Sparse Index Tracking via MM

Prof. Daniel P. Palomar
The Hong Kong University of Science and Technology (HKUST)

MAFS6010R - Portfolio Optimization with R
MSc in Financial Mathematics
Fall 2018-19, HKUST, Hong Kong

Outline

Outline

Investment Strategies

Fund managers follow two basic investment strategies:

Active

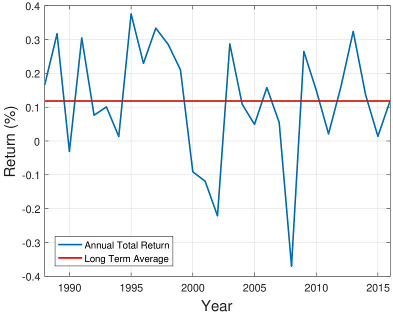
- Assumption: markets are not perfectly efficient.
- Through expertise add value by choosing high performing assets.

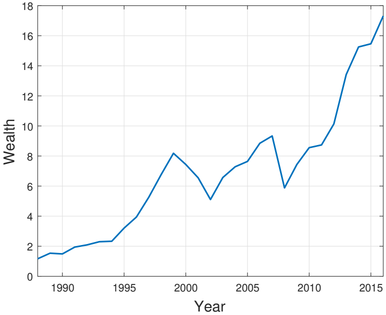
Passive

- Assumption: market cannot be beaten in the long run.
- Conform to a defined set of criteria (e.g. achieve same return as an index).

Passive Investment

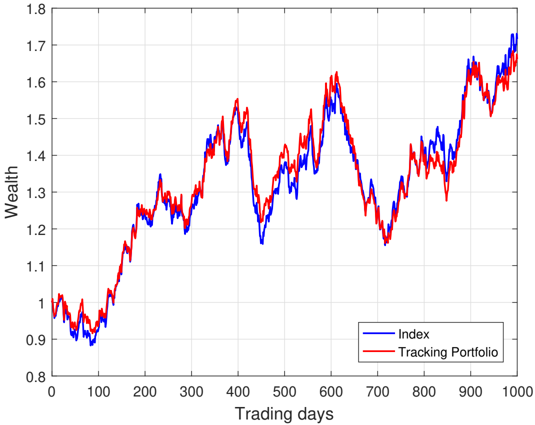
The stock markets have historically risen, e.g. S&P 500:





- Partly misleading: e.g. inflation.
- Still, reasonable returns can be obtained without the active management's risk.
- Makes passive investment more attractive.

Index Tracking



- Index tracking is a popular passive portfolio management strategy.
- **Goal**: construct a portfolio that replicates the performance of a financial index.

Index Tracking

- **Index tracking** or **benchmark replication** is a strategy investment aimed at mimicking the risk/return profile of a financial instrument.
- For practical reasons, the strategy focuses on a **reduced basket** of representative assets.
- The problem is also regarded as portfolio compression and it is intimately related to compressed sensing and ℓ_1-norm minimization techniques.[1,2]
- One example is the replication of an index, e.g., Hang Seng Index, based on a reduced basket of assets.

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

[2] K. Benidis, Y. Feng, and D. P. Palomar, *Optimization Methods for Financial Index Tracking: From Theory to Practice*. Foundations and Trends in Optimization, Now Publishers, 2018.

Definitions

- Price and return of an asset or an index: p_t and $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$
- Returns of an index in T days: $\mathbf{r}^b = [r_1^b, \ldots, r_T^b]^\top \in \mathbb{R}^T$
- Returns of N assets in T days: $\mathbf{X} = [\mathbf{r}_1, \ldots, \mathbf{r}_T]^\top \in \mathbb{R}^{T \times N}$ with $\mathbf{r}_t \in \mathbb{R}^N$
- Assume that an index is composed by a weighted collection of N assets with normalized index weights $\mathbf{b}$ satisfying
  - $\mathbf{b} > \mathbf{0}$
  - $\mathbf{b}^\top \mathbf{1} = 1$
  - $\mathbf{X}\mathbf{b} = \mathbf{r}^b$
- We want to design a (sparse) tracking portfolio $\mathbf{w}$ satisfying
  - $\mathbf{w} \geq \mathbf{0}$
  - $\mathbf{w}^\top \mathbf{1} = 1$
  - $\mathbf{X}\mathbf{w} \approx \mathbf{r}^b$

Full Replication

- How should we select $\mathbf{w}$?

- Straightforward solution: full replication $\mathbf{w} = \mathbf{b}$
  - Buy appropriate quantities of all the assets
  - Perfect tracking

- But it has drawbacks:
  - We may be trying to hedge some given portfolio with just a few names (to simplify the operations)
  - We may want to deal properly with illiquid assets in the universe
  - We may want to control the transaction costs for small portfolios (AUM)

Full Replication

- How should we select $\mathbf{w}$?

- Straightforward solution: full replication $\mathbf{w} = \mathbf{b}$
  - Buy appropriate quantities of all the assets
  - Perfect tracking

- But it has drawbacks:
  - We may be trying to hedge some given portfolio with just a few names (to simplify the operations)
  - We may want to deal properly with illiquid assets in the universe
  - We may want to control the transaction costs for small portfolios (AUM)

Full Replication

- How should we select $\mathbf{w}$?
- Straightforward solution: full replication $\mathbf{w} = \mathbf{b}$
  - Buy appropriate quantities of all the assets
  - Perfect tracking
- But it has drawbacks:
  - We may be trying to hedge some given portfolio with just a few names (to simplify the operations)
  - We may want to deal properly with illiquid assets in the universe
  - We may want to control the transaction costs for small portfolios (AUM)

Sparse Index Tracking

- How can we overcome these drawbacks?
  $\Longrightarrow$ Sparse index tracking.

- Use a small number of assets: $\text{card}(\mathbf{w}) < N$
  - can allow hedging with just a few names
  - can avoid illiquid assets
  - can reduce transaction costs for small portfolios

- Challenges:
  - Which assets should we select?
  - What should their relative weight be?

Sparse Index Tracking

- How can we overcome these drawbacks?
  $\Longrightarrow$ Sparse index tracking.

- Use a small number of assets: $\text{card}(\mathbf{w}) < N$
  - can allow hedging with just a few names
  - can avoid illiquid assets
  - can reduce transaction costs for small portfolios

- Challenges:
  - Which assets should we select?
  - What should their relative weight be?

Sparse Index Tracking

- How can we overcome these drawbacks?
  $\Longrightarrow$ Sparse index tracking.

- Use a small number of assets: $\text{card}(\mathbf{w}) < N$
  - can allow hedging with just a few names
  - can avoid illiquid assets
  - can reduce transaction costs for small portfolios

- Challenges:
  - Which assets should we select?
  - What should their relative weight be?

Outline

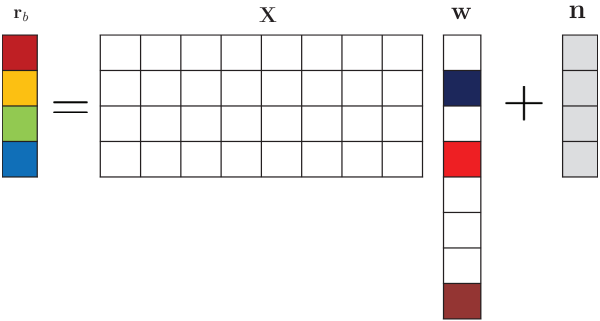
Outline

Sparse Regression

- Sparse regression:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \|\mathbf{r} - \mathbf{X}\mathbf{w}\|_2 + \lambda \|\mathbf{w}\|_0$$

tries to fit the observations by minimizing the error with a sparse solution:



Tracking error

- Recall that $\mathbf{b} \in \mathbb{R}^N$ represents the actual benchmark weight vector and $\mathbf{w} \in \mathbb{R}^N$ denotes the replicating portfolio.
- Investment managers seek to minimize the following **tracking error (TE)** performance measure:

$$\mathrm{TE}\left(\mathbf{w}\right) = \left(\mathbf{w} - \mathbf{b}\right)^T \boldsymbol{\Sigma} \left(\mathbf{w} - \mathbf{b}\right)$$

  where $\boldsymbol{\Sigma}$ is the covariance matrix of the index returns.
- In practice, however, the benchmark weight vector $\mathbf{b}$ may be unknown and the error measure is defined in terms of market observations.
- A common tracking measure is the **empirical tracking error (ETE)**:

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

Formulation for Sparse Index Tracking

- Problem formulation for sparse index tracking:[3]

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\|_2^2 + \lambda\|\mathbf{w}\|_0 \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array} \qquad (1)$$

  - $\|\mathbf{w}\|_0$ is the ℓ_0-"norm" and denotes card($\mathbf{w}$)
  - $\mathcal{W}$ is a set of convex constraints (e.g., $\mathcal{W} = \{\mathbf{w} | \mathbf{w} \geq \mathbf{0}, \mathbf{w}^\top \mathbf{1} = 1\}$)
  - we will treat any nonconvex constraint separately

- Problem (1) is too difficult to deal with directly:
  - Discontinuous, non-differentiable, non-convex objective function.

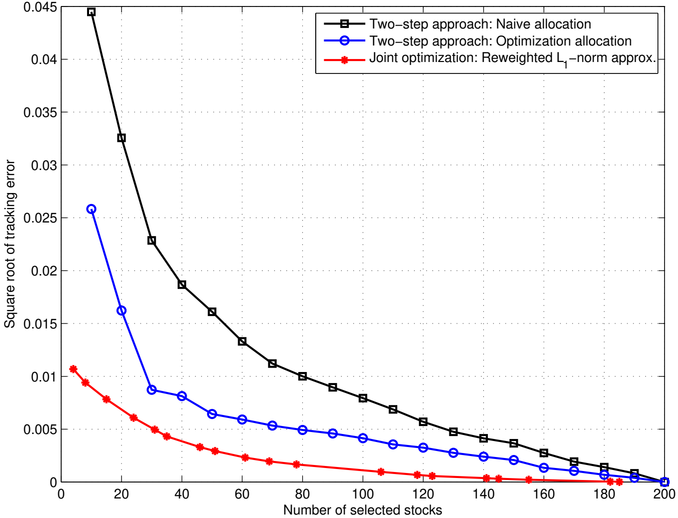
[3] D. Maringer and O. Oyewumi, "Index tracking with constrained portfolios," *Intelligent Systems in Accounting, Finance and Management*, vol. 15, no. 1-2, pp. 57–71, 2007.

Existing Methods

- Two step approach:
  1. stock selection:
     - largest market capital
     - most correlated to the index
     - a combination cointegrated well with the index
  2. capital allocation:
     - naive allocation: proportional to the original weights
     - optimized allocation: usually a convex problem
- Mixed Integer Programming (MIP)
  - practical only for small dimensions, e.g. $\binom{100}{20} > 10^{20}$.
- Genetic algorithms
  - solve the MIP problems in reasonable time
  - worse performance, cannot prove optimality.

Existing Methods

- Two-step approach is much worse than joint optimization:



Outline

Interlude: Majorization-Minimization (MM)

- Consider the following presumably difficult optimization problem:

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x})\\ \text{subject to} & \mathbf{x}\in\mathcal{X},\end{array}$$

  with $\mathcal{X}$ being the feasible set and $f(\mathbf{x})$ being continuous.
- Idea: successively minimize a more managable surrogate function $u\left(\mathbf{x},\mathbf{x}^{(k)}\right)$:

$$\mathbf{x}^{(k+1)}=\arg\min_{\mathbf{x}\in\mathcal{X}} u\left(\mathbf{x},\mathbf{x}^{(k)}\right),$$

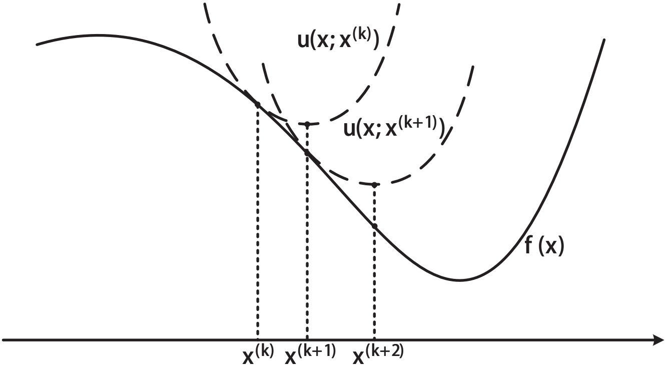
  hoping the sequence of minimizers $\left\{\mathbf{x}^{(k)}\right\}$ will converge to optimal $\mathbf{x}^{\star}$.
- Question: how to construct $u\left(\mathbf{x},\mathbf{x}^{(k)}\right)$?
- Answer: that's more like an art.[4]

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

Interlude on MM: Surrogate/Majorizer

- Construction rule:

$$
\begin{aligned}
u(\mathbf{y},\mathbf{y}) &= f(\mathbf{y}),\ \forall \mathbf{y}\in\mathcal{X} \\
u(\mathbf{x},\mathbf{y}) &\geq f(\mathbf{x}),\ \forall \mathbf{x},\mathbf{y}\in\mathcal{X} \\
u'(\mathbf{x},\mathbf{y};\mathbf{d})\big|_{\mathbf{x}=\mathbf{y}} &= f'(\mathbf{y};\mathbf{d}),\ \forall \mathbf{d} \text{ with } \mathbf{y}+\mathbf{d}\in\mathcal{X} \\
u(\mathbf{x},\mathbf{y}) & \text{ is continuous in } \mathbf{x} \text{ and } \mathbf{y}
\end{aligned}
$$



Interlude on MM: Algorithm

Algorithm MM:

Find a feasible point $\mathbf{x}^0 \in \mathcal{X}$ and set $k = 0$.
repeat
$\quad \mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x} \in \mathcal{X}} u\left(\mathbf{x}, \mathbf{x}^{(k)}\right)$
$\quad k \leftarrow k + 1$
until some convergence criterion is met

Interlude on MM: Convergence

- Under some technical assumptions, every limit point of the sequence $\left\{\mathbf{x}^k\right\}$ is a stationary point of the original problem.
- If further assume that the level set $\mathcal{X}^0 = \{\mathbf{x} | f(\mathbf{x}) \leq f(\mathbf{x}^0)\}$ is compact, then

$$\lim_{k\to\infty} d\left(\mathbf{x}^{(k)}, \mathcal{X}^\star\right) = 0,$$

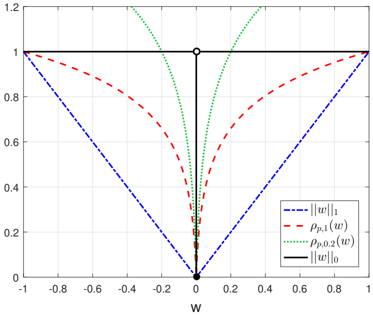
where $\mathcal{X}^\star$ is the set of stationary points.

Outline

Sparse Index Tracking via MM

- Approximation of the ℓ_0-norm (indicator function):

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



- Good approximation in the interval $[-\gamma, \gamma]$.
- Concave for $w \geq 0$.
- So-called folded-concave for $w \in \mathbb{R}$.
- For our problem we set $\gamma = u$, where $u \leq 1$ is an upperbound of the weights (we can always choose $u = 1$).

Approximate Formulation

- Continuous and differentiable approximate formulation:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \text{subject to} & \mathbf{w}\in\mathcal{W} \end{array} \tag{2}$$

- $\boldsymbol{\rho}_{p,u}(\mathbf{w}) = [\rho_{p,u}(w_1),\ldots,\rho_{p,u}(w_N)]^\top$.
- Problem (2) is still non-convex: $\boldsymbol{\rho}_{p,u}(\mathbf{w})$ is concave for $\mathbf{w}\geq\mathbf{0}$.
- We will use MM to deal with the non-convex part.

Majorization of $\rho_{p,\gamma}$

Lemma 1

The function $\rho_{p,\gamma}(w)$, with $w \geq 0$, is upperbounded at $w^{(k)}$ by the surrogate function

$$h_{p,\gamma}(w, w^{(k)}) = d_{p,\gamma}(w^{(k)})w + c_{p,\gamma}(w^{(k)}),$$

where

$$d_{p,\gamma}(w^{(k)}) = \frac{1}{\log(1+\gamma/p)(p+w^{(k)})},$$

$$c_{p,\gamma}(w^{(k)}) = \frac{\log\left(1+w^{(k)}/p\right)}{\log(1+\gamma/p)} - \frac{w^{(k)}}{\log(1+\gamma/p)(p+w^{(k)})}$$

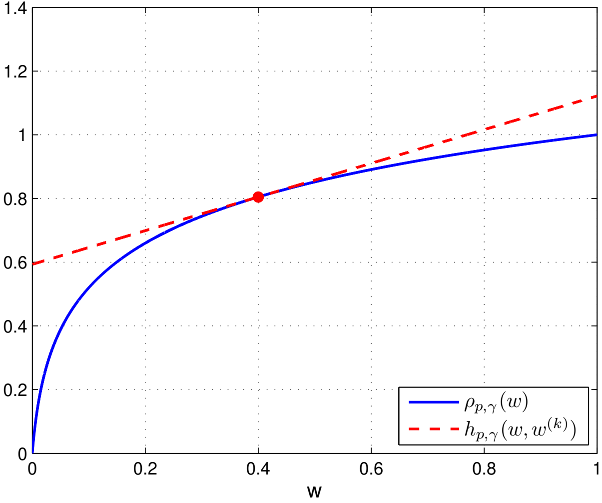
are constants.

Proof of Lemma 1

- The function $\rho_{p,\gamma}(w)$ is concave for $w \geq 0$.
- An upper bound is its first-order Taylor approximation at any point $w_0 \in \mathbb{R}_+$.

$$
\begin{aligned}
\rho_{p,\gamma}(w) &= \frac{\log(1+w/p)}{\log(1+\gamma/p)} \\
&\leq \frac{1}{\log(1+\gamma/p)}\left[\log\left(1+w_0/p\right)+\frac{1}{p+w_0}(w-w_0)\right] \\
&= \underbrace{\frac{1}{\log(1+\gamma/p)(p+w_0)}}_{d_{p,\gamma}} w \\
&\qquad + \underbrace{\frac{\log\left(1+w_0/p\right)}{\log(1+\gamma/p)} - \frac{w_0}{\log(1+\gamma/p)(p+w_0)}}_{b_{p,\gamma}}
\end{aligned}
$$

Majorization of $\rho_{p,\gamma}$



Iterative Formulation via MM

- Now in every iteration we need to solve the following problem:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\right\|_2^2 + \lambda \mathbf{d}_{p,u}^{(k)}{}^\top \mathbf{w} \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array} \tag{3}$$

- $\mathbf{d}_{p,u}^{(k)} = \left[d_{p,u}(w_1^{(k)}), \ldots, d_{p,u}(w_N^{(k)})\right]^\top$.

- Problem (3) is convex (QP).

- Requires a solver in each iteration.

Algorithm LAIT

Algorithm 1: Linear Approximation for the Index Tracking problem (LAIT)

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}$

repeat

  Compute $\mathbf{d}_{p,u}^{(k)}$

  Solve (3) with a solver and set the optimal solution as $\mathbf{w}^{(k+1)}$

  $k \leftarrow k + 1$

until convergence

return $\mathbf{w}^{(k)}$

The Big Picture

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\|\mathbf{w}\|_0 \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

ℓ_0-norm approximation

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

Should we stop here?

- Advantages:
  - ✓ The problem is convex.
  - ✓ Can be solved efficiently by an off-the-shelf solver.
- Disadvantages:
  - × Needs to be solved many times (one for each iteration).
  - × Calling a solver many times increases significantly the running time.
- Can we do something better?
  - ✓ For specific constraint sets we can derive closed-form update algorithms!

Let's rewrite the objective function

- Expand the objective:
$$\frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w} = \frac{1}{T}\mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w}+\left(\lambda\mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top\mathbf{r}^b\right)^\top\mathbf{w}+const.$$

- Further upper-bound it:

Lemma 2

Let $\mathbf{L}$ *and* $\mathbf{M}$ *be real symmetric matrices such that* $\mathbf{M}\succeq\mathbf{L}$*. Then, for any point* $\mathbf{w}^{(k)}\in\mathbb{R}^N$ *the following inequality holds:*

$$\mathbf{w}^\top\mathbf{L}\mathbf{w}\leq\mathbf{w}^\top\mathbf{M}\mathbf{w}+2\mathbf{w}^{(k)\top}\left(\mathbf{L}-\mathbf{M}\right)\mathbf{w}-\mathbf{w}^{(k)\top}\left(\mathbf{L}-\mathbf{M}\right)\mathbf{w}^{(k)}.$$

Equality is achieved when $\mathbf{w}=\mathbf{w}^{(k)}$*.*

Let's majorize the objective function

- Based on Lemma 2:
  - Majorize the quadratic term $\frac{1}{T}\mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\mathbf{w}$.
  - In our case $\mathbf{L}_1 = \frac{1}{T}\mathbf{X}^\top\mathbf{X}$.
  - We set $\mathbf{M}_1 = \lambda_{\max}^{(\mathbf{L}_1)}\mathbf{I}$ so that $\mathbf{M}_1 \succeq \mathbf{L}_1$ *holds*.
- The objective becomes:

$$
\begin{aligned}
&\mathbf{w}^\top\mathbf{L}_1\mathbf{w} + \left(\lambda\mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top\mathbf{r}^b\right)^\top\mathbf{w}\\
&\leq \mathbf{w}^\top\mathbf{M}_1\mathbf{w} + 2\mathbf{w}^{(k)^\top}(\mathbf{L}_1 - \mathbf{M}_1)\,\mathbf{w} - \mathbf{w}^{(k)^\top}(\mathbf{L}_1 - \mathbf{M}_1)\,\mathbf{w}^{(k)}\\
&\qquad + \left(\lambda\mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top\mathbf{r}^b\right)^\top\mathbf{w}\\
&= \lambda_{\max}^{(\mathbf{L}_1)}\mathbf{w}^\top\mathbf{w} + \left(2\left(\mathbf{L}_1 - \lambda_{\max}^{(\mathbf{L}_1)}\mathbf{I}\right)\mathbf{w}^{(k)} + \lambda\mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top\mathbf{r}^b\right)^\top\mathbf{w} + const.
\end{aligned}
$$

Specialized Iterative Formulation

The new optimization problem at the $(k+1)-th$ iteration becomes:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^\top \mathbf{w} + \mathbf{q}_1^{(k)\top} \mathbf{w} \\ \text{subject to} & \left.\begin{array}{l} \mathbf{w}^\top \mathbf{1} = 1, \\ \mathbf{0} \leq \mathbf{w} \leq \mathbf{1}, \end{array}\right\} \mathcal{W} \end{array} \tag{4}$$

where

$$\mathbf{q}_1^{(k)} = \frac{1}{\lambda_{\max}^{(\mathbf{L}_1)}} \left(2 \left(\mathbf{L}_1 - \lambda_{\max}^{(\mathbf{L}_1)} \mathbf{I} \right) \mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} - \frac{2}{T} \mathbf{X}^\top \mathbf{r}^b \right).$$

- Problem (4) can be solved with a closed-form update algorithm.

Solution

Proposition 1

The optimal solution of the optimization problem (4) with $u = 1$ *is:*

$$w_i^\star = \begin{cases} -\frac{\mu + q_i}{2}, & i \in \mathcal{A}, \\ 0, & i \notin \mathcal{A}, \end{cases}$$

with

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

and

$$\mathcal{A} = \{i | \mu + q_i < 0\},$$

where $\mathcal{A}$ *can be determined in* $O(\log(N))$ *steps.*

Algorithm SLAIT

Algorithm 2: Specialized Linear Approximation for the Index Tracking problem (SLAIT)

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}$

repeat

Compute $\mathbf{q}_1^{(k)}$

Solve (4) with Proposition 1 and set the optimal solution as $\mathbf{w}^{(k+1)}$

$k \leftarrow k + 1$

until convergence

return $\mathbf{w}^{(k)}$

The Big Picture

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\|\mathbf{w}\|_0 \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

ℓ_0-norm approximation

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\min.} & \mathbf{w}^\top\mathbf{w}+\mathbf{q}_1^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

The Big Picture

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\right\|_2^2 + \lambda\|\mathbf{w}\|_0 \\ \text{s.t.} & \mathbf{w} \in \mathcal{W} \end{array}$$

ℓ_0-norm approximation

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\right\|_2^2 + \lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \text{s.t.} & \mathbf{w} \in \mathcal{W} \end{array}$$

MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\right\|_2^2 + \lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w} \in \mathcal{W} \end{array}$$

MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min.}} & \mathbf{w}^\top\mathbf{w} + \mathbf{q}_1^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w} \in \mathcal{W} \end{array}$$

Outline

Outline

Holding Constraints

- In practice, the constraints that are usually considered in the index tracking problem can be written in a convex form.
- Exception: holding constraints to avoid extreme positions or brokerage fees for very small orders

$$\mathbf{l} \odot \mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}} \leq \mathbf{w} \leq \mathbf{u} \odot \mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}}$$

- Active constraints only for the selected assets ($w_i > 0$).
- Upper bound is easy: $\mathbf{w} \leq \mathbf{u} \odot \mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}} \iff \mathbf{w} \leq \mathbf{u}$ (convex and can be included in $\mathcal{W}$).
- Lower bound is nasty. 😢

Problem Formulation

The problem formulation with holding constraints becomes (after the ℓ_0-"norm" approximation):

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\|_2^2 + \lambda \mathbf{1}^\top \boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathcal{W}, \\ & \mathbf{l} \odot \mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}} \leq \mathbf{w}. \end{array} \tag{5}$$

- How should we deal with the non-convex constraint?

Penalization of Violations

- Hard constraint $\Longrightarrow$ Soft constraint.
- Penalize violations in the objective.
- A suitable penalty function for a general entry w is (since the constraints are separable):

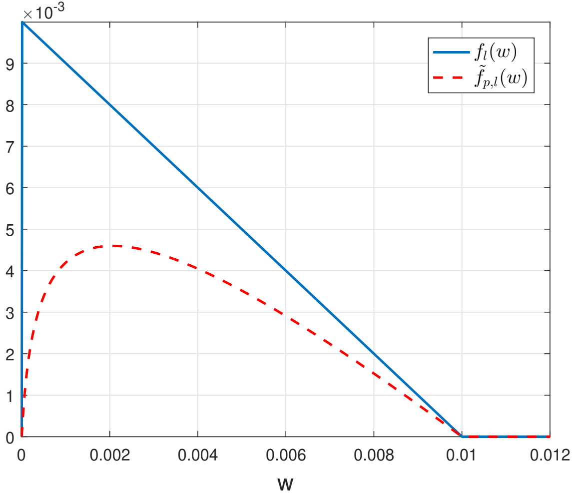
$$f_l(w) = \left(\mathcal{I}_{\{0<w<l\}} \cdot l - w\right)^+ .$$

- Approximate the indicator function with $\rho_{p,\gamma}(w)$. Since we are interested in the interval $[0, l]$ we select $\gamma = l$:

$$\tilde{f}_{p,l}(w) = \left(\rho_{p,l}(w) \cdot l - w\right)^+ .$$

Penalization of Violations

- Penalty functions $f_l(w)$ and $\tilde{f}_{p,l}(w)$ for $l = 0.01, p = 10^{-4}$:



Problem Formulation with Penalty

The penalized optimization problem becomes:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\|_2^2 + \lambda \mathbf{1}^\top \boldsymbol{\rho}_{p,u}(\mathbf{w}) + \boldsymbol{\nu}^\top \tilde{\mathbf{f}}_{p,l}(\mathbf{w}) \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array} \quad (6)$$

- $\boldsymbol{\nu}$ is a parameter vector that controls the penalization.
- $\tilde{\mathbf{f}}_{p,l}(\mathbf{w}) = [\tilde{f}_{p,l}(w_1), \ldots, \tilde{f}_{p,l}(w_N)]^\top$.
- Problem (6) is not convex:
  - $\rho_{p,u}(w)$ is concave $\Longrightarrow$ Linear upperbound with Lemma 1.
  - $\tilde{f}_{p,l}(w)$ is neither convex nor concave. 😢

Outline

Majorization of $\tilde{f}_{p,l}(w)$

Lemma 3

The function $\tilde{f}_{p,l}(w) = (\rho_{p,l}(w) \cdot l - w)^+$ *is majorized at* $w^{(k)} \in [0, u]$ *by the convex function*

$$h_{p,l}(w, w^{(k)}) = \left(\left(d_{p,l}(w^{(k)}) \cdot l - 1\right) w + c_{p,l}(w^{(k)}) \cdot l\right)^+,$$

where $d_{p,l}(w^{(k)})$ *and* $c_{p,l}(w^{(k)})$ *are given in Lemma 1.*

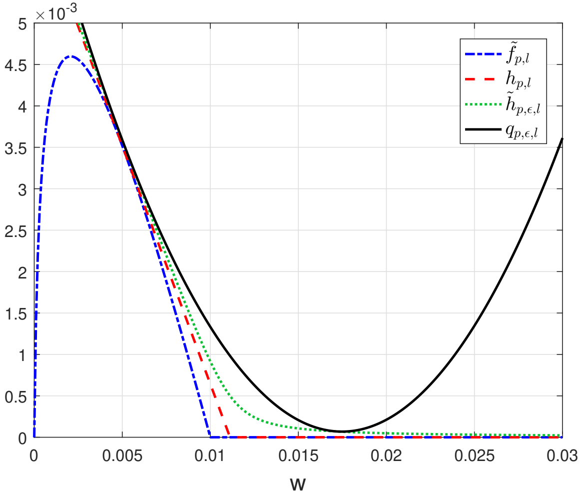
Proof: $\rho_{p,l}(w) \leq d_{p,l}(w^{(k)})w + c_{p,l}(w^{(k)})$ for $w \geq 0$ [Lemma 1].

$$\begin{aligned}
\tilde{f}_{p,l}(w) &= \max\left(\rho_{p,l}(w) \cdot l - w, 0\right) \\
&\leq \max\left((d_{p,l}(w^{(k)})w + c_{p,l}(w^{(k)})) \cdot l - w, 0\right) \\
&= \max\left(\left(d_{p,l}(w^{(k)}) \cdot l - 1\right) w + c_{p,l}(w^{(k)}) \cdot l, 0\right).
\end{aligned}$$

$h_{p,l}(w, w^{(k)})$ is convex as the maximum of two convex functions.

Majorization of $\tilde{f}_{p,l}(w)$

- Observe $\tilde{f}_{p,l}(w)$ and its piecewise linear majorizer $h_{p,l}(w, w^{(k)})$:



Convex Formulation of the Majorization

- Recall our problem:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w})+\boldsymbol{\nu}^\top\tilde{\mathbf{f}}_{p,l}(\mathbf{w})\\ \text{subject to} & \mathbf{w}\in\mathcal{W}.\end{array}$$

- From Lemma 1: $\boldsymbol{\rho}_{p,u}(\mathbf{w})\leq{\mathbf{d}_{p,u}^{(k)}}^\top\mathbf{w}+\mathit{const.}$
- From Lemma 3:

$$\begin{aligned}\tilde{\mathbf{f}}_{p,l}(\mathbf{w})=\left(\boldsymbol{\rho}_{p,l}(\mathbf{w})\cdot\mathbf{l}-\mathbf{w}\right)^+ &\leq\left(\mathrm{Diag}\left(\mathbf{d}_{p,l}^{(k)}\odot\mathbf{l}-\mathbf{1}\right)\mathbf{w}+\mathbf{c}_{p,l}^{(k)}\odot\mathbf{l}\right)^+\\ &=\mathbf{h}_{p,l}(\mathbf{w},\mathbf{w}^{(k)})\end{aligned}$$

- The majorized problem at the $(k+1)$-th iteration becomes:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda{\mathbf{d}_{p,u}^{(k)}}^\top\mathbf{w}+\boldsymbol{\nu}^\top\mathbf{h}_{p,l}(\mathbf{w},\mathbf{w}^{(k)})\\ \text{subject to} & \mathbf{w}\in\mathcal{W}\end{array}\tag{7}$$

- Problem (7) is convex.

Algorithm LAITH

Algorithm 3: Linear Approximation for the Index Tracking problem with Holding constraints (LAITH)

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}$

repeat

  Compute $\mathbf{d}_{p,l}^{(k)}$, $\mathbf{d}_{p,u}^{(k)}$

  Compute $\mathbf{c}_{p,l}^{(k)}$

  Solve (7) with a solver and set the optimal solution as $\mathbf{w}^{(k+1)}$

  $k \leftarrow k + 1$

until convergence

return $\mathbf{w}^{(k)}$

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\|\mathbf{w}\|_0\\ \text{s.t.} & \mathbf{w}\in\mathcal{W},\\ & \mathbf{l}\odot\mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}}\leq\mathbf{w}.\end{array}$$

ℓ_0-norm approximation / soft constraint

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w})\\ & +\boldsymbol{\nu}^\top\tilde{\mathbf{f}}_{p,l}(\mathbf{w})\\ \text{s.t.} & \mathbf{w}\in\mathcal{W}\end{array}$$

MM

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{min}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w}\\ & +\boldsymbol{\nu}^\top\tilde{\mathbf{h}}_{p,l}(\mathbf{w},\mathbf{w}^{(k)})\\ \text{s.t.} & \mathbf{w}\in\mathcal{W}\end{array}$$

Should we stop here?

✓ Again, for specific constraint sets we can derive closed-form update algorithms!

Smooth Approximation of the $(\cdot)^+$ Operator

- To get a closed-form update algorithm we need to majorize again the objective.
- Let us begin with the majorization of the third term, i.e.,

$$\mathbf{h}_{p,l}(\mathbf{w}, \mathbf{w}^{(k)}) = \left(\text{Diag}\left(\mathbf{d}_{p,l}^{(k)} \odot \mathbf{l} - \mathbf{1}\right)\mathbf{w} + \mathbf{c}_{p,l}^{(k)} \odot \mathbf{l}\right)^{+}.$$

✓ Separable: focus only in the univariate case, i.e., $h_{p,l}(w, w^{(k)})$.

× Not smooth: cannot define majorization function at the non-differentiable point.

Smooth Approximation of the $(\cdot)^+$ Operator

- Use a smooth approximation of the $(\cdot)^+$ operator:

$$(x)^+ \approx \frac{x + \sqrt{x^2 + \epsilon^2}}{2},$$

where $0 < \epsilon \ll 1$ controls the approximation.

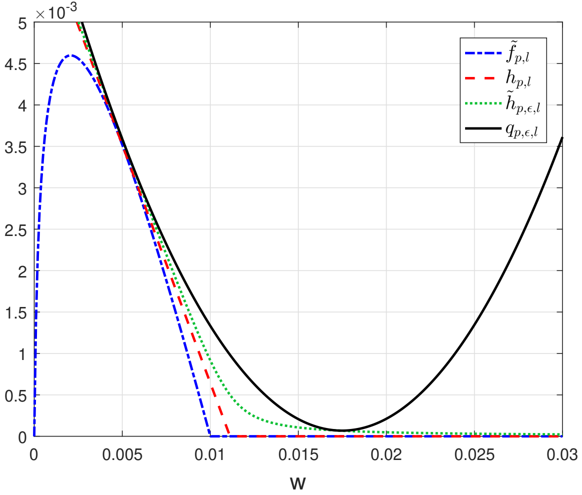
- Apply this to $h_{p,l}(w, w^{(k)}) = \left(\left(d_{p,l}(w^{(k)}) \cdot l - 1 \right) w + c_{p,l}(w^{(k)}) \cdot l \right)^+$:

$$\tilde{h}_{p,\epsilon,l}(w, w^{(k)}) = \frac{\alpha^{(k)} w + \beta^{(k)} + \sqrt{(\alpha^{(k)} w + \beta^{(k)})^2 + \epsilon^2}}{2},$$

where $\alpha^{(k)} = d_{p,l}(w^{(k)}) \cdot l - 1, and \beta^{(k)} = c_{p,l}(w^{(k)}) \cdot l.$

Smooth Majorization of $\tilde{f}_{p,l}(w)$

- Penalty function $\tilde{f}_{p,l}(w)$, its piecewise linear majorizer $h_{p,l}(w, w^{(k)})$, and its smooth approximation $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$:



Quadratic Majorization of $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$

Lemma 4

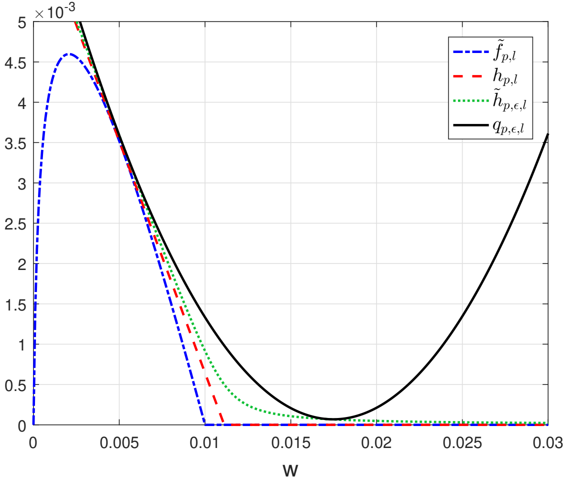
The function $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$ *is majorized at* $w^{(k)}$ *by the quadratic convex function*

$$q_{p,\epsilon,l}(w, w^{(k)}) = a_{p,\epsilon,l}(w^{(k)})w^2 + b_{p,\epsilon,l}(w^{(k)})w + c_{p,\epsilon,l}(w^{(k)}),$$

where $a_{p,\epsilon,l}(w^{(k)}) = \frac{(\alpha^{(k)})^2}{2\kappa}$, $b_{p,\epsilon,l}(w^{(k)}) = \frac{\alpha^{(k)}\beta^{(k)}}{\kappa} + \frac{\alpha^{(k)}}{2}$, *and* $c_{p,\epsilon,l}(w^{(k)}) = \frac{(\alpha^{(k)}w^{(k)})(\alpha^{(k)}w^{(k)}+2\beta^{(k)})+2({\beta^{(k)}}^2+\epsilon^2)}{2\kappa} + \frac{\beta^{(k)}}{2}$ *is an optimization irrelevant constant, with* $\kappa = 2\sqrt{(\alpha^{(k)}w^{(k)} + \beta^{(k)})^2 + \epsilon^2}$.

Proof: Majorize the square root term of $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$ (concave) with its first-order Taylor approximation.

- Penalty function $\tilde{f}_{p,l}(w)$, its piecewise linear majorizer $h_{p,l}(w, w^{(k)})$, its smooth majorizer $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$, and its quadratic majorizer $q_{p,\epsilon,l}(w, w^{(k)})$:



Quadratic Formulation of the Majorization

- Recall our problem:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w}+\boldsymbol{\nu}^\top\tilde{\mathbf{h}}_{p,\epsilon,l}(\mathbf{w},\mathbf{w}^{(k)})\\ \text{subject to} & \mathbf{w}\in\mathcal{W}.\end{array}$$

- From Lemma 4:

$$\tilde{\mathbf{h}}_{p,\epsilon,l}(\mathbf{w},\mathbf{w}^{(k)})\leq\mathbf{w}^\top\mathrm{Diag}\left(\mathbf{a}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}\right)\mathbf{w}+\mathbf{b}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}^\top\mathbf{w}+const.$$

- The majorized problem at the $(k+1)$-th iteration becomes:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^\top\left(\frac{1}{T}\mathbf{X}^\top\mathbf{X}+\mathrm{Diag}\left(\mathbf{a}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}\right)\right)\mathbf{w}\\ & +\left(\lambda\mathbf{d}_{p,u}^{(k)}-\frac{2}{T}\mathbf{X}^\top\mathbf{r}^b+\mathbf{b}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}\right)^\top\mathbf{w}\\ \text{subject to} & \mathbf{w}\in\mathcal{W}\end{array}\qquad(8)$$

Quadratic Formulation of the Majorization

- Problem (8) is a QP that can be solved with a solver, but we can do better.
- Use Lemma 2 to majorize the quadratic part:
  - $\mathbf{L}_2 = \frac{1}{T}\mathbf{X}^\top\mathbf{X} + \text{Diag}\left(\mathbf{a}_{p,\epsilon,l}^{(k)} \odot \boldsymbol{\nu}\right)$
  - $\mathbf{M}_2 = \lambda_{\max}^{(\mathbf{L}_2)}\mathbf{I}$.
- And the final optimization problem at the $(k+1)$-th iteration becomes:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^\top\mathbf{w} + \mathbf{q}_2^{(k)\top}\mathbf{w} \\ \text{subject to} & \mathbf{w} \in \mathcal{W}, \end{array} \tag{9}$$

where

$$\mathbf{q}_2^{(k)} = \frac{1}{\lambda_{\max}^{(\mathbf{L}_2)}}\left(2\left(\mathbf{L}_2 - \lambda_{\max}^{(\mathbf{L}_2)}\mathbf{I}\right)\mathbf{w}^{(k)} + \lambda\mathbf{d}_{p,u}^{(k)} - \frac{2}{T}\mathbf{X}^\top\mathbf{r}^b + \mathbf{b}_{p,\epsilon,l}^{(k)} \odot \boldsymbol{\nu}\right).$$

- Problem (9) can be solved in closed form!

Algorithm SLAITH

Algorithm 4: Specialized Linear Approximation for the Index Tracking problem with Holding constraints (SLAITH)

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}$

repeat

  Compute $\mathbf{q}_2^{(k)}$

  Solve (9) with Proposition 1 and set the optimal solution as $\mathbf{w}^{(k+1)}$

  $k \leftarrow k + 1$

until convergence

return $\mathbf{w}^{(k)}$

The Big Picture

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\|\mathbf{w}\|_0 \\ \text{s.t.} & \mathbf{w}\in\mathcal{W}, \\ & \mathbf{l}\odot\mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}}\leq\mathbf{w}. \end{array}$$

↓ ℓ_0-norm approximation / soft constraint

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{1}^\top\boldsymbol{\rho}_{p,u}(\mathbf{w}) \\ & +\boldsymbol{\nu}^\top\tilde{\mathbf{f}}_{p,l}(\mathbf{w}) \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

→ MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\min} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w}-\mathbf{r}^b\right\|_2^2+\lambda\mathbf{d}_{p,u}^{(k)\top}\mathbf{w} \\ & +\boldsymbol{\nu}^\top\tilde{\mathbf{h}}_{p,l}(\mathbf{w},\mathbf{w}^{(k)}) \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

↑ smooth approx. / MM

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{min.}} & \mathbf{w}^\top\mathbf{w}+\mathbf{q}_2^{(k)\top}\mathbf{w} \\ \text{s.t.} & \mathbf{w}\in\mathcal{W} \end{array}$$

Outline

Extension to Other Tracking Error Measures

In all the previous formulations we used the empirical tracking error (ETE):

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

However, we can use other tracking error measures such as:[5]

- Downside risk:
  $$\mathrm{DR}(\mathbf{w}) = \frac{1}{T}\|(\mathbf{r}^b - \mathbf{X}\mathbf{w})^+\|_2^2,$$
  where $(x)^+ = \max(0, x)$.
- Value-at-Risk (VaR) relative to an index.
- Conditional VaR (CVaR) relative to an index.

[5] K. Benidis, Y. Feng, and D. P. Palomar, *Optimization Methods for Financial Index Tracking: From Theory to Practice*. Foundations and Trends in Optimization, Now Publishers, 2018.

Extension to Downside Risk

- DR($\mathbf{w}$) is convex: can be used directly without any manipulation.
- Interestingly, specialized algorithms can be derived for DR too by properly majorizing it.

Lemma 5

The function $DR(\mathbf{w}) = \frac{1}{T}\left\|(\mathbf{r}^b - \mathbf{X}\mathbf{w})^+\right\|_2^2$ *is majorized at* $\mathbf{w}^{(k)}$ *by the quadratic convex function* $\frac{1}{T}\left\|\mathbf{r}^b - \mathbf{X}\mathbf{w} - \mathbf{y}^{(k)}\right\|_2^2$, *where*

$$\mathbf{y}^{(k)} = -\left(\mathbf{X}\mathbf{w}^{(k)} - \mathbf{r}^b\right)^+.$$

Proof of Lemma 5 (1/4)

For convenience set $\mathbf{z} = \mathbf{r}^b - \mathbf{X}\mathbf{w}$. Then:

$$\mathrm{DR}(\mathbf{w}) = \frac{1}{T}\|(\mathbf{z})^+\|_2^2 = \frac{1}{T}\sum_{i=1}^{T}\tilde{z}_i^2,$$

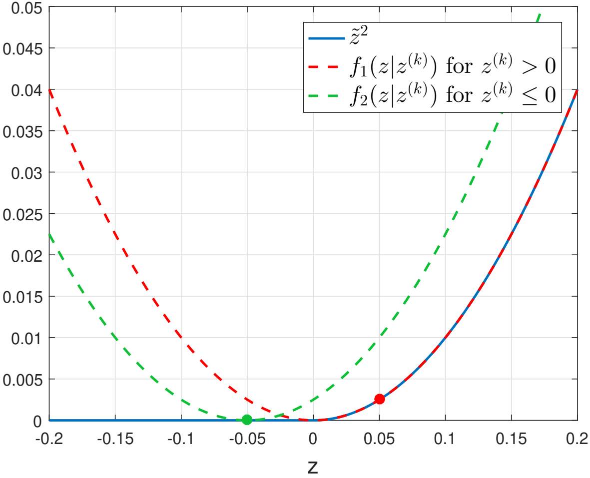
where

$$\tilde{z}_i = \begin{cases} z_i, & \text{if } z_i > 0, \\ 0, & \text{if } z_i \leq 0. \end{cases}$$

- Majorize each $\tilde{z}_i^2$. Two cases:
  - For a point $z_i^{(k)} > 0$, $f_1(z_i|z_i^{(k)}) = z_i^2$ is an upper bound of $\tilde{z}_i^2$, with $f_1(z_i^{(k)}|z_i^{(k)}) = \left(z_i^{(k)}\right)^2 = \left(\tilde{z}_i^{(k)}\right)^2$.
  - For a point $z_i^{(k)} \leq 0$, $f_2(z_i|z_i^{(k)}) = \left(z_i - z_i^{(k)}\right)^2$ is an upper bound of $\tilde{z}_i^2$, with $f_2(z_i^{(k)}|z_i^{(k)}) = \left(z_i^{(k)} - z_i^{(k)}\right)^2 = 0 = \left(\tilde{z}_i^{(k)}\right)^2$.

Proof of Lemma 5 (2/4)

For both cases the proofs are straightforward and they are easily shown pictorially:



Proof of Lemma 5 (3/4)

Combining the two cases:

$$
\begin{aligned}
\tilde{z}_i^2 \leq & \begin{cases} f_1(z_i|z_i^{(k)}), & \text{if } z_i^{(k)} > 0, \\ f_2(z_i|z_i^{(k)}), & \text{if } z_i^{(k)} \leq 0, \end{cases} \\
= & \begin{cases} (z_i - 0)^2, & \text{if } z_i^{(k)} > 0, \\ (z_i - z_i^{(k)})^2, & \text{if } z_i^{(k)} \leq 0, \end{cases} \\
= & (z_i - y_i^{(k)})^2,
\end{aligned}
$$

where

$$
\begin{aligned}
y_i^{(k)} = & \begin{cases} 0, & \text{if } z_i^{(k)} > 0, \\ z_i^{(k)}, & \text{if } z_i^{(k)} \leq 0, \end{cases} \\
= & -(-z_i^{(k)})^+.
\end{aligned}
$$

Proof of Lemma 5 (4/4)

Thus, $\mathrm{DR}(\mathbf{z})$ is majorized as follows:

$$\mathrm{DR}(\mathbf{w}) = \frac{1}{T}\sum_{i=1}^{T}\tilde{z}_i^2 \leq \frac{1}{T}\sum_{i=1}^{T}(z_i - y_i^{(k)})^2 = \frac{1}{T}\|\mathbf{z} - \mathbf{y}^{(k)}\|_2^2.$$

Substituting back $\mathbf{z} = \mathbf{r}^b - \mathbf{X}\mathbf{w}$, we get

$$\mathrm{DR}(\mathbf{w}) \leq \frac{1}{T}\|\mathbf{r}^b - \mathbf{X}\mathbf{w} - \mathbf{y}^{(k)}\|_2^2,$$

where $\mathbf{y}^{(k)} = -(-\mathbf{z}^{(k)})^+ = -(\mathbf{X}\mathbf{w} - \mathbf{r}^b)^+$.

Extension to Other Penalty Functions

- Apart from the various performance measures, we can select a different penalty function.
- We have used only the ℓ_2-norm to penalize the differences between the portfolio and the index.
- We can use the Huber penalty function for robustness against outliers:[6]

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

- The ℓ_1-norm.
- Many more...

[6] K. Benidis, Y. Feng, and D. P. Palomar, *Optimization Methods for Financial Index Tracking: From Theory to Practice*. Foundations and Trends in Optimization, Now Publishers, 2018.

Extension to Huber Penalty Function

Lemma 6

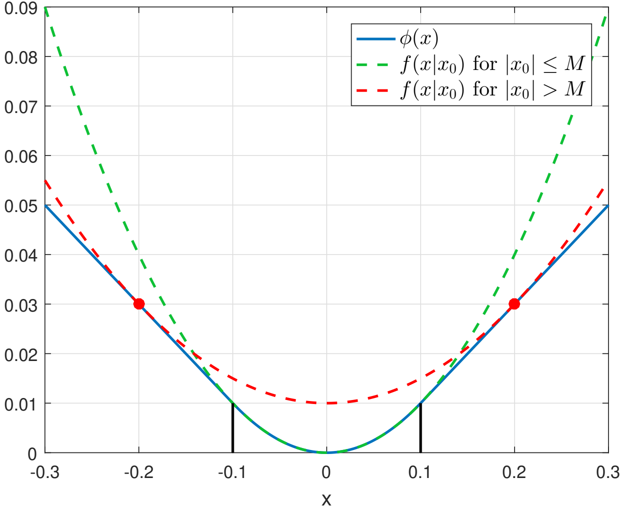
The function $\phi(x)$ is majorized at $x^{(k)}$ by the quadratic convex function $f(x|x^{(k)}) = a^{(k)}x^2 + b^{(k)}$, where

$$a^{(k)} = \begin{cases} 1, & |x^{(k)}| \leq M, \\ \frac{M}{|x^{(k)}|}, & |x^{(k)}| > M, \end{cases}$$

and

$$b^{(k)} = \begin{cases} 0, & |x^{(k)}| \leq M, \\ M(|x^{(k)}| - M), & |x^{(k)}| > M. \end{cases}$$

Extension to Huber Penalty Function



Outline

Set Up

For the numerical experiments we use historical data of two indices.

Table 1: Index Information

Index	Data Period	T_{trn}	Ttst
S&P 500	01/01/10 - 31/12/15	252	252
Russell 2000	01/06/06 - 31/12/15	1000	252

- We use a rolling window approach.
- Performance measure: magnitude of daily tracking error (MDTE)

$$\text{MDTE} = \frac{1}{T - T_{\text{tr}}} \left\| \text{diag}(\mathbf{XW}) - \mathbf{r}^b \right\|_2,$$

where $\mathbf{X} \in \mathbb{R}^{(T-T_{\text{tr}}) \times N}$ and $\mathbf{r}^b \in \mathbb{R}^{T-T_{\text{tr}}}$.

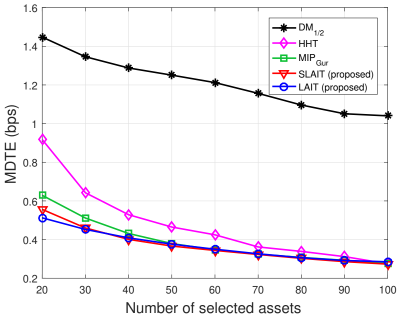
Benchmarks

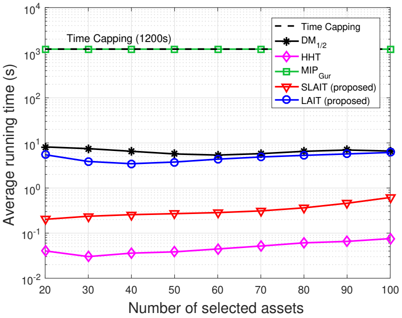
- MIP solution by Gurobi solver (MIP_{Gur}).
- Diversity Method[7] where the $\ell_{1/2}$-“norm” approximation is used ($\text{DM}_{1/2}$).
- Hybrid Half Thresholding (HHT) algorithm[8].

[7] R. Jansen and R. Van Dijk, “Optimal benchmark tracking with small portfolios,” *The Journal of Portfolio Management*, vol. 28, no. 2, pp. 33–39, 2002.

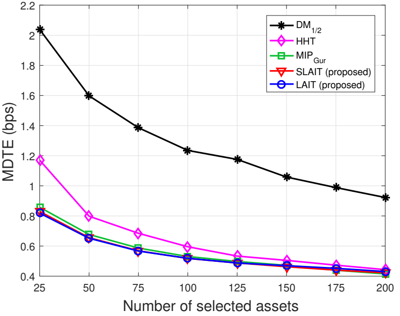
[8] F. Xu, Z. Xu, and H. Xue, “Sparse index tracking based on $L_{1/2}$ model and algorithm,” *arXiv preprint*, 2015.

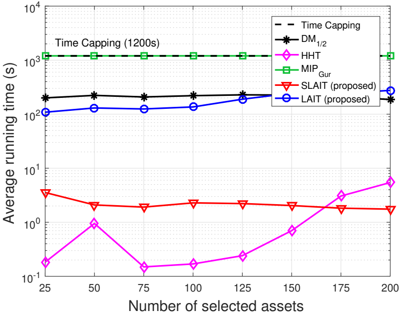
S&P 500 - w/o Holding Constraints



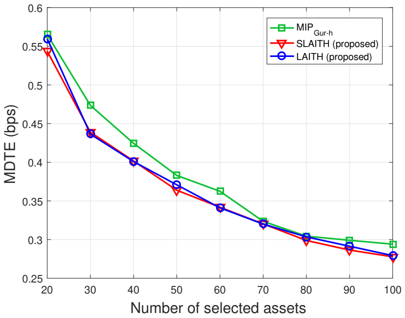


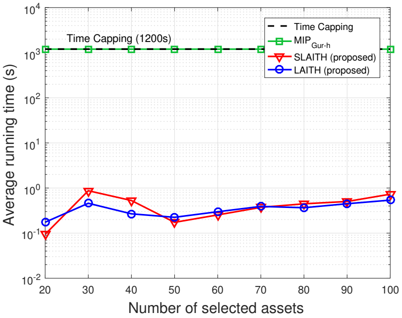
Russell 2000 - w/o Holding Constraints



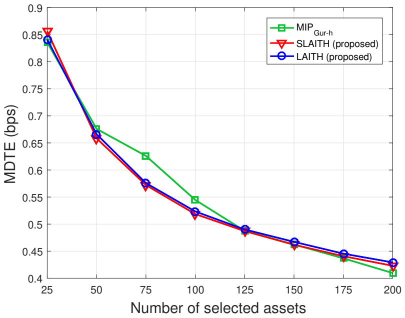


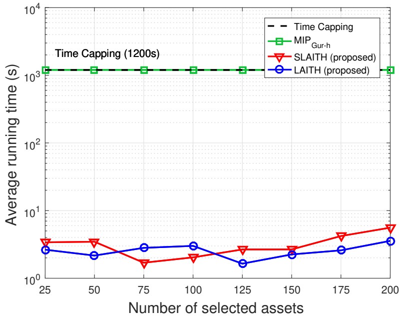
S&P 500 - w/ Holding Constraints



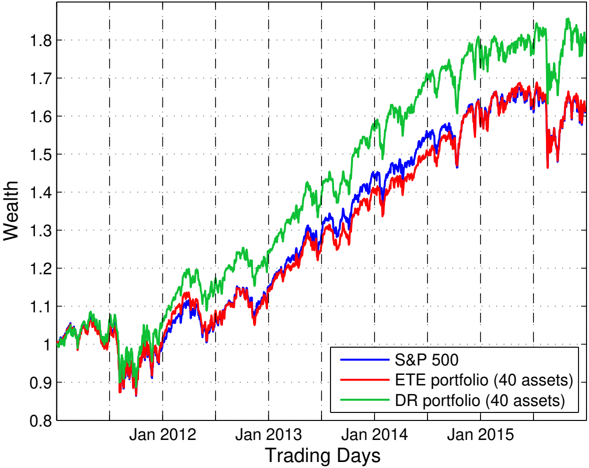


Russell 2000 - w/ Holding Constraints





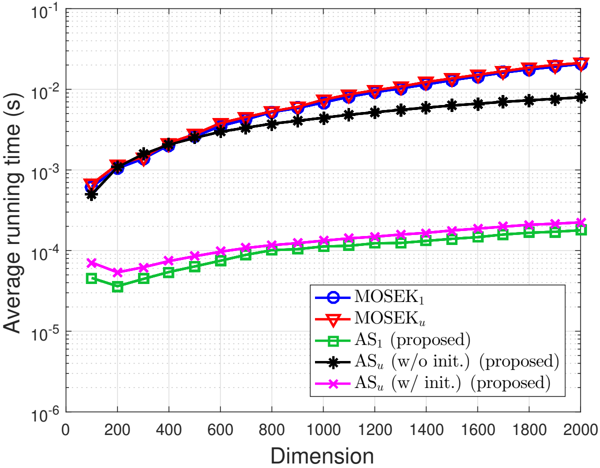
Tracking the S&P 500 index



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

Average Running Time of Proposed Methods

- Comparison of AS_1 and AS_u.[10]



[10]The algorithms MOSEK_1 and MOSEK_u correspond to the solution using the MOSEK solver.

Outline

Conclusions

- We have developed efficient algorithms that promote sparsity for the index tracking problem.
- The algorithms are derived based on the MM framework:
  - Derivation of surrogate functions
  - Majorization of convex problems for closed-form solutions.
- Many possible extensions.
- Same techniques can be used for active portfolio management.
- More generally: if you know how to solve a problem, then inducing sparsity should be a piece of cake!

Thanks

For more information visit:

https://www.danielppalomar.com