The Hilbert Projective Metric, Multi-type Branching Processes and Mathematical Biology: a Model of the Evolution of Resistance

by Christopher Anthony McCarthy

A dissertation submitted to the Graduate Faculty in Mathematics in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York
2010

This manuscript has been read and accepted for the Graduate Faculty in Mathematics in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

Date	Yunping Jiang, Chair of Examining Committee
Date	Frederick P. Gardiner, Examining Committee
Date	Linda Keen, Examining Committee
Date	Roman Kossak, Executive Officer

Supervisory Committee

Yunping Jiang

Frederick P. Gardiner

Linda Keen

Abstract

The Hilbert Projective Metric, Multi-type Branching Processes and Mathematical Biology: a Model of the Evolution of Resistance

by Chris McCarthy

Adviser: Professor Yunping Jiang

Bacteria, viruses, or cancer cells, by means of mutation and replication, are sometimes able to escape the selective pressure exerted by treatment. This is called the development, or evolution, of resistance.

This dissertation is a study of some of the mathematics underlying a model of resistance put forth by Iwasa, Michor, and Nowak (IMN) [48, 49] (2003, 2004).

In the IMN model the pre-treatment phase is modeled as a determinist dynamical system using Eigen and Schuster's quasispecies theory of evolution [29]. It is assumed that at the start of treatment the system has reached an invariant distribution: the quasispecies equilibrium eigenvector.

The equations of the quasispecies theory can be viewed as projections of linear differential equations onto hyperplanes and their asymptotic behavior can be understood via Birkhoff's Projective Contraction Theorem [12], which is related to the Perron-Frobenius Theorem. An understanding of Birkhoff's contraction theorem requires an understanding of the Hilbert Projective Metric and so we develop an extensive collection of useful related results, some novel, about cones, hyperplanes, and the Hilbert Projective Metric.

In the IMN model, the post-treatment phase is modeled as a stochastic multi-type branching process on the various mutant types. The key calculation is the vector of extinction probabilities: the i^{th} entry of the vector being the probability that a process, starting with a single mutant of type i, will eventually go extinct (under the selective pressure of treatment). The techniques for calculating these extinction

probabilities involve the use of multi-type probability generating functions (PGF's).

We prove results about the existence of continuous multi-type PGF's and branching processes. Our proofs involve customizing techniques from the theory of differential equations in complex vector spaces, and then applying results from the theory of several complex variables. We also develop a method to numerically calculate the vector of extinction probabilities.

The pre and post-treatment models are fitted together and the probability of a successful treatment is numerically calculated using a combination of standard techniques from numerical analysis together with insights gained from our examination of the mathematical aspects of the model. Our investigation leads to a phenomena somewhat reminiscent of Eigen's error catastrophe theory.

Acknowledgements

As I finish this work I wish to acknowledge those who contributed to my arriving at this moment. I apologize for any oversights; the chaotic ordering, and my difficulties in translating heartfelt gratitude into plain words.

I would like to thank Yunping Jiang, my advisor, for his many suggestions, patience, and constant support during this research.

I wish to thank my mom and dad for teaching me to think about all sides of every issue and for always being there when I needed them.

I want to thank my dear Eritka for her encouragements, support, caring and culture and more. Taking her advice helped me reach this moment.

I wish to thank my brother Michael, for being a good older brother and not doing away with me on multiple, justifiable occasions; Debbie, and of course, Matthew and Mark who are beyond dear to me; my Aunt Eunice (and Uncle Bud, in memoriam) for their love and kindness, and for the good impact that they have had on my life, more than they may imagine. The same is true, of course, for my cousins Vito (in memoriam), Benny and Claire; Mariette, Nicky, Frank and Carol; Frank Rossi, Norma, Carla, and Randy. Likewise, I wish I could thank my grandparents Eunice Liner and Michelino Rossi in person, but who are now gone; and my grandmother in heaven, Libera Rossi, who I have never met on this earth, but who has influenced me through stories told to me by my mother. A second thanks is extended to my father,

for applying his professional newspaper editor skills to my dissertation (but not these acknowledgements).

I wish to thank Rosa, Rosita, Maria, Natalia and Gideon, for all their varied kindnesses and Netzer, Nitzan, and Noga, for teaching me magical things and capturing my heart.

I would like to thank Professor Józef Dodziuk for having confidence in me and for magically disappearing administrative problems; Robert Landsman, for his many kindnesses; Professor Irene Hueter for being my advisor at the start of this project; the members of my dissertation and-or orals committee, Professors Frederick P. Gardiner, Linda Keen, and, Sandra Hayes, for their graciousness, insightful questions, suggestions and correspondences; the overnight crew at the CUNY Graduate Center, for their patience and good cheer, especially Eyda Balarezo and Mr. Persaud.

I am grateful to Professor Johannes Familton for his deep friendship of many, many years, for his help with mathematics, physics, and chemistry; to Professors Steve Gottlieb, Mkjuma, and little Joyce (who is not a professor yet) for their wonderful friendship and help with mathematics, biology, and computers; to Professors Fred Peskoff and Leonid Khazanov for their friendships, caring, and humor; to my colleagues at BMCC for their support.

I am grateful for my friends throughout the years: Vincent Alliegro and my classmates from PS32; Mitchell Wachtel; Mayer Landau, Ray Thomas and the math students and faculty of Queens Collge of CUNY; and Mabel, Jerry, Tyler and Logan, Charo and little Joe.

Then, there are all my furry friends, who make the world a better place, especially Elky, Wolf, Otto, Fox, Molly, Precious (Boy cat), Ginger (Girl cat), Samantha, Minnie, Sheba, and many others.

My closest friend (for more than half my life) Hayley Greenberg, who taught me all about cats and why I should be a vegetarian, is the inspiration for this dissertation.

New York City, New York Chris McCarthy

September 21, 2010

Contents

List of Figures

Summary of Part I

If viewing this manuscript as a hyperlinked PDF in Adobe Reader see[1].

Chapters 1 and 2 are a collection of results (some original) regarding the Hilbert Projective Metric on Cones, denoted d_H, Birkhoff's Projective Contraction Theorem, and applications of these to geometry, linear maps and differential equations.

C will denote a convex, closed, pointed by the origin, salient cone contained in a Banach Space V. P will denote a linear map of V to itself which also maps C to itself.

The more interesting parts of this work are found in the following sections:

- **Section 2.2 (page 150)**

  Starts by discussing the problem possed to Birkhoff's Projective Contraction Theorem by a non-trivial kernel. This leads to the geometric question of where $\ker P^n$ can be located in the cone C. The answer is that either

  $$C \cap \ker P^n = C \quad \text{or} \quad C \cap \ker P^n \subset \partial C$$

  I have not seen that result elsewhere. The details are contained in the following theorems and corollaries:

[1]It useful to to set navigational preferences to allow returning to previously viewed pages. To do this, go to: Menu, Tools, Customize Tool Bars, Page Navigation Tool Bar, then select ☑ Previous View and ☑ Next View.

Theorem 2.2.3.3 (page 152). *Let P be as usual, a linear map of C to itself, where C is a pointed by the origin, salient, closed, convex cone in a Banach Space V. Let $f, g \in C \setminus \{0\}$ and let n be any positive integer. If $gP^n = 0$ but $fP^n \neq 0$ then $d_H(f, g) = \infty$ and $g \in \partial\left(\mathrm{Span}(f, g) \bigcap C\right) \subset \partial C$.*

Corollary 2.2.3.5 (page 155). *Let n be a positive integer. If $C \cap \ker P^n \neq C$ then $C \cap \ker P^n \subset \partial C$.*

Corollary 2.2.3.6 (page 155). *If there exists an $f \in C \setminus \{0\}$ such that $fP^n \neq 0$ for all integers $n > 0$ then $C \cap (\cup_{n=1}^{\infty} \ker P^n) \subset \partial C$.*

- **Section 2.4.2 (page 186)**

  Discusses the existence of a hyperplane which will intersect each 0-ray [2] in a cone exactly once. Birkhoff in [12] seems to require the existence of such hyperplanes, which results in a loss generality, as such hyperplanes do not exist in the general case. I have not seen Part 3 of the following result elsewhere.

Theorem 2.4.4.2 (page 191). *Suppose that the convex cone C is a subset of V, an arbitrary vector space of finite or infinite dimension, and that C contains at least one non zero vector. Then the following are equivalent.*

1. *There exists a base [3] $\mathcal{B}$ for the cone C.*
2. *There exists a linear functional on V which is strictly positive on $C \setminus \{0\}$.*
3. *There exists a hyperplane H such that H intersects each 0-ray in $C \setminus \{0\}$ exactly once.*

[2] If C is cone then a 0-ray in C is an open ray in C which originates at the origin. Explicitly, if $f \in C$ then the 0-ray $[f] = \{\lambda f : \lambda > 0\}$.

[3] $\mathcal{B}$ is a base for the cone C if $\mathcal{B}$ is a convex subset of $C \setminus \{0\}$ and if for each $x \in C \setminus \{0\}$ there exists a unique $b \in B$ and a unique $\lambda > 0$ such that $x = \lambda b$.

- **Section 2.5 (page 232)**

  Has Lemma 2.5.0.2, below, which is original, as far as I know. It is allows a comparison of $d_H(f,g)$ to $d_V(f^H, g^H)$ when there exists a hyperplane H which intersects each equivalence class [4] of $(C \setminus \{0\}, \sim)$ exactly once. Lemma 2.5.0.2 allows us to concretely compare convergence under d_H with convergence in the hyperplane H under d_V. Lemma 2.5.0.2 also has a useful application to the theory of differential equations.

  Lemma 2.5.0.2 (page 232). *Brief version for Reader's Note: Let H be a hyperplane which intersects each equivalence class of $(C \setminus \{0\},\ \sim)$ exactly once. Let $f, g \in C$ be linearly independent and*

$$b_0 = f^H + t_{min}\left(g^H - f^H\right)$$
$$b_1 = f^H + t_{max}\left(g^H - f^H\right).$$

  Then $b_0, b_1 \in H \cap C \cap Span(f,g) \setminus \{0\}$. If $f', g' \in Span(f,g) \cap C$ are linearly independent then

$$d_V(f'^H, g'^H) \ < \ \frac{d_V(b_0,b_1)}{4}\ d_H(f',g')$$

- **Section 2.6.1 (page 249)**

  Discusses the continuity of the linear map P w.r.t. d_H.

- **Section 2.6.8 (page 270)**

  Discusses an appropriate interpretation of

$$\|fP^n - c\| \ < \ K\ \rho^n$$

[4] Let $f, f' \in C$. We say $f \sim f'$ if $f = \lambda f'$ for some $\lambda > 0$. The equivalence class of f w.r.t. $\sim$ is denoted $[f]$. If H is a hyperplane which intersects each equivalence class of $(C \setminus \{0\}, \sim) = \{[f] : f \in C \setminus \{0\}\}$ exactly once, we define f^H by $f^H = [f] \cap H$.

from Birkhoff's Proof of his Projective Contraction Theorem [12].

- **Section 2.7 (page 273)**

  Culminates in a theorem (Theorem 2.7.3.1 (page 283)) regarding the asymptotic behavior of the projections of certain types of linear ODE's. The proof I give seems to be entirely original. The two main lemmas contained in this section are also of interest.

Lemma 2.7.1.1 (page 273) Projective Additivity Lemma. *Let $x, y, z \in C$ and $d_H(x, z)$ and $d_H(y, z)$ both be finite. Then*

$$d_H(x + y, z) \leq \max\{d_H(x, z),\ d_H(y, z)\}.$$

See Figure 2.9 (page 274).

Lemma 2.7.2.1 (page 277) Poisson Tail Lemma. *Let A be a primitive $n \times n$ non negative matrix; i.e. the entries of A are all non negative and the entries of A^q are all positive for some integer $q > 0$. Let $Y \neq 0$ be a non negative column vector of dimension n and let m be any integer ≥ 0. Then*

$$\lim_{t\to\infty} \frac{\sum_{k=0}^{m} \frac{t^k}{k!}\ A^k Y}{\left\|e^{tA}Y\right\|_1} = 0$$

and

$$\lim_{t\to\infty} \frac{\left\|\sum_{k=m+1}^{\infty} \frac{t^k}{k!}\ A^k Y\right\|_1}{\left\|e^{tA}Y\right\|_1} = 1$$

Theorem 2.7.3.1 (page 283) Projective Linear ODE Theorem. *Let A be a primitive $n \times n$ matrix; i.e. the entries of A are non negative and the*

entries of A^Q are positive for some integer $Q > 0$. Let

$$\dot{X} = AX$$

be a system of linear ODE's with initial condition $X(0) = X_0$, with $X(0) \neq 0$ and non negative. Let v_p be the unique eigenvector of A with all positive entries and l^1 norm 1. Then

$$\lim_{t\to\infty} \frac{X}{\|X\|_1} = v_p,$$

where convergence of this limit is w.r.t. to the Euclidean Metric d_E.

- **Section 2.8 (page 286)**

  Has a nice geometric result (which I thought to be original, until recently) regarding 'd_H circles' in the two simplex $\triangle^2$.

Theorem 2.8.0.2 (page 288) CIRCUMFERENCE = $6r$ THEOREM. *The Hilbert circle $S^1_H(f, r)$ (which looks like a Hexagon in Euclidean Space) has circumference $6r$ w.r.t. d_H.*

Part I

The Hilbert Metric

Introduction to Part I

The formula for the Hilbert Metric, as well as its basic implications for hyperbolic geometry, were not discovered by David Hilbert.

According to John Milnor [70],

> Klein [1871] reinterpreted Beltrami's projective disk model in terms of projective geometry. Following Cayley [1859], he took as his starting point the expression
>
> $$\frac{1}{2}\log\frac{|q-a|\ |b-p|}{|p-a|\ |b-q|} \tag{1}$$
>
> for the non-euclidean distance between two points p, q, as illustrated in Figure 1. (The factor 1/2 is inserted so that curvature will be -1.) Here $|q-a|$ denotes the euclidean distance from a to q. In this paper he introduced the term *hyperbolic geometry* for the non-euclidean geometry of Lobachevsky and Bolyai. [5] , [6] , [7]

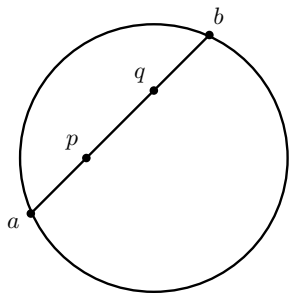


Figure 1: The Cayley formula for non-euclidean distance, in the projective disk model. From [70].

According to Oliver Bletz-Siebert and Thomas Foertsch [15],

> In the late 19th century D. Hilbert informed F. Klein in a letter about

[5] It is worth noting that Hilbert (b. 1862) was 9 years old in 1871.

[6] For Klein [1871] see [59]. Also see Klein's identically titled 1873 essay [60]. For Cayley [1859] see [22].

[7] For an accessible treatment of Cayley [1859] and Klein [1871] see Morris Kline [61]. For an excellent discussion of the connections between the Klein, Hilbert and Poincaré metrics see Beardon [9].

the fact that he had discovered a method to construct metric spaces, which somehow generalizes Klein's model of the real hyperbolic space ([47]).

Hilbert's letter [47] [8] to Klein, published in 1895, showed that the distance formula developed by Cayley and Klein (1) is valid for all bounded convex subsets of $\mathbb{R}^n$. Hilbert dispensed with the scaling factor and wrote the distance formula (1) as

$$\widehat{AB} = l\left\{\frac{\overline{YA}}{\overline{YB}} \cdot \frac{\overline{XB}}{\overline{XA}}\right\} \tag{2}$$

where l stands for logarithm; X, A, B, Y are as suggested by Figure 2; and $\overline{YA}, \overline{YB}, \overline{XB}$, $\overline{XA}$ are the euclidean distances. This formula is now called the Hilbert Metric or the Hilbert Projective Metric, as are its generalizations.

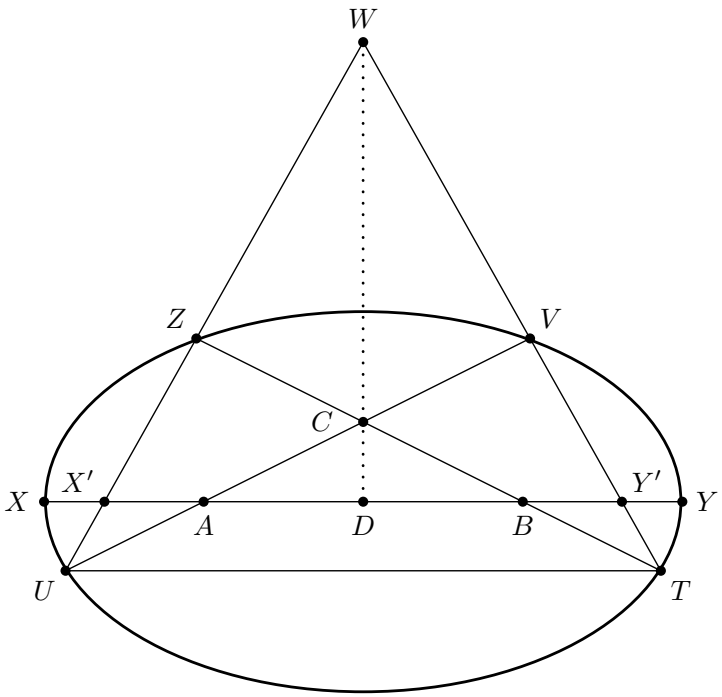


Figure 2: Hilbert used this illustration to show that Klein's metric could be extended to arbitrary convex sets in $\mathbb{R}^n$. The ellipse represents an arbitrary bounded convex set in the plane. From [47].

[8] David Hilbert, *Über die gerade Linie als kürzeste Verbindung zweier Punkte. (Aus einem an Herrn F. Klein gerichten Briefe.)*, Math. Ann. 46 (1895), 91-96. Which can be translated as: *About the straight line as the shortest connection between two points. (From a letter addressed to Mr. Klein.)*

The fractions appearing in the distance formulas (1) and (2) form a projective invariant known as a cross ratio. As Klein, in his 1871 paper [59], wrote [9],

> The logarithm of this cross-ratio multiplied by an arbitrary, but fixed, constant c, is what I call the distance between the two points.

The cross ratio can be defined as follows. Suppose that p_1, p_2, p_3, p_4 are four collinear points living in euclidean space. Their cross ratio is

$$(p_1, p_2; p_3, p_4) = \frac{\Delta_{1,4}}{\Delta_{1,3}} \frac{\Delta_{2,3}}{\Delta_{2,4}} \tag{3}$$

where $\Delta_{i,j}$ is the euclidean distance between p_i and p_j. Using the cross ratio (3) we can re-write the Hilbert Metric (2) as $\widehat{AB} = l(X, Y; A, B)$. See Figure 3.

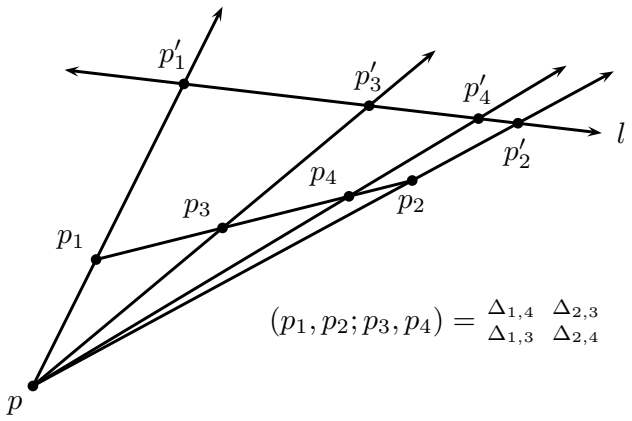


Figure 3: The projective invariance of the cross ratio (Pappus). The projection point is p. The collinear points p_1, p_2, p_3, p_4 are projected onto the line l yielding the points p'_1, p'_2, p'_3, p'_4. In the 4th century B.C. Pappus gave a proof that the cross ratio of four collinear points is invariant with respect to projection [66], [26]; i.e. $(p_1, p_2;\ p_3, p_4) = (p'_1, p'_2;\ p'_3, p'_4)$. More modern proofs of the cross ratio's projective invariance can be found in Milne [69] or Klein [58].

In that letter [47] to Klein, Hilbert used the projective invariance of the cross ratio to prove that the distance formula (2) is a metric and to prove the following result: Suppose the bounded convex set to which the Hilbert Metric (2) is applied

[9] For the English translation of several fundamental papers in hyperbolic geometry, including works by Klein [59], Beltrami, and Poincare, see [86].

is a euclidean triangle, such as TUW, see Figure 2. Then, inside TUW there is a non-degenerate triangle (such as ABC) for which the sum of two sides is equal to the third [10].

Hence, the Hilbert Metric induces a new geometry on euclidean triangles: one in which there are pairs of points for which the shortest connection is not unique. This is quite different than euclidean geometry where the shortest path between two points is always uniquely the straight line connection.

In footnote [11] I give a brief sketch of how Hilbert proved this result and how he proved that the distance formula (2) is a metric.

The utility of the Hilbert Metric extends beyond models of hyperbolic geometry. As P.J.Bushell [19] wrote in 1973,

> The usefulness of Hilbert's metric in algebra and analysis was made clear by Garrett Birkhoff [12] in 1957. Birkhoff showed that the Perron-Frobenius theorem for non-negative matrices and Jentzch's theorem for

[10] "*... und es gibt dann stets Dreiecke, für welche die Summe zweier Seiten gleich der dritten ist.*" Which can be translated as, "... and then there always exist triangles for which the sum of two sides is equal to the third." Hilbert [47].

[11] Adapted from [47]. **Proposition A**. *See Figure 2. In the triangle TUW,* $\widehat{AC} + \widehat{CB} = \widehat{AB}$. **Sketch of Proof.** By projective invariance $\widehat{AC} = l(U,V;A,C) = l(X',Y';A,D)$ and $\widehat{CB} = l(Z,T;C,B) = l(X',Y';D,B)$. For collinear points arranged as in Figure 2: $(X',Y';A,D)\cdot(X',Y';D,B) = \frac{\overline{Y'A}}{\overline{Y'D}}\frac{\overline{X'D}}{\overline{X'A}}\cdot\frac{\overline{Y'D}}{\overline{Y'B}}\frac{\overline{X'B}}{\overline{X'D}} = \frac{\overline{Y'A}}{\overline{Y'B}}\frac{\overline{X'B}}{\overline{X'A}} = (X',Y';A,B)$ and so $\widehat{AC}+\widehat{CB} = l(X',Y';A,D)+l(X',Y';D,B) = l((X',Y';A,D)(X',Y';D,B)) = l(X',Y';A,B) = \widehat{AB}$. □

Adapted from [47]. **Proposition B.** *The Hilbert Metric formula* $\widehat{AB}$ *is in fact a metric.*
Sketch of Proof. See Figure 2. $\widehat{AB} \geq 0$ and $\widehat{AB} = 0 \Leftrightarrow A = B$ both follow from $\overline{XB} \geq \overline{XA}$ and $\overline{YA} \geq \overline{YB}$. The symmetry $\widehat{AB} = \widehat{BA}$ follows from the symmetry of the euclidean metric. The only difficult part is the triangle inequality $\widehat{AB} \leq \widehat{AC} + \widehat{CB}$. If A, B, C are collinear, an easy algebraic argument similar to one given in Proposition A's proof suffices. If A, B, C are non collinear, we note that three non collinear points determine a unique hyperplane in $\mathbb{R}^n$ and so it suffices to prove the triangle inequality in bounded convex planar sets. The lines AC and BC will intersect at C and form an X shape as shown in Figure 2. Assuming that UZ and TV are not parallel we can form the triangle TUW as shown in Figure 2. Then, as in the proof of Proposition A, we use W as the projection point (or if UZ is parallel to TV we use the lines parallel to UZ, TV to guide the projection) and we project the line segments AC and BC onto the line segment AB; and just as in Proposition A's proof: $\widehat{AB}_{TUW} = \widehat{AC} + \widehat{CB}$. We then note that if $0 < s \leq t$ and $h \geq 0$ then $1 \leq \frac{t+h}{s+h} \leq \frac{t}{s}$. This implies $1 \leq (X,Y;A,B) \leq (X',Y';A,B)$ which implies $\widehat{AB} \leq \widehat{AB}_{TUW} = \widehat{AC} + \widehat{CB}$. □

> integral operators with positive kernel could both be proved by an application of the Banach contraction mapping theorem in suitable metric spaces. Birkhoff ... relied heavily on arguments from differential projective geometry.

The "suitable metric spaces" referred to in the above quote are cross sections of certain types of cones [12] to which the Hilbert Projective Metric is applied. Birkhoff [12] starts his definition of the metric by constructing it first on the standard cone $\mathbb{R}^2_{\geq 0}$:

> **2. Projective Metrics on line.** In homogenous coordinates, the first positive quadrant joins $(0,1)$ with $(1,0)$ by "points" (f_1, f_2). This is mapped onto the hyperbolic line $-\infty < u < +\infty$ by the correspondence $\mathrm{Ln}(f_2/f_1) = u$. We define
>
> $$\theta(f,g) = |\mathrm{Ln}(v) - \mathrm{Ln}(u)| = |\mathrm{Ln}(f_2 g_1 / f_1 g_2)|. \tag{4}$$
>
> Since $f_2 g_1 / f_1 g_2$ is the cross-ratio $R(f_2/f_1, g_2/g_1; 0, \infty)$, $\theta(f,g)$ is invariant under all projective transformations mapping the interval $0 < f_2/f_1 < \infty$ onto itself.

Birkhoff's construction is explained in Figure 4. It is worth noting that the ratio f_2/f_1 is the slope of the ray $[f] = \{\lambda f \mid \lambda > 0\}$.

Birkhoff then extends this construction to bounded closed convex cones. For Birkhoff, it seems a cone C is bounded if $\exists$ a hyperplane H which intersects each ray [13] of $C \setminus \{0\}$ exactly once and if the intersection $C \cap H$ is bounded [14]. Birkhoff [12] continued:

[12] Technically, a cone C is a subset of a vector space which is closed under non negative scaling; i.e. $\forall\ \alpha > 0$ we have $\alpha C \subset C$ However, we will additionally assume 0 is in every cone.

[13] If $c \in C$, the ray $[c] = \{\lambda c : \lambda > 0\}$. If we define $c \sim c'$ if $\exists\ \lambda > 0$ such that $c = \lambda c'$ then it is easy to see that $\sim$ is an equivalence relation on C, and the equivalence class of c is $[c]$.

[14] Bounded in the sense of having finite diameter with respect to the vector space L's norm. Birkhoff is implicitly assuming that L is a normed linear space.

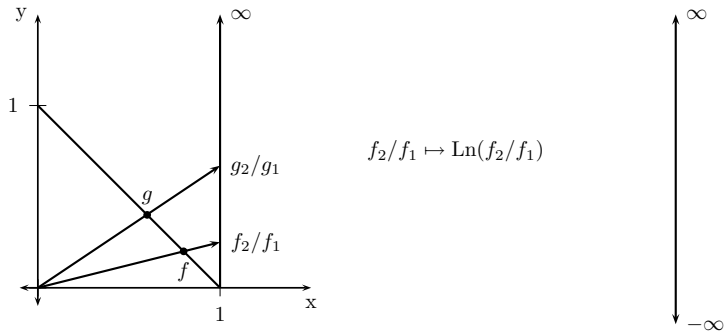


Figure 4: The Hilbert Metric can be applied to the line segment e_1e_2 connecting $e_1 = (1,0)$ to $e_2 = (0,1)$ as it is a bounded convex subset of R^2. By the projective invariance of the cross ratio, the projection of the four collinear points $(1\,,0), f, g, (0,1)$ onto the vertical line $[0,\infty]$, with $(0,1) \mapsto \infty$, as show in this figure, does not alter the value of their cross ratio. So, $\theta(f,g) = |\mathrm{Ln}(f_2g_1/f_1g_2)| = |\mathrm{Ln}R(f_2/f_1, g_2/g_1; 0, \infty)| = |\mathrm{Ln}R(f,g;(1,0),(01))|$, are all equivalent to the Hilbert Metric on the line segment e_1e_2. If we let $m(f) = f_2/f_1 =$ the slope of f, then $\theta(f,g) = |\mathrm{Ln}(m(f)/m(g))|$.

3. Convex cones. Now let C be any bounded closed convex cone of a real vector space L, of finite or infinite dimensions. It is convenient to make a central projection of C onto its (convex) intersection $C \cap H$ with a hyperplane H, cutting each ray of C in exactly one point; we can then discuss C and $C \cap H$ interchangeably, as subspaces of projective space.

Since C is a bounded, closed, convex set, *every line intersects C in a closed segment* [15]. Hence, if $f \neq g$ in H, the intersection of the line $l(f,g)$ with C can be mapped onto the line $0 \leq x \leq \infty$ of §2 [16] so that $fA < gA$ by a *unique* affine transformation A. We define

$$\theta(f,g;C) = \theta(fA, gA). \tag{5}$$

If f or g is a boundary point, $\theta(f,g:C) = \infty$. We call $\theta(f,g;C)$ the

[15] These two C's represent the bounded convex set $C \cap H$. Since, if not, one has the example of C being the cone $\mathbb{R}^2_{\geq 0}$ and l being the line $y = x - 1$. This line $y = x - 1$ is not suitable for Birkhoff's construction of the Hilbert Metric on the cone $\mathbb{R}^2_{\geq 0}$ as it does not intersect all the rays in the the interior of $\mathbb{R}^2_{\geq 0}$.

[16] In §2 there is no explicit mention of the line $0 \leq x \leq \infty$, although it is implicity invoked; see Figure 4.

projective metric associated with C.

The above quote is non-trivial and central to understanding Birkhoff's original work and so here is a brief explanation.

Regarding the second paragraph of the above quote: When Birkhoff writes, "Since C is a bounded, closed, convex set, *every line intersects C in a closed segment.*" the C's should be interpreted as being $C \cap H$. So in that sentence, C being closed, means that $C \cap H$ is closed (which is problematic if H is a non-closed hyperplane). However, $C \cap H$ being closed isn't needed in the proof of $l(f,g) \cap C$ being a line segment. The following argument shows this: We are assuming that the cone C is closed, bounded, and convex. So $C \cap H$ is bounded and convex. Since $f \neq g \in H$, it follows that $l(f,g) \subset H$ and so

$$\underbrace{l(f,g) \cap C}_{\text{closed}} = \underbrace{l(f,g) \cap C \cap H}_{\text{bounded and convex}}$$

is a closed, bounded, convex line segment in L.

Regarding the affine map: The affine map that Birkhoff mentioned in the above quote, but did not describe, is explained in the following construction:

See Figure 5. The two endpoints, b_1, b_2 of the line segment $l(f,g) \cap C \cap H$ are linearly independent and determine a two dimensional vector subspace $\text{Span}(b_1, b_2) \subset L$. There are exactly two non-singular linear transformations which map $\text{Span}(b_1, b_2)$ to $\mathbb{R}^2$ sending the basis $\{b_1, b_2\}$ of $\text{Span}(b_1, b_2)$ to the the standard basis $\{e_1, e_2\}$ of R^2. These two linear transformations, which we will denote by A^+ and A^-, map the line segment $l(f,g) \cap C \cap H$ affinely onto the line segment e_1e_2. Concretely, if $f = f_1b_1 + f_2b_2$ we can let $A^+(f) = (f_1, f_2)$ and $A^-(f) = (f_2, f_1)$. We project the line segment e_1e_2 onto the line $0 \leq u \leq \infty$ via the map $m(x,y) = y/x$ same as in Figure 4.

The inequality $fA < gA$ appearing in the second paragraph should be understood

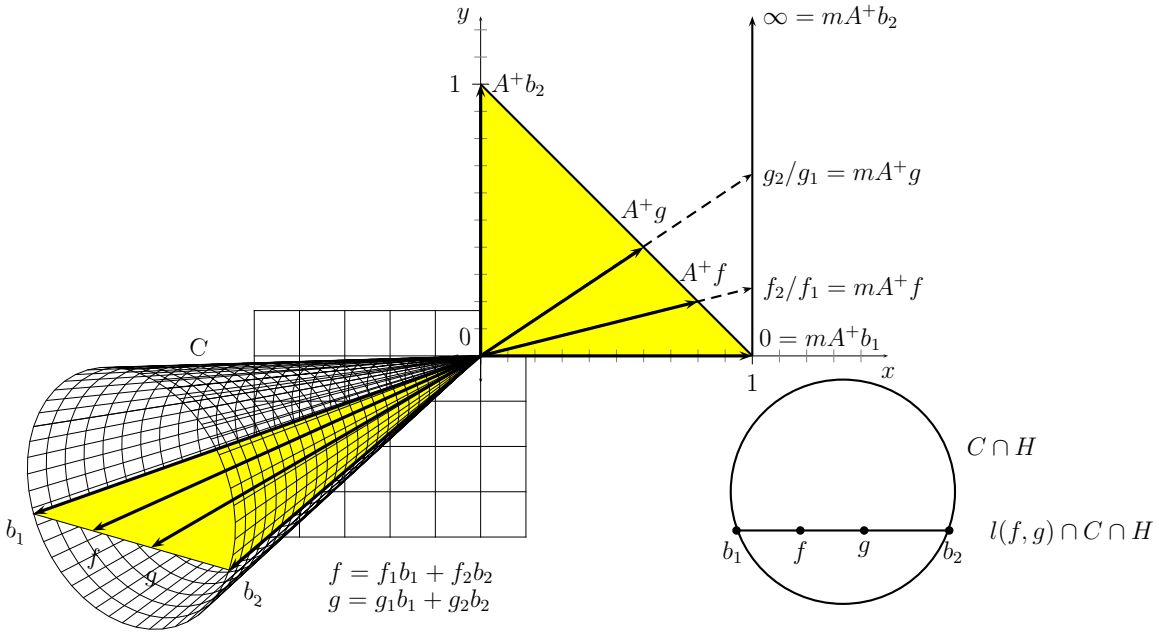


Figure 5: On left, in the cone C, the two endpoints, b_1, b_2 of the line segment $l(f,g) \cap C \cap H$ are linearly independent and determine a two dimensional vector subspace $\mathrm{Span}(b_1, b_2)$. The two dimensional cone $\mathrm{Span}(b_1, b_2) \cap C$ is mapped isomorphically to the standard cone $\mathbb{R}^2_{\geq 0}$ by the linear transformation A^+.The tip of that 2 dim cone, from $l(f,g) \cap C \cap H$ to 0, is shown in yellow. A^+ maps that 'tip' to the yellow triangle shown $\subset \mathbb{R}^2_{\geq 0}$. mA^+f = the slope of A^+f = the height of A^+f's projection to $1 \times [0, \infty]$. $\theta(f, g; C) = |\mathrm{Ln}(A^+f) - \mathrm{Ln}(A^+g)| = |\mathrm{Ln}(f_2g_1/f_1g_2)|$. The hyperplane H which intersects each ray in C exactly once is not shown. On the right is shown $C \cap H$.

as follows. Since,

$$m(A^+(f)) = m(f_1, f_2) = \frac{f_2}{f_1} = \frac{1}{\frac{f_1}{f_2}} = \frac{1}{m(f_2, f_1)} = \frac{1}{m(A^-(f))} \tag{6}$$

either

$$m(A^+(f)) < m(A^+(g)),$$

in which case we let $A = A^+$, or

$$m(A^-(f)) < m(A^-(g)),$$

in which case we let $A = A^-$.

So we have constructed the "*unique* affine transformation A" such that $fA < gA$. Note that Birkhoff preferred to write his transformations to the right of what they are transforming and that the interpretation of $fA < gA$ is as explained.

Birkhoff used A to define θ on $C \cap H$ by setting $\theta(f, g; C) = \theta(fA, gA)$ from (4). On the other hand, if we choose A' to be A^+ or A^- so that A' satisfies $fA' > gA'$, and if instead of A we use A' to affinely map b_1b_2 to $e_1e_2 \subset \mathbb{R}^2$ and to construct the metric, then the metric we obtain will be exactly the same as if we had used A. To see this write f and g in terms of b_1, b_2: $f = f_1b_1 + f_2b_2$ and $g = g_1b_1 + g_2b_2$ [17]. Then:

$$\theta(fA^+, gA^+) = \theta((f_1, f_2),\ (g_1, g_2)) = |\mathrm{Ln}(f_2g_1/f_1g_2)|$$
$$\theta(fA^-, gA^-) = \theta((f_2, f_1),\ (g_2, g_1)) = |\mathrm{Ln}(f_1g_2/f_2g_1)|$$

So $\theta(fA^+, gA^+) = \theta(fA^-, gA^-)$. Hence, we are free to choose the more convenient of A^+ or A^- to calculate $\theta(f, g; C)$.

It is worth emphasizing that the projective metric θ, as constructed by Birkhoff in [12] is defined on $C \cap H$ for some particular H. However, if H is replaced by a similar hyperplane and Birkhoff's construction is applied, the resulting metric will be exactly the same as θ [18]. As a result, the metric θ can be extended in a well defined way to a pseudo metric [19] on $C \setminus \{0\}$ via central projection as follows: if $f \in C \setminus \{0\}$ define $f^H = [f] \cap H$ (i.e. central projection onto H) and then for $f, g \in C \setminus \{0\}$ define

$$\theta(f, g; C) = \theta(f^H, g^H; C) \tag{7}$$

[17] Since $f, g \in b_1b_2$ and b_1, b_2 are linearly independent we must have $f_1, f_2, g_1, g_2 \geq 0$, uniquely determined, and satisfying $f_1 + f_2 = 1;\ g_1 + g_2 = 1$.

[18] Using the projective invariance of the cross ratio with the origin 0 as the projection point, and some topology (for boundary issues), one can show that the projective metric as defined by Birkhoff is independent of our choice of hyperplane H in the following sense. Suppose H' is another hyperplane which intersects each ray in $C \setminus \{0\}$ exactly once (like H) and that $[f] \cap H' = f',\ \ [g] \cap H' = g'$ and that $l(f', g') \cap H' \cap C$ has end points b_1', b_2'. If we calculate $\theta(f', g'; C)$ using b_1', b_2' (which are dependent on H') and we calculate $\theta(f, g; C)$ using b_1, b_2 (which are dependent on H), then the projective distances are the same; i.e. $\theta(f, g; C) = \theta(f', g'; C)$.

[19] A pseudo-metric, d satisfies all the properties of a metric, except $d(x, y) = 0$ does not necessarily imply $x = y$.

where the second θ is as in (5).

The next major step in the evolution of the Hilbert Projective Metric is found in Birkhoff [13] (1962). There, Birkhoff reworked his definition of the projective metric θ, into its modern form, making it "more general and more rigorous," [13, p. 45]. The modern form of the definition is algebraic and based upon a partial order that the cone C induces on L. Here are the details:

If C is a convex salient [20] cone contained in a vector space L and if for $x, y \in L$ we write $x \leq y$ when $y - x \in C$, then $\leq$ partially orders L. For a proof, see [21]. Recall that in [12] Birkhoff required that C be a "bounded closed convex cone". In [13] Birkhoff replaced the requirement of C being bounded with the more general condition of C being salient [22] as this condition (combined with convexity) is sufficient to ensure that $\leq$ partially orders L. Quoting from Birkhoff [13]:

> DEFINITION. In C [23, 24], define $\theta(f, g) = +\infty$ if $\alpha f \geq g$ for no positive scalar α, or if $\beta g \geq f$ for no positive scalar β. Otherwise, let α_0 and β_0 be the least such scalars.. . .
>
> Define
>
> $$\theta(f, g) = \ln(\alpha_0 \beta_0) = \ln \alpha_0 + \ln \beta_0. \tag{8}$$

[20] A cone C is salient if $C \cap (-C) = \{0\}$. It is worth noting that if C is a salient cone then $0 \in C$.

[21] **Proposition C.** If C is convex salient cone then $\leq$ partially orders L.

Proof: Let $x, y, z \in L$. Reflexivity: $x \leq x$ follows from $0 \in C$. Anti-symmetry: $x \leq y, y \leq x \Rightarrow x = y$ follows from C being salient. Transitivity: $x \leq y,\ y \leq z \Rightarrow x \leq z$ follows from C being a convex cone and hence closed under addition (see Proposition D). □

Proposition D. If C is a convex cone then it is closed under addition.

Proof: If $f, g \in C$ then $(f + g)/2$ is on the line segment $fg \subset C$ so, since C closed under non negative scaling, we have $f + g \in C$. □

[22] **Proposition E.** C is bounded implies C is salient.

Proof. If C is bounded $\exists$ a hyperplane H which intersects each ray in $C \setminus \{0\}$ exactly once. Suppose $f, -f \in C \setminus \{0\}$. Then $[f] \cap H$ and $[-f] \cap H$ are both non empty and it easily follows that $[f], [-f] \subset H$, which is a contradiction.

[23] Birkhoff should have written $C \setminus \{0\}$ here, rather than C.

[24] Generally, the cone C is assumed to be closed, convex, and salient. One of the benefits of C being topologically closed is that it forces α_0 and β_0 to be > 0 as the following argument shows. Suppose the convex salient cone C is topologically closed; that $f, g \in C \setminus \{0\}$; and that $\exists$ a sequence of positive numbers α_n with $\alpha_n \to 0$ and $\alpha_n f \geq g$ for each n. Then, $\alpha_n f - g \in C$ for each n. So $\alpha_n f - g \to -g \in C$, as C is closed. Since C is salient $-g, g \in C$ implies $g = 0$. Contradiction. So $\alpha_0 > 0$.

θ, as defined in (8), is a pseudo-metric on $C \setminus \{0\}$ because if f' belongs to the ray $[f] \subset C \setminus \{0\}$ then $\theta(f', f) = 0$. More generally, if we also have $g' \in [g]$ then $\theta(f, g) = \theta(f', g')$. So θ defines a metric on the rays of $C \setminus \{0\}$.

The two very differently constructed θ's give the same distances if C is a bounded convex closed cone [25].

By the early 1970's, it became standard to express the Hilbert Projective Metric as follows, paraphrasing Zabreiko, Krasnosel'skii and Pokornyi [92] (1971):

> Let $x, y \in K$, a cone. Let α be the greatest, and β the smallest numbers for which $\alpha x \leq y \leq \beta x$ holds. Define $\rho(x, y) = \ln(\beta/\alpha)$.

ρ is, of course, just a minor reworking of Birkhoff's definition (8). The interested reader is directed to Bushell [19] (1973) or Nussbaum [77] (1986) for similar constructions of the Hilbert Projective Metric.

Birkhoff constructed the Hilbert Projective Metric so as to prove his "Projective Contraction Theorem," Birkhoff [12] (1957), [13] (1962). Under certain conditions, if a linear map sends a closed, convex, salient cone to itself, then with respect to the Hilbert Projective Metric, this map will be a contraction having a unique fixed point (eigen direction). This can be used to give a proof of the Perron-Frobenius Theorem of linear algebra, [12], or Gaubert and Gunawardena [38] (2004), and has applications to distributions and multiplicative processes (such as branching processes), [13]; and differential equations, Birkhoff and Kotin [14] (1965).

The applications that are of primary importance for this dissertation are distributions; matrix and differential equations; and branching processes as related to population biology and evolutionary dynamics. In Part II those applications will be discussed in depth - the material in Part I (Chapters 1 and 2) is foundational.

[25] In [13] Birkhoff stated without proof that that $R_0 = \alpha_0\beta_0$ is the "cross-ratio of (a, f, g, b) on the projective line L through f and g, where a and b are the ends of the closed segment $L \cap C$." In other words, θ from (8) agrees with θ from (5) when C is a bounded convex closed cone.

Chapter 1 of this dissertation is essentially a collection of results, a tool kit for working with the various forms of the Hilbert Projective Metric. Chapter 2 is the same, but regarding the Projective Contraction Theorem. Almost all the proofs I give are either original (at least in the sense of not being copied); or an expansion of a sketched proof. I believe there are some useful original results as well. That said, the majority of the main results can be found spread out across the literature, but not in one place - and often the proofs given are sketchy for lack of a better word. I felt it would be helpful to have all these results rigorously proven and in one place.

The Hilbert metric and the Projective Contraction Theorem are very useful tools in modern mathematical research and related topics such as population biology and evolutionary dynamics, which we will present in Part II. There are many other noteworthy applications. A few of these in dynamical systems and related topics are those of:

Curt McMullen [67] (2002) for applications to Coexter Groups.

Pierre Ferrero and Bernhard Schmitt [34] (1979) and Carlangelo Liverani [64] (1995) for applications to Ruelle's Perron-Frobenius Theorem on positive transfer operators and the convergence rate.

See also Yunping Jiang [51] (2000) for a survey on Ruelle's Perron-Frobenius Theorem on positive transfer operators and the convergence rate for expanding dynamical systems by using the Hilbert Metric. Using the Hilbert Metric, one can estimate that the rate of the decay of correlation is exponentially fast. That is, the rate of approach of some initial distribution to an invariant one is exponentially fast for good dynamical systems. Some other kinds of the rate of the decay of correlation can be studied also. The reader who is interested in this direction can refer to [34, 64, 51, 32, 33].

Chapter 1

Cones

1.1 Basic definitions for cones

Let V be a linear space (i.e. any vector space). Geometrically speaking, a cone is a subset of V which can be represented as a union of rays emanating from a single source (point). If that source point is considered to be part of the cone, we say the cone is pointed.

We will always assume that the source of the cone is the null vector (the origin 0) of V. This assumption leads to the following easy to state and precise algebraic definition:

Definition 1.1.0.1. A subset C of the vector space V is called a cone if it is closed under positive scaling. If C is closed under non-negative scaling then C is a pointed cone. We will sometimes say pointed by the origin to emphasize that $0 \in C$ and that 0 is the point of the cone.

Unless otherwise noted, all rays will be assumed to be emanating from the origin of V. This assumption leads to the following definitions:

Definition 1.1.0.2. The closed ray emanating from the origin and passing through

the point $v \in V$ is denoted $\overrightarrow{v}$. As a point set:

$$\overrightarrow{v} = \{\lambda v \mid \lambda \geq 0\}.$$

For typographic reasons, the closed ray $\overrightarrow{v}$ may occasionally be denoted as ray(v)

Definition 1.1.0.3. The open ray emanating from the origin and passing through the point $v \in V$ is denoted $[v]$. As a point set:

$$[v] = \{\lambda v \mid \lambda > 0\}.$$

If we wish to emphasize (or remind the reader) that a ray is emanating from $\{0\}$, we will call it a 0-ray. So $\overrightarrow{v}$ would denote a closed 0-ray and $[v]$ would denote an open 0-ray.

Remark 1.1.0.4. The closed 0-ray $\overrightarrow{0}$ and the open 0-ray $[0]$ are equivalent as they both equal the singleton set $\{0\}$. If $v \neq 0$ then $0 \notin [v]$. It is always the case that $\overrightarrow{v} = [v] \cup \{0\}$.

Remark 1.1.0.5. The open 0-rays of V partition V, hence the notation $[v]$ is appropriate. See Section 1.7.1 for details.

Remark 1.1.0.6. The open ray $[v]$ is the smallest cone containing v. The closed ray $\overrightarrow{v}$ is the smallest pointed cone containing v.

Definition 1.1.0.7. The cone opposite to C is denoted $-C$, algebraically $-C = \{-c \mid c \in C\}$. A cone C is salient (bounded) if $C \bigcap -C \subset \{0\}$.

Definition 1.1.0.8. C is convex if $c_1, c_2 \in C$ implies the line segment

$$\overline{c_1 c_2} = \{c_1 + \lambda(c_2 - c_1) \mid \lambda \in [0, 1]\} \subset C.$$

If V is a normed linear space. V's norm $\| \ \|$ turns V into a metric space with $d(v_1, v_2) = \|v_2 - v_1\|$.

Definition 1.1.0.9. C is closed if it is closed w.r.t. V's topology.

Proposition 1.1.0.10. *If C is a salient and closed cone contained in the Banach Space V, then C contains no lines. This proposition is true whether C is pointed or not.*

Proof. Let v_0 and v_1 be any two distinct points on a line L contained in C. Then $L = \{v_0 + (v_1 - v_0)t : t \in \mathbb{R}\}$. Since C is closed under positive scaling the following two sequences:

$$\left\{ \frac{v_0 + (v_1 - v_0)n}{||v_0 + (v_1 - v_0)n||} \right\}_{n=1}^{\infty}, \quad \left\{ \frac{v_0 + (v_1 - v_0)(-n)}{||v_0 + (v_1 - v_0)(-n)||} \right\}_{n=1}^{\infty}$$

are contained in C. Since V is complete and C is closed, these two sequence converge, respectively, to the following two points in C:

$$\frac{v_1 - v_0}{||v_1 - v_0||}, \quad -\frac{v_1 - v_0}{||v_1 - v_0||}.$$

This contradicts C being salient. □

Remark 1.1.0.11. In Proposition 1.1.0.10 (page 21), the requirement that C is closed is required as the following examples show.

Consider the open upper half plane $H = \{(x, y) \in \mathbb{R}^2 : y > 0\}$. H is closed under positive scaling so H is a cone. Since $(x, y) \in H \Leftrightarrow y > 0$ it follows that $-(x, y) = (-x, -y) \notin H$ and so $H \cap -H = \emptyset$. Thus H is a salient cone contained in $\mathbb{R}^2$, a Banach space. However, the cone H contains every line $y = k$ for each $k > 0$.

Similarly, the pointed salient cone $H \cup \{0\} \subset \mathbb{R}^2$ contains every line $y = k$ for each $k > 0$.

1.1.1 The cone's boundary ∂.

Definition 1.1.1.1. Let D be a subset of V. Let hyper$\{D\}$ be the smallest hyperplane such that $D \subset \text{hyper}\{D\}$. Give hyper$\{D\}$ the induced topology. The (linear) boundary of D, ∂D, is D's topological boundary relative to hyper$\{D\}$.

Remark 1.1.1.2. In these notes ∂ will essentially be applied to closed convex subsets of V. Topologically speaking, closed convex sets are manifolds with boundary. Applying ∂ to a closed convex set yields its boundary in terms of the manifold with boundary definition of boundary. The benefit of defining ∂ as we do, in terms of hyperplanes, is that it makes the definition easy to state, easy to apply, and it avoids delving into the machinery and theory of topological manifolds with boundary.

Example 1.1.1.3. Let $a, b \in V$ be two distinct points. If V has dimension greater than 1, then the standard topological boundary of of the line segment $\overline{ab}$ is all of $\overline{ab}$. On the other hand $\partial\overline{ab} = \{a, b\}$; i.e., the end points of $\overline{ab}$.

Example 1.1.1.4. Consider the unit disk D from $\mathbf{R^2}$ embedded in $\mathbf{R^3}$: $D = \{(x, y, 0) \in \mathbf{R^3} \mid x^2 + y^2 \leq 1\}$. The topological boundary of D is all of D. However $\partial D = \{(x, y, 0) \in \mathbf{R^3} \mid x^2 + y^2 = 1\}$, which is the circumference of D.

Proposition 1.1.1.5. *$p \in \partial D$ if and only if $p \in$ hyper$\{D\}$ is such that whenever $p \in U$, with U being any open subset of V, we have:*

$$U \bigcap D \neq \emptyset \quad \textit{and} \quad U \bigcap (\textit{hyper}\{D\} \setminus D) \neq \emptyset.$$

Proof. This an immediate consequence of ∂'s definition. □

Definition 1.1.1.6. Span(D) is the smallest vector space in V such that $D \subset$ Span(D) Alternatively, Span(D) is the set of all finite linear combinations of vectors from D.

Remark 1.1.1.7. If $0 \in D$ then $\text{hyper}\{D\} = \text{Span}(D)$.

Proposition 1.1.1.8. *Let C be a subset of V containing the origin. Let $A \subset C$. Then*

$$\partial(C \cap \mathit{Span}(A)) \ \subset \ (\partial C) \cap \mathit{Span}(A). \tag{1.1}$$

Moreover, the containment in (1.1) can be strict even when C is the cone $\mathbb{R}^2_{\geq 0}$.

Proof. $0 \in C \cap \text{Span}(A)$ so

$$\text{hyper}\{C \cap \text{Span}(A)\} = \text{Span}(C \cap \text{Span}(A)). \tag{1.2}$$

In general, $A \subset A'$ implies $\text{Span}(A) \subset \text{Span}(A')$ so

$$\begin{aligned} &\text{Span}(C \cap \text{Span}(A)) \subset \text{Span}(C) \\ &\text{Span}(C \cap \text{Span}(A)) \subset \text{Span}(\text{Span}(A)) = \text{Span}(A) \qquad \text{so that} \\ &\text{Span}(C \cap \text{Span}(A)) \subset \text{Span}(C) \cap \text{Span}(A). \end{aligned} \tag{1.3}$$

Proposition 1.1.1.5 (page 22), (1.2) and (1.3) yield

$$\begin{aligned} p \in \partial(C \cap \text{Span}(A)) \Rightarrow p &\in \text{hyper}\{C \cap \text{Span}(A)\} \\ &= \text{Span}(C \cap \text{Span}(A)) \\ &\subset \text{Span}(C) \cap \text{Span}(A). \end{aligned}$$

Proposition 1.1.1.5 (page 22) together with (1.2) implies that if $p \in \partial(C \cap \text{Span}(A))$ and if U is any open set containing p then

$$U \bigcap \underbrace{\text{hyper}\{C \cap \text{Span}(A)\}}_{=\text{Span}(C \cap \text{Span}(A))} \setminus \ (C \cap \text{Span}(A)) \neq \emptyset \tag{1.4}$$

But $\text{Span}(C \cap \text{Span}(A)) \subset \text{Span}(A)$ so (1.4) implies:

$$U \bigcap \text{Span}(A) \ \setminus \ (C \cap \text{Span}(A)) \neq \emptyset. \tag{1.5}$$

(1.5) simplifies to

$$U \bigcap \text{Span}(A) \ \setminus \ C \neq \emptyset.$$

Since $A \subset C$ it follows that $\text{Span}(A) \subset \text{Span}(C)$, hence:

$$U \bigcap \text{Span}(C) \ \setminus \ C \neq \emptyset.$$

By Proposition 1.1.1.5 (page 22), if $p \in \partial(C \cap \text{Span}(A))$ and U is any open set about p we must have

$$U \bigcap \{C \cap \text{Span}(A)\} \neq \emptyset,$$

which immediately implies:

$$U \bigcap C \neq \emptyset,$$

So we've shown that if $p \in \partial(C \cap \text{Span}(A))$ and if U is any open set containing p that

$$p \in \text{Span}(C), \ U \bigcap C \neq \emptyset, \text{ and } U \bigcap (\text{Span}(C) \setminus C) \neq \emptyset.$$

which implies $p \in \partial C$. We also have shown that if $p \in \partial(C \cap \text{Span}(A))$ then $p \in \text{Span}(A)$. So

$$(\partial C) \bigcap \text{Span}(A) \subset \partial(C \cap \text{Span}(A))$$

Example:

Let $V = \mathbb{R}^2$; $C = \mathbb{R}^2_{\geq 0}$ and $A = \{(0,1)\}$ so that

$$\text{Span}(A) = \{(0, y) : y \in \mathbb{R}\}.$$

Then

$$\partial(C \cap \mathrm{Span}(A)) = \partial\{(0, y) : y \geq 0\} = \{(0, 0)\}$$

and

$$\partial(C) \cap \mathrm{Span}(A) = \{(0, y) : y \geq 0\}$$

□

1.2 A Collection of Standard Results for Normed Linear Spaces

1.2.1 For Finite Dimensional Normed Linear Spaces

Remarks: We state the following without proof.

1. If W and W' are finite dimensional normed linear spaces over $\mathbf{R}$ (or $\mathbf{C}$ for that matter) and f is an isomorphism from W to W' then f is also a homeomorphism.

2. If W is a subspace of V, then W is a topologically closed subset of V.

3. Finite dimensional normed linear spaces are complete.

For proof see Rudin [83], especially Theorem 1.16 on page 16.

Remark 1.2.1.1. We will make central use of Proposition 1.2.1.2 (page 25), which follows. For that reason, we include its proof. Note that Proposition 1.2.1.2 and the proof we give closely follows Proposition 4.3.1. on page 189 of Bridges [17].

Proposition 1.2.1.2. *Let $(V, \| \ \|)$ be any normed linear space over $\boldsymbol{R}$. Let W be any two dimensional subspace of V and let b_1, b_2 be an ordered basis for W. Then*

$$\begin{aligned} (R^2, \| \ \|_\infty) &\overset{f}{\mapsto} (W, \| \ \|) \\ (t_1, t_2) &\mapsto t_1 b_1 + t_2 b_2 \end{aligned}$$

is an isomorphism of vector spaces; f is bounded (continuous) and has bounded inverse; hence f is a homeomorphism.

Proof. That f is an isomorphism is trivial. Let

$$c = \max\{\|b_1\|, \|b_2\|\}$$

The inequalities

$$\begin{aligned} \|f(t_1,t_2)\| = \|t_1b_1 + t_2b_2\| &\leq |t_1|\,\|b_1\| + |t_2|\,\|b_2\| \\ &\leq |t_1|c + |t_2|c \\ &\leq 2\max\{|t_1|,|t_2|\}c \\ \Rightarrow \quad \frac{\|f(t_1,t_2)\|}{\max\{|t_1|,|t_2|\}} &= \frac{\|f(t_1,t_2)\|}{\|(t_1,t_2)\|_\infty} \leq 2c. \end{aligned}$$

So f is bounded. Now we show that f^{-1} is also bounded. Let

$$S = \left\{(t_1,t_2) \in \mathbf{R^2} \mid \|(t_1,t_2)\|_\infty = 1\right\}.$$

Claim: S is compact.

Proof of Claim: First we prove: $[a,b] \subset \mathbf{R}$ is compact.

Let $\mathcal{C}$ be an open cover of $[a,b]$. Let

$$s = \sup\{t \in [a,b] \mid [a,t] \text{ has a finite subcover from } \mathcal{C}\}.$$

$[a,b]$ is closed and bounded above so $s \in [a,b]$. Since $\mathcal{C}$ covers $[a,b]$ $\exists$ at least one open set in $\mathcal{C}$, say U_s, such that $s \in U_s$. As U_s is open $\exists \varepsilon > 0$ such that $(s-\varepsilon, s+\varepsilon) \subset U_s$. But then the interval $[a, s+\varepsilon]$ has a finite subcover. This contradicts the supremacy of s unless $s = b$. So $[a,b] \subset \mathbf{R}$ is compact and in particular, $[-1,1]$ is compact.

Now we show S is compact. Let $a, b, c, d : [-1, 1] \mapsto S \subset W$ as follows

$$a(t) = (-1, t), \quad b(t) = (t, 1), \quad c(t) = (1, t), \quad d(t) = (t, -1).$$

Clearly $a, b, c,$ and d are continuous. Hence the images of $[-1, 1]$ under a, b, c, d are compact. The finite union of compact sets is compact. Since

$$S = a([-1, 1]) \cup b([-1, 1]) \cup c([-1, 1]) \cup d([-1, 1])$$

it follows that S is compact. So we've proven the claim.

Claim: f^{-1} is bounded.

Proof of Claim: First, let $w \neq 0$ be $\in W$. Since f is an isomorphism there is a unique vector $(w_1, w_2) \in \mathbf{R^2}$ such that

$$f^{-1}(w) = (w_1, w_2)$$

Or, explicitly, since b_1, b_2 is a basis for W we can write $w = w_1 b_1 + w_2 b_2$ with w_1, w_2 uniquely determined. But then $f(w_1, w_2) = w$ and $f^{-1}(w) = (w_1, w_2)$.

The map $x \mapsto \|x\|$ is continuous and f is continuous so the map

$$(t_1, t_2) \mapsto \|f(t_1, t_2)\|$$

is continuous. Since S is compact $\|f(S)\|$ is a compact subset of $\mathbf{R}$.

$0 \notin S$ and S is an isomorphism so $\forall s \in S$ we have $f(s) \neq 0$ and $\|f(s)\| > 0$.

Since $\|f(S)\|$ is a compact subset of $\mathbf{R}$ it contains its inf. So

$$0 < r = \inf \|f(S)\| .$$

Then, since

$$\frac{f^{-1}(w)}{\|f^{-1}(w)\|_\infty} \in S$$

it follows that

$$0 < r < \left\| f\left(\frac{f^{-1}(w)}{\|f^{-1}(w)\|_\infty}\right)\right\| = \frac{\|w\|}{\|f^{-1}(w)\|_\infty} \tag{1.6}$$

Inverting (1.6) we get

$$\frac{\|f^{-1}(w)\|_\infty}{\|w\|} < \frac{1}{r} < \infty. \tag{1.7}$$

Since w was arbitrary (except for being non-zero) (1.7) implies

$$\sup_{\substack{w\in W\\ w\neq 0}} \frac{\|f^{-1}(w)\|_\infty}{\|w\|} \leq \frac{1}{r} < \infty.$$

I.e. f^{-1} is bounded. □

Remark 1.2.1.3. According to the Heine-Borel theorem if S is a closed and bounded subset of R^n it is compact. The unit sphere $S \subset R^n$ is just the inverse image of the number 1 w.r.t. the continuous map $x \mapsto \|x\|$. So S is closed and bounded and hence compact. Consequently, the proof of Proposition 1.2.1.2 (page 25) can be made to work in all finite dimensions.

1.2.2 For Banach Spaces

The following elementary results are quite useful.

Proposition 1.2.2.1. *Let X be a Banach Space. Let $a \in X$ and let*

$$f_a : X \to X \text{ by } f_a(x) = a + x$$

$$f_- : X \to X \text{ by } f_-(x) = -x$$

Then f_a and f_- are isometries and topological homeomorphisms of X to itself. More-

over, $f_a^{-1} = f_{-a}$ *and* $f_-^{-1} = f_-$.

Proof. Let $x, y \in X$

One to one. If $f_a(x) = f_a(y)$ then $a + x = a + y$, which implies $x = y$. Onto. Let $y \in X$ be given. Then $f_a(-a + y) = a + (-a + y) = y$ so f_a is onto. Inverse: $f_{-a}(f_a)(x) = -a + (a + x) = x$ and $f_a(f_{-a})(x) = a + (-a + x) = x$ so $f_a^{-1} = f_{-a}$. Isometry: $\|f_a(x) - f_a(y)\| = \|(a + x) - (a + y)\| = \|x - y\|$. Isometries of metric spaces are always continuous (and one to one), hence f_a and its inverse, f_{-a} are continuous. Hence f_a is a homeomorphism.

One to one. If $f_-(x) = f_-(y)$ then $-x = -y$, which implies $x = y$. Onto. Let $y \in X$ be given. Then $f_-(-y) = -(-y) = y$ so f_a is onto. Inverse. $f_-(f_-)(x) = x$ so $f_-^{-1} = f_-$. Isometry: $\|f_-(x) - f_-(y)\| = \|(-x) - (-y)\| = \|y - x\| = \|x - y\|$. Isometries of metric spaces are always continuous (and one to one), hence f_- and its inverse, which is itself, are continuous. Hence f_- is a homeomorphism. □

Corollary 1.2.2.2. *Let* C *be a closed subset of* X. *Then* $\forall a \in X$ *the sets* $a + C$ *and* $a - C$ *are closed.*

Proof. Since f_a and $f_a \circ f_-$ are homeomorphisms, f_a and $f_a \circ f_-$ map closed sets to closed sets. In particular $f_a C = a + C$ and $f_a \circ f_-(C) = a - C$ are closed. □

1.3 The Cone Partial Order $\leq$ and $m(y/x),\ b(y/x)$

Proposition 1.3.0.3. *If* C *is a pointed convex cone (emanating from the origin) then it is closed under non-negative linear combinations.*

Proof. Let λ_1, λ_2 be non-negative and $c_1, c_2 \in C$. As a cone, C is closed under non-negative scaling, so $\lambda_1 c_1, \lambda_2 c_2 \in C$. Convexity implies $\frac{1}{2}\lambda_1 c_1 + \frac{1}{2}\lambda_2 c_2 \in C$. C is closed w.r.t to non-negative scaling so $2\left(\frac{1}{2}\lambda_1 c_1 + \frac{1}{2}\lambda_2 c_2\right) = \lambda_1 c_1 + \lambda_2 c_2 \in C$. □

Remark 1.3.0.4. Vector subspaces and cones are similar in the following sense; a vector subspace, such as Span(D), is closed under all finite linear combinations. If C is a pointed convex cone then C is closed under all finite non-negative linear combinations.

Definition 1.3.0.5. Let C be a salient cone emanating from the origin. If $f, g \in V$ we write $f \leq g$ if $g - f \in C$.

Example 1.3.0.6. For $x, y \in \mathbf{R} = V$, $C = [0, \infty)$ the relation, $x \leq y$ is the familiar $x \leq y$, where $x \leq y$ means $y - x$ is non-negative.

Proposition 1.3.0.7. *If C is a salient convex cone emanating from the origin the relation $\leq$ partially orders V.*

Proof. Let $f, g \in V$. Reflexivity: $f - f = 0 \in C$ so $f \leq f$. Antisymmetry: if $f \leq g$ and $g \leq f$ then $f-g$ and $g-f = -(f-g) \in C$. Hence $-(f-g) \in C \bigcap -C$. C is salient so $C \bigcap -C = \{0\}$; i.e., $-(f-g) = 0$. So $f = g$. Transitivity: if $f \leq g$ and $g \leq h$ then $f - g, g - h \in C$. By Proposition 1.3.0.3 (page 29), $(f - g) + (g - h) = f - h \in C$. So $f \leq h$. □

Definition 1.3.0.8. Suppose that $x, y \in C$, a salient convex cone emanating from the origin, we define:

$$m(y/x) = \sup\{t \mid tx \leq y\}$$

which is, from the definition of $\leq$, is equivalent to:

$$m(y/x) = \sup\{t \mid y - tx \in C\}$$

Definition 1.3.0.9.

$$b(y/x) = y - m(y/x)x \tag{1.8}$$

Remark 1 on $b(x/y)$: We will show that $b(y/x)$ is on the boundary of C, i.e, in ∂C, as well as on the boundary of $C \bigcap \text{Span}(x, y)$; i.e. in $\partial(C \bigcap \text{Span}(x, y))$.

Remark 2 on $b(x/y)$: In $\mathbf{R^2}$ it is traditional to write the equation of straight line as $y = mx + b$. Keeping this in mind, (1.8) is logically equivalent to

$$y = m(y/x)x + b(y/x).$$

Of course now $x, y, b(y/x)$ are vectors $\in V$. We will show that $m(y/x)$ can be interpreted as kind of slope and that $b(y/x)$ is a boundary intercept, something like the y-intercept in $\mathbf{R^2}$.

Example 1.3.0.10. Let $\mathbf{R} = V$ and $C = [0, \infty)$. If $x, y \in C$, with $x \neq 0$, then $m(y/x) = y/x$ and $b(y/x) = 0$.

1.3.1 A collection of results for $m(y/x), b(y/x)$

Proposition 1.3.1.1. *Let C be a closed salient convex cone emanating from the origin If $x, y \in C$ and $x \neq 0$ then:*

1. *If $\lambda \geq 0$ then $m(\lambda y/y) = \lambda$.*
2. *If $\lambda > 0$ then $m(y/\lambda y) = \frac{1}{\lambda}$.*
3. $0 \leq m(y/x) < \infty$
4. $b(y/x) = y - m(y/x)x \ \in C$
5. $b(y/x) \ \in \ \partial(C \bigcap Span(x, y)) \ \subset \ \partial(C) \bigcap Span(x, y) \subset \ C$
6. *If $x = \lambda y$ with $\lambda > 0$ then $b(y/x) = 0$*
7. *If x, y are linearly independent and $y \in \partial(C \bigcap Span(x, y))$ then $m(y/x) = 0$*
8. $m(y/x) = \max\{t \mid tx \leq y\} = \max\{t \mid y - tx \in C\}$

9. *If* $\alpha, \beta > 0$ *then* $m(\beta y/\alpha x) = \frac{\beta}{\alpha} m(y/x)$.

10. *If* $\alpha, \beta > 0$ *then* $b(\beta y/\alpha x) = \beta b(y/x)$.

Proof. (*1.*) Let $0 < \varepsilon$. Then

$$\begin{aligned} \lambda y - (\lambda + \varepsilon)y = -\varepsilon y &\notin C \quad \text{(as } C \text{ is salient)} \\ \lambda y - (\lambda - \varepsilon)y = \varepsilon y &\in C \quad \text{(as } C \text{ is closed w.r.t. non-negative scaling)} \end{aligned}$$

So $\lambda - \varepsilon \leq m(\lambda y/y) \leq \lambda + \varepsilon$. Letting $\varepsilon \to 0$ we get $m(\lambda y/y) = \lambda$.

(*2.*) This follows immediately from *1.* as, $m(y/\lambda y) = m(\frac{1}{\lambda}\lambda y/\lambda y) = \frac{1}{\lambda}$.

(*3.*) $0 \leq m(y/x)$ since $0 \in \{t \mid y - tx \in C\}$. On the other hand, if $m(y/x) = \infty$, then (*1*) and (*2*) imply y isn't a multiple of x. Moreover, if $m(y/x) = \infty$ then $\exists t_n \to \infty$ such that $y - t_n x \in C \ \forall n$. Since y isn't a multiple of x it follows that $\|y - t_n x\| \neq 0$. Then as C is closed under non-negative scaling, we have:

$$\frac{y - t_n x}{\|y - t_n x\|} \in C \ \forall n \quad \text{and} \quad \frac{x}{\|x\|} \in C.$$

Since C is topologically closed,

$$\lim_{n \to \infty} \frac{y - t_n x}{\|y - t_n x\|} = \frac{-x}{\|x\|} \in C.$$

Which contradicts C being salient. So $m(y/x) < \infty$.

(*4.*) Since $0 \leq m(y/x) < \infty$ we have $b(y/x) = y - m(y/x)x \in V$. Let t_n be a sequence in $\{t \mid y - tx \in C\}$ which converges to $m(y/x)$. So $y - t_n x \in C \ \forall n$. Since

$\|(y - t_n x) - (y - m(y/x)x)\| = |m(y/x) - t_n| \, \|x\|$, we have:

$$\lim_{n\to\infty} y - t_n x = y - m(y/x)x.$$

Since C is closed it follows that $y - m(y/x)x \in C$.

(*5.*) Let U be any open set about $b(y/x)$ and let $\varepsilon > 0$ be small enough so that that $B_\varepsilon(b(y/x)) \subset U$. Let

$$b' = y - \left(m(y/x) + \frac{\varepsilon}{2\,\|x\|} \right) x.$$

$b' \in B_\varepsilon(b(y/x))$ since:

$$\begin{aligned} \|b(y/x) - b'\| &= \left\| (y - m(y/x)x) - \left(y - \left(m(y/x) + \frac{\varepsilon}{2\,\|x\|} \right) x \right) \right\| \\ &= \left\| -\frac{\varepsilon}{2\,\|x\|} x \right\| \\ &= \frac{\varepsilon}{2}. \end{aligned}$$

$b' \notin C$ since:

$$\left(m(y/x) + \frac{\varepsilon}{2\,\|x\|} \right) > m(y/x).$$

Obviously, $b' \in \mathrm{Span}(x, y)$. So $b' \in U \cap \mathrm{Span}(x, y) - C$. On the other hand, by (*4*), $b(y/x) \in C$. so $b(x/y) \in U \cap C$. Since U was an arbitrary open set about $b(y/x)$ we've shown that:

$$b(y/x) \ \in \ \partial(C \bigcap \mathrm{Span}(x, y)).$$

Finally, by Proposition 1.1.1.8, we have:

$$\partial(C \bigcap \mathrm{Span}(x, y)) \ \subset \ (\partial C) \bigcap \mathrm{Span}(x, y).$$

Since C is closed, $\partial C \subset C$.

(*6.*) If $y = \lambda x$ then (*1*) implies $m(\lambda x/x) = \lambda$. Then

$$\begin{aligned} b(y/x) &= y - m(y/x)x \quad \text{so} \\ b(\lambda x/x) &= \lambda x - (\lambda)x = 0 \end{aligned}$$

(*7.*) $\forall(\alpha, \beta) \in \mathbf{R^2}$ let

$$f(\alpha, \beta) = \alpha x + \beta y.$$

Since x, y are linearly independent, Proposition 1.2.1.2 implies that

$$f : \mathbf{R^2} \mapsto \text{Span}(x, y)$$

is a homeomorphism. Let

$$\mathbf{R^2_{>0}} = \{(\alpha, \beta) \in \mathbf{R^2} \mid \alpha, \beta > 0\} = (0, \infty) \times (0, \infty).$$

Since $\mathbf{R^2_{>0}}$ is open in $\mathbf{R^2}$ and f is a homeomorphism,

$$f(\mathbf{R^2_{>0}}) = \{\alpha x + \beta y \mid \alpha, \beta > 0\}$$

is open in $\text{Span}(x, y)$.

C is closed under non-negative linear combinations (Proposition 1.3.0.3 (page 29)), so

$$f(\mathbf{R^2_{>0}}) \subset C.$$

To finish this proof it is convenient to introduce the map $h_{t,b}$.

The map $h_{t,b}(a) = a + t(b - a)$

Definition 1.3.1.2. For all $t \in \mathbf{R}$ and $a, b \in V$

$$h_{t,b}(a) = a + t(b - a)$$

The following lemma is stated without proof. Its proof follows easily from elementary results about normed vector spaces.

Lemma 1.3.1.3. *1. If* $0 \leq t < 1$ *and* $b \in V$*, then*

$$h_{t,b} : a \to a + t(b - a)$$

is a self homeomorphism of V*.*

2. If we fix $a, b \in V$ *with* $a \neq b$ *then*

$$h_{t,b}(a) : [0, 1] \to \overline{ab} \quad \textit{homeomorphically.}$$

Note: $\overline{ab}$ *is the line segment from* a *to* b*.*

3.

$$\begin{aligned} h_{0,b}(a) &= a + 0(b - a) = a \textit{ I.e. the identity map on } V. \\ h_{1,b}(a) &= a + 1(b - a) = b \textit{ I.e. a constant function.} \end{aligned}$$

4. If $c \in \mathcal{C}$*, a convex subset of* V*, then* $\forall\ t \in [0, 1]$

$$h_{t,c}(\mathcal{C}) \subset \mathcal{C}.$$

Proof of Proposition 1.3.1.1 continued. We return to our proof of Proposition 1.3.1.1 part 7 (page 31).

If $m(y/x) \neq 0$ then $y + m(y/x)x \in f(\mathbf{R}^{\mathbf{2}}_{>\mathbf{0}})$.

$f(\mathbf{R}^{\mathbf{2}}_{>\mathbf{0}})$ is contained in C and is a relatively open subset of $\mathrm{Span}(x, y)$. We solve

$$h_{t,b(y/x)}(y + m(y/x)x) = y$$

for t:

$$\begin{aligned} h_{t,b(y/x)}(y + m(y/x)x) &= y \Rightarrow \\ (y + m(y/x)x) + t[b(y/x) - (y + m(y/x)x)] &= y \Rightarrow \\ (y + m(y/x)x) + t[(y - m(y/x)x) - (y + m(y/x)x)] &= y \Rightarrow \\ y + m(y/x)x - 2tm(y/x)x &= y \Rightarrow \\ \frac{1}{2} &= t. \end{aligned}$$

So $h_{\frac{1}{2},b(y/x)}(y + m(y/x)x) = y$. This implies

$$y \in \underbrace{h_{\frac{1}{2},b(y/x)}(f(\mathbf{R}^{\mathbf{2}}_{>\mathbf{0}}))}_{\text{rel. open set in Span(x,y)}} \subset (C \cup \mathrm{Span}(x, y)).$$

Which implies $y \notin \partial(C \cup \mathrm{Span}(x, y))$.

We've shown that x, y linearly independent and $m(y/x) > 0$ implies

$$y \notin \partial(C \cup \mathrm{Span}(x, y)).$$

So if x, y linearly independent and $y \in \partial(C \cup \mathrm{Span}(x, y))$ then $m(y/x) = 0$.

(8.) Follows directly from *4.*

(*9.*) By *2.* we know that $y - m(y/x)x \in C$. Then since C is closed under non-negative scaling we have:

$$\begin{aligned} y - m(y/x)x \in C &\Rightarrow \beta y - \beta m(y/x)x \in C \\ &\Rightarrow \beta y - \beta m(y/x)\frac{\alpha x}{\alpha} \in C \\ &\Rightarrow \beta y - \frac{\beta}{\alpha} m(y/x)\alpha x \in C \\ &\Rightarrow \frac{\beta}{\alpha} m(y/x) \leq m(\beta y/\alpha x). \end{aligned} \tag{1.9}$$

An identical argument shows $\frac{\alpha}{\beta} m(\beta y/\alpha x) \leq m(y/x)$ and hence

$$m(\beta y/\alpha x) \leq \frac{\beta}{\alpha} m(y/x). \tag{1.10}$$

Combining (1.9) and (1.10) gives us $m(\beta y/\alpha x) = \frac{\beta}{\alpha} m(y/x)$.

(*10.*) Using *9.* and $b(y/x) = y - m(y/x)x$ we get

$$\begin{aligned} b\left(\beta y/\alpha x\right) &= \beta y - m(\beta y/\alpha x)\ \alpha x \\ &= \beta y - \frac{\beta}{\alpha} m(y/x)\ \alpha x \\ &= \beta y - \beta m(y/x)\ x \\ &= \beta\left(y - m(y/x)\ x\right) \\ &= \beta b\left(y/x\right) \end{aligned}$$

□

Corollary 1.3.1.4. *Let $x, y \in C$ be linearly independent. Then the following are equivalent.*

1. $m(y/x) = 0$

2. $y = b(y/x)$

3. $y \in \partial(C \cap \mathit{Span}(x, y))$

Proof. These are all immediate consequences of Proposition 1.3.1.1 (page 31). □

1.3.2 The slope $m(b_1, b_2; v)$

As usual, C is closed, salient, pointed convex cone in V. Let $x, y \in C$ and $x \neq 0$. Recall:

$$\begin{aligned} m(y/x) &= \sup\{t \mid y - tx \in C\} \\ b(y/x) &= y - m(y/x)x \qquad \text{so} \\ y &= m(y/x)x + b(y/x) \end{aligned}$$

Definition 1.3.2.1. Let H be a two dimensional vector subspace of V. Let b_1, b_2 be an ordered basis for H and let $v \in H$. So

$$v = v_1 b_1 + v_2 b_2 \qquad \text{with } v_1, v_2 \in \mathbf{R} \text{ uniquely determined.}$$

We define the slope of v relative to the ordered basis b_1, b_2 to be

$$m(b_1, b_2; v) = \frac{v_2}{v_1}.$$

If $v_1 = 0$ the we define $m(b_1, b_2; v) = \infty$. We don't distinguish between $\pm\infty$.

1.3.3 $b(x/y), b(y/x)$ are linearly independent

Lemma 1.3.3.1. *If $x, y \in C$ are linearly independent then $b(x/y), b(y/x)$ are linearly independent.*

Proof. To see that $b(x/y), b(y/x)$ are linearly independent it suffices to show that

$$b(x/y) = x - m(x/y)y$$

is not a positive multiple of

$$b(y/x) = y - m(y/x)x.$$

They can't be negative multiples of each other since C is salient. So suppose $\exists r > 0$ such that

$$x - m(x/y)y = r(y - m(y/x)x).$$

A little algebra shows this is equivalent to

$$(1 + rm(y/x))x - (r + m(x/y)y = 0. \tag{1.11}$$

Since x and y are linearly independent both the coefficients appearing in (1.11), $(1 + rm(y/x))$ and $-(r + m(x/y)$, must be zero. In particular, we must have

$$r + m(x/y) = 0.$$

But then $r = -m(x/y) \leq 0$, a contradiction since in Proposition 1.3.1.1 (page 31) we proved that

$$0 \leq m(x/y) < \infty.$$

So $b(x/y)$ and $b(y/x)$ must be linearly independent. □

1.3.4 $m(b(y/x), b(x/y); y) = \frac{y_{b(x/y)}}{y_{b(y/x)}} = m(y/x)$

Proposition 1.3.4.1. *Let $x, y \in C$ be linearly independent. Then the slope of y relative to the ordered basis $b(x/y), b(y/x)$ is $m(y/x)$. More succinctly:*

$$m(b(y/x), b(x/y); y) = \frac{y_{b(x/y)}}{y_{b(y/x)}} = m(y/x)$$

Proof. x, y are linearly independent. So, by Lemma 1.3.3.1 (page 38), $b(x/y), b(y/x)$ are linearly independent, and form an ordered basis for Span(x, y). We express y (uniquely) in terms of $b(x/y), b(y/x)$ and do the obvious:

$$\begin{aligned} y &= y_{b(x/y)} b(x/y) + y_{b(y/x)} b(y/x) \quad (\text{with } y_{b(x/y)}, y_{b(y/x)} \in \mathbf{R}) \\ &= y_{b(x/y)}(x - m(x/y)y) + y_{b(y/x)}(y - m(y/x)x) \\ &= \underbrace{(y_{b(x/y)} - y_{b(y/x)} m(y/x))}_{=0} x + \underbrace{(y_{b(y/x)} - y_{b(x/y)} m(x/y))}_{=1} y \end{aligned} \tag{1.12}$$

But

$$y_{b(x/y)} - y_{b(y/x)} m(y/x) = 0 \quad \Rightarrow \quad \frac{y_{b(x/y)}}{y_{b(y/x)}} = m(y/x).$$

□

In the following lemma we use the notation and concepts introduced in Proposition 1.3.4.1 (page 40).

Lemma 1.3.4.2. *If $x, y \in C$ are linearly independent then $y_{b(y/x)} > 0$.*

Proof. In Proposition 1.3.1.1 (page 31) we proved that $0 \leq m(x/y) < \infty$. In Proposition 1.3.4.1 (page 40) we proved that $\frac{y_{b(x/y)}}{y_{b(y/x)}} = m(y/x)$. Combining these two propositions and noting that $y \neq 0$ and that $y_{b(x/y)}, y_{b(y/x)} \in \mathbf{R}$ it follows that $y_{b(y/x)} \neq 0$.

If $0 < m(x/y) < \infty$ then $y_{b(x/y)}$ and $y_{b(y/x)}$ have the same sign. But

$$y = y_{b(x/y)} b(x/y) + y_{b(y/x)} b(y/x) \in C$$

and C is salient, so it must be the case that both $y_{b(x/y)}$ and $y_{b(y/x)}$ are positive.

If $0 = m(x/y)$ then $y_{b(x/y)} = 0$ and so $y = y_{b(y/x)}b(y/x)$. But $b(y/x)$ is non-zero and $\in C$, which is salient, so it again must be the case, that $y_{b(y/x)}$ is positive. □

The following innocent looking technical proposition is of key importance.

1.3.5 $0 \leq m(y/x)m(x/y) < 1$

Proposition 1.3.5.1. *If $x, y \in C$ are linearly independent then*

$$0 \leq m(y/x)m(x/y) < 1$$

Proof. Combining Proposition 1.3.3.1 (page 38) and Lemma 1.3.4.2 (page 40) yields

$$y_{b(x/y)} = y_{b(y/x)}m(y/x). \tag{1.13}$$

Combining (1.13) with equation (1.12), from the proof of Lemma 1.3.3.1, yields:

$$\begin{aligned} y_{b(y/x)} - y_{b(x/y)}m(x/y) &= 1 \\ y_{b(y/x)} - y_{b(y/x)}m(y/x)m(x/y) &= 1 \\ y_{b(y/x)}(1 - m(y/x)m(x/y)) &= 1 \end{aligned}$$

By Lemma 1.3.4.2, we know that $0 < y_{b(y/x)} < \infty$. Hence,

$$1 - m(y/x)m(x/y) = \frac{1}{y_{b(y/x)}} > 0.$$

In particular,

$$\begin{aligned} 1 - m(y/x)m(x/y) &> 0 \\ 1 &> m(y/x)m(x/y) \end{aligned}$$

On the other hand, $m(y/x), m(x/y)$ are non-negative by Proposition 1.3.1.1 (page 31). Hence $1 > m(y/x)m(x/y) \geq 0$. □

1.4 $\mathbf{Span}(x, y) \bigcap C = \{\alpha b(y/x) + \beta b(x/y) \mid \alpha, \beta \geq 0\}$

Theorem 1.4.0.2. *Let $x, y \in C$ be linearly independent, then*

$$\mathit{Span}(x, y) \bigcap C = \{\alpha b(y/x) + \beta b(x/y) \mid \alpha, \beta \geq 0\}$$

Proof. By Lemma 1.3.3.1 (page 38), $b(y/x), b(x/y)$ are linearly independent. Since $b(y/x), b(x/y) \in \text{Span}(x, y)$ it is immediate that $\text{Span}(x, y) = \text{Span}(b(y/x), b(x/y))$. So

$$\text{Span}(x, y) \bigcap C = \text{Span}(b(y/x), b(x/y)) \bigcap C.$$

So it suffices to prove the result for $\text{Span}(b(y/x), b(x/y)) \bigcap C$.

Since C is a cone and $b(y/x), b(x/y) \in C$, all non-negative linear multiples of $b(y/x), b(x/y)$ are $\in C$. Hence,

$$\{\alpha b(y/x) + \beta b(x/y) \mid \alpha, \beta \geq 0\} \subseteq \text{Span}(x, y) \bigcap C.$$

Conversely, suppose $w \in \text{Span}(x, y) \bigcap C$. Then w can be written in the form:

$$w = \alpha b(y/x) + \beta b(x/y).$$

Both α and β can't be negative, since that would obviously violate C being salient. In fact, if one of α, β is negative and the other is zero, that would also obviously violate C being salient, since $b(x/y)$ and $b(y/x)$ are $\in C$. So the only difficult case is if one of α, β is negative and the other is positive.

So suppose that

$$w = (-\alpha_w)b(y/x) + \beta_w b(x/y) \in C.$$

with both α_w and β_w positive. This will lead to a contradiction of Proposition 1.3.5.1 (page 41), as the following calculations will show.

$$\begin{aligned} w &= (-\alpha_w)b(y/x) + \beta_w b(x/y) \\ &= (-\alpha_w)(y - m(y/x)x) + \beta_w(x - m(x/y)y) \\ &= (\underbrace{\alpha_w m(y/x) + \beta_w}_{>0})x - (\underbrace{\alpha_w + \beta_w m(x/y)}_{>0})y \end{aligned}$$

Since C is closed under non-negative scaling we have:

$$\frac{1}{\alpha_w m(y/x) + \beta_w} w = x - \frac{\alpha_w + \beta_w m(x/y)}{\alpha_w m(y/x) + \beta_w} y \in C.$$

Recall that $m(x/y) = \sup\{t \in \mathbf{R} \mid x - ty \in C\}$. So we must have:

$$\frac{\alpha_w + \beta_w m(x/y)}{\alpha_w m(y/x) + \beta_w} \leq m(x/y). \tag{1.14}$$

Multiplying both sides of the inequality in (1.14) by $(\alpha_w m(y/x) + \beta_w) > 0$ indicates:

$$\begin{aligned} \alpha_w + \beta_w m(x/y) &\leq m(x/y)(\alpha_w m(y/x) + \beta_w) \\ \alpha_w + \beta_w m(x/y) &\leq m(x/y)\alpha_w m(y/x) + \beta_w m(x/y) \\ \alpha_w &\leq m(x/y)m(y/x)\alpha_w \quad (\text{but } \alpha_w > 0 \text{ so}) \\ 1 = \frac{\alpha_w}{\alpha_w} &\leq m(x/y)m(y/x) \end{aligned} \tag{1.15}$$

Inequality (1.15) contradicts Proposition 1.3.5.1 (page 41), which states:

$$0 \leq m(x/y)m(y/x) < 1.$$

Finally, an identical argument shows that if

$$w = \alpha_w b(y/x) + (-\beta_w) b(x/y) \in C.$$

with both α_w and β_w positive, then Proposition 1.3.5.1 is again violated. □

1.4.1 $\mathbf{ray}(b(x/y)) \bigcup \mathbf{ray}(b(y/x)) = \partial\left(\mathbf{Span}(x,y) \bigcap C\right)$

Recall:

Definition 1.1.0.2: The closed ray emanating from the origin and passing through the point v, is denoted $\overrightarrow{v}$. As a point set:

$$\overrightarrow{v} = \{\lambda v \mid \lambda \geq 0\}.$$

For typographic reasons, $\overrightarrow{v}$ is occasionally denoted ray(v).

Corollary 1.4.1.1. *If $x, y \in C$ are linearly independent then*

$$\overrightarrow{b(x/y)} \bigcup \overrightarrow{b(y/x)} = \partial\left(\mathit{Span}(x,y) \bigcap C\right)$$

Proof. By Lemma 1.3.3.1, $b(x/y)$ and $b(y/x)$ are linearly independent. Since $b(x/y)$ and $b(y/x)$ are $\in \mathrm{Span}(x,y)$ it follows that

$$\begin{aligned} \mathrm{Span}(x,y) &= \mathrm{Span}(b(x/y), b(y/x)) \\ &= \{\alpha b(x/y) + \beta\ b(y/x) \mid \alpha, \beta \in \mathbf{R}\}. \end{aligned}$$

The linear independence of $b(x/y)$ and $b(y/x)$ together with Proposition 1.2.1.2 imply

that:

$$\begin{aligned} \kappa : \mathrm{Span}(x,y) &\rightarrow \mathbf{R^2} \text{ by} \\ \kappa(\alpha b(x/y) + \beta\ b(y/x)) &= (\alpha, \beta) \end{aligned}$$

is a homeomorphism and an isomorphism. By Theorem 1.4.0.2

$$\mathrm{Span}(x,y) \bigcap C = \{\alpha b(x/y) + \beta\ b(y/x) \mid \alpha, \beta \geq 0\}.$$

Then,

$$\begin{aligned} \kappa\left(\mathrm{Span}(x,y) \bigcap C\right) &= \{\alpha\kappa(b(x/y)) + \beta\ \kappa(b(y/x)) \mid \alpha, \beta \geq 0\} \\ &= \{\alpha(1,0) + \beta\ (0,1) \mid \alpha, \beta \geq 0\} \\ &= \mathbf{R^2_{\geq 0}}. \end{aligned}$$

From elementary topology we know that:

$$\partial \mathbf{R^2_{\geq 0}} = \overrightarrow{(1,0)} \bigcup \overrightarrow{(0,1)}.$$

I.e.

$$\begin{aligned} \overrightarrow{(1,0)} &= \{(\lambda, 0\} \in \mathbf{R^2} \mid \lambda \geq 0\}. \\ \overrightarrow{(0,1)} &= \{(0, \lambda\} \in \mathbf{R^2} \mid \lambda \geq 0\}. \end{aligned}$$

Finally, since homeomorphisms take boundaries to boundaries; i.e.

$$\kappa^{-1}\partial = \partial\kappa^{-1},$$

and since isomorphisms take non-trivial 0-rays to non-trivial 0-rays, we have:

$$\begin{aligned}
\kappa^{-1}\partial\left(\mathbf{R}^{\mathbf{2}}_{\geq\mathbf{0}}\right) &= \kappa^{-1}\overrightarrow{(1,0)}\bigcup\kappa^{-1}\overrightarrow{(0,1)} \\
\partial\kappa^{-1}\left(\mathbf{R}^{\mathbf{2}}_{\geq\mathbf{0}}\right) &= \overrightarrow{\kappa^{-1}(1,0)}\bigcup\overrightarrow{\kappa^{-1}(0,1)} \\
\partial\Big(\mathrm{Span}(x,y)\bigcap C\Big) &= \overrightarrow{b(x/y)}\bigcup\overrightarrow{b(y/x)} \qquad (1.16)
\end{aligned}$$

Note regarding (1.16): $\partial\left(\mathrm{Span}(x,y)\bigcap C\right)$ is the boundary of $(\mathrm{Span}(x,y)\bigcap C)$ relative to $\mathrm{Span}(x,y)$. □

1.4.2 Results for $\mathbf{Span}(x,y)\bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha,\beta \geq 0\}$

Proposition 1.4.2.1. *Let $x, y \in C$ be linearly independent, then the following are equivalent statements about $b_1, b_2 \in V$:*

1. *$\{b_1, b_2\} = \{\lambda_1 b(x/y), \lambda_2 b(y/x)\}$ for some particular pair of numbers $\lambda_1, \lambda_2 > 0$. I.e. b_1 is a positive multiple of one of $b(x/y)$ or $b(y/x)$ and b_2 is a positive multiple of the other.*

2. *$Span(x,y)\bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha,\beta \geq 0\}$.*

3. *$\overrightarrow{b_1}\bigcup\overrightarrow{b_2} = \partial\left(Span(x,y)\bigcap C\right)$*

4. *$b_1, b_2 \in \partial\left(Span(x,y)\bigcap C\right)$ and are linearly independent*

Proof. *1 ⇒ 2.* Trivial consequence of Theorem 1.4.0.2 (page 42).

2 ⇒ 3. Since x, y are linearly independent and $\in C$, b_1 and b_2 must also be linearly independent. Then the same proof as given for Corollary 1.4.1.1 (page 44).

3 ⇒ 4. All that needs to be shown is that b_1 and b_2 are linearly independent. $b(x/y), b(y/x) \in \partial\left(\mathrm{Span}(x,y)\bigcap C\right)$ and they are linearly independent. So $b(x/y)$ and $b(y/x)$ can't both be in the same 0-ray, $\overrightarrow{b_1}$ (or $\overrightarrow{b_2}$). Hence b_1 and b_2 must be linearly independent.

4 ⇒ 1. By Corollary 1.4.1.1 (page 44), $\overrightarrow{b(x/y)} \bigcup \overrightarrow{b(y/x)} = \partial\left(\text{Span}(x,y) \bigcap C\right)$. b_1 and b_2 are linearly independent so they both can't be in the same 0-ray, $\overrightarrow{b(x/y)}$ (or $\overrightarrow{b(y/x)}$). The result follows. □

1.4.3 Definition of ends

Definition 1.4.3.1. Let $x, y \in C$ be linearly independent. Any pair b_1, b_2 as described in Proposition 1.4.2.1 (page 46), are called ends for x, y.

Remark 1.4.3.2. Later, in Proposition 1.6.1.1 (page 53) we will prove that if $f, g \in C$ are linearly independent then there exists a pair of ends b_1, b_2 for f, g.

Proposition 1.4.3.3. *Suppose $\alpha, \beta > 0$. If b_1, b_2 are a pair of ends for x, y then b_1, b_2 are a pair of ends for $\alpha x, \beta y$.*

Proof. Definition 1.4.3.1 (page 47) of ends and Proposition 1.4.2.1 part 1 (page 46) tell us that if b_1, b_2 are a pair of ends for x, y then there exists $\lambda_1, \lambda_2 > 0$ such that

$$\{b_1, b_2\} = \{\lambda_1 b(x/y),\ \lambda_2 b(y/x)\}.$$

By part 10 of Proposition 1.3.1.1 (page 31)

$$b(\beta y/\alpha x) = \beta b(y/x) \qquad \text{and} \qquad b(\alpha x/\beta y) = \alpha b(x/y).$$

So

$$\{b_1, b_2\} = \left\{ \frac{\lambda_1}{\alpha}\ b(\alpha x/\beta y),\ \frac{\lambda_2}{\beta}\ b(\beta y/\alpha x) \right\},$$

which implies b_1, b_2 are a pair of ends for $\alpha x, \beta y$. □

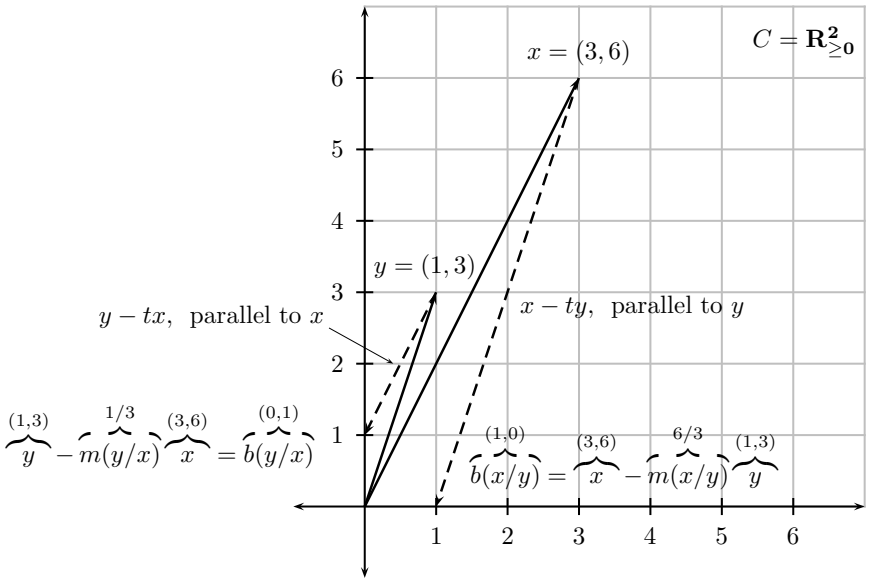


Figure 1.1: Calculating $m(y/x)$ in $\mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}}$, the standard 2 dimensional cone.

1.4.4 Calculating $m(y/x)$ in the cone $C = \mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}}$

Example 1.4.4.1. See Figure 1.1. Let $V = \mathbf{R^2}$; $x = (3,6)$; $y = (1,3)$; and

$$C = \mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}} = \underbrace{\{(\alpha, \beta) \in \mathbf{R^2} \mid \alpha, \beta \geq 0\}}_{[0,\infty)\times[0,\infty)};$$

Then

$$\begin{aligned} m(y/x) &= \sup\{t \mid y - tx \in C\} \\ &= \sup\{t \mid (1,3) - t(3,6) \in \mathbf{R}^{\mathbf{2}}_{\geq \mathbf{0}}\} \\ &= \sup\{t \mid 1 - 3t \geq 0 \text{ and } 3 - 6t \geq 0\} \\ &= 1/3. \end{aligned} \tag{1.17}$$

Similarly,

$$\begin{aligned} m(x/y) &= \sup\{t \mid x - ty \in C\} \\ &= \sup\{t \mid (3,6) - t(1,3) \in \mathbf{R^2_{\geq 0}}\} \\ &= \sup\{t \mid 3 - t \geq 0 \text{ and } 6 - 3t \geq 0\} \\ &= 6/3. \end{aligned} \tag{1.18}$$

According to Proposition 1.3.5.1,

$$0 \leq m(y/x)m(x/y) < 1.$$

In our example:

$$m(y/x)m(x/y) = \frac{1}{3}\,\frac{6}{3} = \frac{2}{3}.$$

Next we calculate $b(y/x)$ and $b(x/y)$.

$$\begin{aligned} b(y/x) &= y - m(y/x)x \\ &= (1,3) - 1/3(3,6) = (0,1) \\ b(x/y) &= x - m(x/y)y \\ &= (3,6) - 6/3(1,3) = (1,0) \end{aligned}$$

Note: x and y were precisely chosen so that $b(x/y)$ would equal $(1,0)$ and $b(y/x)$ would equal $(0,1)$. Usually $b(x/y)$ and $b(y/x)$ do not equal $(1,0)$ or $(0,1)$.

We express x and y in terms of $b(x/y)$ and $b(y/x)$:

$$\begin{aligned} \underbrace{x}_{(3,6)} &= \underbrace{x_{b(x/y)}}_{3}\underbrace{b(x/y)}_{(1,0)} + \underbrace{x_{b(x/y)}}_{6}\underbrace{b(y/x)}_{(0,1)} \\ \underbrace{y}_{(1,3)} &= \underbrace{y_{b(x/y)}}_{1}\underbrace{b(x/y)}_{(1,0)} + \underbrace{y_{b(x/y)}}_{3}\underbrace{b(y/x)}_{(0,1)} \end{aligned}$$

By Definition 1.3.2.1,

$$m(b(y/x), b(x/y); y) = \frac{y_{b(x/y)}}{y_{b(y/x)}} \tag{1.19}$$

and

$$m(b(x/y), b(y/x); x) = \frac{x_{b(y/x)}}{x_{b(x/y)}} \tag{1.20}$$

In our example, equation (1.19) becomes

$$m(\underbrace{b(y/x)}_{(0,1)}, \underbrace{b(x/y)}_{(1,0)}; \underbrace{y}_{(1,3)}) = \frac{y_{b(x/y)}}{y_{b(y/x)}} = \frac{1}{3},$$

which equals $m(y/x)$, see (1.17); equation (1.20) becomes

$$m(\underbrace{b(x/y)}_{(1,0)}, \underbrace{b(y/x)}_{(0,1)}; \underbrace{x}_{(3,6)}) = \frac{x_{b(y/x)}}{x_{b(x/y)}} = \frac{6}{3},$$

which equals m(x/y), see (1.18) Of course this exactly what is predicted by Proposition 1.3.4.1 (page 40), that if $x, y \in C$ are linearly independent, then:

$$\begin{aligned} m(b(x/y), b(y/x); x) = \frac{x_{b(y/x)}}{x_{b(x/y)}} &= m(x/y) \\ m(b(x/y), b(y/x); x) = \frac{x_{b(y/x)}}{x_{b(x/y)}} &= m(x/y). \end{aligned}$$

1.5 Some results for $f \in C \setminus \partial C$ and $b \in \partial C$

Proposition 1.5.0.2. *Let C be a closed, convex, salient, pointed by the origin cone in a Banach Space V. Let $f \in C \setminus \partial C$ and $b \in \partial C$, so $b = 0$ is acceptable. Then*

1. *If $c \in C$ and $\alpha \in (0, \infty)$ then $c + f \in C \setminus \partial C$*
2. *If $\alpha \in (0, \infty)$ then αf and $c + \alpha f \in C \setminus \partial C$.*
3. $$b - tf \in C \Leftrightarrow t \leq 0.$$

4.

$$m(b/f) = \sup\{t \mid b - tf \in C\} = 0$$
$$b(b/f) = b - m(b/f)f = b.$$

5. If $\beta > 0$ then $\beta b \in \partial C$.

6. f, b are linearly independent if $b \neq 0$.

7. $m(b/f)\ m(f/b) = 0$.

Proof. 1. $f + 2c \in C \subset V$ so the map

$$h_{t,f+2c} : v \in V \to v + t(f + 2c - v)$$

is a self-homeomorphism of V provided $0 \leq t < 1$. See Lemma 1.3.1.3. If $0 \leq t < 1$ then $h_{t,f+2c}(\text{Span}(C)) = \text{Span}(C)$. If $0 \leq t \leq 1$ then $h_{t,f+2b}(C) \subset C$.

Since $f \in C \setminus \partial C$ there exists an open set U containing f such that

$$U \cap (\text{Span}(C) \setminus C) = \emptyset.$$

But then, for $0 \leq t < 1$ we have

$$\begin{aligned}
h_{t,f+2c}\left(U \cap (\text{Span}(C) \setminus C)\right) &= h_{t,f+2c}(\emptyset) = \emptyset \\
h_{t,f+2c}(U) \cap h_{t,f+2c}\left(\text{Span}(C) \setminus C\right) &= \emptyset \\
h_{t,f+2c}(U) \cap \left(h_{t,f+2c}\left(\text{Span}(C)\right) \setminus h_{t,f+2c}\left(C\right)\right) &= \emptyset \\
\underbrace{h_{t,f+2c}(U)}_{\text{open}} \cap \left(\text{Span}(C) \setminus \underbrace{h_{t,f+2c}\left(C\right)}_{\subset C}\right) &= \emptyset \\
\underbrace{h_{t,f+2c}(U)}_{\text{open}} \cap (\text{Span}(C) \setminus C) &= \emptyset \qquad (1.21)
\end{aligned}$$

Let $t = 1/2$ then

$$h_{1/2,f+2c}(f) = f + (1/2)(f + 2c - f) = c + f$$

and (1.21) tell us that

$$\underbrace{h_{1/2,f+2c}(U)}_{\text{open set about } c+f} \cap (\text{Span}(C) \setminus C) = \emptyset.$$

So that $c + f \notin \partial C$.

proof of 2. Similar to the proof of 1, but instead of using $h_{1/2,f+2c}$, use the self homeomorphism of V

$$h_\alpha : v \in V \to \alpha v \tag{1.22}$$

which, like $h_{1/2,f+2c}$, has the properties that $h_\alpha(\text{Span}(C)) = \text{Span}(C)$ and $h_\alpha(C) \subset C$. In fact $h_\alpha(C) = C$. Then use the same argument as given in in the proof of part 1 to show that $\alpha f \in C \setminus \partial C$. Then, by part 1 of this proposition $c + \alpha f \in C \setminus \partial C$.

proof of 3. Suppose that $b - tf \in C$ for some particular $t > 0$. By part 2 of this proposition, letting $t = \alpha$,

$$b = (b - tf) + (tf) \in C \setminus \partial C.$$

This contradicts that $b \in \partial C$.

Conversely, if $t \leq 0$ then $t = -|t|$ and

$$b - tf = b - -|t|f = b + |t|f \in C.$$

proof of 4. Trivial consequence of part 3 of this proposition.

proof of 5. Since C is closed and $b \in \partial C$ it follows that $b \in C$. Since $\beta > 0$ and $b \in C$ it follows that $\beta b \in C$. If $\beta b \notin \partial C$ then $\beta b \in C \setminus \partial C$. But then by part 2 of

this proposition, with β playing the role of α,

$$\frac{1}{\beta}\ \beta b = b \in C \setminus \partial C.$$

Which contradicts that $b \in \partial C$. So $\beta b \in \partial C$.

proof of 6. Since $f \in C \setminus \partial C$ it follows that $f \neq 0$. Since C is closed and $b \in \partial C$ it follows that $b \in C$. If f, b are linearly dependent and $b \neq 0$ then

$$b = \alpha f \tag{1.23}$$

for some $\alpha > 0$. We must have $\alpha \geq 0$ since C is salient and $f, b \in C$. That $\alpha \neq 0$ follows from our assumption that $b \neq 0$.

By part 2 of this proposition $\alpha f \in C \setminus \partial C$ but then (1.23) contradicts $b \in \partial C$.

proof of 7. By part 3 of Proposition 1.3.1.1 (page 31) if $x, y \in C$ and $x \neq 0$ then

$$0 \leq m(y/x) < \infty. \tag{1.24}$$

Since $b \neq 0$, if follows from (1.24) that $0 \leq m(f/b) < \infty$. By part 4 of this proposition $m(b/f) = 0$. So $m(b/f)m(f/b) = 0$. □

1.6 The Hilbert Projective Metric d_H

1.6.1 $f, g \in C$ linearly independent $\Rightarrow \exists$ ends b_1, b_2

Proposition 1.6.1.1. *Let $f, g \in C$ be linearly independent. Then there exists*

$$b_1, b_2 \in \partial \left(Span(f, g) \bigcap C \right) \subset \partial C \cap Span(f, g), \tag{1.25}$$

linearly independent, such that

$$Span(f,g)\bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha, \beta \geq 0\} \tag{1.26}$$

So b_1, b_2 are ends for f, g.

Note that the subscripts $1, 2$ represent the ordering that we are giving to the basis vectors $\{b_1, b_2\}$.

Proof. Direct consequence of Theorem 1.4.0.2, Proposition 1.4.2.1 (page 46) and Proposition 1.1.1.8 (page 23). □

1.6.2 Definition of d_H via b_1, b_2.

Equation (1.26) implies that we can represent f and g uniquely in terms of b_1, b_2. In particular, that

$$\begin{aligned} f &= f_1 b_1 + f_2 b_2 \\ g &= g_1 b_1 + g_2 b_2 \end{aligned} \tag{1.27}$$

with $f_1, f_2, g_1, g_2 \geq 0$ and uniquely determined. Based upon this representation, we have the following definition (which is shown to be well defined in Proposition 1.6.4.3 (page 58):

Definition 1.6.2.1. For $f, g \in C$ and linearly independent,

$$d_H(f,g) = \left| \ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| \in \mathbf{R} \cup \infty.$$

where we define $|\ln(0)| = |\ln(\infty)| = \infty$.

If $f, g \in C \setminus \{0\}$ are linearly dependent we define

$$d_H(f,g) = 0.$$

So d_H is defined on in $C \setminus \{0\}$.

Remark 1.6.2.2. We directly define $d(f, g) = 0$ to be zero when f and g are linearly dependent to avoid having to deal with the following issue:

When f and g are linearly dependent it means that $f = \lambda g$ for some $\lambda > 0$. Then $\text{Span}(f, g)$ is a line and (1.26) does not hold. Of course if $b_1, b_2 \in \partial C$ are linearly independent and $f, g \in \text{Span}(b_1, b_2)$ then (1.27) would hold uniquely and we would have

$$\left|\ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| = \left|\ln\left(\frac{f_2}{f_1}\frac{\lambda f_1}{\lambda f_2}\right)\right| = 0.$$

1.6.3 What about 0?

Unless otherwise noted, we will assume that d_H is only defined on $C \setminus \{0\}$. As that is the standard. See for example Bushell [19].

However, we take a few moments to suggest a way to extend d_H from $C \setminus \{0\}$ to all of C.

According to Bushell [19], $(\mathbb{R}^n_{\geq 0}, \sim)$ will be a complete metric space with d_H taking values in $[0, \infty]$. In our Section 1.12.3 (page 123) we prove some general results about when $(C \setminus \{0\}, \sim)$ will be complete. Moreover, one of the key parts in our exposition is Birkhoff's Projective Contraction Mapping Theorem [12] – and that result relies on $(C \setminus \{0\}, \sim)$ being complete w.r.t. d_H.

So if we are going to extend d_H to 0; i.e. if we are going define $d_H(0, f)$ for $f \in C \setminus \{0\}$ in such a way that d_H remains metric, then we are going to have to deal with $(C \setminus \{0\}, \sim)$ being complete.

Proposition 1.6.3.1. *If $(C \setminus \{0\}, \sim)$ is complete w.r.t. d_H and d_H is to be extended to 0 then there exists a $k > 0$ such that $d_H([0], [f]) > k$ for all $f \in C \setminus \{0\}$.*

Proof. Suppose not. Then there exists a sequence of $[f_n] \in (C \setminus \{0\}, \sim)$ such that $d_H([0], [f_n]) < 2^{-n}$. The sequence of $[f_n]$ is cauchy and in $(C \setminus \{0\}, \sim)$ and so there

exists an $[f] \in (C \setminus \{0\}, \sim)$ to which the sequence of $[f_n]$ converges. On the other hand, clearly the sequence of $[f_n]$ converges to $[0]$. If d_H is a metric (which separates points and has unique limits) then $[0] = [f]$, which is a contradiction. $\square$

One way we could extend d_H to 0 would be to pick some $c \in C \setminus \{0\}$ and some fixed $k \in (0, \infty]$ and to define for each $f \in C \setminus \{0\}$

$$\begin{aligned} d_H(0, f) &= k + d_H(c, f) \\ d_H(0, 0) &= 0 \end{aligned} \tag{1.28}$$

Intuitively we can think of this extension as if 0 is an island off the coast of $C \setminus \{0\}$ and that to reach any point $f \in C \setminus \{0\}$ we always have to take a bridge of fixed length k which connects 0 to $C \setminus \{0\}$ at the point c.

Proposition 1.6.3.2. *(1.28) extends d_H to all of C*

Proof. Let $f, h \in C \setminus \{0\}$. The triangle inequality follows from

$$d_H(0, h) = k + d_H(c, h) \leq k + d(c, f) + d(f, h) = d(0, f) + d(f, h)$$

and

$$d(f, h) \leq d(f, c) + d(c, h) < k + d(f, c) + k + d(c, h) = d(f, 0) + d(0, h)$$

The other axioms are trivial. $\square$

Remark 1.6.3.3. 1. The problem with using (1.28) to extend d_H to all of C is that it requires us to single out a special $[c] \in (C \setminus \{0\}, \sim)$ that will be the closest point to 0. However I can not think of a particularly compelling reason why any one particular 0-ray should be considered the projectively closest point to $[0]$. That said, perhaps a choice for c would be an axis of symmetry or center

of mass. The main purpose of (1.28) is to show that an easily definable non-constant extension of d_H exists.

2. By Proposition 1.6.3.1 any metric which extends d_H to include the origin will make the origin an isolated point.

3. According to Bushell [19], the projective d_H distance from the "interior of C" to its "boundary" excluding 0 is ∞. We prove this result in Proposition 1.6.6.1 (page 62). Since $0 \in \partial C$, this result suggests one define $d_H(f, 0) = \infty$ whenever $f \in C \setminus \partial C$.

4. Similarly, suppose $f, g \in C$ are linearly independent. Let $b_1, b_2 \in \partial(\text{Span}(f, g) \cap C)$ be linearly independent (and so be a pair of ends for f, g). By Proposition 1.6.4.4 (page 59) $d_H(b_1, b_2) = \infty$. The triangle inequality

$$d_H(b_1, b_2) \leq d_H(b_1, 0) + d_H(0, b_2)$$

implies at least one of $d_H(b_1, 0), d_H(0, b_2)$ is ∞. This suggests one define $d_H(b, 0) = \infty$ if $b \neq 0 \in \partial C$.

We will only occasionally make use of the following extension of d_H.

Definition 1.6.3.4. Based upon the above Remark 1.6.3.3, we will define $d_H(f, g)$ for when one or both of f, g are zero as follows. If $f \in C \setminus \{0\}$ then then $d_H(f, 0) = d_H(0, f) = \infty$ and $d_H(0, 0) = 0$.

We wish to emphasize that unless otherwise noted $d_H(f, g)$ will not be defined if f or g is 0.

1.6.4 d_H is well defined and $d_H(b_1, b_2) = \infty$

Proposition 1.6.4.1. *If f, g are linearly independent then $\frac{f_2}{f_1}\frac{g_1}{g_2}$ is not indeterminate; i.e. it can't be equal to $0 \cdot \infty$ or $\infty \cdot 0$ or contain a $\frac{0}{0}$.*

Proof. If $\frac{f_2}{f_1}\frac{g_1}{g_2} = 0 \cdot \infty$ then $f_2 = g_2 = 0$ implying that f, g are linearly dependent.

Similarly, if $\frac{f_2}{f_1}\frac{g_1}{g_2} = \infty \cdot 0$ then $f_1 = g_1 = 0$ implying that f, g are linearly dependent.

If $\frac{f_2}{f_1}$ (or $\frac{g_1}{g_2}$) is of the form $\frac{0}{0}$ then $f = 0$ (or $g = 0$)and 0 is linearly dependent on all other vectors. □

Remark 1.6.4.2. If f and g are linearly dependent and $f_1, f_2, g_1, g_2 > 0$ then $\lambda f_1 = g_1$ and $\lambda f_2 = g_2$ for some $\lambda > 0$. Then

$$\frac{f_2}{f_1}\frac{g_1}{g_2} = \frac{f_2}{f_1}\frac{\lambda f_1}{\lambda f_2} = 1$$

and

$$\left|\ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| = |\ln(1)| = 0 \tag{1.29}$$

which agrees with our defining $d_H(f, g) = 0$ for the case f and g linearly dependent and non-zero.

Proposition 1.6.4.3. *$d_H(f, g)$ is well defined; i.e. it is independent of the choice of b_1, b_2 used in its definition, provided that b_1 and b_2 are $\in \partial\left(Span(f, g) \bigcap C\right)$ and are linearly independent.*

Proof. Suppose that b_1' and b_2' are $\in \partial\left(\text{Span}(f, g) \bigcap C\right)$ and are linearly independent. Proposition 1.4.2.1 (item *1*) (page 46) applied twice, once to $\{b_1, b_2\}$ and once to $\{b_1', b_2'\}$, implies that as sets

$$\{b_1, b_2\} = \{\lambda_1 b_1', \lambda_2 b_2'\},$$

with $\lambda_1, \lambda_2 > 0$. So exactly one of the following two cases will occur:

1. $b_1 = \lambda_1 b_1'$ and $b_2 = \lambda_2 b_2'$
2. $b_1 = \lambda_2 b_2'$ and $b_2 = \lambda_1 b_1'$

If case *1* occurs then,

$$\begin{aligned} f &= f_1 b_1 + f_2 b_2 = f_1 \lambda_1 b'_1 + f_2 \lambda_2 b'_2 \\ g &= g_1 b_1 + g_2 b_2 = g_1 \lambda_1 b'_1 + g_2 \lambda_2 b'_2 \end{aligned}$$

which in turn implies that w.r.t. the ordered basis b'_1, b'_2 that:

$$\begin{aligned} d_H(f,g) &= \left| \ln \left(\frac{f_2 \lambda_2}{f_1 \lambda_1} \frac{g_1 \lambda_1}{g_2 \lambda_2} \right) \right| \\ &= \left| \ln \left(\frac{f_2}{f_1} \frac{g_1}{g_2} \right) \right| \\ &= d_H(f,g) \text{ in terms of } b_1, b_2. \end{aligned}$$

If case *2* occurs then,

$$\begin{aligned} f = f_1 b_1 + f_2 b_2 &= f_1 \lambda_2 b'_2 + f_2 \lambda_1 b'_1 \\ &= f_2 \lambda_1 b'_1 + f_1 \lambda_2 b'_2 \\ g = g_1 b_1 + g_2 b_2 &= g_1 \lambda_2 b'_2 + g_2 \lambda_1 b'_1 \\ &= g_2 \lambda_1 b'_1 + g_1 \lambda_2 b'_2 \end{aligned}$$

which in turn implies that w.r.t. the ordered basis b'_1, b'_2 that:

$$\begin{aligned} d_H(f,g) &= \left| \ln \left(\frac{f_1 \lambda_2}{f_2 \lambda_1} \frac{g_2 \lambda_1}{g_1 \lambda_2} \right) \right| \\ &= \left| \ln \left(\frac{f_1}{f_2} \frac{g_2}{g_1} \right) \right| \\ &= \left| - \ln \left(\frac{f_2}{f_1} \frac{g_1}{g_2} \right) \right| \\ &= d_H(f,g) \text{ in terms of } b_1, b_2. \end{aligned}$$

□

Corollary 1.6.4.4. *Let $f, g \in C$ be linearly independent. Let $b_1, b_2 \in \partial\left(Span(f,g) \bigcap C\right)$ be linearly independent, so that b_1, b_2 are a pair of ends for f, g. Then b_1, b_2 are also a pair of ends for themselves and $d_H(b_1, b_2) = \infty$.*

Proof. Since $b_1, b_2 \in \partial\left(\mathrm{Span}(f,g) \bigcap C\right) \subset \mathrm{Span}(f,g)$ are linearly independent we have $\mathrm{Span}(f,g) = \mathrm{Span}(b_1, b_2)$ which implies $\partial\left(\mathrm{Span}(f,g) \bigcap C\right) = \partial\left(\mathrm{Span}(b_1,b_2) \bigcap C\right)$. So $b_1, b_2 \in \partial\left(\mathrm{Span}(b_1,b_2) \bigcap C\right)$ are linearly independent and so are ends to themselves. So, by Proposition 1.6.4.3 (page 58) b_1, b_2 can be used to calculate $d_H(b_1, b_2)$: Let $f = b_1$ and $g = b_2$ we have

$$\begin{aligned} b_1 = f &= f_1 b_1 + f_2 b_2 = 1 b_1 + 0 b_2 \\ b_2 = g &= g_1 b_1 + g_2 b_2 = 0 b_1 + 1 b_2 \end{aligned}$$

so since

$$\begin{aligned} d_H(f,g) &= \left| \ln\left(\frac{f_2}{f_1} \frac{g_1}{g_2} \right) \right| \text{ we have} \\ d_H(b_1,b_2) &= \left| \ln\left(\frac{0}{1} \frac{0}{1} \right) \right| = \infty \end{aligned}$$

□

1.6.5 $d_H(f,g) = |\ln(m(f/g)\ m(g/f))| = \ln\left(\frac{1}{m(f/g)\ m(g/f)}\right)$

Lemma 1.6.5.1. *If $f, g \in C \setminus \{0\}$ then*

$$d_H(f,g) \ = \ |\ln(m(f/g)\ m(g/f))| \ = \ \ln\left(\frac{1}{m(f/g)\ m(g/f)} \right).$$

Proof. If f, g are linearly dependent, then since $f, g \in C \setminus \{0\}$, and C is salient, we must have $\lambda f = g$ for some $\lambda > 0$. By Proposition 1.3.1.1 (page 31) (parts *1.* and

2.), making use again of $f, g \in C \setminus \{0\}$, we have:

$$\begin{aligned} |\ln(m(f/g)\ m(g/f))| &= |\ln(m(f/\lambda f)\ m(\lambda f/f))| \\ &= \left|\ln\left(\frac{1}{\lambda}\frac{\lambda}{1}\right)\right| \\ &= |\ln(1)| \\ &= 0 \\ &= d_H(f,g) \text{ if } f, g \text{ are linearly dependent and non-zero.} \end{aligned} \tag{1.30}$$

Moreover, by (1.30) it is clear that $m(f/g)m(g/f) = 1$ so

$$\frac{1}{m(f/g)m(g/f)} = 1 \text{ and so } |\ln(m(f/g)m(g/f))| = \ln\left(\frac{1}{m(f/g)m(g/f)}\right) = 0.$$

If f, g are linearly independent, it is a little more complicated.

Proposition 1.6.4.3 (page 58), tells us that $d_H(f,g)$ is independent of our choice of the ordered basis b_1, b_2 which we express f and g in, provided $b_1, b_2 \in \partial\left(\text{Span}(f,g)\bigcap C\right)$ and are linearly independent. Lemma 1.3.3.1 (page 38) and Corollary 1.4.1.1 (page 44) imply that $b(f/g)$ and $b(g/f)$ are linearly independent and $\in \partial\left(\text{Span}(f,g)\bigcap C\right)$ So let

$$b_1 = b(f/g) \text{ and let } b_2 = b(g/f).$$

By Proposition 1.3.4.1 (page 40), if $f, g \in C$ are linearly independent then

$$\begin{aligned} m(b(f/g), b(g/f); f) &= \frac{f_{b(g/f)}}{f_{b(f/g)}} = m(f/g) \\ &= \frac{f_2}{f_1} \\ m(b(g/f), b(f/g); g) &= \frac{g_{b(f/g)}}{g_{b(g/f)}} = m(g/f) \\ &= \frac{g_1}{g_2} \end{aligned}$$

So

$$\begin{aligned} |\ln(m(f/g)\ m(g/f))| &= \left|\ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| \\ &= d_H(f,g). \end{aligned}$$

By Proposition 1.3.5.1 (page 41) if $f, g \in C$ are linearly independent then

$$0 \leq m(f/g)m(g/f) < 1$$

and so

$$|\ln(m(f/g)m(g/f))| = \ln\left(\frac{1}{m(f/g)\ m(g/f)}\right).$$

□

1.6.6 $f \in C \setminus \partial C$ and $b \neq 0 \in \partial C$ implies $d_H(f,b) = \infty$

As usual, let C be a closed, convex, salient, pointed by the origin cone in a Banach Space V.

Proposition 1.6.6.1. *If $f \in C \setminus \partial C$ and $b \neq 0 \in \partial C$ then $d_H(f,b) = \infty$.*

Proof. By part 7 of Proposition 1.5.0.2 (page 50)

$$m(b/f)m(f/b) = 0. \tag{1.31}$$

By Lemma 1.6.5.1 (page 60), if $f, g \in C \setminus \{0\}$ then $d_H(f,g) = |\ln(m(f/g)\ m(g/f))|$. So, using (1.31), we get

$$d_H(f,b) = |\ln(m(b/f)\ m(f/b))| = |\ln(0)| = \infty.$$

□

1.6.7 $d_H(f,g)$ is an extended pseudo metric on C

Definition 1.6.7.1. A pseudo metric d satisfies all the conditions of being a metric except the condition $d(x,y) = 0 \Rightarrow x = y$. An extended metric (or extended pseudo metric) is a metric (or pseudo metric) which takes values in $\mathbb{R}_{\geq 0} \cup \{\infty\}$, the assignment $d(x,y) = \infty$ is permitted if $x \neq y$.

Lemma 1.6.7.2. *Let $f, g \in C \setminus \{0\}$.*

If f,g are linearly independent then $0 \leq m(f/g)\ m(g/f) < 1$.

If f,g are linearly dependent then $m(f/g)\ m(g/f) = 1$.

Proof. Proposition 1.3.5.1 (page 41) and the first part of the proof of Lemma 1.6.5.1 (page 60). □

Theorem 1.6.7.3. *$d_H(f,g)$ is a pseudo metric on $C \setminus \{0\}$ taking values in $\mathbb{R}_{\geq 0} \cup \{\infty\}$.*

Proof. Let $f, g, h \in C \setminus \{0\}$.

1. $d_H(f,g) \geq 0$

Absolute value is always ≥ 0.

2. $d_H(f,f) = 0$

f, f are linearly dependent.

3. $d_H(f,g) = d_H(g,f)$

If f, g are linearly dependent then $d_H(f,g) = d_H(g,f) = 0$. If f, g are linearly

independent then:

$$\begin{aligned} d_H(f,g) &= \left|\ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| \\ &= \left|-\ln\left(\frac{f_2}{f_1}\frac{g_1}{g_2}\right)\right| \\ &= \left|-\ln\left(\frac{f_1}{f_2}\frac{g_2}{g_1}\right)\right| \\ &= d_H(g,f). \end{aligned}$$

Note that $|\ln(0)| = |\ln(1/0)| = \infty$.

4. The triangle inequality: $d_H(f,h) \leq d_H(f,g) + d_H(g,h)$

By Lemma 1.6.5.1 (page 60),

$$d_H(f,g) = |\ln(m(f/g)\ m(g/f))|. \tag{1.32}$$

Recalling Definition 1.3.0.8,

$$\begin{aligned} m(f/g) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid f - tg \in C\} \\ m(g/f) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid g - tf \in C\} \\ m(g/h) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid g - th \in C\} \\ m(h/g) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid h - tg \in C\} \\ m(f/h) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid f - th \in C\} \\ m(h/f) &\stackrel{def}{=} \sup\{t \in \mathbf{R} \mid h - tf \in C\}. \end{aligned}$$

By Proposition 1.3.1.1 (page 31) (item *4.*)

$$\begin{aligned} f - m(f/g)g &= b(f/g) \\ g - m(g/f)f &= b(g/f) \\ h - m(h/g)g &= b(h/g) \\ g - m(g/h)h &= b(g/h) \end{aligned}$$

are all $\in C$. Since C is a cone it is closed under non-negative linear combinations. Hence,

$$\begin{aligned} \overbrace{f - m(f/g)g}^{b(f/g)} + m(f/g)\overbrace{(g - m(g/h)h)}^{b(g/h)} &= f - [m(f/g)\ m(g/h)]\ h\ \in C \\ &\Rightarrow\ m(f/g)\ m(g/h) \leq m(f/h) \quad (1.33) \\ \overbrace{h - m(h/g)g}^{b(h/g)} + m(h/g)\overbrace{(g - m(g/f)f)}^{b(g/f)} &= h - [m(h/g)\ m(g/f)]\ h\ \in C \\ &\Rightarrow\ m(g/f)\ m(h/g) \leq m(h/f). \quad (1.34) \end{aligned}$$

By Lemma 1.6.7.2 (page 63),

$$\begin{aligned} 0 \leq\ & m(f/h)\ m(h/f) \leq 1 \\ 0 \leq\ & m(f/g)\ m(g/f) \leq 1 \\ 0 \leq\ & m(g/h)\ m(h/g) \leq 1. \end{aligned}$$

Combining these with the inequalities (1.33) and (1.34) imply:

$$\begin{aligned}
0 \leq m(f/g)\ m(g/h)\ m(h/g)\ m(g/f) &\leq m(f/h)\ m(h/f) \leq 1 \\
-\infty \leq \ln(m(f/g)\ m(g/h)\ m(h/g)\ m(g/f)) &\leq \ln(m(f/h)\ m(h/f)) \leq 0 \\
-\infty \leq \underbrace{\ln(m(f/g)\ m(g/f))}_{\leq 0} + \underbrace{\ln(m(g/h)\ m(h/g))}_{\leq 0} &\leq \ln(m(f/h)\ m(h/f)) \leq 0 \\
&\qquad \text{multiply by } -1 \text{ to get } |\ | \\
\infty \geq -\ln(m(f/g)\ m(g/f)) + -\ln(m(g/h)\ m(h/g)) &\geq -\ln(m(f/h)\ m(h/f)) \geq 0 \\
\infty \geq |\ln(m(f/g)\ m(g/f))| + |\ln(m(g/h)\ m(h/g))| &\geq |\ln(m(f/h)\ m(h/f))| \geq 0 \\
d_H(f,g) + d_H(g,h) &\geq d_H(f,h).
\end{aligned}$$

□

Corollary 1.6.7.4. *$d_H(f,g)$ is a pseudo metric on C taking values in $\mathbb{R}_{\geq 0} \cup \{\infty\}$.*

Proof. By Theorem 1.6.7.3 (page 63) it suffices to show that the pseudo metric axioms holds in the presence of 0. Let $f, g, h \in C$.

1. $d_H(f,g) \geq 0$

We defined

$$d_H(f,g) = \begin{cases} 0, & f = g = 0; \\ \infty, & 0 \in \{f,g\} \text{ and } f \neq g. \end{cases}$$

2. $d_H(f,f) = 0$

We defined $d_H(0,0) = 0$.

3. $d_H(f,g) = d_H(g,f)$

The definition of d_H is symmetric in the presence of zero.

4. The triangle inequality: $d_H(f,h) \leq d_H(f,g) + d_H(g,h)$

If f and h are both equal to 0 there is nothing to show since then $d_H(f,h) = 0$. If $f \neq 0$ but $h = 0$ then $d_H(f,h) = \infty$ and we must show that at least one of $d_H(f,g), d_H(g,h)$ is zero. If $g = 0$ then $d_H(f,g) = \infty$ and we are done. If $g \neq 0$ then $d_H(g,h) = \infty$ and we done. □

1.7 $(C, \sim)$ the Projective Space of C

We are ultimately interested in the Projective Space of C, in particular rays originating at the origin and contained in C.

1.7.1 Basics about Linear Maps, $\sim$, 0-Rays, and Eigenvectors

We depart briefly from our construction of the Hilbert Metric to introduce some useful basic concepts.

We recall Definition 1.1.0.3 (page 20): Let V be a vector space and suppose $x \in V$. An open 0-ray in the direction of x is denoted $[x]$, algebraically:

$$[x] = \{\alpha x : \alpha > 0\}.$$

Remark 1.7.1.1. We allow for the trivial open 0-ray $\{0\} = [0]$.

The term "0-ray" is to remind the reader that our rays are emanating from $\{0\}$.

Remark 1.7.1.2. We will show that the open 0-rays are the equivalence classes of an equivalence relation and so the square bracket notation is appropriate.

Definition 1.7.1.3. If $x, y \in V$ we define $x \sim y$ if $x = \alpha y$ for some real $\alpha > 0$.

Proposition 1.7.1.4. *The relation $x \sim y$ is an equivalence relation on V. The equivalence classes of $\sim$ are the open 0-rays. The open 0-rays partition V.*

Proof. The following argument shows that $\sim$ is reflexive, symmetric, and transitive, and hence is an equivalence relation on V. Let $x, y, z \in V$.

1. Reflexivity: $x = 1x$ implies $x \sim x$

2. Symmetry: If $x \sim y$ then $\exists\ \alpha > 0$ such that $x = \alpha y$. But then $y = \frac{1}{\alpha}x$ which implies $y \sim x$.

3. Transitivity: If $x \sim y$ and $y \sim z$ then $\exists\ \alpha_{x/y}, \alpha_{y/z} > 0$ such that $x = \alpha_{x/y}y$ and $y = \alpha_{y/z}z$. But then $x = \alpha_{x/y}\alpha_{y/z}z$ which implies $x \sim z$.

The following argument shows that the equivalence classes of $\sim$ are the open 0-rays.

$$\begin{aligned}\text{The equivalence class of x } &= \{x' \in V : x \sim x'\} \\ &= \{x' \in V : x' = \alpha x \text{ for some } \alpha > 0\} \\ &= \{\alpha x : \alpha > 0\} \\ &= [x].\end{aligned}$$

The equivalence classes of any equivalence relation form a partition, hence the open 0-rays partition V. □

Definition 1.7.1.5. Notation. If L is any linear map of V to itself and $x \in V$ then we will write $[x]L$ or $[xL]$ for $L([x])$. This is the notation that Birkhoff uses in [12].

Proposition 1.7.1.6. *If L is any linear map of V to itself and $x \in V$ then, as sets, $L([x]) = [L(x)]$. I.e.*

$$[x]L = [xL]$$

So L takes open 0-rays to open 0-rays.

Proof.

$$\begin{aligned}
[x]L &= L([x]) \\
&= L(\{\alpha x : \alpha > 0\}) \\
&= \{L(\alpha x) : \alpha > 0\} \\
&= \{\alpha L(x) : \alpha > 0\} \\
&= [L(x)] \\
&= [xL].
\end{aligned}$$

□

We have the standard definition:

Definition 1.7.1.7. If L is any linear map of V to itself then $c \in V \setminus \{0\}$ is called an eigenvector of L with eigenvalue λ_c if $L(c) = \lambda_c c$.

Proposition 1.7.1.8. *If L is any linear map of V to itself with eigenvector c and corresponding real eigenvalue λ_c, then, as sets, $L([c]) = sgn(\lambda_c)[c]$, where sgn is the sign function.*

Proof. By Proposition 1.7.1.6 (page 68), $L([c]) = [L(c)]$. Then:

$$\begin{aligned}
L([c]) = [L(c)] = [\lambda_c c] &= \{\alpha\lambda_c c : \alpha > 0\} \\
&= \frac{\lambda_c}{|\lambda_c|}\{\alpha|\lambda_c|c : \alpha > 0\} \\
&= \frac{\lambda_c}{|\lambda_c|}\{\alpha c : \alpha > 0\} \\
&= \operatorname{sgn}(\lambda_c)\{\alpha c : \alpha > 0\} \\
&= \operatorname{sgn}(\lambda_c)[c].
\end{aligned}$$

□

Corollary 1.7.1.9. *c is an eigenvector of L with positive eigenvalue if and only if the open 0-ray $[c]$ is a fixed by L.*

Proof. This corollary follows immediately from Proposition 1.7.1.8 (page 69) since if λ_c is positive then $\text{sgn}(\lambda_c) = 1$. □

Remark 1.7.1.10. By Proposition 1.7.1.6 (page 68) we can view L as being an endomorphism of the set of open 0-rays. Viewing L in this way, c is an eigenvector of L with positive eigenvalue λ_c if and only if the open 0-ray $[c]$ is a fixed point of L.

Definition 1.7.1.11. Let V be a vector space and let $\mathcal{C}$ be any subset of V. $\mathcal{C}$ is closed under positive scaling if $x \in \mathcal{C}$ and $\alpha > 0$ then $\alpha x \in \mathcal{C}$. More succinctly, $\mathcal{C}$ is closed under positive (resp. non-negative) scaling if $\alpha\mathcal{C} \subset \mathcal{C}$ for all $\alpha > 0$ (resp. for all $\alpha \geq 0$).

Remark 1.7.1.12. If $\mathcal{C}$ is closed under positive (resp. non-negative) scaling then $\mathcal{C}$ is a cone (resp. pointed cone). See Definition 1.1.0.1 (page 19).

Proposition 1.7.1.13. *If V is a vector space and $\mathcal{C} \subset V$ is closed under positive scaling then for each $\alpha > 0$ it is the case that $\alpha\mathcal{C} = \mathcal{C}$. So if C is a cone or a pointed cone and $\alpha > 0$ then $\alpha C = C$.*

Proof. If $\alpha > 0$ then $\frac{1}{\alpha} > 0$. So if $\mathcal{C}$ is closed under positive scaling then $\alpha\mathcal{C} \subset \mathcal{C}$ and

$$\frac{1}{\alpha}\,\mathcal{C} \subset \mathcal{C}.$$

But then

$$\alpha\,\frac{1}{\alpha}\,\mathcal{C} = \mathcal{C} \subset \alpha\mathcal{C}.$$

So $\alpha\mathcal{C} = \mathcal{C}$. □

Remark 1.7.1.14. If $\mathcal{C}$ is closed under positive scaling and $x \in \mathcal{C}$ then the open ray $[x] \subset \mathcal{C}$

Definition 1.7.1.15. Let $\mathcal{C} \subset V$ be closed under positive scaling. The quotient set $\{[x] : x \in \mathcal{C}\}$ written as either $(\mathcal{C}, \sim)$ or $\mathcal{C}/\sim$ is the set of all equivalence classes of $\sim$ contained in $\mathcal{C}$. So $(\mathcal{C}, \sim)$ consists of all the open 0-rays of $\mathcal{C}$. We may occasionally use the symbol π to represent the quotient map of $\mathcal{C}$ onto $(\mathcal{C}, \sim)$. In particular if $f \in \mathcal{C}$ then $\pi(f) = [f] \in (\mathcal{C}, \sim)$

Remark 1.7.1.16. The relation $\sim$ as defined in Definition 1.7.1.3 (page 67) is an equivalence relation on any non-empty subset $\mathcal{C} \subset V$. However, if $\mathcal{C}$ is not closed under positive scaling, then some of the equivalence classes formed out of $\mathcal{C}$ will only be proper subsets of open 0-rays. However, we want the equivalence classes to be open 0-rays. So, to insure this, the relation $\sim$ will only be applied to subsets of V which are closed under positive scaling; i.e. cones and pointed cones, vector subspaces of V, and V. It is worth mentioning that, technically speaking, vector (sub)spaces are pointed cones, at least based upon the definition of a pointed cone as being any subset of V which is closed under non-negative scaling.

1.7.2 Basics about the projective space of C

Unless otherwise noted, C will always denote a pointed, salient, closed, convex cone contained in the Banach Space V.

Building on Section 1.7.1 (page 67) we have:

Proposition 1.7.2.1. *Let $f \in C$ then:*

1. $[f] \subset C$.
2. *If $\kappa > 0$ then $[\kappa f] = [f]$.*
3. *If P is linear on V, then $[f]P = [fP]$. So P induces a map on $(C, \sim)$, which we will also call P, defined by $P : [f] \mapsto [fP]$. We sometimes write $P([f])$ for $[fP]$*

4. If $f \neq 0$ then $[f] = \overrightarrow{f} \setminus \{ 0\}$. Note $\overrightarrow{f} = \{\alpha f \mid \alpha \geq 0\}$.

5. $[0] = \{0\}$.

Proof. These results are mostly trivial.

(*1.*) C is a cone so it is closed under non-negative (and hence positive) scaling.

(*2.*)

$$[\kappa f] = \{\lambda(\kappa f) \mid \lambda > 0\} = \{(\lambda\kappa) f \mid \lambda > 0\} = \{\lambda f \mid \lambda > 0\} = [f]$$

(*3.*)

$$P(\pi(f)) = P(\{\lambda f \mid \lambda > 0\}) = \{\lambda P(f) \mid \lambda > 0\} = \pi(Pf)$$

(*4.*) Follows trivially from definition of $\overrightarrow{f} = \{\lambda f \mid \lambda \geq 0\}$.

(*5.*) Follows trivially from definition of $[f] = \{\lambda f \mid \lambda > 0\}$. □

Proposition 1.7.2.2. *Let $f, g \in C$ and let $f' \in [f]$ and let $g' \in [g]$. Then*

$$d_H(f, g) = d_H(f', g').$$

Proof. There are four cases.

Case 1: If $f = g = 0$ then $f' = g' = 0$, hence,

$$d_H(f, g) = d_H(0, 0) = d_H(f', g') = 0.$$

Case 2: If $f = 0$ and $g \neq 0$ then $f' = 0$ and $g' = \beta g$ for some $\beta > 0$. Then

$$d_H(f, g) = d_H(0, g) = \infty \quad \text{and} \quad d_H(f', g') = d_H(0, g') = \infty$$

Case 3: $f \neq 0$ and $g = 0$. This is just Case 2., with f and g reversed.

Case 4: If $f, g \in C \setminus \{0\}$ then $f' = \alpha f$ and $g' = \beta g$ for some $\alpha, \beta > 0$. Since f, g, f', g' are non-zero, we can invoke Proposition 1.3.1.1 (page 31), part (*9.*):

$$\begin{aligned} m(f'/g') &= m(\alpha f/\beta g) = \frac{\alpha}{\beta} m(f/g) \\ m(g'/f') &= m(\beta g/\alpha f) = \frac{\beta}{\alpha} m(g/f). \end{aligned}$$

Then, using Lemma 1.6.5.1,

$$\begin{aligned} d_H(f', g') &= |\ln (m(f'/g')\ m(g'/f'))| \\ &= |\ln (m(\alpha f/\beta g)\ m(\beta g/\alpha f))| \\ &= \left| \ln \left(\frac{\alpha}{\beta} m(f/g)\ \frac{\beta}{\alpha} m(g/f) \right) \right| \\ &= |\ln (m(f/g)\ m(g/f))| \\ &= d_H(f, g). \end{aligned}$$

□

1.8 d_H on $(C, \sim)$

Definition 1.8.0.3. We define d_H on $(C,\ \sim)$. Let $f, g \in C$ then

$$d_H([f], [g]) = d_H(f', g')$$

where f' is any element of $[f]$ and g' is any element of $[g]$.

Remark 1.8.0.4. By Proposition 1.7.2.2 (page 72), Definition 1.8.0.3 is independent of the choice of f' and g' and so $d_H([f], [g])$ is well defined. In particular,

$$d_H([f], [g]) = d_H(f, g)$$

is true (and well defined).

1.8.1 $f, g \in C \setminus \partial C$ then $d_H(f,g) < \infty$

Proposition 1.8.1.1. *If $f, g \in C \setminus \partial(C)$ then $d_H(f,g) < \infty$.*

Proof. Since $0 \in \partial(C)$ it follows that $f, g \neq 0$. There are two cases:

Case 1: If f and g are linearly dependent (and non-zero), we have, by definition $d_H(f,g) = 0$,

Case 2: Let f, g be linearly independent. They determine a linearly independent pair of ends b_1, b_2. W.r.t. these ends

$$\begin{aligned} f &= f_1 b_1 + f_2 b_2 \\ g &= g_1 b_1 + g_2 b_2 \end{aligned}$$

with $f_1, f_2, g_1, g_2 \geq 0$ and with

$$d_H(f,g) = \left| \ln \left(\frac{f_2}{f_1} \frac{g_1}{g_2} \right) \right|.$$

Suppose that $d_H(f,g) = \infty$. Then at least one of the f_1, f_2, g_1, g_2 are zero. This implies that f or g is a positive multiple of b_1 or b_2. However, Proposition 1.4.2.1 (page 46), and Proposition 1.3.1.1 (page 31) part *5.*, tells us that $\overrightarrow{b_1} \bigcup \overrightarrow{b_2} \subset \partial(C)$. Contradiction. □

1.8.2 The Equivalence Relation: $f \equiv_m g$ iff $d_H(f,g) < \infty$

Definition 1.8.2.1. Let $f, g \in C$ We define

$$f \equiv_m g \text{ iff } d_H(f,g) < \infty$$

Proposition 1.8.2.2. *$\equiv_m$ is an equivalence relation on C.*

Proof. Let $f, g, h \in C$.

1. $d_H(f, f) = 0 < \infty$. So $f \equiv_m f$.

2. $d_H(f, g) = d_H(f, g)$ so $f \equiv_m g$ implies $g \equiv_m f$.

3. If $f \equiv_m g$ and $g \equiv_m h$ then $d_H(f, g), d_H(g, h) < \infty$. Then since $d_H(f, h) \leq d_H(f, g) + d(g, h)$ we have $d_H(f, h) < \infty$. So $f \equiv_m h$.

□

Proposition 1.8.2.3. *Let $f \in C$. Then $f' \in [f] \Leftrightarrow d_H(f', f) = 0$.*

Proof. If $f = 0$ the result in both directions is a direct consequence of the definition of d_H. So suppose $f \neq 0$.

Suppose $f' \in [f]$. Then $f' = \lambda f \neq 0$ for some $\lambda > 0$. Then by the definition of d_H for linearly dependent non-zero members of C we have $d_H(f, f') = 0$.

In the other direction, suppose that $d_H(f, f') = 0$. If f', f are linearly dependent, then $f' = \lambda f$ for some $\lambda \in \mathbb{R}$. Since C is salient, $\lambda \geq 0$. Since we are assuming at this point in the proof that $f \neq 0$, it must be the case that $\lambda > 0$. But then $f' \in [f]$.

If f', f are linearly independent, then we can write f', f in terms of a pair of linearly independent ends b_1, b_2 as $f = f_1 b_1 + f_2 b_2$ and $f' = f'_1 b_1 + f'_2 b_2$. But then

$$\begin{gathered} d_H(f, f') = \left| \ln \left(\frac{f_2}{f_1} \frac{f'_1}{f'_2} \right) \right| = 0 \quad \Rightarrow \\ \frac{f_2}{f_1} \frac{f'_1}{f'_2} = 1 \quad \Rightarrow \\ \frac{f_2}{f_1} = \frac{f'_2}{f'_1} \end{gathered}$$

which implies $f' = \lambda f$, which contradicts linear independence. □

1.8.3 Partition of C by d_H

Theorem 1.8.3.1. ***(Partition of C by d_H Theorem.)*** *Let X be an equivalence class in $\equiv_m$. Then X is a convex salient cone and $X \cup \{0\}$ is a convex salient pointed by the origin cone.*

Proof. See[1]. Case 1: $0 \in X$. Then $X = \{0\}$, and the result follows trivially.

Case 2: $0 \notin X$.

Let $f \in X$. Then $\lambda f \neq 0 \ \ \forall\ \lambda > 0$. Hence $d_H(f, \lambda f) = 0 \ \ \forall\ \lambda > 0$. So X is closed under positive scaling and $X \cup \{0\}$ is closed under non-negative scaling. So X is a cone and $X \cup \{0\}$ is a pointed-by-the-origin cone.

Next we show that X is convex:

To show that X is convex we must show that if $f, g \in X$ then the line segment $\overline{fg} \ = \ f + t(g - f),\ t \in [0, 1]$ is contained in X.

If $f, g \in X$ then they are both non-zero.

Suppose f, g are linearly dependent. Then we can write $g = \lambda f$ for some $\lambda > 0$ and

$$f + t(g - f) \ \ = \ \ f + t(\lambda f - f) \ \ = \ \ (1 + t(\lambda - 1))f.$$

But $(1+t(\lambda-1))$ is between 1 and λ if $t \in [0, 1]$. So if $t \in [0, 1]$ then $(1+t(\lambda-1)) > 0$ which implies $d_H((1 + t(\lambda - 1))f, f)$ is well defined and equal to zero. So the line segment $\overline{fg} = f + t(g - f),\ t \in [0, 1]$ is contained in X.

Suppose $f, g \in X$, with f, g linearly independent. By Proposition 1.6.1.1 (page 53) there exists a pair of ends, b_0, b_1 for f, g. In terms of b_0, b_1 we have

$$f = f_0 b_0 + f_1 b_1$$

$$g = g_0 b_0 + g_1 b_1$$

[1]See Theorem 1.11.3.2 (page 105) for a more elegant proof that the 'component' X is a cone. The proof of Theorem 1.11.3.2 uses the $\alpha x \leq y \leq \beta x$ formulation of d_H.

with $f_0, f_1, g_0, g_1 \geq 0$. However, since $f, g \in X$ we have $d_H(f,g) < \infty$ and so actually

$$f_0, f_1, g_0, g_1 > 0,$$

see[2]. In terms of b_0, b_1 we have

$$f + t(g-f) \quad = \quad (f_0 + t(g_0 - f_0))b_0 + (f_1 + t(g_1 - f_1))b_1.$$

Since $f_0, f_1 > 0$ if $d_H(f, f + t(g-f)) = \infty$ then

$$(f_0 + t(g_0 - f_0)) = 0 \qquad \text{or} \qquad (f_1 + t(g_1 - f_1)) = 0.$$

Suppose

$$(f_0 + t(g_0 - f_0)) = 0.$$

$f_0 > 0$, so if $(f_0 + t(g_0 - f_0)) = 0$ we must have $(g_0 - f_0) \neq 0$. But then

$$\frac{f_0}{g_0 - f_0} = t.$$

But

$$\begin{aligned} 0 < f_0 < g_0 \;&\Rightarrow\; t = \frac{f_0}{g_0 - f_0} > 1 \\ 0 < g_0 < f_0 \;&\Rightarrow\; t = \frac{f_0}{g_0 - f_0} < 0. \end{aligned}$$

In either case $t \notin [0,1]$. So if $t \in [0,1]$ then $(f_0 + t(g_0 - f_0)) \neq 0$ and similarly $(f_1 + t(g_1 - f_1)) \neq 0$. So $t \in [0,1]$ implies $d_H(f, f + t(g-f)) < \infty$ and so the line segment $\overline{fg} \;=\; f + t(g-f),\; t \in [0,1]$ is contained in X (whether f,g are linearly

[2] $d_H(f,g) = \ln\left(\frac{f_1}{f_0} \; \frac{g_0}{g_1}\right)$ so $d_H(f,g) = \infty$ implies at least one of f_0, f_1, g_0, g_1 is 0. See Definition 1.6.2.1 (page 54) of d_H.

independent or not). So we've proven that X is convex.

Next we show $X \cup \{0\}$ is convex.

To show that $X \cup \{0\}$ is convex, we must show that if $f, g \in X \cup \{0\}$ then $f + t(g - f) \in X \cup \{0\}$ for all $t \in [0, 1]$. Since we know that X is convex, we only have to consider the line segments from 0 to points in $X \cup \{0\}$.

If both f and g are zero then $f + t(g - f) = 0$ and we are done. If $f = 0$ but $g \neq 0$ then $f + t(g - f)$ becomes tg. If $t > 0$ then $d_H(tg, g) = 0$ so $tg \in X$. If $t = 0$ then $tg = 0 \in X \cup \{0\}$.

If $g = 0$ but $f \neq 0$ then $f + t(g - f)$ becomes $(1 - t)f$. If $0 \leq t < 1$ then $d_H((1 - t)f, f) = 0$ so $(1 - t)f \in X$. If $t = 1$ then $(1 - t)f = 0 \in X \cup \{0\}$.

So we've proven that $X \cup \{0\}$ is convex.

C is salient and X and $X \cup \{0\} \subset C$ so X and $X \cup \{0\}$ are salient. Recall S is salient if $S \cap -S = \{0\}$. □

Remark 1.8.3.2. We can think of the partitioning of C via $\equiv_m$ as a sort of boundary operator. E.g. $C = \mathbb{R}^2_{\geq 0}$ decomposes via $\equiv_m$ into 4 equivalence classes. They are the interior of $\mathbb{R}^2_{\geq 0}$, the open 0-ray $[(0, 1)]$, the open 0-ray $[(1, 0)]$, and $[0] = [(0, 0)]$. Via $\sim$ these equivalence classes can be identified with the open line segment in $\mathbb{R}^2$ between the points $(1, 0)$ and $(0, 1)$, the points $(1, 0)$ and $(0, 1)$, and the origin $(0, 0)$.

1.8.4 Hilbert Projective Metric Theorem for Cones

Proposition 1.8.4.1. *Let d be a pseudo metric on X.*

For $x, x' \in X$ let $x \equiv x'$ if $d(x, x') = 0$ and denote the equivalence class of x by $[x]$.

Then $\equiv$ is a well defined equivalence relation. Let $d([x], [y]) = d(x, y)$, then d is a metric on $(X, \equiv)$.

If d were an extended pseudo metric on X then d would be an extended metric on $(X, \equiv)$.

Proof. First we show: $\equiv$ is an equivalence relation. $x \equiv x$ and $x \equiv y \Rightarrow y \equiv x$ are trivial. If $x \equiv y$ and $y \equiv z$ then $d(x,z) \leq d(x,y) + d(y,z) = 0 + 0 = 0$, which implies $x \equiv z$.

Next we show: d is well defined on $(X, \equiv)$. Suppose $x' \in [x]$ and $y' \in [y]$. Then $d(x',y') \leq d(x',x) + d(x,y) + d(y,y') = d(x,y)$. Similarly $d(x,y) \leq d(x',y')$. So $d(x,y) = d(x',y')$.

Finally we show that d is a metric on $(X, \equiv)$. If $d([x],[y]) = d(x,y) = 0$ then $x \equiv y$ by the definition of $\equiv$.

The other conditions: $d([x],[x]) = 0$. $d([x],[y]) = d([y],[x])$ and the triangle inequality are inherited from d being a pseudo metric on X.

The same exact proof works for if d is an extended pseudo metric on X. □

Theorem 1.8.4.2. ***(Hilbert Projective Metric Theorem for Cones.)*** *Let C be a closed convex salient pointed cone (with point $\{0\}$). Let X be an equivalence class in $\equiv_m$.*

1. *d_H is an extended metric on $(C, \sim)$ and $(C \setminus \{0\}, \sim)$.*

2. *d_H is a true metric on the set $(X, \sim)$.*

3. *d_H is a true metric on $(C \setminus \{0\}, \sim)$.*

Proof. Parts 1 and 2 of this theorem are an immediate consequence of Proposition 1.8.4.1 (page 78) and Proposition 1.8.2.3 (page 75). Part 3 of this theorem follows from Part 2 of this theorem and Proposition 1.8.1.1 (page 74). □

Remark 1.8.4.3. We can extend our construction of d_H to certain 'bounded' convex sets.

Note 1. If $K \subset V$ is a bounded convex set we could isometrically embed K in the Banach Space $\mathbb{R} \times V$ via $K \to \{1\} \times K$ and then form the cone of $\{1\} \times K$ over

the single point $(0,0) = 0 \in \mathbb{R} \times V$, we would name this cone C, and then apply the machinery developed in Chapter 1 to produce d_H for $K = (C, \sim)$.

Note 2. Or d_H could be developed in a more direct manner, by replacing $\text{Span}(f, g)$ with the hyperplane (i.e. the line) generated by (i.e. passing though) f and g. If f, g aren't in ∂K such a line can be shown to meet the boundary of K in two points (or ends), say b_1, b_2 and then $d_H(f, g)$ can be developed in terms of the cross ratio of b_1, f, g, b_2 using machinery developed in Chapter 1.

Note 3. If C is a convex salient cone, and we can project C onto a hyperplane H that cuts each ray of C exactly once, the image of this projection, which is just $C \cap H$ will be bounded and convex. Then one could develop d_H for $C \cap H$ in the manner outlined in Note 2.

1.9 The Hilbert Projective Metric via $\alpha f \leq g \leq \beta f$

The Hilbert Metric definition discussed in this section seems to have originated in Birkhoff [13] (1962). The version we present here is what seems to be the most common modern formulation, see e.g. [92]. We first prove some technical results, then we address the metric's definition.

1.9.1 α_{fg}, β_{fg} and $\frac{1}{\beta_{fg}} = \sup\{t \in \mathbb{R} : f - tg \in C\} = \alpha_{gf}$

Definition 1.9.1.1. With respect to $f, g \in C \setminus \{0\}$, given in that order,

$$\begin{aligned} \alpha_{fg} &= \sup\{t : tf \leq g\} = \sup\{t : g - tf \in C\} \\ \beta_{fg} &= \inf\{t : g \leq tf\} = \inf\{t : tf - g \in C\} \end{aligned}$$

where we take $\inf \emptyset = \infty$.

Lemma 1.9.1.2. *Let $f, g \in C \setminus \{0\}$ then*

$$\frac{1}{\beta_{fg}} \;=\; \frac{1}{\inf\{t \in \mathbb{R} : tf - g \in C\}} \;=\; \sup\{t \in \mathbb{R} : f - tg \in C\} \;=\; \alpha_{gf}. \tag{1.35}$$

Proof. If $\beta_{fg} = \inf\{t \in \mathbb{R} : tf - g \in C\} = \infty$ then $\nexists\ t' > 0$ such that $t'f - g \in C$. This implies that

$$\sup\{t \in \mathbb{R} : f - tg \in C\} = 0 \tag{1.36}$$

because if $\sup\{t \in \mathbb{R} : f - tg \in C\} > 0$ then there exists a $t' > 0$ such that $f - t'g \in C$ but then $(1/t')f - g \in C$. Contradiction. So if $\beta_{fg} = \infty$ then (1.35) is true.

If $\beta_{fg} = \inf\{t \in \mathbb{R} : tf - g \in C\} < \infty$ we showed above that $0 < \beta_{fg}$ and $\beta_{fg}f - g \in C$ (because C is closed). But then $f - (1/\beta_{fg})g \in C$ so that

$$\sup\{t \in \mathbb{R} : f - tg \in C\} \geq 1/\beta_{fg}. \tag{1.37}$$

If the above inequality (1.37) is strict, then there is an $s > 0$ such that $1/s > 1/\beta_{fg}$ and $f - (1/s)g \in C$. But then $sf - g \in C$ with $s < \beta_{fg}$ which contradicts

$$\beta_{fg} = \inf\{t \in \mathbb{R} : tf - g \in C\}.$$

So (1.37) must be an equality. Reversing the order of f, g in the definition of α_{fg}, Definition 1.9.1.1, yields the last equality in (1.35). So (1.35) has been proved. □

Corollary 1.9.1.3. *Let $f, g \in C \setminus \{0\}$ then*

$$\frac{1}{\inf\{t \in \mathbb{R} : tg - f \in C\}} = \sup\{t \in \mathbb{R} : g - tf \in C\} = \alpha_{fg}.$$

Proof. Simply switch f and g and apply Lemma 1.9.1.2 (page 80). □

Corollary 1.9.1.4. *Let $f, g \in C \setminus \{0\}$ then*

$$\frac{\beta_{fg}}{\alpha_{fg}} = \frac{\inf\{t \in \mathbb{R} : tg - f \in C\}}{\sup\{t \in \mathbb{R} : f - tg \in C\}} \tag{1.38}$$

Proof. Lemma 1.9.1.2 (page 80):

$$\frac{1}{\beta_{fg}} = \frac{1}{\inf\{t \in \mathbb{R} : tf - g \in C\}} = \sup\{t \in \mathbb{R} : f - tg \in C\}.$$

combined with Corollary 1.9.1.3 (page 81):

$$\frac{1}{\inf\{t \in \mathbb{R} : tg - f \in C\}} = \sup\{t \in \mathbb{R} : g - tf \in C\} = \alpha_{fg}.$$

yields (1.38). □

1.9.2 Theorem: $d_H(f,g) = \ln\left(\frac{1}{m(f/g)m(g/f)}\right) = \ln\left(\frac{\beta_{fg}}{\alpha_{fg}}\right) = \ln\left(\frac{\beta_{gf}}{\alpha_{gf}}\right)$

Theorem 1.9.2.1. *Let $f, g \in C \setminus \{0\}$. Then*

1. $$m(g/f) \;=\; \alpha_{fg} \;=\; \frac{1}{\beta_{gf}}$$

2. $$m(f/g) \;=\; \frac{1}{\beta_{fg}} \;=\; \alpha_{gf}$$

3. $$d_H(f,g) \;=\; \ln\left(\frac{1}{m(f/g)m(g/f)}\right) \;=\; \ln\left(\frac{\beta_{fg}}{\alpha_{fg}}\right) \;=\; \ln\left(\frac{\beta_{gf}}{\alpha_{gf}}\right)$$

Proof. Recall Definition 1.3.0.8 (page 30):

$$m(f/g) \;=\; \sup\{t : tg \leq f \in C\} \;=\; \sup\{t : f - tg \in C\} \tag{1.39}$$

$$m(g/f) \;=\; \sup\{t : tf \leq g \in C\} \;=\; \sup\{t : g - tf \in C\} \tag{1.40}$$

and recall Definition 1.9.1.1 (page 80):

$$\alpha_{fg} \;=\; \sup\{t : tf \le g\} \;=\; \sup\{t : g - tf \in C\} \tag{1.41}$$

$$\beta_{fg} \;=\; \inf\{t : g \le tf\} \;=\; \inf\{t : tf - g \in C\} \tag{1.42}$$

where we take $\inf \emptyset = \infty$. It is immediate that (1.40) = (1.41), so

$$m(g/f) \;=\; \alpha_{fg}. \tag{1.43}$$

By switching the order of f, g in (1.43) we get

$$m(f/g) \;=\; \alpha_{gf}. \tag{1.44}$$

Lemma 1.9.1.2 (page 80) together with (1.39) imply

$$\frac{1}{\beta_{fg}} \;=\; \sup\{t \in \mathbb{R} : f - tg \in C\} \;=\; m(f, g). \tag{1.45}$$

By switching the order of f, g in (1.45) we get

$$\frac{1}{\beta_{gf}} \;=\; m(g, f). \tag{1.46}$$

Combining (1.43) with (1.46) proves Part 1:

$$m(g/f) \;=\; \alpha_{fg} \;=\; \frac{1}{\beta_{gf}}. \tag{1.47}$$

Combining (1.44) with (1.45) proves Part 2:

$$m(f/g) \;=\; \frac{1}{\beta_{fg}} \;=\; \alpha_{gf}. \tag{1.48}$$

By Lemma 1.6.5.1 (page 60)

$$d_H(f,g) \;=\; \ln\!\left(\frac{1}{m(f/g)\; m(g/f)}\right). \tag{1.49}$$

Substituting (1.47) and (1.48) for $m(g/f)$ and $m(f/g)$ into (1.49) yields part 3. □

1.9.3 d_H via $\alpha f \leq g \leq \beta f$

See [3]. The formulation of the Hilbert Metric given below, within Theorem 1.9.3.2, seems to be the most convenient one for many calculations.

Let C be closed, salient, pointed by the origin, convex cone contained in a Banach Space V (as usual).

For convenience, we restate and relabel Definition 1.3.0.5 (page 30):

Definition 1.9.3.1. Suppose $u, v \in V$ we will write

$$u \leq v \;\text{ if }\; v - u \in C.$$

By Proposition 1.3.0.7 (page 30) we can say that V is partially ordered by C.

Recall that an extended metric is a metric except that the value ∞ is allowed; a pseudo metric d satisfies all the standard axioms of being a metric except that $d(f,g) = 0$ does not necessarily imply $f = g$.

Theorem 1.9.3.2. ***The Hilbert Projective Metric***

1. *Let $f, g \in C \setminus \{0\}$. Let α, β be the largest and smallest (non-negative) real numbers such that*

$$\alpha f \leq g \leq \beta f, \tag{1.50}$$

[3] The material covered in Section 1.9.3, defining $d_H(f,g)$ via $\alpha f \leq g \leq \beta f$ is standard. I have included it in the interests of this work being more or less self-contained. The proofs given are my own, but ultimately they are quite close to those given in the standard treatments, e.g. in Bushell [19] or the papers cited therein.

assuming they exist. A largest α *such that* $\alpha f \leq g$ *always exists. It might be the case that no* β *exists such that* $g \leq \beta f$*. If this is the case, we define* $\beta = \infty$*. It is always the case that*

$$0 \leq \alpha < \infty \qquad 0 < \beta \leq \infty \qquad \alpha \leq \beta.$$

We can also write

$$\begin{aligned} \alpha &= \sup\{t \in \mathbb{R} : tf \leq g\} = \sup\{t \in \mathbb{R} : g - tf \in C\} \\ \beta &= \inf\{t \in \mathbb{R} : g \leq tf\} = \inf\{t \in \mathbb{R} : tf - g \in C\} \end{aligned}$$

where, in the case of β*, we take* $\inf \emptyset = \infty$*.*

We "define" d_H *on* $C \setminus \{0\}$ *by*

$$d_H(f, g) = \ln\left(\frac{\beta}{\alpha}\right) \in [0, \infty].$$

Then d_H *so defined is an extended pseudo metric on* $C \setminus \{0\}$*.*

An equivalent formulation of d_H *is:*

$$d_H(f, g) = \ln\left(\frac{\inf\{t \in \mathbb{R} : tf - g \in C\}}{\sup\{t \in \mathbb{R} : g - tf \in C\}}\right)$$

with the convention: if $\{t \in \mathbb{R} : tf - g \in C\} = \emptyset$ *its infimum is* ∞*.*

We "define" d_H *on* $(C \setminus \{0\}, \sim)$ *by*

$$d_H([f], [g]) = d_H(f, g).$$

d_H*, so defined, is well defined on* $(C \setminus \{0\}, \sim)$*.*

I.e. if $f \sim f'$ and $g \sim g'$ so that

$$f' = \lambda_1 f \quad \textit{and} \quad g' = \lambda_2 g,$$

for some $\lambda_1, \lambda_2 > 0$, then

$$\alpha' = \frac{\lambda_2}{\lambda_1}\ \alpha \qquad \beta' = \frac{\lambda_2}{\lambda_1}\ \beta \tag{1.51}$$

and

$$d_H(f, g) = d_H(f', g').$$

2. *d_H acts as an extended metric on $(C \setminus \{0\}, \sim)$.*

3. *Let $u \in C \setminus \{0\}$. We define the component of u to be*

$$C_u = \{f \in C \setminus \{0\} : d_H(f, u) < \infty\}$$

The components of $C \setminus \{0\}$ form a partition.

d_H is a true metric on

$$(C_u, \sim) = \{[f] : f \in C_u\}$$

Proof. Proof of Part *1.*

Regarding α: Let

$$S_{g-tf} = \{t \in \mathbb{R} : g - tf \in C\},$$

so that $\alpha = \sup_t S_{g-tf}$. Since S_{g-tf} is just the inverse image of C under the continuous map $t \to g - tf$, S_{g-tf} is closed in $\mathbb{R}$. It is immediate that $\alpha \geq 0$ since $g - 0f \in C$. On the other hand, S_{g-tf} must be bounded above; if not, then there exists a sequence of positive numbers $t_n \to \infty$ such that $g - t_n f \in C$. But then, since C is closed under non-negative scaling, we must have $\frac{1}{t_n}\ g - f \in C$. Since C is topologically closed, it

follows that:

$$\lim_{t_n \to \infty} \frac{1}{t_n}\, g - f = -f \in C,$$

contradicting that C is salient. So S_{g-tf} is bounded above. Since S_{g-tf} is closed, non empty (it contains 0), and bounded above, it contains its own supremum, which we denote as α. Moreover $0 \leq \alpha < \infty$.

Regarding β: Let

$$S_{tf-g} = \{t \in \mathbb{R} : tf - g \in C\},$$

so that $\beta = \inf_t S_{g-tf}$, with the convention that $\inf \emptyset = \infty$. C is salient and closed under addition so if $t \in S_{tf-g}$ then $t > 0$. S_{tf-g} is closed in $\mathbb{R}$ since S_{tf-g} is the inverse image of the C w.r.t. to the continuous map $t \to tf - g$ and C is topologically closed. Thus if $S_{tf-g} \neq \emptyset$ then S_{tf-g} is a closed non-empty set bounded below by zero (but not containing zero), hence it contains its infimum, which is denoted β and $\beta > 0$. If $S_{tf-g} = \emptyset$ we of course declare $\beta = \infty$. So $0 < \beta \leq \infty$.

Regarding the metric axioms and proving that $\alpha \leq \beta$:

1. (Non-negativity) $0 \leq \alpha < \infty$ and $0 < \beta \leq \infty$ so the ratio β/α is never indeterminate. In particular: if $\alpha = 0$ and/or $\beta = \infty$ the ratio $\beta/\alpha = \infty$ and in this case

$$d_H(f, g) = \ln(\beta/\alpha) = \infty.$$

If $0 < \alpha, \beta < \infty$ then we have, using the definition of α, β, (1.50),

$$\alpha f \leq g \leq \beta f.$$

This implies $\beta f - \alpha f = (\beta - \alpha) f \in C$. But C is salient, so $\beta - \alpha \geq 0$ proving $\beta \geq \alpha$. So if $0 < \alpha, \beta < \infty$ the ratio β/α is real and in fact is ≥ 1 and we have

$$0 \leq d_H(f, g) = \ln(\beta/\alpha) < \infty.$$

2. (Symmetry) As a consequence of Corollary 1.9.1.4 (page 81)

$$\begin{aligned} d_H(f,g) &= \ln\left(\frac{\beta}{\alpha}\right) \\ &= \ln\left(\frac{\inf\{t \in \mathbb{R} : g - tf \in C\}}{\sup\{t \in \mathbb{R} : f - tg \in C\}}\right) \\ &= d_H(g,h). \end{aligned}$$

3. (Triangle Inequality) Let $f, g, h \in C \setminus \{0\}$.

If $d_H(f,g)$ and/or $d_H(g,h) = \infty$ then automatically $d_H(f,h) \leq d_H(f,g) + d_H(g,h)$ and there is nothing to show.

If $0 \leq d_H(f,g), d(g,h) < \infty$ then $\exists\ \alpha, \beta, \alpha', \beta' \in (0,\infty)$ such that

$$d_H(f,g) = \ln(\beta/\alpha) \text{ and } \alpha f \leq g \leq \beta f$$
$$d_H(g,h) = \ln(\beta'/\alpha') \text{ and } \alpha' g \leq h \leq \beta' g.$$

But then

$$\alpha'\alpha f \leq \alpha' g \leq h \leq \beta' g \leq \beta'\beta f,$$

which implies

$$\alpha'\alpha f \leq h \leq \beta'\beta f.$$

Then, by definition,

$$\alpha'\alpha \leq \sup\{t \in \mathbb{R} : tf \leq h\} = \sup\{t \in \mathbb{R} : h - tf \in C\}$$
$$\beta'\beta \geq \inf\{t \in \mathbb{R} : h \leq tf\} = \inf\{t \in \mathbb{R} : tf - h \in C\}.$$

But then:

$$\frac{\inf\{t \in \mathbb{R} : h \leq tf\}}{\sup\{t \in \mathbb{R} : tf \leq h\}} \leq \frac{\beta'\beta}{\alpha'\alpha},$$

which implies:

$$\begin{aligned} d_H(f,h) &= \ln\left(\frac{\inf\{t\in\mathbb{R} : h \le tf\}}{\sup\{t\in\mathbb{R} : tf \le h\}}\right) \\ &\le \ln\left(\frac{\beta'\beta}{\alpha'\alpha}\right) = \\ &= \ln\left(\frac{\beta}{\alpha}\right) + \ln\left(\frac{\beta'}{\alpha'}\right) \\ &= d_H(f,g) + d(g,h). \end{aligned}$$

So we've proven that d_H satisfies the axioms of being an extended pseudo-metric.

Claim: d_H is well defined on $(C \setminus \{0\}, \sim)$.

Proof of claim: Suppose $f \sim f'$ and $g \sim g'$ so that $f' = \lambda_1 f$ and $g' = \lambda_2 g$ for some $\lambda_1, \lambda_2 > 0$. Let α, α' be the largest numbers such that $\alpha f \le g$ and $\alpha' f' \le g'$. But then α' is the largest number such that $\alpha'\lambda_1 f \le \lambda_2 g$. I.e.

$$\alpha' = \sup\{t : \lambda_2 g - t\lambda_1 f \in C\} = \sup\left\{t : g - t\frac{\lambda_1}{\lambda_2} f \in C\right\} = \frac{\lambda_2}{\lambda_1}\ \alpha.$$

Let β, β' be the smallest numbers (assuming they exist) such that $g \le \beta f$ and $g' \le \beta' f'$. But then β' is the smallest number such that $\lambda_2 g \le \beta'\lambda_1 f$. I.e.

$$\beta' = \inf\{t : t\lambda_1 f - \lambda_2 g \in C\} = \inf\left\{t : t\frac{\lambda_1}{\lambda_2}\ f - g \in C\right\} = \frac{\lambda_2}{\lambda_1}\ \beta.$$

If $\beta = \infty$ then $\beta' = \infty$, if not, then let t' be such that $tf' - g' \in C$. But then $t'\lambda_1 f - \lambda_2 g \in C$ which implies $t'\left(\lambda_1/\lambda_2\right) f - g \in C$. But then $t'\left(\lambda_1/\lambda_2\right) \in \{t : tf - g \in C\}$ which contradicts $\beta = \infty$. So if $\beta = \infty$ then $\beta' = \infty$ and so we can write in all cases $\beta' = \frac{\lambda_2}{\lambda_1}\ \beta$. But then

$$\frac{\beta}{\alpha} = \frac{\beta'}{\alpha'} \ \Rightarrow\ d_H(f,g) = d_H(f',g').$$

Proof of Part *2.*

4. (Identity of Indiscernibles) We show that $d_H([f],[g]) = 0$ if and only if $[f] = [g]$.

Suppose that $d_H([f],[g]) = d_H(f,g) = 0$. Then $\ln(\beta/\alpha) = 1$ which implies $\beta = \alpha \in (0,\infty)$ so we have $\alpha f \leq g \leq \alpha f$. But then

$$(g - \alpha f), \quad (\alpha f - g) \in C.$$

Since C is salient $g - \alpha f = 0$, so $g = \alpha f$; i.e. $[g] = [f]$.

Suppose that $[f] = [g]$. Then $f = \lambda g$ for some $\lambda > 0$. Then

$$\beta = \inf\{t : tf - g \in C\} = \inf\{t : t\lambda g - g \in C\} = \frac{1}{\lambda}$$
$$\alpha = \sup\{t : g - tf \in C\} = \sup\{t : g - t\lambda g \in C\} = \frac{1}{\lambda}$$

which implies

$$d_H([f],[g]) = d_H(f,g) = \ln(\beta/\alpha) = \ln(\lambda/\lambda) = 0.$$

So we've proven that d_H is an extended metric on $(C \setminus \{0\}, \sim)$.

Proof of Part *3.*

For $f, g \in C \setminus \{0\}$ we define

$$f \equiv g \text{ if } d_H(f,g) < \infty.$$

Claim: $\equiv$ is an equivalence relation on $C \setminus \{0\}$.

Proof of Claim. Let $f, g, h \in C \setminus \{0\}$. Then

1. (Reflexivity) $f \equiv f$ because $d_H(f,f) = 0$.
2. (Symmetry) $f \equiv g \Rightarrow g \equiv f$ since $d_H(f,g) = d_H(g,f)$
3. (Transitivity) $f \equiv g$ and $g \equiv h$ means $d_H(f,g), d_H(g,h) < \infty$ but then $d_H(f,h) \leq d_H(f,g) + d_H(g,h) < \infty$ which means $f \equiv h$.

So the claim is proved.

The component of $u \in C \setminus \{0\}$, C_u, is the equivalence class of u w.r.t. $\equiv$. $d_H([f],[g]) = d(f,g)$ is finite if $f, g \in C_u$ so d_H satisfies all the axioms of a true metric on $(C_u, \sim)$ for each $u \in C \setminus \{0\}$. □

1.9.4 What about 0?

Recall Theorem 1.50 (page 84) and its defining of d_H:

Let α, β be the largest and smallest (non-negative) real numbers such that

$$\alpha f \leq g \leq \beta f, \tag{1.52}$$

assuming they exist. A largest α such that $\alpha f \leq g$ always exists. It might be the case that no β exists such that $g \leq \beta f$. If this is the case, we define $\beta = \infty$. It is always the case that

$$0 \leq \alpha < \infty \qquad 0 < \beta \leq \infty \qquad \alpha \leq \beta.$$

We "define" d_H on $C \setminus \{0\}$ by

$$d_H(f,g) = \ln\left(\frac{\beta}{\alpha}\right) \in [0, \infty].$$

We investigate what happens if we allow f or g or both $= 0$.

Suppose $g = 0$ and $f \in C \setminus \{0\}$, then (1.52) becomes

$$\alpha f \leq 0 \leq \beta f.$$

The largest α such that $\alpha f \leq 0$ is the largest α such that $0 - \alpha f \in C$. Since C is salient and $f \neq 0$ it follows that $\alpha = 0$.

The smallest β such that $0 \leq \beta f$ is the smallest β such that $\beta f - 0 \in C$. Since C

is salient and $f \neq 0$ it follows that $\beta = 0$. This differs from when $f, g \in C \setminus \{0\}$ as in that case we always have $0 < \beta \leq \infty$.

So $\beta/\alpha = 0/0$ which is indeterminate. And so declaring that

$$d_H(f, 0) = \ln(\beta/\alpha) = \ln(0/0)$$

makes no sense.

On the other hand, suppose $f = 0$ and $g \in C \setminus \{0\}$, then (1.52) becomes

$$\alpha 0 \leq g \leq \beta 0.$$

The largest α such that $\alpha 0 \leq g$ is the largest α such that $g - \alpha 0 \in C$. Since there is no upper bound to the α which satisfy $g - \alpha 0 \in C$ we have $\alpha = \infty$. This differs from when $f, g \in C \setminus \{0\}$ as in that case we always have $0 \leq \alpha < \infty$.

The smallest β such that $g \leq \beta 0$ is the smallest β such that $\beta 0 - g \in C$. Since C is salient and $g \neq 0$ it follows that no such β exists (which can also happen when $f, g \in C \setminus \{0\}$). The definition of β sets $\beta = \infty$ in such cases.

But then $\beta/\alpha = \infty/\infty$ which is indeterminate. And so declaring that

$$d_H(0, g) = \ln(\beta/\alpha) = \ln(\infty/\infty)$$

makes no sense.

So how can we define $d_H(0, f)$ in way which is consistent with d_H on $C \setminus \{0\}$ and which does not destroy the metric properties of d_H? See Section 1.6.3 (page 55) for a discussion of that topic and my suggestion: to define $d_H(f, 0) = \infty$ if $f \in C \setminus \{0\}$.

That said, unless otherwise noted, we will assume that d_H is only defined on $C \setminus \{0\}$, as that is the standard approach, see Bushell [19].

1.9.5 Defining d_H via b_1, b_2 is equivalent to via $\alpha f \leq g \leq \beta f$.

Theorem 1.9.5.1. *The definition of d_H via ends b_1, b_2, Definition 1.6.2.1 (page 54), is equivalent to defining d_H via $\alpha f \leq g \leq \beta f$, definition given within Theorem 1.9.3.2 (page 84).*

Proof. This is actually proven in part 3 of Theorem 1.9.2.1 (page 82). □

1.10 d_H and the linear structure

1.10.1 $\alpha, \beta, \gamma, \delta > 0 \Rightarrow d_H(\alpha f + \beta g,\ \gamma f + \delta g) < \infty$

Lemma 1.10.1.1. *Let $c_1, c_2 \in C \setminus \{0\}$, not necessarily linearly independent. Let the real numbers $\alpha_x, \beta_x, \alpha_y, \beta_y > 0$ and let*

$$x = \alpha_x c_1 + \beta_x c_2$$

$$y = \alpha_y c_1 + \beta_y c_2.$$

Then $d_H(x, y) < \infty$.

Proof. If $x, y \in C \setminus \{0\}$ then Lemma 1.6.5.1 (page 60) implies

$$d_H(x, y) = |\ln(m(x/y)\ m(y/x))|$$

and Proposition 1.3.1.1 part 3 (page 31) implies

$$0 \leq m(y/x) < \infty.$$

So to show $d_H(x, y) < \infty$ it suffices to show that $m(x/y)$ and $m(y/x)$ are both positive.

$$\begin{aligned} m(y/x) &= \sup\{t : y - tx \in C\} \\ &= \sup\{t : (\alpha_y c_1 + \beta_y c_2) - t(\alpha_x c_1 + \beta_x c_2) \in C\} \\ &= \sup\{t : (\alpha_y - t\alpha_x)c_1 + (\beta_y - t\beta_x)c_2 \in C\} \\ &\geq \min\left\{\frac{\alpha_y}{\alpha_x}, \frac{\beta_y}{\beta_x}\right\} \\ &> 0 \end{aligned}$$

$$
\begin{aligned}
m(x/y) &= \sup\{t : x - ty \in C\} \\
&= \sup\{t : (\alpha_x c_1 + \beta_x c_2) - t(\alpha_y c_1 + \beta_y c_2) \in C\} \\
&= \sup\{t : (\alpha_x - t\alpha_y)c_1 + (\beta_x - t\beta_y)c_2 \in C\} \\
&\geq \min\left\{\frac{\alpha_x}{\alpha_y}, \frac{\beta_x}{\beta_y}\right\} \\
&> 0
\end{aligned}
$$

□

1.10.2 $z = \lambda_x x + \lambda_y y \;\Rightarrow\; d_H(x,y) = d_H(x,z) + d_H(z,y)$

Proposition 1.10.2.1. *Let $x, y \in C \setminus \{0\}$ satisfy $d_H(x,y) < \infty$. Note that x, y are not necessarily linearly independent. Let λ_x, λ_y be any non-negative real numbers (but both not zero), and let*

$$z = \lambda_x x + \lambda_y y,$$

so that $z \in (Span(x,y) \bigcap C) \setminus \{0\}$. Then $d_H(x,y) = d_H(x,z) + d_H(z,y)$.

Proof. If x, y are linearly dependent then y and z are multiples of x and it is immediate that $d_H(x,y) = d_H(x,z) = d_H(z,y) = 0$ and the result follows trivially.

If x, y are linearly independent, then, by Proposition 1.6.1.1 (page 53), there exists a pair of ends, b_1, b_2 for x, y in $\mathrm{Span}(x,y) \bigcap C$.

Express x, y in terms of the b_1, b_2:

$$
\begin{aligned}
x &= x_1 b_1 + x_2 b_2 \\
y &= y_1 b_1 + y_2 b_2.
\end{aligned}
$$

Hence,

$$\begin{aligned} z &= \lambda_x x + \lambda_y y \\ &= \lambda_x(x_1 b_1 + x_2 b_2) + \lambda_y(y_1 b_1 + y_2 b_2) \\ &= (\lambda_x x_1 + \lambda_y y_1) b_1 + (\lambda_x x_2 + \lambda_y y_2) b_2 \end{aligned}$$

and

$$d_H(x, y) = \left| \ln \left(\frac{x_2}{x_1} \; \frac{y_1}{y_2} \right) \right| \tag{1.53}$$

$$d_H(x, z) = \left| \ln \left(\frac{x_2}{x_1} \; \frac{\lambda_x x_1 + \lambda_y y_1}{\lambda_x x_2 + \lambda_y y_2} \right) \right| \tag{1.54}$$

$$d_H(z, y) = \left| \ln \left(\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1} \; \frac{y_1}{y_2} \right) \right|. \tag{1.55}$$

The following two lemmas will help us to finish the proof.

Lemma 1.10.2.2. *Suppose that b_1, b_2 are a pair of ends for x, y and that*

$$x = x_1 b_1 + x_2 b_2$$

$$y = y_1 b_1 + y_2 b_2.$$

If $d_H(x, y) < \infty$ and x, y are linearly independent then

$$x_1, y_1, x_2, y_2 > 0.$$

Proof. By the definition of d_H, see Definition 1.6.2.1 (page 54):

$$d_H(x, y) = \left| \ln \left(\frac{x_2}{x_1} \; \frac{y_1}{y_2} \right) \right|$$

Suppose $x_1 = 0$. Then

$$d_H(x,y) = \left|\ln\begin{pmatrix} x_2 & y_1 \\ 0 & y_2 \end{pmatrix}\right| = \infty$$

unless $x_2 = 0$ or $y_1 = 0$.

If $x_2 = 0$ then $x = 0$, contradicting x, y linearly independent. If $y_1 = 0$ then x, y are linearly dependent, again contradicting x, y linearly independent. So $x_1 \neq 0$.

Similar arguments (symmetry, relabeling) show that y_1, x_2, and y_2 can't be zero.

□

The following lemma is a general result about non-negative numbers.

Lemma 1.10.2.3. *Suppose that $x_1, x_2, y_1, y_2 > 0$ and that $a, b \geq 0$ but not both zero, and that*

$$\frac{x_2}{x_1} \leq \frac{y_2}{y_1}.$$

Then

$$\frac{x_2}{x_1} \leq \frac{ax_2 + by_2}{ax_1 + by_1} \leq \frac{y_2}{y_1}.$$

Proof.

$$\begin{aligned}
\frac{ax_2 + by_2}{ax_1 + by_1} - \frac{x_2}{x_1} &= \frac{(ax_2 + by_2)x_1}{(ax_1 + by_1)x_1} - \frac{x_2(ax_1 + by_1)}{x_1(ax_1 + by_1)} \\
&= \frac{(ax_2 + by_2)x_1 - x_2(ax_1 + by_1)}{(ax_1 + by_1)x_1} \\
&= \frac{(ax_2x_1 + by_2x_1 - ax_1x_2 - by_1x_2)}{(ax_1 + by_1)x_1} \\
&= \frac{b(y_2x_1 - y_1x_2)}{(ax_1 + by_1)x_1}
\end{aligned}$$

Since $\frac{x_2}{x_1} \leq \frac{y_2}{y_1}$ and $x_1, y_1 > 0$ we have

$$y_1x_2 - y_2x_1 \leq 0 \text{ and hence } y_2x_1 - y_1x_2 \geq 0.$$

Hence,

$$\frac{ax_2 + by_2}{ax_1 + by_1} - \frac{x_2}{x_1} = \frac{b(y_2x_1 - y_1x_2)}{(ax_1 + by_1)x_1} \geq 0$$

and so

$$\frac{ax_2 + by_2}{ax_1 + by_1} \geq \frac{x_2}{x_1}.$$

A similar argument shows that

$$\frac{y_2}{y_1} \geq \frac{ax_2 + by_2}{ax_1 + by_1}.$$

□

We return to our proof of Proposition 1.10.2.1:

Since we are assuming that $d_H(x, y) < \infty$ and x, y are linearly independent, Lemma 1.10.2.2 (page 96) implies that $x_1, x_2, y_1, y_2 > 0$. We are also assuming that $\lambda_x, \lambda_y \geq 0$ with at most one of them equaling 0.

Without loss of generality, by relabeling if necessary, we can suppose

$$\frac{x_2}{x_1} \leq \frac{y_2}{y_1}.$$

Then Lemma 1.10.2.3 (page 97) implies that

$$\frac{x_2}{x_1} \leq \frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1} \leq \frac{y_2}{y_1}. \tag{1.56}$$

Multiplying (1.56) by $\frac{x_1}{x_2}$ yields

$$1 = \frac{x_1}{x_2}\,\frac{x_2}{x_1} \quad \leq \quad \frac{x_1}{x_2}\,\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1} \quad \leq \quad \frac{x_1}{x_2}\,\frac{y_2}{y_1}. \tag{1.57}$$

Applying ln to (1.57), and recalling (1.53) and (1.54), we get

$$0 = \ln(1) \quad \leq \quad \underbrace{\ln\left(\frac{x_1}{x_2}\,\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1}\right)}_{= d_H(x,z)} \quad \leq \quad \underbrace{\ln\left(\frac{x_1}{x_2}\,\frac{y_2}{y_1}\right)}_{= d_H(x,y)}. \tag{1.58}$$

Multiplying (1.56) by $\frac{y_1}{y_2}$ yields

$$\frac{y_1}{y_2}\,\frac{x_2}{x_1} \quad \leq \quad \frac{y_1}{y_2}\,\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1} \quad \leq \quad \frac{y_1}{y_2}\,\frac{y_2}{y_1} = 1. \tag{1.59}$$

Inverting (1.59) yields

$$\frac{y_2}{y_1}\,\frac{x_1}{x_2} \quad \geq \quad \frac{y_2}{y_1}\,\frac{\lambda_x x_1 + \lambda_y y_1}{\lambda_x x_2 + \lambda_y y_2} \quad \geq \quad \frac{y_2}{y_1}\,\frac{y_1}{y_2} = 1. \tag{1.60}$$

Applying ln to (1.60) and recalling (1.53) and (1.55), we get

$$\underbrace{\ln\left(\frac{y_2}{y_1}\,\frac{x_1}{x_2}\right)}_{= d_H(x,y)} \quad \geq \quad \underbrace{\ln\left(\frac{y_2}{y_1}\,\frac{\lambda_x x_1 + \lambda_y y_1}{\lambda_x x_2 + \lambda_y y_2}\right)}_{= d_H(z,y)} \quad \geq \quad \ln(1) = 0. \tag{1.61}$$

Using (1.58) and (1.61) we get

$$\begin{aligned} d_H(x,z) + d(z,y) &= \ln\left(\frac{x_1}{x_2}\,\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1}\right) \;+\; \ln\left(\frac{y_2}{y_1}\,\frac{\lambda_x x_1 + \lambda_y y_1}{\lambda_x x_2 + \lambda_y y_2}\right) \\ &= \ln\left(\frac{x_1}{x_2}\,\frac{\lambda_x x_2 + \lambda_y y_2}{\lambda_x x_1 + \lambda_y y_1}\,\frac{y_2}{y_1}\,\frac{\lambda_x x_1 + \lambda_y y_1}{\lambda_x x_2 + \lambda_y y_2}\right) \\ &= \ln\left(\frac{x_1}{x_2}\,\frac{y_2}{y_1}\right) \\ &= d_H(x,y). \end{aligned}$$

□

1.11 $\leq,\ E_u,\ |x|_u,\ < x, z >$

As usual, C will denote a closed, convex, pointed by the origin salient cone contained in a Banach Space, which will be denoted by V or X.

1.11.1 A collection of results for $\leq$

By Proposition 1.3.0.7 (page 30) the cone C partially orders X if we define $\forall x, y \in X$,

$$x \leq y \text{ if } y - x \in C.$$

The following are easy consequences of the definition of $\leq$:

Lemma 1.11.1.1. *Let $w, x, y, z_1, z_2 \in X$, X a Banach Space, and let the closed, convex, salient, pointed by the origin cone $C \subset X$.*

1. *$0 \leq x$ implies $x - 0 = x \in C$*
2. *$x \leq 0$ implies $0 - x = -x \in C$.*

   Note. *C is salient; i.e. $C \cap -C = \{0\}$, so if $x \leq 0$ then $x \notin C$ unless $x = 0$.*
3. *$x \leq y$, then $-y \leq -x$.*
4. *$x \leq y$ and $\alpha > 0$ then $\alpha x \leq \alpha y$.*
5. *$x \leq y$ and $\alpha > 0$ then $-\alpha y \leq -\alpha x$.*
6. *$z_1 \leq z_2$ implies $w + z_1 \leq w + z_2$.*
7. *Suppose for $i = 1, 2$ that $x_i, y_i, z_i \in X$. If*

$$x_1 \leq y_1 \leq z_1$$

$$x_2 \leq y_2 \leq z_2$$

then $(x_1 + x_2) \leq (y_1 + y_2) \leq (z_1 + z_2)$.

Proof. Parts 1 and 2 are trivial.

Proof of 3. $-x - (-y) = y - x \in C$ since $x \leq y$.

Proof of 4. $\alpha y - \alpha x = \alpha(y - x) \in C$ since $y - x \in C$ and C is closed under non-negative scaling.

Proof of 5. $(-\alpha x) - (-\alpha y) = \alpha(y - x) \in C$ since $y - x \in C$ and C is closed under non-negative scaling.

Proof of 6. $(w + z_2) - (w + z_1) = z_2 - z_1 \in C$

Proof of 7. We prove $x_1 + x_2 \leq y_1 + y_2$:

$(y_1 + y_2) - (x_1 + x_2) = (y_1 - x_1) + (y_2 - x_2)$. But $x_1 \leq y_1$ and $x_2 \leq y_2$ imply $(y_1 - x_1),\ (y_2 - x_2) \in C$. C is closed under addition, so $(y_1 - x_1) + (y_2 - x_2) \in C$. So $x_1 + x_2 \leq y_1 + y_2$.

Similarly, $y_1 + y_2 \leq z_1 + z_2$. Since $\leq$ is a partial order, this implies $x_1 + x_2 \leq y_1 + y_2 \leq z_1 + z_2$. □

The following notation is useful:

Definition 1.11.1.2.

$$< x, z > = \{y \in X : x \leq y \leq z\} = \{y \in X : y - x \in C \text{ and } z - y \in C\}$$

Lemma 1.11.1.3. *If* $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ *and* $\alpha > 0$ *then*

$$\alpha \sup\{\|x\| : 0 \leq x \leq u\} = \sup\{\|x\| : 0 \leq x \leq \alpha u\} \tag{1.62}$$

$$= \sup\{\|x\| : -\alpha u \leq x \leq 0\} \tag{1.63}$$

$$= \sup\{\|x\| : -\alpha u \leq x \leq \alpha u\} \tag{1.64}$$

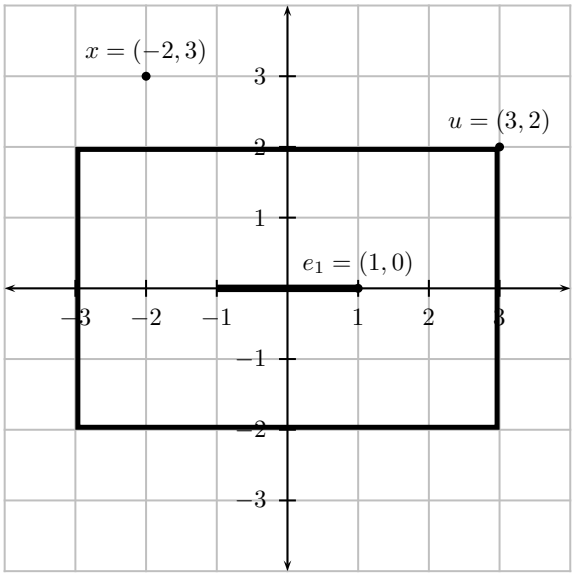


Figure 1.2: Two examples of $E_u^\alpha = (\alpha u - C) \cap (-\alpha u + C)$. Example 1: Let $C = \mathbb{R}^2_{\geq 0}$, $u = (3, 2)$, and $\alpha = 1$. Then E_u^α is the closed rectangular region shown. Note $|x|_u = \inf\{\alpha > 0 : -\alpha u \leq x \leq \alpha u\} = \max_i\{|x_i|/|u_i| : u_i \neq 0\} = \max\left\{\frac{|-2|}{|3|}, \frac{|3|}{|2|}\right\} = 3/2$ since $C = \mathbb{R}^2_{\geq 0}$. Notice that if we scale the rectangular region by $3/2$, its boundary will contain x. Example 2: Again, let $C = \mathbb{R}^2_{\geq 0}$ and $\alpha = 1$, but now let $u = e_1 = (1, 0)$, then E_u^α is the closed line segment shown. See Definitions 1.11.2.2 (page 103) and 1.11.3.6 (page 108).

Proof. $\{0 \leq x \leq u\} = < 0, u >$, so

Proof of (1.62): $x \in < 0, u > \Leftrightarrow \alpha x \in < 0, \alpha u >$

Proof of (1.63): $x \in < 0, u > \Leftrightarrow -x \in < -\alpha u, 0 >$

Proof of (1.64): $x \in < -\alpha u, \alpha u > \Leftrightarrow x \in < -\alpha u, 0 > \cup < 0, \alpha u > .$

□

1.11.2 E_u and E_u^α

The following two definitions are standard. See for example [76] or [92].

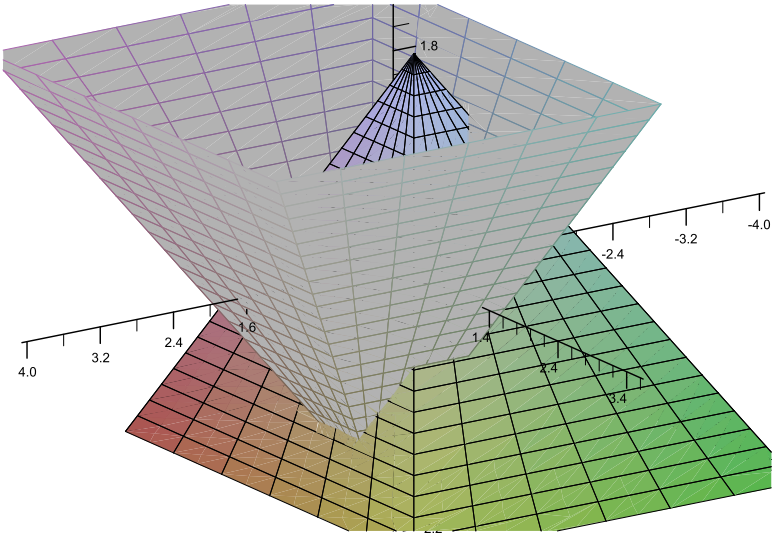


Figure 1.3: A 3 dimensional example of the formation of $E_u^\alpha = (\alpha u - C) \cap (-\alpha u + C)$: Let C be the cone $C = \{(zx, zy, z) : (x, y) \in [0,1] \times [0,1], z \geq 0\}$, let $u = (.3, .7, 2)$, and let $\alpha = 1$. In Maple, $(u - C)$ and $(-u + C)$ are plotted and shown above. (Actually only parts of $(u - \partial C)$ and $(-u + \partial C)$ are plotted as C itself is solid and infinite.) See Definition 1.11.2.2 (page 103) for more on E_u^α.

Definition 1.11.2.1. Let $u \in C \setminus \{0\}$. We define E_u as follows:

$$E_u = \{x \in X : \exists \alpha > 0,\ -\alpha u \leq x \leq \alpha u\}$$

See Figures 1.2 (page 102), 1.3 (page 103)[4] and Definition 1.11.2.2 (page 103).

Definition 1.11.2.2. Let $u \in C \setminus \{0\}$ and let $\alpha > 0$ be $\in \mathbf{R}$.

$$E_u^\alpha = \{x \in X : -\alpha u \leq x \leq \alpha u\}$$

The following technical lemma regarding E_u^α and $< 0, u >$ is useful.

Lemma 1.11.2.3. *Let $\alpha \in \mathbf{R}$ be ≥ 0 and $u \in C \setminus \{0\}$. Then*

1. *$\{x \in X : x \leq \alpha u\} = \alpha u - C$ is closed in X.*

[4]The following Maple commands were used to produce Figure 1.3 (page 103): $C := (x, z) \rightarrow [z * \text{piecewise}(0 \leq x < 1, x, 1 \leq x < 2, 1, 2 \leq x < 3, 3 - x, 3 \leq x \leq 4, 0), z * \text{piecewise}(0 \leq x < 1, 0, 1 \leq x < 2, x - 1, 2 \leq x < 3, 1, 3 \leq x \leq 4, 4 - x), z]$; $u := [.3, .7, 2]$;
plot3d($\{-C(x, z) + u, C(x, z) - u\}, x = 0..4, z = 0..4.2,$ axes = normal);

2. $\{x \in X : -\alpha u \leq x\} = -\alpha u + C$ *is closed in* X.

3. $E_u^\alpha = \{x \in X : -\alpha u \leq x \leq \alpha u\} = \underbrace{\{x \in X : -\alpha u \leq x\}}_{-\alpha u + C} \bigcap \underbrace{\{x \in X : x \leq \alpha u\}}_{\alpha u - C}$ *is closed in* X. *(See Figure 1.2 (page 102).)*

4. $< 0, u >= \{x \in X : 0 \leq x \leq u\}$ *is closed in* X.

Proof. First we show that $\{x \in X : x \leq \alpha u\} = \alpha u - C$

$$\begin{aligned} x \leq \alpha u &\Leftrightarrow \alpha u - x \in C \\ &\Leftrightarrow \alpha u - x = c \text{ for some } c \in C \\ &\Leftrightarrow \alpha u - c = x \text{ for some } c \in C \\ &\Leftrightarrow x \in \alpha u - C. \end{aligned}$$

C is closed so Corollary 1.2.2.2 (page 29) guarantees that $\alpha u - C$ and $\alpha u + C$ is closed. Hence *1.* and *2.* are proven.

The intersection of two closed sets is closed, so $\{x \in X : -\alpha u \leq x \leq \alpha u\}$ is closed. Hence *3.* is proven.

Finally,

$$< 0, \alpha u >= \underbrace{\{x \in X : 0 \leq x\}}_{\text{closed}} \bigcap \underbrace{\{x \in X : x \leq \alpha u\}}_{\text{closed}}.$$

Hence *4.* is proven. □

1.11.3 The component C_u, more on E_u, and $|\ |_u$

The following definition/notation is standard. See for example [76].

Definition 1.11.3.1. Let $u \in C \setminus \{0\}$.

$$C_u = \{c \in C \setminus \{0\} : d_H(u, c) < \infty\}.$$

C_u is called the component of C containing u. C_u consists of the elements of $C \setminus \{0\}$ which are a finite distance from u.

Theorem 1.11.3.2. *C_u is closed under positive linear combinations. C_u is a convex salient cone.*

Proof. Let $\lambda_f, \lambda_g \in \mathbb{R}$ both be positive. Suppose $f, g \in C_u$ so that

$$d_H(u, f) = d_H(u, \lambda_f f) \qquad \text{and} \qquad d_H(u, g) = d_H(u, \lambda_g g)$$

are both finite. Then there exists finite $\alpha_f, \beta_f, \alpha_g, \beta_g > 0$ such that

$$\begin{aligned} \alpha_f u \ &\leq \ \lambda_f f \ \leq \ \beta_f u \\ \alpha_g u \ &\leq \ \lambda_g g \ \leq \ \beta_g u. \text{ But then} \\ (\alpha_f + \alpha_g) u \ &\leq \ \lambda_f f + \lambda_g g \ \leq \ (\beta_f + \beta_g) u \end{aligned}$$

by Lemma 1.11.1.1 part 7 (page 100). This implies

$$d_H\left(\lambda_f f + \lambda_g g, \ u\right) < \infty. \tag{1.65}$$

The general case follows from finite induction.

(1.65) implies C_u is a convex cone. Since $C_u \subset C$ and C is salient it follows that C_u is salient. □

Remark 1.11.3.3. We also proved that C_u is a convex salient cone in Theorem 1.8.3.1 (page 76). That proof (page 76) was longer and much less elegant than this proof. Defining d_H via $\alpha f \leq g \leq \beta f$ has its advantages.

Remark 1.11.3.4. Usually C_u is not closed.

Theorem 1.11.3.5. *Let $u \in C \setminus \{0\}$. Then $E_u = Span(C_u)$.*

Proof. $x \in E_u$ implies $\exists \alpha > 0$ such that $-\alpha u \leq x \leq \alpha u$. This in turn implies $\exists c_1, c_2 \in C$ such that $\alpha u - x = c_1$ and $x - (-\alpha u) = c_2$. We write this nicely as

$$\alpha u - x = c_1 \text{ and } \alpha u + x = c_2. \tag{1.66}$$

Adding the two equations from (1.66) we get

$$2\alpha u = c_1 + c_2 \text{ which implies } u = \frac{1}{2\alpha}\ c_1 + \frac{1}{2\alpha}\ c_2. \tag{1.67}$$

Let

$$c = 2c_1 + c_2. \tag{1.68}$$

By (1.67), (1.68) and Lemma 1.10.1.1 (page 94),

$$d_H(u, c) \ = \ d_H\left(\frac{1}{2\alpha}\ c_1 + \frac{1}{2\alpha}\ c_2,\ 2c_1 + c_2\right) \ < \ \infty \tag{1.69}$$

since the the coefficients of c_1, c_2 in (1.69) are all positive. Hence

$$c \in C_u.$$

Using the definition of c, see (1.68), and the first equation in (1.67) we write c_1 as

$$\begin{aligned} c_1 &= (\underbrace{2c_1 + c_2}_{c}) - (\underbrace{c_1 + c_2}_{2\alpha u}) \\ &= c - 2\alpha u. \end{aligned}$$

From the first equation in (1.66) and using $c_1 = c - 2\alpha u$ we express x as linear

combination of u, c:

$$\begin{aligned} x &= \alpha u - c_1 \\ &= \alpha u - (c - 2\alpha u) \\ &= (3\alpha)u - c. \end{aligned}$$

Since $u, c \in C_u$ we have shown that $x \in \text{Span}(u, c) \subset \text{Span}(C_u)$.

$$E_u = \{x \in X : \exists \alpha > 0,\ -\alpha u \leq x \leq \alpha u\}$$

Let $x \in \text{Span}(C_u) \setminus \{0\}$. This means we can express x as a finite linear combination of elements from C_u,

$$x = \underbrace{\sum_{i=1}^{n^+} \beta_i^+ c_i}_{c} - \underbrace{\sum_{i=1}^{n^-} \beta_i^- d_i}_{d}$$

with n^+, n^- non-negative integers, the $c_i, d_i \in C_u$, and the $\beta_i^+, \beta_i^- > 0$. By Theorem 1.11.3.2 (page 105) $c, d \in C_u \cup \{0\}$. So, assuming that $c, d \neq 0$ there exist finite $\alpha_c, \beta_c, \alpha_d, \beta_d > 0$ such that

$$\alpha_c u \leq c \leq \beta_c u \tag{1.70}$$

$\alpha_d u \leq d \leq \beta_d u$. But then, by Lemma 1.11.1.1, part 3 (page 100):

$$-\beta_d u \leq -d \leq -\alpha_d u. \tag{1.71}$$

By Lemma 1.11.1.1, part 7 (page 100) if we sum (1.70) and (1.71) the result is

$$(\alpha_c - \beta_d)u \leq c - d \leq (\beta_c - \alpha_d)u. \tag{1.72}$$

Claim 1: $(\beta_c - \alpha_d)u \leq |\beta_c - \alpha_d|u$.

Proof of claim 1: If $(\beta_c - \alpha_d) \geq 0$ there is nothing to prove. So suppose $(\beta_c - \alpha_d) <$

0. Then $(\beta_c - \alpha_d)u \leq |\beta_c - \alpha_d|u$ because

$$|\beta_c - \alpha_d|u - (\beta_c - \alpha_d)u = |\beta_c - \alpha_d|u - -|\beta_c - \alpha_d|u = 2|\beta_c - \alpha_d|u \in C$$

because $u \in C$ and C is closed w.r.t. non-negative scaling.

Claim 2: $-|\alpha_c - \beta_d|u \leq (\alpha_c - \beta_d)u$.

Proof of claim 2: If $(\alpha_c - \beta_d) \leq 0$ there is nothing to prove. So suppose $(\beta_c - \alpha_d) > 0$. Then $-|\alpha_c - \beta_d|u \leq (\alpha_c - \beta_d)u$ because

$$(\alpha_c - \beta_d)u - -|\alpha_c - \beta_d|u = 2|\alpha_c - \beta_d|u \in C$$

because $u \in C$ and C is closed w.r.t. non-negative scaling.

Combining (1.72) with claims 1 and 2, we have

$$-|\alpha_c - \beta_d|u \leq (\alpha_c - \beta_d)u \leq c - d \leq (\beta_c - \alpha_d)u \leq |\beta_c - \alpha_d|u. \tag{1.73}$$

Let $\alpha = \max\{|\alpha_c - \beta_d|, |\beta_c - \alpha_d|, 1\}$. Then $\alpha > 0$ and

$$-\alpha u \leq (x = c - d) \leq \alpha u.$$

So $x \in E_u$. □

The following definition/notation is standard. See for example [76].

Definition 1.11.3.6. If $x \in E_u$ then we define

$$|x|_u = \inf\{\alpha > 0 : -\alpha u \leq x \leq \alpha u\}$$

See Figure 1.2 (page 102).

Proposition 1.11.3.7. $|\ |_u$ *is a norm on* E_u. *See* [5].

Proof. Let $k \in \mathbb{R}$ and $x, y \in E_u$.

1. (Positive definiteness: $|x|_u \geq 0$ and $|x|_u = 0 \Leftrightarrow x = 0$.)

If $x \in E_u$ there exists some $\alpha > 0$ such that $-\alpha u \leq x \leq \alpha u$ so $|x|_u \geq 0$. C is closed under positive scaling so for all $\alpha > 0$ we have the relation: $-\alpha u \leq 0 \leq \alpha u$. This implies $|0|_u = 0$. If $|x|_u = 0$ then $\alpha u - x \in C$ for every $\alpha > 0$. C is closed so if we let $\alpha \to 0$ we get $-x \in C$. Similarly, if $|x|_u = 0$ then $x - \alpha u \in C$ for every $\alpha > 0$. C is closed so if we let $\alpha \to 0$ we get $x \in C$. Since C is salient $x = 0$.

2. (Positive homogeneity: $|kx|_u = |k|\ |x|_u$.)

If k = 0 then, by 1. $|kx|_u = 0 = k|x|_u$ and we are done. If $k \neq 0$, then by Lemma 1.11.1.1, parts 4 and 5 (page 100), we have:

$$\begin{aligned} k > 0 \Rightarrow \quad & -\alpha u \leq x \leq \alpha u \quad \Leftrightarrow \quad -k\alpha u \leq kx \leq k\alpha u. \\ k < 0 \Rightarrow \quad & -\alpha u \leq x \leq \alpha u \quad \Leftrightarrow \quad -k\alpha u \geq kx \geq k\alpha u. \end{aligned}$$

This implies $|kx|_u = |k|\ |x|_u$

3. (Triangle inequality: $|x + y|_u \leq |x|_u + |y|_u$.)

Given $\epsilon > 0$ there exists α_x and α_y such that

$$-\alpha_x u \ \leq \ x \ \leq \ \alpha_x u \tag{1.74}$$

$$-\alpha_y u \ \leq \ y \ \leq \ \alpha_y u \tag{1.75}$$

and

$$\begin{aligned} |x|_u \ &\leq \alpha_x \ \leq \ |x|_u + \epsilon/2 \\ |y|_u \ &\leq \alpha_y \ \leq \ |y|_u + \epsilon/2. \end{aligned}$$

[5] Nussbaum [76] mentions this in his comments preceding Theorem 1.2, but gives no proof.

By Lemma 1.11.1.1, part 7, summing (1.74) (1.75) will yield

$$-(\alpha_x + \alpha_y)u \leq x + y \leq (\alpha_x + \alpha_y)u$$

which implies

$$\begin{aligned} |x+y|_u &\leq (\alpha_x + \alpha_y) \\ &\leq |x|_u + \epsilon/2 + |y|_u + \epsilon/2 \\ &= |x|_u + |y|_u + \epsilon. \end{aligned}$$

Letting $\epsilon \to 0$ we get $|x+y|_u \leq |x|_u + |y|_u$. □

1.11.4 Example: $|\ |_u$ on $C = \mathbb{R}^n_{\geq 0}$.

Let $C = \mathbb{R}^n_{\geq 0}$ and let $u = (u_1, u_2, \ldots, u_n) \in C \setminus \{0\}$. Suppose $x \in E_u$. Then $\exists\ \alpha > 0$ such that

$$-\alpha u_j \leq x_j \leq \alpha u_j \ \ \forall\ j = 1, 2, \ldots, n \quad (\text{ i.e. } x \in E_u^\alpha)$$

and so $0 \leq |x_j| \leq \alpha u_j$. Consequently, for those j for which $u_j = 0$, we have $x_j = 0$; and for those j for which $u_j > 0$ we must have $0 \leq |x_j/u_j| \leq \alpha$, so that

$$|x|_u = \max\{|x_j/u_j| : u_j > 0,\ j = 1, 2, \ldots, n\}.$$

Remark 1.11.4.1. Geometrically, we can think of $|x|_u$ as what we must scale the (possibly n dimensional) rectangle E_u^1 so that x will become situated on its boundary. See Figure 1.2 (page 102).

1.12 Completeness

1.12.1 Introduction to completeness w.r.t. d_H

As usual, C will denote a closed, convex, pointed-by-the-origin salient cone contained in a Banach Space, which will be denoted by V or X.

Definition 1.12.1.1. The following definitions are useful, especially for [92, Lemma 1], which follows: Let $R \subset C \setminus \{0\}$.

1. R being a component means if $x, y \in R$ then $d_H(x, y) < \infty$
2. The metric space $\hat{R}$ is $(R, \sim)$ with the metric d_H.
3. $\hat{R}$ being normal means for every $x \in R$ that the conical segment

$$< 0, x >= \{z : 0 \leq z \leq x\}$$

is bounded in the norm [6].

Proposition 1.12.1.2. *Let $u \in C \setminus \{0\}$. The component $C_u \subset X$ is normal if and only if $< 0, u >$ is bounded w.r.t. X's norm.*

Proof. If C_u is normal then the definition of normal, Definition 1.12.1.1 (page 111), immediately implies that $< 0, u >$ is bounded w.r.t. X's norm (since $u \in C_u$).

Suppose $< 0, u >$ is bounded w.r.t. X's norm and $x \in C_u$. Since $x \in C_u$ we have $d_H(x, u) < \infty$ and so there exists $\alpha, \beta \in (0, \infty)$ such that $\alpha u \leq x \leq \beta u$. Let B = an upper bound for $\|< 0, u >\|_X = \{\|z\|_X : z \in < 0, u >\}$.

Claim: βB is an upper bound for $\|< 0, \beta u >\|_X$.

Proof of Claim: Suppose $z \in < 0, \beta u >$. Then $z \in C$ and $\beta u - z \in C$. Since $\beta > 0$ we have $\beta^{-1} z,\ u - \beta^{-1} z \in C$ and so $\beta^{-1} z \in < 0, u >$. This implies $\|\beta^{-1} z\|_X < B$. But then $\|z\|_X < \beta B$. So the claim is proved.

[6] This definition of normal is taken word for word from [92, Zabreiko, Krasnosel'skii, and Pokornyi]. Note $< 0, x >= \{z : 0 \leq z \leq x\} = \{z \in C : x - z \in C\} = C \cap (x - C)$.

Since $0 \leq x \leq \beta u$ we have $< 0, x > \subset < 0, \beta u >$ which implies $\|< 0, x >\|_X \leq \|< 0, \beta u >\|_X \leq \beta B$. Since $x \in C_u$ was arbitrary C_u is normal. □

In 1971 Zabreiko, Krasnosel'skii, and Pokornyi [92, Lemma 1] proved the following result about when

$$\widehat{R} = (R, \sim)$$

will be complete relative to d_H:

[92, Lemma 1. Zabreiko, Krasnosel'skii, and Pokornyi] *The metric space $\hat{R}$ is complete if and only if the component R is normal.*

In 1989 Nussbaum [76, Theorem 1.2] gave a modified version of [92, Lemma 1] which we reproduce here with some minor changes [7]:

[76, Theorem 1.2. Nussbaum] *Let C be pointed, closed, convex, salient cone in a Banach space X. Suppose that $u \in C \setminus \{0\}$ and let C_u denote the component of C containing u; i.e.*

$$C_u = \{x \in C : d_H(x, u) < \infty\}.$$

Let $\widehat{C}_u = (C_u, \sim)$ Then the following statements are equivalent:

1. *$(\widehat{C}_u, d_H) = ((C_u, \sim), d_H)$ is a complete metric space.*

2. *$\sup\{\|x\| : 0 \leq x \leq u\} < \infty$. See* [8].

3. *E_u is a complete normed linear space with respect to $|x|_u$.*

[7] Nussbaum [76, Theorem 1.2] does not use the notation $(C_u, \sim)$ or $\widehat{C}_u$. Instead he uses the notation $\Sigma = \{x \in C_u : \|x\| = 1\}$, identifying $[x] \in (C_u, \sim)$ with $\frac{x}{\|x\|}$, where $\|\ \|$ is the underlying original norm on X (the one that makes X into a Banach Space). We will use the notation $\widehat{C}_u$ in deference to the notation originally used in [92, Zabreiko, Krasnosel'skii, and Pokornyi], or our usual notation, $(C_u, \sim)$.

[8] By Proposition 1.12.1.2 (page 111) the condition $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ is equivalent to C_u being normal.

See Definition 1.11.2.1 (page 102) of E_u and Definition 1.11.3.6 (page 108) of $|x|_u$.

Nussbaum gives no proof of [76, Theorem 1.2]. Instead he cites [92, Lemma 1]. This is fair as the proof given in [92] can be modified to fit Nussbaum's version. The only problem is that the proof in [92] leaves out many details, some of which we supply over the next few sections.

Remark 1.12.1.3. If we allow for d_H to be extended to 0 by defining $d_H(0, f) = \infty$ for $f \in C \setminus \{0\}$ then $\widehat{R} = [0]$ will be complete and satisfy [76, Theorem 1.2. Nussbaum]. So, in [76, Theorem 1.2. Nussbaum] we could make the condition $u \in C$, rather than $u \in C \setminus \{0\}$. See Section 1.6.3 (page 55) for more details on extending d_H to the origin.

1.12.2 $\prod_{n=1}^{\infty} \left(1 + \frac{1}{2^n}\right) < e$

In this section we prove some technical results.

Lemma 1.12.2.1. *Let $x \in \mathbf{R}$. Then*

$$1 + x \leq e^x.$$

The equality holds only if $x = 0$.

Proof.

$$\begin{aligned} e^x - (1 + x) &= \sum_{n=2}^{\infty} \frac{x^n}{n!} && (1.76) \\ &= \sum_{n=1}^{\infty} \left(\frac{x^{2n}}{(2n)!} + \frac{x^{2n+1}}{(2n+1)!} \right) \\ &= \sum_{n=1}^{\infty} \frac{x^{2n}}{(2n)!} \left(1 + \frac{x}{2n+1} \right) && (1.77) \end{aligned}$$

If $x > 0$ it is easy to see that each term in the infinite series (1.76), is positive. If $x = 0$ then both $1 + x$ and $e^x = 1$. If x satisfies $-1 < x < 0$, then each term in the

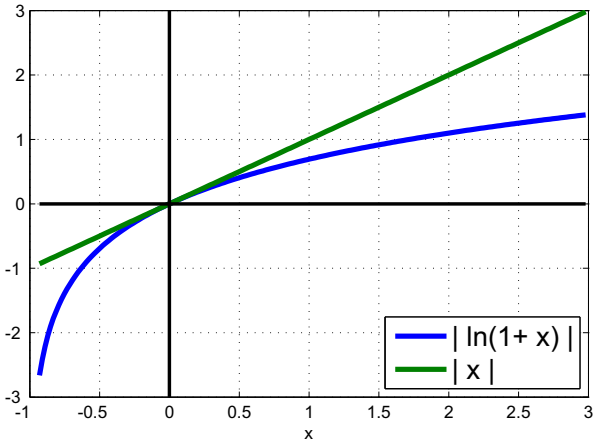


Figure 1.4: If $x > -1$ then $\ln(1+x) < x$.

infinite series (1.77), is positive since

1. $x \neq 0$ implies $x^{2n} > 0$

2. $-1 < x < 0$ implies:

$$1 + \frac{x}{2n+1} > 0$$

Finally, if $x < -1$, then $(1+x) < 0 < e^x$. □

Corollary 1.12.2.2. *If $x > -1$. Then*

$$\ln(1+x) \leq x.$$

The equality holds only if $x = 0$. Moreover, $\ln(1+x)$ has the same sign as x, so when x is positive $0 < \ln(1+x) < x$; but when $x > -1$ is negative $|x| < |\ln(1+x)|$. If $|x| < 1$ and $n > 0$ then $\ln(1+x^n) \leq x^n$ with equality holding only if $x = 0$.

See Figure 1.4 (page 114).

Proof. By Lemma 1.12.2.1 (page 113),

$$1 + x \leq e^x$$

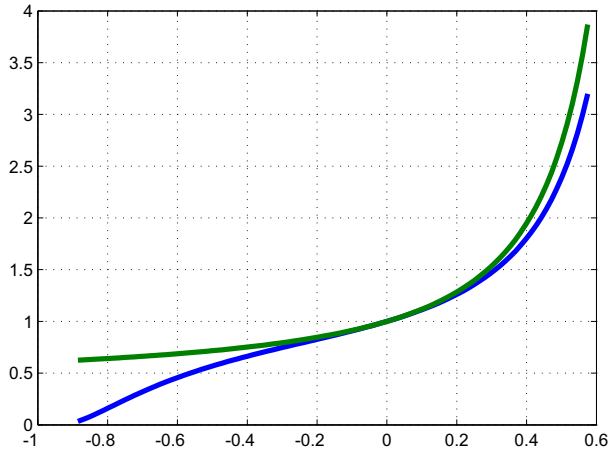


Figure 1.5: The green line is the graph of $e^{\frac{x}{1-x}}$. Beneath the green line is the graph of the product $\prod_{n=1}^{100}(1+x^n)$ rendered in blue.

with equality holding only if $x = 0$. Since ln is monotonically increasing

$$\ln(1+x) \leq \ln\left(e^x\right) = x$$

with equality holding only if $x = 0$. This corollary's requirement that $x > -1$ is due to the domain of ln being the positive reals, this forces the domain of $\ln(x+1)$ to be $x > -1$. The other assertions follow trivially from $\ln(x) > 0 \Leftrightarrow x > 1$; $\ln(x) < 0 \Leftrightarrow 0 < x < 1$; $n > 0, |x| < 1 \Rightarrow |x^n| < 1$. □

Lemma 1.12.2.3. *Let $|x| < 1$. Then the infinite product*

$$\prod_{n=1}^{\infty}(1+x^n) = (1+x)\left(1+x^2\right)\left(1+x^3\right)\ldots$$

converges, moreover

$$0 < \prod_{n=1}^{\infty}(1+x^n) \leq e^{\frac{x}{1-x}},$$

with equality holding only if $x = 0$.

See Figure 1.5 (page 115).

Proof. The result follows relatively easily once we prove that

$$\sum_{n=1}^{\infty} \ln\left(1+x^n\right)$$

is convergent for $|x|<1$. There are three cases, $x\in(-1,0)$, $x=0$, and $x\in(0,1)$.

Case 1: $x\in(-1,0)$.

$$\begin{aligned} x\in(-1,0) \text{ and } n \text{ odd} &\Rightarrow x^n\in(-1,0) \\ &\Rightarrow 1+x^n\in(0,1) \\ &\Rightarrow \ln(1+x^n)\in(-\infty,0) \qquad (1.78) \\ x\in(-1,0) \text{ and } n \text{ even} &\Rightarrow x^n\in(0,1) \\ &\Rightarrow 1+x^n\in(1,2) \\ &\Rightarrow \ln(1+x^n)\in(0,\ln(2)) \qquad (1.79) \end{aligned}$$

So the infinite series $\sum_{n=1}^{\infty}\ln\left(1+x^n\right)$ is alternating. The Alternating Series test guarantees convergence if

1. the terms being summed alternate in sign; i.e. if the terms are labeled a_n, then $a_n/a_{n+1}<0$,
2. $|a_n|\to 0$ monotonically if $n>N$ for some integer $N>0$.

Condition *1*, that the series alternates, has already been verified, see (1.78), (1.79). Condition *2*, is the subject of the following claim.

Claim: If $x\in(-1,0)$ there exists an integer $N_x>0$, dependent on x, such that, with the possible exception of the first N_x terms,

$$\left|\ln\left(1+x^n\right)\right|\to 0 \text{ monotonically.}$$

Proof of Claim: First, suppose that $n > 0$ is even, so that $n + 1$ is odd. Then (1.78), (1.79) imply:

$$\ln(1 + x^n) > 0 \text{ and } 0 < \ln(1 + x^{n+1}).$$

Hence:

$$
\begin{aligned}
|\underbrace{\ln(1+x^n)}_{>0}| > |\underbrace{\ln(1+x^{n+1})}_{<0}| &\Leftrightarrow \\
\ln(1+x^n) > -\ln(1+x^{n+1}) &\Leftrightarrow \\
1+x^n > \frac{1}{1+x^{n+1}} &\Leftrightarrow \\
(1+x^n)(1+x^{n+1}) > 1 &\Leftrightarrow \\
1+x^{n+1}+x^n+x^nx^{n+1} > 1 &\Leftrightarrow \\
x^{n+1}+x^n+x^nx^{n+1} > 0 &\Leftrightarrow \\
\underbrace{x^n}_{>0}(x+1+x^{n+1}) > 0 &\Leftrightarrow \\
x+1+x^{n+1} > 0 &\Leftrightarrow \\
x+x^{n+1} > -1 &\Leftrightarrow \\
x(1+x^n) > -1 &\Leftrightarrow \\
\underbrace{(-x)}_{0<(-x)<1}\underbrace{(1+x^n)}_{1<(1+x^n)<2} < 1 &\Leftrightarrow \\
1+x^n < \frac{1}{(-x)} &\Leftrightarrow \\
x^n < \frac{1}{(-x)} - 1 &\Leftrightarrow \\
x^n < \frac{1}{(-x)} - \frac{(-x)}{(-x)} &\Leftrightarrow \\
x^n < \frac{1+x}{(-x)} &\Leftrightarrow \\
\ln\left(|x|^n\right) = \ln\left(x^n\right) < \ln\left(\frac{1+x}{(-x)}\right) &\Leftrightarrow \\
n\ln\left(|x|\right) < \ln\left(\frac{1+x}{(-x)}\right) &\Leftrightarrow \\
n > \frac{\ln\left(\frac{1+x}{(-x)}\right)}{\ln\left(|x|\right)} &\Leftrightarrow \\
n > \frac{\ln(1+x)-\ln(-x)}{\ln\left(|x|\right)} &\Leftrightarrow \\
n > \frac{\ln(1+x)-\ln(|x|)}{\ln\left(|x|\right)} = N_x.
\end{aligned}
$$

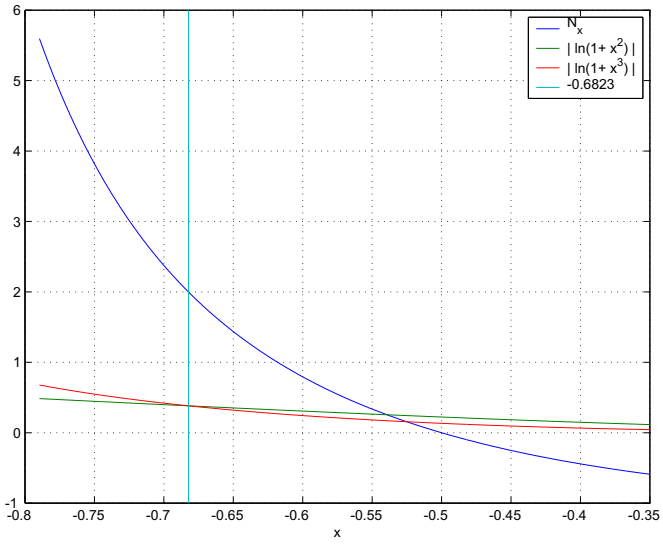


Figure 1.6: The case of n even. Let $x \in (-1, 0)$. If the positive even integer n is bigger than N_x, represented by the dark blue line, then $|\ln(1 + x^n)| > |\ln(1 + x^{n+1})|$. The red and green lines illustrate the case $n = 2$. The red line is the graph of $|\ln(1 + x^2)|$ and the green line is the graph of $|\ln(1+x^3)|$. These two lines cross each other at $x = -0.6823$, represented by the vertical light blue line, when $|\ln(1 + x^2)| = |\ln(1 + x^3)|$. Of course $N_{-0.6823} = 2$. Notice that N_x is a decreasing function of x.

So we have shown that if n is even and $n > N_x$ then

$$|\ln(1 + x^n)| > |\ln(1 + x^{n+1})|. \tag{1.80}$$

See Figure 1.6 (page 119) for illustrative example of how $|\ln(1 + x^n)|$ compares with $|\ln(1 + x^{n+1})|$ for $n = 2$ and $x \in (-0.80, -0.35)$.

Now, suppose that $n > 0$ is odd, so that $n + 1$ is even. Then (1.78), (1.79) imply:

$$\ln(1 + x^n) < 0 \text{ and } \ln(1 + x^{n+1}) > 0.$$

We proceed in the same manner as before.

$$
\begin{aligned}
|\underbrace{\ln(1+x^n)}_{<0}| &> |\underbrace{\ln(1+x^{n+1})}_{>0}| \Leftrightarrow \\
-\ln(1+x^n) &> \ln(1+x^{n+1}) \Leftrightarrow \\
\frac{1}{1+x^n} &> 1+x^{n+1} \Leftrightarrow \\
1 &> (1+x^n)(1+x^{n+1}) \Leftrightarrow \\
1 &> 1+x^{n+1}+x^n+x^n x^{n+1} \Leftrightarrow \\
0 &> x^{n+1}+x^n+x^n x^{n+1} \Leftrightarrow \\
0 &> \underbrace{x^n}_{>0}(x+1+x^{n+1}) \Leftrightarrow \\
0 &> x+1+x^{n+1} \Leftrightarrow \\
-1 &> x+x^{n+1} \Leftrightarrow \\
-1 &> x(1+x^n) \Leftrightarrow \\
1 &< \underbrace{(-x)}_{0<(-x)<1}\underbrace{(1+x^n)}_{0<(1+x^n)<1}
\end{aligned}
$$

So we have shown that for all odd $n > 0$ that

$$|\ln(1+x^n)| > |\ln(1+x^{n+1})|. \tag{1.81}$$

Combining (1.80),(1.81) we get: $n > N_x$ implies

$$|\ln(1+x^n)| > |\ln(1+x^{n+1})|.$$

So we've proven the claim that $x \in (-1,0)$ implies $|\ln(1+x^n)| \to 0$ monotonically, if $n > 0$ satisfies $n > N_x$.

Note. It is illustrative to see how the sequence $|\ln(1+x^n)|$ behaves for a fixed value of $x \in (-1,0)$. In Figure 1.7 (page 121) the sequence $|\ln(1+(-.9086974)^n)|$

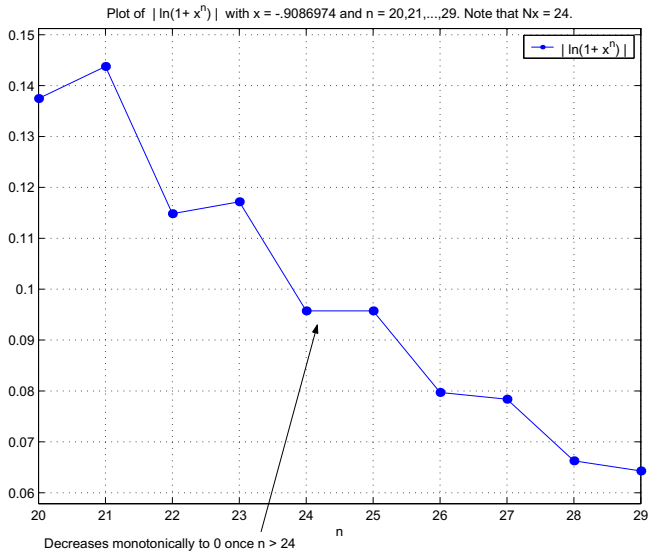


Figure 1.7: The sequence $|\ln(1+x^n)|$ when $x = -.9086974$. Once $n > N_x = 24$ the sequence $|\ln(1+x^n)| \to 0$ monotonically. The solid blue dots, $\bullet$, represent the points $(n, |\ln(1+x^n)|)$ with $n = 20, 21, \ldots 29$. The blue line segments connecting those points are drawn to help illustrate that once $n > N_x = 24$ the sequence $|\ln(1+x^n)| \to 0$ monotonically, but for $n < 24$ monotonicity is lacking. Note the difference in behavior as n goes from even to odd, as compared to when n goes from odd to even.

is plotted for $n = 20, 21, \ldots, 29$. The plot shows how the behavior of that sequence changes at $N_x = 24$.

Thus we've proved that the two conditions of the Alternating Series Test are satisfied when $x \in (-1, 0)$. Hence, when $x \in (-1, 0)$ the series $\sum_{n=1}^{\infty} \ln(1+x^n)$ converges.

If $x = 0$ the infinite series $\sum_{n=1}^{\infty} \ln(1+x^n)$ converges to 0 since each term in the series is zero; i.e. $\ln(1+x^n) = \ln(1) = 0$.

If $x \in (0, 1)$ then $\ln(1+x^n) > 0$ and the partial sums $\sum_{n=1}^{N} \ln(1+x^n)$ form an increasing sequence. So, to show convergence, it suffices to show that the partial sums have an upper bound. Corollary 1.12.2.2 (page 114), together with $x \in (0, 1)$ imply:

$$0 \leq \sum_{n=1}^{N} \ln(1+x^n) < \sum_{n=1}^{N} x^n < \sum_{n=1}^{\infty} x^n = \frac{x}{1-x}. \tag{1.82}$$

So we've shown $\sum_{n=1}^{\infty} \ln(1+x^n)$ converges for $|x| < 1$.

Next we show, if $|x| < 1$ and $x \neq 0$ then

$$\sum_{n=1}^{\infty} \ln(1+x^n) < \frac{x}{1-x}.$$

We can see this as follows. By Corollary 1.12.2.2 (page 114), we have $\ln(1+x^n) < x^n$. This implies:

$$0 < x - \ln(1+x) \text{ and } 0 \leq \sum_{n=2}^{\infty} x^n - \sum_{n=2}^{\infty} \ln(1+x^n).$$

Consequently,

$$\begin{aligned}
\frac{x}{1-x} - \sum_{n=1}^{\infty} \ln(1+x^n) &= \sum_{n=1}^{\infty} x^n - \sum_{n=1}^{\infty} \ln(1+x^n) \\
&= \left(x + \sum_{n=2}^{\infty} x^n \right) - \left(\ln(1+x^n) + \sum_{n=2}^{\infty} \ln(1+x^n) \right) \\
&= \left(\underbrace{x - \ln(1+x)}_{0>0} \right) + \left(\underbrace{\sum_{n=2}^{\infty} x^n - \sum_{n=2}^{\infty} \ln(1+x^n)}_{\geq 0} \right) \\
&> 0 \text{ (if } |x| < 1 \text{ and } x \neq 0).
\end{aligned}$$

If $x \neq 0$ and $|x| < 1$ we have

$$\frac{x}{1-x} > \sum_{n=1}^{\infty} \ln(1+x^n).$$

Then, since the function e^x is monotonically increasing and continuous, we have:

$$\begin{aligned} e^{\frac{x}{1-x}} &> e^{\sum_{n=1}^{\infty} \ln(1+x^n)} \\ &= e^{\lim_{N\to\infty} \sum_{n=1}^{N} \ln(1+x^n)} \\ &= \lim_{N\to\infty} e^{\sum_{n=1}^{N} \ln(1+x^n)} \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} e^{\ln(1+x^n)} \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} (1+x^n) \\ &= \prod_{n=1}^{\infty} (1+x^n). \end{aligned}$$

If $x = 0$, we trivially have $e^{\frac{x}{1-x}} = \prod_{n=1}^{\infty} (1 + x^n) = 1$. □

Corollary 1.12.2.4.

$$\prod_{n=1}^{\infty} \left(1 + \frac{1}{2^n}\right) < e$$

Proof. When $x = 1/2$

$$\frac{x}{1-x} = 1,$$

the result then follows from Lemma 1.12.2.3 (page 115). □

1.12.3 Proof of Completeness Theorem for d_H

We now are ready to address Nussbaum's version of Zabreiko, Krasnosel'skii, and Pokornyi's [92, Lemma 1.] result [9] on completeness w.r.t. d_H which is introduced in Section 1.12.1 (page 111):

[9] [92, Zabreiko, Krasnosel'skii, and Pokornyi, Lemma 1] *The metric space $\hat{R}$ is complete if and only if the component R is normal.* Normal means for every $x \in R \subset X$, X a Banach Space, that the conical segment $< 0, x >= \{z : 0 \leq z \leq x\}$ is bounded w.r.t. X's norm.

Theorem 1.12.3.1. *Let C be pointed, closed, convex, salient cone in a Banach space X. Suppose that $u \in C \setminus \{0\}$ and let C_u denote the component of C containing u; i.e.*

$$C_u = \{x \in C : d_H(x, u) < \infty\}.$$

Let $\widehat{C}_u = (C_u, \sim)$. Then the following statements are equivalent:

1. *$(\widehat{C}_u, d_H) = ((C_u, \sim), d_H)$ is a complete metric space.*
2. *$\sup\{\|x\| : 0 \leq x \leq u\} < \infty$.*
3. *E_u is a complete normed linear space with respect to $|x|_u$.*

See Definition 1.11.2.1 (page 102) of E_u and Definition 1.11.3.6 (page 108) of $|x|_u$.

For us, the important direction is $2 \Rightarrow 1$:

$$\sup\{\|x\| : 0 \leq x \leq u\} < \infty \Rightarrow \widehat{C}_u \text{ complete w.r.t. } d_H;$$

and we provide a detailed proof of that direction. Our proof is based on the original proof appearing in [92, Zabreiko, Krasnosel'skii, and Pokornyi, Lemma 1]. The difference is we provide details. See [10].

Proof. We prove $2 \Rightarrow 1$. The proof is lengthy and will be broken into lemmas and propositions.

Proposition 1.12.3.2. $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ *implies E_u is a complete normed linear space with respect to $|x|_u$.*

Remark. Proposition 1.12.3.2 and a brief version of the following proof appear in Krasnosel'skii [62, Theorem 1.3].

[10] Unfortunately, I am unable to follow a part of the proof in [92, Zabreiko, Krasnosel'skii, and Pokornyi, Lemma 1] in in the direction $1 \Rightarrow 2$: i.e. $\widehat{C}_u$ complete w.r.t. $d_H \Rightarrow \sup\{\|x\| : 0 \leq x \leq u\} < \infty$. See [92, Lemma 1.].

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be cauchy in E_u w.r.t. $|\ |_u$.

Then $\exists\ \alpha_n > 0,\ \alpha_n \to 0$, such that $m > 0$ implies

$$-\alpha_n u \leq x_{n+m} - x_n \leq \alpha_n.$$

Hence

$$x_{n+m} - x_n \in < -\alpha_n u, \alpha_n u >$$

and so by Lemma 1.11.1.3 (page 101), (1.62) and (1.64)

$$\|x_{n+m} - x_n\| \leq \alpha_n \sup\{\|x\| : 0 \leq x \leq u\}.$$

Since $\sup\{\|x\| : 0 \leq x \leq u\}$ is fixed, $m > 0$ is arbitrary, and $\alpha_n \to 0$, it follows that $\{x_n\}_{n=1}^{\infty}$ is cauchy in X. X is complete so x_n converges w.r.t. $\|\ \|$ to some $x^* \in X$. Since

$$-\alpha_n u + C \text{ and } \alpha_n u - C$$

are closed in X, and since

$$< -\alpha_n u, \alpha_n u >= (-\alpha_n u + C) \bigcap (\alpha_n u - C)$$

it follows that $< -\alpha_n u, \alpha_n u >$ is closed in X. Hence $x^* \in < -\alpha_n u, \alpha_n u > \subset E_u$. So E_u is complete. □

Proposition 1.12.3.3. *If* $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ *and* $x_n \to x$ *in* E_u *w.r.t. the u-norm, then* $x_n \to x$ *in* E_u *w.r.t.* X*'s usual norm,* $\|\ \|$.

Proof. Since $x_n \to x$ in E_u w.r.t. the u-norm there exist $\alpha_n > 0,\ \alpha_n \to 0$ such that

$$-\alpha_n u \leq x - x_n \leq \alpha_n u.$$

But then

$$\|x - x_n\| \leq \alpha_n \sup\{\|x\| : 0 \leq x \leq u\}.$$

So $x_n \to x$ w.r.t. X's usual norm. □

The following proposition, Proposition 1.12.3.4, finishes the proof of Theorem 1.12.3.1 in the $2 \Rightarrow 1$ direction. It also proves, with the help of Proposition 1.12.1.2, that C_u normal implies $(C_u, \sim)$ is complete w.r.t. d_H.

Proposition 1.12.3.4. $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ *implies* $\widehat{C}_u$ *is complete with respect to* d_H.

Notation: $\widehat{C}_u = (C_u, \sim)$ *and* $\widehat{x}_n = [x_n]$.

Proof. Preliminaries. Let $\{\widehat{x}_n\}_{n=1}^{\infty}$ be cauchy in $\widehat{C}_u$.

Since $\{\widehat{x}_n\}_{n=1}^{\infty}$ is cauchy, to prove that $\{\widehat{x}_n\}_{n=1}^{\infty}$ converges, it suffices to show that that a subsequence of $\{\widehat{x}_n\}_{n=1}^{\infty}$ converges. Moreover, since $\{\widehat{x}_n\}_{n=1}^{\infty}$ is cauchy, no matter how fast $\varepsilon_k \to 0$, provided each $\varepsilon_k > 0$, we can always find a subsequence $\{\widehat{x}_{n_k}\}_{n_k=1}^{\infty}$ satisfying:

$$d_H(\widehat{x}_{n_k}, \widehat{x}_{n_{k+1}}) < \varepsilon_k.$$

It turns out, that if $\varepsilon_k = \ln\left(1 + (1/2)^k\right)$ then it is relatively easy to show convergence.

However, to avoid unnecessarily complicated subscripting we will not work with the subsequence $\{\widehat{x}_{n_k}\}_{n_k=1}^{\infty}$. Rather, we will prove the following:

If $\{\widehat{x}_n\}_{n=1}^{\infty}$ is a cauchy sequence in $\widehat{C}_u$ which satisfies

$$d_H(\widehat{x}_n, \widehat{x}_{n+1}) < \varepsilon_n = \ln\left(1 + \frac{1}{2^n}\right).$$

Then $\{\widehat{x}_n\}_{n=1}^{\infty}$ converges.

Our strategy will be to positively scale each of the x_n in such a way so that the resulting sequence $\{z_n\}_{n=1}^{\infty}$ is cauchy in E_u w.r.t. the u-norm. The assumption that $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ implies that E_u is complete w.r.t. the u-norm. So the

aforementioned cauchy sequence $\{z_n\}_{n=1}^{\infty}$ will converge to some $z^* \in E_u$. Finally, it will be shown that $\widehat{z}^* \in \widehat{C}_u$ and that $\widehat{x}_n \to \widehat{z^*}$.

The Proof proper. Suppose that $\{\widehat{x}_n\}_{n=1}^{\infty}$ is a cauchy sequence in $\widehat{C}_u$ and that $d_H(\widehat{x}_n, \widehat{x}_{n+1}) < \varepsilon_n$.

Since $x_1 \in C_u$ there exists $\alpha_0, \beta_0 > 0$ such that

$$\alpha_0 u \leq x_1 \leq \beta_0 u \tag{1.83}$$

So

$$u \leq \frac{1}{\alpha_0}\, x_1 \tag{1.84}$$

For $n = 1, 2, 3, \ldots$ we have $d_H(\widehat{x}_n, \widehat{x}_{n+1}) < \varepsilon_n$. So, for $n = 1, 2, 3, \ldots$ $\exists\ \alpha_n, \beta_n > 0$ with

$$\alpha_n x_n \leq x_{n+1} \leq \beta_n x_n \text{ and } \ln(\beta_n/\alpha_n) < \varepsilon_n \tag{1.85}$$

The condition $\ln(\beta_n/\alpha_n) < \varepsilon_n$ implies that $\frac{\beta_n}{\alpha_n} < e^{\varepsilon_n}$. This in turn implies that

$$\frac{\beta_n}{\alpha_n} x_n \leq e^{\varepsilon_n} x_n. \tag{1.86}$$

This suggests that we divide the first set of inequalities in (1.85) by α_n. Doing so we get,

$$x_n \leq \frac{1}{\alpha_n}\, x_{n+1} \leq \frac{\beta_n}{\alpha_n} x_n,$$

which implies, due to the inequality (1.86), that

$$x_n \leq \frac{1}{\alpha_n}\, x_{n+1} \leq e^{\varepsilon_n} x_n. \tag{1.87}$$

Dividing (1.87) by $\frac{1}{\alpha_0 \alpha_1 \ldots \alpha_{n-1}}$ we get

$$\frac{1}{\alpha_0 \alpha_1 \ldots \alpha_{n-1}}\, x_n \leq \frac{1}{\alpha_0 \alpha_1 \ldots \alpha_{n-1}}\, \frac{1}{\alpha_n}\, x_{n+1} \leq \frac{1}{\alpha_0 \alpha_1 \ldots \alpha_{n-1}}\, e^{\varepsilon_n} x_n. \tag{1.88}$$

Keeping in mind (1.88), we define for $n = 1, 2, 3, \ldots$

$$z_n = \frac{1}{\alpha_0 \alpha_1 \ldots \alpha_{n-1}} \, x_n$$

By (1.84) and the definition of z_n we have:

$$0 \leq u \leq \frac{1}{\alpha_0} \, x_1 = z_1 \tag{1.89}$$

From (1.88) it is clear that for $n = 1, 2, \ldots$ that we have:

$$z_n \leq z_{n+1} \leq e^{\varepsilon_n} z_n. \tag{1.90}$$

Combining (1.89) with iterates of "$z_n \leq z_{n+1}$" implies that for all $n \geq 1$ and for all $m \geq 0$,

$$0 \leq u \leq z_n \leq z_{n+m} \tag{1.91}$$

If $n \geq 3$, (1.90) implies:

$$z_n \leq e^{\varepsilon_{n-1}} z_{n-1} \text{ and } z_{n-1} \leq e^{\varepsilon_{n-2}} z_{n-2}. \tag{1.92}$$

Combining the two inequalities from (1.92) we get

$$z_n \leq e^{\varepsilon_{n-1}} e^{\varepsilon_{n-2}} z_{n-2}. \tag{1.93}$$

Continuing in this fashion it is clear from (1.92) and (1.93) that

$$z_n \leq e^{\varepsilon_{n-1}} e^{\varepsilon_{n-2}} \ldots e^{\varepsilon_1} z_1, \tag{1.94}$$

provided $n \geq 3$. It is useful to define $\varepsilon_0 = 0$ so that $e^{\varepsilon_0} = 1$ and then to write (1.94)

as

$$z_n \leq e^{\varepsilon_{n-1}} e^{\varepsilon_{n-2}} \dots e^{\varepsilon_1} e^{\varepsilon_0} z_1, \tag{1.95}$$

If $n = 2$, (1.95) becomes:

$$z_2 \leq e^{\varepsilon_1} e^{\varepsilon_0} z_1 = e^{\varepsilon_1} z_1$$

which is true by (1.90). If $n = 1$, (1.95) becomes:

$$z_1 \leq e^{\varepsilon_0} z_1 = z_1$$

which is trivially true. So (1.95) is true for all $n \geq 1$.

Subtracting z_n from the terms in (1.90) yields

$$0 \leq z_{n+1} - z_n \leq (e^{\varepsilon_n} - 1) z_n. \tag{1.96}$$

Let $m \geq 1$, $n \geq 1$, change n to k in (1.96), and then sum the inequality in (1.96) from n to $n + m - 1$. This results in:

$$0 \leq \underbrace{\sum_{k=n}^{n+m-1} (z_{k+1} - z_k)}_{(z_{m+n}) - (z_n)} \leq \sum_{k=n}^{n+m-1} (e^{\varepsilon_k} - 1) z_k \tag{1.97}$$

Note. The first sum in (1.97) is telescoping.

By (1.95)

$$z_k \leq \left(\prod_{j=0}^{k-1} e^{\varepsilon_j} \right) z_1 \tag{1.98}$$

for all $k \geq 1$. Substituting (1.98) into (1.97) yields

$$0 \leq z_{n+m} - z_n \leq \sum_{k=n}^{n+m-1} \left((e^{\varepsilon_k} - 1) \left(\prod_{j=0}^{k-1} e^{\varepsilon_j} \right) z_1 \right), \tag{1.99}$$

provided $n, m \geq 1$.

Let $\delta_n = e^{\varepsilon_n} - 1$ for $n = 0, 1, 2, \ldots$, then

$$1 + \delta_n = e^{\varepsilon_n} \text{ and } \ln(1 + \delta_n) = \varepsilon_n. \tag{1.100}$$

We rewrite (1.99) in terms of δ_n: for $n, m \geq 1$ we have,

$$\begin{aligned} 0 \leq z_{n+m} - z_n &\leq \sum_{k=n}^{n+m-1} \left(\delta_k \left(\prod_{j=0}^{k-1} (1 + \delta_j) \right) z_1 \right) \\ &= \left(\sum_{k=n}^{n+m-1} \delta_k \left(\prod_{j=0}^{k-1} (1 + \delta_j) \right) \right) z_1. \end{aligned} \tag{1.101}$$

Furthermore, since each $\delta_k \geq 0$ and since each $1 + \delta_j \geq 1$, it follows from (1.101) that for $n, m \geq 1$:

$$0 \leq z_{n+m} - z_n \leq \left(\sum_{k=n}^{\infty} \left(\delta_k \prod_{j=0}^{\infty} (1 + \delta_j) \right) \right) z_1. \tag{1.102}$$

Remark. It is worthwhile to forget, for just a moment, that we have already defined ε_n. If the condition on ε_n was simply that $\varepsilon_n \to 0$, this would guarantee that $\delta_n = e^{\varepsilon_n} - 1 \to 0$, but it wouldn't insure the convergence of the infinite series and product appearing in (1.102). However, if $\delta_n \to 0$ sufficiently fast, for example if $\delta_n = (1/2)^n$, then the infinite series and product in (1.102) converges, as shown below.

Recall, for $n = 1, 2, 3, \ldots$ that $\varepsilon_n = \ln\left(1 + \frac{1}{2^n}\right) = \ln(1 + \delta_n)$, so $\delta_n = (1/2)^n$. When $n = 0$, we've defined $\varepsilon_0 = 1$ so $\delta_0 = 0$.

By Corollary 1.12.2.4,

$$\prod_{j=0}^{\infty} (1 + \delta_j) = (1 + 0) \prod_{j=1}^{\infty} \left(1 + \frac{1}{2^j} \right) < e, \tag{1.103}$$

and by elementary algebra, for $n \geq 1$, we have,

$$\sum_{k=n}^{\infty} \delta_k = \sum_{k=n}^{\infty} (1/2)^k = \frac{(1/2)^n}{1-(1/2)} = (1/2)^{n-1}. \tag{1.104}$$

So the infinite series and product appearing in (1.102) converge. Pluging (1.103) and (1.104) into (1.102) yields:

$$0 \leq z_{n+m} - z_n \leq (1/2)^{n-1}\ e\ z_1, \tag{1.105}$$

which holds for $n \geq 1$ and for all $m \geq 0$. When $m = 0$, (1.105) is trivially true.

By (1.105), z_n is cauchy in E_u w.r.t. the u-norm. We are assuming

$$\sup\{\|x\| : 0 \leq x \leq u\} < \infty.$$

By Proposition 1.12.3.2, $\sup\{\|x\| : 0 \leq x \leq u\} < \infty$ guarantees that E_u is complete w.r.t. the u-norm. Hence, $z_n \to z^*$ for some $z^* \in E_u$, w.r.t. the u-norm.

Since $z_n, u \in E_u$ we have $z_n - u \in E_u$. Since $u \leq z_n$ we have $z_n - u \in C$. We've shown that $z_n \to z^*$ w.r.t. the u-norm. This implies that $z_n - u \to z^* - u$ w.r.t the u-norm. Then, by Proposition 1.12.3.3, we have $z_n - u \to z^* - u$ w.r.t X's usual norm, $\|\ \|$. But C is closed w.r.t $\|\ \|$, so $z^* - u \in C$. Hence $u \leq z^*$. Moreover, since $0 \leq u$, it follows that $0 \leq z^*$; i.e. $z^* \in C$.

Since $z_n \to z^*$ w.r.t. to the u-norm, $\exists \alpha'_n > 0$ such that $\alpha'_n \to 0$ and

$$-\alpha'_n u \leq z_n - z^* \leq \alpha'_n u.$$

This implies:

$$z^* - \alpha'_n u \leq z_n \leq z^* + \alpha'_n u.$$

But $u \leq z^*$, so

$$z^* - \alpha'_n z^* \leq z^* - \alpha'_n u \leq z_n \leq z^* + \alpha'_n u \leq z^* + \alpha'_n z^*.$$

So,

$$(1 - \alpha'_n)z^* \leq z_n \leq (1 + \alpha'_n)z^*.$$

Hence

$$d_H(\widehat{z}_n, \widehat{z}^*) \leq \ln\left(\frac{1 + \alpha'_n}{1 - \alpha'_n}\right) < \infty$$

which goes to zero. Hence, $\widehat{z}_n \to \widehat{z}^*$ w.r.t d_H.

$\widehat{x}_n \in \widehat{C}_u$, so

$$d_H(\widehat{x}_n, \widehat{u}) < \infty.$$

z_n is a positive multiple of x_n, so $\widehat{z}_n = \widehat{x}_n$. So

$$d_H(\widehat{z}_n, \widehat{u}) = d_H(\widehat{x}_n, \widehat{u}) < \infty.$$

Consequently,

$$d_H(\widehat{z}^*, \widehat{u}) \leq d_H(\widehat{z}_n, \widehat{z}^*) + d_H(\widehat{z}_n, \widehat{u}) < \infty.$$

So $\widehat{z}^* \in \widehat{C}_u$.

Since $\widehat{z}_n = \widehat{x}_n$, and since $\widehat{z}_n \to \widehat{z}^*$ w.r.t d_H, it follows that $\widehat{x}_n \to \widehat{z}^*$ w.r.t d_H. □

So the proof of Theorem 1.12.3.1 in the $2 \Rightarrow 1$ direction is finished. □

1.12.4 Using normality to show completeness

Proposition 1.12.4.1. *Suppose C has the following property:*

$$\textit{if } \; x, y \in C \;\; \textit{and} \;\; x \leq y \;\; \textit{then} \;\; \|x\| \leq \|y\| \tag{1.106}$$

Then, if $u \in C \setminus \{0\}$, *the component* $C_u = \{x \in C : d_H(x,u) < \infty\}$ *will be normal; i.e.*

$$\sup\{\|x\| : 0 \leq x \leq u\} < \infty \tag{1.107}$$

and $(C_u, \sim)$ *will be complete w.r.t.* d_H.

Proof. If x, u satisfies $0 \leq x \leq u$ (see (1.107)), then $x, u \in C$ and $x \leq u$, so property (1.106) implies $\|x\| \leq \|u\|$. But then

$$\sup\{\|x\| : 0 \leq x \leq u\} \leq \|u\| < \infty$$

and so, by Proposition 1.12.1.2 (page 111), C_u is normal. Hence, by Theorem 1.12.3.1, $(C_u, \sim)$ is complete. □

Corollary 1.12.4.2. *Let* $C = R^n_{\geq 0}$. *Let* $u \in C \setminus \{0\}$. *Then* $(C_u, \sim)$ *is complete w.r.t.* d_H.

Proof. See [11]. Suppose $x, y \in C$ and $x \leq y$.

Then $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ with $x_i, y_i \geq 0; \quad i = 1, \ldots, n$.

If $x \leq y$ then $y - x \in C$. So $y_i - x_i \geq 0$ which implies $y_i \geq x_i$. But then $\|y\| \geq \|x\|$.

Corollary 1.12.4.1 (page 132), implies $(C_u, \sim)$ is complete w.r.t. d_H. □

Corollary 1.12.4.3. *Let* L *be the Banach Space of continuous functions (on some compact set) with the norm* $\|f\| = \sup |f(x)|$. *Let* $L^+ \subset L$ *be the cone of non-negative functions. Let* $u \in L^+ \setminus \{0\}$. *Then* $(C_u, \sim)$ *is complete w.r.t.* d_H. *See* [12].

Proof. It is easy to see that L^+ is a convex, salient, pointed by the origin cone. Clearly $f \leq g$, meaning $g - f \in L^+$, implies $\|f\| \leq \|g\|$. Corollary 1.12.4.1 (page 132) implies $(C_u, \sim)$ is complete w.r.t. d_H. □

[11] For an alternative proof, based upon the compacteness of closed balls in $\mathbb{R}^n$, see Proposition 1.12.4.4 (page 134) and its corollary, Corollary 1.12.4.5 (page 135).

[12] Birkhoff uses the set of identically positive functions contained in L^+ as an example of a connected component w.r.t. d_H. See Example 2 [12]. Note $\|f\| = \|f\|_\infty$.

Proposition 1.12.4.4. *Suppose that for each $r > 0$ the closed ball of radius r, B_r, is compact. Then every component C_u is normal and thus complete.*

Proof. Suppose that $u \in C \setminus \{0\}$ and the component C_u is not normal.

Then there exists a sequence $\{c_n\}_{n=1}^{\infty}$ with $0 \leq c_n \leq u$ such that $\|c_n\| \to \infty$ as $n \to \infty$. We can assume that each $\|c_n\| \geq 1$.

$0 \leq c_n \leq u$ implies $c_n,\ u - c_n \in C$. C is closed under positive scaling so

$$\frac{c_n}{\|c_n\|},\quad \frac{u - c_n}{\|c_n\|} \in C. \tag{1.108}$$

The norm applied to (1.108) yields:

$$\left\| \frac{c_n}{\|c_n\|} \right\| = 1 \tag{1.109}$$

and

$$\begin{aligned} \left\| \frac{u - c_n}{\|c_n\|} \right\| &= \left\| \frac{u}{\|c_n\|} - \frac{c_n}{\|c_n\|} \right\| \\ &= \left\| \frac{u}{\|c_n\|} + -\frac{c_n}{\|c_n\|} \right\| \\ &\leq \left\| \frac{u}{\|c_n\|} \right\| + \left\| -\frac{c_n}{\|c_n\|} \right\| \\ &= \frac{\|u\|}{\|c_n\|} + \frac{\|c_n\|}{\|c_n\|} \\ &\leq \|u\| + 1. \end{aligned} \tag{1.110}$$

Let $\|u\| + 1 = r$. (1.108) combined with (1.110) implies

$$\frac{u}{\|c_n\|} - \frac{c_n}{\|c_n\|} \in B_r \cap C.$$

B_r is compact and C is closed so $B_r \cap C$ is compact. The compactness of $B_r \cap C$

implies the infinite sequence

$$\left\{ \frac{u}{\|c_n\|} - \frac{c_n}{\|c_n\|} \right\}_{n=1}^{\infty}$$

has a convergent subsequence

$$\left\{ \frac{u}{\|c_{n_k}\|} - \frac{c_{n_k}}{\|c_{n_k}\|} \right\}_{k=1}^{\infty} \tag{1.111}$$

whose limit exists and is in $C \cap B_r$. $\|c_{n_k}\| \to \infty$ as $k \to \infty$ and $\|u\|$ is fixed so

$$\lim_{k\to\infty} \left(\frac{u}{\|c_{n_k}\|} - \frac{c_{n_k}}{\|c_{n_k}\|} \right) \;=\; \lim_{k\to\infty} -\frac{c_{n_k}}{\|c_{n_k}\|} \;=\; -\lim_{k\to\infty} \frac{c_{n_k}}{\|c_{n_k}\|} \;\in\; C \cap B_r. \tag{1.112}$$

Let $S_1 = \{v \in V : \|v\| = 1\}$ be the unit sphere in V. B_1 is compact (by assumption) and thus closed. S_1 is B_1 minus the open unit ball. So S_1 is closed. C is closed so $S_1 \cap C$ is closed.

(1.108) combined with (1.109) implies

$$\frac{c_{n_k}}{\|c_{n_k}\|} \;\in\; S_1 \cap C. \tag{1.113}$$

$S_1 \cap C$ is closed so (1.113) combined (1.112) implies that there exists a vector $c \in S_1 \cap C$ such that

$$\lim_{k\to\infty} \frac{c_{n_k}}{\|c_{n_k}\|} = c$$

(1.112) also implies that $-c \in C \cap B_r$. So $c \in C \cap -C$. Since C is salient, $c = 0$. However $c \in S_1$ so $\|c\| = 1$ and so $c \neq 0$. So we arrive at a contradiction.

Consequently the component C_u must have been normal. So, by Theorem 1.12.3.1 (page 123), $(C_u, \sim)$ is complete. □

Corollary 1.12.4.5. *Let C be any convex, salient, closed, pointed-by-the-origin cone in $\mathbb{R}^n$. Let $u \in C \setminus \{0\}$ Then $(C_u, \sim)$ is complete w.r.t. d_H.*

Proof. In $\mathbb{R}^n$ the closed balls of radius r are compact. Apply Proposition 1.12.4.4 (page 134). □

Chapter 2

Linear Maps and the Hilbert Projective Metric d_H

2.1 The linear map P as a fractional linear transformation.

2.1.1 Introduction

Let P be a linear map of the Banach Space V to itself which also takes the closed, salient, pointed cone $C \subset V$ to itself. Following Birkhoff's notation, we will also let P represent the map of $C/\sim$ to itself which is induced by the linear map P. See [1].

We want to know under which conditions will P on $(C \setminus \{0\})/\sim$ be a contraction map w.r.t. the Hilbert Projective Metric d_H.

The following section is collection of technical results which will help us to understand the action of P on two dimensional cones.

[1] If $f \in C$ then the equivalence class of f w.r.t $\sim$ is the ray $[f] = \{\lambda f : \lambda > 0\}$. When P is acting on a vector $f \in V$ we write fP. When P is acting on an equivalence class $[f] \in C/\sim$ we write $P([f])$ and define $P([f]) = [fP]$. If $f' \in [f]$ then $f' = \lambda' f$ for some $\lambda' > 0$. But then using the linearity of P (the linear map P on V), we get $P([f']) = [(f')P] = \{\lambda((\lambda' f)P) : \lambda > 0\} = \{\lambda\lambda'(fP) : \lambda > 0\} = \{\lambda(fP) : \lambda > 0\} = [fP]$. So the induced map P is well defined.

2.1.2 Theorems about P on Span$(f,g)\bigcap C$

Theorem 2.1.2.1. *Let P be a linear map of the Banach Space V to itself which also takes the cone C to itself. We assume that C is a closed pointed convex salient cone whose point is the origin of V. Suppose $f,g \in C$ are linearly independent and that $fP, gP \in C$ are linearly independent.*

1. *There exists $b_1, b_2 \in \partial\left(Span(f,g)\bigcap C\right)$ which are linearly independent, and there exists $b_{1'}, b_{2'} \in \partial\left(Span(fP,gP)\bigcap C\right)$ which are linearly independent, such that*

$$Span(f,g)\bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha,\beta \geq 0\} \tag{2.1}$$

$$Span(fP,gP)\bigcap C = \{\alpha b_{1'} + \beta b_{2'} \mid \alpha,\beta \geq 0\}. \tag{2.2}$$

   Moreover

$$P : Span(f,g)\bigcap C \to Span(fP,gP)\bigcap C. \tag{2.3}$$

2. *With respect to the ordered basis b_1, b_2 of $Span(f,g)$ and the ordered basis $b_{1'}, b_{2'}$ of $Span(fP,gP)$, the following two equations*

$$b_1 P = db_{1'} + bb_{2'} \tag{2.4}$$

$$b_2 P = cb_{1'} + ab_{2'} \tag{2.5}$$

   uniquely determine $a,b,c,d \in \mathbb{R}$. Moreover, it is the case that $a,b,c,d \geq 0$ and $ad - bc \neq 0$.

3. *Suppose that $x \in Span(f,g)\bigcap C$. Then, w.r.t. the ordered basis b_1, b_2 of $Span(f,g)$ there exists uniquely $x_1, x_2 \geq 0$ such that*

$$x = x_1 b_1 + x_2 b_2.$$

$xP \in Span(fP, gP) \bigcap C$, and w.r.t. the ordered basis $b_{1'}, b_{2'}$ of $Span(fP, gP)$ there exists uniquely $(xP)_{1'}, (xP)_{2'} \geq 0$ such that

$$xP = (xP)_{1'} b_{1'} + (xP)_{2'} b_{2'}.$$

We can calculate $(xP)_{1'}, (xP)_{2'}$ by simple matrix multiplication

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} dx_1 + cx_2 \\ bx_1 + ax_2 \end{pmatrix} = \begin{pmatrix} (xP)_{1'} \\ (xP)_{2'} \end{pmatrix}. \tag{2.6}$$

The matrix of a, b, c, d appearing in the LHS of Equation 2.6 is uniquely determined (i.e. independent of x) once b_1, b_2 and $b_{1'}, b_{2'}$ are chosen.

4. *Suppose that $x, y \in Span(f, g) \bigcap C \setminus \{0\}$ so that (by part 1 of this Theorem)*

$$x = x_1 b_1 + x_2 b_2$$
$$y = y_1 b_1 + y_2 b_2$$

with $x_1, x_2, y_1, y_2 \geq 0$. Then $0 < d_H(x, y) < \infty$ if and only if x, y are linearly independent and $x_1, x_2, y_1, y_2 > 0$.

5. *Recall that we are assuming that fP, gP are linearly independent.*

   a) If $x, y \in Span(f, g) \bigcap C \setminus \{0\}$ and $0 < d_H(x, y) < \infty$ then xP, yP are linearly independent and

$$\begin{aligned} &0 < (xP)_{1'}, (xP)_{2'}, (yP)_{1'}, (yP)_{2'} < \infty \text{ and so} \\ &0 < \frac{x_2}{x_1}, \frac{y_2}{y_1}, \frac{(xP)_{2'}}{(xP)_{1'}}, \frac{(yP)_{2'}}{(yP)_{1'}} < \infty. \end{aligned} \tag{2.7}$$

   Yielding

$$0 < d_H(xP, yP) < \infty.$$

b) The equivalence relation $\sim$, where $x \sim y$ if $x = \lambda y$ for some $\lambda > 0$, partitions $Span(f,g) \bigcap C \setminus \{0\}$ into equivalence classes. The equivalence class of x is denoted $[x]$ so $[x] = \{\lambda x : \lambda > 0\}$. Geometrically, $[x]$ is an open ray originating at the origin.

Once we have specified a choice of b_1, b_2, we can express each $x \in Span(f,g) \bigcap C \setminus \{0\}$ uniquely in the form $x = x_1 b_1 + x_2 b_2$ and associate $[x]$ with the ratio $x_2/x_1 \in [0,\infty]$. In particular, the map

$$m : \Big(Span(f,g) \bigcap C \setminus \{0\}\Big) / \sim \mapsto [0,\infty] \qquad (2.8)$$
$$\text{by } m([x]) = x_2/x_1$$

is well defined (i.e. invariant w.r.t. the vectors in $[x]$ and determinate) and is bijective [2]*. Note, $m([b_1]) = 0/1 = 0$ and $m([b_2]) = 1/0 = \infty$. See* [3]

c) The function $\gamma(s,t) = ([b_1 + sb_2], [b_1 + tb_2])$ maps the set

$$\{(s,t) \in (0,\infty) \times (0,\infty) \quad | \quad s \neq t\}$$

bijectively to

$$\Big\{([u],[v]) \in \Big(Span(f,g) \bigcap C \setminus \{0\}\Big) / \sim \quad \Big| \quad 0 < d_H([u],[v]) < \infty\Big\}.$$

Before we prove Theorem 2.1.2.1 we have the following definition and remarks:

Definition 2.1.2.2. We call a linearly independent pair of vectors b_1, b_2 ends for a cone K if $K = \{\alpha b_1 + \beta b_2 : \alpha, \beta \geq 0\}$.

[2] Intuitively, we are mapping the ray $[x]$ to its slope x_2/x_1 relative to a coordinate system determined by b_1, b_2. See Figure 2.1 (page 150).

[3] Birkhoff calls $[0,\infty]$ with the Hilbert Metric $d_H(x,y) = |\ln(x/y)|$ for $x, y \in [0,\infty]$ the hyperbolic line. See [12]. It can be shown that not only is m a bijection, but that it is actually an isometry.

Remark 2.1.2.3. The vectors b_1, b_2, resp. $b_{1'}, b_{2'}$, are ends for the 2 dimensional cones described in the first part of this theorem. I.e. from (2.1) and (2.2) we have the two dimensional cones:

$$\begin{aligned} \text{Span}(f,g)\bigcap C &= \{\alpha b_1 + \beta b_2 \mid \alpha, \beta \geq 0\} \\ \text{Span}(fP,gP)\bigcap C &= \{\alpha b_{1'} + \beta b_{2'} \mid \alpha, \beta \geq 0\}. \end{aligned}$$

Remark 2.1.2.4. It can be shown that if b_1, b_2 are ends for a cone K then $b_1, b_2 \in \partial K$.

Proof. 1. If $f, g \in C$ are linearly independent and $fP, gP \in C$ are linearly independent, Theorem 1.4.0.2 (page 42) and Proposition 1.4.2.1 (page 46) imply that there exists $b_1, b_2 \in \partial\left(\text{Span}(f,g)\bigcap C\right)$ which are linearly independent, and that there exists

$$b_{1'}, b_{2'} \in \partial\left(\text{Span}(fP,gP)\bigcap C\right)$$

which are also linearly independent, such that

$$\begin{aligned} \text{Span}(f,g)\bigcap C &= \{\alpha b_1 + \beta b_2 \mid \alpha, \beta \geq 0\} && (2.9) \\ \text{Span}(fP,gP)\bigcap C &= \{\alpha b_{1'} + \beta b_{2'} \mid \alpha, \beta \geq 0\}. && (2.10) \end{aligned}$$

Since P is linear, P takes linear combinations of f, g to linear combinations of fP, gP. This, together with $P : C \to C$, implies (2.3).

2. $b_1, b_2 \in \text{Span}(f,g)\bigcap C$ so b_1, b_2 are linear combinations of f, g. Since P is linear, b_1P, b_2P are linear combinations of fP, gP. This, combined with $P : C \to C$, implies that $b_1P, b_2P \in \text{Span}(fP,gP)\bigcap C$. Then, by Equation (2.10) $\exists\ a, b, c, d \geq 0$ such that Equations (2.4) and (2.5) hold.

Next we show that $ad - bc \neq 0$.

Suppose not, suppose $ad = bc$. If $a \neq 0$ and we multiply Equation (2.4) by a and

Equation (2.5) by b we get

$$ab_1P = adb_{1'} + abb_{2'} \tag{2.11}$$

$$bb_2P = bcb_{1'} + bab_{2'}. \tag{2.12}$$

But then $ab_1P = bb_2P$ and so

$$b_1P = \frac{b}{a}\, b_2P$$

which implies that b_1P, b_2P are linearly dependent.

If $a = 0$ then ad - bc = 0 implies b and/or $c = 0$.

If $a = b = 0$, then Equations (2.4) and (2.5) become

$$b_1P = db_{1'} \tag{2.13}$$

$$b_2P = cb_{1'} \tag{2.14}$$

which implies that b_1P, b_2P are linearly dependent.

If $a = c = 0$, then Equations (2.4) and (2.5) become

$$b_1P = db_{1'} + bb_{2'} \tag{2.15}$$

$$b_2P = 0. \tag{2.16}$$

which implies b_1P, b_2P are linearly dependent.

So ad-bc = 0 implies that b_1P, b_2P are linearly dependent in all cases. But then fP, gP are linearly dependent, since P is linear and f and g are linear combinations of b_1, b_2. But this contradicts our assumption that fP, gP are linearly independent.

So it must have been the case that $ad - bc \neq 0$.

Finally a,b,c,d are unique since $b_{1'}, b_{2'}$ are linearly independent.

3. Part 1 of this theorem implies, for $x \in \text{Span}(f,g) \cap C$, the existence and uniqueness and non-negativity of the representations

$$x = x_1 b_1 + x_2 b_2$$
$$xP = (xP)_{1'} b_{1'} + (xP)_{2'} b_{2'}.$$

Part 2 of this theorem implies $\exists\ a, b, c, d \geq 0$ such that

$$b_1 P = d b_{1'} + b b_{2'} \tag{2.17}$$
$$b_2 P = c b_{1'} + a b_{2'}. \tag{2.18}$$

Equations (2.17) and (2.18) imply for $x \in \text{Span}(f,g) \cap C$ that

$$\begin{aligned} xP &= (x_1 b_1 + x_2 b_2) P \\ &= x_1 (d b_{1'} + b b_{2'}) + x_2 (c b_{1'} + a b_{2'}) \\ &= (x_1 d + x_2 c) b_{1'} + (x_1 b + x_2 a) b_{2'} \qquad (2.19) \\ &= (xP)_{1'}\ b_{1'} + (xP)_{2'}\ b_{2'}. \qquad (2.20) \end{aligned}$$

But then

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 d + x_2 c \\ x_1 b + x_2 a \end{pmatrix} = \begin{pmatrix} (xP)_{1'} \\ (xP)_{2'} \end{pmatrix}.$$

Hence Equation (2.6) holds.

4. We first assume that $0 < d_H(x,y) < \infty$. Definition 1.6.2.1 (page 54) of $d_H(x,y)$ depends upon whether x, y are linearly independent: if x, y are linearly dependent and non-zero then $d_H(x,y) = 0$. Thus, $0 < d_H(x,y)$ implies x, y are linearly independent.

If x, y are linearly independent the definition of $d_H(x,y)$ requires that we express x, y in terms of an ordered basis b_1^*, b_2^* for $\text{Span}(x,y)$, with $b_1^*, b_2^* \in \partial\left(\text{Span}(x,y) \bigcap C\right)$.

Such a basis exists by Theorem 1.4.0.2 (page 42) and Proposition 1.4.2.1 (page 46). We can use b_1, b_2 from (2.9) of Part 1. of this theorem for b_1^*, b_2^* since $\text{Span}(x, y) = \text{Span}(f, g)$. See [4].

Since $x, y \in \text{Span}(f, g) \bigcap C$, (2.9) gives us

$$x = x_1 b_1 + x_2 b_2 \tag{2.21}$$

$$y = y_1 b_1 + y_2 b_2 \tag{2.22}$$

with $x_1, x_2, y_1, y_2 \geq 0$. Definition 1.6.2.1 (page 54) of d_H, together with (2.21) and (2.22), and our assumption that $0 < d_H(x, y) < \infty$, yield

$$d_H(x, y) = \left| \ln \left(\frac{x_2}{x_1} \frac{y_1}{y_2} \right) \right| < \infty.$$

Now we will show that $x_1, x_2, y_1, y_2 > 0$. First, suppose $x_1 = 0$. If this is the case then $x_2 \neq 0$, since we are assuming that $x, y \neq 0$. Since x, y are linearly independent, we can't have $y_1 = 0$ if $x_1 = 0$. Since $d(x, y) < \infty$ we can't have $y_2 = 0$. But then $x_2, y_1, y_2 > 0$ and

$$\frac{x_2}{x_1} \frac{y_1}{y_2} = \frac{x_2}{0} \frac{y_1}{y_2} = \infty \Rightarrow d(x, y) = \infty,$$

which contradicts $d(x, y) < \infty$. So $x_1 \neq 0$. Similar arguments show that all of $x_1, x_2, y_1, y_2 > 0$.

Next, we assume that x, y are linearly independent and $x_1, x_2, y_1, y_2 > 0$ w.r.t. b_1, b_2. But then, as above, Definition 1.6.2.1 (page 54) of d_H implies that

$$d_H(x, y) = \left| \ln \left(\frac{x_2}{x_1} \frac{y_1}{y_2} \right) \right|. \tag{2.23}$$

[4] $x, y \in \text{Span}(f, g)$ implies $\text{Span}(x, y) \subset \text{Span}(f, g)$, but x, y are linearly independent, so $\text{Span}(x, y) = \text{Span}(f, g)$.

Direct substitution of $x_1, x_2, y_1, y_2 > 0$ into (2.23) immediately yields

$$0 \leq d(x, y) < \infty.$$

If $0 = d(x, y)$ then Equation (2.23) and $x_1, x_2, y_1, y_2 > 0$ imply

$$\begin{aligned} \frac{x_2}{x_1}\frac{y_1}{y_2} = 1 \Rightarrow \\ \frac{x_2}{x_1} = \frac{y_2}{y_1}. \end{aligned} \tag{2.24}$$

But then

$$\begin{aligned} \frac{1}{x_1}x = \frac{x_1}{x_1}b_1 + \frac{x_2}{x_1}b_2 = b_1 + \frac{x_2}{x_1}b_2 \\ \frac{1}{y_1}y = \frac{y_1}{y_1}b_1 + \frac{y_2}{y_1}b_2 = b_1 + \frac{y_2}{y_1}b_2, \end{aligned}$$

combined with (2.24), implies that

$$\frac{1}{x_1}x = \frac{1}{y_1}y,$$

which implies x, y are linearly dependent. But that contradicts our assumption that x, y are linearly independent. So $0 = d(x, y)$ is not possible with our assumptions.

5a. By Part 4 of this theorem the assumption $0 < d_H(x, y) < \infty$ implies that x, y are linearly independent and $x_1, x_2, y_1, y_2 > 0$. Since the pair x, y is linearly independent and $x, y \in \text{Span}(f, g)$ it follows that $\text{Span}(x, y) = \text{Span}(f, g)$. The linearity of P then implies $\text{Span}(xP, yP) = \text{Span}(fP, gP)$, which in turn implies the pair xP, yP is linearly independent.

By Part 3 of this theorem

$$(xP)_{1'} = dx_1 + cx_2 \tag{2.25}$$

$$(xP)_{2'} = bx_1 + ax_2. \tag{2.26}$$

Equation (2.25) can be used to prove $(xP)_{1'} > 0$ as follows. Since $xP \in \text{Span}(f, g) \bigcap C$ we have $(xP)_{1'} \geq 0$. Since $x_1, x_2 > 0$, it is immediate from Equation (2.25) that $(xP)_{1'} = 0$ only if $d = c = 0$. But $ad - bc \neq 0$ by part 2 of this theorem, so both d and c can't be 0.

Using $x_1, x_2, y_1, y_2 > 0$ and $ad - bc \neq 0$, identical arguments show that

$$(xP)_{2'}, (yP)_{1'}, (yP)_{2'} > 0.$$

Since xP, yP are linearly independent and

$$(xP)_{1'}, (xP)_{2'}, (yP)_{1'}, (yP)_{2'} > 0,$$

applying Part 4 of this theorem to $xP, yP \in \text{Span}(xP, yP) \bigcap C \setminus \{0\}$ (with $b_{1'}, b_{2'}$ playing the role of b_1, b_2) gives us

$$0 < d_H(xP, yP) = \left| \ln \left(\frac{(xP)_{2'}}{(xP)_{1'}} \frac{(yP)_{1'}}{(yP)_{2'}} \right) \right| < \infty.$$

5b. By Proposition 1.7.1.4 (page 67) the relation $\sim$ is an equivalence relation having equivalence classes of the form $[x] = \{\lambda x : \lambda > 0\}$. We show that

$$m : \Big(\text{Span}(f, g) \bigcap C \setminus \{0\} \Big) / \sim \mapsto [0, \infty] \text{ by } m([x]) = x_2/x_1 \tag{2.27}$$

is a bijection:

m is well defined since if

$$[x], [y] \in \left(\text{Span}(f, g) \bigcap C \setminus \{0\}\right) / \sim$$

then, by part 3 of this theorem

$$x = x_1 b_1 + x_2 b_2 \tag{2.28}$$

$$y = y_1 b_1 + y_2 b_2 \tag{2.29}$$

with $x_1, x_2, y_1, y_2 \geq 0$ and uniquely determined (w.r.t. b_1, b_2). Moreover, since neither x (nor y) is 0, we can not have both x_1, x_2 being 0, (nor for that matter can we have both y_1, y_2 being 0). So the ratios x_2/x_1 (and y_2/y_1) are determinate and non-negative so x_2/x_1 (and y_2/y_1) $\in [0, \infty]$. We of course take $r/0 = \infty$ whenever $r \in (0, \infty)$.

If we have $[x] = [y]$ then $x = \lambda y$ for some $\lambda > 0$. Then (2.28) and (2.29) become

$$x = \lambda y_1 b_1 + \lambda y_2 b_2 \tag{2.30}$$

$$y = y_1 b_1 + y_2 b_2 \tag{2.31}$$

and so

$$m([x]) = \frac{\lambda y_2}{\lambda y_1} = \frac{y_2}{y_1} = m([y])$$

which proves m is well defined and into $[0, \infty]$.

To prove that m is onto $[0, \infty]$ we note that

$$m([b_1]) = m([1b_1 + 0b_2]) = \frac{0}{1} = 0$$
$$m([b_2]) = m([0b_1 + 1b_2]) = \frac{1}{0} = \infty$$

and if $r \in (0, \infty)$ then

$$m\left([1b_1 + rb_2]\right) = \frac{r}{1} = r.$$

To prove that m is one to one we note that if

$$x_2/x_1 = m\left([x]\right) = m\left([y]\right) = y_2/y_1 \in [0, \infty)$$

then neither x_1 nor y_1 are zero and we can divide x by x_1 and y by y_1, which yields

$$(1/x_1)x = 1b_1 + (x_2/x_1)b_2 \tag{2.32}$$

$$(1/y_1)y = 1b_1 + (y_2/y_1)b_2. \tag{2.33}$$

But then, since $(x_2/x_1) = (y_2/y_1)$ equations (2.32) and (2.33) imply $(1/x_1)x = (1/y_1)y$ which implies $x = (x_1/y_1)y$; i.e. $[x] = [y]$. If

$$x_2/x_1 = m\left([x]\right) = m\left([y]\right) = y_2/y_1 = \infty$$

then both x_1 and y_1 equal 0 (and $x_2, y_2 > 0$). This implies that $x = x_2b_2$ and $y = y_2b_2$, which in turn implies $x = (x_2/y_2)y$; i.e. $[x] = [y]$. So m is one to one. So we've proved m is bijective as desired.

5c. To see that γ is into we note that if

$$(s, t) \in \{\, (s', t') \in (0, \infty) \times (0, \infty) \;\mid\; s' \neq t' \,\}$$

that the pair

$$b_1 + sb_2,\ b_1 + tb_2$$

is linearly independent since b_1, b_2 is linearly independent and $s \neq t$; we also note that the coefficients of $b_1 + sb_2, b_1 + tb_2$ w.r.t. b_1, b_2 are $1, s, 1, t$, which are all strictly

positive. But then by part 4 of this theorem we have $0 < d_H(b_1 + sb_2, b_1 + tb_2) < \infty$.

To show that γ is surjective we note that, by Part 4 of this theorem, if $0 < d_H([u],[v]) < \infty$ then u, v must be linear independent and, w.r.t. b_1, b_2, that the coefficients of u, v must satisfy $0 < u_1, u_2, v_1, v_2$. But then $1/u_1$ and $1/v_1$ are both > 0 and so

$$[u] = [u_1 b_1 + u_2 b_2] = [(1/u_1)(u_1 b_1 + u_2 b_2)] = [b_1 + u_2/u_1 b_2] \tag{2.34}$$

$$[v] = [v_1 b_1 + v_2 b_2] = [(1/v_1)(v_1 b_1 + v_2 b_2)] = [b_1 + v_2/v_1 b_2] \tag{2.35}$$

which implies $\gamma(u_2/u_1, v_2/v_1) = ([u],[v])$ which shows surjectivity - provided we can show that

$$(u_2/u_1, v_2/v_1) \in \{\, (s',t') \in (0,\infty) \times (0,\infty) \;\mid\; s' \neq t' \,\}. \tag{2.36}$$

But $0 < u_1, u_2, v_1, v_2$ implies $0 < u_2/u_1, v_2/v_1 < \infty$ and (2.34) and (2.35) along with the linear independence of u, v implies $u_2/u_1 \neq v_2/v_1$. So (2.36) is true.

To show that γ is injective we note that if

$$(s,t), (s',t') \in \{\, (s',t') \in (0,\infty) \times (0,\infty) \;\mid\; s' \neq t' \}$$

and $\gamma(s,t) = \gamma(s',t')$ then

$$[b_1 + sb_2] = [b_1 + s'b_2] \text{ and } [b_1 + tb_2] = [b_1 + t'b_2].$$

But $[b_1 + sb_2] = [b_1 + s'b_2]$ implies $b_1 + sb_2 = \lambda(b_1 + s'b_2)$ for some $\lambda > 0$. But since b_1, b_2 are linearly independent it must be that $\lambda = 1$ and $s = s'$. Similarly $t = t'$. So we have proven that γ is injective, and hence bijective. $\square$

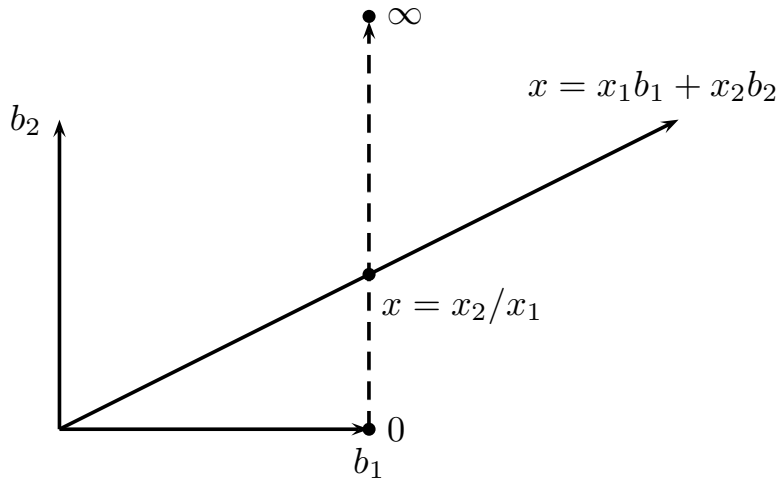


Figure 2.1: This figure illustrates the geometry of the two dimensional cone $\mathrm{Span}(f,g) \bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha, \beta \geq 0\}$ and the projection of the vector $x = x_1 b_1 + x_2 b_2 \in \mathrm{Span}(f,g) \bigcap C \setminus \{0\}$ on to the extended line $\{b_1 + x b_2 : x \in [0,\infty]\}$ (the dashed line). Note that the vectors f, g are not drawn. Via this projection one can identify the vector x and the ray $[x] = \{\lambda x : \lambda > 0\}$ with the point $b_1 + x_2/x_1\, b_2$ on the extended (dashed) line as well as with the number $x = x_2/x_1 \in [0,\infty]$. One defines the Hilbert metric on $[0,\infty]$ by $d_H(x,y) = |\ln(x/y)|$ for $x, y \in [0,\infty]$. It follows immediately from definition 1.6.2.1 (page 54) of d_H on $(\mathrm{Span}(f,g) \bigcap C \setminus \{0\}) / \sim$ and Theorem 2.1.2.1 (page 138) that $(\mathrm{Span}(f,g) \bigcap C \setminus \{0\}) / \sim$ (with d_H) is isometric to $[0,\infty]$ (with d_H).

2.2 $C \cap \ker P^n$ and d_H

2.2.1 Birkhoff's Projective Contraction Theorem and $\ker P$

Birkhoff's Projective Contraction Theorem appears on p. 222 of his widely cited 1957 paper, Extensions of Jentzsch's Theorem [12]:

THEOREM 1 (PROJECTIVE CONTRACTION THEOREM) *Let $N(P^r; C) < 1$ for some r, and let C be complete relative to $\theta(f, g; C)$. Then, for any $f \in C$, the sequence of fP^n converges geometrically to a unique fixpoint (characteristic ray) $c \in C$.* See [5].

Consider the following (trivial) "counter-example" to the above Projective Contraction Theorem: If $f \in \ker P^m \cap C$ then $fP^n \to 0$, not c. There is an easy (and

[5] $d_H = \theta$. See Section 2.3 (page 155) for a detailed discussion of $N(P; C)$ including a brief discussion of Birkhoff's notation.

obvious) way out of this dilemma: we can replace

$$\textit{for any } f \in C \qquad \text{with} \qquad \textit{for any } \ f \in C \setminus \bigcup_{n=1}^{\infty} \ker P^n$$

The following section explores this.

2.2.2 Nontrivial Kernel Counter-example

The following easy example shows it is possible to find a linear map P with $N(P^1; C) = 0$ and a non-zero vector $f \in C$ such that $[f]P^n \to [0]$.

If we wish to accept convergence to $[0]$ then we lose uniqueness of the fixpoint c.

Here is the example:

Example. Let $P : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map that projects $\mathbb{R}^2$ onto the x axis. I.e.

$$(x, y)P = (x, 0).$$

Let $C = \mathbb{R}^2_{\geq 0}$, obviously $CP \subset C$. The vectors $b_1 = (1, 0)$ and $b_2 = (0, 1)$ are a pair of ends for C. In terms of b_1, b_2: $C \cap \ker P = [b_2] \cup [0]$ and $CP = [b_1] \cup [0]$.

We show:

$$N(P; C) = \sup \left\{ \frac{d_H(fP, gP)}{d_H(f, g)} \;\middle|\; f, g \in C \setminus \{0\}, \;\; 0 < d_H(f, g) < \infty \right\} = 0.$$

If $f, g \in \ker P \cap C \setminus \{0\}$ then $f, g \in [b_2]$ and so $d_H(f, g) = 0$ and so the pair f, g is not a factor in calculating $N(P; C)$. If $f \in \ker P \cap C \setminus \{0\}$ and $gP \neq 0$ then $f = (0, f_2)$ with $f_2 \neq 0$ and $g = (g_1, g_2)$ with $g_1 \neq 0$. By the definition of d_H, Definition 1.6.2.1,

$$d_H(f, g) = \left| \ln \begin{pmatrix} f_2 & g_1 \\ 0 & g_2 \end{pmatrix} \right| = \infty$$

and so again the pair f, g is not a factor in calculating $N(P; C)$. The remaining possibility is that $f, g \in C \setminus \ker P$. But in this case, if $0 < d_H(f, g) < \infty$ then $d_H(fP, gP) = d_H(b_1, b_1) = 0$.

So $N(P; C) = 1$.

One can imagine that Birkhoff considered it obvious that we should assume P satisfies $\ker P \cap C = \{0\}$ or that we should simply disregard those f which $fP^m = 0$ for some $m \geq 1$.

2.2.3 $\ker P^n = \{f \in C : fP^n = 0\}$ and d_H

Consideration of the counter-example appearing in Section 2.2.2 (page 151) has led me to the following interesting results, in particular Theorem 2.2.3.3 (page 152) and its Corollaries 2.2.3.5 (page 155) and 2.2.3.5 (page 155), which I have not seen elsewhere.

Definition 2.2.3.1. Let L be a linear map of the vector space V to itself. Then

$$\begin{aligned} \ker L^0 &= \{0\} \\ \ker L^n &= \{v \in V : vL^n = 0\} \\ \bigcup_{n=1}^{\infty} \ker L^n &= \{v \in V : vL^n = 0 \text{ for some positive integer } n\} \end{aligned}$$

Proposition 2.2.3.2. *Let n be any non-negative integer then*

$$\ker L^n \subset \ker L^{n+1}.$$

Proof. Trivial. $vL^n = 0$ implies $vL^nL = 0L = 0$. □

The following four results are original as far as I know.

Theorem 2.2.3.3. *Let P be, as usual, a linear map of C to itself, where C is a pointed by the origin, salient, closed, convex cone in a Banach Space V.*

Let $f, g \in C \setminus \{0\}$ and let n be any positive integer. If

$$gP^n = 0 \text{ but } fP^n \neq 0 \tag{2.37}$$

then $d_H(f, g) = \infty$ and

$$g \in \partial \left(Span(f, g) \bigcap C \right) \subset \partial C. \tag{2.38}$$

Proof. By (2.37) $f, g \in C \setminus \{0\}$ are not multiples of each other and so they are linearly independent. By Theorem 1.4.0.2 (page 42) and Proposition 1.4.2.1 (page 46) there exists ends

$$b_1, b_2 \in \partial \left(\text{Span}(f, g) \bigcap C \right), \tag{2.39}$$

linearly independent, such that

$$\text{Span}(f, g) \bigcap C = \{\alpha b_1 + \beta b_2 \mid \alpha, \beta \geq 0\}$$

and so we can write

$$\begin{aligned} f &= f_1 b_1 + f_2 b_2 \\ g &= g_1 b_1 + g_2 b_2 \end{aligned}$$

for some $f_1, f_2, g_1, g_2 \geq 0$ uniquely determined. Note $b_1, b_2 \in C \setminus \{0\}$. The linearity of P together with (2.37) implies

$$fP^n = f_1 b_1 P^n + f_2 b_2 P^n \neq 0 \tag{2.40}$$

$$gP^N = g_1 b_1 P^n + g_2 b_2 P^n = 0. \tag{2.41}$$

(2.40) implies at least one of b_1P^n, b_2P^n is non-zero. (2.41) implies

$$g_1 b_1 P^n = -g_2 b_2 P^n. \tag{2.42}$$

If both $g_1, g_2 \neq 0$ then $0 < \frac{g_2}{g_1} < \infty$ and (2.42) implies

$$b_1 P^n = -\frac{g_2}{g_1}\, b_2 P^n \tag{2.43}$$

which implies both $b_1P^n, b_2P^n \neq 0$ (since we know that at least one of them is non-zero). This is impossible since $b_1P^n, b_2P^n \in C$ so

$$-\frac{g_2}{g_1}\, b_2 P^n = b_1 P^n \in C \setminus \{0\} \quad \text{and} \quad \frac{g_2}{g_1}\, b_2 P^n \in C \setminus \{0\}$$

which violates C is salient. So either g_1 or $g_2 = 0$.

Since f, g are linearly independent if $g_1 = 0$ then $f_1 \neq 0$ if $g_2 = 0$ then $f_2 \neq 0$. But in either case

$$d_H(f, g) = \left| \ln \left(\frac{f_2}{f_1} \frac{g_1}{g_2} \right) \right| = \infty.$$

Since g_1 or $g_2 = 0$ it follows that g is multiple of b_1 or b_2 and so by Proposition 1.4.2.1

$$g \in \partial \left(\mathrm{Span}(f, g) \bigcap C \right). \tag{2.44}$$

Proposition 1.1.1.8 (page 23) implies

$$\partial \left(\mathrm{Span}(f, g) \bigcap C \right) \subset \partial C. \tag{2.45}$$

Combining (2.44) with (2.45) yields (2.38). □

Corollary 2.2.3.4. *Let $f, g \in C \setminus \{0\}$. If $g \in \ker P$ but $f \notin \ker P$ then $d_H(f, g) = \infty$.*

Proof. Let $n = 1$ in Theorem 2.2.3.3 (page 152). □

The following corollaries are quite interesting as they answer the geometric question of where $\ker P^n$ can be located in the cone C. The surprising answer is that either $C \cap \ker P^n = C$ or $C \cap \ker P^n \subset \partial C$.

Corollary 2.2.3.5. *Let n be a positive integer. If $C \cap \ker P^n \neq C$ then $C \cap \ker P^n \subset \partial C$.*

Proof. If $C \cap \ker P^n \neq C$ then there exists a $f \in C \setminus \{0\}$ such that $fP^n \neq 0$. Let $g \in C \cap \ker P^n$. If $g = 0$ then $g \in \partial C$ since $0 \in \partial C$. If $g \neq 0$ then the f, g satisfy the conditions of Theorem 2.2.3.3 (page 152) so $g \in \partial C$. □

Corollary 2.2.3.6. *If there exists an $f \in C \setminus \{0\}$ such that $fP^n \neq 0$ for all integers $n > 0$ then*

$$C \cap (\cup_{n=1}^{\infty} \ker P^n) \subset \partial C.$$

Proof. Trivial consequence of Corollary 2.2.3.5 (page 155). □

2.3 $N(P;C) < 1$ iff CP has finite diameter w.r.t. d_H

A goal of this section is to provide a proof of Lemma 1., p221 of Birkhoff [12], which word for word is:

Lemma 2.3.0.7. LEMMA 1 *Birkhoff [12].*

If the transform of CP of C under P has finite diameter $\triangle$ under $\theta(f, g; C)$ then

$$N(P;C) = \tanh(\triangle/4) < 1.$$

It is to be understood that CP, in the above lemma actually refers to $CP \setminus \{0\}$; that $CP = \{cP : c \in C\}$, where $P : c \mapsto cP$ and that

$$N(P;C) = N(P) = \sup_{0<\theta(f,g)<\infty} \frac{\theta(fP, gP)}{\theta(f, g)} \tag{2.46}$$

where $\theta(f,g;C) = \theta(f,g) = d_H(f,g)$, using Birkhoff's notation. Or,

$$N(P;C) = N(P) = \sup_{0<d_H(f,g)<\infty} \frac{d_H(fP,gP)}{d_H(f,g)} \tag{2.47}$$

where we tacitly understand $f, g, fP, gP \in C \setminus \{0\}$ and that there exists at least two $f, g \in C$ such that $0 < d_H(fP, gP) < \infty$. This will always happen unless P collapses C to the origin or to a single line. We can take $N(P) = 0$ in those cases.

2.3.1 Maximizing $\frac{d_H(uP,vP)}{d_H(u,v)}$

In light of the definition of N(P;C), (2.46), (2.47) (page 156) we first consider

$$\frac{d_H(P[u], P[v])}{d_H([u],[v])}$$

with

$$[u],[v] \in \left(\text{Span}(f,g) \bigcap C \setminus \{0\}\right) / \sim \text{ and } 0 < d_H([u],[v]) < \infty.$$

Remark 2.3.1.1. We will work with the equivalence classes $[u], [v]$ etc, rather than the vectors themselves, as $d_H(u,v) = d_H([u],[v])$.

Lemma 2.3.1.2.

$$\sup_{\substack{[u],[v]\in(Span(f,g)\bigcap C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{ \frac{d_H(P[u],P[v])}{d_H([u],[v])} \right\} = \sup_{\substack{x,y\in(0,\infty) \\ x\neq y}} \left\{ \frac{\left|\ln\left(\frac{b+ax}{d+cx}\right) - \ln\left(\frac{b+ay}{d+cy}\right)\right|}{|\ln(x) - \ln(y)|} \right\}.$$

Proof. Let b_1, b_2 be a pair of ends for $\text{Span}(f,g) \bigcap C \setminus \{0\}$ and let $b_{1'}, b_{2'}$ be a pair of ends for $\text{Span}(fP,gP) \bigcap C \setminus \{0\}$ then $(\text{Span}(f,g) \bigcap C \setminus \{0\}) / \sim$ is bijective with $[0,\infty]$ by $[u] \mapsto u_2/u_1$, or going the other direction, by say $m \in [0,\infty) \mapsto [1b_1 + mb_2]$ and $m = \infty \mapsto [b_2]$. (We can think of m as being like the slope of the ray $[1b_1 + mb_2]$.)

If $[u],[v] \in (\mathrm{Span}(f,g) \bigcap C \setminus \{0\}) / \sim$ and $0 < d_H([u],[v]) < \infty$ then the definition of d_H immediately implies that neither $[u]$ nor $[v]$ is $[b_1]$ or $[b_2]$.

So

$$\sup_{\substack{[u],[v] \in (\mathrm{Span}(f,g) \bigcap C \setminus \{0\})/\sim \\ 0 < d_H([u],[v]) < \infty}} \left\{ \frac{d_H(P[u], P[v])}{d_H([u],[v])} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{d_H(P[b_1 + xb_2], P[b_1 + yb_2])}{d_H([b_1 + xb_2], [b_1 + yb_2])} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{d_H([(d+cx)b_{1'} + (b+ax)b_{2'}], [(d+cy)b_{1'} + (b+ay)b_{2'}])}{d_H([1b_1 + xb_2], [1b_1 + yb_2])} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{d_H([(d+cx)b_{1'} + (b+ax)b_{2'}], [(d+cy)b_{1'} + (b+ay)b_{2'}])}{d_H([1b_1 + xb_2], [1b_1 + yb_2])} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{d_H\left(\frac{b+ax}{d+cx}, \frac{b+ay}{d+cy}\right)}{d_H\left(\frac{x}{1}, \frac{y}{1}\right)} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{\left|\ln\left(\frac{b+ax}{d+cx} / \frac{b+ay}{d+cy}\right)\right|}{\left|\ln\left(\frac{x}{1} / \frac{y}{1}\right)\right|} \right\} =$$

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{\left|\ln\left(\frac{b+ax}{d+cx}\right) - \ln\left(\frac{b+ay}{d+cy}\right)\right|}{|\ln(x) - \ln(y)|} \right\}. \tag{2.48}$$

□

We abuse notation slightly and (using Birkhoff's convention) define $P(x) = \frac{ax+b}{cx+d}$. Then (2.48) becomes

$$\sup_{\substack{x,y \in (0,\infty) \\ x \neq y}} \left\{ \frac{|\ln(P(x)) - \ln(P(y))|}{|\ln(x) - \ln(y)|} \right\}. \tag{2.49}$$

Letting [6] $f(x) = \ln(P(x))$ and $g(x) = \ln(x)$, (2.49) becomes

$$\sup_{\substack{x,y\in(0,\infty)\\ x\neq y}} \left\{ \frac{|f(x) - f(y)|}{|g(x) - g(y)|} \right\}. \tag{2.50}$$

Dividing the numerator and denominator appearing in (2.50) by $|x - y|$ suggests we consider maximizing the ratio of derivatives

$$\frac{f'(x)}{g'(x)}.$$

2.3.2 Calculus Proposition

The following result, about the ratio of derivatives, is just elementary calculus. I have not seen this result elsewhere, but due to its elementary and useful nature, I am sure it can not be original.

Proposition 2.3.2.1. *Suppose $f(x), g(x)$ are real valued functions differentiable on $(0, \infty)$. Suppose $g'(x) > 0$ $\forall x \in (0, \infty)$. Then for all $0 < x, y < \infty$, with $x \neq y$,*

$$\frac{|f(x) - f(y)|}{|g(x) - g(y)|} \leq \sup_{z\in(0,\infty)} \left| \frac{f'(z)}{g'(z)} \right|. \tag{2.51}$$

In fact,

$$\sup_{0<x\neq y<\infty} \frac{|f(x) - f(y)|}{|g(x) - g(y)|} = \sup_{z\in(0,\infty)} \left| \frac{f'(z)}{g'(z)} \right| \tag{2.52}$$

Proof. Let

$$K = \sup_{z\in(0,\infty)} \left| \frac{f'(z)}{g'(z)} \right| \in [0, \infty].$$

If $K = \infty$ then (2.51) is immediately true.

So suppose $K < \infty$. Let $0 < x < y < \infty$. Then, since $g' > 0$, we must have

[6] $f, g \in C$ should not be confused with the functions $f(x), g(x)$. They have nothing to do with each other.

$(g(y) - g(x)) > 0$, as well as

$$\begin{aligned} |f'(x)| &\leq Kg'(x) \Rightarrow \\ \left|\int_x^y f'\right| \leq \int_x^y |f'| &\leq \int_x^y Kg' \Rightarrow \\ |f(y) - f(x)| &\leq K(g(y) - g(x)) = K|g(y) - g(x)| \Rightarrow \\ \frac{|f(y) - f(x)|}{|g(y) - g(x)|} &\leq K. \end{aligned}$$

So (2.51) is proved.

Let

$$J = \sup_{0<x\neq y<\infty} \left|\frac{f(y) - f(x)}{g(y) - g(x)}\right| \in [0, \infty]. \tag{2.53}$$

Suppose that $J < K$. Then there exists $z_0 \in (0, \infty)$ such that $0 \leq \left|\frac{f'(z_0)}{g'(z_0)}\right| < \infty$ and such that

$$J < \left|\frac{f'(z_0)}{g'(z_0)}\right| \leq K. \tag{2.54}$$

We can consider

$$\frac{f(z_0) - f(z_0 + h)}{h} \text{ and } \frac{g(z_0) - g(z_0 + h)}{h}$$

to be functions of h, with

$$\begin{aligned} \lim_{h\to 0} \frac{f(z_0) - f(z_0 + h)}{h} &= f'(z_0) \in (-\infty, \infty) \\ \lim_{h\to 0} \frac{g(z_0) - g(z_0 + h)}{h} &= g'(z_0) \in (0, \infty). \end{aligned}$$

Then, since the limit as $h \to 0$ of $\frac{g(z_0)-g(z_0+h)}{h} = g'(z_0) \neq 0$, elementary calculus tells us that

$$\lim_{h\to 0} \frac{f(z_0) - f(z_0 + h)}{g(z_0) - g(z_0 + h)} = \lim_{h\to 0} \frac{\frac{f(z_0)-f(z_0+h)}{h}}{\frac{g(z_0)-g(z_0+h)}{h}} = \frac{\lim_{h\to 0} \frac{f(z_0)-f(z_0+h)}{h}}{\lim_{h\to 0} \frac{g(z_0)-g(z_0+h)}{h}} = \frac{f'(z_0)}{g'(z_0)}. \tag{2.55}$$

Let

$$D = \left|\frac{f'(z_0)}{g'(z_0)}\right| - J. \tag{2.56}$$

By (2.54), $D > 0$. By (2.55), there exists an $h_0 > 0$ such that

$$D > \left|\frac{f'(z_0)}{g'(z_0)} - \frac{f(z_0) - f(z_0 + h_0)}{g(z_0) - g(z_0 + h_0)}\right|. \tag{2.57}$$

But then, since $|a - b| \geq |\,|a| - |b|\,|$ if $a, b \in (-\infty, \infty)$, (2.57) implies

$$\begin{aligned} D > \left|\frac{f'(z_0)}{g'(z_0)} - \frac{f(z_0) - f(z_0 + h_0)}{g(z_0) - g(z_0 + h_0)}\right| &\geq \left|\,\left|\frac{f'(z_0)}{g'(z_0)}\right| - \left|\frac{f(z_0) - f(z_0 + h_0)}{g(z_0) - g(z_0 + h_0)}\right|\,\right| \\ &\geq \left|\frac{f'(z_0)}{g'(z_0)}\right| - \left|\frac{f(z_0) - f(z_0 + h_0)}{g(z_0) - g(z_0 + h_0)}\right|. \end{aligned} \tag{2.58}$$

But then (2.56) and (2.58) imply

$$\left|\frac{f(z_0) - f(z_0 + h_0)}{g(z_0) - g(z_0 + h_0)}\right| > \left(\left|\frac{f'(z_0)}{g'(z_0)}\right| - D\right) = J = \sup_{0<x\neq y<\infty}\left|\frac{f(y) - f(x)}{g(y) - g(x)}\right|,$$

which is a contradiction. So $J = K$, proving (2.52). □

2.3.3 The calculation of $\sup\left|\frac{f'(x)}{g'(x)}\right|$

For notational convenience, we let

$$q(x) = \left|\frac{\frac{d}{dx}\ln(P(x))}{\frac{d}{dx}\ln(x)}\right|.$$

Lemma 2.3.3.1. *Let* $\lambda = \ln\left(\frac{ad}{bc}\right)$. *Then*

$$\sup_{0<x<\infty} q(x) = \tanh\left(\frac{|\lambda|}{4}\right).$$

Proof. Proposition 2.3.2.1 (page 158) combined with the calculations leading to (2.49)

and (2.50) tell us that

$$\begin{aligned}\sup_{\substack{[u],[v]\in(\mathrm{Span}(f,g)\bigcap C\setminus\{0\})/\sim\\ 0<d_H([u],[v])<\infty}}\left\{\frac{d_H(P[u],P[v])}{d_H([u],[v])}\right\} &= \sup_{\substack{x,y\in(0,\infty)\\ x\neq y}}\left\{\frac{|\ln(P(x))-\ln(P(y))|}{|\ln(x)-\ln(y)|}\right\}\\ &= \sup_{0<x<\infty}\left|\frac{\frac{d}{dx}\ln(P(x))}{\frac{d}{dx}\ln(x)}\right|\\ &= \sup_{0<x<\infty} q(x).\end{aligned}$$

In this section we find $\sup_{0<x<\infty}\left|\frac{\frac{d}{dx}\ln(P(x))}{\frac{d}{dx}\ln(x)}\right|$, or more succinctly put, we find $\sup_{0<x<\infty} q(x)$.

Using calculus we get:

$$\begin{aligned}\frac{d}{dx}\ln(P(x)) &= \frac{d}{dx}\ln\left(\frac{ax+b}{cx+d}\right)\\ &= \frac{cx+d}{ax+b}\,\frac{(cx+d)a-(ax+b)c}{(cx+d)^2}\\ &= \frac{cx+d}{ax+b}\,\frac{ad-bc}{(cx+d)^2}\text{ and}\\ &= \frac{ad-bc}{(ax+b)(cx+d)}\text{ and}\\ \frac{d}{dx}\ln(x) &= \frac{1}{x}.\end{aligned}$$

So

$$\frac{\frac{d}{dx}xP}{\frac{d}{dx}\ln(x)} = \frac{(ad-bc)x}{(ax+b)(cx+d)}.$$

We are assuming that $ad-bc\neq 0$. Recall $a,b,c,d\geq 0$. So if $x\in(0,\infty)$ then $(ax+b)(cx+d)>0$. So $x\in(0,\infty)$ implies

$$q(x)=\left|\frac{(ad-bc)x}{(ax+b)(cx+d)}\right|=\frac{|ad-bc|\,x}{acx^2+(ad+bc)x+bd}\in(0,\infty). \tag{2.59}$$

Clearly $q(x)$ is differentiable on $(0,\infty)$. Differentiating $q(x)$ yields

$$\begin{aligned}
\frac{d}{dx}\ q(x) &= \frac{d}{dx}\ \frac{|ad-bc|x}{(ax+b)(cx+d)} \\
&= |ad-bc|\ \frac{d}{dx}\ \frac{x}{acx^2+(ad+bc)x+bd} \\
&= |ad-bc|\ \frac{\left(\frac{d}{dx}x\right)(acx^2+(ad+bc)x+bd) - x\frac{d}{dx}(acx^2+(ad+bc)x+bd)}{(ax+b)^2(cx+d)^2} \\
&= |ad-bc|\ \frac{1(acx^2+(ad+bc)x+bd) - x(2acx+(ad+bc))}{(ax+b)^2(cx+d)^2} \\
&= |ad-bc|\ \frac{acx^2+(ad+bc)x+bd-2acx^2-(ad+bc)x}{(ax+b)^2(cx+d)^2} \\
&= |ad-bc|\ \frac{bd-acx^2}{(ax+b)^2(cx+d)^2}. \qquad (2.60)
\end{aligned}$$

By (2.60), if $x \in (0,\infty)$ then

$$\text{sign}\left(\frac{d}{dx}\ q(x)\right) = \text{sign}\left(bd-acx^2\right). \qquad (2.61)$$

We break finding $\sup_{0<x<\infty} q(x)$ into 4 cases: $\{bd=0, bd>0\} \times \{ac=0, ac>0\}$.

1. If $bd=0$ and $ac=0$ then, by (2.61),

$$\text{sign}\left(\frac{d}{dx}\ q(x)\right) = \text{sign}\left(bd-acx^2\right) = \text{sign}\,(0) = 0.$$

So on $(0,\infty)$ the function $q(x)$ is constant. Moreover, $ad-bc \neq 0$ and $a,b,c,d \geq 0$ (combined with $bd=0$ and $ac=0$) imply exactly one of the following two cases

(a) $b=c=0$ and $a,d>0$ so that $ad>0$ and $bc=0$,

(b) $d=a=0$ and $b,c>0$ so that $ad=0$ and $bc>0$.

Since $q(x)$ is constant

$$\begin{aligned}\sup_{0<x<\infty} q(x) &= q(x) \\ &= \frac{|ad-bc|x}{acx^2+(ad+bc)x+bd} \\ &= \frac{|ad-bc|}{ad+bc} \\ &= \begin{cases} \frac{|ad|}{ad}, & \text{if } b=c=0 \text{ and } a,d>0; \\ \frac{|bc|}{bc}, & \text{if } d=a=0 \text{ and } b,c>0. \end{cases} \\ &= 1.\end{aligned}$$

2. If $bd = 0$ and $ac > 0$ then, by (2.61),

$$\text{sign}\left(\frac{d}{dx}\ q(x)\right) = \ \text{sign}\left(bd - acx^2\right) = \ \text{sign}\left(-acx^2\right) = +.$$

So on $(0,\infty)$ the function $q(x)$ is monotonically decreasing. Moreover, $ad-bc \neq 0$ and $a,b,c,d \geq 0$ (combined with $bd = 0$ and $ac > 0$) imply exactly one of the following two cases

(a) $b = 0$ and $a, c, d > 0$ so that $ad > 0$ and $bc = 0$,

(b) $d = 0$ and $a, b, c > 0$ so that $ad = 0$ and $bc > 0$.

The continuity of $q(x)$ implies

$$\begin{aligned}
\sup_{0<x<\infty} q(x) &= \lim_{x\to 0}\ q(x) \\
&= \lim_{x\to 0}\ \frac{|ad-bc|x}{acx^2+(ad+bc)x+bd} \\
&= \lim_{x\to 0}\ \frac{|ad-bc|x}{acx^2+(ad+bc)x} \\
&= \frac{|ad-bc|}{ad+bc} \\
&= \begin{cases} \frac{|ad|}{ad}, & \text{if } b=0 \text{ and } a,c,d>0; \\ \frac{|bc|}{bc}, & \text{if } d=0 \text{ and } a,b,c>0. \end{cases} \\
&= 1.
\end{aligned}$$

3. If $bd > 0$ and $ac = 0$ then, by (2.61),

$$\text{sign}\left(\frac{d}{dx}\ q(x)\right) = \ \text{sign}\left(bd - acx^2\right) = \ \text{sign}\,(bd) = +.$$

So on $(0,\infty)$ the function $q(x)$ is monotonically increasing. Moreover, $ad-bc \neq 0$ and $a,b,c,d \geq 0$ (combined with $bd > 0$ and $ac = 0$) imply exactly one of the following two cases

(a) $a = 0$ and $b,c,d > 0$ so that $ad = 0$ and $bc > 0$,

(b) $c = 0$ and $a,b,d > 0$ so that $ad > 0$ and $bc = 0$.

The continuity of $q(x)$ implies

$$\begin{aligned}
\sup_{0<x<\infty} q(x) &= \lim_{x\to\infty} q(x) \\
&= \lim_{x\to\infty} \frac{|ad-bc|x}{acx^2+(ad+bc)x+bd} \\
&= \lim_{x\to\infty} \frac{|ad-bc|x}{(ad+bc)x+bd} \\
&= \frac{|ad-bc|}{ad+bc} \\
&= \begin{cases} \frac{|ad|}{ad}, & \text{if } c=0 \text{ and } a,b,d>0; \\ \frac{|bc|}{bc}, & \text{if } a=0 \text{ and } b,c,d>0. \end{cases} \\
&= 1
\end{aligned}$$

4. If $bd >$ and $ac > 0$ then, by (2.61),

$$\operatorname{sign}\left(\frac{d}{dx}\, q(x)\right) = \operatorname{sign}\left(bd-acx^2\right) = \begin{cases} +, & \text{if } 0<x<(bd/ac)^{1/2}; \\ 0, & \text{if } x=(bd/ac)^{1/2}; \\ -, & \text{if } (bd/ac)^{1/2}<x<\infty. \end{cases}$$

So on $(0,\infty)$ the function $q(x)$ has exactly one supremum (which also the maximum), which occurs when $x=(bd/ac)^{1/2}$. So

$$
\begin{aligned}
\sup_{0<x<\infty} q(x) = q\left((bd/ac)^{1/2}\right) &= \left|\frac{(ad-bc)(\frac{bd}{ac})^{1/2}}{(a\left(\frac{bd}{ac}\right)^{1/2}+b)(c\left(\frac{bd}{ac}\right)^{1/2}+d)}\right| \\
&= \left|\frac{(ad-bc)\left(\frac{bd}{ac}\right)^{1/2}}{ac\frac{bd}{ac}+(ad+bc)\left(\frac{bd}{ac}\right)^{1/2}+bd}\right| \\
&= \left|\frac{(ad-bc)\left(\frac{bd}{ac}\right)^{1/2}}{(ad+bc)\left(\frac{bd}{ac}\right)^{1/2}+2bd}\right| \\
&= \left|\frac{(ad-bc)\left(\frac{bd}{ac}\right)^{1/2}}{(ad+bc)\left(\frac{bd}{ac}\right)^{1/2}+2bd}\right| \text{ (mult. by } \frac{\left(\frac{bd}{ac}\right)^{-1/2}}{\left(\frac{bd}{ac}\right)^{-1/2}}) \\
&= \left|\frac{(ad-bc)}{(ad+bc)+2bd\left(\frac{bd}{ac}\right)^{-1/2}}\right| \text{ (mult. by } \frac{\frac{1}{bc}}{\frac{1}{bc}}) \\
&= \left|\frac{\frac{ad}{bc}-1}{(\frac{ad}{bc}+1)+2\frac{bd}{bc}\left(\frac{ac}{bd}\right)^{1/2}}\right| \\
&= \left|\frac{\frac{ad}{bc}-1}{(\frac{ad}{bc}+1)+2\left(\frac{d^2ac}{c^2bd}\right)^{1/2}}\right| \\
&= \left|\frac{\frac{ad}{bc}-1}{(\frac{ad}{bc}+1)+2\left(\frac{ad}{bc}\right)^{1/2}}\right| \text{ (let } \frac{ad}{bc}=\nu) \\
&= \left|\frac{\nu-1}{(\nu+1)+2\nu^{1/2}}\right| \\
&= \left|\frac{\nu-1}{\nu+2\nu^{1/2}+1}\right| \text{ (which is } \leq 1) \\
&= \left|\frac{\nu-1}{(\nu^{1/2}+1)^2}\right| \\
&= \left|\frac{(\nu^{1/2}-1)(\nu^{1/2}+1)}{(\nu^{1/2}+1)^2}\right| \\
&= \left|\frac{(\nu^{1/2}-1)}{(\nu^{1/2}+1)}\right| \text{ (mult. by } \frac{\nu^{-1/4}}{\nu^{-1/4}}) \\
&= \left|\frac{\nu^{1/4}-\nu^{-1/4}}{\nu^{1/4}+\nu^{-1/4}}\right|.
\end{aligned}
$$

Let $\lambda = \ln\left(\frac{ad}{bc}\right)$. Then

$$e^{\lambda} = \frac{ad}{bc} = \nu.$$

Making these substitutions we get:

$$\begin{aligned}\left|\frac{\nu^{1/4}-\nu^{-1/4}}{\nu^{1/4}+\nu^{-1/4}}\right| &= \left|\frac{e^{\lambda/4}-e^{-\lambda/4}}{e^{\lambda/4}+e^{-\lambda/4}}\right| \\ &= \left|\tanh\left(\frac{\lambda}{4}\right)\right|.\end{aligned}$$

Note $\tanh(x) = \frac{e^x-e^{-x}}{e^x+e^{-x}}$. Note $\tanh(x)$ is an odd function. Note if $x \geq 0$ then $0 \leq \tanh(x)$. So $|\tanh(x)| = \tanh(|x|)$. Hence, if $ad, bc > 0$ then

$$\sup_{0<x<\infty} q(x) = \max_{0<x<\infty} q(x) = \tanh\left(\frac{|\lambda|}{4}\right).$$

Let us consider the first three cases once more – those cases in which at least one of bd, ac is zero. Since $ad-bc \neq 0$ the quotient $\frac{ad}{bc}$ is not indeterminate. So, if at least one of bd, ac is zero, then $\frac{ad}{bc} = 0$ or ∞.

But then

$$|\lambda| = \left|\ln\left(\frac{ad}{bc}\right)\right| = \infty \tag{2.62}$$

and $\tanh\left(\frac{|\lambda|}{4}\right) = \tanh(\infty) = 1$, which is exactly the result obtained in the first three cases, that is that $\sup_{0<x<\infty} q(x) = 1$. So in the first 3 cases (as well as the forth) we ultimately have the same simple formula

$$\sup_{0<x<\infty} q(x) = \tanh\left(\frac{|\lambda|}{4}\right).$$

□

The next sections show that $|\lambda| = \left|\ln\left(\frac{ad}{bc}\right)\right|$ is the diameter of $P\left(\left(\mathrm{Span}(f,g)\bigcap C \setminus \{0\}\right)/\sim\right)$.

2.3.4 A notational clarification of P.

Notational clarification. The letter P can refer to three different, but related functions:

1. The linear map $P : V \to V$ which also takes $C \subset V$ into C.

2. The induced map $P : (V/\sim) \to (V/\sim)$ defined by $[v] \mapsto [vP]$. We can think of this P as taking the ray through v to the ray through vP.

3. The map $P : [0, \infty] \to [0, \infty]$ defined as $P(x) = \frac{ax+b}{cx+d}$. Where $a, b, c, d \geq 0$ are determined by the linear map P on $\mathrm{Span}(f, g) \subset V$ and by the choice of ends $b_1, b_2 \in \mathrm{Span}(f, g) \cap C$ and $b_{1'}, b_{2'} \in \mathrm{Span}(fP, gP) \cap C$. We can think of this P as taking the slope of the ray [v] w.r.t. b_1, b_2 to the slope of the ray [vP] w.r.t. $b_{1'}, b_{2'}$.

By context one can tell to which map the letter P is referring.

2.3.5 On P, slopes, d_H, and diameter.

Proposition 2.3.5.1. *$P(x) = \frac{ax+b}{cx+d}$ maps the slope of the ray $[u]$ w.r.t. b_1, b_2 to the slope of the ray $P[u]$ w.r.t. $b_{1'}, b_{2'}$. Also,*

$$\begin{aligned} d_H(P[u], P[v]) &= \left| \ln \left(\frac{m(P[u])}{m(P[v])} \right) \right| \\ &= \left| \ln \left(\frac{P(u_2/u_1)}{P(v_2/v_1)} \right) \right|. \end{aligned} \tag{2.63}$$

Proof. Recall the following.

Suppose $u, v \in \mathrm{Span}(f, g) \cap C \setminus \{0\}$ are linearly independent. We uniquely have, in terms of the linearly independent ends b_1, b_2,

$$u = u_1 b_1 + u_2 b_2$$

$$v = v_1 b_1 + v_2 b_2.$$

The "slope" of $[u]$ w.r.t. b_1, b_2 is given by $m([u]) = u_2/u_1$. This combined with the

definition of d_H, see Definition 1.6.2.1 (page 54), yields

$$d_H([u],[v]) = \left|\ln\left(\frac{u_2}{u_1}\ \frac{v_1}{v_2}\right)\right| = \left|\ln\left(\frac{m([u])}{m([v])}\right)\right|. \tag{2.64}$$

The induced map P applied to $[u]$ is defined as $P[u] = [uP]$. The 'slope' of $[uP]$ w.r.t. $b_{1'}, b_{2'}$ is given by, (see (2.48)),

$$m([uP]) = P(u_2/u_1) = \frac{a(u_2/u_1)+b}{b(u_2/u_1)+c}.$$

So if

$$x = u_2/u_1 = m([u]) = \text{ the slope of } [u], \tag{2.65}$$

then

$$P(x) = \frac{ax+b}{bx+c} = P(m([u])) = \text{ the slope of } P[u]. \tag{2.66}$$

So $P(x) = \frac{ax+b}{cx+d}$ maps the slope of the ray $[u]$ to the slope of the ray $P[u]$. Of course all slopes are relative to the ends b_1, b_2 and $b_{1'}, b_{2'}$. It follows from (2.64),(2.65), and (2.66) that

$$\begin{aligned} d_H(P[u],P[v]) &= \left|\ln\left(\frac{m(P[u])}{m(P[v])}\right)\right| \\ &= \left|\ln\left(\frac{P(u_2/u_1)}{P(v_2/v_1)}\right)\right| \end{aligned} \tag{2.67}$$

which is (2.63). □

Definition 2.3.5.2. The **diameter** of $P\left(\left(\mathrm{Span}(f,g)\cap C\setminus\{0\}\right)/\sim\right)$ which we denote by $\Delta_{P,f,g,C}$, or just Δ when its meaning is clear, is given by

$$\Delta_{P,f,g,C} = \sup\left\{d_H(P[u],P[v]) : [u],[v] \in \left(\mathrm{Span}(f,g)\cap C\setminus\{0\}\right)/\sim\right\}.$$

From Theorem 2.1.2.1 (page 138) we know that if f, g are linear independent then

$[0,\infty]$ is bijective with

$$(\mathrm{Span}(f,g)\cap C\setminus\{0\})/\sim$$

via the slope function m. More precisely, the map

$$[u]\in(\mathrm{Span}(f,g)\cap C\setminus\{0\})/\sim\ \mapsto\quad m([u])=u_2/u_1=x\in[0,\infty],$$

is bijective. This, together with (2.67) immediately yields the diameter formula

$$\Delta_{P,f,g,C}=\left|\ln\left(\frac{\sup_{0\le x\le\infty}P(x)}{\inf_{0\le y\le\infty}P(y)}\right)\right|. \tag{2.68}$$

We have the following proposition.

Proposition 2.3.5.3. *Let $f,g\in V\cap C$ be linearly independent. If fP,gP are also linear independent then $P(x)=\frac{ax+b}{cx+d}$ is monotonic on $[0,\infty]$ and the diameter $\Delta_{P,f,g,C}$ of $P\left((Span(fP,gP)\cap C\setminus\{0\})/\sim\right)$ is given by the formula*

$$\Delta_{f,g,C,P}=\left|\ln\left(\frac{b}{d}\ \frac{c}{a}\right)\right|.$$

Proof. From Theorem 2.1.2.1 (page 138) we know that if $[fP]\neq[gP]$ then $ad-bc\neq 0$. Since $ad-bc\neq 0$ and since $a,b,c,d\ge 0$ it follows that at most one of c,d are zero; at most one of b,d are zero; at most one of a,b are zero; and at most one of a,c are zero. So if $x\in(0,\infty)$ then $(ax+b),(cx+d)>0$. This implies $P(x)=\frac{ax+b}{cx+d}$ is differentiable, finite and strictly positive on $(0,\infty)$. We differentiate $P(x)=\frac{ax+b}{cx+d}$:

$$\begin{aligned}\frac{d}{dx}P &= \frac{d}{dx}\frac{ax+b}{cx+d}\\ &= \frac{(cx+d)a-(ax+b)c}{(cx+d)^2}\\ &= \frac{ad-bc}{(cx+d)^2}.\end{aligned}$$

Since $(cx+d) > 0$ as explained immediately above, $P'(x)$ is finite on $(0, \infty)$. Moreover, the sign of $P'(x)$ is equal to $ad - bc \neq 0$. So $P(x)$ is monotonic on $(0, \infty)$.

Since at most one of b, d is zero, the ratio b/d is not indeterminate, hence

$$\lim_{x \to 0} P(x) = \lim_{x \to 0} \frac{ax + b}{cx + d} = P(0) = b/d \in [0, \infty]. \tag{2.69}$$

Since at most one of a, c is zero, the ratio a/c is not indeterminate, hence

$$\lim_{x \to \infty} P(x) = \lim_{x \to \infty} \frac{ax + b}{cx + d} = P(\infty) = a/c \in [0, \infty]. \tag{2.70}$$

If $ad - bc < 0$ then $P'(x)$ is negative and $P(x)$ is monotonically decreasing on $(0, \infty)$. This, together with equations (2.69) and (2.70) imply

$$\inf_{x \in [0,\infty]} P(x) = \lim_{x \to 0} P(x) = b/d \in [0, \infty)$$

$$\sup_{x \in [0,\infty]} P(x) = \lim_{x \to \infty} P(x) = a/c \in (0, \infty]$$

and the strict inequality $b/d < a/c$.

If $ad - bc > 0$ then $P'(x)$ is positive and $P(x)$ is monotonically increasing on $(0, \infty)$. This, together with equations (2.69) and (2.70) imply

$$\inf_{x \in [0,\infty]} P(x) = \lim_{x \to \infty} P(x) = a/c \in [0, \infty)$$

$$\sup_{x \in [0,\infty]} P(x) = \lim_{x \to 0} P(x) = b/d \in (0, \infty]$$

and the strict inequality $a/c < b/d$.

Finally we use the diameter formula, (2.68).

If $ad - bc < 0$ then

$$\Delta_{P,f,g,C} = \left| \ln \left(\frac{\sup_{0 \leq x \leq \infty} P(x)}{\inf_{0 \leq y \leq \infty} P(y)} \right) \right| = \left| \ln \left(\frac{a/c}{b/d} \right) \right| = \left| \ln \left(\frac{a}{c} \, \frac{d}{b} \right) \right| = \left| \ln \left(\frac{b}{d} \, \frac{c}{a} \right) \right|.$$

If $ad - bc > 0$ then

$$\Delta_{P,f,g,C} = \left| \ln \left(\frac{\sup_{0 \leq x \leq \infty} P(x)}{\inf_{0 \leq y \leq \infty} P(y)} \right) \right| = \left| \ln \left(\frac{b/d}{a/c} \right) \right| = \left| \ln \left(\frac{b}{d} \ \frac{c}{a} \right) \right|.$$

□

Corollary 2.3.5.4. *Suppose fP, gP are linearly independent, so that $ad - bc \neq 0$.*

1. *If none of a, b, c, d are zero, then $0 < \Delta_{P,f,g,C} < \infty$.*

2. *If at least one of a, b, c, d is zero, then $\Delta_{P,f,g,C} = \infty$.*

3. *If $\Delta_{P,f,g,C} = \infty$ then at least one of $[b_1 P], [b_2 P]$ must equal at least one of $[b_{1'}], [b_{2'}]$.*

Proof. Part *1.* Immediate consequence of Proposition 2.3.5.3 (page 170).

Part *2.* Immediate consequence of (2.62) (page 167) combined with Proposition 2.3.5.3 (page 170).

Part *3.* Note that by Parts 1. and 2. of this Corollary we must have at least one of a, b, c, d equal to zero. In Theorem 2.1.2.1 (page 138), Equations (2.4), (2.5) define a, b, c, d via

$$\begin{aligned} b_1 P &= d b_{1'} + b b_{2'} \\ b_2 P &= c b_{1'} + a b_{2'}. \end{aligned}$$

Hence Part 3. is proven. □

2.3.6 Birkhoff's Lemma 1: $N(P; C) = \tanh(\triangle/4)$

We finally are in striking distance of of a proof of Lemma 1, p. 221 of Birkhoff [12], which word for word is:

Lemma 1. *If the transform of CP of C under P has finite diameter* $\triangle$ *under* $\theta(f,g;C)$ *then*

$$N(P;C) = \tanh\left(\triangle/4\right) < 1.$$

But first two propositions:

Proposition 2.3.6.1. *As usual, let* C *be a closed, convex, pointed-by-the-origin cone in a Banach Space* V *and let* P *be linear map of* V *to itself such that* $CP \subset C$.

If $f, g \in C \setminus \{0\}$ *are linearly independent and* fP, gP *are linearly independent, then*

$$\sup_{\substack{[u],[v]\in(Span(f,g)\bigcap C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{ \frac{d_H(P[u],P[v])}{d_H([u],[v])} \right\} = \tanh\left(\frac{\Delta_{f,g,C,P}}{4}\right)$$

where $\Delta_{P,f,g,C}$ *is the diameter of* $P\left(\left(Span(f,g)\cap C\setminus\{0\}\right)/\sim\right)$ *w.r.t.* d_H.

Proof.

$$\sup_{\substack{[u],[v]\in(\mathrm{Span}(f,g)\bigcap C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{ \frac{d_H(P[u],P[v])}{d_H([u],[v])} \right\} = \tag{2.71}$$

$$\sup_{\substack{x,y\in(0,\infty) \\ x\neq y}} \left\{ \frac{\left|\ln\left(\frac{b+ax}{d+cx}\right) - \ln\left(\frac{b+ay}{d+cy}\right)\right|}{\left|\ln(x) - \ln(y)\right|} \right\} = \tag{2.72}$$

$$\sup_{x,y\in(0,\infty)x\neq y} \left\{ \frac{\left|\ln\left(P(x)\right) - \ln\left(P(y)\right)\right|}{\left|\ln(x) - \ln(y)\right|} \right\} = \tag{2.73}$$

$$\sup_{0<x<\infty} \left| \frac{\frac{d}{dx}\ln(P(x))}{\frac{d}{dx}\ln(x)} \right| = \tag{2.74}$$

$$\tanh\left(\frac{|\lambda|}{4}\right) = \tag{2.75}$$

$$\tanh\left(\frac{\Delta_{f,g,C,P}}{4}\right).$$

The first equality (2.71) follows from Lemma 2.3.1.2 (page 156). The second equality (2.72) follows from using Birkhoff's notation; i.e. by defining $P(x) = \frac{ax+b}{cx+d}$. The third equality (2.73) follows from our Calculus Proposition 2.3.2.1 Equality (2.52)

(page 158). The fourth equality (2.74) follows from Lemma 2.3.3.1 (page 160) – recall we are letting $\lambda = \ln\left(\frac{ad}{bc}\right)$. The fifth equality (2.75) follows from Proposition 2.3.5.3 (page 170). Recall $\Delta_{P,f,g,C}$ is the diameter of $P\left(\left(\mathrm{Span}(f,g) \cap C \setminus \{0\}\right)/\sim\right)$ with respect to d_H; see Definition 2.3.5.2 (page 169). □

The proof of the following proposition is basically Birkhoff's argument (however I've included the details), see Lemma 1, Birkhoff [12].

Proposition 2.3.6.2.

$$N(P;C) = \sup_{\substack{[u],[v]\in(C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{ \frac{d_H(P[u],P[v])}{d_H([u],[v])} \right\} = \tanh\left(\frac{\Delta_{C,P}}{4}\right) \tag{2.76}$$

where $\Delta_{P,C}$ is the diameter of $P\left(\left(C \setminus \{0\}\right)/\sim\right)$ w.r.t. d_H.

Proof. The first equality in Equation (2.76) is just the definition of $N(P;C)$. Let

$$S = \sup_{\substack{[u],[v]\in(C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{ \frac{d_H(P[u],P[v])}{d_H([u],[v])} \right\}$$

and let $u_n, v_n \in C \setminus \{0\}$, $n = 1, 2, \ldots$ be such that $0 < d_H([u_n],[v_n]) < \infty$ and such that

$$\frac{d_H(P[u_n],P[v_n])}{d_H([u_n],[v_n])}$$

is a monotonically increasing sequence whose limit is S. Since $0 < d_H([u_n],[v_n])$ it follows that u_n, v_n are linearly independent. Fix n and let u_n, v_n play the roll of f, g

in Proposition 2.3.6.1 (page 173). So for each n

$$\begin{aligned}\frac{d_H(P[u_n],P[v_n])}{d_H([u_n],[v_n])} &\leq \sup_{\substack{[u],[v]\in(\mathrm{Span}(u_n,v_n)\cap C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{\frac{d_H(P[u],P[v])}{d_H([u],[v])}\right\} \\ &= \tanh\left(\frac{\Delta_{u_n,v_n,C,P}}{4}\right) \\ &\leq \tanh\left(\frac{\Delta_{C,P}}{4}\right)\end{aligned}$$

with the last inequality follows from tanh being monotonically increasing. So

$$S \leq \tanh\left(\frac{\Delta_{C,P}}{4}\right). \tag{2.77}$$

Next, let $u_n, v_n \in C\setminus\{0\}$, $n = 1, 2, \ldots$ be such that $0 < d_H([u_n],[v_n]) < \infty$ and such that

$$d_H(P[u_n],P[v_n])$$

is a monotonically increasing sequence whose limit is $\Delta_{C,P}$. Since tanh is continuous, monotonically increasing and bounded above by 1 we have

$$\lim_{n\to\infty} \tanh\left(\frac{d_H(P[u_n],P[v_n])}{4}\right) = \tanh\left(\frac{\Delta_{C,P}}{4}\right). \tag{2.78}$$

Since $0 < d_H([u_n],[v_n])$ it follows that u_n, v_n are linearly independent. Fix n and let u_n, v_n play the roll of f, g in Proposition 2.3.6.1 (page 173). So for each n

$$\begin{aligned}\tanh\left(\frac{d_H(P[u_n],P[v_n])}{4}\right) &\leq \tanh\left(\frac{\Delta_{u_n,v_n,C,P}}{4}\right) \\ &= \sup_{\substack{[u],[v]\in(\mathrm{Span}(u_n,v_n)\cap C\setminus\{0\})/\sim \\ 0<d_H([u],[v])<\infty}} \left\{\frac{d_H(P[u],P[v])}{d_H([u],[v])}\right\} \\ &\leq S.\end{aligned}$$

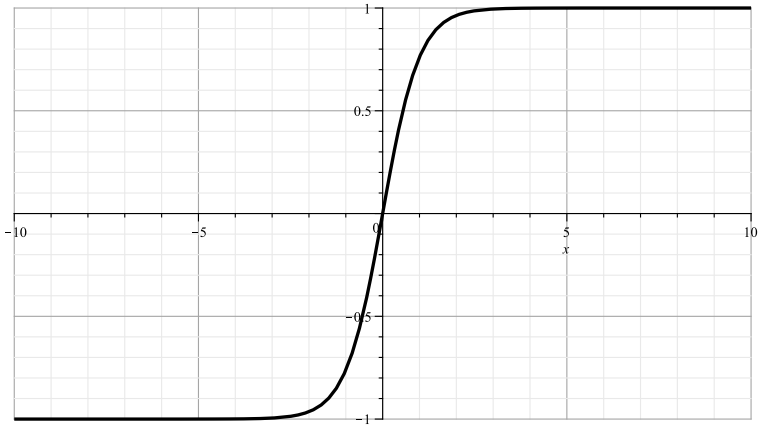


Figure 2.2: A plot of tanh.

So, by (2.78),

$$\tanh\left(\frac{\Delta_{C,P}}{4}\right) \leq S. \tag{2.79}$$

Combining (2.77) with (2.79) finishes the proof. □

We finally prove Lemma 1., p221 of Birkhoff [12], which word for word is:

Lemma 1. *If the transform of CP of C under P has finite diameter* $\triangle$ *under* $\theta(f, g; C)$ *then*

$$N(P; C) = \tanh\left(\triangle/4\right) < 1.$$

Proof. If $0 \leq \Delta_{C,P} < \infty$ then

$$0 \leq \tanh\left(\frac{\Delta_{C,P}}{4}\right) < 1. \tag{2.80}$$

Note $\tanh(x)$ is strictly monotonically increasing, see Figure 2.2 (page 176), and

$$\lim_{x\to\infty} \tanh(x) = 1.$$

By Proposition 2.3.6.2 (page 174)

$$N(P;C) = \tanh\left(\frac{\Delta_{C,P}}{4}\right). \tag{2.81}$$

Combining (2.80) with (2.81) finishes the proof. □

2.3.7 Special case: P maps Span(b_1, b_2) to itself. Eigenvectors in $\mathbb{R}^2$.

Proposition 2.3.7.1. *Let V be a Banach Space and suppose that $b_1, b_2 \in V$ are linearly independent. Define*

$$C_{12} = \{\alpha b_1 + \beta b_2 : \alpha, \beta \geq 0\}. \tag{2.82}$$

Suppose that the linear map P, defined on at least Span(b_1, b_2), maps C_{12} to itself. Then there exists $a, b, c, d \geq 0$ and unique such that

$$b_1 P = d b_1 + b b_2 \tag{2.83}$$

$$b_2 P = c b_1 + a b_2 \tag{2.84}$$

If $x = x_1 b_1 + x_2 b_2 \in C_{12}$ then $xP = (xP)_1 b_1 + (xP)_2 b_2 \in C_{12}$ We can calculate $(xP)_1, (xP)_2$ via matrix multiplication:

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} dx_1 + cx_2 \\ bx_1 + ax_2 \end{pmatrix} = \begin{pmatrix} (xP)_1 \\ (xP)_2 \end{pmatrix} \tag{2.85}$$

Suppose $a, b, c, d > 0$. Let

$$x_+ = \frac{-(d-a) + \sqrt{(d-a)^2 + 4cb}}{2c}$$

then $x_+ > 0$ *and the vector*

$$b_1 + x_+ b_2 \tag{2.86}$$

is an eigenvector pointing in the unique eigen-direction of P *in* C_{12}*. This eigenvector* $b_1 + x_+ b_2$ *will have eigenvalue*

$$d + x_+ c > 0. \tag{2.87}$$

Let $u, v \in C \setminus \{0\}$ be such that $0 < d_H(u, v) < \infty$. Then u, v are linearly independent and for some linearly independent pair $f, g \in C \setminus \{0\}$ we have $[u], [v] \in (\mathrm{Span}(f, g) \bigcap C \setminus \{0\}) / \sim$. In fact, we could even let $u = f$, and $v = g$.

Proof. The first part of this proposition (up to Equation (2.85)) is essentially a special case of the first three parts of Theorem 2.1.2.1 (page 138). However it is worthwhile to sketch an independent proof which uses only algebra:

Equations (2.83),(2.84) with $a, b, c, d \geq 0$ are a trivial consequence of the definition of C_{12} and P mapping C_{12} to itself. The uniqueness of a, b, c, d follow from the linear independence of the pair b_1, b_2. The linearity of P, together with (2.83) and (2.84), imply

$$\begin{aligned} xP &= (x_1 b_1 + x_2 b_2) P \\ &= (x_1 (b_1 P) + x_2 (b_2 P) \\ &= (x_1 (d b_1 + b b_2) + x_2 (c b_1 + a b_2) \\ &= \underbrace{(x_1 d + x_2 c)}_{(xP)_1} b_1 + \underbrace{(x_1 b + x_2 a)}_{(xP)_2} b_2 \end{aligned} \tag{2.88}$$

which yields the matrix Equation (2.85).

We solve the eigenvector equation

$$xP = \lambda x \tag{2.89}$$

by applying the 'slope' function m to both sides of (2.89) to get

$$m([xP]) = m([\lambda x]) = m([x]). \tag{2.90}$$

Note, the 'slope' $m([\lambda x]) = \frac{\lambda x_2}{\lambda x_1}$ is indeterminate; i.e. is $0/0$, if and only if $\lambda x = 0$. See [7]. If x is an eigenvector of P in C_{12} then $\lambda x \neq 0$ for the following reason: We are assuming that $a, b, c, d > 0$. So if $x \in C_{12} \setminus \{0\}$ is an eigenvector of P, simple matrix multiplication implies that both $(xP)_1$ and $(xP)_2$ are > 0, but then λ, x_1 and x_2 are all > 0. Hence if x is an eigenvector in C_{12} none of the slopes in (2.90) will be indeterminate; in fact they will all be positive.

We derive a formula for $m([xP])$, the slope of $[xP]$, in terms of a, b, c, d and $m([x])$, the slope of $[x]$.

$$\begin{aligned}
m([xP]) &= \frac{(xP)_2}{(xP)_1} \\
&= \frac{x_1 b + x_2 a}{x_1 d + x_2 c} \\
&= \frac{x_1 b + x_2 a}{x_1 d + x_2 c} \\
&= \frac{b + (x_2/x_1)a}{d + (x_2/x_1)c} \\
&= \frac{b + m([x])a}{d + m([x])c} \quad (\text{let } x_* = m([x]) = x_2/x_1\ ) \\
&= \frac{ax_* + b}{cx_* + d} \\
&\doteq P(x_*).
\end{aligned}$$

So (2.90) becomes

$$P(x_*) = x_* \tag{2.91}$$

Every solution of x_* of (2.91) corresponds to an eigenvector $b_1 + x_* b_2$ of P. The

[7] Provided $v = v_1 b_1 + v_2 b_2 \neq 0$ the ratio $m([v]) = v_2/v_1 \in \mathbb{R} \cup \infty$ is well defined. We take $v_2/0 = \infty$ provided $v_2 \in \mathbb{R} \setminus \{0\}$.

solutions of (2.91) yield all the determinate, finite non-zero slopes of eigenvectors of P in $\text{Span}(b_1, b_2)$. As discussed within this proof a few paragraphs back, the eigenvectors of P in C_{12} will only have positive slopes and positive eigenvalues. So solving (2.91) for x_* will yield to us all eigenvectors of P in C_{12}.

$$\begin{aligned} P(x_*) &= x_* \quad \Leftrightarrow \\ \frac{ax_* + b}{cx_* + d} &= x_* \quad \Leftrightarrow \\ ax_* + b &= x_*(cx_* + d) \quad \Leftrightarrow \\ 0 &= {cx_*}^2 + (d-a)x_* - b \quad \Leftrightarrow \\ x_* &= \frac{-(d-a) \pm \sqrt{(d-a)^2 + 4cb}}{2c}. \end{aligned} \tag{2.92}$$

Let

$$x_+ = \frac{-(d-a) + \sqrt{(d-a)^2 + 4cb}}{2c} > 0. \tag{2.93}$$

$$x_- = \frac{-(d-a) - \sqrt{(d-a)^2 + 4cb}}{2c} < 0. \tag{2.94}$$

Since we must have non-negative slope (if the eigenvector is to be in C_{12}) we have shown $b_1 + x_+ b_2$ is an eigenvector pointing in the unique eigen-direction of P in C_{12}. Finally we calculate its eigenvalue λ.

Combining Equation (2.88) and (2.89) yields

$$\underbrace{(x_1 d + x_2 c)b_1 + (x_1 b + x_2 a)b_2}_{xP} = \underbrace{\lambda x_1 b_1 + \lambda x_2 b_2}_{\lambda x}$$

which implies, by the linear independence of b_1, b_2 that

$$x_1 d + x_2 c = \lambda x_1 \text{ and } x_1 b + x_2 a = \lambda x_2. \tag{2.95}$$

If x is an eigenvector of P in C_{12}, then as discussed above, $x_1, x_2 > 0$. Dividing the first part of (2.95) by x_1 and the second part of (2.95) by x_2 yields

$$(x_2/x_1)c + d = \lambda \text{ and } (x_1/x_2)b + a = \lambda. \tag{2.96}$$

As $x_2/x_1 = x_+ > 0$, the first part of (2.96) yields

$$cx_+ + d = \lambda > 0. \tag{2.97}$$

□

The following remarks refer to Proposition 2.3.7.1 (page 177).

Remark 2.3.7.2. It can be shown that the vector $b_1 + x_- b_2$ will be an eigenvector of P outside of C_{12}. Its eigenvalue $d + x_- c$ can be negative, zero, or positive depending on a, b, c, d.

Remark 2.3.7.3. Note that $ad - bc$ might equal 0 since we aren't assuming that $b_1 P, b_2 P$ are linearly independent.

Remark 2.3.7.4. It is worth noting that if x is an eigenvector of P then the ray $[x]$ is fixed by P (if $\lambda > 0$), reversed by P (if $\lambda < 0$), or sent to $[0]$ by P (if $\lambda = 0$).

Remark 2.3.7.5. If $\ker P \cap C_{12} = \{0\}$ there will be an eigenvector in C_{12} by the Brouwer Fixed Point Theorem as the following argument shows. Suppose $\ker P \cap C_{12} = \{0\}$. Then P induces a continuous mapping of the closed line segment $\overline{b_1 b_2}$ to itself, where $\overline{b_1 b_2}$ is the line segment connecting b_1 to b_2, explicitly

$$\overline{b_1 b_2} = \{\alpha b_1 + \beta b_2 : \alpha, \beta \geq 0, \alpha + \beta = 1\} \subset C_{12}.$$

The map induced by P that takes $\overline{b_1 b_2}$ to itself is

$$x \in \overline{b_1 b_2} \mapsto \frac{1}{(xP)_1 + (xP)_2}\, xP \in \overline{b_1 b_2}. \tag{2.98}$$

Since $\overline{b_1 b_2}$ is homeomorphic to the closed unit ball in $\mathbb{R}$, the Brouwer Fixed Point Theorem guarantees that the induced map (2.98), has at least one fixed point. The fixed point of that map is an eigenvector of P with eigenvalue $(xP)_1+(xP)_2 \neq 0$ (since $\ker P \cap C_{12} = \{0\}$). On the other hand, if $\ker P \cap C_{12}$ contains non-zero vectors, then these non-zero vectors would be eigenvectors with eigenvalue zero. So we've shown that in either case, that P will have at least one eigenvector in C_{12}.

Remark 2.3.7.6. If we allow for some of a, b, c, d to be zero, we can give examples where the eigenvector will not be unique (e.g. if P is the identity map), or could be b_1, or b_2.

Remark 2.3.7.7. With the assumptions of this section, that in terms of d_H, x is an eigenvector of P with positive eigenvalue λ if and only if

$$\begin{aligned} 0 = d_H(\lambda x, x) = d_H(xP, x) &= |\ln(P(m(x))/m(x))| \\ \Leftrightarrow P(m(x))/m(x) &= 1. \end{aligned}$$

where $m(x) = x_2/x_1$ and $P(x) = \frac{ax+b}{cx+d}$.

Example. Consider the matrix

$$A = \begin{pmatrix} 5 & 1 \\ 4 & 2 \end{pmatrix} = \begin{pmatrix} d & c \\ b & a \end{pmatrix}.$$

Using formula (2.86) we calculate an eigenvector for A that points in the unique eigen-direction of A in the first quadrant of $\mathbb{R}^2$

$$b_1 + x_+ b_2 = b_1 + \frac{-(d-a) + \sqrt{(d-a)^2 + 4cb}}{2c}\, b_2 \quad = b_1 + \frac{-(5-2) + \sqrt{(5-2)^2 + 4 \cdot 1 \cdot 4}}{2 \cdot 1}\, b_2$$
$$= b_1 + \frac{-3 + \sqrt{25}}{2}\, b_2$$
$$= b_1 + b_2.$$

So $x_+ = 1$. The eigenvalue λ corresponding to the eigenvector $b_1 + x_+ b_2$ is computed using formula (2.87), we get $\lambda = d + cx_+ = 5 + 1 = 6$. A quick check shows that $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of A with eigenvalue 6.

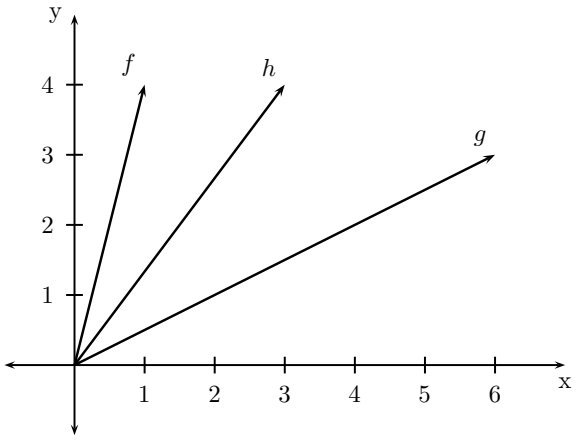


Figure 2.3: $h \in \text{cone}\{f, g\}$.

2.4 Cones and Hyperplanes

2.4.1 The cone $\{f, g\}$

Definition 2.4.1.1. Let $f, g \in C \setminus \{0\}$. Then $\text{cone}\{f, g\} = \{\alpha f + \beta g : \alpha\beta \geq 0\}$.

It is easy to see that $\text{cone}\{f, g\}$ will be a cone that inherits many of the properties of the cone C. In particular, convexity, salience. Clearly $\text{cone}\{f, g\}$ is closed.

Lemma 2.4.1.2. *Let $f, g, h \in Span(f, g) \bigcap C$ be pairwise linearly independent. Then at least one of the following are true:*

$$f \in cone\{g, h\} \text{ or}$$

$$g \in cone\{f, h\} \text{ or}$$

$$h \in cone\{f, g\}.$$

Proof. See Figure 2.3 (page 184). Suppose

$$f \notin \text{cone}\{g, h\} \text{ and } g \notin \text{cone}\{f, h\}.$$

Since f, g, h are pairwise linearly independent and in $\text{Span}(f, g)$ it follows that

$$\text{Span}(f, g) = \text{Span}(g, h) = \text{Span}(f, h).$$

Hence, we can uniquely write f in terms of g and h, and g in terms of f and h:

$$\begin{aligned} f &= \alpha g + \beta h \\ g &= \alpha' f + \beta' h. \end{aligned} \tag{2.99}$$

The signs of α, β (and α', β') must be mixed since $f \notin \text{cone}\{g, h\}$ (and $g \notin \text{cone}\{f, h\}$) imply the signs can not both be positive. None of $\alpha, \alpha', \beta, \beta'$ can be zero, since f, h and g, h are linearly independent. C is a salient cone so both signs can't be negative. Since the signs of α, β (and α', β') must be mixed and non-zero, both of the ratios $-\frac{\alpha}{\beta}$ and $-\frac{\alpha'}{\beta'}$ are positive.

Manipulation of (2.99) yields:

$$\begin{aligned} \frac{1}{\beta} f + -\frac{\alpha}{\beta} g &= h \\ -\frac{\alpha'}{\beta'} f + \frac{1}{\beta'} g &= h. \end{aligned}$$

Since f, g are linearly independent, we must have, by uniqueness,

$$\frac{1}{\beta} = -\frac{\alpha'}{\beta'} \text{ and } -\frac{\alpha}{\beta} = \frac{1}{\beta'}$$

so

$$\underbrace{\left(-\frac{\alpha'}{\beta'}\right)}_{\text{positive}} f + \underbrace{\left(-\frac{\alpha'}{\beta'}\right)}_{\text{positive}} g = h$$

so $h \in \text{cone}\{f, g\}$. □

2.4.2 Cone Hyperplane Intersection Lemmas

Many of the following results require the existence of a hyperplane H which intersects each equivalence class (i.e. each 0-ray) of $(C \setminus \{0\}, \sim)$ exactly once; see [8].

Birkhoff [12], in his proof of his Projective Contraction Mapping Theorem, uses and states (or requires) the existence of such a plane. The existence of such a plane is not required in the proof, as I show in this work. However, the existence of such a hyperplane is useful in relating convergence with respect to d_H to convergence with respect to d_V.

There seems to be a relatively small amount of literature on the existence of such a plane H. However, see [3].

There is a large body of literature on closely related topics: on the separation of convex bodies by hyperplanes (the various separation versions of the Hann-Banach Theorem) and on the support of cones and convex sets by hyperplanes (see for example [2] and [57]). The various results tend to be of a non-trivial nature.

I show below that there exists a linear functional L such that $L > 0$ on $C \setminus \{0\}$ if and only if there exists a hyperplane H which intersects each 0-ray exactly once. See Lemmas 2.4.2.1 (page 187) and 2.4.2.2 (page 188) below. Charalambos and Tourky prove a similar result for cone bases [9] in Theorem 1.47 on page 40 of [3], using a similar, but not identical argument. These results, combined with results by Klee [57],[56], on linear functionals, to be discussed further below, give a partial answer to when one should one expect such a hyperplane to exist.

In the example that we are ultimately concerned with – that is, when the cone C is simply $\mathbb{R}^n_{\geq 0}$, the hyperplane H consisting of all $x = \sum_i^n x_i e_i$ such that $\sum_i^n x_i =$

[8] Recall that if $f \in C \setminus \{0\}$. We say f is projectively equivalent to f', written $f \sim f'$, if and only if $f' = \lambda f$ for some $\lambda > 0$. The equivalence class of f, denoted $[f]$, is sometimes called a 0-ray for the obvious reason.

[9] $\mathcal{B}$ is a base for the cone C if $\mathcal{B}$ is a convex subset of $C \setminus \{0\}$ and if for each $x \in C \setminus \{0\}$ there exists a unique $b \in B$ and a unique $\lambda > 0$ such that $x = \lambda b$.

1 intersects each positive ray $[c]$ exactly once, at the point $c/||c||_1$; see [10]. It is worthwhile noting that this hyperplane H is just the solution of

$$L(x) = (1, 1, 1, \ldots, 1) \cdot x = 1.$$

Lemma 2.4.2.1. *Suppose that C is a subset of V, an arbitrary vector space of finite or infinite dimension; that $\alpha C \subset C$ for every $\alpha > 0$, and that there exists a linear functional L from V to the reals such that $0 < L(c)$ for each $c \in C \setminus \{0\}$.*

Let $H = \{x \in V : L(x) = 1\}$ and for each $c \in C \setminus \{0\}$ let

$$\frac{1}{L(c)}\ c = c^H$$

be the central projection of c onto H. Recall that $[c] = \{\lambda c : \lambda > 0\}$ is the equivalence class of c with respect to $\sim$ and that geometrically $[c]$ is a ray in C; for details see Section 1.7.1. Then

1. *$C \bigcap -C \subset \{0\}$.*
2. *Suppose $c, c' \in C \setminus \{0\}$ then $c^H = (c')^H$ if and only if $[c] = [c']$. Moreover, $[c^H] = [c]$.*
3. *H intersects each ray $[c]$ in $C \setminus \{0\}$ once and only once: $[c] \cap H = \{c^H\}$.*
4. *$c^H + \ker(L) = H$ so that H is a hyperplane in V.*

Proof. 1. If $x \in C \bigcap -C \setminus \{0\}$ then $-x \in C \bigcap -C \setminus \{0\}$. But then both $L(x) > 0$ and $L(-x) = -L(x) > 0$, which is impossible.

[10] If the vectors in $C = \mathbb{R}^n_{\geq 0}$ with integer components are interpreted as being frequencies, then vectors in $H \cap C$ can be interpreted as being distributions whose components are relative frequencies. I.e. if $c = \{c_i\}_{i=1}^n \in H \cap C$ then $\sum c_i = 1$ with each $c_i \geq 0$. This interpretation is what motivates our study of cones, hyperplanes, and Birkhoff's Projective Contraction Mapping Theorem.

2. Let $c, c' \in C \setminus \{0\}$ so $L(c), L(c') > 0$. If

$$\begin{aligned} c^H &= (c')^H \quad \text{then} \\ \frac{1}{L(c)}\ c &= \frac{1}{L(c')}\ c' \quad \text{then} \\ c' &= \frac{L(c')}{L(c)}\ c \quad \text{then} \\ [c'] &= [c]. \end{aligned}$$

On the other hand, suppose $c' \in [c]$, then there exists an $\lambda > 0$ such that $c' = \lambda c$. But then

$$(c')^H = \frac{1}{L(c')}\ c' = \frac{1}{L(\lambda c)}\ \lambda c = \frac{1}{L(c)}\ c = c^H.$$

Since $c^H = \frac{1}{L(c)}\ c$, $c^H \in [c]$, but then $[c^H] = [c]$.

3. A little algebra yields

$$L\left(c^H\right) = L\left(\frac{1}{L(c)}\ c\right) = \frac{1}{L(c)}\ L(c) = 1,$$

so $c^H \in H$. Since $c^H \in [c]$ we have $c^H \in H \cap [c]$.

On the other hand, if $c' \in H \cap [c]$ then $c' \in [c^H] = [c]$. So $c' = \lambda c^H$ for some $\lambda > 0$. Since $c' \in H$ we have $L(c') = 1$. But then

$$1 = L(c') = L(\lambda c^H) = \lambda L(c^H) = \lambda,$$

so $c' = c^H$. Hence $H \cap [c] = \{c^H\}$.

4. $\ker(L)$ is a vector subspace of V so $c^H + \ker(L)$ is a hyperplane. The following argument shows that $c^H + \ker(L) = H$: Suppose that $k \in \ker(L)$. Then $L(c^H + k) = L(c^H) + L(k) = 1 + 0 = 1$. So $c^H + \ker(L) \subset H$. On the other hand, suppose that $x \in H$. Then $L(x - c^H) = L(x) - L(c^H) = 1 - 1 = 0$. So $x - c^H \in \ker(L)$. But then $x = c^H + (x - c^H) \in c^H + \ker(L)$. So $H \subset c^H + \ker(L)$. □

Lemma 2.4.2.2. *Suppose that C is a subset of V, an arbitrary vector space of finite or infinite dimension; that C contains at least one non-zero vector; that $\alpha C \subset C$ for every $\alpha > 0$, and that there exists a hyperplane H which intersects every ray $[c] = \{\lambda c : \lambda > 0\}$ in $C \setminus \{0\}$ exactly once. Then there exists a linear functional L mapping V to the reals such that $L(c) > 0$ for every $c \in C \setminus \{0\}$.*

Proof. Since H is hyperplane there exists a vector $v \in V$ and a subspace $W \subset V$ such that $H = v + W$. Let $c \in H \cap C \setminus \{0\}$, such a c exists by this Lemma's main assumption.

Claim 1: $-c + H$ is a vector subspace of V. Proof: $c = v + w_c$ for some $w_c \in W$. Hence $-c + H = -(v + w_c) + v + W = w_c + W = W$.

Claim 2: $c \notin -c + H$. Proof: if $c \in -c + H$ then $c = -c + h_c$ for some $h_c \in H$. but then $2c = h_c \in [c]$. That means $c, 2c \in H \cap [c]$. This contradicts that each ray intersects H exactly once. So claim 2 is proven.

Let $\mathcal{B}_{-c+H}$ be a Hamel basis [11] for $-c + H$. Since $c \notin -c + H$, which is a vector subspace, the set $\mathcal{B}_{-c+H} \cup \{c\}$ forms a basis for $\mathrm{Span}(-c + H, c)$. Since $H \subset \mathrm{Span}(-c + H, c)$ it follows that $C \subset \mathrm{Span}(-c + H, c)$.

If $\mathrm{Span}(-c + H, c) = V$, let $\mathcal{B} = \mathcal{B}_{-c+H} \cup \{c\}$. If $\mathrm{Span}(-c + H, c)$ is a proper vector subspace of V we can extend the basis $\mathcal{B}_{-c+H} \cup \{c\}$ to a basis $\mathcal{B}$ for V. See Theorem 4.72 in [73] regarding extensions of bases.

If $x \in V$ we define $L(x)$ as follows. We can write x uniquely as a finite linear combination of basis elements from $\mathcal{B}$:

$$x = \alpha_x c + \sum_{i=1}^{n_x} \alpha_i b_i.$$

Define $L(x) = \alpha_x$.

If $d \in C \setminus \{0\}$ then, by this Lemma's assumption, $[d] \cap H$ is a single element,

[11] A Hamel Basis for a vector space X is a linearly independent set whose span (meaning all finite linear combinations) is X. The Hamel Basis is just the regular algebraic basis.

which we'll call d^H. So $d^H = [d] \cap H$. Since $d^H \in H = c + (-c + H)$ we must have $d^H = c + \sum_{i=1}^{n_d} \alpha_i b_i$ with $b_i \in \mathcal{B}_{-c+H}$. Since $[d] = [d^H]$ there exists an $\alpha > 0$ such that $d = \alpha_d d^H$. So $d = \alpha_d d^H = \alpha_d c + \alpha_d \sum_{i=1}^{n_d} \alpha_i b_i$.

So $L(d) = \alpha_d > 0$. □

2.4.3 Some general notes about hyperplanes

The following standard result about the smallest hyperplane generated by subset $\mathcal{S}$ of a vector space V is useful.

Lemma 2.4.3.1. *Let $\mathcal{S}$ be any subset of a vector space V. Let s_0 be any fixed element of $\mathcal{S}$ and let $W = Span\{s_1 - s_2 : s_1, s_2 \in \mathcal{S}\}$. Then $s_0 + W$ is the smallest hyperplane containing $\mathcal{S}$.*

Proof. W is a vector subspace of V. So $s_0 + W$ is a hyperplane. $\mathcal{S} \subset s_0 + W$ since $\forall\ s \in \mathcal{S}$ we have $s - s_0 \in W$ and so $s = s_0 + (s - s_0) \in s_0 + W$. If H' is any hyperplane which contains $\mathcal{S}$ then $H' = h' + W'$ for some fixed element $h' \in H'$ and vector subspace $W' \subset V$. Since $\mathcal{S} \subset H'$ we have $s_0 = h' + w'$ for some $w' \in W'$. But then

$$s_0 + W' = h' + w' + W' = h' + W' = H' \text{ so that}$$
$$W' = -s_0 + H'.$$

Since $\mathcal{S} \subset H'$ we have $-s_0 + \mathcal{S} \subset -s_0 + H' = W'$. This implies $(-s_0 + s_1) - (-s_0 + s_2) = s_1 - s_2 \in W'$ for each pair $s_1, s_2 \in \mathcal{S}$. So $W \subset W'$. So

$$s_0 + W \subset s_0 + W' = H'.$$

□

Lemma 2.4.3.2. *Let f, g be linearly independent vectors in the vector space V. Suppose that $x, y, z \in Span(f, g)$ and that they do not lie on the same line; i.e. they are not collinear. Then if H is a hyperplane and $x, y, z \in H$, then $Span(f, g) \cap H = Span(f, g)$.*

Proof. Let H_{xyz} be the smallest hyperplane containing x, y, z. By Lemma 2.4.3.1 (page 190)

$$H_{xyz} = x + \text{Span}(y - x, z - x, y - z).$$

But $y - z = (y - x) - (z - x)$ so $\text{Span}(y - x, z - x, y - z) = \text{Span}(y - x, z - x)$, which implies

$$H_{xyz} = x + \text{Span}(y - x, z - x).$$

Since x, y, z are not collinear and $x, y, z \in \text{Span}(f, g)$ it follows that $y - x, z - x$ form a basis for $\text{Span}(f, g)$. So $\text{Span}(y - x, z - x) = \text{Span}(f, g)$. But then

$$\begin{aligned} H_{xyz} &= x + \text{Span}(y - x, z - x) \\ &= x + \text{Span}(f, g) \\ &= \text{Span}(f, g). \end{aligned}$$

By Lemma 2.4.3.1 (page 190) $H_{xyz} \subset H$, so $\text{Span}(f, g) \cap H = \text{Span}(f, g)$. □

2.4.4 Cone Bases and Intersecting Hyperplanes Theorem

Definition 2.4.4.1. $\mathcal{B}$ is a base for the cone C if $\mathcal{B}$ is a convex subset of $C \setminus \{0\}$ and if for each $x \in C \setminus \{0\}$ there exists a unique $b \in B$ and a unique $\lambda > 0$ such that $x = \lambda b$.

Theorem 2.4.4.2. *Suppose that the convex cone C is a subset of V, an arbitrary vector space of finite or infinite dimension, and that C contains at least one non-zero vector. Then the following are equivalent.*

1. *There exists a base $\mathcal{B}$ for the cone C.*

2. *There exists a linear functional on V which is strictly positive on $C \setminus \{0\}$.*

3. *There exists a hyperplane H such that H intersects each 0-ray in $C \setminus \{0\}$ exactly once.*

Additionally:

If 1. holds then the smallest hyperplane $H_{\mathcal{B}}$ containing the base $\mathcal{B}$ will intersect each 0-ray exactly once. Let b_0 be any element of $\mathcal{B}$; let $W = Span\{b_1 - b_2 : b_1, b_2 \in \mathcal{B}\}$; then $H_{\mathcal{B}} = b_0 + W$ and $H_{\mathcal{B}} \cap C = \mathcal{B}$.

Proof. 1. is equivalent to 2. is proven by Aliprantis and Tourky in Theorem 1.47 on page 40 of [3]. 2. is equivalent to 3. is proven in our Lemmas 2.4.2.1 (page 187) and 2.4.2.2 (page 188).

Now we prove the "Additionally" part (in a roundabout way).

In their proof of Theorem 1.47 of [3], Aliprantis and Tourky show that if $\mathcal{B}$ is a base for the cone C, then there exists a linear functional L on V which is strictly positive on $C \setminus \{0\}$ and which takes the value of 1 on $\mathcal{B}$, see [12]. So the hyperplane $H_1 = \{v \in V : L(v) = 1\}$ contains $\mathcal{B}$. We also have $\mathcal{B} \subset C$. So $\mathcal{B} \subset H_1 \cap C$.

On the other hand, if $c \in C \setminus \{0\}$ then there exists a unique $b_c \in \mathcal{B}$ and a unique $\lambda > 0$ such that $c = \lambda b_c$, so $L(c) = L(\lambda b_c) = \lambda$. So if $c \in H_1 \cap C$ then $L(c) = 1$, which, by the uniqueness of λ and b_c, forces $c = b_c \in \mathcal{B}$. So $H_1 \cap C \subset \mathcal{B}$ and $H_1 \cap [c] = b_c$. So $H_1 \cap C = \mathcal{B}$.

By Lemma 2.4.3.1 (page 190), $H_{\mathcal{B}} = b_0 + W$ is the smallest hyperplane containing $\mathcal{B}$ so $H_{\mathcal{B}} \subset H_1$ and $H_{\mathcal{B}} \cap C \subset H_1 \cap C = \mathcal{B}$. On the other hand, $\mathcal{B} \subset H_{\mathcal{B}} \cap C$. So $H_{\mathcal{B}} \cap C = \mathcal{B}$.

Since $H_1 \cap [c] = b_c$ and $\mathcal{B} \subset H_{\mathcal{B}} \subset H_1$ we have $H_{\mathcal{B}} \cap [c] = b_c$. So $H_{\mathcal{B}}$ intersects each 0-ray of $C \setminus \{0\}$ exactly once. □

[12] Aliprantis and Tourky use the convexity of $\mathcal{B}$ to prove that L is linear.

Corollary 2.4.4.3. *If C is a salient closed cone in a finite dimensional vector V space then there exists a hyperplane H such that H intersects each 0-ray in C and $H \cap C$ is compact.*

Proof. According to Aliprantis and Tourky [3], Cor 3.8, Klee [56] proved that every closed [salient] cone C in a finite dimensional vector space has a compact base, $\mathcal{B}$. From our perspective, by Theorem 2.4.4.2 (page 191), there exists a hyperplane H, which intersects each 0-ray in C exactly once and $H \cap C = \mathcal{B}$. □

Remark 2.4.4.4. Klee's paper, "Separation Properties of Convex Cones" [57], much referenced in the literature, shows that a closed convex cone C in a separable [13], normed, linear space will have associated to it a linear functional which is strictly positive on $C \setminus \{0\}$. However, the following example shows that given an arbitrary salient cone C in a Banach space V, we can not always find a strictly positive linear functional on $C \setminus \{0\}$. By Theorem 2.4.4.2 (page 191), this means we can not always find a hyperplane H which intersect every 0-ray in $C \setminus \{0\}$ exactly once.

2.4.5 A cone that no hyperplane intersects each 0-ray exactly once

Example. We can't always find a hyperplane H which intersects every 0-ray of a closed cone C exactly once even if the underlying vector space is Banach, as the following example, based upon Problem 6, page 42 of [3], shows. By our Lemmas 2.4.2.1 (page 187) and 2.4.2.2 (page 188) the existence of such a hyperplane is equivalent to the existence of a non-zero linear functional on V which is strictly positive on $C \setminus \{0\}$.

Let $V = B(\Omega) =$ the set of all bounded functions from $\Omega =$ an uncountable set, to $\mathbb{R}$. V equipped with the sup norm, see [14], becomes an l^∞ Banach Space. Let $C =$

[13] X separable means there exists a countable dense subset in X.
[14] If $\phi \in B(\Omega)$ then $||\phi||_\infty = \sup_{\omega \in \Omega} |\phi(\omega)|$.

all the bounded non-negative functions from Ω to $\mathbb{R}$. For each $A \subset \Omega$ let

$$\chi_A(x) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$$

Then $\chi_A \in C$ for each $A \subset \Omega$. Note $\chi_\emptyset = 0$ and if $A \neq \emptyset$ then $||\chi_A||_\infty = 1$. If A, B are subsets of Ω then

$$\chi_{A \cup B} = \chi_A + \chi_B - \chi_{A \cap B}. \tag{2.100}$$

Suppose L is any linear functional on V; that $A \subset B \subset \Omega$; and that F is any finite subset of Ω. Then (2.100) implies

$$L(\chi_B) = L(\chi_{B \setminus A \, \cup \, A}) = L(\chi_{B \setminus A}) + L\left(\chi_A\right) \tag{2.101}$$

and

$$L(\chi_F) = \sum_{\omega \in F} L(\chi_\omega). \tag{2.102}$$

Let us suppose that the linear functional $L > 0$ on all of $C \setminus \{0\}$. If A is strictly contained in B then (2.101) implies $L(\chi_A) < L(\chi_B)$ (so L is strictly monotonic with respect to χ). Let $S_{1/n} = \{\omega \in \Omega : L(\chi_\omega) > 1/n\}$. If the cardinality of $S_{1/n}$ is infinite, then (2.102) combined with the strict monotonicity of L implies $L(\chi_{S_{1/n}})$ is infinite – which is impossible, since $L : V \to \mathbb{R}$; i.e. $L(\chi_{S_{1/n}})$ must be real. But then $S = \bigcup_{n=1}^{\infty} S_{1/n}$ is at most countable, and so $\Omega \setminus S \neq \emptyset$. So suppose $\omega_0 \in \Omega \setminus S$, but then $L(\chi_{\omega_0}) = 0$, which contradicts $L > 0$ on $C \setminus \{0\}$.

Definition 2.4.5.1. Let C be a cone. Suppose that H is a hyperplane which intersects each 0-ray; i.e. each equivalence class $[c] \in (C \setminus \{0\}, \sim)$ exactly once. We denote the unique element of $[c] \cap H$ as c^H. We call the map $c \mapsto c^H$ central projection onto H, although actually the map is only into H, it is onto $(C \setminus \{0\}) \cap H$.

2.4.6 The cone $\mathbb{R}^n_{\geq 0}$ in $\mathbb{R}^n$

The canonical example of a finite dimensional complete normed linear space is $\mathbb{R}^n$ with the euclidean norm. The canonical (and most important to us) cone in $\mathbb{R}^n$ is the cone of non-negative vectors and is denoted $\mathbb{R}^n_{\geq 0}$. Explicitly:

Definition 2.4.6.1. Let e_i be the i^{th} standard basis vector for $\mathbb{R}^n$. I.e. e_i has a 1 in the i^{th} argument and $0's$ in the other $n-1$ arguments. Then

$$\mathbb{R}^n_{\geq 0} = \left\{ \sum_{i=1}^{n} x_i e_i : x_i \geq 0 \right\}.$$

Definition 2.4.6.2. $d_E(x, y)$ is the standard Euclidean distance between x and y.

Theorem 2.4.6.3. *$\mathbb{R}^n_{\geq 0}$ is a closed, salient, convex cone. Moreover,*

$$\mathbb{R}^n_{\geq 0} = int\left(\mathbb{R}^n_{\geq 0}\right) \bigcup \partial_{top}\, \mathbb{R}^n_{\geq 0} \tag{2.103}$$

where

$$int\left(\mathbb{R}^n_{\geq 0}\right) = \left\{ \sum_{i=1}^{n} x_i e_i \;\middle|\; x_i > 0 \right\}$$

is the topological interior [15] *of $\mathbb{R}^n_{\geq 0}$ with respect to the usual topology on $\mathbb{R}^n$ and*

$$\partial_{top}\, \mathbb{R}^n_{\geq 0} = \left\{ \sum_{i=1}^{n} x_i e_i \;\middle|\; x_i \geq 0 \text{ and at least one of the } x_i = 0 \right\}$$

is the topological boundary [16] *of $\mathbb{R}^n_{\geq 0}$ with respect to the usual topology on $\mathbb{R}^n$. The above union, (2.103), is disjoint.*

In the proof of Theorem 2.4.6.3, which follows, $\mathbb{R}^n_{\geq 0}$ is represented as the intersection of n closed supporting half spaces. That such a representation exists is not

[15] The topological interior of a set $A = \text{int}(A) = \{a \in A : \exists \text{ an open set } U_a \text{ with } a \in U_a \subset A\}$.

[16] The topological boundary of a set $A \subset X$, where X is a topological space $= \partial_{top}(A) = \{p \in X :$ whenever U_p is an open set containing p then $U_p \cap (X \setminus A) \neq \emptyset$ and $U_p \cap A \neq \emptyset\}$.

surprising as every closed convex set is representable as the possibly infinite intersection of supporting half spaces [44].

Proof. The projection maps $\pi_i : \mathbb{R}^n \to \mathbb{R}$ by $\pi_i\left(\sum_{i=1}^n x_i e_i\right) = x_i$ are continuous for $i = 1, 2, \ldots n$. So the "closed" half planes $\pi_i^{-1}([0, \infty))$ are closed sets. Hence $\mathbb{R}^n_{\geq 0} = \cap_{i=1}^n \pi_i^{-1}([0, \infty))$ is closed. That $\mathbb{R}^n_{\geq 0}$ is also salient, convex, and a cone is trivial and no proof of that will be given.

Let

$$I = \left\{ \sum_{i=1}^n x_i e_i \;\middle|\; x_i > 0 \right\} \text{ and } B = \left\{ \sum_{i=1}^n x_i e_i \;\middle|\; x_i \geq 0 \text{ and at least one of the } x_i = 0 \right\}.$$

Since $\mathbb{R}^n_{\geq 0}$ is closed, it contains its boundary, i.e:

$$\partial_{top}\, \mathbb{R}^n_{\geq 0} \subset \mathbb{R}^n_{\geq 0}. \tag{2.104}$$

Similarly, the "open" half planes $\pi_i^{-1}((0, \infty))$ are open sets. Hence $I = \cap_{i=1}^n \pi_i^{-1}((0, \infty))$ is open. I being open and entirely contained in $\mathbb{R}^n_{\geq 0}$, together with (2.104), implies

$$\partial_{top}\, \mathbb{R}^n_{\geq 0} \subset \mathbb{R}^n_{\geq 0} \setminus I = B. \tag{2.105}$$

The equality, $\mathbb{R}^n_{\geq 0} \setminus I = B$, appearing in (2.105), follows from $\mathbb{R}^n_{\geq 0}$ being the disjoint union of I and B.

Let $\epsilon > 0$ be given and let $b \in B$. $\pi_j(b) = 0$ for at least one particular integer $j \in 1, 2, \ldots n$. The vector b', defined by $\pi_i(b') = \pi_i(b)$ for $i \neq j$, and $\pi_j(b') = -\epsilon/2$ is in the open ϵ ball centered at b, but it is not in $\mathbb{R}^n_{\geq 0}$. So $B \subset \partial_{top}\, \mathbb{R}^n_{\geq 0}$. This together with (2.105) yields $B = \partial_{top}\, \mathbb{R}^n_{\geq 0}$. Since $\text{int}\left(\mathbb{R}^n_{\geq 0}\right) \subset \mathbb{R}^n_{\geq 0} = I \cup B$ and since I is open, $B = \partial_{top}\, \mathbb{R}^n_{\geq 0}$ implies $I = \text{int}\left(\mathbb{R}^n_{\geq 0}\right)$. The union (2.103) is obviously disjoint. □

2.4.7 The hyperplane H_1 and the simplex $\triangle^{n-1} = H_1 \cap \mathbb{R}^n_{\geq 0}$

Definition 2.4.7.1. The hyperplane $H_1 \subset \mathbb{R}^n$ is defined as follows

$$H_1 = \left\{ x \in \mathbb{R}^n \middle| x = \sum_{i=1}^{n} x_i e_i,\ x_i \in \mathbb{R},\ \sum_{i=1}^{n} x_i = 1 \right\}. \tag{2.106}$$

The following are useful and obvious:

1. If $[x] \in \left(\mathbb{R}^n_{\geq 0} \setminus \{0\}, \sim\right)$ then H_1 intersects $[x]$ exactly once: $[x] \cap H_1 = \frac{x}{\|x\|_1}$.

2. The central projection of $x \in \mathbb{R}^n_{\geq 0} \setminus \{0\}$ onto H_1 is the map

$$x \mapsto x^{H_1} = \frac{x}{\|x\|_1},$$

where $\|x\|_1$ is the l^1 norm of x

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|,$$

since

$$\left\| \frac{x}{\|x\|_1} \right\|_1 = 1.$$

3. If the arguments of $x \in$ consist of frequencies, then $\frac{x}{\|x\|_1}$ gives the relative frequencies, or distribution of x.

4. $H_1 \cap \mathbb{R}^n_{\geq 0}$ is the standard (or unit) $n-1$ simplex from algebraic topology:

$$\triangle^{n-1} = \left\{ (t_1, t_2, \ldots, t_n) \ \middle|\ \sum_{i=1}^{n} = 1 \text{ and } t_i \geq 0 \text{ for all } i \right\} = H_1 \cap \mathbb{R}^n_{\geq 0}. \tag{2.107}$$

5. $H_1 \cap \mathbb{R}^n_{\geq 0}$ is a cone base [17] for the cone $\mathbb{R}^n_{\geq 0}$.

Theorem 2.4.7.2. *If $n \geq 2$ then the diameter of $H_1 \cap \mathbb{R}^n_{\geq 0} = \sqrt{2}$. See* [18].

[17] For the definition of cone base see Definition 2.4.4.1 (page 191).

[18] The diameter of a subset A of Euclidean Space $= \sup\{d_E(x, y) : x, y \in A\}$.

Proof. This is a consequence of the standard result that the diameter of a simplex is the maximum distance between its vertices. For a proof of this see p. 120 of Hatcher's *Algebraic Topology* [43].

The distance between the vertices of the simplex $\triangle^{n-1} = H_1 \cap \mathbb{R}^n_{\geq 0}$ are all $\sqrt{2}$ since the vertices of $\triangle^{n-1}$ are just the standard basis vectors $e_i,\ i = 1, 2, \ldots, n$ and $d_E(e_i, e_j) = \sqrt{(1-0)^2 + (0-1)^2 + (n-2)(0-0)^2} = \sqrt{2}$ if $i \neq j$. So the theorem is proven. Also see Barnette [8] for more general results involving vertices and algorithms for computing the diameter in polytopes. □

2.4.8 The line $H \cap \mathbf{Span}(f, g)$, linear independence, t_{min}, t_{max}

Lemma 2.4.8.1. *Let H be a hyperplane in V which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once and suppose $f, g \in C$ are linearly independent. Then*

$$H \cap Span(f, g) = \{f^H + t(g^H - f^H) : t \in \mathbb{R}\}$$

which is the line passing through f^H and g^H. Moreover, $0 \notin H$ and if

$$a^H, b^H, c^H \in H \bigcap Span(f, g) \bigcap C$$

then a^H, b^H, c^H are collinear.

Proof. By Lemma 2.4.3.1 (page 190) the smallest hyperplane containing f^H and g^H is the line

$$\{f^H + t(g^H - f^H) : t \in \mathbb{R}\} = l(f^H, g^H).$$

So $l(f^H, g^H) \subset H \cap \mathrm{Span}(f, g)$.

On the other hand, if it is the case that

$$z \in H \cap \mathrm{Span}(f, g) \setminus l(f^H, g^H)$$

then f^H, g^H, z are three non-collinear vectors in $H \cap \text{Span}(f,g)$ and Lemma 2.4.3.2 tells us that $H \cap \text{Span}(f,g) = \text{Span}(f,g)$. But then $H \cap [f] = [f]$, which contradicts that H intersects each equivalence of $(C \setminus \{0\}, \sim)$ exactly once. So $l(f^H, g^H) = H \cap \text{Span}(f,g)$.

Let $c \in C \setminus \{0\}$. Since H is a hyperplane, if H contains both c and 0, then H contains the line passing though c and 0. But then $[c] \subset H$, which contradicts that H intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. So $0 \notin H$.

□

Lemma 2.4.8.2. *Let H be a hyperplane which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Then $0 \notin H$. Moreover, if $v, w \in H$ and $v \neq w$ then v, w are linearly independent.*

Proof. Claim: $0 \notin H$. Proof of Claim: If $v, w \in H$, and H is any hyperplane, then $\{v + \kappa(w - v) : \kappa \in \mathbb{R}\} \subset H$. So if $w = 0 \in H$ then

$$\{v + \kappa(0 - v) : \kappa \in \mathbb{R}\} = \{(1 - \kappa)v : \kappa \in \mathbb{R}\} = \{\kappa v : \kappa \in \mathbb{R}\} \subset H. \tag{2.108}$$

Since we are assuming that H intersects each equivalence in $(C \setminus \{0\}, \sim)$ exactly once, if $f \in C \setminus \{0\}$ then $H \cap [f] = f^H$. But then, as $f^H \in H$, we have, by (2.108), that $[f] \subset \{\kappa f^H : \kappa \in \mathbb{R}\} \subset H$. But that contradicts that H intersecting each equivalence class in $C \setminus \{0\}$ exactly once. So $0 \notin H$ and the claim is proved.

Claim: if $v, w \in H,\ v \neq w$ then v, w are linearly independent. Proof of Claim: If $v \neq w$ are linearly dependent then $\kappa' v = w$ for some $\kappa' \in \mathbb{R}$. But then, for all $\kappa \in \mathbb{R}$ we have

$$v + \kappa(w - v) = v + \kappa(\kappa' v - v) = (1 + \kappa(\kappa' - 1))v \in H. \tag{2.109}$$

If $\kappa = -\frac{1}{\kappa' - 1}$ then (2.109) implies $0 \in H$. However, we have just proved that $0 \notin H$ so we must have that $\kappa' = 1$, in which case $v = w$. So if $v \neq w$, then v,w must be linearly independent. □

Theorem 2.4.8.3. *Let H be a hyperplane in V which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once; let $f, g \in C$ be linearly independent; let*

$$t_{min} = \min\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\}$$
$$t_{max} = \max\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\}.$$

Then both t_{min} and t_{max} are finite with

$$-\infty < t_{min} \leq 0 \text{ and } 1 \leq t_{max} < \infty.$$

Moreover,

$$H \bigcap Span(f, g) \bigcap C = \{f^H + t(g^H - f^H) : t \in [t_{min}, t_{max}]\} \tag{2.110}$$

So $H \bigcap Span(f, g) \bigcap C$ is a closed and bounded line segment with end points

$$f^H + t_{min}(g^H - f^H) \quad \text{and} \quad f^H + t_{min}(g^H - f^H)$$

having finite length, with respect to d_V, given by

$$(t_{max} - t_{min}) \left\| g^H - f^H \right\|_V = (t_{max} - t_{min}) d_V(g^H, f^H)$$

where d_V is the metric induced by V's norm $\| \ \|_V$.

Proof. Since $g^H \in C$ we must have $t_{max} \geq 1$.

If t_{max} is unbounded, then for t arbitrarily large we will have

$$\begin{aligned} f^H + t(g^H - f^H) &\in C \\ \Rightarrow \frac{1}{t}\left(f^H + t(g^H - f^H)\right) &\in C \\ \Rightarrow \frac{1}{t}\ f^H + \left(g^H - f^H\right) &\in C \\ \Rightarrow g^H - f^H &\in C \end{aligned}$$

with the last implication holding because C is closed. Since f, g are linear independent $g^H - f^H \neq 0$. So $g^H - f^H \in \text{Span}(f, g) \bigcap C \setminus \{0\}$. Since $g^H - f^H \in C \setminus \{0\}$ we must have that H intersects $[g^H - f^H]$ exactly once: let λ be the unique positive real number such that $H \bigcap [g^H - f^H] = \lambda(g^H - f^H)$.

By Lemma 2.4.8.1 (page 198)

$$\begin{aligned} H \bigcap \text{Span}(f, g) =& \{f^H + t(g^H - f^H) : t \in \mathbb{R}\} \text{ so} \\ \lambda(g^H - f^H) = H \bigcap [g^H - f^H] \in& \{f^H + t(g^H - f^H) : t \in \mathbb{R}\} \end{aligned}$$

so there should exist a $t_* \in \mathbb{R}$ such that

$$\lambda(g^H - f^H) = f^H + t_*(g^H - f^H).$$

But then $(\lambda - t_*)g^H = (1 + \lambda - t_*)f^H$ which is impossible because f^H, g^H are linearly independent. Hence we reach a contradiction. Hence t_{max} must have been bounded above; i.e. $t_{max} < \infty$.

Since $f^H \in C$ we must have $t_{min} \leq 0$. A similar argument to that given for t_{max} will show that $-\infty < t_{min}$.

Since C is closed $f^H + t_{min}(g^H - f^H)$ and $f^H + t_{max}(g^H - f^H)$ are both in C.

Since C is convex the line segment

$$\{f^H + t(g^H - f^H) : t \in [t_{min}, t_{max}]\} \subset C.$$

By Lemma 2.4.8.1 (page 198)

$$H \bigcap \text{Span}(f, g) = \{f^H + t(g^H - f^H) : t \in \mathbb{R}\}$$

so

$$H \bigcap \text{Span}(f, g) \bigcap C = \{f^H + t(g^H - f^H) : t \in [t_{min}, t_{max}]\}.$$

Finally,

$$\begin{aligned} d_V\left(f^H + t_{min}(g^H - f^H), f^H + t_{max}(g^H - f^H)\right) = \\ = \left\|\left(f^H + t_{min}(g^H - f^H)\right) - \left(f^H + t_{max}(g^H - f^H)\right)\right\| \\ = \left\|(t_{min} - t_{max})(g^H - f^H)\right\| \\ = (t_{max} - t_{min})\left\|(g^H - f^H)\right\|. \end{aligned}$$

□

2.4.9 The Hilbert Projective Metric $d_H(f, g)$ using $\alpha f \leq g \leq \beta f$

The following definition of the Hilbert Projective Metric is very useful. For proofs of the assertions made in this section and a more detailed treatment, please see Section 1.9 (page 80).

Definition 2.4.9.1. Let $f, g \in C \setminus \{0\}$ and let α, β be the greatest and least values for which

$$\alpha f \leq g \leq \beta f$$

is true. Here $a \leq b$ means $b - a \in C$. In terms of sets

$$\alpha = \sup\{t \in \mathbb{R} : g - tf \in C\} \quad \text{and} \quad \beta = \inf\{t \in \mathbb{R} : tf - g \in C\} \tag{2.111}$$

where for β we take $\beta = \infty$ if $\nexists\ \beta$ such that $\beta f - g \in C$.

The Hilbert Projective Metric on $C \setminus \{0\}$ is defined by

$$d_H(f, g) = \ln\left(\frac{\beta}{\alpha}\right) \in [0, \infty].$$

We define d_H on $(C \setminus \{0\}, \sim)$ by $d_H([f], [g]) = d_H(f, g)$. Note. d_H is not actually a true metric on $C \setminus \{0\}$, see item 2. immediately below.

Let $u \in C \setminus \{0\}$ and let $C_u = \{f \in C \setminus \{0\} : d_H(f, u) < \infty\}$. In Theorem 1.9.3.2 (page 84) we prove the following assertions:

1. $$\alpha \in [0, \infty) \qquad \beta \in (0, \infty] \quad \text{and} \quad 0 \leq \alpha \leq \beta \leq \infty.$$

2. d_H is an extended pseudo-metric on $C \setminus \{0\}$ meaning all the axioms of a true metric space hold except that d_H can take the value ∞ (extended) and $d_H(f, g) = 0$ does not automatically imply $f = g$ (pseudo).

3. d_H is an extended metric on $(C \setminus \{0\}, \sim)$. In particular if $f \sim f'$ and $g \sim g'$ then $d_H(f, g) = d_H(f', g')$.

4. If we define $f \equiv g$ if $d_H(f, g) < \infty$ then $\equiv$ is an equivalence relation on $C \setminus \{0\}$ and C_u is the $\equiv$ equivalence class containing u.

5. d_H is a true metric on $(C_u, \sim)$ for each $u \in C \setminus \{0\}$.

2.4.10 Technical Lemma regarding t_{max}, t_{min} and d_H

Lemma 2.4.10.1. *Let H be a hyperplane in V which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Let $f, g \in C$ be linearly independent; and let f^H (resp. g^H) be the unique intersection point of H with the equivalence class of f (resp. g). Define*

$$\begin{aligned} t_{min} &= \min\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\} \\ t_{max} &= \max\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\} \\ \alpha &= \sup\{t \in \mathbb{R} : g^H - tf^H \in C\} \\ \beta &= \inf\{t \in \mathbb{R} : tf^H - g^H \in C\} \end{aligned}$$

so that

$$d_H(f^H, g^H) = \ln\left(\frac{\beta}{\alpha}\right).$$

Then:

1. $1 \leq t_{max} < \infty$.

2. $$\frac{1}{t_{max}}\ \left(f^H + t_{max}\left(g^H - f^H\right)\right) = g^H - \alpha f^H \in C.$$

3. $$\alpha = 1 - \frac{1}{t_{max}} = \frac{t_{max} - 1}{t_{max}}.$$

4. $$0 \leq \alpha < 1.$$

5. $-\infty < t_{min} \leq 0$.

6. *If* $t_{min} < 0$, *then*

$$\frac{-1}{t_{min}} \left(f^H + t_{min} \left(g^H - f^H \right) \right) = \beta f^H - g^H \in C.$$

7. *If* $t_{min} < 0$, *then*

$$\beta = 1 - \frac{1}{t_{min}} = \frac{t_{min} - 1}{t_{min}}.$$

8. *If* $t_{min} < 0$, *then*

$$1 \leq \beta < \infty.$$

9. *If* $t_{min} = 0$, *then* $sf^H - g^H \notin C$ *for all* $s \in \mathbb{R}$ *and so* $\beta = \infty$.

It is important to emphasize that the values of t_{min}, t_{max}, α, *and* β *depend on* f, g, H, *and* C. *The result*

$$0 \leq \alpha < 1 \qquad \textit{and} \qquad 1 \leq \beta \leq \infty$$

holds for f^H, g^H *and is not true in general for* f, g.

Remark 2.4.10.2. Regarding Lemma 2.4.10.1 above. Note that $d_H \left(f^H, g^H \right) = \ln(\beta/\alpha)$ where α, β are the smallest, largest "numbers" satisfying $\alpha f^H \leq g^H \leq \beta f^H$. "Numbers" is in quotes, since β might be ∞. The smallest, largest 'numbers' α', β' satisfying $\alpha' f \leq g \leq \beta' f$ will be a fixed multiple of α, β. It will of course be the case that $\beta/\alpha = \beta'/\alpha'$. See (1.51), Theorem 1.9.3.2 (page 84) for details.

The proof of Lemma 2.4.10.1 follows:

Proof. **The relationship between** t_{max} **and** α**.** By Theorem 2.4.8.3 (page 199),

$$f^H + t_{max} \left(g^H - f^H \right) \in C \text{ and } 1 \leq t_{max} < \infty,$$

so

$$\frac{1}{t_{max}}\left(f^H + t_{max}\left(g^H - f^H\right)\right) = \frac{1}{t_{max}}\, f^H + \left(g^H - f^H\right) \tag{2.112}$$
$$= g^H - \left(1 - \frac{1}{t_{max}}\right) f^H \in C.$$

So

$$\left(1 - \frac{1}{t_{max}}\right) \leq \alpha.$$

The map $s \mapsto g^H - sf^H$ parameterizes a line in V passing through g^H (when $s = 0$) and $g^H - \alpha f^H$ when $s = \alpha$. By the definition of α there exists a sequence $\{s_n\}$ with $s_0 = 0 \leq s_1 \leq \ldots \leq \alpha$ such that $g^H - s_n f^H \in C$ and $s_n \to \alpha$. Since C is closed and $s \mapsto g^H - sf^H$ is continuous $g^H - \alpha f^H \in C$. Since C is convex the points $g^H - sf^H \in C$ for all $s \in [0, \alpha]$.

Now suppose that

$$\left(1 - \frac{1}{t_{max}}\right) < \alpha.$$

Then $\exists\ \epsilon > 0$ such that

$$\left(1 - \frac{1}{t_{max} + \epsilon}\right) < \alpha.$$

But then

$$g^H - \left(1 - \frac{1}{t_{max} + \epsilon}\right) f^H \in C. \tag{2.113}$$

Since

$$\left(1 - \frac{1}{t_{max} + \epsilon}\right) = \left(\frac{t_{max} + \epsilon - 1}{t_{max} + \epsilon}\right),$$

we can rewrite the relationship in (2.113) as

$$g^H - \left(\frac{t_{max} + \epsilon - 1}{t_{max} + \epsilon}\right) f^H \in C. \tag{2.114}$$

Since C is closed under positive scaling, multiplying (2.114) by $(t_{max}+\epsilon)$ yields

$$(t_{max}+\epsilon)\, g^H - (t_{max}+\epsilon-1)\, f^H = f^H + (t_{max}+\epsilon)\left(g^H - f^H\right) \in C.$$

But this contradicts the definition of t_{max}. So we must have

$$\left(1 - \frac{1}{t_{max}}\right) = \alpha \tag{2.115}$$

and so, considering (2.112), we have

$$\frac{1}{t_{max}}\ \left(f^H + t_{max}\left(g^H - f^H\right)\right) = g^H - \alpha f^H.$$

By Theorem 2.4.8.3 (page 199)

$$1 \leq t_{max} < \infty$$

and by (2.115), $\left(1 - \frac{1}{t_{max}}\right) = \alpha$, so $0 \leq \alpha < 1$.

The relationship between t_{min} and β.

By Theorem 2.4.8.3 (page 199),

$$f^H + t_{min}\left(g^H - f^H\right) \in C \ \text{ and } \ -\infty < t_{min} \leq 0,$$

so, assuming that $t_{min} < 0$,

$$\begin{aligned}\frac{-1}{t_{min}}\ \left(f^H + t_{min}\left(g^H - f^H\right)\right) &= \frac{-1}{t_{min}}\ f^H + \frac{-t_{min}}{t_{min}}\left(g^H - f^H\right) \\ &= \frac{-1}{t_{min}}\ f^H - \left(g^H - f^H\right) \\ &= \left(1 - \frac{1}{t_{min}}\right)\ f^H - g^H \in C. \end{aligned} \tag{2.116}$$

So, assuming that $t_{min} < 0$,

$$\beta \leq \left(1 - \frac{1}{t_{min}}\right).$$

Consider $B = \{s \in \mathbb{R} : sf^H - g^H \in C\}$. Since C is salient and closed under positive scaling and addition, it follows that $B \subset (0, \infty)$. With the assumption of $t_{min} < 0$ we have $\left(1 - \frac{1}{t_{min}}\right) \in B$, hence B is non-empty, and so $\beta = \inf B$ with $0 \leq \beta \leq \left(1 - \frac{1}{t_{min}}\right)$. Moreover, if $s' > 0$ and $s \in B$, then

$$s'f^H + (sf^H - g^H) \in C$$

(since convex cones are closed under addition). This implies that

$$s' + s \in B \ \forall \ s' > 0 \tag{2.117}$$

provided $s \in B$. If B is non-empty, it must contain a decreasing sequence $s_n \to \beta$. So if B is non-empty, by (2.117), B must contain the intervals $[s_n, \infty)$ and so

$$(\beta, \infty) \subset B \text{ hence } \{sf^H - g^H : s \in (\beta, \infty)\} \subset C. \tag{2.118}$$

The continuous map $\gamma : s \mapsto sf^H - g^H$ parameterizes a line in V that passes through $\beta f^H - g^H$ when $s = \beta$. Moreover, by (2.118), this line is in C for $s > \beta$. Since C is closed $\gamma^{-1}(C) = B$ is closed and since B is bounded below by 0, $\inf B = \beta \in B$. So

$$B = [\beta, \infty) \subset [0, \infty).$$

Now suppose that

$$\begin{aligned}\beta &< \left(1 - \frac{1}{t_{min}}\right) = \left(\frac{t_{min} - 1}{t_{min}}\right) \\ &= \underbrace{\left(\frac{1 - t_{min}}{-t_{min}}\right)}_{>1, \text{ since } t_{min}<0.} .\end{aligned}$$

Then $\exists\ \epsilon > 0$ such that

$$\beta < \underbrace{\left(\frac{1 - t_{min} + \epsilon}{-t_{min} + \epsilon}\right) < \left(\frac{1 - t_{min}}{-t_{min}}\right)}_{2^{nd} \text{ inequality}}. \tag{2.119}$$

To see that the 2^{nd} inequality in (2.119) is true for all $\epsilon > 0$, recall that we are assuming (in this part of the proof) that $t_{min} < 0$; then let $1 - t_{min} = x$ and $-t_{min} = y$ so that $x, y > 0$ and $y < x$. Then

$$\begin{aligned}\frac{x + \epsilon}{y + \epsilon} < \frac{x}{y} &\Leftrightarrow (x + \epsilon)y < x(y + \epsilon) \\ &\Leftrightarrow xy + \epsilon y < xy + \epsilon x \\ &\Leftrightarrow \epsilon y < \epsilon x \\ &\Leftrightarrow y < x.\end{aligned}$$

But then, since $[\beta, \infty) = B$, it follows from the the 1^{st} inequality in (2.119) that $\left(\frac{1-t_{min}+\epsilon}{-t_{min}+\epsilon}\right) \in B$ so that

$$\left(\frac{1 - t_{min} + \epsilon}{-t_{min} + \epsilon}\right) f^H - g^H \in C. \tag{2.120}$$

Since C is closed under positive scaling, multiplying (2.120) by $(-t_{min}+\epsilon)$ yields

$$\begin{aligned}(1-t_{min}+\epsilon)\,f^H-(-t_{min}+\epsilon)\,g^H &= f^H-(t_{min}-\epsilon)\,f^H+(t_{min}-\epsilon)\,g^H\\ &= f^H+(t_{min}-\epsilon)\left(g^H-f^H\right)\in C.\end{aligned}$$

But $(t_{min}-\epsilon)<t_{min}$, which contradicts the definition of t_{min}.

So, when $t_{min}<0$, we must have

$$\beta=\left(\frac{1-t_{min}}{-t_{min}}\right)=\left(1-\frac{1}{t_{min}}\right)>1;$$

and by (2.116), we must have

$$\frac{-1}{t_{min}}\ \left(f^H+t_{min}\left(g^H-f^H\right)\right)=\beta f^H-g^H.$$

Now let us consider the remaining possibility for t_{min}, that $t_{min}=0$.

Suppose that $t_{min}=0$ and that there exists $s'>0$ such that $s'f^H-g^H\in C$ so that $\beta\neq\infty$. (Since C is salient, if such an s' exists, it must be >0.) We will show the existence of such an s' contradicts that $t_{min}=0$.

Since C is closed under positive scaling and addition we have, for all $s>s'$

$$sf^H-g^H=(s-s')f^H+(s'f^H-g^H)\in C. \tag{2.121}$$

Scaling sf^H-g^H in (2.121) by $\frac{1}{s}$ yields

$$f^H-\frac{1}{s}g^H\in C\ \text{ if } s>s'. \tag{2.122}$$

Fixing s and multiplying $f^H-\frac{1}{s}g^H$ by (all real values of) k yields the line

$$k\left(f^H-\frac{1}{s}g^H\right),\ k\in\mathbb{R}$$

which may or may not intersect the line

$$f^H + t\left(g^H - f^H\right),\ t \in \mathbb{R}.$$

The intersection occurs if and only if there exist values of k, t which make

$$k\left(f^H - \frac{1}{s}g^H\right) = f^H + t\left(g^H - f^H\right) \tag{2.123}$$

true. We rewrite (2.123) as:

$$kf^H + \frac{-k}{s}g^H = (1-t)f^H + tg^H.$$

Since f^H, g^H are linearly independent we can equate the coefficients of f^H and g^H:

$$k = 1 - t \tag{2.124}$$

$$\frac{-k}{s} = t. \tag{2.125}$$

Equations (2.124) and (2.125) yield:

$$\begin{aligned} k = 1 - \frac{-k}{s} &\Rightarrow\ sk = s + k \\ &\Rightarrow\ sk - k = s \\ &\Rightarrow\ (s-1)k = s \\ &\Rightarrow\ k = \frac{s}{s-1}. \end{aligned}$$

Let $s > \max\{1, s'\}$. Then by (2.122) $f^H - \frac{1}{s}g^H \in C$ and $k = \frac{s}{s-1} > 0$ so that $k\left(f^H - \frac{1}{s}g^H\right) \in C$ so that the intersection determined by equation (2.123) occurs in

C. That is,

$$f^H + t\left(g^H - f^H\right) \in C \text{ if } s > \max\{1, s'\} \text{ and if } \frac{-k}{s} = t \tag{2.126}$$

however, $k, s > 0$ implies

$$\frac{-k}{s} = t < 0$$

and so (2.126) contradicts our assumption that $t_{min} = 0$ since

$$t_{min} = \min\{t : f^H + t\left(g^H - f^H\right) \in C\}.$$

Hence, if $t_{min} = 0$ there does not exist an s' such that $s'f^H - g^H \in C$ and so $\beta = \infty$. □

2.4.11 Main Theorem for $d_H,\ t_{min},\ t_{max},\ \alpha,\ \beta,\ b_0,\ b_1$

The following Theorem uses the same notation and definitions as given in Lemma 2.4.10.1 (page 204).

Theorem 2.4.11.1. *Let H be a hyperplane in V which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Let $f, g \in C$ be linearly independent; let f^H and g^H be the central projections of f and g onto H. Let t_{max} (resp. t_{min}) be the least (resp. greatest) value of t such that $f^H + t\left(g^H - f^H\right) \in C$. Let*

$$\alpha = \sup\{t \in \mathbb{R} : g^H - tf^H \in C\}$$
$$\beta = \inf\{t \in \mathbb{R} : tf^H - g^H \in C\}$$
$$b_0 = f^H + t_{min}\left(g^H - f^H\right)$$
$$b_1 = f^H + t_{max}\left(g^H - f^H\right).$$

Then

$$\frac{1}{1-t_{min}}\ b_0 = b(f^H/g^H), \qquad 0 < \frac{1}{1-t_{min}} \leq 1, \qquad [b_0] = \left[b\left(f^H/g^H\right)\right]$$
$$\frac{1}{t_{max}}\ b_1 = b\left(g^H/f^H\right), \qquad 0 < \frac{1}{t_{max}} \leq 1, \qquad [b_1] = \left[b\left(g^H/f^H\right)\right]$$

and the following 8 results hold:

1. $$t_{max} = \frac{d_V\left(f^H, b_1\right)}{d_V(f^H, g^H)}, \qquad 1 \leq t_{max} < \infty.$$

2. $$t_{min} = -\frac{d_V\left(f^H, b_0\right)}{d_V(f^H, g^H)}, \qquad -\infty < t_{min} \leq 0.$$

3. *If* $t_{min} < 0$, *then*

$$\beta = 1 - \frac{1}{t_{min}} = \frac{t_{min} - 1}{t_{min}}, \qquad 1 < \beta < \infty.$$

If $t_{min} = 0$, *then*

$$\beta = \infty.$$

4. $$\alpha = 1 - \frac{1}{t_{max}} = \frac{t_{max} - 1}{t_{max}}, \qquad 0 \leq \alpha < 1.$$

5. $$0 \leq \alpha < 1 < \beta \leq \infty \quad \textit{and} \quad 1 < \frac{\beta}{\alpha} \leq \infty.$$

If we take

$$\frac{t_{min} - 1}{t_{min}} = \infty \quad \textit{if} \quad t_{min} = 0 \quad \textit{and}$$
$$\frac{t_{max}}{t_{max} - 1} = \infty \quad \textit{if} \quad t_{max} = 1,$$

we have

$$\frac{\beta}{\alpha} = \frac{t_{min} - 1}{t_{min}} \frac{t_{max}}{t_{max} - 1}.$$

6.

$$\frac{\beta}{\alpha} = \frac{d_V\left(b_0, g^H\right)}{d_V\left(b_0, f^H\right)} \frac{d_V\left(b_1, f^H\right)}{d_V\left(b_1, g^H\right)}$$

7.

$$\begin{aligned} d_H\left(f^H, g^H\right) &= \ln\left(\frac{\beta}{\alpha}\right) \\ &= \ln\left(\frac{t_{min} - 1}{t_{min}} \frac{t_{max}}{t_{max} - 1}\right) \\ &= \ln\left(\frac{d_V\left(b_0, g^H\right)}{d_V\left(b_0, f^H\right)} \frac{d_V\left(b_1, f^H\right)}{d_V\left(b_1, g^H\right)}\right). \end{aligned}$$

8. $d_V(b_0, b_1) = (t_{max} - t_{min})\, d_V\left(f^H, g^H\right)$.

9. $b_0^H = b_0$ *and* $b_1^H = b_1$.

10. b_0, b_1 *are a pair of ends for* f^H, g^H *(and equivalently, for* f, g*).*

Proof. See Figure 2.4 (page 223). Much of this theorem is a consequence of Lemma 2.4.10.1 (page 204).

First we prove

$$\frac{1}{t_{max}}\, b_1 = b\left(g^H / f^H\right) \qquad \text{and} \qquad [b_1] = \left[b\left(g^H / f^H\right)\right]$$

In this theorem we have defined b_1 as

$$b_1 = f^H + t_{max}\left(g^H - f^H\right). \tag{2.127}$$

By part 1 of Lemma 2.4.10.1 (page 204) we have $1 \leq t_{max} < \infty$. Dividing (2.127) by

t_{max} yields

$$\frac{1}{t_{max}}\ b_1\ =\ \frac{1}{t_{max}}\ \left(f^H + t_{max}\left(g^H - f^H\right)\right). \tag{2.128}$$

Part 2 of Lemma 2.4.10.1 (page 204) give us

$$\frac{1}{t_{max}}\ \left(f^H + t_{max}\left(g^H - f^H\right)\right) =\ g^H - \alpha f^H. \tag{2.129}$$

The definition of α, see [19], given in this theorem (and in Lemma 2.4.10.1 (page 204)) is the same as the definition of $m(y/x)$, given in Definition 1.3.0.8 (page 30), and so we have the equality

$$\alpha = \sup\{t \in \mathbb{R} : g^H - tf^H \in C\} = m\left(g^H/f^H\right). \tag{2.130}$$

(2.130) implies

$$g^H - \alpha f^H = g^H - m\left(g^H/f^H\right) f^H. \tag{2.131}$$

The definition of $b(y/x)$, given in Definition 1.3.0.9 (page 30) is

$$y - m(y/x)\ x = b(y/x), \tag{2.132}$$

which immediately implies

$$g^H - m\left(g^H/f^H\right) f^H =\ b\left(g^H/f^H\right). \tag{2.133}$$

The four telescoping equalities (2.128), (2.129), (2.131), and (2.133) give us

$$\frac{1}{t_{max}}\ b_1 =\ b\left(g^H/f^H\right). \tag{2.134}$$

[19]The definition of $\alpha = \sup\{t \in \mathbb{R} : g^H - tf^H \in C\}$ given here, in Theorem 2.4.11.1, matches the definition of α given in the "$\alpha f \leq g \leq \beta f$ definition of $d_H(f, g)$" discussed in Section 1.9 (page 80). Of course in Theorem 2.4.11.1 we are working with f^H and g^H.

By Part 1 of Lemma 2.4.10.1 (page 204) we know that $1 \leq t_{max} < \infty$. But then $0 < \frac{1}{t_{max}} \leq 1$ and Equation (2.134) implies $[b_1] = \left[b\left(g^H/f^H\right)\right]$.

Next we prove:

$$\frac{1}{1-t_{min}}\, b_0 = b(f^H/g^H) \qquad \text{and} \qquad [b_0] = \left[b\left(f^H/g^H\right)\right].$$

Recall $t_{min} = \inf\{t \mid f^H + t(g^H - f^H) \in C\}$ and that by part 5 of Lemma 2.4.10.1 (page 204), $-\infty < t_{min} \leq 0$.

Case 1. $t_{min} = 0$.

Suppose there exists an $s > 0$ such that $f^H - sg^H \in C$. Then, since C is closed under positive scaling and addition,

$$\begin{aligned}
\frac{1}{s}\, f^H - g^H &\in C \\
\frac{1}{s}\, f^H - g^H + f^H &\in C \\
f^H - sg^H + sf^H &\in C \\
f^H - s(g^H - f^H) &\in C \\
f^H + (-s)(g^H - f^H) &\in C
\end{aligned}$$

implies $t_{min} \leq (-s)$. So if $t_{min} = 0$, then there does not exist an $s > 0$ such that $f^H - sg^H \in C$. So $m(f^H/g^H) = 0$ (see [20]) which implies

$$b\left(f^H/g^H\right) = f^H = b_0; \tag{2.135}$$

see [21].

Case 2. $t_{min} < 0$.

[20] Definition 1.3.0.8 (page 30): $m(y/x) = \sup\{t \in \mathbb{R} : y - tx \in C\}$.
[21] Definition 1.3.0.9 (page 30): $y - m(y/x)\, x = b(y/x)$.

By part 8 of Lemma 2.4.10.1 (page 204), if $t_{min} < 0$ then

$$1 \leq \beta < \infty.$$

So β is positive finite and is the smallest number for which $\beta f^H - g^H \in C$. C is closed under positive scaling so $f^H - \frac{1}{\beta}\, g^H \in C$, which in turn implies $\frac{1}{\beta}$ is the largest number t for which $f^H - t\, g^H \in C$. I.e. $m(f^H/g^H) = \frac{1}{\beta}$ and

$$\underbrace{b\left(f^H/g^H\right) = f^H - m(f^H/g^H)\, g^H}_{\text{Definition of } b(f^H/g^H)} \;\; = \;\; f^H - \frac{1}{\beta}\, g^H. \tag{2.136}$$

By Part 6 of Lemma 2.4.10.1 (page 204), if $t_{min} < 0$, then

$$\frac{-1}{t_{min}}\;\left(f^H + t_{min}\left(g^H - f^H\right)\right) = \beta f^H - g^H. \tag{2.137}$$

Since $b_0 = f^H + t_{min}\left(g^H - f^H\right)$ we can rewrite (2.137) as

$$\frac{-1}{t_{min}}\, b_0 \;\; = \;\; \beta f^H - g^H. \tag{2.138}$$

Since $\beta \neq 0$ we can divide both sides of (2.137) and then use (2.136) to get

$$\frac{-1}{t_{min}\;\beta}\, b_0 = f^H - \frac{1}{\beta} g^H = b(f^H/g^H). \tag{2.139}$$

By part 7 of Lemma 2.4.10.1 (page 204), if $t_{min} < 0$, then

$$\beta = \frac{t_{min} - 1}{t_{min}},$$

so

$$\frac{-1}{t_{min}\;\beta} = \frac{-1}{t_{min}\;\frac{t_{min}-1}{t_{min}}} = \frac{-1}{t_{min} - 1} = \frac{1}{1 - t_{min}}. \tag{2.140}$$

(2.140) allows us to write (2.139) as

$$\frac{1}{1-t_{min}}\ b_0 = b(f^H/g^H). \tag{2.141}$$

Equality (2.141) was derived under the assumption that $t_{min} < 0$. However, if we plug $t_{min} = 0$ into (2.141) we get $b_0 = b(f^H/g^H)$ which is the correct result for $t_{min} = 0$; see Equality (2.135). So Equality (2.141) holds for all possible values of t_{min}; i.e. $t_{min} \leq 0$.

Moveover, since $t_{min} \leq 0$ we have

$$0 < \frac{1}{1-t_{min}} \leq 1$$

and so

$$[b_0] = [b(f^H/g^H)].$$

Now we prove the remaining parts of this theorem:

1. By Lemma 2.4.10.1 (page 204) we have $1 \leq t_{max} < \infty$, so

$$\begin{aligned} d_V\left(b_1, f^H\right) &= d_V\left(f^H + t_{max}\left(g^H - f^H\right), f^H\right) \\ &= \left\|\left(f^H + t_{max}\left(g^H - f^H\right)\right) - f^H\right\|_V \\ &= \left\|t_{max}\left(g^H - f^H\right)\right\|_V \\ &= t_{max}\left\|g^H - f^H\right\|_V \\ &= t_{max} d_V\left(f^H, g^H\right) \\ \Rightarrow t_{max} &= \frac{d_V\left(f^H, b_1\right)}{d_V\left(f^H, g^H\right)}. \end{aligned}$$

2. By Lemma 2.4.10.1 (page 204) we have $-\infty < t_{min} \leq 0$, so

$$\begin{aligned} d_V\left(b_0, f^H\right) &= d_V\left(f^H + t_{min}\left(g^H - f^H\right), f^H\right) \\ &= \left\|\left(f^H + t_{min}\left(g^H - f^H\right)\right) - f^H\right\|_V \\ &= \left\|t_{min}\left(g^H - f^H\right)\right\|_V \\ &= -t_{min}\left\|g^H - f^H\right\|_V \\ &= -t_{min} d_V\left(f^H, g^H\right) \\ \Rightarrow t_{min} &= -\frac{d_V\left(f^H, b_0\right)}{d_V\left(f^H, g^H\right)}. \end{aligned}$$

3. and *4.* are proven in Lemma 2.4.10.1 (page 204).

5. In Lemma 2.4.10.1 (page 204), we proved that

$$0 \leq \alpha = \frac{t_{max} - 1}{t_{max}} < 1 \quad \text{and} \quad 1 < \beta = \frac{t_{min} - 1}{t_{min}} \leq \infty,$$

and so the rest of Part *5* follows.

6.

$$\begin{aligned} \frac{\beta}{\alpha} &= \frac{t_{min} - 1}{t_{min}} \frac{t_{max}}{t_{max} - 1} \\ &= \frac{-\frac{d_V\left(f^H, b_0\right)}{d_V\left(f^H, g^H\right)} - 1}{-\frac{d_V\left(f^H, b_0\right)}{d_V\left(f^H, g^H\right)}} \frac{\frac{d_V\left(f^H, b_1\right)}{d_V\left(f^H, g^H\right)}}{\frac{d_V\left(f^H, b_1\right)}{d_V\left(f^H, g^H\right)} - 1} \\ &= \frac{-d_V\left(f^H, b_0\right) - d_V\left(f^H, g^H\right)}{-d_V\left(f^H, b_0\right)} \frac{d_V\left(f^H, b_1\right)}{d_V\left(f^H, b_1\right) - d_V\left(f^H, g^H\right)}. \end{aligned} \tag{2.142}$$

The following remarks will allow us to simplify (2.142).

f^H, g^H, b_0, b_1 are collinear. As t increases, the parametrization

$$t \mapsto f^H + t(g^H - f^H)$$

shows that f^H, g^H, b_0, b_1 are reached on the line $\{f^H + t(g^H - f^H) \in V : t \in \mathbb{R}\}$ in the order b_0, f^H, g^H, b^1 since

$$t_{min} \mapsto b_0 \qquad 0 \mapsto f^H \qquad 1 \mapsto g^H \qquad t_{max} \mapsto b_1$$

and by Lemma 2.4.10.1 (page 204)

$$t_{min} \ \leq \ 0 \ < \ 1 \ \leq \ t_{max}.$$

This indicates

$$d_V\left(b_0, f^H\right) + d_V\left(f^H, g^H\right) = d_V\left(b_0, g^H\right) \tag{2.143}$$

$$d_V\left(f^H, b_1\right) - d_V\left(f^H, g^H\right) = d_V\left(g^H, b_1\right) \tag{2.144}$$

which the following argument explicitly proves. First we prove (2.143):

$$\begin{aligned} d_V\left(b_0, f^H\right) &= \left\|\left(f^H + t_{min}(g^H - f^H)\right) - \left(f^H\right)\right\|_V \\ &= -t_{min}\left\|g^H - f^H\right\|_V \quad \text{since} \quad -t_{min} = |t_{min}| \\ d_V\left(f^H, g^H\right) &= \left\|g^H - f^H\right\|_V \end{aligned}$$

and so

$$\begin{aligned} d_V\left(b_0, f^H\right) + d_V\left(f^H, g^H\right) &= (1 - t_{min})\left\|g^H - f^H\right\|_V \\ &= \left\|(1 - t_{min})\left(g^H - f^H\right)\right\|_V \\ &= \left\|\left(g^H - f^H\right) - t_{min}\left(g^H - f^H\right)\right\|_V \\ &= \left\|\left(g^H\right) - \left(f^H + t_{min}\left(g^H - f^H\right)\right)\right\|_V \\ &= d_V\left(g^H, b_0\right). \end{aligned}$$

Hence, Equation (2.143) is proven. Next we prove (2.144):

$$\begin{aligned} d_V\left(f^H, b_1\right) &= \left\|\left(f^H\right) - \left(f^H + t_{max}\left(g^H - f^H\right)\right)\right\|_V \\ &= t_{max}\left\|g^H - f^H\right\|_V \quad \text{since} \quad \underbrace{t_{max}}_{\geq 1} = |t_{max}| \\ d_V\left(f^H, g^H\right) &= \left\|g^H - f^H\right\|_V \end{aligned}$$

and so

$$\begin{aligned} d_V\left(f^H, b_1\right) - d_V\left(f^H, g^H\right) &= (t_{max} - 1)\left\|g^H - f^H\right\|_V \\ &= \left\|(t_{max} - 1)\left(g^H - f^H\right)\right\|_V \\ &= \left\|\left(f^H + t_{max}\left(g^H - f^H\right)\right) - \left(g^H\right)\right\|_V \\ &= d_V\left(b_1, g^H\right). \end{aligned}$$

Hence, Equation (2.144) is proven. We rewrite (2.142) using Equations (2.143) and (2.144):

$$\begin{aligned} \frac{\beta}{\alpha} &= \frac{-d_V\left(b_0, g^H\right)}{-d_V\left(b_0, f^H\right)} \frac{d_V\left(b_1, f^H\right)}{d_V\left(b_1, g^H\right)} \\ &= \frac{d_V\left(b_0, g^H\right)}{d_V\left(b_0, f^H\right)} \frac{d_V\left(b_1, f^H\right)}{d_V\left(b_1, g^H\right)}. \end{aligned}$$

7. is an immediate consequence of Parts *5* and *6* of this theorem and Definition 2.4.9.1 (page 202).

8. $d_V(b_0, b_1) = \|b_1 - b_0\|_V = (t_{max} - t_{min})\, d_V\left(f^H, g^H\right)$.

9. Lemma 2.4.3.1 (page 190) implies that both b_0 and b_1 are in H. We are assuming that H intersects each 0-ray of $C \setminus \{0\}$ exactly once. So we must have $b_0^H = b_0$ and $b_1^H = b_1$.

10. We have already proved in (2.141) (page 218) and (2.134) (page 215) that

$$\frac{1}{1-t_{min}}\ b_0 = b\left(f^H/g^H\right) \qquad \text{and} \qquad \frac{1}{t_{max}}\ b_1 = b\left(g^H/f^H\right)$$

with $0 < \frac{1}{1-t_{min}}, \frac{1}{t_{max}} \leq 1$. So the pair b_0, b_1 satisfies part 1 of Proposition 1.4.2.1 (page 46) and so are ends for f^H, g^H.

$\alpha f^H = f$ and $\beta g^H = g$ for some $\alpha, \beta > 0$. So, by Proposition 1.4.3.3 (page 47), b_0, b_1 will also be ends for f, g. □

2.4.12 Theorem relating d_H to d_V

The following theorem is important because it allows us to readily calculate d_H provided a hyperplane exists which intersects each 0-ray in $C \setminus \{0\}$.

Theorem 2.4.12.1. *This theorem is a continuation of Theorem 2.4.11.1 (page 212).*

In particular, C is a closed, convex, salient, pointed-by-the-origin cone in Banach Space V. Let H be a hyperplane in V which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Let $f, g \in C$ be linearly independent; let f^H and g^H be the central projections of f and g onto H. Let t_{max} (resp. t_{min}) be the least (resp. greatest) value of t such that $f^H + t\left(g^H - f^H\right) \in C$. Let

$$b_0 = f^H + t_{min}\left(g^H - f^H\right)$$
$$b_1 = f^H + t_{max}\left(g^H - f^H\right)$$

and let $D = d_V(b_0, b_1)$ and $c(t) = b_0 + \frac{t}{D}(b_1 - b_0)$. Then the following hold:

1.

$$c(t) = \left(1 - \frac{t}{D}\right) b_0 + \frac{t}{D}\ b_1.$$

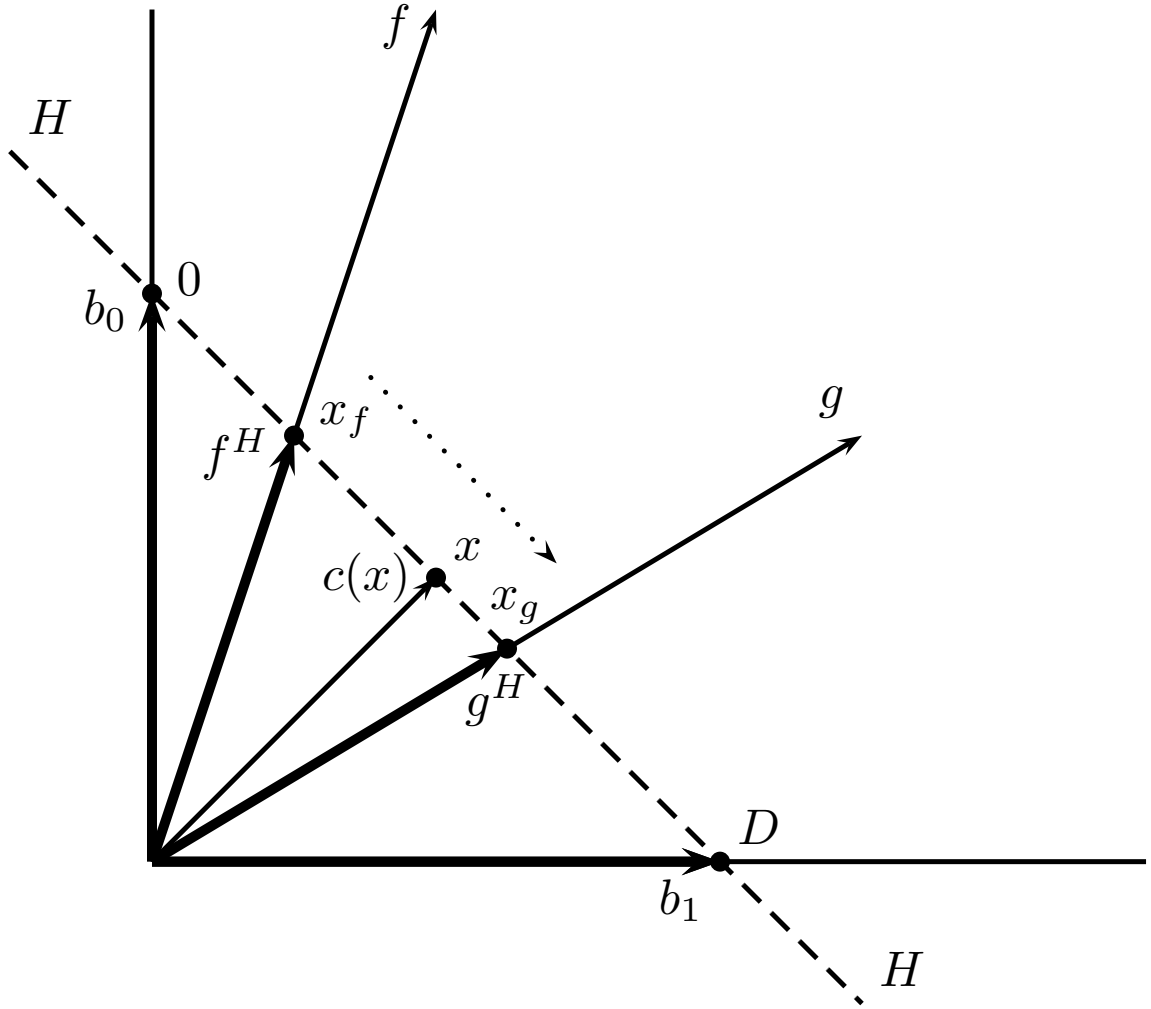


Figure 2.4: The cone $\mathbb{R}^2_{\geq 0} = C$. The hyperplane H intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. f^H and g^H are the projections of f and g into H. In this example, the line $t \mapsto f^H + t\left(g^H - f^H\right)$ parameterizes the hyperplane H. The dotted arrow shows the direction of the parametrization as t increases. $b_0 = f^H + t_{min}\left(g^H - f^H\right)$ and $b_1 = f^H + t_{max}\left(g^H - f^H\right)$. $D = d_V(b_0, b_1)$. $c : x \in [0, D] \subset \mathbb{R} \mapsto \left(1 - \frac{x}{D}\right) b_0 + \frac{x}{D}\ b_1 \in H \bigcap C$. c is an isometry. The dotted arrow shows the direction of increasing x.

2. *c is an isometry of $\mathbb{R}$ to V.*

3. *Let $\overline{b_0 b_1}$ be the line segment connecting $b_0 f$ to b_1. As sets*

$$\overline{b_0 b_1} = \{f^H + t(g^H - f^H) : t \in [t_{min}, t_{max}]\} = c([0, D]). \tag{2.145}$$

$$\{f^H + t(g^H - f^H) : t \in \mathbb{R}\} = c(\mathbb{R}). \tag{2.146}$$

4. *$Span(f, g) \cap C \cap H = c([0, D])$.*

5. *If $x, y \in [0, D]$ with $x < y$ and if*

$$s_{min} = \inf\{s \in \mathbb{R} : c(x) + s(c(y) - c(x)) \in C\}$$
$$s_{max} = \sup\{s \in \mathbb{R} : c(x) + s(c(y) - c(x)) \in C\},$$

then

$$s_{min} = \frac{-x}{y - x}$$
$$s_{max} = \frac{D - x}{y - x}$$

and

$$c(x) + s_{min}(c(y) - c(x)) = b_0$$
$$c(x) + s_{max}(c(y) - c(x)) = b_1.$$

6. *If $x \neq y$ then $c(x), c(y)$ are linearly independent. If $x, y \in (0, D)$ and $x < y$ then*

$$d_H(c(x), c(y)) = \ln\left(\frac{d_V(b_0, c(y))}{d_V(b_0, c(x))} \frac{d_V(b_1, c(x))}{d_V(b_1, c(y))}\right) = \ln\left(\frac{y}{x} \frac{D - x}{D - y}\right) \in (0, \infty). \tag{2.147}$$

If $x, y \in (0, D)$ and $y < x$ then

$$d_H(c(x), c(y)) = \ln\left(\frac{d_V\left(b_0, c(x)\right)}{d_V\left(b_0, c(y)\right)} \frac{d_V\left(b_1, c(y)\right)}{d_V\left(b_1, c(x)\right)}\right) = \ln\left(\frac{x}{y} \frac{D-y}{D-x}\right) \in (0, \infty). \tag{2.148}$$

7. *If $x \neq y$ and $x = 0$ and/or $y = D$ then $d_H(c(x), c(y)) = \infty$.*

8. *Let*

$$\frac{-t_{min}}{t_{max} - t_{min}} D = x_f \quad \text{and} \quad \frac{1 - t_{min}}{t_{max} - t_{min}} D = x_g.$$

Then $x_f, x_g \in [0, D]$ are the unique real numbers such that

$$c(x_f) = f^H \quad \text{and} \quad c(x_g) = g^H.$$

Also, $0 \leq x_f < x_g \leq D$ and

$$0 < x_f < x_g < D \quad \text{if and only if} \quad 0 < d_H(f^H, g^H) < \infty.$$

Proof. 1. Is trivial.

2. Plugging $x_1, x_2 \in \mathbb{R}$ into

$$c(x) = \left(1 - \frac{x}{D}\right) b_0 + \frac{x}{D}\ b_1$$

and doing the obvious, we get

$$\begin{aligned}
d_V(c(x_1), c(x_2)) = \\
d_V\left(\left(1-\frac{x_1}{D}\right)b_0 + \frac{x_1}{D}\ b_1,\ \left(1-\frac{x_2}{D}\right)b_0 + \frac{x_2}{D}\ b_1\right) = \\
\left\|\left(\left(1-\frac{x_1}{D}\right)b_0 + \frac{x_1}{D}\ b_1\right) - \left(\left(1-\frac{x_2}{D}\right)b_0 + \frac{x_2}{D}\ b_1\right)\right\|_V = \\
\left\|\frac{x_2-x_1}{D}b_0 - \frac{x_2-x_1}{D}b_1\right\|_V = \\
\left|\frac{x_2-x_1}{D}\right| \|b_0 - b_1\|_V = \\
\left|\frac{x_2-x_1}{D}\right| D = \\
|x_2 - x_1| = d_{\mathbb{R}}(x_1, x_2).
\end{aligned}$$

So c is an isometry from $\mathbb{R}$ to the line $c(\mathbb{R}) \subset V$.

3.

$$\begin{aligned}
c(t) &= \left(1-\frac{t}{D}\right)b_0 + \frac{t}{D}\ b_1 \\
&= \left(1-\frac{t}{D}\right)\left(f^H + t_{min}(g^H - f^H)\right) + \frac{t}{D}\left(f^H + t_{max}(g^H - f^H)\right) \\
&= f^H + \left(1-\frac{t}{D}\right)t_{min}\left(g^H - f^H\right) + \frac{t}{D}\ t_{max}\left(g^H - f^H\right) \\
&= f^H + \left(\left(1-\frac{t}{D}\right)t_{min} + \frac{t}{D}\ t_{max}\right)\left(g^H - f^H\right) \\
&= f^H + \left(t_{min} + \frac{t}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right). \qquad (2.149)
\end{aligned}$$

As t goes from 0 to D (resp. from $-\infty$ to ∞), the coefficient of $\left(g^H - f^H\right)$ in (2.149) goes monotonically from t_{min} to t_{max} (resp. from $-\infty$ to ∞). So (2.145) (resp. (2.146)) is proven.

4. By Equality (2.110) of Theorem 2.4.8.3 (page 199), and Equality (2.145) of this

theorem (proven in Part 3.) we have

$$H \cap \mathrm{Span}(f,g) \cap C = \{f^H + t(g^H - f^H) : t \in [t_{min}, t_{max}]\} = c([0,D]). \tag{2.150}$$

5. From (2.149)

$$c(t) = f^H + \left(t_{min} + \frac{t}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right)$$

so

$$\begin{aligned}
& c(x) + s\,(c(y) - c(x)) && (2.151)\\
&= (1-s)\ c(x) + s\ c(y)\\
&= (1-s)\left(f^H + \left(t_{min} + \frac{x}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right)\right) +\\
&\qquad s\left(f^H + \left(t_{min} + \frac{y}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right)\right)\\
&= f^H + (1-s)\left(t_{min} + \frac{x}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right) +\\
&\qquad s\left(t_{min} + \frac{y}{D}\ (t_{max} - t_{min})\right)\left(g^H - f^H\right)\\
&= f^H + \left(t_{min} + \left((1-s)\frac{x}{D} + s\frac{y}{D}\right)\ (t_{max} - t_{min})\right)\left(g^H - f^H\right)\\
&= f^H + \underbrace{\left(t_{min} + (x + s(y-x))\ \frac{t_{max} - t_{min}}{D}\right)}_{\text{monotonically increasing with respect to } s \text{ because } x<y}\ \left(g^H - f^H\right). && (2.152)
\end{aligned}$$

So, by the above algebra, (2.151) to (2.152), we must have that

$$\begin{aligned}
s_{min} &= \inf\{s \in \mathbb{R} :\ c(x) + s\,(c(y) - c(x)) \in C\}\\
&= \inf\left\{ s \in \mathbb{R} \ \middle|\ f^H + \underbrace{\left(t_{min} + (x + s(y-x))\ \frac{t_{max} - t_{min}}{D}\right)}_{\text{monotonically increasing with respect to } s \text{ because } x<y}\ \left(g^H - f^H\right) \in C \right\}\\
&= \text{ that unique } s \text{ such that } t_{min} = t_{min} + (x + s(y-x))\ \frac{t_{max} - t_{min}}{D}.
\end{aligned}$$

Hence

$$0 = x + s_{min}(y - x)$$
$$s_{min} = \frac{-x}{y - x}.$$

In a similar fashion, s_{max} will be that unique s such that

$$t_{max} = t_{min} + (x + s(y - x)) \ \frac{t_{max} - t_{min}}{D}. \quad \text{Hence,}$$
$$D = x + s_{max}(y - x)$$
$$s_{max} = \frac{D - x}{y - x}.$$

From (2.151) and (2.152) it is obvious that

$$b_0 = c(x) + s_{min}\left(c(y) - c(x)\right) \tag{2.153}$$
$$b_1 = c(x) + s_{max}\left(c(y) - c(x)\right).$$

6. According to Lemma 2.4.8.2 (page 199) if $v, w \in H$ are not equal, they are linearly independent. By Part 5, $c(x), c(y) \in H$. By Part 2, c is an isometry, so if $x \neq y$ then $c(x) \neq c(y)$. So if $x \neq y$ then $c(x), c(y)$ are linearly independent.

Let us suppose that $0 \leq x < y \leq D$ so that $c(x), c(y) \in C$ are linearly independent If we let

$$b_0' = c(x)^H + s_{min}\left(c(y)^H - c(x)^H\right)$$
$$b_1' = c(x)^H + s_{max}\left(c(y)^H - c(x)^H\right).$$

Then by Part 7 of Theorem 2.4.11.1 (page 212), which requires linear independence,

we have

$$d_H\left(c(x)^H, c(y)^H\right) = \ln\left(\frac{d_V\left(b_0', c(y)^H\right)}{d_V\left(b_0', c(x)^H\right)}\ \frac{d_V\left(b_1', c(x)^H\right)}{d_V\left(b_1', c(y)^H\right)}\right) \tag{2.154}$$

As a consequence of Part 5 (2.150), $c(x)^H = c(x)$ and $c(y)^H = c(y)$; as a consequence of Part 4 (2.153), $b_0' = b_0$ and $b_1' = b_1$. Also, $c(0) = b_0$ and $c(D) = b_1$. So, with these substitutions, and using that c is an isometry, proved in Part 2 of this theorem, (2.154) becomes

$$\begin{aligned} d_H\left(c(x), c(y)\right) &= \ln\left(\frac{d_V\left(b_0, c(y)\right)}{d_V\left(b_0, c(x)\right)}\ \frac{d_V\left(b_1, c(x)\right)}{d_V\left(b_1, c(y)\right)}\right) \\ &= \ln\left(\frac{d_V\left(c(0), c(y)\right)}{d_V\left(c(0), c(x)\right)}\ \frac{d_V\left(c(D), c(x)\right)}{d_V\left(c(D), c(y)\right)}\right) \\ &= \ln\left(\frac{|0-y|}{|0-x|}\ \frac{|D-x|}{D-y|}\right) \\ &= \ln\left(\frac{y}{x}\ \frac{D-x}{D-y}\right). \end{aligned} \tag{2.155}$$

We are assuming that $x, y \in (0, D)$ and that $x < y$ so $1 < \frac{D-x}{D-y}, \frac{y}{x} < \infty$. Hence (2.155) implies that $0 < d_H\left(c(x), c(y)\right) < \infty$. So (2.147) is proven.

If $x > y$, if we switch x with y and note that d_H is symmetric, we get the (2.148), which finishes our proof of Part 6.

7. If $x \neq y$ and $x = 0$ and/or $y = D$, simply plugging these values into (2.155) yields $\ln(\infty) = \infty$ without any indeterminacy.

8. Let $x_f, x_g \in [0, D]$ be the unique real numbers for which

$$\left(1 - \frac{x_f}{D}\right) b_0 + \frac{x_f}{D}\ b_1 = f^H \quad \text{and} \quad \left(1 - \frac{x_g}{D}\right) b_0 + \frac{x_g}{D}\ b_1 = g^H.$$

From the definitions of b_0 and b_1 it is clear that $x_f < x_g$ and that $x_f, x_g \in [0, D]$.

That said, we show this explicitly and derive formulas for x_f and x_g. Recall

$$t_{min} = \min\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\}, \qquad b_0 = f^H + t_{min}(g^H - f^H)$$
$$t_{max} = \max\{t \in \mathbb{R} : f^H + t(g^H - f^H) \in C\}, \qquad b_1 = f^H + t_{max}(g^H - f^H)$$

so

$$\begin{aligned}
f^H &= \left(1 - \frac{x_f}{D}\right) b_0 + \frac{x_f}{D}\, b_1 \\
&= \left(1 - \frac{x_f}{D}\right) \left(f^H + t_{min}\left(g^H - f^H\right)\right) \\
&\quad + \left(\frac{x_f}{D}\right) \left(f^H + t_{max}\left(g^H - f^H\right)\right) \\
&= f^H + \left(1 - \frac{x_f}{D}\right) t_{min}\left(g^H - f^H\right) \\
&\quad + \left(\frac{x_f}{D}\right) t_{max}\left(g^H - f^H\right) \\
\Rightarrow \quad 0 &= \left(1 - \frac{x_f}{D}\right) t_{min} + \left(\frac{x_f}{D}\right) t_{max} \\
&= t_{min} - \frac{x_f}{D} t_{min} + \frac{x_f}{D} t_{max} \\
\Rightarrow \quad -t_{min} &= \frac{x_f}{D}\left(t_{max} - t_{min}\right) \\
\Rightarrow \quad \frac{-t_{min}}{t_{max} - t_{min}} D &= x_f.
\end{aligned}$$

Similarly for g^H:

$$\begin{aligned}
g^H &= \left(1 - \frac{x_g}{D}\right) b_0 + \frac{x_g}{D}\ b_1 \\
&= \left(1 - \frac{x_g}{D}\right) \left(f^H + t_{min}\left(g^H - f^H\right)\right) \\
&\quad + \left(\frac{x_g}{D}\right) \left(f^H + t_{max}\left(g^H - f^H\right)\right) \\
&= f^H + \left(1 - \frac{x_g}{D}\right) t_{min}\left(g^H - f^H\right) + \left(\frac{x_g}{D}\right) t_{max}\left(g^H - f^H\right) \\
\Rightarrow \quad g^H - f^H &= \left(\left(1 - \frac{x_g}{D}\right) t_{min} + \left(\frac{x_g}{D}\right) t_{max}\right)\left(g^H - f^H\right) \\
\Rightarrow \quad 1 &= \left(\left(1 - \frac{x_g}{D}\right) t_{min} + \left(\frac{x_g}{D}\right) t_{max}\right) \\
&= t_{min} - \frac{x_g}{D} t_{min} + \frac{x_g}{D} t_{max} \\
\Rightarrow \quad 1 - t_{min} &= \frac{x_g}{D}\left(t_{max} - t_{min}\right) \\
\Rightarrow \quad \frac{1 - t_{min}}{t_{max} - t_{min}} D &= x_g.
\end{aligned}$$

By Theorem 2.4.8.3 (page 199)

$$-\infty \ < \ t_{min} \ \leq \ 0 \ < \ 1 \ \leq t_{max} \ \leq \ \infty. \tag{2.156}$$

We subtract t_{min} from (2.156) and then multiply the result by

$$\frac{D}{t_{max} - t_{min}}$$

to get

$$\begin{aligned}
t_{min} - t_{min} \leq \quad & 0 - t_{min} \quad < 1 - t_{min} \quad \leq t_{max} - t_{min} \\
0 \leq \quad & -t_{min} \quad < 1 - t_{min} \quad \leq t_{max} - t_{min} \quad \Rightarrow \\
0 \leq \quad & \underbrace{\frac{-t_{min}}{t_{max} - t_{min}} D}_{x_f} < \underbrace{\frac{1 - t_{min}}{t_{max} - t_{min}} D}_{x_g} \ \leq D.
\end{aligned} \tag{2.157}$$

So we have proven that

$$0 \leq x_f < x_g \leq D.$$

If $d_H(f,g) < \infty$, then Part 5 of Theorem 2.4.11.1 (page 212) implies $t_{min} < 0$ and $1 < t_{max}$. If we plug $t_{min} < 0$ and $t_{max} > 1$ into (2.157) we get

$$0 < x_f < x_g < D.$$

Conversely, if $0 < x_f < x_g < D$, then (2.155) with $x = x_f$ and $y = x_g$ yields $1 < d_H(c(x_f), c(x_g)) = d_H(f^H, g^H) < \infty$. □

2.5 If $K \geq D/4$ then $d_V(f^H, g^H) < K d_H(f,g)$

Lemma 2.5.0.2 (page 232), immediately below, is completely original, as far as I know. It is allows a comparison of $d_H(f,g)$ to $d_V(f^H, g^H)$ when there exists a hyperplane H which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Lemma 2.5.0.2 allows us to concretely compare convergence under d_H with convergence in the hyperplane H under d_V. Lemma 2.5.0.2 also has a useful application to the theory of differential equations.

Lemma 2.5.0.2. *This Lemma is a continuation of Theorem 2.4.11.1 (page 212) and Theorem 2.4.12.1 (page 222). See those theorems for the full set of assumptions and notation.*

Let H be a hyperplane which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once.

Let $f, g \in C$ be linearly independent and

$$b_0 = f^H + t_{min}\left(g^H - f^H\right)$$
$$b_1 = f^H + t_{max}\left(g^H - f^H\right).$$

Then

1. $b_0, b_1 \in H \cap C \cap Span(f,g) \setminus \{0\}$ *are a pair of ends for* f, g.

2. *Let* $D = d_V(b_0, b_1)$. *If* $K \geq D/4$ *and* $f', g' \in Span(f,g) \cap C$ *are linearly independent then*

$$d_V(f'^H, g'^H) \ < \ \frac{D}{4}\, d_H(f', g') \ \leq \ K d_H(f', g'). \tag{2.158}$$

In particular,

$$d_V(f^H, g^H) < \frac{D}{4}\, d_H(f, g), \tag{2.159}$$

and (regardless of K*)*

$$\frac{4}{D} < \frac{d_H(f', g')}{d_V(f'^H, g'^H)}. \tag{2.160}$$

3. $K = D/4$ *is the smallest* K *for which (2.158) will be true in the following sense. If* $K < D/4$ *then there exists a linearly independent pair* $f', g' \in Span(f,g) \cap C$ *such that* $d_V(f'^H, g'^H) > K d_H(f', g')$.

4. *Let* $f' \in Span(f,g) \cap C \setminus \{0\}$ *and let* $c(t) = b_0 + \frac{t}{D}(b_1 - b_0)$. *Then there exists a unique* $x_{f'} \in [0, D] \subset \mathbb{R}$ *such that*

$$f'^H = b_0 + \frac{x_{f'}}{D}(b_1 - b_0) = c(x_{f'}).$$

Moreover, $c : [0, D] \to Span(f,g) \cap H \cap C \setminus \{0\}$ *isometrically with respect to* d_E *(on* $[0, D]$*) and* d_V. *Note:* d_E *is the standard Euclidean metric on* $\mathbb{R}$.

Let $\{f_n\}_{n=1}^{\infty}$ *be a sequence in* $Span(f,g) \cap C \setminus \{0\}$ *such that* f_n^H *converges to* f'^H *with respect to the metric* d_V. *If* $[f'] \neq [b_0], [b_1]$*; or equivalently if* $x_{f'} \neq 0, D$, *then*

$$\frac{4}{D} \leq \lim_{f_n^H \to f'^H} \frac{d_H(f', f_n)}{d_V(f'^H, f_n^H)} = \frac{D}{x_{f'}(D - x_{f'})} < \infty.$$

(So $f_n \to f'$ with respect to d_H.)

If x_* is such that $0 < x_* < x_{f'}$ and we let $f^* \in [c(x_*)]$ then

$$\lim_{f_n^H \to f'^H} \frac{d_H(f', f_n)}{d_V(f'^H, f_n^H)} = \left. \frac{d}{dx}\, d_H(f^*, c(x)) \right|_{x = x_{f'}}.$$

If $f' \in [\frac{1}{2}\, b_0 + \frac{1}{2}\, b_1]$, so that $x_{f'} = D/2$, then

$$\frac{4}{D} = \lim_{f_n^H \to f'^H} \frac{d_H(f', f_n)}{d_V(f'^H, f_n^H)} \tag{2.161}$$

and there is no other $f' \in Span(f, g) \cap C \setminus \{0\}$ for which (2.161) holds.

Proof. See Figures 2.4 (page 223), 2.5 (page 243), 2.6 (page 244), 2.7 (page 245), and 2.8 (page 246).

Proof of Part 1.

Part 1 is just a restatement of parts 9 and 10 of Theorem 2.4.11.1 (page 212).

Proof of (2.158) of Part 2.

We identify the line segment $\overline{b_0 b_1} = \text{Span}(f, g) \cap C \cap H$ with line segment $[0, D] \subset \mathbb{R}$ by the isometry

$$c(x) = \left(1 - \frac{x}{D}\right) b_0 + \frac{x}{D}\, b_1. \tag{2.162}$$

The proof that c maps the real line $\mathbb{R}$ to the line $c\,(\mathbb{R}) \subset V$ isometrically and that $c([0, D]) = \overline{b_0 b_1} = \text{Span}(f, g) \cap C \cap H$ is given in the first 4 parts of Theorem 2.4.12.1 (page 222).

Since $d_V(f'^H, g'^H)$ is finite and non-zero (non-zero due to their linear independence), (2.158) is trivially true if $d_H(f', g') = \infty$. So we will assume that $d_H(f', g') < \infty$.

If $d_H(f', g') < \infty$ then by Theorem 2.4.12.1 (page 222) there exist two unique

numbers in $(0, D)$, $x_{f'}$ and $x_{g'}$ such that $c(x_{f'}) = f'^H$ and $c(x_{g'}) = g'^H$. We can assume that $x_{f'} < x_{g'}$ without loss of generality by relabeling (which we can do since metrics are symmetric) if necessary.

Let $x \in (0, D)$ satisfy $x_{f'} < x$, so x could be $x_{g'}$. Then Theorem 2.4.12.1 Equality (2.147) (page 224) implies [22],

$$d_H(f', c(x)) = d_H(c(x_{f'}), c(x)) = \ln\left(\frac{x}{x_{f'}}\,\frac{D - x_{f'}}{D - x}\right). \tag{2.163}$$

Multiplying (2.163) by K yields

$$\begin{aligned} K d_H(f', c(x)) &= K d_H(c(x_{f'}), c(x)) \qquad (2.164)\\ &= K \ln\left(\frac{x}{x_{f'}}\,\frac{D - x_{f'}}{D - x}\right)\\ &= K \ln\left(\frac{D - x_{f'}}{x_{f'}}\,\frac{x}{D - x}\right)\\ &= K \ln\left(\frac{D - x_{f'}}{x_{f'}}\right) + K \ln\left(\frac{x}{D - x}\right). \text{ We also have}\\ d_V(f^H, c(x)) &= x - x_{f'}, \quad \text{since } c \text{ is an isometry.} \end{aligned}$$

$K d_H(f', c(x))$ and $d_V(f'^H, c(x))$ are both functions of x. Both map $[x_{f'}, D)$ into $[0, \infty)$. Both equal 0 at $x = x_{f'}$ because

$$\begin{aligned} c(x_{f'}) = f'^H &\;\Rightarrow\; K d_H(f', c(x_{f'})) = K d_H(f', f'^H) = 0\\ c(x_{f'}) = f'^H &\;\Rightarrow\; d_V(f'^H, c(x_{f'})) = d_V(f'^H, f'^H) = 0. \end{aligned}$$

To prove (2.158) of Part 2 of this lemma we will show that $K \geq D/4$ implies

$$\frac{d}{dx}\, K d_H(f', c(x)) \;>\; \frac{d}{dx}\, d_V(f'^H, c(x)) = 1 \tag{2.165}$$

[22] In Theorem 2.4.12.1 Equality (2.147) (page 224) substitute $x_{f'}$ from this lemma for x; and substitute x from this lemma for y. Equation 2.163 follows.

if $x \in (x_{f'}, D)$ and $x \neq D/2$. The Fundamental Theorem of Calculus applied to (2.165) will then give us $Kd_H(f, c(x)) > d_V(f, c(x))$ because the value of an integrand at a single unique point (e.g. at x = D/2, or more generally, on a set of measure zero) does not effect the value of the integral.

We differentiate (2.164) with respect to x:

$$\begin{aligned}
\frac{d}{dx}\ Kd_H(f', c(x)) &= K\frac{d}{dx}\ \ln\left(\frac{x}{D-x}\right) \\
&= K\frac{D-x}{x}\ \frac{(D-x)(1)-(x)(-1)}{(D-x)^2} \\
&= K\frac{D-x}{x}\ \frac{D}{(D-x)^2} \\
&= K\frac{D}{x(D-x)}. \qquad (2.166) \\
\frac{d}{dx}\ d_V(f'^H, c(x)) &= \frac{d}{dx}\ (x - x_{f'}) = 1 \qquad (2.167)
\end{aligned}$$

See [23]. Combining (2.166) with (2.167) yields:

$$\frac{d}{dx}\ Kd_H(f', c(x)) > \frac{d}{dx}\ d_V(f'^H, c(x)) \qquad (2.168)$$

if and only if

$$\begin{aligned}
\frac{KD}{x(D-x)} &> 1 \\
KD &> x(D-x) \\
x^2 - Dx + KD &> 0 \qquad (2.169)
\end{aligned}$$

and

$$\frac{d}{dx}\ Kd_H(f', c(x)) = \frac{d}{dx}\ d_V(f'^H, c(x)) \qquad (2.170)$$

[23] (2.166) implies that $\frac{d}{dx}\ d_H(f, c(x))$ is essentially independent of f'. (2.166) implies that $\frac{d}{dx}\ d_V(f^H, c(x))$ is independent of f^H.

if and only if

$$x^2 - Dx + KD = 0. \tag{2.171}$$

Using the quadratic formula we find the roots of $x^2 - Dx + KD = 0$:

$$\begin{aligned} x &= \frac{D \pm \sqrt{D^2 - 4KD}}{2} \\ &= \frac{D \pm \sqrt{D(D - 4K)}}{2}. \end{aligned} \tag{2.172}$$

(2.172) implies that if $K > D/4$ then $x^2 - Dx + KD = 0$ will have no real roots and if $K = D/4$ that $x^2 - Dx + KD = 0$ will have a single repeated root at $x = D/2$.

Since $x^2 - Dx + KD$ is concave up, it follows that

$$x^2 - Dx + KD > 0 \text{ if } \underbrace{K > D/4}_{\text{condition 1}} \text{ or if } \underbrace{K = D/4 \text{ and } x \neq D/2}_{\text{condition 2}}. \tag{2.173}$$

Combining the two conditions in (2.173) yields:

$$x^2 - Dx + KD > 0 \text{ if } K \geq D/4 \text{ and } x \neq D/2. \tag{2.174}$$

So (2.174), with the help of (2.168), (2.169) and (2.170), (2.171), implies:

$$\frac{d}{dx}\, Kd_H(f', c(x)) > \frac{d}{dx}\, d_V(f'^H, c(x)) \text{ if } K \geq D/4 \text{ and } x \neq D/2. \tag{2.175}$$

As mentioned earlier, see (2.165) and the comments following it, applying the Fundamental Theorem of Calculus to (2.175) yields

$$Kd_H(f', c(x)) > d_V(f'^H, c(x)) \text{ if } K \geq D/4,$$

and so (2.158) of Part 2 of this lemma is proven.

(2.159) follows immediately from (2.158) if we let $f' = f$ and $g' = g$ and $K = \frac{D}{4}$.

(2.160) follows immediately from (2.158) if we let $K = D/4$.

So the rest of Part 2 of this lemma is proven.

Proof of 3.

If $0 < K < D/4$, then $D > D - 4K > 0$ and

$$I = \left(\frac{D - \sqrt{D(D-4K)}}{2}, \ \frac{D + \sqrt{D(D-4K)}}{2} \right) \tag{2.176}$$

is an interval of length $0 < \sqrt{D(D-4K)} < D$ centered at $D/2$ and contained in $(0, D)$. It is worth noting that the center of I corresponds to the center of [0,D], which is the point $D/2$.

The calculations starting at (2.168), with the appropriate inequalities reversed, imply if $x_{f'}, x \in I$, with $x_{f'} < x$, then

$$\frac{d}{dx} \ K d_H(f', c(x)) < \frac{d}{dx} \ d_V(f'^H, c(x)). \tag{2.177}$$

The Fundamental Theorem of Calculus applied to (2.177) yields:

$$K d_H(f', c(x)) < d_V(f'^H, c(x)) \tag{2.178}$$

if

$$\frac{D - \sqrt{D(D-4K)}}{2} \ < \ x_{f'} \ < \ x \ < \ \frac{D + \sqrt{D(D-4K)}}{2}.$$

So let $x_{f'}$ and $x_{g'}$ be two numbers that satisfy

$$\frac{D - \sqrt{D(D-4K)}}{2} \ < \ x_{f'} \ < \ x_{g'} \ < \ \frac{D + \sqrt{D(D-4K)}}{2}$$

and let $f' = c(x_{f'})$ and $g' = c(x_{g'})$. By Theorem 2.4.12.1 (page 222) $f', g' \in \text{Span}(f, g) \cap H \ \cap C \setminus \{0\}$. I.e. $f' = f'^H$ and $g' = g'^H$. Since c is an isometry

$c(x_{f'}) \neq c(x_{g'})$. Since H intersects each equivalence class of $(C \setminus \{0\},\ \sim)$ exactly once and $f' \neq g'$ it follows that f', g' are linearly independent. Finally, by (2.178)

$$K d_H(f', g') < d_V(f'^H, g'^H). \tag{2.179}$$

So Part 3 of this lemma is proven.

Proof of part 4.

The existence and uniqueness of an $x_{f'} \in [0, D]$ such that

$$f'^H = b_0 + \frac{x_{f'}}{D}(b_1 - b_0) = c(x_{f'})$$

is proven in Theorem 2.4.12.1 (page 222). Similarly, corresponding to each f_n^H is a unique number in $x_{f_n^H} \in [0, D]$ such that $c(x_{f_n^H}) = f_n^H$. Also proven in Theorem 2.4.12.1 is that the map c is an isometry. So the condition f_n^H converges to f'^H implies $x_{f_n^H}$ converges to $x_{f'}$.

Convergence to $x_{f'}$ from above:

From (2.164) (page 235), if $0 < x_{f'} < x < D$, then

$$\begin{aligned} d_H(f', c(x)) &= d_H(c(x_{f'}), c(x)) \\ &= \ln\left(\frac{x}{x_{f'}}\ \frac{D - x_{f'}}{D - x}\right) \qquad (2.180) \\ &= \ln\left(\frac{D - x_{f'}}{x_{f'}}\ \frac{x}{D - x}\right) \\ &= \ln\left(\frac{D - x_{f'}}{x_{f'}}\right) + \ln\left(\frac{x}{D - x}\right) \qquad (2.181) \end{aligned}$$

and

$$d_V(f^H, c(x)) = x - x_{f'}. \tag{2.182}$$

If in (2.181) and (2.182) we let $x = (x_{f'} + \Delta_x)$, with $0 < \Delta_x < (D - x_{f'})$, and then

take their quotient, we get, after rearranging the terms and making use of

$$\ln(a/b) = -\ln(b/a),$$

$$\begin{aligned}\frac{d_H(f', c(x_{f'} + \Delta_x))}{d_V(f'^H, c(x_{f'} + \Delta_x))} &= \frac{\ln\left(\frac{D-x_{f'}}{x_{f'}}\right) + \ln\left(\frac{(x_{f'}+\Delta_x)}{D-(x_{f'}+\Delta_x)}\right)}{(x_{f'} + \Delta_x) - x_{f'}} \\ &= \frac{\ln\left(\frac{(x_{f'}+\Delta_x)}{D-(x_{f'}+\Delta_x)}\right) - \ln\left(\frac{x_{f'}}{D-x_{f'}}\right)}{\Delta_x}. \end{aligned} \tag{2.183}$$

Of course (2.183) is the equation for the slope of a secant line to the graph $\left(x, \ln\left(\frac{x}{D-x}\right)\right)$, its limit, as $\Delta_x \to 0$, being a derivative.

Convergence to $x_{f'}$ from below:

By Theorem 2.4.12.1 Equation (2.148) (page 225), if $0 < x < x_{f'}$ then

$$\begin{aligned} d_H(f', c(x)) &= d_H(c(x_{f'}), c(x)) \\ &= -\ln\left(\frac{x}{x_{f'}}\ \frac{D - x_{f'}}{D - x}\right) \end{aligned} \tag{2.184}$$

and since c is an isometry,

$$d_V(f^H, c(x)) = -(x - x_{f'}). \tag{2.185}$$

(2.184) is the negative of (2.180) and (2.185) is the negative of (2.182). As we did in the convergence from above case, we replace x with $(x_{f'} + \Delta_x)$, except now $0 > \Delta_x > -x_{f'}$. We take the quotient (2.184) over (2.185). The two negatives in the quotient cancel out. The result is identical in form to (2.183).

Convergence to $x_{f'}$ from above or below:

If $x_{f'} \in (0, D)$ then

$$\left.\frac{d}{dx}\ \ln\left(\frac{x}{D-x}\right)\right|_{x = x_{f'}}$$

exists and so

$$\begin{aligned}
\lim_{\Delta_x \to 0^+} \frac{\ln\left(\frac{(x_{f'}+\Delta_x)}{D-(x_{f'}+\Delta_x)}\right) - \ln\left(\frac{x_{f'}}{D-x_{f'}}\right)}{\Delta_x} &= \lim_{\Delta_x \to 0^-} \frac{\ln\left(\frac{(x_{f'}+\Delta_x)}{D-(x_{f'}+\Delta_x)}\right) - \ln\left(\frac{x_{f'}}{D-x_{f'}}\right)}{\Delta_x} \\
&= \left.\frac{d}{dx}\ \ln\left(\frac{x}{D-x}\right)\right|_{x = x_{f'}} \\
&= \left.\frac{D}{x(D-x)}\right|_{x = x_{f'}} \\
&= \frac{D}{x_{f'}(D - x_{f'})}. \qquad (2.186)
\end{aligned}$$

We return to the sequence f_n^H which converges to f'^H.

Since c is an isometry, the convergence of f_n^H to f'^H implies the convergence of $x_{f_n^H}$ to $x_{f'}$. So we can write each $x_{f_n^H}$ as $x_{f'} + \Delta_{x_n}$ with $\Delta_{x_n} \to 0$ as $n \to \infty$.

But then (2.183), when combined with

$$c(x_{f_n^H}) = c(x_{f'} + \Delta_{x_n}) = f_n^H, \qquad [f_n] = [f_n^H],$$

and (2.186), give us:

$$\lim_{n\to\infty}\ \frac{d_H(f', c(x_{f'} + \Delta_{x_n}))}{d_V(f'^H, c(x_{f'} + \Delta_{x_n}))} = \lim_{f_n^H \to f'^H}\ \frac{d_H(f', f_n)}{d_V(f'^H, f_n^H)} = \frac{D}{x_{f'}(D - x_{f'})}.$$

If $x_* \in (0, x_{f'})$ and $f^* \in [c(x_*)]$, then by (2.166) (page 236)

$$\begin{aligned}\lim_{f_n^H \to f'^H} \frac{d_H(f', f_n)}{d_V(f'^H, f_n^H)} &= \left. \tfrac{d}{dx} \ln\left(\tfrac{x}{D-x}\right) \right|_{x = x_{f'}} \\ &= \left. \tfrac{D}{x(D-x)} \right|_{x = x_{f'}} \\ &= \left. \tfrac{d}{dx}\, d_H(f^*, c(x)) \right|_{x = x_{f'}} .\end{aligned}$$

To minimize

$$\frac{d}{dx}\, d_H(f^*, c(x)) = \frac{D}{x(D-x)} \tag{2.187}$$

we set the derivative of (2.187) equal to zero and solve for x:

$$\frac{d^2}{dx^2}\, d_H(f^*, c(x)) = -\frac{D(D-2x)}{x^2(D-x)^2} = 0.$$

We get uniquely $x = D/2$. Then we use the second derivative test from Calculus, which in this case will involve a third derivative

$$\frac{d^3}{dx^3}\, d_H(f^*, c(x)) = \frac{2D(D^2 - 3Dx + 3x^2)}{x^3(D-x)^3}$$

into which we plug $D = D/2$. We get $\frac{32}{D^3}$ which is positive, so $\frac{d}{dx}\, d_H(f^*, c(x))$ has its minimum at $x = D/2$.

Plugging $x = D/2$ into $\frac{d}{dx}\, d_H(f^*, c(x))$, which is (2.187), yields a minimum of $4/D$.

Plugging $x = \frac{D}{2}$ into c yields $c(D/2) = .5b_0 + .5b_1$.

□

The following useful theorem uses the notation from, and is a consequence of, Lemma 2.5.0.2 (page 232). It proves that if $H \cap C$ is bounded in V by r, then

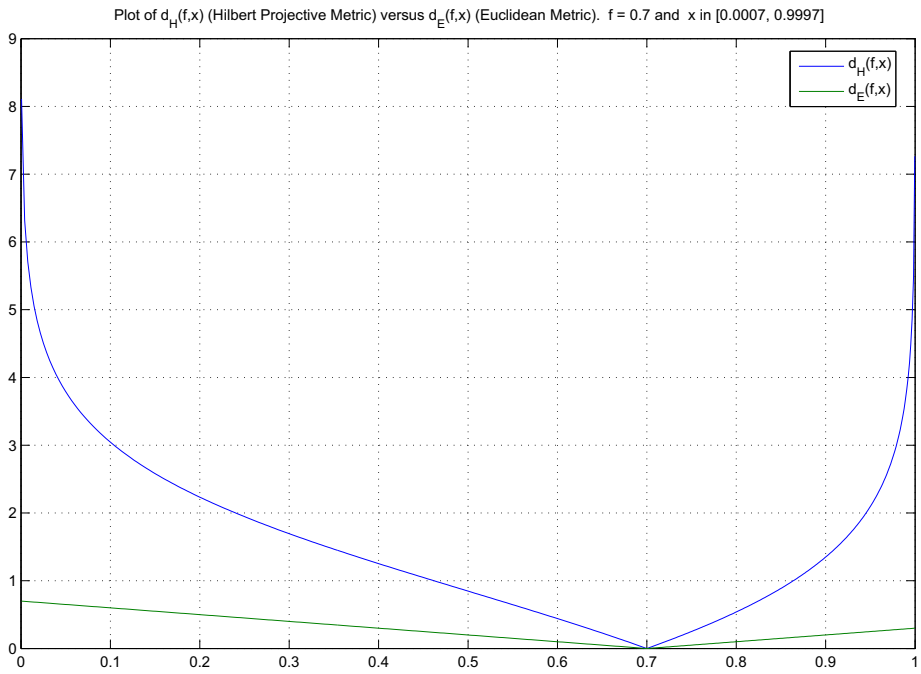


Figure 2.5: The Hilbert Projective Metric, d_H, is compared to the Euclidean Metric, d_E, on the interval $[0, 1]$ at the point $f = 0.7 \in [0, 1]$. Notice that $d_H(f, x) > d_E(f, x)$ for all $x \neq f$.

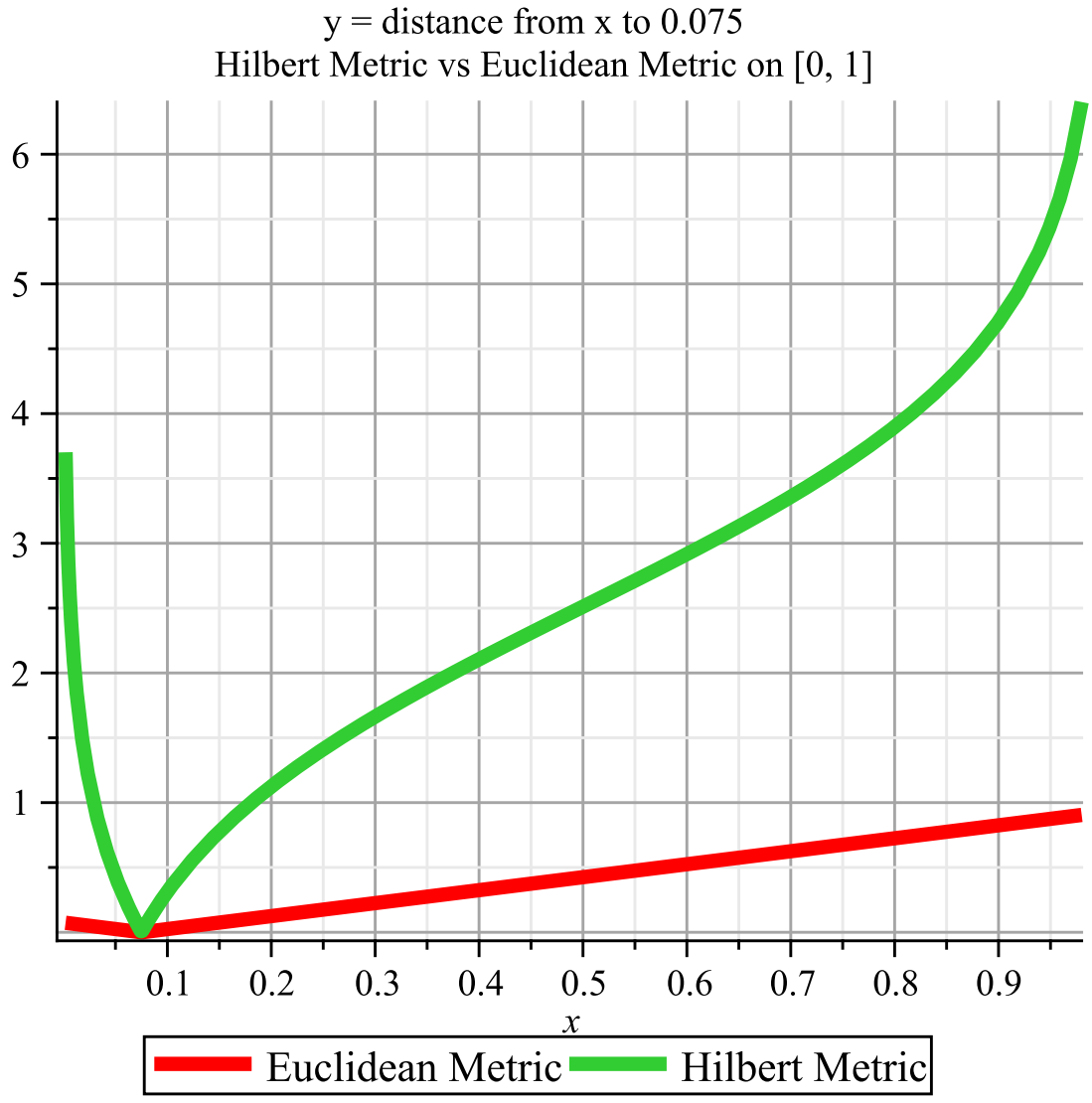


Figure 2.6: The Hilbert Projective Metric, d_H, is compared to the Euclidean Metric, d_E, on the interval $[0,1]$ at the point $f = 0.075 \in [0,1]$. Notice that $d_H(f,x) > d_E(f,x)$ for all $x \neq f$; that the concavity changes at $D/2 = 0.5$; that the positive slopes of d_H have their minimum at $D/2$.

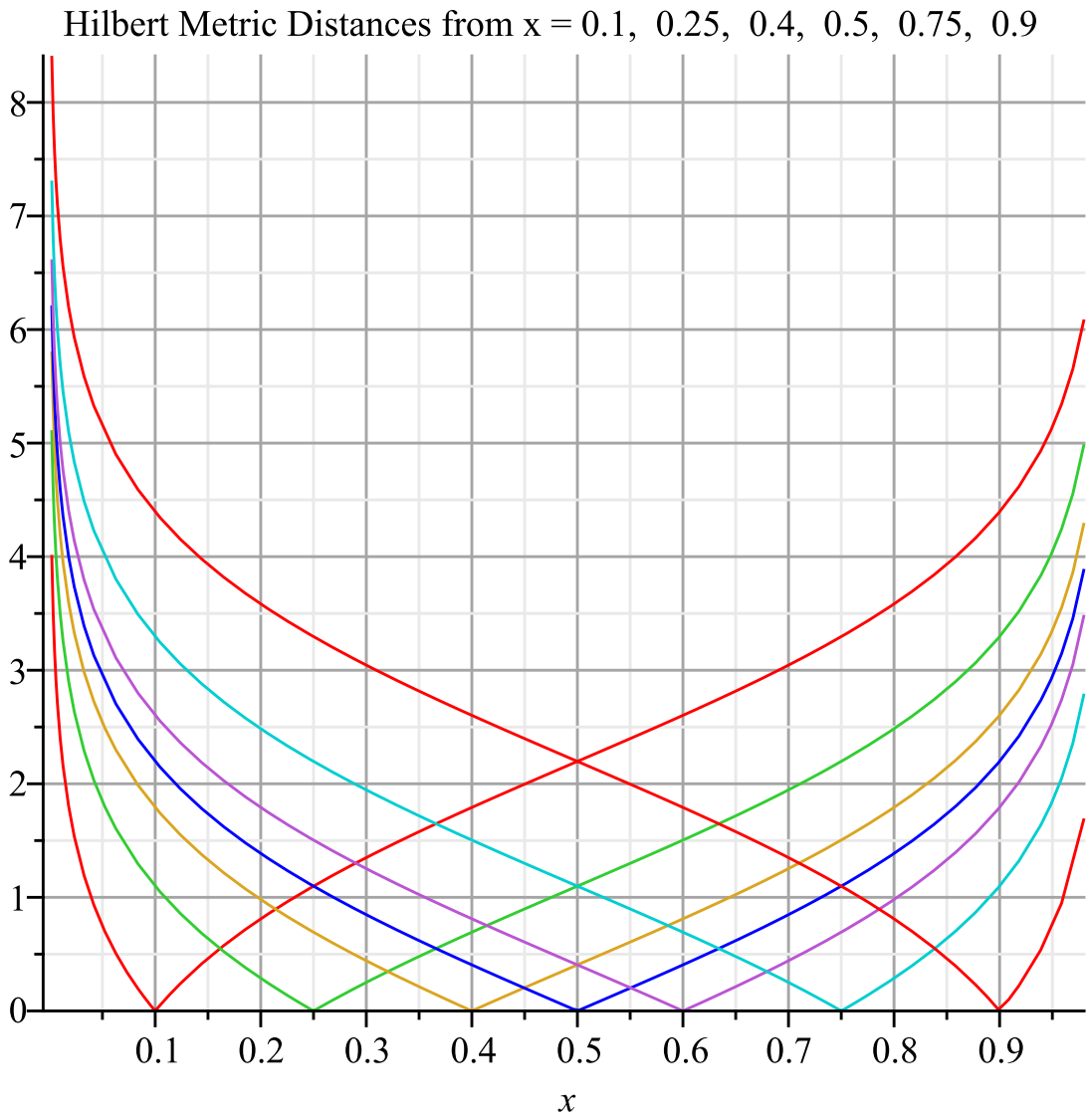


Figure 2.7: The Hilbert Projective Metric d_H. Distances to various points on the interval $[0, 1]$.

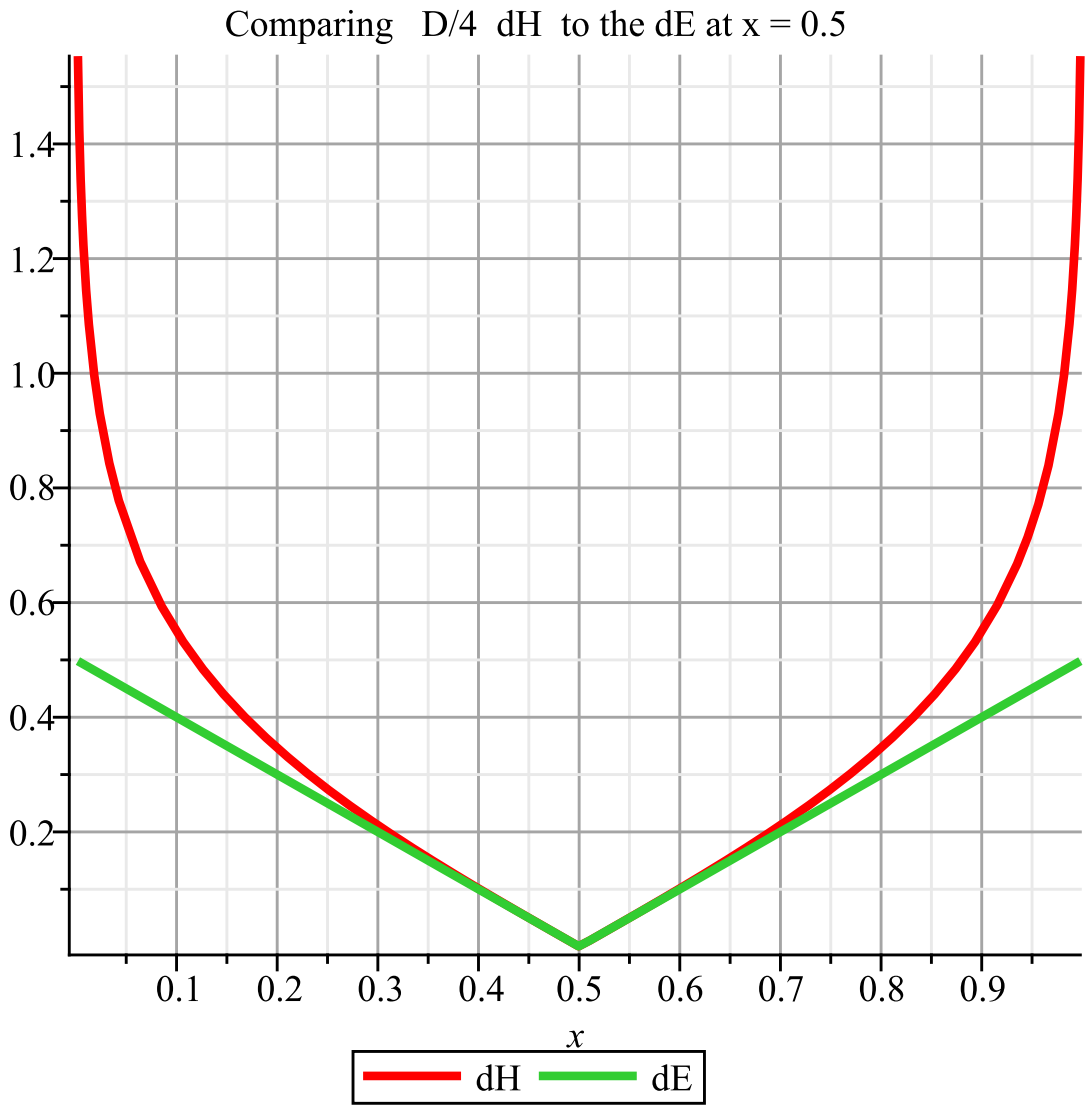


Figure 2.8: $\frac{D}{4}d_H$ is compared to d_E with respect to distances to $D/2$. Notice the slopes of d_H and d_E correspond at $D/2$ and we are taking $D = 1$.

$d_V(f^H, g^H) < \frac{r}{2}\ d_H(f, g)$.

Theorem 2.5.0.3. *Suppose that H is a hyperplane which intersects each equivalence class of $(C \setminus \{0\}, \sim)$ exactly once. Moreover, suppose that*

$$H \cap C \subset \overline{B_r(0)} = \{x \in V : \|x\|_V \leq r\}$$

Let $f, g \in C \setminus \{0\}$ and let f^H, g^H be the central projections of f, g to H.

If $f^H \neq g^H$ then

$$d_V(f^H, g^H) < \frac{r}{2}\ d_H(f, g). \tag{2.188}$$

If $f^H = g^H$ then

$$d_V(f^H, g^H) = d_H(f, g) = 0. \tag{2.189}$$

Proof. If $f^H \neq g^H$ then f, g are linearly independent. By Lemma 2.5.0.2 (page 232)

$$d_V(f^H, g^H) < \frac{D}{4}\ d_H(f, g) \tag{2.190}$$

where $D = d_V(b_0, b_1)$ and where b_0, b_1 are a pair of ends for f, g in

$$\mathrm{Span}(f, g) \cap H \bigcap C \setminus \{0\}.$$

Since we are assuming that $H \cap C \subset \overline{B_r(0)}$ the triangle inequality give us

$$D = d_V(b_0, b_1) \leq d_V(b_0, 0) + d_V(0, b_1) \leq 2r.$$

Since $D \leq 2r$, (2.190) implies

$$d_V(f^H, g^H) < \frac{r}{2}\ \ d_H(f, g).$$

So (2.188) is proven.

If If $f^H = g^H$ then $d_V(f^H, g^H) = 0$ and f, g linearly dependent. f, g are linearly dependent implies $d_H(f, g) = 0$. So (2.189) is proven. □

Corollary 2.5.0.4. *Let $C = \mathbb{R}^n_{\geq 0},\ n \geq 2$ be the standard cone in $\mathbb{R}^n$. Let*

$$\triangle^{n-1} = \underbrace{\{x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \ : \ \|x\|_1 = 1\}}_{\text{hyperplane } H_1} \cap \underbrace{\mathbb{R}^n_{\geq 0}}_{\text{cone } C}$$

be the standard $n-1$ simplex [24].

If $f, g \in \mathbb{R}^n_{\geq 0} \setminus \{0\}$ then the central projections of f, g to H_1 are

$$f^{H_1} = \frac{f}{\|f\|_1} \text{ and } g^{H_1} = \frac{g}{\|g\|_1}.$$

If $f^{H_1} \neq g^{H_1}$ then

$$d_E\left(\frac{f}{\|f\|_1}, \frac{g}{\|g\|_1}\right) \ < \ \frac{\sqrt{2}}{4}\, d_H(f, g) \ < \ d_H(f, g) \tag{2.191}$$

since $\frac{\sqrt{2}}{4} = 0.353553391$.

If $f^{H_1} = g^{H_1}$ then

$$d_E\left(\frac{f}{\|f\|_1}, \frac{g}{\|g\|_1}\right) = d_H(f, g) = 0. \tag{2.192}$$

Proof. $V = \mathbb{R}^n$ = Euclidean space, so $d_V = d_E$. See (2.107) (page 197) for results pertaining to H_1 and $\triangle^{n-1}$.

If $f^{H_1} \neq g^{H_1}$ then f, g are linearly independent and Lemma 2.5.0.2 (page 232) implies

$$d_E\left(f^{H_1}, g^{H_1}\right) < \frac{D}{4}\, d_H(f, g) \tag{2.193}$$

[24] $\|x\|_1 = \sum_{i=1}^n |x_i|$. See (2.107) (page 197).

where $D = d_V(b_0, b_1)$ and where b_0, b_1 are a pair of ends for f, g contained in

$$\text{Span}(f, g) \cap H_1 \cap C \setminus \{0\}.$$

Since $b_0, b_1 \subset H_1 \cap C = \triangle^{n-1}$ and since Theorem 2.4.7.2 (page 197) guarantees

$$\text{diameter}\left(\triangle^{n-1}\right) = \sqrt{2}$$

when $n \geq 2$, we must have

$$D = d_V(b_0, b_1) \leq \text{ diameter}(\triangle^{n-1}) = \sqrt{2}.$$

So (2.193) implies

$$d_E\left(f^{H_1}, g^{H_1}\right) < \frac{\sqrt{2}}{4}\, d_H\left(f, g\right).$$

So (2.191) is proven.

If $f^{H_1} = g^{H_1}$ then f, g are linearly dependent and $d_H(f, g) = 0$. Since $f^{H_1} = g^{H_1}$ immediately implies $d_E(f, g) = 0$, we have proven (2.192). □

2.6 Birkhoff's Projective Contraction Theorem

2.6.1 The induced map P is always continuous

The following lemma will be used in our proof of Birkhoff's Projection Contraction Theorem [12] in Section 2.6.4 (page 260) below.

Birkhoff required that P be bounded so that the map on $(C \setminus \{0\}, \sim)$ induced by P would be continuous. However, the condition of being bounded is unnecessary as the following lemma, Lemma 2.6.1.1 shows.

Bushell [19] in his Theorem 3.1 shows that if a linear map A maps the interior of

a cone K, denoted $\overset{\circ}{K}$, to itself, then $d_H(Ax, Ay) \leq d_H(x, y)$ if $x, y \in \overset{\circ}{K}$. He requires that the cone K have non-empty interior.

Our result, Lemma 2.6.1.1, is related to Bushell's, but it is a little more general. First of all, we do not require our cone C to have a non-empty interior. Secondly our method focuses on the problem of $\ker P$. I.e. if $g \in \ker P$ then $d_H(fP, gP)$ is not defined (since d_H is not defined at 0). By addressing this issue we obtain the most general result possible; i.e. if $d_H(fP, gP)$ is defined, meaning if $f, g \notin \ker P$, then $d_H(fP, gP) \leq d_H(f, g)$. Later, in Proposition 2.6.2.1 (page 257) we show that $\overset{\circ}{K} \subset C \setminus \ker P$. We also show, here in Lemma 2.6.1.1, that by extending our definition of d_H to include 0, by defining $d(f, 0) = \infty$ if $f \in C \setminus \{0\}$, then $d_H(fP, gP) \leq d_H(f, g)$ holds for all of C and that relative to this extended definition of d_H the map induced by P is continuous on all $(C, \sim)$. Additionally, our goal is different, as we utilize our results to address issues of continuity.

Lemma 2.6.1.1. *Let P be any linear map that maps V to itself and satisfies $CP \subset C$.*

1. *Let $f, g \in C \setminus \ker P$ then $d_H(fP, gP) \leq d_H(f, g)$.*

2. *P as an induced map, maps*

$$(C \setminus \ker P,\ \sim) \ \rightarrow \ (C \setminus \{0\},\ \sim)$$

   continuously with respect to d_H.

3. *For each positive integer n, P as an induced map, maps*

$$(C \setminus \ker P^n, \sim) \rightarrow \left(C \setminus \ker P^{n-1}, \sim\right)$$

   continuously with respect to d_H. Note $\ker P^0 = \{0\}$ as we are taking P^0 to be the identity map.

4. *P as an induced map, maps*

$$(C \setminus \cup_{n=1}^{\infty} \ker P^n,\ \sim)$$

to itself continuously w.r.t d_H.

5. *If we extend the definition of d_H so that $d_H(0,0) = 0$ and*

$$d_H(f,0) = d_H(0,f) = \infty$$

for each $f \in C \setminus \{0\}$, then

   (a) *for all $f, g \in C$, $d_H(fP, gP) \leq d_H(f,g)$.*

   (b) *P, as an induced map, maps $(C, \sim)$ to itself continuously with respect to the extended version of d_H.*

Proof.

Proof of Part 1. We will use the $\alpha f \leq g \leq \beta f$ definition for d_H which is found inside Theorem 1.9.3.2 (page 84). We rewrite the definition here for convenience:

Let $f, g \in C \setminus \{0\}$ and let α, β be the largest and smallest non-negative real numbers such that

$$\alpha f \leq g \leq \beta f,$$

assuming they exist. If no real number β exists such that $g \leq \beta f$, we set $\beta = \infty$. We can always find an $\alpha \in \mathbb{R}_{\geq 0}$ such that $\alpha f \leq g$. It is always the case that

$$0 \leq \alpha < \infty \qquad 0 < \beta \leq \infty \qquad \alpha \leq \beta. \tag{2.194}$$

if $f, g \in C \setminus \{0\}$. One defines

$$d_H(f,g) = \ln(\beta/\alpha) \in [0, \infty]. \tag{2.195}$$

In Section 1.9.4 (page 91) we show that if we try to apply the $\alpha f \leq g \leq \beta f$ definition of $d_H(f,g)$ when one of f or $g = 0$ then β/α will be indeterminate. What is of relevance to this proof is to realize that d_H, with its usual definition, is defined only on $C \setminus \{0\}$. So, for this reason the induced map of P can, at best, be continuous with respect to d_H only if we restrict its domain, so that its domain and range stay within $C \setminus \{0\}$. In other words, when its domain is $C \setminus \ker P$.

So, let $f, g \in C \setminus \ker P$ hence

$$fP, gP \in C \setminus \{0\}.$$

The following notation is useful for our proof. Let

$$\alpha_{fg} = \sup\{t : g - tf \in C\}$$
$$\beta_{fg} = \inf\{t : tf - g \in C\},$$

so (2.195) becomes

$$d_H(f,g) = \ln(\beta_{fg}/\alpha_{fg})\,.$$

Since P is linear and $CP \subset C$,

$$\{t : g - tf \in C\} \subset$$
$$\{t : (g - tf)P \in CP\} \subset$$
$$\{t : (g - tf)P \in C\} =$$
$$\{t : gP - t(fP) \in C\}.$$

So

$$\sup\{t : g - tf \in C\} \leq \sup\{t : gP - t(fP) \in C\}$$

$$\alpha_{fg} \leq \alpha_{fPgP}. \tag{2.196}$$

Since $f, g, fP, gP \in C \setminus \{0\}$ and since (2.196) holds, (2.194) implies

$$0 \ \leq \ \alpha_{fg} \ \leq \ \alpha_{fPgP} \ < \ \infty.$$

We develop a similar relationship for the β:

$$\{t : tf - g \in C\} \subset$$

$$\{t : (tf - g)P \in CP\} \subset$$

$$\{t : (tf - g)P \in C\} =$$

$$\{t : t(fP) - gP \in C\}.$$

So

$$\inf\{t : tf - g \in C\} \geq \inf\{t : gP - t(fP) \in C\}$$

$$\beta_{fg} \geq \beta_{fPgP}. \tag{2.197}$$

Since $f, g, fP, gP \in C \setminus \{0\}$ and since (2.197) holds, (2.194) implies

$$0 \ < \ \beta_{fPgP} \ \leq \ \beta_{fg} \ \leq \ \infty.$$

But

$$0 \;<\; \beta_{fPgP} \;\leq\; \beta_{fg} \;\leq\; \infty$$

$$0 \;\leq\; \alpha_{fg} \;\leq\; \alpha_{fPgP} \;<\; \infty$$

implies

$$\frac{\beta_{fPgP}}{\alpha_{fPgP}} \leq \frac{\beta_{fg}}{\alpha_{fg}}.$$

So $d_H(fP, gP) \leq d_H(f, g)$.

Proof of Part 2. If $f, g \in C \setminus \ker P$ then by Part 1 of this lemma we have

$$d_H(fP, gP) \leq d_H(f, g)$$

so

$$d_H([f]P, [g]P) \leq d_H([f], [g]) \tag{2.198}$$

because $d_H(f, g) = d_H([f], [g])$ by the definition of d_H on $(C \setminus \{0\}, \sim)$.

To show continuity at $[f] \in (C \setminus \ker P, \sim)$ it suffices to show that given a finite $\epsilon > 0$ we can find a finite $\delta > 0$ such that

$$g \in C \setminus \ker P \;\text{ and }\; d_H([f], [g]) < \delta \;\text{ then }\; d_H([f]P, [g]P) < \epsilon.$$

By Part 1 of this lemma, or (2.198), it suffices to let $\delta = \epsilon$. So P (or more accurately, the map induced by P, say 'P', see [25]) is continuous on $(C \setminus \ker P, \sim)$.

Claim: 'P' maps

$$(C \setminus \ker P,\ \sim) \;\rightarrow\; (C \setminus \{0\},\ \sim)\,.$$

[25] For lack of better notation sometimes quotes will appear around P, as in 'P', to emphasize that we are dealing with the map induced by P. I.e. 'P' maps $[f]$ to $[fP] = [f]P$. Other times P alone will be used for the induced map. There should be no confusion and the meaning of P should be clear from the context.

Proof of claim: P maps C to itself. So if $f \in C \setminus \ker P$ then $fP \in C$ and $fP \neq 0$ so

$$[f]P = [fP] \in (C \setminus \{0\}, \sim).$$

Proof of Part 3.

Let $v \in V$ and let n be a positive integer. If $vP^{n-1} = 0$ then

$$vP^n = vP^{n-1}P = 0P = 0,$$

so

$$\ker P^{n-1} \subset \ker P^n$$

and

$$(C \setminus \ker P^n, \sim) \subset \left(C \setminus \ker P^{n-1}, \sim\right) \subset \ldots \subset (C \setminus \ker P, \sim). \tag{2.199}$$

By part 2 of this lemma the map induced by P is continuous on $(C \setminus \ker P, \sim)$. By (2.199)

$$(C \setminus \ker P^n, \sim) \subset (C \setminus \ker P, \sim).$$

So the map induced by P is continuous on $(C \setminus \ker P^n, \sim)$.

Claim: 'P' maps

$$(C \setminus \ker P^n, \sim) \to \left(C \setminus \ker P^{n-1}, \sim\right).$$

Proof of Claim: P maps C to itself. So if $f \in C \setminus \ker P^n$ then $fP \in C$ and $fP^n \neq 0$ so

$$([f]P)P^{n-1} = [f]P\ P^{n-1} = [f]P^n = [fP^n] \neq [0]$$

so

$$[f]P \in (C \setminus \ker P^{n-1}, \sim).$$

Proof of Part 4.

$$\ker P \ \subset \ \cup_{n=1}^{\infty} \ker P^n$$

so

$$(C \setminus \cup_{n=1}^{\infty} \ker P^n, \sim) \ \subset \ (C \setminus \ker P, \sim). \tag{2.200}$$

By Part 2 of this lemma 'P' is continuous on $(C \setminus \ker P, \sim)$ so by (2.200) 'P' is continuous on $(C \setminus \cup_{n=1}^{\infty} \ker P^n, \sim)$.

Claim: 'P' maps

$$(C \setminus \cup_{n=1}^{\infty} \ker P^n, \sim)$$

to itself.

Proof of Claim: P maps C to itself. So if

$$f \in C \setminus \cup_{n=1}^{\infty} \ker P^n$$

then $fP \in C$ and $fP^n \neq 0$ for any positive integer n. But then $(fP)P^m = fP^{m+1} \neq 0$ for any positive integer m. So

$$[f]P \in (C \setminus \cup_{n=1}^{\infty} \ker P^n, \ \sim).$$

Proof of Part 5(a). See [26].

By part 1 of this lemma $f, g \in C \setminus \ker P$ implies $d_H(fP, gP) \leq d_H(f, g)$. So to prove 5(a) all that remains to be shown is that if one or both of f, g are $\in C \cap \ker P$

[26] In part of the proof of part 5 we will be using the extended definition of d_H. In the rest of this paper, unless otherwise noted, we will exclusively be using the standard definition of d_H, which of course, is defined only on $C \setminus \{0\}$.

then

$$d_H(fP, gP) \leq d_H(f, g). \tag{2.201}$$

We show this by considering the possible cases:

If $f \in C \setminus \ker P$ and $g \in \ker P \cap C \setminus \{0\}$ then, by Corollary 2.2.3.4 (page 154), $d_H(f, g) = \infty$ and so (2.201) is immediately true.

If $f \in C \setminus \ker P$ and $g = 0$ then, by the extended definition of d_H we have $d_H(f, g) = \infty$ and (2.201) is immediately true.

If both $f, g \in C \cap \ker P$ then $fP = gP = 0$ and so (2.201) is immediately true.

Part 5(b) is a trivial consequence of 5(a).

□

2.6.2 More on the topology of $\ker P$ and ∂C

Proposition 2.6.2.1. *As usual, let C be a closed, convex, salient, pointed by the origin cone contained in a Banach Space V. Let P be any linear map that maps V to itself and satisfies $CP \subset C$. We have*

1. *$(\ker P \cap C \setminus \{0\},\ \sim)$ and $(C \setminus \ker P,\ \sim)$ are both open and closed in $(C \setminus \{0\},\ \sim)$ with respect to d_H.*

2. *If there exists an $f \in C$ such that $fP \neq 0$ then*

   (a)

   $$\ker P \cap C \ \subset\ \partial C.$$

   (b)

   $$C \setminus \partial C \ \subset\ C \setminus \ker P.$$

3. *The topological interior [27] of C, $\overset{\circ}{C}$ is such that*

[27] $a \in$ the topological interior of a set A if there exists an open set U_a such that $a \in U_a \subset A$.

(a)

$$\left(\overset{\circ}{C} \;=\; C \setminus \partial_{top}\, C\right) \;\subset\; (C \setminus \partial C) \;\subset\; (C \setminus \ker P)\,.$$

Regarding ∂_{top} *see* [28]. *Regarding* ∂ *see* [29].

(b) If there exists an $f \in C$ *such that* $fP \neq 0$ *then*

$$\overset{\circ}{C} \subset\; C \setminus \partial C \;\subset\; C \setminus \ker P.$$

Proof.

Proof of Part 1. By Corollary 2.2.3.4 (page 154) if $f, g \in C \setminus \{0\}$ and $f \in C \setminus \ker P$ and $g \in \ker P$ then $d_H(f, g) = \infty$. So every open ball of finite radius about $[f] \in (C \setminus \ker P, \sim)$ will not have any elements from $(C \setminus \{0\} \cap \ker P,\ \sim)$ in it. By symmetry, every open ball of finite radius about $[g] \in (C \setminus \{0\} \cap \ker P,\ \sim)$ will not have any elements from $(C \setminus \ker P,\ \sim)$ in it.

Since $(\ker P \cap C \setminus \{0\},\ \sim)$ and $(C \setminus \{0\} \setminus \ker P,\ \sim)$ are complements of each other relative to $(C \setminus \{0\},\ \sim)$, part 1 of this proposition is proven.

Proof of Part 2.

(a) is a trivial consequence of Corollary 2.2.3.5 (page 155).

(b) is a trivial consequence of part 2(a) of this proposition.

Proof of Part 3.

(a) The following two trivial results are useful enough to be labeled as lemmas. They will combine to provide a proof of 3(a):

Lemma 2.6.2.2. *Let* $A \subset X$, X *a topological space, then*

$$\overset{\circ}{A} = A \;\setminus \partial_{\,top}\; A.$$

[28] $v \in$ the topological boundary of A, denoted $\partial_{\text{top}}\, A$, if whenever $v \in U_v$, an open set about v, both $U_v \cap A$ and $U_v \cap V \setminus A$ are non-empty. Equivalently $\partial_{\text{top}}\, A = \overline{A} \setminus \overset{\circ}{A}$.

[29] Let $A \subset V$ we have defined ∂A to be the topological boundary of A relative to the subspace topology on hyper$\{A\}$. It is the case that $\partial A \subset \partial_{\text{top}} A$.

Proof. If $a \in \overset{\circ}{A}$ there exists an open set U_a such that $a \in U_a \subset A$. This implies $a \in A \setminus \partial_{\text{top}} A$. Conversely, if $a \in A \setminus \partial_{\text{top}} A$ then $a \in A$ and there exists an open set U_a such that $a \in U_a \subset A$. This implies $a \in \overset{\circ}{A}$. □

Lemma 2.6.2.3. *Let $A \subset V$, V being a Banach Space. Then $\partial A \subset \partial_{top} A$.*

Proof. If $a \in \partial_A$ and U_a is an open set in V that contains a, both $U_a \cap A$ and $U_a \cap \text{hyper}\{A\} \setminus A$ are non-empty. But then $U_a \cap V \setminus A$ is non-empty as well. □

We return to proving 3(a). Lemma 2.6.2.2 (page 258) implies

$$\overset{\circ}{C} = C \setminus \partial_{\text{top}} C. \tag{2.202}$$

Lemma 2.6.2.3 (page 259) implies $\partial C \subset \partial_{\text{top}} C$, so

$$C \setminus \partial_{\text{top}} C \subset C \setminus \partial C. \tag{2.203}$$

Combining this proposition's 2(b), (2.202) and (2.203) proves 3(a).

(b) is a trivial consequence of 2(b) and 3(a) of this proposition.

□

2.6.3 $N(PP') \leq N(P)N(P')$

We will use Birkhoff's notation, θ, for the Hilbert Metric d_H:

$$\theta(f, g; C) = \theta(f, g) = d_H(f, g)$$

in this and the following section. Recall

$$N(P; C) = N(P) = \sup_{0<\theta(f,g)<\infty} \frac{\theta(fP, gP)}{\theta(f, g)}$$

(see Section 2.3 (page 155).

Birkhoff [12] states the following without proof (which we supply):

Lemma 2.6.3.1. $N(PP') \leq N(P)N(P')$

Proof.

$$N(P) = \sup_{0<\theta(f,g)<\infty} \frac{\theta(fP, gP)}{\theta(f, g)}$$

Let f_n, g_n be such that

$$N(PP') = \lim_{n\to\infty} \frac{\theta(f_nPP', g_nPP')}{\theta(f_n, g_n)}$$

then for each n:

$$\begin{aligned}\frac{\theta((f_nP)P', (g_nP)P')}{\theta((f_n, g_n)} &= \frac{\theta((f_nP)P', (g_nP)P')}{\theta(f_nP, g_nP)} \ \frac{\theta((f_nP), (g_nP))}{\theta(f_n, g_n)} \\ &\leq N(P')N(P).\end{aligned}$$

□

2.6.4 Proof of Birkhoff's Projective Contraction Theorem

We finally come to Birkhoff's Projective Contraction Theorem [12] which is, word for word:

Theorem 2.6.4.1. *Birkhoff [12]*

Theorem 1 (Projective Contraction Theorem). *Let $N(P^r; C) < 1$ for some r, and let C be complete relative to $\theta(f, g; C)$. Then for any $f \in C$, the sequence fP^n converges geometrically to a unique fixpoint (characteristic ray) $c \in C$.*

Remark 2.6.4.2. This proof is fairly close to Birkhoff's – however the following modifications or clarifications are required:

1. Birkhoff uses the notation $\theta(f, g; C)$ to refer to the Hilbert Projective Metric on the cone C. I.e. $\theta(f, g; C) = d_H(f, g)$. In this proof we will keep Birkhoff's notation.

2. We will assume in this proof that $fP^n \neq 0$ for all positive integers n as otherwise counter examples exist, see Section 2.2.2 (page 151).

3. We will always remove the origin from subsets of C if we are taking their θ 'hyperbolic' diameter. The reason for this is that θ is not defined at the origin.

4. We will interpret "let C be complete" as meaning that $(C \setminus \{0\}, \sim)$ is complete [30] with respect to θ.

5. Recall that $\theta = d_H$ is an extended metric on $(C \setminus \{0\}, \sim)$ meaning that it may take the value $+\infty$ for some pairs $[f], [g] \in (C \setminus \{0\}, \sim)$. Being extended does not interfere [31] with the definition of completeness.

Proof. If the linear map P maps C to itself, then P^r is linear and maps C to itself. So, by Lemma 1 of Birkhoff [12] or by our Section 2.3.6 (page 172)

$$N(P^r; C) = \tanh(\triangle/4) \in [0, 1] \tag{2.204}$$

where $\triangle$ is the $d_H = \theta$ 'hyperbolic' diameter of CP^r.

By elementary properties of tanh we have

$$0 \leq \tanh(\triangle/4) < 1 \qquad \text{if and only if} \qquad 0 \leq \triangle < \infty.$$

[30] $(C \setminus \{0\}, \sim)$ is complete means that if $\{[a_n]\}_1^\infty$ is a cauchy sequence in $(C \setminus \{0\}, \sim)$ with respect to $\theta = d_H$ then there exists an $[a] \in (C \setminus \{0\}, \sim)$ such that $[a_n]$ converges to $[a]$ with respect to $\theta = d_H$.

[31] If $(C \setminus \{0\}, \sim)$ is complete then for each $f \in C \setminus \{0\}$ the set $C_f = \{\, [g] \in (C \setminus \{0\}, \sim) \mid d_H([f], [g]) < \infty \,\}$ is complete because if $\{[g_n]\}_1^\infty \subset C_f$ is cauchy then $\lim_{n\to\infty}[g_n] = [g]$ for some $[g] \in (C \setminus \{0\}, \sim)$. But $[g]$ can't be infinitely far from the $[g_n]$, so $[g] \in C_f$. Since C_f is complete it is closed.

So the condition $N(P^r;\ C) < 1$ implies $\triangle$ is finite.

Let j be any positive integer. $CP \subset C$ implies $CP^{r+j} \subset CP^r$. So both $fP^r, fP^{r+j} \in CP^r$. This implies $\theta(fP^r, fP^{r+j};\ C) \leq \triangle < \infty$.

Let n be any positive integer bigger than r and let q be the largest integer such that $qr \leq n$. So

$$qr \ \leq n \ < \ (q+1)r \ = \ qr + r$$

which implies

$$0 \ \leq \ n - qr \ < \ r.$$

Also, since $n > r$, we have $q \geq 1$.

The inequalities (2.205) to (2.206) make use of the definition [32] of $N(P;C)$:

$$N(P;C) = N(P) = \sup_{0<\theta(f,g)<\infty} \frac{\theta(fP, gP)}{\theta(f,g)}$$

and Lemma 2.6.3.1 (page 260):

$$N(PP') \leq N(P)N(P').$$

$$\theta(fP^n, fP^{n+j};\ C) = \theta(fP^r\ \underbrace{P^r \ldots P^r}_{q-1}\ P^{n-qr},\ fP^{r+j}\ \underbrace{P^r \ldots P^r}_{q-1}\ P^{n-qr};\ C) \quad (2.205)$$

$$\leq N(\underbrace{P^r \ldots P^r}_{q-1}\ P^{n-qr};\ C)\ \ \theta(fP^r, fP^{r+j};\ C)$$

$$\leq (\ \underbrace{N(P^r;C)}_{<1}\ )^{q-1}\ \ \underbrace{N(P^{n-qr};\ C)}_{\leq 1}\ \ \underbrace{\theta(fP^r, fP^{r+j};\ C)}_{\leq \triangle < \infty}$$

$$\leq N(P^r;\ C)^{q-1}\ \ \theta(fP^r, fP^{r+j};\ C)$$

$$\leq N(P^r;\ C)^{q-1}\ \triangle \quad (2.206)$$

[32] $N(P;C)$ is discussed in Section 2.3 (page 155)

The inequalities (2.205) to (2.206) imply that the sequence $\{[fP^n]\}_{n=1}^{\infty}$ is cauchy in $(C\backslash\{0\}, \sim)$ with respect to $\theta = d_H$. We are assuming that $(C\backslash\{0\}, \sim)$ is complete with respect to $\theta = d_H$. So the sequence $\{[fP^n]\}_{n=1}^{\infty}$ converges to a unique $[c] \in (C\backslash\{0\}, \sim)$. See [33].

We are assuming that $fP^n \neq 0$ for each positive integer n. It follows that $fP^n \in C \setminus \ker P$ for each non-negative integer n. See [34].

We are assuming that $(C \setminus \{0\}, \sim)$ is complete. By Lemma 2.6.1.1 (page 250) $(C \setminus \ker P,\ \sim)$ is a closed. So $(C \setminus \ker P,\ \sim)$ is complete. So $[c] \in (C \setminus \ker P,\ \sim)$.

The following argument shows that $[c] = [c]P$. See [35].

By Lemma 2.6.1.1 (page 250) the map induced by P is continuous with respect to $\theta = d_H$ on $(C \setminus \ker P,\ \sim)$. So

$$\begin{aligned}
\theta(c, cP; C) &= \theta(\lim_{n\to\infty} fP^n, (\lim_{n\to\infty} fP^n)P; C) \\
&= \theta(\lim_{n\to\infty} fP^n, \lim_{n\to\infty} (fP^nP); C) \\
&= \theta(c, c; C) \\
&= 0 \\
\Rightarrow [c] &= [cP]. \qquad (2.207)
\end{aligned}$$

Birkhoff writes in his proof $\|fP^n - c\| < K\rho^n$. However that statement, as written, seems to need clarification; see Section 2.6.5 (page 266) below. For now we will show:

$$\theta(fP^n, c; C) < K_1\ \rho^n \qquad (2.208)$$

by reusing the argument given at the beginning of this proof; see line (2.205).

[33] Birkhoff simply writes $c \in C$ identifying $[c]$ with c. Since $\theta([f],[g];\ C) = \theta(f, g;\ C)$ for all $f, g \in C \setminus \{0\}$ this identification is quite valid when working with θ. We take c to be a specific, but arbitrarily chosen element in $\lim_{n\to\infty}[fP^n]$. So (obviously) $c \in [c]$.

[34] $fP^n \neq 0 \Rightarrow fP^{n-1} \in C \setminus \ker P$. As $n > 0$ is arbitrary, we have $fP^n \in C \setminus \ker P$ for all non-negative integers n.

[35] $[c]P = [cP]$ see Section 1.7.1 (page 67).

Let n,q, r be as before, recall q is the largest integer such that $qr \leq n$. We define R implicitly by $n = qr + R$ so $0 \leq R < r$.

By (2.207) we have $[c] = [cP]$ so $[c] = [cP^r] = [cP^n]$. So

$$\begin{aligned}
\theta(fP^n,\ c;\ C) &= \theta(fP^n,\ cP^n;\ C) \\
&= \theta(fP^r\ \underbrace{P^r \dots P^r}_{q-1}\ P^{n-qr},\ cP^r\ \underbrace{P^r \dots P^r}_{q-1}\ P^{n-qr};\ C) \\
&\leq (\ \underbrace{N(P^r,\ C)}_{<1}\ )^{q-1}\ \underbrace{N(P^{n-qr};\ C)}_{\leq 1}\ \underbrace{\theta(fP^r, cP^r;\ C)}_{\leq \triangle \text{ since } fP^r, cP^r \in CP^r \setminus \{0\}} \\
&\leq N(P^r, C)^{q-1}\ \triangle. \qquad (2.209)
\end{aligned}$$

The exponent $q-1$ in (2.209), immediately above, increases by 1 each time n increases by r, since q is the largest integer such that $qr \leq n$. Since $n = qr + R$ with $0 \leq R < r$, we have:

$$\frac{n}{r} = q + \frac{R}{r} \text{ with } 0 \leq \frac{R}{r} < 1. \qquad (2.210)$$

If $N(P^r;\ C) = 0$ then (2.209) implies $\theta(fP^n, c; C) = 0$ and we're done. So we will assume that $N(P^r;\ C) \neq 0$. Since $0 < N(P^r;\ C) < 1$ it follows from (2.210) that

$$\frac{R}{r} - 1 < 0 \ \text{ and } \ N(P^r, C)^{\frac{R}{r}-1} > 1$$

so

$$\begin{aligned} N(P^r,C)^{-2}\left(N(P^r,C)^{1/r}\right)^n &= N(P^r,C)^{\frac{n}{r}-2} \\ &= N(P^r,C)^{q+\frac{R}{r}-2} \\ &= N(P^r,C)^{q-1+\frac{R}{r}-1} \\ &= N(P^r,C)^{q-1}\underbrace{N(P^r,C)^{\frac{R}{r}-1}}_{>1} \\ \Rightarrow \underbrace{N(P^r,C)^{1-\frac{R}{r}}}_{<1} N(P^r,C)^{-2}\left(N(P^r,C)^{1/r}\right)^n &= N(P^r,C)^{q-1} \\ \Rightarrow N(P^r,C)^{q-1} &< N(P^r,C)^{-2}\left(N(P^r,C)^{1/r}\right)^n. \end{aligned} \tag{2.211}$$

We showed in (2.209) that

$$\theta(fP^n,\ c;\ C) \leq N(P^r,C)^{q-1}\ \triangle. \tag{2.212}$$

Plugging (2.211) into (2.212) yields

$$\theta(fP^n,\ c;\ C) < N(P^r,C)^{-2}\ \left(\underbrace{N(P^r,C)^{1/r}}_{\rho}\right)^n\ \triangle. \tag{2.213}$$

Rearranging (2.213) yields

$$\theta(fP^n,\ c;\ C) < \underbrace{N(P^r,C)^{-2}\ \triangle}_{=\ K_1\ <\ \infty}\ \left(\underbrace{N(P^r,C)^{1/r}}_{\rho}\right)^n. \tag{2.214}$$

So we've shown:

$$\theta(fP^n,c;C) < K_1\ \rho^n \tag{2.215}$$

where $\rho = N(P^r,C)^{1/r}$, and $K_1 = N(P^r,C)^{-2}\triangle\ <\ \infty$.

We finish by copying the last few lines of Birkhoff's proof (Birkhoff [12]):

The uniqueness of [the fixed point] c is immediate since, since $cP = C$ and $c^*P = c^*$ imply

$$\theta(c, c^*; C) = \theta(cP, c^*P; C) \leq N(P^r, C)\theta(c, c^*; C).$$

Since $N(P^r, C) < 1$, this implies $\theta(c, c^*; C) = 0$ [hence $c = c^*$].

More precisely, it implies $[c] = [c*]$ and we are done. □

2.6.5 Clarification of $\|fP^n - c\| < K\rho^n$ in Birkhoff's Projective Contraction Theorem.

We have shown in the proof of Birkhoff's Projection Contraction Theorem, see inequality (2.214) (page 265), that for all $n > r$ that

$$\theta(fP^n, c; C) < K_1\ \rho^n \tag{2.216}$$

where $0 \leq \rho = N(P^r, C)^{1/r} < 1$, and $K_1 = N(P^r, C)^{-2}\triangle\ <\ \infty$.

However, Birkhoff states in his proof of his Projective Contraction Theorem, that

$$\|fP^n - c\| < K\rho^n \tag{2.217}$$

where $\rho = N(P^r, C)^{1/r}$, and $K < \infty$.

The problem is this: $C \subset V$, where V is (complete) normed linear space with norm $\|\ \|$. So it makes sense to write $\|fP^n - c\|$. However, Birkhoff's statement, that $\|fP^n - c\| < K\rho^n$, see (2.208), is clearly not true since c is of fixed length, and fP^n can grow without bound for many linear maps, P, even if P is bounded.

Example 1. Let $C = \mathbb{R}^2_{>0}$, with $\|(x, y)\| = \sqrt{x^2 + y^2}$, the usual Euclidean

Norm. Let

$$P = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

and let $f = c = (1, 1)$.

Note, c is in the unique positive eigen direction of P. The eigenvalue corresponding to c is 3. Hence, $cP^n = fP^n = 3^n(1, 1)$ and

$$\|fP^n - c\| = \|3^n(1, 1) - (1, 1)\| = \left\|3^{n-1}(1, 1)\right\| = 3^{n-1}\sqrt{2}$$

.

So the question arises, what is the correct way to interpret fP^n, c, and $\|fP^n - c\|$ so that $\|fP^n - c\| < K\rho^n$ is true.

One possible solution to this dilemma would be to treat the fP^n as vectors in $C \subset V$, but to identify c with the line $l_c = \{tc \in V : t \in \mathbb{R}\}$. Then $\|fP^n - c \|$ might be calculated as

$$\|fP^n - c \| = \min_{t \in \mathbb{R}} \|fP^n - tc \| = d(fP^n, l_c),$$

where $d(fP^n, l_c)$ is the distance from the "point" fP^n to the line l_c. It turns out that this possible solution won't work as the following calculations will show.

2.6.6 Distance formula: a point to a line in $\mathbb{R}^n$ passing through the origin

It is easy enough to compute the Euclidean distance d from a point

$$x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

to a line l_v, where l_v is the line in the direction v passing through the origin.

The line l_v is just the set of points $l_v = \{tv : t \in R\}$, so the distance from a typical point tv, on the line l_v, to the point x, is:

$$d(x, tv) = \sqrt{\sum_{i=1}^{n} (x_i - tv_i)^2}.$$

To minimize $d(tv, x)$ with respect to t, it suffices to minimize the $\sum_{i=1}^{n}(x_i - tv_i)^2$, since the square root function is increasing. Setting the derivative equal to zero, we find the value for t which will minimize $d(x, tv)$:

$$\begin{aligned}
\frac{d}{dt} \sum_{i=1}^{n}(x_i - tv_i)^2 &= 2\sum_{i=1}^{n}(x_i - tv_i)(-v_i); \\
0 &= 2\sum_{i=1}^{n}(x_i - tv_i)(-v_i) \\
&= \sum_{i=1}^{n}(x_i - tv_i)(v_i) \\
&= \sum_{i=1}^{n} x_i v_i - t\sum_{i=1}^{n} v_i v_i \\
\Rightarrow \quad \frac{\sum_{i=1}^{n} x_i v_i}{\sum_{i=1}^{n} v_i^2} &= t.
\end{aligned}$$

So we get the distance formula

$$d(x, l_v) = \min_{t \in \mathbb{R}} d(x, tv) = \sqrt{\sum_{k=1}^{n} \left(x_k - v_k \frac{\sum_{i=1}^{n} x_i v_i}{\sum_{i=1}^{n} v_i^2} \right)^2} \tag{2.218}$$

giving the minimum distance from the point x to the line l_v which passes through the origin and is in the v direction.

2.6.7 The distance from fP^n to a line in $\mathbb{R}^2$

We apply formula (2.218), which gives the the distance between a point and a line, to the point $fP^n = (x_n, y_n) \in \mathbb{R}^2$ and the line $l_c = l_{(1,1)}$.

$$\begin{aligned} d\left(fP^n, l_c\right) = d\left((x_n, y_n), l_{(1,1)}\right) &= \sqrt{\left(x_n - 1\frac{x_n 1 + y_n 1}{1^2 + 1^2}\right)^2 + \left(y_n - 1\frac{x_n 1 + y_n 1}{1^2 + 1^2}\right)^2} \\ &= \sqrt{\left(x_n - \frac{x_n + y_n}{2}\right)^2 + \left(y_n - \frac{x_n + y_n}{2}\right)^2} \\ &= \sqrt{\left(\frac{x_n - y_n}{2}\right)^2 + \left(\frac{y_n - x_n}{2}\right)^2} \\ &= \sqrt{2\left|\frac{x_n - y_n}{2}\right|^2} \\ &= \sqrt{2}\left|\frac{x_n - y_n}{2}\right|. \end{aligned} \tag{2.219}$$

Example 2. Let $C = \mathbb{R}^2_{>0}$, with the usual Euclidean Norm; let

$$P = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix},$$

the same as in Example 1 (page 266). Let $f = (x_0, y_0)$ and let $fP^n = (x_n, y_n)$ for $n = 1, 2, \ldots$ and let $c = (1, 1)$. As remarked in Example 1, c is in the unique positive eigen direction of P. P acts on the vector $f = (x_0, y_0)$ as follows:

$$(x_0, y_0)P = (2x_0 + 1y_0,\ 1x_0 + 2y_0) = (x_1, y_1),$$

but then

$$x_1 - y_1 = (2x_0 + 1y_0) - (1x_0 + 2y_0) = x_0 - y_0$$

and by induction

$$x_n - y_n = x_0 - y_0.$$

So, (2.219) becomes

$$\begin{aligned} d\left(fP^n, l_c\right) &= d\left((x_n, y_n), l_{(1,1)}\right) \\ &= \sqrt{2}\ \left|\frac{x_n - y_n}{2}\right| \\ &= \sqrt{2}\ \left|\frac{x_0 - y_0}{2}\right| \end{aligned}$$

which is constant with respect to n. For concreteness, if $f = (x_0, y_0) = (2, 1)$ then

$$\begin{aligned} d\left(fP^n, l_c\right) &= \sqrt{2}\ \left|\frac{x_0 - y_0}{2}\right| \\ &= \sqrt{2}\ \left|\frac{2 - 1}{2}\right| \\ &= \frac{\sqrt{2}}{2} \end{aligned}$$

regardless of n. So fP^n does not converge to the line l_c with respect to the Euclidean metric even though $[fp^N]$ converges geometrically to $[c]$ with respect to the Hilbert Projective Metric $\theta = d_H$.

2.6.8 An appropriate interpretation of $\|fP^n - c\| < K\rho^n$.

Birkhoff [12] requires the existence of a hyperplane H in V, "cutting each ray of C in exactly one point; we can then discuss C and $C \bigcap H$ interchangeably as subspaces of projective space."

It turns out that there does not always exist such a hyperplane. See Section 2.4.5 (page 193) for a counter-example and see Theorem 2.4.4.2 (page 191) for conditions under which such a hyperplane will exist. If such a hyperplane H exists:

Fix H and identify fP^n and c with their central projections

$$(fP^n)^H = [fP^n] \cap H \qquad \text{and} \qquad c^H = [c] \cap H.$$

Since $H \subset V$, V's norm, $\| \ \|_V$, is defined on H, and so

$$\left\| (fP^n)^H - c^H \right\|_V = d_V \left((fP^n)^H \, , \ c^H \right) \tag{2.220}$$

is well defined. Of course $\left\| (fP^n)^H - c^H \right\|_V$ will depend upon the choice of H.

Let

$$D_{H \cap C} = \text{ diameter} \, (H \cap C)$$

with respect to the norm of V. Lemma 2.5.0.2 (page 232), implies that if fP^n, c are linearly independent then

$$d_V((fP^n)^H, c^H) < \frac{D_{H \cap C}}{4} \ d_H(fP^n, c); \tag{2.221}$$

and of course if they are linearly dependent, then

$$d_V((fP^n)^H, c^H) = d_H(fP^n, c) = 0. \tag{2.222}$$

In our proof of Birkhoff's Projective Contraction Theorem [12], we showed, starting at (2.208) (page 263), that

$$\theta(fP^n, c; C) = \underbrace{d_H(fP^n, c) \ < K_1 \ \rho^n}_{\text{see (2.225) below}}. \tag{2.223}$$

with

$$\rho = N(P^r, C)^{1/r} < 1, \qquad K_1 = N(P^r, C)^{-2} \triangle \ < \ \infty, \tag{2.224}$$

and with $\triangle = \text{diameter} \, (CP^r)$ with respect to d_H.

Combining (2.220), (2.221), (2.222) and (2.223) yields

$$\underbrace{\left\|(fP^n)^H - c^H\right\|_V \leq \frac{D_{H\cap C}}{4}\ \overbrace{d_H(fP^n, c)}^{\text{See (2.223)}}}_{\text{(2.220) combined with (2.221), (2.222)}} < \frac{D_{H\cap C}}{4}\ K_1\ \rho^n. \tag{2.225}$$

Making use of (2.224) we rewrite (2.225) as

$$\begin{aligned}\left\|(fP^n)^H - c^H\right\|_V &< \left(\frac{D_{H\cap C}}{4}\ \frac{\triangle}{N(P^r, C)^2}\right)\ \left(N(P^r, C)^{1/r}\right)^n \\ &= K\rho^n,\end{aligned} \tag{2.226}$$

where we are letting

$$K = \left(\frac{D_{H\cap C}}{4}\ \frac{\triangle}{N(P^r, C)^2}\right). \tag{2.227}$$

Regarding the finiteness of K.

If $N(P^r, C) = 0$ it means that P^r collapses C to single ray, in which case

$$\left\|(fP^n)^H - c^H\right\|_V = 0.$$

Birkhoff assumes that $N(P^r, C) < 1$ in his Projective Contraction Theorem [12]. This assumption implies that $\triangle < \infty$, see[36].

So for K, as defined in (2.227) to be finite it suffices that $0 < N(P^r, C) < 1$ and for $D_{H\cap C}$ = diameter of $H \cap C$ with respect to d_V to be finite.

The case of most interest to us occurs when $C = \mathbb{R}^n_{\geq 0},\ n \geq 2$ and when the hyperplane H is

$$H_1 = \left\{ x \in \mathbb{R}^n \,\middle|\, x = \sum_{i=1}^n x_i e_i,\ x_i \in \mathbb{R},\ \sum_{i=1}^n x_i = 1 \right\}.$$

[36] $N(P^r, C) < 1 \Rightarrow \triangle < \infty$ by Lemma 1 of Birkhoff [12] or by our Section 2.3.6 (page 172) and our comments following (2.204) (page 261).

In this case $D_{C\cap H_1} = \sqrt{2}$ by Theorem 2.4.7.2 (page 197).

2.7 Projective Linear ODE Theorem

Note that in the context of cones, the relation $x \leq y$ means $y - x \in C$.

Proposition 2.7.0.1. *Suppose that*

$$x \leq y \text{ and } x' \leq y'$$

then $(x + x') \leq (y + y')$

Proof.

$$y - x \in C \text{ and } y' - x' \in C.$$

So, since C is closed under addition we have

$$(y - x) + (y' - x') = (y + y') - (x + x') \in C.$$

□

2.7.1 Projective Additivity Lemma

I have not seen the following two result elsewhere.

Lemma 2.7.1.1. Projective Additivity Lemma. *Let* $x, y, z \in C$ *and* $d_H(x, z)$ *and* $d_H(y, z)$ *both be finite. Then*

$$d_H(x + y, z) \leq \max\{d_H(x, z),\ d_H(y, z)\}$$

See Figure 2.9 (page 274).

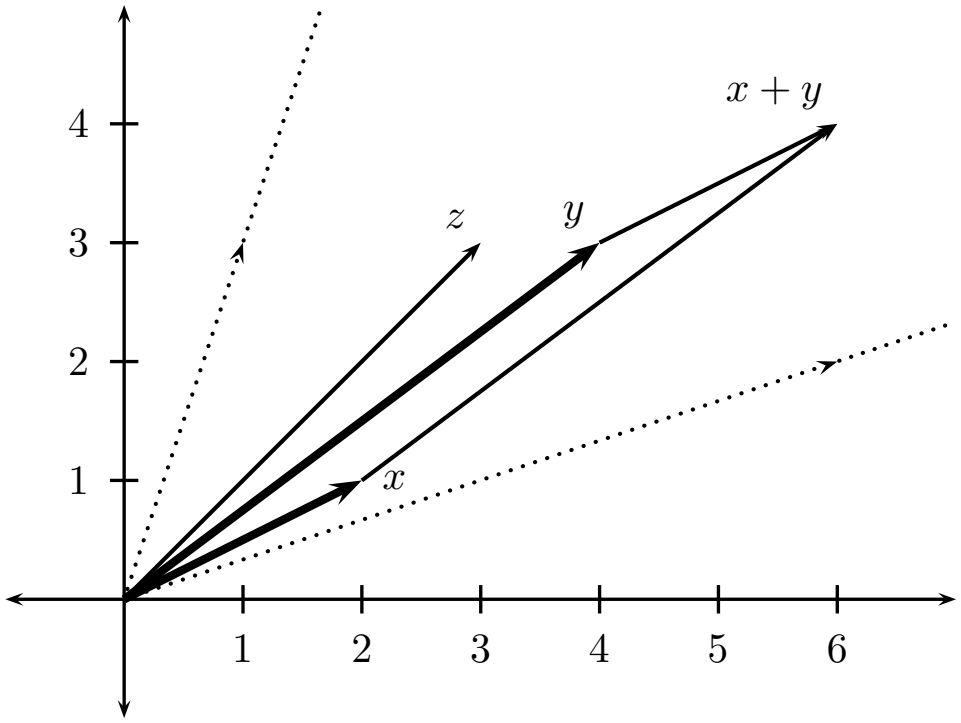


Figure 2.9: Projective Additivity in the standard cone $\mathbb{R}^2_{>0}$. The dotted rays in the directions $(1,3)$ and $(6,2)$ indicate the boundary of $B^H_{\ln(3)}(z)$, the ball of radius $\ln(3)$ centered at $z=(3,3)$. Distances are measured by the Hilbert Project Metric, d_H. If $f,g \in \mathbb{R}^2_{\geq 0}$ then calculating d_H is straight forward and given by the formula $d_H\left((f_x,f_y),(g_x,g_y)\right) = \left|\ln\left(\frac{f_y}{f_x}\ \frac{g_x}{g_y}\right)\right|$. Hence, $d_H(z,(1,3)) = \left|\ln\left(\frac{3}{3}\ \frac{1}{3}\right)\right| = \ln(3)$, and $d_H(z,(6,2)) = \ln\left(\frac{3}{3}\ \frac{6}{2}\right) = \ln(3)$. Note $\ln(3) \approx 1.1$. $B^H_{\ln(3)}(z)$ is a convex cone, closed under addition and positive scaling (Corollary 2.7.1.2 (page 276)). Since both $x=(2,1)$ and $y=(4,3) \in B^H_{\ln(3)}(z)$ it follows that $x+y=(6,4) \in B^H_{\ln(3)}(z)$. We also can see, as guaranteed by Lemma 2.7.1.1 (page 273), that $d_H(x+y,z) \leq \max\{d_H(x,z),d_H(y,z)\} = d_H(x,z)$. In fact, $d_H(x,z) = \left|\ln\left(\frac{1}{2}\ \frac{3}{3}\right)\right| = \ln(2) \approx 0.69$. $d_H(y,z) = \left|\ln\left(\frac{3}{4}\ \frac{3}{3}\right)\right| = \ln(4/3) \approx 0.29$. $d_H(x+y,z) = \left|\ln\left(\frac{4}{6}\ \frac{3}{3}\right)\right| = \ln(3/2) \approx 0.41$.

Proof. Let $\alpha_x,\ \alpha_y,\ \alpha_{x+y}$ be the supremums and $\beta_x,\ \beta_y,\ \beta_{x+y}$ be the infimums for which

$$\alpha_x z \leq x \leq \beta_x z \tag{2.228}$$

$$\alpha_y z \leq y \leq \beta_y z \tag{2.229}$$

$$\alpha_{x+y}\, z \leq x + y \leq \beta_{x+y}\, z$$

hold, with the proviso, that $\inf\{\emptyset\} = \infty$. Since we are assuming that $d_H(x, z)$ and $d_H(y, z)$ are both finite it follows that $\alpha_x,\ \alpha_y,\ \beta_x,\ \beta_y$ are all finite positive numbers. Then, since C is topologically closed, we have

$$x - \alpha_x z,\ y - \alpha_y z,\ \beta_x z - x,\ \beta_y z - y \in C,$$

so the relationships in (2.228) and (2.229) actually hold for the numbers $\alpha_x,\ \alpha_y,\ \beta_x,\ \beta_y$ themselves. Proposition 2.7.0.1 (page 273) applied to the sum of (2.228) and (2.229) yields

$$\alpha_x z + \alpha_y z \leq x + y \leq \beta_x z + \beta_y z. \tag{2.230}$$

(2.230) combined with the definitions of α_{x+y} and β_{x+y} imply:

$$(\alpha_x + \alpha_y) \leq \alpha_{x+y}$$

$$(\beta_x + \beta_y) \geq \beta_{x+y}.$$

So

$$\frac{\beta_{x+y}}{\alpha_{x+y}} \leq \frac{\beta_x + \beta_y}{\alpha_x + \alpha_y}. \tag{2.231}$$

If $\alpha_x, \beta_x, \alpha_y, \beta_y \in (0, \infty)$ then

$$\frac{\beta_x + \beta_y}{\alpha_x + \alpha_y} \leq \frac{\beta_x}{\alpha_x} \Leftrightarrow \tag{2.232}$$

$$\alpha_x(\beta_x + \beta_y) \leq \beta_x(\alpha_x + \alpha_y) \Leftrightarrow$$

$$\alpha_x\beta_x + \alpha_x\beta_y \leq \beta_x\alpha_x + \beta_x\alpha_y \Leftrightarrow$$

$$\alpha_x\beta_y \leq \beta_x\alpha_y \Leftrightarrow$$

$$\frac{\beta_y}{\alpha_y} \leq \frac{\beta_x}{\alpha_x}. \tag{2.233}$$

If we reverse the direction of all the inequalities, starting at (2.232) and ending at (2.233), the logical equivalences will still be true. Hence,

$$\min\left\{\frac{\beta_x}{\alpha_x},\ \frac{\beta_y}{\alpha_y}\right\} \leq \frac{\beta_x + \beta_y}{\alpha_x + \alpha_y} \leq \max\left\{\frac{\beta_x}{\alpha_x},\ \frac{\beta_y}{\alpha_y}\right\}. \tag{2.234}$$

Combining (2.231) and (2.234) yields

$$\frac{\beta_{x+y}}{\alpha_{x+y}} \leq \max\left\{\frac{\beta_x}{\alpha_x},\ \frac{\beta_y}{\alpha_y}\right\}. \tag{2.235}$$

The ln function is monotonically increasing and so (2.235) implies

$$\begin{aligned} d_H(x+y, z) = \ln\left(\frac{\beta_{x+y}}{\alpha_{x+y}}\right) &\leq \max\left\{\ln\left(\frac{\beta_x}{\alpha_x}\right),\ \ln\left(\frac{\beta_y}{\alpha_y}\right)\right\} \\ &= \max\left\{d_H(x, z),\ d_H(y, z)\right\}. \end{aligned}$$

□

Corollary 2.7.1.2. *Let $B_r^H(z)$ be the open Hilbert Projective ball of radius r centered at $z \in C,\ z \neq 0$. Suppose that $x, y \in B_r^H(z)$ and $\alpha, \beta \geq 0$ with at least one of $\alpha, \beta > 0$. Then*

$$\alpha x + \beta y \in B_r^H(z).$$

So $B_r^H(z)$ is a convex cone closed under addition and positive scaling.

See Figure 2.9 (page 274).

Proof. If $\alpha > 0$ then $d_H(\alpha x, z) = d_H(x, z) < r$ and so $\alpha x \in B_r^H(z)$. A similar argument shows $\beta > 0$ implies $\beta y \in B_r^H(z)$. If both α and $\beta > 0$, then both αx and $\beta y \in B_r^H(z)$, implying both $d_H(\alpha x, z)$ and $d_H(\beta y, z) < r$. But then Lemma 2.7.1.1 (page 273) immediately implies $d_H(\alpha x + \beta y, z) < r$. □

2.7.2 Poisson Tail Lemma

I have not seen the following lemma elsewhere.

Lemma 2.7.2.1. Poisson Tail Lemma. *Let A be a primitive $n \times n$ non-negative matrix; i.e. the entries of A are all non-negative and the entries of A^q are all positive for some integer $q > 0$. Let $Y \neq 0$ be a non-negative column vector of dimension n and let m be any integer ≥ 0. Then*

$$\lim_{t\to\infty} \frac{\sum_{k=0}^{m} \frac{t^k}{k!} A^k Y}{\|e^{tA}Y\|_1} = 0 \tag{2.236}$$

and

$$\lim_{t\to\infty} \frac{\left\|\sum_{k=m+1}^{\infty} \frac{t^k}{k!} A^k Y\right\|_1}{\|e^{tA}Y\|_1} = 1. \tag{2.237}$$

Note: A nice illustration of (2.236) is found in Figure 2.10 (page 278). In Figure 2.10 we consider the one dimensional case where the matrix A is just a single non-negative number λ so that e^{At} is just $e^{\lambda t}$. We take $Y = 1$. If we expand $e^{\lambda t}$:

$$e^{\lambda t} = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!}$$

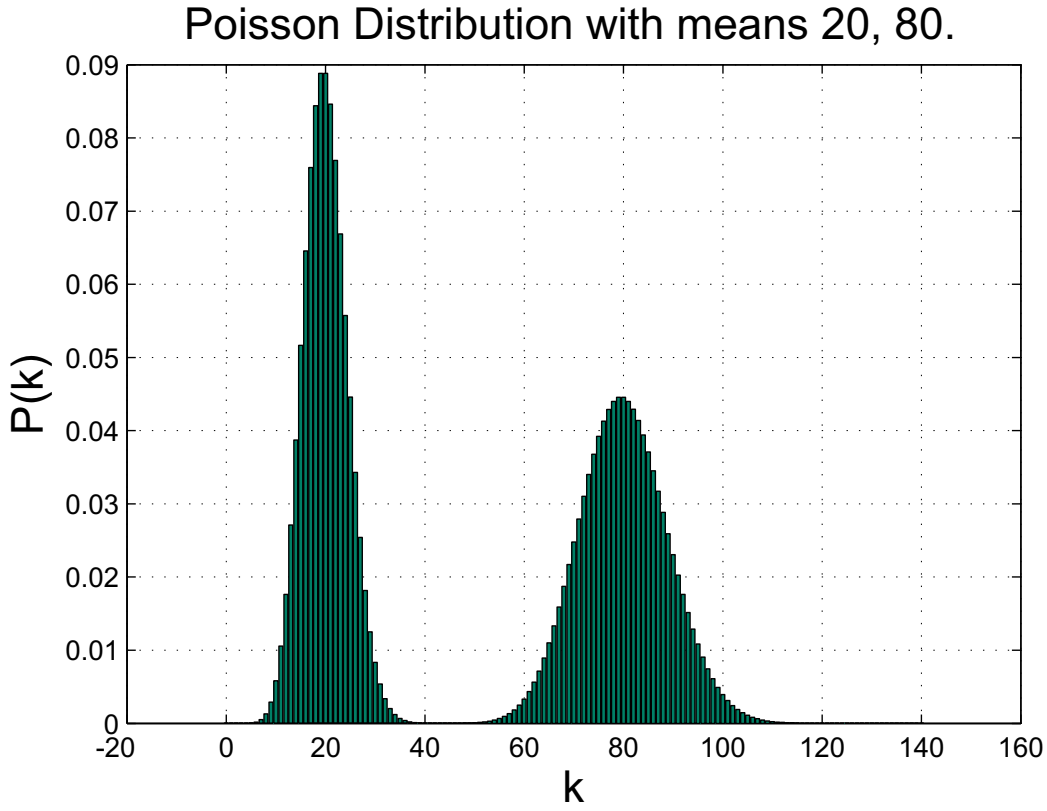


Figure 2.10: Poisson Distribution $P(k) = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t}$. Two Poisson Distributions plotted. One with a mean of $\lambda t = 20$ (left), and one with a mean of $\lambda t = 80$ (right). Notice that almost all the mass is centered around the mean and that there is almost no mass in the tails.

and then divide this expansion of $e^{\lambda t}$ by $e^{\lambda t}$, we get

$$1 = \frac{e^{\lambda t}}{e^{\lambda t}} = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}. \tag{2.238}$$

The interesting thing is that the k^{th} term of (2.238), $e^{-\lambda t}\frac{(\lambda t)^k}{k!}$, is $P(X = k)$, where X is a random variable with mean λt and Poisson Distribution. In Figure 2.10 we plot the Poisson Distributions having means $\lambda t = 20$ and $\lambda t = 80$. From the figure it seems reasonable to guess that as $t \to \infty$ forces $\lambda t \to \infty$, the mass in the left tail,

$$\frac{\sum_{k=0}^{m} \frac{t^k}{k!}\ \lambda^k}{e^{\lambda t}}$$

goes to zero. The proof of this is to simply apply l'Hopital's rule m times.

Proof. The following argument shows that Lemma 2.7.2.1 is true.

Let a_{ij}^k represents the ij entry in the matrix A^k. Note: $A^0 = I$, the identity matrix, $a_{ij}^0 = \delta_{ij}$ and $a_{ij}^k \neq (a_{ij})^k$. It is convenient to use matrix notation to describe the column vector Y, so the i^{th} entry of $Y = y_{i1}$.

Since we are assuming that A is primitive there exists a $q > 0$ such that all the entries of A^q are positive. Let $r > 0$ be an integer such that $qr > m$. Then the matrix $A^{qr} = (A^q)^r$ has all positive entries. Let $qr = Q$, so $Q > m$, A^Q has all positive entries; explicitly $a_{ij}^Q > 0$ for all $i, j, \quad 1 \leq i, j \leq n$.

Since we are assuming that $Y \neq 0$ is non-negative, $Y_{j1} > 0$ for at least one value of j. But then the i^{th} row (and entry) of the column vector $A^Q Y$ is

$$\left(A^Q Y\right)_{i1} = \left(\sum_{j=1}^{n} a_{ij}^Q Y_{j1}\right) > 0 \quad \forall\ i,\ i = 1, 2, \ldots, n. \tag{2.239}$$

So we've shown that all n arguments of the vector $A^Q Y$ are positive.

Since $a_{ij}^k \geq 0$ for all $i, j; \quad 1 \leq i, j \leq n$, and since $Y_{i1} \geq 0$ for all $i; \quad 1 \leq i \leq n$, it

follows from (2.239) that for $t > 0$

$$\begin{aligned}
\left\|e^{tA}Y\right\|_1 &= \left\|\left(\sum_{k=0}^{\infty}\frac{t^k}{k!}\ A^k\right)Y\right\|_1 \\
&= \left\|\sum_{k=0}^{\infty}\left(\frac{t^k}{k!}\ A^kY\right)\right\|_1 \\
&= \sum_{i=1}^{n}\left|\sum_{k=0}^{\infty}\left(\frac{t^k}{k!}\ A^kY\right)_{i1}\right| \\
&= \sum_{i=1}^{n}\left|\sum_{k=0}^{\infty}\frac{t^k}{k!}\ \sum_{j=1}^{n}a_{ij}^kY_{j1}\right| \\
&= \sum_{i=1}^{n}\left(\sum_{k=0}^{\infty}\frac{t^k}{k!}\ \sum_{j=1}^{n}a_{ij}^kY_{j1}\right) \\
&\geq \left(\frac{t^Q}{Q!}\sum_{j=1}^{n}a_{ij}^QY_{j1}\right) \\
&> 0.
\end{aligned}$$

But then

$$\begin{aligned}
0 &\leq \left(\frac{\sum_{k=0}^{m}\frac{t^k}{k!}\ A^kY}{\left\|e^{tA}Y\right\|_1}\right)_{i1} \\
&= \frac{\sum_{k=0}^{m}\frac{t^k}{k!}\ \sum_{j=1}^{n}a_{ij}^kY_{j1}}{\sum_{i=1}^{n}\left|\sum_{k=0}^{\infty}\frac{t^k}{k!}\ \sum_{j=1}^{n}a_{ij}^kY_{j1}\right|} \\
&\leq \frac{\sum_{k=0}^{m}\frac{t^k}{k!}\ \sum_{j=1}^{n}a_{ij}^kY_{j1}}{\frac{t^Q}{Q!}\left(\sum_{j=1}^{n}a_{ij}^QY_{j1}\right)}.
\end{aligned} \tag{2.240}$$

To find the $\lim_{t\to\infty}$ of the last expression in (2.240) we can apply l'Hopital's rule until the limit becomes apparent. This will happen after m' iterations, with $m' \leq m$, since the numerator is a polynomial of degree at most m, and the denominator is a polynomial of degree $Q > m$. See the note following this proof for an example illustrating why the need to introduce the integer m' with $0 \leq m' \leq m$. After

differentiating the denominator m' times the denominator becomes

$$\frac{t^{Q-m'}}{(Q-m')!}\underbrace{\left(\sum_{j=1}^{n} a_{ij}^{Q} Y_{j1}\right)}_{>0}.$$

Since $Q > m$ and $m \geq m'$ it follows that $Q - m' > 0$. So we have

$$\lim_{t\to\infty} \frac{t^{Q-m'}}{(Q-m')!}\underbrace{\left(\sum_{j=1}^{n} a_{ij}^{Q} Y_{j1}\right)}_{>0} = \infty.$$

Putting this all together yields:

$$\begin{aligned}
0 &\leq \lim_{t\to\infty} \left(\frac{\sum_{k=0}^{m} \frac{t^k}{k!}\, A^k Y}{\left\|e^{tA}Y\right\|_1}\right)_{i1} \\
&\leq \lim_{t\to\infty} \frac{\sum_{k=0}^{m} \frac{t^k}{k!} \sum_{j=1}^{n} a_{ij}^{k} Y_{j1}}{\frac{t^Q}{Q!}\left(\sum_{j=1}^{n} a_{ij}^{Q} Y_{j1}\right)} \\
&= \lim_{t\to\infty} \frac{\sum_{j=1}^{n} a_{ij}^{m'} Y_{j1}}{\frac{t^{Q-m'}}{(Q-m')!}\left(\sum_{j=1}^{n} a_{ij}^{Q} Y_{j1}\right)} \\
&= \frac{\sum_{j=1}^{n} a_{ij}^{m'} Y_{j1}}{\infty} \\
&= 0,
\end{aligned}$$

which implies (2.236) of this lemma.

The proof of (2.237)

$$\lim_{t\to\infty} \frac{\left\|\sum_{k=m+1}^{\infty} \frac{t^k}{k!}\, A^k Y\right\|_1}{\left\|e^{tA}Y\right\|_1} = 1$$

follows from the following observation about $\|\ \|_1$:

Let $X = \sum_{i=1}^{n} x_i e_i$ and $Y = \sum_{i=1}^{n} y_i e_i \in \mathbb{R}^n$, with e_i being the standard basis vectors for $\mathbb{R}^n$, and suppose that for each i that $y_i \geq x_i \geq 0$. Then $\|Y - X\|_1 =$

$\|Y\|_1 - \|X\|_1$. The proof of this observation is:

$$\begin{aligned}
\|Y - X\|_1 &= \sum_{i=1}^{n} |y_i - x_i| \\
&= \sum_{i=1}^{n} (y_i - x_i) \\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \\
&= \|Y\|_1 - \|X\|_1 .
\end{aligned}$$

Applying this observation to (2.237) yields:

$$\begin{aligned}
\lim_{t\to\infty} \frac{\left\|\sum_{k=m+1}^{\infty} \frac{t^k}{k!}\ A^k Y\right\|_1}{\|e^{tA}Y\|_1} &= \lim_{t\to\infty} \left(\frac{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!}\ A^k Y\right\|_1}{\|e^{tA}Y\|_1} - \frac{\left\|\sum_{k=0}^{m} \frac{t^k}{k!}\ A^k Y\right\|_1}{\|e^{tA}Y\|_1} \right) \\
&= \lim_{t\to\infty} \left(\frac{\left\|e^{tA}Y\right\|_1}{\|e^{tA}Y\|_1} - \frac{\left\|\sum_{k=0}^{m} \frac{t^k}{k!}\ A^k Y\right\|_1}{\|e^{tA}Y\|_1} \right) \\
&= 1 - \lim_{t\to\infty} \frac{\left\|\sum_{k=0}^{m} \frac{t^k}{k!}\ A^k Y\right\|_1}{\|e^{tA}Y\|_1} \\
&= 1,
\end{aligned}$$

by (2.236). □

Note about why $m' \leq m$. It is possible to find a non-negative primitive matrix A; an integer m; a vector Y; and index i such that $(A^m Y)_{i1} = 0$. For example if A is the 3×3 matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{then} \quad A^2 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

so we see that A is a non-negative primitive matrix. If we choose $m = 1, i = 2$, and

$$Y = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{then} \quad A^1 Y = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

so we see that $\left(A^1 Y\right)_{21} = 0$. With these choices of A, m, i, Y and $Q = 2$ Finding the

$$\lim_{t \to \infty} \left(\frac{\sum_{k=0}^{m} \frac{t^k}{k!} A^k Y}{\left\| e^{tA} Y \right\|_1} \right)_{i1}, \tag{2.241}$$

which comes from line (2.240) of the above proof, requires we apply l'Hopital's rule $m' = 0$ times (zero times). We can see this explicitly as (2.241) becomes, with our choices for A, i, etc:

$$\begin{aligned} \lim_{t \to \infty} \frac{\frac{t^0}{0!}\left(a_{21}^0 Y_{11} + a_{21}^0 Y_{21}\right) + \frac{t^1}{1!}\left(a_{21}^1 Y_{11} + a_{21}^1 Y_{21}\right)}{\frac{t^2}{2!}\left(a_{21}^2 Y_{11} + a_{21}^2 Y_{21}\right)} &= \\ \lim_{t \to \infty} \frac{\frac{t^0}{0!}\left(0 \cdot 0 + 1 \cdot 1\right) + \frac{t^1}{1!}\left(1 \cdot 0 + 0 \cdot 1\right)}{\frac{t^2}{2!}\left(1 \cdot 0 + 1 \cdot 1\right)} &= \\ \lim_{t \to \infty} \frac{\frac{t^0}{0!}\left(1\right) + \frac{t^1}{1!}\left(0\right)}{\frac{t^2}{2!}\left(1\right)} &= \\ 0. \end{aligned} \tag{2.242}$$

No need for l'Hopital's rule. In particular, $m = 1$, yet the degree of the polynomial appearing in the numerator of (2.242) is 0; i.e. $m' = 0 < m = 1$. So we apply l'Hopital's rule $m' = 0$ times in this example.

2.7.3 Projective Linear ODE Theorem

I have not seen the following theorem elsewhere, though somewhat similar results certainly exist; see especially [14]. The proof seems original.

Theorem 2.7.3.1. PROJECTIVE LINEAR ODE THEOREM. *Let A be a primitive*

$n \times n$ matrix; i.e. the entries of A are non-negative and the entries of A^Q are positive for some integer $Q > 0$. Let

$$\dot{X} = AX$$

be a system of linear ODE's with initial condition $X(0) = X_0$, with $X(0) \neq 0$ and non-negative. Let v_p be the unique eigenvector of A with all positive entries and l^1 norm 1. Then

$$\lim_{t \to \infty} \frac{X}{\|X\|_1} = v_p,$$

where convergence of this limit is with respect to to the Euclidean Metric d_E.

Proof. Let $\epsilon > 0$ be given.

By Birkhoff's Projective Contraction Theorem, Section 2.6.4 (page 260), there exists an integer $K_{\epsilon/3} > 0$ such that $k \geq K_{\epsilon/3}$ implies

$$d_H(A^k X_0, v_p) < \frac{\epsilon}{3}. \tag{2.243}$$

The Hilbert Projective Metric is invariant under positive scaling, so (2.243) implies

$$d_H\left(\frac{t^k}{k!} A^k X_0, v_p\right) < \frac{\epsilon}{3} \tag{2.244}$$

if $t > 0$ and $k \geq K_{\epsilon/3}$. The inequality (2.244) implies, with the help of Lemma 2.7.1.1 (page 273), that for all integers $M \geq K_{\epsilon/3}$ and for all $t > 0$ that

$$d_H\left(\sum_{k=K_{\epsilon/3}}^{M} \frac{t^k}{k!} A^k X_0, v_p\right) < \frac{\epsilon}{3}.$$

From Corollary 2.5.0.4 (page 248) we know that $d_E(f, g) \leq d_H(f, g)$ if $f, g \in \triangle^{n-1}$,

and so for all $M \geq K_{\epsilon/3}$ and for all $t > 0$ we must have

$$d_E\left(\frac{\sum_{k=K_{\epsilon/3}}^{M} \frac{t^k}{k!}\ A^k X_0}{\left\|\sum_{k=K_{\epsilon/3}}^{M} \frac{t^k}{k!}\ A^k X_0\right\|_1}, v_p\right) < \frac{\epsilon}{3}. \tag{2.245}$$

A is primitive and X_0 is non-zero non-negative so that $A^k X_0 \neq 0$ for all integers $k > 0$. $e^{tA} X_0$ is convergent so $\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!}\ A^k X_0$ is convergent. Then, since $y \to y/\left\|y\right\|_1$ is continuous at all non-zero y, there exists an integer $M_{t,\epsilon/3} > 0$, dependent on $t > 0$ such that

$$d_E\left(\frac{\sum_{k=K_{\epsilon/3}}^{M_{t,\epsilon/3}} \frac{t^k}{k!}\ A^k X_0}{\left\|\sum_{k=K_{\epsilon/3}}^{M_{t,\epsilon/3}} \frac{t^k}{k!}\ A^k X_0\right\|_1}, \frac{\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!}\ A^k X_0}{\left\|\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!}\ A^k X_0\right\|_1}\right) < \frac{\epsilon}{3}. \tag{2.246}$$

Combining (2.245) and (2.246) yields for $t > 0$:

$$d_E\left(v_p, \frac{\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right) \leq 2\epsilon/3. \tag{2.247}$$

By the previous lemma, Lemma 2.7.2.1 (page 277), there exists a $t_{\epsilon/3} > 0$ such that $t > t_{\epsilon/3}$ implies

$$\left\| \frac{\sum_{k=0}^{(K_{\epsilon/3})-1} \frac{t^k}{k!}\ A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!}\ A^k X_0\right\|_1} \right\|_E < \frac{\epsilon}{3}. \tag{2.248}$$

The following inequality will be useful:

$$\begin{aligned} d_E(X+Y, Z) &= \left\|X + Y - Z\right\|_E \\ &\leq \left\|X\right\|_E + \left\|Y - Z\right\|_E \\ &= d_E(X, 0) + d_E(Y, Z). \end{aligned} \tag{2.249}$$

Finally, putting together the above arguments, in particular the inequalities (2.247),

(2.248), and (2.249), we have for all $t > t_{\epsilon/3}$:

$$\begin{aligned}
d_E\left(v_p, \frac{X(t)}{\|X(t)\|_1}\right) &= d_E\left(v_p, \frac{\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right) \\
&= d_E\left(v_p, \frac{\sum_{k=0}^{(K_{\epsilon/3})-1} \frac{t^k}{k!} A^k X_0 + \sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right) \\
&= d_E\left(v_p, \frac{\sum_{k=0}^{(K_{\epsilon/3})-1} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1} + \frac{\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right) \\
&\leq \underbrace{d_E\left(0, \frac{\sum_{k=0}^{(K_{\epsilon/3})-1} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right)}_{\leq\ \epsilon/3} + \underbrace{d_E\left(v_p, \frac{\sum_{k=K_{\epsilon/3}}^{\infty} \frac{t^k}{k!} A^k X_0}{\left\|\sum_{k=0}^{\infty} \frac{t^k}{k!} A^k X_0\right\|_1}\right)}_{\leq\ 2\epsilon/3} \\
&\leq \epsilon.
\end{aligned}$$

□

2.8 Circumference $= 6r$ Theorem

I had thought that the following nice result about the radius of "circles" in a triangle to which the Hilbert Metric is applied, Theorem 2.8.0.2 (page 288), was original [37], but it was not. However, the result and (my) proof is short and I hope charming, and so I've included it.

Let $C = \mathbf{R}^{\mathbf{3}}_{\geq \mathbf{0}}$. We identify $(C \setminus \{0\}, \sim)$ with the 2 simplex [38]

$$\triangle^2 = \{(x, y, z) \in \mathbf{R}^{\mathbf{3}}_{\geq \mathbf{0}} : x, y, z \geq 0, x + y + z = 1\}$$

[37] I thought that my "discovery" of Theorem 2.8.0.2 (page 288) was original as many searches for Hilbert Circles turned up nothing related. By accident, recently, I came across an article by B.B. Phadke [80] (1975) proving Theorem 2.8.0.2. Phadke nowhere mentions that he is using the Hilbert Metric, which he ultimately is; rather he talks about "straight Desarguesian chord spaces". Hence my difficulty.

[38] See Definition 2.4.7.1 part 4 (page 197).

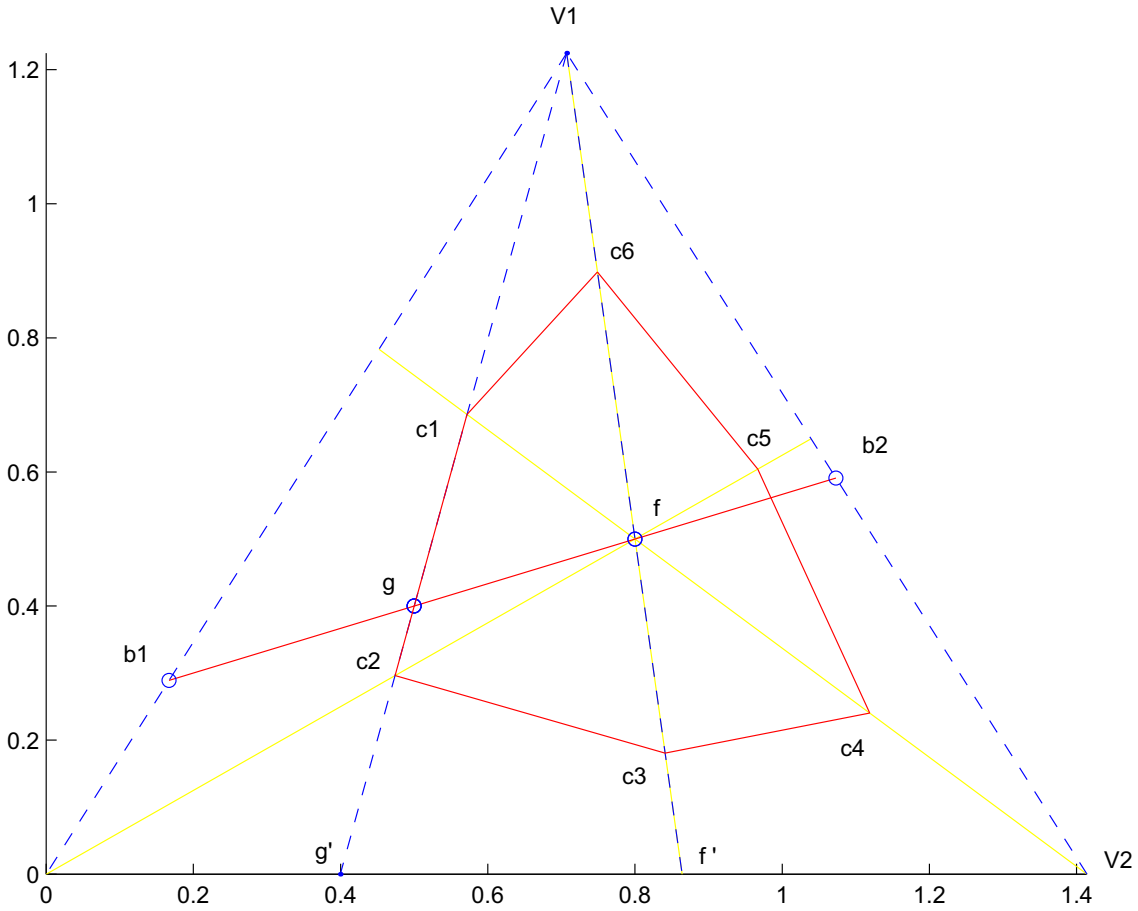


Figure 2.11: Circumference $= 6r$ Theorem (page 288). In this illustration, the "circle" of d_H radius r, denoted $S^1_H(f, r)$, appears to our euclidean eyes as the red hexagon inside the triangle, T, having vertices $0, V_1, V_2$. The triangle T is actually the 2 simplex $\triangle^2 \subset \mathbb{R}^3$, which we have embedded in $\mathbb{R}^2$ for illustrative purposes.

which we in turn identify with the triangle T, shown in Figure 2.11 (page 287), and which we've embedded in $\mathbf{R^2}$. Let $f \in T$ and let $S^1_H(f,r)$ be the circle of d_H radius r in T.

Birkhoff [12] defines the projective metric, $d_H(f,g)$, which he denotes by $\theta(f,g)$, via the cross ratio $R(f_2/f_1, g_2/g_1;\ 0,\infty)$. Explicitly

$$d_H(f,g) = \theta(f,g) = |\ln(R(f_2/f_1, g_2/g_1;\ 0,\infty))| = |\ln(f_2 g_1/f_1 g_2)|$$

In Theorem 2.8.0.2 (page 288) we apply the projective invariance of the cross ratio to Figure 2.11 (page 287) to show that 'circles' of d_H radius r in the 2 simplex always have a circumference of length $6r$ and that 'd_H circles' in $\triangle^2$ are hexagons when viewed as subsets of Euclidean space.

This result is not at all obvious from the more modern definition of d_H [39].

Theorem 2.8.0.2. CIRCUMFERENCE $= 6r$ THEOREM. *The circle $S^1_H(f,r)$ in $\triangle^2$ of d_H radius r and center f has circumference $6r$ with respect to d_H.*

Proof. See Figure 2.11. $S^1_H(f,r)$ is the red hexagon. Let $r = d_H(f,g) = |\ln(b1, b2; f, g)|$. In the cross ratio formulation of d_H we choose the origin to be the point of projection. However, the cross ratio is invariant with respect to changing the point of projection provided the 4 points are all on the same line and the point of projection is not on that line [69]. Let's calculate the cross ratio $(b1, b2 : f, g)$ using $V1$ as the projection point. This makes it easy to see that all the points on $\overline{C1C2}$ are distance r from f. Then notice that if we change our point of projection from $V1$ to $V2$ it is clear that all the points on $\overline{C2C3}$ are distance r from f. Let's once again use $V1$ as the projection point. It is immediate that the length of $\overline{C2C3} = r$. There are six such line segments. □

[39] See Section 1.9.3 (page 84) where for the definition of d_H via $\alpha f \leq g \leq \beta f$.

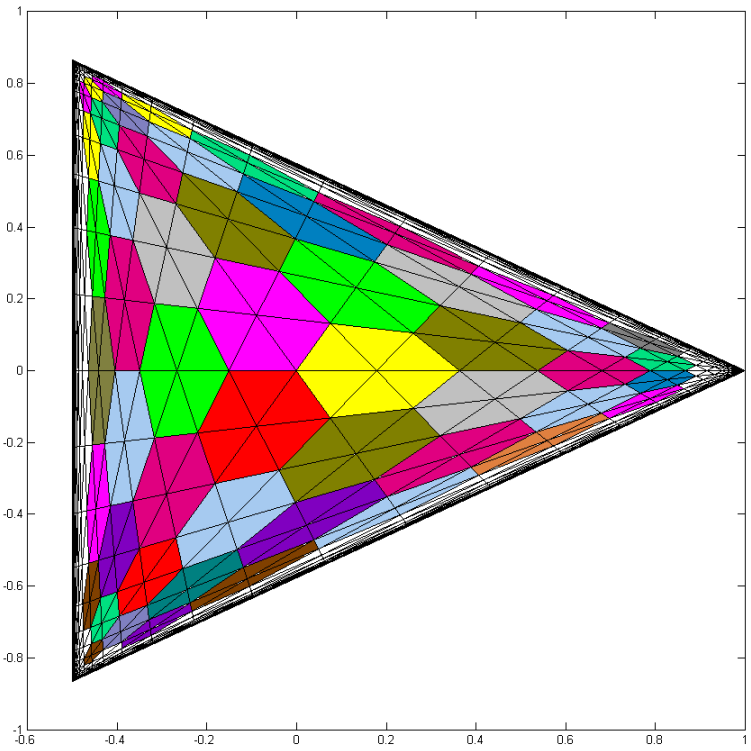


Figure 2.12: Tiling the two simplex with congruent (same sized) Hilbert circles; i.e. each hexagon is a circle of fixed radius r relative to d_H. See Figure 2.11 (page 287).

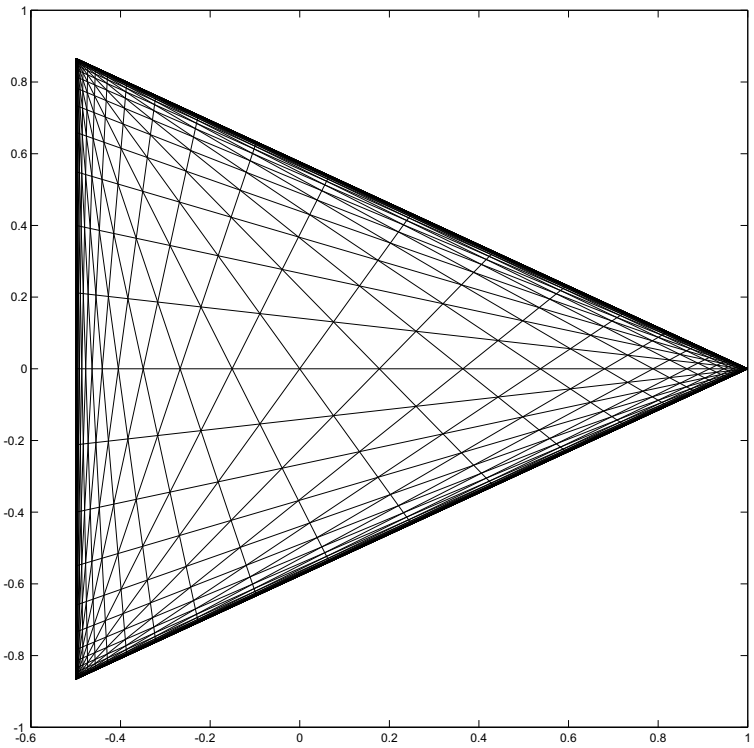


Figure 2.13: Tiling the two simplex with congruent (same sized) Equilateral Triangles relative to d_H. The lines terminating at each vertex are all parallel relative to d_H and uniformly spaced. If 6 of these triangles have a common vertex, the union of the six triangles will form a Hilbert circle. This was shown more explicitly in Figure 2.12 (page 289) and Figure 2.11 (page 287).

Part II

Mathematical Biology

Introduction to Part II

Part II of this dissertation is about a mathematical model whose goal is to estimate the probability that a viral or bacterial infection, or a cancer, will develop resistance to treatment.

The model was introduced in 2003 and 2004 by Yoh Iwasa [40], Franziska Michor [41], and Martin Nowak [42] in two papers, *Evolutionary Dynamics of Escape from Biomedical Intervention* [48] and *Evolutionary Dynamics of Invasion and Escape* [49].

In the model the development of resistance is divided into pre and post-treatment phases. The pre-treatment phase is modeled as a deterministic linear dynamical system. The post-treatment phase is modeled as a stochastic branching process. As the invaders (the viruses, bacteria, or cancer cells) replicate, they may produce offspring with genetic mutations. So, as the invaders colonize their host by replicating themselves, their population will come to consist of "wild" types and "mutant" types. The wild types have the same genotype as the original invaders. The various types may replicate at different rates – some being more and some being less "fit". It is assumed, for simplicity, that all viruses, bacteria, or cancer cells operate independently of each other.

Pre-treatment phase of the model. During the pre-treatment phase, if certain assumptions are made, the relative frequency distribution of the descendants' various genetic types approach an invariant distribution known as quasispecies equilib-

[40]Yoh Iwasa: Professor of Theoretical Biology and Director of the Mathematical Biology Lab, Department of Biology, Kyushu University. The model [48, 49] was developed while Yoh Iwasa was at the Institute for Advanced Study, Princeton. Yoh Iwasa over 200 publications in Mathematical Biology.

[41]Franziska Michor: Memorial Sloan-Kettering Cancer Center's computational biology lab (2007 – 2010). Dana-Farber Cancer Institute and Harvard School of Public Health (2010). PhD. Harvard (2005), Advisor Martin Nowak.

[42]Martin Nowak: Director of the Program for Evolutionary Dynamics at Harvard University since 2003. Studied under Peter K. Schuster (Chemistry) and Karl Sigmund (Mathematics) at the University of Vienna. In 1998 M. Nowak moved from Oxford (where he held a professorship) to Princeton to establish the program in Theoretical Biology at the Institute for Advanced Study.

rium [30, 29]. In the model it is assumed that treatment is not given until quasispecies equilibrium is reached. The quasispecies equilibrium (the invariant distribution) turns out to be an eigenvector and can be calculated using various techniques.

Post-treatment phase of the model.

During the post-treatment phase [43] the treatment (e.g. a drug) exerts a selective pressure on the descendants of the invaders. The treatment is assumed to be successful [44] against the wild type and against all the mutant types – except for one special type called an "escape mutant."

It is useful to enumerate the genetic types of the invader's descendants. The wild type is denoted 0, and the other mutant types are denoted $1, 2, 3, \ldots, m$, with the number m being reserved for the escape mutant.

Using techniques from multi-type Galton-Watson branching processes [6, 42], the probability is calculated that a single individual of type i, $i = 0, 1, 2, \ldots, m$ will escape extinction [45] in spite of the drug treatment.

Combining Pre and Post-treatment calculations. The dot product of the pre-treatment quasispecies distribution vector and the post-treatment escape probabilities vector yields a probability p,

$$p = \sum_{i=0}^{m} \underbrace{P(\text{ selected is of type } i)}_{\text{from quasispecies equilibrium}} \; P(\text{ escape } | \text{ being of type } i).$$

p is the probability that if one individual virus, bacteria, or cancer cell is selected randomly when treatment begins that the selected individual will create a line of descendants that escapes extinction.

[43] The post-treatment phases starts as soon as treatment is started.

[44] A successful treatment (against a particular type) is one which drives the reproductive ratio (of that particular type) to below 1; meaning each generation (of that particular type) is smaller than the next.

[45] An individual virus, bacteria, or cancer cell is said to escape extinction for as long as at least one of its descendants, in possibly mutated form, remains alive. An escape mutant would be said to be resistant to treatment.

If N is the total population size, including all the different types, and we invoke independence, we have that $N \cdot p$ is the mean number of mutants expected to avoid extinction (to escape).

It is assumed that escaping is rare and the model uses the Poisson Distribution with mean $N \cdot p$ to calculate the probability that there will be zero escapes.

Iterative Processes. Both the pre and the post treatment phases are modeled using iterative processes. This is not surprising as one can (mathematically) think of growth, reproduction, and evolution as being examples of iterative processes; a collection of rules and processes repeatedly applied.

For example, a cell splits, then the daughter cells split, and so on. The result is, at least initially, exponentially fast growth. It is worth observing that this exponentially fast growth is what allows multi-cellular organisms to contain trillions of cells. If growth were linearly fast, it would simply take too long to grow so many cells.

The dynamics of the pre-treatment phase can be understood using the projective machinery developed by Birkhoff [12, 13] to understand iterates of linear maps [46]. This machinery augments the approach taken in Iwasa, Michor, and Nowak in [48, 49] and provides a rigorous basis for the pre-treatment part of the model.

The mathematics of the post-treatment phase is also iterative. However, the machinery applied are multi-type generating function techniques; which involve some theory from several complex variables and ODE's in $\mathbb{C}^n$. We give a rigorous explanation of one way to calculate the vector of escape probabilities different from what was used by Iwasa, Michor, and Nowak in [48, 49].

These approaches are somewhat new in the study of mathematical biology for this particular direction in research. Moreover, bringing complex analysis into the treatment makes this study more interesting and promising.

[46] The projective machinery developed by Birkhoff to understand iterates of linear maps was the focus of Part I of this dissertation.

Part II.A.

Pre Treatment: Quasispecies Equilibrium

Introduction to Part II.A.

Charles Darwin was quite aware that altruism [47] posed a problem to his theory of evolution. From Darwin's *On the origin of species* [25, pp. 228 – 230] (1860):

> No doubt many instincts of very difficult explanation could be opposed to the theory of natural selection – cases, in which we cannot see how an instinct could have originated.... I will not enter here on these several cases, but will confine myself to one special difficulty, which at first appeared to me insuperable, and actually fatal to the whole theory [of natural selection]. I allude to neuters or sterile females in insect-communities; for these neuters often differ widely in instinct and in structure from both the males and fertile females, and yet, from being sterile, they cannot propagate their kind....
>
> This difficulty, though appearing insuperable, is lessened, or, as I believe, disappears, when it is remembered that selection may be applied to the family, as well as to the individual and may thus gain the desired end.

As Edward O. Wilson [48] put it in *Sociobiology*[90] (1980), "To save his own theory, Darwin introduced the idea of natural selection operating at the level of family [49] rather than of the single organism."

One model that represents a species (or more precisely, closely related DNA/RNA templates) as an evolving network (or family) is the quasispecies model. According

[47] Altruistic behavior, from the perspective of population biology, is behavior that fails to increase the fitness (the reproductive success rate) of the actor, but benefits the fitness of others. For brief and accessible overviews of altruism and evolution, see Wilson [91] (2005) or Foster, Wenseleers, and Ratnieks [36] (2006).

[48] Edward 0. Wilson (b. 1929) is a distinguished award winning biologist, author, and the Pellegrino University Research Professor, Emeritus, in Entomology for the Department of Organismic and Evolutionary Biology at Harvard University. His specialty is myrmecology, which is the study of ants.

[49] A "family", using modern vocabulary, is a collection of genetically similar, but not necessarily identical individuals. Alternatively, one might think of a family as a type of network, with the nodes representing individuals, or individual genotypes, and the genetic distance between the individuals, being the weights.

to Claus O. Wilke [89] (2005):

> Quasispecies theory has its origin in a seminal paper written by Eigen [50] in 1971 [27], in which he studied the error-prone self-replication of biological macromolecules, primarily with the goal of understanding the origin of life [51].

By the late 1970s Manfred Eigen, working with Peter Schuster [52], had coined the term quasispecies and had developed a mathematical model for this, based upon their knowledge of reaction rate equations in chemistry; Eigen and Schuster [30] (1979) [53]. According to Martin Nowak [54] and Robert May [75, Chapter 8] (2000):

> The word quasispecies is a mystery for most biologists, possibly because it was invented by chemists. In the 1970's, Manfred Eigen and Peter Schuster developed a chemical theory for the origin of life. They described how populations of RNA molecules could reproduce themselves. They noted that the spontaneous chemical replication of such comparatively simple molecules was much less accurate than the genetic replication of any organism alive.... Consequently a population of RNA molecules that was the result of such an inaccurate replication process would not be absolutely homogeneous, but a mixture of RNA molecules with different nucleotide sequences....

[50] Manfred Eigen (b. 1927) is a German biophysicist who won the 1967 Nobel Prize in Chemistry, "for studies of extremely fast chemical reactions, effected by disturbing the equilibrium by means of very short pulses of energy". He is also the former Director of the Max Planck Institute for Biophysical Chemistry.

[51] Life, meaning the complex DNA/RNA molecules which carry genetic hereditary information.

[52] Peter Schuster (b. 1941) Austrian theoretical chemist. Manfred Eigen's post doctoral student 1968 - 69. Head of the Institute of Theoretical Chemistry at the Universitt Wien, 1973-1992, 1996 - 2009; President of the Austrian Academy of Sciences, 2006 - 2009.

[53] [30], which is Eigen and Schuster's 1979 book, *The Hypercycle, a Principle of Natural Self-Organization*, is not widely available. However, Eigen, McCaskill, and Schuster's 1988 article *Molecular Quasi-Species* [29] is readily available, well written, and covers much of the same material as [30].

[54] Martin Nowak earned his PhD, in part under Peter Schuster.

> Chemists refer to an ensemble of equal molecules as a 'species'. For example the species of H_20 molecules. In contrast, a species of RNA molecules, derived from inaccurate reproduction, is not an ensemble of *identical* molecules. Hence the term 'quasispecies'....
>
> Eigen and Schuster go on to argue that the target of natural selection is not the fittest sequence, but the quasispecies.

We come to the specifics of the definition of the quasispecies. We quote directly from Eigen, McCaskill, and Schuster's 1988 article *Molecular Quasi-Species* [55] in the Journal of Physical Chemistry [29]. Keep in mind that these authors are chemists writing for chemists:

> The molecular quasi-species model describes the physicochemical[56] organization of monomers[57] into an ensemble of heteropolymers[58] with combinatorial complexity[59] by ongoing template polymerization[60].... The quasi-species itself represents the stationary distribution[61] of macromolecular sequences maintained by chemical reactions effecting error-prone replication.... It is obtained deterministically, by mass-action kinetics[62], as

[55]The originators of quasispecies theory wrote quasi-species, whereas in the more current literature, the hyphen is dropped. We will write quasispecies, without the hyphen, except when directly quoting a source which hyphenates quasispecies.

[56]Physicochemical, meaning a chemical model based upon the laws of physics.

[57]Monomer, meaning a "smaller" or "simpler" chemical unit or molecule, often having the possibility of being joined with other monomers to form polymers. In the current context, the monomers are DNA/RNA nucleotides.

[58]Heteropolymer, meaning a chemical made up different types of monomers. In the current context DNA/RNA.

[59]Combinatorial complexity, as there are 4 basic nucleotides being assembled into sequences of DNA (or RNA), and as they can be assembled in different orders, the number of possible sequences is 4^n with n being the length of the sequence in terms of nucleotides joined.

[60]Template polymerization, as in DNA/RNA template self-replication

[61]The authors show that under certain conditions that the relative frequencies of the templates becomes a stable (invariant) distribution. This invariant distribution is what they are calling the quasi-species. In some literature, this invariant distribution is called the quasispecies equilibrium.

[62]Mass action kinetics approximates the rate of a chemical reaction to be proportional to the quantity of the reacting substances. It is based upon collision theory. The constants of proportionality are called "chemical rate coefficients."

> the dominant eigenvalue[63] of a value matrix, W, which is derived directly from chemical rate coefficients, but it also exhibits stochastic features, being composed to a significant fraction of unique individual macromolecular sequences[64]. The quasi-species model demonstrates how macromolecular information originates through specific non-equilibrium autocatalytic reactions[65] and thus forms a bridge between reaction kinetics and molecular evolution.... Experimental data obtained from test-tube evolution of poly-nucleotides and from studies of natural virus populations support the quasi-species model.

Briefly, if $c(t) = (c_1(t), c_2(t), \ldots, c_m(t))$ is the vector of concentrations of the m possible templates, with t being time, then

$$\dot{c}(t) = cW, \tag{2.250}$$

where the off diagonal entries of W, $W_{ik} > 0$ are the mutation rates $k \to i$ and the diagonal entries, $W_{ii} > 0$ are the, "fitness factors of conventional population

[63]Actually the authors seem to have meant that the quasispecies is the eigenvector corresponding to the dominant eigenvalue. See p. 6885 of the same article [29], where the authors write:

> The stationary sequence distribution, thus determined by the dominant eigenvector l_0, is called the quasi-species. It consists of a master sequence I_0, which is the most frequent sequence and commonly has the maximum selective value, and a mutant distribution centered around the master. The value matrix W, according to ... has exclusively positive entries, and the Perron-Frobenius theorem [79] applies: the largest eigenvalue λ_0, is non-degenerate, and all components of the vector l_0 are positive. Hence, it fulfills all requirements to describe a mixture of chemical compounds – here the distribution of the master sequence and its mutants.

[64]For the case of DNA or RNA strands, there are 4^n possible sequences of length n, for each n. So the number of *possible* "macromolecular sequences" (meaning possible DNA or RNA sequences) will typically be vastly larger than the number of macromolecular sequences actually present. As a result, many of the possible sequences will not be present – and for some the sequences that are present, there might only be one or two copies. For the sequences, for which only a small number of copies are present, the laws of large numbers or averaging, will not apply, and so "stochastic features" may be important for those cases.

[65]Non-equilibrium autocatalytic reactions, in this context, means that as the self replicating DNA or RNA templates replicate, the relative concentrations of the templates change (i.e. non-equilibrium), and the system evolves.

dynamics"[29]. Finally, "The stationary sequence distribution... determined by the dominant eigenvector [of W], is called the quasi-species" [29].

In Chapter 3 we discuss quasispecies theory and develop Equation (2.250), which, perhaps due to an oversight, does not appear in Eigen, McCaskill, and Schuster's [29].

In Chapter 4 we discuss the pre-treatment phase of the model introduced in 2003 and 2004 by Iwasa, Michor, and Nowak in [48, 49] which is based directly on the quasispecies theory of Eigen and Schuster [30, 29]. We relate quasispecies theory to the projective geometry and machinery developed by Birkhoff in [12, 13] and which we discuss in Part I of this dissertation.

Chapter 3

Quasispecies equilibrium (Pre-Treatment)

Introduction to Chapter 3

In Chapter 3 we discuss quasispecies theory originated by Eigen and Schuster [27, 30, 29]. We then relate quasispecies theory to projective geometry and the machinery developed by Birkhoff in [12, 13] and discussed by us in Part I of this dissertation.

3.1 The Quasispecies Model

The quasispecies is a collection of closely related self-replicating templates (DNA/RNA modeled as binary strings) [1]. We can think of the templates as being emersed in a solution containing the necessary components for self replication. As the templates replicate their concentration levels change depending on replication rates and error or mutation rates [29].

[1] This model could be classified as a toy type model since the actual nucleotides: Guanine (G) which pairs with Cytosine (C); and Adenine (A) which pairs with Thymine (T) (or Uracil (U) in RNA), do not appear in this model. By pairing adjacent 0 and 1's in the strings, we would recover (at least from an information theory sense) some of the properties of the ATCG DNA model.

The parameters of the model [29] [2]:

$c_i(t)$ is the **concentration** (units/vol) of the type i template.

A_i is the **replication rate** constant for the i type.

D_i is the **degradation rate** constant for the i type.

Without mutation we have

$$\frac{d}{dt}\, c_i(t) = (A_i - D_i)c_i(t).$$

The units of A_i and D_i are $\frac{1}{\text{time}}$, but we could also think of

$$\text{the units of } A_i \text{ as } \frac{\frac{\text{new units}}{\text{units}}}{\text{time}} \text{ and the units of } D_i \text{ as } \frac{\frac{\text{degraded units}}{\text{units}}}{\text{time}}.$$

$Q_{i,j}$ is called the **quality factor** and it is the **probability** that a type i template will replicate a type j template [3]. We have $\sum_{j=1}^{n} Q_{i,j} = 1$, assuming there are n different template types. $Q_{i,j}$ is unitless.

If $i \neq j$ then $A_i Q_{i,j}$ is called the **mutation rate** from i to j. We introduce the matrix W with entries,

$$W_{i,j} = \begin{cases} A_i Q_{i,j} = W_{i,j}, & i \neq j \ ; \\ A_i Q_{i,i} - D_i, & i = j. \end{cases} \tag{3.1}$$

The $W_{i,i}$ would be called fitness factors in traditional population dynamics. We get

[2] [29] Manfred Eigen, John McCaskill, and Peter Schuster, *Molecular Quasi-Species*, Journal of Physical Chemistry **92** (1988), 6881–6891.

[3] In the notation used by Eigen, McCaskill, and Schuster $A_i Q_{j,i}$ would represent the mutation rate from i type to j type.

the rate equation

$$\begin{aligned}\frac{d}{dt}\, c_i(t) &= c_i(t)(A_i Q_{i,i} - D_i) + \sum_{j=1, j\neq i}^{n} c_j(t) A_j Q_{j,i} \\ &= c_i(t) W_{i,i} + \sum_{j=1, j\neq i}^{n} c_j(t) W_{j,i} \\ &= \sum_{j=1}^{n} c_j(t) W_{j,i}\ .\end{aligned}$$

In terms of the matrix W, and vectors $c = (c_1(t), c_2(t), \ldots, c_n(t))$ and $\dot{c}$, we have

$$\dot{c} = cW, \tag{3.2}$$

or more explicitly:

$$(\dot{c}_2(t), \dot{c}_1(t), \ldots, \dot{c}_n(t)) = (c_1(t), c_2(t), \ldots, c_n(t)) \begin{pmatrix} W_{1,1} & W_{1,2} & \ldots & W_{1,n} \\ W_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ W_{n,1} & \ldots & \ldots & W_{n,n} \end{pmatrix}. \tag{3.3}$$

In the model we are more concerned with the relative concentrations, and we will denote the relative concentration of the i type template by $x_i(t)$. So

$$x_i(t) = \frac{c_i(t)}{\sum_{j=1}^{n} c_j(t)}.$$

The solution $c(t)$ to the rate equation $\dot{c} = cW,\ c(0) = c_0$ is just $c_0 e^{tW}$, which is a trajectory in $\mathbb{R}^n$. Physically, of course, $c(t)$ must be in $\mathbb{R}^n_{\geq 0}$, and mathematically this will be the case since the off diagonal entries of W are non-negative.

It is convenient to assume that $W \geq 0$ as a negative diagonal entry indicates a non-viable type. Since the off-diagonal entries are by construction non-negative, the

assumption that $W \geq 0$ is the assumption that for each i, that $A_i Q_{i,i} - D_i \geq 0$.

3.2 Projective Systems of Differential Equations

We introduce some of the ideas from Projective Geometry (from Part I of this dissertation) to the quasispecies model. The relationship between Birkhoff's work [12] and the quasispecies model does not seem to be widely recognized.

At the end of this section a concrete example is discussed and illustrated in Figures 3.1 through 3.7.

3.2.1 Relative Concentration as a Projection

Let $c(x,t) =$ the flow $c(x,t) = xe^{tW}$ determined by the linear ODE (3.2). The relative concentration vector $x(t) = \frac{c(t)}{\|c(t)\|_1}$ is just the projection of $c(t)$ onto the n-1 simplex

$$\triangle_{n-1} = \{x \in \mathbb{R}^n_{\geq 0} : \|x\|_1 = 1\}.$$

In terms of flows, $c(x,t) \rightarrow c(x,t)/\|c(x,t)\|_1$, this projection induces a naturally defined vector field on $\triangle_{n-1}$:

Let $c_0 \in \mathbb{R}^n_{\geq 0}$. Since $c(c_0,t) = c_0 e^{tW}$ we have $c(\lambda c_0, t) = \lambda c_0 e^{tW}$ so if we represent the projective equivalence class of $v \in \mathbb{R}^n_{\geq 0}$ by $[v]$ then $c([c_0],t) = [c_0 e^{tW}] = [c(c_0,t)]$. In other words, the flow $c(c_0,t)$ induced by the differential equation $\dot{c}(t) = c(t)W$ maps projective equivalence classes to projective equivalence classes, and so $c(*,t)$ is a flow on projective space of $\mathbb{R}^n_{\geq 0}$. Identifying the projective space of $\mathbb{R}^n_{\geq 0}$ with $\triangle_{n-1}$ we can ask what does the induced vector field on $\triangle_{n-1}$ look like? We differentiate w.r.t. to t and suppress the initial condition c_0, now assumed to be in $\triangle_{n-1}$, unless

needed:

$$\begin{aligned}
\frac{d}{dt}x(t) &= \frac{d}{dt}\ \frac{c(t)}{\|c(t)\|_1} \\
&= \frac{d}{dt}\ \frac{c(t)}{\sum_{i=1}^{n} c_i(t)}v \\
&= \frac{\dot{c}(t)\sum_{i=1}^{n} c_i(t) - c(t)\sum_{i=1}^{n}\dot{c}_i(t)}{\left(\sum_{i=1}^{n} c_i(t)\right)^2} \\
&= \frac{c(t)W\sum_{i=1}^{n} c_i(t) - c(t)\sum_{i=1}^{n}\left(c(t)W\right)_i}{\left(\sum_{i=1}^{n} c_i(t)\right)\left(\sum_{i=1}^{n} c_i(t)\right)} \\
&= x(t)W - x(t)\sum_{i=1}^{n}\left(x(t)W\right)_i\ . \qquad (3.4)
\end{aligned}$$

It is worth noting that all the c, c_i terms canceled, leaving only x, x_i terms, as should happen if the evolution of the relative concentrations are only dependent on the relative concentrations. This is what we should expect as c is invariant w.r.t. projection.

Quasispecies Equilibrium

From (3.4) The vector field vanishes when

$$0 = x(t)W - x(t)\sum_{i=1}^{n}\left(x(t)W\right)_i\ .$$

If a vector field vanishes at a point x_*, then the the constant trajectory $x(t) = x_*$ is the solution when x_* is the initial condition. Since x_* is a probability distribution and is constant with respect to time, it is an example of an invariant distribution. In the context of quasispecies theory, x_* is called the quasispecies equilibrium.

The equations

$$0 = x_* W - x_* \sum_{i=1}^{n} (x_* W)_i$$
$$x_* W = \underbrace{\left(\sum_{i=1}^{n} (x_* W)_i \right)}_{\text{scalar}} x_*$$

imply that the (quasispecies equilibrium) solution x_* is an eigenvector of W.

Mean Excess Production

From (3.4), if

$$\dot{x}(t) = x(t)W - x(t) \sum_{i=1}^{n} (x(t)W)_i$$

then

$$\dot{x}_i(t) = (x(t)W)_i - x_i(t) \sum_{j=1}^{n} (x(t)W)_j \,.$$

Rearranging terms yields:

$$\begin{aligned} \dot{x}_i &= \sum_{j=1}^{n} x_j W_{ji} - x_i \sum_{k=1}^{n} \sum_{j=1}^{n} x_j W_{jk} \\ &= \left(\sum_{j=1, j\neq i}^{n} x_j W_{ji} \right) + (x_i W_{ii}) - x_i \sum_{j=1}^{n} x_j \sum_{k=1}^{n} W_{jk}. \end{aligned} \tag{3.5}$$

We focus on $x_i \sum_{j=1}^{n} x_j \sum_{k=1}^{n} W_{jk}$; make the substitutions:

$$W_{i,j} = \begin{cases} A_i Q_{i,j}, & i \neq j \ ; \\ A_i Q_{i,i} - D_i, & i = j, \end{cases}$$

and get:

$$
\begin{aligned}
x_i \sum_{j=1}^{n} x_j \sum_{k=1}^{n} W_{jk} &= x_i \sum_{j=1}^{n} x_j \left(\sum_{k=1}^{n} W_{jk} \right) \\
&= x_i \sum_{j=1}^{n} x_j \left(\left(\sum_{k=1,k\neq j}^{n} A_j Q_{jk} \right) + (A_j Q_{jj} - D_j) \right) \\
&= x_i \sum_{j=1}^{n} x_j \left(\left(\sum_{k=1}^{n} A_j Q_{jk} \right) + (-D_j) \right) \\
&= x_i \sum_{j=1}^{n} x_j \left(\left(A_j \underbrace{\sum_{k=1}^{n} Q_{jk}}_{=1} \right) - D_j \right) \\
&= x_i \underbrace{\sum_{j=1}^{n} x_j \left(A_j - D_j \right)}_{=E(t)} \\
&= x_i E(t).
\end{aligned}
$$

We plug this back into the differential equation (3.5) for $\dot{x}_i$ and get:

$$
\begin{aligned}
\dot{x}_i &= \left(\sum_{j=1,j\neq i}^{n} x_j W_{ji} \right) + (x_i W_{ii}) - x_i \sum_{j=1}^{n} x_j \sum_{k=1}^{n} W_{jk} \\
&= \left(\sum_{j=1,j\neq i}^{n} x_j W_{ji} \right) + (x_i W_{ii}) - x_i E(t) \\
&= x_i \left(W_{ii} - E(t) \right) + \sum_{j=1,j\neq i}^{n} x_j W_{ji}.
\end{aligned}
$$

$E(t)$ is the mean excess production [4] (divided by the total concentration) [29, 30] [5]. Neglecting the mutation terms, $\sum_{j=1,j\neq i}^{n} x_j W_{ji}$, or better, assuming that they are very small, we see that the relative concentration x_i increases if its fitness W_{ii} is greater than the average excess production $E(t)$, and decreases otherwise. This implies that

[4] $c_j(t)\,(A_j - D_j)$ is the excess production at time t of the j type.

[5] Eigen and Schuster derived and named $E(t)$ in their work, e.g. [29, 30]. They do not bother with the projective geometry aspects.

(at low enough mutation rates) the distribution of the x_i will become more and more weighted with the fittest replicators, which in turn will increase $E(t)$ which will in turn filter for ever higher replication rates. Without mutation, the distribution would eventually only consist of templates with the highest fitness.

Example. Recall, from (3.1), that

$$W_{i,j} = \begin{cases} A_i Q_{i,j} = W_{i,j}, & i \neq j \ ; \\ A_i Q_{i,i} - D_i, & i = j. \end{cases}$$

Letting

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.925 & 0.075 \\ 0.1 & 0.9 \end{pmatrix}, \quad \text{and} \quad D = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.7 \end{pmatrix}$$

we get $W = AQ - D$ with

$$W = \begin{pmatrix} 3 & .3 \\ .3 & 2 \end{pmatrix}.$$

Consider the dynamical system $\dot{c} = cW$. This example is discussed and illustrated in Figures 3.1 through 3.7. Note that the positive eigenvector of W with L^1 norm 1 is (0.7830951895, 0.2169048105).

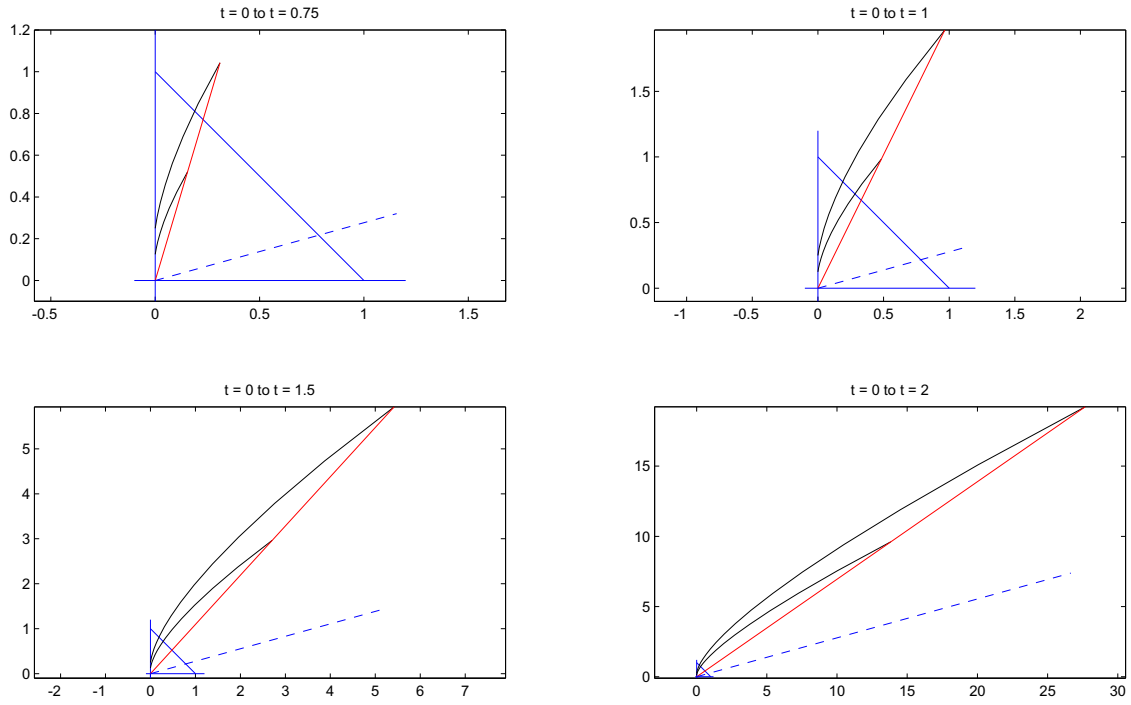


Figure 3.1: The differential equation $\dot{c}(t) = c(t)W$ has solution $c(0)e^{tW}$. These four graphs show the evolution of two solutions, one with initial condition $c(0) = (0, 0.125)$ and the other with initial condition $c(0) = (0, 0.25)$. Both solutions are plotted in all four graphs and are shown in black. Note, if we fix $c_0 \in \mathbb{R}^2_{\geq 0}$, then for each $t > 0$ $\{c(\lambda c_0, t) : \lambda > 0\}$ is a single projective point, this is indicated in the graphs by a straight red line. The dashed blue line shows the unique positive eigen direction of W. In these four graphs we plot the solutions from $t = 0$ to $t = 0.75, 1.00, 1.50$, and 2.00.

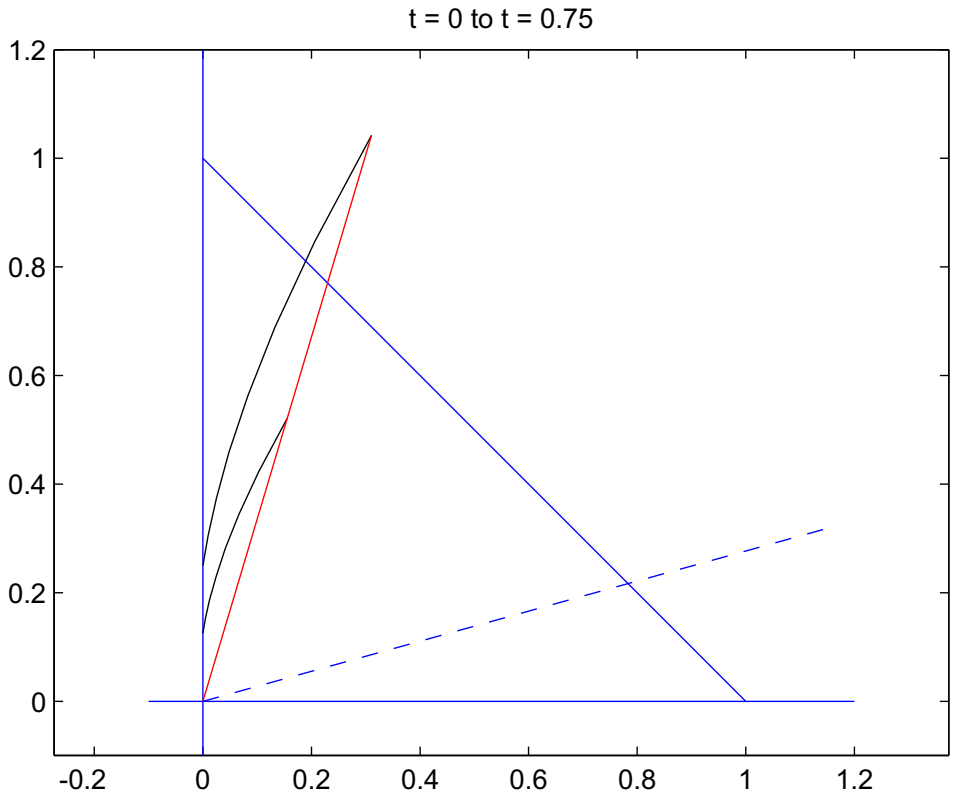


Figure 3.2: The differential equation $\dot{c}(t) = c(t)W$. This graph is an enlargement of the (1,1) subplot of Figure 3.1. In this graph we plot the two solutions (black) from $t = 0$ to $t = 0.75$.

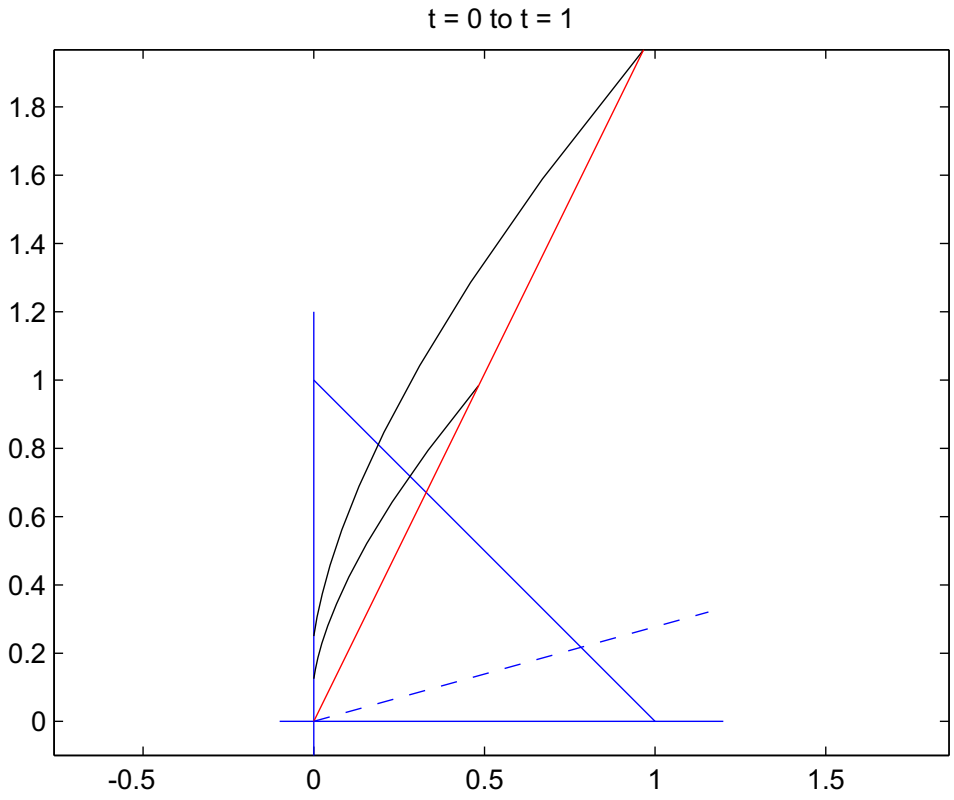


Figure 3.3: The differential equation $\dot{c}(t) = c(t)W$. This graph is an enlargement of the (1,2) subplot of Figure 3.1. In this graph we plot the two solutions (black) from $t = 0$ to $t = 1.00$.

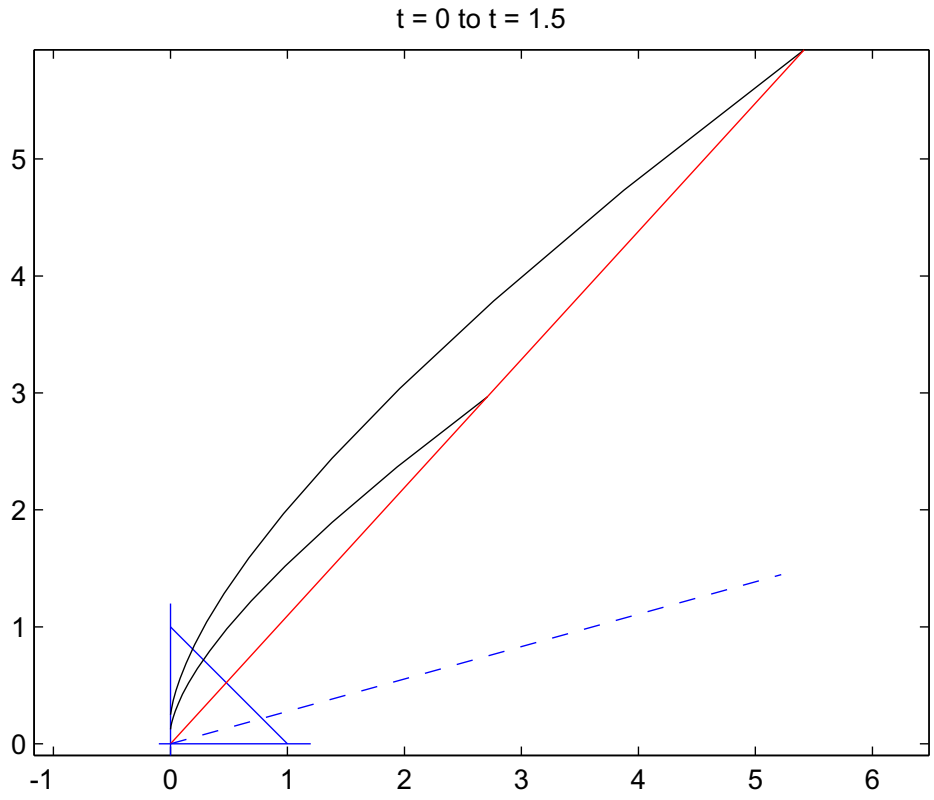


Figure 3.4: The differential equation $\dot{c}(t) = c(t)W$. This graph is an enlargement of the (2,1) subplot of Figure 3.1. In this graph we plot the two solutions (black) from $t = 0$ to $t = 1.50$.

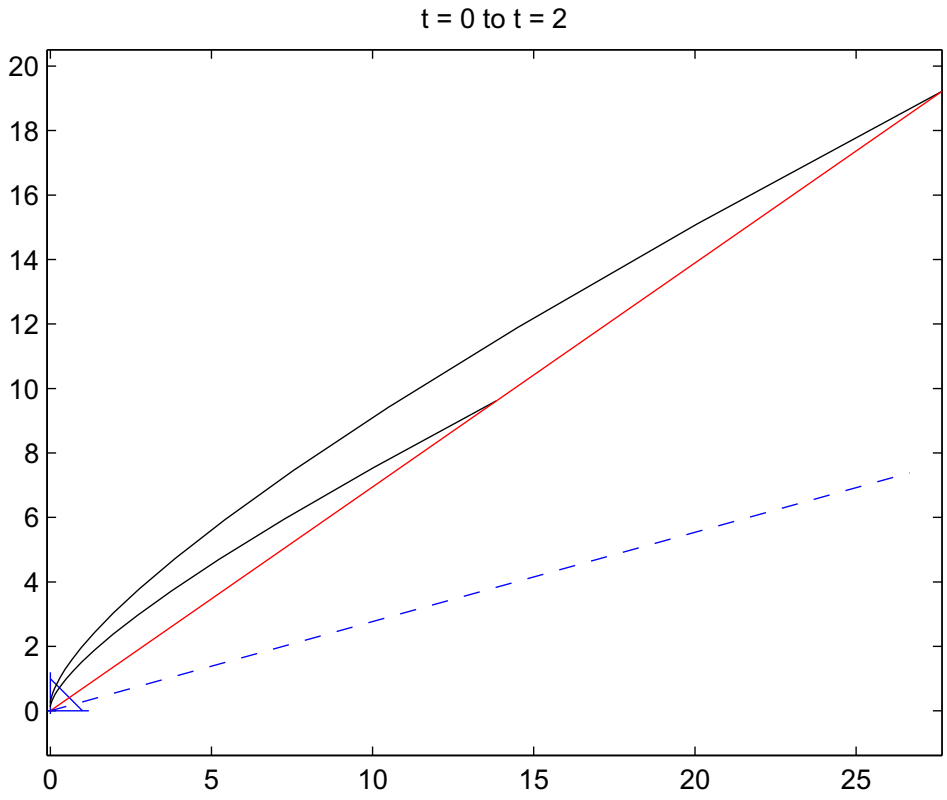
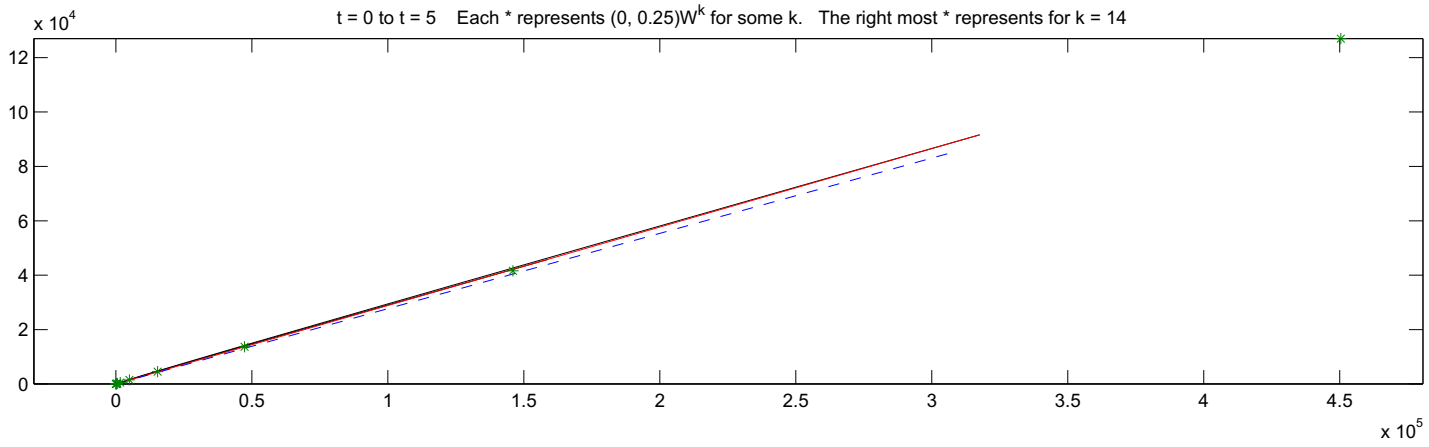


Figure 3.5: The differential equation $\dot{c}(t) = c(t)W$. This graph is an enlargement of the (2,1) subplot of Figure 3.1. In this graph we plot the two solutions (black) from $t = 0$ to $t = 2.00$.



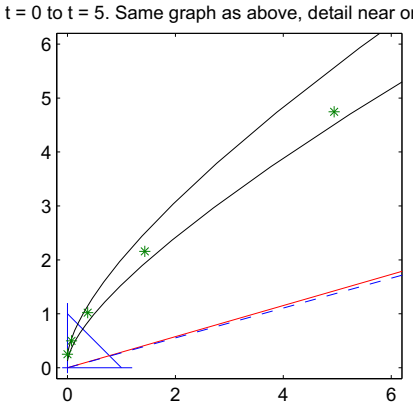


Figure 3.6: The above illustrations are a continuation of the example shown in Figure 3.1. In the top graph we indicate the long term behavior of $c(t)$ by plotting $c((0, 0.25), t)$ and $c((0, 0.125), t)$ as t goes from 0 to 5. On the same graph we also plot, using the symbol $*$, the discrete system, $(0, 0.25)W^k$ for $k = 0, 1, \ldots, 14$. Let v_p be the unique positive eigenvector of W satisfying $\|v_p\| = 1$. The graph shows clearly that as $t \to \infty$ that $\frac{c(t)}{\|c(t)\|_1} \to v_p$ (the projection of the red line approaches the projection of the dashed blue line) and as $k \to \infty$ that $\frac{(0,0.25)W^k}{\left\|(0,0.25)W^k\right\|_1} \to v_p$ (the projection of the $*$'s approach the projection of the dashed blue line). In this example $v_p = (0.7831, 0.2169)$. The one simplex $\triangle_1$ is indicated by the solid blue line segment connecting $(1, 0)$ to $(0, 1)$. The bottom graph is a detail from the top graph near the origin.

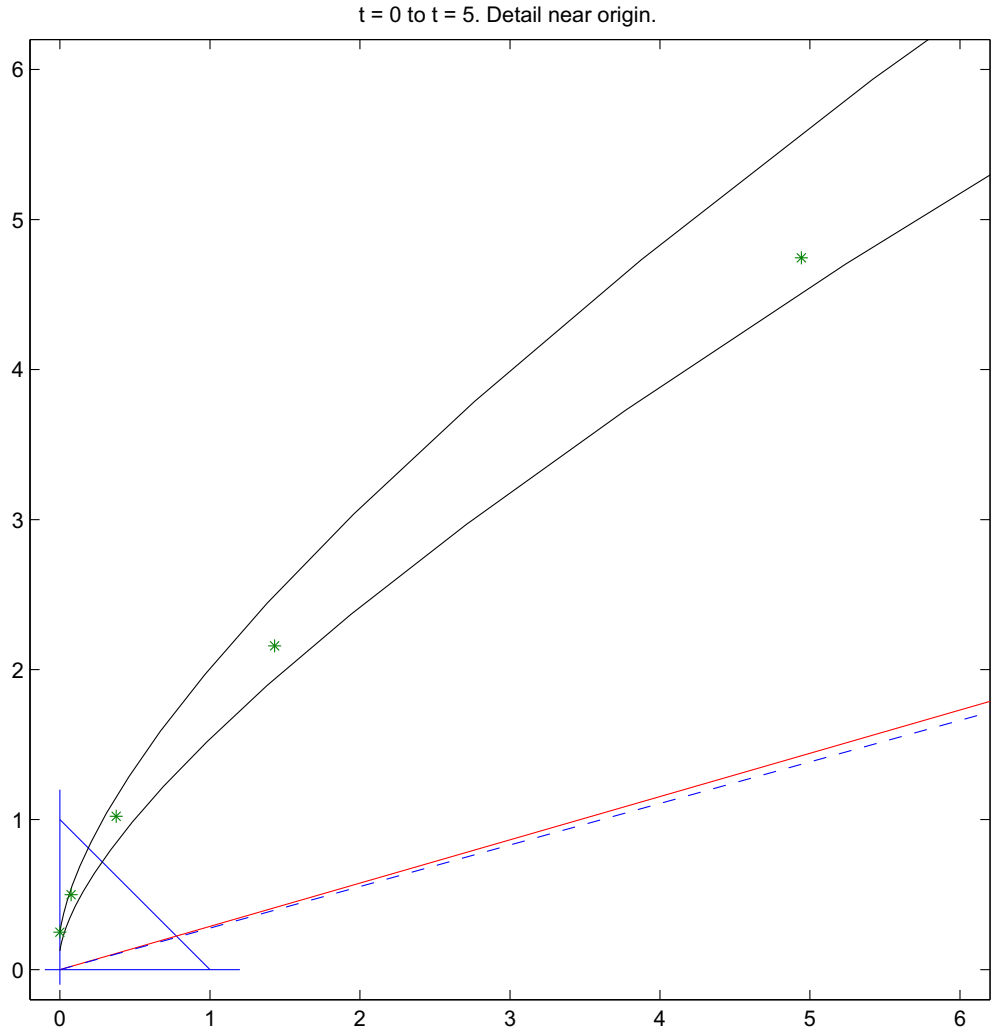


Figure 3.7: The differential equation $\dot{c}(t) = c(t)W$. The above graph is a continuation of the example shown in Figure 3.1 and is an enlargement of the (2,1) subplot of Figure 3.6. In the above graph we plot the two solutions over the interval $t = 0$ to 5 and then magnify the region close to the origin.

Chapter 4

Evolutionary Dynamics of Invasion and Escape

4.1 Introduction

Evolutionary Dynamics of Invasion and Escape [48, 49] (2003, 2004) were written by Yoh Iwasa, Franziska Michor, and Martin Nowak. Their papers present a beautifully creative model for invasion (infection) and escape (the development of resistance to medical treatment by drugs).

Their model ingeniously combines the quasispecies theory of Eigen and Schuster (for the pre-treatment phase) with a multi-type Galton-Watson process (for the post treatment phase).

The mathematics presented in Iwasa, Michor, and Nowak [48, 49] are rich, but terse, and we have found it difficult to give rigorous proofs of some details. Moreover their numerical calculations, e.g. those found in [49, p. 209, Table 1] and [48, p. 2576, Figure 3], are difficult to replicate because it seems that the mutation rate or other parameters were left out.

In 2006, in [10], Beerenwinkel, Eriksson, and Sturmfels [1] reworked some of the

[1]Bernd Sturmfels is a Professor of Mathematics, Statistics and Computer Science at University of

results in Iwasa, Michor, and Nowak's papers [48, 49] in a very mathematically sophisticated manner.

We note that, from a mathematical perspective, the quasispecies equilibrium is an invariant distribution, and can be understood in terms of the machinery developed in Part I of this dissertation:

Birkhoff's Projective Contraction Theorem [12] and its extensions imply directly that the iterates of a primitive matrix, say P, applied to a non-negative vector $X = X^1$ will converge to a unique positive eigen-direction.

Suppose our dynamical system for the pre-treatment phase is given in terms of a primitive matrix P acting on a population vector X:

$$X^{(n)} = XP^{n-1}$$

rather than as a system of differential equations. In Section 2.6.8 (page 270) in a more general setting than needed here, we prove that the central projections of $X^{(n)},\ n = 1, 2, \ldots$, onto the hyperplane of vectors whose coordinates sum to 1, will converges to the unique positive eigenvector of P whose coordinates sum to 1; i.e. to the quasispecies equilibrium.

On the other hand, suppose the quasispecies is developed in terms of a system of differential equations,

$$\dot{c} = cW \tag{4.1}$$

with W being a primitive matrix, as is implied in the Iwasa, Michor, and Nowak model [49] and in the Eigen and Schuster model [29]. Then the iterative machinery of Birkhoff still applies. Using the Hilbert Projective Metric and the machinery developed by Birkhoff in [12, 13], we prove in Part I, in Theorem 2.7.3.1 (page 283),

California, Berkeley. Niko Beerenwinkel was a post doctoral student at Berkeley and Harvard, and is currently with the Computational Biology group at ETH Zurich. Nicholas Eriksson's mathematics PhD adviser was Stumfels.

that: all positive trajectories in the system $\dot{c} = cW$ will, as time $\to \infty$, approach the same eigen-direction, the unique positive eigen-direction of W, which when centrally projected, yields the quasispecies equilibrium distribution.

Essentially, calculating the quasispecies equilibrium is calculating an eigenvector, and doing so numerically is its own subject – and somewhat outside of the scope of this dissertation.

In the rest of Chapter 4 we discuss some of the mathematics from the pretreatment phase of Iwasa, Michor, and Nowak's model.

4.2 Pretreatment Distribution of templates.

See Figure 4.1 (page 319). In [49], Iwasa, Nowak, and Michor calculate the quasispecies distribution relative to the wild type, rather than to total concentration. So $x(t) = \frac{c(t)}{c_0(t)}$ and the authors project the the dynamical system $\dot{c} = cW$, see (3.3) (page 303), into the hyperplane $c_0 = 1$ instead of into $\|c\|_1 = 1$, Note, c_0 refers to the first (the 0^{th}) coordinate which is the wild type's coordinate.

$$\begin{aligned}\frac{d}{dt}\left(\frac{c(t)}{c_0(t)}\right) &= \frac{\dot{c}c_0}{c_0c_0} - \frac{\dot{c_0}c}{c_0c_0} \\ &= xW - x_0W_{00}x\end{aligned}$$

but then

$$\begin{aligned}\dot{x}_i &= \sum x_jW_{ji} - x_0W_{00}x_i \\ &= x_iW_{ii} + \left(\sum_{j\neq i} x_jW_{ji}\right) - x_0W_{00}x_i \\ &= x_i\left(W_{ii} - x_0W_{00}\right) + \left(\sum_{j=1,j\neq i} x_jW_{ji}\right) + x_0W_{0i}.\end{aligned}$$

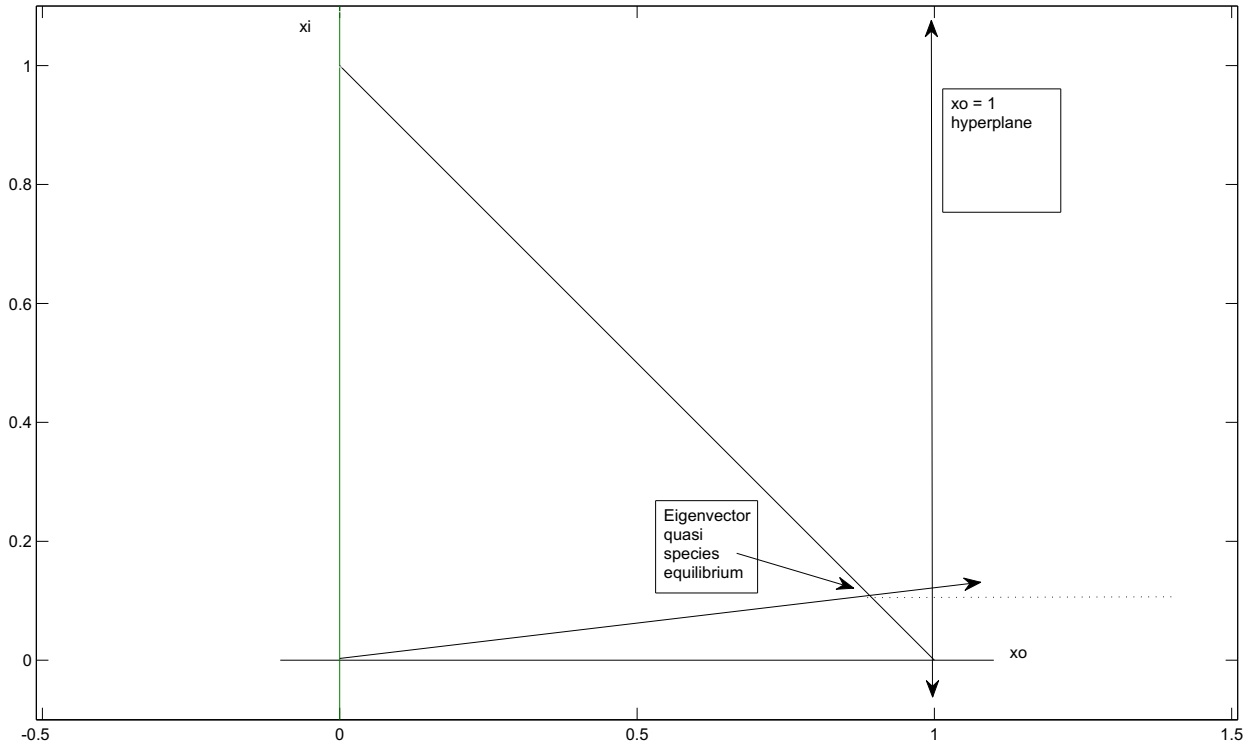


Figure 4.1: Regarding the differential equation $\dot{c}(t) = c(t)W$ with $W > 0$. The above illustration shows the projection of the positive eigenvector of W into the n-1 simplex $\triangle_{n-1}$ and into the hyperplane $x_0 = 1$. If the mutation rate is very low, then the positive eigenvector will be very close to the x_0 direction. Note: if there is no mutation, then the population will (in terms of relative frequencies) become exclusively 0 type. For small mutation rates one can approximate the quasispecies equilibrium by its projection in the hyperplane $x_0 = 1$.

The authors approximate the quasispecies distribution by the eigenvector of W having its x_0 coordinate equal to 1. They do not justify this and the actual quasi-species distribution would have $\| \ \|_1 = 1$.

The authors state that in their model $W_{00} = 1$. Another way to achieve the effect of having $W_{00} = 1$ would be to scale the matrix W to $\frac{W}{W_{00}}$, when $W_{00} \neq 1$. The resulting matrix $\frac{W}{W_{00}}$ has the same eigenvectors as W and its 00 entry is 1. Alternatively, one could simply divide all the differential equations by W_{00}. We are setting them equal to zero; after all, this division would not effect the value of the eigenvector either.

The projective version of that system (equivalently the effect of dividing by W_{00}) would be (setting $x_0 = 1$):

$$\dot{x}_i = x_i \left(\frac{W_{ii}}{W_{00}} - 1 \right) + \left(\sum_{j=1, j \neq i} x_j \frac{W_{ji}}{W_{00}} \right) + \frac{W_{0i}}{W_{00}}.$$

However, the authors simply set $w_0 = W_{00} = 1$ and $w_i = W_{ii} < w_0$. They also treat W_{ji} as $Q_{ij} = u_{ij}$; i.e. setting $A_i = 1$. The authors give the differential equation

$$\dot{x}_i = -x_i \left(1 - w_i\right) + \left(\sum_{j=1, j \neq i}^{n} x_j u_{ji} \right) + x_0 u_{0i} \quad \text{and} \quad x_0 \approx 1. \tag{4.2}$$

However, they do not explain these derivations. They may simply have thought along the following lines: w_i is the proportion of the i type to survive and thus be able to replicate. So $1 - w_i$ is the proportion which die. So $\dot{x}_i$ is proportionally decreased by $1 - w_i$, hence the $-x_i(1 - w_i)$ term. The terms $\left(\sum_{j=1, j \neq i}^{n} x_j u_{ji} \right) + x_0 u_{0i}$ represents the influx from mutation.

4.3 Approximation of Quasispecies Equilibrium.

Next Iwasa, Michor, and Nowak [48, 49] sketch a derivation of a formula to approximate the projection of the quasispecies equilibrium (eigenvector) in the hyperplane $x_0 = 1$. See Figure 4.1. I.e. they develop a formula to approximately solve Equation (4.2):

$$0 = -x_i\,(1 - w_i) + \left(\sum_{j=1, j\neq i}^{n} x_j u_{ji}\right) + x_0 u_{0i} \quad \text{and} \quad x_0 \approx 1.$$

Iwasa, Michor, and Nowak do this using approximation techniques based on "conventional Rayleigh-Schrödinger perturbation theory" found in [29, Section 4].

Eigen and Schuster [2] used Rayleigh-Schrödinger perturbation theory to approximate the quasispecies equilibrium eigenvector. They used perturbation theory because they were thinking in terms of large matrices. If there are n loci where mutations can occur it means that there are 2^n mutant types [3] that W will be a $2^n \times 2^n$ matrix. When n is large, 2^n is huge, and so perturbation theory is used.

In the examples we will calculate $n \leq 7$ and $2^7 = 128$ does not seem too large. In the examples discussed in Iwasa, Michor, and Nowak [48, 49] $n \leq 5$.

We will not dwell further on perturbation theory or the techniques used by Iwasa, Michor, and Nowak to solve (4.2). However, Rayleigh-Schrödinger perturbation theory, and more generally, techniques to find the dominant (positive) eigenvector of huge positive matrices, would be an interesting topic for further study.

For material on perturbation theory, the interested reader is directed to [52, 21, 85] and especially Eigen, McCaskill, Schuster's 1988 paper [29, Section 4]. For a mathematically rigorous treatment of some of the methods employed by Iwasa, Michor, and Nowak in [48, 49], the interested reader is directed to Beerenwinkel, Eriksson and Sturmfels' 2006 paper [10].

[2] See Eigen, McCaskill, and Schuster [29, p. 6885, Section 4]

[3] See Section 5.3.3 (page 359) for a discussion of the binary aspects of the model. Basically, if there are n possible mutations, then the various types are representable as strings of length n of 0's and 1's, a 0 indicating no mutation and a 1 indicating a mutation.

4.3.1 Calculation of quasispecies equilibrium

As explained in the introduction to this chapter, particularly in the discussion surrounding (4.1) about $\dot{c} = cW$, one can find the quasispecies equilibrium vector by simply finding the unique positive eigen-direction of W, see [4].

We write $\dot{c} = cW$:

$$(\dot{c}_2(t), \dot{c}_1(t), \ldots, \dot{c}_n(t)) = (c_1(t), c_2(t), \ldots, c_n(t)) \begin{pmatrix} W_{1,1} & W_{1,2} & \ldots & W_{1,n} \\ W_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ W_{n,1} & \ldots & \ldots & W_{n,n} \end{pmatrix} \tag{4.3}$$

as

$$\dot{c} = c\hat{W}U^T$$

where: $\hat{W}$ = a $2^n \times 2^n$ diagonal reproductive fitness matrix whose $\hat{W}_{ii}$ entry is defined by the rate equation:

$$\dot{c}_i = c_i\hat{W}_{ii}$$

if the system has no mutation; where U is the $2^n \times 2^n$ stochastic matrix of mutation rates u_{ij}, with u_{ij} giving the mutation probabilities $i \to j$; where U^T is the transpose of the matrix U. With these substitutions we can rewrite (4.3):

$$\dot{c} = (c_1, c_2, \ldots, c_n) \begin{pmatrix} \hat{W}_{1,1} & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & \hat{W}_{n,n} \end{pmatrix} \begin{pmatrix} u_{1,1} & u_{1,2} & \ldots & u_{1,n} \\ u_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ u_{n,1} & \ldots & \ldots & u_{n,n} \end{pmatrix}^T . \tag{4.4}$$

[4] In the Eigen and Schuster quasispecies model, "the value matrix W, according to eq 10 and 11, has exclusively positive entries," Eigen, McCaskill, and Schuster [29, p. 6885, bottom left].

Applying the transpose operator to (4.4) we get

$$\dot{c}^T = \left(U^T\right)^T \ \hat{W}^T \ c^T = U \ \hat{W} \ c^T.$$

It suffices to find the dominant right (positive) eigenvector of $U \ \hat{W}$; i.e. a positive column vector v such that

$$U \ \hat{W} \ v = \lambda v.$$

We'll make the same simplifications as in the IMN model [5], that

$$u_{ij} = u^{h(i,j)}(1-u)^{n-h(i,j)}$$

where $h(i,j) =$ the Hamming distance between the base 2 representations of i and j.

4.3.2 Matlab

The following Matlab m-file calculates the quasispecies equilibrium as discussed in Section 4.3.1 (page 322). It calls a function "hamdist.m," whose code can be found in Section 5.3.4 (page 360).For code execution speed see [6].

```
% QuasiSpeciesAug2010.m     m-file name
% calculates the quasispecies equilibrium  eigenvector
% outputs quasispecies equilibrium vector in order
%
% ----- user input -----------------------------------
n = 5;  %choose n = number of mutations to escape
        %choose 2^n = m+1 pretreatment reproductive ratios
wVector = ones(1,2^n,'double');        % default Wii = 1.0
```

[5] See Section 5.3.3 (page 359) for details on the binary aspect of the IMN and quasispecies models.

[6] On a Dell Vostro 1500 laptop, containing an Intel T8100 dual core processor running at 2.1 GHz, the code runs fast (in a second or so) for systems with up to about $n = 5$ mutations. When $n = 9$ there are $2^9 = 512$ different types and the square matrices involved in the calculation contain $512^2 = 262,144$ entries. When n = 9 the run time is about 35 seconds.

```
wVector(1,1) = 2;    % define Woo > Wii [w0 w1 w2 w3 w4 . . . ]
                     % type 0 = matlab 1, type i = matlab i+1
                     % type m = 2^n = 1 is matlab 2^n
u = .005;            % choose mutation rate
% ------ end user input ---------------------------------
U = zeros(2^n, 2^n);    % create U=u_ij matrix
for i=1:2^n
    for j = 1:2^n
        U(i,j) = u^hamdist(i-1,j-1,n)*(1-u)^(n-hamdist(i-1,j-1,n));
    end
end
W = zeros(2^n, 2^n);   % create W diagonal vector
for i=1:2^n
    W(i,i) = wVector(1,i);
end
digits(64)     % accurate display of answers, used with vpa
QSindex = 0;   % initialize QSindex, which is the index number of qs
[EigVects EigValues]= eig(U*W); % Matlab finds Eigenvectors, values
for j =1:2^n                   % pick out which one is Quasispecies
        if abs( sum(  sign(EigVects(:,j)))) == 2^n
            QSindex = j;
        end
end
QS = vpa(EigVects(:,QSindex)/sum(EigVects(:,QSindex))) %normalize QS
% UW = U*W;               % optional check that QS really is QS
% QS - UW*QS/sum(UW*QS)   % should be very close to zero (slows program)
```

The output of the above code, the quasispecies equilibrium vector in column vector form, is shown below. The genotypes of the entries, from top to bottom, are in the usual binary code ordering:

$$00000,\ 00001,\ 00010,\ 00011,\ 00100, \ldots, 11110,\ 11111.$$

QS =

.95074384962175162439024234117823652923107147216796875000000000000

.95566629506103218522072850760196160990744829177856445312500000000e-2

.95566629506103114438664292151770496275275945663452148437500000000e-2

.14409941040115513454303308726878185552777722477912902832031250000e-3

.95566629506103114438664292151770496275275945663452148437500000000e-2

.14409941040115508033292446299356015515513718128204345703125000000e-3

.14409941040115518875314171154400355590041726827621459960937500000e-3

.31383306560022953690810513377451584915434068534523248672485351560e-5

.95566629506103114438664292151770496275275945663452148437500000000e-2

.14409941040115508033292446299356015515513718128204345703125000000e-3

.14409941040115518875314171154400355590041726827621459960937500000e-3

.31383306560022928279822095748441412865759048145264387130737304690e-5

.14409941040115513454303308726878185552777722477912902832031250000e-3

.31383306560022932514986832019943108207371551543474197387695312500e-5

.31383306560022945220481040834448194232209061738103628158569335940e-5

.90983507628810500071080981897286843462779870606027543544769287110e-7

.95566629506103114438664292151770496275275945663452148437500000000e-2

.14409941040115513454303308726878185552777722477912902832031250000e-3

.14409941040115518875314171154400355590041726827621459960937500000e-3

.31383306560022928279822095748441412865759048145264387130737304690e-5

.14409941040115518875314171154400355590041726827621459960937500000e-3

.31383306560022932514986832019943108207371551543474197387695312500e-5

.31383306560022936750151568291444803548984054941684007644653320310e-5

.90983507628810486836191181048844045520240797486621886491775512700e-7

.14409941040115518875314171154400355590041726827621459960937500000e-3

.31383306560022932514986832019943108207371551543474197387695312500e-5

.31383306560022936750151568291444803548984054941684007644653320310e-5

.90983507628810473601301380200401247577701724367216229438781738280e-7

.31383306560022919809492623205438022182534041348844766616821289060e-5

.90983507628810473601301380200401247577701724367216229438781738280e-7

.90983507628810500071080981897286843462779870606027543544769287110e-7

```
.32970886840212373718911821645271773006946602890820940956473350 52e-8
```

We will see this code again when we glue the pre and the post-treatment phases of the model together.

Part II.B.

Post Treatment: Branching Processes

Introduction to Part II.B.

According to David G. Kendall's [7] 1974 presidential address to the London Mathematical Society [53]:

> Branching-process theory, the reader may like to be reminded, is that part of mathematics which deals with the growth and decay of populations of objects which multiply and replace one another, generation by generation, according to rules in which chance plays a prominent part. In the earliest work these objects were always human males, and interest was focussed on the rate of diminution of the stock of family names ('surnames'). In contemporary applications the objects might be heterozygotes carrying a mutant gene, customers waiting in a queueing system, or neutrons in a nuclear reactor, to mention only three of the more important examples.

Up until relatively recently, the origin of the theory of branching processes was attributed to Francis Galton [8] and Henry Watson [9] for their work on the extinction of family names: Galton [37] (1873); Watson and Galton [88] (1874).

For example, according to Krishna B. Athreya and P. E. Ney's classic text *Branching Processes* [6] (1972):

> The study of branching processes has a long history, which, as might be expected is closely interwoven with a number of applications in the physical and biological science. The original problem[10], which was introduced by Francis Galton in 1873 ... and first successfully attacked by the

[7] David Kendall (English mathematician, statistician 1918 – 2007) is considered the father of modern probability theory in Britain [55].

[8] Francis Galton (English scientist, 1822 - 1911) was a cousin of Charles Darwin. Galton, a polymath, explored Africa and made contributions to anthropology, meteorology, the study of genetics and heredity, biometrics, psychology, and statistics [18].

[9] Henry William Watson (English mathematician, 1827 – 1903), main mathematical interests: mathematical physics [18].

[10] The original problem posed by Francis Galton, from *Education Times* [37] (April 1873):

> PROBLEM 4001: A large nation, of whom we will only concern ourselves with adult males, N in number, and who each bear separate surnames colonise a district. Their

> Reverend Henry Watson in that year [Watson and Galton [88] (1874)], was in fact concerned with the extinction of family names in the British peerage.

In actuality, the theory of branching processes seems to have been originated 28 years earlier by the French mathematician Irénée-Jules Bienaymé [11] [11] (1845) working on an almost identical question as the one Galton posed. This earlier origin for branching processes was discovered in 1972 by Christopher C. Heyde and Eugene Seneta [45].

Bienaymé did not give an explicit method of solution, rather he gave results, but the manner in which he reported his results make it seem likely that he used techniques somewhat similar to those developed by Watson and Galton in [88] (1874) – the same techniques that are used today to solve the problem. The key insight was to represent the distribution of the number of male children who survive to adulthood as a generating function. Quoting from Watson's analysis in [88]:

> Let then $\frac{a_0}{100}$, $\frac{a_1}{100}$, $\frac{a_2}{100}$, etc., up to $\frac{a_q}{100}$, be denoted by the symbols t_0, t_1, t_2, etc., up to t_q, be the chances in the first and each succeeding generation of any individual man, in any generation having no son, one son, two sons, and so on, who reach adult life...
>
> Now if any surname have p representatives in any generation it follows from the ordinary theory of chances that the chance of that same surname having s representatives in the next succeeding generation is the coefficient

law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five. Find (1) what proportion of their surnames will have become extinct after r generations; and (2) how many instances there will be of the surname being held by m persons.

[11] Irénée-Jules Bienaymé (French mathematician and statistician, 1796 – 1878). For an account of Irénée-Jules Bienaymé contributions to mathematics and place in history, the interested readers is directed to Heyde and Senata's book *I. J. Bienaymé. Statistical theory anticipated* [46] (1979).

of x^s in the expansion of the multinomial

$$(t_0 + t_1 x + t_2 x^2 +, \text{ etc. } + t_q x^q)^p$$

Watson realized that if he lets $f(x) = (t_0 + t_1 x + t_2 x^2 +, \text{ etc. } + t_q x^q)$ then the iterates of $f(x)$ yield the distribution of sons who live to adulthood, maintaining their family name in succeeding generations; and that the "constant term" of the i^{th} iteration represents the probability of the family name being extinct at the time of i^{th} generation.

After his brilliant insights, Watson makes an error in his analysis, and comes to the false conclusion that the probability of extinction generally approaches 1 given enough time [40].

This is in contrast to Bienaymé's correct analysis. We quote Heyde and Senata's [46] English translation of Bienaymé [11] (1845) [12]:

> The analysis also shows clearly that if the mean ratio [of families which survive in each generation] is greater than unity, the probability of extinction of families with the passing of time no longer reduces to certainty.

So, as Kendall [53] writes:

> [Unlike Galton and Watson] Bienaymé had not merely discussed what we have come to think of as the Galton-Watson problem, but had done so in such a way as to make it plain that he was in possession of the whole Criticality Theorem, for $m < 1$, $m = 1$, and $m > 1$ [m being the mean ratio].

In Chapter 5 we discuss discrete and continuous multi-type Galton-Watson branching processes. These are similar to the one dimensional case investigated by Bienaymé,

[12] see also [40, p. 378]

Galton, and Watson. However, one has the complication of multiple types and the possibility of mutation from one type to another. In the continuous case, rather than having discrete generations, one allows the time needed for one generation to go to zero. The continuous case, as a result, involves differential equations. In Chapter 5 we rigorously prove some existence theorems related to multi-type branching processes using multi-type generating functions, techniques from several complex variables and results about differential equations in $\mathbb{C}^n$.

The motivation for our investigation of multi-type Galton-Watson branching processes is that they are used to model (and then calculate extinction probabilities in) the stochastic post-treatment phase of the Iwasa, Michor, and Nowak model [48, 49].

In Chapter 5 we also discuss some numerical methods regarding calculating extinction probabilities. Chapter 5 also sees us joining the pre and post treatment phases.

Chapter 6 contains a result regarding differential equation in $\mathbb{C}^n$ which is used in Chapter 5, but due to its length and independent nature, was given its own chapter.

Branching processes are important in applied mathematics. The interested reader is directed to the following resources for current applications to biology:

Marek Kimmel and Davdi E. Axelrod's *Branching Processes in Biology* [54] (2002), which includes sections on the polymerase chain reaction [13], various types of mutating biological systems, and telomere shortening [14].

Patsy Haccou, Peter Jagers [15], and Vladimir A. Vatutin's *Branching processes: variation, growth, and extinction of populations* [41] (2007), which includes chapters on modeling measles outbreaks and reverse branching processes (coalescent processes).

[13] Polymerase chain reaction (or PCR) is used to amplify a small quantity of DNA into quantities which can be analyzed or otherwise used. Kary Banks Mullis received the 1993 Nobel Prize for developing the PCR technique.

[14] Telomeres are repetitive sections of DNA on the ends of the chromosome which protect the chromosome from degradation during replication. They shorten with each replication and are an active area of research for those studying cancer and aging processes.

[15] See also Peter Jagers' *Branching Processes with Biological Applications* [50] (1975).

Chapter 5

Post-treatment: Branching Processes

Introduction to Chapter 5

In Chapter 5 we discuss multi-type Galton-Watson branching processes. We rigorously prove some existence theorems related to multi-type branching processes using multi-type generating functions, techniques from several complex variables and results about differential equations in $\mathbb{C}^n$. We discuss the usage of multi-type Galton-Watson branching processes in Iwasa, Michor, and Nowak's model [48, 49] to calculate extinction probabilities. In Chapter 5 we also discuss some numerical methods regarding calculating extinction probabilities in the post-treatment phase. Chapter 5 also sees us joining the pre and post treatment phases.

5.1 Introduction to the Branching Process

It is best to use an example from biology and to dispense with the stochastic aspects until later. We have a single cell. It splits, then its progeny split, and so on. See

Figure 5.1. If we let

$$N(g) = \text{ population size of the } g^{th} \text{ generation,}$$

and the system runs like a clock, then

$$N(g) = 2^g.$$

More generally, if we start with a population of $N(0)$ cells, then

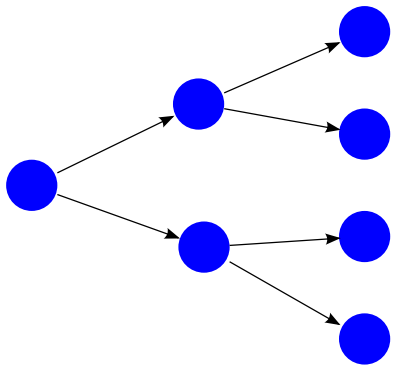


Figure 5.1: Branching Process.

$$N(g) = N(0)\ 2^g. \tag{5.1}$$

We can turn (5.1) into a continuous model parameterized by time t by allowing g in (5.1) to take on non integer values

$$N(t) = N(0)\ 2^{\frac{\text{generation}}{\text{unit of time}}\ t}. \tag{5.2}$$

We can also build the same model starting from the differential equation

$$\frac{d}{dt}N(t) = kN(t). \tag{5.3}$$

One interprets the differential equation (5.3) to mean the growth rate is proportional to the population size. The solution of the differential equation (5.3) is of course

$$N(t) = N(0)e^{kt}. \tag{5.4}$$

We set (5.4) equal to (5.2) and take the ln of both sides

$$\begin{aligned} N(0)e^{kt} &= N(0)\ 2^{\frac{\text{generation}}{\text{unit of time}}\ t} \\ \ln(N(0)) + kt &= \ln(N(0)) + \left(\frac{\text{generation}}{\text{unit of time}}\ln(2)\right)\ t \end{aligned} \tag{5.5}$$

Both sides of (5.5) are straight lines w.r.t. t; both lines having the same 'y-intercept', $\ln(N(0))$. The slopes of the two lines are

$$k \text{ and } \left(\frac{\text{generation}}{\text{unit of time}}\ln(2)\right).$$

So to equate the two models, it suffices to set

$$k = \frac{\text{generation}}{\text{unit of time}}\ln(2).$$

From an experimental perspective usually one finds k in (5.4) by curve fitting, e.g. by plotting the population at time t on log paper, as the "pure" model, $N(g) = N(0)\ 2^g$ doesn't take into consideration cell death and other factors. See[1].

5.1.1 Aside about the Hilbert Projective Metric

See Figure 5.2. The Hilbert Projective Metric on $\mathbb{R}^2_{>0}$ can be based upon the projection of rays onto the hyperbolic line $\mathbb{R}_{>0}$, where the hyperbolic distance d_H between

[1]There is a large body of literature on the experimental determination of growth rate parameters. For a typical research article by experimental biologists on this topic see [16]. For a mathematical treatment of population models, see the classic book, *Mathematical Biology I. An Introduction*, Murray [72] (2002).

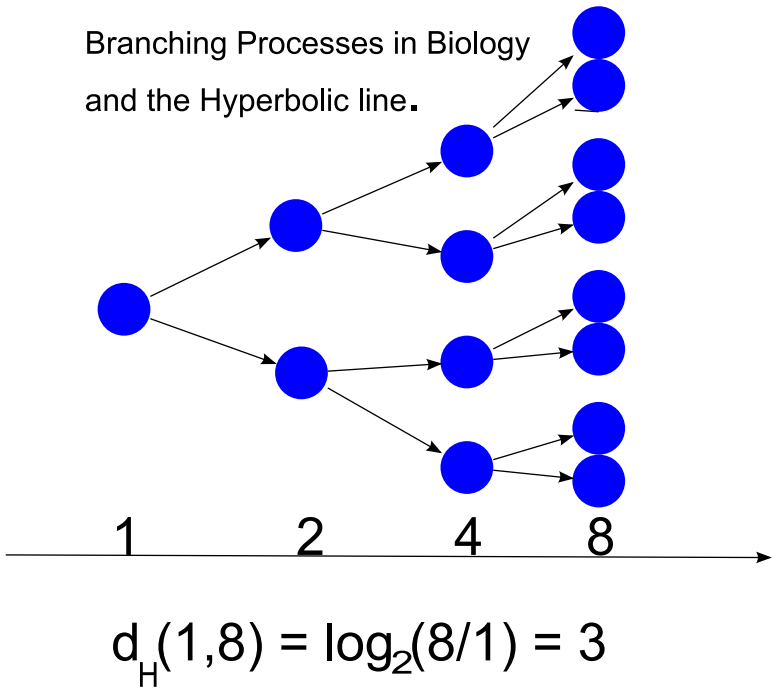


Figure 5.2: Branching Process and the Hyperbolic Line.

x and y in $\mathbb{R}_{>0}$ has been defined to be $\ln(x/y)$ where $x \geq y$.

It is common in biology to look at log plots of the population size to empirically find the growth rate constant. It is also common to talk about the number of generations separating two species. However, the following interpretation of the hyperbolic line and its metric d_H does not seem to have made its way into the standard treatments of population dynamics.

Consider a population growing deterministically according to the above branching process modeled by Equation (5.1) with growth rate R, rather than 2, so that

$$N(g) = N(0)R^g.$$

If we use $\log_R$ instead of $\log_e = \ln$ to define d_H we get

$$d_H(N(g_1), N(g_2)) = |\log_R (R^{g_1}/R^{g_2})| = \left|\log_R \left(R^{g_1-g_2}\right)\right| = |g_1 - g_2|$$

which is just the number of generations separating the two populations $N(g_1)$ and $N(g_2)$.

The point of this small insight is that there is a natural connection between the population dynamics and the metric d_H.

5.1.2 Stochastic Branching Processes

If one repeats an experiment and gets different outcomes, it is a sign that there is either a hidden or unmeasured variable that is changing from one experiment to the next; or (at least in quantum theory) that stochasticity is an unavoidable aspect of the system. In such situations one can turn to stochastic models that will yield a distribution of possible outcomes, which if the model is well done should match the distribution of the experimental results.

The deterministic branching process discussed in Section 5.1 can be turned into a stochastic process by assigning the following probabilities to the possible outcomes for each generation:

$$\begin{aligned} p_0 &= \text{ probability the cell will die} \\ p_1 &= \text{probability the cell will neither die nor replicate} \\ p_2 &= \text{probability the cell will split into two cells.} \end{aligned}$$

To retrieve the deterministic model of Section 5.1 we can set $p_0 = p_1 = 0$ and $p_2 = 1$. It will be useful to encode the distribution p_0, p_1, p_2 as the probability generating function (PGF)

$$f(x) = p_0 + p_1 x + p_2 x^2. \tag{5.6}$$

f determines the branching process since $f^{(n)}(x)$ gives the PGF for the population size for the n^{th} generation.

The expected number of offspring in each generation will be

$$\frac{df}{dx}(1) = p_1 + 2p_2.$$

We can also have a continuous process, modeled on (5.6), which we can develop with a generating function as follows. Let

$$f_{\Delta t}(x) = \Delta t \cdot D + (1 - (D + R) \cdot \Delta t)\, x + \Delta t \cdot Rx^2.$$

Factoring out the Δt discrete time step [2] we get

$$f_{\Delta t}(x) = \underbrace{\left(D - (D + R)x + Rx^2\right)}_{F(x)} \Delta t + x = F(x)\Delta t + x. \tag{5.7}$$

Applying the Picard Iteration - Euler Line process to (5.7), letting $\Delta t \to 0$, yields the solution $f(x, t)$ to the differential equation

$$\frac{\partial}{\partial t} f(x, t) = F(x).$$

$f(x, t)$ will be a probability generating function describing the stochastic evolution of the system. We make this argument rigorous in Section 5.2 (page 341) for the case $D = 1$.

It is worth mentioning that the probability of extinction for this process is 1 if $D \geq R$ and $\frac{D}{R}$ if $D < R$. We get this result by solving the quadratic equation $F(x) = 0$, see [3]. We can think of D as being a death rate and R as being a reproductive rate.

[2] We are assuming that Δt is sufficiently small that $0 < \Delta t \leq 1/(D + R)$.

[3] See Theorem 5.3.1.2 (page 354) and its proof for a justification of this result.

We calculate the expected number of offspring for (5.7) in the time interval Δt:

$$\frac{\partial}{\partial x} f_{\Delta t}(1) = (R - D)\Delta t + 1. \tag{5.8}$$

In (5.8) we replace Δt with g/n where $g =$ the average time needed for one generation. Iterating (5.8), with the aforementioned substitution, n times, will yield the expected population size after g time has passed (if the process was a discrete time process):

$$\left((R - D)\frac{g}{n} + 1\right)^n. \tag{5.9}$$

Using l'Hôpital's rule [4, Chapter 10] together with the ln function we get from (5.9):

$$\lim_{n\to\infty} \left((R - D)\frac{g}{n} + 1\right)^n = e^{(R-D)g}.$$

So, the units of R and D should be $1/g$, see [4]. If r is the (unit-less) mean reproductive ratio for the time period of length g, then we should have

$$e^{R-D} = r$$

which implies

$$R - D = \ln r.$$

5.1.3 Mean Life Span = 1/D

If we somehow follow a single individual template, how long should we expect that particular template to live?

[4] See Section 3.1 (page 301) for a similar result about units regarding the quasispecies model for the pre-treatment phase of the IMN model.

Since:

$$P(\text{dies at } 1\ \Delta t) = (\Delta t \cdot D)$$
$$P(\text{dies at } 2\ \Delta t) = (1 - \Delta t \cdot D)\,(\Delta t \cdot D)$$
$$P(\text{dies at } 3\ \Delta t) = (1 - \Delta t \cdot D)^2\,(\Delta t \cdot D).$$

The expected life span will be:

$$\begin{aligned} E(\text{lifespan}) &= \sum_{i=1}^{\infty} (i\Delta t)\,(1 - \Delta t \cdot D)^{i-1}\,(\Delta t \cdot D) \\ &= (\Delta t)^2 D \sum_{i=1}^{\infty} i\,(1 - \Delta t \cdot D)^{i-1}\,. \end{aligned}$$

We have:

$$\begin{aligned} \frac{d}{dx}\ \sum_{i=0}^{\infty} x^i &= \sum_{i=0}^{\infty} i x^{i-1} \\ &= \sum_{i=1}^{\infty} i x^{i-1} \end{aligned}$$

and

$$\begin{aligned} \frac{d}{dx}\ \sum_{i=0}^{\infty} x^i &= \frac{d}{dx}\ \frac{1}{1-x} \\ &= \frac{1}{(1-x)^2} \end{aligned}$$

so $\sum_{i=1}^{\infty} ix^{i-1} = \frac{1}{(1-x)^2}$. Letting $x = (1 - \Delta t \cdot D)$ we get:

$$\begin{aligned} E(\text{lifespan}) &= (\Delta t)^2 D \sum_{i=1}^{\infty} i \left(1 - \Delta t \cdot D\right)^{i-1} \\ &= (\Delta t)^2 D \ \frac{1}{\left(1 - (1 - \Delta t \cdot D)\right)^2} \\ &= (\Delta t)^2 D \ \frac{1}{(\Delta t \cdot D)^2} \\ &= 1/D. \end{aligned}$$

So, in the Branching Process model discussed in the previous section, the average template lasts 1/D before degrading.

5.1.4 Definition of the Multi-Type Branching Process

Definition 5.1.4.1. (Athreya) [6, 7] See [5]. A continuous k-dimensional multi-type branching process is a Markov Process [6]

$$\{X(t,\omega);\ t \geq 0,\ \omega \in \Omega\}, \quad \omega : \mathbb{R}_{\geq 0} \to \mathbb{Z}^k_{\geq 0}, \quad X(t,\omega) = \omega(t),$$

that satisfies the following property given in generating function form:

Let $\mathbf{i}, \mathbf{j}$ be k-dimensional non-negative indices [7]. The probability generating function giving the transition probabilities for the system going from state $\mathbf{i}$ at time $t = 0$

[5] In the biological model we can think of the random variable $X(t,\omega)$ as being the population size vector for the k different types at time $t \geq 0$ if the system's evolution is given by ω. Each $\omega \in \Omega$ is one possible development of the system. Also, $P(\mathbf{i}, \mathbf{j}, t) = P\left(X(t_0 + t) = \mathbf{j} \mid X(t_0) = \mathbf{i}\right)$ for all $t, t_0 \geq 0$.

[6] See Norris [74] (1999) for an excellent introduction to Markov chains.

[7] For this definition we will think of $\mathbf{i}$ being fixed and $\mathbf{j}$ varying over all possible k-dimensional non-negative indices. We will use $\mathbf{j}$ in this capacity twice, on both sides of the equal sign. This is the notation used by Athreya; it is slightly confusing at first glance.

to state $\mathbf{j}$ at time t satisfies:

$$g_{\mathbf{i}}(s;0,t) = g_{\mathbf{i}}(s,t) = \sum_{\mathbf{j}} P(\mathbf{i},\mathbf{j},t)s^{\mathbf{j}} = \prod_{r=1}^{k}\left(\sum_{\mathbf{j}} P_{\mathbf{e}_r,\mathbf{j}}(t)s^{\mathbf{j}}\right)^{i_r}$$

for each possible index $\mathbf{i}$ where

$$\mathbf{i} = (i_1, i_2, \ldots, i_r, \ldots, i_k) \text{ is fixed, } \mathbf{e}_r = \left(0, \ldots, 0, \underbrace{1}_{r^{th} \text{ entry}}, 0, \ldots, 0\right),$$

and

$$s^{\mathbf{i}} = \prod_{r=1}^{k} s_r^{i_r}.$$

Again, the generating function gives all the transition probabilities $\rightarrow \mathbf{j}$ assuming the system is in state $\mathbf{i}$ initially. The Markov Property implies that

$$g_{\mathbf{i}}(s;0,t) = g_{\mathbf{i}}(s;t_1,t_1+t).$$

It is important to emphasize that each type gives rise to independent lines of descent. In the model that we are focused on, the various types are cells, bacteria, virus, templates, or particles.

5.2 The Iwasa, Michor, and Nowak (IMN) Model and Existence

The stochastic post-treatment phase of the Iwasa, Michor, and Nowak (IMN) model [48, 49] is modeled as a continuous time Galton-Watson branching process.

In their model the "continuous time" branching process is defined in terms of "discrete time" multi-type generating functions, each having a time step of Δt. The time step Δt is allowed to go to zero (somehow, hopefully) yielding a continuous Galton-

Watson process. Using techniques from several complex variables and differential equations in $\mathbb{C}^n$ we show that this indeed the case. In particular, we prove below that in the limit, as $\Delta t \to 0$, the discrete multi-type generating functions appearing in their model become continuous ones, determining a continuous multi-type Galton Watson branching process. Our [8] approach seems somewhat novel.

In the "discrete time" probability generating functions in the IMN model, a single cell of type j will in a time step of Δt:

1. Die with probability Δt
2. Replicate one additional template with probability $R_j \Delta t$
3. The one additional template will be of type i with probability $R_j u_{ji} \Delta t$. Note $\sum_i u_{ji} = 1$.
4. Nothing happens with probability $1 - (1 + R_j)\Delta t$

This yields a family of infinitesimally defined discrete generating functions $g_{\Delta t,j}(z)$, one for each j and each Δt

$$\begin{aligned} g_{\Delta t,j}(z) &= \Delta t + (1 - (1 + R_j)\Delta t)\, z_j + R_j \Delta t\ z_j \sum_{i=1}^{n} u_{ji} z_i \\ &= \underbrace{\left(1 - (1 + R_j) z_j + R_j z_j \sum_{i=1}^{n} u_{ji} z_i\right)}_{F_j(z)} \Delta t + z_j \end{aligned}$$

The generating functions $g_j(z,t)$ are built out of the above family.

We can think of $g_{\Delta t,j}(z)$ as an approximation of g_j expanded at $t = 0$. We can think of this, from the perspective of the model, as taking a sliver of time sufficiently

[8] I am deeply indebted to my advisor Yunping Jiang for suggesting we consider using complex variables when I could not prove this result otherwise.

small that there is not enough time for more than one replication; or the probability of there being more than one replication is of the order $(\Delta t)^2$. Explicitly determining $g_j(z,t)$ can be mathematically difficult, depending on the complexity of the various $g_{\Delta t,j}(z)$. So a continuous process may be defined in terms of $g_{\Delta t,j}(z)$ $j = 1, 2, \ldots, n$, rather than explicitly be given.

The following discussion is a somewhat original construction of the continuous time $g_j(z,t)$ and proof of its existence in terms of the $g_{\Delta t,j}(z)$.

The iteration scheme. Define the family of vector valued generating functions indexed by Δt:

$$\begin{aligned}
G^0_{\Delta t}(z) &= z \\
G^1_{\Delta t}(z) &= (g_{\Delta t,1}(z), \ldots, g_{\Delta t,n}(z)) \\
&= (F_1(z), \ldots, F_n(z))\,\Delta t + z \\
&= F(z)\Delta t + z \\
G^2_{\Delta t}(z) &= G^1_{\Delta t}\left(G^1_{\Delta t}(z)\right) \\
&= F(G^1_{\Delta t}(z))\Delta t + G^1_{\Delta t}(z) \\
G^{n+1}_{\Delta t}(z) &= G^1_{\Delta t}\left(G^n_{\Delta t}(z)\right) \\
&= F(G^n_{\Delta t}(z))\Delta t + G^n_{\Delta t}(z).
\end{aligned}$$

$G^n_{\Delta t}(z)$ is a vector valued generating function; i.e. each component is a generating function. Its j^{th} component $G^n_{\Delta t,j}(z)$ has Taylor expansion in generalized powers of z, the coefficients of which give the probability that staring with 1 of the j types, one will have after n iterations (after $n\Delta t$ time) a population mixture with the powers of z. I.e. If n = 2, then the coefficient of $z_1^3 z_2^{30}$ of will be the probability that there are 3 of the type 1 and 30 of the type 2. This is, of course, if the population reproduces and dies according to the $g_{\Delta t,i}(z)$ discrete basic transition probabilities.

We can suggestively write $G^n_{\Delta t}(z)$ as $G_{\Delta t}(z, n\Delta t)$.

In the Euler Process for numerically solving ODE's $\dot{y} = F(y,t);\ y(0) = y_0$ the iteration scheme is:

$$y_{n+1} = F\left(y_n, n\Delta t\right)\Delta t + y_n.$$

But if F doesn't explicitly depend on t, we can write the Euler scheme as:

$$y_{n+1} = F\left(y_n\right)\Delta t + y_n.$$

To show that the y_n are dependent on the step size Δt we write $y_{\Delta t}(n)$. The Euler Line $Y_{\Delta t}(t)$ is the piecewise linear in t map:

$$Y_{\Delta t}(t) = \begin{cases} y_{\Delta t}(n) & \text{if } t = n\Delta t\ ; \\ (1-s)\ y_{\Delta t}(n) + s\ y_{\Delta t}(n+1) & \text{if } t = n\Delta t + s,\ 0 < s < \Delta t. \end{cases}$$

Clearly, iterating $G_{\Delta t}(z)$ is the Euler Iteration. The Euler Lines, in terms of G are:

$$G_{\Delta t}(z,t) = \begin{cases} G_{\Delta t}(z, n\Delta t) = G^n_{\Delta t}(z) & \text{if } t = n\Delta t\ ; \\ (1-s)\ G_{\Delta t}(z, n\Delta t) + s\ G_{\Delta t}(z, (n+1)\Delta t) & \text{if } t = n\Delta t + s,\ 0 < s < \Delta t. \end{cases}$$

We show that the Euler Lines $G_{\Delta t}(z,t)$ defined above are generating functions:

For each $t = n\Delta t$, $G_{\Delta t}(z,t)$ will be a vector valued generating function since the composition of generating functions is again a generating function.

If $t = n\Delta t + s,\ 0 < s < \Delta t$, $G_{\Delta t}(z,t)$ will again be a vector valued generating function since, if G and G' are generating functions, then $(1-s)G + sG'$ will also be a generating function.

One can show, using Picard Iteration, that the Euler Lines converge uniformly on a compact neighborhood to a continuous function (solution). In Chapter 6 (page 375) we prove this result for Euler Lines in $\mathbb{C}^n$.

So the Euler Lines $G_{\Delta t}(z,t)$ converge uniformly to some $G(z,t)$. Moreover, each Euler Line is analytic (in several complex variables since it is an absolutely convergent Taylor Series on $\{(|z_1|,\ldots,|z_n|) < (1,\ldots,1)\}$). By Wierestraus theory, in several complex variables, the uniform limit of analytic functions will be analytic and the coefficients of the Taylor series of the functions in the sequence will converge to the coefficients of the limit function, see [87]. So we have shown that $G(z,t)$ is a vector value generating function since it has Taylor Series; its coefficients are non-negative; and $G((1,1,\ldots,1),t) = (1,1,\ldots,1)$, which follows from

$$G_{\Delta t}((1,1,\ldots,1),t) = G_{\Delta t}((1,1,\ldots,1),n\Delta t) = (1,1,\ldots,1).$$

Since we have shown that $G(z,t) = (g_1(z,t),\ldots,g_n(z,t))$ is a vector valued continuous multi-type probability generating function, and since $g_1(z,t),\ldots,g_n(z,t)$ uniquely determines the multi-type Galton-Watson branching process (due to the independence of the particles), we are done.

See [9].

5.3 Calculating Extinction Probabilities

5.3.1 Calculating Extinction Probabilities (Theory)

In this section we will develop a technique to calculate the extinction probabilities for the continuous multi-type Galton-Watson process used in the stochastic post-treatment phase of the Iwasa, Michor, and Nowak (IMN) model [48, 49].

Note, that we will use the convention of vectors and vector valued functions being set in **bold** typeface.

The Galton-Watson process is developed in IMN [49, p. 312] as follows:

[9] The above argument, is done in the 1 dimensional case in Section 5.5 (page 371), with a few extra details which help to relate this construction to the more standard one given in [6].

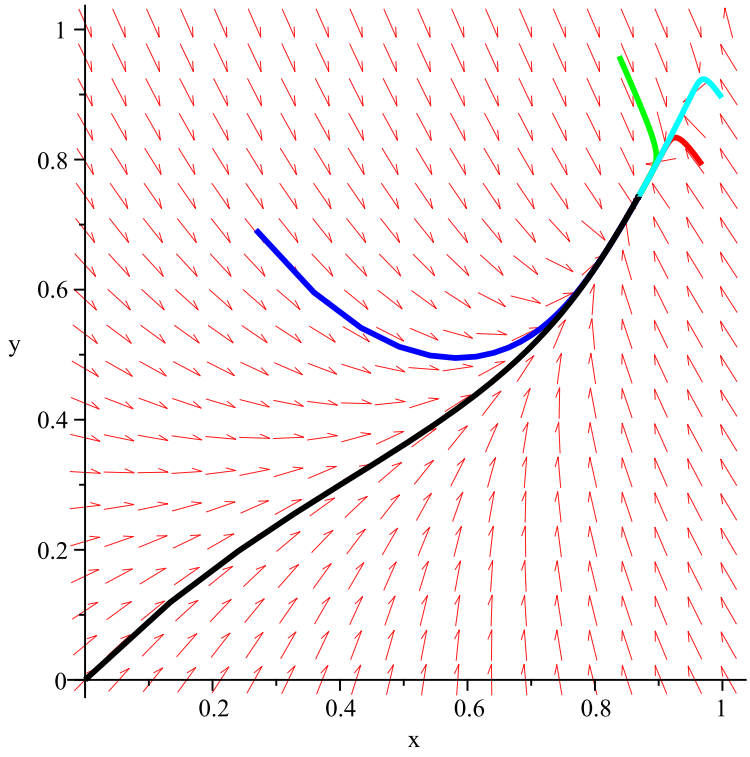


Figure 5.3: Finding extinction probabilities. The black path starting at $(0\,,0)$ terminates at the vector of extinction probabilities. The red arrows indicate the direction of the vector field $\mathbf{F}(\mathbf{z})$, where $\frac{\partial}{\partial t}\mathbf{G}(\mathbf{z},t)=\mathbf{F}(\mathbf{z})$ and where for all $\Delta t>0$ sufficiently small, $\mathbf{F}(\mathbf{z})\Delta t+\mathbf{z}$ is a vector valued multi-type PGF. In general, for such a system, any trajectory; e.g. the various colored trajectories illustrated, which originates in $[0,1]^{m+1}$ will terminate at the vector of extinction of probabilities, unless the starting point is $(1,1,\ldots,1)=\mathbf{1}$. If the starting point is $\mathbf{1}$, then the trajectory is stationary and remains at $\mathbf{1}$ and thus terminates at the vector of extinction probabilities only if extinction is 100% certain for every type. Similarly, iterating $\mathbf{F}(\mathbf{z})\Delta t+\mathbf{z}$ from any starting value of $\mathbf{z}\in[0,1]^{m+1}\setminus\mathbf{1}$ will yield a sequence of points which converge to the vector of extinction probabilities, with $\mathbf{1}$ being stationary again. The vector of extinction probabilities is also stationary. See Theorem 5.3.1.2 (page 354). For the Maple commands used to create this figure see Section 5.3.5 (page 365).

The authors use continuous time probability generating functions (PGF's), $g_j(\mathbf{z}, t)$, to encode the probabilities of the process. The PGF, $g_j(\mathbf{z}, t)$, encodes the probabilities for a process which starts with a single 'cell' or 'template' of type j at time $t = 0$. The $g_j(\mathbf{z}, t)$ are not explicitly defined. Instead $g_j(\mathbf{z}, t + \Delta t)$ is described in terms of $g_j(\mathbf{z}, t)$, see equation (5.10) below.

Equation (5.10) is copied exactly from [49, p. 312] and contains typographic errors in the subscripts of R and in some of the g's. We correct these errors in Equation (5.11).

$$g_j\left(\mathbf{z}, t + \Delta t\right) = \Delta t \bullet 1 + R_i \Delta t g_j\left(\mathbf{z}, t\right) \left\{ \sum_{i=1}^{n} u_{ji} g_j\left(\mathbf{z}, t\right) \right\} + \left(1 - \left(1 + R_i\right) \Delta t\right) g_j\left(\mathbf{z}, t\right). \tag{5.10}$$

See [10] for a note about the $\bullet$ notation. We fix the typographic errors in (5.10) by changing R_i to R_j and some of the g_j to g_i so as to be consistent with the IMN model and subsequent equations appearing in [49]. We get:

$$g_j\left(\mathbf{z}, t + \Delta t\right) = \Delta t \bullet 1 + R_j \Delta t g_j\left(\mathbf{z}, t\right) \left\{ \sum_{i=1}^{n} u_{ji} g_i\left(\mathbf{z}, t\right) \right\} + \left(1 - \left(1 + R_j\right) \Delta t\right) g_j\left(\mathbf{z}, t\right). \tag{5.11}$$

In [49] Iwasa, Michor, and Nowak justify (5.10), or rather (5.11) by stating, more or less [11], that

$$g_{\mathbf{e}_i + \mathbf{e}_j}\left(\mathbf{z}, t\right) = g_{\mathbf{e}_i}\left(\mathbf{z}, t\right)\ g_{\mathbf{e}_j}\left(\mathbf{z}, t\right) = g_i\left(\mathbf{z}, t\right)\ g_j\left(\mathbf{z}, t\right) \tag{5.12}$$

where $g_{\mathbf{e}_i + \mathbf{e}_j}$ is the generating function for the system staring with one template of type i and one of type j. Then they use (5.11) to obtain a system of non linear

[10] The $\Delta t \bullet 1$ notation used in (5.10) is due to the 1 representing $z_1^0 z_2^0 \ldots z_n^0$. The coefficient of $z_1^0 z_2^0 \ldots z_n^0$ corresponds to extinction.

[11] Actually Iwasa, Michor, and Nowak in [49] write (5.12) in terms of expectation functions which are equivalent to (5.12).

differential equations (copied from [49, p. 212]):

$$\frac{\mathrm{d}}{\mathrm{dt}}\, g_j = (1-g_j) + \left\{ u_{jm}(g_m - 1) + \sum_{\substack{i\neq j \\ i\neq m}} u_{ji}(g_i - 1) + (g_j - 1)\left(1 - \sum_{i\neq j} u_{ji}\right)\right\} g_j R_j. \tag{5.13}$$

We should be using $\frac{\partial}{\mathrm{dt}}$ instead of $\frac{\mathrm{d}}{\mathrm{dt}}$; however, we are keeping IMN's notation. In the IMN model the templates have types $0, 1, \ldots, m$, so the 'type' index j runs from 0 to m. Iwasa, Michor, and Nowak switch to that convention in (5.13). The vector of extinction probabilities will be a stationary point of (5.13) [12]. So Iwasa, Michor, and Nowak in [49] set the left hand sides (the derivatives) in (5.13) equal to zero and then use various, sometimes difficult to follow methods to approximate a solution.

Here we develop an alternative method for finding the vector of extinction probabilities. As we do this, we interpret what g_j means in the context of the IMN model and give an alternative justification of (5.11).

Equation (5.13) is complicated so we derive a simpler version using (5.11) and the definition of the partial derivative:

$$\begin{aligned}
\frac{\partial}{\partial t}\, g_j(\mathbf{z}, t) &= \lim_{\Delta t \to 0} \frac{g_j(\mathbf{z}, t+\Delta t) - g_j(\mathbf{z}, t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\Delta t \bullet 1 + R_j \Delta t g_j \left\{\sum_{i=0}^{m} u_{ji} g_i\right\} + (1 - (1+R_j)\,\Delta t)\, g_j - g_j}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{\left(1 + R_j g_j \left\{\sum_{i=0}^{m} u_{ji} g_i\right\} - (1+R_j)\, g_j\right) \Delta t}{\Delta t} \\
&= 1 + R_j g_j \left\{\sum_{i=0}^{m} u_{ji} g_i\right\} - (1+R_j)\, g_j \\
&= \underbrace{1 - (1+R_j)\, g_j + R_j g_j \left\{\sum_{i=0}^{m} u_{ji} g_i\right\}}_{F_j(g_0, g_1, \ldots, g_m)}.
\end{aligned} \tag{5.14}$$

[12] See Theorem 5.3.1.2 for a proof of this.

In consideration of (5.14) we define

$$F_j(\mathbf{z}) = F_j(z_0, z_1, \ldots, z_m) = 1 - (1 + R_j)\, z_j + R_j z_j \left\{ \sum_{i=0}^{m} u_{ji} z_i \right\} \tag{5.15}$$

so that we can write (5.14) as

$$\frac{\partial}{\partial t}\, g_j = F_j(g_0, g_1, \ldots, g_m) \tag{5.16}$$

and (5.11) as

$$g_j(\mathbf{z}, t + \Delta t) = F_j(g_0(\mathbf{z}, t), g_1(\mathbf{z}, t), \ldots, g_m(\mathbf{z}, t))\Delta t + g_j(\mathbf{z}, t). \tag{5.17}$$

We will use (5.16) and (5.17) later.

We let $t = 0$ and note that $g_j\,(\mathbf{z}, 0) = z_0^0 z_1^0 \ldots z_j^1 \ldots z_n^0 = z_j$ so (5.11) implies:

$$\begin{aligned}
g_j\,(\mathbf{z}, \Delta t) &= \Delta t \bullet 1 + R_j \Delta t g_j\,(\mathbf{z}, 0) \left\{ \sum_{i=1}^{n} u_{ji} g_i\,(\mathbf{z}, 0) \right\} + (1 - (1 + R_j)\,\Delta t)\, g_j\,(\mathbf{z}, 0) \\
&= \Delta t \bullet 1 + R_j \Delta t z_j \left\{ \sum_{i=1}^{n} u_{ji} z_i \right\} + (1 - (1 + R_j)\,\Delta t)\, z_j \\
&= \Delta t \bullet 1 + (1 - (1 + R_j)\,\Delta t)\, z_j + R_j \Delta t z_j \left\{ \sum_{i=1}^{n} u_{ji} z_i \right\} \qquad (5.18) \\
&= \underbrace{\left(1 - (1 + R_j)\, z_j + R_j z_j \left\{ \sum_{i=1}^{n} u_{ji} z_i \right\} \right)}_{F_j(\mathbf{z})} \Delta t + z_j \\
&= F_j(\mathbf{z})\Delta t + z_j\ . \qquad (5.19)
\end{aligned}$$

We will use (5.19) later.

Using (5.18), which comes from (5.11) when t = 0, we interpret g_j in terms of the IMN model: (5.18) implies that during an interval Δt of short du-

ration [13], a single template or cell of type j only has sufficient time to do one of the following three things:

1. Die with probability Δt.

2. Replicate one additional template with probability $R_j \Delta t$. More specifically, replicate one additional template of type i with probability $R_j u_{ji} \Delta t$.

   The additional template and the original template will then do nothing; i.e. they will not die or replicate, until the Δt time step ends.

   u_{ji} is the probability that a type j template will, when it replicates, replicate a type i template. We have $\sum_i u_{ji} = 1$. See [14].

3. Do nothing, neither die, nor replicate. This must happen with probability

$$1 - (1 + R_j)\Delta t.$$

We return to finding the vector of extinction probabilities.

If we fix Δt so that

$$0 \leq \Delta t \leq \frac{1}{\max_{j=0,\ldots,m}(1 + R_j)} \tag{5.20}$$

then $g_j(\mathbf{z}, \Delta t),\ j = 0, 1, \ldots, m$ acts as a 'discrete time' PGF; see [15].

[13] See (5.20) (page 350) for an upper bound on Δt. From a modeling perspective, Δt is implicitly assumed to be sufficiently small that there is essentially zero probability that the original template can replicate more than once, or can replicate and then die during Δt. Moreover, it is assumed that there is not sufficient time remaining in Δt for any offspring produced in that interval to replicate, or to die. Alternatively, we could think think of (5.18) as a quadratic (second degree) approximation.

[14] During an interval of duration Δt a single template of type j replicating a single template of type i has probability $R_j u_{ji} \Delta t$. This outcome, $z_j^1 z_i^1$, together with its probability, $R_j u_{ji} \Delta t$, is represented in (5.18) as the term $R_j \Delta t z_j\ u_{ji} z_i$. The z_j represents the original j-type template that continues to live after it produces a single i-type template; the i-type template that was produced is represented by z_i. u_{ji} is called the mutation probability and obviously we must have $\sum_i u_{ji} = 1$.

[15] If we fix Δt sufficiently small, then for each $j,\ j = 0, \ldots, m,\ g_j(\mathbf{z}, \Delta t)$, from (5.18), describes a "discrete time step" multi-type probability generating function (PGF) provided its coefficients are all between 0 and 1 and sum to 1. This will happen provided $\Delta t \geq 0$ is sufficiently small that Δt satisfies $\max_{j=0,1,\ldots,m}(1 + R_j)\Delta t \leq 1$. This is equivalent to condition (5.20).

We define

$$\mathbf{G}(\mathbf{z},t) = (g_0(\mathbf{z},t),\ldots,g_m(\mathbf{z},t))\,. \tag{5.21}$$

By (5.19):

$$\begin{aligned}\mathbf{G}(\mathbf{z},\Delta t) &= (g_0(\mathbf{z},\Delta t),\ldots,g_m(\mathbf{z},\Delta t))\\ &= (F_0(\mathbf{z})\Delta t + z_0,\ldots,F_m(\mathbf{z})\Delta t + z_m)\\ &= \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z}. \end{aligned}\tag{5.22}$$

(5.22) defines a 'vector' of discrete time multi-type PGF's, provided Δt is fixed and satisfies (5.20). It is sometimes convenient to write $\mathbf{G}_{\Delta t}(\mathbf{z})$ for $\mathbf{G}(\mathbf{z},\Delta t)$.

We give an alternative justification of (5.11) if we accept (5.11) at $t = 0$.

In general, for continuous multi-type Galton-Watson branching processes, if

$$\mathbf{f}(\mathbf{s},t) = (f_1(\mathbf{s},t),\ldots,f_n(\mathbf{s},t))$$

with f_i being the the multi-type PGF for when the process starts with one particle of type i; and if $t,u \geq 0$ then

$$\mathbf{f}(\mathbf{s},t+u) = \mathbf{f}(\mathbf{f}(\mathbf{s},u),t). \tag{5.23}$$

See [7, p. 349, Eq. (3b)]. So, by (5.23) and (5.22) [16]:

$$\begin{aligned}\mathbf{G}(\mathbf{z},t+\Delta t) &= \mathbf{G}(\mathbf{G}(\mathbf{z},t),\Delta t)\\ &= \mathbf{G}\left((g_0(\mathbf{z},t),\ldots,g_m(\mathbf{z},t)\right),\Delta t)\\ &= \mathbf{F}((g_0(\mathbf{z},t),\ldots,g_m(\mathbf{z},t)))\Delta t + (g_0(\mathbf{z},t),\ldots,g_m(\mathbf{z},t))\,. \end{aligned}\tag{5.24}$$

[16] (5.22) comes from (5.11) if we set $t = 0$.

By (5.17), the j^{th} argument of (5.24) is $g_j(\mathbf{z}, t + \Delta t)$ as defined in (5.11). So if we except (5.11) at $t = 0$, then (5.11) must hold whenever $t \geq 0$.

We return again to finding the vector of extinction probabilities.

The following theorem about discrete time multi-type Galton-Watson processes will set up our method for finding the extinction probabilities in the IMN model. The PGF's $g_0, g_1, \ldots, g_m$ appearing in the following theorem are not necessarily the ones defined in the IMN model.

Theorem 5.3.1.1. ***Harris*** *[42]. Let* $\mathbf{M}$ *be the matrix of first moments* [17] *for a discrete time multi-type Galton Watson process. Suppose there are* $m + 1$ *types in the process, denoted* $0, 1, \ldots, m$*, and that the process is determined by the multi-type PGF's* [18]

$$g_0(\mathbf{z}), g_1(\mathbf{z}), \ldots, g_m(\mathbf{z}).$$

Let $G(\mathbf{z}) = (g_0(\mathbf{z}), g_1(\mathbf{z}), \ldots, g_m(\mathbf{z}))$*. Suppose the process is positively regular* [19] *and not singular* [20]*. Let* ρ *be the dominant eigenvalue of* $\mathbf{M}$ *and let* $\mathbf{q} = (q_0, q_1, \ldots, q_m)$ *be the vector of extinction probabilities, so that* q_i *is the probability that the line of descent of a single individual 'cell' of type* i *will eventually become extinct. Then*

[17] **Matrix of first moments** [42, Definition 4.1, p. 36]: the square matrix $\mathbf{M} = m_{ij}$ with

$$m_{ij} = \frac{\partial g_i}{\partial z_j}(\mathbf{1})$$

is called the matrix of first moments. $\mathbf{1} = (1, 1, \ldots, 1)$.

[18] $g_i(\mathbf{z})$ is the PGF for the state of a system after its first time step, if the initial state of the system was a single cell of type i. $\mathbf{z} = (z_0, z_1, \ldots, z_m)$.

[19] **Positively regular** [42, Definition 5.2, p. 38]: A multi-type Galton-Watson process is positively regular if the $\mathbf{M}^n > 0$ for some n where $\mathbf{M} = m_{ij}$ is the matrix of first moments with

$$m_{ij} = \frac{\partial g_i}{\partial z_j}(\mathbf{1}).$$

Note, if $\mathbf{M}$ is non-negative (which it is) and if $\mathbf{M}^n > 0$, then $\mathbf{M}$ is said to be primitive. If $\mathbf{M}$ is primitive, then by the Perron-Frobenius Theorem there exists a simple (non-repeated) eigenvalue $\rho > 0$ such that $\rho > |\rho'|$ if ρ' is any other eigenvalue of $\mathbf{M}$. We will call ρ the dominant eigenvalue of $\mathbf{M}$.

[20] **Singular** [42, Definition 6.2, p. 39]: A multi-type Galton-Watson process is singular if each of the generating functions $g_j(\mathbf{z})$ is linear in $\mathbf{z} = z_0, z_1, \ldots, z_m$ with no constant terms. I.e. each object

1. *If* $\rho \leq 1$, *then* $\mathbf{q} = \mathbf{1}$.

2. *If* $\rho > 1$, *then* $\mathbf{0} \leq \mathbf{q} < \mathbf{1}$, *and* $\mathbf{q}$ *satisfies* $\mathbf{q} = \mathbf{G}(\mathbf{q})$.

3. *If* $\boldsymbol{q}'$ *is any vector in the unit cube* [21] *other than* $\mathbf{1}$ *then*

$$\lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{q}') = \mathbf{q}.$$

4. *The only solutions of the equation* $\mathbf{z} = \mathbf{G}(\mathbf{z})$ *in the unit cube are* $\mathbf{q}$ *and* $\mathbf{1}$ *with the possibility* $\mathbf{q} = \mathbf{1}$.

Proof. Harris proves 1. and 2. in Theorem 7.1 (p. 41); 3. in Theorem 7.2 (p. 42); and 4. in Corollary 1. of Theorem 7.2 (p. 42), all in in Chapter II of [42]. □

The following theorem is designed for the continuous time multi-type Galton-Watson branching process found in the IMN model [48]; and more generally, for any continuous branching process which can be defined via

$$\mathbf{G}(\mathbf{z}, t + \Delta t) = \mathbf{F}(\mathbf{G}(\mathbf{z}, t))\Delta t + \mathbf{G}(\mathbf{z}, t) \qquad \text{PGF} \tag{5.25}$$

provided $\Delta t > 0$ is sufficiently small. When $t = 0$ (5.25) becomes

$$\mathbf{G}(\mathbf{z}, \Delta t) = \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z} \qquad \text{PGF .} \tag{5.26}$$

Actually $\mathbf{G}$ satisfying (or being defined by) (5.26) for Δt small, is sufficient for the following theorem to be applicable. The second part of this theorem and its proof seem somewhat original; however, related results are certainly in the literature [6, 42].

---

has exactly one child. E.g. the process on two types defined by

$$g_0(\mathbf{z}) = .2z_0 + .8z_1$$
$$g_1(\mathbf{z}) = z_1$$

is considered singular.

[21] Unit cube $= [0, 1]^{m+1}$.

Theorem 5.3.1.2. *Suppose there are $m+1$ types, denoted $0, 1, \ldots, m$, in a continuous multi-type Galton Watson process and that the process is determined by the multi-type continuous PGF's* [22]

$$g_0(\mathbf{z}, t), g_1(\mathbf{z}, t), \ldots, g_m(\mathbf{z}, t).$$

Let

$$\mathbf{G}(\mathbf{z}, t) = (g_0(\mathbf{z}, t), g_1(\mathbf{z}, t), \ldots, g_m(\mathbf{z}, t))$$

and let $\widetilde{\mathbf{q}} = (\widetilde{q}_0, \widetilde{q}_1, \ldots, \widetilde{q}_m)$ be the vector of extinction probabilities, so that $\widetilde{q}_i$ is the probability that the line of descent of a single individual 'cell' of type i will eventually become extinct in this continuous process. Then $\forall\ t \geq 0$

$$\mathbf{G}(\widetilde{\mathbf{q}}, t) = \widetilde{\mathbf{q}} \quad \textit{and} \quad \frac{\partial}{\partial t}\, \mathbf{G}(\widetilde{\mathbf{q}}, t) = \mathbf{0}. \tag{5.27}$$

Suppose

$$\frac{\partial}{\partial t}\, \mathbf{G}(\mathbf{z}, t) = \mathbf{F}(\mathbf{z}). \tag{5.28}$$

Further suppose there exists a $\delta > 0$ such that $0 \leq \Delta t \leq \delta$ implies that

$$\mathbf{H}_{\Delta t}(\mathbf{z}) = \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z} \tag{5.29}$$

is a vector valued discrete time multi-type probability generating function yielding a positively regular, not singular process. Let $\mathbf{q}$ be the vector of extinction probabilities for this discrete process. Then

$$\mathbf{q} = \widetilde{\mathbf{q}} \tag{5.30}$$

[22] $g_i(\mathbf{z}, t)$ is the PGF for the state of a system which initially is a single cell of type i, so $g_i(\mathbf{z}, 0) = z_i$. $\mathbf{z} = (z_0, z_1, \ldots, z_m)$.

and so if $\mathbf{q}'$ *is any vector in the unit cube* $[0,1]^{m+1}$ *other than* $\mathbf{1}$ *then*

$$\lim_{n\to\infty} \mathbf{H}_{\Delta t}^{(n)}(\mathbf{q}') = \widetilde{\mathbf{q}}. \tag{5.31}$$

Proof. To prove the left side of (5.27); i.e. to prove $\mathbf{G}(\widetilde{\mathbf{q}}, t) = \widetilde{\mathbf{q}}$:

By (5.23)(page 351) [23] we have for any fixed $\Delta t > 0$

$$\mathbf{G}(\mathbf{0}, n\Delta t) = \underbrace{\mathbf{G}^{(n)}(\mathbf{0}, \Delta t) = \mathbf{G}\left(\mathbf{G}^{(n-1)}(\mathbf{0}, \Delta t), \Delta t\right)}_{\mathbf{G}^{(n)}(\mathbf{z}, \Delta t) \text{ is recursively defined by setting } \mathbf{G}^{(0)}(\mathbf{z}, t) = \mathbf{z}} \tag{5.32}$$

and since

$$\widetilde{\mathbf{q}} = \lim_{t\to\infty} \mathbf{G}(\mathbf{0}, t) = \lim_{n\to\infty} \mathbf{G}(\mathbf{0}, n\Delta t) = \lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{0}, \Delta t) \tag{5.33}$$

it follows from the continuity of $\mathbf{G}(\mathbf{z}, \Delta t)$ and (5.33) that

$$\begin{aligned} \mathbf{G}(\widetilde{\mathbf{q}}, \Delta t) &= \mathbf{G}\left(\lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{0}, \Delta t), \Delta t\right) \\ &= \lim_{n\to\infty} \mathbf{G}\left(\mathbf{G}^{(n)}(\mathbf{0}, \Delta t), \Delta t\right) \\ &= \lim_{n\to\infty} \mathbf{G}^{(n+1)}(\mathbf{0}, \Delta t) \\ &= \widetilde{\mathbf{q}}. \end{aligned} \tag{5.34}$$

So the left side of (5.27) is proven and the right side of (5.27), which is $\frac{\partial}{\partial t}\mathbf{G}(\widetilde{\mathbf{q}}, t) = \mathbf{0}$, follows trivially.

To prove (5.30); i.e. to prove $\mathbf{q} = \widetilde{\mathbf{q}}$:

Combining (5.27) and (5.28) leads to $\mathbf{F}(\widetilde{\mathbf{q}}) = \mathbf{0}$. But then $\mathbf{H}_{\Delta t}(\widetilde{\mathbf{q}}) = \widetilde{\mathbf{q}}$. This implies, by Part 4. of the previous theorem, Theorem 5.3.1.1 (page 352), that $\widetilde{\mathbf{q}} = \mathbf{q}$ or $\mathbf{1}$.

If $\widetilde{\mathbf{q}} \neq \mathbf{1}$ or $\mathbf{q} = \mathbf{1}$ we're done. So suppose $\mathbf{q} \neq \widetilde{\mathbf{q}} = \mathbf{1}$. Then there exists some j

[23] Or see e.g. Athreya [7, p. 349, Eq. (3b)].

in $0, 1, \ldots m$, such that $q_j < \widetilde{q}_j = 1$. Then, since

$$\lim_{t\to\infty} \; g_j(\mathbf{0}, t) = \widetilde{q}_j = 1$$

there exists some $t^* > 0$ such that

$$g_j(\mathbf{0}, t^*) > q_j. \tag{5.35}$$

Note that $g_j(\mathbf{0}, t^*)$ is the j^{th} component of $\mathbf{G}(\mathbf{0}, t^*)$.

We will derive a contradiction of (5.35) using the Picard Iteration – Euler Line method from the existential theory of ODE's . Let us restrict our attention to uniform partitions of $[0, t^*]$ of the form

$$0 = 0\Delta t < 1\Delta t < 2\Delta t < \cdots < N_{\Delta t}\Delta t = t^*, \qquad t_i = i\Delta t \;\; i = 0, 1, \ldots N_{\Delta t} \tag{5.36}$$

with $N_{\Delta t} = t^*/\Delta t$.

The polygonal Euler Lines [24], with vertices [25]

$$\mathbf{H}_{\Delta t}^{(n)}(0), \qquad n = 0, 1, \ldots, N_{\Delta t}$$

[24]For more details on Picard Iteration and Euler Lines see Chapter 6 (page 375) and for their application to iterates of PGF's see Section 5.2 (page 341).

[25]The vertices $\mathbf{y}_n, \; n = 0, 1, \ldots, N_{\Delta t},$ of the polygonal Euler Line corresponding to the differential equation (5.28) and (5.37)

$$\frac{\partial}{\partial t}\, \mathbf{G}(\mathbf{z}, t) = \mathbf{F}(\mathbf{z}) \quad \text{with I.C. } \mathbf{G}(\mathbf{0}, 0) = \mathbf{0}.$$

and to the Δt partition of $[0, t^*]$ (5.36) are

$$\begin{aligned} \mathbf{y}_0 = \mathbf{0}, \qquad \mathbf{y}_n &= \mathbf{F}(\mathbf{y}_{n-1})(t_n - t_{n-1}) + \mathbf{y}_{n-1}, \qquad n = 1, 2, \ldots, N_{\Delta t} \\ &= \mathbf{F}(\mathbf{y}_{n-1})\Delta t + \mathbf{y}_{n-1} \\ &= \mathbf{H}_{\Delta t}(\mathbf{y}_{n-1}) \\ &= \mathbf{H}_{\Delta t}^{(n)}(\mathbf{0}) \end{aligned}$$

since, as defined in (5.29), $\mathbf{H}_{\Delta t}(\mathbf{z}) = \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z}$.

See Section 6.1.2 (page 376) for details on partitions and Euler Lines.

will converge uniformly on $[0, t_*]$ to $\mathbf{G}(\mathbf{0}, t)$, the solution of (5.28)

$$\frac{\partial}{\partial t}\,\mathbf{G}(\mathbf{z}, t) = \mathbf{F}(\mathbf{z}) \quad \text{with I.C. } \mathbf{G}(\mathbf{0}, 0) = \mathbf{0} \tag{5.37}$$

as $\Delta t \to 0$.

So, as $\Delta t \to 0$ the final vertex, $\mathbf{H}_{\Delta t}^{(N_{\Delta t})}(\mathbf{0})$, of the polygonal Euler Line corresponding to the Δt partition of $[0, t^*]$ (5.36), converges to $\mathbf{G}(\mathbf{0}, t^*)$.

In particular, as $\Delta t \to 0$, the j_{th} component of $\mathbf{H}_{\Delta t}^{(N_{\Delta t})}(\mathbf{0})$, which we will denote

$$h_{j,\Delta t}^{(N_{\Delta t})}(\mathbf{0})$$

converges to the j_{th} component of $\mathbf{G}(\mathbf{0}, t^*)$, which is $g_j(\mathbf{0}, t^*)$.

Recalling that we are assuming that $g_j(\mathbf{0}, t^*) > q_j$, so we have

$$\lim_{\Delta t \to 0} h_{j,\Delta t}^{(N_{\Delta t})}(\mathbf{0}) = g_j(\mathbf{0}, t^*) > q_j$$

So, there exists a $\Delta t > 0$ such that

$$h_{j,\Delta t}^{(N_{\Delta t})}(\mathbf{0}) > q_j. \tag{5.38}$$

(5.38) is a contradiction for the following reason:

By Part 3. of Theorem 5.3.1.1 (page 352):

$$\lim_{n \to \infty}\, \mathbf{H}_{\Delta t}^{(n)}(\mathbf{0}) = \mathbf{q} \tag{5.39}$$

but

$$\mathbf{H}_{\Delta t}^{(n)}(\mathbf{0}) \leq \mathbf{H}_{\Delta t}^{(n+1)}(\mathbf{0}),$$

because every coefficient of $\mathbf{H}_{\Delta t}(\mathbf{z})$ is non-negative, so

$$\mathbf{H}_{\Delta t}^{(N_{\Delta t})}(\mathbf{0}) \leq \mathbf{q}$$

and

$$h_{j,\Delta t}^{(N_{\Delta t})}(\mathbf{0}) \leq q_j$$

which contradicts (5.38):

$$h_{j,\Delta t}^{(N_{\Delta t})}(\mathbf{0}) > q_j.$$

Finally, we can prove (5.31): that if $\mathbf{q}' \in [0,1]^{m+1}$ and $\mathbf{q}' \neq \mathbf{1}$ then

$$\lim_{n\to\infty} \mathbf{H}_{\Delta t}^{(n)}(\mathbf{q}') = \widetilde{\mathbf{q}}. \tag{5.40}$$

Since $\widetilde{\mathbf{q}} = \mathbf{q}$ (5.40) follows from from Part 3. of Theorem 5.3.1.1 (page 352). □

5.3.2 Calculating Extinction Probabilities (Numerically)

We showed in Section 5.3.1 (page 345) that in the IMN model [48, 49] that the continuous process and the vector of extinction probabilities is determined by

$$\mathbf{G}(\mathbf{z}, \Delta t) = (g_0(\mathbf{z}, \Delta t), g_1(\mathbf{z}, \Delta t), \ldots, g_m(\mathbf{z}, \Delta t))$$

where

$$g_j(\mathbf{z}, \Delta t) = \Delta t + (1 - (1 + R_j)\,\Delta t)\, z_j + R_j \Delta t z_j \left\{ \sum_{i=1}^{n} u_{ji} z_i \right\}. \tag{5.41}$$

Theorems 5.3.1.1 and 5.3.1.2 tell us we calculate the vector of extinction probabilities by simply iteration of $\mathbf{G}(\mathbf{z}, \Delta t)$ provided $\mathbf{G}$ determines a positively regular and not singular process. See Theorem 5.3.1.1 for definitions. Provided $u_{ij} > 0$, a trivial calculation verifies 'positively regular'; 'not singular' is obviously true provided $\Delta t > 0$.

So, in consideration of the Section 5.3.1 and Theorems 5.3.1.1 and 5.3.1.2 we can calculate the extinction probabilities for the IMN model by iterating

$$\mathbf{G}(\mathbf{q}', \Delta t)$$

starting from any $\mathbf{q}' \in [0,1]^{m+1},\ \mathbf{q}' \neq \mathbf{1}$; provided Δt satisfies (5.20) (page 350) which we reproduce here:

$$0 \leq \Delta t \leq \frac{1}{\max_{j=0,\ldots,m}(1+R_j)}. \tag{5.42}$$

The iteration should be fairly stable, as any starting point in the unit cube (except $\mathbf{1}$) will, in theory, lead to the correct solution. So, if due to round off errors, less than exact values occur in the calculation, the process should, I imagine, still lead to the correct solution as $\mathbf{G}$ is well-behaved and continuous. However, a sophisticated discussion of the convergence rate, optimal algorithms, and rigorously dealing with the issue of round-off error is beyond the scope of this dissertation. Moreover, as our iteration scheme is basically the Euler method for solving differential equations, switching to the higher order Runge-Kutter methods [26] may offer an improvement in convergence speed. These topics of computation and numerical analysis would make for interesting further study.

5.3.3 Binary aspect of genotypes in the IMN model

In the IMN model it is assumed that there are $1, 2, 3, \ldots, n$ loci where a mutation can occur. If a mutation is present (or not) at a locus, that locus is said to be in state 1 (or state 0). So, if there are n loci, then there are 2^n possible genotypes. The wild

[26] See Butcher [20] for an extensive treatment of Euler and Runge-Kutter numerical methods of solving differential equations.

type has no mutations and genotype

$$\underbrace{00\cdots0}_{n \text{ zeros}}.$$

The escape mutant has all the possible mutations giving it genotype

$$\underbrace{11\cdots1}_{n \text{ ones}}$$

Let $i \in 0, 1, \ldots, m = 2^n - 1$. The i mutant type will be the mutant having genotype the base 2 representation of i.

Example: Suppose $n = 3$. So there are three loci were mutations can occur and 2^3 different types. Type $0 = 000$; type $1 = 001$, type $2 = 010$, ..., type $7 = 111$.

A simplifying assumption in the IMN model is that each single mutation is equally likely and that all mutations are independent. So if we let $u =$ probability of a single mutation, then

$$u_{ij} = u^{h(i,j)}(1-u)^{n-h(i,j)} \tag{5.43}$$

where $h(i, j) =$ the Hamming distance between the base 2 representations of i and j.

Example: Suppose $n = 3$. Then

$$h(1,7) = h(001,111) = 2 \text{ and } u_{1,7} = u^2(1-u)^1.$$
$$h(0,7) = h(000,111) = 3 \text{ and } u_{0,7} = u^3(1-u)^0.$$

5.3.4 Matlab

The following Matlab m-files implement an algorithm to calculate the extinction and escape probabilities. The algorithm assumes the simplifications expressed in (5.43) (page 360).

The first m-file is the main one. It calls three other functions which are given at

the end of this section. However, the user only needs to interact with the main m-file.

```
% IterateG.m
% This m-file iterates the multi-type generating function
% GenFunctionG from the IMN model. GenFunctionG is constructed
% from parameters input below:
'IterateG'      % displays name of this m-file
% --------- user input ----------------------------------------
its = 8000;        % number of times to iterate process
u = 0.005;          % mutation probability per loci per replication
n = 5;              %  number of loci that can mutate
dt = .15;           % time step 'Delta t' < max 1/(1 + maxR)
R = ones(2^n,1);    % reproductive fitnesses  vector, default is all 1's
R(2^n,1) = 2;       % reproductive fitnesses of escape mutant
% --------- end user input ------------------------------------
U = uMutationMatrix(u,n);  % create mutation matrix u_ij
Z = zeros(2^n,1);  % initiate column vector of extintion probabilities defaults = 0
Z(2^n,1) = 1/R(2^n,1); % extinction probability of escape mutant ...
                                                  if no mutation is 1/Rm
EscapeProbs = zeros(2^n,1); % column vector to hold escape probabilities defaults = 0
for i = 1:its               % iteration process for extinction probabilities
    Z = GenFunctionG(dt,Z,n,R,U);
end
for i = 1: 2^n        % filling vector to hold escape probabilities
   Zescape(i,1) = 1 - Z(i,1);
end
['After ' num2str(its) ' iterations: ESCAPE probabilities ...
                                  (in binary order 000, 001, etc)']
digits(64)            % to display high accuracy numbers with vpa( )
vpa(Zescape(:,1))     % display Escape Probs in binary order 000, 001, 010, etc
```

The output of the above code, the vector of escape probabilities is displayed below. The genotypes are in binary order.

```
ans = IterateG

ans = After 8000 iterations: ESCAPE probabilities (in binary order 000, 001, etc)

ans =

 .49958493075084042800426686881110072135925292968750000000000000000e-2

 .59886992663560345562245856854133307933807373046875000000000000000e-2

 .59886992663552574001073480758350342512130737304687500000000000000e-2

 .80189912600627799577068799408152699470520019531250000000000000000e-2

 .59886992663557014893171981384512037038803100585937500000000000000e-2

 .80189912600625579131019549095071852207183837890625000000000000000e-2

 .80189912600625579131019549095071852207183837890625000000000000000e-2

 .13621319934506215254543803894193843007087707519531250000000000000e-1

 .59886992663549243331999605288729071617126464843750000000000000000e-2

 .80189912600625579131019549095071852207183837890625000000000000000e-2

 .80189912600618917792871798155829310417175292968750000000000000000e-2

 .13621319934505993209938878862885758280754089355468750000000000000e-1

 .80189912600618917792871798155829310417175292968750000000000000000e-2

 .13621319934505993209938878862885758280754089355468750000000000000e-1

 .13621319934506215254543803894193843007087707519531250000000000000e-1

 .40080255502995432159707434038864448666572570800781250000000000000e-1

 .59886992663557014893171981384512037038803100585937500000000000000e-2

 .80189912600625579131019549095071852207183837890625000000000000000e-2

 .80189912600618917792871798155829310417175292968750000000000000000e-2

 .13621319934505882187636416347231715917587280273437500000000000000e-1

 .80189912600625579131019549095071852207183837890625000000000000000e-2

 .13621319934505882187636416347231715917587280273437500000000000000e-1

 .13621319934506104232241341378539800643920898437500000000000000000e-1

 .40080255502995432159707434038864448666572570800781250000000000000e-1

 .80189912600618917792871798155829310417175292968750000000000000000e-2

 .13621319934505327076124103768961504101753234863281250000000000000e-1

 .13621319934505882187636416347231715917587280273437500000000000000e-1

 .40080255502995432159707434038864448666572570800781250000000000000e-1

 .13621319934505438098426566284615546464920043945312500000000000000e-1
```

```
.40080255502995432159707434038864448666572570800781250000000000000e-1
.40080255502995432159707434038864448666572570800781250000000000000e-1
   .48836891396559956035616778535768389701843261718750000000000000000
```

Notice that the escape mutant has escape probability of 0.4883. Without mutation, the probability of escape for the escape mutant is $1/R_m$ assuming $R_m > 1$. In the above m-file, $R_m = 2$ and $1/R_m = 1/2 = 0.5$, which differs from 0.4883 by about .0116. The probability of escape for the escape mutant is decreased due to back mutation.

Iwasa, Michor, and Nowak in [48, 49] are a little confusing on this issue of back mutation. For example they use a mutation rate [48, p. 2577, top right] of

$$u_{ij} = u_{ij}^h (1-u)^{n-h_{ij}}$$

which we have incorporated into the above m-file. On the other hand, they will approximate the escape probability of the escape mutant to be $1 - 1/R_m$ – which ignores back mutation. One should keep this in mind while reading Iwasa, Michor, and Nowak [48, 49].

The next m-file is a function, it contains the code for the generating function.

```
% GeneratingFunctionG.m       = name of this function m-file
% multitype generating function for Iwasa Michor Nowak model
%
% dt = time step Delta t
% the Z vector is of length 2^n, for the 2^n different types
% n = the number of loci where a mutation can occur
% R vector is of length 2^n containing the reproductive ratios
% the U matrix is 2^n x 2^n containing the mutation rates u_ij
% the  output vector GenOut is of length 2^n
%
function GenOut =  GenFunctionG(dt,Z,n,R,U)
```

```
GenOut = zeros(2^n,1);
for j = 1 : 2^n
    GenOut(j) = dt + (1 - (1 + R(j))*dt)*Z(j) + R(j)*dt*Z(j)*(U(j,:)*Z);
end
```

The next m-file is the function which creates the mutation matrix U. It assumes uniform mutation probability u at each loci and independence of mutation; see (5.43) (page 360).

```
% uMutationMatrix   = function m-file name
% Creates a 2^n x 2^n mutation matrix; i.e. U
% assuming uniform locus mutation rate u.
% n = the number of loci where mutations can happen
%
function matrixOut = uMutationMatrix(u, n)
matrixOut = zeros(2^n, 2^n);
for i = 1 : 2^n
  for j = 1 : 2^n
   matrixOut(i,j) = u^(hamdist(i-1,j-1,n))*(1-u)^(n-hamdist(i-1,j-1,n));
  end
end
```

The next function m-file calculates the Hamming distance between two decimal numbers representing the cell's type.

```
% hamdist    = function m-file name
% calculates the Hamming distance between the padded
% binary representation of two decimal numbers dec1, dec2
% n = the number of loci where a mutation can happen
%
function out = hamdist(dec1, dec2, n)
bin1 = dec2bin(dec1,n);
bin2 = dec2bin(dec2,n);
out = 0;
```

```
for i = 1 : n
    if bin1(i) ~= bin2(i)
        out = out + 1;
    end
end
```

5.3.5 Maple

The following Maple commands generate Figure 5.3 (page 346).

```
with(DEtools):

u := .9; Rzero := 6.2*(1/10); Rone := 24*(1/10); dt := 1/(Rzero+Rone+10);

sys := {diff(x(t), t) =
        1-(1+Rzero)*x(t)+Rzero*x(t)*((1-u)*x(t)+u*y(t)),
        diff(y(t), t) = 1-(1+Rone)*y(t)+Rone*y(t)*((1-u)*y(t)+u*x(t))};
DEplot(sys, [x(t), y(t)],
      t = 0 .. 150, x = 0 .. 1.01, y = 0 .. 1.01,
      [[x(0) = .9680, y(0) = .792], [x(0) = .2680, y(0) = .692],
       [x(0) = .8380, y(0) = .959], [x(0) = .999, y(0) = .895],
       [x(0) = 0, y(0) = 0]],
      numpoints = 1000, linecolor = [red, blue, green, cyan, black]);
```

5.4 Combining Pre and Post-treatment Calculations in the IMN Model

The goal of the Iwasa, Michor, and Nowak (IMN) model [48, 49] is to calculate the probability that a treatment will be successful. As we mentioned in the Introduction to Part II:

The dot product of the pre-treatment quasispecies distribution vector and the post-treatment escape probabilities vector yields a probability p,

$$p = \sum_{i=0}^{m} \underbrace{P(\text{ selected is of type } i)}_{\text{from quasispecies equilibrium}} \quad \underbrace{P(\text{ escape } \mid \text{ being of type } i)}_{q_i \text{ from the multi-type branching process}} . \tag{5.44}$$

p is the probability that if one individual virus, bacteria, or cancer cell is selected randomly when treatment begins that the selected individual will create a line of descendants that escapes extinction. In [48, 49] the notation used is

$$P(\text{ selected is of type } i) = x_i \quad \text{and} \quad P(\text{ escape } \mid \text{ being of type } i) = \xi_i.$$

so, in terms of x_i and xi_i, (5.44) becomes

$$p = \sum_{i=0}^{m} x_i \xi_i.$$

If N is the total population size, including all the different types, and we invoke independence, we have that

$$N \cdot p$$

is the mean number of mutants expected to avoid extinction, to escape.

In the IMN model it is assumed that for the typical 'cell', randomly selected, that escaping is rare and the model uses the Poisson Distribution, with mean

$$\lambda = N \cdot p = N \sum_{i=0}^{m} x_i \xi_i$$

to calculate the probability that there will be zero escapes from extinction. According to the Poisson Distribution, the probability that k cells will escape and avoid extinction is

$$P(k) = \frac{\lambda^k}{k!} \, e^{-\lambda}.$$

So the probability that there will be 0 escapes, that all cells' lines of descent will go extinct, that the treatment will be successful is

$$P(0) = \frac{\lambda^0}{0!}\, e^{-\lambda} = e^{-\lambda} = e^{-N\sum_{i=1}^{n} x_i \xi_i}.$$

The probability of at least one escape, meaning that the treatment is a failure, is

$$1 - e^{-N\sum_{i=1}^{n} x_i \xi_i}.$$

In the next section we combine our numerical technics and produce Matlab code which calculates the probabilities discussed in this section.

5.4.1 Matlab

The following Matlab code (TreatmentSuccessProbabilityAug2010.m) carries out the calculations discussed in Section 5.4 (page 365). The user inputs [27]:

1. The number n of mutations separating the wild type from the escape mutant.

   Typical values: $1 \leq n \leq 5$.

2. The mutation rate u at each locus, which is assumed to be uniform [28].

   Typical values for u would be $0 \leq u \leq 10^{-1}$ with $u = 0.00003$ suggested for the HIV virus and 10^{-3} suggested for cancer cells in Iwasa, Michor, and Nowak [48, pp. 2575 – 2576] (2003).

3. The population size at start of treatment, N with $1 \leq N \leq 10^{12}$ being typical.

[27] The values listed here for the user inputs were gleaned from Iwasa, Michor, and Nowak [48, 49] (2003, 2004).

[28] It is assumed that for the mutation matrix $U = \{u_{ij}\}$, that $u_{ij} = u^{h_{ij}}(1-u)^{n-h_{ij}}$ where h_{ij} is the Hamming distance, in binary, between the i-type and the j-type. It is fairly easy to recode U to change the mutation scheme.

4. The $\hat{W}_{ii}$, which are the pretreatment reproductive rate constants assuming no mutation:

   Generally we will suppose that

   $$\hat{W}_{00} > \hat{W}_{ii} \geq 0 \text{ for } i = 1, 2, \ldots m.$$

   Typical values: $\hat{W}_{00} = 1.5,\ \hat{W}_{ii} = .75, \quad i \neq 0.$

5. The R_j, which are the reproductive ratios during the post treatment phase:

   Typical values $1 < R_m < 2$ and $0 \leq R_j < 1$ for $j = 0, 1, \ldots m - 1$.

The program then calculates:

1. The quasispecies equilibrium distribution (the eigenvector); i.e. the "x_i".
2. The vector of extinction probabilities; i.e. the "ξ_i", which are called q_i and $\widetilde{q}_i$ in Section 5.3 (page 345).
3. The probability a randomly chosen "cell" will escape from extinction: $p = \sum_{i=0}^{m} x_i \xi_i$.
4. The number of expected escapes λ assuming the population at the start of treatment is size N: $\lambda = N \cdot p$.
5. The probability of zero escapes; i.e. probability the treatment is successful.
6. The probability of at least one escape; i.e. the probability the treatment fails.

```
% TreatmentSuccessProbabilityAug2010.m     m-file name
% calculates the success probability
% by combining pre and post treatment models
% outputs Probability of success
%
% ----- user input  -----------------------------------
```

```
n = 5;            %choose n = number of mutations to escape
u = .00015;         % mutation rate during pre-treatment
N = 10^12;         % number of pathogens at time of treatment
% ----- user pre-treatment parameters -------------------
%choose 2^n = m+1 pretreatment reproductive ratios
wVector = .6*ones(1,2^n,'double');        % default Wii = 1.0
wVector(1,1) = 1.6;    % wVector(1,1) = wild type
                     % define Woo > Wii [w0 w1 w2 w3 w4 . . . ]
                     % type 0 = matlab 1, type i = matlab i+1
                     % type m = 2^n = 1 is matlab 2^n
% --------- user input post treatment -----------------------
its = 8000;         % number of times to iterate generating function
R = .9*ones(2^n,1);    % reproductive fitnesses  vector, default is all 1's
R(2^n,1) = 1.8;       % reproductive fitnesses of escape mutant R_m > 1
% --------- end user input ----------------------------------------
digits(64)     % accurate display of answers, used with vpa
U = uMutationMatrix(u,n);     % create mutation matrix u_ij
% -------------------------------
W = zeros(2^n, 2^n);   % create W diagonal vector
for i=1:2^n
    W(i,i) = wVector(1,i);
end
QSindex = 0;   % initialize QSindex, which is the index number of qs
[EigVects EigValues]= eig(U*W); % Matlab finds Eigenvectors, values
for j =1:2^n                   % pick out which one is Quasispecies
        if abs( sum(  sign(EigVects(:,j)))) == 2^n
            QSindex = j;
        end
end
QS = vpa(EigVects(:,QSindex)/sum(EigVects(:,QSindex))) %normalize QS column vec
% -------------------------------
dt = .9/(1 + max(R));  % time step 'Delta t' < max 1/(1 + maxR)
Z = zeros(2^n,1);      % initiate column vector of extinction ...
```

```
                                        probabilities defaults = 0
Z(2^n,1) = 1/R(2^n,1); %  extinction probability of escape ...
                            mutant if no mutation is 1/Rm
                        %  this slightly speeds up convergence
for i = 1:its           % iteration process for extinction probabilities
    Z = GenFunctionG(dt,Z,n,R,U);
end
EscapeProbs = zeros(2^n,1);  % column vector to hold escape ...
                                      probabilities defaults = 0
for i = 1: 2^n         % filling vector to hold escape probabilities
   Zescape(i,1) = 1 - Z(i,1);
end
['After ' num2str(its) ' iterations: ESCAPE probabilities ...
                               (in binary order 000, 001, etc)']
EscapeProbs = vpa(Zescape(:,1))     % display Escape Probs ...
                                        in binary order 000, 001, 010, etc
% ------------------------------
p = vpa(dot(QS,EscapeProbs))   % probability a randomly chosen cell will escape
ExpectedEscapes = vpa(N*p)     % expected number of escapes for pop of size N
ProbZeroEscapes = vpa(exp(-N*p)) % treatment is successful
ProbAtLeastOneEscape = vpa(1 - ProbZeroEscapes) % treatment is a failure
```

The output of the above program, for the parameters as shown, was in part:

```
p =                .33879061305413791194064454 6 ...
                           64775903693736740966557716628209888 97e-12
ExpectedEscapes = .33879061305413791194064454 6 ...
                           64775903693736740966557716628209888 97
ProbZeroEscapes = .71263164923292297796603018 3 ...
                           52101310912022130249910697446060417 76
ProbAtLeastOneEscape =  .287368350767077022033 ...
                           96981647898689087977869750089302553939582 24
```

This program can be easily extended to produce plots relevant to various assumptions, see Chapter 7 (page 399).

5.5 Chapter Appendix: $g(z,t)$ Existence in 1D

In this section we give an alternative derivation of the Kolmogorov backwards equation appearing on page 106, equation (5), of *Branching Processes*, Athreya and Ney [6] (1972). The Kolmogorov equations are the standard way of constructing the continuous process.

I have not seen our derivation elsewhere. Completing the derivation required an insight from Complex Function Theory, which was provided by Yunping Jiang.

5.5.1 About a

We can imagine that we are keeping track of many "cells." We record the times for each of these cells to "die." What we mean here by die is that the cell either dies or is transformed into k progeny, $k = 0, 2, 3, \ldots$. We define

$$a = \frac{1 \text{ transformation}}{\text{average lifespan}}.$$

5.5.2 Derivation

We will use the generating function from Iwasa, Michor, and Nowak [49] for concreteness. Let the probability generating function for the number of offspring produced in time step h be given by:

$$P(s,h) = [h] + [1 - (1+R)h]s + [Rh]s^2.$$

Note that the term $[(1 - (1 + R)h]$ represents the probability that a cell will not be transformed. Also note the initial condition, $P(s,0) = s$; i.e. the population consists of one cell at time $t = 0$. Collecting terms with h we get:

$$P(s,h) = h[1 - (1+R)s + Rs^2] + s.$$

If we iterate $P(s,h)$ with respect to s twice, we get $P(P(s,h),h) = Z_h(s,2h)$ where $Z_h(s,t)$ is the population size generating function based upon iterates of $P(s,h)$. Continuing in this way we get:

$$\begin{aligned} Z_h(s,(n+1)h) &= \underbrace{P(P\cdots P(P}_{n+1}(s,h),h)\cdots h)h) \\ &= P(\underbrace{P\cdots P(P}_{n}(s,h),h)\cdots h)h) \\ &= P(Z_h(s,nh),h) \\ &= h[(1-(1+R)\ Z_h(s,nh) + R\ Z_h^2(s,nh)] + Z_h(s,nh). \end{aligned} \tag{5.45}$$

Let

$$f(s) = \frac{1}{1+R}[1+Rs^2]$$

be the conditional probability generating function for the number of offspring a transforming (dying) cell leaves behind, either no children; i.e. the cell dies with probability $1/(1+R)$, or the cell splits; i.e. 2 children, with probability $R/(1+R)$. Then

$$(1+R)(f(s)-s) = [1-(1+R)s+Rs^2]$$

Letting $a = 1+R$ we get

$$a(f(s)-s) = [1-(1+R)s+Rs^2].$$

Define:

$$u(s) = a(f(s)-s) = [1-(1+R)s+Rs^2].$$

But (5.45) is simply the Euler Method for numerically solving ODE's. See Chapter 6 (page 375). So as $h \to 0$, (5.45) approaches the solution of

$$\frac{d}{dt}\ Z(s,t) = 1 - (1+R)\ Z(s,t) + R\ Z^2(s,t)$$

i.e.

$$\frac{d}{dt}\ Z(s,t) = u(Z(s,t)).$$

Athreya and Ney in [6] use the notation $F(s,t)$ instead of $Z(s,t)$.

To see that $Z(s,t)$ must be a generating function for each $t > 0$ it suffices to show that $\forall\ t > 0$:

1. $Z(s,t)$ is analytic w.r.t. s in neighborhood of zero.

2. The coefficients of the Taylor expansion of $Z(s,t)$ w.r.t. s are non-negative and sum to 1.

The Euler Line $P_h(s,t)$ with time step $h > 0$ is piecewise linear w.r.t. $t \geq 0$. Moreover, for each $t \geq 0$, $P_h(s,t)$ is a generating function and so analytic if $|s| < 1$. An easy consequence of Picard's Existence and Uniqueness Theorem for first order ODE's:

$$\dot{x} = v(x,t) \in \mathbb{R}^n,\ x(0) = x_0 \in \mathbb{R}^n \tag{5.46}$$

is that the Euler Lines for (5.46) converge uniformly to their unique solution. Picard's Theorem tells us that the solution, $x(t)$ is continuous in t, and the solutions are continuous with respect to the initial conditions x. (This is a slight abuse of notation.)

So, we have proven, as $h \to 0$ the Euler Lines $P_h(s,t)$ converge uniformly (in a neighborhood of zero) to a continuous in s,t function $Z(s,t)$.

The Weierstrass Theorem guarantees if f_n is a sequence of functions analytic in an open set about z, then if they converge uniformly to some limit, say f, then f will also be analytic in an open set about z.

So we must in fact have that $Z(s,t)$ is not only continuous, but in fact analytic in a neighborhood of zero. This in turn implies that $Z(s,t)$ has a Taylor Expansion.

Moreover, The Weierstrass Theorem, or at least its proof as presented in Ahlfors [1] page 177, shows that the Taylor coefficients of the functions f_n in the sequence converge to the Taylor coefficients of the limit f of that sequence.

As $P_h(s,t)$ is a generating function, the Taylor coefficients of $P_h(s,t)$ are all non-negative, and so any limit of their Taylor coefficients will be non-negative.

So we have shown that the Taylor coefficients of $Z(s,t)$ exist and are all non-negative.

Since $P_h(s,t)$ is a generating function $P_h(1,t) = 1$ identically. But then

$$\lim_{h \to 0} P_h(1,t) = 1 = Z(1,t).$$

So $Z(s,t)$ is a generating function.

Chapter 6

Differential Equations

6.1 Some results on Differential Equations in $\mathbb{C}^n$.

Note. In this section, some results from the theory of ODE's in $\mathbb{C}^n$ are derived. We basically follow Coddington & Levinson [23] (Chap. 1, Sec. 1 and Sec. 2) but with a number of customizations to suit our purposes. For an alternative approach, see Arnold [5].

6.1.1 Initial Assumptions and the Differential Equation

Let:

1. $D \subset \mathbb{C}^n$ be compact.

2. $f : D \times [t_0, \infty) \to \mathbb{C}^n$ be bounded [1] by $M > 0$ and Lipschitz [2] with Lipschitz constant L.

[1] Bounded by M. If $(z,t) \in D \times [t_0, \infty)$ then $||f(z,t)|| \leq M$.

[2] Lipschitz. If $(z_1,t_1), (z_2,t_2) \in D \times [t_0, \infty)$ then

$$||f(z_1,t_1) - f(z_2,t_2)|| \leq L||(z_1,t_1) - (z_2,t_2)||.$$

Lipschitz implies continuous.

3. $\exists\ \delta_f > 0$ such that if $0 \leq s \leq \delta_f$ and $(z,t) \in D \times [t_0, \infty)$ then

$$E_s(z,t) = z + f(z,t)s \in D. \tag{6.1}$$

4. Let $\xi : D \to D$ continuously. Define the differential equation

$$(\mathrm{DE}_1) \quad \frac{\partial}{\partial t}\varphi(z,t) = f(\varphi(z,t)) \text{ with initial condition } \varphi(z,t_0) = \xi(z).$$

6.1.2 Partitions and the Euler Lines

Let $p = \{t_0, t_1, \ldots, t_m\}$ be a finite partition of $[t_0, t_m]$ meaning that $t_0 < t_1 < \cdots < t_m$. Define $|p\,| = \max\{|t_k - t_{k-1}| : k = 1, 2, \ldots, m\}$.

Let f, D, t_0, δ_f, and ξ be as in Subsection 6.1.1. Let p be a partition of $[t_0, t_m]$ with $|p\,| < \delta_f$. The Euler Line φ_p with initial condition ξ is defined inductively:

$$\varphi_p(z,t_0) = \xi(z) \tag{6.2}$$

$$\varphi_p(z,t) = \varphi_p(z,t_{k-1}) + f\left(\varphi_p(z,t_{k-1}), t_{k-1}\right)(t - t_{k-1}) \text{ if } \ t \in (t_{k-1}, t_k]. \tag{6.3}$$

By (6.1), $\varphi_p : D \times [t_0, t_m] \to D$.

6.1.3 ϵ approximate solution

Let f, D, t_0, δ_f, and ξ be as in Subsection 6.1.1. Let I be an interval in $\mathbb{R}$ that contains t_0. An ϵ approximate solution, φ of DE_1 on the strip $D \times I$ satisfies:

1. $\varphi : D \times I \to \mathbb{C}^n$ continuously.

2. $\varphi(z,t_0) = \xi(z)$.

3. $$S = \left\{ t \in I : \frac{\partial \varphi}{\partial t}(z,t) \text{ does not exist for some } z \in D \right\} \text{ is finite .}$$

4. $\frac{\partial \varphi}{\partial t}(z,t)$ is continuous on $D \times (I \setminus S)$.

5.
$$\left|\left|\frac{\partial \varphi}{\partial t}(z,t) - f(\varphi(z,t),t)\right|\right| \leq \epsilon, \quad \forall\ (z,t) \in D \times (I \setminus S)\,. \tag{6.4}$$

Lemma 6.1.3.1. *Let $\varphi_p : D \times [t_0, t_m] \to D$ be as defined in Subsection 6.1.2. In particular note that $|p\,| < \delta_f$. Then φ_p satisfies the first four conditions of being an ϵ approximate solution. Moreover, if $t \in (t_{k-1}, t_k)$ then*

$$\frac{\partial \varphi}{\partial t}(z,t) = f\left(\varphi_p(z, t_{k-1}), t_{k-1}\right). \tag{6.5}$$

Proof. Note, the condition of f being Lipschitz or bounded is not required for this proof. f being continuous is sufficient.

1. The continuity of φ_p will be proved by induction on k, $k = 1, 2, \ldots, m$.

Let $i, j \in \{0, 1, 2, \ldots, m\}$ with $i \leq j$. Denote φ_p restricted to $D \times [t_i, t_j]$ by $\varphi_{p,[t_i,t_j]}$. Note, $\varphi_{p,[t_0,t_m]} = \varphi_p$.

If $k = 1$ then $\varphi_{p,[t_0,t_{k-1}]}(z,t) = \xi(z)$ is continuous. So let us assume that $\varphi_{p,[t_0,t_{k-1}]}$ is continuous for some $k \in \{1, 2, \ldots, m\}$. We will show that this implies that $\varphi_{p,[t_0,t_k]}$ is continuous.

$E_s(z,t) = z + f(z,t)s$ is a continuous function of $(z,t,s) \in D \times [t_0, \infty) \times \mathbb{R}$. We define $g_{[t_{k-1},t_k]} : D \times [t_{k-1}, t_k] \to D$:

$$\begin{aligned} g_{[t_{k-1},t_k]}(z,t) &= E_{t-t_{k-1}}(\hat{\varphi}_p(z, t_{k-1}), t_{k-1}) \\ &= \hat{\varphi}_p(z, t_{k-1}) + f(\hat{\varphi}_p(z, t_{k-1}), t_{k-1})(t - t_{k-1}). \end{aligned} \tag{6.6}$$

$g_{[t_{k-1},t_k]}$ is continuous since $E_s(z,t)$ is continuous. Note, the range of $g_{[t_{k-1},t_k]}$ is con-

tained in D since $|p\,| < \delta_f$. Next, we show:

$$g_{[t_{k-1},t_k]}(z,t) = \varphi_{p,[t_{k-1},t_k]}(z,t). \tag{6.7}$$

First, if $t = t_{k-1}$ then $t - t_{k-1} = 0$ in (6.6) and so

$$g_{[t_{k-1},t_k]}(z,t_{k-1}) = \varphi_p(z,t_{k-1}) + 0 = \varphi_{p,[t_{k-1},t_k]}(z,t_{k-1}).$$

If $t \in (t_{k-1},t_k]$ then the definition of $g_{[t_{k-1},t_k]}(z,t)$ exactly matches the definition of $\varphi_p(z,t)$, see (6.3) of Subsection 6.1.2. So (6.7) is true, which implies $\varphi_{p,[t_{k-1},t_k]}$ is continuous, since $g_{[t_{k-1},t_k]}$ is continuous.

So both $\varphi_{p,[t_0,t_{k-1}]}$ and $\varphi_{p,[t_{k-1},t_k]}$ are continuous. They agree on their overlap $D \times \{t_{k-1}\}$ since they are restrictions of the same map, φ_p. It now follows, by Lemma 6.2.1.1, that

$$\left(\varphi_{p,[t_0,t_k]} \cup \varphi_{p,[t_{k-1},t_k]}\right)(z,t) = \varphi_{p,[t_0,t_k]}(z,t)$$

is continuous on $D \times [t_0,t_k]$.

Finally, by induction, $\varphi_{p,[t_0,t_m]} = \varphi_p$ is continuous on all $D \times [t_0,t_m]$.

2. $\varphi_p(z,t_0) = \xi(z)$ by construction. (See (6.2) in Subsection 6.1.2.)

3. We show that if $t \in [t_0,t_m] \setminus p$ then $\varphi_p'(z,t) = f(\varphi_p(z,t_{k-1}),t_{k-1})$, where $k \in \{1,2,\ldots,m\}$ is that unique k such that $t \in (t_{k-1},t_k)$, and where:

$$\varphi_p'(z,t) = \frac{\partial \varphi_p}{\partial t}(z,t) = \lim_{h\to 0} \frac{\varphi_p(z,t+h) - \varphi_p(z,t)}{h}. \tag{6.8}$$

Let $h \in \mathbb{R}$ and let $|h|$ be sufficiently small that h satisfies $t_{k-1} - t < h < t_k - t$, or equivalently that $t + h \in (t_{k-1},t_k)$. We apply the definition of φ_p (see (6.3) in

Subsection 6.1.2) to (z,t) and $(z,t+h)$:

$$\varphi_p(z,t+h) = \varphi_p(z,t_{k-1}) + f\left(\varphi_p(z,t_{k-1}),t_{k-1}\right)(t+h-t_{k-1}) \tag{6.9}$$

$$\varphi_p(z,t) = \varphi_p(z,t_{k-1}) + f\left(\varphi_p(z,t_{k-1}),t_{k-1}\right)(t-t_{k-1}), \tag{6.10}$$

and then subtract (6.10) from (6.9):

$$\varphi_p(z,t+h) - \varphi_p(z,t) = f\left(\varphi_p(z,t_{k-1}),t_{k-1}\right)h. \tag{6.11}$$

Dividing both sides of (6.11) by h and taking the limit as $h \to 0$ yields the desired result:

$$\frac{\partial \varphi_p}{\partial t}(z,t) = f\left(\varphi_p(z,t_{k-1}),t_{k-1}\right), \qquad (z,t) \in D \times (t_{k-1},t_k). \tag{6.12}$$

So φ_p satisfies condition 3 of being an ϵ solution if we take $S = p$.

4. By (6.12), $\varphi_p'(z,t)$ is continuous on each of the relatively open sets $D \times (t_{k-1},t_k)$. As we can write $D \times ([t_0,t_m] \setminus p)$ as the (disjoint) union

$$D \times ([t_0,t_m] \setminus p) = \bigcup_{k=1}^{m} D \times (t_{k-1},t_k)$$

it follows that $\varphi_p'(z,t)$ is continuous on $D \times ([t_0,t_m] \setminus p)$. □

Lemma 6.1.3.2. *If $(z,t),(z,t') \in D \times [t_0,t_m]$ we have*

$$||\varphi_p(z,t) - \varphi_p(z,t')|| < M|t-t'|.$$

Proof. If $t = t'$ the result is trivial. So, by relabeling if necessary, we can assume that $t' < t$. The (inductive) definition φ_p (see (6.3) in Subsection 6.1.2) implies for

$t \in (t_{k-1}, t_k],\ k = 1, 2, \ldots, m$, that:

$$\varphi_p(z,t) = \xi(z) + \left(\sum_{i=1}^{k-1} f(\varphi_p(z,t_{i-1}),t_{i-1})(t_i - t_{i-1}) \right) + f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t_{k-1}) \tag{6.13}$$

(if we use the convention $\sum_{i=a}^{b} = 0$ if $b < a$).

Since $t' < t$ then either $t' = t_0$ or $t' \in (t_{j-1}, t_j]$, for some $j,\ 1 \leq j \leq k$. If $t' = t_0$ then equation (6.13) is also true if we set $j = 1$. Equation (6.13) applied to t and t', implies that:

$$\begin{aligned} \varphi_p(z,t) - \varphi_p(z,t') = & \\ & \left(\sum_{i=j}^{k-1} f(\varphi_p(z,t_{i-1}),t_{i-1})(t_i - t_{i-1}) \right) + \\ & f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t_{k-1}) - \\ & f(\varphi_p(z,t_{j-1}),t_{j-1})(t' - t_{j-1}). \end{aligned} \tag{6.14}$$

If $j = k$ then the sum in the parentheses in (6.14) is zero and (6.14) becomes:

$$\begin{aligned} \varphi_p(z,t) - \varphi_p(z,t') = & \\ & f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t_{k-1}) - f(\varphi_p(z,t_{k-1}),t_{k-1})(t' - t_{k-1}) \\ & = f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t'), \end{aligned}$$

which implies:

$$\begin{aligned} ||\varphi_p(z,t) - \varphi_p(z,t')|| &= ||f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t')|| \\ &\leq M|t - t'|. \end{aligned}$$

If $j < k$, then the sum in the parentheses in (6.14) is not zero and (6.14) becomes:

$$\varphi_p(z,t) - \varphi_p(z,t') = f(\varphi_p(z,t_{j-1}),t_{j-1})(t_j - t_{j-1}) + \left(\sum_{i=j+1}^{k-1} f(\varphi_p(z,t_{i-1}),t_{i-1})(t_i - t_{i-1})\right) + f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t_{k-1}) - f(\varphi_p(z,t_{j-1}),t_{j-1})(t' - t_{j-1}),$$

which can be simplified to:

$$\varphi_p(z,t) - \varphi_p(z,t') = f(\varphi_p(z,t_{j-1}),t_{j-1})(t_j - t') + \left(\sum_{i=j+1}^{k-1} f(\varphi_p(z,t_{i-1}),t_{i-1})(t_i - t_{i-1})\right) + f(\varphi_p(z,t_{k-1}),t_{k-1})(t - t_{k-1}).$$

The triangle inequality then yields:

$$||\varphi_p(z,t) - \varphi_p(z,t')|| \leq ||f(\varphi_p(z,t_{j-1}),t_{j-1})||\,(t_j - t') + \left(\sum_{i=j+1}^{k-1} ||f(\varphi_p(z,t_{i-1}),t_{i-1})||\,(t_i - t_{i-1})\right) + ||f(\varphi_p(z,t_{k-1}),t_{k-1})||(t - t_{k-1}).$$

Substituting in $||f|| < M$ yields:

$$||\varphi_p(z,t) - \varphi_p(z,t')|| \leq M\,(t_j - t') \; + \overbrace{\left(\sum_{i=j+1}^{k-1} M\,(t_i - t_{i-1})\right)}^{M\,(t_{k-1}-t_j)} + M(t - t_{k-1})$$

$$= M(t - t')$$

$$= M|t - t'|.$$

□

Lemma 6.1.3.3. *Let $\epsilon > 0$ be given. There exists a $\delta_\epsilon > 0$ such that if $(z,t), (z',t') \in D \times [t_0, t_m]$ and $|t - t'|, ||z - z'|| < \delta_\epsilon$ then $||f(z,t) - f(z',t')|| < \epsilon$.*

Let $|p\,| < \min\{\delta_f,\, \delta_\epsilon,\, \delta_\epsilon/M\}$. Then φ_p satisfies all five conditions of being an ϵ approximate solution.

Proof. Since f is continuous and $D \times [t_0, t_m]$ is compact, the Heine-Cantor Theorem (see Theorem 6.2.2.1 of Subsection 6.2) guarantees that f is uniformly continuous on $D \times [t_0, t_m]$ and hence the existence of a δ_ϵ such that if $(z,t), (z',t') \in D \times [t_0, t_m]$ and

$$|t - t'|, ||z - z'|| < \delta_\epsilon, \text{ then } ||f(z,t) - f(z',t')|| < \epsilon. \tag{6.15}$$

So the first part of this lemma is proven.

By Lemma 6.1.3.1 the first four conditions of being an ϵ approximate solution are satisfied. Note Lemma 6.1.3.1 relies on $|p\,| < \delta_f$.

We prove that φ_p satisfies the fifth condition of being an ϵ solution (i.e., equation (6.4)) by proving:

$$\left|\left|\varphi_p'(z,t) - f(\varphi(z,t),t)\right|\right| \leq \epsilon, \quad \forall\ (z,t) \in D \times ([t_0, t_m] \setminus p) \tag{6.16}$$

and noting that p is a finite set.

If $(z,t) \in D \times ([t_0, t_m] \setminus p)$ then $t \in (t_{k-1}, t_k)$ for some $k \in \{1, 2, \ldots, m\}$. But then

$$t - t_{k-1} < |p\,| < \min\{\delta_\epsilon / M, \delta_\epsilon\}. \tag{6.17}$$

Lemma 6.1.3.2 with (6.17) implies:

$$||\varphi_p(z,t) - \varphi_p(z,t_{k-1})|| < M|t - t_{k-1}| < \delta_\epsilon. \tag{6.18}$$

Combining (6.17) and (6.18):

$$|t - t_{k-1}|,\ ||\varphi_p(z,t) - \varphi_p(z,t_{k-1})|| < \delta_\epsilon \text{ if } (z,t) \in D \times (t_{k-1}, t_k). \tag{6.19}$$

By Lemma 6.1.3.1, if $(z,t) \in D \times (t_{k-1}, t_k)$, then

$$\varphi_p'(z,t) = f\left(\varphi_p(z,t_{k-1}), t_{k-1}\right)$$

and so

$$\left|\left|\varphi_p'(z,t) - f\left(\varphi_p(z,t), t\right)\right|\right| = ||f\left(\varphi_p(z,t_{k-1}), t_{k-1}\right) - f\left(\varphi_p(z,t), t\right)|| < \epsilon.$$

The inequality in (6.20) follows from (6.19) and (6.15). □

The following inequality (about the rate at which two ϵ approximate solutions that are close at one time spread apart) is useful. It appears as Theorem 2.1, on page 8, of Coddington & Levinson [23]. It is reproduced here (with some slight customizations).

Theorem 6.1.3.4. *Suppose f is Lipschitz with constant k. Let $\varphi_1(z,t)$ and $\varphi_2(z,t)$ be ϵ_1 and ϵ_2 approximate solutions of the DE_1 on the interval $[a,b]$. Let $\tau \in [a,b]$ and suppose that*

$$||\varphi_1(z,\tau) - \varphi_2(z,\tau)|| \leq \delta. \tag{6.20}$$

Let $\epsilon = \epsilon_1 + \epsilon_2$. *If* $t \in [a, b]$ *then*

$$||\varphi_1(z,t) - \varphi_2(z,t)|| \leq \delta\, e^{k|t-\tau|} + \frac{\epsilon}{k}\left(e^{k|t-\tau|} - 1\right). \tag{6.21}$$

Proof. Notation. In this proof, we will represent $\frac{\partial}{\partial t}$ by priming.

We will do the case $\tau \leq t \leq b$.

By the definition of ϵ approximate solution $\varphi_i'(z,t)$ is defined on $D \times [a,b]$ except for, at most, finitely many t in the following sense. There is a finite set $S \subset [a,b]$ with the property that if $\varphi_i'(z,t)$ is not defined at (z,t) then $t \in S$.

Since φ_i is an ϵ_i approximate solution it is the case that

$$||\varphi_i'(z,s) - f(\varphi_i(z,s),s)|| \leq \epsilon_i \quad (i = 1,2). \tag{6.22}$$

Integrating both sides of (6.22) yields for $(i = 1, 2)$:

$$\begin{aligned}
\epsilon_i(t-\tau) &\geq \int_\tau^t ||\varphi_i'(z,t) - f(\varphi_i(z,s),s)||\, ds \\
&\geq \left|\left| \int_\tau^t \varphi_i'(z,t) - f(\varphi_i(z,s),s)\, ds \right|\right| \\
&= \Bigg|\Bigg| \underbrace{\varphi_i(z,t) - \varphi_i(z,\tau) - \int_\tau^t f(\varphi_i(z,s),s)\, ds}_{\alpha_i} \Bigg|\Bigg|.
\end{aligned} \tag{6.23}$$

Since $||\alpha_1|| + ||\alpha_2|| \geq ||\alpha_1 - \alpha_2||$ and since $||a - b|| \geq \Big|\,||a|| - ||b||\,\Big| \geq ||a|| - ||b||$, we

get from (6.23):

$$\begin{aligned}
\overbrace{\epsilon_1(t-\tau)+\epsilon_2(t-\tau)}^{\epsilon(t-\tau)} &\geq \left|\left|\varphi_1(z,t)-\varphi_1(z,\tau)-\int_\tau^t f(\varphi_1(z,s),s)\,ds\right|\right| \\
&+ \left|\left|\varphi_2(z,t)-\varphi_2(z,\tau)-\int_\tau^t f(\varphi_2(z,s),s)\,ds\right|\right| \\
&\geq \Big|\Big| (\varphi_1(z,t)-\varphi_2(z,t))-(\varphi_1(z,\tau)-\varphi_2(z,\tau)) \\
&\quad - \int_\tau^t [f(\varphi_1(z,s),s)-f(\varphi_2(z,s),s)]\,ds\Big|\Big| \\
&\geq \Big|\Big| (\varphi_1(z,t)-\varphi_2(z,t))\Big|\Big| - \Big|\Big| (\varphi_1(z,\tau)-\varphi_2(z,\tau)) \\
&\quad - \int_\tau^t [f(\varphi_1(z,s),s)-f(\varphi_2(z,s),s)]\,ds\Big|\Big|,
\end{aligned}$$

which implies:

$$\begin{aligned}
\epsilon(t-\tau)+\Big|\Big|\overbrace{(\varphi_1(z,\tau)-\varphi_2(z,\tau))}^{a}-\overbrace{\int_\tau^t [f(\varphi_1(z,s),s)-f(\varphi_2(z,s),s)]\,ds}^{b}\Big|\Big| \\
\geq \Big|\Big| (\varphi_1(z,t)-\varphi_2(z,t))\Big|\Big|.
\end{aligned}$$

Then, using $||a||+||b||\geq||a-b||$ and $\int||g||\geq\left|\left|\int g\right|\right|$, we obtain:

$$\begin{aligned}
\epsilon(t-\tau)+\overbrace{\left|\left|\varphi_1(z,\tau)-\varphi_2(z,\tau)\right|\right|}^{r(z,\tau)}+\int_\tau^t ||f(\varphi_1(z,s),s)-f(\varphi_2(z,s),s)||\,ds \\
\geq \underbrace{\left|\left|\varphi_1(z,t)-\varphi_2(z,t)\right|\right|}_{r(z,t)}.
\end{aligned}$$

Letting $r(z,t)=||\varphi_1(z,t)-\varphi_2(z,t)||$ (which is defined on $[\tau,b]$) and using that f is

Lipschitz with constant k, we get:

$$\begin{aligned} ||f(\varphi_1(z,s),s) - f(\varphi_2(z,s),s)|| &\leq k\,||(\varphi_1(z,s),s) - (\varphi_2(z,s),s)|| \\ &= k\,||\varphi_1(z,s) - \varphi_2(z,s)|| \\ &= k \cdot r(z,s) \end{aligned}$$

and

$$\epsilon(t-\tau) + r(z,\tau) + k\overbrace{\int_\tau^t r(z,s)\,ds}^{R(z,t)} \geq r(z,t).$$

We let $R(z,t) = \int_\tau^t r(z,s)\,ds$ (defined on $D \times [\tau, b]$), so that on $D \times (\tau, b)$, $R'(z,t) = r(z,t)$. By assumption $r(z,\tau) = ||\varphi_1(z,\tau) - \varphi_2(z,\tau)|| \leq \delta$. With these substitutions and reversing the order in the inequality (for readability) we get:

$$\overbrace{r(z,t)}^{R'(z,t)} \leq \overbrace{k\int_\tau^t r(z,s)\,ds}^{k\cdot R(z,t)} + \overbrace{r(z,\tau)}^{\leq\delta} + \epsilon(t-\tau) \tag{6.24}$$

which implies:

$$R'(z,t) - k \cdot R(z,t) \leq \delta + \epsilon(t-\tau).$$

Multiplying both sides by $e^{-k(t-\tau)}$ gives:

$$e^{-k(t-\tau)}R'(z,t) - e^{-k(t-\tau)}k \cdot R(z,t) \leq \delta e^{-k(t-\tau)} + \epsilon(t-\tau)e^{-k(t-\tau)}. \tag{6.25}$$

We integrate (6.25) and make good use of integration by parts, $\int u dv = uv - \int v du$.

Integration of the first term in the left hand side of the inequality (6.25) gives

$$\begin{aligned} \int_\tau^t \overbrace{e^{-k(s-\tau)}}^{u}\overbrace{R'(z,s)\,ds}^{dv} &= e^{-k(s-\tau)}R(z,s)\Big|_\tau^t - \int_\tau^t R(z,s)(-k)e^{-k(s-\tau)}\,ds \\ &= e^{-k(s-\tau)}R(z,s)\Big|_\tau^t + k\int_\tau^t R(z,s)e^{-k(s-\tau)}\,ds, \end{aligned}$$

since $du = -ke^{-k(s-\tau)}\,ds$ and $v = R(z,s)$.

So the integral of the left hand side of (6.25) is:

$$\begin{aligned}\int_\tau^t e^{-k(s-\tau)}R'(z,s) - e^{-k(s-\tau)}k\cdot R(z,s)\,ds = \\ e^{-k(s-\tau)}R(z,s)\Big|_\tau^t + k\int_\tau^t R(z,s)e^{-k(s-\tau)}\,ds - \int_\tau^t e^{-k(s-\tau)}k\cdot R(s,t)\,ds \\ = e^{-k(s-\tau)}R(z,s)\Big|_\tau^t = e^{-k(t-\tau)}R(z,t) - \underbrace{R(z,\tau)}_{=0} = e^{-k(t-\tau)}R(z,t).\end{aligned}$$

Integration of the first term in the right hand side of the inequality (6.25) easily gives

$$\begin{aligned}\int_\tau^t \delta e^{-k(s-\tau)}\,ds &= -\frac{\delta}{k}e^{-k(s-\tau)}\Big|_\tau^t \\ &= \left(-\frac{\delta}{k}e^{-k(t-\tau)}\right) - \left(-\frac{\delta}{k}\cdot 1\right) \\ &= \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right).\end{aligned}$$

Integration of the second term in the right hand side of the inequality (6.25) is done via integration by parts:

$$\begin{aligned}\int_\tau^t \overbrace{\epsilon\cdot(s-\tau)}^{u}\overbrace{e^{-k(s-\tau)}\,ds}^{dv} &= \epsilon\,(s-\tau)\cdot\frac{-1}{k}\,e^{-k(s-\tau)}\Big|_\tau^t - \int_\tau^t \frac{-1}{k}\,e^{-k(s-\tau)}\cdot\epsilon\,ds \\ &= \epsilon\,(s-\tau)\cdot\frac{-1}{k}\,e^{-k(s-\tau)}\Big|_\tau^t + \frac{\epsilon}{k}\int_\tau^t e^{-k(s-\tau)}\,ds \\ &= \epsilon\,(s-\tau)\cdot\frac{-1}{k}\,e^{-k(s-\tau)}\Big|_\tau^t - \frac{\epsilon}{k^2}\,e^{-k(s-\tau)}\Big|_\tau^t \\ &= -\frac{\epsilon\,(t-\tau)}{k}\,e^{-k(s-\tau)} - \frac{\epsilon}{k^2}\left(e^{-k(t-\tau)} - 1\right)\end{aligned}$$

since $du = \epsilon\,ds$ and $v = \frac{-1}{k}\,e^{-k(s-\tau)}$.

So the integral of the right hand side of (6.25) is:

$$\int \delta e^{-k(s-\tau)} + \epsilon(s-\tau)e^{-k(s-\tau)}\,ds =$$

$$\begin{aligned}
&= \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right) - \frac{\epsilon\,(t-\tau)}{k}\,e^{-k(s-\tau)} - \frac{\epsilon}{k^2}\left(e^{-k(t-\tau)} - 1\right) \\
&= \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right) - \frac{\epsilon\,k\,(t-\tau)}{k^2}\,e^{-k(s-\tau)} - \frac{\epsilon}{k^2}\left(e^{-k(t-\tau)} - 1\right) \\
&= \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right) - \frac{\epsilon}{k^2}\,k\,(t-\tau)\,e^{-k(s-\tau)} - \frac{\epsilon}{k^2}e^{-k(t-\tau)} + \frac{\epsilon}{k^2} \\
&= \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right) - \frac{\epsilon}{k^2}\,e^{-k(s-\tau)}\,\left(k\,(t-\tau) + 1\right) + \frac{\epsilon}{k^2}.
\end{aligned}$$

Combining our expressions for the left and right hand sides of (6.25) we get the inequality:

$$e^{-k(t-\tau)}R(z,t) = \frac{\delta}{k}\left(1 - e^{-k(t-\tau)}\right) - \frac{\epsilon}{k^2}\,e^{-k(s-\tau)}\,\left(k\,(t-\tau) + 1\right) + \frac{\epsilon}{k^2}.$$

We move the $e^{-k(t-\tau)}$ to the other side:

$$R(z,t) = \frac{\delta}{k}\left(e^{k(t-\tau)} - 1\right) - \frac{\epsilon}{k^2}\,\left(k\,(t-\tau) + 1\right) + \frac{\epsilon}{k^2}\,e^{k(t-\tau)}.$$

We plug the above expression for $R(z,t)$ into (6.24):

$$\begin{aligned}
r(z,t) &\leq \overbrace{k\int_{\tau}^{t} r(z,s)\,ds}^{k\cdot R(z,t)} + \overbrace{r(z,\tau)}^{\leq\delta} + \epsilon(t-\tau) \\
&\leq k\cdot R(z,t) + \delta + \epsilon(t-\tau) \\
&= k\left(\frac{\delta}{k}\left(e^{k(t-\tau)}-1\right) - \frac{\epsilon}{k^2}\left(k\left(t-\tau\right)+1\right) + \frac{\epsilon}{k^2}\,e^{k(t-\tau)}\right) + \delta + \epsilon(t-\tau) \\
&= \delta\left(e^{k(t-\tau)}-1\right) - \frac{\epsilon}{k}\left(k\left(t-\tau\right)+1\right) + \frac{\epsilon}{k}\,e^{k(t-\tau)} + \delta + \epsilon(t-\tau) \\
&= \delta e^{k(t-\tau)} - \frac{\epsilon}{k} + \frac{\epsilon}{k}\,e^{k(t-\tau)} \\
&= \delta e^{k(t-\tau)} + \frac{\epsilon}{k}\left(e^{k(t-\tau)}-1\right).
\end{aligned}$$

Since we defined $r(z,t) = ||\varphi_1(z,t) - \varphi_2(z,t)||$ we are done for the case $\tau \leq t \leq b$.

The case of $a \leq t \leq \tau$ is proven in an identical manner except for minor changes such as in the limits of integration.

□

The following theorem and its proof are based on Theorem 2.2, on page 10, of Coddington & Levinson [23]. Our version is somewhat different, being a global, rather than local result.

Theorem 6.1.3.5. *Let $t_0, t' \in \mathbb{R}$ with $t_0 < t'$.*

For each integer $n \geq 1$ let p_n partition $[t_0, t']$ and satisfy $|p_n| < \min\{\delta_f,\, \delta_{\epsilon_n},\, \delta_{\epsilon_n}/M\}$. Suppose that the $\epsilon_n \to 0$. Then:

1. *The polygonal Euler Iterates φ_{p_n}, converge uniformly on $D \times [t_0, t']$ to a continuous function $\varphi_{[t_0,t']}$.*

2. *$\varphi_{[t_0,t']}(z,t)$ is the unique continuous solution of DE_1 on $D \times [t_0, t']$.*

3. *The DE_1 has a unique continuous solution φ valid on $D \times [t_0, \infty)$. It is an extension of $\varphi_{[t_0,t']}(z,t)$.*

4. *Let φ be the unique solution of DE_1 with I.C. $\varphi(z,0) = z$ that is valid on $D \times [0,\infty)$. Let $t, s \geq 0$ then $\varphi(z, s+t) = \varphi(\varphi(z,s),t)$ so that φ is a semi-flow* [3].

Proof. **1. Uniform Convergence of φ_p.** First we show that the polygonal Euler Iterates $\varphi_{p_n}(z,t)$ converge uniformly on $D \times [t_0, t]$ to a continuous function $\varphi_{[t_0,t']}(z,t)$.

By Lemma 6.1.3.3 for each $n \geq 1$, $\varphi_{p_n}(z,t)$ is an ϵ_n solution. Theorem 6.1.3.4 applied φ_{p_n} and φ_{p_m} guarantees that if for some $(z,\tau) \in D \times [t_0, t']$ we have

$$||\varphi_{p_n}(z,\tau) - \varphi_{p_m}(z,\tau)|| \leq \delta_{n,m} \tag{6.26}$$

for some $\delta_{n,m} \geq 0$. Then we have for all $(z,t) \in D \times [t_0, t']$:

$$||\varphi_{p_n}(z,t) - \varphi_{p_m}(z,t)|| \leq \delta_{n,m}\, e^{k|t-\tau|} + \frac{\epsilon_n + \epsilon_m}{k}\left(e^{k|t-\tau|} - 1\right). \tag{6.27}$$

where $k > 0$ is the Lipschitz constant for f on $D \times [t_0, t']$. ((6.27) is just inequality (6.21) rewritten in terms of φ_{p_n} and φ_{p_m}.) But by the definition of φ_p we have:

$$\varphi_{p_n}(z,t_0) = \varphi_{p_m}(z,t_0) = \xi(z)$$

so letting $\tau = t_0$, (6.26) becomes:

$$||\varphi_{p_n}(z,t_0) - \varphi_{p_m}(z,t_0)|| \leq \delta_{n,m} = 0.$$

If we set $\alpha = t' - t_0$, then for all $t \in [t_0, t']$, we have $e^{k|t-\tau|} = e^{k|t-t_0|} \leq e^{k\alpha}$. With these replacements inequality (6.27) becomes:

$$||\varphi_{p_n}(z,t) - \varphi_{p_m}(z,t)|| \leq \frac{\epsilon_n + \epsilon_m}{k}\left(e^{k\alpha} - 1\right), \quad \forall\ (z,t) \in D \times [t_0, t']. \tag{6.28}$$

[3] Let X be a set. $\theta : X \times [0,\infty) \to X$ is a semi-flow if whenever $t_1, t_2 \geq 0$ and $x \in X$ we have $\theta(x, t_1 + t_2) = \theta(\theta(x,t_1), t_2)$ and $\theta(x,0) = x\ \forall\ x \in X$.

$\mathbb{C}^n$ is complete so (6.28) implies that the sequence φ_{p_n} uniformly converges to a unique function we'll call $\varphi_{[t_0,t']}$. Each φ_{p_n} is continuous by Lemma 6.1.3.1. So by the standard topological result (see Lemma 6.2.2.2) the function $\varphi_{[t_0,t']}$ must be continuous.

So we've proven that φ_{p_n} converges uniformly to a continuous function $\varphi_{[t_0,t']}$.

2a. Solution of DE_1. Next we prove that $\varphi_{[t_0,t']}$ is a solution of DE_1.

For each n the polygonal Euler Iterate $\varphi_{p_n}(z,t)$ is an ϵ_n solution by Lemma 6.1.3.3, moreover in the proof of that lemma we showed that except at possibly $t \in p_n$, that $\frac{\partial}{\partial t}\varphi_{p_n} = \varphi'_{p_n}$ exists, and that

$$||\varphi'_{p_n}(z,t) - f(\varphi_{p_n}(z,t),t)|| < \epsilon_n, \qquad \forall\ (z,t) \in D \times ([t_0,t'] \setminus p_n)\,. \tag{6.29}$$

With this in mind we define:

$$\Delta_{p_n}(z,t) = \begin{cases} \varphi'_{p_n}(z,t) - f(\varphi_{p_n}(z,t),t), & \text{if } (z,t) \in D \times ([t_0,t'] \setminus p_n)\ ; \\ 0, & \text{if } (z,t) \in D \times p_n. \end{cases}$$

The equality

$$\varphi'_{p_n}(z,s) = f(\varphi_{p_n}(z,s),s) + \overbrace{\varphi'_{p_n}(z,s) - f(\varphi_{p_n}(z,s),s)}^{\Delta_{p_n}(z,s)} \tag{6.30}$$

holds except when $s \in p_n$, a finite set (of measure zero w.r.t. ds). Equation (6.30) and the fundamental theorem of calculus yield for $(z,t) \in D \times [t_0,t']$:

$$\varphi_{p_n}(z,t) = \xi(z) + \int_{t_0}^{t} f(\varphi_{p_n}(z,s),s) + \Delta_{p_n}(z,s)\,ds. \tag{6.31}$$

From the first part of this proof we know that the sequence of $\varphi_{p_n}(z,t)$, see (6.31), converges uniformly to a continuous function which we've named $\varphi_{[t_0,t']}$. In this part

of the proof we concentrate on the right hand side of (6.31).

Since φ_{p_n} converges uniformly to $\varphi_{[t_0,t']}$ on $D \times [t_0, t']$ it follows that $(\varphi_{p_n}(z,t),t)$ converges uniformly to $(\varphi_{[t_0,t']}(z,t),t)$ on $D \times [t_0,t']$. Since $D \times [t_0,t']$ is compact, f is uniformly continuous on $D \times [t_0,t']$. It then follows from general topological considerations, see Lemma 6.2.2.3, that $f(\varphi_{p_n}(z,t),t)$ converges uniformly to $f(\varphi_{[t_0,t']}(z,t),t)$ on $D \times [t_0,t']$.

Moreover, (6.29) implies that Δ_{p_n} converges uniformly to the constant function 0 on $D \times [t_0,t'] \setminus p_n$. On $D \times p_n$ we have $\Delta_{p_n} = 0$. So on all of $D \times [t_0,t']$ we have Δ_{p_n} converges uniformly to the constant function 0.

$f(\varphi_{p_n}(z,t),t)$ and $f(\varphi_{[t_0,t']}(z,t),t)$ are continuous on the compact set $D \times [t_0,t']$ and hence are integrable for each $z \in D$ w.r.t. $t \in [t_0,t']$. Δ_{p_n} is continuous on $D \times ([t_0,t'] \setminus p_n)$ and bounded on $D \times [t_0,t']$. So $\Delta_{p_n}(z,t)$ is integral for each $z \in D$ w.r.t. $t \in [t_0,t']$.

So $f(\varphi_{p_n}(z,t),t) + \Delta_{p_n}(z,t)$ is integrable and uniformly converges to $f(\varphi(z,t),t)$ which is also integrable.

The first part of this proof and (6.31) imply the first and second equalities of (6.32) below. The uniform convergence of $f(\varphi_{p_n}(z,t),t) + \Delta_{p_n}(z,t)$ to $f(\varphi_{[t_0,t']}(z,t),t)$ and their integrability implies (by Lemma 6.2.2.4) the last equality in (6.32):

$$\begin{aligned} \varphi_{[t_0,t']}(z,t) &= \lim_{n\to\infty} \varphi_{p_n}(z,t) \\ &= \lim_{n\to\infty} \left(\xi(z) + \int_{t_0}^{t} f(\varphi_{p_n}(z,s),s) + \Delta_{p_n}(z,s)\, ds \right) \\ &= \xi(z) + \int_{t_0}^{t} f(\varphi_{[t_0,t']}(z,s),s)\, ds. \end{aligned} \tag{6.32}$$

But then $\varphi_{[t_0,t']}(z,t)$ satisfies DE_1, by the Fundamental Theorem of Calculus (differentiate the last part of (6.32)).

2b. Uniqueness. Uniqueness follows immediately from Theorem 6.1.3.4: if φ_1 and φ_2 are two solution of DE_1 then they are ϵ solutions with $\epsilon = 0$. I.e., $\epsilon_1 = \epsilon_2 = 0$. Moreover, $\delta = 0$ since $\varphi_1(z, t_0) = \varphi_2(z, t_0)$. But then (6.21) becomes $||\varphi_1(z, t) - \varphi_2(z, t)|| \leq 0$ on $D \times [t_0, t']$, which implies $\varphi_1 = \varphi_2$. Hence uniqueness.

3. Extension to $D \times [t_0, \infty)$. We define $\varphi(z, t)$ on $D \times [t_0, \infty)$ as follows. Let $z \in D$. Then $\varphi(z, t_0) = \xi(z)$. If $t_a > t_0$ then $\varphi(z, t_a) = \varphi_{[t_0,t_a]}(z, t_a)$.

Let $t_b \geq t_0$. By the uniqueness part of this Theorem, for each t_a, $t_0 \leq t_a \leq t_b$ we know that $\varphi_{[t_0,t_b]}$ restricted to $D \times [t_0, t_a]$ is equal to $\varphi_{[t_0,t_a]}$. So φ restricted $D \times [t_0, t_b]$ is equal to $\varphi_{[t_0,t_b]}$. So φ satisfies DE_1 on $D \times [t_0, t_b]$. Since $t_b \in \mathbb{R}$ was arbitrary other than $t_b \geq t_0$ we have shown that φ satisfies DE_1 on all of $D \times [t_0, \infty)$. Finally, if φ_2 also satisfies DE_1 on $D \times [t_0, \infty)$ then, by uniqueness, φ_2 restricted to $D \times [t_0, t_b]$ must equal $\varphi_{[t_0,t_b]}$ which in turn equals φ restricted to $[t_0, t_b]$, by our previous comments. Since $t_b \in \mathbb{R}$ was arbitrary other than $t_b \geq t_0$ we have shown that φ is unique.

4. Semi-flow. Let φ be the unique solution of DE_1 with I.C. $\varphi(z, 0) = z$ that is valid on $D \times [0, \infty)$. Let $t, s \geq 0$. Define $\hat{\varphi}(z, t) = \varphi(\varphi(z, s), t)$ and $\breve{\varphi}(z, t) = \varphi(z, s+t)$. Then

$$\hat{\varphi}'(z, t) = f(\varphi(\varphi(z, s), t), t) = f(\hat{\varphi}(z, t), t)$$
$$\hat{\varphi}(z, 0) = \varphi(\varphi(z, s), 0) = \varphi(z, s)$$

and

$$\breve{\varphi}'(z, t) = f(\varphi(\varphi(z, s), t), t) = f(\breve{\varphi}(z, t), t)$$
$$\breve{\varphi}(z, 0) = \varphi(z, s + 0) = \varphi(z, s).$$

So both $\breve{\varphi}$ and $\breve{\varphi}$ solve ODE_1 with I.C. $z \to \varphi(z,s)$ at $t=0$, and both are defined on $D \times [0,\infty)$. Hence by the uniqueness part of item 3 of this theorem $\breve{\varphi}(z,t) = \breve{\varphi}(z,t)$ and so $\varphi(z,s+t) = \varphi(\varphi(z,s),t)$. Moreover, $\varphi(z,0) = z$, by our construction in this part of the theorem, so φ is a semi-flow

□

6.2 Standard Results from Topology and Analysis

The following standard results from topology and real analysis are useful. My proofs of these results are included for reference and completeness, also see [71, 84, 82].

6.2.1 Standard Topological Results

Lemma 6.2.1.1. *Let $\Gamma \subset X$ a topological space. Suppose $f_a : \Gamma \times [t_a, t_b] \to Y$ and $f_c : \Gamma \times [t_b, t_c] \to Y$ are both continuous and that f_a and f_c agree on $\Gamma \times \{t_b\}$. Define*

$$(f_a \cup f_c)(x,t) = \begin{cases} f_a(x,t), & \text{if } (x,t) \in \Gamma \times [t_a, t_b] \ ; \\ f_c(x,t), & \text{if } (x,t) \in \Gamma \times [t_b, t_c]. \end{cases}$$

Then $(f_a \cup f_c) : \Gamma \times [t_a, t_c] \to Y$ is well defined and continuous.

Proof. See Figure 6.1. By symmetry it suffices to show that $(f_a \cup f_c)$ is continuous at each $(x,t) \in \Gamma \times [t_a, t_b]$.

To show continuity at the point (x,t) it suffices to show that if W is an open set in Y containing $(f_a \cup f_c)(x,t)$, then there is an open set V in $\Gamma \times [t_a, t_c]$ containing (x,t) such that $(f_a \cup f_a)(V) \subset W$.

First, let $(x,t) \in \Gamma \times [t_a, t_b)$ and suppose that $(f_a \cup f_b)(x,t) \in W$ open in Y. Since f_a is continuous on $\Gamma \times [t_a, t_b]$ there exists an open set V in $\Gamma \times [t_a, t_b]$ having the property that $(x,t) \in V$ and $f_a(V) \subset W$. We can take V to be constructed out of

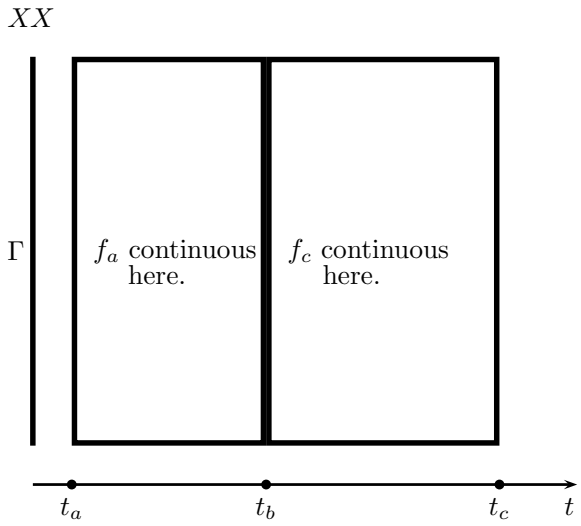


Figure 6.1: $(f_a \cup f_c)(x,t)$ is continuous if f_a is continuous on $\Gamma \times [t_a, t_b]$ and f_c is continuous on $\Gamma \times [t_b, t_c]$ and $f_a = f_c$ on $\Gamma \times \{t_b\}$. See Lemma 6.2.1.1.

basic open sets; that is, to be of the form: $V = (U \cap \Gamma) \times (t_\alpha, t_\beta)$, with $x \in U$ open in X and $t \in (t_\alpha, t_\beta)$. But $V \cap (\Gamma \times [t_b, t_c]) = \emptyset$, so $(f_a \cup f_c)(V) = f_a(V) \subset W$.

Next, let $(x, t_b) \in \Gamma \times \{t_b\}$ and suppose that $(f_a \cup f_b)(x,t) \in W$ open in Y.

Since f_a is continuous on $\Gamma \times [t_a, t_b]$ there exists an open set V_1 in $\Gamma \times [t_a, t_b]$ having the property that $(x, t_b) \in V_1$ and $f_a(V) \subset W$. We can take V to be constructed out of basic open sets; that is, to be of the form: $V = (U_1 \cap \Gamma) \times (t_1, t_b]$, with $x \in U_1$ open in X.

Since f_c is continuous on $\Gamma \times [t_b, t_c]$ there exists an open set V_2 in $\Gamma \times [t_b, t_c]$ having the property that $(x, t_b) \in V_2$ and $f_c(V_2) \subset W$. We can take V_2 to be constructed out of basic open sets; that is, to be of the form: $V_2 = (U_2 \cap \Gamma) \times [t_b, t_2)$, with $x \in U_2$ open in X.

Consider $V = (U_1 \cap U_2 \cap \Gamma) \times (t_1, t_2)$. Clearly $(x, t_b) \in V$ and V is open in $\Gamma \times [t_a, t_c]$. Using the definition of $(f_a \cup f_c)$ and the assumption that f_a agrees with

f_c on $\Gamma \times \{t_b\}$ we have:

$$\begin{aligned}(f_a \cup f_c)(V) &= (f_a \cup f_c)\left((U_1 \cap U_2 \cap \Gamma) \times (t_1, t_2)\right) \\ &= f_a\left((U_1 \cap U_2 \cap \Gamma) \times (t_1, t_b]\right) \bigcup f_c\left((U_1 \cap U_2 \cap \Gamma) \times [t_b, t_2)\right) \\ &\subset W.\end{aligned}$$

□

6.2.2 Standard Results about Uniform Continuity and Convergence

Theorem 6.2.2.1. *Heine – Cantor Theorem. Let X and Y be metric spaces and let X be compact. Let $g : X \to Y$ continuously. Then given $\epsilon > 0$ $\exists$ $\delta_\epsilon > 0$ such that if $x_a, x_b \in X$ with $d(x_a, x_b) < \delta_\epsilon$ then $d(g(x_a), g(x_b)) < \epsilon$.*

Proof. Suppose not. Then there exists a sequence $(x_n, x'_n) \in X \times X$ such that $d(x_n, x'_n) \to 0$ for all n, but $d(g(x_n), g(x'_n)) > \epsilon$.

By compactness we can extract a sub sequence of the x_n, say x_{n_k} which converges to some $x \in X$. By the same reasoning, a subsequence of the x'_{n_k}, say $x'_{n_{k_j}}$ converges to some $x' \in X$.

Since a subsequence of a convergent sequence converges to the same limit as the original sequence, we have $x_{n_{k_j}}$ converges to x.

Since $d(x_n, x'_n) \to 0$ it must be the case that $x = x'$. So both $x'_{n_{k_j}}$ and $x_{n_{k_j}}$ converge to x. By the continuity of g, both $g(x'_{n_{k_j}})$ and $g(x'_{n_{k_j}})$ converge to $g(x)$. Contradicting $d(g(x_n), g(x'_n)) > \epsilon$ for all n. □

Lemma 6.2.2.2. *Let f_n be a sequence of continuous functions from X to Y which converge uniformly to f. Then f is continuous.*

Proof. Let $\epsilon > 0$ be given. Since $f_n \to f$ uniformly there exists an $N > 0$ such that

$n \geq N$ implies that $d(f_n(x), f(x)) < \epsilon/3$ for all $x \in X$. Pick any $m > N$. Since f_m is continuous at x there exists a $\delta_{x,\epsilon/3,f_m} > 0$ such that $y \in X$ and $d(x,y) < \delta_{x,\epsilon/3,f_m}$ implies $d(f_m(x), f_m(y)) < \epsilon/3$.

So let $d(x,y) < \delta_{x,\epsilon/3,f_m}$ then

$$d(f(x), f(y)) \leq \overbrace{d(f(x), f_m(x))}^{<\epsilon/3} + \overbrace{d(f_m(x), f_m(y))}^{<\epsilon/3} + \overbrace{d(f_m(y), f(y))}^{<\epsilon/3} < \epsilon.$$

So f is continuous.

□

Lemma 6.2.2.3. *Let f_n be a sequence of continuous functions from X to Y which converge uniformly. Let g map Y to Z be uniformly continuous. Then $g \circ f_n$ converges uniformly to $g \circ f$.*

Proof. Let $\epsilon > 0$ be given. Since g is uniformly continuous on Y there exists a $\delta_\epsilon > 0$ such that $y_1, y_2 \in Y$ and $d(y_1, y_2) < \delta_\epsilon$ implies $d(g(y_1), g(y_2)) < \epsilon$. Since the sequence of f_n converges uniformly there exists an $N > 0$ such that $m, m' \geq N$ implies that $d(f_m(x), f_{m'}(x)) < \delta_\epsilon$ for all $x \in X$. But then if $m, m' \geq N$ it is the case that $d(g \circ f_m(x), g \circ f_{m'}(x)) < \epsilon$ for all $x \in X$. □

Lemma 6.2.2.4. *Let f_n be a sequence of integrable functions from X to Y which converge uniformly to f an integrable function. Let Y have norm $||\ ||$. Let $\mu(X) =$ the measure of X be finite. Then*

$$\int_X f \, d\mu = \int_X \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu.$$

Proof. Let $\epsilon > 0$ be given. Let $0 < \mu(X) < \infty$ be the measure of X. Since the sequence of f_n converges uniformly there exists an $N > 0$ such that $m \geq N$ implies

that $||f - f_n|| < \frac{\epsilon}{\mu(X)}$ for all $x \in X$. But then if $m \geq N$ it is the case that

$$\begin{aligned}\left|\left|\int_X f - f_n \, d\mu\right|\right| &\leq \int_X ||f - f_n|| \, d\mu \\ &\leq \int_X \frac{\epsilon}{\mu(X)} \, d\mu \\ &= \frac{\epsilon}{\mu(X)} \cdot \mu(X) \\ &= \epsilon.\end{aligned}$$

□

Chapter 7

Appendix: Graphs of Treatment Success Probabilities

7.1 Error Catastrophe and u

We use the Matlab applications developed earlier to produce graphs of "the probability of successful treatment" versus "mutation rate u" for the Iwasa, Michor, Nowak model [48, 49] [1]. See Figures 7.1 (page 409) and 7.2 (page 410).

Some of the graphs are startling for their abrupt transitions and step-like features: If we let the "single digit mutation probability" u go from 0 to 1 the graph of the success probability is step-like, with values of essentially 0 or 1, except in abrupt transition zones having sharp thresholds.

We see at high mutation rates that the treatment is successful with probability 1. This result differs from what one expects from Iwasa, Michor, and Nowak's development of their model [48, 49] as they discount back mutation and assume u is small. They view the process of the development of resistance in terms of mutating to the escape mutant, which will then escape with probability of $1 - 1/R_m$. This is

[1] We will assume that the mutation rate $i \to j$, denoted u_{ij}, is $u^{h_{ij}}(1-u)^{n-h_{ij}}$ where h_{ij} is the Hamming distance between the base 2 representations of i and j.

the correct escape probability for the escape mutant [2] if one does not permit (back) mutation, or approximately true, generally, for very low mutation rates.

The sort of effects we find in our graphs, abrupt transitions, and special effects at high mutation rates, are similar to those discovered by Manfred Eigen in his quasispecies and error catastrophe theory. Error catastrophe theory was originated by Manfred Eigen in 1971 [3]

According to Eigen [28] (2002):

> The term error catastrophe is of a descriptive nature and lacks a clear-cut definition. A catastrophe is usually triggered if certain tolerances are exceeded. For replication, there is indeed such a limiting value of error or mutation rate that must not be surpassed if the wild type [4] is to be kept stable. We call this limit the error threshold. Why is it a sharply defined limit? Why does the efficiency of replication not vary monotonically with the error rate [5]?

Error catastrophe theory is not just of interest to theoreticians, it has become important in the search for antiviral drugs. Eigen [28] (2002) writes:

> The term error catastrophe, originally introduced in the theory of molecular evolution [27], has become fashionable among virologists. In a recent paper in PNAS [24], it was suggested, on the basis of quantitative sequence studies, that ribavirin, a common antiviral drug, by its mutagenic action drives poliovirus into an error catastrophe of replication, thereby turning a productive infection into an abortive one.

[2] See the paragraph preceding Equation (5.8) (page 338) and let $D = 1$.

[3] Error catastrophe theory is an aspect of quasispecies theory [27, 30, 29]. Quasispecies theory, along with the error catastrophe theory, was originated by Eigen in his 1971 paper [27].

[4] "Wild type," in the above quote, is the virus with the genotype which allows it to produce the greatest number of offspring.

[5] The error rate is u.

Developing successful anti-viral drugs by pushing the virus' mutation rate into the high mutation zone to the right of the well runs counter to intuition about resistance. I.e., pathogens mutate to avoid the immune response and drug treatment. There are real risks involved. In particular: pushing a virus, one that mutates slowly enough to be treatable, into the bottom-of-the-well region where it can not be successfully treated.

The graphs produced in this section, and the issues they raise, are further areas for interesting and potentially useful research.

7.2 The mathematics of the graphs' step-like behavior and abrupt transitions

In the Iwasa, Michor, Nowak model [48, 49] the probability that the treatment will be successful [6] (meaning 0 escapes) is:

$$P(0) = \frac{\lambda^0}{0!}\, e^{-\lambda} = e^{-\lambda} = e^{-N \sum_{i=1}^{n} x_i \xi_i} \tag{7.1}$$

where P is the Poisson distribution with mean

$$\lambda = N \sum_{i=0}^{m} x_i \xi_i. \tag{7.2}$$

N is the the number of virus, bacteria, or cancer cells present at the start of treatment; $\{x_i\}_0^m$ is the quasispecies equilibrium distribution from the pre-treatment phase; and ξ_i is the escape probability for a single particle of type i.

Since $N, \{x_i\}_0^m > 0$, Equations (7.1), (7.2) imply that

$$P(0) = 1 \;\Leftrightarrow\; \{\xi_i\}_0^m = \mathbf{0} \;\Leftrightarrow\; \mathbf{q} = \mathbf{1},$$

[6] See Section 5.4 (page 365) for more details.

where we write $\mathbf{q}$ for the vector of extinction probabilities, with $\mathbf{1} - \mathbf{q} = \{\xi_i\}_0^m$.

By Theorem 5.3.1.1 (page 352) [7] if a discrete branching process is positively regular and not singular then

$$\mathbf{q} = \mathbf{1} \;\Leftrightarrow\; \rho \leq 1 \tag{7.3}$$

where ρ is the dominant eigenvalue of $\mathbf{M}$, the matrix of first moments. We showed in Theorem 5.3.1.2 (page 354) that if a continuous branching process is defined in terms of a discrete process,

$$\mathbf{G}(\mathbf{z}, \Delta t) = \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z}, \tag{7.4}$$

then the vector of extinction probabilities in the continuous and discrete processes are identical [8]. We showed that the continuous branching process in the IMN model [48, 49] [9] can be defined as in (7.4). This was shown by Equation (5.18) (page 349), which we reproduce here:

$$\begin{aligned} g_i(\mathbf{z}, \Delta t) &= \Delta t + (1 - (1 + R_i)\Delta t)z_i + z_i R_i \Delta t \sum_{j=0}^{m} u_{i,j} z_j; \qquad i = 0, 1, \ldots, m \\ &= \underbrace{\left(1 - z_i(1 + R_i) + z_i R_i \sum_{j=0}^{m} u_{i,j} z_j\right)}_{F_i(\mathbf{z})} \Delta t + + z_i; \qquad i = 0, 1, \ldots, m \end{aligned}$$

The relevant matrix of first moments is

$$\mathbf{M} = \left\{ \frac{\partial g_i}{\partial z_j}(\mathbf{1}, \Delta t) \right\} = \{m_{ij}\},$$

[7] Theorem 5.3.1.1 (page 352) is a collection of results about discrete processes found in Harris [42].
[8] Actually 'discrete processes', since $\mathbf{G}(\mathbf{z}, \Delta t) = \mathbf{F}(\mathbf{z})\Delta t + \mathbf{z}$ is a different process for each Δt.
[9] We showed this in Section 5.3 (page 345).

which we calculate here: if $i = j$

$$\begin{aligned} m_{ii} = \frac{\partial g_i}{\partial z_i}(\mathbf{1}, \Delta t) &= (1 - (1 + R_i)\Delta t) + 2\Delta t \cdot R_i u_{ii} + \Delta t \cdot R_i \sum_{\substack{j=0 \\ i \neq j}}^{m} u_{ij} \\ &= 1 - (1 + R_i)\Delta t + 2R_i u_{ii}\Delta t + R_i(1 - u_{ii})\Delta t \\ &= 1 - (1 + R_i)\Delta t + \underbrace{R_i u_{ii}\Delta t + R_i u_{ii}\Delta t}_{2R_i u_{ii}\Delta t} + R_i(1 - u_{ii})\Delta t \\ &= 1 - \Delta t - R_i\Delta t + R_i u_{ii}\Delta t + R_i\Delta t \\ &= 1 - \Delta t + R_i u_{ii}\Delta t \end{aligned} \tag{7.5}$$

and if $i \neq j$

$$m_{ij} = \frac{\partial g_i}{\partial z_j}(\mathbf{1}, \Delta t) = R_i u_{ij}\Delta t. \tag{7.6}$$

Define the matrix $\mathbf{A} = \{a_{ij}\}$, where $a_{ij} = R_i u_{ij}$. Then (7.5) and (7.6) imply:

$$m_{ij} = \begin{cases} 1 - \Delta t + a_{ii}\Delta t, & \text{if } i = j; \\ a_{ij}\Delta t, & \text{if } i \neq j. \end{cases} \tag{7.7}$$

Writing (7.7) in matrix form we get

$$\mathbf{M} = (1 - \Delta t)\mathbf{I} + \Delta t \cdot \mathbf{A}. \tag{7.8}$$

Let $\mathbf{R} = R_{ij}$ be the diagonal matrix with $R_{ii} = R_i$. Let $\mathbf{U} = u_{ij}$ be the mutation matrix. Then $\mathbf{A} = \mathbf{RU}$ and (7.8) becomes

$$\mathbf{M} = (1 - \Delta t)\mathbf{I} + \Delta t \cdot \mathbf{RU}. \tag{7.9}$$

From (7.9) it is clear that ρ'_{RU} is an eigenvalue of $\mathbf{RU}$ if and only if

$$1 - \Delta t + \rho'_{RU}\Delta t \tag{7.10}$$

is an eigenvalue of **M**. So the dominant eigenvalue of **M** is 1 if and only if the dominant eigenvalue of **RU** is 1.

Let ρ_{RU} be the dominant eigenvalue of **RU** and let ρ be the dominant eigenvalue of **M**. By (7.10), which is an increasing function with respect to ρ'_{RU}, we have

$$\rho = 1 - \Delta t + \rho_{RU}\Delta t$$

and so

$$\rho \in (0,1) \Leftrightarrow \rho_{RU} \in (0,1) \Leftrightarrow \rho_{RU} < \rho \tag{7.11}$$

$$\rho = 1 \Leftrightarrow \rho_{RU} = 1 \Leftrightarrow \rho_{RU} = \rho$$

$$\rho \in (1,\infty) \Leftrightarrow \rho_{RU} \in (1,\infty) \Leftrightarrow \rho_{RU} > \rho. \tag{7.12}$$

7.2.1 The behavior of ρ_{RU} and ρ when u is near the end points of [0,1]

The following argument shows that it is always the case that $\rho > 1$ for small values of u:

When $u = 0$ the matrix $\mathbf{U} = \mathbf{I}$ and so $\mathbf{RU} = \mathbf{R}$, which is diagonal. This implies that ρ_{RU} will be the maximal (diagonal) entry of **R**, which by assumption is $R_m > 1$. By (7.12) it follows that $\rho > 1$. The result then follows from the continuity of eigenvalues [63, p. 130], or by the Gershgorin circle theorem [10] [68, p. 498].

The following argument shows that if u is near 1 then

$$\rho_{RU} \approx \max_{i=0,\ldots,m} \sqrt{R_i R_{c(i)}} \tag{7.13}$$

[10] Gerschgorin, S. Über die Abgrenzung der Eigenwerte einer Matrix. Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk 7, 749-754, 1931 [39].

where $c(i)$ is the complement of i with respect to i's representation as binary number of length n [11]:

Basically, $u = 1$ means a type replicates its complement with perfect fidelity [12]. This allows us to quickly write **RU** and to solve the eigenvalue – eigenvector equation. For concreteness, we do this for the case when $n = 2$ and the complementary types 00, 11:

$$\begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & R_{00} \\ 0 & 0 & R_{01} & 0 \\ 0 & R_{10} & 0 & 0 \\ R_{11} & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{pmatrix} \sqrt{\frac{R_{00}}{R_{11}}} \\ 0 \\ 0 \\ \pm 1 \end{pmatrix} = \pm\sqrt{R_{00}R_{11}} \begin{pmatrix} \sqrt{\frac{R_{00}}{R_{11}}} \\ 0 \\ 0 \\ \pm 1 \end{pmatrix} .$$

In general, each pair of complementary types, i and $c(i)$, will yield a pair of linearly independent eigenvectors, e.g.

$$\left(0, \ldots, 0, \underbrace{\sqrt{\frac{R_i}{R_{c(i)}}}}_{i^{th} \text{ position}}, 0, \ldots, 0, \underbrace{\pm 1}_{c(i) \text{ position}}, 0, \ldots, 0 \right)^T$$

having eigenvalues

$$\pm\sqrt{R_i R_{c(i)}} .$$

Moreover, the eigenvectors from one complementary pair will be linearly independent of the eigenvectors of the other complementary pairs. Since there are at most 2^n linearly independent eigenvectors for **RU** we've accounted for all the possibilities. The desired result then follows from the continuity of eigenvalues [63, p. 130].

[11] For example, if $n = 4$ then $c(2) = c(0010) = 1101 = 13$.

[12] For example: if $n = 4$ then a type 0010 will always replicate a type 1101.

We are assuming for $i = 0, 1, \ldots, m-1$ that $0 < R_i < 1 < R_m$. So for

$$i \in 1, \ldots, m-1 \quad \text{we have} \quad 0 < \sqrt{R_i R_{c(i)}} < 1.$$

So if $u \in [0, 1]$ is sufficiently close to 1 and

$$R_m R_0 > 1 \text{ (resp. } < 1) \quad \text{we have} \quad \rho_{RU}, \rho > 1 \text{ (resp. } < 1), \tag{7.14}$$

which follows from (7.13) and (7.11), (7.12).

Note: whether or not $\rho \leq 1$ for some value of $u \in [0, 1]$ depends upon R.

7.2.2 The explanation of the graphs' step-like features and abrupt transitions

We are now ready to explain the step like features and abrupt transitions seen in Figures 7.1 (page 409), 7.2 (page 410), and 7.3 (page 411).

If $\rho_{RU} \leq 1$ then, by (7.11), $\rho \leq 1$ and so (7.3) (page 402) implies that the vector of extinction probabilities $\mathbf{q} = \mathbf{1}$ and so all the escape probabilities will be 0. So the probability of successful treatment will be 1.

On the other hand, if $\rho_{RU} > 1$ then, by (7.12), $\rho > 1$ and so (7.3) (page 402) implies that the vector of extinction probabilities, $\mathbf{q}$, is not $\mathbf{1}$. So the escape probabilities, $\xi_i = 1 - q_i$, will not all be 0. But then the escape probability for a cell randomly selected at the start of treatment

$$\sum_{i=0}^{m} x_i \xi_i > 0$$

since all the $x_i > 0$ [13]. See Equations (7.1), (7.2). Since

$$P(\text{successful treatment}) = e^{-N \sum x_i \xi_i} \tag{7.15}$$

it follows that if

$$N > \frac{-\ln(0.01)}{\sum_{i=0}^{m} x_i \xi_i} \tag{7.16}$$

then

$$P(\text{successful treatment}) < 0.01.$$

See [14] , [15] . Unless of course $\rho_{RU} \leq 1$, in which case

$$P(\text{successful treatment}) = 1$$

as discussed above.

So the graph of P(successful treatment) versus u will have a step-like shape and abrupt transitions if N satisfies (7.16) for each $u \in [0,1]$ and $\mathbf{R}$ is such that $\exists\, u \in [0,1]$ implying $\rho_{RU} \leq 1$. Determining whether N satisfies this condition is non-trivial, the same goes for $\mathbf{R}$.

See Figures 7.3 (page 411) and 7.4 (page 412).

7.2.3 Further research

One direction for further research would be to use the Gershgorin circle theorem to formulate sufficient conditions on $\mathbf{R}$ to insure $\rho_{RU} \leq 1$ for some value of u. Additionally, it seems that ρ_{RU} as a function of $u \in [0,1]$ is concave up and catenary shaped.

[13] All the $x_i > 0$ since $x = (x_0, \ldots, x_m)$ is the quasispecies equilibrium eigenvector of W from the pre-treatment part of the model. The matrix W is assumed to be > 0 and so the vector $x > 0$ by Birkhoff's Projective Contraction Theorem [12], which we discuss in Part I of this dissertation.

[14] To obtain (7.16) we set P(successful treatment) = .01 in Equation (7.15). The value of 0.01 was chosen for concreteness.

[15] By (7.14), if $R_0 R_m > 1$ and N satisfies (7.16) then for u near 1 the P(successful treatment) $< .01$. See u near 1 in Figure 7.3 (page 411).

See Figure 7.3 (page 411). However, proving this seems difficult. One approach towards proving this might involve approximating ρ_{RU} by using Birkhoff's Projective Contraction Theorem [12]; i.e. one has

$$\lim_{k\to\infty} \frac{\left\|(\mathbf{RU})^{k+1}v\right\|_1}{\left\|(\mathbf{RU})^{k}v\right\|_1} = \rho$$

for any $v > 0$.

A second direction for research lies in developing the theory discussed in this dissertation to accurately model or interpret data from evolutionary biology. See, for example, [65] (2010) on bacterial evolution and mutation rates.

A third and important direction would be in the area of disease modeling and treatment, especially as related to diseases which are sometimes resistant to treatment due to mutation. Some current research includes: [81] (2008) regarding AIDS, HIV and error catastrophe; [31] (2010) on resistant malaria strains in Africa; and [78, 35] (both 2010) on clinical and theoretical issues relating to resistance and mutation in cancer.

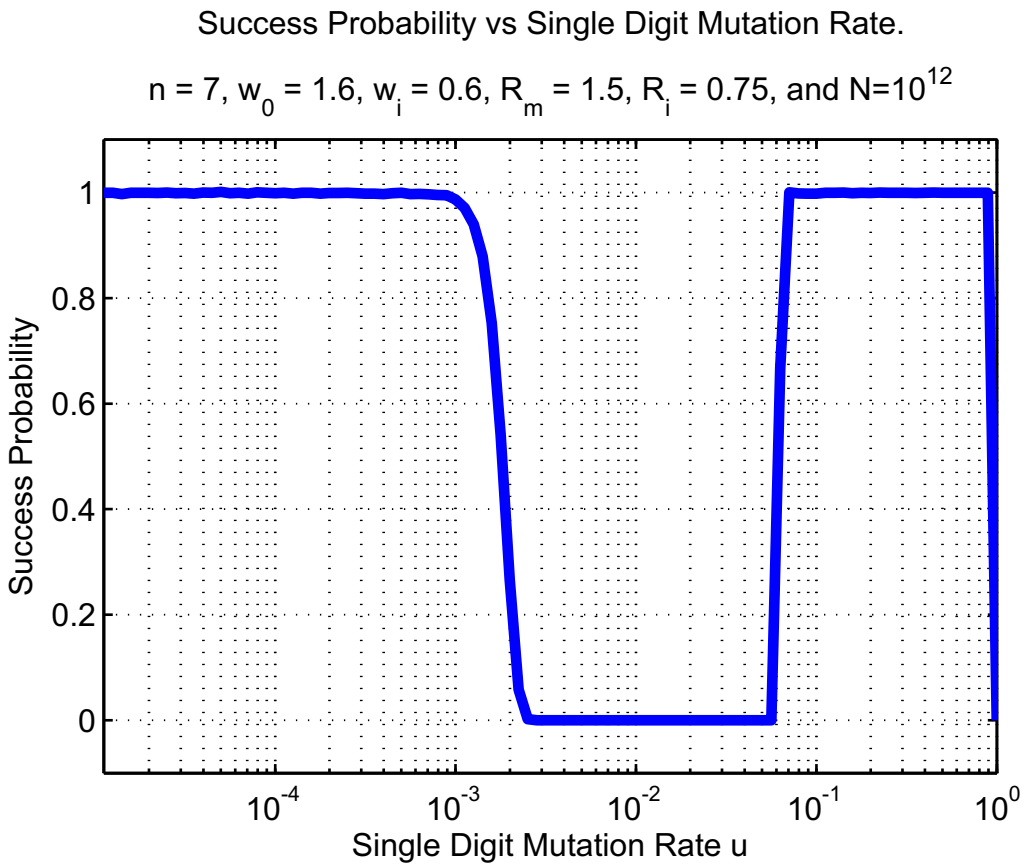


Figure 7.1: Non-uniform pre-treatment quasispecies equilibrium distribution. The graph shows "Probability of treatment success" on the y-axis, versus the "single digit mutation probability u" on a log x-axis. At low mutation rates during the pre-treatment phase the virus doesn't mutate fast enough to produce sufficient escape mutants. So when treatment is applied it is successful with probability 1. At the bottom of the well, the mutation rate is high enough that escape mutants are sufficiently produced and the treatment fails (i.e., is successful with probability 0). If the mutation rate is high enough (to the right of well), many escape mutants are produced. But their offspring are often non-escape types and are killed by the treatment. The escape mutants themselves eventually succumb to stochastic effects, as each has a small chance of dying with each clock tick. The result is in the presence of high mutation rates the treatment is successful with probability 1. When the single digit mutation rate $\approx 1 = 10^0$ the complement type is being replicated with high fidelity. In this example, when $u = 0.99$ the quasispecies equilibrium distribution is 31% wild type and 51% escape mutant: the wild type replicates escape mutants faster than the escape mutant replicates wild types, so the escape mutant type is somewhat absorbing, hence it actually dominates the pre-treatment quasispecies equilibrium distribution. See QuasiSpeciesAug2010.m in 4.3.2 (page 323). Also, notice the rapid transitions, with clear thresholds, indicated by the well's steep walls. These effects seem consistent with Eigen's "error catastrophe" theory [27]. The parameters are as shown in the title. Note: $W_0 = 1.6,\ W_i = 0.6\ \forall\ i = 1, \ldots, m$ and $R_i = 0.75\ \forall\ i = 0, \ldots, m-1$ and $R_m = 1.5$.

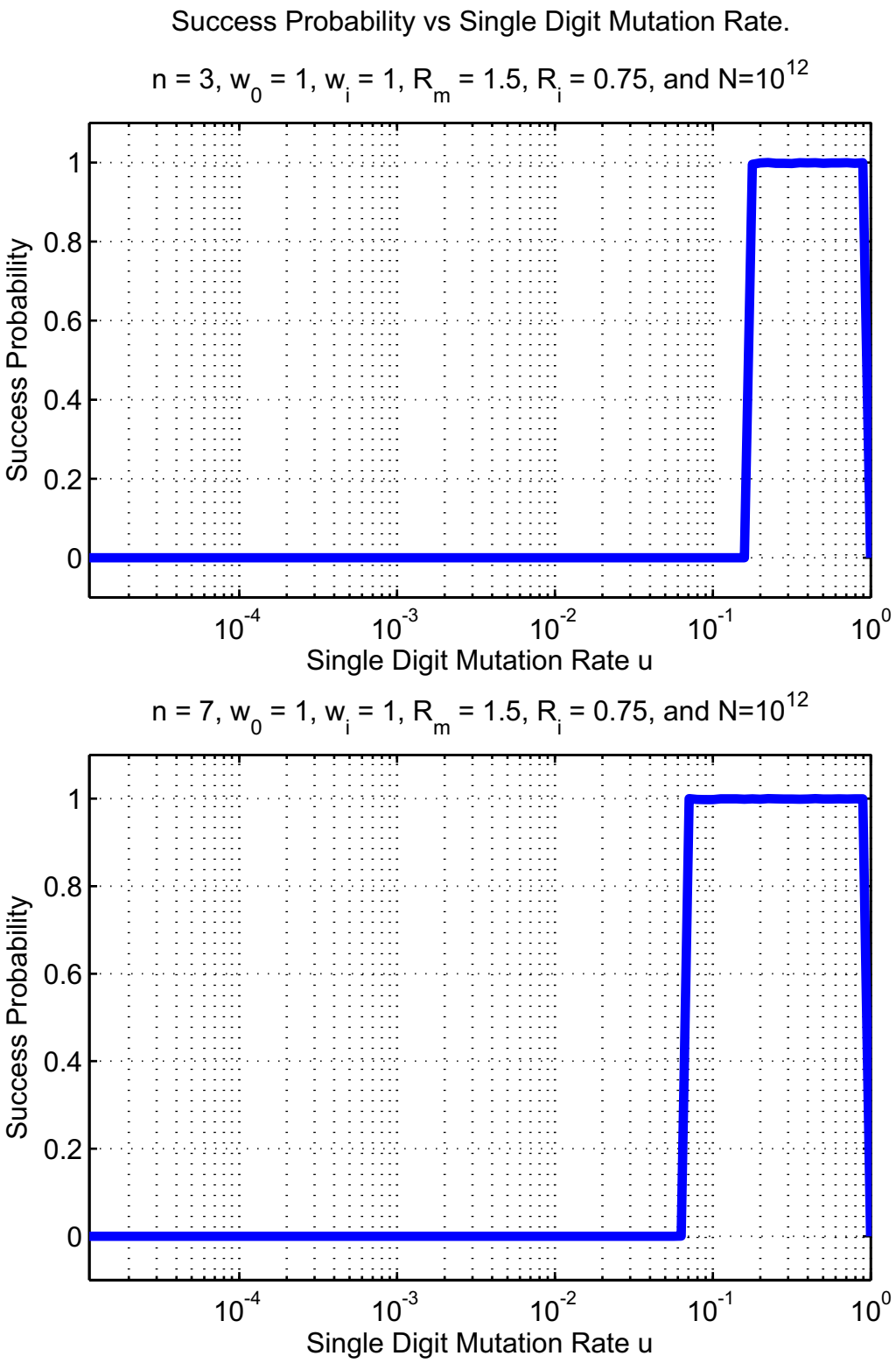


Figure 7.2: The initial distribution of mutant types is uniform. So at the start of treatment there are sufficient escape mutants to overcome treatment unless the mutation rate is sufficiently high. Number of mutation loci: top $n = 3$, bottom $n = 7$. All other parameters the same.

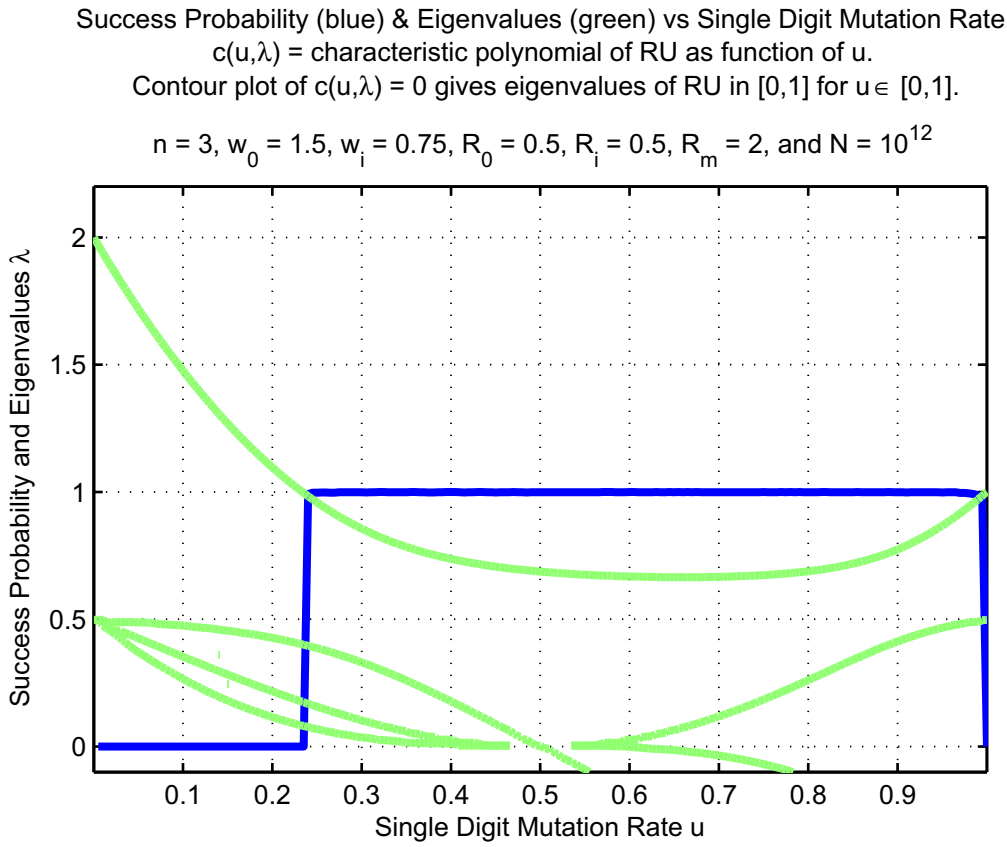


Figure 7.3: Contour plot (green lines) of characteristic equation $c(u, \lambda) = 0$ for **RU** superimposed upon plot of "success probability" versus "single digit mutation rate u" (blue line).

The Matlab code for this graph is contained in the m-file SuccessProbabilityANDeigenvaluePlot.m, which can be found at the end of Section 7.2.4 (page 413).

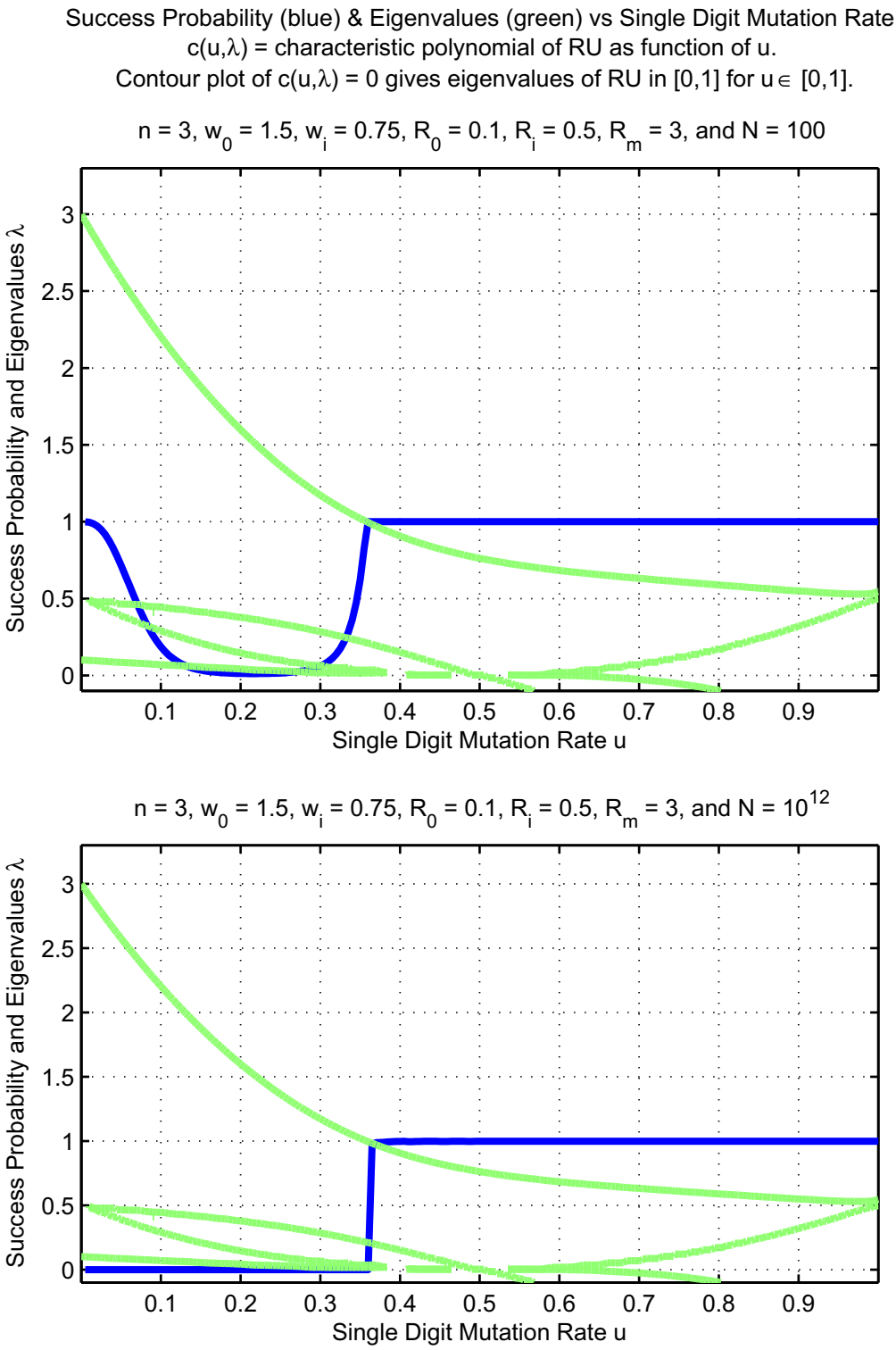


Figure 7.4: Contour plot (green lines) of characteristic equation $c(u, \lambda) = 0$ for **RU** superimposed upon plot of "success probability" versus "single digit mutation rate u" (blue line). All parameters are the same in both graphs except that $N = 100$ for the top one and $N = 10^{12}$ for bottom one (which has a step-like graph). Note $\sqrt{R_0 R_m} = \sqrt{(0.1)(3)} = 0.55$.

7.2.4 Matlab

The following m-file was used to create Figures 7.1 (page 409) and 7.2 (page 410).

The functions, GenFunctionG and uMutationMatrix, found in the code below, are given in Section 5.3.4 (page 360).

```
% PlotTreatmentSuccessProbVSmutationRate.m     m-file name
% plots treatment success vs u  (x axis is log plotted)
%
% ----- user input  ----------------------------------
plotpoints = 40;        % number of points to plot
exponentRangeMin = -5;  % u starts at 10^(-5)
exponentRangeMax = -.1; % u finishes at 10^(-.1)
n = 7;                  %choose n = number of mutations to escape
N = 10^12;
% ----- user pre-treatment parameters -------------------
               %choose 2^n = m+1 pretreatment reproductive ratios
wVector = ones(1,2^n,'double');        % default Wii = 1.0
wVector(1,1) = 1.0;  % wVector(1,1) = wild type
                     % define Woo > Wii [w0 w1 w2 w3 w4 . . . ]
                     % type 0 = matlab 1, type i = matlab i+1
                     % type m = 2^n = 1 is matlab 2^n
% --------- user input post treatment -----------------------
its = 3000;          % number of times to iterate generating function
R = .75*ones(2^n,1); % reproductive fitnesses  vector, default is all 1's
R(2^n,1) = 1.5;      % reproductive fitnesses of escape mutant R_m > 1
% --------- end user input ----------------------------------------
digits(64)     % accurate display of answers, used with vpa
SuccessProbVectorY = zeros(1,plotpoints);
uVector = zeros(1,plotpoints);
u = 0;
for i = 1:plotpoints
    u = 10^(exponentRangeMin + i*((exponentRangeMax- exponentRangeMin)/plotpoints))
```

```
    uVector(1,i) = u;
    sucessProb = TreatmentSuccessProbabilityFunction(n,u,N,wVector,its,R)
    SuccessProbVectorY(1,i) = sucessProb;
end
figure(1)
semilogx(uVector, SuccessProbVectorY, 'LineWidth',4)
t=({['Success Probability vs Single Digit Mutation Rate.'];
    ['n = ', num2str(n), ', w_{0} = ', num2str( wVector(1,1)), ', w_{i} = ',
      num2str( wVector(1,2)), ', R_{m} = ',
      num2str( R(2^n,1)), ', R_{i} = ',
      num2str( R(2,1)), ', and N=10^{12}' ]});
title(t)
xlabel('Single Digit Mutation Rate u')
ylabel('Success Probability')
axis([0 1 -.1 1.1])
grid on
```

The following m-file is a function used in the above code.

```
function out = TreatmentSuccessProbabilityFunction(n,u,N,wVector,its,R)
% based on TreatmentSuccessProbabilityAug2010.m  <-- see for definitions
% calculates the success probability
% by combining pre and post treatment models
% outputs Probability of success
%
digits(64)     % accurate display of answers, used with vpa
U = uMutationMatrix(u,n);     % create mutation matrix u_ij
% ------------------------------
W = zeros(2^n, 2^n);   % create W diagonal vector
for i=1:2^n
    W(i,i) = wVector(1,i);
end
QSindex = 0;   % initialize QSindex, which is the index number of qs
[EigVects EigValues]= eig(U*W); % Matlab finds Eigenvectors, values
```

```
for j =1:2^n                    % pick out which one is Quasispecies
        if abs( sum(  sign(EigVects(:,j)))) == 2^n
            QSindex = j;
        end
end
QS = vpa(EigVects(:,QSindex)/sum(EigVects(:,QSindex))); %normalize QS column vec
% -------------------------------
dt = .9/(1 + max(R));  % time step 'Delta t' < max 1/(1 + maxR)
Z = zeros(2^n,1);      % initiate column vector of extinction ...
                                    probabilities defaults = 0
Z(2^n,1) = 1/R(2^n,1); %  extinction probability of escape mutant ...
                                if no mutation is 1/Rm
                       %  this slightly speeds up convergence
for i = 1:its                % iteration process for extinction probabilities
    Z = GenFunctionG(dt,Z,n,R,U);
end
EscapeProbs = zeros(2^n,1);  % column vector to hold escape ...
                                     \'{e} probabilities defaults = 0
for i = 1: 2^n        % filling vector to hold escape probabilities
   Zescape(i,1) = 1 - Z(i,1);
end
EscapeProbs = vpa(Zescape(:,1));
% -------------------------------
p = vpa(dot(QS,EscapeProbs));   % probability a randomly chosen cell will escape
ExpectedEscapes = vpa(N*p);     % expected number of escapes for pop of size N
ProbZeroEscapes = vpa(exp(-N*p)); % treatment is successful
%ProbAtLeastOneEscape = vpa(1 - ProbZeroEscapes) % treatment is a failure
out = ProbZeroEscapes;
```

The following m-file was used to create Figures 7.3 (page 411) and 7.4 (page 412).

The function m-file TreatmentSuccessProbabilityFunction.m, whose code is given immediately above, is used in the following program.

```
% SuccessProbabilityANDeigenvaluePlot.m     m-file name
```

```
% Plots probability of treatment success vs single digit mutation rate u
% and overlays eigenvalue contour plot
%-----------------------------------------
clear
hold off
% ----- user input  ---------------------------------
plotpoints = 200;     % number of points to plot
umin =    10^(-7);    % minimum u
umax =  .9999999999;  % maximum u
n = 3;   % n = number of mutations to escape
Nexp = 12;  % usually 12; when n = 2, let N = 100;
N = 10^Nexp;  % N = population size start of treatment
% ------warning: N needs to be manually adjusted in title below!
% ----- user pre-treatment parameters ----------------------
% ------ choose 2^n = m+1 pretreatment reproductive ratios -
wVector = .75*ones(1,2^n,'double');   % default Wii = 1.0
wVector(1,1) = 1.5;  % wVector(1,1) = wild type
                     % define Woo > Wii [w0 w1 w2 w3 w4 . . . ]
                     % type 0 = matlab 1, type i = matlab i+1
                     % type m = 2^n = 1 is matlab 2^n
% --------- user input post treatment -----------------------
its = 9000;         % number of times to iterate generating function
R = .5*ones(2^n,1); % reproductive fitnesses  vector, default is all 1's
Rm = 2;  % escape mutant reproductive ratio
R(2^n,1) = Rm;      % reproductive fitnesses of escape mutant R_m > 1
% R(1,1) = .1;      % Ro = 0.1  (wild type's)
% ------ defaults
LamMax = 1.1* Rm;  % maximum eigenvalue, assuming Rm dominates
LamMin = -0.1;     % minimum eigenvalue
Lam = LamMin: (LamMax - LamMin)/plotpoints: LamMax; % y direction .01
exponentRangeMin = -5;  % u starts at 10^(-5)
exponentRangeMax = -.00001;% u finishes at 10^(-.1)
% --------- end user input ---------------------------------------
```

```
% -------- program code -------------------------
U = umin: (umax - umin)/plotpoints : umax;  %the u's to plot
Rmatrix = zeros(2^n,2^n);   % Constructing the R matrix
for i = 1:2^n
    Rmatrix(i,i) = R(i);
end
digits(64)     % accurate display of answers, used with vpa
SuccessProbVectorY = zeros(1,plotpoints); % vector to hold success probs
uVector = zeros(1,plotpoints); % vector to hold u values
for i = 1:plotpoints  % success probabilities calculated here
    u = umin + (i/plotpoints)*(umax - umin)
    uVector(1,i) = u;
    sucessProb = TreatmentSuccessProbabilityFunction(n,u,N,wVector,its,R)
    SuccessProbVectorY(1,i) = sucessProb;
end
figure(1)  %first we plot success probabilities
plot(uVector, SuccessProbVectorY, 'LineWidth',4)
axis([umin umax LamMin LamMax])
grid on
hold on    % hold image since we have more plotting to do
sizeU = size(U);  % accessing size of the matrix U
lenU =sizeU(1,2);
sizeLam = size(Lam); % accessing size of Lam
lenLam =sizeLam(1,2);
UU = ones(lenLam, lenU); % cols of UU are x coords from U
for i=1:lenLam    %% build the UU matrix
    for j = 1:lenU
        UU(i,j) = U(j);
    end
end
LamLam = ones(lenLam, lenU); % rows are y coords from Lam
for i=1:lenLam  % build the LamLam matrix
    for j = 1:lenU
```

```
        LamLam(i,j) = Lam(i);
    end
end
ZZ = ones(lenLam, lenU); % rows are y coords from Lam
for i=1:lenLam  %build the ZZ matrix
    for j = 1:lenU
        ZZ(i,j) = polyval( poly(Rmatrix* uMutationMatrix(UU(i,j),n) ),...
            LamLam(i,j));
    end
end
v = [0 0];  % draw single contour line for c(u,lambda) = 0
[C,h] = contour(UU,LamLam,ZZ,v);  % produce contour map
set(h,'LineWidth',3)
grid on
t=({['Success Probability vs Single Digit Mutation Rate.'];...
  ['c(u,\lambda) = characteristic polynomial of RU as function of u.'...
   ' Contour plot of c(u,\lambda) = 0 gives eigenvalues of RU'...
   ' in [0,1] for u \in [0,1]'];...
  ['n = ', num2str(n), ', w_{0} = ',...
   num2str( wVector(1,1)),...
   ', w_{i} = ', num2str( wVector(1,2)),...
   ', R_{m} = ', num2str( R(2^n,1)),...
   ', R_{i} = ', num2str( R(2,1)),...
   ', and N=10^{12}' ]}); % N needs to be manually entered here!
title(t)
xlabel('Single Digit Mutation Rate u')
ylabel('Success Probability and eigenvalues \lambda')
hold off
```

7.3 Image Processing Application

We illustrate Birkhoff's Projective Contraction Theorem [12] by having Matlab apply the matrix W to each pixel appearing in an image [16], see Figure 7.5 (page 419).

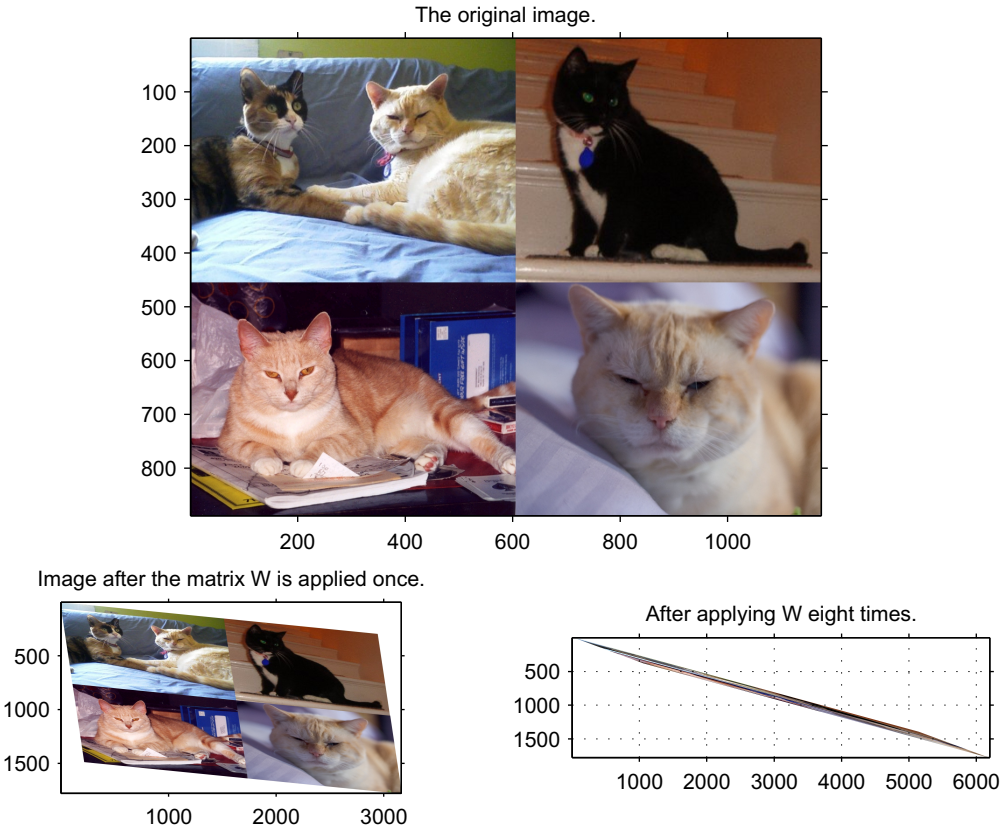


Figure 7.5: The coordinate system used in image processing typically addresses pixels with the positive y direction pointing down. The matrix W is defined in (7.17) (page 419). The positive eigenvector of W, with L^1 norm 1 is $(0.7830951895,\ 0.2169048105)$ yielding a ratio y/x of $0.2169048105/0.7830951895 = 0.2770$. If W is applied eight times, then the ratio y/x, for the tip of the image is $1777/6200 = 0.2866$.

[16] We saw the matrix

$$W = \begin{pmatrix} 3 & .3 \\ .3 & 2 \end{pmatrix} \tag{7.17}$$

applied in Figures 3.1 through 3.7 on pages 309 to 315. Cat models in clockwise order, starting with Quadrant I: Samantha, Precious, Ginger, and Minnie (on left).

Bibliography

[1] Lars V. Ahlfors, *Complex analysis, third edition*, McGraw-Hill, Inc., New York, 1979.

[2] Charalambos D. Aliprantis and Kim C. Border, *Infinite dimensional analysis*, third ed., Springer, Berlin, 2006, A hitchhiker's guide. MR MR2378491 (2008m:46001)

[3] Charalambos D. Aliprantis and Rabee Tourky, *Cones and duality*, Graduate Studies in Mathematics, vol. 84, American Mathematical Society, Providence, RI, 2007. MR MR2317344 (2008k:46012)

[4] Howard Anton, *Calculus*, John Wiley and Sons, Inc., New York, 1980.

[5] V. I. Arnold, *Ordinary Differential Equations*, The M.I.T. Press, Cambridge, Mass.-London, 1973, Translated from the Russian and edited by Richard A. Silverman. MR MR0361233 (50 #13679)

[6] Krishna B. Athreya and Peter E. Ney, *Branching processes*, Springer-Verlag, New York, 1972, Die Grundlehren der mathematischen Wissenschaften, Band 196. MR MR0373040 (51 #9242)

[7] Krishna Balasundaram Athreya, *Some results on multitype continuous time markov branching processes*, The Annals of Mathematical Statistics **39** (April 1968), no. 2, 347–357.

[8] David Barnette, *An upper bound for the diameter of a polytope*, Discrete Mathematics **10** (1974), no. 1, 9 – 13.

[9] A. F. Beardon, *The Klein, Hilbert and Poincaré metrics of a domain*, J. Comput. Appl. Math. **105** (1999), no. 1-2, 155–162, Continued fractions and geometric function theory (CONFUN) (Trondheim, 1997). MR MR1690583 (2000e:51030)

[10] Niko Beerenwinkel, Nicholas Eriksson, and Bernd Sturmfels, *Evolution on distributive lattices*, J. Theoret. Biol. **242** (2006), no. 2, 409–420. MR MR2272562 (2007h:92051)

[11] I.J. Bienaymé, *De la loi de multiplication et de la duree des families*, Soc. Philomat. Paris Extraits, Ser. **5** (1845), 37 – 39.

[12] Garrett Birkhoff, *Extensions of Jentzsch's Theorem*, Transactions of the American Mathematical Society **85** (1957), no. 1, 219–227.

[13] ———, *Uniformly semi-primitive multiplicative processes*, Trans. Amer. Math. Soc. **104** (1962), 37–51. MR MR0146100 (26 #3626)

[14] Garrett Birkhoff and Leon Kotin, *Essentially positive systems of linear differential equations*, Bull. Amer. Math. Soc. **71** (1965), 771–772. MR MR0179414 (31 #3662)

[15] Oliver Bletz-Siebert and Thomas Foertsch, *The Euclidean rank of Hilbert geometries*, Pacific J. Math. **231** (2007), no. 2, 257–278. MR MR2346496 (2008j:53073)

[16] Fleming, H.P., Breidt, F., Romik, T.L., *A rapid method for the determination of bacterial growth kinetics*, Journal of Rapid Methods and Automation in Microbiology **3** (1994), 59–68.

[17] Douglas S. Bridges, *Foundations of real and abstract analysis*, Springer, New York, 1997.

[18] Michael G. Bulmer, *Francis Galton: Pioneer of Heredity and Biometry*, Johns Hopkins Univ. Press, Baltimore, 2003.

[19] P. J. Bushell, *Hilbert's metric and positive contraction mappings in a Banach space*, Arch. Rational Mech. Anal. **52** (1973), 330–338. MR MR0336473 (49 #1247)

[20] J. C. Butcher, *Numerical methods for ordinary differential equations*, John Wiley & Sons Ltd., Chichester, 2003. MR 1993957 (2004e:65069)

[21] Frederick W. Byron, Jr. and Robert W. Fuller, *Mathematics of classical and quantum physics*, Dover Publications Inc., New York, 1992, Corrected reprint of the 1969 (Vol. 1) and 1970 (Vol. 2) originals. MR 1189553 (93h:00001)

[22] Arthur Cayley, *Sixth memoir upon quantics*, Phil. Trans. **149** (1859), 61–91.

[23] Earl A. Coddington and Norman Levinson, *Theory of ordinary differential equations*, McGraw-Hill Book Company, Inc., New York-Toronto-London, 1955. MR MR0069338 (16,1022b)

[24] Shane Crotty, Craig E. Cameron, and Raul Andino, *RNA virus error catastrophe: Direct molecular test by using ribavirin*, Proceedings of the National Academy of Sciences (PNAS) **98** (June 5, 2001), no. 12, 6895 – 6900.

[25] Charles Darwin, *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*, J. Murray, London, 1860.

[26] Max Dehn, *Mathematics, 200 B.C.–600 A.D.*, The American Mathematical Monthly **51** (Mar., 1944), no. 2, 149–157.

[27] Manfred Eigen, *Self organization of matter and the evolution of biological macromolecules*, Die Naturwissenschaften **58** (1971).

[28] ———, *Error catastrophe and antiviral strategy*, Proceedings of the National Academy of Sciences (PNAS) **99** (October 15, 2002), no. 21, 13374–13376.

[29] Manfred Eigen, John McCaskill, and Peter Schuster, *Molecular Quasi-Species*, Journal of Physical Chemistry **92** (1988), 6881–6891.

[30] Manfred Eigen and Peter Schuster, *The Hypercycle, a Principle of Natural Self-Organization*, Springer-Verlag, Berlin, 1979.

[31] Teferi Eshetu, Nicole Berens-Riha1, Sintayehu Fekadu, Zelalem Tadesse, Robert Grkov, Michael Hölscher, Thomas Löcher, and Isabel Barreto Miranda, *Different mutation patterns of Plasmodium falciparum among patients in Jimma University Hospital, Ethiopia*, Malaria Journal **226** (August 7, 2010), no. 9, Published Online.

[32] Aihua Fan and Yunping Jiang, *On Ruelle-Perron-Frobenius operators. I. Ruelle theorem*, Comm. Math. Phys. **223** (2001), no. 1, 125–141. MR MR1860762 (2002i:37029a)

[33] ———, *On Ruelle-Perron-Frobenius operators. II. Convergence speeds*, Comm. Math. Phys. **223** (2001), no. 1, 143–159. MR MR1860763 (2002i:37029b)

[34] Pierre Ferrero and Bernhard Schmitt, *Ruelle's Perron-Frobenius theorem and projective metrics*, Coll. Math. Soc. J'ane Bolyai **27** (1979).

[35] Jasmine Foo and Franziska Michor, *Evolution of resistance to anti–cancer therapy during general dosing schedules*, Journal of Theoretical Biology **263** (2010), 179–188.

[36] Kevin R. Foster, Tom Wenseleers, and Francis L.W. Ratnieks, *Kin selection is the key to altruism*, TRENDS in Ecology and Evolution **21** (2006), no. 2, 57–60.

[37] Francis Galton, *Problem 4001: On the extinction of surnames*, Educational Times **26** (April 1873), 17.

[38] Stéphane Gaubert and Jeremy Gunawardena, *The Perron-Frobenius theorem for homogeneous, monotone functions*, Trans. Amer. Math. Soc. **356** (2004), no. 12, 4931–4950 (electronic). MR MR2084406 (2006d:15038)

[39] Semyon Aranovich Gershgorin, *Über die abgrenzung der eigenwerte einer matrix*, Izv. Akad. Nauk. USSR Otd. Fiz.–Mat. Nauk **7** (1931), 749–754.

[40] Charles M. Grinstead and James Laurie Snell, *Introduction to probability*, American Mathematical Society, Providence, RI, 1998.

[41] Patsy Haccou, Peter Jagers, and Vladimir A. Vatutin, *Branching processes: variation, growth, and extinction of populations*, Cambridge Studies in Adaptive Dynamics, Cambridge University Press, Cambridge, 2007. MR MR2429372 (2009h:92064)

[42] Theodore E. Harris, *The theory of branching processes*, Dover Phoenix Editions, Dover Publications Inc., Mineola, NY, 2002, Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)]. MR MR1991122

[43] Allen Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. MR MR1867354 (2002k:55001)

[44] Michiel Hazewinkel, *Encyclopaedia of mathematics*, Springer-Verlag, Berlin Heidelberg New York, 2002.

[45] C. C. Heyde and E. Seneta, *Studies in the History of Probability and Statistics. XXXI. The Simple Branching Process, a Turning Point Test and a Fundamental Inequality: A Historical Note on I. J. Bienaymé*, Biometrika **59** (1972), no. 3, 680–683.

[46] ———, *I. J. Bienaymé. Statistical theory anticipated*, Springer-Verlag, New York, 1977, Studies in the History of Mathematics and Physical Sciences, No. 3. MR MR0462888 (57 #2855)

[47] David Hilbert, *Über die gerade Linie als kürzeste Verbindung zweier Punkte. (Aus einem an Herrn F. Klein gerichten Briefe.)*, Math. Ann. **46** (1895), 91–96.

[48] Yoh Iwasa, Franziska Michor, and Martin A. Nowak, *Evolutionary dynamics of escape from biomedical intervention*, Proc R Soc B **270** (2003), 2573–2578.

[49] ———, *Evolutionary dynamics of invasion and escape*, J. Theoret. Biol. **226** (2004), no. 2, 205–214. MR MR2069303

[50] Peter Jagers, *Branching processes with biological applications*, Wiley-Interscience [John Wiley & Sons], London, 1975, Wiley Series in Probability and Mathematical Statistics—Applied Probability and Statistics. MR MR0488341 (58 #7890)

[51] Yunping Jiang, *Nanjing Lecture Notes In Dynamical Systems. Part One: Transfer Operators in Thermodynamical Formalism*, FIM Publication Series, ETH-Zurich, June 2000.

[52] Tosio Kato, *Perturbation theory for linear operators*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Reprint of the 1980 edition. MR 1335452 (96a:47025)

[53] David G. Kendall, *The genealogy of genealogy: branching processes before (and after) 1873*, Bull. London Math. Soc. **7** (1975), no. 3, 225–253, With a French appendix containing Bienaymé's paper of 1845. MR MR0426186 (54 #14132)

[54] Marek Kimmel and David E. Axelrod, *Branching processes in biology*, Interdisciplinary Applied Mathematics, vol. 19, Springer-Verlag, New York, 2002. MR MR1903571 (2003b:60004)

[55] John Kingman, *David George Kendall. 15 January 1918 – 23 October 2007*, Biogr. Mems Fell. R. Soc **55** (2009), 121 – 138, first published online 14 May 2009.

[56] V. L. Klee, Jr., *Extremal structure of convex sets*, Arch. Math. (Basel) **8** (1957), 234–240. MR MR0092112 (19,1065a)

[57] V. L. Jr. Klee, *Separation properties of convex cones*, Proceedings of the American Mathematical Society **6** (1955), no. 2, 313–318.

[58] Felix Klein, *Elementary mathematics from an advanced standpoint*, Dover Publications Inc., Mineola, NY, 2004, Geometry, Translated from the third German edition and with a preface by E. R. Hendrik and C. A. Noble, Reprint of the 1949 translation. MR MR2078728 (2005c:01029)

[59] ———, *Ueber die sogenannte Nicht-Euklidische Geometrie*, Math. Ann. **4** (December, 1871), no. 4, 573–625.

[60] ———, *Ueber die sogenannte Nicht-Euklidische Geometrie. Zweiter Aufsatz*, Math. Ann. **6** (June, 1873), no. 2, 112–145. MR MR1509812

[61] Morris Kline, *Mathematical thought from ancient to modern times. Vol. 3*, second ed., The Clarendon Press Oxford University Press, New York, 1990. MR MR1058203 (91i:01003c)

[62] M.A. Krasnosel'skii, *Positive solutions of operator equations*, P. Noordhff Ltd., Groningen, The Netherlands, 1964, QA320 .K6813.

[63] Peter D. Lax, *Linear algebra and its applications, Volume 10*, second ed., Wiley-Interscience, New York, 2007.

[64] Carlangelo Liverani, *Decay of correlations*, Ann. of Math. (2) **142** (1995), no. 2, 239–301. MR MR1343323 (96e:58090)

[65] Ern Loh, Jesse J. Salk, and Lawrence A. Loeb, *Optimization of DNA polymerase mutation rates during bacterial evolution*, Proceedings of the National Academy of Sciences USA (PNAS) **107** (January 19, 2010), no. 3, 1154 – 1159.

[66] Elena Anne Marchisotto, *The Theorem of Pappus: A bridge between algebra and geometry*, The American Mathematical Monthly **109** (Jun. - Jul., 2002), no. 6, 497–516.

[67] Curtis T. McMullen, *Coxeter groups, Salem numbers and the Hilbert metric*, Publ. Math. Inst. Hautes Études Sci. (2002), no. 95, 151–183. MR MR1953192 (2004b:20054)

[68] Carl Meyer, *Matrix analysis and applied linear algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000, With 1 CD-ROM (Windows, Macintosh and UNIX) and a solutions manual (iv+171 pp.). MR 1777382

[69] John James Milne, *An elementary treatise on cross-ratio geometry, with historical notes*, Cambridge: The University Press, 1911.

[70] John Milnor, *Hyperbolic geometry: the first 150 years*, Bull. Amer. Math. Soc. (N.S.) **6** (1982), no. 1, 9–24. MR MR634431 (82m:57005)

[71] James R. Munkres, *Topology: a first course*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1975. MR 0464128 (57 #4063)

[72] James Dickson Murray, *Mathematical biology I. An introduction, Third Edition*, Springer-Verlag, New York, 2002.

[73] Arch W. Naylor and George R. Sell, *Linear operator theory in engineering and science*, second ed., Applied Mathematical Sciences, vol. 40, Springer-Verlag, New York, 1982. MR MR672108 (83j:46001)

[74] James R. Norris, *Markov Chains*, Cambridge series in statistical and probabilistic mathematics, Cambridge Univ. Press., Cambridge, 1999.

[75] Martin A. Nowak and Robert M. May, *Virus dynamics*, Oxford University Press, Oxford, 2000, Mathematical principles of immunology and virology. MR MR2009143

[76] Roger D. Nussbaum, *Iterated nonlinear maps and Hilbert's projective metric, II*, Memoirs of the AMS **79** (1989), no. 401.

[77] Roger D. Nussbaum and Cormac Walsh, *A metric inequality for the Thompson and Hilbert geometries*, Journal of Inequalities in Pure and Applied Mathematics **5** (2004), no. 3.

[78] So Yeon Park, Mithat Gnen, Hee Jung Kim, Franziska Michor, and Kornelia Polyak, *Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype*, The Journal of Clinical Investigation **120** (2010), 636 – 644.

[79] Oskar Perron, *Zur Theorie der Matrices*, Math. Ann. **64** (1907), no. 2, 248–263. MR MR1511438

[80] B. B. Phadke, *A triangular world with hexagonal circles*, Geometriae Dedicata **3** (1974/75), 511–520. MR MR0367819 (51 #4061)

[81] Satish K Pillai, Joseph K Wong, and Jason D Barbour, *Turning up the volume on mutational pressure: Is more of a good thing always better? (A case study of HIV-1 Vif and APOBEC3)*, Retrovirology **26** (2008), no. 5, Published Online.

[82] H. L. Royden, *Real Analysis*, third ed., Macmillan Publishing Company, New York, 1988. MR 1013117 (90g:00004)

[83] Walter Rudin, *Functional Analysis*, McGraw Hill, New York, 1973.

[84] ———, *Real and Complex Analysis*, third ed., McGraw-Hill Book Co., New York, 1987. MR 924157 (88k:00002)

[85] Michael E. Starzak, *Mathematical Methods in Chemistry and Physics*, Plenum Press, New York, 1989.

[86] John Stillwell, *Sources of hyperbolic geometry*, History of Mathematics, vol. 10, American Mathematical Society, Providence, RI, 1996. MR MR1402697 (97k:01071)

[87] Joseph L. Taylor, *Several Complex Variables with Connections to Algebraic Geometry and Lie groups*, Graduate Studies in Mathematics, vol. 46, American Mathematical Society, Providence, RI, 2002. MR MR1900941 (2004b:32001)

[88] Henry William Watson and Francis Galton, *On the probability of the extinction of families*, J. Anthropol. Inst. Great B. and Ireland **4** (1874), 138–144.

[89] Claus O. Wilke, *Quasispecies theory in the context of population genetics*, `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1208876/`, August 2005, Published Online by BioMed Central Evolutionary Biology.

[90] Edward O. Wilson, *Sociobiology. The Abridged Edition.*, Belknap Press of Harvard University Press, Cambridge, Mass, 1980.

[91] ———, *Kin Selection as the Key to Altruism: Its Rise and Fall*, Social Research **72** (Spring 2005), no. 1, 159–166.

[92] P.P. Zabreiko, M.A. Krasnosel'skii, and Yu. V. Pokornyi, *On a class of linear positive operators*, Functional Analysis and its Applications **5** (1971), no. 4, 272–279.

Autobiographical Information

Chris McCarthy hiking near Bear Mountain, NY. Self portrait.

Born in New York City. Raised in Flushing, New York.