Very Large-Scale Linear Programming: A Case Study in Combining Interior Point and Simplex Methods

R.E. Bixby J.W. Gregory
I.J. Lustig R.E. Marsten
D.F. Shanno

May, 1991

TR91-11

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 1991		**00-00-1991 to 00-00-1991**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Very Large-Scale Linear Programming: A Case Study in Combining Interior Point and Simplex Methods	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**24**	

Very Large-Scale Linear Programming: A Case Study in Combining Interior Point and Simplex Methods

Robert E. Bixby[1] John W. Gregory[2] Irvin J. Lustig[3]
Roy E. Marsten[4] David F. Shanno[5]

May, 1991[6]

[1]Department of Mathematical Sciences, Rice University, Houston, TX 77251. Research of this author is partially supported by NSF Grant CCR-8815914 and the Air Force Office of Scientific Research under Grant AFOSR-90-0273.

[2]Industry, Science and Technology Department, Cray Research, Inc., 655-E Lone Oak Drive, Eagan, MN 55121

[3]Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ 08544, currently Visiting Scholar, Center for Research in Parallel Computation, Rice University, Houston, TX 77251

[4]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332. Research of this author is partially supported by the Office of Naval Research under Grant N09014-88-16-0104.

[5]Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ 08903. Research of this author is partially supported by the Air Force Office of Scientific Research, Air Force System Command under Grant AFOSR-87-0215.

Abstract

Experience with solving a 12,753,313 variable linear program is described. This problem is the linear programming relaxation of a set partitioning problem arising from an airline crew scheduling application. A scheme is described that requires successive solutions of small subproblems, yielding a procedure that has little growth in solution time in terms of the number of variables. Experience using the simplex method as implemented in CPLEX, an interior point method as implemented in OB1, and a hybrid interior point/simplex approach is reported. The resulting procedure illustrates the power of an interior point/simplex combination for solving very large-scale linear programs.

1 Introduction

Recent improvements in implementations of the simplex method as well as developments in interior point methods have changed our concept of large-scale linear programming. In this study, experience in solving the linear programming relaxation of a large set partitioning problem on a CRAY Y-MP[1] supercomputer is reported. The linear program has 837 rows with 12,753,313 columns and 99,845,826 nonzeros and arises from airline crew scheduling. Using a number of techniques, the solution time is reduced to approximately 4 minutes with a 3.5 minute preprocessing time.

The basic procedure (called *sifting*) is a kind of "column generation" method. Its application for airline crew scheduling was proposed by John Forrest [4], but in principle, it can be applied to many problem classes. Computational experience is reported for this procedure using two different methods to solve the linear programs that are generated: CPLEX,[2] an implementation of the simplex method by Bixby [3], and OB1, an implementation of the primal-dual predictor-corrector interior point method by Lustig, Marsten and Shanno [9]. In addition, a hybrid interior point-simplex implementation is explored.

In Section 2, the characteristics of the set partitioning problems are discussed and in Section 3, the sifting procedure is described. Section 3.2 describes a new pricing rule. Sections 4 and 5 describe our experience with the sifting procedure using the simplex method and an interior point method, respectively. This experience led to the development of a hybrid scheme, using both the interior point method and the simplex method. It is described in Section 6. This hybrid scheme illustrates how interior point methods and the simplex method can be effectively combined, exploiting the relative advantages of each method.

2 A Set Partitioning Problem

Set partitioning models have the form [14]:

$$\begin{array}{lrcl} \text{minimize} & c^T x & & \\ \text{subject to} & Ax & = & e, \\ & x & \geq & 0, \\ & x & & \text{integer.} \end{array} \tag{1}$$

where $A \in \{0,1\}^{m \times n}$, $e = (1,1,\ldots,1)^T \in \Re^m$, and $c \in \Re^n$. In the problems studied, $c > 0$. The impetus for this work was the "American Airlines Challenge." American Airlines uses a set partitioning model of the flight crew scheduling problem [6]. (For a recent discussion of crew scheduling, see Barutt and Hull [2].) In order to stimulate fresh thinking on the solution of such models, American generated an instance with 837 rows and approximately 12,750,000 columns. They then made the data available to researchers. The ultimate challenge is to obtain an optimal integer solution. As a step toward that goal, the sifting procedure was

[1] CRAY and CRAY Y-MP are trademarks of Cray Research, Inc.

[2] CPLEX is a trademark of CPLEX Optimization Inc.

Number of Columns $\hat{n}$	Nonduplicate Columns	Method	Iterations	CRAY CPU Time (secs)
25,000	17,937	CPLEX Dual SE	302	1.4
25,000	17,937	CPLEX Primal SE	1768	22.6
25,000	17,937	OB1	16	17.3
50,000	35,331	CPLEX Dual SE	509	4.8
50,000	35,331	CPLEX Primal SE	8233	179.7
50,000	35,331	OB1	18	22.9
100,000	68,428	CPLEX Dual SE	1215	27.3
100,000	68,428	OB1	21	34.1
200,000	134,556	CPLEX Dual SE	2549	116.6
200,000	134,556	OB1	30	67.7

Table 1: Direct solution of problems with CPLEX and OB1 (SE is Steepest Edge). Problems are generated by taking the first $\hat{n}$ columns from the input file, where $\hat{n} = (25000, 50000, 100000, 200000)$.

developed and used to solve the associated linear programming relaxation (obtained by dropping integrality in (1)).

Airline crew scheduling problems exhibit a number of interesting characteristics. Due to the way the columns are generated [6], the matrix A contains many duplicate columns. Duplicate columns are columns that have nonzeros in exactly the same rows, but possibly different cost coefficients. Eliminating all but one of these columns can significantly reduce the size of the problem without affecting the optimal solution.

In typical airline crew scheduling problems there are an average of about 10 ones per column, with a maximum of 18 in the instance provided by American. The cost coefficients c are highly variable (in the range from zero to 10,000 in the instance provided). On the other hand, the right-hand-side is constant. Thus, (1) is typically highly degenerate, but its dual is not. This fact plays a significant role in the solution strategies used.

Primarily because of degeneracy, but also because of slow convergence even when objective function progress is being made, relaxations of set partitioning problems have long been considered difficult. Even small numbers of columns can result in a problem that is troublesome for standard implementations of the primal simplex method. Recent developments of dual simplex methods and interior point methods have largely eliminated that difficulty, at least for moderate sized instances. As Table 2 shows, these approaches achieve substantial improvements over even a rather strong version of primal simplex using steepest-edge pricing. However, the growth of solution times in Table 2 with problem size also demonstrates quite clearly that direct application of even these latter approaches is not sufficient to deal with the full 12.75 million variable problem. This is especially true since solving the integer program will require multiple solutions of the linear programming relaxation.

Fortunately, set-partitioning models have an important property that allows them to be solved by non-direct methods. Let $\mathcal{W} \subset \{1, \ldots, n\}$ be a subset of columns such that the

linear program

$$\begin{array}{lrcl} \text{minimize} & c_{\mathcal{W}}^T x_{\mathcal{W}} & & \\ \text{subject to} & A_{\mathcal{W}} x_{\mathcal{W}} & = & e, \\ & x_{\mathcal{W}} & \geq & 0, \end{array} \tag{2}$$

is feasible. Then this problem is an instance of (1), since columns in $A_{\mathcal{W}}$ have the same structure as *all* columns in A. It is this property along with the fact that n is *significantly larger* than m that makes "sifting" natural for (1), and also makes sifting natural for a number of other problems such as the traveling salesman problem and path formulations of appropriate multicommodity flow problems.

3 Sifting

As noted in the introduction, the sifting procedure applied here was suggested by Forrest [4], who termed it *sprint.*

Consider the general linear program

$$\begin{array}{lrcl} \text{minimize} & c^T x & & \\ \text{subject to} & Ax & = & b, \\ & x & \geq & 0, \end{array} \tag{3}$$

with $A \in \Re^{m \times n}$, $c \in \Re^n$, $b \in \Re^m$. The dual of (3) is

$$\begin{array}{lrcl} \text{maximize} & b^T y & & \\ \text{subject to} & A^T y & \leq & c. \end{array} \tag{4}$$

The sifting procedure begins by taking a "working set" of columns $\mathcal{W} \subset \{1, \ldots, n\}$ such that

$$\begin{array}{lrcl} \text{minimize} & c_{\mathcal{W}}^T x_{\mathcal{W}} & & \\ \text{subject to} & A_{\mathcal{W}} x_{\mathcal{W}} & = & b, \\ & x_{\mathcal{W}} & \geq & 0, \end{array} \tag{5}$$

is feasible. (This assumption is not essential.) Let π^* be an optimal solution to

$$\begin{array}{lrcl} \text{maximize} & b^T \pi & & \\ \text{subject to} & A_{\mathcal{W}}^T \pi & \leq & c_{\mathcal{W}}, \end{array} \tag{6}$$

the dual of (5), and let $x_{\mathcal{W}}^*$ be an optimal solution of (5). Then the vector $x^T = ((x_{\mathcal{W}}^*)^T, 0) \in \Re^n$ is optimal for (3) if

$$c - A^T \pi^* \geq 0. \tag{7}$$

The sifting procedure suggested by these observations may be summarized as follows:

Given the linear program (3) and a set $\mathcal{W}$ such that (5) is feasible:
Solve (5) obtaining x^* and π^*.
while $(c - A^T\pi^* \not\geq 0)$ **do** (major iteration)
  Choose $\mathcal{P} \subset \{1, \ldots, n\} \setminus \mathcal{W}$. (price)
  Set $\mathcal{W} \leftarrow \mathcal{W} \cup \mathcal{P}$. (augment problem)
  (Optionally) If $\mathcal{W}$ is too big,
    reduce the size of $\mathcal{W}$. (purge)
  Solve (5) obtaining x^* and π^*. (solve)
end while

In the following sections, the key ideas in implementing the above procedure are discussed:

1. The efficient computation of $A^T\pi^*$ given π^* is discussed in Section 3.1.

2. Choosing a good set of candidates for $\mathcal{P}$ is discussed in Section 3.2, where a new method for normalizing reduced costs is introduced.

3. Controlling the size of $\mathcal{W}$ via control of the size of $\mathcal{P}$ and purges of $\mathcal{W}$ is discussed in Section 3.3.

4. Finally, the specific algorithms used to solve (5) are treated. That discussion is broken into three parts, the simplex method (Section 4), an interior point method (Section 5), and a hybrid interior point/simplex approach (Section 6). These three sections are preceded by a discussion (Section 3.4) of a number of issues applicable to each of the three solution methods.

3.1 Vectorization of the Pricing Operation

For each major iteration of the sifting procedure, the values $A^T\pi^*$ are computed. However, on a vector supercomputer, the usual direct procedure to compute $A^T\pi^*$ can exhibit poor performance if most of the columns in A are short. To get effective vectorization, the algorithm described by Forrest and Tomlin [5] is used.

Let ℓ_j represent the number of nonzeros in column j. Since $\ell_j \leq m$, a simple bucket sort (using $O(n)$ time) can be used to sort the columns according to increasing column length. Given this ordering, let n_ℓ represent the number of columns with ℓ nonzeros and let j_ℓ represent the first column (in the new ordering) with ℓ nonzeros. Then it follows that columns $j_\ell, j_\ell + 1, \ldots, j_\ell + n_\ell - 1$ all have the same column length. To compute $d_j = A_j^T\pi^*$, for $j_\ell \leq j \leq j_\ell + n_\ell - 1$ and a fixed value of ℓ, the following procedure is used, where i_{kj} represents the row number of the k^{th} nonzero in column j:

for $k = 1$ **to** l **do**
  for $j = j_\ell$ **to** $j_\ell + n_\ell - 1$ **do**
    $d_j = d_j + \pi^*_{i_{kj}} A_{i_{kj}j}$

This loop is repeated for each value of ℓ. The loop vectorizes well if the nonzeros of each column and the row indices i_{kj} are stored consecutively. This happens when the values n_ℓ are large as compared to the maximum value of ℓ_j. This procedure was found to be approximately ten times faster than the usual direct procedure for computing d_j by

$$d_j = \sum_{k=1}^{\ell_j} \pi^*_{i_{kj}} A_{i_{kj}j}. \tag{8}$$

The inner loop of the new procedure is making a calculation with a large vector of average length $n_\ell/10$, as opposed to the small vectors of average length 10 used in the direct procedure.

3.2 A New Pricing Rule

The initial experiments with solving the full problem used the following procedure that chooses a target number t of columns j with small reduced cost $c_j - A_j^T\pi^*$:

procedure standard_price:

Step 1. Compute $\bar{c} = c - A^T\pi^*$.

Step 2. Find $\mathcal{P} = \{j : \bar{c}_j < 0\}$.

Step 3. If $\mathcal{P}$ has more than t columns, then sort (in ascending order) the values $\bar{c}_j$ for $j \in \mathcal{P}$ to find the t columns to keep in $\mathcal{P}$ with the most negative values of $\bar{c}_j$.

Step 1 was carried out as described in Section 3.1. Step 2 used the CRAY subroutine `WHENFLE` and the sort in step 3 was done by using the CRAY `ORDERS` routine. The set $\mathcal{P}$ determined by **standard_price** was added to the current working set $\mathcal{W}$.

For the particular set partitioning problems considered, this pricing procedure produced adequate performance. However, in working directly with small versions of these problems, it was observed that the steepest-edge pricing rule [7] with the simplex method produced much better results than simply using reduced costs. However, the steepest-edge rule is inherently based on the simplex method. Our desire was to use a column selection strategy that is as independent as possible of the algorithm used to solve the subproblem. An analysis of the problem yielded a new pricing rule that performed well for these problems.

Given an optimal solution π^* to the dual (6) of (5), let

$$\lambda = \min_{1 \le j \le n} \left\{ \frac{c_j}{A_j^T\pi^*} : A_j^T\pi^* > 0 \right\}. \tag{9}$$

Then, since $c \ge 0$, it is clear that $\lambda \ge 1$ if and only if $\bar{c}_j = c_j - A_j^T\pi^* \ge 0$ for all j. Further, if $A_j^T\pi^* \le 0$ for all j, then $y = \pi^*$ is optimal for the dual problem (4). It follows from the definition of λ that $A^T(\lambda\pi^*) \le c$, so that the vector $\lambda\pi^*$ provides a feasible solution $y = \lambda\pi^*$ to

$$\begin{aligned} \text{maximize} \quad & e^T y \\ \text{subject to} \quad & A^T y \le c, \end{aligned} \tag{10}$$

Problem Name	Major Iterations	
	standard_price	**lambda_price**
1mil	12	9
2mil	11	9
4mil	10	11
8mil	15	12
13mil	17	12

Table 2: Major Iterations for 2 Pricing Rules

the dual of the linear programming relaxation of (1). In fact, the value $\lambda e^T \pi^*$ provides a lower bound on the objective value for the full primal linear programming relaxation of (1). This bound may be useful in other procedures that use column generation on classes of linear programs with nonnegative costs.

Intuitively, the constraints in (10) yielding the smallest values of $c_j / A_j^T \pi^*$ are most likely to be active at the optimal solution. This idea suggests the following procedure for selecting a target t number of columns:

procedure lambda_price:

Step 1. Compute $d = A^T \pi^*$.

Step 2. Compute $\bar{c} = c - d$.

Step 3. Find $\mathcal{P} = \{j : \bar{c}_j < 0\}$.

Step 4. If $\mathcal{P}$ has t or less columns, then exit.

Step 5. Compute $\lambda_j = c_j / d_j$ for columns $j \in \mathcal{P}$.

Step 6. Sort (in ascending order) the values λ_j for $j \in \mathcal{P}$.

Step 7. Set $\mathcal{P}$ to the t columns with the lowest values of λ_j.

Note that if $j \in \mathcal{P}$, then $c_j - A_j^T \pi^* < 0$ which implies that $\lambda_j < 1$ and $A_j^T \pi^* > 0$ so that the above steps are well defined. This pricing rule should prefer columns with low absolute cost as well as columns with more nonzeros. Experimentation with the initial pricing rule indicated that these columns should be preferred as these types of columns are predominant in the optimal solution.

As compared to the **standard_price** procedure, there is a small amount of extra work represented by step 5 in **lambda_price**. This step vectorizes well and only added a small cost to the pricing procedure. This cost was small compared to the savings generated by the reduction in the number of outer iterations. The new pricing procedure generally seems to pick better columns for the sifting procedure. Table 2 indicates the number of major iterations of the sifting procedure using the two different pricing procedures. CPLEX was used to solve the subproblems. The different problems are defined in Section 3.4.

Degeneracy and Multiple Dual Optimality

It is well known that the linear programming relaxations of set partitioning problems exhibit massive primal degeneracy. For the sifting procedure, the most important implication is the presence of multiple dual optimal solutions. The particular choice of optimal dual solution influences the selection of new columns.

If the subproblem (5) is solved by the simplex method, the optimal basis chosen is essentially random, changing the qualitative nature of an optimal dual solution. If the subproblem (5) is solved by an interior point method, the solution will be in the relative interior of the face of optimal dual solutions. The particular dual solution π^* chosen depends on the initial interior point used by the solution algorithm.

3.3 Problem Size Control

One of the most difficult decisions in implementing the sifting procedure is control of the cardinality of $\mathcal{W}$. We chose to specify a sequence $(t^1, t^2, \ldots)$ of target values, such that the cardinaltiy of $\mathcal{P}$ is restricted to be smaller than t^k on the k^{th} iteration of the sifting procedure. In addition, it is important to occasionally purge columns from $\mathcal{W}$ as an additional control on the size of the subproblems. The details of these procedures for each of the three solution methods are discussed in the appropriate sections. Unfortunately, one of the weaknesses of the algorithm as implemented is that these sequences were determined by experimentation for the specific problem being studied. However, once a reasonable sequence was determined, it was applied *unchanged* to all problem instances. Thus, although the largest problem solved has approximately 12.75 million columns, we would choose exactly the same sequence for a problem of larger size.

It may seem more natural to simply build the subproblems by including in $\mathcal{P}$ all columns with negative reduced cost. That approach is simply impractical. For example, 10 percent of the columns have negative reduced costs on the first iteration.

3.4 Engineering the Problem

Our experiments were run on one processor of a CRAY Y-MP computer with 128 megawords of memory running UNICOS 6.0.11 with the `cft77` Fortran compiler version 5.0.0 and the `scc` C compiler version 3.0.0. At most 28 megawords were required to solve any of the problems. In attempting to solve such a linear program, a number of tasks are required to prepare the problem for solution.

The data was provided in 7 ASCII (machine independent format) files. Each line of each file corresponds to one column in the constraint matrix. A simple preprocessor was used to read in this file in 250,000 column sections and write out 51 binary files consisting of a slightly modified image of the original data. Because of the nature of the ASCII files, 39 sections had exactly 250,000 columns, 10 sections had 250,302 columns, 1 section had 249,998 columns and 1 section had 250,297 columns. This process of converting the data to a binary format is essentially a required input step. Moreover, it is reasonable to assume that the data could be provided in this binary format initially. In fact, the time to read

Problem Name	Original Columns	Nonduplicate Columns	Number of Sections s	Optimal Objective
`1mil`	1,000,000	673,310	4	54325.363
`2mil`	1,999,998	1,341,904	8	51725.598
`4mil`	3,999,998	2,682,175	16	50413.578
`8mil`	7,999,998	5,374,005	32	48994.020
`13mil`	12,753,313	8,549,879	51	48400.129

Table 3: Problem Statistics

in the 7 ASCII files (approximately 1400 CRAY Y-MP CPU seconds) exceeds the solution time! The particular format of these binary files stored each section of data as 3 vectors, a cost vector, the vector ℓ of nonzero counts per column, and a vector of the indices of rows with nonzeros. These three vectors are written to each file as separate binary vectors in order to enhance vectorization.

For the problem that was provided, the first 1000 columns constitute an initial feasible solution. This set of columns was used as the initial $\mathcal{W}$ for all solution methods. The performance of the sifting method on this model is very insensitive to the size of this initial feasible set.

To reduce the cost of computing $A^T\pi^*$ on each iteration, duplicate columns were removed from each of the 51 sections. Let $\mathcal{S} \subset \{1, \ldots, n\}$ denote the columns within some particular section. Given $\mathcal{S}$, all of the nonduplicate columns in $\mathcal{S}$ are found before sorting the columns of $\mathcal{S}$ by the number of nonzeros per column, as described in Section 3.1. Note that this step does not remove all nonduplicates from A, as duplicate columns may still exist across section boundaries. Because of this, duplicate columns were eliminated among $\mathcal{W}$ on each major iteration of the sifting procedure before the subproblem (5) was optimized.

For the 12,753,313 column problem provided by American, duplicates were removed from each of the 51 sections reducing the size of each section to an average size of 167,645 columns. This process consumed 212 CRAY Y-MP CPU seconds, plus an estimated 103 additional overhead seconds for input of the binary data files and output of the binary non-duplicate files, as described below. Table 3 presents statistics for the sequence of 5 problems generated from the initial data that were used for the tests in this study and the names we assigned for reference to each problem. The value s is the number of sections used to solve the problem. Note that the sizes of the problems are not exact multiples of 1,000,000 because of the way the original data sets were provided. In particular, one of the original files contained only 1,999,998 columns, rather than the expected 2,000,000.

The size of each section was chosen large enough to get good vectorization performance during the pricing operation and small enough so as not to consume too much memory when loaded from external storage. Each section of nonduplicate columns is stored in an external file. On the CRAY Y-MP, experiments were conducted with three choices for the locations of the external files. The first choice was a UNICOS 5.0 file system, the second a UNICOS 6.0 file system, and the third a solid-state storage device (SSD) that essentially is

a RAM-disk. It was found that using SSD to store the external files produced a negligible overhead for the sifting procedure, and hence all further results are reported assuming SSD usage. Each section of nonduplicate files is stored as a binary file on the SSD. The values n_ℓ, i_{kj} (Section 3.1) and the cost vector c are stored as binary vectors in these files.

There is one final issue to discuss in connection with nonduplicates. Since no duplicates were initially found across section boundaries, duplicate columns were removed from $\mathcal{W}$ before the problem (5) was solved. Hence, the effective number of columns added is unpredictable and decreases as s increases. As a consequence, the following instrumentation was added. On the first iteration of the sifting procedure, t^1/s columns are chosen from each section. Duplicate columns are removed from $\mathcal{P}$ and the number of new nonduplicate columns added, p^1, is computed. The value $r = p^1/t^1$ is then determined. Successive target sizes t^k for $k > 1$ are then adjusted to new values $\bar{t}^k = t^k/r$ to regulate the growth in $\mathcal{W}$. On each further iteration, $\bar{t}^k/s$ columns are then chosen from each section using the procedure described in Section 3.2.

4 Using CPLEX

To test the application of the simplex method in the sifting procedure, the CPLEX optimizer [3], version 2.0, was used. Much effort has been put into CPLEX to get good vectorization performance from various aspects of the code. Some modifications of CPLEX were made for this application. To conserve approximately 6 percent of CPU time, routines accessing the matrix $A_{\mathcal{W}}$ assumed that nonzeroes in the matrix had the value 1. In addition, because the bases of the problem have extremely dense factorizations, the refactorization interval was set at 175.

Extensive experimentation with CPLEX on various versions of these problems reveal that steepest-edge pricing is more appropriate than standard reduced-cost pricing for both the primal and dual simplex methods. At the beginning of the sifting procedure, the initial feasible solution consisting of the first 1000 columns is highly degenerate and no starting basis is available. However, the basis consisting of the m artificial variables is dual feasible, making the dual simplex method effective for solving the first few subproblems. Since a large number of columns are added on these initial iterations, using an advanced starting basis with the dual simplex method is not appropriate due to the large dual infeasibility incurred by the addition of the new columns. After the initial degenerate feasible point is passed by the sifting procedure, the optimal bases found by solving (5) are less degenerate, and a primal simplex method using the previously optimal basis works well. Note that this procedure is unstable on smaller instances of the problem since the initial degenerate point may not be passed. In this case, (5) is reoptimized from an artificial start each time, and the pricing information varies due to multiple dual optimality, as discussed in Section 3.2.

The use of an artificial variable basis to start the dual simplex algorithm provides some insight about the dual variables π when using the simplex algorithm to solve the subproblems. For these problems, it was found that the optimal bases tended to have a large number of artificial variables, which are degenerate. (For the largest problem (`13mil`), 96 artificial variables are basic at the optimal solution.) Artificial variables also affect the prices for the

Problem Name	Size of $\mathcal{W}$		Major Its.	CPLEX Its.	CPU Time (secs)			
	Max.	Final			Misc.	Price	Optim.	Total
1mil	24,637	7,778	9	11,411	1.3	3.0	176.0	179.0
2mil	24,911	7,579	9	12,617	1.4	5.8	199.7	205.5
4mil	25,275	14,256	11	16,769	2.3	13.5	293.0	308.8
8mil	26,539	6,826	12	20,297	1.9	31.3	325.6	358.8
13mil	26,637	7,556	12	19,396	2.2	47.4	333.9	383.5

Table 4: Results with CPLEX

sifting procedure. If the artificial variable on row i is basic, it is easy to see that $\pi_i^* = 0$ and hence row i contributes nothing to the selection of new columns. Thus, it is worth considering the assignment of a penalty M as the cost of each artificial variable, as in the OB1 runs (Section 5). In this case, the artificial variable may still remain basic with $\pi_i = M$. For the simplex method, no difference was found in the performance of the sifting procedure when a penalty M was used as compared to no penalty.

With CPLEX, the deletion strategy aimed at keeping the subproblems relatively small. A large number of columns requires more simplex iterations as well as more work per iteration since full pricing is being used. The size of a problem was limited to 33,000 columns. If a problem exceeded that size, then all columns $j \in \mathcal{W}$ with a reduced cost $\bar{c}_j > 100$ were deleted. (For example, in Table 5 on the third major iteration, 18,666 colums were added to a problem of size 18,977. Then 17,843 columns were removed because the problem was too large. This was followed by the removal of 8,884 duplicate columns.) Furthermore, as soon as an optimal solution to (5) was found with a smaller objective value than the initial feasible solution, columns were purged (using the same rule). Since the sifting algorithm used the primal simplex algorithm with steepest-edge pricing after this point, the philosophy is to keep the subproblems small.

4.1 CPLEX: Computational Results

The sequence $(t^1, t^2, \ldots, t^{10})$ was taken to be $(16000, 16000, 8000, 8000, 4000, 2000, 2000, 1000, 1000, 500)$, with $t^k = 1000$ for $k > 10$. After removal of duplicates, the first linear program solved in each case had 851 columns. Table 4 contains the summary of the results with CPLEX. The columns indicating the size of $\mathcal{W}$ represent the maximum and final sizes of the linear programs (5) that were solved. The number of major iterations is the number of iterations in the sifting procedure. The number of CPLEX iterations is the total number of simplex iterations needed to solve each subproblem. CPU Time represents the CRAY Y-MP time to solve the problem, broken into miscellaneous time (for example, the time to remove duplicates from each subproblem), pricing time and optimization time. The pricing time includes the time required to add the columns in $\mathcal{P}$ to the CPLEX data structures.

Table 5 contains an iteration log for solving the problem 13mil. The second column indicates the size of $\mathcal{W}$ on each iteration. The CPLEX algorithm is the dual simplex or

Major Iter	Cols in Prob	CPLEX Alg.	CPLEX iters	Objval	Optim. time	Price time	Cols added	Cols purg.	Dups del.
1	851	Dual	108	57448.000	0.190	5.960	15963	0	9122
2	7692	Dual	346	57448.000	1.500	4.581	37332	7080	18967
3	18977	Dual	1654	57448.000	21.497	4.625	18666	17843	8884
4	10916	Dual	1544	57448.000	16.993	4.795	18666	0	8595
5	20987	Dual	3981	57448.000	69.532	3.888	9333	0	3683
6	26637	Dual	6218	52035.965	129.514	3.584	4641	24725	2267
7	4286	Primal	1829	49835.305	26.073	3.529	4641	0	2272
8	6655	Primal	1876	48856.961	35.269	3.500	2049	0	1309
9	7395	Primal	1356	48438.438	23.684	3.466	300	0	158
10	7537	Primal	432	48400.691	7.333	3.288	34	0	17
11	7554	Primal	50	48400.129	1.455	3.100	2	0	0
12	7556	Primal	2	48400.129	0.839	3.076	0	0	0

Table 5: Iteration log for CPLEX on `13mil`

primal simplex method, each with steepest-edge pricing. The number of CPLEX iterations is the total number of iterations to solve the corresponding subproblems. The objective value is the final objective value of the subproblem (5). The optimization time is the time required by CPLEX to solve each subproblem, while the pricing time is the time required to price all 12.75 million columns and add columns to the CPLEX data structure.

5 Using OB1

The implementation of OB1 [9] of the primal-dual predictor-corrector interior point method used for these experiments was modified to take advantage of the special characteristics of the problem. In each iteration of the interior point method, it is necessary to compute $\pi^T A_W$ twice. To do this efficiently, each subproblem was preprocessed to order the columns by column count, using the same method as described in section 3.1. In addition, all references in the code to specific nonzero values in A_W were eliminated, since the problem is assumed to have all nonzero values in A_W equal to 1.

All interior point methods for linear programming form a matrix of the form $(A\Theta A^T)$ on each iteration, where Θ is some diagonal matrix. This step is followed by a Cholesky factorization. For better vectorization performance, it was found that reordering A_W to increase the sparsity of the Cholesky factor L was not worth the effort as $(A_W \Theta A_W^T)$ is almost totally dense. Hence, reordering was not used, and a special version of a Cholesky factorization that assumed a dense matrix replaced the default sparse factorization routines. This factorization routine achieved an average performance on the CRAY Y-MP of 240 Megaflops (Millions of Floating Point Operations per Second) on a single processor capable of a theoretical maximum of 333 Megaflops.

An analysis of the performance of OB1 solving the subproblems revealed that the formation of $(A_W \Theta A_W^T)$ was requiring a significant amount of the processor time. An analysis of this computation led to an algorithm that vectorizes well for this task. For this discussion, let $\hat{n}$ represent the number of variables in the linear program (5) and let $\hat{A} = A_W$. Then

$$\hat{A}\Theta\hat{A}^T = \sum_{j=1}^{\hat{n}} \Theta_j \hat{A}_j \hat{A}_j^T. \tag{11}$$

Note that the nonzero components of $(\hat{A}_j \hat{A}_j^T)$ are exactly equal to one. The matrix $(\hat{A}\Theta\hat{A}^T)$ is symmetric so that only the diagonal and the lower triangular part needs to be computed. Suppose $\hat{A}_j$ has ℓ_j nonzeros at positions $i_{1j}, i_{2j}, \ldots, i_{\ell_j j}$. The diagonal $D \in \Re^m$ of $(\hat{A}\Theta\hat{A}^T)$ is computed as follows:

procedure compute_diagonal:
$D \leftarrow 0.$
for $j = 1$ to $\hat{n}$ **do**
  for $k = 1$ to ℓ_j **do**
    $D_{i_{kj}} \leftarrow D_{i_{kj}} + \Theta_j$

The inner loop of this procedure vectorizes (although with short vector lengths, adversely affecting peak performance) since the values $i_{1j}, i_{2j}, \ldots, i_{\ell_j j}$ are unique.

To efficiently compute the lower triangular part of $(\hat{A}\Theta\hat{A}^T)$, an address list similar to the one described by Adler, et.al. [1] is used. However, the nature of the matrix dictates that the "elementary products" are easier to store since each is equal to 1.0. To achieve the best possible vectorization, the address list is computed in a special way. Let $Q \in \Re^{m \times m}$ be a two-dimensional lower triangular matrix with a zero diagonal. For $i_1 > i_2$, define $Q_{i_1 i_2}$ to be the number of columns in $\hat{A}$ containing both i_1 and i_2. It is easy to see that the lower triangle of Q is equal to the lower triangle in the matrix $(\hat{A}\hat{A}^T)$. Let

$$\hat{Q} = \max \{ Q_{i_1 i_2} : 1 \leq i_2 < i_1 \leq m \}, \tag{12}$$

so that $\hat{Q}$ is the maximum element in Q. Define T_q, for $1 \leq q \leq \hat{Q}$, by

$$T_q = \sum_{1 \leq i_2 < i_1 \leq m} I(Q_{i_1 i_2} \geq q), \tag{13}$$

where I is an indicator function defined by

$$I(Q_{i_1 i_2} \geq q) = \begin{cases} 1 & \text{if } Q_{i_1 i_2} \geq q; \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Hence T_q is a count of the number of elements in Q greater than or equal to q.

The address list R consists of $\hat{Q}$ sections, each with T_q ordered triples, $1 \leq q \leq \hat{Q}$. The ordered triples (i_r, k_r, j_r) in section R_q, $1 \leq r \leq T_q$, are defined by an inductive rule such that $(i_r, k_r, j_r) \in R_q$ if and only if the following three conditions are satisfied:

1. $Q_{i_r k_r} \geq q$;

2. $\hat{A}_{i_r j_r} = 1$ and $\hat{A}_{k_r j_r} = 1$; and,

3. $(i_r, k_r, j_r) \notin R_t$, for $t < q$.

Hence, R_1 is a list of all of the nonzero positions in Q, with each element of Q having one of the columns j_r that contribute to it. The list R_2 is a list of all of the nonzero positions that have 2 or more columns contributing to Q, using a different column from the one present in R_1.

Given the address lists, it is easy to then compute the lower triangular part of $(\hat{A}\Theta\hat{A}^T)$ using the following procedure:

procedure compute_AAT:
$(\hat{A}\Theta\hat{A}^T)_{i_1 i_2} \leftarrow 0$ for $1 \leq i_2 < i_1 \leq m$.
for $q = 1$ to $\hat{Q}$ **do**
  for $r = 1$ to T_q **do**
    $(\hat{A}\Theta\hat{A}^T)_{i_r k_r} \leftarrow (\hat{A}\Theta\hat{A}^T)_{i_r k_r} + \Theta_{j_r}$

The inner loop of this procedure vectorizes well and substantially reduces the time it takes to compute $(\hat{A}\Theta\hat{A}^T)$. Note that this matrix is actually stored as a vector and that the location of element (i_r, k_r) is easily computed since $(\hat{A}\Theta\hat{A}^T)$ is assumed to be dense and is stored as a long vector. In fact, the address list stores this location. If $\mathcal{W}$ has $\bar{n}$ columns, then the length of the address list (the size of R) is expected to be $\bar{\ell}^2\bar{n}/2$, where $\bar{\ell}$ is the average of $\ell_1, \ell_2, \ldots, \ell_{\bar{n}}$.

When using the interior point method within sifting, it was found that deleting columns was not beneficial. Deleting many columns caused most of those columns to re-enter the problem on a later iteration. Hence, growth in the size of $\mathcal{W}$ was allowed and controlled. It seems that the dual solution π^* generated on a particular iteration k yielded important information regarding the dual feasibility of the problem. Discarding columns that forced this dual feasibility to be maintained was not useful due to the fact that the dual solution is in the relative interior of a dual face of the subproblem. This contrasts to using the simplex method, where the extreme point nature of the dual solution seemed to allow column deletion without any difficulty.

The predictor-corrector interior point method was started with a primal solution of $x_j = 100.0$ for each $j \in \mathcal{W}$, with a dual feasible solution of $\pi = 0$ and $z = c_{\mathcal{W}}$. The large value of x_j improves the performance of the predictor-corrector method. No attempts at implementing a restart strategy along the lines of Lustig, et.al. [10] or Mitchell [13] were successful. This is viewed as an avenue for future research.

Since these problems are highly rank deficient, a set of m artificial columns with a cost of $10,000$ were included in the problems solved by OB1 on each iteration of the sifting procedure. While the presence of these columns assists in keeping $(\hat{A}\Theta\hat{A}^T)$ better conditioned, these columns also influence the pricing information obtained by OB1. As remarked earlier, the presence of these columns did not affect the performance of the sifting procedure with CPLEX.

Problem Name	Final Size of $\mathcal{W}$	Major Its.	OB1 Its.	CPU Time (secs)			
				Misc.	Price	Optim.	Total
1mil	28,211	8	197	1.5	3.6	230.2	235.3
2mil	32,541	8	210	1.7	7.0	257.8	266.5
4mil	36,415	8	204	1.9	14.3	251.6	267.9
8mil	41,680	9	242	2.4	29.5	306.6	338.5
13mil	52,368	9	239	3.0	48.8	326.7	378.5

Table 6: Results with OB1

Major Iter	Cols in Prob	OB1 iters	Objval	Optim. time	Price time	Cols added	Dups del.
1	1688	9	57448.000	8.528	7.191	16983	10934
2	7737	13	57448.000	13.317	6.104	16269	7453
3	16553	18	57448.000	19.467	6.378	30855	13898
4	33510	32	52316.105	41.113	5.745	30855	12791
5	51574	34	48815.031	48.720	5.662	1747	1044
6	52277	33	48431.762	48.084	5.620	244	169
7	52352	33	48400.703	50.498	4.160	26	15
8	52363	32	48400.129	46.224	4.752	37	32
9	52368	35	48400.129	50.736	3.196	0	0

Table 7: Iteration Log for OB1 solving 13mil

5.1 OB1: Computational Results

In this section, computational results are presented for the sifting procedure using OB1 to solve the subproblems. The values used were $t^1 = 17,000$, $t^2 = 5800$ and $t^k = 11000$ for $k \geq 3$. Table 6 presents the results for the sifting procedure using the interior point method to solve the subproblems. The final size of $\mathcal{W}$ is also the maximum size since no deletions were done on any of the problems. The number of major iterations remained essentially constant. The times represent similar values as in Section 4.1.

Table 7 illustrates the iteration log for solving the problem 13mil using OB1 as the solver. It is interesting to note that the sifting procedure with OB1 requires fewer iterations to get past the objective value corresponding to the initial point. The prices obtained by OB1 are insensitive to the degeneracy of the initial problem, although they are sensitive to the initial interior point used by the primal-dual algorithm. The time to solve each problem rises with the number of columns and is fairly predictable. The pricing time per iteration for this method is higher due to a higher overhead required to add columns to the OB1 data structures than the CPLEX data structures.

6 A Hybrid Method

A careful examination of the above computational results reveals that the interior point method spends a predictable amount of time solving each subproblem. It performs quite well in the beginning, but towards the end, the same amount of time is spent to solve a new subproblem when only a few columns are added. In contrast, the simplex method has difficulty making progress away from the initial highly degenerate solution, but as fewer columns are added after this point is passed, the simplex method spends little time re-optimizing the new problem starting from an optimal basis for the previous subproblem. This led to the development of a hybrid procedure, where the interior point method solves 5 linear programs and is then followed by the application of the simplex method. In effect, OB1 is being used to find a good starting point for the simplex method.

A difficulty with this method is the identification of a starting basis for the simplex method from an interior point solution. If all columns with $x_j > \epsilon$ (where ϵ is a suitably chosen tolerance) are selected to be basic columns, and the dual solution is ignored, too much information is lost. Essentially, the solution provided by the interior point method indicates something about the dual feasibility of all the columns in $\mathcal{W}$, and discarding columns for which dual feasibility is attained does not work well.

Therefore, the procedure described by Megiddo [12] was used to identify an optimal basis. The implementation of this procedure is described by Lustig [8]. First, the procedure identifies a crash basis, leaving all nonbasic variables at the value determined by the interior point method. This solution is purified to a primal optimal solution by lowering each positive nonbasic x_j to its bound. The work per iteration of this method is equivalent to the work per iteration of the primal simplex method. The second part of the procedure maintains the dual feasibility of the interior point method by doing a similar procedure to the dual linear program. Each of these iterations is equivalent to a step of the dual simplex method. The procedure is guaranteed to converge in at most m pivots. For the results reported here, the routines implemented by Lustig [8], using XMP [11] and LA05 [15], were employed. Both of these codes are not suitable for good vectorization. Better results would be obtained by implementing this crossover algorithm within CPLEX.

Purifying the primal and dual optimal solutions in this way maintained dual feasibility relative to the columns already in $\mathcal{W}$. All columns with $\bar{c}_j \leq 100$, $j \in \mathcal{W}$, were then passed to the simplex method, effectively paring the problem down to an acceptable size. This procedure worked well and substantially reduced the total computation time.

6.1 OB1 and CPLEX: Computational Results

In this section, computational results are presented for the hybrid scheme described above. For each problem solved, OB1 was used to solve the first 5 subproblems, with $t^1 = 25,500$ and $t^k = 11,000$ for $2 \leq k \leq 4$. (Note that these target values differ from those used in Section 5.1, where total problem growth is a more serious consideration.) After the fifth linear program was solved, the crossover scheme was invoked followed by one solve of a small linear program via CPLEX to get an initial factorization. After this initial solve, $t^6 = t^7 = 800$ and $t^k = 400$ for $k \geq 8$. Note that t^5 is considered irrelevant due to the crossover scheme.

Problem Name	Size of $\mathcal{W}$			Major Its.	OB1 Its.	Crossover Its.	CPLEX Its.
	Last OB1	First CPLEX	Last CPLEX				
1mil	33,456	2457	3899	9	119	335	228
2mil	37,185	2984	4250	9	116	421	358
4mil	41,479	3539	4592	9	116	387	478
8mil	42,176	4228	6774	12	126	336	1389
13mil	45,083	5217	6980	10	111	327	1268

Table 8: Statistics for Hybrid Scheme

Problem Name	CPU Time (secs)					
	Misc.	Price	OB1	Cross	CPLEX	Total
1mil	1.3	3.1	145.4	21.1	8.7	179.6
2mil	1.4	6.0	145.4	22.9	10.7	186.4
4mil	1.6	12.3	146.8	24.7	13.8	199.2
8mil	1.9	31.9	163.1	25.9	35.9	258.7
13mil	2.0	42.6	140.3	30.5	33.6	249.0

Table 9: CPU Times for Hybrid Scheme

Table 8 shows various statistics for this approach, while Table 9 shows CPU time. In Table 8, the size of the last linear program solved by OB1 is given, along with sizes for the first and last linear programs solved by CPLEX. The total number of major iterations includes 5 interior point iterations, followed by 1 crossover iteration, followed finally by a sequence of CPLEX solves. In Table 9, the miscellaneous time includes miscellaneous tasks such as the deletion of duplicates before each subproblem is solved. The pricing time is the time to price out all of the columns including the time required to add the columns to the appropriate data structures for OB1 or CPLEX. The crossover time is the time used by the OB1/XMP routines to compute an optimal basis. The final CPLEX time is the time used by CPLEX to finish the procedure.

Table 10 gives an iteration log for solving the problem 13mil with the hybrid scheme. After OB1 solves the first 5 subproblems, the crossover procedure is invoked. The crossover requires only 327 iterations, but as previously noted, these iterations are expensive because of the lack of vectorization in XMP and LA05. CPLEX is then handed a small problem and an optimal basis. This basis is factorized and checked for optimality. CPLEX then solves the remainder of the subproblems.

The computational performance of the hybrid scheme is clearly superior to the pure interior point or pure simplex procedures and could be further improved by doing the crossover procedure within CPLEX. Moreover, the hybrid procedure is clean and a more robust prod-

Major Iter	Cols in Prob	Method	iters	Objval	Optim. time	Price time	Cols added	Dups del.
1	1688	OB1	9	57448.000	8.549	7.262	25449	14638
2	12499	OB1	14	57448.000	15.324	6.041	25857	12135
3	26221	OB1	24	57448.000	28.582	6.493	25857	12774
4	39304	OB1	33	50734.910	44.399	5.750	13018	7239
5	45083	OB1	31	48530.461	43.398	0.000	0	0
*	45083	Crossover	327	48530.461	30.538			
6	5217	CPLEX	0	48530.461	3.105	3.555	1832	1000
7	6049	CPLEX	297	48510.117	9.572	3.532	1798	1022
8	6825	CPLEX	648	48418.199	13.754	3.571	382	238
9	6969	CPLEX	316	48400.129	6.244	3.259	18	7
10	6980	CPLEX	7	48400.129	0.972	3.145	0	0

Table 10: Iteration log for problem `13mil` with hybrid scheme

uct. It is not subject to the unpredictable behavior of the simplex method in the earlier "degenerate" phase, and is not sensitive to the total number of iterations at the end in the same way as OB1, where an additional pass can cost almost 1 minute of additional computation time.

References

[1] I. ADLER, N. KARMARKAR, M. G. RESENDE, AND G. VEIGA, *Data structures and programming techniques for the implementation of Karmarkar's algorithm*, ORSA Journal on Computing, 1 (1989).

[2] J. BARUTT AND T. HULL, *Airline crew scheduling: Supercomputers and algorithms*, SIAM News, 23 (1990).

[3] R. E. BIXBY, *Implementing the simplex method: The initial basis*, Tech. Rep. TR90-32, Rice University, Department of Mathematical Sciences, Houston TX, 1990.

[4] J. J. FORREST, *Mathematical programming with a library of optimization subroutines.* Presentation, ORSA/TIMS Joint National Meeting, New York, October 1989.

[5] J. J. FORREST AND J. A. TOMLIN, *Vector processing in simplex and interior methods for linear programming*, Annals of Operations Research, 22 (1990), pp. 71–100.

[6] I. GERSHKOFF, *Optimizing flight crew schedules*, Interfaces, 19 (1989), pp. 29–43.

[7] D. GOLDFARB AND J. K. REID, *A practicable steepest-edge simplex algorithm*, Mathematical Programming, 12 (1977), pp. 361–371.

[8] I. J. LUSTIG, *An implementation of a strongly polynomial time algorithm for basis recovery.* (In preparation).

[9] I. J. LUSTIG, R. E. MARSTEN, AND D. F. SHANNO, *On implementing Mehrotra's predictor-corrector interior point method for linear programming*, Technical Report SOR 90-3, Princeton University, Department of Civil Engineering and Operations Research, Princeton, NJ, 1990. To appear in *SIAM Journal of Optimization.*

[10] ——, *Starting and restarting the primal-dual interior point method*, Technical Report SOR 90-14, Princeton University, Department of Civil Engineering and Operations Research, Princeton, NJ, 1990.

[11] R. E. MARSTEN, *The design of the XMP linear programming library*, ACM Transactions on Mathematical Software, 7 (1981), pp. 481–497.

[12] N. MEGIDDO, *On finding primal- and dual-optimal bases*, ORSA Journal on Computing, 3 (1991), pp. 63–65.

[13] J. E. MITCHELL, *An interior point column generation method for linear programming with shifted barriers*, RPI Technical Report 191, Rensselaer Polytechnic Institute, Department of Mathematical Sciences, Troy, NY, 1990.

[14] G. L. NEMHAUSER AND L. A. WOLSEY, *Integer and Combinatorial Optimization*, John Wiley and Sons, New York, NY, 1988.

[15] J. K. REID, *A sparsity-exploiting variant of the bartels-golub decomposition for linear programming bases*, Mathematical Programming, 24 (1982), pp. 55–69.