15.070

Discrete Probability and Stochastic Processes

Portfolio Optimization with Portfolio Size Constraints

Gabriel Deo, Mouad M'Ghari & Tristan Magnin

Tuesday, May 16

Contents

1 Introduction

1.1 Background

Portfolio optimization is a fundamental problem in investment management, aiming to construct an optimal investment portfolio by balancing the trade-off between expected return and risk. Traditionally, portfolio optimization has been approached as an optimization problem with a linear objective representing expected returns and convex constraints representing risk measures (Markowitz, 1952) [6]. However, in many practical scenarios, it is desirable to incorporate additional constraints to limit the number of stocks included in the portfolio, resulting in a sparse optimization problem.

1.2 Research Questions

This report investigates the impact of sparsity constraints on portfolio optimization and explores two key research questions. First, does the optimization approach find optimal solutions similar to compressive sensing, a field that deals with sparse signal recovery [2]? Second, does the model exhibit any computational hardness akin to the Overlap Gap Property [3]? Finally, we aim to test these ideas using real-life data on mean returns and risks, providing a practical evaluation of the proposed methodology.

By addressing these research questions, we seek to enhance our understanding of portfolio optimization with sparsity constraints and shed light on its computational properties and real-life applicability.

2 Overlap Gap Property

We present in this section the second independent part of the project, which starts to tackle the question of the presence - or not - of the Overlap Gap Property in Sparse Portfolio Optimization.
A similar problem was solved in Gamarnik and Li(2018) [4], they proved that the problem of finding a $r \times r$ submatrix of an $n \times n$ matrix with *iid* Gaussian entries which has a large average value exhibits the *Overlap Gap Property*.
More precisely, the largest average value of a $r \times r$ submatrix is $2(1 + o(1))\sqrt{\log n/r}$, and it was shown that the overlap of submatrices with average value asymptotically $(1 + o(1))\alpha\sqrt{2\log n/r}$ was a *discontinuous* set of $[0,1]^2$ for $\alpha > \alpha^*$ a certain threshold (see 2.1 for more details).
Our research is inspired by their work and aims at generalizing their result to a weighted-average value of a Gaussian Random Matrix.

Most of the results presented in this section are presented without being proved rigorously, we will rather focus on giving a proof sketch and a path for getting to a complete solution. As our research is still in progress at the time of submission, we will precisely point out the parts that require further work to be rigorously proven.

2.1 Portfolio Optimization Problem

Consider a set of d stocks with *iid* Gaussian returns : $\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_d)$.
Here we assume that $\boldsymbol{\mu} = \mathbb{1} \triangleq (1, \ldots, 1)^\intercal$.

Based on N observations of the stocks returns $\mathbf{r}_1, \dots, \mathbf{r}_N$, we define the sample covariance matrix:

$$\hat{\Sigma} \triangleq \frac{1}{N} \sum_{t=1}^{N} (\mathbf{r}_t - \boldsymbol{\mu})(\mathbf{r}_t - \boldsymbol{\mu})^\top$$

We recall the classical mean-variance portfolio optimization problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}^\top \Sigma \mathbf{w} \\ \text{s.t.} \quad & \sum_i w_i = 1 \end{aligned} \tag{1}$$

where $\sigma_\mathrm{p} = \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}$ is the volatility of the portfolio, and $\mu_\mathrm{p} = \sum_i w_i = \sum_i w_i \mu_i$ is both the sum of the weights and the return of the portfolio.

2.2 Main Theorem

We recall here a simplified version of the OGP theorem proven in Gamarnik and Li(2018) [4]. For any set $I \subseteq [d]$, we denote by A_I the squared submatrix of A indexed on I.
Let $\mathcal{O}_r(\alpha, y, \delta)$ be the set of pairs of $r \times r$ submatrices A_I, A_J that have average value within $[(\alpha \pm \delta)\sqrt{2 \log n / r}]$ and such that $|I \cap J|/r \in [y \pm \delta]$, where we used the notation $[a \pm b] \triangleq [a - b, a + b]$.

Theorem 2.1. *Fix* $\alpha \in (1, \sqrt{2})$.

$$\begin{gathered} \forall \varepsilon > 0, y \in (0,1), \exists \delta_0 \in (0, \alpha), n_0 > 0 \quad \textit{s.t.} \quad \textit{for } k = O(\log n) \textit{ and } \delta < \delta_0 : \\ \left| \frac{\log \mathbb{E}[|\mathcal{O}_k(\alpha, y, \delta)|]}{k \log n} - f(\alpha, y) \right| < \varepsilon \quad \textit{asymptotically, when } n \to +\infty \\ \textit{where } f \textit{ is a function that exhibits the Overlap Gap Property on } [0,1]^2 \end{gathered} \tag{2}$$

Our goal is now to prove a similar theorem for an overlap set $\mathcal{O}_r(\alpha, y, \delta)$ adapted to our portfolio optimization problem. It has been shown that the Overlap Gap Property is an indirect indicator of NP-hardness, as it appears in many models exhibiting algorithmic hardness, and it aligns with the transition from tractable to hard problems [3]. Thus, proving that our problem also exhibits the Overlap Gap Property for some values of the portfolio volatility would indirectly suggest a hardness regime in the Sparse Portfolio Optimization Problem.
First, preliminary work on the distribution of the optimal weighted average submatrix is needed.

2.3 Sparse Portfolio Optimization Problem

In the remaining part of this section, we assume the case of an estimated matrix $\hat{\Sigma}$ with a known vector of returns which is an unbiased estimator of Σ. The optimization problem with sparsity constraint rewrites in the following - more manageable - way:

$$\begin{aligned} \min_{I \in \mathcal{P}_r([d])} \min_{\mathbf{w}} \quad & \frac{1}{2}\mathbf{w}_I^\top \hat{\Sigma}_I \mathbf{w}_I \\ \text{s.t.} \quad & \sum_{i \in I} w_i = 1 \end{aligned} \tag{3}$$

Where we define $\mathbf{w}_I$ to be the projection of $\mathbf{w}$ into the set of coordinate I of cardinality r and $\hat{\Sigma}_I$ to be the submatrix of $\hat{\Sigma}$ using the lines and columns in I.

Noticing that we have two imbricated optimization problems, we solve the inside one directly taking into account the constraint. That gives the solution $\mathbf{w}_I^* = \dfrac{\hat{\Sigma}_I^{-1}\mathbb{1}}{\mathbb{1}^\top \hat{\Sigma}_I^{-1}\mathbb{1}}$ and the optimal variance is $v_I^* = \dfrac{1}{\mathbb{1}^\top \hat{\Sigma}_I^{-1}\mathbb{1}}$. The distribution of this variance is given by a $\chi^2(N-r+1)$ as shown in [1]. We will approximate this by $\chi^2(N)$ as we study the regime $N \to +\infty$, r constant, and $d \to \infty$ with $d = o(N)$. This approximation is not rigorously taken care of here and needs to be proven in a sense that remains to be defined for completeness. As the sparsity parameter r remains constant in our setup and the dimension of the matrix (number of assets) $d \to \infty$, we will assume the minimum

$$\min_{I \in \mathcal{P}_r([d])} v_I^*$$

to be the minimum of $\binom{d}{r}$ iid random variables with distribution $\chi^2(N)$.

We therefore define $X_i \sim \chi^2(N)$ iid for $i \in S$ with $|S| = s$ and study $\min_{i \in S} X_i$. We can study this distribution by assuming $X_i = \sum_{k=1}^N Z_{i,k}^2$ with iid standard normal $Z_{i,k}$'s. We notice that $\mathbb{V}[Z_{i,k}^2] = 2$. As we are taking the asymptotic in N, we rewrite

$$X_i = \sqrt{2N} \underbrace{\left(\frac{1}{\sqrt{N}} \sum_{k=1}^N \frac{Z_{i,k}^2 - 1}{\sqrt{2}} \right)}_{\xrightarrow[N\to\infty]{d} \mathcal{N}(0,1)} + N$$

By using the formula of the maximum (symmetric for the minimum) of N iid standard normal variables (true with high probability), we asymptotically estimate

$$\min_{i \in S} X_i - N \sim -\sqrt{2N \log s}$$

where $\sim$ means that the ratio of the two sides tends towards 1. This estimation is under the iid assumption and uses an unjustified swap of limits, these are yet to be made rigorous. In our asymptotic setup, we have $\log s = \log \binom{d}{r} \sim r \log d$. Thus, we make the following assumption:

$$\min_{I \in \mathcal{P}_r([d])} \mathbf{w}_I^{*\top} \hat{\Sigma}_I \mathbf{w}_I^* - N \sim -\underbrace{\sqrt{2Nr \log d}}_{\triangleq \theta_N} \tag{4}$$

with high probability as $N \to +\infty$.
This result provides us with an estimation of the minimum of the weighted-average value of a $r \times r$ submatrix. The expression has a form similar to the one from the initial Largest Submatrix Problem (recall, $2\sqrt{\log n/k}$). This encourages us to now consider the overlap set for weighted-averages values asymptotically close to $\alpha\sqrt{2Nr \log d}$ for $\alpha \in (0,1)$ and compute the expectation of the Overlap Set for such values. In fact, we apply the same approach as Gamarnik and Li to study a potential Overlap Gap Property and a related phase transition.

2.4 Expectation of the Overlap Set

We can define the overlap set: Let $\mathcal{O}_r(\alpha, y, \delta)$ be the set of pairs of $r \times r$ submatrices A_I, A_J for $I, J \in \mathcal{P}_r([d])$ for wich the optimal variances verify: $v_I^*, v_J^* \in [N - (\alpha \pm \delta)\theta_N]$ and such that $|I \cap J|/r \in [y \pm \delta]$.

By the straightforward counting of such pairs of matrices, using the linearity of expectation, we get:

$$\mathbb{E}[|\mathcal{O}_r(\alpha, y, \delta)|] = \sum_{i \in [(y \pm \delta)r]} \binom{d}{i}\binom{d-i}{r-i}\binom{d-r}{r-i} \times \mathbf{P}[v^*_{I_i}, v^*_{J_i} \in [N - (\alpha \pm \delta)\theta_N]] \tag{5}$$

where $|I_i \cap J_i| = i$. In order to reproduce the idea used in [4], we write $I = K \sqcup I_0$ and $J = K \sqcup J_0$ with $|K| = i$ and $I_0 \cap J_0 = \varnothing$. We now want to decompose the variances $v^*_I = v_K + v_I^{(0)}$ and $v^*_J = v_K + v_J^{(0)}$ where $v_K, v_I^{(0)}, v_J^{(0)}$ are independent and follow known distributions. The caveat here is the non-separability of the expression $v^*_I = \dfrac{1}{\mathbb{1}^\top \hat{\Sigma}_I^{-1} \mathbb{1}}$ as it was in the largest submatrix problem. We can however notice that the matrix $\hat{\Sigma}_I$ concentrates around its mean $\mathbf{I}_r$ and the function:

$$\begin{aligned} f \colon \mathrm{GL}_r(\mathbb{R}) &\to \mathbb{R} \\ M &\mapsto (\mathbb{1}^\top M^{-1} \mathbb{1})^{-1} \end{aligned}$$

is differentiable over its domain. We compute its gradient at I_r which is $\nabla f(\mathbf{I}_r) = \dfrac{1}{r^2}\mathbb{1}\mathbb{1}^\top$. Since the value of f is invariant under the action of the symmetric group of order r that applies a common permutation to the lines and the columns, we can assume that the common indices in K are the first ones in the covariance matrices $\hat{\Sigma}_I$ and $\hat{\Sigma}_J$. We define the following submatrices: $\hat{\Sigma}_I = \begin{pmatrix} S & T_I \\ T_I^\top & C_I \end{pmatrix}$ and $\hat{\Sigma}_J = \begin{pmatrix} S & T_J \\ T_J^\top & C_J \end{pmatrix}$. If we define $\mathbb{1}_k$ to be the ones vector of $\mathbb{R}^k$, we have the following Taylor expansions:

$$\begin{aligned} v^*_I &= \frac{1}{r^2}\mathbb{1}_{r-i}^\top S \mathbb{1}_{r-i} + \frac{2}{r^2}\mathbb{1}_{r-i}^\top T_I \mathbb{1}_i + \frac{1}{r^2}\mathbb{1}_i^\top C_I \mathbb{1}_i + \mathcal{O}(\|\Sigma_I - \mathbf{I}_r\|) \\ v^*_J &= \frac{1}{r^2}\mathbb{1}_{r-i}^\top S \mathbb{1}_{r-i} + \frac{2}{r^2}\mathbb{1}_{r-i}^\top T_J \mathbb{1}_i + \frac{1}{r^2}\mathbb{1}_i^\top C_J \mathbb{1}_i + \mathcal{O}(\|\Sigma_J - \mathbf{I}_r\|) \end{aligned}$$

As we did for v^*_I, $\mathbb{1}_i^\top S \mathbb{1}_i$ has distribution $\chi^2(N - r + i + 1)$ that we approximate by $\chi^2(N)$, $\mathbb{1}_i^\top C_I \mathbb{1}_i$ has distribution $\chi^2(N - i + 1)$ that we approximate by $\chi^2(N)$, and

$$\mathbb{1}_{r-i}^\top T_J \mathbb{1}_i = \sqrt{i(r-i)} \sum_{t=1}^N \tilde{Z}_{i,t}\tilde{Z}_{r-i,t} = \frac{\sqrt{i(r-i)}}{4} \sum_{t=1}^N (\tilde{Z}_{i,t} + \tilde{Z}_{r-i,t})^2 - (\tilde{Z}_{i,t} - \tilde{Z}_{r-i,t})^2$$

where $\tilde{Z}_{i,t}, \tilde{Z}_{r-i,t}$ are iid standard Gaussians which implies $(\tilde{Z}_{i,t} + \tilde{Z}_{r-i,t})^2, (\tilde{Z}_{i,t} - \tilde{Z}_{r-i,t})^2$ are independent $\chi^2(2)$ and therefore

$$\mathbb{1}_{r-i}^\top T_J \mathbb{1}_i \overset{d}{=} \frac{\sqrt{i(r-i)}}{4}(X - Y)$$

where $X, Y \sim \chi^2(2N)$ are independent. If we can abandon the negligible terms in v^*_I, v^*_J, we can break them into a common term and separated terms for which we know the distribution. We define: $\alpha \triangleq \frac{1}{r^2}\mathbb{1}_{r-i}^\top S \mathbb{1}_{r-i}, \beta_I \triangleq \frac{2}{r^2}\mathbb{1}_{r-i}^\top T_I \mathbb{1}_i, \gamma_I \triangleq \frac{1}{r^2}\mathbb{1}_i^\top C_I \mathbb{1}_i$ and the analog for J. That would allow us to write:

$$\begin{aligned} \mathbf{P}[v^*_{I_i}, v^*_{J_i} \in [N - (\alpha \pm \delta)\theta_N]] &\approx \mathbf{P}[\alpha + \beta_I + \gamma_I, \alpha + \beta_J + \gamma_J \in [N - (\alpha \pm \delta)\theta_N]] \\ &= \Gamma\left(\frac{N}{2}\right)^{-1} 2^{-\frac{N}{2}} \int_0^\infty \mathbf{P}[\beta_I + \gamma_I, \beta_J + \gamma_J \in [N - (\alpha \pm \delta)\theta_N - a]] da \end{aligned}$$

Afterwards and before making all our previous claim rigorous, two caveats remain to deal with :

- Deal with the dependency of T_I and T_J to break apart our probability in a product (we conjecture that this is solvable in our case by using the convergence of $\hat{\Sigma}$ to $\mathbf{I}_r$ in order to make these two terms negligible)
- Find a way to approximate this integral closely enough in order to recover the asymptotic behavior of $\mathbb{E}[|\mathcal{O}_r(\alpha, y, \delta)|]$.

3 Temp: Covariance computation

Since we're dealing with multivariate gaussian and chi-2 variables, the computation of the probability boils down to the computation of the covariance of our random variables.
Using the notation $t_I \triangleq \mathbb{1}_{r-i}^\top T_I \mathbb{1}_i$ and extending it to all the variables concerned, and $Z_t^{(r)} = (Z_{1,t}, \dots, Z_{r,t})$ for a vector of r Gaussian variables *iid*, we have:

1. $t_I = \sum_{t=1}^{N} \sum_{k=1}^{i} \sum_{l=i+1}^{r} Z_{k,t} Zl, t = \sqrt{i(r-i)} \sum_t Z_t^{(i)} Z_t^{(r-i)}$ with $Z_t^{(i)}$ and $Z_t^{(r-i)}$ *iid*
2. $t_J = \sum_{t=1}^{N} \sum_{k=1}^{i} \sum_{l=i+1}^{r} Z_{k,t} \tilde{Z} l, t = \sqrt{i(r-i)} \sum_t Z_t^{(i)} \tilde{Z}_t^{(r-i)}$ where the distinction between Z and $\tilde{Z}$ *iid* comesfrom the non-int
3. $c_I = (r-i) \sum_t (Z^2)_t^{(r-i)}$
4. $c_J = (r-i) \sum_t (\tilde{Z}^2)_t^{(r-i)}$
5. $s = i \sum_t (Z^2)_t^{(i)}$

We now compute a list of the covariances between those variables (recall that the Gaussian variables are also serially independent):

1. $\mathrm{Cov}(t_I, t_I) = \mathrm{Cov}(t_J, t_J) = i(r-i) \sum_{t,t'} \mathbb{E}(Z^2)_t^{(r-i)} (Z^2)_{t'}^{(i)} = i(r-i) N^2$
2. $\mathrm{Cov}(c_I, c_I) = \mathrm{Cov}(c_J, c_J) = (r-i)^2 \sum_{t,t'} \mathbb{E}(Z^2)_t^{(r-i)} (Z^2)_{t'}^{(r-i)} = (r-i)^2 \sum_t \left(\mathbb{E}(Z^4)_t^{(r-i)} + \sum t' \neq t \mathbb{E}(Z^2)_t^{(r-i)} (Z^2)_{t'}^{(r-i)} \right) = (r-i)^2 N(N+2)$
3. $\mathrm{Cov}(s, s) = i^2 \sum_{t,t'} \mathbb{E}(Z^2)_t^{(i)} (Z^2)_{t'}^{(i)} = i^2 N(N+2)$
4. $\mathrm{Cov}(t_I, t_J) = i(r-i) \sum_{t,t'} \mathbb{E} Z_t^{(i)} Z_t^{(r-i)} Z_{t'}^{(i)} \tilde{Z}_{t'}^{(r-i)} = 0$
5. $\mathrm{Cov}(c_I, c_J) = (r-i)^2 \sum_{t,t'} \mathbb{E}(Z^2)_t^{(r-i)} (\tilde{Z}^2)_{t'}^{(r-i)} = (r-i)^2 N^2$
6. $\mathrm{Cov}(s, t_I) = \mathrm{Cov}(s, t_J) = i\sqrt{i(r-i)} \sum_{t,t'} \mathbb{E}(Z^2)_t^{(i)} Z_{t'}^{(r-i)} Z_{t'}^{(i)} = 0$
7. $\mathrm{Cov}(s, c_I) = \mathrm{Cov}(s, c_J) = i(r-i) \sum_{t,t'} \mathbb{E}(Z^2)_t^{(i)} (Z^2)_{t'}^{(r-i)} = i(r-i) N^2$
8. $\mathrm{Cov}(t_I, c_I) = \mathrm{Cov}(t_J, c_J) = (r-i)\sqrt{i(r-i)} \sum_{t,t'} \mathbb{E} Z_t^{(i)} Z_t^{(r-i)} (Z^2)_{t'}^{(r-i)} = 0$

9. $\mathrm{Cov}(t_I, c_J) = \mathrm{Cov}(t_J, c_I) = (r-i)\sqrt{i(r-i)} \sum_{t,t'} \mathbb{E} Z_t^{(i)} \tilde{Z}_t^{(r-i)} (Z^2)_{t'}^{(r-i)} = 0$

Hence

$$\mathrm{Cov}(v_I^*, v_J^*) = \mathrm{Cov}(s + t_I + c_I, s + t_J + c_J) = i^2 N(N+2) + 2i(r-i)N^2 + (r-i)^2 N^2 = r^2 N^2 + 2i^2 N$$

and

$$\mathrm{Var}(v_I *) = \mathrm{Var}(v_J^*) = \mathrm{Cov}(s+t_I+c_I, s+t_I+c_I) = i^2 N(N+2) + 3i(r-i)N^2 + (r-i)^2 N(N+2) = r^2 N^2 + 2(i^2 + (r-i)^2)N + i(r-i)N^2$$

4 Sparse Portfolio Optimization

4.1 Mathematical Formulation

The classic long-only Mean-Variance optimization problem takes the following form:

$$\max_{w} \hat{r}'w$$
$$\text{s.t.} \begin{cases} w'\hat{\Sigma}w \leq \sigma^2_{\max} \\ \sum_{i=1}^{n} w_i = 1 \\ w_i \geq 0 \end{cases}$$

This problem constitutes a second-order cone programming problem. However, since the estimated covariance matrix $\hat{\Sigma}$ is positive semidefinite, the problem can be reformulated as a quadratic programming problem, which solves in $O(n^3)$, as proved in Goldfarb & Liu, 1990 [5].

We take this as a comparative benchmark when comparing it to our sparse problems. The "default formulation of our sparse optimization problem is the following:

$$\max_{w,z} \hat{r}'w$$
$$\text{s.t.} \begin{cases} w'\hat{\Sigma}w \leq \sigma^2_{\max} \\ \sum_{i=1}^{n} w_i = 1 \\ w_i \leq z_i \\ \sum_{i=1}^{n} z_i \leq s \\ w_i \geq 0 \\ z_i \in \{0, 1\} \end{cases}$$

4.2 Methodology

We designed a Python library to process market data and create our portfolio optimization models to run on it. The library allows you to write your own models and run them on selected data automatically, which gave us flexibility to test various variations of our problem. The code, along with the data and results it generated, are included in the zip file.

The first component of the library is the `MarketLoader` class, which we used to load weekly returns on a variety of securities. For our tests on n securities, we used the first $n - 4$ tickers of the Russell 3000 index, along with 4 treasury bonds with different maturities. Including the treasury securities ensures feasibility of the optimization problem.

We used the Adjusted Close prices to calculate returns. Given our primary objective of comparing runtimes, our focus did not center on forecast quality. For the estimation of expected returns, we adopted a simple 1-year moving average approach. Similarly, we computed the covariance matrix using a 1-year rolling sample covariance. It is important to acknowledge that these estimation methods may exhibit limited predictive power within our model. In order to design a successful strategy, it would be necessary to rework these estimates.

4.3 Main results

We tested the variant on different securities numbers n (which give the time complexity), and on different sparsity rates s. The securities number ranged from 128 to 2048 (sith $\sqrt{2}$ power increments), and the sparsity ratio ranged from 1% to 32% (with ×2 product increments).

We also tested the model at 5 different time steps consecutively. Since the runtime is highly stochastic in this case, this allowed us to have somewhat of an average runtime on all 5 dates. Due to computational limitations however, we couldn't run it on more than 5 dates, but it would be recommendable to try to run it on several dozen dates to have a more meaningful measure of the average runtime.

Our solver doesn't find the exact optima, as in any mixed-integer programming problem (due to the discrete nature of the algorithm), but rather finds an ϵ approximate solution. By default, Gurobi sets the tolerance ϵ at 10^{-6}, and we will keep it as such in all further computations.

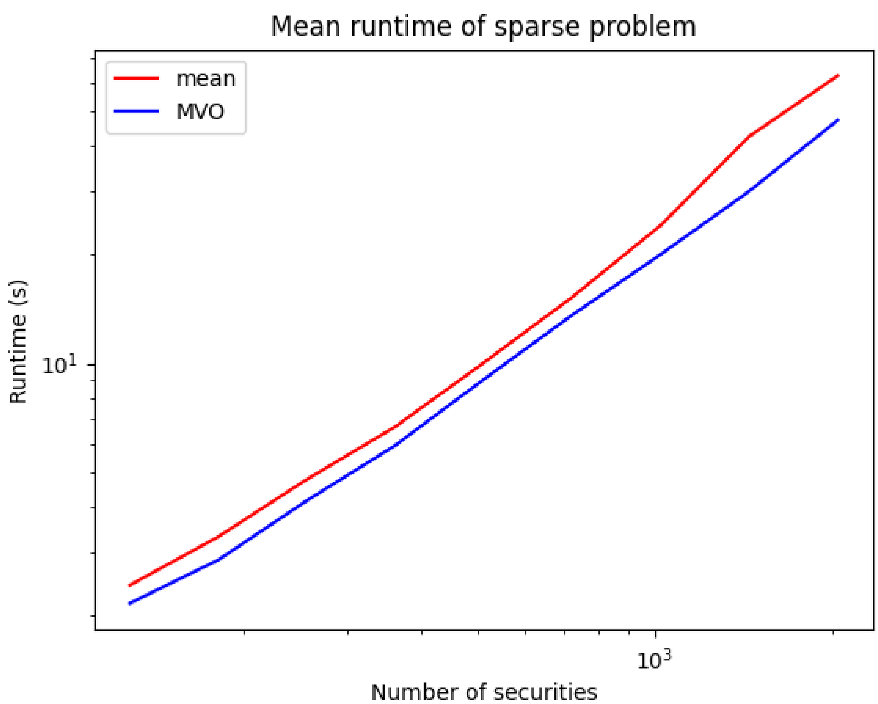

Figure 1: Mean time-complexity of the default sparsity problem

As seen in (2), each sparse model takes strictly longer than the MVO model to run, as expected (as they each add constraints to the MVO). However, our results seem to indicate that the sparsity problem follows similar complexity to the MVO problem ($O(n^3)$).

In analysing the results, we find that this might be caused by the fact that even in the basic MVO setting, the optimal weights are found to be highly sparse. In the $n = 128$ setting for example, the MVO only allocates strictly positive weights to 9 different assets. The ℓ_2 distance of the optima of any pair of sparsity model (including MVO) is of the order 10^{-6}, and is therefore probability linked mainly to the tolerance error of the solver.

In particular, because Gurobi uses a branch-and-bound algorithm to solve MIQP problems by default,

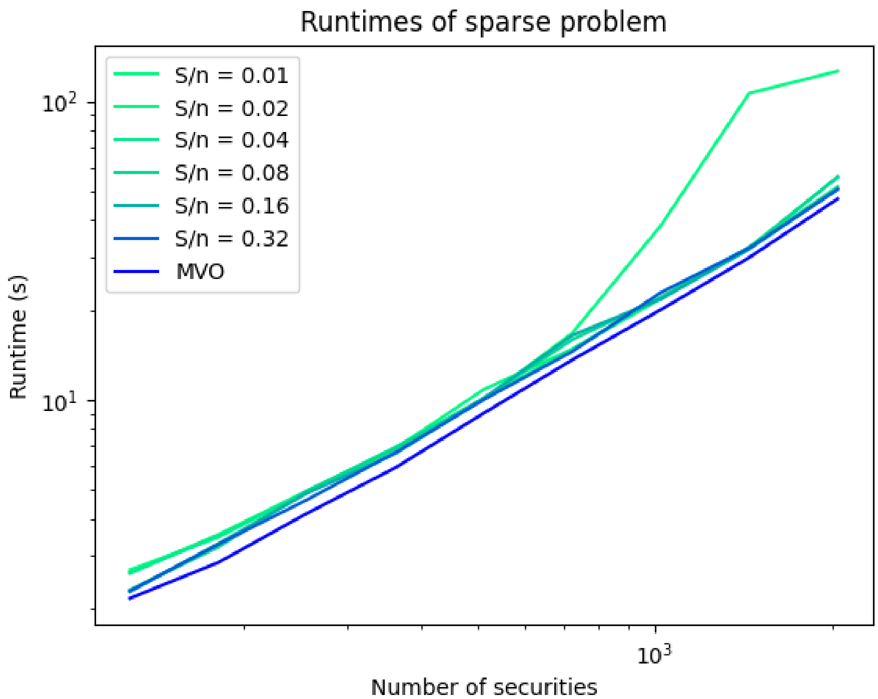

Figure 2: Time-complexity of the default sparsity problem at each sparsity level

the added integer constraints will have very little impact on runtime in general because the solutions are the same with or without the constraints. In particular, when the integer constraints are relaxed, after the first iteration of the QP problem, the optimum found already satisfies the integer constraints, so the branch-and-bound algorithm will find a solution in one iteration. Therefore, the integer constraints won't affect runtime at all (in complexity).

The fact that the original solution is sparse represents a problem here, because it doesn't allow us to conclude anything significant regarding the complexity of the sparse optimization problem. This might come from the long only constraint. In particular, because our covariance (and expected returns) estimates are likely very bad because of the small sample size (52), it is probable that the model focuses on a few outliers.

To mediate this problem, we tried another approach, detailed below.

4.4 Long-short model

We computed runtimes on a variant of the default model, on the same sparsities and securities sets as before, but we allowed for short selling. Specifically, we switched the $w_i \geq 0$ constraint for a $|w_i| \leq 1$, in order to avoid having highly sparse optima. Not that this also adds the extra constraints $w_i \geq -z_i \forall i = 1, ..., n$, but this can only increase the complexity by a constant factor.

This model, although poorly accurate in practice, represents the other end of the sparsity spectrum. The optimal solutions are all extremely non-sparse, consisting almost only of values in $\{-1, 1\}$. Therefore, the sparse solution is likely to be very far away from the relaxed problem's solution.

In this case, the runtimes are much more volatile, and seems to be NP-hard at first sight as indicated

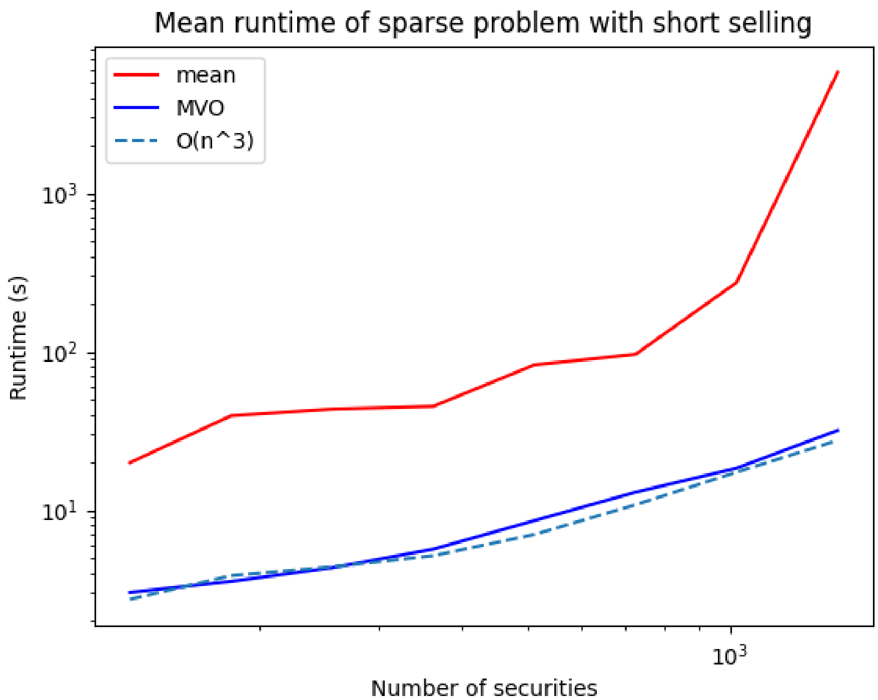

Figure 3: Mean time-complexity of the sparsity problem allowing for short selling

Figure 4: Time-complexity of each sparsity problem allowing for short selling

by (3). This can be explained by how "lucky" the branch-and bound algorithm is when finding an optimal solution. We know that at worse, the branch-and-bound algorithm used by Gurobi is NP hard, and this seems to coincide with the results we are getting here.

However, upon further analysis, we see that the hike in time complexity is in fact caused by a single sparsity level (s = 16%) (see 4). All other sparsity level seem to follow normal polynomial times. Since the case s = 16% is between the MVO case (s = 100%), and the lower sparsity levels, it would seem counter-intuitive if this sparsity level indeed represents an NP-hard problem. Therefore, since all estimates are very volatile here, it might be the case that the branch-and-bound got very "unlucky" in this specific iteration, but that the average runtime is in fact polynomial. Further analysis (requiring larger computational power) would be needed to find a better estimate of the time complexity.

5 Conclusion

In this project, we focused on understanding the sparse portfolio optimization, its complexity and how it relates to the well known convex mean variance optimization explored by Markowitz. We conduct this analysis through an empirical numerical analysis and a theoretical study of the possible exhibition of the Overlap Gap Property (OGP) in this problem.

Computationally, the problem seems to be NP-hard at worst, but polynomial on average. Although in the default model proposed specifically, the problem looks to be of the same order as the MVO problem, this likely comes from the fact that the solution to the relaxed (MVO) problem is often the same as its sparse counterpart. Therefore, the branch-and-bound process used to solve the MIQP problem is resolved in a single iteration, which doesn't affect runtime. However, when allowing for short-selling (i.e. making the solution to the MVO problem non-sparse), we find that the problem seems to be NP hard in some case, but polynomial on average. Nevertheless, larger computational capacities would be necessary to confirm this hypothesis.

On the theoretical side, we followed the path of [4] to find the behavior of the size of the overlap sets. This nonrigorous first analysis showed some complications compared to the largest submatrix problem as we needed to study local optima for the new parameter $\mathbf{w}$ that were not conserved from one submatrix to another. This led us to use a non-linear explicit solution of the local optimization problems that was non separable and a Taylor expansion allowed us to create some separation. These first steps clear the way for further analysis of the expected size of the overlap sets but some challenges lay ahead as the dependency of the different terms in the separated variances and the computations of the non-gaussian integrals. We are continuing our investigation in this direction.

References

[1] Taras Bodnar, Stepan Mazur, and Krzysztof Podgórski. Singular inverse wishart distribution and its application to portfolio theory. *Journal of Multivariate Analysis*, 143:314–326, 2016.

[2] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[3] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

[4] David Gamarnik and Quan Li. Finding a large submatrix of a Gaussian random matrix. *Ann. Statist.*, 46(6A):2511–2561, 2018.

[5] Donald Goldfarb and Shucheng Liu. An o (n 3 l) primal interior point algorithm for convex quadratic programming. *Mathematical programming*, 49(1-3):325–340, 1990.

[6] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.