

Risk-Parity Portfolio

Prof. Daniel P. Palomar

The Hong Kong University of Science and Technology (HKUST)

MAFS6010R- Portfolio Optimization with R

MSc in Financial Mathematics

Fall 2018-19, HKUST, Hong Kong

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Motivation

- The Markowitz portfolio has never been embraced by practitioners, among other reasons because
 - ① variance is not a good measure of risk in practice since it penalizes both the unwanted high losses and the desired low losses: the solution is to use **alternative measures for risk, e.g., VaR and CVaR**,
 - ② it is highly sensitive to parameter estimation errors (i.e., to the covariance matrix Σ and especially to the mean vector μ): solution is **robust optimization**,
 - ③ it only considers the risk of the portfolio as a whole and ignores the risk diversification (i.e., concentrates risk too much in few assets, this was observed in the 2008 financial crisis): solution is the **risk-parity portfolio**.

👉 *We will here address the risk diversification among the assets by totally changing the portfolio formulation (and somehow address the sensitivity w.r.t. parameters).*

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Returns

- Let us denote the log-returns of N assets at time t with the vector $\mathbf{r}_t \in \mathbb{R}^N$.
- The time index t can denote any arbitrary period such as days, weeks, months, 5-min intervals, etc.
- \mathcal{F}_{t-1} denotes the previous historical data.
- Econometrics aims at modeling \mathbf{r}_t conditional on \mathcal{F}_{t-1} .
- \mathbf{r}_t is a multivariate stochastic process with conditional mean and covariance matrix denoted as¹

$$\boldsymbol{\mu}_t \triangleq \mathbb{E}[\mathbf{r}_t \mid \mathcal{F}_{t-1}]$$

$$\boldsymbol{\Sigma}_t \triangleq \text{Cov}[\mathbf{r}_t \mid \mathcal{F}_{t-1}] = \mathbb{E}\left[(\mathbf{r}_t - \boldsymbol{\mu}_t)(\mathbf{r}_t - \boldsymbol{\mu}_t)^T \mid \mathcal{F}_{t-1}\right].$$

¹Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends in Signal Processing, Now Publishers, 2016.

I.I.D. Model

- For simplicity we will assume that \mathbf{r}_t follows an i.i.d. distribution (which is not very inaccurate in general).
- That is, both the conditional mean and conditional covariance are constant

$$\begin{aligned}\mu_t &= \mu, \\ \Sigma_t &= \Sigma.\end{aligned}$$

- Very simple model, however, it is one of the most fundamental assumptions for many important works, e.g., the Nobel prize-winning Markowitz portfolio theory².

²H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Parameter Estimation

- Consider the i.i.d. model:

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{w}_t,$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the mean and $\mathbf{w}_t \in \mathbb{R}^N$ is an i.i.d. process with zero mean and constant covariance matrix $\boldsymbol{\Sigma}$.

- The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ have to be estimated in practice based on T observations.
- The simplest estimator is the sample estimator:
 - sample mean estimator: $\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t$
 - sample covariance matrix: $\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1} \sum_{t=1}^T (\mathbf{r}_t - \hat{\boldsymbol{\mu}})(\mathbf{r}_t - \hat{\boldsymbol{\mu}})^T$.
- Many more sophisticated estimators exist, namely: shrinkage estimators, Black-Litterman estimators, etc.

Parameter Estimation

- The parameter estimates $\hat{\mu}$ and $\hat{\Sigma}$ are only good for large T , otherwise the estimation error is unacceptable.
- For instance, the sample mean is particularly a very inefficient estimator, with very noisy estimates.³
- In practice, T cannot be large enough due to either:
 - unavailability of data or
 - lack of stationarity of data.
- As a consequence, the estimates contain too much estimation error and a portfolio design (e.g., Markowitz mean-variance) based on those estimates can be fatal.
- Indeed, this is why Markowitz portfolio and other extensions are rarely used by practitioners.

³A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Portfolio Return

- Suppose the budget is B dollars.
- The portfolio $\mathbf{w} \in \mathbb{R}^N$ denotes the normalized weights of the assets such that $\mathbf{1}^T \mathbf{w} = 1$ (then $B\mathbf{w}$ denotes dollars invested in the assets).
- For each asset, the initial wealth is Bw_i and the end wealth is

$$Bw_i(p_{i,t}/p_{i,t-1}) = Bw_i(R_{it} + 1).$$

- Then the portfolio return is

$$R_t^p = \frac{\sum_{i=1}^N Bw_i(R_{it} + 1) - B}{B} = \sum_{i=1}^N w_i R_{it} \approx \sum_{i=1}^N w_i r_{it} = \mathbf{w}^T \mathbf{r}_t$$

- The portfolio expected return and variance are $\mathbf{w}^T \boldsymbol{\mu}$ and $\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$, respectively.⁴

⁴G. Cornuejols and R. Tütüncü, *Optimization Methods in Finance*. Cambridge University Press, 2006.

Performance Measures

- Expected return: $\mathbf{w}^T \boldsymbol{\mu}$
- Volatility: $\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}$
- Sharpe Ratio (SR): expected return per unit of risk

$$\text{SR} = \frac{\mathbf{w}^T \boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}$$

where r_f is the risk-free rate (e.g., interest rate on a three-month U.S. Treasury bill).

- Information Ratio (IR):

$$\text{IR} = \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}$$

- Drawdown: decline from a historical peak of the cumulative profit $X(t)$: $D(T) = \max \left\{ 0, \max_{t \in (0, T)} X(t) - X(T) \right\}$
- VaR (Value at Risk)
- ES (Expected Shortfall) or CVaR (Conditional Value at Risk)

Practical Constraints

- Capital budget constraint:

$$\mathbf{w}^T \mathbf{1} = 1.$$

- Long-only constraint:

$$\mathbf{w} \geq 0.$$

- Market-neutral constraint:

$$\mathbf{w}^T \mathbf{1} = 0.$$

- Turnover constraint:

$$\|\mathbf{w} - \mathbf{w}_0\|_1 \leq u$$

where \mathbf{w}_0 is the currently held portfolio.

Practical Constraints

- Holding constraint:

$$\mathbf{l} \leq \mathbf{w} \leq \mathbf{u}$$

where $\mathbf{l} \in \mathbb{R}^N$ and $\mathbf{u} \in \mathbb{R}^N$ the lower and upper bounds of turnover of each asset, respectively.

- Cardinality constraint:

$$\|\mathbf{w}\|_0 \leq K.$$

- Leverage constraint:

$$\|\mathbf{w}\|_1 \leq 2.$$

- In finance, the expected return $\mathbf{w}^T \boldsymbol{\mu}$ is very relevant as it quantifies the average benefit.
- However, in practice, the average performance is not enough to characterize an investment and one needs to control the probability of going bankrupt.
- Risk measures control how risky an investment strategy is.
- The most basic measure of risk is given by the variance⁵: a higher variance means that there are large peaks in the distribution which may cause a big loss.
- There are more sophisticated risk measures such as downside risk, VaR, ES, etc.

⁵H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Mean-Variance Tradeoff

- The mean return $\mathbf{w}^T \boldsymbol{\mu}$ and the variance (risk) $\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$ constitute two important performance measures.
- Usually, the higher the mean return the higher the variance and vice-versa.
- Thus, we are faced with two objectives to be optimized: it is a multi-objective optimization problem.
- They define a fundamental mean-variance tradeoff curve (Pareto curve).
- The choice of a specific point in this tradeoff curve depends on how aggressive or risk-averse the investor is.

Markowitz mean-variance portfolio (1952)

- The idea of the Markowitz framework⁶ is to find a trade-off between the expected return $\mathbf{w}^T \boldsymbol{\mu}$ and the risk of the portfolio measured by the variance $\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T \boldsymbol{\mu} - \lambda \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1 \end{array}$$

where $\mathbf{w}^T \mathbf{1} = 1$ is the capital budget constraint and λ is a parameter that controls how risk-averse the investor is.

- This is a convex QP with only one linear constraint which admits a closed-form solution:

$$\mathbf{w}^* = \frac{1}{2\lambda} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} + \nu^* \mathbf{1}),$$

where ν^* is the optimal dual variable $\nu^* = \frac{2\lambda - \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}}$.

⁶H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Global Minimum Variance Portfolio (GMVP)

- The global minimum variance portfolio (GMVP) ignores the expected return and focuses on the risk only:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1.\end{array}$$

- It is a simple convex QP with solution

$$\mathbf{w}_{\text{GMVP}} = \frac{1}{\mathbf{1}^T \mathbf{\Sigma}^{-1} \mathbf{1}} \mathbf{\Sigma}^{-1} \mathbf{1}.$$

- It is widely used in academic papers for simplicity of evaluation and comparison of different estimators of the covariance matrix $\mathbf{\Sigma}$ (while ignoring the estimation of $\boldsymbol{\mu}$).

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Drawbacks of Markowitz's formulation

- The Markowitz portfolio has never been embraced by practitioners, among other reasons because
 - ① variance is not a good measure of risk in practice since it penalizes both the unwanted high losses and the desired low losses: the solution is to use **alternative measures for risk, e.g., VaR and CVaR**,
 - ② it is highly sensitive to parameter estimation errors (i.e., to the covariance matrix Σ and especially to the mean vector μ): solution is **robust optimization**,
 - ③ it only considers the risk of the portfolio as a whole and ignores the risk diversification (i.e., concentrates risk too much in few assets, this was observed in the 2008 financial crisis): solution is the **risk-parity portfolio**.

👉 *We will here address the risk diversification among the assets by totally changing the portfolio formulation (and somehow address the sensitivity w.r.t. parameters).*

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Motivation

- The Markowitz portfolio has never been embraced by practitioners, among other reasons because
 - it only considers the risk of the portfolio as a whole and ignores the risk diversification (i.e., concentrates risk too much in few assets, this was observed in the 2008 financial crisis)
 - it is highly sensitive to the estimation errors in the parameters (i.e., small estimation errors in the parameters may change completely the designed portfolio)
- 🖐 Recently, the alternative **risk parity portfolio design** has been receiving significant attention from both the theoretical and practical sides because
 - diversifies the risk, instead of the capital, among the assets
 - less sensitive to parameter estimation errors.

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Risk Contribution

- Given a portfolio $\mathbf{w} \in \mathbb{R}^N$ and the return covariance matrix $\mathbf{\Sigma}$, the portfolio volatility is:

$$\sigma(\mathbf{w}) = \sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}.$$

- Following Euler's theorem, the volatility can be decomposed as follows:

$$\sigma(\mathbf{w}) = \sum_{i=1}^N w_i \frac{\partial \sigma}{\partial w_i} = \sum_{i=1}^N \frac{w_i (\mathbf{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}}$$

- Risk contribution from asset i to the total risk $\sigma(\mathbf{w})$:

$$\frac{\partial \sigma}{\partial w_i} = \frac{w_i (\mathbf{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}}$$

- Normalized risk contribution: $w_i (\mathbf{\Sigma} \mathbf{w})_i / \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}$
- Other measures of risk like VAR and CVaR can also be decomposed following Euler's theorem.

Risk-Parity Portfolio

- Idea: equalize the risks contributions:

$$w_i \frac{\partial f(\mathbf{w})}{\partial w_i} = w_j \frac{\partial f(\mathbf{w})}{\partial w_j} \quad \forall i, j.$$

- Risk budgeting is a more general concept. Given a risk budget vector $\mathbf{b} = [b_1, \dots, b_N]^T > \mathbf{0}$, $\mathbf{1}^T \mathbf{b} = 1$, the risk budgeting portfolio should satisfy

$$w_i \frac{\partial f(\mathbf{w})}{\partial w_i} = b_i f(\mathbf{w}) \quad \forall i.$$

- Risk parity portfolio is a special case of the risk budgeting portfolio with $\mathbf{b} = \mathbf{1}/N$.

Risk-Parity Portfolio

- For the volatility we can write
 - risk parity: $w_i (\boldsymbol{\Sigma} \mathbf{w})_i = w_j (\boldsymbol{\Sigma} \mathbf{w})_j$
 - risk budgeting: $w_i (\boldsymbol{\Sigma} \mathbf{w})_i = b_i \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$
- Assuming that $\boldsymbol{\Sigma}$ is diagonal and with the constraints $\mathbf{1}^T \mathbf{w} = 1$ and $\mathbf{w} \geq \mathbf{0}$, the risk budgeting portfolio is

$$w_i = \frac{\sqrt{b_i} / \sqrt{\Sigma_{ii}}}{\sum_{k=1}^N \sqrt{b_k} / \sqrt{\Sigma_{kk}}}, \quad i = 1, \dots, N.$$

- However, for non-diagonal $\boldsymbol{\Sigma}$ or with other additional constraints, a closed-form solution does not exist in general and some optimization procedures have to be constructed.

Risk-Parity Formulation

- Maillard et al.⁷ aimed at solving:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \sum_{i,j=1}^N \left(w_i (\boldsymbol{\Sigma} \mathbf{w})_i - w_j (\boldsymbol{\Sigma} \mathbf{w})_j \right)^2 \\ & \text{subject to} && \mathbf{1}^T \mathbf{w} = 1. \end{aligned}$$

- The idea is to try to achieve equal risk contributions $\frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}$ by penalizing the differences between the terms $w_i (\boldsymbol{\Sigma} \mathbf{w})_i$.
- This is a simplified formulation:

$$\begin{aligned} & \underset{\mathbf{w}, \theta}{\text{minimize}} && \sum_{i=1}^N (w_i (\boldsymbol{\Sigma} \mathbf{w})_i - \theta)^2 \\ & \text{subject to} && \mathbf{1}^T \mathbf{w} = 1. \end{aligned}$$

⁷S. Maillard, T. Roncalli, and J. Teïletche, "The properties of equally weighted risk contribution portfolios," *Journal of Portfolio Management*, vol. 36, no. 4, pp. 60–70, 2010.

More Risk-Parity Formulations

- Bruder and Roncalli⁸ proposed to solve:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \sum_{i=1}^N \left(\frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}} - b_i \right)^2 \\ & \text{subject to} && \mathbf{w}^T\mathbf{1} = 1, \end{aligned}$$

where $b_i = \frac{1}{N}$.

- More generally, one can equalize the risk contribution by setting arbitrary proportions (as opposed to equal contributions):

$$\mathbf{b} = [b_1, \dots, b_N]^T > \mathbf{0}, \quad \mathbf{1}^T\mathbf{b} = 1.$$

- One more formulation:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \sum_{i,j=1}^N \left(\frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{b_i} - \frac{w_j(\boldsymbol{\Sigma}\mathbf{w})_j}{b_j} \right)^2 \\ & \text{subject to} && \mathbf{1}^T\mathbf{w} = 1. \end{aligned}$$

⁸B. Bruder and T. Roncalli, "Managing risk exposures using the risk budgeting approach," *University Library of Munich, Germany, Tech. Rep.*, 2012.

And Yet More Risk-Parity Formulations

- One more formulation:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \sum_{i=1}^N \left(w_i (\boldsymbol{\Sigma} \mathbf{w})_i - b_i \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \right)^2 \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1.\end{array}$$

- And one more:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \sum_{i=1}^N \left(\frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}} - b_i \sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}} \right)^2 \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1.\end{array}$$

Yet Even More Risk-Parity Formulations

- What about this one:

$$\begin{aligned} & \underset{\mathbf{w}, \theta}{\text{minimize}} && \sum_{i=1}^N \left(\frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{b_i} - \theta \right)^2 \\ & \text{subject to} && \mathbf{1}^T \mathbf{w} = 1. \end{aligned}$$

- More formulations can be found in the book:
T. Roncalli, *Introduction to Risk Parity and Budgeting*. CRC Press, 2013.

General Problem Formulation

A more general risk parity formulation is⁹:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && U(\mathbf{w}) \triangleq \sum_{i=1}^N (g_i(\mathbf{w}))^2 + \lambda F(\mathbf{w}) \\ & \text{subject to} && \mathbf{1}^T \mathbf{w} = 1, \quad \mathbf{w} \in \mathcal{W} \end{aligned}$$

where

- $\sum_{i=1}^N (g_i(\mathbf{w}))^2$: risk concentration measurement, e.g.,
$$g_i(\mathbf{w}) \triangleq \frac{w_i (\boldsymbol{\Sigma} \mathbf{w})_i}{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}} - \frac{1}{N},$$
- $F(\mathbf{w})$: preference, e.g., 0 , $-\mu^T \mathbf{w}$, $-\mu^T \mathbf{w} + \nu \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$,
- $\lambda \geq 0$: trade-off parameter,
- $\mathbf{1}^T \mathbf{w} = 1, \mathbf{w} \in \mathcal{W}$: capital budget & other convex constraints.

Challenge: the problem is highly nonconvex due to $\sum_{i=1}^N (g_i(\mathbf{w}))^2$.

⁹Y. Feng and D. P. Palomar, "SCRIP: Successive convex optimization methods for risk parity portfolios design," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5285–5300, 2015.

Unified Problem Formulation

- The previous general formulation contains the risk term $R(\mathbf{w}) = \sum_{i=1}^N (g_i(\mathbf{w}))^2$, which can be written in a compact way to represent the many formulations presented before.
- Define $\mathbf{M}_i \in \mathbb{R}^{N \times N}$ as a sparse matrix with its i -th row equal to that of the covariance matrix Σ .
- Examples:

- $R(\mathbf{w}) = \sum_{i,j=1}^N \left(w_i(\Sigma \mathbf{w})_i - w_j(\Sigma \mathbf{w})_j \right)^2$ corresponds to

$$g_{i,j}(\mathbf{w}) = \mathbf{w}^T (\mathbf{M}_i - \mathbf{M}_j) \mathbf{w}$$

- $R(\mathbf{w}) = \sum_{i=1}^N (w_i(\Sigma \mathbf{w})_i - \theta)^2$ corresponds to

$$g_i(\mathbf{w}) = \mathbf{w}^T \mathbf{M}_i \mathbf{w} - \theta$$

- $R(\mathbf{w}) = \sum_{i=1}^N \left(\frac{w_i(\Sigma \mathbf{w})_i}{\mathbf{w}^T \Sigma \mathbf{w}} - b_i \right)^2$ corresponds to

$$g_i(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \Sigma \mathbf{w}} - b_i$$

Unified Problem Formulation

- More examples:

- $R(\mathbf{w}) = \sum_{i,j=1}^N \left(\frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{b_i} - \frac{w_j(\boldsymbol{\Sigma}\mathbf{w})_j}{b_j} \right)^2$ corresponds to

$$g_{i,j}(\mathbf{w}) = \mathbf{w}^T(\mathbf{M}_i/b_i - \mathbf{M}_j/b_j)\mathbf{w}$$

- $R(\mathbf{w}) = \sum_{i=1}^N (w_i(\boldsymbol{\Sigma}\mathbf{w})_i - b_i\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w})^2$ corresponds to

$$g_i(\mathbf{w}) = \mathbf{w}^T(\mathbf{M}_i - b_i\boldsymbol{\Sigma})\mathbf{w}$$

- $R(\mathbf{w}) = \sum_{i=1}^N \left(\frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} - b_i\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}} \right)^2$ corresponds to

$$g_i(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{M}_i\mathbf{w}}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} - b_i\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$$

- $R(\mathbf{w}) = \sum_{i=1}^N \left(\frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{b_i} - \theta \right)^2$ corresponds to

$$g_i(\mathbf{w}) = \mathbf{w}^T\mathbf{M}_i\mathbf{w}/b_i - \theta$$

Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

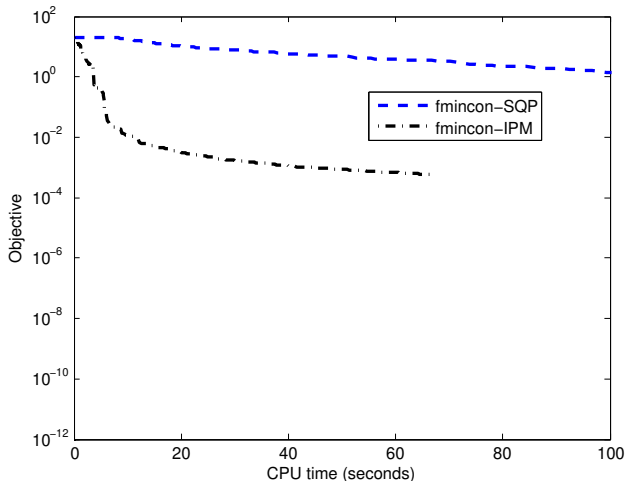
Numerical Solving Approach

- Some off-the-shelf nonlinear numerical optimization methods¹⁰ are typically used, e.g.,
 - Sequential Quadratic Programming (SQP)
 - Interior Point Methods (IPM).
- For such risk-parity portfolio problems, they
 - may be very slow, and
 - get stuck at some unsatisfactory points.
- Because the **structure of the objective is not explored**.

¹⁰J. Nocedal and S. J. Wright., *Numerical Optimization*, Second. Springer Verlag, 2006.

Numerical Example: Slow Convergence

Off-the-shelf nonlinear solvers have slow convergence for the risk-parity portfolio problem:



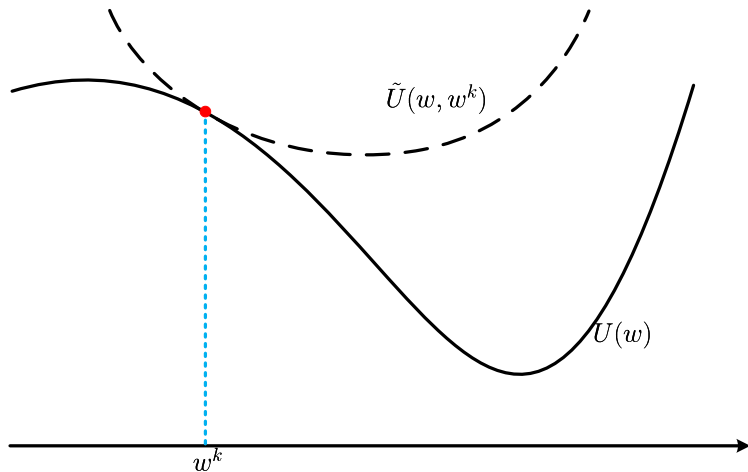
Successive Convex Approximation (SCA)

- **Basic idea:** solving a difficult problem via solving a sequence of simpler problems.
- Minimize $U(\mathbf{w})$ over $\mathbf{w} \in \overline{\mathcal{W}}$ via SCA method¹¹:
 - **Construction of Approximation:** finding $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$ that approximates the function $U(\mathbf{w})$ at the point \mathbf{w}^k and
 - $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$: uniformly strongly convex & cont. differentiable
 - $\nabla \tilde{U}(\mathbf{w}; \mathbf{w}^k)$: Lipschitz continuous on $\overline{\mathcal{W}}$
 - $\nabla \tilde{U}(\mathbf{w}; \mathbf{w}^k)|_{\mathbf{w}=\mathbf{w}^k} = \nabla U(\mathbf{w})|_{\mathbf{w}=\mathbf{w}^k}$
 - **Minimization:** minimizing $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$ to get the update

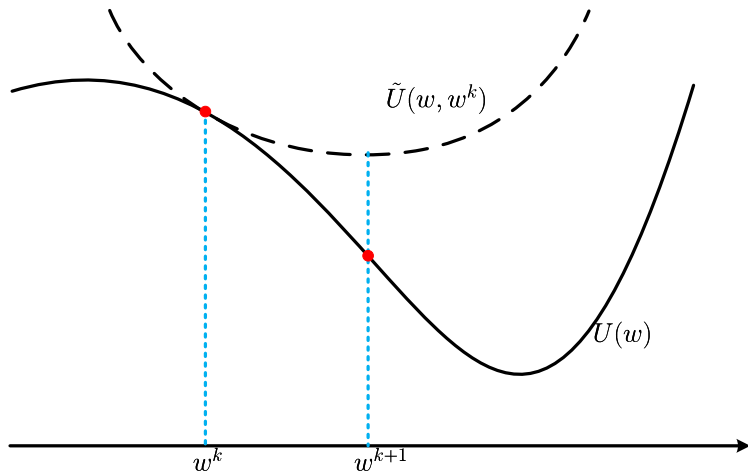
$$\mathbf{w}^{k+1} \triangleq \arg \min_{\mathbf{w} \in \overline{\mathcal{W}}} \tilde{U}(\mathbf{w}; \mathbf{w}^k).$$

¹¹G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, 2014.

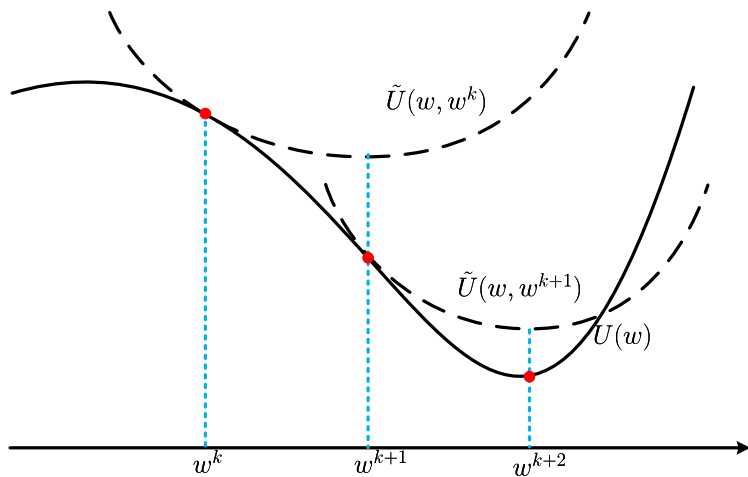
Construction of Approximation



Minimization



One More Iteration



Classical Methods as SCA

- **(Unconstrained) gradient descent:** Set

$$\tilde{U}(\mathbf{w}; \mathbf{w}^k) = U(\mathbf{w}^k) + \nabla U(\mathbf{w}^k)^T (\mathbf{w} - \mathbf{w}^k) + \frac{1}{2\alpha^k} \|\mathbf{w} - \mathbf{w}^k\|_2^2.$$

Setting the derivative w.r.t. \mathbf{w} to zero yields:

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \alpha^k \nabla U(\mathbf{w}^k).$$

- **(Unconstrained) Newton's method:** Set

$$\begin{aligned} \tilde{U}(\mathbf{w}; \mathbf{w}^k) &= U(\mathbf{w}^k) + \nabla U(\mathbf{w}^k)^T (\mathbf{w} - \mathbf{w}^k) \\ &\quad + \frac{1}{2\alpha^k} (\mathbf{w} - \mathbf{w}^k)^T \nabla^2 U(\mathbf{w}^k) (\mathbf{w} - \mathbf{w}^k). \end{aligned}$$

Setting the derivative w.r.t. \mathbf{w} to zero yields:

$$\mathbf{w}^{k+1} = \mathbf{w}^k - \alpha^k \left(\nabla^2 U(\mathbf{w}^k) \right)^{-1} \nabla U(\mathbf{w}^k).$$

SCA for Risk Parity Portfolio Design

- Recall the objective

$$U(\mathbf{w}) = \sum_{i=1}^N (g_i(\mathbf{w}))^2 + \lambda F(\mathbf{w}).$$

- At the k -th iteration \mathbf{w}^k , set $\tau > 0$ and construct

$$\tilde{U}(\mathbf{w}, \mathbf{w}^k) = \overbrace{\sum_{i=1}^N \left(g_i(\mathbf{w}^k) + \left(\nabla g_i(\mathbf{w}^k) \right)^T (\mathbf{w} - \mathbf{w}^k) \right)^2}^{P(\mathbf{w}; \mathbf{w}^k) \triangleq} + \frac{\tau}{2} \left\| \mathbf{w} - \mathbf{w}^k \right\|_2^2 + \lambda F(\mathbf{w})$$

- IDEA: linearizing nonconvex functions $g_i(\mathbf{w})$ inside the least square \implies quadratic convex $P(\mathbf{w}; \mathbf{w}^k)$ approximates

$$R(\mathbf{w}) = \sum_{i=1}^N (g_i(\mathbf{w}))^2, \text{ with } \nabla P(\mathbf{w}, \mathbf{w}^k) |_{\mathbf{w}=\mathbf{w}^k} = \nabla R(\mathbf{w}) |_{\mathbf{w}=\mathbf{w}^k}.$$

Problem Reformulation

- $P(\mathbf{w}; \mathbf{w}^k)$ can be rewritten more compactly as

$$P(\mathbf{w}; \mathbf{w}^k) = \|\mathbf{A}^k (\mathbf{w} - \mathbf{w}^k) + \mathbf{g}(\mathbf{w}^k)\|^2$$

where

$$\begin{aligned}\mathbf{A}^k &\triangleq \left[\nabla g_1(\mathbf{w}^k), \dots, \nabla g_N(\mathbf{w}^k) \right]^T, \\ \mathbf{g}(\mathbf{w}^k) &\triangleq \left[g_1(\mathbf{w}^k), \dots, g_N(\mathbf{w}^k) \right]^T.\end{aligned}$$

- We can further expand $P(\mathbf{w}; \mathbf{w}^k)$ as

$$\begin{aligned}P(\mathbf{w}; \mathbf{w}^k) &= (\mathbf{w} - \mathbf{w}^k)^T (\mathbf{A}^k)^T \mathbf{A}^k (\mathbf{w} - \mathbf{w}^k) + \mathbf{g}(\mathbf{w}^k)^T \mathbf{g}(\mathbf{w}^k) \\ &\quad + 2\mathbf{g}(\mathbf{w}^k)^T \mathbf{A}^k (\mathbf{w} - \mathbf{w}^k)\end{aligned}$$

Problem Reformulation

- The QP approximation problem at the k -th iteration is

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \tilde{U}(\mathbf{w}, \mathbf{w}^k) = \frac{1}{2} \mathbf{w}^T \mathbf{Q}^k \mathbf{w} + \mathbf{w}^T \mathbf{q}^k + \lambda F(\mathbf{w}) \\ & \text{subject to} && \mathbf{1}^T \mathbf{w} = 1, \quad \mathbf{w} \in \mathcal{W}. \end{aligned} \quad (1)$$

where

$$\begin{aligned} \mathbf{Q}^k &\triangleq 2 \left(\mathbf{A}^k \right)^T \mathbf{A}^k + \tau \mathbf{I}, \\ \mathbf{q}^k &\triangleq 2 \left(\mathbf{A}^k \right)^T \mathbf{g}(\mathbf{w}^k) - \mathbf{Q}^k \mathbf{w}^k, \end{aligned}$$

- This problem can be solved directly with a solver or, depending on the constraints in \mathcal{W} , one may derive simpler closed-form solutions.
- For example, if we only have equality constraints in the form $\mathbf{C}\mathbf{w} = \mathbf{c}$, then from the KKT optimality conditions the optimal solution is found as $\hat{\mathbf{w}}^k = -(\mathbf{Q}^k)^{-1}(\mathbf{q}^k + \mathbf{C}^T \boldsymbol{\lambda}^k)$ where $\boldsymbol{\lambda}^k = -\left(\mathbf{C}(\mathbf{Q}^k)^{-1} \mathbf{C}^T\right)^{-1} \left(\mathbf{C}(\mathbf{Q}^k)^{-1} \mathbf{q}^k + \mathbf{c}\right)$.

Sequential Numerical Algorithm

Algorithm 1: Successive Convex optimization for Risk Parity portfolio (SCRIP).

Set $k = 0$, $\mathbf{w}^0 \in \overline{\mathcal{W}}$, $\tau > 0$, $\{\gamma^k\} \in (0, 1]$

repeat

Solve QP problem (1) to get the optimal solution $\hat{\mathbf{w}}^k$ (global minimum)

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \gamma^k (\hat{\mathbf{w}}^k - \mathbf{w}^k)$$

$$k \leftarrow k + 1$$

until convergence

return \mathbf{w}^k

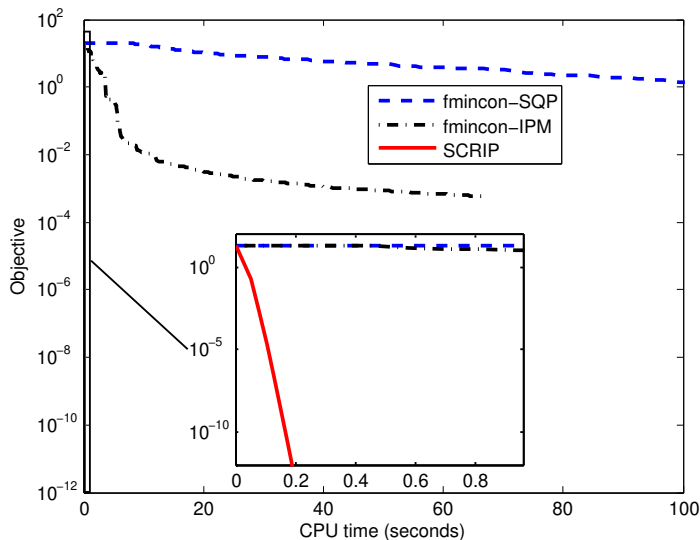
- More advanced algorithms can be found in Y. Feng and D. P. Palomar, "SCRIP: Successive convex optimization methods for risk parity portfolios design," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5285–5300, 2015.

Proposition 1

Under some technical conditions, suppose $\tau > 0$, $\gamma^k \in (0, 1]$, $\gamma^k \rightarrow 0$, $\sum_k \gamma^k = +\infty$ and $\sum_k (\gamma^k)^2 < +\infty$, and let $\{\mathbf{w}^k\}$ be the sequence generated by Algorithm 1. Then, either Algorithm 1 converges in a finite number of iterations to a stationary point or every limit of $\{\mathbf{w}^k\}$ (at least one such point exists) is a stationary point.

Numerical example

Fast algorithms based on successive convex approximation (SCA):



Outline

1 Introduction

2 Warm-Up: Markowitz Portfolio

- Signal model
- Markowitz formulation
- Drawbacks of Markowitz portfolio

3 Risk-Parity Portfolio

- Problem formulation
- Algorithms via SCA

4 Conclusions

Conclusions

- We have reviewed the Markowitz portfolio formulation and understood that it has many practical flaws that make it impractical. Indeed, it is not used by practitioners.
- We have learned about the risk-parity portfolio formulation.
- We have explored the numerical resolution of such problems via successive convex approximation (SCA) methods.
- The performance of risk-parity portfolio versus Markowitz portfolio is much improved.
- Side result: we have learned how to develop efficient numerical algorithms based on SCA.

Thanks

For more information visit:

<https://www.danielpalomar.com>

