Since $\operatorname{Tr}\left(\bar{\mathbf{M}}_i\bar{\mathbf{W}}\right) = \operatorname{vec}\left(\bar{\mathbf{M}}_i\right)^T\operatorname{vec}\left(\bar{\mathbf{W}}\right)$, where for matrix $\mathbf{M}$, $\operatorname{vec}(\mathbf{M})$ is a column vector consisting of all the columns of $\mathbf{M}$ stacked, then problem (5) can be reformulated as

$$
\begin{aligned}
&\underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\operatorname{minimize}} && \operatorname{vec}\left(\bar{\mathbf{W}}\right)^T\bar{\mathbf{M}}\operatorname{vec}\left(\bar{\mathbf{W}}\right) \\
&\text{subject to} && \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\
& && \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\
& && \mathbf{c}^T\bar{\mathbf{w}} = 1,
\end{aligned} \tag{10}
$$

where $\bar{\mathbf{M}} = \sum_{i=1}^{p}\operatorname{vec}\left(\bar{\mathbf{M}}_i\right)\operatorname{vec}\left(\bar{\mathbf{M}}_i\right)^T$. Now, the objective function in (10) becomes a quadratic function of $\bar{\mathbf{W}}$, however, this problem is still hard to solve. We then consider the application of the MM trick on this problem (10) by introducing the following result.

Lemma 1 ([15, Lemma 1]). *Let $\mathbf{L}$ be a $K \times K$ symmetric matrix and $\mathbf{M}$ be another $K \times K$ symmetric matrix such that $\mathbf{M} \succeq \mathbf{L}$. Then for any point $\mathbf{x}_0 \in \mathbb{R}^K$, the quadratic function $\mathbf{x}^T\mathbf{L}\mathbf{x}$ is majorized by $\mathbf{x}^T\mathbf{M}\mathbf{x} + 2\mathbf{x}_0^T\left(\mathbf{L} - \mathbf{M}\right)\mathbf{x} + \mathbf{x}_0^T\left(\mathbf{M} - \mathbf{L}\right)\mathbf{x}_0$ at $\mathbf{x}_0$.*

According to Lemma 1, given $\bar{\mathbf{W}}^{(k)} = \bar{\mathbf{w}}^{(k)}\bar{\mathbf{w}}^{(k)T}$ at iteration k, we know the objective function of problem (10) is majorized by the following majorization function at $\bar{\mathbf{W}}^{(k)}$:

$$
\begin{aligned}
& u\left(\bar{\mathbf{W}}, \bar{\mathbf{W}}^{(k)}\right) \\
=\ & \psi\left(\bar{\mathbf{M}}\right)\operatorname{vec}\left(\bar{\mathbf{W}}\right)^T\operatorname{vec}\left(\bar{\mathbf{W}}\right) \\
& + 2\operatorname{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T\left(\bar{\mathbf{M}} - \psi\left(\bar{\mathbf{M}}\right)\mathbf{I}\right)\operatorname{vec}\left(\bar{\mathbf{W}}\right) \\
& + \operatorname{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T\left(\psi\left(\bar{\mathbf{M}}\right)\mathbf{I} - \bar{\mathbf{M}}\right)\operatorname{vec}\left(\bar{\mathbf{W}}^{(k)}\right),
\end{aligned} \tag{11}
$$

where $\psi\left(\bar{\mathbf{M}}\right)$ is a scalar depending on $\bar{\mathbf{M}}$ and satisfying $\psi\left(\bar{\mathbf{M}}\right)\mathbf{I} \succeq \bar{\mathbf{M}}$. Since the first term $\operatorname{vec}\left(\bar{\mathbf{W}}\right)^T\operatorname{vec}\left(\bar{\mathbf{W}}\right) = \left(\bar{\mathbf{w}}^T\bar{\mathbf{w}}\right)^2 = \nu^2$ and the last term only depends on $\bar{\mathbf{W}}^{(k)}$, they are just two constants.

Remark 2. On the choosing of $\psi\left(\bar{\mathbf{M}}\right)$, according to Lemma 1, it is obvious that $\psi\left(\bar{\mathbf{M}}\right)$ can be easily chosen to be $\lambda_{max}\left(\bar{\mathbf{M}}\right) = \left\|\bar{\mathbf{M}}\right\|_2$. In the implementation of the algorithm, although $\left\|\bar{\mathbf{M}}\right\|_2$ only need to be computed once for the whole algorithm, it is still not computationally easy to get, so we can try to get an alternative for $\left\|\bar{\mathbf{M}}\right\|_2$. Since for any matrix $\mathbf{M} \in \mathbb{R}^{p \times q}$, we have $\left\|\mathbf{M}\right\|_2 \leq \left\|\mathbf{M}\right\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{q}|m_{ij}|^2}$, $\psi\left(\bar{\mathbf{M}}\right)$ can be chosen to be $\left\|\bar{\mathbf{M}}\right\|_F$ which is bigger than $\left\|\bar{\mathbf{M}}\right\|_2$ but much easier to compute.

After ignoring the constants in (11), the majorized problem of problem (10) is given by

$$
\begin{aligned}
&\underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\operatorname{minimize}} && \operatorname{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T\left(\bar{\mathbf{M}} - \psi\left(\bar{\mathbf{M}}\right)\mathbf{I}\right)\operatorname{vec}\left(\bar{\mathbf{W}}\right) \\
&\text{subject to} && \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\
& && \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\
& && \mathbf{c}^T\bar{\mathbf{w}} = 1,
\end{aligned} \tag{12}
$$

which can be further written as

$$
\begin{aligned}
&\underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\operatorname{minimize}} && \sum_{i=1}^{p}\operatorname{Tr}\left(\bar{\mathbf{M}}_i\bar{\mathbf{W}}^{(k)}\right)\operatorname{Tr}\left(\bar{\mathbf{M}}_i\bar{\mathbf{W}}\right) \\
& && -\psi\left(\bar{\mathbf{M}}\right)\operatorname{Tr}\left(\bar{\mathbf{W}}^{(k)}\bar{\mathbf{W}}\right) \\
&\text{subject to} && \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\
& && \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\
& && \mathbf{c}^T\bar{\mathbf{w}} = 1.
\end{aligned} \tag{13}
$$

By changing $\bar{\mathbf{W}}$ back to $\bar{\mathbf{w}}$, problem (13) becomes

$$
\begin{aligned}
&\underset{\bar{\mathbf{w}}}{\operatorname{minimize}} && \bar{\mathbf{w}}^T\bar{\mathbf{H}}^{(k)}\bar{\mathbf{w}} \\
&\text{subject to} && \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\
& && \mathbf{c}^T\bar{\mathbf{w}} = 1,
\end{aligned} \tag{14}
$$

where in the objective function, $\bar{\mathbf{H}}^{(k)}$ is defined in this way $\bar{\mathbf{H}}^{(k)} = \sum_{i=1}^{p}\left(\bar{\mathbf{w}}^{(k)T}\bar{\mathbf{M}}_i\bar{\mathbf{w}}^{(k)}\right)\bar{\mathbf{M}}_i - \psi\left(\bar{\mathbf{M}}\right)\bar{\mathbf{w}}^{(k)}\bar{\mathbf{w}}^{(k)T}$. Finally, we can undo the change of variable $\bar{\mathbf{w}} = \mathbf{L}^T\mathbf{w}$, obtaining

$$
\begin{aligned}
&\underset{\mathbf{w}}{\operatorname{minimize}} && \mathbf{w}^T\mathbf{H}^{(k)}\mathbf{w} \\
&\text{subject to} && \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\
& && \mathbf{1}^T\mathbf{w} = 1,
\end{aligned} \tag{15}
$$

where in the objective function, $\mathbf{H}^{(k)} = \sum_{i=1}^{p}\left(\mathbf{w}^{T(k)}\mathbf{M}_i\mathbf{w}^{(k)}\right)\mathbf{M}_i - \psi\left(\bar{\mathbf{M}}\right)\mathbf{M}_0\mathbf{w}^{(k)}\mathbf{w}^{(k)T}\mathbf{M}_0$.

Observe that in the majorization problem (15), the objective function becomes quadratic in the variable $\mathbf{w}$ rather than quartic as in the original problem (5).

C. Solving the Majorization Problem in MM

Thus, we just need to solve (15) at every iteration. In order to solve (15), one possible approach is via semidefinite programming (SDP). By using matrix lifting $\mathbf{W} = \mathbf{w}\mathbf{w}^T$, the problem (15) can be reformulated as

$$\begin{array}{ll} \underset{\mathbf{W}}{\text{minimize}} & \text{Tr}\left(\mathbf{H}^{(k)}\mathbf{W}\right) \\ \text{subject to} & \text{Tr}\left(\mathbf{M}_0\mathbf{W}\right) = \nu \\ & \text{Tr}\left(\mathbf{E}\mathbf{W}\right) = 1 \\ & \mathbf{W} \succeq \mathbf{0}, \end{array} \qquad (16)$$

where $\mathbf{E} = \mathbf{1}\mathbf{1}^T$. The above SDP is empirically observed to always give a rank one solution. However, since the computational complexity of solving an SDP is high and it needs to be solved at every iteration, it is not amenable for solving the problem (15).

Here we introduce the solving approach by reformulating (15) into a generalized trust region subproblem [16]. Considering $\mathbf{w} = \mathbf{w}_0 + \mathbf{F}\mathbf{x}$ where $\mathbf{w}_0$ is any vector satisfying $\mathbf{1}^T\mathbf{w}_0 = 1$ and $\mathbf{F}$ is the kernel of $\mathbf{1}^T$ satisfying $\mathbf{1}^T\mathbf{F} = \mathbf{0}$ and a semi-unitary matrix satisfying $\mathbf{F}^T\mathbf{F} = \mathbf{I}$. Let us define $\mathbf{N}^{(k)} = \mathbf{F}^T\mathbf{H}^{(k)}\mathbf{F}$, $\mathbf{p}^{(k)} = \mathbf{F}^T\mathbf{H}^{(k)}\mathbf{w}_0$, $b^{(k)} = \mathbf{w}_0^T\mathbf{H}^{(k)}\mathbf{w}_0$, $\mathbf{N}_0 = \mathbf{F}^T\mathbf{M}_0\mathbf{F}$ which is positive definite, $\mathbf{p}_0 = \mathbf{F}^T\mathbf{M}_0\mathbf{w}_0$, and $b_0 = \mathbf{w}_0^T\mathbf{M}_0\mathbf{w}_0$, then the problem in (15) is equivalent to

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{x}^T\mathbf{N}^{(k)}\mathbf{x} + 2\mathbf{p}^{(k)T}\mathbf{x} + b^{(k)} \\ \text{subject to} & \mathbf{x}^T\mathbf{N}_0\mathbf{x} + 2\mathbf{p}_0^T\mathbf{x} + b_0 = \nu, \end{array} \qquad (17)$$

which is a nonconvex quadratically constrained quadratic programming. Problems of this type (17) are called generalized trust region subproblems (GTRS) [16]. Such problems are usually nonconvex but possess necessary and sufficient optimality conditions. So efficient solving methods can be derived based on these conditions.

According to Theorem 3.2 in [16], the optimality conditions for the primal and dual variables $(\mathbf{x}^\star, \xi^\star)$ of problem (17) are given as follows:

$$\begin{cases} (\mathbf{N} + \xi^\star\mathbf{N}_0)\,\mathbf{x}^\star + \mathbf{p} + \xi^\star\mathbf{p}_0 = 0 \\ \mathbf{x}^{\star T}\mathbf{N}_0\mathbf{x}^\star + 2\mathbf{p}_0^T\mathbf{x}^\star + b_0 - \nu = 0 \\ \mathbf{N} + \xi^\star\mathbf{N}_0 \succeq \mathbf{0}, \end{cases} \qquad (18)$$

where the superscripts for the parameters are neglected hereafter. We assume $\mathbf{N} + \xi\mathbf{N}_0 \succ \mathbf{0}$, then we can get the optimal solution is given by

$$\mathbf{x}\left(\xi\right) = -\left(\mathbf{N} + \xi\mathbf{N}_0\right)^{-1}\left(\mathbf{p} + \xi\mathbf{p}_0\right), \qquad (19)$$

and ξ is the unique solution of the following equation with definition on the interval $\mathcal{I}$

$$\phi\left(\xi\right) = 0, \quad \xi \in \mathcal{I}, \qquad (20)$$

where the function $\phi\left(\xi\right)$ is defined by

$$\phi\left(\xi\right) = \mathbf{x}\left(\xi\right)^T\mathbf{N}_0\mathbf{x}\left(\xi\right) + 2\mathbf{p}_0^T\mathbf{x}\left(\xi\right) + b_0 - \nu, \qquad (21)$$

and the interval $\mathcal{I}$ consists of all ξ for which $\mathbf{N} + \xi\mathbf{N}_0 \succ \mathbf{0}$, which implies that

$$\mathcal{I} = \left(-\lambda_{\min}\left(\mathbf{N}, \mathbf{N}_0\right), \infty\right), \qquad (22)$$

where $\lambda_{min}\left(\mathbf{N}, \mathbf{N}_0\right)$ is the minimum generalized eigenvalue of matrix pair $(\mathbf{N}, \mathbf{N}_0)$. From Theorem 5.2 in [16], we know $\phi\left(\xi\right)$ is strictly decreasing on $\mathcal{I}$, then a simple line search algorithm like bisection algorithm can be used to find the

optimal ξ over $\mathcal{I}$. In the beginning of the derivation, we assume $\mathbf{N} + \xi\mathbf{N}_0$ to be positive definite. This is a reasonable assumption since the case when $\xi = -\lambda_{min}\left(\mathbf{N}, \mathbf{N}_0\right)$ is very rare to happen theoretically and practically.

D. Summary of the MM-Based Algorithm

So, finally in order to solve the original problem (5), we just need to iteratively solve a sequence of GTRS problems in (17). In this paper, we call this MM-based algorithm iteratively reweighted GTRS (IRGTRS). We summarize this algorithm in the following.

Algorithm 1 IRGTRS - Iteratively Reweighted Generalized Trust Region Subproblem algorithm for the MRP design problem in (5).

Require: $k = 0$, p, ν, and initial value $\mathbf{w}^{(0)}$.
1: Compute $\bar{\mathbf{M}} = \sum_{i=1}^p \text{vec}\left(\bar{\mathbf{M}}_i\right)\text{vec}\left(\bar{\mathbf{M}}_i\right)^T$ and $\psi\left(\bar{\mathbf{M}}\right)$;
2: **repeat**
3:     Compute $\mathbf{H}^{(k)} = \sum_{i=1}^p\left(\mathbf{w}^{T(k)}\mathbf{M}_i\mathbf{w}^{(k)}\right)\mathbf{M}_i - \psi\left(\bar{\mathbf{M}}\right)\mathbf{M}_0\mathbf{w}^{(k)}\mathbf{w}^{(k)T}\mathbf{M}_0$;
4:     Compute $\mathbf{w}^{(k+1)}$ by solving the GTRS problem (17);
5:     $k = k + 1$;
6: **until** convergence

IV. NUMERICAL EXPERIMENTS

In this section, we will conduct some numerical experiments to study the performance of our proposed algorithm using synthetic data.

We use a cointegrated system [17] to generated $M = 6$ time series for log-prices with cointegration rank $r = 5$. Within the 5 cointegration relations, we estimate $N = r = 5$ spreads. We divide the sample path into 2 overlapped sliding windows, and every window is divided into two stages: in-sample training stage and out-of-sample backtesting/trading stage. In the training stage, $T_{in}^{wnd} = 12 \times 22 = 264$ samples corresponding to 12 months trading days are used to design the MRP. In the trading stage, $T_{out}^{wnd} = 6 \times 22 = 132$ samples corresponding to 6 months trading days are used for simulated trading experiments. In total, the length of the generated sample path is $T = T_{in}^{wnd} + 2 \times T_{out}^{wnd} = 528$.

We apply the designed spread z_t to a mean-reversion trading strategy. We first define a trading threshold δ for the spread. Here we model z_t as a Gaussian white noise, then according to the result in [1], the maximum trading profit is attained when δ is set to be $0.75 \times \text{Std}\left(z_t\right)$, where $\text{Std}\left(\cdot\right)$ refers to the standard deviation. Knowing the long-run equilibrium of z_t to be μ_z, a simple trading framework is designed as follows: open a long position (denoted as 1), i.e., buy the portfolio z_t at time t_1 when $z_{t_1} \leq \mu_z - \delta$ and unwind the position (denoted as 0) if $\mu_z \leq z_{t_1+\tau_1} < \mu_z + \delta$ or change to a short position (denoted as -1) if $z_{t_1+\tau_1} \geq \mu_z + \delta$ at time $t_1 + \tau_1$ which is the first time that z_t satisfies either of the above conditions; open a short position, i.e., short-sell the portfolio z_t at time t_2 if $z_{t_2} \geq \mu_z + \delta$ and unwind the position if $\mu_z - \delta < z_{t_2+\tau_2} \leq$

μ_z or change to a long position if $z_{t_2+\tau_2} \leq \mu_z - \delta$ at time $t_2 + \tau_2$ which is the first time that z_t satisfies either of the above conditions. Based on this trading scheme, one example is shown in Figure 1.

We also compare the cumulative profit and loss (P&L), the return on investment (ROI), and the Sharpe ratio of the ROI of our designed portfolio with every underlying spread and the portfolio designed using method in [10]. Cumulative P&L measures the return performance from the trading. ROI at time t is defined as the P&L at time t normalized by the gross investment deployed (that is the gross exposure to the market including the long position investment and the short position investment). In the computation of Sharpe ratios, we set the risk-free return to 0. Simulation results show that our designed portfolio can outperform all the underlying spreads and the design method in [10] with a higher final cumulative P&L and a higher Sharpe ratio. The comparison results for two cases are given in Figure 2.

A more detailed discussion on this mean-reverting portfolio design problem will be available in a journal version in preparation.
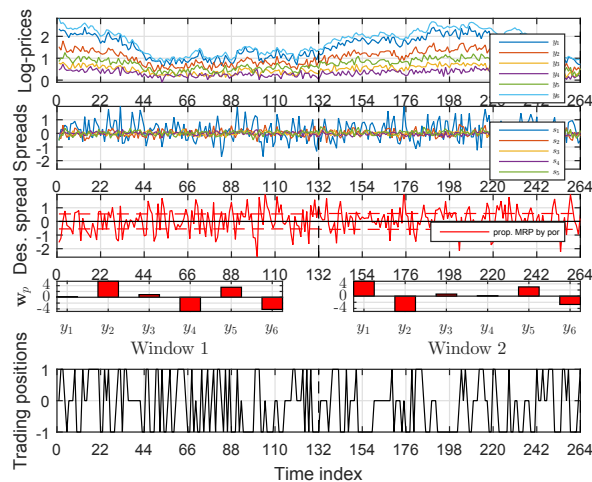
Fig. 1. Log-prices, spreads, a designed MRP using portmanteau statistics $\mathrm{por}_z(\mathbf{w}, 3)$, portfolio weights $\mathbf{w}_p$ (the weight on the underlying assets) and trading positions for two trading windows.

V. CONCLUSIONS

The mean-reverting portfolio design problem arising from statistical arbitrage has been considered in this paper. We have formulated the problem by optimizing a mean-reversion criterion characterizing the mean-reversion strength of the portfolio spread, and at the same time taking into consideration the variance of the spread and an investment budget constraint. Then an efficient algorithm based on MM method has been proposed to solve the problem. Numerical results show that our proposed method is able to generate consistent positive profits and significantly outperform the underlying spreads and the design method in the literature.
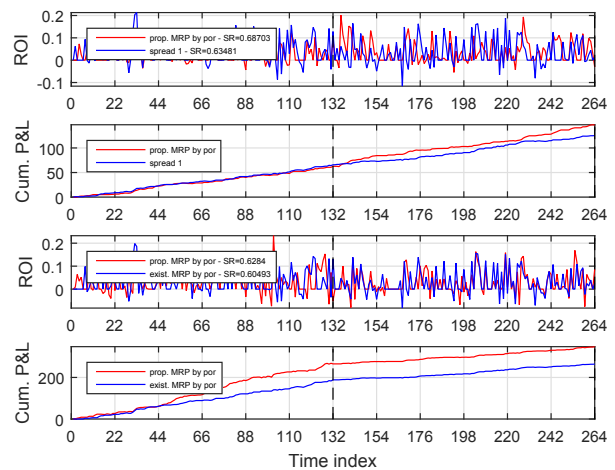
Fig. 2. Comparisons of ROIs, Sharpe ratios of ROIs, and cumulative P&Ls between the MRP designed using our proposed method (prop. MRP by por) with one underlying spread (spread 1) and one benchmark method in [10] (exist. MRP by por).

REFERENCES

[1] G. Vidyamurthy, *Pairs Trading: quantitative methods and analysis*. John Wiley & Sons, 2004, vol. 217.

[2] D. S. Ehrman, *The handbook of pairs trading: Strategies using equities, options, and futures*. John Wiley & Sons, 2006, vol. 240.

[3] R. Bookstaber, *A Demon of Our Own Design: Markets, Hedge Funds, and the Perils of Financial Innovation*. John Wiley & Sons, 2007.

[4] A. Pole, *Statistical arbitrage: algorithmic trading insights and techniques*. John Wiley & Sons, 2011, vol. 411.

[5] B. I. Jacobs and K. N. Levy, *Market Neutral Strategies*. John Wiley & Sons, 2005, vol. 112.

[6] J. G. Nicholas, *Market Neutral Investing: Long/Short Hedge Fund Strategies*. Bloomberg Press, 2000.

[7] R. F. Engle and C. W. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica: Journal of the Econometric Society*, pp. 251–276, 1987.

[8] S. Johansen, "Modelling of cointegration in the vector autoregressive model," *Economic modelling*, vol. 17, no. 3, pp. 359–373, 2000.

[9] A. d'Aspremont, "Identifying small mean-reverting portfolios," *Quantitative Finance*, vol. 11, no. 3, pp. 351–364, 2011.

[10] M. Cuturi and A. d'Aspremont, "Mean reversion with a variance threshold," in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 271–279.

[11] G. E. Box and G. C. Tiao, "A canonical analysis of multiple time series," *Biometrika*, vol. 64, no. 2, pp. 355–365, 1977.

[12] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[13] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[14] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, p. 1, 2016.

[15] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3998–4009, 2015.

[16] J. J. Moré, "Generalizations of the trust region problem," *Optimization methods and Software*, vol. 2, no. 3-4, pp. 189–209, 1993.

[17] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer, 2007.