Index Tracking in Finance

Konstantinos Benidis and Daniel P. Palomar

The Hong Kong University of Science and Technology (HKUST)

ELEC 5470 - Convex Optimization
Fall 2017-18, HKUST, Hong Kong

Outline

Outline

Investment Strategies

Fund managers follow two basic investment strategies:

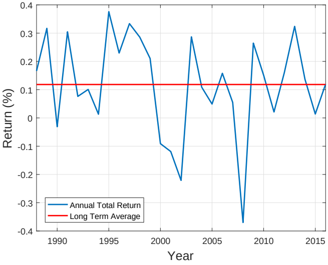
Active

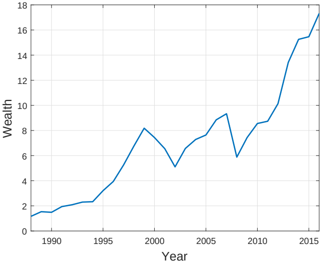
- Assumption: markets are not perfectly efficient.
- Through expertise add value by choosing high performing assets.

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).

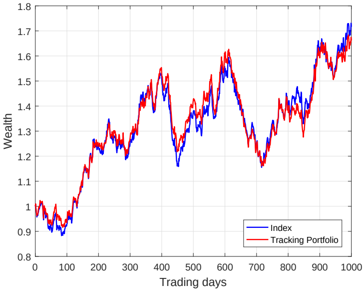
The stock markets have historically risen, e.g. S&P 500.





- 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.

Definitions

- Price (at time t) of an asset or an index: p_t
- Net return: $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$
- Returns of an index in a period of T days: $\mathbf{r}^b = [r_1^b, \dots, r_T^b]^\top \in \mathbb{R}^T$
- Returns of N assets in a period of T days: $\mathbf{X} = [\mathbf{r}_1, \dots, \mathbf{r}_T]^\top \in \mathbb{R}^{T \times N}$
  - $\mathbf{r}_t \in \mathbb{R}^N$ with $t \in [1, T]$

Definitions

- Assume that an index is composed by a weighted collection of N assets.
- $\mathbf{b} \in \mathbb{R}_+^N$: normalized index weights
  - $\mathbf{b} > \mathbf{0}$
  - $\mathbf{b}^T\mathbf{1} = 1$
  - $\mathbf{X}\mathbf{b} = \mathbf{r}^b$
- A portfolio $\mathbf{w} \in \mathbb{R}_+^N$ is defined to be the proportion of the money we allocate in each asset.
- $\mathbf{w} \in \mathbb{R}_+^N$: tracking portfolio we wish to design
  - $\mathbf{w} \geq \mathbf{0}$
  - $\mathbf{w}^T\mathbf{1} = 1$

Full Replication

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

$\Longrightarrow$ Straightforward solution: Full replication.

Full Replication

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

$\Longrightarrow$ Straightforward solution: Full replication.

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

Full Replication

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

  $\Longrightarrow$ Straightforward solution: Full replication.

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

- Limitations:
  - Transaction costs: increase as $\text{card}(\mathbf{w})$ increases.
  - Illiquid assets: cannot buy/sell an asset easily or in market price.

Sparse Index Tracking

- How can we overcome there limitations?

  $\Longrightarrow$ Sparse index tracking.

Sparse Index Tracking

- How can we overcome there limitations?

  $\Longrightarrow$ Sparse index tracking.

- Use a small number of assets: $\mathsf{card}(\mathbf{w}) < N$
  - Reduce transaction costs.
  - Avoid illiquid assets.
  - Tradeoff: imperfect tracking.

Sparse Index Tracking

- How can we overcome there limitations?

  $\Longrightarrow$ Sparse index tracking.

- Use a small number of assets: card$(\mathbf{w}) < N$
  - Reduce transaction costs.
  - Avoid illiquid assets.
  - Tradeoff: imperfect tracking.

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

Existing Methods

- Two step approach:
  - Stock selection:
    - Largest market capital.
    - Most correlated to the index.
    - A combination cointegrated well with the index.
  - Capital allocation:
    - Naive allocation: proportional to the original weights.
    - Optimized allocation: usually a convex problem.
- Mixed Integer Programming
  - 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.

Outline

Tracking Error Measure

A common tracking measure is the **empirical tracking error (ETE)**:

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

- $\text{ETE}(\mathbf{w}) = \begin{cases} 0, & \text{if } \mathbf{w} = \mathbf{b}, \\ +, & \text{otherwise.} \end{cases}$

Goal: Construct a sparse tracking portfolio with minimum ETE.

Problem Formulation

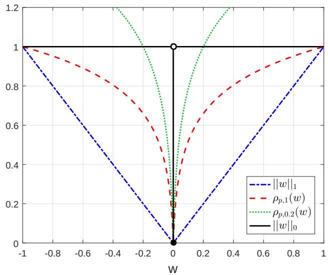
$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{1}{T}\left\|\mathbf{X}\mathbf{w} - \mathbf{r}^b\right\|_2^2 + \lambda\|\mathbf{w}\|_0 \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array} \tag{1}$$

- $\mathcal{W}$: A set of general **convex** constraints.
  - We assume that $\{\mathbf{w}|\mathbf{w} \geq \mathbf{0}, \mathbf{w}^\top\mathbf{1} = 1\} \subseteq \mathcal{W}$.
  - We will state separately any non-convex constraints.
- The optimization problem (1) is too difficult to deal with directly:
  - Discontinuous, non-differentiable, non-convex objective function.

- 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 $[-\gamma, \gamma]$.



- For our problem we set $\gamma = u$, where $u \leq 1$ is an upperbound of the weights.
  - If $\{\mathbf{w} | \mathbf{w} \leq u\mathbf{1}\} \nsubseteq \mathcal{W}$ (no upperbound constraint) then implicitly $u = 1$.

Approximate Formulation

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

- $\boldsymbol{\rho}_{p,u}(\mathbf{w}) = [\rho_{p,u}(w_1), \ldots, \rho_{p,u}(w_N)]^\top$.
- Problem (2) is continuous and differentiable for $\mathbf{w} \geq \mathbf{0}$.
- 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 $w_0 \in \mathbb{R}_+$.

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

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.

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

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

The Big Picture

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

ℓ_0-norm approximation

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

The Big Picture

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

ℓ_0-norm approximation

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

MM

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

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!

- We consider the following convex set parametrized by u:

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

- Expand the objective:

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

- The problem becomes:

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

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:

$$\mathbf{w}^\top \mathbf{L} \mathbf{w} \le \mathbf{w}^\top \mathbf{M} \mathbf{w} + 2\mathbf{w}^{(k)\top} (\mathbf{L} - \mathbf{M})\, \mathbf{w} - \mathbf{w}^{(k)\top} (\mathbf{L} - \mathbf{M})\, \mathbf{w}^{(k)}.$$

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

Proof:

$$\begin{aligned}
\mathbf{w}^\top \mathbf{L} \mathbf{w} &= \mathbf{w}^{(k)\top} \mathbf{L} \mathbf{w}^{(k)} + 2\left(\mathbf{w} - \mathbf{w}^{(k)}\right)^\top \mathbf{L} \mathbf{w}^{(k)} + \left(\mathbf{w} - \mathbf{w}^{(k)}\right)^\top \mathbf{L} \left(\mathbf{w} - \mathbf{w}^{(k)}\right) \\
&\le \mathbf{w}^{(k)\top} \mathbf{L} \mathbf{w}^{(k)} + 2\left(\mathbf{w} - \mathbf{w}^{(k)}\right)^\top \mathbf{L} \mathbf{w}^{(k)} + \left(\mathbf{w} - \mathbf{w}^{(k)}\right)^\top \mathbf{M} \left(\mathbf{w} - \mathbf{w}^{(k)}\right) \\
&= \mathbf{w}^\top \mathbf{M} \mathbf{w} + 2\mathbf{w}^{(k)\top} (\mathbf{L} - \mathbf{M})\, \mathbf{w} - \mathbf{w}^{(k)\top} (\mathbf{L} - \mathbf{M})\, \mathbf{w}^{(k)}
\end{aligned}$$

- 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:

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

Specialized Iterative Formulation

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

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

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.
- Consider two cases of $\mathcal{W}_u$:
  1. $u = 1$ (we can discard the upper bound - easier KKT)
  2. $u < 1$

AS_1: Active Set for $u = 1$

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

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

and

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

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

AS$_u$: Active Set for $u < 1$

Proposition 2

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

$$w_i^\star = \begin{cases} u, & i \in \mathcal{B}_1, \\ -\frac{\mu+q_i}{2}, & i \in \mathcal{B}_2, \\ 0, & i \notin \mathcal{B}_1 \cup \mathcal{B}_2, \end{cases}$$

with

$$\mu = -\frac{\sum_{i\in\mathcal{B}_2} q_i + 2 - \textit{card}(\mathcal{B}_1)2u}{\textit{card}(\mathcal{B}_2)},$$

$$\mathcal{B}_1 = \{i | \mu + q_i \leq -2u\},$$

$$\mathcal{B}_2 = \{i | -2u < \mu + q_i < 0\},$$

where $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *can be determined in* $O(N\log(N))$ *steps.*

SLAIT

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

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}_u$
Compute $\lambda_{\text{max}}^{(\mathbf{L}_1)}$
repeat
  Compute $\mathbf{q}_1^{(k)}$
  Solve (4) with AS_1 or AS_u and set the optimal solution as $\mathbf{w}^{(k+1)}$
  $k \leftarrow k + 1$
until convergence
return $\mathbf{w}^{(k)}$

The Big Picture

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

ℓ_0-norm approximation

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

MM

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

The Big Picture

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

ℓ_0-norm approximation

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

MM

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

MM (for the set $\mathcal{W}_u$)

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^\top\mathbf{w} + \mathbf{q}_1^{(k)\top}\mathbf{w} \\ \text{subject to} & \mathbf{w} \in \mathcal{W}_u \end{array}$$

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 ($w_i > 0$).
- $\mathbf{w} \leq \mathbf{u} \odot \mathcal{I}_{\{\mathbf{w}>\mathbf{0}\}} \iff \mathbf{w} \leq \mathbf{u}$ (convex).
  - Can be included in $\mathcal{W}$.

Problem Formulation

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

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

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

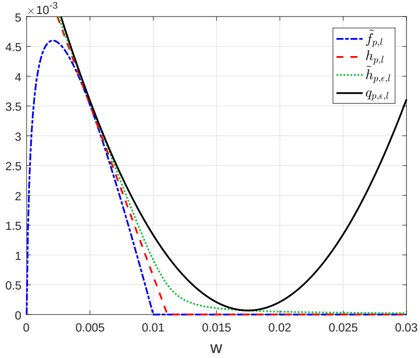
Penalization of Violations

- 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 for the interval $[0, l]$ we select $\gamma = l$.

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



- Penalty function $\tilde{f}_{p,l}(w)$ for $l = 0.01$, $p = 10^{-4}$.
- $h_{p,l}(w)$, $\tilde{h}_{p,\epsilon,l}(w)$, $q_{p,\epsilon,l}(w)$?

The optimization problem becomes:

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

- $\boldsymbol{\nu}$ 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 [Lemma 1].
  - $\tilde{f}_{p,l}(w)$ is neither convex nor concave.

Lemma 3

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

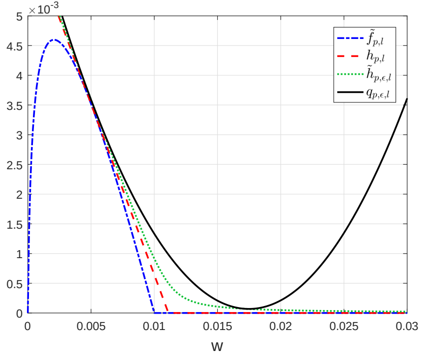
$$h_{p,l}(w, w^{(k)}) = \left(\left(d_{p,l}(w^{(k)}) \cdot l - 1\right) w + c_{p,l}(w^{(k)}) \cdot l\right)^{+},$$

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]. Thus:

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

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



- Penalty function $\tilde{f}_{p,l}(w)$ for $l = 0.01$, $p = 10^{-4}$.
- $h_{p,l}(w)$: linear upperbound of $\tilde{f}_{p,l}(w)$.
- $\tilde{h}_{p,\epsilon,l}(w)$, $q_{p,\epsilon,l}(w)$?

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

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

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

- $\boldsymbol{\rho}_{p,u}(\mathbf{w})\leq\mathbf{d}_{p,u}^{(k)\top}\mathbf{w}+const.$ [Lemma 1]
- $\tilde{\mathbf{f}}_{p,l}(\mathbf{w})=(\boldsymbol{\rho}_{p,l}(\mathbf{w})\cdot\mathbf{l}-\mathbf{w})^+\leq\left(\text{Diag}\left(\mathbf{d}_{p,l}^{(k)}\odot\mathbf{l}-\mathbf{1}\right)\mathbf{w}+\mathbf{c}_{p,l}^{(k)}\odot\mathbf{l}\right)^+ = \mathbf{h}_{p,l}(\mathbf{w},\mathbf{w}^{(k)})$ [Lemma 3]
- Problem (7) is convex.

LAITH

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

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

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

The Big Picture

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

The Big Picture

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

ℓ_0-norm approximation / soft constraint

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

The Big Picture

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

ℓ_0-norm approximation / soft constraint

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

MM

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

Should we stop here?

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

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

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

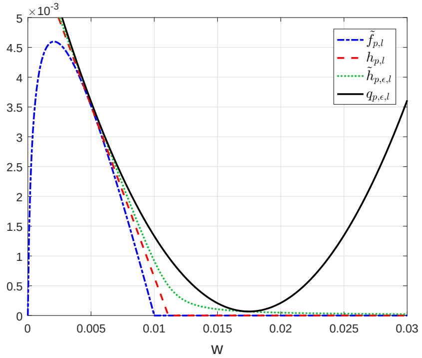
$$(x)^+ \approx \frac{x + \sqrt{x^2 + \epsilon^2}}{2},$$

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 $\beta^{(k)} = c_{p,l}(w^{(k)}) \cdot l$.



- Penalty function $\tilde{f}_{p,l}(w)$ for $l = 0.01$, $p = 10^{-4}$.
- $h_{p,l}(w)$: linear upperbound of $\tilde{f}_{p,l}(w)$.
- $\tilde{h}_{p,\epsilon,l}(w)$: smooth approximation of $h_{p,l}(w)$ for $\epsilon = 10^{-3}$.
- $q_{p,\epsilon,l}(w)$?

- Now that the function is smooth we can derive a majorizer.

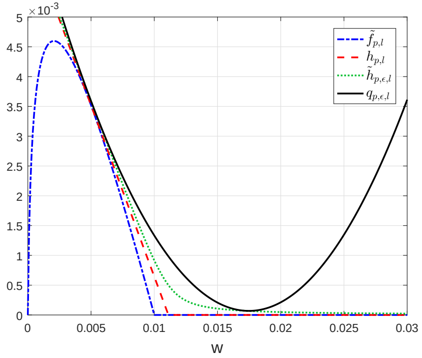
Lemma 4

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*

$$\begin{aligned} a_{p,\epsilon,l}(w^{(k)}) &= \frac{(\alpha^{(k)})^2}{2\kappa}, \\ b_{p,\epsilon,l}(w^{(k)}) &= \frac{\alpha^{(k)}\beta^{(k)}}{\kappa} + \frac{\alpha^{(k)}}{2}, \end{aligned}$$

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 $\tilde{f}_{p,l}(w)$ for $l = 0.01$, $p = 10^{-4}$.
- $h_{p,l}(w)$: linear upperbound of $\tilde{f}_{p,l}(w)$.
- $\tilde{h}_{p,\epsilon,l}(w)$: smooth approximation of $h_{p,l}(w)$ for $\epsilon = 10^{-3}$.
- $q_{p,\epsilon,l}(w)$: quadratic majorizer of $\tilde{h}_{p,\epsilon,l}(w)$.

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

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

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

- $\tilde{\mathbf{h}}_{p,\epsilon,l}(\mathbf{w})\leq\mathbf{w}^\top\text{Diag}\left(\mathbf{a}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}\right)\mathbf{w}+\mathbf{b}_{p,\epsilon,l}^{(k)}\odot\boldsymbol{\nu}^\top\mathbf{w}+const.$ [Lemma 4]

- Problem (8) is a QP.
- Use Lemma 2 to majorize the quadratic part:
  - $\mathbf{L}_2 = \frac{1}{T}\mathbf{X}^\top\mathbf{X} + \mathrm{Diag}\left(\mathbf{a}_{p,\epsilon,l}^{(k)} \odot \boldsymbol{\nu}\right)$.
  - $\mathbf{M}_2 = \lambda_{\max}^{(\mathbf{L}_2)}\mathbf{I}$.

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

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

where

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

SLAITH

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

Set $k = 0$, choose $\mathbf{w}^{(0)} \in \mathcal{W}_u$
Compute $\lambda_{\text{max}}^{(\mathbf{L}_2)}$
repeat
  Compute $\mathbf{q}_2^{(k)}$
  Solve (9) with AS$_1$ or AS$_u$ and set the optimal solution as $\mathbf{w}^{(k+1)}$
  $k \leftarrow k + 1$
until convergence
return $\mathbf{w}^{(k)}$

The Big Picture

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

ℓ_0-norm approximation / soft constraint

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

MM

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

The Big Picture

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

ℓ_0-norm approximation / soft constraint

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

MM

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

MM (for the set $\mathcal{W}_u$) / smooth approximation

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

Extension to Other Tracking Error Measures

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

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

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

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

Extension to Downside Risk

- $\mathrm{DR}(\mathbf{w})$ is convex: can be used directly without any manipulation.
- Interestingly, if we consider the set $\mathcal{W}_u$, specialized algorithms can be derived for the DR too.

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:

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

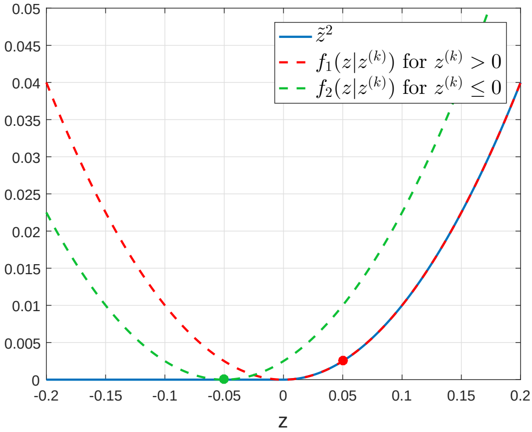
where

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

- 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:

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

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

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

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

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 in the portfolio and the index.
- We can use the Huber penalty function for robustness:

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

- The ℓ_1-norm.
- Many more...

Huber

Lemma 6

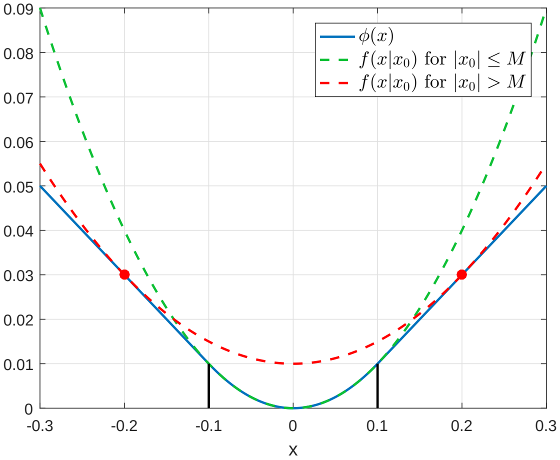
The function $\phi(x)$ *is majorized at* $x^{(k)}$ *by the quadratic convex function* $f(x|x^{(k)}) = a^{(k)}x^2 + b^{(k)}$, *where*

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

and

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

Huber


$\phi(x)$
$f(x|x_0)$ for $|x_0| \leq M$
$f(x|x_0)$ for $|x_0| > M$
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
x

Outline

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

- S&P 500
- Russell 2000

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

Table 1: Index information.

- Rolling window.
- Performance measure: magnitude of daily tracking error (MDTE)

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

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 [Jansen et al., 2002] where the $\ell_{1/2}$-norm approximation is used ($\text{DM}_{1/2}$).
- Hybrid Half Thresholding (HHT) algorithm [F. Xu et al., 2015].

S&P 500 - w/o holding constraints

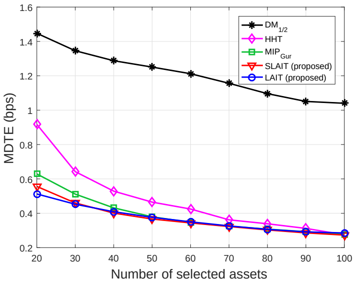


Figure 1: Magnitude of daily tracking error.

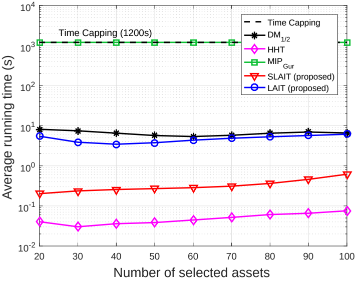


Figure 2: Average running time.

Russell 2000 - w/o holding constraints

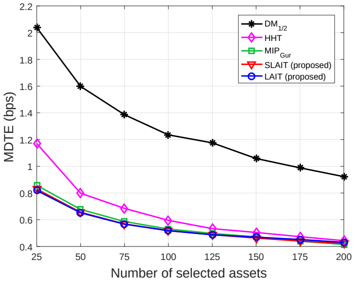


Figure 3: Magnitude of daily tracking error.

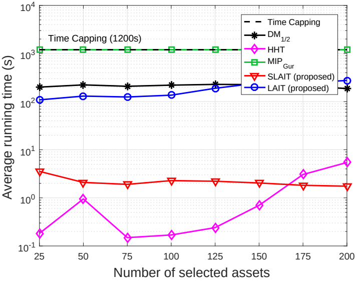


Figure 4: Average running time.

S&P 500 - w/ holding constraints

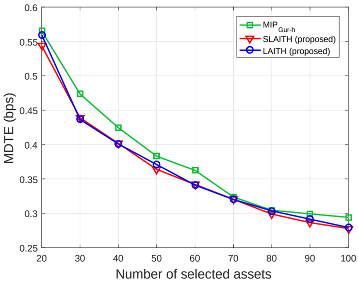


Figure 5: Magnitude of daily tracking error.

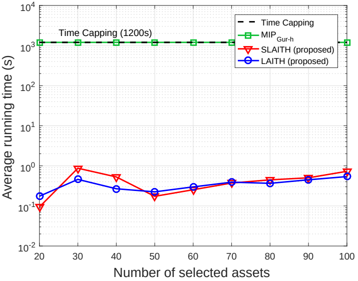


Figure 6: Average running time.

Russell 2000 - w/ holding constraints

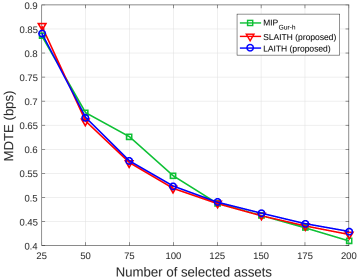


Figure 7: Magnitude of daily tracking error.

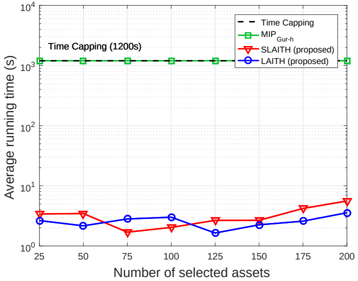


Figure 8: Average running time.

Wealth

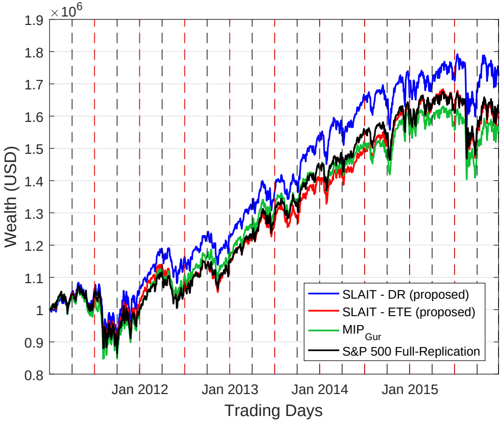


Figure 9: S&P 500 and out-of-sample tracking portfolio wealth with $\text{card}(\mathbf{w}) = 40$ (apart from the full-replication portfolio). Vertical black dashed lines: redesign. Vertical red dashed lines: rebalancing.

Average Running Time of AS_1 - AS_u

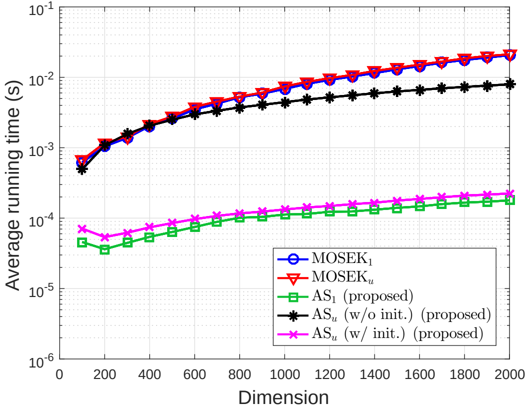


Figure 10: Comparison of AS_1 and AS_u. The algorithms $MOSEK_1$ and $MOSEK_u$ correspond to the solution using the MOSEK solver.

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.

References

Benidis, K., Feng, Y., and Palomar, D. P. (2017).
Sparse Portfolios for High-Dimensional Financial Index Tracking.
IEEE Transactions on Signal Processing (accepted).

Xu, F., Xu, Z., and Xue, H., (2015).
Sparse index tracking based on $L_{1/2}$ model and algorithm.
arXiv preprint.

Jansen, R., and Van Dijk, R. (2002).
Optimal benchmark tracking with small portfolios.
The Journal of Portfolio Management vol. 28, no. 2, pp. 33–39.

Sun, Y., Babu, P., and Palomar, D. P. (2016).
Majorization-minimization algorithms in signal processing, communications, and machine learning.
IEEE Transactions on Signal Processing,, vol. 65, pp. 794–816, Feb.

Beasley, J. E., Meade, N., and Chang, T.-J. (2003).
An evolutionary heuristic for the index tracking problem.
European Journal of Operational Research, vol. 148, no. 3, pp. 621–643.

Scozzari, A., Tardella, F., Paterlini, S., and, Krink, T. (2013).
Exact and heuristic approaches for the index tracking problem with UCITS constraints.
Annals of Operations Research, vol. 205, no. 1, pp. 235–250.

Thanks

For more information visit:

http://www.danielppalomar.com