A Modified Schur-Complement Method for Handling Dense Columns in Interior-Point Methods for Linear Programming

KNUD D. ANDERSEN
CORE, Universite Catholique de Louvain

The main computational work in interior-point methods for linear programming (LP) is to solve a least-squares problem. The normal equations are often used, but if the LP constraint matrix contains a nearly dense column the normal-equations matrix will be nearly dense. Assuming that the nondense part of the constraint matrix is of full rank, the Schur complement can be used to handle dense columns. In this article we propose a modified Schur-complement method that relaxes this assumption. Encouraging numerical results are presented.

Categories and Subject Descriptors: G.1.3 [**Numerical Analysis**]: Numerical Linear Algebra—*linear systems (direct and iterative methods)*; G.1.6 [**Numerical Analysis**]: Optimization—*linear programming*

General Terms: Algorithms, Performance

Additional Key Words and Phrases: Cholesky factorization, interior-point methods

1. INTRODUCTION

The most time consuming part in interior-point methods for linear programming (LP) is to find a solution of the following linear system:

$$\begin{bmatrix} D^{-1} & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}, \tag{1}$$

where $A \in \Re^{m \times n}$, $x \in \Re^n$, $b \in \Re^n$, and $D \in \Re^{n \times n}$ is a positive diagonal matrix. The linear system (1) can be reduced to a system of normal

This work was supported by a Danish SNF Research Studentship. Some of it was carried out while the author was a visiting scholar at the Computer Science Department of Stanford University.
Author's address: CORE, Universite Catholique de Louvain, 34 Voie Roman Du Pays, B-1348 Louvain-la-Neuve, Belgium; email: kda@core.ucl.ac.be.

equations:

$$ADA^Tx = ADb. \tag{2}$$

We will not discuss methods for solving (1), but only direct methods for (2). We will assume that A has full row rank, so that ADA^T is a positive definite symmetric matrix. Alternatives for solving this system are the QR factorization of $D^{1/2}A^T$, an LU factorization of ADA^T using row and column interchanges, or a Cholesky factorization of ADA^T. The latter is equal to the LU factorization choosing the pivots on the diagonal. The most successful in practice is the Cholesky factorization $ADA^T = LL^T$, e.g., see Saunders [1994]. The success of the Cholesky factorization in interior-point methods is due to its good numerical properties and the fact that it can be implemented very efficiently. If ADA^T is sparse, the rows of A can usually be reordered once such that the Cholesky factorization of ADA^T is sparse. However a drawback of (2) and the normal-equations approach is that if one or more columns of A is nearly dense, ADA^T and L will be very dense. Clearly such columns have to be handled in a special way. Several methods have been used to do this:

(1) *Schur Complement.* The Schur complement is used to eliminate dense columns from A, such that the remaining part of (2) is sparse [Choi et al. 1990; Marxen 1989].
(2) *Preconditioned Conjugate Gradient.* A PCG method where the preconditioner is a Cholesky factorization of (2), without the dense columns [Adler et al. 1989; Gill et al. 1986].
(3) *Splitting.* "Stretch" the dense columns into two or more pieces [Vanderbei 1991].

According to Saunders [1994], splitting is the most efficient way of handling dense columns in the normal-equations system. The problem with the Schur complement has been lack of numerical stability [Lustig et al. 1991; 1992], which is very important in interior-point methods. In Section 2 we describe a modified Schur complement, and Section 3 gives implementations details. Finally we present results in Section 4.

2. THEORY

In the following we assume that the Cholesky factorization is well known. Let

$$A = [\hat{A}E],$$

where E is a subset of dense columns from A. We assume that $\hat{A}$ has full row rank and that there are few columns in E. We want to solve

$$AA^Tx = \hat{A}\hat{A}^Tx + EE^Tx = b. \tag{3}$$

$$L^T L = \hat{A}\hat{A}^T \tag{8}$$

$$LV = E \tag{9}$$

$$Lp = b \tag{10}$$

$$(V^T V + I)r = V^T p \tag{11}$$

$$L^T x = p - Vr \tag{12}$$

Fig. 1. The Schur-complement method.

Note that without loss of generality the diagonal matrix D can be omitted, because it is only a scaling of the columns in A. Equation (3) is rewritten as

$$\begin{bmatrix} \hat{A}\hat{A}^T & E \\ E^T & -I \end{bmatrix} \begin{bmatrix} x \\ r \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}. \tag{4}$$

By assumption, $\hat{A}\hat{A}^T$ has full rank, so the Cholesky factorization $LL^T = \hat{A}\hat{A}^T$ exists, where L is lower triangular. The linear system (4) is replaced by the system

$$\begin{bmatrix} LL^T & E \\ E^T & -I \end{bmatrix} \begin{bmatrix} x \\ r \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}. \tag{5}$$

If $LV = E$ and $Lp = b$ we obtain

$$L^T x + Vr = p \tag{6}$$

$$V^T L^T x - r = 0 \tag{7}$$

This way of finding a solution to (3) is known as a Schur-complement method (or update), e.g., see Choi et al. [1990] and Heath [1982, 230–231]. The method is formally stated in Figure 1.

The vector r can be found cheaply from (11), because of the low dimension of $(V^T V + I)$. Cholesky factorization of $(V^T V + I)$ is one possibility. Since (11) is equivalent to the least-squares problem

$$\min \left\| \begin{bmatrix} V \\ I \end{bmatrix} r - \begin{bmatrix} p \\ 0 \end{bmatrix} \right\|_2,$$

an alternative is to use a QR factorization of the matrix

$$[V^T\ I]^T.$$

For numerical accuracy and convenience we have tried a third approach, based on an LU factorization of $V^TV + I$ with row and column interchanges for stability. (We use a similar factorization later.) To improve the accuracy of r and L^Tx, we used iterative refinement [Golub and Loan 1989, 253–254] on (11)–(12) and on the full system (3). This approach was tested on LP problems with dense columns and proved to work well in most cases. However, it failed as $\hat{A}\hat{A}^T$ approached a singular matrix. The other two factorizations would eventually fail for the same reason.

In Lustig et al. [1991], L is replaced by the Cholesky factor of a modified system,

$$LL^T = \hat{A}\hat{A}^T + \rho I,$$

where ρ is a positive scalar that ensures the Cholesky factor is well defined. Iterative refinement is then used to compute each column of V. This is computationally expensive, and again it does not provide a stable method for solving (3).

Instead of introducing an arbitrary perturbation to the problem, we propose to modify the linear system that is solved. The modified system has the same solution as the original problem, but is better conditioned. We know that we can make $\hat{A}\hat{A}^T$ of full rank if we add something to the diagonal. Using this observation we rewrite (3) as

$$\begin{bmatrix} (\hat{A}\hat{A}^T + FF^T) & E & F \\ E^T & -I & 0 \\ F^T & 0 & I \end{bmatrix} \begin{bmatrix} x \\ r \\ s \end{bmatrix} = \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}, \tag{8}$$

where $F \in R^{m \times k}$ is a matrix with only one element different from zero in every column, so that the introduction of F only changes the diagonal of $\hat{A}\hat{A}^T$. Therefore the revised matrix has a Cholesky factor with the same sparsity structure as the old. Our modified Schur-complement method can now be outlined as in Figure 2.

In the procedure for calculating the Cholesky factorization (14) we check for rank deficiency by monitoring the size of the diagonal elements. If a diagonal element is small we add something to it and introduce a new column in F.

Dr. J. Gondzio has pointed out that Stewart [1974] describes how to handle null pivots arising from a fixed (specified) pivot order on a square matrix of full rank. The method described here is related to Stewart's approach.

3. IMPLEMENTATION DETAILS

For general implementation details of a sparse Cholesky factorization, see Andersen and Andersen [1996]. We will only give implementation details related to the modified Schur-complement method in Figure 2.

$$LL^T = \hat{A}\hat{A}^T + FF^T \tag{14}$$

$$LV = E \tag{15}$$

$$LW = F \tag{16}$$

$$Lp = b \tag{17}$$

$$C = \begin{bmatrix} V^T \\ W^T \end{bmatrix} \begin{bmatrix} V & W \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix} \tag{18}$$

$$C \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} V^T p \\ W^T p \end{bmatrix} \tag{19}$$

$$L^T x = p - Vr - Ws \tag{20}$$

Fig. 2. The modified Schur-complement method.

Normally one would assume that V is fully dense. This is true if the columns in E are fully dense. However, as the problem size increases and as columns with more than say 40 nonzeros are dropped, then the columns in E may be sparse, but dense in comparison with the columns in $\hat{A}$. Note that the number of nonzeros per column in an LP model is typically in the range 3–10. We have therefore chosen to treat V as a sparse matrix. V's symbolic structure is calculated only once, as for the Cholesky factor L. Assuming L is known, the nonzero structure of the ith column of V can be found by solving $LV_i = E_i$ symbolically. This clearly gives a worst-case estimate of nonzeros in V_i.

The matrix W is treated the same as V.

To solve system (19) we find an LU factorization of the matrix C using row and column interchanges for stability. A QR factorization of C can also be used, but note that the linear system (19) cannot be formulated as a least-squares problem unless W is empty. In practice it is important for satisfactory numerical accuracy to use iterative refinement on (19)–(20).

The number of columns in F is restricted to at most the number of columns in E. A column is added permanently to F if $l_{ii} < tol_{add}$, where l_{ii} is the ith diagonal element of L. This is found to work well in practice.

Another important issue is to choose which columns should be in E. We keep the number of columns in E relatively small to avoid possible numerical instability and because it is inefficient to drop many columns.

Let n be the number of columns in A and define $\beta = |A|/n$, where $|A|$ is the number of nonzeros in A. Let A_j be the jth column in A, and define $n_j = |A_j|$. E is obtained in the following way:

$$E^1 = \{A_j \mid n_j > 3\beta, \quad n_j \geq \gamma, \quad i = 1, \ldots, n\}$$
$$E^2 = \{A_j \mid n_j > 10\beta, \quad n_j \geq \gamma, \quad i = 1, \ldots, n\}$$
if (columns in E^1) $\leq \eta$
  $E = E^1$
else
  $E = E^2$

where γ ensures that every column in E has a certain amount of nonzeros, and η ensures that the number of columns is not too large. The result is that only relatively dense columns are dropped, and only if A is sparse.

The parameters we have used are $\gamma = 40$, $\eta = 10$, and $tol_{add} = 10^{-10}$.

4. RESULTS

The ideas presented in Section 2 have been implemented and are part of the APOS system (which is a joint work with E. D. Andersen). The codes are OPTMZE, a primal-dual interior-point code for linear programming (this code is part of the XPRESS modeling system from Dash Associates Ltd.), and GOPT (Global Optimizer), a Newton method for minimizing the sum of L1 or L2 norms subject to linear constraints [Andersen 1995]. Both codes use the same Cholesky factorization, described in Andersen and Andersen [1996]. The codes are written in ANSI C and have been ported to several different computers including PCs, an HP 9000/715 workstation, a SUN SPARCstation 10, different SGI workstations, and a CONVEX C3240 vector computer.

All testing presented here was done on a CONVEX C3240 at Odense University. The method was tested on all problems with dense columns that we had access to. This includes some problems from the netlib test set [Gay 1985] and some from another source. The method was also tested on two minimum sum of norms (MSN) problems using GOPT. These are problems "ssu" and "cl399."

The results are shown in Tables I and II. $|AA^T|/2$ is the number of nonzeros in AA^T on and below the diagonal; $|L|$ is the number of nonzeros in L; $|V|$ is the number of nonzeros in V and W; "V col" is the number of columns in V including columns in W; $|W|$ is the number of columns in F; m is the number of rows in A; NSU stands for No Schur Update and WSU for With Schur Update; CPU is total time in CPU seconds including cross-over [Andersen and Ye 1994] to a basic feasible solution and presolving [Andersen and Andersen 1995]; and ITER is the total number of iterations. We see in Table I that V is sparse. For p2, the upper bound on the number of nonzeros in V is 15,763 $\times$ 51 = 803,913, but there are only 29,646 nonzeros, a density of 4%. Storing V sparsely will clearly give a large reduction in memory. If we look at L and ignore the storage for V (it is

Table I. Problem Sizes with WSU and without NSU Schur Update

	$\|AA^T\|/2$		$\|L\|$			$\|V\|$	V col	W col	
Name	NSU	WSU	NSU	WSU	NSU/WSU	WSU	WSU	WSU	m
fit1d	291	291	296	296	1	0	0	0	24
fit1p	196878	627	196878	627	314	8215	24	0	627
fit2d	321	321	324	324	1	0	0	0	25
fit2p	4501500	3000	4501500	3000	1500	36784	25	0	3000
israel	10963	4273	11224	4442	2.5	435	3	0	163
pilot	59844	56641	197346	186783	1.1	4729	12	3	1362
p1	492797	48077	1112409	199145	5.6	5916	6	3	4901
p2	232686	65826	1339156	163840	8.2	29646	51	1	15763
p3	2381	1823	3208	2615	1.2	666	12	3	203
seba(max)	51165	949	53492	1035	51.7	3762	14	0	446
ssu (MSN)	405450	18188	405450	45732	8.9	900	2	1	900
cl399 (MSN)	3103949	3024548	57678407	30001436	1.9	12924	2	1	319466

actually very small in comparison with the nonzeros in L), the reduction in nonzeros is between 2 and 10 for most of the problems where dense columns occur. We see that the reduction in CPU time (see Table II) for the same problems is in the range 1.2 to 6. Clearly for these problems it is beneficial to handle the dense columns separately. Only for two problems did the CPU time increase (pilot and p3), which is consistent with the fact that the reduction in nonzeros in L is low. In all cases standard accuracy is achieved, and only on p1 and ssu did it cost more iterations to get the same accuracy; see Table II. We further observe in Table I that the number of columns in W is very low. The cl399 problem is interesting because the column that is dropped has a density of only (401/319,466) * 100 = 0.13%, but it is dense in comparison with the other columns in the problem. The reduction in L is somewhat small (below a factor of 2). However, a reduction in nonzeros from 57 million to 30 million would make a critical difference on many computers.

The method was also used on fit1d and fit2d because they are dual problems of fit1p and fit2p. We see that the time for fit1d and fit2d is nearly the same as for fit1p and fit2p using the Schur complement. This means that for some LP problems the dense columns can be avoided by finding a solution to the dual problem. However, this cannot be done in general for nonlinear problems such as the two *MSN* examples.

5. CONCLUSION

In this article we propose a modified Schur-complement approach for handling dense columns when a least-squares problem is solved using the normal equations. As noted by other authors, this approach can be used to handle a few dense columns in large sparse LP problems efficiently. However, our modified Schur-complement approach handles the numerical rank or near rank deficiency better than the original approach at a minor cost. This conclusion is confirmed by our (limited) numerical results.

Table II. Run-Times and Number of Iterations with WSU and without NSU Schur Update

	CPU			ITER	
Name	NSU	WSU	NSU/WSU	NSU	WSU
fit1d	10.7	11.5	0.93	21	21
fit1p	115.7	12.2	9.48	16	16
fit2d	104.0	111.1	0.94	22	22
fit2p	11770.2	83.2	141.47	20	20
israel	7.7	6.0	1.28	24	24
pilot	283.5	317.5	0.89	43	43
p1	2720.3	438.4	6.20	43	48
p2	2056.7	944.8	2.18	29	28
p3	3.9	4.3	0.91	16	16
seba(max)	33.5	10.1	3.31	21	21
ssu (MSN)	298.7	37.6	7.9	17	23
cl399 (MSN)	100084	36003	2.8	40	35
Total	117491	37979	3.1		

ACKNOWLEDGMENTS

I thank Erling D. Andersen for reading earlier drafts of this article and the Associate Editor Michael Saunders for doing much work on this article.

REFERENCES

ADLER, I., KARMARKAR, N., RESENDE, M., AND VEIGA, G. 1989. An implementation of Karmarkar's algorithm for linear programming. *Math. Program. 44*, 297–235.

ANDERSEN, E. AND ANDERSEN, K. 1995. Presolving in linear programming. *Math. Program. 71*, 221–245.

ANDERSEN, E. AND ANDERSEN, K. 1996. The APOS LP solver. Tech Rep., CORE, Univ. Catholique de Louvain, Belgium.

ANDERSEN, K. 1996. An efficient Newton barrier method for minimizing a sum of Euclidean norms. *SIAM J. Optim. 6*, 1, 74–95.

ANDERSEN, E. AND YE, Y. 1994. Combining interior-point and pivoting algorithms for linear programming. Tech. Rep., Dept. of Management Sciences, The Univ. of Iowa, Ames, Iowa. In preparation.

CHOI, I., MONMA, C., AND SHANNO, D. 1990. Further development of a primal-dual interior point method. *ORSA J. Comput. 2*, 304–311.

GAY, D. 1985. Electronic mail distribution of linear programming test problems. *COAL Newslett. 13*, 10–12.

GILL, P., MURRAY, W., SAUNDERS, M., TOMLIN, J., AND WRIGHT, M. 1986. On the projected Newton barrier methods for linear programming and an equivalence to Karmarkar's projective method. *Math. Program. 36*, 183–209.

GOLUB, G. AND LOAN, C. V. 1989. *Matrix Computations*. 2nd ed. The John Hopkins University Press, Baltimore, Md.

HEATH, M. 1982. Some extensions of an algorithm for sparse linear least squares problems. *SIAM J. Sci. Stat. Comput. 3*, 2, 223–237.

LUSTIG, I., MARSTEN, R., AND SHANNO, D. 1991. Computational experience with a primal-dual interior point method for linear programming. *Lin. Algebra Appl. 20*, 191–222.

LUSTIG, I., MARSTEN, R., AND SHANNO, D. 1992. On implementing Mehrotra's predictor-corrector interior-point method for linear programming. *SIAM J. Optim. 2*, 3, 435–449.

MARXEN, A. 1989. Primal barrier methods for linear programming. Rep. SOL 89-6, Dept. of Operations Research, Stanford Univ., Stanford, Calif.

SAUNDERS, M. 1994. Major Cholesky would feel proud. *ORSA J. Comput. 6*, 1, 94–105.

STEWART, G. 1974. Modifying pivot elements in Gaussian elimination. *Math. Comput. 28*, 126, 537–642.

VANDERBEI, R. 1991. Splitting dense columns in sparse linear systems. *Lin. Algebra Appl. 152*, 107–117.

Received September 1994; revised January, March, July, and October 1995; accepted November 1995