Disclaimer: The author in no way claims that the method presented in this paper is optimal, or even correct. Instead, the purpose of the paper is to explain mathematical aspects being the foundation of today's mathematical finance.

Fredrik Lundin
✉ fredriklundin50@gmail.com

Investment strategy - a mathematical introduction
Part I - The problem formulation

1 Introduction

When considering what investments to do, one might tend to read about the companies, read about the industry, look at historic performance and try to get a general feeling for how wise it would be to invest in that company. While that could be a sufficient approach, in this paper, I present some of the key metrics that could simplify our decision making. The purpose is to explain mathematical fundementals which could simplify decision making of investments - not only at the stock market, the concepts apply also for other type of investments.

2 Considerations

When studying the historic performance, one might look at how the *price* of the stock developed over time. The price should be related to the *profits* or *margin* experienced during these years. That is, if we see an ever increasing stock price but the profit or margin stays flat each year, there could be reasons to be careful.

In this paper, we will assume that market automatically corrects the price if the profit is not sufficiently large. Therefore, we will only look at the *price* of the stock. Also, a stock typically pays out dividends, which are to be considered in the overall investment strategy optimisation.

So, this far, we see that we want a stock with increasing price, and large dividends. The market adjusts the price in relation to the profit. However, we need to adjust the price so that we are always measuring the development in relative terms (%) and not absolute terms ($).

Manually, by studying individual graphs, we are capable of finding stocks with increasing prices, and high dividends. However, by such manual study, it is difficult getting an overview of how many stocks relate to each other. This is an important aspect to consider. It is often done by finding the *covariance* between stocks, that is, how they vary with each other. Note that the covariance does not tell us anything about the price, meaning the following two scenarios can yield exactly the same "good" covariance:

1. Two stocks steadily increase year after year.
2. Two stocks steadily decrease

year after year.

Therefore, we need a method to incorporate the price in our consideration. While it is valuable to have a low variance, we still want to have a good price development. In addition, we add the dividends as an extra weight to counter for ending up in the abovementioned second scenario.

3 The mathematics

We consider N number of stocks - $S_0, S_1, ..., S_N$. Each stock has a time series of prices $p_{n,t}$, where n is the index of the stock and t is the time. We define the *rate of return* of an investment as the difference in price between two consecutive observations.

Mathematically, this is defined as $r_n(t) = \big(p_{t,n+1} - p_{t,n}\big)/p_{t,n}$, where the number 1 is the difference in time between the two observations. Given the data used in this report, the number 1 represents 1 *day*. The *rate of return* is now our method to measure the price development over time, adjusted for the price of the stock. Therefore, it measures the price development in relative terms.

We define the average rate of return of stock S_n, as:

$$\mathbb{E}_t[r_n(t)] = \frac{1}{T}\sum_{t=0}^{T-1} r_n(t),$$

where $t = 0$ is the first time of the time series, $T-1$ is the last time of the time series.

We define covariance of the rate of return between two stocks, for example S_0 and S_1, as

$$C_{\mathbf{r}_0,\mathbf{r}_1} = \mathbb{E}_t\big[(\mathbf{r}_0 - \mathbb{E}[\mathbf{r}_0])(\mathbf{r}_1 - E[\mathbf{r}_1])\big].$$

One could see that if we want to calculate the covariance between $\mathbf{r}_0$ and itself, we get

$$C(\mathbf{r}_0,\mathbf{r}_0) = \mathbb{E}_t\big[(\mathbf{r}_0 - \mathbb{E}[\mathbf{r}_0])(\mathbf{r}_0 - \mathbb{E}[\mathbf{r}_0])\big] = \mathbb{E}\big[(\mathbf{r}_0 - E[\mathbf{r}_0])^2\big], \quad (1)$$

which is also know as the *variance*. We want to calculate the covariance between all stocks. That is, how does $\mathbf{r}_0$ vary with all the stocks, including itself - $\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{N-1}$? Then we continue and see how $\mathbf{r}_1$ varies with all the other stocks, and continues the procedure for all stocks up until S_{N-1}. Using equation (1), we can form the following matrix of covariance values

$$\mathbf{C} = \begin{bmatrix} C_{r_0,r_0} & C_{r_0,r_1} & \dots & C_{r_0,r_{N-1}} \\ C_{r_1,r_0} & C_{r_1,r_1} & \dots & C_{r_1,r_{N-1}} \\ \vdots & \vdots & \ddots & \vdots \\ C_{r_{N-1},r_0} & C_{r_{N-1},r_1} & \dots & C_{r_{N-1},r_{N-1}} \end{bmatrix}.$$

Mathematically, we define our investment through the use of a vector. The elements of that vector is the ratio of the total number of capital being invested. That is, we have the following definition of our placements:

$$\mathbf{w} = \begin{bmatrix} w_0 & w_1 & \dots & w_{N-1} \end{bmatrix}, \quad (2)$$

where $w_n, n = 0, 1, ..., N-1$ is the *weight* put to stock S_n. In total, all weights add up to 1, since they are ratios of the total capital being invested. Therefore, the capital placed at stock S_n is given by $w_n I$, where I is the total amount of capital being invested.

For example, it can turn out that it is optimal to put all capital to one single stock. That stock would then get the weight $w_n = 1$, and subsequently $100\,\%$ of the total capital goes to that stock.

Having defined the basic mathematics being used, it is now time to define the optimisation problem. We want to invest in stocks having the following properties:

1. The average **rate of return**, adjusted for the stocks' prices, is as high as possible.
2. The **dividends**, adjusted for the stocks' prices, is as high as possible.
3. The **covariance** between the selected stocks is as low as possible.

When adjusting a metric for the price of a stock, we simply divide with the stock's price today. In an attempt to combine the three wishes in the above list, we can define the following equation:

$$\min_{\mathbf{w}} \overbrace{\sum_{n=0}^{N-1} \frac{1}{T} \sum_{t=0}^{T-1} r_n(t)}^{\text{Rate of return}} + \overbrace{\sum_{n=0}^{N-1} \sum_{t=0}^{T-1} \frac{d_n}{p_{n,t=0}}}^{\text{Dividends}} - \overbrace{\mathbf{w}^\intercal \mathbf{C} \mathbf{w}}^{\text{To be explained}} \quad . \tag{3}$$

Note that the expression for the dividends only relates the dividends to the price at time $t = 0$, since that is the date the investment was assumed to be performed. Also, note that we summarise all the stocks' dividends at all times (typically several years). Equation (3) can be rewritten in a slightly more compact form, namely:

$$\min_{\mathbf{w}} \sum_{n=0}^{N-1} \mathbb{E}[\mathbf{r}_n] + \frac{d_n}{p_{n,t=0}} - \mathbf{w}^\intercal \mathbf{C} \mathbf{w}. \tag{4}$$

Now, the covariance matrix $\mathbf{C}$ somehow is represented in the equation as a *number*, since all the summations represent *numbers*, and since the end result of the minimisation is a *number*. While this is true, it is not enough to explain why we define that *number* as $\mathbf{w}^T \mathbf{C} \mathbf{w}$. Instead, we want to think of it as follows.

First of all, we want to try different vectors $\mathbf{w}$, and find the one which gives the minimum in equation (4). The covariane matrix is known at this stage, and represents the variance between stocks. Given the current placement vector, i.e. the current $\mathbf{w}$, we want to emphasise good relationships between stocks, and deemphasise bad relationships between stocks. This is given by the term $\mathbf{Cw}$, which represents a transformation/altering of the weights to account for the covariance of the underlying rate of returns. This operation will either end up in one of these scenarios:

1. It will not do very much, meaning the current choice of placement is wise, given the covariance of the stocks' rate of returns.

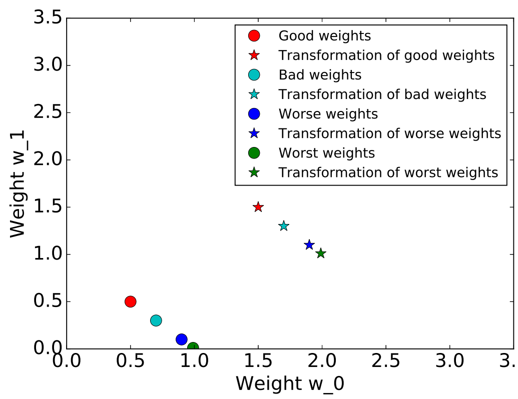
2. It will change the placement vector much, meaning the current choice of placement is bad, given the covariance of the stocks' rate of returns.

However, we do not have any way of measuring either of scenario 1. or 2. This is where the term $\mathbf{w}^\intercal$ comes into play. Loosely speaking, an equation on the form $\mathbf{w}^\intercal(\mathbf{Cw})$ measures how similar $\mathbf{w}$ and $\mathbf{Cw}$ are. More accurately, it expresses how the current weights $\mathbf{w}$ would amplify the transformed weights $\mathbf{Cw}$, in various directions. Therefore, we would like to think of $\mathbf{w}^\intercal\mathbf{Cw}$ in the following way:

- If we first let the covariance matrix transform the weights to make them worse because of the covariance. Then we see if the current $\mathbf{w}$ would further amplify that transformation.

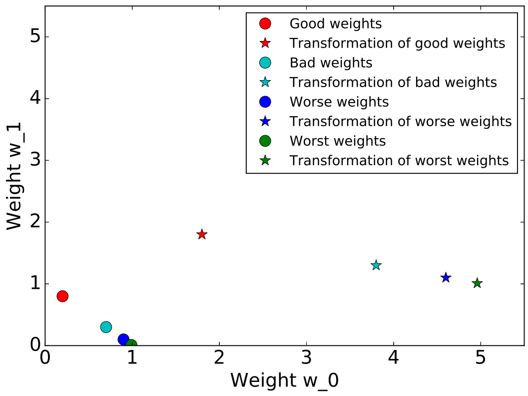
We want the transformation, defined by the covariance matrix, to make as small impact as possible on our weights. If we can not alternate the weights in any way making the transformation have less impact, it means that the current weights are optimal from a covariance perspective. This is achieved by placing our weights so that large covariances get small weights, and vice versa.

If we have a covariance matrix $\mathbf{C} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, then the transformation $\mathbf{Cw}$ would look like the following.



Here a selection of equal weights, i.e. $w_0 = w_1 = 0.5$ would be optimal. However, it is not very obious. The following figure has a covariance matrix $\mathbf{C} = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix}$, and so it really tries to push the weights in the direction of the largest variance (5). The bad selection of original weights, where $w_0 \approx 1$

would then amplify the transformation $\mathbf{Cw}$, which is exactly what is not wanted.



If that is the case, our current placements are only multiplied by a number when we try to transform it by the covariance matrix. That is the worst possible placement given the covariance matrix. We are not interested in exactly how much the transformation changes the weights. Instead, it is enough to see if the current weights are the choice being impacted the least of all possible choices of weights.

Using the weights defined in equation (), we can get the following equation typically found in literature:

$$\min_{\mathbf{w}} f(\mathbf{w}) = \overbrace{-\left(\sum_{n=0}^{N-1} w_n \mathbb{E}[\mathbf{r}_n]\right)}^{=-\mathbf{w}^\intercal \mu} + \mathbf{w}^\intercal \mathbf{C}\mathbf{w}. \tag{5}$$

We see that we want the average return to be as large as possible, while at the same time balancing so that the covariance does not become too large. Also, we see that there are two extreme cases - one is that of maximising the expected return, and the other is that of minimising the variance.

Now, we should introduce some constraints. We require that the weights in total are equal to 1, that is, $\sum_{n=0}^{N-1} w_n = 1$. Otherwise, they can not represent ratios of our total capital being invested. Mathematically, this can be written as $\mathbf{w}^\intercal\mathbf{1}$. Also, in optimisation, it is convenient to always have the right hand side of the equation being equal to 0. Therefore, we have $\mathbf{w}^\intercal\mathbf{1} - 1 = 0$.

We also would like each weight to be non-negative, that is, $w_n \geq 0$, for each $n \in [0, N-1]$. A negative weight would correspond to having a short position in a stock, which we will not accept in our model. For minimisation, it is convenient to work with $\leq$ conditions, meaning we get $-w_n \leq 0, \forall n$. In vector form this can be written as $-\mathbf{w} \leq \mathbf{0}$. Note that we had to put a minus sign in front of the weight to change from the $\geq$ sign to the $\leq$ sign.

One could also imagine having a requirement that the value of the portfolio at the end time is larger than at the start time. However, one could look at graphs later and see that there is positive expected return at an annual basis, and so it will be studied manually instead. Finally, we have the model expressed as the optimisation problem to

$$\begin{aligned} \min_{\mathbf{w}} \quad & -\mathbf{w}^\intercal \mu + \mathbf{wCw} && \text{(6a)}\\ \text{subject to} \quad & \mathbf{w}^\intercal\mathbf{1} - 1 = 0, && \text{(6b)}\\ & -\mathbf{w} \leq \mathbf{0}. && \text{(6c)} \end{aligned}$$

4 Pause and reflection

I would argue that the most important part is to understand how to formulate optimisation problem, not to actually solve them. If one is able to express the problem in the right way, there are typically several software solvers which can efficiently solve the problem. To solve the problem can be automated and performed much better than what many persons can implement themselves. Therefore, the takeaway of this paper should be mainly how to express ideas of various types of problems in a way so that software can then take over. Nevertheless, I here present and implement a way of solving equation (6a) - (6c). The purpose of doing so, is to provide extra understanding of what really happens when we alternate our investments, that is, $\mathbf{w}$.

4.1 Some words on the data

The data used in the paper is the *price*, which was changed into a *rate of return* - $r_n(t)$. When owning a stock for multiple periods, one could say that the capital is re-invested at each period (without cost). Consider the following example.

Day	Rate of return	Price
1	0 (start)	5
2	50	$(1+0.5)\cdot 5 = 7.5$
3	-50	$(1+0.5)\cdot 7.5 = 3.75$
	Average: 0	

The price after re-investments ("compounding") is then expressed as

$$5\cdot(1.0+0.0)\cdot(1.0+0.5)\cdot(1.0-0.5) = 5\cdot 1\cdot 1.5\cdot 0.5 = 3.75. \quad (7)$$

Therefore, the average return is $(3.75-5.00)/5.00 = -0.25 = -25\,\%$. This is a large difference to the average of 0 seen in the table. The correct way of handling the data, is to use it in a multiplicative fashion. However, calculations then become more difficult, and so we will use the data in an additive fashion, which was here shown to be *incorrect*. This seems to be an area in itself to study. Perhaps - if the changes in rate of return are much smaller than $50\,\%$, it will be acceptable to use an additive approach, since:

$$(1+x)(1+y) \approx 1\cdot\Big(1+\overbrace{\frac{x+y}{2}}^{\text{Mean value}}\Big),$$

if both x and y are small. Given the data, we will see that this is in fact true, and so x and y will be small numbers varying around 0. For example, we consider the case when x and y are $3\,\%$, which are quite large daily rate of returns. Then, we have:

$$(1+0.03)\cdot(1+0.03) = 1.06091,$$

which is approximately equal to $1 \cdot \left(1+\frac{0.03+0.03}{2}\right) = 1.03$. The error is

$$(1.0609 - 1.03)/1.0609 = 0.029 = 2.9\,\%,$$

compared to

$$(3.75 - 5.00)/3.75 \approx -0.33 = -33\,\%$$

which was previously seen.

Now, we know that the way of handling data is not correct. Therefore, we have to assume that the rate of returns are small so that the error between our way and the correct way can be a bit smaller than more extreme cases where it can be really wrong.

5 Solving the problem

One way of solving equation (6a), is to randomly generate many different placement vectors, $\mathbf{w}$, which would fulfill equation (6b) and (6c). Then we can measure the value of the function. If we repeat this many time, and always store the currently best value of $f(\mathbf{w})$, we could get *some sort of solution*. However, it is very inefficient and especially when we want to consider many different stocks, that is, if $\mathbf{w}$ has many dimensions.

Unfortunately, the constraints (6b) and (6c) make a potential solution to the problem to not be so stragithforward. Therefore, we could for example transform the problem (6a) - (6c) to something that is much easier to handle. If we can get rid of the constraints, and at the same time ensure that the problem is *convex*, everything becomes much easier. One way of doing this, is to integrate the constraints to the cost function. When doing so, we would like to put a large penalty/cost to the constraints not being fulfilled.

We define the equality constraint in equation (6b) as $h(\mathbf{w}) = \mathbf{w}^{\intercal} - 1 = 0$. We also define the inequality constraints (6c), as a series of functions $g_n(\mathbf{w}) = -w_n,\, n \in [0, N-1]$. Furthermore, we define a penalty function $\psi(-)$.

If the equality constraint (6b) is not fulfilled, we want $\psi(h(\mathbf{w}))$ to be large. Also, the further away it is from being fulfilled, we want the value to be even larger. One way of achieving this, is by using a quadratic penalty function $\psi(x) = x^2$.

Similar logic applies for the inequality constraint (6c). However, here we only want to penalise values being positive, that is, $g_i(\mathbf{w}) = -w_n > 0$. This can be expressed as $\max(0, g_i(\mathbf{w}))$.

If $g_i(\mathbf{w}) = -w_n \leq 0$, everything is fine, and we instead select 0, giving $\psi(0) = 0^2 = 0$ as penalty. However, if $-w_n > 0$, we have a quadratic penalty $\psi(g_i(\mathbf{w})) = (-w_n)^2$.

By doing this, we can now select *any* values for $\mathbf{w}$ but bad selections will be penalised much harder than good selections. What is important, is that twe have gone from a constrained problem to an unconstrained. Now, the problem is that to

$$\min_{\mathbf{w}} f(\mathbf{w}) = -\mathbf{w}^\intercal \mu + \mathbf{w}^\intercal \mathbf{C}\mathbf{w} + \quad (8)$$

$$v\Big(h^2(\mathbf{w}) + \sum_{n=0}^{N-1} \big(\max(0, -w_n)\big)^2\Big), \quad (9)$$

where v was introduced to give even more penalty to solutions $\mathbf{w}$ not fulfilling (6b) and (6c).

Equation (8) is unconstrained and easier to solve, and we will use a gradient descent approach. This means that we want to see at the current guess $\mathbf{w}_{\text{guess}}$, in which direction the cost $f(\mathbf{w}_{\text{guess}})$ grows. Then we go in the opposite direction of that growth, in order to reduce the cost. Generally speaking, this is going to provide the optimal solution after a while, given that the cost function is *convex*.