Acta Numerica (1996), *pp.* 149–190

Eigenvalue optimization

Adrian S. Lewis *
Department of Combinatorics and Optimization
University of Waterloo
Ontario N2L 3G1, Canada
E-mail: aslewis@orion.uwaterloo.ca
WWW: http://orion.uwaterloo.ca/~aslewis

Michael L. Overton †
Computer Science Department
Courant Institute of Mathematical Sciences
New York University
New York, NY 10012–1110, USA
E-mail: overton@cs.nyu.edu
WWW: http://cs.nyu.edu/cs/faculty/overton

Dedicated to Terry Rockafellar on the occasion of his sixtieth birthday.

Optimization problems involving eigenvalues arise in many different mathematical disciplines. This article is divided into two parts. Part I gives a historical account of the development of the field. We discuss various applications that have been especially influential, from structural analysis to combinatorial optimization, and we survey algorithmic developments, including the recent advance of interior-point methods for a specific problem class: semidefinite programming. In Part II we primarily address optimization of convex functions of eigenvalues of symmetric matrices subject to linear constraints. We derive a fairly complete mathematical theory, some of it classical and some of it new. Using the elegant language of conjugate duality theory, we highlight the parallels between the analysis of invariant matrix norms and weakly invariant convex matrix functions. We then restrict our attention further to linear and semidefinite programming, emphasizing the parallel duality theory and comparing primal-dual interior-point methods for the two problem classes. The final section presents some apparently new variational results about eigenvalues of nonsymmetric matrices, unifying known characterizations of the spectral abscissa (related to Lyapunov theory) and the spectral radius (as an infimum of matrix norms).

The dedication reflects the key role that convex analysis plays in the optimization of eigenvalues.

* Supported in part by the Natural Sciences and Engineering Research Council of Canada.
† Supported in part by the U.S. National Science Foundation under grant CCR-9401136.

CONTENTS

PART I: THE HISTORY

1. The shape of the strongest column

In 1773, Lagrange posed the following problem: determine the shape of the strongest axially symmetric column with prescribed length, volume and boundary conditions. The mathematical statement of this problem relies on earlier work of J. Bernoulli and Euler. The latter, in 1744, established the buckling load of such a column as the least eigenvalue of a self-adjoint fourth-order differential operator. Consequently, Lagrange's problem requires the maximization of this least eigenvalue, over all possible functions defining the cross-sectional area of the column.

Lagrange's problem, so easily stated, proved extraordinarily resistant to many attempts at its solution. Many authors made substantial contributions as well as serious errors. Lagrange must have the credit for posing the problem, yet several errors led to his incorrect conclusion that it is solved by the uniform column. The first to offer a correct solution was Clausen in 1851, in the case of clamped-free boundary conditions. The solution has the cigar shape shown in Fig. 1(a), where the cross-sectional area of the column is plotted as a function of its length. Clausen's paper is known primarily through later work of Pearson, who introduced many errors in an attempt to simplify the results.

Lagrange's problem then lay dormant for a century before it was taken up in a modern treatment by J. Keller in 1960. Keller established the solution, shown in Fig. 1(b), in the case of hinged-hinged boundary conditions. Then Tadjbakhsh and Keller (1962) offered solutions in the case of clamped-hinged and clamped-clamped boundary conditions. These are shown in Figs 1 (c) and (d) respectively. A conspicuous feature in both cases is the vanishing of the cross-sectional area at an internal point.

These solutions went unchallenged for fifteen years. Then Olhoff and Rasmussen (1977) claimed that the Tadjbaksh–Keller (TK) clamped-clamped solution was incorrect, because its solution procedure implicitly assumed that the least eigenvalue associated with the optimal solution is simple (that is, has multiplicity one). The solution offered by Olhoff and Rasmussen (OR), displayed in Fig. 1 (e), has a double least eigenvalue. However, no proof of the validity of this column was offered, nor were details of their numerical approximation procedure. Consequently, the issue remained quite controversial, with some authors defending the TK solution, and others, notably Masur and Seiranian, offering evidence for the OR solution. Recently, Cox and Overton (1992) gave the first proof of existence of a solution to the clamped-clamped problem, as well as the first proof that the OR solution indeed satisfies the Clarke (1983) first-order necessary conditions for optimality. In addition, Cox and Overton (1992) offered the first systematic numerical results using direct optimization techniques that take into account the possibility of a multiple eigenvalue. Both the theoretical contributions and the numerical techniques of Cox and Overton (1992) rely on the theory of convex analysis and its generalizations due to Rockafellar (1970) and Clarke (1983).

However, following in the footsteps of their illustrious predecessors in more ways than one, Cox and Overton also introduced a new error, claiming in an appendix that the TK clamped-hinged solution was also incorrect. Rather than believing their own numerical evidence, albeit uncertain given the vanishing of the cross-sectional area at an internal point and the corresponding absence of an existence proof (Cox and Overton 1992, p. 315), they placed

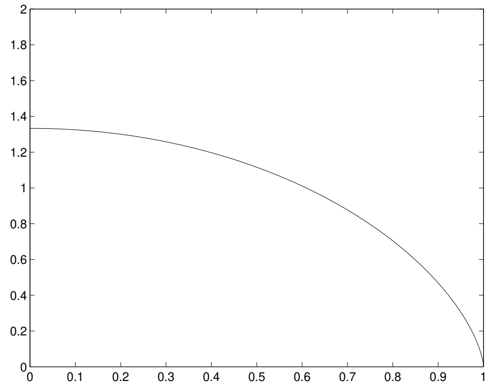

Fig. 1(a) Optimal clamped-free column

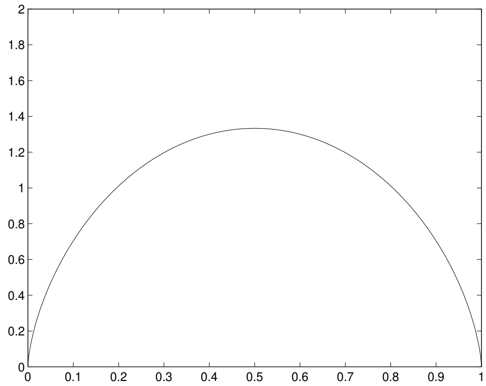

Fig. 1(b) Optimal hinged-hinged column

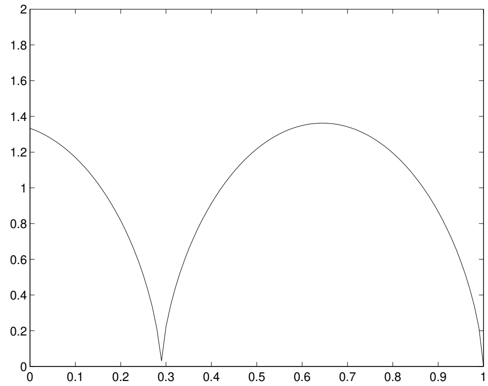

Fig. 1(c) Optimal clamped-hinged column

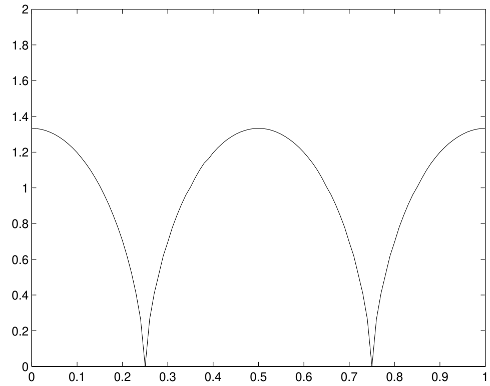

Fig. 1(d) TK clamped-clamped column

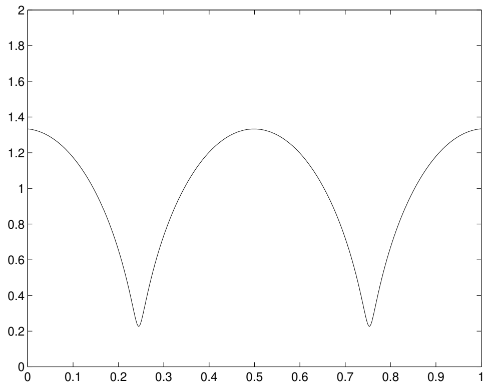

Fig. 1(e) Optimal clamped-clamped column faith in a mathematical proof that contained a simple scaling error: the irony will doubtless be enjoyed by the readers of this journal. That the TK clamped-hinged solution is indeed correct (though the TK clamped-clamped solution is not) has now been established beyond doubt by Cox and Maddocks (1996). For more details, see the article by Cox in *Math. Intelligencer* (Cox 1992), accompanied by illustrations of the strongest columns on the cover, and also the follow-up discussion (Kirmser and Hu 1993, Cox 1993).

2. Optimal partitioning of graphs

Our next example of eigenvalue optimization could not be more different in character to the strongest column problem. Consider a nonnegative edge-weighting of the complete (undirected) graph on the vertex set $\{1, 2, \ldots, n\}$. We can associate any such weighting with an $n \times n$ symmetric matrix W with diagonal entries all zero and off-diagonal entries all nonnegative: entry W_{ij} is just the weight on the edge (i, j). Given integers $d_1 \geq d_2 \geq \cdots \geq d_k > 0$, with sum n, consider the problem of partitioning the vertex set into k subsets such that the ith subset contains exactly d_i vertices and the sum of weights of edges between subsets is minimized. Equivalently, the sum of the weights of edges whose endpoints are both inside the same subset is to be maximized. This problem is NP-hard. However, Donath and Hoffman (1973) suggested the clever idea of deriving bounds on the solution by means of eigenvalue optimization. (For other approaches to graph partitioning that exploit eigenvalues, though not necessarily eigenvalue optimization, see the early work of Fiedler (1973) and the recent survey paper of Pothen (1996).)

Denote the characteristic (column) vector for the ith subset by $x^i \in \mathbb{R}^n$: thus x^i_r is 1 if vertex r is in subset i, and is 0 otherwise. Let X be the $n \times k$ matrix $[x^1, x^2, \ldots, x^k]$. Then, by construction, $X^T X = \mathrm{Diag}(d_1, d_2 \ldots, d_k)$, and we seek to maximize $\sum W_{ij}(XX^T)_{ij}$, or equivalently, the trace of WXX^T. Since for any matrices A and B we have $\mathrm{tr}\, AB = \mathrm{tr}\, BA$, we can write the partitioning problem as:

$$\begin{aligned} &\max_{X \in \mathbb{R}^{n \times k}} \mathrm{tr}\, X^T W X \\ &\text{subject to } \; X^T X = \mathrm{Diag}(d) \text{ and } X_{ij} \in \{0, 1\}. \end{aligned} \tag{2.1}$$

Now let us replace the variable matrix X by making the normalized definitions $y^i = d_i^{1/2} x^i$ and $Y = [y^1, y^2, \ldots, y^k]^T$. With this change of variable, the optimization problem becomes

$$\begin{aligned} &\max_{Y \in \mathbb{R}^{n \times k}} \sum_{i=1}^{k} d_i (y^i)^T W y^i \\ &\text{subject to } \; Y^T Y = I \text{ and } \sqrt{d_i d_j} Y_{ij} \in \{0, 1\}. \end{aligned} \tag{2.2}$$

The last constraint is the integrality constraint, which makes the problem difficult. So, let us *relax* the problem by dropping this constraint. As we shall see (in Lemma 10.2), the relaxed problem is solved by taking the columns of Y to be an orthonormal set of eigenvectors for the largest k eigenvalues of W: each y^i should be an eigenvector corresponding to $\lambda_i(W)$, the ith largest eigenvalue of W (counting multiplicities). The ordering is important since, by assumption, the d_i are given in descending order. Because the relaxation was obtained by dropping the integer constraint, the quantity $\sum_{i=1}^k d_i\lambda_i(W)$ is an upper bound for the optimal value of the problem (2.1).

Now we come to the key point: *a tighter upper bound can be obtained using eigenvalue optimization.* The diagonal elements of XX^T are all one, so we can replace the objective function of problem (2.1) by the trace of $(W+D)XX^T$ for any diagonal matrix D with zero trace. Equivalently, after the change of variables, we replace W by $W+D$ in the objective function of (2.2). Different choices of D give different relaxations when the integer constraint is dropped, and therefore different upper bounds. Thus D can be chosen to improve the upper bound, by *minimizing the weighted sum of the largest eigenvalues* of $W+D$, that is

$$G(D) = \sum_{i=1}^{k} d_i\lambda_i(W+D),$$

over all diagonal matrices D with zero trace.

Donath and Hoffman reasoned that since the function G is convex (as we shall see in Section 10), the task of minimizing G should be tractable. This turned out to be a more mathematically interesting and challenging problem than they anticipated at the time, as we shall now discuss.

3. Multiple eigenvalues, optimality conditions, and algorithms

Multiple eigenvalues had not been expected in the problem of Lagrange because, in all but the clamped-clamped case, the structure of the differential operator makes it impossible for the least eigenvalue, say $\hat\lambda_1$, to have multiplicity greater than one. If one considers more general eigenvalue optimization problems, however, it is clear that maximizing a least eigenvalue (equivalently minimizing a greatest eigenvalue) will potentially lead to coalescence of eigenvalues. Of course, *minimizing* a least eigenvalue has the opposite effect. The latter occurs, for example, in Rayleigh's problem of finding the shape of the two-dimensional drum with the least natural frequency. Mathematically, this means finding the shape of the domain that minimizes $\hat\lambda_1$, the least eigenvalue of the Laplacian. The least eigenvalue is necessarily simple, and the solution is a circle. An interesting variation is to find the shape that minimizes the ratio $\hat\lambda_1/\hat\lambda_2$. This was considered by Payne, Pólya and Weinberger (1956): they conjectured that the solution is a circle, but this was proved only recently (Ashbaugh and Benguria 1991). In this case, a double eigenvalue plays a role, because $\hat{\lambda}_2$ and $\hat{\lambda}_3$ coalesce at the solution. Eigenvalue optimization problems for plates (modelled by fourth-order differential operators in two dimensions) are also of interest, but these have received relatively little attention. All of these problems are difficult because they are infinite-dimensional and the operators depend on the variables in a complicated way. For the remainder of this article we confine our attention to matrix problems with linear dependence on the variables.

The Donath–Hoffman approach to graph partitioning requires minimizing a weighted sum of the largest eigenvalues of a matrix, the variables being simply the diagonal elements. This work led to a paper of Cullum, Donath and Wolfe (1975) that is remarkable for two significant contributions. The first was the development of an optimality condition using convex analysis, emphasizing the issue of multiple eigenvalues. Specifically, the authors recognized and addressed the fact that the sum-of-eigenvalues function, although convex, is not a differentiable function at points where the eigenvalues coalesce. The second contribution of Cullum et al. (1975) was the development of a convergent algorithm to find a minimizer. The significance of this work was not appreciated for some ten years or so. Then Fletcher (1985) revived interest in the problem, inspiring further analytical improvements by Overton and Womersley (1993) and Hirriart-Urruty and Ye (1995). These results are now largely subsumed by a more general but concise approach due to Lewis (1996*a*), presented in Part II of this survey. Specifically, rather general composite functions of the form $h \circ \lambda$ are considered, where λ is the eigenvalue map from symmetric matrix space to $\mathbb{R}^n$, and h is *any* convex function that is symmetric with respect to its arguments. A duality theory for this class of functions will be given in some detail, building on the fundamental results of convex analysis due to Rockafellar as well as key matrix theoretic results of von Neumann and others. Composite eigenvalue optimization includes semidefinite programming (SDP), a generalization of linear programming that has received much attention in the last few years.

The SDP problem is to minimize a linear function of a symmetric matrix variable subject to linear and positive semidefinite constraints on the matrix. Typically, SDPs have solutions with multiple zero eigenvalues. Semidefinite constraints have been considered in many contexts; two early papers are Bellman and Fan (1963) and Craven and Mond (1981). In fact, SDP was the variant of eigenvalue optimization that was primarily addressed by Fletcher (1985), introducing a new algorithmic approach and emphasizing the issues of multiple eigenvalues and quadratic convergence. This led to the computational work on minimizing a maximum eigenvalue due to Overton (1988, 1992) and the associated second-order convergence analysis (a complicated issue in the presence of multiple eigenvalues) given by Overton and Womersley (1995) and Shapiro and Fan (1995). However, since many eigenvalue optimization problems can be rephrased as equivalent SDPs, this work has now been largely overshadowed by the sudden advance of interior-point methods for SDP, to which we now turn.

4. Interior-point methods and polynomial-time algorithms

Linear programming (LP) was established as a discipline in the 1940s by Dantzig. The LP problem is to minimize a linear function subject to linear equality and inequality constraints on the variables, a problem which, remarkably, had largely escaped earlier attention, with the exception of some work on systems of linear inequalities by Fourier and Motzkin. As well as introducing the problem class, Dantzig gave an algorithm for solving LPs: the simplex method. Duality played a key role from the beginning, originating in a famous conversation between Dantzig and von Neumann at Princeton in 1947; see Dantzig (1991). The highly efficient simplex method went essentially unchallenged for 30 years, although it was known that, in the worst case, it required computation time exponential in the problem size. In 1979 Khaciyan showed that the ellipsoid method of Nemirovskii and Shor could be used to guarantee the solution of LPs in polynomial time. The ellipsoid method proved to be impractical, but it inspired the work of Karmarkar (1984), which established the interior-point framework as a practical, polynomial-time approach to solving LP. In the 10 years since, a profusion of interior-point methods for LP have been proposed, implemented and theoretically analysed; see the surveys by Lustig, Marsten and Shanno (1994), Gonzaga (1992) and Wright (1992). It is now generally accepted that the primal-dual interior-point method due to Monteiro and Adler (1989) and Kojima, Mizuno and Yoshise (1989) has substantial theoretical and practical advantages over the other interior-point methods, including Karmarkar's method.

As we already noted, the difference between LP and SDP is that, in the latter case, the variable is a symmetric matrix and the inequality constraint is a semidefinite matrix constraint. In the case that the matrix is constrained to be diagonal, SDP reduces to LP. There is no simplex method for SDP, because the feasible region is not polyhedral. In the late 1980s, Nesterov and Nemirovskii extended many of the interior-point methods and theoretical results from LP to a much broader class of convex programming problems, including SDP; see Nesterov and Nemirovskii (1994). Alizadeh (1991, 1995) and Karmarkar and Thakur (1992) also independently proposed such a generalization for SDP, a key component being the 'log determinant' barrier function. In the last three years there has been a burst of activity in the development of interior-point methods for SDP. Some of the most recent work, namely the derivation of a primal-dual interior-point method for SDP, will be discussed in Section 14. See Vandenberghe and Boyd (1996) for a survey article on SDP, including many applications not discussed here.

We now briefly discuss two important application areas that have successfully exploited the success of interior-point methods for SDP.

5. Polynomial-time approximations to NP-hard graph problems

The availability of polynomial-time algorithms for semidefinite programming has led to great interest by the combinatorial optimization community in provably good polynomial-time approximations to NP-hard problems. We consider one example.

As in Section 2, consider the complete graph with vertex set $\{1, 2, \ldots, n\}$ and edges (i, j) with associated nonnegative weights W_{ij}. The max-cut problem is to divide the vertices into two sets, V_1 and V_2, such that the weighted sum of edges crossing from one set to the other is maximized. This is *not* the same as the graph partitioning problem with $k = 2$ since the number of vertices in each set is not preassigned. The max-cut problem is NP-hard, although the min-cut (max-flow) problem can be solved by standard fast algorithms. (The min-cut problem is trivial if one does not specify that V_1 and V_2 must be nonempty). The max-cut problem can be expressed as

$$\max_{x_1, x_2, \ldots, x_n \in \mathbb{R}} \left\{ \sum_{1 \le i \le j \le n} W_{ij}(1 - x_i x_j) : |x_i| = 1 \text{ for all } i \right\}, \tag{5.1}$$

where we adopt the convention that $x_i = 1$ means $i \in V_1$ and $x_i = -1$ means $i \in V_2$. Now consider the modified problem

$$\max_{x^1, x^2, \ldots, x^n \in \mathbb{R}^n} \left\{ \sum_{1 \le i \le j \le n} W_{ij}(1 - (x^i)^T x^j) : \|x^i\| = 1 \text{ for all } i \right\}, \tag{5.2}$$

where $\|\cdot\|$ denotes the Euclidean norm. If the vectors $x^1, \ldots, x^n$ solving problem (5.2) all happen to be parallel, then they can be associated with the scalar solutions $x_i = \pm 1$ to problem (5.1), and the max-cut problem is solved. Of course, this is very unlikely to occur. However, given any fixed optimal solution of problem (5.2), we can generate a cut for the graph by cutting the unit ball in half, and then assigning vertex i to set V_1 or V_2 according to which half of the ball contains the vector x^i. Goemans and Williamson (1996) recently established the surprising fact that, if one makes the division of the unit ball in the appropriate way, the resulting cut in the graph is an approximate solution of the max-cut problem with an objective value within a factor of 1.14 of the optimal value. Notice that problem (5.2) is equivalent to the SDP

$$\min_{Z \succeq 0} \{\operatorname{tr} WZ : Z_{ii} = 1, \text{ for each } i\},$$

the variable Z being a symmetric matrix associated with the vectors x^i by the equation $Z = X^T X$, where X is the matrix $[x^1, x^2, \ldots, x^n]$, and where $Z \succeq 0$ denotes the semidefinite constraint.

To summarize, the max-cut problem, which is NP-hard, is provably solvable within a factor of 1.14 in polynomial time, via the solution of a semidefinite program. For more on the max-cut problem, see the survey by Poljak and Tuza (1993) and the recent thesis of Helmberg (1994). For other applications of SDP and eigenvalue optimization to combinatorial optimization, see Grötschel, Lovász and Schriver (1988, Chapter 9), Mohar and Poljak (1993) and Rendl and Wolkowicz (1992).

6. Linear matrix inequalities in system and control theory

The title of this section is also the title of a recent book (Boyd, Ghaoui, Feron and Balakrishnan 1994). A *linear matrix inequality* (LMI) is generally understood to mean a positive semidefinite or definite constraint on a matrix depending affinely on parameters: as such, an LMI is simply the constraint of an SDP. However, the term is also sometimes used to describe more general matrix inequality constraints, especially bounds on the eigenvalues of a pencil (those scalars λ satisfying $\det(A - \lambda B) = 0$, where the matrices A and B are symmetric and depend affinely on parameters, and B is positive definite). The application of LMIs to control theory has its origins in the work of Lyapunov in the 1890s and Yakubovitch in the 1960s.

The impact of LMIs on system and control theory is hard to overstate: it is fair to say that the field has been revolutionized by the realization that optimization problems with LMI constraints can be effectively solved using interior-point methods. We give no further details here since the relevant material is available in Boyd et al. (1994).

7. Non-Lipschitz eigenvalue optimization

Up to this point we have discussed eigenvalue optimization for symmetric matrices and self-adjoint operators, which have real eigenvalues and orthonormal sets of eigenvectors. Eigenvalues of nonsymmetric matrices and operators also play many roles in applied mathematics, though it is well known that their potential sensitivity to perturbation requires caution. Stability issues arise in many applications, with instability generally associated with eigenvalues whose real parts are nonnegative. Indeed, the widespread use of symmetric linear matrix inequalities in system and control theory is, in part, motivated by stability issues for nonsymmetric matrices, via Lyapunov theory and its generalizations. It is therefore natural to consider direct application of

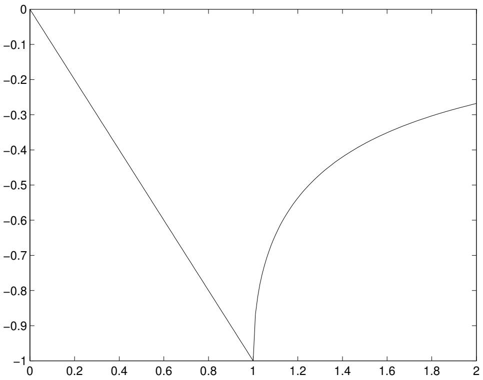

Fig. 2 Spectral abscissa for the damped linear oscillator

optimization theory to functions of eigenvalues of nonsymmetric matrices. However, this is quite complicated because of the non-Lipschitz behaviour of the eigenvalues.

The damped linear oscillator provides a simple example of eigenvalue optimization in the nonsymmetric case. Consider the ordinary differential equation, for a given real b,

$$y''(t) + 2by'(t) + y(t) = 0. \tag{7.1}$$

Noting that the vector $z(t) = [y(t)\ y'(t)]^T$ satisfies the first-order system $z'(t) = A(b)z(t)$ where $A(b) = \begin{bmatrix} 0 & 1 \\ -1 & -2b \end{bmatrix}$, the initial value problem may be solved in terms of the eigenvalues and eigenvectors of $A(b)$. The effectiveness of the damping is measured by the spectral abscissa of $A(b)$ (that is, the largest real part of the eigenvalues of $A(b)$): we denote this function by $\alpha(b)$. Now $\alpha(b) = -b + \operatorname{Re}\sqrt{b^2-1}$, so, since the spectral abscissa achieves its minimum at $b = 1$, equation (7.1) is said to be over(under)damped if $b > 1$ ($b < 1$), and critically damped if $b = 1$. The function $\alpha(b)$ is plotted in Fig. 2. Note that α is not a Lipschitz function of b. The sharply different behaviour of the function α on the two sides of the minimizer occurs because, on one side, a double eigenvalue splits into a real pair, while on the other side, it splits into a complex conjugate pair. In both cases the changes in the eigenvalues are non-Lipschitz, but only in the former case do the real parts have non-Lipschitz behaviour. The optimal damping factor $b = 1$ yields a matrix $A(b)$ with an eigenvalue having algebraic multiplicity two, but geometric multiplicity one, and thus with a nontrivial Jordan block.

A similar phenomenon is well known from the analysis of the successive overrelaxation (SOR) iterative method for solving systems of linear equations;

see Ortega (1972). The critical value of the overrelaxation parameter is determined by an eigenvalue optimization problem in one variable. Over- and underrelaxation are well known to have very different consequences, again because of the presence of a nontrivial Jordan block at the minimizing point.

Of course, non-Lipschitz eigenvalue optimization problems also arise in more than one variable. Cox and Overton (1996) treat a generalization of the damped linear oscillator, namely the damped wave equation. Ringertz (1996) considers applications to stability issues for aircraft design.

Optimality conditions for non-Lipschitz eigenvalue optimization are rather complicated and beyond the scope of this article. For the present state of the art, see Burke and Overton (1994) and Overton and Womersley (1988).

Indeed, consider the following, far simpler question. Suppose A is a nonsymmetric matrix with multiple eigenvalues, and consider the eigenvalues of the perturbed matrix $A + \epsilon B$, where the matrix B is arbitrary and ϵ is a scalar perturbation parameter. How can we quantify the leading terms of the expansions of these eigenvalues in fractional powers of ϵ? When A has nontrivial Jordan structure, the behaviour of the eigenvalues under perturbation is quite complicated. Apparently, the only book that addresses this issue is Baumgärtel (1985), building on results of Lidskii and others published in the Russian literature in the 1960s, but remaining largely unknown in the West. See Moro, Burke and Overton (1995) for discussion of Lidskii's results and connections with the classical Newton diagram.

In the final section of this article we derive some apparently new variational results for functions of eigenvalues of nonsymmetric matrices. One special case amounts to a characterization of the spectral abscissa as the optimal value of a *symmetric* matrix eigenvalue optimization problem, a result well known to control theorists and one which may be viewed as a quantitative version of Lyapunov theory. Another special case implies the well known result that the spectral radius may be characterized as the infimum of all submultiplicative matrix norms. These results suggest a possible approach to non-Lipschitz eigenvalue optimization by means of symmetric eigenvalue optimization.

PART II: THE MATHEMATICS

8. Conjugacy

Convex analysis is an elegant and powerful tool for studying duality in optimization. Particularly for linearly constrained problems, it provides a concise and flexible framework. We begin by summarizing the relevant ideas.

Let E be a Euclidean space, by which we mean a finite-dimensional, real inner-product space. We could, of course, always identify E with $\mathbb{R}^n$, but a less concrete approach helps our future development. We call a real function f on E a *prenorm* if it is continuous, and satisfies

- *homogeneity:* $f(\alpha x) = |\alpha| f(x)$ for all real α and points x in E
- *positivity:* $f(x) > 0$ for all nonzero points x in E.

A norm is then just a prenorm satisfying the triangle inequality. For a prenorm f, we can define a real function f^D on E by

$$f^D(y) = \max\{\langle x, y\rangle : f(x) = 1\}.$$

The function f^D is actually a norm: we call it the *dual norm* of f.

Theorem 8.1. (von Neumann, 1937) A prenorm f is a norm if and only if $f = f^{DD}$.

The reader may consult Horn and Johnson (1985) for these ideas.

In optimization it is very convenient to consider *extended-real* functions $f : E \to [-\infty, +\infty]$. We call such a function *convex* (respectively *closed, polyhedral*) if its *epigraph* $\{(x, r) \in E \times \mathbb{R} : f(x) \leq r\}$, is a convex (respectively closed, polyhedral) set. The *domain* of f is the set

$$\operatorname{dom} f = \{x \in E : f(x) < +\infty\};$$

if this set is nonempty and if f never takes the value $-\infty$, then f is called *proper*. For any extended-real function f we can define an extended-real function f^* on E by

$$f^*(y) = \sup\{\langle x, y\rangle - f(x) : x \in E\}.$$

The function f^* is always closed and convex: we call it the *(Fenchel) conjugate* of f. The basic reference for these and later convex-analytic ideas is Rockafellar (1970). Our definition of a closed function is slightly different from that of Rockafellar (1970): the definitions coincide for proper functions.

Theorem 8.2. (Fenchel–Hörmander, 1949) Suppose the extended-real function f is proper. Then f is closed and convex if and only if $f = f^{**}$. In this case, f^* is also proper.

The ideas of dual norms and conjugate functions are closely related: if f is a norm then a short calculation shows

$$(f^2/2)^* = (f^D)^2/2. \tag{8.1}$$

The first-order behaviour of a function $f : E \to (-\infty, +\infty]$ at a point x in its domain is fundamental to any study of optimality conditions and algorithms. For convex f this behaviour is encapsulated in the *subdifferential*

$$\partial f(x) = \{y \in E : \langle y, z - x\rangle \leq f(z) - f(x) \text{ for all } z \text{ in } E\}.$$

Specifically, the directional derivative of f at x in a direction $w \in E$ is given by the formula

$$f'(x;w) = \sup\{\langle w, y\rangle : y \in \partial f(x)\}.$$

In particular, f is differentiable at x exactly when its subdifferential there is a singleton ($\partial f(x) = \{\nabla f(x)\}$). If $f(x)$ is infinite, we define $\partial f(x) = \emptyset$.

Immediately from its definition, we can relate the subdifferential to the conjugate:

$$y \in \partial f(x) \quad \Leftrightarrow \quad f(x) + f^*(y) = \langle x, y\rangle. \tag{8.2}$$

Using the Fenchel–Hörmander Theorem (Theorem 8.2), we deduce that for a proper closed convex function f, the subdifferential map can be 'inverted':

$$y \in \partial f(x) \quad \Leftrightarrow \quad x \in \partial f^*(y). \tag{8.3}$$

Example 8.3. (cones) One benefit of convex analysis is the possibility of studying a subset K of E through its *indicator function*

$$\delta_K(x) = \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{otherwise.} \end{cases}$$

This function is convex (closed) exactly when K is convex (closed). Suppose K is a *cone*: that is, $\mathbb{R}_+ K = K$. Then we deduce immediately that the function δ_K^* is just δ_{K^-}, the indicator function of the *polar* cone

$$K^- = \{y \in E : \langle x, y\rangle \leq 0 \text{ for all } x \text{ in } K\}.$$

The Fenchel–Hörmander Theorem (Theorem 8.2) then shows that a cone K is closed and convex exactly when $K^{--} = K$. From the subdifferential property (8.2) we deduce the 'complementarity' condition

$$y \in \partial \delta_K(x) \quad \Leftrightarrow \quad x \in K,\ y \in K^-, \text{ and } \langle x, y\rangle = 0. \tag{8.4}$$

In particular, if the space E is $\mathbb{R}^n$ and the cone K is the nonnegative orthant $\mathbb{R}^n_+$, then the polar K^- is $-\mathbb{R}^n_+$, and for vectors x and y in $\mathbb{R}^n_+$ we deduce

$$y \in \partial \delta_{\mathbb{R}^n_+}(x) \quad \Leftrightarrow \quad x_j \geq 0,\ y_j \leq 0, \text{ and } x_j \text{ or } y_j = 0 \text{ for each } j. \tag{8.5}$$

When f is a norm, the subdifferential property (8.2) has a simple analogue. For nonzero points x in E, an easy calculation shows

$$y \in \partial f(x) \quad \Leftrightarrow \quad f(x) = \langle x, y\rangle \text{ and } f^D(y) = 1, \tag{8.6}$$

while $\partial f(0) = \{y \in E : f^D(y) \leq 1\}$.

The duality theory of linearly-constrained convex optimization is particularly transparent in this framework. We will always consider $\mathbb{R}^m$ as a Euclidean space of column vectors, with the standard inner product. Given a linear map $A : E \to \mathbb{R}^m$, we define the *adjoint* map $A^* : \mathbb{R}^m \to E$ by the property

$$y^T(Ax) = \langle A^* y, x\rangle \text{ for all points } x \text{ in } E \text{ and } y \text{ in } \mathbb{R}^m.$$

Suppose the function $f : E \to (-\infty, +\infty]$ is closed, convex and proper, fix a vector b in $\mathbb{R}^m$, and consider the pair of optimization problems,

$$\begin{aligned} \text{Primal:} \quad \rho &= \inf\{f(x) : x \in E,\ Ax = b\}; \\ \text{Dual:} \quad \delta &= \sup\{y^T b - f^*(A^* y) : y \in \mathbb{R}^m\}. \end{aligned}$$

The following result is derived from theory due to Rockafellar, dating from 1963 (Rockafellar 1970). We say the primal problem is *superconsistent* if there is a point $\hat{x}$ in $\mathrm{int}(\mathrm{dom}\, f)$ satisfying $A\hat{x} = b$, and we say the dual problem is *superconsistent* if there is a point $\hat{y}$ in $\mathbb{R}^m$ with $A^*\hat{y}$ in $\mathrm{int}(\mathrm{dom}\, f^*)$. By 'consistent' we mean the same properties with 'int' omitted.

Theorem 8.4. (Fenchel Duality)

(i) *Weak duality*: $\rho \geq \delta$.

(ii) *Dual attainment*: if the primal is superconsistent, then $\rho = \delta$, and δ is attained, if finite. Furthermore, if A is surjective, then, for any real α, the set

$$\{y \in \mathbb{R}^m : y^T b - f^*(A^* y) \geq \alpha\}$$

is compact.

(iii) *Primal attainment*: if the dual is superconsistent, then $\rho = \delta$, and ρ is attained if finite. Furthermore, for any real α, the set

$$\{x \in \mathbb{R}^n : f(x) \leq \alpha,\ Ax = b\}$$

is compact.

(iv) *Polyhedrality*: if f is polyhedral and either problem is consistent, then the other problem is attained, if finite, and $\rho = \delta$.

(v) *Complementary slackness*: suppose $\rho = \delta$. Then points $\overline{x}$ and $\overline{y}$ are optimal for the primal and the dual problems respectively, if and only if $A\overline{x} = b$ and $A^*\overline{y} \in \partial f(\overline{x})$.

The *complementary slackness* condition $A^*\overline{y} \in \partial f(\overline{x})$ is equivalent to $\overline{x} \in \partial f^*(A^*\overline{y})$, by the inversion formula (8.3). If in addition f^* is differentiable at $A^*\overline{y}$ then the primal solution $\overline{x}$ must therefore be $\nabla f^*(A^*\overline{y})$. In these circumstances we are thus able to recover a primal optimal solution by solving the dual problem.

A nice exercise is to apply the Fenchel Duality Theorem (Theorem 8.4) and Example 8.3 to the 'cone optimization problem'

$$\inf\{\langle c, x\rangle : Ax = b,\ x \in K\},$$

for a convex cone K and an element c of E. This model (*cf.* Nesterov and Nemirovskii 1994) subsumes both linear and semidefinite programming, which we discuss later.

9. Invariant norms

The theoretical foundations of eigenvalue optimization parallel the better-known theory of invariant matrix norms pioneered by von Neumann. A brief sketch of this theory's salient features is therefore illuminating. For clarity we consider only square, real matrices.

We consider the Euclidean space M_n of $n \times n$ real matrices, where the inner product is defined by $\langle X, Y \rangle = \operatorname{tr} X^T Y$. The singular values of a matrix X in M_n we denote $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X)$. In this way we define the 'singular value map' $\sigma : M_n \to \mathbb{R}^n$.

We denote the groups of $n \times n$ permutation and orthogonal matrices by P_n and O_n respectively. We call a function f on $\mathbb{R}^n$ *symmetric* if, for any point x in $\mathbb{R}^n$ and any matrix Q in P_n, we have $f(Qx) = f(x)$. We say a norm ϕ on M_n is *(orthogonally) invariant* if, for any matrices X in M_n, and U and V in O_n, we have $\phi(UXV) = \phi(X)$.

For a vector x in $\mathbb{R}^n$, we denote the diagonal matrix with diagonal entries $x_1, x_2, \ldots, x_n$ by $\operatorname{Diag} x$. Clearly, for any invariant norm ϕ on M_n, the real function g on $\mathbb{R}^n$ defined by $g(x) = \phi(\operatorname{Diag} x)$ is a symmetric norm that is also *absolute*: $g((|x_1|, |x_2|, \ldots, |x_n|)^T) = g(x)$ for all vectors x in $\mathbb{R}^n$. Such norms are called *symmetric gauges*. The original norm ϕ is just the composite function $g \circ \sigma$. A beautiful result of von Neumann shows that this property characterizes invariant norms.

Theorem 9.1. (von Neumann, 1937) Invariant matrix norms are exactly those composite functions of the form $g \circ \sigma$, where g is a symmetric gauge.

For our purposes, almost more important than the result is the proof technique. Naturally, it relies heavily on the existence of an 'ordered singular value decomposition' for any matrix X:

$$X = U(\operatorname{Diag} \sigma(X))V \text{ for some orthogonal } U \text{ and } V.$$

If a second matrix Y satisfies $Y = U(\operatorname{Diag} \sigma(Y))V$, then we say X and Y have a *simultaneous ordered singular value decomposition*. Von Neumann's key step was the following result, of substantial independent interest.

Lemma 9.2. (von Neumann, 1937) Any $n \times n$ real matrices X and Y satisfy the inequality

$$\operatorname{tr} X^T Y \leq \sigma(X)^T \sigma(Y);$$

equality holds if and only if X and Y have a simultaneous ordered singular value decomposition.

Equipped with this (nontrivial) result, von Neumann's characterization (Theorem 9.1) follows from a beautifully transparent duality argument. For an absolute, symmetric prenorm g on $\mathbb{R}^n$, we first use Lemma 9.2 to deduce that the prenorm $g \circ \sigma$ satisfies

$$(g \circ \sigma)^D = g^D \circ \sigma. \tag{9.1}$$

Hence if g is actually a symmetric gauge, applying this formula twice and using Theorem 8.1, we deduce

$$(g \circ \sigma)^{DD} = (g^D \circ \sigma)^D = g^{DD} \circ \sigma = g \circ \sigma,$$

and, by Theorem 8.1, $g \circ \sigma$ must be a norm. The result is now easy to see.

Lemma 9.2 also greatly facilitates the calculation of subdifferentials. The following result, due to Ziętak (1993) (*cf.* Watson (1992)) follows immediately from the Lemma, the subdifferential characterization (8.6), and the duality formula (9.1) (*cf.* Lewis (1995*a*)).

Theorem 9.3. (Ziętak, 1994) If g is a symmetric gauge, then matrices X and Y satisfy $Y \in \partial(g \circ \sigma)(X)$ if and only if they have a simultaneous ordered singular value decomposition and satisfy $\sigma(Y) \in \partial g(\sigma(X))$.

Such techniques help reveal the intimate geometric connections between the two norms g and $g \circ \sigma$. For example, $g \circ \sigma$ is strict (respectively smooth) if and only if g is: see Arazy (1981) and Ziętak (1988). Furthermore, the facial structure of the unit ball of $g \circ \sigma$ can be derived from that of g (de Sá 1994*a*, 1994*b*, 1994*c*).

Example 9.4. (invariant approximation) Given a subspace of matrices and an invariant norm $g \circ \sigma$ (where g is a symmetric gauge), suppose we wish to approximate, in the norm $g \circ \sigma$, a given matrix by a matrix from the given subspace. We can rewrite this problem, for a suitable choice of matrices A_i and reals b_i (for $i = 1, 2, \ldots, m$), as

$$\inf_{X \in M_n} \{(g(\sigma(X)))^2/2 : \operatorname{tr} A_i^T X = b_i \text{ for each } i\}. \tag{9.2}$$

By the Fenchel Duality Theorem (Theorem 8.4) and the dual norm equation (8.1), both this problem and its dual

$$\sup_{y \in \mathbb{R}^m} \left\{ \sum_i b_i y_i - \frac{1}{2} \left(g^D \left(\sigma \left(\sum_i y_i A_i \right) \right) \right)^2 \right\} \tag{9.3}$$

have optimal solutions, with equal optimal values: the form of the dual is a consequence of the duality formula (9.1). If the norm g is strict (that is, the unit sphere $\{x : g(x) = 1\}$ contains no line segments) then its dual norm g^D is smooth (see for example Deville, Godefroy and Zizler (1993, II.1.6)), whence so is $g^D \circ \sigma$: Ziętak's Theorem (Theorem 9.3) provides a simple formula for $\nabla(g^D \circ \sigma)$ in terms of ∇g^D. Then the dual problem (9.3) is an unconstrained, smooth, concave maximization, and if the vector $\overline{y}$ is a solution, then the unique primal optimal solution is given by

$$X = g^D(\sigma(\overline{Y}))U(\mathrm{Diag}\,\nabla g^D(\sigma(\overline{Y})))V,$$

where $\overline{Y} = \sum_i \overline{y}_i A_i$, and U and V are any orthogonal matrices for which $\overline{Y} = U(\mathrm{Diag}\,\sigma(\overline{Y}))V$.

10. Functions of eigenvalues

We turn next to our principal interest: variational properties of eigenvalues. Our development mimics that of the previous section. An invariant matrix function is simply an absolute, symmetric function of the singular values. Analogously, a function of a symmetric matrix X that is invariant under transformations $X \mapsto U^T X U$, for all orthogonal matrices U, must be a symmetric function of the eigenvalues of X.

We consider the Euclidean space S_n of $n \times n$ real symmetric matrices, where the inner product is defined by $\langle X, Y\rangle = \mathrm{tr}\, XY$. We denote the eigenvalues of a matrix X in S_n by $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$. In this way we define the 'eigenvalue map' $\lambda : M_n \to \mathbb{R}^n$.

We say a function ψ on S_n is *weakly (orthogonally) invariant* if, for any matrices X in S_n and U in O_n, we have $\psi(U^T X U) = \psi(X)$. Clearly, for any weakly invariant convex function ψ on S_n, the extended-real function h on $\mathbb{R}^n$ defined by $h(x) = \psi(\mathrm{Diag}\, x)$ is symmetric and convex. Remarkably, just like von Neumann's Theorem (Theorem 9.1), this property is actually a characterization.

Theorem 10.1. (Davis, 1957) Functions on S_n that are weakly invariant and convex are exactly those composite functions of the form $h \circ \lambda$, where the function $h : \mathbb{R}^n \to [-\infty, +\infty]$ is symmetric and convex.

For proofs of this result, see Davis (1957), Martínez-Legaz (1995) and Lewis (1996*c*). A rather different characterization when the function h is differentiable may be found in Friedland (1981).

To pursue our analogy, we sketch a revealing, duality-based proof when the functions are closed. It begins with an analogue of von Neumann's Lemma (Lemma 9.2), for symmetric matrices. The inequality is actually an easy consequence of von Neumann's; the condition for equality is due to Theobald (1975). We say that two matrices X and Y in S_n have a *simultaneous ordered spectral decomposition* if there is an orthogonal matrix U with $X = U^T(\mathrm{Diag}\,\lambda(X))U$ and $Y = U^T(\mathrm{Diag}\,\lambda(Y))U$.

Lemma 10.2. (von Neumann–Theobald) Any $n \times n$ real symmetric matrices X and Y satisfy the inequality

$$\mathrm{tr}\, XY \leq \lambda(X)^T \lambda(Y);$$

equality holds if and only if X and Y have a simultaneous ordered spectral decomposition.

As in the singular value case, this inequality is the key tool. We first use it to prove that any extended-real symmetric function h satisfies

$$(h \circ \lambda)^* = h^* \circ \lambda. \tag{10.1}$$

Hence if h is also closed, proper and convex, then applying this formula twice and using the Fenchel–Hörmander Theorem (Theorem 8.2), we deduce

$$(h \circ \lambda)^{**} = (h^* \circ \lambda)^* = h^{**} \circ \lambda = h \circ \lambda,$$

and, by Theorem 8.2, $h \circ \lambda$ must be convex. Theorem 10.1 is now easy to see.

Very much as in the invariant norm case, the von Neumann–Theobald Lemma (Lemma 10.2) helps in the computation of subdifferentials. Using the Lemma, the subdifferential characterization (8.2), and the conjugacy formula (10.1), we obtain the following result (Lewis 1996*a*).

Theorem 10.3. (Lewis, 1996) If the function $h : \mathbb{R}^n \to (-\infty, \infty]$ is symmetric and convex, then matrices X and Y satisfy $Y \in \partial(h \circ \lambda)(X)$ if and only if they have a simultaneous ordered spectral decomposition and satisfy $\lambda(Y) \in \partial h(\lambda(X))$.

There are similar results for smooth and nonsmooth, nonconvex functions (Lewis 1996*b*, Tsing, Fan and Verriest 1994). Special versions of some of these ideas appeared independently in Barbara and Crouzeix (1994).

As in the invariant norm case, geometric/analytic properties of the two functions h and $h \circ \sigma$ are intimately related: strict convexity and smoothness are examples (Lewis 1996*a*). Furthermore, if the convex subset C of $\mathbb{R}^n$ is *symmetric* (that is, $PC = C$ for all matrices P in P_n), then by applying Davis's Theorem (Theorem 10.1), to the function $\delta_C \circ \lambda$ we see that the matrix set $\lambda^{-1}(C) = \{X \in S_n : \lambda(X) \in C\}$ is also convex: the extremal and facial structure of $\lambda^{-1}(C)$ may be deduced from that of C (Lewis 1996*a*, Lewis 1995*b*). Similar examples appear in Seeger (1996) and Martínez-Legaz (1995).

The parallel between the invariant norm case in the previous section and the development in this section is not accidental. There is a deeper, algebraic structure underlying both theorems (Lewis 1995*c*, Lewis 1996*c*).

Example 10.4. (semidefinite cone) Starting with the indicator function of the positive orthant, $\delta_{\mathbb{R}^n_+}$, the composite function $\delta_{\mathbb{R}^n_+} \circ \lambda$ is just the indicator function of the cone of positive semidefinite matrices. We denote this cone S_n^+, and for matrices X and Y in S_n, we write $X \succeq Y$ if $X - Y \in S_n^+$. The conjugacy formula (10.1) and Example 8.3 show Fejer's result that the positive semidefinite cone is 'self-dual' (that is, $(S_n^+)^- = -S_n^+$), since

$$\delta_{(S_n^+)^-} = \delta^*_{S_n^+} = (\delta_{\mathbb{R}^n_+} \circ \lambda)^* = \delta^*_{\mathbb{R}^n_+} \circ \lambda = \delta_{-\mathbb{R}^n_+} \circ \lambda = \delta_{-S_n^+}.$$

Furthermore, if matrices X and Y in S_n^+ satisfy $\operatorname{tr} XY = 0$, then in fact they must satisfy $XY = 0$. To see this, note that from the complementarity condition (8.4) and the self-duality of S_n^+, we deduce $-Y \in \partial\delta_{S_n^+}(X)$. By the subdifferential characterization, Theorem 10.3, X and $-Y$ have a simultaneous ordered spectral decomposition, and $\lambda(-Y) \in \partial\delta_{\mathbb{R}^n_+}(\lambda(X))$, whence (by relation (8.5)) $\lambda_j(X)\lambda_j(-Y) = 0$ for each j. Thus for some orthogonal matrix U,

$$\begin{aligned} -XY &= (U^T(\operatorname{Diag}\lambda(X))U)(U^T(\operatorname{Diag}\lambda(-Y))U) \\ &= U^T(\operatorname{Diag}[\lambda_j(X)\lambda_j(-Y)])U = 0. \end{aligned}$$

Example 10.5. (logarithmic barrier) For vectors x and y in $\mathbb{R}^n$, we write $x > y$ if $x_j > y_j$ for each index j. For matrices X and Y in S_n, we write $X \succ Y$ if $X - Y$ is positive definite. Define a symmetric closed convex function $h : \mathbb{R}^n \to (-\infty, +\infty]$ by

$$h(x) = \begin{cases} -\sum_j \log x_j & \text{if } x > 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{10.2}$$

(Henceforth we will interpret $\log\alpha$ as $-\infty$ for any nonpositive real α.) The corresponding matrix function is

$$(h \circ \lambda)(X) = \begin{cases} -\log\det X & \text{if } X \succ 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{10.3}$$

(Analogously, we henceforth interpret $\log\det X$ as $-\infty$ unless the symmetric matrix X is positive definite.) By Davis's Theorem (Theorem 10.1), this function is convex (and in fact essentially strictly convex, since h is; see Lewis (1996*a*). Using Theorem 10.3, a simple exercise shows, for positive definite X,

$$\nabla(h \circ \lambda)(X) = -X^{-1}. \tag{10.4}$$

Since $h^*(y) = -n + h(-y)$, we deduce from the conjugacy formula (10.1),

$$(h \circ \lambda)^*(Y) = \begin{cases} -n - \log\det(-Y) & \text{if } 0 \succ Y, \\ +\infty & \text{otherwise.} \end{cases} \tag{10.5}$$

In this example we see the intimate connection between the functions (10.2) and (10.3), two of the 'self-concordant barriers' fundamental to the development of Nesterov and Nemirovskii (1994). This connection suggests the following interesting question (Tunçel 1995): if the function h is a self-concordant barrier, is the same true of the matrix function $h \circ \lambda$?

Example 10.6. (BFGS updates – Fletcher, 1991) Given a matrix H in S_n which is positive definite, and vectors s and b in $\mathbb{R}^n$, we consider the primal problem

$$\inf_{X \in S_n} \{\operatorname{tr} H^{-1}X - \log\det X : Xs = b,\ X \succ 0\}. \tag{10.6}$$

Using the framework of the Fenchel Duality Theorem (Theorem 8.4), and formula (10.5), the dual problem is

$$\sup_{y\in\mathbb{R}^m} \{b^T y + \log\det(H^{-1} - (ys^T + sy^T)/2)\} + n. \tag{10.7}$$

If $s^T b > 0$, standard quasi-Newton theory shows the primal problem (10.6) is superconsistent, and choosing $\hat{y} = 0$ shows the dual problem (10.7) is also superconsistent. Thus the primal and dual problems are both attained, by the Fenchel Duality Theorem, and routine calculation using the gradient formula (10.4) shows that the unique primal optimal solution is the 'BFGS update' of the 'Hessian approximation' H, subject to the 'secant equation' $Xs = b$ (Fletcher 1991, Lewis 1996*a*).

Example 10.7. (eigenvalue sums) For an integer k between 0 and n, define a symmetric closed convex function h on $\mathbb{R}^n$ by

$$h(x) = \text{sum of the } k \text{ largest } x_j. \tag{10.8}$$

The corresponding matrix function is the sum of the k largest eigenvalues,

$$(h\circ\lambda)(X) = \sum_{j=1}^{k} \lambda_j(X).$$

A calculation shows the conjugate of h is the indicator function of the set

$$\left\{ z\in\mathbb{R}^n : \sum_{j=1}^{n} z_j = k,\ 0\le z_j\le 1 \text{ for each } j \right\},$$

so by the conjugacy formula (10.1), the conjugate of $h\circ\lambda$ is the indicator function of the matrix set

$$H = \{Y\in S_n : \operatorname{tr} Y = k,\ I\succeq Y\succeq 0\}. \tag{10.9}$$

For given matrices $A^1, A^2, \ldots, A^m$ in S_n and a vector b in $\mathbb{R}^m$, consider the optimization problem

$$\inf_{X\in S_n} \left\{ \sum_{j=1}^{k} \lambda_j(X) : \operatorname{tr} A^i X = b_i \text{ for each } i \right\};$$

cf. Fletcher (1985), Overton and Womersley (1993), Hirriart-Urruty and Ye (1995) and Pataki (1995). In the Fenchel Duality framework the dual problem is therefore

$$\sup_{y\in\mathbb{R}^m} \left\{ b^T y : \sum_{i=1}^{m} y_i A^i \in H \right\}, \tag{10.10}$$

where the set H is given by equation (10.9).

Rather more generally, suppose the vector d in $\mathbb{R}^n$ has nonincreasing components. For any vector x in $\mathbb{R}^n$, let $\overline{x}$ denote the vector with components x_j rearranged into nonincreasing order. Then the function $h(x) = d^T\overline{x}$ is symmetric, closed and convex (since $h(x) = \max_{Q \in P_n}\{d^T Qx\}$). The corresponding matrix function is exactly the weighted sum of eigenvalues appearing in the graph partitioning problem in Section 2, namely $(h \circ \lambda)(W) = d^T\lambda(W)$.

11. Linear programming

An important area of eigenvalue optimization is semidefinite programming (SDP). Since the analogies with ordinary linear programming (LP) are very close, we begin by outlining the relevant classical theory.

For given vectors $c, a^1, a^2, \dots, a^m$ in $\mathbb{R}^n$, and b in $\mathbb{R}^m$, the primal linear program we study is

$$\rho_0 = \inf_{x \in \mathbb{R}^n_+}\{c^T x : (a^j)^T x = b_j \text{ for each } j\}. \tag{11.1}$$

Using the framework of Theorem 8.4 (with objective function $f(x) = c^T x + \delta_{\mathbb{R}^n_+}(x)$), we obtain the dual problem

$$\delta_0 = \sup_{y \in \mathbb{R}^m}\left\{b^T y : c \geq \sum_i y_i a^i\right\}. \tag{11.2}$$

By polyhedrality, we immediately see from the Fenchel Duality Theorem (Theorem 8.4) that if either the primal or dual problem is consistent, then $\rho_0 = \delta_0$, and both values are attained if finite. This is the classical linear programming duality theorem. The *complementary slackness* condition ((v) in Theorem 8.4) states that primal feasible $\overline{x}$ in $\mathbb{R}^n$ and dual feasible $\overline{y}$ in $\mathbb{R}^m$ are both optimal if and only if

$$\left(c - \sum_i \overline{y}_i a^i\right)^T \overline{x} = 0,$$

or, equivalently, $\overline{x}_j(c - \sum_i \overline{y}_i a^i)_j = 0$, for each index $j = 1, 2, \dots, n$.

If we penalize the primal constraint $x \in \mathbb{R}^n_+$ using the logarithmic barrier (10.2) with a small positive parameter μ, we obtain the new primal problem

$$\rho_\mu = \inf_{x \in \mathbb{R}^n}\left\{c^T x - \mu \sum_j \log x_j : (a^i)^T x = b_i \text{ for each } i\right\},$$

and the dual problem is

$$\delta_\mu = \sup_{y \in \mathbb{R}^m}\left\{b^T y + \mu \sum_j \log\left(c_j - \sum_i y_i a^i_j\right)\right\} + n\mu(\log \mu - 1).$$

The Fenchel Duality Theorem now needs a regularity condition. We assume the following:

(i) *Primal superconsistency:* some vector $\hat{x} > 0$ in $\mathbb{R}^n$ satisfies

$$(a^i)^T \hat{x} = b_i, \text{ for each } i.$$

(ii) *Dual superconsistency:* some vector $\hat{y}$ in $\mathbb{R}^m$ satisfies $c > \sum_i \hat{y}_i a^i$.
(iii) *Independence:* the vectors $a^1, a^2, \ldots, a^m$ are linearly independent.

Assumptions (i) and (ii) guarantee $\rho_\mu = \delta_\mu$, by the Duality Theorem, and both values are attained. The primal objective is (essentially) strictly convex; assumption (iii) ensures the dual objective is too. Hence the primal and dual both have unique optimal solutions, $x = x^\mu$ in $\mathbb{R}^n$ and $y = y^\mu$ in $\mathbb{R}^m$ respectively, and by the complementary slackness condition, they are the unique solution of the system

$$(a^i)^T x = b_i, \text{ for each } i, \tag{11.3}$$

$$x_j \left(c_j - \sum_i y_i a^i_j \right) = \mu, \text{ for each } j, \tag{11.4}$$

$$x \geq 0 \text{ and } c \geq \sum_i y_i a^i. \tag{11.5}$$

Notice that when $\mu = 0$ these conditions reduce to the complementary slackness conditions for the original linear program.

The trajectory $\{(x^\mu, y^\mu) : \mu > 0\}$ is called the *central path.* From equations (11.3) and (11.4), we deduce the *duality gap*

$$c^T x^\mu - b^T y^\mu = n\mu. \tag{11.6}$$

Thus, using the *weak duality* inequality (Theorem 8.4(i)), we see that the feasible solutions x^μ and y^μ approach optimality:

$$\lim_{\mu \downarrow 0} c^T x^\mu = \rho_0 = \delta_0 = \lim_{\mu \downarrow 0} b^T y^\mu.$$

But our regularity assumptions (i), (ii) and (iii) then imply, using the Fenchel Duality compactness results, that the central path (x^μ, y^μ) stays bounded for small positive μ. Any limit point (x^0, y^0) must satisfy $c^T x^0 = b^T y^0$ (by the duality gap formula (11.6)), whence x^0 and y^0 are optimal for the primal and dual respectively. In fact, a more careful argument shows that the limit point (x^0, y^0) is unique: the vectors x^0 and y^0 are the 'analytic centres' of the optimal faces for the primal and dual problems respectively; see Megiddo (1989) and McLinden (1980). (Given a polytope $P = \{z \in L : z \geq 0\}$, where L is an affine subspace and P contains a point $z > 0$, the analytic centre of P is the unique minimizer of the logarithmic barrier $-\sum_j \log(z_j)$ over all $z \in P$.)

12. Semidefinite programming

In the previous section we outlined some of the classical theory of linear programming, from a Fenchel Duality perspective. In this section we describe the parallel theory of semidefinite programming: with a little caution, the development is largely identical.

For given matrices $C, A^1, A^2, \ldots, A^m$ in S_n and a vector b in $\mathbb{R}^m$, the primal semidefinite program is

$$\rho_0 = \inf_{X \in S_n^+} \{\operatorname{tr} CX : \operatorname{tr} A^i X = b_i \text{ for each } i\}. \tag{12.1}$$

Calculating the conjugate of the objective function $f(X) = \operatorname{tr} CX + \delta_{S_n^+}(X)$ is easy, using Example 10.4, and we arrive at the dual problem

$$\delta_0 = \sup_{y \in \mathbb{R}^m} \left\{ b^T y : C \succeq \sum_i y_i A^i \right\}. \tag{12.2}$$

Despite its simple form, the primal-dual pair of semidefinite programs is a remarkably flexible model. For example, it is easy to see how to rewrite the dual of the eigenvalue sum problem (see (10.10)) as a dual semidefinite program. Many other examples appear in Nesterov and Nemirovskii (1994).

The primal and dual problems are not polyhedral. As we have seen, a linear program with finite optimal value must have an optimal solution, and its optimal value must equal the optimal value of the dual linear program. By contrast, these properties may fail for semidefinite programs. Hence we assume some regularity conditions at the outset:

(i) *Primal superconsistency:* some matrix $\hat{X} \succ 0$ in S_n satisfies

$$\operatorname{tr} A^i \hat{X} = b_i, \text{ for each } i.$$

(ii) *Dual superconsistency:* some vector $\hat{y}$ in $\mathbb{R}^m$ satisfies $C \succ \sum_i \hat{y}_i A^i$.

(iii) *Independence:* the matrices $A^1, A^2, \ldots, A^m$ are linearly independent.

With these assumptions, we see from the Fenchel Duality Theorem (Theorem 8.4) that the optimal values are equal, $\rho_0 = \delta_0$, and both are attained. The complementary slackness condition states that primal feasible $\overline{X}$ in S_n and dual feasible $\overline{y}$ in $\mathbb{R}^m$ are both optimal if and only if

$$\operatorname{tr}\left(C - \sum_i \overline{y}_i A^i\right)\overline{X} = 0.$$

As observed in Example 10.4 this condition is equivalent to $(C - \sum_i \overline{y}_i A^i)\overline{X} = 0$, and therefore it implies that $\overline{X}$ and $\overline{Z}$ have a simultaneous ordered spectral decomposition, where $\overline{Z} = C - \sum_i \overline{y}_i A^i$. In other words, there is an orthogonal matrix $\overline{U}$ such that

$$\begin{array}{rcll} \overline{X} & = & \overline{U}^T(\mathrm{Diag}\,\lambda(\overline{X}))\overline{U}, & \text{and} \\ -\overline{Z} & = & \overline{U}^T(\mathrm{Diag}\,\lambda(-\overline{Z}))\overline{U}. & \end{array} \tag{12.3}$$

Thus the complementary slackness condition implies

$$\lambda_j(\overline{X})\lambda_j(-\overline{Z}) = 0, \text{ for each } j = 1, 2, \ldots, n. \tag{12.4}$$

The minus signs are present because of the convention that λ maps a matrix to its eigenvalues in descending order. Note that, except in special cases, we expect both $\overline{X}$ and $\overline{Z}$ to have a multiple zero eigenvalue.

As with linear programming, we next consider a penalized version of the primal problem, using the logarithmic barrier (10.3) with a small positive parameter μ. We obtain the new primal problem

$$\rho_\mu = \inf_{X \in S_n} \left\{ \mathrm{tr}\, CX - \mu \log\det X : \mathrm{tr}\, A^i X = b_i \text{ for each } i \right\}.$$

The corresponding dual problem is

$$\delta_\mu = \sup_{y \in \mathbb{R}^m} \left\{ b^T y + \mu \log\det\left(C - \sum_i y_i A^i \right) \right\} + n\mu(\log\mu - 1).$$

By the Fenchel Duality Theorem (Theorem 8.4), $\rho_\mu = \delta_\mu$, and both values are attained. Both objective functions are (essentially) strictly convex, so the primal and dual problems have unique optimal solutions $X = X^\mu$ in S_n and $y = y^\mu$ in $\mathbb{R}^m$ respectively: by the complementary slackness condition and the gradient characterization (10.4), they uniquely solve the system

$$\mathrm{tr}\, A^i X = b_i, \text{ for each } i, \tag{12.5}$$

$$X\left(C - \sum_i y_i A^i \right) = \mu I, \tag{12.6}$$

$$X \succeq 0 \text{ and } C \succeq \sum_i y_i A^i. \tag{12.7}$$

The trajectory $\{(X^\mu, y^\mu) : \mu > 0\}$ is called the central path. Points on the central path have the *duality gap*

$$\mathrm{tr}\, CX^\mu - b^T y^\mu = n\mu.$$

As with linear programming, this guarantees that as μ decreases to zero the solutions X^μ and y^μ approach optimality. Once again, the Fenchel Duality Theorem shows that the central path (X^μ, y^μ) stays bounded for small positive μ, and any limit point (X^0, y^0) must be a pair of optimal solutions for the original primal and dual problems.

Condition (12.6) implies that, just like the solution pair $\overline{X}$ and $\overline{Z}$, X^μ and $C - \sum_i y_i^\mu A^i$ have a simultaneous ordered spectral decomposition. When $\mu = 0$, equation (12.6) reduces to the complementary slackness condition for the original semidefinite program.

Notice that, with the choices $C = \operatorname{Diag} c$ and $A^i = \operatorname{Diag} a^i$ for each i, the semidefinite theory developed in this section collapses to the linear theory of the previous section.

13. Strict complementarity and nondegeneracy

Let us go back to the linear programming problem and its dual, (11.1) and (11.2). We say a primal-dual solution $(\overline{x}, \overline{y})$ satisfies the *strict complementarity condition* if, for each j, exactly one of the two statements $\overline{x}_j = 0$ and $(c - \sum \overline{y}_i a^i)_j = 0$ holds. We say a strictly complementary solution is *nondegenerate* if the vector $\overline{x}$ has exactly m nonzero components and the corresponding m rows of the matrix $[a^1, a^2, \ldots, a^m]$ are linearly independent. It is well known that these conditions guarantee that $\overline{x}$ is the unique optimal solution of the primal problem (11.1) and that $\overline{y}$ is the unique optimal solution of the dual problem (11.2). Furthermore, these conditions hold 'generically' for a linear program: roughly speaking, this means that they hold with probability one, given randomly generated linear programs with associated nonempty feasible regions.

The situation is less clear in semidefinite programming. There is no difficulty with the idea of strict complementarity: we say a primal-dual solution $(\overline{X}, \overline{y})$ for (12.1) and (12.2) satisfies the *strict complementarity condition* if, for each index j, exactly one of the two statements $\lambda_j(\overline{X}) = 0$ and $\lambda_j(-\overline{Z}) = 0$ holds, where $\overline{Z} = C - \sum_i \overline{y}_i A^i$. Let r denote the rank of $\overline{X}$ and let s denote the rank of $\overline{Z}$; then strict complementarity holds if and only if $r + s = n$. Nondegeneracy conditions are more complicated and are discussed by Alizadeh, Haeberly and Overton (1996*a*) and Shapiro (1996). Assume that strict complementarity holds and let $\overline{U} = [\overline{U}_1\ \overline{U}_2]$ be the orthogonal matrix of eigenvectors which simultaneously diagonalizes $\overline{X}$ and $\overline{Z}$ (see equations (12.3)), with the first r columns (collected in $\overline{U}_1$) corresponding to nonzero eigenvalues of $\overline{X}$ and the last s columns (collected in $\overline{U}_2$) corresponding to nonzero eigenvalues of $\overline{Z}$. Then the appropriate nondegeneracy assumptions are the following two conditions, motivated by studying the primal and the dual separately: first, that the matrices

$$\begin{bmatrix} \overline{U}_1^T A^i \overline{U}_1 & \overline{U}_1^T A^i \overline{U}_2 \\ \overline{U}_2^T A^i \overline{U}_1 & 0 \end{bmatrix}, \quad \text{for } i = 1, 2, \ldots, m,$$

are linearly independent in the space S_n and second, that the matrices

$$\overline{U}_1^T A^i \overline{U}_1, \quad \text{for } i = 1, 2, \ldots, m,$$

Table 1. *Number of occurrences of rank(X) in 1000 randomly generated problems with $n = 10$ and various values of m.*

m	0	1	2	3	4	5	6	7	8	9	10
5		297	703								
10		0	494	506	0						
15			18	712	270	0					
20				100	813	87					
25				1	325	667	7				

span the space S_r. It is shown by Alizadeh et al. (1996*a*) that the strict complementarity and nondegeneracy conditions imply uniqueness of the primal and dual solutions, and also that the conditions are indeed generic properties of SDP, meaning roughly that they hold with probability one for an optimal solution pair, given random data with feasible solutions. An immediate consequence is the existence of generic bounds on the optimal solution matrix ranks r and s, and therefore on the multiplicity of the zero eigenvalues. Let $k^{\overline{2}}$ denote $k(k+1)/2$, and let $\sqrt[\overline{2}]{k} = \lfloor t \rfloor$, where t is the positive real root of $t^{\overline{2}} = k$. Then generic bounds on the rank of the primal optimal solution matrix $\overline{X}$ are given by

$$n - \sqrt[\overline{2}]{n^{\overline{2}} - m} \quad \leq \quad r \quad \leq \quad \sqrt[\overline{2}]{m}.$$

For further discussion of related issues, see Pataki (1995).

Experiments reported in Alizadeh et al. (1996*a*) show clearly that, given randomly generated data, the rank r is far more likely to lie in the centre of its range than near the end points. This is demonstrated by Table 1, which shows, for $n = 10$ and various choices of m, how many times the primal rank r occurred during 1000 runs with different random data. The zeros indicate possible values of r which did not occur, while the blanks indicate generically impossible values.

A natural question is: what is the underlying probability distribution for the primal solution rank r? We consider this to be a very interesting open question.

Table 1 also shows, incidentally, the reliability of the numerical method used to obtain the results: accurate solutions to 5000 different randomly generated problems were obtained without a single failure. (As with linear programming, it is easy to check the optimality of a solution pair, simply by checking primal and dual feasibility and the complementary slackness condition.) We now sketch the ideas behind the primal-dual interior-point method used to obtain these results.

14. Primal-dual interior-point methods

We begin again with the case of linear programming. The basic idea of the primal-dual interior-point method is to generate a sequence of iterates $(x^{(k)}, y^{(k)}) \in \mathbb{R}^n \times \mathbb{R}^m$ (for $k = 1, 2, \ldots$) approximating a sequence of points lying on the central path and converging to an optimal solution as $k \to \infty$. Briefly, this approximation is achieved by applying, at the kth iteration, one step of Newton's method to (11.3) and (11.4), a system of $n + m$ linear and quadratic equations in the $n+m$ variables x_j, y_i. Here μ is a positive number, fixed at the kth iteration to a value $\mu^{(k)}$, with $\mu^{(k)} \to 0$ as $k \to \infty$. (If we also introduce the equations $z_j = c_j - \sum_i y_i a^i_j$, substituting these into (11.4) to obtain $x_j z_j = \mu$, and treating z_j, $j = 1, \ldots, n$, as independent variables, Newton's method yields an equivalent iteration.) The Newton step is defined by the linear system

$$\begin{bmatrix} A^T & 0 \\ Z^{(k)} & -X^{(k)}A \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} b - A^T x^{(k)} \\ (\mu^{(k)} I - X^{(k)} Z^{(k)})e \end{bmatrix}, \tag{14.1}$$

where A is the $n \times m$ matrix $[a^1, \ldots, a^m]$, $X^{(k)}$ and $Z^{(k)}$ are respectively the diagonal matrices $\operatorname{Diag} x^{(k)}$ and $\operatorname{Diag}(c - Ay^{(k)})$, I is the $n \times n$ identity matrix and e is the n-vector whose components are all one. Block Gauss elimination reduces this system to

$$\left(A^T (Z^{(k)})^{-1} X^{(k)} A\right) \Delta y = b - A^T (x^{(k)} - w^{(k)}) \tag{14.2}$$

$$\Delta x = w^{(k)} + (Z^{(k)})^{-1} X^{(k)} A \Delta y, \tag{14.3}$$

where $w^{(k)} = (\mu^{(k)} (Z^{(k)})^{-1} - X^{(k)})e$. New iterates are then obtained by $x^{(k+1)} = x^{(k)} + \alpha \Delta x$, $y^{(k+1)} = y^{(k)} + \beta \Delta y$, where steplengths α and β are chosen so that $x^{(k+1)} > 0$ and $c - Ay^{(k+1)} > 0$. A value $\mu^{(k+1)} < \mu^{(k)}$ is then chosen and the iterative step repeated. Different rules for reducing the parameter μ and choosing the steplengths α and β give different variants of the algorithm, with some specific rules known to guarantee a solution with prescribed accuracy in polynomial time. The original references for this method are Monteiro and Adler (1989) and Kojima et al. (1989).

We shall not discuss the global convergence theory. However, the following well known result is important for understanding the local convergence and numerical stability of the algorithm. It analyses the condition numbers of the two key matrices defining the algorithm at points on the central path.

Proposition 14.1 Suppose that $(\overline{x}, \overline{y})$ solves the LP primal-dual pair (11.1) and (11.2), with both the strict complementarity and nondegeneracy conditions holding. Then, using $X^\mu = \operatorname{Diag} x^\mu$ and $Z^\mu = \operatorname{Diag}(c - Ay^\mu)$, where (x^μ, y^μ) lies on the central path defined by (11.3), (11.4) and (11.5), the condition numbers of the matrices

$$\begin{bmatrix} A^T & 0 \\ Z^\mu & -X^\mu A \end{bmatrix} \quad \text{and} \quad A^T(Z^\mu)^{-1}X^\mu A \tag{14.4}$$

are both bounded independent of μ as $\mu \downarrow 0$.

Proof. It is well known that the assumptions guarantee that $(\overline{x}, \overline{y})$ is the unique solution of the linear program and consequently, as discussed in Section 11, also the limit point of the central path (x^μ, y^μ) as $\mu \downarrow 0$. Without loss of generality, we may take

$$\overline{x} = \begin{bmatrix} \tilde{x} \\ 0 \end{bmatrix}, \qquad c - A\overline{y} = \begin{bmatrix} 0 \\ \tilde{z} \end{bmatrix}, \qquad A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

where all partitionings are from n rows into m and $n-m$ rows respectively, with $\tilde{x} > 0$, $\tilde{z} > 0$, and A_1 nonsingular. Then the first matrix in (14.4) converges to

$$\begin{bmatrix} A_1^T & A_2^T & 0 \\ 0 & 0 & (-\operatorname{Diag}\tilde{x})A_1 \\ 0 & \operatorname{Diag}\tilde{z} & 0 \end{bmatrix}$$

as $\mu \downarrow 0$. This matrix can be permuted into a block upper triangular matrix with nonsingular diagonal blocks A_1^T, $-\operatorname{Diag}(\tilde{x})A_1$, $\operatorname{Diag}\tilde{z}$. The second matrix does not have a limit, but $\mu A^T(Z^\mu)^{-1}X^\mu A$ has the limit

$$A_1^T(\operatorname{Diag}\tilde{x})^2 A_1$$

by virtue of (11.4), which completes the proof. □

Let us refer to the first matrix in (14.4) as the *block Jacobian* matrix and to the second as the *Schur complement*. The consequence of the bounded condition number of the block Jacobian is that, given strict complementarity and nondegeneracy assumptions, the primal-dual interior-point method for linear programming has a quadratic rate of local convergence as long as the parameter μ is reduced sufficiently fast and the steplengths α and β are chosen sufficiently close to one. See Zhang, Tapia and Dennis (1992) for details. (Even without nondegeneracy assumptions, certain superlinear convergence properties still hold; see Zhang and Tapia (1993).) The consequence of the bounded condition number of the Schur complement is that, again under the given assumptions, there is no numerical difficulty with factorizing the matrix in (14.2) as $\mu \downarrow 0$. This is not necessarily the case in the absence of nondegeneracy assumptions, a fact that is a subject of some current interest (Wright 1995).

Now let us turn to semidefinite programming. As in linear programming, the essential idea of the primal-dual interior-point method is to generate a sequence of iterates $(X^{(k)}, y^{(k)}) \in S_n \times \mathbb{R}^m$ approximating a sequence of points on the central path, converging to a solution as $k \to \infty$. However, it is not clear in this case how to apply Newton's method. The key difficulty is that the left-hand side of (12.6) is *not* symmetric, so equations (12.5) and (12.6) do not map $S_n \times \mathbb{R}^m$ to itself; consequently, Newton's method is not directly applicable. The cleanest solution seems to be to replace (12.6) by

$$X\left(C - \sum_i y_i A^i\right) + \left(C - \sum_i y_i A^i\right) X = 2\mu I. \tag{14.5}$$

That (12.6) implies (14.5) is immediate. That the converse holds for $X \succeq 0$ is easily seen by premultiplying (14.5) by U^T and postmultiplying by U, where the orthogonal matrix U diagonalizes X. Application of Newton's method to (12.5) and (14.5) leads to a very effective method for semidefinite programming called the $XZ{+}ZX$ method by Alizadeh, Haeberly and Overton (1996*b*); this method was used to generate the results shown in Table 1 in the previous section. On average, each problem was solved in less than 10 iterations, a property which is, in practice, almost independent of the problem dimension. Other variants of the primal-dual interior-point method given by Helmberg, Rendl, Vanderbei and Wolkowicz (1996), Kojima, Shindoh and Hara (1994), Nesterov and Todd (1996) and Vandenberghe and Boyd (1995) give similar performance, but the $XZ+ZX$ method is especially robust with respect to changes in the rules for reducing μ and choosing the steplengths α, β (Alizadeh et al. 1996*b*). It is proved by Alizadeh et al. (1996*b*) that, given the SDP strict complementarity and nondegeneracy assumptions stated in the previous section, the *first* part of Theorem 14.1, namely that the condition number of the block Jacobian is bounded, extends from LP to hold also for the $XZ{+}ZX$ method for SDP, but the *second does not*, that is, the condition number of the corresponding Schur complement matrix is unbounded for SDP, even with nondegeneracy assumptions. Consequently, the $XZ + ZX$ method *is* locally quadratically convergent, in contrast to other variants of the primal-dual interior-point method for SDP, given the nondegeneracy assumptions and appropriate μ reduction and steplength rules. However, under the same conditions, the method *is not* necessarily numerically stable as $\mu \downarrow 0$, since the condition number of the linear system which must be solved at each iteration is $O(1/\mu)$. Indeed, this was observed numerically by Alizadeh et al. (1996*b*): generally, it was possible to compute results accurate to only about the square root of the machine precision, given random data. The same difficulty applies to other variants of the primal-dual interior-point method as well as the $XZ+ZX$ method. By contrast, there is no difficulty solving modest-sized randomly generated linear programs to machine precision accuracy using the LP primal-dual interior-point method.

As in LP, if we use the substitution $Z = C - \sum_i y_i A^i$ in (12.6), introducing Z as an independent variable and $Z = C - \sum_i y_i A^i$ as an additional equation, Newton's method yields an equivalent iteration.

An appealing alternative primal-dual interior-point iteration for SDP is based on the following idea. Instead of treating the variable X (or X and Z) directly, recall from the discussion in Section 12 that for (X^μ, y^μ) to lie on the central path, X^μ and $Z = C - \sum_i y_i^\mu A^i$ must have a simultaneous ordered spectral decomposition. Therefore, consider the following set of variables: an orthogonal matrix U, which diagonalizes both X and $Z = C - \sum_i y_i A^i$, together with the eigenvalues of X and Z, say ξ_j and ζ_j, $j = 1, \ldots, n$. Equations (12.5) and (12.6) then reduce to

$$\operatorname{tr}\left(U^T A^i U (\operatorname{Diag} \xi_i)\right) = b_i, \text{ for each } i,$$

$$(\operatorname{Diag} \zeta_j) + \sum_i y_i U^T A^i U = U^T C U,$$

and

$$\xi_j \zeta_j = \mu, \text{ for each } j.$$

Borrowing a technique used by Friedland, Nocedal and Overton (1987), Overton (1988) and Overton and Womersley (1995), the orthogonal matrix U can be parametrized by $U = \exp(S) = I + S + \frac{1}{2}S^2 + \cdots$, where S is skew-symmetric, making the application of Newton's method straightforward. This leads to a method which, though it apparently has poor global convergence properties, is at present able to compute more accurate solutions than any other SDP interior-point method (Alizadeh et al. 1996*b*).

The eigenvalue optimization method of Overton (Overton 1988, Overton and Womersley 1995) is easily extended to apply to SDP. This method does not share the global convergence properties known for the interior-point methods. However, it can be used as an effective technique to obtain highly accurate solutions when an interior-point method reaches its limiting accuracy. The same presumably applies to Fletcher's method (Fletcher 1985), though this has not been tested. These methods are more difficult to describe because they use second derivatives, a complicated issue in the presence of multiple eigenvalues. They need second derivatives to achieve quadratic convergence because they are based on an appropriate form of Newton's method in the dual space only. These Newton methods use primal information to construct the second derivative of an appropriate Lagrangian function, but they are not primal-dual methods. A really remarkable property of the $XZ + ZX$ primal-dual interior-point method for SDP is that, in exact arithmetic, it generically achieves quadratic convergence with only first-order primal and dual information, even though the constraints are not polyhedral.

Primal-dual interior-point methods for LP have been used to solve very large problems; the best methods are generally thought to be superior to the simplex method, except for special problem classes. However, at present the implementation of interior-point methods for SDP has been limited to small problems, or problems with block-diagonal structure. If C and the A^i are block-diagonal with the same block structure, then, without loss of generality, the primal matrix X can be taken to have the same block structure: indeed, LP is a special case with block sizes all one. Consequently, the primal-dual interior-point methods for SDP can be implemented very efficiently if the block sizes are not large. However, if C and the matrices A^i have a more general sparse structure, then even if $C-\sum_i y_i^\mu A^i$ is sparse, the corresponding primal matrix $X^\mu = \mu(C - \sum_i y_i^\mu A^i)^{-1}$ is generally dense. For example, this is the case when C and the A^i are tridiagonal. In this situation, it is possible that an interior-point method based only on dual information is preferable to a primal-dual method. It may also be worth reconsidering some older and simpler first-order methods (Cullum et al. 1975, Overton 1992, Schramm and Zowe 1992).

15. Nonlinear semidefinite programming and eigenvalue optimization

The primal semidefinite program (12.1) permits only linear constraints; likewise the constraint in the dual program (12.2) is a semidefinite constraint on an affine matrix function $C - \sum_i y_i A^i$. In many applications, one finds eigenvalue optimization problems with nonlinear constraints, or with matrix functions depending nonlinearly on the variables. Such problems are, of course, substantially more difficult and a detailed discussion is beyond the scope of this article. However, we make two remarks.

First, although much of the duality theory described above fails to extend to the nonlinear case, some results are possible. Instead of subdifferentials, one may introduce the Clarke generalized gradient (Clarke 1983). A suitable chain rule yields first-order optimality conditions, though these are generally only necessary, not sufficient, conditions for optimality (Cox and Overton 1992, Lewis 1996*b*, Overton 1992). Second-order optimality conditions may also be derived (Shapiro 1996).

Second, some of the essential ideas of interior-point methods can be extended to nonlinear, nonconvex problems. Specifically, the logarithmic barrier function remains a very useful tool (Ringertz 1995). Whether primal-dual methods have an important role to play in the nonlinear case is not clear. However, the main idea remains valid, namely the application of Newton's method to a perturbed form of the optimality conditions, which, as in the linear case, involve a complementarity condition.

16. Eigenvalues of nonsymmetric matrices

The eigenvalues of a real symmetric matrix, which we described by the function $\lambda : S_n \to \mathbb{R}^n$, are Lipschitz functions of the matrix elements. Our development in Section 10 and our analysis of semidefinite programming depend heavily on the symmetry of the matrices. A completely parallel theory holds for complex Hermitian matrices. However, the eigenvalues of a real nonsymmetric or a general complex matrix are, in general, non-Lipschitz functions of the matrix elements.

In this section we give some apparently new variational results for functions of eigenvalues of nonsymmetric matrices. One special case characterizes the spectral abscissa of a nonsymmetric matrix, in a quantitative version of Lyapunov theory, while another special case yields a well known characterization of the spectral radius.

We can order the complex numbers $\mathbb{C}$ *lexicographically*: in this order, one complex number, z, dominates another, w if either $\operatorname{Re} z > \operatorname{Re} w$, or $\operatorname{Re} z = \operatorname{Re} w$ and $\operatorname{Im} z > \operatorname{Im} w$. For a matrix X in the vector space of $n \times n$ complex matrices, $M_n(\mathbb{C})$, let us denote the eigenvalues of X by $\lambda_1(X), \lambda_2(X), \ldots, \lambda_n(X)$, counted by multiplicity and ordered lexicographically. In this way we can extend the eigenvalue function λ to the space $M_n(\mathbb{C})$. If the matrices X and Z in $M_n(\mathbb{C})$ are *similar* (that is, some matrix L satisfies $Z = LXL^{-1}$), then we write $X \sim Z$.

Proposition 16.1 If the function $F : M_n(\mathbb{C}) \to [-\infty, +\infty]$ satisfies

$$F(X) \geq F(\operatorname{Diag} \lambda(X)) \text{ for all } X \text{ in } M_n(\mathbb{C}), \tag{16.1}$$

and if, for some matrix Y in $M_n(\mathbb{C})$, the function F is upper semicontinuous at $\operatorname{Diag} \lambda(Y)$, then

$$F(\operatorname{Diag} \lambda(Y)) = \inf_{Z \sim Y} F(Z). \tag{16.2}$$

Proof. If Z is similar to Y, then $\lambda(Z) = \lambda(Y)$, whence by inequality (16.1) we obtain $F(Z) \geq F(\operatorname{Diag} \lambda(Z)) = F(\operatorname{Diag} \lambda(Y))$. Thus $F(\operatorname{Diag} \lambda(Y)) \leq \inf_{Z \sim Y} F(Z)$.

On the other hand, by Schur's Theorem (Horn and Johnson 1985, Theorem 2.3.1), there is a unitary matrix Q and an upper triangular matrix T with main diagonal $\lambda(Y)$, satisfying $QYQ^* = T$. For positive real t, let D_t denote the matrix $\operatorname{Diag}(t, t^2, \ldots, t^n)$. As t approaches $+\infty$, we have

$$(D_t Q) Y (D_t Q)^{-1} = D_t T D_t^{-1} \to \operatorname{Diag} \lambda(Y),$$

and since F is upper semicontinuous at $\operatorname{Diag} \lambda(Y)$, we deduce

$$\inf_{Z \sim Y} F(Z) \leq \limsup_{t \to +\infty} F((D_t Q) Y (D_t Q)^{-1}) \leq F(\operatorname{Diag} \lambda(Y)).$$

Equation (16.2) follows. □

The key technique in this proof, using diagonal similarity transformations to reduce the strictly upper triangular part of the Schur triangular form, is well known: see, for example, Horn and Johnson (1985, Lemma 5.6.10). Notice that if Y is not diagonalizable, the infimum in (16.2) may not be attained.

The following two propositions begin to look reminiscent of the material in Sections 9 and 10. Indeed, the complex versions of von Neumann's Lemma (Lemma 9.2) and the von Neumann–Theobald Lemma (Lemma 10.2) may be used to prove the propositions (although we quote intermediate results). For a vector z in $\mathbb{C}^n$, we write $\operatorname{Re} z$ and $|z|$ for the vectors with entries $\operatorname{Re} z_j$ and $|z_j|$ respectively.

Proposition 16.2 If the function $h : \mathbb{R}^n \to [-\infty, +\infty]$ is symmetric and convex, then any matrix X in $M_n(\mathbb{C})$ satisfies the inequality

$$h\left(\tfrac{1}{2}\lambda(X + X^*)\right) \geq h(\operatorname{Re}(\lambda(X))). \tag{16.3}$$

Proof. Since h is 'Schur convex' (see Marshall and Olkin (1979)), it suffices to show that the inequalities

$$\tfrac{1}{2}\sum_{j=1}^{k} \lambda_j(X + X^*) \geq \sum_{j=1}^{k} \operatorname{Re} \lambda_j(X)$$

hold for each index $k = 1, 2, \ldots, n$, with equality for $k = n$. This is exactly Horn and Johnson (1991, (3.3.33)). □

Proposition 16.3 If g is a symmetric gauge on $\mathbb{R}^n$, then any matrix X in $M_n(\mathbb{C})$ satisfies the inequality

$$g(\sigma(X)) \geq g(|\lambda(X)|). \tag{16.4}$$

Proof. By Horn and Johnson (1985, Theorem 7.4.45), it suffices to show that the inequalities

$$\sum_{j=1}^{k} \sigma_j(X) \geq \sum_{j=1}^{k} |\lambda_{\pi(j)}(X)|$$

hold for each index $k = 1, 2, \ldots, n$, where π is any permutation for which $|\lambda_{\pi(j)}(X)|$ is nonincreasing in j. But this is precisely a result of Weyl (Horn and Johnson 1991, Theorem 3.3.13). □

The analogy between the previous two propositions is clear if we recall that the components of $\sigma(X)$ are just $(\lambda_i(X^*X))^{1/2}$.

Putting together the three previous propositions, we arrive at the main result of this section.

Theorem 16.4

(a) Suppose the function $h : \mathbb{R}^n \to [-\infty, +\infty]$ is convex and symmetric. If the matrix Y in $M_n(\mathbb{C})$ has $\operatorname{Re} \lambda(Y)$ in $\operatorname{int}(\operatorname{dom} h)$, then it satisfies

$$h(\operatorname{Re} \lambda(Y)) = \inf_{Z \sim Y} h\left(\tfrac{1}{2}\lambda(Z + Z^*)\right).$$

(b) Suppose g is a symmetric gauge on $\mathbb{R}^n$. Then any matrix Y in $M_n(\mathbb{C})$ satisfies

$$g(|\lambda(Y)|) = \inf_{Z \sim Y} g(\sigma(Y)).$$

Proof.

(a) We choose $F(X) = h(\frac{1}{2}\lambda(X+X^*))$ in Proposition 16.1. Inequality (16.1) follows from Proposition 16.2, and since the convex function h must be continuous on the interior of its domain (Rockafellar 1970, Theorem 10.1) and λ is continuous, it follows that F is continuous at $\operatorname{Diag}\lambda(Y)$.
(b) We choose $F(X) = g(\sigma(X))$ in Proposition 16.1.

□

Example 16.5. (spectral abscissa) The *spectral abscissa* of a matrix Y in $M_n(\mathbb{C})$ is $\operatorname{Re}\lambda_1(Y)$. Applying Theorem 16.4(a) with the function h defined by $h(x) = \max_j x_j$, we obtain, for any matrix Y in $M_n(\mathbb{C})$,

$$\text{spectral abscissa of } Y = \tfrac{1}{2} \inf_{Z \sim Y} \lambda_1(Z + Z^*), \tag{16.5}$$

or equivalently

$$\text{spectral abscissa of } Y = \tfrac{1}{2} \inf_{L : \det L \neq 0} \lambda_1(LYL^{-1} + L^{-*}Y^*L^*). \tag{16.6}$$

We can interpret the spectral abscissa characterization (16.5) as a quantitative version of the Lyapunov Stability Theorem. We say a matrix A is *positive stable* if all its eigenvalues have strictly positive real part.

Corollary 16.6. (Lyapunov, 1947) For any matrix A in $M_n(\mathbb{C})$, the following statements are equivalent.

(a) The matrix A is positive stable.
(b) There is a matrix B similar to A for which $B + B^*$ is positive definite.
(c) There is a positive definite matrix W for which $WA + A^*W$ is positive definite.

Proof. The matrix A is positive stable exactly when the spectral abscissa of $-A$ is strictly negative. The equivalence of parts (a) and (b) now follows from the characterization (16.5). The equivalence of parts (b) and (c) follows by observing that W is positive definite if and only if $W = L^*L$ for some invertible L. □

We can give a third form of the spectral abscissa characterization (16.5), (16.6) using the notation of generalized eigenvalue problems. For Hermitian matrices H and W, let $\lambda_1(H, W)$ denote the largest real μ for which there is a nonzero vector x in $\mathbb{C}^n$ satisfying $Hx = \mu Wx$. With this notation, we have, as an immediate consequence of (16.6),

$$\text{spectral abscissa of } Y = \tfrac{1}{2} \inf_{W \succ 0} \lambda_1(WY + Y^*W, W). \tag{16.7}$$

This result is apparently well known in the control theory community, although we are not aware of a standard reference.

Other quantitative results related to the Lyapunov Theorem may be stated by making different choices for the function h in Theorem 16.4(a). Two interesting examples follow.

Example 16.7. (products of eigenvalue real parts) Letting the function h be the logarithmic barrier function (10.2), we have, for a positive stable matrix Y,

$$\prod_{j=1}^{n} \operatorname{Re} \lambda_j(Y) = \sup\{\det \tfrac{1}{2}(Z+Z^*) : Z \sim Y,\ Z+Z^* \succ 0\}.$$

Example 16.8. (sums of eigenvalue real parts) Choosing the function $h(x)$ to be the sum of the k largest x_j (see (10.8)), we obtain

$$\operatorname{Re} \sum_{j=1}^{k} \lambda_j(Y) = \tfrac{1}{2} \inf_{Z \sim Y} \sum_{j=1}^{k} \lambda_j(Z+Z^*).$$

Let us now turn to Theorem 16.4(b):

Example 16.9. (spectral radius) The *spectral radius* of a matrix X in $M_n(\mathbb{C})$ is $\max_j |\lambda_j(X)|$. Applying Theorem 16.4(b) with the symmetric gauge $g(\cdot) = \|\cdot\|_\infty$, we obtain, for any matrix Y in $M_n(\mathbb{C})$,

$$\text{spectral radius of } Y = \inf_{Z \sim Y} \sigma_1(Z) \tag{16.8}$$

(recalling that σ_1 denotes the largest singular value), or equivalently

$$\text{spectral radius of } Y = \inf_{L : \det L \neq 0} \sigma_1(LYL^{-1}). \tag{16.9}$$

A norm f on $M_n(\mathbb{C})$ is *submultiplicative* if $f(AB) \leq f(A)f(B)$ for all $A, B \in M_n(\mathbb{C})$. Clearly, the function $f(Y) = \sigma_1(LYL^{-1})$ is a submultiplicative matrix norm. Furthermore, it is easy to check that the spectral radius of any matrix Y cannot exceed the value of a submultiplicative matrix norm of Y: to prove this, choose $A = Y$ and B such that every column of B is the eigenvector of Y corresponding to the spectral radius. Consequently, equation (16.9) proves the well known fact that the spectral radius is the infimum of all submultiplicative matrix norms (Horn and Johnson 1985, Lemma 5.6.10).

More generally, we have the following example.

Example 16.10. (sums of eigenvalue moduli) Choosing $g(x)$ to be the sum of the k largest $|x_j|$, we obtain

$$\text{sum of } k \text{ largest } |\lambda_j(Y)| = \inf_{Z \sim Y} \sum_{j=1}^{k} \sigma_j(Z).$$

Theorem 16.4 suggests a simple approach to nonsymmetric eigenvalue optimization which, to some extent, avoids the technical difficulties associated with the non-Lipschitz nature of the problem. Given a function $A : \mathbb{R}^m \to M_n(\mathbb{C})$ and a symmetric convex function $h : \mathbb{R}^n \to [-\infty, +\infty]$, consider the optimization problem

$$\inf_{w \in \mathbb{R}^m} \{h(\operatorname{Re} \lambda(A(w)))\}.$$

Using Theorem 16.4(a) we can rephrase this as

$$\inf_{w \in \mathbb{R}^m,\, L \in M_n(\mathbb{C})} \{(h \circ \lambda)(\tfrac{1}{2}(Z + Z^*)) : Z = LA(w)L^{-1},\ L \text{ invertible}\}.$$

Likewise, given a function $f : \mathbb{R}^m \to \mathbb{R}$, the problem

$$\inf_{w \in \mathbb{R}^m} \{f(w) : \operatorname{Re} \lambda(A(w)) \geq 0\}$$

can be rewritten as

$$\inf_{w \in \mathbb{R}^m,\, L \in M_n(\mathbb{C})} \{f(w) : LA(w)L^{-1} + L^{-*}A(w)^*L^* \succeq 0,\ L \text{ invertible}\}.$$

At the expense of introducing the extra variable matrix L, we have reduced these problems to symmetric eigenvalue optimization. Indeed, this idea (using an equivalent Lyapunov formulation based on (16.7)) is exploited in the application of linear matrix inequalities to system and control theory (Boyd et al. 1994); for applications to structural mechanics, see Ringertz (1996).

A similar technique could be applied to the problem

$$\inf_{w \in \mathbb{R}^m} \{g(|\lambda(A(w))|)\},$$

for a symmetric gauge g, this time using Theorem 16.4(b).

However, we caution that there are at least two difficulties with this approach. The first is the expense of introducing so many extra variables (the entire matrix L) into the optimization problem. The second is that the infimum is not likely to be achieved for many interesting applications, a fact that is likely to cause serious difficulties with ill-conditioning.

Acknowledgements. We thank Farid Alizadeh, Steve Cox and Levent Tunçel for various helpful comments. We especially thank Jim Burke and Jean-Pierre Haeberly for contributing unpublished key observations, giving parallel characterizations for the spectral radius and the spectral abscissa, which led to the results of the final section.

REFERENCES

F. Alizadeh (1991), Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices, PhD thesis, University of Minnesota.

F. Alizadeh (1995), 'Interior point methods in semidefinite programming with applications to combinatorial optimization', *SIAM J. Optim.* **5**, 13–51.

F. Alizadeh, J.-P. A. Haeberly and M. L. Overton (1996*a*), 'Complementarity and nondegeneracy in semidefinite programming', *Mathematical Programming (Series B).* To appear.

F. Alizadeh, J.-P. A. Haeberly and M. L. Overton (1996*b*), Primal-dual interior-point methods for semidefinite programming: convergence rates, stability, and numerical results. In preparation.

J. Arazy (1981), 'On the geometry of the unit ball of unitary matrix spaces', *Integral Equations and Operator Theory* **4**, 151–171.

M. S. Ashbaugh and R. D. Benguria (1991), 'Proof of the Payne–Pólya–Weinberger conjecture', *Bulletin of the American Math Society* **25**, 19–29.

A. Barbara and J.-P. Crouzeix (1994), 'Concave gauge functions and applications', *ZOR – Mathematical Methods of Operations Research* **40**, 43–74.

H. Baumgärtel (1985), *Analytic Perturbation Theory for Matrices and Operators*, Birkhäuser, Boston.

R. Bellman and K. Fan (1963), On systems of linear inequalities in Hermitian matrix variables, in *Convexity*, American Mathematical Society, Providence, pp. 1–11. Proceedings of Symposia in Pure Mathematics VII.

S. Boyd, L. E. Ghaoui, E. Feron and V. Balakrishnan (1994), *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia.

J. V. Burke and M. L. Overton (1994), 'Differential properties of the spectral abscissa and the spectral radius for analytic matrix–valued mappings', *Nonlinear Analysis, Theory, Methods and Applications* **23**, 467–488.

F. H. Clarke (1983), *Optimization and Nonsmooth Analysis*, Wiley, New York. Reprinted by SIAM, Philadelphia, 1990.

S. J. Cox (1992), 'The shape of the ideal column', *Mathematical Intelligencer* **14**, 16–31.

S. J. Cox (1993), 'Steven Cox responds', *Mathematical Intelligencer* **15**, 68.

S. J. Cox and J. H. Maddocks (1996), Optimal design of the nonlinear elastica. In preparation.

S. J. Cox and M. L. Overton (1992), 'The optimal design of columns against buckling', *SIAM J. Math. Anal.* **23**, 287–325.

S. J. Cox and M. L. Overton (1996), 'Perturbing the critically damped wave equation', *SIAM J. Math. Anal.* To appear.

B. D. Craven and B. Mond (1981), 'Linear programming with matrix variables', *Linear Algebra and its Applications* **38**, 73–80.

J. Cullum, W. E. Donath and P. Wolfe (1975), 'The minimization of certain nondifferentiable sums of eigenvalues of symmetric matrices', *Mathematical Programming Study* **3**, 35–55.

G. B. Dantzig (1991), Linear programming, in *History of Mathematical Programming: A Collection of Personal Reminiscences* (J. K. Lenstra, A. H. G. Rinnooy Kan and A. Schrijver, eds), North-Holland, Amsterdam, pp. 19–31.

C. Davis (1957), 'All convex invariant functions of Hermitian matrices', *Archiv der Mathematik* **8**, 276–278.

E. M. de Sá (1994*a*), 'Exposed faces and duality for symmetric and unitarily invariant norms', *Linear Algebra and its Applications* **197,198**, 429–450.

E. M. de Sá (1994*b*), 'Faces and traces of the unit ball of a symmetric gauge function', *Linear Algebra and its Applications* **197,198**, 349–395.

E. M. de Sá (1994*c*), 'Faces of the unit ball of a unitarily invariant norm', *Linear Algebra and its Applications* **197,198**, 451–493.

R. Deville, G. Godefroy and V. Zizler (1993), *Smoothness and Renormings in Banach Spaces*, Longman, Harlow, UK.

W. E. Donath and A. J. Hoffman (1973), 'Lower bounds for the partitionings of graphs', *IBM Journal of Research and Development* **17**, 420–425.

M. Fiedler (1973), 'Algebraic connectivity of graphs', *Czech. Math. Journal* **23**, 298–305.

R. Fletcher (1985), 'Semi-definite constraints in optimization', *SIAM J. Control Optim.* **23**, 493–513.

R. Fletcher (1991), 'A new variational result for quasi-Newton formulae', *SIAM J. Optim.* **1**, 18–21.

S. Friedland (1981), 'Convex spectral functions', *Linear and multilinear algebra* **9**, 299–316.

S. Friedland, J. Nocedal and M. L. Overton (1987), 'The formulation and analysis of numerical methods for inverse eigenvalue problems', *SIAM J. Numer. Anal.* **24**, 634–667.

M. X. Goemans and D. P. Williamson (1996), 'Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming', *J. Assoc. Comput. Mach.* To appear.

C. C. Gonzaga (1992), 'Path-following methods for linear programming', *SIAM Review* **34**, 167–224.

M. Grötschel, L. Lovász and A. Schriver (1988), *Geometric Algorithms and Combinatorial Optimization*, Springer, New York.

C. Helmberg (1994), An Interior Point Method for Semidefinite Programming and Max-Cut Bounds, PhD thesis, Technische Universität Graz, Austria.

C. Helmberg, F. Rendl, R. Vanderbei and H. Wolkowicz (1996), 'An interior-point method for semidefinite programming', *SIAM J. Optim.* To appear.

J.-B. Hirriart-Urruty and D. Ye (1995), 'Sensitivity analysis of all eigenvalues of a symmetric matrix', *Numer. Math.* **70**, 45–72.

R. A. Horn and C. R. Johnson (1985), *Matrix Analysis*, Cambridge University Press, UK.

R. A. Horn and C. R. Johnson (1991), *Topics in Matrix Analysis*, Cambridge University Press, UK.

N. Karmarkar (1984), 'A new polynomial time algorithm for linear programming', *Combinatorica* **4**, 373–395.

N. Karmarkar and S. Thakur (1992), An interior-point approach to a tensor optimization problem with application to upper bounds in integer quadratic optimization problems, in *Integer Programming and Combinatorial Optimization* (E. Balas, G. Cornuejils and R. Kannan, eds), pp. 406–419. Proceedings of a conference held at Carnegie Mellon University by the Math. Programming Society.

P. G. Kirmser and K.-K. Hu (1993), 'The shape of the ideal column reconsidered', *Mathematical Intelligencer* **15**, 62–67.

M. Kojima, S. Mizuno and A. Yoshise (1989), A primal-dual interior point method for linear programming, in *Progress in Mathematical Programming, Interior-Point and Related Methods* (N. Megiddo, ed.), Springer, New York, pp. 29–47.

M. Kojima, S. Shindoh and S. Hara (1994), Interior-point methods for the monotone linear complementarity problem in symmetric matrices. Technical Report B-282, Department of Information Sciences, Tokyo Inst. of Technology.

A. S. Lewis (1995*a*), 'The convex analysis of unitarily invariant functions', *J. Convex Anal.* **2**, 173–183.

A. S. Lewis (1995*b*), Eigenvalue-constrained faces. Technical Report CORR 95-22, University of Waterloo. Submitted to Linear Algebra and its Applications.

A. S. Lewis (1995*c*), Von Neumann's lemma and a Chevalley-type theorem for convex functions on Cartan subspaces. Technical Report CORR 95-19, University of Waterloo. Submitted to Transactions of the American Mathematical Society.

A. S. Lewis (1996*a*), 'Convex analysis on the Hermitian matrices', *SIAM J. Optim.* **6**, 164–177.

A. S. Lewis (1996*b*), 'Derivatives of spectral functions', *Mathematics of Operations Research.* To appear.

A. S. Lewis (1996*c*), 'Group invariance and convex matrix analysis', *SIAM J. Matrix Anal. Appl.* To appear.

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

A. W. Marshall and I. Olkin (1979), *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York.

J.-E. Martínez-Legaz (1995), On convex and quasi-convex spectral functions. Technical report, Universitat Autònoma de Barcelona.

L. McLinden (1980), 'An analogue of Moreau's proximation theorem with applications to nonlinear complementarity problems', *Pacific Journal of Mathematics* pp. 101–161.

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

B. Mohar and S. Poljak (1993), Eigenvalues in combinatorial optimization, in *Combinatorial and graph-theoretic problems in linear algebra* (R. Brualdi, S. Friedland and V. Klee, eds), Vol. 50, Springer. IMA Volumes in Mathematics and its Applications.

R. Monteiro and I. Adler (1989), 'Interior path-following primal-dual algorithms: Part I: Linear programming', *Mathematical Programming* **44**, 27–41.

J. Moro, J. V. Burke and M. L. Overton (1995), On the Lidskii–Vishik–Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. Technical Report 710, NYU Comp. Sci. Dept. Submitted to *SIAM J. Matr. Anal. Appl.*

Y. Nesterov and A. Nemirovskii (1994), *Interior Point Polynomial Methods in Convex Programming*, SIAM, Philadelphia.

Y. Nesterov and M. Todd (1996), 'Self-scaled barriers and interior-point methods for convex programming', *Mathematics of Operations Research.* To appear.

N. Olhoff and S. Rasmussen (1977), 'On single and bimodal optimum buckling loads of clamped columns', *Int. J. Solids Struct.* **9**, 605–614.

J. M. Ortega (1972), *Numerical Analysis: A Second Course*, Academic Press, New York and London. Reprinted by SIAM, Philadelphia, 1990.

M. L. Overton (1988), 'On minimizing the maximum eigenvalue of a symmetric matrix', *SIAM J. Matrix Anal. Appl.* **9**, 256–268.

M. L. Overton (1992), 'Large-scale optimization of eigenvalues', *SIAM J. Optim.* **2**, 88–120.

M. L. Overton and R. S. Womersley (1988), 'On minimizing the spectral radius of a nonsymmetric matrix function – optimality conditions and duality theory', *SIAM J. Matrix Anal. Appl.* **9**, 473–498.

M. L. Overton and R. S. Womersley (1993), 'Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices', *Mathematical Programming* **62**, 321–357.

M. L. Overton and R. S. Womersley (1995), 'Second derivatives for optimizing eigenvalues of symmetric matrices', *SIAM J. Matrix Anal. Appl.* **16**, 697–718.

G. Pataki (1995), On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. Technical Report MSRR-604 (revised), Graduate School of Industrial Administration, Carnegie–Mellon University. Submitted to *Math. of Operations Research.*

L. E. Payne, G. Pólya and H. F. Weinberger (1956), 'On the ratio of consecutive eigenvalues', *Journal of Mathematics and Physics* **35**, 289–298.

S. Poljak and Z. Tuza (1993), Max-cut problem – a survey. Technical report, Charles University, Prague.

A. Pothen (1996), Graph partitioning algorithms in scientific computing, in *Parallel Numerical Algorithms* (D. Keyes, A. Sameh and V. Venkatakrishnan, eds), Kluwer. To appear.

F. Rendl and H. Wolkowicz (1992), 'Applications of parametric programming and eigenvalue maximization to the quadratic assignment problem', *Mathematical Programming* **53**, 63–78.

U. T. Ringertz (1995), 'An algorithm for optimization of nonlinear shell structures', *International Journal for Numerical Methods in Engineering* **38**, 299–314.

U. T. Ringertz (1996), Eigenvalues in optimum structural design, in *Proceedings of an IMA workshop on Large-Scale Optimization* (A. R. Conn, L. T. Biegler, T. F. Coleman and F. Santosa, eds). To appear.

R. T. Rockafellar (1970), *Convex Analysis*, Princeton University Press, Princeton University.

H. Schramm and J. Zowe (1992), 'A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results', *SIAM J. Optim.* **2**, 121–152.

A. Seeger (1996), 'Convex analysis of spectrally-defined matrix functions', *SIAM J. Optim.* To appear.

A. Shapiro (1996), 'First and second order analysis of nonlinear semidefinite programs', *Mathematical Programming (Series B)*. To appear.

A. Shapiro and M. K. H. Fan (1995), 'On eigenvalue optimization', *SIAM J. Optim.* **3**, 552–568.

I. Tadjbakhsh and J. B. Keller (1962), 'Strongest columns and isoperimetric inequalities for eigenvalues', *J. Appl. Mech.* **29**, 159–164.

C. M. Theobald (1975), 'An inequality for the trace of the product of two symmetric matrices', *Mathematical Proceedings of the Cambridge Philosophical Society* **77**, 265–266.

N.-K. Tsing, M. K. H. Fan and E. I. Verriest (1994), 'On analyticity of functions involving eigenvalues', *Linear Algebra and its Applications* **207**, 159–180.

L. Tunçel (1995). Private communication.

L. Vandenberghe and S. Boyd (1995), 'Primal-dual potential reduction method for problems involving matrix inequalities', *Mathematical Programming (Series B)* **69**, 205–236.

L. Vandenberghe and S. Boyd (1996), 'Semidefinite programming', *SIAM Review* **38**, 49–95.

G. A. Watson (1992), 'Characterization of the subdifferential of some matrix norms', *Linear Algebra and its Applications* **170**, 33–45.

M. H. Wright (1992), Interior methods for constrained optimization, in *Acta Numerica*, Cambridge University Press, UK, pp. 341–407.

S. J. Wright (1995), 'Stability of linear equations solvers in interior-point methods', *SIAM J. Matrix Anal. Appl.* **16**, 1287–1307.

Y. Zhang and R. A. Tapia (1993), 'A superlinearly convergent polynomial primal-dual interior-point algorithm for linear programming', *SIAM J. Optim.* **3**, 118–133.

Y. Zhang, R. A. Tapia and J. E. Dennis (1992), 'On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms', *SIAM J. Optim.* **2**, 304–324.

K. Ziętak (1988), 'On the characterization of the extremal points of the unit sphere of matrices', *Linear Algebra and its Applications* **106**, 57–75.

K. Ziętak (1993), 'Subdifferentials, faces and dual matrices', *Linear Algebra and its Applications* **185**, 125–141.