AN ALGORITHM FOR COMPUTING RISK PARITY WEIGHTS

FLORIN SPINU

ABSTRACT. Given a portfolio of assets (or trading strategies), the risk budget allocation problem seeks the long-only portfolio, fully invested in those assets, with the following property: the contribution of each asset to the risk of the portfolio equals a pre-determined weight. In the particular case when the pre-determined weights are all equal, the solution is the risk parity portfolio. This problem reduces to a non-linear equation which does not have an explicit solution, unless assets are pairwise uncorrelated. However if the matrix covariance is non-degenerate, a solution always exists and is unique. The purpose of this paper is to: a) prove the existence and uniqueness of the solution; b) spell out the relation to the zero-correlation solution; c) give a provably convergent, iterative algorithm (based on Newton's method) that computes the solution. In practice, the algorithm is very efficient even in large dimension: for the risk parity problem in dimension 1400, it typically converges after less than 6 iterations.

1. INTRODUCTION

For $N \geq 2$ a positive integer, let $\mathbb{R}^N_+ = \{x = (x_1, \dots, x_N)^T \in \mathbb{R}^N, x_i > 0\}$ the positive cone, and $\mathcal{S}^N_+$ the set of symmetric, positive definite $N \times N$ matrices. For $x, y \in \mathbb{R}^N$, $xy := (x_1y_1, \dots, x_Ny_N)^T$, and $\text{diag}(x)$ is the diagonal $N \times N$ matrix with x_i on the diagonal. For $x \in \mathbb{R}^N_+$, $x^{-1} := (1/x_1, \dots, 1/x_N) \in \mathbb{R}^N_+$. We will also make use of the norms $\|x\|_2 := (\sum_i x_i^2)^{1/2}$, $\|x\|_\infty := \max_i |x_i|$, and $\|x\|_C := (x^TCx)^{1/2}$.

For a given $C \in \mathcal{S}^N_+$ and $b \in \mathbb{R}^N_+$, we consider the following equation

$$Cx = bx^{-1}, \quad x \in \mathbb{R}^N_+ \ . \tag{1}$$

Theorem 1.1. *The equation (1) has a unique solution $x^* \in \mathbb{R}^N_+$.*

As a consequence, we have the following corollary (cf. [BR12]):

Corollary 1.2. *The vector $x^* / \sum_i x_i^*$ is the unique solution to the system*

$$\frac{(Cx)_i x_i}{x^TCx} = \frac{b_i}{\sum_{j=1}^N b_j}, \quad 1 \leq i \leq N \ , \tag{2}$$

satisfying

$$x_i > 0, \quad \sum_{i=1}^N x_i = 1 \ . \tag{3}$$

OMERS Capital Markets, 200 Bay St., Toronto, ON M5J 2J2, Canada (fspinu@gmail.com).

1.1. **Financial interpretation.** Given a set of N financial assets with covariance matrix C, a long-only, fully invested portfolio in the N assets is determined by the vector of allocation weights $x = (x_1, \dots, x_N)^T$ satisfying (3). (x_i represents the fraction of total capital invested in asset i.) The risk of the portfolio is usually defined as the standard deviation of the portfolio returns

$$\sigma_P(x) := \sqrt{x^T C x} \ . \tag{4}$$

As a homogeneous function of degree one, $\sigma_P(x)$ satisfies Euler's identity

$$\sum_{i=1}^{N} \frac{1}{\sigma_P} x_i \frac{\partial \sigma_P}{\partial x_i}(x) = 1 \ . \tag{5}$$

The i^{th} summation term is considered to be the *contribution* of the asset i to the risk of the total portfolio. Since

$$\frac{1}{\sigma_P} x_i \frac{\partial \sigma_P}{\partial x_i}(x) = \frac{(Cx)_i x_i}{x^T C x}, \tag{6}$$

equation (2) matches the risk contribution of asset i with the predetermined weight $b_i / \sum_j b_j$. For this reason, the solution to (2) is referred to as the *risk budget allocation* in [BR12]. When the weights b_i are equal, the term *risk parity* was introduced earlier in [Q05]. The reader is referred to these papers for an in-depth discussion of the properties of portfolios constructed with these weights.

2. Proof of Theorem 1.1

The proof, as well as the choice of the function below, is similar to the one in [BR12]. Let

$$F(x) := \frac{1}{2} x^T C x - \sum_{i=1}^{N} b_i \log(x_i) \ . \tag{7}$$

The gradient and Hessian of F are given by

$$\nabla F(x) = Cx - bx^{-1}, \quad \nabla^2 F(x) = C + \mathrm{diag}(bx^{-2}) \ . \tag{8}$$

In particular, a solution to (1) is a critical point of F. Since F is strictly convex on $\mathbb{R}^N$ ($\nabla^2 F(x)$ is positive definite), it has at most one critical point (this proves the uniqueness part of the Theorem). To prove that F has at least one critical point, it suffices to show that F has a global minimum $x^* \in \mathbb{R}^N_+$. This is a consequence of the fact that $F(x) \to +\infty$ as $x \to \partial \mathbb{R}^N_+$ (Lemma 4.1 in the Appendix).

2.1. **Proof of Corollary 1.2.** If x^* satisfies (1), $\frac{x^*}{\sum_i x_i^*}$ satisfies (2) and (3). Conversely, if x satisfies (2), $\lambda^{-1} x$ satisfies (1), with $\lambda = \frac{(x^T C x)^{1/2}}{(\sum_i b_i)^{1/2}}$. This implies that $\lambda^{-1} x = x^*$, the unique solution to (1). Furthermore, the constraint (3) implies that $\lambda = \frac{1}{\sum_i x_i^*}$, which determines x uniquely.

3. Newton's algorithm

3.1. **Covariance-correlation weight transfer.** Let D the diagonal of C, and $R = D^{-1/2}CD^{-1/2}$ the associated *correlation matrix*: R is positive definite with 1 on the diagonal. If x^* is the solution to (1), $y^* := D^{1/2}x^*$ is the unique solution to the equation

$$Ry = by^{-1} \tag{9}$$

Therefore, we may assume from now on that C is a correlation matrix:

$$C \in \mathcal{S}_+^N, \quad C_{ii} = 1 \; . \tag{10}$$

In particular, $|C_{ij}| \leq 1$, $\forall i, j$, by the Cauchy-Schwartz inequality.

The next observation is that if x^* satisfies (1), then $t^{1/2}x^*$ satisfies the same equation with the vector b replaced by tb. Therefore we may rescale the vector of weights b so that

$$\min_{1 \leq i \leq N} b_i = 1 \; . \tag{11}$$

In what follows, we refer to [N98, Chap.4] for the notion of *self-concordant* functions. The set of self-concordant functions is closed under addition and multiplication by a scalar greater than one, and it counts among its members all the quadratic functions, as well as the functions $-\log(x_i)$. We conclude that

Proposition 3.1. *Under the assumptions (10) and (11), $F(x)$ is self-concordant.*

3.2. **Convex optimization.** We remarked in Section 2 that the solution to (1) is the global minimum of the function F, defined at (7). This allows us to translate (1) into the convex optimization problem

$$x^* = \mathrm{argmin}_{x \in \mathbb{R}_+^N} F(x) \; . \tag{12}$$

The iteration step in Newton's algorithm is

$$\Delta x := (\nabla^2 F(x))^{-1} \nabla F(x) \; . \tag{13}$$

If x_0 is close enough to x^*, the sequence

$$x_{k+1} = x_k - \Delta x_k, \quad k \geq 0, \tag{14}$$

converges quadratically to x^*. A difficulty arises when an appropriate choice of the initial point x_0 is not available. Since F is self-concordant this issue can be dealt with as in [N98, Chap 4.]. Let

$$\lambda_F(x) := (\nabla F(x), \Delta x)^{1/2} \tag{15}$$

As long as $\lambda_k > \lambda_* := (3 - \sqrt{5})/2$, the above iteration is replaced with

$$x_{k+1} = x_k - \frac{1}{1 + \lambda_k} \Delta x_k, \quad \lambda_k := \lambda_F(x_k) \; . \tag{16}$$

This is the *damped phase* of the algorithm, and is guided by the key estimate [N98, Thm 4.10]

$$F(x_{k+1}) \leq F(x_k) - \omega(\lambda_k), \tag{17}$$

where $\omega(t) := t - \log(1 + t)$. This inequality guarantees a decrease in the objective function of at least $\omega(\lambda_*)$ per iteration. In particular, less than

$$\frac{F(x_0) - F(x^*)}{\omega(\lambda_*)} \tag{18}$$

iterations are required to arrive at $\lambda_k < \lambda_*$, at which point the *quadratic phase* sets in (Newton's full step (14) is used).

We can improve on the generic recipe by exploiting the explicit form of the function F. The following theorem is proved in the Appendix.

Theorem 3.2. *For $x \in \mathbb{R}^N_+$, let $\delta := \max_{1\le i\le N} \frac{|(\Delta x)_i|}{x_i}$. With $h_* = \frac{1}{1+\delta}$, we have*

$$F(x + h_*\Delta x) \le F(x) - (\lambda_F^2(x)/\delta^2)\,\omega(\delta)\ . \tag{19}$$

Consequently, we may replace the iteration step (16) with

$$x_{k+1} = x_k - \frac{1}{1+\delta_k}\Delta x_k, \quad \delta_k := \|\frac{\Delta x_k}{x_k}\|_\infty\ . \tag{20}$$

This step is greater than (16), since $\delta_k < \lambda_k$, while the decrease in the objective function is at least as good as (17): by Lemma 4.3 in the Appendix,

$$\frac{\lambda_k^2}{\delta_k^2}\,\omega(\delta_k) \ge \omega(\lambda_k)\ .$$

The effect is a significant reduction in the number of damped iterations (see below). Let $\mathbf{u}_1 := (1, \dots, 1)^T$ and $S := \sum_i b_i$. We conclude with the following

Theorem 3.3. *Let $x_0 := \frac{\sqrt{\sum_i b_i}}{\sqrt{\mathbf{u}_1^T C \mathbf{u}_1}}\,\mathbf{u}_1$, $\lambda_* := 0.95 \times \frac{3-\sqrt{5}}{2}$, and $Tol > 0$ a termination threshold. Consider the following algorithm: for $k \ge 0$,*

- *Compute*
  $u_k := \nabla F(x_k) = Cx_k - bx_k^{-1}$
  $H_k := \nabla^2 F(x_k) = C + \mathrm{diag}(bx_k^{-2})$
  $\Delta x_k := H_k^{-1} u_k$
  $\delta_k := \|\Delta x_k / x_k\|_\infty$
  $\lambda_k := \sqrt{u_k^T \Delta x_k}$
- *(Damped Phase) While $\lambda_k > \lambda_*$, do*
  $x_{k+1} = x_k - \frac{1}{1+\delta_k}\Delta x_k$.
- *(Quadratic Phase) While $\lambda_k > Tol$, do*
  $x_{k+1} = x_k - \Delta x_k$.

The number of iterations is less than $9.4\,S\log(\kappa_C N)$ in the damped phase, and $(\log_2\log_2(1/Tol) + 2.6)$ in the quadratic phase, with κ_C the condition number of C. The terminal value x_{end} of the algorithm satisfies the bound $\|x_{end} - x^*\|_C \le \frac{Tol}{1-Tol}$.

Proof. The bound on the number of iterations of the quadratic phase is standard, and can be derived from [N98, Thm 4.1.12]. To bound the number of damped iterations, we start from the upper bound (18). First, we remark that $x_0^T C x_0 = (x^*)^T C x^* = S$. Let λ_1 and λ_N be the smallest and the largest eigenvalue of C, respectively. Since $\mathbf{u}_1^T C \mathbf{u}_1 \le \lambda_N \|\mathbf{u}_1\|_2^2 = N\lambda_N$, we have $F(x_0) = \frac{1}{2}S - S\log(\frac{S^{1/2}}{(\mathbf{u}_1^T C\mathbf{u}_1)^{1/2}}) \le \frac{1}{2}S - \frac{1}{2}S\log(\frac{S}{\lambda_N N})$. On the other hand, $S = (x^*)^T C x^* \ge \lambda_1 \|x^*\|_2^2$. This implies $x_i^* \le \sqrt{S/\lambda_1}, \forall i$, hence $F(x^*) \ge \frac{1}{2}S - \frac{1}{2}S\log(\frac{S}{\lambda_1})$. Taking the difference, we obtain $F(x_0) - F(x^*) \le \frac{1}{2}S\log(\kappa_C N)$, where $\kappa_C = \frac{\lambda_N}{\lambda_1}$. Finally $\frac{1}{2}\frac{1}{\omega(\lambda_*)} = 9.385 < 9.4$. The last term satisfies $\|x_k - x^*\|_C \le \|x_k - x^*\|_{x_k} \le \frac{Tol}{1-Tol}$ (cf. [N98, Thm 4.1.11]).

3.3. **Performance.** a) For $N = 50$ and $Tol = 10^{-6}$, the theoretical upper bound for the number of quadratic iterations is 8, while the stated upper bound for the number of damped iterations can be quite large (depending on S). We find in practice however that the total number of iterations is significantly lower than that: for each of the 10^7 trials with random weights b (uniform distribution) and random covariance matrix C (Wishart distribution), the algorithm terminated after less than 16 iterations. b) We have also tested the algorithm on the risk parity problem in dimension $N = 1400$. That is, $b_i = 1/N$ and $Tol = 10^{-6}$. We found that the number of iterations required was less than 6 for each of 2×10^5 trials with random covariance matrix C.

4. Appendix

Lemma 4.1. *$F(x)$ tends to $+\infty$ as x approaches the boundary of $\mathbb{R}^N_+$:*

$$\lim_{x \to \partial \mathbb{R}^N_+} F(x) = +\infty \ . \tag{21}$$

Proof. Let λ_1 be the lowest eigenvalue of C. Since $x^T C x \geq \lambda_1 \sum_i x_i^2$,

$$F(x) \geq \sum_{i=1}^{N} (\frac{1}{2}\lambda_1 x_i^2 - b_i \log(x_i)) =: \sum_{i=1}^{N} g_i(x_i) \ , \tag{22}$$

with $g_i(t) = \frac{1}{2}\lambda_1 t^2 - b_i \log(t)$. The function g_i is convex when $t > 0$, has a global minimum at $t_i^* = \sqrt{b_i/\lambda_1}$, and satisfies

$$\lim_{t \to +\infty} g_i(t) = \lim_{t \to 0+} g_i(t) = +\infty \ . \tag{23}$$

Therefore for each i, $F(x) \geq g_i(x_i) + \sum_{j \neq i} g_j(t_j^*)$.

Let $B > 0$ an arbitrary threshold. By (23) there exist α_i,$\beta_i > 0$ (depending on B) such that

$$x_i \notin [\alpha_i, \beta_i] \Rightarrow F(x) \geq B \ . \tag{24}$$

Therefore $F(x) \geq B$ outside the box $\prod_{i=1}^{N} [\alpha_i, \beta_i]$, and this proves the Lemma.

Lemma 4.2. $\omega(x) > \frac{x^2}{2(x+1)}$, *when* $x > 0$.

Proof. Let $f(x) := \omega(x) - \frac{x^2}{2(x+1)}$. Since $f'(x) = \frac{x^2}{2(x+1)^2} > 0$, f is increasing. Therefore $f(x) > f(0) = 0$, when $x > 0$.

Lemma 4.3. *The function* $\frac{\omega(x)}{x^2}$ *is decreasing on* $(0, \infty)$.

Proof. $\frac{d}{dx}(\frac{\omega(x)}{x^2}) = \frac{2}{x^3}[\frac{x^2}{2(x+1)} - \omega(x)] < 0$, by Lemma 4.2.

4.1. **Proof of Theorem 3.2.** We follow the argument in [N98, Chap.4] and adapt it to our case. From there, we borrow the notations

$$\|\mathbf{u}\|_x := (\mathbf{u}^T \nabla^2 F(x) \mathbf{u})^{1/2}, \quad \lambda_F(x) := \|\Delta x\|_x \ . \tag{25}$$

The starting point is the second order Taylor expansion, with x, $x + \mathbf{u} \in \mathbb{R}^N_+$

$$F(x + \mathbf{u}) - F(x) - \mathbf{u}^T \nabla F(x) = \int_0^1 \int_0^1 t \, \mathbf{u}^T [\nabla^2 F(x + ts \, \mathbf{u})] \mathbf{u} \, ds dt \ . \tag{26}$$

Let $\mathbf{u} := -h\Delta x$. To ensure $x + \mathbf{u} \in \mathbb{R}^N_+$, we need $0 < h < \frac{1}{\delta}$, where

$$\delta := \max_{1 \le i \le N} \frac{|(\Delta x)_i|}{x_i} \; . \tag{27}$$

With $\lambda := \lambda_F(x)$, $\mathbf{u}^T \nabla F(x) = -h\lambda^2$ and (26) can be re-written as

$$F(x - h\Delta x) - F(x) + \lambda^2 h = \int_0^1 \int_0^1 h^2 t \|\Delta x\|^2_{x - hts\Delta x} \, dsdt \; . \tag{28}$$

We remark that $\lambda^2 = (\Delta x)^T C(\Delta x) + \sum_i b_i \frac{\Delta x_i^2}{x_i^2}$ and

$$\|\Delta x\|^2_{x - h\Delta x} = (\Delta x)^T C(\Delta x) + \sum_{i=1}^N b_i \frac{\Delta x_i^2}{x_i^2} \frac{1}{(1 - h\frac{\Delta x_i}{x_i})^2} \; .$$

Since $|1 - h\frac{\Delta x_i}{x_i}| \ge 1 - h\delta$, it follows that

$$\|\Delta x\|^2_{x - h\Delta x} \le \frac{\lambda^2}{(1 - h\delta)^2}, \quad 0 < h < \frac{1}{\delta} \; .$$

Using this estimate, the right-hand side of (28) is less than

$$\int_0^1 \int_0^1 \frac{h^2 t \lambda^2}{(1 - hts\delta)^2} \, dsdt = \frac{\lambda^2}{\delta^2} \, \omega_*(h\delta),$$

where (cf. [N98]) $\omega_*(t) := -t - \log(1 - t)$. Hence (28) yields

$$F(x) - F(x - h\Delta x) \ge \lambda^2 h - \frac{\lambda^2}{\delta^2} \omega_*(h\delta), \quad \forall h \in (0, \frac{1}{\delta}) \; . \tag{29}$$

Let $Q(h)$ denote the function on the right-hand side of the above inequality. Since

$$Q'(h) = \lambda^2 \left(1 - \frac{h}{1 - h\delta} \right), \quad Q''(h) = -\frac{\lambda^2}{(1 - h\delta)^2} < 0 \; ,$$

it follows that Q is concave on $(0, \frac{1}{\delta})$ and achieves its maximum at the critical point

$$h_* := \frac{1}{1 + \delta} \; . \tag{30}$$

Finally, one can check that $Q(h_*) = (\lambda^2/\delta^2)\omega(\delta)$. We conclude that

$$F(x) - F(x - h_*\Delta x) \ge Q(h_*) = (\lambda^2/\delta^2) \, \omega(\delta) \; .$$

References

[BV04] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Seventh printing with corrections, 2009.

[BR12] B. Bruder, T. Roncalli, *Managing Risk Exposures using the Risk Budgeting Approach*, 2012, http://mpra.ub.uni-muenchen.de/37749 (online).

[N98] Yu. Nesterov, *Introductory Lectures on Convex Programming*, Volume I: Basic course, 1998 (online).

[Q05] E.Qian, *Risk Parity Portfolios*, research paper, PanAgora, 2005.