



WORKING PAPER 12/2021 (STATISTICS)

Portfolio Selection with a Rank-deficient Covariance Matrix

Mårten Gulliksson, Anna Oleynik and Stepan Mazur

ISSN 1403-0586

Örebro University School of Business
SE-701 82 Örebro, Sweden

Portfolio Selection with a Rank-deficient Covariance Matrix

Mårten Gulliksson^(a), Anna Oleynik^(b), and Stepan Mazur^(c)

^(a)School of Science and Technology, Örebro University, 70182 Örebro, Sweden

^(b)Department of Mathematics, Univeristy of Bergen, 5020 Bergen, Norway

^(c)School of Busniess, Örebro University, 70182 Örebro, Sweden

June 2, 2021

Abstract

In this paper, we consider optimal portfolio selection when the covariance matrix of the asset returns is rank-deficient. For this case, the original Markowitz' problem does not have a unique solution. The possible solutions belong to either two subspaces namely the range- or nullspace of the covariance matrix. The former case has been treated elsewhere but not the latter. We derive an analytical unique solution, assuming the solution is in the null space, that is risk-free and has minimum norm. Furthermore, we analyse the iterative method which is called the discrete functional particle method in the rank-deficient case. It is shown that the method is convergent giving a risk-free solution and we derive the initial condition that gives the smallest possible weights in the norm. Finally, simulation results on artificial problems as well as real-world applications verify that the method is both efficient and stable.

Keywords: Mean–variance portfolio; Rank-deficient covariance matrix; Linear ill-posed problems; Second order damped dynamical systems

1 Introduction

Modern portfolio theory plays an important role in finance for both researchers and practitioners. Its fundamental goal is to provide efficient ways for the allocation of investments among different assets in the portfolio. The basic concepts of modern portfolio theory were developed in [21] by Markowitz, who introduced a mean-variance portfolio optimization procedure in which investors incorporate their preferences towards variance and

their expectation of the return on different assets. The original Markowitz' problem is formulated as follows

$$\min_{\mathbf{w}} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^\top \mathbf{1} = 1, \quad \mathbf{w}^\top \boldsymbol{\mu} = q, \quad (1)$$

where \mathbf{w} is a vector of portfolio weights, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are mean vector and covariance matrix of the asset returns, $\mathbf{1}$ denotes the vector of ones, and q stands for the expected rate of return on the portfolio. In practice, both parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unobservable and, therefore, need to be estimated from historical data. The most common estimators are sample mean vector and sample covariance matrix (see, for example, [25], [13], [7], [3], [19], among many others). The sample covariance matrix is positive definite when the number of observations is greater than the portfolio size, otherwise, it is positive semi-definite and rank-deficient¹.

If $\boldsymbol{\Sigma}$ is a positive definite matrix, then the optimization problem in (1) has a unique (global) minimum given by

$$\mathbf{w} = \frac{c - bq}{ac - b^2} \boldsymbol{\Sigma}^{-1} \mathbf{1} + \frac{aq - b}{ac - b^2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \quad (2)$$

where $a = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}$, $b = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$, and $c = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$. Meanwhile, if $\boldsymbol{\Sigma}$ is rank-deficient, the minimum is not necessarily unique and the solution given in (2) is not valid since the regular inverse of $\boldsymbol{\Sigma}$ does not exist. In [23], it has been shown that if, in addition to the constraints in (1), one would assume that \mathbf{w} does not belong to the null space of $\boldsymbol{\Sigma}$, $\mathcal{N}(\boldsymbol{\Sigma})$, the minimum would be achieved at

$$\mathbf{w} = \frac{c - bq}{ac - b^2} \boldsymbol{\Sigma}^+ \mathbf{1} + \frac{aq - b}{ac - b^2} \boldsymbol{\Sigma}^+ \boldsymbol{\mu} \quad (3)$$

where $a = \mathbf{1}^\top \boldsymbol{\Sigma}^+ \mathbf{1}$, $b = \mathbf{1}^\top \boldsymbol{\Sigma}^+ \boldsymbol{\mu}$, $c = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^+ \boldsymbol{\mu}$, and $\boldsymbol{\Sigma}^+$ denotes the Moore-Penrose inverse of $\boldsymbol{\Sigma}$. Note that the expression in (3) is used in several papers with a focus on the statistical analysis ([9], [10], [11], [1]). However, there is no obvious motivation for the additional constraint $\mathbf{w} \notin \mathcal{N}(\boldsymbol{\Sigma})$, except the apparent similarity between solutions given in (3) and (2). Moreover, the expression in (3) is not a solution to the classical Markowitz' problem formulated in (1) unless this constraint is imposed. Further on, we refer to this solution as a Range Space (RS) solution.

¹Under the assumption that the asset returns are independent and identically multivariate normally distributed, the sample mean vector has multivariate (singular) normal distribution while the sample covariance matrix has (singular) Wishart distribution (see [22] and [9]). Moreover, the vector which is proportional to the product of the (generalized) inverse sample covariance matrix and sample mean vector represent the sample estimator of the tangency portfolio weights. It leads us to the product of the (singular) inverse Wishart matrix and (singular) normal vector which is studied by [12], [5], [6], [20], [4], and [8].

The focus of this paper is on the rank-deficient case. We contribute to the existing literature by deriving an analytical solution to (1) that can be obtained via the singular value decomposition (SVD). This solution lies in the null space of Σ and, therefore, is a global (non-unique) minimum of (1). In addition, this global minimum corresponds to the smallest portfolio weight norm among all the other minima. For these reasons, we call the solution as a risk-free minimum norm solution or, when comparing to the RS solution (3), as a null space minimum norm (NSMN) solution to (1). There are disadvantages of using the SVD approach. Firstly, the SVD may be computationally costly. Secondly, if there is a need to find many portfolios even smaller size SVDs might increase the computational cost too much. Consequently, it might be of value to use an iterative method. Here, our second contribution to the existing literature comes in. We further develop the iterative method presented in [16], called the Discrete Functional Particle Method (DFPM), and analyze its convergence properties for rank-deficient Σ . We show that DFPM always converges to a risk-free solution and, for a specific choice of the initial value, the same minimum norm solution to (1) as mentioned above.

The structure of the paper is as follows. In Section 2, we present the analytical form of the risk-free minimum norm solution. Section 3 discuss the solution obtained via the DFPM. Sections 4 and 5 consist of a simulation and empirical studies, respectively.

2 Risk-free minimum norm solution

We begin by stating the conditions for a unique solution of (1). It is then convenient to use a more compact form of our problem

$$\min_{\mathbf{w}} \mathbf{w}^\top \Sigma \mathbf{w} \quad \text{s.t.} \quad \mathbf{B} \mathbf{w} = \mathbf{c}, \quad (4)$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{1}^\top \\ \boldsymbol{\mu}^\top \end{bmatrix} \in \mathbb{R}^{2 \times p}, \quad \mathbf{c} = \begin{bmatrix} 1 \\ q \end{bmatrix} \in \mathbb{R}^2. \quad (5)$$

From now on, we assume that \mathbf{B} has full rank, i.e., $\boldsymbol{\mu}$ is not parallel to $\mathbf{1}$. It is straight forward to treat the case of \mathbf{B} rank-deficient. For the analysis to come, we use the singular value decomposition of Σ ,

$$\Sigma = \mathbf{V}_\Sigma \mathbf{D}_\Sigma \mathbf{V}_\Sigma^\top, \quad (6)$$

where $\mathbf{V}_\Sigma \in \mathbb{R}^{p \times p}$ is orthonormal and $\mathbf{D}_\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0)$, $\sigma_j > 0$, $j = 1, \dots, r$ with r the rank of Σ . We will also need the following partition

$$\mathbf{V}_\Sigma = [\mathbf{V}_{\Sigma 1} \quad \mathbf{V}_{\Sigma 2}], \quad \mathbf{D}_\Sigma = \begin{bmatrix} \mathbf{D}_{\Sigma 1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (7)$$

where $\mathbf{V}_{\Sigma_1} \in \mathbb{R}^{p \times r}$, $\mathbf{V}_{\Sigma_2} \in \mathbb{R}^{p \times (p-r)}$, $\mathbf{D}_{\Sigma_1} = \text{diag}(\sigma_1, \dots, \sigma_r)$, and $\mathbf{0}$ stands for the matrix of zeros. Due to this partition, the decomposition (6) can be written in reduced form

$$\mathbf{\Sigma} = \mathbf{V}_{\Sigma_1} \mathbf{D}_{\Sigma_1} \mathbf{V}_{\Sigma_1}^\top. \quad (8)$$

In the next lemma, we show that if the rank of $\mathbf{\Sigma}$ is less than $p - 2$, a solution to (1) is non-unique.

Lemma 2.1. *The solution to the optimization problem given in (1) is not unique if $r < p - 2$.*

Proof. Using the KKT-conditions with the Lagrange function $L = \mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} + \boldsymbol{\lambda}^\top (\mathbf{B} \mathbf{w} - \mathbf{c})$, the solution is given by the linear system

$$\begin{bmatrix} \mathbf{\Sigma} & \mathbf{B}^\top \\ \mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{c} \end{bmatrix}, \quad (9)$$

where the elements in $\boldsymbol{\lambda}$ contains the two Lagrange parameters. Substituting the SVD given in (6) into (9), we get

$$\begin{bmatrix} \mathbf{D}_{\Sigma} & \mathbf{V}^\top \mathbf{B}^\top \\ \mathbf{B} \mathbf{V} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}^\top \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{c} \end{bmatrix}. \quad (10)$$

Next, using the partition (7), we can write

$$\begin{bmatrix} \mathbf{D}_{\Sigma_1} & \mathbf{0} & (\mathbf{B} \mathbf{V}_{\Sigma_1})^\top \\ \mathbf{0} & \mathbf{0} & (\mathbf{B} \mathbf{V}_{\Sigma_2})^\top \\ \mathbf{B} \mathbf{V}_{\Sigma_1} & \mathbf{B} \mathbf{V}_{\Sigma_2} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\Sigma_1}^\top \mathbf{w} \\ \mathbf{V}_{\Sigma_2}^\top \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{c} \end{bmatrix}. \quad (11)$$

The matrix $\mathbf{B} \mathbf{V}_{\Sigma_2}$ is of size $2 \times (n - r)$. Therefore, if $r < p - 2$, i.e. $p - r > 2$, the matrix in (9) will be rank-deficient and, consequently, the solution will not be unique. The lemma is proved. \square

Since we are only interested in the case of a non-unique solution, we from now on assume that $r < p - 2$.

The next lemma gives us a condition for a solution to be risk-free, i.e., $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} = 0$.

Lemma 2.2. *Let $r < p - 2$. Then $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} = 0$ if and only if $\mathbf{w} = \mathbf{V}_{\Sigma_2} \mathbf{y}$, where \mathbf{V}_{Σ_2} is defined in the partition (7) and $\mathbf{y} \in \mathbb{R}^{p-r}$.*

Proof. Using the SVD of $\mathbf{\Sigma}$ in (6) and (7), we have

$$\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} = \mathbf{w}^\top \mathbf{V}_{\Sigma_1} \mathbf{D}_{\Sigma_1} \mathbf{V}_{\Sigma_1}^\top \mathbf{w} = \boldsymbol{\xi}^\top \mathbf{D}_{\Sigma_1} \boldsymbol{\xi} = \sum_{j=1}^r \xi_j^2 \sigma_j,$$

where $\boldsymbol{\xi} = \mathbf{V}_{\Sigma_1}^\top \mathbf{w}$. Therefore, $\mathbf{w}^\top \mathbf{\Sigma} \mathbf{w} = 0$ implies $\boldsymbol{\xi} = \mathbf{0}$ and $\mathbf{V}_{\Sigma_1}^\top \mathbf{w} = \mathbf{0}$. Since $\mathbf{V}_{\Sigma_1}^\top \mathbf{V}_{\Sigma_2} = \mathbf{0}$, we get $\mathbf{w} = \mathbf{V}_{\Sigma_2} \mathbf{y}$. The other implication in the lemma is obvious. The proof of the lemma is complete. \square

Next, we show how to choose a risk-free solution with a minimal ℓ_2 norm. For this, we need yet another SVD decomposition. We introduce $\mathbf{Y} = \mathbf{B}\mathbf{V}_{\Sigma_2} \in \mathbb{R}^{2 \times (p-r)}$ and define its decomposition as

$$\mathbf{Y} = \mathbf{U}_{\mathbf{Y}}\mathbf{D}_{\mathbf{Y}}\mathbf{V}_{\mathbf{Y}}^\top, \quad (12)$$

or in the reduced form,

$$\mathbf{Y} = \mathbf{U}_{\mathbf{Y}_1}\mathbf{D}_{\mathbf{Y}_1}\mathbf{V}_{\mathbf{Y}_1}^\top, \quad (13)$$

where

$$\mathbf{V}_{\mathbf{Y}} = [\mathbf{V}_{\mathbf{Y}_1} \quad \mathbf{V}_{\mathbf{Y}_2}], \quad \mathbf{D}_{\mathbf{Y}_1} = \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix}, \quad \gamma_j > 0, \quad j \in \{1, 2\}, \quad (14)$$

and $\mathbf{V}_{\mathbf{Y}_1} \in \mathbb{R}^{(p-r) \times 2}$, $\mathbf{V}_{\mathbf{Y}_2} \in \mathbb{R}^{(p-r) \times (p-r-2)}$. Let $\mathcal{N}(\mathbf{A})$ stands for the null space of the matrix \mathbf{A} . We can now formulate our main theorem.

Theorem 2.1. *Let $r < p - 2$ and \mathbf{V}_{Σ_2} is defined as in (7), $\mathbf{Y} = \mathbf{B}\mathbf{V}_{\Sigma_2}$ and $\mathbf{V}_{\mathbf{Y}_2}$ is given as in (14). Then*

$$\mathbf{w} = \mathbf{V}_{\Sigma_2}\mathbf{Y}^+\mathbf{c} + \mathbf{V}_{\Sigma_2}\mathbf{x}, \quad \mathbf{x} \in \mathcal{N}(\mathbf{Y}), \quad (15)$$

is a solution to (1) and it gives zero risk. The minimum norm solution among (15), or specifically, the solution to

$$\begin{aligned} & \min_{\mathbf{w} \in \mathbb{R}^p} \frac{1}{2} \|\mathbf{w}\|_2^2 \\ & \text{s.t. } \mathbf{w}^\top \Sigma \mathbf{w} = 0 \\ & \mathbf{B}\mathbf{w} = \mathbf{c}, \end{aligned} \quad (16)$$

is given as

$$\mathbf{w}_{\min} = (\mathbf{I} - (\mathbf{V}_{\Sigma_2}\mathbf{V}_{\mathbf{Y}_2})(\mathbf{V}_{\Sigma_2}\mathbf{V}_{\mathbf{Y}_2})^+)\mathbf{Y}^+\mathbf{c}. \quad (17)$$

Proof. From Lemma 2.2 the problem (1) simplifies to finding $\mathbf{y} \in \mathbb{R}^{p-r}$ that satisfies the equation

$$\mathbf{B}\mathbf{V}_{\Sigma_2}\mathbf{y} = \mathbf{c}, \quad \text{or equivalently, } \mathbf{Y}\mathbf{y} = \mathbf{c}, \quad (18)$$

which is a system of two equations with $p-r$ unknowns. It is straightforward to see that any solution to (18) can be written in the following form

$$\mathbf{y} = \mathbf{Y}^+\mathbf{c} + \mathbf{x}, \quad \mathbf{x} \in \mathcal{N}(\mathbf{Y}).$$

Then, as $\mathbf{w} = \mathbf{V}_{\Sigma_2}\mathbf{y}$, we arrive at the expression in (15).

Next, we observe that $\mathbf{x} \in \mathcal{N}(\mathbf{Y})$ is equivalent to $\mathbf{x} = \mathbf{V}_{\mathbf{Y}_2}\mathbf{v}$, $\mathbf{v} \in \mathbb{R}^{p-r-2}$, and therefore

$$\mathbf{w} = \mathbf{V}_{\Sigma_2} \mathbf{Y}^+ \mathbf{c} + \mathbf{V}_{\Sigma_2} \mathbf{V}_{\mathbf{Y}_2} \mathbf{v}. \quad (19)$$

Thus, the constrained optimization in (16) simply could be written as the unconstrained problem

$$\min_{\mathbf{v} \in \mathbb{R}^{p-r-2}} \frac{1}{2} \|\mathbf{V}_{\Sigma_2} (\mathbf{B} \mathbf{V}_{\Sigma_2})^+ \mathbf{c} + \mathbf{V}_{\Sigma_2} \mathbf{V}_{\mathbf{Y}_2} \mathbf{v}\|_2^2. \quad (20)$$

The linear least squares problem in (20) is overdetermined and has the unique solution expressed as

$$\mathbf{v} = -(\mathbf{V}_{\Sigma_2} \mathbf{V}_{\mathbf{Y}_2})^+ \mathbf{V}_{\Sigma_2} \mathbf{Y}^+ \mathbf{c}. \quad (21)$$

Finally, we insert (21) into (19) to obtain (17). The theorem is proved. \square

The computational complexity to calculate \mathbf{w}_{\min} in (17) is mainly due to the SVDs of Σ , \mathbf{Y} , and $\mathbf{V}_{\Sigma_2} \mathbf{V}_{\mathbf{Y}_2}$. Since the SVD of an $r_1 \times r_2$ matrix is approximately $2r_1 r_2^2 + 11r_2^3$ (see [24]), we get a total cost of

$$N_{TOT} = 13p^3 + 4(p-r)^2 + 11(p-r)^3 + 2p(p-r-2)^2 + (p-r-2)^3.$$

An interesting special case is when $r \ll p$. Expanding factors we get

$$N_{TOT} = 27p^3 - 40p^2r - 10p^2 + 38pr^2 + 12pr + 20p - 12r^3 - 2r^2 - 12r - 8$$

with leading term, the main computational cost, $27p^3$. Thus, we have a complexity of order $\mathcal{O}(p^3)$, and for large p it is of interest to use an iterative method of lower complexity. In the next section, we will present such an iterative method that generally has $\mathcal{O}(p^2)$ in complexity.

Another aspect is that an iterative method can be used to get an approximate solution just stopping the iterations early as well as producing a set of possibly interesting portfolios.

3 The Discrete Functional Particle Method

Our approach for solving optimization problem stated in (1) iteratively is the Discrete Functional Particle method (DFPM), see [16] and [15] for details. DFPM is based on formulating a damped dynamical system and we will show that the stationary solution is the solution to (1) or with a specific initial value the solution given in (16).

This method can either be formulated by first eliminating the constraints in (1) or formulating an additional dynamical system for the constraints. We choose the former here and leave the latter for future research.

The null space of \mathbf{B} can be formulated using the SVD decomposition

$$\mathbf{B} = \mathbf{U}_{\mathbf{B}} \begin{bmatrix} \mathbf{D}_{\mathbf{B}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{\mathbf{B}_1}^\top \\ \mathbf{V}_{\mathbf{B}_2}^\top \end{bmatrix},$$

where $\mathbf{V}_{\mathbf{B}_2} \in \mathbb{R}^{p \times (p-2)}$ spans $\mathcal{N}(\mathbf{B})$. Thus, any solution to (1) can be written in the form

$$\mathbf{w} = \mathbf{V}_{\mathbf{B}_2} \mathbf{u} + \mathbf{g}, \quad (22)$$

where \mathbf{g} is any solution to $\mathbf{B}\mathbf{w} = \mathbf{c}$. Since \mathbf{B} has only two rows, the matrix $\mathbf{V}_{\mathbf{B}_2}$ can be found by doing a reduced (or thin) SVD of \mathbf{B} or a reduced QR-decomposition of \mathbf{B}^\top together with Gram-Schmidt orthogonalization at a cost of order $\mathcal{O}(p^2)$.

Let, e.g., $\mathbf{g} = \mathbf{B}^+ \mathbf{c}$, then (1) can be written as

$$\min_{\mathbf{u}} \frac{1}{2} (\mathbf{V}_{\mathbf{B}_2} \mathbf{u} + \mathbf{g})^\top \boldsymbol{\Sigma} (\mathbf{V}_{\mathbf{B}_2} \mathbf{u} + \mathbf{g}), \quad (23)$$

or

$$\min_{\mathbf{u}} \left(\frac{1}{2} \mathbf{u}^\top \mathbf{V}_{\mathbf{B}_2}^\top \boldsymbol{\Sigma} \mathbf{V}_{\mathbf{B}_2} \mathbf{u} + \mathbf{g}^\top \boldsymbol{\Sigma} \mathbf{V}_{\mathbf{B}_2} \mathbf{u} + \frac{1}{2} \mathbf{g}^\top \boldsymbol{\Sigma} \mathbf{g} \right). \quad (24)$$

Calculating \mathbf{g} can be done efficiently by a QR-decomposition of \mathbf{B}^\top with a cost of $\mathcal{O}(p^2)$. We can exclude the last term in (24) as it is constant and introduce

$$\mathbf{M} = \mathbf{V}_{\mathbf{B}_2}^\top \boldsymbol{\Sigma} \mathbf{V}_{\mathbf{B}_2} \text{ and } \mathbf{d} = \mathbf{V}_{\mathbf{B}_2}^\top \boldsymbol{\Sigma} \mathbf{g}, \quad (25)$$

to simplify the notation. Finally, we get the unconstrained problem

$$\min_{\mathbf{u} \in \mathbb{R}^{p-2}} \frac{1}{2} \mathbf{u}^\top \mathbf{M} \mathbf{u} + \mathbf{u}^\top \mathbf{d}. \quad (26)$$

Here, $\mathbf{M} \in \mathbb{R}^{(p-2) \times (p-2)}$ and $\mathbf{d} \in \mathbb{R}^{p-2}$.

The idea is to solve (26) by finding the stationary solution, say \mathbf{u}^* , to the second order damped dynamical system

$$\begin{aligned} \frac{d^2 \mathbf{u}}{dt^2} + \eta \frac{d\mathbf{u}}{dt} &= -\mathbf{d} - \mathbf{M}\mathbf{u}, \quad \eta > 0, \\ \mathbf{u}(0) &= \mathbf{u}_0, \quad \frac{d\mathbf{u}}{dt}(0) = \mathbf{v}_0. \end{aligned} \quad (27)$$

We remind the reader about our assumption on the rank of $\boldsymbol{\Sigma}$, i.e., $r < p - 2$. This implies that \mathbf{M} is rank-deficient as well with the rank $s \leq r$ and a solution to (26) is not unique. We show here that as $t \rightarrow \infty$, the solution to (27) converges to \mathbf{u}^* which corresponds to a zero risk solution \mathbf{w} via (22). Moreover, we show that there is a choice of the initial condition \mathbf{u}_0 , that ensures that \mathbf{u}^* converges to the risk-free minimum norm solution. Let us note that if $\boldsymbol{\Sigma}$ is positive definite, then so is \mathbf{M} , giving $\mathbf{u}^* = -\mathbf{M}^{-1} \mathbf{d}$. We refer to [16] for this special case.

3.1 Stationary solutions

Let the SVD of $\mathbf{M} \in \mathbb{R}^{(p-2) \times (p-2)}$ of rank s be defined as

$$\mathbf{M} = \mathbf{V}_M \mathbf{D}_M \mathbf{V}_M^\top = \mathbf{V}_{M1} \mathbf{D}_{M1} \mathbf{V}_{M1}^\top \quad (28)$$

where

$$\mathbf{V}_M = [\mathbf{V}_{M1} \quad \mathbf{V}_{M2}] , \quad \mathbf{D}_M = \begin{bmatrix} \mathbf{D}_{M1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad (29)$$

$\mathbf{V}_M \in \mathbb{R}^{(p-2) \times (p-2)}$, $\mathbf{V}_{M1} \in \mathbb{R}^{(p-2) \times s}$, $\mathbf{V}_{M2} \in \mathbb{R}^{(p-2) \times (p-2-s)}$, $\mathbf{D}_M \in \mathbb{R}^{(p-2) \times (p-2)}$, and $\mathbf{D}_{M1} \in \mathbb{R}^{s \times s}$. Here, the diagonal matrix \mathbf{D}_{M1} contains the non-zero singular values of \mathbf{M} , i.e., $\mathbf{D}_{M1} = \text{diag}(\gamma_1, \dots, \gamma_s)$. Substituting (28) into (27) and introducing the change of variables $\mathbf{z} = \mathbf{V}_M^\top \mathbf{u}$, we transform (27) into

$$\frac{d^2 \mathbf{z}}{dt^2} + \eta \frac{d\mathbf{z}}{dt} = -\mathbf{V}_M^\top \mathbf{d} - \mathbf{D}_M \mathbf{z}. \quad (30)$$

From (25), $\mathbf{d} \in \mathcal{R}(\mathbf{V}_{\Sigma_2}^\top)$ and as $\mathcal{R}(\mathbf{V}_{\Sigma_2}^\top) = \mathcal{R}(\mathbf{M}) = \mathcal{R}(\mathbf{V}_{M1})$, we have $\mathbf{d} = \mathbf{V}_{M1} \mathbf{h}$ for some $\mathbf{h} \in \mathbb{R}^s$. As $\mathbf{V}_M^\top \mathbf{V}_{M1} = [\mathbf{I} \quad \mathbf{0}]^\top$, (30) can be partitioned as

$$\frac{d^2 \mathbf{z}_1}{dt^2} + \eta \frac{d\mathbf{z}_1}{dt} = -\mathbf{h}_1 - \mathbf{D}_{M1} \mathbf{z}_1, \quad (31a)$$

$$\frac{d^2 \mathbf{z}_2}{dt^2} + \eta \frac{d\mathbf{z}_2}{dt} = 0, \quad (31b)$$

where the sizes of the partitions should be evident. System (31) was already considered in [16] but the following results are new.

Lemma 3.1. *Let \mathbf{u}_0 and \mathbf{v}_0 be fixed initial conditions. Then a solution of (27) approaches to*

$$\mathbf{u}^* = \mathbf{V}_{M1} \mathbf{D}_{M1}^{-1} \mathbf{V}_{M1}^\top \mathbf{d} + \mathbf{V}_{M2} \mathbf{V}_{M2}^\top \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right). \quad (32)$$

Proof. A solution of (31a)-(31b) is given as

$$\begin{aligned} \mathbf{z}_1(t) &= e^{-\Lambda_1 t} \mathbf{a}_1 + e^{-\Lambda_2 t} \mathbf{a}_2 - \mathbf{D}_{M1}^{-1} \mathbf{h}_1, \\ \mathbf{z}_2(t) &= \mathbf{a}_3 + \mathbf{a}_4 e^{-\eta t}, \end{aligned} \quad (33)$$

with $\mathbf{a}_i, i = 1, \dots, 4$, that are given by the initial conditions,

$$\Lambda_j = \begin{bmatrix} -\eta/2 \pm \sqrt{\eta^2/4 - \gamma_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -\eta/2 \pm \sqrt{\eta^2/4 - \gamma_s} \end{bmatrix}$$

where index j refers to the two different roots. Further, we have

$$\mathbf{u}(t) = \mathbf{V}_M \mathbf{z}(t) = \mathbf{V}_M \begin{bmatrix} \mathbf{z}_1(t) \\ \mathbf{z}_2(t) \end{bmatrix} = \mathbf{V}_{M_1} \mathbf{z}_1(t) + \mathbf{V}_{M_2} \mathbf{z}_2(t). \quad (34)$$

Since $\mathbf{V}_{M_2}^\top \mathbf{V}_{M_1} = \mathbf{0}$ and $\mathbf{V}_{M_2}^\top \mathbf{V}_{M_2} = \mathbf{I}$, we have

$$\mathbf{V}_{M_2}^\top \mathbf{u}(t) = \mathbf{V}_{M_2}^\top \mathbf{V}_{M_1} \mathbf{z}_1(t) + \mathbf{V}_{M_2}^\top \mathbf{V}_{M_2} \mathbf{z}_2(t) = \mathbf{z}_2(t). \quad (35)$$

Inserting the initial condition we conclude that

$$\mathbf{z}_2(t) = \mathbf{V}_{M_2}^\top \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right) - \frac{1}{\eta} \mathbf{V}_{M_2}^\top \mathbf{v}_0 e^{-\eta t},$$

and, therefore,

$$\mathbf{u}^* = \lim_{t \rightarrow \infty} \mathbf{V}_M \mathbf{z}(t) = \mathbf{V}_M \begin{bmatrix} -\mathbf{D}_{M_1}^{-1} \mathbf{h}_1 \\ \mathbf{V}_{M_2}^\top \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right) \end{bmatrix}. \quad (36)$$

After completing the matrix multiplication, we obtain the result. \square

Note that \mathbf{u}^* in (32) consists of two terms lying in perpendicular subspaces. Therefore, a simple way of decreasing the norm of \mathbf{u}^* is to choose $\mathbf{u}_0 + \mathbf{v}_0/\eta = \mathbf{0}$ but that does not necessarily give the smallest \mathbf{w} . Next, we establish which initial conditions would lead to the minimal norm \mathbf{w} .

Given \mathbf{u}^* , we can easily calculate

$$\mathbf{w}^* = \mathbf{V}_{B_2} \mathbf{u}^* + \mathbf{g}. \quad (37)$$

We see from Lemma 3.1 that the stationary solution is dependent on the initial condition \mathbf{u}_0 . Therefore, we are able to choose \mathbf{u}_0 such that the norm of \mathbf{w}^* is as small as possible.

Theorem 3.1. *Let \mathbf{M} be of rank $s \leq r$, \mathbf{w}^* given by (37) and (32), $\mathbf{F} = \mathbf{V}_{B_2} \mathbf{V}_{M_2} \mathbf{V}_{M_2}^\top$, and $\mathbf{f} = \mathbf{V}_{B_2} \left(\mathbf{V}_{M_1} \mathbf{D}_{M_1}^{-1} \mathbf{V}_{M_1}^\top \right) \mathbf{d} + \mathbf{g}$. Then*

$$\min_{(\mathbf{u}_0, \mathbf{v}_0)} \|\mathbf{w}^*\|_2, \quad (38)$$

is attained at

$$\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 = -\mathbf{F}^+ \mathbf{f} \quad (39)$$

and it is equal to

$$\mathbf{w}_+^* = (\mathbf{I} - \mathbf{F}\mathbf{F}^+) \mathbf{f}.$$

In addition, $\mathbf{w}_+^* = \mathbf{w}_{\min}$, where \mathbf{w}_{\min} is given by (17), i.e., \mathbf{w}_+^* solves (16).

Proof. For \mathbf{F} and \mathbf{f} chosen as above and (32), we have

$$\begin{aligned}\mathbf{w}^* &= \mathbf{V}_{\mathbf{B}_2} \left(\mathbf{V}_{\mathbf{M}_1} \mathbf{D}_{\mathbf{M}_1}^{-1} \mathbf{h} + \mathbf{V}_{\mathbf{M}_2} \mathbf{V}_{\mathbf{M}_2}^\top (\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0) \right) + \mathbf{g} \\ &= \mathbf{F} \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right) + \mathbf{f}.\end{aligned}\tag{40}$$

The norm is then

$$\|\mathbf{w}^*\|_2 = \left\| \mathbf{F} \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right) + \mathbf{f} \right\|_2.\tag{41}$$

Consequently, we get (39) and calculate

$$\mathbf{w}_+^* = -\mathbf{F} \left(\mathbf{u}_0 + \frac{1}{\eta} \mathbf{v}_0 \right) + \mathbf{f} = \mathbf{F}\mathbf{F}^+ \mathbf{f} - \mathbf{f} = (\mathbf{I} - \mathbf{F}\mathbf{F}^+) \mathbf{f}.$$

To prove the last claim, we notice that \mathbf{w}^* lies in the null space of $\mathbf{\Sigma}$. Furthermore, it satisfies $\mathbf{B}\mathbf{w} = \mathbf{c}$ by construction. Since, \mathbf{w}_+^* is minimal, it solves (16). It completes the proof of the theorem. \square

Remark 3.1. From Theorem 3.1 one choice of the initial condition is

$$\mathbf{u}_0 = -\mathbf{F}^+ \mathbf{f} \text{ and } \mathbf{v}_0 = \mathbf{0},$$

which we will assume for simplicity further on.

While Theorem 3.1 gives a way to find an optimally small \mathbf{w}^* it also requires the SVD of $\mathbf{\Sigma}$ and \mathbf{M} which increase the computational complexity considerably. However, we would like to emphasise that any choice of \mathbf{u}_0 and \mathbf{v}_0 results in a risk-free solution. Indeed, numerical tests in Section 4 show that DFPM is a competitive method in many cases.

3.2 Symplectic Euler with optimal parameters

In order to solve (27) through an numerical iterative method, we start by writing (27) as the first order system

$$\begin{aligned}\frac{d\mathbf{u}}{dt} &= \mathbf{v}, \\ \frac{d\mathbf{v}}{dt} &= -\eta\mathbf{v} - \mathbf{d} - \mathbf{M}\mathbf{u}.\end{aligned}\tag{42}$$

with $\mathbf{u}(0) = \mathbf{u}_0$ and $\mathbf{v}(0) = \mathbf{0}$. The latter is assumed for simplicity and does not affect further analysis, see also Remark 3.1.

We first note that (42) may be numerically solved with a small error using any stable and consistent numerical method if the time step Δt is small enough. However, we are only interested in reaching the stationary solution

as efficiently as possible. As shown in [15], an excellent choice for this purpose is the symplectic Euler method. Thus, for some initial conditions $\mathbf{u}(0) = \mathbf{u}_0$ and $\dot{\mathbf{u}}(0) = \mathbf{v}_0$ we use the symplectic Euler method on (42) defined as

$$\begin{aligned}\mathbf{v}_{k+1} &= (I - \Delta t \eta) \mathbf{v}_k - \Delta t (\mathbf{d} + \mathbf{M} \mathbf{u}_k), \\ \mathbf{u}_{k+1} &= \mathbf{u}_k + \Delta t \mathbf{v}_{k+1}, \quad k = 0, 1, \dots,\end{aligned}\tag{43}$$

or more concisely $\boldsymbol{\xi}_{k+1} = \mathbf{G} \boldsymbol{\xi}_k + \mathbf{b}$, where

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} - \Delta t^2 \mathbf{M} & \Delta t (1 - \Delta t \eta) \mathbf{I} \\ \Delta t \mathbf{I} & (1 - \Delta t \eta) \mathbf{I} \end{bmatrix}, \quad \boldsymbol{\xi}_k = \begin{bmatrix} \mathbf{v}_k \\ \mathbf{u}_k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -\Delta t^2 \mathbf{d} \\ -\Delta t \mathbf{d} \end{bmatrix}.\tag{44}$$

We emphasize that symplectic methods are very well suited for the case when \mathbf{M} has non-negative eigenvalues, see [15]. In order to ensure fast convergence of the method one must choose a time step Δt and damping factor η such that $\|\mathbf{G}\|$ is as small as possible.

Theorem 3.2. *Assume that \mathbf{M} has rank s and let $\gamma_i, i = 1, \dots, s$, denote its non-zero singular values with $\gamma_1 \geq \dots \geq \gamma_s$. Then symplectic Euler (44) with parameters*

$$\Delta t = \frac{2}{\sqrt{\gamma_s} + \sqrt{\gamma_1}}, \quad \eta = \frac{2\gamma_s \gamma_1}{\sqrt{\gamma_s} + \sqrt{\gamma_1}}\tag{45}$$

is convergent, i.e.,

$$\lim_{k \rightarrow \infty} \mathbf{u}_k = -\mathbf{V}_{\mathbf{M}1} \mathbf{D}_{\mathbf{M}1}^{-1} \mathbf{V}_{\mathbf{M}1}^\top \mathbf{d} + \mathbf{V}_{\mathbf{M}2} \mathbf{V}_{\mathbf{M}2}^\top \mathbf{u}_0.\tag{46}$$

Moreover, Δt and η are optimal in the sense that they are the solution to the problem

$$\min_{\Delta t, \eta} \max_{1 \leq i \leq 2p} |\mu_i(\mathbf{G})|,$$

where $\mu_i(\mathbf{G})$ are the eigenvalues of \mathbf{G} .

Proof. By using (28), we define the transformations $\tilde{\mathbf{u}}_k = \mathbf{V}_{\mathbf{M}} \mathbf{u}_k$ and $\tilde{\mathbf{v}}_k = \mathbf{V}_{\mathbf{M}} \mathbf{v}_k$ resulting into $\tilde{\boldsymbol{\xi}}_{k+1} = \tilde{\mathbf{G}} \tilde{\boldsymbol{\xi}}_k + \tilde{\mathbf{b}}$, where

$$\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{V}_{\mathbf{M}}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\mathbf{M}}^\top \end{bmatrix} \mathbf{G} \begin{bmatrix} \mathbf{V}_{\mathbf{M}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{\mathbf{M}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \Delta t^2 \mathbf{D}_{\mathbf{M}} & \Delta t (1 - \Delta t \eta) \mathbf{I} \\ \Delta t \mathbf{D}_{\mathbf{M}} & (1 - \Delta t \eta) \mathbf{I} \end{bmatrix},$$

and obvious similar definitions of $\tilde{\boldsymbol{\xi}}_{k+1}, \tilde{\mathbf{b}}$. Since, $\gamma_{s+1} = \dots = \gamma_{p-2} = 0$ we can partition $\tilde{\mathbf{G}}$ further using the reduced SVD of \mathbf{M} so that

$$\tilde{\mathbf{G}} = \begin{bmatrix} \mathbf{I} - \Delta t^2 \mathbf{D}_{\mathbf{M}1} & \mathbf{0} & \Delta t (1 - \Delta t \eta) \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \Delta t (1 - \Delta t \eta) \mathbf{I} \\ \Delta t \mathbf{D}_{\mathbf{M}1} & \mathbf{0} & (1 - \Delta t \eta) \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (1 - \Delta t \eta) \mathbf{I} \end{bmatrix}.$$

Due to the structure of $\tilde{\mathbf{G}}$, we can solve for the last $p - 2 - s$ components of $\tilde{\mathbf{u}}_k$, denoted by $\tilde{\mathbf{u}}_2^{(k)}$, and eliminate the corresponding elements in $\tilde{\mathbf{v}}_k$ (that will play no further role). In other words the iterations $\tilde{\boldsymbol{\xi}}_{k+1} = \tilde{\mathbf{G}}\tilde{\boldsymbol{\xi}}_k + \tilde{\mathbf{b}}$ will separate into one part with the iteration matrix

$$\tilde{\mathbf{G}}_1 = \begin{bmatrix} \mathbf{I} - \Delta t^2 \mathbf{D}_{M1} & \Delta t(1 - \Delta t \eta) \mathbf{I} \\ \Delta t \mathbf{D}_{M1} & (1 - \Delta t \eta) \mathbf{I} \end{bmatrix}$$

and the constant part $\tilde{\mathbf{u}}_2^{(k)}$ now to be determined. We have

$$\mathbf{u}_k = \mathbf{V}_M \tilde{\mathbf{u}}^{(k)} = \mathbf{V}_M \begin{bmatrix} \tilde{\mathbf{u}}_1^{(k)} \\ \tilde{\mathbf{u}}_2^{(k)} \end{bmatrix} = \mathbf{V}_{M1} \tilde{\mathbf{u}}_1^{(k)} + \mathbf{V}_{M2} \tilde{\mathbf{u}}_2^{(k)} \quad (47)$$

that gives

$$\mathbf{V}_{M2}^\top \tilde{\mathbf{u}}^{(k)} = \mathbf{V}_{M2}^\top \mathbf{V}_{M1} \tilde{\mathbf{u}}_1^{(k)} + \mathbf{V}_{M2}^\top \mathbf{V}_{M2} \tilde{\mathbf{u}}_2^{(k)} = \tilde{\mathbf{u}}_2^{(k)} \quad (48)$$

since $\mathbf{V}_{M2}^\top \mathbf{V}_{M1} = \mathbf{0}$ and $\mathbf{V}_{M2}^\top \mathbf{V}_{M2} = \mathbf{I}$. Inserting the initial condition we conclude that $\tilde{\mathbf{u}}_2^{(k)} = \mathbf{V}_{M2}^\top \mathbf{u}_0$ and, therefore,

$$\begin{aligned} \mathbf{u}^* &= \lim_{k \rightarrow \infty} \mathbf{V}_M \tilde{\mathbf{u}}^{(k)} = \lim_{k \rightarrow \infty} \left(\mathbf{V}_{M1} \tilde{\mathbf{u}}_1^{(k)} + \mathbf{V}_{M2} \tilde{\mathbf{u}}_2^{(k)} \right) \\ &= \mathbf{V}_{M2} \mathbf{V}_{M2}^\top \mathbf{u}_0 + \lim_{k \rightarrow \infty} \mathbf{V}_{M1} \tilde{\mathbf{u}}_1^{(k)}. \end{aligned} \quad (49)$$

Following already known results in [15], we get

$$\lim_{k \rightarrow \infty} \mathbf{V}_{M1} \tilde{\mathbf{u}}_1^{(k)} = -\mathbf{V}_{M1} \mathbf{D}_{M1}^{-1} \mathbf{V}_{M1}^\top \mathbf{d}$$

which concludes the proof. \square

Attaining estimates of γ_1/γ_s and the rank for the optimal parameters can be done by standard condition and rank estimators at a cost of $\mathcal{O}(p^2)$, [26].

4 Simulation Study

There are several aspects that we illustrate in this simulation study. In Section 4.1 we compare the outcomes of three different approaches of finding a portfolio, namely, the analytic Range Space (RS), see (3), and Null Space Minimum Norm (NSMN), see (17), approaches, and the iterative DFPM with zero initial value $\mathbf{u}_0 = 0$, see Section 3. In particular, we investigate how the portfolio risk and portfolio weights' norm vary for each method depending on the problem size, rank of $\boldsymbol{\Sigma}$, and its condition number. In Section 4.2 we touch upon robustness of these different approaches in relation to rank underestimation due to the sampling of $\boldsymbol{\Sigma}$.

The setup for the simulation study is as follows. Each element of the mean vector $\boldsymbol{\mu}$ is drawn from the uniform distribution on the interval $[-0.01, 0.01]$. The expected rate of return on the portfolio q is equal to the return on a naive equal-weighted (EW) portfolio, i.e., $q = \mathbf{1}^\top \boldsymbol{\mu}/p$. The population covariance matrix $\boldsymbol{\Sigma}$ of rank r was obtained as $\boldsymbol{\Sigma} = \sum_{i=1}^r \mathbf{x}_i \mathbf{x}_i^\top$, where $\mathbf{x}_i \in N(\mathbf{0}, \mathbf{I}), i = 1, \dots, r$, i.e., $\mathbf{x}_i \in \mathbb{R}^p$ are normally distributed vectors with mean vector $\mathbf{0}$ and covariance matrix \mathbf{I} . In Section 4.1, the condition number of $\boldsymbol{\Sigma}$, defined by $\kappa = \sigma_1/\sigma_r$, was then set through a similarity transformation.

In order to define the convergence criterium we set $V(\mathbf{u}) = \mathbf{u}^\top \mathbf{M} \mathbf{u}/2 + \mathbf{u}^\top \mathbf{d}$ and $\nabla V(\mathbf{u}) = \mathbf{M} \mathbf{u} + \mathbf{d}$. Then, given a tolerance ε we end the iterations at iteration k if $\|\nabla V(\mathbf{u}_k)\|_2/V(\mathbf{u}_k) < \varepsilon$. Here, we set the tolerance $\varepsilon = 10^{-6}$. In accordance with [15], the number of iterations is proportional to the square root of the condition number, which is observed in all the numerical tests.

In order to estimate method performances, we solve 1000 problems with random matrices $\boldsymbol{\Sigma}$ for a specific set of p, r and κ , and random $\boldsymbol{\mu}$. The results presented as scatter- and box-plots. In addition we provide the tables with mean estimates for the norm and the risk.

4.1 Comparison of the RS, NSMN, and DFPM solutions

Here, we solve 1000 problems (1) with random matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ for a specific set of p, r and κ with three different approaches: The RS and NSMN methods, and DFPM. For the DFPM we set the initial value $\mathbf{u}_0 = \mathbf{0}$. Thus, the method provides an approximation of a null space solution (and, therefore, risk-free solution), but is not expected to converge to the NSMN solution.

We consider four different cases of the condition number $\kappa = \sigma_1/\sigma_r \in \{1.01, 10^2, 10^3, 10^5\}$. The dimension of the portfolio, p , was set to 70 and 140, while the rank r of the covariance matrix $\boldsymbol{\Sigma}$ is 60. More tests were performed with other choices of p and r . However, as the results were very similar to the ones presented here, we do not include them.

In Figures 1 and 2 we plot the portfolio weights' norm vs the portfolio risk for different solutions types and different condition numbers. In Figures 3 and 4 the norm and the corresponding risk as box plots are displayed. The red boxes correspond to the NSMN solution, yellow to the DFPM solutions, and blue to the RS solutions.

Our first observation is that the risk - norm distributions corresponding to the NSMN and DFPM solutions are not sensitive to the condition number κ , which is not the case for the RS approach. For the RS solution, the larger condition number gives a lower risk and higher norm variance. Secondly, the norm distribution for the NSMN approach and DFPM is flatter for a lower p to r ratio ($p = 70, r = 60$), see Figure 1.

Figure 3 illustrates well the trade off between getting a smaller norm and smaller risk. While the the RS solution norm is the smallest, among the approaches, for the lower p to r ratio ($p = 70, r = 60$), the risk is the

highest. This is rather expected, as the NSMN solution minimizes the norm among zero-risk solutions and the DFPM solution is only approximating a risk-free solution without minimizing the norm. Interestingly, for the high p to r ratio ($p = 140$, $r = 60$), see Figure 4, the NSMN solution gives both the smallest norm and zero risk. The DFPM solution norm is typically larger than the norm of the NSMN solution, which is expected.

In Table 1, the portfolio weights' norm and portfolio risk are shown as the mean of the 1000 randomly generated problems with $p = 70$ and $r = 60$ for four different cases of κ . Similarly, in Table 2 we list the mean estimation of the norm and risk for $p = 140$ and $r = 60$, for the four different κ . We would like to point out here, that the risk for both DFPM and NSMN is close to zero and of the order of computational accuracy (floating double precision). For that reason, the computed risk can be negative, see Tables 1 and 2. In order to use logarithmic scale in the plots, we plot absolute values of the computed risk in Figures 3 and 2.

In Figure 5 we give an example of the portfolio weights' norm and portfolio risk for the DFPM iterations. Here, the smallest norm is not attained at the stationary point, but at an intermediate iteration where the risk is rather high. Further note that there is a consistent relation between the norm of weights and risks for the different initial values; blue is right of black is right of red. From other numerical experiments, this seems to be generally the case but we have not been able to prove that.

4.2 How rank deficiency of Σ effects different solutions

Let Σ have the rank $r < p - 2$. The reason for the rank deficiency could be (i) low sampling number $N = r + 1$ compared to the portfolio size, or (ii) collinearity of data samples.

Thus, assume that Σ_r is a covariance matrix obtained from the limited data, and Σ_R is the true covariance matrix, then $\Sigma_r \rightarrow \Sigma_R$ if the number of samples $N \rightarrow \infty$. Here $r \leq R \leq p$, and $r < p - 2$. In another words, the matrix $\Sigma = \Sigma_r$ in (1) is an (under-sampled) approximation of Σ_R , and neither \mathbf{w} in (3) nor (17) are true solutions of the problem when $\Sigma = \Sigma_R$.

Let \mathbf{w}_r denote a solution to (1) with $\Sigma = \Sigma_r$. We computed the risk associated with \mathbf{w}_r , $\mathbf{w}_r^\top \Sigma_R \mathbf{w}_r$ to see if any of the RS, NSMN or DFPM solutions consistently give higher risk.

Here, the matrices Σ_r and Σ_R obtained as follows. Assume $\mathbf{x}_i \in N(\mathbf{0}, \mathbf{I})$, $i = 1, \dots, R$, i.e., $\mathbf{x}_i \in \mathbb{R}^p$ are normally distributed vectors with mean vector $\mathbf{0}$ and covariance matrix \mathbf{I} . Here we perform a simulation study where $\Sigma_r = \sum_{i=1}^r \mathbf{x}_i \mathbf{x}_i^\top$ and $\Sigma_R = \Sigma_r + \sum_{i=r+1}^R \mathbf{x}_i \mathbf{x}_i^\top$. We do not monitor the condition numbers for Σ_r and Σ_R since we do not consider it important enough to perform the theoretical derivation for that, which can be quite elaborate. In Figure 6, we plot the risk associated with different solutions for $r = 60$, $p = 140$, and varying R . We chose $R \in [61, 80]$ to illustrated

what seems to be a generic case for this construction of Σ_r and Σ_R . That is, solution obtained using the NSMN theoretical approach gives the smallest risk that other methods as for any R . When $R - r$ is small, DFPM method with zero initial guess gives better results, but as R increases, the method becomes less reliable. Finally, as R grows, the risk is increasing for all considered methods.

In Figure 7, we plot the relative risk, that is, $\mathbf{w}_r^\top \Sigma_R \mathbf{w}_r / \|\mathbf{w}_r\|^2$. This plot illustrates that the higher risk obtained using the DFPM seems to be due to the larger magnitude of the weights \mathbf{w} .

We conclude that all methods do have a predictive capacity, i.e., the original weight vector does not give a much larger risk when adding non-correlated data.

5 Empirical Study

To get a better understanding of the results obtained in Sections 2 and 3, we consider examples with real datasets. In this study, we compare the performance of NSMN, DFPM, RS and EW portfolios. We make use of the log return weekly data from S&P 500 of 250 stocks for the period from 2016W34 to 2021W11. The analysis is performed for different sample sizes such as $N \in \{30, 60, 120, 240\}$. The expected rate of return on the portfolio q is set to the return on the equal-weighted portfolio. For the DFPM, we set the same the convergence criterion, the tolerance and the maximum number of iterations as in Section 4.

Since the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of the logarithmic stock returns are unknown, we should estimate them using historical data. For that, we use the corresponding sample estimators. In this case, the sample estimator of the covariance matrix $\boldsymbol{\Sigma}$ is singular because the number of stocks $p = 250$ is greater than the number of observations in the sample $N \in \{30, 60, 120, 240\}$. Having estimated $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we can build optimal portfolio strategies using NSMN, DFPM and RS methods. These methods are compared with naive EW portfolio strategy. In Table 3, we present the values of the portfolio weights' norm and portfolio risk for four approaches. The portfolio weights' norm of RS and EW approaches is lower than for NSMN and DFPM for all considered sample sizes. The highest portfolio weights' norm is observed for DFPM except for the case $N = 240$ where the NSMN approach has a slightly higher norm. For the portfolio risk, solutions obtained via NSMN and DFPM approaches deliver much lower portfolio risk in comparison with RS and EW methods. Moreover, NSMN and DFPM attain risks close to zero for all considered cases. It can be observed that some values of the portfolio risk obtained via the NSMN approach are negative. This happens due to numerical errors. Overall, NSMN and DFPM strategies have much lower portfolio risk and higher portfolio weight's norm in

comparison with RS and EW strategies.

6 Conclusions

We have analyzed the classical Markowitz' problem for a rank-deficient covariance matrix assuming the solutions lie in the nullspace. Thus, the solutions will be risk-free but not unique. We derive a unique risk-free solution that has minimum norm through the use of several SVD decompositions which is our main theoretical result. Our analysis then fills a gap so that all possible relevant solutions of the classical Markowitz' problem are derived and analysed. We also show the dependence on the size of the risk and the norm of the weights depending on the condition number of the covariance matrix (quotient between largest and smallest non-zero singular value). Finally, we illustrate the fairly strong predictive capacity of our methods, i.e., in most cases they still give solutions with a smaller risk than the range space solution when adding more (non-correlated) data to the model.

The other part of our results is a thorough analysis of the iterative method DFPM for the rank-deficient case. This was partly covered in [16] but without a strict convergence analysis. Here, we show that DFPM always converges to a risk-free solution. Further, we derive an initial condition such that DFPM converges to the unique minimum norm solution. Numerical results verify the theoretical findings through simulations on random generated problems as well as real world data. Moreover, the numerical results show several advantages using DFPM such as being more efficient for fairly well conditioned problems and the possibility to pick any one of the iterates that for some reason has preferable properties.

References

- [1] Alfelt, G., & Mazur, S. (2020). On the mean and variance of the estimated tangency portfolio weights for small samples. Working paper, 8, Örebro University School of Business.
- [2] Best, M. J. (2017). *Quadratic Programming with Computer Programs*. Chapman and Hall/CRC.
- [3] Bauder, D., Bodnar, T., Mazur, S., & Okhrin, Y. (2018). Bayesian inference for the tangent portfolio. *International Journal of Theoretical and Applied Finance*, 21(8), 1850054.
- [4] Bodnar, T., Mazur, S., Muhinyuza, S., & Parolya, N. (2018). On the product of a singular Wishart matrix and a singular Gaussian vector in high dimension. *Theory of Probability and Mathematical Statistics*, 99(2), 37–50.

- [5] Bodnar, T., Mazur, S., & Okhrin, Y. (2013). On the exact and approximate distributions of the product of a Wishart matrix with a normal vector. *Journal of Multivariate Analysis*, 122, 70–81.
- [6] Bodnar, T., Mazur, S., & Okhrin, Y. (2014). Distribution of the product of singular Wishart matrix and normal vector. *Theory of Probability and Mathematical Statistics*, 91, 1–15.
- [7] Bodnar, T., Mazur, S., & Okhrin, Y. (2017). Bayesian estimation of the global minimum variance portfolio. *European Journal of Operational Research*, 256(1), 292–307.
- [8] Bodnar, T., Mazur, S., & Parolya, N. (2019). Central limit theorems for functionals of large dimensional sample covariance matrix and mean vector in matrix-variate skewed model. *Scandinavian Journal of Statistics*, 46(2), 636–660.
- [9] Bodnar, T., Mazur, S., & Podgórski, K. (2016). Singular inverse Wishart distribution and its application to portfolio theory. *Journal of Multivariate Analysis*, 143, 314–326.
- [10] Bodnar, T., Mazur, S., & Podgórski, K. (2017). A test for the global minimum variance portfolio for small sample and singular covariance. *AStA Advances in Statistical Analysis*, 101, 253–265.
- [11] Bodnar, T., Mazur, S., Podgórski, K., & Tyrcha, J. (2019). Tangency portfolio weights in small and large dimensions: Estimation and test theory. *Journal of Statistical Planning and Inference*, 201, 40–57.
- [12] Bodnar, T., & Okhrin, Y. (2011). On the product of inverse Wishart and normal distributions with applications to discriminant analysis and portfolio theory. *Scandinavian Journal of Statistics*, 38, 311–331.
- [13] Bodnar, T., & Schmid, W. (2011). On the exact distribution of the estimated expected utility portfolio weights: Theory and applications. *Statistics & Risk Modeling*, 28, 319–342.
- [14] Edvardsson, S., Neuman, M., Edström, P., & Olin, H. (2015). Solving equations through particle dynamics. *Computer Physics Communications*, 197, 169–181.
- [15] Gulliksson, M., Ögren, M., Oleynik, A., & Zhang, Y. (2018). *Damped Dynamical Systems for Solving Equations and Optimization Problems*. In: Sriraman B. (eds) Handbook of the Mathematics of the Arts and Sciences. Springer, Cham.
- [16] Gulliksson, M., & Mazur, S. (2020). An iterative approach to ill-conditioned optimal portfolio selection. *Computational Economics*, 56, 773–794.

- [17] Harier, E., Lubich, C., & Wanner, G. (2006). *Geometrical Numerical Integration*. Springer.
- [18] Lundgren, J., Rönqvist, M., & Värbrand, P. (2010). *Optimeringslära*. Studentlitteratur.
- [19] Javed, F., Mazur, S., & Ngailo, E. (2021). Higher order moments of the estimated tangency portfolio weights. *Journal of Applied Statistics*, 48(3), 517–535.
- [20] Kotsiuba, I., & Mazur, S. (2015). On the asymptotic and approximate distributions of the product of an inverse Wishart matrix and a Gaussian random vector. *Theory of Probability and Mathematical Statistics*, 93, 95–104.
- [21] Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7, 77–91.
- [22] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.
- [23] Pappas, D., Kiriakopoulos, K., & Kaimakamis, G. (2010). Optimal portfolio selection with singular covariance matrix. *International Mathematical Forum*, 5(47), 2305—2318.
- [24] Trefethen, L. N., & Bau III, D. (1997). *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia.
- [25] Okhrin, Y., & Schmid, W. (2006). Distributional properties of portfolio weights. *Journal of Econometrics*, 134, 235–256.
- [26] Hager, W. W. (1984). Condition estimates. *SIAM Journal on Scientific and Statistical Computing*, 5(2), 311–316.

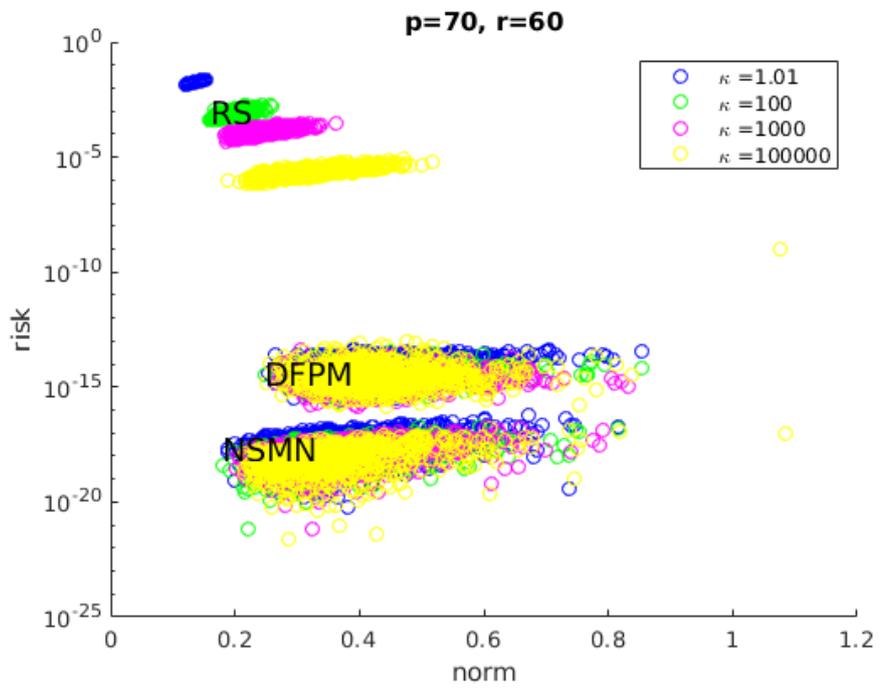


Figure 1: Scatter plot of portfolio weights' norm vs portfolio risk for three different approaches, with $p = 70$, $r = 60$, and four different condition numbers.

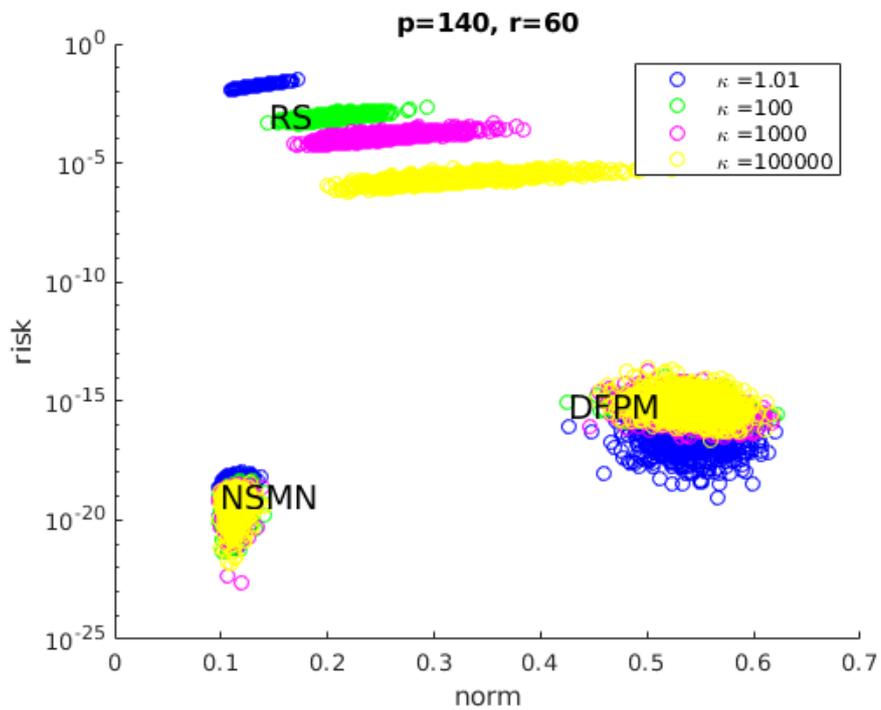


Figure 2: Scatter plot of portfolio weights' norm vs portfolio risk for three different approaches, with $p = 140$, $r = 60$, and four different condition numbers.

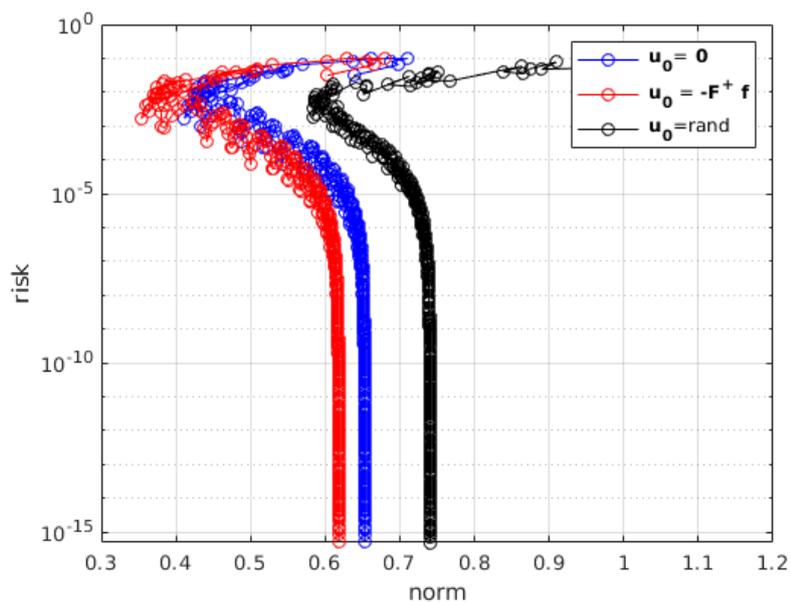


Figure 5: Iterates from DFPM with three different choices of the initial conditions for $p = 70$, $r = 60$ and $\kappa = 10^3$.

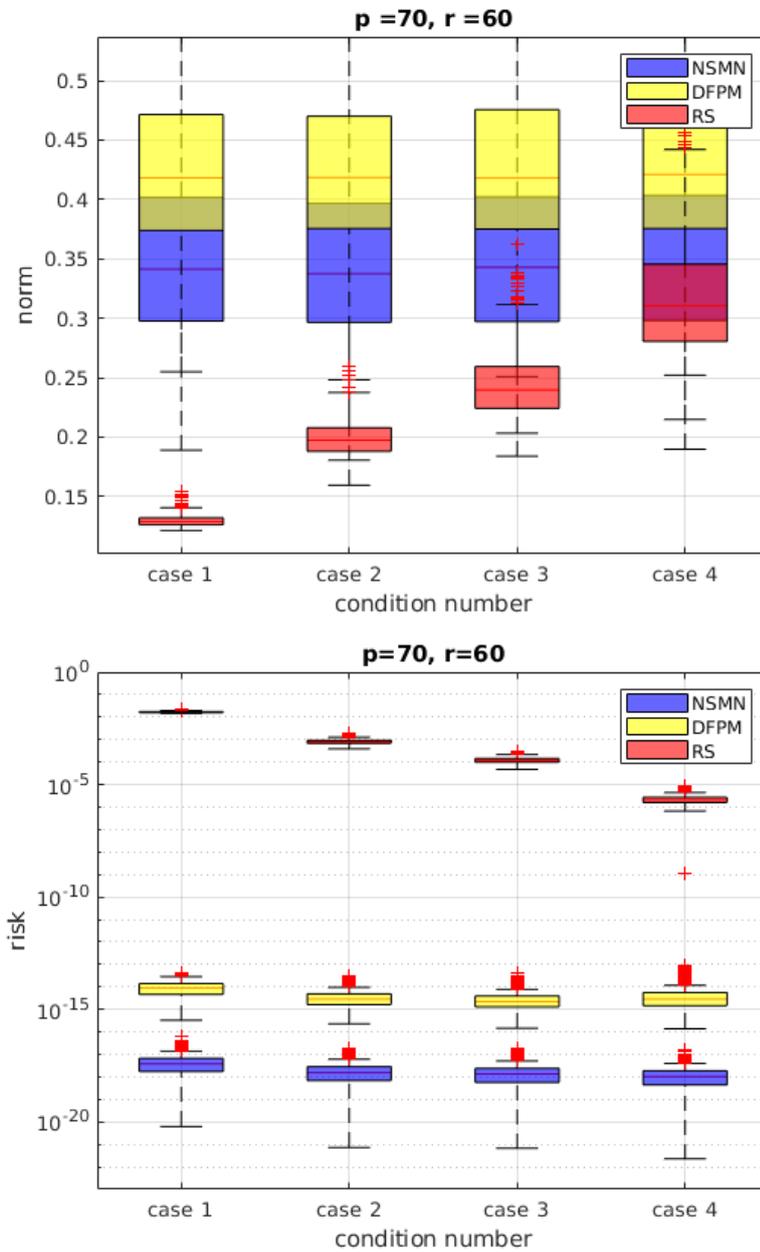


Figure 3: Boxplot of the portfolio weights' norm and related portfolio risk, for the test case I. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

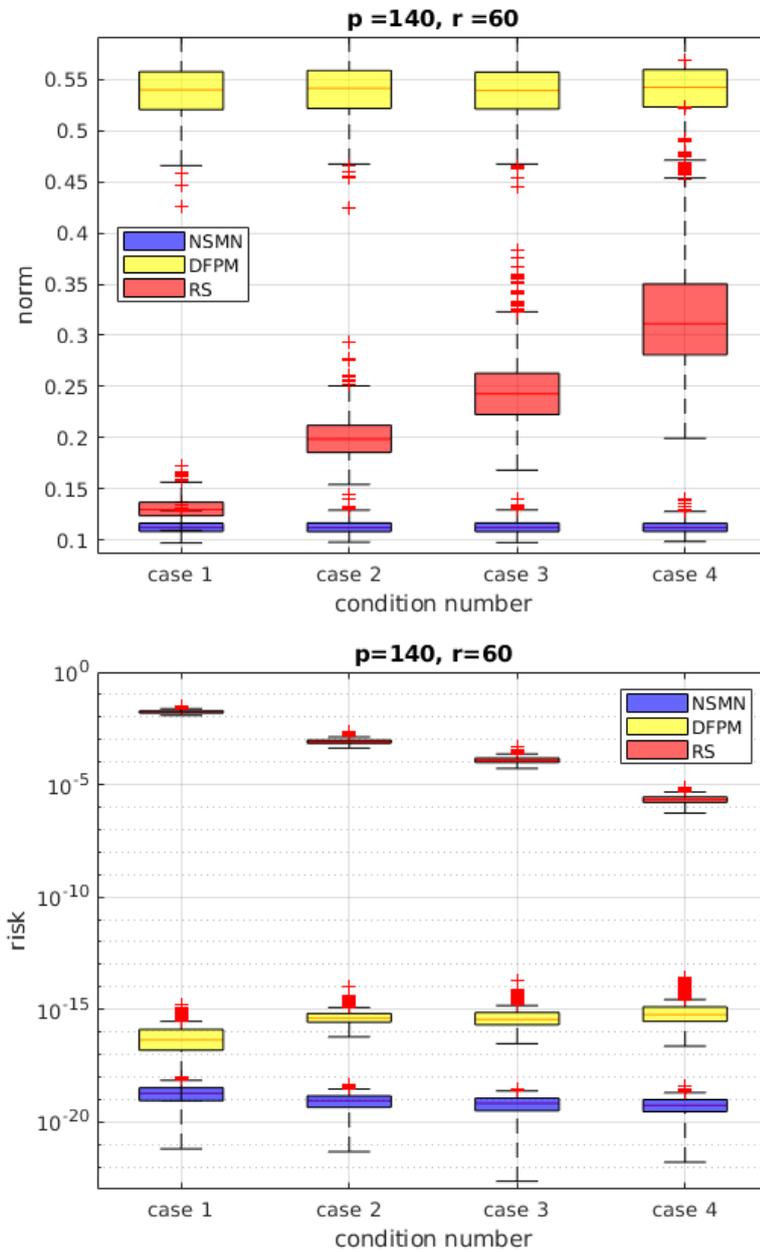


Figure 4: Boxplot of the portfolio weights' norm and related portfolio risk, for the test case II. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the '+' symbol.

	Norm			Risk		
	NSMN	DFPM	RS	NSMN	DFPM	RS
case 1 ($\kappa = 1.01$)	3.58×10^{-1}	4.32×10^{-1}	1.29×10^{-1}	2.95×10^{-19}	1.04×10^{-14}	1.67×10^{-2}
case 2 ($\kappa = 10^2$)	3.57×10^{-1}	4.31×10^{-1}	1.99×10^{-1}	1.95×10^{-19}	3.84×10^{-15}	8.20×10^{-4}
case 3 ($\kappa = 10^3$)	3.60×10^{-1}	4.31×10^{-1}	2.43×10^{-1}	-1.09×10^{-19}	3.27×10^{-15}	1.26×10^{-4}
case 4 ($\kappa = 10^5$)	3.60×10^{-1}	4.33×10^{-1}	3.16×10^{-1}	4.77×10^{-20}	1.06×10^{-12}	2.26×10^{-6}

Table 1: Mean estimation (out of 1000 random simulations) of the portfolio weights' norm and the corresponding portfolio risk, for different condition numbers. Here $p = 70$ and $r = 60$.

	Norm			Risk		
	NSMN	DFPM	RS	NSMN	DFPM	RS
case 1 ($\kappa = 1.01$)	1.13×10^{-1}	5.40×10^{-1}	1.31×10^{-1}	-6.09×10^{-21}	1.05×10^{-16}	1.71×10^{-2}
case 2 ($\kappa = 10^2$)	1.13×10^{-1}	5.40×10^{-1}	2.00×10^{-1}	4.84×10^{-21}	5.64×10^{-16}	8.33×10^{-4}
case 3 ($\kappa = 10^3$)	1.12×10^{-1}	5.39×10^{-1}	2.46×10^{-1}	-1.14×10^{-21}	6.63×10^{-16}	1.30×10^{-4}
case 4 ($\kappa = 10^5$)	1.12×10^{-1}	5.41×10^{-1}	3.19×10^{-1}	-1.59×10^{-22}	1.16×10^{-15}	2.34×10^{-6}

Table 2: Mean estimation (out of 1000 random simulations) of the portfolio weights' norm and the corresponding portfolio risk, for different condition numbers. Here $p = 140$ and $r = 60$.

N	Norm				Risk			
	NSMN	DFPM	RS	EW	NSMN	DFPM	RS	EW
30	0.2070	0.8580	0.1689	0.0632	-8.2309×10^{-21}	1.6626×10^{-18}	2.9360×10^{-4}	5.1693×10^{-4}
60	0.3534	0.7191	0.2699	0.0632	4.2065×10^{-20}	6.2424×10^{-17}	2.1292×10^{-4}	0.0020
120	0.5728	0.6665	0.4813	0.0632	-1.0835×10^{-19}	1.1570×10^{-14}	6.2364×10^{-5}	0.0011
240	3.9050	3.8987	2.1748	0.0632	-5.6200×10^{-20}	2.0725×10^{-11}	3.8010×10^{-6}	6.4904×10^{-4}

Table 3: Norm and risk of $p = 250$ stocks and different sample sizes $N \in \{30, 60, 120, 240\}$ obtained using NSMN, DFPM, RS and EW approaches.

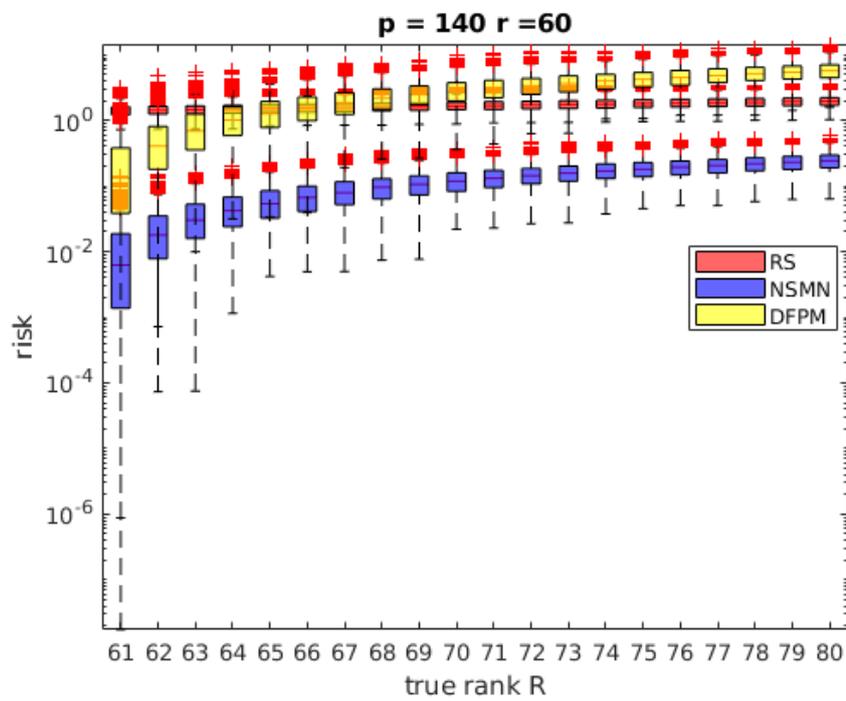


Figure 6: Risk $\mathbf{w}_r^\top \Sigma_R \mathbf{w}_r$ calculated with \mathbf{w}_r obtained using different methods: RS, NSMN, and DFPM. Here $r = 60$, $p = 140$, and R varies between 61 and 80. Number of random simulations is 1000.

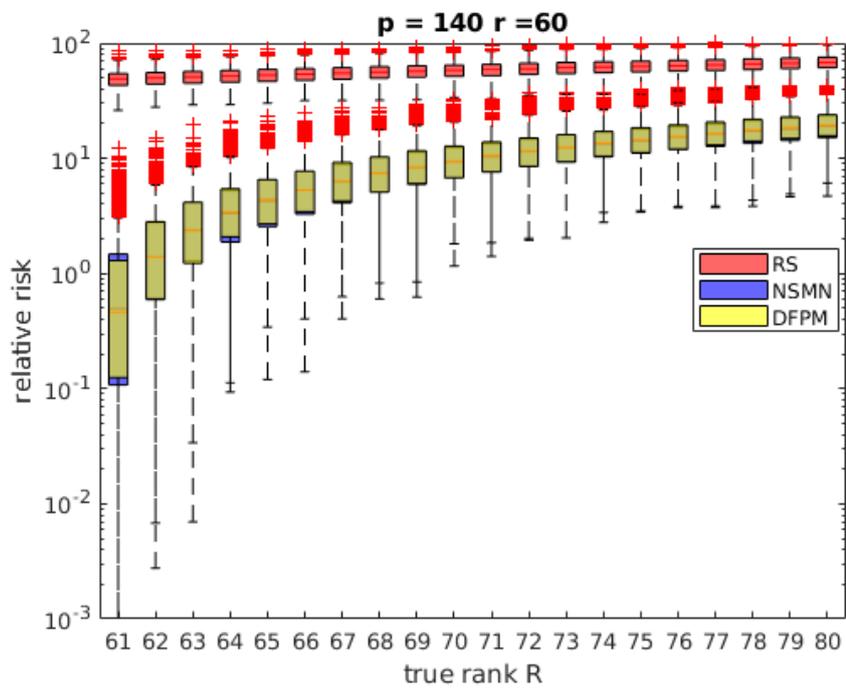


Figure 7: Relative risk $\mathbf{w}_r^\top \Sigma_R \mathbf{w}_r / \|\mathbf{w}_r\|^2$ calculated with \mathbf{w}_r obtained using different methods: RS, NSMN, and DFPM. Here $r = 60$, $p = 140$, and R varies between 61 and 80. Number of random simulations is 1000.