Efficient Computation of Risk and Jacobians for Risk-Parity Portfolio

Convex Group - HKUST

2018-09-04

Contents

This note contains compact expressions for the risk and corresponding Jacobians for efficient implementation

1 General risk-parity portfolio formulation

The risk-parity portfolio formulation is of the form [1]:

$$\begin{array}{ll} \underset{\mathbf{w}}{\mathsf{minimize}} & R(\mathbf{w}) \\ \mathsf{subject\ to} & \mathbf{1}^T\mathbf{w} = 1, \quad \mathbf{w} \in \mathcal{W}, \end{array}$$

where the risk term is of the form

$$R(\mathbf{w}) = \sum_{i,j=1}^{N} (g_{ij}(\mathbf{w}))^2$$

or simply

$$R(\mathbf{w}) = \sum_{i=1}^{N} (g_i(\mathbf{w}))^2.$$

This problem can be solved directly with some nonlinear solver (for which we need to be able to compute the risk term $R(\mathbf{w})$ (even better if the gradient is computed) as well as with the Successive Convex Approximation (SCA) method developed in [2]. The algorithm iteratively solves a sequence of QP problems of the form:

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

where

$$\begin{aligned} \mathbf{g}\left(\mathbf{w}^k\right) &\triangleq \left[g_1\left(\mathbf{w}^k\right), \ldots, g_N\left(\mathbf{w}^k\right)\right]^T \\ \mathbf{A}^k\left(\mathbf{w}^k\right) &\triangleq \left[\nabla g_1\left(\mathbf{w}^k\right), \ldots, \nabla g_N\left(\mathbf{w}^k\right)\right]^T, \\ \mathbf{Q}^k &\triangleq 2\left(\mathbf{A}^k\right)^T\mathbf{A}^k + \tau\mathbf{I}, \\ \mathbf{q}^k &\triangleq 2\left(\mathbf{A}^k\right)^T\mathbf{g}\left(\mathbf{w}^k\right) - \mathbf{Q}^k\mathbf{w}^k. \end{aligned}$$

To effectively implement the SCA method we need efficient computation of the risk contribution terms contained in $\mathbf{g}(\mathbf{w})$ and their gradients contained in the Jacobian matrix $\mathbf{A}(\mathbf{w}) = [\nabla g_1(\mathbf{w}), \ldots, \nabla g_N(\mathbf{w})]^T$.

Notation:

- Define the i-th risk contribution: $x_i = w_i(\mathbf{\Sigma w})_i$ (in R: `x <- w*(Sigma %*% w)`)
- Gradient of $R(\mathbf{w})$: $\nabla_{\mathbf{w}} R = \left[\frac{\partial R}{\partial w_1}, \ldots, \frac{\partial R}{\partial w_N}\right]^T$
- Jacobian of $\mathbf{w}$: $\mathsf{J}_{\mathbf{w}}\mathbf{x} = \frac{\partial \mathbf{x}}{\partial \mathbf{w}^T}$ (note that the Jacobian of a scalar function is the traspose of the gradient)
- For the single index case: $\mathbf{g}(\mathbf{w}) = [g_1(\mathbf{w}), \ldots, g_N(\mathbf{w})]^T$
- For the double index case: $\mathbf{G}(\mathbf{w}) = (g_{ij}(\mathbf{w}))$ and $\mathbf{g}(\mathbf{w}) = \mathsf{vec}(\mathbf{G}(\mathbf{w}))$
- M-notation [2]: for unification purposes, one can express the risk contributions and risk term in terms of the matrix $\mathbf{M}_i$, whose i-th row is equal to that of $\mathbf{\Sigma}$ and is zero elsewhere. Then it follows that $x_i = w_i(\mathbf{\Sigma w})_i = \mathbf{w}^T\mathbf{M}_i\mathbf{w}$ (although for computational purposes it is far more efficient to use former expression of x_i than the latter).

2 Original formulation with double summation

Let's focus on one specific risk expression:

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

which can be efficiently coded as `risk <- 2*N*sum(x^2) - 2*sum(x)^2`.

Let's compute now the gradient of $R(\mathbf{w})$:

- $\frac{\partial R}{\partial x_i} = 4(Nx_i - \sum_i x_i) \implies \nabla_{\mathbf{x}} R = 4(N\mathbf{x} - (\mathbf{1}^T\mathbf{x})\mathbf{1})$
- $\frac{\partial x_i}{\partial w_j} = w_i\mathbf{\Sigma}_{ij} + \delta_{ij}(\mathbf{\Sigma w})_i \implies \mathsf{J}_{\mathbf{w}}\mathbf{x} = \mathsf{Diag}(\mathbf{w})\mathbf{\Sigma} + \mathsf{Diag}(\mathbf{\Sigma w})$
- chain rule: using Jacobians is $\mathsf{J}_{\mathbf{w}}R = \mathsf{J}_{\mathbf{x}}R \cdot \mathsf{J}_{\mathbf{w}}\mathbf{x}$, using gradients is $(\nabla_{\mathbf{w}}R)^T = (\nabla_{\mathbf{x}}R)^T \cdot \mathsf{J}_{\mathbf{w}}\mathbf{x}$ or, more conveniently, $\nabla_{\mathbf{w}}R = (\mathsf{J}_{\mathbf{w}}\mathbf{x})^T \cdot \nabla_{\mathbf{x}}R$:

$$\nabla_{\mathbf{w}}R = 4(\mathbf{\Sigma}\mathsf{Diag}(\mathbf{w}) + \mathsf{Diag}(\mathbf{\Sigma w}))(N\mathbf{x} - (\mathbf{1}^T\mathbf{x})\mathbf{1}),$$

  which can be coded as `risk_grad <- 4*(Sigma*w + diag(Sigma %*% w)) %*% (N*x-sum(x)*rep(1, N))`. Another way to code it is:
  `v <- (N*x-sum(x)*rep(1, N)) risk_grad <- 4*(Sigma %*% (w*v) + (Sigma %*% w)*v)`

However, if we are interested in implenting the SCA method, this is not enough. Then we need an expression for the risk contributions contained in $\mathbf{g}$ as well as its Jacobian matrix $\mathbf{A}$.

The risk deviations are

$$g_{ij}(\mathbf{w}) = w_i(\mathbf{\Sigma w})_i - w_j(\mathbf{\Sigma w})_j = x_i - x_j,$$

which can be efficiently coded as `g <- rep(x, times = N) - rep(x, each = N)`. So another way to compute $R(\mathbf{w})$ is with `sum(g^2)`, but it's not as efficient as the previous computation since $\mathbf{g}$ has N^2 elements.

Matrix $\mathbf{A}$ is more involved to compute. Using the M-notation: $\nabla g_{ij} = (\mathbf{M}_i + \mathbf{M}_i^T + \mathbf{M}_j + \mathbf{M}_j^T)\mathbf{w}$ (recall that $\mathbf{A} = [\nabla g_{11}, \ldots, \nabla g_{NN}]^T$). But we need an efficient way to compute this...

This derivation is in the making:

Observations:

1. each terms $\nabla g_{ij} = (\mathbf{M}_i + \mathbf{M}_i^T + \mathbf{M}_j + \mathbf{M}_j^T)\mathbf{w}$ is symmetric, i.e., $\nabla g_{ij} = \nabla g_{ji}$

2. we need $\mathbf{A}$ only through the terms $\mathbf{A}^T\mathbf{A}$ (which are symmetric) and $\mathbf{A}^T\mathbf{g}$, which can simplify things a lot

3. can we compute efficiently $(\mathbf{M}_i + \mathbf{M}_i^T)\mathbf{w}$ for all $i = 1, \ldots, N$ at once?. Yes!!:

$$[\mathbf{M}_1\mathbf{w}, \ldots, \mathbf{M}_N\mathbf{w}] = \mathsf{Diag}(\mathbf{\Sigma w})$$

   and

$$\left[\mathbf{M}_1^T\mathbf{w}, \ldots, \mathbf{M}_N^T\mathbf{w}\right] = \mathbf{\Sigma}\mathsf{Diag}(\mathbf{w})$$

So that

$$(\mathbf{M}_i + \mathbf{M}_i^T)\mathbf{w} = [\mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w}) + \boldsymbol{\Sigma}\mathsf{Diag}(\mathbf{w})]_{:,i}$$

and

$$\nabla g_{ij} = [\mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w}) + \boldsymbol{\Sigma}\mathsf{Diag}(\mathbf{w})]_{:,i} + [\mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w}) + \boldsymbol{\Sigma}\mathsf{Diag}(\mathbf{w})]_{:,j}$$
$$[\nabla g_{ij}]_k = \delta_{ki}(\boldsymbol{\Sigma}\mathbf{w})_i + \boldsymbol{\Sigma}_{k,i}w_i + \delta_{kj}(\boldsymbol{\Sigma}\mathbf{w})_j + \boldsymbol{\Sigma}_{k,j}w_j$$

4. TBD:

$$[\mathbf{A}^T\mathbf{A}]_{kl} = \sum_{i,j}[\nabla g_{ij}]_k[\nabla g_{ij}]_l$$

Consider one of the cross terms after the multiplication:

$$\sum_{i,j}\boldsymbol{\Sigma}_{k,i}w_i \cdot \boldsymbol{\Sigma}_{l,j}w_j =$$

3 Formulation with single summation

Consider now the risk expression with a single index:

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

where $b_i = \frac{1}{N}$. This can be efficiently coded as `risk <- sum((x/sum(x)-b)^2)`.

Let's compute now the gradient of $R(\mathbf{w})$:

- first, w.r.t. $\mathbf{x}$:

$$\begin{aligned}\frac{\partial R}{\partial x_j} &= 2\sum_i\left(\frac{x_i}{\mathbf{1}^T\mathbf{x}} - b_i\right)\left(\frac{\delta_{ij}}{\mathbf{1}^T\mathbf{x}} - \frac{x_i}{(\mathbf{1}^T\mathbf{x})^2}\right)\\ &= 2\sum_i\left(\frac{x_i}{\mathbf{1}^T\mathbf{x}} - b_i\right)\frac{\delta_{ij}}{\mathbf{1}^T\mathbf{x}} - 2\sum_i\left(\frac{x_i}{\mathbf{1}^T\mathbf{x}} - b_i\right)\frac{x_i}{(\mathbf{1}^T\mathbf{x})^2}\\ &= \frac{2}{\mathbf{1}^T\mathbf{x}}\left(\left(\frac{x_j}{\mathbf{1}^T\mathbf{x}} - b_j\right) - \left(\frac{\mathbf{x}}{\mathbf{1}^T\mathbf{x}} - \mathbf{b}\right)^T\frac{\mathbf{x}}{(\mathbf{1}^T\mathbf{x})^2}\right)\end{aligned}$$

and

$$\nabla_{\mathbf{x}}R = \frac{2}{\mathbf{1}^T\mathbf{x}}\left(\left(\frac{\mathbf{x}}{\mathbf{1}^T\mathbf{x}} - \mathbf{b}\right) - \mathbf{1}\cdot\frac{\left(\mathbf{x}/(\mathbf{1}^T\mathbf{x}) - \mathbf{b}\right)^T\mathbf{x}}{(\mathbf{1}^T\mathbf{x})^2}\right)$$

- $\frac{\partial x_i}{\partial w_j} = w_i\boldsymbol{\Sigma}_{ij} + \delta_{ij}(\boldsymbol{\Sigma}\mathbf{w})_i \implies \mathsf{J}_{\mathbf{w}}\mathbf{x} = \mathsf{Diag}(\mathbf{w})\boldsymbol{\Sigma} + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w})$
- chain rule: using Jacobians is $\mathsf{J}_{\mathbf{w}}R = \mathsf{J}_{\mathbf{x}}R \cdot \mathsf{J}_{\mathbf{w}}\mathbf{x}$, using gradients is $(\nabla_{\mathbf{w}}R)^T = (\nabla_{\mathbf{x}}R)^T \cdot \mathsf{J}_{\mathbf{w}}\mathbf{x}$ or, more conveniently, $\nabla_{\mathbf{w}}R = (\mathsf{J}_{\mathbf{w}}\mathbf{x})^T \cdot \nabla_{\mathbf{x}}R$:

$$\nabla_{\mathbf{w}}R = \frac{2}{\mathbf{1}^T\mathbf{x}}(\boldsymbol{\Sigma}\mathsf{Diag}(\mathbf{w}) + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w}))\left(\left(\frac{\mathbf{x}}{\mathbf{1}^T\mathbf{x}} - \mathbf{b}\right) - \mathbf{1}\cdot\frac{\left(\mathbf{x}/(\mathbf{1}^T\mathbf{x}) - \mathbf{b}\right)^T\mathbf{x}}{(\mathbf{1}^T\mathbf{x})^2}\right),$$

which can be coded as

```r
sum_x <- sum(x)
x_b <- x/sum_x - b
v <- x_b - rep(1, N) * (t(x_b) %*% x)/(sum_x^2)
risk_grad <- (2/sum_x) * (Sigma %*% (w*v) + (Sigma %*% w)*v)
```

However, if we are interested in implenting the SCA method, this is not enough. Then we need an expression for the risk contributions contained in $\mathbf{g}$ as well as its Jacobian matrix $\mathbf{A}$.

The risk deviations are

$$g_i(\mathbf{w}) = \frac{x_i}{\mathbf{1}^T\mathbf{x}} - b_i,$$

which can be efficiently coded as `g <- x/sum(x) - b`.

Matrix $\mathbf{A} = [\nabla g_1, \ldots, \nabla g_N]^T$ is more involved to compute (we will avoid the M-notation as it gets too involved). The partial derivative is

$$\frac{\partial g_i}{\partial x_j} = \frac{\delta_{ij}}{\mathbf{1}^T\mathbf{x}} - \frac{x_i}{(\mathbf{1}^T\mathbf{x})^2}$$

and the gradient is

$$\nabla_{\mathbf{x}} g_i = \frac{1}{\mathbf{1}^T\mathbf{x}}\left(\mathbf{e}_i - \frac{\mathbf{x}}{\mathbf{1}^T\mathbf{x}}\right)$$

where $\mathbf{e}_i$ is the i-canonical vector (all-zero except a 1 at the i-th position). We can then write the Jacobian as

$$\mathsf{J}_{\mathbf{x}}\mathbf{g} = \frac{1}{\mathbf{1}^T\mathbf{x}}\left(\mathbf{I} - \frac{\mathbf{1}\otimes\mathbf{x}^T}{\mathbf{1}^T\mathbf{x}}\right),$$

where $\otimes$ denotes Kronecker product. Finally we can use the Jacobian chain rule (recall that $\mathsf{J}_{\mathbf{w}}\mathbf{x} = \mathsf{Diag}(\mathbf{w})\boldsymbol{\Sigma} + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w})$):

$$\mathbf{A} \triangleq \mathsf{J}_{\mathbf{w}}\mathbf{g} = \frac{1}{\mathbf{1}^T\mathbf{x}}\left(\mathbf{I} - \frac{\mathbf{1}\otimes\mathbf{x}^T}{\mathbf{1}^T\mathbf{x}}\right)\cdot\left(\mathsf{Diag}(\mathbf{w})\boldsymbol{\Sigma} + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w})\right),$$

which can be more efficiently expressed as

$$\mathbf{A} = \frac{1}{\mathbf{1}^T\mathbf{x}}\left(\left(\mathsf{Diag}(\mathbf{w})\boldsymbol{\Sigma} + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w})\right) - \frac{1}{\mathbf{1}^T\mathbf{x}}\mathbf{1}\otimes\left(\mathbf{x}^T\left(\mathsf{Diag}(\mathbf{w})\boldsymbol{\Sigma} + \mathsf{Diag}(\boldsymbol{\Sigma}\mathbf{w})\right)\right)\right).$$

This can be efficiently coded as

```
sum_x <- sum(x)
Mat <- t(Sigma*w) + diag(Sigma %*% w)
A <- (1/sum_x) * (Mat - (1/sum_x)*matrix((t(x) %*% Mat), N, N, byrow = TRUE)
```

References

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

[2] 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.