PRECONDITIONING PRIMAL-DUAL INTERIOR POINT METHODS FOR LINEAR PROGRAMMING*

TIM REES† AND CHEN GREIF†

Abstract. We investigate the application of a class of preconditioners to the problem of solving linear systems arising from primal-dual interior point algorithms in linear programming. The preconditioner developed in [2, 3] is generalized to the nonsymmetric case and we show that it has the same clustering properties. The preconditioner has the attractive property of improved eigenvalue clustering with increased ill-conditioning of the (1,1) block of the saddle-point matrix. We analyze its spectral characteristics, utilizing projections onto the null space of the constraint matrix. We then present a practical application to the symmetric case and demonstrate performance of the preconditioner on LP problems from the NETLIB test suite.

Key words. preconditioners, saddle-point systems, primal-dual interior point methods

AMS subject classifications. 65F10

1. Introduction.

Consider the linear programming problem

$$\min_{x\in\mathbb{R}^n} c^T x, \text{ subject to: } Ax = b,\ x \geq 0.$$

The cost vector c is $n \times 1$, A is the $m \times n$ constraint matrix ($m < n$), and b is an $m \times 1$ vector. We assume that $\text{rank}(A) = m$. Linear programs are convex problems, so satisfaction of the *Karush-Kuhn-Tucker* (KKT) conditions [7, p. 366] guarantees global optimality of a solution. These conditions require that $Ax = b$, $A^T\lambda + s = c$, and $x^T s = 0$, where $\lambda \in \mathbb{R}^m$, and $s \in \mathbb{R}^n$ with $s \geq 0$, are the dual and dual slack variables respectively. By defining

$$F(x, \lambda, s) = \begin{pmatrix} x^T s \\ A^T\lambda + s - c \\ Ax - b \end{pmatrix},\ x,\ s\ \geq 0,$$

linear programming can be treated as a root finding problem for F, since any root satisfies the KKT conditions. Solutions are found through a sequence of steps, with search directions from a feasible point (x_k, λ_k, s_k) given by the Newton step:

$$\begin{pmatrix} X & S & 0 \\ I & 0 & A^T \\ 0 & A & 0 \end{pmatrix} \begin{pmatrix} \Delta s \\ \Delta x \\ \Delta\lambda \end{pmatrix} = \begin{pmatrix} -XSe \\ 0 \\ 0 \end{pmatrix}. \tag{1.1}$$

This will be referred to as the step equation [7]. Here X is the diagonal matrix with $X_{ii} = x_i$, S is the diagonal matrix with $S_{ii} = s_i$, and $e = [1, \ldots, 1]^T$. Note that the matrix in equation (1.1) can be treated as a nonsymmetric $(2n + m) \times (2n + m)$ saddle-point system with an indefinite $(1, 1)$ block.

Allowance for infeasible iterates and upper bounds on primal variables is typical in primal-dual interior point methods [7, 9]. These modifications lead to matrix equations that can be reduced to the same block 3×3 structure of the linear system (1.1).

*This work was supported in part by the Natural Sciences and Engineering Research Council of Canada.

†Department of Computer Science, The University of British Columbia, Vancouver B.C. V6T 1Z4, Canada, {trees, greif}@cs.ubc.ca. Work by the first author was done in partial fulfillment of an MSc degree.

Primal-dual algorithms are often based on Mehrotra's algorithm [5], which requires solutions of the step equation twice at each iteration. This illustrates the importance of efficiently solving (1.1). A difficulty here is that as LP solutions are approached, increasing complementarity of x and s makes the matrix very ill-conditioned. Typically, equation (1.1) is reduced to a block 2×2 system:

$$\begin{pmatrix} -D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \end{pmatrix} = \begin{pmatrix} XSe \\ 0 \end{pmatrix} \tag{1.2}$$

by substituting $\Delta s = -X^{-1}S\Delta x - XSe$, and letting $D = X^{-1}S$. This is a symmetric saddle-point system with special properties. The $(1,1)$ block, D, is diagonal, and near the LP solution it approaches singularity. When the LP solution occurs at a unique vertex, m values of D tend to zero, while $n-m$ become very large. Even when the LP solution is not unique D is generally seen to be very poorly conditioned. System (1.2) can be transformed to have a positive definite $(1,1)$ block by multiplying both sides by -1. Note that with infeasible iterates, entries in x and s only approach zero, so D will become arbitrarily ill-conditioned but not singular.

Because D is diagonal, the 2×2 system can easily be reduced to the normal equations, $AD^{-1}A^T\Delta\lambda = ADX^{-1}Se$. This approach is used by LIPSOL and PCx [11, 1]. However, since D is ill-conditioned, if a direct solve such as the sparse Cholesky decomposition is employed, special precautions are necessary to deal with the presence of extreme pivots [6].

The ill-conditioning of D, as well as the need to solve very large linear programs motivates our work; in this paper we are concerned with solving the block 2×2 system with a preconditioned iterative solver. An iterative solution to the step equation has some advantages. Since solving an LP involves a sequence of steps, it is not necessary that at each iteration we solve the step equation exactly. Solving iteratively allows us to relax the accuracy of intermediate solutions, but still converge to the optimal value. Secondly, if storage is a concern, then an iterative approach offers potentially substantial savings.

One of the main features of the proposed preconditioner is that the more ill-conditioned D is, the better a minimum residual solver, such as MINRES, performs. Therefore, it is particularly effective in the last few iterations of the LP solver, where serious numerical difficulties may arise and effective preconditioning is essential. Our approach is based on augmentation of the (1,1) block using a weight matrix. On the one hand, this requires a larger computational effort per iteration. On the other hand, a proper choice of the weight matrix may lead to a much better conditioned (1,1) block and may give rise to faster and more robust solution methods. The viability of our approach is demonstrated throughout the paper both analytically and numerically.

The remainder of this paper is organized as follows. In Section 2 we present a preconditioner for nonsymmetric KKT systems with indefinite $(1,1)$ blocks. In Section 3 the preconditioner is presented to deal with (1.2) and we provide eigenvalue analysis on the projected space. In Section 4 we explain how the parameter of the preconditioner can be chosen. Section 5 describes the performance of the preconditioner on problems from the NETLIB LP test suite. In Section 6 we draw some conclusions and outline future work.

2. Preconditioning the nonsymmetric system. We present a preconditioner for nonsymmetric saddle-point systems with indefinite $(1,1)$ blocks. As noted earlier, the block 3×3 step equation (1.1) is such a system. The generalized preconditioner,

though, will probably be most appropriate for systems that lack the simple structure of equation (1.1).

Consider the general saddle-point system

$$\mathcal{G} = \begin{pmatrix} F & A^T \\ B & 0 \end{pmatrix},$$

where F is $n \times n$, and A and B are $m \times n$. We assume $\mathcal{G}$ is non-singular, and both A^T and B have full rank. We allow F to be nonsymmetric and indefinite and assume it has p linearly independent null vectors.

We define a preconditioner for this system by:

$$\mathcal{M}_G = \begin{pmatrix} F + A^T W^{-1} B & 0 \\ 0 & W \end{pmatrix},$$

which is a generalization of the preconditioner introduced in [2, 3]. We assume that F, A, and B are such that $\mathcal{M}_G$ is non-singular; a necessary condition for this is that the column space of A^T and null space of B do not intersect. Spectral properties of the preconditioned matrix $\mathcal{M}_G^{-1}\mathcal{G}$ are characterized by the following theorem.

THEOREM 2.1. *Consider a saddle-point system $\mathcal{G}$ and preconditioner $\mathcal{M}_G$ as above. Let F have p linearly independent null vectors. Then the matrix $\mathcal{M}_G^{-1}\mathcal{G}$ has eigenvalues $\lambda = 1$ with algebraic multiplicity n, and $\lambda = -1$ with algebraic multiplicity p.*

Proof. The generalized eigenvalue problem can be written as

$$\begin{pmatrix} F & A^T \\ B & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} F + A^T W^{-1} B & 0 \\ 0 & W \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

From the second row, we see that $y = \frac{1}{\lambda} W^{-1} B x$. This is legitimate by the assumption $\mathcal{G}$ and $\mathcal{M}_G$ are non-singular. This can be substituted into the first row and reduced to:

$$(\lambda^2 - \lambda)F + (\lambda^2 - 1)A^T W^{-1} B x = 0.$$

We observe that if x is a null vector of B then $(\lambda^2 - \lambda)Fx = 0$. Clearly $Fx \neq 0$ or else $\mathcal{G}$ would be singular. Since the column space of A^T and the null space of B span $\mathbb{R}^n$, an additional m eigenvalues of $\lambda = 1$ can be seen. These have eigenvectors with upper n-components v chosen from m linearly independent vectors in the column space of A^T. To be precise, if we let $QR = A^T$ be the QR decomposition of A^T, and e_i be the i^{th} column vector of the identity matrix, then $(x, y) = (Qe_i, W^{-1}BQe_i)$ is an eigenvector with $\lambda = 1$.

We also see that $\lambda = -1$ with multiplicity p by inspection. Every null vector of F forms the upper n-component of an eigenvector of $\mathcal{M}_G^{-1}\mathcal{G}$, since Fx and $A^T W^{-1} B x$ cannot be zero simultaneously. If we let p_i denote the i^{th} null vector of F, then $(x, y) = (p_i, -W^{-1}Bp_i$ is an eigenvector with $\lambda = -1$. ∎

While the approach laid out in this section is applicable to a large class of problems, in the particular case of linear programming it is possible to derive similar preconditioners for a symmetric formulation. We now turn our attention to the symmetric case.

3. A block-diagonal positive definite preconditioner. Consider now the linear system $\mathcal{K}x = b$, where

$$\mathcal{K} = \begin{pmatrix} D & A^T \\ A & 0 \end{pmatrix}, \tag{3.1}$$

and D is diagonal, positive definite, and ill-conditioned ($\mathcal{K}$ is obtained by multiplying both sides of equation (1.2) by -1). As mentioned in the introduction, the ill-conditioning arises from the complementarity of x and s (in D) creating large and small entries on the diagonal of $\mathcal{K}$. We consider a preconditioner of the form

$$\mathcal{M}_1 = \begin{pmatrix} D + A^T W^{-1} A & 0 \\ 0 & W \end{pmatrix}. \tag{3.2}$$

which was introduced in [2, 3]. The following theorem characterizes the spectral properties of $\mathcal{M}_1^{-1}\mathcal{K}$. Its proof can be seen as a special case of Theorem 2.1 (see also [2, Thm.2.2]), however in the context of optimization it is useful to derive spectral properties on the reduced space of the constraint matrix. Below we give such a proof, which is similar in spirit to the proof for the constraint preconditioners given in [4].

THEOREM 3.1. *Assume D is positive semi-definite, with nullity p. Assume A is of full rank. The preconditioned matrix $\mathcal{M}_1^{-1}\mathcal{K}$ has eigenvalues $\lambda = 1$ with multiplicity n, and $\lambda = -1$ with multiplicity p. The remaining eigenvalues lie in the interval $(-1, 0)$.*

Proof. The eigenvalues of the preconditioned matrix $\mathcal{M}^{-1}\mathcal{K}$ can be found through the generalized eigenvalue problem

$$\begin{pmatrix} D & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} = \lambda \begin{pmatrix} D + A^T W^{-1} A & 0 \\ 0 & W \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}. \tag{3.3}$$

We transform this system to observe its behavior on the null space of A.

Let $QR = A^T$ be the QR factorization of A^T. Since A^T is $n \times m$, Q is $n \times m$, and R is $m \times m$. Define Z to be a $n \times (n-m)$ orthogonal basis for the null space of A. Since $Z \cup Q$ forms an orthogonal basis for $\mathbb{R}^n$, any vector $x \in \mathbb{R}^n$ can be written as: $x = Zx_z + Qx_q$.

We define $P = \begin{pmatrix} Z & Q & 0 \\ 0 & 0 & I \end{pmatrix}$ to be $(n+m) \times (n+m)$. We express $\hat{x} = Z\hat{x}_z + Q\hat{x}_q$, and let $v = (x_z, x_q, y)^T$ where $Pv = (\hat{x}_z, \hat{x}_q, \hat{y})^T$. The generalized eigenvalue problem can then be written as $P^T\mathcal{K}Pv = \lambda P^T\mathcal{M}Pv$. This yields:

$$\begin{pmatrix} Z^TDZ & Z^TDQ & 0 \\ Q^TDZ & Q^TDQ & R \\ 0 & R^T & 0 \end{pmatrix} \begin{pmatrix} x_z \\ x_q \\ y \end{pmatrix} = \lambda \begin{pmatrix} Z^TDZ & Z^TDQ & 0 \\ Q^TDZ & Q^TDQ + RW^{-1}R^T & 0 \\ 0 & 0 & W \end{pmatrix} \begin{pmatrix} x_z \\ x_q \\ y \end{pmatrix}. \tag{3.4}$$

Observe that by setting $\lambda = 1$, the system reduces to:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -RW^{-1}R^T & R \\ 0 & R^T & -W \end{pmatrix} \begin{pmatrix} x_z \\ x_q \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Evidently there are $n - m$ corresponding eigenvectors that can be written in the form $(x_z, x_q, y) = (e_i, 0, 0)$. In addition, m linearly independent eigenvectors can be written in the form: $(x_z, x_q, y) = (0, e_i, W^{-1}R^Te_i)$.

Now consider $\lambda = -1$. Equation (3.4) reduces to:

$$\begin{pmatrix} 2Z^TDZ & 2Z^TDQ & 0 \\ 2Q^TDZ & 2Q^TDQ + RW^{-1}R^T & R \\ 0 & R^T & W \end{pmatrix} \begin{pmatrix} x_z \\ x_q \\ y \end{pmatrix} = 0.$$

Any vector $x^* = Zx_z^* + Qx_q^*$ in the null space of D obeys $D(Zx_z^* + Qx_q^*) = 0$. There are p such vectors, so p linearly independent eigenvectors of the form $(x_z, x_q, y) = (x_z^*, x_q^*, -W^{-1}R^Tx_q^*)$ will satisfy equation (3.4) with $\lambda = -1$.

To derive an expression for the remaining eigenvalues, $\lambda \neq \pm 1$, we reduce equation (3.4) to an eigenvalue problem in x_q. From (3.4), $y = \frac{1}{\lambda}W^{-1}R^Tx_q$ and $x_z = -(Z^TDZ)^{-1}Z^TDQx_q$. Substituting these values into the second line of (3.4) and simplifying yields the generalized eigenvalue problem:

$$\lambda \mathcal{S} x_q = RW^{-1}R^Tx_q,$$

where

$$\mathcal{S} = C^T(Z^TDZ)^{-1}C - Q^TDQ - RW^{-1}R^T,$$

with $C = Z^TDQ$. The remaining eigenvalues satisfy this equation and lie in the interval $(-1, 0)$, as proved in [2]. □

The structure of D in LP problems makes $\mathcal{M}_1$ very worthy of consideration. As LP solutions are approached and the nullity of D increases (up to m), convergence of a Krylov solver can be expected in just 2 iterations (in the absence of round-off errors). The preconditioner actually performs better as $\mathcal{K}$ becomes more ill-conditioned. Another important point to note is that when W is symmetric positive definite, so is $\mathcal{M}_1$. This is because the null space of D and A cannot intersect or $\mathcal{K}$ would be singular. Positive definiteness of $\mathcal{M}_1$ makes it usable with MINRES. Let us now describe how the choice of W can affect the convergence of MINRES.

4. Choosing W. We are taking a system with a diagonal $(1, 1)$ block and iteratively solving it with a preconditioner whose $(1, 1)$ block may be dense. For small problems this is not a practical approach, and it is hard to compete with the sparse Cholesky decomposition of the normal equations. However, as problems become larger iterative methods will be required.

The critical decision in the application of preconditioners $\mathcal{M}_1, \mathcal{M}_G$ is the selection of W. Recall that each step of a preconditioned Krylov solver requires the solution of a system of the form $\mathcal{M}z = r$. The preconditioner $\mathcal{M}_1$ must be chosen so that solving $Wy = a$ and $(D + A^TW^{-1}A)z = c$ are both relatively easy. For the latter, even if $A^TW^{-1}A$ is effectively dense, matrix vector products with it can be done cheaply with no extra storage.

We considered a variety of W matrices, and focussed on a simple scaled identity. Inverting W is thus trivial, and systems of the form $(D + A^TW^{-1}A)z = c$ can be tackled with a conjugate gradient approach.

Experiments showed that with $W = \gamma I$, a large γ did not create the desired effect of augmentation and gave poor results. Increasing $\frac{1}{\gamma}$ to the order of the largest entry in D, though, showed much better performance. This can be explained as follows. Since $W^{-1} = \frac{1}{\gamma}I$, and D is a diagonal matrix with extremely large entries, a small γ can scale $\frac{1}{\gamma}A^TA$ to have a norm close to D. When the two matrices do not have similar norms, D will dominate and the solution of $(D + \frac{1}{\gamma}A^TA)z = c$ will be inaccurate. This

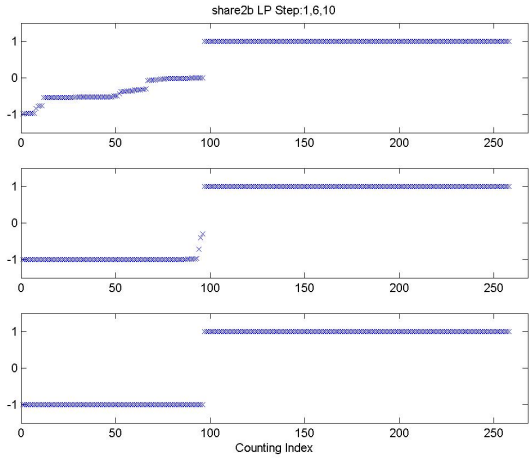

FIG. 5.1. *Eigenvalues of $\mathcal{M}_1^{-1}\mathcal{K}$ at different steps of the LP solution for "share2b". Note clustering to $\lambda = \pm 1$ occurs quickly, typically within a few LP steps.*

was especially true of the larger problems ($n + m > 2000$). On the other hand, if γ is too small the augmented $(1, 1)$ block is dominated by the rank deficient $A^T A$. This demonstrates the importance of making a sensible choice of γ.

5. Numerical Experiments. Numerical experiments were conducted on the NETLIB suite of LP problems. Testing was done in MATLAB, using LIPSOL [10]. MINRES is appropriate as the iterative solver because the preconditioner is positive definite but the KKT system is indefinite.

Preliminary experiments confirmed that the preconditioned matrices $\mathcal{M}_1^{-1}\mathcal{K}$ did exhibit the expected eigenvalue clustering. These tests were necessary because interior point methods do not generate feasible iterates with singular D matrices. With D only approximately singular it was uncertain whether the full eigenvalue clustering of $\mathcal{M}_1^{-1}\mathcal{K}$ would occur. Figure 5.1 depicts the eigenvalues of $\mathcal{M}_1^{-1}\mathcal{K}$ at three different steps of the solving process for problem "share2b". Typically the clustering improves with iterations, and only a few outlying eigenvalues remain within the first ten steps of a problem. For problems without unique solutions, D will not approach nullity m, so full clustering cannot be expected.

Experiments were conducted to study how varying W in $\mathcal{M}_1$ would affect convergence of MINRES. Diagonal W matrices were investigated. We found that convergence to the solution of the step equation occurred in far fewer iterations as γ was decreased. On the other hand, if γ is taken too small numerical difficulties can occur. Taking $1/\gamma = \max(D)$ consistently gave good results and was used as a basic heuristic for choosing γ. This is demonstrated in Figure 5.2, where our choice of γ clearly results in the fewest MINRES iterations throughout the LP process.

Studies were also conducted to see how varying the error tolerance of MINRES would affect overall convergence to the correct LP solution. This is a very important question because relaxing the accuracy of intermediate solutions in the step process is a key advantage of an iterative approach. Generally we found that for many problems a loose error tolerance of 10^{-2} was sufficient. In Table 5.1 we demonstrate the total number of MINRES iterations required for several NETLIB problems. A loose error

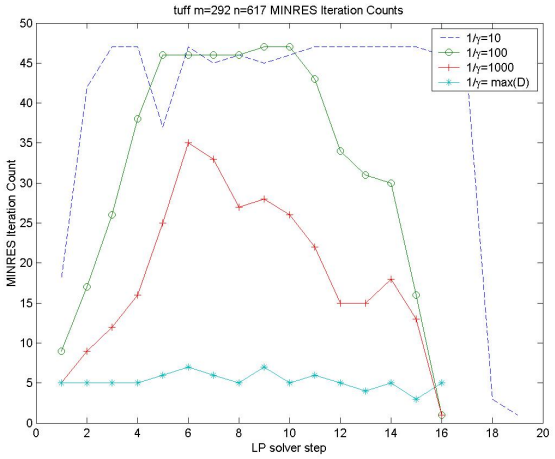

FIG. 5.2. *MINRES iteration counts with various* $W = \gamma I$. *Problem "tuff" has* $m = 292$, $n = 617$ *after LIPSOL preprocessing. MINRES error tolerance was set to* 10^{-8}.

tolerance often leads to a reduction in MINRES iterations, typically requiring only a few extra LP steps. The total MINRES iteration counts (summed over all LP steps) in Table 5.1 demonstrate the extra cost associated with having a tighter error tolerance.

MINRES Tol	10^{-2}	10^{-4}	10^{-6}	10^{-8}
adlittle(m=55, n=137)	19	28	39	44
bandm(m=269, n=436)	45	90	138	213
grow15(m=300, n=645)	19	22	30	46
scorpion(m=375, n=453)	78	88	126	353
scsd1(m=77, n=760)	11	20	20	22
25fv47(m=798, n=1854)	DNC	54	83	121

TABLE 5.1

Total MINRES iteration counts for NETLIB problems with different error tolerances. "DNC" indicates the solution did not converge.

The method did not converge for every problem, as can be seen from the table. Larger problems often required stricter error tolerances. We note that for a few small problems the solver did not perform well and we suspect that for those problems a more sophisticated choice of W may be required. Experiments were run for approximately 30 NETLIB problems, entries in Table 5.1 are a representative subset of our results.

6. Conclusions. We have studied a new preconditioner for linear programming problems, and have demonstrated its merits in several aspects. The preconditioner is well suited for saddle-point systems with a highly singular (1,1) block; in fact close to convergence, where ill-conditioning is at its prime, convergence of MINRES is the fastest and is theoretically guaranteed to occur within two iterations at most (in the absence of roundoff errors). We have also provided spectral analysis on the null space of the constraint matrix.

The value of the parameter W is crucial and we have pointed out a way to make that choice by using a scaled identity matrix based on the entries of D. This sensible

choice has been shown to reduce the iteration counts throughout, demonstrated by way of "a flattened hump" in Figure 5.2. Furthermore, we have shown that applying an inexact version of MINRES throughout the iteration, with a convergence tolerance as low as 0.01 substantially reduces the overall amount of computational work.

Future work will focus on applying this approach to other classes of optimization problems. LP problems have the special property that the normal equations can be explicitly formed because D is diagonal. In quadratic programming and nonlinear optimization, this is generally not the case and reduction to the normal equations is not possible. In these problems the gains of preconditioning of the sort we are proposing may likely be very visible.

REFERENCES

[1] J. Czyzyk, S. Mehrotra, and S. J. Wright. PCx User Guide. Technical Report OTC 96/011, Optimization Technology Center, 1996.
[2] C. Greif and D. Schötzau. Preconditioners for saddle point linear systems with highly singular $(1, 1)$ blocks. *ETNA, Accepted for publication*, 2005.
[3] C. Greif and D. Schötzau. Preconditioners for the discretized time-harmonic maxwell equations in mixed form. *Submitted to SIAM Journal on Numerical Analysis*, 2005.
[4] C. Keller, N. I. M. Gould, and A. J. Wathen. Constraint preconditioning for indefinite linear systems. *SIAM J. Matrix Anal. Appl.*, 2000.
[5] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM journal on Optmization*, 2:575–601, 1992.
[6] E. Ng and B. W. Peyton. Block sparse cholesky algorithms on advanced uniprocessor computers. *SIAM Journal on Scientific Computing 14*, pages 617–629, 1995.
[7] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New york, 1999.
[8] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA, 1986.
[9] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, PA, 1997.
[10] Y. Zhang. Solving large-scale linear programs by interior-point methods under the matlab environment. *Optimization Methods and Software*, 1998.
[11] Y. Zhang. User's Guide to LIPSOL. *Optimization Methods and Software*, 11-12:385–396, 1999.