

PORTFOLIO OPTIMIZATION WITH ASSET SELECTION AND RISK PARITY CONTROL

Yiyong Feng and Daniel P. Palomar

Dept. of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

After the 2008 financial crisis, risk management has become more important than performance management and an alternative portfolio design, referred to as risk parity portfolio, has been receiving significant attention from both theoretical and practical fields due to its advantage in diversification of (ex-ante) risk contributions among assets. Usually, this approach results in a portfolio with nonzero weights in all the assets. Investors, however, could not lay out the capital among all the assets listed on the markets, which results in unrealistically high transaction costs, and therefore, reduction of the return of the designed portfolio. To overcome this drawback, in this paper, we propose a method to jointly select only some of the assets and distribute the capital among the selected assets such that the risk is diversified enough.

Index Terms—Asset Selection, Portfolio Optimization, Risk Parity, Successive Convex Approximation.

1. INTRODUCTION

For over fifty years, the mean-variance portfolio optimization framework [1–3], i.e., minimizing the mean-variance trade-off of a portfolio, has been well-researched in the academic field. However, this framework tends to output a portfolio with risk excessively concentrated over a few assets, which goes against the common sense of diversification as a way to reduce the risk. The portfolio that diversifies the capital via minimizing the mean-variance trade-off does not necessarily diversify risk [4]. Serious issues, which might not occur during normal times, would happen if a financial crisis were to happen, because such a concentrated portfolio would probably incur huge losses.

Around 2005, Qian [5, 6] first showed that equal risk contributions (ERC) actually lead to a diverse enough portfolio, and the (ex-ante) risk contributions (RCs) (i.e., the risks computed using historical data) are not only a mathematical measurement, but also good indicators of the (ex-post) loss contributions of the assets (i.e., the observed risks and losses in the future), especially when there exist large losses. According to this observation, the way to avoid a potential huge loss is to distribute the RCs. However, the risk parity portfolio did not attract too much attention before the 2008 financial crisis until Maillard et al. [7] first analyzed the properties of the ERC portfolio and showed that it is a trade-off between the minimum variance (MV) and equal weight (EW) portfolios. Following that, there are more works on different formulations [8–12] or numerical methods for computing the risk parity portfolio [13–16]. Meanwhile, the risk parity portfolio has found its application in various applications, e.g., risk-based indexation, alternative assets management, portfolio with multi-asset classes, etc., and the recent book [4] serves as a good summary on both the theoretical foundations and various applications of the risk parity portfolio.

Another practical issue in portfolio design is the transaction cost. The transaction cost in general is assumed to be concave [17], e.g., Lobo et al. [18] considered a constant plus linear cost function. Thus, to reduce the transaction cost, it is beneficial to invest only part of

all the assets and then invest a significant amount of money in each selected asset. The first idea was to regularize (or to constrain) the ℓ_1 -norm of the portfolio weights since it is well-known that the ℓ_1 -norm is convex and can promote sparsity at the same time [19, 20]. And later people also considered some nonconvex (e.g., ℓ_0 -norm, ℓ_q -norm with $0 < q < 1$, logarithmic penalty, etc.) regularizations (or constraints), see [21, 22].

Unfortunately, the risk parity conditions always result in a portfolio with nonzero weights in all the assets [4, 7], which implies considerable transaction costs and may attenuate the portfolio performance significantly.

To the best of our knowledge, the combination of selecting assets and risk diversification has not been investigated in portfolio design. This motivates us to consider them jointly for the portfolio designs and the main contributions are i) we first propose a problem formulation allowing us to jointly select the assets and diversify the risk for portfolio optimization, which is a highly nonconvex problem, and ii) we further develop an efficient iterative solving algorithm dealing with the proposed problem.

2. RISK PARITY PORTFOLIO BACKGROUND

2.1. Risk contribution

Suppose there are n assets with random returns $\mathbf{r} \in \mathbb{R}^n$, and the mean vector and (positive definite) covariance matrix are denoted as $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$. We use $\mathbf{w} \in \mathbb{R}^n$ to denote the normalized portfolio (e.g., $\mathbf{w}^T \mathbf{1} = 1$), which describes how the total capital budget is to be allocated over the assets. To study the risk parity portfolio, we need to properly define a quantity which can represent the “RC” of each asset to the whole portfolio.

In this paper, we focus on the portfolio volatility¹, i.e., $\sigma(\mathbf{w}) \triangleq \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$, as the risk measurement. Since the portfolio volatility can always be decomposed as follows:

$$\sum_{i=1}^n w_i \frac{\partial \sigma(\mathbf{w})}{\partial w_i} = \sum_{i=1}^n w_i \frac{(\boldsymbol{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} = \sigma(\mathbf{w}), \quad (1)$$

then each term $w_i \frac{\partial \sigma}{\partial w_i}$ can be regarded as the RC of the i -th asset since the summation of all the terms is the total risk.

2.2. Risk parity portfolio

A risk parity portfolio is a portfolio such that each asset has the same RC. That, given the portfolio volatility $\sigma(\mathbf{w})$, the risk parity portfolio should satisfy

$$w_i \frac{(\boldsymbol{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} = w_j \frac{(\boldsymbol{\Sigma} \mathbf{w})_j}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}, \quad \forall i, j. \quad (2)$$

Based on (1), the relationship (2) can be rewritten as

$$w_i \frac{\partial \sigma(\mathbf{w})}{\partial w_i} = \frac{1}{n} \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}, \quad \forall i. \quad (3)$$

¹The RCs are also well defined for other risk measures, e.g., Value-at-Risk (VaR) and Conditional VaR, but not variance. For further information, see [4] and references therein.

This work was supported by the Hong Kong RGC 16207814 research grant.

Furthermore, multiplied by $\sigma(\mathbf{w})$, the relationships (2) and (3) can be simplified, respectively, as follows

$$w_i(\Sigma \mathbf{w})_i = w_j(\Sigma \mathbf{w})_j, \quad (4)$$

$$w_i(\Sigma \mathbf{w})_i = \frac{1}{n} \mathbf{w}^T \Sigma \mathbf{w}. \quad (5)$$

Given the above relationships (2)-(5), we can have various functions, denoted as $g_i(\mathbf{w})$ (e.g., $g_i(\mathbf{w}) \triangleq w_i(\Sigma \mathbf{w})_i$, or $\frac{w_i(\Sigma \mathbf{w})_i}{\sqrt{\mathbf{w}^T \Sigma \mathbf{w}}}$, or $\frac{w_i(\Sigma \mathbf{w})_i}{\mathbf{w}^T \Sigma \mathbf{w}}$), to represent the RCs. More such functions can be found in [16, Table I].

3. PROBLEM FORMULATION

In this section, we will propose a portfolio optimization formulation considering asset selection and risk parity jointly and highlight the challenges of solving it.

3.1. Portfolio optimization with asset selection and risk parity control

Now we propose the following portfolio optimization problem formulation with asset selection and risk parity control:

$$\begin{aligned} & \underset{\mathbf{w}, \theta}{\text{minimize}} && U(\mathbf{w}) \triangleq F(\mathbf{w}) + \lambda_1 \|\mathbf{w}\|_0 + \lambda_2 R(\mathbf{w}, \theta) \\ & \text{subject to} && \mathbf{w}^T \mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (6)$$

where

- $F(\mathbf{w}) \triangleq -\nu \mathbf{w}^T \boldsymbol{\mu} + \mathbf{w}^T \Sigma \mathbf{w}$ is the mean variance trade-off objective with $\nu \geq 0$ the trade-off parameter.
- $\|\mathbf{w}\|_0 \triangleq \sum_{i=1}^n \mathbb{1}_{\{w_i \neq 0\}}$ is the ℓ_0 -norm that regularizes the cardinality of the portfolio weights.
- $R(\mathbf{w}, \theta)$ measures the risk concentration and has the form

$$R(\mathbf{w}, \theta) \triangleq \sum_{i=1}^n (g_i(\mathbf{w}) - \theta)^2 \mathbb{1}_{\{w_i \neq 0\}} \quad (7)$$

in which $g_i(\mathbf{w})$ is any smooth nonconvex differentiable function that measures the RC of the i -th asset of the portfolio and θ is a (scalar) variable denoting average RCs of the selected assets (i.e., for $w_i \neq 0$). The smaller the quantity $R(\mathbf{w}, \theta)$ is, the more uniform the risk is distributed among the selected assets.

- $\lambda_1, \lambda_2 \geq 0$ are the regularization trade-off parameters.
- $\mathbf{w}^T \mathbf{1} = 1$ denotes the capital budget constraint.
- \mathcal{W} is a convex set that denotes the market regulations and investor's profiles, e.g., turnover constraints, holding constraints, tracking error constraints, etc.

Observing the objective of (6), we can see that the first regularization $\|\mathbf{w}\|_0$ regularizes the number of stocks and promotes sparsity, and the second regularization $R(\mathbf{w}, \theta)$ tends to distribute the risk among the selected assets. If $\lambda_1 > 0$, denote $\alpha \triangleq \lambda_2 / \lambda_1 \geq 0$ and the objective of (6) can be rewritten as

$$\begin{aligned} U(\mathbf{w}) &= F(\mathbf{w}) + \lambda_1 \sum_{i=1}^n \mathbb{1}_{\{w_i \neq 0\}} + \lambda_2 \sum_{i=1}^n (g_i(\mathbf{w}) - \theta)^2 \mathbb{1}_{\{w_i \neq 0\}} \\ &= F(\mathbf{w}) + \lambda_1 \sum_{i=1}^n (1 + \alpha (g_i(\mathbf{w}) - \theta)^2) \mathbb{1}_{\{w_i \neq 0\}}. \end{aligned} \quad (8)$$

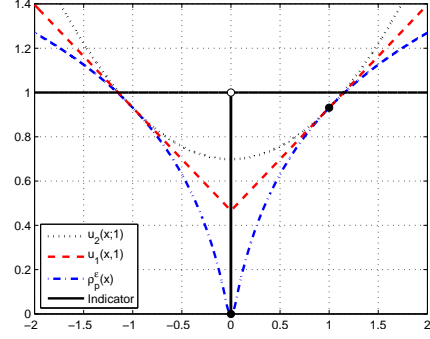


Fig. 1. The indicator function, the smooth approximation $\rho_p^\epsilon(x)$ with $p = 0.2$ and $\epsilon = 0.05$, and the first-order linear and second-order quadratic approximation functions $u_1(x; x^0)$ and $u_2(x; x^0)$ at the approximation point $x^0 = 1$.

The second term in (8) is a weighted ℓ_0 -norm of \mathbf{w} with the weights $1 + \alpha (g_i(\mathbf{w}) - \theta)^2$ representing the deviation of the RC of the i -th asset from the average level θ . Thus, the larger the risk deviation is, the heavier weight is put to the cardinality penalty. This makes sense since we want to jointly select the assets and achieve risk diversification among them.

3.2. Challenges

Since each function $g_i(\mathbf{w})$ is highly nonconvex and the indicator function $\mathbb{1}_{\{w_i \neq 0\}}$ is highly nonconvex, nondifferentiable, and discontinuous, clearly $R(\mathbf{w}, \theta)$ is nonconvex and nondifferentiable, and the problem (6) is hard to deal with.

4. SOLVING APPROACH

In the following, we deal with the problem of interest (6). Note that there are two main difficulties in the objective function: the nonconvexity and discontinuity (caused by the nondifferentiable indicator function). To deal with these main difficulties, in this section, we first approximate problem (6) with a (nonconvex but) differentiable problem and then derive a successive convex approximation (SCA) based solving approach for the approximation problem. The proposed iterative algorithm is guaranteed to converge to a stationary point of the approximation problem.

4.1. Smooth approximation problem

Following [23], we approximate the indicator function by the following smooth approximation:

$$\rho_p^\epsilon(x) = \begin{cases} \frac{x^2}{2\epsilon(p+\epsilon) \log(1+1/p)}, & |x| \leq \epsilon \\ \frac{\log(1+|x|/p) - \log(1+\epsilon/p) + \frac{\epsilon}{2(p+\epsilon)}}{\log(1+1/p)}, & |x| > \epsilon \end{cases} \quad (9)$$

where $p > 0$ and $\epsilon > 0$ are controlling parameters, and when they both goes to zero, $\rho_p^\epsilon(x)$ converges to the indicator function. Fig. 1 shows the approximation when $p = 0.2$ and $\epsilon = 0.05$. The smaller ϵ is, the better the indicator function is approximated around the point $x = 0$. Indeed, when ϵ converges to zero, $\rho_p^\epsilon(x)$ converges to the (nondifferentiable) function $\frac{\log(1+|x|/p)}{\log(1+1/p)}$ which coincides with the approximation in [24, 25]. In practice, we can set ϵ to be very small, e.g., $\epsilon = 10^{-8}$, to achieve satisfactory approximation and avoid the numerical nondifferentiability at $x = 0$.

Then moving the indicator function inside the square of $R(\mathbf{w}, \theta)$ (so that later we can construct a convex quadratic approximation

more easily) and replacing each indicator function $\mathbb{1}_{\{w_i \neq 0\}}$ in problem (6) with the approximation $\rho_p^\epsilon(w_i)$ yield the following approximation problem:

$$\begin{aligned} \underset{\mathbf{w}, \theta}{\text{minimize}} \quad & \tilde{U}(\mathbf{w}) \triangleq F(\mathbf{w}) + \lambda_1 \sum_{i=1}^n \rho_p^\epsilon(w_i) \\ & + \lambda_2 \underbrace{\sum_{i=1}^n ((g_i(\mathbf{w}) - \theta) \rho_p^\epsilon(w_i))^2}_{\tilde{R}(\mathbf{w}, \theta) \triangleq} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (10)$$

where $\tilde{U}(\mathbf{w})$ is a continuous and differentiable approximation of $U(\mathbf{w})$, however, it is still nonconvex. In the following, we mainly focus on solving the approximation problem (10) instead and develop fast iterative numerical algorithms based via SCA.

For technical reasons, we make the following assumptions:

- (A1) $\overline{\mathcal{W}}_1 \triangleq \{\mathbf{w}^T \mathbf{1} = 1\} \cap \mathcal{W}$ is nonempty, closed, and convex;
- (A2) \tilde{R} and each g_i are C^1 on an open set containing $\overline{\mathcal{W}}_1$;
- (A3) $\nabla \tilde{R}$ is Lipschitz continuous on $\overline{\mathcal{W}}_1$ with constant L_R ;
- (A4) $F(\mathbf{w})$ is continuous and convex on $\overline{\mathcal{W}}_1$;
- (A5) $U(\mathbf{w})$ is coercive with respect to $\overline{\mathcal{W}}_1$.

Note that the above assumptions are standard and are satisfied by a large class of functions. For instance, A3 is satisfied automatically if $\overline{\mathcal{W}}_1$ is bounded, and A4 is satisfied by all the standard F used in portfolio design. Assumption A5 guarantees that the sequence generated by the solving approach later is bounded, and if $\overline{\mathcal{W}}_1$ is bounded A5 is trivially satisfied. For the portfolio design in the real markets, the feasible set will always be bounded due to some practical constraints, e.g., turnover constraints, holding constraints, tracking error constraints, etc., [26], and the above assumptions are easily satisfied.

4.2. Solving approach via SCA

The idea of SCA is to approximate the original (possibly nonconvex) function at each iteration point by a solvable convex problem to get an update. Suppose now we are at the k -th iteration point (\mathbf{w}^k, θ^k) , the updates of θ and \mathbf{w} based on SCA are designed as follows.

4.2.1. Updating θ

When \mathbf{w} is fixed to \mathbf{w}^k , minimizing problem (10) with respect to θ equals to the following unconstrained scalar minimization problem

$$\underset{\theta}{\text{minimize}} \quad \sum_{i=1}^n ((g_i(\mathbf{w}^k) - \theta) \rho_p^\epsilon(w_i^k))^2. \quad (11)$$

This is a univariate weighted least-square problem, which is strongly convex, and setting the derivative of the objective to zero directly yields the optimal solution in closed-form, which is given by the steps 2-3 in Alg. 1.

4.2.2. Updating \mathbf{w}

To update \mathbf{w} , we need to first construct a local convex approximation of the objective of the problem (10) w.r.t. \mathbf{w} at the point \mathbf{w}^k . Let us consider it term by term. Since $F(\mathbf{w})$ is already convex, we keep it as it is and only consider the nonconvex terms.

Approximating $\sum_{i=1}^n \rho_p^\epsilon(w_i)$. We apply two different ideas here. *Method 1: First-order linear approximation.* The first idea is to follow iterative weighted ℓ_1 -norm approximation method [24, 25] to construct a piecewise linear approximation at the k -th point w_i^k :

$$\rho_p^\epsilon(w_i) \approx u_1(w_i; w_i^k) = d_1(w_i^k) |w_i| + c_{1i}^k \quad (12)$$

where $d_1(x) = \begin{cases} \frac{|x|}{\epsilon(p+\epsilon) \log(1+1/p)}, & |x| \leq \epsilon \\ \frac{1}{(|x|+p) \log(1+1/p)}, & |x| > \epsilon \end{cases}$ and c_{1i}^k is a properly chosen value so that the equality holds at w_i^k . Thus, $\sum_{i=1}^n \rho_p^\epsilon(w_i)$ can be approximated by a weighted ℓ_1 -norm function:

$$\sum_{i=1}^n \rho_p^\epsilon(w_i) \approx \|\mathbf{D}_1^k \mathbf{w}\|_1 + \sum_{i=1}^n c_{1i}^k \quad (13)$$

where $\mathbf{D}_1^k \triangleq \text{Diag}([d_1(w_1^k), d_1(w_2^k), \dots, d_1(w_n^k)])$ and $\sum_{i=1}^n c_{1i}^k$ is a constant given w^k and can be disregarded at each iteration.

Method 2: Second-order quadratic approximation. Following [23], the second idea is to approximate $\rho_p^\epsilon(w_i)$ by a quadratic convex function at the k -th point w_i^k as follows:

$$\rho_p^\epsilon(w_i) \approx u(w_i; w_i^k) = d_2(w_i^k) (w_i)^2 + c_{2i}^k \quad (14)$$

where $d_2(x) = \begin{cases} \frac{1}{2\epsilon(p+\epsilon) \log(1+1/p)}, & |x| \leq \epsilon \\ \frac{1}{2|x|(|x|+p) \log(1+1/p)}, & |x| > \epsilon \end{cases}$ is a weight function and c_{2i}^k is a properly chosen value so that the equality holds at w_i^k . Then we have that $\sum_{i=1}^n \rho_p^\epsilon(w_i)$ is approximated by a convex quadratic function

$$\sum_{i=1}^n \rho_p^\epsilon(w_i) \approx \|\mathbf{D}_2^k \mathbf{w}\|_2^2 + \sum_{i=1}^n c_{2i}^k \quad (15)$$

where $\mathbf{D}_2^k \triangleq \text{Diag}([\sqrt{d_2(w_1^k)}, \sqrt{d_2(w_2^k)}, \dots, \sqrt{d_2(w_n^k)}])$ and they share the same derivative w.r.t. \mathbf{w} at point \mathbf{w}^k and $\sum_{i=1}^n c_{2i}^k$ is a constant given w^k and can be disregarded at each iteration.

Fig. 1 shows an example how the first-order and second-order approximations $u_1(x; x^0)$ and $u_2(x; x^0)$ approximate the nonconvex function $\rho_p^\epsilon(x)$ at the point $x^0 = 1$. Generally speaking, the first-order approximation is that it approximates the indicator better, however, the second-order approximation may result simpler update expressions for some special case constraints (e.g., it results in closed-form update step for linear equality constraints).

Approximating $\tilde{R}(\mathbf{w}, \theta)$. Defining $\tilde{g}_i(\mathbf{w}, \theta) \triangleq (g_i(\mathbf{w}) - \theta) \rho_p^\epsilon(w_i)$, we have $\tilde{R}(\mathbf{w}, \theta) = \sum_{i=1}^n (\tilde{g}_i(\mathbf{w}, \theta))^2$. Following the idea of [16], we can linearize the functions \tilde{g}_i inside the square and get the following approximation:

$$P(\mathbf{w}, \theta) \triangleq \sum_{i=1}^n \left(\tilde{g}_i(\mathbf{w}^k, \theta) + (\nabla \tilde{g}_i(\mathbf{w}^k, \theta))^T (\mathbf{w} - \mathbf{w}^k) \right)^2$$

where

$\nabla \tilde{g}_i(\mathbf{w}, \theta) \triangleq \rho_p^\epsilon(w_i) \cdot \nabla g_i(\mathbf{w}) + ((g_i(\mathbf{w}) - \theta) \cdot \nabla \rho_p^\epsilon(w_i)) \cdot \mathbf{e}_i$ is the derivative of $\tilde{g}_i(\mathbf{w}, \theta)$ w.r.t. \mathbf{w} . Here the vector $\mathbf{e}_i \in \mathbb{R}^n$ denotes the column vector with only the i -th element being one and zero elsewhere. It is easy check that $\tilde{R}(\mathbf{w}, \theta)$ and $P(\mathbf{w}, \theta)$ have the same derivative w.r.t. \mathbf{w} at point \mathbf{w}^k , that is, $\nabla \tilde{R}(\mathbf{w}, \theta)|_{\mathbf{w}=\mathbf{w}^k} = \nabla P(\mathbf{w}, \theta)|_{\mathbf{w}=\mathbf{w}^k}$, where $\nabla \tilde{R}(\mathbf{w}, \theta)$ and $\nabla P(\mathbf{w}, \theta)$ denote the partial gradient of $\tilde{R}(\mathbf{w}, \theta)$ and $P(\mathbf{w}, \theta)$ w.r.t. to \mathbf{w} , respectively.

Updating problem. When θ is fixed to θ^k , replacing $\sum_{i=1}^n \rho_p^\epsilon(w_i)$ and $\tilde{R}(\mathbf{w}, \theta^k)$ with the right hand side of (13) or (15) and $P(\mathbf{w}, \theta^k)$, respectively, and removing the constant terms, the problem (10) can be approximated at \mathbf{w}^k by the following convex problem:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & F(\mathbf{w}) + \lambda_1 \|\mathbf{D}_o^k \mathbf{w}\|_o + \lambda_2 P(\mathbf{w}, \theta^k) + \tau \|\mathbf{w} - \mathbf{w}^k\|_2^2 \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{1} = 1, \quad \mathbf{w} \in \mathcal{W}, \end{aligned} \quad (16)$$

where the proximal term $\|\mathbf{w} - \mathbf{w}^k\|_2^2$ with $\tau > 0$ is added for convergence reasons, and $o = 1$ or $o = 2$ denotes the first-order or second-order approximation in (13) or (15), respectively.

Supposing $F(\mathbf{w})$ is convex, for nonempty convex set $\overline{\mathcal{W}}_1$ (recall that $\overline{\mathcal{W}}_1 = \{\mathbf{w}^T \mathbf{1} = 1\} \cap \mathcal{W}$) and $\tau > 0$, for either $o = 1$ or $o = 2$, the problem (16) is strongly convex and can be solved by the existing efficient solvers (e.g., MOSEK [27], CPLEX [28], etc).

4.2.3. Iterative algorithm and convergence

Alg. 1 summarizes the previous derived iterative solving procedure.

Algorithm 1 SCA for portfolio optimization under asset selection and risk parity control.

Input: $k = 0, \mathbf{w}^0 \in \overline{\mathcal{W}}, \theta^0 = \sum_{i=1}^n x_i^0 g_i(\mathbf{w}^0), \tau > 0, \{\gamma^k\} > 0$
Output: a stationary point of problem (10)

1: **repeat**

$$2: x_i^k = \frac{(\rho_p^e(w_i^k))^2}{\sum_{j=1}^n (\rho_p^e(w_j^k))^2}$$

$$3: \hat{\theta} = \sum_{i=1}^n x_i^k g_i(\mathbf{w}^k)$$

$$4: \theta^{k+1} = \theta^k + \gamma^k (\hat{\theta}^k - \theta^k)$$

5: Solve (16) to get the optimal solution $\hat{\mathbf{w}}^k$

$$6: \mathbf{w}^{k+1} = \mathbf{w}^k + \gamma^k (\hat{\mathbf{w}}^k - \mathbf{w}^k)$$

7: $k \leftarrow k + 1$

8: **until** convergence

Proposition 1. Under assumptions A1-A5, suppose $\tau > 0, \gamma^k \in (0, 1], \gamma^k \rightarrow 0, \sum_k \gamma^k = +\infty$ and $\sum_k (\gamma^k)^2 < +\infty$, and let $\{\mathbf{w}^k\}$ be the sequence generated by Alg. 1. Then either Alg. 1 converges in a finite number of iterations to a stationary point of (10) or every limit of $\{\mathbf{w}^k\}$ (at least one such point exists) is a stationary point of (10).

Proof. Under assumptions A1-A5 and given $\tau > 0$ and γ^k as above, since θ can be found in closed form for any given \mathbf{w} and it is easy to check that for any fixed θ the approximated problem (16) is a partial linear approximation of (10) with a quadratic uniformly strongly convex proximal term. That is, [29, Assumptions A1-A4] and [29, condition (b) in Theorem 3] are satisfied, and the proof of Prop. 1 follows directly from [29, Theorem 3]. \square

In Alg. 1, steps 4 and 6 are used to speed up the convergence numerically and one practical rule of choosing γ^k is: given $\gamma^0 \in (0, 1]$, let $\gamma^k = \gamma^{k-1} (1 - \zeta \gamma^{k-1})$, $k = 1, 2, \dots$ where $\zeta \in (0, 1)$ is a given constant [29, 30]. This rule in general enjoys really fast numerical convergence speed, e.g., see [16, 29–31].

5. NUMERICAL RESULTS

In this section, we conduct a simple synthetic example to show that indeed the proposed methods can efficiently select the assets and diversify the RCs jointly. For simplicity, let us consider $n = 10$ uncorrelated assets with increasing volatilities $\sigma_1 = 1\%, \sigma_2 = 2\%, \dots, \sigma_n = 10\%$, and the covariance matrix is $\Sigma = \text{Diag}([\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2])$. We consider risk performance portfolios: i) the EW portfolio, i.e., $w_i = 1/n$, ii) the ERC portfolio, i.e., $w_i(\Sigma \mathbf{w})_i = w_j(\Sigma \mathbf{w})_j$, which reduces to $w_i \propto 1/\sigma_i$, iii) the MV portfolio, $\mathbf{w} \propto \Sigma^{-1} \mathbf{1}$ which turns to be $w_i \propto 1/\sigma_i^2$, and iv) two proposed portfolio given by Alg. 1: one is using the first-order linear approximation, i.e., $o = 1$, where $\nu = 0, g_i(\mathbf{w}) = w_i(\Sigma \mathbf{w})_i, p = 0.002, \epsilon = 10^{-8}, \lambda_1 = 0.1$, and $\lambda_2 = 4$, the other is using the second-order quadratic

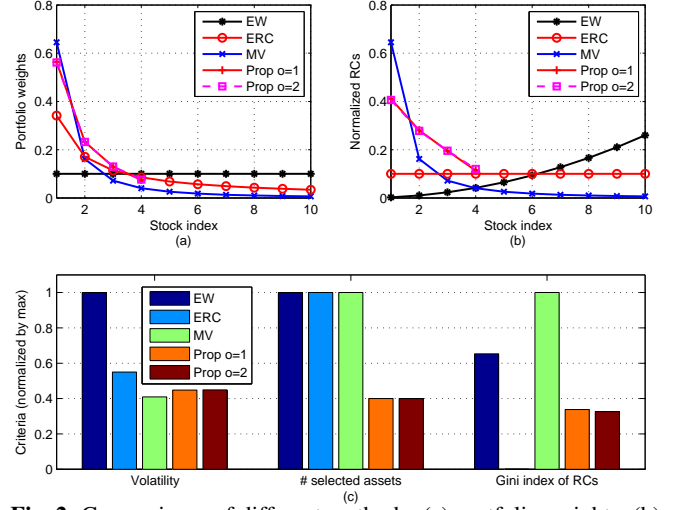


Fig. 2. Comparisons of different methods: (a) portfolio weights; (b) normalized of RCs, and (c) different portfolio criteria, i.e., portfolio volatility, number of selected stocks, and Gini index of the RCs.

approximation, i.e., $o = 2$, with all the parameters unchanged except that $\lambda_1 = 2^{-4}$. The covariance matrix is scaled up by 10^4 to avoid any numerical issues.

Intuitively, if all the volatilities are close to each other, all the three existing methods, i.e., the EW, ERC, and MV portfolios, tend to provide similar weights. Here, we expect they provide very different weights in this numerical example because the asset volatilities vary significantly from 1% to 10%. Fig. 2(a) shows the portfolio weights of all the methods and we can see that the MV portfolio concentrates most of its capital on the first asset with lowest volatility, the ERC portfolio concentrates less, and the EW portfolio is uniform. The weight concentration of the both proposed methods are between the MV and ERC portfolios, however, with only four out of ten assets selected (see the right panel of Fig. 2(c)).

Next, let us investigate the risk diversification. Fig. 2(b) shows the normalized asset RCs such that summation of the normalized RCs given by each method is one. We see the MV portfolio concentrates most of the risk on the first asset with lowest volatility. This makes sense since it intends to achieve as the minimum variance. The EW portfolio concentrates more on volatile assets, and the ERC portfolio distributes the RCs uniformly. Compared with MV and EW portfolios, the proposed methods are less risk concentrated while using fewer assets (see the right panel of Fig. 2(c)).

Fig. 2(c) summarizes three important criteria, i.e., portfolio volatility, number of selected stocks, and Gini index of the RCs (for this criterion see [4]), normalized by their maximum values, respectively. For all these three criteria, the smaller the better. Interestingly, the proposed methods with different (i.e., first-order or second-order) approximations perform virtually the same in every aspect and almost have the same volatility as the MV portfolio but select fewer assets and diversify the risk enough. If we incorporate some low correlations among the assets, we can have similar results as Fig. 2(c). Thus, we can conclude that the proposed methods can efficiently select the assets and diversify the risk.

6. CONCLUSION

In this paper, we have proposed a portfolio optimization formulation with asset selection and risk parity control and a simple and efficient sequential solving approach based on SCA. The numerical results show that the proposed formulation jointly achieves asset selection and risk diversification.

7. REFERENCES

- [1] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
- [2] —, "The optimization of a quadratic function subject to linear constraints," *Naval Research Logistics Quarterly*, vol. 3, no. 1-2, pp. 111–133, 1956.
- [3] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1968.
- [4] T. Roncalli, *Introduction to Risk Parity and Budgeting*. CRC Press, 2013.
- [5] E. Qian, "Risk parity portfolios: Efficient portfolios through true diversification," *Panagora Asset Management*, Sept., 2005.
- [6] —, "On the financial interpretation of risk contribution: Risk budgets do add up," *Journal of Investment Management*, vol. 4, no. 4, p. 41, 2006.
- [7] S. Maillard, T. Roncalli, and J. Teïletche, "The properties of equally weighted risk contribution portfolios," *Journal of Portfolio Management*, vol. 36, no. 4, pp. 60–70, 2010.
- [8] X. Bai, K. Scheinberg, and R. Tutuncu, "Least-squares approach to risk parity in portfolio selection," *Available at SSRN 2343406*, 2013.
- [9] T. Roncalli and G. Weisang, "Risk parity portfolios with risk factors," *Available at SSRN 2155159*, 2012.
- [10] B. Bruder and T. Roncalli, "Managing risk exposures using the risk budgeting approach," University Library of Munich, Germany, Tech. Rep., 2012.
- [11] K. Boudt, P. Carl, and B. G. Peterson, "Asset allocation with conditional Value-at-Risk budgets," *Journal of Risk*, vol. 15, no. 3, pp. 39–68, 2013.
- [12] M. Haugh, G. Iyengar, and I. Song, "A generalized risk budgeting approach to portfolio construction," *Available at SSRN 2462145*, 2014.
- [13] D. B. Chaves, J. C. Hsu, F. Li, and O. Shakernia, "Efficient algorithms for computing risk parity portfolio weights," *The Journal of Investing*, vol. 21, pp. 150–163, 2012.
- [14] F. Spinu, "An algorithm for computing risk parity weights," *Available at SSRN 2297383*, 2013.
- [15] T. Griveau-Billion, J.-C. Richard, and T. Roncalli, "A fast algorithm for computing high-dimensional risk parity portfolios," *arXiv preprint arXiv:1311.4057*, 2013.
- [16] Y. Feng and D. P. Palomar, "SCRIP: Successive convex optimization methods for risk parity portfolios design," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5285–5300, Oct. 2015.
- [17] H. Konno and A. Wiyayanayake, "Portfolio optimization problem under concave transaction costs and minimal transaction unit constraints," *Mathematical Programming*, vol. 89, no. 2, pp. 233–250, 2001.
- [18] M. S. Lobo, M. Fazel, and S. Boyd, "Portfolio optimization with linear and fixed transaction costs," *Annals of Operations Research*, vol. 152, no. 1, pp. 341–365, 2007.
- [19] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, 2015.
- [20] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, "Sparse and stable markowitz portfolios," *Proceedings of the National Academy of Sciences*, vol. 106, no. 30, pp. 12 267–12 272, 2009.
- [21] B. Fastrich, S. Paterlini, and P. Winker, "Cardinality versus q-norm constraints for index tracking," *Quantitative Finance*, no. ahead-of-print, pp. 1–14, 2012.
- [22] —, "Constructing optimal sparse portfolios using regularization methods," *Computational Management Science*, pp. 1–18, 2013.
- [23] J. Song, P. Babu, and D. P. Palomar, "Sparse generalized eigenvalue problem via smooth optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1627–1642, April 2015.
- [24] B. K. Sriperumbudur, D. A. Torres, and G. R. G. Lanckriet, "Sparse eigen methods by dc programming," in *Proceedings of the 24th International Conference on Machine Learning*. ACM, 2007, pp. 831–838.
- [25] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
- [26] F. Fabozzi, S. Focardi, and P. Kolm, *Quantitative Equity Investing: Techniques and Strategies*. Wiley, 2010.
- [27] MOSEK, "The MOSEK optimization toolbox for MATLAB manual," Tech. Rep., 2013. [Online]. Available: <http://www.mosek.com>
- [28] IBM ILOG CPLEX, "V12. 1: User's manual for CPLEX," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.
- [29] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641–656, Feb. 2014.
- [30] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, April 2015.
- [31] A. Daneshmand, F. Facchinei, V. Kungurtsev, and G. Scutari, "Hybrid random/deterministic parallel algorithms for convex and nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 3914–3929, Aug. 2015.