A New Primal-Dual Interior-Point Method for Semidefinite Programming*

Farid Alizadeh† Jean-Pierre A. Haeberly‡ Michael L. Overton§

Abstract

Semidefinite programming (SDP) is a convex optimization problem in the space of symmetric matrices. Primal-dual interior-point methods for SDP are discussed. These generate primal and dual matrices X and Z which commute only in the limit. A new method is proposed which iterates in the space of commuting matrices.

Let $S\Re^{n\times n}$ denote the set of real symmetric $n \times n$ matrices. The standard inner product on this space is $A \bullet B = \mathbf{tr}\ AB = \sum_{i,j} a_{ij}b_{ij}$. By $X \succeq 0$, where $X \in S\Re^{n\times n}$, we mean that X is positive semidefinite. Consider the *semidefinite programming problem* (SDP)

$$\min \quad C \bullet X \tag{1}$$

$$\text{s.t.} \quad A_i \bullet X = b_i \quad i = 1, \dots, m; \quad X \succeq 0. \tag{2}$$

Here C and A_i, $i = 1, \dots, m$, are all fixed matrices in $S\Re^{n\times n}$, and the unknown variable X also lies in $S\Re^{n\times n}$. The semidefinite constraint on X is said to be *nonsmooth*, since it is equivalent to a bound constraint on the least eigenvalue of X, which is *not* a differentiable function of X. The constraint is, however, convex. If the constraints are chosen to enforce X to be diagonal one obtains linear programming (LP) as a special case of SDP.

There is a complete duality theory for SDP, which is quite analogous[1] to the well known duality theory for linear programming (LP). The dual of SDP is

$$\max \quad b^T y \tag{3}$$

$$\text{s.t.} \quad Z + (\textstyle\sum_{i=1}^m y_i A_i) = C; \quad Z \succeq 0 \tag{4}$$

where Z is a dual slack matrix variable, which also lies in $S\Re^{n\times n}$. The SDP primal-dual pair enjoy many of the same properties as in LP. If either is unbounded, the other is infeasible, and if one has a finite optimal value, the other does also, with the same objective value. A complementary slackness result also holds for SDP: if X and y, Z are respectively primal and dual optimal, then $XZ = 0$. Notice that, in contrast with LP, component-wise multiplication in the complementary slackness theorem is replaced by matrix-matrix multiplication. Proofs of all these duality results and references to relevant literature are given in [2].

*Computer Science Dept. Report 659, Courant Institute of Mathematical Sciences, New York University. To appear in: Proceedings of the Fifth SIAM Conference on Applied Linear Algebra, Snowbird, Utah, June 1994.

†International Computer Science Institute, Berkeley, CA. Supported in part by NSF postdoctoral grant CDA-9211106.

‡Mathematics Department, Fordham University, Bronx, NY.

§Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, NY. Supported in part by National Science Foundation Grant CCR-9101649.

[1] A constraint qualification is required to establish these properties in general. It is sufficient to assume that the primal feasible region has nonempty relative interior. See [2] for details.

Several authors [1,2,6,8,9,10] have observed that the interior point methods which have been so successful for LP may also be applied to solve SDP and related problems. Nesterov and Nemirovskii [6] is the primary reference for theoretical properties of interior point algorithms for general convex programs. Alizadeh [1,2] argues that the various algorithmic techniques and mathematical proofs which have been developed for LP may be generalized in a systematic way to apply to SDP. Instead of the barrier term $-\sum_{j=1}^{n}\log x_j$ used in LP, one introduces the barrier term $-\log\det X$ to correspond to the semidefinite constraint $X \succeq 0$. Provided the initial guess X satisfies the semidefinite constraint, the barrier term prevents subsequent values from leaving the positive semidefinite cone.

Primal-dual interior point methods are of particular interest, since these have been shown to be very efficient for solving LP (see e.g. [3,4,11,12]). One iteration of the primal-dual method can be derived by applying Newton's method to three equations: primal feasibility, dual feasibility, and complementarity/centering. The primal and dual feasibility equations are the equality constraints in (2), (4). The complementarity/centering equation is $XZ = \mu I$, where μ is a parameter to be driven to zero. In contrast to LP, X and Z are full matrices, not diagonal matrices, so the left hand side XZ is not necessarily symmetric. Consequently, direct application of Newton's method to $XZ = \mu I$ leads to nonsymmetric corrections ΔX and ΔZ, which is not acceptable. Instead, one may apply Newton's method to any of the following three symmetric matrix equations

$$Z = \mu X^{-1} \tag{5}$$

$$X = \mu Z^{-1} \tag{6}$$

$$XZ + ZX = 2\mu I. \tag{7}$$

These equations have been used by various authors. For example, (6) is used by [8], while the algorithm of [10] implicitly uses a combination of (5) and (6). The third form (7), which has the attraction of symmetry in X and Z, does not seem to have been used previously. It is also attractive because in the special case of LP, it is generally agreed that the form $XZ = \mu I$ is preferable to formulations involving X^{-1} or Z^{-1}. Newton linearization of (7) gives

$$X(\Delta Z) + (\Delta X)Z + Z(\Delta X) + (\Delta Z)X = 2\mu I - XZ - ZX. \tag{8}$$

Combining this equation with the primal and dual feasibility equations we see that one step of the primal-dual Newton method can be written[2]

$$\begin{bmatrix} 0 & \mathcal{A}^T & I \\ \mathcal{A} & 0 & 0 \\ E & 0 & F \end{bmatrix} \begin{bmatrix} \operatorname{vec}\Delta X \\ \Delta y \\ \operatorname{vec}\Delta Z \end{bmatrix} = \begin{bmatrix} \operatorname{vec} C - \mathcal{A}^T y - \operatorname{vec} Z \\ b - \mathcal{A}\operatorname{vec} X \\ \operatorname{vec}(2\mu I - XZ - ZX) \end{bmatrix}, \tag{9}$$

where $E = Z \otimes I + I \otimes Z$ and $F = X \otimes I + I \otimes X$. If (5) or (6) is used instead of (7), the only change in the left-hand side occurs in the blocks E and F, which become, respectively, $E = \mu X^{-1} \otimes X^{-1}$, $F = I$ (in the case of (5)) and $E = I$, $F = \mu Z^{-1} \otimes Z^{-1}$ (for (6)).

As in LP, this 3 by 3 block matrix equation (9) can be reduced by block Gauss elimination. Using the third block equation, either ΔX or ΔZ can be eliminated in terms

[2] The operator "vec" maps the matrix space $S\Re^{n\times n}$ into the corresponding vector space $\Re^{n^2}$, so that $(\operatorname{vec} A)^T(\operatorname{vec} B) = A \bullet B$. The matrix $\mathcal{A}$ consists of the m rows $(\operatorname{vec} A_i)^T$ $i = 1, \ldots, m$. The notation $A \otimes B$ means the Kronecker product. The block matrix equation contains duplicate rows because of symmetry. Consequently, in practice, the vec operator is modified to generate vectors of length $n(n+1)/2$, not n^2. The Kronecker product is likewise also modified to exploit symmetry. Thus (9) consists of $n(n+1) + m$ equations in $n(n+1) + m$ variables.

of the other. This leaves a 2 by 2 block matrix equation of order $\frac{n(n+1)}{2} + m$. Application of block Gauss elimination then reduces the problem to that of factoring the m by m matrix $\mathcal{A}G\mathcal{A}^T$, where G is either $E^{-1}F$ or $F^{-1}E$. In cases (5) and (6) the matrix G can be computed in $O(n^3)$ operations by observing that the inverse of a Kronecker product is the Kronecker product of the inverses. Likewise in case (7) the spectral decompositions of E and F can be explicitly written in terms of those of Z, X. For large problems, it may be better not to form the matrix $\mathcal{A}G\mathcal{A}^T$ at all, but solve the corresponding least squares problem by conjugate gradients or LSQR [10].

Once ΔX, Δy and ΔZ are computed, the variables are updated by $X \leftarrow X + \alpha\Delta X$, $y \leftarrow y + \beta\Delta y$, and $Z \leftarrow Z + \beta\Delta Z$, where α and β are respectively primal and dual steplengths. The method of [10] uses a rule based on work of [6] which guarantees global convergence. Alternatively one can, as in LP, take steps $\alpha = \min(1, \tau\hat{\alpha})$ and $\beta = \min(1, \tau\hat{\beta})$, where τ is a number close to 1, and $\hat{\alpha}$ and $\hat{\beta}$ are respectively steps to the boundary of the feasible regions $X \succeq 0$ and $Z \succeq 0$. This can be computed explicitly [8]: $\hat{\alpha}$ is the inverse of the maximum eigenvalue of $-R^{-T}(\Delta X)R^{-1}$, where $X = R^TR$.

We have experimented with all these versions of the primal-dual method, using various techniques to reduce μ. We have found no clear advantage to any one of the three forms of the complementarity/centering condition. We find that taking steps close to the boundary, e.g. with $\tau = 0.99$, has much faster convergence than using a Nesterov–Nemirovskii line search as in [10], but is less reliable. The SDP algorithms are much more prone to "getting stuck" than corresponding LP algorithms, for reasons that are not clear. The rate of convergence is significantly improved by using Mehrotra's LP predictor-corrector method [5]. This method uses additional terms of the Taylor approximation to the complementarity/centering condition and is particularly easy to implement using (7), since its expansion is so convenient. However, problems with reliability remain.

We now outline a new primal-dual interior point method for SDP. The idea is to iterate in the space of *commuting matrices* X and Z. Consider again the complementarity/centering matrix equation $XZ = \mu I$. For matrices X and Z to satisfy this equation, they must commute with each other, i.e. share a common set of eigenvectors. Let us therefore *replace* X and Z respectively by

$$QXQ^T \quad \text{and} \quad QZQ^T$$

where Q is an n by n *orthogonal* matrix and X and Z are now *diagonal* matrices. Thus the new diagonal matrices X and Z consist of eigenvalues of the original X and Z while the columns of Q are the common set of orthogonal eigenvectors. The primal and dual feasibility equations then become

$$Q^TA_iQ \bullet X = b_i \quad i = 1, \ldots, m \quad \text{and} \quad Z + (\sum_{i=1}^{m} y_i Q^TA_iQ) = Q^TCQ. \tag{10}$$

The complementarity/centering condition reduces to the *diagonal equation* $XZ = \mu I$. The total number of variables and unknowns is reduced from $n^2 + n + m$ to $\frac{n(n+1)}{2} + n + m$, since the orthogonal matrix Q has $\frac{n(n-1)}{2}$ degrees of freedom. The price paid for the diagonalization is the nonlinear appearance of the variable Q in the feasibility equations.

We now wish to apply Newton's method to (10) together with $XZ = \mu I$. In order to do so, we use a technique from [7], namely *parameterize the orthogonal matrix Q by an exponential transformation.* Specifically, we shall replace X, y and Z by, respectively,

$X + \Delta X$, $y + \Delta y$ and $Z + \Delta Z$, and we shall replace Q by

$$Qe^S = Q(I + S + \frac{1}{2}S^2 + \cdots)$$

where S is *skew-symmetric* (so that e^S is orthogonal). The Newton step is then derived by approximating e^S by $I + S$, while the predictor-corrector method uses additional terms of the expansion. Let $B_i = Q^T A_i Q$, and let $H = (\sum_{i=1}^m y_i B_i) - Q^T C Q$. Linearizing, we see that Newton's method applied to these equations is the following. Here the variables are ΔX, Δy, ΔZ and the skew symmetric matrix S; remember B_i, X, y, Z and Q are fixed at their current values, and that X, Z are diagonal.

$$B_i \bullet \Delta X + (B_i S - S B_i) \bullet X = b_i - B_i \bullet X, \qquad i = 1, ..., m \tag{11}$$

$$(\sum_{i=1}^m \Delta y_i B_i) + HS - SH + \Delta Z = -H - Z \tag{12}$$

$$X \Delta Z + Z \Delta X = \mu I - XZ \tag{13}$$

As usual, we can eliminate ΔX or ΔZ using (13). Equation (11) can be rewritten

$$B_i \bullet \Delta X + \mathbf{tr}\,(XB_i - B_i X)S = b_i - B_i \bullet X.$$

The usual block Gauss elimination to reduce the linear system to size m by m is applicable, provided we can efficiently solve systems of the form

$$HS - SH + \Delta Z = M$$

for the skew-symmetric matrix S and the diagonal matrix ΔZ, given a symmetric matrix right-hand side M. This is a linear system of $\frac{n(n+1)}{2}$ equations in $\frac{n(n+1)}{2}$ variables which has an interesting structure which we have not seen before, but which may well have arisen in other applications. It has the character of a Lyapunov equation and can be solved in approximately $O(n^3)$ operations using a spectral decomposition of H followed by the factorization of the Hadamard product matrix $P^T \circ P^T$, where P is an orthogonal eigenvector matrix for H.

Once ΔX, Δy, ΔZ and S are computed, a simple ratio test determines primal and dual steplengths for the diagonal matrices ΔX and ΔZ. The step Δy is scaled by the dual steplength. We use the geometric mean of the primal and dual steplengths to scale S. We also incorporate Mehrotra's predictor-corrector modification, which can be applied without difficulty in the context of the new method.

An implementation of the new primal-dual method in Matlab shows fast local convergence. Results for a typical random problem with $n = 15$ and $m = 50$, starting with nearly feasible points, are shown in Table 1. The number of primal variables in this problem is $\frac{n(n+1)}{2} = 120$. The second and third columns in the table show primal and dual infeasibility, i.e. the norm of the residual of the two equations in (10). The last column displays the duality gap $\mathbf{tr}\,XZ$. Primal and dual feasibility is achieved only in the limit, because of the nonlinear variable Q in (10). The factors of 100 in the final convergence rate are a consequence of using $\tau = 0.99$. Faster convergence can be achieved by taking steps closer to the boundary. Various methods can be used to reduce μ. We use the rule used by [10] at first, switching to the predictor-corrector rule when primal and dual infeasibility

Iteration	Primal Infeas.	Dual Infeas.	Duality Gap
1	2.784e-02	8.129e-03	1.198e+01
2	2.169e-02	7.881e-03	9.560e+00
3	1.332e-02	6.700e-03	7.628e+00
4	8.417e-03	5.601e-03	6.084e+00
5	1.660e+00	5.529e-01	5.768e-01
6	6.138e-02	6.726e-02	4.644e-01
7	3.944e-04	3.141e-03	3.697e-01
8	1.484e-04	9.674e-03	6.716e-03
9	4.209e-08	5.356e-04	8.068e-05
10	4.217e-12	5.524e-06	8.071e-07

Table 1

Sample Convergence of New Primal-Dual Method for SDP

drop below a threshold value 0.01. This explains the irregular behavior at iteration 5, where the switch took place in this case.

We think the new method has great potential, but many questions need to be studied. In particular, the issues of global convergence and of exploiting sparsity in the data have not been addressed as yet.

References

[1] F. Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices.* PhD thesis, University of Minnesota, 1991.

[2] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 1994. To appear.

[3] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point method for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming, Interior-Point and Related Methods*, pages 29–47, New York, 1989. Springer-Verlag.

[4] I.J. Lustig, R.E. Marsten, and D.F. Shanno. Interior point methods for linear programming: Computational state of the art. *ORSA Journal on Computing*, 6:1–14, 1994.

[5] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, 2:575–601, 1992.

[6] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming.* SIAM, 1994.

[7] M.L. Overton and R.S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. Computer Science Dept. Report 626, Courant Institute of Mathematical Sciences, NYU, 1993. Submitted to *SIAM J. Matrix Anal. Appl.*

[8] F. Rendl, R.J. Vanderbei, and H. Wolkowicz. Max-min eigenvalue problems, primal-dual interior point algorithms, and trust region subproblems. Technical Report CORR 93-30, University of Waterloo, 1993. To appear in Proceedings of NATO Conference on Nonlinear Programming, Il Ciocco, Italy, Sept. 1993.

[9] U.T. Ringertz. Optimal design of nonlinear shell structures. Report, The Aeronautical Research Institute of Sweden, 1991.

[10] L. Vandenberghe and S. Boyd. Primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 1994. To appear.

[11] R.J. Vanderbei and T.J. Carpenter. Symmetric indefinite systems for interior-point methods. *Mathematical Programming*, 58:1–32, 1993.

[12] Y. Zhang, R.A. Tapia, and J.E. Dennis. On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms. *SIAM Journal on Optimization*, 2:304–324, 1992.