TECHNICAL REPORT 1930
July 2005

Multiobjective Optimization on Function Spaces: A Kolmogorov Approach

J. C. Allen
D. Acero

SSC San Diego

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **JUL 2005**	2. REPORT TYPE	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Multiobjective Optimization on Function Spaces: A Kolmogorov Approach**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center,53560 Hull Street,San Diego,CA,92152-5001**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**62**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

TECHNICAL REPORT 1930
July 2005

Multiobjective Optimization on Function Spaces: A Kolmogorov Approach

J. C. Allen
D. Acero

SPAWAR
Systems Center
San Diego

SSC San Diego
San Diego, CA 92152-5001

Executive Summary

This report originated in the H^∞ Research Initiative of the Office of Naval Research and the ILIR Program of SPAWAR Systems Center San Diego. These programs migrated H^∞ Engineering into fleet applications, specifically wideband impedance matching and wideband amplifier optimization. Research in these applications produced several papers [24], [23], [3], [4], four patents, a book [2], and sparked the Defense Advanced Research Projects Agency's interest in Digital H^∞ Engineering.

As the applications coalesced, a general principle underlying these optimization problems became apparent—that solutions of these optimization problems could be characterized by the Kolmogorov Criterion. This report makes explicit that the Kolmogorov Criterion can specialize with sufficient detail to yield concrete and computationally viable tests that identify solutions to difficult optimization problems.

Specifically, the classical "equal-ripple" characterization of best polynomial approximation is generalized to nonlinear polynomial optimization, and then generalized again to multiobjective polynomial optimization. Thus, results in polynomial optimization stretching over this last century readily fit into a single framework and are illustrated with applications in filter design and control theory. In addition to the finite-dimensional polynomials, the Kolmogorov Criterion also applies to the infinite-dimensional disk algebra. The disk algebra is basic to signal processing and control theory. Many engineering problems in these disciples are optimization problems on the disk algebra. The Kolmogorov Criterion readily characterizes the minimizers of these nonlinear optimization problems.

By making explicit the Kolmogorov Criterion and working specific examples, this report equips researchers with a general approach to optimization on spaces of functions and a collection of accessible research problems.

Contents

List of Figures

List of Tables

1 The Mathematical Summary

Let Z be a compact subset of the complex numbers $\mathbb{C}$. Let $C(Z, \mathbb{C})$ denote the complex-valued functions that are continuous on Z. A real-valued function $\Gamma : Z \times \mathbb{C} \to \mathbb{R}$ is called a *performance* function. A continuous performance function induces a continuous *objective* function $\gamma : C(Z, \mathbb{C}) \to \mathbb{R}$:

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in Z\}.$$

Let $\mathcal{H}$ denote a subset of $C(Z, \mathbb{C})$. Minimization of this objective function γ on $\mathcal{H}$ is the general optimization problem:

$$\inf\{\gamma(h) : h \in \mathcal{H} \subset C(Z, \mathbb{C})\}.$$

Important for both theory and computation is recognizing solutions to this minimization problem. Specifically, *if you were handed a minimizer*

$$h_{\min} := \operatorname{argmin}\{\gamma(h) : h \in \mathcal{H}\},$$

could you recognize that $h_{\min}$ *was a minimum of* γ*?* Recognizing such minimizers is the *characterization* problem. The multiobjective characterization problem has $\Gamma : Z \times \mathbb{C} \to \mathbb{R}^M$ and minimizes the corresponding vector-valued function:

$$\gamma(h) := \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \\ \vdots \\ \gamma_M(h) \end{bmatrix} \qquad \gamma_m(h) := \sup\{\Gamma_m(z, h(z)) : z \in Z\}.$$

We consider the characterization problem for the following subspaces:

- Polynomials $\mathcal{P}^N$
- Disk algebra $\mathcal{A}(\mathbf{D})$

The Kolmogorov Criterion provides an easy route to the necessary conditions that characterize a minimizer while the interpolating properties of the subspaces complete the sufficiency arguments.

For optimization on the polynomials, a new characterization of polynomial minimizers is obtained. This characterization is a substantial extension of the well-known "equal-ripple" theorem of polynomial approximation [11]. Applications to nonlinear approximation and spectral factorization illustrate this result.

For optimization on the infinite-dimensional disk algebra, we recapture Helton and Merino's [17] flatness and winding number characterization of minimizers. The point of this recapitulation is to show flexibility of the Kolmogorov approach. Applications to impedance matching and control theory illustrate the characterization and bring us to minimizing multiple objective functions.

For multiobjective optimization, a new identification of polynomial minimizers is obtained by spreading the equal-ripple result over the multiple objective functions. However, a "phase-splitting" phenomenon confounds the sufficiency argument. Nevertheless, this new theory is sufficient to explore the set of all possible minima and uncover a surprisingly fine structure.

In summary, Kolmogorov Criterion is a computational framework for exploring optimization theory in general with sufficient detail to deliver specific results on optimization on function spaces.

2 Notation and Preliminaries

The real numbers are denoted by $\mathbb{R}$. Real N-dimensional space is denoted by $\mathbb{R}^N$. The closed positive cone of $\mathbb{R}^N$ is denoted by $\mathbb{R}^N_+$. The complex numbers are denoted by $\mathbb{C}$. Complex N-dimensional space is denoted by $\mathbb{C}^N$. Throughout this report, Z denotes a compact subset of $\mathbb{C}$. The open unit disk

$$\mathbf{D} := \{z \in \mathbb{C} : |z| < 1\}$$

has the unit circle $\mathbf{T}$ as boundary

$$\mathbf{T} := \{z \in \mathbb{C} : |z| = 1\}.$$

If E is a Banach space with norm $\| \circ \|_E$, $C(Z, E)$ denotes the set of continuous functions $h : Z \to E$ with norm

$$\|h\|_\infty := \sup\{\|h(z)\|_E : z \in Z\}.$$

In more detail,

- $C(Z, \mathbb{R})$ denotes the set of continuous real-valued functions on Z
- $C(Z, \mathbb{C})$ denotes the set of continuous complex-valued functions on Z
- $C(Z, \mathbb{C}^M)$ denotes the continuous $\mathbb{C}^M$-valued functions on Z

If $\Gamma \in C(Z \times \mathbb{C}, \mathbb{R})$ is continuous, it lifts to the mapping $\widetilde{\Gamma} : C(Z, \mathbb{C}) \to C(Z, \mathbb{R})$

$$\widetilde{\Gamma}(h; z) = \Gamma(z, h(z))$$

that induces the *objective* function $\gamma : C(Z, \mathbb{C}) \to \mathbb{R}$

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in Z\}.$$

The associated *critical set* of $h \in C(Z, \mathbb{C})$ is denoted

$$\mathrm{crit}[\gamma(h)] := \{z \in Z : \gamma(h) = \Gamma(z, h(z))\}.$$

If $\mathcal{H}$ is a subspace of $C(Z, \mathbb{C})$, a *nonzero* $\Delta h \in \mathcal{H}$ is called a *direction of nonincrease* [9] or *direction of descent* for γ provided for all $t > 0$ sufficiently small

$$\gamma(h + t\Delta h) \leq \gamma(h).$$

The function $h \in \mathcal{H}$ is called a *local minimum* for γ provided for all $\Delta h \in \mathcal{H}$ sufficiently small there holds

$$\gamma(h + \Delta h) \geq \gamma(h).$$

A Taylor's expansion in $C(Z, \mathbb{C})$ is needed. Following Helton's notation, recall the derivative on $\mathbb{C}$ has the form [22]

$$\partial = \frac{\partial}{\partial z} = \frac{1}{2}\left\{\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right\}.$$

If $\Gamma : \mathbb{C} \to \mathbb{R}$ is C^2, Taylor's expansion is

$$\begin{aligned}\Gamma(h + \Delta h) &= \Gamma(h) + \frac{\partial \Gamma}{\partial x}(h)\Delta u + \frac{\partial \Gamma}{\partial y}(h)\Delta v + \mathcal{O}[|\Delta h|^2] \\ &= \Gamma(h) + 2\Re[\partial\Gamma(h)\Delta h] + \mathcal{O}[|\Delta h|^2],\end{aligned}$$

where $\Delta h = \Delta u + i\Delta v \in \mathbb{C}$. The Omega Lemma lifts this expansion to the corresponding expansion for $\widetilde{\Gamma}$ operating on $C(Z, \mathbb{C})$.

Lemma 1 (Omega) *[1] Let E and F be Banach spaces. Let $U \subseteq E$ be open. Assume $g : U \subseteq E \to F$ is a C^r map ($r > 0$) with first variation $Dg : E \to F$. Let M be a compact topological space. Then the map $\widetilde{g} : C(M, U) \to C(M, F)$ defined by*

$$\widetilde{g}(h; m) := g(h(m))$$

is also C^r. The derivative of $\widetilde{g}$ at $h \in C(M, U)$ is denoted $D\widetilde{g}(h)$ and is the linear map $D\widetilde{g}(h) : C(M, E) \to C(M, F)$

$$D\widetilde{g}(h)[\Delta h; m] := Dg(h(m))[\Delta h(m)].$$

The only modification needed to get Taylor's expansion is to account for the fact that the domain of Γ is $Z \times \mathbb{C}$. Let

$$\partial_1\Gamma(z_1, z_2) = \frac{\partial \Gamma}{\partial z_1}(z_1, z_2), \qquad \partial_2\Gamma(z_1, z_2) = \frac{\partial \Gamma}{\partial z_2}(z_1, z_2).$$

Lemma 2 (Γ) *Let $Z \subset \mathbb{C}$ be compact. Let $U \subseteq \mathbb{C}$ be an open subset containing Z. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^r $(r > 0)$ with first variation*

$$D\Gamma(z_1, z_2) = [\partial_1 \Gamma(z_1, z_2) \;\; \partial_2 \Gamma(z_1, z_2)].$$

Then the map $\widetilde{\Gamma} : C(Z, \mathbb{C}) \to C(Z, \mathbb{R})$ defined by

$$\widetilde{\Gamma}(h; z) := \Gamma(z, h(z))$$

is also C^r. The derivative of $\widetilde{\Gamma}$ at $h \in C(Z, \mathbb{C})$ is the linear map $D\widetilde{\Gamma}(h) : C(Z, \mathbb{C}) \to C(Z, \mathbb{R})$

$$D\widetilde{\Gamma}(h)[\Delta h; z] := 2\Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)].$$

The Taylor expansion exists on $C(Z, \mathbb{C})$ as

$$\Gamma(z, h(z) + \Delta h(z)) = \Gamma(z, h(z)) + 2\Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)] + \mathcal{O}[\|\Delta h\|_\infty^2],$$

where $\mathcal{O}[\|\Delta h\|_\infty^2]$ does not depend on $z \in Z$.

Proof: Let $\mathbf{h} \in C(Z, \mathbb{C}^2)$ be written as

$$\mathbf{h}(z) = \begin{bmatrix} h_1(z) \\ h_2(z) \end{bmatrix} = \begin{bmatrix} u_1(z) + iv_1(z) \\ u_2(z) + iv_2(z) \end{bmatrix}$$

and with the corresponding notation for $\Delta \mathbf{h}(z)$. The Omega Lemma gives that $\widetilde{\Gamma} : C(Z, \mathbb{C}^2) \to C(Z, \mathbb{R})$ defined by $\widetilde{\Gamma}(\mathbf{h}; z) := \Gamma(h_1(z), h_2(z))$ is C^r with derivative

$$\begin{aligned} D\widetilde{\Gamma}(\mathbf{h})[\Delta \mathbf{h}; z] &= D\Gamma(\mathbf{h}(z)) \Delta \mathbf{h}(z) \\ &= \left[\frac{\partial \Gamma}{\partial x_1}(\mathbf{h}(z)) \;\; \frac{\partial \Gamma}{\partial y_1}(\mathbf{h}(z)) \;\; \frac{\partial \Gamma}{\partial x_2}(\mathbf{h}(z)) \;\; \frac{\partial \Gamma}{\partial y_2}(\mathbf{h}(z)) \right] \begin{bmatrix} \Delta u_1(z) \\ \Delta v_1(z) \\ \Delta u_2(z) \\ \Delta v_2(z) \end{bmatrix} \\ &= 2\Re[\partial_1 \Gamma(\mathbf{h}(z)) \Delta h_1(z)] + 2\Re[\partial_2 \Gamma(\mathbf{h}(z)) \Delta h_2(z)]. \end{aligned}$$

Restrict $\widetilde{\Gamma}$ to the affine space

$$\mathcal{M} = \{\mathrm{id}\} \times C(Z, \mathbb{C}) = \left\{ \begin{bmatrix} z \\ h(z) \end{bmatrix} : h \in C(Z, \mathbb{C}) \right\}.$$

$\mathcal{M}$ has tangent space

$$T\mathcal{M} = \{0\} \times C(Z, \mathbb{C}) = \left\{ \begin{bmatrix} 0 \\ \Delta h(z) \end{bmatrix} : \Delta h \in C(Z, \mathbb{C}) \right\}.$$

$\widetilde{\Gamma}$ restricted to $\mathcal{M}$ has derivative

$$D(\widetilde{\Gamma}|\mathcal{M})(\mathbf{h})[\Delta\mathbf{h}; z] = D\Gamma(\mathbf{h})\begin{bmatrix} 0 \\ \Delta h(z) \end{bmatrix} = 2\Re[\partial_2\Gamma(z, h(z))\Delta h(z)].$$

Taylor's expansion follows from this first variation. ///

The end-of-proof symbol is "///." On occasion, a point x will be "added" to a subset B of a vector space:

$$x + B = \{x + b : b \in B\}.$$

Likewise, the sum of sets A and B of a vector space is denoted

$$A + B = \{a + b : a \in A, b \in B\}.$$

Table 1: Summary of notation.

Variable	Description
$\mathbb{R}$	real numbers
$\mathbb{R}^N$	real N-dimensional space
$\mathbb{R}_+$	non-negative real numbers
$\mathbb{R}_+^N$	positive cone of $\mathbb{R}^N$
$\mathbb{C}$	complex numbers
$\mathbb{C}^N$	complex N-space
$\mathbf{D}$	open unit disk in $\mathbb{C}$
$\mathbf{T}$	unit circle
$C(Z, E)$	continuous E-valued functions on the compact set Z
$\mathcal{P}^N$	real polynomials of degree not exceeding N
$\mathcal{A}(\mathbf{D})$	disk algebra
$\text{crit}[\gamma(h)]$	critical set of $\gamma(h)$
$\overline{z}$	complex conjugate of z
///	end-of-proof symbol

3 The Kolmogorov Criterion

The Kolmogorov Criterion characterizes optimal points of the best approximation problem and the minimizers of convex functions. For brevity, the Kolmogorov Criterion is stated only for the best approximation problem while the text develops the criterion for the nonlinear minimization problems.

Theorem 1 (Kolmogorov Criterion) *[7, pages 6–11]. Let X be a Bananch space with dual space X^*. Let K a convex subset of X. The following are equivalent:*

(a) *$k_0 \in K$ is a best approximation to $x \in X$:*

$$\|x - k_0\| = \inf\{\|x - k\| : k \in K\}.$$

(b) *There exists an $x^* \in X^*$ that has unit norm*

$$\|x^*\| = 1,$$

that supports the error function

$$\langle x^*, x - k_0 \rangle = \|x - k_0\|,$$

and belongs to the negative cone of K

$$0 \geq \Re[\langle x^*, k \rangle] \quad (k \in K).$$

For nonlinear functions, the first variation "almost" convexifies the problem. However, the nonlinearity splits the necessary and sufficient conditions of the Kolmogorov Criterion. The necessary condition for optimization on $\mathcal{P}^N$ and $\mathcal{A}(\mathbf{D})$ is the easy part of the Kolmogorov Criterion [7, pages 6–11]. Although the result holds for arbitrary sets using tangent and contingent cones, we state it only for subspaces.

Lemma 3 (Descent) *Let $Z \subset \mathbb{C}$ be compact. Let $\mathcal{H}$ be a closed linear subspace of $C(Z, \mathbb{C})$. Let U be an open subset containing Z. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^2. Define $\gamma : \mathcal{H} \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in Z\}.$$

Assume $\mathcal{H}$ is boundedly compact. If $h \in \mathcal{H}$ is not a local minimum, there exists a nonzero $\Delta h \in \mathcal{H}$ such that

$$0 \geq \Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)] \quad (z \in \mathrm{crit}[\gamma(h)]).$$

Proof: If $h \in \mathcal{H}$ not a local minimum, there exists a nonzero sequence $\{\Delta h_n\} \subset \mathcal{H}$ converging to zero such that $\gamma(h + \Delta h_n) \leq \gamma(h)$. Set $t_n := \|\Delta h_n\|_\infty > 0$ and $u_n := t_n^{-1} \Delta h_n$. Compactness of $\mathcal{H}$ implies that the bounded sequence $\{u_n\}$ contains a convergent subsequence. By relabeling, let $u_n \to \Delta h \in \mathcal{H}$. Because u_n has unit norm, Δh cannot be zero. For all $z \in \text{crit}[\gamma(h)]$, Lemma 2 provides the expansion:

$$\begin{aligned}\gamma(h + \Delta h_n) &\geq \Gamma(z, h(z) + \Delta h_n(z)) \\ &= \Gamma(z, h(z)) + 2\Re\left[\partial_2 \Gamma(z, h(z)) \Delta h_n(z)\right] + \mathcal{O}[t_n^2] \\ &= \gamma(h) + 2\Re\left[\partial_2 \Gamma(z, h(z)) \Delta h_n(z)\right] + \mathcal{O}[t_n^2].\end{aligned}$$

Subtract $\gamma(h)$ from both sides, divide by $t_n > 0$ to get

$$0 \geq \Re\left[\partial_2 \Gamma(z, h(z)) u_n(z)\right] + \mathcal{O}[t_n].$$

Letting $n \to \infty$ gives the result. ///

The Descent Lemma (Lemma 3) has a clean proof that reveals why boundedly compact supplies a "direction of descent." It also supplies various points-of-departure for more sophisticated results. For example, a minimization test is obtained, provided the "=" in the "≥" is handled with care.

Lemma 4 (Minimum Test) *Let $Z \subset \mathbb{C}$ be compact. Let $\mathcal{H}$ be a closed linear subspace of $C(Z, \mathbb{C})$. Let U be an open subset containing Z. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^2. Define $\gamma : \mathcal{H} \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in Z\}.$$

Let $h \in \mathcal{H}$. If there exists a $\Delta h \in \mathcal{H}$ such that

$$0 > \Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)] \qquad (z \in \text{crit}[\gamma(h)]),$$

$h \in \mathcal{H}$ cannot be a local minimum for γ.

Proof: Compactness of Z and continuity give the existence of a $\delta > 0$ such that $\Re\left[\partial_2 \Gamma(z, h) \Delta h\right] \leq -\delta < 0$ on $\text{crit}[\gamma(h)]$. Continuity gives an open neighborhood U of $\text{crit}[\gamma(h)]$ such that for all $z \in U$ there holds:

$$\Re\left[\partial_2 \Gamma(z, h) \Delta h(z)]\right] \leq -\delta/2 < 0.$$

Then for $z \in U$ and for $t > 0$ sufficiently small there holds

$$\begin{aligned}\Gamma(z, h(z) + t\Delta h(z)) &= \Gamma(z, h(z)) + t2\Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)] + \mathcal{O}[t^2] \\ &\leq \gamma(h) - \delta t + \mathcal{O}[t^2] \\ &< \gamma(h).\end{aligned}$$

The first equality is obtained by taking $t > 0$ so small that $t\Delta h \in B(0, \epsilon)$ and applying Lemma 2. The first inequality follows from the δ bound on U. The last inequality follows by taking $t > 0$ small enough so that the first-order term dominates the second-order term. For $z \in Z \setminus U$, continuity forces $\Gamma(z, h(z)) < \gamma(h)$. Continuity of Γ and compactness of $Z \setminus U$ imply

$$\Gamma(z, h(z) + t\Delta h(z)) < \gamma(h)$$

for $t > 0$ sufficiently small. Thus, $\gamma(h + t\Delta h) < \gamma(h)$ for all $t > 0$ sufficiently small. Consequently, h cannot be a local minimum of γ. ///

The Minimum Test (Lemma 4) tells us that $h \in \mathcal{H}$ cannot be a local minimum if we can find a $\Delta h \in \mathcal{H}$ that "interpolates" the first variation $\partial_2\Gamma(h)$ on the critical set $\mathrm{crit}[\gamma(h)]$. Conversely, if $h \in \mathcal{H}$ is a local minimum, no such interpolator can exist. That is, $\Re[\partial_2\Gamma(z, h(z))\Delta h(z)]$ must assume positive and negative values on $\mathrm{crit}[\gamma(h)]$ for any $\Delta h \in \mathcal{H}$. Put another way, a local minimum will force $\partial_2\Gamma(z, h(z))\Delta h(z)$ to wind around zero. Thus, even at this abstract level, the winding numbers appear in the characterization of minima.

The Descent Lemma (Lemma 3) uses a "$\leq$". The Minimum Test (Lemma 4) needs a "$<$" . The necessary and sufficient conditions fail on the "$=$". The bulk of our efforts are devoted to bridging this gap. The basic idea is to exploit the interpolating properties of the subspaces. The polynomials are the classic interpolating space.

4 Optimization on $\mathcal{P}^N$

It is instructive to consider the minimization problem for the real polynomials $\mathcal{P}^N$ in $C([0, 1], \mathbb{R})$. Let $\Gamma : [0, 1] \times \mathbb{R} \to \mathbb{R}$ be C^2. The complex derivative is unnecessary but adapt the notation as follows:

$$\partial_2\Gamma(x_1, x_2) := \frac{\partial\Gamma}{\partial x_2}(x_1, x_2).$$

The open set condition in Lemma 2 becomes $[0, 1] \subset U \subset \mathbb{R}$ so that non-differentiability at 0 or 1 is not an issue. Define the mapping $\gamma : C([0, 1], \mathbb{R}) \to \mathbb{R}$ by

$$\gamma(h) := \sup\{\Gamma(x, h(x)) : x \in [0, 1]\}.$$

Suppose $h \in \mathcal{P}^N$ is a local minimum. The Descent Lemma gives that no $\Delta h \in \mathcal{P}^N$ exists such that

$$\partial_2\Gamma(x, h(x))\Delta h(x) < 0 \qquad (x \in \mathrm{crit}[\gamma(h)]).$$

The interpolating properties of the polynomials force a classical support and alignment condition.

Lemma 5 (Support) *Let U be an open subset containing $[0,1]$. Let $\Gamma : U \times \mathbb{R} \to \mathbb{R}$ be C^2. Define the mapping $\gamma : C([0,1], \mathbb{R}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(x, h(x)) : x \in [0,1]\}.$$

Suppose $h \in \mathcal{P}^N$ is a local minimum of $\gamma : \mathcal{P}^N \to \mathbb{R}$. Assume $\partial_2\Gamma(x, h(x)) \neq 0$ for $x \in \mathrm{crit}[\gamma(h)]$. Then $|\mathrm{crit}[\gamma(h)]| \geq N+2$

Proof: If $\mathrm{crit}[\gamma(h)]$ contains $N+1$ points or less, there exists a $\Delta h \in \mathcal{P}^N$ such that

$$\Delta h(x) = -\mathrm{sign}(\partial_2\Gamma(x, h(x))) \quad (x \in \mathrm{crit}[\gamma(h)]).$$

This forces

$$\partial_2\Gamma(x, h(x))\Delta h(x) < 0 \quad (x \in \mathrm{crit}[\gamma(h)]).$$

By the Minimum Test (Lemma 4), $h \in \mathcal{P}^N$ is not a local minimum. This contradiction forces at least $N+2$ points into $\mathrm{crit}[\gamma(h)]$. ///

Lemma 6 (Alignment) *Let U be an open subset containing $[0,1]$. Let $\Gamma : U \times \mathbb{R} \to \mathbb{R}$ be C^2. Define the mapping $\gamma : C([0,1], \mathbb{R}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(x, h(x)) : x \in [0,1]\}.$$

Suppose $h \in \mathcal{P}^N \subset C([0,1], \mathbb{R})$ is a local minimum of $\gamma : \mathcal{P}^N \to \mathbb{R}$. Assume $\partial_2\Gamma(x, h(x)) \neq 0$ for $x \in \mathrm{crit}[\gamma(h)]$. Then $\partial_2\Gamma(x, h(x))$ admits an alternating sequence of length $N+2$ *on* $\mathrm{crit}[\gamma(h)]$. *That is, there are at least $N+2$ points $x_n \in \mathrm{crit}[\gamma(h)]$*

$$0 \leq x_1 < \ldots x_2 \ldots < x_{N+2} \leq 1$$

such that

$$\mathrm{sign}(\partial_2\Gamma(x_n, h(x_n))) = -\mathrm{sign}(\partial_2\Gamma(x_{n+1}, h(x_{n+1}))).$$

Proof: This standard argument is from Cheney [11]. If $\partial_2\Gamma(x, h(x))$ alternates only $N+1$ times on $\mathrm{crit}[\gamma(h)]$, then a polynomial Δh with N zeros placed at the sign changes of $\partial_2\Gamma(x, h(x))$ and by multiplication by ± 1 will have opposite sign as $\partial_2\Gamma(x, h(x))$ on $\mathrm{crit}[\gamma(h)]$. Thus, $\partial_2\Gamma(x, h(x))\Delta h(x) < 0$ on $\mathrm{crit}[\gamma(h)]$. By the Minimum Test (Lemma 4), $h \in \mathcal{P}^N$ is not a local minimum. ///

The satisfying property of Haar spaces is that this condition is strong enough to force a useful converse. The proof reveals how the inequality furnished by the Descent Lemma must be folded into the *strict* inequality required by the Minimum Test (Lemma 4).

Corollary 1 *Let U be an open subset containing $[0,1]$. Let $\Gamma : U \times \mathbb{R} \to \mathbb{R}$ be C^2. Define the mapping $\gamma : C([0,1], \mathbb{R}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(x, h(x)) : x \in [0,1]\}.$$

Assume $h \in \mathcal{P}^N$ and $\partial_2\Gamma(x, h(x)) \neq 0$ for $x \in \operatorname{crit}[\gamma(h)]$. Then the following are equivalent:

(a) *$h \in \mathcal{P}^N$ is a local minimum.*

(b) *$\partial_2\Gamma(x, h(x))$ admits alternating sequence of length $N+2$ on $\operatorname{crit}[\gamma(h)]$.*

Proof: We have (a)⇒(b) so we need to prove (b)⇒(a). Suppose (b) holds but (a) is not true. The Descent Lemma (Lemma 3) provides a *nonzero* $\Delta h \in \mathcal{P}^N$ such that

$$\partial_2\Gamma(x, h(x))\Delta h(x) \leq 0$$

on $\operatorname{crit}[\gamma(h)]$. Because $\partial_2\Gamma(x, h(x))$ is continuous, does not vanish on $\operatorname{crit}[\gamma(h)]$, and alternates $N+2$ times on $0 \leq x_1 < \ldots x_{N+2} \leq 1$, Δh is forced to have at least one zero in each interval $[x_n, x_{n+1}]$ for $n = 1, \ldots, N+1$. If Δh was simply continuous, Δh could have as few as floor$(N/2) + 1$ zeros. This configuration happens when the zeros in each interval are common to adjacent end points. Figure 1 illustrates this phenomenon for $N = 3$. The alternating sequence is marked with the arrows. The graph of Δh is schematically shown by the curved lines. The figure shows how Δh can satisfy the inequality with only two zeros. However, Δh is a polynomial so the zeroes are at least second order. Thus, the third-order polynomial Δh has four zeros. More generally, any equality in $\partial_2\Gamma(x_n, h(x_n))\Delta h(x_n) = 0$ still forces $N+1$ zeros into $\Delta h \in \mathcal{P}^N$. This contradicts $0 \neq \Delta h$. Then (a) must be true. ///

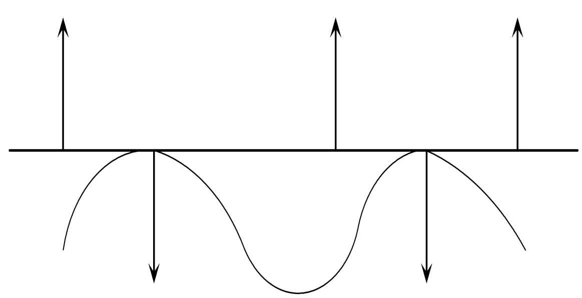

Figure 1: Minimal number of zeros of a continuous Δh for an alternating sequence of length 5.

What Corollary 1 demonstrates is that the Kolmogorov Criterion is a general optimization technique that is strong enough to specialize to specific problems—namely, nonlinear polynomial optimization.

4.1 Nonlinear Approximation of $\exp(-x)$

The exponential function is a classical test of approximation schemes. This section assesses one nonlinear approximation scheme of the exponential function: find a polynomial

$$h(x) = h_0 + h_1 x + h_2 x^2 + h_3 x^3 \in \mathcal{P}^3$$

that fits the exponential function as follows:

$$e^{-x} \approx h(x)^{-1}.$$

One approach chooses the performance function:

$$\Gamma(x, h(x)) = (e^{-x} - h(x)^{-1})^2.$$

The objective function is

$$\gamma(h) := \sup\{\Gamma(x, h(x)) : x \in [0, 1]\}.$$

The goal is to minimize the worst fit

$$\min\{\gamma(h) : h \in \mathcal{H}\}$$

over the subset $\mathcal{H} \subset \mathcal{P}^3$ consisting of those polynomials that never vanish on the unit interval. Although $\mathcal{H}$ is nonlinear, it is an open set of $\mathcal{P}^3$. As an open set, $\mathcal{H}$ admits enough local linear space structure to apply the Kolmogorov Theory. The variation of the performance function is

$$\partial_2\Gamma(x, h) = \partial_h(e^{-x} - h^{-1})^2 = 2(e^{-x} - h^{-1})h^{-2}.$$

Corollary 1 applies, provided the gradient $\partial_2\Gamma(x, h)$ does not vanish on the critical set of $\gamma(h)$: if $x \in \operatorname{crit}[\gamma(h)]$,

$$\partial_2\Gamma(x, h(x)) \neq 0 \iff 2(e^{-x} - h(x)^{-1})h(x)^{-2} \neq 0.$$

The error term cannot vanish because $h(x)$ is not a perfect fit to $\exp(x)$. The rational function $h(x)^{-2}$ cannot vanish on the unit interval. Consequently, no constraints on $h(x)$ really exist, except that $h(x)$ never vanishes on the unit interval. Therefore, Corollary 1 applies to characterize local minima—the gradient $\partial_2\Gamma(x, h)$ has an alternating sequence of length 5. Specifically,

$$0 \leq x_1 < x_2 < x_3 < x_4 < x_5 \leq 1$$

must exist in $\operatorname{crit}[\gamma(h)]$ such that

$$\operatorname{sign}(\partial_2\Gamma(x_n, h(x_n))) = -\operatorname{sign}(\partial_2\Gamma(x_{n+1}, h(x_{n+1}))).$$

Because

$$\text{sign}(\partial_2\Gamma(x, h(x))) = \text{sign}(e^{-x} - h(x)^{-1}),$$

a local minimum is characterized whenever the error term $e^{-x} - h(x)^{-1}$ alternates in sign on crit$[\gamma(h)]$. Figure 2 illustrates such an alternating sequence close to a local minimum. The coefficients of this near-local minimum are listed on the right of the plot:

$$h_{\min}(x) = 0.9998 + 1.0088x + 0.4453x^2 + 0.2629x^3.$$

The solid red segments mark those $x \in [0, 1]$ in the 95% neighborhood of crit$[\gamma(h)]$:

$$|e^{-x} - h_{\min}(x)^{-1}| > 0.95 \times \gamma(h_{\min}).$$

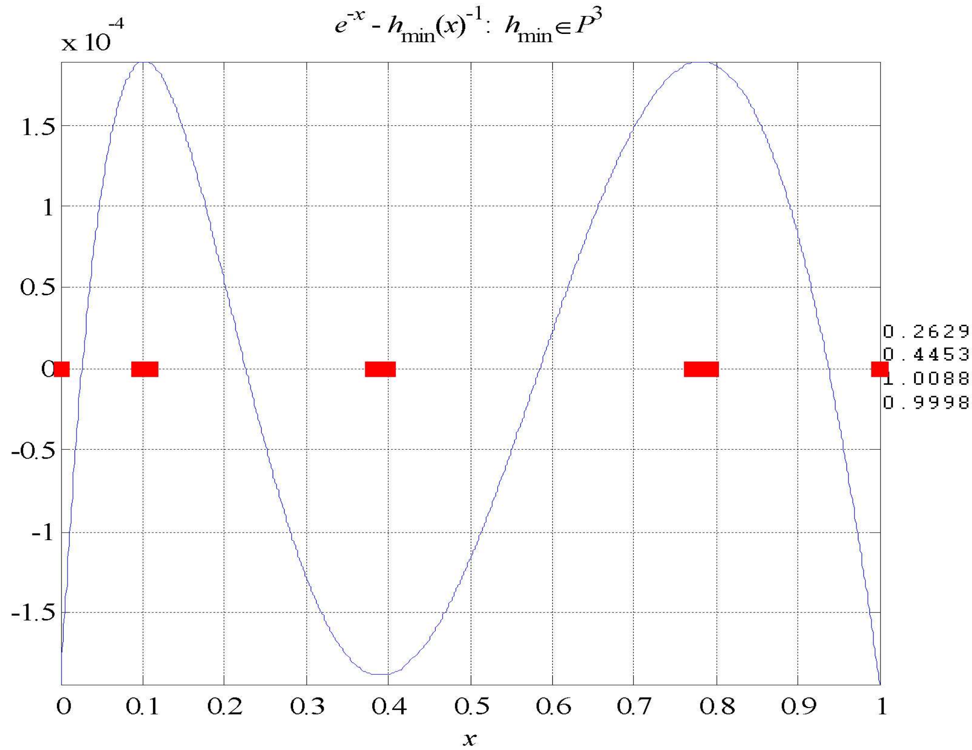

Figure 2: Error curve at near-local minimum.

For completeness, slices of the error surface at $h_{\min}$ are also plotted. The error surface is the graph of

$$h_0 + h_1 x + h_2 x^2 + h_3 x^3 \mapsto \gamma(h)$$

and needs five dimensions to plot. By varying only two coefficients of $h_{\min}$, we can see a three-dimensional slice of this error. Figures 3, 4, and 5 show these slices of the error function around $h_{\min}$. The plots reveal two general features. First, the minimum looks unique. Second, the error surface has non-differentiable “creases” that run through the minimum. Both features have theoretical and numerical consequences. Section 9 discusses these consequences and opportunities for research with the remainder of the research topics.

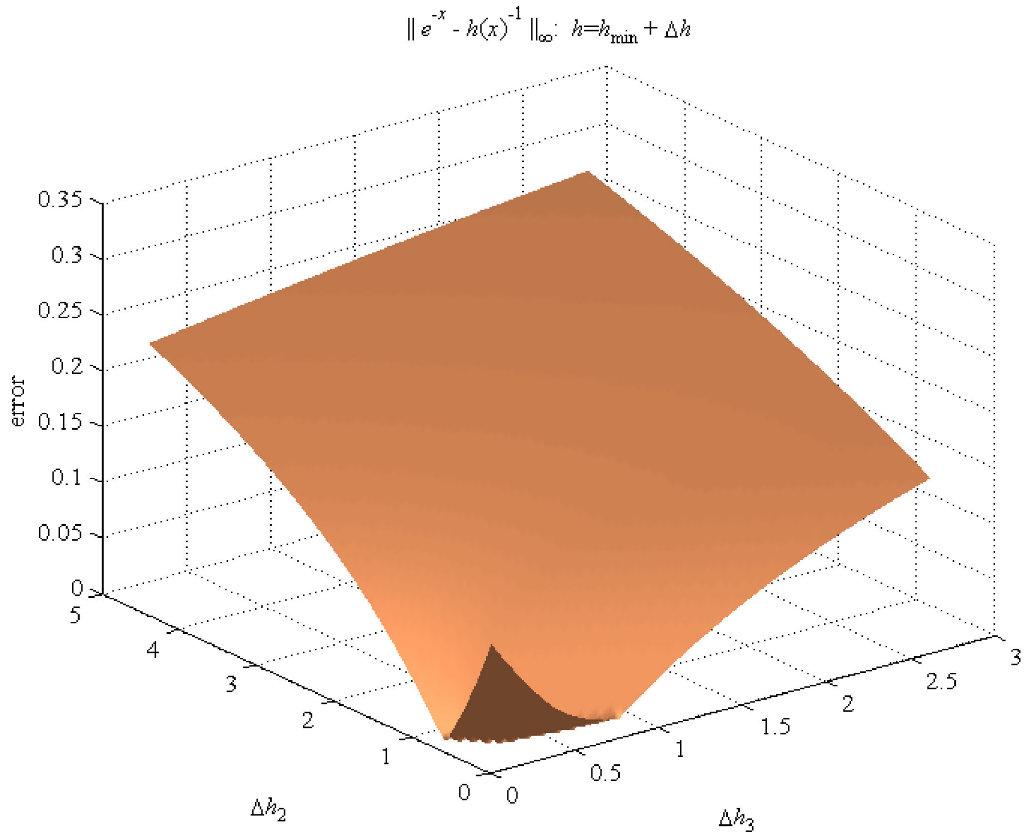

Figure 3: h_2-h_3 slice of the error surface at a local minimum.

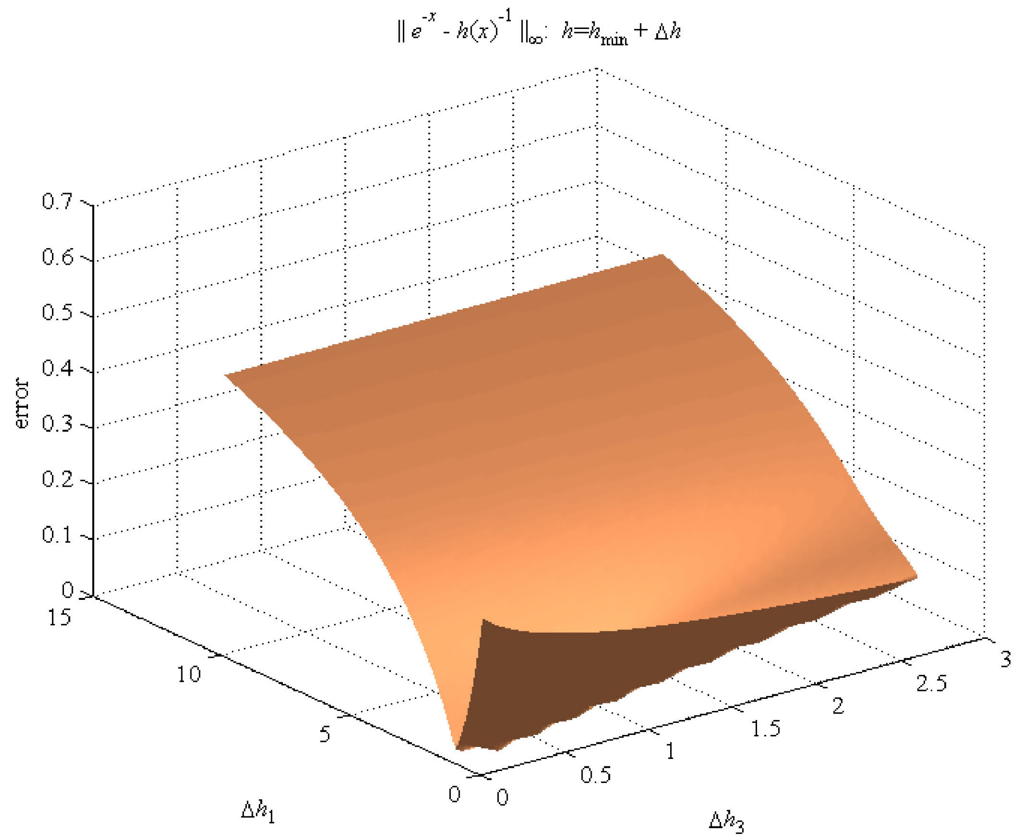

Figure 4: h_1-h_3 slice of the error surface at a local minimum.

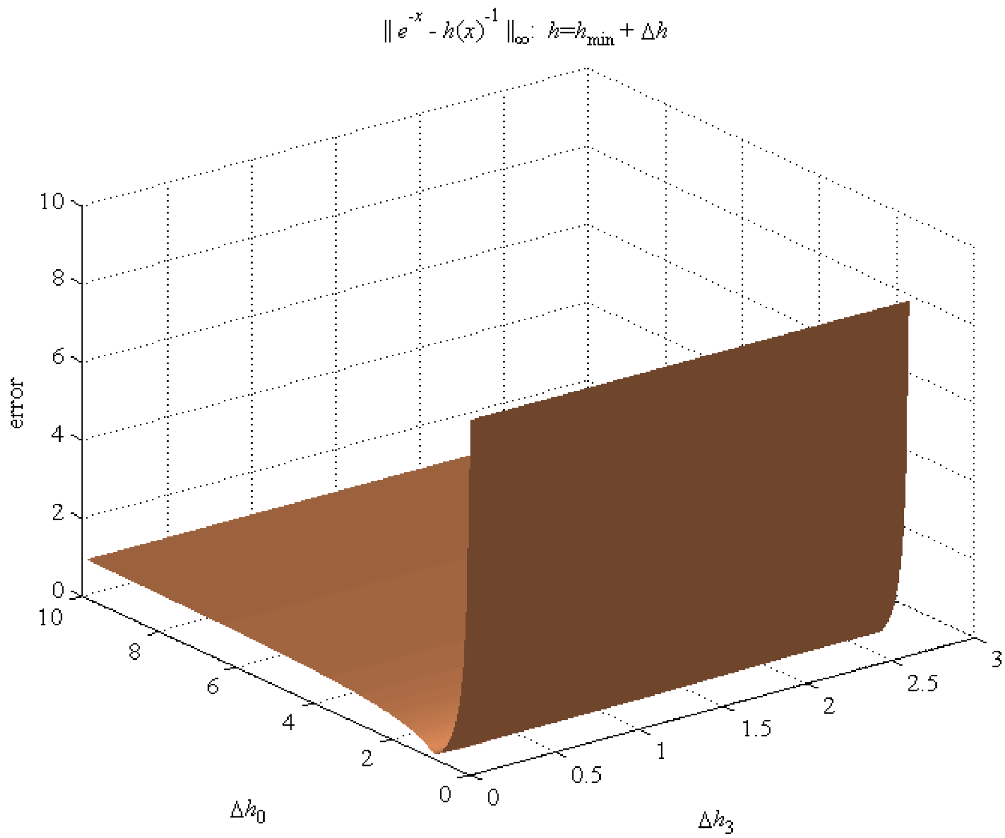

Figure 5: h_0-h_3 slice of the error surface at a local minimum.

4.2 Approximating Outer Functions

Computing outer functions is a common task in applied harmonic analysis [5] and signal processing [19]. The problem is to find a polynomial $h(z)$ that approximates a positive-valued function $q(z) \geq 0$ as follows:

$$q(z) \approx \exp(\Re[h(z)]) \qquad (z \in \mathbf{T}).$$

Although the general problem is on the unit circle, applications restrict $q(z)$ to be "real" [17, Eq. 1.1]:

$$q(z) = \overline{q(\overline{z})}.$$

Because this real symmetry is inherited by best approximations [21], [17], it suffices to approximate using polynomials with real coefficients. If $h(z)$ has real coefficients h_n, we can expand $h(z)$ as follows:

$$\begin{aligned}
\Re[h(z)] &= \sum_{n=0}^{N} h_n \Re[z^n] \qquad (z = e^{i\theta}) \\
&= \sum_{n=0}^{N} h_n \cos(n\theta) \\
&= \sum_{n=0}^{N} h_n T_n(x) \qquad (x = \cos(\theta)),
\end{aligned}$$

where T_n is the Chebyshev polynomial. Therefore, the real polynomial approximation of real functions on the unit circle is equivalent to approximation on the real interval $[-1, 1]$ by real polynomials. Consider the real outer function

$$g(z) = (z - a)^{-1}$$

with real pole a exterior to the unit disk. The magnitude of $g(z)$ on the unit circle is the target function:

$$q(z) = |g(z)|.$$

Although we are starting with the answer, any real $0 < q \in C(\mathbf{T})$ is approximated to arbitrary precision by $\exp(\Re[h(z)])$, where $h(z)$ is a real polynomial. Introduce the performance function

$$\Gamma(z, h(z)) = |q(z) - \exp(\Re[h(z)])|^2$$

and the objective function

$$\gamma(h) = \sup\{\Gamma(z, h(z)) : z \in \mathbf{T}\}.$$

The goal is to minimize the worst fit over the polynomials:

$$\min\{\gamma(h) : h \in \mathcal{P}^N\}.$$

To apply Corollary 1, switch to the real formalism:

$$\Gamma(z, h(z)) = (Q(x) - \exp(H(x)))^2 \qquad \begin{cases} Q(x) = q(z), \\ H(x) = \Re[h(z)], \\ x = \Re[z] = \cos(\theta) \end{cases} .$$

The variation of the performance function is

$$\partial_2\Gamma(x, h) = \partial_H(Q(x) - \exp(H))^2 = -2(Q(x) - \exp(H))\exp(H).$$

Because $\exp(H) > 0$, the following are equivalent:

- $\partial_2\Gamma(x, h(x))$ admits alternating sequence of length $N+2$ on $\mathrm{crit}[\gamma(h)]$;
- $Q(x) - \exp(H)$ admits alternating sequence of length $N+2$ on $\mathrm{crit}[\gamma(h)]$.

Figure 6 compares the outer function and an approximation from the polynomials of degree 6.

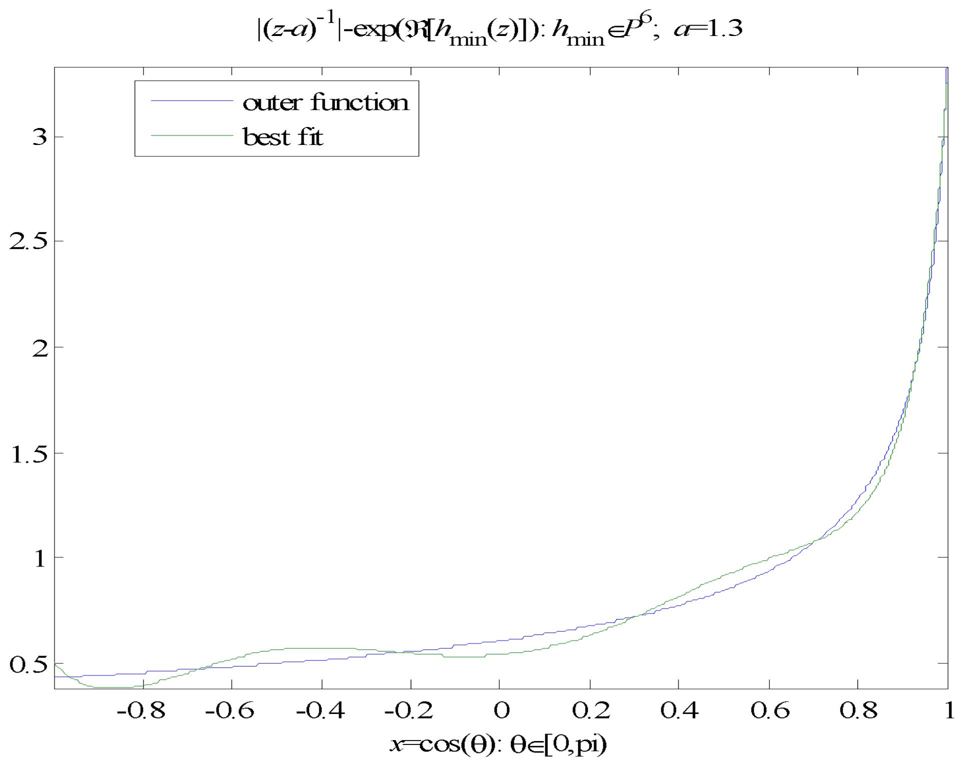

Figure 6: Outer function approximation.

Figure 7 displays the error. The right side of the plot lists the coefficients of $h_{\min}$. The solid red segments mark those $x \in [-1, 1]$ in the 90% neighborhood of crit[$\gamma(h)$]:

$$\Gamma(z, h_{\min}(z)) > 0.90 \times \gamma(h_{\min}).$$

The last two segments run together on the 90% neighborhood. Close examination of the plot shows that the error curve does alternate in sign eight times. Consequently, $h_{\min}$ is a nearly optimum minimizer.

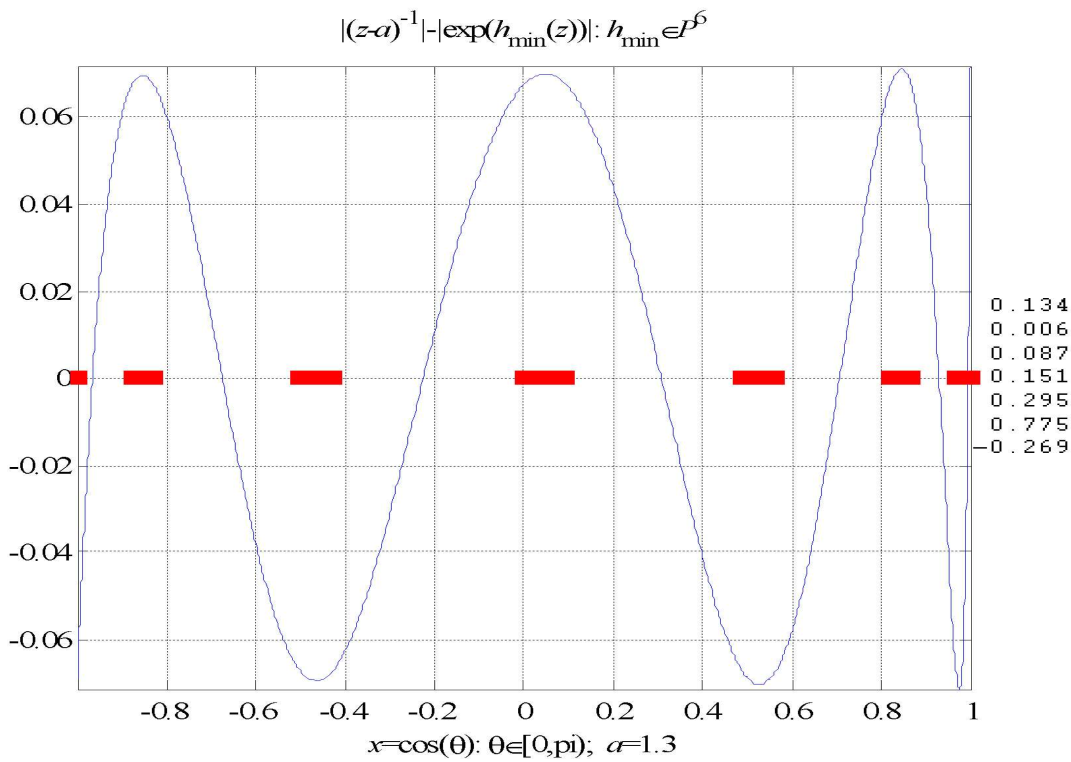

Figure 7: Outer function error.

5 Optimization on $\mathcal{A}(\mathbf{D})$

Helton and Merino [17] characterized disk algebra minimizers in their book. More importantly, their book discusses several computer programs that compute these minimizers. This section shows that the Kolmogorov Criterion also characterizes local minimizers of the disk algebra. The point is not to reinvent the results of Helton and Merino but to show that the Kolmogorov Criterion provide a general framework for optimization problems that also encompass the disk algebra. The disk algebra

$$\mathcal{A}(\mathbf{D}) := H^{\infty}(\mathbf{D}) \cap C(\mathbf{T}, \mathbb{C}) = 1 \dotplus z \dotplus z^2 \dotplus z^3 \dotplus \ldots$$

is essentially the space of polynomials on the unit disk. Analogous to the previous results on polynomial optimization, we will see that the support and alignment conditions readily follow from the Kolmogorov Criterion. Mergelyan's Theorem gives us our support condition.

Theorem 2 (Mergelyan) *[22, page 423] If K is a compact set in $\mathbb{C}$ with connected component, if $f \in C(K, \mathbb{C})$ is analytic on the interior of K, and if $\epsilon > 0$, there exists a polynomial $p(z)$ such that $\|f - p\|_{C(K,\mathbb{C})} < \epsilon$.*

If the critical set $\text{crit}[\gamma(h)]$ in not the entire unit circle $\mathbf{T}$, Mergelyan's Theorem forces the existence of a $\Delta h \in \mathcal{A}(\mathbf{D})$ that matches the performance function's variation:

$$\Delta h(z) = -\overline{\partial_2 \Gamma(z, h(z))} \quad (z \in \text{crit}[\gamma(h)]).$$

If the variation does not vanish on the critical set,

$$0 > \Re[\partial_2 \Gamma(z, h(z)) \Delta h(z)] \quad (z \in \text{crit}[\gamma(h)]),$$

the Minimum Test (Lemma 4) states that $h \in \mathcal{A}(\mathbf{D})$ cannot be a local minimum. Conversely, if $h \in \mathcal{A}(\mathbf{D})$ is a local minimum, the critical set is the entire unit circle.

Lemma 7 (Support) *Let U be an open subset containing $\mathbf{T}$. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^2. Define $\gamma : \mathcal{A}(\mathbf{D}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in \mathbf{T}\}.$$

Suppose $h \in \mathcal{A}(\mathbf{D})$ is a local minimum of γ. Assume $|\partial_2 \Gamma(z, h(z))| > 0$ on $\text{crit}[\gamma(h)]$. Then $\text{crit}[\gamma(h)] = \mathbf{T}$.

Proof: If $\text{crit}[\gamma(h)]$ is not the entire unit circle $\mathbf{T}$, the continuity of $\partial_2 \Gamma(z, h(z))$ permits an application of Mergelyan's Theorem: there exists a $\Delta h \in \mathcal{A}(\mathbf{D})$ such that

$$\Delta h(z) = -\overline{\partial_2 \Gamma(z, h(z))} \quad (z \in \text{crit}[\gamma(h)]).$$

Then $\Re[\partial_2\Gamma(z, h(z))\Delta h(z)] < 0$ on $\operatorname{crit}[\gamma(h)]$. The inequality is strict because $\partial_2\Gamma(z, h(z))$ is assumed never to vanish. The Minimum Test (Lemma 4) asserts that $h \in \mathcal{A}(\mathbf{D})$ cannot be a minimum of γ. This contradiction forces $\operatorname{crit}[\gamma(h)] = \mathbf{T}$. ///

Alignment is a little more tricky. The basic idea was pointed out in Section 3. A local minimum will force $\partial_2\Gamma(z, h(z))\Delta h(z)$ to wind around zero. For $w \in C(\mathbf{T}, \mathbb{C})$, let $\operatorname{Wind}[w(z), \mathbf{T}]$ algebraically count the number of times $w(z)$ winds around 0. The "alternating condition" of the real polynomials turns into a *positive winding number* at a local minimum:

$$\operatorname{Wind}[\partial_2\Gamma(z, h(z)), \mathbf{T}] > 0.$$

The trick is to link the phase of $\partial_2\Gamma(z, h(z))$ to elements of $\mathcal{A}(\mathbf{D})$. The Poisson integral is the starting point. For a complex Borel measure μ on $\mathbf{T}$, the harmonic extension of μ is its Poisson integral [22, page 252–255]:

$$P[\mu](z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Re\left[\frac{e^{it}+z}{e^{it}-z}\right] d\mu(t) \qquad (z \in \mathbf{D}).$$

Theorem 3 (Harmonic Extension) *[22, page 254] Let $h \in C(\mathbf{T}, \mathbb{C})$ and define $H[h]$ on $\overline{\mathbf{D}}$ by*

$$H[h](re^{i\theta}) = \begin{cases} h(e^{i\theta}) & r = 1 \\ P[h](re^{i\theta}) & r \in [0,1) \end{cases}.$$

Then $H[h] \in C(\overline{\mathbf{D}})$.

Thus, functions in $C(\mathbf{T}, \mathbb{C})$ admit harmonic extensions to $\mathbf{D}$ that are continuous on $\overline{\mathbf{D}}$. Closely related is the corresponding analytic extension.

Theorem 4 *[22, page 255] Suppose u is a* real-valued *function continuous on $\overline{\mathbf{D}}$ and harmonic on $\mathbf{D}$. Then (on $\mathbf{D}$) u is the Poisson integral of its restriction to $\mathbf{T}$ and the real part of the analytic function*

$$h(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{it}+z}{e^{it}-z} u(e^{it})dt \qquad (z \in \mathbf{D}).$$

Lemma 8 (Phase Alignment) *Let U be an open subset containing $\mathbf{T}$. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^2. Define $\gamma : \mathcal{A}(\mathbf{D}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in \mathbf{T}\}.$$

Suppose $h \in \mathcal{A}(\mathbf{D})$ is a local minimum of γ. Assume $|\partial_2\Gamma(z, h(z))| > 0$ on $\operatorname{crit}[\gamma(h)]$. Then $\operatorname{Wind}[\partial_2\Gamma(z, h(z)), \mathbf{T}] > 0$.

Proof: Suppose $-k = \text{Wind}[\partial_2\Gamma(z,h(z)),\mathbf{T}] \le 0$. Continuity gives that k is finite. Thus, the differential error term has phase like $\overline{z}^k$. Set

$$v(e^{i\theta}) = -\text{Arg}[z^k\partial_2\Gamma(z,h(z))](e^{i\theta}).$$

Then $v \in C(\mathbf{T})$. Use Theorem 3 to extend v to a real function continuous on $\overline{\mathbf{D}}$ and harmonic on $\mathbf{D}$. Use Theorem 4 to extend v as the imaginary part of a analytic function $g = u + iv$ on $\mathbf{D}$. For $0 < r < 1$, define $g_r(z) = g(rz)$ on $\mathbf{D}$. Then $g_r \in \mathcal{A}(\mathbf{D})$. Its imaginary part v_r converges to v as $r \to 1$. Set $\Delta h_r = \exp(g_r)$. Then Δh_r belongs to $\mathcal{A}(\mathbf{D})$ and so does $z^k\Delta h_r(z)$. Then as $r \to 1$ there holds

$$\Re\left[\partial_2\Gamma(z,h(z))z^k\Delta h_r(z)\right] = \Re\left[|\partial_2\Gamma(z,h(z))|\, e^{-iv(z)}|\Delta h(z)|e^{iv_r(z)}\right] > 0.$$

Then the Minimum Test (Lemma 4) asserts that h cannot be a minimum of γ. This contradiction forces the winding number to be strictly positive. ///

The upshot of Lemmas 7 and 8 is that if $h \in \mathcal{A}(\mathbf{D})$ is a local minimum:

- $\Gamma(z,h(z))$ is constant on $\mathbf{T}$ or $\text{crit}[\gamma(h)] = \mathbf{T}$.
- $\text{Wind}[\partial_2\Gamma(z,h(z)),\mathbf{T}] > 0$.

As with the polynomials, the disk algebra has enough structure to force a converse—provided the differential does not vanish. To see this, suppose $\text{crit}[\gamma(h)] = \mathbf{T}$ and $\text{Wind}[\partial_2\Gamma(z,h(z)),\mathbf{T}] > 0$ but $h \in \mathcal{A}(\mathbf{D})$ is **not** a local minimum. By the Descent Lemma, there exists a nonzero $\Delta h \in \mathcal{A}(\mathbf{D})$ such that

$$0 \ge \Re[\partial_2\Gamma(z,h(z))\Delta h(z)] \quad (z \in \mathbf{T}). \tag{1}$$

Figure 8 illustrates the geometry. For fixed $z \in \mathbf{T}$, the complex vector $\partial_2\Gamma(z,h(z))$ determines the solid half-plane consisting of all $\Delta h \in \mathbb{C}$ that satisfy Equation (1). Figure 8 also plots the conjugate $\overline{\partial_2\Gamma(z,h(z))}$. The plot shows Δh must belong to the negative cone

$$\overline{\partial_2\Gamma}(h)^{\ominus}(e^{i\theta}) := \{\mathbf{v} \in \mathbb{R}^2 : 0 \ge \overline{\partial_2\Gamma(e^{i\theta},h(e^{i\theta}))}^T\mathbf{v}\},$$

where we switch to real coordinates in the negative cone. Referring again to Figure 8, we see if $\partial_2\Gamma(z,h(z))$ winds positively around zero, the negative cone $\overline{\partial\Gamma}(h)^{\ominus}(z)$ must wind negatively around zero. Because Δh belongs to this cone that winds negatively around zero, Δh must also wind negatively around zero (provided it that never vanishes on $\mathbf{T}$). But Δh belongs to $\mathcal{A}(\mathbf{D})$, so Δh must have a non-negative winding number. These contradictory winding numbers for Δh imply that Δh cannot exist as a "direction of descent." Thus. the positive winding number of the differential forces $h \in \mathcal{A}(\mathbf{D})$ to be a local minimum. This is the geometric idea of the winding number. The technical part is to handle when Δh does have zeros on $\mathbf{T}$. The following result summarizes our Kolmogorov approach that captures a slightly weaker result than obtained by Helton and Merino in 1998.

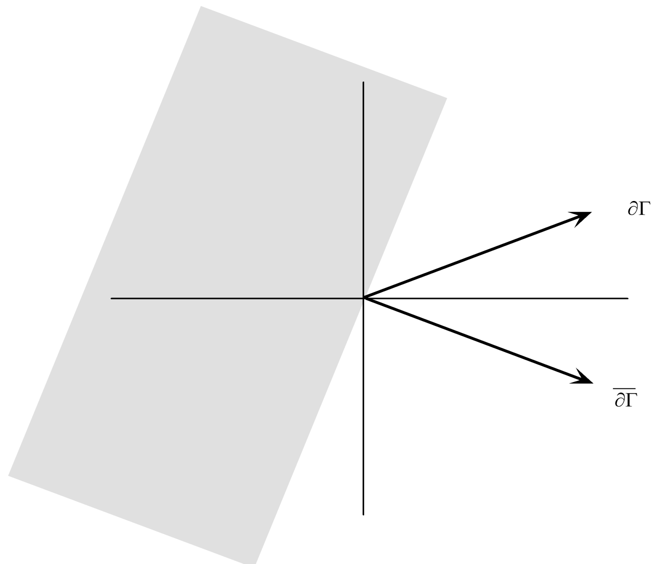

Figure 8: Solid half-plane marking $\Delta h \in \mathbb{C}$ for which $\Re[\partial_2\Gamma\Delta h] \leq 0$.

Corollary 2 *[17, Theorem 9.3.1] Let U be an open subset containing $\mathbf{T}$. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}$ be C^2. Define $\gamma : \mathcal{A}(\mathbf{D}) \to \mathbb{R}$ by*

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in \mathbf{T}\}.$$

Assume $|\partial_2\Gamma(z, h(z))| > 0$ on $\mathbf{T}$. Then the following are equivalent:

(a) *$h \in \mathcal{A}(\mathbf{D})$ is a local minimum of γ.*

(b) *$\Gamma(z, h(z))$ is constant on $\mathbf{T}$ and $\mathrm{Wind}[\partial_2\Gamma(z, h(z)), \mathbf{T}] > 0$.*

Proof: Lemmas 7 and 8 give that (a) implies (b). For the converse, assume (b) holds but that (a) is not true: that $h \in \mathcal{A}(\mathbf{D})$ is not a local minimum. The Descent Lemma (Lemma 3) provides a nonzero $\Delta h \in \mathcal{A}(\mathbf{D})$ such that $0 \geq \Re[\partial_2\Gamma(z, h(z))\Delta h(z)]$ on $\mathbf{T}$. Let

$$k = \mathrm{Wind}[\partial_2\Gamma(z, h(z)), \mathbf{T}] > 0.$$

Then Γ being C^2 with a nonzero variation permits us to write

$$\partial_2\Gamma(e^{i\theta}, h(e^{i\theta})) = |\partial_2\Gamma(e^{i\theta}, h(e^{i\theta}))| e^{ik\theta} e^{iv(e^{i\theta})},$$

where $v(e^{i\theta})$ is real and continuous. In terms of the inequality, there holds for all $z \in \mathbf{T}$:

$$0 \geq \Re[|\partial_2\Gamma(z, h(z))| e^{iv(z)} \Delta h(z)]. \tag{2}$$

Use Theorem 3 to extend v to a real function continuous on $\overline{\mathbf{D}}$ and harmonic on $\mathbf{D}$. Use Theorem 4 to extend v as the imaginary part of a analytic function $g = u + iv$ on $\mathbf{D}$. Observe $\exp(g)\Delta h \in H^\infty(\mathbf{D})$ so that

$$0 = \frac{1}{2\pi}\int_{\pi}^{\pi} e^{ik\theta} e^{g(e^{i\theta})} \Delta h(e^{i\theta}) d\theta.$$

Take the real part of both sides to get

$$0 = \frac{1}{2\pi}\int_{\pi}^{\pi} \Re[e^{ik\theta} e^{iv(e^{i\theta})} \Delta h(e^{i\theta})]\ e^{u(e^{i\theta})} d\theta.$$

Equation (2) gives that the "real part" of the integrand is negative so that

$$0 = \Re[e^{ik\theta} e^{iv(e^{i\theta})} \Delta h(e^{i\theta})] \quad \text{a.e.}$$

Continuity implies that equality holds everywhere. ///

5.1 Hyperbolic Approximation to $(z-a)^{-1}$

A canonical problem in H^∞ Engineering is computing the hyperbolic distance from the disk algebra to a given function [14]. The pseudo-hyperbolic distance[1] ρ between two elements g, $h \in \mathbf{D}$ is [27, page 58]:

$$\rho(g,h) := \left| \frac{g-h}{1-g\overline{h}} \right|. \tag{3}$$

Fix $g \in L^\infty(\mathbf{T})$ and assume $\|g\|_\infty < 1$. Let h vary over the disk algebra with $\|h\|_\infty < 1$. The pseudo-hyperbolic distance between $g(z)$ and $h(z)$ defines the performance function:

$$\Gamma(z, h(z)) := \left| \frac{g(z)-h(z)}{1-g(z)\overline{h(z)}} \right|^2; \quad (z = e^{i\theta}).$$

The objective function is

$$\gamma(h) := \sup\{\Gamma(z, h(z)) : z \in \mathbf{T}\}.$$

The minimization problem is

$$\inf\{\gamma(h) : h \in \mathcal{A}(\mathbf{D})\}.$$

[1] Also known as the Poincaré hyperbolic distance function and is a metric on $\mathbf{D}$ [26].

Corollary 2 characterizes local solutions by the flatness and winding conditions:

- $\Gamma(z, h_{\min}(z))$ is constant on $\mathbf{T}$,
- $\text{Wind}[\partial_2\Gamma(z, h(z)), \mathbf{T}] > 0$,

provided $g \in C^2$. For example, to make a function not in the disk algebra, put a pole at $0 \leq a < 1$ and set

$$g(z) = \frac{1}{2}\frac{1-a}{z-a}.$$

The scaling puts $g(z)$ into the unit ball of $L^\infty(\mathbf{T})$. Figure 9 shows the image of $g(z)$ in the unit disk. The goal is to approximate $g(z)$ from the disk algebra in the pseudo-hyperbolic metric.

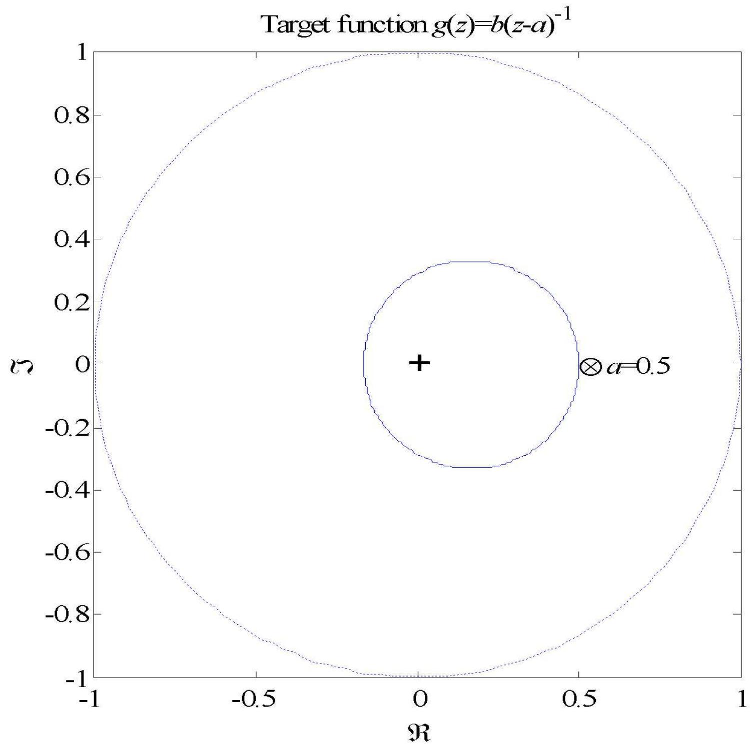

Figure 9: $g(z)$ for $z \in \mathbf{T}$.

The best approximation is a constant:

$$h_{\min} = 0.1883.$$

That is, $h_{\min}$ minimizes the pseudo-hyperbolic distance $g(z)$ to the disk algebra. Figure 10 plots the complex error curve:

$$\rho_{\mathbb{C}}(g(z), h_{\min}(z)) = \frac{g(z) - h_{\min}(z)}{1 - g(z)\overline{h_{\min}(z)}}.$$

The figure shows that the error is circular. Helton insightfully saw that this circularity of the error transplanted Nehari's Theorem from the complex plane to the disk equipped with the hyperbolic metric. The full power of Helton's insight becomes apparent when he extended this result to a nonlinear performance function—this error curve actually encodes a flatness condition and the winding number condition.

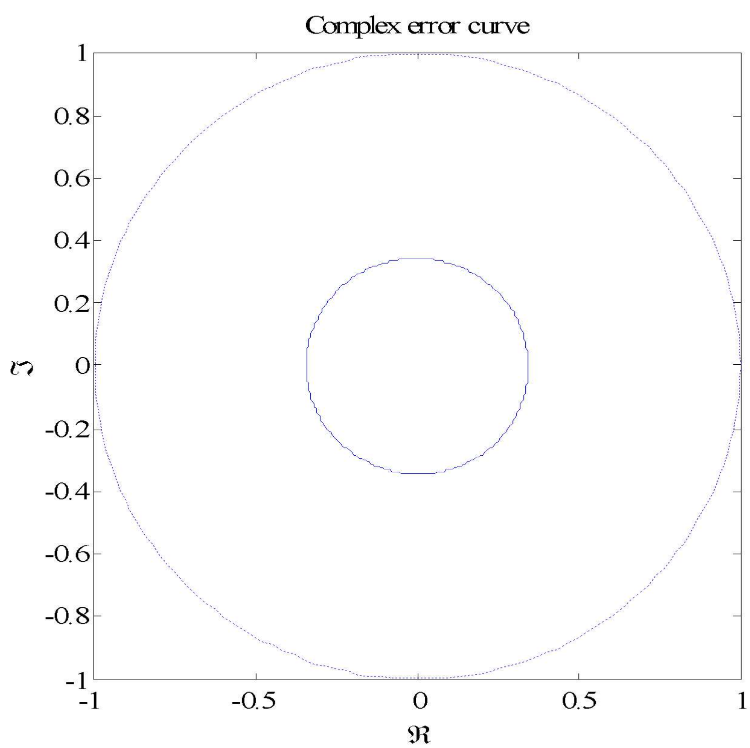

Figure 10: Complex error curve of $\rho_{\mathbb{C}}(g(z), h_{\min}(z))$ for $z = \exp(j\theta)$.

Figure 11 plots the error curve at the minimizer:

$$\Gamma(z, h_{\min}(z))^{1/2} := \left| \frac{g(z) - h_{\min}(z)}{1 - g(z)\overline{h_{\min}(z)}} \right|; \quad (z = e^{i\theta}).$$

Examination of the vertical axis shows that $\Gamma(z, h_{\min}(z))$ is numerically *flat*:

$$\Gamma(z, h_{\min}(z)) = \text{constant}.$$

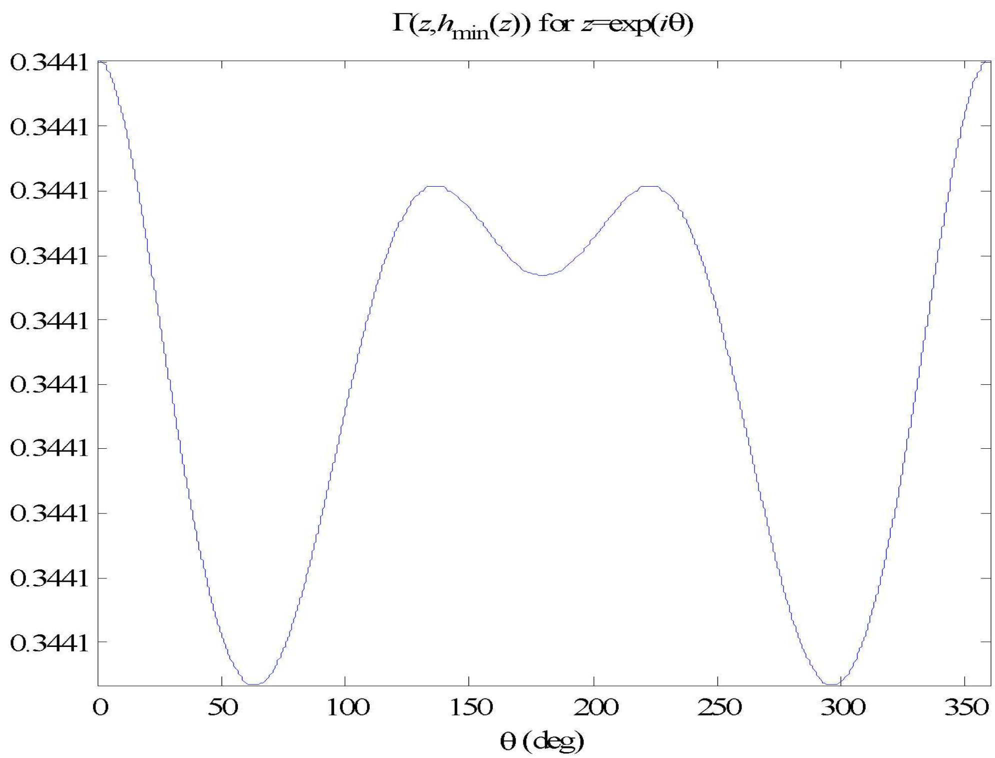

Figure 11: Flatness of $\Gamma(z, h_{\min}(z))$.

The differential of the performance function at the minimizer is

$$\partial_2\Gamma(z, h_{\min}(z)) = \frac{(\overline{h_{\min}(z)} - \overline{g(z)})(1 - \overline{g(z)}g(z))}{(1 - g(z)\overline{h_{\min}(z)})(1 - \overline{g(z)}h_{\min}(z))^2}.$$

Figure 12 plots this differential. The differential winds once around zero, which is the winding condition:

$$\mathrm{Wind}[\partial_2\Gamma(z, h_{\min}(z))] \geq 1.$$

As expected, the winding number of the differential is positive at the minimizer. What is unexpected is that the differential is also flat.

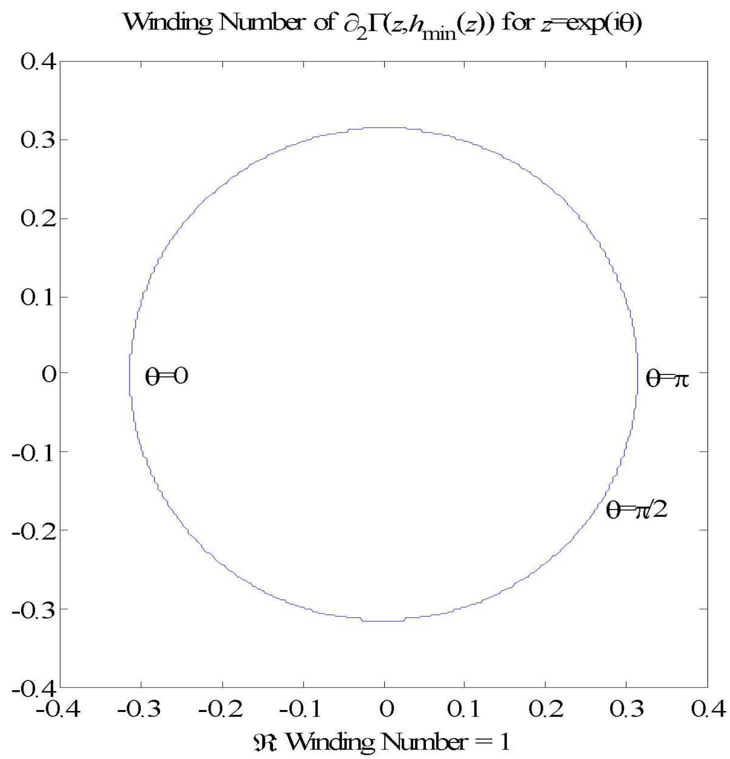

Figure 12: Differential $\partial_2\Gamma(z, h_{\min}(z))$ at the minimizer.

5.2 Helton's Example

Helton and Merino [17, page 142] offer a computer solution to the minimization problem:

$$\gamma_{\mathcal{A}} := \inf\{\gamma(h) : h \in \mathcal{A}(\mathbf{D})\}$$

on the disk algebra for the performance function

$$\Gamma(z, h(z)) = |0.8 + (z^{-1} + h(z))^2|^2. \tag{4}$$

The power of Helton and Merino's solution overcomes the infinite dimensional nature this minimization problem by using Nehari's Theorem. They estimate that

$$1.0005821 \approx \gamma_{\mathcal{A}} = \inf\{\gamma(h) : h \in \mathcal{A}(\mathbf{D})\}.$$

Our approach is absolutely pedestrian—simply approximate the disk algebra by the polynomials. This is a typical engineering approach because the engineer typically has only a finite number of parameters to synthesize a solution. This engineering approach becomes less pedestrian by comparing the suboptimal result against the best bound of Helton and Merino [17]. Benchmarking engineering solutions against the best bound is becoming common in impedance matching [13], [24], [4], [3]; amplifier optimization [14], [2]; and control problems [15], [17].

For example, Figure 13 plots the performance function of Equation (4) evaluated on the minimizer restricted to the polynomials of degree 11. The plot shows that this minimum is relatively close to the best bound:

$$1.0005821 \approx \gamma_{\mathcal{A}} < \inf\{\gamma(h) : h \in \mathcal{P}^{11}\} = 1.0149.$$

Whether this suboptimal solution is "good enough" is the decision that an engineer must make.

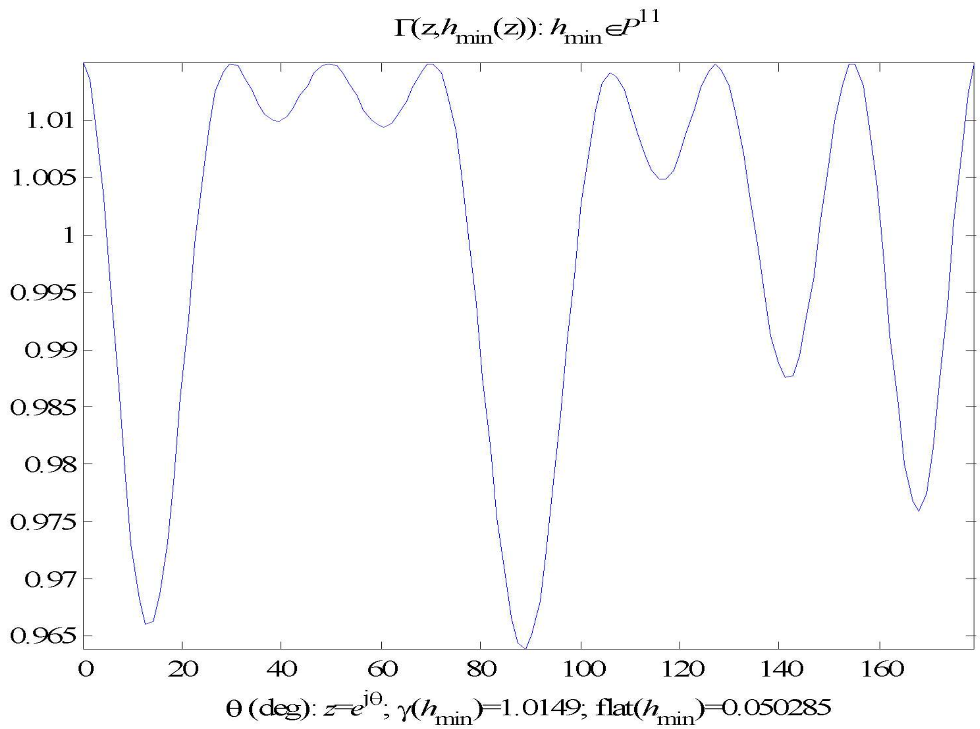

Figure 13: Flatness of the performance function.

For completeness, Figure 14 plots the variation of the performance function:

$$\partial_2\Gamma(z, h(z)) = 2(z^{-1} + h(z)))(0.8 + \overline{(z^{-1} + h(z))}^2).$$

The winding number of the variation is

$$\text{Wind}[\partial_2\Gamma(z, h_{\min}(z))] = 1$$

so that the alignment condition is satisfied. The relative flatness and alignment of $h_{\min}$ led us to suspect that $h_{\min}$ is close to the disk algebra minimizer

$$h_{\mathcal{A}} := \text{argmin}\{\gamma(h) : h \in \mathcal{A}(\mathbf{D})\}.$$

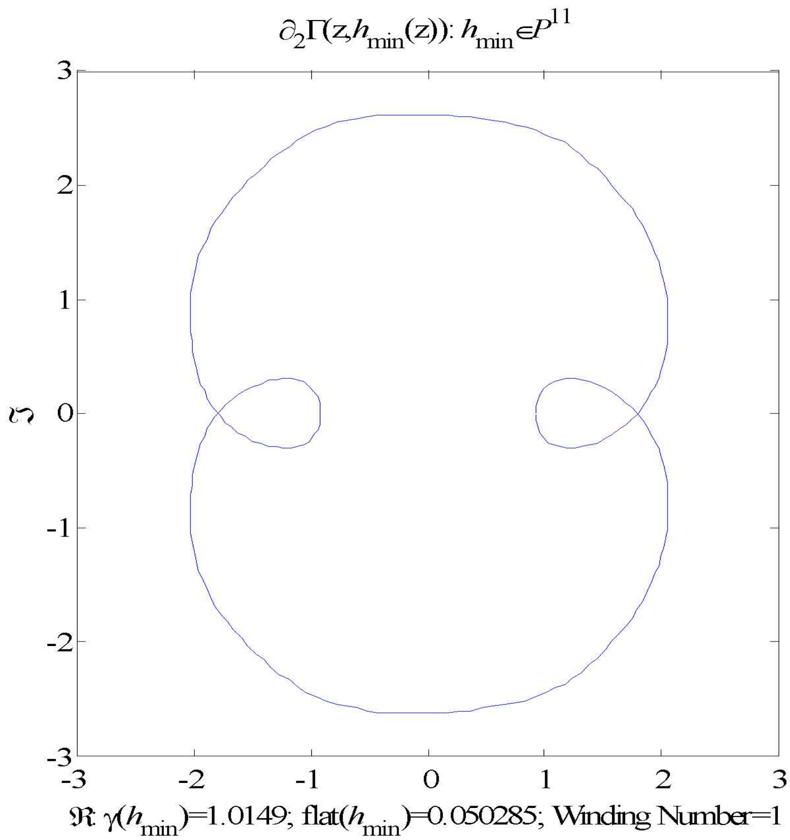

Figure 14: Winding Number of $\partial_2\Gamma(z, h_{\min}(z))$.

Moreover, the minimizers computed from increasing the degree of the polynomials should be "converging" to the disk algebra minimizer:

$$h_N = \operatorname{argmin}\{\gamma(h) : h \in \mathcal{P}^N\} \to h_{\mathcal{A}}.$$

Figure 15 exemplifies this belief by plotting the performance $\gamma(h_{\min,N})$ as a function of the degree N of the polynomials. The plot shows that near optimal performance is achieved on the polynomials of degree $N \geq 25$. Thus, knowing the best bound from the Helton-Merino computations provides the critical stopping point. Indeed, the minimum at $N = 29$ is starting to creep beneath the Helton-Merino bound. It is not that the Helton-Merino bound is incorrect—this creep is caused by over-interpolating on the finite samples of the unit circle [24]. Nevertheless, Figure 15 graphically raises the fundamental question:

Question 1 *How do the finite-dimensional but realizable minimizers h_N approximate the disk-algebra minimizer $h_{\mathcal{A}}$?*

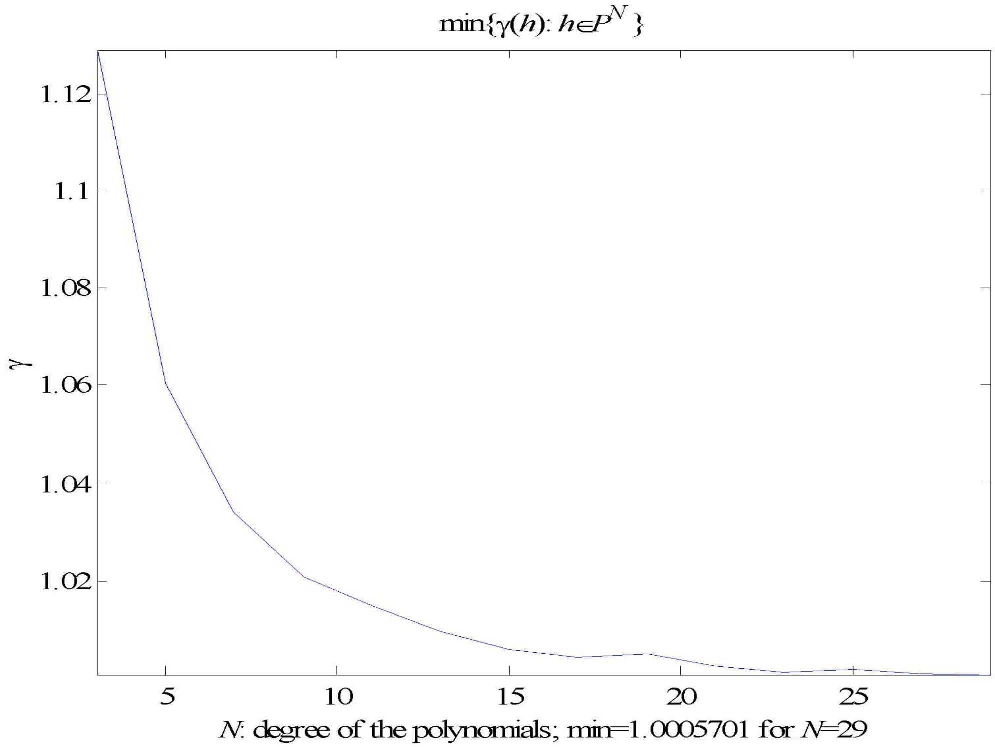

Figure 15: Performance as a function of the degree of the polynomials.

In particular, relating the rate of converge of $h_N \to h_{\mathcal{A}}$ to the smoothness of $\Gamma(z, h(z))$ offers a fascinating research opportunity to insert extrapolation techniques into H^∞ Theory. Likewise, the flatness and winding number conditions offer additional measurements of the quality of a suboptimal solution, which raises classic question regarding a suboptimal solution:

$$\gamma(h_{\Delta\gamma}) \leq \gamma_{\mathcal{A}} + \Delta\gamma.$$

Assuming convergence does happen, Question 2 asks:

Question 2 *How fast does $h_{\Delta\gamma}$ converge to $h_{\mathcal{A}}$ as $\Delta\gamma \to 0$?*

However, the far more useful question is far more difficult, particularly when $h_{\Delta\gamma}$ is known only on a finite number of points on the unit circle:

Question 3 *Suppose $\{z_k\}$ is a dense sampling of the unit circle;*

$$z_k = e^{jk/K} \quad (k = 0, \ldots, K-1),$$

where $K \gg 1$. Assume on this sampling of the unit circle,

$$\Gamma(z_k, h_{\Delta\gamma}(z_k)) \leq \gamma_{\mathcal{A}} + \Delta\gamma.$$

How far is $h_{\Delta\gamma}$ from $h_{\mathcal{A}}$?

These questions are the standard ones. Helton and Merino [17, page 141] exploit the flatness condition to measure the quality of a suboptimal solution

$$\mathrm{flat}(h_{\min}) := 1 - \frac{\sup\{\Gamma(z, h_{\min}(z))\}}{\inf\{\Gamma(z, h_{\min}(z))\}}.$$

As the flatness tends to zero, the performance function tends to a constant value. So, the performance of a suboptimal solution and its flatness are multiple criteria for the quality of this numerical solution.

For example, Figure 16 displays the performance of a suboptimal solution from the polynomials of degree 11. The figure shows a worse performance than reported from Figure 13, but better flatness.

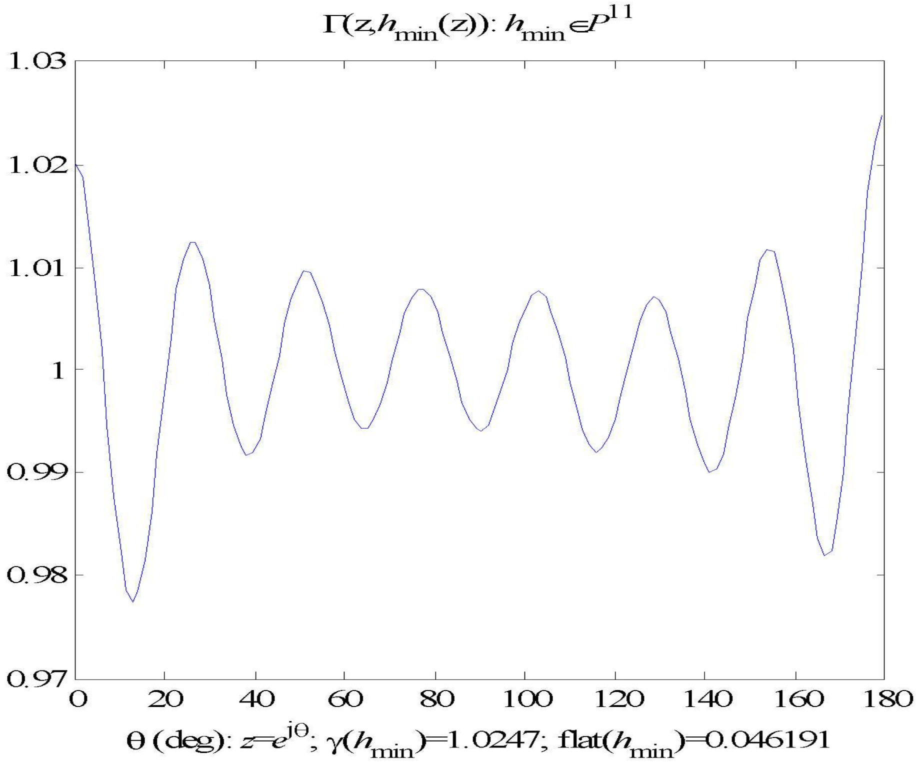

Figure 16: Flatness of the performance function.

Table 2 summarizes the performance of the suboptimal solutions from the polynomials of degree 11. The table shows that one solution attains a smaller objective but worse flatness. Consequently, the engineer can trade off the objective function against the flatness function. The formal mathematical approach to these engineering trade-offs is multiobjective optimization.

Table 2: Assessing suboptimal solutions from $\mathcal{P}^{11}$.

	Table 12.1 [17]	Figure 13	Figure 16
Performance	1.000582	1.0149	1.0274
Flatness	0.0020	0.0503	0.0462

6 Multiobjective Optimization

Multiobjective optimization is a powerful tool to trade off competing objectives. The objective functions are stacked in a vector and vector-valued optimization is undertaken. The beauty of this approach is that the impossible problems of simultaneously rationalizing units and adjusting scaling factors is avoided. Introduce the *partial order* on $\mathbb{R}^N$ by declaring

$$\mathbf{u} \leq \mathbf{v} \quad \Longleftrightarrow \mathbf{v} - \mathbf{u} \in \mathbb{R}_+^N,$$

where $\mathbb{R}_+^N$ denotes the closed positive orthant

$$\mathbb{R}_+^N := \{\mathbf{y} \in \mathbb{R}^N : y_n \geq 0\}.$$

Let $\gamma : X \subseteq \mathbb{R}^M \to \mathbb{R}^N$ denote the mapping

$$\gamma(\mathbf{x}) := \begin{bmatrix} \gamma_1(\mathbf{x}) \\ \gamma_2(\mathbf{x}) \\ \vdots \\ \gamma_M(\mathbf{x}) \end{bmatrix}.$$

Each γ_m is called an *objective function* so that γ is called a *multiobjective function*. We want to solve the vector-valued minimization of γ on X. Boyd and Vandenberghe [6, page 20] have generalized the notion of a "minimizer." Denote the *image of X under γ by*

$$\gamma(X) := \{\gamma(\mathbf{x}) : \mathbf{x} \in X\}.$$

Any $\gamma(\mathbf{x}) \in \gamma(X)$ is called a *minimum element* of $\gamma(X)$, provided

$$\gamma(\mathbf{x}) \leq \gamma(\mathbf{x}')$$

for all $\mathbf{x}' \in X$. A convenient notation for this inequality between a point $\gamma(\mathbf{x})$ and the set of all the $\gamma(\mathbf{x}')$'s is

$$\gamma(\mathbf{x}) \leq \gamma(X).$$

The key to a minimal element is that the inequality holds on all the image $\gamma(X)$. Equivalently, this inequality states that attaching the positive orthant at $\gamma(\mathbf{x})$ will subsume all the elements in $\gamma(X)$:

$$\gamma(X) \subseteq \gamma(\mathbf{x}) + \mathbb{R}_+^N.$$

Figure 17 illustrates the geometry of the minimum element in $\mathbb{R}^2$ and offers a strong geometric proof that the minimum element is unique.

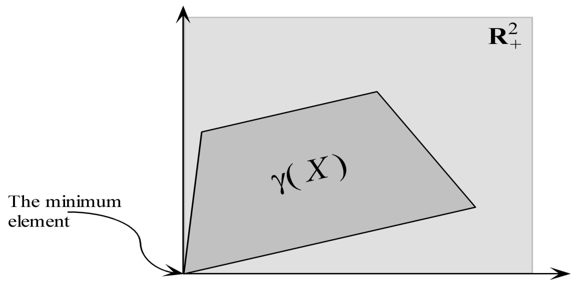

Figure 17: The minimum element.

Not all sets admit a minimum element. More commonly, we look for *minimal elements* as illustrated in Figure 18. Any $\gamma(\mathbf{x}) \in \gamma(X)$ is a minimal element of $\gamma(X)$, provided [6, page 21]:

$$\gamma(\mathbf{y}) \leq \gamma(\mathbf{x}) \implies \gamma(\mathbf{y}) = \gamma(\mathbf{x}).$$

Figure 18 shows this is equivalent[2] to

$$\left(\gamma(\mathbf{x}) - \mathbb{R}^N_+\right) \bigcap \gamma(X) = \{\gamma(\mathbf{x})\}.$$

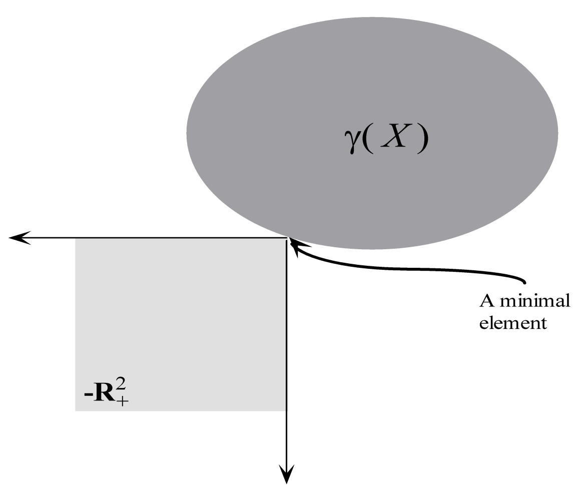

Figure 18: A minimal element.

[2]More restrictive is the notion of *weak minimizers* [10]: $(\gamma(\mathbf{x}) - \mathrm{int}[\mathbb{R}^N_+]) \cap \gamma(X) = \emptyset$.

These definitions occur in the range of $\gamma : X \subseteq \mathbb{R}^M \to \mathbb{R}^N$. In the domain of γ, any $\mathbf{x} \in X$ is called *Pareto optimal*, provided $\gamma(\mathbf{x})$ is a minimal element of $\gamma(X)$ [6, page 102]. Figure 19 illustrates all minimal elements, or the images of the Pareto optima, as the dark line on the boundary of $\gamma(X)$. Regardless of the shape of $\gamma(X)$, finding its Pareto set is fundamental. From Das and Dennis [12]:

> "Very often in engineering applications, the desired result helpful in facilitating design is a whole collection of Pareto optimal points, representative of the entire spectrum of efficient solutions. Thus, ideally, the desired solution is the entire Pareto optimal set."

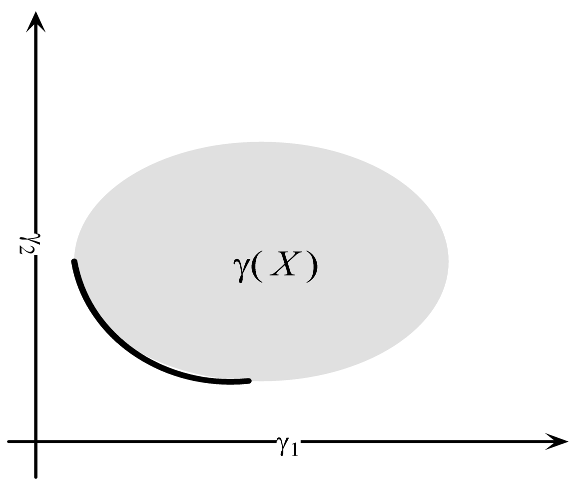

Figure 19: The minimal elements.

In summary,

Computing all Pareto optima is the Fundamental Goal of Multiobjective Optimization.

Of the many multiobjective optimization schemes, the *Goal Attainment Method* is well-suited for a wide range of applications. Figure 20 illustrates the method. The user specifies a vector of *design goals* γ_u such that

$$\gamma_u \leq \gamma(X)$$

and a vector of non-negative *weights* $\mathbf{w}$. The minimizer attempts to shoot from γ_u along the direction of the weight vector $\mathbf{w}$ and hit the boundary of $\gamma(X)$. The stopping point, if it exists, may be a minimal element of $\gamma(X)$.

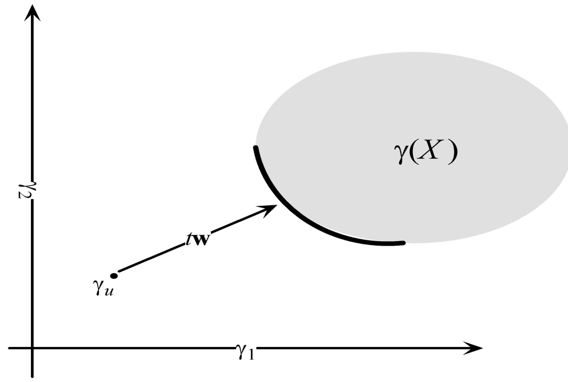

Figure 20: The Goal Attainment Method.

The Goal Attainment Method [8]. Given $\gamma : X \subseteq \mathbb{R}^M \to \mathbb{R}^N$. Select a design goal $\gamma_u \leq \gamma(X)$. Select a weight vector $\mathbf{w} \in \mathbb{R}_+^N$.

$$\text{minimize}\{t \in \mathbb{R}\}$$

subject to $\mathbf{x} \in X$ and

$$\gamma(\mathbf{x}) - t\mathbf{w} \leq \gamma_u.$$

Das and Dennis [12] introduce the normal-bound intersection (NBI) method for computing the Pareto set using the *global* minimizers:

$$\mathbf{x}_n = \text{argmin}\{\gamma_n(\mathbf{x}) : \mathbf{x} \in X\} \quad (n = 1, \dots, N).$$

These minimizers determine the *utopic point*

$$\gamma_\diamond := \begin{bmatrix} \gamma_1(\mathbf{x}_1) \\ \gamma_2(\mathbf{x}_2) \\ \vdots \\ \gamma_N(\mathbf{x}_N) \end{bmatrix}$$

that is a pseudo-minimum of $\gamma(X)$. Figure 21 shows that the utopic point is within the "line of sight" of the Pareto points by shooting along the weight vector $\mathbf{w} \geq 0$. The claim is that by setting $\gamma_\diamond = \gamma_u$ and varying the weight vector $\mathbf{w} \geq 0$, a superset of the Pareto set can be computed. In practice, we can only sample this superset. Das and Dennis [12] point out that this sampling may not be uniformly distributed. Their key claim is that the NBI method produces a uniform sampling of the Pareto set. Thus, practical questions can be raised about the efficacy of the multiobjective minimizers.

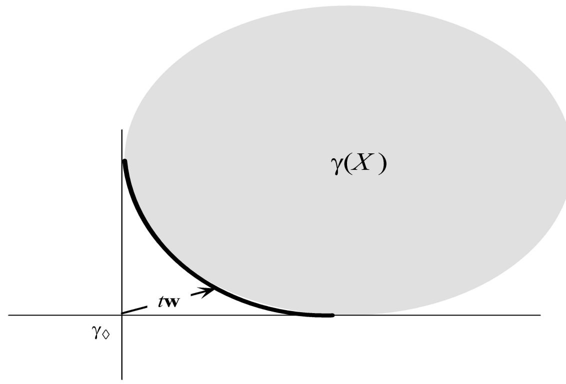

Figure 21: The utopic point $\gamma_\diamond$.

Figure 18 shows that Pareto optimal is a global definition. All of the image

$$\gamma(\mathcal{H}) = \{\gamma(\mathbf{x}) : \mathbf{x} \in X\}$$

must be tested. A local notion of Pareto optimal will be needed.

Definition 1 *An element $h_p \in \mathcal{H}$ is called locally Pareto-optimal, provided a neighborhood U of h_p exists such that for all $h \in U$*

$$\gamma(h) \leq \gamma(h_p) \implies \gamma(h) = \gamma(h_p).$$

Equivalently, $h_p \in \mathcal{H}$ is **not** locally Pareto-optimal if a sequence $\Delta h_k \in \mathcal{H}$ exists that converges to zero and satisfies

$$\gamma_m(h + \Delta h_k) \leq \gamma_m(h_p) \quad (m = 1, \ldots, M),$$

with strict inequality for at least one index. The image of a Pareto point lies on the boundary of $\gamma(\mathcal{H})$. The image of a local Pareto point may actually lie in the interior of $\gamma(\mathcal{H})$.

7 The Multiobjective Kolmogorov Criterion

The Kolmogorov Criterion generalizes to the multiobjective problem. The first result is that a direction of descent exists at points that are not locally Pareto-optimal.

Lemma 9 (Multiobjective Descent) *Let $Z \subset \mathbb{C}$ be compact. Let $\mathcal{H}$ be a closed linear subspace of $C(Z, \mathbb{C})$. Let U be an open subset containing Z. Let $\Gamma : U \times \mathbb{C} \to$*

$\mathbb{R}^M$ *be* C^2*. Define* $\gamma : \mathcal{H} \to \mathbb{R}^M$ *by*

$$\gamma(h) := \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \\ \vdots \\ \gamma_M(h) \end{bmatrix} \qquad \gamma_m(h) := \sup\{\Gamma_m(z, h(z)) : z \in Z\}.$$

Assume that $\mathcal{H}$ *is boundedly compact. If* $h \in \mathcal{H}$ *is not locally Pareto-optimal, a nonzero* $\Delta h \in \mathcal{H}$ *exists such that*

$$0 \geq \Re[\partial_2 \Gamma_m(x, h(x))\Delta h(x)] \quad (x \in \mathrm{crit}[\gamma_m(h)])$$

for $m = 1, \ldots, M$*.*

Proof: If $h \in \mathcal{H}$ is not locally Pareto-optimal, a sequence $\{\Delta h_k\} \in \mathcal{H}$ exists that converges to zero for which

$$\gamma_m(h + \Delta h_k) \leq \gamma_m(h)$$

with strict inequality in at least one index. Let $t_k := \|\Delta h_k\|_\infty$ and set $u_k := t_k^{-1} \Delta h_k$. By selecting a subsequence, the bounded compactness of $\mathcal{H}$ asserts the existence of a limit point: $u_k \to \Delta h \in \mathcal{H}$. By construction, Δh is nonzero. For all $x \in \mathrm{crit}[\gamma_m(h)]$, Lemma 2 provides the expansion:

$$\begin{aligned} \gamma_m(h + \Delta h_k) &\geq \Gamma_m(x, h(x) + \Delta h_k(x)) \\ &= \Gamma_m(x, h(x)) + \Re[\partial_2 \Gamma_m(x, h(x))\Delta h_k(x)] + \mathcal{O}[t_k^2] \\ &= \gamma_m(h) + \Re[\partial_2 \Gamma_m(x, h(x))\Delta h_k(x)] + \mathcal{O}[t_k^2]. \end{aligned}$$

Subtract $\gamma_m(h)$ from both sides, then divide by $t_k > 0$ to get

$$0 \geq \Re[\partial_2 \Gamma_m(x, h(x)) u_k(x)] + \mathcal{O}[t_k].$$

Taking the limit as $k \to \infty$ gives

$$0 \geq \Re[\partial_2 \Gamma_m(x, h(x))\Delta h(x)] \quad (x \in \mathrm{crit}[\gamma_m(h)])$$

for $m = 1, \ldots, M$. ///

Roughly speaking, this lemma provides a “candidate” for a direction of descent at those points not locally Pareto-optimal. The quotes are used because the linearization may have enough information to dominate the function in the neighborhood of the point. If the derivative does not vanish, this problem is eliminated and the following Multiobjective Minimization Test is available.

Lemma 10 (Multiobjective Minimization Test) *Let $Z \subset \mathbb{C}$ be compact. Let $\mathcal{H}$ be a closed linear subspace of $C(Z, \mathbb{C})$. Let U be an open subset containing Z. Let $\Gamma : U \times \mathbb{C} \to \mathbb{R}^M$ be C^2. Define $\gamma : \mathcal{H} \to \mathbb{R}$ by*

$$\gamma(h) := \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \\ \vdots \\ \gamma_M(h) \end{bmatrix} \qquad \gamma_m(h) := \sup\{\Gamma_m(z, h(z)) : z \in Z\}.$$

Let $h \in \mathcal{H}$. If there exists a $\Delta h \in \mathcal{H}$ such that

$$0 > \Re[\partial_2 \Gamma_m(x, h(x)) \Delta h(x)] \qquad (x \in \mathrm{crit}[\gamma_m(h)]),$$

for $m = 1, \ldots, M$, then $h \in \mathcal{H}$ is not locally Pareto-optimum for γ.

Examination of both results reveals a new phenomenon. For clarity, consider the real-valued case in $C([0,1], \mathbb{R})$. Suppose $h \in \mathcal{H}$ admits a direction of descent $\Delta h \in \mathcal{H}$. The MultiDescent Lemma (Lemma 9) forces

$$0 \geq \partial_2 \Gamma_m(x, h(x)) \Delta h(x) \quad x \in \mathrm{crit}[\gamma_m(h)]$$

for $m = 1, \ldots, M$. What if two critical sets share a common element? Specifically, suppose $x_\pm \in \mathrm{crit}[\gamma_{m_1}(h)] \cap \mathrm{crit}[\gamma_{m_2}(h)]$ with differing signs:

$$0 > \partial_2 \Gamma_{m_1}(x_\pm, h(x_\pm)) \text{ and } 0 < \partial_2 \Gamma_{m_2}(x_\pm, h(x_\pm)).$$

This forces $\Delta h(x_\pm) = 0$. This *phase splitting of the differential* requires some consideration and is best approached through an example on the polynomials.

8 Multiobjective Optimization on $\mathcal{P}^N$

Phase splitting forces additional constraints that depend on local smoothness. Examining a few examples is worthwhile before setting out a general theory.

8.1 Approximating $\exp(\pm x)$

The problem is simultaneous polynomial approximation: Find a polynomial

$$h(x) = h_0 + h_1 x + h_2 x^2 + h_3 x^3$$

that fits the exponential function and its reciprocal:

$$\begin{aligned} e^x &\approx h(x) \\ e^{-x} &\approx h(x)^{-1}. \end{aligned}$$

Choose the performance functions as follows:

$$\begin{aligned} \Gamma_1(x, h(x)) &= (e^x - h(x))^2 \\ \Gamma_2(x, h(x)) &= (e^{-x} - h(x)^{-1})^2. \end{aligned}$$

The objective functions are

$$\begin{aligned} \gamma_1(h) &:= \sup\{\Gamma_1(x, h(x)) : x \in [0,1]\} \\ \gamma_2(h) &:= \sup\{\Gamma_2(x, h(x)) : x \in [0,1]\}. \end{aligned}$$

The goal is to minimize the multiobjective function

$$\gamma(h) := \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \end{bmatrix}$$

on the nonlinear subset $\mathcal{H} \subset \mathcal{P}^3$ consisting of those polynomials that never vanish on the unit interval.

Figure 22 sketches out $\gamma(\mathcal{H})$. The blue dots plot the value of $\gamma(h)$ on random polynomials $h \in \mathcal{H}$. The red square is the starting point for the numerical minimizing method. The diamond marks where the method terminated.

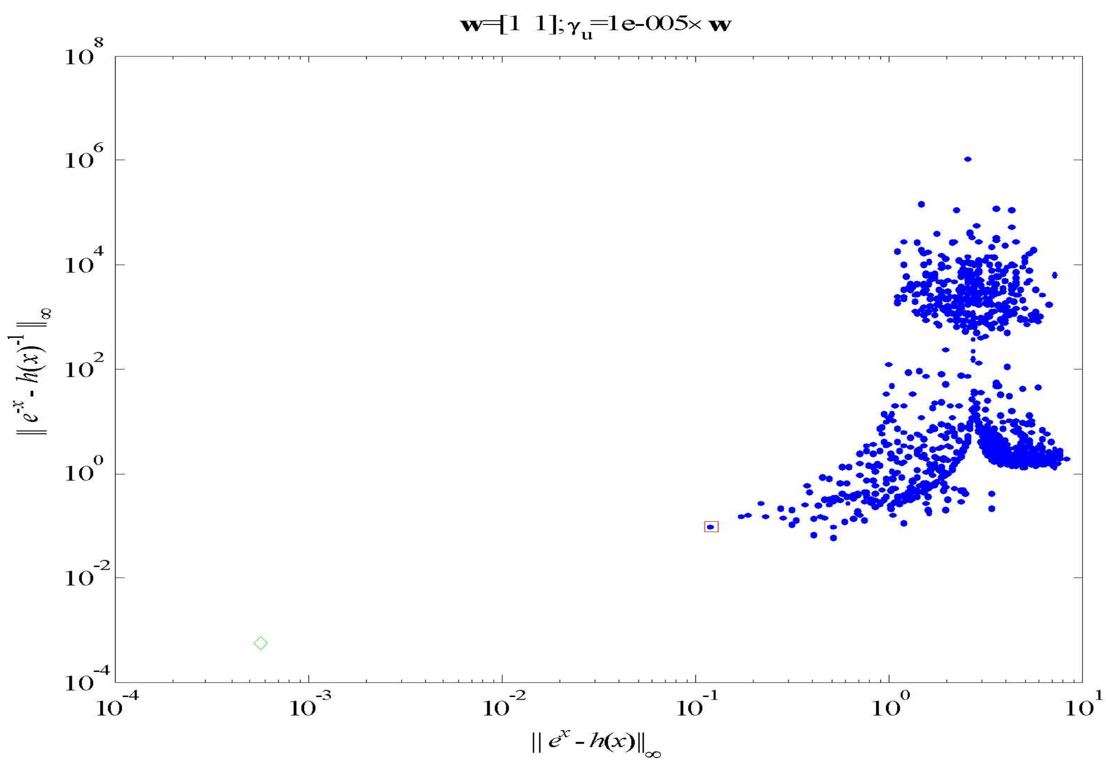

Figure 22: Random values of $\gamma(h)$.

The minimizer was computed using the Goal Attainment method with equal weights and goals:

$$\gamma(h) - t\mathbf{w} \le \gamma_u; \quad \mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \gamma_u = 10^{-5} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Figure 23 plots the error curves at this numerical minimizer. To understand this plot, compute the partials of each performance function:

$$\begin{aligned} \partial_2 \Gamma_1(x, h(x)) &= -2(e^x - h(x)) \\ \partial_2 \Gamma_2(x, h(x)) &= +2(e^{-x} - h(x)^{-1})h(x)^{-2}. \end{aligned}$$

A minimizer of $\gamma_1(h)$ on $\mathcal{H} \subset \mathcal{P}^3$ is characterized whenever $\partial_2 \Gamma_1(x, h(x))$ exhibits an alternation sequence of length 5, which forces the error function $-(e^x - h(x))$ to alternate five times. The upper panel of Figure 23 shows that this alternation—$\gamma_1(h_\diamond)$ is at its minimal value. Because of the unicity of best polynomial approximations, any nonzero perturbation of $h_\diamond$ degrades the performance γ_1 [11]:

$$0 \neq \Delta h \in \mathcal{P}^3 \implies \gamma_1(h_\diamond) < \gamma_1(h_\diamond + \Delta h).$$

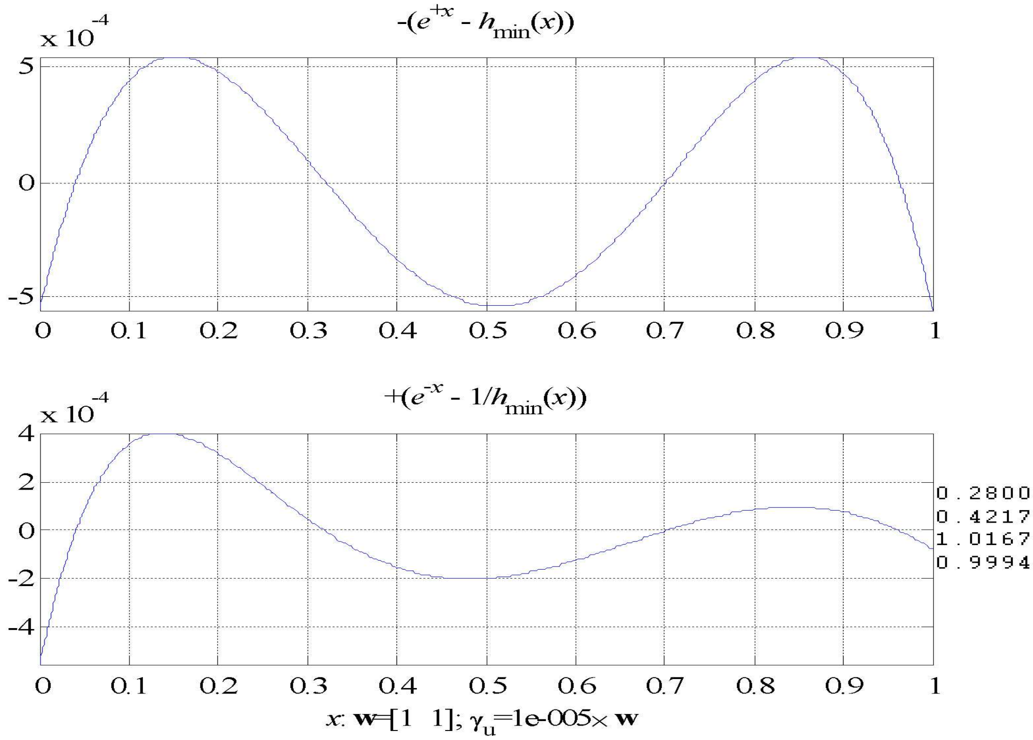

Figure 23: Error curve at local minimum $h_\diamond$.

Consequently, any nonzero perturbation of $h_\diamond$ that improves the second objective:

$$\gamma_2(h_\diamond + \Delta h) < \gamma_2(h_\diamond)$$

must degrade the first objective, that is, $\gamma(h_\diamond)$ is a minimal element of $\gamma(\mathcal{H})$ and $h_\diamond$ is Pareto optimal. Likewise, a minimizer of $\gamma_2(h)$ on $\mathcal{H} \subset \mathcal{P}^3$ is also characterized when the error function $(e^{-x} - h(x)^{-1})$ alternates five times. The lower panel of Figure 23 shows that $\gamma_2(h_\diamond)$ has only one critical point—at the endpoint of the unit interval. Although not an example of phase splitting, the figure does show that the critical sets of the individual objective functions can easily have common elements.

8.2 Characterization

For multiobjective optimization on the polynomials, the alternating condition now expands to include all the critical sets of the objective functions while phase splitting forces zeros into the "tangent space" of the objective function. In the polynomials, the alternation and phase splitting balance out. Define

$$\mathrm{crit}[\gamma(h)] = \bigcup_{m=1}^{M} \mathrm{crit}[\gamma_m(h)].$$

Let $\mathrm{crit}_\pm[\gamma(h)]$ denote those critical points for which the differential phase splits. Formally, $x_\pm \in \mathrm{crit}_\pm[\gamma(h)]$ provided $x_\pm \in \mathrm{crit}[\gamma_{m_1}(h)] \cap \mathrm{crit}[\gamma_{m_2}(h)]$ and

$$0 > \partial_2\Gamma_{m_1}(x_\pm, h(x_\pm))\partial_2\Gamma_{m_2}(x_\pm, h(x_\pm)).$$

Lemma 11 *Suppose* $\Gamma : [0,1] \times \mathbb{R} \to \mathbb{R}^M$ *is* C^2*. Define* $\gamma : C([0,1], \mathbb{R}) \to \mathbb{R}^M$ *by*

$$\gamma(h) := \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \\ \vdots \\ \gamma_M(h) \end{bmatrix} \qquad \gamma_m(h) := \sup\{\Gamma_m(x, h(x)) : x \in [0,1]\}.$$

Let $h \in \mathcal{P}^N$*. Assume* $\partial_2\Gamma_m(h) \neq 0$ *on* $\mathrm{crit}[\gamma_m(h)]$*. On* $\mathrm{crit}[\gamma_m(h)] \setminus \mathrm{crit}_\pm[\gamma(h)]$*, define*

$$s(x) := \begin{cases} 1 & x \in \mathrm{crit}[\gamma_m(h)] \quad \partial_2\Gamma_m(x, h(x)) > 0 \\ -1 & x \in \mathrm{crit}[\gamma_m(h)] \quad \partial_2\Gamma_m(x, h(x)) < 0 \end{cases}.$$

If $s(x)$ *alternates at least* $N + 2 - |\mathrm{crit}_\pm[\gamma(h)]|$ *times, then* $h \in \mathcal{P}^N$ *is a local Pareto point of* γ*.*

Proof: Let $N_c := |\text{crit}_{\pm}[\gamma(h)]|$ and assume $s(x)$ alternates at least $N + 2 - N_c$ times. Suppose that $h \in \mathcal{P}^N$ is a **not** locally Pareto-optimal. Lemma 9 furnishes $\Delta h \in \mathcal{P}^N$ that is nonzero and

$$0 \geq \partial_2 \Gamma_m(x, h(x))\Delta h(x) \quad (x \in \text{crit}[\gamma_m(h)])$$

for $m = 1, \ldots, M$. Because $\partial_2 \Gamma_m(x, h(x))$ does not vanish on $\text{crit}[\gamma_m(h)]$, it follows that Δh must vanish on $\text{crit}_{\pm}[\gamma(h)]$. Factor $\Delta h(x)$ as

$$\Delta h(x) = p(x)\Delta\widetilde{h}(x)$$

where $p \in \mathcal{P}^{N_c}$ contains the zeros of $\text{crit}_{\pm}[\gamma(h)]$ and $\Delta\widetilde{h}(x) \in \mathcal{P}^{N+2-N_c}$ is zero-free on $\text{crit}_{\pm}[\gamma(h)]$. However, $0 \geq s(x)\Delta\widetilde{h}(x)$ on $\text{crit}[\gamma(h)] \setminus \text{crit}_{\pm}[\gamma(h)]$. The $N + 2 - N_c$ alternations of $s(x)$ force $\Delta\widetilde{h}$ to have at least $N + 1 - N_c$ zeros. Consequently, $\Delta\widetilde{h}$ must be zero, which forces $\Delta h = 0$ and contradicts that Δh is nonzero. Thus, $h \in \mathcal{H}$ must be a local Pareto point. ///

The beauty of Lemma 11 is that all the critical sets contribute to the alternating sequence—decreased by the phase splitting. Although phase splitting obscures the converse, we have enough machinery to explore the Pareto sets.

8.3 The Pareto Set of $\exp(\pm x)$

Section 6 pointed out that the fundamental problem of multiobjective optimization is computing the Pareto set of $\gamma : \mathcal{H} \to \mathbb{R}^N$. Recall that the Pareto set resides in $\mathcal{H}$. Consequently, the Pareto set depends on the parameterization of $\mathcal{H}$ and γ. From this computational point of view, the Pareto set is difficult to visualize and to use in engineering trade-offs. In contrast, *Pareto image*—the set of all the Pareto points mapped by γ into $\mathbb{R}^N$—is far more practical and computationally available. Typically, the range of γ has low dimension ($N \leq 3$) so that the engineer can see the performance and make decisions about trade-offs.

How to get the Pareto image when its Pareto set is unknown is an excellent question. Because the Pareto image consists of the minimal elements of $\gamma(\mathcal{H})$, we can "sketch" $\gamma(\mathcal{H})$ by randomly sampling $h \in \mathcal{H}$ and plotting the random points $\gamma(h)$. As the sampling gets denser, the image $\gamma(\mathcal{H})$ starts to fill in and the boundary containing the minimal elements starts to appear.

For example, consider the objective function of Section 8.1 that approximates the exponential function and its reciprocal by polynomials $h \in \mathcal{P}^3$. The objective function

$$\gamma(h) = \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \end{bmatrix}$$

consists of the performance functions

$$\begin{aligned}\Gamma_1(x, h(x)) &= (e^x - h(x))^2 \\ \Gamma_2(x, h(x)) &= (e^{-x} - h(x)^{-1})^2\end{aligned}$$

that have variation

$$\begin{aligned}\partial_2\Gamma_1(x, h(x)) &= -2(e^x - h(x)) \\ \partial_2\Gamma_2(x, h(x)) &= +2(e^{-x} - h(x)^{-1})h(x)^{-2}.\end{aligned}$$

The domain $\mathcal{H}$ of γ is the nonvanishing polynomials of $\mathcal{P}^3$. Figure 24 sketches the image of γ. The plot is a closeup of Figure 22. The blue dots in the upper right are $\gamma(h)$ for random polynomials in $\mathcal{P}^3$. The red square marks the starting point for the minimizing method. The green diamonds are the method's terminal points.

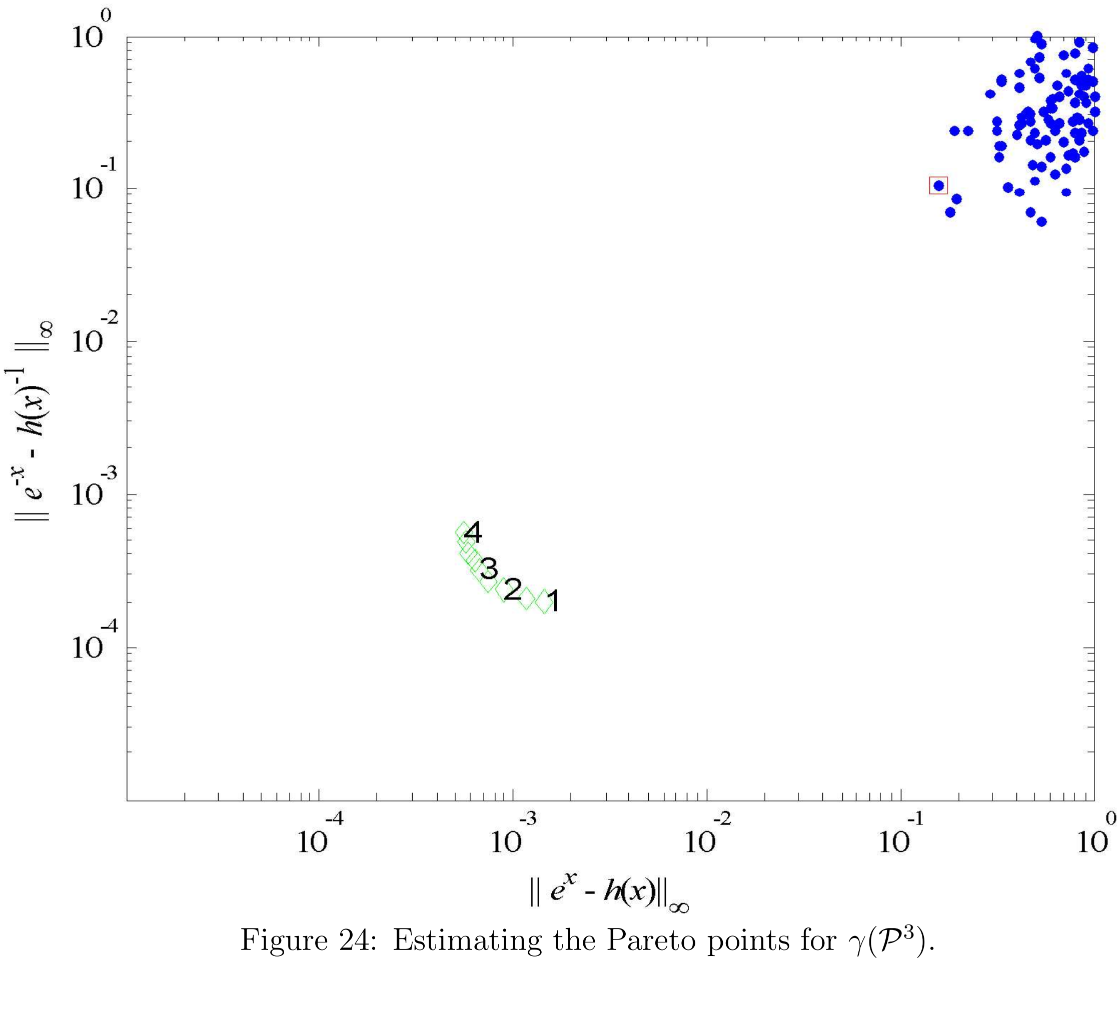

Figure 24: Estimating the Pareto points for $\gamma(\mathcal{P}^3)$.

The Goal Attainment Method computed the minimizers using the weight vector

$$\mathbf{w} = \begin{bmatrix} \cos(\theta_w) \\ \sin(\theta_w) \end{bmatrix}.$$

As θ_w sweeps from 0° to 90°, the Goal Attainment Method sweeps out what are numerically local Pareto with images marked by the green diamonds. This numerical approximation of the Pareto image allows an engineer to see the trade-off between the objective functions. Figure 24 also numbers selected points. The following plots discuss the Pareto condition for each numbered point.

Figure 25 shows the error curves of Point #1. The red segments mark the critical set regions. The numbers on the right are the coefficients of the polynomial. The lower plot exhibits an alternating sequence of length 5. Lemma 11 observes that this polynomial is indeed locally Pareto-optimal.

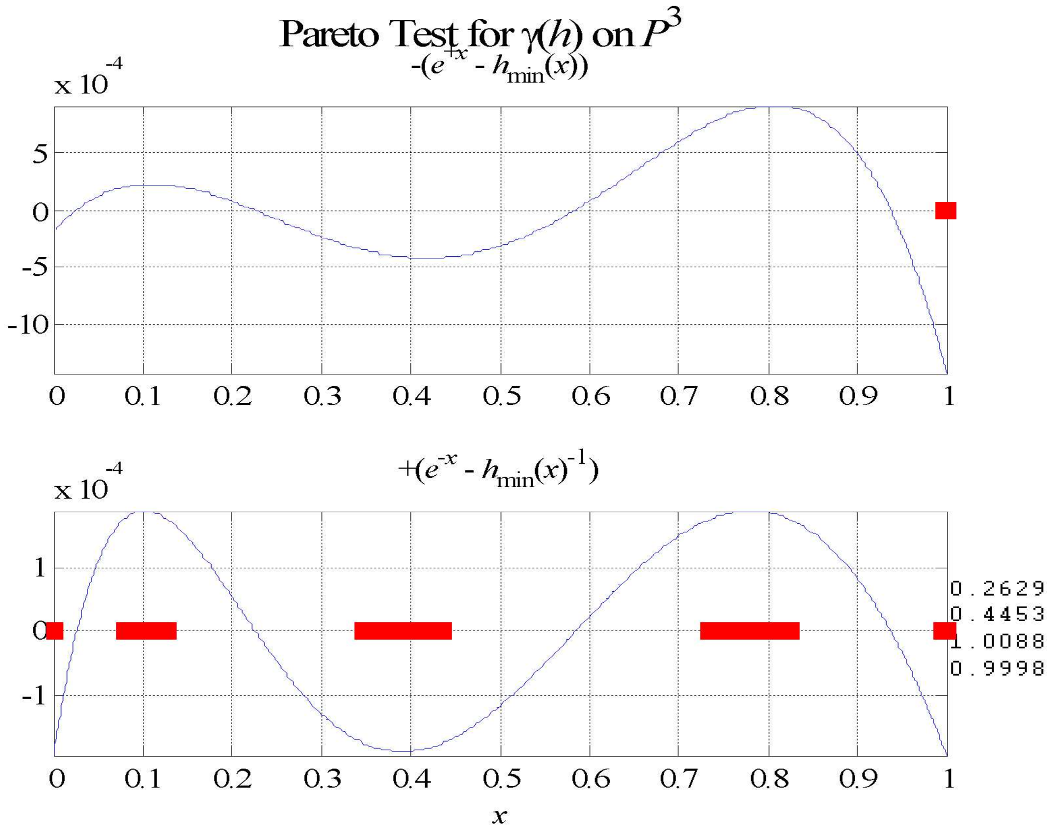

Figure 25: Pareto Test for Point #1.

Figure 26 shows the error curves of Point #2. The lower plot now exhibits an alternating sequence of length 3 while the upper plot picks up the alternating sequence 2—in phase with the lower plot—to get a generalized alternating sequence of length 5. Lemma 11 observes that this polynomial (coefficients listed on the right) is locally Pareto-optimal.

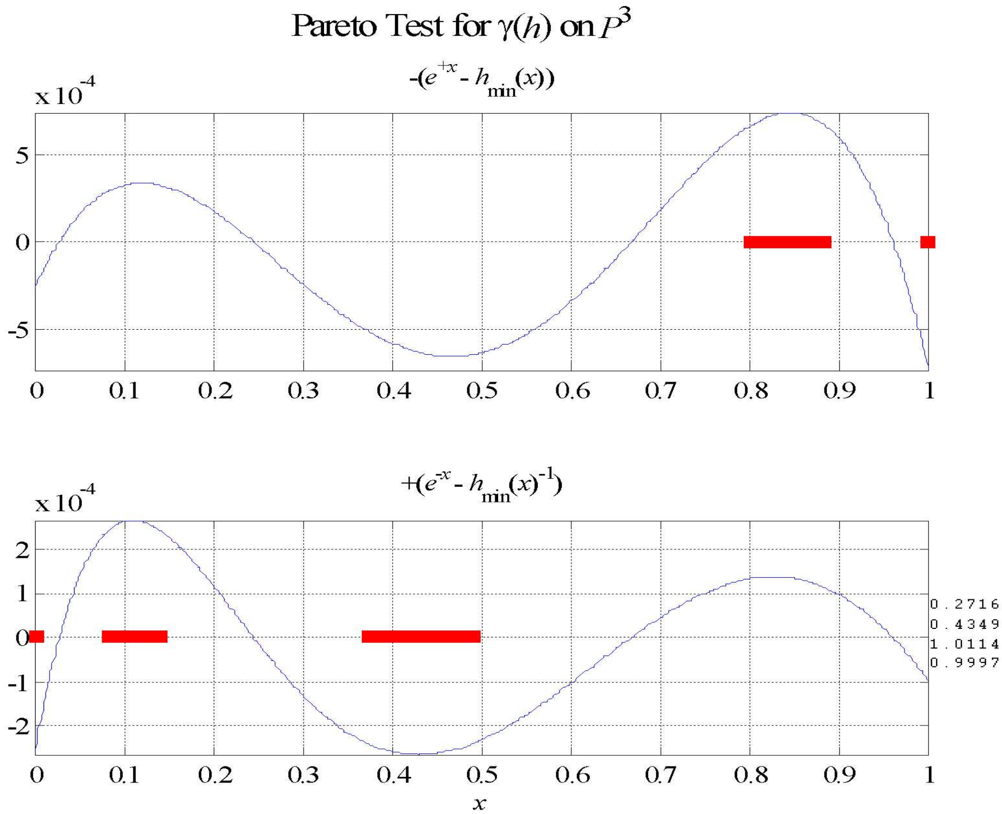

Figure 26: Pareto Test for Point #2.

Figure 27 shows the error curves of Point #3. Here, the alternating sequence splits between the two error curves. This plot is a splendid example of Lemma 11. Because this generalized alternating sequence has length 5, Lemma 11 verifies that the polynomial under test is locally Pareto-optimal.

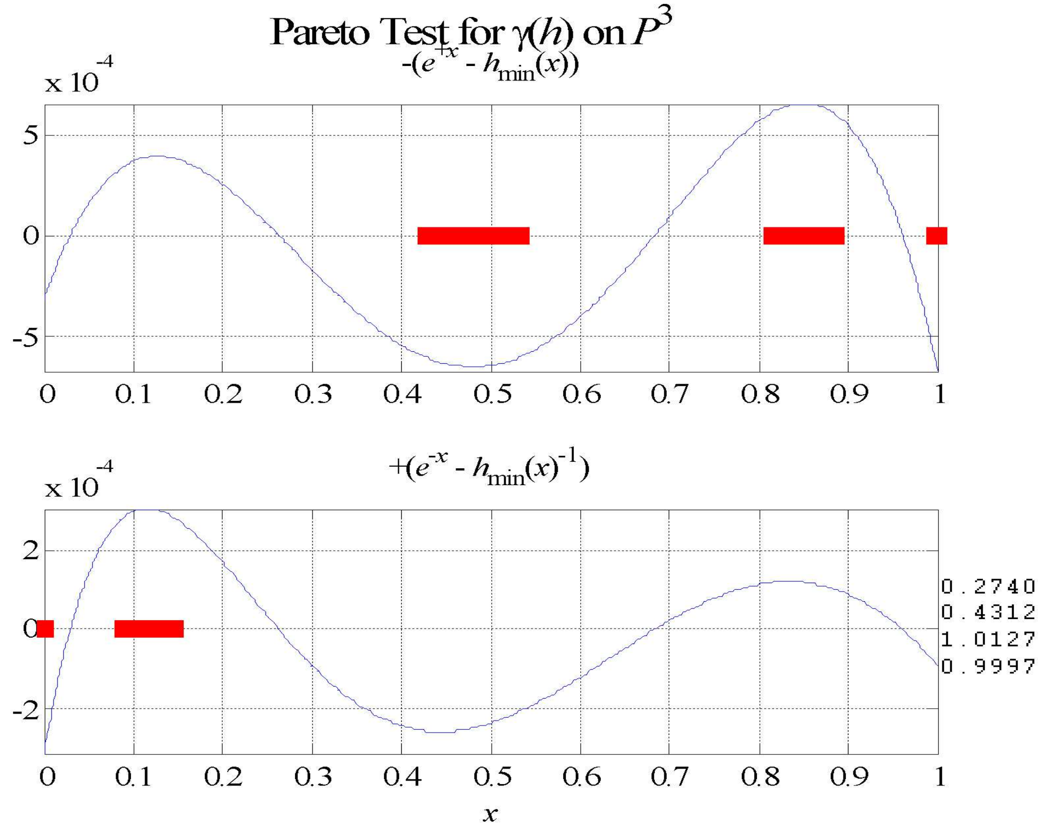

Figure 27: Pareto Test for Point #3.

Figure 28 shows the error curves of Point #4 and shows that the alternating sequence resides in the upper plot. Looking at all these plots in sequence, we see the alternating sequence starting in the lower plot (Point #1), splitting between the lower and upper plots (Points #2 and #3), and moving into the upper plot (Point #4). Lemma 11 applies in all cases and verifies that the polynomials under test are locally Pareto-optimal.

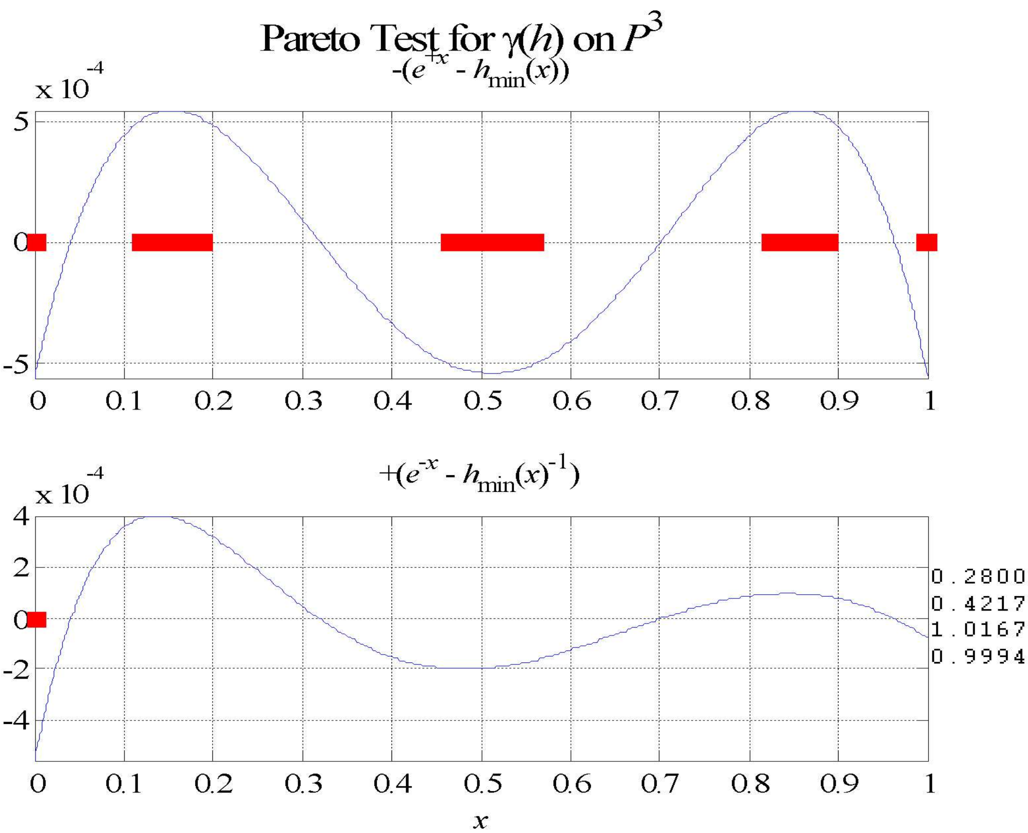

Figure 28: Pareto Test for Point #4.

9 The Kolmogorov Approach

The Kolmogorov approach to optimization is a general method that yields surprisingly concrete results when applied to the objective function

$$\gamma(h) = \sup\{\Gamma(z, h(z)) : z \in Z\}.$$

Section 4 demonstrated the effectiveness of the Kolmogorov approach for characterizing the minimizers of $\gamma(h)$ on the polynomials. The classical alternating condition for polynomial approximation is generalized to a new alternating condition from $\partial_2\Gamma(z, h(z))$. On the polynomials, we have essentially a finite-dimensional and real-valued minimization problem.

In contrast, Section 5 applies the Kolmogorov approach to the disk algebra—an infinite-dimensional domain consisting of complex-valued functions. The Kolmogorov approach readily characterizes the minimizers in the disk algebra. Although this result belongs to Helton and Merino [17], this sections shows that the Kolmogorov approach provides a general method to attack these minimization problems. This section also showed how to link the polynomial minimizers to the disk algebra bounds obtained by Helton and Merino [17]. This approach allows the engineer to "benchmark" these suboptimal solutions against an absolute best bound. This benchmarking is a splendid example of how pure mathematics can enhance traditional engineering [2], [4], [3]. Indeed, nothing drives an engineer to seek an optimal solution as striving against a "best bound."

Not only does the Kolmogorov approach give the basic results for these minimization problems, it generalizes to minimization problems of several objective functions. Section 6 lifts the single-objective minimization problem to the multiobjective minimization problem. Section 7 develops the multiobjective Kolmogorov approach Section 8 applies this approach to multiobjective optimization on the polynomials. The new alternating condition of the single-objective case is generalized to a new alternating condition that sweeps over the objective functions. The technical complication that stymies a complete characterization is the possibility of "phase splitting." Nevertheless, there exists enough theory to identify locally optimal minimizers.

The Kolmogorov approach should also apply to multiobjective optimization on the disk algebra. Indeed, Helton and Whittlesey [18], Helton and Vityaev [16], and Helton and Merino [17] generalize the single-objective minimization problem to a multiobjective problem on the disk algebra.

We conclude by making explicit a fruitful point of contact between the disk algebra and the finite-dimensional spaces of polynomials in a robust filter-design problem. The problem is to make a trade-off between realizing a transfer function and its associated sensitivity to design parameters [25].

For example, the transducer power gain G_T of a low-pass ladder can be parameterized as

$$G_T(h; j\omega) := \frac{1}{1 + |h(j\omega)|^2},$$

where $h(j\omega)$ is a real-valued polynomial and ω is the radial frequency. By making a bilinear transform and a slight abuse of notation, the problem can be put on the unit circle:

$$G_T(h; z) := \frac{1}{1 + |h(z)|^2}, \qquad (z = e^{j\theta})$$

where $h(z)$ is a real-valued polynomial. The optimization of the transducer power gain is the problem of finding a polynomial $h(z)$ so that $G_T(h)$ follows a user-specified design:

$$G_T(h; z) \approx G_{T,u}(z).$$

We specify a Gaussian filter as plotted in Figure 29. The goal is to build a low-pass ladder with a gain G_T that is close to the Gaussian filter. One measure of "filter error" is

$$\Gamma_1(z, h(z)) = (G_T(h; z) - G_{T,u}(z))^2; \qquad (z = e^{j\theta}).$$

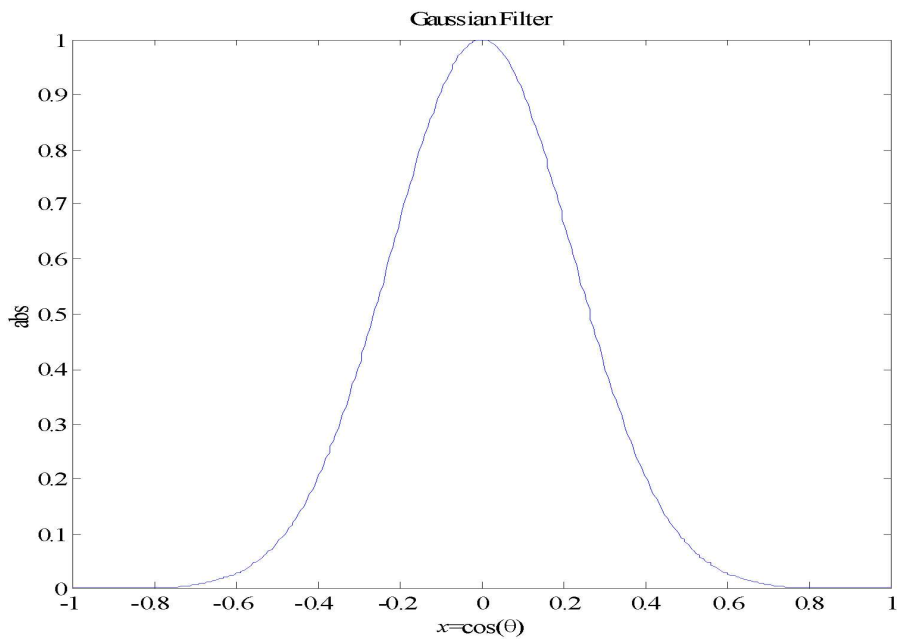

Figure 29: User-specified Gaussian filter.

Allied with the design of the low-pass ladder is its sensitivity—the variation of the gain as a function of its parameterizing polynomial:

$$G(h + \Delta h) = G(h) + 2\Re[\partial_h G(h)\Delta h] + \ldots.$$

With the variation of the gain given as

$$\partial_h G(h) = -G(h)^2\overline{h},$$

one measure of the "sensitivity" of the design is then

$$\Gamma_2(j\omega, h(j\omega)) = |\Re[\partial_h G(h; j\omega)]|^2 = |\Re[G(h)^2\overline{h}|^2.$$

Consequently, the problem of minimizing

$$\gamma(h) = \begin{bmatrix} \gamma_1(h) \\ \gamma_2(h) \end{bmatrix}$$

is the problem of finding a filter of minimum sensitivity that is closest to the specified design.

Figure 30 plots random samples of $\gamma(h)$ in the Filter-Sensitivity plane for the polynomials of third degree ($h \in \mathcal{P}^3$) as the blue dots. The red square is the starting point for the minimizing method. The green squares are the numerical minimizers.

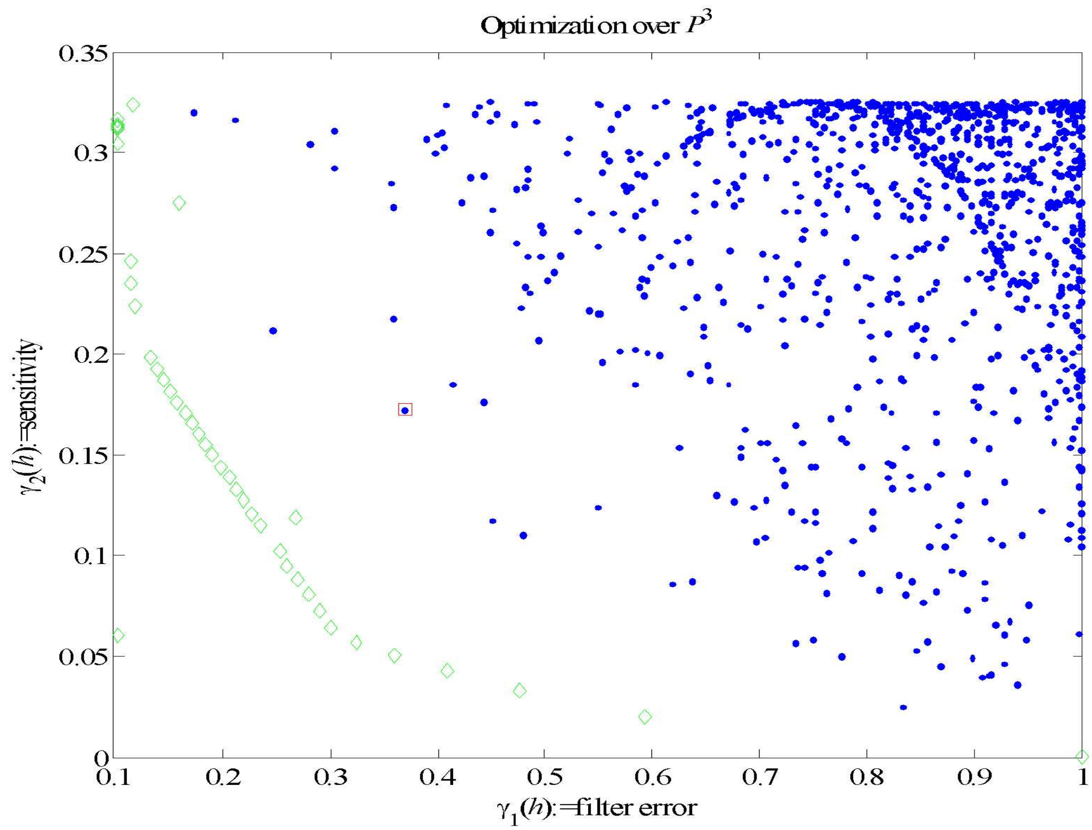

Figure 30: Pareto surface in the Filter-Sensitivity plane.

The Goal Attainment Method computes these minimizers using the weight vector

$$\mathbf{w} = \begin{bmatrix} \cos(\theta_w) \\ \sin(\theta_w) \end{bmatrix}.$$

As θ_w sweeps from $0°$ to $90°$, the Goal Attainment Method sweeps out numerical approximations to local Pareto points with images marked by the green diamonds.

This numerical approximation of the Pareto image shows an engineer the trade-off between the filter accuracy and sensitivity. The curve generally shows that sensitivity increases as the error decreases. The curve also hints at the existence of fascinating fine structure. The single point that has near optimal gain and low sensitivity certainly attracts the attention of an engineer and brings the rest of the "connected" Pareto image into question.

Figure 31 increases the degree of the polynomials from 3 to 6. The plot reveals that the Pareto image does have a fine structure—a fine structure of "high-performance" points.

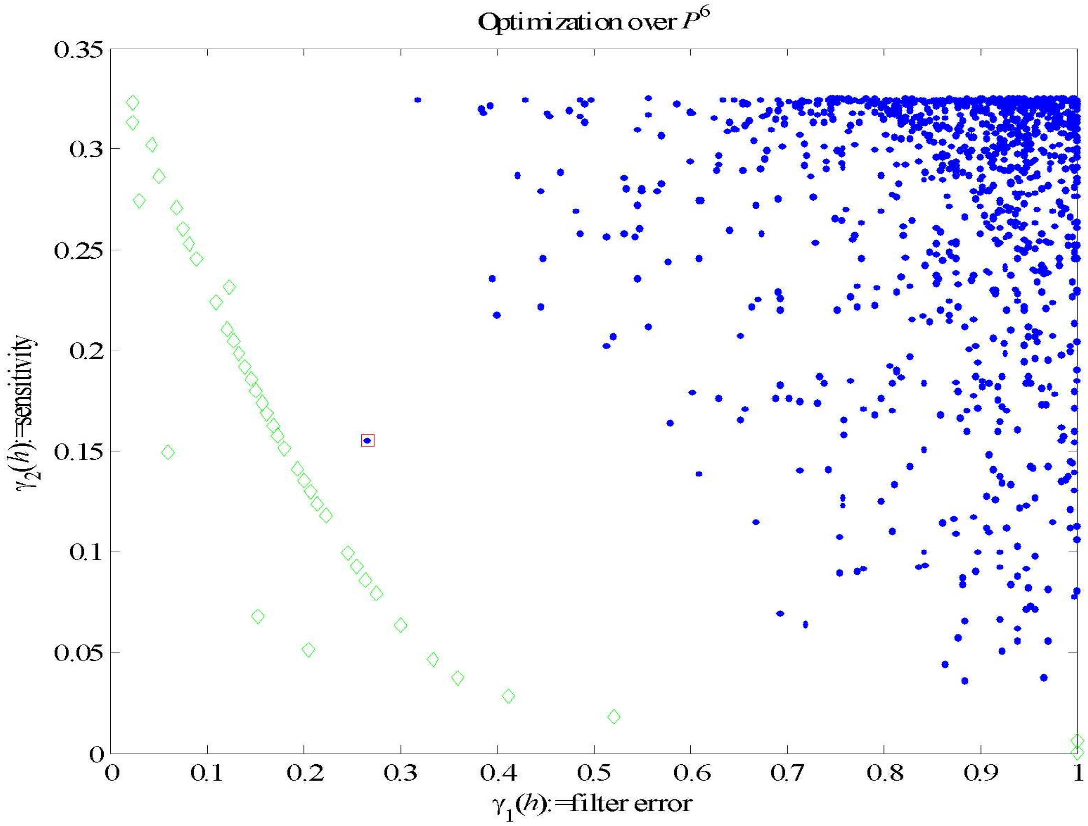

Figure 31: Pareto surface for $\mathcal{P}^6$.

When both plots are put in the context of optimizing over a family of polynomials $\mathcal{P}^N$ for $N \to \infty$, two issues become apparent. First is the problem of determining if a given point belongs to the Pareto image. This problem is specific to the multiobjective optimization for polynomials and the general *characterization problem* raised in the beginning of this report—can an answer be recognized? The second issue puts both plots in the context of the best bounds that follow from multiobjective optimization on the disk algebra. What would be very helpful for the engineer is a plot of the best possible bounds attainable on the disk algebra. This "ultimate Pareto image" would bound all the polynomial cases and let the engineer trade off filter performance as a function of degree. Thus, this simple filter design problem is an excellent point-of-departure for research in multiobjective optimization.

References

[1] Abraham, Ralph, Jerrold E. Marsden, and Tudor Ratiu [1983] *Manifolds, Tensor Analysis and Applications*, Addison-Wesley, Reading, MA.

[2] Allen, J. [2004] *H^∞ Engineering and Amplifier Optimization*, Birkhauser, Boston.

[3] Allen, J. C., Lance Koyama, and David F. Schwartz [2000] H^∞ Broadband Antenna Matching: Case Studies *17th Annual Review of Progress in Applied Computational Electromagnetics*, pages 622–625. Monterey, CA

[4] Allen, J. C. and Dennis Healy [2003] Nehari's Theorem and Electric Circuits, in *Modern Signal Processing*, Edited by Daniel Rockmore & Dennis Healy, Springer-Verlag.

[5] Ball, Joseph A. and J. William Helton [1979] Interpolation with Outer Functions and Gain Equalization in Amplifiers, *International Symposium on Mathematical Theory of Networks and Systems*, Delft University of Technology, The Netherlands.

[6] Boyd, Stephen and Lieven Vandenberghe [1999] *Convex Optimization*, preprint from Stanford University.

[7] Braess, Dietrich [1986] *Nonlinear Approximation Theory*, Springer-Verlag, Berlin.

[8] Branch, Mary Ann and Andrew Grace [1996], *Optimization Toolbox*, The MathWorks Inc.

[9] Censor, Yair [1977] Pareto Optimality in Multiobjective Problems, *Applied Mathematics and Optimization*, Volume 4, pages 41–59.

[10] Chen, G. Y. and B. D. Craven [1994] Existence and Continuity of Solutions for Vector Optimization, *Journal of Optimization Theory and Applications*, Volume 81, Number 3, pages 459–468.

[11] Cheney, E. W. [1982] *Introduction to Approximation Theory*, second edition, Chelsea Publishing Company.

[12] Das, Indraneel and J. E. Dennis [1998] Normal-Boundary Intersecting: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems, *SIAM Journal of Optimization*, Volume 8, Number 3, pages 631–657.

[13] Helton, J. W. [1981] Broadbanding Gain Equalization Directly From Data, *IEEE Transactions on Circuits and Systems*, CAS-28(12), pages 1125–1137.

[14] Helton, J. W. [1982] Non-Euclidean Functional Analysis and Electronics, *Bulletin of the American Mathematical Society*, Volume 7, Number 1, pages 1–64.

[15] Helton, J. W. [1985] Worst Case Analysis in the Frequency Domain: The H^∞ Approach to Control, *IEEE Transactions on Automatic Control*, AC-30(12), pages 1154–1170.

[16] Helton, J. W. and A. E. Vityaev [1997] Analytic Functions Optimizing Competing Constraints, *SIAM Journal of Mathematical Analysis*, Volume 30, Number 3, pages 749–767.

[17] Helton, J. W. and O. Merino [1998] *Classical Control Using $H^\infty(\mathbb{C}_+)$ Methods*, SIAM, Philadelphia.

[18] Helton, J. W. and Marshall A. Whittlesey [2000] Global Uniqueness Tests for H^∞ Optima, *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia.

[19] Jonckheere, Edmond A. and J. W. Helton [1985] Power Spectrum Reduction by Optimal Hankel Norm Approximation of the Phase of the Outer Spectral Factor, *IEEE Transactions on Automatic Control*, AC-30(12), pages 1192-1201.

[20] Light, W. A. and E. W. Cheney [1985] *Approximation Theory in Tensor Product Spaces*, Springer-Verlag, Berlin.

[21] Meinardus, G. [1967] *Approximation of Function*, Springer-Verlag, Berlin.

[22] Rudin, Walter [1974] *Real and Complex Analysis*, McGraw-Hill, New York.

[23] Schwartz, D. and J. Allen [2004] H^∞ Approximation with Point Constraints Applied to Impedance Estimation, *Circuits, Systems, and Signal Processing*, 16(5), pages 507–522.

[24] Schwartz, D. and J. Allen [2004] Wideband Impedance Matching: H^∞ Performance Bounds, *IEEE Transactions on Circuits and Systems II: Express Briefs*, 51(7), pages 364–368

[25] Strycula, E. C. and N. K. Bose [1974] Theory and Applications of Simultaneous Realization of Transfer and Sensitivity Functions, *IEEE Transactions on Circuits and Systems*, CAS-21(3), pages 382–387.

[26] Ungar, Abraham A. [1999] The Hyperbolic Pythagorean Theorem, in the Poincaré Disk Model of Hyperbolic Geometry, *The American Mathematical Monthly*, Volume 106, Number 8, pages 759–763.

[27] Zhu, Kehe [1990] *Operator Theory in Function Spaces*, Marcel Dekker, Inc., New York, NY.