SOLVING REGULARIZED LINEAR PROGRAMS USING BARRIER METHODS AND KKT SYSTEMS*

MICHAEL A. SAUNDERS† AND JOHN A. TOMLIN‡

Abstract. We discuss the solution of regularized linear programs using a primal-dual barrier method. Our implementation is based on indefinite Cholesky-type factorizations of full and reduced KKT systems. Regularization improves the stability of the Cholesky factors and allows infeasibility to be detected, but it affects the efficiency of Cross-over to Simplex (to obtain a basic solution to the original problem). We explore these effects by running OSL on the larger Netlib examples.

Key words. barrier methods, interior methods, linear programming, quadratic programming, regularization, KKT systems, Cholesky factors

AMS subject classifications. 90C05, 90C06, 90C20, 65F05, 65F50

1. Introduction. We consider primal-dual interior methods (barrier methods) for solving sparse linear programs of the form

$$\underset{x}{\text{minimize}} \quad c^T x \quad \text{subject to} \quad Ax = b, \quad l \le x \le u. \tag{1}$$

Most of the computational work lies in solving large indefinite systems of linear equations (KKT systems) to obtain search directions. We focus on techniques for making these solves stable and efficient.

Following Vanderbei [Van95], we employ sparse Cholesky-type factorizations rather than more stable indefinite solvers (which are typically less efficient [FM93, DR95]). To make best use of existing and future Cholesky packages, we perturb or "regularize" the LP problem as in Gill *et al.* [GMPS94]; see (4) below. This gives KKT matrices of the form

$$K = \begin{pmatrix} -H & A^T \\ A & \delta^2 I \end{pmatrix}, \qquad H \equiv D_x + \gamma^2 I, \tag{2}$$

where γ and δ are specified scalars and D_x is a positive semidefinite diagonal matrix that changes every iteration. If γ and δ are sufficiently positive, the triangular factorization

$$PKP^T = LDL^T, \qquad D \text{ diagonal but indefinite} \tag{3}$$

exists for any permutation P, even in the presence of rounding error. Thus, P may be chosen purely to preserve sparsity, and the same P may be used for all iterations. In effect, K can be treated as if it were positive definite.

We have three aims in the present work. First, we explore the use of *reduced KKT systems* [GMPS94]. These are formed by pivoting "manually" on some of the diagonals of H (associated with sparse columns in A) and applying a sparse Cholesky

* IBM Research Report RJ 10064 and Stanford SOL Report 96-4, December 1996.
Presented at the 5th SIAM Conference on Optimization, Victoria, BC, Canada, May 20–22, 1996. This research was partially supported by Office of Naval Research grants N00014-94-C-007 and N00014-96-1-0274 and by National Science Foundation grant DMI-9500668.

† Department of EES & Operations Research, Stanford University, Stanford, CA 94305-4023 (mike@SOL-michael.stanford.edu).

‡ IBM Almaden Research Center, San Jose, CA 95120-6099 (tomlin@almaden.ibm.com).

package to the remaining indefinite matrix. We find that some such reduction is usually more efficient than *full* reduction to the familiar matrix $AH^{-1}A^T + \delta^2 I$, but the optimal choice is problem-dependent.

Second, we vary the regularization parameters γ and δ over a wide range, to determine values that give stable performance in practice. On scaled problems with machine precision $\epsilon \approx 10^{-16}$, we find that $\gamma = \delta = 10^{-3}$ is always reliable and that smaller values often suffice. In addition, $\delta = 1$ proves effective on infeasible models.

Finally, we examine the effect of regularization on the "Cross-over to Simplex", i.e., solution of the original (unperturbed) problem by the simplex method [Dan63], starting from the barrier solution. We find that "sufficiently stable" values of γ and δ do not affect Cross-over greatly in most cases.

2. Regularized LP. Most of our discussion applies to regularized linear programs of the form

$$\begin{array}{lll} \underset{x,\,p}{\text{minimize}} & c^Tx + \frac{1}{2}\|\gamma x\|^2 + \frac{1}{2}\|p\|^2 & \\ \text{subject to} & Ax + \delta p = b, & l \le x \le u, \end{array} \tag{4}$$

where the scalars γ and δ are specified and are usually "small". We assume that the problem has been scaled so that $\|A\| \approx 1$, $\|b\| \approx 1$ and $\|c\| \approx 1$. In most of our experiments, γ and δ range from 10^{-2} to 10^{-6}.

Problem (4) is really a convex quadratic program. Throughout, we could replace the term $\frac{1}{2}\|\gamma x\|^2$ by $\frac{1}{2}x^TQx$, where Q is symmetric and positive definite. The terms $\gamma^2 I$ below would become Q. We envisage $Q = Q_0 + \gamma^2 I$, where Q_0 is positive semidefinite with perhaps many empty rows and columns.

Also, the term δp could be replaced by Mp for any matrix M. All terms $\delta^2 I$ below would become MM^T. For example, M could be diagonal with all diagonal entries larger than some positive δ. Our statements about stability and sparsity are true for that case.

Note that setting $\delta = 1$ leads to a meaningful algorithm for solving bound-constrained least-squares problems of the form

$$\underset{x}{\text{minimize}} \quad c^Tx + \frac{1}{2}\left\| \begin{pmatrix} A \\ \gamma I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2 \quad \text{subject to} \quad l \le x \le u. \tag{5}$$

This has proved successful in the context of Basis Pursuit methods for de-noising images [Chen95].

Finally, note that (4) is a completely general form for linear programs, suitable for use with the industry-standard MPS format. The matrix A generally contains many unit columns associated with slack variables on inequality rows, and x includes such variables. Slacks on equality rows are specifically excluded (but the δ term covers all rows).

In LOQO [Van94], Vanderbei includes a slack for every row of A, with zero upper and lower bounds on equality rows. Algorithmically, this provides much of the effect of our δ regularization without perturbing the problem. (However, the effect diminishes as the solution is approached.) The γ regularization could be included explicitly because LOQO handles quadratic programs, but it is not yet standard practice for linear programs. Precautions must therefore be taken in LOQO when P is chosen in the Cholesky-type factorization (3).

3. The Newton equations. Following Megiddo [Meg89], Mehrotra [Meh90], Lustig *et al.* [LMS92], Forrest and Tomlin [FT92], Kojima *et al.* [KMM93] and others, we apply an infeasible primal-dual predictor-corrector algorithm to problem (4). The nonlinear equations defining the central trajectory are $p = \delta y$ and

$$
\begin{array}{rcll}
x - s & = & l, & \\
x + t & = & u, & \\
SZe & = & \mu e, & \mu > 0, \\
TWe & = & \mu e, & s, t, z, w > 0, \\
A^T y + z - w & = & c + \gamma^2 x, & \\
Ax + \delta^2 y & = & b, &
\end{array} \tag{6}
$$

where e is a vector of ones, $S = \mathrm{diag}(s_j)$, and similarly for T, Z, W. (If l and u contain infinite entries, the corresponding equations are omitted.)

The primal-dual algorithm uses Newton's method to generate search directions from equations of the form

$$
\begin{array}{rcccl}
\Delta x - \Delta s & = & \hat{u} & = & (l + s) - x, \\
\Delta x + \Delta t & = & \hat{v} & = & (u - t) - x \\
S\Delta z + Z\Delta s & = & g & = & \mu e - Sz, \\
T\Delta w + W\Delta t & = & h & = & \mu e - Tw, \\
-\gamma^2 \Delta x + A^T \Delta y + \Delta z - \Delta w & = & d & = & c + \gamma^2 x - A^T y - z + w, \\
A\Delta x + \delta^2 \Delta y & = & r & = & b - Ax - \delta^2 y.
\end{array} \tag{7}
$$

Eliminating Δs and Δt gives

$$
\begin{pmatrix} I & -I & -\gamma^2 I & A^T \\ S & & Z & \\ & T & -W & \\ & & A & \delta^2 I \end{pmatrix}
\begin{pmatrix} \Delta z \\ \Delta w \\ \Delta x \\ \Delta y \end{pmatrix}
=
\begin{pmatrix} d \\ \hat{g} \\ \hat{h} \\ r \end{pmatrix}, \tag{8}
$$

where $\hat{g} = g + Z\hat{u}$ and $\hat{h} = h - W\hat{v}$. As shown in a companion paper [ST96], (8) may be reduced in a reasonably stable manner to the KKT system

$$
\begin{pmatrix} -H & A^T \\ A & \delta^2 I \end{pmatrix}
\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}
=
\begin{pmatrix} \bar{d} \\ r \end{pmatrix}, \tag{9}
$$

where $H = S^{-1}Z + T^{-1}W + \gamma^2 I$ and $\bar{d} = d - S^{-1}\hat{g} + T^{-1}\hat{h}$. The eliminated variables are then recovered from

$$
\begin{array}{rcl}
q & = & d + \gamma^2 \Delta x - A^T \Delta y, \\
(S + T)\Delta w & = & \hat{g} + \hat{h} - Sq + (W - Z)\Delta x, \\
\Delta z & = & \Delta w + q, \\
\Delta s & = & \Delta x - \hat{u}, \\
\Delta t & = & \hat{v} - \Delta x,
\end{array} \tag{10}
$$

Individual equations simplify in (9)–(10) if a component of l or u is infinite, but all cases lead to a similar KKT system (9).

4. The effect of γ and δ. Regularization of a problem generally implies some benefit in terms of the problem itself or methods for solving it (e.g., uniqueness of the solution, or algorithmic simplicity).

Here, if γ and δ are both positive, problem (4) is feasible for any data (assuming $l \le u$), and the primal and dual solutions are bounded and unique. Similarly, the Newton equations have a nonsingular Jacobian, and the KKT systems are nonsingular. One is tempted to conclude that *some* degree of regularization is only sensible. Indeed, this is why it was used in [GMPS94].

In practice, moderate regularization produces most of the benefits if the unperturbed problem is feasible. Our results are mainly for that case. Infeasible examples are discussed in §8.

5. Least-squares formulation. When $\gamma > 0$ and $\delta > 0$, the KKT system (9) can be written as the least-squares problem

$$\min \left\| \begin{pmatrix} DA^T \\ \delta I \end{pmatrix} \Delta y - \begin{pmatrix} D\bar{d} \\ r/\delta \end{pmatrix} \right\|, \qquad r_{LS} \equiv D(\bar{d} - A^T \Delta y), \tag{11}$$

where $D \equiv H^{-1/2}$ and r_{LS} is the associated residual vector. Thus, regularization allows us to analyze the KKT systems using known theory about least-squares problems (e.g., [Bjo96]).

5.1. Sensitivity. The sensitivity of Δy to data perturbations depends on the condition of the associated normal matrix N. We have

$$\begin{aligned} N &= AD^2A^T + \delta^2 I, \\ D^2 &= (S^{-1}Z + T^{-1}W + \gamma^2 I)^{-1}, \\ \|N\| &\le \|A\|^2/\gamma^2 + \delta^2, \\ \|N^{-1}\| &\le 1/\delta^2, \\ \mathrm{cond}(N) &\le \|A\|^2/(\gamma\delta)^2 + 1 \;\approx\; \|A\|^2/(\gamma\delta)^2, \end{aligned}$$

and it is sensible to compute Δy with machine precision ϵ as long as $\mathrm{cond}(N) \ll 1/\epsilon$. With $\|A\| \approx 1$, we can expect Δy to be well defined as long as

$$\gamma\delta \gg \sqrt{\epsilon}. \tag{12}$$

The residual vector is less sensitive than Δy to perturbations. Since $\Delta x = -Dr_{LS}$, it seems likely that Δx will also be well defined if (12) holds.

5.2. Stability. Let the KKT system (9) be denoted by

$$Kv = d, \qquad K = \begin{pmatrix} -(D_x + \gamma^2 I) & A^T \\ A & \delta^2 I \end{pmatrix}, \tag{13}$$

and let $\bar{v}$ be the computed solution (used to form the search directions). Also, let *relerr* $\equiv \|v - \bar{v}\|/\|v\|$ denote the relative error in $\bar{v}$. Our particular method for solving with K depends greatly on γ and δ being sufficiently large. For analyses of the accuracy attainable *without* regularization, see Wright [Wri95, Wri96].

As described in the Introduction, we can allow a black-box Cholesky package to compute the indefinite Cholesky-type factorization $PKP^T = LDL^T$ (3) for any ordering P. The necessary stability analysis follows from [GV79, GSS96, Sau96]. In particular, when the factors are used to solve $Kv = d$, *relerr* is bounded in terms of an "effective condition number" of the form

$$\text{Econd}(K) \approx \|A\|^2/(\gamma\delta)^2. \tag{14}$$

With $\|A\| \approx 1$, this means that $\bar{v}$ will have at least some digits of precision if (12) is again satisfied: $\gamma\delta \gg \sqrt{\epsilon}$. When $\gamma = \delta$ and $\epsilon = 10^{-16}$, typical values are as follows:

(15)

γ, δ	Bound on *relerr*
10^{-2}	10^{-8}
10^{-3}	10^{-4}
4×10^{-4}	10^{-2}
10^{-4}	1

We see that the accuracy of the computed search directions may fall rapidly with the size of the regularization parameters. The value $\gamma = \delta = (100\epsilon)^{1/4} \approx 4 \times 10^{-4}$ appears to be as small as one should risk, while $\gamma = \delta = 10^{-3}$ gives a more comfortable margin (four digits of accuracy) with only slighter greater perturbation of the LP.

A concern is that (15) bounds the *norm* of the error, but not necessarily the relative error in individual components of the search direction. Analysis along the lines of [Wri95, Wri96] may be needed, but empirically the individual errors appear to be sufficiently small (because the steplengths and rate of convergence remain good).

6. Reduced KKT systems. The KKT system (9) can often be solved by forcing a block pivot on all of (diagonal) H and allowing a black-box Cholesky package to process the resulting normal equations $AH^{-1}A^T\Delta y = AH^{-1}\bar{d} + r$. This has been standard procedure for most interior-point LP codes (which do not employ regularization), and its numerical stability without regularization is analyzed by Wright [Wri95]. For regularized LP problems, it is clearly stable if γ and δ are sufficiently large, even when A is rank deficient. However, it may be unsatisfactory when $AH^{-1}A^T$ or the Cholesky factor L are excessively dense—commonly as a result of A containing one or more relatively dense columns.

Reduced KKT systems [GMPS94] are a compromise between the full KKT system and the normal equations approach, formed by block pivoting on *part* of H (say H_S). When the regularization parameters are large enough, this partition can be based solely on the sparsity of the associated columns of A.

Let A be partitioned as $(\,A_S \;\; A_D\,)$, where the columns of A_D contain *ndense* or more nonzeros, and partition H, Δx and $\bar{d}$ accordingly. Pivoting on H_S (the first part of H) gives a *reduced KKT system* of the form

$$K_r \begin{pmatrix} \Delta x_D \\ \Delta y \end{pmatrix} = \begin{pmatrix} \bar{d}_D \\ \hat{r} \end{pmatrix}, \qquad K_r \equiv \begin{pmatrix} -H_D & A_D^T \\ A_D & A_S H_S^{-1} A_S^T + \delta^2 I \end{pmatrix}, \tag{16}$$

where $\hat{r} = r + A_S H_S^{-1} \bar{d}_S$. A black-box factorization $PK_rP^T = LDL^T$ may be used to solve for $(\Delta x_D, \Delta y)$. Finally we solve $H_S \Delta x_S = A_S^T \Delta y - \bar{d}_S$.

Acceptable values for *ndense* and P can be specified empirically for most problems. However, a more elaborate procedure might be useful in some situations. For example, *ndense* $= 100$ may be successful for most cases (treating most columns of A as sparse), but if ordering and symbolic factorization of K_r indicate that L will exceed available storage, values such as 50, 20, 15, 10, 5, 1 could be tried in turn. Some intermediate value will probably be optimal. We report on experiments with the *ndense* parameter in §7.1.

7. Numerical results. Our regularized LP implementation is based on OSL, the IBM Optimization Subroutine Library [OSL]. It is the same implementation that we used for experiments on stable reduction of the Newton-type equations (described in a companion paper [ST96]). The following substantial modifications were required to the OSL primal-dual predictor-corrector barrier code [FT92] and the Cholesky factorization routines:

- Incorporation of regularization parameters to accommodate the LP form (4).
- Solution of full or reduced KKT systems (with normal equations becoming a special case).
- Computation of sparse LDL^T factors rather than LL^T, where D is diagonal but indefinite.

Though it would have been possible to investigate the behavior of regularized LPs while retaining the normal equations approach, we felt it important to implement reduced KKT systems in order to handle dense columns directly. The approach also extends naturally to quadratic programs.

As test problems we used a subset of the Netlib collection [Gay85], namely all problems with more than 1000 equations (with the exception of *stocfor3*, which is in inconvenient form). In Tables 1–4, the problems are sorted according to the number of nonzeros in A. All models were scaled but not presolved, to simplify comparison with other implementations.

The tests were performed on an IBM RS/6000 workstation (model 550) with machine precision $\epsilon \approx 10^{-16}$.

7.1. Reduced KKT systems. We first report on the use of reduced KKT systems (16). The "Dense Column threshold" *ndense* affects the number of nonzeros in the Cholesky factor L. Table 1 gives results for $ndense = 0, 5, 10, 20, 50, \infty$. Note that $ndense = \infty$ leads to the normal equations, and $ndense = 0$ corresponds to working with the full KKT system, *except* that we have always treated the logical (unit) columns of A as sparse (and pivoted on their diagonal of H), regardless of the value of *ndense*.

The ordering of the reduced KKT matrices was performed using the original Watson Graph Partitioning package of Gupta [Gup96]. The minimum values for each problem are marked by an asterisk. No single value of *ndense* is best for all problems (despite the large number of ties), but a value of 20 seems to be best (or nearly best) overall and we adopt it from now on. Note that problem *fit2p* has some completely dense columns (3000 elements) and thus will not solve satisfactorily when $ndense = \infty$, since L is dense.

7.2. Regularization. The next important question to be determined numerically is appropriate values for the regularization parameters γ and δ. In these experiments we set both parameters to a common value, to place equal emphasis on minimizing $\|x\|$ and $\|y\|$. We expect to observe two phenomena:

- As γ and δ decrease, the condition of the reduced KKT systems will deteriorate and at some point the indefinite Cholesky factorization may fail.
- As γ and δ increase, the solution of the regularized problem will be further from that of the underlying LP.

Hence, the aim must be to find values that are large enough to maintain stability without perturbing the original problem too much. As discussed above, the value $\gamma = \delta = 4 \times 10^{-4}$ should be (just) safe on scaled data. We also experimented with values of 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3} and 10^{-2}.

Table 1
*Nonzeros in LDL^T factors of the reduced KKT matrix K_r. Smallest values are marked by *.*

	Dense Column threshold					
	0	5	10	20	50	∞
ship12s	23270	6517	6418*	6418*	6418*	6418*
ship12l	34807	11591	12009*	12009*	12009*	12009*
sierra	21626	13906*	13906*	13906*	13906*	13906*
sctap2	24986	17932	15377*	15377*	15377*	15377*
sctap3	35253	24387	22304*	22304*	22304*	22304*
stocfor2	30998	30407	31463	25577*	25577*	25577*
ganges	30714	22848	22848	20802*	20802*	20802*
fit2p	50584	37084*	37084*	37084*	37084*	–
80bau3b	80221	45015	44690	42406*	42406*	42406*
woodw	113895	82924	53968	51237	49435*	49435*
greenbea	148645	149939	88942	76707*	77340	77340
greenbeb	139265	140612	89623	79917	77666*	77666*
bnl2	97878	101447	88817*	88817*	88817*	88817*
cycle	74362	72038	56481	53628*	54951	54951
degen3	169368	166857	160581	127189	124127*	124127*
d2q06c	132616	119412	102579	92249	88024*	88024*
pilot	237222	224441	220709	204847	192903*	198864
pilot87	499841	498203	472463	479501	458917	431253*
dfl001	2430093	1800383	1569817*	1597468	1597468	1597468

Barrier iterations were terminated when the relative gap between the (regularized) primal and dual objectives was below 10^{-8}, or the total complementarity $s^Tz + t^Tw$ (suitably normalized) was below 10^{-12}. The barrier algorithm was also summarily terminated if the Cholesky factorization failed; i.e., if the element to be pivoted on fell below 10^{-12}. In all such cases we have examined so far, the prospective pivot has cancelled to zero.

7.3. Cross-over. After termination of the barrier algorithm, the final primal and dual solution is used by the default OSL Cross-over procedure to obtain an optimal solution and basis for the *original* LP. The time and number of basis changes in Cross-over seemed the most practical way to determine the "closeness" of the regularized solution to a true basic optimum.

Tables 2–4 give iteration counts and times for the Barrier and Cross-over portions of each run. We also give the number of correct digits in the objective value c^Tx for the regularized LP solution (compared to the true optimal objective). This number is marked by an ^{f} if the barrier algorithm terminated with a pivot failure in the Cholesky factorization. We see that there is no failure when γ, $\delta \geq 4 \times 10^{-4}$, as our analysis led us to expect.

The smaller and easier problems in Table 2 are remarkably insensitive to the parameter values, as are most problems in Table 3 (with the exception of *80bau3b*). The larger problems in Table 4 show more sensitivity (particularly *degen3*, *pilot87* and *dfl001*), though surprisingly, the interior regularized solution often seems to provide a good starting solution for Cross-over even when it has been halted by pivot failure.

Again it is difficult to find one ideal set of values for the parameters. Setting $\gamma = \delta = 10^{-2}$ is very safe, but may be too great a perturbation for some problems. (It

Table 2
Effect of γ, δ on Barrier time and Cross-over to Simplex.

		Barrier		Obj	Cross-over		Total
	γ, δ	Itns	Time	Digits	Itns	Time	Time
ship12s	10^{-6}	19	4.1	10.	138	.7	5.4
	10^{-5}	22	4.6	8.	138	.7	5.9
	10^{-4}	24	5.0	5.	138	.7	6.3
	4×10^{-4}	25	5.2	4.	138	.7	6.6
	10^{-3}	25	5.1	3.	138	.7	6.5
	10^{-2}	24	5.0	9.	94	.4	6.0
ship12l	10^{-6}	19	7.2	9.	170	1.3	9.7
	10^{-5}	19	7.2	7.	170	1.3	9.7
	10^{-4}	25	9.1	5.	170	1.3	11.6
	4×10^{-4}	27	9.8	4.	170	1.3	12.3
	10^{-3}	26	9.4	3.	169	1.3	11.9
	10^{-2}	25	9.1	1.	177	1.9	12.2
sierra	10^{-6}	22	5.0	7.	539	1.3	7.0
	10^{-5}	22	5.0	6.	539	1.3	7.1
	10^{-4}	24	5.4	4.	541	1.3	7.5
	4×10^{-4}	23	5.2	2.	539	1.3	7.3
	10^{-3}	23	5.2	2.	527	1.3	7.2
	10^{-2}	23	5.1	1.	519	1.6	7.5
sctap2	10^{-6}	14	3.1	10.	382	.8	4.4
	10^{-5}	14	3.1	10.	380	.7	4.4
	10^{-4}	16	3.4	7.	382	.7	4.7
	4×10^{-4}	17	3.6	5.	247	.6	4.8
	10^{-3}	18	3.8	5.	227	.6	5.0
	10^{-2}	21	4.3	7.	209	.5	5.4
sctap3	10^{-6}	15	4.5	10.	321	1.5	6.7
	10^{-5}	15	4.5	10.	321	1.5	6.7
	10^{-4}	17	4.9	6.	523	1.2	6.9
	4×10^{-4}	19	5.4	5.	321	1.1	7.2
	10^{-3}	20	5.6	4.	320	1.0	7.4
	10^{-2}	22	6.1	2.	304	1.0	7.9
stocfor2	10^{-6}	22	7.2	9.	1103	4.5	12.3
	10^{-5}	22	7.2	10.	1210	4.8	12.6
	10^{-4}	22	7.2	8.	1211	4.8	12.7
	4×10^{-4}	22	7.2	6.	1211	4.7	12.6
	10^{-3}	23	7.4	6.	1211	4.8	12.9
	10^{-2}	23	7.4	3.	1256	5.0	13.1
ganges	10^{-6}	19	4.3	7.	452	2.6	7.5
	10^{-5}	19	4.2	5.	451	2.5	7.3
	10^{-4}	20	4.5	3.	453	2.5	7.5
	4×10^{-4}	21	4.6	2.	455	2.6	7.7
	10^{-3}	21	4.6	2.	439	2.6	7.8
	10^{-2}	24	5.3	9.	220	1.0	6.8

is disastrous here for *dfl001*.) Values of 10^{-4} or smaller lead to failure on some models and give relatively small advantages in terms of closeness of the optimal solutions even when there is no failure. A value of 4×10^{-4} is (just) safe but seems to give little advantage in closeness of the solution over a choice of 10^{-3} in most cases. The latter seems to be the safest choice.

Table 3
Effect of γ, δ on Barrier time and Cross-over to Simplex.

		Barrier		Obj	Cross-over		Total
	γ, δ	Itns	Time	Digits	Itns	Time	Time
fit2p	10^{-6}	23	21.4	10.	6225	282.8	308.2
	10^{-5}	23	21.4	9.	6232	282.7	308.1
	10^{-4}	23	21.4	7.	6225	282.8	308.3
	4×10^{-4}	23	21.4	6.	6232	282.7	308.0
	10^{-3}	23	21.4	5.	6232	282.6	308.1
	10^{-2}	21	19.6	3.	7271	269.3	293.0
80bau3b	10^{-6}	18	16.0	$0.^f$	15602	66.4	84.7
	10^{-5}	19	16.8	$0.^f$	15254	61.3	80.3
	10^{-4}	25	21.1	$0.^f$	14185	56.5	79.8
	4×10^{-4}	56	43.1	6.	1402	11.4	56.9
	10^{-3}	64	48.9	5.	1393	11.6	62.7
	10^{-2}	50	37.7	4.	1737	13.7	53.7
woodw	10^{-6}	28	20.1	7.	3941	61.2	83.7
	10^{-5}	29	20.6	6.	3673	55.1	78.1
	10^{-4}	34	23.8	4.	3319	51.2	77.3
	4×10^{-4}	32	22.5	3.	3344	48.2	73.0
	10^{-3}	31	21.9	2.	3458	54.2	78.4
	10^{-2}	34	23.8	0.	3813	60.9	87.0
greenbea	10^{-6}	37	28.9	3.	1561	8.4	39.4
	10^{-5}	42	32.4	3.	1534	8.3	42.8
	10^{-4}	40	31.0	3.	1523	8.3	41.4
	4×10^{-4}	40	31.0	2.	1611	10.6	43.7
	10^{-3}	39	30.3	2.	1847	17.1	49.4
	10^{-2}	37	28.9	1.	5127	53.8	84.8
greenbeb	10^{-6}	47	35.5	8.	1744	11.4	48.9
	10^{-5}	49	36.8	6.	1777	11.5	50.2
	10^{-4}	46	34.8	4.	1759	11.8	48.6
	4×10^{-4}	50	37.4	3.	1761	12.1	51.5
	10^{-3}	52	38.8	3.	1744	11.7	52.5
	10^{-2}	45	34.1	2.	2875	24.6	60.7
bnl2	10^{-6}	39	27.0	7.	935	3.8	32.3
	10^{-5}	37	25.6	6.	952	3.9	31.0
	10^{-4}	37	25.6	5.	933	3.9	31.0
	4×10^{-4}	38	26.3	4.	908	3.9	31.8
	10^{-3}	38	26.2	3.	931	4.1	31.9
	10^{-2}	35	24.3	2.	1116	7.4	33.3
cycle	10^{-6}	32	16.5	10.	1075	4.9	22.8
	10^{-5}	46	22.7	9.	1102	6.4	30.4
	10^{-4}	39	19.6	4.	1265	7.1	28.1
	4×10^{-4}	38	19.2	4.	657	3.6	24.1
	10^{-3}	41	20.5	4.	1279	8.0	29.8
	10^{-2}	36	18.2	0.	1261	6.6	26.1

It should be remembered that other circumstances may argue for a different choice of γ and δ. For example, these experiments were carried out on un-presolved models with *ndense* = 20. Many of the models are significantly rank deficient (without regularization). This can be substantially remedied by an appropriate Presolve, leading to less likelihood of pivot failure. Similarly, the existence of several “dense” columns

Table 4
Effect of γ, δ on Barrier time and Cross-over to Simplex.

		Barrier		Obj	Cross-over		Total
	γ, δ	Itns	Time	Digits	Itns	Time	Time
degen3	10^{-6}	22	26.7	$2.^f$	1499	11.7	39.7
	10^{-5}	18	22.9	$3.^f$	1033	7.4	31.6
	10^{-4}	26	30.6	5.	1030	8.0	39.8
	4×10^{-4}	26	29.9	6.	948	7.4	38.6
	10^{-3}	26	29.9	5.	948	7.4	38.6
	10^{-2}	23	27.1	3.	1088	8.5	36.9
d2q06c	10^{-6}	35	30.6	9.	884	7.4	40.2
	10^{-5}	32	28.1	8.	902	7.4	37.7
	10^{-4}	33	28.9	6.	898	7.5	38.5
	4×10^{-4}	34	29.8	5.	900	8.0	40.0
	10^{-3}	34	29.8	4.	908	7.8	39.7
	10^{-2}	33	28.9	2.	1210	15.5	46.6
pilots	10^{-6}	48	79.7	8.	766	17.7	100.2
	10^{-5}	50	82.9	6.	661	14.0	99.6
	10^{-4}	44	73.5	4.	589	11.2	87.5
	4×10^{-4}	40	67.4	3.	615	14.9	85.0
	10^{-3}	39	65.8	3.	640	17.4	86.0
	10^{-2}	33	56.5	3.	1275	37.7	97.0
pilot87	10^{-6}	39	208.5	$4.^f$	1256	55.8	268.7
	10^{-5}	39	208.5	$5.^f$	876	34.2	247.0
	10^{-4}	38	199.1	$6.^f$	860	36.1	239.6
	4×10^{-4}	40	208.9	5.	852	27.2	240.6
	10^{-3}	38	199.0	4.	921	29.2	232.6
	10^{-2}	35	184.5	2.	1149	48.0	236.9
dfl001	10^{-6}	34	1129.1	$2.^f$	7398	195.9	1328.0
	10^{-5}	34	1129.8	$2.^f$	6557	167.3	1299.9
	10^{-4}	47	1533.4	5.	5567	112.5	1648.7
	4×10^{-4}	47	1503.5	6.	5887	124.0	1630.4
	10^{-3}	44	1411.0	5.	7321	163.5	1577.4
	10^{-2}	46	1472.0	2.	29602	1033.1	2508.1

in some of the more difficult models (when $ndense = 20$) means that there is a chance of these being permuted to the front of the reduced KKT matrix, leading to pivots on "naked" δ^2 values. This might be avoided by a larger choice of *ndense* (though the aim of regularization is to make any ordering safe). Finally, the parameter choice might be influenced by the relative efficiency of the Cross-over procedure. We have made no attempt to use any of the OSL tuning parameters to affect this, and instead used the defaults.

8. Infeasible problems. If a problem is *known* to be infeasible, we can set $\delta = 1$ and solve the least-squares problem (5). Most emphasis is then placed on trying to satisfy $Ax = b$ (with $l \leq x \leq u$).

We experimented with the non-trivial infeasible Netlib problems, using $\gamma = 10^{-3}$ and two values of δ in order to obtain a comparison. Table 5 summarizes the results.

With $\delta = 1$, we see that the barrier method converged remarkably quickly in all cases, much as we would expect with feasible problems. It appears that this is an efficient method of confirming infeasibility.

With $\delta = 10^{-3}$, all problems behaved satisfactorily except *klein3*. For that problem, the iterates exhibited the symptoms of being not sufficiently interior. (The search directions were computed with reasonable accuracy, but they were not good directions. Many short steps were taken and the iteration limit was exceeded.) Note however that Cross-over finished quite quickly.

Thus in practice, a question remains about models that are being solved for the first time. Infeasibility is best detected by setting $c = 0$, $\delta = 1$ and $\gamma = 10^{-3}$, 10^{-4} or 10^{-5} (say), but if the problem is actually feasible, the final point is not very useful for optimizing the true objective. Conversely, the true c and $\gamma = \delta = 10^{-3}$ are effective if the problem is feasible, but otherwise there is a risk of slow convergence.

Table 5
Results for infeasible models with $\delta = 1$ and 10^{-3} (when $\gamma = 10^{-3}$).

	Barrier Itns		Barrier Time		Cross-over Itns		Cross-over Time	
	$\delta = 1$	10^{-3}	$\delta = 1$	10^{-3}	$\delta = 1$	10^{-3}	$\delta = 1$	10^{-3}
ceria3d	16	46	8.3	30.1	1188	1432	13.4	13.4
chemcom	15	77	1.0	7.6	135	234	.4	.3
cplex1	20	22	8.4	9.1	1255	1899	5.5	10.4
forest	17	23	.2	.3	62	20	.1	.0
gosh	28	52	82.4	142.5	1987	3378	69.7	35.9
gran	22	23	15.4	16.0	2339	899	39.2	17.1
greenbea	35	49	27.4	37.1	505	1855	2.2	13.2
klein2	20	52	1.1	3.4	47	54	.3	.4
klein3	22	100	2.8	17.8	79	87	1.0	1.0
pang	20	27	1.4	1.9	299	156	.9	.3
pilot4i	20	38	2.6	5.7	1010	248	7.1	.8
qual	16	21	.9	1.2	340	177	1.5	.8
reactor	20	40	1.4	3.7	109	122	.3	.2
refinery	16	45	.8	3.1	282	154	.9	.3
vol1	16	20	.9	1.0	382	165	3.9	.4

9. Conclusions. Our experiments show that regularized LPs can be solved effectively by interior methods, using Cholesky-type LDL^T factorizations of full or reduced KKT systems without resorting to more stable indefinite solvers. In general, adequate stability is achieved without severely perturbing the problem (as measured by the effort required by Cross-over to obtain a basic optimal solution).

Use of reduced KKT systems is a way to limit the size of the fundamental linear systems to be solved, while overcoming the difficulties that arise for problems with dense columns. Use of arbitrary density thresholds, while satisfactory in practice, clearly leaves room for improvement. Recognition of the KKT structure as an integral part of the ordering and symbolic factorization is an area we plan to pursue in further research.

For infeasible LP problems (1), the regularized problem (4) is sometimes useful. To confirm infeasibility, the choice $\delta = 1$ appears to be effective if $c = 0$ or if the barrier solution is followed by Cross-over. If $c \neq 0$, setting $\delta = 1$ provides a solution to the least-squares problem (5), but we cannot be sure of satisfying $Ax = b$ closely even if a feasible point exists.

For general LP problems, the homogeneous algorithms of Ye *et al.* [YTM94, XHY96] appear to be very successful in allowing for infeasibility. Incorporation of regularization into that approach seems another promising line of future research.

REFERENCES

[Bjo96] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, PA, 1996.

[Chen95] S. Chen, *Basis Pursuit*, Ph.D. thesis, Dept of Statistics, Stanford University, Stanford, CA, 1995.

[Dan63] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.

[DR95] I. S. Duff and J. K. Reid, *MA47, a Fortran code for direct solution of indefinite symmetric systems of linear equations*, Report RAL 95-001, Rutherford Appleton Laboratory, Oxfordshire, England, 1995.

[FT92] J. J. H. Forrest and J. A. Tomlin, *Implementing interior point linear programming methods in the Optimization Subroutine Library*, IBM Systems J., 31(1), pp. 26–38, 1992.

[FM93] R. Fourer and S. Mehrotra, *Solving symmetric indefinite systems in an interior-point method for linear programming*, Math. Prog., 62 (1993), pp. 15–39.

[Gay85] D. M. Gay, *Electronic mail distribution of linear programming test problems*, Math. Programming Society COAL Newsletter, 13, pp. 10–12, 1985.

[GMPS94] P. E. Gill, W. Murray, D. B. Ponceleón, and M. A. Saunders, *Solving reduced KKT systems in barrier methods for linear programming*, in G. A. Watson and D. Griffiths (eds.), *Numerical Analysis 1993*, Pitman Research Notes in Mathematics 303, Longmans Press, pp. 89–104, 1994.

[GSS96] P. E. Gill, M. A. Saunders and J. R. Shinnerl, *On the stability of Cholesky factorization for quasi-definite systems*, SIAM J. Mat. Anal., 17(1) (1996), pp. 35–46.

[GV79] G. H. Golub and C. F. Van Loan, *Unsymmetric positive definite linear systems*, Linear Alg. and its Appl., 28 (1979), pp. 85–98.

[Gup96] A. Gupta, *Graph partitioning based sparse matrix orderings for interior-point algorithms*, Research Report RC 20467, IBM Watson Research Center, Yorktown Heights, NY, May 1996.

[KMM93] M. Kojima, N. Megiddo and S. Mizuno, *A primal-dual infeasible-interior-point algorithm for linear programming*, Math. Programming 61 (1993), pp. 263–280.

[LMS92] I. J. Lustig, R. E. Marsten and D. F. Shanno, *On implementing Mehrotra's predictor-corrector interior point method for linear programming*, SIAM J. Optim., 2 (1992), pp. 435–449.

[Meg89] N. Megiddo, *Pathways to the optimal set in linear programming*, in N. Megiddo (ed.), *Progress in Mathematical Programming, Interior-Point and Related Methods*, Springer, NY (1989), pp. 131–158.

[Meh90] S. Mehrotra, *On the implementation of a (primal-dual) interior point method*, Report 90-03, Dept of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, 1990.

[OSL] IBM Corporation, *Optimization Subroutine Library: Guide and Reference*, Document SC23-0519-2 (1991), available through IBM branch offices.

[Sau96] M. A. Saunders, *Cholesky-based methods for sparse least squares: The benefits of regularization*, in L. Adams and J. L. Nazareth (eds.), *Linear and Nonlinear Conjugate Gradient-Related Methods*, SIAM, Philadelphia, 1996, pp. 92–100.

[ST96] M. A. Saunders and J. A. Tomlin, *Stable reduction to KKT systems in barrier methods for linear and quadratic programming*, Report SOL 96-3, Dept of EES&OR, Stanford University, Stanford, CA, 1996.

[Van94] R. J. Vanderbei, *LOQO: An interior point code for quadratic programming*, Report SOR-94-15, Dept of Civil Engineering and Operations Research, Princeton University, Princeton, NJ.

[Van95] ———, *Symmetric quasi-definite matrices*, SIAM J. Optim., 5(1) (1995), pp. 100–113.

[Wri95] S. J. Wright, *Stability of linear equations solvers in interior-point methods*, SIAM J. Mat. Anal. Appls., 16 (1995), pp. 1287–1307.

[Wri96] S. J. Wright, ———, *Stability of augmented system factorizations in interior-point methods*, SIAM J. Mat. Anal. Appls., to appear.

[XHY96] X. Xu, P.-F. Hung and Y.Ye, *A simplified homogeneous and self-dual linear programming algorithm and its implementation*, Annals Oper. Res., 62 (1996), pp. 151–171.

[YTM94] Y. Ye, M. J. Todd and S. Mizuno, *An* $O(\sqrt{n}L)$*-iteration homogeneous and self-dual linear programming algorithm*, Math. Oper. Res., 19(1) (1994), pp. 53–67.