

# Ambient Gaussian Landmarks and a Unit-Circle Obstruction to Nyström Approximation

Alper Ferudun\*

24 October 2026

## Abstract

We consider the cross-kernel Nyström approximation for  $k(x, y) = 1 + x^\top y$  when the landmarks are independent standard Gaussian points in the ambient space  $\mathbb{R}^n$ . For a fixed unit query, we derive an exact chi-square representation for the intercept leverage and a finite-sample lower bound on the diagonal approximation error. If  $m = o(n)$  landmarks are used, this error converges in probability to 2, the full diagonal kernel value. The data may lie on a compact smooth algebraic unit circle, where the restricted kernel has rank three and three suitably chosen on-circle landmarks give exact reconstruction. More generally, if  $m/n \rightarrow \alpha \in [0, 1]$ , the ambient-Gaussian approximant converges uniformly on that circle to  $\alpha k$ . The sublinear regime contradicts the global-constant polynomial-kernel tail bound proposed in Oberwolfach Report 44/2014. It does not rule out modified sampling rules, other low-rank approximations, or the report's separate Gaussian-kernel question.

## 1 The sampling question and the scope of the result

Király's contribution to Oberwolfach Report 44/2014 [1, pp. 2499–2501] discusses the cross-kernel approximation

$$\tilde{k}_Z(x, y) = K(x, Z)K(Z, Z)^{-1}K(Z, y), \quad K(x, Z) = (k(x, z_j))_{j=1}^m. \quad (1)$$

For the inhomogeneous polynomial kernel  $k(x, y) = (1 + x^\top y)^d$ , the displayed conjecture on p. 2501 asks for global positive constants  $C, c$  such that

$$\mathbb{P}\{|k(x, y) - \tilde{k}_Z(x, y)| \geq \varepsilon\} \leq C\varepsilon^{-c\lambda}, \quad m \geq \lambda Dd \log n, \quad \lambda \geq 1, \quad (2)$$

when  $x, y$  are fixed points of an algebraic manifold of dimension  $D$  in  $\mathbb{R}^n$ , and the  $z_j$  have independent standard normal ambient coordinates. We use the natural logarithm; changing its base only changes a constant factor. No restriction to on-manifold landmarks or to  $\varepsilon < 1$  is stated in the displayed proposal. The report also invites a suitably modified variant. We refute (2) with its stated global constants, not that open-ended request for alternative theorems.

Only  $d = 1$  is needed. For  $n \geq 2$ , let

$$\mathcal{C}_n = \{x \in \mathbb{R}^n : x_1^2 + x_2^2 = 1, x_3 = \cdots = x_n = 0\}. \quad (3)$$

---

\*Mercury Software GmbH. Email: [alper@mercurycodelab.com](mailto:alper@mercurycodelab.com). GitHub: <https://github.com/AlperTheKing>. Unrefereed preprint, version 1.0. Licensed under CC BY 4.0.

This is a compact smooth real algebraic manifold of dimension one and degree two. Take  $x = y = e_1$ , fixed before the landmarks are drawn. Thus neither growing query norms nor points depending on the sample are involved.

The projection interpretation and exact feature-span criterion for cross-kernel reconstruction are classical; see Király, Kreuzer and Theran [2, Theorem 3.4]. Altschuler and Parrilo [3] establish improved kernel approximation ranks on algebraic varieties, using constructions different from the ambient-Gaussian rule in (1). Those results are compatible with the obstruction below. Our proofs are elementary and self-contained. We make no absolute novelty or priority claim. AI assistance was used in developing the argument, checking finite identities and preparing this manuscript; the result is not formally verified or independently refereed.

## 2 Projection and intercept leverage

Let  $Z \in \mathbb{R}^{m \times n}$  have independent  $N(0, 1)$  entries. Unless stated otherwise,  $1 \leq m \leq n$ . Write  $\mathbf{1}$  for the all-ones vector of length  $m$ , and set

$$A = [\mathbf{1} \ Z], \quad W = ZZ^\top, \quad P = A^\top(AA^\top)^{-1}A, \quad \phi(x) = (1, x^\top)^\top.$$

The Gaussian matrix  $Z$  has row rank  $m$  almost surely. Consequently  $W$  and  $AA^\top = W + \mathbf{1}\mathbf{1}^\top$  are positive definite almost surely. All inverses used below therefore exist. The matrix  $P$  is an orthogonal projector of rank  $m$ . The feature identity  $k(x, y) = \phi(x)^\top \phi(y)$  gives

$$\tilde{k}_Z(x, y) = \phi(x)^\top P \phi(y), \quad R_Z(x) := k(x, x) - \tilde{k}_Z(x, x) = \phi(x)^\top (I - P) \phi(x) \geq 0. \quad (4)$$

Index the intercept coordinate by 0, and write

$$P = \begin{pmatrix} p & b^\top \\ b & H \end{pmatrix}.$$

**Lemma 1.** *The intercept leverage satisfies*

$$p \stackrel{d}{=} \frac{m}{m + X}, \quad X \sim \chi_{n-m+1}^2. \quad (5)$$

If  $n > m + 1$ , then

$$\mathbb{E}p \leq \frac{m}{n - m - 1}. \quad (6)$$

*Proof.* Put  $T = \mathbf{1}^\top W^{-1} \mathbf{1}$ . The Sherman–Morrison identity gives  $p = T/(1 + T)$ . Choose a deterministic orthogonal row rotation taking  $\mathbf{1}$  to  $\sqrt{m} e_1$ . The rotated  $Z$  still has independent standard normal entries. Write its first row as  $g$  and its remaining  $m - 1$  rows as  $B$ . For  $m > 1$ ,  $B$  has row rank  $m - 1$  almost surely. The Schur-complement formula gives

$$((ZZ^\top)^{-1})_{11} = (g(I - B^\top(BB^\top)^{-1}B)g^\top)^{-1}$$

in the rotated coordinates. The matrix in parentheses between  $g$  and  $g^\top$  is the orthogonal projector onto  $\ker B$ , of dimension  $n - m + 1$ . Conditional on  $B$ , the row  $g$  is an independent standard Gaussian vector. The denominator therefore has distribution  $\chi_{n-m+1}^2$ . For  $m = 1$ , take  $B$  empty and the projector equal to  $I_n$ . This proves  $T \stackrel{d}{=} m/X$ , and hence (5).

For  $\nu > 2$ , integration against the chi-square density gives

$$\mathbb{E}[(\chi_\nu^2)^{-1}] = \frac{2^{\nu/2-1} \Gamma(\nu/2 - 1)}{2^{\nu/2} \Gamma(\nu/2)} = \frac{1}{\nu - 2}.$$

Since  $m/(m + X) \leq m/X$ , (6) follows. □

**Lemma 2.** For every deterministic unit vector  $u \in \mathbb{R}^n$ ,

$$\mathbb{E}[\phi(u)^\top P \phi(u)] = \frac{m}{n} + \left(1 - \frac{1}{n}\right) \mathbb{E}p. \quad (7)$$

*Proof.* For every orthogonal  $Q \in \mathbb{R}^{n \times n}$ ,  $ZQ$  has the same distribution as  $Z$ . Thus  $P$  is invariant in distribution under conjugation by  $\text{diag}(1, Q)$ . In particular,  $\mathbb{E}b = 0$  and  $\mathbb{E}H$  is a scalar multiple of  $I_n$ . Since  $\text{tr } P = m$ , its value is  $\mathbb{E}H = (m - \mathbb{E}p)I_n/n$ . Expanding  $\phi(u)^\top P \phi(u) = p + 2b^\top u + u^\top H u$  proves the identity.  $\square$

### 3 A finite bound and the counterexample

**Theorem 3.** If  $n > m + 1$ ,  $u$  is a deterministic unit vector, and  $0 < \varepsilon < 2$ , then

$$\mathbb{P}\{R_Z(u) \geq \varepsilon\} \geq 1 - \frac{1}{2 - \varepsilon} \left( \frac{m}{n} + \frac{m}{n - m - 1} \right). \quad (8)$$

In particular, if  $m = m_n = o(n)$ , then  $R_Z(u) \rightarrow 2$  in probability, even for  $u = e_1 \in \mathcal{C}_n$ .

*Proof.* Let  $S = \phi(u)^\top P \phi(u) \geq 0$ . Equations (6) and (7) imply

$$\mathbb{E}S \leq \frac{m}{n} + \frac{m}{n - m - 1}.$$

Because  $R_Z(u) = 2 - S$ , Markov's inequality applied to  $S$  proves (8). When  $m = o(n)$ ,  $\mathbb{E}S \rightarrow 0$ , so  $S \rightarrow 0$  in probability and  $R_Z(u) \rightarrow 2$ .  $\square$

**Corollary 4.** No positive constants  $C, c$ , uniform in the stated parameters, can satisfy (2) for ambient standard Gaussian landmarks. This already fails for the circles  $\mathcal{C}_n$ , with  $D = d = 1$ ,  $x = y = e_1$ , and  $\varepsilon = 3/2$ .

*Proof.* One explicit sequence is

$$n = j^2, \quad m = j, \quad \lambda = \frac{j}{2 \log j}, \quad j \geq 5.$$

Then  $\lambda \geq 1$ ,  $m = \lambda \log n$ , and Theorem 3 gives

$$\mathbb{P}\{R_Z(e_1) \geq 3/2\} \geq 1 - \frac{2}{j} - \frac{2j}{j^2 - j - 1} \rightarrow 1.$$

The proposed upper bound  $C(3/2)^{-cj/(2 \log j)}$  tends to zero, a contradiction.

The obstruction does not rely on a moving  $\lambda$  in an asymptotic assertion. Given any proposed  $C, c > 0$ , choose a fixed  $\lambda \geq 1$  so large that  $C(3/2)^{-c\lambda} < 1/2$ . Then set  $m = \lceil \lambda \log n \rceil$ . For all sufficiently large  $n$ , the inverse exists almost surely and  $n > m + 1$ . The lower probability still tends to one, contradicting the same fixed upper bound.  $\square$

*Remark 5.* The threshold  $3/2$  is below the unit-query diagonal value 2. If one restricts (2) to  $0 < \varepsilon < 1$ , its right side increases with  $\lambda$ , so it is no longer a useful concentration bound of the stated kind. Independently of that notation issue, Theorem 3 shows that any proposed vanishing error guarantee for this sampling rule must account for ambient dimension. The relative diagonal error  $R_Z(e_1)/k(e_1, e_1)$  tends to one.

## 4 The aspect-ratio limit

The failure is not confined to sublinear landmark counts.

**Theorem 6.** *Suppose  $n \rightarrow \infty$ ,  $1 \leq m_n \leq n$ , and  $m_n/n \rightarrow \alpha \in [0, 1]$ . Then*

$$\sup_{x,y \in \mathcal{C}_n} |\tilde{k}_Z(x,y) - \alpha k(x,y)| \rightarrow 0 \quad \text{in probability.} \quad (9)$$

*Proof.* Let  $\nu = n - m + 1$ . For  $X \sim \chi_\nu^2$ ,  $\mathbb{E}X = \nu$  and  $\text{Var } X = 2\nu$ . Consequently  $X/n - \nu/n \rightarrow 0$  in probability. Lemma 1 yields

$$p = \frac{m/n}{m/n + X/n} \rightarrow \alpha \quad \text{in distribution and hence in probability,}$$

where the displayed equality is interpreted in distribution. The denominator converges to one, including at  $\alpha = 0$  and  $\alpha = 1$ .

For completeness, rotational invariance also controls the spatial block  $H$ . Put  $a = \text{tr } H/n = (m - p)/n$ . For a uniform vector  $v$  on the unit sphere in  $\mathbb{R}^n$ ,

$$\mathbb{E}v_i^4 = \frac{3}{n(n+2)}, \quad \mathbb{E}v_i^2 v_j^2 = \frac{1}{n(n+2)} \quad (i \neq j).$$

These identities follow from rotational symmetry, which gives  $\mathbb{E}v_1^4 = 3\mathbb{E}v_1^2 v_2^2$ , and  $\mathbb{E}(\sum_i v_i^2)^2 = 1$ . Conditional on any fixed symmetric matrix  $H$ , expansion in an eigenbasis therefore gives

$$\mathbb{E}_v \left( v^\top H v - \frac{\text{tr } H}{n} \right)^2 = \frac{2}{n(n+2)} \left( \text{tr } H^2 - \frac{(\text{tr } H)^2}{n} \right).$$

Introduce an independent uniform orthogonal rotation and use the invariance of  $(p, H)$ . For every deterministic unit  $u$ , this proves

$$\mathbb{E}(u^\top H u - a)^2 \leq \frac{2m}{n(n+2)}. \quad (10)$$

Indeed  $0 \preceq H \preceq I_n$  and  $\text{tr } H^2 \leq \text{tr } H = m - p \leq m$ .

The identity  $P^2 = P$  gives  $\|b\|^2 = p(1 - p) \leq 1/4$ . Isotropy of  $b$  then implies

$$\mathbb{E}(b^\top u)^2 = \frac{\mathbb{E}[p(1 - p)]}{n} \leq \frac{1}{4n}. \quad (11)$$

Apply (10) to  $e_1, e_2, (e_1 + e_2)/\sqrt{2}, (e_1 - e_2)/\sqrt{2}$ . Since  $a \rightarrow \alpha$ , the upper-left  $2 \times 2$  spatial compression tends entrywise in probability to  $\alpha I_2$ . Equation (11) controls its intercept cross terms. Together with  $p \rightarrow \alpha$ , the  $3 \times 3$  compression of  $P$  to the intercept and first two spatial coordinates tends in operator norm to  $\alpha I_3$ . For every  $x \in \mathcal{C}_n$ , the vector  $\phi(x)$  lies in this three-dimensional space and has norm  $\sqrt{2}$ . The supremum in (9) is therefore at most twice that operator-norm difference, proving the claim.  $\square$

## 5 A rank-three control and what remains open

**Proposition 7.** *For the same kernel restricted to  $\mathcal{C}_n$ , the on-circle landmarks  $e_1, -e_1, e_2$  give exact reconstruction for every pair  $x, y \in \mathcal{C}_n$ . The restricted kernel has rank exactly three.*

*Proof.* The feature vectors of the three landmarks, restricted to the intercept and first two coordinates, are the rows of

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

whose determinant is  $-2$ . Their span thus contains every  $\phi(x)$  for  $x \in \mathcal{C}_n$ , so the feature projector is the identity on all query features. This proves exact reconstruction. The Gram matrix of these three features is nonsingular, while all circle features lie in a three-dimensional space, which proves the rank assertion.  $\square$

This proposition is an elementary instance of the known feature-span criterion [2, Theorem 3.4], not a new general Nyström theorem. It separates representational rank from the effectiveness of an ambient landmark distribution. Our counterexample uses a circle in a two-dimensional linear subspace; an added requirement to preprocess the data’s affine span would change the algorithm being analyzed.

The report’s invitation to find modified polynomial-kernel guarantees is not exhausted by a counterexample to its displayed bound. Nor do we analyze nonlinear polynomial degrees systematically, alternative reweighting, regularized estimators, or the separate Gaussian-kernel question. The precise completed claim is the failure of the stated global-constant ambient-Gaussian guarantee, supplemented by the finite bound and aspect-ratio law above.

## References

- [1] F. J. Király, Algebraic Combinatorial Methods in Statistics and Machine Learning, in *Mini-Workshop: Asymptotic Statistics on Stratified Spaces*, Oberwolfach Report 44/2014, pp. 2499–2501, [doi:10.4171/OWR/2014/44](https://doi.org/10.4171/OWR/2014/44).
- [2] F. J. Király, M. Kreuzer and L. Theran, *Learning with Cross-Kernels and Ideal PCA*, 2014, [arXiv:1406.2646](https://arxiv.org/abs/1406.2646).
- [3] J. M. Altschuler and P. A. Parrilo, Kernel approximation on algebraic varieties, *SIAM Journal on Applied Algebra and Geometry* **7** (2023), 1–28, [doi:10.1137/21M1425050](https://doi.org/10.1137/21M1425050). Preprint: [arXiv:2106.02755](https://arxiv.org/abs/2106.02755).