

James V. Burke · Michael L. Overton

## Variational analysis of non-Lipschitz spectral functions

Received: September 1999 / Accepted: January 2001

Published online March 22, 2001 – © Springer-Verlag 2001

**Abstract.** We consider spectral functions  $f \circ \lambda$ , where  $f$  is any permutation-invariant mapping from  $\mathbf{C}^n$  to  $\mathbf{R}$ , and  $\lambda$  is the eigenvalue map from the set of  $n \times n$  complex matrices 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). We show that a necessary condition for  $Y$  to be a subgradient of an eigenvalue function  $f \circ \lambda$  at  $X$  is that  $Y^*$  commutes with  $X$ . We also give a number of other necessary conditions for  $Y$  based on the Schur form and the Jordan form of  $X$ . In the case of the spectral abscissa, we refine these conditions, and we precisely identify the case where subdifferential regularity holds. We conclude by introducing the notion of a semistable program: maximize a linear function on the set of square matrices subject to linear equality constraints together with the constraint that the real parts of the eigenvalues of the solution matrix are non-positive. Semistable programming is a nonconvex generalization of semidefinite programming. Using our analysis, we derive a necessary condition for a local maximizer of a semistable program, and we give a generalization of the complementarity condition familiar from semidefinite programming.

**Key words.** nonsmooth analysis – eigenvalue function – spectral abscissa – spectral radius – semistable program – stability

### 1. Introduction

Let  $\mathbf{M}^n$  denote the Euclidean space of  $n \times n$  complex matrices, with real inner product

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

and norm  $\|X\| = \langle X, X \rangle^{1/2}$ . 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  $\lambda_1(X), \dots, \lambda_n(X)$ , repeated according to multiplicity and ordered lexicographically so that, if  $k < \ell$ , then either  $\operatorname{Re} \lambda_k(X) > \operatorname{Re} \lambda_\ell(X)$ , or  $\operatorname{Re} \lambda_k(X) = \operatorname{Re} \lambda_\ell(X)$  with  $\operatorname{Im} \lambda_k(X) \geq \operatorname{Im} \lambda_\ell(X)$ . Thus we uniquely define the eigenvalue map

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

J.V. Burke: Department of Mathematics, University of Washington, Seattle, WA 98195, USA,  
e-mail: burke@math.washington.edu

M.L. Overton: Courant Institute of Mathematical Sciences, New York University, New York, NY 10012,  
USA, e-mail: overton@cs.nyu.edu

*Mathematics Subject Classification (2000):* 15A42, 34D05, 49J52, 49K99, 90C26

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 [4–6] as well as other work done by the authors with R.S. Womersley [22] and J. Moro [20]. 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 [16–18].

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 composite form

$$f \circ \lambda : \mathbf{M}^n \rightarrow [-\infty, +\infty], \quad (1.1)$$

where the only restriction on the function  $f : \mathbf{C}^n \rightarrow [-\infty, +\infty]$  is that it must be invariant under permutation of its argument components. Thus, the lexicographic order used to define  $\lambda$  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  $\lambda$  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*

$$\alpha = (\max \operatorname{Re}) \circ \lambda$$

and the *spectral radius*

$$\rho = (\max \operatorname{mod}) \circ \lambda$$

where  $\operatorname{mod}(x) = |x|$  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 Ph.D. thesis [8]. 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 gradient) in the smooth case. In more recent years, attention has turned to the nonconvex-set-valued maps introduced and analyzed by Mordukhovich [19], Kruger and Mordukhovich [13] and Ioffe [11], and it is just such a map that forms the centerpiece of the comprehensive book by Rockafellar and Wets [23, Chap. 8]. Following Lewis [18], 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 [23, Chap. 8] for more details. Let  $\phi : \mathbf{E} \rightarrow [-\infty, +\infty]$ , 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  $\phi$  at  $x$  (written  $y \in \hat{\partial}\phi(x)$ ) if

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

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

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

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

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

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

A vector  $y \in \mathbf{E}$  is a *horizon subgradient* of  $\phi$  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),

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

where by  $s_i \downarrow 0$ , we mean a sequence of positive real numbers decreasing to zero.

It follows from the definition that  $\hat{\partial}\phi(x)$ , the set of regular subgradients of  $\phi$  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  $\phi$  is a convex function,  $\hat{\partial}\phi = \partial\phi$  and coincides with the ordinary subdifferential of convex analysis.

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

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

where  $\tilde{y}$  is any element of  $\hat{\partial}\phi(x)$  [23, Theorem 3.6]. Directly from the definitions, we have

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

Regularity is a key notion in nonsmooth analysis, going back to [8]. We say that  $\phi$  is *subdifferentially regular* at  $x$  if [23, 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  $\phi$  at  $x$  in the direction  $w$  is

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

We have, immediately from the definition, that

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

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  $\phi(x) = \operatorname{Re} x$ , 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  $\{y : |y| \leq 1\}$  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 equivalence of  $\mathbf{R}^2$  and  $\mathbf{C}$  is conveniently captured by the linear transformation  $\Theta : \mathbf{R}^2 \rightarrow \mathbf{C}$  defined by

$$\Theta(x) = x_1 + \sqrt{-1}x_2,$$

where  $\sqrt{-1}$  denotes the imaginary unit. We have

$$\Theta^{-1}\mu = \Theta^*\mu = \begin{bmatrix} \operatorname{Re} \mu \\ \operatorname{Im} \mu \end{bmatrix},$$

where  $\Theta^*$  is the adjoint of  $\Theta$  with respect to the real inner product

$$\langle \mu, v \rangle = \operatorname{Re}(\overline{\mu}v). \tag{1.10}$$

Let  $\gamma : \mathbf{R}^2 \rightarrow \mathbf{R}$  be given, and define  $\kappa : \mathbf{C} \rightarrow \mathbf{R}$  by the composition

$$\kappa = \gamma \circ \Theta^*.$$

If  $\gamma$  is differentiable at  $\Theta^*\mu$ , the chain rule gives us

$$\kappa'(\mu) = \Theta\nabla\gamma(\Theta^*\mu), \tag{1.11}$$

where  $\nabla\gamma$  denotes the gradient of  $\gamma$ , and, if  $\gamma$  is twice differentiable,

$$\kappa''(\mu)v = \Theta\nabla^2\gamma(\Theta^*\mu)\Theta^*v, \tag{1.12}$$

where  $\nabla^2\gamma$  denotes the Hessian of  $\gamma$ . If  $\gamma$  is continuously differentiable at  $\Theta^*\mu$ , then  $\kappa'$  is continuous at  $\mu$ , and we say that  $\kappa$  is  $C^1$  in the real sense at  $\mu$ . If  $\gamma$  is twice continuously differentiable at  $\Theta^*\mu$ , then  $\kappa''$  is continuous at  $\mu$ , and we say that  $\kappa$  is  $C^2$  in the real sense at  $\mu$ . Application of Taylor's theorem to  $\kappa$  gives us the following lemma, which will be useful later.

**Lemma 1.1.** *Let  $\mu \in \mathbf{C}$  and define  $\kappa, \kappa'$  and  $\kappa''$  as above, with  $\kappa$  being  $C^2$  in the real sense at  $\mu$ . Suppose  $\kappa'(\mu) \neq 0$ . Let  $v, \omega \in \mathbf{C}$  satisfy  $v = \pm\sqrt{-1}\kappa'(\mu)$  and  $\omega = \delta\kappa'(\mu)$ , where*

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

Then

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

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

Having introduced the basic notation that we need, we now give an overview of the paper.

In Sect. 2, we prove a necessary condition for a matrix  $Y$  to be a subgradient or horizon subgradient of *any* spectral function  $f \circ \lambda$  at  $X$ , namely:  $Y^*$  must commute with  $X$ . An immediate consequence is that there is a unitary matrix that simultaneously triangularizes  $X$  and  $Y^*$ . This result generalizes one established by Lewis [18] in the Hermitian matrix setting, namely: the Hermitian matrices  $X$  and  $Y$  must commute and must therefore be unitarily simultaneously diagonalizable. In Sect. 3, we take this one step further, showing that the *diagonal of the triangular form* of  $Y$  is actually a *subgradient* (or horizon subgradient) of  $f$  at  $\lambda(X)$ . Again, this generalizes a result of Lewis in the Hermitian setting, where it is the end of the story; all subgradients and horizon subgradients are completely characterized by this condition. In the general setting, it is just the beginning.

In Sect. 4 we introduce the Jordan form. A semisimple eigenvalue is one for which all corresponding Jordan blocks have size one, and a nonderogatory eigenvalue is one for which there is only one Jordan block, whose size equals the eigenvalue multiplicity. We give a detailed necessary condition for  $Y$  to be a subgradient or horizon subgradient of  $f \circ \lambda$ , based on the fact that the matrices that commute with a Jordan form have a block structure with triangular Toeplitz blocks. We give stronger results in the cases that the subgradient is regular, or the eigenvalues are nonderogatory.

We obtain further conditions characterizing the entries of the triangular Toeplitz blocks in Sect. 5. Here we restrict the spectral function  $f \circ \lambda$  by  $f = g \circ h_\kappa$ , where the smooth function  $h_\kappa$  maps all eigenvalues by the same complex-to-real map  $\kappa$ , and the (not necessarily smooth) function  $g$  maps a real vector to a real scalar, and is invariant under permutations of its argument components. Further results along this line are obtained for “spectral max functions” in Sect. 6, where we assume that  $g$  is the max function. These functions include the spectral abscissa and the spectral radius.

In Sect. 7, we further restrict our attention to the spectral abscissa  $\alpha$ , and we completely characterize all regular subgradients of  $\alpha$ . For example, the only regular subgradient of  $\alpha$  at  $X = 0$  is  $(1/n)I$ , where  $I$  is the identity matrix, while the regular subgradients of  $\alpha$  at  $X$  given by an  $n$  by  $n$  upper Jordan block consist of the lower triangular Toeplitz matrices with the restriction that the diagonal entry is  $1/n$  and the first subdiagonal entry has nonnegative real part. Section 8 considers all subgradients and horizon subgradients of  $\alpha$ . In this section, we prove the most important result in the paper: the spectral abscissa is subdifferentially regular at  $X$  if and only if all active eigenvalues of  $X$  (those with real part equal to the maximum real part) are nonderogatory. In particular,  $\alpha$  is subdifferentially regular at  $X$  given by an  $n$  by  $n$  Jordan block, and hence the set of subgradients is the same as the set of regular subgradients just described. We also completely characterize the subgradients and horizon subgradients of  $\alpha$  at  $X$  when all active eigenvalues of  $X$  are semisimple. For example, the subgradients of  $\alpha$  at  $X = 0$  are exactly those matrices whose eigenvalues are real, nonnegative, and sum to one, and the horizon subgradients of  $\alpha$  at  $X = 0$  are the nilpotent matrices. Neither of these subgradient sets is convex.

In Sect. 9 we draw analogies and contrasts between these results and the well known results in the Hermitian setting, where, for example, the (necessarily regular)

subgradients of the convex function “max eigenvalue” at  $X = 0$  consist of all matrices that are positive semidefinite and have trace one.

Finally, in Sect. 10 we introduce semistable programming, a nonconvex generalization of semidefinite programming. Using our analysis, we derive a necessary condition for a local maximizer of a semistable program, and give a generalization of the complementarity condition familiar from semidefinite programming.

By  $\text{Diag}(x)$  we mean the diagonal matrix constructed from the vector  $x$ , while  $\text{diag}(X)$  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 [18, Theorem 3], where a closely related result is given for spectral functions on the space of Hermitian matrices. Instead of [18, 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 [1]. Although a proof of the formula for  $T_X$  is not to be found in [1], one is easily constructed by generalizing the proof of [18, Theorem 1] to the non-Hermitian case. The rest of the proof follows exactly as in the proof of [18, Theorem 3].

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 2.1.** *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*

$$R = U^*XU \quad \text{and} \quad S = U^*YU \tag{2.1}$$

*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.*

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

*Proof.* The existence of the simultaneously triangularizing unitary matrix  $U$  follows from [9, Thm 2.3.3]. For the ordering, see Lemma A.1 in Appendix A.

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

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

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

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

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

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

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

The following lemmas will be useful.

**Lemma 2.1.** *Suppose that*

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

*where  $R$  and  $\tilde{R}$  are both upper triangular and both have the block triangular structure given in (2.3) and the diagonal restriction given in (2.4), with  $\mu_1, \dots, \mu_p$  distinct and ordered lexicographically. 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  $\text{diag}(R) = \text{diag}(\tilde{R})$ . 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 [10, 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  $T^{(12)} = 0$ .

In the following, by  $X^{-*}$  we mean  $(X^*)^{-1} = (X^{-1})^*$ .

**Lemma 2.2.** *Let the assumptions of Lemma 2.1 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

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

*Proof.* Since, by Lemma 2.1,  $T^*$  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  $j$ th block, there is a permutation matrix  $Q^{(j)}$  such that

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

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

An immediate consequence of Lemma 2.2 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  $\operatorname{diag}(R) = \lambda(X)$ , for some unitary matrix  $U$ , as in Corollary 2.1. Then

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

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



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

$$f(\text{diag}(R) + z) = (f \circ \lambda)(U(R + \text{Diag}(z))U^*) \tag{3.1}$$

$$\begin{aligned} &= (f \circ \lambda)(X + U\text{Diag}(z)U^*) \\ &\geq (f \circ \lambda)(X) + \langle Y, U\text{Diag}(z)U^* \rangle + o(\|z\|) \end{aligned} \tag{3.2}$$

$$= f(\text{diag}(R)) + \langle \text{diag}(S), z \rangle + o(\|z\|) \tag{3.3}$$

so  $\text{diag}(S) \in \hat{\partial} f(\text{diag}(R)) = \hat{\partial} f(\lambda(X))$ . 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  $X_i \rightarrow X$ , with  $f(\lambda(X_i)) \rightarrow f(\lambda(X))$  and a sequence of regular subgradients  $Y_i \in \hat{\partial}(f \circ \lambda)(X_i)$ , with  $Y_i \rightarrow Y$ . By Corollary 2.1 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  $\text{diag}(R_i) \rightarrow \text{diag}(R) = \lambda(X)$ . (This does not imply that  $\text{diag}(R_i)$  is lexicographically ordered.) From an identical argument to (3.1)–(3.3), we have

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

Since the set of all unitary matrices is compact, we can also assume  $U_i \rightarrow \tilde{U}$ , 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,  $\text{diag}(\tilde{R}) = \text{diag}(R)$ , 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 2.1 and 2.2 to obtain the existence of a permutation matrix  $Q$  satisfying

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

Taking limits in (3.4) yields

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

By [23, 10.7, p. 428] or [18, Proposition 2],

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

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

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

Both the statement and the proof of this result were inspired by Lewis [18, 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

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

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

Here, as in the previous section,  $\mu_1, \dots, \mu_p$  denote the distinct eigenvalues of  $X$ . Each  $J_k^{(j)}$  is a Jordan block of size  $m_k^{(j)} \times m_k^{(j)}$  for the eigenvalue  $\mu_j$ . The multiplicity of  $\mu_j$  is

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

The size of the largest Jordan block for  $\mu_j$  is denoted

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

An eigenvalue  $\mu_j$  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  $\mu_j$  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 [1]. 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  $\mu_j$  is semisimple, the corresponding chains have length one, so the generalized eigenvectors are actually eigenvectors.

We also define

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

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*

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

where  $W_{rs}^{(j)}$  is a rectangular  $m_r^{(j)} \times m_s^{(j)}$  lower triangular Toeplitz matrix,  $r = 1, \dots, q^{(j)}$ ,  $s = 1, \dots, q^{(j)}$ ,  $j = 1, \dots, p$ . By this we mean that the value of the  $k, \ell$  entry in each  $W_{rs}^{(j)}$  depends only on the difference  $k - \ell$  (is constant along the diagonals), and is zero if  $k < \ell$  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 [14, Sect. 12.4] for a proof and [1, Sect. 4.2] or [22] for illustrations.

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

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

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 4.1.** *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  $j$ ,  $S^{(j)}$  and  $W^{(j)}$  have the same eigenvalues, namely, the diagonal entries of  $S^{(j)}$ . Furthermore, if  $\mu_j$  is nonderogatory,  $S^{(j)}$  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

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

where  $T = U^*P$ . Applying Lemma 2.2 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.

For regular subgradients, there is a much stronger result.

**Theorem 4.2.** *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*

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

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

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

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

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  $r$ th row and  $s$ th column of the entire matrix  $W$  and let  $\beta$  be this nonzero value. Let  $Z = PVP^{-1}$ , where all components of  $V$  are zero except the  $r, s$  component, which is set to  $\beta$ . 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

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

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  $1 \leq k < k' \leq q^{(j)}$ , and each  $\ell$  satisfying  $1 \leq \ell \leq \min(m_k^{(j)}, m_{k'}^{(j)})$ , 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  $\ell$ . Without loss of generality we may assume  $k' = k + 1$ . Let  $r$  be the integer such that the  $r$ th diagonal entry of the entire matrix  $W$  is in the last diagonal position of



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

If an eigenvalue  $\mu_j$  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  $\mu_j$  is derogatory. In Sect. 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 4.2.** *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*

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

with

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

*Proof.* Since  $Y$  is regular,  $W^{(j)}$ , like  $S^{(jj)}$ , is lower triangular. Therefore, by Corollary 4.1 and Lemma 2.2, we know there exists a permutation matrix  $Q$  satisfying

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

This shows that  $\text{diag}(W) = \text{diag}(S)$ , since, from Theorem 4.2, any permutation matrix  $Q$  satisfying  $Q \text{diag}(J) = \text{diag}(J)$  also satisfies  $Q \text{diag}(W) = \text{diag}(W)$ . We know that  $\text{diag}(S) \in \hat{\partial} f(\lambda(X))$  from Theorem 3.1. The last statement is an immediate consequence.

### 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

$$f \circ \lambda = g \circ h_\kappa \circ \lambda \tag{5.1}$$

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

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

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

**Theorem 5.1.** *Let (5.2) hold, where  $\kappa$  is  $C^1$  in the real sense at  $\mu_j, j = 1, \dots, p$ , and let*

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

*Suppose that  $Kz = 0$  implies  $z = 0$  for all  $z \in \partial^\infty g(h_\kappa(\lambda(X)))$ , i.e. for all horizon subgradients of  $g$  at  $h_\kappa(\lambda(X))$ . (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  $g \circ h_\kappa \circ \lambda$  at  $X$ , i.e.*

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

*respectively, with  $R = U^*XU$  upper triangular,  $S = U^*YU$  lower triangular, and  $\text{diag}(R) = \lambda(X)$ , for some unitary matrix  $U$ , as in Corollary 2.1. Then there exists*

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

*respectively, satisfying*

$$\text{diag}(S) = Kw.$$

*Proof.* Applying Theorem 3.1 with  $f = g \circ h_\kappa$ , we find

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

The result therefore follows from applying the basic chain rule for subgradients [23, Theorem 10.6] to  $g \circ h_\kappa$ .

An important special case is:

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

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

*and*

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

*each  $\sigma_j$  being repeated  $m^{(j)}$  times. Then*

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

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

*Proof.* This is an immediate consequence of Corollary 4.1 and Theorem 5.1.

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

**Theorem 5.2.** *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  $\sigma_j$  by (5.3) and the vector  $\sigma$  by (5.4). Then*

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

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

$$\text{diag}(S) \in \hat{\partial}(g \circ h_\kappa)(\lambda(X)).$$

The result therefore follows from [23, Exercise 10.7], together with Corollary 4.2.

Theorem 5.2 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  $\kappa$  is  $C^2$  in the real sense at  $\mu \in \mathbf{C}$  if its second derivative  $\kappa''$ , defined in (1.12), is continuous at  $\mu$ .

**Theorem 5.3.** *Let (5.2) hold and suppose that  $\kappa$  is  $C^2$  in the real sense at  $\mu_j$  for  $j = 1, \dots, p$ . Assume that  $\kappa'(\mu_j) \neq 0$ ,  $j = 1, \dots, p$ , and suppose also that  $g$  is Lipschitz at  $h_\kappa(\lambda(X))$ . Let  $X$  have the Jordan form (4.1), (4.2), and suppose that  $Y$  is a **regular** subgradient of  $g \circ h_\kappa \circ \lambda$  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*

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

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  $\sigma_j$  is real from Theorem 5.2, and  $\eta_j$  is real by definition. Suppose that (5.6) does not hold, for some eigenvalue  $\mu_j$  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

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



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

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

Since  $\kappa'(\mu_j) \neq 0$ , we may apply Lemma 1.1, 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

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

and thus  $Y$  is not a regular subgradient of the spectral function  $g \circ h_{\kappa} \circ \lambda$ .

Theorem 5.2 and Theorem 5.3 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.  $\theta_3^{(j)}, \dots, \theta_{m_k}^{(j)}$ . We prove this in the case of the spectral abscissa in Sect. 7.

### 6. Spectral max functions

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

Let us define the *active set*

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

An eigenvalue  $\mu_j$  is said to be active if  $j \in \mathcal{A}$ , and inactive otherwise. We now show that if an eigenvalue  $\mu_j$  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:<sup>1</sup>

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

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

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

---

<sup>1</sup> The original reference is [1, Theorem 4.4]. A detailed proof may be found in [3].

Since  $P_i \rightarrow P$ , we have  $L_i \rightarrow J$ , 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 6.1.** *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  $g \circ h_\kappa \circ \lambda$  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  $\mu_j$  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  $j$ th 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  $\mu_j$ . Therefore, by continuity of eigenvalues,  $g \circ h_\kappa \circ \lambda$  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  $X_i \rightarrow X$  and  $Y_i \rightarrow Y$  with

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

By Lemma 6.1, there exists  $P_i \rightarrow P$  such that (6.2) holds. Since the Jordan form of  $X_i$  has the block diagonal form (6.3), Theorem 4.2 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  $\mu_j$  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

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

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

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 5.1 (where we assume that all  $\mu_j$  are nonderogatory, in the case where  $Y$  is a subgradient) or Theorem 5.2 (where we assume that  $Y$  is a regular subgradient), that the  $\sigma_j$  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  $\mu_j$  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  $\mu_j$  (a right eigenvector of  $X$ ) and  $u^*$  is the row of  $P^{-1}$  associated with  $\mu_j$  (a left eigenvector of  $X$ ); note that  $u^* v = 1$ . In this case,  $g \circ h_\kappa \circ \lambda$  is differentiable at  $X$ , with gradient  $Y$ , as is well known.

Likewise, when  $Y$  is a horizon subgradient of a spectral max function, the assumptions of Corollary 5.1 imply that, instead of (6.5), we have

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

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

We can be more specific: if  $\kappa = \text{Re}$ , so that  $g \circ h_\kappa \circ \lambda$  is the spectral abscissa, then  $\kappa'(\mu_j) = 1$ , and so  $\theta_1^{(j)} = \sigma_j$ ,  $j = 1, \dots, p$ . 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

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

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

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

Strictly speaking, in the spectral radius case, Corollary 5.1 and Theorem 5.2 do not apply if any eigenvalue  $\mu_j$  is zero; however, in view of Theorem 6.1, 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 5.3. In the spectral abscissa case, with  $\kappa = \text{Re}$ , we have  $\kappa'(\mu) = 1$  and  $\kappa''(\mu)v = 0$ , so condition (5.6) reduces to

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

(In this case, the proof of Theorem 5.3 simplifies considerably, since  $\tau_{\pm}(t) - \mu_j$  are imaginary and therefore Lemma 1.1 is not needed.) In the spectral radius case, where  $\kappa = \text{mod}$ , we have, for  $\mu_j \neq 0$ ,

$$\begin{aligned} \kappa'(\mu_j) &= \mu_j/|\mu_j| \quad \text{and} \quad \kappa''(\mu_j) \frac{\sqrt{-1}\mu_j}{|\mu_j|} = \\ &= \frac{-(\text{Im } \mu_j)^3 - (\text{Re } \mu_j)^2(\text{Im } \mu_j) + \sqrt{-1}(\text{Im } \mu_j)^2(\text{Re } \mu_j) + \sqrt{-1}(\text{Re } \mu_j)^3}{|\mu_j|^4}, \end{aligned}$$

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

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

### 7. The regular subgradients of the spectral abscissa

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

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

where  $g$  is the “max” function and  $h_{\kappa}$  maps the eigenvalues to their real parts, i.e.  $\kappa$  in (5.2) is the function  $\text{Re}$ . With this choice of spectral function, the active set of eigenvalues at  $X$  is given by

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

We shall show that the necessary conditions, derived in the previous sections, for  $Y$  to be a regular subgradient of  $\alpha$  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  $\zeta$  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:

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

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

$$\alpha = a \circ \Phi,$$

where  $\Phi : \mathbf{M}^n \rightarrow \mathcal{P}_n$  is defined by

$$\Phi(X) = \det(\zeta I - X).$$

Before we state the main theorem about regular subgradients of the spectral abscissa, we need two results. The first of these concerns the directional derivative of the differentiable map  $\Phi$ .

**Lemma 7.1.** *Let  $X \in \mathbf{M}^n$  and  $Z \in \mathbf{M}^n$  be given, and assume  $X$  has Jordan form (4.1), (4.2). Define the polynomials  $p(\zeta)$  and  $q(\zeta)$  by*

$$p(\zeta) = \Phi(X) = \det(\zeta I - X) = \prod_{j=1}^p (\zeta - \mu_j)^{m^{(j)}}, \tag{7.3}$$

and

$$q(\zeta) = \Phi'(X; Z) = \lim_{t \rightarrow 0} \frac{\Phi(X + tZ) - \Phi(X)}{t}. \tag{7.4}$$

Define

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

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

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

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  $M : \mathbf{R} \rightarrow \mathbf{M}^n$ , we have<sup>2</sup>

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

as long as  $M(0)$  is nonsingular [15, Chapter 9]. Since the derivative we are evaluating is a polynomial in  $\zeta$ , we may assume  $\zeta$  is not an eigenvalue of  $X$  without loss of generality. Therefore, we obtain

$$q(\zeta) = -p(\zeta) \text{tr}(\zeta I - X)^{-1} Z \tag{7.5}$$

$$= -p(\zeta) \text{tr} P (\zeta I - J)^{-1} P^{-1} Z \tag{7.6}$$

$$= -p(\zeta) \text{tr}(\zeta I - J)^{-1} V \tag{7.7}$$

$$= -p(\zeta) \sum_{j=1}^p \text{tr} \left( (\zeta - \mu_j) I - N^{(j)} \right)^{-1} V^{(jj)}. \tag{7.8}$$

The proof is completed by using (7.3) and noting that

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

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

---

<sup>2</sup> Equivalently, via the ordinary chain rule, the complex gradient of  $\det(M)$  is  $(\det(M)M^{-1})^*$ , for any nonsingular matrix  $M$ .

The next result concerns the subderivative of the abscissa map  $a$ . Recall that the subderivative was defined in (1.8).

**Theorem 7.1.** *Let  $p(\zeta)$ ,  $q(\zeta)$ ,  $V^{(jj)}$  and  $N^{(j)}$  be defined as in Lemma 7.1. Then*

$$da(p(\zeta))(q(\zeta)) = \infty \tag{7.9}$$

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

$$\operatorname{Re} \operatorname{tr}(N^{(j)} V^{(jj)}) \leq 0, \quad \operatorname{Im} \operatorname{tr}(N^{(j)} V^{(jj)}) = 0, \tag{7.10}$$

$$\operatorname{tr} \left( (N^{(j)})^\ell V^{(jj)} \right) = 0, \quad \ell = 2, \dots, n^{(j)} - 1. \tag{7.11}$$

On the other hand, if (7.10) and (7.11) hold for all  $j \in \mathcal{A}$ , then

$$da(p(\zeta))(q(\zeta)) = \max \left\{ \frac{\operatorname{Re} \operatorname{tr} V^{(jj)}}{m^{(j)}} : j \in \mathcal{A} \right\}. \tag{7.12}$$

*Proof.* The proof is a consequence of [7, Corollary 1.7], using Lemma 7.1.

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

**Theorem 7.2.** *Let  $X$  have Jordan form (4.1), (4.2). Then  $\hat{\partial}\alpha(X)$ , the set of regular subgradients of the spectral abscissa  $\alpha$  at  $X$ , is the set of matrices  $Y$  satisfying (4.7), (4.8), (6.4), (6.8) and (6.10).*

*Proof.* That these conditions are necessary for  $Y$  to be a regular subgradient was proved in Theorem 4.2, Theorem 5.2, Theorem 5.3, and Theorem 6.1. Now suppose that  $Y$  satisfies these conditions. We must prove that  $Y$  is a regular subgradient, i.e., using (1.9), that

$$\langle Y, Z \rangle \leq d\alpha(X)(Z), \quad \forall Z \in \mathbf{M}^n. \tag{7.13}$$

We first apply the basic chain rule of [23, Theorem 10.6] to obtain the subderivative inequality

$$d\alpha(X)(Z) = d(a \circ \Phi)(X)(Z) \geq da(\Phi(X))(\nabla\Phi(X)Z),$$

where, following [23], we use  $\nabla\Phi$  to denote the Jacobian of the differentiable map  $\Phi$ . Let  $p(\zeta)$  and  $q(\zeta)$  be defined by (7.3) and (7.4). Since  $\Phi(X) = p(\zeta)$  and  $\nabla\Phi(X)Z = \Phi'(X; Z) = q(\zeta)$ , we have

$$d\alpha(X)(Z) \geq da(p(\zeta))(q(\zeta)). \tag{7.14}$$

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

$$\langle Y, Z \rangle = \langle W, V \rangle = \sum_{j \in \mathcal{A}} \sum_{\ell=1}^{n^{(j)}} \operatorname{Re} \left( \overline{\theta_\ell^{(j)}} \operatorname{tr}(N^{(j)})^{\ell-1} V^{(jj)} \right). \tag{7.15}$$

It follows from Theorem 7.1 that if any of (7.10), (7.11) are violated, then (7.9) must hold, and so (7.14) shows that (7.13) holds trivially. On the other hand, suppose that

(7.10), (7.11) hold for all  $j \in \mathcal{A}$ , implying that (7.12) holds. Using these conditions, together with (7.15), (6.8) and (6.10), we have

$$\begin{aligned} \langle Y, Z \rangle &= \sum_{j \in \mathcal{A}} \left( \theta_1^{(j)} \operatorname{Re} \operatorname{tr} V^{(jj)} + \operatorname{Re} \theta_2^{(j)} \operatorname{Re} \operatorname{tr} (N^{(j)} V^{(jj)}) \right) \\ &\leq \sum_{j \in \mathcal{A}} m^{(j)} \theta_1^{(j)} \frac{\operatorname{Re} \operatorname{tr} V^{(jj)}}{m^{(j)}} \\ &\leq da(p(\zeta))(q(\zeta)). \end{aligned}$$

Combining this with (7.14) gives (7.13), as desired.

It follows from Theorem 7.2 that if  $Y$  is a regular subgradient of  $\alpha$  at  $X$  and  $\mu_j$  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 7.1.** *Suppose  $X$  is the  $n$  by  $n$  zero matrix. Then the spectral abscissa  $\alpha$  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 [22, 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 Sect. 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  $\theta_2^{(j)}, \dots, \theta_{n^{(j)}}^{(j)}$  for  $j \in \mathcal{A}$ , except  $\operatorname{Re} \theta_2^{(j)} \geq 0$ . More specifically, we have:

**Corollary 7.2.** *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).

### 8. The subgradients and horizon subgradients of the spectral abscissa

In this section we consider *all* subgradients and horizon subgradients of the spectral abscissa  $\alpha$ , 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  $\alpha$  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  $\alpha$  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 4.1, Theorem 5.1 and Theorem 6.1, 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.

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 5.1 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 5.3 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).

**Theorem 8.1.** *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  $\alpha$  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  $X_i \rightarrow X$  and  $Y_i \in \hat{\partial}\alpha(X_i)$  with  $Y_i \rightarrow Y$ . We wish to show that  $\theta_2^{(j)}$ , the subdiagonal entry in the Toeplitz matrix  $W^{(j)}$ , satisfies (6.10) for all eigenvalues  $\mu_j$  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  $\theta_2^{(j)}$ , with  $\text{Re } \theta_2^{(j)} < 0$ .

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

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

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  $D_i \rightarrow I$ , each differing from  $I$  only in the  $r$ th 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

$$\langle Y_i, Z_i \rangle \rightarrow \langle P^* Y P^{-*}, V \rangle = \langle W, V \rangle = -\text{Re } \theta_2^{(j)} > 0. \tag{8.1}$$



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

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

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

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

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

$$\operatorname{Re} \tau_{\pm}(t) = \operatorname{Re} v_i^{(1)}$$

for all  $t \geq 0$  and 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, for any fixed  $i$ , we have

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

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

$$\operatorname{Re} v_i^{(1)} - t \frac{\operatorname{Re} v_i^{(1)} - \operatorname{Re} v_i^{(2)}}{|v_i^{(1)} - v_i^{(2)}|^2} + o(t).$$

Consequently, in both cases (1) and (2), we have, for all  $i$ ,

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

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  $\alpha$  at  $X_i$ , and we have our desired contradiction.

When  $Y$  is a horizon subgradient, the proof is almost identical. Instead of  $Y_i \rightarrow Y$  we have  $s_i Y_i \rightarrow Y$ , 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.

We are now ready for the main result of this section, characterizing regularity of the spectral abscissa. Recall from Sect. 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 8.2.** *The spectral abscissa  $\alpha$  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  $\alpha$  at  $X$ . Since all active eigenvalues are nonderogatory,  $W$  in (4.4) satisfies (4.5) as well as (6.4). Corollary 5.1 shows that (6.8) must hold, and Theorem 8.1 shows that (6.10) must also be satisfied. Furthermore, Theorem 7.2 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 7.2), and subdifferential regularity is proved.

For the converse, suppose that  $X$  has an active derogatory eigenvalue  $\mu_j$ , so that  $q^{(j)} \geq 2$ . Let  $\beta_i$  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 7.2, the regular subgradients of  $\alpha$  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  $\alpha$  at  $X$ . However,  $\tilde{E}$  is not a regular subgradient of  $\alpha$  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  $\mu_j$ .) Therefore,  $\alpha$  is not subdifferentially regular at  $X$ .

*Example 8.1.* Let  $X$  be the Jordan block

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

Theorem 7.2 shows that

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

and therefore, by definition,

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

It is instructive to consider a specific sequence

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

where  $\epsilon_i > 0$ ,  $\epsilon_i \rightarrow 0$  as  $i \rightarrow \infty$ . 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 7.2, that

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

Let  $Y_i \in \hat{\partial}\alpha(X_i)$ , with  $\sigma_i$  being the corresponding value of  $\sigma$  in (8.3). If  $Y_i \rightarrow Y$ , then its bottom left entry must converge to a limit: any imaginary limit is possible, but since  $\epsilon_i \rightarrow 0$ , we must have  $\sigma_i \rightarrow \frac{1}{2}$ . Thus, the subgradient  $Y$  is regular (an element of  $\hat{\partial}\alpha(X)$ ). On the other hand, if  $s_i Y_i \rightarrow Y$ , 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 8.2 shows that these properties hold for every sequence  $X_i \rightarrow X$ , i.e.,  $\alpha$  is subdifferentially regular at  $X$ .

Theorem 8.2 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  $f = \max \text{Re}$ . This is in contrast with the Hermitian case discussed in [18].

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

**Theorem 8.3.** *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*

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

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  $\alpha$  at  $X$  consists of the matrices  $Y$  satisfying the same block condition on  $W$ , with  $W^{(j)} = 0$  if  $j \notin \mathcal{A}$ , and 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, with  $W^{(j)} = 0$  for  $j \notin \mathcal{A}$ , and although  $W$  may not be diagonalizable, there exists a sequence  $W_i \rightarrow W$  with  $W_i$  diagonalizable, say

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

where  $T_i$  is block diagonal and  $D_i$  is diagonal, with  $T_i^{(j)} = I$  and  $D_i^{(j)} = 0$  for  $j \notin \mathcal{A}$ . 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

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

where  $K_i$  is diagonal with distinct diagonal entries, all having the same real part, and converging to zero. Since the active blocks of  $J$  are multiples of the identity and  $T_i$  is block diagonal,  $X_i \rightarrow X$  if  $K_i$  is chosen to converge to zero sufficiently fast (relative to  $\|T_i\| \|T_i^{-1}\|$ ). Thus, by Theorem 7.2,

$$Y_i = P^{-*} W_i P^* = P^{-*} T_i^{-*} D_i T_i^* P^* \tag{8.4}$$

is a regular subgradient of  $\alpha$  at  $X_i$ , for all  $i$ . Since  $Y_i \rightarrow Y$ , it follows that  $Y$  is a subgradient of  $\alpha$  at  $X$ . The proof for the horizon subgradients is almost identical: now the eigenvalues of  $W$  are zero, so  $D_i \rightarrow 0$ , 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  $s_i Y_i \rightarrow Y$ , the latter is a horizon subgradient of  $\alpha$  at  $X$ .

In particular, we have:

**Corollary 8.2.** *Suppose  $X$  is the  $n$  by  $n$  zero matrix. Then the set of subgradients of the spectral abscissa  $\alpha$  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  $\lambda$  is Lipschitz on  $\mathbf{H}^n$  (see e.g. [12, Theorem II.6.10]). Variational properties of  $\lambda$  on  $\mathbf{H}^n$  have been extensively studied, especially in the recent work of Lewis [18]. Indeed, the general results given here in Sects. 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 Sect. 2 and the Jordan form of Sect. 4 are the same. More specifically, the following is true:

- the eigenvalues  $\mu_j, j = 1, \dots, p$ , 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  $\text{Diag}(\lambda(X))$

Furthermore, since all eigenvalues are real, the spectral abscissa of  $X$  is the maximum eigenvalue  $\mu_1$ , and hence the active set  $\mathcal{A}$  defined in (7.2) is  $\{1\}$ ; i.e., only  $\mu_1$  is active.

Let

$$\omega : \mathbf{H}^n \rightarrow \mathbf{R}$$

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

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

where, as earlier, each  $W^{(j)}$  is  $m^{(j)} \times m^{(j)}$ . Then, as is well known, e.g. [16], 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  $\alpha$  defined on  $\mathbf{M}^n$ . Let  $\hat{\mathbf{H}}_n$  be the subspace of  $\mathbf{M}^n$  consisting of all Hermitian matrices, and let  $\iota : \mathbf{H}^n \rightarrow \mathbf{M}^n$  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  $\iota$  is the linear operator which maps a matrix to its Hermitian part, i.e.

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

Since

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

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 8.3 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 7.2, 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 [23, 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  $\alpha$  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:

$$\max_{X \in \mathbf{M}^n} \quad \langle C, X \rangle \tag{10.1}$$

$$\text{subject to} \quad \langle A_k, X \rangle = b_k, \quad k = 1, \dots, m$$

$$\text{and} \quad \alpha(X) \leq 0, \tag{10.2}$$

where  $C \in \mathbf{M}^n$ ,  $A_k \in \mathbf{M}^n$ ,  $k = 1, \dots, m$ , and  $b \in \mathbf{R}^m$ . We call this a *semistable program*. The second constraint imposes the condition that all eigenvalues of  $X$  lie in the left-half plane or on the imaginary axis. We call such matrices *semistable*. Semistable programs have many potential applications in stability and control theory. 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 maximum is NP-hard [2,21]. 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 maximizer 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

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

where

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

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

$$\tau(x) = \begin{cases} 0 & , \text{ if } x = 0, \\ +\infty & , \text{ otherwise,} \end{cases}$$

and

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

The theorem is now proved by applying the composite Lagrange multiplier rule [23, Example 10.8 together with Proposition 10.3]. The proof uses the fact that 0 is never an element of  $\partial\alpha(X)$  (see Corollary 8.1.) There are two cases: one where the constraint qualification described in [23, Example 10.8] holds, and one where it does not hold. The conclusion follows in both cases, with  $\eta = 1$  in the first case and  $\eta = 0$  in the second case.

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  $\eta$  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 negative real part. Such a matrix  $X$  is a local maximizer in the trivial case that  $C$  lies in the range of the  $A_k, k = 1, \dots, m$ . However, this case is of little interest, since the spectral abscissa constraint (10.2) is not active. Accordingly, let us change the definition of active set from (7.2) to one more suitable for semistable programming, namely

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

The two definitions are equivalent except in the trivial case that the spectral abscissa constraint is inactive. We then have

**Theorem 10.2 (First-order necessary conditions, Fritz John type, details).** *Suppose that a matrix  $X$  is a local maximizer of (10.1)–(10.2), with  $X$  having Jordan form (4.1), (4.2). Then there exist 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  $Y$ . In the trivial case that (10.2) is inactive, we take  $Y = 0$ .

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 maximizer 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 left-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 left-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 2.1 and Theorem 10.2. The second diagonal block of  $S$  vanishes because of Lemma 2.2 and Corollary 4.1: (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  $(\text{diag}(R))_\ell \text{diag}(S)_\ell$ ,  $\ell = 1, \dots, n$ . Thus we get imaginary eigenvalues for the first diagonal block (imaginary times real) and, more specifically, zero eigenvalues for the second diagonal block (complex times zero).



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  $\Lambda_2$  diagonal, real and strictly negative, and  $\Omega_1$  diagonal, real, and nonnegative. Thus, Theorem 10.3 generalizes the well known notion of complementarity in semidefinite programming.

*Acknowledgements.* It is a pleasure to thank Adrian Lewis for his interest in this work and for many helpful conversations, leading among other things to simplified proofs of Theorem 7.2 and Lemma A.2. We also thank the referees for reading the paper and making several helpful suggestions. This work was supported in part by U.S. National Science Foundation Grants DMS-9971852 and CCR-9731777, and U.S. Department of Energy Contract DE-FG02-98ER25352.

### A. The Schur factorization

The lemmas presented here are variations on standard results for the Schur factorization [9, Sect. 2.3]. They are surely known, but we were unable to find them in the literature.

**Lemma A.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  $\mu$  be an eigenvalue of  $A$  and set  $\mathcal{E}_\mu = \text{Null}(\mu I - A)$ . 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 [9, 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 [9] to establish the result.

The other result that we need concerns the continuity of the Schur factorization of a perturbation of a Jordan block.

**Lemma A.2.** *Let  $J \in \mathbf{M}^n$  be an upper Jordan block, i.e., a single block of the form  $J_k^{(j)}$  in (4.2). For all  $\epsilon > 0$ , there exists  $\delta > 0$  such that, if  $\|E\| < \delta$ , there exists a unitary matrix  $U$  with  $\|U - I\| < \epsilon$  and  $U^*(J + E)U$  upper triangular.*

*Proof.* This proof is due to A.S. Lewis; it is more elementary than our original proof. Suppose that the result does not hold. Then there exists  $\epsilon > 0$  and  $E_i \rightarrow 0$  such that, for each  $i$ , if  $U$  is unitary with  $U^*(J + E_i)U$  upper triangular, then  $\|U - I\| \geq \epsilon$ . Choose  $U_i$  unitary such that  $U_i^*(J + E_i)U_i$  is upper triangular for all  $i$ . By compactness, we

can assume without loss of generality that  $U_i$  converges to a limit  $U$ , which must be unitary and such that  $U^*JU$  is upper triangular. Lemma A.3 (below) shows that  $U$  must therefore be diagonal. Therefore,  $UU_i^*(J + E_i)U_iU^*$  is upper triangular for all  $i$ . This is a contradiction, since  $UU_i^* \rightarrow I$ .

**Lemma A.3.** *Suppose that  $T = P^{-1}JP$  is upper triangular, where  $P$  is nonsingular and  $J$  is an upper Jordan block as in Lemma A.2. Then  $P$  is also upper triangular. If  $P$  is also unitary, it must be diagonal.*

*Proof.* Since the diagonal of  $J$  is constant, we may take it to be zero without loss of generality. Consequently, the eigenvalues of  $T$  are zero, and so  $T$  must be strictly upper triangular. From  $JP = PT$ , we have

$$p_{i,k} = \sum_{j < k} p_{i-1,j} t_{j,k} \quad \text{for } i > 1.$$

It follows by induction on  $k$  that  $p_{i,k} = 0$  for all  $i > k \geq 1$ . Thus,  $P$  is upper triangular. It is well known (and easily proved by induction) that a unitary triangular matrix must be diagonal.

## References

1. Arnold, V.I. (1971): On matrices depending on parameters. *Russ. Math. Surveys* **26**, 29–43
2. Blondel, V., Tsitsiklis, J.N. (1997): NP-Hardness of some linear control design problems. *SIAM J. Control Optim.* **35**, 2118–2127
3. Burke, J.V., Lewis, A.S., Overton, M.L. (2001): Optimal stability and eigenvalue multiplicity. *Foundations Comput. Math.*, DOI: 10.1007/s102080010008
4. Burke, J.V., Overton, M.L. (1992): On the subdifferentiability of functions of a matrix spectrum I: Mathematical foundations; II: Subdifferential formulas. In: Giannessi, F., ed., *Nonsmooth Optimization: Methods and Applications*, pp. 11–29. Gordon and Breach, Philadelphia, 1992. Proceedings of a conference held in Erice, Italy, June 1991
5. Burke, J.V., Overton, M.L. (1992): Stable perturbations of nonsymmetric matrices. *Linear Alg. Appl.* **171**, 249–273
6. Burke, J.V., Overton, M.L. (1994): Differential properties of the spectral abscissa and the spectral radius for analytic matrix-valued mappings. *Nonlinear Analysis, Theory, Methods and Applications* **23**, 467–488
7. Burke, J.V., Overton, M.L. (2001): Variational analysis of the abscissa map for polynomials. *SIAM J. Control Optim.* **39**, 1651–1676
8. Clarke, F.H. (1973): Necessary conditions for nonsmooth problems in optimal control and the calculus of variations. PhD thesis, University of Washington
9. Horn, R.A., Johnson, C.R. (1985): *Matrix Analysis*. Cambridge University Press, Cambridge, U.K.
10. Horn, R.A., Johnson, C.R. (1991): *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, U.K.
11. Ioffe, A.D. (1981): Sous-différentielles approchées de fonctions numériques. *Comptes Rendus de l'Académie des Sciences de Paris* **292**, 675–678
12. Kato, T. (1984): *Perturbation Theory for Linear Operators*. Springer-Verlag, New York, second edition
13. Kruger, A.Y., Mordukhovich, B.S. (1980): Extremal points and the Euler equation in nonsmooth analysis. *Doklady Akademii Nauk BSSR (Belorussian Academy of Sciences)* **24**, 684–687
14. Lancaster, P., Tismenetsky, M. (1985): *The Theory of Matrices*. Academic Press, New York, London
15. Lax, P.D. (1997): *Linear Algebra*. John Wiley, New York
16. Lewis, A.S. (1996): Convex analysis on the Hermitian matrices. *SIAM J. Optim.* **6**, 164–177
17. Lewis, A.S. (1996): Derivatives of spectral functions. *Math. Oper. Res.* **21**, 576–588
18. Lewis, A.S. (1999): Nonsmooth analysis of eigenvalues. *Math. Program.* **84**, 1–24
19. Mordukhovich, B.S. (1976): Maximum principle in the problem of time optimal response with nonsmooth constraints. *J. Appl. Math. Mech.* **40**, 960–969

20. Moro, J., Burke, J.V., Overton, M.L. (1997): On the Lidskii-Vishik-Lyusternik perturbation theory for eigenvalues of matrices with arbitrary Jordan structure. *SIAM J. Matrix Anal. Appl.* **18**, 793–817
21. Nemirovskii, A. (1993): Several NP-hard problems arising in robust stability analysis. *Math. Control Signals Systems* **6**, 99–105
22. Overton, M.L., Womersley, R.S. (1988): On minimizing the spectral radius of a nonsymmetric matrix function – optimality conditions and duality theory. *SIAM J. Matrix Anal. Appl.* **9**, 473–498
23. Rockafellar, R.T., Wets, R.J.B. (1998): *Variational Analysis*. Springer-Verlag, New York