Diameter-Free Pointwise Gaussian KDE in Near-Additive Query Time

Bolzano team*

July 2026

Abstract

For a nonempty finite set $X \subset \mathbb{R}^d$, consider the Gaussian kernel-density estimate

$$K_X(y) = \frac{1}{|X|} \sum_{x \in X} \exp\left(-\frac{\|x-y\|_2^2}{\sigma^2}\right).$$

We construct a randomized data structure which, for every fixed query $y \in \mathbb{R}^d$, returns an additive-ε approximation to $K_X(y)$ with probability greater than $1-\delta$. Its query time is

$$O\big((d+\varepsilon^{-2})\log(d+1)\log(2/\delta)\big)$$

and its stored representation has $O((d+\varepsilon^{-2})\log(2/\delta))$ real coordinates. No bounded-domain or effective-diameter parameter occurs. The feature map is built from independent blocks $W = \sqrt{D}\,HGHB$, where D is the next power of two at least d, H is the normalized Walsh–Hadamard matrix, G is diagonal Gaussian, and B is diagonal Rademacher. Although the rows within one block are correlated, Gaussian damping and an exact signed Walsh-autocorrelation identity give a norm-uniform bound $\operatorname{Var} Y(z) \le 5/D$ for every fixed displacement z. This estimate lifts to an arbitrary KDE average without independence across data points. The guarantee is pointwise in the query; fixed finite query batches follow by a union bound, whereas uniform or adaptive queries require different control.

1 Introduction

For a finite set $X \subset \mathbb{R}^d$ and a kernel k, kernel-density estimation asks for

$$\frac{1}{|X|} \sum_{x \in X} k(x, y)$$

at query points y. For the Gaussian kernel, dense independent random Fourier features give a dimension-free statistical estimate, but evaluating all dense projections costs $O(d/\varepsilon^2)$. The algorithmic target is instead near-additive time in the input dimension and the Monte Carlo scale: $\widetilde{O}(d+\varepsilon^{-2})$.

*Dissemination note. All results produced by Bolzano are publicly available at bolzano.app. When a result appears especially promising, we give it a preliminary review and contact the authors of the source paper, inviting them to use it in any way they consider appropriate. To avoid excessive unsolicited email, we reserve this outreach for a small number of cases. This note presents a result that either fell just below our threshold for direct outreach or about which we contacted the source-paper authors but received no response. We are disseminating it more broadly in the hope that it may still be useful to the community.

1.1 Main result and proof mechanism

Theorem 2.1 attains this target under the standard pointwise randomized guarantee. After one randomized preprocessing of X, every fixed query y has an estimate $\widetilde{E}(y)$ satisfying

$$\mathbb{P}\left[\left|\frac{1}{|X|}\sum_{x\in X} k_\sigma(x,y) - \widetilde{E}(y)\right| < \varepsilon\right] > 1-\delta,$$

where the probability is over the preprocessing randomness. The query time is

$$O\big((d+\varepsilon^{-2})\log(d+1)\log(2/\delta)\big),$$

with no restriction on the diameter of the data or query domain.

The construction evaluates a correlated batch of D Gaussian Fourier features using two Walsh–Hadamard transforms. For a padded displacement $z \in \mathbb{R}^D$, one block returns

$$Y(z) = \frac{1}{D}\sum_{i=1}^{D}\cos((Wz)_i), \qquad W = \sqrt{D}\,HGHB.$$

Each row of W is marginally standard Gaussian, so the estimator is unbiased. Conditional on B, the covariance of rows separated by the Walsh index ℓ is the signed autocorrelation

$$\rho_\ell = \sum_r b_r b_{r+\ell} z_r z_{r+\ell}.$$

The Gaussian cosine covariance contributes $e^{-\|z\|^2}(\cosh\rho_\ell - 1)$, while the random signs give the exact second-moment identity

$$\mathbb{E}_B \sum_{\ell\neq 0} \rho_\ell^2 = 2\big(\|z\|_2^4 - \|z\|_4^4\big).$$

A damping inequality then yields $\operatorname{Var} Y(z) \le 5/D$, uniformly over $\|z\|_2$. Minkowski's inequality transfers the same variance bound to the empirical kernel average. Independent blocks and median amplification complete the data structure.

The paper makes three precise contributions.

(1) Theorem 2.1 gives diameter-free pointwise Gaussian KDE in near-additive query time and records preprocessing and space bounds.

(2) Proposition 3.6 proves the norm-uniform structured-feature variance bound, and Lemma 4.1 lifts it to arbitrary datasets without an independence assumption across their points.

(3) Proposition 5.1 gives an explicit one-block witness showing why the sign cancellation is not uniform after the preprocessing randomness has been fixed.

1.2 Prior bounds and the motivating question

Wagner [1, Def. 1.1] uses exactly the pointwise quantifier order above. If data and queries lie in a region of diameter Δ, write $\Delta_\sigma = \Delta/\sigma$ for the effective diameter. The Gaussian-KDE bounds summarized there are

$$O(d/\varepsilon^2), \qquad \widetilde{O}(d+\varepsilon^{-4}), \qquad \widetilde{O}(d+\Delta_\sigma^2/\varepsilon^2),$$

coming respectively from random Fourier features [3], Fast Johnson–Lindenstrauss-transform-composed random Fourier features [2, Thm. E.2 and Cor. E.3], and Fastfood-type structured features [4]; see [1, Sec. 2.1 and Table 1] for the parameter regimes. Wagner's Theorem 1.2 gives

$$\widetilde{O}(d + \varepsilon\Delta_\sigma^2 + \varepsilon^{-3})$$

through a fast spherical embedding [1, Thm. 1.2]. The paper's discussion asks whether the ideal near-additive bound can have only polylogarithmic, or no, dependence on the effective diameter.

Question 1.1 (Wagner, Section 7)**.** Is a Gaussian KDE query-time bound $\widetilde{O}(d+\varepsilon^{-2})$ possible with only polylogarithmic, or no, dependence on the effective diameter Δ_σ? More generally, what is the optimal trade-off between ε and Δ_σ?

Theorem 2.1 answers the displayed existence question for the pointwise model [1, Sec. 7]. It does not assert simultaneous success over a continuum of queries or robustness to queries selected after the random feature maps are revealed.

1.3 Structured random features

The matrix used here is a permutation-free, unscaled Fastfood block. For the standard block, which also contains a random permutation, Le, Sarlós, and Smola proved unbiasedness and analyzed variance [4, Lem. 3 and Thms. 4–5]. Their calculation already contains the conditional Gaussian covariance formula involving cosh. For n features and normalized displacement length α, however, their Theorem 4 has the additional term $C(\alpha)/n$, where

$$C(\alpha) = 6\alpha^4 \left(e^{-\alpha^2} + \frac{\alpha^2}{3} \right).$$

This grows on the order of α^6 and does not imply a diameter-free bound. The additional step supplied here is a norm-uniform completion of that variance analysis: Gaussian damping is combined with the exact sum of squared signed Walsh autocorrelations.

Feng, Hu, and Liao claim for signed circulant random features that random row signs reduce the variance to that of independent RFF features [5, Thm. 2, Eq. (12), and Cor. 2, Eq. (13)]. Their independence inference from zero covariance is invalid because the signed projections form a mixture of Gaussian pairs rather than a jointly Gaussian pair. A direct witness is available. Let $d \geq 2$, let C be a $d \times d$ circulant matrix generated by independent $N(0, \tau^2)$ entries for some $\tau > 0$, put $P = \operatorname{diag}(s_1, \ldots, s_d)C$, and take $v = (t/\sqrt{d})\mathbf{1}$. Every coordinate of Cv is the same nondegenerate Gaussian variable Z, hence

$$\frac{1}{d} \sum_{j=1}^{d} \cos((Pv)_j) = \cos Z.$$

Here $\nu^2 = \operatorname{Var} Z = t^2\tau^2$, and this average has variance

$$\frac{1}{2} \left(1 - e^{-\nu^2} \right)^2,$$

not this quantity divided by d. Thus that statement does not supply the diameter-free theorem proved here.

Broader structured-feature frameworks cover Fastfood, circulant, Toeplitz, and low-displacement-rank matrices [6, Secs. 3.2–4]. Their stated bounds are controlled by coherence and structural parameters and are formulated for bounded inputs; they do not yield the global norm-uniform one-block variance estimate used below. The remainder of the paper proves the main theorem and then isolates the pointwise scope of the sign cancellation.

2 The pointwise KDE theorem

Theorem 2.1 (Pointwise Gaussian KDE without diameter dependence)**.** *Let $d \geq 1$, let $\varnothing \neq X \subset \mathbb{R}^d$ be finite, let $\sigma > 0$, and let*

$$k_\sigma(x, y) = \exp\left(-\frac{\|x-y\|_2^2}{\sigma^2}\right).$$

For every $0 < \varepsilon < 1$ and $0 < \delta < 1$, there is a randomized KDE data structure such that, for every fixed query $y \in \mathbb{R}^d$,

$$\mathbb{P}\left[\left|\frac{1}{|X|}\sum_{x \in X} k_\sigma(x, y) - \widetilde{E}(y)\right| < \varepsilon\right] > 1 - \delta.$$

Its query time is

$$O\big((d + \varepsilon^{-2}) \log(D+1) \log(2/\delta)\big),$$

where D is the next power of two at least d. It uses $O((d + \varepsilon^{-2}) \log(2/\delta))$ stored real numbers. In particular the query time is $\widetilde{O}(d + \varepsilon^{-2})$ and has no dependence on the effective diameter Δ_σ. Its preprocessing time is

$$O\big(|X|(d + \varepsilon^{-2}) \log(D+1) \log(2/\delta)\big).$$

The theorem is proved in Sections 3 and 4. Padding vectors by zeros allows us to work in $\mathbb{R}^D$, where $D = 2^m$ and $d \leq D < 2d$ for $d \geq 1$. The query-time bound counts ordinary arithmetic operations and evaluations of sine and cosine. As in standard random-feature KDE structures, the data structure stores the empirical feature mean.

3 A fast Gaussian block

Let H be the normalized $D \times D$ Walsh–Hadamard matrix. Index rows and columns by the group $\mathbb{F}_2^m$, so

$$H_{ij} = D^{-1/2}\chi_i(j), \qquad \chi_i(j) = (-1)^{i \cdot j}.$$

Let $B = \operatorname{diag}(b_r)$, where the b_r are independent random signs, and let $G = \operatorname{diag}(g_j)$, where the g_j are independent standard Gaussian random variables. All variables in B and G are mutually independent. Define

$$W = \sqrt{D}\, HGHB.$$

For a fixed displacement $z \in \mathbb{R}^D$, define the one-block estimator

$$Y(z) = \frac{1}{D}\sum_{i \in \mathbb{F}_2^m} \cos((Wz)_i).$$

The goal of this section is to prove

$$\mathbb{E}Y(z) = e^{-\|z\|_2^2/2}, \qquad \operatorname{Var}(Y(z)) \leq \frac{5}{D}.$$

Lemma 3.1 (Row marginals)**.** *For every row index i, conditional on B, the ith row w_i of W is distributed as $N(0, I_D)$.*

Proof. Let h_j^T denote the jth row of H. The vectors $h_j^T B$ form an orthonormal basis, because H is orthogonal and B is diagonal with signs. The ith row of W is

$$w_i^T = \sqrt{D}\, e_i^T HGHB = \sum_{j \in \mathbb{F}_2^m} \chi_i(j)\, g_j\, h_j^T B.$$

This is a Gaussian linear combination of an orthonormal basis with independent $N(0,1)$ coefficients, up to deterministic signs. Hence $w_i \sim N(0, I_D)$ conditional on B. $\square$

Lemma 3.2 (Walsh covariance parameters)**.** *Fix $z \in \mathbb{R}^D$, set $t = \|z\|_2$, and put $a = HBz$. For $R_i = (Wz)_i$, conditional on B the vector $R = (R_i)_i$ is centered Gaussian,* $\mathrm{Var}(R_i \mid B) = t^2$*, and for $i \neq k$,*

$$\mathrm{Cov}(R_i, R_k \mid B) = \rho_{i+k},$$

where

$$\rho_\ell = \sum_{r \in \mathbb{F}_2^m} b_r b_{r+\ell} z_r z_{r+\ell}.$$

For $\ell = 0$ this formula gives $\rho_0 = t^2$.

Proof. Since $a = HBz$,

$$R_i = \sqrt{D} \sum_j H_{ij} g_j a_j.$$

Thus

$$\mathrm{Cov}(R_i, R_k \mid B) = D \sum_j H_{ij} H_{kj} a_j^2.$$

For $i = k$ this is $\sum_j a_j^2 = \|z\|_2^2$, because HB is orthogonal. For $i \neq k$, write $\ell = i + k$ in $\mathbb{F}_2^m$. Expanding $a_j = D^{-1/2} \sum_r \chi_j(r) b_r z_r$ gives

$$\begin{aligned} D \sum_j H_{ij} H_{kj} a_j^2 &= \sum_j \chi_{i+k}(j) a_j^2 \\ &= D^{-1} \sum_{r,s} b_r b_s z_r z_s \sum_j \chi_j(\ell + r + s). \end{aligned}$$

Walsh orthogonality makes the inner sum equal to D exactly when $s = r + \ell$ and zero otherwise. The displayed formula follows. $\square$

Lemma 3.3 (Cosine covariance identity)**.** *For centered jointly Gaussian random variables U, V with common variance t^2 and covariance ρ,*

$$\mathrm{Cov}(\cos U, \cos V) = e^{-t^2}(\cosh \rho - 1).$$

Consequently,

$$\mathrm{Var}(Y(z)) = \frac{(1 - e^{-t^2})^2}{2D} + \frac{e^{-t^2}}{D}\, \mathbb{E}_B \sum_{\ell \neq 0} (\cosh \rho_\ell - 1).$$

Proof. The identity $\cos u \cos v = (\cos(u+v) + \cos(u-v))/2$ and the formula $\mathbb{E} \cos Z = e^{-\mathrm{Var}(Z)/2}$ for centered Gaussian Z give

$$\mathbb{E}[\cos U \cos V] = \frac{1}{2} e^{-(2t^2 + 2\rho)/2} + \frac{1}{2} e^{-(2t^2 - 2\rho)/2} = e^{-t^2} \cosh \rho.$$

Also $\mathbb{E}\cos U = \mathbb{E}\cos V = e^{-t^2/2}$, proving the covariance identity.

Condition on B and apply the identity to the Gaussian vector $R = (Wz)_i$. For each fixed ℓ, exactly D ordered pairs (i,k) satisfy $i + k = \ell$. Therefore

$$\mathrm{Var}_G(Y(z) \mid B) = \frac{e^{-t^2}}{D}\sum_{\ell}(\cosh \rho_\ell - 1).$$

The conditional mean is $e^{-t^2/2}$ by Lemma 3.1, independent of B, so the law of total variance introduces no additional term. The $\ell = 0$ term is

$$\frac{e^{-t^2}}{D}(\cosh t^2 - 1) = \frac{(1 - e^{-t^2})^2}{2D},$$

which gives the formula. □

Lemma 3.4 (Signed autocorrelation moment)**.** *For fixed $z \in \mathbb{R}^D$ and $t = \|z\|_2$,*

$$|\rho_\ell| \le t^2 \quad \textit{for every } \ell,$$

and

$$\mathbb{E}_B \sum_{\ell \neq 0} \rho_\ell^2 = 2(t^4 - \|z\|_4^4) \le 2t^4.$$

Proof. The deterministic bound follows from Cauchy–Schwarz:

$$|\rho_\ell| \le \sum_r |z_r||z_{r+\ell}| \le \left(\sum_r z_r^2\right)^{1/2}\left(\sum_r z_{r+\ell}^2\right)^{1/2} = t^2.$$

For $\ell \neq 0$, expand

$$\rho_\ell^2 = \sum_{r,s} z_r z_{r+\ell} z_s z_{s+\ell} b_r b_{r+\ell} b_s b_{s+\ell}.$$

The expectation over independent signs is nonzero exactly when the unordered pairs $\{r, r+\ell\}$ and $\{s, s+\ell\}$ coincide, which happens for $s = r$ or $s = r + \ell$. Thus

$$\mathbb{E}_B \rho_\ell^2 = 2\sum_r z_r^2 z_{r+\ell}^2.$$

Summing over nonzero ℓ gives

$$\mathbb{E}_B \sum_{\ell \neq 0} \rho_\ell^2 = 2\sum_r z_r^2 \sum_{\ell \neq 0} z_{r+\ell}^2 = 2\sum_r z_r^2(t^2 - z_r^2) = 2(t^4 - \|z\|_4^4).$$

□

Lemma 3.5 (Gaussian damping)**.** *For every $a \ge 0$ and every real u with $|u| \le a$,*

$$e^{-a}(\cosh u - 1) \le \frac{2u^2}{(1+a)^2}.$$

Proof. The case $a = 0$ is immediate. Assume $a > 0$. The identity

$$\cosh s - 1 = s^2 \int_0^1 (1 - r)\cosh(rs)\, dr$$

shows that $(\cosh s - 1)/s^2$ is nondecreasing for $s > 0$. Hence, for $|u| \leq a$,

$$\cosh u - 1 \leq u^2 \frac{\cosh a - 1}{a^2}.$$

Multiplying by e^{-a} and using

$$e^{-a}(\cosh a - 1) = \frac{(1 - e^{-a})^2}{2}$$

gives

$$e^{-a}(\cosh u - 1) \leq \frac{u^2}{2} \left(\frac{1 - e^{-a}}{a} \right)^2.$$

If $0 < a \leq 1$, then $(1 - e^{-a})/a \leq 1 \leq 2/(1 + a)$. If $a \geq 1$, then $(1 - e^{-a})/a \leq 1/a \leq 2/(1 + a)$. This proves the displayed inequality. $\square$

Proposition 3.6 (One-block variance)**.** *For every fixed* $z \in \mathbb{R}^D$*,*

$$\mathbb{E}Y(z) = e^{-\|z\|_2^2/2}, \qquad \mathrm{Var}(Y(z)) \leq \frac{5}{D}.$$

Proof. The expectation follows from Lemma 3.1: each row of W is marginally $N(0, I_D)$, so each summand has expectation $e^{-\|z\|_2^2/2}$.

For the variance, use Lemma 3.3. The diagonal term is at most $1/(2D)$. For the off-diagonal term, Lemma 3.4 gives $|\rho_\ell| \leq t^2$, so Lemma 3.5, with $a = t^2$ and $u = \rho_\ell$, gives

$$\begin{aligned} \frac{e^{-t^2}}{D} \mathbb{E}_B \sum_{\ell \neq 0} (\cosh \rho_\ell - 1) &\leq \frac{2}{D(1 + t^2)^2} \mathbb{E}_B \sum_{\ell \neq 0} \rho_\ell^2 \\ &\leq \frac{4t^4}{D(1 + t^2)^2} \leq \frac{4}{D}. \end{aligned}$$

Thus $\mathrm{Var}(Y(z)) \leq 1/(2D) + 4/D \leq 5/D$. $\square$

4 From pairwise estimation to KDE

For the original bandwidth σ, pad every vector by zeros and set

$$u_x = \sqrt{2}\, x/\sigma \in \mathbb{R}^D.$$

For one block W, define the feature map

$$\phi_W(x) = \frac{1}{\sqrt{D}} \big(\cos(W u_x), \sin(W u_x)\big) \in \mathbb{R}^{2D}.$$

Then

$$\phi_W(x)^T \phi_W(y) = \frac{1}{D} \sum_{i=1}^{D} \cos((W(u_x - u_y))_i) = Y(u_x - u_y).$$

By Proposition 3.6, this is an unbiased estimator of $k_\sigma(x, y)$.

Lemma 4.1 (KDE variance lift)**.** *For one block and a fixed query y, define*

$$\widehat{E}_1(y) = \frac{1}{|X|}\sum_{x\in X} Y(u_x - u_y).$$

Then

$$\mathbb{E}\widehat{E}_1(y) = \frac{1}{|X|}\sum_{x\in X} k_\sigma(x,y), \qquad \operatorname{Var}(\widehat{E}_1(y)) \le \frac{5}{D}.$$

Proof. The expectation follows from

$$e^{-\|u_x-u_y\|_2^2/2} = e^{-\|x-y\|_2^2/\sigma^2} = k_\sigma(x,y).$$

For the variance, write $p_x = 1/|X|$ and

$$Z_x = Y(u_x - u_y) - e^{-\|u_x-u_y\|_2^2/2}.$$

Then

$$\widehat{E}_1(y) - \mathbb{E}\widehat{E}_1(y) = \sum_{x\in X} p_x Z_x.$$

Minkowski's inequality in L_2 gives

$$\left\|\sum_x p_x Z_x\right\|_{L_2} \le \sum_x p_x \|Z_x\|_{L_2} \le \sum_x p_x\sqrt{5/D} = \sqrt{5/D}.$$

Squaring gives the variance bound. No independence among the Z_x is required. □

Proof of Theorem 2.1. Let

$$q = \max\left\{1, \left\lceil \frac{20}{\varepsilon^2 D}\right\rceil\right\}.$$

One constant-success copy of the data structure uses q independent blocks $W_1,\ldots,W_q$, each with fresh mutually independent sign and Gaussian diagonals, and the concatenated feature map

$$\phi(x) = \frac{1}{\sqrt{q}}(\phi_{W_1}(x),\ldots,\phi_{W_q}(x)).$$

The stored vector is the empirical feature mean

$$F_X = \frac{1}{|X|}\sum_{x\in X}\phi(x).$$

On query y, the copy returns $F_X^T\phi(y)$, equivalently the average of the q independent one-block KDE estimates.

By Lemma 4.1 and independence of the blocks,

$$\operatorname{Var}(F_X^T\phi(y)) \le \frac{5}{qD} \le \frac{\varepsilon^2}{4}.$$

Chebyshev's inequality gives failure probability at most $1/4$ for additive error ε. Choose an odd integer $L \ge C\log(2/\delta)$, for a sufficiently large absolute constant C, use L mutually independent constant-success copies, and return the median of their estimates. A standard Chernoff bound shows that the median fails with probability at most $\delta/2$. In particular, its success probability is at least $1-\delta/2 > 1-\delta$, as required by the strict inequality in the source definition.

For one block, computing $Wu_y = \sqrt{D}\,HGHBu_y$ uses two fast Walsh–Hadamard transforms and diagonal multiplications, for time $O(D\log D)$. The trigonometric evaluations and inner-product contribution cost $O(D)$. Since

$$qD \le D + 20\varepsilon^{-2}$$

and $D < 2d$ for $d \ge 1$, one constant-success copy has query time

$$O\big((d+\varepsilon^{-2})\log(D+1)\big).$$

Median amplification multiplies this by $O(\log(2/\delta))$. The concatenated feature mean and the random diagonal entries require $O((d+\varepsilon^{-2})\log(2/\delta))$ stored real numbers. Computing and accumulating these features for every point of X takes

$$O\big(|X|(d+\varepsilon^{-2})\log(D+1)\log(2/\delta)\big)$$

preprocessing time. The construction and proof never use a diameter bound on the data or query region, so no Δ_σ dependence appears. □

Corollary 4.2 (Fixed finite query batches)**.** *Let $y_1,\dots,y_s$ be fixed independently of the preprocessing randomness. Running the data structure with per-query failure probability δ/s answers all s queries correctly with probability at least $1-\delta$, with per-query time multiplied only by the factor $O(\log(2s/\delta))$ in the median step.*

Proof. Apply Theorem 2.1 to each fixed y_j with failure probability δ/s and take a union bound. □

5 Why the statement is pointwise

The proof above uses cancellation in the random signs B for each fixed displacement z before the preprocessing randomness is sampled. It does not imply that a single realized block has small variance uniformly over all possible directions. The following calculation makes the distinction explicit.

Proposition 5.1 (A one-block nonuniformity witness)**.** *Fix a realization of the sign matrix B in one Fastfood block, leaving G random. For any $t>0$, define $z\in\mathbb{R}^D$ by*

$$z_r = t\,b_r/\sqrt{D}.$$

Then $\|z\|_2 = t$, every nonzero signed autocorrelation satisfies $\rho_\ell = t^2$, and

$$Y(z) = \cos(tg_0),$$

where g_0 is the Gaussian diagonal entry indexed by the trivial Walsh character.

Proof. The norm identity is immediate. For $\ell \ne 0$,

$$\rho_\ell = \sum_r b_r b_{r+\ell}\left(tb_r/\sqrt{D}\right)\left(tb_{r+\ell}/\sqrt{D}\right) = t^2.$$

Also $Bz = (t/\sqrt{D})\mathbf{1}$, so $HBz = te_0$, where e_0 is the coordinate of the trivial Walsh character. Therefore

$$Wz = \sqrt{D}\,HG(te_0) = \sqrt{D}\,tg_0He_0 = tg_0\mathbf{1},$$

and every cosine summand equals $\cos(tg_0)$. In particular,

$$\operatorname{Var}_G(Y(z)) = \frac{1}{2}\left(1 - e^{-t^2}\right)^2,$$

which is bounded away from zero for fixed $t > 0$ independently of D. $\square$

Thus the theorem should not be read as a uniform guarantee over a continuum of queries after the randomness is fixed, nor as an adaptive guarantee for queries chosen after inspecting the random structure or as a function of earlier answers. This matches the pointwise quantifiers in [1, Def. 1.1]. Proposition 5.1 concerns one realized block only. It is not a lower bound for the average of independent blocks or for the median-amplified data structure in Theorem 2.1; it only shows that the signed autocorrelation cancellation cannot be upgraded to a uniform one-block claim.

References

[1] Tal Wagner. New bounds for kernel sums via fast spherical embeddings. To appear at the International Conference on Machine Learning (ICML), 2026. arXiv:2605.01263v1 [cs.DS], 2 May 2026.

[2] Arturs Backurs, Zinan Lin, Sepideh Mahabadi, Sandeep Silwal, and Jakub Tarnawski. Efficiently computing similarities to private datasets. In *The Twelfth International Conference on Learning Representations*, 2024. OpenReview:HMe5CJv9dQ.

[3] Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Advances in Neural Information Processing Systems*, volume 20, pages 1177–1184, 2007. NeurIPS proceedings record.

[4] Quoc V. Le, Tamás Sarlós, and Alexander J. Smola. Fastfood—computing Hilbert space expansions in loglinear time. In *Proceedings of the 30th International Conference on Machine Learning*, PMLR 28(3):244–252, 2013. PMLR:v28/le13.

[5] Chang Feng, Qinghua Hu, and Shizhong Liao. Random feature mapping with signed circulant matrix projection. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 3490–3496, 2015. IJCAI proceedings record.

[6] Krzysztof Choromanski and Vikas Sindhwani. Recycling randomness with structure for sublinear time kernel expansions. In *Proceedings of the 33rd International Conference on Machine Learning*, PMLR 48:2502–2510, 2016. PMLR:v48/choromanski16.