Mean Variance Optimization at Large Scale using R and RMosek

Zhongnan Fang

March 27, 2017

Abstract

When we need to do portfolio optimization with large number of stocks, the classical mean variance optimization will become very slow. In this report, I try to modify the classical mean-variance optimization problem into conic quadratic optimization problem, which is suitable for Mosek optimizer. The final result shows that, the two year monthly rebalance optimization for 1695 stocks costs 5.6 seconds on average. Considering the data cleaning process, the total calculation time is around 9.45 seconds.

1 Factorization of the Covariance

Since conic quodratic formulations are well-suited structure for MOSEK optimizer. In this section, I will derive the conic formulations of Markowitz portfolio selection, which is the factoriztion of the covariance Σ_r. The covariance matrix can be factored as

$$\bar{\Sigma_r} = G^T G \tag{1}$$

where the G matrix depends on the choice of factorization. For any such factorization, we can then express the risk as a Euclidean norm, i.e.,

$$(\omega^0 + x)^T \bar{\Sigma_r}(\omega^0 + x) = (\omega^0 + x)^T G^T G(\omega^0 + x) = \|G(\omega^0 + x)\|, \tag{2}$$

which is used in the conic formulations in the following sections.
The sample estimate is defined as a factorization (or a Grammian matrix)

$$\bar{\Sigma_r} \triangleq X^T X \tag{3}$$

where

$$\bar{X} \triangleq \frac{1}{\sqrt{N-1}}(X - e\bar{r}^T) \tag{4}$$

denotes a scaled zero-mean version of the data-matrix, and e is a vector of all ones. We can therefore express the risk directly in terms of the (normalized) data-matrix $\bar{X}$ as

$$(\omega^0 + x)^T \bar{\Sigma_r}(\omega^0 + x) = (\omega^0 + x)^T X^T X(\omega^0 + x) = \|X(\omega^0 + x)\|, \tag{5}$$

The factored risk expression in (5) is the one most frequently used in Markowitz portfolio optimization, and makes no assumptions about the dimensions or the rank of $X - it$ can be employed independent of whether we have more observations than assets (i.e., whether X has more rows than columns). Moreover, it is not required that X has full rank.

2 Mean Variance Optimization under Conic Quadratic Optimization

Given the traditional Markowitz portfolio optimization problem, where we minimize the risk (in this case by minimizing the square root of the variance) given a fixed number of expected return. The optimization structure is like,

$$\begin{aligned} &minimize \quad \|RX(\omega^0 + x)\| && (6)\\ &subject \quad to \quad \bar{r}^T(\omega^0 + x) = t, && (7)\\ &e^T x = 0 && (8) \end{aligned}$$

The additional constraint $e^T x = 0$ is the self-financing constraint, states that the total amount of reinvestment at each month should be zero.
Then we try to write the (6) (7) (8) as the conic quadratic problem. First, we use $g(x)$ as our objective function. Then, we

can rewrite the problem as:

$$\begin{aligned} & minimize \quad f && (9) \\ & subject \quad to \quad \bar{r}^T(\omega^0 + x) = t && (10) \\ & \|RX(\omega^0 + x)\| \leq f && (11) \\ & e^T x = 0 && (12) \end{aligned}$$

The problem can finally be formulated as a standard conic quadratic optimization problem suited for MOSEK by including an additional equality constraint as

$$\begin{aligned} & minimize \quad f && (13) \\ & subject \quad to \quad \bar{r}^T(\omega^0 + x) = t && (14) \\ & R(\omega^0 + x) = g && (15) \\ & f \geq \|g\| && (16) \\ & e^T x = 0 && (17) \end{aligned}$$

Besides this, I also include the constraint that allows shorting. Since it is not very practical to find 3000 high liquid stocks with good data quality, I download the major stock daily data from Capital IQ. The total number of stocks is around 1800 and I finally get 1695 stocks after cleaning the data.

By using the Mosek optimizer, I can find the significant different on speed compared to classical optimization packages. Also, since we decompose the original covariance matrix, we can do faster calculation under conic quadratic optimization problem.

The expected outputs contain two parts, first one is the rebalancing weight x, which is the optimized weight changes on each stock. The second one is the minimized risk at that time, which is the smallest risk we can get given the expected return at that month. In this report, I didn't consider the rounding issue and transaction cost, which cannot be ignored in the practice. We may dig out more details about that later.

The final result shows that, on average, the typical mean variance optimization can be finished within 10 seconds, for two year 1695 monthly rebalance portfolio. I believe that, if we extend the time window to more than two years and increase the number of stocks, we can still get the optimized answer in a short period of time.