A Fast Successive QP Algorithm for General Mean-Variance Portfolio Optimization

Shengjie Xiu, Xiwen Wang, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—The mean and variance of portfolio returns are the standard quantities to measure the expected return and risk of a portfolio. Efficient portfolios that provide optimal trade-offs between mean and variance warrant consideration. To express a preference among these efficient portfolios, investors have put forward many mean-variance portfolio (MVP) formulations which date back to the classical Markowitz portfolio. However, most existing algorithms are highly specialized to particular formulations and cannot be generalized for broader applications. Therefore, a fast and unified algorithm would be extremely beneficial. In this paper, we first introduce a general MVP problem formulation that can fit most existing cases by exploring their commonalities. Then, we propose a widely applicable and provably convergent successive quadratic programming algorithm (SCQP) for the general formulation. The proposed algorithm can be implemented based on only the QP solvers and thus is computationally efficient. In addition, a fast implementation is considered to accelerate the algorithm. The numerical results show that our proposed algorithm significantly outperforms the state-of-the-art ones in terms of convergence speed and scalability.**

***Index Terms*—Mean-variance portfolios, successive quadratic programming algorithm, active set methods, Pareto frontier.**

I. Introduction

MEAN-VARIANCE analysis, pioneered by Harry Markowitz's publication in 1952 [1], is a breakthrough in modern portfolio theory. It starts a new era of financial research using quantitative tools and becomes one of the most widely-used investment decision rules among academics and practitioners. Mean-variance analysis assumes that the expected return and risk of a portfolio can be fully measured by the mean and variance of the portfolio return. It accords with the more general expected utility maximization when returns are assumed to be normally distributed [2]. In addition, it is the theoretical support of the capital asset pricing model (CAPM) developed by Sharpe [3] and Lintner [4].

According to mean-variance analysis, rational risk-averse investors would pursue efficient portfolios that provide the highest expected return subject to a certain level of risk. The set of efficient portfolios with different risk aversion levels is known as the efficient frontier. However, the number of efficient portfolios is usually infinite, which is of low applicability in practice. Therefore, to achieve a specific efficient portfolio corresponding to individual risk appetite, investors solve optimization problems based on well-designed trade-offs between means and variances. These problems are collectively referred to as mean-variance portfolio (MVP) problems, and their optimization process is called MVP optimization [5].

Many MVP problem formulations have been proposed in the literature. Markowitz portfolio, which optimizes a risk-adjusted return in the form of a quadratic utility, serves as a starting point [1], [6]. It has been popularly employed because it is simple for theoretical analysis and numerical optimization. Nonetheless, this formulation cannot fully characterize the preferences of a wide range of investors, so it has been extended in several directions. First, the risk-adjusted return is measured by other metrics like the Sharpe ratio [7], and the corresponding maximum Sharpe ratio portfolio (MSRP) also belongs to the efficient frontier [8]. Second, investors adopt other utility functions, including exponential and logarithmic ones [9], to represent preferences over portfolio returns. In this case, efficient portfolios could be obtained by applying mean-variance approximations to expected utility [10], [11]. Third, practical linear and quadratic constraints, such as the restrictions on desired return or risk, are included to take more investment guidelines into account [5], [12]. Since various MVP formulations contain different problem structures, a broadly applicable and numerically efficient algorithm is favored by academic research and the financial industry. However, there are two significant challenges in algorithm design.

Challenge #1: Existing efficient algorithms are problem-dependent and thus difficult to generalize. Markowitz portfolio gains popularity partly because it can be directly solved by efficient quadratic program (QP) solvers. Nevertheless, involving more intricate functional forms and constraints usually results in more complicated problems that cannot benefit from the same computational advantage of QPs. For example, given an MSRP which is a fractional program (FP) instead, we have to resort to FP algorithms such as the bisection method, Dinkelbach's algorithm [13], and the quadratic transform [14]. However, these algorithms are limited to certain types of FPs and cannot be applied to other problems [15]. As another example, given a worst-case robust global maximum return portfolio (worst-case robust GMRP), the majorization-minimization (MM) algorithm efficiently solves it via a sequence of quadratic upper-bound

Manuscript received 28 November 2022; revised 30 May 2023; accepted 4 July 2023. Date of publication 18 July 2023; date of current version 8 August 2023. This work was supported by the Hong Kong GRF under Grant 16207820. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Kunal Narayan Chaudhury. *(Corresponding author: Shengjie Xiu.)*

The authors are with the Hong Kong University of Science and Technology (HKUST), Clear Water Bay, Kowloon, Hong Kong (e-mail: sxiu@connect.ust.hk; xwangew@connect.ust.hk; palomar@ust.hk).

Digital Object Identifier 10.1109/TSP.2023.3295180

problems [16], [17]. Nonetheless, this construction of quadratic bounds relies on the structure of ℓ_2-norm, and upper-bound problems may not be easily identifiable unless specific convexity/concavity structures are contained in objective functions [18], [19]. Therefore, the scope of applicability of MM is also limited.

Challenge #2: General algorithms are inefficient in practice. We could turn to traditional off-the-shelf optimization methods to solve MVP problems. For example, feasible interior-point methods (FIP) [20] and feasible sequential quadratic programming (FSQP) [21] are commonly used. However, they are computationally expensive in most cases [22], especially for nonconvex problems [23]. Metaheuristic methods can be applied, including differential evolution and genetic algorithms. Nonetheless, they are inefficient as they require many objective function evaluations and barely exploit problem-specific information [24]. Although general solvers are available for some MVPs, it is widely accepted that QP algorithms are faster, more reliable, and easier to use.

In response to the challenges in designing an efficient and unified algorithm for MVP optimization, this work presents the following major contributions:

- We provide a general formulation that can characterize most of the existing MVP formulations.
- We propose an algorithm that solves different MVPs with various structures by uniformly solving a sequence of QPs. It can reuse the well-developed QP solvers, and has provable convergence and a fast implementation.
- We present extensive results to show that our proposed algorithm is more efficient and scalable than the state-of-the-art methods.

The rest of the paper is organized as follows. We begin with the general MVP formulation and its motivating examples in Section II. Then, in Section III, we propose our successive QP algorithm for the general problem formulation. The detailed analysis of the proposed algorithm is provided in Section IV, and the fast implementation that exploits the underlying sparsity pattern is developed in Section V. Section VI justifies the proposed algorithm's broad applicability and performance with comprehensive experiments. In Section VII, we provide a case study in parameter uncertainty to illustrate the practical usage of the proposed algorithm in portfolio backtesting. Finally, Section VIII concludes this paper.

II. Problem Formulation

Firstly, we introduce the notation and related background. We denote by $\mathbf{r} \in \mathbb{R}^N$ the returns of N assets. The mean vector and covariance matrix of the returns are represented by $\boldsymbol{\mu} \in \mathbb{R}^N$ and $\boldsymbol{\Sigma} \in \mathbb{S}^N_{++}$, respectively. We let $\mathbf{w} \in \mathbb{R}^N$ denote the portfolio weights. Then, the expected return and risk of this portfolio are individually measured by the mean ($\mathbf{w}^\mathsf{T}\boldsymbol{\mu}$) and variance ($\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}$) of the portfolio return ($\mathbf{w}^\mathsf{T}\mathbf{r}$).

The classical Markowitz portfolio can be obtained by solving the following QP problem:

$$\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \quad -\mathbf{w}^\mathsf{T}\boldsymbol{\mu} + \frac{\alpha}{2}\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}, \tag{1}$$

where $\alpha \geq 0$ is a risk aversion parameter that determines the trade-off between expected return and risk. $\mathcal{W}$ is the convex feasible set of $\mathbf{w}$ and is defined in its simplest form as

$$\mathcal{W} = \{\mathbf{w} \geq \mathbf{0},\ \mathbf{1}^\mathsf{T}\mathbf{w} = 1\}, \tag{2}$$

which includes the long-only and capital budget constraints [6]. In practice, this set may contain other constraints such as turnover control, upper bound limits, leverage constraints, etc. For illustrative purposes, we focus on the constraints in (2).

The original formulation proposed by Markowitz has been extended in two aspects. The first one is amending the mean-variance framework with a broad class of objective functions based on different combinations of means and variances to express investors' preferences. In practice, we may have multiple estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, denoted as $\{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \cdots, \boldsymbol{\mu}_p\}$ and $\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \cdots, \boldsymbol{\Sigma}_q\}$, by using different estimation methods or market regimes. For instance, one may use shrinkage estimators to address concerns around parameter uncertainty in the sample mean and covariance matrix (e.g., [25], [26]). Therefore, we can obtain multiple estimates of expected return and risk, denoted as vector-valued functions $\mathbf{x}(\mathbf{w}) : \mathcal{W} \to \mathbb{R}^p$ and $\mathbf{y}(\mathbf{w}) : \mathcal{W} \to \mathbb{R}^q_{++}$,

$$\begin{aligned} \mathbf{x}(\mathbf{w}) &= [x_1(\mathbf{w}), \ldots, x_p(\mathbf{w})], \\ \mathbf{y}(\mathbf{w}) &= [y_1(\mathbf{w}), \ldots, y_q(\mathbf{w})], \end{aligned} \tag{3}$$

where

$$x_i(\mathbf{w}) = \mathbf{w}^\mathsf{T}\boldsymbol{\mu}_i, \quad y_j(\mathbf{w}) = \mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}_j\mathbf{w}. \tag{4}$$

To represent the family of objective functions of means $\mathbf{x}(\mathbf{w})$ and variances $\mathbf{y}(\mathbf{w})$, we consider such functions in their most general form $F(\mathbf{x}, \mathbf{y}) : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$. The second aspect consists in adding minimum expected return or bearable maximum risk as practical constraints.

Therefore, the general MVP can be formulated as

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & f(\mathbf{w}) \triangleq F(\mathbf{x}(\mathbf{w}), \mathbf{y}(\mathbf{w})) \\ \mathsf{subject\ to} & \left.\begin{array}{l} x_i(\mathbf{w}) \geq a_i,\ i = 1, \ldots, p \\ y_j(\mathbf{w}) \leq b_j,\ j = 1, \ldots, q \\ \mathbf{w} \in \mathcal{W}, \end{array}\right\} \triangleq \mathcal{K} \end{array} \tag{$\mathcal{P}$}$$

where

- $f(\mathbf{w}) : \mathcal{K} \to \mathbb{R}$ is a continuously differentiable and possibly nonconvex function.
- $\mathcal{K}$ is the feasible set that contains at least one strictly feasible point, where a_i is the lower limit on the expected return, and b_j is the upper limit on the risk. The constraints on the expected return and risk are named mean-variance constraints. They can be written in a compact vector-valued function $\mathbf{g}(\mathbf{w}) : \mathcal{W} \to \mathbb{R}^p \times \mathbb{R}^q$ satisfying

$$\mathbf{g}(\mathbf{w}) \triangleq \begin{bmatrix} \mathbf{g}_x(\mathbf{w}) \\ \mathbf{g}_y(\mathbf{w}) \end{bmatrix} \leq \mathbf{0}, \tag{5}$$

with

$$[\mathbf{g}_x(\mathbf{w})]_i \triangleq a_i - x_i(\mathbf{w}),\ [\mathbf{g}_y(\mathbf{w})]_j \triangleq y_j(\mathbf{w}) - b_j. \tag{6}$$

TABLE I
FORMULATIONS OF WELL-RESEARCHED MVPS UNDER THE FRAMEWORK OF $\mathcal{P}$

Portfolio	$F(\mathbf{x}(\mathbf{w}), \mathbf{y}(\mathbf{w}))$	$\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$	Problem class
Markowitz portfolio	$-x(\mathbf{w}) + \frac{\alpha}{2} y(\mathbf{w})$	-	QP
MSRP	$-(x(\mathbf{w}) - r_f)/\sqrt{y(\mathbf{w})}$	-	FP
	$\sqrt{y(\mathbf{w})}/(x(\mathbf{w}) - r_f)$	-	
MGSRP	$-(x(\mathbf{w}) - r_f)/y(\mathbf{w})^\beta$	-	
Worst-case robust GMRP	$-x(\mathbf{w}) + \alpha\sqrt{y(\mathbf{w})}$	-	SOCP
Expected utility portfolio	$-U(x(\mathbf{w})) - \frac{1}{2}U^{''}(x(\mathbf{w}))y(\mathbf{w})$	-	Depends on U
Kelly portfolio	$-\log(1 + x(\mathbf{w})) + \frac{1}{2}\frac{y(\mathbf{w})}{(1+x(\mathbf{w}))^2}$	-	-
Return-constrained Markowitz portfolio	$y(\mathbf{w})$	$x(\mathbf{w}) \geq a$	QP
Risk-constrained Markowitz portfolio	$-x(\mathbf{w})$	$y(\mathbf{w}) \leq b$	QCQP

Compared with Problem (1), which well-developed QP solvers can efficiently solve, $\mathcal{P}$ is usually more complicated because of the substantial flexibility regarding the objective function f (i.e., F) and the inclusion of mean-variance constraints. Therefore, our goal is to efficiently deal with $\mathcal{P}$ by solving Problem (1) iteratively.

As the risk-return trade-off characterized by F should be reasonable, we require the following natural assumption.

Assumption 1: For each x_i and y_j that exists in F, we have

$$\nabla_{x_i} F(\mathbf{x}, \mathbf{y}) < 0, \quad \nabla_{y_j} F(\mathbf{x}, \mathbf{y}) > 0.$$

where $\nabla_{x_i} F(\mathbf{x}, \mathbf{y})$ and $\nabla_{y_j} F(\mathbf{x}, \mathbf{y})$ denote the partial gradient of F evaluated at x_i and y_j, respectively.[1]

Assumption 1 implies that function $F(\mathbf{x}, \mathbf{y})$ is decreasing with respect to x_i and increasing with respect to y_j. In other words, a higher expected return is always better, while more risk is always worse. It is a direct generalization of the continuity–monotonicity–finiteness axiom of the two-moment decision model [27].

A. Motivating Examples of General MVP

The general MVP formulation $\mathcal{P}$ can be easily customized to many well-researched portfolios. Representative portfolios and their classes of optimization problems are listed in Table I and elaborated next.[2]

1) On Objective Function F: Investors' preferences for different risk-return trade-off strategies can be modeled with different F. We provide some examples as follows.

MSRP: Sharpe ratio evaluates the expected return earned over the risk-free rate r_f per unit of volatility [7]. To find the portfolio that provides the maximum Sharpe ratio, we minimize the negative Sharpe ratio written as F_{SR},

$$F_{\text{SR}}(x(\mathbf{w}), y(\mathbf{w})) = -\frac{x(\mathbf{w}) - r_f}{\sqrt{y(\mathbf{w})}}, \tag{7}$$

or, alternatively, the inverse of Sharpe ratio. MSRP is a special case of the maximum generalized Sharpe ratio portfolio (MGSRP) introduced in [28] with the objective function F_{GSR},

$$F_{\text{GSR}}(x(\mathbf{w}), y(\mathbf{w})) = -\frac{x(\mathbf{w}) - r_f}{y(\mathbf{w})^\beta}, \tag{8}$$

where $\beta \geq 1/2$ is a risk aversion parameter. The corresponding optimization problems are all FPs.

Worst-case robust GMRP: The estimate of the expected return $\hat{\boldsymbol{\mu}}$ is inevitably subject to estimation error. The true $\boldsymbol{\mu}$ may be assumed to fall within an uncertainty ellipsoid shaped by $\boldsymbol{\Sigma}$, i.e., $\mathcal{U}_{\boldsymbol{\mu}} = \{\boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \alpha\boldsymbol{\Sigma}^{1/2}\mathbf{u} \,|\, \|\mathbf{u}\|_2 \leq 1\}$, where $\alpha > 0$ is a predefined parameter. Then, the objective function of the worst-case robust GMRP is written as $f_{\text{WC}}(\mathbf{w}) = -\min_{\boldsymbol{\mu}\in\mathcal{U}_{\boldsymbol{\mu}}} \mathbf{w}^\mathsf{T}\boldsymbol{\mu}$, which can be reformulated as the following F_{WC} according to [29],

$$F_{\text{WC}}(x(\mathbf{w}), y(\mathbf{w})) = -x(\mathbf{w}) + \alpha\sqrt{y(\mathbf{w})}. \tag{9}$$

The resulting optimizing problem can be recast as a second-order cone program (SOCP).

Expected utility portfolio: A utility function U measures the relative satisfaction with portfolio return. In this setting, the expected utility portfolio is attained by maximizing the expected utility, i.e.,

$$\underset{\mathbf{w}\in\mathcal{W}}{\text{maximize}} \;\mathbb{E}\left[U\left(\mathbf{w}^\mathsf{T}\mathbf{r}\right)\right]. \tag{10}$$

It is common and effective to apply mean-variance approximations to the expected utility [10], [30], i.e., performing Taylor expansion at the point $\mathbf{r} = \boldsymbol{\mu}$ and ignoring the moments greater than the second one:

$$\begin{aligned}\mathbb{E}\left[U\left(\mathbf{w}^\mathsf{T}\mathbf{r}\right)\right] &\approx U\left(\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right) + U^{'}\left(\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)\mathbb{E}\left[\mathbf{w}^\mathsf{T}\left(\mathbf{r} - \boldsymbol{\mu}\right)\right] \\ &\quad + \frac{1}{2}U^{''}\left(\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)\mathbb{E}\left[\left(\mathbf{w}^\mathsf{T}\left(\mathbf{r} - \boldsymbol{\mu}\right)\right)^2\right] \\ &= U\left(\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right) + \frac{1}{2}U^{''}\left(\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}. \end{aligned} \tag{11}$$

Therefore, the approximated expected utility as the objective function can be expressed as F_{EU},

$$F_{\text{EU}}(x(\mathbf{w}), y(\mathbf{w})) = -U(x(\mathbf{w})) - \frac{1}{2}U^{''}(x(\mathbf{w}))\,y(\mathbf{w}). \tag{12}$$

Kelly portfolio: To achieve maximum growth of wealth, Kelly portfolio maximizes the expected value of the logarithm of portfolio return [31], [32]. It coincides with the expected utility portfolio using the logarithmic utility function $U(\mathbf{w}^\mathsf{T}\boldsymbol{r}) = \log(1 + \mathbf{w}^\mathsf{T}\boldsymbol{r})$. Accordingly, the objective function F_{KL} of the Kelly portfolio is given by

[1] x_i (or y_j) exists in F iff $\nabla_{x_i} F$ (or $\nabla_{y_j} F$) does not always equal to 0.

[2] Scalar functions $x(\mathbf{w})$ and $y(\mathbf{w})$ denote a single expected return and a single risk of a portfolio, respectively.

$$F_{\mathrm{KL}}\left(x\left(\mathbf{w}\right), y\left(\mathbf{w}\right)\right) = -\log\left(1 + x\left(\mathbf{w}\right)\right) + \frac{1}{2}\frac{y\left(\mathbf{w}\right)}{\left(1 + x\left(\mathbf{w}\right)\right)^2}, \tag{13}$$

following the approximation (12), as shown in [33].

2) On Mean-Variance Constraints: The mean-variance constraints present in the following examples.

Alternatives of Markowitz portfolio: Markowitz portfolio (1) has two alternative formulations, which lead to the same efficient frontier. One minimizes the risk given a lower limit on the return, known as a return-constrained Markowitz portfolio:

$$\begin{aligned} &\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} && y\left(\mathbf{w}\right) \\ &\mathsf{subject\ to} && x\left(\mathbf{w}\right) \geq a. \end{aligned} \tag{14}$$

The other one maximizes the return given an upper limit on the risk, known as a risk-constrained Markowitz portfolio. It is formulated as the quadratically constrained quadratic program (QCQP):

$$\begin{aligned} &\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{maximize}} && x\left(\mathbf{w}\right) \\ &\mathsf{subject\ to} && y\left(\mathbf{w}\right) \leq b. \end{aligned} \tag{15}$$

We have seen that the general formulation $\mathcal{P}$ has a number of instances whose problem structures are distinct from each other. However, their relationship with a common QP representation can be explored, thus leading to an efficient successive QP algorithm introduced in the next section.

III. Proposed Algorithm

In this section, we propose an algorithm to solve $\mathcal{P}$ via a sequence of QP surrogate problems.

A. Algorithmic Framework

Unlike $\mathcal{P}$ which implicitly specifies investors' risk-return trade-off, the QP surrogate problem that we are interested in applies a straightforward characterization of this trade-off. It adopts the following formulation

$$\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \quad -\left(\boldsymbol{\lambda}_x + \boldsymbol{\eta}_x\right)^{\mathsf{T}}\mathbf{x}\left(\mathbf{w}\right) + \left(\boldsymbol{\lambda}_y + \boldsymbol{\eta}_y\right)^{\mathsf{T}}\mathbf{y}\left(\mathbf{w}\right), \tag{16}$$

whose objective function is a weighted combination of the means and variances. Specifically, $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_x; \boldsymbol{\lambda}_y]$ is the non-negative weight that characterizes the mean-variance trade-off reflected in f, and $\boldsymbol{\eta} = [\boldsymbol{\eta}_x; \boldsymbol{\eta}_y]$ is the non-negative weight that controls the magnitude of impact from $\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$. Without loss of generality, we assume that Problem (16) has a unique solution $\hat{\mathbf{w}}(\boldsymbol{\lambda}, \boldsymbol{\eta})$ at every iterate in our algorithm. By verifying the optimality conditions, $\hat{\mathbf{w}}(\boldsymbol{\lambda}, \boldsymbol{\eta})$ coincides with the stationary solution of $\mathcal{P}$ when $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are correctly chosen (see Section IV). Therefore, instead of solving the complicated $\mathcal{P}$ directly, we propose that by solving a sequence of QP surrogate problems (16), the weights could be dynamically adjusted such that $\hat{\mathbf{w}}(\boldsymbol{\lambda}, \boldsymbol{\eta})$ converges to the stationary solution of $\mathcal{P}$.

As mentioned, we already have efficient QP solvers for Problem (1). Interestingly, Problem (16) can be rewritten in the form of (1) as follows

$$\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \quad -\mathbf{w}^{\mathsf{T}}\bar{\boldsymbol{\mu}} + \frac{1}{2}\mathbf{w}^{\mathsf{T}}\bar{\boldsymbol{\Sigma}}\mathbf{w}, \tag{17}$$

where

$$\bar{\boldsymbol{\mu}} = \sum_{i=1}^{p}\left(\lambda_{x,i} + \eta_{x,i}\right)\boldsymbol{\mu}_i, \bar{\boldsymbol{\Sigma}} = \sum_{j=1}^{q} 2\left(\lambda_{y,j} + \eta_{y,j}\right)\boldsymbol{\Sigma}_j. \tag{18}$$

Therefore, solving Problem (16) shares the same computational convenience as dealing with Problem (1). Overall, this algorithmic framework can be efficient due to the low computational cost of QP surrogate problems with the undermentioned fast implementation presented in Section V and the simplicity of the weight updates introduced next.

We begin with a systematic manner that updates $\boldsymbol{\lambda}$ by iteratively approximating f using quadratic functions $\tilde{f}$. Given the current iterate $\mathbf{w}^k$, according to the successive convex approximation (SCA) framework [22], [34], we propose to optimize $\mathcal{P}$ by iteratively solving the following problem

$$\underset{\mathbf{w}\in\mathcal{K}}{\mathsf{minimize}} \ \tilde{f}\left(\mathbf{w}; \mathbf{w}^k\right) \triangleq \nabla_{\mathbf{x}} F\left(\mathbf{x}^k, \mathbf{y}^k\right)^{\mathsf{T}}\mathbf{x}\left(\mathbf{w}\right) + \nabla_{\mathbf{y}} F\left(\mathbf{x}^k, \mathbf{y}^k\right)^{\mathsf{T}}\mathbf{y}\left(\mathbf{w}\right), \tag{19}$$

where we denote $\mathbf{x}^k = \mathbf{x}(\mathbf{w}^k)$ and $\mathbf{y}^k = \mathbf{y}(\mathbf{w}^k)$ for notational simplicity. If $\mathcal{K} = \mathcal{W}$, i.e., the mean-variance constraints are not present, Problem (19) coincides with Problem (16) when $\boldsymbol{\lambda}$ is chosen as

$$\boldsymbol{\lambda}_x^k = -\nabla_{\mathbf{x}} F\left(\mathbf{x}^k, \mathbf{y}^k\right), \quad \boldsymbol{\lambda}_y^k = \nabla_{\mathbf{y}} F\left(\mathbf{x}^k, \mathbf{y}^k\right), \tag{20}$$

and $\boldsymbol{\eta}$ is ignored. In this way, $\mathcal{P}$ can be handled via a sequence of QP problems. However, if quadratic constraints are included, Problem (19) is still a QCQP, which has a much higher complexity than a QP.

To further recast Problem (19) as a sequence of QP (16), we apply the partial relaxation method [35]. Given Problem (19) with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^k$, we relax its mean-variance constraints and then optimize its Lagrangian dual problem

$$\underset{\boldsymbol{\eta}\geq\mathbf{0}}{\mathsf{maximize}} \ h\left(\boldsymbol{\eta}; \boldsymbol{\lambda}^k\right), \tag{21}$$

where h is the Lagrangian dual function of (19), and $\boldsymbol{\eta}$ is the Lagrange multiplier with respect to $\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$. At the lth iteration, the gradient projection method updates $\boldsymbol{\eta}^{l+1}$ by

$$\boldsymbol{\eta}^{l+1} = \left[\boldsymbol{\eta}^l + \alpha^l \nabla h(\boldsymbol{\eta}^l; \boldsymbol{\lambda}^k)\right]_+, \tag{22}$$

where α^l is a step-size and $[\cdot]_+ = \max\{0, \cdot\}$ denotes projection for nonnegative constraints. Following [36], the gradient of h can be computed as

$$\nabla h(\boldsymbol{\eta}^l; \boldsymbol{\lambda}^k) = \mathbf{g}(\hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l)). \tag{23}$$

Obviously, each evaluation of ∇h is attained by a QP (16). Therefore, Problem (19) is solved by alternatively conducting a QP (16) to compute ∇h and taking an inexpensive gradient projection step.

Algorithm 1 summarizes the proposed algorithm for $\mathcal{P}$ that consists of solving a sequence of QP (16), which is referred to as SCQP (SucCessive QP). It adopts a double-loop scheme: the outer loop decomposes $\mathcal{P}$ into problems (19) with quadratic objectives and iteratively updates $\boldsymbol{\lambda}$; the inner loop further recasts

TABLE II
CLOSED-FORM UPDATES OF $\boldsymbol{\lambda}$ AND $\boldsymbol{\eta}$ FOR WELL-RESEARCHED MVPS

Portfolio	Updates of $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$	
Markowitz portfolio	$\lambda_x \leftarrow 1$	$\lambda_y \leftarrow \alpha/2$
MSRP*	$\lambda_x \leftarrow 1$	$\lambda_y \leftarrow (x^k - r_f)/(2y^k)$
GMSRP*	$\lambda_x \leftarrow 1$	$\lambda_y \leftarrow \beta(x^k - r_f)/y^k$
Worst-case robust GMRP	$\lambda_x \leftarrow 1$	$\lambda_y \leftarrow \alpha/(2\sqrt{y^k})$
Expected utility portfolio	$\lambda_x \leftarrow U^{'}(x^k) + U^{'''}(x^k)y^k/2,$	$\lambda_y \leftarrow -U^{''}(x^k)/2$
Kelly portfolio	$\lambda_x \leftarrow 1/(1+x^k) + y^k/(1+x^k)^3$	$\lambda_y \leftarrow 1/(2(1+x^k)^2)$
Return-constrained Markowitz portfolio	$\eta_x \leftarrow [\eta_x + \alpha\left(a - x\left(\mathbf{w}^\star\left(\lambda_y, \eta_x\right)\right)\right)]_+$	$\lambda_y \leftarrow 1$
Risk-constrained Markowitz portfolio	$\eta_y \leftarrow [\eta_y + \alpha\left(y\left(\mathbf{w}^\star\left(\lambda_x, \eta_y\right)\right) - b\right)]_+$	$\lambda_x \leftarrow 1$

$^*\boldsymbol{\lambda}$ is scaled by a constant factor so that $\lambda_x \leftarrow 1$.

Algorithm 1 SCQP Algorithm for $\mathcal{P}$.

Input: $k = 0$, $\mathbf{w}^0 \in \mathcal{K}$, $\boldsymbol{\eta}^0 \geq \mathbf{0}$, and $\{\alpha^l\}, \{\gamma^k\} \in (0, 1]$;
repeat
  $\boldsymbol{\lambda}_x^k = -\nabla_{\mathbf{x}} F\left(\mathbf{x}^k, \mathbf{y}^k\right)$, $\boldsymbol{\lambda}_y^k = \nabla_{\mathbf{y}} F\left(\mathbf{x}^k, \mathbf{y}^k\right)$;
  if $\mathcal{K} = \mathcal{W}$ (no mean-variance constraints) **then**
    Compute $\hat{\mathbf{w}}^k = \hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \mathbf{0})$ via QP (16);
  else
    Set $l = 0$;.
    repeat
      Compute $\hat{\mathbf{w}}^k = \hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l)$ via QP (16);
      $\boldsymbol{\eta}^{l+1} = [\boldsymbol{\eta}^l + \alpha^l \mathbf{g}(\hat{\mathbf{w}}^k)]_+$;
      $l \leftarrow l + 1$;
    until convergence
  end if
  $\mathbf{w}^{k+1} = \mathbf{w}^k + \gamma^k(\hat{\mathbf{w}}^k - \mathbf{w}^k)$;
  $k \leftarrow k + 1$;
until convergence
Output: A stationary solution $\mathbf{w}^{k+1}$ of $\mathcal{P}$.

Problem (19) into QP surrogate problems (16) and iteratively updates $\boldsymbol{\eta}$. We remark again that Algorithm 1 can be generalized when $\mathcal{W}$ includes other constraints. Technical details of SCQP are deferred to Section IV.

Algorithm 1 can be simplified in two scenarios. First, if $\mathcal{K} = \mathcal{W}$, i.e., $\mathcal{P}$ does not include mean-variance constraints, we get rid of the difficulty mentioned in solving Problem (19), and thus we can set $\boldsymbol{\eta}$ to zero and ignore its update. Second, if the objective function f already satisfies the form of $\tilde{f}$, we maintain $\boldsymbol{\lambda}$ without tuning. Though $\mathcal{P}$ allows different problem structures that result in various difficulties in numerical optimization, with the proposed algorithm, we only require to focus on QP surrogate problems with simple updates of $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$. The applicability of the proposed algorithm will be described in detail in the next subsection.

B. Applicability of Proposed Algorithm

The proposed algorithm is broadly applicable in two aspects. The first aspect is that when dealing with different instances of $\mathcal{P}$, the modification only resides in the updates of $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$. Some closed-form updates are summarized in Table II, and we note that they are easily derived and almost computationally free. The second aspect is that SCQP, as a more general algorithmic framework, includes some problem-dependent algorithms as specific cases.

On the Connection to Quadratic Transform for MSRP: Under Assumption 1, the MSRP problem is equivalent to[3]

$$\underset{\mathbf{w} \in \mathcal{W}}{\text{minimize}} \quad -\frac{\left(x\left(\mathbf{w}\right) - r_f\right)^2}{y\left(\mathbf{w}\right)}. \tag{24}$$

Problem (24) can be solved by the quadratic transform [14] via a sequence of quadratic subproblems given as

$$\underset{\mathbf{w} \in \mathcal{W}}{\text{minimize}} \quad -\frac{2\left(\mathbf{x}^k - r_f\right)}{\mathbf{y}^k}\left(x\left(\mathbf{w}\right) - r_f\right) + \left(\frac{\mathbf{x}^k - r_f}{\mathbf{y}^k}\right)^2 y\left(\mathbf{w}\right). \tag{25}$$

Note that Problem (25) is in the form of Problem (16), and the weight updates coincide with that in SCQP after scaling. Hence, the quadratic transform is a special case of SCQP.

On the Connection to Dinkelbach's Algorithm for MGSRP: To solve the MGSRP problem whose objective is (8) given $\beta = 1$, Dinkelbach's algorithm [13] deals with the following surrogate problem

$$\underset{\mathbf{w} \in \mathcal{W}}{\text{minimize}} \quad -\left(x\left(\mathbf{w}\right) - r_f\right) + \left(\frac{\mathbf{x}^k - r_f}{\mathbf{y}^k}\right) y\left(\mathbf{w}\right), \tag{26}$$

equivalent to the one in SCQP. Therefore, SCQP can be readily specialized as Dinkelbach's algorithm.

On the Connection to MM for Worst-Case Robust GMRP: To solve the worst-case robust GMRP problem whose objective is (9), the MM algorithm [16] iteratively approximates the ℓ_2-norm and constructs a sequence of upper-bound problems

$$\underset{\mathbf{w} \in \mathcal{W}}{\text{minimize}} \quad -x\left(\mathbf{w}\right) + \frac{\alpha}{2}\left(\frac{y\left(\mathbf{w}\right)}{\sqrt{\mathbf{y}^k}} + \sqrt{\mathbf{y}^k}\right). \tag{27}$$

We notice that Problem (27) matches Problem (16) with the same updates of weights as SCQP, which means MM can be interpreted as a special case of SCQP.

We remark again that the existing algorithms are only preferred in specific cases when their surrogate problems are easy to compute. In contrast, our algorithmic framework is more accessible because it can always solve $\mathcal{P}$ via a series of simple QP surrogate problems. We refer the reader to Sections IV and V for more detailed analysis and design, and Section VI for the performance in representative applications.

[3]The satisfaction of Assumption 1 indicates $x(\mathbf{w}) > r_f$.

IV. Analysis of Proposed Algorithm

In the previous section, we have offered an overview of the proposed algorithm to solve $\mathcal{P}$. In this section, we provide more details and insights. In Subsection IV-A, we show how $\mathcal{P}$ and the proposed algorithm strongly relate to Pareto optimality. In Subsections IV-B and IV-C, we elaborate the technical details of constructing QP surrogate problems, and in Subsection IV-D, we provide the convergence analysis.

A. Insight From Multiobjective Optimization

We begin by interpreting general MVP optimization using the tools from multiobjective optimization. Given multiple estimates of expected return and risk in (3), portfolio selection can be treated as a multiobjective optimization problem

$$\underset{\mathbf{w}\in\mathcal{W}}{\text{minimize}}\ \{-x_1(\mathbf{w}),\ldots,-x_p(\mathbf{w}),y_1(\mathbf{w}),\ldots,y_q(\mathbf{w})\}. \quad (28)$$

Ideally, investors attempt to obtain the portfolio $\mathbf{w}\in\mathcal{W}$ that simultaneously maximizes all $x_i(\mathbf{w})$ and minimizes all $y_j(\mathbf{w})$. However, such a portfolio does not usually exist due to competing objectives. Therefore, Pareto optimal solutions, a.k.a. efficient portfolios, are favored by investors. Given a Pareto optimal solution, we cannot find any other solution that improves one objective (i.e., higher return or lower risk) without degrading at least one of the other objectives. The set of Pareto optimal solutions constitutes the Pareto frontier. Although $\mathcal{P}$ can represent various risk-return preferences by different formulations, its design should obey a rational risk-return tradeoff discipline and produce a Pareto optimal solution. For ease of interpretation, we denote the stationary solution of $\mathcal{P}$ as $\mathbf{w}^\star$ and make the following assumption.

Assumption 2: For each x_i and y_j that does not exist in F, we have

$$x_i(\mathbf{w}^\star)=a_i,\quad y_j(\mathbf{w}^\star)=b_j.$$

The above assumption requires that for all x_i and y_j that does not exist in F, their corresponding mean-variance constraints must be active at $\mathbf{w}^\star$. Otherwise, these estimates can be safely removed from $\mathcal{P}$ without affecting $\mathbf{w}^\star$. We note that this assumption is not required by the proposed algorithm. Then, the connection between $\mathcal{P}$ and Pareto optimality can be established as follows.

Lemma 1: Under Assumptions 1 and 2, every stationary solution of $\mathcal{P}$ is a Pareto optimal solution of Problem (28).

Fig. 1 depicts the relationship between several MVPs and the Pareto frontier. It is based on a bi-objective optimization problem with a single expected return $x(\mathbf{w})$ and a single risk $y(\mathbf{w})$. The feasible objective region in gray contains all possible objective vectors that could be achieved by any $\mathbf{w}\in\mathcal{W}$. Therefore, its upper left boundary, highlighted with a dashed curve, is the Pareto frontier. We observe that all the exhibited portfolios reside along the Pareto frontier. By changing the parameter α of the worst-case robust portfolio, we can also achieve different Pareto optimal solutions.

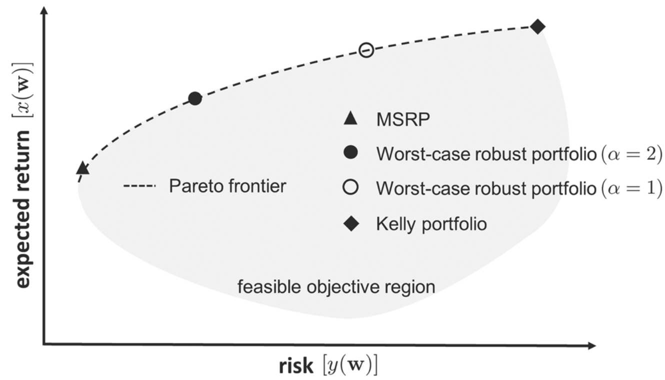


Fig. 1. MVPs in the risk-return objective space for illustrative purposes.

We now show how the QP surrogate problem (16) relates to the Pareto frontier. To characterize the Pareto optimal solutions, the weighting method solves the following problem

$$\underset{\mathbf{w}\in\mathcal{W}}{\text{minimize}}\ \ -\mathbf{v}_x^\mathsf{T}\mathbf{x}(\mathbf{w})+\mathbf{v}_y^\mathsf{T}\mathbf{y}(\mathbf{w}), \quad (29)$$

where $\mathbf{v}_x$ and $\mathbf{v}_y$ are non-negative coefficients. The weighing coefficients have the physical meaning of reflecting the relative importance of the objectives [37]. Previous studies have demonstrated that Problem (29) has two properties: every Pareto optimal solution of (28) can be found by (29) given proper $\mathbf{v}_x$ and $\mathbf{v}_y$; the unique solution of (29) is always Pareto optimal [38]. In our case, the surrogate problem (16) coincides with Problem (29) when $\mathbf{v}_x$ and $\mathbf{v}_y$ are chosen as

$$\mathbf{v}_x=\boldsymbol{\lambda}_x+\boldsymbol{\eta}_x,\quad \mathbf{v}_y=\boldsymbol{\lambda}_y+\boldsymbol{\eta}_y. \quad (30)$$

Therefore, the surrogate problem (16) inherits the properties of Problem (29). Based on the above findings, we reveal our proposed algorithm's insight. By dynamically adjusting $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$, SCQP tracks the Pareto frontier as every point in its solution sequence $\{\hat{\mathbf{w}}^k\}$ is Pareto optimal. When $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ converge, the algorithm terminates at the Pareto optimal solution that corresponds to the stationary solution of $\mathcal{P}$. This insight distinguishes our proposed algorithm from the general state-of-the-art methods. As the Pareto frontier has a number of attractive features, SCQP can benefit from them and has the fast implementation given in Section V.

Fig. 2 exhibits how iterates of FIP, FSQP, and SCQP shift in the risk-return objective space when solving $\mathcal{P}$. Specifically, for FIP and FSQP whose subproblems have no guarantee of Pareto optimal solutions, usually they do not achieve Pareto optimality before termination.

B. Outer Loop: Making Objective Function Quadratic

In Section III, we iteratively approximate f with the quadratic function $\tilde{f}$ in the form of the weighted combination of means and variances. When the outer loop converges to its fixed point, denoted by $\mathbf{w}^\star$ with a slight abuse of notation, we obtain

$$(\mathbf{z}-\mathbf{w}^\star)^\mathsf{T}\nabla\tilde{f}(\mathbf{w}^\star;\mathbf{w}^\star)\geq 0,\ \text{for all}\ \mathbf{z}\in\mathcal{K}, \quad (31)$$

from the optimality conditions of Problem (19). It is easy to observe that $\mathbf{w}^\star$ is also a stationary solution of $\mathcal{P}$ whose stationary

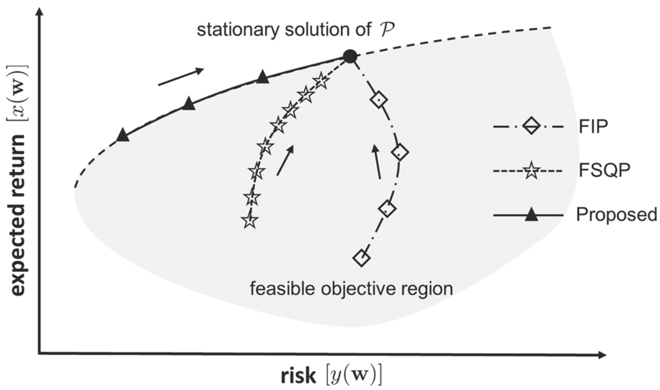


Fig. 2. Iterates of FIP, FSQP, and SCQP plotted in the risk-return objective space for illustrative purposes.

condition matches (31) as $\nabla f(\mathbf{w}^\star) = \nabla \tilde{f}(\mathbf{w}^\star; \mathbf{w}^\star)$. Therefore, the outer loop should converge only when it meets the stationary solution of $\mathcal{P}$. The rationality of using successive quadratic approximation with $\tilde{f}$ to solve $\mathcal{P}$ can be formally established in the SCA framework [22], [34] when the following additional assumption is made.

Assumption 3: At least one $y_j(\mathbf{w})$ exists in F. In addition, f has a Lipschitz continuous gradient on $\mathcal{K}$.

Note that Assumption 3 is quite standard and is satisfied by a large class of practical problems. Therefore, $\tilde{f}$ has the following properties: for all $\mathbf{w}^k \in \mathcal{K}$,

- $\tilde{f}(\mathbf{w}; \mathbf{w}^k)$ is differentiable and $\nabla f(\mathbf{w}^k) = \nabla \tilde{f}(\mathbf{w}^k; \mathbf{w}^k)$;
- $\tilde{f}(\mathbf{w}; \mathbf{w}^k)$ is strongly convex on $\mathcal{K}$.

These properties confirm that $\tilde{f}$ is a suitable approximant in the SCA framework, which requires consistency in the first-order derivative of f and strong convexity [17]. $\boldsymbol{\lambda}$ computed as the partial derivatives in (20) represents the intended preferences of f over the objectives, aligning with the interpretation of the weighing coefficients in Problem (29).

We note that there are other approximants under the SCA framework. They may have outstanding performance in risk parity portfolio optimization [23] and high-order portfolio optimization [39]. However, in our case, the proposed $\tilde{f}$ is preferred due to the following advantages. First, it does not require a proximal-like regularization term that adds to the strong convexity for convergence as it is strongly convex in nature by taking a positive weighted sum of $y_j(\mathbf{w})$. Second, it grasps the benefits of tracking the Pareto frontier for a much more efficient implementation.

C. Inner Loop: Dealing With Mean-Variance Constraints

When the mean-variance constraints exist in Problem (19), we apply the partial relaxation of $\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$ and then obtain the partial Lagrangian written as

$$L(\mathbf{w}, \boldsymbol{\eta}; \boldsymbol{\lambda}^k) = \tilde{f}(\mathbf{w}; \mathbf{w}^k) - \boldsymbol{\eta}_x^\mathsf{T} (\mathbf{x}(\mathbf{w}) - \mathbf{a}) + \boldsymbol{\eta}_y^\mathsf{T} (\mathbf{y}(\mathbf{w}) - \mathbf{b}), \tag{32}$$

where $\boldsymbol{\eta}_x \in \mathbb{R}_+^p$ and $\boldsymbol{\eta}_y \in \mathbb{R}_+^q$ are the Lagrangian multipliers associated with the constraints $\mathbf{x}(\mathbf{w}) \geq \mathbf{a}$ and $\mathbf{y}(\mathbf{w}) \leq \mathbf{b}$, respectively [35]. The Lagrangian dual function h is defined as the infimum of L over $\mathbf{w} \in \mathcal{W}$:

$$h(\boldsymbol{\eta}; \boldsymbol{\lambda}^k) = \inf_{\mathbf{w} \in \mathcal{W}} L(\mathbf{w}, \boldsymbol{\eta}; \boldsymbol{\lambda}^k), \tag{33}$$

where the infimum is achieved at $\tilde{\mathbf{w}}(\boldsymbol{\eta}; \boldsymbol{\lambda}^k)$.

The strong duality holds as Problem (19) is convex and has at least one strictly feasible point [40]. Therefore, instead of solving the primal problem, we can optimize its dual problem (21) using the gradient projection method. According to [36], the gradient of h can be computed as

$$\nabla h(\boldsymbol{\eta}; \boldsymbol{\lambda}^k) = \mathbf{g}(\tilde{\mathbf{w}}(\boldsymbol{\eta}; \boldsymbol{\lambda}^k)). \tag{34}$$

An important feature is that the partial minimization problem in (33) coincides with our QP surrogate problem (16), i.e., we have $\tilde{\mathbf{w}}(\boldsymbol{\eta}; \boldsymbol{\lambda}^k) = \hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta})$. In other words, each ∇h can be simply obtained from a QP (16) as shown in (23).

The Lagrangian multiplier $\boldsymbol{\eta}$ has the following explanation. Assume η_j corresponds to the risk constraint $y_j(\mathbf{w}) \leq b_j$. If this constraint is satisfied with strict inequality $y_j(\mathbf{w}) < b_j$, η_j will decrease according to (22) and (23). From the view of the weighing method, the weight on $y_j(\mathbf{w})$ will decrease, which means more risk $y_j(\mathbf{w})$ is preferred in exchange for a better risk-return trade-off in the next iteration. Otherwise, if the risk constraint is violated with $y_j(\mathbf{w}) > b_j$, η_j will increase, and thus the concern about the risk $y_j(\mathbf{w})$ will rise. The above interpretation can also be applied to the expected return constraint $x_i(\mathbf{w}) \geq a_i$.

D. Convergence Analysis

In this subsection, we provide the convergence of the proposed SCQP. We note that the convergence of the outer loop is based on that of the inner loop.

To ensure the convergence of the inner loop, we choose α^l according to the Armijo rule along the projection arc [36]. To be more specific, selecting $\alpha^{\text{init}} > 0$, and $\sigma, \beta \in (0,1)$, α^l is chosen to be the largest element in $\{\alpha^{\text{init}} \beta^j\}_{j=0,1,\ldots}$ satisfying

$$h(\boldsymbol{\eta}^{l+1}; \boldsymbol{\lambda}^k) - h(\boldsymbol{\eta}^l; \boldsymbol{\lambda}^k) \geq \sigma \mathbf{g}(\hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l))^\mathsf{T} (\boldsymbol{\eta}^{l+1} - \boldsymbol{\eta}^l), \tag{35}$$

where

$$\boldsymbol{\eta}^{l+1} = \left[\boldsymbol{\eta}^l + \alpha^l \mathbf{g}(\hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l))\right]_+. \tag{36}$$

Then, the convergence of the inner loop can be established as follows.

Proposition 1: Under Assumption 1, suppose α^l is chosen according to Armijo rule along the projection arc, then the sequence $\{\hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l)\}_{l=1}^\infty$ generated by the inner loop of Algorithm 1 converges to the optimal solution of Problem (19).

Proof: The convergence of the gradient projection method follows directly from [36], Proposition 2.3.3]. It states that every limit point of $\{\boldsymbol{\eta}^l\}_{l=1}^\infty$ is a stationary solution of the dual problem (21). Given the strong duality holds, the generated sequence $\{\hat{\mathbf{w}}(\boldsymbol{\lambda}^k, \boldsymbol{\eta}^l)\}_{l=1}^\infty$ converges to the optimal solution of Problem (19). ■

Based on the convergence of the inner loop, we next analyze the convergence of the outer loop.

Proposition 2: Under Assumption 1 and 3, suppose $\gamma^k \in (0,1]$, $\gamma^k \to 0$ and $\sum_k \gamma^k = +\infty$. Then, either Algorithm 1

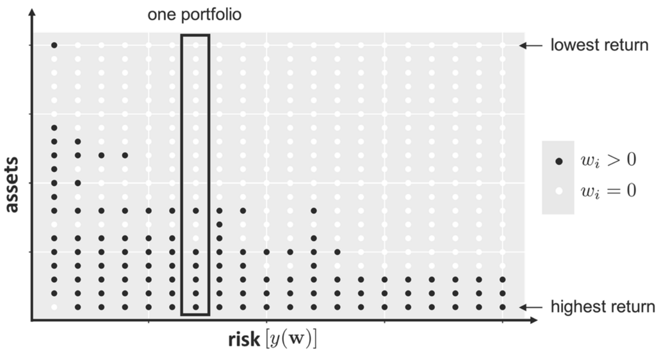


Fig. 3. Sparsity pattern in efficient portfolios.

converges in a finite number of iterations to a stationary solution of $\mathcal{P}$ or every limit point of the solution sequence $\{\mathbf{w}^k\}_{k=1}^{\infty}$ (at least one such point exists) is a stationary solution of $\mathcal{P}$.

Proof: Note that problem $\mathcal{P}$ has a feasible region $\mathcal{K}$ that is closed, bounded, convex and nonempty. Thus, [34, Assumptions A1–A4] hold, and the proof of Proposition 3 follows directly from [34, Theorem 3]. ∎

Thus, by proving Propositions 2 and 3, we prove the convergence of the proposed algorithm.

V. Fast Implementation

Since the updates of $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are generally cheap as mentioned, the major cost of the proposed algorithm comes from solving the sequence of QP surrogate problems (16). Although open-source QP solvers are well established and considered efficient for a single problem, solving a sequence of QP problems is not a small workload. Therefore, it is necessary and attractive to explore the hidden structures of SCQP for potential significant speed-up. In this section, we first introduce the sparsity pattern related to the Pareto frontier, and then propose a novel active-set strategy to take advantage of this sparsity pattern for a fast implementation of SCQP.

A. Sparsity Pattern on Pareto Frontier

One of the crucial features of the Pareto frontier is the sparsity pattern. Previous studies have shown that Pareto optimal solutions are naturally sparse under the long-only constraint [41], [42]. One explanation is that the long-only constraint acts like the ℓ_1-norm penalty which is extensively used to promote sparsity [43].

To better illustrate the sparsity pattern, Fig. 3 visualizes the efficient portfolios of $N = 20$ given different levels of risk aversion, which is based on the trade-off between a single expected return $x(\mathbf{w})$ and a single risk $y(\mathbf{w})$ in the S & P 500 market. Two observations from Fig. 3 should be noted. First, the sparsity is remarkable, and the portfolios are concentrated on a small number of assets with high returns, especially when the risk tolerance is high. Second, the sparsity pattern between any two neighboring portfolios is similar.

The above two observations have important algorithmic implications. The first observation motivates us to solve each QP surrogate problem (16) via an equivalent dimension-reduced problem defined on only a small portion of assets corresponding to $w_i > 0$. Because the complexity of the QP solver is around $\mathcal{O}(N^3)$ [44], [45], solving the dimension-reduced problem can be computationally much cheaper than directly handling the original problem, especially when N is large. The second observation inspires that the solution of the previous QP can be used to "warm start" the solution of the next so that QP algorithms benefit from the similar sparsity pattern and thus have faster convergence. The following subsection details how we achieve the above speedup.

B. New Active-Set Strategy for QP

We focus on the general form of the QP surrogate problem (16) given as

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^N}{\text{minimize}} \quad & q(\mathbf{w}) = \mathbf{c}^\mathsf{T}\mathbf{w} + \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{H}\mathbf{w} \\ \text{subject to} \quad & \mathbf{A}\mathbf{w} = \mathbf{b}, \quad \mathbf{l} \le \mathbf{w} \le \mathbf{u}, \end{aligned} \tag{37}$$

where $\mathbf{H} \in \mathbb{S}^N_{++}$, $\mathbf{A} \in \mathbb{R}^{M\times N}$, $\mathbf{b} \in \mathbb{R}^M$, and $\mathbf{l}, \mathbf{u} \in \mathbb{R}^N$ that specify bounds on the variables. Specifically, suppose $\mathbf{A} = \mathbf{1}^\mathsf{T}$, $\mathbf{b} = 1$, $\mathbf{l} = \mathbf{0}$, and $\mathbf{u} = +\infty$, the set of constraints in Problem (37) specializes to $\mathcal{W}$. Empirically, a large proportion of the variables w_i will touch the boundaries, i.e., $w_i = l_i$ or $w_i = u_i$, when they meet the optimality, which results in sparsity in our case. Hence, if we know such variables in advance, we can instead search for the solution of Problem (37) on a subspace spanned by the remaining variables or, equivalently, solve the following small-sized problem

$$\begin{aligned} \underset{\mathbf{w}\in\mathbb{R}^N}{\text{minimize}} \quad & \mathbf{c}^\mathsf{T}\mathbf{w} + \frac{1}{2}\mathbf{w}^\mathsf{T}\mathbf{H}\mathbf{w} \\ \text{subject to} \quad & \mathbf{A}\mathbf{w} = \mathbf{b}, \quad l_i \le w_i \le u_i,\ i \notin \bar{L}\cup\bar{U}, \\ & w_i = l_i,\ i \in \bar{L}, \quad w_i = u_i,\ i \in \bar{U}, \end{aligned} \tag{38}$$

where $\bar{L}\cup\bar{U}$ is called a working set, and $\bar{L}\cap\bar{U} = \emptyset$. Problem (38) is generally low-cost because it can be reduced to an $N - |\bar{L}\cup\bar{U}|$ dimensional QP by eliminating the variables fixed on the bound. By analyzing the difference in optimality conditions of Problems (37) and (38), solving the reduced problem is equivalent to solving the original problem if and only if the following conditions are satisfied

$$\beta_i^l \ge 0, \quad \forall i \in \bar{L}, \quad \beta_i^u \ge 0, \quad \forall i \in \bar{U}, \tag{39}$$

where β_i^l and β_i^u are the Lagrangian multipliers corresponding to $w_i = l_i$ and $w_i = u_i$.

The above findings further inspire our new active-set strategy in two aspects. First, we hope that the size of the working set can be as large as possible for the sake of computational efficiency. Second, if the current subproblem (38) cannot solve (37), the violation of condition (39) instructs the update of the working set. The detailed procedure is given as follows.

We begin with a feasible subproblem (38) with a large working set $\bar{L}^0\cup\bar{U}^0$. At the kth iteration, we compute $\mathbf{w}^k$ and $(\boldsymbol{\beta}^l, \boldsymbol{\beta}^u)$ as the primal and dual optimal solutions of the subproblem (38) based on the current working set $\bar{L}^k\cup\bar{U}^k$. If condition (39) is satisfied, the algorithm terminates as $\mathbf{w}^k$ meets the optimality conditions of Problem (37). Otherwise,

Algorithm 2 New Active-set Strategy for Problem (37).

Input: $\bar{L}^0$ and $\bar{U}^0$;

for $k = 0, 1, 2, \ldots$ **do**
  Compute $\mathbf{w}^k$, $\boldsymbol{\beta}^l$, and $\boldsymbol{\beta}^u$ via (38) given $\{\bar{L}^k, \bar{U}^k\}$;
  if $\min(\boldsymbol{\beta}^l) < 0$ or $\min(\boldsymbol{\beta}^u) < 0$ **then**
    $\bar{L}^{k+1} = \bar{L}^k \setminus \{i \,|\, \beta_i^l < 0\}$;
    $\bar{U}^{k+1} = \bar{U}^k \setminus \{i \,|\, \beta_i^u < 0\}$;
  else
    Stop;
  end if
end for

Output: The optimal solution $\mathbf{w}^k$ of Problem (37).

the objective function $q(\mathbf{w})$ may be further decreased by relaxing the bounded variables corresponding to negative multipliers. Hence, the working set for the subsequent iteration should be adjusted as $\bar{L}^{k+1} = \bar{L}^k \setminus \{i \,|\, \beta_i^l < 0\}$ and $\bar{U}^{k+1} = \bar{U}^k \setminus \{i \,|\, \beta_i^u < 0\}$. We formally describe this new active-set strategy in Algorithm 2. Its convergence is summarized in the following proposition.

Proposition 3: The sequence $\{\mathbf{w}^k\}$ generated by Algorithm 2 converges to the optimal solution of Problem (37) within $|\bar{L}^0 \cup \bar{U}^0|$ iterations.

Proof: See Appendix A. ∎

Warm Starting for a Sequence of QP: In practice, Problem (37) can be solved in much fewer iterations than the theoretical result in Proposition 4. When dealing with a sequence of related QPs, the number of iterations can be further reduced using warm starting. This is because the optimal working set stays mostly the same from one QP to the next because of similar data input, manifesting as the similar sparsity pattern in our case. To take advantage of this prior knowledge, the new active-set strategy applies the warm starting that uses the optimal working set of the former QP as an initial guess. In numerical experiments, Algorithm 2 with the warm starting typically solves each QP surrogate problem (16) in less than three iterations, significantly reducing the cost of SCQP.

VI. Numerical Experiments

We evaluate the applicability and performance of SCQP by conducting experiments on representative MVPs.

A. Experiment Set-Up

1) Real Market Data: To evaluate the performance of SCQP, we perform experiments on historical daily price time series data. Each dataset contains N stocks randomly chosen from the S & P 500 index, and a time period of $5N$ continuous trading days is randomly picked over the long period from 2008-12-01 to 2018-12-01. All results are averaged over 20 independent realizations.

2) Benchmarks: The benchmarks for different MVPs are usually different. Specific problems can be solved either individually or jointly by the following methods:

- ECOS: an open-source conic optimization software using the interior point method [46]; efficient for small and medium-sized SOCP problems.
- MOSEK: a commercial conic optimization software using the interior point method [47]; performs closely to other commercial solvers like GUROBI and CPLEX; efficient for large-scale SOCP problems.
- Dinkelbach's algorithm & quadratic transform: iterative methods for FP problems proposed in [13], [14].
- Majorization-minimization (MM): an iterative method that solves difficult optimization problems by solving a series of upper-bound surrogate problems [16], [17].
- NLopt: a standard library for nonlinear optimization [48]; applicable to general problems.
- DEoptim and GA: libraries of differential evolution and genetic algorithm [49], [50]; applicable to general problems.

Note that all the benchmarks safeguard the convergence, and thus our comparisons focus on their convergence speed. It is generally difficult to have a unified termination criterion for all methods as interior-point solvers focus on the tolerance of primal and dual feasibility while the other methods do not. Therefore, we set the termination criterion of ECOS and MOSEK to their default options, and terminate the other methods when

$$|\mathbf{w}^{k+1} - \mathbf{w}^k| \leq 10^{-6}. \tag{40}$$

3) Implementation Details: All experiments were carried out on 3.40GHz Intel Xeon Gold 6246R machines with 80G RAM running R 3.6.3. The inner QP solver of SCQP is the open-source quadprog [51]. In addition, we allow DEoptim and GA to run in parallel mode with the same initial population of size 2000 as the starting population, while the other methods run on only one core.

B. Application I: Worst-Case Robust GMRP

We consider the worst-case robust GMRP formulated as

$$\underset{\mathbf{w} \in \mathcal{W}}{\mathsf{minimize}} \quad -\mathbf{w}^\mathsf{T}\boldsymbol{\mu} + \alpha\sqrt{\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}}, \tag{41}$$

where $\alpha > 0$ is the predefined parameter. Without loss of generality, we consider $\alpha = 1$. Problem (41) does not include the mean-variance constraint $\mathbf{g}(\mathbf{w}) \leq \mathbf{0}$, so the inner loop of SCQP is not required. It is suitable for evaluating the efficiency of the outer loop in dealing with convex objective functions. Problem (41) can be converted in the standard way into a SOCP. Thus, SOCP solvers (ECOS and MOSEK), NLopt, DEoptim, and GA serve as benchmarks.

Fig. 4(a) shows the convergence of different methods in one realization of Problem (44) given $N = 200$. The gap is defined as the absolute difference between the objective value and the minimum one obtained by all the methods. The plot shows that the proposed SCQP takes substantially less time than the other methods to reach the gap of 10^{-9}. We also observe that SCQP consumes fewer iterations. Besides, ECOS and MOSEK converge with similar performance as they share

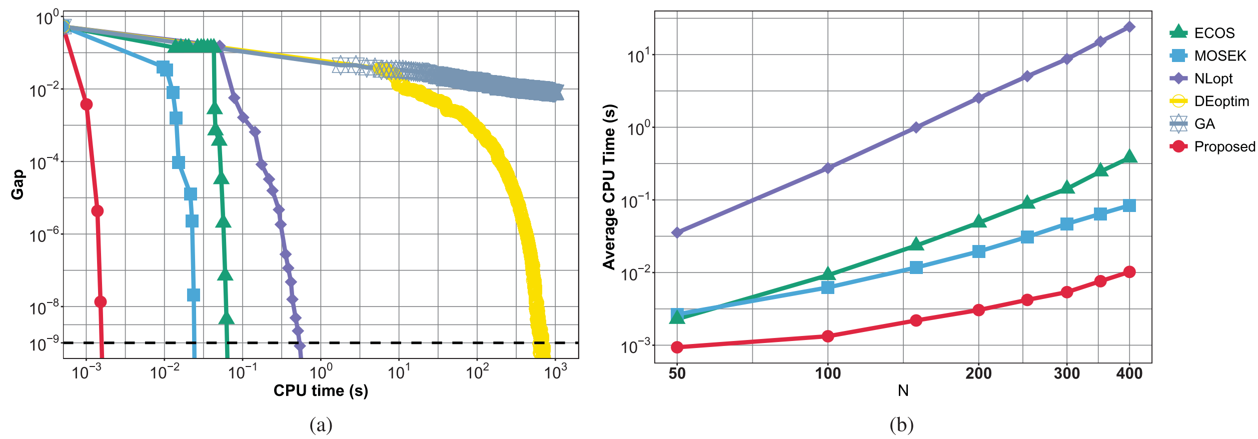


Fig. 4. Numerical experiments on solving worst-case robust GMRP (41). (a) Convergence of algorithms in a realization with $N = 200$. (b) Timing results of algorithms with different N.

TABLE III
EMPIRICAL TIME COMPLEXITY ORDER

Application	Empirical time complexity order $\mathcal{O}(N^c)$
I: (41)	Proposed (1.155) < MOSEK (1.691) < ECOS (2.453) < NLopt (3.134)
II: (42)	Proposed (0.944) < MM (1.620) < NLopt (3.241)
III: (43)	Proposed (1.137) < MOSEK (1.615) < ECOS (2.365) < NLopt (3.140)
IV: (44)	Proposed (0.768) < Dinkelbach (1.786) < QT (1.853) < NLopt (3.085)

similar interior point methods. As for the metaheuristic methods, DEoptim converges to an accurate solution but has a much slower rate, while GA cannot obtain good solutions within 1000s.

Fig. 4(b) compares the CPU time of different methods across different problem sizes N. DEoptim and GA are omitted because they are time-consuming in this application. The results show that SCQP consistently outperforms all the other established solvers concerning computation time. More specifically, when $N = 50$, the proposed algorithm is $\sim 2.5\times$ faster than ECOS; when $N \geq 100$, the proposed algorithm has a $4.7 \sim 8.7\times$ faster speed than MOSEK. Another observation is that ECOS performs better than MOSEK in small problems ($N = 50$) and worse when the problem dimension becomes larger, which is also verified by [46].

Table III presents the empirical time complexity order of different methods using $\mathcal{O}(N^c)$ notation. They are obtained by fitting the average CPU time vs N curves (in a log-log scale) with linear functions. We observe that SCQP has the lowest empirical time complexity which is nearly linear. Besides, NLopt has an empirical time complexity around $\mathcal{O}(N^3)$, which coincides with the discussion in [48]. Combined with Fig. 4(b), these results show that SCQP is more scalable than the benchmarks. The major reason is that the sparsity pattern becomes prominent in high-dimensional problems, while SCQP benefits from it using the new active-set strategy for low computational cost.

C. Application II: Kelly Portfolio

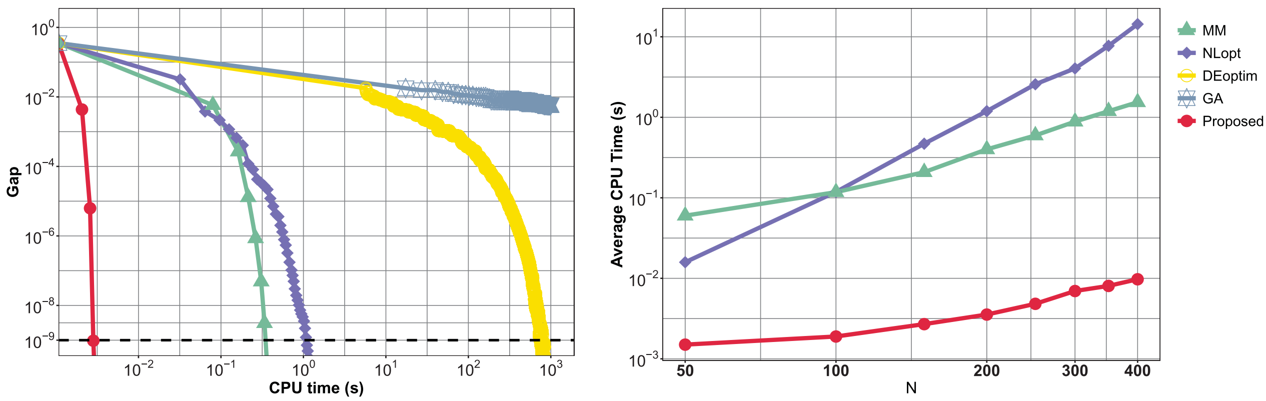
In the first application, we show that the outer loop efficiently deals with convex objective functions. Next, we focus on an MVP with a nonconvex objective function. Recall the Kelly portfolio formulated as

$$\underset{\mathbf{w}\in\mathcal{W}}{\text{minimize}} \quad -\log\left(1+\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)+\frac{1}{2}\frac{\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}}{\left(1+\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)^2}. \tag{42}$$

The challenge of solving Problem (42) is that it is a nonconvex problem with a ratio term in the objective function. Similar to Problem (41), the inner loop of SCQP is also not required.

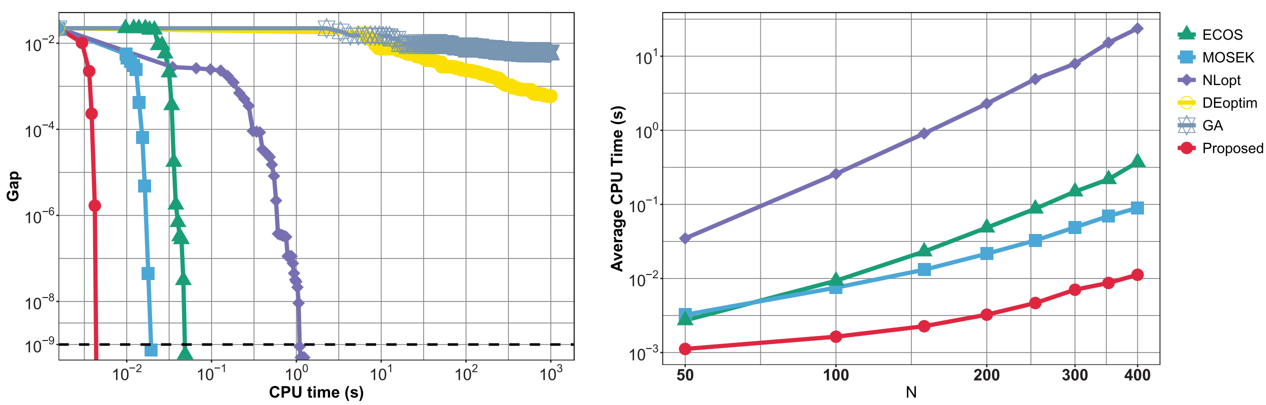
It is difficult to apply Dinkelbach's algorithm as Problem (42) is not a traditional single-ratio FP. Alternatively, we introduce the MM method that recasts the problem to a sequence of convex conic subproblems in Appendix B. In addition, NLopt, DEoptim, and GA are also applicable. Fig. 5(a) shows the convergence of different methods in one realization of Problem (42) given $N = 200$. The proposed SCQP converges to the gap of 10^{-9} in very few iterations and exhibits superior performance over the benchmarks. MM and NLopt are slower than SCQP by at least two orders of magnitude. Similar to the first application, DEoptim and GA have slow convergence or cannot obtain appropriate solutions within the time limit. The results show the high efficiency of the proposed algorithm in dealing with nonconvex objective functions.

Fig. 5(b) compares the CPU time of SCQP, MM, and NLopt for different N. As expected, the proposed SCQP algorithm shows a significant gain. When $N = 50$, the convergence speed achieved by SCQP is $\sim 10.5\times$ higher than that of NLopt. This difference tends to grow in favor of SCQP as the problem size increases. Besides, MM is more competitive than NLopt when more assets are considered. This may be because MM requires calling the inner SOCP solver, and the setup time of the solver is less significant compared with the solving time when the problem size is large. Table III also shows that SCQP enjoys



(a) (b)

Fig. 5. Numerical experiments on solving Kelly portfolio (42). (a) Convergence of algorithms in a realization with $N = 200$. (b) Timing results of algorithms with different N.



(a) (b)

Fig. 6. Numerical experiments on solving risk-constrained Markowitz portfolio (43). (a) Convergence of algorithms in a realization with $N = 200$. (b) Timing results of algorithms with different N.

the best scalability with the lowest empirical time complexity order.

D. Application III: Risk-Constrained Markowitz Portfolio

The third application we consider is the risk-constrained Markowitz portfolio formulated as

$$\begin{aligned} \underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \quad & -\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu} \\ \mathsf{subject\ to} \quad & \mathbf{w}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{w} \le b. \end{aligned} \tag{43}$$

Since its objective function satisfies the form of $\tilde{f}$, only the inner loop of SCQP is required. Therefore, it is suitable for evaluating the efficiency of the inner loop in dealing with mean-variance constraints. Given that the equally weighted portfolio $\bar{\mathbf{w}} = \mathbf{1}/n$ achieves the risk $r = \bar{\mathbf{w}}^{\mathsf{T}}\boldsymbol{\Sigma}\bar{\mathbf{w}}$, here we set the risk limit as $b = r$.

Problem (43) is in a standard SOCP form. Therefore, we include SOCP solvers (ECOS and MOSEK), NLopt, DEoptim, and GA as benchmarks. Fig. 6(a) shows the convergence of different methods in one realization of Problem (43) given $N = 200$. The result indicates that SCQP is far better than the benchmarks regarding convergence speed. Compared with Fig. 4(a), metaheuristic methods perform worse due to the difficulty in handling nonlinear constraints.

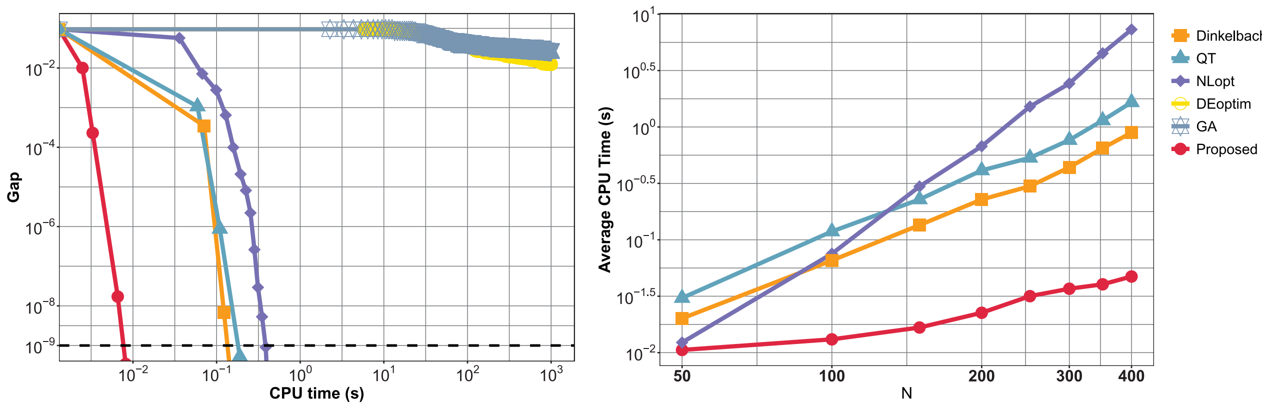
Fig. 6(b) compares the CPU time of all methods except DEoptim and GA across different problem sizes N. Compared with Fig. 4(b), we have two observations. First, SCQP is still much faster than the benchmarks, especially when the dimension is high. This fact shows that the inner loop handles the mean-variance constraints efficiently, and taking advantage of sparsity is necessary. Second, SOCP solvers and NLopt have consistent performance in optimizing different SOCP problems. These observations are also corroborated by the similar empirical time complexity order in Applications I and III, shown in Table III.

E. Application IV: Long-Term MSRP With Short-Term Goals

The fourth application we consider is the maximization of the long-term Sharpe ratio along with the goals of running over the short-term market. It is formulated as

$$\begin{aligned} \underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \quad & -\frac{\mathbf{w}^{\mathsf{T}}\boldsymbol{\mu}_1}{\sqrt{\mathbf{w}^{\mathsf{T}}\boldsymbol{\Sigma}_1\mathbf{w}}} \\ \mathsf{subject\ to} \quad & \mathbf{w}^{\mathsf{T}}\boldsymbol{\mu}_2 \ge a, \quad \mathbf{w}^{\mathsf{T}}\boldsymbol{\Sigma}_2\mathbf{w} \le b. \end{aligned} \tag{44}$$

Without loss of generality, $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are estimated from the whole period of $5N$ trading days, while $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}_2$ are estimated from the latest $2N$ trading days. Given that the equally



(a) (b)

Fig. 7. Numerical experiments on solving long-term MSRP with short-term goals (44). (a) Convergence of algorithms in a realization with $N = 200$. (b) Timing results of algorithms with different N.

weighted portfolio $\bar{\mathbf{w}} = \mathbf{1}/n$ achieves the short-term expected return $e = \bar{\mathbf{w}}^\mathsf{T}\boldsymbol{\mu}_2$ and risk $r = \bar{\mathbf{w}}^\mathsf{T}\boldsymbol{\Sigma}_2\bar{\mathbf{w}}$, we set $a = 1.2e$ and $b = 0.8r$, representing the interest in higher return and lower risk than the short-term market.

Dinkelbach's algorithm and the quadratic transform (QT) solve Problem (44). Besides, NLopt, DEoptim, and GA also serve as benchmarks. Fig. 7(a) shows the convergence of different methods in one realization of Problem (44) given $N = 200$. We observe that the proposed SCQP reaches the gap of 10^{-9} faster than Dinkelbach's algorithm and the quadratic transform by more than one order of magnitude.

Fig. 7(b) shows the average CPU time across different problem sizes N. We observe that SCQP has the best performance, consistently outperforming the benchmarks. Compared with the first three applications that solely require the inner or outer loop, SCQP takes longer solving time in this case because both loops are required. Moreover, the performance of Dinkelbach's algorithm is fairly close to that of the quadratic transform. One possible reason is that both FP algorithms share the same algorithmic framework that recasts Problem (44) into a sequence of conic subproblems. In addition, SCQP is far more efficient than NLopt when the dimension increases. The empirical time complexity order in Table III further supports our findings.

VII. Practical Usage in Portfolio Backtesting: Parameter Uncertainty Case Study

Portfolio backtesting involves running investment strategies using historical prices to evaluate their performance in the past. However, this process poses two significant challenges. Firstly, executing trades for a large number of stocks over several decades is computationally intensive. Secondly, evaluating numerous trading strategies using different solvers is less manageable. Our proposed algorithm can efficiently address these challenges in a unified manner in mean-variance settings.

To demonstrate the universality and efficiency of our method in handling different formulations, we present a case study that focuses on evaluating various strategies to deal with parameter uncertainty. It is important to note that the goal of this case study is not to provide an exhaustive comparison.

As previous studies have shown, mean-variance portfolios face the issue of parameter estimation error as the number of assets N increases and approaches the sample size T. For comparison purposes, we assume that the mean is known (e.g., obtained from factor models) and evaluate different methods to enhance the robustness of the maximum Sharpe ratio portfolio (MSRP)

$$\underset{\mathbf{w}\in\mathcal{W}}{\text{minimize}} \quad -\frac{\mathbf{w}^\mathsf{T}\boldsymbol{\mu}}{\sqrt{\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}}}. \tag{45}$$

Three class of approaches are considered.

1) Shrinkage Methods: Shrinkage covariance estimators are proposed to mitigate the undesired impact of sample covariance matrix.[4] We examine two shrinkage covariance estimators, namely the linear Ledoit-Wolf estimator [25] and the nonlinear Ledoit-Wolf estimator [26]. Our framework supports both of these estimators, as discussed in Section II.

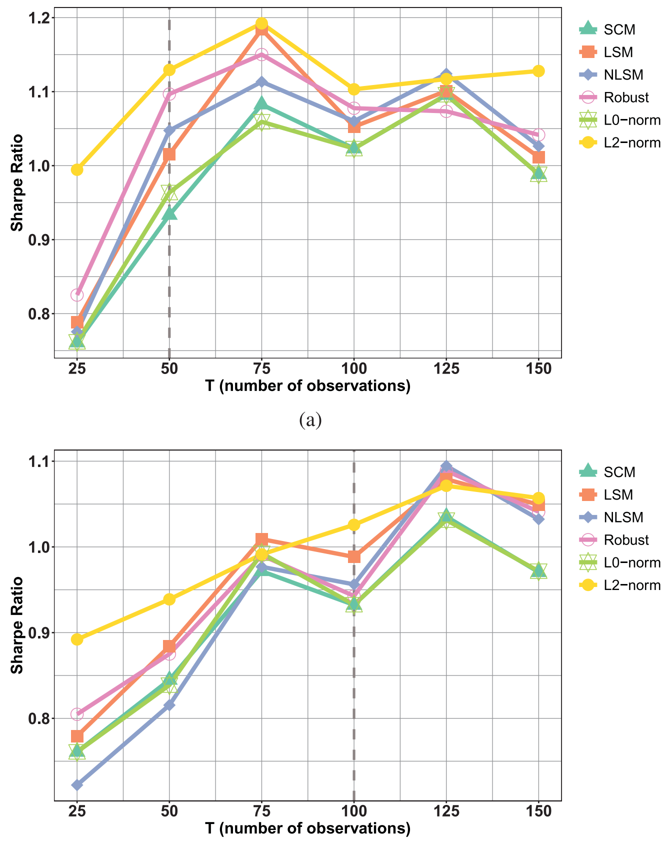
2) Robust Optimization: Another approach is robust optimization [54], which involves defining an uncertainty set of the covariance matrix and optimizing for the worst-case scenario. In particular, this method assumes the uncertainty set to be a sphere $\mathcal{U}_\mathbf{X} = \{\mathbf{X} | \|\mathbf{X} - \hat{\mathbf{X}}\|_F \leq \delta\}$ when dealing with a noisy data matrix $\hat{\mathbf{X}}$. As described in [55], [56], the worst-case MSRP can be formulated as

$$\begin{aligned}\underset{\mathbf{w}\in\mathcal{W}}{\text{minimize}} &- \frac{\mathbf{w}^\mathsf{T}\boldsymbol{\mu}}{\underset{\mathbf{X}\in\mathcal{U}_\mathbf{X}}{\max}\sqrt{\mathbf{w}^\mathsf{T}\frac{1}{T}\mathbf{X}^\mathsf{T}\mathbf{X}\mathbf{w}}} \\ &= -\frac{\mathbf{w}^\mathsf{T}\boldsymbol{\mu}}{\sqrt{\mathbf{w}^\mathsf{T}\hat{\boldsymbol{\Sigma}}\mathbf{w}} + \frac{\delta}{\sqrt{T}}\|\mathbf{w}\|_2}.\end{aligned} \tag{46}$$

Our proposed algorithm can solve this problem by setting $\boldsymbol{\Sigma}_1 = \hat{\boldsymbol{\Sigma}}$ (sample covariance matrix) and $\boldsymbol{\Sigma}_2 = \mathbf{I}$ in (3), (4), and Problem $\mathcal{P}$.

3) Weight Constraints: The third approach aims to solve Problem (45) based on sample covariance matrix but with additional constraints on the portfolio weights. Specifically, researchers have proposed using ℓ_0-norm, ℓ_1-norm, and ℓ_2-norm constraints/regularization [57], [58]. We focus on the

[4] Shrinkage methods can also be applied to portfolio weights [52], [53].



(a)

(b)

Fig. 8. Out-of-sample Sharpe ratio of different methods. (a) Number of assets $N = 50$. (b) Number of assets $N = 100$.

ℓ_0-norm regularized MSRP, which involves a regularization term $\lambda\|\mathbf{w}\|_0$ in the objective of Problem (45). Our proposed algorithm can solve this problem in a reweighted ℓ_1-norm minimization scheme [59]. Additionally, we investigate the ℓ_2-norm constrained MSRP, which includes an additional constraint $\|\mathbf{w}\|_2 \leq \varepsilon$ in Problem (45).

We conduct portfolio backtesting on real financial data to evaluate the out-of-sample performance of different methods for robustification. These methods are denoted as:

- SCM: Problem (45) based on the sample covariance matrix;
- LSM: Problem (45) based on the linear shrinkage covariance matrix [25];
- NLSM: Problem (45) based on the nonlinear shrinkage covariance matrix [26];
- Robust: Worst-case robust MSRP (46) with $\delta = 1$.
- ℓ_0-norm: ℓ_0-norm regularized MSRP with $\lambda = 10^{-4}$.
- ℓ_2-norm: ℓ_2-norm constrained MSRP with $\varepsilon = 0.25$.

We use a dataset of 15-year historical daily price data of stocks in the S & P 500, spanning from 2008-01-01 to 2022-12-31. To ensure the reliability of our findings, we randomly select 100 datasets, each containing the price of $N = 50$ (also $N = 100$) stocks over 500 continuous trading days. We then conduct rolling-window backtesting on each dataset using the R package portfolioBacktest [60] on the same platform as described in Section VI.

Fig. 8 shows the out-of-sample Sharpe ratio of various methods for robustification, with Fig. 8(a) displaying the results for $N = 50$ and Fig. 8(b) for $N = 100$. We observe that all methods perform relatively worse when $T \leq N$, indicating the presence of significant parameter uncertainty.

Our findings suggests that LSM, Robust, and the ℓ_2-norm method consistently outperform SCM. However, the ℓ_0-norm method, which produces sparser portfolios, may perform worse than SCM because it is more sensitive to out-of-sample observations. Additionally, NLSM exhibits worse performance than LSM when $T \leq N$, which aligns with the results reported in [26]. From the perspective of ease of use, shrinkage estimators are more convenient than other methods as they do not require tuning of hyperparameters.[5]

To clarify, the purpose of the above comparison is not to provide a comprehensive analysis of various robust variations but rather to demonstrate how our proposed algorithm can support different formulations. It is worth noting that each robust variation needs to be solved approximately $13,000$ times per experiment, which can be a significant computational challenge for existing solvers. However, our proposed algorithm can complete each experiment within an hour, demonstrating its practical value in portfolio design and evaluation.

VIII. Conclusion

In this paper, we proposed and analyzed a successive QP algorithm for general MVP optimization. The main advantage is that no matter what problem structures are contained in different MVPs, the proposed algorithm only requires solving a sequence of QP surrogate problems which already have well-developed efficient solvers. In addition, by exploiting the underlying sparsity pattern of this algorithm, we proposed the fast implementation that can further reduce the computational cost. The theoretical convergence analysis has been established, and comprehensive experiments reveal that the proposed algorithm has a higher convergence speed and better scalability than the state-of-the-art methods.

Appendix

A. Proof for Proposition 4

We begin with the following technical lemma.

Lemma 2: For any two adjacent iterations in Algorithm 2, we have $(\bar{L}^{k+1} \cup \bar{U}^{k+1}) \subset (\bar{L}^k \cup \bar{U}^k)$ together with $q(\mathbf{w}^{k+1}) < q(\mathbf{w}^k)$.

Proof: $\mathbf{w}^k$ is the solution of Problem (38) with the KKT condition

$$\beta_i^l \geq 0, \quad \beta_i^u \geq 0, \quad i \notin \bar{L}^k \cup \bar{U}^k, \tag{47}$$

so the “violated” assets with $\beta_i^l < 0$ or $\beta_i^u < 0$ must belong to the working set. Relaxing the strict equality bound constraints on these variables results in a smaller working set, i.e., $(\bar{L}^{k+1} \cup \bar{U}^{k+1}) \subset (\bar{L}^k \cup \bar{U}^k)$. Thus, we have

[5]The implementation of shrinkage methods is detailed in [25, Theorem 2.1] and [26, Appendix E].

$$\left\{\mathbf{w}\left|\begin{array}{ll}\mathbf{A}\mathbf{w}=\mathbf{b}, & l_i \le w_i \le u_i, \quad i \notin \bar{L}^k \cup \bar{U}^k \\ w_i = l_i, & i \in \bar{L}^k, \quad w_i = u_i, \quad i \in \bar{U}^k\end{array}\right.\right\}$$
$$\subset \left\{\mathbf{w}\left|\begin{array}{ll}\mathbf{A}\mathbf{w}=\mathbf{b}, & l_i \le w_i \le u_i, \quad i \notin \bar{L}^{k+1} \cup \bar{U}^{k+1} \\ w_i = l_i, & i \in \bar{L}^{k+1}, \quad w_i = u_i, \quad i \in \bar{U}^{k+1}\end{array}\right.\right\}, \quad (48)$$

which means $\mathbf{w}^k$ is also a feasible point of the subproblem (38) in the $(k+1)$th iteration. Also, $\mathbf{w}^k \neq \mathbf{w}^{k+1}$, or we would not have "violated" assets. Therefore, each iteration must decrease the objective function, i.e., $q(\mathbf{w}^{k+1}) < q(\mathbf{w}^k)$. ■

We are now ready to prove Proposition 4. First of all, the proposed algorithm is convergent. This is guaranteed by the monotone decrease of the size of the working set, according to Lemma 5. As the minimum size of the working set is bounded by zero (i.e., optimality achieved), the algorithm converges to the optimal solution within finite steps. At each iteration, at least one index is dropped from the working set. Hence, the number of iterations is bounded by $|\bar{L}^0 \cup \bar{U}^0|$.

B. MM Algorithm for Kelly Portfolio

To solve Problem (42), we first apply the quadratic transform on the ratio term and reformulates it as

$$\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \left(-\log\left(1+\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right) + a_1\sqrt{\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}} - \frac{a_1^2}{2}\left(1+\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right)^2\right), \quad (49)$$

where a_1 is an auxiliary variable iteratively updated by

$$a_1 = \sqrt{\left(\mathbf{w}^k\right)^\mathsf{T} \boldsymbol{\Sigma}\mathbf{w}^k} \Big/ \left(1+\boldsymbol{\mu}^\mathsf{T}\mathbf{w}^k\right)^2. \quad (50)$$

As shown by [61], (49) is an upper-bound problem of the primal problem (42). Further, we apply the first-order Taylor expansion to the nonconvex quadratic term of the objective function in (49). It constructs a convex upper-bound problem of (49) given by

$$\underset{\mathbf{w}\in\mathcal{W}}{\mathsf{minimize}} \left(-\log\left(1+\mathbf{w}^\mathsf{T}\boldsymbol{\mu}\right) + a_1\sqrt{\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}} - a_1^2\left(\boldsymbol{\mu}+\mathbf{a}_2\right)^\mathsf{T}\mathbf{w} + c\right), \quad (51)$$

where c is a constant, and $\mathbf{a}_2$ is another auxiliary variable computed as

$$\mathbf{a}_2 = \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}\mathbf{w}^k. \quad (52)$$

Overall, this algorithm solves Problem (42) by solving a sequence of upper-bound problems (51), and thus it can be interpreted as an MM algorithm.

References

[1] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] H. Levy and M. Levy, "Prospect theory and mean-variance analysis," *Rev. Financial Stud.*, vol. 17, no. 4, pp. 1015–1041, 2004.
[3] W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk," *J. Finance*, vol. 19, no. 3, pp. 425–442, 1964.
[4] J. Lintner, "Security prices, risk, and maximal gains from diversification," *J. Finance*, vol. 20, no. 4, pp. 587–615, 1965.
[5] F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi, *Robust Portfolio Optimization and Management*. Hoboken, NJ, USA: Wiley, 2007.
[6] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. New Haven, CT, USA: Yale Univ. Press, 1959.
[7] W. F. Sharpe, "Mutual fund performance," *J. Bus.*, vol. 39, no. 1, pp. 119–138, 1966.
[8] R. C. Merton, "An analytic derivation of the efficient portfolio frontier," *J. Financial Quantitative Anal.*, vol. 7, no. 4, pp. 1851–1872, 1972.
[9] J. G. Kallberg and W. T. Ziemba, "Comparison of alternative utility functions in portfolio selection problems," *Manage. Sci.*, vol. 29, no. 11, pp. 1257–1276, 1983.
[10] H. Levy and H. M. Markowitz, "Approximating expected utility by a function of mean and variance," *Amer. Econ. Rev.*, vol. 69, no. 3, pp. 308–317, 1979.
[11] H. Levy, *The Capital Asset Pricing Model in the 21st Century: Analytical, Empirical, and Behavioral Perspectives*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[12] P. N. Kolm, R. Tütüncü, and F. J. Fabozzi, "60 years of portfolio optimization: Practical challenges and current trends," *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 356–371, 2014.
[13] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, 1967.
[14] K. Shen and W. Yu, "Fractional programming for communication systems—Part I: Power control and beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[15] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Found. Trends Commun. Inf. Theory*, vol. 11, nos. 3–4, pp. 185–396, 2015.
[16] 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. 2017.
[17] G. Scutari and Y. Sun, "Parallel and distributed successive convex approximation methods for big-data optimization," in *Multi-Agent Optimization*. New York, NY, USA: Springer International Publishing, 2018, pp. 141–308.
[18] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.
[19] Y. Yang, G. Scutari, D. P. Palomar, and M. Pesavento, "A parallel decomposition method for nonconvex stochastic multi-agent optimization problems," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2949–2964, Jun. 2016.
[20] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, 1999.
[21] C. T. Lawrence and A. L. Tits, "A computationally efficient feasible sequential quadratic programming algorithm," *SIAM J. Optim.*, vol. 11, no. 4, pp. 1092–1118, 2001.
[22] 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.
[23] Y. Feng and D. P. Palomar, "SCRIP: Successive convex optimization methods for risk parity portfolio design," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5285–5300, Oct. 2015.
[24] S. Sivanandam and S. Deepa, "Genetic algorithms," in *Introduction to Genetic Algorithms*. Berlin, Germany: Springer, 2008, pp. 15–37.
[25] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.
[26] O. Ledoit and M. Wolf, "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *Ann. Statist.*, vol. 48, no. 5, pp. 3043–3065, 2020.
[27] D. Johnstone and D. Lindley, "Mean–variance and expected utility: The Borch paradox," *Statistical Sci.*, vol. 28, no. 2, pp. 223–237, 2013.
[28] Z. Landsman, U. Makov, and T. Shushi, "A generalized measure for the optimal portfolio selection problem and its explicit solution," *Risks*, vol. 6, no. 1, p. 19, 2018.
[29] M. S. Lobo and S. Boyd, "The worst-case risk of a portfolio," *Tech. Rep.*, Stanford University, USA, 2000. Accessed: Jul. 1, 2023. [Onlilne]. Available: http://stanford.edu/~boyd/papers/pdf/risk_bnd.pdf.
[30] H. Markowitz, "Mean–variance approximations to expected utility," *Eur. J. Oper. Res.*, vol. 234, no. 2, pp. 346–355, 2014.
[31] J. L. Kelly Jr., "A new interpretation of information rate," *Bell Syst. Tech. J.*, vol. 35, no. 4, pp. 917–926, Jul. 1956.
[32] E. O. Thorp, "Portfolio choice and the Kelly criterion," in *Stochastic Optimization Models in Finance*. Amsterdam, The Netherlands: Elsevier, 1975, pp. 599–619.

[33] L. B. Pulley, "Mean-variance approximations to expected logarithmic utility," *Oper. Res.*, vol. 31, no. 4, pp. 685–696, 1983.
[34] 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.
[35] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.
[36] D. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.
[37] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: New insights," *Structural Multidisciplinary Optim.*, vol. 41, no. 6, pp. 853–862, 2010.
[38] K. Miettinen, *Nonlinear Multiobjective Optimization*. Boston, MA, USA: Kluwer, 1999.
[39] R. Zhou and D. P. Palomar, "Solving high-order portfolios via successive convex approximation algorithms," *IEEE Trans. Signal Process.*, vol. 69, pp. 892–904, 2021.
[40] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[41] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, "Sparse and stable Markowitz portfolios," *Proc. Nat. Acad. Sci.*, vol. 106, no. 30, pp. 12267–12272, 2009.
[42] Y. Tian, X. Zhang, C. Wang, and Y. Jin, "An evolutionary algorithm for large-scale sparse multiobjective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 380–393, Apr. 2020.
[43] I. Kondor, G. Papp, and F. Caccioli, "Analytic solution to variance optimization with no short positions," *J. Statistical Mechanics, Theory Exp.*, vol. 2017, no. 12, 2017, Art. no. 123402.
[44] Y. Ye and E. Tse, "An extension of Karmarkar's projective algorithm for convex quadratic programming," *Math. Program.*, vol. 44, no. 1, pp. 157–179, 1989.
[45] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 4, pp. 4412–4419.
[46] A. Domahidi, E. Chu, and S. Boyd, "ECOS: An SOCP solver for embedded systems," in *Proc. Eur. Contr. Conf. (ECC)*, Piscataway, NJ, USA: IEEE Press, 2013, pp. 3071–3076.
[47] E. D. Andersen and K. D. Andersen, "The MOSEK interior point optimizer for linear programming: An implementation of the homogeneous algorithm," in *High Performance Optimization*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000, pp. 197–232.
[48] S. G. Johnson. "The NLopt nonlinear-optimization package." GitHub. Accessed: Jul. 1, 2023. [Online]. Available: http://github.com/stevengj/nlopt
[49] K. Mullen, D. Ardia, D. L. Gil, D. Windover, and J. Cline, "DEoptim: An R package for global optimization by differential evolution," *J. Statistical Softw.*, vol. 40, no. 6, pp. 1–26, 2011.
[50] L. Scrucca, "GA: A package for genetic algorithms in R," *J. Statistical Softw.*, vol. 53, no. 4, pp. 1–37, 2013.
[51] D. Goldfarb and A. Idnani, "A numerically stable dual method for solving strictly convex quadratic programs," *Math. Program.*, vol. 27, no. 1, pp. 1–33, 1983.
[52] T. Bodnar, N. Parolya, and W. Schmid, "Estimation of the global minimum variance portfolio in high dimensions," *Eur. J. Oper. Res.*, vol. 266, no. 1, pp. 371–390, 2018.
[53] T. Bodnar, Y. Okhrin, and N. Parolya, "Optimal shrinkage-based portfolio selection in high dimensions," *J. Bus. Econ. Statist.*, vol. 41, no. 1, pp. 140–156, 2022.
[54] D. Goldfarb and G. Iyengar, "Robust portfolio selection problems," *Math. Oper. Res.*, vol. 28, no. 1, pp. 1–38, 2003.
[55] S. A. Vorobyov, A. B. Gershman, Z.-Q. Luo, and N. Ma, "Adaptive beamforming with joint robustness against mismatched signal steering vector and interference nonstationarity," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 108–111, Feb. 2004.
[56] Y. Feng et al., "A signal processing perspective on financial engineering," *Found. Trends® Signal Process.*, vol. 9, nos. 1–2, pp. 1–231, 2016.
[57] B. Fastrich, S. Paterlini, and P. Winker, "Constructing optimal sparse portfolios using regularization methods," *Comput. Manage. Sci.*, vol. 12, pp. 417–434, 2015.
[58] V. DeMiguel, L. Garlappi, F. J. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Manage. Sci.*, vol. 55, no. 5, pp. 798–812, 2009.
[59] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.
[60] R. Zhou and D. P. Palomar, "Portfoliobacktest: Automated backtesting of portfolios over multiple datasets," R Package Version 0.3.1, 2021. [Online]. Available: https://cran.rstudio.com/web/packages/portfolioBacktest/index.html
[61] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach," *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 2164–2177, 2019.

Shengjie Xiu received the bachelor's degree in electronic engineering from Sun Yat-sen University in Guangzhou, China, in 2020. He is currently pursuing the Ph.D. degree in electronic and computer engineering with the Hong Kong University of Science and Technology. His research interests include intraday signal modeling for financial assets and efficient algorithms for portfolio optimization and algorithmic trading.

Xiwen Wang received the B.Sc. degree in electronic information of science and technology from Nanjing University, Nanjing, China, in June 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic and Computer Engineering with the Hong Kong University of Science and Technology. His research interests include convex optimization and fast algorithms with applications in financial engineering, machine learning, and operational research.

Daniel P. Palomar (Fellow, IEEE) 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 with the Department of Electronic & Computer Engineering and the Department of Industrial Engineering & Decision Analytics 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 optimization theory, graph methods, and signal processing in financial systems and big data analytics. He was a recipient of a 2004/06 Fulbright Research Fellowship, the 2004, 2015, and 2020 (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 the 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."