CSCI 1951-G – Optimization Methods in Finance
Part 04:
Building Index Funds with Integer Programming

February 23, 2018
Stock market index

Heard of S&P500, Dow Jones Industrial Avg, NASDAQ composite? They are *Stock market indices*:

measure the *composite value* $V$ of a set $S$ of stocks

$$V = \sum_{\text{stock } i \in S} w_i p_i$$

**price-weighted:** $w_i = 1$, for each stock $i \in S$

**market-share:** $n_i$: # of *outstanding shares* for stock $i \in S$,

$$w_i = n_i / \sum_{\text{stock } i \in S} n_i$$

**market-capitalization:** $v_i = n_i p_i$: total *value of outstanding shares* for stock $i \in S$,

$$w_i = v_i / \sum_{\text{stock } i \in S} v_i$$

An index is a *statistic*, not an asset (you cannot invest in an index)
Index fund

An investment strategy aiming to provide the “same” returns as an index.

I.e., a portfolio of stocks whose behavior, collectively, replicates (tracks) that of the index:
- each stock makes up a fraction of the portfolio
- the stocks and their fractions are chosen at the beginning
- the stocks and their fractions are fixed over time
- Exception: periodic rebalancing (adjustment) of index

Easy and bad way to build an index fund

Buy each stock in the index proportionally to its weight

Bad because…we end up with few shares of many stocks: inefficient and costly (transaction costs, fractional shares)
Motivation

• No one hits \textit{above-index} returns \textit{consistently}.
• Why not just trying to \textit{always hit the average}?

“What we need is a no-load, \textit{minimum management-fee} mutual fund that simply buys the hundreds of stocks making up the \textit{broad stock-market averages} and \textit{does no trading} from security to security in an attempt to catch the winners.” (1973)
Index funds were insulted and ridiculed at the beginning, now they are very successful.

The first index fund by John Bogle (1975) was called *un-American*. He founded Vanguard, the largest mutual fund company in the US.

“[I couldn’t] believe that the great mass of investors are going to be satisfied with *receiving just average returns*”

Fidelity Investments Chairman Edward Johnson

- Most investors are average people
- They should actually be fine with average returns!
Index funds (cont.)

There are hundreds of thousands of index funds, because there are that many indices, and you can always define your own index.

In 2014: 20.2% of equity mutual fund assets in the US were index funds.

2007-14: +$ 1 trillion in net new cash in index funds and ETFs (exchange traded funds)

2017-14; -$ 659 billion in net outflow of actively managed mutual funds.
Managed vs unmanaged investment strategies

- Index funds are *unmanaged* investment strategies: we buy stocks at the beginning and we hold them, no matter what.

- *Managed* strategies involve continuously looking for overvalued stocks to sell and undervalued stocks to buy: they look for stocks that will *outperform the markets*, i.e., the indices.

Which one is better?

**Answer (It’s not clear!)**

- *Theoretical ground and empirical evidence* suggest that *unmanaged strategies are better*

- *There is also evidence to the opposite*
Why may index funds be better?

Efficient Market Hypothesis (Eugene Fama)

“[Stock] prices fully reflect all available information”

The prices reflect information to the point that marginal returns that could be obtained by exploiting additional information do not exceed marginal costs.

- Fund managers are *aggressively continuously looking* for undervalued/overvalued stocks
- Any information about the *real value* of a company will be *almost immediately reflected* in the stock price
- The fast adjustment makes it impossible to detect potentially market-outperforming stocks, not to mention act on them
- So why even try to find them?
Empirical evidence

- Actively managed strategies have *often underperformed* with respect to many indices
- An index fund has *lower management costs* (you do not have to pay someone a lot for picking stocks well)

“Put 10% […] in short-term government bonds and 90% *in a very low-cost S&P 500 index fund*. […] long-term results from this policy will be superior to those attained by most investors […] who employ high-fee managers.” Warren Buffett, 2014
Why may index funds not be better?

- *Theoretical* critiques to the Efficient Market Hypothesis, including challenges to assuming *rational investors*. If interested, look for *behavioral finance*

- *Empirical* evidence:
  - Indices announce in advance when they are going to rebalance and how;
  - Creates changes in prices that actively managed fund managers exploit to extract value while index funds can’t use this information
How to build an index fund

Easy way to build an index fund

Buy each stock in the index proportionally to its weight.

This is bad because we end up with few shares of many stocks, which has a cost in practice;

Our Goal

We want to pick *a few* stocks that closely replicate the whole index.
Data and settings

- The index contains $n$ stocks
- Pick $q$ stocks for the index fund, with $q \ll n$
- For each pair $(i, j)$ of stocks, let $\rho_{ij}$ be their similarity:
  \[ \rho_{ii} = 1 \quad \rho_{ij} \leq 1. \]

  E.g., the correlation of their returns (other measures are possible)
- In practice, $\rho_{ij}$ is estimated from historical data: it may include the sizes of the companies, their sector, the historical behavior of their stock prices, …
An Integer Programming formulation

- **Clustering approach**: each stock $i$ will be *represented by one and only one* stock $j$ in the index fund
- Pick the $q$ representative stock that *maximize the similarity* between the $n$ stocks and the representatives
- Let $y_j$ be 1 if stock $j$ is chosen as representative, 0 otherwise
- Let $x_{ij} = 1$ if stock $j$ is the representative for stock $i$, 0 otherwise

$$
\begin{align*}
\text{max} & \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij} \\
\text{s.t.} & \quad \sum_{j=1}^{n} y_j = q \\
& \quad \sum_{j=1}^{n} x_{ij} = 1 \text{ for } i = 1, \ldots, n \\
& \quad x_{ij} \leq y_j \text{ for } i = 1, \ldots, n, j = 1, \ldots, n \\
& \quad x_{ij}, y_j \in \{0, 1\} \text{ for } i = 1, \ldots, n, j = 1, \ldots, n
\end{align*}
$$
Proportions

**Question**

*How many shares of each representative stock should we buy?*

**Answer**

*A number proportional to the value of the stocks it represents*

- After having selected the $q$ representative, we compute, for each representative $j$, the weight

  \[ w_j = \sum_{i=1}^{n} v_i x_{ij}, \]

  where $v_i$ is the total market value of the stock $i$

- The fraction of the index fund to invest into stock $j$ is proportional to $w_j$:

  \[ p_j = \frac{w_j}{\sum_{i=1}^{n} w_i} \]
Issue with this formulation

\[
\begin{align*}
\text{max} & \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij} \\
\text{s.t.} & \quad \sum_{j=1}^{n} y_{j} = q \\
& \quad \sum_{j=1}^{n} x_{ij} = 1 \text{ for } i = 1, \ldots, n \\
& \quad x_{ij} \leq y_{j} \text{ for } i = 1, \ldots, n, j = 1, \ldots, n \\
& \quad x_{ij}, y_{j} \in \{0, 1\} \text{ for } i = 1, \ldots, n, j = 1, \ldots, n
\end{align*}
\]

Question

Let \( n = 500 \) and \( q = 20 \)

How many variables has this formulation?
How many constraints?

Solving many LP-relaxations of this problem is not a viable option
### Question

*What is the goal of solving the LP-relaxation during the branch-and-bound algorithm?*

### Answer

*Obtaining upper/lower bounds to the optimal integral solution*

- We want to obtain the bounds *fast*
- We obtained the LP-relaxation by relaxing the integrality constraint
- What about relaxing also the other constraints?
Lagrangian relaxation

Idea to speed up the computation of bounds

- Remove some of the constraints that specify the feasible region
- Add those constraints into the objective function, *penalizing their violation*

Consider the ILP:

\[
\begin{align*}
\text{max } & \quad c^T x \\
A_1 x & \leq b_1 \\
A_2 x & \leq b_2
\end{align*}
\]

We move (2) into the objective function, penalizing its violation:

\[
L(\lambda) = \max c^T x + \lambda^T (b_2 - A_2 x) \\
A_1 x & \leq b_1
\]

where \(\lambda\) is a vector of non-negative weights

The objective value is penalized when \(A_2 x \not\leq b_2\), and “rewarded” otherwise
Lagrangian relaxation (cont.)

\[ L(\lambda) = \max c^T x + \lambda^T (b_2 - A_2 x) \]
\[ A_1 x \leq b_1 \]

**Definition**

The function \( L(\lambda) \) is known as the *Lagrangian function*, the modified optimization problem as *Lagrangian relaxation*, and the weights \( \lambda \) as *Lagrangian multipliers*.

**Question**

*What can we say about the optimal solution of the Lagrangian relaxation vs. the optimal solution of the original ILP?*

**Theorem**

*The optimal solution of the Lagrangian relaxation is an upper bound to the optimal solution of the original ILP.*
The Lagrangian relaxation of the index fund ILP problem is:

\[
L(\lambda) = \max \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij} x_{ij} + \sum_{i=1}^{n} \lambda_i \left(1 - \sum_{j=1}^{n} x_{ij}\right)
\]

s.t. \[\sum_{j=1}^{n} y_j = q\]

\[x_{ij} \leq y_j \text{ for } i = 1, \ldots, n, j = 1, \ldots, n\]

\[x_{ij}, y_j \in \{0, 1\} \text{ for } i = 1, \ldots, n, j = 1, \ldots, n\]

How many constraints has this formulation?

We cut the number of constraints in half: solving the (LP-relaxation of the) Lagrangian relaxation will be much faster than solving the LP-relaxation of the original problem

In the homework you’ll see how to get an ever better formulation.