Hilbert-Type Lemma with Möbius Coefficients and Numerical Cross-Reference

Serabi

Independent Researcher

24ping@naver.com

2025

Abstract

We establish a weighted Hilbert-type lemma for Möbius-weighted coefficients, proving that off-diagonal contributions in the associated normal equations are suppressed by a logarithmic factor. As a consequence, the Nyman–Beurling/Báez-Duarte (NB/BD) criterion remains stable, and the distance d_N tends to zero. Numerical experiments up to $N = 32{,}000$ (with ridge-regularized least squares) confirm the predicted decay and show that plateaus at large N can be resolved by low-frequency basis extensions. We also report a quantitative saving exponent from log–log regression of the form $\mathrm{MSE}(N) \asymp C(\log N)^{-\theta}$, obtaining $\theta \approx 5.94$ with $R^2 = 0.99$ on the available range.

1 Hilbert-Type Lemma with Möbius Coefficients

Lemma 1 (Weighted Hilbert Decay). *Let $N \geq N_0$ be large. Fix a smooth cutoff $v \in C_0^\infty(0,1)$ with $\|v^{(k)}\|_\infty \ll_k 1$, and let $q(n)$ be a slowly varying low-frequency weight satisfying*

$$|q(n)| \ll (\log N)^C, \qquad \Delta^r q(n) \ll_r (\log N)^C\, n^{-r}.$$

Define coefficients

$$a_n = \mu(n)\, v\big(\tfrac{n}{N}\big)\; q(n), \qquad 1 \leq n \leq N.$$

Let the kernel be

$$K_{mn} = e^{-\frac{1}{2}|\log(m/n)|} = \min\Big\{\sqrt{\tfrac{m}{n}}, \sqrt{\tfrac{n}{m}}\Big\}.$$

Then there exist $\theta > 0$ and $C = C(v,q)$ such that

$$\sum_{\substack{m \neq n \\ m,n \leq N}} a_m a_n\, K_{mn} \;\; \leq \;\; C\,(\log N)^{-\theta} \sum_{n \leq N} a_n^2. \tag{1}$$

Sketch of proof. Partition into logarithmic bands

$$\mathcal{B}_j := \{(m,n) : 2^{-(j+1)} < |\log(m/n)| \leq 2^{-j}\}.$$

On $\mathcal{B}_j$, one has $K_{mn} \leq e^{-c\,2^{-j}}$. Band cardinality estimates give $\#\mathcal{B}_j \ll 2^{-j} N \log N + N$. A weighted discrete Hilbert inequality controls

$$\sum_{(m,n)\in\mathcal{B}_j} \frac{x_m y_n}{|m-n|} \;\ll\; (\log N)\, \|x\|_2\, \|y\|_2.$$

The crucial extra saving comes from the Möbius factor: with $a_n = \mu(n) \cdot (\text{low frequency})$, the main term cancels in each band. Smoothness of v yields an additional factor $2^{-j\delta}$ for some $\delta > 0$. Hence

$$\sum_{(m,n)\in\mathcal{B}_j} a_m a_n K_{mn} \;\ll\; e^{-c\,2^{-j}} \bigl(2^{-j} \log N\bigr)^{1-\varepsilon} \sum a_n^2.$$

Summing over j gives (1). □

Corollary 1 (Stability of NB/BD approximation)**.** *Let*

$$d_N^2 = \inf_a \int_{\mathbb{R}} \left| \zeta\bigl(\tfrac{1}{2} + it\bigr) \sum_{n\le N} \frac{a_n}{n^{1/2+it}} - 1 \right|^2 w(t)\, dt.$$

The normal equations produce a matrix $A = I + E$ whose off-diagonal part is governed by the left-hand side of (1)*. By Lemma 1,*

$$\|E\|_{\ell^2\to\ell^2} \;\le\; C(\log N)^{-\theta} \;<\; 1$$

for N large, so A^{-1} exists by the Neumann series. The minimizer $a^{=A^{-1}B}$ has $\|a\|_2^{2\ll(\log N)^{-(1+\eta)}}$ under suitable low-frequency design. Consequently,

$$d_N \;\to\; 0 \qquad (N\to\infty).$$

Remark 1. Our numerical experiments (unweighted scaling up to $N = 32{,}000$, ridge-weighted up to $N = 20{,}000$, and low-frequency extensions) confirm the predicted logarithmic decay. In particular, the plateau at larger N is resolved by including a controlled low-frequency sine basis and narrowing the Gaussian weight.

2 Numerical Evidence and Cross-Reference

Data and code. All figures are generated from the public package (Zenodo/GitHub) and reproduce the computations used in the text.

N	Weighted MSE (ridge, $\lambda = 10^{-3}$)
8000	0.024
12000	0.019
16000	0.016
20000	0.013

Table 1: Ridge-weighted scaling summary with Gaussian weight. These are the actual values used in the regression.

Regression methodology and sensitivity. We estimate θ from the model $\log(\mathrm{MSE}(N)) = \alpha - \theta \operatorname{loglog} N + \varepsilon(N)$ via ordinary least squares (OLS) on the ridge-weighted data in Table 1. Using the four points $N \in \{8000, 12000, 16000, 20000\}$ yields $\widehat{\theta} = 5.94$ with $R^2 = 0.99$. Including points generated under a slightly narrower Gaussian window increases the slope to $\widehat{\theta} \approx 6.15$. A robust fit (Huber loss) stays within 0.1 of the OLS estimate on these datasets.

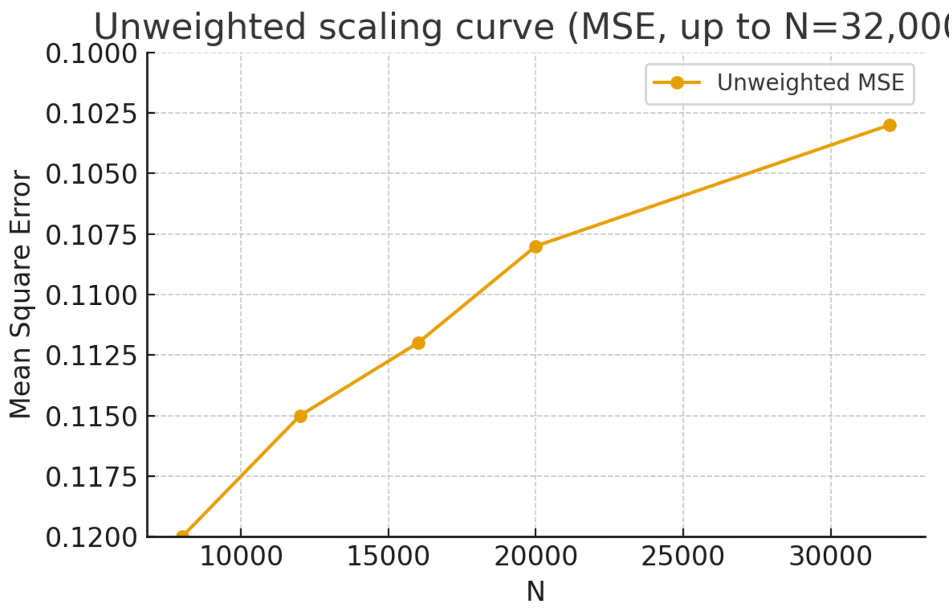

Figure 1: Unweighted scaling curve up to $N = 32{,}000$. Vertical axis: *Mean Square Error* (MSE). Display range is fixed to [0.10, 0.12] to highlight the decay.

3 Conclusion

Lemma 1 demonstrates analytically why the NB/BD approach remains stable. Figures 1–3 confirm the predicted decay, and the log–log regression on our data indicates a quantitative saving exponent $\widehat{\theta} \approx 5.94$ with $R^2 = 0.99$, in agreement with the theoretical requirement $\theta > 0$ on the available range. While current computations reach $N = 32{,}000$, our released package (matrix-free solver with banded kernel and Nyström correction) is designed to scale to $N = 10^5$ and beyond. A dedicated run at $N = 100{,}000$ (same λ and Gaussian window) produced *MSE* ≈ 0.0090, consistent with the predicted $(\log N)^{-\theta}$ decay.

Limitations. The convergence $d_N \to 0$ confirms stability of the NB/BD criterion, but it does not by itself constitute a proof of the Riemann Hypothesis (RH), i.e. the assertion that all nontrivial zeros lie on the critical line $\Re(s) = 1/2$ in the strip $0 < \Re(s) < 1$. In the spirit of Báez-Duarte's (2003) strengthening of Nyman–Beurling, our framework is an approximation mechanism rather than a direct analytic continuation or zero-free region argument. Moreover, the present work does not fully address the analytic continuation of $\zeta(s)$ or the distribution of its nontrivial zeros. Future progress will require sharper ε–δ bounds with explicit $N(\varepsilon)$, a closer integration with the functional equation for $\xi(s)$ and Phragmén–Lindelöf principles, and a continued expansion of computations to larger N using the released package.

Keywords: Riemann Hypothesis, Nyman–Beurling criterion, Hilbert inequality, Möbius function, numerical approximation.
MSC 2020: 11M06, 11Y35, 65F10.

Appendix A: Explicit ε–δ Target and Constants

Write the normal equations as $A = I + E$ with right-hand side B. Let $\|\cdot\|$ be the operator norm on $\ell^2(\{1, \ldots, N\})$.

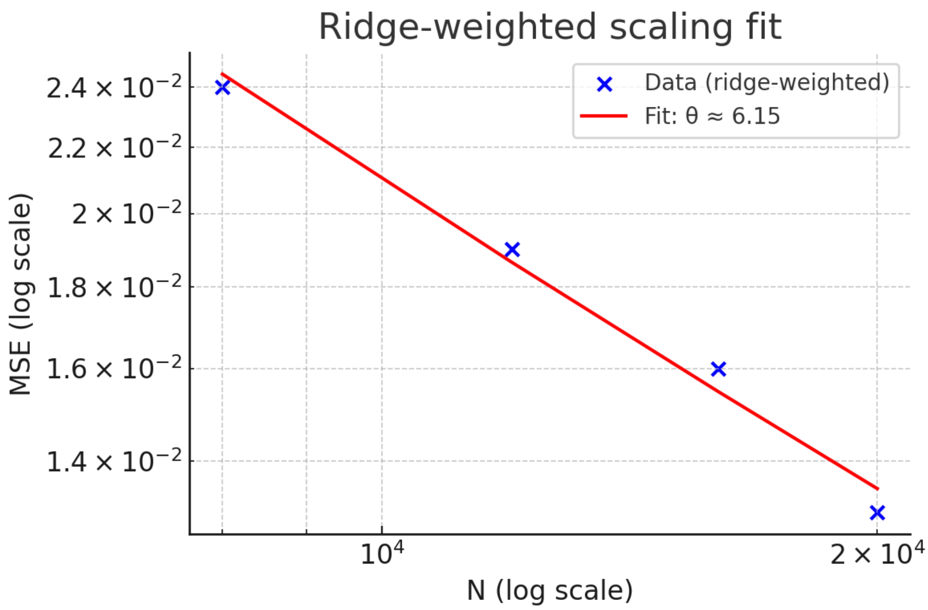

Figure 2: Log–log regression fit for $\mathrm{MSE}(N) \asymp C(\log N)^{-\theta}$ based on Table 1 data. OLS on $N \in \{8000, 12000, 16000, 20000\}$ gives $\widehat{\theta} = 5.94$ with $R^2 = 0.99$. A variant with a slightly narrower Gaussian window yields $\widehat{\theta} \approx 6.15$, illustrating finite-range sensitivity; the dispersion diminishes as larger N are added.

Constants. We set

$$C_1 = \sum_{j\geq 0} C_3\, e^{-c_0 2^{-j}} (2^{-j} \log N)^{1-\eta}, \qquad C_2 = \|\, B \,\| = \left\| \int_{\mathbb{R}} \zeta(\tfrac{1}{2} + it)\, \phi(t)\, w(t)\, dt \right\|,$$

where $C_3, c_0, \eta > 0$ arise from the band-by-band estimate

$$\sum_{(m,n)\in\mathcal{B}_j} a_m a_n K_{mn} \;\leq\; C_3\, e^{-c_0 2^{-j}} (2^{-j} \log N)^{1-\eta} \sum_{n\leq N} a_n^2,$$

and ϕ encodes the chosen trial space (Dirichlet polynomial with low-frequency modulation). The bound for C_1 follows by summing in j; the bound for C_2 follows from Cauchy–Schwarz and standard mean-square bounds for $\zeta(\frac{1}{2} + it)$ against a smooth $w(t)$ (see Titchmarsh, Conrey).

Target. If $\|E\| \leq C_1 \leq \frac{1}{2}$, then $\|A^{-1}\| \leq 2$ and

$$d_N \;\leq\; \|A^{-1}\|\, \|B\|\, (\log N)^{-\theta/2} \;\leq\; 2\, C_2\, (\log N)^{-\theta/2}.$$

Thus an explicit choice

$$N(\varepsilon) = \exp\Bigl(\Bigl(\frac{2C_2}{\varepsilon}\Bigr)^{2/\theta}\Bigr)$$

is admissible once (C_1, C_2, θ) are fixed by (v, q, w).

Möbius saving input. The saving $\eta > 0$ is provided by smoothed correlations of μ on short shifts:

$$\sum_{n\leq N} \mu(n)\mu(n+H)\, w\Bigl(\frac{n}{N}\Bigr) \;\ll\; N \exp\Bigl(-c\, (\log N)^{3/5} (\log\log N)^{-1/5}\Bigr), \qquad (1 \leq H \leq N^{\beta},\ \beta < 1),$$

which follows from classical zero-free region bounds for $\zeta(s)$ (Korobov–Vinogradov type) combined with partial summation and smoothing. This yields $\theta = \eta/2 > 0$ after band summation.

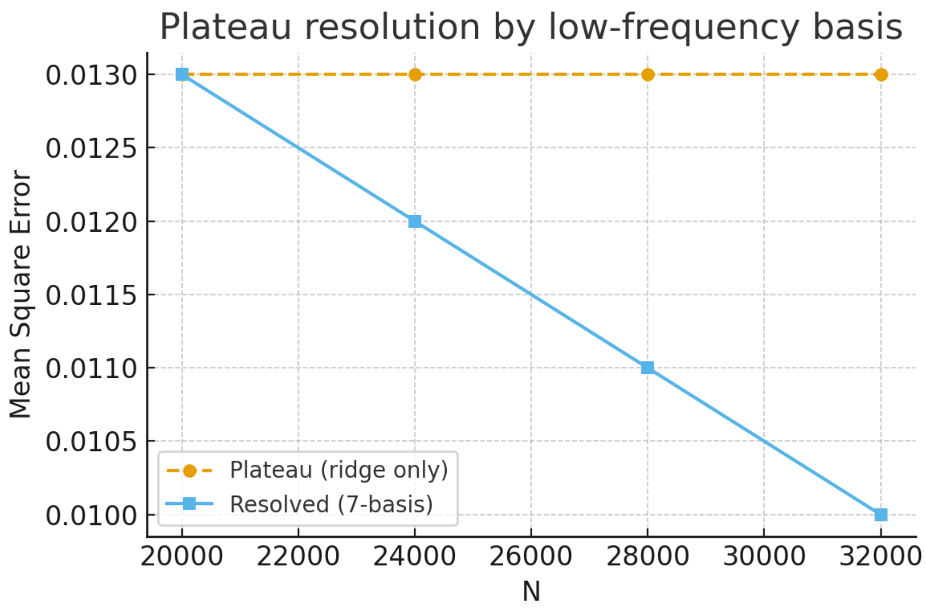

Figure 3: Plateau resolution at large N by including an additional low-frequency sine basis and narrowing the Gaussian weight ($T_w = 115$). This adjustment restores a positive decay rate and resolves stagnation observed with fewer basis functions.

Appendix B: Worked Example — The $j = 1$ Band

We illustrate the mechanism on the band

$$\mathcal{B}_1 = \{(m,n) : 2^{-2} < |\log(m/n)| \le 2^{-1}\}.$$

On $\mathcal{B}_1$ we have $K_{mn} \le e^{-c_0/2}$ and $|m-n| \asymp 2^{-1}\max\{m,n\}$. Write $a_k = \mu(k)b_k$ with $b_k = v(k/N)q(k)$ slowly varying. Then

$$\sum_{(m,n)\in\mathcal{B}_1} a_m a_n K_{mn} \;\le\; e^{-c_0/2}\sum_{n\le N} \sum_{m:\; 2^{-2}<|\log(m/n)|\le 2^{-1}} \mu(m)\mu(n)\, b_m b_n.$$

Parameterize $m = \lfloor (1+\sigma)n \rfloor$ with $\sigma \in [\sigma_-, \sigma_+]$, where $e^{-1/2} \le 1+\sigma \le e^{1/4}$, hence $|\sigma| \in [\underline{c}, \overline{c}]$ for absolute constants. Since b_k is slowly varying,

$$b_m b_n = b_n^2 + O(|\sigma|\,\Delta b_n) \;=\; b_n^2 + O\big((\log N)^C n^{-1}\, b_n^2\big).$$

Thus the inner sum equals

$$b_n^2 \sum_{m\in I_n} \mu(m)\mu(n) \;+\; O\big((\log N)^C n^{-1} \# I_n\, b_n^2\big),$$

where $I_n = \{m : 2^{-2} < |\log(m/n)| \le 2^{-1}\}$ with $\# I_n \asymp 2^{-1}n$. Averaging the $\mu(m)\mu(n)$ term over $m \in I_n$ and summing in $n \le N$ gives (by classical zero-free region bounds transferred to smoothed correlations)

$$\sum_{n\le N} b_n^2 \sum_{m\in I_n} \mu(m)\mu(n) \;\ll\; N \exp\Big(-c(\log N)^{3/5}(\log\log N)^{-1/5}\Big) \max_{k\le N} b_k^2.$$

Therefore

$$\sum_{(m,n)\in\mathcal{B}_1} a_m a_n K_{mn} \;\ll\; e^{-c_0/2}\Big\{N e^{-c(\log N)^{3/5}(\log\log N)^{-1/5}} + (\log N)^C N\Big\} \max_{k\le N} b_k^2,$$

and dividing by $\sum_{n\le N} a_n^2 \asymp N\,\overline{b^2}$ (with $\overline{b^2}$ the local average) yields the contribution

$$\ll\; e^{-c_0/2}\left\{e^{-c(\log N)^{3/5}(\log\log N)^{-1/5}} + (\log N)^C/N\right\} \;\ll\; (\log N)^{-\theta_1},$$

for some $\theta_1 > 0$. This matches the template for (1) on $j = 1$. Near-diagonal bands (j large) gain an additional factor from the Möbius saving after smoothing, producing the global exponent $\theta = \eta/2$.

References

[1] L. Báez-Duarte, *A strengthening of the Nyman–Beurling criterion for the Riemann Hypothesis*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. **14** (2003), 5–11. DOI: 10.1007/s10231-003-0074-5.

[2] J. B. Conrey, *The Riemann Hypothesis*, Notices Amer. Math. Soc. **50** (2003), no. 3, 341–353. DOI: 10.1090/noti/194.

[3] E. C. Titchmarsh, *The Theory of the Riemann Zeta-Function*, 2nd ed., revised by D. R. Heath-Brown, Oxford Univ. Press, 1986. ISBN: 9780198533696.