# Minimizing Tracking Error While Restricting the Number of Assets [*]

Thomas F. Coleman, Yuying Li

Department of Computer Science, Cornell University, Ithaca, New York, USA

Jay Henniger

Center for Applied Mathematics, Cornell University, Ithaca, New York, USA


Contact: Prof. Thomas F. Coleman

Phone: (607) 255-9203, Email: coleman@tc.cornell.edu

November 23, 2004

**Abstract**

Tracking error minimization is commonly used by the traditional passive fund managers as well as alternative portfolio (for example, hedge fund) managers. We propose a graduated non-convexity method to minimize portfolio tracking error with the total number of assets no greater than a specified integer $K$. The solution of this tracking error minimization problem is the global minimizer of the sum of the tracking error function and the discontinuous counting function. We attempt to track the globally minimal tracking error portfolio by approximating the discontinuous counting function with a sequence of continuously differentiable non-convex functions, a graduated non-convexity process. We discuss the advantages of this approach, present numerical results, and compare it with two methods from recent literature.

# Index Tracking

Portfolio managers evaluate the performance of their portfolios by comparing it to a benchmark, e.g., the index portfolio. Holding relatively few assets prevents a portfolio from holding very small and illiquid positions and limits administration and transaction costs. A practical problem for passive portfolio management is the index tracking problem of finding a portfolio of a small number of stocks which minimizes a chosen measure of index tracking error; for example, consider minimizing the index tracking error with the portfolio size no greater than a specified number of instruments $K$. Tracking error minimization is also frequently used by hedge fund managers as part of a long-short hedging strategy. In this case, the tracking error measures the difference from a pre-selected benchmark long portfolio or short portfolio and tracking error minimization yields the hedging positions. Although our discussion is illustrated with the index tracking example, our proposed method is applicable to any tracking error minimization problem subject to a constraint on the total number of assets.

This tracking error minimization problem, with a restriction on the total number of assets, is NP-hard and consequently heuristic methods have been suggested. As described in Jansen and Dijk [2002], a simple heuristic algorithm that is common for solving the cardinality-constrained index tracking problem can be illustrated as follows. As an example consider the problem of choosing a portfolio consisting of 25 stocks to track the S&P500 index. Suppose that

the tracking error function is $\text{TE}_{\text{JD}}(x) = (x - w)^T Q (x - w)$, which is used in Jansen and Dijk [2002]. Here, $x_i$ is the percentage of the portfolio invested in stock $i, 1 \leq i \leq n$, $w$ is the vector of percentage weights of the stocks in the index, and $Q$ is the (positive definite) covariance matrix of the stock returns. This measure of tracking error, along with two others, are discussed further in the next section. The simple heuristic algorithm consists of the following steps. First, one solves the quadratic programming problem: choose the best weights $x_i$ of the 500 stocks in the S&P500 to minimize the tracking error (so $x = w$ is optimal initially). Then, remove the 25 stocks that are weighted smallest in this solution, and solve the problem of finding the best portfolio of the remaining 475 stocks to minimize the tracking error. This is also a quadratic programming problem. Continue in this way until only 25 stocks remain. In general, this algorithm could proceed by removing any number of stocks after each solution, say 10 stocks or 1 stock at a time. Besides being ad hoc, a disadvantage of this heuristic method is that it may require solving many index tracking sub-problems with hundreds of variables.

Another heuristic method for the index tracking problem is proposed by Beasley, Meade and Chang [1999]. They use a population heuristic (genetic algorithm) to search for a good tracking portfolio by imposing the cardinality constraint explicitly. In this case, all members of the population of tracking portfolios have the desired number of instruments. This heuristic approach admits a very general problem formulation, allowing the imposition of a limit

2

on transaction costs (assuming some initial tracking portfolio is held, and re-balancing of the initial portfolio is needed) as well as limiting the maximum or minimum holding of any stock in the portfolio and the use of virtually any measure of tracking error.

Salkin and Meade [1989] investigate tracking an index by constructing a portfolio that matches the sector-weightings of the index. They also consider using the relative market capitalizations of the stocks in the index as the relative holdings in the tracking portfolio. However, restriction on the total number of stocks in the portfolio is not considered in their paper.

Mathematically, a tracking error minimization problem subject to a cardinality constraint can be formulated as computing the global minimizer of an objective function involving a measure of tracking error and a discontinuous counting function $\sum_{i=1}^{n} \Lambda(x_i)$, where $\Lambda(x_i)$ equals 1 if $x_i \neq 0$ and 0 otherwise. In addition, simple constraints (typically linear) may exist. The tracking error minimization problem is difficult to solve since there is an exponential number of local minimizers, with each one corresponding to an optimal tracking portfolio from a fixed subset of stocks. Jansen and Dijk [2002] present the idea of solving the index tracking problem by approximating the discontinuous counting function $\Lambda(z)$ by a sequence of *continuous but not continuously differentiable* functions. To implement this idea they use a penalty function approach and choose one approximation function from the sequence to approximate the counting function.

In this case, however, the lack of differentiability of the selected approximation to the counting function causes some difficulty. In particular, the first and second derivatives of the objective function are not well-behaved when one or more of the holdings $x_i$ are close to zero. This is problematic because many of the stock holdings are expected to approach zero when the desired total number of stocks, $K$, is small. In particular, under reasonable assumptions (see Appendix C for details) the reduced Hessian matrix can be arbitrarily ill-conditioned at or near solutions to the cardinality-constrained index tracking problem. This method is described in detail in the Results section, where computational results are presented.

In this paper, we propose to solve the tracking error minimization problem subject to a cardinality constraint by approximating the discontinuous function $\Lambda(x_i)$ by a sequence of *continuously differentiable* non-convex piecewise quadratic functions which approach $\Lambda(x_i)$ in the limit. To further describe this approach, consider the convex tracking error function for example. The proposed method begins by solving a convex programming problem without the cardinality constraint and computes its global minimizer. Then, from this minimizer, a sequence of local minima of approximations to the tracking error minimization problem is tracked, using the minimizer of the previous approximation problem as a starting point. In each successive approximation to the tracking error minimization problem, additional negative curvature is introduced to the objective function through the approximation to the count-

ing function. Our proposed method is an adaptation of the known graduated non-convexity method for tracking the global minimum for the image reconstruction problem (see Blake and Zisserman [1987]).

## Tracking Error Minimization

Let $x_i$ represent the percentage weight of asset $i$ in the portfolio $x$. A tracking error minimization problem subject to a constraint on the total number of assets can be formulated as a constrained *discontinuous* optimization problem,

$$
\begin{aligned}
\min_{x \in \Re^n} \ & \text{TE}(x) \\
\text{subject to} \quad & \sum_{i=1}^{n} \Lambda(x_i) \leq K \\
& \sum_{i=1}^{n} x_i = 1 \\
& x \geq 0
\end{aligned}
\tag{1}
$$

where $\Lambda(x_i) = 1$ if $x_i \neq 0$ and $\Lambda(x_i) = 0$ otherwise. Here we discuss the problem in terms of the percentage holding $x$, noting that the formulation in terms of the actual units, $y$, is similar. The function $\text{TE}(x)$ measures the tracking error and the cardinality constraint, $\sum_{i=1}^{n} \Lambda(x_i) \leq K$, can be interpreted as enforcing an upper bound on the administration costs (modeled as a linear function of the total number of stocks). Solving this problem is NP-hard and, not surprisingly, all of the existing methods for the tracking error minimization problem subject to a cardinality constraint are heuristic in nature.

To further develop the index tracking problem mathematically, let $y_i$ be the number of units of asset $i$ in the portfolio $y$. To simplify notation, we will describe some measures of tracking error as functions of $x$ and others as functions of $y$. We note that the relationship between $x$ and $y$ at time $t$ is

$$x_i = \frac{S_{it}y_i}{S_t^T y} \tag{2}$$

where $S_{it}$ is the price of stock $i$ at time $t$ and $S_t$ is the vector of time $t$ stock prices.

There are a few different ways of measuring tracking error. Beasley, Meade and Chang [1999] measure the tracking error based on historical stock and index prices as follows:

$$\text{TE}_{\text{BMC}}(x) \stackrel{\text{def}}{=} \frac{1}{T}\left(\sum_{t=1}^{T}|r_t(y) - R_t|^p\right)^{\frac{1}{p}} \tag{3}$$

where $T$ is total number of periods, $R_t = \frac{\ln(I_t)}{\ln(I_{t-1})}$ is the return of the index at the period $[t-1, t]$,

$$r_t(x) \stackrel{\text{def}}{=} \frac{\ln(S_t y)}{\ln(S_{t-1}y)}$$

and $S_t \in \Re^n, I_t$ are the stock prices and index price at $t$ respectively. Note that the tracking error function $\text{TE}_{\text{BMC}}(y)$ is not convex. For our computational results and for the results presented from Beasley et al [1999], $p = 2$ is used.

A similar measure of tracking error, denoted here by $\text{TE}_{\text{SM}}(y)$, is used by Salkin and Meade [1989]; it is obtained by setting $R_t = \frac{I_t - I_{t-1}}{I_{t-1}}$ as the return

of the index during the period $[t-1, t]$ and

$$r_t(x) \stackrel{\text{def}}{=} \frac{S_t y - S_{t-1} y}{S_{t-1} y}$$

The tracking error functions used by Beasley et al [1999] and Salkin and Meade [1989] explicitly penalize any deviation of each period return (daily, weekly, etc) of a tracking portfolio from the return on the index. Furthermore, deviations in each time period (each day or week) are counted equally towards the total tracking error, and these measures are sensitive to the choice of time period. For example, assume that there is a total of two weeks, the tracking portfolio underperforms the index by 1% during the first week, and it outperforms the index by 1% in the next week. Then the two-week return on the tracking portfolio and the index are quite similar, but both $\text{TE}_{\text{BMC}}(y)$ and $\text{TE}_{\text{SM}}(y)$ (with one week per period) can be large. If the time period were doubled, to two weeks, then the two errors would offset each other and $\text{TE}_{\text{BMC}}(y)$ and $\text{TE}_{\text{SM}}(y)$ would be smaller than that corresponding to one week per period. In this respect, both $\text{TE}_{\text{BMC}}(y)$ and $\text{TE}_{\text{SM}}(y)$ depend on the time period selected for calculating returns.

Another frequently used definition of tracking error measures the active risk of a portfolio based on the covariance matrix of the stock returns (see Beckers [1998]); this definition is used in Jansen and Dijk [2002] and we denote it by

$$\text{TE}_{\text{JD}}(x) \stackrel{\text{def}}{=} (x-w)^T Q (x-w) \tag{4}$$

where $w$ denotes the stock weights for the index and $Q$ is the covariance matrix of the stock returns. Here, as before, $x_i$ represents the percentage weight of

asset $i$ in the portfolio $x$ and $w_i$ is similarly defined as the percentage weight of asset $i$ in the index portfolio. Note that the function $\text{TE}_{\text{JD}}(x)$ is convex and is mathematically more appealing. This active risk is a direct function of the extent to which stocks are weighted differently to their weights in the index. This measure of tracking error has a convenient financial interpretation if the historical covariance matrix $Q$ is assumed to be accurate for that of future returns. In this case, if the tracking error $(\text{TE}_{\text{JD}}(x))^{1/2} = 1\%$, then you can expect the return on your tracking portfolio to be within $\pm 1\%$ of the return on the index about $67\%$ of the time in one observation period, and to be within $\pm 2\%$ about $95\%$ of the time.

## Tracking Error Minimization Via Graduated Non-Convexity

Standard optimization software does not apply to the index tracking problem (1) directly since the cardinality constraint function is discontinuous. One possible way of overcoming this difficulty is to consider a sequence of approximations which approach the tracking error minimization problem in the limit. How well this type of method works depends on what the sequence of approximations is and how the sequence of approximations approaches the original tracking error minimization problem.

For simplicity, one may consider (see Jansen and Dijk [2002] for example)

an equivalent form of the tracking error minimization problem for (1):

$$\min_{x \in \Re^n} \left( \mathrm{TE}(x) + \mu \sum_{i=1}^{n} \Lambda(x_i) \right)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \tag{5}$$

$$x \geq 0$$

where $\mu \geq 0$ is a penalty parameter. Here $\mathrm{TE}(x)$ is a smooth function which measures the tracking error for a general portfolio benchmarking problem. By varying $\mu > 0$, solutions of (5) yield optimal tracking portfolios of different number of assets.

To handle the discontinuity introduced by the counting function $\Lambda(x_i)$, Jansen and Dijk [2002] approximate $\Lambda(x_i)$ by $x_i^p$, where $p \geq 0$ is small. Specifically, $p = 0.5$ is used. Thus the following problem is solved

$$\min_{x \in \Re^n} \left( \mathrm{TE}_{\mathrm{JD}}(x) + \mu \sum_{i=1}^{n} x_i^p \right)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \tag{6}$$

$$x \geq 0$$

The motivation behind this method is that $x_i^p$ converges to $\Lambda(x_i)$ as $p$ converges to zero. As mentioned before, one of the difficulties of this approach is that, when $p < 1$, the objective function is not everywhere differentiable and standard optimization software is not guaranteed to yield a minimizer of the problem (6); this is illustrated in the Results section.

We attempt to track the portfolio of the globally minimal tracking error by first computing the minimizer of the tracking error function without the

cardinality constraint, i.e., a solution to the continuous optimization problem

$$\min_{x \in \Re^n} \ \text{TE}(x)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1$$

$$x \geq 0$$

Note that if the tracking error function is continuously differentiable and convex, e.g., $\text{TE}(x) = \text{TE}_{JD}(x)$, a global minimizer can be computed from standard optimization software. Starting from this minimizer, a sequence of approximations $\{\mathcal{P}_k\}_{k=1,2,\dots}$ to the tracking error minimization problem (1) is solved by approximating the counting function $\Lambda(x_i)$ with *continuously differentiable* piecewise quadratic functions with graduated non-convexity; the solution of the approximation problem $\mathcal{P}_{k-1}$ is used as the staring point for the approximation problem $\mathcal{P}_k$.

Next, we motivate and describe our proposed method; we show the connection of the proposed method to the graduated non-convexity technique used in image reconstruction (Blake and Zisserman [1987]) in appendix B.

To motivate our approximations $\{\mathcal{P}_k\}_{k=1,2,\dots}$ to the tracking error problem (1), let us first approximate the discontinuous counting function $\Lambda(z)$ by the following *continuous* function $h_\lambda(z)$

$$h_\lambda(z) = \begin{cases} \lambda z^2 & \text{if } |z| \leq \sqrt{\frac{1}{\lambda}} \ , \\ 1 & \text{otherwise.} \end{cases}$$

where $\lambda > 0$ is a large constant (which is set to $10^8$ in our computations). The function $h_\lambda(z)$ is illustrated in Exhibit 1; it is used in image segmentation to

approximate counting the number of edges in an image.

Now problem (1) can be formulated as a *continuous* but non-differentiable mathematical programming problem.

$$\min_{x \in \Re^n} \left( \text{TE}(x) + \mu \sum_{i=1}^{n} h_\lambda(x_i) \right)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \tag{7}$$

$$x \geq 0$$

In appendix A, we illustrate how this formulation can be derived from a penalty function formulation and a line elimination technique similar to that used in image reconstruction (Blake and Zisserman [1987]).

The above minimization problem (7) is not everywhere differentiable and it has many local minimizers. We consider the following graduated non-convexity method to attempt to track the global minimizier of (7), based on a similar method used in Blake and Zisserman [1987].

We approximate the non-differentiable function $h_\lambda(z)$ by the continuously differentiable function $g_\lambda(z; \rho)$ below:

$$g_\lambda(z; \rho) = \begin{cases} \lambda z^2 & \text{if } |z| \leq q \\ 1 - \frac{\rho}{2}(|z| - r)^2 & \text{if } q \leq |z| < r \\ 1 & \text{otherwise} \end{cases}$$

$$r^2 = (\frac{2}{\rho} + \frac{1}{\lambda}), \quad q = \frac{1}{\lambda r}$$

Here $\rho > 0$ is a parameter. The function $g_\lambda(z; \rho)$ is a piecewise quadratic with a concave quadratic piece for $z \in (q, r)$. Let $\{\rho_k\}$ be a given monotonically

increasing sequence which converges to $+\infty$. As $\rho_k$ increases, the curvature of the quadratic function defining $g_\lambda(z; \rho)$ for $z \in [q_k, r_k]$ becomes more negative, introducing a graduated nonconvexity. In addition, as $\rho_k \to +\infty$, $r_k, q_k$ converge to $\sqrt{\frac{1}{\lambda}}$ and the functions $g_\lambda(z, \rho_k)$ approach $h_\lambda(z)$. Exhibit 2 illustrates how the sequence of approximations $g_\lambda(z; \rho)$ approaches the function $h_\lambda(z)$ as $\rho$ increases.

Substituting $g_\lambda(z; \rho)$ for $h_\lambda(z)$ in (7), the following sequence $\{\mathcal{P}_k\}$ of approximations to the tracking error minimization problem arises:

$$\min_{x \in \Re^n} \left( \mathrm{gnc}_k(x) = \mathrm{TE}(x) + \mu \sum_{i=1}^n g_\lambda(x_i; \rho_k) \right)$$

$$\text{subject to} \quad \sum_{i=1}^n x_i = 1 \tag{8}$$

$$x \geq 0$$

To appreciate why this is a reasonable process to track the global minimizer of the tracking error minimization problem subject to a cardinality constraint, consider a convex tracking error function and suppose that there are no other constraints for simplicity. For a sufficiently small $\rho > 0$, the objective function $\mathrm{gnc}_k(x)$ of (8) remains convex, where

$$\mathrm{gnc}_k(x) \stackrel{\mathrm{def}}{=} \mathrm{TE}(x) + \mu \sum_{i=1}^n g_\lambda(x_i; \rho_k) \tag{9}$$

Thus, for a small $\rho_k$, the solution to $\mathcal{P}_k$ (8) is the unique global minimizer. For sufficiently small $\rho$, Approximations $\mathrm{gnc}_k(x)$, can be regarded as multi-dimensional convex envelopes of the objective function of (5) with respect to all asset subsets, see Exhibit 3. We start with minimizing the tracking

error function; each subsequent approximation introduces increasingly more negative curvature to the objective function $gnc_k(z)$ through $g_\lambda(x; \rho_k)$. The negative curvature interacts with the positive curvature of the tracking error function to ensure that optimal tracking portfolios of subsets of stocks are reachable via minimizing $gnc_k(x)$. The minimizer of $gnc_{k-1}(x)$ is then used as a starting point to compute a minimizer for the subsequent approximation $gnc_k(x)$. As $\rho_k$ converges to $+\infty$, the approximate problems approach the tracking error minimization problem (7).

Exhibit 3 illustrates this process for a one-dimensional function $TE(x) = \frac{1}{2}(x+2)^2$. Without loss of generality, this process is depicted without constraints. In the top-left subplot of Exhibit 3 we see the original non-convex function $TE(x) + \mu h_\lambda(x)$, a convex approximation (corresponding to $\rho_k = 0.001$), and its minimizer. Increasing $\rho$ to 1 we see the next approximation (in this example, still convex) to the original function in the top-right subplot. With the minimizer of the first approximation function as a starting point, the minimizer of the new approximation, which is very close to the global minimizer, is computed. In the bottom two subplots ($\rho_k = 10$ and 1000 respectively ) we see how the approximating functions $gnc_k(x)$ approach the original function as $\rho_k$ increases. From this illustration we see that the proposed process first considers large-scale features of the original function and gradually focuses in on features of a smaller scale.

The proposed graduated non-convexity process starts with minimizing the

tracking error without any limit on the total number of instruments, i.e.,

$$\min_{x \in \Re^n} \mathrm{TE}(x)$$

subject to
$$\sum_{i=1}^{n} x_i = 1 \tag{10}$$

$$x \geq 0$$

When the tracking error function is convex, e.g., $\mathrm{TE_{JD}}$, problem (10) is a convex programming problem and a global minimizer can be computed. Under the assumption of convexity, if the minimizer $x^*$ of (10) satisfies the condition $\sum \Lambda(x_i^*) \leq K$, then our proposed method is guaranteed to yield $x^*$ as the solution.

We have so far described the proposed graduated non-convexity method with respect to the tracking error minimization formulation (5), in which the size of the optimal tracking portfolio is chosen by varying the parameter $\mu$. Typically a portfolio manager explicitly wants to obtain a tracking portfolio with an upper bound $K$ on the number of stocks. Using the formulation (5), it is necessary to experiment with different values of $\mu$ in order to generate a tracking portfolio of the desired number of stocks. To compute a tracking portfolio of the desired size directly, we consider the following exact penalty formulation of the tracking problem (1) with $\sum_{i=1}^{n} h_\lambda(x_i)$ replacing $\sum_{i=1}^{n} \Lambda_i(x_i)$,

$$\min_{x \in \Re^n} \left( \mathrm{TE}(x) + \mu \max(\sum_{i=1}^{n} h_\lambda(x_i) - K, 0) \right)$$

subject to
$$\sum_{i=1}^{n} x_i = 1 \tag{11}$$

$$x \geq 0$$

where $\mu$ is an exact penalty parameter corresponding to the cardinality constraint. For a sufficiently large $\mu$, e.g., $\mu = 100$ in all our subsequent computations, the minimizer of Tracking error minimization formulation (11) yields a tracking portfolio of no more than $K$ stocks.

The graduated non-convexity approximation problem $\mathcal{P}_k$ can similarly be generated:

$$\min_{x \in \Re^n} \left( \text{TE}(x) + \mu \max(\sum_{i=1}^n g_\lambda(x_i; \rho_k) - K, 0) \right)$$

$$\text{subject to} \quad \sum_{i=1}^n x_i = 1 \tag{12}$$

$$x \geq 0$$

Intuitively, when $\sum_{i=1}^n g_\lambda(x_i; \rho_k) > K$, decreasing the objective function of (12) leads to the decrease of the objective function $\text{gnc}_k(x)$ of (8). Moreover, for initial approximate problems $\mathcal{P}_k$ with small $\rho_k$, $\sum_{i=1}^n g_\lambda(x_i; \rho_k) < K$. As more negative curvature is gradually introduced into the objective function by $g_\lambda(x; \rho_k)$, the function $\text{TE}(x) + \sum_{i=1}^n g_\lambda(x_i; \rho_k)$ is gradually decreased as described above to track the global minimizer. The graduated non-convexity process is terminated when, for all $i$, either $(x_k)_i < q$ (the $i$th stock is not in the tracking portfolio) or $(x_k)_i > r$ (the $i$th stock is in the tracking portfolio). This computational procedure is described in Exhibit 4.

It may seem, from the description in Exhibit 4, the proposed graduated non-convexity method requires solving an excessive number of approximation problems $\mathcal{P}_k$. This is, in fact, not the case. Firstly, we note that, for many updates of parameter $\rho_k$, the minimizer of the problem $\mathcal{P}_{k-1}$ remains the min-

15

imizer of $\mathcal{P}_k$. Secondly, the minimizer of $\mathcal{P}_{k-1}$ is a good starting point of $\mathcal{P}_k$ even when it is not a minimizer for $\mathcal{P}_k$; thus a small number of iterations are typically required to reach the minimizer of $\mathcal{P}_k$.

# Computational Results

To illustrate, we present numerical results using several publicly available historical data sets for equity index tracking. We compare the quality of the solutions produced by the proposed graduated non-convexity method to those of the exact optimal solutions (computed via brute force) when $K$ is very small. We also compare the graduated non-convexity method to the population heuristic method from Beasley et. al [1999] and the method from Jansen and Dijk [2002], which uses a continuous but not differentiable function to approximate the counting function $\sum_{i=1}^{n} \Lambda(x_i)$.

To test the proposed graduated non-convexity (GNC) algorithm described in Exhibit 4, we use the data sets made publicly available by Beasley, Meade and Chang [1999]. These data sets consist of weekly price observations on stocks in five indices from different world markets during the period of March 1992 to September 1997. Both stock price data and index price data are included in the data sets for the market indices Hang Seng, DAX 100, FTSE 100, S&P 100, and Nikkei 225. In each case, stocks were dropped from the data set if they were not present during the entire period of observation. Exhibit 6 describes the data sets where the number of stocks, $n$, and the number of

16

weeks of data, $m$, are given for each set of historical price data.

First, we present computational results using the quadratic tracking error function $\text{TE}_{\text{JD}}(x)$. Since index weights are not given in the data sets, we artificially generate the index from equally weighting the stocks, i.e., $w_i = 1/n, \ 1 \leq i \leq n$.

Exhibit 7 presents (annualized) optimal quadratic tracking errors $(TE_{\text{JD}}(x))^{\frac{1}{2}}$ obtained using the graduated non-convexity method (GNC) with $K = 25$. The annualized tracking errors are given for two different GNC solutions corresponding to two different sequences $\{\rho_k\}$ that define the optimization subproblems. In particular the $\rho$ update rules $\rho_{k+1} = 1.2\rho_k$ and $\rho_{k+1} = 2\rho_k$ are compared. Because the GNC method solves a different sequence of subproblems in each case, the solutions achieved from these two rules are different. In Exhibit 7 we also give the number $N$ of optimization subproblems which are solved for each data set and each update rule. In most cases, we observe that the $\rho_{k+1} = 1.2\rho_k$ update rule (which imposes the cardinality consideration "more gradually") typically yields a solution with slightly smaller tracking error. However, in two cases we see that the "less gradual" update rule $\rho_{k+1} = 2\rho_k$ gives a slightly better tracking portfolio. This illustrates a sensitivity of the quality of the solution on the updating rule for the parameter $\rho$. This is not surprising since there are many local minimizers corresponding to tracking portfolios with different stocks, and they can achieve similar tracking errors. Theoretically one would like to update $\rho_k$ as gradually as possible; but

17

this needs to be balanced with the computational time that can be afforded. The update rule $\rho_{k+1} = 1.2\rho_k$ seems to work best for a range of $K$ and different measures of tracking error and the GNC results in this paper, except for the results in Exhibit 7, all use this rule.

For both rules in Exhibit 7 we see that, using no more than 25 stocks, a annualized tracking error of approximately 2% is obtained for all the data sets except Hang Seng (for which a smaller tracking error of 1.27% is achieved). As mentioned before, we can interpret the results as the standard deviation of the difference between the return on the tracking portfolio and the return on the index. For example, consider the computed 25 stock tracking portfolio on the 225 stock Nikkei index and let $R_I$ denote the annual percent return of the index. Then we expect the tracking portfolio to have annual return $R_I \pm 2\%$ about 67 percent of the time.

Although the graduated non-convexity method gives solution portfolios with reasonably small tracking errors, we would like to assess how close the tracking error of the portfolio computed by GNC is to the global minimum of the cardinality constrained tracking error minimization problem. However, a brute-force computation of the optimal solution for $K = 25$ using the S&P 100 data set, for example, would require solving over $1.3 \times 10^{23}$ (small) optimization sub-problems. While this is impractical, we can compare the results of our GNC algorithm to the optimal solution for very small K. Exhibit 8 compares, for $K = 3$, the tracking error achieved by GNC with the global minimum

tracking error computed by a brute-force method. It can be observed that the GNC method produces nearly optimal tracking portfolio for these tests: the accuracy of the tracking error achieved by the GNC method, compared with the global minimum tracking error, ranges from 86% to 95%.

In the rest of this section, we compare the graduated non-convexity method with the method used by Jansen and Dijk [2002] and the method used by Beasley, Meade and Chang [1999].

The method proposed by Jansen and Dijk [2002] is similar to our proposed GNC algorithm in that both methods attempt to solve the index tracking problem by approximating the discontinuous counting function by continuous functions. However, the two methods are fundamentally different. Jansen and Dijk [2002] approximate the discontinuous counting function using a non-differentiable function while we approximate the counting function by a continuously differentiable function. In addition, the proposed GNC method includes a graduated non-convexity process of solving a sequence of continuously differentiable optimization problems (which can be solved using a constrained minimization approach or a penalty function approach) to track the global minimizer.

In Jansen and Dijk [2002] the following minimization problem is first solved for a fixed small $p$ ($p = 0.5$),

$$\min_{x \in \Re^n} \quad \left( \mathrm{TE}_{\mathrm{JD}}(x) + \mu \sum_{i=1}^{n} x_i^p \right)$$
$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \tag{13}$$

19

$$x \geq 0$$

From a computed solution to (13), a set $I = \{i : x_i > \epsilon\}$ is identified for some small threshold value $\epsilon$ and a smaller quadratic tracking portfolio programming problem is solved with the $x_i$ fixed at zero for $i \notin I$, i.e.,

$$\min_{x \in \Re^n} \ \text{TE}_{\text{JD}}(x)$$
$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1$$
$$\forall i \notin I, x_i = 0 \qquad (14)$$
$$x \geq 0$$

One of the difficulties due to the use of the function $z^p$ to approximate the counting function $\Lambda(z)$ is that $z^p$ is not differentiable when $z = 0$ and $p < 1$. If a standard optimization method for a continuously differentiable optimization problem is used to solve (13), convergence is not guaranteed. In addition, solving (13) presents numerical difficulties, especially for larger values of $\mu$ which typically are required for tracking portfolios with small number of assets. In particular, the reduced Hessian matrix can be arbitrarily ill-conditioned near a solution to (13). Larger values of $\mu$ exacerbate this problem by magnifying the ill conditioning from the function $x_i^p$. A more detailed discussion of this ill-conditioning is given in the appendix. The lack of a theoretical convergence property and ill-conditioning issues means that the computational software may yield non-minimizers and, depending on the starting point, optimization software can terminate at different approximations. For example, with

$\mu = 0.005$, the annualized tracking error $\left(\mathrm{TE}_{\mathrm{JD}}(x)\right)^{\frac{1}{2}}$ ranges from 0.0228 with 28 stocks to 0.0274 with 23 stocks when 10 different starting points, randomly perturbed from the index portfolio, are used to solve for 10 solutions (8 were distinct). Perturbations of about 20 percent were randomly chosen to generate the 10 starting portfolios.

In spite of the numerical difficulties in the method of Jansen and Dijk [2002], Exhibit 9 compares the average tracking errors from the method of Jansen and Dijk [2002] (using different starting points) for several different choices of $\mu$ and the proposed GNC method on the S&P 100 data set. For each $\mu$, the reported tracking error and the number of stocks for Jansen and Dijk's method are the averages of 15 solutions of (14) using 15 different randomly generated initial portfolios and $p = 0.5$. For comparison, for each $\mu$, the result from a single invocation of the GNC method with $K$ closest to the corresponding average portfolio size is reported in the right columns.

From Exhibit 9 we see that the results from the GNC method are 8 to 15 percent better than the average results from the solutions of the method described by Jansen and Dijk [2002]. In our investigation we were unable to reliably use the Jansen and Dijk method to solve the index tracking problem for a portfolio smaller than about 10 stocks (using the S&P data set). The GNC method, on the other hand, handles small portfolio selection well (as in Exhibit 8 where we use $K = 3$.)

Finally, we compare the proposed GNC algorithm to the population heuris-

tic algorithm by Beasley, Meade and Chang [1999]. To do this, we use $K = 10$ and we choose the tracking error measure $\mathrm{TE_{SM}}(x)$ for the GNC computations. This is slightly different from the tracking error $\mathrm{TE_{BMC}}(x_*)$ but we report the tracking error measure $\mathrm{TE_{BMC}}(x_*)$ corresponding to our computed solution for comparison with the results in Beasley et. al. [1999]. In Exhibit 10 we compare the GNC results described above with the results from Beasley, Meade, and Chang with transaction cost limit $\gamma = 0.01$. In contrast to results using quadratic tracking error, here actual index price data, instead of the index weights, is used for these calculations. Because of changes in the composition of the index over the period of observation and stocks which were dropped from the data sets due to missing observations, there is no portfolio with a tracking error of zero as measured by $\mathrm{TE_{BMC}}$. To give some idea of the smallest possible tracking error, we include in Exhibit 10 the approximate solution to the tracking error problem with no cardinality constraint, which is computed in Step 0 of the algorithm in Exhibit 4. Note that all solutions from Beasley et. al. [1999] are 10-stock portfolios.

From the results in Exhibit 10 we see that the GNC tracking error results are 11% to 44% smaller than the BMC results. Of note, however, is that the BMC results satisfy an additional constraint. Specifically, the BMC results satisfy a transaction cost limit constraint which effectively permits turnover of only half of the value of the initial position in 10 stocks. We did not impose this constraint on our method, which lends our method an advantage in the results

in Exhibit 10; nonetheless comparison to the optimal unconstrained tracking error (no cardinality constraint) suggests that the GNC method produced reasonably good tracking portfolios. We also note that $\text{TE}_{\text{BMC}}(x)$ is a non-convex function, so the GNC method we describe no longer starts with a convex approximation (the tracking error is non-convex in this case). Instead, we start with the non-convex function $\text{TE}_{\text{BMC}}(x)$ and move towards the (also non-convex) function $\text{TE}_{\text{BMC}}(x) + \mu \sum_{i=1}^{n} h_\lambda(x_i)$. It is interesting to see that the GNC method is successful in sequentially approximating the cardinality constraint even when the tracking error function is not convex.

## Concluding Remarks

The problem of tracking error minimization with a constraint on the total number of assets in the tracking portfolio is an important problem for both passive as well as dynamic fund managers. Finding the optimal tracking portfolio with a fixed number of stocks K is an NP-hard problem and heuristic approaches are common (Beasley et. all [1999], Jansen and Dijk [2002]). We propose a graduated non-convexity method by approximating the discontinuous counting functions using a sequence of continuously differentiable piecewise quadratic functions with increasingly more negative curvature. When the tracking error is measured with a convex function, this graduated non-convexity (GNC) method starts with the optimal tracking portfolio without considering the cardinality constraint and gradually moves towards a solution

satisfying the constraint on the total number of assets. This is more appealing from a theoretical perspective than a purely heuristic approach (as in Beasley et. al. [1999]) or solving a single approximation to the cardinality-constrained problem (as in Jansen and Dijk [2002]).

In addition to being mathematically more appealing, we have illustrated that the GNC method gives good computational results for the index tracking problem. When compared with results from Jansen and Dijk [2002] the GNC method gives results that are 8% to 15% better on average. For small $K$ where an optimal solution can be found by exhaustion (Exhibit 8), the GNC method solution is within 4 to 16 percent of the optimal solution. In the case where the tracking error is a non-convex function, we have seen that the GNC method is able to gradually impose the cardinality constraint to arrive at a solution that is an improvement over that obtained by a heuristic approach (Beasley et. al. [1999]).

# References

[1] Beasley, J. B., N. Meade, and T. J. Chang. Index tracking. Technical report, Imperial College, 1999. ( http://mscmga.ms.ic.ac.uk/jeb/orlib/indtrackinfo.html).

[2] Beckers, S. A survey of risk measurement theory and practice. In C. Alexander, editor, *Risk Managment and Analysis*, pp. 39-59. John Wiley

& Sons Ltd, 1998.

[3] Blake, A. , and A. Zisserman. *Visual Reconstruction*. Cambridge, 1987.

[4] Jansen, R. , and R. V. Dijk. Optimal benchmark tracking. *The Journal of Portfolio Management*, Winter 2002, pp. 33-39.

[5] Meade, N. , and G. R. Salkin. Index funds - construction and performance measurement. *The Journal of the Operational Research Society*, 40 (1989), pp. 871-879.

# A    Line Elimination

We now show that the approximation problem (7) to the original tracking error minimization problem (5) can be derived using a penalty function technique for the constraint and an integer variable elimination technique (similar to the line elimination used for image segmentation in Blake and Zisserman [1987]).

From the index tracking problem (5), we first introduce the equality constraint $x_i(1 - l_i) = 0$ to handle the discontinuous function $\Lambda(x_i)$ and formulate the problem (5) equivalently as a mixed integer programming problem

$$\min_{x \in \Re^n, l_i \in \{0,1\}} \left( \text{TE}(x) + \mu \sum_{i=1}^{n} l_i \right)$$

$$\text{subject to} \quad x_i(1 - l_i) = 0, \quad i = 1, \cdots, n \tag{15}$$

$$\sum_{i=1}^{n} x_i = 1$$

$$x \geq 0.$$

To handle the nonlinear constraint $x_i(1 - l_i) = 0$, we consider a quadratic penalty function formulation of (15):

$$\min_{x \in \Re^n} \left( \text{TE}(x) + \mu \min_{l_i \in \{0,1\}} \left( \sum_{i=1}^{n} l_i + \lambda \sum_{i=1}^{n} x_i^2(1 - l_i) \right) \right)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \tag{16}$$

$$x \geq 0$$

where $\lambda > 0$ is a large penalty parameter associated with the quadratic penalty function for $x_i(1 - l_i) = 0$. Note that $1 - l_i \geq 0$.

26

Similar to the line elimination technique in Blake and Zisserman [1987], we explicitly solve

$$\min_{l_i \in \{0,1\}} \left( \sum_{i=1}^{n} l_i + \lambda \sum_{i=1}^{n} x_i^2 (1 - l_i) \right)$$

to eliminate the (line) integer variable $l_i$ and obtain

$$h_\lambda(x_i) = \min_{l_i \in \{0,1\}} \left( l_i + \lambda x_i^2 (1 - l_i) \right)$$

where

$$h_\lambda(z) = \begin{cases} \lambda z^2 & \text{if } |z| \leq \sqrt{\frac{1}{\lambda}} \, , \\ 1 & \text{otherwise.} \end{cases}$$

The line elimination idea and the function $h_\lambda(z)$ are illustrated in Exhibit 5.

Thus the problem (16) is now formulated as a *continuous* but nondifferentiable programming problem.

$$\min_{x \in \Re^n} \left( \text{TE}(x) + \mu \sum_{i=1}^{n} h_\lambda(x_i) \right)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = 1 \qquad (17)$$

$$x \geq 0$$

Note that we can regard $h_\lambda(x)$ as an approximation to $\Lambda(x)$ for a large $\lambda > 0$.

# B  Connections to Image Analysis

The proposed GNC method for tracking error minimization subject to a constraint on the total number of assets is similar to the GNC method for image

27

reconstruction (see Blake and Zisserman [1987]) to generate an image which is faithful to the original noisy image and has invariance properties such as optical blurring and noise. We show here how the tracking error minimization problem is related to the image reconstruction problem. To illustrate, we consider the quadratic tracking error and assume that one can hold both long positions and short positions.

In Appendix A, we have shown that the tracking error minimization problem (15), using a penalty approach, can be approximated by

$$\min_{x \in \Re^n} \left( (x-w)^T Q(x-w) + \mu \min_{l_i \in \{0,1\}} \left( \sum_{i=1}^n l_i + \lambda \sum_{i=1}^n x_i^2 (1-l_i) \right) \right)$$

subject to $\quad \sum_{i=1}^n x_i = 1 \hfill (18)$

for a positive definite matrix $Q$ (so the problem is convex if the cardinality constraint is removed), a vector of data $w$, and a subset $I \subseteq \{1, 2, \ldots, n\}$. For our tracking error minimization problem, $Q$ is the covariance matrix, $w$ represents the percent composition of the index, $K$ is the upper bound on the number of stocks in our tracking portfolio, and $I = \{1, 2, \ldots, n\}$. We will show that this optimization problem is a generalization of a problem arising naturally in the study of image analysis. Particularly, this problem generalizes the problem of solving for the state of a weak elastic string (as in Blake and Zisserman [1987]) which is related to the problem of detecting edges in images.

Rename the two penalty parameters $\mu$ and $\mu\lambda$ to $\alpha$ and $\lambda$ in (18) to get

$$\min_{x \in \Re^n} \left( (x-w)^T Q(x-w) + \min_{l_i \in \{0,1\}} \left( \alpha \sum_{i=1}^n l_i + \lambda \sum_{i=1}^n x_i^2 (1-l_i) \right) \right)$$

subject to $\qquad \displaystyle\sum_{x \in \Re^n} x_i = 1$ $\hfill$ (19)

The weak elastic string problem can be formulated as follows. Consider the problem of finding a piecewise smooth function $u(x)$ that best fits some data $d(x)$. One approach is to look for a solution $u(x)$ that represents a weak elastic string. That is, an elastic string (so there is a resistance to too much "bending") that may have a number of breaks (step discontinuities) in it. In computing a solution, breaks in the string are penalized (otherwise, after discretizing the problem, you could simply take $u(x_i) = d(x_i)$, with a break at each node). The total energy of the weak elastic string is modeled as the sum of three energy components, one component measuring the faithfulness of $u(x)$ to the data $d(x)$, one component measuring the elastic energy of the string (this component prohibits excessive bending), and a final component penalizing breaks in the string,. If the problem is discretized over a grid $\{u_i : i = 1, 2, ..., n\}$ then the energy function is

$$\sum_{i=1}^{n} (u_i - d_i)^2 + \lambda \sum_{i=1}^{n-1} (u_{i+1} - u_i)^2 (1 - l_i) + \alpha \sum_{i=1}^{n-1} l_i \qquad (20)$$

with line variables $l_i \in \{0, 1\}$ such that there is a break in the string in the interval $[x_{i-1}, x_i]$ if $l_i = 1$ and no break if $l_i = 0$. Minimizing the energy function (20) yields a solution to the edge detection problem in one-dimensional image analysis where discontinuities in the solution $u$ correspond to edges in an image. We wish to show that this problem fits into the more general framework of tracking error minimization problem (18).

Consider the tracking error minimization problem (19): introduce change of variables $x_i = u_{i+1} - u_i$ for $i = 1, 2, \ldots, n-1$ and $x_n = -u_n$. The equality $\sum_{i=1}^{n} x_i = 1$ becomes $u_1 = -1$. Hence the tracking error minimization problem (19) becomes the unconstrained minimization problem

$$\min_{u \in \Re^n} \left( (x-w)^T Q(x-w) + \min_{l_i \in \{0,1\}} \left( \alpha \sum_{i=1}^{n} l_i + \lambda \sum_{i=1}^{n-1} (u_{i+1} - u_i)^2 (1 - l_i) \right) + \lambda u_n^2 (1 - l_n) \right) \tag{21}$$

Now consider the $n \times n$ upper triangular matrix $V$ with every element on or above the diagonal equal to $-1$. That is,

$$V = - \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{22}$$

$V$ is clearly nonsingular, so let $w$ solve $Vw = d$ where $d$ is the data vector in the weak elastic string problem. Now, notice that $Vx = u$. That is,

$$- \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} u_2 - u_1 \\ u_3 - u_2 \\ \vdots \\ u_n - u_{n-1} \\ -u_n \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \\ u_n \end{pmatrix} \tag{23}$$

30

So with $Q = V^T V$ and $w$ satisfying $Vw = d$, (21) becomes

$$\min_{u \in \Re^n} \left( (u-d)^T (u-d) + \min_{l_i \in \{0,1\}} \left( \alpha \sum_{i=1}^{n} l_i + \lambda \sum_{i=1}^{n-1} (u_{i+1} - u_i)^2 (1 - l_i) + \lambda u_n^2 (1 - l_n)) \right) \right) (24)$$

which differs from (20) only by the addition of the term $\lambda u_n^2 (1 - l_n)$ and $u_1 = 1$.

In Blake and Zisserman [1987], it has been shown that, for the isolated discontinuity image analysis problem, GNC solves the global minimization problem (20) correctly. It would be interesting to investigate whether (24) can be solved correctly by GNC under similar conditions; an affirmative answer would show that, there exists some tracking error minimization problem (18) for which the GNC algorithm yields a global minimizer.

## C  Ill-conditioning in problem (13)

The method proposed in Jansen and Dijk [2002] is to solve the constrained optimization problem

$$\min_{x \in \Re^n} \left( f(x) \stackrel{\text{def}}{=} (x-w)^T Q(x-w) + \mu \sum_{i=1}^{N} x_i^p \right) \qquad (25)$$

$$\text{subject to} \quad \sum_{i=1}^{N} x_i = 1 \qquad (26)$$

$$x \geq 0 \qquad (27)$$

for a small value of $p$ and a positive penalty parameter $\mu$ ($p = 0.5$ is used in Jansen and Dijk [2002]) and in our implementation of their method, described

above.) The method is designed to give a solution where many of the variables $x_i$ are close to zero, thus selecting a small tracking portfolio. When any $x_i$ is close to zero, however, the gradient and Hessian of the objective function $f(x)$ in (25) have elements of very large magnitude. In particular

$$\frac{\partial f}{\partial x_i} = 2Q_i^T(x - w) + \mu p x_i^{p-1} \tag{28}$$

where $Q_i$ is column $i$ of the covariance matrix. So for $p < 1$, $\frac{\partial f}{\partial x_i} \to \infty$ as $x_i \to 0$. Further,

$$\frac{\partial^2 f}{\partial x_i^2} = 2Q_{ii} + \mu p(p - 1)x_i^{p-2} \tag{29}$$

so for $p < 1$, $\frac{\partial^2 f}{\partial x_i^2} \to -\infty$ as $x_i \to 0$. We now examine the condition number of the reduced Hessian as elements of $x$ approach to zero.

Let $H$ be the Hessian of the objective function $f(x)$ in (25) at some $x \geq 0$. Let the set $\Omega$ be the set of all $i$ such that $x_i = 0$, i.e., $\Omega$ is the set of all variables at which the constraint $x \geq 0$ (27) is active. Write $\bar{H}$ for the matrix $H$ with row and column $i$ removed for all $i \in \Omega$. Let $H_r = Z^T \bar{H} Z$ where $Z$ is a matrix whose columns form a basis for the null space of the linear constraint matrix (26) for the non-active variables. Then $H_r$ is the reduced Hessian with respect to the constraints (26) and (27). To simplify notation, set $x = \bar{x}$, and re-number as $1, ..., n$ the remaining variables in $x$ (and the remaining rows and columns in $\bar{H}$). Denote the condition number of a matrix $A$ as $\kappa(A) = \|A\| \|A^{-1}\|$ where $\|A\|$ is the 1-norm of $A$.

With these definitions, if $0 < p < 1$ and there exist integers $j$ and $k$ and real $\epsilon > 0$ such that $x_j \geq \epsilon$ and $x_k \geq \epsilon$ as some $x_i \to 0$ then $\kappa(H_r) \to \infty$. In other words, as any solution is approached where one stock holding is becoming nearly zero and at least two stock holdings are bounded away from zero the reduced Hessian becomes increasingly ill-conditioned. Note that this is the situation that we expect to encounter in a financially interesting solution to the tracking error minimization problem.

To see that $\kappa(H_r) \to \infty$ under these conditions, suppose that we have $x_j \geq \epsilon$ and $x_k \geq \epsilon$ as $x_q \to 0$. First note the equivalent definitions of the matrix 1-norm:

$$\|H\| = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |H_{ij}| = \max_{\|x\|_1 = 1} \|Hx\|_1 \tag{30}$$

Note also that the Hessian matrix $H$ can be written as $H = Q + D$ (and $\bar{H} = \bar{Q} + \bar{D}$) where $Q$ is the covariance matrix and $D$ is the diagonal matrix

$$D = -\mu p(1-p) \begin{pmatrix} x_1^{p-2} & 0 & 0 & \ldots & 0 \\ 0 & x_2^{p-2} & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & x_{N-1}^{p-2} & 0 \\ 0 & \ldots & 0 & 0 & x_N^{p-2} \end{pmatrix} \tag{31}$$

Note that $p - 2 < 0$ and that $Q$ is a matrix of constants, i.e., $Q$ is independent

33

of $x$. Now let $e_i$ denote column $i$ of the identity matrix and choose $e_i - e_k$ for $i = 1, 2, ..., k-1, k+1, k+2, ..., n$ as $n - 1$ vectors which form a basis for the null space of the linear constraint matrix. It is straightforward to verify that these vectors are linearly independent and span the null space. This choice of basis will make it easy to see that $\kappa(H_r) \to \infty$.

So with

$$
Z = \begin{pmatrix}
1 & & & & & & \\
 & 1 & & & & & \\
 & & \ddots & & & & \\
 & & & 1 & & & \\
-1 & -1 & \cdots & -1 & -1 & \cdots & -1 & -1 \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1 & \\
 & & & & & & & 1
\end{pmatrix} \text{row k}
$$

(32)

let $Q_r = Z^T \bar{Q} Z$ and note that $Q_r$ is a constant matrix. The matrix $D_r = Z^T \bar{D} Z$ has diagonal elements

$$-\mu p(1-p) \begin{pmatrix} x_1^{p-2} + x_k^{p-2} & & & & & \\ & \ddots & & & & \\ & & x_{k-1}^{p-2} + x_k^{p-2} & & & \\ & & & x_{k+1}^{p-2} + x_k^{p-2} & & \\ & & & & \ddots & \\ & & & & & x_n^{p-2} + x_k^{p-2} \end{pmatrix}$$

$$(33)$$

and each off-diagonal element has value $-\mu p(1-p)x_k^{p-2}$.

Now because $x_q \to 0$, consider column $q$ of $H_r = Q_r + D_r$. One element

in this column is $Q_r(q,q) + -\mu p(1-p)(x_q^{p-2} + x_k^{p-2})$ where $Q_r(q,q)$ denotes

entry $(q,q)$ of $Q_r$. Observe that $x_q^{p-2} \to \infty$ as $x_q \to 0$ because $p < 1$. So

$\sum_{i=1}^n |H_r(i,q))| \to \infty$ as $x_q \to 0$ and hence $\|H_r\| \to \infty$ as $x_q \to 0$. To prove

$\kappa(H_r) \to \infty$ as $x_q \to 0$ it suffices to show that $\|H_r^{-1}\|$ is bounded away from

zero.

To see this, first note that $\|H_r^{-1}\| \geq \|H_r^{-1}e_j\|$ by (30) and $\sum_{i=1}^{n-1} H_r(j,i)H_r^{-1}(i,j) =$

1. The elements $H_r(j,i)$ (the elements in row $j$ of $H_r$) are bounded in abso-

lute value because the elements in row $j$ of $Q_r$ are constant, and the ele-

ments in row $j$ of $D_r$ are $-\mu p(1-p)x_k^{p-2}$ (all elements except $D_r(j,j)$) and

$-\mu p(1-p)(x_k^{p-2} + x_j^{p-2})$ ($D_r(j,j)$). But by assumption $x_k \geq \epsilon$ and $x_j \geq \epsilon$, so

the elements of row $j$ of $D_r$ are bounded in absolute value by $2\mu p(1-p)\epsilon^{p-2}$.

So let $z \in \Re$ be an upper bound for the absolute value of elements in row $j$ of

$H_r = Q_r + D_r$. Now we must have $|H_r^{-1}(i,j)| \geq \frac{1}{(n-1)z}$ for some $i$. Otherwise, if for all $i$ $|H_r^{-1}(i,j)| < \frac{1}{(n-1)z}$ we have

$$1 = \sum_{i=1}^{n-1} H_r(j,i) H_r^{-1}(i,j) \leq \sum_{i=1}^{n-1} z|H_r^{-1}(i,j)| < \sum_{i=1}^{n-1} z\frac{1}{(n-1)z} = 1 \qquad (34)$$

which is a contradiction. So for some $i$, $|H_r^{-1}(i,j)| \geq \frac{1}{(n-1)z}$ and

$$\|H_r^{-1}\|_1 \geq \|H_r^{-1} e_j\|_1 \geq \frac{1}{(n-1)z}. \qquad (35)$$

So as $x_q \to 0$ we have $\|H_r\| \to \infty$ and $\|H_r^{-1}\| \geq \frac{1}{(n-1)z}$, hence $\kappa(H_r) \to \infty$.
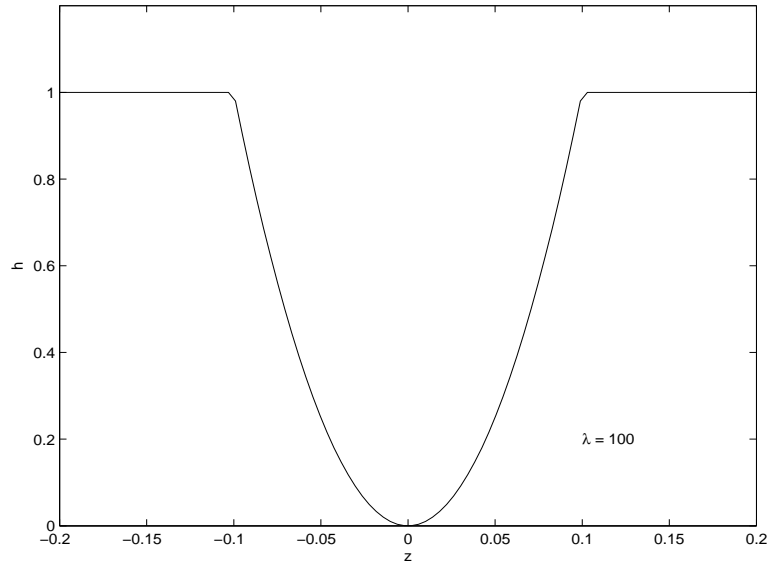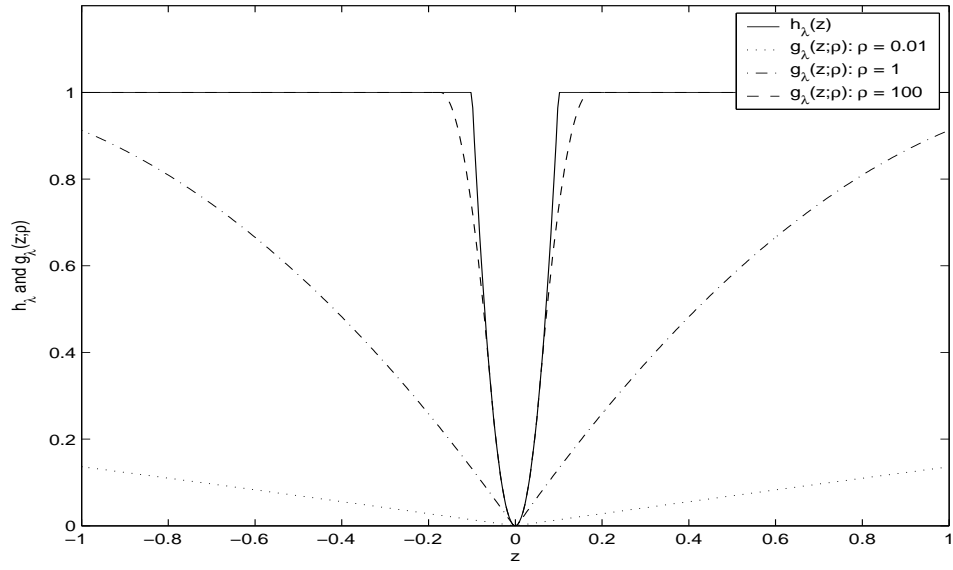
Exhibit 1: Function $h_\lambda(z)$: $\lambda = 100$

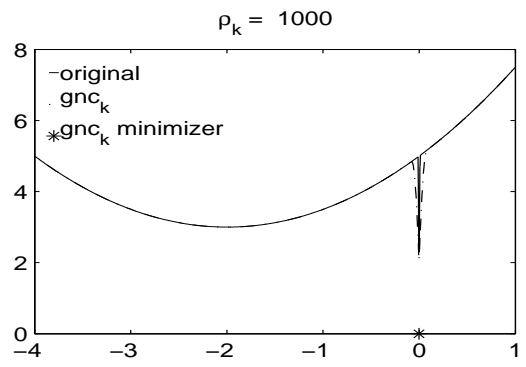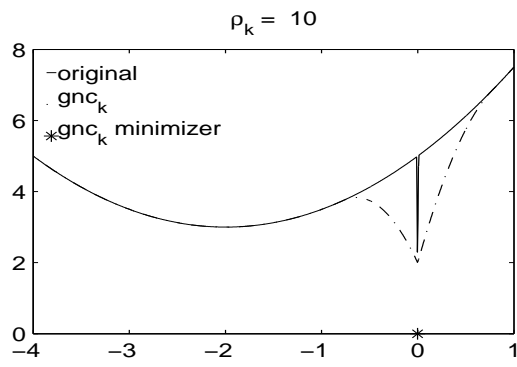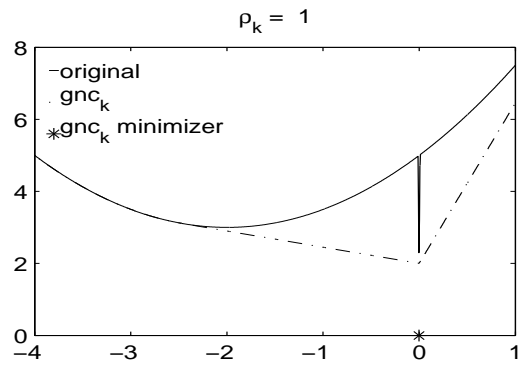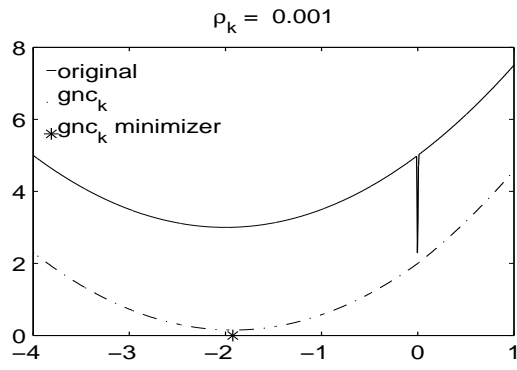Exhibit 2: Graduated Non-convexity Approximations

Exhibit 3: Tracking the Gloabl Minimizer: Graduated Non-convexity

Approximations

**Algorithm.** Let $\lambda > 0$ be a large constant and $\{\rho_k\}$ be a monotonically increasing sequence which converges to $+\infty$.

**Step 0** Compute a minimizer to the tracking error minimization problem (10) without cardinality constraint. Let $k = 1$.

**Step 1** Compute a solution to (12), the problem $\mathcal{P}_k$, using the solution of the approximation at $\mathcal{P}_{k-1}$ as a starting point

**Step 2** If, for all $i$, either $(x_k)_i \leq q_k$ or $(x_k)_i \geq r_k$, terminate. Otherwise, $k \leftarrow k + 1$ and go to Step 1.

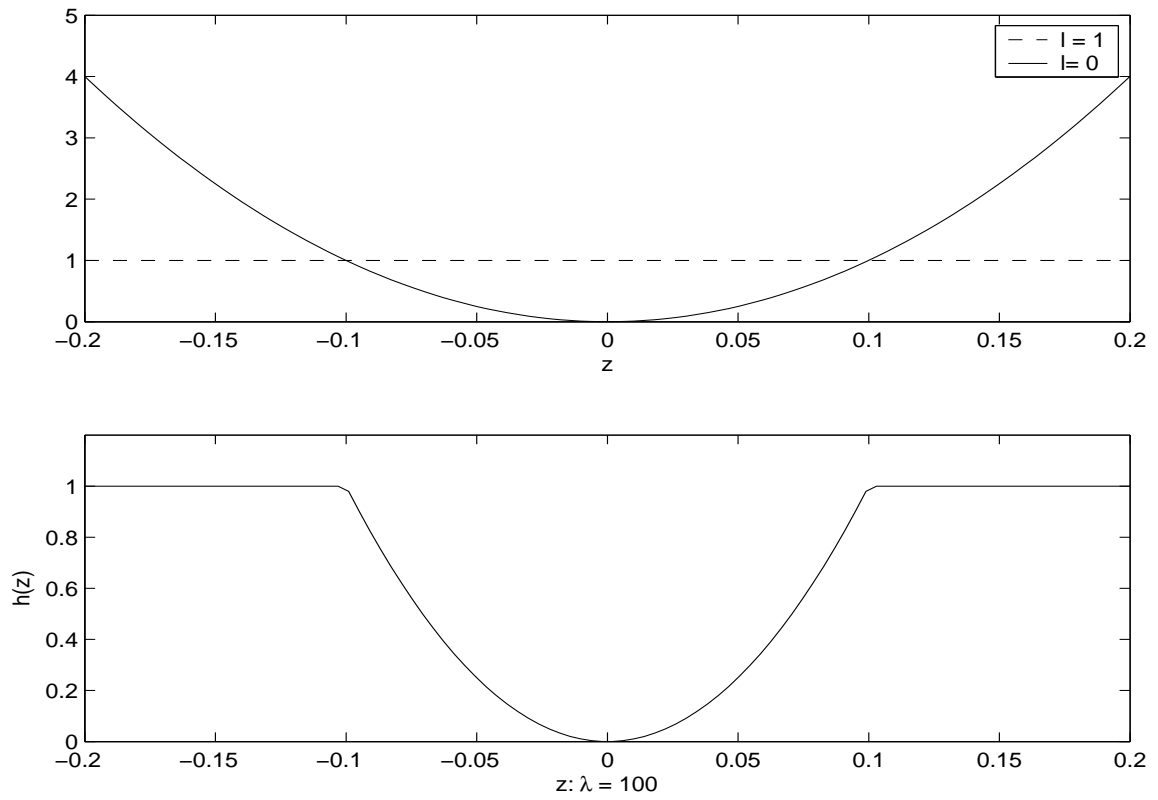Exhibit 4: A Graduated Non-Convexity Method for Index Tracking Problem

Exhibit 5: Line Elimination

| Data set | n | m |
|---|---|---|
| Hang Seng | 31 | 291 |
| DAX 100 | 85 | 291 |
| FTSE 100 | 89 | 291 |
| S&P 100 | 98 | 291 |
| Nikkei 225 | 225 | 291 |

Exhibit 6: Data Sets

| Data set | n | $\rho_{k+1} = 1.2\rho_k$ | N | $\rho_{k+1} = 2\rho_k$ | N |
|----------|-----|------|-----|------|-----|
| Hang Seng | 31 | .0127 | 71 | .0140 | 19 |
| DAX 100 | 85 | .0225 | 71 | .0251 | 19 |
| FTSE 100 | 89 | .0210 | 71 | .0205 | 19 |
| S&P 100 | 98 | .0219 | 71 | .0233 | 19 |
| Nikkei 225 | 225 | .0206 | 72 | .0198 | 20 |

Exhibit 7: Annualized Optimal Tracking Errors Using GNC with $K = 25$

| K=3 | | | |
|---|---|---|---|
| Data set | n | GNC $(\text{TE}_{\text{JD}})^{1/2}$ | optimal $(\text{TE}_{\text{JD}})^{1/2}$ |
| Hang Seng | 31 | .0961 | .0869 |
| DAX 100 | 85 | .0850 | .0732 |
| FTSE 100 | 89 | .0924 | .0825 |
| S&P 100 | 98 | .0891 | .0854 |
| Nikkei 225 | 225 | .0881 | .0810 |

Exhibit 8: Comparing Tracking Errors using GNC with the Global Minimum

Tracking Errors

| $\mu$ | Avg. # of assets | Avg $(\text{TE}_{\text{JD}}(x))^{1/2}$ | GNC portfolio size | GNC $(\text{TE}_{\text{JD}}(x))^{1/2}$ |
|-------|------------------|----------------------------------------|--------------------|-----------------------------------------|
| .02 | 9.64 | .0504 | 10 | .0437 |
| .0105 | 15.08 | .0373 | 15 | .0344 |
| .005 | 25.29 | .0245 | 25 | .0218 |

Exhibit 9: Comparison of Annualized Tracking Error: average from (13) vs.

GNC

| Data set | n | GNC unconstrained | GNC ($k = 10$) | BMC ($K = 10$) |
|---|---|---|---|---|
| Hang Seng | 31 | $9.94 \times 10^{-5}$ | $2.37 \times 10^{-4}$ | $4.21 \times 10^{-4}$ |
| DAX | 85 | $2.91 \times 10^{-4}$ | $3.79 \times 10^{-4}$ | $4.28 \times 10^{-4}$ |
| FTSE | 89 | $6.31 \times 10^{-5}$ | $3.90 \times 10^{-4}$ | $5.42 \times 10^{-4}$ |
| S&P | 98 | $4.21 \times 10^{-5}$ | $3.67 \times 10^{-4}$ | $4.60 \times 10^{-4}$ |
| Nikkei | 225 | $1.31 \times 10^{-5}$ | $3.53 \times 10^{-4}$ | $4.59 \times 10^{-4}$ |

Exhibit 10: GNC Results and BMC (Beasley et. al. [1999]) Results Using

$$\mathrm{TE}_{\mathrm{BMC}}(x)$$