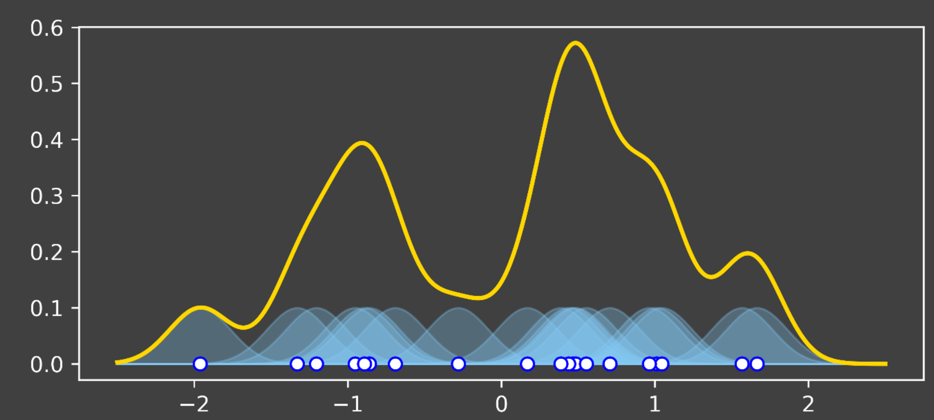
Efficient Interpolation of Density Estimators

Paxton Turner (MIT), Jingbo Liu (UIUC), Philippe Rigollet (MIT)

Motivation

Fast evaluation of kernel density estimators

$$\hat{f}(y) = \frac{1}{n}\sum_{j=1}^{n} K_h(X_j - y)$$



- Statistically accurate, computationally slow
- Many known speed-ups:
  Fast Gauss transform, locality sensitive hashing, coresets, binning, interpolation

Question: Given an accurate estimator, can we convert to a computationally tractable form?

Problem

Given: good estimator $\hat{f}$ for unknown f ← Hölder smooth of order β

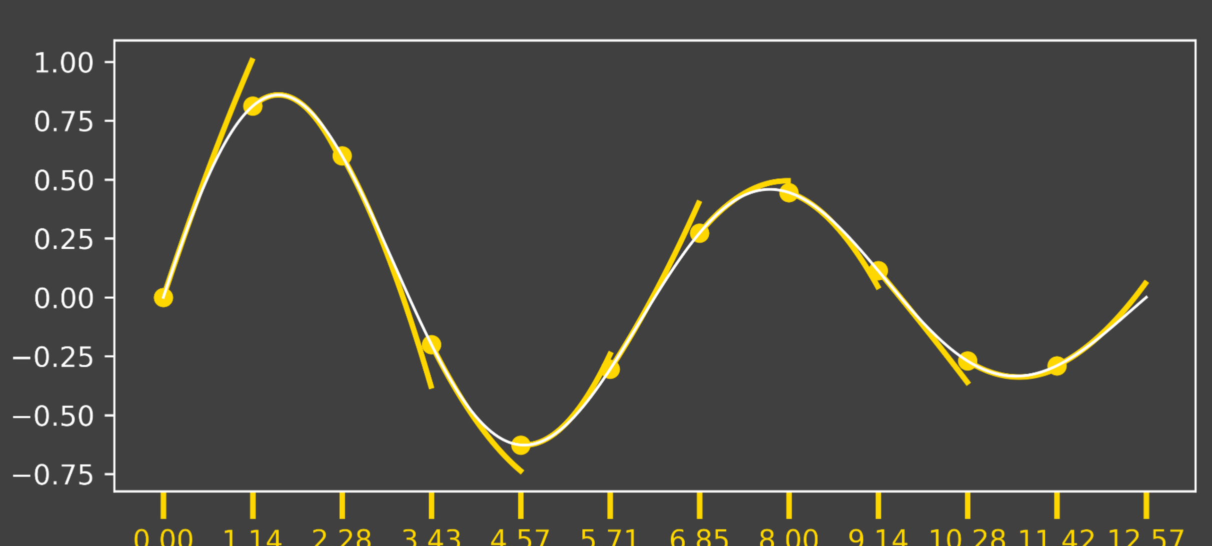
Goal: Convert $\hat{f}$ to $\hat{g}$ satisfying

1. (Accurate) $\hat{g}$ is a good estimator for f
2. (Low-space) $\hat{g}$ can be stored efficiently
3. (Fast) $\hat{g}$ can be queried efficiently

Our approach

Fact: $f \approx$ degree β piecewise polynomial

Strategy: recover these polynomials from $\hat{f}$

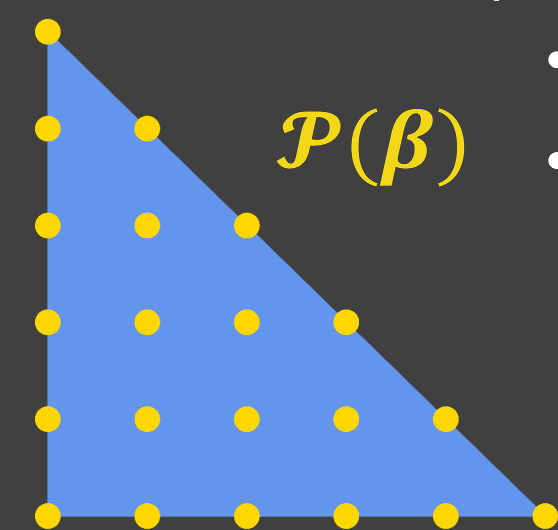


Principal Lattice Interpolation

Definition (principal lattice) Let $\beta \in \mathbb{Z}$. Define

$$\mathcal{P}(\beta) = \{x \in \mathbb{R}^d : \beta x \in \mathbb{Z}^d_{\geq 0} \; and \; \textstyle\sum_{i=1}^d x_i \leq 1\}$$

Example ($d = 2, \beta = 5$)



Properties

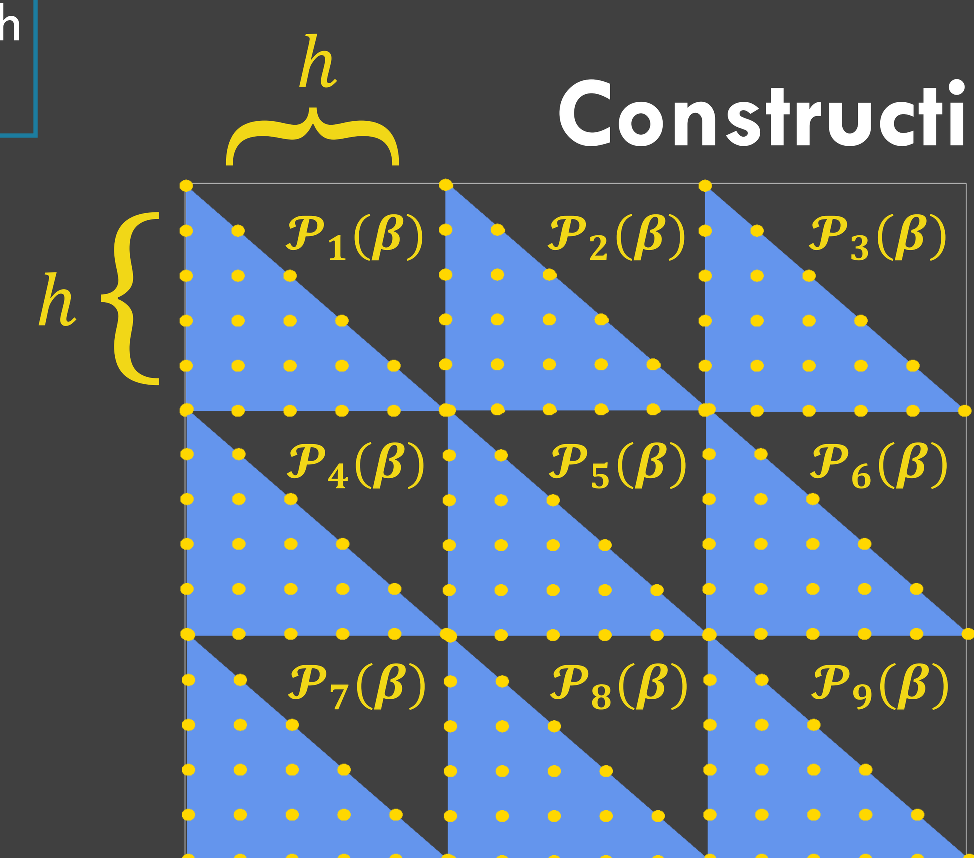
- $|\mathcal{P}(\beta)| = \binom{\beta+d}{\beta}$
- Every $x \in \mathcal{P}(\beta)$ has associated linear functions $\{h_j^x\}_{j=1}^{\beta}$ such that $\forall x' \in \mathcal{P}(\beta)$, $\prod_{j=1}^{\beta} h_j^x(x') = \mathbf{1}(x' = x)$

Theorem (Chung-Yao) The degree β polynomial

$$q(y) = \textstyle\sum_{x\in\mathcal{P}(\beta)} a_x \prod_{j=1}^{\beta} h_j^x(y)$$

satisfies $q(x) = a_x \; \forall x \in \mathcal{P}(\beta)$.

Construction of $\hat{g}$



For each box B_j, compute the Chung-Yao interpolant $q_j(y)$ on $\mathcal{P}_j(\beta)$

↓

$$\hat{g}(y) = \textstyle\sum_{B_j} \mathbf{1}(y \in B_j) \sum_{x\in\mathcal{P}_j(\beta)} \hat{f}(x) \prod_{j=1}^{\beta} h_j^x(y)$$

Analysis: Suppose

$$|\hat{f}(x) - f(x)| \lesssim \varepsilon \quad \forall j, x \in \mathcal{P}_j(\beta)$$

Let $h \asymp \varepsilon^{d/\beta}$. If $t(x)$ = degree β Taylor expansion,

$$|f(x) - t(x)| \lesssim \varepsilon$$

Thus

$$|\hat{g}(y) - t(y)| \leq \textstyle\sum_{x\in\mathcal{P}_j(\beta)} |\hat{f}(x) - t(x)| \, |h_j^x(y)| \lesssim \varepsilon$$

Results

Theorem Suppose $\hat{f}$ is pointwise minimax optimal:

$$\sup_{y\in[0,1]^d} \mathbb{P}\left[\,|\hat{f}(y) - f(y)| > t\, n^{-\frac{\beta}{2\beta+d}}\right] \lesssim e^{-ct^2}$$

Then $\hat{g}$ has these properties.

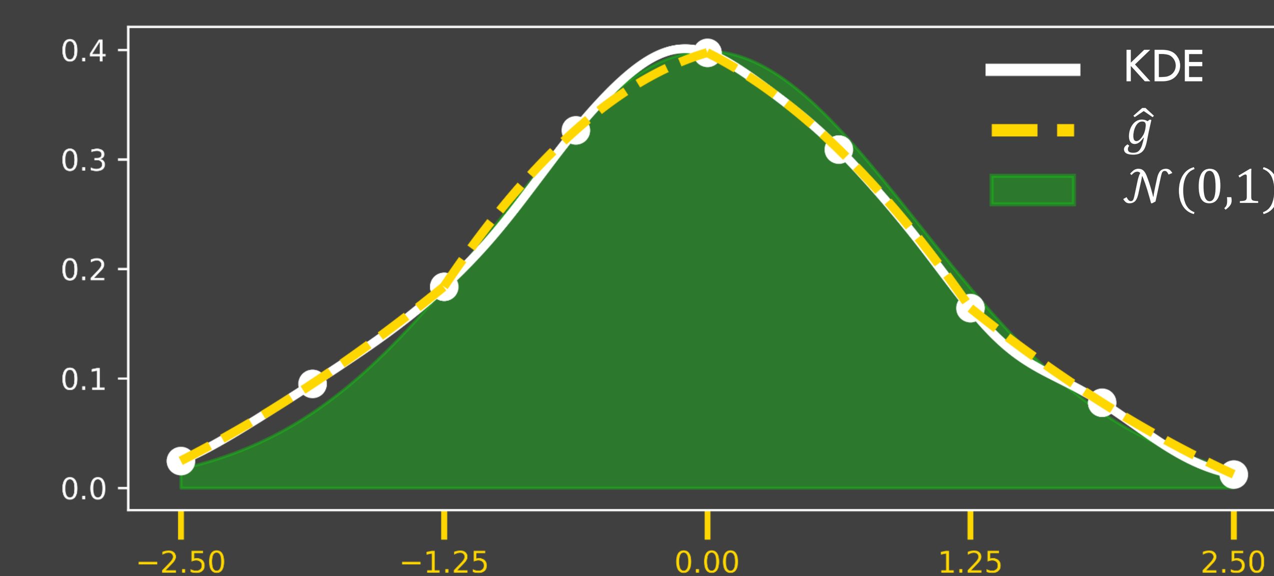
- Sublinear space: $\tilde{O}_{\beta,d}(n^{\frac{d}{2\beta+d}})$
- Near-constant query time: $\tilde{O}_{\beta,d}(1)$
- Near-minimax error:

$$\|f - \hat{g}\|_\infty \lesssim \tilde{O}_{\beta,d}(n^{-\frac{\beta}{2\beta+d}})$$

Remarks

1. Space, error bounds are near-optimal
2. Pointwise accuracy is weaker than sup-norm minimax optimality
3. Applies to other nonparametric estimators

Question: Adaptive fast evaluation & compression?



References

Backurs, Indyk, & Wagner. 2019. Space and Time Efficient Kernel Density Estimation in High Dimensions. Advances in Neural Information Processing Systems (NEURIPS).

Charikar & Siminelakis. 2017. Hashing based-estimators for kernel density in high dimensions. Annual Symposium on Foundations of Computer Science (FOCS).

Chung & Yao. 1977. On Lattices Admitting Unique Lagrange Interpolations. SIAM Journal on Numerical Analysis.

Kogure. 1998. Effective interpolations for kernel density estimators. Journal of Nonparametric Statistics.