Sparse Index Tracking via MM

Prof. Daniel P. Palomar

ELEC5470/IEDA6100A - Convex Optimization
The Hong Kong University of Science and Technology (HKUST)
Fall 2020-21
1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
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 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 $\ell_1$-norm minimization techniques (Benidis et al. 2018a)$^1$, (Benidis et al. 2018b)$^2$.

- One example is the replication of an index, e.g., Hang Seng Index, based on a reduced basket of assets.

---


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: $r^b = [r^b_1, \ldots, r^b_T]^\top \in \mathbb{R}^T$

Returns of $N$ assets in $T$ days: $X = [r_1, \ldots, r_T]^\top \in \mathbb{R}^{T \times N}$ with $r_t \in \mathbb{R}^N$

Assume that an index is composed by a weighted collection of $N$ assets with normalized index weights $\mathbf{b}$ satisfying

- $\mathbf{b} > 0$
- $\mathbf{b}^\top \mathbf{1} = 1$
- $X \mathbf{b} = r^b$

We want to design a (sparse) tracking portfolio $\mathbf{w}$ satisfying

- $\mathbf{w} \geq 0$
- $\mathbf{w}^\top \mathbf{1} = 1$
- $X \mathbf{w} \approx r^b$
How should we select $w$?

Straightforward solution: full replication $w = b$
- 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)
How should we select $w$?

**Straightforward solution: full replication $w = b$**
- 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)
How should we select $w$?

Straightforward solution: full replication $w = b$
- 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?
  - Sparse index tracking.

- Use a small number of assets: $\text{card}(w) < N$
  - 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?
  📈 Sparse index tracking.

- Use a small number of assets: \( \text{card}(w) < N \)
  - 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?
  - 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

1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
Outline

1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
Sparse regression:

\[
\text{minimize } \| r - Xw \|_2 + \lambda \|w\|_0
\]

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

$$
\text{TE}(\mathbf{w}) = (\mathbf{w} - \mathbf{b})^T \Sigma (\mathbf{w} - \mathbf{b})
$$

where $\Sigma$ 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):

$$
\text{ETE}(\mathbf{w}) = \frac{1}{T} \left\| \mathbf{Xw} - \mathbf{r}_b \right\|^2_2
$$
Formulation for sparse index tracking

- Problem formulation for sparse index tracking (Maringer and Oyewumi 2007):

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda \|w\|_0 \\
\text{subject to} & \quad w \in \mathcal{W}
\end{align*}
\]

- \(\|w\|_0\) is the \(\ell_0\)–"norm" and denotes \(\text{card}(w)\)
- \(\mathcal{W}\) is a set of convex constraints (e.g., \(\mathcal{W} = \{w | w \geq 0, w^\top 1 = 1\}\))
- we will treat any nonconvex constraint separately

- This problem is too difficult to deal with directly:
  - Discontinuous, non-differentiable, non-convex objective function.

---

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

The two-step approach is much worse than joint optimization:
Outline

1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
Interlude: Majorization-Minimization (MM)

- Consider the following presumably difficult optimization problem:

\[
\begin{align*}
\text{minimize } & \quad f(x) \\
\text{subject to } & \quad x \in \mathcal{X},
\end{align*}
\]

with \( \mathcal{X} \) being the feasible set and \( f(x) \) being continuous.

- Idea: successively minimize a more manageable surrogate function \( u(x, x^{(k)}) \):

\[
\begin{align*}
x^{(k+1)} &= \underset{x \in \mathcal{X}}{\arg \min} u(x, x^{(k)}) ,
\end{align*}
\]

hoping the sequence of minimizers \( \{x^{(k)}\} \) will converge to optimal \( x^* \).

- Question: how to construct \( u(x, x^{(k)}) \)?

Answer: that’s more like an art (Sun et al. 2017).\(^4\)

Construction rule:

\[
\begin{align*}
  u(y, y) &= f(y), \quad \forall y \in \mathcal{X} \\
  u(x, y) &\geq f(x), \quad \forall x, y \in \mathcal{X} \\
  u'(x, y; d) \big|_{x=y} &= f'(y; d), \quad \forall d \text{ with } y + d \in \mathcal{X} \\
  u(x, y) &\text{ is continuous in } x \text{ and } y
\end{align*}
\]
Algorithm MM

Set $k = 0$ and initialize with a feasible point $x^0 \in \mathcal{X}$.

repeat
  $x^{(k+1)} = \arg\min_{x \in \mathcal{X}} \ u(x, x^{(k)})$
  $k \leftarrow k + 1$
until convergence

return $x^{(k)}$
Interlude on MM: Convergence

- Under some technical assumptions, every limit point of the sequence \( \{x^k\} \) is a stationary point of the original problem.

- If further assume that the level set \( \mathcal{X}^0 = \{ x | f(x) \leq f(x^0) \} \) is compact, then

\[
\lim_{k \to \infty} d(x^{(k)}, \mathcal{X}^*) = 0,
\]

where \( \mathcal{X}^* \) is the set of stationary points.

D. Palomar (HKUST)
Index Tracking 22 / 86
Approximation of the $\ell_0$-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 $u = 1$).
Continuous and differentiable approximate formulation:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \| Xw - r^b \|_2^2 + \lambda^T \rho_{p,u}(w) \\
\text{subject to} & \quad w \in \mathcal{W}
\end{align*}
\]

where \( \rho_{p,u}(w) = [\rho_{p,u}(w_1), \ldots, \rho_{p,u}(w_N)]^T \).

This problem is still non-convex: \( \rho_{p,u}(w) \) is concave for \( w \geq 0 \).

We will use MM to deal with the non-convex part.
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

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

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

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 $w_0 \in \mathbb{R}_+$.

$$
\rho_{p,\gamma}(w) = \frac{\log(1 + w/p)}{\log(1 + \gamma/p)} \\
\leq \frac{1}{\log(1 + \gamma/p)} \left[ \log \left( 1 + \frac{w_0}{p} \right) + \frac{1}{p + w_0} (w - w_0) \right] \\
= \frac{1}{\log(1 + \gamma/p)(p + w_0)} w \\
\underbrace{d_{p,\gamma}}_{1} + \underbrace{\frac{\log \left( 1 + \frac{w_0}{p} \right)}{\log(1 + \gamma/p)}}_{b_{p,\gamma}} - \frac{w_0}{\log(1 + \gamma/p)(p + w_0)}
$$
Majorization of $\rho_{p,\gamma}$
Iterative formulation via MM

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

$$\min_w \quad \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda d_{p,u}^{(k)} w$$

subject to $$w \in \mathcal{W}$$

where $$d_{p,u}^{(k)} = \begin{bmatrix} d_{p,u}(w_1^{(k)}), \ldots, d_{p,u}(w_N^{(k)}) \end{bmatrix}^\top$$.

This problem is convex (actually a QP).

Requires a solver in each iteration.
Algorithm 1: Linear Approximation for the Index Tracking problem (LAIT)

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

repeat

- Compute $\mathbf{d}_p \mathbf{u}^{(k)}$
- $\mathbf{w}^{(k+1)} = \arg \min_{\mathbf{w} \in \mathcal{W}} \frac{1}{T} \| \mathbf{Xw} - \mathbf{r}^b \|_2^2 + \lambda \mathbf{d}_p \mathbf{u}^{(k)} \mathbf{T} \mathbf{w}$
- $k \leftarrow k + 1$

until convergence

return $\mathbf{w}^{(k)}$
The big picture

\[
\min_{w} \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda \|w\|_0
\]
\[\text{s.t.} \quad w \in \mathcal{W}\]

\[\ell_0\text{-norm approximation}\]

\[
\min_{w} \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda 1^T \rho_{p,u}(w)
\]
\[\text{s.t.} \quad w \in \mathcal{W}\]

\[
\min_{w} \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda d_{p,u}^{(k)}^T w
\]
\[\text{s.t.} \quad w \in \mathcal{W}\]
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:
  \[
  \frac{1}{T} \| Xw - r^b \|^2 + \lambda d_{\rho, u}^{(k)} w = \frac{1}{T} w^T X^T X w + \left( \lambda d_{\rho, u}^{(k)} - \frac{2}{T} X^T r^b \right)^T w + \text{const}.
  \]

- Further upper-bound it:

**Lemma 2**

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

\[
w^T L w \leq w^T M w + 2w^{(k)^T} (L - M) w - w^{(k)^T} (L - M) w^{(k)}.
\]

Equality is achieved when \( w = w^{(k)} \).
Let’s majorize the objective function

- Based on Lemma 2:
  - Majorize the quadratic term $\frac{1}{T} w^T X^T X w$.
  - In our case $L_1 = \frac{1}{T} X^T X$.
  - We set $M_1 = \lambda_{\text{max}}^{(L_1)} I$ so that $M_1 \succeq L_1$ holds.

- The objective becomes:

$$w^T L_1 w + \left( \lambda d_{p,u}^{(k)} - \frac{2}{T} X^T r^b \right)^T w \leq w^T M_1 w + 2w^{(k)^T} (L_1 - M_1) w - w^{(k)^T} (L_1 - M_1) w^{(k)}$$

$$+ \left( \lambda d_{p,u}^{(k)} - \frac{2}{T} X^T r^b \right)^T w$$

$$= \lambda_{\text{max}}^{(L_1)} w^T w + \left( 2 \left( L_1 - \lambda_{\text{max}}^{(L_1)} I \right) w^{(k)} + \lambda d_{p,u}^{(k)} - \frac{2}{T} X^T r^b \right)^T w + \text{const.}$$
Specialized iterative formulation

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

\[
\begin{align*}
\text{minimize} \quad & \mathbf{w}^\top \mathbf{w} + \mathbf{q}_1^{(k)} \mathbf{w} \\
\text{subject to} \quad & \mathbf{w}^\top \mathbf{1} = 1, \\
& 0 \leq \mathbf{w} \leq 1, \\
\end{align*}
\]

where

\[
\mathbf{q}_1^{(k)} = \frac{1}{\lambda_{\text{max}}^{(L_1)}} \left( 2 \left( \mathbf{L}_1 - \lambda_{\text{max}}^{(L_1)} \mathbf{1} \right) \mathbf{w}^{(k)} + \lambda \mathbf{d}_{p,u}^{(k)} - \frac{2}{T} \mathbf{X}^\top \mathbf{r}^b \right).
\]

- This problem can be solved with a closed-form update algorithm.
Proposition 1

The optimal solution to the previous problem with $u = 1$ is:

$$w^*_i = \begin{cases} -\frac{\mu + q_i}{2}, & i \in A, \\ 0, & i \notin A, \end{cases}$$

with

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

and

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

where $A$ can be determined in $O(\log(N))$ steps.
Algorithm 2: Specialized Linear Approximation for the Index Tracking problem (SLAIT)

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

repeat
  \begin{itemize}
  \item Compute \( q_1^{(k)} \)
  \item \( \mathbf{w}^{(k+1)} = \arg\min_{\mathbf{w} \in \mathcal{W}} \mathbf{w}^T \mathbf{w} + q_1^{(k)T} \mathbf{w} \)
  \item \( k \leftarrow k + 1 \)
  \end{itemize}
until convergence

return \( \mathbf{w}^{(k)} \)
The big picture

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda \| w \|_0 \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]

\(\ell_0\)-norm approximation

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda^T p_{\rho,u}(w) \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda^T d_{\rho,u}^{(k)} w \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]
The big picture

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda \|w\|_0 \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]

**\( \ell_0 \)-norm approximation

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda^T \rho_{p,u}(w) \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]

\[ \begin{align*}
\min_{w} & \quad w^T w + q^{(k)}_1^T w \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]

**MM**

\[ \begin{align*}
\min_{w} & \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda \left[ (F) R^T + (F) \right]^T w \\
\text{s.t.} & \quad w \in \mathcal{W}
\end{align*} \]
Outline

1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
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

\[
I \odot I_{\{w > 0\}} \leq w \leq u \odot I_{\{w > 0\}}
\]

- Active constraints only for the selected assets \((w_i > 0)\).

- Upper bound is easy: \(w \leq u \odot I_{\{w > 0\}} \iff w \leq u\) (convex and can be included in \(\mathcal{W}\)).

- Lower bound is nasty. 😞
Problem formulation

The problem formulation with holding constraints becomes (after the $\ell_0$-“norm” approximation):

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \| Xw - r_b \|_2^2 + \lambda \mathbf{1}^\top \rho_{\rho,u}(w) \\
\text{subject to} & \quad w \in \mathcal{W}, \\
& \quad l \odot \mathcal{I}_{\{w > 0\}} \leq w.
\end{align*}
\]

- How should we deal with the non-convex constraint?
Penalization of violations

- Hard constraint $\implies$ 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 $\gamma = l$:

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

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

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda^T \rho_{p,u}(w) + \nu^T \tilde{f}_{p,l}(w) \\
\text{subject to} & \quad w \in \mathcal{W}
\end{align*}
\]

where \( \nu \) is a parameter vector that controls the penalization and 
\( \tilde{f}_{p,l}(w) = [\tilde{f}_{p,l}(w_1), \ldots, \tilde{f}_{p,l}(w_N)]^T \).

- This problem is not convex:
  - \( \rho_{p,u}(w) \) is concave \( \implies \) Linear upperbound with Lemma 1.
  - \( \tilde{f}_{p,l}(w) \) is neither convex nor concave.
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

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

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].

$$\tilde{f}_{p,l}(w) = \max (\rho_{p,l}(w) \cdot l - w, 0)$$

$$\leq \max \left( (d_{p,l}(w^{(k)})w + c_{p,l}(w^{(k)})) \cdot l - w, 0 \right)$$

$$= \max \left( (d_{p,l}(w^{(k)}) \cdot l - 1) w + c_{p,l}(w^{(k)}) \cdot l, 0 \right).$$

$h_{p,l}(w, w^{(k)})$ 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 $h_{p,l}(w, w^{(k)})$: 

![Graph showing majorization relationship between $\tilde{f}_{p,l}$, $h_{p,l}$, and $q_{p,c,l}$]
Convex formulation of the majorization

- Recall our problem:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda \mathbf{1}^\top \rho_{p,u}(w) + \nu^\top \tilde{f}_{p,l}(w) \\
\text{subject to} & \quad w \in \mathcal{W}.
\end{align*}
\]

- From Lemma 1: \( \rho_{p,u}(w) \leq \mathbf{d}_{p,u}^{(k)}^\top w + \text{const.} \)

- From Lemma 3:

\[
\tilde{f}_{p,l}(w) = (\rho_{p,l}(w) \cdot \mathbf{1} - w)^+ \leq \left( \text{Diag} \left( \mathbf{d}_{p,l}^{(k)} \circ \mathbf{1} - \mathbf{1} \right) w + \mathbf{c}_{p,l}^{(k)} \circ \mathbf{1} \right)^+ = h_{p,l}(w, w^{(k)})
\]

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

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda \mathbf{d}_{p,u}^{(k)}^\top w + \nu^\top h_{p,l}(w, w^{(k)}) \\
\text{subject to} & \quad w \in \mathcal{W}
\end{align*}
\]

which is convex.
Algorithm LAITH

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

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

repeat

- Compute $d_{p,l}^{(k)}$, $d_{p,u}^{(k)}$
- Compute $c_{p,l}^{(k)}$
- $w^{(k+1)} = \arg\min_{w \in \mathcal{W}} \frac{1}{T} \|Xw - r^b\|_2^2 + \lambda d_{p,u}^{(k)\top} w + \nu^{\top} h_{p,l}(w, w^{(k)})$
- $k \leftarrow k + 1$

until convergence

return $w^{(k)}$
The big picture

\[
\begin{align*}
\min_{\mathbf{w}} & \quad \frac{1}{T} \left\| \mathbf{Xw} - \mathbf{r}^b \right\|_2^2 + \lambda \left\| \mathbf{w} \right\|_0 \\
\text{s.t.} & \quad \mathbf{w} \in \mathcal{W}, \\
& \quad \mathbf{l} \odot \mathcal{I}\{\mathbf{w} > 0\} \leq \mathbf{w}.
\end{align*}
\]

\(\ell_0\)-norm approximation / soft constraint

\[
\begin{align*}
\min_{\mathbf{w}} & \quad \frac{1}{T} \left\| \mathbf{Xw} - \mathbf{r}^b \right\|_2^2 + \lambda \mathbf{1}^\top \rho_{p,u}(\mathbf{w}) \\
& \quad + \nu^\top \tilde{\mathbf{f}}_{p,l}(\mathbf{w}) \\
\text{s.t.} & \quad \mathbf{w} \in \mathcal{W}
\end{align*}
\]

\[
\begin{align*}
\min_{\mathbf{w}} & \quad \frac{1}{T} \left\| \mathbf{Xw} - \mathbf{r}^b \right\|_2^2 + \lambda \mathbf{d}_p^\top \mathbf{w} \\
& \quad + \nu^\top \tilde{\mathbf{h}}_{p,l}(\mathbf{w}, \mathbf{w}^{(k)}) \\
\text{s.t.} & \quad \mathbf{w} \in \mathcal{W}
\end{align*}
\]
Again, for specific constraint sets we can derive closed-form update algorithms!
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.,

\[
\begin{align*}
    h_{p,l}(\mathbf{w}, \mathbf{w}^{(k)}) &= \left( \text{Diag} \left( \mathbf{d}_{p,l}^{(k)} \odot \mathbf{l} - \mathbf{1} \right) \right) \mathbf{w} + \mathbf{c}_{p,l}^{(k)} \odot \mathbf{l}^{+}.
\end{align*}
\]

- **Separable**: focus only in the univariate case, i.e., \( h_{p,l}(\mathbf{w}, \mathbf{w}^{(k)}) \).
- **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 $\tilde{h}_{p,\epsilon,l}(w, w^{(k)})$:  

![Graph showing smooth majorization and approximations](image_url)
**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

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

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.
Quadratic majorization of $\tilde{f}_{p,l}(w)$

- 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:
  \[
  \begin{align*}
  \text{minimize} & \quad \frac{1}{T} \| Xw - r^b \|^2_2 + \lambda d_{p,u}^{(k)\top} w + \nu^{\top} \tilde{h}_{p,\epsilon,l}(w, w^{(k)}) \\
  \text{subject to} & \quad w \in \mathcal{W}
  \end{align*}
  \]

- From Lemma 4:
  \[
  \tilde{h}_{p,\epsilon,l}(w, w^{(k)}) \leq w^{\top} \text{Diag} \left( a_{p,\epsilon,l}^{(k)} \odot \nu \right) w + b_{p,\epsilon,l}^{(k)} \odot \nu^{\top} w + \text{const}.
  \]

- The majorized problem at the \((k + 1)\)-th iteration becomes:
  \[
  \begin{align*}
  \text{minimize} & \quad w^{\top} \left( \frac{1}{T} X^{\top} X + \text{Diag} \left( a_{p,\epsilon,l}^{(k)} \odot \nu \right) \right) w \\
  & \quad + \left( \lambda d_{p,u}^{(k)} - \frac{2}{T} X^{\top} r^b + b_{p,\epsilon,l}^{(k)} \odot \nu \right)^{\top} w \\
  \text{subject to} & \quad w \in \mathcal{W}
  \end{align*}
  \]

- This problem is a QP that can be solved with a solver, but we can do better.
Quadratic formulation of the majorization

- Use Lemma 2 to majorize the quadratic part:
  
  \[
  \mathbf{L}_2 = \frac{1}{T} \mathbf{X}^\top \mathbf{X} + \text{Diag}\left( \mathbf{a}_p^{(k)} \odot \mathbf{\nu} \right)
  \]
  
  \[
  \mathbf{M}_2 = \lambda_{\text{max}}^{(L_2)} \mathbf{I}.
  \]

- And the final optimization problem at the \((k+1)\)-th iteration becomes:

  \[
  \begin{align*}
  \text{minimize} & \quad \mathbf{w}^\top \mathbf{w} + \mathbf{q}_2^{(k)} \mathbf{w} \\
  \text{subject to} & \quad \mathbf{w} \in \mathcal{W},
  \end{align*}
  \]

  where

  \[
  \mathbf{q}_2^{(k)} = \frac{1}{\lambda_{\text{max}}^{(L_2)}} \left( 2 \left( \mathbf{L}_2 - \lambda_{\text{max}}^{(L_2)} \mathbf{I} \right) \mathbf{w}^{(k)} + \lambda \mathbf{d}_p,u - \frac{2}{T} \mathbf{X}^\top \mathbf{r}^b + \mathbf{b}_p,e,l \odot \mathbf{\nu} \right).
  \]

- This problem can be solved in closed form!
Algorithm SLAITH

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

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

repeat
- Compute $q^{(k)}_2$
- Solve $w^{(k+1)} = \arg\min_{w \in \mathcal{W}} w^T w + q^{(k)}_2 w^T$, using Proposition 1.
- $k \leftarrow k + 1$

until convergence

return $w^{(k)}$
The big picture

\[ \min_w \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda \|w\|_0 \]
\[ \text{s.t.} \quad w \in \mathcal{W}, \quad I \odot I_{\{w > 0\}} \leq w. \]

\( \ell_0 \)-norm approximation / soft constraint

\[ \min_w \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda 1^T \rho_{p,u}(w) \]
\[ + \nu^T f_{p,i}(w) \]
\[ \text{s.t.} \quad w \in \mathcal{W} \]

smooth approx. / MM

\[ \min_w \quad \frac{1}{T} \left\| Xw - r^b \right\|_2^2 + \lambda d_{p,u}^{(k)}^T w \]
\[ + \nu^T f_{p,i}(w, w^{(k)}) \]
\[ \text{s.t.} \quad w \in \mathcal{W} \]
Extension to other tracking error measures

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

\[ \text{ETE}(w) = \frac{1}{T} \| r^b - Xw \|^2. \]

However, we can use other tracking error measures such as (Benidis et al. 2018b):\(^5\)

- **Downside risk:**
  \[ \text{DR}(w) = \frac{1}{T} \| (r^b - Xw)^+ \|^2, \]
  where \((x)^+ = \max(0, x)\).
- **Value-at-Risk (VaR) relative to an index.**
- **Conditional VaR (CVaR) relative to an index.**

---

DR(\(w\)) 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 \(\text{DR}(w) = \frac{1}{T} \| (r^b - Xw)^+ \|^2\) is majorized at \(w^{(k)}\) by the quadratic convex function

\[
\frac{1}{T} \| r^b - Xw - y^{(k)} \|^2,
\]

where \(y^{(k)} = - \left( Xw^{(k)} - r^b \right)^+ \).
Proof of Lemma 5 (1/4)

For convenience set \( z = r^b - Xw \). Then:

\[
\text{DR}(w) = \frac{1}{T} \| (z)^+ \|_2^2 = \frac{1}{T} \sum_{i=1}^{T} \tilde{z}_i^2,
\]

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 \( z_i^{(k)} \leq 0 \), \( f_2(z_i|z_i^{(k)}) = (z_i - z_i^{(k)})^2 \) 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:

![Graph showing the functions $\tilde{z}^2$ and $f_1(z|z^{(k)})$ for $z^{(k)} > 0$ and $f_2(z|z^{(k)})$ for $z^{(k)} \leq 0$.]
Proof of Lemma 5 (3/4)

Combining the two cases:

\[
\tilde{z}_i^2 \leq \begin{cases} 
  f_1(z_i | z_i^{(k)}), & \text{if } z_i^{(k)} > 0, \\
  f_2(z_i | z_i^{(k)}), & \text{if } z_i^{(k)} \leq 0,
\end{cases}
\]

\[
= \begin{cases} 
  (z_i - 0)^2, & \text{if } z_i^{(k)} > 0, \\
  (z_i - z_i^{(k)})^2, & \text{if } z_i^{(k)} \leq 0,
\end{cases}
\]

\[
=(z_i - y_i^{(k)})^2,
\]

where

\[
y_i^{(k)} = \begin{cases} 
  0, & \text{if } z_i^{(k)} > 0, \\
  z_i^{(k)}, & \text{if } z_i^{(k)} \leq 0,
\end{cases}
\]

\[
= - (-z_i^{(k)})^+.
\]
Thus, $\text{DR}(z)$ is majorized as follows:

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

Substituting back $z = r^b - Xw$, we get

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

where $y^{(k)} = -(z^{(k)})^+ = -(Xw - r^b)^+$. 
Extension to other penalty functions

- Apart from the various performance measures, we can select a different penalty function.
- We have used only the $\ell_2$-norm to penalize the differences between the portfolio and the index.
- We can use the Huber penalty function for robustness against outliers (Benidis et al. 2018b):

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

- The $\ell_1$-norm.
- Many more...

---

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

\[ \phi(x) \]

\[ f(x|\mathbf{x}_0) \text{ for } |\mathbf{x}_0| \leq M \]

\[ f(x|\mathbf{x}_0) \text{ for } |\mathbf{x}_0| > M \]
Outline

1 Introduction

2 Sparse Index Tracking
   - Problem formulation
   - Interlude: Majorization-Minimization (MM) algorithm
   - Resolution via MM

3 Holding Constraints and Extensions
   - Problem formulation
   - Holding constraints via MM
   - Extensions

4 Numerical Experiments

5 Conclusions
Set up

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

Table 1: Index Information

<table>
<thead>
<tr>
<th>Index</th>
<th>Data Period</th>
<th>$T_{\text{trn}}$</th>
<th>Ttst</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>01/01/10 - 31/12/15</td>
<td>252</td>
<td>252</td>
</tr>
<tr>
<td>Russell 2000</td>
<td>01/06/06 - 31/12/15</td>
<td>1000</td>
<td>252</td>
</tr>
</tbody>
</table>

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

\[
\text{MDTE} = \frac{1}{T - T_{\text{trn}}} \| \text{diag}(XW) - r^b \|_2,
\]

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

We will use the following benchmark methods:

- MIP solution by Gurobi solver (MIP$_{Gur}$).
- Diversity Method (Jansen and Van Dijk 2002)$^7$ where the $\ell_{1/2}$-“norm” approximation is used (DM$_{1/2}$).
- Hybrid Half Thresholding (HHT) algorithm (Xu et al. 2015)$^8$.

---


S&P 500 - w/o holding constraints

[Graphs showing the relationship between MDTE (bps) and the number of selected assets, as well as the average running time (s) for different methods.]
Russell 2000 - w/o holding constraints

![Graph 1: MDTE (bps) vs Number of selected assets](image1)

- **DM$_{1/2}$**
- **HHT**
- **MIP$_{Gur}$**
- **SLAIT (proposed)**
- **LAIT (proposed)**

![Graph 2: Average running time (s) vs Number of selected assets](image2)

- **Time Capping**
- **DM$_{1/2}$**
- **HHT**
- **MIP$_{Gur}$**
- **SLAIT (proposed)**
- **LAIT (proposed)**
S&P 500 - w/ holding constraints

![Graph showing MDTE (bps) vs. Number of selected assets](image1)

![Graph showing Average running time (s) vs. Number of selected assets](image2)
Russell 2000 - w/ holding constraints

**Graph 1:**
- Title: MDTE (bps) vs Number of selected assets
- X-axis: Number of selected assets
- Y-axis: MDTE (bps)
- Legend:
  - MIP\textsubscript{Gur-h}
  - SLAITH (proposed)
  - LAITH (proposed)

**Graph 2:**
- Title: Average running time (s) vs Number of selected assets
- X-axis: Number of selected assets
- Y-axis: Average running time (s)
- Legend:
  - Time Capping
  - MIP\textsubscript{Gur-h}
  - SLAITH (proposed)
  - LAITH (proposed)
Average running time of proposed methods

- Comparison of $AS_1$ and $AS_u$:
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


