SD-KDE: Score-Debiased Kernel Density Estimation

Elliot L. Epstein   Rajat Dwaraknath
Thanawat Sornwanee   John Winnicki   Jerry Weihong Liu

NeurIPS 2025

Motivation

- KDE estimates an unknown probability density by smoothing sample data
- The bandwidth controls the balance between bias and variance
- The score function is the gradient of the log-density, pointing toward regions where probability increases most rapidly
- Diffusion models use the score to reverse noise and recover sharp samples
- SD-KDE applies the same idea to correct KDE's over-smoothing, reducing bias

Method and Theory

Algorithm: One-Step Score Debiasing

1. Given data x_i, score estimator $\hat{s}$, kernel K, bandwidth h
2. Shift each point: $\tilde{x}_i = x_i + \delta\, \hat{s}(x_i)$
3. Estimate density: $\hat{p}(x) = \frac{1}{nh^d} \sum_i K\left(\frac{x - \tilde{x}_i}{h}\right)$

Optimal choices: $\delta = \frac{h^2}{2}, \quad h_* = \Theta(n^{-1/(d+8)})$

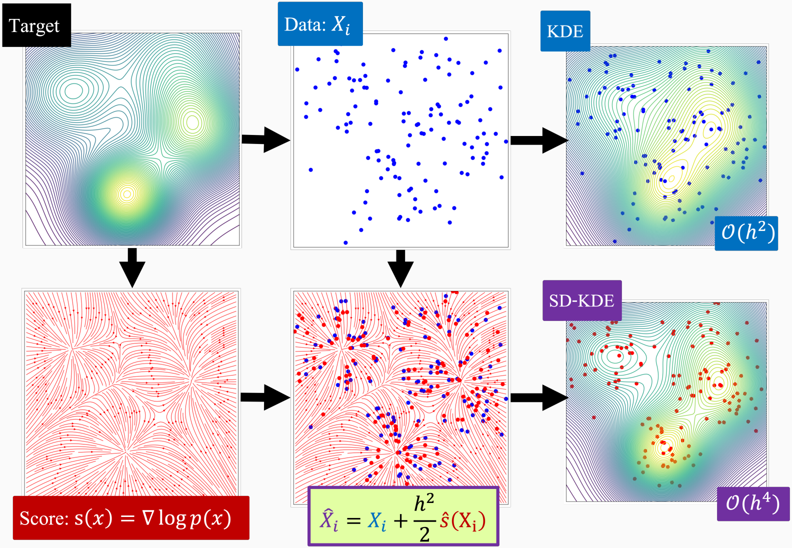
Theoretical Guarantee

- Bias order reduces from $\mathcal{O}(h^2)$ to $\mathcal{O}(h^4)$; variance remains $\mathcal{O}(1/(nh^d))$
- Balancing terms yields the optimal bandwidth $h_* = n^{-1/(d+8)}$
- Provably improves AMISE convergence rate: from $\Theta(n^{-4/(d+4)})$ to $\Theta(n^{-8/(d+8)})$

Intuition

A short step along the score sharpens samples, canceling KDE's leading bias while keeping variance unchanged

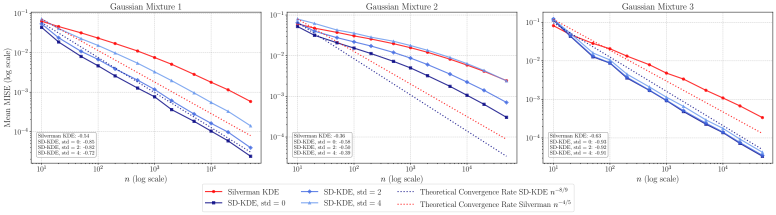
Algorithm Illustration



SD-KDE moves points along the score direction before smoothing, reducing bias without increasing variance

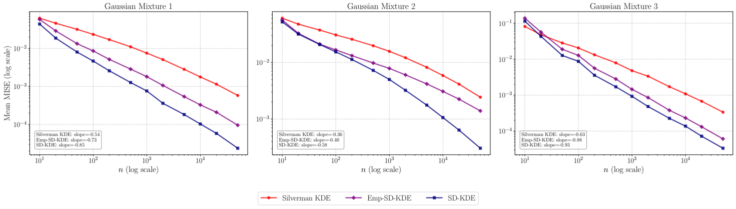
Experiments: 1D Gaussian Mixtures

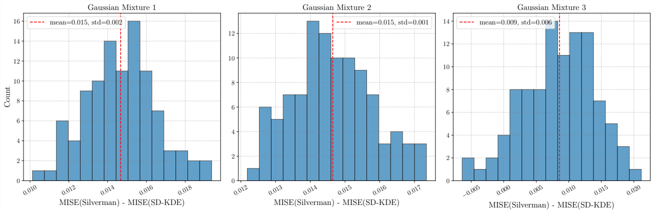
Setup: Three 1D Gaussian mixtures; baseline Silverman KDE, MISE averaged over 50 seeds



- SD-KDE shows faster asymptotic scaling than Silverman KDE
- Slopes match theory ($n^{-8/9}$)
- Robust to noisy score estimates

Empirical Score and Consistency





⇒ You can approximate the score using KDE itself, no score oracle required

⇒ SD-KDE consistently outperforms Silverman across datasets and seeds

Key Takeaways

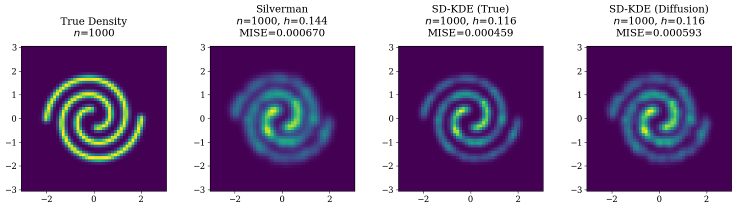
Theoretical Contribution

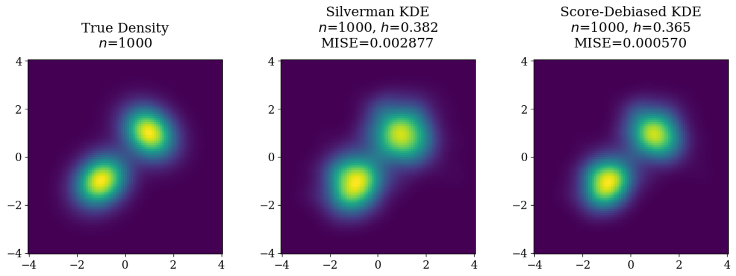
- Bias reduction: removes leading $\mathcal{O}(h^2)$ term using a one-step score shift
- Convergence: provably improves KDE rate from $\Theta(n^{-4/(d+4)})$ to $\Theta(n^{-8/(d+8)})$ in AMISE

Empirical Contribution

- Practical algorithm: simple, one-step KDE debiasing
- No true score needed: approximate directly from data
- Consistent gains: outperforms Silverman KDE across experiments

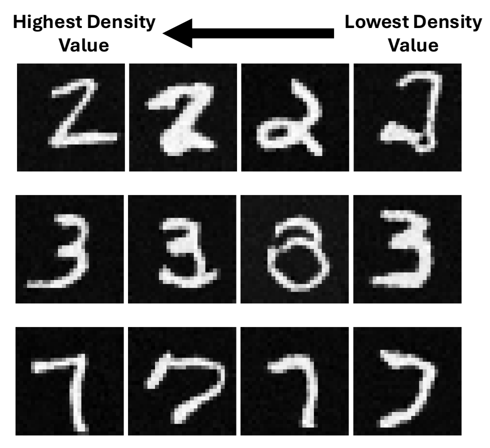
Appendix: 2D Synthetics Experiments





SD-KDE outperforms Silverman KDE with oracle or learned diffusion scores

Appendix: Latent-Space Density Ordering on MNIST



Ranking by SD-KDE-estimated density correlates with image realism