Risk Parity Portfolio

Prof. Daniel P. Palomar

ELEC5470/IEDA6100A - Convex Optimization
Hong Kong University of Science and Technology (HKUST)
Fall 2019-20

Outline

Outline

Motivation

Markowitz's portfolio has never been fully embraced by practitioners, among other reasons (Zhao et al. 2019)[1] 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 $\mathbf{\Sigma}$ and especially to the mean vector $\boldsymbol{\mu}$): solution is **robust optimization** and **improved parameter estimation**,

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 address here the risk diversification among the assets by properly redefining the portfolio formulation.*

---

[1]Z. Zhao, R. Zhou, D. P. Palomar, and Y. Feng, "Portfolio optimization," *submitted, 2019*.

Outline

Outline

Returns

- Let us denote the log-returns of N assets at time t with the vector $\mathbf{r}_t \in \mathbb{R}^N$ (i.e., $r_{it} = \log p_{i,t} - \log p_{i,t-1}$).
- Note that the log-returns are almost the same as the linear returns $R_{it} = \frac{p_{i,t} - p_{i,t-1}}{p_{i,t-1}}$, i.e., $r_{it} \approx R_{it}$.
- 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 as (Feng and Palomar 2016)[2]

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

[2]Y. 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 innacurate in general).

- That is, both the **conditional mean and conditional covariance are constant**:
$$\boldsymbol{\mu}_t = \boldsymbol{\mu},$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}.$$

- Very simple model, however, it is one of the most fundamental assumptions for many important works, e.g., the Nobel prize-winning Markowitz portfolio theory (Markowitz 1952)[3].

---

[3]H. 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 estimators are the sample estimators:
  - **sample mean**: $\quad \hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{r}_t$
  - **sample covariance matrix**: $\quad \hat{\boldsymbol{\Sigma}} = \frac{1}{T-1} \sum_{t=1}^{T} (\mathbf{r}_t - \hat{\boldsymbol{\mu}})(\mathbf{r}_t - \hat{\boldsymbol{\mu}})^T.$

- 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 (Meucci 2005)[4].

- 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 severely affected (Chopra and Ziemba 1993)[5].

- Indeed, this is why Markowitz portfolio and other extensions are rarely used by practitioners.

---

[4]A. Meucci, *Risk and Asset Allocation*. Springer, 2005.
[5]V. Chopra and W. Ziemba, "The effect of errors in means, variances and covariances on optimal portfolio choice," *Journal of Portfolio Management*, 1993.

Outline

Portfolio return

- Suppose the capital budget is B dollars.
- The portfolio $\mathbf{w} \in \mathbb{R}^N$ denotes the normalized dollar weights of the N assets such that $\mathbf{1}^T\mathbf{w} = 1$ (so $B\mathbf{w}$ denotes dollars invested in the assets).
- For each asset i, 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.

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

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

  where r_f is the risk-free rate (e.g., interest rate on a three-month U.S. Treasury bill).
- **Information Ratio (IR)**: SR with $r_f = 0$.
- **Drawdown**: decline from a historical peak of the cumulative profit $X(t)$:

$$D(T) = \max_{t \in [0, T]} X(t) - X(T)$$

- **VaR (Value at Risk)**: quantile of the loss.
- **ES (Expected Shortfall) or CVaR (Conditional Value at Risk)**: expected value of the loss above some quantile.

Practical constraints

- **Capital budget constraint**:

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

- **Long-only constraint**:

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

- **Dollar-neutral or self-financing constraint**:

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

- **Holding constraint**:

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

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

Practical constraints

- **Leverage constraint**:

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

- **Cardinality constraint**:

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

- **Turnover constraint**:

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

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

- **Market-neutral constraint**:

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

Risk control

- 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 **variance** (Markowitz 1952)[6]: 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.

---

[6]H. 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) $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$ (equivalently, the standard deviation or **volatility** $\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$) 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 agressive or risk-averse the investor is.

Mean-variance tradeoff
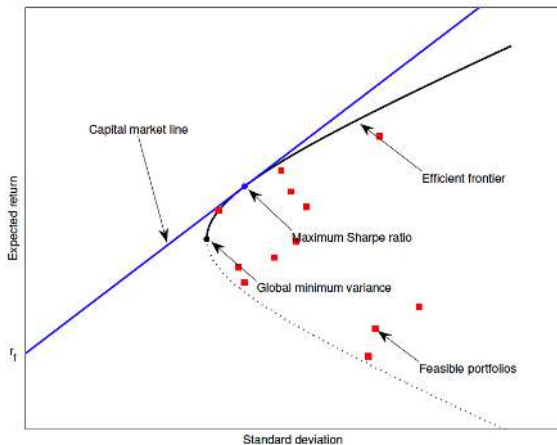
Markowitz mean-variance portfolio (1952)

- The idea of the **Markowitz mean-variance portfolio (MVP)** (Markowitz 1952)[7] 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 $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$:

$$\underset{\mathbf{w}}{\text{maximize}} \quad \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$$
$$\text{subject to} \quad \mathbf{1}^T\mathbf{w} = 1$$

  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 quadratic problem (QP) with only one linear constraint which admits a closed-form solution:

$$\mathbf{w}_{\text{MVP}} = \frac{1}{2\lambda}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \nu\mathbf{1}\right),$$

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

---

[7]H. 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:

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

- It is a simple convex QP with solution

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

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

Outline

Drawbacks of Markowitz's formulation

Markowitz's portfolio has never been fully embraced by practitioners, among other reasons (Zhao et al. 2019)[8] 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 $\boldsymbol{\Sigma}$ and especially to the mean vector $\boldsymbol{\mu}$): solution is **robust optimization** and **improved parameter estimation**,

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 address here the risk diversification among the assets by properly redefining the portfolio formulation.*

---

[8] Z. Zhao, R. Zhou, D. P. Palomar, and Y. Feng, "Portfolio optimization," *submitted, 2019.*

Lack of diversification of Markowitz portfolio

Markowitz mean-variance portfolio (MVP) is typically concentrated in very few assets, while GMVP is more diversified (but not totally):

Portfolio allocation

Outline

Motivation

- The Markowitz mean-variance portfolio has never been fully embraced by practitioners, among other reasons (Zhao et al. 2019)[9] 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)
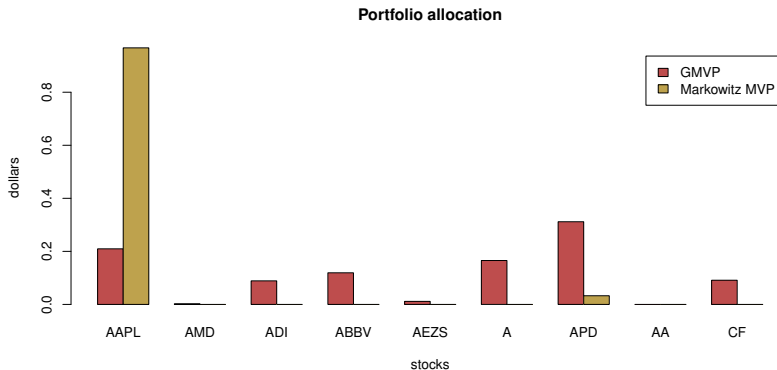 - it is highly sensitive to the estimation errors in the parameters (i.e., small estimation errors in the parameters may change completely the designed portfolio) (Chopra and Ziemba 1993)[10]

- Although portfolio management did not change much during the 40 years after the seminal works of Markowitz and Sharpe, the development of risk budgeting techniques marked an important milestone in deepening the relationship between risk and asset management.

[9]Z. Zhao, R. Zhou, D. P. Palomar, and Y. Feng, "Portfolio optimization," *submitted,* 2019.

[10]V. Chopra and W. Ziemba, "The effect of errors in means, variances and covariances on optimal portfolio choice," *Journal of Portfolio Management,* 1993.

Motivation

- Since the global financial crisis in 2008, **risk management** has particularly become more important than performance management in portfolio optimization
  - risk parity became a popular financial model after the global financial crisis in 2008 (Asness et al. 2012; Qian 2005).

- The alternative **risk parity portfolio design** has been receiving significant attention from both the theoretical and practical sides because it
  - diversifies the risk, instead of the capital, among the assets
  - is less sensitive to parameter estimation errors.

- Today, pension funds and institutional investors are using this approach in the development of smart indexing and the redefinition of long-term investment policies.

From "dollar" to risk diversification

- **Risk parity** is an approach to portfolio management that focuses on **allocation of risk** rather than allocation of capital.

- The **risk parity approach** asserts that when asset allocations are adjusted to the same risk level, the portfolio can achieve a higher Sharpe ratio and can be **more resistant to market downturns**.

- While the minimum variance portfolio tries to minimize the variance (with the disadvantage that a few assets may be the ones contributing most to the risk), the risk parity portfolio tries to constrain **each asset** (or asset class, such as bonds, stocks, real estate, etc.) to **contribute equally to the portfolio overall volatility**.

From "dollar" to risk diversification

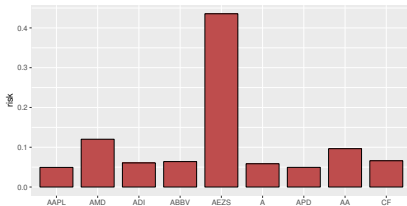- The term "risk parity" was coined by Edward Qian from PanAgora Asset Management (Qian 2005) and was then adopted by the asset management industry.

- Some of its theoretical components were developed in the 1950s and 1960s but the **first risk parity fund, called the "All Weather" fund**, was pioneered by Bridgewater Associates LP in 1996.

- Interest in the risk parity approach has increased since the late 2000s financial crisis as the risk parity approach fared better than traditionally constructed portfolios.

- **Some portfolio managers have expressed skepticism** about the practical application of the concept and its effectiveness in all types of market conditions but others point to its performance during the financial crisis of 2007-2008 as an indication of its potential success.
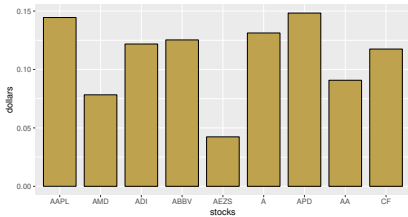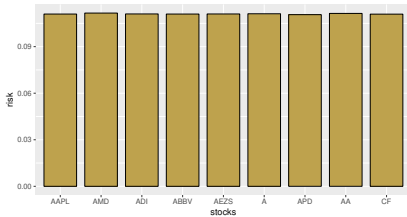
From "dollar" to risk diversification
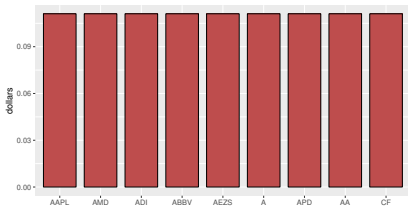
Outline

Risk contribution

- One of the important concepts in portfolio management is quantifying the risk of individual components to the total portfolio risk
- 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

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

- The **marginal risk contribution (MRC)** of the ith asset to the total risk $\sigma(\mathbf{w})$ is defined as

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

  - measures the sensitivity of the portfolio volatility to the ith asset weight
  - MRC can be defined based on other risk measures, like VaR and CVaR.

Risk contribution

- The **risk contribution (RC)** from the ith asset to the total risk $\sigma(\mathbf{w})$ is defined as
$$\mathsf{RC}_i = w_i \frac{\partial \sigma}{\partial w_i} = \frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}}$$

- Observe that (from Euler's theorem)
$$\sum_{i=1}^{N} \mathsf{RC}_i = \sigma(\mathbf{w}).$$

- The **relative risk contribution (RRC)** is defined as the ratio of its RC to the total portfolio risk $\sigma(\mathbf{w})$:
$$\mathsf{RRC}_i = \frac{\mathsf{RC}_i}{\sigma(\mathbf{w})} = \frac{w_i(\boldsymbol{\Sigma}\mathbf{w})_i}{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$$

so that $\sum_{i=1}^{N} \mathsf{RRC}_i = 1$.

Risk parity portfolio (RPP)

- Goal: to allocate the weights so that all the assets contribute the same amount of risk, effectively "equalizing" the risk.
- The **risk parity portfolio (RPP)** or **equal risk portfolio (ERP)** equalizes the risk contributions:

$$RC_i = \sigma(\mathbf{w})/N$$

or

$$RRC_i = 1/N.$$

- Note the parallel with the **equal weight portfolio (EWP)** (aka uniform portfolio):

$$w_i = 1/N.$$

- While the EWP equalizes the capital allocation $w_i = 1/N$, the RPP equalizes the risk allocation $RRC_i = 1/N$.

Risk contribution of EWP

Portfolio allocation of EWP
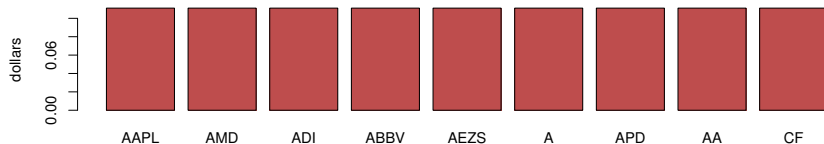
Relative risk contribution of EWP

Risk contribution of RPP

Portfolio allocation of RPP
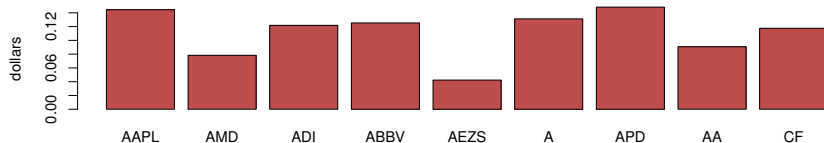
Relative risk contribution of RPP

Risk budgeting portfolio (RBP)

- The RPP aims at allocating the total risk evenly across the assets.
- More generally, the **risk budgeting portfolio (RBP)** allocates the risk according to the risk profile determined by the weights $\mathbf{b}$ (with $\mathbf{1}^T\mathbf{b} = 1$ and $\mathbf{b} \geq \mathbf{0}$):

$$\text{RC}_i = b_i\sigma(\mathbf{w})$$

or

$$\text{RRC}_i = b_i.$$

- We can rewrite $\text{RRC}_i = \frac{w_i(\mathbf{\Sigma w})_i}{\mathbf{w}^T\mathbf{\Sigma w}} = b_i$ simply as

$$w_i(\mathbf{\Sigma w})_i = b_i\mathbf{w}^T\mathbf{\Sigma w}, \qquad i = 1, \ldots, N.$$

- Obviously, RPP is a special case of RBP with $b_i = 1/N$.
- We will consider the more general RBP and we will generally call it RPP with some abuse of terminology
- In general, finding a risk parity portfolio is not trivial.

Risk contribution of RBP

Risk budgeting portfolio with budget $\mathbf{b} \propto (2, 2, 2, 1, 1, 1, 1, 1, 1)$:
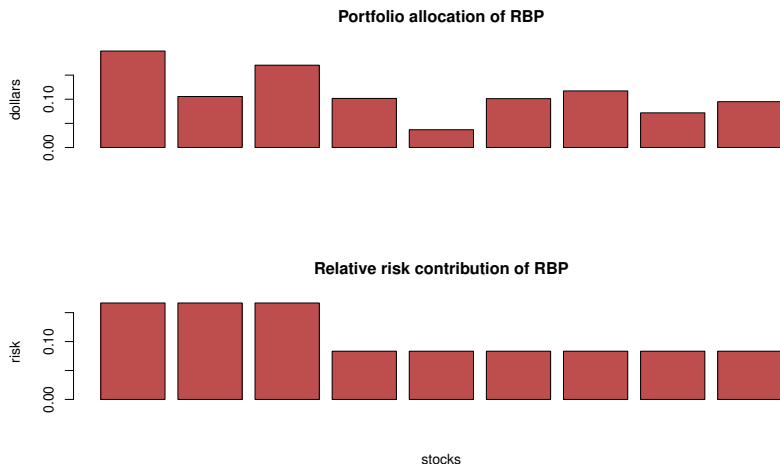
Portfolio allocation of RBP

Relative risk contribution of RBP

Outline

RPP: The diagonal case

- Suppose that the covariance matrix of the returns is diagonal, $\mathbf{\Sigma} = \text{Diag}(\boldsymbol{\sigma}^2)$, and that the portfolio has the constraints $\mathbf{1}^T\mathbf{w} = 1$ and $\mathbf{w} \geq \mathbf{0}$.

- We can then write the risk parity/budgeting constraints $w_i(\mathbf{\Sigma w})_i = b_i \mathbf{w}^T\mathbf{\Sigma w}$ as

$$w_i^2\sigma_i^2 = b_i\sum_{j=1}^{N} w_j^2\sigma_j^2$$

or simply

$$w_i^2\sigma_i^2 \propto b_i$$

which leads to

$$w_i \propto \sqrt{b_i}/\sigma_i.$$

- Observe that the portfolio is inversely proportional to the assets volatilities.

RPP: The diagonal case

- The RPP in the diagonal case is then

$$w_i = \frac{\sqrt{b_i}/\sigma_i}{\sum_{j=1}^{N} \sqrt{b_j}/\sigma_j}, \qquad i = 1, \ldots, N.$$

or, in terms of $\boldsymbol{\Sigma}$,

$$w_i = \frac{\sqrt{b_i}/\sqrt{\Sigma_{ii}}}{\sum_{j=1}^{N} \sqrt{b_j}/\sqrt{\Sigma_{jj}}}, \qquad i = 1, \ldots, N.$$

- 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.

- The previous diagonal solution can be used even when $\boldsymbol{\Sigma}$ is not diagonal and is then called **naive risk budgeting portfolio**.

Risk contribution of naive RPP

The risk contribution of the naive RPP is not perfectly equalized (as expected):

Relative risk contribution
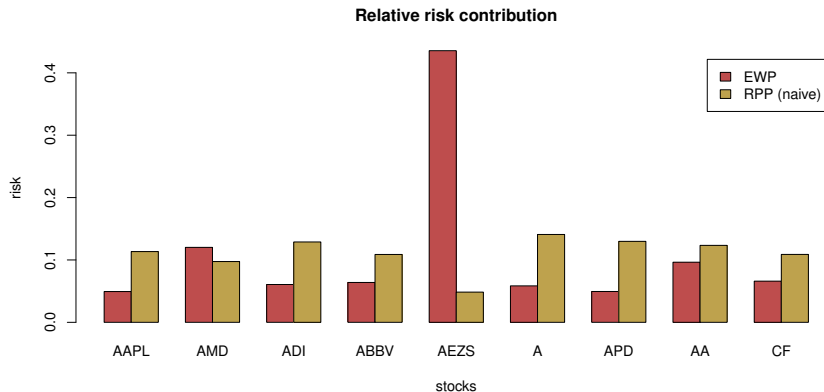
Inverse volatility portfolio

- Similar to RPP, the aim of inverse volatility portfolio (IVP) is to control the portfolio risk.
- The IVP is defined as

$$\mathbf{w} = \frac{\boldsymbol{\sigma}^{-1}}{\mathbf{1}^T \boldsymbol{\sigma}^{-1}}$$

where $\boldsymbol{\sigma}^2 = \text{Diag}(\boldsymbol{\Sigma})$.

- Lower weights are given to high volatility assets and higher weights to low volatility assets
- IVP is also called "equal volatility" portfolio since the weighted constituent assets have equal volatility:

$$\text{sd}(w_i r_i) = w_i \sigma_i = 1/N.$$

- Observe that the IVP coincides with the **naive risk parity portfolio**.

Outline

RPP: Unveiling the hidden convexity

- Consider the risk budgeting equations for an arbitrary covariance matrix $\boldsymbol{\Sigma}$:

$$w_i(\boldsymbol{\Sigma}\mathbf{w})_i = b_i\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}, \qquad i = 1, \ldots, N$$

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

- If we define $\mathbf{x} = \mathbf{w}/\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$, then we can rewrite the risk budgeting equations as $x_i(\boldsymbol{\Sigma}\mathbf{x})_i = b_i$ or, more compactly in vector form, as

$$\boldsymbol{\Sigma}\mathbf{x} = \mathbf{b}/\mathbf{x}$$

with $\mathbf{x} \geq \mathbf{0}$ and we can always recover the portfolio by normalizing: $\mathbf{w} = \mathbf{x}/(\mathbf{1}^T\mathbf{x})$.

- At this point, we can use a nonlinear multivariate root finder for $\boldsymbol{\Sigma}\mathbf{x} = \mathbf{b}/\mathbf{x}$. For example, in R we can use the package **rootSolve**.

Risk contribution of vanilla RPP

The risk contribution of the vanilla RPP is perfectly equalized (unlike the naive diagonal design):

Relative risk contribution
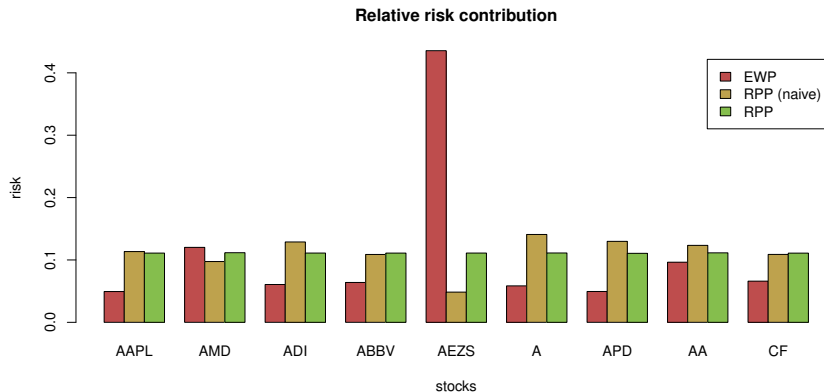
RPP: Unveiling the hidden convexity

- Interestingly, Spinu (2013)[11] realized that precisely the risk budgeting equation $\boldsymbol{\Sigma}\mathbf{x} = \mathbf{b}/\mathbf{x}$ corresponds to the gradient of the convex function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \mathbf{b}^T\log(\mathbf{x})$ set to zero:

$$\nabla f(\mathbf{x}) = \boldsymbol{\Sigma}\mathbf{x} - \mathbf{b}/\mathbf{x} = \mathbf{0}.$$

- This is precisely the optimality condition for the minimization of $f(\mathbf{x})$.
- Thus, we can finally formulate the risk budgeting problem as the following convex optimization problem:

$$\underset{\mathbf{x}\geq\mathbf{0}}{\text{minimize}} \quad \frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x} - \mathbf{b}^T\log(\mathbf{x})$$

which has optimality condition $\boldsymbol{\Sigma}\mathbf{x} = \mathbf{b}/\mathbf{x}$.

---

[11]F. Spinu, "An algorithm for computing risk parity weights," *SSRN*, 2013. [Online]. Available: https://ssrn.com/abstract=2297383.

RPP: Unveiling the hidden convexity

- Griveau-Billion et al. (2013)[12] proposed a slightly different formulation (also convex):

$$\underset{\mathbf{x} \geq \mathbf{0}}{\text{minimize}} \quad \sqrt{\mathbf{x}^T \mathbf{\Sigma} \mathbf{x}} - \mathbf{b}^T \log(\mathbf{x})$$

  with optimality condition $\frac{\mathbf{\Sigma} \mathbf{x}}{\sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}} = \mathbf{b}/\mathbf{x}$ or $\frac{\mathbf{\Sigma} \mathbf{x}}{\sigma} = \mathbf{b}/\mathbf{x}$.

- It looks like the optimal solution is not what we want, but after a careful inspection we can conclude that it is just a different normalization factor from $\mathbf{w}$.

- Simply define $\tilde{\mathbf{x}} = \mathbf{x}/\sigma^{1/2} = \mathbf{w}/\sigma^{3/2}$ to obtain the optimality condition

$$\mathbf{\Sigma} \tilde{\mathbf{x}} = \mathbf{b}/\tilde{\mathbf{x}}$$

  from which we can recover the portfolio by normalizing: $\mathbf{w} = \tilde{\mathbf{x}}/(\mathbf{1}^T \tilde{\mathbf{x}})$.

---

[12] T. Griveau-Billion, J.-C. Richard, and T. Roncalli, "A fast algorithm for computing high-dimensional risk parity portfolios," *SSRN*, 2013. [Online]. Available: https://ssrn.com/abstract=2325255.

RPP: Unveiling the hidden convexity

- Kaya and Lee (2012)[13] proposed yet another reformulation in convex form as the solution to

$$\underset{\mathbf{x} \geq \mathbf{0}}{\text{maximize}} \quad \mathbf{b}^T \log(\mathbf{x})$$
$$\text{subject to} \quad \sigma(\mathbf{x}) \leq \sigma_0.$$

- Ignoring the nonnegativity constraint, the Lagrangian of this constrained convex optimization problem is

$$L(\mathbf{x}; \lambda) = \mathbf{b}^T \log(\mathbf{x}) + \lambda(\sigma_0 - \sqrt{\mathbf{x}^T \mathbf{\Sigma} \mathbf{x}})$$

with gradient

$$\nabla_{\mathbf{x}} L(\mathbf{x}; \lambda) = \mathbf{b}/\mathbf{x} - \lambda \frac{\mathbf{\Sigma} \mathbf{x}}{\sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}}$$

- Defining $\tilde{\mathbf{x}} = (\lambda^{1/2}/\sigma^{1/2})\mathbf{x}$, we can rewrite $\nabla_{\mathbf{x}} L(\mathbf{x}; \lambda) = \mathbf{0}$ as

$$\mathbf{b}/\tilde{\mathbf{x}} = \mathbf{\Sigma} \tilde{\mathbf{x}}$$

which is the desired risk parity/budgeting condition.

[13]H. Kaya and W. Lee, "Demystifying risk parity," *Neuberger Berman*, 2012.

Solving the RPP problem

- A direct way is to attempt to directly solve the nonlinear equations $\mathbf{\Sigma x} = \mathbf{b}/\mathbf{x}$ with a nonlinear multivariate root finder:
  - in R we can use the function multiroot from the package **rootSolve**
  - in Matlab we can use the function fsolve.
- An indirect way is to solve some of the previous convex formulations:

$$\underset{\mathbf{x}\geq\mathbf{0}}{\text{minimize}} \quad \frac{1}{2}\mathbf{x}^T\mathbf{\Sigma x} - \mathbf{b}^T\log(\mathbf{x})$$

- Unfortunately, these convex problems do not conform with the classes most solvers embrace (i.e., LP, QP, QCQP, SOCP, SDP, GP, etc.).
- We can still solve them with a general-purpose solver:
  - in R we can use the function optim
  - in Matlab we can use the function fmincon
- But if we really aim for speed and computational efficiency, there are simple iterative algorithms that can be tailored to the problem at hand, like the **cyclical coordinate descent algorithm** and the **Newton algorithm**.

RPP: Newton method

- Gradient and Newton methods are the most fundamental numerical methods for optimization (Boyd and Vandenberghe 2004).
- The gradient method obtains the iterates based on the gradient ∇f of the objective function $f(\mathbf{x})$ as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \mu \nabla f(\mathbf{x}^{(k)})$$

  but has a slow convergence.
- The Newton method also incorporates the Hessian H:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - H^{-1}(\mathbf{x}^{(k)}) \nabla f(\mathbf{x}^{(k)})$$

  obtaining much faster convergence.
- In practice, one may use the backtracking method to properly adjust the step size of each iteration.
- For our function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{\Sigma} \mathbf{x} - \mathbf{b}^T \log(\mathbf{x})$, the gradient and Hessian are given by

$$\nabla f(\mathbf{x}) = \mathbf{\Sigma}\mathbf{x} - \mathbf{b}/\mathbf{x}$$

$$H(\mathbf{x}) = \mathbf{\Sigma} + \text{Diag}(\mathbf{b}/\mathbf{x}^2).$$

Block coordinate descent (BCD)

The BCD method (aka Gauss-Seidel method) minimizes the function $f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$ with respect to each block of variables one by one in a sequential manner (Bertsekas 1999)[14].

Algorithm 1: BCD

Set $k = 0$ and initialize $\mathbf{x}^{(0)}$
repeat

- Solve sequentially for $i = 1, \ldots, N$:

$$\mathbf{x}_i^{(k+1)} = \arg\min_{\mathbf{x}_i} f\left(\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}_i, \mathbf{x}_{i+1}^{(k)}, \ldots, \mathbf{x}_N^{(k)}\right)$$

- $k \leftarrow k + 1$

until convergence
return $\mathbf{x}^{(k)}$

---

[14] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

Convergence of BCD

Proposition 1:

If $f(\mathbf{x})$ is continuously differentiable and each minimization has a unique solution, then every limit point of the algorithm is a stationary point (optimal point for a convex problem).

RPP: Cyclical coordinate descent algorithm

- The cyclical coordinate descent algorithm is a particular case of the BCD method where $f(\mathbf{x})$ is minimized in a cyclical manner with respect to each element of the variable $\mathbf{x} = (x_1, x_2, \ldots, x_N)$.

- The minimization of $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} - \mathbf{b}^T\log(\mathbf{x})$ with respect to x_i is (denote $\mathbf{x}_{-i} = (x_1, \cdots, x_{i-1}, 0, x_{i+1}, \cdots, x_N)$)

$$\underset{x_i \geq 0}{\text{minimize}} \quad \frac{1}{2}x_i^2\sigma_i^2 + x_i(\mathbf{x}_{-i}^T\mathbf{\Sigma}_{:,i}) - b_i\log x_i$$

with gradient $\nabla_i f = x_i\sigma_i^2 + (\mathbf{x}_{-i}^T\mathbf{\Sigma}_{:,i}) - b_i/x_i$.

- Setting the gradient to zero gives us the second order equation

$$x_i^2\sigma_i^2 + x_i(\mathbf{x}_{-i}^T\mathbf{\Sigma}_{:,i}) - b_i = 0$$

with positive solution given by

$$x_i^\star = \frac{-(\mathbf{x}_{-i}^T\mathbf{\Sigma}_{:,i}) + \sqrt{(\mathbf{x}_{-i}^T\mathbf{\Sigma}_{:,i})^2 + 4\sigma_i^2 b_i}}{2\sigma_i^2}.$$

Outline

RPP: General formulation

- The previous methods are based on a convex reformulation of the problem so they are guaranteed to converge to the optimal risk budgeting solution.
- However, they can only be employed for the simplest risk budgeting formulation with a simplex constraint set (i.e., $\mathbf{1}^T\mathbf{w} = 1$ and $\mathbf{w} \geq \mathbf{0}$).
- They cannot be used if
  - we have other constraints like allowing shortselling or box constraints: $l_i \leq w_i \leq u_i$
  - on top of the risk budgeting constraints $w_i(\boldsymbol{\Sigma}\mathbf{w})_i = b_i\,\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$ we have other objectives like maximizing the expected return $\mathbf{w}^T\boldsymbol{\mu}$ or minimizing the overall variance $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$.
- In those more general cases, we need more sophisticated formulations, which unfortunately are not convex.
- In the R programming language there is a package called **riskParityPortfolio** that can solve very efficiently all the formulations.
- We will overview the different general formulations and the solution methods.

RPP formulations

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

- Maillard et al. (2010)[15] aimed at solving:

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

- This is a simplified formulation with a single-index summation (objective only has N terms instead of N^2):

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

---

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

RBP formulations

- This formulation is again based on the double-index summation with budgets:

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

- This one on a single-index summation:

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

RBP formulations

- Bruder and Roncalli (2012)[16] proposed a formulation based on the RRC:

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

- This one is instead based on the RC:

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

- This one is also similar:

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

---

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

RPP: References

- Two standard textbooks (Qian 2016; Roncalli 2013):

  📕 *T. Roncalli*, Introduction to Risk Parity and Budgeting. *CRC Press, 2013.*

  📕 *E. Qian*, Risk Parity Fundamentals. *CRC Press, 2016.*

- A unified general formulation and advanced algorithms can be found in (Feng and Palomar 2015, 2016):

  📄 *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.*

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

  - A software implementation of the algorithms is available in the R package **riskParityPortfolio**.

Unified RPP problem formulation

A more general risk parity formulation is (Feng and Palomar 2015)[17]

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \sum_{i=1}^{N} g_i(\mathbf{w})^2 + \lambda F(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array}$$

where

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

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

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

Risk concentration term

- The previous general formulation contains the **risk concentration 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 ith row equal to that of the covariance matrix $\boldsymbol{\Sigma}$.
- Examples:
  - $R(\mathbf{w}) = \sum_{i,j=1}^{N} \left(w_i (\boldsymbol{\Sigma}\mathbf{w})_i - w_j (\boldsymbol{\Sigma}\mathbf{w})_j \right)^2$ corresponds to
  $$g_{i,j}(\mathbf{w}) = \mathbf{w}^T (\mathbf{M}_i - \mathbf{M}_j) \mathbf{w}$$
  - $R(\mathbf{w}) = \sum_{i=1}^{N} \left(w_i (\boldsymbol{\Sigma}\mathbf{w})_i - \theta \right)^2$ corresponds to
  $$g_i(\mathbf{w}) = \mathbf{w}^T \mathbf{M}_i \mathbf{w} - \theta$$
  - $R(\mathbf{w}) = \sum_{i=1}^{N} \left(\frac{w_i (\boldsymbol{\Sigma}\mathbf{w})_i}{\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w}} - b_i \right)^2$ corresponds to
  $$g_i(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{M}_i \mathbf{w}}{\mathbf{w}^T \boldsymbol{\Sigma}\mathbf{w}} - b_i.$$

Risk concentration term

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

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

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

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

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

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

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

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

Solving the unified nonconvex RPP problem

- Recall the unified nonconvex RPP formulation:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \sum_{i=1}^{N} g_i(\mathbf{w})^2 + \lambda F(\mathbf{w})$$
$$\text{subject to} \quad \mathbf{w} \in \mathcal{W}.$$

- We can solve this with some general-purpose multivariate optimization solver:
  - in R we can use the function optim
  - in Matlab we can use the function fmincon

- However, for our RPP problem, such off-the-shelf nonlinear numerical optimization methods can be slow and may get stuck at some unsatisfactory points.
- This is because the **structure of the objective is not exploited**.
- We can develop some tailored numerical algorithm with much faster convergence speed and computational efficiency; in particular, we will use the framework of **successive convex approximation (SCA)**.

Slow convergence of general-purpose solvers
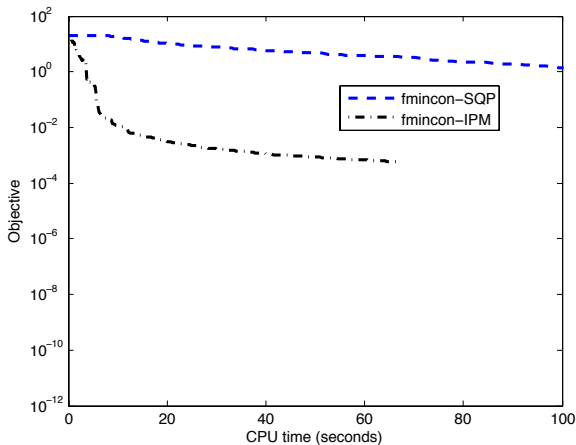
Successive Convex Approximation (SCA)

- Consider our difficult nonconvex problem:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & U(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathcal{W}. \end{array}$$

- Basic idea of SCA: solve a difficult problem via **solving a sequence of simpler problems**.
- Minimize $U(\mathbf{w})$ over $\mathbf{w} \in \mathcal{W}$ via SCA (Scutari et al. 2014)[18]:
 - **Approximation**: find $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$ that approximates the function $U(\mathbf{w})$ at the point $\mathbf{w}^k$ with
    - $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$: uniformly strongly convex & cont. differentiable
    - $\nabla \tilde{U}(\mathbf{w}; \mathbf{w}^k)$: Lipschitz continuous on $\mathcal{W}$
    - $\nabla \tilde{U}(\mathbf{w}; \mathbf{w}^k)|_{\mathbf{w}=\mathbf{w}^k} = \nabla U(\mathbf{w})|_{\mathbf{w}=\mathbf{w}^k}$
  - **Minimization**: minimize $\tilde{U}(\mathbf{w}; \mathbf{w}^k)$ to get the update

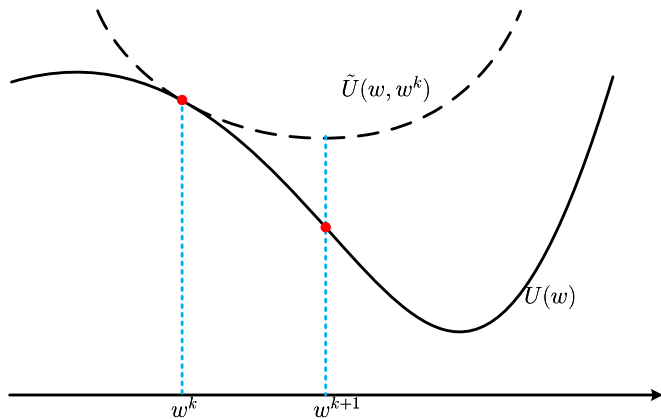$$\mathbf{w}^{k+1} \triangleq \arg \min_{\mathbf{w} \in \mathcal{W}} \tilde{U}(\mathbf{w}; \mathbf{w}^k).$$

---

[18]G. 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 Processing*, vol. 62, no. 3, pp. 641–656, 2014.
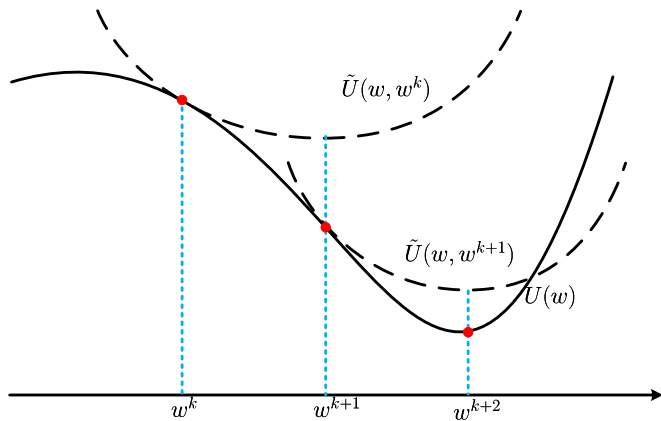
Construction of approximation
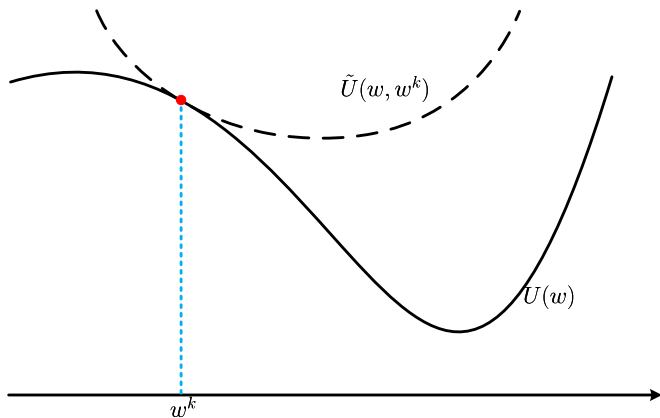
Minimization

Classical methods as SCA

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

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

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

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

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

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

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

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

SCA for RPP optimization

- Recall the objective

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

- At the kth iteration $\mathbf{w}^k$, linearize $g_i(\mathbf{w})$ to construct

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

- Idea: lineare nonconvex functions $g_i(\mathbf{w})$ inside the square leads to quadratic convex $P\left(\mathbf{w}; \mathbf{w}^k\right)$ that approximates $R(\mathbf{w}) = \sum_{i=1}^{N} g_i(\mathbf{w})^2$, with $\nabla P\left(\mathbf{w}, \mathbf{w}^k\right)|_{\mathbf{w}=\mathbf{w}^k} = \nabla R(\mathbf{w})|_{\mathbf{w}=\mathbf{w}^k}$.

SCA for RPP optimization

- $P\left(\mathbf{w}; \mathbf{w}^k\right) = \sum_{i=1}^{N} \left(g_i\left(\mathbf{w}^k\right) + \nabla g_i\left(\mathbf{w}^k\right)^T \left(\mathbf{w} - \mathbf{w}^k\right)\right)^2$ can be rewritten more compactly as

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

where

$$\mathbf{A}^k \triangleq \left[\nabla g_1\left(\mathbf{w}^k\right), \ldots, \nabla g_N\left(\mathbf{w}^k\right)\right]^T,$$
$$\mathbf{g}\left(\mathbf{w}^k\right) \triangleq \left[g_1\left(\mathbf{w}^k\right), \ldots, g_N\left(\mathbf{w}^k\right)\right]^T.$$

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

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

SCA for RPP optimization

- The quadratic program (QP) approximation problem at the kth iteration is

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

where

$$\mathbf{Q}^k \triangleq 2\left(\mathbf{A}^k\right)^T\mathbf{A}^k + \tau\mathbf{I},$$

$$\mathbf{q}^k \triangleq 2\left(\mathbf{A}^k\right)^T\mathbf{g}\left(\mathbf{w}^k\right) - \mathbf{Q}^k\mathbf{w}^k,$$

- This problem can be solved direclty with a QP 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 $\mathbf{Cw} = \mathbf{c}$, 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 $\boldsymbol{\lambda}^k = -\left(\mathbf{C}(\mathbf{Q}^k)^{-1}\mathbf{C}^T\right)^{-1}\left(\mathbf{C}(\mathbf{Q}^k)^{-1}\mathbf{q}^k + \mathbf{c}\right)$.

RPP: Sequential numerical algorithm

Algorithm 2: Successive Convex optimization for RIsk Parity portfolio (SCRIP)

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

repeat

- Solve QP problem to get the optimal solution $\hat{\mathbf{w}}^k$ (global minimum)
- $\mathbf{w}^{k+1} = \mathbf{w}^k + \gamma^k \left(\hat{\mathbf{w}}^k - \mathbf{w}^k \right)$
- $k \leftarrow k + 1$

until convergence
**return** $\mathbf{w}^k$

More advanced algorithms can be found in (Feng and Palomar 2015):

📄 *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.*

Convergence analysis

Proposition 2:

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

Fast numerical convergence of SCA
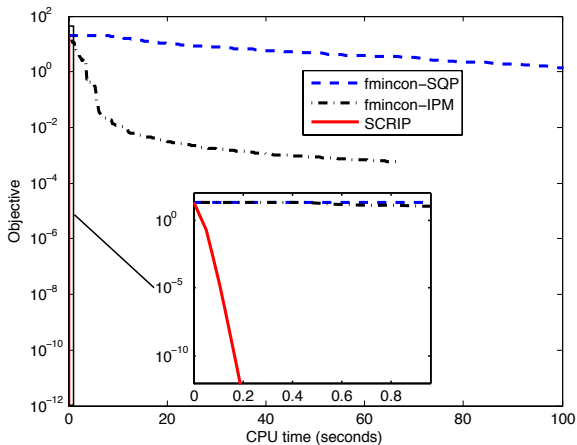
Outline

Conclusions

- We have reviewed the Markowitz portfolio formulation and understood that it has many practical flaws that make it impractical. Indeed, it has not been embraced by practitioners.

- We have learned about the risk parity portfolio formulation.

- We have explored different numerical methods for the risk parity portfolio:
  - the closed-form solution for the naive diagonal formulation
  - many algorithms for the vanilla convex formulation
  - the successive convex approximation (SCA) method for the general nonconvex formulation.

- The performance of risk parity portfolio vs. Markowitz portfolio is much improved.

- Side result: we have learned how to develop efficient numerical algorithms for nonconvex problems based on SCA.

Thanks

For more information visit:

https://www.danielppalomar.com

References I

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2012). Leverage aversion and risk parity. *Financial Analysts Journal*, *68*(1), 47–59.

Bertsekas, D. P. (1999). *Nonlinear programming*. Athena Scientific.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Bruder, B., & Roncalli, T. (2012). *Managing risk exposures using the risk budgeting approach*. University Library of Munich, Germany.

Chopra, V., & Ziemba, W. (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*.

Feng, Y., & Palomar, D. P. (2015). SCRIP: Successive convex optimization methods for risk parity portfolios design. *IEEE Trans. Signal Processing*, *63*(19), 5285–5300.

References II

Feng, Y., & Palomar, D. P. (2016). *A Signal Processing Perspective on Financial Engineering*. Foundations; Trends in Signal Processing, Now Publishers.

Griveau-Billion, T., Richard, J.-C., & Roncalli, T. (2013). A fast algorithm for computing high-dimensional risk parity portfolios. *SSRN*. https://ssrn.com/abstract=2325255

Kaya, H., & Lee, W. (2012). Demystifying risk parity. *Neuberger Berman*.

Maillard, S., Roncalli, T., & Teiletche, J. (2010). The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management*, *36*(4), 60–70.

Markowitz, H. (1952). Portfolio selection. *J. Financ.*, *7*(1), 77–91.

Meucci, A. (2005). *Risk and asset allocation*. Springer.

Qian, E. (2005). Risk parity portfolios: Efficient portfolios through true diversification. *PanAgora Asset Management*.

References III

Qian, E. E. (2016). *Risk parity fundamentals*. CRC Press.

Roncalli, T. (2013). *Introduction to risk parity and budgeting*. CRC Press.

Scutari, G., Facchinei, F., Song, P., Palomar, D. P., & Pang, J.-S. (2014). Decomposition by partial linearization: Parallel optimization of multi-agent systems. *IEEE Trans. Signal Processing*, *62*(3), 641–656.

Spinu, F. (2013). An algorithm for computing risk parity weights. *SSRN*. https://ssrn.com/abstract=2297383

Zhao, Z., Zhou, R., Palomar, D. P., & Feng, Y. (2019). Portfolio optimization. *submitted*.