Vast Portfolio Selection With Submodular Norm Regularizations

Zepeng Zhang and Ziping Zhao
School of Information Science and Technology
ShanghaiTech University, Shanghai 201210, China
Email: zhangzp1@shanghaitech.edu.cn, zhaoziping@shanghaitech.edu.cn

Abstract—Portfolio selection (or portfolio optimization) has been a fundamental problem in the financial investment world since the modern portfolio theory (a.k.a. mean-variance analysis) was introduced by Harry Markowitz in 1952. The goal of portfolio selection is to assign different portions of dollars to the underlying assets according to a certain investment target. In practice, to overcome the high sensitivity to inevitable estimation errors of the input parameters, regularization techniques have been introduced for portfolio stabilization. Besides, regularization techniques have been shown to be capable of achieving some other goals like selecting sparse portfolios and grouping assets whose returns exhibit colinearity. While achieving such merits, acclaimed regularizers like the ℓ_1-norm may bring an adverse effect on the original portfolio design target. In this paper, a novel framework for vast portfolio selection is proposed via the lens of submodular set functions, which can select a sparse portfolio according to the influence that each asset exerts on the overall portfolio risk. An efficient and convergent algorithm based on the alternating direction method of multipliers is developed for problem resolution. The superiority of the proposed portfolio selection framework is demonstrated with numerical simulations on real market data.

I. INTRODUCTION

Portfolio selection (or portfolio optimization) is a fundamental problem in finance, which aims at assigning different portions of dollars to the underlying assets according to a certain investment target. In the early days, portfolio selection was commonly done based on the personal expertise of investors. In 1952, Markowitz ushered in the modern era of portfolio selection by introducing the mean-variance analysis [1], leading to the renowned mean-variance portfolio (MVP). The idea of MVP later becomes the backbone of the large majority of portfolio selection frameworks in the financial industry [2]. Despite its mathematical elegance, the vanilla MVP is prohibited by practitioners in the real financial world since the resulting portfolio weights are highly sensitive to the inevitable estimation errors in the asset expected returns and the asset return covariance matrices [3], which will result in unstable portfolio weights. Apart from that, the existence of multicollinearity in asset returns can cause extremely long and short positions with vanilla MVP, leading to high gross exposure of the portfolio [4]. Furthermore, the above issues can be more pronounced in the vast portfolio selection cases (i.e., the asset universe is large) [5].

To tame the troublesome instability and colinearity issues in vanilla MVPs, regularization procedures have been adopted for portfolio selection [6]–[8]. With proper regularizers, the regularized MVPs (RMVPs) can not only stabilize the portfolio weights but also promote sparsity (i.e., only select a subset of assets) which is essential for vast portfolio selection [3]. One of the most famous regularizers therein is the ℓ_1-norm that is largely used in portfolio selection and many others [9]–[11], with which the RMVPs can realize sparsity and stabilize the allocation. However, the ℓ_1-norm becomes invalid in the presence of the long-only constraint (i.e., portfolio weights must be nonnegative) as it degenerates to a constant. Besides, the RMVP with ℓ_1-norm will randomly select one or several among the assets that are highly correlated [12], resulting in inconsistent portfolio weights over different investment periods and cause high portfolio turnovers and hence high transaction costs [13]. To overcome the defects of the ℓ_1-norm, some weighted ℓ_1-norms have been introduced. E.g., the sorted ℓ_1-norm [14], [15] becomes prevailing in the portfolio selection field since with it the RMVPs can not only realize stable and sparse portfolios, but also group the highly correlated assets to which the same weights will be assigned [16]. Besides that, the sorted ℓ_1-norm still works in the long-only case. Note that many other regularizers have also been used in portfolio selection, such as the ℓ_2-norm (with which RMVPs can stabilize the allocation but cannot promote sparsity) [17], the combination of the ℓ_1-norm and the ℓ_∞-norm (a variation of the sorted ℓ_1-norm) [12], and so on.

Although some issues of the vanilla MVP can be alleviated with regularizations, it is acknowledged that the inclusion of the regularizers like the ℓ_1-norm will inevitably influence the original objective (e.g., portfolio risk minimization), which is a burning question in the field of portfolio selection [16]. To address this, we consider constructing regularizers for RMVP from a previously unconsidered perspective, that is, developing regularizers via the lens of submodular set functions. It can be seen that the aforementioned widely-used regularizers are all defined from the continuous optimization perspective, while the asset selection procedure involved in vast portfolio selection can be naturally regarded as a discrete optimization problem [18]. Therefore, regularizers can also be developed with set functions from the discrete optimization standpoint [19], through which new regularizers can be obtained.

In this paper, a submodular set function that can indicate the risk of sparse portfolios (i.e., the risk control set function) is

induced from the global minimum variance portfolio (GMVP), based on which a novel regularizer called risk control norm (RCN) is proposed. Regularized by RCN, RMVPs can select sparse portfolios according to the influence that each asset exerts on the overall portfolio risk. To solve the resulting portfolio selection problems, an efficient and convergent algorithm based on the alternating direction method of multipliers (ADMM) [20] is further developed, which is scalable and especially amenable to vast portfolio selection scenarios. It should be noted that the RCN can also be applied to many other portfolio selection problems or under additional portfolio constraints and the proposed ADMM algorithm is applicable with slight modifications. Furthermore, many existing regularizers such as the ℓ_1-norm and sorted ℓ_1-norm can be obtained through submodular set functions and new regularizers can be developed by following the similar scheme of designing RCN. Numerical simulations on real market data will showcase the superiority of the proposed RCN regularized portfolio selection framework.

II. PORTFOLIO SELECTION VIA THE RISK CONTROL NORM

A. The Risk Control Set Function $\sigma_\star^2(\mathcal{K})$

Consider N available financial assets in the market with asset returns at time t denoted by $\mathbf{r}_t = [r_{1,t}, \ldots, r_{N,t}]^T$. For a portfolio defined by $\mathbf{w} = [w_1, \ldots, w_N]^T$ that represents the proportion of dollars invested on the N assets, its expected return is given by

$$\mu(\mathbf{w}) = \mathbb{E}\left[\mathbf{w}^T \mathbf{r}_t\right] = \mathbf{w}^T \mathbb{E}\left[\mathbf{r}_t\right] = \mathbf{w}^T \boldsymbol{\mu},$$

where $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_N]^T$ is the expected return vector of the N assets. And the variance of the portfolio, commonly called portfolio risk, is calculated by

$$\sigma^2(\mathbf{w}) = \mathbb{E}\left[\left|\mathbf{w}^T \mathbf{r}_t - \mu(\mathbf{w})\right|^2\right] = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w},$$

where $\boldsymbol{\Sigma} = \mathbb{E}\left[\left(\mathbf{r}_t - \boldsymbol{\mu}\right)\left(\mathbf{r}_t - \boldsymbol{\mu}\right)^T\right]$ is the covariance matrix of the N asset returns.

Based on the portfolio risk $\sigma^2(\mathbf{w})$, the renowned GMVP problem [1] is given by

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{w} = 1, \end{aligned} \tag{1}$$

where $\mathbf{1}^T \mathbf{w} = 1$ denotes the portfolio budget constraint with $\mathbf{1}$ denoting the all-one vector.

Lemma 1 ([21]). *The optimal solution of the GMVP problem* (1) *is* $\mathbf{w}^\star = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1}\mathbf{1}}$.

Denote $\mathcal{N} = \{1, \ldots, N\}$ as the index set of all the available assets. Suppose we target at designing a sparse portfolio and use $\mathcal{K} = \{j_1, \ldots, j_K\} \subseteq \mathcal{N}$ with $K = |\mathcal{K}|$ (the cardinality of $\mathcal{K}$) to denote the index set of the active assets in the sparse portfolio. Then, we can obtain the active asset weight vector $\mathbf{w}_\mathcal{K} \in \mathbb{R}^K$, which is a subvector of $\mathbf{w}$. Accordingly, the active asset return vector and the active asset-active asset covariance matrix can be defined as $\boldsymbol{\mu}_\mathcal{K} \in \mathbb{R}^K$ (a subvector of $\boldsymbol{\mu}$) and $\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}} \in \mathbb{R}^{K \times K}$ (a submatrix of $\boldsymbol{\Sigma}$), respectively.

Based on Lemma 1, the risk of a sparse optimal GMVP for a given cardinality level K is given by

$$\begin{aligned} \sigma_\star^2(\mathcal{K}) &= \mathbf{w}_\mathcal{K}^{\star T} \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}} \mathbf{w}_\mathcal{K}^\star = \frac{\mathbf{1}_K^T \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}^{-1}}{\mathbf{1}_K^T \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}^{-1}\mathbf{1}_K} \cdot \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}} \cdot \frac{\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}^{-1}\mathbf{1}_K}{\mathbf{1}_K^T \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}^{-1}\mathbf{1}_K} \\ &= \frac{1}{\mathbf{1}_K^T \boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}^{-1}\mathbf{1}_K}, \end{aligned}$$

where the all-one vector $\mathbf{1}_K \in \mathbb{R}^K$. The $\sigma_\star^2(\mathcal{K})$ is a set function, which can be used to control the risk of the portfolio $\mathbf{w}$ by properly selecting the active asset in $\mathcal{N}$. In this paper, $\sigma_\star^2(\mathcal{K})$ will be named as the risk control set function. For portfolio optimization problems, their design objectives can be jointly optimized with $\sigma_\star^2(\mathcal{K})$ to select sparse portfolios, through which sparse portfolios can be selected according to the influence that each asset exerts on the overall portfolio risk. As a matter of fact, directly minimizing $\sigma_\star^2(\mathcal{K})$ can make the problem resolution prohibitive due to the high computational complexity resulting from its combinatorial nature. In the next section, we will resort to find a convex continuous substitution for the risk control set function.

B. The Risk Control Norm $\phi(\mathbf{w})$

We first give several useful results.

Definition 2. [Lovász extension [22]] Given a set function F with $F(\emptyset) = 0$ and $\mathbf{w} \in \mathbb{R}^N$, its Lovász extension $f : \mathbb{R}^N \to \mathbb{R}$ is given by

$$f(\mathbf{w}) = \sum_{n=1}^N \left(F\left(\mathcal{K}_n\right) - F\left(\mathcal{K}_{n-1}\right)\right)\left|w_{j_n}\right|,$$

where $(j_1, \ldots, j_N)$ denotes one permutation of $\{1, \ldots, N\}$ such that $|w_{j_1}| \geq \ldots \geq |w_{j_N}|$ and $\mathcal{K}_n = \{j_1, \ldots, j_n\}$ [1].

Definition 3. [Submodular set functions [23]] Consider $\mathcal{N} = \{1, \ldots, N\}$ with $N = |\mathcal{N}|$ and its power set (i.e., the set of all subsets) denoted by $2^\mathcal{N}$, a set function $F : 2^\mathcal{N} \to \mathbb{R}$ is said to be submodular if and only if, for any subsets $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{N}$ and an element $c \in \mathcal{N} \backslash \mathcal{B}$, it follows that $F(\{\mathcal{A}, c\}) - F(\mathcal{A}) \geq (F(\{\mathcal{B}, c\}) - F(\mathcal{B}))$.

Proposition 4 ([24]). *Assume submodular set function F is non-decreasing and strictly positive for all singletons, then it's Lovász extension $f(\mathbf{w})$ is a norm and it is the convex hull of $F(\mathrm{supp}(\mathbf{w}))$ on the unit ℓ_∞-ball with $\mathrm{supp}(\mathbf{w})$ denoting the support of $\mathbf{w}$.*

In this section, we will try to find a continuous surrogate of $\sigma_\star^2(\mathcal{K})$, where a natural candidate is its convex hull [25]. In general, computing the convex hull of a set function is NP-hard, while the convex hull of a submodular set function that satisfies the condition in Proposition 4 can be easily obtained through its Lovász extension [22]. However, it can be proved that the $\sigma_\star^2(\mathcal{K})$ is not submodular, hence we will first find a submodular substitute for $\sigma_\star^2(\mathcal{K})$. Observing that

$$\sigma_\star^2(\mathcal{K}) \leq \|\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}\|_2 \leq \|\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{K}}\|_F \leq \|\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{N}}\|_F \tag{2}$$

holds for all $\mathcal{K}$, where $\boldsymbol{\Sigma}_{\mathcal{K}\mathcal{N}} \in \mathbb{R}^{K \times N}$ is defined as the active asset-all asset covariance matrix (a submatrix of $\boldsymbol{\Sigma}$). And it can

---

[1]The subscript n in notation $\mathcal{K}_n$ has been uesd to denote the cardinality level of the set, i.e., $|\mathcal{K}_n| = n$.

$$L_\rho(\mathbf{w}, \mathbf{z}, u(x), \mathbf{v}(\mathbf{y})) = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu} + \gamma \sum_{n=1}^{N} \lambda_{j_n} |z_{j_n}| + x \left(\mathbf{1}^T \mathbf{w} - 1 \right) + \mathbf{y}^T \left(\mathbf{w} - \mathbf{z} \right) + \frac{\rho}{2} \left[\left(\mathbf{1}^T \mathbf{w} - 1 \right)^2 + \| \mathbf{w} - \mathbf{z} \|_2^2 \right]$$

$$= \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu} + \gamma \sum_{n=1}^{N} \lambda_{j_n} |z_{j_n}| + \frac{\rho}{2} \left[\left(\mathbf{1}^T \mathbf{w} - 1 + u \right)^2 + \| \mathbf{w} - \mathbf{z} + \mathbf{v} \|_2^2 \right] + const.,$$

$$\text{(AL)}$$

---

be verified $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{N}}\|_F$ is actually submodular while $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{K}}\|_2$ and $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{K}}\|_F$ are not, hence we will choose $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{N}}\|_F$ as the submodular substitute for $\sigma_\star^2(\mathcal{K})$. Furthermore, since $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{N}}\|_F$ satisfies all the conditions in Proposition 4, its Lovász extension $\phi(\mathbf{w})$ is a convex norm, which is given by

$$\phi(\mathbf{w}) = \sum_{n=1}^{N} \left(\|\mathbf{\Sigma}_{\mathcal{K}_n \mathcal{N}}\|_F - \|\mathbf{\Sigma}_{\mathcal{K}_{n-1}\mathcal{N}}\|_F \right) |w_{j_n}| = \sum_{n=1}^{N} \lambda_{j_n} |w_{j_n}|,$$

$$\text{(RCN)}$$

where $\lambda_{j_n} = \|\mathbf{\Sigma}_{\mathcal{K}_n \mathcal{N}}\|_F - \|\mathbf{\Sigma}_{\mathcal{K}_{n-1}\mathcal{N}}\|_F$.

In this paper, we will name $\phi(\mathbf{w})$ as the risk control norm (RCN) since it measures the portfolio risk, and the RCN will be incorporated as a convex surrogate of $\sigma_\star^2(\mathcal{K})$ in the vast portfolio selection problems as a regularizer. It is also worth mentioning that the above scheme of inducing the RCN from the portfolio risk can also be adopted to develop regularizers based on other criteria. Besides, some existing popular regularizers like the commonly used ℓ_1-norm and the sorted ℓ_1-norm in portfolio selection literature can also be explained from the viewpoint of submodular set functions [26].

It should be noted that although $\phi(\mathbf{w})$ takes the form of a weighted ℓ_1-norm, different from the traditional weighted ℓ_1-norm with heuristic constant weights, the weights of $\phi(\mathbf{w})$ is adaptively derived from $\|\mathbf{\Sigma}_{\mathcal{K}\mathcal{N}}\|_F$, representing the influence that each asset exerts on the overall portfolio risk. To be more specific, in $\sigma_\star^2(\mathcal{K})$ the asset with lower risk will be assigned with lower weight, and vice versa. In view of this, under a given sparsity level the RCN regularized portfolio tends to choose assets that will contribute lower risks to the whole portfolio, while the popular ℓ_1-norm and sorted ℓ_1-norm cannot attain this property. To explain this in another way, if we exchange the portfolio allocation weights of two assets (assume the portfolio allocation weights of them are different), the values of ℓ_1-norm and sorted ℓ_1-norm will not change, while the value of RCN will change if the risks of these two assets are different, through which the superiority of RCN is obvious. This conjecture becomes the main motivation for designing this RCN for vast portfolio selection problems which will be detailed in the next section.

C. Risk Control Norm Regularized Portfolio Optimization

In this paper we mainly investigate the RMVP optimization problem regularized by RCN which is given by

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu} + \gamma \phi(\mathbf{w}) \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{w} = 1, \end{aligned}$$

$$\text{(RMVP-RCN)}$$

where η and γ are tuning parameters. As a special case, if $\eta = 0$ it becomes the regularized GMVP (RGMVP)

problem. Problem (RMVP-RCN) is convex, for which efficient algorithms with global optimality guarantee can be derived.

It also need to be mentioned that apart from the RMVP and RGMVP problems, the RCN can also be extended to other portfolio selection problems like the index tracking portfolio [27] and the hedge portfolio [28] and it can also be applied under additional portfolio constraints [29].

III. ADMM ALGORITHM FOR RCN REGULARIZED PORTFOLIO OPTIMIZATION

In the section, we will develop a problem-tailored and scalable algorithm to tackle problem (RMVP-RCN). It should be noted that although problem (RMVP-RCN) is convex, it cannot be programmed by the disciplined convex programming language CVX [30] or other off-the-shelf solvers due to the special structure of λ_{j_n}. To handle this issue, an efficient algorithm based on ADMM [20] will be developed. Besides, due to its ability in variable splitting, ADMM is efficient in "arbitrary-scale" optimization problems, which is beneficial to the vast portfolio selection problems.

For the algorithm development, an auxiliary variable $\mathbf{z}$ is firstly introduced and then problem (RMVP-RCN) can be rewritten as follows:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} - \eta \mathbf{w}^T \boldsymbol{\mu} + \gamma \phi(\mathbf{z}) \\ \text{subject to} \quad & \mathbf{1}^T \mathbf{w} = 1, \mathbf{w} = \mathbf{z}. \end{aligned}$$

$$\text{(3)}$$

The augmented Lagrangian function of (3) is then given in (AL), where $\rho > 0$ is a constant penalty parameter, the scaled dual variables are accordingly defined as $u \triangleq \frac{1}{\rho} x$ and $\mathbf{v} \triangleq \frac{1}{\rho} \mathbf{y}$ respectively, and $const.$ represents the constant term. Based on ADMM, the variables $(\mathbf{w}, \mathbf{z}, u, \mathbf{v})$ will be updated cyclically by minimizing (AL).

The w-minimization step. To minimize L_ρ over $\mathbf{w}$, we first find a majorized function of (AL).

Lemma 5 (Quadratic Majorization [13]). *Let* $\mathbf{A} \in \mathbb{R}^N$, *at any given point* $\mathbf{w}^{(t)} \in \mathbb{R}^N$ *the following relation holds* $\mathbf{w}^T \mathbf{A} \mathbf{w} \leq \psi \mathbf{w}^T \mathbf{w} + 2 \mathbf{w}^T (\mathbf{A} - \psi \mathbf{I}) \mathbf{w}^{(t)} + \mathbf{w}^{(t)T} (\psi \mathbf{I} - \mathbf{A}) \mathbf{w}^{(t)}$, *with the equality attained at* $\mathbf{w} = \mathbf{w}^{(t)}$ *and the constant* ψ *is greater than the largest eigenvalue of* $\mathbf{A}$.

Based on Lemma 5, a majorized function of (AL) with respect to variable $\mathbf{w}$ is

$$\begin{aligned} & \overline{L}_\rho^{(t)}(\mathbf{w}, \mathbf{z}^{(t)}, u^{(t)}, \mathbf{v}^{(t)} \mid \mathbf{w}^{(t)}) \\ = & \psi \mathbf{w}^T \mathbf{w} + 2 \mathbf{w}^T (\mathbf{\Sigma} + \frac{\rho}{2} \mathbf{1} \mathbf{1}^T - \psi \mathbf{I}) \mathbf{w}^{(t)} - \eta \mathbf{w}^T \boldsymbol{\mu} \\ & + \rho(u^{(t)} - 1) \mathbf{1}^T \mathbf{w} + \frac{\rho}{2} \| \mathbf{w} - \mathbf{z} + \mathbf{v} \|_2^2 + const. \\ = & (\psi + \frac{\rho}{2}) \| \mathbf{w} - \mathbf{m}^{(t)} \|_2^2 + const., \end{aligned}$$

$$\text{(4)}$$

2101

where $\psi \geq \lambda_{\max}\left(\mathbf{\Sigma} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T\right)$ and $\mathbf{m}^{(t)} \triangleq \frac{-1}{2\psi+\rho}\left[2\left(\mathbf{\Sigma} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T - \psi\mathbf{I}\right)\mathbf{w}^{(t)} - \eta\boldsymbol{\mu} + \rho\left(u^{(t)} - 1\right)\mathbf{1} + \rho\left(\mathbf{v}^{(t)} - \mathbf{z}^{(t)}\right)\right]$. Instead of minimizing the original problem (AL), problem (4) will be minimized to update $\mathbf{w}^{(t+1)}$ which is given by

$$\mathbf{w}^{(t+1)} = \underset{\mathbf{w}}{\text{argmin}}\left\{(\psi + \frac{\rho}{2})\|\mathbf{w} - \mathbf{m}^{(t)}\|_2^2 + cst.\right\}$$

$$= \frac{-1}{2\psi + \rho}\left[2(\mathbf{\Sigma} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T - \psi\mathbf{I})\mathbf{w}^{(t)} - \eta\boldsymbol{\mu} \qquad (5)\right.$$

$$\left. + \rho(u^{(t)} - 1)\mathbf{1} + \rho(\mathbf{v}^{(t)} - \mathbf{z}^{(t)})\right].$$

The z-minimization step. The augmented Lagrangian function L_ρ in terms of variable $\mathbf{z}$ is

$$L_\rho(\mathbf{w}^{(t+1)}, \mathbf{z}, u^{(t)}, \mathbf{v}^{(t)}) = \frac{\rho}{2}\|\mathbf{z} - \mathbf{n}^{(t)}\|^2 + \gamma\phi\left(\mathbf{z}\right) + cst.,$$

where $\mathbf{n}^{(t)} \triangleq \mathbf{w}^{(t+1)} + \mathbf{v}^{(t)}$. By replacing $\phi\left(\mathbf{z}\right)$ with its dual norm, the updating for variable $\mathbf{z}$ is a proximal step [31] as

$$\begin{array}{ll} \underset{\mathbf{z}}{\text{minimize}} & \underset{\boldsymbol{\kappa} \in \mathbb{R}^n}{\max} \quad \frac{\rho}{2}\|\mathbf{z} - \mathbf{n}^{(t)}\|^2 + \gamma \sum_{n=1}^N \kappa_n |z_{j_n}| \\ \text{subject to} & \mathbf{1}^T\boldsymbol{\kappa} \leq \|\mathbf{\Sigma}_{\mathcal{N}\mathcal{N}}\|_F, \end{array} \qquad (6)$$

which can be solved via decomposition algorithms [24].

Dual variable update. The updating rule for the dual variables u and $\mathbf{v}$ are accordingly given by

$$\begin{aligned} u^{(t+1)} &= u^{(t)} + \mathbf{1}^T\mathbf{w}^{(t+1)} - 1, \\ \mathbf{v}^{(t+1)} &= \mathbf{v}^{(t)} + \mathbf{w}^{(t+1)} - \mathbf{z}^{(t+1)}. \end{aligned} \qquad (7)$$

In summary, variables $(\mathbf{w}, \mathbf{z}, u, \mathbf{v})$ will be updated cyclically until some convergence criteria are attained. The overall algorithm is outlined in Algorithm 1.

Algorithm 1 The ADMM Alg. for Prob. (RMVP-RCN)

Input: $\mathbf{\Sigma}$, ρ, γ, and $\psi \geq \lambda_{\max}\left(\mathbf{\Sigma} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T\right)$.
Initialize $\mathbf{w}^{(0)} = \frac{1}{N}\mathbf{1}$, $\mathbf{z}^{(0)} = \mathbf{v}^{(0)} = \mathbf{0}$, $u^{(0)} = 0$, $t = 0$.
Repeat
  1) Updating variable $\mathbf{w}$ via Eq. (5);
  2) Updating variable $\mathbf{z}$ by solving Prob. (6);
  3) Updating variables u and $\mathbf{v}$ via Eq. (7);
  4) $t \leftarrow t + 1$
Until the termination criteria are satisfied.
Output: $\mathbf{w}^{(t)}$.

We also outline the convergence result in the following.
Lemma 6 ([32])**.** *The ADMM Algorithm 1 converges globally to the optimal solutions of problem* (RMVP-RCN).
Remark 7. The proposed ADMM can be extended to some other RCN regularized portfolio selection problems and/or applied under additional portfolio constraints (e.g., leverage, turnover, self-financing, holding) with slight modifications.

IV. NUMERICAL SIMULATIONS

In this section, the superiority of RCN will be demonstrated based on the historical data of S&P 500 stocks from the U.S. market. The RGMVP and RMVP selection problems are considered. In-sample and out-of-sample performance are given in Section IV-A and Section IV-B. In the out-of-sample performance evaluation stage, the rolling-window scheme will

be used, which has a lookback window with length of 252 days (i.e., trading days for one year) for obtaining the optimal portfolio and a window length of 22 days (i.e., trading days for one month) to test the portfolio performance, that is, the portfolio will be updated per month based on the one-year data ahead of that month.

A. RGMVP with RCN and ℓ_1-Norm

Sparsity promoting property of the RGMVPs. We first check the sparsity promoting property of RGMVP with RCN and ℓ_1-norm. The available asset universe is chosen randomly from the S&P 500 with $N = 50$ and a time span of 252. The sparsity levels of RGMVP, RGMVP-ℓ_1-norm, and RGMVP-RCN with different γ controlling the cardinality of $\mathbf{w}$ are depicted in Fig. 1. Both RGMVP-ℓ_1-norm and RGMVP-RCN can achieve sparse portfolios, while for RGMVP-ℓ_1-norm, the sparsity level stops increasing when γ is larger than a breakpoint, which is due to the regularizer "degenerates" to a constant [16].
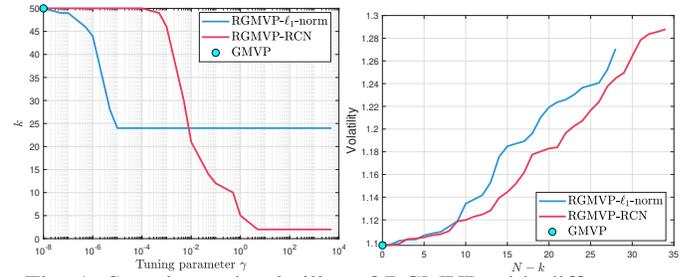

Fig. 1: Sparsity and volatility of RGMVP with different γ.

Volatility profiles of the RGMVPs. The comparison of volatility profiles (i.e., portfolio risk) of the RGMVP-RCN and the RGMVP-ℓ_1-norm is presented in Fig. 1. We can see that the volatility of RGMVP-RCN and RGMVP-ℓ_1-norm is larger than the volatility of the GMVP portfolio, while RGMVP-RCN achieves lower volatility than RGMVP-ℓ_1-norm does under the same sparsity level.

We further examine how the out-of-sample performance scale with the asset universe where we have chosen $N = 15$, 30, 50, and 100. Comparison results between RGMVP-RCN and RGMVP-ℓ_1-norm in terms of volatility are listed in Table I, where the sparsity level and volatility are averaged over six months and $*$ indicating that such sparsity level cannot be achieved. It can be seen that RGMVP-RCN can achieve lower volatility than RGMVP-ℓ_1-norm does in all cases.

B. RMVP with RCN and ℓ_1-Norm

In the context of RMVP problems, the comparisons between RMVP-ℓ_1-norm and RMVP-RCN in terms of the cumulative returns over 96 months with the same sparsity level (i.e., the sparsity ratios are the same and we have chosen 10% here) are presented. The out-of-sample results are shown in Fig. 2. The comparison is benchmarked by the MVP and S&P 500, which demonstrates RMVP-RCN can achieve a certain sparsity level with lower volatility than RMVP-ℓ_1-norm in the long run.

V. CONCLUSIONS

In this paper, a novel regularizer named risk control norm is brought up via the lens of submodular set functions, based on which a general vast risk control norm regularized portfolio

TABLE I: Out-of-sample performance in terms of volatility of different assets universe scales.

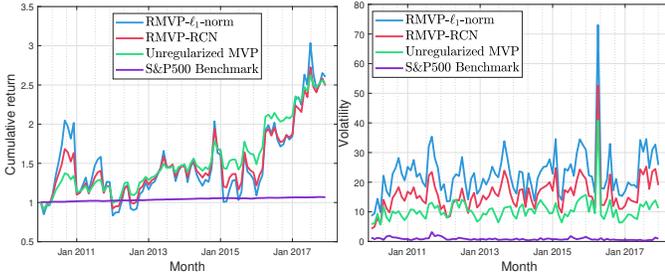| N	RGMVP	Sparsity Level					
		0%	$\sim 10\%$	$\sim 30\%$	$\sim 50\%$	$\sim 70\%$	$\sim 90\%$
15	RCN	25.64%	$\sim 25.88\%$	$\sim 26.64\%$	$\sim 27.20\%$	$\sim 27.61\%$	$\sim 30.38\%$
	ℓ_1-norm	25.64%	$\sim 25.89\%$	$\sim 26.83\%$	$\sim 27.33\%$	*	*
30	RCN	23.77%	$\sim 23.83\%$	$\sim 24.57\%$	$\sim 24.91\%$	$\sim 25.77\%$	$\sim 26.17\%$
	ℓ_1-norm	23.77%	$\sim 24.09\%$	$\sim 25.31\%$	$\sim 26.09\%$	*	*
50	RCN	22.57%	$\sim 22.66\%$	$\sim 23.88\%$	$\sim 25.02\%$	$\sim 25.60\%$	$\sim 27.28\%$
	ℓ_1-norm	22.57%	$\sim 22.80\%$	$\sim 24.39\%$	$\sim 25.69\%$	*	*
100	RCN	21.45%	$\sim 21.45\%$	$\sim 21.85\%$	$\sim 22.54\%$	$\sim 24.03\%$	*
	ℓ_1-norm	21.45%	$\sim 21.52\%$	$\sim 22.00\%$	$\sim 23.39\%$	$\sim 24.92\%$	*

Fig. 2: Cumulative return and volatility of RMVP over 8 years.

selection problem is proposed. The proposed risk control norm can be inteprated as an adaptively weighted ℓ_1-norm, based on which the regularized portfolio optimization problem can select portfolios according to the influence that each asset exerts on the overall portfolio risk. An efficient algorithm based on ADMM is developed for problem resolution. The superiority of the proposed method is demonstrated through numerical simulations on real data. Some possible future lines of research are applications to other portfolios like the risk-parity portfolio and the index tracking portfolio.

REFERENCES

[1] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] P. N. Kolm, R. Tütüncü, and F. J. Fabozzi, "60 years of portfolio optimization: Practical challenges and current trends," *European Journal of Operational Research*, vol. 234, no. 2, pp. 356–371, 2014.
[3] 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.
[4] R. Jagannathan and T. Ma, "Risk reduction in large portfolios: Why imposing the wrong constraints helps," *The Journal of Finance*, vol. 58, no. 4, pp. 1651–1683, 2003.
[5] J. Fan, J. Zhang, and K. Yu, "Vast portfolio selection with gross-exposure constraints," *Journal of the American Statistical Association*, vol. 107, no. 498, pp. 592–606, 2012.
[6] B. Bruder, N. Gaussel, J.-C. Richard, and T. Roncalli, "Regularization of portfolio allocation," *Available at SSRN 2767358*, 2013.
[7] M. Carrasco and N. Noumon, "Optimal portfolio selection using regularization," Citeseer, Tech. Rep., 2011.
[8] Z. Zhao, R. Zhou, and D. P. Palomar, "Optimal mean-reverting portfolio with leverage constraint for statistical arbitrage in finance," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1681–1695, 2019.
[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
[10] Z. Zhao and D. P. Palomar, "Robust maximum likelihood estimation of sparse vector error correction model," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2017, pp. 913–917.
[11] ——, "Sparse reduced rank regression with nonconvex regularization," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 811–815.
[12] H. D. Bondell and B. J. Reich, "Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR," *Biometrics*, vol. 64, no. 1, pp. 115–123, 2008.
[13] Z. Zhao and D. P. Palomar, "Large-scale regularized portfolio selection via convex optimization," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.
[14] X. Zeng and M. A. Figueiredo, "Decreasing weighted sorted ℓ_1 regularization," *IEEE Signal Processing Letters*, vol. 21, no. 10, pp. 1240–1244, 2014.
[15] M. Bogdan, E. Van Den Berg, C. Sabatti, W. Su, and E. J. Candès, "SLOPE: adaptive variable selection via convex optimization," *The annals of applied statistics*, vol. 9, no. 3, p. 1103, 2015.
[16] P. J. Kremer, S. Lee, M. Bogdan, and S. Paterlini, "Sparse portfolio selection via the sorted ℓ_1 norm," *Journal of Banking & Finance*, vol. 110, p. 105687, 2020.
[17] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[18] J. Lee, *A first course in combinatorial optimization*. Cambridge University Press, 2004, vol. 36.
[19] F. Bach, R. Jenatton, J. Mairal, G. Obozinski *et al.*, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.
[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
[21] Y. Feng and D. P. Palomar, "A signal processing perspective on financial engineering," *Foundations and Trends in Signal Processing*, p. 1, 2016.
[22] L. Lovász, "Submodular functions and convexity," in *Mathematical programming the state of the art*. Springer, 1983, pp. 235–257.
[23] S. Fujishige, *Submodular functions and optimization*. Elsevier, 2005.
[24] F. Bach *et al.*, "Learning with submodular functions: A convex optimization perspective," *Foundations and Trends® in Machine Learning*, vol. 6, no. 2-3, pp. 145–373, 2013.
[25] M. Berman, A. R. Triki, and M. B. Blaschko, "The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4413–4421.
[26] K. Minami, "Degrees of freedom in submodular regularization: A computational perspective of Stein's unbiased risk estimate," *Journal of Multivariate Analysis*, vol. 175, p. 104546, 2020.
[27] J. E. Beasley, N. Meade, and T.-J. Chang, "An evolutionary heuristic for the index tracking problem," *European Journal of Operational Research*, vol. 148, no. 3, pp. 621–643, 2003.
[28] D. Giamouridis and I. D. Vrontos, "Hedge fund portfolio construction: A comparison of static and dynamic approaches," *Journal of Banking & Finance*, vol. 31, no. 1, pp. 199–217, 2007.
[29] Z. Zhao and D. P. Palomar, "Mean-reverting portfolio with budget constraint," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2342–2357, 2018.
[30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.
[31] G. Obozinski and F. Bach, "A unified perspective on convex structured sparsity: Hierarchical, symmetric, submodular norms and beyond," Tech. Rep., 2016.
[32] M. Hong and Z.-Q. Luo, "On the linear convergence of the alternating direction method of multipliers," *Mathematical Programming*, vol. 162, no. 1-2, pp. 165–199, 2017.