This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

Mean-Reverting Portfolio With Budget Constraint

Ziping Zhao, *Student Member, IEEE*, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—This paper considers the mean-reverting portfolio (MRP) design problem arising from statistical arbitrage (a.k.a. pairs trading) in the financial markets. It aims at designing a portfolio of underlying assets by optimizing the mean reversion strength of the portfolio, while taking into consideration the portfolio variance and an investment budget constraint. Several specific design problems are considered based on different mean reversion criteria. Efficient algorithms are proposed to solve the problems. Numerical results on both synthetic and market data show that the proposed MRP design methods can generate consistent profits and outperform the traditional design methods and the benchmark methods in the literature.**

***Index Terms*—Portfolio optimization, statistical arbitrage, pairs trading, mean reversion, cointegration, algorithmic trading, quantitative trading, nonconvex optimization, majorization.**

I. Introduction

PAIRS trading [2]–[5], also known as spread trading [6]–[9], is a famous investment and trading strategy pioneered by scientists Gerry Bamberger and David Shaw, as well as the quantitative trading group led by Nunzio Tartaglia at Morgan Stanley in the mid 1980s. As indicated by the name, it is a trading strategy that focuses on a pair of assets at the same time rather than a single one. Investors or arbitrageurs embracing this strategy do not need to forecast the absolute price of every single asset within one trading pair, which by nature is difficult, but only the relative price of this pair. As a contrarian investment strategy, in order to arbitrage from the market, investors should buy the under-priced asset and short-sell the over-priced one. Profits will be locked in after the trading positions are unwound when the relative mispricing of the pair corrects itself in the future.

More generally, pairs trading with only two trading assets falls into the umbrella of statistical arbitrage [10]–[13], also referred to as *stat. arb.*, where the underlying trading basket could consist of three or more financial assets of many kinds such as equities, options, bonds, futures, commodities, etc. Statistical arbitrage opportunities exist as a result of the market inefficiency. Since such strategies can hedge the overall market or systematic risk, and profits do not depend on the movements and conditions of the general financial markets, it is also a kind of market neutral strategy [14], [15]. Nowadays, statistical arbitrage is widely used by many parties in the financial markets, e.g., institutional investors, hedge funds, proprietary trading firms, and individual investors [16].

This work was supported by the Hong Kong RGC 16208917 research grant. Z. Zhao is supported by the Hong Kong PhD Fellowship Scheme (HKPFS).

The authors are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Kowloon, Hong Kong (e-mail: ziping.zhao@connect.ust.hk; palomar@ust.hk).

Part of the results in this paper were preliminary presented at [1].

There are many ways to construct a trading basket, where the cointegration-based method is a prominent one. In [17], [18], the authors first came up with the concept of "cointegration" to describe the linear stationary and hence mean-reverting relationship of the underlying nonstationary time series which are named to be cointegrated. Later, the cointegrated vector autoregressive model was proposed to incorporate such cointegration relations in time series modeling [19], [20]. Empirical and technical analyses show that such relations exist in different financial markets and can be used to get arbitrage opportunities [21]–[25]. Taking the prices of common stocks for example, it is generally known that the stock price can be modeled as a nonstationary random walk process which is hard to predict. However, since companies in the same financial sectors or industries usually share similar fundamental characteristics, their stock prices very often move in company with each other under the same trend, and cointegration relations can be established therefrom for arbitrage. Illustrative examples are the two American famous consumer staple companies Coca-Cola and PepsiCo and the two energy companies Ensco and Noble Corporation. Examples for other financial assets, to name a few, are the future contract prices of E-mini S&P 500 and E-mini Dow, the ETF prices of SPDR S&P 500 and SPDR DJIA, the US dollar foreign exchange rates for different countries, the swap rates for US interest rates of different maturities, and so on.

Mean reversion is a classic indicator of predictability in financial markets. Assets in one cointegration relation can be used to form a portfolio or basket and traded based upon the mean reversion property therein. We call such a designed portfolio a mean-reverting portfolio (MRP) or sometimes a long-short portfolio which is also named a "spread". An asset that naturally shows stationarity is a spread as well, e.g., the option implied volatility for stocks. The profits of statistical arbitrage come directly from trading on the mean reversion of a spread around its long-run equilibrium. MRPs in practice are usually constructed using heuristic or statistical methods. Traditional statistical methods are the Engle-Granger ordinary least squares (OLS) method [18] and the Johansen model-based method [19]. In practice, inherent correlations may exist among different spreads. For example, the spreads estimated from the Johansen method which essentially forms a "cointegration subspace". When having multiple MRPs, instead of trading them separately neglecting the possible connections, a natural and interesting question is whether we can design an optimized MRP based on the underlying spreads which could outperform every single one. In this paper, this issue is addressed.

Designing one MRP by choosing proportions of various assets falls within the umbrella of portfolio optimization or asset allocation problem [26]. Portfolio optimization is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

important in portfolio management as well as in algorithmic trading in the financial industry. The seminal paper [27] by Markowitz in 1952 laid on the foundations of what is now popularly referred to as the mean-variance portfolio and the modern portfolio theory. Given a collection of financial assets, the mean-variance portfolio design problem is aimed at finding a tradeoff between the expected return and the risk. Different from that, to design a mean-reverting portfolio, there are two main factors to consider: i) the designed MRP should exhibit a strong mean reversion indicating that it has frequent mean-crossing points and hence bring in trading opportunities, and ii) the designed MRP should exhibit sufficient but controlled variance so that each trade can provide enough profit while controlling the probability that the expected mean reversion equilibrium does not break down. In [28], the author first proposed to design an MRP by optimizing a criterion characterizing the mean reversion strength, and portfolios for swaps and foreign exchange rates were designed. Later, authors in [29], [30] realized that solving the MRP design problem in [28] could result in a portfolio with very low variance, then the variance control was taken into consideration and also new desirable mean reversion criteria were proposed with portfolios for option implied volatilities designed.

The methods proposed in [28]–[30] are general and tractable for MRP design. However, they are all carried out by imposing an ℓ_2-norm constraint on the portfolio weights. The ℓ_2-norm has a physical meaning of power constraint in wireless communications and used as a similarity constraint in radar signal processing, but its practical significance in financial applications is unclear since the ℓ_2-norm on portfolio weights do not carry a physical meaning in a financial context. In practice, for portfolio design, the constraint on portfolio weights should represent the investment policy and allocation [31]. So, in this paper, we propose to use the investment budget constraints which explicitly represent the budget allocation for different assets.

In [28], [29], semidefinite programming relaxation (SDR) methods were used to solve the nonconvex MRP design problems. SDR also has the drawback of squaring the number of variables, which lifts the problem to much higher dimension. Besides that, not every proposed problem formulation in [29] has a tight SDR with zero duality gap, which makes it hard to justify the resulting solution properties. After solving an SDR, randomization-based rank reduction methods, e.g., [32], are typically applied in order to recover a rank-1 feasible solution from a tight SDR for the original problem, which are computationally costly in general. To solve our problem formulations, instead of resorting to SDR, more efficient solving algorithms are developed.

To make it clear, the contributions of this paper are summarized as follows.

- Based on the mean reversion criteria in [29], [30], the MRP design problem is formulated with a variance constraint and an investment budget constraint (not an ℓ_2-norm constraint). Two commonly used budget constraints are considered, namely, the dollar neutral constraint and the net budget constraint.
- Efficient algorithms are proposed for problem solving. For some problems, after reformulations they can be readily tackled by solving a quadratically constrained quadratic programming (QCQP), specifically, a generalized eigenvalue problem (GEVP) or a generalized trust region subproblem (GTRS) depending on the constraints.
- Other MRP design problems are efficiently solved based on the majorization-minimization (MM) method by solving a sequence of QCQPs, which are named iteratively reweighted GEVP (IRGEVP) or iteratively reweighted GTRS (IRGTRS). Due to the power of MM, more efficient algorithms, named extended IRGEVP (E-IRGEVP) and extended IRGTRS (E-IRGTRS), are also proposed by solving a quadratically constrained linear programming (QCLP) with a closed-form solution at each iteration.
- The complexity per iteration and convergence properties, like monotonic decreasingness and convergence to a stationary point, are analyzed for the MM-based algorithms.

The remaining sections of this paper are organized as follows. In Section II, we briefly introduce the MRP. In Section III, the MRP design problem is formulated based on some mean reversion criteria and two investment budget constraints. Section IV introduces the GEVP and GTRS algorithms. The MM-based algorithms are elaborated in Section V with algorithm complexity and convergence analysis given in Section VI. The numerical performance is evaluated in Section VII and, finally, the concluding remarks are drawn in Section VIII.

Notation: Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and italics denote scalars. The notations $\mathbf{1}$ and $\mathbf{I}$ denote an all-one vector and an identity matrix with proper size, respectively. $\mathbb{R}$ denotes the real field with $\mathbb{R}_+$ denoting positive real numbers and $\mathbb{R}^N$ denoting the N-dimensional real vector space. $\mathbb{N}$ denotes the natural field. $\mathbb{S}^K$ denotes the $K \times K$-dimensional symmetric matrices. Superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote the matrix transpose and inverse operator, respectively. For nonsingular matrices, superscript $(\cdot)^{-T}$ denotes the matrix inverse and transpose operator. $x_{i,j}$ denotes the (ith, jth) element of matrix $\mathbf{X}$ and x_i denotes the ith element of vector $\mathbf{x}$. $\mathbf{X} \succeq \mathbf{0}$ ($\mathbf{X} \succ \mathbf{0}$) means $\mathbf{X}$ is a positive semidefinite (definite) matrix. $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. $\mathrm{vec}(\cdot)$ denotes the vectorization of a matrix. $\otimes$ denotes the Kronecker product of two matrices.

II. Mean-Reverting Portfolio and Mean Reversion Trading

For a financial asset, e.g., a common stock, a future contract, an ETF, or a portfolio of them, its price at time index or holding period $t \in \mathbb{N}$ is denoted by $p_t \in \mathbb{R}_+$, and the corresponding logarithmic price or log-price $y_t \in \mathbb{R}$ is computed as $y_t = \log(p_t)$, where $\log(\cdot)$ is the natural logarithm function. An illustrative example of the log-prices for two security assets denoted as $[y_1, y_2]$ is shown in Figure 1.

For one single asset, the (cumulative) return at time t for τ holding periods is defined as

$$r_t(\tau) = \frac{p_t - p_{t-\tau}}{p_{t-\tau}}, \tag{1}$$

where τ denotes the period length and is usually omitted when the length is one. Then we can have

$$\begin{aligned} r_t(\tau) &\approx \log(p_t) - \log(p_{t-\tau}) \\ &= y_t - y_{t-\tau}, \end{aligned} \tag{2}$$

where the approximation follows from $\log(1+x) \approx x$ for small x, which is valid for the usual trading intervals. Here, the return $r_t(\tau)$ as a rate of return is used to measure the aggregate amount of profits or losses (in percentage) of an investment strategy on one asset over a time period τ.

In order to make an profitable investment (i.e., with a positive return) in the financial markets, the investors need either to buy an asset before its price is going up or to sell an asset before its price is going down. However, in many cases, the asset price is hard to predict. It is usually difficult for people to decide the time point to make an investment on the asset.

In statistical arbitrage strategy, rather than investing on a single asset, people invest on a portfolio of assets at the same time. Such a portfolio or spread is stationary and thus easy to choose the time for investment. In practice, spreads can be naturally stationary like option implied volatilities, designed using methods like technical or fundamental analysis, or constructed based on statistical models. In Figure 1, a spread designed from two security assets is shown.

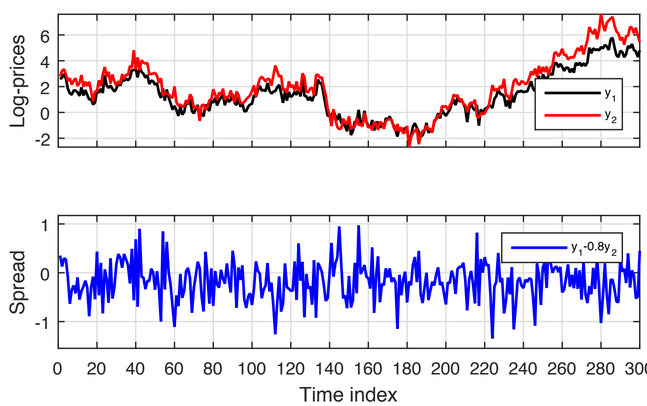

Fig. 1. An illustrative example of log-prices for two assets and the spread.

A. Mean-Reverting Portfolio (MRP)

Different spreads may possess different mean reversion and variance properties in nature. Our objective is to design an MRP to combine such spreads into an improved overall spread with better properties. Suppose there exist N spreads denoted by $\mathbf{s}_t = [s_{1,t}, s_{2,t}, \ldots, s_{N,t}]^T$. We denote the designed mean-reverting portfolio (MRP) by the portfolio weight or hedge ratio $\mathbf{w} = [w_1, w_2, \ldots, w_N]^T$, then the resulting MRP (or spread) is given by

$$z_t = \mathbf{w}^T \mathbf{s}_t = \sum_{n=1}^{N} w_n s_{n,t}, \tag{3}$$

where vector $\mathbf{w}$ indicates the market value proportion invested on the underlying spreads[1]. For $n = 1, 2, \ldots, N$, $w_n > 0$, $w_n < 0$, and $w_n = 0$ mean a long position (i.e., it is bought), a short position (i.e., it is short-sold or, more plainly, borrowed and sold), and no position on the spread, respectively.

When the spread s_t is composed with other underlying financial assets (say, spread from the cointegration model [33]), we can further have the relation between the designed MRP and the underlying financial assets. If a collection of M assets is considered with their log-prices denoted by $\mathbf{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{M,t}]^T$, and a portfolio is defined by the weights $\mathbf{w}_s = [w_{s,1}, w_{s,2}, \ldots, w_{s,M}]^T$, its (log-price) spread s_t is accordingly given by $s_t = \mathbf{w}_s^T \mathbf{y}_t$. Then if N such spreads are consider as in (3), we can get the resulting MRP as

$$z_t = \mathbf{w}_p^T \mathbf{y}_t = \sum_{m=1}^{M} w_{p,m} y_{m,t}, \tag{4}$$

where $\mathbf{w}_p = \mathbf{W}_s \mathbf{w}$ denoting the portfolio weight directly defined on the underlying assets and $\mathbf{W}_s = [\mathbf{w}_{s_1}, \mathbf{w}_{s_2}, \ldots, \mathbf{w}_{s_N}]$.

It is worth noting that an MRP can be interpreted as a synthesized stationary asset. The spread accordingly means the log-price for this MRP, which is much easier to profit from (i.e., to arbitrage) compared to the underlying component assets. The trading strategy to make profits from an MRP is called the mean reversion trading, which is precisely to trade on the mean reversion property of the spread around its equilibrium, i.e., to buy this MRP when the price is lower than its equilibrium and to sell it when the price is higher than its equilibrium.

B. Mean Reversion Trading

In this paper, we use a simple mean reversion trading strategy where the trading signals, i.e., to buy, to sell, or simply to hold, are designed based on simple event triggers. The trading is carried out on the designed spread z_t which is tested to be unit-root stationary. A trading position (a long position denoted by 1 and a short position by -1) is a state for investment and it is opened when the spread z_t is away from its equilibrium μ_z by a predefined trading threshold Δ and closed (denoted by 0) when z_t crosses its equilibrium μ_z. (A common variation is to close the position after the spread crosses the equilibrium by more than another threshold Δ'.) The time period from position opening to position closing is defined as a trading period.

In order to get a standard trading rule, we use the *z-score*, which is a normalized spread measuring the distance to the spread equilibrium in units of standard deviations as follows:

$$\tilde{z}_t = \frac{z_t - \mu_z}{\sigma_z}, \tag{5}$$

where μ_z and σ_z are the mean and the standard deviation of the spread z_t and computed over an in-sample look-back period in practice. For $\tilde{z}_t$, we have $\mathsf{E}[\tilde{z}_t] = 0$ and $\mathsf{Std}[\tilde{z}_t] = 1$.

[1]If the spread is designed based on asset price p_t instead of the log-price, $\mathbf{w}$ indicates the asset amount proportion measured in shares.

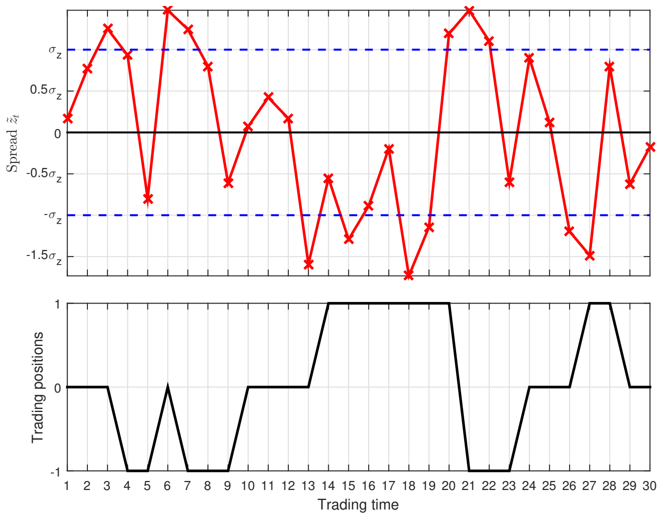

Fig. 2. An example for mean reversion trading (trading threshold $\Delta = \sigma_z$).

Then, we can define the threshold as $\Delta = d \times \sigma_z$, for some value of d (e.g., $d = 1$).

In the trading stage, based on the trading position and observed (normalized) spread value at holding period t, we can get the trading actions at the next consecutive holding period $t+1$. The mean reversion trading strategy is summarized in Table I and a simple trading example based on this strategy is illustrated in Figure 2.

Based on this trading scheme, we can get the profit and loss (P&L) for the MRP which measures the payoff and is also the amount of profits or losses (in units of dollars) of an investment on the portfolio for some holding periods. Within one trading period, if a long position is opened on an MRP at time t_o and closed at time t_c, then the multi-period P&L of this MRP at time t ($t_o \leq t \leq t_c$) accumulated from t_o is computed as $\text{P\&L}_t(\tau) = \mathbf{w}_p^T \mathbf{r}_t(\tau) = \mathbf{w}_p^T \mathbf{r}_t(t - t_o)$,[2] where $\tau = t - t_o$ denotes the length of the holding period, and $\mathbf{r}_t(\tau) = [r_{1,t}(\tau), r_{2,t}(\tau), \ldots, r_{M,t}(\tau)]^T$ is the return vector. More generally, the cumulative P&L of this MRP at time t for τ ($0 \leq \tau \leq t - t_o$) holding periods is defined as

$$\text{P\&L}_t(\tau) = \mathbf{w}_p^T \mathbf{r}_t(t - t_o) - \mathbf{w}_p^T \mathbf{r}_{t-\tau}(t - \tau - t_o), \quad (6)$$

where we define $\mathbf{r}_t(0) = \mathbf{0}$. Then we have the single-period P&L (e.g., daily P&L, monthly P&L) denoted by $\text{P\&L}_t$ at time t (i.e., $\tau = 1$) is computed as

$$\text{P\&L}_t = \mathbf{w}_p^T \mathbf{r}_t(t - t_o) - \mathbf{w}_p^T \mathbf{r}_{t-1}(t - 1 - t_o). \quad (7)$$

If, instead, a short position is opened on this MRP, then multi-period P&L is $\text{P\&L}_t(\tau) = \mathbf{w}_p^T \mathbf{r}_{t-\tau}(t - \tau - t_o) - \mathbf{w}_p^T \mathbf{r}_t(t - t_o)$ and the single-period P&L is $\text{P\&L}_t = \mathbf{w}_p^T \mathbf{r}_{t-1}(t - 1 - t_o) - \mathbf{w}_p^T \mathbf{r}_t(t - t_o)$. About the portfolio P&L calculation within the trading periods, we have the following lemma.

Lemma 1 (P&L Calculation for Mean Reversion Trading). *Within one trading period, if the price change of every asset in an MRP is small enough, then the P&L in (6) can be approximately calculated by the change of the log-price spread z_t. Specifically,*

1) for a long position on the MRP, $\text{P\&L}_t(\tau) \approx z_t - z_{t-\tau}$*; and*

2) for a short position on the MRP, $\text{P\&L}_t(\tau) \approx z_{t-\tau} - z_t$.

Proof: See Appendix A. ■

In fact, Lemma 1 reveals the philosophy behind the MRP design problem and also the mean reversion trading by showing the connection between the log-price spread value and the computation of the portfolio return.

III. Problem Formulation for MRP Design

The traditional mean-variance portfolio which is based on the Nobel prize-winning Markowitz portfolio theory [27], [34] aims at finding a desired trade-off between return and risk, with the latter being measured by the variance. For the mean-reverting portfolio design, we formulate the problem by optimizing a mean reversion criterion quantifying the mean reversion strength [29], [30], while controlling its variance and imposing an investment budget constraint.

A. Mean Reversion Criteria

In this section, we introduce several mean reversion criteria that can characterize the mean reversion strength of the designed spread z_t. We start by defining the ith order (lag-i) autocovariance matrix for a stochastic process $\mathbf{s}_t$ as

$$\mathbf{M}_i = \mathsf{Cov}\left(\mathbf{s}_t, \mathbf{s}_{t+i}\right) = \mathsf{E}\left[\left(\mathbf{s}_t - \mathsf{E}[\mathbf{s}_t]\right)\left(\mathbf{s}_{t+i} - \mathsf{E}[\mathbf{s}_{t+i}]\right)^T\right],$$

where $i \in \mathbb{N}$. Specifically, when $i = 0$, $\mathbf{M}_0$ stands for the (positive definite) covariance matrix of $\mathbf{y}_t$.

Since for any random process $\mathbf{s}_t$, we can always get its centered form as $\tilde{\mathbf{s}}_t = \mathbf{s}_t - \mathsf{E}[\mathbf{s}_t]$, without loss of generality, we use $\mathbf{s}_t$ to denote its centered counterpart $\tilde{\mathbf{s}}_t$ in the following.

1) Predictability Statistics $\text{pre}(\mathbf{w})$*:* Consider a centered univariate stationary autoregressive process $z_t = \hat{z}_{t-1} + \epsilon_t$, where $\hat{z}_{t-1}$ is the prediction of z_t based on the information up to time $t-1$, and ϵ_t denotes a white noise independent from $\hat{z}_{t-1}$. The predictability statistics [35] is defined as

$$\text{pre} = \frac{\sigma_{\hat{z}}^2}{\sigma_z^2}, \quad (8)$$

where $\sigma_z^2 = \mathsf{E}[z_t^2]$ and $\sigma_{\hat{z}}^2 = \mathsf{E}[\hat{z}_{t-1}^2]$. If we define $\sigma_\epsilon^2 = \mathsf{E}[\epsilon_t^2]$, then we have $\sigma_z^2 = \sigma_{\hat{z}}^2 + \sigma_\epsilon^2$ in the denominator. When pre is small, the variance of ϵ_t dominates that of $\hat{z}_{t-1}$, and z_t behaves like a white noise; when pre is large, the variance of $\hat{z}_{t-1}$ dominates that of ϵ_t, and z_t can be well predicted by $\hat{z}_{t-1}$. The predictability statistics is usually used to measure how close a random process is to a white noise.

Based on this criterion, in order to design a spread z_t as close as possible to a white noise process, we need to minimize pre in (8). For $z_t = \mathbf{w}^T \mathbf{s}_t$, we assume the spread $\mathbf{s}_t$ follows a centered vector autoregressive model of order 1 (VAR(1)) as $\mathbf{s}_t = \mathbf{A}\mathbf{s}_{t-1} + \mathbf{e}_t$, where $\mathbf{A}$ is the autoregressive coefficient and $\mathbf{e}_t$ denotes a white noise independent from $\mathbf{s}_{t-1}$. Then

[2]Here $\mathbf{w}_p$ defines the real dollar values for the underlying assets, which is the portfolio weights scaled up by the investment budget.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

TABLE I
TRADING POSITIONS, NORMALIZED SPREAD, AND TRADING ACTIONS OF A MEAN REVERSION TRADING STRATEGY

Trading Position at t	Normalized Spread $\tilde{z}_t$	Action(s) Taken within Holding Period $t+1$	Trading Position at $t+1$
1	$+d \leq \tilde{z}_t$	Close the long position & Open a short position	-1
	$0 \leq \tilde{z}_t < +d$	Close the long position	0
	$\tilde{z}_t < 0$	No action	1
0	$+d \leq \tilde{z}_t$	Open a short position	-1
	$-d < \tilde{z}_t < +d$	No action	0
	$\tilde{z}_t \leq -d$	Open a long position	1
-1	$0 < \tilde{z}_t$	No action	-1
	$-d < \tilde{z}_t \leq 0$	Close the short position	0
	$\tilde{z}_t \leq -d$	Close the short position & Open a long position	1

we can get $\mathbf{A} = \mathbf{M}_1^T \mathbf{M}_0^{-1}$. Premultiplying the VAR(1) by $\mathbf{w}$ and defining $\hat{z}_{t-1} = \mathbf{w}^T \mathbf{A} \mathbf{s}_{t-1}$ and $\epsilon_t = \mathbf{w}^T \mathbf{e}_t$, we have $\sigma_z^2 = \mathbf{w}^T \mathbf{M}_0 \mathbf{w}$ and $\sigma_{\hat{z}}^2 = \mathbf{w}^T \mathbf{T} \mathbf{w}$ with $\mathbf{T} = \mathbf{A} \mathbf{M}_0 \mathbf{A}^T = \mathbf{M}_1^T \mathbf{M}_0^{-1} \mathbf{M}_1$. This also applies to high order models VAR(p) ($p > 1$) through proper reparametrization [36]. Then the predictability statistics for z_t is computed as

$$\mathrm{pre}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{T} \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}. \tag{9}$$

2) Portmanteau Statistics $\mathrm{por}(p, \mathbf{w})$*:* The portmanteau statistics of order p [37] for a centered univariate stationary process z_t is defined as

$$\mathrm{por}(p) = \sum_{i=1}^{p} \rho_i^2, \tag{10}$$

where ρ_i is the ith order (lag-i) autocorrelation of z_t defined as $\rho_i = \mathsf{E}[z_t z_{t+i}] / \mathsf{E}[z_t^2]$. The portmanteau statistics is used to test whether a random process is close to a white noise. From (10), we have $\mathrm{por}(p) \geq 0$ and the minimum is attained by a white noise, i.e., the portmanteau statistics for a white noise process is 0 for any p.

Based on this criterion, in order to get a spread z_t close to a white noise process, we need to minimize $\mathrm{por}_z(p)$ for a prespecified order p. For an MRP $z_t = \mathbf{w}^T \mathbf{s}_t$, the $\rho_i = \mathbf{w}^T \mathsf{E}[\mathbf{s}_t \mathbf{s}_{t+i}^T] \mathbf{w} / \mathbf{w}^T \mathsf{E}[\mathbf{s}_t \mathbf{s}_t^T] \mathbf{w} = \mathbf{w}^T \mathbf{M}_i \mathbf{w} / \mathbf{w}^T \mathbf{M}_0 \mathbf{w}$. Then we can get the expression for $\mathrm{por}(p, \mathbf{w})$ as

$$\mathrm{por}(p, \mathbf{w}) = \sum_{i=1}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2. \tag{11}$$

3) Crossing Statistics $\mathrm{cro}(\mathbf{w})$ *and Penalized Crossing Statistics* $\mathrm{pcro}(p, \mathbf{w})$*:* Crossing statistics (zero-crossing rate) of a centered stationary process z_t is defined as $\mathrm{zcr} = 1/(T-1) \sum_{t=2}^{T} \mathbf{1}_E(z_t)$, where the indicator function $\mathbf{1}_E(z_t) = \begin{cases} 1, & z_t \in E \\ 0, & z_t \notin E \end{cases}$ with event $E = \{z_t z_{t-1} \leq 0\}$. It is used to test the probability that a process crosses its mean per unit of time. According to [38], [39], for a centered stationary Gaussian process, $\mathrm{zcr} = 1/\pi \arccos(\rho_1)$.

Based on this criterion, in order to get a spread z_t having many zero-crossings, we can minimize ρ_1. So for a spread $z_t = \mathbf{w}^T \mathbf{s}_t$, we define the crossing statistics as

$$\mathrm{cro}(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}. \tag{12}$$

In [29], besides minimizing $\mathrm{cro}(\mathbf{w})$, it is also proposed to ensure the absolute high order autocorrelations $|\rho_i|$'s ($i = 2, \ldots, p$) are small which can result in good performance. In this paper, we denote this criterion as the penalized crossing statistics of order p as

$$\mathrm{pcro}(p, \mathbf{w}) = \frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} + \eta \sum_{i=2}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2, \tag{13}$$

where η is a positive prespecified penalization factor.

B. Investment Budget Constraint

In this paper, two types of budget constraints are considered, namely, dollar neutral constraint and net budget constraint.

The dollar neutral constraint, denoted by $\mathcal{W}_0$, means the net investment or net portfolio position is zero; in other words, all the long positions are financed by the short positions, commonly termed self-financing.[3] The portfolio in this case is called zero-cost portfolio. It is represented mathematically by

$$\mathcal{W}_0 = \left\{ \mathbf{1}^T \mathbf{w} = 0 \right\}. \tag{14}$$

The net budget constraint, denoted by $\mathcal{W}_1$, means the net investment or net portfolio position is nonzero and equal to the current budget which is normalized to one.[4] It is represented mathematically by

$$\mathcal{W}_1 = \left\{ \mathbf{1}^T \mathbf{w} = 1 \right\}. \tag{15}$$

It is worth noting that the two trading spreads defined by $\mathbf{w}^T \mathbf{y}_t$ and $-\mathbf{w}^T \mathbf{y}_t$ are naturally the same, because in statistical arbitrage the actual investment not only depends on $\mathbf{w}$, which defines a spread, but also on whether a long or short position is taken on this spread later in the trading stage.

C. General MRP Design Problem Formulation

To make the illustration for the MRP design problem clear in the following, we denote the mean reversion criterion in a compact form as $F(\mathbf{w})$ that takes all the aforementioned criteria into account as follows:

$$F(\mathbf{w}) = \xi \frac{\mathbf{w}^T \mathbf{H} \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} + \zeta \left(\frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2 + \eta \sum_{i=2}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2, \tag{16}$$

[3]The dollar neutral constraint generally cannot be satisfied by the traditional design methods, like methods in [18] and [19], and the methods in [29].

[4]The net portfolio position can be positive or negative under net budget constraint. Since the problem formulation in (17) is invariant to the sign of $\mathbf{w}$, only the case that budget is normalized to positive 1 is considered.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

which particularizes to i) $\text{pre}(\mathbf{w})$, when $\xi = 1$, $\mathbf{H} = \mathbf{T}$, and $\zeta = \eta = 0$; ii) $\text{por}(p, \mathbf{w})$, when $\xi = 0$, and $\zeta = \eta = 1$; iii) $\text{cro}(\mathbf{w})$, when $\xi = 1$, $\mathbf{H} = \mathbf{M}_1$, and $\zeta = \eta = 0$; and iv) $\text{pcro}(p, \mathbf{w})$, when $\xi = 1$, $\mathbf{H} = \mathbf{M}_1$, $\zeta = 0$, and $\eta > 0$. The matrices $\mathbf{M}_i$'s in (16) are assumed symmetric without loss of generality since they can always be symmetrized. As mentioned before, the variance of the spread should be controlled to a certain level which is represented as $\mathsf{Var}\left[\mathbf{w}^T\mathbf{s}_t\right] = \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu$ with ν being a predefined positive constant. Due to this variance constraint, the denominators in $F(\mathbf{w})$ can be reduced. Denoting the portfolio investment budget constraint by $\mathcal{W}$, the general MRP design problem can be formulated as follows:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{mininize}} & \underbrace{\xi \mathbf{w}^T\mathbf{H}\mathbf{w} + \zeta\left(\mathbf{w}^T\mathbf{M}_1\mathbf{w}\right)^2 + \eta\sum_{i=2}^{p}\left(\mathbf{w}^T\mathbf{M}_i\mathbf{w}\right)^2}_{\triangleq f(\mathbf{w})} \\ \text{subject to} & \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i,\ (i = 0, 1). \end{array} \tag{17}$$

The MRP design problem (17) is a nonconvex smooth constrained optimization problem [40] with highly nonconvex (quartic or quadratic) objective function and nonconvex constraint set. To solve the problem, efficient, effective, and convergent algorithms are designed in the following sections.

IV. Problem Solving via GEVP and GTRS Algorithms

In this section, solving methods for the MRP design problem formulations using $\text{pre}(\mathbf{w})$ and $\text{cro}(\mathbf{w})$ (i.e., (17) with $\zeta = \eta = 0$) are introduced.

A. *GEVP: Solving Algorithm for MRP Design Using* $\text{pre}(\mathbf{w})$ *and* $\text{cro}(\mathbf{w})$ *with* $\mathbf{w} \in \mathcal{W}_0$

We recast the relevant problems in (17) as follows:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\mathbf{H}\mathbf{w} \\ \text{subject to} & \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\ & \mathbf{1}^T\mathbf{w} = 0, \end{array} \tag{18}$$

By rewriting the constraint $\mathbf{1}^T\mathbf{w} = 0$ as $\mathbf{w}^T\mathbf{1}\mathbf{1}^T\mathbf{w} = 0$ (since the problem is invariant to a sign change in $\mathbf{w}$) and using the matrix lifting technique (i.e., $\mathbf{W} = \mathbf{w}\mathbf{w}^T$), the above problem can be solved by the following SDR:

$$\begin{array}{ll} \underset{\mathbf{W}}{\text{minimize}} & \text{Tr}(\mathbf{H}\mathbf{W}) \\ \text{subject to} & \text{Tr}(\mathbf{M}_0\mathbf{W}) = \nu \\ & \text{Tr}\left(\mathbf{1}\mathbf{1}^T\mathbf{W}\right) = 0 \\ & \mathbf{W} \succeq \mathbf{0}. \end{array} \tag{19}$$

Although problem (18) is nonconvex, it has no duality gap [41], [42]. In other words, by solving the SDR (19), a rank-1 solution for $\mathbf{W}$ always exists which is a feasible global optimal solution for (18).

As an alternative to the SDR procedure, the optimal solution for (18) can be efficiently solved by reformulating it as a nonconvex QCQP. Considering $\mathbf{w} = \mathbf{F}\mathbf{x}$, where $\mathbf{F}$ is a left-invertible matrix that lies on the null space of $\mathbf{1}^T$ (i.e., $\mathbf{1}^T\mathbf{F} = \mathbf{0}$), we define $\mathbf{N} = \mathbf{F}^T\mathbf{H}\mathbf{F}$ and $\mathbf{N}_0 = \mathbf{F}^T\mathbf{M}_0\mathbf{F}$, then problem (18) is equivalent to the following one:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{x}^T\mathbf{N}\mathbf{x} \\ \text{subject to} & \mathbf{x}^T\mathbf{N}_0\mathbf{x} = \nu. \end{array} \tag{20}$$

This QCQP problem is also known as a generalized eigenvalue problem (GEVP) [43] which can be efficiently solved by tailored algorithms. We choose the steepest descent algorithm in [44] to solve it, which is summarized in Algorithm 1.

Algorithm 1 GEVP - Algorithm for MRP design problems using $\text{pre}(\mathbf{w})$ and $\text{cro}(\mathbf{w})$ with $\mathbf{w} \in \mathcal{W}_0$.

Require: $\mathbf{N}$, $\mathbf{N}_0$, and ν.

1: Set $k = 0$ and $\mathbf{x}^{(0)} \in \{\mathbf{x} \mid \mathbf{x}^T\mathbf{N}_0\mathbf{x} = \nu\}$;
2: **repeat**
3: $\quad R(\mathbf{x}^{(k)}) = \mathbf{x}^{(k)T}\mathbf{N}\mathbf{x}^{(k)}/\mathbf{x}^{(k)T}\mathbf{N}_0\mathbf{x}^{(k)}$;
4: $\quad \mathbf{d}^{(k)} = \mathbf{N}\mathbf{x}^{(k)} - R(\mathbf{x}^{(k)})\mathbf{N}_0\mathbf{x}^{(k)}$;
5: $\quad \hat{\mathbf{x}} = \mathbf{x}^{(k)} + \tau\mathbf{d}^{(k)}$ with τ minimizing $R(\mathbf{x}^{(k)} + \tau\mathbf{d}^{(k)})$;
6: $\quad \mathbf{x}^{(k+1)} = \sqrt{\nu}\hat{\mathbf{x}}/\sqrt{\hat{\mathbf{x}}^T\mathbf{N}_0\hat{\mathbf{x}}}$;
7: $\quad k = k + 1$;
8: **until** convergence

B. *GTRS: Solving Algorithm for MRP Design Using* $\text{pre}(\mathbf{w})$ *and* $\text{cro}(\mathbf{w})$ *with* $\mathbf{w} \in \mathcal{W}_1$

The relevant problems in (17) can be rewritten as

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\mathbf{H}\mathbf{w} \\ \text{subject to} & \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\ & \mathbf{1}^T\mathbf{w} = 1, \end{array} \tag{21}$$

Again, problem (21) can be solved by the SDR. It can also be efficiently solved as a QCQP. Considering $\mathbf{w} = \mathbf{F}\mathbf{x} + \mathbf{w}_0$ where $\mathbf{F}$ is defined as before and $\mathbf{w}_0$ is (any) particular solution of $\mathbf{1}^T\mathbf{w} = 1$, and defining $\mathbf{N} = \mathbf{F}^T\mathbf{H}\mathbf{F}$, $\mathbf{p} = \mathbf{F}^T\mathbf{H}\mathbf{w}_0$, $b = \mathbf{w}_0^T\mathbf{H}\mathbf{w}_0$, $\mathbf{N}_0 = \mathbf{F}^T\mathbf{M}_0\mathbf{F}$, $\mathbf{p}_0 = \mathbf{F}^T\mathbf{M}_0\mathbf{w}_0$, and $b_0 = \mathbf{w}_0^T\mathbf{M}_0\mathbf{w}_0$, the problem (21) is equivalent to the following nonconvex QCQP:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{x}^T\mathbf{N}\mathbf{x} + 2\mathbf{p}^T\mathbf{x} + b \\ \text{subject to} & \mathbf{x}^T\mathbf{N}_0\mathbf{x} + 2\mathbf{p}_0^T\mathbf{x} + b_0 = \nu. \end{array} \tag{22}$$

This QCQP is specially named generalized trust region subproblem (GTRS) [45], [46]. Such problem is usually nonconvex but possesses necessary and sufficient optimality conditions. Efficient solving algorithms for global optimal solution based on the matrix pencil technique can be derived. According to Theorem 3.2 in [45], the optimality conditions for the primal and dual variables $(\mathbf{x}^\star, \xi^\star)$ of problem (22) are given as follows:

$$\begin{cases} (\mathbf{N} + \xi^\star\mathbf{N}_0)\,\mathbf{x}^\star + \mathbf{p} + \xi^\star\mathbf{p}_0 = 0, \\ \mathbf{x}^{\star T}\mathbf{N}_0\mathbf{x}^\star + 2\mathbf{p}_0^T\mathbf{x}^\star + b_0 - \nu = 0, \\ \mathbf{N} + \xi^\star\mathbf{N}_0 \succeq \mathbf{0}. \end{cases} \tag{23}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

Algorithm 2 GTRS - Algorithm for MRP design problems using $\mathrm{pre}(\mathbf{w})$ and $\mathrm{cro}(\mathbf{w})$ with $\mathbf{w} \in \mathcal{W}_1$.

Require: $\mathbf{N}$, $\mathbf{N}_0$, $\mathbf{p}$, $\mathbf{p}_0$, b_0, and ν.

Compute $\lambda_{\min}(\mathbf{N}, \mathbf{N}_0)$.
Set $k = 0$ and $\xi^{(0)} \in (-\lambda_{\min}(\mathbf{N}, \mathbf{N}_0), \infty)$;
repeat
  $\mathbf{x}^{(k)} = -(\mathbf{N} + \xi^{(k)}\mathbf{N}_0)^{-1}(\mathbf{p} + \xi^{(k)}\mathbf{p}_0)$;
  $\phi(\xi^{(k)}) = \mathbf{x}^{(k)T}\mathbf{N}_0\mathbf{x}^{(k)} + 2\mathbf{p}_0^T\mathbf{x}^{(k)} + b_0 - \nu$;
  Update $\xi^{(k+1)}$ by a line search algorithm;
  $k = k + 1$;
until convergence

Assuming $\mathbf{N} + \xi\mathbf{N}_0 \succ \mathbf{0}$,[5] the optimal solution is given by

$$\mathbf{x}(\xi) = -(\mathbf{N} + \xi\mathbf{N}_0)^{-1}(\mathbf{p} + \xi\mathbf{p}_0), \tag{24}$$

and ξ is the unique solution for equation $\phi(\xi) = 0$, where

$$\phi(\xi) = \mathbf{x}(\xi)^T \mathbf{N}_0\mathbf{x}(\xi) + 2\mathbf{p}_0^T\mathbf{x}(\xi) + b_0 - \nu, \tag{25}$$

and $\xi \in \mathcal{I}$. The interval $\mathcal{I} = \{\xi \mid \mathbf{N} + \xi\mathbf{N}_0 \succ \mathbf{0}\}$, which implies $\mathcal{I} = (-\lambda_{\min}(\mathbf{N}, \mathbf{N}_0), \infty)$, where $\lambda_{\min}(\mathbf{N}, \mathbf{N}_0)$ is the minimum generalized eigenvalue of matrix pair $(\mathbf{N}, \mathbf{N}_0)$. According to Theorem 5.2 in [45], the function $\phi(\xi)$ is strictly decreasing on $\mathcal{I}$. So based on this property, a one dimensional search method (e.g., bisection algorithm) can be used to find the optimal ξ over $\mathcal{I}$. The algorithm for solving problem (22) is summarized in Algorithm 2.

V. Problem Solving via MM-Based Algorithms

In this section, we first discuss the MM method briefly, and then two solving algorithms for the MRP design problems using $\mathrm{por}(p, \mathbf{w})$ (i.e., (17) with $\xi = 0$ and $\zeta = \eta = 1$) and $\mathrm{pcro}(p, \mathbf{w})$ (i.e., (17) with $\xi = 1$, $\mathbf{H} = \mathbf{M}_1$, $\zeta = 0$ and $\eta > 0$) are derived based on the MM method together with the GEVP and GTRS algorithms in Section IV.

A. The MM Method

The MM method [47], [48] refers to majorization-minimization for minimization problems or minorization-maximization for maximization problems. It is also known as the successive upper bound minimization method [49], [50].

For an optimization problem given as follows:

$$\begin{array}{ll} \underset{\mathbf{x}}{\mathsf{minimize}} & f(\mathbf{x}) \\ \mathsf{subject\ to} & \mathbf{x} \in \mathcal{X}, \end{array} \tag{26}$$

where the constraint set $\mathcal{X} \subseteq \mathbb{R}^N$ and no assumption is on the convexity of $f(\mathbf{x})$ and $\mathcal{X}$, instead of dealing with the original problem which could be difficult to tackle directly, the MM method solves a series of simple subproblems with surrogate functions that majorize the original objective function $f(\mathbf{x})$ over the set $\mathcal{X}$.

More specifically, starting from an initial feasible point $\mathbf{x}^{(0)}$, the MM method produces a sequence $\{\mathbf{x}^{(k)}\}$ according to the following update rule:

$$\mathbf{x}^{(k)} \in \arg\min_{\mathbf{x} \in \mathcal{X}} \overline{f}\left(\mathbf{x}, \mathbf{x}^{(k-1)}\right), \tag{27}$$

where $\mathbf{x}^{(k-1)}$ is the point generated by the update rule at the $(k-1)$th iteration and $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ is called the majorizing function of $f(\mathbf{x})$ at $\mathbf{x}^{(k)}$.

As to claim convergence for the MM method, the function $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ should satisfy the following assumptions:

$$\begin{aligned} &\text{A1)}\ \overline{f}\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}\right) = f\left(\mathbf{x}^{(k)}\right),\ \forall \mathbf{x}^{(k)} \in \mathcal{X}, \\ &\text{A2)}\ \overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right) \geq f(\mathbf{x}),\ \forall \mathbf{x}, \mathbf{x}^{(k)} \in \mathcal{X}, \\ &\text{A3)}\ \overline{f}'\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}; \mathbf{d}\right) = f'\left(\mathbf{x}^{(k)}; \mathbf{d}\right),\ \forall \mathbf{d}\ \text{s.t.}\mathbf{x}^{(k)} + \mathbf{d} \in \mathcal{X}, \end{aligned} \tag{28}$$

where $f'\left(\mathbf{x}^{(k)}; \mathbf{d}\right)$ stands for the directional derivative of $f(\mathbf{x})$ at $\mathbf{x}^{(k)}$ along the direction $\mathbf{d}$, i.e.,

$$f'\left(\mathbf{x}^{(k)}; \mathbf{d}\right) = \liminf_{t \to 0} \frac{f\left(\mathbf{x}^{(k)} + t\mathbf{d}\right) - f\left(\mathbf{x}^{(k)}\right)}{t};$$

similarly, $\overline{f}'\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}; \mathbf{d}\right)$ is the directional derivative for $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ at $\mathbf{x}^{(k)}$ along $\mathbf{d}$; and $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ is assumed continuous in both $\mathbf{x}$ and $\mathbf{x}^{(k)}$.[6] For convex $\mathcal{X}$, the proof of convergence to a d(irectional)-stationary point is established in [49], i.e., the limit point $\mathbf{x}^{(\infty)}$ of $\{\mathbf{x}^{(k)}\}$ satisfies

$$f'\left(\mathbf{x}^{(\infty)}; \mathbf{d}\right) \geq 0,\ \forall \mathbf{d}\ \text{s.t.}\ \mathbf{x}^{(\infty)} + \mathbf{d} \in \mathcal{X}. \tag{29}$$

For a nonconvex set $\mathcal{X}$, to claim stationarity convergence, the A3) in (28) should be modified as

$$\text{A3}')\ \overline{f}'\left(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}; \mathbf{d}\right) = f'\left(\mathbf{x}^{(k)}; \mathbf{d}\right),\ \forall \mathbf{d} \in \mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(k)}\right), \tag{30}$$

where in this case $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ and $f(\mathbf{x})$ are defined on the whole space $\mathbb{R}^N$ and $\mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(k)}\right)$ means the Bouligand tangent cone of $\mathcal{X}$ at $\mathbf{x}^{(k)} \in \mathcal{X}$. Then, the limit point $\mathbf{x}^{(\infty)}$ of $\{\mathbf{x}^{(k)}\}$ can be proved to be a B(ouligand)-stationary point satisfying

$$f'\left(\mathbf{x}^{(\infty)}; \mathbf{d}\right) \geq 0,\ \forall \mathbf{d} \in \mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(\infty)}\right), \tag{31}$$

where the expression $\mathbf{d} \in \mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(\infty)}\right)$ means there exist a sequence of points $\{\mathbf{x}^{(k)}\} \in \mathcal{X}$ converging to $\mathbf{x}^{(\infty)}$ and a sequence of positive scalars $\{\tau^{(k)}\}$ converging to 0 such that $\mathbf{d} = \lim_{k\to\infty} \frac{\mathbf{x}^{(k)} - \mathbf{x}^{(\infty)}}{\tau^{(k)}}$. For more details of B-stationarity, please refer to [51], [52].

Although the definition for the majorizing functions $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ gives us a great deal of choosing flexibility, they must be properly chosen so as to make the iterative update in (27) easy to compute while maintaining a fast convergence over the iterations. In the following, we are going to solve the MRP design problem based on the MM method.

[5] The limiting case $\mathbf{N} + \xi\mathbf{N}_0$ being singular (i.e., $\xi = -\lambda_{\min}(\mathbf{N}, \mathbf{N}_0)$) can be treated separately. The assumption here is reasonable since the case when $\xi = -\lambda_{\min}(\mathbf{N}, \mathbf{N}_0)$ is very rare to occur theoretically and practically.

[6] Note that if $f(\mathbf{x})$ and $\overline{f}\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$ are both continuously differentiable, then A1) and A2) imply A3).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

B. IRGEVP and IRGTRS: Solving Algorithms for MRP Design Using $\text{por}(p, \mathbf{w})$ and $\text{pcro}(p, \mathbf{w})$

From (17), the MRP deign problems using $\text{por}(p, \mathbf{w})$ and $\text{pcro}(p, \mathbf{w})$ can be written as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & f(\mathbf{w}) = \xi \mathbf{w}^T \mathbf{M}_1 \mathbf{w} + \zeta \left(\mathbf{w}^T \mathbf{M}_1 \mathbf{w}\right)^2 \\ & + \eta \sum_{i=2}^{p} \left(\mathbf{w}^T \mathbf{M}_i \mathbf{w}\right)^2 \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i, \ (i = 0, 1). \end{aligned} \tag{32}$$

Problem (32) is nonconvex with nonconvex quartic objective function, nonconvex quadratic equality constraint and convex linear constraint. In order to solve this problem via MM method, the key step is to find a majorizing function of the objective such that the majorized subproblem is easy to solve.

To compute a majorizing function, the following mathematical manipulations are necessary. We first get the Cholesky decomposition of $\mathbf{M}_0$ which is given as $\mathbf{M}_0 = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower triangular matrix with positive diagonal elements. We further define $\bar{\mathbf{w}} = \mathbf{L}^T \mathbf{w}$, $\bar{\mathbf{M}}_i = \mathbf{L}^{-1} \mathbf{M}_i \mathbf{L}^{-T}$, $\bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T$, and the set $\mathcal{W}$ is mapped to $\bar{\mathcal{W}}$ under the linear transformation $\mathbf{L}$. Then using $\bar{\mathbf{w}}^T \mathbf{A} \bar{\mathbf{w}} = \text{Tr}\left(\mathbf{A}\bar{\mathbf{W}}\right)$, problem (32) can be rewritten as

$$\begin{aligned} \underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\text{minimize}} \quad & \xi \text{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{W}}\right) + \zeta \Big(\text{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{W}}\right)\Big)^2 \\ & + \eta \sum_{i=2}^{p} \Big(\text{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{W}}\right)\Big)^2 \\ \text{subject to} \quad & \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\ & \bar{\mathbf{w}}^T \bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i, \ (i = 0, 1). \end{aligned} \tag{33}$$

Since $\text{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{W}}\right) = \text{vec}\left(\bar{\mathbf{M}}_i\right)^T \text{vec}\left(\bar{\mathbf{W}}\right)$ ($\mathbf{M}_i$'s are assumed symmetric), problem (33) can be reformulated as follows:

$$\begin{aligned} \underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\text{minimize}} \quad & \xi \text{vec}\left(\bar{\mathbf{M}}_1\right)^T \text{vec}\left(\bar{\mathbf{W}}\right) + \text{vec}\left(\bar{\mathbf{W}}\right)^T \bar{\mathbf{M}} \text{vec}\left(\bar{\mathbf{W}}\right) \\ \text{subject to} \quad & \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\ & \bar{\mathbf{w}}^T \bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i, \ (i = 0, 1), \end{aligned} \tag{34}$$

where in the objective function

$$\bar{\mathbf{M}} \triangleq \zeta \text{vec}\left(\bar{\mathbf{M}}_1\right) \text{vec}\left(\bar{\mathbf{M}}_1\right)^T + \eta \sum_{i=2}^{p} \text{vec}\left(\bar{\mathbf{M}}_i\right) \text{vec}\left(\bar{\mathbf{M}}_i\right)^T. \tag{35}$$

Specifically, we have the expressions for portmanteau statistics $\text{por}(p, \mathbf{w})$ (i.e., $\zeta = 1$ and $\eta = 1$) and penalized crossing statistics $\text{pcro}(p, \mathbf{w})$ (i.e., $\zeta = 0$ and $\eta > 0$) as follows:[7]

$$\bar{\mathbf{M}} = \begin{cases} \sum_{i=1}^{p} \left(\mathbf{L} \otimes \mathbf{L}\right)^{-1} \text{vec}\left(\mathbf{M}_i\right) \text{vec}\left(\mathbf{M}_i\right)^T \left(\mathbf{L} \otimes \mathbf{L}\right)^{-T}, \\ \qquad \text{for } \text{por}(p, \mathbf{w}); \\ \eta \sum_{i=2}^{p} \left(\mathbf{L} \otimes \mathbf{L}\right)^{-1} \text{vec}\left(\mathbf{M}_i\right) \text{vec}\left(\mathbf{M}_i\right)^T \left(\mathbf{L} \otimes \mathbf{L}\right)^{-T}, \\ \qquad \text{for } \text{pcro}(p, \mathbf{w}). \end{cases}$$

The objective function in problem (34) becomes quadratic in variable $\bar{\mathbf{W}}$; however, this problem is still hard to solve due to the rank-1 constraint $\bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T$. We then consider applying the MM idea on this problem based on the following result.

Lemma 2. *Let $\mathbf{A} \in \mathbb{S}^K$ and $\mathbf{B} \in \mathbb{S}^K$ such that $\mathbf{B} \succeq \mathbf{A}$. At any point $\mathbf{x}_0 \in \mathbb{R}^K$, the quadratic function $\mathbf{x}^T \mathbf{A} \mathbf{x}$ is majorized by $\mathbf{x}^T \mathbf{B} \mathbf{x} + 2\mathbf{x}_0^T \left(\mathbf{A} - \mathbf{B}\right) \mathbf{x} + \mathbf{x}_0^T \left(\mathbf{B} - \mathbf{A}\right) \mathbf{x}_0$.*

Proof: It follows from $\left(\mathbf{x} - \mathbf{x}_0\right)^T \left(\mathbf{B} - \mathbf{A}\right) \left(\mathbf{x} - \mathbf{x}_0\right) \geq 0$, when $\mathbf{B} \succeq \mathbf{A}$ for any $\mathbf{x}_0$. ■

According to Lemma 2, at the $(k+1)$th iteration with point $\bar{\mathbf{W}}^{(k)}$, the second term (quadratic in $\bar{\mathbf{W}}$) in the objective function of problem (34) can be majorized by the following function:

$$\begin{aligned} & u_1\left(\bar{\mathbf{W}}, \bar{\mathbf{W}}^{(k)}\right) = \psi(\bar{\mathbf{M}}) \text{vec}\left(\bar{\mathbf{W}}\right)^T \text{vec}\left(\bar{\mathbf{W}}\right) \\ & + 2\text{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \left(\bar{\mathbf{M}} - \psi(\bar{\mathbf{M}})\mathbf{I}\right) \text{vec}\left(\bar{\mathbf{W}}\right) \\ & + \text{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \left(\psi(\bar{\mathbf{M}})\mathbf{I} - \bar{\mathbf{M}}\right) \text{vec}\left(\bar{\mathbf{W}}^{(k)}\right), \end{aligned} \tag{36}$$

where $\psi(\bar{\mathbf{M}})$ only depends on matrix $\bar{\mathbf{M}}$ and satisfies $\psi(\bar{\mathbf{M}})\mathbf{I} \succeq \bar{\mathbf{M}}$. On the choice of $\psi(\bar{\mathbf{M}})$ in (36), according to Lemma 2, it is obvious that $\psi(\bar{\mathbf{M}})$ can be chosen as the spectral norm of $\bar{\mathbf{M}}$, i.e., $\left\|\bar{\mathbf{M}}\right\|_2 = \lambda_{\max}\left(\bar{\mathbf{M}}\right)$. In the implementation of the algorithm, although $\lambda_{\max}\left(\bar{\mathbf{M}}\right)$ only needs to be computed once for the whole algorithm, it still may not be computationally easy to get. Then since $\left\|\bar{\mathbf{M}}\right\|_F \geq \left\|\bar{\mathbf{M}}\right\|_2$, we can choose $\psi(\bar{\mathbf{M}}) = \left\|\bar{\mathbf{M}}\right\|_F$, which is easier for computation.

In the majorizing function, the first term and the last term are just two constants irrelevant of the optimization variable $\bar{\mathbf{W}}$, since the first term $\text{vec}\left(\bar{\mathbf{W}}\right)^T \text{vec}\left(\bar{\mathbf{W}}\right) = \left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right)^2 = \nu^2$, and the last term only depends on $\bar{\mathbf{W}}^{(k)}$. After replacing the second term by its majorizing function (36) in problem 34 and ignoring the constants, the majorized problem is given by

$$\begin{aligned} \underset{\bar{\mathbf{w}}, \bar{\mathbf{W}}}{\text{minimize}} \quad & \xi \text{vec}\left(\bar{\mathbf{M}}_1\right)^T \text{vec}\left(\bar{\mathbf{W}}\right) \\ & + 2\text{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \left(\bar{\mathbf{M}} - \psi(\bar{\mathbf{M}})\mathbf{I}\right) \text{vec}\left(\bar{\mathbf{W}}\right) \\ \text{subject to} \quad & \bar{\mathbf{W}} = \bar{\mathbf{w}}\bar{\mathbf{w}}^T \\ & \bar{\mathbf{w}}^T \bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i, \ (i = 0, 1), \end{aligned} \tag{37}$$

where the objective function becomes linear in the variable $\bar{\mathbf{W}}$ rather than quadratic as in (33) and (34). Further, by changing variable $\bar{\mathbf{W}}$ back to $\mathbf{w}$, we can get the overall majorizing function for (32) and the majorized subproblem in $\mathbf{w}$ which is given in the following lemma.

Lemma 3. *The final majorizing function of $f(\mathbf{w})$ in (32) is*

$$\begin{aligned} & \overline{f}_1\left(\mathbf{w}, \mathbf{w}^{(k)}\right) = \mathbf{w}^T \mathbf{H}^{(k)} \mathbf{w} + 2\psi(\bar{\mathbf{M}})\nu^2 \\ & - \zeta \left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)^2 - \eta \sum_{i=2}^{p} \left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)^2, \end{aligned} \tag{38}$$

[7]It follows from $\text{vec}\left(\mathbf{ABC}\right) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \text{vec}\left(\mathbf{B}\right)$ and $\mathbf{A}^{-1} \otimes \mathbf{B}^{-1} = \left(\mathbf{A} \otimes \mathbf{B}\right)^{-1}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

Algorithm 3 IRGEVP and IRGTRS - Algorithms for MRP design problems using $\mathrm{por}(p, \mathbf{w})$ and $\mathrm{pcro}(p, \mathbf{w})$.

Require: p, $\mathbf{M}_i$ with $i = 1, \ldots, p$, and $\nu > 0$.

1: Set $k = 0$ and $\mathbf{w}^{(0)} \in \mathcal{W}$;
2: Compute $\bar{\mathbf{M}}$ in (35) and $\psi(\bar{\mathbf{M}})$;
3: **repeat**
4: Compute $\mathbf{H}^{(k)}$ in (39);
5: Update $\mathbf{w}^{(k+1)}$ by solving
6: 1) the GEVP in (20) for $\mathbf{w} \in \mathcal{W}_0$; or
7: 2) the GTRS in (22) for $\mathbf{w} \in \mathcal{W}_1$;
8: $k = k + 1$;
9: **until** convergence

where $\mathbf{H}^{(k)}$ *is defined as follows:*

$$\begin{aligned}\mathbf{H}^{(k)} \triangleq\, & \xi\mathbf{M}_1 + 2\zeta\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_1\mathbf{w}^{(k)}\right)\mathbf{M}_1 \\ & + 2\eta\sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_i\mathbf{w}^{(k)}\right)\mathbf{M}_i \\ & - 2\psi(\bar{\mathbf{M}})\mathbf{M}_0\mathbf{w}^{(k)}\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_0.\end{aligned} \tag{39}$$

More specifically, for portmanteau statistics $\mathrm{por}(p, \mathbf{w})$ *(i.e.,* $\xi = 0$, $\zeta = 1$ *and* $\eta = 1$*) and penalized crossing statistics* $\mathrm{pcro}(p, \mathbf{w})$ *(i.e.,* $\xi = 1$, $\zeta = 0$ *and* $\eta > 0$*), we have*

$$\mathbf{H}^{(k)} = \begin{cases} 2\sum_{i=1}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_i\mathbf{w}^{(k)}\right)\mathbf{M}_i & \\ -2\psi(\bar{\mathbf{M}})\mathbf{M}_0\mathbf{w}^{(k)}\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_0, & \textit{for } \mathrm{por}(p, \mathbf{w}); \\ \mathbf{M}_1 + 2\eta\sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_i\mathbf{w}^{(k)}\right)\mathbf{M}_i & \\ -2\psi(\bar{\mathbf{M}})\mathbf{M}_0\mathbf{w}^{(k)}\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_0, & \textit{for } \mathrm{pcro}(p, \mathbf{w}). \end{cases}$$

Thus, the majorized problem for problem (32) *is given by*

$$\begin{aligned}\underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T\mathbf{H}^{(k)}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i,\ (i = 0, 1),\end{aligned} \tag{40}$$

Proof: See Appendix B. ■

Lemma 3 shows that the objective function in the majorized problem (40) is a quadratic upperbound of that in the original problem (32). Depending on the specific form of $\mathcal{W}$, subproblem (40) can be efficiently solved for a global optimal solution by using an GEVP or an GTRS problem discussed in Section IV.

Finally, in order to handle the original nonconvex problem (32) directly, we just need to iteratively solve a sequence of QCQPs (i.e., GEVPs or GTRSs). We name these MM-based algorithms iteratively reweighted GEVP (IRGEVP) and iteratively reweighted GTRS (IRGTRS), respectively, which are summarized in Algorithm 3.

C. E-IRGEVP and E-IRGTRS: Solving Algorithms for MRP Design Using $\mathrm{por}(p, \mathbf{w})$ *and* $\mathrm{pcro}(p, \mathbf{w})$

In Section V-B, based on algorithms IRGEVP or IRGTRS, the MRP design problems can be efficiently resolved by solving a nonconvex QCQP at every iteration. However, instead of dealing with a QCQP, it would be much desirable if we could get a closed-form solution for the majorized problem at each iteration. In fact, this target can be attained and the whole procedure is discussed in the following.

Instead of introducing the algorithm derivation from the original problem (32), for simplicity we start from the majorized problem (40) in Section V-B. Problem (40) is equivalent to (53) which is recast as follows:

$$\begin{aligned}\underset{\bar{\mathbf{w}}}{\text{minimize}} \quad & \bar{\mathbf{w}}^T\bar{\mathbf{H}}^{(k)}\bar{\mathbf{w}} \\ \text{subject to} \quad & \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i,\ (i = 0, 1).\end{aligned} \tag{41}$$

Based on Lemma 2, at the $(k+1)$th iteration with iterate $\bar{\mathbf{w}}^{(k)}$, the objective function in (41) (quadratic in $\bar{\mathbf{w}}$) can be majorized by the following majorizing function:

$$\begin{aligned}u_2\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right) =\, & \psi(\bar{\mathbf{H}}^{(k)})\left(\bar{\mathbf{w}}^T\bar{\mathbf{w}}\right) \\ & + 2\left(\bar{\mathbf{w}}^{(k)}\right)^T\left(\bar{\mathbf{H}}^{(k)} - \psi(\bar{\mathbf{H}}^{(k)})\mathbf{I}\right)\bar{\mathbf{w}} \\ & + \left(\bar{\mathbf{w}}^{(k)}\right)^T\left(\psi(\bar{\mathbf{H}}^{(k)})\mathbf{I} - \bar{\mathbf{H}}^{(k)}\right)\bar{\mathbf{w}}^{(k)},\end{aligned} \tag{42}$$

where $\psi(\bar{\mathbf{H}}^{(k)})$ can be chosen as $\left\|\bar{\mathbf{H}}^{(k)}\right\|_F$, and the first and the last terms are constants. Dropping the constants in (42), the majorized problem for (41) is given as follows:

$$\begin{aligned}\underset{\bar{\mathbf{w}}}{\text{minimize}} \quad & \left(\bar{\mathbf{w}}^{(k)}\right)^T\left(\bar{\mathbf{H}}^{(k)} - \psi(\bar{\mathbf{H}}^{(k)})\mathbf{I}\right)\bar{\mathbf{w}} \\ \text{subject to} \quad & \bar{\mathbf{w}}^T\bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i,\ (i = 0, 1).\end{aligned} \tag{43}$$

By changing variable $\bar{\mathbf{w}}$ back to $\mathbf{w}$, we can get the overall majorizing function and the majorized subproblem given in the following lemma.

Lemma 4. *The two majorization steps in* (36) *and* (42) *can be shown as one overall majorization at point* $\mathbf{w}^{(k)}$ *for problem* (32) *with the majorizing function given as follows:*

$$\begin{aligned}\overline{f}_2\left(\mathbf{w}, \mathbf{w}^{(k)}\right) =\, & 2\left(\mathbf{e}^{(k)}\right)^T\mathbf{w} - \left(\mathbf{w}^{(k)}\right)^T\mathbf{H}^{(k)}\mathbf{w}^{(k)} \\ & + 2\psi(\bar{\mathbf{H}}^{(k)})\nu - \zeta\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_1\mathbf{w}^{(k)}\right)^2 \\ & - \eta\sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T\mathbf{M}_i\mathbf{w}^{(k)}\right)^2 + 2\psi(\bar{\mathbf{M}})\nu^2,\end{aligned} \tag{44}$$

where

$$\mathbf{e}^{(k)} \triangleq \left(\mathbf{H}^{(k)} - \psi(\bar{\mathbf{H}}^{(k)})\mathbf{M}_0\right)\mathbf{w}^{(k)}. \tag{45}$$

Thus, the final majorized problem for problem (32) *becomes*

$$\begin{aligned}\underset{\mathbf{w}}{\text{minimize}} \quad & \left(\mathbf{e}^{(k)}\right)^T\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i,\ (i = 0, 1).\end{aligned} \tag{46}$$

Proof: See Appendix C. ■

Lemma 4 shows after using the MM trick twice, the objective function in problem (46) becomes a linear upperbound in variable $\mathbf{w}$ of the original problem (32). By the trick used to get problems (20) and (22), we can eliminate the linear constraint in (46). Then, it becomes a QCLP which has a closed-form solution rather than the QCQP derived from (40).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

Algorithm 4 E-IRGEVP and E-IRGTRS - Algorithms for MRP design problems using $\mathrm{por}(p, \mathbf{w})$ and $\mathrm{pcro}(p, \mathbf{w})$.

Require: p, $\mathbf{M}_i$ with $i = 1, \ldots, p$, and $\nu > 0$.
1: Set $k = 0$ and $\mathbf{w}^{(0)} \in \mathcal{W}$;
2: Compute $\bar{\mathbf{M}}$ in (35) and $\psi(\bar{\mathbf{M}})$;
3: **repeat**
4: Compute $\bar{\mathbf{H}}^{(k)}$ in (52), $\psi(\bar{\mathbf{H}}^{(k)})$, and $\mathbf{e}^{(k)}$ in (45);
5: Update $\mathbf{w}^{(k+1)}$ with a closed-form solution according to Lemma 5;
6: $k = k + 1$;
7: **until** convergence

Lemma 5. *Based on Lagrange duality, problem* (46) *has a closed-form solution. Specifically, for* $\mathbf{w} \in \mathcal{W}_0$,

$$\mathbf{w}^\star = -\left(\frac{\nu}{\left(\mathbf{e}^{(k)}\right)^T \mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)}}\right)^{\frac{1}{2}} \mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)},$$

and for $\mathbf{w} \in \mathcal{W}_1$,

$$\mathbf{w}^\star = -\left(\frac{\nu - \mathbf{w}_0^T\mathbf{M}_0\mathbf{w}_0 + \mathbf{w}_0^T\mathbf{M}_0\mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{M}_0\mathbf{w}_0}{\left(\mathbf{e}^{(k)}\right)^T \mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)}}\right)^{\frac{1}{2}} \times \mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)} - \mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{M}_0\mathbf{w}_0 + \mathbf{w}_0,$$

where $\mathbf{F}$ *satisfies* $\mathbf{1}^T\mathbf{F} = \mathbf{0}$, *and* $\mathbf{w}_0$ *satisfies* $\mathbf{1}^T\mathbf{w}_0 = 1$.

Proof: See Appendix D. ■

Finally, the MRP design problem (32) is solved iteratively by a closed-form update at each iteration. Just to make a connection with IRGEVP and IRGTRS, these algorithms are named extended IRGEVP (E-IRGEVP) and extended IRGTRS (E-IRGTRS) which are summarized in Algorithm 4.

VI. Complexity and Convergence Analysis

A. Complexity Analysis

For Algorithms 1 and 2 (i.e., GEVP and GTRS) in Section IV, the per-iteration computational cost mainly comes from the matrix multiplication with complexity of $\mathcal{O}\left(N^3\right)$. The algorithm converges to the global optimal solution of the original problem (18) or (21). For the MM-based Algorithms 3 and 4 (i.e., IRGEVP, IRGTRS, E-IRGEVP and E-IRGTRS) in Section V, the per-iteration computational cost comes from the Cholesky decomposition or matrix multiplication, so the complexity is still of $\mathcal{O}\left(N^3\right)$.

B. Convergence Analysis

The algorithms IRGEVP and IRGTRS given in Algorithm 3 and algorithms E-IRGEVP and E-IRGTRS given in Algorithm 4 are all based on the general MM method, thus according to Section V-A, we know that the sequence of objective values $\left\{f\left(\mathbf{w}^{(k)}\right)\right\}$ generated by these algorithms is nonincreasing. The original optimization problem (32) is a constrained minimization problem and the objective function f is bounded below, thus the sequence $\left\{f\left(\mathbf{w}^{(k)}\right)\right\}$ is guaranteed to converge to a finite value. Then based on the B-stationarity defined in Section V-A, we can further give the convergence property for the sequence $\left\{\mathbf{w}^{(k)}\right\}$ generated by the MM-based algorithms in the following result.

Proposition 6. *Every limit point, denoted by* $\mathbf{w}^{(\infty)}$, *of the sequence* $\left\{\mathbf{w}^{(k)}\right\}$ *generated by the MM-based algorithms (i.e., Algorithm 3 and Algorithm 4) is a B-stationary point of problem* (32).

Proof: See Appendix E. ■

VII. Numerical Experiments

A statistical arbitrage strategy involves several steps of which the MRP design is a central one. Here, we divide the whole strategy into four sequential steps, namely: assets pool construction, MRP design, unit-root test, and mean reversion trading. In the first step, we select a collection of possibly cointegrated asset candidates to construct an asset pool, on which we will not elaborate in this paper. In the second step, based on the candidate assets from the asset pool, MRPs are designed using either traditional design methods like Engle-Granger OLS method [18] and Johansen method [53] or the proposed methods in this paper. In the third step, unit-root test procedures like Augmented Dickey-Fuller test [54] and Phillips-Perron test [55] are applied to test the stationarity or mean reversion property of the designed MRPs. In the fourth step, MRPs passing the unit-root tests will be traded based on a designed mean reversion trading strategy.

In this section, we first illustrate several performance metrics for the portfolio investment. Then the performance of our proposed MRP design methods in Sections IV and V is evaluated using both synthetic data and real market data are shown accordingly.

A. Performance Metrics

In this paper, we employ the following performance metrics for the numerical experiments.

1) Portfolio Return Measures: In Section II, we have defined the multi-period P&L $\mathrm{P\&L}_t(\tau)$ and single-period P&L $\mathrm{P\&L}_t$. Since there is no trading conducted between two trading periods, the P&L measures (both the multi-period P&L and single-period P&L) are simply defined to be 0. In the following, based on the P&L definition, we give the following useful portfolio return measures.

a) Cumulative P&L: In order to measure the cumulative return performance for an MRP, we define the cumulative P&L (not compounding) in one trading from time t_1 to t_2 as

$$\mathrm{Cum.\ P\&L}\left(t_1, t_2\right) = \sum_{t=t_1}^{t_2} \mathrm{P\&L}_t. \tag{47}$$

b) Return On Investment (ROI): Since different MRPs may have different leverage properties due to $\mathbf{w}_p$, we introduce another portfolio return measure (rate of return) called return on investment (ROI). Within one trading period, the ROI at time t ($t_o \leq t \leq t_c$) is defined to be the single-period P&L at time t normalized by the gross investment deployed which is $\left\|\mathbf{w}_p\right\|_1$ (that is the gross investment exposure to the market

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

including the long position investment and the short position investment) written as

$$\mathrm{ROI}_t = \frac{\mathrm{P\&L}_t}{\|\mathbf{w}_p\|_1}. \tag{48}$$

Like the P&L measures, between two trading periods, ROI_t is defined to be 0.

2) Sharpe Ratio (SR): The Sharpe ratio (SR) [56] is a measure for calculating risk-adjusted return. It describes how much excess return one can receive for the extra volatility (square root of variance). The annualized Sharpe ratio of ROI (or, equivalently, Sharpe ratio of P&L) for a trading stage from time t_1 to t_2 is defined as follows:

$$\mathrm{SR}_{\mathrm{ROI}}(t_1, t_2) = \sqrt{252}\frac{\mu_{\mathrm{ROI}}}{\sigma_{\mathrm{ROI}}}, \tag{49}$$

where $\mu_{\mathrm{ROI}} = 1/(t_2 - t_1)\sum_{t=t_1}^{t_2} \mathrm{ROI}_t$, $\sigma_{\mathrm{ROI}} = \left[1/(t_2 - t_1)\sum_{t=t_1}^{t_2} (\mathrm{ROI}_t - \mu_{\mathrm{ROI}})^2\right]^{1/2}$, t denotes day, and the factor $\sqrt{252}$ relates the daily SR to the annualized SR (assuming 252 trading days per year). In the computation of the SR, we set the risk-free return to be 0, in which case it reduces to the information ratio.

3) Transaction Cost: The transaction or trading costs refer to brokerage commissions, stamp fees, bid-ask spreads, financing costs, and so on. In our experiments, we assume the transaction cost to be fixed as 35 basis points (BPs), i.e., 0.35%, per trade when opening or closing a trading position, then the round-trip transaction cost is 70 BPs.

B. Synthetic Data Experiments

For synthetic data experiments, we generate the sample path of log-prices for M financial assets using a multivariate cointegrated system model [36], where there are r long-run cointegration relations and $M - r$ common trends. We divide the sample path into two stages: in-sample training stage and out-of-sample backtesting or trading stage. All the parameters like spread equilibrium μ_z, trading threshold Δ, and portfolio weight $\mathbf{w}$ are decided in the training stage. The out-of-sample performance of our design methods are tested in the trading stage. In the synthetic experiments, we set $M = 6$ and $r = 5$ and only show the performance of the MRP design methods under net budget constraint $\mathcal{W}_1$. We estimate $N = 5$ spreads using the generated sample path by the OLS and the Johansen method. Based on these five spreads, an MRP is designed as $z_t = \mathbf{w}^T\mathbf{s}_t$. The simulated log-prices and the spreads for the trading stage are shown in Figure 3.

In Figure 4, our proposed problem formulation (denoted as IRGTRS (prop.) and E-IRGTRS (prop.)) is compared to the benchmark formulation in [29] (denoted as SDR (bench.)). To ensure a fair comparison, the net investment budget (i.e., $\mathbf{1}^T\mathbf{w}$) and the variance of the spread (i.e., $\mathbf{w}^T\mathbf{M}_0\mathbf{w}$) are set to be the same for all the methods. From the simulation results, the proposed MRP design problem formulation can attain a lower objective function value. The proposed problem formulation is also solved using the SDR method (denoted as SDR (prop.)) with comparison to the MM-based algorithms (denoted as IRGTRS (prop.) and E-IRGTRS (prop.)) in Figure 4. From the

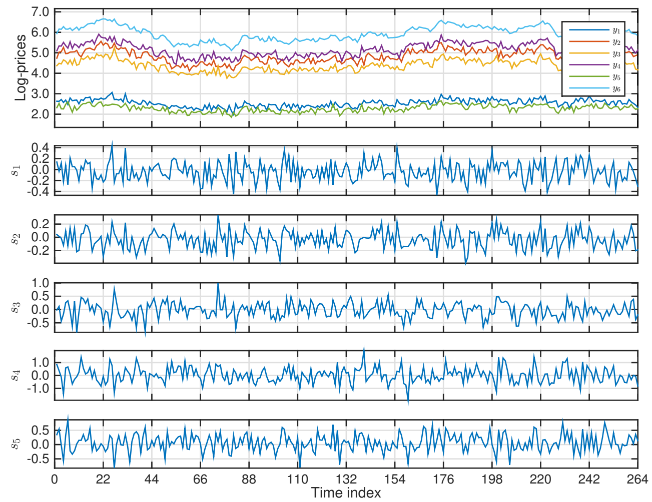

Fig. 3. Log-prices and five estimated spreads. (The sample length for in-sample training is chosen to be $5 \times 12 \times 22$, and the sample length for out-of-sample trading is 12×22.)

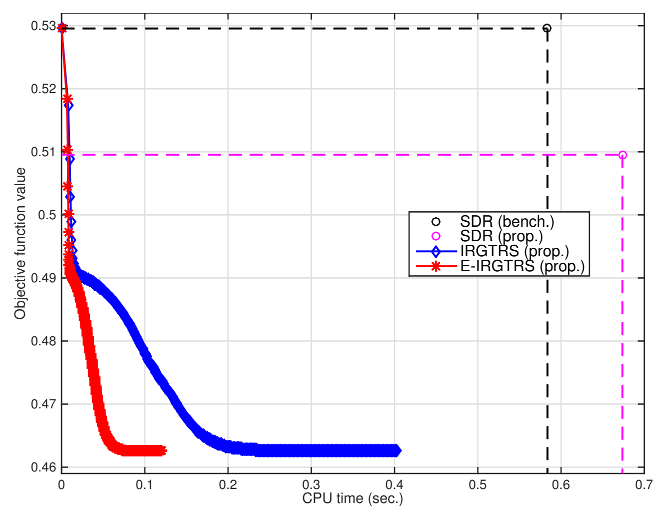

Fig. 4. Numerical convergence of objective function value for $\mathrm{pcro}(5, \mathbf{w})$.

convergence results, the MM-based algorithms are better than the SDR methods in terms of converging solution property and the time.

The performance of the MRPs designed using our proposed methods are compared with those of one underlying spread and the method in [29] based on $\mathrm{pcro}(5, \mathbf{w})$ and $\mathrm{pre}(\mathbf{w})$, which are shown in Figure 5 and Figure 6. From our simulations, we can conclude that our designed MRPs do generate consistent positive profits. And simulation results also show that our designed portfolios can outperform the underlying spreads and the MRPs designed using methods in [29] with higher Sharpe ratios of ROIs and higher cumulative P&Ls.

C. Market Data Experiments

We also test our methods using real market data from the Standard & Poor's 500 (S&P 500) Index, which is usually considered as one of the best representatives for the U.S. stock markets. The data are retrieved from Google

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

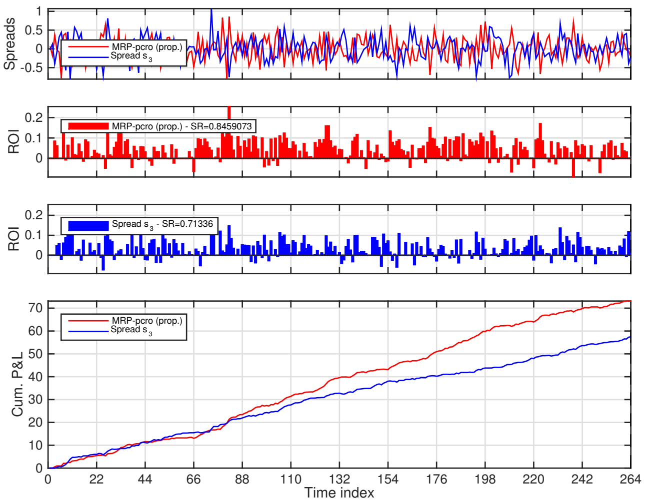

Fig. 5. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP-pcro (prop.) with one underlying spread denoted as Spread s_3.

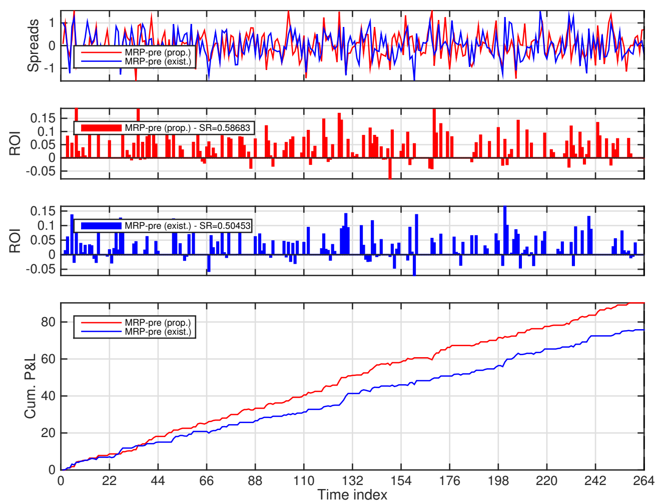

Fig. 6. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP-pre (prop.) and one existing benchmark method in [29] denoted as MRP-pre (exist.).

Finance[8] and adjusted daily closing stock prices are employed. We first choose stock candidates which are possibly cointegrated to form stock asset pools. One stock pool is $\{\mathsf{APA}, \mathsf{AXP}, \mathsf{CAT}, \mathsf{COF}, \mathsf{FCX}, \mathsf{IBM}, \mathsf{MMM}\}$, where the stocks are denoted by their ticker symbols. Three spreads are constructed from this pool based on the Johansen method. Then MRP design methods are employed and unit-root tests are used to test their tradability. The log-prices of the stocks and the log-prices for the three spreads are shown in Figure 7. Based on the mean reversion trading framework mentioned before, one trading experiment is carried out from February 1st, 2012 to June 30th, 2014.

In Figure 8, we compare the performance of our designed MRP with the underlying spread s_1. The log-prices for the designed spreads, and the out-of-sample performance like ROIs, Sharpe ratios of ROIs, and cumulative P&Ls are reported. It is

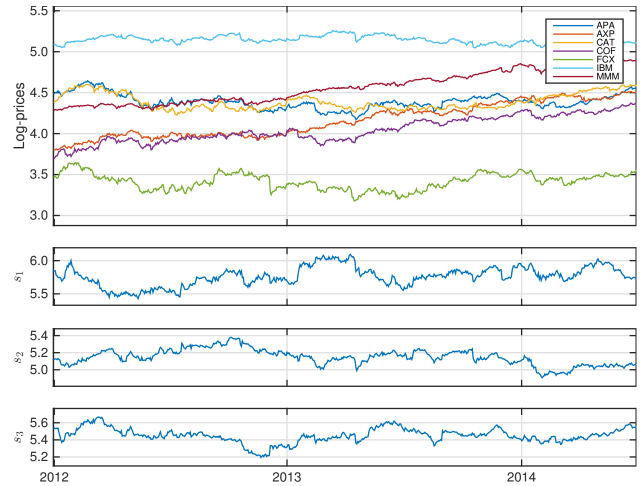

Fig. 7. Log-prices for $\{\mathsf{APA}, \mathsf{AXP}, \mathsf{CAT}, \mathsf{COF}, \mathsf{FCX}, \mathsf{IBM}, \mathsf{MMM}\}$ and three spreads s_1, s_2, and s_3.

shown that using our method, the designed MRP can achieve a higher Sharpe ratio and a better final cumulative return. We also compare our proposed design method with the method in [29] based on the mean reversion criterion $\mathrm{por}(3, \mathbf{w})$ where the investment budget and the portfolio variance are set to be the same. From Figure 9, we can see that our proposed method can outperform the benchmark method through a mean reversion trading design with a higher Sharpe ratio and a higher final cumulative return performance.

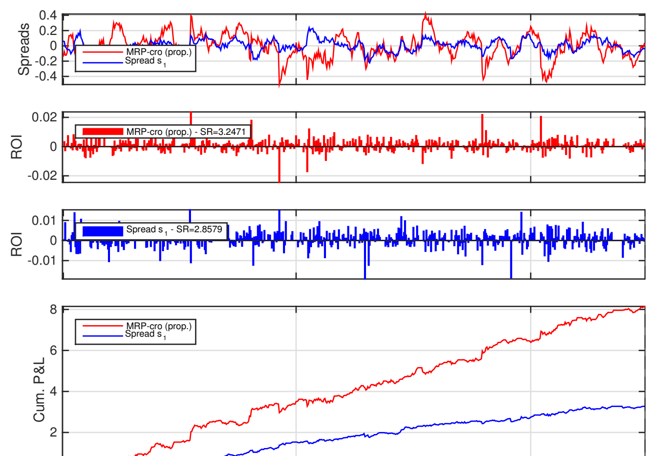

Fig. 8. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP-cro (prop.) with one underlying spread denoted as Spread s_1.

VIII. Conclusions

The mean-reverting portfolio design problem arising from statistical arbitrage has been considered in this paper. We have formulated the MRP design problem as the optimization of a mean reversion criterion characterizing the mean reversion strength of the portfolio and, at the same time, taking into consideration the variance of the portfolio and an investment budget constraint. Several specific optimization problems have

[8]https://www.google.com/finance

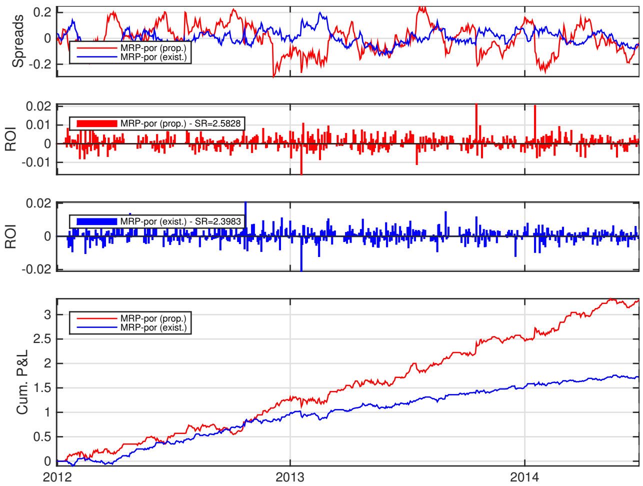

Fig. 9. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP-por (prop.) and one existing benchmark method in [29] denoted as MRP-por (exist.).

been considered based on the general design idea and efficient algorithms have been derived for problem solving. Numerical results show that our proposed methods are able to generate consistent positive profits and outperform the the design methods in literature.

Appendix A
Proof for Lemma 1

Since the spread of an MRP is defined as $z_t = \mathbf{w}_p^T \mathbf{y}_t$, then the multi-period P&L at time t for τ holding periods for a long position on the MRP is given by

$$\begin{aligned}
&\mathrm{P\&L}_t(\tau)\\
=&\mathbf{w}_p^T \mathbf{r}_t (t - t_o) - \mathbf{w}_p^T \mathbf{r}_{t-\tau} (t - \tau - t_o)\\
=&\sum_{m=1}^{M} \left(w_{p,m} r_{m,t} (t - t_o) - w_{p,m} r_{m,t-\tau} (t - \tau - t_o) \right)\\
=&\sum_{m=1}^{M} \left(w_{p,m} \left(\frac{p_{m,t}}{p_{m,t_o}} - 1 \right) - w_{p,m} \left(\frac{p_{m,t-\tau}}{p_{m,t_o}} - 1 \right) \right)\\
\approx&\sum_{m=1}^{M} w_{p,m} \left[\log (p_{m,t}) - \log (p_{m,t_o}) \right]\\
&\qquad - \sum_{m=1}^{M} w_{p,m} \left[\log (p_{m,t-\tau}) - \log (p_{m,t_o}) \right]\\
=&\sum_{m=1}^{M} w_{p,m} y_{m,t} - \sum_{m=1}^{M} w_{p,m} y_{m,t-\tau}\\
=&z_t - z_{t-\tau}.
\end{aligned}$$

Similarly, for a short position on the MRP, the $\mathrm{P\&L}_t(\tau)$ is computed as $z_{t-\tau} - z_t$.

Appendix B
Proof for Lemma 3

It is easy to see that, based on Lemma 2, only the second term of $f(\mathbf{w})$ in problem (32) is majorized. Then the overall majorizing function for $f(\mathbf{w})$ at $\mathbf{w}^{(k)}$ can be attained through replacing the second term by its majorizing function.

Replacing the the second term in the objective function of problem (34) by $u_1\left(\bar{\mathbf{W}}, \bar{\mathbf{W}}^{(k)}\right)$ in (36) and substituting $\bar{\mathbf{M}}$ in (35) back into the function, we get the following overall majorizing function in variable $\bar{\mathbf{W}}$ as follows:

$$\begin{aligned}
&\bar{f}\left(\bar{\mathbf{W}}, \bar{\mathbf{W}}^{(k)}\right) = \xi \mathrm{vec}\left(\bar{\mathbf{M}}_1\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}\right) + u_1\left(\bar{\mathbf{W}}, \bar{\mathbf{W}}^{(k)}\right)\\
&= \xi \mathrm{vec}\left(\bar{\mathbf{M}}_1\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}\right) + \psi(\bar{\mathbf{M}}) \mathrm{vec}\left(\bar{\mathbf{W}}\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}\right)\\
&+ 2\zeta \left[\mathrm{vec}\left(\bar{\mathbf{M}}_1\right) \mathrm{vec}\left(\bar{\mathbf{M}}_1\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right) \right]^T \mathrm{vec}\left(\bar{\mathbf{W}}\right)\\
&+ 2\eta \left[\sum_{i=2}^{p} \mathrm{vec}\left(\bar{\mathbf{M}}_i\right) \mathrm{vec}\left(\bar{\mathbf{M}}_i\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right) \right]^T \mathrm{vec}\left(\bar{\mathbf{W}}\right)\\
&- 2\psi(\bar{\mathbf{M}}) \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}\right)\\
&+ \psi(\bar{\mathbf{M}}) \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)\\
&- \zeta \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \mathrm{vec}\left(\bar{\mathbf{M}}_1\right) \mathrm{vec}\left(\bar{\mathbf{M}}_1\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)\\
&- \eta \sum_{i=2}^{p} \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right)^T \mathrm{vec}\left(\bar{\mathbf{M}}_i\right) \mathrm{vec}\left(\bar{\mathbf{M}}_i\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}^{(k)}\right).
\end{aligned} \tag{50}$$

Then, by undoing the matrix lifting, i.e., changing variable $\bar{\mathbf{W}}$ back to $\bar{\mathbf{w}}$, and using $\mathrm{vec}\left(\bar{\mathbf{M}}_i\right)^T \mathrm{vec}\left(\bar{\mathbf{W}}\right) = \mathrm{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{W}}\right) = \bar{\mathbf{w}}^T \bar{\mathbf{M}}_i \bar{\mathbf{w}}$, we can get the majorizing function in $\bar{\mathbf{w}}$ given by

$$\begin{aligned}
&\bar{f}_1\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right) = \xi \mathrm{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{w}} \bar{\mathbf{w}}^T\right) + \psi(\bar{\mathbf{M}}) \mathrm{Tr}\left(\bar{\mathbf{w}} \bar{\mathbf{w}}^T \bar{\mathbf{w}} \bar{\mathbf{w}}^T\right)\\
&+ 2\zeta \mathrm{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T\right) \mathrm{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{w}} \bar{\mathbf{w}}^T\right)\\
&+ 2\eta \sum_{i=2}^{p} \left[\mathrm{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T\right) \mathrm{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{w}} \bar{\mathbf{w}}^T\right) \right]\\
&- 2\psi(\bar{\mathbf{M}}) \mathrm{Tr}\left(\bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}} \bar{\mathbf{w}}^T\right)\\
&+ \psi(\bar{\mathbf{M}}) \mathrm{Tr}\left(\bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T\right)\\
&- \zeta \mathrm{Tr}\left(\bar{\mathbf{M}}_1 \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T\right)^2 - \eta \sum_{i=2}^{p} \mathrm{Tr}\left(\bar{\mathbf{M}}_i \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T\right)^2\\
&= \bar{\mathbf{w}}^T \bar{\mathbf{H}}^{(k)} \bar{\mathbf{w}} + \psi(\bar{\mathbf{M}}) \left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right)^2 + \psi(\bar{\mathbf{M}}) \left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}}^{(k)}\right)^2\\
&- \zeta \left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_1 \bar{\mathbf{w}}^{(k)}\right)^2 - \eta \sum_{i=2}^{p} \left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_i \bar{\mathbf{w}}^{(k)}\right)^2,
\end{aligned} \tag{51}$$

where in the objective function, $\bar{\mathbf{H}}^{(k)}$ is defined as follows:

$$\begin{aligned}
\bar{\mathbf{H}}^{(k)} \triangleq\ & \xi \bar{\mathbf{M}}_1 + 2\zeta \left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_1 \bar{\mathbf{w}}^{(k)}\right) \bar{\mathbf{M}}_1\\
&+ 2\eta \sum_{i=2}^{p} \left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_i \bar{\mathbf{w}}^{(k)}\right) \bar{\mathbf{M}}_i\\
&- 2\psi(\bar{\mathbf{M}}) \bar{\mathbf{w}}^{(k)} \left(\bar{\mathbf{w}}^{(k)}\right)^T.
\end{aligned} \tag{52}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

Dropping the constants in $\overline{f}_1\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right)$, problem (37) becomes

$$\begin{array}{ll} \underset{\bar{\mathbf{w}}}{\mathsf{minimize}} & \bar{\mathbf{w}}^T \bar{\mathbf{H}}^{(k)} \bar{\mathbf{w}} \\ \mathsf{subject\ to} & \bar{\mathbf{w}}^T \bar{\mathbf{w}} = \nu \\ & \bar{\mathbf{w}} \in \bar{\mathcal{W}}_i,\ (i = 0, 1)\,. \end{array} \tag{53}$$

From $\overline{f}_1\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right)$, by changing variable $\bar{\mathbf{w}}$ back to $\mathbf{w}$ based on $\bar{\mathbf{w}} = \mathbf{L}^T \mathbf{w}$ and considering the constraint $\mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu$, we have the majoring function in variable $\mathbf{w}$ given as follows:

$$\begin{aligned} &\overline{f}_1\left(\mathbf{w}, \mathbf{w}^{(k)}\right) = \mathbf{w}^T \mathbf{H}^{(k)} \mathbf{w} + \psi(\bar{\mathbf{M}})\left(\mathbf{w}^T \mathbf{M}_0 \mathbf{w}\right)^2 \\ &+ \psi(\bar{\mathbf{M}})\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_0 \mathbf{w}^{(k)}\right)^2 - \zeta\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)^2 \\ &- \eta \sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)^2 \\ &= \mathbf{w}^T \mathbf{H}^{(k)} \mathbf{w} + 2\psi(\bar{\mathbf{M}})\nu^2 - \zeta\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)^2 \\ &- \eta \sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)^2, \end{aligned} \tag{54}$$

where $\mathbf{H}^{(k)}$ in the objective function is given by

$$\begin{aligned} \mathbf{H}^{(k)} \triangleq\ & \xi \mathbf{M}_1 + 2\zeta\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)\mathbf{M}_1 \\ & + 2\eta \sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)\mathbf{M}_i \\ & - 2\psi(\bar{\mathbf{M}})\mathbf{M}_0 \mathbf{w}^{(k)}\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_0. \end{aligned}$$

Finally, based on $\overline{f}_1\left(\mathbf{w}, \mathbf{w}^{(k)}\right)$, the majorized problem is given as follows:

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & \mathbf{w}^T \mathbf{H}^{(k)} \mathbf{w} \\ \mathsf{subject\ to} & \mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i,\ (i = 0, 1)\,. \end{array} \tag{55}$$

Appendix C
Proof for Lemma 4

The proof is similar to that for Lemma 3. Since the majorization step in Lemma 4 can be regarded as a second majorization for the majorizing function given in Lemma 3 (i.e., $\overline{f}_2\left(\mathbf{w}, \mathbf{w}^{(k)}\right) \geq \overline{f}_1\left(\mathbf{w}, \mathbf{w}^{(k)}\right)$), we can just start the proof from the first majorizing function in (51).

First, replacing the first term of $\overline{f}_1\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right)$ by its majorizing function $u_2\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right)$ in (42), we have

$$\begin{aligned} &\overline{f}_2\left(\bar{\mathbf{w}}, \bar{\mathbf{w}}^{(k)}\right) = \psi(\bar{\mathbf{H}}^{(k)})\left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right) + 2\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{H}}^{(k)} \bar{\mathbf{w}} \\ &- 2\psi(\bar{\mathbf{H}}^{(k)})\left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}}\right) + \psi(\bar{\mathbf{H}}^{(k)})\left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}}\right) \\ &- \left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{H}}^{(k)} \bar{\mathbf{w}}^{(k)} + \psi(\bar{\mathbf{M}})\left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right)^2 \\ &+ \psi(\bar{\mathbf{M}})\left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{w}}^{(k)}\right)^2 - \zeta\left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_1 \bar{\mathbf{w}}^{(k)}\right)^2 \\ &- \eta \sum_{i=2}^{p}\left(\left(\bar{\mathbf{w}}^{(k)}\right)^T \bar{\mathbf{M}}_i \bar{\mathbf{w}}^{(k)}\right)^2. \end{aligned} \tag{56}$$

Then, we change the variable $\bar{\mathbf{w}}$ back to $\mathbf{w}$, consider the constraint $\mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu$, and get the majoring function in variable $\mathbf{w}$ as follows:

$$\begin{aligned} &\overline{f}_2\left(\mathbf{w}, \mathbf{w}^{(k)}\right) = \psi(\bar{\mathbf{H}}^{(k)})\left(\mathbf{w}^T \mathbf{M}_0 \mathbf{w}\right) + 2\left(\mathbf{w}^{(k)}\right)^T \mathbf{H}^{(k)} \mathbf{w} \\ &- 2\psi(\bar{\mathbf{H}}^{(k)})\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_0 \mathbf{w}\right) + \psi(\bar{\mathbf{H}}^{(k)})\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_0 \mathbf{w}\right) \\ &- \left(\mathbf{w}^{(k)}\right)^T \mathbf{H}^{(k)} \mathbf{w}^{(k)} + \psi(\bar{\mathbf{M}})\left(\mathbf{w}^T \mathbf{M}_0 \mathbf{w}\right)^2 \\ &+ \psi(\bar{\mathbf{M}})\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_0 \mathbf{w}^{(k)}\right)^2 - \zeta\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)^2 \\ &- \eta \sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)^2 \\ &= 2\left(\mathbf{e}^{(k)}\right)^T \mathbf{w} - \left(\mathbf{w}^{(k)}\right)^T \mathbf{H}^{(k)} \mathbf{w}^{(k)} + 2\psi(\bar{\mathbf{H}}^{(k)})\nu \\ &- \zeta\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_1 \mathbf{w}^{(k)}\right)^2 - \eta \sum_{i=2}^{p}\left(\left(\mathbf{w}^{(k)}\right)^T \mathbf{M}_i \mathbf{w}^{(k)}\right)^2 \\ &+ 2\psi(\bar{\mathbf{M}})\nu^2 \end{aligned} \tag{57}$$

where

$$\mathbf{e}^{(k)} \triangleq \left(\mathbf{H}^{(k)} - \psi(\bar{\mathbf{H}}^{(k)})\mathbf{M}_0\right)\mathbf{w}^{(k)}.$$

Finally, the majorized subproblem is accordingly given in the following way:

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & \left(\mathbf{e}^{(k)}\right)^T \mathbf{w} \\ \mathsf{subject\ to} & \mathbf{w}^T \mathbf{M}_0 \mathbf{w} = \nu \\ & \mathbf{w} \in \mathcal{W}_i,\ (i = 0, 1)\,. \end{array} \tag{58}$$

Appendix D
Proof for Lemma 5

We show the proof for the case $\mathbf{w} \in \mathcal{W}_1$, and the other case follows accordingly. We first check the regularity conditions (or constraint qualifications). Problem (46) is equivalent to the following convex problem

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & \left(\mathbf{e}^{(k)}\right)^T \mathbf{w} \\ \mathsf{subject\ to} & \mathbf{w}^T \mathbf{M}_0 \mathbf{w} \leq \nu \\ & \mathbf{1}^T \mathbf{w} = 1, \end{array} \tag{59}$$

since the objective is linear and the optimal solution $\mathbf{w}^\star$ is always attained in the boundary of the quadratic constraint set. Slater's regularity condition holds for (59), i.e., it is strictly feasible. By variable changing $\mathbf{w} = \mathbf{F}\mathbf{x} + \mathbf{w}_0$, with $\mathbf{N}_0 = \mathbf{F}^T \mathbf{M}_0 \mathbf{F}$, $\mathbf{p}_0 = \mathbf{F}^T \mathbf{M}_0 \mathbf{w}_0$, and $b_0 = \mathbf{w}_0^T \mathbf{M}_0 \mathbf{w}_0$, we have

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & \left(\mathbf{e}^{(k)}\right)^T \mathbf{F}\mathbf{x} \\ \mathsf{subject\ to} & \mathbf{x}^T \mathbf{N}_0 \mathbf{x} + 2\mathbf{p}_0^T \mathbf{x} + b_0 \leq \nu. \end{array} \tag{60}$$

The Karush-Kuhn-Tucker (KKT) conditions for primal and dual variable pair $(\mathbf{x}^\star, \lambda^\star)$ can be written as

$$\begin{cases} 2\lambda^\star \mathbf{N}_0 \mathbf{x}^\star + 2\lambda^\star \mathbf{p}_0 + \mathbf{F}^T \mathbf{e}^{(k)} = 0, \\ \mathbf{x}^{\star T} \mathbf{N}_0 \mathbf{x}^\star + 2\mathbf{p}_0^T \mathbf{x}^\star + b_0 \leq \nu, \\ \lambda^\star \geq 0, \\ \lambda^\star\left(\mathbf{x}^{\star T} \mathbf{N}_0 \mathbf{x}^\star + 2\mathbf{p}_0^T \mathbf{x}^\star + b_0 - \nu\right) = 0. \end{cases}$$

By solving the KKT conditions, we have

$$\begin{aligned}&\mathbf{x}^\star = \\ &-\left(\frac{\nu - \mathbf{w}_0^T\mathbf{M}_0\mathbf{w}_0 + \mathbf{w}_0^T\mathbf{M}_0\mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{M}_0\mathbf{w}_0}{\left(\mathbf{e}^{(k)}\right)^T\mathbf{F}\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)}}\right)^{\frac{1}{2}} \\ &\times\left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{e}^{(k)} - \left(\mathbf{F}^T\mathbf{M}_0\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{M}_0\mathbf{w}_0,\end{aligned}$$

and accordingly have $\mathbf{w}^\star = \mathbf{F}\mathbf{x}^\star + \mathbf{w}_0$.

Appendix E
Proof for Proposition 6

From the derivation of the MM-based algorithms (IRGEVP and IRGTRS in Algorithm 3 as well as E-IRGEVP and E-IRGTRS in Algorithm 4), we know that the objective function $f(\mathbf{w})$ in (32) is majorized by functions $\overline{f}_1\left(\mathbf{w},\mathbf{w}^{(k)}\right)$ in Lemma 3 and $\overline{f}_2\left(\mathbf{w},\mathbf{w}^{(k)}\right)$ in Lemma 4 at $\mathbf{w}^{(k)}$ over the constraint $\mathcal{W} = \left\{\mathbf{w}^T\mathbf{M}_0\mathbf{w} = \nu\right\} \cap \mathcal{W}_i$, $(i = 0, 1)$. In the following, for the purpose of easy explanation, $\overline{f}_1\left(\mathbf{w},\mathbf{w}^{(k)}\right)$ and $\overline{f}_2\left(\mathbf{w},\mathbf{w}^{(k)}\right)$ will be jointly denoted as $\overline{f}\left(\mathbf{w},\mathbf{w}^{(k)}\right)$.

Based on (27) and (28) in Section V-A, we can get the objective function value is monotonically nonincreasing at each iteration, i.e.,

$$\begin{aligned}f\left(\mathbf{w}^{(k+1)}\right) &\overset{(a)}{\leq} \overline{f}\left(\mathbf{w}^{(k+1)},\mathbf{w}^{(k)}\right) \\ &\overset{(b)}{\leq} \overline{f}\left(\mathbf{w}^{(k)},\mathbf{w}^{(k)}\right) \overset{(c)}{=} f\left(\mathbf{w}^{(k)}\right), \quad \forall k \in \mathbb{N},\end{aligned}$$

where (a) and (c) follow from the A2) and A1) in (28), respectively, and (b) follows from (27). It implies $\left\{f\left(\mathbf{w}^{(k)}\right)\right\}$ is a nonincreasing sequence, i.e.,

$$f\left(\mathbf{w}^{(0)}\right) \geq f\left(\mathbf{w}^{(1)}\right) \geq f\left(\mathbf{w}^{(2)}\right) \geq \ldots.$$

Assume that there exists a subsequence $\left\{\mathbf{w}^{(k_j)}\right\}$ converging to a limit point $\mathbf{w}^{(\infty)}$. We first have

$$\begin{aligned}&\overline{f}\left(\mathbf{w}^{(k_{j+1})},\mathbf{w}^{(k_{j+1})}\right) = f\left(\mathbf{w}^{(k_{j+1})}\right) \leq f\left(\mathbf{w}^{(k_j+1)}\right) \\ &\leq \bar{f}\left(\mathbf{w}^{(k_j+1)},\mathbf{w}^{(k_j)}\right) \leq \bar{f}\left(\mathbf{w},\mathbf{w}^{(k_j)}\right), \forall \mathbf{w} \in \mathcal{W}.\end{aligned}$$

Letting $j \to \infty$, we can further obtain

$$\bar{f}\left(\mathbf{w}^{(\infty)},\mathbf{w}^{(\infty)}\right) \leq \bar{f}\left(\mathbf{w},\mathbf{w}^{(\infty)}\right), \forall \mathbf{w} \in \mathcal{W},$$

i.e., $\mathbf{w}^{(\infty)}$ is the global minimizer of $\bar{f}\left(\mathbf{w},\mathbf{w}^{(\infty)}\right)$ over $\mathcal{W}$. Based on the B-stationarity defined in Section V-A, we have

$$\bar{f}'\left(\mathbf{w}^{(\infty)},\mathbf{w}^{(\infty)};\mathbf{d}\right) \geq 0, \forall \mathbf{d} \in \mathcal{T}_{\mathcal{W}}\left(\mathbf{w}^{(\infty)}\right).$$

Then, according to the A3) in (28), we have

$$f'\left(\mathbf{w}^{(\infty)};\mathbf{d}\right) \geq 0, \forall \mathbf{d} \in \mathcal{T}_{\mathcal{W}}\left(\mathbf{w}^{(\infty)}\right),$$

which implies $\mathbf{w}^{(\infty)}$ is a B-stationary point of problem (32).

References

[1] Z. Zhao and D. P. Palomar, "Mean-reverting portfolio design via majorization-minimization method," in *Proc. the 50th Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2016, pp. 1530–1534.

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

[3] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: Performance of a relative-value arbitrage rule," *Review of Financial Studies*, vol. 19, no. 3, pp. 797–827, 2006.

[4] D. S. Ehrman, *The Handbook of Pairs Trading: Strategies using equities, options, and futures*. John Wiley & Sons, 2006, vol. 240.

[5] R. Bookstaber, *A Demon of Our Own Design: Markets, hedge funds, and the perils of financial innovation*. John Wiley & Sons, 2007.

[6] D. Butterworth and P. Holmes, "Inter-market spread trading: Evidence from UK index futures markets," *Applied Financial Economics*, vol. 12, no. 11, pp. 783–790, 2002.

[7] S.-J. Kim, J. Primbs, and S. Boyd, "Dynamic spread trading," *Unpublished Working Paper*, 2008.

[8] T. Kanamura, S. T. Rachev, and F. J. Fabozzi, "A profit model for spread trading with an application to energy futures," *The Journal of Trading*, vol. 5, no. 1, pp. 48–62, 2010.

[9] M. Cummins and A. Bucca, "Quantitative spread trading on crude oil and refined products markets," *Quantitative Finance*, vol. 12, no. 12, pp. 1857–1875, 2012.

[10] M. Whistler, *Trading Pairs: Capturing profits and hedging risk with statistical arbitrage strategies*. John Wiley & Sons, 2004, vol. 216.

[11] C. Alexander and A. Dimitriu, "Indexing and statistical arbitrage," *The Journal of Portfolio Management*, vol. 31, no. 2, pp. 50–63, 2005.

[12] A. Pole, *Statistical Arbitrage: Algorithmic trading insights and techniques*. John Wiley & Sons, 2011, vol. 411.

[13] S. F. LeRoy and J. Werner, *Principles of Financial Economics*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[14] J. G. Nicholas, *Market Neutral Investing: Long/Short hedge fund strategies*. Bloomberg Press, 2000.

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

[16] C. Krauss, "Statistical arbitrage pairs trading strategies: Review and outlook," *Journal of Economic Surveys*, vol. 31, no. 2, pp. 513–545, 2017.

[17] C. W. Granger, "Cointegrated variables and error correction models," Unpublished USCD Discussion Paper 83-13a, Tech. Rep., 1983.

[18] 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.

[19] S. Johansen, "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models," *Econometrica: Journal of the Econometric Society*, pp. 1551–1580, 1991.

[20] Z. Zhao and D. P. Palomar, "Robust maximum likelihood estimation of sparse vector error correction model," in *Proc. the 2017 5th IEEE Global Conf. on Signal and Information Processing,*, Montreal, QB, Canada, Nov. 2017, pp. 913–917.

[21] P. Draper and J. K. Fung, "A study of arbitrage efficiency between the FTSE-100 index futures and options contracts," *Journal of Futures Markets*, vol. 22, no. 1, pp. 31–58, 2002.

[22] G. Hong and R. Susmel, "Pairs-trading in the Asian ADR market," *University of Houston, Unpublished Manuscript*, 2003.

[23] M. S. Perlin, "Evaluation of pairs-trading strategy at the Brazilian financial market," *Journal of Derivatives & Hedge Funds*, vol. 15, no. 2, pp. 122–136, 2009.

[24] M. Avellaneda and J.-H. Lee, "Statistical arbitrage in the US equities market," *Quantitative Finance*, vol. 10, no. 7, pp. 761–782, 2010.

[25] S. Drakos, "Statistical arbitrage in S&P500," *Journal of Mathematical Finance*, vol. 6, no. 01, p. 166, 2016.

[26] J. L. Farrell and W. J. Reinhart, *Portfolio Management: Theory and application*. McGraw-Hill, 1997.

[27] H. M. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

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

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

[30] ——, "Mean-reverting portfolios," in *Financial Signal Processing and Machine Learning*, A. N. Akansu, S. R. Kulkarni, and D. M. Malioutov, Eds. John Wiley & Sons, 2016, ch. 3, pp. 23–40.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSP.2018.2799193, IEEE Transactions on Signal Processing

[31] F. J. Fabozzi, S. M. Focardi, and P. N. Kolm, *Quantitative Equity Investing: Techniques and strategies*. John Wiley & Sons, 2010.

[32] Y. Huang and D. P. Palomar, "Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1093–1108, Jan. 2014.

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

[34] H. M. Markowitz, "The optimization of a quadratic function subject to linear constraints," *Naval Research Logistics Quarterly*, vol. 3, no. 1-2, pp. 111–133, 1956.

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

[36] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2007.

[37] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *J. Amer. Statist. Assoc.*, vol. 65, no. 332, pp. 1509–1526, 1970.

[38] N. D. Ylvisaker, "The expected number of zeros of a stationary gaussian process," *Ann. Math. Statist.*, vol. 36, no. 3, pp. 1043–1046, 1965.

[39] B. Kedem and S. Yakowitz, *Time Series Analysis by Higher Order Crossings*. Piscataway, NJ, USA: IEEE Press, 1994.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[41] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 3, pp. 844–860, 2006.

[42] Y. Huang and D. P. Palomar, "Rank-constrained separable semidefinite programming with applications to optimal beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 664–678, Sep. 2010.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[44] J. Song, P. Babu, and D. P. Palomar, "Sparse generalized eigenvalue problem via smooth optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1627–1642, Jan. 2015.

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

[46] T. K. Pong and H. Wolkowicz, "The generalized trust region subproblem," *Computational Optimization and Applications*, vol. 58, no. 2, pp. 273–322, 2014.

[47] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.

[48] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Aug. 2016.

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

[50] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Dec. 2016.

[51] J. Pang, "Partially B-regular optimization and equilibrium problems," *Math. Oper. Res.*, vol. 32, no. 3, pp. 687–699, 2007.

[52] J.-S. Pang, M. Razaviyayn, and A. Alvarado, "Computing B-stationary points of nonsmooth DC programs," *Math. Oper. Res.*, vol. 42, no. 1, pp. 95–118, Feb. 2017.

[53] S. Johansen, "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, vol. 12, no. 2, pp. 231–254, 1988.

[54] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Statist. Assoc.*, vol. 74, no. 366a, pp. 427–431, 1979.

[55] P. C. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988.

[56] W. F. Sharpe, "The Sharpe ratio," *The Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, 1994.

Ziping Zhao (S'16) received the B.Eng. degree in Electronics and Information Engineering (with highest honor) from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2014. Currently, he is pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong. He was an awardee of the Hong Kong PhD Fellowship Scheme (HKPFS).

His research interests are in optimization, machine learning, and signal processing methods, with applications in data analytics, financial engineering, and communication networks.

Daniel P. Palomar (S'99-M'03-SM'08-F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively, and was a Fulbright at Princeton University.

He is a Professor in the Department of Electronic and Computer Engineering at the Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. Since 2013 he is a Fellow of the Institute for Advance Study (IAS) at HKUST. He had previously held several research appointments, namely, at King's College London (KCL), London, UK; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona, Spain; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome "La Sapienza", Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory and signal processing to financial systems, big data systems, and communication systems.

Dr. Palomar is an IEEE Fellow, a recipient of a 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (co-author) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015-16 HKUST Excellence Research Award, the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC), the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.

He has been a Guest Editor of the IEEE Journal of Selected Topics in Signal Processing 2016 Special Issue on "Financial Signal Processing and Machine Learning for Electronic Trading", an Associate Editor of IEEE Transactions on Information Theory and of IEEE Transactions on Signal Processing, a Guest Editor of the IEEE Signal Processing Magazine 2010 Special Issue on "Convex Optimization for Signal Processing," the IEEE Journal on Selected Areas in Communications 2008 Special Issue on "Game Theory in Communication Systems," and the IEEE Journal on Selected Areas in Communications 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks."