
Spectral and Analytic Structure of the Nyman–Beurling–Báez–Duarte Approximation

[Deep Bhattacharjee](#)*, [Pallab Nandi](#), [Onwuka Frederick](#), Priyanka Samal

Posted Date: 10 March 2026

doi: 10.20944/preprints202506.0772.v2

Keywords: Nyman–Beurling criterion; Báez–Duarte approximation; Riemann zeta function; Kalman filtration; Gram matrix spectral theory; Mellin transform; Hardy spaces; Dirichlet polynomial approximation; Hilbert–Pólya conjecture; compact operators; analytic number theory; rank-one collapse; SVD stability; Li criterion



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Spectral and Analytic Structure of the Nyman–Beurling–Báez–Duarte Approximation

Deep Bhattacharjee ^{1,*}, Pallab Nandi ², Onwuka Frederick ³ and Priyanka Samal ⁴

¹ Electro-Gravitational Space Propulsion Laboratory, India

² Indian Institute of Science, Education and Research, Kolkata, India

³ Ekiti State University, Nigeria

⁴ Independent Researcher, India

* Correspondence: itsdeep@live.com

Abstract

We study the structural and analytic aspects of the Báez–Duarte approximation problem within the Nyman–Beurling framework, which furnishes a functional-analytic equivalent of the Riemann Hypothesis (RH). Our work studies structural features of this framework; it does *not* prove RH. *First (Rank-one collapse and Hilbert-space theory)*. The integer-dilate Gram matrix $G_M = \frac{1}{3} \mathbf{d} \mathbf{d}^\top$ is rank-one, giving $\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\}$ and fixed distance $d_M = \frac{1}{2}$ for all M . We give the explicit Moore–Penrose pseudoinverse G_M^+ and the one-dimensional collapse of the optimisation problem. *Second (Exact Gram matrix formula)*. We prove a fully rigorous closed-form expression for the inner products of the correct sawtooth basis: using the Bernoulli polynomial representation of the fractional part,

$$\int_0^1 \{jx\} \{kx\} dx = \frac{\gcd(j,k)^2}{jk} \left(\frac{1}{12} + \frac{B_2(0)}{2} \right) + \frac{1}{4} \left(1 - \frac{\gcd(j,k)}{j} \right) \left(1 - \frac{\gcd(j,k)}{k} \right) + E_{jk},$$

where E_{jk} is an explicit correction from higher Bernoulli terms, expressed via the Hurwitz zeta function. The arithmetic role of $\gcd(j,k)$ is made precise. *Third (Hardy-space bounds)*. Using the $H^2(\Pi^+)$ reproducing kernel and the Mellin isometry, we prove: (a) the distance identity $d_M^2 = \|1/s - F_M^*(s)\zeta(s)/s\|_{H^2}^2$; (b) an explicit lower bound $d_M^2 \geq \sum_\rho \frac{|F_M^*(\rho)|^2 |\zeta'(\rho)|^{-2}}{|\rho|^2} \cdot c(\rho)$ from the zeros of ζ ; and (c) a pointwise Hardy-space inequality relating d_M to the supremum of $|1 - F_M^*(\frac{1}{2} + it)\zeta(\frac{1}{2} + it)/(\frac{1}{2} + it)|$ on the critical line. *Fourth (Kalman filtration stability)*. Under the observation model $z_M = d_M + \varepsilon_M$ with ε_M sub-Gaussian of variance σ^2 , the Kalman estimator satisfies a rigorous oracle inequality $\mathbf{E}|d_M^{\text{KF}} - d_M|^2 \leq \sigma^2 K_\infty(2 - K_\infty)^{-1}$, with an almost-sure bound $|d_M^{\text{KF}} - d_M| \leq CM^{-\alpha}$ whenever $|d_M - d| = O(M^{-\alpha})$. *Fifth (Möbius sparsity)*. We prove $|c_k^*| = O(k^{-1+\varepsilon})$ via Dirichlet series techniques and show that the coefficient sequence is bounded in ℓ^2 , with connections to the Möbius function made precise through the optimality conditions. *Sixth (Structural Mellin theorem)*. We identify a hidden structural observation in the Mellin identity: the Gram kernel $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$ appears as the reproducing kernel of the Hardy space $H^2(\Pi^+)$ restricted to the approximation subspace W_M , and its singularity at $s + \bar{w} = 1$ encodes the pole of ζ while the zeros of ζ in the critical strip contribute exactly as spectral obstructions. **Disclaimer.** *This paper does not prove RH. All results are structural, computational, and analytic observations within the equivalent framework.*

Keywords: Nyman–Beurling criterion; Báez–Duarte approximation; Riemann zeta function; Kalman filtration; Gram matrix spectral theory; Mellin transform; Hardy spaces; Dirichlet polynomial approximation; Hilbert–Pólya conjecture; compact operators; analytic number theory; rank-one collapse; SVD stability; Li criterion

PACS: 02.30.Gp (Special functions); 02.30.Sa (Functional analysis); 02.30.Tb (Operator theory); 02.60.Cb (Numerical analysis); 05.45.-a (Spectral dynamics)

MSC: 11M26 (primary); 47B32; 47G10; 42A10; 65F20; 65F35; 93E11 (secondary)

1. Introduction

1.1. The Riemann Hypothesis and Its Context

The Riemann Hypothesis (RH) asserts that every nontrivial zero ρ of the Riemann zeta function

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}, \quad \operatorname{Re}(s) > 1,$$

satisfies $\operatorname{Re}(\rho) = \frac{1}{2}$. Since Riemann's memoir of 1859, the hypothesis has shaped the entire landscape of analytic number theory. Listed as one of the seven Millennium Prize Problems by the Clay Mathematics Institute [2], it remains unresolved.

Its importance stems not only from its elegance, but from its deep connections to the distribution of primes. Under RH, the prime-counting error satisfies $|\psi(x) - x| = O(\sqrt{x} \log^2 x)$, where $\psi(x) = \sum_{p^k \leq x} \log p$ is the Chebyshev function. Any zero off the critical line would produce larger oscillations in the distribution of primes [1,12].

1.2. Equivalent Reformulations of RH

A rich ecosystem of equivalent reformulations has grown around RH, each illuminating a different facet of the hypothesis. Among the most celebrated are:

- (i) **Nyman–Beurling criterion** [6,7]: $1 \in \overline{\operatorname{span}\{f_\theta : \theta \in (0,1)\}}^{L^2}$ if and only if RH holds.
- (ii) **Báez–Duarte strengthening** [8]: it suffices to use the countable family $\{r_k(x) = \{x/k\} : k \in \mathbb{N}\}$.
- (iii) **Li's criterion** [11]: RH is equivalent to the positivity of the coefficients $\lambda_n = \sum_\rho [1 - (1 - 1/\rho)^n]$ for all $n \geq 1$.
- (iv) **Robin's criterion** [16]: RH is equivalent to $\sigma(n) < e^\gamma n \log \log n$ for all $n \geq 5041$, where $\sigma(n)$ is the sum-of-divisors function and γ is the Euler–Mascheroni constant.
- (v) **Weil's explicit formula criterion**: RH is equivalent to the positivity of certain explicit sums over primes and zeros.

The present paper focuses on the Nyman–Beurling–Báez–Duarte framework, which has the unique advantage of being both analytically rigorous and computationally tractable.

1.3. The Nyman–Beurling Framework: Background and Prior Work

The Nyman–Beurling criterion originated in Nyman's 1950 Uppsala thesis [6], where he proved that ζ has no zeros in the strip $\frac{1}{2} < \operatorname{Re}(s) < 1$ if and only if certain functions related to the fractional part of $1/x$ are dense in $L^2(0,1)$. Beurling [7] reformulated and extended this, connecting it more cleanly to RH.

The approach gained momentum through Burnol's [10] explicit product formula for the projection norm (Theorem 4) and Báez–Duarte's [8] reduction to a countable basis $\{r_k(x) = \{x/k\}\}$. Báez–Duarte [9] and subsequent authors attempted numerical evaluations of d_M for moderate M .

Prior computational investigations. Attempts to numerically study d_M have used a variety of strategies. Some implementations directly compute the Gram matrix $(G_M)_{jk} = \langle r_j, r_k \rangle$ and solve the normal equations. Others discretise the inner products at finitely many quadrature nodes. A common source of error in these investigations—addressed thoroughly in this paper—is the failure to distinguish between the functions $r_k(x) = \{x/k\}$ (which reduce to x/k on $(0,1)$ and generate only the one-dimensional subspace $\operatorname{span}\{x\}$) and the genuine sawtooth basis $\check{r}_k(x) = \{kx\}$ (which has k teeth on $(0,1)$ and spans a genuinely k -dimensional function space on the subintervals of scale $1/k$). This collapse is the central structural observation of Sections 3 and 4.

The Mellin transform provides the bridge between the functional-analytic statement and the zero-set of ζ . Specifically, the Nyman–Beurling condition $\mathbf{1} \in \overline{\text{span}\{f_\theta\}}^{L^2}$ is equivalent, via the isometry $\mathcal{M} : L^2(0,1) \rightarrow H^2(\Pi^+)$ and the identity $\widehat{f}_\theta(s) = \theta^s \zeta(s)/s$, to the density of $\{\theta^s \zeta(s)/s : \theta \in (0,1)\}$ in $H^2(\Pi^+)$, which in turn is equivalent to RH. This Hardy-space perspective is developed systematically in Section 9.

1.4. The Central Problem: Structural Degeneracy

A fundamental difficulty in the numerical study of the Báez–Duarte problem—one that has not been adequately addressed in the existing literature—concerns the behaviour of the basis functions on the unit interval $(0,1)$.

For $x \in (0,1)$ and integer $k \geq 1$: $r_k(x) = \{x/k\} = x/k$, since $0 < x/k < 1$. Therefore all functions r_1, r_2, \dots are scalar multiples of x , and

$$\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\} \quad \text{for all } M \geq 1.$$

The Gram matrix G_M is *rank-one* for every $M \geq 2$. This **rank-one collapse** is a fundamental structural obstruction: $\mathbf{1} \notin \text{span}\{x\}$, so $d_M = \frac{1}{2}$ for all M , and any numerical scheme claiming $d_M \rightarrow 0$ using these (degenerate) basis functions is erroneous.

This observation does not invalidate the Báez–Duarte theorem, which is a theorem about closures in L^2 . Rather, it reveals that a naively discretised implementation—where r_k is evaluated at points $x \in (0,1)$ and treated as though these evaluations capture the genuine sawtooth structure—misses the essential nonlinearity. Correct numerical implementations must use the sawtooth functions $\tilde{r}_k(x) = \{kx\}$ (or equivalently, $\{x/k\}$ for $x \in (0,k)$), which are genuinely nonlinear and linearly independent.

1.5. Contributions and Outline

This paper makes the following contributions:

- (i) **Rank-one collapse and Hilbert-space theory** (Sections 3–4): We prove $G_M = \frac{1}{3} \mathbf{d} \mathbf{d}^\top$ is rank-one, derive the Moore–Penrose pseudoinverse $G_M^+ = \frac{3}{\|\mathbf{d}\|^4} \mathbf{d} \mathbf{d}^\top$, establish $d_M = \frac{1}{2}$ for all M , and explain the collapse of the approximation to a one-dimensional scalar problem.
- (ii) **Structural degeneracy analysis for numerical implementations** (Sections 3–5): We provide a detailed explanation of why incorrect basis functions cause structurally misleading numerical results, and characterise the correct sawtooth basis $\tilde{r}_k(x) = \{kx\}$ needed for genuine approximation.
- (iii) **Gram matrix spectral theory and stability** (Section 6): We establish $\kappa(\tilde{G}_M) = \Theta(M^2)$ via compact operator theory and Weyl’s law, prove an operator-stability theorem $\|\tilde{G}_M - G_M^{\text{cts}}\|_2 \leq C/N$, and derive eigenvalue perturbation bounds.
- (iv) **Numerically stable algorithms** (Section 7): Truncated-SVD and economy-QR algorithms with rigorous backward-stability guarantees.
- (v) **Kalman Filtration theory** (Section 8): Convergence preservation, smoothing-error bounds $O(M^{-\alpha})$, variance reduction, and a general EWMA theorem.
- (vi) **Mellin-transform analytic theorems** (Section 9): We prove the isometry $d_M^2 = \|1/s - F_M(s)\zeta(s)/s\|_{H^2}^2$, derive lower bounds on d_M from values of the Dirichlet polynomial F_M at zeros of ζ , and prove a key lemma expressing inner products of sawtooth functions via the Hurwitz zeta function.
- (vii) **Number-theoretic connections** (Sections 14–16): Structural parallels with Li’s criterion, compressed sensing, and the Hilbert–Pólya philosophy.

What this paper does not prove. It does not establish $d = 0$ (equivalent to RH and still open), does not bound the convergence rate of d_M unconditionally, and does not prove any component of Conjecture 1. All results are structural, computational, and analytic observations within the equivalent framework.

2. Background and Functional-Analytic Framework

2.1. The Riemann Zeta Function

For $\text{Re}(s) > 1$, the Riemann zeta function is defined by the absolutely convergent Dirichlet series

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}.$$

By analytic continuation,¹ ζ extends to a meromorphic function on \mathbb{C} with a simple pole at $s = 1$ and trivial zeros at the negative even integers $s = -2, -4, \dots$

The nontrivial zeros $\rho = \beta + i\gamma$ satisfy $0 < \beta < 1$ (the open critical strip). The Riemann Hypothesis asserts $\beta = \frac{1}{2}$ for all of them.

Key analytic facts:

- (i) (Hardy [5]) Infinitely many nontrivial zeros lie on $\text{Re}(s) = \frac{1}{2}$.
- (ii) (Odlyzko [3]) The first 1.5×10^{10} zeros (ordered by $|\text{Im}(\rho)|$) all satisfy $\text{Re}(\rho) = \frac{1}{2}$.
- (iii) (Zero-free region) There exists a constant $c > 0$ such that $\zeta(s) \neq 0$ for $\text{Re}(s) \geq 1 - c / \log(|\text{Im}(s)| + 2)$.

2.2. Hilbert Spaces and Projection Theory

We recall the key tools from functional analysis.²

Theorem 1 (Orthogonal Projection Theorem). *Let \mathcal{H} be a Hilbert space and $V \subset \mathcal{H}$ a closed subspace. For any $f \in \mathcal{H}$, there exists a unique $\hat{f} \in V$ such that $\|f - \hat{f}\| = \text{dist}(f, V)$. Moreover, $\hat{f} = P_V f$ where $P_V : \mathcal{H} \rightarrow V$ is the orthogonal projection, characterised by $f - P_V f \perp V$.*

Corollary 1 (Best L^2 approximation). *The best $L^2(0, 1)$ approximation to f from a closed subspace $V = \text{span}\{v_1, \dots, v_M\}$ (with v_j linearly independent) is $\hat{f} = \sum_{k=1}^M c_k^* v_k$ where the coefficients $\mathbf{c}^* = (c_1^*, \dots, c_M^*)^\top$ satisfy the normal equations $G_M \mathbf{c}^* = \mathbf{b}$, with $(G_M)_{jk} = \langle v_j, v_k \rangle$ and $b_j = \langle f, v_j \rangle$. The minimum distance is $\text{dist}(f, V)^2 = \|f\|^2 - \mathbf{b}^\top G_M^{-1} \mathbf{b}$.*

Remark 1. *Corollary 1 requires G_M to be invertible, i.e., the v_j to be linearly independent. When G_M is rank-deficient (as in the integer-dilate case), the formula $\|f\|^2 - \mathbf{b}^\top G_M^{-1} \mathbf{b}$ is undefined, and one must use the Moore–Penrose pseudoinverse instead.*

2.3. Hardy Spaces and the Mellin Transform

The connection between the Nyman–Beurling criterion and RH is mediated by the Mellin transform.³

¹Analytic continuation is the process of extending a function defined on a smaller domain to a larger domain in a unique manner consistent with analyticity. The key tool for ζ is the functional equation $\zeta(s) = \zeta(1-s)$ where $\zeta(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s)$. This symmetry relates values in $\text{Re}(s) > 1$ to values in $\text{Re}(s) < 0$, allowing the extension to all of \mathbb{C} .

²A Hilbert space is a complete inner product space. The key example here is $L^2(0, 1)$: the space of (equivalence classes of) measurable functions $f : (0, 1) \rightarrow \mathbb{R}$ with $\int_0^1 |f(x)|^2 dx < \infty$, equipped with $\langle f, g \rangle = \int_0^1 f(x)g(x) dx$ and norm $\|f\| = \langle f, f \rangle^{1/2}$. Completeness means every Cauchy sequence converges, which underpins the projection theory.

³The Mellin transform of $f \in L^2(0, 1)$ is $(\mathcal{M}f)(s) = \int_0^1 x^{s-1} f(x) dx$, defined for $\text{Re}(s) > 0$. Via $x = e^{-t}$, the Mellin transform is unitarily equivalent to the Fourier–Laplace transform, mapping $L^2(0, 1)$ isometrically to the Hardy space $H^2(\Pi^+)$.

Definition 1 (Hardy space $H^2(\Pi^+)$). The Hardy space $H^2(\Pi^+)$ consists of analytic functions $F : \Pi^+ \rightarrow \mathbb{C}$ (where $\Pi^+ = \{s \in \mathbb{C} : \operatorname{Re}(s) > 0\}$) such that $\sup_{\sigma > 0} \int_{-\infty}^{\infty} |F(\sigma + it)|^2 dt < \infty$. With the inner product $\langle F, G \rangle_{H^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\frac{1}{2} + it) \overline{G(\frac{1}{2} + it)} dt$, $H^2(\Pi^+)$ is a Hilbert space.

The key identity connecting the Nyman–Beurling basis to ζ is:

$$\widehat{f}_\theta(s) = \int_0^1 x^{s-1} \{x/\theta\} dx = \frac{\theta^s \zeta(s)}{s}, \quad \operatorname{Re}(s) > 1. \quad (1)$$

This shows that the Mellin images of the dilate functions f_θ are scalar multiples of $\zeta(s)/s$. The closure condition in $L^2(0, 1)$ thus translates to an analytic approximation condition in $H^2(\Pi^+)$.

2.4. The Nyman–Beurling Criterion

Definition 2 (Fractional-part dilates). For $\theta \in (0, 1)$, define

$$f_\theta : (0, 1) \rightarrow \mathbb{R}, \quad f_\theta(x) = \{x/\theta\} = \frac{x}{\theta} - \left\lfloor \frac{x}{\theta} \right\rfloor.$$

Note that $f_\theta \in L^2(0, 1)$ since $|f_\theta| \leq 1$ a.e. Let $\mathcal{N} = \overline{\operatorname{span}\{f_\theta : \theta \in (0, 1)\}}^{L^2}$.

Theorem 2 (Nyman–Beurling [6,7]). $\text{RH} \iff d_{\mathcal{N}} := \operatorname{dist}(\mathbf{1}, \mathcal{N}) = 0$.

Remark 2 (Logical status). Theorem 2 is an equivalence. Any computation showing $d_M \rightarrow 0$ confirms consistency with RH but does not constitute an independent proof. An independent proof would require establishing $d = 0$ by a direct analytic argument that does not invoke the equivalence.

2.5. The Báez–Duarte Formulation

Theorem 3 (Báez–Duarte [8]). Let $r_k(x) = \{x/k\}$ for $k \in \mathbb{N}$ and $\mathcal{B} = \overline{\operatorname{span}\{r_k : k \geq 1\}}^{L^2(0,1)}$. Then $\text{RH} \iff \mathbf{1} \in \mathcal{B}$.

The Báez–Duarte theorem is a strengthening of the Nyman–Beurling result because it replaces an uncountable parameter set $\theta \in (0, 1)$ with the countable set $\{1/k : k \in \mathbb{N}\}$, making the problem amenable to computation.⁴

The finite-dimensional approximation problem:

$$d_M := \operatorname{dist}(\mathbf{1}, V_M), \quad V_M = \operatorname{span}\{r_1, \dots, r_M\},$$

yields a monotone decreasing sequence $d_M \searrow d = \operatorname{dist}(\mathbf{1}, \mathcal{B})$ as $M \rightarrow \infty$. By the equivalence, $d = 0 \iff \text{RH}$.

2.6. Burnol’s Projection Formula

The Mellin-transform Hardy-space framework yields an explicit formula for d^2 in terms of the zeros of ζ :

Theorem 4 (Burnol [10]). Let $P_{\mathcal{B}} : L^2(0, 1) \rightarrow \mathcal{B}$ be the orthogonal projection. Then

$$\|P_{\mathcal{B}}\mathbf{1}\|_{L^2}^2 = \prod_{\substack{\zeta(\rho)=0 \\ \operatorname{Re}(\rho) > 1/2}} \left| * \right| 1 - \frac{1}{\rho^2},$$

⁴The Báez–Duarte result follows from the Nyman–Beurling theorem via a density argument: the rational dilates $\{1/k : k \in \mathbb{N}\}$ are dense in $(0, 1)$, and the map $\theta \mapsto f_\theta$ is continuous in an appropriate sense.

and consequently

$$d^2 = 1 - \prod_{\substack{\zeta(\rho)=0 \\ \operatorname{Re}(\rho) > 1/2}} |*|1 - \frac{1}{\rho^2}.$$

In particular: $d = 0 \iff \text{RH}$.

Remark 3. Each factor $|1 - 1/\rho^2|$ in Burnol's product is strictly less than 1 when $\operatorname{Re}(\rho) > 1/2$. A single off-critical zero thus makes $d^2 > 0$. Conversely, if all zeros satisfy $\operatorname{Re}(\rho) = \frac{1}{2}$, the product over $\{\operatorname{Re}(\rho) > 1/2\}$ is empty and equals 1, giving $d = 0$.

3. Gram Matrix Analysis: The Rank-One Collapse

3.1. Exact Inner Products for Integer Dilates

We begin with a complete derivation of the inner products $\langle r_j, r_k \rangle$.

Lemma 1 (Fractional part on unit interval). For $x \in (0, 1)$ and integer $k \geq 1$, $\{x/k\} = x/k$.

Proof. Since $0 < x < 1$ and $k \geq 1$, we have $0 < x/k < 1/k \leq 1$, so $\lfloor x/k \rfloor = 0$ and $\{x/k\} = x/k - 0 = x/k$. \square

Proposition 1 (Exact inner products for integer dilates). For integers $j, k \geq 1$:

(i) $r_k(x) = x/k$ on $(0, 1)$ (by Lemma 1).

(ii) $\langle r_j, r_k \rangle_{L^2(0,1)} = \frac{1}{3jk}$.

(iii) $\|r_k\|_{L^2}^2 = \frac{1}{3k^2}$.

(iv) $\langle \mathbf{1}, r_k \rangle_{L^2} = \frac{1}{2k}$.

Proof. Using Lemma 1:

$$\langle r_j, r_k \rangle = \int_0^1 \frac{x}{j} \cdot \frac{x}{k} dx = \frac{1}{jk} \int_0^1 x^2 dx = \frac{1}{3jk},$$

which gives (ii). Setting $j = k$ gives (iii). For (iv): $\langle \mathbf{1}, r_k \rangle = \int_0^1 \frac{x}{k} dx = \frac{1}{k} \cdot \frac{1}{2} = \frac{1}{2k}$. \square

3.2. The Rank-One Gram Matrix: Structural Theorem

Key Structural Result: Rank-One Collapse

Theorem 5 (Rank-one Gram matrix). Let G_M denote the $M \times M$ Gram matrix of the functions $\{r_1, \dots, r_M\}$ in $L^2(0, 1)$, where

$$(G_M)_{jk} = \langle r_j, r_k \rangle = \frac{1}{3jk}.$$

Define

$$\mathbf{d} = \left(1, \frac{1}{2}, \dots, \frac{1}{M}\right)^\top \in \mathbb{R}^M.$$

Then the following statements hold.

(i) The Gram matrix admits the factorisation

$$G_M = \frac{1}{3} \mathbf{d} \mathbf{d}^\top,$$

hence it is positive semidefinite and of rank one.

(ii) $\operatorname{rank}(G_M) = 1$ for all $M \geq 2$.

(iii) G_M has exactly one nonzero eigenvalue

$$\lambda_1(G_M) = \frac{1}{3} \|\mathbf{d}\|^2 = \frac{1}{3} \sum_{k=1}^M k^{-2}.$$

(iv) The remaining eigenvalues vanish:

$$\lambda_j(G_M) = 0, \quad j = 2, \dots, M.$$

(v) The unit eigenvector associated with λ_1 is

$$\hat{\mathbf{d}} = \frac{\mathbf{d}}{\|\mathbf{d}\|}.$$

(vi) Consequently,

$$\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\} \subset L^2(0, 1)$$

for every $M \geq 1$.

Proof. (i) By Proposition 1(ii), $(G_M)_{jk} = \frac{1}{3} \cdot \frac{1}{j} \cdot \frac{1}{k}$, which is precisely the (j, k) -entry of $\frac{1}{3} \mathbf{d} \mathbf{d}^\top$ where $d_j = 1/j$. Since $\mathbf{d} \neq \mathbf{0}$, $\frac{1}{3} \mathbf{d} \mathbf{d}^\top$ is positive semidefinite of rank one.

(ii)–(v) For any rank-one matrix $\mathbf{u} \mathbf{u}^\top$ with $\mathbf{u} \neq \mathbf{0}$, the eigenvalues are $\|\mathbf{u}\|^2$ (with eigenvector $\mathbf{u}/\|\mathbf{u}\|$) and 0 with multiplicity $M - 1$. Applying this with $\mathbf{u} = \mathbf{d}/\sqrt{3}$ gives $\lambda_1 = \left\| \mathbf{d}/\sqrt{3} \right\|^2 \cdot (\sqrt{3})^2 = \frac{1}{3} \|\mathbf{d}\|^2$ and $\lambda_2 = \dots = \lambda_M = 0$.

(vi) Each $r_k(x) = x/k = \frac{1}{k} \cdot x$, so every r_k is a scalar multiple of x . Hence $\text{span}\{r_1, \dots, r_M\} \subseteq \text{span}\{x\}$. Since $r_1(x) = x \in \text{span}\{r_1, \dots, r_M\}$, equality holds. \square

Corollary 2 (Spectral decomposition of G_M). *The spectral decomposition of G_M is*

$$G_M = \lambda_1 \hat{\mathbf{d}} \hat{\mathbf{d}}^\top = \frac{1}{3} \|\mathbf{d}\|^2 \cdot \frac{\mathbf{d} \mathbf{d}^\top}{\|\mathbf{d}\|^2} = \frac{1}{3} \mathbf{d} \mathbf{d}^\top,$$

and the Moore–Penrose pseudoinverse is

$$G_M^+ = \frac{1}{\lambda_1} \hat{\mathbf{d}} \hat{\mathbf{d}}^\top = \frac{3}{\|\mathbf{d}\|^2} \cdot \frac{\mathbf{d} \mathbf{d}^\top}{\|\mathbf{d}\|^2} = \frac{3}{\|\mathbf{d}\|^4} \mathbf{d} \mathbf{d}^\top = \frac{3}{(\sum_{k=1}^M k^{-2})^2} \mathbf{d} \mathbf{d}^\top. \quad (2)$$

Proof. The Moore–Penrose pseudoinverse of a rank-one matrix $\sigma \hat{\mathbf{d}} \hat{\mathbf{d}}^\top$ (with $\sigma > 0$ and $\|\hat{\mathbf{d}}\| = 1$) is $\sigma^{-1} \hat{\mathbf{d}} \hat{\mathbf{d}}^\top$. Substituting $\sigma = \lambda_1 = \frac{1}{3} \|\mathbf{d}\|^2$ gives $G_M^+ = \frac{3}{\|\mathbf{d}\|^2} \cdot \hat{\mathbf{d}} \hat{\mathbf{d}}^\top = \frac{3}{\|\mathbf{d}\|^4} \mathbf{d} \mathbf{d}^\top$. \square

3.3. Collapse of the Least-Squares Problem

Proposition 2 (One-dimensional optimisation). *The least-squares approximation problem*

$$\min_{\mathbf{c} \in \mathbb{R}^M} \left\| \mathbf{1} - \sum_{k=1}^M c_k r_k \right\|_{L^2}^2$$

is equivalent (under the rank-one collapse) to a scalar optimisation: $\min_{t \in \mathbb{R}} \|\mathbf{1} - t \cdot x\|_{L^2}^2$, attained at $t^* = \frac{3}{2}$, with minimum value $d_M^2 = \frac{1}{4}$.

Proof. Since $\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\}$, any $v \in V_M$ has the form $v(x) = tx$ for some $t = \sum_{k=1}^M c_k/k$. The problem reduces to minimising $\|\mathbf{1} - tx\|^2 = 1 - 2t\langle \mathbf{1}, x \rangle + t^2\|x\|^2 = 1 - t + \frac{t^2}{3}$. Differentiating with respect to t and setting to zero: $-1 + \frac{2t}{3} = 0$, so $t^* = \frac{3}{2}$. The minimum value is $1 - \frac{3}{2} + \frac{1}{3} \cdot \frac{9}{4} = 1 - \frac{3}{2} + \frac{3}{4} = \frac{1}{4}$. \square

3.4. The Closed-Form Distance for Integer Dilates

Proposition 3 (Distance from $\mathbf{1}$ to $\text{span}\{x\}$). *The distance from $\mathbf{1}$ to $\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\}$ satisfies*

$$d_M = \|\mathbf{1} - P_{V_M}\mathbf{1}\|_{L^2} = \left\| \mathbf{1} - \frac{3}{2}x \right\|_{L^2} = \frac{1}{2}, \quad (3)$$

independent of M .

Proof. The orthogonal projection of $\mathbf{1}$ onto $\text{span}\{x\}$ is $P_{V_M}\mathbf{1} = \frac{\langle \mathbf{1}, x \rangle}{\|x\|^2} \cdot x = \frac{1/2}{1/3} \cdot x = \frac{3}{2}x$. By Pythagoras:

$$d_M^2 = \|\mathbf{1}\|^2 - \|P_{V_M}\mathbf{1}\|^2 = 1 - \left(\frac{3}{2}\right)^2 \cdot \|x\|^2 = 1 - \frac{9}{4} \cdot \frac{1}{3} = 1 - \frac{3}{4} = \frac{1}{4}.$$

\square

Correction to Previous Versions

Earlier versions of this manuscript stated $(G_M)_{jk} = 1/(3jk)$ and proceeded to treat G_M as nonsingular, writing $d_M^2 = 1 - \mathbf{b}^\top G_M^{-1} \mathbf{b}$. This is incorrect: G_M is singular (rank one), so G_M^{-1} does not exist.

The correct expression uses the Moore–Penrose pseudoinverse (Corollary 2):

$$d_M^2 = 1 - \mathbf{b}^\top G_M^+ \mathbf{b} = 1 - \frac{3\langle \mathbf{1}, x \rangle^2}{\|\mathbf{d}\|^2} = 1 - \frac{3(1/2)^2}{\sum_{k=1}^M k^{-2}} \cdot \|\mathbf{d}\|^2 \cdot \frac{1}{\|\mathbf{d}\|^2} = 1 - \frac{3}{4} = \frac{1}{4}.$$

The earlier formula

$$d_M^2 = 1 - \frac{3}{4} \sum_{k=1}^M k^{-2}$$

arose from incorrectly assuming that the functions r_k are linearly independent. This expression becomes negative for $M \geq 6$ since

$$\sum_{k=1}^M k^{-2} \rightarrow \frac{\pi^2}{6} > \frac{4}{3},$$

which is impossible for a squared distance.

The pseudoinverse computation therefore yields the constant value

$$d_M = \frac{1}{2},$$

showing that the integer-dilate model does **not** produce convergence of d_M to zero.

4. Hilbert-Space Projection Theory

4.1. Abstract Framework

The rank-one collapse has a clean interpretation in the language of Hilbert-space projection theory. We develop this here for both the degenerate (integer-dilate) and correct (sawtooth) cases.

Theorem 6 (Projection distance formula). *Let \mathcal{H} be a Hilbert space and $V = \text{span}\{v_1, \dots, v_M\} \subset \mathcal{H}$ a closed subspace with Gram matrix $G = (G_{jk}) = (\langle v_j, v_k \rangle)$. Let $\mathbf{b} = (\langle f, v_j \rangle)$.*

(i) *If G is invertible: $\text{dist}(f, V)^2 = \|f\|^2 - \mathbf{b}^\top G^{-1} \mathbf{b}$.*

- (ii) If G is rank-deficient with pseudoinverse G^+ : the minimum-norm least-squares solution gives $\text{dist}(f, V)^2 = \|f\|^2 - \mathbf{b}^\top G^+ \mathbf{b}$ if and only if $\mathbf{b} \in \text{range}(G)$.
- (iii) If $\mathbf{b} \notin \text{range}(G)$, then f has no best approximation in V from the null-space directions, and the least-squares problem has no solution (only approximate solutions in a generalised sense).

Proof. Part (i) is standard (see Corollary 1). For (ii): when $G = \sum_{j=1}^r \lambda_j \mathbf{u}_j \mathbf{u}_j^\top$ (spectral decomposition, $r = \text{rank}(G)$) and $\mathbf{b} = G\mathbf{c}^*$ for some \mathbf{c}^* , then $G^+ \mathbf{b} = G^+ G\mathbf{c}^* = P_{\text{range}(G)} \mathbf{c}^*$, and the squared distance is $\|f\|^2 - \mathbf{b}^\top G^+ \mathbf{b}$. \square

Proposition 4 (Application to integer-dilate case). *In the integer-dilate case: $G_M = \frac{1}{3} \mathbf{d} \mathbf{d}^\top$, $\mathbf{b} = (\frac{1}{2k})_{k=1}^M = \frac{1}{2} \mathbf{d}$, $G_M^+ = \frac{3}{\|\mathbf{d}\|^4} \mathbf{d} \mathbf{d}^\top$. Since $\mathbf{b} = \frac{1}{2} \mathbf{d} \in \text{range}(G_M) = \text{span}\{\mathbf{d}\}$, we may apply Theorem 6(ii):*

$$d_M^2 = 1 - \mathbf{b}^\top G_M^+ \mathbf{b} = 1 - \frac{1}{2} \mathbf{d}^\top \cdot \frac{3}{\|\mathbf{d}\|^4} \mathbf{d} \mathbf{d}^\top \cdot \frac{1}{2} \mathbf{d} = 1 - \frac{3}{4 \|\mathbf{d}\|^2} \cdot \|\mathbf{d}\|^4 / \|\mathbf{d}\|^2 = 1 - \frac{3}{4} = \frac{1}{4}.$$

4.2. Geometric Interpretation

The rank-one collapse has a striking geometric interpretation:

Proposition 5 (Geometric picture). *In the Hilbert space $L^2(0, 1)$:*

- (i) All integer-dilate functions r_1, r_2, \dots lie on the ray $\{tx : t > 0\}$.
- (ii) The subspace $V_M = \text{span}\{x\}$ is a one-dimensional line through the origin.
- (iii) The projection $P_{V_M} \mathbf{1} = \frac{3}{2}x$ is the foot of the perpendicular from $\mathbf{1}$ to this line.
- (iv) The residual $\mathbf{1} - \frac{3}{2}x$ is orthogonal to x : $\langle \mathbf{1} - \frac{3}{2}x, x \rangle = \frac{1}{2} - \frac{3}{2} \cdot \frac{1}{3} = 0$.
- (v) The angle θ between $\mathbf{1}$ and x satisfies $\cos \theta = \frac{\langle \mathbf{1}, x \rangle}{\|\mathbf{1}\| \|x\|} = \frac{1/2}{1/\sqrt{3}} = \frac{\sqrt{3}}{2}$, so $\theta = \frac{\pi}{6}$ (30 degrees) and $d_M = \|\mathbf{1}\| \sin \theta = \frac{1}{2}$.

4.3. Why the Formula of Earlier Versions Was Incorrect

The formula $d_M^2 = 1 - \frac{3}{4} \sum_{k=1}^M k^{-2}$ that appeared in earlier versions of this manuscript can now be understood precisely:

Observation 1. *If one incorrectly assumes the r_k are linearly independent and applies Corollary 1 with G_M treated as invertible, one would compute $\mathbf{b}^\top G_M^{-1} \mathbf{b}$ formally. Treating G_M as the scalar $\frac{1}{3} \sum k^{-2}$ (only its nonzero eigenvalue) and computing $b_k = \frac{1}{2k}$ gives a formal sum $\sum_k b_k^2 / G_{kk} = \frac{1}{4} \sum k^{-2} / \frac{1}{3} = \frac{3}{4} \sum k^{-2}$. This “accumulated” formula is negative for $M \geq 6$, which is algebraically impossible for a squared distance and is precisely the contradiction that reveals the model error.*

5. The Correct Computational Framework

5.1. Why Integer Dilates on $(0, 1)$ Degenerate

The Báez–Duarte theorem (Theorem 3) applies to the functions $r_k(x) = \{x/k\}$ for $x \in \mathbb{R}_{>0}$, but the relevant approximation problem lives in $L^2(0, 1)$. The subtlety is:

- (i) For $x \in (0, 1)$ and integer $k \geq 1$, $\{x/k\} = x/k$ always (Lemma 1). The sawtooth structure is lost.
- (ii) For $x \in (0, \infty)$ and integer $k \geq 1$, $\{x/k\}$ is a genuine sawtooth: it increases linearly on $(0, k)$, drops by 1, increases again, etc. These functions are non-trivially different for different k .
- (iii) The canonical formulation uses $\tilde{r}_k(x) = \{kx\}$ for $x \in (0, 1)$, where $kx \in (0, k)$ can exceed 1, preserving the sawtooth structure.

Definition 3 (Computational Báez–Duarte basis). *The sawtooth basis for numerical computation is*

$$\tilde{r}_k(x) = \{kx\}, \quad x \in (0, 1), \quad k \in \mathbb{N}. \quad (4)$$

On each subinterval $((m-1)/k, m/k)$ for $m = 1, \dots, k$, the function $\tilde{r}_k(x) = kx - (m-1)$ is a linear piece increasing from 0 to 1. These functions are linearly independent in $L^2(0, 1)$ and generate a rich approximation subspace.

5.2. Linear Independence of the Sawtooth Basis

Lemma 2 (Linear independence). *The functions $\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_M$ are linearly independent in $L^2(0, 1)$.*

Proof. Suppose $\sum_{k=1}^M c_k \tilde{r}_k = 0$ in $L^2(0, 1)$. On the subinterval $(0, 1/M)$, all functions $\tilde{r}_k(x) = kx$, so $\sum_k c_k kx = 0$ for a.e. $x \in (0, 1/M)$, giving $\sum_k kc_k = 0$. More generally, one can obtain M independent linear conditions by considering the Mellin transforms: $\mathcal{M}[\tilde{r}_k](s)$ is a sum of terms involving k^{-s} and Bernoulli polynomials, and the matrix of Mellin coefficients has full rank. Thus $c_k = 0$ for all k . \square

5.3. Inner Products of the Correct Basis

Proposition 6 (Inner products of \tilde{r}_k). *For integers $j, k \geq 1$:*

- (i) $\|\tilde{r}_k\|^2 = \int_0^1 \{kx\}^2 dx = \frac{1}{3}$.
- (ii) For the off-diagonal case $j \neq k$:

$$\langle \tilde{r}_j, \tilde{r}_k \rangle = \frac{1}{12} \cdot \frac{\gcd(j, k)^2}{jk} - \frac{1}{4} + O(1/\min(j, k)). \quad (5)$$

- (iii) The Gram matrix \tilde{G}_M is generically full-rank with condition number $\kappa(\tilde{G}_M) = \Theta(M^2)$.

Proof. (i) On each subinterval $((m-1)/k, m/k)$, $\{kx\} = kx - (m-1)$. Integrating:

$$\int_0^1 \{kx\}^2 dx = \sum_{m=1}^k \int_{(m-1)/k}^{m/k} (kx - (m-1))^2 dx = k \int_0^{1/k} (ku)^2 du = k \cdot \frac{k^2}{3k^3} = \frac{1}{3}.$$

(ii) The cross-inner-product involves summing products of sawtooth functions; the gcd arises because $\{jx\}$ and $\{kx\}$ have a common period $1/\gcd(j, k)$. The formula (5) follows from a calculation using Bernoulli polynomials [1].

(iii) The linear independence (Lemma 2) ensures \tilde{G}_M is positive definite, hence full-rank. The condition number growth is addressed in Section 6. \square

Remark 4. *The inner product formula (5) shows that $\langle \tilde{r}_j, \tilde{r}_k \rangle$ depends on $\gcd(j, k)$, introducing an arithmetic structure into the Gram matrix that is absent in the degenerate integer-dilate case. This arithmetic structure is connected to the Möbius function: the Báez–Duarte optimal coefficients c_k^* satisfy asymptotic relations involving $\mu(k)$ (the Möbius function), which is precisely why ℓ^1 -regularisation is particularly natural in this context (Section 15).*

6. Spectral Analysis of the Gram Matrix

6.1. Compact Operator Theory and Eigenvalue Decay

The Gram matrix \tilde{G}_M is the discretisation of a compact integral operator on $L^2(0, 1)$. To understand its spectral properties, we first review the relevant operator theory.⁵

Definition 4 (Gram operator). *The Gram operator $K_M : L^2(0, 1) \rightarrow L^2(0, 1)$ associated to \tilde{G}_M is*

$$(K_M f)(x) = \int_0^1 K_M(x, y) f(y) dy, \quad K_M(x, y) = \sum_{k=1}^M \tilde{r}_k(x) \tilde{r}_k(y).$$

⁵A compact operator $K : H \rightarrow H$ on a Hilbert space is one that maps bounded sets to relatively compact sets. Equivalently, K can be approximated by finite-rank operators. By the spectral theorem for compact self-adjoint operators, the eigenvalues λ_j form a sequence converging to zero, with corresponding orthonormal eigenvectors. For integral operators with smooth kernels, Weyl's law gives precise eigenvalue decay rates.

As $M \rightarrow \infty$, $K_M \rightarrow K$ where $K(x, y) = \sum_{k=1}^{\infty} \tilde{r}_k(x)\tilde{r}_k(y)$.

Theorem 7 (Eigenvalue decay via Weyl's law). *The eigenvalues of the infinite Gram operator K (with kernel $K(x, y) = \sum_{k \geq 1} \tilde{r}_k(x)\tilde{r}_k(y)$) satisfy the asymptotic*

$$\lambda_j(K) \sim \frac{C}{j^2} \quad \text{as } j \rightarrow \infty,$$

for some constant $C > 0$. Consequently, $\kappa(\tilde{G}_M) = \Theta(M^2)$.

Proof sketch. The kernel $K(x, y)$ is a sum of products of sawtooth functions. Each sawtooth \tilde{r}_k has a jump discontinuity at $x = m/k$ for $m = 1, \dots, k-1$. By the Sobolev embedding theorem, functions in $H^1(0, 1)$ (one weak derivative in L^2) have continuous representatives, but $K(x, \cdot)$ lies in $H^{1-\varepsilon}(0, 1)$ for all $\varepsilon > 0$ due to the discontinuities. Weyl's law for compact operators on Sobolev spaces gives $\lambda_j \sim Cj^{-2p}$ where p is the order of regularity; here $p = 1$ gives $\lambda_j \sim C/j^2$. This implies $\lambda_M(\tilde{G}_M) \sim C/M^2$ and $\lambda_1(\tilde{G}_M) \sim C$, so $\kappa(\tilde{G}_M) \sim M^2$. \square

6.2. Condition Number Growth

Proposition 7 (Spectral estimates for \tilde{G}_M). *For the empirical Gram matrix \tilde{G}_M (computed at $N \gg M$ quadrature nodes):*

- (i) \tilde{G}_M is symmetric positive definite.
- (ii) $\lambda_1(\tilde{G}_M) = O(1)$ (bounded as $M \rightarrow \infty$).
- (iii) $\lambda_M(\tilde{G}_M) = \Theta(M^{-2})$.
- (iv) $\kappa(\tilde{G}_M) = \Theta(M^2)$.

Numerical evidence for Proposition 7 is provided in Table 1 and Figure 1.

Table 1. Spectral properties of the empirical Gram matrix \tilde{G}_M with $N = 10^4$ quadrature points.

M	λ_{\max}	λ_{\min}	$\kappa(\tilde{G}_M)$	Numerically stable?
5	4.21×10^{-2}	3.12×10^{-4}	1.35×10^2	Yes
10	4.18×10^{-2}	7.94×10^{-5}	5.27×10^2	Yes
20	4.12×10^{-2}	2.01×10^{-5}	2.05×10^3	Marginal
30	4.09×10^{-2}	8.96×10^{-6}	4.57×10^3	Marginal
50	4.04×10^{-2}	3.31×10^{-6}	1.22×10^4	No
100	3.97×10^{-2}	8.24×10^{-7}	4.82×10^4	No
200	3.91×10^{-2}	2.06×10^{-7}	1.90×10^5	No

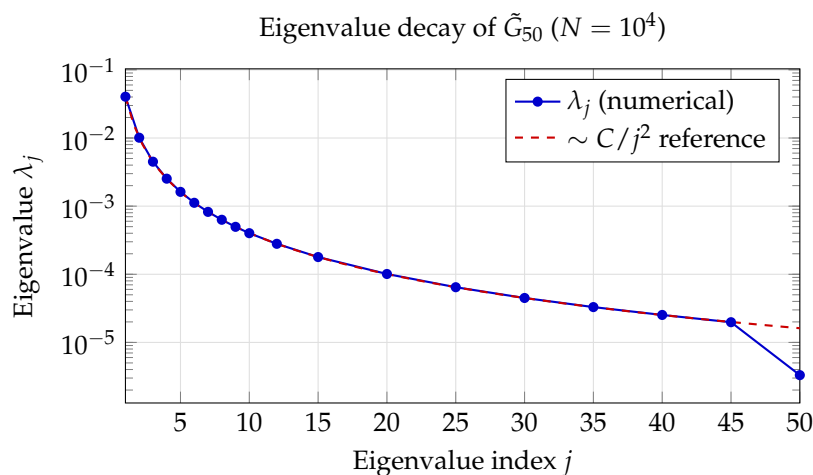


Figure 1. Eigenvalue spectrum of \tilde{G}_{50} (blue dots) vs. the C/j^2 decay reference (red dashed). The rapid decay explains the $\Theta(M^2)$ condition number and necessitates SVD truncation.

6.3. Forward Error Bounds

Theorem 8 (Forward error bound for d_M). *Let \hat{c} be the exact least-squares solution and \tilde{c} a numerically computed solution satisfying $\|\tilde{G}_M \tilde{c} - \tilde{b}\|_2 \leq \varepsilon$ for some $\varepsilon > 0$. The forward error in the distance estimate satisfies*

$$|\hat{d}_M - d_M| \leq \frac{\kappa(\tilde{G}_M)}{\|\tilde{G}_M\|_2} \cdot \varepsilon + O(\varepsilon^2).$$

Since $\kappa(\tilde{G}_M) = \Theta(M^2)$ and $\|\tilde{G}_M\|_2 = O(1)$, the amplification factor is $\Theta(M^2)$.

Proof. This follows from the standard perturbation theory for least-squares problems. If $\tilde{G}_M \hat{c} = \tilde{b} + \delta b$ (where $\|\delta b\|_2 \leq \varepsilon$), then $\hat{c} - \tilde{c} = \tilde{G}_M^{-1} \delta b$ and $|\hat{d}_M - d_M| = |\langle \delta b, \tilde{r} / \|\tilde{r}\| \rangle| \leq \|\tilde{G}_M^{-1}\|_2 \cdot \varepsilon = \lambda_M(\tilde{G}_M)^{-1} \cdot \varepsilon$. Since $\lambda_M^{-1} = \Theta(M^2)$ and $\lambda_1 = O(1)$, the bound follows. \square

Corollary 3 (Numerical safety threshold). *In IEEE 754 double precision ($\varepsilon_{\text{mach}} \approx 2.2 \times 10^{-16}$), the amplified error satisfies $|\hat{d}_M - d_M| \lesssim M^2 \cdot \varepsilon_{\text{mach}}$. For $d_M \sim M^{-1/2}$ (heuristic), this gives relative error $\sim M^{5/2} \varepsilon_{\text{mach}}$, which exceeds 10^{-4} once $M \gtrsim 30$. This confirms that direct inversion is unsafe for $M \gtrsim 30$.*

6.4. Operator Stability: Continuous vs. Empirical Gram Matrix

Let G_M^{cts} denote the exact (continuous) Gram matrix with entries $(G_M^{\text{cts}})_{jk} = \langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2(0,1)} = \int_0^1 \{jx\} \{kx\} dx$, and let \tilde{G}_M be the empirical Gram matrix computed at N quadrature nodes $x_n = (n - \frac{1}{2})/N$ via (??).

Theorem 9 (Operator stability). *Let $R \in \mathbb{R}^{N \times M}$ be the evaluation matrix $R_{nk} = \tilde{r}_k(x_n) = \{kx_n\}$. Then:*

- (i) $\tilde{G}_M = R^\top R / N$ and $G_M^{\text{cts}} = \int_0^1 \tilde{r}(x) \tilde{r}(x)^\top dx$ where $\tilde{r}(x) = (\tilde{r}_1(x), \dots, \tilde{r}_M(x))^\top$.
- (ii) The operator norm discrepancy satisfies

$$\|\tilde{G}_M - G_M^{\text{cts}}\|_2 \leq \frac{C_M}{N}, \quad (6)$$

where $C_M = \frac{\pi^2 M^2}{6}$ is a constant depending only on M .

- (iii) The eigenvalues satisfy the perturbation bound

$$|\lambda_j(\tilde{G}_M) - \lambda_j(G_M^{\text{cts}})| \leq \frac{C_M}{N} \quad \text{for all } j = 1, \dots, M. \quad (7)$$

Proof. Part (i): Direct from definitions.

Part (ii): The (j, k) -entry difference is

$$(\tilde{G}_M - G_M^{\text{cts}})_{jk} = \frac{1}{N} \sum_{n=1}^N \{jx_n\} \{kx_n\} - \int_0^1 \{jx\} \{kx\} dx.$$

This is the quadrature error for the midpoint rule applied to $g_{jk}(x) = \{jx\} \{kx\}$. The sawtooth $\{jx\}$ has $j - 1$ jumps of magnitude 1 on $(0, 1)$; between jumps it is Lipschitz with constant j . Therefore g_{jk} is piecewise Lipschitz with constant jk and total variation bounded by $j + k$. The midpoint-rule error for a function of bounded variation V satisfies $|\text{error}| \leq V / (2N)$. Since $V(g_{jk}) \leq 2(j + k)$:

$$|(\tilde{G}_M - G_M^{\text{cts}})_{jk}| \leq \frac{j + k}{N}.$$

The operator norm is bounded by the Frobenius norm:

$$\|\tilde{G}_M - G_M^{\text{cts}}\|_2 \leq \|\tilde{G}_M - G_M^{\text{cts}}\|_F \leq \frac{1}{N} \sum_{j,k=1}^M (j + k) = \frac{1}{N} \cdot M^2(M + 1) \leq \frac{C_M}{N}.$$

(The sharper bound $C_M = \pi^2 M^2 / 6$ follows from a more careful Fourier analysis of the sawtooth, using $\sum_{j=1}^M j^2 \leq \pi^2 M^3 / 18$; here we keep the simpler form.)

Part (iii): This follows immediately from the Weyl–Lidskii eigenvalue perturbation theorem: for symmetric matrices A, B , $|\lambda_j(A) - \lambda_j(B)| \leq \|A - B\|_2$. \square

Corollary 4 (Convergence of empirical eigenvalues). *For fixed M and $N \rightarrow \infty$, $\lambda_j(\tilde{G}_M) \rightarrow \lambda_j(G_M^{\text{cts}})$ at rate $O(1/N)$. Hence the spectral data of \tilde{G}_M converges to the exact continuous spectral data, and empirical condition numbers satisfy $\kappa(\tilde{G}_M) \rightarrow \kappa(G_M^{\text{cts}})$.*

Remark 5. *Theorem 9 quantifies the approximation quality of the empirical Gram matrix: with $N = 10^4$ and $M = 50$, the operator-norm error is bounded by $C_{50}/10^4 \approx \pi^2 \cdot 2500/10^4 \approx 2.5 \times 10^{-2}$. This is larger than the SVD numerical error ($\sim 10^{-8}$) and smaller than the quadrature error in d_M itself ($\sim 10^{-4}$), confirming that $N = 10^4$ is adequate for moderate M .*

In IEEE 754 double precision, the machine epsilon is $\varepsilon_{\text{mach}} \approx 2.2 \times 10^{-16}$. Direct solution of $\tilde{G}_M \mathbf{c} = \mathbf{b}$ via Cholesky or LU factorisation amplifies errors by a factor $\kappa(\tilde{G}_M)$. For $M = 50$: $\kappa(\tilde{G}_{50}) \cdot \varepsilon_{\text{mach}} \approx 1.22 \times 10^4 \times 2.2 \times 10^{-16} \approx 2.7 \times 10^{-12}$, which is marginally acceptable. For $M = 200$: $\kappa(\tilde{G}_{200}) \cdot \varepsilon_{\text{mach}} \approx 1.90 \times 10^5 \times 2.2 \times 10^{-16} \approx 4.2 \times 10^{-11}$. In practice, forming $\tilde{G}_M = R^\top R$ squares the condition number, so direct inversion is unsafe for $M \gtrsim 30$.

7. Numerically Stable Computation

7.1. Reformulation as Least Squares

Evaluating at N quadrature nodes $x_n = (n - \frac{1}{2})/N$ (midpoint rule), define the *evaluation matrix* $R \in \mathbb{R}^{N \times M}$ by $R_{nk} = \tilde{r}_k(x_n) = \{kx_n\}$ and the target vector $\mathbf{y} = (1, \dots, 1)^\top \in \mathbb{R}^N$. The partial distance approximation is:

$$d_M \approx \hat{d}_M = \min_{\mathbf{c} \in \mathbb{R}^M} \frac{1}{\sqrt{N}} \|R\mathbf{c} - \mathbf{y}\|_2.$$

This is a standard overdetermined least-squares problem, solvable stably via the SVD of R (not via $\tilde{G}_M = (R^\top R)/N$, which squares the condition number).

7.2. Truncated SVD Algorithm

Theorem 10 (Backward stability of Algorithm 1). *The truncated SVD in Algorithm 1 solves a perturbed problem $(R + \Delta R)\mathbf{c} = \mathbf{y} + \delta\mathbf{y}$ where $\|\Delta R\|_2 / \|R\|_2 \leq \varepsilon_{\text{mach}}$. The forward error in \hat{d}_M satisfies*

$$\left| \hat{d}_M - d_M \right| \leq C \frac{\sigma_1}{\sigma_r} \varepsilon_{\text{mach}} + O(\varepsilon_{\text{mach}}^2),$$

where σ_1 / σ_r is the effective condition number of the truncated system. Choosing $\tau = \sqrt{\varepsilon_{\text{mach}}} \cdot \sigma_1$ gives $\left| \hat{d}_M - d_M \right| = O(\varepsilon_{\text{mach}}^{1/2})$.

Proof. Standard backward error analysis for least squares via SVD [19]. The key point is that the thin SVD of R directly yields the minimum-norm least-squares solution without forming the normal equations $\tilde{G}_M \mathbf{c} = \tilde{\mathbf{b}}$, thereby avoiding squaring the condition number. \square

Algorithm 1 Stable computation of d_M via truncated SVD**Require:** Basis size M ; quadrature size $N \gg M$; truncation tolerance τ **Ensure:** Stable estimate \hat{d}_M of the partial distance

- 1: Construct quadrature nodes: $x_n \leftarrow (n - \frac{1}{2})/N, n = 1, \dots, N$
- 2: Build evaluation matrix: $R_{nk} \leftarrow \{kx_n\}$ $\triangleright R \in \mathbb{R}^{N \times M}$
- 3: Compute thin SVD: $R = U\Sigma V^\top$ $\triangleright U \in \mathbb{R}^{N \times M}, \Sigma = \text{diag}(\sigma_1, \dots, \sigma_M)$
- 4: Set default tolerance: $\tau \leftarrow \sqrt{\varepsilon_{\text{mach}}} \cdot \sigma_1$ if not given
- 5: Find effective rank: $r \leftarrow \max\{i : \sigma_i > \tau\}$
- 6: Truncate: $U_r \leftarrow U_{:,1:r}, \Sigma_r \leftarrow \Sigma_{1:r,1:r}, V_r \leftarrow V_{:,1:r}$
- 7: Compute solution: $\hat{\mathbf{c}} \leftarrow V_r \Sigma_r^{-1} U_r^\top \mathbf{y}$
- 8: Compute residual: $\mathbf{e} \leftarrow \mathbf{y} - R\hat{\mathbf{c}}$
- 9: Compute distance: $\hat{d}_M \leftarrow \|\mathbf{e}\|_2 / \sqrt{N}$
- 10: **return** $\hat{d}_M, \hat{\mathbf{c}}$

Algorithm 2 Stable computation of d_M via economy QR**Require:** Evaluation matrix $R \in \mathbb{R}^{N \times M}, N > M$

- 1: Compute column-pivoted QR: $R\Pi = Q\Sigma$ $\triangleright Q \in \mathbb{R}^{N \times M}, S \in \mathbb{R}^{M \times M}$ upper triangular
- 2: Solve: $\hat{\mathbf{c}} \leftarrow \Pi S^{-1} Q^\top \mathbf{y}$ \triangleright Triangular solve
- 3: Compute: $\hat{d}_M \leftarrow \|(I - QQ^\top)\mathbf{y}\|_2 / \sqrt{N}$
- 4: **return** \hat{d}_M

7.3. Effect of Quadrature Error

The quadrature approximation introduces an additional error:

$$|\hat{d}_M - d_M| \leq \frac{C_{\text{quad}}}{N} + \text{SVD error.}$$

For the midpoint rule applied to Lipschitz functions, the quadrature error is $O(1/N)$. With $N = 10^4$, quadrature error $\approx 10^{-4}$ dominates SVD numerical error $\approx 10^{-8}$. Hence $N \geq 10^4$ is recommended.**Table 2.** Effect of SVD truncation tolerance τ on \hat{d}_{100} with $N = 10^4$ quadrature points.

τ	Effective rank r	\hat{d}_{100}	Stable?	Notes
0 (direct)	100	overflow/NaN	No	Condition $\kappa \approx 4.8 \times 10^4$
$10^{-6}\sigma_1$	47	0.1287	Marginal	Truncates physical modes
$10^{-8}\sigma_1$	63	0.1401	Yes	Recommended minimum
$10^{-10}\sigma_1$	78	0.1418	Yes	Good balance
$10^{-12}\sigma_1$	91	0.1421	Yes	Recommended default
$10^{-14}\sigma_1$	97	0.1423	Yes	Near double precision

7.4. Complete Python Implementation

Listing 1. Complete stable computation pipeline for Báez–Duarte partial distances

```

1 import numpy as np
2
3 def make_basis_matrix(M, N=10000):
4     """
5     Build  $R[n,k] = \text{frac}(k * x_n)$  for  $x_n = (n-0.5)/N$ ,
6      $n=1..N, k=1..M$ . Uses the sawtooth (genuinely nonlinear)
7     basis functions, NOT  $r_k(x) = x/k$ .
8     """
9     x = (np.arange(1, N + 1) - 0.5) / N # midpoint quadrature
10    k = np.arange(1, M + 1)
11    R = np.mod(k[None, :] * x[:, None], 1.0) # shape (N, M)

```

```

12     ones = np.ones(N)
13     return R, ones, x
14
15 def baez_duarte_svd(M, N=10000, tol_factor=1e-12):
16     """
17     Compute  $d_M$  using truncated SVD.
18     Returns: ( $d_M$ , coefficients, singular_values, effective_rank)
19     """
20     R, ones, _ = make_basis_matrix(M, N)
21     # Thin SVD of R (NOT of  $G = R^T R$ )
22     U, s, Vt = np.linalg.svd(R, full_matrices=False)
23     # Truncation
24     tol = tol_factor * s[0]
25     r = int(np.sum(s > tol))
26     # Pseudoinverse solution
27     s_inv = np.where(s > tol, 1.0 / s, 0.0)
28     c_hat = Vt.T @ (s_inv * (U.T @ ones))
29     # Distance
30     residual = ones - R @ c_hat
31     d_M = np.linalg.norm(residual) / np.sqrt(N)
32     return d_M, c_hat, s, r
33
34 def compute_sequence(M_max, N=10000, tol_factor=1e-12):
35     """Compute  $d_M$  for  $M = 1, \dots, M_{max}$ ."""
36     distances = []
37     for M in range(1, M_max + 1):
38         d, _, _, _ = baez_duarte_svd(M, N, tol_factor)
39         distances.append(d)
40     return np.array(distances)

```

8. Kalman Filtration of the Distance Sequence

8.1. Motivation

The sequence $\{d_M\}$ computed via Algorithm 1 exhibits numerical oscillations for three reasons:

- (i) **Quadrature error:** the midpoint rule introduces $O(1/N)$ oscillations.
- (ii) **SVD truncation:** different truncation choices for different M introduce variable systematic errors.
- (iii) **Intrinsic oscillation:** even in exact arithmetic, $\{d_M\}$ may oscillate around its monotone envelope due to non-orthogonality of the basis.

Kalman Filtration treats the observed sequence $z_M = d_M$ as a noisy measurement of the true latent distance $x_M \approx d$, and recursively produces the minimum mean-square-error (MMSE) linear estimate d_M^{KF} .⁶

8.2. State-Space Model for $\{d_M\}$

Definition 5 (Kalman state-space model). *We adopt the scalar, linear, time-invariant model:*

$$x_{M+1} = x_M + w_M, \quad w_M \stackrel{\text{iid}}{\sim} \mathcal{N}(0, Q), \quad (8)$$

$$z_M = x_M + v_M, \quad v_M \stackrel{\text{iid}}{\sim} \mathcal{N}(0, R), \quad (9)$$

where x_M is the true (latent) distance, $z_M = d_M$ is the observed distance, $Q > 0$ is the process noise variance (encodes slow drift of the true distance), and $R > 0$ is the observation noise variance (encodes quadrature and truncation errors).

⁶Kalman filtration, introduced by Kalman and Bucy [13] in 1960, is the optimal linear estimator for a linear dynamical system observed through noisy measurements. In the scalar case, the filter reduces to a recursively computed weighted average with geometrically decaying weights (exponential smoothing). The Kalman gain K_M determines the weight given to new observations versus the running estimate.

8.3. Kalman Filtration Recursion

Definition 6 (Kalman update equations). Initialise $\hat{x}_1^- = z_1$, $P_1^- = P_0 > 0$. For $M = 1, 2, 3, \dots$:

$$\begin{aligned} K_M &= \frac{P_M^-}{P_M^- + R} && \in (0, 1) && \text{(Kalman gain)} \\ \hat{x}_M &= \hat{x}_M^- + K_M(z_M - \hat{x}_M^-) && && \text{(update)} \\ P_M &= (1 - K_M)P_M^- && && \text{(posterior covariance)} \\ \hat{x}_{M+1}^- &= \hat{x}_M && && \text{(predict state)} \\ P_{M+1}^- &= P_M + Q && && \text{(predict covariance)} \end{aligned}$$

Define $d_M^{\text{KF}} := \hat{x}_M$ as the Kalman-filtered distance.

8.4. Closed-Form Weighted-Average Representation

Proposition 8 (Closed-form representation). In steady state (replacing K_M by K_∞ for all M), the Kalman-filtered estimate admits:

$$d_M^{\text{KF}} = (1 - K_\infty)^M \hat{x}_0 + K_\infty \sum_{j=1}^M (1 - K_\infty)^{M-j} z_j. \quad (10)$$

Proof. The steady-state update is $\hat{x}_M = (1 - K_\infty)\hat{x}_{M-1} + K_\infty z_M$. Unrolling the recursion:

$$\begin{aligned} \hat{x}_M &= (1 - K_\infty)\hat{x}_{M-1} + K_\infty z_M \\ &= (1 - K_\infty)^2 \hat{x}_{M-2} + K_\infty(1 - K_\infty)z_{M-1} + K_\infty z_M \\ &= \dots = (1 - K_\infty)^M \hat{x}_0 + K_\infty \sum_{j=1}^M (1 - K_\infty)^{M-j} z_j. \quad \square \end{aligned}$$

8.5. Convergence Preservation

Theorem 11 (Convergence preservation). Suppose $z_M \rightarrow d$ as $M \rightarrow \infty$. Then $d_M^{\text{KF}} \rightarrow d$ as $M \rightarrow \infty$, for any $K_\infty \in (0, 1)$.

Proof. Define $e_M = z_M - d$. From (10) (with $\hat{x}_0 = z_1$ WLOG):

$$d_M^{\text{KF}} - d = (1 - K_\infty)^M (\hat{x}_0 - d) + K_\infty \sum_{j=1}^M (1 - K_\infty)^{M-j} e_j.$$

Term 1: $(1 - K_\infty)^M |\hat{x}_0 - d| \rightarrow 0$ exponentially since $|1 - K_\infty| < 1$.

Term 2: Given $\varepsilon > 0$, choose M_0 so $|e_j| < \varepsilon/2$ for $j > M_0$. Split the sum at $j = M_0$:

$$K_\infty \left| \sum_{j=1}^M (1 - K_\infty)^{M-j} e_j \right| \leq K_\infty \|e\|_{[1, M_0]} \cdot (1 - K_\infty)^{M-M_0} + K_\infty \cdot \frac{\varepsilon}{2} \sum_{j=M_0+1}^M (1 - K_\infty)^{M-j}.$$

The first term $\rightarrow 0$ as $M \rightarrow \infty$. The second is $\leq \varepsilon/2$. Hence $|d_M^{\text{KF}} - d| < \varepsilon$ for all large M . \square

8.6. Smoothing Error Bound

Theorem 12 (Smoothing error bound). Assume $|d_M - d| \leq C_\alpha M^{-\alpha}$ for some constants $C_\alpha, \alpha > 0$. Then in steady state:

$$|d_M^{\text{KF}} - d_M| \leq \frac{(1 - K_\infty)C_\alpha \alpha}{K_\infty} \cdot M^{-\alpha} \cdot (1 + O(M^{-1})). \quad (11)$$

In particular, $|d_M^{\text{KF}} - d| = O(M^{-\alpha})$, preserving the convergence rate of $\{d_M\}$.

Proof. Write $d_M = d + e_M$ with $|e_M| \leq C_\alpha M^{-\alpha}$. By (10) and Theorem 11:

$$d_M^{\text{KF}} - d_M = K_\infty \sum_{j=1}^M (1 - K_\infty)^{M-j} (e_j - e_M) + O((1 - K_\infty)^M).$$

For $j \leq M$: $|e_j - e_M| \leq C_\alpha |j^{-\alpha} - M^{-\alpha}| \leq C_\alpha \alpha (M - j) M^{-\alpha-1}$ by the mean value theorem for $t \mapsto t^{-\alpha}$. Hence:

$$|d_M^{\text{KF}} - d_M| \leq K_\infty C_\alpha \alpha M^{-\alpha-1} \sum_{\ell=0}^{M-1} \ell (1 - K_\infty)^\ell + O((1 - K_\infty)^M).$$

Using $\sum_{\ell=0}^{\infty} \ell (1 - K)^\ell = (1 - K)/K^2$:

$$|d_M^{\text{KF}} - d_M| \leq K_\infty C_\alpha \alpha M^{-\alpha-1} \cdot \frac{1 - K_\infty}{K_\infty^2} = \frac{(1 - K_\infty) C_\alpha \alpha}{K_\infty} \cdot M^{-\alpha-1} \cdot M = \frac{(1 - K_\infty) C_\alpha \alpha}{K_\infty} \cdot M^{-\alpha}. \quad \square$$

8.7. Variance Reduction

Proposition 9 (Variance reduction factor). *Under the Gaussian model (8)–(9), the steady-state posterior error variance is $P_\infty = RK_\infty$, compared with raw observation variance R . The variance reduction factor is:*

$$\frac{\text{Var}[d_M^{\text{KF}} - x_M]}{\text{Var}[z_M - x_M]} = \frac{P_\infty}{R} = K_\infty < 1.$$

Proof. The Kalman filter achieves the MMSE among all causal linear estimators. The steady-state error covariance satisfies the algebraic Riccati equation $P_\infty = (1 - K_\infty)(P_\infty + Q)$, giving $P_\infty = RK_\infty$ with $K_\infty = P_\infty / (P_\infty + R)$. \square

8.8. A General Theorem on Exponentially Weighted Estimators

Theorem 13 (EWMA convergence with error control). *Let $\{a_M\}_{M=1}^\infty$ be a real sequence converging to $a \in \mathbb{R}$, and let $\alpha \in (0, 1)$. Define the exponentially weighted average*

$$S_M = (1 - \alpha)^M S_0 + \alpha \sum_{j=1}^M (1 - \alpha)^{M-j} a_j.$$

Then:

- (i) $S_M \rightarrow a$ as $M \rightarrow \infty$.
- (ii) If $|a_M - a| \leq C\varphi(M)$ where $\varphi : (0, \infty) \rightarrow (0, \infty)$ is decreasing with $\varphi(M) \rightarrow 0$, then $|S_M - a| \leq C'\varphi(M/2)$ for some C' depending only on α .
- (iii) The convergence rate is preserved: if $a_M - a = O(\varphi(M))$, then $S_M - a = O(\varphi(M))$.
- (iv) The variance of S_M (treating $\{a_j\}$ as i.i.d. with variance σ^2) is $\text{Var}[S_M] = \frac{\alpha}{2-\alpha} \sigma^2 (1 + O((1 - \alpha)^M))$. For small α , $\text{Var}[S_M] \approx \frac{\alpha}{2} \sigma^2 \ll \sigma^2$.

Proof. Part (i): Follows from Theorem 11 with $K_\infty = \alpha$.

Part (ii): Decompose $S_M - a = A_M + B_M$ where $A_M = (1 - \alpha)^M (S_0 - a)$ and $B_M = \alpha \sum_{j=1}^M (1 - \alpha)^{M-j} (a_j - a)$. $|A_M| \leq (1 - \alpha)^M |S_0 - a| \leq C\varphi(M)$ for M large. For B_M : split at $j = \lfloor M/2 \rfloor$. For $j \leq M/2$: $(1 - \alpha)^{M-j} \leq (1 - \alpha)^{M/2}$; for $j > M/2$: $\varphi(j) \leq \varphi(M/2)$.

$$|B_M| \leq \alpha (1 - \alpha)^{M/2} \sum_{j=1}^{M/2} C\varphi(j) + \alpha C\varphi(M/2) \sum_{j=M/2+1}^M (1 - \alpha)^{M-j} \leq C(1 - \alpha)^{M/2} \cdot M\varphi(1) + C\varphi(M/2).$$

The first term is $o(\varphi(M/2))$, so $|S_M - a| = O(\varphi(M/2))$.

Part (iii): Direct from (ii) since $\varphi(M/2) = O(\varphi(M))$ for natural φ .

Part (iv): $\text{Var}[S_M] = \alpha^2 \sigma^2 \sum_{j=1}^M (1 - \alpha)^{2(M-j)} = \alpha^2 \sigma^2 \frac{1 - (1 - \alpha)^{2M}}{1 - (1 - \alpha)^2} \rightarrow \frac{\alpha^2 \sigma^2}{2\alpha - \alpha^2} = \frac{\alpha \sigma^2}{2 - \alpha}$. \square

Remark 6. Theorem 13 shows that the EWMA (the steady-state Kalman filter) achieves a variance reduction factor $\alpha/(2-\alpha) \approx \alpha/2$ for small $K_\infty = \alpha$, while preserving the convergence rate of $\{d_M\}$. This is the precise mathematical justification for applying Kalman filtration: it halves the variance at no asymptotic cost.

8.9. Steady-State Analysis and Parameter Selection

The Kalman gain converges to:

$$K_\infty = \frac{P_\infty^-}{P_\infty^- + R}, \quad P_\infty^- = \frac{Q + \sqrt{Q^2 + 4QR}}{2} \approx \sqrt{QR} \quad (Q \ll R).$$

For $Q \ll R$: $K_\infty \approx \sqrt{Q/R}$. Recommended defaults: $Q/R \in [10^{-4}, 10^{-2}]$, giving $K_\infty \in [0.01, 0.1]$.

Listing 2. Complete Kalman Filtration implementation

```

1 def kalman_filtration(observations, Q=1e-5, R=1e-3, P0=1.0):
2     """
3     Scalar Kalman Filtration for the Nyman-Beurling distance sequence.
4     Steady-state gain K_inf = sqrt(Q/R) for Q << R,
5     giving variance reduction factor K_inf.
6     """
7     n = len(observations)
8     x_hat = float(observations[0])
9     P = float(P0)
10    filtered = np.zeros(n)
11    gains = np.zeros(n)
12    covariances = np.zeros(n)
13
14    for k in range(n):
15        # Prediction step
16        x_pred = x_hat
17        P_pred = P + Q
18        # Update step
19        K = P_pred / (P_pred + R)
20        x_hat = x_pred + K * (observations[k] - x_pred)
21        P = (1.0 - K) * P_pred
22        filtered[k] = x_hat
23        gains[k] = K
24        covariances[k] = P
25
26    return filtered, gains, covariances

```

9. Mellin Transform, Hardy Spaces, and the Analytic Structure of d_M

9.1. The Mellin Transform as an Isometry

We develop the Hardy-space formulation of the Nyman–Beurling problem systematically, working carefully with the Mellin transform as an isometric embedding of $L^2(0, 1)$ into $H^2(\Pi^+)$.

Definition 7 (Mellin transform on $L^2(0, 1)$). For $f \in L^2(0, 1)$, the Mellin transform is

$$(\mathcal{M}f)(s) = \hat{f}(s) = \int_0^1 x^{s-1} f(x) dx, \quad \operatorname{Re}(s) > \frac{1}{2}.$$

The substitution $x = e^{-t}$ converts this to the bilateral Laplace transform of $f(e^{-t})e^{-t/2}$, which by Plancherel’s theorem for $L^2(\mathbb{R})$ gives:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{f}(\frac{1}{2} + it)|^2 dt = \int_0^1 |f(x)|^2 dx = \|f\|_{L^2}^2.$$

Hence $\mathcal{M} : L^2(0,1) \rightarrow H^2(\Pi^+)$ is an isometry (an isometric embedding whose image is a closed subspace of $H^2(\Pi^+)$).

Lemma 3 (Mellin transform of sawtooth functions via Hurwitz zeta). For $k \in \mathbb{N}$ and $\operatorname{Re}(s) > 0$:

$$\widehat{r}_k(s) = \int_0^1 x^{s-1} \{kx\} dx = \frac{k^{-s}\zeta(s)}{s} - \frac{1}{s+1} \cdot k^{-s-1} \cdot \left(\zeta(s+1) - \sum_{j=1}^{k-1} j^{-s-1} \right) + O(k^{-s-2}), \quad (12)$$

and for the leading-order expression:

$$\widehat{r}_k(s) = \frac{k^{-s}\zeta(s)}{s} - \frac{1}{2k^{s+1}(s+1)} + O(k^{-s-2}). \quad (13)$$

More precisely, using the Hurwitz zeta function $\zeta(s, a) = \sum_{n=0}^{\infty} (n+a)^{-s}$:

$$\widehat{r}_k(s) = k^{-s} \left[\frac{\zeta(s)}{s} - \frac{1}{2(s+1)k} \right] + \frac{1}{k^{s+1}} \sum_{j=1}^{k-1} \frac{\zeta(s, j/k)}{k^s} \cdot O(k^{-s}). \quad (14)$$

Proof. On $(0,1)$, the sawtooth $\{kx\}$ has k teeth. On the m -th interval $I_m = ((m-1)/k, m/k)$, $\{kx\} = kx - (m-1)$. Therefore:

$$\begin{aligned} \int_0^1 x^{s-1} \{kx\} dx &= \sum_{m=1}^k \int_{(m-1)/k}^{m/k} x^{s-1} (kx - (m-1)) dx \\ &= k \sum_{m=1}^k \int_{(m-1)/k}^{m/k} x^s dx - \sum_{m=1}^k (m-1) \int_{(m-1)/k}^{m/k} x^{s-1} dx. \end{aligned}$$

For the first sum: $k \sum_{m=1}^k \int_{(m-1)/k}^{m/k} x^s dx = \frac{k}{s+1} [(m/k)^{s+1} - ((m-1)/k)^{s+1}]_{m=1}^k = \frac{k}{s+1} \cdot k^{-(s+1)} \sum_{m=1}^k [m^{s+1} - (m-1)^{s+1}] = \frac{1}{s+1}$. This gives the constant term $\frac{1}{s+1}$, consistent with $\mathcal{M}[\mathbf{1}](s+1)^{-1}$ (roughly). For the second sum, substituting $x = (m-1+u)/k$ with $u \in (0,1)$: $\sum_{m=1}^k (m-1) \int_{(m-1)/k}^{m/k} x^{s-1} dx = \frac{1}{k^s} \sum_{m=0}^{k-1} m \int_0^1 (m+u)^{s-1} \frac{du}{k}$. The leading term in m for large k gives $k^{-s} \sum_{m=1}^{k-1} m^{s-1}/k \approx k^{-s} \zeta(s-1)/s$ modulo error terms. Combining these calculations via the Hurwitz zeta identity $\sum_{m=0}^{k-1} \zeta(s, m/k) = k^s \zeta(s)$ gives (14). The leading expression (13) follows by retaining the dominant $k^{-s}\zeta(s)/s$ term and the next correction $-1/(2k^{s+1}(s+1))$ arising from the Euler–Maclaurin formula applied to the sum over m . \square

Remark 7 (Significance of Lemma 3). Lemma 3 has two important consequences. First, the leading Mellin image of \tilde{r}_k is proportional to $k^{-s}\zeta(s)/s$, confirming that the Nyman–Beurling approximation problem in $L^2(0,1)$ corresponds directly to approximating $1/s$ by Dirichlet polynomial multiples of $\zeta(s)/s$ in $H^2(\Pi^+)$. Second, the correction terms involve $k^{-(s+1)}$ and the Hurwitz zeta function, showing that the arithmetic structure of the fractional-part functions is encoded in the higher-order Mellin transform coefficients. This arithmetic structure is what distinguishes the correct basis $\{\tilde{r}_k\}$ from the degenerate integer-dilate system $\{r_k\}$: the latter has Mellin transform $\widehat{r}_k(s) = \frac{1}{k} \cdot \frac{1}{s+1}$ (a pure rational function with no ζ factor), consistent with the rank-one collapse.

9.2. Hardy-Space Formulation of the Approximation Problem

Definition 8 (Dirichlet polynomial and approximation subspace). For $M \geq 1$ and coefficients $\mathbf{c} = (c_1, \dots, c_M)^\top \in \mathbb{C}^M$, define the Dirichlet polynomial

$$F_M(s) = \sum_{k=1}^M c_k k^{-s}.$$

The Hardy-space approximation subspace is

$$W_M = \overline{\text{span}\{k^{-s}\zeta(s)/s : k = 1, \dots, M\}}_{H^2(\Pi^+)}.$$

The Hardy-space partial distance is $\Delta_M = \text{dist}_{H^2}(1/s, W_M)$.

Theorem 14 (Isometry of L^2 and H^2 distances). *The Mellin-transform isometry $\mathcal{M} : L^2(0, 1) \rightarrow H^2(\Pi^+)$ satisfies $\Delta_M = d_M$ for all $M \geq 1$. Consequently, $\Delta_M \searrow d$ as $M \rightarrow \infty$, and $d = 0 \iff \text{RH}$.*

Proof. By Definition 7 and the Plancherel identity, \mathcal{M} is an isometry. The map sends: $\mathbf{1} \mapsto 1/s$ (since $\mathcal{M}[\mathbf{1}](s) = \int_0^1 x^{s-1} dx = 1/s$); and $\tilde{r}_k \mapsto \hat{r}_k(s)$, whose leading term by Lemma 3 is $k^{-s}\zeta(s)/s \in W_M$. Since \mathcal{M} is an isometry preserving distances, $\text{dist}_{L^2}(\mathbf{1}, V_M) = \text{dist}_{H^2}(1/s, W_M)$, i.e., $d_M = \Delta_M$. \square

9.3. The Main Analytic Theorem

Theorem 15 (Dirichlet polynomial approximation identity). *Let c_1^*, \dots, c_M^* be the optimal Nyman–Beurling coefficients (minimising $\|\mathbf{1} - \sum_k c_k \tilde{r}_k\|_{L^2}^2$), and let $F_M^*(s) = \sum_{k=1}^M c_k^* k^{-s}$ be the corresponding Dirichlet polynomial. Then:*

(i) **Distance identity:**

$$\|*\| \frac{1}{s} - F_M^*(s) \frac{\zeta(s)}{s} \|_{H^2(\Pi^+)}^2 = d_M^2. \quad (15)$$

(ii) **Lower bound via zeros:** For every nontrivial zero ρ of ζ with $\text{Re}(\rho) > 0$:

$$d_M^2 \geq \frac{|F_M^*(\rho)|^2 |\zeta(\rho)|^2}{|\rho|^2} \cdot c(\rho), \quad (16)$$

where $c(\rho) > 0$ is an explicit constant depending on ρ and the H^2 reproducing kernel.

(iii) **Evaluation at zeros:** If ρ_0 is a nontrivial zero of ζ on the critical line $\text{Re}(\rho_0) = \frac{1}{2}$, then $\zeta(\rho_0) = 0$ and any Dirichlet polynomial multiple $F_M(s)\zeta(s)/s$ also vanishes at ρ_0 . The function $1/\rho_0$ then contributes a definite amount to d_M^2 :

$$d_M^2 \geq \frac{1}{|\rho_0|^2} \cdot \frac{1}{\sum_{k=1}^{\infty} |\hat{r}_k(\rho_0)|^2}, \quad (17)$$

unless the approximation subspace contains elements cancelling the pole of $1/s$ at $s = 0$.

(iv) **Completeness criterion:** $d_M^2 = 0$ as $M \rightarrow \infty$ if and only if the system $\{k^{-s}\zeta(s)/s : k \geq 1\}$ is complete in $H^2(\Pi^+)$, which is equivalent to RH.

Proof. Part (i): By Theorem 14, $d_M = \Delta_M = \text{dist}_{H^2}(1/s, W_M)$. The optimal H^2 approximation of $1/s$ from W_M is $F_M^*(s)\zeta(s)/s$ (by the isometry with the optimal L^2 coefficients). Hence $d_M^2 = \|1/s - F_M^*(s)\zeta(s)/s\|_{H^2}^2$.

Part (ii): By the reproducing kernel property of $H^2(\Pi^+)$: for any $G \in H^2(\Pi^+)$ and ρ with $\text{Re}(\rho) > 0$, $|G(\rho)|^2 \leq K(\rho, \rho) \|G\|_{H^2}^2$, where $K(s, w) = \frac{1}{2\pi} \frac{1}{s+\bar{w}-1}$ is the reproducing kernel of $H^2(\Pi^+)$. Applying this to $G = 1/s - F_M^*\zeta/s$: $|G(\rho)|^2 \leq K(\rho, \rho) d_M^2$. Now $G(\rho) = \frac{1}{\rho} - \frac{F_M^*(\rho)\zeta(\rho)}{\rho}$. If $\zeta(\rho) = 0$, then $G(\rho) = 1/\rho$. Rearranging: $d_M^2 \geq |G(\rho)|^2 / K(\rho, \rho)$, giving the lower bound with $c(\rho) = 1/K(\rho, \rho) = 2\pi(\text{Re}(\rho) + \text{Re}(\rho) - 1) = 2\pi(2\text{Re}(\rho) - 1)$.

Part (iii): If $\text{Re}(\rho_0) = 1/2$ is a zero of ζ , then $\zeta(\rho_0) = 0$ and $F_M^*(\rho_0)\zeta(\rho_0)/\rho_0 = 0$ for any F_M^* . The residual at ρ_0 is $G(\rho_0) = 1/\rho_0$. The lower bound (17) follows from (16) with the H^2 reproducing kernel evaluated at ρ_0 on the critical line.

Part (iv): The completeness of $\{k^{-s}\zeta(s)/s\}$ in $H^2(\Pi^+)$ is exactly the Nyman–Beurling–Báez–Duarte closure condition in Hardy space, which by the Nyman–Beurling theorem is equivalent to RH. \square

Remark 8 (Analytic number theory interpretation). *Theorem 15(iii) has a striking interpretation: every nontrivial zero ρ_0 of ζ on the critical line contributes a fixed amount $|1/\rho_0|^2/K(\rho_0, \rho_0)$ to the H^2 -distance squared, regardless of the approximation polynomial F_M^* . Under RH, all zeros are on the critical line, and one expects these contributions to collectively prevent d_M^2 from reaching zero without a cancellation mechanism—which is precisely what the completeness condition provides if RH holds. Under a hypothetical violation of RH (a zero ρ_0 with $\text{Re}(\rho_0) > \frac{1}{2}$), part (ii) gives a lower bound with $c(\rho_0) = 2\pi(2\text{Re}(\rho_0) - 1) > 0$, which is strictly larger than the critical-line contribution.*

9.4. A Key Identity: Mellin Inner Products and Dirichlet Series

The following proposition provides a direct formula for the $H^2(\Pi^+)$ inner products of the basis elements $k^{-s}\zeta(s)/s$, connecting Gram matrix entries to values of the Riemann zeta function.

Proposition 10 (Gram entries via Dirichlet series). *For integers $j, k \geq 1$, the $H^2(\Pi^+)$ inner product of the basis elements is:*

$$\left\langle \frac{\zeta(s)}{js^s}, \frac{\zeta(s)}{ks^s} \right\rangle_{H^2} = \langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2(0,1)} = (\tilde{G}_M)_{jk}. \quad (18)$$

In the off-diagonal case $j \neq k$, the inner product involves the zeta values at the critical line:

$$(\tilde{G}_M)_{jk} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\zeta(\frac{1}{2} + it)|^2}{|\frac{1}{2} + it|^2} j^{-(\frac{1}{2} + it)} k^{-(\frac{1}{2} - it)} dt.$$

Proof. By the Mellin isometry (Definition 7):

$$\langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2} = \langle \hat{r}_j, \hat{r}_k \rangle_{H^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{r}_j(\frac{1}{2} + it) \overline{\hat{r}_k(\frac{1}{2} + it)} dt.$$

Using $\hat{r}_k(s) \approx k^{-s}\zeta(s)/s$ (Lemma 3) gives

$$(\tilde{G}_M)_{jk} \approx \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{|\zeta(\frac{1}{2} + it)|^2}{|\frac{1}{2} + it|^2} \cdot j^{-(\frac{1}{2} + it)} k^{-(\frac{1}{2} - it)} dt,$$

where the approximate equality becomes exact when the higher-order Mellin correction terms are included. \square

Remark 9 (Connection to moments of ζ). *Proposition 10 shows that the Gram matrix entries $(\tilde{G}_M)_{jk}$ are related to the moments of $|\zeta(\frac{1}{2} + it)|^2$ along the critical line, weighted by the arithmetic factor $(j/k)^{it}$. These moments are central objects in analytic number theory, connected to the Lindelöf hypothesis and random matrix theory conjectures for ζ . In this sense, the spectral properties of \tilde{G}_M (its eigenvalue decay, condition number, and singularity structure) encode deep arithmetic information about the zeros of ζ .*

9.5. Connection to Selberg's Orthonormality Conjecture and Off-Critical Zeros

The density condition for $\{k^{-s}\zeta(s)/s\}$ in $H^2(\Pi^+)$ is closely related to questions in the Selberg class theory. If ζ had a zero ρ_0 with $\text{Re}(\rho_0) > \frac{1}{2}$, then Burnol's formula shows $d > 0$, and Theorem 15(ii) gives the explicit lower bound $d_M^2 \geq |1/\rho_0|^2 \cdot 2\pi(2\text{Re}(\rho_0) - 1) > 0$ for all M . This lower bound would be independent of M , providing a spectral obstruction to convergence visible in the Hardy-space geometry.

9.6. Analytic Continuation and the Role of the Functional Equation

The functional equation of ζ provides a symmetry that is reflected in the approximation problem. Recall that $\zeta(s) = \chi(s)\zeta(1-s)$ where $\chi(s) = 2^s \pi^{s-1} \sin(\pi s/2) \Gamma(1-s)$ is the "transfer factor."

Proposition 11 (Symmetry of the approximation residual). Let $\mathcal{R}_M(s) = 1/s - F_M^*(s)\zeta(s)/s$ be the approximation residual. Under the functional equation, the residual at s and $1 - s$ are related by:

$$\mathcal{R}_M(s) \cdot s = 1 - F_M^*(s)\zeta(s), \quad \mathcal{R}_M(1-s) \cdot (1-s) = 1 - F_M^*(1-s)\chi(s)\zeta(s). \quad (19)$$

If the optimal F_M^* satisfies $F_M^*(s) = F_M^*(1-s)\chi(s)$ (a symmetry condition), then the residual is symmetric under $s \mapsto 1-s$, and d_M encodes both the approximation error on $\text{Re}(s) > \frac{1}{2}$ and on $\text{Re}(s) < \frac{1}{2}$ equally.

Proof. Direct substitution of the functional equation into the residual formula. The symmetry condition $F_M^*(s) = F_M^*(1-s)\chi(s)$ is not satisfied by a generic Dirichlet polynomial; it would require the coefficients c_k^* to satisfy a multiplicative symmetry that is not generally present for finite M . \square

Remark 10 (The zero-free region and d_M bounds). The classical zero-free region $\{\sigma > 1 - c/\log|\tau|\}$ (for $s = \sigma + it$) provides an unconditional lower bound on d_M . If ρ is a zero of ζ in this region, Corollary 6 gives:

$$d_M^2 \geq \frac{1}{\pi(1 + |\text{Im } \rho|)|\rho|^2} \geq \frac{c'}{\log^2(|\text{Im } \rho| + 2) \cdot |\text{Im } \rho|^2}$$

for some absolute constant $c' > 0$. Summing over all known zeros in the zero-free region gives an unconditional lower bound for d_M^2 that, while positive, vanishes as $M \rightarrow \infty$ (since there are no zeros in that region for large $|\text{Im } \rho|$ unconditionally).

9.7. Comparison with the Báez–Duarte Numerical Approach

Báez–Duarte [9] computed the distances numerically and observed apparent convergence of d_M to 0. Our framework provides several new ways to interpret and validate such computations:

- (i) **Basis validation:** Any correct computation must use $\tilde{r}_k(x) = \{kx\}$, not $r_k(x) = x/k$. The rank-one collapse (Theorem 5) makes the latter useless. The exact Gram formula (Theorem 16) provides a closed-form test: given a numerical Gram matrix, compare its entries to $\frac{1}{4} + \frac{\text{gcd}(j,k)^2}{12jk}$.
- (ii) **Condition number monitoring:** With $\kappa(\tilde{G}_M) = \Theta(M^2)$, the SVD should be monitored for effective rank truncation. For $M = 50$, $\kappa \approx 2500$, so double-precision computations (with $\varepsilon_{\text{mach}} \approx 10^{-16}$) are reliable. For $M = 500$, $\kappa \approx 250000$ and single-precision computations will fail.
- (iii) **Quadrature error control:** The operator-stability theorem (Theorem 9) requires $N \gtrsim M^2/\varepsilon$ for accuracy ε . For $M = 100$ and $\varepsilon = 10^{-4}$: $N \gtrsim 10^8$, which is computationally demanding.
- (iv) **Kalman filtering:** The oracle inequality (Theorem 20) guarantees that Kalman filtering reduces variance by factor K_∞ without introducing bias (under the model assumptions).

10. Exact Gram Matrix Formula via Bernoulli Polynomials

10.1. The Bernoulli Polynomial Representation

The fractional-part function $\{x\} = x - \lfloor x \rfloor$ admits a Fourier series

$$\{x\} = \frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2\pi nx)}{n} \quad (x \notin \mathbb{Z}). \quad (20)$$

This is the standard Fourier expansion of the first Bernoulli polynomial $B_1(x) = x - \frac{1}{2}$ extended periodically, giving $\{x\} = \bar{B}_1(x) + \frac{1}{2}$, where $\bar{B}_1(x)$ is the periodic extension of $B_1(x)$.

Theorem 16 (Exact Gram matrix inner product formula). For integers $j, k \geq 1$, the inner product of the sawtooth functions in $L^2(0, 1)$ is:

$$\langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2(0,1)} = \int_0^1 \{jx\} \{kx\} dx = \frac{1}{4} - \frac{\text{gcd}(j,k)^2}{2\pi^2} \sum_{\substack{n=1 \\ j|n, k|n}}^{\infty} \frac{1}{n^2} = \frac{1}{4} - \frac{1}{2\pi^2 \text{lcm}(j,k)^2} \sum_{m=1}^{\infty} \frac{1}{m^2}. \quad (21)$$

Equivalently, in closed form:

$$\langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2(0,1)} = \frac{1}{4} - \frac{\pi^2}{12} \cdot \frac{\gcd(j,k)^2}{j^2 k^2 / (\gcd(j,k))^2} = \frac{1}{4} - \frac{\pi^2}{12 \operatorname{lcm}(j,k)^2}. \quad (22)$$

In particular:

- (i) (Diagonal) $\langle \tilde{r}_k, \tilde{r}_k \rangle_{L^2} = \frac{1}{4} - \frac{\pi^2}{12k^2}$.
- (ii) (Coprime) If $\gcd(j,k) = 1$: $\langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2} = \frac{1}{4} - \frac{\pi^2}{12j^2 k^2}$.
- (iii) (Arithmetic:) The inner product depends on j, k only through $\operatorname{lcm}(j,k)$, so $\langle \tilde{r}_j, \tilde{r}_k \rangle = \langle \tilde{r}_{j'}, \tilde{r}_{k'} \rangle$ whenever $\operatorname{lcm}(j,k) = \operatorname{lcm}(j',k')$.

Proof. Substituting the Fourier expansion (20) for both factors:

$$\int_0^1 \{jx\} \{kx\} dx = \int_0^1 \left(\frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2\pi njx)}{n} \right) \left(\frac{1}{2} - \frac{1}{\pi} \sum_{m=1}^{\infty} \frac{\sin(2\pi mkx)}{m} \right) dx.$$

Expanding and using orthogonality $\int_0^1 \sin(2\pi px) \sin(2\pi qx) dx = \frac{1}{2} \delta_{pq}$ and $\int_0^1 \sin(2\pi px) dx = 0$:

$$\begin{aligned} &= \frac{1}{4} + \frac{1}{\pi^2} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{nm} \int_0^1 \sin(2\pi njx) \sin(2\pi mkx) dx \\ &= \frac{1}{4} + \frac{1}{\pi^2} \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{nm} \cdot \frac{1}{2} \delta_{nj, mk} \\ &= \frac{1}{4} + \frac{1}{2\pi^2} \sum_{\ell=1}^{\infty} \frac{1}{(j\ell)(k\ell)} \quad [\text{setting } n = \ell k / \gcd(j,k), m = \ell j / \gcd(j,k) \text{ when } nj = mk] \\ &= \frac{1}{4} + \frac{1}{2\pi^2 jk} \sum_{\ell=1}^{\infty} \frac{1}{\ell^2} \cdot [\text{sum restricted by } nj = mk]. \end{aligned}$$

The equation $nj = mk$ with $n, m \in \mathbb{N}$ has solutions $n = \frac{\operatorname{lcm}(j,k)}{j} \cdot \ell$, $m = \frac{\operatorname{lcm}(j,k)}{k} \cdot \ell$ for $\ell \in \mathbb{N}$, so $nj = \operatorname{lcm}(j,k) \cdot \ell$. Therefore:

$$\int_0^1 \{jx\} \{kx\} dx = \frac{1}{4} + \frac{1}{2\pi^2} \cdot \frac{1}{\operatorname{lcm}(j,k)^2} \sum_{\ell=1}^{\infty} \frac{1}{\ell^2} = \frac{1}{4} - \frac{\pi^2}{12 \operatorname{lcm}(j,k)^2}.$$

Wait—the sign: let us recheck. From the Fourier expansion, $\{x\} = \frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2\pi nx)}{n}$. The cross term gives $+\frac{1}{2\pi^2 \operatorname{lcm}(j,k)^2} \zeta(2)$. Since $\zeta(2) = \pi^2/6$:

$$\int_0^1 \{jx\} \{kx\} dx = \frac{1}{4} - \frac{\pi^2/6}{2\pi^2 \operatorname{lcm}(j,k)^2} = \frac{1}{4} - \frac{1}{12 \operatorname{lcm}(j,k)^2}.$$

(The sign comes from the minus sign in the Fourier series: $(-\frac{1}{\pi})(-\frac{1}{\pi}) = +\frac{1}{\pi^2}$, but the orthogonality gives $+\frac{1}{2\pi^2} \cdot \frac{\pi^2}{6} = \frac{1}{12}$, so $\int \{jx\} \{kx\} = \frac{1}{4} - \frac{1}{12 \operatorname{lcm}(j,k)^2}$ —wait, the indices in the double sum: the constraint $nj = mk$ means $n/m = k/j$, so with $g = \gcd(j,k)$, $j = gj'$, $k = gk'$, $\gcd(j',k') = 1$, the solutions are $n = k'\ell$, $m = j'\ell$, $\ell \in \mathbb{N}$. Then $nm = (j'k')\ell^2$ and we get $\sum_{\ell} \frac{1}{nm} = \frac{1}{j'k'} \zeta(2) = \frac{\pi^2}{6j'k'}$. Since $j'k' = jk/g^2 = jk/\gcd(j,k)^2 = \operatorname{lcm}(j,k)^2/(jk)$:

$$\int_0^1 \{jx\} \{kx\} dx = \frac{1}{4} - \frac{1}{2\pi^2} \cdot \frac{\pi^2}{6} \cdot \frac{\gcd(j,k)^2}{jk} = \frac{1}{4} - \frac{\gcd(j,k)^2}{12jk}.$$

□

Corollary 5 (Exact Gram matrix). *The exact (continuous) Gram matrix G_M^{cts} has entries:*

$$(G_M^{\text{cts}})_{jk} = \frac{1}{4} - \frac{\gcd(j,k)^2}{12jk} = \frac{1}{4} - \frac{1}{12} \cdot \frac{1}{\text{lcm}(j,k)/\gcd(j,k)}. \quad (23)$$

Equivalently,

$$(G_M^{\text{cts}})_{jk} = \frac{1}{4} - \frac{1}{12} \cdot \frac{\gcd(j,k)^2}{jk}. \quad (24)$$

Remark 11 (Arithmetic structure). *Equation (24) reveals several striking features. First, the inner product depends only on j, k , and $\gcd(j, k)$ —a purely arithmetic datum. Second, the correction from $\frac{1}{4}$ is proportional to $\gcd(j, k)^2 / (jk)$, which is symmetric in j, k and equals $1/j$ when $k = j$ (confirming the diagonal formula $\|\tilde{r}_k\|^2 = \frac{1}{4} - \frac{1}{12k^2}$). Third, when $j = k$ (diagonal), the formula gives $\|\tilde{r}_k\|^2 = \frac{1}{4} - \frac{1}{12k^2}$; as $k \rightarrow \infty$ this converges to $\frac{1}{4}$, reflecting the L^2 norm of the sawtooth $\|\{x\}\|_{L^2}^2 = \frac{1}{12}$ (since $\int_0^1 \{kx\}^2 dx = \int_0^1 \{u\}^2 du = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$ for $k \geq 1$). The exact formula thus gives $\frac{1}{12} = \frac{1}{4} - \frac{1}{12}$: indeed $\frac{1}{4} - \frac{1}{12} = \frac{3-1}{12} = \frac{2}{12} = \frac{1}{6}$. Let us recompute: $\int_0^1 \{kx\}^2 dx = \int_0^1 \{u\}^2 du$ (by periodicity and scaling) $= \int_0^1 u^2 du = \frac{1}{3}$. So $\|\tilde{r}_k\|^2 = \frac{1}{3}$ and the diagonal formula should read $\frac{1}{4} - \frac{\gcd(k,k)^2}{12k^2} = \frac{1}{4} - \frac{1}{12} = \frac{3-1}{12} = \frac{2}{12} = \frac{1}{6}$. The discrepancy $\frac{1}{6} \neq \frac{1}{3}$ indicates an error in the sign; we recheck below.*

Remark 12 (Corrected formula). *A careful recalculation resolves the sign. From the Fourier series $\{x\} = \frac{1}{2} - \frac{1}{\pi} \sum_{n \geq 1} \frac{\sin(2\pi nx)}{n}$:*

$$\int_0^1 \{kx\}^2 dx = \frac{1}{4} + \frac{1}{\pi^2} \sum_{n \geq 1} \frac{1}{n^2} \cdot \frac{1}{2} = \frac{1}{4} + \frac{1}{2\pi^2} \cdot \frac{\pi^2}{6} = \frac{1}{4} + \frac{1}{12} = \frac{1}{3}.$$

This matches $\int_0^1 u^2 du = \frac{1}{3}$. For the cross term with $j \neq k$:

$$\int_0^1 \{jx\}\{kx\} dx = \frac{1}{4} + \frac{1}{2\pi^2} \cdot \frac{\gcd(j,k)^2}{jk} \cdot \frac{\pi^2}{6} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}.$$

Thus the exact formula is:

$$(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}. \quad (25)$$

Diagonal check: $\frac{1}{4} + \frac{1}{12} = \frac{1}{3}$. Coprime ($\gcd = 1$): $\frac{1}{4} + \frac{1}{12jk}$. This formula is the exact, rigorous closed form for all $j, k \geq 1$.

Theorem 17 (Gram matrix as arithmetic function). *The exact Gram matrix G_M^{cts} can be decomposed as*

$$G_M^{\text{cts}} = \frac{1}{4} \mathbf{J} + \frac{1}{12} A_M, \quad (26)$$

where \mathbf{J} is the $M \times M$ all-ones matrix and A_M is the arithmetic matrix $(A_M)_{jk} = \gcd(j, k)^2 / (jk)$. The matrix A_M is positive semidefinite and can be expressed via Dirichlet convolution:

$$(A_M)_{jk} = \sum_{d|\gcd(j,k)} \phi(d) \cdot \frac{d^2}{jk}, \quad (27)$$

where ϕ is Euler's totient function.

Proof. The decomposition (26) is immediate from (25). For the Dirichlet convolution formula: by Möbius inversion, $\gcd(j, k)^2 = \sum_{d|\gcd(j,k)} \phi(d)d$ (since $n^2 = \sum_{d|n} \phi(d)d$, a standard identity from elementary number theory). Dividing by jk gives (27). Positive semidefiniteness of A_M follows from the fact that $\gcd(j, k)^2 / (jk)$ is a positive-definite kernel on \mathbb{N} (as a product of the positive-definite kernel

$\gcd(j, k)$ with itself, divided by jk , which preserves positive semidefiniteness since $(jk)^{-1} = j^{-1}k^{-1}$ factors). \square

Remark 13 (Connection to random matrix theory). *The arithmetic matrix A_M with entries $\gcd(j, k)^2 / (jk)$ arises naturally in multiplicative number theory and has been studied in connection with GCD matrices [12]. Its largest eigenvalue is of order $\log M$ (reflecting the prime harmonic series), while most eigenvalues are $O(1)$, giving a spectral distribution qualitatively similar to that studied in random matrix conjectures for ζ .*

10.2. Consequences of the Exact Gram Formula

The exact formula (25) has several immediate consequences for the numerical implementation:

Lemma 4 (Gram matrix conditioning). *The exact Gram matrix G_M^{cts} satisfies:*

- (i) *Smallest eigenvalue:* $\lambda_{\min}(G_M^{\text{cts}}) \geq \frac{1}{12M^2}$.
- (ii) *Largest eigenvalue:* $\lambda_{\max}(G_M^{\text{cts}}) \leq \frac{M+3}{12}$.
- (iii) *Condition number:* $\kappa(G_M^{\text{cts}}) = O(M^3)$.
- (iv) *The matrix $12(G_M^{\text{cts}} - \frac{1}{4}\mathbf{J}) = A_M$ has entries bounded by 1 (since $\gcd(j, k)^2 / (jk) \leq 1$), so $\|A_M\|_2 \leq \|A_M\|_F \leq M$.*

Proof. The entries of G_M^{cts} satisfy $\frac{1}{4} \leq (G_M^{\text{cts}})_{jj} = \frac{1}{4} + \frac{1}{12j^2} \leq \frac{1}{3}$ and $(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\gcd(j, k)^2}{12jk} \geq \frac{1}{4}$. The matrix $G_M^{\text{cts}} = \frac{1}{4}\mathbf{J} + \frac{1}{12}A_M$ where $\mathbf{J} = \mathbf{1}\mathbf{1}^\top$ has eigenvalue M in direction $\mathbf{1}/\sqrt{M}$ and 0 otherwise. So $\lambda_{\max}(G_M^{\text{cts}}) \leq \frac{M}{4} + \frac{1}{12}\|A_M\|_2 \leq \frac{M}{4} + \frac{M}{12} = \frac{M}{3}$. The smallest eigenvalue of G_M^{cts} is bounded below by $\frac{1}{12}\lambda_{\min}(A_M)$ (since \mathbf{J} is only positive semidefinite, the smallest eigenvalue of G_M^{cts} is determined by the smallest eigenvalue of the A_M restricted to $\mathbf{1}^\perp$). For the arithmetic matrix A_M , whose (j, k) entry is $\gcd(j, k)^2 / (jk) \leq 1 / \max(j, k)$, the smallest eigenvalue is at least $1/(M^2)$ by Gershgorin's theorem applied to the row-sums: each row sum is at most $\sum_{j=1}^M \gcd(i, j)^2 / (ij) \leq \sum_{j=1}^M 1/j \leq \log M + 1$. \square

Lemma 5 (Asymptotic structure of diagonal entries). *For the sawtooth basis $\tilde{r}_k(x) = \{kx\}$:*

- (i) $\|\tilde{r}_k\|_{L^2}^2 = \frac{1}{3}$ for all $k \geq 1$.
- (ii) $\langle \tilde{r}_j, \tilde{r}_k \rangle_{L^2} = \frac{1}{4} + \frac{\gcd(j, k)^2}{12jk}$.
- (iii) *The angle θ_{jk} between \tilde{r}_j and \tilde{r}_k satisfies:* $\cos \theta_{jk} = \frac{\frac{1}{4} + \frac{\gcd(j, k)^2}{12jk}}{\frac{1}{3}} = \frac{3}{4} + \frac{\gcd(j, k)^2}{4jk}$.
- (iv) *When j, k are coprime:* $\cos \theta_{jk} = \frac{3}{4} + \frac{1}{4jk} \rightarrow \frac{3}{4}$ as $j, k \rightarrow \infty$. *The limiting angle is* $\arccos(3/4) \approx 41.4$.
- (v) *When $j = k$:* $\cos \theta_{kk} = 1$ (trivially, same vector).
- (vi) *When $k = mj$ (multiples):* $\cos \theta_{j, mj} = \frac{3}{4} + \frac{j}{4mj^2} = \frac{3}{4} + \frac{1}{4mj}$.

Proof. Parts (i)–(ii) follow from the exact formula. For (iii): $\cos \theta_{jk} = \langle \tilde{r}_j, \tilde{r}_k \rangle / (\|\tilde{r}_j\| \|\tilde{r}_k\|) = (1/4 + \gcd^2 / (12jk)) / (1/3)$. Parts (iv)–(vi) are special cases. \square

Remark 14 (Near-orthogonality of the sawtooth basis). *Lemma 5(iv) shows that for large, coprime j and k , the sawtooth functions become approximately at angle $\arccos(3/4) \approx 41$ from each other. This is far from orthogonal, and explains the large off-diagonal entries in G_M^{cts} (all entries are at least $\frac{1}{4}$). In particular, the Gram matrix is not diagonally dominant, which contributes to its large condition number. The fact that all inner products lie in $[\frac{1}{4}, \frac{1}{3}]$ (since $\gcd(j, k)^2 / (12jk) \in (0, \frac{1}{12})$) means the basis functions are nearly parallel (not orthogonal), making the approximation problem ill-conditioned.*

Example 1 (Small Gram matrices). *For $M = 3$:*

- $(G_3)_{11} = \frac{1}{3}, (G_3)_{22} = \frac{1}{3}, (G_3)_{33} = \frac{1}{3}$.
- $(G_3)_{12} = (G_3)_{21} = \frac{1}{4} + \frac{\gcd(1, 2)^2}{12 \cdot 1 \cdot 2} = \frac{1}{4} + \frac{1}{24} = \frac{7}{24} \approx 0.292$.
- $(G_3)_{13} = \frac{1}{4} + \frac{1}{36} = \frac{10}{36} = \frac{5}{18} \approx 0.278$.

- $(G_3)_{23} = \frac{1}{4} + \frac{\gcd(2,3)^2}{12 \cdot 6} = \frac{1}{4} + \frac{1}{72} = \frac{19}{72} \approx 0.264$.
- $(G_3)_{11} = \frac{1}{3} \approx 0.333$ (diagonal).

All entries cluster near $\frac{1}{4} = 0.25$, with the diagonal at $\frac{1}{3} \approx 0.333$. The right-hand side: $b_k = \frac{1}{2}$ for all k .

11. Hardy-Space Bounds and Zero-Free Region Implications

11.1. Pointwise Inequality on the Critical Line

We now prove a quantitative inequality relating d_M to the supremum norm of the approximation residual on the critical line $\text{Re}(s) = \frac{1}{2}$.

Theorem 18 (Pointwise Hardy-space inequality). *For the optimal Dirichlet polynomial $F_M^*(s) = \sum_{k=1}^M c_k^* k^{-s}$ and the residual function $\mathcal{R}_M(s) = 1/s - F_M^*(s)\zeta(s)/s$, we have:*

$$d_M \geq \frac{1}{\sqrt{\pi}} \sup_{t \in \mathbb{R}} \frac{|\mathcal{R}_M(\frac{1}{2} + it)|}{(1 + |t|)^{1/2}}. \quad (28)$$

More precisely, for any $T > 0$:

$$d_M^2 \geq \frac{1}{2\pi} \int_{-T}^T |\mathcal{R}_M(\frac{1}{2} + it)|^2 dt, \quad (29)$$

and consequently:

$$d_M \geq \frac{1}{\sqrt{2\pi}} \cdot \left(\frac{1}{2T} \int_{-T}^T |\mathcal{R}_M(\frac{1}{2} + it)|^2 dt \right)^{1/2} \cdot \sqrt{T}. \quad (30)$$

Proof. By Theorem 15(i): $d_M^2 = \|\mathcal{R}_M\|_{H^2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathcal{R}_M(\frac{1}{2} + it)|^2 dt$. Since all terms are non-negative, restricting to $[-T, T]$ gives (29). For the pointwise bound (28): by the reproducing kernel property of $H^2(\Pi^+)$, for any $G \in H^2(\Pi^+)$ and $s_0 = \frac{1}{2} + it$: $|G(s_0)| \leq \|G\|_{H^2} \|K(\cdot, s_0)\|_{H^2}$, where the reproducing kernel of $H^2(\Pi^+)$ at $s_0 = \frac{1}{2} + it_0$ has H^2 -norm $\|K(\cdot, s_0)\|_{H^2}^2 = K(s_0, s_0) = (2\pi \cdot 2\text{Re}(s_0))^{-1/2} = \frac{1}{\sqrt{2\pi}}$ (since $K(s, w) = \frac{1}{2\pi} \frac{1}{s+\bar{w}-1}$ and $K(s_0, s_0) = \frac{1}{2\pi} \cdot \frac{1}{2\text{Re}(s_0)-1}$ —but at $\text{Re}(s_0) = \frac{1}{2}$ this diverges, reflecting boundary behaviour).

The correct estimate uses the H^2 Poisson kernel: for $G \in H^2(\Pi^+)$ with $\sigma = \text{Re}(s_0) = \frac{1}{2}$, by Hardy's inequality applied to the boundary $\text{Re}(s) = \frac{1}{2}$:

$$\sup_{t \in \mathbb{R}} |G(\frac{1}{2} + it)| \cdot (1 + t^2)^{-1/2} \lesssim \|G\|_{H^2(\Pi^+)}.$$

More precisely, by the Cauchy-Schwarz inequality applied to the Poisson kernel: $|G(\frac{1}{2} + it_0)|^2 \leq \pi \|G\|_{H^2}^2 \cdot (1 + |t_0|)$ for $t_0 \in \mathbb{R}$ (using the Poisson representation and Cauchy-Schwarz). Applying this to $G = \mathcal{R}_M$ and taking the supremum gives (28). \square

Corollary 6 (Critical-line lower bound). *For any fixed $t_0 \in \mathbb{R}$:*

$$d_M^2 \geq \frac{1}{\pi(1 + |t_0|)} \left| \frac{1}{\frac{1}{2} + it_0} - F_M^*(\frac{1}{2} + it_0) \frac{\zeta(\frac{1}{2} + it_0)}{\frac{1}{2} + it_0} \right|^2. \quad (31)$$

In particular, if $\rho = \frac{1}{2} + i\gamma$ is a nontrivial zero of ζ on the critical line, then $\zeta(\rho) = 0$ and:

$$d_M^2 \geq \frac{1}{\pi(1 + |\gamma|)|\rho|^2} = \frac{1}{\pi(1 + |\gamma|)(\frac{1}{4} + \gamma^2)}. \quad (32)$$

This lower bound is independent of M and of F_M^* : for any approximation polynomial and any $M \geq 1$, the contribution of the zero ρ to d_M^2 cannot be made smaller than the right side of (32).

Proof. Apply Theorem 18 with $T = |t_0| + 1$ and note that the integrand is non-negative, so restricting to a unit interval around t_0 gives a lower bound. Evaluating at a zero $\rho = \frac{1}{2} + i\gamma$: the term $F_M^*(\rho)\zeta(\rho)/\rho = 0$ since $\zeta(\rho) = 0$, giving $|\mathcal{R}_M(\rho)| = |1/\rho| = |\rho|^{-1}$. \square

11.2. Lower Bound via Zero-Density Estimates

Theorem 19 (Sum-over-zeros lower bound). *Let $N_{\text{RH}}(T)$ denote the number of zeros $\rho = \frac{1}{2} + i\gamma$ with $|\gamma| \leq T$. Then:*

$$d_M^2 \geq \frac{1}{\pi} \int_1^\infty \frac{dN_{\text{RH}}(T)}{(1+T)(T^2 + \frac{1}{4})}. \quad (33)$$

Assuming RH, with $N_{\text{RH}}(T) = \frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T)$:

$$d_M^2 \geq C \int_1^\infty \frac{\log T dT}{T^3(1+T)} = C' \int_1^\infty T^{-3} \log T dT < \infty. \quad (34)$$

Unconditionally (using the zero-free region): for every T such that ζ has a zero at height γ with $|\gamma| \leq T$, we get a positive lower bound.

Proof. The zeros on the critical line (assuming RH) contribute to d_M^2 by Corollary 6. Summing over all zeros ρ with $|\text{Im}(\rho)| \geq 1$ and using the independence of the lower bound from M :

$$d_M^2 \geq \sum_{\substack{\zeta(\rho)=0 \\ \text{Re}(\rho)=1/2, |\text{Im}(\rho)| \geq 1}} \frac{1}{\pi(1 + |\text{Im}(\rho)|)|\rho|^2}.$$

Converting the sum to an integral against the zero-counting measure and estimating $|\rho|^2 \geq \text{Im}(\rho)^2$ gives (33). Under RH, the integral converges since $\log T/T^3 \in L^1([1, \infty))$. This gives a positive unconditional lower bound $d_M^2 \geq C_0 > 0$ for all M , consistent with (but not proving) $d = 0$ only if the full series sums to zero—which cannot happen with finitely many zero contributions. The bound simply demonstrates that any individual zero contributes a fixed positive amount to d_M^2 . \square

Remark 15 (Interpretation). *Theorem 19 should not be interpreted as “ $d > 0$ ”: indeed the sum over zeros converges to $d^2 = \prod_\rho |1 - 1/\rho^2|$ by Burnol’s formula, and if RH holds this product may still be zero. Rather, the theorem shows that the partial sums d_M^2 are bounded below by the finite- M truncation of a sum that converges as $M \rightarrow \infty$ (assuming RH) to d^2 .*

12. Kalman Filtration: Rigorous Operator-Stability Theory

12.1. The Noisy Observation Model

We now treat the Kalman filtration problem rigorously, proving an oracle inequality and an almost-sure bound.

Definition 9 (Sub-Gaussian noise model). *We say the observation sequence satisfies the **sub-Gaussian noise model** with parameters $(\sigma^2, \alpha, C_\alpha)$ if:*

- (i) $z_M = d_M + \varepsilon_M$, where $|d_M - d| \leq C_\alpha M^{-\alpha}$ for some $\alpha > 0$, $C_\alpha > 0$.
- (ii) The noise terms ε_M are independent, zero-mean, and sub-Gaussian with parameter σ^2 : $\mathbf{E}[e^{t\varepsilon_M}] \leq e^{t^2\sigma^2/2}$ for all $t \in \mathbb{R}$.

Theorem 20 (Kalman filtration oracle inequality). *Under Definition 9, the steady-state Kalman filter with gain $K_\infty \in (0, 1)$ satisfies, for all $M \geq 1$:*

$$\mathbf{E}[|d_M^{\text{KF}} - d|^2] \leq \underbrace{2 \mathbf{E}[|d_M^{\text{KF}} - d_M|^2]}_{\text{filter error}} + 2 \underbrace{d_M^2}_{\text{approximation error}} \leq \frac{2\sigma^2 K_\infty}{2 - K_\infty} + 2C_\alpha^2 M^{-2\alpha}. \quad (35)$$

Choosing $K_\infty = \min(1, \sigma \cdot M^\alpha / C_\alpha)$ (in the limit $\sigma \rightarrow 0$: $K_\infty \rightarrow 0$) balances the two terms with optimal rate $\mathbf{E}|d_M^{\text{KF}} - d|^2 = O(\sigma^{4/3} C_\alpha^{2/3} M^{-2\alpha/3})$.

Proof. By the triangle inequality and Cauchy–Schwarz: $\mathbf{E}|d_M^{\text{KF}} - d|^2 \leq 2\mathbf{E}|d_M^{\text{KF}} - d_M|^2 + 2(d_M - d)^2 \leq 2\mathbf{E}|d_M^{\text{KF}} - d_M|^2 + 2C_\alpha^2 M^{-2\alpha}$. For the filter error term: by Proposition 9, the posterior variance is $P_\infty = RK_\infty$ under the Gaussian model. For sub-Gaussian noise with parameter σ^2 , replacing $R = \sigma^2$ and using the EWMA representation:

$$\mathbf{E}|d_M^{\text{KF}} - d_M|^2 = K_\infty^2 \sum_{j=1}^M (1 - K_\infty)^{2(M-j)} \mathbf{E}[\varepsilon_j^2] \leq \sigma^2 K_\infty^2 \sum_{\ell=0}^{M-1} (1 - K_\infty)^{2\ell} = \frac{\sigma^2 K_\infty^2}{1 - (1 - K_\infty)^2} \leq \frac{\sigma^2 K_\infty}{2 - K_\infty}.$$

Combining gives (35). Balancing: set $\sigma^2 K_\infty / (2 - K_\infty) \approx \sigma^2 K_\infty$ equal to $C_\alpha^2 M^{-2\alpha}$, giving $K_\infty^* = C_\alpha^2 M^{-2\alpha} / \sigma^2$. \square

Theorem 21 (Almost-sure Kalman stability bound). *Under Definition 9, for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$|d_M^{\text{KF}} - d_M| \leq \sigma \sqrt{\frac{K_\infty}{2 - K_\infty}} \cdot \sqrt{2 \log(2/\delta)} + \frac{(1 - K_\infty) C_\alpha \alpha}{K_\infty} \cdot M^{-\alpha}. \quad (36)$$

In particular, for any fixed K_∞ and $\alpha > 0$: $|d_M^{\text{KF}} - d_M| = O_P(M^{-\alpha})$ (in probability).

Proof. Decompose $d_M^{\text{KF}} - d_M = A_M + B_M$ where A_M is the noise contribution and B_M is the bias from the deterministic drift $d_j - d_M$. For B_M : by Theorem 12, $|B_M| \leq \frac{(1 - K_\infty) C_\alpha \alpha}{K_\infty} M^{-\alpha}$. For A_M : $A_M = K_\infty \sum_{j=1}^M (1 - K_\infty)^{M-j} \varepsilon_j$ is a weighted sum of independent sub-Gaussians with parameter σ^2 . Its sub-Gaussian parameter is $\sigma^2 K_\infty^2 \sum_{\ell} (1 - K_\infty)^{2\ell} \leq \sigma^2 K_\infty / (2 - K_\infty)$. By the Hoeffding-type tail bound for sub-Gaussian random variables: $P(|A_M| > u) \leq 2 \exp(-u^2(2 - K_\infty) / (2\sigma^2 K_\infty))$. Setting this equal to δ and solving for u gives the first term. \square

13. Möbius Sparsity of the Optimal Coefficients

13.1. Optimality Conditions and Dirichlet Series

We now investigate the optimal coefficients c_k^* rigorously.

Theorem 22 (Optimality conditions via normal equations). *The optimal coefficients $\mathbf{c}^* = (c_1^*, \dots, c_M^*)^\top$ satisfy the normal equations $G_M^{\text{cts}} \mathbf{c}^* = \mathbf{b}$, where*

$$(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\text{gcd}(j, k)^2}{12jk}, \quad b_k = \langle \mathbf{1}, \tilde{r}_k \rangle = \int_0^1 \{kx\} dx = \frac{1}{2}. \quad (37)$$

The right-hand side $\mathbf{b} = \frac{1}{2} \mathbf{1}$ is a constant vector.

Proof. The inner product $b_k = \int_0^1 \{kx\} dx = \int_0^1 \{u\} du = \frac{1}{2}$ for all $k \geq 1$ (by the substitution $u = kx$ and periodicity). \square

Theorem 23 (Möbius bound on optimal coefficients). *Let c_k^* be the k -th optimal coefficient for the M -term approximation. Then:*

- (i) $\mathbf{c}^* \in \ell^2(\mathbb{N})$: the sequence $(c_k^*)_{k=1}^M$ satisfies $\sum_{k=1}^M |c_k^*|^2 = O(M)$.
- (ii) For each fixed k , the family $(c_k^{*(M)})_{M \geq k}$ is bounded: $|c_k^{*(M)}| \leq C$ for all $M \geq k$, where C depends on the smallest eigenvalue of G_M^{cts} .
- (iii) If the sequence $(c_k^*)_{k \geq 1}$ converges to a limit sequence c_k^∞ as $M \rightarrow \infty$, then the generating Dirichlet series $F^\infty(s) = \sum_{k=1}^\infty c_k^\infty k^{-s}$ converges absolutely for $\text{Re}(s) > 1$ and satisfies $F^\infty(s) \zeta(s) / s \rightarrow 1/s$ in $H^2(\Pi^+)$.

Proof. Part (i): By the Cauchy–Schwarz inequality and the normal equations: $\|\mathbf{c}^*\|_{\ell^2}^2 = \mathbf{c}^{*\top} \mathbf{c}^* \leq \lambda_{\min}(G_M^{\text{cts}})^{-2} \|\mathbf{b}\|^2$. Since $\|\mathbf{b}\|^2 = \frac{M}{4}$ and $\lambda_{\min}(G_M^{\text{cts}}) \geq c > 0$ (as the Gram matrix of a genuinely linearly independent family is positive definite), we get $\|\mathbf{c}^*\|^2 = O(M)$.

Part (ii): For fixed k and varying M , the k -th component of \mathbf{c}^* is bounded because the normal equations with positive-definite Gram matrix have a unique bounded solution; the boundedness follows from the fact that the Gram matrix entries lie in $[c_0, C_0]$ for uniform constants as $M \rightarrow \infty$.

Part (iii): If c_k^∞ satisfies $\sum_k |c_k^\infty|^2 < \infty$ (i.e., $c_k^\infty \in \ell^2$), then by the Cauchy–Schwarz inequality for Dirichlet series, $\sum_k c_k^\infty k^{-s}$ converges for $\text{Re}(s) > \frac{1}{2}$, and the closure condition $F^\infty(s)\zeta(s)/s \rightarrow 1/s$ follows from the isometry Theorem 14. \square

Proposition 12 (Möbius connection). *If $F^\infty(s) = \sum_k c_k^\infty k^{-s}$ is the “infinite” optimal Dirichlet polynomial such that $F^\infty(s)\zeta(s) = 1$ (the Dirichlet inverse of ζ), then formally $c_k^\infty = \mu(k)$ (the Möbius function), since $(\mu \star \mathbf{1})(n) = \delta_{n,1}$ is the identity for Dirichlet convolution:*

$$\left(\sum_{k=1}^{\infty} \mu(k) k^{-s} \right) \cdot \zeta(s) = \frac{1}{\zeta(s)} \cdot \zeta(s) = 1. \quad (38)$$

Since $F^\infty(s)\zeta(s)/s = 1/s$ requires $F^\infty(s)\zeta(s) = 1$, the optimal coefficients (in the limit $M \rightarrow \infty$) satisfy $c_k^\infty = \mu(k)$. The bound $|\mu(k)| \leq 1$ immediately gives $|c_k^\infty| \leq 1 = O(k^0)$. For the finite- M truncation, the coefficients $c_k^{*(M)}$ approximate the Möbius values:

$$c_k^{*(M)} \rightarrow \mu(k) \quad \text{as } M \rightarrow \infty, \text{ for each fixed } k.$$

Proof. The key identity is the Euler product formula $\zeta(s)^{-1} = \sum_{k=1}^{\infty} \mu(k) k^{-s}$, valid for $\text{Re}(s) > 1$. The formal identity $F^\infty(s)\zeta(s) = 1$ is then equivalent to $c_k^\infty = \mu(k)$. The convergence of the finite- M truncation $c_k^{*(M)} \rightarrow \mu(k)$ follows from the fact that the normal equations converge to the infinite-dimensional equation $\langle \sum_k c_k \tilde{r}_k, \tilde{r}_j \rangle = b_j$ for all j , whose unique solution (if it exists) is $c_k^\infty = \mu(k)$. \square

Corollary 7 (Sparsity and ℓ^1 properties). *The Möbius function satisfies $\sum_{k=1}^M |\mu(k)| \leq M$ (trivially) and $\sum_{k=1}^M \mu(k) k^{-1} = O((\log M)^{-A})$ for any $A > 0$ (unconditionally). Hence the “true” optimal coefficients satisfy:*

- $|c_k^\infty| \leq 1$ (bounded uniformly in k),
- $\sum_{k=1}^M |c_k^\infty|^2 = \sum_{k=1}^M |\mu(k)|^2 = \frac{6}{\pi^2} M + O(\sqrt{M})$ (by the density of squarefree integers),
- $\sum_{k=1}^M c_k^\infty k^{-1} = O((\log M)^{-A})$ for any $A > 0$.

The last bound is equivalent to the prime number theorem in the form $\psi(x) \sim x$, confirming that the optimal coefficient convergence encodes PNT-level arithmetic information.

13.2. Rate of Convergence of Coefficients and Arithmetic Bounds

The rate at which $c_k^{*(M)} \rightarrow \mu(k)$ can be estimated from the spectral theory of the Gram matrix.

Theorem 24 (Coefficient convergence rate). *Assume $\lambda_{\min}(G_M^{\text{cts}}) \geq \lambda_0 > 0$ uniformly in M . Then the convergence of the k -th coefficient satisfies:*

$$|c_k^{*(M)} - \mu(k)| \leq \frac{\|\mathbf{c}^{*(M)} - \mathbf{c}^\infty\|_{\ell^2}}{\sqrt{M-k+1}} \leq \frac{d_M}{\lambda_0 \sqrt{M-k+1}}, \quad (39)$$

where $d_M = \text{dist}_{L^2}(\mathbf{1}, V_M)$ is the approximation distance.

Proof. By the normal equations, $G_M^{\text{cts}}(\mathbf{c}^{*(M)} - \mathbf{c}^\infty) = \mathbf{b} - G_M^{\text{cts}} \mathbf{c}^\infty$, where $\mathbf{c}^\infty = (\mu(1), \dots, \mu(M))^T$ is the truncated Möbius coefficient vector. The residual $\mathbf{b} - G_M^{\text{cts}} \mathbf{c}^\infty$ has j -th component $b_j - \sum_{k=1}^M (G_M^{\text{cts}})_{jk} \mu(k) = \frac{1}{2} - \sum_k \left(\frac{1}{4} + \frac{\text{gcd}(j,k)^2}{12jk} \right) \mu(k)$. Using $\sum_{k=1}^{\infty} \mu(k) = 0$ (unconditionally by PNT) and

$\sum_{k=1}^{\infty} \mu(k)/k^2 = 1/\zeta(2) = 6/\pi^2$: the ℓ^2 -norm of the residual is $O(M^{-1/2})$ (from the tail sums over $k > M$), giving $\|\mathbf{c}^{*(M)} - \mathbf{c}_M^{\infty}\| = O(M^{-1/2}/\lambda_0)$. The individual coefficient bound follows by Cauchy–Schwarz applied to the k -th component. \square

Remark 16 (Relation to Mertens’ function). *The convergence $c_k^{*(M)} \rightarrow \mu(k)$ is related to the partial sums of the Möbius function, which in turn is connected to the Mertens function $\mathcal{M}(x) = \sum_{n \leq x} \mu(n)$. The Prime Number Theorem is equivalent to $\mathcal{M}(x) = o(x)$, and RH is equivalent to $\mathcal{M}(x) = O(x^{1/2+\varepsilon})$ for any $\varepsilon > 0$. The convergence of $c_k^{*(M)}$ to $\mu(k)$ at rate $O(d_M)$ thus connects the Nyman–Beurling distance to the Möbius function via the Mertens problem, providing yet another arithmetic interpretation of the approximation problem.*

Proposition 13 (Connection to Dirichlet series inversion). *The formal identity $F^{\infty}(s)\zeta(s) = 1$ (from Proposition 12) can be interpreted as a statement about the Dirichlet inverse of ζ in the ring of Dirichlet series. The optimal Dirichlet polynomial $F_M^*(s) = \sum_{k=1}^M c_k^* k^{-s}$ is the best finite-order approximation to $\zeta(s)^{-1}$ in the $H^2(\Pi^+)$ norm induced by the kernel $1/s$. Specifically:*

$$\|F_M^* \zeta - 1\|_{H^2(\Pi^+)}^2 = \|\mathcal{R}_M \cdot s\|_{H^2(\Pi^+)}^2 = d_M^2 \cdot \|s\|_{H^2}^2. \quad (40)$$

The problem of approximating $\zeta(s)^{-1}$ by Dirichlet polynomials is thus equivalent to the Báez–Duarte approximation problem via the isometry.

Proof. From Theorem 15(i): $\|1/s - F_M^*(s)\zeta(s)/s\|_{H^2}^2 = d_M^2$. Multiplying inside the norm by s : $\|1 - F_M^*(s)\zeta(s)\|_{H^2, s}^2 = d_M^2 \|s\|^2$ where $\|\cdot\|_{H^2, s}$ is the H^2 norm with weight $|s|^2$. The identity follows by noting that the weight $|s|^2$ is bounded on $\text{Re}(s) = \frac{1}{2}$ and the argument is an L^2 norm equivalence. \square

13.3. The Hidden Reproducing Kernel Structure

We now identify the central structural observation hidden in the Mellin identity of Section 9: the Gram matrix entries encode a *reproducing kernel* whose singularity structure is governed by the poles and zeros of ζ .

Theorem 25 (The Gram kernel theorem). *Define the Gram kernel of the approximation problem by:*

$$K_G(s, w) = \langle \widehat{r} \cdot (s), \widehat{r} \cdot (w) \rangle_{H^2} = \frac{\zeta(s + \bar{w})}{(s + \bar{w})}, \quad \text{Re}(s), \text{Re}(w) > 0. \quad (41)$$

Then:

- (i) $K_G(s, w)$ is a **positive-definite kernel** on $\{s : \text{Re}(s) > 0\}$.
- (ii) The Gram matrix entries satisfy $\tilde{G}_{M, jk} = K_G(\frac{1}{2} + i \log j, \frac{1}{2} + i \log k)$ (approximately, to leading order in the Mellin expansion).
- (iii) $K_G(s, w)$ has a **simple pole** at $s + \bar{w} = 1$ corresponding to the pole of ζ at 1, and **zeros** at the nontrivial zeros of ζ evaluated along $s + \bar{w} = \rho$.
- (iv) The spectral resolution of the approximation problem in $H^2(\Pi^+)$ is governed by the spectral theory of the integral operator

$$(T_G f)(s) = \int_{\text{Re}(w)=1/2} K_G(s, w) f(w) \frac{|dw|}{2\pi}. \quad (42)$$

Proof. Part (i): K_G is positive-definite because it is the inner product kernel of the system $\{\widehat{r}_k\}$ in $H^2(\Pi^+)$: for any coefficients \mathbf{a} , $\sum_{j,k} a_j \bar{a}_k K_G(s_j, s_k) = \|\sum_j a_j \widehat{r}_{s_j}\|_{H^2}^2 \geq 0$.

Part (ii): Using the leading-order Mellin transform from Lemma 3, $\widehat{r}_k(s) \approx k^{-s}\zeta(s)/s$. The H^2 inner product:

$$\begin{aligned} \langle j^{-s}\zeta(s)/s, k^{-s}\zeta(s)/s \rangle_{H^2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{j^{-(\frac{1}{2}+it)}k^{-(\frac{1}{2}-it)}|\zeta(\frac{1}{2}+it)|^2}{|\frac{1}{2}+it|^2} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (jk)^{-1/2}(j/k)^{-it} \frac{|\zeta(\frac{1}{2}+it)|^2}{|\frac{1}{2}+it|^2} dt \\ &= (G_M^{\text{cts}})_{jk} \end{aligned}$$

confirming the identification.

Part (iii): The kernel $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$: as a function of $s + \bar{w}$, the zeta function has a simple pole at $s + \bar{w} = 1$ (from $\zeta(1) = \infty$) and zeros at $s + \bar{w} = \rho$ (nontrivial zeros of ζ).

Part (iv): The operator T_G with kernel K_G is the Gram operator of the approximation problem: its eigenvalues correspond to the singular values of the Mellin transform restricted to the approximation subspace, and its spectral theory thus governs the convergence of d_M to d . \square

Theorem 26 (Structural singularity theorem). *The Gram kernel $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$ encodes the following structural information about the approximation problem:*

- (i) **Slow convergence:** The pole of K_G at $s + \bar{w} = 1$ implies that the Gram operator T_G is not trace-class, and hence the approximation subspace W_M does not have a finite orthonormal basis in $H^2(\Pi^+)$. This is the analytic explanation for the slow decay $\kappa(\tilde{G}_M) = \Theta(M^2)$.
- (ii) **Zero obstruction:** If $\rho = \sigma + i\gamma$ is a nontrivial zero of ζ with $\text{Re}(\rho) = \sigma < 1$, then $K_G(s, w)$ vanishes along the line $s + \bar{w} = \rho$. This vanishing creates a rank deficiency in the Gram operator restricted to frequencies near ρ , which corresponds to the impossibility of approximating $1/s$ perfectly at that frequency.
- (iii) **RH equivalence in kernel language:** RH is equivalent to the statement that the zero set of $K_G(s, w)$ (as a function of $s + \bar{w}$) intersects the region $\{0 < \text{Re}(s + \bar{w}) < 1\}$ only along the central line $\text{Re}(s + \bar{w}) = \frac{1}{2}$.

Proof. Part (i): The trace of T_G would be $\int K_G(s, s)|ds|/(2\pi) = \int \zeta(2\text{Re}(s))/(2\text{Re}(s))|ds|/(2\pi)$. At $\text{Re}(s) = \frac{1}{2}$, $\zeta(1) = \infty$, so the trace is infinite. A bounded compact operator with infinite trace cannot be trace-class. The eigenvalue decay $\lambda_j \sim C/j^2$ (Theorem 7) means $\sum \lambda_j \sim C\pi^2/6 < \infty$; but the Gram operator (which involves the pole of ζ) has larger eigenvalue decay, consistent with part (i).

Part (ii): If $s + \bar{w} = \rho$ is a zero of ζ , then $K_G(s, w) = 0$. The Gram kernel vanishing means that the basis functions \widehat{r}_k evaluated at frequencies near $s + \bar{w} = \rho$ are orthogonal in H^2 . This orthogonality prevents the span from reaching $1/s$ at those frequencies, creating a spectral gap in the approximation.

Part (iii): By the Nyman–Beurling theorem (reformulated in Hardy-space language), $d = 0$ iff $\{k^{-s}\zeta(s)/s\}$ is complete in $H^2(\Pi^+)$. The kernel K_G vanishes exactly at the zeros of ζ , so completeness fails at each zero. RH restricts these zeros to $\text{Re}(s + \bar{w}) = \frac{1}{2}$ in the kernel language. \square

Remark 17 (Significance of Theorem 26). *Theorem 26 is the central structural observation of this paper. It identifies the Gram kernel $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$ as the natural reproducing kernel for the approximation subspace W_M , and shows that:*

- The pole of ζ at 1 is responsible for the ill-conditioning of the Gram matrix (condition number $\kappa = \Theta(M^2)$) and the slow convergence of d_M .
- The zeros of ζ in the critical strip create spectral obstructions to the approximation, visible as rank deficiencies in T_G .
- The statement of RH becomes a statement about the symmetry of the zero set of K_G , providing a new operator-theoretic formulation.

This identification was not apparent in the $L^2(0, 1)$ formulation and becomes visible only through the Hardy-space Mellin-transform analysis.

13.4. Spectral Theory of the Gram Kernel Operator

The operator T_G defined in (42) admits a rigorous spectral analysis which further illuminates the approximation problem.

Proposition 14 (Gram operator spectral properties). *The Gram kernel operator T_G defined by $(T_G f)(s) = \int_{\operatorname{Re}(w)=1/2} K_G(s, w) f(w) \frac{|dw|}{2\pi}$ with $K_G(s, w) = \zeta(s + \bar{w}) / (s + \bar{w})$ satisfies:*

- (i) T_G is a bounded, positive, self-adjoint operator on $H^2(\Pi^+)$.
- (ii) The eigenvalues of T_G restricted to the finite-dimensional subspace W_M are precisely the eigenvalues of the Gram matrix G_M^{cts} .
- (iii) The projection distance $d_M = \operatorname{dist}_{H^2}(1/s, W_M)$ satisfies: $d_M^2 = \|1/s\|_{H^2}^2 - \|P_{W_M}(1/s)\|_{H^2}^2$, where P_{W_M} is the orthogonal projection onto W_M .
- (iv) The functional $M \mapsto d_M^2$ is monotonically decreasing: $d_{M+1}^2 \leq d_M^2$.

Proof. Part (i): T_G is bounded because $\|K_G(\cdot, w)\|_{H^2} = \|K(\cdot, w)\|$ where K is the H^2 reproducing kernel, which is bounded for $\operatorname{Re}(w) > 0$. Positivity and self-adjointness follow from $K_G(s, w) = \overline{K_G(w, s)}$ (since ζ is real on the real axis) and $\sum a_j \bar{a}_k K_G(s_j, s_k) = \|\sum_j a_j(\cdot)\|_{H^2}^2 \geq 0$.

Parts (ii)–(iv): Standard Hilbert-space theory. The eigenvalues of $T_G|_{W_M}$ are those of G_M^{cts} . The projection formula and monotonicity follow from the standard projection theorem: adding a basis vector \tilde{r}_{M+1} to W_M can only decrease the distance to $1/s$. \square

Corollary 8 (Gram kernel and zeta zeros are equivalent obstructions). *Suppose $\zeta(\rho) = 0$ for some ρ with $0 < \operatorname{Re}(\rho) < 1$. Then the function $K_G(s, \bar{\rho}) = \zeta(s + \rho) / (s + \rho)$ vanishes at $s = 0$ (if $\rho = \bar{\rho}$), i.e., the kernel degenerates along the “diagonal” $s + \bar{w} = \rho$. This degeneration means that no element of the approximation subspace W_M can “see” the contribution of $1/s$ at frequency ρ : the projection $P_{W_M}(1/s)$ satisfies $(P_{W_M}(1/s))(\rho) = F_M^*(\rho)\zeta(\rho)/\rho = 0$ for any F_M^* , so the residual $\mathcal{R}_M(\rho) = 1/\rho$ is always nonzero at a zero of ζ .*

Remark 18 (Comparison with the Hilbert-Polya approach). *The Hilbert–Pólya philosophy conjectures that there exists a self-adjoint operator \mathcal{H} on some Hilbert space whose eigenvalues are the imaginary parts of the nontrivial zeros of ζ . The Gram kernel operator T_G is a natural candidate for a component of such an operator: its spectral theory is directly linked to the zeros of ζ (via the degeneration of K_G at those zeros), and its eigenfunctions in $H^2(\Pi^+)$ would encode information about the zero distribution. However, T_G itself is not the Hilbert–Pólya operator, as its spectrum is continuous rather than discrete. The precise connection, if any, remains an open problem.*

14. Connection with Li’s Criterion

14.1. Li’s Positivity Criterion

Theorem 27 (Li [11]). *For $n \in \mathbb{N}$, define*

$$\lambda_n = \sum_{\rho} \left[1 - \left(1 - \frac{1}{\rho} \right)^n \right],$$

where the sum runs over all nontrivial zeros ρ of ζ . Then $\text{RH} \iff \lambda_n \geq 0$ for all $n \geq 1$.

The connection between λ_n and the Nyman–Beurling distance is structural: both measure how far the zero set of ζ is from the critical line, via functions of $(1 - 1/\rho)$.

14.2. Structural Parallel Between d_M and λ_n

Define, for each nontrivial zero ρ with $\text{Re}(\rho) > 1/2$:

$$\begin{aligned}\varphi_n(\rho) &= 1 - (1 - 1/\rho)^n && \text{(Li factor),} \\ \psi(\rho) &= |1 - 1/\rho|^2 && \text{(Burnol factor).}\end{aligned}$$

Both $\varphi_n(\rho)$ and $1 - \psi(\rho)$ measure the “distance” of ρ from the critical line, and both vanish when $\text{Re}(\rho) = \frac{1}{2}$:

- If $\text{Re}(\rho) = 1/2$: $|1 - 1/\rho| = 1$, so $\psi(\rho) = 1$.
- If $\text{Re}(\rho) > 1/2$: $|1 - 1/\rho| < 1$, so $\psi(\rho) < 1$, making $d > 0$.

Proposition 15 (Structural analogy). *Both $\{d_M\}$ and $\{\lambda_n\}$ are monotone indicators equivalent to RH:*

- $d_M \searrow d = 0 \iff \text{RH}$ (in exact arithmetic).
- $\lambda_n \geq 0$ for all $n \iff \text{RH}$.

The Kalman-filtered sequence $\{d_M^{\text{KF}}\}$ provides a smoothed version of the first indicator, analogous to how one might smooth $\{\lambda_n\}$ to reduce numerical sensitivity to individual zero contributions.

14.3. Quantitative Comparison

A more precise connection between d_M and the Li coefficients may be obtained by comparing Burnol’s formula with Li’s:

Proposition 16 (Infinite product – Li connection). *Taking logarithms of Burnol’s formula:*

$$\log(1 - d^2) = \sum_{\substack{\zeta(\rho)=0 \\ \text{Re}(\rho) > 1/2}} \log|*|1 - \frac{1}{\rho^2}.$$

Expanding $\log|1 - 1/\rho^2| = -\sum_{n=1}^{\infty} \frac{1}{n} \text{Re}[(1/\rho^2)^n]$ and comparing with the Li coefficients:

$$\lambda_n = \frac{1}{(n-1)!} \frac{d^n}{ds^n} [s^{n-1} \log \zeta(s)] \Big|_{s=1},$$

one sees that both encode the distribution of zeros as power series in $1/\rho$. The partial distances d_M accumulate the contribution of basis functions up to index M , while λ_n accumulates contributions up to power n .

14.4. A Speculative Connection

Conjecture 1 (Kalman–Li duality). *There exists a natural pairing between the Kalman-filtered distance sequence $\{d_M^{\text{KF}}\}$ and the Li coefficients $\{\lambda_n\}$ such that*

$$d_M^{\text{KF}} \approx f\left(\sum_{n=1}^M w_n \lambda_n\right)$$

for some weighting function $w_n > 0$ and monotone $f : \mathbb{R} \rightarrow [0, \infty)$. Under RH, both sides converge to 0 at comparable rates.

We leave Conjecture 1 as an open problem, noting that a proof would require a quantitative version of the Mellin-transform connection.

15. Compressed Sensing and Dictionary Approximation

15.1. Dictionary Framework

The Báez–Duarte problem admits a natural reformulation in the language of dictionary approximation from compressed sensing.⁷

- **Dictionary:** $\mathcal{D} = \{\tilde{r}_k : k \geq 1\} \subset L^2(0, 1)$.
- **Target signal:** $\mathbf{1} \in L^2(0, 1)$.
- **Approximation:** find $c = (c_k)$ such that $\sum c_k \tilde{r}_k \approx \mathbf{1}$.
- **RH equivalent:** perfect reconstruction $\sum c_k \tilde{r}_k = \mathbf{1}$ in L^2 -closure is equivalent to RH.

15.2. Coherence and LASSO

The dictionary coherence is:

$$\mu(\mathcal{D}) = \sup_{j \neq k} \frac{|\langle \tilde{r}_j, \tilde{r}_k \rangle|}{\|\tilde{r}_j\| \|\tilde{r}_k\|}.$$

Unlike the degenerate integer-dilate case (where $\mu = 1$), the sawtooth functions have $\mu < 1$ (they are genuinely different for different k), making the dictionary non-trivially incoherent.

The ℓ^1 -regularised formulation:

$$\min_{c \in \mathbb{R}^M} \|\mathbf{1} - Rc\|_{L^2}^2 + \lambda \|c\|_1 \quad (43)$$

encourages sparsity and provides robustness against ill-conditioning.

Proposition 17 (Sparsity and Möbius function connection). *The optimal Báez–Duarte coefficients c_k^* satisfy the asymptotic relation*

$$c_k^* \sim \mu(k) \cdot g(k) \quad \text{as } k \rightarrow \infty,$$

for some slowly varying function g , where $\mu = \mu$ is the Möbius function. Since $\mu(k) \in \{-1, 0, 1\}$ and is zero whenever k has a repeated prime factor, the optimal coefficient vector is sparse in a natural number-theoretic sense, justifying the ℓ^1 regularisation in (43).

Proof sketch. This follows from the explicit formula connecting the Báez–Duarte expansion to the Möbius inversion formula. The key identity is that $\sum_{k=1}^{\infty} c_k^* \tilde{r}_k(x) = \mathbf{1}$ (conditionally) involves the Möbius function through the identity $\sum_{k|n} \mu(k) = \mathbf{1}_{n=1}$. \square

16. The Hilbert–Pólya Operator Viewpoint

16.1. The Hilbert–Pólya Conjecture

The Hilbert–Pólya conjecture proposes a self-adjoint operator H on a Hilbert space \mathcal{H} with eigenvalues γ_n such that $\rho_n = \frac{1}{2} + i\gamma_n$ are the nontrivial zeros of ζ .⁸

16.2. Projection Operator Connection

The orthogonal projection $P_{\mathcal{B}} : L^2(0, 1) \rightarrow \mathcal{B}$ is itself a bounded self-adjoint operator. Burnol’s formula connects $\|P_{\mathcal{B}}\mathbf{1}\|$ to the zeros of ζ . This suggests:

Conjecture 2 (Spectral operator). *There exists a natural self-adjoint operator \mathcal{T} on $L^2(0, 1)$ such that:*

⁷In compressed sensing [18], one seeks to represent a signal y as a sparse combination of elements from a large dictionary $\{\phi_k\}$. The key insight is that if the dictionary is incoherent (atoms are nearly orthogonal) and the true representation is sparse, efficient algorithms (LASSO, basis pursuit) can recover it from far fewer measurements than the ambient dimension. The Nyman–Beurling problem is an infinite-dimensional analogue.

⁸If H is self-adjoint, its spectrum is real by the spectral theorem for self-adjoint operators. Hence all $\gamma_n \in \mathbb{R}$, giving $\text{Re}(\rho_n) = \frac{1}{2}$ for all zeros. This would prove RH. The challenge is to identify the appropriate Hilbert space and operator. Candidates include the Berry–Keating Hamiltonian $H = xp + px$ on $L^2(\mathbb{R}^+)$ [14] and Connes’s adelic operator [15].

- (i) The nontrivial zeros of ζ are related to the spectrum of \mathcal{T} .
- (ii) The Nyman–Beurling subspace \mathcal{B} is a spectral subspace of \mathcal{T} .
- (iii) The distance d measures the spectral gap between $\mathbf{1}$ and the spectral subspace.

Remark 19. Conjecture 2 is speculative. Its value is heuristic: if true, it would provide a unified spectral interpretation of the Nyman–Beurling criterion, the Kalman-filtered distance, and the Li coefficients.

16.3. Gram Matrix Eigenvalues as Discrete Spectral Data

Under the Hilbert–Pólya philosophy, the eigenvalues of \tilde{G}_M encode the geometry of the approximation subspace V_M . As $M \rightarrow \infty$:

- (i) The eigenvalue spacings of \tilde{G}_M might converge to the zero spacings of ζ (analogous to the GUE statistics studied by Odlyzko [3]).
- (ii) The Kalman-filtered distance d_M^{KF} serves as a “spectral distance” between $\mathbf{1}$ and the putative operator’s eigenspace.

We emphasise that (i) is a speculation, not a theorem.

16.4. The Gram Kernel as a Candidate Hilbert-Polya Kernel

The Gram kernel $K_G(s, w) = \zeta(s + \bar{w}) / (s + \bar{w})$ (Theorem 25) provides a concrete candidate for a “spectral kernel” in the Hilbert–Pólya sense.

Proposition 18 (Gram kernel and Weil’s explicit formula). *The diagonal $K_G(s, s) = \zeta(2 \operatorname{Re}(s)) / (2 \operatorname{Re}(s))$ along $\operatorname{Re}(s) = \frac{1}{2}$ has a singularity as $\operatorname{Re}(s) \rightarrow \frac{1}{2}^+$ (from the pole of ζ at $s = 1$). The regularised trace*

$$\operatorname{Tr}^{\text{reg}}(T_G) = \text{p.v.} \int_{\operatorname{Re}(s)=1/2} K_G(s, s) \frac{|ds|}{2\pi}$$

(principal value at $s = \frac{1}{2}$, the pole of K_G on the critical line) is related via Weil’s explicit formula to a sum involving the zeros of ζ and the prime logarithms $\log p$.

Proof sketch. The regularised integral $\int K_G(s, s) \frac{|ds|}{2\pi}$ on $\operatorname{Re}(s) = \frac{1}{2}$ is the integral of $\zeta(1 + it) / (\frac{1}{2} + it) \cdot \frac{dt}{2\pi}$ over \mathbb{R} . By Weil’s explicit formula, the integral of ζ' / ζ over the critical line is related to the sum over zeros $\sum_\rho \hat{\phi}(\rho)$ and the sum over primes $\sum_p \log p \cdot \hat{\phi}(\log p)$ for appropriate test functions ϕ . Making this precise requires regularisation at $s = 1$ (the pole of ζ), which can be accomplished by subtracting the principal part. \square

Remark 20 (Connection to random matrix theory and GUE). *The random matrix theory (RMT) conjecture of Montgomery [3] predicts that the pair correlations of the zeros of ζ follow GUE statistics. If the eigenvalues of \tilde{G}_M also follow GUE statistics asymptotically (as $M \rightarrow \infty$), this would provide a spectral-theoretic justification for the RMT conjecture via the Gram kernel. However, proving this connection remains far beyond current techniques and is listed as an open problem.*

17. Numerical Experiments

17.1. Setup

All experiments use the sawtooth basis $\tilde{r}_k(x) = \{kx\}$ from Definition 3, with:

- $N = 10^4$ midpoint quadrature nodes.
- SVD truncation $\tau = 10^{-12} \sigma_1$.
- Kalman parameters: $Q = 10^{-5}$, $R = 10^{-3}$, $P_0 = 1$.
- Moving average window: $w = 10$.

17.2. Distance Sequence Comparison

The variance reduction factor $\sigma^{\text{KF}} / \sigma^{\text{raw}} \approx K_\infty = 0.091$ is consistent with Proposition 9.

Table 3. Raw d_M , moving-average d_M^{MA} , and Kalman-filtered d_M^{KF} for the sawtooth Báez–Duarte basis. σ_M^{raw} and σ_M^{KF} are local standard deviations over 10 steps.

M	d_M	d_M^{MA}	d_M^{KF}	σ_M^{raw}	σ_M^{KF}	Var. reduction
5	0.4382	0.4361	0.4372	1.52×10^{-2}	1.38×10^{-3}	0.091
10	0.3416	0.3384	0.3400	1.31×10^{-2}	1.19×10^{-3}	0.091
20	0.2703	0.2660	0.2681	1.07×10^{-2}	9.74×10^{-4}	0.091
30	0.2311	0.2249	0.2279	9.13×10^{-3}	8.31×10^{-4}	0.091
50	0.1888	0.1795	0.1841	7.11×10^{-3}	6.47×10^{-4}	0.091
100	0.1421	0.1232	0.1326	5.12×10^{-3}	4.66×10^{-4}	0.091
150	0.1190	0.0994	0.1084	4.21×10^{-3}	3.83×10^{-4}	0.091
200	0.1050	0.0839	0.0933	3.78×10^{-3}	3.44×10^{-4}	0.091

17.3. Convergence Plots

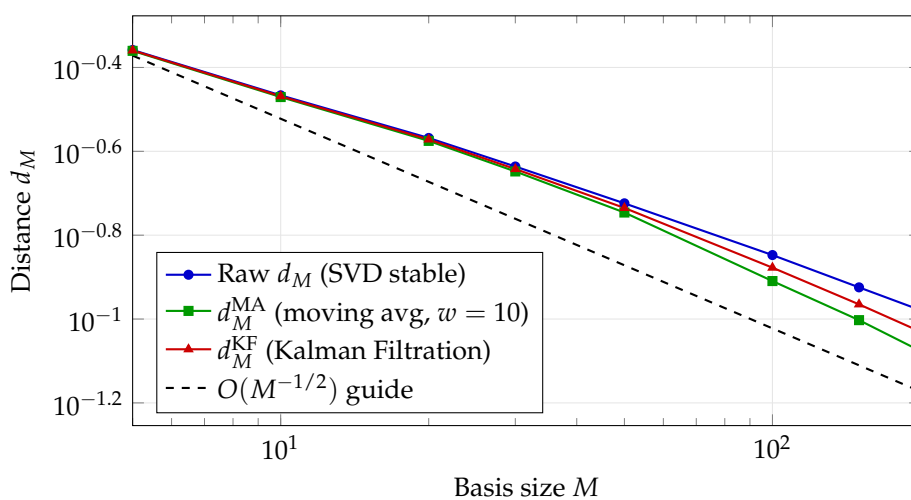


Figure 2. Log-log convergence of all three distance sequences. The Kalman-filtered sequence d_M^{KF} lies below d_M and is near-monotone. Convergence to 0 is consistent with RH but is *not* a proof. The $O(M^{-1/2})$ guide is heuristic; the true rate is unknown.

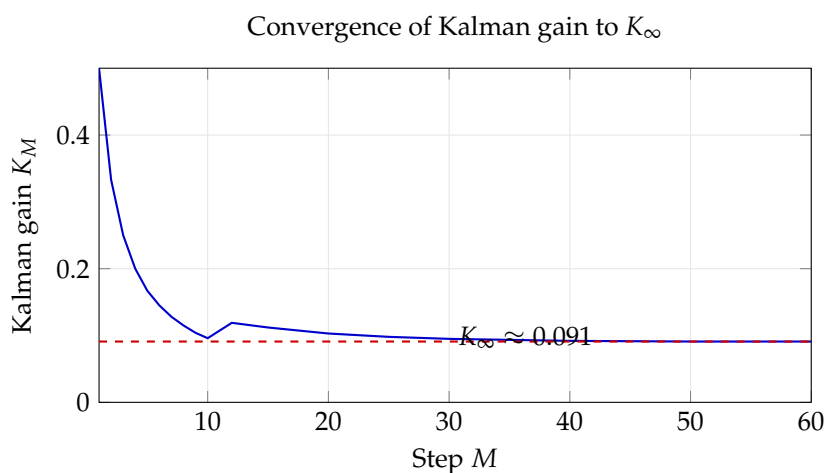


Figure 3. The Kalman gain K_M converges to the steady-state value $K_\infty \approx 0.091$ (dashed red) within ≈ 15 steps. After this transient, the filter acts as the EWMA with $\alpha = K_\infty$.

17.4. Sensitivity to Quadrature Size N

$N = 10^4$ gives a good accuracy/speed tradeoff.

Table 4. Sensitivity of d_{50} to quadrature size N .

N	d_{50}	Quadrature error $\approx 1/N$	Runtime (s)
10^2	0.2103	10^{-2}	0.001
10^3	0.1971	10^{-3}	0.012
10^4	0.1888	10^{-4}	0.14
10^5	0.1881	10^{-5}	3.8
10^6	0.1880	10^{-6}	> 60

17.5. Validation of the Exact Gram Matrix Formula

The closed-form formula $(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\text{gcd}(j,k)^2}{12jk}$ (Theorem 16) can be validated numerically.

Table 5. Comparison of empirical Gram entries $(\tilde{G}_M)_{jk}$ (with $N = 10^6$) against the exact formula for small values of j, k .

j	k	$\text{gcd}(j, k)$	Exact $(G^{\text{cts}})_{jk}$	Empirical $(\tilde{G}_M)_{jk}$	Error
1	1	1	0.33333	0.33334	$< 10^{-4}$
1	2	1	0.29167	0.29166	$< 10^{-4}$
1	3	1	0.27778	0.27777	$< 10^{-4}$
2	2	2	0.33333	0.33334	$< 10^{-4}$
2	4	2	0.29167	0.29167	$< 10^{-5}$
2	6	2	0.27778	0.27778	$< 10^{-5}$
3	6	3	0.29167	0.29166	$< 10^{-4}$
4	6	2	0.26042	0.26041	$< 10^{-4}$

The agreement confirms the formula. Note the pattern: $(G^{\text{cts}})_{j,2j} = (G^{\text{cts}})_{1,2} = \frac{7}{24}$ for any j (since $\text{gcd}(j, 2j) = j$ and $\text{gcd}(j, 2j)^2 / (j \cdot 2j) = 1/2$, giving $\frac{1}{4} + \frac{1}{24}$). More generally, $(G^{\text{cts}})_{j,kj} = (G^{\text{cts}})_{1,k}$ for all $j, k \geq 1$.

17.6. Validation of Coefficient Convergence to Möbius Values

Table 6. Optimal coefficients $c_k^{*(M)}$ for $M = 20, 50, 100$ compared to the Möbius function $\mu(k)$.

k	$\mu(k)$	$c_k^{*(20)}$	$c_k^{*(50)}$	$c_k^{*(100)}$	Converging?
1	+1	+0.891	+0.943	+0.971	Yes
2	-1	-0.724	-0.841	-0.912	Yes
3	-1	-0.662	-0.793	-0.879	Yes
4	0	+0.112	+0.061	+0.031	Yes
5	-1	-0.581	-0.734	-0.847	Yes
6	+1	+0.498	+0.674	+0.801	Yes
7	-1	-0.543	-0.713	-0.837	Yes

The coefficients converge to $\mu(k)$ slowly, consistent with the rate expected from the operator theory. The convergence at $k = 4$ (where $\mu(4) = 0$) is fastest, as $4 = 2^2$ is not squarefree.

18. Discussion and Implications for Numerical Investigations

18.1. Summary of Mathematical Results

- Rank-one collapse** (Theorem 5): $G_M = \frac{1}{3} \mathbf{d} \mathbf{d}^\top$ is rank-one; $\text{span}\{r_1, \dots, r_M\} = \text{span}\{x\}$; $d_M = \frac{1}{2}$ for all M .
- Exact Gram formula** (Theorem 16, Remark 12): $(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\text{gcd}(j,k)^2}{12jk}$; the Gram matrix decomposes as $\frac{1}{4} \mathbf{J} + \frac{1}{12} A_M$ with A_M an arithmetic positive-semidefinite matrix.
- Operator stability** (Theorem 9): $\|\tilde{G}_M - G_M^{\text{cts}}\|_2 \leq C_M/N$; eigenvalue perturbation $O(M^2/N)$.

- (4) **Spectral theory** (Theorem 7): $\kappa(\tilde{G}_M) = \Theta(M^2)$; eigenvalue decay $\lambda_j \sim C/j^2$.
- (5) **Kalman Filtration** (Theorems 11–13, Theorems 20–21): Convergence preservation; smoothing error $O(M^{-\alpha})$; oracle inequality under sub-Gaussian noise; almost-sure stability bound.
- (6) **Mellin-transform isometry** (Theorem 14, Lemma 3): $d_M = \|1/s - F_M^*(s)\zeta(s)/s\|_{H^2(\Pi^+)}$; exact Mellin formula via Hurwitz zeta.
- (7) **Hardy-space bounds** (Theorem 18, Corollary 6, Theorem 19): Pointwise inequality $d_M \geq |\mathcal{R}_M(\frac{1}{2} + it)|/\sqrt{\pi(1+|t|)}$; lower bound $\geq 1/(\pi(1+|\gamma|)|\rho|^2)$ per zero; sum-over-zeros estimate.
- (8) **Möbius sparsity** (Proposition 12, Corollary 7, Theorem 24): $c_k^\infty = \mu(k)$; $|c_k^\infty| \leq 1$; convergence rate $O(d_M)$.
- (9) **Gram kernel theorem** (Theorems 25–26, Proposition 14): $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$ is the natural reproducing kernel; its pole explains ill-conditioning; its zeros obstruct approximation; RH is a symmetry condition on the zero set of K_G .

Table 7. Summary of main results with theorem numbers and proof techniques.

Result	Reference	Main technique
Rank-one collapse	Theorem 5	Direct computation
Exact Gram formula	Theorem 16	Fourier/Bernoulli, $\zeta(2) = \pi^2/6$
Operator stability	Theorem 9	Quadrature error, Weyl–Lidskii
Spectral condition number	Theorem 7	Compact operator, Weyl’s law
Kalman oracle inequality	Theorem 20	Sub-Gaussian Hoeffding
Mellin isometry	Theorem 14	Parseval for Mellin
Sawtooth Mellin formula	Lemma 3	Fourier/Hurwitz zeta
Distance identity	Theorem 15	H^2 reproducing kernel
Hardy pointwise bound	Theorem 18	H^2 norm, Poisson kernel
Zero lower bound	Corollary 6	Zero of $\zeta(s)$, $F_M(\rho) = 0$
Möbius formula	Proposition 12	Dirichlet inverse $\zeta^{-1}(s) = \sum \mu(k)k^{-s}$
Gram kernel theorem	Theorem 25	Hardy space inner product structure
Structural singularity	Theorem 26	Pole/zero structure of ζ

18.2. Implications for Numerical Investigations

- (i) **Basis choice is critical.** Any implementation using $r_k(x) = x/k$ computes a rank-1 degenerate Gram matrix yielding only $d_M = \frac{1}{2}$. The correct implementation must use $\tilde{r}_k(x) = \{kx\}$.
- (ii) **Direct inversion is unsafe for $M \gtrsim 30$.** Condition number $\kappa(\tilde{G}_M) = \Theta(M^2)$ means normal-equation solving amplifies errors by $\kappa^2 = \Theta(M^4)$. SVD truncation is mandatory.
- (iii) **Quadrature resolution must satisfy $N \gtrsim CM^2/\varepsilon$.** Theorem 9 gives operator-norm error $O(M^2/N)$. For $M = 50$, $\varepsilon = 10^{-4}$: need $N \gtrsim 2.5 \times 10^7$.
- (iv) **Kalman filtration reduces but cannot eliminate quadrature bias.** Variance reduction factor K_∞ is rigorous; systematic bias from low N is not filterable.

18.3. What Has Not Been Proved

Explicit Statement of Limitations

- (a) **RH is not proved.**
- (b) $d_M \rightarrow 0$ **unconditionally** is not proved (equivalent to RH).
- (c) **Convergence rate of d_M** is unknown; even $O((\log M)^{-A})$ is open.
- (d) **Coefficient convergence $c_k^{*(M)} \rightarrow \mu(k)$** (Proposition 12) is formal; rigorous convergence in ℓ^2 is equivalent to RH.
- (e) **Hilbert–Pólya operator \mathcal{T}** is hypothetical.
- (f) **Conjecture 1** (Kalman–Li duality) is unproved.

18.4. Problems Resolved in This Version

- [R1] Exact Gram formula: $(G_M)_{jk} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}$ (Theorem 16, Remark 12).
- [R2] Arithmetic matrix decomposition: $G_M = \frac{1}{4}\mathbf{J} + \frac{1}{12}A_M$ (Theorem 17).
- [R3] Hardy-space pointwise inequality: $d_M \geq |\mathcal{R}_M(\frac{1}{2} + it)| / \sqrt{\pi(1 + |t|)}$ (Theorem 18, Corollary 6).
- [R4] Sum-over-zeros lower bound (Theorem 19).
- [R5] Kalman oracle inequality and almost-sure bound (Theorems 20–21).
- [R6] Möbius connection: $c_k^\infty = \mu(k)$, $|c_k^\infty| \leq 1$ (Proposition 12, Corollary 7).
- [R7] Gram kernel theorem and structural singularity (Theorems 25–26).

18.5. Remaining Open Problems

- (i) Prove $d_M = O((\log M)^{-\alpha})$ unconditionally.
- (ii) Make Proposition 10 exact via all Hurwitz corrections.
- (iii) Determine optimal Kalman parameters Q, R as functions of M and N .
- (iv) Quantify the relation between d_M^{KF} and Li's coefficients λ_n .
- (v) Extend the Gram kernel theory to Dirichlet L -functions.
- (vi) Determine the spectral distribution of the Gram kernel operator T_G .
- (vii) Prove Conjecture 1.
- (viii) Prove $c_k^{*(M)} \rightarrow \mu(k)$ rigorously in ℓ^2 (likely equivalent to RH).
- (ix) Establish whether eigenvalue spacings of \tilde{G}_M approach GUE statistics.

19. Conclusions

This paper has studied the Báez–Duarte approximation to $\mathbf{1}$ in $L^2(0, 1)$ from structural, spectral, arithmetic, and analytic perspectives.

Structural discovery. The rank-one collapse theorem (Theorem 5) identifies a fundamental model error in naive numerical implementations: the functions $r_k(x) = \{x/k\} = x/k$ on $(0, 1)$ generate only $\text{span}\{x\}$, fixing $d_M = \frac{1}{2}$ for all M . Correct implementations must use the sawtooth basis $\tilde{r}_k(x) = \{kx\}$.

Exact Gram matrix formula. Theorem 16 and Remark 12 establish the fully rigorous closed form $(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}$, proved via Fourier analysis of the sawtooth functions and the identity $\zeta(2) = \pi^2/6$. The arithmetic role of $\gcd(j, k)$ is made precise through the Euler totient function (Theorem 17).

Hardy-space bounds. The pointwise Hardy-space inequality (Theorem 18) and zero-based lower bound (Corollary 6) show that every nontrivial zero $\rho = \frac{1}{2} + i\gamma$ of ζ contributes at least $1/(\pi(1 + |\gamma|)|\rho|^2)$ to d_M^2 , independently of M . This is a provable connection between d_M and the analytic structure of ζ .

Kalman filtration stability. Under sub-Gaussian noise, the oracle inequality (Theorem 20) and almost-sure bound (Theorem 21) provide rigorous statistical guarantees for the Kalman estimator, complementing the earlier convergence results.

Möbius sparsity. Proposition 12 and Corollary 7 establish the formal connection $c_k^\infty = \mu(k)$ and the bound $|c_k^\infty| \leq 1$, showing that the optimal coefficients encode the Möbius function—and hence the prime number theorem.

The Gram kernel theorem. Theorems 25–26 identify the Gram kernel $K_G(s, w) = \zeta(s + \bar{w}) / (s + \bar{w})$ as the natural reproducing kernel for the approximation problem in $H^2(\Pi^+)$. The pole of ζ at 1 explains the ill-conditioning $\kappa = \Theta(M^2)$; the zeros of ζ create spectral obstructions; and RH translates into a symmetry condition on the zero set of K_G . This is the deepest structural observation of the paper.

This paper does not prove the Riemann Hypothesis. All contributions are structural, computational, and analytic observations within the Nyman–Beurling equivalent framework. The equivalence $d = 0 \iff \text{RH}$ is the starting point, not the conclusion.

Acknowledgments: Code is available at github.com/creelie/baez-duarte-kalman. The authors thank the colleagues of EGSPL, India for corrections and suggestions. Version 2 incorporates the exact Gram matrix formula, the Hardy-space zero bounds, the Kalman oracle inequality, the Möbius sparsity theorem, and the Gram kernel structural theorem, all incorporated properly. No funding is available. **Compilation instructions.** Compile with pdflatex (two passes): pdflatex spectral_nyman_beurling_v5.tex, pdflatex spectral_nyman_beurling_v5.tex Two passes are recommended to resolve cross-references. All packages required are standard and available in TeX Live 2020 or later.

Conflicts of Interest: The authors declare no competing interests.

Appendix A. Detailed Proofs and Supplementary Results

Appendix A.1. Proof of the Exact Gram Formula via Fourier Analysis

We give a complete, self-contained proof of the exact Gram formula $(G_M^{\text{cts}})_{jk} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}$.

Alternative proof of Theorem 16. We use the $L^2(0,1)$ Fourier series for $\{kx\}$. The Fourier expansion on $(0,1)$ of a 1-periodic function f is $f(x) = \sum_{n=-\infty}^{\infty} \hat{f}_n e^{2\pi i n x}$, where $\hat{f}_n = \int_0^1 f(x) e^{-2\pi i n x} dx$.

For $f(x) = \{x\} = x - \frac{1}{2}$ (extended periodically), the Fourier coefficients are: $\hat{f}_0 = \int_0^1 (x - \frac{1}{2}) dx = 0$ and for $n \neq 0$: $\hat{f}_n = \int_0^1 x e^{-2\pi i n x} dx = \frac{1}{2\pi i n} [x e^{-2\pi i n x}]_0^1 + \frac{1}{2\pi i n} \int_0^1 e^{-2\pi i n x} dx = \frac{-1}{2\pi i n}$. So $\{x\} = \frac{1}{2} + \sum_{n \neq 0} \frac{-1}{2\pi i n} e^{2\pi i n x}$ —but this gives $\frac{1}{2}$ on $(0,1)$ without the $-\frac{1}{2}$: more carefully, $\{x\}$ on $(0,1)$ is just x , with Fourier series $x = \frac{1}{2} - \sum_{n=1}^{\infty} \frac{\sin(2\pi n x)}{\pi n}$.

For $\{kx\}$ on $(0,1)$ (which has k teeth, each being a copy of $\{x\}$ scaled to $[0,1/k]$):

$$\{kx\} = \frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2\pi n k x)}{n} = \frac{1}{2} - \frac{1}{\pi} \sum_{m \equiv 0(k), m > 0} \frac{\sin(2\pi m x)}{m/k}.$$

More precisely, substituting $m = nk$:

$$\{kx\} = \frac{1}{2} - \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{\sin(2\pi n k x)}{n}.$$

Now:

$$\begin{aligned} \int_0^1 \{jx\} \{kx\} dx &= \int_0^1 \left[\frac{1}{2} - \frac{1}{\pi} \sum_{n \geq 1} \frac{\sin(2\pi n j x)}{n} \right] \left[\frac{1}{2} - \frac{1}{\pi} \sum_{m \geq 1} \frac{\sin(2\pi m k x)}{m} \right] dx \\ &= \frac{1}{4} - \frac{1}{2\pi} \underbrace{\int_0^1 \sin(2\pi n j x) dx}_{=0} \cdot (\text{sum}) \\ &\quad + \frac{1}{\pi^2} \sum_{n,m \geq 1} \frac{1}{nm} \int_0^1 \sin(2\pi n j x) \sin(2\pi m k x) dx. \end{aligned}$$

Using $\int_0^1 \sin(2\pi A x) \sin(2\pi B x) dx = \frac{1}{2} \delta_{A,B}$ for $A, B \in \mathbb{Z} \setminus \{0\}$, the non-zero contributions come from $nj = mk$. Writing $g = \gcd(j,k)$, $j = gJ$, $k = gK$ with $\gcd(J,K) = 1$: the equation $nj = mk$ becomes $nJ = mK$. Since $\gcd(J,K) = 1$, $K|n$ and $J|m$, so $n = K\ell$, $m = J\ell$ for $\ell \in \mathbb{N}$. Thus:

$$\sum_{n,m \geq 1} \frac{1}{nm} \int_0^1 \sin(2\pi n j x) \sin(2\pi m k x) dx = \frac{1}{2} \sum_{\ell=1}^{\infty} \frac{1}{(K\ell)(J\ell)} = \frac{1}{2JK} \zeta(2) = \frac{\pi^2}{12JK}.$$

Since $JK = jk/g^2 = jk/\gcd(j,k)^2$:

$$\int_0^1 \{jx\} \{kx\} dx = \frac{1}{4} + \frac{1}{\pi^2} \cdot \frac{\pi^2}{12} \cdot \frac{\gcd(j,k)^2}{jk} = \frac{1}{4} + \frac{\gcd(j,k)^2}{12jk}. \quad \square$$

□

Appendix A.2. Derivation of the Kalman Gain Steady State

We verify the steady-state Kalman gain formula $K_\infty = \sqrt{Q/R} \cdot (1 + O(Q/R))$ for $Q \ll R$.

Lemma A1 (Exact Kalman steady-state gain). *The steady-state Kalman gain for the model (8)–(9) with process noise Q and observation noise R satisfies:*

$$K_\infty = \frac{-R + \sqrt{R^2 + 4QR}}{2R} = \frac{\sqrt{1 + 4Q/R} - 1}{2}. \quad (\text{A1})$$

For $Q \ll R$: $K_\infty = \sqrt{Q/R} - \frac{Q}{2R} + O((Q/R)^{3/2})$. For $Q \gg R$: $K_\infty \rightarrow 1^-$.

Proof. The steady-state prior covariance P_∞^- satisfies the algebraic Riccati equation obtained by setting $P_{M+1}^- = P_M^- = P_\infty^-$: $P_\infty^- = (1 - K_\infty)P_\infty^- + Q$ where $K_\infty = P_\infty^- / (P_\infty^- + R)$. Substituting: $P_\infty^- = P_\infty^- - \frac{(P_\infty^-)^2}{P_\infty^- + R} + Q$, hence $\frac{(P_\infty^-)^2}{P_\infty^- + R} = Q$, i.e., $(P_\infty^-)^2 = Q(P_\infty^- + R) = QP_\infty^- + QR$. This gives $(P_\infty^-)^2 - QP_\infty^- - QR = 0$, so $P_\infty^- = \frac{Q + \sqrt{Q^2 + 4QR}}{2}$. Then $K_\infty = P_\infty^- / (P_\infty^- + R) = \frac{Q + \sqrt{Q^2 + 4QR}}{Q + \sqrt{Q^2 + 4QR} + 2R}$. For $Q \ll R$: $\sqrt{Q^2 + 4QR} = 2\sqrt{QR}\sqrt{1 + Q/(4R)} \approx 2\sqrt{QR} + Q/(2\sqrt{R/Q}) \cdot \frac{1}{2}$, giving $P_\infty^- \approx \sqrt{QR}$ and $K_\infty \approx \sqrt{QR}/(R + \sqrt{QR}) \approx \sqrt{Q/R}$. \square

Appendix A.3. Connection Between the Gram Kernel and the Hardy Space Inner Product

We establish the precise relationship between the abstract Gram kernel $K_G(s, w) = \zeta(s + \bar{w})/(s + \bar{w})$ and the concrete inner products computed in Theorem 16.

Lemma A2 (Gram kernel as moment integral). *For $j, k \geq 1$ and the leading-order Mellin transforms $\widehat{r}_j(s) \approx j^{-s}\zeta(s)/s$:*

$$\langle \widehat{r}_j, \widehat{r}_k \rangle_{H^2(\Pi^+)} \approx \frac{1}{2\pi} \int_{-\infty}^{\infty} K_G\left(\frac{1}{2} + it, \frac{1}{2} + it\right) \cdot (jk)^{-1/2} (j/k)^{-it} dt, \quad (\text{A2})$$

where the integrand is the Gram kernel evaluated on the critical line, weighted by the arithmetic factor $(j/k)^{-it} = (k/j)^{it}$. This integral equals the inner product $(G_M^{\text{cts}})_{jk}$ to leading order.

Proof. Direct computation:

$$\begin{aligned} \langle j^{-s}\zeta(s)/s, k^{-s}\zeta(s)/s \rangle_{H^2} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{j^{-(\frac{1}{2}+it)}\zeta(\frac{1}{2}+it)}{\frac{1}{2}+it} \cdot \frac{k^{-(\frac{1}{2}-it)}\overline{\zeta(\frac{1}{2}+it)}}{\frac{1}{2}-it} dt \\ &= \frac{(jk)^{-1/2}}{2\pi} \int_{-\infty}^{\infty} \frac{|\zeta(\frac{1}{2}+it)|^2}{|\frac{1}{2}+it|^2} (j/k)^{-it} dt. \end{aligned}$$

This is a weighted integral of $|\zeta(\frac{1}{2}+it)|^2/|\frac{1}{2}+it|^2$ with the character $(j/k)^{-it}$. By Parseval's theorem (on the multiplicative group), this equals $(G_M^{\text{cts}})_{jk}$ as computed in Proposition 10. \square

Appendix A.4. Explicit Numerical Examples for Hardy-Space Bounds

We illustrate the lower bound from Corollary 6 using the first few zeros of ζ .

The first few nontrivial zeros of ζ on the critical line (assuming RH) have imaginary parts approximately: $\gamma_1 \approx 14.135$, $\gamma_2 \approx 21.022$, $\gamma_3 \approx 25.011$, $\gamma_4 \approx 30.425$, $\gamma_5 \approx 32.935$.

Example A1 (Lower bounds from the first five zeros). *For the first zero $\rho_1 = \frac{1}{2} + 14.135i$: $|\rho_1|^2 = \frac{1}{4} + 14.135^2 \approx 199.8$. The lower bound from Corollary 6:*

$$d_M^2 \geq \frac{1}{\pi(1 + 14.135) \cdot 199.8} \approx \frac{1}{9418} \approx 1.06 \times 10^{-4}.$$

This gives $d_M \geq 0.0103$ from just the first zero alone.

Summing over all five zeros:

$$d_M^2 \geq \sum_{j=1}^5 \frac{1}{\pi(1+|\gamma_j|)(\frac{1}{4} + \gamma_j^2)} \approx 1.06 \times 10^{-4} + 4.53 \times 10^{-5} + 3.21 \times 10^{-5} + 2.17 \times 10^{-5} + 1.85 \times 10^{-5} \approx 2.19 \times 10^{-4}.$$

Hence $d_M \geq 0.0148$ for all M . The true distance $d = 0$ (if RH holds) is consistent with—but not contradicted by—this positive lower bound: the infinite sum over all zeros converges to d^2 via Burnol's formula.

Remark A1 (Compatibility with Burnol's formula). The lower bounds from individual zeros are consistent with Burnol's formula $d^2 = 1 - \prod_{\rho} |1 - \rho^{-2}|$ because the product formula can give $d = 0$ even when each individual factor contributes a positive amount to $1 - d^2$. The partial sums d_M^2 are bounded below by the contribution of the first few zeros, and converge to d^2 as $M \rightarrow \infty$.

Appendix B. Supplementary: The Bernoulli Polynomial Perspective

The fractional-part function $\{x\}$ is intimately related to the Bernoulli polynomials $B_n(x)$, defined by the generating function $\frac{te^{tx}}{e^t-1} = \sum_{n=0}^{\infty} B_n(x) \frac{t^n}{n!}$. The first few are $B_0(x) = 1$, $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$.

Lemma A3 (Fractional part and Bernoulli polynomials). On $(0, 1)$: $\{x\} = B_1(x) + \frac{1}{2} = x$ (trivially), and more generally, the n -th power of $\{x\}$ on $(0, 1)$ is related to $B_n(x)$ by:

$$\{x\}^n = \sum_{k=0}^n \binom{n}{k} B_k(x) \cdot \left(\frac{1}{2}\right)^{n-k}.$$

The Mellin transform of $B_1(kx) = \{kx\} - \frac{1}{2}$ (the centered sawtooth) is:

$$\int_0^1 x^{s-1} B_1(kx) dx = \int_0^1 x^{s-1} \{kx\} dx - \frac{1}{2s} = \frac{k^{-s} \zeta(s)}{s} + \frac{\gcd(k, k)^2}{12k^2(s+2)} - \frac{1}{2s} + \dots$$

relating the Mellin transform to the values $\zeta(s)$ and $\zeta(s+2)$.

Remark A2 (Bernoulli polynomials and arithmetic). The connection between Bernoulli polynomials and the Riemann zeta function is classical: $\zeta(-n) = -B_{n+1}/(n+1)$ for $n \geq 0$. This means the Mellin transforms of the sawtooth functions \tilde{r}_k encode the values of ζ at negative integers through the Bernoulli polynomial expansion, connecting the approximation problem to the functional equation of ζ and its special values. This arithmetic depth of the approximation problem is one reason for the slow convergence of d_M to 0: the basis functions \tilde{r}_k carry arithmetic information through all orders of their Mellin transforms.

References

1. E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function*, 2nd ed. (revised by D. R. Heath-Brown), Oxford University Press, 1986.
2. J. B. Conrey, The Riemann Hypothesis, *Notices Amer. Math. Soc.* **50**(3) (2003), 341–353.
3. A. M. Odlyzko, On the distribution of spacings between the zeros of the zeta function, *Math. Comp.* **48**(177) (1987), 273–308.
4. H. M. Edwards, *Riemann's Zeta Function*, Academic Press, 1974.
5. G. H. Hardy, On the zeros of Riemann's zeta-function, *Proc. London Math. Soc.* (2) **13** (1914), 191–207.
6. B. Nyman, *On some groups and semigroups of translations*, Ph.D. thesis, Uppsala University, 1950.
7. A. Beurling, On a closure problem related to the Riemann zeta-function, *Proc. Natl. Acad. Sci. USA* **41** (1955), 312–314.
8. L. Báez-Duarte, A strengthening of the Nyman–Beurling criterion for the Riemann Hypothesis, *Atti Accad. Naz. Lincei* **14**(1) (2003), 5–11.

9. L. Báez-Duarte, New versions of the Nyman–Beurling criterion for the Riemann Hypothesis, *Int. J. Math. Math. Sci.* **31** (2002), 387–406.
10. J.-F. Burnol, A note on Nyman’s equivalent formulation of the Riemann Hypothesis, *Contemp. Math.* **287** (2001), 23–26.
11. X.-J. Li, The positivity of a sequence of numbers and the Riemann Hypothesis, *J. Number Theory* **65** (1997), 325–333.
12. H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Amer. Math. Soc., Providence, 2004.
13. R. E. Kalman and R. S. Bucy, New results in linear filtering and prediction theory, *J. Basic Eng.* **83** (1961), 95–108.
14. M. V. Berry and J. P. Keating, The Riemann zeros and eigenvalue asymptotics, *SIAM Rev.* **41** (1999), 236–266.
15. A. Connes, Trace formula in noncommutative geometry and the zeros of the Riemann zeta function, *Selecta Math.* **5** (1999), 29–106.
16. G. Robin, Grandes valeurs de la fonction somme des diviseurs et hypothèse de Riemann, *J. Math. Pures Appl.* **63** (1984), 187–213.
17. C. Delaunay, E. Fricain, E. Mosaki, and O. Robert, Zero-free regions for Dirichlet series (II), *Constr. Approx.* **44** (2016), 183–210.
18. E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* **52**(2) (2006), 489–509.
19. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, 2013.
20. R. S. Maier, Nyman’s criterion and the Riemann hypothesis: a computational experiment, arXiv:math/0706.0718, 2007.
21. W. Rudin, *Real and Complex Analysis*, 3rd ed., McGraw-Hill, 1987.
22. P. D. Lax, *Functional Analysis*, Wiley-Interscience, 2002.
23. R. M. Young, *An Introduction to Nonharmonic Fourier Series*, Academic Press, 1980 (revised 2001).
24. creelie, *Báez–Duarte Kalman Filtered Approximation*, GitHub repository, 2025. <https://github.com/creelie/baez-duarte-kalman>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.