VARIATIONAL ANALYSIS OF NON-LIPSCHITZ SPECTRAL FUNCTIONS

JAMES V. BURKE AND MICHAEL L. OVERTON

ABSTRACT. We consider spectral functions $f \circ \lambda$, where f is any permutation-invariant mapping from $\mathbf{C}^n$ to $\mathbf{R}$, and λ is the eigenvalue map from $\mathbf{C}^{n\times n}$ to $\mathbf{C}^n$, ordering the eigenvalues lexicographically. For example, if f is the function "maximum real part", then $f \circ \lambda$ is the spectral abscissa, while if f is "maximum modulus", then $f \circ \lambda$ is the spectral radius. Both these spectral functions are continuous, but they are neither convex nor Lipschitz. For our analysis, we use the notion of subgradient extensively analyzed in *Variational Analysis*, R.T. Rockafellar and R. J.-B. Wets (Springer, 1998), which is particularly well suited to the variational analysis of non-Lipschitz spectral functions. We derive a number of necessary conditions for subgradients of spectral functions. For the spectral abscissa, we give both necessary and sufficient conditions for subgradients, and precisely identify the case where subdifferential regularity holds. We conclude by introducing the notion of semistable programming: minimizing a linear function of a matrix subject to linear constraints, together with the constraint that the eigenvalues of the matrix all lie in the right half-plane or on the imaginary axis. This is a generalization of semidefinite programming for non-Hermitian matrices. Using our analysis, we derive a necessary condition for a local minimizer of a semistable program, and give a generalization of the complementarity condition familiar from semidefinite programming.

1. INTRODUCTION

Let $\mathbf{M}^n$ denote the Euclidean space of $n \times n$ complex matrices. For any $X \in \mathbf{M}^n$, the n eigenvalues of X are the n roots of its characteristic polynomial $\det(\zeta I - X)$. We denote these by $\lambda_1(X), \ldots, \lambda_n(X)$, repeated according to multiplicity and ordered lexicographically so that, if $k < \ell$, then either Re $\lambda_k(X) >$ Re $\lambda_\ell(X)$, or Re $\lambda_k(X) =$ Re $\lambda_\ell(X)$ with Im $\lambda_k(X) \geq$ Im $\lambda_\ell(X)$. Thus we uniquely define the eigenvalue map

$$\lambda : \mathbf{M}^n \to \mathbf{C}^n.$$

This paper considers variational properties of functions of the eigenvalue map. It builds on two foundations. On the one hand, it extends earlier work of the authors [BO92a, BO92b, BO94] as well as other work done by the authors with R.S. Womersley [OW88] and J. Moro [MBO97]. On the other hand, its approach is very much inspired by the beautiful recent work of Adrian Lewis on analysis of eigenvalues for the Hermitian (and real symmetric) matrix case [Lew96a, Lew96b, Lew99].

Following Lewis, we define a *spectral function* (equivalently, an *eigenvalue function*) as an extended-real-valued function of the eigenvalue map, writing it in the

Date: September 28, 1999. Submitted to *Math Programming*.

This work was supported in part by National Science Foundation Grants DMS-9971852 and CCR-9731777, and U.S. Department of Energy Contract DE-FG02-98ER25352.

composite form

$$f \circ \lambda : \mathbf{M}^n \to [-\infty, +\infty], \tag{1.1}$$

where the only restriction on the function $f : \mathbf{C}^n \to [-\infty, +\infty]$ is that it must be invariant under permutation of its argument components. Thus, the lexicographic order used to define λ has no influence on the value of $f \circ \lambda$. This implies that, if f is continuous on $\mathbf{C}^n$, then $f \circ \lambda$ is continuous on $\mathbf{M}^n$ (though λ is not), since the unordered n-tuple of roots of a polynomial is a continuous function of its coefficients. Spectral functions of great interest in applications include the *spectral abscissa*

$$(\max\ \mathrm{Re}) \circ \lambda$$

and the *spectral radius*

$$(\max\ \mathrm{mod}) \circ \lambda$$

where $\mathrm{mod}(x) = |x|$ for $x \in \mathbf{C}$. Although these spectral functions are continuous, they are neither convex nor Lipschitz on $\mathbf{M}^n$. For example, let $t \in \mathbf{R}$ and consider

$$X(t) = \begin{bmatrix} 0 & 1 \\ t & 0 \end{bmatrix}$$

whose eigenvalues are $\pm\sqrt{t}$. We have

$$\alpha(X(t)) = \sqrt{t} \ \text{ if } t \geq 0; \quad 0 \text{ if } t \leq 0$$

and

$$\rho(X(t)) = \sqrt{|t|}.$$

The development of tools for studying the variational properties of general nonconvex functions has been a very active area of research for 25 years, beginning with Clarke's analysis of locally Lipschitz functions [Cla73]. Clarke's generalized gradient is a convex-set-valued map, reducing to the well known subdifferential of convex analysis in the convex case, and to a singleton (the derivative) in the smooth case. In more recent years, attention has turned to the nonconvex-set-valued map advocated by Mordukhovich [Mor76] and Ioffe [Iof81], and extensively analyzed by Rockafellar and Wets [RW98, Chap. 8]. Following Lewis [Lew99], we confine our attention to this map, defining subgradients and horizon subgradients accordingly. As we shall demonstrate, this choice is very well suited to variational analysis of non-Lipschitz spectral functions.

We now introduce the necessary notation; see [RW98, Chap. 8] for more details. Let $\phi : \mathbf{E} \to [-\infty, +\infty]$, where $\mathbf{E}$ is a finite-dimensional Euclidean space, real or complex, with the real inner product $\langle \cdot, \cdot \rangle$, and let $x \in \mathbf{E}$ be such that $\phi(x) < \infty$. A vector $y \in \mathbf{E}$ is a *regular subgradient* of ϕ at x (written $y \in \hat{\partial}\phi(x)$) if

$$\liminf_{z \to 0} \frac{\phi(x+z) - \phi(x) - \langle y, z \rangle}{\|z\|} \geq 0. \tag{1.2}$$

A vector $y \in \mathbf{E}$ is a *subgradient* of ϕ at x (written $y \in \partial\phi(x)$) if there exist sequences x_i and y_i in $\mathbf{E}$ satisfying

$$x_i \to x \tag{1.3}$$

$$\phi(x_i) \to \phi(x) \tag{1.4}$$

$$y_i \in \hat{\partial}\phi(x_i) \tag{1.5}$$

$$y_i \to y. \tag{1.6}$$

A vector $y \in \mathbf{E}$ is a *horizon subgradient* of ϕ at x (written $y \in \partial^\infty \phi(x)$) if $y = 0$ or there exist sequences $x_i, y_i \in \mathbf{E}$ satisfying (1.3), (1.4), and (1.5), but, instead of (1.6),

$$s_i y_i \to y, \quad s_i \downarrow 0,$$

where by $s_i \downarrow 0$, we mean $s_i \to 0$ and $s_i \in \mathbf{R}_+$, the nonnegative real numbers.

It follows from the definition that $\hat{\partial}\phi(x)$, the set of regular subgradients of ϕ at x, is closed and convex (though possibly empty). The set of subgradients, $\partial\phi(x)$, is not necessarily convex. For example, if $\mathbf{E} = \mathbf{R}$ and $\phi(x) = -|x|$, then $\hat{\partial}\phi(0) = \emptyset$, and $\partial\phi(0) = \{-1, 1\}$. For the same example, $\partial^\infty\phi(0) = \{0\}$. For the function $\phi(x) = |x|^{1/2}$, we have $\hat{\partial}\phi(0) = \partial\phi(0) = \partial^\infty\phi(0) = \mathbf{R}$, while for the function $\phi(x) = x^{1/3}$, we have $\hat{\partial}\phi(0) = \partial\phi(0) = \emptyset$ and $\partial^\infty\phi(0) = \mathbf{R}_+$. If ϕ is a convex function, $\hat{\partial}\phi = \partial\phi$ and coincides with the ordinary subdifferential of convex analysis.

We shall also need the concepts of horizon cone and subdifferential regularity, which we define, to avoid unnecessary complication, under the assumption that ϕ is continuous and has at least one regular subgradient at x. In this case, since $\hat{\partial}\phi(x)$ is nonempty, closed and convex, the horizon cone of $\hat{\partial}\phi(x)$ is defined by

$$\hat{\partial}\phi(x)^\infty = \Big\{ y \ : \ \tilde{y} + ty \in \hat{\partial}\phi(x) \quad \forall t \in \mathbf{R}_+ \Big\} \tag{1.7}$$

where $\tilde{y}$ is *any* element of $\hat{\partial}\phi(x)$ [RW98, Theorem 3.6]. Directly from the definitions, we have

$$\hat{\partial}\phi(x) \subseteq \partial\phi(x) \quad \text{and} \quad 0 \in \hat{\partial}\phi(x)^\infty \subseteq \partial^\infty\phi(x).$$

We say that ϕ is *subdifferentially regular* at x if [RW98, Corollary 8.11]

$$\hat{\partial}\phi(x) = \partial\phi(x) \quad \text{and} \quad \hat{\partial}\phi(x)^\infty = \partial^\infty\phi(x).$$

Finally, the *subderivative* of ϕ at x in the direction w is

$$d\phi(x)(w) = \liminf_{\substack{t \downarrow 0 \\ w' \to w}} \frac{\phi(x + tw') - \phi(x)}{t}. \tag{1.8}$$

We have, immediately from the definition, that

$$\hat{\partial}\phi(x) = \{ y \ : \ \langle y, w \rangle \le d\phi(x)(w), \quad \forall w \in \mathbf{E} \}. \tag{1.9}$$

In Appendix A, we establish the following useful lemma:

Lemma 1.1. *Let $\phi : \mathbf{E} \to [-\infty, +\infty]$, with $\phi(x)$ finite, and let $w \in \mathbf{E}$ with $w \neq 0$. Set*

$$\Gamma(x, w) = \left\{ \gamma \ \middle|\ \begin{array}{c} \gamma : [0, \delta] \to \mathbf{E} \text{ is continuous for some } \delta > 0, \\ \text{with } \gamma(0) = x \text{ and } \gamma'_+(0) = w \end{array} \right\}, \tag{1.10}$$

where

$$\gamma'_+(0) = \lim_{t \downarrow 0} \frac{\gamma(t) - \gamma(0)}{t}.$$

Then

$$d\phi(x)(w) = \inf_{\gamma \in \Gamma(x, w)} d(\phi \circ \gamma)(0)(\|w\|),$$

where the infimum is attained.

An immediate corollary is:

Corollary 1.2. *Let $\phi : \mathbf{E} \to [-\infty, +\infty]$, with $\phi(x)$ finite. Then*

$$\hat{\partial}\phi(x) = \{y \,:\, \langle y, w\rangle \le d(\phi\circ\gamma)(0)(1), \quad \forall \gamma \in \Gamma(x,w) \text{ with } \|w\| = 1\}\ .$$

Our primary interest is in the case where $\mathbf{E}$ is a complex space. It is important to note that the definitions given above are independent of whether we regard $\mathbf{E}$ as a complex space, say $\mathbf{C}^n$, or the corresponding real space, $\mathbf{R}^{2n}$. For example, if $\mathbf{E} = \mathbf{C}$ and $\phi(x) = \text{Re } x$, then $\partial\phi(x) = \hat{\partial}\phi(x) = \{1\}$, while if $\phi(x) = |x|$, then $\partial\phi(x) = \hat{\partial}\phi(x) = \{x/|x|\}$ for $x \neq 0$ and $\{y : |y| \le 1\}$ for $x = 0$. Thus, our use of a complex domain is purely for convenience; all results could be stated equivalently using a real domain.

The convenient interplay between $\mathbf{C}$ and $\mathbf{R}^2$ is illustrated by the following pair of lemmas, which we shall need later. First we need to define some notation. The inner product on $\mathbf{C}^n$ is defined by

$$\langle x, y\rangle = \text{Re } x^* y = \text{Re } \sum_j \overline{x}_j y_j.$$

where x^* denotes the conjugate transpose of the vector x, with $\overline{x}_j$ denoting the complex conjugate of x_j. The norm $\|x\|$ denotes the Euclidean vector norm. The imaginary unit is denoted $\sqrt{-1}$.

Let $\gamma : \mathbf{R}^2 \to \mathbf{R}$ be given, and let us define a function $\kappa : \mathbf{C} \to \mathbf{R}$ by

$$\kappa(\mu) = \gamma(x) \quad \text{where} \quad x = \begin{bmatrix} \text{Re } \mu \\ \text{Im } \mu \end{bmatrix}.$$

If γ is differentiable, with gradient $\nabla\gamma$, we define $\kappa' : \mathbf{C} \to \mathbf{C}$ by

$$\kappa'(\mu) = (\nabla\gamma(x))_1 + \sqrt{-1}(\nabla\gamma(x))_2. \tag{1.11}$$

If γ is continuously differentiable at x, then κ' is continuous at μ, and we say that κ is differentiable in the real sense. If γ is twice differentiable, with Hessian ∇^2, we define $\kappa'' : \mathbf{C}^2 \to \mathbf{C}$ by

$$\begin{aligned} \kappa''(\mu;\nu) &= (\nabla^2\gamma(x))_{11}\text{Re } \nu + (\nabla^2\gamma(x))_{12}\text{Im } \nu \\ &+ \sqrt{-1}\,(\nabla^2\gamma(x))_{12}\text{Re } \nu + \sqrt{-1}\,(\nabla^2\gamma(x))_{22}\text{Im } \nu. \end{aligned} \tag{1.12}$$

If γ is twice continuously differentiable at x, then $\kappa''(\cdot,\nu)$ is continuous at μ, and we say that κ is twice differentiable at μ in the real sense.

Taylor's theorem tells us that

$$\gamma(x+sv) = \gamma(x) + s\langle\nabla\gamma(x), v\rangle + \frac{1}{2}s^2\big\langle v, \nabla^2\gamma(x)v\big\rangle + o(s^2) \tag{1.13}$$

for $v \in \mathbf{R}^2$ and $s \in \mathbf{R}$. It follows that

$$\kappa(\mu+s\nu) = \kappa(\mu) + s\langle\kappa'(\mu), \nu\rangle + \frac{1}{2}s^2\langle\nu, \kappa''(\mu,\nu)\rangle + o(s^2) \tag{1.14}$$

for $\nu \in \mathbf{C}$ and $s \in \mathbf{R}$. We therefore have:

Lemma 1.3. *Let $x \in \mathbf{R}^2$ and let $\gamma : \mathbf{R}^2 \to \mathbf{R}$ be twice continuously differentiable at x, with gradient $\nabla\gamma$ and Hessian $\nabla^2\gamma$. Let $x \in \mathbf{R}^2$ and suppose $\nabla\gamma(x) \neq 0$. Let $v, w \in \mathbf{R}^2$ satisfy $v \perp \nabla\gamma(x)$ and $w = \delta\nabla\gamma(x)$, where*

$$\delta = -\frac{v^T\nabla^2\gamma(x)v}{2\|\nabla\gamma(x)\|^2}.$$

Then

$$\gamma(x + sv + s^2 w) = \gamma(x) + o(s^2),$$

for $s \in \mathbf{R}$.

Proof. Apply (1.13). QED

Lemma 1.4. *Let the assumptions of the previous lemma hold and define* κ, κ' *and* κ'' *as above. Let* $\nu, \omega \in \mathbf{C}$ *satisfy* $\nu = \pm\sqrt{-1}\kappa'(\mu)$ *and* $\omega = \delta\kappa'(\mu)$, *where*

$$\delta = -\frac{\langle \nu, \kappa''(\mu, \nu)\rangle}{2|\nu|^2} \in \mathbf{R}.$$

Then

$$\kappa(\mu + s\nu + s^2\omega) = \kappa(\mu) + o(s^2),$$

for $s \in \mathbf{R}$.

Proof. This is the complex analogue of Lemma 1.3, and is proved using (1.14). QED

We conclude the introduction with some matrix notation. The inner product on $\mathbf{M}^n$ is defined by

$$\langle X, Y\rangle = \mathrm{Re}\ \mathrm{tr}\ X^*Y = \mathrm{Re}\sum_{r,s} \overline{x}_{rs} y_{rs}.$$

By X^{-*} we mean $(X^*)^{-1} = (X^{-1})^*$. By $\|X\|$ we mean $\sqrt{\langle X, X\rangle}$, though any norm would suffice. By $\mathrm{Diag}(x)$ we mean the diagonal matrix constructed from the vector x, while $\mathrm{diag}(X)$ is the vector constructed from the diagonal entries in the matrix X. The identity matrix is denoted I, and the vector whose components are all one is denoted e; their dimensions will be evident from the context.

2. Commutativity and the Schur Form

The following result is essential for all subsequent analysis.

Theorem 2.1. *If* Y *is a subgradient or horizon subgradient of a spectral function* $f \circ \lambda$ *at* X, *then* $Y^*X = XY^*$.

Proof. We follow the proof in [Lew99, Theorem 3], where a closely related result is given for spectral functions on the space of Hermitian matrices. Instead of [Lew99, Theorem 1], the result we need here is that the orbit of X, that is the set of matrices similar to X, is a submanifold whose tangent space at X is given by

$$T_X = \{XZ - ZX : Z \in \mathbf{M}^n\}$$

and whose normal space at X is given by

$$(T_X)^\perp = \{Y \in \mathbf{M}^n : XY^* = Y^*X\}.$$

This fact is presented in [Arn71]. Although a proof of the formula for T_X is not to be found in [Arn71], one is easily constructed by generalizing the proof of [Lew99, Theorem 1] to the non-Hermitian case. The rest of the proof follows exactly as in the proof of [Lew99, Theorem 3]. QED

A unitary matrix U transforms X into *Schur form* if U^*XU is upper triangular. An immediate corollary of Theorem 2.1 is the existence of a unitary matrix U which simultaneously transforms both X and Y^* to Schur form:

Corollary 2.2. *If Y is a subgradient or horizon subgradient of a spectral function $f \circ \lambda$ at X, then there exists a unitary matrix U which simultaneously triangularizes X and Y^*, i.e. such that*

$$R = U^* X U \quad \text{and} \quad S = U^* Y U \tag{2.1}$$

are respectively upper and lower triangular. Furthermore, U can be chosen so that the diagonal components of R appear in any desired order, e.g.

$$\operatorname{diag}(R) = \lambda(X). \tag{2.2}$$

Proof. The existence of the simultaneously triangularizing unitary matrix U follows from [HJ85, Thm 2.3.3]. For the ordering, see Lemma B.1 in Appendix B. QED

To go further we must establish some more notation. Let $\mu_1, \ldots, \mu_p$ be the *distinct* eigenvalues of X, ordered lexicographically. Thus $\lambda(X)$ is a vector whose components are the μ_j, repeated according to multiplicity. Let $m^{(j)}$ be the multiplicity of the eigenvalue μ_j. Given a Schur form $R = U^* X U$, where U is unitary and $\operatorname{diag}(R) = \lambda(X)$, we may partition R into the block upper triangular form

$$R = \begin{bmatrix} R^{(11)} & \cdots & R^{(1p)} \\ & \ddots & \vdots \\ & & R^{(pp)} \end{bmatrix} \tag{2.3}$$

where, for each j, $R^{(jj)}$ is upper triangular and

$$\operatorname{diag}(R^{(jj)}) = \mu_j e \in \mathbf{C}^{m^{(j)}}, \tag{2.4}$$

i.e., all diagonal components of $R^{(jj)}$ equal μ_j. It will be convenient to also partition S, satisfying (2.1), conformally, as

$$S = \begin{bmatrix} S^{(11)} & & \\ \vdots & \ddots & \\ S^{(p1)} & \cdots & S^{(pp)} \end{bmatrix}, \tag{2.5}$$

where, for each j, $S^{(jj)}$ is lower triangular.

The following lemmas will be useful.

Lemma 2.3. *Suppose that*

$$T^{-1} R T = \tilde{R}$$

where R and $\tilde{R}$ are both upper triangular, and the diagonal components of R and $\tilde{R}$ are both ordered lexicographically, i.e., $\operatorname{diag}(R) = \operatorname{diag}(\tilde{R})$. *Let R and $\tilde{R}$ have the block triangular structure given in* (2.3),(2.4). *Then T has the same block triangular structure. Furthermore, if T is unitary, it is not only block triangular, but block diagonal.*

Proof. The proof recursively applies the result for the following partitioning:

$$R = \begin{bmatrix} R^{(11)} & R^{(12)} \\ 0 & R^{(22)} \end{bmatrix}, \quad \tilde{R} = \begin{bmatrix} \tilde{R}^{(11)} & \tilde{R}^{(12)} \\ 0 & \tilde{R}^{(22)} \end{bmatrix},$$

where the diagonal blocks are square with dimensions n_1 and n_2 respectively, and where no diagonal entry in $R^{(11)}$ appears on the diagonal of $R^{(22)}$. Recall that

$\mathrm{diag}(R) = \mathrm{diag}(\tilde{R})$. Let

$$T = \begin{bmatrix} T^{(11)} & T^{(12)} \\ T^{(21)} & T^{(22)} \end{bmatrix}.$$

Since $RT = T\tilde{R}$, we have

$$R^{(22)}T^{(21)} = T^{(21)}\tilde{R}^{(11)}.$$

Since $\tilde{R}^{(11)}$ and $R^{(22)}$ have no common diagonal entry, we conclude, applying [HJ91, p. 270], that $T^{(21)} = 0$. This shows that T is block triangular. Furthermore, it follows immediately from the definition that if T is unitary, we also have $T^{(12)} = 0$. QED

Lemma 2.4. *Let the assumptions of Lemma 2.3 hold, and assume also that*

$$T^*ST^{-*} = \tilde{S},$$

where S and $\tilde{S}$ both have the block structure shown in (2.5)*, with $S^{(jj)}$ (but not necessarily $\tilde{S}^{(jj)}$) lower triangular for each j. Then, for each j, the blocks $S^{(jj)}$ and $\tilde{S}^{(jj)}$ have the same eigenvalues. Furthermore, if $\tilde{S}^{(jj)}$ is also lower triangular for each j, then there exists a permutation matrix Q such that*

$$Q\,\mathrm{diag}(R) = \mathrm{diag}(R) = \mathrm{diag}(\tilde{R}) \quad \text{and} \quad Q\,\mathrm{diag}(S) = \mathrm{diag}(\tilde{S}).$$

Proof. Since, by Lemma 2.3, T^* is block lower triangular, we have, for each j,

$$(T^{(jj)})^* S^{(jj)} (T^{(jj)})^{-*} = \tilde{S}^{(jj)}, \tag{2.6}$$

so the eigenvalues of $S^{(jj)}$ and $\tilde{S}^{(jj)}$ are the same. If the matrices are lower triangular, their eigenvalues appear on the diagonals. Hence, for the jth block, there is a permutation matrix $Q^{(j)}$ such that

$$Q^{(j)}\mathrm{diag}(S^{(jj)}) = \mathrm{diag}(\tilde{S}^{(jj)}). \tag{2.7}$$

Now set Q to be the block diagonal permutation matrix whose jth block is $Q^{(j)}$. Multiplication by Q leaves $\mathrm{diag}(R)$ invariant since each of the diagonal blocks of R has constant diagonal entries, so the proof is complete. QED

An immediate consequence of Lemma 2.4 is that, although S in (2.1) may not be unique, its diagonal entries are uniquely determined, up to permutations within blocks.

3. A General Necessary Condition for Subgradients of $f \circ \lambda$ in terms of Subgradients of f

The following is a key result.

Theorem 3.1. *Let Y be a subgradient or horizon subgradient of a spectral function $f \circ \lambda$ at X, i.e.*

$$Y \in \partial(f \circ \lambda)(X) \quad \text{or} \quad Y \in \partial^\infty(f \circ \lambda)(X)$$

respectively, with $R = U^*XU$ upper triangular, $S = U^*YU$ lower triangular, and $\mathrm{diag}(R) = \lambda(X)$*, for some unitary matrix U, as in Corollary 2.2. Then*

$$\mathrm{diag}(S) \in \partial f(\lambda(X)) \quad \text{or} \quad \mathrm{diag}(S) \in \partial^\infty f(\lambda(X))$$

respectively. Furthermore, if Y is a regular subgradient of $f \circ \lambda$, then $\mathrm{diag}(S)$ *is a regular subgradient of f.*

Proof. First suppose that Y is a regular subgradient. Then

$$\begin{aligned} f(\operatorname{diag}(R)+z) &= (f\circ\lambda)(U(R+\operatorname{Diag}(z))U^*) && (3.1)\\ &= (f\circ\lambda)(X+U\operatorname{Diag}(z)U^*) \\ &\geq (f\circ\lambda)(X)+\langle Y, U\operatorname{Diag}(z)U^*\rangle + o(\|z\|) && (3.2)\\ &= f(\operatorname{diag}(R))+\langle \operatorname{diag}(S), z\rangle + o(\|z\|) && (3.3)\end{aligned}$$

so $\operatorname{diag}(S) \in \hat{\partial} f(\operatorname{diag}(R)) = \hat{\partial} f(\lambda(X))$. Here (3.1) and (3.3) hold because the eigenvalues of a triangular matrix appear on its diagonal, and (3.2) follows directly from the definition (1.2).

Now assume only that Y is a subgradient, not necessarily regular, so there is a sequence of matrices $X_i \to X$, with $f(\lambda(X_i)) \to f(\lambda(X))$ and a sequence of regular subgradients $Y_i \in \hat{\partial}(f\circ\lambda)(X_i)$, with $Y_i \to Y$. By Corollary 2.2 there exists a sequence of unitary matrices U_i with

$$R_i = U_i^* X_i U_i \quad \text{and} \quad S_i = U_i^* Y_i U_i$$

respectively upper and lower triangular for all i. Furthermore, the freedom in the simultaneous triangularization procedure allows us to choose the order of the diagonal components in R_i so that $\operatorname{diag}(R_i) \to \operatorname{diag}(R) = \lambda(X)$. (This does not imply that $\operatorname{diag}(R_i)$ is lexicographically ordered.) From an identical argument to (3.1)–(3.3), we have

$$\operatorname{diag}(S_i) \in \hat{\partial} f(\operatorname{diag}(R_i)). \tag{3.4}$$

Since the set of all unitary matrices is compact, we can also assume $U_i \to \tilde{U}$, which, while not necessarily the same as U, is also a simultaneously triangularizing matrix. Let $\tilde{R} = \tilde{U}^* X \tilde{U}$ and $\tilde{S} = \tilde{U}^* Y \tilde{U}$; by construction, $\operatorname{diag}(\tilde{R}) = \operatorname{diag}(R)$, and $\tilde{R}$ and $\tilde{S}$ are respectively upper and lower triangular. We have

$$\hat{U}^* R \hat{U} = \tilde{R} \quad \text{and} \quad \hat{U}^* S \hat{U} = \tilde{S}$$

where $\hat{U} = U^*\tilde{U}$ is unitary, allowing us to apply Lemmas 2.3 and 2.4 to obtain the existence of a permutation matrix Q satisfying

$$Q\operatorname{diag}(R) = \operatorname{diag}(R) = \operatorname{diag}(\tilde{R}) \quad \text{and} \quad Q\operatorname{diag}(S) = \operatorname{diag}(\tilde{S}).$$

Taking limits in (3.4) yields

$$\operatorname{diag}(\tilde{S}) \in \partial f(\operatorname{diag}(\tilde{R})). \tag{3.5}$$

By [RW98, 10.7, p. 428] or [Lew99, Proposition 2],

$$V\operatorname{diag}(\tilde{S}) \in \partial f(V\operatorname{diag}(\tilde{R})),$$

for any unitary matrix V. Choosing $V = Q^T$ completes the proof.

The proof for the horizon subgradients is identical: instead of $Y_i \to Y$, we have $s_i Y_i \to Y$, where $s_i \downarrow 0$, and so instead of (3.5), we obtain $\operatorname{diag}(\tilde{S}) \in \partial^\infty f(\operatorname{diag}(\tilde{R}))$. QED

Both the statement and the proof of this result were inspired by Lewis [Lew99, Proposition 5], where a related result was proved for the Hermitian case.

4. Necessary Conditions Based on the Jordan Form

A nonsingular matrix P transforms X to *Jordan form* if

$$P^{-1}XP = J = \begin{bmatrix} J^{(1)} & & \\ & \ddots & \\ & & J^{(p)} \end{bmatrix}, \text{ where } J^{(j)} = \begin{bmatrix} J_1^{(j)} & & \\ & \ddots & \\ & & J_{q^{(j)}}^{(j)} \end{bmatrix}, \tag{4.1}$$

$$\text{with } J_k^{(j)} = \begin{bmatrix} \mu_j & 1 & & & \\ & \cdot & \cdot & & \\ & & \cdot & \cdot & \\ & & & \cdot & 1 \\ & & & & \mu_j \end{bmatrix}, \quad k = 1, \ldots, q^{(j)}, \quad j = 1, \ldots, p. \tag{4.2}$$

Each $J_k^{(j)}$ is a Jordan block of size $m_k^{(j)} \times m_k^{(j)}$ for the eigenvalue μ_j. The multiplicity of μ_j is

$$m^{(j)} = \sum_{k=1}^{q^{(j)}} m_k^{(j)}.$$

The size of the largest Jordan block for μ_j is denoted

$$n^{(j)} = \max_{k=1,\ldots,q^{(j)}} m_k^{(j)}.$$

An eigenvalue μ_j is said to be *nonderogatory* if $q^{(j)} = 1$ and *semisimple* if $n^{(j)} = 1$. These cases coincide if and only if $m^{(j)} = 1$, in which case μ_j is said to be *simple.* The set of matrices with a given Jordan block structure defines a submanifold of $\mathbf{M}^n$ whose properties are well known [Arn71]. Nonderogatory Jordan structures are the most generic.

We note that

$$XP = PJ, \quad \text{and} \quad P^{-1}X = JP^{-1}.$$

Therefore, for each Jordan block $J_k^{(j)}$, the corresponding block of $m_k^{(j)}$ columns of P (respectively rows of P^{-1}) contains a chain of $m_k^{(j)}$ generalized right (respectively left) eigenvectors of X. The first column (respectively last row) in this block is a right (respectively left) eigenvector. When μ_j is semisimple, the corresponding chains have length one, so the generalized eigenvectors are actually eigenvectors.

We also define

$$N^{(j)} = J^{(j)} - \mu_j I, \quad j = 1, \ldots, p. \tag{4.3}$$

The matrix $N^{(j)}$ is called the nilpotent part of $J^{(j)}$, since $(N^{(j)})^{n^{(j)}} = 0$.

Theorem 4.1. *If Y is a subgradient or horizon subgradient of a spectral function $f \circ \lambda$ at X, then any P satisfying* (4.1), (4.2) *also satisfies*

$$P^*YP^{-*} = W = \begin{bmatrix} W^{(1)} & & \\ & \ddots & \\ & & W^{(p)} \end{bmatrix}, \; W^{(j)} = \begin{bmatrix} W_{11}^{(j)} & \cdots & W_{1q^{(j)}}^{(j)} \\ \vdots & \vdots & \vdots \\ W_{q^{(j)}1}^{(j)} & \cdots & W_{q^{(j)}q^{(j)}}^{(j)} \end{bmatrix}, \tag{4.4}$$

where $W_{rs}^{(j)}$ is a rectangular $m_r^{(j)} \times m_s^{(j)}$ lower triangular Toeplitz matrix, $r = 1, \ldots, q^{(j)}$, $s = 1, \ldots, q^{(j)}$, $j = 1, \ldots, p$. By this we mean that, for each $W_{rs}^{(j)}$, the value of *the k,ℓ entry in $W_{rs}^{(j)}$ depends only on the difference $k-\ell$ (is constant along the diagonals), and is zero if $k<l$ or $m_r^{(j)}-k>m_s^{(j)}-\ell$ (is zero above the main diagonal, drawn either from the top left of the block, or from the bottom right).*

Proof. The proof follows immediately from the fact that the matrices commuting with the Jordan form J are exactly the matrices W described in the theorem statement; see [LT85, Sec. 12.4] for a proof and [Arn71, Sec. 4.2] or [OW88] for illustrations. QED

It follows immediately that if an eigenvalue μ_j is nonderogatory ($q^{(j)}=1$), then $W^{(j)}$ is lower triangular Toeplitz, i.e.,

$$W^{(j)} = \begin{bmatrix} \theta_1^{(j)} & & & & \\ \theta_2^{(j)} & \cdot & & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \\ \theta_{m^{(j)}}^{(j)} & \cdot & \cdot & \theta_2^{(j)} & \theta_1^{(j)} \end{bmatrix}, \tag{4.5}$$

for some $\theta_\ell^{(j)}$, $\ell=1,\dots,m^{(j)}$.

We can relate the conditions on subgradients derived from the Schur and Jordan forms as follows.

Corollary 4.2. *Let Y be any subgradient or horizon subgradient of a spectral function $f\circ\lambda$, satisfying* (2.1), (2.2), (2.3), (2.4), (2.5) *as well as* (4.1), (4.2), (4.4). *Then, for each j, $S^{(jj)}$ and $W^{(j)}$ have the same eigenvalues, namely, the diagonal entries of $S^{(jj)}$. Furthermore, if μ_j is nonderogatory, $S^{(jj)}$ and $W^{(j)}$ are both lower triangular with the same constant diagonal entry.*

Proof. We have

$$X = URU^* = PJP^{-1}, \quad Y = USU^* = P^{-*}WP^*$$

so

$$T^{-1}RT = J, \quad T^*ST^{-*} = W, \tag{4.6}$$

where $T=U^*P$. Applying Lemma 2.4 with $\tilde{R}=J$ and $\tilde{S}=W$ gives the desired result. The last statement is an immediate consequence of the fact that $W^{(j)}$ is lower triangular Toeplitz in the nonderogatory case. QED

For regular subgradients, there is a much stronger result.

Theorem 4.3. *If Y is a* **regular** *subgradient of a spectral function $f\circ\lambda$ at X, then any P satisfying* (4.1), (4.2) *also satisfies*

$$P^*YP^{-*} = W = \begin{bmatrix} W^{(1)} & & \\ & \ddots & \\ & & W^{(p)} \end{bmatrix}, \; W^{(j)} = \begin{bmatrix} W_{11}^{(j)} & & \\ & \ddots & \\ & & W_{q^{(j)}q^{(j)}}^{(j)} \end{bmatrix}, \tag{4.7}$$

$$\text{where } W_{kk}^{(j)} = \begin{bmatrix} \theta_1^{(j)} & & & & \\ \theta_2^{(j)} & \cdot & & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \\ \theta_{m_k^{(j)}}^{(j)} & \cdot & \cdot & \theta_2^{(j)} & \theta_1^{(j)} \end{bmatrix}, k=1,\dots,q^{(j)}, \quad j=1,\dots,p, \tag{4.8}$$

for some $\theta_\ell^{(j)}$, $\ell = 1, \ldots, n^{(j)}$, $j = 1, \ldots, p$. *Thus, for each* j, $W^{(j)}$ *is block diagonal with (square) lower triangular Toeplitz blocks, and, furthermore, the entries on the diagonals of the Toeplitz blocks are constant not only within each block, but also across all* $q^{(j)}$ *blocks. Finally,*

$$W^{(j)} = \sum_{\ell=1}^{n^{(j)}} \theta_\ell^{(j)} \left(\left(N^{(j)}\right)^*\right)^{\ell-1}, \quad j = 1, \ldots, p, \tag{4.9}$$

where $N^{(j)}$ *is defined in* (4.3).

Proof. Suppose that for some j, $W^{(j)}$ has a nonzero entry in an off-diagonal block of (4.4); suppose this occurs in the rth row and sth column of the entire matrix W and let β be this nonzero value. Let $Z = PVP^{-1}$, where all components of V are zero except the r, s component, which is set to β. Then

$$\langle Y, Z\rangle = \langle W, V\rangle = |\beta|^2 > 0.$$

The eigenvalues of $X + tZ$ are the same as the eigenvalues of X for all $t \in \mathbf{R}$, so

$$\liminf_{t\to 0} \frac{(f\circ\lambda)(X + tZ) - (f\circ\lambda)(X) - \langle Y, tZ\rangle}{\|tZ\|} = -\frac{\langle Y, Z\rangle}{\|Z\|} < 0. \tag{4.10}$$

Thus Y is not a regular subgradient of $f \circ \lambda$ at X (substituting tZ for z in (1.2)). This proves that the off-diagonal blocks of $W^{(j)}$ are zero. That the diagonal blocks are lower triangular and Toeplitz is known from Theorem 4.1.

We must now show that, for each j, and each pair k, k', with $1 \le k < k' \le q^{(j)}$, and each ℓ satisfying $1 \le \ell \le \min(m_k^{(j)}, m_{k'}^{(j)})$, the constant entry on the diagonal $\ell - 1$ positions below the main diagonal of $W_{kk}^{(j)}$ equals the constant entry on the diagonal $\ell - 1$ positions below the main diagonal of $W_{k'k'}^{(j)}$. Suppose this is not the case for some j, k, k' and ℓ. Without loss of generality we may assume $k' = k + 1$. Let r be the integer such that the rth diagonal entry of the entire matrix W is in the last diagonal position of $W_{kk}^{(j)}$ and, therefore, the $(r + 1)$th diagonal entry of W is in the first diagonal position of $W_{k+1,k+1}^{(j)}$. Now consider the case $\ell = 1$, so that the diagonals in question are the main diagonals of the blocks $W_{kk}^{(j)}$ and $W_{k+1,k+1}^{(j)}$, with constant values β_1 and β_2 respectively, with $\beta_1 \neq \beta_2$. Suppose further that $\mathrm{Re}\beta_1 > \mathrm{Re}\beta_2$. Let $Z = PVP^{-1}$, where V has all zero components except

$$\begin{bmatrix} v_{r,r} & v_{r,r+1} \\ v_{r+1,r} & v_{r+1,r+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}. \tag{4.11}$$

We have

$$\langle Y, Z\rangle = \langle W, V\rangle = \mathrm{Re}\beta_1 - \mathrm{Re}\beta_2 > 0. \tag{4.12}$$

Both eigenvalues of (4.11) are zero, so, since r and $r + 1$ correspond to different Jordan blocks of J corresponding to the same eigenvalue μ_j, the eigenvalues of $X + tZ$ are the same as the eigenvalues of X for all t. Therefore (4.10) holds, and Y is not a regular subgradient of $f \circ \lambda$ at X. If $\mathrm{Re}\beta_1 < \mathrm{Re}\beta_2$, we reverse the sign of V and make the same conclusion. If the real parts of β_1 and β_2 are the same, their imaginary parts must differ, and so we multiply V by $\pm\sqrt{-1}$ and reach the same conclusion. This completes the proof for the case $\ell = 1$, showing that the constant on the main diagonals of $W_{kk}^{(j)}$ is the same for all $k = 1, \ldots, q^{(j)}$.

We now generalize this argument to the case $\ell > 1$. Consider the diagonals $\ell - 1$ positions below the main diagonals of the blocks $W_{kk}^{(j)}$ and $W_{k+1,k+1}^{(j)}$, with constant values β_1 and β_2 respectively, with $\beta_1 \neq \beta_2$. Suppose again that $\mathrm{Re}\beta_1 > \mathrm{Re}\beta_2$. Let $Z = PVP^{-1}$, where V has all zero components except in the four entries whose row index is either r or $r + \ell$ and whose column index is either $r - \ell + 1$ or $r + 1$. Let the two nonzero entries in row r have the value 1 and the two nonzero entries in row $r + \ell$ have the value -1, so that (4.12) holds. We must now determine the eigenvalues of $X + tZ$. The part of $J + tV$ which needs examination is the following diagonal block (of dimension 2ℓ):

$$\begin{bmatrix} \mu_j & 1 & & & & & & \\ & \ddots & \ddots & & & & & \\ & & \ddots & 1 & & & & \\ t & & & \mu_j & t & & & \\ & & & & \mu_j & 1 & & \\ & & & & & \ddots & \ddots & \\ & & & & & & \ddots & 1 \\ -t & & & & -t & & & \mu_j \end{bmatrix}. \tag{4.13}$$

Consideration of the characteristic polynomial shows that all eigenvalues of (4.13) equal μ_j, for all t. Therefore, the eigenvalues of $X + tZ$ are the same as the eigenvalues of X for all t, (4.10) holds, and Y is not a regular subgradient of $f \circ \lambda$ at X. As earlier, if it is not the case that $\mathrm{Re}\beta_1 > \mathrm{Re}\beta_2$, the proof is modified by scaling the choice of V appropriately.

The final statement is an immediate consequence of the definition of the nilpotent matrix $N^{(j)}$. QED

If an eigenvalue μ_j is nonderogatory, i.e. $q^{(j)} = 1$, the structure on $W^{(j)}$ imposed by (4.4) and that imposed by (4.7) are the same, but the latter is more restrictive if μ_j is derogatory. In Section 8, we shall see that, in the derogatory case, subgradients do not necessarily satisfy the more restrictive block diagonal condition required for regular subgradients.

The condition on the regular subgradients, derived from the Jordan form, can now be related to the condition derived from the Schur form.

Corollary 4.4. *Let Y be a* **regular** *subgradient of a spectral function $f \circ \lambda$ at X, and assume* (2.1)*,* (2.2)*,* (2.3)*,* (2.4)*,* (2.5) *as well as* (4.1)*,* (4.2)*,* (4.7)*,* (4.8) *all hold. Then*

$$\mathrm{diag}(W) = \mathrm{diag}(S) \in \partial f(\lambda(X)),$$

with

$$\mathrm{diag}(S^{(jj)}) = \theta_1^{(j)} e \in \mathbf{C}^{m^{(j)}}, \quad j = 1, \ldots, p.$$

Proof. Since Y is regular, $W^{(j)}$, like $S^{(jj)}$, is lower triangular. Therefore, by Corollary 4.2 and Lemma 2.4, we know there exists a permutation matrix Q satisfying

$$Q\,\mathrm{diag}(J) = \mathrm{diag}(J) \quad \text{and} \quad Q\,\mathrm{diag}(W) = \mathrm{diag}(S).$$

This shows that $\mathrm{diag}(W) = \mathrm{diag}(S)$, since, from Theorem 4.3, any permutation matrix Q satisfying $Q\,\mathrm{diag}(J) = \mathrm{diag}(J)$ also satisfies $Q\,\mathrm{diag}(W) = \mathrm{diag}(W)$. We know that $\operatorname{diag}(S) \in \partial f(\lambda(X))$ from Theorem 3.1. The last statement is an immediate consequence. QED

5. Further Decomposition of the Spectral Function

In order to state additional necessary conditions that subgradients must satisfy, we assume the spectral function $f \circ \lambda$ can be decomposed further as

$$f \circ \lambda = g \circ h \circ \lambda \tag{5.1}$$

where $g : \mathbf{R}^n \to [-\infty, +\infty]$ and $h : \mathbf{C}^n \to \mathbf{R}^n$, with g invariant under permutations of its argument components and h mapping each of its argument components by the same complex-to-real function, i.e.,

$$h_\ell(\lambda) = \kappa(\lambda_\ell), \quad \ell = 1, \ldots, n \tag{5.2}$$

where $\kappa : \mathbf{C} \to \mathbf{R}$. For example, if $g = \max$ and $\kappa = \operatorname{Re}$, the composite function is the spectral abscissa, while if $g = \max$ and $\kappa = \operatorname{mod}$, it is the spectral radius. Recall from Section 1 that κ is continuously differentiable at $\mu \in \mathbf{C}$ in the real sense if its derivative κ' defined in (1.11) is continuous at μ. A chain rule then gives the following:

Theorem 5.1. *Let* (5.2) *hold, where κ is continuously differentiable at μ_j in the real sense, $j = 1, \ldots, p$, and let*

$$K = \operatorname{diag}\left([\kappa'(\lambda_1(X)), \ldots, \kappa'(\lambda_n(X))]^T\right).$$

Suppose that $Kz = 0$ implies $z = 0$ for all $z \in \partial^\infty g(h(\lambda(X)))$, i.e. for all horizon subgradients of g at $h(\lambda(X))$. (This is true, for example, if K is nonsingular, or if g is convex and finite-valued.) Let Y be a subgradient or horizon subgradient of $g \circ h \circ \lambda$ at X, i.e.

$$Y \in \partial(g \circ h \circ \lambda)(X) \quad \text{or} \quad Y \in \partial^\infty(g \circ h \circ \lambda)(X)$$

respectively, with $R = U^*XU$ upper triangular, $S = U^*YU$ lower triangular, and $\operatorname{diag}(R) = \lambda(X)$*, for some unitary matrix U, as in Corollary 2.2. Then*

$$\operatorname{diag}(S) = Kw$$

where

$$w \in \partial g(h(\lambda(X))) \subseteq \mathbf{R}^n \quad \text{or} \quad w \in \partial^\infty g(h(\lambda(X))) \subseteq \mathbf{R}^n$$

respectively.

Proof. Applying Theorem 3.1 with $f = g \circ h$, we find

$$\operatorname{diag}(S) \in \partial(g \circ h)(\lambda(X)) \quad \text{or} \quad \operatorname{diag}(S) \in \partial^\infty(g \circ h)(\lambda(X)).$$

The result therefore follows from applying the basic chain rule for subgradients [RW98, Theorem 10.6] to $g \circ h$. QED

An important special case is:

Corollary 5.2. *Let the assumptions of Theorem 5.1 hold, with X and Y also satisfying* (4.1), (4.2) *and* (4.4)*. Suppose that, for $j = 1, \ldots, p$, $\kappa'(\mu_j) \neq 0$ and μ_j is nonderogatory, so that* (4.5) *holds. Define*

$$\sigma_j = \frac{\theta_1^{(j)}}{\kappa'(\mu_j)}, \quad j = 1, \ldots, p \tag{5.3}$$

and

$$\sigma = [\sigma_1, \dots, \sigma_1, \dots, \sigma_p, \dots, \sigma_p]^T, \tag{5.4}$$

each σ_j being repeated $m^{(j)}$ times. Then

$$\sigma \in \partial g(h(\lambda(X))) \subseteq \mathbf{R}^n \quad \text{or} \quad \sigma \in \partial^\infty g(h(\lambda(X))) \subseteq \mathbf{R}^n \tag{5.5}$$

respectively (according to whether Y is a subgradient or a horizon subgradient).

Proof. This is an immediate consequence of Corollary 4.2 and Theorem 5.1. QED

We obtain a similar result for regular subgradients without the nonderogatory assumption:

Theorem 5.3. *Let the assumptions of Theorem 5.1 hold, with X and Y also satisfying* (4.1), (4.2) *and* (4.4) *respectively. Assume also that Y is a* **regular** *subgradient, so that* (4.4) *reduces to* (4.7), (4.8)*. Suppose that $\kappa'(\mu_j) \neq 0$, for $j = 1, \dots, p$, and define σ_j by* (5.3) *and the vector σ by* (5.4)*. Then*

$$\sigma \in \hat{\partial} g(h(\lambda(X))) \subseteq \mathbf{R}^n.$$

Proof. Applying Theorem 3.1 again with $f = g \circ h$, we find

$$\operatorname{diag}(S) \in \hat{\partial}(g \circ h)(\lambda(X)).$$

The result therefore follows from [RW98, Exercise 10.7], together with Corollary 4.4. QED

Theorem 5.3 gives a condition on the *diagonal* components of the matrices $W_{kk}^{(j)}$ in (4.8) that must hold if Y is to be a regular subgradient. We now give an additional necessary condition on the *subdiagonal* components of the $W_{kk}^{(j)}$, again for the case of regular subgradients. Recall that κ is twice continuously differentiable at $\mu \in \mathbf{C}$ in the real sense if its second derivative $\kappa''(\cdot, \nu)$, defined in (1.12), is continuous at μ.

Theorem 5.4. *Let* (5.2) *hold and suppose that κ is twice continuously differentiable at μ_j in the real sense, for $j = 1, \dots, p$. Assume that $\kappa'(\mu_j) \neq 0$, $j = 1, \dots, p$, and suppose also that g is Lipschitz at $h(\lambda(X))$. Let X have the Jordan form* (4.1), (4.2)*, and suppose that Y is a* **regular** *subgradient of $g \circ h \circ \lambda$ at X, so that conditions* (4.7) *and* (4.8) *hold. Then a further necessary condition is that, for each $j = 1, \dots, p$ with $n^{(j)} \geq 2$, we have*

$$\left\langle \theta_2^{(j)}, \kappa'(\mu_j)^2 \right\rangle \geq -\sigma_j \eta_j, \tag{5.6}$$

where

$$\sigma_j = \frac{\theta_1^{(j)}}{\kappa'(\mu_j)} \quad \text{and} \quad \eta_j = \langle \sqrt{-1}\kappa'(\mu_j), \kappa''(\mu_j, \sqrt{-1}\kappa'(\mu_j)) \rangle. \tag{5.7}$$

Proof. First note that σ_j is real from Theorem 5.3, and η_j is real by definition. Suppose that (5.6) does not hold, for some eigenvalue μ_j with $n^{(j)} \geq 2$. Let r be an integer such that the row $r+1$, column r component of the matrix W is in block $W_{kk}^{(j)}$, for some k with $m_k^{(j)} \geq 2$, this component therefore having the value $\theta_2^{(j)}$. Let $Z = PVP^{-1}$ where V has all zero components except

$$\begin{bmatrix} v_{r,r} & v_{r,r+1} \\ v_{r+1,r} & v_{r+1,r+1} \end{bmatrix} = \begin{bmatrix} \delta\kappa'(\mu_j) & 0 \\ -\kappa'(\mu_j)^2 & \delta\kappa'(\mu_j) \end{bmatrix} \tag{5.8}$$

where

$$\delta = -\frac{\eta_j}{2|\kappa'(\mu_j)|^2}. \tag{5.9}$$

Thus

$$\langle Y, Z\rangle = \langle W, V\rangle = -\sigma_j\eta_j - \left\langle \theta_2^{(j)}, \kappa'(\mu_j)^2 \right\rangle, \tag{5.10}$$

which is positive by assumption. The only eigenvalues of $X + tZ$ not equal to a corresponding eigenvalue of X are eigenvalues of the 2 by 2 matrix

$$\begin{bmatrix} \mu_j + t\delta\kappa'(\mu_j) & 1 \\ -t\kappa'(\mu_j)^2 & \mu_j + t\delta\kappa'(\mu_j) \end{bmatrix}.$$

These eigenvalues are

$$\tau_\pm(t) = \mu_j + t\delta\kappa'(\mu) \pm \sqrt{-1}\sqrt{t}\kappa'(\mu_j). \tag{5.11}$$

Since $\kappa'(\mu_j) \neq 0$, we may apply Lemma 1.4, identifying $\sqrt{t}$ with s, to conclude that

$$\kappa(\tau_\pm(t)) = \kappa(\mu_j) + o(t). \tag{5.12}$$

Since g is Lipschitz, we therefore have

$$\liminf_{t\downarrow 0} \frac{(g\circ h\circ\lambda)(X+tZ) - (g\circ h\circ\lambda)(X) - \langle Y, tZ\rangle}{\|tZ\|} = -\frac{\langle Y, Z\rangle}{\|Z\|} < 0,$$

and thus Y is not a regular subgradient of the spectral function $g \circ h \circ \lambda$. QED

Theorem 5.3 and Theorem 5.4 respectively give conditions on the main diagonal and the subdiagonal of W that must hold if the associated matrix Y is a regular subgradient. There is, in general, no restriction on the *lower* subdiagonal components of $W_{kk}^{(j)}$, i.e. $\theta_3^{(j)}, \dots, \theta_{m_k^{(j)}}^{(j)}$. We prove this in the case of the spectral abscissa in Section 7.

6. Spectral Max Functions

An important class of spectral functions consists of those that can be expressed in the form (5.1), where $g : \mathbf{R}^n \to \mathbf{R}$ is the ordinary "max" function. We call these spectral max functions.

Let us define the *active set*

$$\mathcal{A} = \{j : \max(h(\lambda(X))) = \kappa(\mu_j)\}. \tag{6.1}$$

An eigenvalue μ_j is said to be active if $j \in \mathcal{A}$, and inactive otherwise. We now show that if an eigenvalue μ_j is inactive, the block $W^{(j)}$ in (4.4) must be zero. This is obvious for regular subgradients, but to prove this in general we need the following useful tool:

Lemma 6.1. *(Arnold* [Arn71, Theorem 4.4]*) Let X have Jordan form* (4.1), (4.2), *and let $X_i \to X$. Then there exists $P_i \to P$ such that for i sufficiently large,*

$$P_i^{-1}X_iP_i = L_i = \begin{bmatrix} L_i^{(1)} & & \\ & \ddots & \\ & & L_i^{(p)} \end{bmatrix} \tag{6.2}$$

where $L_i^{(j)}$ has dimension $m^{(j)} \times m^{(j)}$.

Since $P_i \to P$, we have $L_i \to J$, but L_i is not, in general, the Jordan form of X_i. This would not be possible because the Jordan form is not continuous. However, the transformation that takes L_i into Jordan form necessarily respects the block diagonal structure in (6.2). Thus, the Jordan form of X_i is displayed by

$$Q_i^{-1} P_i^{-1} X_i P_i Q_i \tag{6.3}$$

where Q_i and Q_i^{-1} do not necessarily converge, but have the same block diagonal structure as (6.2).

We are now ready to prove:

Theorem 6.2. *Let* (5.2) *hold, where* g *is the max function. Define* $\mathcal{A}$ *as in* (6.1). *Let* Y *be a subgradient or a horizon subgradient of* $g \circ h \circ \lambda$ *at* X, *so that* (4.4) *holds. Then, for* $1 \leq j \leq p$,

$$j \notin \mathcal{A} \Rightarrow W^{(j)} = 0. \tag{6.4}$$

Proof. First suppose that Y is a regular subgradient. Suppose also that μ_j is an inactive eigenvalue, i.e. with $j \notin \mathcal{A}$, and that $W^{(j)} \neq 0$. Let

$$Z = PVP^{-1}$$

with V chosen to have one nonzero entry, in its jth diagonal block, in the same position as a nonzero entry of $W^{(j)}$, so that $\langle W, V\rangle$ is positive. Thus, for $t \in \mathbf{R}$ sufficiently small, all eigenvalues of $X+tZ$ are identical to corresponding eigenvalues of X, except the eigenvalues corresponding to μ_j. Therefore, by continuity of eigenvalues, $g \circ h \circ \lambda$ is identical at $X + tZ$ and X, for sufficiently small t. This yields a contradiction of the form (4.10).

Now suppose that Y is any subgradient, so that there is a sequence $X_i \to X$ and $Y_i \to Y$ with

$$Y_i \in \hat{\partial}(g \circ h \circ \lambda)(X_i).$$

By Lemma 6.1, there exists $P_i \to P$ such that (6.2) holds. Since the Jordan form of X_i has the block diagonal form (6.3), Theorem 4.3 shows that

$$W_i = Q_i^* P_i^* Y_i P_i^{-*} Q_i^{-*}$$

has a block diagonal structure that respects the block diagonal structure shown in (6.2). Now suppose μ_j is not an active eigenvalue of X. By eigenvalue continuity, the eigenvalues of $L_i^{(j)}$ cannot be active eigenvalues of X_i for i sufficiently large. Therefore, since Y_i is regular, the corresponding block $W_i^{(j)}$ must be zero, for i sufficiently large. Since W_i and Q_i both have a block diagonal structure consistent with (6.2), and since

$$Q_i^{-*} W_i Q_i^* = P_i^* Y_i P_i^{-*} \to W,$$

it follows that $W^{(j)} = 0$. The proof for horizon subgradients is identical. QED

We now consider how the results of the previous section specialize to the case of spectral max functions. The max function g is convex, so all its subgradients are regular, and its only horizon subgradient is zero. Using the well known formula for the subgradients of g, we therefore have, under the assumptions of Corollary 5.2 (where we assume that all μ_j are nonderogatory, in the case where Y is a subgradient) or Theorem 5.3 (where we assume that Y is a regular subgradient), that the σ_j defined in (5.3) satisfy

$$\sigma_j \in \mathbf{R}, \quad \sigma_j \geq 0, \quad \sum_{j \in \mathcal{A}} m^{(j)} \sigma_j = 1, \tag{6.5}$$

with

$$\sigma_j = 0, \quad \text{for } j \notin \mathcal{A}. \tag{6.6}$$

(In fact, (6.6) is a consequence of (5.3) and (6.4).) In particular, if $\mathcal{A}$ contains only one index, say j, then

$$\sigma_j = \frac{1}{m^{(j)}}.$$

If we further assume that this eigenvalue μ_j is simple (and therefore nonderogatory) the only possible value for Y is

$$Y = \kappa'(\mu_j) u v^*,$$

where v is the column of P associated with μ_j (a right eigenvector of X) and u^* is the row of P^{-1} associated with μ_j (a left eigenvector of X); note that $u^* v = 1$. In this case, $g \circ h \circ \lambda$ is differentiable at X.

Likewise, when Y is a horizon subgradient of a spectral max function, the assumptions of Corollary 5.2 imply that, instead of (6.5), we have

$$\sigma_j = 0, \quad j = 1, \ldots, p. \tag{6.7}$$

This follows because zero is the only horizon subgradient of the max function.

We can be more specific: if $\kappa = \mathrm{Re}$, so that $g \circ h \circ \lambda$ is the spectral abscissa, then $\kappa'(\mu_j) = 1$, and so $\theta_1^{(j)} = \sigma_j, \quad j = 1, \ldots, p$. In this case, (6.5) reduces to

$$\theta_1^{(j)} \in \mathbf{R}, \quad \theta_1^{(j)} \geq 0, \quad \sum_{j \in \mathcal{A}} m^{(j)} \theta_1^{(j)} = 1. \tag{6.8}$$

and (6.7) reduces to

$$\theta_1^{(j)} = 0, \quad j = 1, \ldots, p. \tag{6.9}$$

On the other hand, if $\kappa = \mathrm{mod}$, so that $g \circ h \circ \lambda$ is the spectral radius, then $\kappa'(\mu_j) = \mu_j / |\mu_j|$, so

$$\theta_1^{(j)} = \frac{\sigma_j \mu_j}{|\mu_j|}, \quad j = 1, \ldots, p.$$

Strictly speaking, in the spectral radius case, Corollary 5.2 and Theorem 5.3 do not apply if any eigenvalue μ_j is zero; however, in view of Theorem 6.2, it is easy to extend them to cover this case as long as at least one eigenvalue is nonzero. The spectral radius case where all eigenvalues are zero is exceptional.

Now let us turn to Theorem 5.4. In the spectral abscissa case, with $\kappa = \mathrm{Re}$, we have $\kappa'(\mu) = 1$ and $\kappa''(\mu, \nu) = 0$, so condition (5.6) reduces to

$$\mathrm{Re}\ \theta_2^{(j)} \geq 0. \tag{6.10}$$

(In this case, the proof of Theorem 5.4 simplifies considerably, since $\tau_\pm(t) - \mu_j$ are imaginary and therefore Lemma 1.4 is not needed.) In the spectral radius case, where $\kappa = \mathrm{mod}$, we have, for $\mu_j \neq 0$, $\kappa'(\mu_j) = \mu_j / |\mu_j|$ and

$$\kappa''(\mu_j, \frac{\sqrt{-1}\mu_j}{|\mu_j|}) = \frac{-(\mathrm{Im}\ \mu_j)^3 - (\mathrm{Re}\ \mu_j)^2 (\mathrm{Im}\ \mu_j) + \sqrt{-1}(\mathrm{Im}\ \mu_j)^2 (\mathrm{Re}\ \mu_j) + \sqrt{-1}(\mathrm{Re}\ \mu_j)^3}{|\mu_j|^4},$$

so $\eta_j = 1/|\mu_j|$, and condition (5.6) reduces to

$$\left\langle \theta_2^{(j)}, \mu_j^2 \right\rangle \geq -\sigma_j |\mu_j|.$$

7. The Regular Subgradients of the Spectral Abscissa

In this section we specialize the discussion further to the spectral abscissa

$$\alpha = g \circ h \circ \lambda, \tag{7.1}$$

where g is the "max" function and h maps the eigenvalues to their real parts, i.e. κ in (5.2) is the function Re. With this choice of spectral function, the active set of eigenvalues at X is given by

$$\mathcal{A} = \{j : \alpha(X) = \mathrm{Re}\ \mu_j\}. \tag{7.2}$$

We shall show that the necessary conditions, derived in the previous sections, for Y to be a regular subgradient of α at X, are also sufficient conditions; that is, these conditions completely characterize $\hat{\partial}\alpha(X)$.

Let $\mathcal{P}_n$ denote the space of polynomials in ζ of degree n or less. Define the *abscissa* of a monic polynomial $p \in \mathcal{P}_n$ to be the maximum of the real parts of its roots:

$$a(p) = \max\{\mathrm{Re}\ \zeta\ :\ p(\zeta) = 0\},$$

and extend the definition of a to the linear space $\mathcal{P}_n$ by defining it to be ∞ for polynomials which are not monic. The spectral abscissa of a matrix is the abscissa of its characteristic polynomial, i.e.,

$$\alpha(X) = a(\det(\zeta I - X)).$$

Before we state the main theorem about regular subgradients of the spectral abscissa, we need two key results. The first of these concerns the characteristic polynomial of a matrix depending on a single parameter.

Lemma 7.1. *Let $A(t)$ denote a continuous curve in $\mathbf{M}^n$ satisfying*

$$A(t) = X + tZ + o(t), \tag{7.3}$$

so that $A \in \Gamma(X, Z)$, where Γ is defined in (1.10)*. By the derivative, with respect to t, of a polynomial whose coefficients are functions of t, we mean the polynomial obtained by differentiating each coefficient. Define the polynomials $p(\zeta)$ and $q(\zeta)$ by*

$$p(\zeta) = \det(\zeta I - A(t))|_{t=0} = \det(\zeta I - X) = \prod_{j=1}^{p} (\zeta - \mu_j)^{m^{(j)}}, \tag{7.4}$$

$$q(\zeta) = \frac{d}{dt} \det(\zeta I - A(t)) \bigg|_{t=0}. \tag{7.5}$$

Let X have Jordan form (4.1), (4.2)*, let*

$$V = P^{-1} Z P$$

and let $V^{(jj)}$ be the $m^{(j)} \times m^{(j)}$ diagonal block of V corresponding to the block $J^{(j)}$ of J. (Note that V is not necessarily block diagonal.) Then

$$q(\zeta) = -\sum_{j=1}^{p} \left[\prod_{\substack{k=1 \\ k \neq j}}^{p} (\zeta - \mu_k)^{m^{(k)}} \right] \left[\sum_{\ell=1}^{n^{(j)}} \mathrm{tr} \left(\left(N^{(j)} \right)^{\ell-1} V^{(jj)} \right) (\zeta - \mu_j)^{m^{(j)} - \ell} \right].$$

where $N^{(j)}$ *is defined in* (4.3).

Proof. The determinant of a matrix is a differentiable, complex-valued spectral function whose derivative is well known. For a smooth matrix function $M : \mathbf{R} \to \mathbf{M}^n$, we have[1]

$$\frac{d}{dt} \det\left(M(t)\right)\bigg|_{t=0} = \det(M(0)) \operatorname{tr}\left((M(0))^{-1} \frac{d}{dt} M(t)\bigg|_{t=0} \right)$$

as long as $M(0)$ is nonsingular [Lax97, Chapter 9]. Since the derivative we are evaluating is a polynomial in ζ, we may assume ζ is not an eigenvalue of X without loss of generality. Therefore, we obtain

$$\begin{aligned} q(\zeta) &= -p(\zeta) \operatorname{tr} \left(\zeta I - X\right)^{-1} Z &(7.6)\\ &= -p(\zeta) \operatorname{tr} P\left(\zeta I - J\right)^{-1} P^{-1} Z &(7.7)\\ &= -p(\zeta) \operatorname{tr} \left(\zeta I - J\right)^{-1} V &(7.8)\\ &= -p(\zeta) \sum_{j=1}^{p} \operatorname{tr} \left((\zeta - \mu_j) I - N^{(j)} \right)^{-1} V^{(jj)}. &(7.9) \end{aligned}$$

The proof is completed by using (7.4) and noting that

$$\left(I - \gamma N^{(j)}\right)^{-1} = I + \gamma N^{(j)} + \cdots + \gamma^{n^{(j)}-1} \left(N^{(j)}\right)^{n^{(j)}-1}$$

for any scalar γ, since $(N^{(j)})^{n^{(j)}} = 0$. QED

The next result concerns the subderivative of the abscissa of the characteristic polynomial. Recall that the subderivative of a function was defined in (1.8).

Theorem 7.2. *Let* $p(\zeta)$ *and* $q(\zeta)$ *be defined as in Lemma 7.1. Then*

$$da(p(\zeta))(q(\zeta)) = \infty \tag{7.10}$$

if any of the following conditions is violated for any $j \in \mathcal{A}$*:*

$$\operatorname{Re} \operatorname{tr} \left(N^{(j)} V^{(jj)}\right) \leq 0, \quad \operatorname{Im} \operatorname{tr} \left(N^{(j)} V^{(jj)}\right) = 0, \tag{7.11}$$

$$\operatorname{tr} \left(\left(N^{(j)}\right)^{\ell} V^{(jj)} \right) = 0, \quad \ell = 2, \ldots, n^{(j)} - 1. \tag{7.12}$$

On the other hand, if (7.11) *and* (7.12) *hold for all* $j \in \mathcal{A}$*, then*

$$da(p(\zeta))(q(\zeta)) = \max \left\{ \frac{\operatorname{Re} \operatorname{tr} V^{(jj)}}{m^{(j)}} : j \in \mathcal{A} \right\}. \tag{7.13}$$

Proof. The proof is a consequence of [BO99, Corollary 1.7], using Lemma 7.1. QED

We are now in a position to present the main result of this section.

Theorem 7.3. *Let* X *have Jordan form* (4.1), (4.2). *Then* $\hat{\partial}\alpha(X)$*, the set of regular subgradients of the spectral abscissa* α *at* X*, is the set of matrices* Y *satisfying* (4.7), (4.8), (6.4), (6.8) *and* (6.10).

[1] Equivalently, via the ordinary chain rule, the complex gradient of $\det(M)$ is $(\det(M)M^{-1})^*$, for any nonsingular matrix M.

Proof. That these conditions are necessary for Y to be a regular subgradient was proved in Theorem 4.3, Theorem 5.3, Theorem 5.4, and Theorem 6.2. Now suppose that Y satisfies these conditions. We must prove that Y is a regular subgradient.

By Corollary 1.2, to show that Y is a regular subgradient of α at X, it is sufficient to show that

$$\langle Y, Z\rangle \leq d(\alpha \circ A)(0)(1), \tag{7.14}$$

for all matrix curves of the form (7.3) where $\|Z\| = 1$. We have

$$\begin{aligned} d\left(\alpha \circ A\right)(0)(1) &= \liminf_{t \downarrow 0} \frac{\alpha\left(A(t)\right) - \alpha\left(A(0)\right)}{t} \\ &= \liminf_{t \downarrow 0} \frac{a\left(\det\left(\zeta I - A(t)\right)\right) - a\left(\det\left(\zeta I - A(0)\right)\right)}{t} \\ &= \liminf_{t \downarrow 0} \frac{a\left(p(\zeta) + tq(\zeta) + o(t)\right) - a\left(p(\zeta)\right)}{t} \qquad (7.15) \\ &\geq da\left(p(\zeta)\right)\left(q(\zeta)\right). \qquad (7.16) \end{aligned}$$

Here (7.15) follows from the definitions of $p(\zeta)$ and $q(\zeta)$ in (7.4) and (7.5), and (7.16) is a consequence of Lemma 1.1.

Using (4.7), (4.8), or equivalently (4.9), as well as (6.4), we have

$$\langle Y, Z\rangle = \langle W, V\rangle = \sum_{j\in\mathcal{A}} \sum_{\ell=1}^{n^{(j)}} \mathrm{Re}\left(\overline{\theta_\ell^{(j)}}\ \mathrm{tr}\left(N^{(j)}\right)^{\ell-1} V^{(jj)}\right). \tag{7.17}$$

If any of (7.11), (7.12) are violated, then (7.10) must hold, and so (7.16) shows that (7.14) holds trivially. On the other hand, suppose that (7.11), (7.12) hold for all $j \in \mathcal{A}$, implying that (7.13) holds. Using these conditions, together with (7.17), (6.8) and (6.10), we have

$$\begin{aligned} \langle Y, Z\rangle &= \sum_{j\in\mathcal{A}} \left(\sigma_j \mathrm{Re\ tr}\ V^{(jj)} + \mathrm{Re}\ \theta_2^{(j)} \mathrm{Re\ tr}\left(N^{(j)} V^{(jj)}\right)\right) \\ &\leq \sum_{j\in\mathcal{A}} \sigma_j \mathrm{Re\ tr}\ V^{(jj)} \\ &\leq da(p(\zeta))(q(\zeta)). \end{aligned}$$

Combining this with (7.16) gives (7.14), as desired. QED

It follows from Theorem 7.3 that if Y is a regular subgradient of α at X and μ_j is semisimple, then $W^{(j)}$ must be a multiple of I. If all active eigenvalues of X are semisimple, W must be diagonal. In particular, we have:

Corollary 7.4. *Suppose X is the n by n zero matrix. Then the spectral abscissa α has only one regular subgradient at X. Specifically,*

$$\hat{\partial}\alpha(0) = \left\{\frac{1}{n} I\right\}.$$

This result was to some extent anticipated in [OW88, Theorem 4.3], though the result there is weaker and stated in the spectral radius context (for a nonzero semisimple eigenvalue). This stands in marked contrast to the well known result for the Hermitian case: see Section 9.

If at least one of the active eigenvalues of X is not semisimple, i.e. has a Jordan block of order greater than one, $\hat{\partial}\alpha(X)$ is unbounded, since there is no restriction on the values of $\theta_2^{(j)}, \ldots, \theta_{n^{(j)}}^{(j)}$ for $j \in \mathcal{A}$, except Re $\theta_2^{(j)} \geq 0$. More specifically, we have:

Corollary 7.5. *Let X have Jordan form* (4.1), (4.2) *for some P. Then $\hat{\partial}\alpha(X)^\infty$, the horizon cone of $\hat{\partial}\alpha(X)$, is the set of matrices Y satisfying* (4.7), (4.8), (6.4), (6.9) *and* (6.10). *If all active eigenvalues of X are semisimple, the only matrix in this set is $Y = 0$; on the other hand, if at least one active eigenvalue of X is not semisimple, $\hat{\partial}\alpha(X)^\infty$ is unbounded.*

Proof. The proof follows immediately from the definition of the horizon cone in (1.7). QED

8. The Subgradients and Horizon Subgradients of the Spectral Abscissa

In this section we consider *all* subgradients and horizon subgradients of the spectral abscissa α, giving complete characterizations in the nonderogatory and semisimple cases.

We begin with a corollary of results proved earlier:

Corollary 8.1. *Let X satisfy* (4.1), (4.2) *for some P. A necessary condition for Y to be a subgradient of the spectral abscissa α at X is that* (4.4) *and* (6.4) *hold, where the eigenvalues of Y (equivalently of W) are all real, nonnegative, and sum to one. Furthermore, a necessary condition for Y to be a horizon subgradient of α at X is that* (4.4) *and* (6.4) *hold, and that Y (equivalently W) is nilpotent (all its eigenvalues are zero).*

Proof. The first statement follows from Corollary 4.2, Theorem 5.1 and Theorem 6.2, using the subdifferential of the max function. The second follows in the same way, since the only horizon subgradient of the max function is zero. QED

To go further, we consider two cases separately: (1) all active eigenvalues of X are *nonderogatory*, and (2) all active eigenvalues are *semisimple*.

In the nonderogatory case, Corollary 5.2 shows that if Y is a subgradient or horizon subgradient, satisfying (4.4) and (4.5), then (6.8) must hold, thus characterizing the diagonal components of $W^{(j)}$. We now turn our attention to the subdiagonal condition (6.10), showing that it applies to all subgradients and horizon subgradients, not just regular subgradients, under the nonderogatory assumption. We conjecture that Theorem 5.4 can be extended in this way for all spectral functions of the form $g \circ h \circ \lambda$, but, to avoid unnecessary complication, we generalize it only for the spectral abscissa, with active set defined by (7.2). Even in this case, the proof depends on machinery not needed by any of our other results; despite some effort, we have not succeeded in finding a simpler proof.

Theorem 8.2. *Let X have Jordan form* (4.1), (4.2), *and suppose that all active eigenvalues of X are nonderogatory. Let Y be a subgradient or horizon subgradient of α at X, satisfying* (4.4) *and* (4.5). *Then* (6.10) *holds for all j with $m^{(j)} \geq 2$.*

Proof. First suppose that Y is a subgradient. Then there exist sequences $X_i \to X$ and $Y_i \in \hat{\partial}\alpha(X_i)$ with $Y_i \to Y$. We wish to show that $\theta_2^{(j)}$, the subdiagonal entry in the Toeplitz matrix $W^{(j)}$, satisfies (6.10) for all eigenvalues μ_j with $m^{(j)} \geq 2$: suppose that this is not the case for some j. Let r be an integer such that the row

$r+1$, column r component of W is in block $W^{(j)}$, this component therefore having the value $\theta_2^{(j)}$.

By [Arn71, Theorem 4.4], there exists a sequence $P_i \to P$ such that (6.2) holds. Consider the sequence $P_i^{-1}X_iP_i \to J$. By applying Lemma B.2 in Appendix B to each diagonal block of $P_i^{-1}X_iP_i$ separately, we see that there exists a sequence of unitary matrices $U_i \to I$ such that

$$U_i^* P_i^{-1} X_i P_i U_i = T_i \to J,$$

where T_i is upper triangular for all i. Thus the eigenvalues of X_i appear on the diagonal of T_i. Furthermore, there exists a sequence of diagonal matrices $D_i \to I$, each differing from I only in the rth diagonal position, such that

$$D_i^{-1} T_i D_i = \tilde{T}_i,$$

with the row r, column $r+1$ component of $\tilde{T}_i$ exactly equal to one; this is possible because the corresponding entry in J is one. Let

$$Z_i = P_i U_i D_i V D_i^{-1} U_i^* P_i^{-1},$$

where V has all zero components except that the row $r+1$, column r entry is -1. We have

$$\langle Y_i, Z_i \rangle \to \left\langle P^* Y P^{-*}, V \right\rangle = \langle W, V \rangle = -\mathrm{Re}\ \theta_2^{(j)} > 0, \tag{8.1}$$

the inequality holding by our hypothesis that (6.10) does not hold.

Now consider the eigenvalues of $X_i + tZ_i$, where $t \in \mathbf{R}^+$. Since

$$X_i + tZ_i = P_i U_i D_i (\tilde{T}_i + tV) D_i^{-1} U_i^* P_i^{-1},$$

the only eigenvalues of $X_i + tZ_i$ not equal to a corresponding eigenvalue of X_i are eigenvalues of the 2 by 2 matrix

$$\begin{bmatrix} \nu_i^{(1)} & 1 \\ -t & \nu_i^{(2)} \end{bmatrix},$$

where $\nu_i^{(1)}$ and $\nu_i^{(2)}$ are respectively the rth and $(r+1)$th diagonal entries of $\tilde{T}_i$, which are eigenvalues of X_i. These eigenvalues are

$$\tau_{\pm}(t) = \frac{\nu_i^{(1)} + \nu_i^{(2)}}{2} \pm \frac{1}{2}\sqrt{\left(\nu_i^{(1)} - \nu_i^{(2)}\right)^2 - 4t}.$$

By considering a subsequence if necessary, we may assume that either (1) $\nu_i^{(1)} = \nu_i^{(2)}$ for all i, or (2) $\nu_i^{(1)} \neq \nu_i^{(2)}$ for all i. In the first case,

$$\mathrm{Re}\ \tau_{\pm}(t) = \nu_i^{(1)}$$

for all i, and hence the spectral abscissa difference quotient

$$\alpha(X_i + tZ_i) - \alpha(X_i)$$

is zero for all $t \geq 0$ and all i. In the second case, we have

$$\tau_{\pm}(t) = \frac{\nu_i^{(1)} + \nu_i^{(2)}}{2} \pm \left(\frac{\nu_i^{(1)} - \nu_i^{(2)}}{2} - \frac{t}{\nu_i^{(1)} - \nu_i^{(2)}} \right) + o(t).$$

Suppose without loss of generality that Re $\nu_i^{(1)} \geq$ Re $\nu_i^{(2)}$. Then the maximum of the real parts of the two eigenvalues $\tau_\pm(t)$ is

$$\text{Re } \nu_i^{(1)} - t \, \frac{\text{Re } \nu_i^{(1)} - \text{Re } \nu_i^{(2)}}{\left|\nu_i^{(1)} - \nu_i^{(2)}\right|^2} + o(t).$$

Consequently, in both cases (1) and (2), we have

$$\liminf_{t \downarrow 0} \frac{\alpha(X_i + tZ_i) - \alpha(X_i)}{t} \leq 0$$

and therefore, using (8.1) and choosing i sufficiently large,

$$\liminf_{t \downarrow 0} \frac{\alpha(X_i + tZ_i) - \alpha(X_i) - \langle Y_i, tZ_i\rangle}{\|tZ_i\|} < 0.$$

Thus, Y_i is not a regular subgradient of α at X_i, and we have our desired contradiction.

When Y is a horizon subgradient, the proof is almost identical. Instead of $Y_i \to Y$ we have $s_i Y_i \to Y$, with $s_i \downarrow 0$, so $\langle Y_i, Z_i\rangle$ in (8.1) must be multiplied by s_i. The contradiction is then obtained exactly as before. QED

We are now ready for the main result of this section, characterizing regularity of the spectral abscissa. Recall from Section 1 that a function is subdifferentially regular at X if all its subgradients at X are regular and all its horizon subgradients at X are contained in the horizon cone of the set of regular subgradients.

Theorem 8.3. *The spectral abscissa α is subdifferentially regular at X if and only if all active eigenvalues of X are nonderogatory.*

Proof. First suppose that all active eigenvalues of X are nonderogatory. Let X have Jordan form (4.1), (4.2), and let Y be a subgradient of α at X. Since all active eigenvalues are nonderogatory, W in (4.4) satisfies (4.5) as well as (6.4). Corollary 5.2 shows that (6.8) must hold, and Theorem 8.2 shows that (6.10) must also be satisfied. Furthermore, Theorem 7.3 tells us that the conditions just described are exactly those characterizing regular subgradients, so Y must be regular. This proves $\partial\alpha(X) = \hat{\partial}\alpha(X)$. If Y is a horizon subgradient, the same conditions hold except that, instead of (6.8), we have (6.9). Thus $\partial^\infty\alpha(X) = \hat{\partial}\alpha(X)^\infty$ (see Corollary 7.5), and subdifferential regularity is proved.

For the converse, suppose that X has an active derogatory eigenvalue μ_j, so that $q^{(j)} \geq 2$. Let β_i be a real, positive sequence converging to zero, and define

$$X_i = P(J + \beta_i E)P^{-1} \tag{8.2}$$

where E is zero except in the $m_1^{(j)}$ diagonal positions corresponding to the Jordan block $J_1^{(j)}$, where the entries in E are one. The matrix X_i has only one active eigenvalue, namely $\mu_j + \beta_i$, with multiplicity $m_1^{(j)}$, and the expression on the right-hand side of (8.2) is the Jordan form of X_i (although the diagonal entries of $J + \beta_i E$ may not be lexicographically ordered). Consequently, from Theorem 7.3, the regular subgradients of α at X_i include the matrix

$$\tilde{E} = \frac{1}{m_1^{(j)}} P^{-*} E P^*.$$

Since this remains true for all $\beta_i > 0$, $\tilde{E}$ is a subgradient of α at X. However, $\tilde{E}$ is not a regular subgradient of α at X, because it does not satisfy (4.7), (4.8). (It satisfies the block partitioning requirement, but not the condition that the diagonal entries be the same across all blocks corresponding to μ_j.) Therefore, α is not subdifferentially regular at X. QED

Example. Let X be the Jordan block

$$X = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Theorem 7.3 shows that

$$\hat{\partial}\alpha(X) = \left\{ \begin{bmatrix} 1/2 & 0 \\ \tau & 1/2 \end{bmatrix} : \mathrm{Re}\ \tau \geq 0 \right\}.$$

and therefore, by definition,

$$\hat{\partial}\alpha(X)^\infty = \left\{ \begin{bmatrix} 0 & 0 \\ \tau & 0 \end{bmatrix} : \mathrm{Re}\ \tau \geq 0 \right\}.$$

It is instructive to consider a specific sequence

$$X_i = \begin{bmatrix} \epsilon_i\sqrt{-1} & 1 \\ 0 & -\epsilon_i\sqrt{-1} \end{bmatrix} \to X$$

where $\epsilon_i > 0$, $\epsilon_i \to 0$ as $i \to \infty$. The matrix X_i has distinct eigenvalues, so its Jordan form is diagonal, and does not converge to J. Both eigenvalues are active for all $\epsilon_i > 0$. It is easily verified, using the Jordan form of X_i and Theorem 7.3, that

$$\hat{\partial}\alpha(X_i) = \left\{ \begin{bmatrix} \sigma & 0 \\ (\sigma - \frac{1}{2})\sqrt{-1}/\epsilon_i & 1 - \sigma \end{bmatrix} : \quad \sigma \in [0,1] \right\}. \tag{8.3}$$

Let $Y_i \in \hat{\partial}\alpha(X_i)$, with σ_i being the corresponding value of σ in (8.3). If $Y_i \to Y$, then its bottom left entry must converge to a limit: any imaginary limit is possible, but since $\epsilon_i \to 0$, we must have $\sigma_i \to \frac{1}{2}$. Thus, the subgradient Y is regular (an element of $\hat{\partial}\alpha(X)$). On the other hand, if $s_i Y_i \to Y$, with $s_i \downarrow 0$, then the diagonal entries of Y are both zero (Y is nilpotent), i.e., the horizon subgradient Y is an element of the recession cone $\hat{\partial}\alpha(X)^\infty$. Theorem 8.3 shows that these properties hold for *every* sequence $X_i \to X$, i.e., α is subdifferentially regular at X.

Theorem 8.3 also demonstrates that $f \circ \lambda$ may not be subdifferentially regular at X, even if f is subdifferentially regular at $\lambda(X)$, as is the case for the convex function $f = \max\ \mathrm{Re}$. This is in contrast with the Hermitian case discussed in [Lew99].

We now turn to semisimple eigenvalues, for which the subdifferential properties of the spectral abscissa are quite different from the nonderogatory case.

Theorem 8.4. *Let X have Jordan form* (4.1), (4.2) *for some P, and suppose that all active eigenvalues of X are semisimple, so that $J^{(j)} = \mu_j I$, for all $j \in \mathcal{A}$. Then the set of subgradients of the spectral abscissa at X consists of those matrices Y satisfying*

$$P^*YP^{-*} = W = \begin{bmatrix} W^{(1)} & & \\ & \ddots & \\ & & W^{(p)} \end{bmatrix},$$

each $W^{(j)}$ being $m^{(j)} \times m^{(j)}$, where $W^{(j)} = 0$ if $j \notin \mathcal{A}$, and the eigenvalues of Y (equivalently of W) are all real, nonnegative, and sum to one. Furthermore, the set of horizon subgradients of α at X consists of the matrices Y satisfying the same block structure, but such that Y (equivalently W) is nilpotent (all its eigenvalues are zero).

Proof. That the conditions stated here are necessary for Y to be a subgradient has already been established in Corollary 8.1; since all active eigenvalues are semisimple, the nonzero $W_{rs}^{(j)}$ in (4.4) are all scalars and hence are trivially Toeplitz. We need to prove that the given conditions are also sufficient. Let Y satisfy these conditions. The matrix $W = P^*YP^{-*}$ is block diagonal by assumption, and although it may not be diagonalizable, there exists a sequence $W_i \to W$ with W_i diagonalizable, say

$$W_i = T_i^{-*} D_i T_i^*,$$

where T_i is block diagonal and D_i is diagonal. By scaling and shifting D_i, equivalently W_i, we may assume that the diagonal entries of D_i are real, nonnegative and sum to one, without changing the limit W. Now define

$$X_i = PT_i\,(J + K_i)\,T_i^{-1}P^{-1},$$

where K_i is diagonal with distinct diagonal entries, all having the same real part, and converging to zero. Since the blocks of J are multiples of the identity and T_i is block diagonal, $X_i \to X$ if K_i is chosen to converge to zero sufficiently fast (relative to $\|T_i\|\|T_i^{-1}\|$). Thus, by Theorem 7.3,

$$Y_i = P^{-*}W_iP^* = P^{-*}T_i^{-*}D_iT_i^*P^* \tag{8.4}$$

is a regular subgradient of α at X_i, for all i. Since $Y_i \to Y$, it follows that Y is a subgradient of α at X. The proof for the horizon subgradients is almost identical: now the eigenvalues of W are zero, so $D_i \to 0$, but we can assume its entries are real, nonnegative, and sum to s_i, with $s_i \downarrow 0$. The left-hand side of (8.4) must then be multiplied by s_i; then Y_i is a regular subgradient as before, and since $s_iY_i \to Y$, the latter is a horizon subgradient of α at X. QED

In particular, we have:

Corollary 8.5. *Suppose X is the n by n zero matrix. Then the set of subgradients of the spectral abscissa α at X is the set of all matrices whose eigenvalues are real, nonnegative and sum to one, and the set of horizon subgradients is the set of all nilpotent matrices.*

Note that nonzero horizon subgradients arise in both the nonderogatory and semisimple cases, if any active eigenvalue of X has multiplicity greater than one.

9. The Hermitian Case

Let $\mathbf{H}^n$ denote the Euclidean space of $n \times n$ Hermitian matrices, i.e. those matrices X satisfying $X^* = X$. It is well known that the eigenvalues of $X \in \mathbf{H}^n$ are real and that the eigenvalue map λ is Lipschitz on $\mathbf{H}^n$ (see e.g. [Kat84, Theorem II.6.10]). Variational properties of λ on $\mathbf{H}^n$ have been extensively studied, especially in the recent work of Lewis [Lew99]. Indeed, the general results given here in Sections 2 and 3 are direct extensions of Lewis' results. In this section we make some further remarks about how the results given above specialize in the Hermitian case.

Let $X = X^*$. Then the Schur form of Section 2 and the Jordan form of Section 4 are the same. More specifically, the following is true:

- the eigenvalues μ_j, $j = 1, \ldots, p$, are real and semisimple, i.e. all $m_k^{(j)}$ equal one
- any unitary matrix U transforming X to Schur form R also transforms X to Jordan form J; we may therefore assume without loss of generality that P in (4.1) is unitary
- the Schur form R and the Jordan form J are the same, namely the diagonal matrix $\mathrm{Diag}(\lambda(X))$

Furthermore, since all eigenvalues are real, the spectral abscissa of X is the maximum eigenvalue μ_1, and hence the active set $\mathcal{A}$ defined in (7.2) is $\{1\}$; i.e., only μ_1 is active.

Let

$$\omega \,:\, \mathbf{H}^n \to \mathbf{R}$$

be the maximum eigenvalue function on $\mathbf{H}^n$. It is well known that ω is convex. Let us define a set $\mathcal{Y}$, depending on another set $\mathcal{W}$, by

$$\mathcal{Y}(\mathcal{W}) = \Bigg\{ Y \quad : \quad P^*YP = \begin{bmatrix} W^{(1)} & & \\ & \ddots & \\ & & W^{(p)} \end{bmatrix}, \qquad W^{(1)} \in \mathcal{W}, \quad W^{(j)} = 0 \text{ for } j = 2, \ldots, p \Bigg\}. \tag{9.1}$$

where, as earlier, each $W^{(j)}$ is $m^{(j)} \times m^{(j)}$. Then, as is well known, e.g. [Lew96a], we have for all $X \in \mathbf{H}^n$,

$$\hat{\partial}\omega(X) = \partial\omega(X) = \mathcal{Y}(\mathcal{W}), \quad \partial^\infty\omega(X) = \{0\}\,, \tag{9.2}$$

where $\mathcal{W}$ is the set of $m^{(1)} \times m^{(1)}$ positive semidefinite Hermitian matrices with trace one.

It is instructive to investigate whether (9.2) can be recovered from our characterization of the subgradients of the spectral abscissa α defined on $\mathbf{M}^n$. Let $\hat{\mathbf{H}}_n$ be the subspace of $\mathbf{M}^n$ consisting of all Hermitian matrices, and let $\iota \,:\, \mathbf{H}^n \to \mathbf{M}^n$ be the canonical embedding of $\mathbf{H}^n$ into $\mathbf{M}^n$, so that

$$\iota(\mathbf{H}^n) = \hat{\mathbf{H}}_n.$$

It is straightforward to show that the adjoint of ι is the linear operator which maps a matrix to its Hermitian part, i.e.

$$\iota^* Z = \frac{1}{2}\left(Z + Z^*\right), \quad \text{for } Z \in \mathbf{M}^n. \tag{9.3}$$

Since

$$\omega = \alpha \circ \iota \,:\, \mathbf{H}^n \to \mathbf{R}, \tag{9.4}$$

we have

$$\hat{\partial}\omega(X) = \partial\omega(X) = \hat{\partial}(\alpha \circ \iota)(X) = \partial(\alpha \circ \iota)(X) \tag{9.5}$$

for all $X \in \mathbf{H}^n$.

Now let $X \in \mathbf{M}^n$ satisfy $X = X^*$. Since all eigenvalues of X are semisimple, Theorem 8.4 shows that

$$\partial\alpha(X) = \mathcal{Y}(\tilde{\mathcal{W}}) \tag{9.6}$$

where $\tilde{\mathcal{W}}$ is the set of all (not necessarily Hermitian) $m^{(1)} \times m^{(1)}$ matrices whose eigenvalues are real, nonnegative, and sum to one. However, from Theorem 7.3, only one subgradient in $\partial\alpha(X)$ is regular, namely, when $W^{(1)}$ is a multiple of the identity. Thus

$$\hat{\partial}\alpha(X) = \mathcal{Y}\left(\left\{\frac{1}{m^{(1)}}I\right\}\right). \tag{9.7}$$

We now apply the basic chain rule of [RW98, Theorem 10.6] to the composition $\alpha \circ \iota$. Since X is Hermitian, we obtain

$$\hat{\partial}(\alpha \circ \iota)(X) \supseteq \iota^* \hat{\partial}\alpha(X) \tag{9.8}$$

and

$$\partial(\alpha \circ \iota)(X) \subseteq \iota^* \partial\alpha(X). \tag{9.9}$$

Comparing (9.2) with (9.7) and using (9.5) shows that, in fact, the inclusion (9.8) is strict. On the other hand, comparison of (9.2) and (9.6) shows that the sets on the left and right-hand side of (9.9) are the same (using (9.3)). Because α is not regular at X (unless $m^{(1)} = 1$), this equality condition could not be concluded from the chain rule. This suggests that a version of the chain rule with weaker hypotheses could be useful in this context. Similar remarks hold for the chain rule for the set of horizon subgradients.

10. Semistable Programming

We conclude by giving an example of an important class of optimization problems that can be treated by our analysis. Consider the problem:

$$\begin{aligned} \min_{X \in \mathbf{M}^n} \quad & \langle C, X\rangle && (10.1)\\ \text{subject to} \quad & \langle A_k, X\rangle = b_k, \quad k = 1, \ldots, m \\ \text{and} \quad & \alpha(-X) \le 0, && (10.2) \end{aligned}$$

where $C \in \mathbf{M}^n$, $A_k \in \mathbf{M}^n$, $k = 1, \ldots, m$, and $b \in \mathbf{R}^m$. We call this a *semistable program*. The second constraint imposes the condition that the spectral abscissa of $-X$ is nonpositive, or in other words that all eigenvalues of X lie in the right-half plane or on the imaginary axis. We call such matrices *semistable*. Semistable programs have many potential applications in stability and control theory. The choice of the right-half plane instead of the left-half plane is purely for notational reasons; in applications, stability is usually associated with the left-half plane. If the domain of the semistable program is restricted to the Hermitian matrices, the problem reduces to a semidefinite program.

Semistable programs are, of course, not convex, and it is known that finding the global minimum is NP-hard [BT97, Nem93]. However, *local* optimality conditions may be addressed by means of the analysis developed in this paper. Here, we give a first-order necessary condition for local optimality. Other optimality conditions may also be derived but we leave these for future work.

Theorem 10.1. *(First-order necessary conditions, Fritz John type) If a matrix X is a local minimizer of* (10.1)–(10.2)*, then there exists a scalar $\eta \in \mathbf{R}_+$, a matrix $Y \in \mathbf{M}^n$ and a vector $y \in \mathbf{R}^m$, not all zero, satisfying*

$$\eta C = Y + \sum_{k=1}^{m} y_k A_k, \text{ and} \tag{10.3}$$

$$Y \in \text{pos } \partial\alpha(-X) \cup \partial^\infty \alpha(-X), \tag{10.4}$$

where, for a nonempty $\Omega \subset \mathbf{M}^n$,

$$\text{pos } \Omega = \{t\, Z : \; Z \in \Omega, \; t \in \mathbf{R}_+\}.$$

Proof. The semistable program (10.1)–(10.2) is equivalent to the problem

$$\min_{X \in \mathcal{X}} \langle C, X\rangle + \tau(F(X))$$

where

$$F(X) = [\langle A_1, X\rangle - b_1, \ldots, \langle A_m, X\rangle - b_m]^T \in \mathbf{R}^m,$$

τ is the indicator function defined by, for $x \in \mathbf{R}^m$,

$$\tau(0) = 0, \qquad \tau(x) = \infty \text{ if } x \neq 0,$$

and

$$\mathcal{X} = \{X \;\; : \;\; \alpha(-X) \leq 0\}.$$

Applying the composite Lagrange multiplier rule [RW98, Example 10.8 together with Proposition 10.3] gives the result. QED

The matrix Y is called the *dual matrix.* We leave for future work consideration of the appropriate constraint qualification that would guarantee $\eta > 0$, allowing the elimination of η and therefore the upgrading of the Fritz John condition to one of Karush-Kuhn-Tucker type.

Notice that since $0 \in \partial^\infty \alpha(-X)$ for all X, it is not necessary to exclude the case that all eigenvalues of X have strictly positive real part. Such a matrix X is a local minimizer in the trivial case that C lies in the range of the A_k, $k = 1, \ldots, m$. However, this case is of little interest, since the spectral abscissa constraint is not active. Accordingly, let us change the definition of active set to one suitable for semistable programming, using

$$\mathcal{A} = \{j : \text{Re } \mu_j = 0\}. \tag{10.5}$$

With the understanding that this definition replaces (7.2), we then have:

Theorem 10.2. *(First-order necessary conditions, Fritz John type, details) Suppose that a matrix X is a local minimizer of* (10.1)–(10.2)*, with X having Jordan form* (4.1), (4.2)*. Then there exists a scalar $\eta \in \mathbf{R}_+$, a matrix $Y \in \mathbf{M}^n$ and a vector $y \in \mathbf{R}^m$, not all zero, satisfying* (10.3)*, the Toeplitz block condition* (4.4)*, and the active set condition* (6.4)*. Furthermore, the eigenvalues of Y (equivalently of W) are real and nonnegative.*

Proof. This is a consequence of Theorem 10.1 and Corollary 8.1. Notice that the trace condition on the sum of the eigenvalues no longer appears as a necessary condition; the positive multiplier implicit in the "pos" operator has been absorbed into Y. QED

We now generalize the notion of complementarity, familiar from semidefinite programming, to semistable programming.

Theorem 10.3. *(Complementarity) Suppose that a matrix X is a local minimizer of (10.1)–(10.2), and Y is a dual matrix whose existence is guaranteed by Theorem 10.2. Then the eigenvalues of X are in the right-half plane, the eigenvalues of Y are on the nonnegative real axis, and the eigenvalues of XY^* are on the imaginary axis. More specifically, there exist U unitary and P nonsingular such that*

$$U^*XU = R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, \quad U^*YU = S = \begin{bmatrix} S_{11} & 0 \\ S_{21} & 0 \end{bmatrix} \tag{10.6}$$

and

$$P^{-1}XP = J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}, \quad P^*YP^{-*} = W = \begin{bmatrix} W_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{10.7}$$

with $RS^* = S^*R$ and $JW^* = W^*J$, where R_{11} and R_{22} are upper triangular, S_{11} is lower triangular, and J consists of Jordan blocks, with the eigenvalues of R_{11} (and of J_1) on the imaginary axis, the eigenvalues of R_{22} (and of J_2) strictly in the right-half plane, and the eigenvalues of S_{11} (and of W_1) on the nonnegative real axis.

Proof. The block partitioning corresponds to the active set partitioning, with the eigenvalues of R_{11} (and J_1) being the active eigenvalues. The proof of (10.6) and (10.7) follows from Corollary 2.2 and Theorem 10.2. The second diagonal block of S vanishes because of Lemma 2.4 and Corollary 4.2: (4.6) implies (2.6), identifying S^{jj} with the second diagonal block of S and $\tilde{S}^{jj}$ with the second diagonal block of W, which is zero (by Theorem 10.2).

The eigenvalues of JW^* and of XY^* are the same as those of RS^*, namely, its diagonal entries, since RS^* is upper triangular. These eigenvalues are the pairwise products $(\mathrm{diag}(R))_\ell \mathrm{diag}(S))_\ell$, $\ell = 1, \dots, n$. Thus we get imaginary eigenvalues for the first diagonal block (real times imaginary) and, more specifically, zero eigenvalues for the second diagonal block (complex times zero). QED

If X and Y are both Hermitian positive semidefinite, the statement that XY has imaginary eigenvalues is equivalent to the statement $XY = 0$. More specifically, both (10.6) and (10.7) reduce to

$$U^*XU = \Lambda = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_2 \end{bmatrix}, \quad U^*YU = \Omega = \begin{bmatrix} \Omega_1 & 0 \\ 0 & 0 \end{bmatrix}$$

with U unitary and Λ_2 and Ω_1 both diagonal, real, and nonnegative (in fact Λ_2 strictly positive). Thus, Theorem 10.3 generalizes the well known notion of complementarity in semidefinite programming.

Acknowledgments. It is a pleasure to thank Adrian Lewis for his interest in this work and for many helpful conversations.

Appendix A. Proof of Lemma 1.1

We restate the definition of the subderivative and the statement of Lemma 1.1 with a slight change of notation.

Let $\phi : \mathbf{E} \to [-\infty, +\infty]$, with $\phi(\tilde{x})$ finite. The subderivative of ϕ at $\tilde{x}$ in the direction $\tilde{w}$ is

$$d\phi(\tilde{x})(\tilde{w}) = \liminf_{\substack{\tau \downarrow 0 \\ w \to \tilde{w}}} \frac{\phi(\tilde{x} + \tau w) - \phi(\tilde{x})}{\tau}.$$

Lemma A.1. *Let $\phi : \mathbf{E} \to [-\infty, +\infty]$, with $\phi(\tilde{x})$ finite, and let $\tilde{w} \in \mathbf{E}$ with $\tilde{w} \neq 0$. Set*

$$\Gamma(\tilde{x}, \tilde{w}) = \left\{ \gamma \;\middle|\; \begin{array}{c} \gamma : [0,\delta] \to \mathbf{E} \text{ is continuous for some } \delta > 0, \\ \text{with } \gamma(0) = \tilde{x} \text{ and } \gamma'_+(0) = \tilde{w} \end{array} \right\},$$

where

$$\gamma'_+(0) = \lim_{t \downarrow 0} \frac{\gamma(t) - \gamma(0)}{t}.$$

Then

$$d\phi(\tilde{x})(\tilde{w}) = \inf_{\gamma \in \Gamma(\tilde{x}, \tilde{w})} d(\phi \circ \gamma)(0)(\|\tilde{w}\|),$$

where the infimum is attained.

Proof. Clearly the left-hand quantity is less than or equal to the right-hand quantity, so we need only establish the reverse inequality. In order to do so, we construct a curve $\gamma \in \Gamma(\tilde{x}, \tilde{w})$ for which $d(\phi \circ \gamma)(0)(\|\tilde{w}\|) \leq d\phi(\tilde{x})(\tilde{w})$. Let (τ_k, w^k), $k = 1, 2, \dots$, be a sequence that attains the "lim inf" in the definition of $d\phi(\tilde{x})(\tilde{w})$. Without loss of generality, assume that $\|\tilde{w}\| = 1$, $\|w^k\| = 1$, and the τ_k converge to zero superlinearly. Set $x_k = \tilde{x} + \tau_k w^k$, for $k = 1, 2, \dots$, so that $\tau_k = \|x^k - \tilde{x}\|$. Let $s^k = x^k - x^{k+1}$ and $\sigma_k = \|s^k\|$. We have

$$\frac{x^k - \tilde{x}}{\|x^k - \tilde{x}\|} = w^k \to \tilde{w}. \tag{A.1}$$

Since the convergence is superlinear we have

$$\frac{\tau_{k+1}}{\tau_k} \to 0, \tag{A.2}$$

$$\delta = \sum_{i=1}^{\infty} \sigma_i < \infty,$$

$$\lim_{k \to \infty} \frac{\sigma_k}{\tau_k} = 1, \quad \text{and} \tag{A.3}$$

$$\lim_{k \to \infty} \sum_{i=k+1}^{\infty} \frac{\tau_i}{\tau_k} = 0. \tag{A.4}$$

Define $\gamma : [0, \delta] \to \mathbf{E}$ as the piecewise linear trajectory connecting the points x^k in succession, with $\gamma(0) = \tilde{x}$ and $\gamma(\delta) = x_1$, i.e.

$$\gamma(t) = x^k + \frac{t - t_k}{\sigma_k} s^k, \quad t \in [t_k, t_{k+1}]$$

where

$$t_k = \sum_{i=k}^{\infty} \sigma_i, \quad k = 1, 2, \dots .$$

We now show that $\gamma'_+(0) = \tilde{w}$. For $t \in [t_k, t_{k+1}]$,

$$\|\frac{\gamma(t) - \tilde{x}}{t} - \tilde{w}\| \leq \|\frac{s^k}{\sigma_k} - \tilde{w}\| + \frac{1}{t_{k+1}} \|x^k - \tilde{x} - \frac{t_k}{\sigma_k} s^k\|. \tag{A.5}$$

Observe that

$$\frac{s^k}{\sigma_k} \to \tilde{w}, \tag{A.6}$$

using (A.1), (A.2), and (A.3), since

$$\frac{s^k}{\sigma_k} = \frac{\tau_k}{\sigma_k} w^k - \frac{\tau_{k+1}}{\tau_k}\frac{\tau_k}{\sigma_k} w^{k+1}.$$

It remains to show that the second term on the right-hand side of (A.5) converges to zero. Noting that $t_k - t_{k+1} = \sigma_k$, it is easily shown that

$$\frac{1}{t_{k+1}}\left(x^k - \tilde{x} - t_k \frac{s^k}{\sigma_k}\right) = \frac{\tau_{k+1}}{t_{k+1}} w^{k+1} - \frac{s^k}{\sigma_k}. \tag{A.7}$$

Also,

$$\frac{t_k}{\tau_k} = \frac{\sigma_k}{\tau_k} + \sum_{i=k+1}^{\infty} \frac{\sigma_i}{\tau_k} \to 1,$$

by (A.3), since

$$\sum_{i=k+1}^{\infty} \frac{\sigma_i}{\tau_k} \le \sum_{i=k+1}^{\infty} \frac{\tau_{i+1} + \tau_i}{\tau_k} \to 0,$$

by (A.4). Therefore, using (A.1) and (A.6) we see that both terms on the right-hand side of (A.7) converge to $\tilde{w}$. This shows that both terms on the right-hand side of (A.5) converge to zero, yielding $\gamma'_+(0) = \tilde{w}$. Hence

$$\begin{aligned} d\phi(\tilde{x})(\tilde{w}) &= \lim_{k\to\infty} \frac{\phi(x^k) - \phi(\tilde{x})}{\|x^k - \tilde{x}\|} \\ &\ge \liminf_{t\downarrow 0} \frac{t}{\|\gamma(t) - \tilde{x}\|} \frac{\phi(\gamma(t)) - \phi(\tilde{x})}{t} \\ &= \liminf_{t\downarrow 0} \frac{1}{\|\gamma'_+(0)\|} \frac{\phi(\gamma(t)) - \phi(\tilde{x})}{t} \\ &= d(\phi \circ \gamma)(0)(1), \end{aligned}$$

completing the proof. QED

Appendix B. The Schur Factorization

The two lemmas presented here are variations on standard results for the Schur factorization [HJ85, Section 2.3].

The first result is surely known, but we were unable to find it in the literature.

Lemma B.1. *Suppose $A \in \mathbf{M}^n$ and $B \in \mathbf{M}^n$ commute, i.e. $AB = BA$. Then there exists a unitary matrix $U \in \mathbf{M}^n$ such that both U^*AU and U^*BU are upper triangular and the eigenvalues of A appear on the diagonal of U^*AU in any prescribed order.*

Proof. We begin by showing that every eigenvalue of A has an associated eigenvector that is also an eigenvector for B. Let μ be an eigenvalue of A and set $\mathcal{E}_\mu = \mathrm{Null}\,(\mu I - A)$. For any $v \in \mathcal{E}_\mu$, we have

$$ABv = BAv = \mu Bv$$

so that $Bv \in \mathcal{E}_\mu$. Therefore, $\mathcal{E}_\mu$ is a B–invariant subspace. Consequently, by [HJ85, p. 51], there is an eigenvector of B in $\mathcal{E}_\mu$.

This fact can now be used in conjunction with the proofs of Theorems 2.3.1 and 2.3.3 in [HJ85] to establish the result. QED

The second result concerns the continuity of the Schur factorization of a perturbation of a Jordan block.

Lemma B.2. *Let $J \in \mathbf{M}^n$ be a Jordan block, i.e., a single block of the form $J_k^{(j)}$ in* (4.2)*. Then there exists $\delta_0 > 0$ such that, if $\|E\| = \delta \le \delta_0$, there exists a unitary matrix U with $U^*(J+E)U$ upper triangular and $\|U - I\| = O(\delta^{1/n!})$.*

Proof. Without loss of generality, we may assume the diagonal entries of J are zero. The proof is by induction, and is based on the standard proof of existence of the Schur form [HJ85, Theorem 2.3.1]. There is nothing to prove when $n = 1$. Suppose the lemma is true for $n = k-1$; we sketch an argument that it must then be true for $n = k$. The first step is to observe that any eigenvalue μ of the k by k matrix $J + E$ must satisfy $\mu = O(\delta^{1/k})$ [Kat84, Section II.1.2]. Let u be an eigenvector corresponding to an eigenvalue μ, with $u = [\eta\ v^*]^*$, where $\eta \in \mathbf{C}$ and $v \in \mathbf{C}^{k-1}$. We have

$$\mu \begin{bmatrix} \eta \\ v \end{bmatrix} = (J+E) \begin{bmatrix} \eta \\ v \end{bmatrix} = \begin{bmatrix} v \\ 0 \end{bmatrix} + E \begin{bmatrix} \eta \\ v \end{bmatrix}. \tag{B.1}$$

If $\eta = 0$, we may normalize v so that $\|v\| = 1$, but then (B.1) contradicts the fact that $\mu = O(\delta^{1/k})$ (for δ sufficiently small). Therefore, we may choose $\eta = 1$. Defining $\tilde{E}$ to be the $k-1$ by $k-1$ matrix obtained by dropping the first column and the last row from E, and setting $F = \begin{bmatrix} 0 & 0 \\ \mu I & 0 \end{bmatrix}$, where the identity block has order $k-1$, we find, dropping the last equation from (B.1), that

$$(I + \tilde{E} - F)v = O(\mu) + O(\delta), \quad \text{with } \|\tilde{E}\| + \|F\| = O(\mu) + O(\delta).$$

Since $\mu = O(\delta^{1/k})$, the standard Banach lemma tells us that $\|v\| = O(\delta^{1/k})$. Applying the Gram-Schmidt orthogonalization process to the k vectors $u, e_2, \ldots, e_k$, where $u = [1\ v^*]^*$ and e_j is the jth column of the k by k identity matrix, results in a unitary matrix U_1 with $\|U_1 - I\| = O(\delta^{1/k})$ and

$$U_1^*(J+E)U_1 = \begin{bmatrix} \mu & y^* \\ 0 & K \end{bmatrix}$$

where K is an $O(\delta^{1/k})$ perturbation of a $k-1$ by $k-1$ Jordan block. Applying the inductive hypothesis, we obtain

$$U_2^* K U_2 = T$$

where T is upper triangular and

$$\|U_2 - I\| = O(\delta^{1/k})^{1/(k-1)!} = O(\delta^{1/k!}).$$

The desired unitary matrix which triangularizes $J + E$ is therefore

$$U = U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}$$

which satisfies

$$\|U - I\| = O(\|U_1 - I\| + \|U_2 - I\|) = O(\delta^{1/k!}).$$

QED

Numerical experiments suggest that it may be possible to improve the modulus of continuity to $\|U - I\| = O(\delta^{1/n})$.

References

[Arn71] V.I. Arnold. On matrices depending on parameters. *Russian Mathematical Surveys*, 26:29–43, 1971.

[BO92a] J.V. Burke and M.L. Overton. On the subdifferentiability of functions of a matrix spectrum I: Mathematical foundations; II: Subdifferential formulas. In F. Giannessi, editor, *Nonsmooth Optimization : Methods and Applications*, pages 11–29, Philadelphia, 1992. Gordon and Breach. Proceedings of a conference held in Erice, Italy, June 1991.

[BO92b] J.V. Burke and M.L. Overton. Stable perturbations of nonsymmetric matrices. *Linear Algebra and its Applications*, 171:249–273, 1992.

[BO94] J.V. Burke and M.L. Overton. Differential properties of the spectral abscissa and the spectral radius for analytic matrix–valued mappings. *Nonlinear Analysis, Theory, Methods and Applications*, 23:467–488, 1994.

[BO99] J.V. Burke and M.L. Overton. The subderivative of the abscissa mapping for polynomials, 1999. In preparation.

[BT97] V. Blondel and J.N. Tsitsiklis. NP-Hardness of some linear control design problems. *SIAM Journal on Control and Optimization*, 35:2118–2127, 1997.

[Cla73] F.H. Clarke. *Necessary conditions for nonsmooth problems in optimal control and the calculus of variations*. PhD thesis, University of Washington, 1973.

[HJ85] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1985.

[HJ91] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1991.

[Iof81] A.D. Ioffe. Sous-différentielles approchées de fonctions numériques. *Comptes Rendus de l'Académie des Sciences de Paris*, 292:675–678, 1981.

[Kat84] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, second edition, 1984.

[Lax97] P. D. Lax. *Linear Algebra*. John Wiley, New York, 1997.

[Lew96a] A.S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6:164–177, 1996.

[Lew96b] A.S. Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 21:576–588, 1996.

[Lew99] A.S. Lewis. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84:1–24, 1999.

[LT85] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, New York and London, 1985.

[MBO97] J. Moro, J.V. Burke, and M.L. Overton. On the Lidskii-Vishik-Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. *SIAM Journal on Matrix Analysis and Applications*, 18:793–817, 1997.

[Mor76] B.S. Mordukhovich. Maximum principle in the problem of time optimal response with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, 40:960–969, 1976.

[Nem93] A. Nemirovskii. Several NP-hard problems arising in robust stability analysis. *Math. Control Signals Systems*, 6:99–105, 1993.

[OW88] M.L. Overton and R.S. Womersley. On minimizing the spectral radius of a nonsymmetric matrix function – optimality conditions and duality theory. *SIAM Journal on Matrix Analysis and Applications*, 9:473–498, 1988.

[RW98] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*. Springer-Verlag, 1998.

Department of Mathematics, University of Washington, Seattle, WA 98195
E-mail address: `burke@math.washington.edu`

Courant Institute of Mathematical Sciences, New York University, New York, NY 10012
E-mail address: `overton@cs.nyu.edu`