Perturbation Theory of Polyhedral Seminorms Under Linear Constraints

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Alexander J. Gutierrez

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

Doctor of Philosophy

Prof. Gilad Lerman

August, 2020

Acknowledgements

Thanks to the the NSF Graduate Research Fellowship and NSF Award DMS-18-30418 which financially supported this thesis.

I am profoundly grateful to have had the chance to attend graduate school at the University of Minnesota. The environment is truly exceptional both academically and—uniquely to my knowledge—from a community perspective. It would be an undertaking of a similar scope to the one I have just survived to fully catalogue my debt to everyone whose support has allowed for this achievement—but the following paragraphs attempt to make some of it clear.

In particular I'd like to thank my mentors during my undergraduate career: Dr. Eric Kostelich who first gave me a chance to "do" math and Drs. Rodrigo Platte and Anne Gelb with whom I was also lucky enough to collaborate with. I was priviliged to work with another fantastic group of people in graduate school starting with my adviser Gilad Lerman—who always managed to keep things in perspective—and my collaborator on the work that became this thesis, Sam Stewart.

The work spun out of earlier research—I have been blessed in my career here with an embarassment of great mentors and role models and the person that has filled those roles for the entirety of my graduate school experience has been Jarvis Haupt. I also owe a debt of gratitude to the rest of my collaborators on that work:

Di Xiao, Mike Mullen, Albert Jang, Taylor Froelich, and Mike Garwood. I am also sincerely grateful to my other committee members: Richard McGehee and William Leeb for their sharp questions, comments, and flexibility throughout the hectic end of my time here in Minnesota.

Beyond the academic component—- I would not have finished without the strong community in both the math and electrical engineering departments—thank you to all my friends in both. I was also lucky to have constant support from my parents and sisters throughout the experience and from Olivia and Sally during some of the hardest parts.

Abstract

A recent paper by Klatte and Kummer provides a new characterization for the regularity of the $\arg\min$ set of an optimization problem with respect to perturbations. In this thesis we prove that this characterization applies to a broad class of widely used minimization problems. In particular, this work applies to both ℓ_1 and (anisotropic) discrete total variation minimization problems under linear constraints.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The utility of optimization tools in scientific fields demands three prerequisites: (a) the ability to model the underlying problem in an optimization framework, (b) the ability to find the optimal value, and (c) the knowledge that this minimizer is *robust* with respect to inevitable errors in the data or the model. Although the utility of (c) is predicated on the success of parts (a) and (b) it is just as integral to the scientific project. Recent years have seen a rebirth of interest in examining this form of robustness for a wide variety of optimization problems, and a robust set of tools has been developed to study it. In this thesis we will examine a particular form of robustness, known as *calmness* that a certain minimization problem possesses. This minimization problem arises naturally in an applied problem in the field of Magnetic Resonance Imaging [1]. In this thesis we will motivate and prove that this minimization problem is "calm" (see Section 2.3 for definitions).

To do this we will first review related literature in Chapter 2, then introduce and use Hoffman's lemma to prove that the constraint set of the minimization problem is calm in Chapter 3. Finally, we will use this result to state and prove our main theorem in Chapter 4 before ending with discussion and future directions in Chapter 5.

1.1 Notation

Let $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex fields, respectively. We denote vectors using lower-case boldface letters ($\mathbf{u}, \mathbf{x}, \boldsymbol{\theta}$), matrices using upper-case boldface letters ($\mathbf{A}, \mathbf{B}$), and unbolded characters (τ, t, Ω) refer to scalars and scalar-valued functions. Calligraphic notation ($\mathscr{U}, \mathscr{V}$) denotes vector (or affine) subspaces and $\mathbf{P}_{\mathscr{U}}$ stands for the (orthogonal) projection matrix onto the subspace $\mathscr{U}$ (similarly for a closed convex $C \subset \mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$ write $\mathbf{P}_C(\mathbf{x})$ to denote the projection of $\mathbf{x}$ onto C). We use an asterisk * to denote conjugate transpose, so that the $\mathbf{u}^*\mathbf{v}$ is the standard inner product on $\mathbb{R}^n$. Let $\|\cdot\| = \|\cdot\|_2$ represent the standard Euclidean norm on vectors and the induced norm on matrices and let $\|\mathbf{A}\|_F$ denote the Frobenius norm of a matrix, and $N(\mathbf{A})$ the null space of $\mathbf{A}$.

Write the ball of radius δ centered at a vector $\mathbf{x}$ (or matrix $\mathbf{A}$) as $B_\delta(\mathbf{x})$ (respectively, $B_\delta(\mathbf{A})$). Given a natural number n we use $[n]$ to refer to the set $\{1, 2, \cdots, n\}$. Given a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a subset $J \subseteq [n]$ we write $\mathbf{A}_J$ to denote the submatrix of $\mathbf{A}$ with columns indexed by J. All convex functions, including the ones mentioned in the literature, will be proper.

Chapter 2

Related Work in Parametric Optimization

In this chapter we discuss related work in the context of this thesis. To fix notation, consider the problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \Omega(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{Ax} = \mathbf{b} \\ & \|\mathbf{x}\| \leq R \end{aligned} \tag{P($\mathbf{A}$)}$$

for various relevant classes of functions Ω and under perturbations to $\mathbf{A}$. The problem falls under the more general class of parametric optimization, which deals with minimization problems of the form:

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ f(\mathbf{x}, t) \quad \text{s.t.} \quad \mathbf{x} \in M(t) := \left\{\mathbf{x} \in \mathbb{R}^n :\ g_j(\mathbf{x}, t) \leq 0, j = 1, \cdots, m\right\}. \tag{P(t)}$$

For a fixed t_0 and t' close to t_0, the field of parametric optimization is concerned with: (a) the relationship between $f(\mathbf{x}, t)$ and $F(\mathbf{x}, t')$ and (b) the existence of a continuous path of minimizers $\mathbf{x}(t)$ to (P(t)). See [2] for a general overview of the topic and [3] for an overview with an emphasis on the non-smooth case.

In the remainder of this section we survey relevant results from this field in the context of our own contributions. The section is organized in, roughly, increasing order of generality.

2.1 The ℓ^2 case

We begin our survey with the well-studied and much simpler case $\Omega(\mathbf{x}) = \|\mathbf{x}\|_2$. In this case, the optimization problem (P($\mathbf{A}$)) in fact has a closed-form solution that is linear in $\mathbf{b}$, where the linear operator is $\mathbf{A}^\dagger$, the *Moore-Penrose* pseudoinverse of the matrix $\mathbf{A}$. Parametric optimization is interested in a perturbation theory for $\mathbf{A}^\dagger$, guaranteeing that whenever $\mathbf{A}$ and $\widetilde{\mathbf{A}}$ are close, the corresponding solutions $\mathbf{A}^\dagger\mathbf{b}$ and $\widetilde{\mathbf{A}}^\dagger\mathbf{b}$ will be close. The problem was first studied by Stewart in [4] and Wedin in [5] who proved the foundational results in the area. In particular, [4] showed that a pseudoinverse is not even a continuous function of the entries of the matrix when the rank of the involved matrices is non-constant. In particular, if $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m < n$ and $\mathbf{rank}\,\mathbf{A} < m$ then there exists a sequence of matrices $\mathbf{E}_n$ with $\|\mathbf{E}\| \to 0$ such that $\left\|\mathbf{A}^\dagger - (\mathbf{A}+\mathbf{E}_n)^\dagger\right\| \to \infty$ [4]. Nevertheless, it is possible to show that, under the constraint $\mathbf{rank}\,\mathbf{A} = \mathbf{rank}(\mathbf{A}+\mathbf{E}_n)$, the pseudoinverse is continuous: $(\mathbf{A}+\mathbf{E}_n)^\dagger \to \mathbf{A}^\dagger$. Under this condition, [5] proves the following result:

Theorem 1 (Wedin's Theorem (1973) [5])**.** Let $|||\cdot|||$ be a unitarily invariant matrix norm and let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ (in this theorem the matrices can be either "fat" or "tall". If

	Arbitrary	Spectral	Frobenius
$\mathbf{rank}\,\mathbf{A} < \min(m,n)$	3	$\frac{1+\sqrt{5}}{2}$	$\sqrt{2}$
$\mathbf{rank}\,\mathbf{A} = \min(m,m)$	2	$\sqrt{2}$	1
$\mathbf{rank}\,\mathbf{A} = m = n$	1	1	1

Table 2.1: Values of μ in Wedin's Theorem 1

$\mathbf{rank}\,\mathbf{A} = \mathbf{rank}\,\mathbf{B}$ then,

$$\left|\left|\left|\mathbf{B}^\dagger - \mathbf{A}^\dagger\right|\right|\right| \le \mu \left\|\mathbf{A}^\dagger\right\|_2 \left\|\mathbf{B}^\dagger\right\|_2 |||\mathbf{B}-\mathbf{A}|||,$$

where μ is given in Table 2.1.

This result is useful both as a baseline for what we might be able to hope for in the harder problem, (P(**A**)). For example, the fact that $\mathbf{A} \to \mathbf{A}^\dagger$ is discontinuous at points where $\mathbf{rank}(\mathbf{A}) < m$ justifies our constraint that $\mathbf{A}$ must have full rank in order to get a result of this type (cf to the main result of this work, Theorem 4).

This line of research has a long line of history extending it to more general problems with a more refined analysis. To the author's knowledge the current state of the art in this regard appears in [6], where the author proves two very interesting theorems in a slightly more general setting. In it, and in the setting of this paper, he proves the following result:

Theorem 2 (Theorem 3.1 in [6])**.** Let $\mathbf{A}_0$, $\mathbf{A}_1 = \mathbf{A}+\mathbf{E} \in \mathbb{R}^{m\times n}$ with $m < n$ and and consider a perturbation to the data $\mathbf{b}_1 = \mathbf{b}_0 + \mathbf{e}$. Further suppose that $\|\mathbf{E}\| \cdot \left\|\mathbf{A}^\dagger\right\| < 1$ and that $\mathbf{rank}\,\mathbf{A}_0 = \mathbf{rank}\,\mathbf{A}_1$. Denote by $\mathbf{x}_i = \mathbf{x}^*(\mathbf{A}_i, \mathbf{b}_i)$ for $i \in \{0,1\}$ the optimal solution to problem (P(**A**)) with matrix $\mathbf{A}_i$ and data $\mathbf{b}_i$. Define the relative perturbation sizes as $\epsilon_{\mathbf{A}} = \|\mathbf{E}\| / \|\mathbf{A}_0\|$ and $\epsilon_{\mathbf{b}} = \|\mathbf{e}\| / \|\mathbf{b}_0\|$ and let $\kappa = \kappa(\mathbf{A}_0) = \|\mathbf{A}_0\| \left\|\mathbf{A}_0^\dagger\right\|$ be the spectral condition number of $\mathbf{A}_0$. Then

$$\frac{\|\mathbf{x}_1-\mathbf{x}_0\|}{\|\mathbf{x}_0\|} \leq \kappa\left[\left(1+\frac{1}{1-\kappa\epsilon_{\mathbf{A}}}\right)\epsilon_{\mathbf{A}}+\frac{1}{1-\kappa\epsilon_{\mathbf{A}}}\frac{\|\mathbf{b}\|}{\|\mathbf{A}_0\|\,\|\mathbf{x}_0\|}\epsilon_{\mathbf{b}}\right]. \tag{2.1.1}$$

This theorem is particularly interesting in the setting of this paper because it shows the interaction between perturbations of the data and perturbations of the forward operator. In fact, the theorem as stated in [6] includes further interaction terms for when $\mathbf{A}\mathbf{x}=\mathbf{b}$ is not feasible (but under rank constraints) and for perturbations of the origin (i.e., minimization of $\left\|\mathbf{x}-\mathbf{p}\right\|$ for $\mathbf{p}$ also possibly perturbed).

Additionally, the same paper shows that a weaker form of regularity is still possible even when the perturbation is not rank preserving. Define the function

$$F(\mathbf{A},\mathbf{b})=\left\{\mathbf{x}\in\mathbb{R}^n : \mathbf{A}^T\mathbf{A}\mathbf{x}=\mathbf{A}^T\mathbf{b}\right\}, \tag{2.1.2}$$

to be the constraint set of problem (P(**A**)) defined as a function of its parameters (here we write it in normal equation format to cover the case when the linear equations are not satisfiable). Let $\mathbf{B}$ be a matrix defined by any subset of the columns of $\mathbf{A}$ that has full rank, and let $\mathbf{C}$ be the remaining columns so that, without loss of generality, $\mathbf{A}=[\mathbf{B},\mathbf{C}]$. Denote $\widetilde{\kappa}=\|\mathbf{A}\|\cdot\left\|\mathbf{B}^\dagger\right\|$. Using this notation:

Theorem 3 (Theorem 4.1 in [6])**.** Let the parameters $\mathbf{A}_0, \mathbf{A}_1, \epsilon_{\mathbf{A}}, \mathbf{b}_0, \mathbf{b}_1$, and $\epsilon_{\mathbf{b}}$ be as in Theorem 2. Further, assume $\widetilde{\kappa}\epsilon_{\mathbf{A}}<1$. Then, given the solution $\mathbf{x}_1=\mathbf{x}^*(\mathbf{A}_1,\mathbf{b}_1)$ to the perturbed problem there is an $\mathbf{x}\in F(\mathbf{A}_0,\mathbf{b}_0)$ such that

$$\frac{\|\mathbf{x}_1-\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\widetilde{\kappa}}{1-\widetilde{\kappa}\epsilon_{\mathbf{A}}}\left[\left(1+\frac{\widetilde{\kappa}}{1-\widetilde{\kappa}\epsilon_{\mathbf{A}}}\frac{\|\mathbf{b}-\mathbf{A}\mathbf{x}\|}{\|\mathbf{A}\|\,\|\mathbf{x}\|}\right)\epsilon_{\mathbf{A}}+\frac{\|\mathbf{b}\|}{\|\mathbf{A}\|\,\|\mathbf{x}\|}\epsilon_{\mathbf{b}}\right]. \tag{2.1.3}$$

Note that in the case that $\mathbf{Ax} = \mathbf{b}$ is consistent this reduces to the bound:

$$\frac{\|\mathbf{x}_1 - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\widetilde{\kappa}}{1 - \widetilde{\kappa}\epsilon_{\mathbf{A}}} \left(\epsilon_{\mathbf{A}} + \epsilon_{\mathbf{b}}\right). \tag{2.1.4}$$

This theorem is a notable improvement over Theorem 1 in that it shows the interaction between various types of perturbations, both to the data and to the underlying operator. It is further interesting to note that it manages to get rid of the constant μ from that theorem (which is tight in some cases), at the expense of a denominator.

2.2 Implicit function theorem

Stability of the form given by Theorems 1 and 2 are true in general whenever the conditions of the implicit function theorem are satisfied (for example, when the problem is strongly convex under proper constraint qualifications such as Slater's condition). Indeed, consider the problem:

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) \\ &\text{subject to} \quad \mathbf{A}(t)\mathbf{x} = \mathbf{b} \quad \text{for } t \in [0,1]. \end{aligned} \tag{2.2.1}$$

where f is strongly convex, $\mathbf{b}$ lies in the range of $\mathbf{A}(t)$, and $\mathbf{A}(t)$ is differentiable. Define the Lagrangian associated with (2.2.1) by

$$\mathscr{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda}^*(\mathbf{Ax} - \mathbf{b}).$$

Then, for each t, the KKT conditions give necessary and sufficient conditions for optimality. The stationarity condition states that a point $(\mathbf{x}^*, \boldsymbol{\lambda}^*) = (\mathbf{x}^*(t), \boldsymbol{\lambda}^*(t))$ is optimal if

$$0 = \begin{pmatrix} \frac{\partial \mathscr{L}}{\partial \mathbf{x}} \\ \\ \frac{\partial \mathscr{L}}{\partial \boldsymbol{\lambda}} \end{pmatrix} (\mathbf{x}^*, \boldsymbol{\lambda}^*). \tag{2.2.2}$$

Differentiating this constraint with respect to the parameter t yields a linear system in terms of the Hessian $\nabla^2 \mathscr{L}$:

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \frac{\partial \mathscr{L}}{\partial \mathbf{x}} \\ \\ \frac{\partial \mathscr{L}}{\partial \boldsymbol{\lambda}} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 \mathscr{L}}{\partial \mathbf{x}^2} & \frac{\partial^2 \mathscr{L}}{\partial \boldsymbol{\lambda} \partial \mathbf{x}} \\ \\ \frac{\partial^2 \mathscr{L}}{\partial \mathbf{x} \partial \boldsymbol{\lambda}} & \frac{\partial^2 \mathscr{L}}{\partial \boldsymbol{\lambda}^2} \end{pmatrix} \begin{pmatrix} \frac{\mathrm{d}\mathbf{x}^*}{\mathrm{d}t} \\ \\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}t} \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \mathscr{L}}{\partial t \partial \mathbf{x}} \\ \\ \frac{\partial^2 \mathscr{L}}{\partial t \partial \boldsymbol{\lambda}} \end{pmatrix}. \tag{2.2.3}$$

When the conditions of the implicit function theorem are satisfied (e.g., when f is strongly convex) the Hessian matrix will be invertible which shows that the solution $\mathbf{x}^*$ is differentiable with respect to changes in t. See [7] for the foundational results in this area.

2.3 More general situations

Some recent work has analyzed related problems with respect to more general objective functions. For example, [8] examines the perturbation theory of generalized inverses, where they seek the "generalized inverse" that minimizes the p-norms $\|\cdot\|_p$ (in contrast to the Moore-Penrose pseudoinverse, which minimizes $\|\cdot\|_2$) and proves some concentration results on the Frobenius norms of those inverses. Note that this is a very weak form of regularity in the context of Problem (P($\mathbf{A}$))—two matrices may have similar Frobenius norms while acting very differently as linear operators.

The work most closely related to this thesis, and the literature on which it most relies can be found in work by Cánovas et al. [9] and extended by Klatte and Kummer [10] on the calmness of the argmin mapping in relatively general settings. The main theorem of this work shows that certain equivalent conditions to calmness first proved in [10] apply to a specific setting of particular interest. Both of these works are concerned with proving that certain multi-valued mappings are "calm." Let us proceed by defining the notions that will be needed.

In what follows, let $F : \mathscr{T} \rightrightarrows \mathscr{Y}$ be a set-valued mapping (also known as a multifunction) between normed vector spaces. Given a vector $\mathbf{x} \in \mathbb{R}^n$ and a set $K \subseteq \mathbb{R}^n$ let $d(\mathbf{x}, K)$ denote the infimal distance from $\mathbf{x}$ to K:

$$d(\mathbf{x}, K) = \inf_{\mathbf{y} \in K} \|\mathbf{x} - \mathbf{y}\|.$$

By an abuse of notation we will also denote the Hausdorff distance between two sets $J, K \subseteq \mathbb{R}^n$ with the same notation

$$d(J, K) = \max\left(\sup_{\mathbf{x} \in J} d(\mathbf{x}, K), \sup_{\mathbf{y} \in K} d(\mathbf{y}, J)\right).$$

Definition 1 (Lipschitzian)**.** We say that a multifunction is *Lipschitzian with constant* C if there exists a constant C such that for any $\mathbf{t}, \mathbf{t}' \in \mathscr{T}$

$$d(F(\mathbf{t}), F(\mathbf{t}')) \leq C \|\mathbf{t} - \mathbf{t}'\|.$$

Definition 2 (Calmness)**.** We call a set valued mapping $F(\mathbf{t})$ *calm* at $(\mathbf{t}_0, \mathbf{x}_0) \in \mathbf{Graph}\, F$ if there exists $\epsilon, \delta, C > 0$ such that for any $\mathbf{t} \in B_\delta(\mathbf{t}_0)$ and $x \in B_\epsilon(x_0) \cap F(\mathbf{t})$ we have

$$d(x, F(\mathbf{t}_0)) \leq C \|\mathbf{t} - \mathbf{t}_0\|$$

This definition is sometimes called "half-Lipschitz" since it is not global but rather fixes one of the end points.

Definition 3 (Lipschitz lower semicontinuous)**.** A set-valued map $F(\mathbf{t})$ is *Lipschitz lower semicontinuous* at $(\mathbf{t}_0, \mathbf{x}_0) \in \mathbf{Graph}\, F$ if there exists $C, \delta > 0$ such that for all $\mathbf{t} \in B_\delta(\mathbf{t}_0)$

$$d(\mathbf{x}_0, F(\mathbf{t})) \leq C \|\mathbf{t} - \mathbf{t}_0\|$$

Definition 4 (Aubin / Pseudo Lipschitz)**.** A set-valued map $F(\mathbf{t})$ is said to have the Aubin / pseudo-Lipschitz property at $(\mathbf{t}_0, \mathbf{x}_0) \in \mathbf{Graph}\, F$ if there exists $\epsilon, C, \delta > 0$ such that for all $\mathbf{t}, \mathbf{t}' \in B_\delta(\mathbf{t}_0)$ and $\mathbf{x} \in B_\epsilon(\mathbf{x}_0) \cap F(\mathbf{t})$,

$$d(\mathbf{x}, F(\mathbf{t}')) \leq C \left\|\mathbf{t} - \mathbf{t}'\right\|.$$

It is clear from the definitions that the Aubin property implies both lower Lipschitz semicontinuity as well as calmness.

Definition 5 (Locally upper Lipschitz semicontinuous)**.** We call a set valued mapping $F(\mathbf{t})$ *locally upper Lipschitz semicontinuous* if there exists $C > 0$ such that for all $\mathbf{t}_0$ there exists a $\delta > 0$ such that for all $\mathbf{t} \in B_\delta(\mathbf{t}_0)$ we have for all $\mathbf{x} \in F(\mathbf{t})$

$$d(\mathbf{x}, F(\mathbf{t}_0)) \leq C \|\mathbf{t} - \mathbf{t}_0\|$$

With this background we can now state related literature on the problem. The notion of calmness was first formulated by Rockafellar and was later explored by Clarke [11] and in the thesis of Thibault [12]. Its fundamental importance as a notion of regularity for optimization problems set off exploration of a variety of related notions some of which, unfortunately, are also known as "calm" (see, e.g., [13]). Of particular interest in this literature is Burke's work [14] showing the relationship between calmness and the viability of exact penalization and work by Henrion and Outrata [15] showing the calmness of an additive perturbation to a closed constraint set.

It turns out that a multifunction F having the Aubin property is equivalent to the *metric regularity* of the inverse function F^{-1} (see, for example, Section 3.8 of the monograph by Dontchev and Rockafellar, [16]) while calmness corresponds to *metric subregularity*. These notions seem to have developed distinctly from the notion of calmness, see the survey by Ioffe [17] for an overview and work by Ioffe, Jourani, and Thibault for some of the foundational results [18, 19, 20].

Besides the two works [9, 10] mentioned at the beginning of this section, this work is most closely related to a long line of work largely spearheaded by Robinson dealing with the properties of polyhedral mappings. Call a function polyhedral if its graph is the union of finitely many polyhedra. In 1969 Walkup and Wets provided a characterization of polyhedra as the only closed convex subsets for which intersections with an affine transformation are Lipschitzian [21]. The work [22] crucially relied on this result to show a necessary and sufficient condition for a polyhedral mapping to be Lipschitz lower semicontinuous. In 1987 Mangasarian and Shiau examined a close relative to Problem (P(**A**)) in which they instead examined perturbations to the data **b** [23]. In 2000, Jourani proved that a polyhedral mapping into a Hilbert space is locally upper Lipschitz semicontinuous [24], while in 2007 Robinson identified the precise class of positive semidefinite transformations for which the solutions to a certain type of affine variational inequalities defined over polyhedral multifunctions will be Lipschitzian. Finally, in 2008 Lu and Robinson [25] extended the earlier work by Mangasarian and Shiau et al. to a significantly more general setting.

Chapter 3

Subspace Perturbation Theory

In this chapter we will discuss the theory of (invariant) subspace perturbations as it applies to ($P(\mathbf{A})$). To recall notation, consider a matrix $\mathbf{A}_0$ and perturbation $\mathbf{A}_1 = \mathbf{A}_0 + \mathbf{E}$. The field seeks to understand how the singular (resp. eigen) values and vectors of $\mathbf{A}_0$ are related to their counterparts of $\mathbf{A}_1$.

The problem has a rich history, perhaps beginning with Jordan's definition of principal angles between subspaces in [26] and continuing with Weyl's foundational result on the perturbation of eigenvalues [27] and the foundational Davis-Kahan theorem for the pertubation of eigenspaces [28]. In this chapter we will first provide the relevant definitions and background to understand our main result and then show that the affine space defined by $\mathbf{A}_0\mathbf{x} = \mathbf{b}$ is well-behaved with respect to perturbations to $\mathbf{A}_0$.

3.1 Principal angles

We begin by defining the so-called "principal" or "canonical" angles between affine spaces and various properties about them that we will need later.

Definition 6 (Angle between vectors)**.** Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then we recall that the principal angle between two vectors to be $\theta \in [0, \pi/2]$ such that

$$\cos\theta = \left|\mathbf{u}^*\mathbf{v}\right|.$$

Note that this is well-defined by Cauchy-Schwarz and zero if and only if **u** and **v** are co-linear. This definition extends to subspaces by recursively picking the largest possible inner product:

Definition 7 (Principal angles between subspaces)**.** Let $\mathscr{U}, \mathscr{V}$ be subspaces in $\mathbb{R}^n$, with dimensions $\dim \mathscr{U} = m$, $\dim \mathscr{V} = n$, and let $r = \min(m, n)$. Then we define the **principal angles between $\mathscr{U}$ and $\mathscr{V}$** to be the vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_r) \in [0, \pi/2]^r$ given by

$$\begin{aligned} \cos\theta_k = \max_{\mathbf{u}\in\mathscr{U}, \mathbf{v}\in\mathscr{V}} \quad & \left|\mathbf{u}^*\mathbf{v}\right| \\ \text{subject to} \quad & \|\mathbf{u}\| = \|\mathbf{v}\| = 1, \\ & \mathbf{u} \perp \mathbf{u}_i, \quad \mathbf{v} \perp \mathbf{v}_i \qquad \text{for all } i = 1, \cdots, k-1. \end{aligned}$$

We will denote the map going from subspaces to principal angles by

$$\Theta(\mathscr{U}, \mathscr{V}) = \boldsymbol{\theta}.$$

In words, the kth principal angle between subspaces $\mathscr{U}$, $\mathscr{V}$ is defined to be the largest possible inner product between vectors in $\mathscr{U}$ and $\mathscr{V}$ that is orthogonal to the

$k-1$ largest principal directions. From this coordinate-free definition it is clear that the angle of principal vectors is well-defined and that it yields a nonincreasing vector $\boldsymbol{\theta}$. The vectors $\{\mathbf{u}_1, \cdots, \mathbf{u}_r\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ are referred to as the **principal vectors** between $\mathscr{U}$ and $\mathscr{V}$, although they are not necessarily unique. Note that if $(\mathbf{u}_i, \mathbf{v}_i) \leftrightarrow 0$ are principal angles corresponding to the angle zero then we can insist without loss of generality that $\mathbf{u}_i = \mathbf{v}_i$

in a small enough neighborhood of $\mathbf{x}_0$ there are In practice it is often convenient to work with subspaces which are already in coordinates. To that end [29] develops an equivalent definition or formula for the angles between subspaces in given bases:

Theorem 1 (SVD gives principal angles in coordinates, [29])**.** Let $\mathscr{X}$ and $\mathscr{Y}$ be subspaces of $\mathbb{R}^n$ and let $\mathbf{X}$ and $\mathbf{Y}$ be corresponding orthonormal bases. Define $\mathbf{U}\Sigma\mathbf{V} = \mathbf{X}^*\mathbf{Y}$ to be the singular value decomposition of $\mathbf{X}^*\mathbf{Y}$. Then the singular values Σ are the cosines of the principal angles:

$$\cos\theta_k = \Sigma_{kk} \tag{3.1.1}$$

It is easy to see from this relationship that the map Θ is symmetric, that it is unitarily invariant (since the singular values are), and that it satisfies a nice algebra with respect to complements and inclusions (see [30] and [31]). Recent interesting work has shown—among much else—that any metric measuring the distance between two subspaces that is invariant to rotations can only depend on the principal angles between the two subspaces and the difference in dimension (c.f. Theorem 2.2 and Proposition 6.2 in [32]).

An equivalent definition of one of the principal angles will sometimes be useful.

Given a subspace $\mathcal{U}$ and a vector $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$, define the distance from $\mathbf{x}$ to $\mathcal{U}$ as

$$\sin\theta(\mathbf{x}, \mathcal{U}) = \min_{\mathbf{u}\in\mathcal{U}} \|\mathbf{u} - \mathbf{x}\|_2 .$$

Rewriting this via projection we find that

$$\sin\theta(\mathbf{x}, \mathcal{U}) = \|(\mathbf{I} - \mathbf{P}_{\mathcal{U}})\mathbf{x}\|_2 .$$

This basic idea can be extended to measure distances between subspaces as follows:

Definition 8 (Sine of angle between subspaces)**.** Given subspaces $\mathcal{U}$ and $\mathcal{V}$ define the (sine of the) angle between them by

$$\|\sin\theta(\mathcal{U}, \mathcal{V})\| = \|(\mathbf{I} - \mathbf{P}_{\mathcal{U}})\mathbf{P}_{\mathcal{V}}\| \tag{3.1.2}$$

It is not hard to see that the angle θ defined by this definition is exactly the smallest angle defined by the previous definition, see again [30] and Theorem 2.2 of [31] for nice overviews. Finally, we can define the principal angle between affine spaces by simply translating each of the affine spaces to the origin and measuring the angle there.

3.2 Davis-Kahan $\sin\theta$ theorem

In this section we present a version of a Davis-Kahan-like theorem restricted to the setting that we care about and use it to prove calmness of the set $\{\mathbf{x} : \mathbf{A}_0\mathbf{x} = \mathbf{0}\}$. We follow convention in describing these results as forms of Davis-Kahan, since they deal with the perturbation of eigenvectors or singular vectors. The original form of the Davis-Kahan theorem [28] deals with Hermitian matrices and the result we will actually use will be a simplification of a result in [33].

We begin with a simplified version of Davis-Kahan.

Theorem 2 (Davis-Kahan [28])**.** Let $\mathbf{A}_0, \mathbf{A}_1$ be symmetric matrices of rank q with $\|\mathbf{A}_0 - \mathbf{A}_1\| < \sigma_{\min}(\mathbf{A}_0)$. Then

$$\|\sin(\theta(N(\mathbf{A}_0, N(\mathbf{A}_1))))\| \leq \frac{1}{\sigma_{\min}(\mathbf{A}_0)} \|\mathbf{A}_0 - \mathbf{A}_1\|. \tag{3.2.1}$$

In words, this bounds the null space of a matrix with

In fact, an almost identical version of this theorem holds for singular vectors of general (non-square) matrices (see [33], equation (3.12)) at the expense of significantly more complicated proof:

Theorem 3 (Davis-Kahan for the null space of fat matrices)**.** Let $\mathbf{A}_0 \in \mathbb{R}^{m\times n}$ with $m < n$ and $\mathbf{rank}\mathbf{A}_0 = m$. Let $\mathbf{A}_1 = \mathbf{A}_0 + \mathbf{E}$ also have rank m, and denote by $\sigma_m(\mathbf{A}_0) > 0$ the smallest singular value of $\mathbf{A}_0$. Then

$$\|\sin\theta\,(N(\mathbf{A}_0), N(\mathbf{A}_1))\| \leq 2 \cdot \frac{\|\mathbf{A}_0 - \mathbf{A}_1\|}{\sigma_{\min}(\mathbf{A}_0)}. \tag{3.2.2}$$

Proof. See [33]. □

We can use this version to show that the constraint mapping $M(\mathbf{A}_0) = \{\mathbf{x} : \mathbf{A}_0\mathbf{x} = 0\}$ is calm when the data is zero.

Theorem 4. Let $\mathbf{b} = \mathbf{0}$. Then the constraint mapping $M(\mathbf{A})$ is calm at $(\mathbf{A}_0, \mathbf{x}_0)$ when $\mathbf{A}_0$ has full-rank and for any $\mathbf{x}_0 \in M(\mathbf{A}_0)$.

Proof. By Davis-Kahan if we take $\epsilon = \|\mathbf{x}_0\|$ and $\delta < \frac{1}{2\sigma_{\min}(\mathbf{A}_0)}$ then for any $\mathbf{A}_1 \in B_\delta(\mathbf{A}_0)$

and $\mathbf{x}_1 \in M(\mathbf{A}_1) \cap B_\epsilon(\mathbf{x}_0)$

$$d(\mathbf{x}_1, M(\mathbf{A}_0)) \leq \frac{2}{\sigma_{\min}(\mathbf{A}_0)} \|\mathbf{A}_0 - \mathbf{A}_1\| .$$

□

In the following section we will generalize this to the case when $\mathbf{b} \neq 0$.

3.3 Hoffman's lemma

Extending the previous work based on Davis-Kahan is not immediately trivial when the constraint spaces $M(\mathbf{A}_0)$, $M(\mathbf{A}_1)$ are affine. In fact, in this scenario it is not even clear that the two spaces intersect (we would expect them not to, in general, unless $m < \frac{n}{2}$ by counting their dimensions) and so it is unclear how knowing the sine of the angle between them as affine spaces might apply to yield the kind of bounds we want. In this section, we use Hoffman's lemma [34] to close this gap.

Lemma 1 (Hoffman's Lemma)**.** Let P be a non-empty polyhedron given by

$$P := \left\{\mathbf{z} \in \mathbb{R}^n : \mathbf{E}\mathbf{z} \leq \mathbf{b}\right\},$$

where $\mathbf{E} \in \mathbb{R}^{m\times n}$ and let $\mathbf{x}$ be any other point in $\mathbb{R}^n$. Then there exists a constant $c = c(\mathbf{E}) > 0$ such that

$$d(\mathbf{x}, P) \leq c \cdot \|(\mathbf{E}\mathbf{x} - \mathbf{b})_+)\| , \tag{3.3.1}$$

where $(\mathbf{y})_+$ denotes the (entrywise) maximum $\max(\mathbf{y}, \mathbf{0})$.

We will denote the optimal such value of $c(\mathbf{E})$ by $H = H(\mathbf{E})$

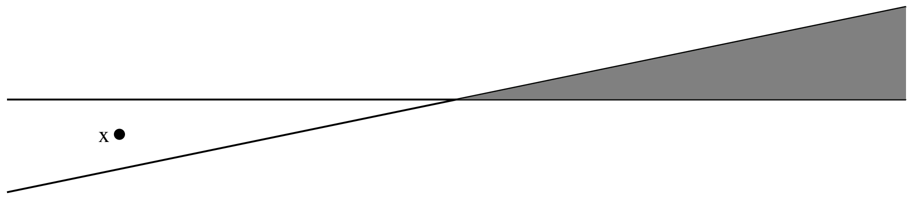

Figure 3.1: An "ill-conditioned" polyhedron (in the sense that the angle between subspaces is small). The shaded gray section represents the polyhedron with the lines the boundaries of the corresponding half spaces.

For a visualization of the geometry of Hoffman's lemma see Figure 3.1. Note that the point $\mathbf{x}$ almost satisfies both of the inequalities (i.e., it is close to being in both half spaces) but it is far from the polyhedron itself. It is clear from this figure that the constant $c(\mathbf{E})$ can get arbitrarily large as the slopes of these lines get closer.

Originally a completely geometric result (as it is still presented here), it now extends to much more complicated systems of convex inequalities via Fenchel duality [35]. It is known to be equivalent to Farkas' lemma [24], one of the theoretical workhorses of the optimization community. There is a long line of very interesting work on how to define or compute the optimal constant $c(\mathbf{E})$. In the absence of equality constraints, we can define

$$H(\mathbf{E}) = \max\left\{ \left\| \mathbf{E}_I^{-1} \right\| : |I| = m \text{ and } \mathbf{E}_I \text{ nonsingular} \right\}, \tag{3.3.2}$$

and show that $H(\mathbf{E}) = c(\mathbf{E})$ is sharp [36]. See [37] and [38] for a discussion on $c(\mathbf{E})$, [39] for very interesting work on the sensitivity of $H(\mathbf{E})$ to perturbations in a more general setting, along with recent work [40, 41] establishing the equivalence (under certain conditions) between Hoffman's constant and two other well-studied quantities in other areas of the literature: the chi measure and Renegar's distance to ill-posedness. We will begin by using Hoffman's lemma to bound the normalized distance between two affine spaces.

Note that an affine space is a polyhedral set. Therefore,

Lemma 2. Let $\mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{m\times n}$ be matrices with $\mathbf{b} \in \mathbb{R}^m$ and define the corresponding solution sets $M(\mathbf{A}_i) = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}_i\mathbf{x} = \mathbf{b}\right\}$. Then, for all $\mathbf{x}_1 \in M(\mathbf{A}_1)$

$$\frac{d(\mathbf{x}_1, \mathbf{A}_0)}{\|\mathbf{x}_1\|} \le H(\mathbf{A}_0)\cdot\|\mathbf{A}_0 - \mathbf{A}_1\|\,. \tag{3.3.3}$$

Proof. Note that $\mathbf{A}_1\mathbf{x}_1 = \mathbf{b}$ and so

$$\begin{aligned}\mathbf{A}_0\mathbf{x}_1 - \mathbf{b} &= \mathbf{A}_0\mathbf{x}_1 - \mathbf{b} - (\mathbf{A}_1\mathbf{x}_1 - \mathbf{b})\\ &= (\mathbf{A}_0 - \mathbf{A}_1)\,\mathbf{x}_1\end{aligned}$$

and that $\|(\mathbf{A}_0\mathbf{x}_1 - \mathbf{b})_+\| \le \|\mathbf{A}_0\mathbf{x}_1 - \mathbf{b}\|$. Therefore, by Hoffman's lemma

$$\begin{aligned}d(\mathbf{x}_1, M(\mathbf{A}_0)) &\le H(\mathbf{A}_0)\cdot\|(\mathbf{A}_0\mathbf{x}_1 - \mathbf{b})_+\|\\ &\le H(\mathbf{A}_0)\,\|\mathbf{x}_1\|\cdot\|\mathbf{A}_1 - \mathbf{A}_0\|\,.\end{aligned}$$

□

Using the condition in (P(**A**)) that $\|\mathbf{x}_1\| \le R$ we can immediately use this theorem to extend Theorem 4 to the case when $\mathbf{b} \ne 0$ as long as both problems are feasible. It is sufficient for both $\mathbf{A}_0$ and $\mathbf{A}_1$ to be full-rank for both problems to be feasible.

3.3.1 Analogue between Davis-Kahan and Hoffman

Given their similarities, one might wonder about the exact nature of the relationship between Davis-Kahan and Hoffman. For example, could Hoffman be used to prove (a version of) Davis-Kahan? There are some parallels. For example, it is readily apparent that

$$\frac{1}{\sigma_{\min}(\mathbf{A}_0)} \leq H(\mathbf{A}_0) \tag{3.3.4}$$

from the form of (3.3.2). But this bound can be arbitrarily bad, as the following example shows:

Example 1. The matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \epsilon \end{bmatrix}$$

has $H(\mathbf{A}) = \frac{1}{\epsilon}$ while $1/\sigma_{\min}(\mathbf{A}) = 1$.

The constant for Davis-Kahan is tight in the worst case. A recent line of literature has explored the *average case* bounds for Davis-Kahan using techniques from random matrix theory and careful analysis involving the interactions between the principal angles of the perturbation and the original eigenspaces (see [42, 43] and citations, for example). It would be interesting to also know the average-case behavior of Hoffman's constant. As a first step in this direction, consider Figure 3.2, which plots the smallest singular value of a random matrix $\mathbf{A} \in \mathbb{R}^{2\times 4}$ where $\mathbf{A}_{i,j} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1)$. Note that it appears to possess some regularity in these random instances.

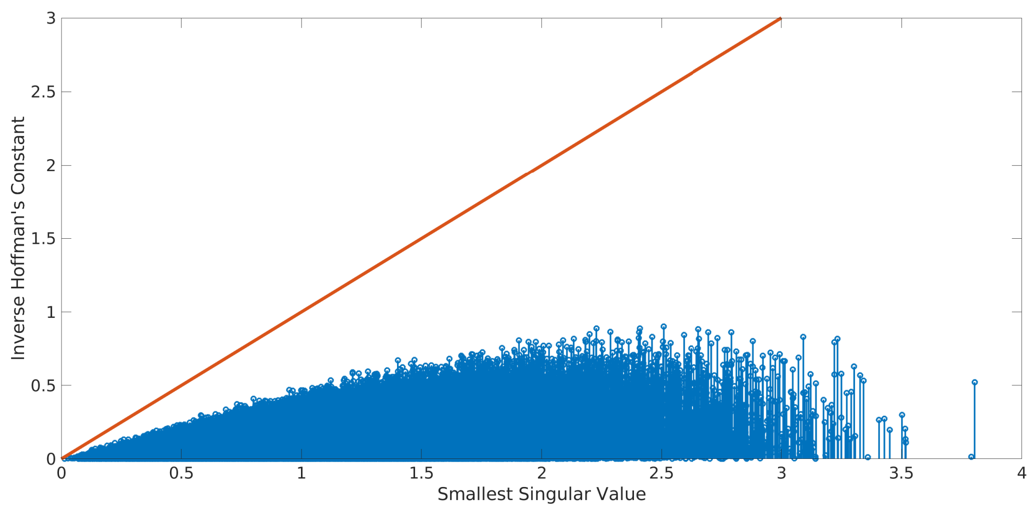

Figure 3.2: A stem plot displaying the smallest singular value of a random matrix $\mathbf{A}$ on the x-axis and $1/H(\mathbf{A})$ the inverse Hoffman constant on the y-axis. The line $y = x$ is included in red for reference.

Chapter 4

Calmness of the Armgin

In this chapter we state and prove our main result. We begin with a definition.

Definition 9 (General position)**.** Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ with $m < n$. We will say that the *columns of* $\mathbf{A}$ *are in general position* if for any subset I of size m the submatrix $\mathbf{A}_I$ will be invertible.

Note, for example, that $\mathbf{A}$ in general position is much stronger than $\mathbf{A}$ having full row-rank.

Theorem 4. let $\mathbf{A}_0 \in \mathbb{R}^{m\times n}$ with the columns of $\mathbf{A}_0$ in general position and let $\Omega(\cdot)$ be a (semi)norm with polyhedral sublevel sets. Define the solution map

$$\begin{aligned} S(\mathbf{A}) := \arg\min \quad & \Omega(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \|\mathbf{x}\| \leq R. \end{aligned} \tag{4.0.1}$$

Then, for any $\mathbf{x}_0 \in S(\mathbf{A}_0)$, S is calm at $(\mathbf{A}_0, \mathbf{x}_0)$. In particular, for $\mathbf{A}_0, \mathbf{A}_1$ sufficiently close, there exist solutions $\mathbf{x}_0 \in S(\mathbf{A}_0), \mathbf{x}_1 \in S(\mathbf{A}_1)$ such that

$$\|\mathbf{x}_1 - \mathbf{x}_0\| \leq C \|\mathbf{A}_0 - \mathbf{A}_1\|. \tag{4.0.2}$$

Note that this theorem applies to certain functions $\Omega(\mathbf{x})$ which are *not* strictly convex. As motivation consider the following functions to which the theorem applies.

Example 2 (Examples of applicable Ω)**.** Note that Theorem 4 applies to the following functions Ω.

1. $\Omega(\mathbf{x}) = \|\mathbf{x}\|_1$. ℓ_1 minimization methods have been utilized by both Galileo (1632) and Laplace (1793). It has long been known that ℓ_1 minimization in certain settings facilitates a certain sense of robustness: for example ℓ_1 minimization is an estimate of the median of a distribution, while ℓ_2 minimization estimates the mean.

   Its popularity in modern times is just as often related to its sparsity-promoting qualities in certain contexts. This line of work was first explored by Santosa and Symes in 1986 [44] when they proposed ℓ_1 minimization as a method of recovering sparse spike trains. It was further generalized and theoretically grounded by Donoho and Starks in 1989 [45]. Over the next decade research into ℓ_1 minimization and closely related methods exploded in popularity as a succession of results showed that ℓ_1 minimization is in fact equivalent to the NP Hard problem of ℓ_0 minimization in a variety of situations and that these situations can be good models for real problems (see [45, 46, 47] for a smattering of results of this flavor). Over the following decade a general theory known as compressed sensing grew out of these developments (see [48, 49, 50, 51] for the early and foundational work in the area and [52] for a very nice book on the topic). See Figures 4.1 and 4.2 for a two dimensional cartoon displaying the geometry of the underlying $\Omega(\mathbf{x}) = \|\mathbf{x}\|_1$ and $\Omega(\mathbf{x}) = \|\mathbf{x}\|_2$ problems.

2. $\Omega(\mathbf{x})$ is the (anisotropic) discrete total variation seminorm. Let $\mathbf{X} \in \mathbb{R}^{n\times n}$ be a matrix representing an image (in grayscale). Then we define the anisotropic total variation of $\mathbf{X}$ to be

$$\Omega(\mathbf{X}) := \sum_{i,j=1}^{n-1} \left|\mathbf{X}_{i+1,j} - \mathbf{X}_{i,j}\right| + \left|\mathbf{X}_{i,j+1} - \mathbf{X}_{i,j}\right| 1.$$

This seminorm was first proposed as a means of noise reduction in natural images by Rudin, Osher, and Fatemi in [53] (it is sometimes known as the ROF functional in their honor). Assuming that images have low total variation has proved to be a popular model for natural images essentially ever since, both as an object of theoretical research (see [54, 55], for example) as well as a tool in practical applications (see [56, 57], for example). Note that in this problem we only deal with the anisotropic version of total variation, as the isotropic version (which replaces our absolute values with an ℓ^2 norm) does not lead to polyhedral sublevel sets. See [54] and citations for more background on the difference.

The proof of this theorem will require checking the conditions of several technical lemmas from [10], which we reproduce below in the current setting. Define the sublevel set of Ω as

$$F(\alpha) = \{\mathbf{x} : \Omega(\mathbf{x}) \le \alpha\}, \tag{4.0.3}$$

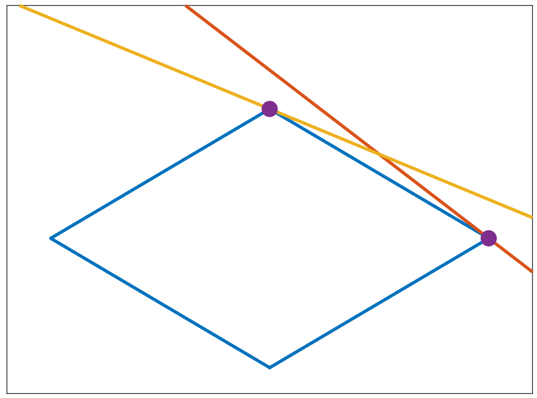

Figure 4.1: The geometry of the $\|\mathbf{x}\|_1$ problem for $n = 2$. The ℓ_1 ball is displayed intersecting with two different affine spaces representing $M(\mathbf{A}_0)$ and $M(\mathbf{A}_1)$. Note that despite the small difference of angle between these two affine spaces the respective $\arg\min$s are far apart.

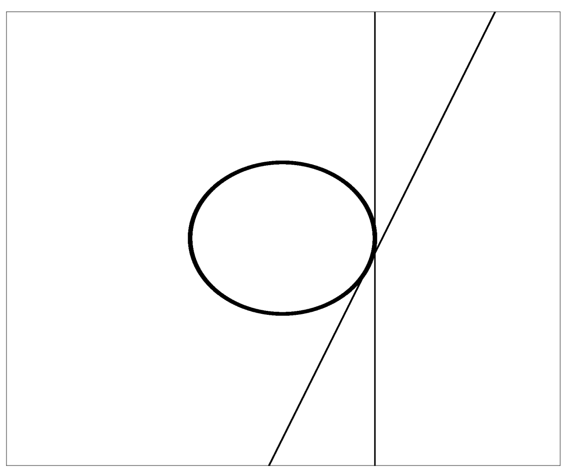

Figure 4.2: The geometry of the $\|\mathbf{x}\|_2$ problem for $n = 2$. The ℓ_2 ball is once again displayed intersecting with the two different affine spaces representing $M(\mathbf{A}_0)$ and $M(\mathbf{A}_1)$. Note, in contrast with Figure 4.1, how well-behaved the argument minimizer is (as explained by the theory in Section 2.1).

and the intersection map:

$$L(\mathbf{A},\alpha) = M(\mathbf{A}) \cap F(\alpha), \tag{4.0.4}$$

where we recall that the map $M(\mathbf{A}) = \left\{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\right\}$ is the constraint set. Define $\phi(\mathbf{A})$ to be the infimum value function for a given matrix $\mathbf{A}$

$$\phi(\mathbf{A}) := \inf_{\mathbf{x}} \{\Omega(\mathbf{x}) | \mathbf{x} \in M(\mathbf{A})\}.$$

Our proof of Theorem 4 incorporates the following characterizations of calmness.

Lemma 3 (Klatte and Kummer [10], Theorem 4.1)**.** The intersection map $L(\mathbf{A},\alpha) = M(\mathbf{A}) \cap F(\alpha)$ is calm at a point $(\mathbf{A},\alpha,\mathbf{x}) \in \mathbf{Graph}\, L$ if the following four conditions hold:

1. M is calm at $(\mathbf{A},\mathbf{x})$,
2. F is calm at $(\alpha,\mathbf{x})$,
3. F^{-1} is Aubin at $(\mathbf{x},\alpha)$, and
4. $Q(\alpha) = M(\mathbf{A}) \cap F(\alpha)$ is calm at $(\alpha,\mathbf{x})$.

Lemma 4 (Klatte and Kummer [10], Theorem 3.1)**.** Let $\mathbf{A}_0$ have full row rank and let $\mathbf{x}_0 \in S(\mathbf{A}_0)$ be a minimizer of (4.0.1). If the following two conditions hold:

1. the constraint set $M(\mathbf{A})$ is calm and Lipschitz lower semicontinuous at $(\mathbf{A}_0,\mathbf{x}_0)$ and
2. the multifunction $L(\mathbf{A},\alpha)$ is calm at $((\mathbf{A}_0,\phi(\mathbf{A}_0)),\mathbf{x}_0)$,

then S is calm at $(\mathbf{A}_0,\mathbf{x}_0)$.

Both of these results are restrictions to the current setting of general facts about argmin mappings, and their proofs can be found in [10]. The first, Lemma 3 is a general result on when the intersection (in this setting, L) of two set-valued mappings is calm in terms of the subsets that make up the intersection. The second roughly states that when the growth of the constraint set is controlled and the intersection of the constraint set with the sublevel sets does not explode then the argmin mapping S is nicely behaved.

Thus, to prove Theorem 4 it remains to check the hypotheses of these theorems. Note that we have already shown that the constraint set mapping is calm in Chapter 3 Section 3.3 using Hoffman's lemma and again in Section 3.2 when $\mathbf{b} = 0$. We will next prove an auxiliary lemma which will help in showing that $F(\alpha)$ is calm.

Lemma 5. If Ω has polyhedral sublevel sets then there exists $c > 0$ such that for all $\mathbf{x}_0$ with $\Omega(\mathbf{x}_0) \neq 0$ we can choose $\mathbf{v}_{\mathbf{x}_0} \in \partial\Omega(\mathbf{x}_0)$ such that

$$c \leq \left\| \mathbf{v}_{\mathbf{x}_0} \right\|.$$

Proof. Let $\mathbf{x}_0 \in \mathbb{R}^n$ with $\Omega(\mathbf{x}_0) > 0$ and let $\mathbf{v} \in \partial\Omega(\mathbf{x}_0)$. We claim that a vector $\mathbf{v}$ is a subgradient at the point $\mathbf{x}_0$ if and only if it is the normal of a supporting hyperplane to the sublevel set $F(\alpha)$ at $\mathbf{x}_0$. Indeed, let $\Omega(\mathbf{x}_0) = \alpha$. By assumption, since $\Omega(\mathbf{x}_0) \neq 0$, $\mathbf{0} \notin \partial\Omega(\mathbf{x}_0)$. Let $\mathbf{x} \in F(\alpha)$. Since $\mathbf{v}$ is a subgradient, $\Omega(\mathbf{x}) \geq \Omega(\mathbf{x}_0) + \langle \mathbf{v}, \mathbf{x} - \mathbf{x}_0 \rangle$. Therefore

$$\begin{aligned} \langle \mathbf{v}, \mathbf{x} \rangle &\leq \langle \mathbf{v}, \mathbf{x}_0 \rangle + \overbrace{(\Omega(\mathbf{x}) - \Omega(\mathbf{x}_0))}^{\leq 0 \text{ since } \mathbf{x} \in F(\alpha)} \\ &\leq \langle \mathbf{v}, \mathbf{x}_0 \rangle, \end{aligned}$$

and thus $\mathbf{v}$ is the normal of a supporting hyperplane to $F(\alpha)$.

Conversely, if we have a supporting hyperplane to $F(\alpha)$ at $\mathbf{x} \in F(\alpha)$ then it is defined by some equality $\langle \mathbf{v}, \mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{x} \rangle$ and by convexity for all $\mathbf{x} \in \mathscr{F}[(]\alpha)$, $\langle \mathbf{v}, \mathbf{x} \rangle \le \langle \mathbf{v}, \mathbf{x}_0 \rangle$. Furthermore, this extends to a supporting hyperplane on $\mathbf{epi}\,\Omega$, the epigraph of Ω, by considering the vector $[\mathbf{v}, -1]^T$. Since this vector supports $\mathbf{epi}\,\Omega$ at $[\mathbf{x}_0, \Omega(\mathbf{x}_0)]^T$,

$$\begin{bmatrix} \mathbf{v} & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \le \begin{bmatrix} \mathbf{v} & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \Omega(\mathbf{x}_0) \end{bmatrix}$$

for all $\begin{bmatrix} \mathbf{x} & t \end{bmatrix}^T$ in $\mathbf{epi}\,\Omega$. In particular, this is true for $t = \Omega(\mathbf{x})$, which implies, after rearranging, that $\Omega(\mathbf{x}) \ge \Omega(\mathbf{x}_0) + \mathbf{v}^*(\mathbf{x} - \mathbf{x}_0)$ and so $\mathbf{v} \in \partial\Omega(\mathbf{x}_0)$.

Since $F(\alpha)$ is polyhedral it can be defined by the intersection of finitely many half spaces given by a supporting hyperplane, say with normal vectors $\{\mathbf{n}_i\}_{i=1}^k$. Thus, we can choose c in the lemma to be $c := \min_i \|\mathbf{n}_i\|$. □

Note that Lemma 5 is closely related to the so-called "sticky face lemma" [58], which describes the local regularity with the inverse of the normal-cone operator of a polyhedral convex set.

Lemma 6. The map $F(\alpha)$ is calm at $(\alpha_0, \mathbf{x}_0)$ for any $\mathbf{x}_0$ such that $\Omega(\mathbf{x}_0) \neq 0$.

Proof. Fix $(\alpha_0, \mathbf{x}_0)$. To prove calmness we need to show that given an α_1 sufficiently close to α_0 and a point $\mathbf{x}_1 \in F(\alpha_1)$ we can bound the distance between $\mathbf{x}_1$ and $F(\alpha_0)$. But not that if $\alpha_1 < \alpha_0$ then we have that $F(\alpha_1) \subset F(\alpha_0)$ and so the result is trivial.

Therefore, consider the case $\alpha_1 > \alpha_0$ and let $\mathbf{x} \in F(\alpha_1)$. We will construct a $y \in F(\alpha_0)$ such that

$$\|x - y\| \le \frac{1}{c} |\alpha_1 - \alpha_0|,$$

with the c defined by Lemma 5. Choose $v \in \partial(-\Omega(\mathbf{x})) = -\partial\Omega(\mathbf{x})$ such that $\|v\| \ge c$. By definition of the (negative) subdifferential, we have that

$$\langle \mathbf{v}, \mathbf{y} - \mathbf{x}, \leq \rangle \Omega(\mathbf{x}) - \Omega(\mathbf{y}). \tag{4.0.5}$$

Define $\Delta\alpha = \alpha_1 - \alpha_0$ and let $\mathbf{y} = \mathbf{x} + \Delta\alpha \frac{\mathbf{v}}{\|\mathbf{v}\|^2}$ Then, expanding the inequality in Equation (4.0.5) we find that

$$\Delta\alpha \leq \Omega(\mathbf{x}) - \Omega(\mathbf{y}).$$

But by assumption $\Omega(\mathbf{x}) \leq \alpha_1$, so it follows that $\Omega(\mathbf{y}) \leq \alpha_0$, as desired. Therefore,

$$\left\| \mathbf{y} - \mathbf{x} \right\| = \frac{\Delta\alpha}{\|v\|} \leq \frac{1}{c} |\alpha_0 - \alpha_1|.$$

□

Note that in this proof we actually proved something stronger than was required: the c we chose was independent of the point $(\alpha_0, \mathbf{x}_0)$. This is the second condition needed to invoke Lemma 3.

It remains to prove the remaining three conditions needed to invoke Lemmas 3 and 4. Let us next show the third condition of Lemma 3 along with an easy proof that Hoffman's constant is continuous in the current setting, which will be of use in proving that the constraint mapping $M(\mathbf{A})$ is Lipschitz lower semicontinuous.

Lemma 7. The map F^{-1} is Aubin at all points.

Proof. Define

$$\alpha = \Omega(\mathbf{x}) \in F^{-1}(\mathbf{x}) \text{ and } \alpha' = \Omega(\mathbf{x}') \in F^{-1}(\mathbf{x}').$$

Without loss of generality assume that $\alpha' \geq \alpha$. Then

$$F^{-1}(\mathbf{x}') = [\alpha', \infty) \subseteq F^{-1}(\mathbf{x}) = [\alpha, \infty).$$

Thus it suffices to prove that the endpoints are calm. That is, the map $\mathbf{x}' \mapsto \alpha'$ is calm. Since all convex functions are locally Lipschitz it follows that

$$\left\|\alpha' - \alpha\right\| \leq K \left\|\mathbf{x} - \mathbf{x}'\right\|.$$

□

Lemma 8. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be in general position.

Then for any $\epsilon > 0$ there exists a $\delta > 0$ such that for all $\mathbf{B} \in B_\delta(\mathbf{A})$ we have

$$H(\mathbf{B}) \leq (1+\epsilon) H(\mathbf{A}).$$

Proof. Assume without loss of generality that $\epsilon < \frac{1}{3}$. Let $J \subset [n]$ be a subset of cardinality m such tat $H(\mathbf{A}_0) = \left\|\mathbf{A}_J^{-1}\right\| = \frac{1}{\sigma_{\min}(\mathbf{A}_J)}$. Now let $I \subset [n]$ be an arbitrary subset with $|I| = m$.

Since $\mathbf{A}$ is full-rank, $\mathbf{A}_I$ is invertible and thus $\left\|\mathbf{A}_I^{-1}\right\| \leq \left\|\mathbf{A}_J^{-1}\right\|$ by the definition of Hoffman's constant. That is, $\sigma_{\min}(\mathbf{A}_I) \geq \sigma_{\min}(\mathbf{A}_J)$. Choose $\delta < \epsilon/(1-\epsilon)\sigma_{\min}(\mathbf{A}_J)$ and let $\mathbf{B} \in B_\delta(\mathbf{A})$. Then it is not hard to see that $\mathbf{B}_I \in B_\delta(\mathbf{A}_J)$. Therefore, by Weyl's inequality it follows that $\left|\sigma_{\min}(\mathbf{A}_I) - \sigma_{\min}(\mathbf{A}_J)\right| < \delta$. Rearranging,

$$\frac{1}{\sigma_{\min}(\mathbf{B}_I)} \leq \frac{1}{\sigma_{\min}(\mathbf{A}_I) - \delta} \leq (1+\epsilon) H(\mathbf{A}),$$

where the last inequality follows by our choice of ϵ and since I was an arbitrary subset of the correct size. □

Note that the above proof critically relies on the fact that $\mathbf{A}$ is full rank (else $\mathbf{A}_I$ may not be invertible). It is easy to come up with examples where the converse is not true.

Example 3. Define the matrix $\mathbf{A} = \mathbb{R}^{2\times 3}$ by

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

Then note that $H(\mathbf{A}) = 1$ but that the sequence

$$\mathbf{B}_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1/k \end{bmatrix}$$

has $\|\mathbf{A} - \mathbf{B}_k\| = \frac{1}{k}$ and $H(\mathbf{B}_k) \to \infty$.

We will use Lemma 8 to show that the constraint set is Lipschitz lower semicontinuous.

Theorem 5. The constraint set mapping $M(\mathbf{A})$ is Lipschitz lower semicontinuous at $(\mathbf{A}_0, \mathbf{x}_0) \in \textbf{Graph}\, M(\mathbf{A})$ for $\mathbf{A}$ with columns in general position.

Proof. Let $\mathbf{A}_0$ have columns in general position and let $\mathbf{x}_0 \in \{\mathbf{x} : \mathbf{A}_0\mathbf{x} = \mathbf{b}\}$. Then Lemma 2 proves that, for any $\mathbf{A}_1$ sufficiently close, the bound

$$d(\mathbf{x}_0, M(\mathbf{A}_1)) \leq H(\mathbf{A}_1)\, \|\mathbf{x}_0\|\, \|\mathbf{A}_0 - \mathbf{A}_1\|$$

holds. Lemma 8 then shows that for any $\epsilon > 0$ there exists a $\delta > 0$ such that for all $\mathbf{A}_1 \in B_\delta(\mathbf{A}_0)$

$$H(\mathbf{A}_1) \leq (1+\epsilon)H(\mathbf{A}_0).$$

Combining these two results for $\mathbf{A}_1$ sufficiently close to $\mathbf{A}_0$ we find that

$$d(\mathbf{x}_1, M(\mathbf{A}_1)) \leq (1+\epsilon)H(\mathbf{A}_0)R\,\|\mathbf{A}_0 - \mathbf{A}_1\|\,.$$

□

The final thing to be shown is just that the intersection map is calm. First define the *infimal value function* of the optimization problem (4.0.1):

$$\varphi(\mathbf{A}_0) = \inf_{\mathbf{x}}\{\Omega(\mathbf{x})|\mathbf{x} \in M(\mathbf{A})\}\,. \tag{4.0.6}$$

Lemma 9. The map

$$Q(\alpha) = M(\mathbf{A}_0) \cap F(\alpha_0)$$

is calm at $(\varphi(\mathbf{A}_0), \mathbf{x}_0) \in \textbf{Graph}\, H(\alpha)$.

Proof. Define $\alpha_0 = \varphi(\mathbf{A}_0)$. Note that if $\alpha_1 < \alpha_0$ then $Q(\alpha_1) = \emptyset$ and so the calmness condition is trivially satisfied. Therefore, assume that $\alpha_1 > \alpha_0$. We are concerned with the maximum distance a point $\mathbf{x}_1 \in Q(\alpha_1)$ can be from $Q(\alpha_0)$. By monotonicity of $F(\alpha)$ it follows that $Q(\alpha_0) \subset Q(\alpha_1)$ and by definition of $Q(\alpha)$ and convexity it will be a closed, convex subset of $M(\alpha)$.

Since $\Omega(\mathbf{x})$ has polyhedral sublevel sets, it follows that $M(\mathbf{A}_0)$ intersects $F(\alpha_0)$ only on the boundary of $F(\alpha_0)$ (else α_0 would not be minimal). Therefore, $Q(\alpha_0) = M(\mathbf{A}_0) \cap F(\alpha_0)$ is a subset of a k-face of $F(\alpha_0)$, for some dimension k.

Every k face of $F(\alpha)$ is the intersection of finitely many halfspaces (in fact, $2(n-k)$ of them are enough). Write the affine planes that are the boundaries of these half spaces as $\{\mathcal{U}_i\}_{i=1}^r$ and let this representation be minimal.

Let $\{\mathbf{u}_1^i, \mathbf{u}_2^i, \cdots, \mathbf{u}_k^i\}, \{\mathbf{v}_1^i, \mathbf{v}_2^i, \mathbf{v}_k^i\}$, and $\{\boldsymbol{\theta}_1^i, \boldsymbol{\theta}_2^i, \cdots, \theta_k^i\}$ be principal vectors and principal angles associated with $\mathcal{U}_i$ and $M(\mathbf{A}_0)$ for each $i \in [r]$. Define θ^i to be the smallest non-zero principal angle between $\mathcal{U}_i$ and $M(\mathbf{A}_0)$ and let $\theta = \min_i \theta^i$.

Let $\mathbf{x}_1 \in Q(\alpha_1)$ and consider the projections $\mathbf{y} := \mathbf{P}_{Q(\alpha_0)}(\mathbf{x}_1)$ and $\mathbf{z} := \mathbf{P}_{F(\alpha_0)}(\mathbf{x}_1)$. Then by definition of θ it follows that the angle between $\mathbf{x}_1 - \mathbf{y}$ and $\mathbf{x}_1 - \mathbf{z}$ is greater than or equal to θ. Note also that the proof of Lemma 6 in fact shows that

$$\|\mathbf{x}_1 - \mathbf{z}\| \leq \frac{1}{c} |\alpha_0 - \alpha_1|,$$

for some constant c and for $\mathbf{x}_1$ in a neighborhood of $\mathbf{x}$ such that $\mathbf{x}_1 - \mathbf{z}$ is in the normal cone at $\mathbf{z}$.

For any fixed affine space $\mathcal{U}_i$ if $(\mathbf{u}, \mathbf{v}) \leftrightarrow 0$ are principal vectors corresponding to the principal angle zero (so that $\mathbf{u} = \mathbf{v}$ then $\mathbf{x}_1 - \mathbf{y} \perp \mathbf{u}$. Indeed say without loss of generality that $\langle \mathbf{x}_1 - \mathbf{y}, \mathbf{u} \rangle = a > 0$ then, for some $\lambda \leq 1$, $\mathbf{y} + a\lambda\mathbf{u} \in Q(\alpha_0)$ (since $\mathbf{u} = \mathbf{v}$). Furthermore,

$$\left\|\mathbf{x} - (\mathbf{y} + \alpha\lambda\mathbf{u})\right\|_2^2 = \left\|\mathbf{x}_1 - \mathbf{y}\right\|_2^2 - 2a^2\lambda + a^2\lambda \quad = \left\|\mathbf{x}_1 - \mathbf{y}\right\|_2^2 - 2a^2\left(\lambda - \frac{\lambda^2}{2}\right) \leq \left\|\mathbf{x}_1 - \mathbf{y}\right\|_2^2$$

for all $\lambda \leq 1$, contradicting minimality of $\mathbf{y}$.

Since $\mathbf{x}_1 - \mathbf{y}$ is orthogonal to every principal vector corresponding to the principal angle zero, it follows that the angle between $\mathbf{x}_1 - \mathbf{y}$ and $F(\alpha)$ is lower bounded by θ. We have thus bounded the angle and one of the sides of a right triangle, yielding the bound on the hypotenuse

$$\|\mathbf{x}_1 - \mathbf{z}\| \leq \frac{1}{c\sin\theta}\,|\alpha_0 - \alpha_1|\,.$$

□

All conditions of Lemmas 3 and 4 have now been verified and so the conclusion holds. Thus, applying Lemmas 3 and 4 we have proved Theorem 4.

Chapter 5

Discussion and Future Directions

In this thesis we introduce a class of convex functions: (semi)norms with polyhedral level sets and motivated their use in minimization problems. We state a relatively general result, Theorem 4 on the regularity properties of this minimization when their linear constraints are perturbed and showed where this work fits into the related literature. Finally, we introduce the relevant results and tools and undertake the proof.

The main ingredient that makes this proof work is the fine control of the geometry of the minimization problem inherent in minimizing a function with polyhedral sublevel sets. In particular, Hoffman's lemma can be made to do the brunt of the work in proving our main result, at the expense of inferior constants.

There are several principal directions that this work can be extended. Firstly, this work hints that Hoffman's lemma is intimately related to the Davis-Kahan Theorem. There has been much recent work on calculating optimal constants for the Davis-Kahan Theorem (see Section 3.3.1) but this relationship has not yet been fully explored. Secondly, it would be interesting to explicitly calculate the constants for the functions $\Omega(\mathbf{x}) = \|\mathbf{x}\|_1$ and $\Omega(\mathbf{x})$ the discrete total variation seminorm. It further remains to be seen under what conditions perturbations of the form considered here (to the operator $\mathbf{A}$) can be combined with perturbations to the data $\mathbf{b}$ to still achieve a calm $\arg\min$ mapping. On the other hand, it is possible that the conclusion to the theorem can also be slightly tightened into showing that the argmin mapping in question is locally upper Lipschitz semicontinuous, using idea along the lines of [24], possibly at the expense of inexplicit constants. We are very interested in pursuing this line of work. Finally, it is interesting to speculate on the minimal conditions it is necessary to impose on Problem (4.0.1) in order to achieve the same result. For example, much of the proof goes through immediately if we merely insist that $\Omega(\mathbf{x})$ be polyhedral. Similarly, it may be possible to prove Theorem 4 with $\mathbf{A}$ satisfying a weaker condition than being in general position.

Chapter 6

Support

Bibliography

[1] Alex Gutierrez, Michael Mullen, Di Xiao, Taylor Froelich, Jarvis Haupt, and Michael Garwood. “Reducing the Complexity of Model-Based MRI Reconstructions via Sparsification”. Preprint. 2019.

[2] Georg Still. “Lectures on parametric optimization: An introduction”. In: *Optimization Online* (2018).

[3] Diethard Klatte and Bernd Kummer. *Nonsmooth equations in optimization: regularity, calculus, methods and applications*. Vol. 60. Springer Science & Business Media, 2006.

[4] GW Stewart. “On the continuity of the generalized inverse”. In: *SIAM Journal on Applied Mathematics* 17.1 (1969), pp. 33–45.

[5] Per-Åke Wedin. “Perturbation theory for pseudo-inverses”. In: *BIT Numerical Mathematics* 13.2 (1973), pp. 217–232.

[6] Jiu Ding. “Perturbation analysis for the projection of a point to an affine set”. In: *Linear Algebra and its applications* 191 (1993), pp. 199–212.

[7] Anthony V Fiacco et al. *Introduction to sensitivity and stability analysis in nonlinear programming*. Vol. 90075. Academic press New York, 1983.

[8] Ivan Dokmanic and Rémi Gribonval. “Concentration of the Frobenius norm of generalized matrix inverses”. In: *SIAM Journal on Matrix Analysis and Applications* 40.1 (2019), pp. 92–121.

[9] María J Cánovas, Abderrahim Hantoute, Juan Parra, and F Javier Toledo. “Calmness of the argmin mapping in linear semi-infinite optimization”. In: *Journal of Optimization Theory and Applications* 160.1 (2014), pp. 111–126.

[10] Diethard Klatte and Bernd Kummer. “On calmness of the argmin mapping in parametric optimization problems”. In: *Journal of Optimization Theory and Applications* 165.3 (2015), pp. 708–719.

[11] Frank H Clarke. “A new approach to Lagrange multipliers”. In: *Mathematics of Operations Research* 1.2 (1976), pp. 165–174.

[12] Lionel Thibault. “Propriétés des sous-différentiels de fonctions localement lipschitziennes définies sur un espace de Banach séparable: Applications”. PhD thesis. 1976.

[13] R.T. Rockafellar. “Extensions of Subgradient Calculus with applications to optimization”. In: *Nonlinear Analysis* 9.7 (1985), pp. 665–698.

[14] JV Burke. “Calmness and exact penalization”. In: *SIAM Journal on control and optimization* 29.2 (1991), pp. 493–497.

[15] René Henrion and Jiří V Outrata. “Calmness of constraint systems with applications”. In: *Mathematical Programming* 104.2-3 (2005), pp. 437–464.

[16] Asen L Dontchev and R Tyrrell Rockafellar. “Implicit functions and solution mappings”. In: *Springer Monographs in Mathematics. Springer* 208 (2009).

[17] AD Ioffe. “Metric regularity—a survey part 1. theory”. In: *Journal of the Australian Mathematical Society* 101.2 (2016), pp. 188–243.

[18] Abderrahim Jourani and L Thibault. “Approximations and metric regularity in mathematical programming in Banach space”. In: *Mathematics of Operations Research* 18.2 (1993), pp. 390–401.

[19] A Jourani and L Thibault. “Metric regularity and subdifferential calculus in Banach spaces”. In: *Set-Valued Analysis* 3.1 (1995), pp. 87–100.

[20] Aleksandr Davidovich Ioffe. “Metric regularity and subdifferential calculus”. In: *Russian Mathematical Surveys* 55.3 (2000), p. 501.

[21] David W Walkup and Roger J-B Wets. “A Lipschitzian characterization of convex polyhedra”. In: *Proceedings of the American Mathematical Society* (1969), pp. 167–173.

[22] M Seetharama Gowda and Roman Sznajder. “On the Lipschitzian properties of polyhedral multifunctions”. In: *Mathematical programming* 74.3 (1996), pp. 267–278.

[23] Olvi L Mangasarian and T-H Shiau. “Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems”. In: *SIAM Journal on Control and Optimization* 25.3 (1987), pp. 583–595.

[24] Abderrahim Jourani. “Hoffman’s error bound, local controllability, and sensitivity analysis”. In: *SIAM Journal on Control and Optimization* 38.3 (2000), pp. 947–970.

[25] Shu Lu and Stephen M Robinson. “Variational inequalities over perturbed polyhedral convex sets”. In: *Mathematics of Operations Research* 33.3 (2008), pp. 689–711.

[26] Camille Jordan. “Essai sur la géométrie à n dimensions”. In: *Bulletin de la Société mathématique de France* 3 (1875), pp. 103–174.

[27] Hermann Weyl. “Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)”. In: *Mathematische Annalen* 71.4 (1912), pp. 441–479.

[28] Chandler Davis and William M Kahan. “Some new bounds on perturbation of subspaces”. In: *Bulletin of the American Mathematical Society* 75.4 (1969), pp. 863–868.

[29] Ake Björck and Gene H Golub. “Numerical methods for computing angles between linear subspaces”. In: *Mathematics of Computation* 27.123 (1973), pp. 579–594.

[30] Peizhen Zhu and Andrew V Knyazev. “Angles between subspaces and their tangents”. In: *Journal of Numerical Mathematics* 21.4 (2013), pp. 325–340.

[31] Merico Edward Argentati. “Principal angles between subspaces as related to Rayleigh quotient and Rayleigh Ritz inequalities with applications to eigenvalue accuracy and an eigenvalue solver”. PhD thesis. Citeseer, 2003.

[32] Ke Ye and Lek-Heng Lim. “Schubert varieties and distances between subspaces of different dimensions”. In: *SIAM Journal on Matrix Analysis and Applications* 37.3 (2016), pp. 1176–1197.

[33] Per-Åke Wedin. “Perturbation bounds in connection with singular value decomposition”. In: *BIT Numerical Mathematics* 12.1 (1972), pp. 99–111.

[34] Alan J Hoffman. “On approximate solutions of systems of linear inequalities”. In: *Selected Papers Of Alan J Hoffman: With Commentary*. World Scientific, 2003, pp. 174–176.

[35] James V Burke and Paul Tseng. “A unified analysis of Hoffman′s bound via Fenchel duality”. In: *SIAM Journal on Optimization* 6.2 (1996), pp. 265–282.

[36] Shuzhong Zhang. “Global error bounds for convex conic problems”. In: *SIAM Journal on Optimization* 10.3 (2000), pp. 836–851.

[37] Diethard Klatte and Gisbert Thiere. “Error bounds for solutions of linear equations and inequalities”. In: *Zeitschrift für Operations Research* 41.2 (1995), pp. 191–214.

[38] Diethard Klatte and Gisbert Thiere. “A note of Lipschitz constants for solutions of linear inequalities and equations”. In: *Linear algebra and its applications* 244 (1996), pp. 365–374.

[39] Dominique Azé and Jean-Noël Corvellec. “On the sensitivity analysis of Hoffman constants for systems of linear inequalities”. In: *SIAM Journal on Optimization* 12.4 (2002), pp. 913–927.

[40] Javier F Pena, Juan C Vera, and Luis F Zuluaga. “Equivalences among the chi measure, Hoffman constant, and Renegar’s distance to ill-posedness”. In: *arXiv preprint arXiv:1905.06366* (2019).

[41] Javier Pena, Juan Vera, and Luis Zuluaga. “New characterizations of Hoffman constants for systems of linear constraints”. In: *arXiv preprint arXiv:1905.02894* (2019).

[42] Justin Eldridge, Mikhail Belkin, and Yusu Wang. “Unperturbed: spectral analysis beyond Davis-Kahan”. In: *arXiv preprint arXiv:1706.06516* (2017).

[43] Sean O’Rourke, Van Vu, and Ke Wang. “Random perturbation of low rank matrices: Improving classical bounds”. In: *Linear Algebra and its Applications* 540 (2018), pp. 26–59.

[44] Fadil Santosa and William W Symes. "Linear inversion of band-limited reflection seismograms". In: *SIAM Journal on Scientific and Statistical Computing* 7.4 (1986), pp. 1307–1330.

[45] David L Donoho and Philip B Stark. "Uncertainty principles and signal recovery". In: *SIAM Journal on Applied Mathematics* 49.3 (1989), pp. 906–931.

[46] David L Donoho and Xiaoming Huo. "Uncertainty principles and ideal atomic decomposition". In: *IEEE transactions on information theory* 47.7 (2001), pp. 2845–2862.

[47] Michael Elad and Alfred M Bruckstein. "A generalized uncertainty principle and sparse representation in pairs of bases". In: *IEEE Transactions on Information Theory* 48.9 (2002), pp. 2558–2567.

[48] Emmanuel J Candes and Terence Tao. "Near-optimal signal recovery from random projections: Universal encoding strategies?" In: *IEEE transactions on information theory* 52.12 (2006), pp. 5406–5425.

[49] Emmanuel J Candès, Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509.

[50] David L Donoho et al. "Compressed sensing". In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[51] Jarvis Haupt and Robert Nowak. "Signal reconstruction from noisy random projections". In: *IEEE Transactions on Information Theory* 52.9 (2006), pp. 4036–4048.

[52] Simon Foucart and Holger Rauhut. "A mathematical introduction to compressive sensing". In: *Bull. Am. Math* 54 (2017), pp. 151–165.

[53] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[54] Amir Beck and Marc Teboulle. "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems". In: *IEEE transactions on image processing* 18.11 (2009), pp. 2419–2434.

[55] Deanna Needell and Rachel Ward. "Stable image reconstruction using total variation minimization". In: *SIAM Journal on Imaging Sciences* 6.2 (2013), pp. 1035–1058.

[56] Luminita A Vese and Stanley J Osher. "Modeling textures with total variation minimization and oscillating patterns in image processing". In: *Journal of scientific computing* 19.1-3 (2003), pp. 553–572.

[57] Tony F Chan, Jianhong Shen, and Hao-Min Zhou. "Total variation wavelet inpainting". In: *Journal of Mathematical imaging and Vision* 25.1 (2006), pp. 107–125.

[58] Stephen M Robinson. "A short proof of the sticky face lemma". In: *Mathematical Programming* 168.1-2 (2018), pp. 5–9.