Applications of Convex Optimization in Quantitative Trading and Portfolio Management

A project report

submitted by

Dhyey Dharmendrakumar Mavani

(Computer Science, Mathematics and Statistics major in Amherst College Class of 2025)

under the mentorship of

Dr. Tanya Leise

in partial fulfillment of
the requirements for the course credits of MATH-294: Optimization towards the degree of

Bachelor of Arts (Mathematics)

to the

Department of Mathematics
Amherst College, Amherst, MA

December 2022

ACKNOWLEDGEMENT

As a matter of first importance, I offer my genuine thanks to my supervisor Dr. Tanya Leise, Brian E. Boyle Professor in Mathematics and Computer Science, Amherst College, Amherst, MA. I appreciate her support and help during the project work while teaching the course MATH-294: Optimization during Fall 2022.

Dhyey Mavani

Contents

Chapter 1

Introduction

1.1 Overview and Outline

In this report, I will try to explore the process of finding trades in the markets through convex optimization problems which trades-off expected return, risk, transaction cost and holding cost for example. I will not focus on the α finding and making predictions in the markets, but I would rather focus on how can we evaluate the performance of a trading strategy and methods used to exploit the predictions at hand to the furthest extent. Furthermore, at the end, I will interpret some results of experiments done by authors of "Multi-Period Trading via Convex Optimization" [2] while using their Open Source Python Package named CVXPortfolio [1] for further analysis.

1.2 Historical Background

In 1952, Markowitz first formulated the portfolio selection idea in terms of a convex optimization problem with just the trade-off of risk and return disregarding the cost of trading on an exchange or the like over one or more periods. Later on, Goldsmith re-formulated the ideas to extend them in the world of varying transaction costs in the single-period setting. Then, towards the rise of multi-period trading ideas, the problem boiled down to choosing a sequence of trades to carry out over a set of discrete periods as introduced first by Samuelson and Merton. This work was later extended by many others to continuous time. The work in multi-period trading is based on dynamic programming approach of computer science which serves us well as it updates the beliefs of the system by taking into the consideration the information of trades in a sequential fashion. Because of the huge computational overhead of dynamic programming, and hence a lot of literature is only focused on simplified constraints. For practical purposes, thus it is viable to trade off some accuracy for the ability to get approximate results quickly. Due to advent in technology, the quantitative trading and portfolio management industry started focusing on imposing more constraints to better approximate the parameters involved unlike 2000s when our best bets were separate simplified unconstrained optimizers.

Chapter 2

Preliminaries

In this chapter, I will try to concisely define the basic dynamic model and associated metrics of multi-period trading as proposed by authors of [2] which tells us how a portfolio holdings and associated liquid cash account changes over time due to factors such as trading, investment returns, and transaction & holding costs.

2.1 Asset and Cash Holdings of the Portfolio

Portfolio

Portfolio is defined as a ledger of n assets with an associated liquid cash account over a finite time horizon over 252 trading days with discrete time periods labeled $t = 1, ..., T$. We define a $n+1$ column-vector of holdings as h_t, where $(h_t)_i$ defines, at the beginning of a time period t, the dollar value of asset i in our portfolio if $1 \leq i \leq n$, and the portfolio's liquid cash balance in dollars if $i = n+1$.

Total value, Gross Exposure and Leverage

Total value (or net asset value, NAV) or $\nu_t = \mathbf{1}^{\boldsymbol{T}} h_t$ (assumed to be ≥ 0 here).
Gross Exposure or $||(h_t)_{1:n}|| = |(h_t)_1| + ... + |(h_t)_n|$, which is sum of your asset positions. Leverage of a portfolio is defined as gross exposure divided by value or $\frac{||(h_t)_{1:n}||}{\nu_t}$.

Weights

Weights is a $n+1$ column-vector defined as $w_t = h_t/\nu_t$. Note that leverage can be defined as the $l1$-norm of assets weights (or $||w_{1:n}||_1$)

2.2 Trades

Trade vector

Trade vector u_t is a $n+1$ column vector representing dollar value of trades, where $(u_t)_i > 0$ means we buy asset i and $(u_t)_i < 0$ means we sell asset i if $1 \leq i \leq n$, but if $i = n+1$, it means we put or take-out money from the cash account respectively. Furthermore, let $z_t = u_t/\nu_t$ gives trades normalized by total value.

Post-trade portfolio

Post-trade portfolio's holding vector can be defined as $h_t^+ = h_t + u_t$ and post-trade portfolio normalized by the portfolio value in terms of weights and normalized trades can be represented by the equation $h_t^+/\nu_t = w_t + z_t$

2.3 Transaction and Holding cost

Transaction cost

We define a *separable* transaction cost function as $\phi_t^{trade}(u_t) : \boldsymbol{R}^{n+1} \to \boldsymbol{R}$. Some reasonable transaction cost models are described in detail in [2].

Holding cost

Holding cost associated to holding a post-trade portfolio h_t^+ over t^{th} period is given by $\phi_t^{hold}(h_t^+)$, where the holding cost function is defined as $\phi_t^{hold} : \boldsymbol{R}^{n+1} \to \boldsymbol{R}$. Some reasonable holding cost models are briefly described in [2].

2.4 Self-Financing Condition

This condition states that no external cash is put into or taken out of the portfolio from beginning to the end of a period. This can be mathematically expressed as

$$\mathbf{1}^T u_t + \phi_t^{trade}(u_t) + \phi_t^{hold}(h_t^+) = 0$$

This logically translates to us connecting the cash trade amount to asset trades through relation given by

$$(u_t)_{n+1} = -(\mathbf{1}^T (u_t)_{1:n} + \phi_t^{trade}((h_t + u_t)_{1:n}) + \phi_t^{hold}((u_t)_{1:n}))$$

Normalized self-financing

By rearranging the equations and previous results, the self-financing condition can be represented in terms of weights and normalized trades as

$$\mathbf{1}^T z_t + \phi_t^{trade}(z_t) + \phi_t^{hold}(w_t + z_t) = 0$$

Moreover, we can express the normalized cash trade value in terms of the normalized non-cash asset trade values through relation given by

$$(z_t)_{n+1} = -(\mathbf{1}^T (z_t)_{1:n} + \phi_t^{trade}((w_t + z_t)_{1:n}) + \phi_t^{hold}((z_t)_{1:n}))$$

2.5 Investment

We say that the post-trade portfolio (including cash) are invested for one period, until the start of the next time period. The portfolio at next time period is given by $h_{t+1} = (\mathbf{1} + r_t) \circ h_t^+$, where r_t is a $(n+1)$ column vector of return rate which can be linear or logarithmic for example and $\circ$ denotes element-wise multiplication. It logically follows that we define the next period portfolio value as $\nu_{t+1} = \nu_t + r_t^T h_t + r_t^T u_t - \phi_t^{trade}(u_t) - \phi_t^{hold}(h_t^+)$, hence portfolio realized return is defined as

$$R_t^p = \frac{\nu_{t+1} - \nu_t}{\nu_t} = r_t^T w_t + r_t^T z_t - \phi_t^{trade}(z_t) - \phi_t^{hold}(w_t + z_t)$$

This makes sense as it can be interpreted as the portfolio return over period t equals portfolio returns without trades or holding cost + return on the trades - transaction cost - holding cost. Further algebraic rearrangement gives us that the next-period weights are given by $w_{t+1} = \frac{1}{1+R_t^p}(\mathbf{1} + r_t) \circ (w_t + z_t)$.

2.6 Simulation and Aspects not modeled

With every simplified model comes the limitations in consideration, so ours is none different. We are ignoring some aspects of real trading including External cash, Dividends, Non-instant trading, Imperfect execution, Multi-period Price impact, Trade settlement. Mergers/Acquisitions, Bankruptcies/Dissolution and Trading freezes. We will do simulations to see the evolution of portfolio over the various time periods. Our simulations will require Starting portfolio values, asset trade vectors, transaction cost model parameters, shorting rates and various types of return rates as inputs. We will focus on testing how we would have performed in the past (Back-tests) and in the time of extreme situations (what-if simulations) with added uncertainty in the simulations.

2.7 Performance Metrics

2.7.1 Absolute metrics

Return and Growth Rate

In the case of linear returns we consider Return rate, which is defined as time-averaged realized returns, formally $\overline{R^p} = \frac{1}{T}\sum_{t=1}^{T} R_t^p$. In the case of log-returns we consider the Growth rate, which is defined as $G_t^p = log(\nu_{t+1}/\nu_t) = log(1 + R_t^p)$

Volatility and Risk

The realized volatility is defined as the standard deviation of the portfolio return time series, formally $\sigma^p = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_t^p - \overline{R^p})^2}$. Furthermore, $(\sigma^p))^2$ is often regarded as the quadratic risk measure.

2.7.2 Relative Metrics

Benchmark weights

Benchmark weights are a set of weights against which we compare the weight vector of our portfolio at any given time. For example, in the risk-free cash the benchmark vector is a (n+1) column vector with n 0's and the last entry being 1.

Active and excess return

Active return with respect to a benchmark is defined as $R_t^a = R_t^p - R_t^b$, when we have the case of the benchmark being risk-free cash we have excess return defined as $R_t^e = R_t^p - (r_t)_{n+1}$

Active and excess risk

Active risk (σ^a) with respect to a benchmark is defined as standard deviation of R_t^a, when we have the case of the benchmark being risk-free cash we have excess risk (σ^e) defined in a similar fashion.

Information and Sharpe ratio

Information ratio relative to a benchmark is average of active returns divided by the standard deviation of active returns, formally $IR = \overline{R^a}/\sigma^a$, when we have the special case of the risk-free cash benchmark, we have Sharpe ratio defined as $SR = \overline{R^e}/\sigma^e$

Chapter 3

Single-Period Trading

3.1 Risk-Return Optimization Trading Strategy

Trading strategy of determining z_t can be formulated as a convex optimization problem as follows (as per the idea introduced by Markowitz).

maximize $\hat{R_t^p} - \gamma_t \psi_t(w_t + z_t)$ OR (substituting $\hat{R_t^p}$) $\hat{r_t}^T z_t - \hat{\phi_t}^{trade}(z_t) - \hat{\phi_t}^{hold}(w_t + z_t) - \gamma_t \psi_t(w_t + z_t)$

$s.t.$ $\mathbf{1}^T z_t + \phi_t^{trade}(z_t) + \phi_t^{hold}(w_t + z_t) = 0$ OR $\mathbf{1}^T z_t = 0$ (approximated self-financing constraint)

$\psi_t, z_t, (z_t + w_t), \phi_t^{trade}(z_t)$ and $\phi_t^{hold}(w_t + z_t)$ are Convex or belong to Convex sets

where γ_t is the risk aversion parameter (> 0) and ϕ_t is the risk function.

Risk Measures

There are various kinds of risk measures to factor in the risk in returns. The first one is absolute risk, which is defined in accordance with the $((n+1) \times (n+1))$ variance-covariance matrix Σ_t of stochastic returns r_t also called the traditional quadratic risk measure given as $\psi_t(x) = x^T \Sigma_t x$. As discussed before, we have an analog called active risk which is the w_t^b shifted weight version of absolute risk, formally $\psi_t(x) = (x - w_t^b)^T \Sigma_t (x - w_t^b)$. Moreover, the risk aversion parameter (γ_t) can be loosely considered as the measure of relative importance of risk over return. γ_t is a completely different study on its own, one of the many famous studies in this regard is done by J. Kelly [4] which lead to Kelly's optimal portfolio for log-normal returns where the parameter is estimated to be equal to 1/2, when the objective is to maximize the expected portfolio growth rate. Finally, there are risk models like Factor model, Transformed Risk model and worst-case quadratic risk model which try to take into account different aspects of risk modeling which is an extensive subject of its own.

Forecast Error Risk

Since we are using forecasts in our previously framed optimization problem we will have unavoidable error in forecasts in terms of returns and covariance of returns (the first and second order moments which are being used). Given the uncertainty column vector of ρ in our predictions the return forecast error risk can be represented as $\psi_t(x) = \rho^T|x - w_t^b|$. On the other hand, covariance forecast error risk is the risk associated with perturbations of covariance matrix (Δ) in our predictions. Perturbation model can be represented by a parameter κ in the sense that the non-diagonal entries (or asset correlations) of Σ_t can change by the fraction κ of the respective diagonal entries, formally $|\Delta_{ij}| \leq \kappa(\Sigma_{ii}\Sigma_{jj})^{1/2}$. This further using the idea of worst-case quadratic risk (described more formally in [2]) leads to covariance forecast error risk $\psi_t = (x - w_t^b)^T\Sigma_t(x - w_t^b) + \kappa(\sigma^T|x - w_t^b|)^2$, where $\sigma = (\Sigma_1 1^{1/2}, ..., \Sigma_{nn}^{1/2})$ is the vector of asset volatilities.

Other Constraints

There are two major types of constraints other than the Self-Financing Constraint, namely Holding and Trading constraints. Trading constraints are the conditions that directly regulate the trading activity of assets for example the Turnover limit, limits relative to the trading volume, and specific restrictions like no-buy/no-sell/no-trade. On the other hand, Holding constraints are the conditions that directly regulate the holding of assets in the portfolio for example the long-only portfolio condition, leverage restrictions, limits on asset capitalization, limits on minimum liquid cash balance, no-hold constraint, liquidation loss constraint, concentration limit and stress constraints (restricts the probability of huge losses in the periods of shocks). The constraint which is needed to make our model efficiently solvable comes from the groundwork done by authors of [2], which is the Convexity of functions and sets in hand. We might be tempted to add some obvious-looking constraints to our model, but we should not directly do so as it might tamper the convexity of the optimization problem and hence reducing the efficiency of run-time, instead as the authors of [2] say, we should try to find a combination of alternative convex constraints that achieves the same effect as addition of the non-convex constraint we were tempted to add before.

Please note that in real-world we have priorities among the constraints and in the above framework of optimization model, it is quite easy to make the constraints "soft" by assigning them priority values so that our model only violates a constraint when its necessary in the grand scheme of things.

SPO Model to Exploit the given Forecasts

Now we can write down the SPO model in form of a convex optimization problem as follows...

maximize $\hat{r_t}^T z_t - \gamma_t^{trade}\hat{\phi_t}^{trade}(z_t) - \gamma_t^{hold}\hat{\phi_t}^{hold}(w_t + z_t) - \gamma_t^{risk}\psi_t(w_t + z_t)$

s.t. $\mathbf{1}^T z_t = 0$ (approximated self-financing constraint)

$\psi_t, z_t, (z_t + w_t), \phi_t^{trade}(z_t)$ and $\phi_t^{hold}(w_t + z_t)$ are Convex or belong to Convex sets

where γ_t^{risk} is the risk aversion parameter (> 0) and ϕ_t is the risk function.

Here we introduced the idea of factors such as γ_t^{trade} and γ_t^{hold} which can be treated as hyperparameters or more simply the knobs/dials to turn through backtesting so that we can better model the data generating process (which is the ultimate goal of all unknown DGP models).

Chapter 4

Multi-Period Optimization

4.1 Multi-Period optimization Trading Strategy

Motivation

As we already saw, SPO problem is already difficult to solve in many situations, but it has its limitations in terms of what types of effects it can consider. Our motive behind exploring the MPO technique of making trading decisions is to double down on the idea that the greedy strategy employed by SPO fails if the performance of current period depends on the previous holdings which is very much the case in the real world, which can be solved with Dynamic Programming but is significantly more computationally expensive. The simplest idea to think about MPO here is taking into consideration if the current trades can put us in good/bad positions for future periods. This might be due to the varying trading costs (if it's cheap to trade now, then it makes sense to trade more now and exit out early), signal decay or time -varying return predictions, known future changes in volatility/risk/liquidity (note that the predictability of volume is much better than that of returns), or changing constraints over the multiple periods and setting-up/shutting-down/transfering of a portfolio.

The MPO Problem and its Simplifications

Our MPO problem is solving for a z_t over a planning horizon (H) number of periods into the future, for example t, t+1, t+2, ... , t+H-1. Just for notational clarity for any quantity or function Z, we will represent Z_τ's estimate conditional on the information in the t time periods by $\widehat{Z_{\tau|t}}$. We will go by our natural instinct developed so far by choosing our objective function as the total risk adjusted return over the horizon, which is more formally given by

$$\sum_{\tau=t}^{t+H-1}(\widehat{r_{\tau|t}}^T(w_\tau+z_\tau)-\gamma_\tau\psi_\tau(w_\tau+z_\tau)-\widehat{\phi_\tau}^{hold}(w_\tau+z_\tau)-\widehat{\phi_\tau}^{trade}(z_\tau))$$

We can make this objective solely in terms of w_t and w_{t+1} by using the dynamics equation defined before which tells us how money gets propagated along the periods. This dynamics equation is formally given by

$$w_{t+1} = \frac{1}{1+R_t^p}(1+r_t) \circ (w_t + z_t)$$

Now we will consider adding the risk term $\gamma_t \psi_t(w_t + z_t)$, which makes sense because we assume randomness (independent and identically distributed condition), which gives us that variance of sum is the sum of variances.

As we can see this gets pretty messy pretty quickly, as suggested by authors of [2] we will do a simplification of our dynamics equation. Firstly, it is easy to notice that R_t^p and r_t will be non-significant compared to 1, which means $R_t^p = 0$ and $r_t = 0$ are decent assumptions to make which gives us $w_{t+1} = (w_t + z_t)$. Please note that our self-financing condition can be assumed to be the simplified one meaning $\mathbf{1}^T z_t = 0$ because we are only getting rid of the costs when propagating the portfolio to next periods, but our updated model doesn't undermine itself as it is still considering the effects of inter-period costs which were not captured that well or at all in some situations by the SPO model. Also, replacing everything in terms of z with the terms of w gives us that the self-financing condition can be equivalently expressed as the sum of weights being 1 because the cash and assets are both contained in the different indexes of our vectorized portfolio from 1 to (n+1). This gives us the MPO model in form of a convex optimization problem as follows...

maximize $\sum_{\tau=t+1}^{t+H} (\widehat{r_{\tau|t}}^T w_\tau - \gamma_\tau \psi_\tau w_\tau - \widehat{\phi_\tau}^{hold} w_\tau - \widehat{\phi_\tau}^{trade} (w_\tau - w_{\tau-1}))$
$s.t.$ $\mathbf{1}^T w_\tau = 1$ (approximated self-financing constraint)
$\psi_\tau, w_,(w_\tau - w_{\tau-1}), \phi_\tau^{trade}(w_\tau - w_{\tau-1})$ and $\phi_t^{hold}(w_\tau)$ are Convex or belong to Convex sets
where γ_τ^{risk} is the risk aversion parameter (> 0) and ϕ_τ is the risk function.
here $\tau = t+1, ..., t+H$, where H is the previously defined planning horizon over which we are considering the future impact of our trading decision.

Some further ideas presented are to add terminal constraints such as $w_{t+H} = w^{term} = w_{t+H}^b$, which sets a benchmark. For example, if our terminal/benchmark weight vector stores 0 for all asset weights and 1 for the cash account weight, it means that our new problem will give the solution which will have us liquidate at the end hence saving us from winner's curse positions of high returns, but are expensive to get out of for instance. Furthermore, one can also do multi-scale optimization by adding the conditions that the trade can only occur in certain periods. This will trade off consideration of some opportunities with the span of being able to take into account bigger future period in our MPO.

Chapter 5

Simulations Review and Further Exploration

5.1 Simulations Review

Upon review of the simulations performed by the engineers of the associated Python library named CVXPortfolio, I noticed that our argumentative claim that MPO takes into account the future shocks better than SPO is validated by their results of hyperparameter searches during backtesting on both models and their performance over new data. It was clear that the trading decisions made by MPO model, although more computationally expensive than SPO, gave outcomes of higher return and lower risk, which is ideal for any sane investor even in real-world. As a part of further exploration, I was able to use their library to run some simulations based on already implemented object-oriented methods by them along with the support of Quandl package [5], which allowed me to get access to open-source financial data very efficiently in the form of pandas data frames without the need of much data cleaning.

5.2 Further Work

These days, multi-billion-dollar market-making firms like Jane Street, Akuna Capital and Citadel Securities use advanced stochastic analysis and portfolio optimization theory with more and more intricacies in order to better model the Data Generating Process in the short-term, while on the other hand famous hedge-funds like D. E. Shaw & Co. and Renessaince Technologies use more long-term strategies to trade while missing on the short term opportunities because their models are incredibly complex and require significant time to optimize outcomes. I think it would be interesting to further analyze this trade-off quantitatively in terms of how the strategies and optimization problems differ at core in terms of constraints being imposed.

Bibliography

[1] E. Busseti, S. Diamond, S. Boyd, and Blackrock. CVXPortfolio, 2017. Available at https://github.com/cvxgrp/cvxportfolio.

[2] Boyd, S., Busseti, E., Diamond, S., Kahn, R. N., Koh, K., Nystrup, P., & Speth, J., *Multi-Period Trading via Convex Optimization*, arXiv, https://doi.org/10.48550/arXiv.1705.00109 , (2017).

[3] Calafiore, G.C. and El Ghaoui L., *Optimization Models*, Cambridge University Press , (2014)

[4] J. Kelly, Jr. *A new interpretation of information rate.* IRE Transactions on Information Theory, 1956.

[5] Quandl. WIKI end-of-day data, 2016. Available at https://www.quandl.com/data/WIKI.