Available online at www.sciencedirect.com
SCIENCE @ DIRECT•
An International Journal
computers & mathematics with applications
ELSEVIER Computers and Mathematics with Applications 50 (2005) 1157–1166
www.elsevier.com/locate/camwa

The Cholesky Factorization in Interior Point Methods

C. MÉSZÁROS
Computer and Automation Research Institute
Hungarian Academy of Sciences, Budapest
mcsaba@oplab.sztaki.hu

Abstract—The paper concerns the Cholesky factorization of symmetric positive definite matrices arising in interior point methods. Our investigation is based on a property of the Cholesky factorization which interprets "small" diagonal values during factorization as degeneracy in the scaled optimization problem. A practical, scaling independent technique, based on the above property, is developed for the modified Cholesky factorization of interior point methods. This technique increases the robustness of Cholesky factorizations performed during interior point iterations when the optimization problem is degenerate. Our investigations show also the limitations of interior point methods with the recent implementation technology and floating point arithmetic standard. We present numerical results on degenerate linear programming problems of NETLIB. © 2005 Elsevier Ltd. All rights reserved.

Keywords—Cholesky factorization, Interior point methods, Linear programming.

1. INTRODUCTION

The impressive progress in theory and practice of interior point methods (IPM) in the past 15 years raised several questions. The important practical issue, the stability of computations in IPMs, deserved especially great attention in the literature [1–4]. For most interior point algorithms, the major computational task is to solve systems of linear equations with sparse positive definite matrix, which is done by Cholesky decomposition in practice. One of the most important difficulties for IPMs is the ill-conditioning of these linear systems when the method approaches the optimal solution of the optimization problem [5]. It was observed [2,6] that in the degenerate case, if the Jacobian of the active constraints of an inequality-constrained optimization problem is not of full row rank at the optimum, ill-conditioning can break down the Cholesky factorization. The breakdown of the decomposition usually occurs if very small (perhaps zero) pivot elements during the factorization process arise. This problem was recognized by several authors and different solutions were proposed.

One possibility to overcome this situation is performing a modified Cholesky factorization by skipping factors corresponding to small pivots during numerical computations. This can be performed in different ways either by setting the pivot value to a very large number or by setting the elements in the factor to zero [2].

Supported in part by the Alexander von Humboldt Foundation and by the Hungarian Research Fund OTKA T-043276.

 Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
doi:10.1016/j.camwa.2005.08.016

Another possibility is using regularization or "pivot boosting", i.e., increasing the small pivot values to "reasonably" large ones [7–9]. In [6] it was described how the effect of the regularization can be taken into account by using quadratic penalty terms. On the other hand, the analysis in [10] shows that small regularization causes small perturbation in the solution of the system which can be decreased by standard iterative refinement techniques.

The common property of these approaches is that they are very sensitive of how the "small" pivot values are identified. Up to now, only one approach has been discussed in the literature [2], in which the author suggested using a threshold value, relative to the largest diagonal element of the system to be factorized. Unfortunately, this method depends on the scaling of the problem, and therefore, it is not reliable in practice.

In this paper, we investigate the Cholesky factorization of a matrix arising from a *source matrix* S by computing $SS^\top$. In Section 2, we show that the diagonal values of the Cholesky decomposition of $SS^\top$ give information about the degeneracy of S. This property gives some theoretical support for the modified Cholesky factorization. In Section 3, we describe a scaling independent way for identifying small pivot values during the Cholesky decomposition and discuss the consequences of the modified Cholesky factorization in interior point methods. In Section 4, we show numerical results on degenerate linear programming problems. Section 5 contains the conclusions of our investigation.

2. THE CHOLESKY FACTORIZATION

In this section, we concern the Cholesky factorization

$$SS^\top = LL^\top, \tag{1}$$

where $S \in \Re^{m\times n}$ is of full row rank and L is lower triangular. An important application of this decomposition, such as in interior point methods, is to compute projections to the space spanned by the columns of $S^\top$. Usually, the Cholesky factorization is performed by the following standard procedure:

$$\begin{aligned} &\text{for } i = 1,\dots,m,\\ &\quad L_{ii} = \sqrt{\sum_{k=1}^{n} S_{ik}^2 - \sum_{k=1}^{i-1} L_{ik}^2};\\ &\quad \text{for } j = i+1,\dots,m,\\ &\quad\quad L_{ji} = \frac{\sum_{k=1}^{n} S_{ik}S_{jk} - \sum_{k=1}^{i-1} L_{jk}L_{ik}}{L_{ii}}. \end{aligned} \tag{2}$$

Further on, we will use the following definition [11].

DEFINITION 1. *Let $r_1,\dots,r_k$ be row vectors of the same dimension, and $\mathcal{L}\{r_1,\dots,r_{k-1}\}$ the linear subspace spanned by vectors $r_1,\dots,r_{k-1}$. The distance of r_k from the linear subspace $\mathcal{L}\{r_1,\dots,r_{k-1}\}$ is defined as*

$$d\left(r_k, \mathcal{L}\left\{r_1,\dots,r_{k-1}\right\}\right) = \min_{p_1,\dots,p_{k-1}} \left\| r_k - \sum_{i=1}^{k-1} p_i r_i \right\|, \tag{3}$$

where $p_i \in \Re$ and $\|.\|$ denotes the Euclidean norm.

As it is known, the residual vector, $r_k - \sum_{i=1}^{k-1} p_i r_i$, is orthogonal to $\mathcal{L}\{r_1,\dots,r_{k-1}\}$ and when the vectors $r_1,\dots,r_{k-1}$ are linearly independent, then the optimal $p = (p_1,\dots,p_{k-1})^\top$ is unique and can be obtained as

$$p = \left(S_{k-1}S_{k-1}^\top\right)^{-1} S_{k-1} r_k^\top,$$

where $S_{k-1} = [r_1^\top,\dots,r_{k-1}^\top]^\top$ (see [11]).

Further on, let S_i be the matrix consisting of the first i rows of S, that is $S_i = [r_1^\top, \dots, r_{i-1}^\top]^\top$ and L_i the Cholesky factors of $S_i S_i^\top$ for $i = 1, \dots, m$. It is easy to see that

$$L_k = \begin{bmatrix} L_{k-1} & \\ g_k^\top & \alpha_k \end{bmatrix},$$

where

$$L_{k-1} g_k = S_{k-1} r_k^\top \tag{4}$$

and

$$\alpha_k^2 = r_k^\top r_k - g_k^\top g_k. \tag{5}$$

THEOREM 1. *The k^{th} diagonal value of the Cholesky factorization of $SS^\top$ is equal to the distance of vector r_k from the subspace spanned by vectors $r_1, \dots, r_{k-1}$, that is*

$$\alpha_k = d\left(r_k, \mathcal{L}\left\{r_1, \dots r_{k-1}\right\}\right).$$

Although the above theorem is well known in linear algebra (see, e.g., [12]), we give a short proof which will highlight the connection of the algebraic properties of S and the Orchard-Hays relative tolerance [13] for computing α_k^2 in (5).

PROOF. Let us take a closer look on the norm of the residual

$$\begin{aligned}
\left\| r_k - \sum_{i=1}^{k-1} p_i r_i \right\|^2 &= \left\| r_k^\top - S_{k-1}^\top \left(S_{k-1} S_{k-1}^\top\right)^{-1} S_{k-1} r_k^\top \right\|^2 \\
&= \left\| r_k^\top - S_{k-1}^\top L_{k-1}^{-\top} L_{k-1}^{-1} S_{k-1} r_k^\top \right\|^2 \\
&= \left\| r_k^\top - S_{k-1}^\top L_{k-1}^{-\top} g_k \right\|^2 \\
&= \|r_k\|^2 + \left\langle S_{k-1}^\top L_{k-1}^{-\top} g_k, S_{k-1}^\top L_{k-1}^{-\top} g_k \right\rangle + \left\langle S_{k-1}^\top L_{k-1}^{-\top} g_k, -2 r_k^\top \right\rangle \\
&= \|r_k\|^2 + \left\langle S_{k-1}^\top L_{k-1}^{-\top} g_k, S_{k-1}^\top L_{k-1}^{-\top} g_k - 2 r_k^\top \right\rangle.
\end{aligned}$$

Since $S_{k-1}^\top L_{k-1}^{-\top} g_k - r_k^\top$ is orthogonal to $\mathcal{L}\{r_1, \dots, r_{k-1}\}$, thus

$$\left\langle S_{k-1}^\top L_{k-1}^{-\top} g_k, S_{k-1}^\top L_{k-1}^{-\top} g_k - r_k^\top \right\rangle = 0$$

and

$$\begin{aligned}
\left\| r_k - \sum_{i=1}^{k-1} p_i r_i \right\|^2 &= \|r_k\|^2 - \left\langle S_{k-1}^\top L_{k-1}^{-\top} g_k, r_k^\top \right\rangle \\
&= \|r_k\|^2 - \left\langle g_k, L_{k-1}^{-1} S_{k-1} r_k^\top \right\rangle \\
&= \|r_k\|^2 - \|g_k\|^2,
\end{aligned}$$

which together with (5) proves the theorem. ∎

Since the value of $d(r_k, \mathcal{L}\{r_1, \dots, r_{k-1}\})$ depends on the scaling of r_k, later on we will use the relative distance of r_k and $\mathcal{L}\{r_1, \dots, r_{k-1}\}$ which can be defined for $r_k \neq 0$ vector as

$$\hat{d}\left(r_k, \mathcal{L}\left\{r_1, \dots, r_{k-1}\right\}\right) = \frac{d\left(r_k, \mathcal{L}\left\{r_1, \dots, r_{k-1}\right\}\right)}{\|r_k\|}. \tag{6}$$

Note that $\hat{d}(r_k, \mathcal{L}\{r_1, \dots, r_{k-1}\})^2$ is the Orchard-Hays' relative acceptability measure [13] for α_k^2 in (5). Let us observe the following consequence of Theorem 1 in the practical factorization where computations are performed with the finite ϵ_M relative precision.

COROLLARY 1. *If $\hat{d}(r_k, \mathcal{L}\{r_1, \dots, r_{k-1}\}) < \sqrt{\epsilon_M}$, then the computed α_k has no significant digits.*

PROOF. Since
$$\hat{d}\left(r_k, \mathcal{L}\left\{r_1, \dots, r_{k-1}\right\}\right)^2 = \frac{\|r_k\|^2 - \|g_k\|^2}{\|r_k\|^2} < \epsilon_M,$$
the error when computing $\|r_k\|^2 - \|g_k\|^2$ is of the same (or larger) order of magnitude than the result, and therefore, in the computed value of $\sum_{i=1}^n S_{ik}^2 - \sum_{i=1}^{k-1} L_{ik}^2$ the round-off error is the dominating factor. ∎

The relative precision of today's commonly-used double precision arithmetic is $\epsilon_M \approx 10^{-16}$, and thus for a matrix S in which $\hat{d}(r_k, \mathcal{L}\{r_1, \dots, r_{k-1}\}) < 10^{-8}$ the k^{th} and forthcoming computed Cholesky factors of $SS^\top$ mostly depend on the previous round-off errors and on the sequence of floating point operations performed in (2); therefore, its use provides unreliable results. In the next section, we will investigate this situation in interior point methods.

3. THE MODIFIED CHOLESKY FACTORIZATION AND INTERIOR POINT METHODS

In this section, we introduce a modified version of (2) which handles ill-conditioned cases by skipping the unreliable steps of the factorization. A similar approach was introduced in [2], which differs from ours in the determination of the threshold value of the pivot elements. Our version uses an $\epsilon > 0$ tolerance and an adaptive rule in contrast to the method described in [2] where the threshold value is fixed at the beginning of the factorization. Our modified Cholesky decomposition performs the following steps:

$$\begin{aligned}
\text{for} \quad & i = 1, \dots, m, \\
& g_i = \sum_{k=1}^{i-1} L_{ik}^2, \\
& f_i = \sum_{k=1}^{n} S_{ik}^2, \\
& L_{ii} = \begin{cases} \infty, & \text{if } (1-\epsilon) f_i \le g_i, \\ \sqrt{f_i - g_i}, & \text{otherwise;} \end{cases} \\
\text{for} \quad & j = i+1, \dots, m, \\
& L_{ji} = \frac{\sum\limits_{k=1}^{n} S_{ik} S_{jk} - \sum\limits_{k=1}^{i-1} L_{jk} L_{ik}}{L_{ii}}.
\end{aligned} \tag{7}$$

Let us note that the pivot element is computed in the above algorithm in such a way that the negative and positive terms are accumulated separately and added together at the end with the Orchard-Hays relative tolerance [13]. This idea proved to be numerically quite efficient in other contexts [14].

We selected $\epsilon = 10^{-15}$ which provides at least one significant digit for all the accepted results of $f_i - g_i$ on standard floating point arithmetic. Our procedure skips all factors, for which the computed diagonal (pivot) value is definitely unreliable. Let us note that the original problem in most of the applications is to compute projections onto the space of $S^\top$ and the theorem in Section 2 shows that skipping the rows as described does not hurt the result too much because the rows of S, the distance of which to the subspace spanned by the remaining rows is less than $\sqrt{\epsilon} \approx 3 \cdot 10^{-8}$, are omitted.

In the following part of this section, we will consider the linear programming problem and primal-dual log barrier interior point method for further investigation. It is to be noted, however, that this choice has notational consequences only and one can draw conclusions similar to ours when considering the general nonlinear programming problem or another interior point approach.

Let us consider the linear programming problem as

$$\begin{aligned} \min c^{\top} x, \\ Ax = b, \\ x \geq 0, \end{aligned} \tag{8}$$

where $x, c \in \Re^n$, $A \in \Re^{m \times n}$ is of full row rank, and furthermore, $b \in \Re^m$. The logarithmic barrier problem corresponding to (8) is

$$\begin{aligned} &\min c^{\top} x - \mu \sum_{i=1}^{n} \ln x_i, \\ &Ax = b, \qquad x > 0, \end{aligned} \tag{9}$$

where μ is a positive scalar barrier parameter. A log barrier interior point method approaches the optimal solution of (8) by a sequence of barrier problems (9), while the barrier parameter is decreased to zero. Following the classical introduction of the primal-dual log barrier method, the algorithm can be derived by applying Newton's method to solve the Karush-Kuhn-Tucker system of (9). It is easy to see that the computational task of the resulting method in the k^{th} iteration can be reduced to a solution of a system of linear equations whose matrix is $AD^kA^{\top}$ where D^k is a positive definite diagonal *scaling* matrix and its diagonal values are

$$D^k_{ii} = \frac{\left(x^k_i\right)^2}{\mu} \tag{10}$$

if the k^{th} iterate lies on the central path, i.e., if the iterate is perfectly centered. See [6] for details. Thus, in our case at the k^{th} iteration the source matrix is defined as $S = A\hat{D}^k$, where $\hat{D}^k(\hat{D}^k)^{\top} = D^k$.

3.1. Interior Point Methods and Problem Scaling

In this section, we consider the numerical effects of scaling the LP problem when applying interior point methods. We will consider the behavior of matrix $AD^kA^{\top}$ and its Cholesky factorization $LL^{\top} = AD^kA^{\top}$. Clearly, applying column scaling on problem (8) does not affect the matrix $AD^kA^{\top}$. Scaling the problem column-wise with the full-rank diagonal matrix $T \in R^{n \times n}$ at iteration k results in $\tilde{A} = AT$, $\tilde{x}^k = T^{-1}x^k$, thus $\tilde{D}^k = T^{-2}D^k$ and $\tilde{A}\tilde{D}^k\tilde{A}^{\top} = AD^kA^{\top}$. Row scaling, however, changes the numerical values in the matrix and may affect the modified Cholesky factorization. Let us consider the full-rank diagonal matrix $T \in R^{m \times m}$ and the scaled problem at the k^{th} iteration: $\tilde{A} = TA$, $\tilde{x}^k = x^k$ and thus $\tilde{D}^k = D^k$. Since $\tilde{A}\tilde{D}^k\tilde{A}^{\top} = TAD^kA^{\top}T$, it is easy to see that $\tilde{L} = TL$, and therefore, methods based on fix pivot tolerances, like the one described in [2], may skip different factors depending on the diagonal values in T. Our pivot tolerance at the l^{th} step of the factorization is based on the quantity

$$\hat{d}\left(r_l, \mathcal{L}\left\{r_1, \ldots, r_{l-1}\right\}\right).$$

Since $\mathcal{L}\{r_1, \ldots, r_{l-1}\} = \mathcal{L}\{\tilde{r}_1, \ldots, \tilde{r}_{l-1}\}$ our approach is not sensitive to the row scaling because

$$\hat{d}\left(r_l, \mathcal{L}\left\{r_1, \ldots, r_{l-1}\right\}\right) = \hat{d}\left(\tilde{r}_l, \mathcal{L}\left\{\tilde{r}_1, \ldots, \tilde{r}_{l-1}\right\}\right).$$

The modified Cholesky factorization introduced by Wright [2] examines the value of L_{ii} relative to the largest diagonal element of $SS^{\top}$ in (7). This may cause different pivot rejections with different row scalings. Wright's method is equivalent with ours in the case when S is scaled such that all diagonal entries in $SS^{\top}$ are equal. Since applying row scaling on $A\hat{D}^k$ in every iteration of the interior point method is impractical due to its computational cost, in the context of interior point methods our modified Cholesky factorization is not equivalent with the one described in [2].

3.2. Special Cases

In case the set of columns corresponding to variables having positive value at the optimum of (8) is not of full row rank, i.e., if the optimization problem is primal degenerate, $A\hat{D}^k$ will converge to a singular matrix. In this situation the standard Cholesky factorization will break down because during iterations some row vectors of $A\hat{D}^k$ converge into the subspace spanned by the other row vectors. Our proposed modification of the Cholesky factorization will handle this situation well by skipping the rows in $A\hat{D}^k$ which appear to be linearly dependent form the set of the remaining rows. The analysis in [2] shows that the obtained search directions are sufficiently accurate in this case.

It is to be noted, however, that degeneracy is not the only phenomenon which may break down the Cholesky factorization in interior point methods. Let us consider the optimal complementarity partition of problem (8) as (B, N) where B is the set of indices of variables having nonzero value at the optimum and N is its complement. Furthermore, let A_B denote the matrix consisting of the columns of A whose index belongs to B. It is easy to see that in case A_B is ill-conditioned, $A\hat{D}^k$ will behave in a manner similar to the degenerate case. Let us demonstrate this by the following example:

$$\begin{array}{lccccccccc}
\min & & & & & & & x_4, & & \\
& x_1 & & & & & & & = & 1, \\
& & & x_2 & + & x_3 & & & = & 2, \\
& x_1 & + & x_2 & + & (1+10^{-8})\,x_3 & + & x_4 & = & 3+10^{-8}, \\
& & & & & x_1, x_2, x_3, x_4 \geq 0. & & & &
\end{array}$$

It is easy to see that the unique optimal solution of the above problem is $x = (1,1,1,0)^{\mathsf{T}}$. Let us observe that with perfectly centered primal-dual iterates, $A\hat{D}^k$ converges to

$$\frac{1}{\sqrt{\mu}}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1+10^{-8} & 0 \end{bmatrix},$$

thus, for a μ significantly small, the measure $\hat{d}(r_3, \mathcal{L}\{r_1, r_2\})$ falls bellow 10^{-8} although r_3 does not converge to $\mathcal{L}\{r_1, r_2\})$. Hence, although the optimization problem is not degenerate, the Cholesky factorization may break down. As expected, our state-of-the-art interior point optimizer, called BPMPD [15] (with disabled presolve) was not able to correctly solve the above problem and stopped at the primal iterate $(1, 10^{-8}, 2-10^{-8}, 10^{-12})$ with primal infeasibility of order of 10^{-8}.

We also applied another efficient interior point implementation, PCx version 1.1 [16], available freely from `http://www-fp.mcs.anl.gov/otc/Tools/PCx/Windows/`, on the above problem. The result provided by PCx (with disabled presolve) was $(1, 10^{-3}, 2-10^{-3}, 10^{-11})$ which is also far from the true optimum of the problem. Let us note that, due to the ill-conditioning of A_B, a qualitatively better solution cannot be expected either from simplex-based approaches.

Another situation is that the positive components of the optimal solution can be of different orders of magnitude, i.e., if the problem is poorly scaled. This phenomenon can also cause breakdown of the Cholesky factorizations in interior point methods, even if the matrix of the linear program is well conditioned and the problem is nondegenerate. To demonstrate this case let us consider the following example:

$$\begin{array}{lcccccccc}
\min & & & & & x_3, & & \\
& x_1 & & & & & = & 10^5, \\
& x_1 & & -x_2 & + & x_3 & = & 10^{-5}, \\
& x_1 & + & -x_2 & + & 2x_3 & = & 2\cdot 10^{-5}, \\
& & & x_1, x_2, x_3 \geq 0. & & & &
\end{array}$$

Note that the matrix of the above problem is well conditioned and the unique optimum of the problem is $(10^5, 10^5, 10^{-5})$. It is to be observed that with perfectly centered iterates $A\hat{D}^k$ converges to matrix

$$\frac{1}{\sqrt{\mu}}\begin{bmatrix} 10^5 & 0 & 0 \\ 10^5 & 10^5 & 10^{-5} \\ 10^5 & 10^5 & 2\cdot 10^{-5} \end{bmatrix},$$

thus, $\hat{d}(r_3, \mathcal{L}\{r_1, r_2\})$ can decrease under 10^{-8} during iterations, and therefore, the Cholesky factorization may break down. On this problem our solver (without presolve) was not able to compute sufficiently accurate Cholesky decompositions after the third iteration when the value of x_3 decreased under $5\cdot 10^{-5}$. Similarly, the solver PCx reported inaccurate factorizations during the last iterations.

It is to be noted that this latter situation does not result in numerical problems in the factorization process of the algorithms based on the simplex method, because A_B is well conditioned. The poor scaling of the problem appears in the sensitivity of the optimality tolerances by these approaches.

4. NUMERICAL RESULTS

In this section, we show our numerical results on degenerate real-life linear programs. We implemented the supernodal loop unrolling variant [17] of the left looking Cholesky factorization [18]

Table 1. Results on degenerate Netlib problems.

Problem Name	No. of Rows	Skip. Rows	Abs. Primal Infeas.	Rel. Primal Infeas.	Rel. Duality Gap
80bau3b	1965	3	4.0E−09	1.0E−14	2.1E−09
bnl2	1059	9	1.6E−09	2.5E−12	2.2E−10
boeing1	297	7	1.5E−08	4.5E−15	3.0E−11
boeing2	124	13	3.2E−08	1.5E−13	3.1E−11
capri	137	1	5.5E−10	1.3E−13	1.0E−09
cycle	1026	159	1.0E−09	1.7E−13	1.7E−11
d6cube	402	1	1.8E−10	4.5E−14	7.8E−10
degen2	383	52	9.0E−11	2.8E−12	7.2E−11
degen3	1406	203	5.6E−08	1.3E−12	6.6E−11
dfl001	3791	68	7.7E−09	1.9E−11	7.5E−09
forplan	80	12	2.3E−07	2.7E−14	5.8E−10
greenbea	1168	30	2.9E−06	9.1E−11	1.4E−09
greenbeb	1171	6	4.3E−08	5.6E−11	4.1E−09
lotfi	106	1	1.8E−08	9.4E−15	1.7E−11
maros	494	13	1.4E−08	7.0E−13	1.0E−10
modszk1	447	19	1.5E−11	2.2E−16	1.0E−10
pilot-ja	651	1	8.4E−11	1.2E−13	3.7E−09
qap08	894	41	1.0E−09	1.8E−10	1.1E−10
qap12	3192	141	2.7E−08	9.3E−09	6.0E−10
recipe	62	1	1.0E−10	1.4E−12	5.4E−11
sctap3	1346	1	1.7E−13	1.4E−15	6.6E−11
ship04l	288	2	2.6E−10	8.3E−14	5.2E−11
ship12l	610	1	6.7E−11	2.5E−13	3.0E−10
sierra	1083	5	5.4E−07	7.5E−15	1.2E−10
standata	232	32	1.5E−12	3.1E−15	1.3E−10
standgub	232	32	1.5E−12	3.1E−15	1.3E−10
standmps	336	40	5.8E−11	2.0E−14	1.1E−11

Table 2. Results on degenerate Kennington problems.

Problem Name	No. of Rows	Skip. Rows	Abs. Primal Infeas.	Rel. Primal Infeas.	Rel. Duality Gap
cre-a	2760	60	8.1E−10	1.8E−13	1.4E−10
cre-b	4995	37	7.4E−08	6.4E−12	3.3E−09
cre-c	2168	17	7.1E−10	2.2E−13	1.1E−10
cre-d	3791	19	7.9E−09	1.4E−12	1.1E−09
ken-13	13141	2	1.6E−11	5.9E−16	1.1E−10
pds-10	8410	31	2.4E−08	1.0E−13	2.6E−11

Table 3. Detection of dependent rows.

Problem Name	Relative Tolerance	Wright [2]	Pivoting
bore3d	2	0	2
degen2	2	0	2
degen3	2	1	2
dfl001	13	6	13
qap08	170	56	170
qap12	398	145	398
scorpion	29	12	29
shell	1	1	1
sierra	10	3	10
wood1p	1	0	1
cre-a	5	0	5
cre-b	8	2	8
cre-c	5	0	5
cre-d	8	2	8
ken-07	49	12	49
ken-11	121	15	121
ken-13	169	23	169
pds-02	11	2	11
pds-06	11	2	11
pds-10	11	3	11

with the modification described in the previous section in our primal-dual log barrier solver. In our implementation, we store the inverse of the Cholesky factors; thus, in factors skipped we set the corresponding diagonal element to zero.

Tables 1 and 2 present the problems of Netlib and Kennington problem sets [19] in which our modified Cholesky factorization skipped any rows in our interior point implementation. Figures given include the number of rows after presolve, number of skipped factors in the last decomposition, final absolute and relative primal infeasibilities, and the relative duality gap. The relative primal infeasibility and relative duality gap are defined as

$$\frac{\|Ax - b\|}{\|b\| + 1} \quad \text{and} \quad \frac{b^\top y - c^\top x}{|c^\top x| + 1}$$

for an (x, y) where x is an approximate solution of (8) and y is an approximate solution for the dual problem $\{\max_{y \in \Re^m} b^\top y \mid y^\top A \leq c\}$.

Let us note that the special case, when the constraint matrix of the LP problem is not of full row rank, the problem is degenerate but this degeneracy can be detected and removed at the beginning of the algorithm. In our implementation this is done during the computation of the starting point: we compute one factorization with $D^0 = I$ with drop tolerance $\epsilon = 10^{-12}$. All rows corresponding to dropped factors during the Cholesky decomposition are removed from the optimization problem after checking their feasibility. Let us note that as a result of this procedure, the rows of which distance to the linear space of the remaining rows is less than 10^{-6} are dropped from the problem. During iterations the drop tolerance used is $\epsilon = 10^{-15}$ and rows of the original problem corresponding to skipped factors in the Cholesky decomposition are not removed.

In the next experiment, we consider the problems which have dependent rows. The true number of dependent rows is computed by the pivoting procedure of [20]. In Table 3, we compare the number of dependent rows detected by our approach, by that of [2] with the default tolerance of $\epsilon = 10^{-30}$ and by the pivoting algorithm [20].

The results presented indicate that our interior point implementation, with the modified Cholesky decomposition using our adaptive threshold tolerance, handles degenerate problems and detects rank deficiency reliably. In all cases, our solver was able to achieve the desired 10^{-8} relative duality gap and relative infeasibility level.

5. CONCLUSIONS

In this paper, we investigated the Cholesky factorization of symmetric positive definite systems which arise from a matrix S by computing SS^T. Our investigation was based on a property of the Cholesky decomposition which shows that the k^{th} diagonal element of the Cholesky decomposition is equal to the distance of the k^{th} row of S and the linear space spanned by the preceding rows. We pointed out that this property gives a relationship between the Orchard-Hays relative tolerance when computing the pivot elements of the factorization and the degeneracy of the underlying matrix. Based on our investigations, the use of the Orchard-Hays relative tolerance for computing the pivot elements in the modified Cholesky decomposition was proposed.

Since the Cholesky decomposition is perhaps the most important operation in recent interior point implementations, the practice of IPMs is a natural area for the application of our results. We showed that our modified Cholesky factorization can reliably handle matrices arising from the application of interior point methods on degenerate optimization problems. The main advantage of the proposed pivot criteria over the ones proposed so far is its independence of problem scaling. Unfortunately, degeneracy is not the only source of the breakdown of the Cholesky decomposition in IPMs. We pointed out that nondegenerate problems with well-conditioned constraint set can result in numerical troubles in the Cholesky factorization of interior point methods. We showed situations which are impossible to handle with the modifications of the standard decomposition scheme. We demonstrated that in these cases recent interior point implementations are not able to solve very small optimization problems either. Techniques capable of successfully handling situations like this in the sparse context require further research.

REFERENCES

1. M.H. Wright, Some properties of the Hessian of the logarithmic barrier function, *Math. Programming* **67**, 265–295, (1994).
2. S.J. Wright, Modified Cholesky factorizations in interior-point algorithms for linear programming, *SIAM Journal on Optimization* **9** (4), 1159–1191, (1999).
3. S.J. Wright, Stability of linear algebra computations in interior-point methods for linear programming, *SIAM J. Matrix Analysis and Applications* **18**, 191–222, (1997).
4. S.J. Wright, Effects of finite-precision arithmetic on interior-point methods for nonlinear programming, *SIAM Journal on Optimization* **12** (1), 36–78, (2001).
5. O. Güler, D. den Hertog, C. Roos, T. Terlaky and T. Tsuchiya, Degeneracy in interior point algorithms for linear programming: A survey, *Annals of Operations Research* **46**, 107–138, (1993).
6. E.D. Andersen, J. Gondzio, Cs. Mészáros and X. Xu, Implementation of interior point methods for large scale linear programs, In *Interior Point Methods of Mathematical Programming*, (Edited by T. Terlaky), pp. 189–252, Kluwer Academic, (1996).
7. M.A. Saunders, Cholesky-based methods for sparse least squares: The benefits of regularization, In *Linear and Nonlinear Conjugate Gradient-Related Methods*, (Edited by L. Adams and J.L. Nazareth), pp. 92–100, SIAM, (1996).

8. R.J. Vanderbei, LOQO user's manual, Technical Report SOR-96-07, Princeton University, School of Engineering and Applied Science, Dept. of Civil Engineering and Op. Res., (1996).
9. A. Gupta, WSMP: Watson sparse matrix package, Technical Report RC21888, IBM Research Report, IBM T.J. Watson Research Center, (November 2000).
10. Cs. Mészáros, On free variables in interior point methods, *Optimization Methods and Software* **9**, 121–139, (1997).
11. G.H. Golub and C.F. Van Loan, *Matrix Computations*, Second Edition, The John Hopkins University Press, Baltimore, (1989).
12. C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, (1974).
13. W. Orchard-Hays, *Advanced Linear Programming Computing Techniques*, McGraw-Hill, (1968).
14. I. Maros and Cs. Mészáros, A numerically exact implementation of the simplex method, *Annals of Operations Research* **58**, 3–17, (1995).
15. Cs. Mészáros, The efficient implementation of interior point methods for linear programming and their applications, Ph.D. Thesis, Eötvös Loránd University of Sciences, (1996).
16. J. Czyzyk, S. Mehrotra and S.J. Wright, PCx user guide, Technical Report OTC 96/01, Optimization Technology Center, Argonne National Laboratory and Northwestern University, (1996).
17. Cs. Mészáros, Fast Cholesky factorization for interior point methods of linear programming, *Computers Math. Applic.* **31** (4/5), 49–54, (1996).
18. A. George and J.W.H. Liu, *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ, (1981).
19. D.M. Gay, Electronic mail distribution of linear programming test problems, *COAL Newsletter* **13**, 10–12, (1985).
20. E.D. Andersen, Finding all linearly dependent rows in large-scale linear programming, *Optimization Methods and Software* **6**, 219–227, (1995).