Optimal Mean-Reverting Portfolio With Leverage Constraint for Statistical Arbitrage in Finance

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

***Abstract*—The optimal mean-reverting portfolio (MRP) design problem is an important task for statistical arbitrage, also known as pairs trading, in the financial markets. The target of the problem is to construct a portfolio of the underlying assets (possibly with an asset selection target) that can exhibit a satisfactory mean reversion property and a desirable variance property. In this paper, the optimal MRP design problem is studied under an investment leverage constraint representing the total investment positions on the underlying assets. A general problem formulation is proposed by considering the design targets subject to a leverage constraint. To solve the problem, a unified optimization framework based on the successive convex approximation method is developed. The superior performance of the proposed formulation and the algorithms are verified through numerical simulations on both synthetic data and real market data.**

***Index Terms*—Portfolio optimization, pairs trading, mean reversion strategy, cointegration, algorithmic trading, quantitative trading, asset selection, leverage constraint, nonconvex optimization, sparse optimization, successive convex approximation.**

I. Introduction

STATISTICAL arbitrage [2], also known as *Stat Arb*, is a quantitative investment and trading strategy widely used by many parties in the financial markets, e.g., institutional investors, hedge funds, mutual funds, proprietary trading firms, and individual investors [3]. In statistical arbitrage, the trading basket usually consists of many financial assets of possibly different categories such as equities, options, bonds, futures, commodities, etc. To arbitrage from the markets, investors need to buy the under-priced assets and short-sell or, more plainly, borrow and sell the over-priced ones. The profits will finally be locked in by unwinding the trading positions when the mispricings of the assets correct themselves in the future. Such an investment strategy is usually coined as a contrarian relative-value strategy [4]. In statistical arbitrage, the arbitrage opportunities exist as a consequence of the market inefficiency [5]. As revealed by the name, the design of trading baskets and trading actions largely relies on statistical analysis [6].

Statistical arbitrage dated back to the well-known trading strategy called pairs trading [7]–[9], which was firstly developed at Morgan Stanley by a quantitative trading group under the lead of Nunzio Tartaglia in the mid 1980s in the Wall Street [10]. Pairs trading, as a special scenario, falls into the umbrella of statistical arbitrage and, as indicated by the name, it is often used when there are only two assets in the trading basket. Since statistical arbitrage is able to hedge the overall market or systematic risk, and hence the profits are independent from the movements and the conditions of the prevailing markets (volatile, flat, or falling), it is also named as a market neutral strategy or an absolute return strategy [11], [12].

In statistical arbitrage, the trading basket is used to design a "spread" which characterizes the mis-pricings (also called the "relative pricing") of the underlying assets. The designed spread is stationary, hence mean-reverting, and virtually represents the price for a synthetic mean-reverting asset [13]. In order to make profits, the trading process is carried out based on the mean reversion (MR) behavior of the spread around its statistical equilibrium, and hence named mean reversion trading or convergence trading [14]. For example, a simple mean reversion trading design could be buying the spread when it is below the equilibrium and selling it when it is above the equilibrium. Statistical arbitrage or pairs trading is accordingly also referred to as spread trading in the literature [15]–[17]. In practice, there are many existing methods to design a trading spread based on different philosophies, such as the distance method [18], the cointegration analysis method [7], the factor analysis method [19], the Copulas method [20], the stochastic modeling method [21], [22], and so on. In this paper, we will only focus on the cointegration analysis method where the spread is constructed by a formal time series analysis [23]. The concept of "cointegration" was first come up with by Clive W. J. Granger in [24] and later in [25] to describe the linear stationary relationships within nonstationary time series which are named to be cointegrated. Later, the cointegrated vector autoregressive model was put forward into time series modeling [26], [27] to efficiently estimate the cointegration relations. To honor the discovery of cointegration statistical property in time series, Granger was awarded the Nobel Prize in Economic Sciences in 2003. The cointegration relations have been verified by empirical analyses in many different financial markets to get statistical arbitrage opportunities [28]–[30].

Manuscript received September 2, 2018; revised December 11, 2018; accepted January 5, 2019. Date of publication January 21, 2019; date of current version February 12, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Byonghyo Shim. This work was supported by the Hong Kong RGC 16208917 Research Grant. The work of Z. Zhao was supported by the Hong Kong Ph.D. Fellowship Scheme. This paper was presented in part at the 20th IEEE Statistical Signal Processing Workshop, Freiburg, Germany, June 10–13, 2018 [1]. *(Corresponding author: Daniel P. Palomar.)*

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

Digital Object Identifier 10.1109/TSP.2019.2893862

Traditionally, cointegration analysis methods like the Engle-Granger ordinary least squares (OLS) method [25] and the Johansen method [26] are used to estimate the trading spreads from the underlying assets. An asset that naturally shows stationarity is a spread as well [31], e.g., the option implied volatility for stocks. Inherent correlations, however, may exist among different spreads. For example, when using the Johansen method, although many distinct spreads could be estimated from the same underlying assets, they essentially fall into the same "cointegration space". When having multiple spreads, a natural and interesting question is put forward: Can we design an optimal portfolio of these underlying assets? This question will be addressed in this paper. A portfolio of the mean-reverting spreads is named a mean-reverting portfolio (MRP) or sometimes a long-short portfolio. To design an optimal MRP, two factors should be considered. Firstly, the designed MRP should exhibit a strong MR property so that it has frequent mean-crossing points and hence brings in more trading opportunities. Secondly, the designed MRP should exhibit large variance property so that each trade can provide sufficient profit. These two factors together naturally result in a multi-objective optimization problem, i.e., to find a desirable trade-off between MR and variance.

In [32], the author first proposed to design an MRP by optimizing a criterion characterizing the mean reversion strength. Later, authors in [33] and [34] realized that directly solving the MRP design problem in [32] could result in a portfolio with very low variance, then the variance control was taken into consideration and several new mean reversion criteria were also brought up. However, all the aforementioned MRP design problems were carried out by imposing an ℓ_2-norm constraint on the portfolio weights. In [35], the authors argued that the ℓ_2-norm has a physical meaning of power constraint in many signal processing problems (like beamforming in wireless communications), but its practical significance in the financial context is unclear. As a result, the investment budget constraint (a linear constraint) was firstly proposed in [35] and then in [13]. Compared to [33] and [34], the proposed methods in [13] make the designed portfolio more explainable and practical, in a sense that it explicitly represents the budget allocation on different underlying assets. However, one prominent issue incurred by the methods of using investment budget constraints as in [13] is that the designed portfolio could lead to a very large leverage (i.e., the total dollar position, both in longs and shorts), which makes the methods not always acceptable for practical use in real investments.

In this paper, we are going to propose a new formulation for the optimal MRP design problem by jointly optimizing the two factors (i.e., MR and variance) subject to an investment leverage constraint. To make it clear, the contributions of the paper are summarized as follows.

- A general problem formulation for optimal MRP design is proposed that aims at finding a desirable trade-off between the MR and the variance of the portfolio, while subject to a practical leverage constraint instead of a budget constraint. Different MR criteria and variance criteria are considered in the formulation. The portfolio leverage constraint takes two cases into consideration, namely, the case of cointegration space and the case of naturally stationary assets.
- Besides the MR and variance criteria, the asset selection criterion is further considered in the optimal MRP design problem. Finally, the formulation becomes a constrained nonconvex problem.
- A unified algorithm framework based on the successive convex approximation (SCA) method named SCA-MRP is proposed to solve the MRP design problem, which tackles the original highly nonconvex problem by solving a sequence of easy convex subproblems.
- In order to efficiently solve the convex inner subproblems in SCA-MRP and to address different design cases in practice, several methods are proposed. The Armijo-like backtracking line search method is proposed to accelerate the SCA-MRP algorithm.
- The algorithm complexity and convergence to a stationary point are analyzed for the SCA-MRP algorithm.
- Numerical simulations on both synthetic and real market data are carried out to address the efficacy of the proposed MRP design problem formulation and the algorithms.

The remaining sections of this paper are organized as follows. In Section II, the optimal MRP design problem is briefly introduced. A general problem formulation for the optimal MRP design is given in Section III. Section IV generally introduces the SCA method. The SCA-based algorithm called SCA-MRP is elaborated in Section V and three efficient algorithms to solve the convex subproblems are given in Section VI. The algorithm complexity analysis and convergence analysis are given in Section VII. Numerical performance is evaluated in Section VIII and, finally, concluding remarks are drawn in Section IX.

Notations: Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and italics denote scalars. We denote by $\mathbf{1}$ the all-one vectors and by $\mathbf{I}$ the identity matrices, respectively. We denote by $\mathbb{R}$ ($\mathbb{R}_+$) the real (nonnegative real) numbers. The N-dimensional real vectors are denoted by $\mathbb{R}^N$. We use $\mathbb{N}$ to denote the natural field. The $K \times K$-dimensional symmetric matrices are denoted by $\mathbb{S}^K$. Superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote the matrix transpose operation and the matrix inverse operation, respectively. For a vector $\mathbf{x}$, x_i denotes the ith element. For a matrix $\mathbf{X}$, $x_{i,j}$ denotes the (ith, jth) element. For symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succeq \mathbf{B}$ ($\mathbf{A} \succ \mathbf{B}$) means $\mathbf{A} - \mathbf{B}$ is a positive semidefinite (positive definite) matrix. Other notations will be introduced along this paper when required.

II. Optimal Mean-Reverting Portfolio Design

A. *Mean-Reverting Portfolio (MRP)*

For a financial asset, e.g., a common stock, a future contract, or a portfolio of them, its price at time t is denoted by $p_t \in \mathbb{R}_+$, and then its corresponding logarithmic price or log-price $y_t \in \mathbb{R}$ is given by $y_t \triangleq \log(p_t)$[1]. Let $\mathbf{y}_t \triangleq [y_{1,t}, \ldots, y_{M,t}]^T$ denote the log-prices of M assets. The (log-price) spread s_t is given by $s_t \triangleq \boldsymbol{\beta}^T \mathbf{y}_t$, where $\boldsymbol{\beta} \triangleq [\beta_1, \ldots, \beta_M]^T$ denotes the weights or hedge ratios. Suppose there exists a subspace, termed cointegration

[1]The $\log(\cdot)$ is the natural logarithm function.

space, with N (usually $N \leq M$) cointegration relations defined by $\mathbf{B} \triangleq [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N]$. Then these N spreads are obtained as

$$\mathbf{s}_t \triangleq \mathbf{B}^T \mathbf{y}_t, \tag{1}$$

where every element of $\mathbf{s}_t$ is a spread. Specifically, if the asset log-prices are stationary in nature, every element of $\mathbf{y}_t$ can be defined as a spread, i.e., $\mathbf{s}_t = \mathbf{y}_t$ with $\mathbf{B} = \mathbf{I}$ $(N = M)$.

Different spreads may possess different mean reversion and variance properties in nature. The objective of the MRP design problem is to construct a portfolio of the underlying spreads to attain desirable trading properties. For the N spreads in $\mathbf{s}_t$, the MRP is denoted by the portfolio weight $\mathbf{w} \triangleq [w_1, \ldots, w_N]^T$ with its resulting spread given by

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

Based on (1) and (2), we can further get the spread z_t defined on the underlying assets as follows:

$$z_t \triangleq \mathbf{w}_p^T \mathbf{y}_t, \tag{3}$$

where $\mathbf{w}_p \triangleq \mathbf{B}\mathbf{w} \in \mathbb{R}^M$ denotes the MRP weights on the underlying assets and represents the dollar value proportion invested on the underlying assets. For each asset $m = 1, \ldots, M$, the sign of $w_{p,m}$ indicates the type of positions, namely, $w_{p,m} > 0$ means a long position (i.e., it is bought), $w_{p,m} < 0$ means a short position (i.e., it is short-sold), and $w_{p,m} = 0$ means no position on the asset.

In the following, we continue to introduce some criteria for MR, variance, and asset selection.

B. Mean Reversion (MR) Criteria

Several MR criteria were used in [13], [34] and will be briefly introduced here. We start by defining the ith order (lag-i) autocovariance matrix for the spreads $\mathbf{s}_t$ as $\mathbf{M}_i \triangleq \mathsf{Cov}(\mathbf{s}_t, \mathbf{s}_{t+i}) = \mathsf{E}\left[(\mathbf{s}_t - \mathsf{E}[\mathbf{s}_t])(\mathbf{s}_{t+i} - \mathsf{E}[\mathbf{s}_{t+i}])^T\right]$ with $i \in \mathbb{N}$. Specifically, when $i = 0$, $\mathbf{M}_0$ stands for the (positive definite) covariance matrix of $\mathbf{s}_t$. Since we can always compute the centered form for $\mathbf{s}_t$, as $\tilde{\mathbf{s}}_t = \mathbf{s}_t - \mathsf{E}[\mathbf{s}_t]$, without loss of generality, $\mathbf{s}_t$ will be used to denote $\tilde{\mathbf{s}}_t$ in the following.

1) Predictability Statistics $\mathrm{pre}(\mathbf{w})$*:* 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 at time $t-1$, and ϵ_t denotes a white noise. The predictability statistics [36] is proposed to measure how close a random process is to a white noise and defined by $\mathrm{pre} \triangleq \sigma_{\hat{z}}^2 / \sigma_z^2$, where $\sigma_z^2 \triangleq \mathsf{E}\left[z_t^2\right]$ and $\sigma_{\hat{z}}^2 \triangleq \mathsf{E}\left[\hat{z}_{t-1}^2\right]$. Given the spread $z_t = \mathbf{w}^T \mathbf{s}_t$, the predictability statistics for spread $z_t = \mathbf{w}^T \mathbf{s}_t$ is computed as

$$\mathrm{pre}(\mathbf{w}) \triangleq \frac{\mathbf{w}^T \mathbf{T} \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}, \tag{4}$$

where $\mathbf{T} \triangleq \mathbf{M}_1^T \mathbf{M}_0^{-1} \mathbf{M}_1$. To design a spread z_t as close as possible to white noise, we need to minimize $\mathrm{pre}(\mathbf{w})$.

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 $\mathrm{por}(p) \triangleq \sum_{i=1}^{p} \rho_i^2$, where ρ_i is the ith order autocorrelation of z_t defined as $\rho_i \triangleq \mathsf{E}[z_t z_{t+i}] / \mathsf{E}\left[z_t^2\right]$. The measure $\mathrm{por}(p)$ is used to test whether a random process is close to a white noise. To design a spread z_t close to white noise, we need to minimize $\mathrm{por}_z(p)$ with a prespecified order p. Given $z_t = \mathbf{w}^T \mathbf{s}_t$, we can get $\mathrm{por}(p, \mathbf{w})$ as follows:

$$\mathrm{por}(p, \mathbf{w}) \triangleq \sum_{i=1}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2. \tag{5}$$

3) Crossing Statistics $\mathrm{cro}(\mathbf{w})$ *and Penalized Crossing Statistics* $\mathrm{pcro}(p, \mathbf{w})$*:* The zero-crossing rate for a centered stationary process z_t is defined as $\mathrm{zcr} \triangleq (T-1)^{-1} \sum_{t=2}^{T} \mathbf{1}\{z_t z_{t-1} \leq 0\}$, where

$$\mathbf{1}\{z_t z_{t-1} \leq 0\} \triangleq \begin{cases} 1, & z_t z_{t-1} \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

is the indicator function defined on z_t. It tests the probability that a process crosses its mean per unit of time. According to [38], for a centered stationary Gaussian process, zero-crossing rate is defined as $\mathrm{zcr} = \pi^{-1} \arccos(\rho_1)$. To design a spread exhibiting sufficient zero-crossings, we should minimize ρ_1. So given $z_t = \mathbf{w}^T \mathbf{s}_t$, we define the crossing statistics as

$$\mathrm{cro}(\mathbf{w}) \triangleq \frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}. \tag{6}$$

Besides minimizing the criterion $\mathrm{cro}(\mathbf{w})$, it is also proposed to minimize the absolute high order autocorrelations $|\rho_i|$'s $(i = 2, \ldots, p)$ [13]. Based on $\mathrm{cro}(\mathbf{w})$, the penalized crossing statistics of order p is defined as follows:

$$\mathrm{pcro}(p, \mathbf{w}) \triangleq \frac{\mathbf{w}^T \mathbf{M}_1 \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} + \eta \sum_{i=2}^{p} \left(\frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}} \right)^2, \tag{7}$$

where η is a positive prespecified factor.

C. Variance Criteria

Given a spread $z_t = \mathbf{w}^T \mathbf{s}_t$, its variance is naturally given by $\mathsf{Var}[z_t] = \mathsf{E}\left[z_t^2\right] = \mathbf{w}^T \mathbf{M}_0 \mathbf{w}$. Another criterion we will consider is the standard deviation of z_t which is given by $\mathsf{Std}[z_t] = \sqrt{\mathbf{w}^T \mathbf{M}_0 \mathbf{w}}$.

D. Asset Selection Criterion

In portfolio design problems, allocating capital to all the assets can increase significantly the transaction costs, which motivates to select a subset of assets [39]. To realize this asset selection goal in MRP design, it is desirable to pursue sparsity in the cointegration space $\mathbf{B}$. Based on the ℓ_0-"norm"[2] [40], the asset selection criterion is accordingly given by $\|\mathbf{B}\mathbf{w}\|_0$.

E. Portfolio Leverage Constraint

Constraints on portfolio weights in portfolio design problems represent the investment policy and allocation [41]. As we mentioned in the introduction, the returns from statistical arbitrage

[2] Strictly speaking, it is not a norm. For $\mathbf{x} \in \mathbb{R}^N$, the ℓ_0-"norm" $\|\mathbf{x}\|_0 \triangleq \sum_{i=1}^{N} \mathrm{sgn}(|x_i|)$, where $\mathrm{sgn}(\cdot)$ is the sign function.

are usually small. Hence, investors in practice may want to use leverage to multiply the returns.

In [32]–[34], the ℓ_2-norm $\|\mathbf{w}\|_2$ was considered as a portfolio constraint. The ℓ_2-norm is commonly used as a power constraint in electrical engineering like wireless communications and radar, but its practical significance in financial applications is unclear since imposing the ℓ_2-norm on portfolio weights does not carry any physical meaning in a financial context. To address this issue, in our previous paper [13], the budget constraint $\mathbf{1}^T\mathbf{w} = \mathbf{1}$ was proposed, but still fails to take the "portfolio leverage" into account which is the key practical consideration in portfolio design. In this paper, we will use a general investment leverage constraint given as follows:

$$\mathcal{W} \triangleq \{\mathbf{w} \mid \|\mathbf{B}\mathbf{w}\|_1 \leq L\},$$

where $\mathbf{B}$ is the cointegration space and L means the total investment leverage considering both long and short positions deployed on the underlying financial assets.

III. The Optimal MRP Design Problem

A. Problem Formulation

Considering the three design criteria previously described, i.e., MR criterion, variance criterion, and asset selection criterion, a general problem formulation for the optimal MRP design problem is given as follows:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & F(\mathbf{w}) \triangleq U(\mathbf{w}) + \mu V(\mathbf{w}) + \gamma S(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathcal{W}, \end{array} \tag{8}$$

which is a nonconvex constrained problem. The constant $\mu > 0$ defines the trade-off between the portfolio MR measure and variance preference. The regularizing parameter $\gamma \geq 0$ controls the sparsity level. The three terms $U(\mathbf{w})$, $V(\mathbf{w})$, and $S(\mathbf{w})$ composing the objective are described below in detail.

T1) The mean reversion term $U(\mathbf{w})$ considering different MR criteria can be jointly represented as

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

which contains as specific cases the predictability statistics $\text{pre}(\mathbf{w})$ ($\xi = 1$, $\mathbf{H} = \mathbf{T}$, and $\zeta = \eta = 0$), the portmanteau statistics $\text{por}(p, \mathbf{w})$ ($\xi = 0$ and $\zeta = \eta = 1$), the crossing statistics $\text{cro}(\mathbf{w})$ ($\xi = 1$, $\mathbf{H} = \mathbf{M}_1$, and $\zeta = \eta = 0$), and the penalized crossing statistics $\text{pcro}(p, \mathbf{w})$ ($\xi = 1, \mathbf{H} = \mathbf{M}_1, \zeta = 0$, and $\eta > 0$).

T2) The variance term $V(\mathbf{w})$ can be represented in the following four different forms:

$$V(\mathbf{w}) \triangleq \begin{cases} 1/\left(\mathbf{w}^T\mathbf{M}_0\mathbf{w}\right) & (\text{VarInv}(\mathbf{w})) \\ 1/\sqrt{\mathbf{w}^T\mathbf{M}_0\mathbf{w}} & (\text{StdInv}(\mathbf{w})) \\ -\,\mathbf{w}^T\mathbf{M}_0\mathbf{w} & (\text{VarNeg}(\mathbf{w})) \\ -\,\sqrt{\mathbf{w}^T\mathbf{M}_0\mathbf{w}} & (\text{StdNeg}(\mathbf{w})). \end{cases}$$

T3) The asset selection term $S(\mathbf{w})$ is given by

$$S(\mathbf{w}) \triangleq \|\mathbf{B}\mathbf{w}\|_0 = \textstyle\sum_{m=1}^{M} \text{sgn}\left(|[\mathbf{B}\mathbf{w}]_m|\right),$$

where $\text{sgn}(\cdot)$ is the sign function and $[\mathbf{a}]_m$ denotes the mth element in $\mathbf{a}$.

B. Observations and Insight

In this section, we will focus on the analysis of the optimal MRP design problem formulation in (8). Some observations and insight are given in the following.

Lemma 1: Given any two colinear MRPs: $\mathbf{w}_1$ and $\mathbf{w}_2 \triangleq \alpha\mathbf{w}_1$ ($\mathbf{w}_1 \neq \mathbf{0}$ and $\alpha \neq 0$), we have: i) $U(\mathbf{w}_1) = U(\mathbf{w}_2)$; ii) $V(\mathbf{w}_1) > (<, =)\, V(\mathbf{w}_2)$, when $|\alpha| < (>, =)\, 1$; iii) $S(\mathbf{w}_1) = S(\mathbf{w}_2)$; iv) $F(\mathbf{w}_1) = F(\mathbf{w}_2)$, when $|\alpha| = 1$; and v) $|\alpha|\|\mathbf{B}\mathbf{w}_1\|_1 = \|\mathbf{B}\mathbf{w}_2\|_1$.

Proof: The proof is trivial and hence omitted. ∎

In Lemma 1, the points i)-iii) reveal that increasing the leverage level L on an MRP can only increase its variance, but not change its MR and asset selection properties. The point iv) reveals that two MRPs with the opposite sign of weights $\mathbf{w}$ are essentially the same; or in other words, two trading spreads defined by $\mathbf{w}^T\mathbf{s}_t$ and $-\mathbf{w}^T\mathbf{s}_t$ are the same. This is really to the nature of MRP design, 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.

Based on Lemma 1, we further have the following result.

Lemma 2: Denote the set of optimal solutions of Problem (8) as $\mathcal{W}^\star \triangleq \{\mathbf{w}^\star | F(\mathbf{w}^\star) \leq F(\mathbf{w}), \forall \mathbf{w} \neq \mathbf{0}, \mathbf{w} \in \mathcal{W}\}$, then $\mathcal{W}^\star \subseteq \text{bd}(\mathcal{W})$ (the boundary of set $\mathcal{W}$).

Proof: The proof is trivial and hence omitted. ∎

Lemma 2 essentially reveals the inequality leverage constraint is always active, i.e., the designed MRPs always attain the total leverage L.

As mentioned before, the cointegration matrix $\mathbf{B}$, in practice, is commonly estimated based on time series modeling. However, the matrix $\mathbf{B}$ is not unique [23]. (Assuming the singular value decomposition $\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, the cointegration space (i.e., the column space of $\mathbf{B}$) is given by $\mathcal{R}(\mathbf{U})$). Then based on Lemma 2, another intriguing observation for the MRP design problem (8) is given in the following.

Proposition 3: Suppose there exist $\mathbf{B}, \mathbf{B}' \in \mathcal{R}(\mathbf{U})$ with the corresponding designed optimal MRPs from Problem (8) denoted as $\mathcal{W}_p^\star \triangleq \left\{\mathbf{w}_p^\star | \mathbf{w}_p^\star = \mathbf{B}\mathbf{w}^\star, \forall \mathbf{w}^\star \in \mathcal{W}^\star\right\}$ and $\mathcal{W}_p'^\star \triangleq \left\{\mathbf{w}_p'^\star | \mathbf{w}_p'^\star = \mathbf{B}'\mathbf{w}'^\star, \forall \mathbf{w}'^\star \in \mathcal{W}'^\star\right\}$ respectively, we have $\mathcal{W}_p^\star = \mathcal{W}_p'^\star$.

Proof: See Appendix A. ∎

This result reveals that the optimal MRP $\mathbf{w}_p^\star$ designed from Problem (8) does not depend on the explicit form of $\mathbf{B}$, but instead only on the subspace $\mathcal{R}(\mathbf{U})$.

C. Mild Problem Modifications

The objective function $F(\mathbf{w})$ in (8) is not well-defined at $\mathbf{0}$ making it discontinuous over $\mathcal{W}$. Some mild modifications to $F(\mathbf{w})$ will be introduced in this section. Firstly, since $U(\mathbf{w})$ and $V(\mathbf{w})$ (refer to $\text{VarInv}(\mathbf{w})$ and $\text{StdInv}(\mathbf{w})$) are singular at $\mathbf{0}$, we propose to reduce this "singularity" by defining two modified criteria $U^{\epsilon}(\mathbf{w})$ and $V^{\epsilon}(\mathbf{w})$ as follows:

$$U^{\epsilon}(\mathbf{w}) \triangleq \xi \frac{\mathbf{w}^T\mathbf{H}\mathbf{w}}{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon} + \zeta\left(\frac{\mathbf{w}^T\mathbf{M}_1\mathbf{w}}{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon}\right)^2 + \eta\sum_{i=2}^{p}\left(\frac{\mathbf{w}^T\mathbf{M}_i\mathbf{w}}{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon}\right)^2, \tag{9}$$

and

$$V^{\epsilon}(\mathbf{w}) \triangleq \begin{cases} 1/\left(\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon\right) & (\text{VarInv}(\mathbf{w}))\\ 1/\sqrt{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon} & (\text{StdInv}(\mathbf{w}))\\ -\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon & (\text{VarNeg}(\mathbf{w}))\\ -\sqrt{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon} & (\text{StdNeg}(\mathbf{w})), \end{cases} \tag{10}$$

where $\epsilon > 0$ is a small constant. Secondly, since the sparsity criterion $S(\mathbf{w})$ is nonconvex and discontinuous, the following smooth nonconvex sparsity function will be considered

$$S^{\epsilon}(\mathbf{w}) \triangleq \textstyle\sum_{m=1}^{M}\left[1-\exp\left(-\epsilon^{-1}\left|[\mathbf{Bw}]_m\right|^2\right)\right], \tag{11}$$

where compared to $S(\mathbf{w})$ the function $1-\exp(-\epsilon^{-1}(\cdot)^2)$ is used to approximate $\text{sgn}(\cdot)$ with $\epsilon > 0$ controlling the approximation tightness [42]. Finally, the the modified objective is given as $F^{\epsilon}(\mathbf{w}) \triangleq U^{\epsilon}(\mathbf{w}) + \mu V^{\epsilon}(\mathbf{w}) + \gamma S^{\epsilon}(\mathbf{w})$ which is almost "equivalent" to $F(\mathbf{w})$.

Now, we are ready to discuss the solving procedure for the optimal MRP design problem in (8). We will firstly introduce a general algorithmic framework based on the idea of successively approximating the original nonconvex problem with a series of convex subproblems, and the derived algorithms are expected to be simple and efficient with provable convergence to a stationary point.

IV. The Successive Convex Approximation Method

The successive convex approximation (SCA) method [43] is a general optimization framework especially for solving nonconvex optimization problems. In this paper, we will use the SCA method proposed in [44] which is based on solving a sequence of simpler strongly convex problems, preserves feasibility of the iterates for the original nonconvex problem, and also has guaranteed convergence.

Specifically, an optimization problem is given as follows:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X}, \end{array} \tag{12}$$

where $\mathcal{X} \subseteq \mathbb{R}^N$ is convex and $f(\mathbf{x})$ is nonconvex and (possibly) nonsmooth. In order to solve Problem (12) which could be difficult to tackle directly, starting from an initial feasible point $\mathbf{x}^{(0)}$, the SCA method solves a series of subproblems with surrogate functions $\tilde{f}\left(\mathbf{x};\mathbf{x}^{(k)}\right)$ (or simply denoted as $\tilde{f}^{(k)}(\mathbf{x})$) approximating the original objective $f(\mathbf{x})$ over the set $\mathcal{X}$. A sequence $\left\{\mathbf{x}^{(k)}\right\}$ is generated by the following rules:

$$\begin{cases} \hat{\mathbf{x}}^{(k+1)} = \arg\min_{\mathbf{x}\in\mathcal{X}} \tilde{f}(\mathbf{x};\mathbf{x}^{(k)}) & (a)\\ \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \gamma^{(k)}(\hat{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}). & (b) \end{cases} \tag{13}$$

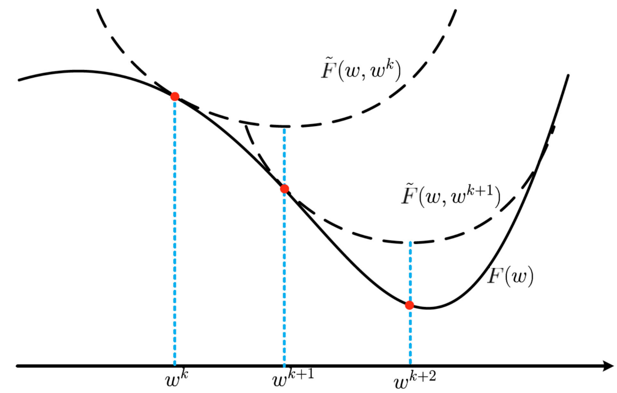


Fig. 1. Solving Problem (8) with objective $F(\mathbf{w})$ by solving a sequence of strongly convex subproblems with quadratic objective functions $\tilde{F}\left(\mathbf{w};\mathbf{w}^{(k)}\right)$ (or $\tilde{F}^{(k)}(\mathbf{w})$) in (14). (Illustration is shown in one dimension.)

The first step is to generate the descent direction (i.e., $\hat{\mathbf{x}}^{(k+1)} - \mathbf{x}^{(k)}$) by solving a best-response problem (a), and the second step is the variable update rule with $\gamma^{(k)}$ to be the step-size.

Convergence to a stationary solution of the original nonconvex optimization problem in (12) can be established under the following mild assumptions on the problem:

A1) $\mathcal{X}$ is nonempty, closed, and convex;
A2) $\nabla_{\mathbf{x}} f(\mathbf{x})$ is $L_{\nabla f}$-Lipschitz continuous on $\mathcal{X}$;
A3) $f(\mathbf{x})$ is coercive on $\mathcal{X}$.

And as to $\tilde{f}\left(\mathbf{x};\mathbf{x}^{(k)}\right)$, the following conditions are needed:

B1) given $\mathbf{x}^{(k)}$, $\tilde{f}\left(\mathbf{x};\mathbf{x}^{(k)}\right)$ is c-strongly convex on $\mathcal{X}$ for some $c>0$, i.e., $\nabla^2_{\mathbf{x}}\tilde{f}\left(\mathbf{x};\mathbf{x}^{(k)}\right) \succeq c\mathbf{I}$;
B2) $\nabla_{\mathbf{x}}\tilde{f}\left(\mathbf{x}^{(k)};\mathbf{x}^{(k)}\right) = \nabla_{\mathbf{x}} f\left(\mathbf{x}^{(k)}\right)$ for all $\mathbf{x}^{(k)} \in \mathcal{X}$;
B3) $\nabla_{\mathbf{x}}\tilde{f}(\mathbf{x};\mathbf{x})$ is continuous for all $\mathbf{x} \in \mathcal{X}$.

It is easy to see that the key point to use SCA is to find a good approximation function $\tilde{f}\left(\mathbf{x};\mathbf{x}^{(k)}\right)$, which could make the best response problem easy to solve and result in a fast convergence. In the following, we are going to apply the SCA method for the optimal MRP design problem in (8).

V. Problem Solving Based on The SCA Method

A. Using SCA For MRP Design

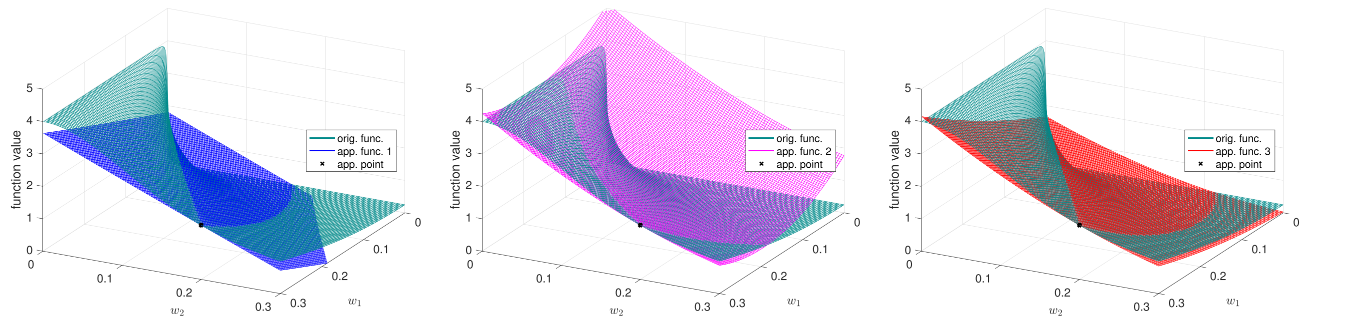
Applying the SCA method to Problem (8), we have the convex approximation function $\tilde{F}^{(k)}(\mathbf{w})$ at the $(k+1)$th iteration for the objective $F^{\epsilon}(\mathbf{w})$ given as follows:

$$\tilde{F}^{(k)}(\mathbf{w}) \triangleq \tilde{U}^{(k)}(\mathbf{w}) + \mu\tilde{V}^{(k)}(\mathbf{w}) + \gamma\tilde{S}^{(k)}(\mathbf{w}) + \tau\|\mathbf{w}-\mathbf{w}^{(k)}\|_2^2, \tag{14}$$

with $\tau \geq 0$ denoting a parameter on the proximal term added for convergence [44]. An illustrative figure for the relation between $F^{\epsilon}(\mathbf{w})$ and $\tilde{F}^{(k)}(\mathbf{w})$ is given in Figure 1.

1) On The Approximation Term $\tilde{U}^{(k)}(\mathbf{w})$: The term $\tilde{U}^{(k)}(\mathbf{w})$ is a convex approximation for the MR term $U^{\epsilon}(\mathbf{w})$. Based on the general idea of SCA, there could be many choices on deriving such an approximation. When $U^{\epsilon}(\mathbf{w})$ is chosen as $\text{pre}(\mathbf{w})$ or $\text{cro}(\mathbf{w})$ (i.e., ratio of quadratic functions), the convex approximation is chosen to be a linearization of the criterion, which is simply given as

$$\tilde{U}^{(k)}(\mathbf{w}) \triangleq \mathbf{b}_U^{(k)T}\mathbf{w}, \tag{15}$$



(a) Approximation function $\tilde{u}_1\left(\mathbf{w};\mathbf{w}^{(k)}\right)$. (b) Approximation function $\tilde{u}_2\left(\mathbf{w};\mathbf{w}^{(k)}\right)$ (c) Approximation function $\tilde{u}_3\left(\mathbf{w};\mathbf{w}^{(k)}\right)$

Fig. 2. Three different convex approximation functions for $u\left(\mathbf{w}\right)$ in Equation (18) at approximation point $\mathbf{w}^{(k)}=(0.3,0.2)^T$.

where $\mathbf{b}_U^{(k)}$ is given in Equation (16). When $U^{\epsilon}\left(\mathbf{w}\right)$ is chosen as $\mathrm{por}\left(p,\mathbf{w}\right)$ or $\mathrm{pcro}\left(p,\mathbf{w}\right)$, in which case a "square of ratio of quadratic functions" term is involved, a nice approximation technique by exploring the convex curvature of $U^{\epsilon}\left(\mathbf{w}\right)$ given in Example 4 will be employed.

Example 4: We first define a function $u\left(\mathbf{w}\right)$ as

$$u\left(\mathbf{w}\right)\triangleq\left(\frac{\mathbf{w}^T\mathbf{M}_i\mathbf{w}}{\mathbf{w}^T\mathbf{M}_0\mathbf{w}+\epsilon}\right)^2. \tag{18}$$

At a given point $\mathbf{w}^{(k)}$, we describe three possible choices of the convex approximation function $\tilde{u}^{(k)}\left(\mathbf{w}\right)$ in Equation (19) together with their visualizations in Figure 2. The function $\tilde{u}_1^{(k)}\left(\mathbf{w}\right)$ is based on the direct linearization of $u\left(\mathbf{w}\right)$ and $\tilde{u}_2^{(k)}\left(\mathbf{w}\right)$ is designed by linearization w.r.t. partial variables.[3] To obtain the third approximation function $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$, $u\left(\mathbf{w}\right)$ is convexified by linearizing the fractional term inside the square operation $\left(\cdot\right)^2$.

By comparing the subfigures (a), (b), and (c) in Figure 2, we can find that $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$ gives a much tighter approximation than the other two.

[3]The details are given in Appendix B.

For $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$, it is easy to verify that the approximation technique ensures it has the same gradient as $u\left(\mathbf{w}\right)$ at $\mathbf{w}^{(k)}$. We can also observe that the matrix $\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)^T$ in $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$ is in fact an approximation to the true Hessian matrix, which is known as an outer product approximation or Levenberg-Marquardt approximation [45]. Thus, it is reasonable to assume that based on $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$ the overall resulting algorithm is able to largely maintain the information on the curvature of the cost function, even if higher order derivatives with respect to the gradient are never explicitly computed. In fact, this is an interesting line of reasoning, which could eventually lead to improved approximations for the cost functions.

Finally, for $\mathrm{por}\left(p,\mathbf{w}\right)$ or $\mathrm{pcro}\left(p,\mathbf{w}\right)$, using the approximation technique for $\tilde{u}_3^{(k)}\left(\mathbf{w}\right)$ in Example 4, we have $\tilde{U}^{(k)}\left(\mathbf{w}\right)$ as follows:

$$\tilde{U}^{(k)}\left(\mathbf{w}\right)\triangleq\mathbf{w}^T\mathbf{A}_U^{(k)}\mathbf{w}+\mathbf{b}_U^{(k)T}\mathbf{w}, \tag{20}$$

where $\mathbf{A}_U^{(k)}$ and $\mathbf{b}_U^{(k)}$ are given in (21) (shown at the bottom of the next page) and (16), respectively.

2) On The Approximation Term $\tilde{V}^{(k)}\left(\mathbf{w}\right)$: The $\tilde{V}^{(k)}\left(\mathbf{w}\right)$ denotes the convex (i.e., linearization) approximation for the

$$\mathbf{b}_U^{(k)}\triangleq 2\xi\left(\mathbf{d}_{0,h}^{(k)}-\mathbf{d}_{h,0}^{(k)}\right)+2\zeta r_1^{(k)}\left(\mathbf{d}_{0,1}^{(k)}-\mathbf{d}_{1,0}^{(k)}\right)+2\eta\sum_{i=2}^{p}r_i^{(k)}\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right) \tag{16}$$

with

$$\begin{aligned}
&r_i^{(k)}\triangleq\left(\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),\ i=1,\ldots,p, && r_h^{(k)}\triangleq\left(\mathbf{w}^{(k)T}\mathbf{H}\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),\\
&\mathbf{d}_{0,i}^{(k)}\triangleq\left(\mathbf{M}_i\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),\ i=1,\ldots,p, && \mathbf{d}_{0,h}^{(k)}\triangleq\left(\mathbf{H}\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),\\
&\mathbf{d}_{i,0}^{(k)}\triangleq\left(r_i^{(k)}\mathbf{M}_0\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),\ i=1,\ldots,p, && \mathbf{d}_{h,0}^{(k)}\triangleq\left(r_h^{(k)}\mathbf{M}_0\mathbf{w}^{(k)}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right),
\end{aligned} \tag{17}$$

$$\begin{aligned}
\tilde{u}_1^{(k)}\left(\mathbf{w}\right)&\triangleq\left(r_i^{(k)}\right)^2+4r_i^{(k)}\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)^T\left(\mathbf{w}-\mathbf{w}^{(k)}\right)\\
\tilde{u}_2^{(k)}\left(\mathbf{w}\right)&\triangleq 4\left(r_i^{(k)}\right)^3-\left(r_i^{(k)}\right)^2-4r_i^{(k)}\left(\mathbf{d}_{i,0}^{(k)}\right)^T\mathbf{w}+2r_i^{(k)}\left(\mathbf{w}^T\mathbf{M}_i\mathbf{w}\right)/\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon\right)\\
\tilde{u}_3^{(k)}\left(\mathbf{w}\right)&\triangleq\left[r_i^{(k)}+2\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)^T\left(\mathbf{w}-\mathbf{w}^{(k)}\right)\right]^2\\
&=\left(r_i^{(k)}\right)^2+4r_i^{(k)}\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)^T\left(\mathbf{w}-\mathbf{w}^{(k)}\right)+4\left(\mathbf{w}-\mathbf{w}^{(k)}\right)^T\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)\left(\mathbf{d}_{0,i}^{(k)}-\mathbf{d}_{i,0}^{(k)}\right)^T\left(\mathbf{w}-\mathbf{w}^{(k)}\right)
\end{aligned} \tag{19}$$

variance term $V^{\epsilon}(\mathbf{w})$ given by

$$\tilde{V}^{(k)}(\mathbf{w}) \triangleq \mathbf{b}_V^{(k)T}\mathbf{w}, \tag{22}$$

where

$$\mathbf{b}_V^{(k)} \triangleq \begin{cases} -2(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon)^{-2}\mathbf{M}_0\mathbf{w}^{(k)} & (\mathrm{VarInv}(\mathbf{w}))\\ -(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon)^{-\frac{3}{2}}\mathbf{M}_0\mathbf{w}^{(k)} & (\mathrm{StdInv}(\mathbf{w}))\\ -2\mathbf{M}_0\mathbf{w}^{(k)} & (\mathrm{VarNeg}(\mathbf{w}))\\ -(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon)^{-\frac{1}{2}}\mathbf{M}_0\mathbf{w}^{(k)} & (\mathrm{StdNeg}(\mathbf{w})).\end{cases}$$

3) On The Approximation Term $\tilde{S}^{(k)}(\mathbf{w})$: The $\tilde{S}^{(k)}(\mathbf{w})$ is the convex approximation for the sparsity term $S^{\epsilon}(\mathbf{w})$. To derive it, we need the following lemma.

Lemma 5: At any point $x^{(k)} \in \mathbb{R}$, a tight upperbound function for $s(x) \triangleq 1-\exp(-\epsilon^{-1}x^2)$ is obtained as follows:

$$s(x) \le \tilde{s}^{(k)}(x) \triangleq \epsilon^{-1}\exp\big(-\epsilon^{-1}(x^{(k)})^2\big)x^2 + 1 - \exp\big(-\epsilon^{-1}(x^{(k)})^2\big)\big(1+\epsilon^{-1}(x^{(k)})^2\big),$$

for $\forall x \in \mathbb{R}$.

Proof: The proof is trivial and hence omitted. ■

From Lemma 5, we have $\nabla_x \tilde{s}^{(k)}(x^{(k)}) = \nabla_x s(x^{(k)})$. Then, based on the function $\tilde{s}^{(k)}(x)$, the approximation for $S(\mathbf{w})$ is accordingly given as follows:

$$\tilde{S}^{(k)}(\mathbf{w}) \triangleq \mathbf{w}^T\mathbf{A}_S^{(k)}\mathbf{w}, \tag{23}$$

with $\mathbf{A}_S^{(k)} \triangleq \epsilon^{-1}\mathbf{B}^T\mathrm{Diag}\left[\exp\left(-\epsilon^{-1}\left(\mathbf{B}\mathbf{w}^{(k)}\odot\mathbf{B}\mathbf{w}^{(k)}\right)\right)\right]\mathbf{B}$, where $\exp(\cdot)$ is taken elementwise and $\mathrm{Diag}[\mathbf{d}]$ is a matrix with diagonal elements formed by $\mathbf{d}$.

Finally, by combining $\tilde{U}^{(k)}(\mathbf{w})$, $\tilde{V}^{(k)}(\mathbf{w})$, and $\tilde{S}^{(k)}(\mathbf{w})$, $\tilde{F}^{(k)}(\mathbf{w})$ is accordingly written as

$$\tilde{F}^{(k)}(\mathbf{w}) \triangleq \mathbf{w}^T\mathbf{A}^{(k)}\mathbf{w} + \mathbf{b}^{(k)T}\mathbf{w}, \tag{24}$$

where $\mathbf{A}^{(k)} \triangleq \mathbf{A}_U^{(k)} + \gamma\mathbf{A}_S^{(k)} + \tau\mathbf{I}$, and $\mathbf{b}^{(k)} \triangleq \mathbf{b}_U^{(k)} + \mu\mathbf{b}_V^{(k)} - 2\tau\mathbf{w}^{(k)}$. Based on the approximation $\tilde{F}^{(k)}(\mathbf{w})$ for $F(\mathbf{w})$, the subproblem to solve in the $(k+1)$th iteration is

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T\mathbf{A}^{(k)}\mathbf{w} + \mathbf{b}^{(k)T}\mathbf{w} \\ \text{subject to} \quad & \|\mathbf{B}\mathbf{w}\|_1 \le L, \end{aligned} \tag{25}$$

which is a convex problem. We can observe that the objective function in Problem (25) is quadratic in variable $\mathbf{w}$ instead of nonconvex in $\mathbf{w}$ as in Problem (8). Since it is convex, this problem can be efficiently solved, and some efficient methods for different cases will be discussed in detail in Section VI.

B. SCA-MRP: The Overall Algorithm

Based on SCA, in order to solve the original nonconvex problem in (8), we just need to iteratively solve a convex subproblem in (25). We name this SCA-based algorithm SCA-MRP and summarize it in Algorithm 1. The algorithm can be guaranteed to converge globally when the step-size $\gamma^{(k)}$ is chosen properly. A practical approach to choosing $\gamma^{(k)}$ is the Armijo-like backtracking line search rule [46], which is given as follows:

$$\begin{aligned} &\text{Given} && \alpha,\beta\in(0,1),\ l=0\\ &\text{While} && \Delta F^{\epsilon}(\mathbf{w}^{(k)}) > -\alpha\beta^l\|\Delta\mathbf{w}^{(k)}\|_2^2\\ & && l=l+1\\ &\text{Let} && \gamma^{(k)}=\beta^l \text{ for } k=0,1,2,\ldots, \end{aligned}$$

where $\Delta F^{\epsilon}\left(\mathbf{w}^{(k)}\right) \triangleq F^{\epsilon}\left(\mathbf{w}^{(k)}+\beta^l\Delta\mathbf{w}^{(k)}\right) - F^{\epsilon}\left(\mathbf{w}^{(k)}\right)$ with $\Delta\mathbf{w}^{(k)} \triangleq \hat{\mathbf{w}}^{(k+1)} - \mathbf{w}^{(k)}$.

Algorithm 1: SCA-MRP: An SCA-Based Algorithm for The Optimal MRP Design Problem (8).

Require: $\mathbf{H}$, $\mathbf{M}_i$ $(i=0,\ldots,p)$, μ, γ, $\mathbf{B}$, L and τ
1: Set $k=0$, $\gamma^{(0)}$ and $\mathbf{w}^{(0)}$.
2: **repeat**
3: Compute $\mathbf{A}^{(k)}$ and $\mathbf{b}^{(k)}$ in (24)
4: $\hat{\mathbf{w}}^{(k+1)} = \arg\min_{\mathbf{w}\in\mathcal{W}} \mathbf{w}^T\mathbf{A}^{(k)}\mathbf{w} + \mathbf{b}^{(k)T}\mathbf{w}$
5: $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \gamma^{(k)}(\hat{\mathbf{w}}^{(k+1)} - \mathbf{w}^{(k)})$
6: $k \leftarrow k+1$
7: **until** some convergence criterion is met

VI. Solving Methods for The Inner Subproblem

In SCA-MRP, we need to solve a sequence of convex subproblems in each iteration (see `Step 4` in Algorithm 1). This inner subproblem has no closed-form solution, but we can resort to the off-the-shelf public or commercial solvers like `SeDuMi` [47], `SDPT3` [48], and `MOSEK` [49] or some popular convex optimization toolboxes (scripting languages) like `YALMIP` [50] and `CVX` [51]. However, as an alternative to the general-purpose solvers and toolboxes, we can also develop problem-specific algorithms to solve this problem more efficiently.

A. Algorithm Based on the ADMM Method

For the sake of notational simplicity, we omit the superscript (k) in the SCA subproblem (25) and recast it as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} \\ \text{subject to} \quad & \|\mathbf{B}\mathbf{w}\|_1 \le L, \end{aligned} \tag{26}$$

where $\mathbf{A} \succ \mathbf{0}$.

The alternating direction method of multipliers (ADMM) is a method that can solve a convex optimization problem by breaking it into smaller parts, each of which are then easier to handle

$$\mathbf{A}_U^{(k)} \triangleq 4\zeta\big(\mathbf{d}_{0,1}^{(k)}\mathbf{d}_{0,1}^{(k)T} + \mathbf{d}_{1,0}^{(k)}\mathbf{d}_{1,0}^{(k)T} - \mathbf{d}_{0,1}^{(k)}\mathbf{d}_{1,0}^{(k)T} - \mathbf{d}_{1,0}^{(k)}\mathbf{d}_{0,1}^{(k)T}\big) + 4\eta\sum_{i=2}^{p}\big(\mathbf{d}_{0,i}^{(k)}\mathbf{d}_{0,i}^{(k)T} + \mathbf{d}_{i,0}^{(k)}\mathbf{d}_{i,0}^{(k)T} - \mathbf{d}_{0,i}^{(k)}\mathbf{d}_{i,0}^{(k)T} - \mathbf{d}_{i,0}^{(k)}\mathbf{d}_{0,i}^{(k)T}\big) \quad \mathbf{d}_{0,i}^{(k)}\text{'s and } \mathbf{d}_{i,0}^{(k)}\text{'s are defined in Eq. (17)} \tag{21}$$

[52]. It has recently been applied on applications in a number of areas. To solve Problem (26) based on ADMM, we first rewrite it by introducing an auxiliary variable $\mathbf{z} = \mathbf{B}\mathbf{w}$, then the problem becomes

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{z}}{\text{minimize}} & \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} \\ \text{subject to} & \|\mathbf{z}\|_1 \leq L,\ \mathbf{B}\mathbf{w} - \mathbf{z} = \mathbf{0}, \end{array} \tag{27}$$

We further define an indicator function for the ℓ_1-norm ball set as $I_{\mathcal{C}}(\mathbf{z}) \triangleq \begin{cases} 0, & \mathbf{z} \in \mathcal{C} \triangleq \left\{\mathbf{z} \middle| \|\mathbf{z}\|_1 \leq L\right\} \\ +\infty, & \text{otherwise}, \end{cases}$. Problem (27) can be written in the following standard ADMM form:

$$\begin{array}{ll} \underset{\mathbf{w},\mathbf{z}}{\text{minimize}} & \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} + I_{\mathcal{C}}(\mathbf{z}) \\ \text{subject to} & \mathbf{B}\mathbf{w} - \mathbf{z} = \mathbf{0}. \end{array} \tag{28}$$

And the augmented Lagrangian is given as follows:

$$\begin{aligned} &\mathcal{L}_\rho(\mathbf{w}, \mathbf{z}, \mathbf{u}(\mathbf{y})) \\ &= \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} + I_{\mathcal{C}}(\mathbf{z}) + \mathbf{y}^T(\mathbf{B}\mathbf{w} - \mathbf{z}) + \frac{\rho}{2}\|\mathbf{B}\mathbf{w} - \mathbf{z}\|_2^2 \\ &= \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} + I_{\mathcal{C}}(\mathbf{z}) + \frac{\rho}{2}\|\mathbf{B}\mathbf{w} - \mathbf{z} + \mathbf{u}\|_2^2 + const., \end{aligned}$$

where $\rho > 0$ is the penalty parameter which serves as the dual update step-size and the scaled dual variable $\mathbf{u} \triangleq \frac{1}{\rho}\mathbf{y}$. Then, the ADMM updates are given in three variable blocks $(\mathbf{w}, \mathbf{z}, \mathbf{u})$ by

$$\begin{cases} \mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w}} \Big\{ \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} \\ \qquad\qquad + \frac{\rho}{2}\left\|\mathbf{B}\mathbf{w} - \mathbf{z}^{(k)} + \mathbf{u}^{(k)}\right\|_2^2 \Big\} \\ \mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}} \left\{ I_{\mathcal{C}}(\mathbf{z}) + \frac{\rho}{2}\left\|\mathbf{z} - \mathbf{B}\mathbf{w}^{(k+1)} - \mathbf{u}^{(k)}\right\|_2^2 \right\} \\ \qquad\quad = \Pi_{\mathcal{C}}(\mathbf{B}\mathbf{w}^{(k+1)} + \mathbf{u}^{(k)}) \\ \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \mathbf{B}\mathbf{w}^{(k+1)} - \mathbf{z}^{(k+1)}, \end{cases}$$

where the $\mathbf{z}$ update step is essentially a projection onto set $\mathcal{C}$ with $\Pi_{\mathcal{C}}(\cdot)$ denoting the projection operator.

Specifically, the update of variable $\mathbf{w}$ amounts to solving a convex quadratic programming (QP) which has the closed-form solution:

$$\mathbf{w}^{(k+1)} = -(2\mathbf{A} + \rho\mathbf{B}^T\mathbf{B})^{-1}(\mathbf{b} + \rho\mathbf{B}^T(\mathbf{u}^{(k)} - \mathbf{z}^{(k)})).$$

By defining $\mathbf{h}^{(k)} \triangleq \mathbf{B}\mathbf{w}^{(k+1)} + \mathbf{u}^{(k)}$, the variable $\mathbf{z}$ update is equivalent to solving

$$\begin{array}{ll} \underset{\mathbf{z}}{\text{minimize}} & \left\|\mathbf{z} - \mathbf{h}^{(k)}\right\|_2^2 \\ \text{subject to} & \|\mathbf{z}\|_1 \leq L, \end{array} \tag{29}$$

which is the classical projection onto the ℓ_1-norm ball problem with efficient algorithms for problem solving [53, Lemma 1][54]. An efficient water-filling-like algorithm based on sorting is given in Algorithm 2.

In Algorithm 2, $\mathrm{sgn}(\cdot)$ is the sign function which extracts the sign of a real number; $\mathrm{abs}(\cdot)$ is the absolute value function; and $b_{(j)}$ $(1 \leq j \leq N)$ denotes the j-th largest element in $\mathbf{b}$. This algorithm gives a water-filling-like closed-form solution to Problem (29). In this ADMM-based algorithm, we have three blocks of variables to minimize, which could possibly be slow for convergence. For the primal variable $\mathbf{w}$ update, we also need to solve a convex QP involving the matrix inversion. In the following, we will develop an alternative algorithm.

Algorithm 2: Euclidean Projection Onto An ℓ_1-Norm Ball (29).

Require: $\mathbf{h}$ and L
if $||\mathbf{h}||_1 \leq L$ **then**
  $\mathbf{z} = \mathbf{h}$
  return $\mathbf{z}$
else
  $\mathbf{a} = \mathrm{sign}(\mathbf{h})$ and $\mathbf{b} = \mathrm{abs}(\mathbf{h})$
  Sort the elements in $\mathbf{b}$ as $b_{(1)} \geq b_{(2)} \geq \cdots \geq b_{(N)}$
  $\rho = \arg\max_{1\leq j\leq N} \left\{ j \,|\, b_{(j)} > \frac{1}{j}\left(\sum_{i=1}^{j} b_{(i)} - L\right) \right\}$
  $\theta = \frac{1}{\rho}\left(\sum_{i=1}^{\rho} b_{(i)} - L\right)$
  $z_j = a_j \max\{b_j - \theta, 0\}, 1 \leq j \leq N$
  return $\mathbf{z}$
end if

B. Algorithm Based on the M-ADMM Method

The following method to solve the convex inner problem in (26) is based on majorized ADMM (M-ADMM) [55]. Compared to the vanilla ADMM, M-ADMM introduces the majorization-minimization (MM) [56] idea to find an upper-bound function for the variable update. By minimizing instead an upperbound function, a cheap closed-form variable update can be achieved in many cases.

To use the M-ADMM method, based on Problem (26), we first define a new variable $\tilde{\mathbf{w}} \triangleq \mathbf{B}\mathbf{w}$. Then, we can equivalently have $\mathbf{w} = \mathbf{B}^\dagger\tilde{\mathbf{w}}$ and $\left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} = \mathbf{0}$,[4] where $\mathbf{B}^\dagger$ is the Moore-Penrose pseudo-inverse of $\mathbf{B}$, and the columns of $\mathbf{B}^\perp$ span the orthogonal complementary subspace of $\mathbf{B}$, respectively. Then by defining $\tilde{\mathbf{A}} \triangleq \left(\mathbf{B}^\dagger\right)^T\mathbf{A}\mathbf{B}^\dagger$ and $\tilde{\mathbf{b}} \triangleq \left(\mathbf{B}^\dagger\right)^T\mathbf{b}$, Problem (26) can be equivalently rewritten in terms of $\tilde{\mathbf{w}}$ as

$$\begin{array}{ll} \underset{\tilde{\mathbf{w}}}{\text{minimize}} & \tilde{\mathbf{w}}^T\tilde{\mathbf{A}}\tilde{\mathbf{w}} + \tilde{\mathbf{b}}^T\tilde{\mathbf{w}} \\ \text{subject to} & \|\tilde{\mathbf{w}}\|_1 \leq L,\ \left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} = \mathbf{0}. \end{array} \tag{30}$$

Based on the indicator function $I_{\mathcal{C}}(\tilde{\mathbf{w}})$ defined before, the above problem can be rewritten in the following form:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \tilde{\mathbf{w}}^T\tilde{\mathbf{A}}\tilde{\mathbf{w}} + \tilde{\mathbf{b}}^T\tilde{\mathbf{w}} + I_{\mathcal{C}}(\tilde{\mathbf{w}}) \\ \text{subject to} & \left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} = \mathbf{0}. \end{array} \tag{31}$$

[4]A simple proof for this is given in Appendix C.

And the augmented Lagrangian for (31) can be written as

$$\mathcal{L}_\rho\left(\tilde{\mathbf{w}}, \mathbf{u}\left(\mathbf{y}\right)\right)$$
$$= \tilde{\mathbf{w}}^T \tilde{\mathbf{A}}\tilde{\mathbf{w}} + \tilde{\mathbf{b}}^T\tilde{\mathbf{w}} + I_{\mathcal{C}}(\tilde{\mathbf{w}}) + \mathbf{y}^T\left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} + \frac{\rho}{2}\left\|\left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}}\right\|_2^2$$
$$= \tilde{\mathbf{w}}^T \tilde{\mathbf{A}}\tilde{\mathbf{w}} + \tilde{\mathbf{b}}^T\tilde{\mathbf{w}} + I_{\mathcal{C}}(\tilde{\mathbf{w}}) + \frac{\rho}{2}\left\|\left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} + \mathbf{u}\right\|_2^2 + const.,$$

where $\rho > 0$ is the penalty parameter and the scaled dual variable $\mathbf{u} = \frac{1}{\rho}\mathbf{y}$. Based on the augmented Lagrangian, it is easy to see that we only have two variable blocks $(\tilde{\mathbf{w}}, \mathbf{u})$ for alternating minimization. Before we drive the variable update rule, we give the following useful result.

Lemma 6 ([56]): Let $\mathbf{A} \in \mathbb{S}^K$ and $\mathbf{B} \in \mathbb{S}^K$ such that $\mathbf{B} \succeq \mathbf{A}$. At any point $\mathbf{x}^{(k)} \in \mathbb{R}^K$, we have $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq \mathbf{x}^T\mathbf{B}\mathbf{x} + 2\mathbf{x}^{(k)T}\left(\mathbf{A} - \mathbf{B}\right)\mathbf{x} + \mathbf{x}^{(k)T}\left(\mathbf{B} - \mathbf{A}\right)\mathbf{x}^{(k)}$.

Then for the $\tilde{\mathbf{w}}$ update, at the $(k+1)$th iteration with iterates $\left(\tilde{\mathbf{w}}^{(k)}, \mathbf{u}^{(k)}(\mathbf{y}^{(k)})\right)$, by taking $\mathbf{M}^{\mathrm{M-ADMM}} \triangleq \tilde{\mathbf{A}} + \frac{\rho}{2}\mathbf{B}^\perp\left(\mathbf{B}^\perp\right)^T$ as $\mathbf{A}$ and choosing $\mathbf{B} = \lambda_{\max}^{\mathrm{M-ADMM}}\mathbf{I}$ where $\lambda_{\max}^{\mathrm{M-ADMM}} \triangleq \lambda_{\max}\left(\mathbf{M}^{\mathrm{M-ADMM}}\right)$ in Lemma 6, we get

$$\mathcal{L}_\rho\left(\tilde{\mathbf{w}}; \tilde{\mathbf{w}}^{(k)}, \mathbf{u}^{(k)}\left(\mathbf{y}^{(k)}\right)\right)$$
$$= \tilde{\mathbf{w}}^T\mathbf{M}^{\mathrm{M-ADMM}}\tilde{\mathbf{w}} + \left(\tilde{\mathbf{b}} + \mathbf{B}^\perp\mathbf{y}^{(k)}\right)^T\tilde{\mathbf{w}} + I_{\mathcal{C}}(\tilde{\mathbf{w}}) + const.$$
$$\leq \lambda_{\max}^{\mathrm{M-ADMM}}\tilde{\mathbf{w}}^T\tilde{\mathbf{w}} + 2\tilde{\mathbf{w}}^{(k)T}\left(\mathbf{M}^{\mathrm{M-ADMM}} - \lambda_{\max}^{\mathrm{M-ADMM}}\mathbf{I}\right)\tilde{\mathbf{w}}$$
$$+ \left(\tilde{\mathbf{b}} + \mathbf{B}^\perp\mathbf{y}^{(k)}\right)^T\tilde{\mathbf{w}} + I_{\mathcal{C}}(\tilde{\mathbf{w}}) + const.$$
$$= \lambda_{\max}^{\mathrm{M-ADMM}}\left\|\tilde{\mathbf{w}} - \mathbf{h}^{(k)}\right\|_2^2 + I_{\mathcal{C}}(\tilde{\mathbf{w}}) + const.,$$

where $\mathbf{h}^{(k)} \triangleq -\left(\left(\lambda_{\max}^{\mathrm{M-ADMM}}\right)^{-1}\mathbf{M}^{\mathrm{M-ADMM}} - \mathbf{I}\right)\tilde{\mathbf{w}}^{(k)} - \frac{1}{2}\left(\lambda_{\max}^{\mathrm{M-ADMM}}\right)^{-1}\left(\tilde{\mathbf{b}} + \mathbf{B}^\perp\mathbf{y}^{(k)}\right)$. Then, the variable updates in M-ADMM are given as follows:

$$\begin{cases} \tilde{\mathbf{w}}^{(k+1)} = \arg\min_{\tilde{\mathbf{w}}}\left\{\left\|\tilde{\mathbf{w}} - \mathbf{h}^{(k)}\right\|_2^2 + I_{\mathcal{C}}\left(\tilde{\mathbf{w}}\right)\right\} = \Pi_{\mathcal{C}}(\mathbf{h}^{(k)}) \\ \mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}}^{(k+1)}. \end{cases}$$

Specifically, for the variable $\tilde{\mathbf{w}}$ update, it is the projection onto the ℓ_1-norm ball problem as in (29). In the M-ADMM algorithm, the number of variable blocks is reduced to 2 compared to the 3 variable blocks in the ADMM algorithm. In fact, when $\mathbf{B} = \mathbf{I}$, by leveraging on this specific structure, more efficient algorithm can be derived.

C. Specialized Algorithm Based on the MM Method

When $\mathbf{B} = \mathbf{I}$, the convex subproblem in (26) is written as

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w} \\ \text{subject to} & \|\mathbf{w}\|_1 \leq L. \end{array} \tag{32}$$

Besides using ADMM and M-ADMM, this problem can be more efficiently solved by the majorization-minimization (MM) method [56]. Using this primal-only method, we can get rid of the dual variable update in ADMM and M-ADMM.

From (32), based on Lemma 6, at the $(k+1)$th iteration with iterate $\mathbf{w}^{(k)}$, the objective function in (32) can be majorized as follows:

$$\mathbf{w}^T\mathbf{A}\mathbf{w} + \mathbf{b}^T\mathbf{w}$$
$$\leq \lambda_{\max}\left(\mathbf{A}\right)\mathbf{w}^T\mathbf{w} + 2\mathbf{w}^{(k)T}\left(\mathbf{A} - \lambda_{\max}\left(\mathbf{A}\right)\mathbf{I}\right)\mathbf{w}$$
$$+ \mathbf{b}^T\mathbf{w} + \mathbf{w}^{(k)T}\left(\lambda_{\max}\left(\mathbf{A}\right)\mathbf{I} - \mathbf{A}\right)\mathbf{w}^{(k)}$$
$$= \lambda_{\max}\left(\mathbf{A}\right)\left\|\mathbf{w} - \mathbf{h}^{(k)}\right\|_2^2 + const.,$$

where $\mathbf{h}^{(k)} \triangleq -\left(\lambda_{\max}^{-1}\left(\mathbf{A}\right)\mathbf{A} - \mathbf{I}\right)\mathbf{w}^{(k)} - \frac{1}{2}\lambda_{\max}^{-1}\left(\mathbf{A}\right)\mathbf{b}$. Then, the subproblem to solve in MM is given by

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \left\|\mathbf{w} - \mathbf{h}^{(k)}\right\|_2^2 \\ \text{subject to} & \|\mathbf{w}\|_1 \leq L, \end{array}$$

which is still a projection onto the ℓ_1-norm ball problem and can be solvedbased on Algorithm 2.

VII. Complexity and Convergence Analysis

A. Complexity Analysis

In this section, we give a detailed discussion on the computational complexity of our proposed algorithms in Section VI. We analyze the per-iteration computational cost of the algorithms proposed to solve the inner convex subproblems, i.e., the ADMM-based algorithm, the M-ADMM-based algorithm, and the MM-based algorithm.

For the ADMM-based algorithm, the computational cost for updating three variable blocks $\mathbf{w}$, $\mathbf{z}$, and $\mathbf{u}$ are analyzed separately. The computational cost for updating $\mathbf{w}$ is $\mathcal{O}(N^3 + MN + 3N^2 + M + 2N)$. For updating $\mathbf{z}$, the cost is $\mathcal{O}(MN + M)$ (to calculate $\mathbf{h}^{(k)}$) plus $\mathcal{O}(M)$ (to do projection). The cost for updating $\mathbf{u}$ is $\mathcal{O}(MN + 2M)$. So, the total cost per iteration $(M \geq N)$ is $\mathcal{O}(N^3 + MN + 3N^2 + M + 2N) + \mathcal{O}(MN + M) + \mathcal{O}(M) + \mathcal{O}(MN + 2M) \approx \mathcal{O}(N^3)$.

In the M-ADMM-based algorithm, for pre-processing, computing the Moore-Penrose pseudo-inverse the $\mathbf{B}^\dagger$ requires complexity of $\mathcal{O}(MN^2)$ and computing the orthogonal compliment $\mathbf{B}^\perp$ needs complexity of $\mathcal{O}(MN^2)$. So computing $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{b}}$ requires complexity of $\mathcal{O}(NM^2 + MN^2)$ and $\mathcal{O}(MN)$, separately. To recover the variable $\mathbf{w}$, i.e., post-processing, still need $\mathcal{O}(MN)$ time. Therefore the cost for computation outside the iterations is $\mathcal{O}(NM^2 + MN^2) + \mathcal{O}(MN) + \mathcal{O}(MN) \approx \mathcal{O}(NM^2)$. To compute $\mathbf{h}^{(k)}$ in each iteration, it contains costs of $\mathcal{O}(2M^2)$ and $\mathcal{O}(M^3)$ to calculate a $M \times M$ matrix (i.e., $\tilde{\mathbf{A}} + \frac{\rho}{2}\left(\mathbf{B}^\perp\right)^T\mathbf{B}^\perp$) and its largest eigenvalue. However, the $\mathcal{O}(M^3)$ complexity can be reduced by simply replacing the largest eigenvalue with some easily computed quantity (like the Frobenius norm since $\|\mathbf{A}\|_F \geq \lambda\left(\mathbf{A}\right)_{\max}$) of that matrix, which only requires cost $\mathcal{O}(2M^2)$. Therefore, the overall cost for calculating $\mathbf{h}^{(k)}$ is $\mathcal{O}(6M^2 + M^2 + 4M - MN)$ and the cost for updating $\tilde{\mathbf{w}}$ is $\mathcal{O}(M)$. Besides, it requires $\mathcal{O}(M^2 + M - MN - N)$ to update $\mathbf{u}$. Then the overall cost for each iteration is $\mathcal{O}(6M^2 + M^2 + 4M - MN) + \mathcal{O}(M) + \mathcal{O}(M^2 + M - MN - N) \approx \mathcal{O}(M^2)$.

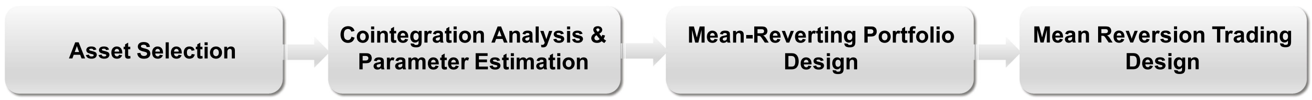


Fig. 3. A system view of the statistical arbitrage trading strategy in finance.

The MM-based algorithm is proposed for the $\mathbf{B} = \mathbf{I}$ case. It also needs pretreatment, which costs $\mathcal{O}(N^3)$ to calculate the maximum eigenvalue of a $N \times N$ matrix and $\mathcal{O}(N^2 + N)$ to calculate the constant part of $\mathbf{h}^{(k)}$. The overall computation cost is of order $\mathcal{O}(N^3)$. In each iteration, to update $\mathbf{w}$, $\mathcal{O}(N^2 + N)$ is needed for constructing the majorization function and $\mathcal{O}(N)$ for projection onto l_1-norm ball. The total cost per iteration is of order $\mathcal{O}(N^2)$.

The three algorithms for solving the subproblem should be properly chosen in order to achieve a better computational performance. The per-iteration computational cost for the ADMM-based and M-ADMM-based algorithms are $\mathcal{O}(N^3)$ and $\mathcal{O}(M^2)$, respectively. So, under the condition $M \geq N^{1.5}$, ADMM is recommended; otherwise, M-ADMM should be more appropriate. Compared to the $\mathcal{O}(N^3)$ complexity in ADMM and the $\mathcal{O}(M^2)$ complexity in M-ADMM for each iteration, MM-based algorithm just need $\mathcal{O}(N^2)$ computation per iteration. The time complexity of the MM-based algorithm is also lower in the preprocessing stage compared with M-ADMM. So the MM-based algorithm is highly recommended when $\mathbf{B} = \mathbf{I}$.

B. Convergence Analysis

The convergence property for the SCA-MRP algorithm is given in the following.

Proposition 7: Under assumptions A1)-A3) and B1)-B3), suppose $\tau \geq 0$, $\gamma^{(k)} \in (0, 1]$, $\gamma^{(k)} \to 0$ and $\sum_k \gamma^{(k)} = +\infty$, and let $\left\{\mathbf{w}^{(k)}\right\}$ be the sequence generated by SCA-MRP. Then either SCA-MRP converges in a finite number of iterations to a stationary solution of Problem (8) or every limit of sequence $\left\{\mathbf{w}^{(k)}\right\}$ (at least one such point exists) is a stationary solution of Problem (8).

Proof: We can first check that the proposed problem satisfies Assumptions A1)-A3) in Section IV. Given $\tau \geq 0$ and $\gamma^{(k)}$ as above, it is easy to check that the approximation function (24) is a strongly convex quadratic function and satisfies Assumptions B1)-B3) in Section IV. Then this result directly follows from the proof in [46, Theorem 2]. ∎

VIII. Numerical Simulations

In this section, we first give a system view of the statistical arbitrage strategy. Then several performance evaluation metrics on portfolio investment will be introduced. The performance of our proposed MRP design problem and the algorithms will finally be given based on both synthetic data and real market data.

A. A Flow Diagram of The Statistical Arbitrage Strategy

We summarize the whole statistical arbitrage strategy as shown in Figure 3.

1) Asset Selection: In this stage, a collection of (possibly cointegrated) asset candidates are selected to construct an asset pool. Conducting this process may require prior knowledge on the underlying financial assets.

2) Parameter Estimation and Cointegration Analysis: The cointegration analysis (say, Engle-Granger two-step test, Johansen test, Phillips-Ouliaris test, etc.) will be conducted to test the hypothesis that there is a statistically significant stationarity connection within the underlying asset prices. Accordingly, a cointegration space will be identified in this stage.

3) Mean-Reverting Portfolio Design: This stage is the focus of this paper. An optimal MRP is designed considering different criteria based on the assets within the identified cointegration space. Unit root test may be applied to test the stationarity of the finally designed spread.

4) Mean Reversion Trading Design: The designed spread will be firstly traded for an in-sample testing period for parameter estimation and trading actions optimization, such as the mean reversion equilibrium, trading threshold, timing of entering a position, lightening up a position, adding to a position, or exiting a position, and so on. After these trading parameters are obtained, the designed MRP can finally be invested for the out-of-sample trading.

B. Performance Evaluation Metrics

Some performance metrics for mean reversion trading used in [13] are briefly introduced in the following.

1) Profit and Loss: We define the profit and loss (P&L) for the MRP at time t as $\text{P\&L}_t \triangleq \mathbf{w}_p^T \mathbf{r}_t$ where the asset returns $\mathbf{r}_t \triangleq \mathbf{y}_t - \mathbf{y}_{t-1}$ (please refer to [13] for further details). P&L measures the amount of profits or losses (in units of dollars) of an investment on the portfolio for one holding period. In order to measure the cumulative return performance, we define the cumulative P&L (not compounding) in one trading from time t_1 to t_2 as Cum. $\text{P\&L}\,(t_1, t_2) \triangleq \sum_{t=t_1}^{t_2} \text{P\&L}_t$.

2) Return On Investment: Different MRPs may have different leverage properties, the return on investment (ROI) is introduced as another measure as the rate of return. Within one trading period, the ROI at time t is defined as $\text{ROI}_t \triangleq \frac{\text{P\&L}_t}{\|\mathbf{w}_p\|_1}$.

3) Sharpe Ratio: The Sharpe ratio (SR) describes how much excess return one can receive for the extra volatility (square root of variance). The annualized SR for a trading stage from time t_1 to t_2 is defined as $\text{SR}_{\text{ROI}}(t_1, t_2) \triangleq \sqrt{252}\frac{\mu_{\text{ROI}}}{\sigma_{\text{ROI}}}$, where $\mu_{\text{ROI}} \triangleq \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \text{ROI}_t$ is the sample return and $\sigma_{\text{ROI}} \triangleq \left[\frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2} \left(\text{ROI}_t - \mu_{\text{ROI}}\right)^2\right]^{\frac{1}{2}}$ is the sample standard deviation, and the factor $\sqrt{252}$ relates the daily SR to the annualized SR (assuming 252 trading days per year).

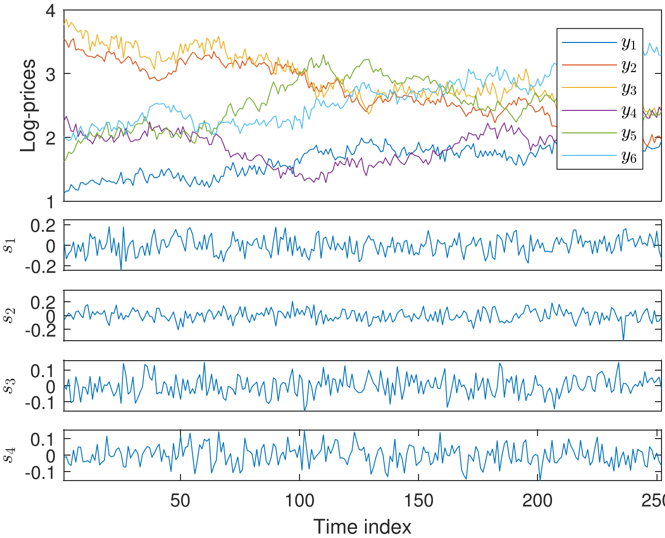


Fig. 4. Synthetic log-prices ($M = 6$) and the spreads ($N = 4$) generated from a VECM model of order 1.

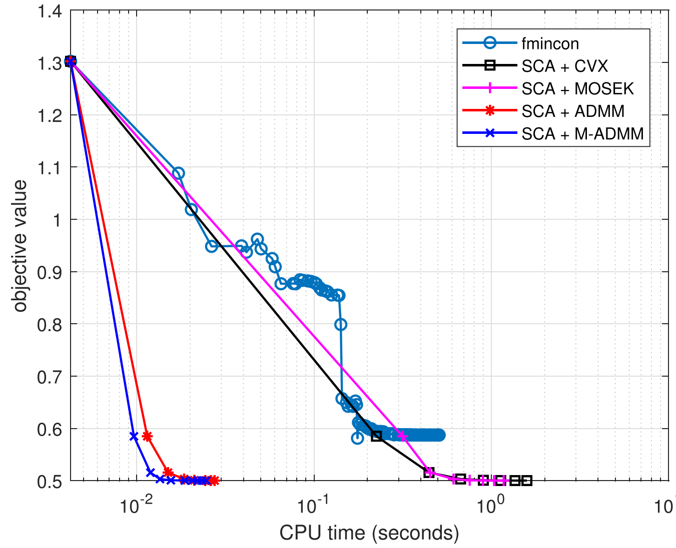
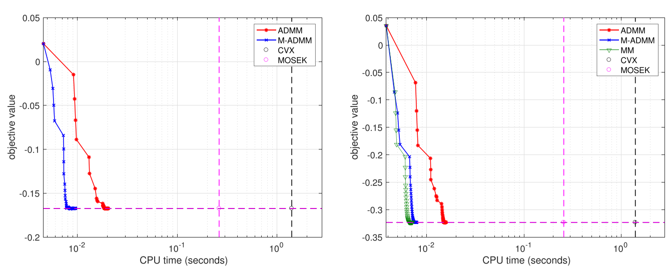


Fig. 6. Convergence of the objective function value for different solving methods for $\mathrm{pro}\,(\mathbf{w}, 3)$ and $L = 1.3$.



(a) Case 1: $M = 6$ and $N = 4$. (b) Case 2: $M = 10$ and $N = 10$.

Fig. 5. Convergence of the objective function value of different solving methods for the inner convex problem.

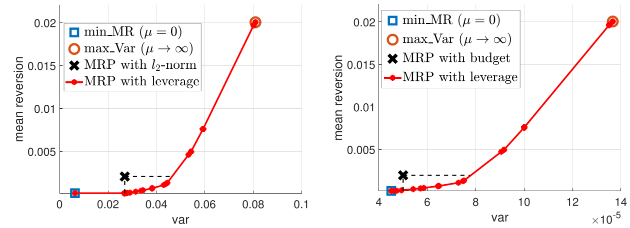
C. Synthetic Data Simulations

In this section, we will first show the superiority of the proposed algorithm SCA-MRP over some off-the-shelf solvers based on synthetic data. Following that, we will show the MRP design problem proposed in this paper is able to design a portfolio attaining a trade-off between MR and variance, which is a practical and desirable property for MRP design, but has never been considered in the literature. The synthetic data is generated using a vector error correction model (VECM) [23], which models the stock log-prices with underlying cointegration relations as shown in Figure 4.

1) Algorithm Performance: We first compare our proposed algorithms for the inner convex problems in SCA-MRP, i.e., the ADMM method, the M-ADMM method, and the MM method. The proposed methods are first compared with the standard off-the-shelf packages CVX and MOSEK in Figure 5. Based on our simulations, the M-ADMM and ADMM algorithms can converge to the optimal solution orders of magnitude faster compared to CVX and MOSEK. In Case 1 of Figure 5, M-ADMM method outperforms ADMM method. And in Case 2, where $\mathbf{B} = \mathbf{I}$, the MM method achieves the best convergence performance in terms of runtime in all the tested algorithms as expected. These convergence results match the complexity analyses given in Section VII.

We now compare the solution of the original problem based on SCA-MRP algorithm with the standard solver fmincon in MATLAB Optimization Toolbox for the MRP design problem where the MR criterion and the variance criterion are chosen as the portmanteau statistics of order 3, i.e., $\mathrm{pro}\,(\mathbf{w}, 3)$ and $\mathrm{VarInv}\,(\mathbf{w})$, respectively. For the SCA-MRP algorithm, the inner convex problem is solved by different proposed algorithms. In the simulations, we use $\alpha = 10^{-5}$ and $\beta = 0.8$ in choosing the stepsize. From Figure 6, it is easy to see that the SCA-MRP algorithms can converge to better local optimal solutions with faster convergence speed compared to fmincon which is a general-purpose solver. Within all the SCA-MRP algorithms, the algorithms with inner problem solved by M-ADMM and ADMM show better convergence performance over those using CVX and MOSEK.

2) Formulation Property: In this section, we will show that our proposed MRP design problem in (8) is more practical and flexible. We compare the design problem model in this paper with the existing problem formulations in [13], [33]–[35]. Given a fixed portfolio variance and a fixed ℓ_1-norm as in [33], [34] or a fixed portfolio budget B as in [13], [35], we compute the MRP $\mathbf{w}$ (denoted as "MRP with ℓ_2-norm" in Figure 7(a) and "MRP with budget" in Figure 7(b)). Since, in real markets, the investment is always guided by the leverage which essentially tells the total amount of money people can employ, we accordingly compute the investment leverage $L = \|\mathbf{Bw}\|_1$ in these cases. Based on this leverage L, we use the newly proposed MRP design problem in (8) to design a series of MRPs (denoted as "MRP with leverage" in Figures 7(a) and 7(b)) where the MR criterion and the variance criterion are chosen as $\mathrm{por}\,(\mathbf{w}, 3)$ and $\mathrm{VarInv}\,(\mathbf{w})$, respectively. We first design the portfolio with the minimal MR denoted as $\mathbf{w}^{\star}_{\mathrm{min\_MR}}$ (corresponding to the case when $\mu \to 0$ in Problem (8)) and the portfolio with the maximal variance denoted as $\mathbf{w}^{\star}_{\mathrm{max\_Var}}$ (corresponding to the



(a) Comparison between the proposed methods and the methods in [33], [34]. (b) Comparison between the proposed methods and the methods in [13], [35].

Fig. 7. The trade-off between mean reversion and variance. (Each point is averaged based on 100 Monte Carlo simulations with random initializations.)

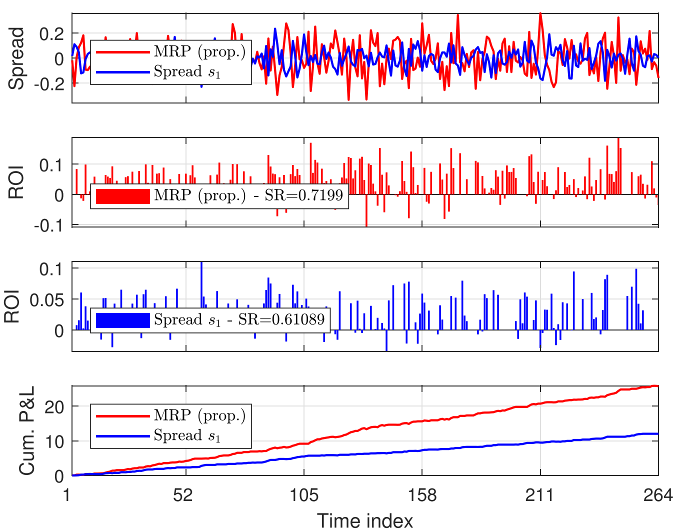


Fig. 8. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP (prop.) and one underlying spread denoted as Spread s_1.

case when $\mu \to \infty$ in Problem (8)). We also plot the path of the designed MRPs by tuning the parameter μ. In both Figure 7(a) and Figure 7(b), it can be found that by tuning parameter μ, for a fixed leverage the newly proposed design problem can easily get a trade-off between MR and variance of the MRP. However, although under the same investment leverage L, the MRP designed from [13], [33]–[35] is suboptimal no mater considering its MR property or variance property.

3) Trading Performance: We test the performance our designed MRP through mean reversion trading. The performance of the designed portfolio based on the proposed problem formulation is compared with spread s_1. Some performance metrics are reported in Figures 8. From the simulations, we can conclude that the MRPs designed based on the proposed formulation is able to generate consistent positive profits and can outperform the underlying spreads with higher Sharpe ratios of ROIs and higher cumulative P&Ls.

D. Real Data Simulations

In this section, we test the proposed problem formulation and algorithms based on real market data. We first select stocks from the Standard & Poor's 500 (S&P 500) Index to

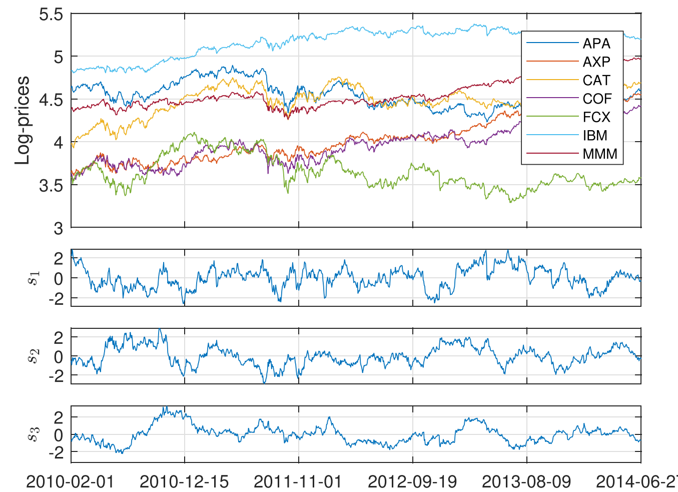


Fig. 9. Log-prices for $\{\mathsf{APA}, \mathsf{AXP}, \mathsf{CAT}, \mathsf{COF}, \mathsf{FCX}, \mathsf{IBM}, \mathsf{MMM}\}$ and three estimated spreads s_1, s_2, and s_3.

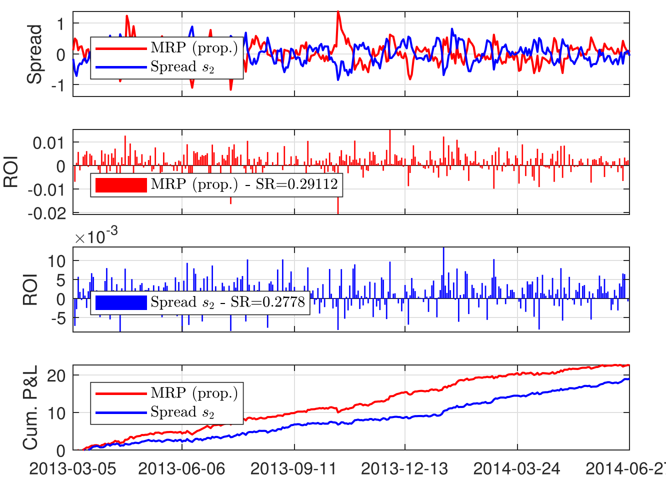


Fig. 10. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP (prop.) and one underlying spread denoted as Spread s_2.

construct an asset pool, which are denoted by their ticker labels as $\{\mathsf{APA}, \mathsf{AXP}, \mathsf{CAT}, \mathsf{COF}, \mathsf{FCX}, \mathsf{IBM}, \mathsf{MMM}\}$. The data is retrieved from Google Finance (https://www.google.com/finance). Then, a VECM model is fitted to identify the cointegration space $\mathcal{R}(\mathbf{B})$. After that, the MRP design problem proposed in this paper is used to design the optimal MRP where the MR is chosen as $\mathrm{pre}(\mathbf{w})$ and the variance criterion is chosen as $\mathrm{VarInv}(\mathbf{w})$. In Figure 9, we show the stock log-prices and spreads constructed from our asset pool. In Figures 10, 11, and 12, we show the performance comparisons between our designed MRP and one underlying spread s_2 and the MRPs from the literature [13], [33]–[35]. The log-prices for the designed spread, and the out-of-sample performance like ROI, Sharpe ratios of ROI, and cumulative P&Ls are reported. The in-sample training (learning) period is chosen from February 1st, 2010 to March 4th, 2013, and the out-of-sample trading (investing) period is from March 5th, 2013 to June 27th, 2014. It is easy

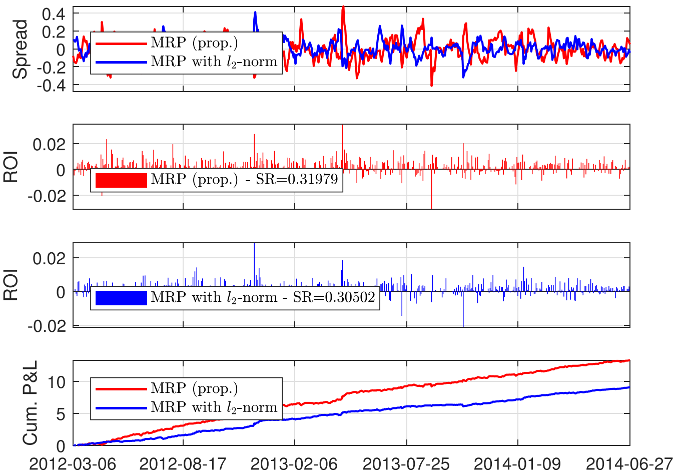


Fig. 11. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP (prop.) and the MRP design from [33], [34] denoted as MRP with ℓ_2-norm.

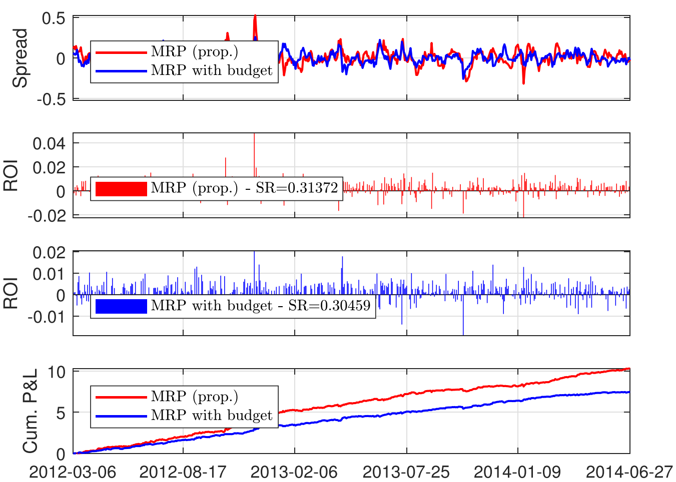


Fig. 12. Comparisons of ROIs, Sharpe ratios, and cumulative P&Ls between the MRP designed using our proposed method denoted as MRP (prop.) and the MRP design from [13], [35] denoted as MRP with budget.

to see the designed optimal MRP can achieve a higher Sharpe ratio and a better final cumulative return performance.

IX. Conclusions

The optimal mean-reverting portfolio design problem arising from statistical arbitrage has been considered in this paper. We have proposed a general problem formulation for MRP design where a trade-off can be attained between the mean reversion and variance of an MRP. Asset selection criterion has been further considered in the problem formulation. A practical investment leverage constraint has been imposed for MRP design. To solve the problem, a unified SCA-based algorithm has been proposed with the inner subproblems efficiently solved by different algorithms. Numerical results have shown that our proposed problem formulation can generate consistent profits and outperform the benchmark methods.

Appendix A
Proof for Proposition 3

Given $\mathbf{B} \in \mathcal{R}(\mathbf{U})$, we assume the optimal MRP from Problem (8) is given by $\mathbf{w}_p^\star = \mathbf{B}\mathbf{w}^\star \in \mathcal{W}_p^\star$ with $\mathbf{w}^\star \in \mathcal{W}^\star$. Accordingly, for another $\mathbf{B}' \in \mathcal{R}(\mathbf{U})$, we have $\mathbf{w}_p'^\star = \mathbf{B}'\mathbf{w}'^\star \in \mathcal{W}_p'^\star$ with $\mathbf{w}'^\star \in \mathcal{W}'^\star$.

Since $\mathbf{B}, \mathbf{B}' \in \mathcal{R}(\mathbf{U})$, there always exists $\mathbf{Q} \succ \mathbf{0}$ such that $\mathbf{B}' = \mathbf{B}\mathbf{Q}$. Also notice that the estimation of parameters in $U(\mathbf{w})$ and $V(\mathbf{w})$ depend on $\mathbf{B}$. Substitute $\mathbf{B}' = \mathbf{B}\mathbf{Q}$ into Problem (8) with variable $\mathbf{w}'$. Defining $\bar{\mathbf{w}} = \mathbf{Q}\mathbf{w}'$ with the optimal set $\bar{\mathcal{W}}^\star$, it is easy to see $\bar{\mathcal{W}}^\star = \mathcal{W}^\star$. Accordingly, we get $\forall \mathbf{w}'^\star \in \mathcal{W}'^\star$, $\mathbf{w}'^\star = \mathbf{Q}^{-1}\bar{\mathbf{w}}^\star = \mathbf{Q}^{-1}\mathbf{w}^\star$ with $\mathbf{w}^\star \in \mathcal{W}^\star$. Then we have $\forall \mathbf{w}_p'^\star \in \mathcal{W}_p'^\star$,

$$\mathbf{w}_p'^\star = \mathbf{B}'\mathbf{w}'^\star = (\mathbf{B}\mathbf{Q})\left(\mathbf{Q}^{-1}\mathbf{w}^\star\right) = \mathbf{B}\mathbf{w}^\star = \mathbf{w}_p^\star,$$

which implies $\mathcal{W}_p'^\star = \mathcal{W}_p^\star$.

Appendix B
On the Derivation of $\tilde{u}_2^{(k)}(\mathbf{w})$

Given $u(\mathbf{w})$ in (18), we define the numerator quadratic function in $(\cdot)^2$ as $t \triangleq \mathbf{w}^T\mathbf{M}_i\mathbf{w}$. Then, with a little abuse of notation, we have

$$u(t, \mathbf{w}) = \left(\frac{t}{\mathbf{w}^T\mathbf{M}_0\mathbf{w} + \epsilon}\right)^2.$$

A linear approximation function for $u(t, \mathbf{w})$ at $\left(t^{(k)}, \mathbf{w}^{(k)}\right)$ is given as follows:

$$\begin{aligned}\tilde{u}_2^{(k)}(t, \mathbf{w}) &= \left(\frac{t^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 \\ &+ 2\left(\frac{1}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 t^{(k)}(t - t^{(k)}) \\ &- 4(t^{(k)})^2\left(\frac{1}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^3 \mathbf{w}^{(k)T}\mathbf{M}_0(\mathbf{w} - \mathbf{w}^{(k)}).\end{aligned}$$

Changing the variables back to $\mathbf{w}$ (i.e., $t = \mathbf{w}^T\mathbf{M}_i\mathbf{w}$ and $t^{(k)} = \mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}$), we have

$$\begin{aligned}&\tilde{u}_2^{(k)}(\mathbf{w}) \\ &= \left(\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 + \frac{1}{\left(\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon\right)^2} \\ &\quad \times 2\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}(\mathbf{w}^T\mathbf{M}_i\mathbf{w} - \mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}) \\ &\quad - 4\left(\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 \frac{\mathbf{w}^{(k)T}\mathbf{M}_0(\mathbf{w} - \mathbf{w}^{(k)})}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon} \\ &= 4\left(\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^3 - \left(\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 \\ &\quad - 4\left(\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\right)^2 \frac{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon} \\ &\quad + 2\frac{\mathbf{w}^{(k)T}\mathbf{M}_i\mathbf{w}^{(k)}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}\frac{\mathbf{w}^T\mathbf{M}_i\mathbf{w}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)} + \epsilon}.\end{aligned}$$

Based on the definitions in (17), we further have

$$\widetilde{u}_2^{(k)}(\mathbf{w}) = 4\left(r_i^{(k)}\right)^3 - \left(r_i^{(k)}\right)^2 - 4r_i^{(k)}\left(\mathbf{d}_{i,0}^{(k)}\right)^T\mathbf{w} + 2r_i^{(k)}\frac{\mathbf{w}^T\mathbf{M}_i\mathbf{w}}{\mathbf{w}^{(k)T}\mathbf{M}_0\mathbf{w}^{(k)}+\epsilon}.$$

Appendix C Proof for The Variable Transformation

Since $\tilde{\mathbf{w}} = \mathbf{B}\mathbf{w}$ (where $\mathbf{B} \in \mathbb{R}^{M\times N}$ with $M \geq N$), we have

$$\tilde{\mathbf{w}} = \begin{bmatrix}\mathbf{B}\ \mathbf{B}^\perp\end{bmatrix}\begin{bmatrix}\mathbf{w}\\ \mathbf{0}\end{bmatrix},$$

where the columns of $\mathbf{B}^\perp$ span the orthogonal complementary of $\mathbf{B}$. Multiplying both sides of the above equation by $\begin{bmatrix}\mathbf{B}\ \mathbf{B}^\perp\end{bmatrix}^T$, we get

$$\begin{bmatrix}\mathbf{B}^T\\ \left(\mathbf{B}^\perp\right)^T\end{bmatrix}\tilde{\mathbf{w}} = \begin{bmatrix}\mathbf{B}^T\\ \left(\mathbf{B}^\perp\right)^T\end{bmatrix}\begin{bmatrix}\mathbf{B}\ \mathbf{B}^\perp\end{bmatrix}\begin{bmatrix}\mathbf{w}\\ \mathbf{0}\end{bmatrix} = \begin{bmatrix}\mathbf{B}^T\mathbf{B} & \mathbf{0}\\ \mathbf{0} & \left(\mathbf{B}^\perp\right)^T\mathbf{B}^\perp\end{bmatrix}\begin{bmatrix}\mathbf{w}\\ \mathbf{0}\end{bmatrix},$$

and then we have

$$\begin{bmatrix}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\\ \left(\left(\mathbf{B}^\perp\right)^T\mathbf{B}^\perp\right)^{-1}\left(\mathbf{B}^\perp\right)^T\end{bmatrix}\tilde{\mathbf{w}} = \begin{bmatrix}\mathbf{w}\\ \mathbf{0}\end{bmatrix}.$$

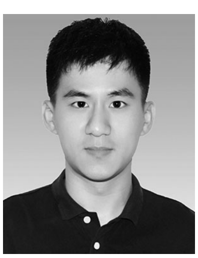
Notice that $\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T$ is the Moore-Penrose pseudo-inverse of $\mathbf{B}$ which can be written as $\mathbf{B}^\dagger$. We get the following equivalence relation

$$\tilde{\mathbf{w}} = \mathbf{B}\mathbf{w} \iff \begin{cases}\mathbf{B}^\dagger\tilde{\mathbf{w}} = \mathbf{w}\\ \left(\mathbf{B}^\perp\right)^T\tilde{\mathbf{w}} = \mathbf{0}\end{cases}.$$

References

[1] Z. Zhao, R. Zhou, Z. Wang, and D. P. Palomar, "Optimal portfolio design for statistical arbitrage in finance," in *Proc. 20th IEEE Statist. Signal Process. Workshop*, Freiburg, Germany, Jun. 2018, pp. 801–805.
[2] A. Pole, *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*, vol. 411. Hoboken, NJ, USA: Wiley, 2011.
[3] C. Krauss, "Statistical arbitrage pairs trading strategies: Review and outlook," *J. Econ. Surv.*, vol. 31, no. 2, pp. 513–545, 2017.
[4] C. Alexander and A. Dimitriu, "Indexing and statistical arbitrage," *J. Portfolio Manage.*, vol. 31, no. 2, pp. 50–63, 2005.
[5] S. F. LeRoy and J. Werner, *Principles of Financial Economics*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[6] Y. Feng and D. P. Palomar, "A signal processing perspective on financial engineering," *Found. Trends Signal Process.*, vol. 9, no. 1/2, pp. 1–231, 2016.
[7] G. Vidyamurthy, *Pairs Trading: Quantitative Methods and Analysis*, vol. 217. Hoboken, NJ, USA: Wiley, 2004.
[8] M. Whistler, *Trading Pairs: Capturing Profits and Hedging Risk With Statistical Arbitrage Strategies*, vol. 216, Hoboken, NJ, USA: Wiley, 2004.
[9] D. S. Ehrman, *The Handbook of Pairs Trading: Strategies Using Equities, Options, and Futures*. Hoboken, NJ, USA: Wiley, 2006.
[10] R. Bookstaber, *A Demon of Our Own Design: Markets, Hedge Funds, and the Perils of Financial Innovation*, vol. 240. Hoboken, NJ, USA: Wile, 2007.
[11] J. G. Nicholas, *Market Neutral Investing: Long/Short Hedge Fund Strategies*. New York, NY, USA: Bloomberg Press, 2000.
[12] B. I. Jacobs and K. N. Levy, *Market Neutral Strategies*, vol. 112. Hoboken, NJ, USA: Wiley, 2005.
[13] Z. Zhao and D. P. Palomar, "Mean-reverting portfolio with budget constraint," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2342–2357, May 2018.
[14] W. Xiong, "Convergence trading with wealth effects: An amplification mechanism in financial markets," *J. Financial Econ.*, vol. 62, no. 2, pp. 247–292, 2001.
[15] D. Butterworth and P. Holmes, "Inter-market spread trading: Evidence from UK index futures markets," *Appl. Financial Econ.*, vol. 12, no. 11, pp. 783–790, 2002.
[16] T. Kanamura, S. T. Rachev, and F. J. Fabozzi, "A profit model for spread trading with an application to energy futures," *J. Trading*, vol. 5, no. 1, pp. 48–62, 2010.
[17] M. Cummins and A. Bucca, "Quantitative spread trading on crude oil and refined products markets," *Quant. Finance*, vol. 12, no. 12, pp. 1857–1875, 2012.
[18] E. Gatev, W. N. Goetzmann, and K. G. Rouwenhorst, "Pairs trading: Performance of a relative-value arbitrage rule," *Rev. Financial Stud.*, vol. 19, no. 3, pp. 797–827, 2006.
[19] M. Avellaneda and J.-H. Lee, "Statistical arbitrage in the US equities market," *Quant. Finance*, vol. 10, no. 7, pp. 761–782, 2010.
[20] Y. Stander, D. Marais, and I. Botha, "Trading strategies with copulas," *J. Econ. Financial Sci.*, vol. 6, no. 1, pp. 83–107, 2013.
[21] R. J. Elliott, J. van der Hoek, and W. P. Malcolm, "Pairs trading," *Quant. Finance*, vol. 5, no. 3, pp. 271–276, 2005.
[22] T. Leung and X. Li, *Optimal Mean Reversion Trading: Mathematical Analysis and Practical Applications*. Singapore: World Scientific, 2016.
[23] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. New York, NY, USA: Springer, 2007.
[24] C. W. Granger, "Cointegrated variables and error correction models," USCD Discussion Paper, Tech. Rep. 83-13a, 1983.
[25] R. F. Engle and C. W. Granger, "Co-integration and error correction: Representation, estimation, and testing," *Econometrica: J. Econometric Soc.*, vol. 55, no. 2, pp. 251–276, 1987.
[26] S. Johansen, "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models," *Econometrica: J. Econometric Soc.*, vol. 59, no. 6, pp. 1551–1580, 1991.
[27] Z. Zhao and D. P. Palomar, "Robust maximum likelihood estimation of sparse vector error correction model," in *Proc. 5th IEEE Global Conf. Signal Inf. Process.*, Montreal, QC, Canada, Nov. 2017, pp. 913–917.
[28] M. S. Perlin, "Evaluation of pairs-trading strategy at the Brazilian financial market," *J. Derivatives Hedge Funds*, vol. 15, no. 2, pp. 122–136, 2009.
[29] S. Drakos, "Statistical arbitrage in S&P500," *J. Math. Finance*, vol. 6, no. 1, pp. 166–177, 2016.
[30] H. Rad, R. K. Y. Low, and R. Faff, "The profitability of pairs trading strategies: Distance, cointegration and copula methods," *Quant. Finance*, vol. 16, no. 10, pp. 1541–1558, 2016.
[31] H. Zhang and Q. Zhang, "Trading a mean-reverting asset: Buy low and sell high," *Automatica*, vol. 44, no. 6, pp. 1511–1518, 2008.
[32] A. d'Aspremont, "Identifying small mean-reverting portfolios," *Quant. Finance*, vol. 11, no. 3, pp. 351–364, 2011.
[33] M. Cuturi and A. d'Aspremont, "Mean reversion with a variance threshold," in *Proc. 30th Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 271–279.
[34] M. Cuturi and A. d'Aspremont, "Mean-reverting portfolios," in *Financial Signal Processing and Machine Learning*, A. N. Akansu, S. R. Kulkarni, and D. M. Malioutov, Eds., New York, NY, USA: Wiley, 2016, ch. 3, pp. 23–40.
[35] Z. Zhao and D. P. Palomar, "Mean-reverting portfolio design via majorization-minimization method," in *Proc. 50th Asilomar Conf. Signals Proc. Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2016, pp. 1530–1534.
[36] G. E. Box and G. C. Tiao, "A canonical analysis of multiple time series," *Biometrika*, vol. 64, no. 2, pp. 355–365, 1977.
[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] B. Kedem and S. Yakowitz, *Time Series Analysis by Higher Order Crossings*. Piscataway, NJ, USA: IEEE Press, 1994.
[39] J. Brodie, I. Daubechies, C. de Mol, D. Giannone, and I. Loris, "Sparse and stable Markowitz portfolios," in *Proc. Nat. Acad. Sci.*, vol. 106, no. 30, pp. 12267–12272, 2009.
[40] F. Bach, R. Jenatton, J. Mairaland, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, 2012.
[41] F. J. Fabozzi, S. M. Focardi, and P. N. Kolm, *Quantitative Equity Investing: Techniques and Strategies*. Hoboken, NJ, USA: Wiley, 2010.

[42] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm,” *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
[43] B. R. Marks and G. P. Wright, “A general inner approximation algorithm for nonconvex mathematical programs,” *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.
[44] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, “Decomposition by partial linearization: Parallel optimization of multi-agent systems,” *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.
[45] S. J. Wright and J. Nocedal, “Numerical optimization,” *Springer Sci.*, vol. 35, no. 67/68, p. 7, 1999.
[46] G. Scutari, F. Facchinei, and L. Lampariello, “Parallel and distributed methods for constrained nonconvex optimization—Part I: Theory,” *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 1929–1944, Apr. 2017.
[47] J. F. Sturm, “Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,” *Optim. Methods Softw.*, vol. 11, no. 1-4, pp. 625–653, 1999.
[48] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, “On the implementation and usage of SDPT3—a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0,” in *Handbook on Semidefinite, Conic and Polynomial Optimization*. New York, NY, USA: Springer, 2012, pp. 715–754.
[49] MOSEK ApS, *“The MOSEK Optimization Toolbox for MATLAB Manual Vers. 7.1 (Revision 28)*,” 2015. [Online]. Available: http://www.mosek.com
[50] J. Lofberg, “YALMIP: A toolbox for modeling and optimization in MATLAB,” in *Proc. IEEE Int. Symp. Comput. Aided Control Syst. Des.*, 2004, pp. 284–289.
[51] M. Grant, S. Boyd, and Y. Ye, “CVX: MATLAB software for disciplined convex programming,” 2008. [Online]. Available: http://cvxr.com/cvx
[52] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[53] D. P. Palomar, “Convex primal decomposition for multicarrier linear MIMO transceivers,” *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4661–4674, Dec. 2005.
[54] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, “Efficient projections onto the ℓ_1-ball for learning in high dimensions,” in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 272–279.
[55] M. Li, D. Sun, and K.-C. Toh, “A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization,” *SIAM J. Optim.*, vol. 26, no. 2, pp. 922–950, 2016.
[56] 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, Feb. 2016.



Ziping Zhao (S’16) received the B.Eng. degree in Electronics and Information Engineering (with highest honors) from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2014. Currently, he is working toward the Ph.D. degree at the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology (HKUST), Hong Kong. He was an awardee of the Hong Kong PhD Fellowship Scheme (HKPFS) and received the Overseas Research Awards at HKUST in 2018.

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

Rui Zhou (S’19) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2017. He is currently working toward the Ph.D. degree at the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong.

His research interests include optimization algorithms, statistical signal processing, machine learning, and financial engineering.

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 Scholar at Princeton University during 2004–2006.

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. 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 and big data analytics.

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.”