ROBUST REGULARIZATION

A.S. Lewis*

September 18, 2002

Key words: robust optimization, nonsmooth analysis, locally Lipschitz, regularization, semi-algebraic, pseudospectrum, robust control, semidefinite representable, prox-regularity
AMS 2000 Subject Classification:
Primary: 90C30, 49K40
Secondary: 65K10, 15A18

Abstract

Given a real function on a Euclidean space, we consider its "robust regularization": the value of this new function at any given point is the maximum value of the original function in a fixed neighbourhood of the point in question. This construction allows us to impose constraints in an optimization problem *robustly*, safeguarding a constraint against unpredictable perturbations in variables or data. After outlining some examples, we consider in particular a function that is locally Lipschitz on the complement of a suitably well-behaved (for example, semi-algebraic or prox-regular) small set, and satisfies a growth condition near the set. We show that, around any given point, the robust regularization is eventually locally Lipschitz once the size of the neighbourhood is sufficiently small. Our result applies in particular to the pseudospectral abscissa of a square matrix, a useful function in robust stability theory.

*Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. `aslewis@sfu.ca`, `http://www.cecm.sfu.ca/~aslewis`. Research supported by NSERC.

1 Introduction and notation

Several concerns motivate the paradigm of robust optimization. The data of the original optimization problem may be uncertain, either due to inaccurate measurements or unpredictable inputs (in robust control theory, for example—see [4]). Numerical efforts to find an optimal choice of the variables have finite precision. Even after choosing the variables, implementing the choice precisely in a concrete model may be impossible (design of digital filters being a typical example [11]).

As a consequence of these unpredictable perturbations, we can imagine evaluating a typical function appearing in a concrete optimization problem not at a precise point but somewhere in a neighbourhood of the nominal point of interest. Since practical optimization problems may be very ill-conditioned, this fuzziness can have disasterous consequences for the quality of an optimal solution. For an extensive exposition of these issues in robust optimization, particularly in the case of linear and quadratic programming, see the recent monograph [2].

As a tool for "robust" variational analysis, we consider the following regularization process. We fix a finite-dimensional normed space $\mathbf{E}$, and consider a function $f : \mathbf{E} \to [-\infty, +\infty]$. We also fix a compact convex neighbourhood C of the origin in $\mathbf{E}$. Typically C is the closed unit ball for some norm. We denote the closed unit ball in $\mathbf{E}$ by B: with no loss of generality, we could choose $\mathbf{E} = \mathbf{R}^n$ and B as the Euclidean ball. For a real *regularization parameter* $\epsilon \geq 0$, we now define the *robust regularization* $f_\epsilon : \mathbf{E} \to \mathbf{R}$ by

$$f_\epsilon(x) = \sup\{f(y) : y - x \in \epsilon C\}.$$

We could interpret this construction, for example, as deriving a version of f that is robust relative to "implementation" errors in the variable x. If we wish to impose the constraint $f(x) \leq 0$ in some optimization problem, then by instead requiring $f_\epsilon(x) \leq 0$, we restrict attention to points x satisfying $f(x) \leq 0$ "robustly": any perturbation to x from the small set ϵC leaves the original constraint satisfied. Similarly, if we want to minimize the function f over some feasible region, then by instead minimizing f_ϵ over the same set, we obtain a "robust" solution of the original problem: a solution that is optimized with respect to all possible perturbations from ϵC to the chosen solution.

As a simple example to keep in mind, consider the function $f : \mathbf{R} \to \mathbf{R}$

defined by

$$f(x) = \begin{cases} -x & \text{if } x < 0 \\ \sqrt{x} & \text{if } x \geq 0. \end{cases} \tag{1.1}$$

If $C = [-1, 1]$, a quick calculation shows

$$f(x) = \begin{cases} \epsilon - x & \text{if } x < \alpha \\ \sqrt{\epsilon + x} & \text{if } x \geq \alpha, \end{cases} \tag{1.2}$$

where

$$\alpha = \frac{1 + 2\epsilon - \sqrt{1 + 8\epsilon}}{2} > -\epsilon.$$

In this work, we begin by discussing some elementary properties of the robust regularization (which can be considered a kind of "deconvolution" [12]). We describe some simple examples, noting in particular that the robust regularization of a strictly convex quadratic function is "semi-definite representable" in the sense of [2], and hence applicable in tractable computational approaches to convex optimization.

For nonconvex functions f, where we can no longer hope to say much about the strict computational tractability of minimization problems involving f, robust regularization nonetheless remains compelling philosophically. It also has a surprising benefit from an analytic and numerical perspective: it often transforms non-Lipschitz functions f into regularizations with enhanced Lipshitz properties. For example, the nonlipschitz function (1.1) has a Lipschitz robust regularization (1.2).

In the main part of this paper, we consider a function f that is locally Lipschitz on the complement of a well-behaved exceptional set A, and satisfies a certain growth condition as we move away from the set A. We show that, around any given point, the robust regularization f_ϵ is locally Lipschitz for all sufficiently small values of the regularization parameter ϵ. The function (1.1) is an example (where we choose the set A to be the origin).

In particular, our result applies, on the space of square complex matrices, to the spectral abscissa function (the largest real part of an eigenvalue). Although this function is highly non-Lipschitz, we show that its robust regularization, the so-called *ϵ-pseudospectral abscissa*, is locally Lipschitz around any given nonderogatory matrix, for all sufficiently small $\epsilon > 0$. Since the pseudospectral abscissa is relatively easy to calculate [6, 7], and indeed is interesting in its own right [17], this Lipschitz property is an attractive feature numerically and analytically.

2 Quadratic examples

We begin with a more-or-less standard example in the spirit of [4, 2]. We write $X \succeq 0$ for a real symmetric matrix X if X is positive semidefinite.

Theorem 2.1 (Euclid norm) *For any real m-by-n matrix A and vector $b \in \mathbf{R}^m$, consider the function $g : \mathbf{R}^n \to \mathbf{R}$ defined by*

$$g(x) = \|Ax + b\|_2,$$

with robust regularization

$$g_\epsilon(x) = \max\{g(y) : \|y - x\|_2 \leq \epsilon\}.$$

Then the following properties are equivalent for any point $(x, t) \in \mathbf{R}^n \times \mathbf{R}$:

(i) $t \geq g_\epsilon(x)$*;*

(ii) *there exists a real μ such that*

$$\begin{bmatrix} tI_m & Ax + b & \epsilon A \\ (Ax + b)^T & t - \mu & 0 \\ \epsilon A^T & 0 & \mu I_n \end{bmatrix} \succeq 0.$$

Proof Applying [2, Thm 4.5.60] shows $t \geq g_\epsilon(x)$ holds if and only if there exist real s and μ satisfying

$$\begin{array}{rcl} t - s & \geq & 0 \\ \begin{bmatrix} sI_m & Ax + b & \epsilon A \\ (Ax + b)^T & s - \mu & 0 \\ \epsilon A^T & 0 & \mu I_n \end{bmatrix} & \succeq & 0, \end{array}$$

and the result now follows immediately. □

Since the matrix in property (ii) above is an affine function of the variables x, t and μ, it follows that the robust regularization g_ϵ is "semidefinite-representable", in the language of [2]. This result allows us to use g_ϵ in building tractable representations of convex optimization problems as semidefinite programs.

An easy consequence of the above result is a representation for the robust regularization of any strictly convex quadratic function.

Corollary 2.2 (quadratics) *For any real positive definite n-by-n matrix H, vector $c \in \mathbf{R}^n$, and scalar d, consider the function $h : \mathbf{R}^n \to \mathbf{R}$ defined by*

$$h(x) = x^T H x + 2c^T x + d.$$

with robust regularization

$$h_\epsilon(x) = \max\{h(y) : \|y - x\|_2 \le \epsilon\}.$$

Then the following properties are equivalent for any point $(x, t) \in \mathbf{R}^n \times \mathbf{R}$:

(i) $t \ge h_\epsilon(x)$;

(ii) *there exists reals s and μ such that*

$$\begin{array}{rcl} t - s^2 + c^T H^{-1} c - d & \ge & 0 \\ \begin{bmatrix} sI_n & H^{1/2}x + H^{-1/2}c & \epsilon H^{1/2} \\ (H^{1/2}x + H^{-1/2}c)^T & s - \mu & 0 \\ \epsilon H^{1/2} & 0 & \mu I_n \end{bmatrix} & \succeq & 0. \end{array}$$

Proof Clearly $t \ge h_\epsilon(x)$ if and only if

$$\|y - x\|_2 \le \epsilon \quad \Rightarrow \quad \|H^{1/2}y + H^{-1/2}c\|_2^2 \le t - d + c^T H^{-1} c.$$

This property in turn is equivalent to the existence of a real s satisfying

$$\begin{array}{rcl} s^2 & \le & t - d + c^T H^{-1} c \ \text{ and} \\ \|y - x\|_2 \le \epsilon & \Rightarrow & \|H^{1/2}y + H^{-1/2}c\|_2 \le s, \end{array}$$

and the result now follows from the preceding theorem. $\square$

Since the quadratic inequality

$$t - s^2 + c^T H^{-1} c - d \ge 0$$

is semidefinite-representable, so is the robust regularization h_ϵ.

3 Elementary properties

We begin our general investigation of robust regularization by collecting some elementary properties. Throughout this section we consider functions $f : E \to [-\infty, +\infty]$.

Proposition 3.1 (convexity) *If the function f is convex then so is the robust regularization f_ϵ for any $\epsilon \geq 0$.*

Proof The robust regularization is a supremum of convex functions:

$$f_\epsilon(x) = \sup\{f(x+z) : z \in \epsilon C\}.$$

The result follows immediately. □

An analogous argument holds for lower semicontinuity.

Proposition 3.2 (closedness) *If the function f is closed then so is the robust regularization f_ϵ for any $\epsilon \geq 0$.*

Suppose we are interested in a set defined by several inequality constraints:

$$f^\gamma(x) \leq 0 \text{ for all } \gamma \in \Gamma.$$

We would naturally describe a point x as "robustly" feasible in our context if every point in the set $x + \epsilon C$ satisfies all the constraints, or equivalently, if

$$f^\gamma_\epsilon(x) \leq 0 \text{ for all } \gamma.$$

We could also describe the original set equivalently by the single constraint

$$\sup_\gamma f^\gamma(x) \leq 0,$$

and in this case robust feasibility would naturally mean

$$\left(\sup_\gamma f^\gamma\right)_\epsilon (x) \leq 0.$$

These two notions of robust feasibility are equivalent, due to the following easy result.

Proposition 3.3 (suprema) *For any collection of functions f^γ indexed by γ, we have*

$$\left(\sup_\gamma f^\gamma\right)_\epsilon = \sup_\gamma f^\gamma_\epsilon.$$

Proof The left hand side evaluated at any point x is

$$\sup_{z \in \epsilon C} \sup_\gamma f^\gamma(x+z)$$

which in turn equals the right hand side, by interchanging the suprema. □

We denote the *attaining set* where the supremum defining the robust regularization

$$f_\epsilon(x) = \sup\{f(y) : y - x \in \epsilon C\}$$

is attained by $M_\epsilon(x)$. Clearly, if f is continuous around x then this set is nonempty and compact for all small $\epsilon \geq 0$.

Our next result notes that the robust regularization really is an approximation of the underlying function.

Proposition 3.4 (approximation) *If the function f is continuous around the point x, then*

$$\lim_{\epsilon \downarrow 0} f_\epsilon(x) = f(x).$$

Proof This is a simple consequence of the continuity of f and the boundedness of C. □

We next make the routine observation that the robust regularization of a continuous function is itself continuous.

Proposition 3.5 (continuity) *Suppose the function f is continuous on the open set $U \subset \mathbf{E}$, and the subset $V \subset U$ is compact. Then for all small $\epsilon \geq 0$ (specifically, providing $V + \epsilon C \subset U$), the robust regularization f_ϵ is continuous on V.*

Proof Compactness shows $V + \epsilon C \subset U$ for all small $\epsilon \geq 0$, so fix such an ϵ. Suppose f is not continuous at the point $x \in V$ so there exists a sequence of points $x_r \to x$ in V and a real $\delta > 0$ such that

$$|f_\epsilon(x_r) - f_\epsilon(x)| \geq \delta \text{ for all } r.$$

For all large r, continuity implies the attaining set $M_\epsilon(x_r)$ is nonempty, so we can choose an element y_r. Since the sequence (x_r) is bounded, and $y_r - x_r$ lies in the bounded set ϵC, the sequence (y_r) is also bounded, so without loss of generality we can suppose it converges to some point $y_\infty \in \mathbf{E}$. Since C is closed, we deduce $y_\infty - x \in \epsilon C$, so

$$f_\epsilon(x) \geq f(y_\infty) = \lim f(y_r) = \lim f_\epsilon(x_r).$$

Fix any point $\bar{y} \in M_\epsilon(x)$, and notice that for all r we have

$$(\bar{y} + x_r - x) - x_r \in \epsilon C.$$

Since f is continous at $\bar{y}$, we deduce

$$f_\epsilon(x_r) \geq f(\bar{y} + x_r - x) \to f(\bar{y}) = f_\epsilon(x).$$

In combination with the paragraph above, this shows $\lim f_\epsilon(x_r) = f_\epsilon(x)$, which is the desired contradiction. $\Box$

The next argument is also straightforward.

Proposition 3.6 (upper semicontinuity) *Suppose the function f is continuous on the open set $U \subset \mathbf{E}$, and the subset $V \subset U$ is compact. Then for all small $\epsilon \geq 0$ (specifically, providing $V + \epsilon C \subset U$) the set-valued "attainment" map $M_\epsilon : V \to U$ has closed graph with nonempty compact images, and is upper semicontinuous: that is, for any point $\bar{x} \in V$, any open neighbourhood of the image $M_\epsilon(\bar{x})$ also contains $M_\epsilon(x)$ for all points $x \in V$ close to $\bar{x}$.*

Proof Continuity shows that the images of M_ϵ are nonempty and compact. We next show that M_ϵ has closed graph. To see this, consider sequences $x_r \to \bar{x}$ in V and $y_r \to \bar{y}$ in U satisfying $y_r \in M_\epsilon(x_r)$ for all r. Since $y_r - x_r$ lies in the closed set ϵC for all r, we deduce $\bar{y} - \bar{x} \in \epsilon C$. Furthermore, by the previous proposition we know

$$f(\bar{y}) = \lim_r f(y_r) = \lim_r f_\epsilon(x_r) = f_\epsilon(\bar{x}),$$

so $\bar{y} \in M_\epsilon(\bar{x})$, as required.

Upper semicontinuity follows from the fact that M_ϵ is locally bounded. We spell out the argument, for completeness. If the result fails, there is an open set W containing $M_\epsilon(\bar{x})$, a sequence $x_r \to \bar{x}$ in V, and points $y_r \in M_\epsilon(x_r)$ lying outside W for all r. As in the previous result, the sequence y_r must be bounded, so without loss of generality we can suppose it converges to some point $\bar{y}$, which must also lie outside W. But since M_ϵ is closed, we have the contradiction $\bar{y} \in M_\epsilon(\bar{x})$. □

The last result of this section shows that we can approximate strict local minimizers of the original function by minimizing the robust regularization.

Proposition 3.7 (local minimizers) *If the function $f : \mathbf{E} \to [-\infty, +\infty]$ has a strict local minimizer relative to the closed set $F \subset \mathbf{E}$ at the point $\bar{x} \in F$, and is continuous around $\bar{x}$, then for all small $\epsilon \geq 0$, the robust regularization f_ϵ has a local minimizer relative to F near $\bar{x}$.*

Proof Since $\bar{x}$ is a strict local minimizer, we can choose arbitrarily small $\delta > 0$ such that $f(x) > f(\bar{x})$ whenever $x \in F$ and $\|x - \bar{x}\| = \delta$. Hence by continuity and compactness there exists a real $\alpha > f(\bar{x})$ such that

$$f_\epsilon(x) \geq f(x) > \alpha \text{ whenever } \|x - \bar{x}\| = \delta,\ x \in F.$$

But by Proposition 3.4 (approximation) we know $f_\epsilon(\bar{x}) < \alpha$ for all small ϵ, and providing δ is small, Proposition 3.5 (continuity) further shows f_ϵ is continuous on $(\bar{x} + \delta B)$. Hence for small ϵ, the minimum value of the continuous function f_ϵ on the compact set $(\bar{x} + \delta B) \cap F$ must be strictly less than α, and so must be attained at a local minimizer of f_ϵ relative to F. □

4 Lipschitz behaviour

We plan to investigate functions f that are locally Lipschitz except on a small exceptional set. Our main tool is the straightforward idea below: under reasonable conditions, the robust regularization f_ϵ inherits the local Lipschitz property.

Proposition 4.1 (Lipschitz behaviour) *Suppose the function f is continuous on the open set $U \subset \mathbf{E}$, and locally Lipschitz on the open set $\Omega \subset U$. Given any point $\bar{x} \in U$ and any small $\epsilon \geq 0$ (specifically, providing $\bar{x} + \epsilon C \subset U$), if the corresponding attaining set satisfies $M_\epsilon(\bar{x}) \subset \Omega$, then the robust regularization f_ϵ is locally Lipschitz around $\bar{x}$.*

Proof Since the attaining set $M_\epsilon(\bar{x})$ is compact, there exists a real $\gamma > 0$ such that the compact set $M_\epsilon(\bar{x}) + 2\gamma B$ is contained in Ω. Choose a Lipschitz constant L for the function f on $M_\epsilon(\bar{x}) + 2\gamma B$. Applying Proposition 3.6 (upper semicontinuity) at $\bar{x}$, there exists a real $\mu \in (0, \gamma/2)$ such that

$$\emptyset \neq M_\epsilon(x) \subset M_\epsilon(\bar{x}) + \gamma B \text{ whenever } \|x - \bar{x}\| \leq \mu.$$

We now show that the regularization f_ϵ has Lipschitz constant L on the set $\bar{x} + \mu B$.

To see this, consider any two points $x_1, x_2 \in \bar{x} + \mu B$. Choose any point $y_1 \in M_\epsilon(x_1)$, so $f_\epsilon(x_1) = f(y_1)$. Since $y_1 - x_1 \in \epsilon C$, we know

$$(y_1 - x_1 + x_2) - x_2 \in \epsilon C,$$

so $f_\epsilon(x_2) \geq f(y_1 - x_1 + x_2)$. But $\|x_1 - x_2\| \leq 2\mu < \gamma$, so

$$y_1 - x_1 + x_2 \in M_\epsilon(x_1) + \gamma B \subset M_\epsilon(\bar{x}) + 2\gamma B.$$

Applying the Lipschitz property shows

$$f_\epsilon(x_2) \geq f(y_1 - x_1 + x_2) \geq f(y_1) - L\|x_1 - x_2\| = f_\epsilon(x_1) - L\|x_1 - x_2\|.$$

Interchanging the roles of x_1 and x_2, we deduce

$$|f(x_1) - f(x_2)| \leq L\|x_1 - x_2\|,$$

as required. □

In the next sections we develop a property guaranteeing that the attaining set is disjoint from the exceptional set of points where f is not locally Lipschitz. We can then apply the above result to deduce that the robust regularization is locally Lipschitz.

5 Nearly radial and semi-algebraic sets

The exceptional sets corresponding to functions of interest to us have a useful tangential property. Loosely speaking, given any point $\bar{x}$ in the exceptional set, at any nearby point there is a tangent vector pointing back approximately towards $\bar{x}$. The formal definition is as follows. The *(Bouligand) tangent cone* (or "contingent cone") to a set $A \subset \mathbf{E}$ at a point $\bar{x} \in A$ is the set

$$T_A(\bar{x}) = \{\lim t_r^{-1}(x_r - \bar{x}) : t_r \downarrow 0,\ x_r \to \bar{x},\ x_r \in A\}$$

(see, for example, [14]).

Definition 5.1 (nearly radial sets) A set $A \subset \mathbf{E}$ is *nearly radial at* a point $\bar{x} \in A$ if

$$\mathrm{dist}\Big(\bar{x}, x + T_A(x)\Big) = o(\|x - \bar{x}\|) \ \text{ as } x \to \bar{x} \text{ in } A$$

(where dist denotes distance). The set A is *nearly radial* if it is nearly radial at every point in A.

We contrast this definition with a stronger property introduced by [16], which is the uniform version of the same idea. This idea was called $o(1)$*-convexity* in [15].

Definition 5.2 (nearly convex sets) A set $A \subset \mathbf{E}$ is *nearly convex* at a point $\bar{x} \in A$ if

$$\mathrm{dist}\Big(y, x + T_A(x)\Big) = o(\|x - y\|) \ \text{ as } x, y \to \bar{x} \text{ in } A.$$

The set A is *nearly convex* if it is nearly convex at every point in A.

Clearly if a set is nearly convex at a point, then it is nearly radial there, but the class of nearly radial sets is considerably broader. For example, the set

$$A = \{x \in \mathbf{R}^2 : x_1 x_2 = 0\}$$

is nearly radial at the origin but not nearly convex there, since as $n \to \infty$ the points $x_n = (n^{-1}, 0)$ and $y_n = (0, n^{-1})$ approach the origin in A and yet

$$\mathrm{dist}(y_n, x_n + T_A(x_n)) = n^{-1} \neq o(\|x_n - y_n\|).$$

It is immediate that convex sets are nearly convex, and hence nearly radial. A straightforward exercise shows that smooth manifolds are also nearly convex, and hence again nearly radial. These observations are both special cases of the following result, rather analogous to [16, Thm 2.2]. A set $A \subset \mathbf{E}$ is *amenable* [14] at a point $\bar{x} \in A$ if there is an open neighbourhood V of $\bar{x}$, a $\mathcal{C}^1$ mapping $F : V \to \mathbf{R}^m$, and a closed convex set $D \subset \mathbf{R}^m$, such that

$$A \cap V = \{x \in V : F(x) \in D\} \ \text{ and } \ N_D(F(\bar{x})) \cap N(\nabla F(\bar{x})^*) = \{0\} \tag{5.3}$$

where $N_D(\cdot)$ denotes the normal cone to D, and $N(\cdot)$ denotes null space. If in fact F is $\mathcal{C}^2$ then we call A *strongly amenable* at $\bar{x}$.

Theorem 5.4 (amenable implies nearly radial) *Suppose the set $A \subset \mathbf{E}$ is amenable at the point $\bar{x} \in A$. Then A is nearly convex (and hence nearly radial) at $\bar{x}$.*

Proof Since A is amenable at $\bar{x}$, we can suppose property (5.3) holds. Suppose without loss of generality $\bar{x} = 0$, and consider a sequences of points $x_r, y_r \to 0$ in the set $A \cap V$. We want to show

$$\mathrm{dist}(y_r, x_r + T_A(x_r)) = o(\|x_r - y_r\|).$$

Without loss of generality we can suppose $x_r \neq y_r$ for all r, and denote the unit vectors $\|x_r - y_r\|^{-1}(x_r - y_r)$ by z_r. We want to prove

$$d_r = \min\{\|w + z_r\| : w \in T_A(x_r)\} \to 0.$$

The unique minimizer $w_r \in T_A(x_r)$ in the above projection problem satisfies

$$\begin{aligned} d_r &= \|w_r + z_r\| \\ w_r + z_r &\in -N_A(x_r) = -\nabla F(x_r)^* N_D(F(x_r)) \\ \langle w_r, w_r + z_r \rangle &= 0, \end{aligned}$$

by [14, Ex 10.26]. Choose vectors $u_r \in -N_D(F(x_r))$ such that

$$w_r + z_r = \nabla F(x_r)^* u_r.$$

We next observe that the sequence of vectors $\{u_r\}$ is bounded. Otherwise, we could choose a subsequence $\{u_{r'}\}$ satisfying $\|u_{r'}\| \to \infty$, and then any limit point of the sequence of unit vectors $\{\|u_{r'}\|^{-1} u_{r'}\}$ must lie in the set $-N_D(F(0)) \cap N(\nabla F(0)^*)$, contradicting property (5.3).

We now have

$$\begin{aligned} 0 \ \leq \ d_r^2 \ &= \ \langle z_r, \nabla F(x_r)^* u_r \rangle \ = \ \langle \nabla F(x_r) z_r, u_r \rangle \\ &= \ \langle \nabla F(x_r) z_r - \|x_r - y_r\|^{-1}[F(x_r) - F(y_r)], u_r \rangle \\ &\qquad + \langle \|x_r - y_r\|^{-1}[F(x_r) - F(y_r)], u_r \rangle. \end{aligned}$$

The first term converges to zero, using the smoothness of the mapping F and the boundedness of sequence $\{u_r\}$. On the other hand, since the set D is convex, we have $F(y_r) - F(x_r) \in T_D(F(x_r))$, and $u_r \in -N_D(F(x_r))$ by assumption, so the second term is nonpositive, and the result follows. $\square$

It is worth comparing these notions to a property that is slightly stronger still: *prox-regularity* (in the terminology of [14]), or $O(2)$-*convexity* [15].

Definition 5.5 (prox-regular sets) A set $A \subset \mathbf{E}$ is *prox-regular* at a point $\bar{x} \in A$ if

$$\mathrm{dist}\Big(y, x + T_A(x)\Big) = O(\|x - y\|^2) \ \text{ as } x, y \to \bar{x} \text{ in } A.$$

Theorem 5.4 (amenable implies nearly radial) is analogous to the fact that strong amenability implies prox-regularity [14, Prop 13.32] (and also to [15, Prop 2.3]).

The class of nearly radial sets is very broad, as the following easy result (which fails for nearly convex sets) emphasizes.

Proposition 5.6 (unions) *If the sets $A_1, A_2, \ldots, A_n$ are each nearly radial at the point $\bar{x} \in \cap_j A_j$, then so is the union $\cup_j A_j$.*

Proof If the result fails, there is a sequence of points $x_r \to \bar{x}$ in $\cup_j A_j$ and real $\epsilon > 0$ such that

$$\mathrm{dist}\Big(\frac{\bar{x} - x_r}{\|\bar{x} - x_r\|}, T_{\cup_j A_j}(x_r)\Big) \geq \epsilon \ \text{ for all } r. \tag{5.7}$$

By taking a subsequence, we can suppose there is an index i such that $x_r \in A_i$ for all r. But then we know

$$\mathrm{dist}\Big(\frac{\bar{x} - x_r}{\|\bar{x} - x_r\|}, T_{A_i}(x_r)\Big) \to 0,$$

which contradicts inequality (5.7), since $T_{A_i}(x_r) \subset T_{\cup_j A_j}(x_r)$. □

A key concept in variational analysis is the idea of Clarke regularity (see for example [9, 10, 14]). We make no essential use of this concept in our development, but it is worth remarking on the relationship (or lack of it) between the nearly radial property and Clarke regularity. Note first that nearly radial sets need not be Clarke regular: the union of the two coordinate axes in $\mathbf{R}^2$ is nearly radial at the origin, for example, but it is not Clarke regular there.

On the other hand, Clarke regular sets need not be nearly radial. Consider, for example, the function $f : \mathbf{R} \to \mathbf{R}$ defined by

$$f(x) = \begin{cases} 2^{-n} - 2^{-n-1}(2 - 2^{n+1}|x|)^{1+2^{-n}} & \text{if } 2^{-n-1} \leq |x| \leq 2^{-n} \ (n \in \mathbf{N}) \\ 0 & \text{if } x = 0. \end{cases}$$

The graph of this even function consists of concave segments on each interval $x \in [2^{-n-1}, 2^{-n}]$, passing through the point $2^{-n}(1,1)$ with left derivative zero, and through the point $2^{-n-1}(1,1)$ with right derivative $1 + 2^{-n}$. A routine calculation now shows that this function is everywhere regular, and hence its epigraph $\mathrm{epi}\, f$ is everywhere Clarke regular. However, $\mathrm{epi}\, f$ is not nearly radial at the origin. To see this, observe that for each $n \in \mathbf{N}$, if we consider the sequence $x_n = 2^{-n}(1,1) \to (0,0)$, then we have

$$T_{\mathrm{epi}\, f}(x_n) = \Big\{(x,y) : y \geq (1 + 2^{1-n}) \max\{x, 0\}\Big\},$$

so

$$\mathrm{dist}(0, x_n + T_{\mathrm{epi}\, f}(x_n)) = \frac{\|x_n\|}{\sqrt{2}},$$

contradicting the definition of a nearly radial set.

One other class of sets is very important for our development, namely *semi-algebraic* sets (those sets defined by polynomial equations and inequalities).

Theorem 5.8 (semi-algebraic sets) *Semi-algebraic sets are nearly radial.*

Proof Suppose the origin lies in a semi-algebraic set $A \subset \mathbf{E}$. We will show that A is nearly radial at the origin.

If the result fails, then there is a real $\delta > 0$ and a sequence of points $y_r \to 0$ in A such that

$$\Big\| u + \frac{y_r}{\|y_r\|} \Big\| > \delta \quad \text{for all } u \in T_A(y_r).$$

Hence for each index r there exists a real $\gamma_r > 0$ such that

$$\Big\| \frac{z - y_r}{\|z - y_r\|} + \frac{y_r}{\|y_r\|} \Big\| > \delta \quad \text{for all } z \in A \text{ such that } 0 < \|z - y_r\| < \gamma_r.$$

Consequently, each point y_r lies in the set

$$A_0 = \Big\{ y \in A : \exists \gamma > 0 \text{ so } \Big\| \frac{z - y}{\|z - y\|} + \frac{y}{\|y\|} \Big\| > \delta \ \forall z \in A \backslash \{y\} \text{ with } \|z - y\| < \gamma \Big\},$$

so $0 \in \mathrm{cl}\, A_0$.

By quantifier elimination (see for example the discussion of the Tarski-Seidenberg Theorem in [3, p. 62]), the set A_0 is semi-algebraic. Hence the

Curve Selection Lemma (see [3, p. 98] and [13]) shows that there is a real-analytic path $p : [0,1] \to \mathbf{E}$ such that $p(0) = 0$ and $p(t) \in A_0$ for all $t \in (0,1]$. For some positive integer k and nonzero vector $g \in \mathbf{E}$ we have, for small $t > 0$,

$$\begin{aligned} p(t) &= gt^k + O(t^{k+1}) \\ p'(t) &= kgt^{k-1} + O(t^k), \end{aligned}$$

and in particular both $p(t)$ and $p'(t)$ are nonzero. For any such t we know

$$\Big\| \frac{z - p(t)}{\|z - p(t)\|} + \frac{p(t)}{\|p(t)\|} \Big\| > \delta$$

for any point $z \in A \setminus \{p(t)\}$ close to $p(t)$. Hence for any real $s \neq t$ close to t we have

$$\Big\| \frac{p(s) - p(t)}{\|p(s) - p(t)\|} + \frac{p(t)}{\|p(t)\|} \Big\| > \delta.$$

Taking the limit as $s \uparrow t$ shows

$$\left\| \frac{p(t)}{\|p(t)\|} - \frac{p'(t)}{\|p'(t)\|} \right\| \geq \delta$$

for all small $t > 0$. But since

$$\lim_{t \downarrow 0} \frac{p(t)}{\|p(t)\|} = \frac{g}{\|g\|} = \lim_{t \downarrow 0} \frac{p'(t)}{\|p'(t)\|},$$

this is a contradiction. $\square$

By contrast, semi-algebraic sets need not be nearly convex. For example, the union of the two coordinate axes in $\mathbf{R}^2$ is semi-algebraic, but it is not nearly convex at the origin.

Semi-algebraic sets have another useful tangential property, which we need later. Let us call a unit vector in $\mathbf{E}$ an *analytic direction* to an arbitrary set $A \subset \mathbf{E}$ at a point $x \in \mathbf{E}$ if it can be written in the form

$$\lim_{t \downarrow 0} \frac{p(t) - p(0)}{\|p(t) - p(0)\|}$$

for some analytic path $p : [0,1] \to A$ such that $p(0) = x$. Geometrically, an analytic direction is simply the direction at x of some analytic path from x

into A. We denote the set of analytic directions by $U_A(x)$: clearly in general we have

$$U_A(x) \subset \{u \in T_A(x) : \|u\| = 1\},$$

so since the tangent cone is always closed we deduce

$$\text{cl}\, U_A(x) \subset \{u \in T_A(x) : \|u\| = 1\}. \tag{5.9}$$

The following result shows that this inclusion is in fact an equality when the set A is semi-algebraic.

Proposition 5.10 (tangent versus analytic directions) *At any point $\bar{x}$ in a semi-algebraic set $A \subset \mathbf{E}$, we have*

$$\text{cl}\, U_A(\bar{x}) = \{u \in T_A(\bar{x}) : \|u\| = 1\}.$$

Proof Denote the right hand side by $\hat{T}$. Without loss of generality, suppose $\bar{x} = 0$. Consider any vector $u \in \hat{T}$, and fix any real $\delta > 0$. By definition, some sequence of points $x_r \to 0$ in A and reals $t_r \downarrow 0$ satisfy $t_r^{-1} x_r \to u$. Hence as $r \to \infty$ we have

$$\lim \frac{x_r}{\|x_r\|} = \lim \frac{x_r}{t_r} \cdot \frac{t_r}{\|x_r\|} = \lim \frac{x_r}{t_r} \cdot \lim \frac{t_r}{\|x_r\|} = u.$$

Consequently the origin lies in the closure of the semi-algebraic set

$$A_1 = \Big\{ x \in A \setminus \{0\} : \Big\| u - \frac{x}{\|x\|} \Big\| < \delta \Big\}.$$

Hence by the Curve Selection Lemma [13, 3], there is a real-analytic path $p : [0,1] \to \mathbf{E}$ such that $p(0) = 0$ and $p(t) \in A_1$ for all $t \in (0,1]$. We deduce $u \in U_A(0) + \delta B$. Since the vector $u \in \hat{T}$ was arbitrary, we have the inclusion $\hat{T} \subset U_A(0) + \delta B$, and since $\delta > 0$ was arbitrary, we deduce $\hat{T} \subset \text{cl}\, U_A(0)$. The reverse inclusion is just (5.9). □

6 Functions with sharp growth

In this section we investigate functions that grow rapidly as we move away from an associated nearly radial set. We first recall an idea from variational analysis. The *subderivative* of the function $f : \mathbf{E} \to [-\infty, +\infty]$ at a point $x \in \mathbf{E}$ (see [14]) in a direction $z \in \mathbf{E}$ is

$$df(x)(z) = \liminf_{t \downarrow 0,\ w \to z} \frac{f(x+tw) - f(x)}{t}.$$

Definition 6.1 (sharp growth) A function $f : \mathbf{E} \to [-\infty, +\infty]$ *grows sharply from* a set $A \subset \mathbf{E}$ at a point $x \in A$ if

$$T_A(x) \subset \mathrm{cl}\{z \in \mathbf{E} : d(-f)(x)(z) < 0\}.$$

The following is an illustrative example to keep in mind.

Proposition 6.2 (distance to manifolds) *If the set $M \subset \mathbf{E}$ is a smooth manifold of dimension strictly less than that of $\mathbf{E}$ around the point $x \in M$ then the distance function $\mathrm{dist}(\cdot, M)$ grows sharply from M at x.*

Proof Choose any unit vector $z \in N_M(\bar{x})$. By [14, Ex 6.8 and 8.53], for any vector $u \in T_M(\bar{x})$ and any real $\delta > 0$ we have

$$\begin{aligned} -d(-\mathrm{dist}(\cdot, M))(x, u + \delta z) &\geq d(\mathrm{dist}(\cdot, M))(x, u + \delta z) \\ &= \mathrm{dist}(u + \delta z, T_M(x)) = \delta > 0. \end{aligned}$$

The result now follows, since $u + \delta z \to u$ as $\delta \downarrow 0$. □

We make some comments about our definition of sharp growth.

1. Since zero is always a tangent vector, if f grows sharply from A at x, then x cannot be a local maximizer of f.

2. If x is not isolated in A, then $T_A(x) \neq \{0\}$. In this case, if A is semi-algebraic, then it suffices to check

$$U_A(x) \subset \mathrm{cl}\{z \in \mathbf{E} : d(-f)(x)(z) < 0\},$$

by Proposition 5.10 (tangent versus analytic directions).

If f grows sharply from A at every point in A, then we simply say that f *grows sharply from* A. For such functions, the result below shows that the attaining set $M_\epsilon(x)$ must be disjoint from A for all small $\epsilon > 0$.

Lemma 6.3 (attaining set) *Consider an open set $U \subset \mathbf{E}$. If the set $A \subset U$ is nearly radial at the point $x \in A$ and the continuous function $f : U \to \mathbf{R}$ grows sharply from A, then the attaining set $M_\epsilon(x)$ is disjoint from A for all small regularization parameters $\epsilon > 0$.*

Proof Suppose without loss of generality that $x = 0$. If the result fails, there is a sequence of parameters ϵ_n decreasing to zero, and a sequence of attaining points $y_n \in M_{\epsilon_n}(0) \cap A$. Since the origin is not a local maximizer, we can suppose each y_n is nonzero.

Choose reals $K > k > 0$ such that $kB \subset C \subset KB$. Since A is nearly radial, for n sufficiently large, there exists a tangent vector $u \in T_A(y_n)$ such that

$$\Big\| u + \frac{y_n}{\|y_n\|} \Big\| < \frac{k}{K}.$$

Since f grows sharply from A at y_n, there exists a direction $z \in \mathbf{E}$ such that $d(-f)(y_n)(z) < 0$ and

$$\Big\| z + \frac{y_n}{\|y_n\|} \Big\| < \frac{k}{K}.$$

By the definition of the subderivative, there exist sequences $t_r \downarrow 0$ in $\mathbf{R}$ and $w_r \to z$ in $\mathbf{E}$ such that

$$\frac{f(y_n + t_r w_r) - f(y_n)}{t_r} \to -d(-f)(y_n)(z) > 0.$$

Hence for sufficiently large r we have the inequalities

$$\begin{aligned} f(y_n + t_r w_r) &> f(y_n) \\ t_r &< \|y_n\| \\ \Big\| w_r + \frac{y_n}{\|y_n\|} \Big\| &< \frac{k}{K}. \end{aligned}$$

This last inequality implies the inclusion

$$y_n + \|y_n\| w_r \in \epsilon_n k B \subset \epsilon_n C,$$

so since the set C is convex we have

$$y_n + t_r w_r \in \epsilon_n C,$$

contradicting the definition of $M_{\epsilon_n}(0)$. □

Of course, if the point $\bar{x} \in \mathbf{E}$ lies outside the closure of any set A, then so must the attaining set $M_\epsilon(\bar{x})$ for all small ϵ.

We are now ready for our main result.

Theorem 6.4 (Lipschitz robust regularization) *Consider an open set $U \subset \mathbf{E}$ and a continuous function $f : U \to \mathbf{R}$ that grows sharply from the closed, nearly radial set $A \subset U$, and is locally Lipschitz on the complement $A^c \cap U$. Given any point $x \in U$, the robust regularization f_ϵ is locally Lipschitz around x for all small $\epsilon > 0$.*

Proof As we noted above, if $x \notin A$ then $M_\epsilon(x) \subset A^c$ for all small $\epsilon \geq 0$. The same holds for $x \in A$, by Lemma 6.3 (attaining set). The result now follows by Proposition 4.1 (Lipschitz behaviour). $\square$

The role of the specified point x in the above is worth emphasizing. Consider, for example, the function $f(x) = \sqrt{\max\{x, 0\}}$ on $\mathbf{R}$. If we choose $A = \{0\}$, the assumptions of the theorem are satisfied, so in particular the robust regularization f_ϵ is locally Lipschitz around zero for all small $\epsilon > 0$. However, for any *fixed* neighbourhood U of zero, it is easy to see that f_ϵ is nonlipschitz on U for all small $\epsilon > 0$. The neighbourhood of zero predicted by the theorem, on which f_ϵ is locally Lipschitz, must shrink as ϵ shrinks.

7 Example: the spectral abscissa

Let $\mathbf{M}^n$ denote the Euclidean space of n-by-n complex matrices, with inner product $\langle X, Y \rangle = \operatorname{Re} \operatorname{tr}(X^* Y)$ for matrices $X, Y \in \mathbf{M}^n$. The *spectral abscissa* $\alpha : \mathbf{M}^n \to \mathbf{R}$ is defined by

$$\alpha(X) = \max\{\operatorname{Re} \lambda : \lambda \text{ an eigenvalue of } X\}.$$

An eigenvalue is *nonderogatory* if it has geometric multiplicity (the dimension of the corresponding eigenspace) equal to one. Thus a nonderogatory eigenvalue corresponds to a single block in the Jordan form. We call an eigenvalue of a matrix *active* if its real part equals the spectral abscissa, and we call a matrix *active-nonderogatory* if all its active eigenvalues are nonderogatory. If a matrix is active-nonderogatory, then it is easy to check that so are all nearby matrices.

We call a matrix *active-simple* if every active eigenvalue is simple (that is, has algebraic multiplicity one). We denote the set of matrices that are not active-simple by A^n: it is routine to verify that this is a closed, semi-algebraic set. Since the dependence of any simple eigenvalue on the matrix is locally analytic, the spectral abscissa is locally Lipschitz on $(A^n)^c$.

On the other hand, the spectral abscissa is not locally Lipschitz around any active-nonderogatory matrix $X \in A^n$ [8]. We can be more specific with our current perspective.

Theorem 7.1 (sharp growth of spectral abscissa) *The spectral abscissa grows sharply from the set A^n of non-active-simple matrices at any active-nonderogatory matrix in A^n.*

Proof Consider an active-nonderogatory matrix $X \in A^n$. Notice X is not isolated in A^n, since since $X + \mu I \in A^n$ for all real μ. By note 2 after Definition 6.1, it suffices to check that any analytic direction $U \in U_{A^n}(X)$ lies in

$$\mathrm{cl}\,\{Z \in \mathbf{M}^n : d(-\alpha)(X)(Z) < 0\}.$$

By definition, there is a real-analytic path $P : [0,1] \to A^n$ such that $P(0) = X$ and

$$\lim_{t \downarrow 0} \frac{P(t) - P(0)}{\|P(t) - P(0)\|} = U.$$

The matrix $P(t)$ is active-nonderogatory for all small $t \geq 0$. If we consider the active eigenvalues for the matrix $P(t)$, listed say by decreasing imaginary parts, there are only finitely many possibilities for the corresponding list of algebraic multiplicities. For all $t > 0$ sufficiently small we can therefore assume that $P(t)$ has a constant number of active eigenvalues, with constant algebraic multiplicities, each one nonderogatary. The set $\mathcal{M}$ of matrices with this property is a manifold [1, 5].

We now appeal to the results of [5], which show that at each matrix $M \in \mathcal{M}$ there exists a horizon subgradient Z of norm one for the spectral abscissa, and Z is orthogonal to $\mathcal{M}$ at M:

$$Z \in \partial^\infty \alpha(M) \subset N_{\mathcal{M}}(M), \quad \|Z\| = 1.$$

Applying this observation at a sequence of matrices $P(t_r)$ where $t_r \downarrow 0$ shows the existence of matrices Z_r satisfying

$$Z_r \in \partial^\infty \alpha(P(t_r)) \subset N_{\mathcal{M}}(P(t_r)), \quad \|Z_r\| = 1.$$

Taking a subsequence, we can suppose Z_r converges to some nonzero matrix Z, which satisfies

$$Z \in \partial^\infty \alpha(X) = \partial \alpha(X)^\infty,$$

since the spectral abscissa function α is subdifferentially regular at X (see [14, Cor 8.11] and [8]). Hence for some matrix $Z_0 \in \mathbf{M}^n$ we have

$$Z_0 + \beta Z \in \partial \alpha(X), \quad \text{for all } \beta \in \mathbf{R}_+.$$

Since the path $p(t)$ lies in the manifold $\mathcal{M}$ for small $t > 0$, we have

$$\langle Z_r, \|p'(t_r)\|^{-1} p'(t_r) \rangle = 0$$

for all large r. Taking the limit as $r \to \infty$ and using l'Hôpital's rule shows $\langle Z, U \rangle = 0$.

Now for any real $\delta > 0$ we have, using [14, Thm 8.30],

$$\begin{aligned} -d(-\alpha)(X)(U + \delta Z) &\geq d\alpha(X)(U + \delta Z) \\ &= \sup\{\langle U + \delta Z, W \rangle : W \in \partial \alpha(X)\} \\ &\geq \sup\{\langle U + \delta Z, Z_0 + \beta Z \rangle : \beta \geq 0\} \\ &= +\infty \end{aligned}$$

and the result now follows. □

The ϵ-*pseudospectral abscissa* of a matrix $X \in \mathbf{M}^n$, denoted $\alpha_\epsilon(X)$ is the largest possible real part of an eigenvalue of a matrix $Y \in \mathbf{M}^n$ lying within a distance ϵ of X (measured in the usual operator norm):

$$\alpha_\epsilon(X) = \max\{\operatorname{Re} \lambda : \lambda \text{ an eigenvalue of } Y,\ \|Y - X\| \leq \epsilon\}.$$

Since

$$\alpha_\epsilon(X) = \max\{\alpha(Y) : \|Y - X\| \leq \epsilon\},$$

the pseudospectral abscissa is exactly the robust regularization of the spectral abscissa, justifying the notation.

The pseudospectral abscissa has important modelling applications (see for example [17]). For example, $\alpha_\epsilon(X)$ often reveals more about the realistic behaviour of the dynamical system $\dot{w} = Xw$ than the spectral abscissa $\alpha(X)$ alone. Furthermore, by using the equivalent definition

$$\alpha_\epsilon(X) = \max\{\operatorname{Re} z : \sigma_{\min}(X - zI) \leq \epsilon\}$$

(where $\sigma_{\min}$ denotes the smallest singular value), the pseudospectral abscissa can be computed relatively easily, along with sensitivity information [6, 7]. This makes it a viable function to use in optimization applications.

Applying the observations of this section, we obtain the following nice application of Theorem 6.4 (Lipschitz robust regularization). An analogous result appears in [6], proved by a very different approach.

Corollary 7.2 (Lipschitz psesudospectral abscissa) *If the matrix $X \in \mathbf{M}^n$ is active-nonderogatory, then the pseudospectral abscissa α_ϵ is locally Lipschitz around X for all small $\epsilon > 0$.*

Spectral abscissa minimization problems typically have active-nonderogatory solutions with multiple eigenvalues [5]. Around such a solution, the spectral abscissa is not Lipschitz. On the other hand, if the solution is a strict local minimizer, for small regularization parameters $\epsilon > 0$ the corresponding pseudospectral abscissa minimization problem has a nearby local minimizer, by Proposition 3.7 (local minimizers). Furthermore, as we have seen above, around any fixed matrix the pseudospectral abscissa is locally Lipschitz for small $\epsilon > 0$. This example demonstrates how robust regularization in optimization, as well as serving a compelling intrinsic purpose, can reduce a non-Lipschitz problem to a Lipschitz setting, which may be more numerically tractable.

Acknowledgements Thanks to Jim Renegar for helpful discussions concerning Theorem 5.8 (semi-algebraic sets), and to Jon Borwein for the statement of Definition 5.1 (nearly radial sets).

References

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

[2] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM, Philadelphia, 2001.

[3] R. Benedetti and J.-J. Risler. *Real algebraic and semi-algebraic sets.* Hermann, Paris, 1990.

[4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, Philadelphia, 1994.

[5] J.V. Burke, A.S. Lewis, and M.L. Overton. Optimal stability and eigenvalue multiplicity. *Foundations of Computational Mathematics*, 1:205–225, 2001.

[6] J.V. Burke, A.S. Lewis, and M.L. Overton. Optimization over pseudospectra. Technical report, Simon Fraser University, 2002. Submitted to SIAM Journal on Matrix Analysis.

[7] J.V. Burke, A.S. Lewis, and M.L. Overton. Robust stability and a criss-cross algorithm for pseudospectra. Technical report, Simon Fraser University, 2002. Submitted to IMA Journal of Numerical Analysis.

[8] J.V. Burke and M.L. Overton. Variational analysis of non-Lipschitz spectral functions. *Mathematical Programming*, 90:317–352, 2001.

[9] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983. Republished as Vol. 5, Classics in Applied Mathematics, SIAM, 1990.

[10] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory.* Springer-Verlag, New York, 1998.

[11] L. El Ghaoui and S.-I. Niculescu. Robust decision problems in engineering: a linear matrix inequality approach. In L. El Ghaoui and S.-I. Niculescu, editors, *Advances in Linear Matrix Inequality Methods in Control*, pages 3–37. SIAM, Philadelphia, 2000.

[12] J.-B Hiriart-Urruty. The deconvolution operation in convex analysis: an introduction. *Cybernetics Systems Analysis*, 30:555–560, 1994.

[13] J.W. Milnor. *Singular Points of Complex Hypersurfaces.* Princeton University Press, Princeton, N.J., 1968.

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

[15] A. Shapiro. Existence and differentiability of metric projections in Hilbert space. *SIAM Journal on Optimization*, 4:130–141, 1994.

[16] A. Shapiro and F. Al-Khayyal. First-order conditions for isolated locally optimal solutions. *Journal of Optimization Theory and Applications*, 77:189–196, 1993.

[17] L.N. Trefethen. Pseudospectra of linear operators. *SIAM Review*, 39:383–406, 1997.