Alex Kokot
PhD Candidate, Department of Statistics

Research Statement
University of Washington

In my research, I incorporate **geometric principles** into the analysis of functionals arising in statistics and machine learning. Key research themes include:

1. Making ML algorithms more **computationally** and **statistically efficient**.
2. Leveraging **geometric structure** in data.
3. Re-imagining classical methods in **modern settings**.

To tackle these problems, I use tools from **functional analysis**, **entropic optimal transport**, **numerical linear algebra**, and **empirical processes**.

Local EGOP Learning. It is commonly speculated that machine learning algorithms more effectively leverage **structured data** than their classical counterparts. I developed a geometric model for such structure via the *supervised noisy manifold hypothesis*, where covariates are concentrated about a low-dimensional manifold, and labels do not depend on orthogonal deviations. This is an instance of our newly introduced setting of *continuous-index learning* (Kokot et al., 2026), generalizing multi-index learning. I study this problem via **kernel smoothing** in an adaptive Mahalanobis metric. The objective is to induce anisotropy, reducing estimator variance by elongating along the normal space and pooling additional low bias data points.

My approach is motivated by works such as Yuan et al. (2025), Radhakrishnan et al. (2022), and Takeda et al. (2007) which suggest metrization by the **expected gradient outerproduct (EGOP)**. Function gradients are tangential to the manifold, thus local EGOPs are effectively low rank. I developed an iterative procedure, Local EGOP Learning, which induces the desired metric degeneration, yielding intrinsic dimensional learning rates. We analyze this by developing a novel framework to assess the utility of kernel anisotropy, relating it to the EGOP matrix via a **Poincaré inequality**. Our iteration scheme can then be formulated as an appropriate autonomous recurrence in the covariance of Gaussian **test distributions** we call *localizations*. Experimentally, we demonstrate that **deep neural networks** achieve a similar metrization, while two-layer neural networks cannot, even in overparameterized regimes. Thus, this local EGOP ansatz presents a promising lens to understand the efficiency gains present in modern machine learning algorithms.

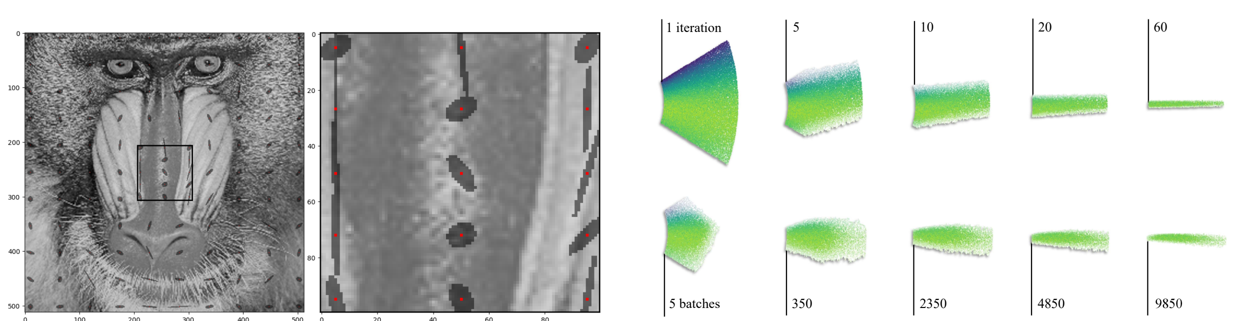


Figure 1: Illustration of results on **Local EGOP Learning**. (Left) Local EGOP metrics adapted to grayscale pixel data. (Right) Localization regions produced by Local EGOP Learning (top) and a deep transformer architecture (bottom) after progressively many iterations/training batches.

Coreset Selection. In Kokot and Luedtke (2025), I took on the task of **distributional compression**. Given n independent draws from a compactly supported $\mathbb{P}$ on $\mathcal{X} \subseteq \mathbb{R}^d$ with empirical distribution $\mathbb{P}_n$, we seek to construct P_m supported on a 'coreset' of m of these observations such that, relative to a specified **divergence** D, the coreset's deviation from the empirical is of the same order as the empirical's deviation from the truth, in that $D(P_m, \mathbb{P}_n) = O(D(\mathbb{P}_n, \mathbb{P}))$ as $m, n \to \infty$. This presents a substantial generalization of typical clustering and coreset selection settings, as the loss D is arbitrary as opposed to classical objectives such as squared residuals. The **Sinkhorn divergence** (Feydy et al., 2019), a type of smoothed K-means, was our leading example.

I developed a general criteria for when D-compression can be asymptotically relaxed to a problem specific **MMD**, culminating in the **Coresets of Order 2** (CO2) algorithm. This reduces the problem to minimizing a **quadratic form** under affine and sparsity constraints, for which we employ **Nyström** approximations (Tropp et al., 2017) in combination with the recombination algorithm (Hayakawa et al., 2023). By verifying **Hadamard differentiability** of the **entropic potentials** in the Gaussian RKHS, we were able to show that only poly-log(n) samples are required to approximate the population distribution up to negligible error.

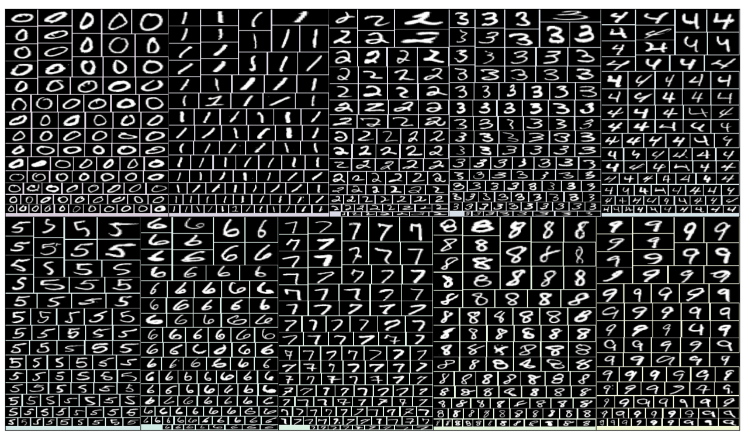
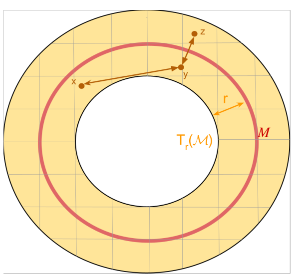


Figure 2: Illustration of results on **Coreset Selection**. A treemap of 1000 digits selected from the MNIST dataset via my proposed approach, Sinkhorn CO2. Image areas proportional to coreset weights, and dimensions scaled to fit a uniform grid.

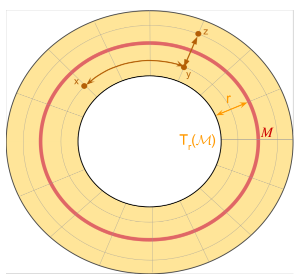
Entropic Optimal Transport (EOT). In my analysis of the Sinkhorn divergence, I verified Hadamard differentiability via an inverse function theorem applied to the **Schrödinger equations**, refining the approaches of Gonzalez-Sanz et al. (2022) and Goldfeld et al. (2022). In current research we are building on this methodology, deriving limits for self-EOT, entropic transport from a distribution to itself, as the regularization $\varepsilon \to 0$. By reparameterizing this equation, I derived explicit first order approximations of an appropriate analogue to the entropic potentials, providing new and rigorous insight to claims as presented in works such as Mordant (2024). The key implication of this result is a refinement of Cheng and Landa (2024), as we show an asymptotic correspondence between diffusion maps and the entropic self-transport coupling, relating our coreset selection algorithm for the Sinkhorn divergence to **spectral clustering**. I plan to submit this research to SIMODS in early February.

Targeted Sampling. Beyond entropic optimal transport, my analysis of the Sinkhorn divergence yields a generic framework for the selection of coresets to optimize arbitrary functionals, and this includes key settings such as **empirical risk minimization**. In ongoing research, these methods are applied to select data to accelerate model fitting. This research has many applications, from identifying corpuses for **fine-tuning**, to **batch selection** in SGD. The resulting methodology leverages recent insights from the setting of "kernel thinning" (Dwivedi and Mackey, 2024), relating these problems to data specific kernels derived via functional Taylor expansions that are **learned actively** during the training process.

Spectral Embeddings. In Kokot et al. (2025), I analyzed **Laplacian spectral embeddings** for manifold data injected with high-dimensional noise. The traditional folklore is that low-dimensional spectral embeddings are insensitive to such contamination. To verify this rigorously, I developed a metric perturbation argument, comparing the induced tube geometry to the Sasaki metric (Sasaki, 1958; Nobis and Wittich, 2019). The Sasaki metric naturally splits the Neumann Laplacian into an intrinsic component and a high frequency perturbation. By leveraging **perturbation theory** for **unbounded operators**, I showed that in the continuum, low frequency eigenfunctions are nearly invariant to deviations away from the manifold. Rather than a stability result, we take the dual point of view: Laplacian spectral embeddings **detect fundamental data structure**, with a dependence that is **not strictly dimensional**.



(a) Induced Metric



(b) Sasaki Metric

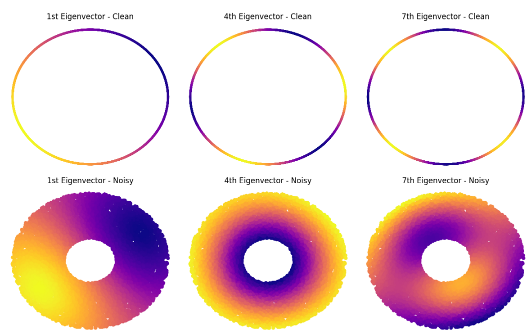


Figure 3: Illustration of results on **Spectral Embeddings**. (Left) The Sasaki and induced metrics, with geodesics depicted between indicated points. (Right) Laplacian eigenvectors on the circle and annulus, demonstrating the loss of manifold information for larger eigenvalues.

Future Work: As a postdoctoral researcher, I will continue my work in the following settings:

Kernel Adaptation: My local EGOP method achieves intrinsic learning rates by anisotropically aligning kernels to affine level-sets. In current work, this is generalized to a continuous diffusion process capable of stretching along smooth contours. This capacity to trace complex manifolds offers new insights for understanding the feature learning capabilities of deep neural networks.

High-Efficiency Coreset Selection: To optimize the training of modern machine learning algorithms, I will advance the efficiency theory for coreset selection targeting asymptotically linear functionals. A key component is the refinement of numerical methods for MMD kernel thinning, a sparse, quadratic optimization problem with affine constraints.

Optimal Transport Methodology: I will develop limiting results for entropic OT to drive concretely motivated statistical methodology. These analyses yield new insights in problems ranging from density estimation to spectral clustering and K-means seeding.

Beyond these core themes, I am eager to expand my research through new collaborations, particularly in **optimization, generative modeling, inverse problems,** and **scientific machine learning**.

References

X. Cheng and B. Landa (2024). "Bi-stochastically normalized graph Laplacian: convergence to manifold Laplacian and robustness to outlier noise". *Information and Inference: A Journal of the IMA* 13.4, iaae026.

R. Dwivedi and L. Mackey (2024). "Kernel thinning". *Journal of Machine Learning Research* 25.152, pp. 1–77.

J. Feydy, T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouve, and G. Peyré (2019). "Interpolating between Optimal Transport and MMD using Sinkhorn Divergences". In: *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2681–2690.

Z. Goldfeld, K. Kato, G. Rioux, and R. Sadhu (2022). "Limit theorems for entropic optimal transport maps and the Sinkhorn divergence". *arXiv preprint arXiv:2207.08683*.

A. Gonzalez-Sanz, J.-M. Loubes, and J. Niles-Weed (2022). "Weak limits of entropy regularized Optimal Transport; potentials, plans and divergences". arXiv: `2207.07427 [math.PR]`.

S. Hayakawa, H. Oberhauser, and T. Lyons (2023). *Sampling-based Nyström Approximation and Kernel Quadrature*. arXiv: `2301.09517 [math.NA]`.

A. Kokot, A. Hemmady, V. Thiyageswaran, and M. Meila (2026). *Local EGOP for Continuous Index Learning*. arXiv: `2601.07061 [stat.ML]`.

A. Kokot and A. Luedtke (2025). "Coreset selection for the Sinkhorn divergence and generic smooth divergences". *arXiv preprint arXiv:2504.20194*. Manuscript submitted to JMLR (May 2025).

A. Kokot, O.-V. Murad, and M. Meila (2025). "The Noisy Laplacian: a Threshold Phenomenon for Non-Linear Dimension Reduction". In: *Forty-second International Conference on Machine Learning*.

G. Mordant (2024). "The entropic optimal (self-) transport problem: Limit distributions for decreasing regularization with application to score function estimation". *arXiv preprint arXiv:2412.12007*.

V. Nobis and O. Wittich (2019). "A Convergence Result for Dirichlet Semigroups on Tubular Neighbourhoods and the Marginals of Conditional Brownian Motion". *arXiv preprint arXiv:1908.01385*.

A. Radhakrishnan, D. Beaglehole, P. Pandit, and M. Belkin (2022). "Mechanism of feature learning in deep fully connected networks and kernel machines that recursively learn features". *arXiv preprint arXiv:2212.13881*.

S. Sasaki (1958). "On the differential geometry of tangent bundles of Riemannian manifolds". *Tohoku Mathematical Journal, Second Series* 10.3, pp. 338–354.

H. Takeda, S. Farsiu, and P. Milanfar (2007). "Kernel regression for image processing and reconstruction". *IEEE Transactions on image processing* 16.2, pp. 349–366.

J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher (2017). *Fixed-Rank Approximation of a Positive-Semidefinite Matrix from Streaming Data*. arXiv: `1706.05736 [cs.NA]`.

G. Yuan, M. Xu, S. Kpotufe, and D. Hsu (2025). "Efficient estimation of the central mean subspace via smoothed gradient outer products". *SIAM Journal on Mathematics of Data Science* 7.3, pp. 1241–1264.