Mathematical Modeling

Lecture 5. Fixed-Point Proximity Algorithm for Portfolio Optimization

Yizun Lin, Zhao-Rong Lai

Department of Mathematics, Jinan University

April 29, 2022

Introduction review portfolio optimization
existing work
short comings
introduce sections

1 Sparse and Stable Markowitz Portfolios

numerical result

Conclusion

Refrence

Portfolio optimization (PO) with machine learning methods has become a prospective approach in advancing the interdiscipline of financial engineering. The first proposal of the mean-variance (MV) approach is given by Markowitz [5], his criterion has become the most popular one for many PO models. In brief, the original MV (OMV) model is

covariance

$$\begin{aligned} \hat{\boldsymbol{w}} = \operatorname*{argmin}_{\boldsymbol{w}\in\mathbb{R}^N} & \ \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s. t.} & \ \boldsymbol{w}^\top \boldsymbol{\mu} = \overset{0.066}{\rho}, \ \boldsymbol{w}^\top \mathbf{1}_N = 1, \ \boldsymbol{w} \geq \mathbf{0}_N, \end{aligned} \tag{1.1}$$

$(w^\top \mathrm{Var}(\mu))$

$A=\begin{bmatrix} \mu^\top \\ 1_n^\top \end{bmatrix}$ $b=\begin{bmatrix} \rho \\ 1 \end{bmatrix}$ $Aw=b$ $Iw \geq 0_N$

expected return (average return)

where $\boldsymbol{w}$ denotes the N-dimensional portfolio (with respect to N assets); $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the expected return and the return covariance of these N assets, respectively; $\boldsymbol{w}^\top \mathbf{1}_N = 1$ ($\mathbf{1}_N$ denotes the vector of N ones) is the self-financing constraint, which indicates that no additional money can be used and full re-investment is compulsory. $\boldsymbol{w}^\top \boldsymbol{\mu} = \rho$ means that the expected portfolio return is fixed at a level of ρ. The objective is to minimize the portfolio variance $\boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}$ (considered as the portfolio risk) at this return level.

Brodie et al. [1] propose the Sparse and Stable Markowitz Portfolios (SSMP) model

every day risk approx to ρ — every weight approx to 0

$$\begin{aligned} \hat{\boldsymbol{w}} = \operatorname*{argmin}_{\boldsymbol{w}\in\mathbb{R}^N} & \left\{ \frac{1}{T} \| \underline{\boldsymbol{R}\boldsymbol{w} - \rho \mathbf{1}_T} \|_2^2 + \underline{\tau \|\boldsymbol{w}\|_1} \right\}, \\ \text{s. t.} & \ \boldsymbol{w}^\top \hat{\boldsymbol{\mu}} \overset{\geq}{=} \rho, \boldsymbol{w}^\top \mathbf{1}_N = 1, \ \boldsymbol{w} \geq \mathbf{0}_N, \end{aligned} \tag{1.2}$$

$\mathrm{prox}_{|\cdot|}(x) = \max\{|x|-1, 0\}\mathrm{sgn}(x)$

can be negative

where $\boldsymbol{R} \in \mathbb{R}^{T\times N}$ is the sample asset return matrix (T trading times and N assets), $\boldsymbol{r}^{(t)}$ denotes the t-th row of $\boldsymbol{R}$ (i.e., the asset returns at time t), $\hat{\boldsymbol{\mu}} := \frac{1}{T}\boldsymbol{R}^\top \mathbf{1}_T$ is a column vector of sample mean returns, $\rho \in \mathbb{R}$ is a given expected return level, $\tau \geq 0$ is the regularization parameter, $\|\cdot\|_2$ is the ℓ^2-norm and $\|\cdot\|_1$ is the ℓ^1-norm. In this model, the portfolio risk is embedded in the quadratic form $\frac{1}{T}\|\boldsymbol{R}\boldsymbol{w} - \rho\mathbf{1}_T\|^2$, which computes the mean squared error of the sample portfolio return $\boldsymbol{r}^{(t)}\boldsymbol{w}$ fitting the given level ρ. Therefore, SSMP tries to obtain a sparse portfolio that minimizes the risk at the return level ρ. To see the relationship between (1.2) and (1.1), one can expand $\frac{1}{T}\|\boldsymbol{R}\boldsymbol{w} - \rho\mathbf{1}_T\|^2$ as $\frac{1}{T}(\boldsymbol{w}^\top \underbrace{\boldsymbol{R}^\top \boldsymbol{R}}_{\Sigma} \boldsymbol{w} - 2\rho\mathbf{1}_T^\top \boldsymbol{R}\boldsymbol{w} + \rho^2 T)$, which is a quadratic function of $\boldsymbol{w}$ with the symmetric matrix $\frac{1}{T}\boldsymbol{R}^\top\boldsymbol{R}$. On the other hand, the sample estimator $\hat{\boldsymbol{\Sigma}}$ for (1.1) is $\frac{1}{T-1}\boldsymbol{R}^\top(\boldsymbol{I}_T - \frac{1}{T}\boldsymbol{1}_T\boldsymbol{1}_T^\top)\boldsymbol{R}$, which is a centralized version of $\frac{1}{T}\boldsymbol{R}^\top\boldsymbol{R}$. By this way, (1.2) essentially follows the Markowitz's criterion.

Note: $\boldsymbol{R}$ can be computed by $\boldsymbol{R} := \boldsymbol{X} - \boldsymbol{1}_{T\times N}$, where $\boldsymbol{X} \in \mathbb{R}^{T\times N}$ is the price relative matrix.

data −1

2 FPPA for Portfolio Optimization

2.1 FPPA for OMV

We first consider FPPA for solving model (1.1). To this end, we rewrite model (1.1) as a more compact form. Let

$$f_1(\boldsymbol{w}) := \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w}, \quad \boldsymbol{w} \in \mathbb{R}^N,$$

$$\boldsymbol{A} := \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{1}_N \end{pmatrix}, \text{ and } \boldsymbol{b} := \begin{pmatrix} \rho \\ 1 \end{pmatrix}.$$

Then model (1.1) can be rewritten by

$$\begin{aligned} \hat{\boldsymbol{w}} &= \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmin}} f_1(\boldsymbol{w}), \\ &\text{s. t. } \boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}, \ \boldsymbol{I}\boldsymbol{w} \geq \boldsymbol{0}_N. \end{aligned} \tag{2.1}$$

In fact, the equality constraint $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{b}$ is equivalent to the following two inequality constraints

$$\boldsymbol{A}\boldsymbol{w} \geq \boldsymbol{b} \text{ and } -\boldsymbol{A}\boldsymbol{w} \geq -\boldsymbol{b}.$$

We further define

$$\boldsymbol{D} := \begin{pmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \\ \boldsymbol{I} \end{pmatrix} \in \mathbb{R}^{(N+4)\times N}, \text{ and } \boldsymbol{d} := \begin{pmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \\ \boldsymbol{0}_N \end{pmatrix} \in \mathbb{R}^{N+4}.$$

Then (2.1) becomes

$$\begin{aligned} \hat{\boldsymbol{w}} &= \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmin}} f_1(\boldsymbol{w}), \\ &\text{s. t. } \boldsymbol{D}\boldsymbol{w} \geq \boldsymbol{d}. \end{aligned} \tag{2.2}$$

We next turn the above model into an unconstrained optimization model. We define the indicator function $\iota_{\boldsymbol{d}} : \mathbb{R}^{N+4} \to \{0\} \cup \{+\infty\}$ by

$$\iota_{\boldsymbol{d}}(\boldsymbol{z}) := \begin{cases} 0, & if\ \boldsymbol{z} \geq \boldsymbol{d}, \\ +\infty, & else. \end{cases}$$

$$\begin{cases} \boldsymbol{x}^{k+1} = \operatorname{prox}_{\beta h}\left(\boldsymbol{x}^k - \beta(\nabla f(\boldsymbol{x}^k) + B^\top \boldsymbol{y}^k)\right), \\ \boldsymbol{y}^{k+1} = \omega(\mathcal{I} - \operatorname{prox}_{\frac{1}{\omega}g})(\frac{1}{\omega}\boldsymbol{y}^k + B(2\boldsymbol{x}^{k+1} - \boldsymbol{x}^k)). \end{cases}$$

$\iota_d(x)$

convex function
$\Uparrow$
convex set

$f(\lambda x_1 + (1-\lambda)x_2) \geq \lambda f(x_1) + (1-\lambda) f(x_2)$

Then model (2.2) is equivalent to

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmin}}\, f_1(\boldsymbol{w}) + \iota_{\boldsymbol{d}}(\boldsymbol{D}\boldsymbol{w}). \tag{2.3}$$

$f(x) + g(Bx) + h(x) \rightarrow$ zero function

$\operatorname{prox}_h = I$

Then the fixed-point proximity algorithm (FPPA) to solve model (2.3) can be given by

$$\begin{cases} \boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \beta(\nabla f_1(\boldsymbol{w}^k) + \boldsymbol{D}^\top \boldsymbol{y}^k), \\ \boldsymbol{y}^{k+1} = \omega(\mathcal{I} - \operatorname{prox}_{\frac{1}{\omega}\iota_{\boldsymbol{d}}})(\frac{1}{\omega}\boldsymbol{y}^k + \boldsymbol{D}(2\boldsymbol{w}^{k+1} - \boldsymbol{w}^k)). \end{cases}$$

From the definition of $\iota_{\boldsymbol{d}}$, we know that $\frac{1}{\omega}\iota_{\boldsymbol{d}} = \iota_{\boldsymbol{d}}$. Hence the above iteration can be rewritten by

$$\begin{cases} \boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \beta(\nabla f_1(\boldsymbol{w}^k) + \boldsymbol{D}^\top \boldsymbol{y}^k), \\ \boldsymbol{y}^{k+1} = \omega(\mathcal{I} - \operatorname{prox}_{\iota_{\boldsymbol{d}}})(\frac{1}{\omega}\boldsymbol{y}^k + \boldsymbol{D}(2\boldsymbol{w}^{k+1} - \boldsymbol{w}^k)). \end{cases} \tag{2.4}$$

Convergence condition:

$$\beta \in \left(0, \frac{2}{L_1}\right), \ \omega \in \left(0, \frac{2(2-\beta L_1)}{4\beta\|\boldsymbol{D}\|_2^2 + L_1(2-\beta L_1)}\right).$$

where L_1 is the Lipschitz constant of ∇f_1. Since $\nabla f_1(\boldsymbol{w}) = 2\boldsymbol{\Sigma}\boldsymbol{w}$ and

$$\|\nabla f_1(\boldsymbol{x}) - \nabla f_1(\boldsymbol{y})\|_2 = 2\|\boldsymbol{\Sigma}\boldsymbol{x} - \boldsymbol{\Sigma}\boldsymbol{y}\|_2 \leq \underbrace{2\|\boldsymbol{\Sigma}\|_2}_{L=2\|\Sigma\|_2}\|\boldsymbol{x} - \boldsymbol{y}\|_2, \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N,$$

we have that

$$L_1 = 2\|\boldsymbol{\Sigma}\|_2.$$

To implement the above FPPA iteration, we still need the closed form of $\operatorname{prox}_{\iota_{\boldsymbol{d}}}$, which can be obtained as follows:

$$\begin{aligned} \operatorname{prox}_{\iota_{\boldsymbol{d}}}(\boldsymbol{z}) &= \underset{\boldsymbol{u}\in\mathbb{R}^{N+4}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{z}\|_2^2 + \iota_{\boldsymbol{d}}(\boldsymbol{u}) \right\} \\ &= \underset{\boldsymbol{u}\geq\boldsymbol{d}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N+4} (u_i - z_i)^2 \right\} \\ &= \max(\boldsymbol{z}, \boldsymbol{d}). \end{aligned}$$

if $z_i \geq d_i$ $u_i = z_i$ distance is 0 $(z_i - z_i)^2 = 0$
if $z_i < d_i$ $u_i = d_i$ that is $(d_i - z_i)^2$
$\downarrow$
$\max(z, d)$

2.2 FPPA for SSMP

Let $f_2(\boldsymbol{w}) := \frac{1}{T}\|\boldsymbol{R}\boldsymbol{w} - \rho\boldsymbol{1}_T\|_2^2$ and $h(\boldsymbol{w}) := \tau\|\boldsymbol{w}\|_1$, $\boldsymbol{w} \in \mathbb{R}^N$. Then model (1.2) can be rewritten by

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmin}}\, f_2(\boldsymbol{w}) + \iota_{\boldsymbol{d}}(\boldsymbol{D}\boldsymbol{w}) + h(\boldsymbol{w}). \tag{2.5}$$

The fixed-point proximity algorithm (FPPA) to solve model (2.5) can be given by

$$\begin{cases} \boldsymbol{w}^{k+1} = \mathrm{prox}_{\beta h}\left(\boldsymbol{w}^k - \beta(\nabla f_2(\boldsymbol{w}^k) + \boldsymbol{D}^\top \boldsymbol{y}^k)\right), \\ \boldsymbol{y}^{k+1} = \omega(\mathcal{I} - \mathrm{prox}_{\frac{1}{\omega}\iota_{\boldsymbol{d}}})(\frac{1}{\omega}\boldsymbol{y}^k + \boldsymbol{D}(2\boldsymbol{w}^{k+1} - \boldsymbol{w}^k)). \end{cases}$$

Convergence condition:

$$\beta \in \left(0, \frac{2}{L_2}\right),\ \omega \in \left(0, \frac{2(2-\beta L_2)}{4\beta\|\boldsymbol{D}\|_2^2 + L_2(2-\beta L_2)}\right),$$

where L_2 is the Lipschitz constant of ∇f_2.

The closed form of $\mathrm{prox}_{\beta h}$ is given by

$$\mathrm{prox}_{\beta h}(\boldsymbol{x}) = \mathrm{prox}_{\beta\tau\|\cdot\|_1}(\boldsymbol{x}) = \left(\mathrm{prox}_{\beta\tau|\cdot|}(x_1),\ \mathrm{prox}_{\beta\tau|\cdot|}(x_2),\ \ldots,\ \mathrm{prox}_{\beta\tau|\cdot|}(x_N)\right)^\top,$$

where $\mathrm{prox}_{\beta\tau|\cdot|}(t) = \max(|t| - \beta\tau, 0)\cdot \mathrm{sgn}(t)$ for $t \in \mathbb{R}$ (see [6]).

Stopping criterion for the iteration: Set the tolerance (e.g. $tol = 10^{-7}$) and the maximum iteration number (e.g. $MaxIter = 10^4$). Repeat the iteration until either $\frac{\|\boldsymbol{w}^k - \boldsymbol{w}^{k-1}\|_2}{\|\boldsymbol{w}^{k-1}\|_2} \le tol$ or the maximum iteration number is achieved.

3 Experimental Setting

Conduct extensive experiments on 6 benchmark data sets from Kenneth R. French's Data Library[1] (a standard and widely-used data library for long-term PO), named FRENCH32, FF25, FF25EU, FF48, FF100 and FF100MEOP. FRENCH32 is the same as the FRENCH data set used in [2]. FF25 contains 25 portfolios (they can also be considered as "assets" in our experiments) formed on BE/ME (book equity to market equity) and investment from the US market. FF25EU contains 25 portfolios formed on ME and prior return from the European market. FF48 contains 48 industry portfolios from the US market. FF100 contains 100 portfolios formed on ME and BE/ME, while FF100MEOP contains 100 portfolios formed on ME and operating profitability, all from the US market. All these data sets are monthly price relative sequences, which is a conventional frequency setting for long-term PO. Their profiles are shown in Table 1.

Adopt the moving-window trading framework [3] in the experiments, which is consistent with practical portfolio management. In brief, a window size T and the initial wealth $S^{(0)} = 1$ are preset for a strategy, then the price relatives in the time window $t = [1 : T]$ are used

[1] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

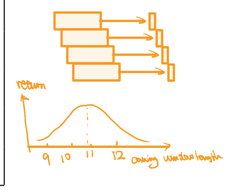
Table 1: Information of 6 benchmark data sets from real-world financial markets.

Data Set	Region	Time	Months	Assets
FRENCH32	US	$Jul/1990 \sim Feb/2020$	356	32
FF25	US	$Jul/1971 \sim Oct/2021$	604	25
FF25EU	EU	$Nov/1990 \sim Oct/2021$	372	25
FF48	US	$Jul/1971 \sim Oct/2021$	604	48
FF100	US	$Jul/1971 \sim Oct/2021$	604	100
FF100MEOP	US	$Jul/1971 \sim Oct/2021$	604	100

Coming window

X months → next month

form 10 months → 11th month



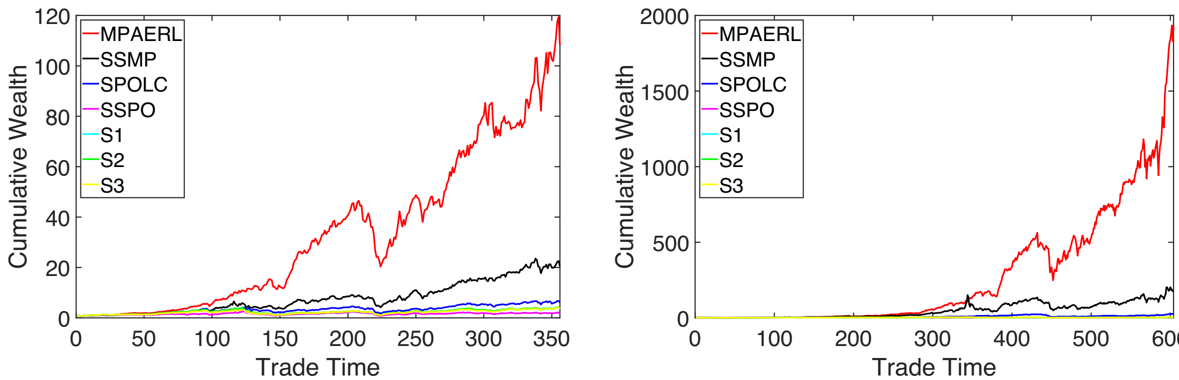
to update the portfolio $\hat{\boldsymbol{w}}^{(T+1)}$ for the next trading period. Then we proceed to $(T+1)$ and update the cumulative wealth $S^{(T+1)} = (\mathbf{x}^{(T+1)} \cdot \hat{\boldsymbol{w}}^{(T+1)})S^{(T)}$. In the next round, the price relatives in the time window $t = [2 : (T+1)]$ are used to update the portfolio $\hat{\boldsymbol{w}}^{(T+2)}$, and the above procedure is repeated, till the last period $\mathscr{T}$ of the investment. The equal-weight portfolio can be used at the beginning where there are insufficient samples to run a strategy. By this way, we obtain a backtest sequence $\{S^{(t)}\}_{t=0}^{\mathscr{T}}$ of cumulative wealths, which can be used to compute several evaluating scores for the investing performance and the risk assessment.

Present the performances of the following evaluating indicators by plots of tables (One can refer to [2, 4]):

- Cumulative Wealth
- Sharpe Ratio
- Maximum Drawdown

Of course, it is better to present more results not limited to the above.

Some sample figures:



(a) FRENCH32

(b) FF100

Figure 1: Cumulative wealths of different strategies with respect to trade time on 2 benchmark data sets.

References

[1] Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris. Sparse and stable Markowitz portfolios. Proceedings of the National Academy of Sciences of the United States of America, 106(30):12267–12272, Jul. 2009.

[2] Zhao-Rong Lai, Liming Tan, Xiaotian Wu, and Liangda Fang. Loss control with rank-one covariance estimate for short-term portfolio optimization. Journal of Machine Learning Research, 21(97):1–37, Jun. 2020.

[3] Zhao-Rong Lai and Haisheng Yang. A survey on gaps between mean-variance approach and exponential growth rate approach for portfolio optimization. ACM Computing Surveys, 55(2):1–36, Mar. 2023. Article No. 25.

[4] Zhao-Rong Lai, Pei-Yi Yang, Liangda Fang, and Xiaotian Wu. Short-term sparse portfolio optimization based on alternating direction method of multipliers. Journal of Machine Learning Research, 19(63):1–28, Oct. 2018.

[5] Harry M. Markowitz. Portfolio selection. Journal of Finance, 7(1):77–91, Mar. 1952.

[6] Charles A Micchelli, Lixin Shen, and Yuesheng Xu. Proximity algorithms for image models: denoising. Inverse Problems, 27(4):045009, 2011.