A 1D Fluid Model on the Circle, an Algorithm for Simulating Dense Crowds, and Stability for Programs with Seminorm Objective and Linear Constraints

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Samuel J. Stewart

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

Doctor of Philosophy

Prof. Vladimir Sverak

April, 2020

Acknowledgements

A profound thanks to my advisor Professor Vladimir Sverak who taught me to think deeply about simple problems and to keep an open mind. Without his patience and expertise, this thesis would not have been possible. I would also like to thank Professor Gilad Lerman for his knowledge and guidance on the work presented in the final chapter. I am indebted to Professor Fadil Santosa, Professor Arnd Scheel, Professor Daniel Spirn, and Professor Richard McGehee for professional and personal mentorship. Many thanks to my colleagues and friends in the department, especially Alex Guterriez, Nicole Bridgland, Sarah Milstein, Harini Chandramouli, Sunita Chepuri, and Vivekanand Dabade. Thanks to Bonny Flemming for her help and patience. And finally, thanks to my parents and siblings, especially our family's other three scientists.

Dedication

To my father, for reminding me to laugh.

Abstract

In this thesis, we describe three contributions made to three different fields. First, we prove local stability of solutions to a 1D model equation of the 3D Euler equations. Second, we describe a model of human crowds where people are modeled by ellipses. Finally, we prove local stability of solutions for a family of convex programs.

Contents

Acknowledgements **i**

Dedication **ii**

Abstract **iii**

List of Figures **vi**

1 Introduction **1**

2 A 1D Fluid Model **3**

2.1 Introduction 3

2.2 Exponential Decay in the Linear Regime 9

2.2.1 Absolute Continuity of μ 19

2.2.2 Exponential Decay to Equilibria 25

2.3 Stability in the Nonlinear Regime 34

2.4 Numerics 42

2.4.1 Numerical Experiments 45

3 Crowd Dynamics **49**

3.1 Introduction 49

3.2 Background 50
3.3 Model 54
3.4 Main Result 58
3.4.1 Elliptical Particles 66
3.5 Numerics 66

4 MRI **74**
4.1 Introduction 75
4.2 Background 77
4.3 The Main Theorem and its Proof 81
4.3.1 $M(\boldsymbol{A})$ is Calm and Lower Lipschitz 82
4.3.2 $L(\boldsymbol{A}, \mu)$ is Calm 83
4.4 Conclusion 92

Bibliography **94**

List of Figures

2.1 Evolution of slightly perturbed $\sin(\theta)$ as $\sin(\theta) + \frac{1}{4}\sin(5\theta)$. Notice how the slope of the tangent line is preserved in time. 46

2.2 The second derivative blows up as the zeros of ω get close. 47

3.1 Projecting onto the osculating circle is a good approximation for projecting onto the ellipse 70

3.2 Renormalizing sends points to the boundary of the ellipse 71

3.3 To compute the signed distance, we find the closest points on the osculating circles to the center of the opposite osculating circle 72

3.4 Two formations of ellipses trying to push past each other in opposite directions 73

4.1 Demonstration of a case where $S(\boldsymbol{A}_0)$ is a face of the unit ℓ_1 ball and $S(\boldsymbol{A})$ for $\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E}$ is a vertex of this ball. Here, the line intersecting the face represents the solution of $\boldsymbol{A}_0\boldsymbol{x} = \boldsymbol{b}$, the other line represents the solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, and the sets $S(\boldsymbol{A}_0)$ and $S(\boldsymbol{A})$ are the intersections of these lines, respectively, with the unit ℓ_1 ball. 78

Chapter 1

Introduction

Our first contribution is to study finite-time blowup to solutions of the model problem

$$\omega_t + [\omega, v] = 0, \quad v_x = H\omega,\ x \in S^1$$

to the 3D Euler equations. In fact, we find that the original numerical evidence of blowup is actually due to conservation of the first derivative at the zeros of ω. In fact, in fractional Sobolev spaces with $s < 3/2$, the solutions converge exponentially to a two-parameter family of equilibria $A\sin(\theta - \theta_0)$. Our main tools are results from spectral theory for unitary operators and complex ODE theory.

Our second contribution is to extend a model from crowd dynamics. In the original model, people are circles with the constraint that they cannot overlap. At each timestep, every person's preferred velocity is projected to the set of legal velocities that preserve the overlap constraint. Our contribution extend the model from circle to ellipses. In the process, we develop a fast algorithm for computing the distance between two ellipses in addition to simplifying dramatically the original model's well-posedness result and then extending it to ellipses. Our main tools are results about differential inclusions.

Our final contribution is to prove that a family of convex programs of the form

$$\min_{Ax=b} \Omega(x)$$

with A full rank and $\Omega(x)$ a seminorm with piecewise constant gradient, is stable (in a set-valued sense) with respect to perturbations in the matrix A. Our main tool are local error bounds inspired by the classical Hoffman's lemma.

Chapter 2

A 1D Fluid Model

2.1 Introduction

In $\mathbb{R}^3$ the Euler equations are given by

$$v_t + v\nabla v + \nabla p = 0, \quad \text{div } v = 0,$$

with v the velocity and p the pressure. A blowup of Navier-Stokes would be evidence that canonical models might be insufficient, whereas blowup of Euler would be relevant for some of the basic ideas in turbulence theory. One potential approach, which dates at least back to Burgers [1], is to obtain insights by studying singularity formation or regularity for 1D model equations. Here we will continue the work of [2] on a 1D model and prove local stability of equilibria, a partial result in ruling out blowup.

We can motivate our model from the Euler equations in vorticity form with

$$\omega_t + v\nabla\omega - \omega\nabla v = 0, \quad \text{div } v = 0, \quad \omega = \text{curl } v.$$

Thus vorticity is related to the velocity by the system

$$\begin{aligned} \text{div } &= 0 \\ \nabla \times v &= \omega \end{aligned}$$

which can be inverted in $\mathbb{R}^3$ to recover the velocity v via a so-called "Biot-Savart" law

$$v = K * \omega$$

with kernel that scales as $K(\lambda x) = \lambda^{-2} K(x)$. Then

$$\nabla v = \nabla K * \omega$$

is essentially a singular integral operator, with scaling $\nabla K(\lambda x) = \lambda^{-3} \nabla K(x)$.

A good analogy for the situation in 1D (with domain being the circle, for simplicity) is thus the Hilbert transform

$$v_\theta = H\omega = \text{p.v.} \int_{S^1} \omega(y) \cot(x - y) dy.$$

If we rewrite the vorticity form of the Euler equations with the convective derivative

$$\frac{D\omega}{Dt} = \omega \nabla v,$$

then the Constantin-Lax-Majda [3] model is natural

$$\omega_t = \omega v_\theta, \quad v_\theta = H\omega. \tag{2.1.1}$$

There are easy-to-construct blowup solutions if we consider the holomorphic extension

$$f(z) = w + iHw.$$

The equation then becomes

$$f = -\frac{i}{2}f^2,$$

or in coordinates $g = \frac{1}{f}$ the constant field

$$\dot{g} = -\frac{i}{2},$$

yielding the solution

$$g(t, z) = -\frac{i}{2}t + g_0(z).$$

In our original coordinates

$$\omega = \text{Re } f = \frac{\omega_0}{(1 - \frac{t}{2}H\omega_0)^2 + \frac{(\omega_0 t)^2}{4}}$$

blows up if and only if the initial data $f_0 = \omega_0 + iH\omega_0$ intersects the positive imaginary axis.

In 1990, De Gregorio [4] added the transport term $u\nabla\omega$ thinking it might regularize solutions. His model

$$\omega_t + v\omega_\theta - \omega v_\theta = 0, \quad v_\theta = H\omega \tag{2.1.2}$$

is also connected to the 2D Boussinesq system in the half-plane (itself a model of axi-symmetric 3D Euler)

$$\begin{aligned} \omega_t + u\omega_x &= \theta_x \\ \theta_t + u\theta_x &= 0, \end{aligned} \tag{2.1.3}$$

with the same Biot-Savart law $u_x = H\omega$ as for (2.1.1), and an additional variable θ representing temperature. For this system, the authors in [1] gave a nonconstructive proof of blowup. The De Gregorio model is similar to (2.1.1) except for the addition of the transport term $u\omega_x$, and it is the same as (2.1.3) (when $\theta = 0$) except for the $u_x\omega$ term. Although both (2.1.3) and (2.1.1) have known blowup solutions, but well-posedness of the De Gregorio model has been open since 1990 [4].

We have developed numerical and theoretical evidence that suggests solutions to (2.1.2) do not blow up but generically converge to a family of functions

$$\Omega_{A,\theta_0} = A\sin(\theta - \theta_0).$$

That these are equilibria is not difficult to see. We can rewrite the equation with the Lie bracket as

$$\omega_t + [v, \omega] = 0.$$

Now define

$$\Lambda v = -H\partial_\theta$$

then we rewrite our equation as

$$\omega_t + [v, \Lambda v] = 0.$$

The equilibria are now clear since

$$\Lambda \sin\theta = -\sin\theta$$

and the Lie bracket gives zero for identical vectors. More generally, the equilibria are

$$\Omega_{n,A,\theta_0} = A\sin(n(\theta - \theta_0)).$$

Given the periodicity of the problem, we used a numerical method with spectral convergence to gain intuition about the behavior of solutions before we began working on the theory. Numerical simulations have been done before (see for example [2]), but in our calculations we observed some new features, such as convergence to a family of equilibria and new conserved quantities.

The regularity might be due to some underlying geometric structure. Indeed, (2.1.2) has many connections to geometry. Changing the coefficient of $u_x\omega$ to 2 from -1 gives the equation for geodesic flow on $\mathrm{Diff}(S^1)$ with an appropriate right-invariant Riemannian metric. The authors of [5] show that this equation has blowup solutions.

Applying a geometric perspective to our original equation, the addition of the transport term to (2.1.1) allows us to write our equation in terms of the Lie bracket as

$$\omega_t + [v, \omega] = 0.$$

Solutions evolve by the pushforward of the initial vector field under a diffeomorphism. This is similar to 3D Euler. Using a mixture of geometric and analytic techniques, we proved local stability of equilibria [6].

Theorem 2.1.1. *Let* $\Omega_{A^\pm,\theta_0^\pm} = A^\pm \sin(\theta - \theta_0^\pm)$. *Then if* $\omega(0) = \omega_0$ *is* $C^2(S^1)$ *close to* $\Omega_{A^\pm,\theta_0^\pm}$, *then there exists unique solution* $\omega(t)$ *in* C^2 *that exists for all time. Moreover,* ω *converges exponentially to* $\Omega_{A^\pm,\theta_0^\pm}$ *in* $H^s(S^1)$ *for all* $s < \frac{3}{2}$ *as* $t \to \pm\infty$, *respectively.*

For initial data ω that is C^2 close to Ω_{A,θ_0}, the solution will have only two roots. At one, say θ, the derivative is positive, and negative at the other. As we will see, the

derivative at the root θ, which may move, is conserved. Thus, assuming the solution converges to *an* equilibrium, this conservation means the solution must converge to $\Omega_{A,\theta}$ with amplitude $A = -\omega_\theta(\theta)$.

We present our main techniques in this thesis. Our proof is involved, so we outline it first. We prove exponential decay of small perturbations to the equilibria $A\sin(\theta - \theta_0)$ in $H^s(S^1)$ by first analyzing the linearized equation. We will see that the linear evolution is an isometry in the critical space $s = \frac{3}{2}$, so a natural first step is to rule out periodic solutions (that is, rule out eigenvalues in the continuous spectra of the linear operator). We then verify weak convergence to equilibria by showing the spectral measure is absolutely continuous.

After studying the linear regime, we prove exponential decay in the nonlinear regime by first improving our weak decay to exponential decay for solutions to the linearized equation in certain weighted spaces. This combined with an energy bound allows us to use perturbation theory to prove exponential decay in the nonlinear regime.

Following publication of our result, Lei et al. [7] proved global well-posedness for a class of initial data using an entirely different technique. For $\omega_0 \geq 0$, they found that $\|\omega\|_{H^1}$ is conserved. This conservation enabled them to obtain the following result.

Theorem 2.1.2 (Theorem 1.1, [7])**.** *Let initial data $\omega_0(\theta)$ be nonnegative with compact support on S^1. Let $k \geq 1$. Furthermore assume that $\sqrt{\omega_0} \in H^k(S^1)$. Choose coordinates so that the velocity field v satisfies*

$$\int_{S^1} v d\theta = 0.$$

Then Eq 2.1.2 is globally well-posed in $C([0,T]; H^k(S^1))$ and

$$\left\| \sqrt{\omega(t,\cdot)} \right\|_{H^k(S^1)}$$

is conserved for all time $t \in [0, T]$.

2.2 Exponential Decay in the Linear Regime

Our goal for this section is to verify weak stability of solutions to Eq 2.1.2 linearized at equilibrium $\sin\theta$ in a critical space, and then verify exponential decay in a family of subcritical spaces. The linearization of Eq 2.1.2 is given by

$$\eta_t = -[\sin\theta,\, v + \eta], \quad v_\theta = H\eta, \quad \int_{S^1} v\, d\theta = 0. \tag{2.2.1}$$

The general approach is to find a space where the evolution is unitary and use the spectral theorem to diagonalize the flow. Then we prove that the spectral measure is absolutely continuous and thus by the Riemann-Lebesgue lemma we see that the solutions converge weakly to zero. The right family of spaces are a kind of fractional Sobolev spaces on S^1.

Definition 2.2.1. We define the fractional Sobolev space

$$H^s(S^1) := \{f(\theta) \in L^2(S^1) \mid \sum_{k=-\infty}^{\infty} |k|^{2s} |f_k|^2 < \infty\}$$

where f_k denotes the kth Fourier coordinate of f.

Note that this space is a Hilbert space with inner product

$$\langle f,\, g\rangle = \sum_{k=-\infty}^{\infty} |k|^{2s} f_k \overline{g}_k$$

We denote

$$\dot{H}^s(S^1) := \{f \in H^s(S^1) \mid f_0 = 0\}$$

We want to find the critical s so that the evolution is unitary in $H^s(S^1)$ but converges weakly to equilibria. The simple model captures the process of choosing such a space

$$f_t = -\sin\theta f_\theta, \tag{2.2.2}$$

Define $L = -\sin\theta\partial_\theta$. One approach is to show that the flow is unitary is to choose s so that L is skew-adjoint in $\dot{H}^s(S^1)$. To see the equivalence between a skew-adjoint L and a unitary flow, note that

$$\frac{d}{dt}\langle \eta,\, \eta\rangle = \langle \dot{\eta},\, \eta\rangle + \langle \eta,\, \dot{\eta}\rangle = 0.$$

implying that

$$\langle L\eta,\, \eta\rangle = -\langle \eta,\, L\eta\rangle$$

In the Fourier basis $e_k = e^{ik\theta}$ we have

$$Le_k = A_k e_{k-1} + B_k e_{k+1}, \quad k \in \mathbb{Z}$$

with

$$A_k = \frac{k}{2}, \quad B_k = -\frac{k}{2},$$

so that in coordinates

$$(Lf)_k = B_{k-1} f_{k-1} + A_{k+1} f_{k+1}.$$

One can check that L is skew adjoint in the space $\dot{H}^{\frac{1}{2}}(S^1)$.

There is another way to see that the flow is unitary. Consider an extension of the solution to the disk as follows. First, recall that if ϕ^t is given by the flow $\dot{\phi} = \sin(\theta)$ on

S^1 then for initial data $f_0 \in \dot{H}^{\frac{1}{2}}(S^1)$ the solution to Eq 2.2.2 is

$$f(\theta, t) = f_0((\phi^t)^{-1}(\theta)).$$

To extend this solution to the disk, we extend ϕ^t holomorphically to a function $\widetilde{\phi}^t : D \to \mathbb{R}$ and extend f_0 harmonically to a function $\widetilde{f_0} : D \to \mathbb{R}$. Call the extension of f to the disk

$$\widetilde{f}(z, t) = \widetilde{f_0}((\widetilde{\phi}^t)^{-1}(z)).$$

One can check that

$$\|f\|^2_{\dot{H}^{\frac{1}{2}}(S^1)} = \left\|\nabla \widetilde{f}\right\|_{L^2(D)}.$$

Calculation shows that the Dirichlet energy on the right-hand side is invariant under conformal automorphisms implying that the flow is unitary in $\dot{H}^{\frac{1}{2}}(S^1)$.

The flow induces several decompositions of $\dot{H}^{\frac{1}{2}}(S^1)$. For example, the flow leaves invariant the decomposition

$$\dot{H}^{\frac{1}{2}}(S^1) = \{f \mid f(\theta) = 0, \theta \in [0, \pi)\} \oplus \{f \mid f(\theta) = 0, \theta \in [\pi, 2\pi)\}.$$

A more convenient splitting is

$$\dot{H}^{\frac{1}{2}}(S^1) = \{f \mid f_k = 0, k \geq 1\} \oplus \{f \mid f_k = 0, k \leq -1\}.$$

Extending to the disk, this splitting corresponds to holomorphic and anti-holomorphic functions. Indeed, one can check that $[L, H] = 0$ in the space of harmonic extensions to the disk. Since the eigenspaces of the Hilbert transform are holomorphic and antiholomorphic functions, then L respects this splitting. We can thus consider without loss of generality the space of holomorphic functions on the disk with boundary data in

$\dot{H}^{\frac{1}{2}}(S^1)$. Denote this space $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$.

The evolution for holomorphic functions is given by

$$\begin{aligned}
\dot{f}_1 &= A_2 f_2 \\
\dot{f}_2 &= B_1 f_1 + A_3 f_3 \\
\dot{f}_3 &= B_2 f_2 + A_4 f_4 \\
\dot{f}_4 &= \dots
\end{aligned}$$

Note that this system is closed and the variable f_1 can be calculated by integration of $A_2 f_2$ once the components $f_2, f_3, \dots$ are known.

A final insight we can obtain from this model problem is the following lemma, which will be useful later.

Lemma 2.2.2. *Let ϑ be a measure with compact support in $\mathbb{R}\backslash\{0\}$. Then the function*

$$f(z) = \int_{-\infty}^{\infty} \left(\frac{1-z}{1+z}\right)^{is} d\vartheta \tag{2.2.3}$$

is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ if and only if ϑ is absolutely continuous, with square integrable density.

Proof. We begin by motivating a certain change of coordinates. Note that $1/2(1-z^2)\partial_z$ is a holomorphic extension of $L = -\sin\theta\partial_\theta$ to the disk. Under the conformal change of coordinates

$$w = \log\left(\frac{1-z}{1+z}\right)$$

from the disk to the strip $\mathcal{O} = \{\frac{-\pi}{2} < \text{Im } z < \frac{\pi}{2}\}$, our vector field $1/2(1-z^2)\partial_z$ pushes forward to ∂_w so that our equation becomes

$$f_t + f_w = 0 \tag{2.2.4}$$

on the strip in the space $\dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ of holomorphic functions on the strip with norm

$$\int_{\mathcal{O}} \left|f'\right|^2 \frac{i}{2} dw \wedge d\overline{w}.$$

Note that $\dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ is unitarily equivalent to $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$. The evolution of Eq 2.2.4 is $f(w-t)$ and is thus diagonalized by the Fourier representation

$$f(w) = \int_{-\infty}^{\infty} \phi(s) e^{isw} ds.$$

Roughly speaking, this is the spectral decomposition of $\dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ by the holomorphic extension of our operator L (after a change of coordinates). Changing coordinates back to the disk results in a spectral decomposition of $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ by eigenfunctions

$$\left(\frac{1-z}{1+z}\right)^{is}$$

of $1/2(1-z^2)\partial_z$, the extension of L to the disk.

To prove the lemma, consider the case when $\phi(s)$ is compactly supported. To verify that $f(z)$ is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ we need only verify that $f(w)$ is in $\dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ since as noted earlier they are unitarily equivalent. We now calculate the norm of $\dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ given by

$$\int_{\mathcal{O}} \left|f'\right|^2 \frac{i}{2} dw \wedge d\overline{w}.$$

Since $\phi(s)$ is compactly supported, we can differentiate under the integral

$$f'(w) = \int_{-\infty}^{\infty} is\phi(s) e^{iws} ds.$$

The compactness of $\phi(s)$ also allows us to take the Fourier transform in w_1. Applying

Parseval's theorem gives

$$\int_{-\infty}^{\infty} \left|f'(w_1 + iw_2)\right|^2 dw_1 = 2\pi \int_{-\infty}^{\infty} s^2 e^{-2w_2\xi} \left|\phi(s)\right|^2 ds$$

Fubini's Theorem allows us to switch the integrals

$$\int_{\mathcal{O}} \left|f'\right|^2 i dw \wedge d\overline{w} = 2\pi \int_{-\infty}^{\infty} \int_{-\pi/2}^{\pi/2} s^2 e^{-2w_2 s} \left|\phi(s)\right|^2 dw_2 ds = 2\pi \int_{-\infty}^{\infty} \xi^2 \sinh(\pi\xi) \left|\hat{f}\right|^2 d\xi.$$

This is finite since $\phi(s)$ is compactly supported.

Now we consider the other direction. Assume that $f(z) \in \dot{\mathcal{H}}^{\frac{1}{2}}(D)$. Then changing coordinates to the strip as above means that $f(w) \in \dot{\mathcal{H}}^{\frac{1}{2}}(\mathcal{O})$ so $f'(w)$ is in $L^2(\mathcal{O})$. Since f is holomorphic in $\mathcal{O}$ this means f' restricted to the real line is square integrable. Thus we have the Fourier representation

$$f'(w_1) = \int_{-\infty}^{\infty} \varphi(s) e^{iw_1 s} ds$$

for some $\varphi \in L^2(\mathbb{R})$. One the other hand, by the compactness of ϑ we can differentiate under the integral and get

$$f'(w_1) = \int_{-\infty}^{\infty} is e^{iw_1 s} ds$$

and so by uniqueness of the Fourier representation $\varphi(s)ds = d\vartheta$. □

Using the insights from this model, we now consider again the linearized evolution to Eq 2.1.2 near the equilibrium $\sin\theta$ given by

$$L = [-\sin\theta, \eta + v], \quad v_x = H\eta, \quad \int_{S^1} v d\theta = 0$$

As in the model problem, L acts in Fourier coordinates as

$$Le_n = A_n e_{n+1} + B_n e_{n-1}$$

with

$$A_n = \frac{1-n}{2}\left(1 - \frac{1}{|n|}\right), \quad B_n = \frac{n+1}{2}\left(1 - \frac{1}{|n|}\right). \tag{2.2.5}$$

which gives the recurrence relation Writing the inner product in Fourier coordinates $\eta_2, \eta_3, \ldots$ gives

$$\langle \eta, \eta \rangle = \sum_{k \geq 2} c_k \eta_k \overline{\eta}_k$$

with $(c_k) \subset (0, \infty)$. Then L acts in these coordinates as

$$(L\eta)_k = A_{k-1}\eta_{k-1} + B_{k+1}\eta_{k+1}. \tag{2.2.6}$$

In a real Hilbert space, if $\langle L\eta, \eta \rangle = 0$ then L is skew-adjoint so we want to find $c_k \in \mathbb{R}$ so that

$$\sum_{k \geq 2} c_k \left(A_{k-1}\eta_{k-1} + B_{k+1}\eta_{k+1}\right) \overline{\eta}_k = 0$$

Setting the real and imaginary parts of the sum to zero gives the condition

$$c_{k+1} = \frac{B_{k+1}}{-A_k} c_k$$

Then solving the recurrence relation

$$c_{k+1} = 6B_{k+1}P(k)c_2$$

explicitly, where we define the fraction

$$P(k) = \prod_{3 \le n \le k} \frac{-B_n}{A_n} = (-1)^{k-3} \frac{(k+1)!}{6(k-1)!}$$

gives

$$c_{k+1} = (k-1)^2(k+1) \sim k^3. \tag{2.2.7}$$

Thus L is skew adjoint in the space with inner product

$$\langle \eta, \eta \rangle = \sum_{k=2}^{\infty} c_k \eta \overline{\eta}_k$$

Such a space is equivalent to $\dot{H}^{\frac{3}{2}}(S^1)$. In fact, as in the model problem, one can consider, without loss of generality, the space $\dot{\mathcal{H}}^{\frac{3}{2}}(D)$ because L again commmutes with H. This is easiest to verify in Fourier coordinates $e_k = e^{ik\theta}$. The Biot-Savart law gives that

$$Le_1 = [\sin\theta, e_1 - e_1] = 0$$

and likewise for e_{-1}. This and the fact that L shifts indices by at most two (see Eq 2.2.6), shows that L leaves invariant holomorphic and antiholomorphic functions.

We can apply a standard formulation of the spectral theorem to find a unitarily equivalent space where our flow is diagonalized. See [8] for more details.

Definition 2.2.3 (Cyclic Vector)**.** Let $L : X \to X$ be a linear operator on Hilbert space X. Then v is a cyclic vector if the span of

$$\{L^n v \mid n \ge 0\}$$

is dense in H.

Theorem 2.2.4. *Let $L : X \to X$ be a skew-adjoint, bounded operator with cyclic vector v on Hilbert space X. Then there exists isometry $U : L^2(\mathbb{R}, \mu) \to X$ such that*

$$U\,L\,U^{-1}\,f = is\,f(s)$$

where μ is a Radon measure.

Lemma 2.2.5. *The operator L has cyclic vector e_2 in $X = \dot{\mathcal{H}}^{\frac{3}{2}}(D)/\langle e_1 \rangle$.*

Proof. Define

$$V_n = \langle L^n e_1, L^{n-1} e_1, \ldots, e_1 \rangle$$

so our goal is to prove that

$$\overline{\bigcup_{n=1} V_n} = H.$$

it suffices to prove that

$$V_n = \langle e_1, \ldots, e_{n+1} \rangle$$

since the basis $(\, e_n \,)$ is dense in X.

The base case $n = 0$ is trivial. Assume that

$$\langle e_1, \ldots, e_n \rangle = V_{n-1}.$$

Then we can write

$$e_n = \alpha_{n-1} L^{n-1} e_1 + \alpha_{n-2} L^{n-2} e_1 + \cdots + \alpha_1 e_1$$

so that

$$Le_n = \alpha_{n-1} L^n e_1 + \alpha_{n-2} L^{n-1} e_1 + \cdots + \alpha_1 L e_1 \in V_n.$$

By definition of L we also have that

$$Le_n = a_{n-1}e_{n-1} - a_n e_{n+1}$$

Since $e_{n-1} \in V_n$ by the inductive hypothesis, then $e_{n+1} \in V_n$. □

Since the spectrum of L is the imaginary axis, the spectral theorem guarantees that we have a Radon measure μ and isometry $U : L^2(\mathbb{R}, \mu) \to X$ so that

$$U^{-1}LUf(s) = isf(s), \quad \mu \text{ almost everywhere}$$

for all $f \in L^2(\mathbb{R}, \mu)$. The flow in $L^2(\mathbb{R}, \mu)$ is thus

$$\eta(s, t) = e^{i\,s\,t}\eta_0(s).$$

Suppose that μ is absolutely continuous so that $\mu = g(s)ds$ for some measurable density g. Recall that we wish to show that $\eta(z, t)$ converges weakly to zero in X. Since $U : L^2(\mathbb{R}, \mu) \to X$ is an isometry, it suffices to check that for every compactly supported ϕ we have

$$\lim_{t\to\infty} \langle \phi,\, \eta(t) \rangle_{L^2(\mathbb{R},\mu)} = 0$$

By definition

$$\langle \phi,\, \eta(t) \rangle_{L^2(\mathbb{R},\mu)} = \int_{\mathbb{R}} \phi(s)\eta(t, s)d\mu(s) = \int_{\mathbb{R}} (\phi(s)\eta_0(s)g(s))\, e^{is\,t}ds = \int_{\mathbb{R}} F(s)e^{is\,t}ds$$

where

$$F(s) = \phi(s)\eta_0(s)g(s) \in L^1(\mathbb{R})$$

The result now follows from the classical Riemann-Lebesgue lemma applied to F.

2.2.1 Absolute Continuity of μ

A first step to showing μ is absolutely continuous is to rule out Dirac masses. Such point masses correspond to eigenfunctions of L in X. We will use complex ODE theory to show that L has no eigenfunctions in X.

Let's derive the eigenfunction equation for holomorphic functions of the disk. The operator L extends to a vector field on the disk as

$$L = [(z^2 - 1)/2\partial_z, (\eta + v)\partial_z].$$

For holomorphic functions, the Hilbert transform simplifies and becomes a local operator (i.e. derivative with respect to z) so that inverting gives

$$\eta(z) = -z\partial_z v.$$

Our final eigenfunction equation is for eigenvalue λ is

$$2\lambda z\eta + (z^2 - 1)(\eta + v)' - 2z^2(\eta + v) = 0, \quad \eta = -zv'. \tag{2.2.8}$$

An equivalent definition in terms of a recurrence on the coefficients η_k is

$$\begin{aligned} \lambda\eta_1 &= A_2\eta_2 \\ \lambda\eta_2 &= A_3\eta_3 \\ \lambda\eta_3 &= B_2\eta_2 + A_4\eta_4 \\ &\cdots \end{aligned} \tag{2.2.9}$$

with A_k, B_k defined by Eq 2.2.5.

We know that L has no kernel in X, so we can assume that $\lambda = is \neq 0$. By

assumption we know that $\eta_0 = v_0 = 0$ (no constant term) so we can write

$$v = zF, \quad \eta = -z(zF)'$$

for $F(z)$ that satisfies

$$z(z^2-1)F'' + (z^2 + 2\lambda z - 3)F' + 2\lambda F = 0.$$

This is a Heun equation [9] with singularities $z = 0, -1, 0, \infty$. We will use the fact that the symmetries of Heun equations include Mobius transformations.

Theorem 2.2.6. *The operator L has no eigenfunctions in X.*

Proof. Standard theory [9] gives local formula for F near the singularities. The eigenfunctions must be holomorphic in the disk so we ignore F that have a singularity at $z = 0$. Solutions with singularities at infinity take the form

$$F(\zeta, is) = AU(\zeta, is) + B[U(\zeta, is)\, \log \zeta + V(\zeta, is)]$$

with $U(\zeta, is), V(\zeta, is)$ holomorphic in ζ, A, B complex constants, and $U(0) = 1$. For these solutions to be holomorphic (even continuous), we must have $B = 0$ since $\log \zeta$ is not continuous across its branch cut. But $F(0) = A\,U(0) = A$, then F is both holomorphic and bounded at infinity. By Liouville's theorem, F must constant, and thus identically zero in X.

Now consider solutions with singularities at $z = 1$. Then the solution can be written near $z = 1$ as

$$F(z, is) = A(1-z)^{2-is}U(z, is) + BV(z, is)$$

with $U(\cdot, is), V(\cdot, is)$ holomorphic and A, B complex constants. When $A \neq 0$, then

$\eta'(z) \notin \dot{H}^{\frac{1}{2}}(S^1)$ so $\eta \notin \dot{\mathcal{H}}^{\frac{3}{2}}(D)$. On the ohter hand, if $A = 0$, then since F is bounded at infinity by assumption, again by Liouville's theorem is constant. An identical argument holds for $z = -1$ with blowup on the order $(1+z)^{2+is}$. □

Having shown there are no eigenfunctions in X and thus no Dirac masses in μ, we now prove (by contradiction) that μ is in fact absolutely continuous. We will construct a function $f(s) \in L^2(\mathbb{R}, \mu)$ such that $\eta(z) = U^{-1} f(s)$ is not in X, contradicting the fact that U is an isometry.

Theorem 2.2.7. *The spectral measure μ is absolutely continuous.*

Proof. We will construct a $\Psi(z, s)$ satisfying

$$L\Psi = is\Psi, \quad \text{for } \mu \text{ almost every } s$$

and $\Psi(z, s) \in \dot{\mathcal{H}}^{\sigma}(D)$ where $\sigma < 1$. Then we will construct a compactly supported $\phi(s)$ so that the function

$$g(z) = \int_{\mathbb{R}} \phi(s)\Psi(z, s)d\mu(s)$$

is not in X, contradicting the fact that U is an isometry.

Represent $\eta = Uf$ by Fourier coordinates $\eta_2, \eta_3, \ldots$. For each $k \geq$ the map $f \mapsto \eta_k(f)$ is a continuous functional on $L^2(\mathbb{R}, \mu)$ and thus has a representation

$$\eta_k(f) = \int_{\mathbb{R}} f(s)G_k(s)d\mu(s)$$

with

$$G_k \in L^2(\mathbb{R}, \mu), \quad \int_{\mathbb{R}} G_k(s)\overline{G}_l(s)d\mu \sim \frac{\delta_{kl}}{c_k} \tag{2.2.10}$$

with c_k defined in Eq 2.2.7. Since

$$\eta_k = \frac{1}{c_k}\langle \eta,\, e_k\rangle_X = \frac{1}{c_k}\langle U^{-1}\eta,\, U^{-1}e_k\rangle_{L^2(\mathbb{R},\mu)} = \frac{1}{c_k}\int_{-\infty}^{\infty} f(s)\overline{U^{-1}e_k(s)}d\mu(s),$$

then

$$G_k = \frac{1}{c_k}\overline{U^{-1}e_k}.$$

Define $G := (G_2, G_3, \ldots)$ to be an element of the linear space of infinite sequences of $L^2(\mathbb{R},\mu)$ functions, we can write

$$\eta = \int_{\mathbb{R}} f(s)G(s)d\mu(s), \quad L\eta = \int_{\mathbb{R}} f(s)LG(s)d\mu(s)$$

with the integrals defined component by component. On the other hand, by the definition of U we have

$$L\eta = \int_{\mathbb{R}} isf(s)G(s)d\mu(s).$$

Since $f \in L^2(\mathbb{R},\mu)$ was arbitrary, we see

$$LG(s) = isG(s), \quad \text{for } \mu \text{ almost every } s. \tag{2.2.11}$$

Now define

$$G_1(s) = \frac{1}{is}A_2G_2(s), \quad \Psi(z,s) = G_1(s)z + G_2(s)z^2 + \cdots,$$

where G_1 comes from the recurrence relation for the eigenfunctions in Eq 2.2.9. Using the orthogonality relations of G_k in Eq 2.2.10 we see that $z \mapsto \Psi(z,s)$ is well-defined in the sense that it is in $\dot{\mathcal{H}}^\sigma(D)$ for $\sigma < 1$ (not necessarily optimal). By Eq 2.2.11 we see

that

$$L\Psi(z,s) = is\Psi(z,s) \quad \text{for } \mu \text{ almost every } s$$

so Ψ satisfies Eq 2.2.8. Let $\Phi(z,is)$ be a solution of the same equation with normalization $\eta_2 = 1$. From earlier discussion, near $z = 1$ we have

$$\Phi(z,\lambda) = A(\lambda,1)(1-z)^{1-\lambda} + W(z,1,\lambda)$$

where A is analytic in λ and W is analytic in $z \in D$ and λ and $W(\cdot,1,is) \in \dot{\mathcal{H}}^{\frac{3}{2}}(D)$ for $\lambda = is \neq 0$.

Then by uniqueness of solutions to Eq 2.2.8 we have

$$\Psi(z,s) = G_2(s)\Phi(z,is), \quad \text{for almost every } s.$$

Thus

$$\Psi(z,s) = G_2(s)A(is)(1-z)^{1-is} + G_2(s)W(z,is), \quad \text{for almost every } s$$

By Lemma 2.2.2 we know for compactly supported measure ϑ the function

$$h(z) = \int_{\mathbb{R}} \left(\frac{1-z}{1+z}\right)^{-is} d\vartheta(s)$$

is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ if and only if $d\vartheta(s)$ is absolutely continuous with square integrable density. Expanding

$$(1+z)^{is} = 2^{is} + (1-z)H(z,s)$$

with H holomorphic in z, we can write

$$h(z) = \int_{\mathbb{R}} 2^{is}(1-z)^{-is} d\vartheta(s) + \int_{\mathbb{R}} (1-z)^{1-is} H(z,s) d\vartheta(s)$$

The second term is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ because we can bound $(1-z)^{1-is}$ uniformly in s. We can conclude that the first term,

$$h_1(z) = \int_{\mathbb{R}} 2^{is}(1-z)^{-is} d\vartheta(s)$$

is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ if and only if $h(z)$ is in $\dot{\mathcal{H}}^{\frac{1}{2}}(D)$ and thus $h_1 \in \dot{\mathcal{H}}^{\frac{1}{2}}(D)$ if and only ϑ is absolutely continuous with square integrable density.

Since by assumption μ is *not* absolutely continuous, choose set E so that $\mu(E) > 0$ but has Lebesgue measure zero. Then restrict to a subset of E such that $|G_2(s)|, |A(is)| > 0$. Such a subset exists because $A(is)$ is analytic (thus vanishes at most at countably many points and we already proved that μ does not have any Dirac masses), and G_2 is the image of e_2, the cyclic vector of X (thus cannot vanish on any set of positive measure). Now define

$$f(s) := \frac{2^{is}}{G_2(s)A(is)}\ 1_E(s).$$

With this choice of $f(s)$, we define $d\vartheta(s) := f(s)d\mu(s)$ a compactly supported measure that is *not* absolutely continuous. Thus $h_1 \notin \dot{\mathcal{H}}^{\frac{1}{2}}(D)$. If we define

$$g(z) = \int_{\mathbb{R}} f(s)\Psi(z,s)d\mu(s),$$

then

$$g'(z) = \int_{\mathbb{R}} f(s)\Psi'(z,s)d\mu(s) \sim \int_{\mathbb{R}} f(s)G_2(s)A(is)(1-z)^{-is}d\mu(s) = h_1(z) \notin \dot{\mathcal{H}}^{\frac{1}{2}}(D).$$

Thus $g \notin \dot{\mathcal{H}}^{\frac{3}{2}}(D)$ contradicting the fact that U is an isometry. □

2.2.2 Exponential Decay to Equilibria

We have established weak convergence to equilibria in the critical space $\dot{\mathcal{H}}^{\frac{3}{2}}(D)$. Now we show that in fact we have exponetial decay in the weighted L^2 space defined as

$$Y := \{f \in L^2(S^1) \mid \int_{-\pi}^{\pi} |f|^2 \, |\sin(\theta/2)|^{-2\gamma} \, d\theta\}$$

where $\gamma \in (3/2, 2)$. We equip Y with norm $\|f\|_Y = \left\|\sin^{-\gamma}(\theta) f(\theta)\right\|_{L^2}$.

If $f \in C^1(S^1) \cap Y$, then near $\theta = 0$, the function f behaves like θ^γ and in particular f vanishes with order $\gamma \in (3/2, 2)$ at 0. Likewise, the derivative f' vanishes with order $\gamma - 1$ at 0. Thus the space Y allows us to control rate of decay of f at $\theta = 0$. Note that exponential decay in this norm means exponential decay of the $k \geq 2$ Fourier modes.

As before, we can write the linear equation on the disk as

$$L = \frac{1}{2}(z^2 - 1)\eta' - z\eta - \frac{1}{2}(z + \frac{1}{z})v = 0, \quad \eta = -zv', \quad v(0,t) = 0.$$

Our goal is to prove that

$$\left\|e^{tL}\eta_0\right\|_Y \leq Ce^{-\beta t} \|\eta_0\|_Y$$

for $\beta < \beta_0 = \gamma - 3/2$ with C depending on β.

The operator L can be split as

$$L = L_0 + K$$

where we define

$$L_0 := \frac{1}{2}(z^2 - 1)\eta' - z\eta,$$

and K the operator

$$K = -\frac{1}{2}(z + \frac{1}{z})v, \quad \eta = -zv', \quad v(0,t) = 0.$$

We can show that $\left\|e^{tL_0}\eta_0\right\|_Y$ decays exponentially and that K is a compact oeprator on the $\eta \in Y$ such that $\eta = -zv'$. By general theory this will allow us to prove that $\left\|e^{tL}\eta\right\|_Y$ decays exponentially if L has no eigenvalues in the region Re $\lambda > -\beta_0$.

Note that we write L_0, K in complex coordinates for convenience. The definitions of L_0, K do not depend on η being holomorphic. On the circle, $L_0\xi = \sin\theta\xi_\theta - \cos\theta\xi$ via the change of coordinates $\eta = z\xi$ and

$$K\eta = (\cos\theta)v, \quad v_\theta = H\eta, \quad \int_{-\pi}^{\pi} v(\theta)d\theta = 0.$$

Lemma 2.2.8. *Let $\beta_0 = \gamma - 3/2$. Then*

$$\left\|e^{tL_0}\eta_0\right\|_Y \leq e^{-\beta_0 t}\left\|\eta_0\right\|_Y$$

for $t \geq 0$.

Proof. The first order PDE

$$\eta_t = L_0\eta$$

can be solved via the method of characteristics by solving

$$z'(t) = \frac{1}{2}(z^2 - 1).$$

This ODE has the flow map

$$\phi_t(x) = \frac{z - \tau}{1 - \tau z}, \quad \tau = \tanh\frac{t}{2}$$

the solution by definition is the pushforward of the initial vector field η_0

$$\eta(z,t) = \phi_t'(\phi_t^{-1}(z))\eta_0(\phi_t^{-1}(z)) \tag{2.2.12}$$

We can expand the left-hand using Eq 2.2.12

$$\|\eta\|_Y = \int_{S^1} \left|\eta_0(\phi_t^{-1}(z))\right|^2 \left|\phi_t'(\phi_t^{-1}(z))\right|^2 |z-1|^{2\gamma}\, d\mathcal{H}(z)$$

where $d\mathcal{H}(z)$ is the 1d Hausdorff measure on the circle. Changing coordinates to $w = \phi_t^{-1}(z)$ (essentially a u-substitution) means the measure transforms as

$$d\mathcal{H}(z) = \left|\phi_t'(w)\right| d\mathcal{H}(w).$$

The weight term in the product becomes

$$|z-1|^{2\gamma} = |\phi(w)-1|^{2\gamma} = \left|\frac{w-\gamma-1+\tau w}{1-\tau w}\right|^{-2\gamma} = \left|\frac{(w-1)(\tau+1)}{1-\tau w}\right|^{-2\gamma}.$$

Our norm in these new coordinates is thus

$$\|\eta\|_Y = \int_{S^1} |\eta_0(w)|^2 \left|\phi_t'(w)\right| \left|\frac{(w-1)(\tau+1)}{1-\tau w}\right|^{-2\gamma} d\mathcal{H}(w).$$

We now will show that

$$\left|\phi_t'(w)\right| \left|\frac{(w-1)(1+\tau)}{1-\tau w}\right|^{-2\gamma} \leq |w-1|^{-2\gamma} (1-\tau)^{2\gamma-3}.$$

Expanding $\phi_t'(w)$ gives

$$\phi_t'(w) = \frac{(1-\tau)(1+\tau)}{(1-\tau w)^2}.$$

Simplifying we now want to prove

$$\frac{(w-1)^{-2\gamma}(\tau+1)^{1-2\gamma}(1-\tau)}{(1-\tau w)^{2-2\gamma}} \leq |w-1|^{-2\gamma}(1-\tau)^{2\gamma-3}.$$

We proceed by bounding above each of the terms in the product. Since $\gamma \in (\frac{3}{2}, 2)$, then the term

$$(\tau+1)^{1-2\gamma} < 1$$

so we can drop it from the product. Note that

$$|1-\tau| \leq |1-\tau w|$$

when w is on the unit circle and $\tau \in [0,1)$. So we can bound the term

$$\frac{1-\tau}{(1-\tau w)^{2-2\gamma}} \leq (1-\tau)^{2\gamma-3}.$$

Thus we have shown that

$$\frac{(w-1)^{-2\gamma}(\tau+1)^{1-2\gamma}(1-\tau)}{(1-\tau w)^{2-2\gamma}} \leq |w-1|^{-2\gamma}(1-\tau)^{2\gamma-3}.$$

This in turn shows that

$$\|\eta\|_Y^2 \leq (1-\tau)^{2\gamma-3}\int_{S^1} |\eta_0(w)|^2 |w-1|^{-2\gamma}\, d\mathcal{H}(w) \leq e^{-2\beta_0 t}\|\eta_0\|_Y^2$$

where $\beta_0 = \gamma - \frac{3}{2}$.

□

Having proven that the decay is exponential for the flow with operator L_0, we want to prove the decay is exponential for the flow with compact perturbation of L_0, that is

$L = L_0 + K$. We first show that K maps $Y \oplus \langle 1, \sin\theta \rangle$ back to itself and is compact and continuous as well. Since K fixes $\langle 1, \sin\theta \rangle$, as one can see by computing in complex coordinates

$$K \cdot 1 = 0, \quad K \cdot z = -\frac{1}{2}(z^2 + 1),$$

then it is enough to show $K : Y \to Y \oplus \langle 1, \sin\theta \rangle$ is compact and continuous.

Since $Kf = \cos(\theta) f$ is continuous, it suffices to verify that H is continuous and that the inverse of $v_\theta = H\eta$ given by the integral operator

$$(Tv)(\theta) = \int_0^\theta v(t)dt$$

is compact.

Showing H is continuous on $Y \oplus \langle 1, \sin\theta \rangle$ has the added benefit of allowing us to exploit the fact that $[L, H] = 0$ and later restrict our attention to holomorphic functions (as we did in the analysis of L in the critical space $\dot{H}^{\frac{3}{2}}(S^1)$) without loss of generality.

Lemma 2.2.9. *The Hilbert transform $H : Y \oplus \langle 1, \sin\theta \rangle \to Y \oplus \langle 1, \sin\theta \rangle$ is continuous.*

Proof. The proof is similar to the proof that the Hilbert transform is continuous on $L^2(\mathbb{R})$. The Hilbert transform leaves the subspace $\{\alpha_1 \cos\theta + \alpha_2 \sin\theta\}$ invariant so the Hilbert transform is automatically continuous on this subspace. We can restrict our attention to the space Y and show that $P_Y H : Y \to Y$ is continuous where $P_Y : Y \oplus \langle 1, \sin\theta \rangle \to Y$ is the projection onto Y from the augmented space $Y \oplus \langle 1, \sin\theta \rangle$. When $f \in C^1$, it is given by

$$P_Y f = f - f(0) - f'(0)\sin\theta$$

The Hilbert transform on the circle is

$$(Hf)(\theta) = \frac{1}{2\pi}\int_S^1 f(v)\cot\left(\frac{\theta - v}{2}\right)dv.$$

We want to compute $P_Y Hf$. Using the definition of the Hilbert transform and the projection P_Y we have

$$(Hf)(0) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} f(v)\cot\left(\frac{v}{2}\right)dv, \quad (Hf)'(0) = -\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sin\theta}{2\sin^2(\frac{v}{2})}dv$$

then

$$(P_Y Hf)(\theta) = \frac{1}{2\pi}f(v)\left(\cot\left(\frac{\theta - v}{2}\right) - \cot\left(\frac{v}{2}\right) - \frac{\sin\theta}{2\sin^2(\frac{v}{2})}\right)dv.$$

The integrand can be simplified as

$$\cot\left(\frac{\theta - v}{2}\right) - \cot\left(\frac{v}{2}\right) - \frac{\sin\theta}{2\sin^2(\frac{v}{2})} = \frac{\sin^2(\frac{\theta}{2})}{\sin^2(\frac{v}{2})}\cot\left(\frac{\theta - v}{2}\right).$$

We know we can write $f \in Y$ as $f = \left|\sin(\frac{\theta}{2})\right|^\gamma g(\theta)$ where $g \in L^2(S^1)$. To verify that $P_Y H$ is continuous from Y to Y, by definition of the weighted space Y, we need only show that for any $g \in L^2(S^1)$ the quantity

$$\begin{aligned}\left|\sin\left(\frac{\theta}{2}\right)\right|^\gamma (P_Y Hf)(\theta) &= \left|\sin(\frac{\theta}{2})\right|^\gamma \frac{1}{2\pi}\int_{-\pi}^{\pi} f(v)\left(\frac{\sin^2(\frac{\theta}{2})}{\sin^2(\frac{v}{2})}\cot\left(\frac{\theta - v}{2}\right)\right)dv \\ &= \left|\sin(\frac{\theta}{2})\right|^\gamma \frac{1}{2\pi}\int_{-\pi}^{\pi} g(v)\left(\frac{\sin^2(\frac{\theta}{2})}{\sin^{2-\gamma}(\frac{v}{2})}\cot\left(\frac{\theta - v}{2}\right)\right)dv \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} g(v)\left(\frac{\sin^{2-\gamma}(\frac{\theta}{2})}{\sin^{2-\gamma}(\frac{v}{2})}\cot\left(\frac{\theta - v}{2}\right)\right)dv\end{aligned}$$

is bounded in $L^2(S^1)$ with respect to $\|g\|_{L^2}$. In other words, that the operator

$$g(\theta) \mapsto \frac{1}{2\pi} \int_{-\pi}^{\pi} g(v) \left(\frac{\sin^{2-\gamma}(\frac{\theta}{2})}{\sin^{2-\gamma}(\frac{v}{2})} \cot\left(\frac{\theta - v}{2} \right) \right) dv$$

is continuous from $L^2(S^1)$ to $L^2(S^1)$. Using the theory of A_p weights [10] we can show that the operator $W : C_0^\infty \to \mathcal{D}'$ given by

$$W(x) = \int_{\mathbb{R}} K(x,y) f(y) dy, \quad K(x,y) = \frac{|x|^\alpha}{|y|^\alpha} \frac{1}{x-y}$$

can be continuously extended as an operator from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$ for $\alpha \in [0, \frac{1}{2})$. One only need verify that $w(x) = |x|^{2\alpha}$ is an A_2 weight. □

Finally, we show that the inverse of $v_\theta = H\eta$ given by the integral operator

$$(Tv)(\theta) = \int_0^\theta v(t) dt$$

is compact using the following general fact.

Lemma 2.2.10. *The linear operator* $G : L^2([0,1]) \to L^2([0,1])$

$$G(x) = \int_0^x g(t) dt$$

is compact.

Proof. The idea is to show that G is Holder continuous with $\alpha = \frac{1}{2}$ and constant $M = 1$. We then show that Holder functions are a compact subset of L^2.

We need only show that the image of the unit ball is relatively compact, so assume

$\|g\|_2 = 1$. By definition of our map we have

$$|G(x_0) - G(x_1)| \leq \int_{x_0}^{x_1} |1 \cdot g(t)|\, dt \leq \|g\|_2 \, |x_1 - x_0|^{\frac{1}{2}} = |x_1 - x_0|^{\frac{1}{2}}$$

where we used Cauchy-Schwartz with one function being g and the other function being the constant 1 function. Thus G is Holder continuous.

Under the L^∞ norm, we have the bound (from the Holder definition)

$$\|G\|_\infty \leq 1$$

so the G are uniformly bounded. They are also equicontinuous. That is

$$|G(x_0) - G(x_1)| < \epsilon$$

if

$$|x_0 - x_1| < \delta = \epsilon^2.$$

By the Arzela-Ascoli theorm this implies that the G form a relatively compact set in $C([0,1])$. Since there is a continuous inclusion from $C([0,1])$ to $L^2([0,1])$ and continuous functions preserve compactness, we have shown that the image of our integral operator is compact in $L^2([0,1])$. □

We can extend this proof to showing T is continuous and compact as an operator from the set of mean-zero functions in Y to the space $Y \oplus \langle 1, \sin(\theta) \rangle$.

Write f as

$$f(\theta) = \left|\sin\left(\frac{\theta}{2}\right)\right|^\gamma g(\theta), \quad g \in L^2(S^1).$$

Then we are proving that

$$Tg(\theta) = \int_0^\theta \frac{w(t)}{w(\theta)} g(t)dt$$

where $w(\theta) = \left|\sin\left(\frac{\theta}{2}\right)\right|^\gamma$, is a compact operator from the subspace of L^2 functions g that satisfy $\int_{-\pi}^{\pi} w(\theta)g(\theta) = 0$ into L^2. Near zero, we have

$$|Tg(\theta)|^2 \leq \|g\|_{L^2}^2 \int_0^\theta \frac{w^2(t)}{w^2(\theta)} dt$$

so that

$$|Tg(\theta)|^2 \lesssim \|g\|_{L^2}^2 |\theta|, \quad |\theta| \leq \theta_0 \tag{2.2.13}$$

for some θ_0. This shows that $Tg(\theta)$ is Holder continuous since. Near $\theta = 0$ we have Eq 2.2.13 and away from zero, we can uniformly bound the $\frac{1}{w(\theta)}$. By the proof of Lemma 2.2.10, we know the Holder continuous functions are compact in $L^2([0,1])$.

Having shown that K is a compact perturbation of L_0, we can now prove a decay estimate for $L = L_0 + K$.

Lemma 2.2.11. *For each $\beta < \beta_0 = \gamma - \frac{3}{2}2$ we have the decay estimate*

$$\left\|e^{tL}\eta\right\|_Y \leq Ce^{-\beta t} \|\eta\|_Y, \quad \eta \in Y$$

where C depends only on β.

Proof. We outline the proof in [6]. By [11, Section 2 of Chapter IV, Corollary 2.11 and Prop 2.12] it suffices to show that there are no eigenvalues of L in the region

$$\{\lambda \mid \text{Re } \lambda > -\beta_0\}.$$

In the case of L on Y, the continuity of $H : Y \to Y$ and the recursive formula for eigenfunctions again allows us to restrict to holomorphic functions and use the same

complex ODE as before. We slightly strengthen the analysis in Lemma 2.2.6 to show there are no eigenfunctions of L in Y. □

Useful in the nonlinear analysis in the next section, Lemma 2.2.11 also holds with gauge $\widetilde{v}(0,t) = 0$. The flow satisfies

$$\eta_t + [\sin\theta, \eta + \widetilde{v}] = 0, \quad \widetilde{v}_\theta = H\eta, \quad \widetilde{v}(0,t) = 0 \tag{2.2.14}$$

Define

$$\widetilde{L} := -[\sin\theta, \eta + \widetilde{v}]. \tag{2.2.15}$$

Lemma 2.2.12. *For $\beta < \beta_0 = \gamma - \frac{3}{2}$ and $\eta_0 \in Y$ we have*

$$\left\| e^{t\widetilde{L}}\eta_0 \right\|_Y \leq C(\beta) e^{-\beta t} \|\eta_0\|_Y$$

Proof. See [6] □

2.3 Stability in the Nonlinear Regime

For the remainder of this section, we consider initial data of the form $\omega_0 = \Omega + \epsilon\eta_0$ where ϵ is small and η_0 is a suitably regular function (as measured by the $\dot{H}^{\frac{3}{2}}(S^1)$ and Y norm. We can rotate coordinates so that $\omega_0(0) = 0$ and change A in $\Omega = -A\sin(\theta - \theta_0)$ so that $\Omega_\theta(0) = \omega_{0\theta}(0)$, giving $\eta_0 \in Y$. A time rescaling and suitable rescaling of A finally allows us to assume that $\Omega = -\sin\theta$ and $\eta_0 \in Y$. Our main condition will be

$$\|\eta_0\|_{\dot{H}^{32}(S^1)} \lesssim 1, \quad \|\eta_0\|_Y \lesssim 1 \tag{2.3.1}$$

We choose gauge (see [6, Eq 2.27]) so that $\eta(0,t) = 0$ for all time. Thus we are studying

$$\eta_t + [\sin\theta, \eta + \widetilde{v}] + \epsilon[\widetilde{v}, \eta] = 0, \quad \widetilde{v}_\theta = H\eta, \quad \widetilde{v}(0,t) = 0 \tag{2.3.2}$$

Local in time existence for $t \in [0,T]$ follows from [2] if

$$\int_0^T \|v_\theta(t)\|_{L^\infty}\, dt < \infty.$$

One can think of this as analogous to the Beale-Kato-Majda criterion in 3D Euler.

Theorem 2.3.1. *There exists an ϵ small enough so that the local solution to Eq 2.3.2 can be continued globally for any initial data satisfying Eq 2.3.1, and the solution satisfies*

$$\|\eta(t)\|_Y \leq 2C_0 e^{-\beta t}, \quad \|\eta(t)\|_{H^{3/2}} \leq 2C_0$$

where C_0 is the constant from Lemma 2.2.12.

Let us outline the proof. First, we show that the quantity

$$\|M\eta(t)\|_{L^2(S^1)},$$

where Mf_k in Fourier coordinates is

$$\widehat{Mf_k} = \sqrt{(k^2-1)(|k|-1)}\hat{f}_k,$$

has a self-improving bound. Note that this quantity is conserved in the linear regime $e^{t\widetilde{L}}$. We then rewrite Eq 2.3.2 as

$$\eta_s = \widetilde{L}\eta + g(s),$$

and then use the boundedness of $\|M\eta(t)\|_{L^2(S^1)}$ to show that $\|g(s)\|_Y$ decays exponentially in time.

Lemma 2.3.2. *Assume η_0 satisfies the conditions in (2.3.1) and*

$$\|\eta(t)\|_Y \leq \Gamma e^{-\beta t}, \quad \|M\eta(t)\|_2 \leq \Gamma$$

for $t \in [0,T]$. Then there exists $c \geq 0$ such that

$$\|M\eta(t)\|_{L^2(S^1)} \leq e^{c\epsilon\Gamma}$$

for $t \in [0,T]$.

Proof. First, a quick outline. Using the conservation of $M\eta$ at the linear level we have

$$\frac{d}{dt}\int_{S^1} \|M\eta(t)\|^2 \, d\theta = -\int_{S^1} 2\epsilon(M[\widetilde{v},\eta])M\eta d\theta$$

If we can obtain the bound on the right-hand side

$$\left|\int_{S^1} 2\epsilon(M[\widetilde{v},\eta])M\eta d\theta\right| \lesssim \|\eta\|_{\dot{H}^\sigma(S^1)a/\langle\sin(\theta),\cos(\theta)\rangle} \|M\eta\|_2^2$$

for all $\eta \in H^2(S^1)$ and for some $\sigma > \frac{1}{2}$, then since

$$\begin{aligned}\|\eta\|_{\dot{H}^\sigma(S^1)/\langle\sin(\theta),\cos(\theta)\rangle} \|M\eta\|_2^2 &\lesssim \epsilon\left(\|M\eta(t)\|_{L^2}^{1-\alpha}\|\eta(t)\|_Y^\alpha + \|\eta\|_Y\right)\|M\eta(t)\|_{L^2}^2 \\ &\lesssim \epsilon\Gamma e^{-\beta\alpha t}\|M\eta(t)\|_{L^2}^2\end{aligned}$$

for α slightly below $\frac{2}{3}$, we will have finally

$$\frac{d}{dt}\int_{S^1} \|M\eta(t)\|^2 \, d\theta \lesssim \epsilon\Gamma e^{-\beta\alpha t}\|M\eta(t)\|_{L^2}^2 .$$

Then the Gronwall inequality

$$\|M\eta(t)\|_{L^2} \leq e^{c\epsilon\Gamma}$$

where c is a constant involving the constants of the above inequalities.

It thus suffices to verify that

$$\left|\int_{S^1} (M[\widetilde{v},\eta])M\eta d\theta\right| \lesssim \|\eta\|_{\dot{H}^\sigma(S^1)/\langle\sin(\theta),\cos(\theta)\rangle} \|M\eta\|_2^2$$

Since the quantity

$$\int_{S^1} 2\epsilon(M[\widetilde{v},\eta])M\eta d\theta$$

is unchanged by additions of $\langle \sin\theta, \cos\theta\rangle$ so it suffices to prove

$$\left|\int_{S^1} 2\epsilon(M[\widetilde{v},\eta])M\eta d\theta\right| \lesssim \|\eta\|_{\dot{H}^\sigma(S^1)} \|M\eta\|_2^2 \tag{2.3.3}$$

Recall that $[\widetilde{v},\eta] = \widetilde{v}\eta_\theta - \widetilde{v}_\theta\eta$. The second term with $\widetilde{v}_\theta\eta$ can be estimated with the standard Sobolev product estimate

$$\|\widetilde{v}_\theta\eta\|_{H^{3/2}} \lesssim \|\widetilde{v}_\theta\|_\infty \|\eta\|_{H^{\frac{3}{2}}} + \|\eta\|_\infty \|\widetilde{v}_\theta\|_{H^{\frac{3}{2}}}$$

The first term with $\widetilde{v}\eta_\theta$ can be estimated via integration by parts

$$M(\widetilde{v}\eta_\theta) = \widetilde{v}M\eta_\theta + [M,\widetilde{v}]\eta_\theta$$

where $[M,\widetilde{v}]$ the commutator of M and multiplication by $\widetilde{v}$. The first term can be estimated

$$\int_{S^1} \widetilde{v}(M\eta)_\theta(M\eta)d\theta = -\frac{1}{2}\int_{S^1} \widetilde{v}_\theta |M\eta|^2 d\theta.$$

The term $[M, \widetilde{v}]\eta_\theta$ can be estimated with the standard Kato-Ponce estimate (see [12])

$$\|[M, \widetilde{v}]\eta_\theta\|_{L^2} \lesssim \|\widetilde{v}_\theta\|_{L^\infty} \|\eta_\theta\|_{H^{\frac{1}{2}}} + \|\widetilde{v}\|_{H^{\frac{3}{2}}} \|\eta_\theta\|_{L^\infty}$$

We can replace $\|\eta_\theta\|_{L^\infty}$ by $\|\eta\|_{H^{\frac{1}{2}}}$ by adding an extra ϵ-derivative to $\|\widetilde{v}\|_{H^{\frac{3}{2}}} \sim \|\eta\|_{H^{\frac{1}{2}}}$ to obtain

$$\|[M, \widetilde{v}]\eta_\theta\|_{L^2} \lesssim \|\widetilde{v}_\theta\|_{L^\infty} \|\eta_\theta\|_{H^{\frac{1}{2}}} + \|\eta\|_{H^\sigma} \|\eta_\theta\|_{H^{\frac{1}{2}}} \lesssim \|\eta\|_{H^\sigma} \|\eta_\theta\|_{H^{1/2}}$$

for some $\sigma > 1/2$. Combining the above estimates gives our desired bound Eq 2.3.3. □

Lemma 2.3.3. *There exists new time coordinates s so that Eq 2.3.2 becomes*

$$\eta_s = \widetilde{L}\eta + g$$

for a $g(s)$ that satisfies

$$\|g(s)\|_Y \lesssim \epsilon\Gamma^2 e^{\frac{-4}{3}\beta t}$$

for $t \in [0, T]$.

Proof. Setting

$$b(t) = \widetilde{v}_\theta(0, t), \quad w = \widetilde{v} - b(t)U, \quad U = -\sin(\theta)$$

so we can rewrite the nonlinear term as

$$[\widetilde{v}, \eta] = [w, \eta] + b(t)[U, \eta].$$

We can thus rewrite Eq 2.3.2 as

$$\eta_t - (1 + \epsilon\, b(t))\widetilde{L}\eta - \epsilon b(t)[U, \widetilde{v} - b(t)U] + \epsilon[\widetilde{v} - b(t)U, \eta] = 0.$$

Dividing through by $(1+\epsilon\, b(t))$ and changing time coordinates to

$$\frac{ds}{dt} = 1 + \epsilon b(t), \quad s(0) = 0$$

gives

$$\eta_s = \widetilde{L}\eta + g(s)$$

where

$$g(s) = \frac{\epsilon b(t)}{1+\epsilon b(t)}[U, \widetilde{v} - b(t)U] - \frac{\epsilon}{1+\epsilon b(t)}[\widetilde{v} - b(t)U, \eta]. \tag{2.3.4}$$

Thus to obtain our decay estiamte

$$\|g(s)\|_Y \lesssim \epsilon\Gamma^2 e^{\frac{-4}{3}\beta t}$$

we will bound the two terms in $g(s)$. For the first term, it suffices to show that

$$\|b(t)\| \lesssim \Gamma e^{-\beta\, t}, \quad \|[U, \widetilde{v} - b(t)U]\|_Y \lesssim \|\eta\|_Y$$

By the continuity of the Hilbert transform on Y we have

$$|b(t)| = |H\eta(0,t)| \lesssim \|\eta_0\|_Y \lesssim \Gamma e^{-\beta\, t}$$

Next we prove

$$\|[U, \widetilde{v} - b(t)U]\|_Y \lesssim \|\eta\|_Y .$$

Since $w(0) = w'(0) = 0$ and from the proof that H is continuous on Y, we see

$$|w(\theta)| \lesssim (\sin\theta/2)^2 \|\eta\|_Y \tag{2.3.5}$$

and thus

$$\|[U, \widetilde{v} - b(t)U]\|_Y = \|[U, w]\|_Y \lesssim \|\eta\|_Y .$$

Finally, for the second term in Eq 2.3.4 we will prove that $\|[\widetilde{v} - b(t)U, \eta]\|_Y = \|[w, \eta]\|_Y$ can be controlled via $\|M\eta\|$ and $\|\eta\|_Y$. We estimate $[w, \eta]$ in Y by estimating separately $w\eta_\theta$ and $w_\theta\eta$. For the latter, we have

$$\begin{aligned}
\|w_\theta\eta\|_Y &\lesssim \|w_\theta\|_{L^\infty} \|\eta\|_Y \\
&= \|H\eta - H\eta(0)U_\theta\|_{L^\infty} \|\eta\|_Y \\
&\lesssim (\|H\eta\|_{L^\infty} + |H\eta(0)|) \|\eta\|_Y \\
&\lesssim \left(\|\eta\|^{1-\alpha} \|\eta\|_{L^2}^{\alpha} + \|\eta\|_Y \right) \|\eta\|_Y \\
&\lesssim \|M\eta\|^{1-\alpha} \|\eta\|_Y^{1+\alpha} + \|\eta\|_Y^2
\end{aligned}$$

for α slightly below $2/3$. To estimate $w\eta_\theta$ note that because of Eq 2.3.5 we have

$$\|w\eta_\theta\|_Y \lesssim \|\eta\|_Y \|\eta_\theta\|_{L^2} \lesssim \|\eta\|_Y \|\eta\|_{L^2}^{1/3} \|\eta\|_{H^{3/2}}^{2/3} \lesssim \|M\eta\|^{2/3} \|\eta\|_Y^{4/3} + \|\eta\|_Y^2 .$$

Combining these bounds we get

$$\|[w, \eta]\|_Y \lesssim \|M\eta\|^{1-\sigma} \|\eta\|_Y^{1+\sigma} + \|\eta\|_Y^2$$

for some $\sigma > 0$. Since $\|\eta\|_Y$ decays exponentially and $\|M\eta\|_2$ is bounded, the above bound means

$$\|[w, \eta]\|_Y = \|[\widetilde{v} - b(t)U, \eta]\|$$

decays exponentially. We have shown that each term of $g(s)$ decays exponentially and thus $\|g(s)\|_Y$ decays exponentially. □

Duhamel's principal tells us that the solution to

$$\eta_s = \widetilde{L}\eta + g(s), \quad \eta(0,\theta) = \eta_0$$

is given by

$$\eta(s,\theta) = e^{s\widetilde{L}}\eta_0 + \int_0^s (e^{tL}g(t))(s,\theta)dt.$$

We have the bounds

$$\left\| e^{s\widetilde{L}}\eta_0 \right\| \leq C_0 e^{-\beta\, s} \left\|\eta_0\right\|_Y, \quad \|g(s)\| \leq \widetilde{c}\epsilon\Gamma^2 e^{\frac{-4}{3}\beta s}$$

from Lemma 2.3.3 and Lemma 2.2.12. Thus from the Duhamel formula

$$\|\eta(s)\|_Y \leq C_0 e^{-\beta s} + \widetilde{c}\epsilon\Gamma^2 \int_0^s C_0 e^{-\beta(s-s')-\frac{4}{3}\beta s'} ds' = C_0(1 + c\epsilon\Gamma^2)e^{-\beta s}$$

Using the fact that

$$e^{-\epsilon\Gamma} \leq \frac{e^{-\beta\, s}}{e^{-\beta\, t}} \leq e^{\epsilon\Gamma}$$

we can change back to the t variable and get

$$\|\eta(t)\|_Y \leq e^{c\epsilon\Gamma}(1 + c\epsilon\Gamma^2)C_0 e^{-\beta\, t}.$$

Finally, we prove Theorem 2.3.1. Consider the local solution $\eta(0) = \eta_0, \eta$ with initial data that satisfies Eq 2.3.1. Choose $\Gamma = 2C_0$ and ϵ small enough so that

$$\max(e^{c\epsilon\Gamma}, e^{c\epsilon\Gamma}(1 + c\epsilon\Gamma^2)C_0) = e^{c\epsilon\Gamma}(1 + c\epsilon\Gamma^2)C_0 < \Gamma.$$

where we have used that $C_0 \geq 1$. Then by Lemma 2.3.2 and Lemma 2.3.3 and by

continuity we have that the local solution satisfies

$$\|\eta(t)\|_Y < 2C_0 e^{-\beta t}, \quad \|\eta(t)\|_{H^{3/2}} < 2C_0$$

for t in some open time interval. Suppose towards a contradiction equality is achieved for a first moment in time T. However, by the choice of ϵ and the bounds in Lemma 2.3.2 and Lemma 2.3.3 this is impossible.

2.4 Numerics

Spectral methods are a family of techniques for solving PDEs with smooth solutions. The idea is to work in Fourier coordinates since the truncation error to a finite set of frequencies $[-N, N]$ will decay exponentially in N. We base the following discussion on [13]. Let P_N be the projection onto the Fourier modes in the range $[-N, N]$. Then we wish to solve

$$P_N(\omega_t + [v, \omega]) = 0.$$

Thus we write

$$\omega_N(\theta) = \sum_{k=-N}^{N} \hat{\omega}_k e^{k\theta i}, \quad \hat{v}_N(\theta) = \sum_{k=1}^{N} v_k e^{k\theta i}.$$

The quadratic terms $P_N([v, \omega]) = P_N(\partial_\theta \omega v) - P_N(\partial_\theta v \omega)$ are take time $\mathcal{O}(N^2)$ to compute. Using the Fast Fourier Transform we can reduce this time to $\mathcal{O}(N \log N)$. Consider a grid $x_j, j = 1, \ldots, N$ on the circle and let u_j denote $u(x_j)$ for a function u. Define the discrete Fourier transform

$$\hat{u}_k = \frac{1}{N} \sum_{j=1}^{N} u_j e^{-ikx_j}, u_j = \sum_{k=-K}^{K} \hat{u}_k e^{ikx_j}.$$

We denote $\mathcal{F}(u), \mathcal{F}^{-1}(u)$ the Fast Fourier Transform that computes these vectors in $\mathcal{O}(N \log N)$ time. Our procedure for computing a multiplication $u \cdot v$ is thus

1. Transform $\hat{u}_k, \hat{v}_m$ to physical space: $u_k = \mathcal{F}^{-1}(\hat{u}_k), v_k = \mathcal{F}^{-1}(\hat{v}_k)$.

2. Multiplying pointwise on the grid to get $w_j = u_j \cdot v_j$.

3. Transform back: $\hat{w}_l = \mathcal{F}(w_j)$

This process introduces so-called aliasing errors where high frequencies wrap to low frequencies on the finite grid $[-N, N]$. Note that the highest wave number that can be represented as a grid function f_j with $j = 1, \ldots, N = 2K$ is K. Thus higher wave numbers $k > K$ are mapped back to the range $[-K, K]$ via

$$k - nN \text{ for some } n \in \mathbb{Z}.$$

To see how this impacts the multiplication process described above, consider the discrete orthogonality relation

$$\frac{1}{N} \sum_{j=1}^{N} e^{ikx_j} e^{imx_j} = \delta_{k,-m+n\cdot N}$$

for some $n \in \mathbb{Z}$ so that $m - n \cdot N$ is in the range $[-K, K]$. One can check that in Step 3, we thus get

$$\hat{w}_l = \sum_{k=-K, l=k+m+n\cdot N, |m| \leq K}^{K} \hat{u}_k \hat{v}_m. \tag{2.4.1}$$

To remove the aliases corresponding to $\ldots + n \cdot N$ we perform the multiplication on a refined grid with $2M$ sample points (we will discuss the optimal choice of M). That is, in Step 1, we zero pad the frequencies $[-M, -K-1]$ and $[K+1, M]$. Step 2 is thus performed on a denser grid with $2M$ sample points. When we transform back in Step 3, we truncate to $[-K, K]$. We will now find the optimal M that keeps $[-K, K]$ alias

free. Note that on the larger grid $[-M, M]$ we have

$$\hat{w}_l = \sum_{k=-M/2, l=k+m+n\cdot 2M, |m|\leq M}^{M} \hat{u}_k \hat{v}_m.$$

To keep $[-K, K]$ alias free, we want that when $m = k = K$ the mode $l = 2K + n \cdot 2M$ is outside $[-K, K]$. That is, we have the inequality $2K - 2M < -K - 1$ which we can rewrite as

$$M \geq \frac{3(K+1)}{2} - 1.$$

giving the so-called 3/2 rule. Below is the code we use in Matlab.

```
% X = [X_1; X_2; ...] fourier modes (column vec)
% Y = [Y_1; Y_2; ...] fourier modes (column vec)
% should be same size
function product = compute_product(X, Y)
    % all this for scaling grid to fix aliasing (2/3 rule)
    K = min(size(X,2), size(Y, 2));
    N2 = 3*K + 1; % size of the new grid

    % resized grid (middle third is zeros) because of the way
       matlab stores coefficients
    % what really happens is that we expand grid on either side
        of the spectrum and pad with zeros
    % F = [X .... 0 0 0 0 0 0 ... flip(conj(X)) ]
    % total size is 3K + 1
    F = N2 * [0 X zeros(1, K) conj(flip(X))];
```

```
14      G = N2 * [0 Y zeros(1, K) conj(flip(Y))];
15
16      % fast way to compute product with FFT O(n log n)
17      product = 1/N2 * fft(ifft(F) .* ifft(G));
18      % product does not respect constant mode staying zero, but
            we project it anyway
19      % we also ignore higher modes
20      product = product(2:(K + 1));
21  end
```

2.4.1 Numerical Experiments

Two important insights from the numerics are stability near equilibria and conservation of $\omega_\theta(\theta(t),t)$ where $\omega(\theta(t),t)=0$. This conservation can then be seen by direct computation and does not depend on a particular Biot-Savart law: simply differentiate the De Gregorio equation with respect to θ and obtain

$$\omega_{\theta t} + v\omega_{\theta\theta} = u_{\theta\theta}\omega$$

Notice that the left side is the time derivative of the conserved quantity

$$M(t) = \partial_\theta \omega(\theta(t),t)$$

and the right side is zero for $\omega(\theta(t),t)$ by definition of $\theta(t)$.

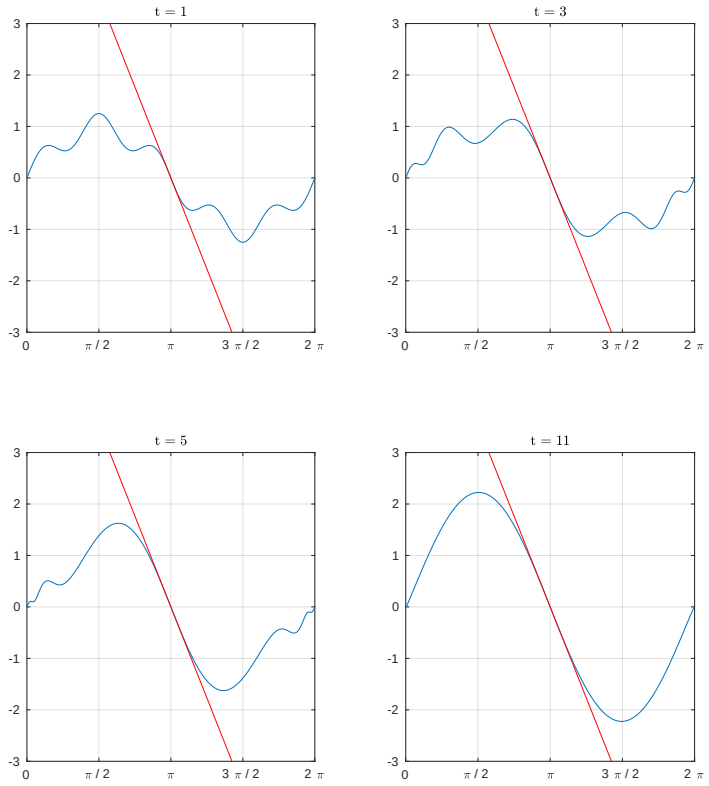


Figure 2.1: Evolution of slightly perturbed $\sin(\theta)$ as $\sin(\theta) + \frac{1}{4}\sin(5\theta)$. Notice how the slope of the tangent line is preserved in time.

This means that as the zeros of ω become close, the second derivative becomes large

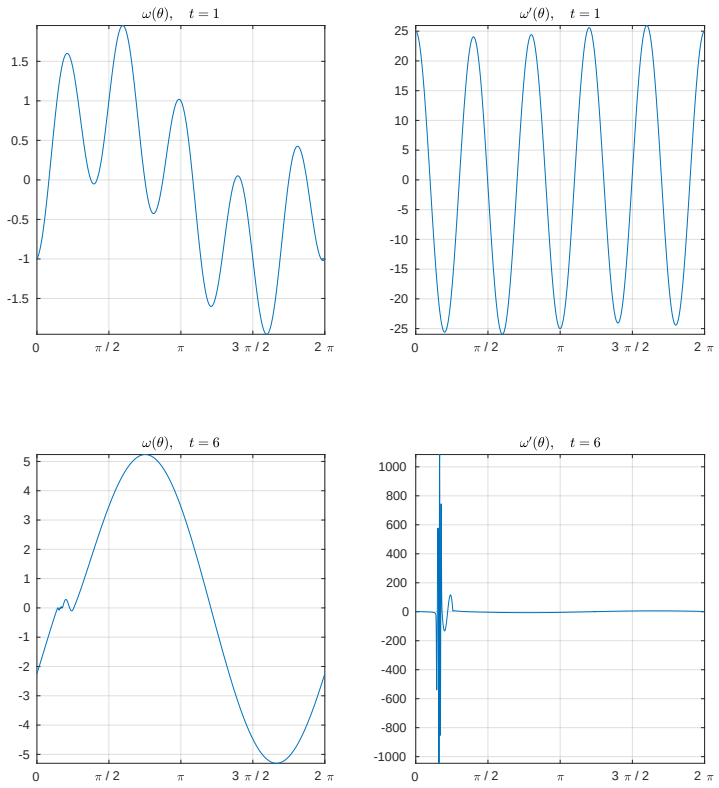


Figure 2.2: The second derivative blows up as the zeros of ω get close.

A general perspective confirms the derivative at the zeros as a conserved quantity. Let $G = \mathrm{Diff}_+(S^1)$ be the infinite dimensional Lie group of orientation preserving diffeomorphisms and $\mathfrak{g}$ its Lie algebra of divergence free vector fields. Formally, the De Gregorio equation is of the form

$$\dot{\xi} = [L(\xi), \xi] = ad_{L(\xi)}(\xi).$$

The trajectory $\xi(t)$ with $\xi(0) = \xi_0$ lies in the adjoint orbit

$$\mathcal{O}_{\xi_0} = \{a \cdot \xi_0 \cdot a^{-1},\ a \in G\}.$$

In our case, the adjoint orbits are the pushfoward of vector fields under diffeomorphisms in G

$$\mathcal{O}_{\omega_0} = \{\phi_{\#}\omega_0,\ \phi \in G\}.$$

Thus we wish to find invariants of orbits of vector fields under orientation preserving diffemorphisms of S^1. In general these invariants are hard to classify. However, the results in Hitchins [14] show that when the zeros of ω are nondegenerate (that is, for zeros θ_i we have $\omega'(\theta_i, 0) \neq 0$) then the orbit invariants are the number of zeros $2m$, the derivatives $\partial_\theta \omega(\theta_i(t), t)$ at each of the zeros $\theta_i(t)$, and the quantity

$$\text{p.v.} \int \frac{d\theta}{\omega}$$

which one can think of as an analogue of the Kelvin-Helmholtz law for classical fluids. Note that the solutions to our equations are contained within these orbits and thus necessarily conserve these quantities. This is not a complete list however, as [7] shows.

Chapter 3

Crowd Dynamics

3.1 Introduction

A variety of models for human crowds have been developed in fields such as computer graphics [15], computational sociology [16], and physics [17]. One can divide these models into two broad categories: agent-based and density-based.

Agent models often treat humans as particles. A popular example is Helbing's "social force" model where people avoid collision via a repulsive force balanced against an alignment force with their desired velocity [16]. Experiments support this model in the low density regime [18]. First order agent based models are popular in the robotics and granular media literature [19], [20]. Here collision avoidance is enforced by a hard constraint. At each time step, the desired velocity of every agent is projected (essentially solving a quadratic minimization problem) onto a set of legal velocities that guarantees no overlap. We will explore this kind of model in this chapter.

It is also worth mentioning the density-based models. Despite some experimental work using video from concerts [21], second-order agent-based models break down in

high density scenarios due to numerical instability. In addition to the first-order models discussed above, anothert approach is to instead treat the crowd as a continuous density. Collision avoidance becomes an incompressibility constraint. Hughes [17] is a classic example. Mean field games where a continuous limit results from taking the number of agents to infinity is another approach [22]. There are also multiscale models where individual agents are embedded in a continuous density with a pressure field that enforces incompressibility [23], [24].

3.2 Background

A number of concepts from convex analysis will be useful. Note that unless otherwise mentioned, we use the standard Euclidean norm on $\mathbb{R}^n$ denoted by $\|\cdot\|$.

Definition 3.2.1. A convex cone C in $\mathbb{R}^n$ is a set of vectors such that if $v_1, v_2 \in C$ and $\lambda_1, \lambda_2 \in \mathbb{R}^+$, then $\lambda_1 v_1 + \lambda_2 v_2 \in C$.

The positive cone of $\mathbb{R}^n$ is the set $\{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \ldots, n\}$.

Definition 3.2.2. Let $v_i \in \mathbb{R}^n, i = 1, \ldots m$ be m vectors and let V be the matrix with columns v_i. Let $C \subset \mathbb{R}^m$ be a convex cone. We say the vectors are linearly independent with respect to C if

$$\text{if } \lambda \in C \text{ such that } V\lambda = 0 \implies \lambda = 0. \tag{3.2.1}$$

The following lemma shows that Condition 3.2.1 guarantees that for any $y \in \mathbb{R}^n$ the set

$$\{\lambda \in C \mid y = V\lambda\}$$

is bounded.

Lemma 3.2.3. *Let V, C be as in Condition 3.2.1. Then there exists $L > 0$ (independent of y, λ) such that for all $y \in \mathbb{R}^n$ and all $\lambda \in C$ such that $V\lambda = y$ we have*

$$\|\lambda\| \leq L \|y\| .$$

Proof. Suppose towards a contradiction that there exists $y_k \to y$ and $\lambda_k \to \lambda$ satisfying $V\lambda_k = y_k$ such that

$$\frac{\|\lambda_k\|}{\|y_k\|} \to \infty. \tag{3.2.2}$$

Define

$$w_k := \frac{\lambda_k}{\|\lambda_k\|}.$$

Since $\|w_k\| = 1$, then w_k has a convergent subsequence $w_{k_m} \to w$ with $\|w\| = 1$. Then by the assumption that $V\lambda_k = y_k$ and Eq 3.2.2 we have

$$\lim_{m\to\infty} V w_{k_m} = \lim_{m\to\infty} \frac{y_k}{\|\lambda_{k_m}\|} = 0. \tag{3.2.3}$$

Since all linear operators are continuous in finite dimensions, we also have

$$\lim_{m\to\infty} V w_{k_m} = V w. \tag{3.2.4}$$

Combining Eq 3.2.2 and Eq 3.2.4 we have

$$Vw = 0.$$

Condition 3.2.1 then implies that

$$w = 0.$$

But this contradicts the fact that $\|w\| = 1$.

□

Definition 3.2.4. Let $C \subset \mathbb{R}^n$ be a convex set. A point $p \in C$ is an *extreme point* if there are no points $x_1, x_2 \in C, x_i \neq p$ such that there exists $\lambda \in (0,1)$ so that

$$p = \lambda x_1 + (1-\lambda)x_2.$$

In other words, there is no open line in C that contains p.

Definition 3.2.5. Let $p_i, i = 1, \ldots, n$ be finitely many points. The convex hull of $p_1, \ldots, p_n$ is defined as

$$\mathrm{CO}(p_1, \ldots, p_n) := \{p_1\lambda_1 + \cdots p_n\lambda_n \mid \lambda_i \in \mathbb{R}^+, \text{ for all } i = 1, \ldots, n \text{ and } \lambda_1 + \cdots + \lambda_n = 1\}$$

Now we define convex functions that can take values in the extended real line. We define the extended real line to be the set $\overline{\mathbb{R}} = (-\infty, \infty]$ where ∞ is allowed as a value with the following conventions:

1. $\alpha + \infty = \infty$ for $\alpha \in \mathbb{R}$.
2. $\alpha \cdot \infty = \infty$ for $\alpha > 0$.
3. $0 \cdot \infty = \infty$.

The domain of a extended real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is dom $f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. We say f is convex if

1. Its domain dom f is convex in $\mathbb{R}^n$.
2. For all $x_1, x_2 \in$ dom f and $\lambda \in (0,1)$

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2).$$

We say f is proper if dom $f \neq \emptyset$. We say f is lower semi-continuous if for every $x_0 \in$ dom f

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

A canonical example of a lower-semicontinuous function is the indicator of a closed convex set C defined as

$$1_C(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

There is a nonsmooth calculus for convex functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Classical concepts include the subdifferential and the normal cone [25].

Definition 3.2.6. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended real-valued function and let $x_0 \in$ dom f. Then the subdifferential at x_0 is

$$\partial f(x_0) = \{v \in \mathbb{R}^n \mid f(y) - f(x_0) \geq \langle v,\, y - x_0 \rangle,\ y \in \text{dom } f\}.$$

Definition 3.2.7. Let $C \subset \mathbb{R}^n$ be a convex set. Then the normal cone at $x_0 \in C$ is

$$N_C(x_0) = \{v \in \mathbb{R}^n \mid \langle v,\, y - x_0 \rangle \leq 0, \forall y \in C\}$$

with the convention that $N_C(x_0) = \emptyset$ for $x_0 \notin C$.

These two definitions are related. Define the indicator function

$$1_C(x) = \begin{cases} \infty & x \notin C \\ 0 & x \in C \end{cases}.$$

Then $\partial 1_C(x_0) = N_C(x_0)$. The definitions above extend to the case when C, f are not convex. See [26] for more details.

Definition 3.2.8. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended real-valued function and let $x_0 \in \text{dom } f$. Then the proximal subdifferential is

$$\partial_P f = \{v \in \mathbb{R}^n \mid \exists r, \alpha > 0,\ f(y) - f(x_0) + r\|y - x_0\|^2 \geq \langle v,\, y - x_0\rangle,\ \forall y \in B(x_0, \alpha).$$

Define the set projection

$$P_C(q) := \{x' \in C \mid \|x' - x\| = d(x, C)\}.$$

Definition 3.2.9. Let $C \subset \mathbb{R}^n$ be a closed set. The proximal normal cone $N_C^P(x_0)$ is defined as

$$N_C^P(x_0) = \{v \in \mathbb{R}^n \mid \text{ there exists } \alpha > 0, \quad x_0 \in P_C(x_0 + \alpha v)\}.$$

3.3 Model

We extend a first-order agent-based model developed by Maury et al. [27]. Each of the N agents is a circle with radius R, position $q_i \in \mathbb{R}^2$, and a desired velocity $U_i(q_i) \in \mathbb{R}^2$. We use the notation $q \in \mathbb{R}^{2N}$ and $U(q) \in \mathbb{R}^{2N}$ to denote vectors of all the N positions and preferred velocities. Define the signed distance between circles i and j

$$D_{ij}(q) := |q_i - q_j| - 2R.$$

The set of legal positions is

$$Q = \{q \in \mathbb{R}^{2N} \mid D_{ij}(q) \geq 0 \text{ for all } i < j\} \tag{3.3.1}$$

Note that this set is not convex already in the case of two agents.

Consider a trajectory $q(t) \in Q, t \in [0, T]$ with initial positions $q(0) = q_0$. Intuitively, to maintain the constraint that $q(t) \in Q$ we would project the preferred velocities $U(q(t))$ to the tangent space of Q at $q(t)$. But Q is not a manifold so instead of a tangent space we consider a tangent cone defined in terms of the normal cone as

$$T_Q^P(q) := \{v \in \mathbb{R}^{2N} \mid \langle v,\, w\rangle \leq 0 \text{ for all } w \in N_Q^P(q)\}.$$

Using the classical decomposition [28] of a Hilbert space into mutually polar cones

$$P_{N_Q^P(q)} + P_{T_Q^P(q)} = I,$$

we can write our evolution as

$$\frac{dq}{dt} = P_{T_Q^P(q)}(U(q(t))) = U(q(t)) - P_{N_Q^P(q)}(U(q(t))), \quad q(0) = q_0. \tag{3.3.2}$$

Standard ODE theory does not apply as $N_Q^P(q)$ does not have even continuous dependence on q. Instead we weaken Eq 3.3.2 to a differential inclusion

$$U(q) \in \frac{dq}{dt} + N_Q^P(q),\ q(0) = q_0. \tag{3.3.3}$$

We will show that this differential inclusion will have a unique, absolutely continuous solution if $U(q)$ is Lipschitz and bounded, and if Q satisfies the following regularity condition

Definition 3.3.1. Let $C \subseteq \mathbb{R}^N$ be a closed set. C is r-prox-regular at $x_0 \in \partial C$ if there exists $r > 0$ such that for all $v \in N_C(x_0), \|v\| = 1$ we have

$$B_r(x_0 + r\, v) \cap C = \emptyset \tag{3.3.4}$$

And equivalent definition is that there exists an $r(x_0) > 0$ such that for all $v \in N_C^P(x_0)$

$$\langle v,\, y - x_0 \rangle \leq \frac{\|v\|}{2r} \|y - x_0\|^2 \tag{3.3.5}$$

for all $y \in C$. This can be seen by writing Eq 3.3.4 as

$$\|y - (x_0 + r\, v/\, \|v\|)\|^2 \geq r^2$$

and expanding.

To give some intuition for this definition, prox-regularity of closed $C \subset \mathbb{R}^n$ ensures that the set projection P_C is locally well-defined. Graphically, it means that that there exists a ball centered at x_0 that is tangent to C.

We say C is uniformly prox-regular if C is r-prox-regular at every $x_0 \in \partial C$ and $r(x_0)$ is uniformly bounded away from zero. This means the set projection P_C is well-defined in a uniform neighborhood of C. Graphically, uniform prox-regularity means that we can roll a tangent ball with constant radius around the boundary of the set C.

Theorem 3.3.2 (Theorem 1, [29]). *Let H be a Hilbert space and $I = [T_0, T] \subset \mathbb{R}$. Let $C : I \to H$ be a set valued map that satisfies*

1. *For each $t \in I$, $C(t)$ is nonempty closed subset of H that is r-prox-regular.*
2. *$C(t)$ varies in an absolutely continuous way, that is, there exists an absolutely continuous function $v : I \to \mathbb{R}$ such that, for any $y \in H$ and $s, t \in I$,*

$$|d(y, C(t)) - d(y, C(s))| \leq |v(t) - v(s)|\,.$$

Let $f : I \times H \to H$ be a measurable map on the interval I such that

1. *for every $\eta > 0$ there exists a non-negative function $k_\eta \in L^1(I, \mathbb{R})$ such that for*

all $t \in I$ and for any $(x, y) \in B_\eta(0) \times B_\eta(0)$,

$$\|f(t,x) - f(t,y)\| \leq k_\eta(t)\,\|x - y\|\,;$$

2. *there exists a non-negative function $\beta \in L^1(I, \mathbb{R})$ such that, for all $t \in I$ and all $x \in \cup_{s\in I} C(s)$,*

$$\|f(t,x)\| \leq \beta(t)(1 + \|x\|).$$

Then, for any $x_0 \in C(T_0)$, the following perturbed sweeping process

$$-\dot{x}(t) \in N^P_{C(t)}(x(t)) + f(t, x(t)) \text{ a.e. } t \in I, \quad x(T_0) = x_0$$

has a unique, absolutely continuous solution $x(t), t \in I$.

This theorem covers our differential inclusion in Eq 3.3.3 with $C(t) = Q$ and $f(t, q(t)) = -U(q(t))$. To prove well-posedness of Eq 3.3.3 it thus suffices to verify that Q is uniformly prox-regular. Maury et al. [27] prove this via a long series of technical lemmas. We offer a shorter proof by noting that Q satisfies a stronger condition and then applying existing theory from [30]. Our approach highlights the geometry and offers a potential path to proving well-posedness for a model with ellipses instead of circles. The geometric insights derived from the existence proof also guide our construction of a fast and robust numerical scheme.

The outline of this chapter is as follows. In Section 3.4 we prove the main result, namely that we have a unique, absolutely continuous solution to Eq 3.3.3. In Section 3.5 we introduce a numerical algorithm for solving an extension of Eq 3.3.3 where we replace circles with ellipses.

3.4 Main Result

Theorem 3.4.1. *Assume $U : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ is Lipschitz and bounded. Let Q be the set in Eq 3.3.1. Let $T > 0$ and $q_0 \in Q$. Then*

$$U(q) \in \frac{dq}{dt} + N_Q^P(q),\ q(0) = q_0$$

has a unique, absolutely continuous solution $q(t), t \in [0, T]$.

As discussed in the previous section, Theorem 3.4.1 is an application of Theorem 3.3.2 if we can prove that Q is uniformly prox-regular. In fact, we show that Q satisfies a stronger condition.

Definition 3.4.2. A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is strongly amenable at $x_0 \in \text{dom } (f)$ if there is an open neighborhood U of x_0 on which f has a representation as $g \circ F$ with $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ a proper, lower semicontinuous, convex function and $F : U \to \mathbb{R}^m$ a C^2 map such that $(\nabla F(x_0))^T$ has columns linearly independent with respect to the cone $N_{\text{dom } g}(F(x_0))$. That is, $(\nabla F(x_0))^T$ satisfies Condition 3.2.1 with respect to the the cone $N_{\text{dom } g}(F(x_0))$. Geometrically, this means that f is a convex functions after a change of coordinates F that preserves the "dimension" of the normal cone cone $N_{\text{dom } g}(F(x_0))$.

Next, we prove a modification of [30, Corollary 2.12] that applies to our definition of prox-regularity.

Lemma 3.4.3. *Let C be a closed set. If 1_C is strongly amenable at a point $x_0 \in \partial C$ then the set C is r-prox-regular at x_0.*

Proof. Let $v \in N_C^P(x_0)$. As noted in [31, p. 8],

$$N_C^P(x_0) = \partial_P 1_C(x_0).$$

By the chain rule for proximal subdifferentials [32] we have

$$\partial_P 1_C(x_0) = (\nabla F(x_0))^T \partial_P g(F(x_0))$$

so $v = (\nabla F(x_0))^T s$ for some $s \in \partial g(F(x_0))$. Since g is convex $\partial_P g = \partial g$ [26], [33], so by definition of the subdifferential

$$1_C(y) - 1_C(x_0) = g(F(y)) - g(F(x_0)) \geq \langle s,\, F(y) - F(x_0)\rangle. \tag{3.4.1}$$

Let $H(x) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ be the Hessian of F given in coordinates

$$(H(x)(v,w))_i = \sum_{j,k} \partial_j \partial_k F_i(x) v_j w_k \tag{3.4.2}$$

We will denote the Hessian of F_i by H_i. Using this notation we can rewrite Eq 3.4.2 as

$$(H(x)(v,w))_i = v^T H_i w.$$

Let $\delta > 0$. By Taylor's theorem for C^2 functions, there exists $c \in B_\delta(x_0)$ such that for all $y \in B_\delta(x_0)$ we have

$$F(y) = F(x_0) + \nabla F(x_0)(y - x_0) + \frac{1}{2} H(c)(y - x_0, y - x_0).$$

Then

$$\begin{aligned}
\langle s,\, F(y) - F(x_0)\rangle &= \langle s,\, \nabla F(x_0)(y - x_0)\rangle + \frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle \\
&= \langle (\nabla F(x_0))^T s,\, y - x_0\rangle + \frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle \\
&= \langle v,\, y - x_0\rangle + \frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle
\end{aligned} \tag{3.4.3}$$

Combining Eq 3.4.3 with Eq 3.4.1 gives

$$1_C(y) - 1_C(x_0) = g(F(y)) - g(F(x_0)) \geq \langle v,\, y - x_0 \rangle + \frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle.$$

Because $y, x_0 \in C$ we have

$$0 \geq \langle v,\, y - x_0 \rangle + \frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle. \tag{3.4.4}$$

and thus

$$-\frac{1}{2}\langle s,\, H(c)(y - x_0, y - x_0)\rangle \geq \langle v,\, y - x_0 \rangle. \tag{3.4.5}$$

With the bound

$$\frac{\|s\|}{2}\left(\sum_{i=1}^{m} \|H_i(c)\|\right)\|y - x_0\|^2 \geq -\frac{1}{2}\langle s,\, H_i(c)(y - x_0, y - x_0)\rangle$$

and Eq 3.4.5, we have

$$\langle v,\, y - x_0 \rangle \leq \frac{\|s\|}{2}\left(\sum_{i=1}^{m} \|H_i(c)\|\right)\|y - x_0\|^2 .$$

By the definition of strong amenability, the matrix $(\nabla F(x_0))^T$ satisfies the hypothesis of Lemma 3.2.3 and thus there exists $L > 0$ such that $\|s\| \leq L\,\|v\|$. Thus with

$$r := \frac{1}{L\left(\sum_i \|H_i(c)\|\right)} \tag{3.4.6}$$

we have

$$\langle v,\, y - x_0 \rangle \leq \frac{1}{2r}\,\|v\|\,\|y - x_0\|^2$$

as desired. □

To apply this theorem to our case, set

$$F(q) = D(q)$$

where D the vector valued function $D(q) : \mathbb{R}^{2N} \to \mathbb{R}^{N(N-1)/2}$ of distances defined by enumerating the pairs of circles $(i, j),\ i < j$ according to the dictionary ordering. We can ignore the points where D is not C^2 since those points are not in the boundary of Q, the only points we are considering.

Set also

$$f(q) = 1_Q(q), \quad g = 1_{\mathbb{R}^{+N(N+1)/2}}(y)$$

In summary, we have decomposed $1_Q(q)$ as

$$f(q) = 1_Q(q) = g \circ F = 1_{\mathbb{R}^{+N(N+1)/2}}(D(q)).$$

This reflects the fact that good constraint coordinates are distances between circles, not positions in space. In the "distance coordinates" the constraint set is simply the positive quadrant of $\mathbb{R}^{N(N+1)/2}$. In some sense we have changed from primal to dual (constraint) coordinates. Indeed, Condition 3.2.1 is an analogue of the Mangasarian-Fromovitz constraint qualification condition from convex optimization. As we will see, dual coordinates are also used in the numerical algorithm.

We have reduced the problem of showing that Q is uniformly prox-regular to showing that $1_Q(q)$ is strongly amenable at all $q \in \partial Q$. It remains to check that for all $q \in \partial Q$ the matrix $(\nabla D(q))^T$ satisfies Condition 3.2.1 with respect to the cone $N_{\mathrm{dom}\ g}(y)$, which we compute next.

Since dom $g = \mathbb{R}^{+N(N+1)/2}$ then for $q \in \partial Q$ we have

$$N_{\text{dom } g}(y) = \{\lambda \in \mathbb{R}^{N(N+1)/2} \mid \lambda_i < 0, \text{ if } y_i > 0 \text{ then } \lambda_i = 0\}.$$

Using the dictionary ordering on $(i, j),\ i < j$ to enumerate the coordinates of $D(q)$, we thus have

$$N_{\text{dom } g}(D(q)) = \{\lambda \in \mathbb{R}^{N(N+1)/2} \mid \lambda_{ij} < 0, \text{ if } D_{ij}(q) > 0 \text{ then } \lambda_{ij} = 0\}$$

In other words, when circles i, j are in contact, we have $\lambda_{ij} < 0$ and otherwise $\lambda_{ij} = 0$. In this sense, the vector $\lambda \in \mathbb{R}^{N(N+1)/2}$ is a vector of Lagrange multipliers that enforce the constraint that the circles cannot overlap.

The rows of the Jacobian $\nabla D \in \mathbb{R}^{N(N+1)/2 \times 2N}$ are the gradients of the individual constraint functions $D_{ij}(q) : \mathbb{R}^{2N} \to \mathbb{R}$. Using the dictionary order indexing on pairs $i, j,\ i < j$ we thus have

$$(\nabla D(q))_{ij} = (0, \ldots, 0, -e_{ij}(q), 0, \ldots, 0, e_{ij}(q), 0, \ldots, 0) \in \mathbb{R}^{2N} \tag{3.4.7}$$

with the unit vectors

$$e_{ij}(q) = \frac{q_j - q_i}{|q_j - q_i|} \tag{3.4.8}$$

so that the ith entry of the i, j row of ∇D is $-e_{ij}(q)$ and the j entry is $e_{ij}(q)$.

Each row of ∇D corresponds to a pair of circles, so number of rows $N(N+1)/2$ surpasses the number of columns N. The rows are thus not linearly independent – but they are linearly independent with respect to the negative cone as the following lemma shows.

Lemma 3.4.4. *Let $\lambda \in N_{dom\ g}(D(q))$. If*

$$(\nabla D(q))^T \lambda = 0,$$

then $\lambda = 0$. Here $\nabla D(q)$ denotes the Jacobian of D.

Proof. Let the notation $i \sim j$ mean circle i is in contact with circle j. By assumption we have $\lambda \in N_{\mathrm{dom}\ g}(D(q))$ and

$$(\nabla D(q))^T \lambda = 0. \tag{3.4.9}$$

Coordinatewise, this gives

$$\sum_{i \neq j} \lambda_{ij} e_{ij} = 0. \tag{3.4.10}$$

Since $\lambda \in N_{\mathrm{dom}\ g}(D(q))$ we have $\lambda_{ij} = 0$ when circle i and j are not in contact so in fact Eq 3.4.10 becomes

$$\sum_{i \neq j, i \sim j} \lambda_{ij} e_{ij} = 0. \tag{3.4.11}$$

Assume at least one $\lambda_{ij} < 0$. Without loss of generality, assume the circles form a connected cluster. By standard theory [25] there exists a q_k that is an extreme point of $\mathrm{CO}(q_1, \ldots, q_N)$. Suppose towards a contradiction that the vectors $\{e_{kj} \mid j \sim k\}$ are linearly dependent so that for some $\lambda \in \mathbb{R}^{N(N+1)/2+}, \lambda \neq 0$ we have

$$\sum_{j \sim k} e_{kj} \lambda_{kj} = \sum_{j \sim k} \frac{q_k - q_j}{2R} \lambda_{kj} = 0. \tag{3.4.12}$$

Let M be the number of tangent circles to k. We can rearrange Eq 3.4.12 and get

$$\left(\sum_{j \sim k} \lambda_{kj} \right) q_k = \sum_{j \sim k} q_j \lambda_{kj}.$$

Then since $\lambda \leq 0$ with at least one $\lambda_{jk} < 0$, dividing both sides by $1/\sum_{j \sim k} \lambda_{kj}$ shows

that q_k is a convex combination of q_j, contradicting the Definition 3.2.4 of extreme point. Hence we have shown that $\lambda_{kj} = 0$ for all contacting circles j in contact with k.

Now remove circle k from the cluster and repeat the above argument. At each step, we remove one circle k and set $\lambda_{ik} = 0$ for all circles i that are in contact with k. Since the circles form a connected cluster, this process can be repeated until we have shown $\lambda = 0$.

□

We have verified that $1_Q(q)$ is strongly amenable for all $q \in \partial Q$. By Proposition 3.4.3 this implies that Q is r-prox-regular at all $q \in \partial Q$. It remains to verify that Q is uniformly prox-regular.

Theorem 3.4.5. *The set Q is uniformly prox regular.*

Proof. Let $q \in \partial Q$. We want to check that r_q given in Eq 3.4.6 can be bounded uniformly away from zero in q. This requires bounding $\left\|\nabla^2 D\right\|$ and the constant L.

We first bound uniformly the term with the Hessian of D. As noted in [27, Proposition 2.15], we have

$$H_i := (\nabla^2 D)_i = (\nabla^2 D_{ij})(h) = \frac{1}{\|q_j - q_i\|}(0, \ldots, 0, -P_{e_{ij}^{\perp}}(h_j - h_i), 0, \ldots, 0, P_{eij^{\perp}}(h_j - h_i), 0, \ldots, 0)$$

with the projection

$$P_{e_{ij}^{\perp}}(h_j - h_i) = (h_j - h_i) - [(h_j - h_i) \cdot e_{ij}]e_{ij}.$$

The Hessian matrix has two eigenvalues, 0 and $1/\|q_j - q_i\|$. For $q \in \partial Q$,

$$\frac{1}{\|q_j - q_i\|} \leq \frac{1}{R}$$

so $\left\|(\nabla^2 D)_i(q)\right\| \leq 1/R$. Thus we have derived the bound

$$\left\|\nabla^2 D\right\| \leq \frac{N}{R}\left\|y-q\right\|^2. \tag{3.4.13}$$

Next we bound L in Eq 3.4.6 in q. Recall that by Lemma 3.4.4 and Lemma 3.2.3 there exists $L_q > 0$ such that for all $v \in N_Q^P(q)$ and all $\lambda \in N_{\text{dom } g}(D(q))$ we have

$$\|\lambda\| \leq L_q \|v\| \tag{3.4.14}$$

Define

$$C := \{v \in N_Q^P(q), \lambda \in N_{\text{dom } g}(D(q)) \mid (\nabla D)^T\lambda = v\}, \quad M := \{\lambda \in N_{\text{dom } g}(D(q)) \mid \|\lambda\| = 1\}$$

Rearranging Eq 3.4.14 means

$$\frac{1}{L_q^2} = \min_C \frac{\|v\|^2}{\|\lambda\|^2} = \min_M \lambda^T(\nabla D)(\nabla D)^T\lambda.$$

Using the explicit formula for ∇D given in Eq 3.4.7 one can see that

$$(\nabla D)(\nabla D)^T$$

depends only on the angles between the vectors e_{ij} defined in Eq 3.4.8. This is a compact set. Since

$$0 < \frac{1}{L_q^2} = \min_M \lambda^T(\nabla D)(\nabla D)^T\lambda,$$

then $1/L_q^2$ uniformly bounded below in q. We have shown that r in Eq 3.4.6 is uniformly bounded away from zero so Q is uniformly prox-regular. □

3.4.1 Elliptical Particles

In the case of elliptical particles, we have an additional rotational degree of freedom. Each of the N ellipses is thus specified by $(q_i, \theta) \in \mathbb{R}^3$ and a preferred velocity that includes preferred rotational velocity, so $U_i(q, \theta) : \mathbb{R}^3 \to \mathbb{R}^3$. Define

$$w = (q_1, \theta_1, q_2, \theta_2, \ldots, q_N, \theta_n), \quad U = (U_1, \ldots, U_N).$$

The set of legal positions is

$$W = \{w \in \mathbb{R}^{3N} \mid D_{ij}(w) \geq 0 \text{ for } i < j\} \tag{3.4.15}$$

where now $D_{ij}(w)$ is the distance between ellipses i and j. We conjecture that differential inclusion

$$U(w(t)) \in \frac{dw}{dt} + N_W^P(w), \quad w(0) = w_0$$

has unique and absolutely continuous solutions though we present here only a numerical algorithm for constructing approximate solutions.

3.5 Numerics

The numerical scheme approximates the constraint set W and then uses finite differences in time to solve Eq 3.4.15. Let $h > 0$. One choice of approximation is to simply truncate the Taylor expansion of D

$$\widetilde{W}(w) := \{\widetilde{w} \in \mathbb{R}^{2N} \mid D(w) + h\nabla D(w) \cdot (\widetilde{w} - w) \geq 0\}$$

The set $\widetilde{W}$ is convex and has tangent cone

$$T_{\widetilde{W}}(w) = \{v \in \mathbb{R}^{2N} \mid D(w) + h\nabla D(w) \cdot v \geq 0\}.$$

Then a finite difference approximation to the solution w is

$$w_{n+1} = w_n + hP_{T_{\widetilde{W}}(w_n)}(U(w_n)).$$

Thus at each timestep we need to solve the minimization problem

$$P_{T_{\widetilde{W}(w_n)}}(U(w_n)) = \underset{u \in T_{\widetilde{W}}(w_n)}{\arg\min} \frac{1}{2} \|U(w_n) - u\|^2$$

where $\|(q,\theta)\|^2 = \|q\|^2 + \alpha\theta^2$ weights rotations by α.

There are many methods for solving quadratic programs with linear inequality constraints, we present the Uzawa method here. Let A be a $n \times n$ symmetric, positive definite matrix, C a $m \times n$ matrix, and $d \in \mathbb{R}^m, b \in \mathbb{R}^n$. We describe how to solve the general problem

$$\min_{Cx \leq d} \frac{1}{2} x^T Ax - b^T x. \tag{3.5.1}$$

The main idea is gradient descent on the dual problem, projecting to the constraint set at each step. As in the well-posedness proof above, dual coordinates dramatically simplify the problem. In dual coordinates, projecting the gradient is simple: simply compute $\lambda_i = \max(\lambda_i, 0)$. The Lagrangian for Problem 3.5.1 is

$$L(x,\lambda) = \frac{1}{2} x^T Ax - b^T x + \lambda \cdot (Cx - d)$$

We want to find a saddle point (x, λ) so that

$$\nabla_{x,\lambda} L(x, \lambda) = 0$$

For fixed λ, the minimum with respect to x is

$$x_\lambda = A^{-1}(b - C^T\lambda)$$

Maximizing L with respect to λ via projected gradient ascent with rate $\rho > 0$ gives

$$\lambda_{k+1} = \max(0,\ \lambda_k + \rho\partial_\lambda L(\lambda_k)) = \max(0,\ \lambda_k + \rho(Cx_{\lambda_k} - d)),$$

gives two sequences λ_k, x_k that will converge to the saddle point

$$\nabla_{x,\lambda} L(x, \lambda) = 0$$

if ρ is not too large [34, Proposition 5]. For our problem, we iteratively compute the sequences

$$x_k = A^{-1}(U(q_n) - h\nabla D^T\lambda_k), \quad \lambda_{k+1} = \max(0, \lambda_k + \rho(\nabla D x_k - D)).$$

In the case of circles, we have $A = I$ and in the case of ellipses

$$A = \begin{bmatrix} I & 0 \\ 0 & L \end{bmatrix}$$

to account for the penalty on large rotations.

Thus the only things we need to compute are the pairwise distances D and ∇D. For

numerical robustness, we need the *signed* distance. In the case of circles, this is easy. In the case of ellipses, it is harder.

We present first the algorithm described in [35] for the easier problem of computing the closest point w on an ellipse E in standard position (long axis aligned with the x axis) to an arbitrary point p. The signed distance is then $\sigma(p)\,\|p - w\|$ where

$$\sigma(p) = \begin{cases} 1 & \text{if } p \notin E \\ -1 & \text{if } p \in E \end{cases}$$

One easy way to compute this is

$$\sigma(p) = \operatorname{sgn}\left(\left(\frac{p_x}{a}\right)^2 + \left(\frac{p_y}{b}\right)^2 - 1\right)$$

since the ellipse is defined by the sublevel set

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 \leq 1. \tag{3.5.2}$$

Another option is

$$\sigma(p) = \operatorname{sgn}(\langle p - w,\, n_w \rangle)$$

where n_w is the normal at w and

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

We can find a normal n to the ellipse at (x, y) by noting that the gradient of the

constraint Eq 3.5.2 orthogonal to the level set (the ellipse). Thus

$$n_w = \left(\frac{2x}{a}, \frac{2y}{b} \right).$$

Returning to the main algorithm, the key trick is to use the osculating circle $C(q)$ to approximate the ellipse at a point q and find the closest point on $C(q)$. When q is close to the actual closest point w on E, then the closest point $v(q)$ on $C(q)$ will also be close to w since E is well-approximated by $C(q)$.

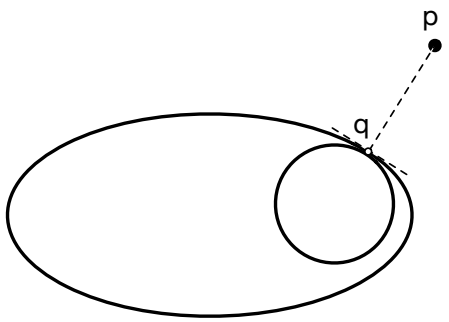


Figure 3.1: Projecting onto the osculating circle is a good approximation for projecting onto the ellipse

Thus, we have reduced the problem to finding the closest point from p on $C(q)$ which can be done with the explicit formula

$$v(q) = C_{\text{center}}(q) + R(q) \frac{p - C_{\text{center}}(q)}{\|p - C_{\text{center}}(q)\|}$$

Fortunately, in the case of the ellipse, we have explicit formula

$$C_{\text{center}}(q) = (a^2 - b^2) \left(\frac{q_x^3}{a^4}, \frac{-q_y^3}{b^4} \right)$$

so that by definition of the osculating circle

$$C_{\text{radius}}(q) = \|q - C_{\text{center}}(q)\| .$$

To recover the actual closest point $w \in E$, we iterate

1. Start at arbitrary q on the ellipse.
2. Find $v(q)$.
3. If $v(q)$ is already on the ellipse E, we're done. If not, choose q_1 to be a "close" (more details below) point on the ellipse to $v(q)$ and start again at step 2.
4. Repeating this process will produce a sequence q_n, r_n such that $q_n \to q, r_n \to q$ where q is the closest point on the ellipse to p.

Since the algorithm finds a fixed point, it terminates when q_{n+1} is within a predefined tolerance of q_n.

We expand on step (2). If we have a point $\widetilde{q}$ that is close to the ellipse E, one way to find a close point q on the ellipse is to simply rescale $\widetilde{q}$ via

$$q = (\frac{\widetilde{q}_x}{a}, \frac{\widetilde{q}_y}{b})$$

Geometrically, we are simply sending $\widetilde{q}$ to the point q that is the intersection of E and the line through the points 0 and $\widetilde{q}$.

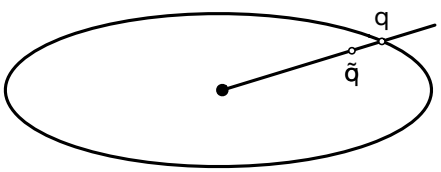


Figure 3.2: Renormalizing sends points to the boundary of the ellipse

The above algorithm finds the closest point on an ellipse in standard position, its long axis aligned with the x-axis. Call the operator that maps from p to its closest point w on the ellipse $P_{\text{std}} : \mathbb{R}^2 \to E$. Then we can extend to projecting on ellipses with rotation matrix R and translation T via the change of coordinates

$$P(p) = R\,P_{\text{std}}(R^{-1}(p - T)) + T \tag{3.5.3}$$

We extend this algorithm to find the closest points between two ellipses E_1, E_2 with osculating circle centers $C^1(q), C^2(r)$ for $q \in \partial E_1, r \in \partial E_2$. Let P_1, P_2 be projections onto E_1, E_2 defined individually as in Eq 3.5.3. Starting with two arbitrary points $q_0 \in E_1, r_0 \in E_2$, we compute the sequence (q_n, r_n)

$$q_n = P_1(C^2(r_n)), \quad r_n = P_2(C^1(q_n)),$$

that will converge to the two closest points.

Figure 3.3 shows why we project the *centers of the osculating circles* onto the opposite ellipse. It allows us to compute the overlap amount, crucial for the signed distance

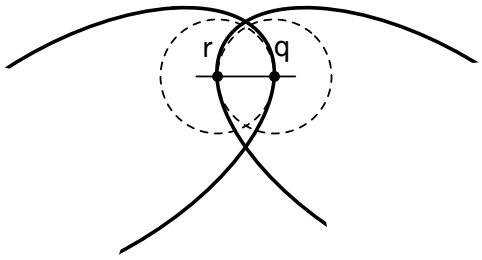


Figure 3.3: To compute the signed distance, we find the closest points on the osculating circles to the center of the opposite osculating circle

Note that from an efficiency point of view, one does not have to fully construct the matrix ∇D since it is sparse. Figure 3.4 shows an example simulation run.

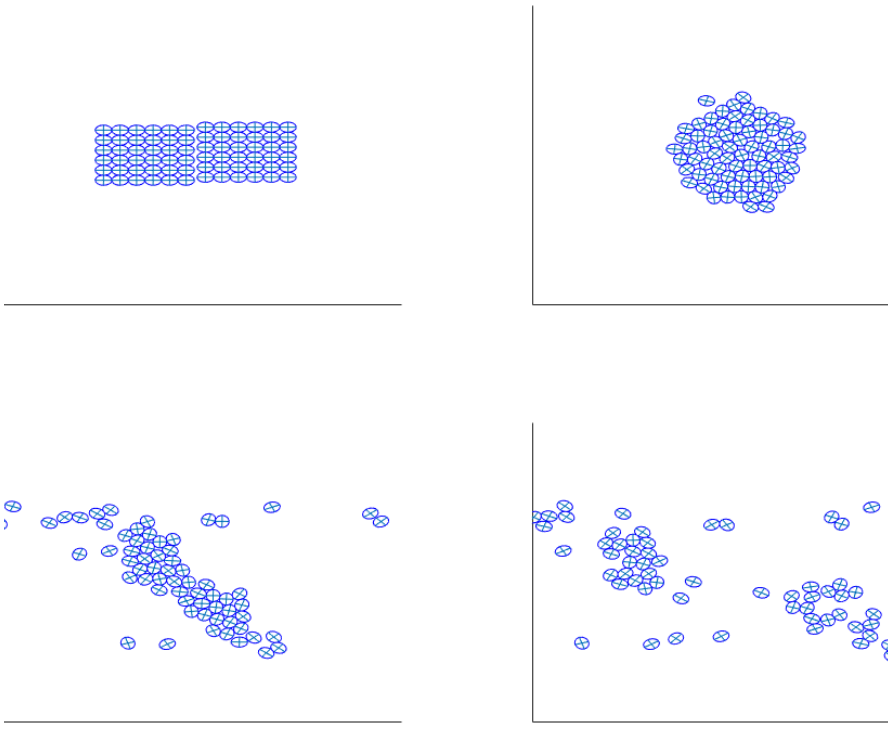


Figure 3.4: Two formations of ellipses trying to push past each other in opposite directions

Chapter 4

MRI

4.1 Introduction

Convex optimization problems with linear constraints $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a full-rank matrix with $m < n$, covers a wide variety of applications. A well-known example is compressed sensing, where one aims to recover a sparse signal given a vector $\boldsymbol{b}$ of reduced number of linear measurements and the corresponding and very special measurement vectors stored as the rows of $\boldsymbol{A}$. Under some restricting assumptions, the unknown signal is the solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with minimal ℓ_1 norm [36]–[40], that is, it is the solution of

$$\min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \|\boldsymbol{x}\|_1 . \tag{4.1.1}$$

A particular case of interest, which is motivated by application to MRI, assumes Fourier measurements [37]. In this case, sampling can be done by taking $m = \mathcal{O}(s \log n)$ random measurements of the Fourier coefficients of the unknown signal $\boldsymbol{x}^*$, where s is an upper bound for the sparseness of $\boldsymbol{x}^*$. These random measurements are dot products of the form $\langle \boldsymbol{x}^*, \boldsymbol{\psi}_k \rangle$, where $\boldsymbol{\psi}_k$ is a randomly selected row of the $n \times n$ discrete Fourier transform matrix. Therefore, $\boldsymbol{A}$ has rows $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ and $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}^*$. Candès, Romberg and Tao [37] showed that with probability 1 the solution of (4.1.1) is unique and recovers $\boldsymbol{x}^*$. This observation extends to other kinds of measurements (see [39] and also [38], [40]).

In applications, the actual measurements of $\mathbf{A}$ and $\boldsymbol{b}$ are noisy, so understanding the reconstruction error under noise is important. Several works [41]–[44] have proven that for very special measurement matrices $\boldsymbol{A}$ and for noisy measurements $\hat{\boldsymbol{b}}$ of the form $\hat{\boldsymbol{b}} = \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{e}$, where $\|\boldsymbol{e}\|_2$ is sufficiently small and $\boldsymbol{x}^*$ has a sufficiently close approximation in ℓ_1 by an s-sparse vector, the solution $\boldsymbol{x}$ of (4.1.1) with $\boldsymbol{b}$ replaced by $\hat{\boldsymbol{b}}$ is sufficiently close to $\boldsymbol{x}^*$ in ℓ_2 norm. Herman and Strohmer [45] extend this theory by allowing possible errors in the measurement matrix $\boldsymbol{A}$.

Many other applications involve error in the matrix $\boldsymbol{A}$. Examples include radar [46], remote sensing [47], telecommunications [48] and source separation [49]. More recently, Gutierrez et al. [50] developed a compression technique for model-based MRI reconstruction. Here there are at least two sources of error, the discretization of a set of ODEs and the search for a sparse basis to represent the simulation results. They use the total variation norm instead of the ℓ^1 norm.

Motivated by these broad model problems we prove a general stability result for the convex optimization problem

$$\min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \Omega(\boldsymbol{x}) \tag{4.1.2}$$

with respect to perturbations in $\boldsymbol{A}$, where Ω is from a particular class of seminorms which includes the ℓ^1 and total variation norms. Even for the special case of ℓ_1 norm, our result differs from the earlier perturbation result of Herman and Strohmer [45] since we do not enforce strong assume we are in the regime of singleton solution sets. We are thus able to eliminate in this case their conditions on the matrix $\boldsymbol{A}$ and prove a kind of local stability result. For general convex Ω, Klatte and Kumar [51] have shown that (4.1.2) is stable with respect to perturbations in $\boldsymbol{b}$. We extend their result to stability in $\boldsymbol{A}$ for a restricted class of Ω.

The paper is organized as follows. In Section 4.2 we define the various notions of set-valued regularity and describe stability results for set-valued functions that are related to our problem. In Section 4.3, we formulate and prove our theorem using tools from subspace perturbation theory, the theory of polyhedral mappings, and the theory of local error bounds for convex inequality systems. Finally, in Section 4.4 we conclude this work and clarify the difficulty of extending our approach to the full class of seminorms.

4.2 Background

We consider the following solution map

$$S(\boldsymbol{A}) = \underset{\boldsymbol{Ax}=\boldsymbol{b}}{\arg\min}\, \Omega(\boldsymbol{x}), \tag{4.2.1}$$

where Ω has polyhedral level sets. Here "polyhedral level sets" means that the unit ball $\{\boldsymbol{x} \mid \Omega(\boldsymbol{x}) \leq 1\}$ is a polyhedron. A polyhedron can be unbounded and is defined as the intersection of finitely many half spaces. Let X and Y be Banach spaces. We say a map $F : X \to Y$ is a polyhedral map if there exists polyhedra $P_i, i = 1, \ldots, N$ such that

$$\text{Graph}F = \{(x, F(x)) \mid x \in X\} = \bigcup_{i=1}^{N} P_i.$$

Note that a seminorm Ω has polyhedral level sets if the map

$$F(\mu) = \{\boldsymbol{x} \mid \Omega(\boldsymbol{x}) \leq \mu\}$$

is a polyhedral map consisting of one polyhedron P_1.

We assume that $\boldsymbol{A}_0 \in \mathbb{R}^{m\times n}$ is a full-rank matrix and study the effect of perturbing $\boldsymbol{A}_0$ on the set-valued solution map $S(\boldsymbol{A})$. Consider, for example, the case where $\Omega(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$ and $S(\boldsymbol{A}_0)$ is an entire face of the ℓ^1 ball. Then a small perturbation in $\boldsymbol{A}_0$ results in a jump from that face to a single vertex as we demonstrate in Figure 4.1. Our definition of regularity must allow such jumps. A good model set-valued function with such a jump is

$$F(t) = \begin{cases} [0,1] & \text{if } t \in (1/2, 1] \\ 0 & \text{if } t \in [0, 1/2). \end{cases}$$

This function is upper semicontinuous. We will state this and other definitions in terms

of Banach spaces X and Y with 2^X denoting all subsets of X.

Definition 4.2.1. A map $F : Y \to 2^X$ is upper semicontinuous at y_0 if for any open set V intersecting $F(y_0)$ there exists a neighborhood U of y_0 such that

$$F(y) \cap V \neq \emptyset, \quad \text{for all } y \in U.$$

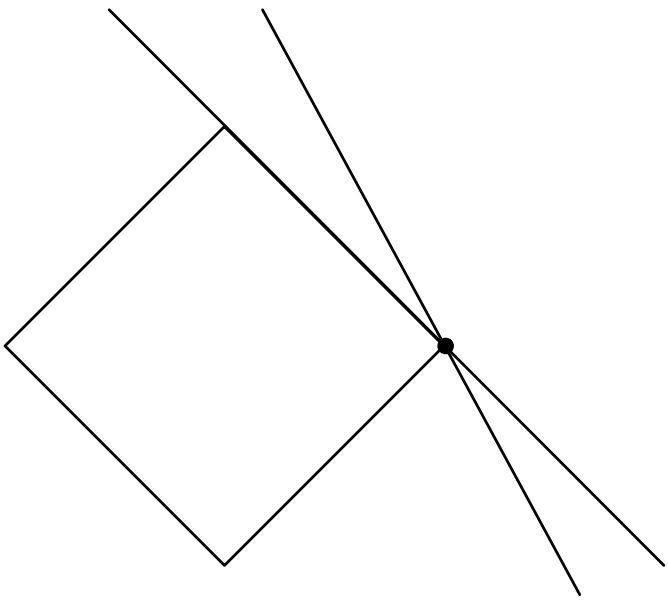


Figure 4.1: Demonstration of a case where $S(\boldsymbol{A}_0)$ is a face of the unit ℓ_1 ball and $S(\boldsymbol{A})$ for $\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{E}$ is a vertex of this ball. Here, the line intersecting the face represents the solution of $\boldsymbol{A}_0\boldsymbol{x} = \boldsymbol{b}$, the other line represents the solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, and the sets $S(\boldsymbol{A}_0)$ and $S(\boldsymbol{A})$ are the intersections of these lines, respectively, with the unit ℓ_1 ball.

It has been known since the 1970s that if Ω is convex, then the solution map $S(\boldsymbol{A})$ is upper semicontinuous [52, Theorem 1.15]. Our result upgrades the upper semicontinuity of $S(\boldsymbol{A})$ to a kind of set-valued Lipschitz regularity, which we define next.

Definition 4.2.2. A map $F : Y \to 2^X$ is is calm at $(y_0, x_0) \in \text{Graph}(F)$ if there exist neighborhoods U and V of y_0 and x_0, respectively, and a constant $L(y_0, x_0)$ such that

for all $y_1 \in U$ and $x_1 \in F(y_1) \cap V$,

$$d(x_1, F(y_0)) \leq L(y_0, x_0) \|y_0 - y_1\|.$$

We note that if $S(\boldsymbol{A})$ is calm at $(\boldsymbol{A}_0, \boldsymbol{x}_0)$ then for all $\boldsymbol{A}_1$ sufficiently close, there exists $\boldsymbol{x}_1 \in S(\boldsymbol{A}_1)$ such that

$$\|\boldsymbol{x}_0 - \boldsymbol{x}_1\| \leq C(\boldsymbol{A}_0) \|\boldsymbol{A}_0 - \boldsymbol{A}_1\|.$$

This means that small perturbations to the measurement matrix $\boldsymbol{A}$ leave at least two solutions $\boldsymbol{x}_0 \in S(\boldsymbol{A}_0)$ and $\boldsymbol{x}_1 \in S(\boldsymbol{A}_1)$ close.

A simple example shows how calmness is stronger than upper semicontinuity. The following map

$$F(t) = \{x \in \mathbb{R} \mid x^2 \leq t\}$$

for $t \in [0, 1]$, is upper semicontinuous but not calm. The distance between $F(y_0)$ and $F(y_1)$ for small y_0 and y_1 grows as $\sqrt{|y_0 - y_1|}$ and in general cannot be controlled by $|y_0 - y_1|$. Take for example $y_0 = 0$ and arbitrary small $t > 0$; then $F(0) = 0$ and $\inf_{x \in F(0)} d(x, F(t)) = d(0, F(t)) = \sqrt{t} \gg t$.

In the general setting of Banach spaces X, Y and Z, $M : Y \to 2^X$ and $f : Z \times Y \to \mathbb{R}$, Klatte and Kumar [51] provide sufficient conditions for calmness at $(y_0, x_0) \in Y \times 2^X$ of

$$S(y) = \underset{z \in M(y)}{\arg\min}\, f(z, y). \tag{4.2.2}$$

The sufficient conditions for calmness of $S(y)$ are requirements on the regularity of the constraint set $M(y)$ and the regularity of the mapping

$$L(y, \mu) = \{x \in M(y) \mid f(x, y_0) \leq \mu\} \tag{4.2.3}$$

for a certain choice of μ. The regularity of these quantities is quantified by calmness and the following notion of lower Lipschitz semicontinuity.

Definition 4.2.3. A set-valued map $F : Y \to 2^X$ is called lower Lipschitz semicontinuous at (y_0, x_0) if there exists $\delta, C > 0$ such that

$$d(x_0, F(y)) \leq C \|y_0 - y_1\|, \quad \text{for all } y \in B(y_0, \delta).$$

Lower Lipschitz semicontinuity is almost identical to calmness except y varies instead of x.

We formulate Theorem 3.1 of Klatte and Kumar [51] on sufficient conditions for calmness of $S(y)$.

Theorem 4.2.4. *Let M, f, S be as in (4.2.2) and L as in (4.2.3). Define*

$$\phi(y) := \min_{x \in M(y)} f(x, y).$$

Let $y_0 \in Y, x_0 \in X$ and $\mu_0 = \phi(y_0)$ and assume that

1. *$M(y)$ is calm and lower Lipschitz semicontinuous at (y_0, x_0).*
2. *$L(y, \mu)$ is calm at $((y_0, \mu_0), x_0)$.*

Then $S(y)$ is calm at (y_0, x_0).

At last, we define a stronger notion of regularity which will be useful later. Unlike lower Lipschitz semicontinuity, both y and y' are allowed to vary.

Definition 4.2.5. A map $F : Y \to 2^X$ is said to have the Aubin property at $(y_0, x_0) \in$ Graph F if there exists $\epsilon, \delta, C > 0$ such that

$$d(x, F(y')) \leq C \|y_0 - y_1\|$$

for all $y, y' \in B_\delta(y_0)$ and all $x \in B_\epsilon(x_0) \cap F(y)$.

4.3 The Main Theorem and its Proof

We formulate the main theorem of this paper and follow up with its proof

Theorem 4.3.1. *Assume $m \leq n$, $\boldsymbol{b} \in \mathbb{R}^m$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and Ω is a seminorm with polyhedral levelsets. Then the set-valued map*

$$S(\boldsymbol{A}) = \underset{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}}{\arg\min}\, \Omega(\boldsymbol{x}) \tag{4.3.1}$$

is calm at $(\boldsymbol{A}_0, \boldsymbol{x}_0)$ for every full-rank $\boldsymbol{A}_0 \in \mathbb{R}^{m \times n}$ such that

$$\mu_0 := \min_{\boldsymbol{A}_0\boldsymbol{x}=\boldsymbol{b}} \Omega(\boldsymbol{x}) > 0$$

and $\boldsymbol{x}_0 \in S(\boldsymbol{A}_0)$.

The assumption that $\mu_0 > 0$ is reasonable from an applied perspective. Take for example the case of compressed sensing when $\Omega(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$. In this case, $\mu_0 = 0$ corresponds to the trivial solution $\boldsymbol{x} = 0$. One can interpret this as saying that our sparsity prior tells us nothing.

In order to prove this theorem, we first prove in Section 4.3.1 that $M(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is calm and lower Lipschitz at $(\boldsymbol{A}_0, \boldsymbol{x}_0)$. We then prove in Section 4.3.2 that the map

$$L(\boldsymbol{A}, \mu) = \{\boldsymbol{x} \in M(\boldsymbol{A}) \mid \Omega(\boldsymbol{x}) \leq \mu\}$$

is calm at $((\boldsymbol{A}_0, \mu_0), \boldsymbol{x}_0)$, where

$$\mu_0 = \min_{\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}} \Omega(\boldsymbol{x}).$$

In view of Theorem 4.2.4, these results imply Theorem 4.3.1.

4.3.1 $M(\boldsymbol{A})$ is Calm and Lower Lipschitz

The following classical property of the Moore-Penrose inverse will be useful. We offer a short proof here.

Lemma 4.3.2. *Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{A}^\dagger$ be the Moore-Penrose inverse of $\boldsymbol{A}$ [53]. Then the following bound holds*

$$d(\boldsymbol{x}, M(\boldsymbol{A})) \leq \left\|\boldsymbol{A}^\dagger\right\| \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|. \tag{4.3.2}$$

Proof. The operator $\boldsymbol{I} - \boldsymbol{A}^\dagger\boldsymbol{A}$ is the projection onto the kernel of $\boldsymbol{A}$ and thus the projection $P_{M(\boldsymbol{A})}$ is given by

$$P_{M(\boldsymbol{A})} = \boldsymbol{A}^\dagger\boldsymbol{b} + (I - \boldsymbol{A}^\dagger\boldsymbol{A}).$$

Therefore,

$$\begin{aligned} d(\boldsymbol{x}, M(\boldsymbol{A})) &= \left\|\boldsymbol{x} - P_{M(\boldsymbol{A})}\right\| \\ &= \left\|\boldsymbol{x} - \boldsymbol{A}^\dagger\boldsymbol{b} - \boldsymbol{x} + \boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{x}\right\| \\ &= \left\|\boldsymbol{A}^\dagger\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^\dagger\boldsymbol{b}\right\| \\ &\leq \left\|\boldsymbol{A}^\dagger\right\| \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|. \end{aligned}$$

□

Let $\boldsymbol{A}_1 \in \mathbb{R}^{n\times m}$ be another matrix. The calmness of $M(\boldsymbol{A})$ at $(\boldsymbol{A}_0, \boldsymbol{x}_0)$ follows from the bound in (4.3.2) because of the simple fact that if $\boldsymbol{x}_1 \in M(\boldsymbol{A}_1)$ then

$$\boldsymbol{A}_0\boldsymbol{x}_1 - \boldsymbol{b} + \underline{0} = \boldsymbol{A}_0\boldsymbol{x}_1 - \boldsymbol{b} - \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{b} = (\boldsymbol{A}_0 - \boldsymbol{A}_1)\boldsymbol{x}_1.$$

Substituting this into (4.3.2) we get

$$d(\boldsymbol{x}_1, M(\boldsymbol{A}_0)) \leq \left\|\boldsymbol{A}_0^\dagger\right\| \|\boldsymbol{x}_1\| \|\boldsymbol{A}_0 - \boldsymbol{A}_1\|, \tag{4.3.3}$$

which proves that $M(\boldsymbol{A})$ is calm for a finite $\epsilon > 0$.

To see that $M(\boldsymbol{A})$ is lower Lipschitz semicontinuous, note that by the SVD construction of $\boldsymbol{A}^\dagger$

$$\left\|\boldsymbol{A}^\dagger\right\| = \frac{1}{\sigma_{\min}(\boldsymbol{A})}. \tag{4.3.4}$$

In general, since the roots of a polynomial vary continuously with repect to the coefficients, the eigenvalues of a matrix vary continuously with respect to that matrix. Thus $\sigma_{\min}(\boldsymbol{A}) = \lambda_{\min}(\boldsymbol{A}\boldsymbol{A}^T)$ depends continuously on $\boldsymbol{A}$. By (4.3.4), $\left\|\boldsymbol{A}^\dagger\right\|$ thus depends continuously on $\boldsymbol{A}$. In particular, for every $\epsilon > 0$ this implies that there exists a $\delta > 0$ such that for $\boldsymbol{A}_1 \in B_\delta(\boldsymbol{A}_0)$

$$\left\|\boldsymbol{A}_1^\dagger\right\| \leq (1+\epsilon)\left\|\boldsymbol{A}_0^\dagger\right\|. \tag{4.3.5}$$

Furthermore, by the continuity of the rank of a matrix, we can choose δ small enough to guarantee that A_1 is also full-rank, like A_0. Swapping $\boldsymbol{A}_1$ and $\boldsymbol{A}_0$ in (4.3.3) and combining it with (4.3.5) gives

$$d(\boldsymbol{x}_0, M(\boldsymbol{A}_1)) \leq (1+\epsilon)\|\boldsymbol{A}_0\|^\dagger \|\boldsymbol{x}_0\| \|\boldsymbol{A}_0 - \boldsymbol{A}_1\|.$$

This bound shows that $M(\boldsymbol{A})$ is lower Lipschitz semicontinuous at $(\boldsymbol{A}_0, \boldsymbol{x}_0)$.

4.3.2 $L(\boldsymbol{A}, \mu)$ is Calm

Note that we can write

$$L(\boldsymbol{A}, \mu) = M(\boldsymbol{A}) \cap F(\mu), \quad \text{where } F(\mu) = \{\boldsymbol{x} \mid \Omega(\boldsymbol{x}) \leq \mu\}. \tag{4.3.6}$$

The following theorem from [51] gives conditions on M and F that ensure L is calm.

Theorem 4.3.3. *Let X, Y ,Z be Banach spaces. Let $F_1 : Y \to 2^X$ be calm at $(y_0, x_0) \in$ Graph F_1, $F_2 : Z \to 2^X$ be calm at $(z_0, x_0) \in$ Graph F_2, and F_1^{-1} have the Aubin property at (x_0, y_0). Finally, let $H(z) = F_1(\boldsymbol{y}_0) \cap F_2(z)$ be calm at (z_0, x_0). Then $L(y, z) = F_1(y) \cap F_2(z)$ is calm at $((y_0, z_0), x_0)$.*

We apply the theorem with $X = \mathbb{R}^n$, Y the space of $m \times n$ matrices, and $Z = \mathbb{R}^+$. We set $F_1 = M$, $F_2 = F$, $H = W$ and define L by (4.3.6). Let $((\boldsymbol{A}_0, \mu_0), \boldsymbol{x}_0) \in$ Graph L. To see that L is calm at this point, we need to check that

1. $F^{-1}(\boldsymbol{x})$ has the Aubin property at $(\boldsymbol{x}_0, \mu_0)$.
2. $F(\mu)$ is calm at $(\mu_0, \boldsymbol{x}_0)$.
3. $W(\mu) = M(\boldsymbol{A}_0) \cap F(\mu)$ is calm at $(\mu_0, \boldsymbol{x}_0)$.

We verify property 1 in Section 4.3.2. We set the background for verifying properties 2 and 3 in Section 4.3.2 and then prove them in Sections 4.3.2 and 4.3.2, respectively.

The main trick is that we have reduced the problem from perturbations in $\boldsymbol{A}$ to the better understood case of perturbations in the right hand side the inequality system $W(\mu) = M(\boldsymbol{A}_0) \cap F(\mu)$.

F^{-1} is Aubin

Since Ω is convex, it is Lipschitz on $\boldsymbol{b}_\delta(\boldsymbol{x}_0)$ for some $\delta > 0$ so that

$$\left|\Omega(\boldsymbol{x}) - \Omega(\boldsymbol{x}')\right| \leq C_\Omega \left\|\boldsymbol{x} - \boldsymbol{x}'\right\| \text{ for all } x, \boldsymbol{x}' \in B_\delta(\boldsymbol{x}_0). \tag{4.3.7}$$

Let $\boldsymbol{x}, x' \in B_\delta(\boldsymbol{x}_0)$ and let $\mu_1 \in F^{-1}(\boldsymbol{x}) = [\Omega(\boldsymbol{x}), \infty)$. Then

$$d(\mu_1, F^{-1}(\boldsymbol{x}')) \leq \left|\mu_1 - \Omega(\boldsymbol{x}')\right| \leq \left|\Omega(\boldsymbol{x}) - \Omega(\boldsymbol{x}')\right|.$$

By (4.3.7) we thus have

$$d(\mu_1, F^{-1}(\boldsymbol{x}')) \leq C_\Omega \left\| \boldsymbol{x} - \boldsymbol{x}' \right\|.$$

Background for Proving the Calmness of F and W

To prove that $F(\mu)$ and $W(\mu)$ are calm, we can take either a geometric or algebraic perspective. The geometric proof is quite simple, while the algebraic proof enables us to compute explicit constants. The geometric perspective for F uses a classic result from the theory of polyhedral mappings [54].

Theorem 4.3.4. *Let $F : \mathbb{R} \to 2^{\mathbb{R}^n}$ be a polyhedral multivalued mapping. Then F is calm at all $(\boldsymbol{y}_0, \boldsymbol{x}_0) \in Graph(F)$.*

The geometric perspective for W uses a notion of angles between subspaces. A good introduction is [55]. Let $\mathcal{U}, \mathcal{V}$ be subspaces in $\mathbb{R}^n$ with dimensions $\dim \mathcal{U} = m, \dim \mathcal{V} = n$, and let $r = \min(m, n)$. Then we define the principal angles between $\mathcal{U}$ and $\mathcal{V}$ to be the vector $\hat{\theta} = (\theta, \cdots, \theta_r) \in [0, \pi/2]^r$ given by the following recursive definition. Define

$$\cos\theta_1 = \max_{\boldsymbol{u} \in \mathcal{U}, \boldsymbol{v} \in \mathcal{V}, \|\boldsymbol{u}\|=1, \|\boldsymbol{v}\|=1} \langle \boldsymbol{u}, \boldsymbol{v} \rangle. \tag{4.3.8}$$

Let $\boldsymbol{u}_1, \boldsymbol{v}_1$ be two unit vectors that solve this maximization problem. Call these the $k = 1$ principal directions. We then define θ_k identically except that we add to (4.3.8) the constraints that $\langle \boldsymbol{u}, \boldsymbol{u}_i \rangle = 0, \langle \boldsymbol{v}, \boldsymbol{v}_i \rangle = 0$ for $i = 1, \ldots, k-1$ where $\boldsymbol{u}_i, \boldsymbol{v}_i$ are the $\boldsymbol{I}$th principal vectors. In other words, the kth principal angle between subspaces $\mathcal{U}, \mathcal{V}$ is defined to be the largest possible angle between vectors in $\mathcal{U}, \mathcal{V}$ that are orthogonal to the subspaces of $\mathcal{U}$ and $\mathcal{V}$ spanned by the $i = 1, \ldots, k-1$ principal directions.

The algebraic approach for F and W relies on bounds on the subgradient of a convex inequality $f(\boldsymbol{x}) \leq 0$.

Definition 4.3.5. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then the subgradient of f at $\boldsymbol{x}$ is defined as

$$\partial f(\boldsymbol{x}) := \{v \in \mathbb{R}^n \mid f(\boldsymbol{z}) - f(\boldsymbol{x}) \geq \langle \boldsymbol{v}, \boldsymbol{z} - \boldsymbol{x} \rangle \text{ for all } z \in \mathbb{R}^n\}.$$

A function f is subdifferentiable at $\boldsymbol{x}$ if $\partial f(\boldsymbol{x}) \neq \emptyset$.

Using the subgradient to bound distances to constraint sets began with Hoffman's [56] original work on error bounds for systems of inequalities $\boldsymbol{A}\boldsymbol{x} \leq b$. He derives a bound analogous to the bound (4.3.3) we prove for the easier case of system of equalities $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Define $P_{\boldsymbol{A},b}$ to be the $\boldsymbol{x}$ that satisfy $\boldsymbol{A}\boldsymbol{x} \leq b$. Assuming this set is nonempty, Hoffman showed that there exists a constant C so that for all $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$d(\boldsymbol{x}, P_{\boldsymbol{A},b}) \leq C[\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}]_+,$$

where $([x]_+)_i = \max(\boldsymbol{x}_i, 0), i = 1, \ldots, n$. In his original paper, he proved the bound using results from conic geometry and then computes the constants ad hoc for a few different norms. Later results generalize this to convex inequalities $f(\boldsymbol{x}) \leq 0$ and compute sharp constants using the subgradient $\partial f(\boldsymbol{x})$ [57]. In our case, we need one of these results for convex inequalities, a kind of local error bound.

Theorem 4.3.6. *Define $Q := \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq 0\}$ and let $\boldsymbol{x}_0 \in \partial Q$. Then there exists $c(f, \boldsymbol{x}_0), \epsilon > 0$ such that*

$$d(\boldsymbol{x}, Q) \leq c(f, \boldsymbol{x}_0)[f(\boldsymbol{x})]_+ \tag{4.3.9}$$

for all $\boldsymbol{x} \in B(\boldsymbol{x}_0, \epsilon)$, if and only if for all sequences $\boldsymbol{x} \to x_0$ with $f(\boldsymbol{x}) > 0$ we have

$$\tau(f, \boldsymbol{x}_0) := \liminf_{\boldsymbol{x} \to x_0} d(0, \partial f(\boldsymbol{x})) > 0. \tag{4.3.10}$$

Moreover $c = \tau(f, \boldsymbol{x}_0)^{-1}$ *is optimal in* (4.3.9).

One can recover Hoffman's inequality and compute the sharp constant by applying the above theorem with

$$f(\boldsymbol{x}) = \max_{i=1,\ldots,m} \boldsymbol{A}_i \cdot \boldsymbol{x} - \boldsymbol{b}_i$$

where $\boldsymbol{A}_i$ are the rows of $\boldsymbol{A}$.

We have described the tools necessary to prove that $F(\mu)$ and $W(\mu)$ are calm from two perspectives: the geoemtric and the algebraic. The geometric proof is quite simple, while the algebraic proof enables us to compute explicit constants. We will present both proofs for both F and W.

F is calm

We begin with the geometric proof. Note that by the scaling $\Omega(\lambda \boldsymbol{x}) = \lambda \Omega(\boldsymbol{x})$ our map

$$F(\mu) = \{\boldsymbol{x} \mid \Omega(\boldsymbol{x}) \leq \mu\} = \{\boldsymbol{x} \mid \Omega(\boldsymbol{x}/\mu) \leq 1\}$$

is simply a scaling of the set $\{\boldsymbol{x} \mid \Omega(\boldsymbol{x}) \leq 1\}$. By assumption, this set is a polyhedron so $F(\mu)$ is a polyhedron for all $\mu \in B_\delta(\mu_0)$ where δ small enough so that $0 \notin B_\delta(\mu_0)$. By definition of F, the graph of F restricted to $\boldsymbol{b}_\delta(\mu_0)$ is thus the polyhedron $F(\mu_0 + \delta)$. Then F is calm at $(\mu_0, \boldsymbol{x}_0)$ by Theorem 4.3.4.

Next we present the algebraic proof with more explicit constants. Note that by assumption $\mu_0 = \Omega(\boldsymbol{x}_0) > 0$. Choose $f(\boldsymbol{x}) = \Omega(\boldsymbol{y}) - \mu$. We know $0 \notin \partial\Omega(\boldsymbol{x}_0)$ for $\Omega(\boldsymbol{x}_0) > 0$ so to verify (4.3.10) it suffices to show that $d(0, \partial f(\boldsymbol{x}))$ is lower semicontinuous. In general, one can check from the definitions that for upper semicontinuous set-valued

mapping P and convex function g the function

$$Q(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in P(\boldsymbol{x})} g(\boldsymbol{y})$$

is lower semicontinuous as a real-valued function. Indeed, $d(0, \partial f(\boldsymbol{x}))$ is of this form with $P(\boldsymbol{x}) = \partial\Omega(\boldsymbol{x})$ and $g(\boldsymbol{y}) = \|\boldsymbol{y}\|_2$. Thus $d(0, \partial f(\boldsymbol{x}))$ is lower semicontinuous and we have a local error bound of the form (4.3.9). Calmness of F now follows by a simple argument. Without loss of generality, let $\mu_1 \geq \mu_0$. Let $\boldsymbol{x}_1 \in F(\mu_1)$. Since $\Omega(\boldsymbol{x}_1) \leq \mu_1$ we can then write

$$\Omega(\boldsymbol{x}_1) - \mu_0 \leq \Omega(\boldsymbol{x}_1) - \mu_0 - (\Omega(\boldsymbol{x}_1) - \mu_1) = \mu_1 - \mu_0.$$

We can now bound the right hand side of (4.3.9) with $\tau^{-1}(\mu_1 - \mu_0)$,

$$d(\boldsymbol{x}_1, W(\mu_0)) \leq \frac{1}{\tau} |\mu_1 - \mu_0|, \quad \text{for all } \boldsymbol{x}_1 \in B_\epsilon(\boldsymbol{x}_0) \cap W(\mu_1).$$

Thus F is calm. Note that in this proof we do not need the polyhedral level set assumption on $\Omega(\boldsymbol{x})$.

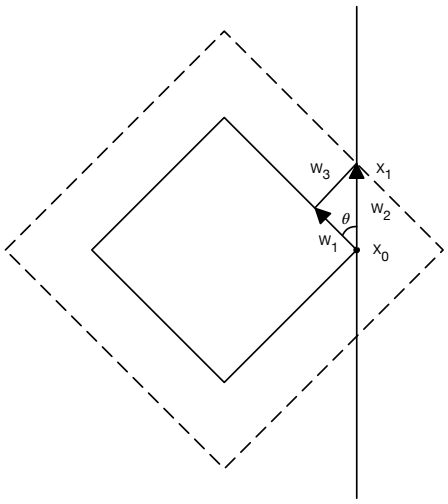
W is calm

Extending the polyhedral mapping proof from Section 4.3.2 is nontrivial because W does not obey a simple scaling law. Instead, we analyze carefully the intersection of the affine set $M(\boldsymbol{A}_0)$ and the polyhedral set $F(\mu_0)$ near $\boldsymbol{x}_0$. The calmness constant for W will be $\sin(\theta)\, c(F)$ where $c(F)$ is the calmness constant from Section 4.3.2 and θ is smallest nonzero principle angle between our affine space $M(\boldsymbol{A}_0)$ and the affine sets defining the faces of $F(\mu_0)$ containing $\boldsymbol{x}_0$.

By assumption $F(\mu_0)$ is a polyhedral set defined the intersection of $n-1$ dimensional

affine spaces. We can choose coordinates so that $\boldsymbol{x}_0 = 0$. Then the affine spaces $\mathcal{U}_1, \ldots, \mathcal{U}_k$ that contain $\boldsymbol{x}_0$ and define $F(\mu_0)$ are in fact subspaces of $\mathbb{R}^n$. Likewise for the set $M(\boldsymbol{A}_0)$. Let $\boldsymbol{w}_1 \in W(\mu_1) \cap B_\epsilon(\boldsymbol{x}_0)$ where ϵ chosen so that the projection $P_{F(\mu_0)}(\boldsymbol{w}_1)$ lies in at least one subspace $\mathcal{U}_i$. Let $\theta^i 0$ be the smallest principle angle (see Section 4.3.2) between this $\mathcal{U}_i$ and $M(\boldsymbol{A}_0)$. Define

$$\boldsymbol{w}_2 := P_{\mathcal{U}_i}(\boldsymbol{w}_1), \quad \boldsymbol{w}_3 := \boldsymbol{w}_1 - \boldsymbol{w}_2 = \boldsymbol{w}_1 - P_{\mathcal{U}_i}(\boldsymbol{w}_1).$$



Since $\boldsymbol{w}_3$ is orthogonal to $\mathcal{U}_i$, the vectors $\boldsymbol{w}_1, \boldsymbol{w}_2$ cannot be in a shared subspace of $\mathcal{U}_i$ and $M(\boldsymbol{A}_0)$. Thus $\boldsymbol{w}_1, \boldsymbol{w}_2$ are orthogonal to all principal directions corresponding to principal angles that are zero. Hence $\theta^i \leq \theta$ where θ is the angle between $\boldsymbol{w}_1, \boldsymbol{w}_2$. Since $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$ form a right triangle

$$\|\boldsymbol{w}_2\| \sin \theta^i \leq \|\boldsymbol{w}_2\| \sin \theta = \|\boldsymbol{w}_3\| .$$

Note that

$$\|\boldsymbol{w}_3\| = \|\boldsymbol{x}_1 - P_{\mathcal{U}_i}(\boldsymbol{x}_1)\| = d(\boldsymbol{x}_1, F(\mu_0)), \quad \|\boldsymbol{w}_2\| = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|$$

so we have the bound

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| \leq \frac{1}{\sin\theta^i} d(\boldsymbol{x}_1, F(\mu_0)). \tag{4.3.11}$$

when $\theta^i > 0$. If $\theta^i = 0$, then the proof reduces to the proof in Section 4.3.2 and we have the bound

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| \leq d(\boldsymbol{x}_1, F(\mu_0)). \tag{4.3.12}$$

Combining (4.3.11) and (4.3.12) we have

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| \leq \frac{2}{1 + \sin\theta^i} d(\boldsymbol{x}_1, F(\mu_0)).$$

From Section 4.3.2 we can bound $d(\boldsymbol{x}_1, F(\mu_0))$ giving finally

$$d(\boldsymbol{x}_1, W(\mu_0)) \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_0\| \leq \frac{2\tau}{1 + \sin\theta^i} \left|\mu_1 - \mu_0\right|,$$

which can be made independent of $\boldsymbol{x}_1$ by taking the min of $\theta_i, i = 1, \ldots, k$.

The algebraic approach is similar to Section 4.3.2. Define

$$f(\boldsymbol{x}) := \max(\Omega(\boldsymbol{x}) - \mu_0, d(\boldsymbol{x}, M(\boldsymbol{A}_0))).$$

Our goal is to verify Condition (4.3.10) and obtain a local error bound of the form (4.3.9) by Theorem 4.3.6. We can write $W(\mu_0)$ as either

$$W(\mu_0) = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \leq 0\}$$

or

$$W(\mu_0) = \arg\min_x f(\boldsymbol{x}).$$

A point $\boldsymbol{x}^*$ is then a solution (with optimal value $\Omega(\boldsymbol{x}) = \mu_0$) if and only if

$$0 \in \partial f(\boldsymbol{x}^*)$$

or equivalently if and only if

$$f(\boldsymbol{x}) \leq 0.$$

Thus $0 \in \partial f(\boldsymbol{x})$ if and only if $f(\boldsymbol{x}) \leq 0$.

It remains to verify that $d(0, \partial f(\boldsymbol{x}))$ is bounded *uniformly* away from zero for $\boldsymbol{x}$ such that $f(\boldsymbol{x}) > 0$. It suffices to show that $d(0, \partial f(\boldsymbol{x}))$ is piecewise constant. By standard properties of the subgradient [25]

$$\partial f(\boldsymbol{x}) = \text{CO } \{\partial\Omega(\boldsymbol{x}) \mid \Omega(\boldsymbol{x}) = f(\boldsymbol{x})\} \cup \{\partial d(\boldsymbol{x}, M(\boldsymbol{A}_0)) \mid d(\boldsymbol{x}, M(\boldsymbol{A}_0)) = f(\boldsymbol{x})\}, \tag{4.3.13}$$

where CO denotes the convex hull. By the fact that $\Omega(\boldsymbol{x})$ has polyhedral level sets, we know $\partial\Omega(\boldsymbol{x})$ is piecewise constant in the sense that $\partial\Omega(\boldsymbol{x})$ outputs only finitely many different sets. Similarly, since $M(\boldsymbol{A}_0)$ is also a polyhedron, $\partial d(\boldsymbol{x}, M(\boldsymbol{A}_0))$ is piecewise constant. (4.3.13) thus shows that $\partial f(\boldsymbol{x})$ is piecewise constant. Hence $d(0, \partial f(\boldsymbol{x}))$ is piecewise constant as a real-valued function.

We thus have a local error bound of the form (4.3.9) with

$$\tau = \liminf_{\boldsymbol{x} \to x_0, f(\boldsymbol{x}) > 0} d(0, \partial f(\boldsymbol{x})).$$

Calmness of W now follows by a simple argument. Without loss of generality, let

$\mu_1 \geq \mu_0$. Let $\boldsymbol{x}_1 \in W(\mu_1)$. Thus

$$f(\boldsymbol{x}_1) = \max(\Omega(\boldsymbol{x}_1) - \mu_0, d(\boldsymbol{x}_1, M(\boldsymbol{A}_0))) = \Omega(\boldsymbol{x}_1) - \mu_0$$

since $\boldsymbol{x}_1 \in M(\boldsymbol{A}_0)$. Since $\Omega(\boldsymbol{x}_1) \leq \mu_1$ we can then write

$$\Omega(\boldsymbol{x}_1) - \mu_0 \leq \Omega(\boldsymbol{x}_1) - \mu_0 - (\Omega(\boldsymbol{x}_1) - \mu_1) = \mu_1 - \mu_0$$

We can now bound the right hand side of (4.3.9) with $\tau^{-1}(\mu_1 - \mu_0)$,

$$d(\boldsymbol{x}_1, W(\mu_0)) \leq \frac{1}{\tau} |\mu_1 - \mu_0|, \quad \text{for all } \boldsymbol{x}_1 \in B_\epsilon(\boldsymbol{x}_0) \cap W(\mu_1).$$

Thus F is calm.

4.4 Conclusion

We have shown that if Ω has polyhedral level sets, then

$$S(\boldsymbol{A}) = \underset{\boldsymbol{Ax}=\boldsymbol{b}}{\arg\min}\, \Omega(\boldsymbol{x})$$

is calm. By tracing the constants in our proof and in Klatte and Kumar [51], we see that a calmness constant for S is

$$\max\left(\left\|\boldsymbol{A}_0^\dagger\right\|, \frac{1}{\tau}\right)\left(1 + \frac{2C_\Omega}{\tau \sin(\theta)}\right) C_\Omega \left\|\boldsymbol{A}_0^\dagger\right\|$$

where C_Ω is the Lipschitz constant of Ω from Section 4.3.2, τ is the constant from Section 4.3.2, and θ is the smallest angle defined in Section 4.3.2.

A key assumption for proving that W is calm is that Ω has polyhedral sublevel sets.

Removing this assumption would allow us to recover known calmness results such as the case $\Omega(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$ [58]. Unfortunately, the intersection W is in fact not calm for $\Omega(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$ as we show below. Therefore our approach needs major change if one wants to extend our result to the full class of seminorms.

We verify our claim above by using the following equivalent definition of calmness is as follows. A map $F : Y \to X$ is calm at $(\boldsymbol{y}_0, \boldsymbol{x}_0)$ if and only if there exists a neighborhood U of $\boldsymbol{x}_0$ and a constant $c(\boldsymbol{x}_0) > 0$ such that

$$d(\boldsymbol{x}, F(\boldsymbol{y}_0)) \leq c\, d(\boldsymbol{y}_0, F^{-1}(\boldsymbol{x}))$$

for all $\boldsymbol{x} \in U$.

Using this definition, we show W is not calm at $(1, (0, -1))$. Since $\Omega(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$, $\mu_0 = 1$ corresponds to the unit disk. Choose $\boldsymbol{A}_0 = (0, -1)$ and $\boldsymbol{b} = 1$ to give a line L passing through $(0, -1)$. Note that

$$d(\boldsymbol{y}_0, F^{-1}(\boldsymbol{x})) = \|\boldsymbol{x}\| - \mu_0 = \sqrt{\boldsymbol{x}_1^2 + \boldsymbol{x}_2^2} - \mu_0.$$

Consequently,

$$\lim_{\boldsymbol{x}_1 \to 0} \frac{d(\boldsymbol{y}_0, F^{-1}(\boldsymbol{x})}{d(\boldsymbol{x}, F(\boldsymbol{y}_0))} = \lim_{\boldsymbol{x}_1 \to 0} \frac{\sqrt{\boldsymbol{x}_1^2 + 1} - 1}{\boldsymbol{x}_1} = 0$$

and this contradicts the assumption that $c(\boldsymbol{x}_0) > 0$.

Bibliography

[1] K. Choi, T. Y. Hou, A. Kiselev, G. Luo, V. Sverak, and Y. Yao, "On the Finite Time Blowup of a 1D Model for the 3D Axisymmetric Euler Equations," *arXiv:1407.4776 [math]*, 2014. [Online]. Available: http://arxiv.org/abs/1407.4776.

[2] H. Okamoto, T. Sakajo, and M. Wunsch, "On a generalization of the Constantin-Lax-Majda equation," *Nonlinearity*, vol. 21, no. 10, p. 2447, 2008.

[3] P Constantin, P. D. Lax, and A Majda, "A simple one-dimensional model for the three-dimensional vorticity equation," *Communications on Pure and Applied Mathematics*, vol. 38, no. 6, pp. 715–724, 1985, ISSN: 00103640, 10970312. DOI: 10.1002/cpa.3160380605. [Online]. Available: http://doi.wiley.com/10.1002/cpa.3160380605.

[4] S. De Gregorio, "On a one-dimensional model for the three-dimensional vorticity equation," *Journal of Statistical Physics*, vol. 59, no. 5-6, pp. 1251–1263, 1990, ISSN: 0022-4715, 1572-9613. DOI: 10.1007/BF01334750. [Online]. Available: http://link.springer.com/10.1007/BF01334750.

[5] M. Bauer, B. Kolev, and S. C. Preston, "Geometric investigations of a vorticity model equation," *arXiv:1504.08029 [math]*, 2015. [Online]. Available: http://arxiv.org/abs/1504.08029.

[6] H. Jia, S. Stewart, and V. Sverak, "On the De Gregorio Modification of the Constantin–Lax–Majda Model," *Archive for Rational Mechanics and Analysis*, vol. 231, no. 2, pp. 1269–1304, 2019, ISSN: 14320673. DOI: 10.1007/s00205-018-1298-1. arXiv: arXiv:1710.02737v1.

[7] Z. Lei, J. Liu, and X. Ren, "On the Constantin-Lax-Majda Model with Convection," 2018. arXiv: 1811.09754. [Online]. Available: http://arxiv.org/abs/1811.09754.

[8] E Kowalski, "Spectral theory in Hilbert spaces," *ETH Zurich*, p. 129, 2009. [Online]. Available: https://people.math.ethz.ch/{~}kowalski/spectral-theory.pdf.

[9] R. Maier, "The 192 solutions of the {Heun} equation," *Mathematics of Computation*, vol. 76, no. 258, pp. 811–843, 2007.

[10] E. M. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2009.

[11] K.-J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations*. Springer Science & Business Media, 1999, vol. 194.

[12] T. Kato and G. Ponce, "Commutator estimates and the euler and navier-stokes equations," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 891–907, 1988, ISSN: 00103640. DOI: 10.1002/cpa.3160410704. [Online]. Available: http://doi.wiley.com/10.1002/cpa.3160410704.

[13] D. Dutykh, "A brief introduction to pseudo-spectral methods: Application to diffusion problems," *arXiv preprint arXiv:1606.05432*, 2016.

[14] N. HITCHIN, "Vector fields on the circle," in *Mechanics, analysis and geometry: 200 years after Lagrange*, Elsevier, 1991, pp. 359–378.

[15] C. Reynolds, "Flocks, Herds, and Schools: A Distributed Behavioral Model Craig," *ACM Computer Graphics*, vol. 21, no. 4, pp. 25–34, 1987, ISSN: 11236337.

[16] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995, ISSN: 1063651X. DOI: 10.1103/PhysRevE.51.4282. arXiv: 9805244v1 [arXiv:cond-mat].

[17] R. L. Hughes, "The Flow of Human Crowds," *Annual Review of Fluid Mechanics*, vol. 35, no. 1, pp. 169–182, 2003, ISSN: 0066-4189. DOI: 10.1146/annurev.fluid.35.101101.161136. [Online]. Available: http://www.annualreviews.org/doi/10.1146/annurev.fluid.35.101101.161136.

[18] I. Karamouzas, B. Skinner, and S. J. Guy, "Universal power law governing pedestrian interactions," *Physical Review Letters*, vol. 113, no. 23, 2014, ISSN: 10797114. DOI: 10.1103/PhysRevLett.113.238701. arXiv: arXiv:1412.1082v1.

[19] J. van den Berg, M. Lin, and D. Manocha, "Reciprocal Velocity Obstacles for Real-Time Multi-Agent Collision Avoidance," *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 1928–1935, 2007.

[20] S. Faure and B. Maury, "Crowd motion from the granular standpoint," *Mathematical Models and Methods in Applied Sciences*, vol. 25, no. 03, pp. 463–493, 2015.

[21] J. L. Silverberg, M. Bierbaum, J. P. Sethna, and I. Cohen, "Collective Motion of Moshers at Heavy Metal Concerts," 2013. DOI: 10.1103/PhysRevLett.110.228701. arXiv: 1302.1886.

[22] A. Lachapelle and M. T. Wolfram, "On a mean field game approach modeling congestion and aversion in pedestrian crowds," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1572–1589, 2011, ISSN: 01912615. DOI: 10.1016/j.trb.2011.07.011.

[23] R. Narain, A. Golas, S. Curtis, and M. C. Lin, "Aggregate dynamics for dense crowd simulation," *ACM Transactions on Graphics*, vol. 28, no. 5, p. 1, 2009, ISSN: 07300301. DOI: 10.1145/1618452.1618468.

[24] E. Cristiani, B. Piccoli, and A. Tosin, "Multiscale Modeling of Granular Flows with Application to Crowd Dynamics," vol. 9, no. 1, pp. 155–182, 2011.

[25] R. T. Rockafellar, *Convex analysis*, 28. Princeton university press, 1970.

[26] F. H. Clarke, "Generalized gradients and applications," *Transactions of the American Mathematical Society*, vol. 205, pp. 247–262, 1975.

[27] B. Maury and J. Venel, "A discrete contact model for crowd motion," *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 45, no. 1, pp. 145–168, 2010, ISSN: 0764-583X. DOI: 10.1051/m2an/2010035.

[28] J. J. Moreau, "Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires," 1962.

[29] J. F. Edmond and L. Thibault, "Relaxation of an optimal control problem involving a perturbed sweeping process," *Mathematical programming*, vol. 104, no. 2-3, pp. 347–373, 2005.

[30] R. A. Poliquin and R. T. Rockafellar, "Prox-regular functions in variational analysis," *Transactions of the American Mathematical Society*, vol. 348, no. 5, pp. 1805–1838, 1996, ISSN: 0002-9947, 1088-6850. DOI: 10.1090/S0002-9947-96-01544-9.

[31] F. Clarke, *Nonsmooth Analysis in Systems and Control Theory.* 2009.

[32] R. A. Poliquin and R. T. Rockafellar, "Amenable functions in optimization," *Nonsmooth optimization: methods and applications ({E}rice, 1991)*, pp. 338–353, 1992.

[33] F. H. Clarke, *Optimization and nonsmooth analysis*. Siam, 1990, vol. 5.

[34] B. Maury, "Version continue de l'algorithme d' Uzawa Continuous version of the Uzawa algorithm," no. September, 2005.

[35] C. Chatfield, *A Simple Method for Distance to Ellipse*, 2017. [Online]. Available: https://wet-robots.ghost.io/simple-method-for-distance-to-ellipse/.

[36] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003, ISSN: 0027-8424. DOI: 10.1073/pnas.0437847100. [Online]. Available: https://doi-org.ezp1.lib.umn.edu/10.1073/pnas.0437847100.

[37] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006, ISSN: 1557-9654. DOI: 10.1109/TIT.2005.862083.

[38] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006, ISSN: 00189448. DOI: 10.1109/TIT.2006.885507. arXiv: 0410542v3 [arXiv:math].

[39] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006. DOI: 10.1002/cpa.20132.

[40] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[41] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE transactions on information theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[42] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[43] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006, ISSN: 1557-9654. DOI: 10.1109/TIT.2005.860430.

[44] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006, ISSN: 1557-9654. DOI: 10.1109/TIT.2005.864420.

[45] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE Journal on Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010, ISSN: 19324553. DOI: 10.1109/JSTSP.2009.2039170. arXiv: 0907.2955.

[46] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE transactions on signal processing*, vol. 57, no. 6, pp. 2275–2284, 2009.

[47] A. C. Fannjiang, T. Strohmer, and P. Yan, "Compressed remote sensing of sparse objects," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 595–618, 2010.

[48] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 655–687, 2008.

[49] T. Blumensath and M. Davies, "Compressed sensing and source separation," in *International Conference on Independent Component Analysis and Signal Separation*, Springer, 2007, pp. 341–348.

[50] A. Gutierrez, "Reducing the Complexity of Model-Based MRI Reconstructions via Sparsification," 2019.

[51] D. Klatte and B. Kummer, "On Calmness of the Argmin Mapping in Parametric Optimization Problems," *Journal of Optimization Theory and Applications*, vol. 165, no. 3, pp. 708–719, 2015, ISSN: 15732878. DOI: 10.1007/s10957-014-0643-2.

[52] D. Klatte and K. Bernd, *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods, and Applications*. Springer, 2006, vol. 60, ISBN: 9788578110796. DOI: 10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3.

[53] S. L. Campbell and C. D. Meyer, *Generalized inverses of linear transformations*. SIAM, 2009.

[54] A. Jourani, "Hoffman's error bound, local controllability, and sensitivity analysis," *SIAM Journal on Control and Optimization*, vol. 38, no. 3, pp. 947–970, 2000.

[55] P. Å. Wedin, "On angles between subspaces of a finite dimensional inner product space," pp. 263–285, 1983. DOI: 10.1007/bfb0062107.

[56] A. Hoffman, "On approximate solutions of systems of linear inequalities," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 4, p. 263, 1952, ISSN: 0091-0635. DOI: 10.6028/jres.049.027.

[57] M. T. Ngai, Huynh, Alexander Kruger, "Stability of error bounds for semi-infinite convex constraint systems," vol. 20, no. 4, pp. 2080–2096, 2010.

[58] J. Ding, "Perturbation analysis for the projection of a point to an affine set," *Linear Algebra and Its Applications*, vol. 191, no. C, pp. 199–212, 1993, ISSN: 00243795. DOI: 10.1016/0024-3795(93)90515-P.