Design of Portfolio of Stocks to Track an Index

Konstantinos Benidis and Daniel P. Palomar

2018-04-25

Contents

This vignette illustrates the design of sparse portfolios that aim to track a financila index with the package `sparseIndexTracking` (with a comparison with other packages) and gives a description of the algorithms used.

1 Comparison with other packages

2 Usage of the package

3 Explanation of the algorithms

3.1 `spIndexTrack()`: Sparse portfolio construction

Assume that an index is composed of N assets. We denote by $\mathbf{r}^b = [r_1^b, \dots, r_T^b]^\top \in \mathbb{R}^T$ and $\mathbf{X} = [\mathbf{r}_1, \dots, \mathbf{r}_T]^\top \in \mathbb{R}^{T\times N}$ the (arithmetic) net returns of the index and the N assets in the past T days, respectively, with $\mathbf{r}_t \in \mathbb{R}^N$ denoting the net returns of the N assets at the t-th day.

The goal of `spIndexTrack()` is the design of a (sparse) portfolio $\mathbf{w} \in \mathbb{R}_+^N$, with $\mathbf{w}^\top\mathbf{1} = 1$, that tracks closely the index, i.e., $\mathbf{X}\mathbf{w} \approx \mathbf{r}^b$, based on [1]. The underlying optimization problem that is solved is

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \text{TE}(\mathbf{w}) + \lambda\|\mathbf{w}\|_0 \\ \text{subject to} & \mathbf{w}^\top\mathbf{1} = 1 \\ & \mathbf{0} \le \mathbf{w} \le u\mathbf{1}, \end{array} \qquad (1)$$

where $\text{TE}(\mathbf{w})$ is a general tracking error (we will see specific tracking errors shortly), λ is a regularization parameter that controls the sparsity of the portfolio, and u is an upper bound on the weights of the portfolio.

The ℓ_0-"norm" is approximated by the continuous and differentiable (for $w \ge 0$) function

$$\rho_{p,u}(w) = \frac{\log(1 + w/p)}{\log(1 + u/p)},$$

where $p > 0$ is a parameter that controls the approximation. This leads to the following approximate problem:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathrm{TE}(\mathbf{w}) + \lambda \mathbf{1}^\top \boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \textsf{subject to} & \mathbf{w}^\top \mathbf{1} = 1 \\ & \mathbf{0} \leq \mathbf{w} \leq u\mathbf{1}, \end{array} \tag{2}$$

where $\boldsymbol{\rho}_{p,u}(\mathbf{w}) = [\rho_{p,u}(w_1), \dots, \rho_{p,u}(w_N)]^\top$.

There are four available tracking errors $\mathrm{TE}(\mathbf{w})$ in `spIndexTrack()`:

- Empirical tracking error (ETE):

$$\mathrm{ETE}(\mathbf{w}) = \frac{1}{T}\left\|\mathbf{r}^b - \mathbf{X}\mathbf{w}\right\|_2^2$$

- Downside risk (DR):

$$\mathrm{DR}(\mathbf{w}) = \frac{1}{T}\left\|(\mathbf{r}^b - \mathbf{X}\mathbf{w})^+\right\|_2^2$$

- Huber empirical tracking error (HETE):

$$\mathrm{HETE}(\mathbf{w}) = \frac{1}{T}\mathbf{1}^\top \boldsymbol{\phi}\left(\mathbf{r}^b - \mathbf{X}\mathbf{w}\right)$$

- Huber downside risk (HDR):

$$\mathrm{HDR}(\mathbf{w}) = \frac{1}{T}\mathbf{1}^\top \boldsymbol{\phi}\left((\mathbf{r}^b - \mathbf{X}\mathbf{w})^+\right)$$

where $\boldsymbol{\phi}(\mathbf{x}) = [\phi(x_1), \dots, \phi(x_T)]^\top$, and

$$\phi(x) = \begin{cases} x^2 & |x| \leq M \\ M(2|x| - M) & |x| > M, \end{cases}$$

with $M > 0$ being the Huber parameter.

Regardless of the selected tracking error measure, problem (2) can be solved via Majorization-Minimization (MM) [2] with an iterative closed-form update algorithm (with iterations denoted by k). It can be shown that all of the above variations boil down to the iterative optimization of the following convex problem:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^\top \mathbf{w} + \mathbf{q}^{(k)}{}^\top \mathbf{w} \\ \textsf{subject to} & \mathbf{w} \in \mathcal{W}_u, \end{array} \tag{3}$$

where

$$\mathcal{W}_u = \left\{\mathbf{w} \middle| \mathbf{w}^\top \mathbf{1} = 1, \mathbf{0} \leq \mathbf{w} \leq u\mathbf{1}\right\},$$

and $\mathbf{q}^{(k)} \in \mathbb{R}^N$.

What differentiates the various tracking errors is the exact form of $\mathbf{q}^{(k)}$ that we need to compute at each iteration k of the algorithm:

$$\begin{aligned} \mathbf{q}_{\mathrm{ETE}}^{(k)} &= \frac{1}{\lambda_{\max}^{(\mathbf{L}_1)}}(2(\mathbf{L}_1 - \lambda_{\max}^{(\mathbf{L}_1)}\mathbf{I})\mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top \mathbf{r}^b), \\ \mathbf{q}_{\mathrm{DR}}^{(k)} &= \frac{1}{\lambda_{\max}^{(\mathbf{L}_1)}}(\frac{2}{T}2(\mathbf{L}_1 - \lambda_{\max}^{(\mathbf{L}_1)}\mathbf{I})\mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} + \frac{2}{T}\mathbf{X}^\top (\mathbf{y}^{(k)} - \mathbf{r}^b)), \\ \mathbf{q}_{\mathrm{HETE}}^{(k)} &= \frac{1}{\lambda_{\max}^{(\mathbf{L}_2)}}(2(\mathbf{L}_2 - \lambda_{\max}^{(\mathbf{L}_2)}\mathbf{I})\mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top \mathrm{Diag}(\mathbf{a}^{(k)})\mathbf{r}^b), \\ \mathbf{q}_{\mathrm{HDR}}^{(k)} &= \frac{1}{\lambda_{\max}^{(\mathbf{L}_3)}}(2(\mathbf{L}_3 - \lambda_{\max}^{(\mathbf{L}_3)}\mathbf{I})\mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} + \frac{2}{T}\mathbf{X}^\top \mathrm{Diag}(\mathbf{b}^{(k)})(\mathbf{c}^{(k)} - \mathbf{r}^b)), \end{aligned}$$

where $\lambda_{\max}^{(\mathbf{A})}$ denotes the maximum eigenvalue of a matrix $\mathbf{A}$, $\mathbf{I}$ denotes the identity matrix, $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with the vector $\mathbf{x}$ at its principal diagonal, and

$$\begin{aligned}
\mathbf{d}_{p,u}^{(k)} &= \left[d_{p,u}(w_1^{(k)}), \ldots, d_{p,u}(w_N^{(k)})\right]^\top, \\
d_{p,u}(w^{(k)}) &= \frac{1}{\log(1+u/p)(p+w^{(k)})}, \\
\mathbf{y}^{(k)} &= -(\mathbf{X}\mathbf{w}^{(k)} - \mathbf{r}^b)^+, \\
\mathbf{a}^{(k)} &= [a([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_1), \ldots, a([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_T)]^\top, \\
a(x) &= \begin{cases} 1 & |x| \leq M \\ \frac{M}{|x|} & |x| > M, \end{cases} \\
\mathbf{b}^{(k)} &= [b([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_1), \ldots, b([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_T)]^\top, \\
b(x) &= \begin{cases} \frac{M}{M-2x} & x < 0 \\ 1 & 0 \leq x \leq M \\ \frac{M}{x} & x > M, \end{cases} \\
\mathbf{c}^{(k)} &= [c([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_1), \ldots, c([\mathbf{r}^b - \mathbf{X}\mathbf{w}^{(k)}]_T)]^\top, \\
c(x) &= \begin{cases} x & x < 0 \\ 0 & x \geq 0, \end{cases} \\
\mathbf{L}_1 &= \frac{1}{T}\mathbf{X}^\top \mathbf{X}, \\
\mathbf{L}_2 &= \frac{1}{T}\mathbf{X}^\top \mathrm{Diag}(\mathbf{a}^{(k)})\mathbf{X}, \\
\mathbf{L}_3 &= \frac{1}{T}\mathbf{X}^\top \mathrm{Diag}(\mathbf{b}^{(k)})\mathbf{X}.
\end{aligned}$$

The following propositions provide a waterfilling structured solution of problem (3), considering two special cases, namely, $u = 1$ and $u < 1$.

Proposition 1 The optimal solution of the optimization problem (3) with $u = 1$ is

$$\mathbf{w}^\star = \left(-\frac{1}{2}(\mu\mathbf{1} + \mathbf{q})\right)^+,$$

with

$$\mu = -\frac{\sum_{i\in\mathcal{A}} q_i + 2}{\mathrm{card}(\mathcal{A})},$$

and

$$\mathcal{A} = \{j | \mu + q_j < 0\},$$

where $\mathcal{A}$ can be determined in $O(\log(N))$ steps. We refer to the iterative procedure of Proposition 1 as $\mathrm{AS}_1(\mathbf{q})$ (Active-Set for $u = 1$).

Proposition 2 The optimal solution of the optimization problem (3) with $u < 1$ is

$$\mathbf{w}^\star = \left(\min\left(-\frac{1}{2}(\mu\mathbf{1} + \mathbf{q}), u\mathbf{1}\right)\right)^+,$$

with

$$\mu = -\frac{\sum_{j\in\mathcal{B}_2} q_j + 2 - \mathrm{card}(\mathcal{B}_1)2u}{\mathrm{card}(\mathcal{B}_2)},$$

and

$$\mathcal{B}_1 = \{j \,|\, \mu + q_j \leq -2u\},$$
$$\mathcal{B}_2 = \{j \,|\, -2u < \mu + q_j < 0\},$$

where $\mathcal{B}_1$ and $\mathcal{B}_2$ can be determined in $O(N \log(N))$ steps. We refer to the iterative procedure of Proposition 2 as $\mathrm{AS}_u(\mathbf{q})$ (Active-Set for general $u < 1$).

The iterative closed-form update algorithm is (where $\mathrm{AS}_{1|u}(\mathbf{q})$ means $\mathrm{AS}_1(\mathbf{q})$ or $\mathrm{AS}_u(\mathbf{q})$):

Algorithm 1

```
1. Set $k = 0$ and choose an initial point $\mathbf{w}^{(0)}$
2. Compute $\mathbf{q}$ according to the selected tracking error
3. Find the optimal solution $\mathbf{w}^\star$ with $\mathrm{AS}_{1|u}(\mathbf{q})$ and set it equal to $\mathbf{w}^{(k+1)}$ 4. $k \leftarrow k + 1$
5. Repeat steps 2-4 until convergence
6. Return $\mathbf{w}^{(k)}$
```

The initial point of the algorithm $\mathbf{w}^{(0)}$ is set by default to $1/N$, unless the user specifies otherwise.

Finally, note that the approximate problem is controlled by the parameter p, and in particular, as $p \to 0$ we get $\rho_{p,u} \to \ell_0$. However, by setting small values to p, it is likely that the algorithm will get stuck to a local minimum. To solve this issue we start with large values for p, i.e., a "loose" approximation, and solve the corresponding optimization problem. Then, we sequentially decrease p, i.e., we "tighten" the approximation, and solve the problem again using the previous solution as an initial point. In practice we are interested only in the last, "tightest" problem. For each problem that is solved (i.e., for fixed p) we utilize an acceleration scheme that increases the convergence speed of the MM algorithm. For details, please refer to [3].

References

[1] K. Benidis, Y. Feng, and D. P. Palomar, "Sparse portfolios for high-dimensional financial index tracking," *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 155–170, Jan. 2018.

[2] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[3] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any em algorithm," *Scandinavian Journal of Statistics*, vol. 35, no. 2, pp. 335–353, 2008.