Rayleigh Portfolios and Penalised Matrix Decomposition

Francois Buet-Golfouse
ucahfbu@ucl.ac.uk
University College London
London, United Kingdom

Hans Roggeman
Islam Utyagulov
hraoyama@gmail.com
islam.utyagulov@gmail.com

ABSTRACT

Since the development and growth of personalised financial services online, effective tailor-made and fast statistical portfolio allocation techniques have been sought after. In this paper, we introduce a framework called *Rayleigh portfolios*, that encompasses many well-known approaches, such as the Sharpe Ratio, maximum diversification or minimum concentration. By showing the commonalities amongst these approaches, we are able to provide a solution to all such optimisation problems via matrix decomposition, and principal component analysis in particular. In addition, thanks to this reformulation, we show how to include sparsity upper bounds in such portfolios, thereby catering for two additional requirements in portfolio construction: robustness and low transaction costs. Importantly, modifications to the usual penalised matrix decomposition algorithms can be applied to other problems in statistics. Finally, empirical applications show promising results.

CCS CONCEPTS

• **Computing methodologies → Dimensionality reduction and manifold learning**.

KEYWORDS

portfolio allocation, generalized PCA, generalized low rank models

ACM Reference Format:
Francois Buet-Golfouse, Hans Roggeman, and Islam Utyagulov. 2022. Rayleigh Portfolios and Penalised Matrix Decomposition. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3487553.3524634

1 INTRODUCTION

Asset allocation generally deals with broad asset classes, such as stocks, bonds and cash (but investors may also wish to include credit default swaps or commodities), which can be fine-tuned to account for sectors (e.g., utilities, financials, etc.), countries (e.g., U.S., U.K. or emerging countries) and capitalisations (small vs large stocks for instance). Markowitz's 1952 mean–variance optimisation [9] quickly became a staple of financial investment and an integral part of financial theory, by taking into account securities' expected returns but also their covariance structure. The key insight was that for a desired level of expected reward, one should minimise the risk incurred to achieve it. By accounting for returns' randomness and correlations, Markowitz laid the foundations of the risk-reward analysis of investment choices under uncertainty.

Different approaches to portfolio construction and optimisation have been proposed [3, 6, 10, 11], to which the reader is referred for an overview and textbook treatment. Our purpose is somewhat more limited and consists of considering a particular class of portfolio metrics that can be construed as explicit risk-reward trade-offs. In particular, we focus on measures such as the Sharpe ratio [8] or maximum diversification [2], which –as we show– can be recast as a generalized principal component analysis. While this is a new and interesting result in itself, it also opens the door to several extensions of, say, the maximum diversification portfolio, the main one being a *sparse* maximum diversification portfolio we propose. For reasons we make clear in the text, we call this category of problems *sparse Rayleigh portfolios*.

Our contributions.

(1) We show that multiple well-known allocation criteria used in practice can be recast as particular instances of a broader framework. This is useful for a better understanding of what each particular choice may entail.
(2) We indicate that this framework can be tackled via matrix decomposition methods, similar to principal component analysis.
(3) We add sparsity upper bounds as a way of controlling total exposure and increasing robustness.
(4) We propose a biconvex algorithm based on penalised matrix decomposition to efficiently solve the sparse Rayleigh ratio problem.

2 ON SOME TRADE-OFF METRICS

Let us first motivate our approach by considering a number of metrics that have been used to measure performance or allocate portfolio weights. Throughout this section, we consider n different securities, whose characteristics over a given time period (say one month) are given by a mean of expected returns $\boldsymbol{\mu}$, a covariance matrix Σ and $\boldsymbol{\sigma}$ a vector of individual standard deviations (which can be obtained from the covariance matrix). A general comparison of the empirical performance derived from different metrics can be found in [13].

2.1 Sharpe ratio

The *Sharpe ratio* [8] is defined as the trade-off between the portfolio's excess return and its volatility and can be expressed as

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T(\boldsymbol{\mu} - r\mathbf{e})}{\sqrt{\mathbf{w}^T\Sigma\mathbf{w}}}. \tag{1}$$

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524634

If the problem is feasible (i.e., there exists $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w}^T(\boldsymbol{\mu} - r\mathbf{e}) > 0$), then, as previously, Eq. 1 can be expressed as

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \boldsymbol{\mu}\boldsymbol{\mu}^T \mathbf{w}}{\mathbf{w}^T \Sigma \mathbf{w}}, \tag{2}$$

where we have set $r = 0$ for convenience.

2.2 Maximum diversification

The maximum diversification portfolio ("MDP") [2] is defined as the set of weights $\mathbf{w} = [w_1, ..., w_n]^T \in \mathbb{R}^n$ which maximises the ratio of sum of individual volatilities to the portfolio volatility, i.e.,

$$\max_{w_1,...,w_n} \frac{\sum_{i=1}^n w_i \sigma_i}{\sqrt{\sum_{i,j} w_i w_j \rho_{i,j} \sigma_i \sigma_j}} \tag{3}$$

We suppose that the problem is feasible, i.e., there exists $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w}^T \boldsymbol{\sigma} > 0$.

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \boldsymbol{\sigma}\boldsymbol{\sigma}^T \mathbf{w}}{\mathbf{w}^T \Sigma \mathbf{w}}. \tag{4}$$

2.3 Minimum concentration

The minimum concentration portfolio, on the other hand, is given by

$$\min_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \mathrm{diag}(\boldsymbol{\sigma})^2 \mathbf{w}}{\mathbf{w}^T \boldsymbol{\sigma}\boldsymbol{\sigma}^T \mathbf{w}}, \tag{5}$$

which can be rewritten as

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \boldsymbol{\sigma}\boldsymbol{\sigma}^T \mathbf{w}}{\mathbf{w}^T \mathrm{diag}(\boldsymbol{\sigma})^2 \mathbf{w}}. \tag{6}$$

It fundamentally considers the ratio of two extremes: the fully correlated case in the numerator and the fully uncorrelated (or diversified) case in the denominator.

2.4 A general formula

Let us point out that the metrics considered so far can finally be rewritten in general form as

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \left[\delta \boldsymbol{\mu}\boldsymbol{\mu}^T + (1-\delta)\boldsymbol{\sigma}\boldsymbol{\sigma}^T\right] \mathbf{w}}{\mathbf{w}^T \left[\gamma \Sigma + (1-\gamma)\mathrm{diag}(\boldsymbol{\sigma})^2\right] \mathbf{w}}, \tag{7}$$

where $\delta, \gamma \in [0, 1]$. In particular, one notes that $\gamma\Sigma + (1-\gamma)\mathrm{diag}(\boldsymbol{\sigma})^T$ can be understood as a shrinkage (or regularised) estimator [7] of the covariance matrix Σ. Parameters δ and γ can be fine-tuned in practice and taken to be (strictly) in $(0, 1)$.

3 RAYLEIGH PORTFOLIOS

3.1 Rayleigh Quotients

In this section, we show that some well-known portfolio allocation problems can be recast in the common form of a ratio of two quadratic forms

$$\max_{\mathbf{w}\in\mathbb{R}^n} \frac{\mathbf{w}^T \mathbf{A} \mathbf{w}}{\mathbf{w}^T \mathbf{B} \mathbf{w}} := R_{\mathbf{A},\mathbf{B}}(\mathbf{w}), \tag{8}$$

where $\mathbf{A}$ is positive semi-definite matrix and $\mathbf{B}$ is positive definite. This particular optimisation problem is actually well-known and corresponds to generalised Rayleigh ratios [12].

Ratios of two quantities have found their use in finance as they are a natural way of expressing a trade-off between risks and rewards.

- *No leverage effect*: Rayleigh ratios are 0-homogeneous in the sense that for any $\alpha \in \mathbb{R}^*$, $R_{\mathbf{A},\mathbf{B}}(\alpha\mathbf{w}) = R_{\mathbf{A},\mathbf{B}}(\mathbf{w})$. This implies that Rayleigh ratios are "leverage-free" and thus enable us to compare portfolios with different leverage effects.
- *Direct trade-off between two metrics*: There is no need to specify additional hyper-parameters, as is the case in the Markowitz set-up where –in one formulation– one tries to maximise the expected return within a certain amount of acceptable variance that needs to be specified.

3.2 Basic Properties

It is also key to notice that all well-defined Rayleigh ratios (i.e., where $\mathbf{B}$ is positive definite) can be "normalised" in the sense that they can be reduced to an equivalent Rayleigh ratio

$$\max_{\mathbf{w}} R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) = \max_{\mathbf{u}} R_{\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2},\mathbf{I}}(\mathbf{u}), \tag{9}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. Thus, maximising a Rayleigh ratio is akin to performing principal component analysis of the matrix $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$.

Finally, since $R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) = R_{\mathbf{A},\mathbf{B}}(-\mathbf{w})$, maximising a Rayleigh ratio gives two symmetric solutions (i.e., a solution up to a sign flip). One must then proceed to choose a particular solution, usually by choosing the direction that maximises a given linear combination (e.g., the portfolio return in the case of the Sharpe ratio or the sum of individual standard deviations in the case of maximum diversification). In other words,

$$\mathbf{w}^* = \mathrm{sign}(\widehat{\mathbf{w}}^T \mathbf{m}) \cdot \widehat{\mathbf{w}}, \tag{10}$$

with $\mathbf{m}$ a "goal" vector (such as $\boldsymbol{\mu}$ or $\boldsymbol{\sigma}$).

3.3 Introducing Sparsity

There are a number of different reasons for introducing sparsity upper bounds (by which we generally mean setting a large number of weights to 0) [5], both financial and technical.

(1) The L^1-norm represents the absolute exposure of a given portfolio. This sets an upper bound.
(2) Sparsity offers a way of selecting a subset of stocks rather than having to hold the whole universe considered in the optimisation problem, thus reducing the number of instruments that need to be tracked.
(3) We can adopt a statistical viewpoint and understand our formulation as a bet on sparsity; this is akin to Occam's razor.
(4) More robust out-of-sample performance due to regularisation.

At a more fundamental level, sparse Rayleigh ratios offer a way of interpolating between individual security selection and full diversification.

4 FINDING SPARSE PORTFOLIOS

Solving generalised Rayleigh ratio problems with sparsity is a challenge that has been partially addressed in the literature, in particular by leveraging the numerous algorithms developed to tackle sparse

PCA [5, 19]. In this section, we develop an algorithm that is close to sparse PCA.

4.1 A Biconvex Problem

We start by stating a slight generalisation of Theorem 2.1 in [19] (the proof –being straightforward– is omitted for the sake of clarity)

Proposition 4.1. *The problem in Eq. 8 is equivalent to choosing* $\mathbf{w}$ *which is a solution to*

$$\min_{\mathbf{w},\mathbf{v}} -\mathbf{w}^T \mathbf{A}^{1/2}\mathbf{v} \ s.t.\ \|\mathbf{w}\|_{\mathbf{B},2} \leq 1, \|\mathbf{v}\|_2 \leq 1. \tag{11}$$

Note that this problem is biconvex (i.e., convex in $\mathbf{w}$ when $\mathbf{v}$ is fixed and vice-versa), which leads to fast algorithms. Furthermore, we can now add sparsity by introducing an additional constraint on the weight vector $\mathbf{w}$ and its L^1 norm as follows

$$\min_{\mathbf{w},\mathbf{v}} -\mathbf{w}^T \mathbf{A}^{1/2}\mathbf{v} \text{ s.t. } \|\mathbf{w}\|_{\mathbf{B},2} \leq 1, \|\mathbf{v}\|_2 \leq 1, \|\mathbf{w}\|_1 \leq c. \tag{12}$$

It is still a biconvex problem and we can thus resort to alternate convex search ("ACS") or alternating minimisation [1, 14] to solve it. More involved algorithms, generally based on linear quadratic programming, are possible in simple cases [3].

4.2 Modified Penalised Matrix Decomposition

To tackle Eq. 12, we thus propose a modified version of the penalised matrix decomposition algorithm proposed in [19].

Algorithm 1 Modified PMD

Require: Matrix $\mathbf{A}$, sparsity parameter c, goal vector $\mathbf{m}$

Select initial values $\mathbf{u}^0$ and $\mathbf{v}^0$

repeat

$\mathbf{v}^{i+1} \leftarrow \frac{\mathbf{A}^{1/2}\mathbf{w}^i}{\sqrt{(\mathbf{w}^i)^T A \mathbf{w}^i}}$

$\mathbf{d}^{i+1} \leftarrow \mathbf{A}^{1/2}\mathbf{v}^{i+1}$

Find $\widehat{\mathbf{w}} \in \arg\min_{\mathbf{w} \text{ s.t. } \|\mathbf{w}\|^2_{\mathbf{B},2} \leq 1, \|\mathbf{w}\|_1 \leq c} -\mathbf{w}^T \mathbf{d}^{i+1}$

$\mathbf{w}^{i+1} \leftarrow \widehat{w}$

until convergence

$\mathbf{w}^* \leftarrow \text{sign}((\mathbf{w}^i)^T \boldsymbol{m})$

return $\mathbf{w}^*$

While each iteration in $\mathbf{v}$ is available in closed-form, this is not the case for $\mathbf{w}$. However, this inner minimisation problem is a simple convex linear quadratic problem, which is thus solved efficiently at each step with a linear quadratric programming solver.

5 NUMERICAL EXPERIMENTS

5.1 Data

To apply our techniques we looked at the out-of-sample performance of minimum concentration and maximum diversification portfolios using daily returns on the component stocks of the FTSE-100 [17], CAC-40 [15] and DAX-30 [16] European equity indices. The stock prices were obtained from Yahoo Finance [18] which provides an API to access the data.

In order to maximize the time covered but minimize any discontinuity due to changing index composition, de-listings, etc., a handful of stocks were removed and the time coverage altered per index. This resulted in coverage for 38 stocks in the CAC, 91 stocks in the FTSE and 28 stocks in the DAX from Feb 2007 to Feb 2019.

This data was then split in two: a period before October 2016 (the in-sample period) and a period starting October 2016 (the out-of-sample performance).

5.2 Results

We obtain sparse vector weights using the period from February 2007 to September 2016 and apply the same weights to the period from October 2016 to February 2019. This is done in order to determine the stability of the optimal measures over prolonged out of sample periods. Our comparison yields encouraging results and shows strong stability for the minimum concentration measure (figure 1) compared to the maximum diversification measure (figure 2). CAC and FTSE display consistency throughout but maximum diversification for DAX is not stable. This instability is also apparent in the in-sample fit for DAX, which does not show a consistent pattern of increasing maximum diversification measure as the L^1 constraint is relaxed. The complications with DAX can be caused by the lower number of components which combine with sparsity requirements and idiosyncratic stock behaviour to generate a more difficult fit and stability. A fact hinted at by the slightlty lower stability of the relationship in CAC versus FTSE (figures 1 and 2).

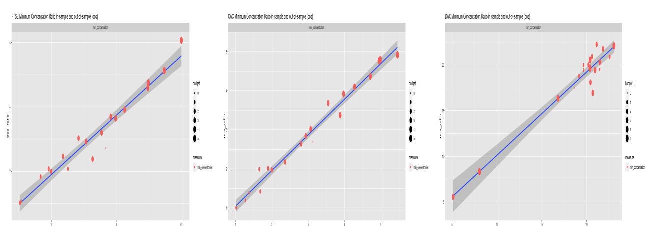

Figure 1: (a) the out of sample minimum concentration ratio for FTSE (obtained using the sparse weights from the in sample fit) regressed on the in sample minimum concentration ratio, each point represents a different level of constraint. (b) the same as (a) but for CAC. (c) the same as (a) but for DAX.

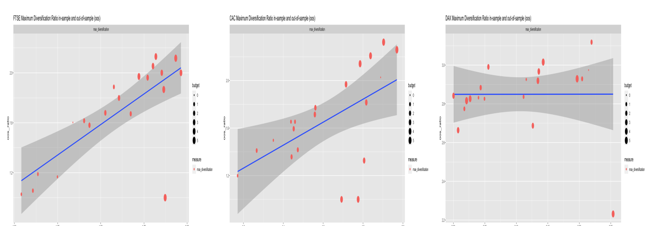

Figure 2: (a) the out of sample maximum diversification ratio for FTSE (obtained using the sparse weights from the in sample fit) regressed on the in sample maximum diversification ratio, each point represents a different level of constraint. (b) the same as (a) but for CAC. (c) the same as (a) but for DAX.

6 EXTENSIONS

While the framework we have described is rather flexible, several extensions are possible, ranging from additional metrics such as the information ratio ("IR") [4]. However, for the sake of brevity, we focus on two ideas related to sparsity.

6.1 Penalising turnover

Let us first consider a slightly different setup in which we *inherit* an existing portfolio allocation $\mathbf{w}_t$ (from, say, period t). To build a new allocation, we still wish to maximise a given Rayleigh ratio $\max_{\mathbf{w}} R_{\mathbf{A},\mathbf{B}}(\mathbf{w})$, but also wish to minimise turnover [10], i.e., the cost of exiting positions and entering new ones. There are multiple ways of modelling turnover, but a simple extension to our previous penalised matrix decomposition approach is to modify the penalty term $\mathcal{P}(\mathbf{w})$ on the vector $\mathbf{w}$ as

$$\mathcal{P}(\mathbf{w}) := \lambda\|\mathbf{w}\|_1 + \mu\|\mathbf{w} - \mathbf{w}_t\|_1, \tag{13}$$

where $\lambda, \mu > 0$ are tunable hyperparameters.

6.2 Region or sector allocation

If we suppose that the weights $\mathbf{w}$ can be decomposed into m groups (for instance sectors or regions) so that the overall allocation vector $\mathbf{w}$ is concatenation of each group's allocation vector $\mathbf{w}_m$:

$$\mathbf{w} = \left[\mathbf{w}^{(1)}, \cdots, \mathbf{w}^{(m)}\right]^T. \tag{14}$$

One of an investor's wishes may be to decide whether or not to get exposure to a sector, as opposed to thinking in terms of exposure to a given security. This can be translated mathematically in terms of a "group LASSO" [5] type of penalty

$$\mathcal{P}(\mathbf{w}) := \lambda\|\mathbf{w}\|_1 + \mu \sum_{l=1}^{m} \sqrt{p_l} \left\|\mathbf{w}^{(l)}\right\|_2, \tag{15}$$

whose objective is to induce sparsity at the group level.

7 CONCLUSION

In this paper, we have established a correspondence between a number of portfolio allocation problems involving –in general– a ratio of risks and rewards (such as Sharpe ratios, maximum diversification or minimum concentration) and Rayleigh ratios. This link is important as it enabled us to connect such optimisation objectives and matrix decomposition.

Matrix decomposition techniques, including principal component analysis or singular value decomposition, have long played an important role in quantitative finance, especially for factor analysis of the cross section of asset returns or the term structure of interest rates.

Our main interest, however, resides in insights from matrix decomposition methods and their generalisations, amongst which, first and foremost, sparsity-inducing techniques. These correspond to a *statistical* imperative, namely robustness and stability, and to a *financial* constraint, in the sense of transaction cost minimisation or exposure management. To induce sparsity, we have revisited three widely used algorithms, sparse canonical correlation analysis, penalised matrix decomposition and sparse principal component analysis, and applied their modified versions to the challenges at hand. (These variations can be more widely applied to mathematically related problems, such as linear discriminant analysis and canonical correlation analysis.)

Empirically, we apply our techniques to three European indices, CAC, FTSE and DAX, between 2016 and 2020. One of the challenges of Cross-validation can be adapted to time series and investment decisions' back-testing

Finally, interesting research avenues are left to be fully explored in this framework, such as the inclusion of non-negative weight constraints ("long-only portfolios"), structured penalties (e.g., group-sparse regularisation to turn on or off certain sectors or countries in the allocation) and robust matrix decomposition techniques (for instance by considering generalised low rank models).

REFERENCES

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex optimization.* Cambridge university press.
[2] Yves Choueifaty and Yves Coignard. 2008. Toward maximum diversification. *The Journal of Portfolio Management* 35, 1 (2008), 40–51.
[3] Gérard Cornuéjols, Javier Peña, and Reha Tütüncü. 2018. *Optimization Methods in Finance* (2 ed.). Cambridge University Press. https://doi.org/10.1017/9781107297340
[4] Thomas H Goodwin. 1998. The information ratio. *Financial Analysts Journal* 54, 4 (1998), 34–43.
[5] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Chapman amp; Hall/CRC.
[6] Philippe Jorion. 1992. Portfolio optimization in practice. *Financial analysts journal* 48, 1 (1992), 68–74.
[7] Olivier Ledoit and Michael Wolf. 2004. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management* 30, 4 (2004), 110–119.
[8] Andrew W Lo. 2002. The statistics of Sharpe ratios. *Financial analysts journal* 58, 4 (2002), 36–52.
[9] Harry M Markowitz. 1968. *Portfolio selection.* Yale university press.
[10] Edward E Qian. 2007. *Quantitative equity portfolio management: modern techniques and applications.* Chapman and Hall/CRC.
[11] Thierry Roncalli. 2013. *Introduction to risk parity and budgeting.* CRC Press.
[12] Gilbert Strang. 2019. *Linear algebra and learning from data.* Wellesley-Cambridge Press Cambridge.
[13] Ludan Theron and Gary van Vuuren. 2018. The maximum diversification investment strategy: A portfolio performance comparison. *Cogent Economics & Finance* 6, 1 (2018), 1427533. https://doi.org/10.1080/23322039.2018.1427533 arXiv:https://doi.org/10.1080/23322039.2018.1427533
[14] Madeleine Udell. 2015. *Generalized low rank models.* Stanford University.
[15] Wikipedia. 2022. CAC 40 — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=CAC%2040&oldid=1064038984. [Online; accessed 04-February-2022].
[16] Wikipedia. 2022. DAX — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=DAX&oldid=1069646364. [Online; accessed 04-February-2022].
[17] Wikipedia. 2022. FTSE 100 Index — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=FTSE%20100%20Index&oldid=1068946404. [Online; accessed 04-February-2022].
[18] Wikipedia. 2022. Yahoo! Finance — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Yahoo!%20Finance&oldid=1068965272. [Online; accessed 04-February-2022].
[19] Daniela M. Witten, Robert Tibshirani, and Trevor Hastie. 2009. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10, 3 (04 2009), 515–534.