J Glob Optim
DOI 10.1007/s10898-015-0305-4

Extracting from the relaxed for large-scale semi-continuous variable nondominated frontiers

Ralph E. Steuer[1] · **Markus Hirschberger**[2] · **Kalyanmoy Deb**[3]

Received: 31 December 2013 / Accepted: 17 April 2015

Abstract Because of size and covariance matrix problems, computing much of anything along the nondominated frontier of a large-scale (1000–3000 securities) portfolio selection problem with semi-continuous variables is a task that has not previously been achieved. But given (a) the speed at which the nondominated frontier of a classical portfolio problem can now be computed and (b) the possibility that there might be overlaps between the nondominated frontier of the classical problem and that of the same problem but with semi-continuous variables, the paper shows how considerable amounts of the nondominated frontier of a large-scale mean-variance portfolio selection problem with semi-continuous variables can be computed in very little time.

Keywords Multiple criteria optimization · Portfolio selection · Buy-in thresholds · Nondominated frontiers · Semi-continuous variables · Parametric quadratic programming

1 Introduction

In finance and operations research there has long been the problem of portfolio selection—the problem of how to allocate one's capital to a pool of n approved securities to maximize return. But since the "return" just mentioned has not yet had time to occur and is thus a random variable, the problem is difficult because it is a stochastic programming problem. But since

✉ Ralph E. Steuer
rsteuer@uga.edu

Kalyanmoy Deb
kdeb@erg.msu.edu

[1] Department of Finance, University of Georgia, Athens, GA 30602-6253, USA

[2] Munich Re, 80802 Munich, Germany

[3] College of Engineering, Michigan State University, East Lansing, MI 48824, USA

Markowitz [17], for addressing the stochastic nature of portfolio selection, the problem has been formulated as a bi-criterion optimization problem with one objective being to minimize "variance" (i.e., the variance of the return random variable) and the other being to maximize "expected return" (i.e., the expected value of the return random variable) as in

$$\begin{aligned} &\min\ \mathbf{x}^T \boldsymbol{\Sigma}\, \mathbf{x} = \sigma^2(\mathbf{x}) \quad \text{variance} \\ &\max\ \boldsymbol{\mu}^T \mathbf{x} = \mu(\mathbf{x}) \quad \text{expected return} \\ &s.t.\ \mathbf{1}^T \mathbf{x} = 1 \\ &\mathbf{A}\mathbf{x} \le \mathbf{b} \\ &x_i \in [0, U] \quad \text{for all } i \end{aligned} \tag{C}$$

where $\boldsymbol{\Sigma}$ is an $n \times n$ covariance matrix, $\mathbf{x} \in \mathbb{R}^n$ is a portfolio composition vector in which x_i is the proportion of capital allocated to security i, and $\boldsymbol{\mu} \in \mathbb{R}^n$ is a vector of individual security expected returns. Concerning the constraints, $\mathbf{1}^T \mathbf{x} = 1$ assures full investment, $\mathbf{A}\mathbf{x} \le \mathbf{b}$ accommodates conditions such as sector constraints (like no more than 20 % of a portfolio is to be invested in oil), and U enforces an upper bound on the amount of investment in any single security. When not vacuous, $\mathbf{A}\mathbf{x} \le \mathbf{b}$ usually adds only a few rows to the model so its presence is mainly for purposes of completeness rather than anything else. The formulation is designated (C) as it is often seen as the *classical* problem of portfolio selection.

Since it is rare for a decision maker in portfolio selection to be able to recognize an optimal solution in the absolute, decision makers typically wind up "backing into a solution." By this we mean settling on a solution, not always because of its greatness, but because it is seen that everything else is worse. While it is known that the solution that optimizes the decision maker's utility function is Pareto optimal, there is, unfortunately, rarely enough a priori information around to compute it directly. This is where the *nondominated frontier*[1] (the set of all Pareto optimal solutions) comes in. Its importance is that it is precisely the set of all candidates for optimality. By being able to see them all at once in frontier form, only then can it be assured that the solution that gets backed into, like it or not, is the decision maker's global optimum.

2 ε-Constraint method

Although there are parametric methods, such as Markowitz's [18] critical line method, for computing the whole continuous curve of the nondominated frontier of (C), the normal process for computing a nondominated frontier is by means of the ε-constraint method.[2] In this method, one of the objectives, typically the expected return objective, is converted to a constraint with an ε right-hand side. For (C), its ε-constraint formulation is

$$\begin{aligned} &\min\ \mathbf{x}^T \boldsymbol{\Sigma}\, \mathbf{x} \\ &s.t.\ \boldsymbol{\mu}^T \mathbf{x} \ge \varepsilon \\ &\mathbf{1}^T \mathbf{x} = 1 \\ &\mathbf{A}\mathbf{x} \le \mathbf{b} \\ &x_i \in [0, U] \quad \text{for all } i \end{aligned} \tag{eC}$$

[1] Also known as the "efficient frontier".

[2] For more about the ε-constraint method than utilized here, see Mavrotas [19] and Miettinen [20].

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

By solving repeatedly for different values of ε, a dotted rendition of the nondominated frontier can be obtained. Requiring the solution of a quadratic programming (QP) problem for each dot to be generated, the time to compute the nondominated frontier of (C) by means of the ε-constraint method depends upon several factors two of which are

1. number of approved securities n in the pool
2. number of dots required to represent the nondominated frontier

Concerning 1, we are now entering an era in which problems with more than 1000 securities eligible for investment are beginning to appear with greater frequency at the large financial services firms. And with Big Data, only more can be expected in the future. It is because of this, and because of difficulties that can arise in problems with more than 1000 securities, that we focus on large-scale problems (between 1000 and 3000 securities) in the this paper. As for 2, in contrast to the dozen or so dots seen in academic examples, in practice the number of dots required is likely to be 50–100 or more, so this is to be kept in mind.

3 Issues concerning the covariance matrix

In addition to the two factors of the previous section, the time to compute the nondominated frontier of (C) by means of the ε-constraint method also further depends upon the factors of

3. whether the covariance matrix $\boldsymbol{\Sigma}$ is positive definite or just positive semi-definite
4. the solver utilized
5. whether the model has been modified to incorporate into it features that require integer variables.

It is largely within these last three factors that the serious problems in the area of large-scale portfolio selection addressed in this paper lie. If in Factor 3 the covariance matrix $\boldsymbol{\Sigma}$ in (C) is positive definite, there are few difficulties. Warm starting in (eC) can be employed with state-of-the-art pivoting-based solvers as in Cplex [7] and representations of the nondominated frontier of (C) with 50–100 or more points can be obtained in times few would object to.

However, in portfolio selection, warm starting is generally only an option in problems with fewer than about 120 securities because warm starting is conditional on $\boldsymbol{\Sigma}$ being positive definite. Typically, covariance matrices in portfolio selection are computed from historical data. This means two things. One is that the resulting covariance matrix can be anticipated to be 100 % dense, and sometimes solvers have a more difficult time solving problems when the covariance matrix is dense than when it is less dense. But the other thing is much more serious. It is that a covariance matrix can only be positive definite (i.e., invertible) when the number of time periods comprising the historical data is greater than the number of securities surveyed. But this is hard when the number of time periods is typically in the range of 12 (monthly data for a year) to 120 (monthly data for 10 years).

While there are procedures for diagonalizing a covariance matrix and thereby making it invertible, information can get lost in the process causing the computed nondominated frontier to vary from the true nondominated frontier in unknown ways, so we do not want to get involved in this situation if at all possible. Thus, for the accurate computation of the nondominated frontiers of large-scale portfolio selection problems, one must be prepared to deal with covariance matrices that are both only positive semi-definite and dense.

Given that (eC) is a quadratic problem, the consequence of a covariance matrix not being positive definite is that an interior-point algorithm is required. Not being a pivoting-based method, this rules out warm-starting and puts us into the world of repetitive optimization

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

Table 1 Cplex times for solving (continuous variable) formulation (eC) for a single instance of ε, as well as for 50 repetitive optimizations of (eC), as n is varied from 1000 to 3000 in problems whose covariance matrices are fully dense and only positive semi-definite

Size	Cplex time for single instance of (eC)	Cplex time for 50 repetitive optimizations of (eC)	Time proportion for subsequent optimizations
$n = 1000$	3.46 s	73.56 s	.413
$n = 1500$	8.67 s	233.58 s	.529
$n = 2000$	17.11 s	484.31 s	.557
$n = 2500$	29.18 s	947.86 s	.645
$n = 3000$	47.61 s	1624.98 s	.676

where the savings one optimization to the next are smaller (discussed shortly). Factor 4 is not to be overlooked because as discussed in Steuer et al. [24], not all interior-point solvers are equally powerful. Consequently, with anything more than a few hundred securities, use of anything less than a solver like Cplex is not recommended.

To see where we are so far, consider Table 1. In the second column, as a function of large-scale n, are the times[3] taken on average by Cplex's barrier (interior-point) algorithm to solve (eC) for a single value of ε. In the third column, as a function of n, are the times taken on average by Cplex's barrier algorithm to compute, by repetitive optimization, a 50-point dotted representation of the nondominated frontier of (C).

The repetitive optimizations are carried out as follows. For each problem, after determining the maximum and minimum values of expected return over the nondominated frontier, the resulting range is divided into 50 equally spaced values. Then, written in Cplex's OPL modeling language, a script is applied to call Cplex in a do-loop type fashion until all repetitive optimizations are completed. While there are not the savings of traditional warm starting, there are nevertheless savings to the repetitive optimization process as indicated in the rightmost column of the table. For instance, consider the .557 figure in that column. What this means, for $n = 2000$, is that after the first optimization, which takes 17.11 s on average, each subsequent repetitive optimization takes on average only 55.7 % of that time. The savings come from the fact that the formulation, except for its ε-value, only has to be read in and laid out in memory once.

With all of this as background, Factor concerns modifications to (C) that interject integer variables into the model to enable the model to handle special features such as semi-continuous variables, which is our interest in this paper. This is because of their practicality. Semi-continuous variables are used to model buy-in thresholds so that for a security to be held in a portfolio it must be held in at least some minimum amount. Such conditions are particularly relevant to mutual funds and pension funds, where in such funds, say, with more than a few billion in assets, and there are many of them,[4] it would almost never be worth the effort to invest in a security without investing in it a few million.

Re-casting (C) and (eC) with semi-continuous variables

$$x_i = 0 \text{ or } x_i \in [L, U] \quad \text{where } L > 0$$

[3] All times in this paper are from an i7-2720 2.20 GHz computer. Sample sizes are 10 throughout.

[4] In the pension fund arena alone, the 300th largest pension fund has assets in excess of \$11 billion (Towers Watson, The World's 300 Largest Pension Funds—Year End 2012, www.towerswatson.com).

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

we have (S)

$$
\begin{aligned}
&\min\ \mathbf{x}^T \boldsymbol{\Sigma}\, \mathbf{x}\\
&\max\ \boldsymbol{\mu}^T \mathbf{x}\\
&s.t.\ \mathbf{1}^T \mathbf{x} = 1\\
&\mathbf{A}\mathbf{x} \le \mathbf{b}\\
&L y_i \le x_i \le U y_i \quad \text{for all } i\\
&y_i \in \{0, 1\} \quad \text{for all } i
\end{aligned}
\tag{S}
$$

and (eS)

$$
\begin{aligned}
&\min\ \mathbf{x}^T \boldsymbol{\Sigma}\, \mathbf{x}\\
&s.t.\ \boldsymbol{\mu}^T \mathbf{x} \ge \varepsilon\\
&\mathbf{1}^T \mathbf{x} = 1\\
&\mathbf{A}\mathbf{x} \le \mathbf{b}\\
&L y_i \le \quad x_i \le U y_i \quad \text{for all } i\\
&y_i \in \{0, 1\} \quad \text{for all } i
\end{aligned}
\tag{eS}
$$

respectively. In addition to the n original continuous variables, the models now have just as many new integer variables. This changes things again quite considerably as no interior-point method can handle a mixed integer quadratic program (MIQP) unless $\boldsymbol{\Sigma}$ is positive definite. This then quickly throws us into a "zone of insolvability" in problems with more than few hundred securities as there will almost always not be enough observations in the historical data to produce a positive definite covariance matrix. Thus, in order to focus on semi-continuous variable problems with n between 1000 and 3000 (in which it is highly unlikely that the covariance matrix will be positive definite), we will henceforth assume that the covariance matrix in (S) and (eS) is not positive definite.

After Mansini and Speranza [16] recognized (eS) as "*NP*-complete," and Chang et al. [6] pointed out the potentialities of simulated annealing, genetic algorithms, and tabu search for developing discretized approximations of the nondominated frontier of (S), a sizeable literature has materialized since on how to use both exact and heuristic procedures for addressing MIQPs in portfolio selection in which $\boldsymbol{\Sigma}$ is not positive definite. Key papers in this literature include Jobst et al. [11], Konno and Wijayanayake [12], Konno and Yamamoto [13,14], Lin and Liu [15], Bartholomew-Biggs and Kane [2], Bonami and Lejeune [3], Anagnostopoulos and Mamanis [1], Woodside-Oriakhi et al. [25], and Xidonas and Mavrotas [26]. A helpful review of this literature is also in Woodside-Oriakhi et al. [25]. But one thing stands out. Useful results on problems with more than about 225 securities, whether with exact or heuristic procedures, are very difficult to obtain. Instead of being confined to only a few hundred securities, we show in this paper how we are able to obtain useful results on the most practical of portfolio problems with up to 3000 securities.

But now, with n between 1000 and 3000 and (eS) supposedly *NP*-complete, how is (eS), for instance, to be solved for a given value of ε? The answer is not to be scared off by the complexity measure. Note that (eC) is the *relaxed* problem for (eS). (eS) is only *NP*-complete in situations where the relaxed problem does not solve (eS). However, should the solution to (eC) for a given ε also satisfy the semi-continuous variable requirements of (eS), then we have solved (eS) for that value of ε, and moreover, this is accomplished within the times listed in the second column of Table 1. But how often can we expect something like this to occur?

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

Table 2 Times taken by CIOS to compute full mathematical specifications of the nondominated frontier of (C)

Size	CIOS time for whole nondominated frontier of (C)
$n = 1000$	3.01 s
$n = 1500$	6.47 s
$n = 2000$	13.66 s
$n = 2500$	21.86 s
$n = 3000$	35.61 s

The likelihood that the solution to (eS) can be obtained from its relaxed problem depends upon L. If L is around .01 or .02, the likelihood is low, but if L is around .001 or .002, the likelihood is high. This is essentially independent of n. Given that a problem with between 1000 and 3000 securities is from a large fund, L would most likely be something like .001 or .002. For example, if we were dealing with the world's 300th largest pension fund ($11 billion), $L = .001$ would imply a buy-in threshold of $11 million. Thus L could easily be as little as .0005 in many funds without being unrealistic.

As for the rest of the paper, in Sect. 4 we spell out the observations upon which the solution strategy of this paper is based. Because detailed knowledge about the nondominated frontier of (C) is necessary to carry out the strategy, this is started in Sect. 5 and continued in Sects. 6 and 7. Section 8 reports on the computational effectiveness of the strategy, and Sect. 9 ends the paper with concluding remarks.

4 Key observations and a strategy

The principles underlying this paper come from the following two key observations. One is based on the fact that the feasible region of a problem with semi-continuous variables [e.g., (S)] is a subset of the feasible region of the same problem but with continuous variables [e.g., (C)]. This means that if any point on the nondominated frontier of (C) is feasible in (S), it is on the nondominated frontier of (S).

The other is that modern implementations of parametric methods can compute the full continuous curve on the nondominated frontier of (C) in remarkably little time. For this we have the recent implementations of Markowitz's [18] critical line method by Stein et al. [23] and Niedermayer and Niedermayer [21], the multiparametric quadratic programming procedure of Faisca et al. [8], and the CIOS parametric quadratic programming implementation specified in Hirschberger et al. [10]. Representative of this research, using CIOS, we have Table 2. Let us now compare the results of Table 2 with those of Table 1. Whereas for $n = 2500$ it takes CIOS 21.86 s to compute the whole nondominated frontier, it takes in Cplex 947.86 s to compute a 50-dot representation of the frontier. Moreover, with regard to the 29.18 s entry in Table 1, CIOS is seen to be able to compute the whole nondominated frontier in less time than Cplex can compute, on average, a single point on it. Furthermore, when CIOS computes a nondominated frontier, it does so in the form of a full *mathematical specification*, so that we can know everything mathematical about it.

Also, a nice thing about knowing a full mathematical specification of the nondominated frontier is that if a dotted representation of it is required, dots can be easily dropped onto the frontier in virtually any pattern in very little time.

With it possible for parts of the nondominated frontier of (C) to supply parts of the nondominated frontier of (S) in problems between 1000 and 3000 securities with L-values

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

appropriate to these problem sizes, the endeavor now is to determine how much of the nondominated frontier of (C) is feasible in (S), to what extent good use can be made of the information, and how long everything takes. For this we have the 4-step strategy:

1. For the (S) of interest, form its corresponding (C).
2. Solve for a full mathematical specification (done in this paper by CIOS) of the continuous curve nondominated frontier of (C).
3. Post-process the mathematical specification (done in this paper in Matlab) of the nondominated frontier of (C) to determine all points along it that satisfy the semi-continuous variable requirements of (S). This typically results in many bits and pieces.
4. Continuing with our post-processing, drop onto the bits and pieces dots to determine how much of a desired dotted representation of the nondominated frontier of (S) can be obtained in this way.

Because Steps 3 and 4 require an in-depth understanding about the structure of the nondominated frontier of (C), such information now follows.

5 Structure of classical nondominated frontier

In this section, we discuss in necessary detail the continuous curve, as in Fig. 1a, that is the nondominated frontier of (C) in *standard-deviation, expected-return* criterion space, and how it is mathematically specified. This is done to be able to extract all of the bits and pieces of the nondominated frontier of (C) that are feasible in (S). Results will later show that the number of bits and pieces can often be over one hundred.

A property of the continuous curve that is the nondominated frontier of (C) is that it is piecewise hyperbolic. That is, it is made up of a connected collection of curved line segments, each coming from a different hyperbola. In Fig. 1b, on the nondominated frontier, we see 14 dots (some of which are hard to distinguish). They define, in this example, the nondominated frontier's 13 hyperbolic segments. The topmost (1st) dot is the upper endpoint of the 1st hyperbolic segment and the bottommost (14th) dot is the lower endpoint the 13th hyperbolic segment.[5] The other dots are where the lower endpoint of one hyperbolic segment connects with the upper endpoint of the next hyperbolic segment coming down the curve. In Fig. 1c are displayed the 13 hyperbolas (some of which are hard to distinguish) that supply the 13 hyperbolic segments. For instance, the most nested hyperbola supplies the 1st hyperbolic segment.

Information (generated by CIOS) that provides a mathematical specification of a classical nondominated frontier is organized as in Tables 3 and 4. The actual entries in the two tables specify the nondominated frontier of Fig. 1, which is that of a 25-security problem produced by the random problem generator, developed in Hirschberger et al. [9], that is built into CIOS.

To illustrate Table 3, consider the row of any hyperbolic segment j. Employing the a_i in the row, the hyperbola that provides the jth hyperbolic segment is given by

$$\sigma = \sqrt{a_0 + a_1\mu + a_2\mu^2} \tag{1}$$

[5] A word about the bottommost hyperbolic "segment" of the nondominated frontier: In portfolio selection there is the minimum standard deviation boundary as shown in Sharpe [22]. It is entirely constructed out of hyperbolic segments, and the upper portion of this boundary is the nondominated frontier, that is, from the global minimum standard deviation point upward. With the global minimum standard deviation point likely falling within the relative interior of one of the hyperbolic segments of the minimum standard deviation boundary, the bottommost hyperbolic segment of the nondominated frontier will normally be observed to be a subset of this generally larger hyperbolic segment.

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

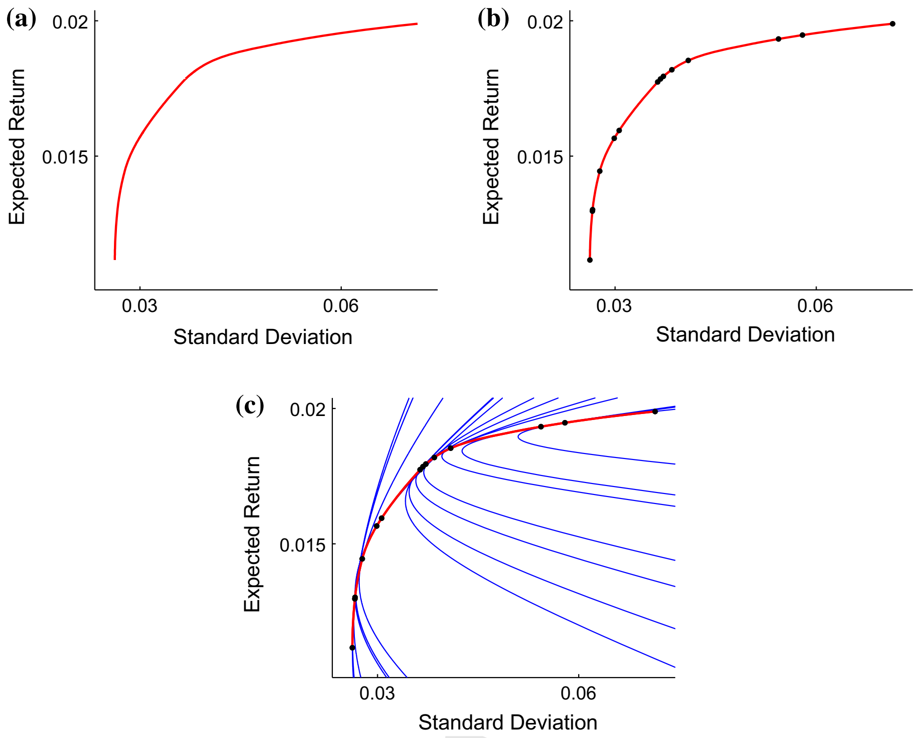

Fig. 1 **a** Nondominated frontier of a classical problem (with $n = 25$), **b** the 14 endpoints of the frontier's 13 hyperbolic segments, **c** the 13 hyperbolas from which the hyperbolic segments are taken

Table 3 Information describing the hyperbolic segments of the nondominated frontier of Fig. 1 where the a_i (must be multiplied by 10^3 before being used) are the parameters of the different hyperbolas, and the μ^{upper} and μ^{lower} specify the expected return ranges over which the different hyperbolas contribute segments to the nondominated frontier

Hyp seg	a_0	a_1	a_2	μ^{upper}	μ^{lower}
1	.0010314	−.1085372	2.862618	.019892	.019473
2	.0004814	−.0520491	1.412221	.019473	.019327
3	.0003874	−.0423224	1.160586	.019327	.018535
⋮	⋮			⋮	
13	.0000014	−.0001350	0.006051	.012966	.011159

Utilizing the μ^{upper} and μ^{lower} values in the row, expression (1) limited to

$$\mu \in [\mu^{lower}, \mu^{upper}]$$

exactly specifies the hyperbolic segment.

Table 4, on the other hand, provides information about the sets of **x**-vectors in decision space that generate the hyperbolic segments of the nondominated frontier. Specifically, the rows of Table 4 are the **x**-vectors that generate one after the other the endpoints of the hyper-

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

Table 4 Information specifying the piecewise linear path of portfolios in **x**-space that generates the piecewise hyperbolic nondominated frontier in criterion space

Endpoint portfolios	x_1	x_2	x_3	$\cdots$	x_{10}	x_{11}	x_{12}	x_{13}	$\cdots$
$\mathbf{x}^1$	.0	.0	.0	$\cdots$	.0	.0	1.00000	.0	$\cdots$
$\mathbf{x}^2$	.0	.0	.0		.22563	.0	.77437	.0	
$\mathbf{x}^3$	.0	.0	.0		.26215	.0	.55793	.0	
$\vdots$	$\vdots$								
$\mathbf{x}^{14}$	.0	.0	.03796		.0	.0	.04907	.21740	

bolic segments coming down the frontier. Consider again hyperbolic segment j. Then the **x**-vector in row j generates its upper endpoint and the **x**-vector in row $j+1$ generates its lower endpoint, with the relative interior of the straight line connecting $\mathbf{x}^j$ with $\mathbf{x}^{j+1}$ generating the relative interior of the hyperbolic segment. In this way, with the jth nondominated hyperbolic segment and the linear line segment $\mathbf{x}^j$ to $\mathbf{x}^{j+1}$ corresponding to one another, it is as it is often said, that the nondominated set of (C) is piecewise hyperbolic in criterion space and piecewise linear in decision space.

6 Nature of the sharing

With the L-values discussed earlier that would be appropriate to problems with 1000–3000 securities, there will almost certainly be overlap between the nondominated frontiers of (C) and (S). That is, there will almost certainly be points on nondominated frontier of (C) whose **x**-vectors also satisfy the semi-continuous variable requirements of (S).

To determine all places of overlap, it is necessary to examine the nondominated frontier of (C) hyperbolic segment by hyperbolic segment. As it turns out, there are thirteen different ways a nondominated hyperbolic segment of (C) can have portions of itself feasible in (S). With no significance given to the order in which shown, they are portrayed in Fig. 2. The solid dots and solid lines portray the different ways endpoints and/or portions of a hyperbolic segment can be feasible in (S), and thus be part of the nondominated frontier of (S). The dots without centers are hyperbolic segment endpoints that are not feasible in (S). For the thirteen different types of nondominated hyperbolic segments of (C), the list below spells out what can be extracted from each of them for the nondominated frontier of (S).

1. Whole hyperbolic segment including both endpoints
2. Upper portion of segment plus both endpoints
3. Lower portion of segment plus both endpoints
4. Middle portion of segment plus both endpoints
5. Middle portion of segment plus only upper endpoint
6. No part of the segment is nondominated in (S)
7. Only middle portion
8. Lower portion of segment plus only lower endpoint
9. Upper portion of segment plus only upper endpoint
10. Middle portion of segment plus only lower endpoint
11. Upper endpoint only

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

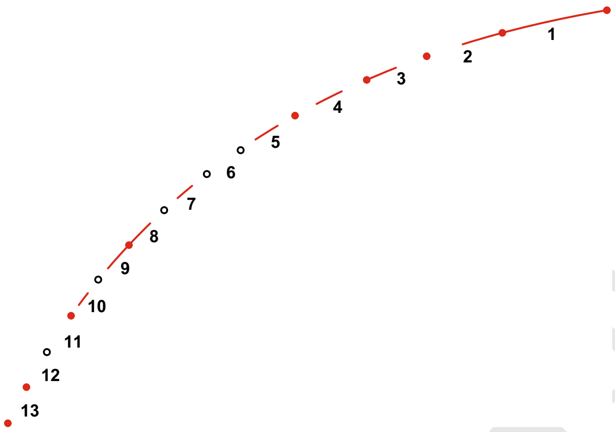

Fig. 2 The 13 different types of hyperbolic segments

Table 5 Percentages of classical problem nondominated frontier hyperbolic segments that fall into the different categories as a function of L

Hyp seg type	$L = .0005$	$L = .0015$	$L = .0025$
1	38.6 %	24.7 %	17.5 %
2	10.7	4.0	1.6
3	14.4	6.9	4.0
4	5.8	1.8	1.0
5	1.3	1.8	1.2
6	8.3	33.1	51.9
7	.6	.8	1.4
8	3.5	5.0	3.9
9	3.7	5.4	3.9
10	1.3	1.2	0.6
11	5.4	7.0	6.2
12	5.4	7.5	6.5
13	1.0	.8	.3

12. Lower endpoint only
13. Only both endpoints

To get an idea of the prevalence of each type of segment, we conducted an experiment on problems with 2000 securities. With the problems averaging 267 nondominated hyperbolic segments each, for three different values of L, we have the results of Table 5. For instance, the 38.6 % figure in the table means that for $L = .0005$, each problem had on average about 103 hyperbolic segments of type 1.

Of course, when $L = .0000$, all hyperbolic segments are of type 1, but as L takes on larger values, there is a shift of hyperbolic segments into category 6. While the 51.9 % figure in the table might not look particularly encouraging, an L-value of .0025 would be large in many situations. For example, in our \$11 billion fund, $L = .0025$ implies a buy-in threshold of \$27.5 million, which would probably be way too high.

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

Note that for $L = .0015$, by combining segment types 1–5 and 7–10, we see that 51.6 % of the segments contribute some continuous portion of themselves to the (S) nondominated frontier. Thus for the realistic values of L in the table, the rate of extraction should be good. We do not report on problems with other numbers of securities because the distributions are roughly the same as a function of L. The only thing that changes is the number of hyperbolic segments, which grows from an average of 220 when $n = 1000$ to an average of 273 when $n = 3000$.

7 Identifying the bits and pieces

In this section we describe how, using L and whatever is in Table 4, the hyperbolic segments of the nondominated frontier of (C) are classified for the purpose of extracting from them all of their endpoints and/or portions that are "L-qualified." This is the term we use from now on for feasible in (S), or in other words, on the nondominated frontier of (S).

Recall that the inverse image set of a each hyperbolic segment of (C) is a linear line segment in **x**-space. Since a linear line segment is the set of all convex combinations of its endpoints, let the *collection* of securities associated with a given hyperbolic segment be those that are positive over the relative interior of its linear line segment. In this way, a given x_i in a collection either remains fixed in value, increases linearly, or decreases linearly over the line segment. Therefore, if for the linear line segment of a given hyperbolic segment there exists an x_i in the collection whose value is strictly between 0 and L at each endpoint, no points along the hyperbolic segment are L-qualified, thus making it of type 6. Also, if every x_i in a collection has values at each endpoint that are $\geq L$, the whole hyperbolic segment is L-qualified, thus making it of type 1. These are easy cases.

For the other types, consider Fig. 3. In the figure, let the line between $\mathbf{x}^h$ and $\mathbf{x}^{h+1}$ denote the linear line segment in **x**-space of hyperbolic segment h. To begin the process of determining the hyperbolic segment's type, let I^h be the index set of all x_i that are less than L at $\mathbf{x}^h$ and greater than L at $\mathbf{x}^{h+1}$. Assume that the sloped line in Fig. 3 with $xlo_i > 0$ is the graph of one such x_i. By just this x_i alone, the portion of the hyperbolic segment associated with the first

$$apart_i = \frac{L - xlo_i}{xhi_i - xlo_i}$$

of the line from $\mathbf{x}^h$ to $\mathbf{x}^{h+1}$ is non L-qualified. Taking into account all $i \in I^h$, at least the portion of the hyperbolic segment associated with the first

$$apart = \max_{i \in I^h}\{apart_i\}$$

of the line from $\mathbf{x}^h$ to $\mathbf{x}^{h+1}$ is non L-qualified. We say "at least" because the same type of thing could be happening from the $\mathbf{x}^{h+1}$ side. In the event that this in not true, a type 8 hyperbolic segment could result. In the event that this is true on the $\mathbf{x}^{h+1}$ side with at least one $xlo_i > 0$, a type 7 hyperbolic segment could result. In the event that on the $\mathbf{x}^{h+1}$ side all $xlo_i = 0$, a type 10 hyperbolic segment could result. Should all $xlo_i = 0$ on both sides, a type 4 hyperbolic segment could result, and so forth.

When coming down the nondominated frontier, it helps to understand what causes a specific hyperbolic segment's lower endpoint. It is a change of basis in the Karush–Kuhn–Tucker system of equations as described in Hirschberger et al. [10]. While a security hitting zero or its upper bound are reasons, the most frequent change is caused by a loss of Pareto

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

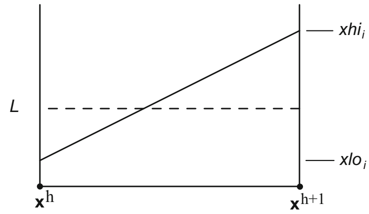

Fig. 3 Determining which parts of a hyperbolic segment h are non L-qualified to classify it

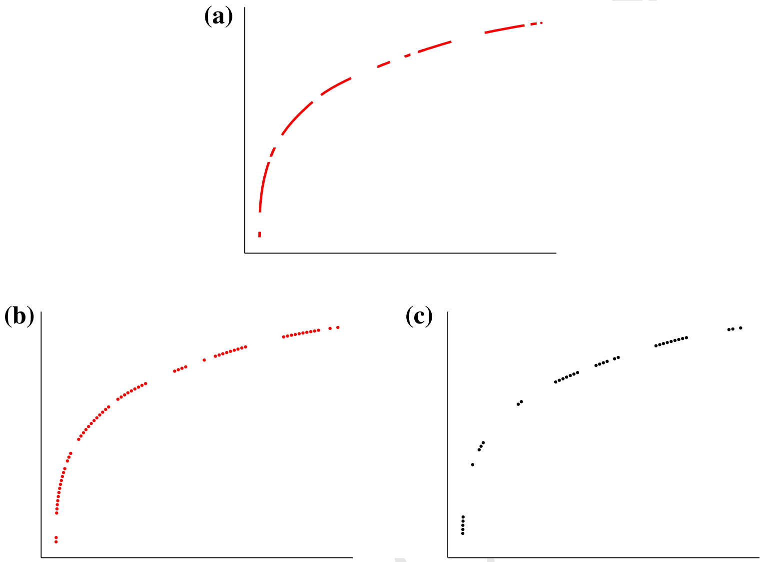

Fig. 4 **a** The 11 bits and pieces of the nondominated frontier of a 40-security (C) that are feasible in its (S) with $L = .012$, **b** out of 100 equally spaced points dropped onto the nondominated frontier of (C), the 64 that fall exactly onto the bits and pieces, **c** the 36 that do not

optimality if one were to continue on the hyperbola of the segment without a basis change.[6] Often when this last condition occurs the collection of securities one hyperbolic segment to the next does not change. This is what gives rise to the number of different hyperbolic segments being 13 rather than a smaller number.

8 Experimental results

We start this section with a small (40-security) example, because with it we can better see what is going on. Figure 4a shows the 11 bits and pieces of the nondominated frontier of (C) that are feasible in its (S) with $L = .012$.

Let us assume that we are contemplating a 100-point dotted representation of the nondominated frontier of (S). Dropping 100 equally spaced dots onto the nondominated frontier of (C), we find that 64 of them fall exactly onto the bits and pieces as in Fig. 4b. This has just saved 64 ε-constraint optimizations of (eS) when attempting to compute the nondominated frontier of (S) in the normal way, and probably several more if one were to accept moving

[6] Or stop in the case of the lower endpoint of the bottommost hyperbolic segment, see footnote 5.

Table 6 Results of experiments on problems with 1000, 2000, 3000 securities and L's as given. In all problems, $U = .04$

L	Item measured	1000	2000	3000
.0025	%Arc Length	48.34	44.26	46.72
	%Biggest Gap	7.55	7.44	6.41
	#Bits & Pieces	56.4	62.5	65.5
	Ave%ArcLengthGap	0.92	0.89	0.81
.0020	%Arc Length	55.46	51.71	53.20
	%Biggest Gap	6.00	6.67	4.50
	#Bits & Pieces	62.0	73.2	73.6
	Ave%ArcLengthGap	0.72	0.66	0.64
.0015	%Arc Length	64.95	59.86	62.55
	%Biggest Gap	4.37	4.42	3.25
	#Bits & Pieces	74.0	91.1	89.2
	Ave%ArcLengthGap	0.47	0.44	0.42
.0010	%Arc Length	73.43	70.39	73.37
	%Biggest Gap	3.46	3.53	2.35
	#Bits & Pieces	85.8	111.0	111.4
	Ave%ArcLengthGap	0.31	0.27	0.24
.0005	%Arc Length	85.53	84.80	86.22
	%Biggest Gap	1.68	1.30	1.09
	#Bits & Pieces	111.9	139.5	138.2
	Ave%ArcLengthGap	0.13	0.11	0.10

some of the points a small amount so they don't just miss falling on a bit or a piece. Figure 4c shows the 36 that do not exactly fall onto the bits and pieces and, barring any movements, would have to be computed in another way, presumably by a heuristic or an evolutionary algorithm (EA). Whereas an MIQP solver is out in situations like this in problems with between 1000 and 3000 securities because of the covariance matrix, there are advantages to heuristics and EAs. An MIQP solver either runs or it doesn't, and when it doesn't you get nothing. But with a heuristic or an EA, you always get something, and the longer you run it, the better that "something" is. (Although what happens in the gaps is not a part of this research, for insights gained from small problems about the non-concavities and discontinuities that can occur in them, see Calvo et al. [4,5]). Note that in Fig. 4c, the *biggest gap* is nine dots or 9 %. Running the problem again but with $L = .008$, we find that 74 % of the *arc length* is now covered and the biggest gap drops to 8 %, changes in the directions expected.

In Table 6 we see the results of experiments conducted over problem sizes from 1000 to 3000 securities. There are no appreciable changes horizontally across the table with regard to %Arc Length (the percent of the nondominated frontier of (C) that is feasible in (S)). However, vertically with this measure, we see significant increases as L decreases.

As for the #Bits & Pieces (number of continuous pieces and isolated endpoints), it increases as we sweep from the upper left to the lower right of the table while the %Biggest Gap and Ave%ArcLengthGap (average percent of the nondominated frontier per gap) figures decrease as we do the same. The three measures provide a guide as to how the amount of information conveyed by the bits and pieces and its dispersion increases as we sweep down and across the table.

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

Looking, for example, into the $L = .0010, n = 1000$ cell, we see on average 85.8 bits and pieces. Should we be attempting a nice, but not necessarily perfectly dispersed, 50-point representation of the nondominated frontier of (S), we should be able to nearly complete the job. Of course there will be one or two missing points due to the %BiggestGap being 3.46 %, but with an average distance between a bit or a piece (coming down the frontier) being 0.31 %, we should be in good shape with the rest of the representation. Note that an equally spaced 50-point representation has an fixed gap between dots of $1/49 = 2.04$ %.

As another example, let us look into the $L = .0005, n = 3000$ cell. Here we see an average of 138.2 bits and pieces. Say we are now thinking of a perfectly equally spaced 100-point dotted representation of the nondominated frontier of (S) (where in such a representation, the average distance between points is $1/99 = 1.01$ %). Then with an average biggest gap of 1.09 % and an average distance between a bit or a piece being 0.10 %, we should not have great difficulty in almost perfectly completing the task.

As for the time savings of the 4-step strategy of this paper for semi-continuous variable nondominated frontiers, let us again consider the $L = .0005, n = 3000$ cell. Going about a nondominated frontier in the normal way, we have on average Cplex taking 47.61 s to compute a single ε-constraint point on the nondominated frontier, but with the 4-step strategy, the 100-point representation discussed above should only take on average 35.61 s plus two or three extra seconds for Matlab to do the post-processing. That is faster than one ε-constraint optimization because of what can be accomplished by parametric QP plus post-processing.

9 Concluding remarks

To put the paper in perspective, previous research, in attempts to deal with minimum transaction sizes and buy-in thresholds, has been unable to report even modest success on the computation of points along the nondominated frontier of a mean-variance semi-continuous variable problem with many more than 225 securities (size of the Japanese NIKKEI index). This is to the best of our knowledge. But in this paper, with between 1000 and 3000 securities and realistic buy-in thresholds, we are typically able to produce a majority of the semi-continuous variable mean-variance nondominated frontier, and moreover, in very little time.

This is possible because the 4-step strategy of the paper involves first using a code like CIOS to compute a full mathematical specification of the nondominated frontier of the relaxed problem. Even for a problem with 3000 securities, this should not take much more than 35-36 s on average. Then, from the mathematical specification, the (relaxed) nondominated frontier just computed is post-processed to determine all parts of it that are on the semi-continuous variable nondominated frontier. This only takes two or three more seconds. The surprise here, with buy-in thresholds appropriate to the size of the problem, is that between 50 and 85 % of the nondominated frontier of a problem with semi-continuous variables can be extracted from its relaxed nondominated frontier. And since that extracted does not come in one strip, but in 50–140 bits and pieces, many times the 4-step strategy is able to come close to creating reasonably full dotted representations of the semi-continuous variable nondominated frontier being sought.

Furthermore, as discussed in Steuer et al. [24], it is more than the time to do just a single nondominated frontier computation that counts. Typically, when refining an asset allocation, one experiments with different pools of securities, different minimum transaction sizes, different upper bounds, and so forth. Hence, nondominated frontier after nondominated

Journal: 10898-JOGO Article No.: 0305 TYPESET DISK LE CP Disp.:2015/4/25 Pages: 16 Layout: Small

frontier may have to be computed to double check, re-confirm, and verify effects. Looking at it from a turnaround point of view, whatever time can be saved will be saved again whenever a new nondominated frontier computation request is made. Thus, this paper contributes because the faster the turnaround time, the better it is for an analyst, and this can only be for the good.

Lastly, it should eventually be possible to extend the semi-continuous variable approach of this paper to include short sales. With regard to Fig. 3, the analysis would then involve a band, a gap, zero, a gap, and a band for each variable. As a consequence of this, the number of different types of hyperbolic segments would increase from the 13 in Sect. 6 to some higher number. Also, considerable computational testing would be required to complete the extension.

Acknowledgments The authors would like to acknowledge helpful comments from the referees.

References

1. Anagnostopoulos, K.P., Mamanis, G.: Multiobjective evolutionary algorithms for complex portfolio optimization problems. CMS **8**(3), 259–279 (2011)
2. Bartholomew-Biggs, M.C., Kane, S.J.: A global optimization problem in portfolio selection. CMS **6**, 329–345 (2009)
3. Bonami, P., Lejeune, M.A.: An exact solution approach for portfolio optimization problems under stochastic and integer constraints. Oper. Res. **57**(3), 650–670 (2009)
4. Calvo, C., Ivorra, C., Liern, V.: The geometry of the efficient frontier of the portfolio selection problem. J. Financ. Decis. Mak. **7**(1), 27–36 (2011)
5. Calvo, C., Ivorra, C., Liern, V.: On the computation of the efficient frontier of the portfolio selection problem. J. Appl. Math. (2012). doi:10.1155/2012/105616
6. Chang, T.-J., Meade, N., Beasley, J.E., Sharaiha, Y.M.: Heuristics for cardinally constrained portfolio optimisation. Comput. Oper. Res. **27**(13), 1271–1302 (2000)
7. Cplex. IBM ILOG CPLEX Optimization Studio, version 12.6 (2013)
8. Faisca, N.P., Dua, V., Pistikopoulos, E.N.: Multiparametric linear and quadratic programming. In: Pistikopoulos, E.N., Georgiadis, M.C., Dua, V. (eds.) Multi-parametric Programming: Volume 1: Theory, Algorithms, and Applications, pp. 3–23. Wiley-VCH Verlag, Weinheim (2007)
9. Hirschberger, M., Qi, Y., Steuer, R.E.: Randomly generating portfolio-selection covariance matrices with specified distributional characteristics. Eur. J. Oper. Res. **177**(3), 1610–1625 (2007)
10. Hirschberger, M., Qi, Y., Steuer, R.E.: Large-scale MV efficient frontier computation via a procedure of parametric quadratic programming. Eur. J. Oper. Res. **204**(3), 581–588 (2010)
11. Jobst, N.B., Horniman, M.D., Lucas, C.A., Mitra, G.: Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. Quant. Finance **1**(5), 1–13 (2001)
12. Konno, H., Wijayanayake, A.: Portfolio optimization under D.C. transaction costs and minimal transaction unit constraints. J. Glob. Optim. **22**(2), 137–154 (2001)
13. Konno, H., Yamamoto, R.: Global optimization versus integer programming in portfolio optimization under nonconvex transaction costs. J. Glob. Optim. **32**(5), 207–219 (2005a)
14. Konno, H., Yamamoto, R.: Integer programming approaches in mean-risk models. CMS **2**(5), 339–351 (2005b)
15. Lin, C.-C., Liu, Y.-T.: Genetic algorithms for portfolio selection problems with minimum transaction lots. Eur. J. Oper. Res. **185**(1), 393–404 (2008)
16. Mansini, R., Speranza, M.G.: Heuristic algorithms for the portfolio selection problem with minimum transaction lots. Eur. J. Oper. Res. **114**(2), 219–233 (1999)
17. Markowitz, H.M.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
18. Markowitz, H.M.: The optimization of a quadratic function subject to linear constraints. Nav. Res. Logist. Q. **3**(1–2), 111–133 (1956)
19. Mavrotas, G.: Effective implementation of the ε-constraint method in multiobjective mathematical programming. Appl. Math. Comput. **213**(2), 455–465 (2009)
20. Miettinen, K.M.: Nonlinear Multiobjective Optimization. Kluwer, Boston (1999)
21. Niedermayer, A., Niedermayer, D.: Applying Markowitz's critical line algorithm. In: Guerard, J.B. (ed.) Handbook of Portfolio Construction, pp. 383–400. Springer, Berlin (2010)
22. Sharpe, W.F.: Portfolio Theory and Capital Markets. McGraw-Hill, New York (2000)

Journal: **10898-JOGO** Article No.: **0305** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2015/4/25** Pages: **16** Layout: **Small**

23. Stein, M., Branke, J., Schmeck, H.: Efficient implementation of an active set algorithm for large-scale portfolio selection. Comput. Oper. Res. **35**(12), 3945–3961 (2008)
24. Steuer, R.E., Qi, Y., Hirschberger, M.: Comparative issues in large-scale mean-variance efficient frontier computation. Decis. Support Syst. **51**(2), 250–255 (2011)
25. Woodside-Oriakhi, M., Lucas, C., Beasley, J.E.: Heuristic algrithms for the cardinality constrained efficient frontier. Eur. J. Oper. Res. **213**, 538–550 (2011)
26. Xidonas, P., Mavrotas, G.: Multiobjective portfolio optimization with non-convex policy constraints: evidence from the Eurostoxx 50. Eur. J. Finance **20**(11), 957–977 (2014)

Journal: **10898-JOGO** Article No.: **0305** TYPESET DISK LE CP Disp.:**2015/4/25** Pages: **16** Layout: **Small**