New bounds for kernel sums via fast spherical embeddings

Tal Wagner

Tel Aviv University CS&AI

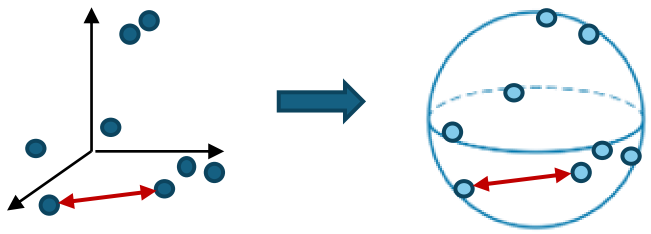
May 2026

Overview

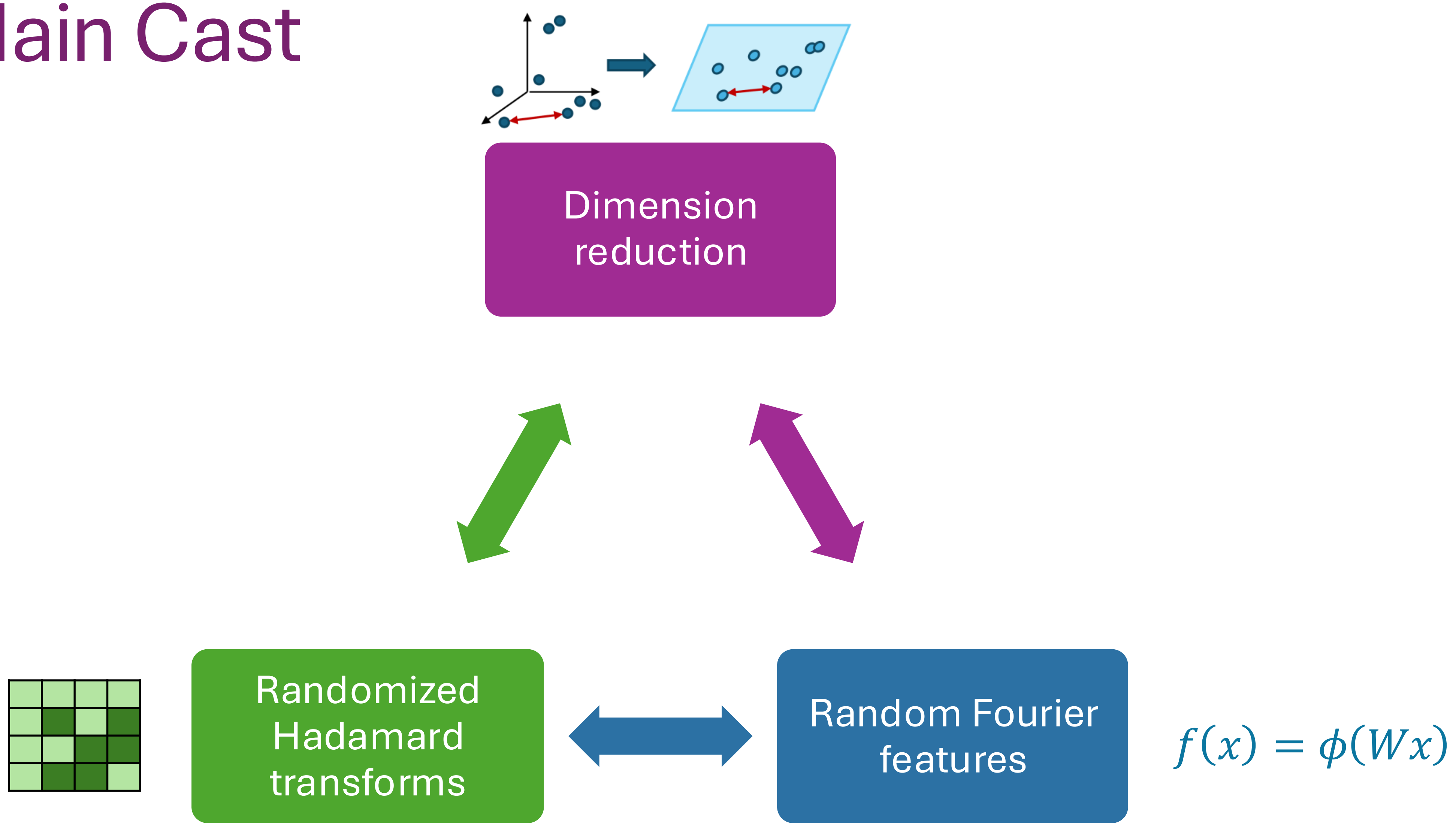
- Main theorem 1: Faster gaussian KDE approximation

$$y \mapsto \frac{1}{|X|} \sum_{x \in X} e^{-\|y-x\|_2^2} \pm \varepsilon$$

- Main theorem 2: Fast spherical embedding
  - Map high dimensional points to sphere in a distance-preserving way



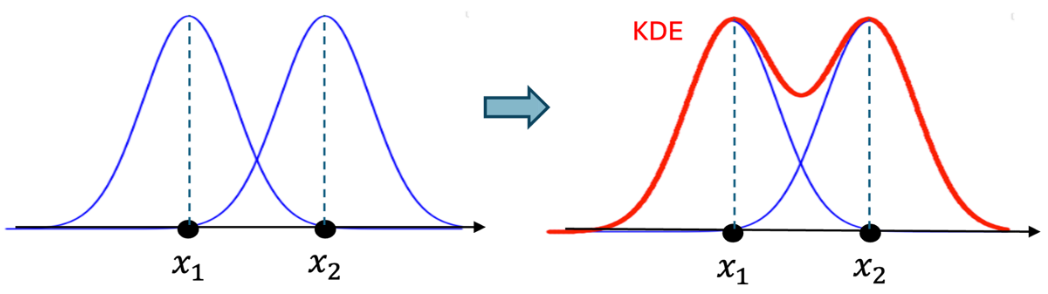
The Main Cast

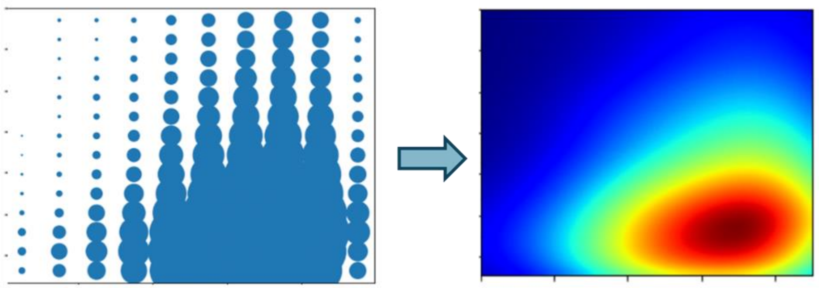


Problem:

Gaussian Kernel Density Estimation

- <u>Input</u>: Finite (arbitrary) set of n points $X \subset \mathbb{R}^d$
- <u>Task</u>: Given a query $y \in \mathbb{R}^d$, return $KDE_X(y) := \frac{1}{|X|} \sum_{x \in X} e^{-\|y-x\|_2^2}$
- **<u>Goal</u>**: Compute this up to error $\pm\varepsilon$ as fast as possible
  - Exact computation time: $O(nd)$

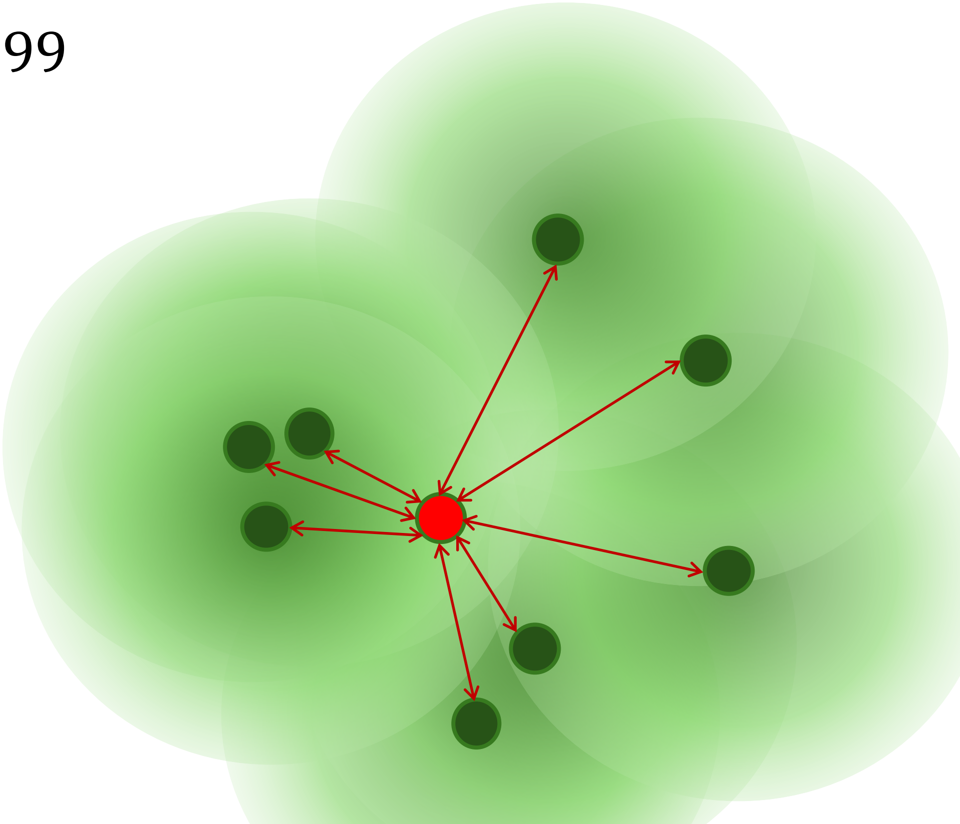




Computational Problem:
KDE Data Structures

- A KDE data structure has two operations:
  - Build (happens once in the beginning): Get $X \subset \mathbb{R}^d$, build data structure
  - Query (happens many times): Get $y \in \mathbb{R}^d$, return $KDE_X(y) \pm \varepsilon$
- Requirement: Each query succeeds with probability 0.99
- **Goal: Minimize query time**

- Assumptions:
  - High dimension: $\exp(d)$ not allowed
  - Bounded range: points have diameter Δ



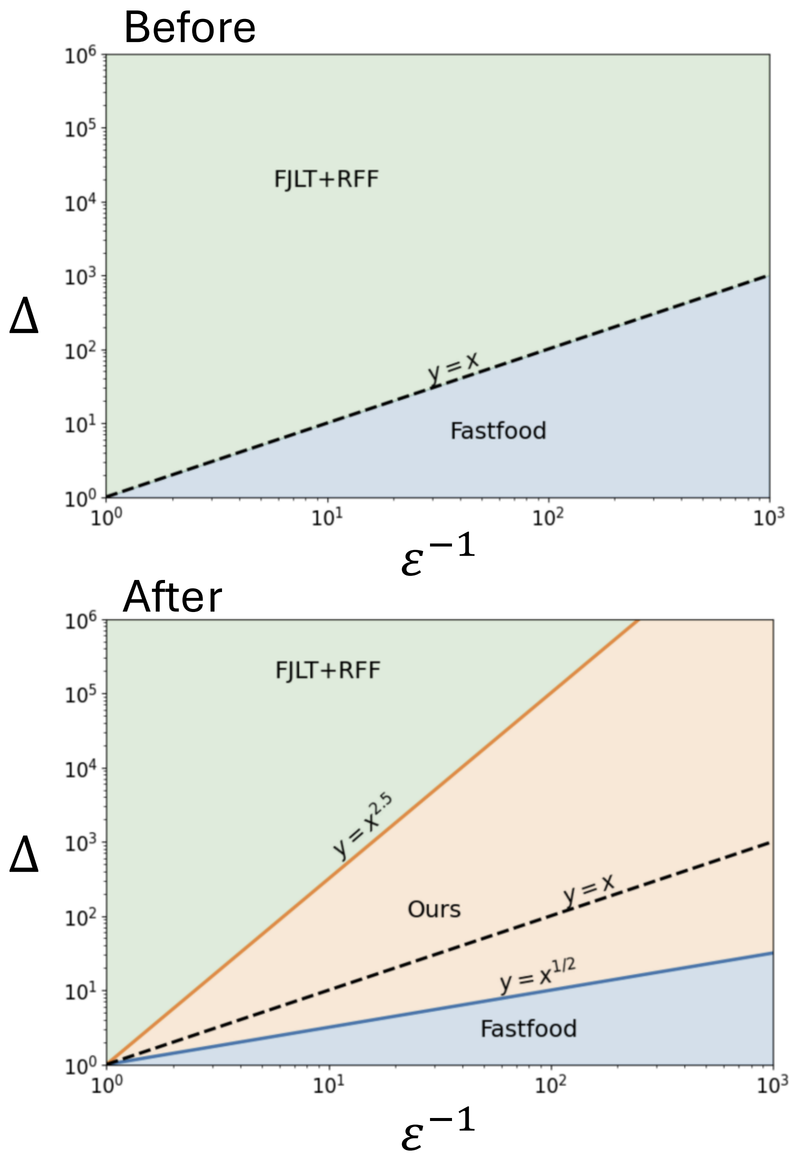
Our Result

Hypothesis:

$\tilde{O}\left(d + \frac{1}{\varepsilon^2}\right)$???

What query time can we get?

- Prior work:
  - $\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{O(d)}$ [Fast Gauss Transform]
  - $O\left(\frac{d}{\varepsilon^2}\right)$ [Random Fourier Features]
  - $\tilde{O}\left(d + \frac{1}{\varepsilon^4}\right)$ [Fast dimension reduction + RFF]
  - $\tilde{O}\left(d + \frac{\Delta^2}{\varepsilon^2}\right)$ ["Fastfood"]
- Ours:
  - $\tilde{O}\left(d + \varepsilon\Delta^2 + \frac{1}{\varepsilon^3}\right)$



Blueprint: Linear Features for KDE

- KDE data structure blueprint:
  - Find a feature map $f: \mathbb{R}^d \to \mathbb{R}^m$ such that $f(x)^T f(y) \approx e^{-\|y-x\|_2^2}$
  - Build: Compute $F(X) = \frac{1}{|X|} \sum_{x \in X} f(x)$
  - Query: Return $F(X)^T f(y) = \frac{1}{|X|} \sum_{x \in X} f(x)^T f(y) \approx \frac{1}{|X|} \sum_{x \in X} e^{-\|y-x\|_2^2} = KDE_X(y)$
- In low dimensions: Fast Gauss Transform [Greengard & Strain '91]:
  - Get f via truncated Hermite (or Taylor) expansion (but has $\exp(d)$)
- In high dimensions?

Random Fourier Features for KDE

- Goal: $f(x)^T f(y) \approx e^{-\|y-x\|_2^2}$
- Let $\phi: \mathbb{R}^m \to \mathbb{R}^{2m}$ be a mapping of vector to sin/cos pairs of its entries:

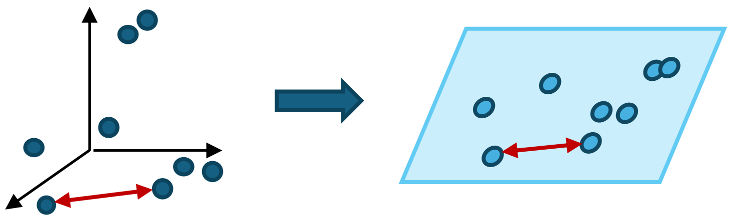
$$v \mapsto \bigoplus_{j=1}^{m} \left(\cos(v_j), \sin(v_j)\right)$$

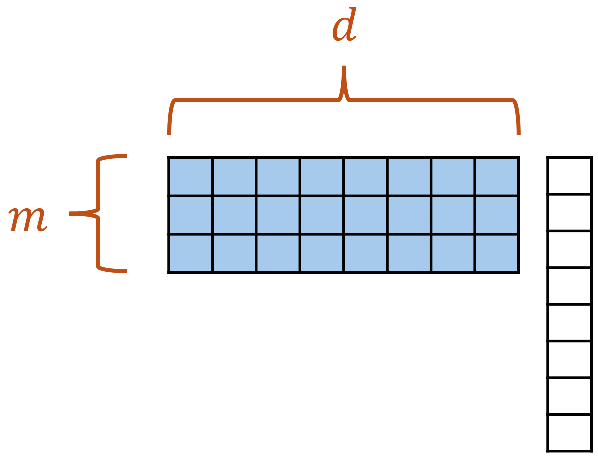
- RFF [Rahimi-Recht'07]: $f(x) = \phi(Wx)$ with iid gaussian $W \in \mathbb{R}^{m \times d}$
  - Need $m \approx 1/\varepsilon^2$, time: $O\left(\frac{d}{\varepsilon^2}\right)$
- How can we speed this up?

Dimension Reduction

[Johnson-Lindenstrauss'84]:

- We can reduce the dimension from d to $m \approx \frac{1}{\varepsilon^2}$ and preserve Euclidean distances up to $(1 \pm \varepsilon)$
- How? Multiply points by $W \in \mathbb{R}^{m \times d}$ with i.i.d. gaussian entries
- Time per point: $O\left(\frac{d}{\varepsilon^2}\right)$ -- same as RFF
- How can we speed this up?

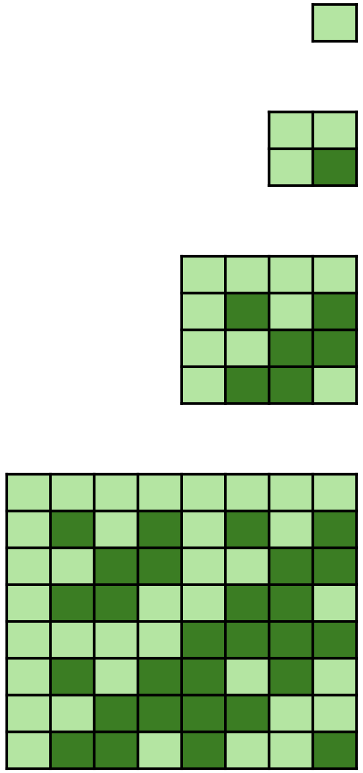




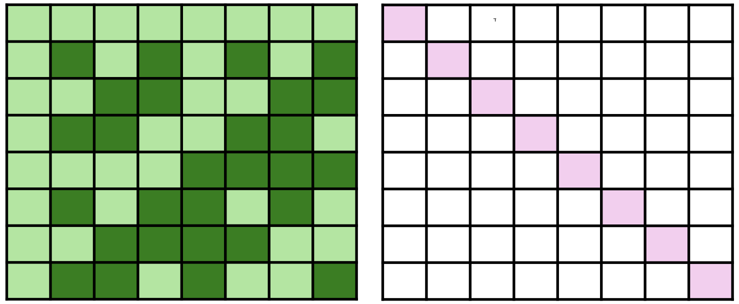
Randomized Hadamard Transform (RHT)

- The Hadamard matrix of order $2^\ell \times 2^\ell$ is defined by induction:

$$H_{2^0} = [1]\,, \qquad H_{2^\ell} = \frac{1}{\sqrt{2}}\begin{bmatrix} H_{2^\ell} & H_{2^\ell} \\ H_{2^\ell} & -H_{2^\ell} \end{bmatrix}$$

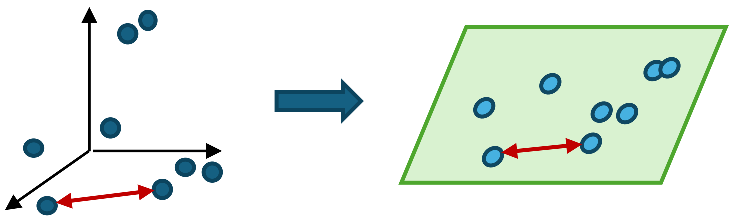


- An **RHT** is H times a random *diagonal* gaussian matrix D:



- Why?
  - HD can be multiplied by a vector in time $O(d \log d)$
  - HD behaves a lot like a full gaussian matrix

Fast Dimension Reduction: FJLT

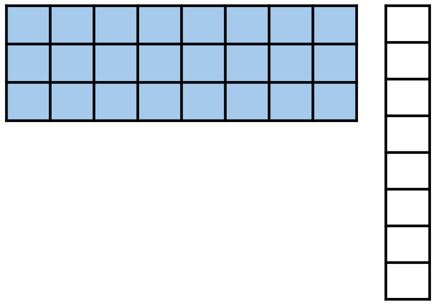


[Ailon-Chazelle'06]:

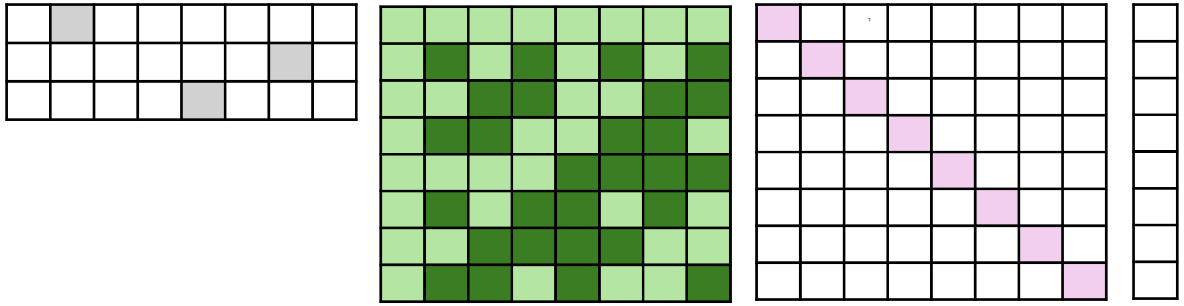
- RHT gives JL in time $\tilde{O}\left(d + \frac{1}{\varepsilon^2}\right)$
- How? Replace W with SHD
  - $S \in \mathbb{R}^{m \times d}$ is a row-sampling matrix

Same dimension reduction guarantee as dense-gaussian JL in faster time

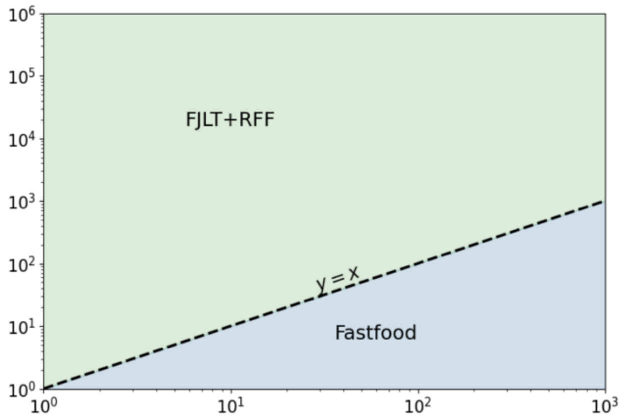
JL



FJLT



Faster RFF with RHT

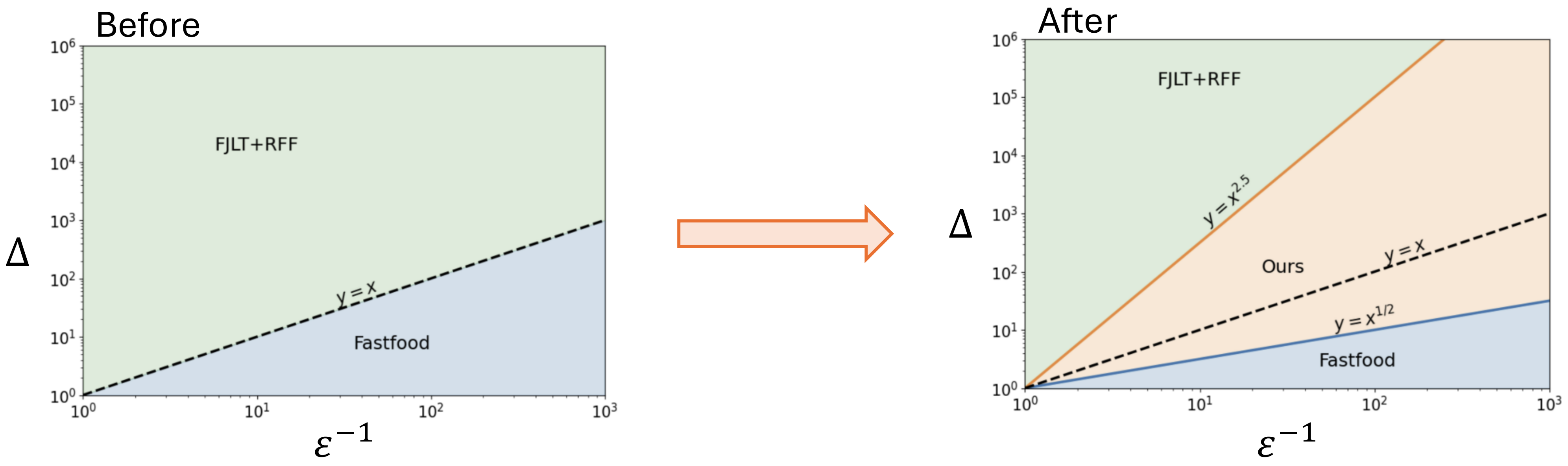


- Goal: $f(x)^T f(y) \approx e^{-\|y-x\|_2^2}$
- RFF: $f(x) = \phi(Wx)$

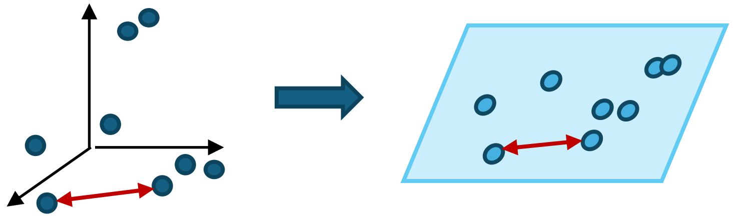
- [Backurs et al. '24]: Use FJLT as preprocessing: $f(x) = \phi(W \cdot SHDx)$
  - Still needs iid gaussian W for RFF, but now it's $\approx \frac{1}{\varepsilon^2} \times \frac{1}{\varepsilon^2}$
  - Time: $\tilde{O}\left(d + \frac{1}{\varepsilon^4}\right)$
- [Le-Sarlos-Smola'13]: "Fastfood": $f(x) = \phi(SHD_2HD_1x)$
  - Uses two iterated RHTs
  - There is theoretical and empirical literature showing iterated RHTs help
  - Cost: Incur dependence on the diameter, hence time: $\tilde{O}\left(d + \frac{\Delta^2}{\varepsilon^2}\right)$

Our Plan of Action

1. Find a way to reduce the diameter Δ
2. Make it fast
3. Apply Fastfood

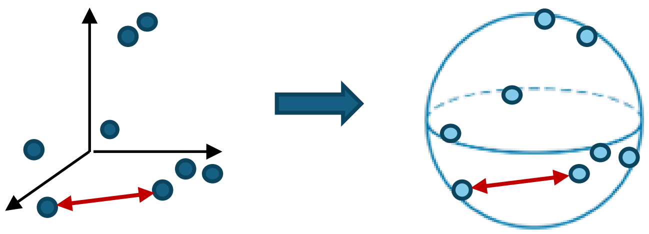


Dimension Reduction for KDE



- JL (or FJLT) is a map $g: \mathbb{R}^d \rightarrow \mathbb{R}^m$ preserves Euclidean distances up to $(1 \pm \varepsilon)$ w.h.p.:
  - Non-Expansion: $\|g(x) - g(y)\|_2^2 \leq (1 + \varepsilon)\|x - y\|_2^2$
  - Non-Contraction: $\|g(x) - g(y)\|_2^2 \geq (1 - \varepsilon)\|x - y\|_2^2$
- *But do we need to preserve all distances for KDE?*
  - Suppose $\|x - y\|_2^2 \geq \ln\left(\frac{1}{\varepsilon}\right)$
  - Then $e^{-\|x-y\|_2^2} \leq \varepsilon$, hence x contributes negligibly to $KDE_X(y)$
  - We just need it to stay negligible: $e^{-\|g(x)-g(y)\|_2^2} \leq \varepsilon$
  - So, we need Non-Collapse: $\|g(x) - g(y)\|_2^2 \geq \ln\left(\frac{1}{\varepsilon}\right)$
  - This is **weaker** than non-contraction, and should allow us to **limit diameter**

"Spherical JL" [Bartal-Recht-Schulman'11]

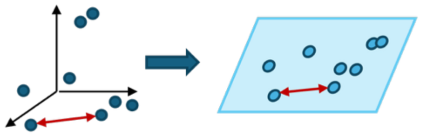


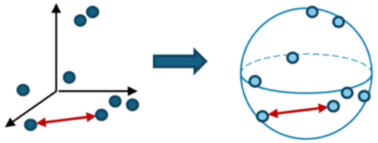
- Theorem [BRS'11] - "Spherical dimension reduction": We can embed the points into the unit sphere in dimension $m \approx \frac{1}{\varepsilon^2}$ such that for every x, y, w.h.p:
  - Non-Expansion: $\|g(x) - g(y)\|_2^2 \leq (1+\varepsilon)\|x - y\|_2^2$
  - If $\|x - y\|_2^2 \leq \varepsilon$, Non-Contraction: $\|g(x) - g(y)\|_2^2 \geq (1-\varepsilon)\|x - y\|_2^2$
  - If $\|x - y\|_2^2 > \varepsilon$, Non-Collapse: $\|g(x) - g(y)\|_2^2 \geq \Omega(\varepsilon)$
- Are we finished?
  - Scale the critical squared-distance threshold $\varepsilon \mapsto \ln(1/\varepsilon)$
  - This scales the squared-diameter to $\widehat{\Delta}^2 = \frac{\ln(1/\varepsilon)}{\varepsilon}$
  - Apply Fastfood and get time $\tilde{O}\left(d + \frac{\widehat{\Delta}^2}{\varepsilon^2}\right) = \tilde{O}\left(d + \frac{1}{\varepsilon^3}\right)$

But how much time does it take?

Faster Spherical Dimension Reduction?

- Problem: Spherical JL takes time $\left(\frac{d}{\varepsilon^2}\right)$ -- as slow as RFF
- In fact, the algorithm is identical to RFF: $g(x) = \phi(Wx)$
  - Different analysis of the same randomized map
- Can we make it faster with RHTs?





RFF time	JL: $x \mapsto Wx$	BRS: $x \mapsto \phi(Wx)$
Faster (RHTs)	FJLT: $x \mapsto SHDx$	??? (our goal)

Our Theorem: Faster Spherical Embedding

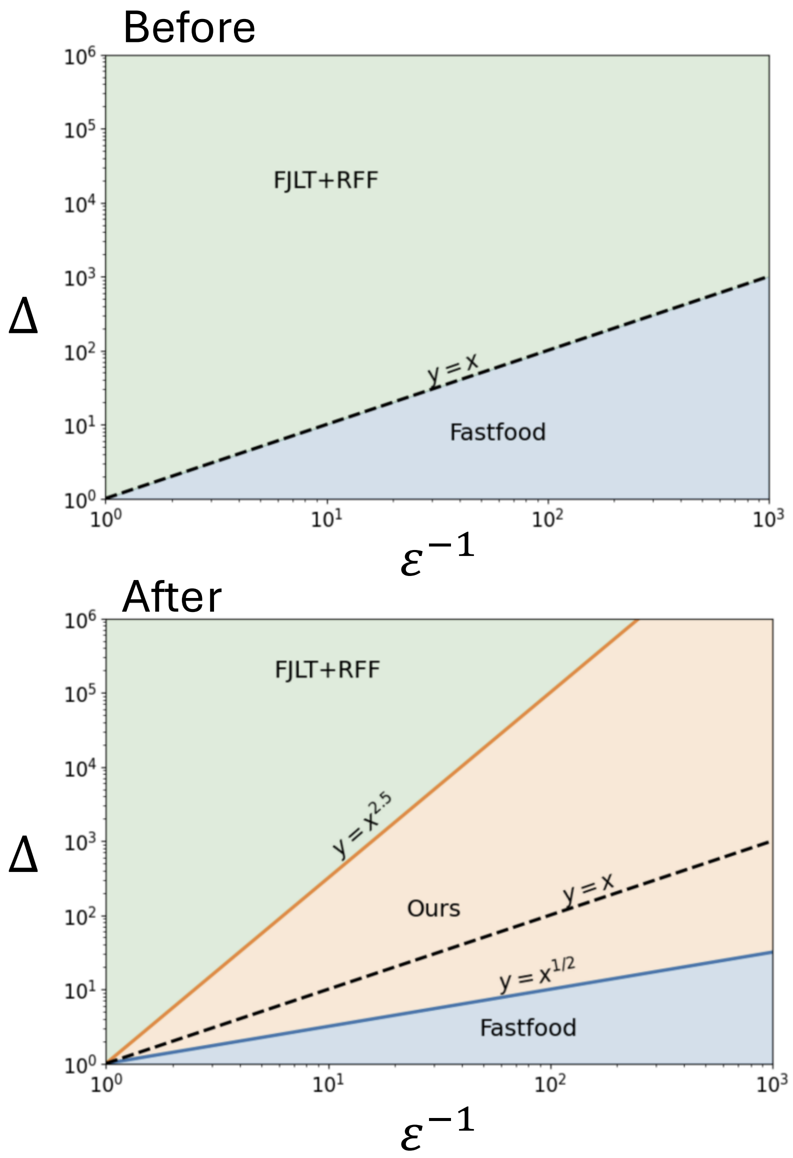
But with fine-print

- We analyze $x \mapsto \phi(HD_2HD_1x)$ as fast spherical embedding
- Theorem: Let $\Lambda > 0$. We can embed the points into the unit sphere in dimension $m \approx \frac{1}{\varepsilon^2}$ and in time $\tilde{O}\left(d + \frac{1}{\varepsilon^2} + \Lambda^2\right)$ such that for every x, y, w.h.p:
  - Non-Expansion: $\|g(x) - g(y)\|_2^2 \le (1+\varepsilon)\|x-y\|_2^2$
  - If $\|x-y\|_2^2 \le \varepsilon$, Non-Contraction: $\|g(x) - g(y)\|_2^2 \ge (1-\varepsilon)\|x-y\|_2^2$
  - If $\Lambda^2 \ge \|x-y\|_2^2 > \varepsilon$, Non-Collapse: $\|g(x) - g(y)\|_2^2 \ge \Omega(\varepsilon)$
- Our final KDE features: $f(x) = \phi\big(\underbrace{SHD_4HD_3}_{\text{Fastfood}} \cdot \underbrace{\phi(HD_2HD_1x)}_{\text{Our new spherical embedding}}\big)$

Four consecutive RHTs

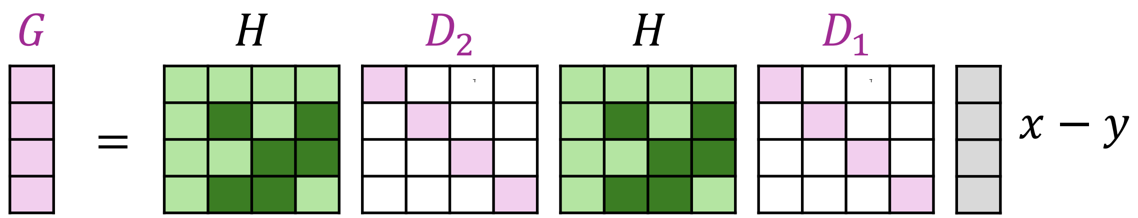
Our Theorem: Faster KDE

- Scale the critical distance² from $\ln(1/\varepsilon)$ to ε
- Input diameter² to fast spherical embedding: $\approx \varepsilon\Delta^2$
- Fast spherical embedding takes time: $\tilde{O}\left(d + \frac{1}{\varepsilon^2} + \varepsilon\Delta^2\right)$
- Output diameter² from fast spherical embedding: $\approx 1/\varepsilon$
- Fastfood takes time: $\tilde{O}\left(d + \frac{1}{\varepsilon^3}\right)$
- **Total time:** $\tilde{O}\left(d + \frac{1}{\varepsilon^3} + \varepsilon\Delta^2\right)$



Gist of Analysis: Concentration for RHTs

- Our spherical embedding map is $\Phi(x) = \phi(HD_2HD_1x)$
- By trig identities: $\|\Phi(x) - \Phi(y)\|_2^2 = \frac{1}{m}\sum_{j=1}^{m}\left(1 - \cos G_j\right)$ where $G_j \sim N(0, \|x - y\|_2^2)$
- But the G_js are dependent: they are polynomials in the entries of D_1 and D_2



- Main tool: Wiener chaos decomposition into hypercontractive terms
- We need a 4th chaos analysis because $\frac{1}{2}G^2 - \frac{1}{24}G^4 \leq 1 - \cos G \leq \frac{1}{2}G^2$

Conclusion

- Main results:
  - New bound for gaussian KDE approximation up to $\pm\varepsilon$
  - New fast spherical embedding theorem (FJLT-analog of Bartal-Recht-Schulman)
- Extensions:
  - Extends to other kernels, e.g., inverse multi-quadratic: $k(x,y) = \left(\frac{1}{1+\|x-y\|_2^2}\right)^{\beta}$
  - Extends to other settings, e.g., differential privacy
    - Five RHTs! $f_{DP}(x) = HD_5 \cdot \phi\big(SHD_4HD_3 \cdot \phi(HD_2HD_1x)\big)$
- Open questions:
  - Optimal bound for gaussian KDE? $\tilde{O}\left(d + \frac{1}{\varepsilon^2}\right)$???
  - More applications of fast spherical embedding?

Thank you