

A General Kernel Framework for Non-CND Distance Measures Using $|\mathcal{D}|$ -Dimensional Sparse Landmark Embeddings

Distance-based kernels are **positive semi-definite (PSD)** only when the squared distance is **conditionally negative definite (CND)**, a condition that geodesic, Wasserstein, edit, and shortest-path distances violate. The **Sparse Landmark Embedding (SLE)** kernel is PSD for **any** distance measure, uses all $|\mathcal{D}|$ training points as landmarks, and keeps Gram matrices sparse and well conditioned.

Marcus M. Noack · Maher B. Alghalayini · Mark D. Risser · Lawrence Berkeley National Laboratory

THE SLE KERNEL IN ONE LINE

Encode the geometry in the **embedding**; inherit positive semi-definiteness from the **kernel**.

$$k_{\text{SLE}}(x, x') = h(\phi(x), \phi(x')), \quad \phi(x) = [b(d(x, x_i); a, r, \beta)]_{i=1}^{|\mathcal{D}|}$$

Each input x becomes a vector of compactly supported bumps b , one centered at every training point x_i . Any standard Euclidean kernel h (RBF, Matérn) applied to that vector is PSD for every distance d .



0
CONDITIONS REQUIRED OF THE DISTANCE

all $|\mathcal{D}|$ points
SERVE AS LANDMARKS · NONE SELECTED

$\approx 9.8\times$ lower κ_2
VS. DENSE EMBEDDING AT $N = 1000$

FROM ANY DISTANCE TO A VALID GAUSSIAN PROCESS

The Obstacle

NON-CND DISTANCES

Natural distances break standard kernels

By Schoenberg's theorem, $\exp(-d^2/\sigma^2)$ is PSD for every length scale $\sigma > 0$ only if d is **Hilbertian**: the input space $(\mathcal{X}, d)$ embeds isometrically in a Hilbert space. Geodesic distances, W_2 (dimension > 1), ℓ_1 (dimension ≥ 2), edit and shortest-path distances are not; Gram matrices can turn indefinite.



Workarounds alter the problem: heat kernels need Laplace–Beltrami eigenpairs, sliced Wasserstein swaps W_2 for a surrogate, Nyström methods must select landmarks.

GEODESIC · WASSERSTEIN · EDIT · GRAPH

The Insight

PSD BY COMPOSITION

Any embedding, any PSD kernel

Let $\phi: \mathcal{X} \rightarrow \mathbb{R}^m$ be any map and h any PSD kernel on $\mathbb{R}^m$. For inputs $x_1, \dots, x_N$ and weights $c \in \mathbb{R}^N$,

$$\sum_{i,j=1}^N c_i c_j h(\phi(x_i), \phi(x_j)) \geq 0,$$

since the $\phi(x_i)$ are ordinary points in $\mathbb{R}^m$. **No property of d enters.** The distance may be asymmetric, violate the triangle inequality, or carry noise (Remark 1).

Raw landmark distances $\phi(x) = [d(x, \ell_j)]_{j=1}^m$ with landmarks ℓ_j couple two problems: few landmarks miss geometry, and many landmarks concentrate distances and ill-condition the Gram matrix K .

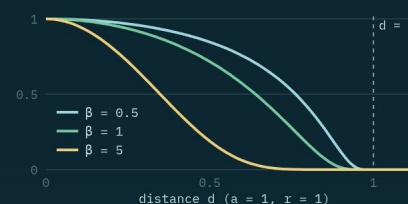
THEOREM 1 · VALID FOR EVERY D

The Embedding

COMPACT SUPPORT · ALL TRAINING POINTS

$$b(d; a, r, \beta) = \begin{cases} a \exp(-\frac{\beta}{1-d^2/r^2} + \beta), & d < r \\ 0, & d \geq r \end{cases}$$

a : amplitude · r : support radius · β : shape. C^∞ , strictly decreasing on $[0, r)$.



Only landmarks within r of x activate, so the expected nonzero count $\mathbb{E} \|\phi(x)\|_0 = \sum_i \mathbb{P}[d(x, x_i) < r]$ depends on r , not on $|\mathcal{D}|$ (Thms 2, 5).

SPARSE · SMOOTH · NO LANDMARK SELECTION

The Guarantees

THEOREMS 1–9 · PROPOSITIONS 1–2

- PSD** For every d ; spatially varying $a(x)$, $r(x)$, $\beta(x)$ keep PSD (non-stationary SLE). Thm 1 · Prop 2
- FIDELITY** $\phi(x) = \phi(x')$ iff the distances to all landmarks within reach agree; no surrogate metric. Prop 1
- SPARSITY** $m p_r \ll m$ expected nonzeros for m landmarks, with $p_r = \mathbb{P}[d(x, x_i) < r]$. Thms 2, 5
- CONDITIONING** Under bounded overlap, $\kappa(K)$ does not grow with $|\mathcal{D}|$; distance concentration is suppressed. Thms 4, 6
- EXPRESSIVITY** Universal on compact $\mathcal{X}$ for injective ϕ ; smoothness inherited from h ; RBF recovered as $r \rightarrow \infty$ for CND d . Thms 3, 7, 8
- COST** $\mathcal{O}(s)$ per kernel evaluation, s = overlapping nonzeros. Thm 9

PROVABLY PSD · WELL CONDITIONED

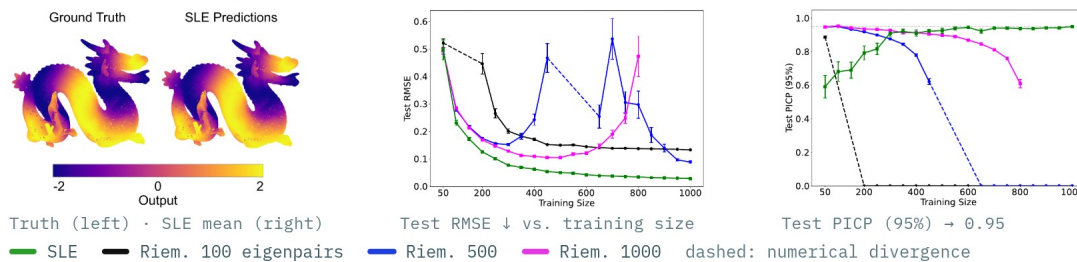
Three benchmarks against published baselines — **calibrated uncertainty on the native geometry**

HEADLINE RESULT R1

Dragon manifold vs. Riemannian Matérn kernel

Borovitskiy et al., NeurIPS 2020 · 100,179 vertices · Matérn $\nu = 3/2$ · 30 datasets per size

Target: sine of the geodesic distance from the snout; 5,000 test points. The baseline truncates a Laplace–Beltrami expansion at l eigenpairs; as the training size n nears l , K turns rank-deficient and coverage collapses. **SLE has no truncation parameter.**



4.0 $\times$ lower RMSE
0.034 vs 0.137 · $n = 800$

6.0 $\times$ lower CRPS
0.015 vs 0.090 · $n = 800$

0.937 vs 0.609
PICP (95%) · $n = 800$

Metric	Model	$n = 400$	$n = 600$	$n = 800$
RMSE ↓	Riem. 100	0.171 $\pm$ 0.006	0.144 $\pm$ 0.001	0.137 $\pm$ 0.001
	Riem. 500	0.242 $\pm$ 0.020	4.257 $\pm$ 0.580	0.297 $\pm$ 0.049
	Riem. 1000	0.109 $\pm$ 0.002	0.120 $\pm$ 0.006	0.473 $\pm$ 0.073
CRPS ↓	SLE	0.062 $\pm$ 0.002	0.042 $\pm$ 0.001	0.034 $\pm$ 0.001
	Riem. 100	0.111 $\pm$ 0.001	0.101 $\pm$ 0.000	0.098 $\pm$ 0.000
	Riem. 500	0.100 $\pm$ 0.018	1.057 $\pm$ 0.147	0.090 $\pm$ 0.006
PICP $\rightarrow$.95	SLE	0.030 $\pm$ 0.001	0.020 $\pm$ 0.001	0.015 $\pm$ 0.000
	Riem. 100	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
	Riem. 500	0.780 $\pm$ 0.007	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
	Riem. 1000	0.912 $\pm$ 0.004	0.869 $\pm$ 0.007	0.609 $\pm$ 0.025
	SLE	0.911 $\pm$ 0.018	0.944 $\pm$ 0.007	0.937 $\pm$ 0.005

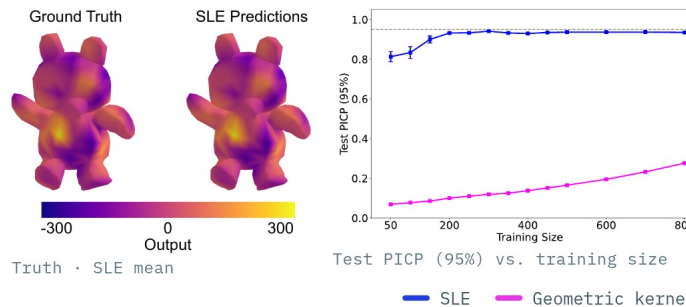
Mean $\pm$ s.e. over 30 datasets (Table 1) · green: closest to target · KPIs vs. best baseline at $n = 800$

R2

Teddy Bear vs. Geometric kernel

Mostowsky et al., JMLR 2025 · 1,598 vertices · Matérn $\nu = 5/2$ · reference code, unmodified

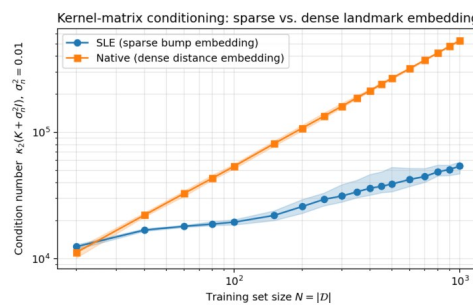
Target: a prior sample from the baseline publication. The baseline is numerically stable, and RMSE agrees within 2.6% at 200–400 points. Its 95% intervals cover only **8–27%** of test values. At $n = 200$ –400, SLE reaches PICP **0.929–0.941** with **19–24% lower CRPS**.



R4

Sparse embeddings stay well conditioned

1D benchmark · identical hyperparameters and nugget $\sigma^2_n = 0.01$ for both embeddings



$\kappa_2(K + \sigma^2_n I)$ vs. training size $N = |\mathcal{D}|$

The 2-norm condition number κ_2 of the regularized Gram matrix grows as $N^{0.38}$ for **SLE** and as $N^{\approx 1}$ for the dense raw-distance embedding, $\approx 9.8\times$ apart at $N = 1000$. Bounded overlap is not enforced in this test.

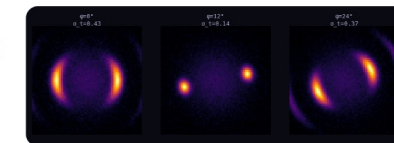
Ablation (same inner kernel and optimizer): the bump embedding gives better-calibrated intervals on the Dragon up to ≈ 650 points and lower RMSE on the Teddy Bear beyond ≈ 400 points.

R3

X-ray scattering images as distributions

vs. Matérn on sliced W_2 (50 projections) · 500 synthetic SAXS images · $\nu = 3/2$

64×64 SAXS patterns of lamellar films; target: effective elastic modulus. SLE on the **exact** W_2 matches or slightly improves the sliced baseline. SLE on the sliced distance is less accurate at every size, so the gain stems from the exact distance.



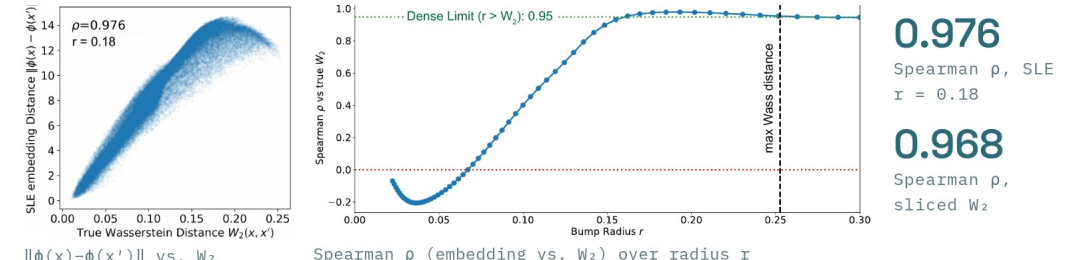
Metric	Model	$n = 150$	$n = 200$	$n = 250$
RMSE ↓	Matérn-sliced	0.489 $\pm$ 0.010	0.445 $\pm$ 0.011	0.358 $\pm$ 0.009
	SLE-W_2	0.467 $\pm$ 0.010	0.396 $\pm$ 0.009	0.334 $\pm$ 0.007
CRPS ↓	Matérn-sliced	0.221 $\pm$ 0.004	0.188 $\pm$ 0.004	0.148 $\pm$ 0.005
	SLE-W_2	0.210 $\pm$ 0.004	0.174 $\pm$ 0.003	0.146 $\pm$ 0.003

PICP (95%): 0.934–0.959 (SLE- W_2) and 0.940–0.963 (Matérn-sliced). 30 draws, Table 3.

R5

No surrogate metric within the bump radius

SAXS distance matrices · all 124,750 image pairs · embedding map only, no fitted model



0.976
Spearman ρ , SLE
 $r = 0.18$

0.968
Spearman ρ , sliced W_2

TAKE-HOME

Positive semi-definiteness comes from the Euclidean kernel h ; the geometry comes from the bump embedding ϕ . Any distance d , CND or not, yields a valid GP kernel with no projection, slicing, or landmark selection, and with sparse, well-conditioned Gram matrices.

LIMITATIONS

Distance profiles cost $\mathcal{O}(N \cdot |\mathcal{D}|)$ distance evaluations for N evaluation points; metric trees reduce this to range queries. Accuracy depends on the radius r , and information beyond r is discarded, so globally governed problems need larger radii.

CODE & DATA · doi.org/10.5281/zenodo.22884130

Implemented in gpCAM. Supported by CAMERA (DOE SC ASCR & BES), the DOE SC ASCR Applied Mathematics and Competitive Portfolios programs, and the DOE SC Early Career Research Program, Contract No. DE-AC02-05CH11231.

