AN ALGORITHM FOR COMPUTING RISK PARITY WEIGHTS

FLORIN SPINU

Abstract. Given a portfolio of assets (or return streams), 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 predetermined weight. When the predetermined weights arel equal the solution is the risk parity portfolio. Mathematically this reduces to to solving the nonlinear equation $Cx = b/x$, where C is a positive definite (covariance) matrix and x and b are column vectors with positive entries. We prove that this equation has unique solution and we give an efficient algorithm to compute it based on Newton's method.

1. Introduction

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 $x = (x_1, \ldots, x_N)$, where $x_i > 0$ represents the fraction of total capital invested in asset i, $\sum_i x_i = 1$. The risk of the portfolio is usually defined as the standard deviation of the portfolio returns, $\sigma_P(x) := \sqrt{x'Cx}$. As a homogeneous function of degree one, $\sigma_P(x)$ satisfies Euler's identity

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

The quantity $MRC_i(x) := \frac{1}{\sigma_P} x_i \frac{\partial \sigma_P}{\partial x_i} = \frac{(Cx)_i x_i}{x'Cx}$ represents the marginal contribution of to portfolio risk of asset i. Note that $\sum_i MRC_i = 1$.

A *risk budget* portfolio (or allocation) is one whose marginal risk contribution from each individual asset that it is invested in matches a predetermined quantity: $MRC_i(x) = b_i$, where $b_i > 0$ are the predetermined weights, $\sum_i b_i = 1$. In other words

$$\sum_i |x_i| = 1, \quad (Cx)_i x_i = b_i x'Cx, \quad 1 \le i \le N \tag{2}$$

The term *risk parity* was introduced in [Q05] in the special case $b_i = \frac{1}{N}$. (The reader is referred to [BR12],[R13] for an in-depth discussion of the properties of portfolios constructed with these weights.)

We use the notation $x^{-1} = (1/x_1, \ldots, x_N)'$ and $xy = (x_1 y_1, \ldots, x_N y_N)'$. The equation (2) is equivalent, after scalar multiplication by $(x'Cx)^{-1/2}$, to

$$Cx = bx^{-1} \ . \tag{3}$$

Throughout the rest of the paper we will be concerned with equation (3). Let $\mathbb{R}_+^N = \{x = (x_1, \ldots, x_N)', x_i > 0\}$ the positive cone.

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

Theorem 1.1. *If C is a non-singular, positive definite matrix, the equation (3) has a unique solution $x^* \in \mathbb{R}^N_+$.*

Corollary 1.2. *The equation (2) has a unique solution in $\mathbb{R}^N_+$, namely $x^*/(\sum_i x_i^*)$.*

The following result clarifies the relevance of the role of the long-only constraint.

Proposition 1.3. *Assume that $\epsilon_i = \pm 1$ is a choice of sign for each coordinate i. The the equation (3) has a unique solution in the cone $sign(x_i) = \epsilon_i$. In particular, equation (3) has exactly 2^N solutions in the set $\{(x_1, \ldots, x_N) \in \mathbb{R}^N : x_i \neq 0, \forall i\}$, one in each of its simply connected components.*

Proof. Let $E = \mathrm{Diag}(\epsilon_i)$ be the $N \times N$ diagonal matrix with ϵ_i on the diagonal. Then $\tilde{C} = ECE$ is positive definite. Let $\tilde{x} \in \mathbb{R}^N_+$ the unique solution to (3) with $\tilde{C}$ instead of C:

$$\tilde{x}_i(\tilde{C}\tilde{x})_i = b_i, \quad \tilde{x}_i > 0, \quad 1 \leq i \leq N \ .$$

Then $x_i = \epsilon_i \tilde{x}_i$ satisfies $Cx = b/x$, and $sign(x_i) = \epsilon_i$. □

2. Proof of Theorem 1.1

The proof, as well as the choice of the utility function, is similar to the one in [BR12]. We define

$$F(x) := \frac{1}{2}x'Cx - \sum_{i=1}^{N} b_i \log(x_i) \ . \tag{4}$$

The gradient and Hessian of F are given by

$$F'(x) = Cx - bx^{-1}, \quad F''(x) = C + \mathrm{diag}(bx^{-2}) \ , \tag{5}$$

where $\mathrm{diag}(d)$ is a diagonal matrix with d on the diagonal. In particular, the solutions to (3) are the critical points of F. Since F is strictly convex on $\mathbb{R}^N$ ($F''(x)$ is positive definite), it has at most one critical point, and 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 following lemma.

Lemma 2.1. *$F(x) \to \infty$ as x approaches the boundary of $\mathbb{R}^N_+$: $\lim_{x \to \partial\mathbb{R}^N_+} F(x) = \infty$.*

Proof. Let λ_1 be the lowest eigenvalue of C. By assumption $\lambda_1 > 0$. We have $x'Cx \geq \lambda_1 \sum_i x_i^2$, hence

$$F(x) \geq \sum_{i=1}^{N} (\frac{1}{2}\lambda_1 x_i^2 - b_i \log(x_i)) = \sum_{i=1}^{N} g_i(x_i) \ , \tag{6}$$

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

$$\lim_{t \to +\infty} g_i(t) = \lim_{t \to 0+} g_i(t) = +\infty \ . \tag{7}$$

In particular, $F(x) \geq g_i(x_i) + \sum_{k \neq i} g_k(t_k^*)$, $\forall i$. Hence $F(x) \geq g_i(x_i) + \mu$, where $\mu := \min_i \sum_{k \neq i} g_k(t_k^*)$. Let $B > 0$ an arbitrary threshold. By (7) there exist $0 < \alpha < \beta$ (depending on B) such that $g_i(t) > B - \mu$, whenever $t \notin [\alpha, \beta]$. This in turn implies that $F(x) > B$ whenever $x \notin \prod_{i=1}^{N} [\alpha, \beta]$. □

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

3. Newton's algorithm

3.1. **Preliminaries.** Let $\mathcal{S}_+^N$ the set of positive definite $N \times N$ matrices. 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 (3), then $y^* := D^{1/2}x^*$ is the unique solution to the equation

$$Ry = by^{-1}, \quad y \in \mathbb{R}_+^N \ . \tag{8}$$

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

$$C \in \mathcal{S}_+^N, \quad C_{ii} = 1, \quad 1 \le i \le N \ . \tag{9}$$

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

The next observation is that if x^* satisfies (3), 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, for purposes which will be transparent later, so that

$$\min_{1 \le i \le N} b_i = 1 \ . \tag{10}$$

3.2. **Convex optimization.** We have established in Section 2 that the solution to (3) is the global minimum of the function F defined at (4). This allows us to translate (3) into the convex optimization problem

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}_+^N} F(x) \ . \tag{11}$$

Given a starting point $x_0 \in \mathbb{R}_+^N$, the sequence generated by Newton's algorithm is defined by (cf. [N98, (4.1.2)]):

$$x_{k+1} = x_k - \Delta x_k, \quad k \ge 0, \tag{12}$$

with the iteration step given by

$$\Delta x_k = [F''(x_k)]^{-1}F'(x_k) \ . \tag{13}$$

If x_0 is close enough to x^*, the sequence converges to x^*. A difficulty arises when such an x_0 is not available. This issue is addressed by Nesterov's theory of self-concordant functions [N98, Chap.4] which we describe briefly below.

3.3. **Self-concordant functions.** Let $\Omega \subset \mathbb{R}^N$ an arbitrary open set. For a convex function $f \in C^3(\Omega)$ we define the norm $\|u\|_x = \langle f''(x)u, u\rangle^{1/2}$.

Definition 3.1. *A convex function $f \in C^3(\Omega)$ is self-concordant if the trilinear form $f'''(x)$ is bounded in the $\|\cdot\|_x$ norm as follows:*

$$|f'''(x)[u,u,u]| \le 2\|u\|_x^3, \quad \forall x \in \Omega, u \in \mathbb{R}^N \tag{14}$$

The set of self-concordant functions is closed under addition and multiplication by a scalar greater than one, and contains all linear and quadratic functions, as well as $-\log(x_i)$. In particular, we have the following proposition.

Proposition 3.2. *Under the assumptions (9) and (10), the function $F(x)$ defined at (4) is self-concordant.*

Nesterov's modification of the Newton algorithm consists of two steps: a) at first, a damped iteration replaces (12) to ensure that a sufficiently small neighborhood of x^* is reached, while staying in the domain of definition of F. b) once this happens, (12) is used. Let

$$\lambda_F(x) := \left\langle [F''(x)]^{-1} F'(x), F'(x) \right\rangle^{1/2} . \tag{15}$$

Let $\lambda_* = 0.95 \times \frac{3-\sqrt{5}}{2}$ (in fact any positive number less than $\frac{3-\sqrt{5}}{2}$ will do).

The *damped iteration* lasts as long as $\lambda_F(x_k) > \lambda_* := \frac{3-\sqrt{5}}{2}$, and the corresponding iteration is

$$x_{k+1} = x_k - \frac{1}{1+\lambda_F(x_k)} \Delta x_k . \tag{16}$$

The key estimate governing this phase is (cf. [N98, Thm 4.10]):

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

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, the *damp phase* terminates after fewer iterations than

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

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

Theorem 3.3. *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

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

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 4.2 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 3.4. *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*
  $u_k := F'(x_k) = C x_k - b x_k^{-1}$
  $H_k := F''(x_k) = C + \mathrm{diag}(b x_k^{-2})$
  $\Delta x_k := H_k^{-1} u_k$
  $\delta_k := \|\Delta x_k / x_k\|_\infty$
  $\lambda_k := \sqrt{u_k^T \Delta x_k}$
- *(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* $\|x_{end} - x^*\|_C \leq \frac{Tol}{1-Tol}$.

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 $\mathbf{u}_1^T C \mathbf{u}_1 \leq \lambda_N \|\mathbf{u}_1\|_2^2 = N\lambda_N$, we have $F(x_0) = \frac{1}{2}S - S\log(\frac{S^{1/2}}{(\mathbf{u}_1^T C \mathbf{u}_1)^{1/2}}) \leq \frac{1}{2}S - \frac{1}{2}S\log(\frac{S}{\lambda_N N})$. On the other hand, $S = (x^*)^T C x^* \geq \lambda_1 \|x^*\|_2^2$. This implies $x_i^* \leq \sqrt{S/\lambda_1}, \forall i$, hence $F(x^*) \geq \frac{1}{2}S - \frac{1}{2}S\log(\frac{S}{\lambda_1})$. Taking the difference, we obtain $F(x_0) - F(x^*) \leq \frac{1}{2}S\log(\kappa_C N)$, where $\kappa_C = \frac{\lambda_N}{\lambda_1}$. Finally $\frac{1}{2}\frac{1}{\omega(\lambda_*)} = 9.385 < 9.4$. The last term satisfies $\|x_k - x^*\|_C \leq \|x_k - x^*\|_{x_k} \leq \frac{Tol}{1-Tol}$ (cf. [N98, Thm 4.1.11]).

3.4. **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. $\omega(x) > \frac{x^2}{2(x+1)}$*, 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 4.2. *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 4.1.

4.1. **Proof of Theorem 3.3.** We follow the argument in [N98, Chap.4] and adapt it to our case. From there, we borrow the notations

$$\|\mathbf{u}\|_x := (\mathbf{u}^T F''(x)\mathbf{u})^{1/2}, \quad \lambda_F(x) := \|\Delta x\|_x \ . \tag{21}$$

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

$$F(x+\mathbf{u}) - F(x) - \mathbf{u}^T F'(x) = \int_0^1 \int_0^1 t\, \mathbf{u}^T [F''(x + ts\,\mathbf{u})]\mathbf{u}\, ds dt \ . \tag{22}$$

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

$$\delta := \max_{1 \leq i \leq N} \frac{|(\Delta x)_i|}{x_i} \ . \tag{23}$$

With $\lambda := \lambda_F(x)$, $\mathbf{u}^T F'(x) = -h\lambda^2$ and (22) can be re-written as

$$F(x - h\Delta x) - F(x) + \lambda^2 h = \int_0^1 \int_0^1 h^2 t \|\Delta x\|^2_{x-hts\Delta x}\, dsdt \; . \tag{24}$$

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

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

Using this estimate, the right-hand side of (24) 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 (24) yields

$$F(x) - F(x - h\Delta x) \geq \lambda^2 h - \frac{\lambda^2}{\delta^2}\omega_*(h\delta), \quad \forall h \in (0, \frac{1}{\delta}) \; . \tag{25}$$

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

$$h_* := \frac{1}{1+\delta} \; . \tag{26}$$

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

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

References

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

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

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

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

[R13] T. Roncalli (2013), *Introduction to Risk Parity and Budgeting*, Chapman and Hall/CRC Financial Mathematics Series.