Rao-Blackwellized Score Matching on Manifolds

Divit Rawal • UC Berkeley, Department of Statistics • SPIGM @ ICML 2026

Denoising Score Matching

Many scientific data are constrained to lower-dimensional manifolds $M \subset \mathbb{R}^D$; e.g., $SO(3)$ for protein data and S^2 for climate data. The latent law is $q\,\mathrm{dVol}_M$, whose intrinsic score is $\nabla_M \log q$. As a distribution in $\mathbb{R}^D$, this law is singular with respect to Lebesgue measure, so the ambient score is not defined. Ambient methods use Denoising Score Matching (DSM), which adds Gaussian noise:

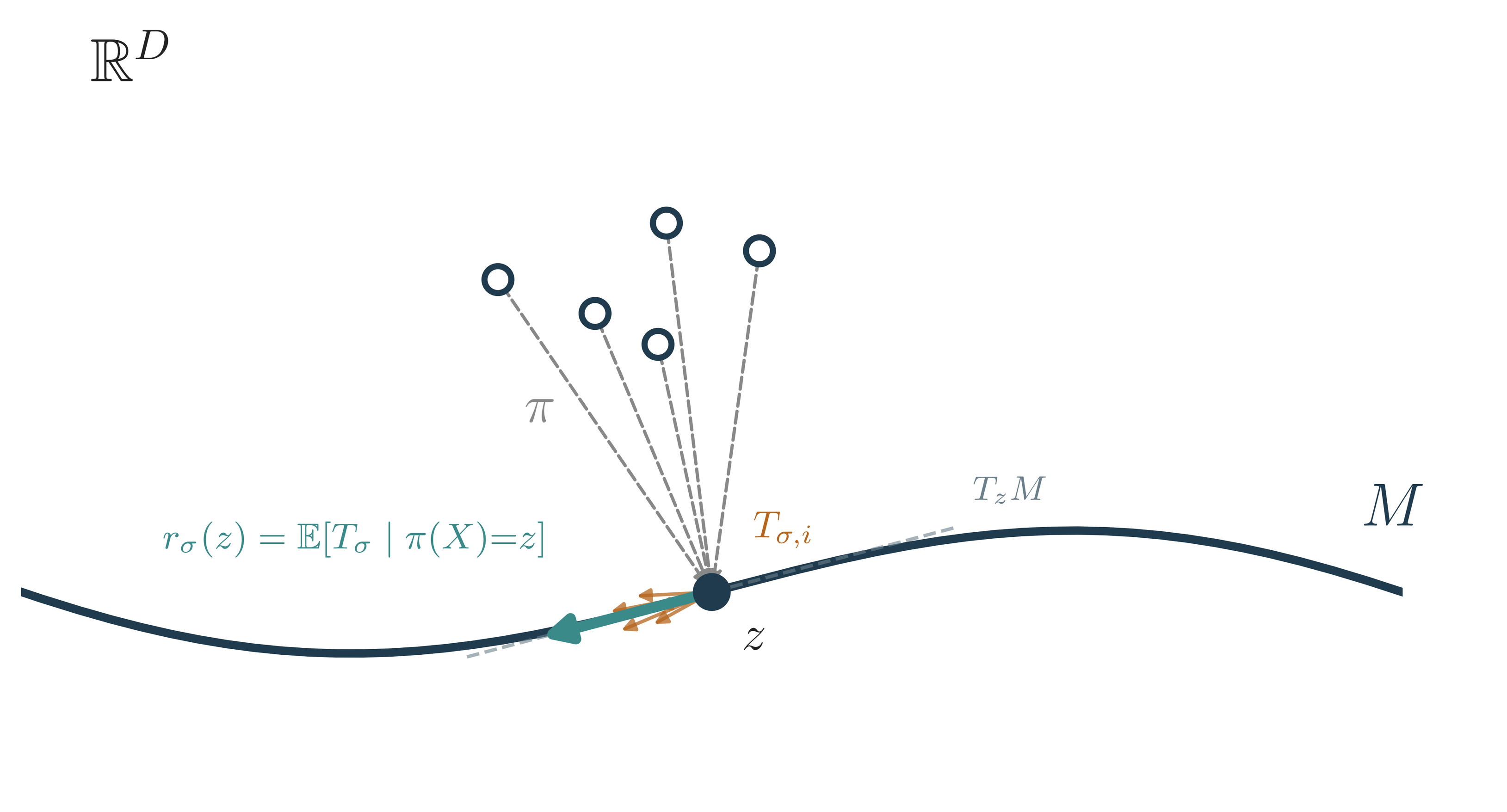
$$X \doteq Z + \sigma\xi, \qquad Z \in M.$$

DSM therefore learns a surrogate that mixes two effects:

- **statistical noise** from the normal fiber around M;
- **geometric bias** from the way M bends inside $\mathbb{R}^D$.

The natural question is: *What object does ambient DSM actually learn on manifold-supported data?*

Setup and Canonical Target



Observations X_i in the normal fiber over z share $\pi(X) = z$; their raw tangent targets $T_{\sigma,i}$ scatter within T_zM with variance d/σ^2, which Rao-Blackwellization averages into $r_\sigma(z)$.

Project each noisy point back to the manifold with the nearest-point map $\pi(X)$. The tangent part of the denoising target is given by

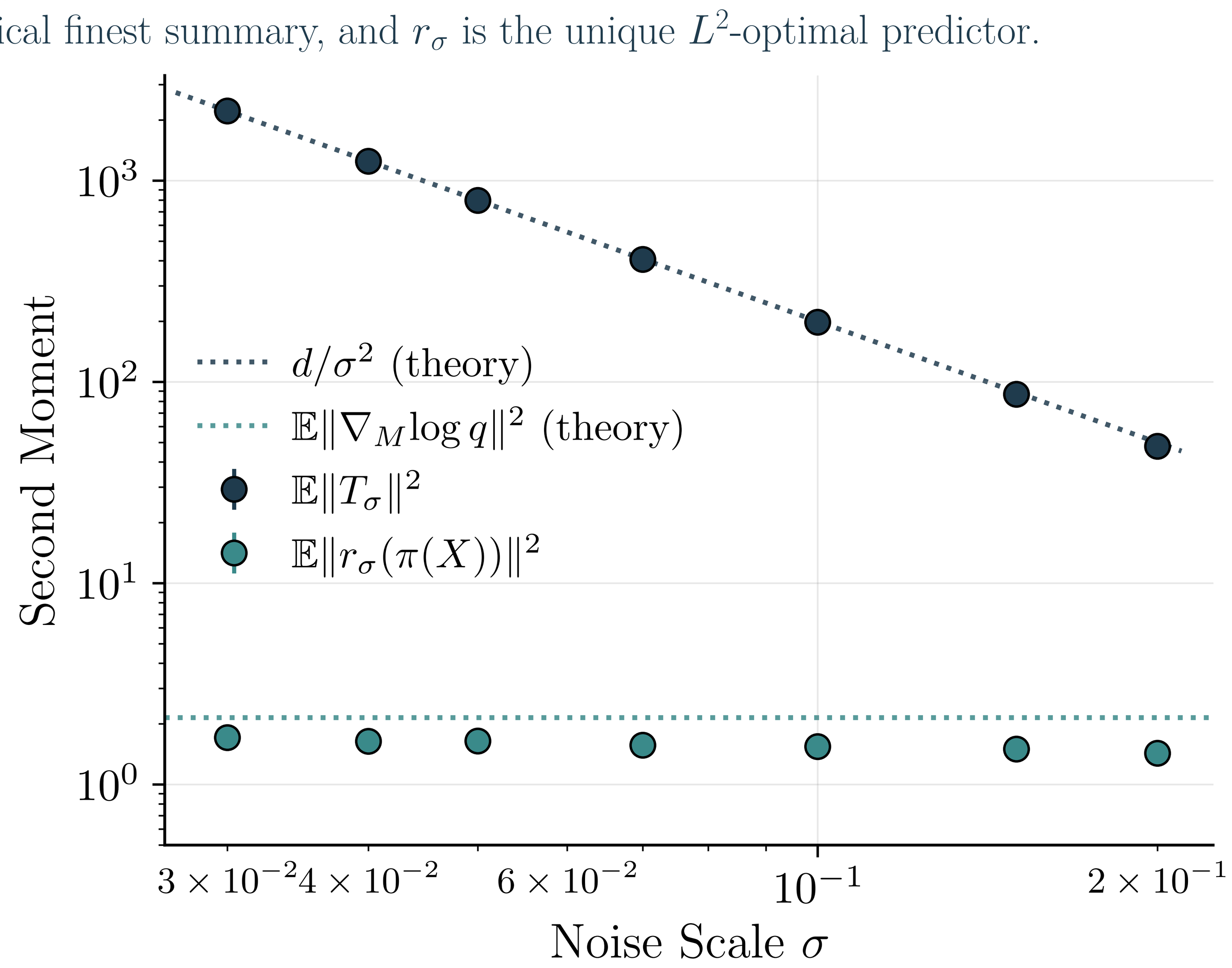
$$T_\sigma \doteq \tfrac{1}{\sigma^2} P_{T(\pi(X))}\big(Z - \pi(X)\big) \ \in\ T_{\pi(X)}M.$$

Now we Rao-Blackwellize: replace the noisy target T_σ by its conditional expectation given the summary $\pi(X)$:

$$r_\sigma(z) \doteq \mathbb{E}[T_\sigma \mid \pi(X) = z]\,.$$

Variance Reduction

The conditional variance exposes the normal-fiber noise:
$\mathrm{Var}(T_\sigma \mid \pi(X) = z) = \dfrac{d}{\sigma^2} + O(1)$. Averaging over that fiber gives r_σ with $O(1)$ variance. This d/σ^2 floor is irreducible: among all summaries S that depend on $\pi(X)$, $\pi(X)$ is the canonical finest summary, and r_σ is the unique L^2-optimal predictor.



Raw target variance scales like d/σ^2; the Rao-Blackwellized target stays flat on S^2.

Bias of Ambient DSM

Rao-Blackwellization removes the exploding variance, but it does not make ambient DSM exactly intrinsic. The finite-noise target is

$$r_\sigma(z) = \nabla_M \log q(z) + \sigma^2\big(b_q(z) + g_M^{\mathrm{ext}}(z)\big) + o(\sigma^2).$$

Two (second-order) biases remain:

- **Intrinsic smoothing bias** b_q: Tweedie/Gaussian-smoothing bias, which exists on any manifold.
- **Extrinsic curvature bias** g_M^{ext}: a curvature-dependent bias caused by adding noise in $\mathbb{R}^D$ instead of on M.

The curvature-bias coefficient is computable from the embedding:

$$g_M^{\mathrm{ext}}(z) = \big(\tfrac{1}{2} W_{H(z)} - \mathrm{Ric}_z^\sharp\big)\nabla_M \log q.$$

Here $W_{H(z)}$ is the Weingarten operator, and $\mathrm{Ric}_z^\sharp$ is the Ricci endomorphism. They measure extrinsic curvature (how the manifold sits in ambient space) and intrinsic curvature (how the manifold curves independently of its embedding), respectively.

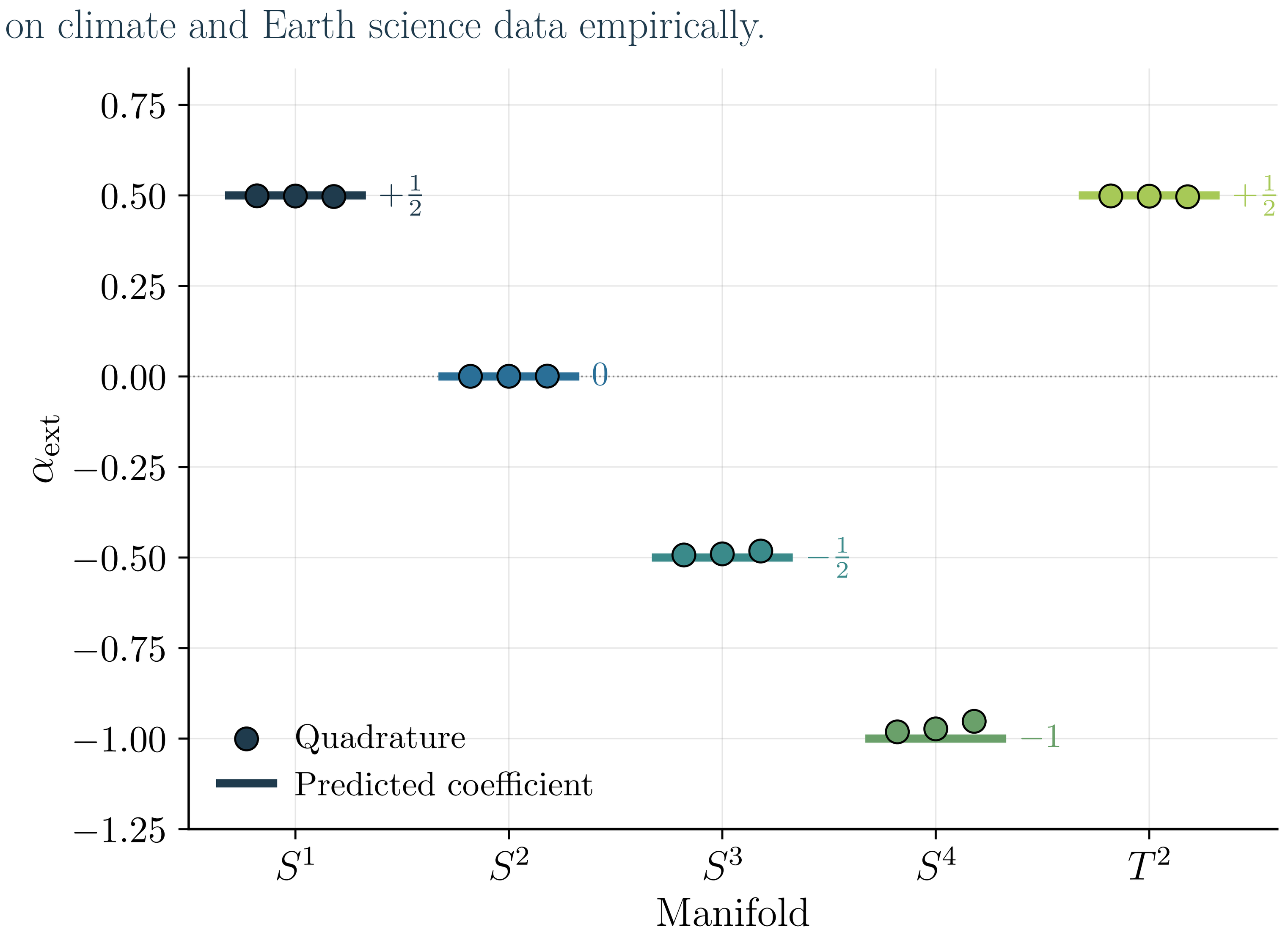
No ambient bending. If M is a flat plane in $\mathbb{R}^D$, then $g_M^{\mathrm{ext}} = 0$ and ambient DSM reduces exactly to ordinary lower-dimensional Gaussian DSM.

Hyperspheres

On the unit d-dimensional hypersphere S^d, $W_{H(z)}$ and $\mathrm{Ric}_z^\sharp$ are proportional. The geometric correction reduces to

$$g_{S^d}^{\mathrm{ext}}(z) = \big(1 - \tfrac{d}{2}\big)\nabla_M \log q(z).$$

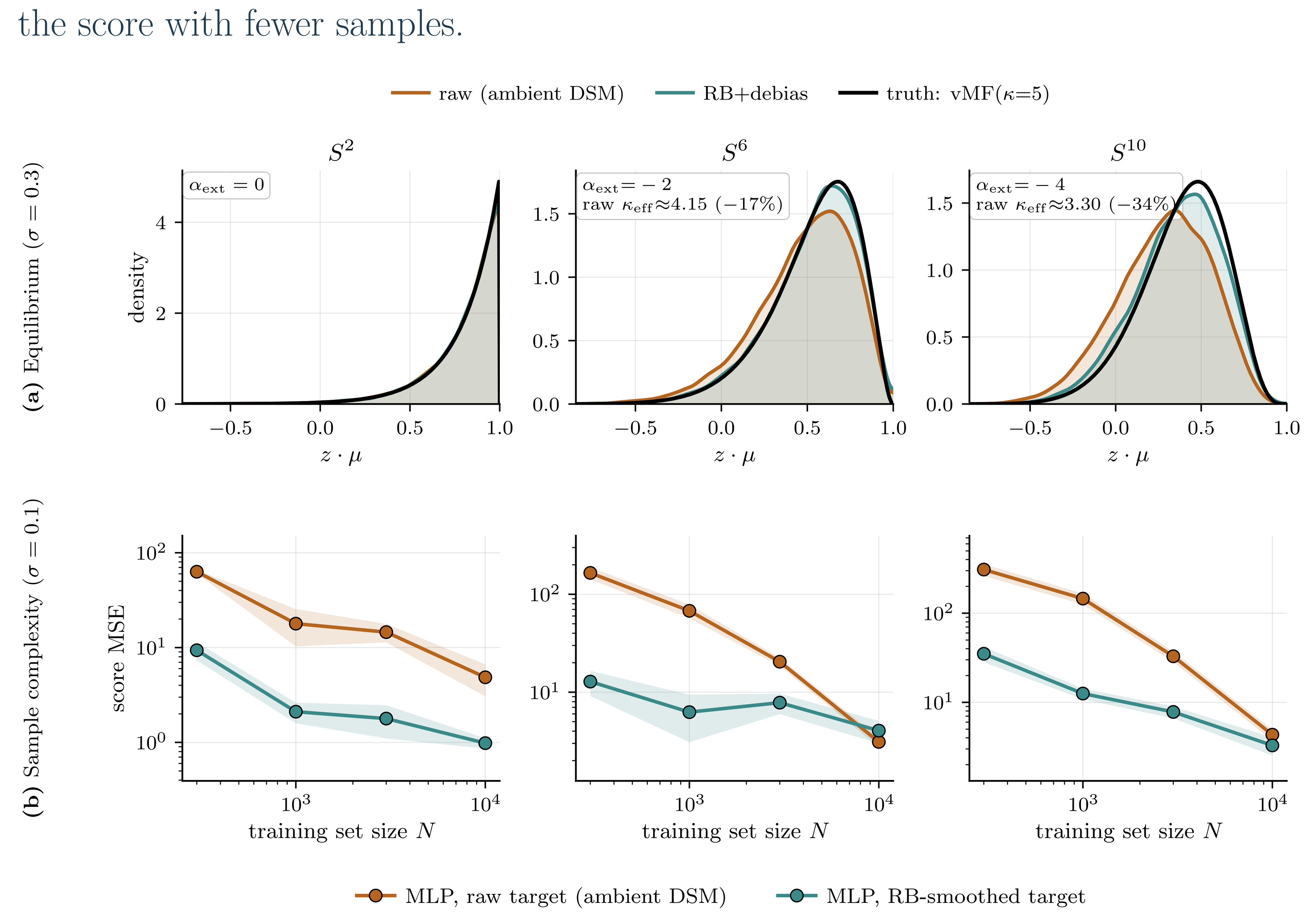
The coefficient α_{ext} by dimension is: $d = 1$: $+\frac{1}{2}$, $d = 2$: 0, $d = 3$: $-\frac{1}{2}$, $d = 4$: -1. Thus, g^{ext} vanishes exactly on S^2, which helps explain why ambient DSM performs so well on climate and Earth science data empirically.



Predicted $\alpha_d \doteq 1 - d/2$ vs. Gauss–Hermite quadrature across S^1–S^4 and T^2 at different values of σ. T^2 is intrinsically flat but extrinsically curved, demonstrating that g^{ext} depends on the manifold's embedding.

Consequences

- **Ambient bias:** raw ambient DSM (orange) becomes increasingly biased as d grows; RB+debias (blue) reduces bias.
- **Sample complexity:** Rao-Blackwellization allows score-matching models to learn the score with fewer samples.



Debiasing fixes ambient bias; Rao-Blackwellization improves score learning.

Full paper at arXiv:2605.25567.

Email: `divit.rawal@berkeley.edu`.
Website: `divitr.github.io`.