Sharp Inequalities between TV and Hellinger Distances for Gaussian Mixtures

Joonhyuk Jung

Feb 4, 2026

THE UNIVERSITY OF CHICAGO

Joonhyuk Jung

1. Seoul National University (2017-2023)
   - B.S. in Statistics and B.A. in Economics
2. The University of Chicago (2024-)
   - 2nd-year Ph.D. Student in Statistics
   - Advisor: Prof. Chao Gao

Please visit https://joonhyuk.com for more details.

Authors

This is a joint work with Prof. Chao Gao. [arXiv:2602.03202]

Recap: Gaussian Location Mixtures

We define d-dimensional Gaussian location mixture by

$$f_\pi(x) = \int_{\mathbb{R}^d} \phi_d(x - \theta)\, d\pi(\theta),$$

where

$$\phi_d(x) = (2\pi)^{-d/2} \exp\left(-\frac{\|x\|_2^2}{2}\right)$$

is the standard Gaussian density.

- Nonparametric density estimation;
- Bayesian inference;
- Clustering [Lin95, Das99].

Recap: f-Divergences

Recall that, in general, we have

$$H^2(p,q) \le \mathrm{TV}(p,q) \le \sqrt{2}H(p,q) \le \sqrt{2\,\mathrm{KL}(p\|q)} \le \sqrt{2\,\chi^2(p\|q)},$$

where we define

$$H^2(p,q) := \frac{1}{2}\int(\sqrt{p}-\sqrt{q})^2,$$
$$\mathrm{TV}(p,q) := \frac{1}{2}\int|p-q|,$$
$$\mathrm{KL}(p\|q) := \int p\log\frac{p}{q},$$
$$\chi^2(p\|q) := \int\frac{(p-q)^2}{q}.$$

Related Works: Hellinger Distance

The Hellinger distance H is a commonly used loss function for density estimation as it is useful in Gaussian location mixture estimation [WS95, KG22].

Why Hellinger? Bounded metric, symmetric, tensorized, etc.

- An upper bound on $H(f_\pi, f_\eta)$ immediately implies an upper bound on $\mathrm{TV}(f_\pi, f_\eta)$.

$$H^2(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta) \leq \sqrt{2}H(f_\pi, f_\eta).$$

Related Works: Hellinger Distance

The Hellinger distance H is a commonly used loss function for density estimation as it is useful in Gaussian location mixture estimation [WS95, KG22].

Why Hellinger? Bounded metric, symmetric, tensorized, etc.

- An upper bound on $H(f_\pi, f_\eta)$ immediately implies an upper bound on $\mathrm{TV}(f_\pi, f_\eta)$.

$$H^2(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta) \le \sqrt{2}H(f_\pi, f_\eta).$$

- The Hellinger control has a direct consequence for bounding the regret of an empirical Bayes estimator, e.g., NPMLE [GW00, GVDV01, JZ09, SG20, SGS25].

Related Works: Information Theory

Suppose $\pi([-M, M]^d) = \eta([-M, M]^d) = 1$ for some $M > 0$. Then, f_π and f_η satisfy

$$H(f_\pi, f_\eta) \asymp \sqrt{\mathrm{KL}(f_\pi \| f_\eta)},$$

up to constant factors depending on M and d [JPW23].

- This allows an entropic characterization of learning in Hellinger of Gaussian mixtures (Will be discussed later).

$$\inf_{\widehat{P}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[H^2 \left(P, \widehat{P} \right) \right] \asymp \inf_{\epsilon > 0} \left(\epsilon^2 + \frac{1}{n} \underbrace{\log N_{H,loc}(\mathcal{P}, \epsilon)}_{\text{local Hellinger entropy}} \right).$$

Resolving an Open Question

- Prior to our contribution, it was questioned whether

$$\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$$

holds for Gaussian mixtures [JPW23].

Resolving an Open Question

- Prior to our contribution, it was questioned whether

$$\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$$

holds for Gaussian mixtures [JPW23].

- That is,

$$H(f_\pi, f_\eta) \overset{?}{\lesssim} \mathrm{TV}(f_\pi, f_\eta)$$

Resolving an Open Question

- Prior to our contribution, it was questioned whether

$$\mathrm{TV}(f_\pi, f_\eta) \overset{?}{\asymp} H(f_\pi, f_\eta)$$

holds for Gaussian mixtures [JPW23].

- That is,

$$H(f_\pi, f_\eta) \overset{?}{\leq} \mathrm{TV}(f_\pi, f_\eta)$$

- We say NO by proving

$$H(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta)^{1 - \dfrac{\Theta(1)}{\log\log(1/\mathrm{TV}(f_\pi, f_\eta))}}$$

and by constructing a sharp example $(\pi_n, \eta_n)_{n=1}^{\infty}$ of this inequality.

Resolving an Open Question

This might be unexpected:

Re: TV and H for Gaussian mixtures

YP Yury Polyanskiy <polyanskiy@gmail.com> Yesterday at 9:56 AM

To: Chao Gao; Cc: zyjia@mit.edu; yihong.wu@yale.edu; +1 more

```
Ah, cool result! Thanks Chao and Joonhyuk. I didn't
expect H\asymp TV is wrong.
```

Our Contribution

When the data set contains a small subset of arbitrary outliers, the density estimation problem can be regarded as misspecified under total variation.

- Too loose for deriving optimal error rates for robust density estimation:

$$H^2(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta) \leq \sqrt{2}H(f_\pi, f_\eta).$$

Our Contribution

When the data set contains a small subset of arbitrary outliers, the density estimation problem can be regarded as misspecified under total variation.

- Too loose for deriving optimal error rates for robust density estimation:

$$H^2(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta) \le \sqrt{2}H(f_\pi, f_\eta).$$

- Our contribution 1: Better inequalities!

$$H^{1+o(1)}(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta) \le \sqrt{2}H(f_\pi, f_\eta).$$

Our Contribution

When the data set contains a small subset of arbitrary outliers, the density estimation problem can be regarded as misspecified under total variation.

- Too loose for deriving optimal error rates for robust density estimation:

$$H^2(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta) \leq \sqrt{2}H(f_\pi, f_\eta).$$

- Our contribution 1: Better inequalities!

$$H^{1+o(1)}(f_\pi, f_\eta) \leq \mathrm{TV}(f_\pi, f_\eta) \leq \sqrt{2}H(f_\pi, f_\eta).$$

- Our contribution 2: The $o(1)$ term is indeed necessary, i.e., unimprovable.

Our Contribution

When the data set contains a small subset of arbitrary outliers, the density estimation problem can be regarded as misspecified under total variation.

- Too loose for deriving optimal error rates for robust density estimation:

$$H^2(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta) \le \sqrt{2}H(f_\pi, f_\eta).$$

- Our contribution 1: Better inequalities!

$$H^{1+o(1)}(f_\pi, f_\eta) \le \mathrm{TV}(f_\pi, f_\eta) \le \sqrt{2}H(f_\pi, f_\eta).$$

- Our contribution 2: The $o(1)$ term is indeed necessary, i.e., unimprovable.
- Our contribution 3: Implications in robust statistics!

Preview: Implication I

Theorem (Minimax rate of robust density estimation)

Consider the data generating process as follows.

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P := (1-\epsilon)P_{f_\pi} + \epsilon Q, \tag{1}$$

Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[H^2\left(f_\pi, \widehat{f}\right)\right] \asymp \epsilon^{2\left(1-\frac{\Theta(1)}{\log(\log(1/\epsilon)\vee e)}\right)}, \tag{2}$$

where the expectation is under (1) *and the supremum is taken over all* Q *and* π *such that* $\mathrm{supp}(\pi) \subseteq [-M, M]^d$, *provided that the sample size satisfies* $n \geq \mathrm{poly}(1/\epsilon)$.

Preview: Implication II

Theorem (Robust regret bound)

Consider the data generating process as follows.

$$X_i \sim (1-\epsilon)N(\theta_i, I_d) + \epsilon Q, \tag{3}$$
$$\theta_1, \ldots, \theta_n \overset{i.i.d.}{\sim} \pi.$$

Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[\mathbb{E}_{X \sim f_\pi} \left\| \widehat{\theta}(X) - \widehat{\theta}^\star(X) \right\|^2\right] \lesssim \epsilon^{2\left(1-\frac{\Theta(1)}{\log(\log(1/\epsilon)\vee e)}\right)} + \left(\frac{1}{n}\right)^{1-o(1)},$$

where the outer expectation is under (3) *and the supremum is taken over all* Q *and* π *such that* $\mathrm{supp}(\pi) \subseteq [-M, M]^d$.

Organization

- Main Results (Formal)
  - Proof Technique: Hermite Polynomials
  - Physical Interpretation: Quantum Harmonic Oscillator
- Sharpness Results (Formal)
- Applications
  - Entropic Characterization of Learning in Total Variation
  - Robust Density Estimation
  - Robust Regret Bound in Empirical Bayes

Main Result (Hellinger distance)

What we want to show:

Corollary (Inequality between TV and Hellinger distances)

Let π and η be probability measures supported on the d-dimensional cube $[-M, M]^d$. Let $\delta > 0$. Then, there exists $C_0 = C_0(\delta, M, d) > 0$, not depending on π or η, such that

$$H(f_\pi, f_\eta) \leq \left(C_0 \vee \mathrm{TV}^{-\alpha(\mathrm{TV}(f_\pi, f_\eta))}(f_\pi, f_\eta) \right) \mathrm{TV}(f_\pi, f_\eta),$$

where we define

$$\alpha(t) := \frac{2 + \delta}{\log\left(\log(1/t) \vee e\right)}, \tag{4}$$

for $t > 0$.

Main Theorem (χ^2-divergence)

The above corollary follows immediately from our main theorem:

Theorem (Inequality between TV distance and χ^2-divergence)

Let π and η be probability measures supported on the d-dimensional cube $[-M, M]^d$. Let $\delta > 0$. Then, there exists $C_0 = C_0(\delta, M, d) > 0$, not depending on π or η, such that

$$\sqrt{\chi^2(f_\pi \| f_\eta)} \leq \left(C_0 \vee \mathrm{TV}^{-\alpha(\mathrm{TV}(f_\pi, f_\eta))}(f_\pi, f_\eta)\right) \mathrm{TV}(f_\pi, f_\eta),$$

where we define $\alpha(\cdot)$ as in (4).

Stronger Main Theorem ($L^2(\phi_d)$-norm)

We indeed prove a stronger result.

Theorem (Inequality between $L^1(\phi_d)$ and $L^2(\phi_d)$ norms)

Let π and η be probability measures supported on the d-dimensional cube $[-2M, 2M]^d$. Let $\delta > 0$. Then, there exists $C_0 = C_0(\delta, M, d) > 0$, not depending on π or η, such that

$$\sqrt{\int \frac{(f_\pi - f_\eta)^2}{\phi_d}} \leq \left(C_0 \vee \mathrm{TV}^{-\alpha(\mathrm{TV}(f_\pi, f_\eta))}(f_\pi, f_\eta)\right) \mathrm{TV}(f_\pi, f_\eta),$$

where we define $\alpha(\cdot)$ as in (4).

Why Stronger?

Remark

Let us define $g := \frac{f_\pi - f_\eta}{\phi_d}$*. Then, we have*

$$\|g\|_{L^2(\phi_d)} = \sqrt{\int_{\mathbb{R}^d} |g|^2 \phi_d} = \sqrt{\int_{\mathbb{R}^d} \frac{(f_\pi - f_\eta)^2}{\phi_d}},$$

$$\|g\|_{L^1(\phi_d)} = \int_{\mathbb{R}^d} |g|\phi_d = \int_{\mathbb{R}^d} |f_\pi - f_\eta| = 2\mathrm{TV}(f_\pi, f_\eta).$$

Moreover, by convexity argument (Jensen and Fubini-Tonelli),

$$\chi^2(f_\pi \| f_\eta) = \int_{\mathbb{R}^d} \frac{(f_\pi - f_\eta)^2}{f_\eta} \le \sup_{\theta \in [-M,M]^d} \int_{\mathbb{R}^d} \frac{(f_\pi(x) - f_\eta(x))^2}{\phi_d(x-\theta)}\, dx.$$

- It only remains to bound $\|g\|_{L^2(\phi_d)}$ from above by $\|g\|_{L^1(\phi_d)}$.

Hermite Polynomial Expansion

Lemma (Hermite polynomial expansion)

For $\theta = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$ *and* $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, *we have*

$$\frac{\phi_d(x-\theta)}{\phi_d(x)} = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \frac{\theta^{\mathbf{k}}}{\sqrt{\mathbf{k}!}} h_{\mathbf{k}}(x),$$

where we define

$$\theta^{\mathbf{k}} := \prod_{j=1}^d \theta_j^{k_j}, \qquad \mathbf{k}! := \prod_{j=1}^d k_j!, \qquad h_{\mathbf{k}}(x) := \prod_{j=1}^d h_{k_j}(x_j).$$

- $\{h_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}_0^d\}$ is an orthonormal basis of $L^2(\phi_d)$.
- For $d = 1$ and $k \in \mathbb{N}_0$, we define $h_k(x) := \frac{(-1)^k}{\sqrt{k!}\phi_1(x)} \frac{d^k}{dx^k} \phi_1(x)$.

Hermite Polynomial Expansion

Lemma (Hermite polynomial expansion)

That is,

$$g(x) := \frac{f_\pi(x) - f_\eta(x)}{\phi_d(x)} = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \frac{\Delta_{\mathbf{k}}}{\sqrt{\mathbf{k}!}} h_{\mathbf{k}}(x),$$

where

$$\Delta_{\mathbf{k}} = \int_{\mathbb{R}^d} \theta^{\mathbf{k}} \, d(\pi - \eta)(\theta).$$

is the difference in $\mathbf{k}$*-th moments between* π *and* η*.*

- By Parseval's theorem, $\|g\|_{L^2(\phi_d)}^2 = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \frac{\Delta_{\mathbf{k}}^2}{\mathbf{k}!}$.

Extremal Problem

Recall that we want to bound $\|g\|_{L^2(\phi_d)}$ from above by $\|g\|_{L^1(\phi_d)}$. What if we consider a proper subspace of $L^2(\phi_d)$?

Definition (Extremal problem)

Define

$$c_{n,d} := \inf \left\{ \|P\|_{L^1(\phi_d)} : P \in \Pi_n^d, \|P\|_{L^2(\phi_d)} = 1 \right\}, \qquad (5)$$

where Π_n^d is the set of real polynomials of total degree $\leq n$ in d variables.

- $\{h_{\mathbf{k}} : \mathbf{k} \in \mathbb{N}_0^d, |\mathbf{k}| \leq n\}$ is an orthonormal basis of the finite-dimensional subspace $\Pi_n^d \leq L^2(\phi_d)$.
- $|\mathbf{k}| := k_1 + \cdots + k_d$, $\dim \Pi_n^d = \binom{n+d}{n}$.

Proof Sketch of the Main Theorem ($L^2(\phi_d)$-norm)

We decompose $g = q + r$, where

$$q = \sum_{|\mathbf{k}|\le n} \frac{\Delta_{\mathbf{k}}}{\sqrt{\mathbf{k}!}} h_{\mathbf{k}} \in \Pi_n^d, \qquad r = \sum_{|\mathbf{k}|>n} \frac{\Delta_{\mathbf{k}}}{\sqrt{\mathbf{k}!}} h_{\mathbf{k}}.$$

Then,

$$\begin{aligned}
\|g\|_{L^1(\phi_d)} &\ge \|q\|_{L^1(\phi_d)} - \|r\|_{L^1(\phi_d)} \\
&\ge c_{n,d}\|q\|_{L^2(\phi_d)} - \|r\|_{L^2(\phi_d)} \\
&\ge \underbrace{c_{n,d}}_{e^{-O(n)}}\|g\|_{L^2(\phi_d)} - \underbrace{2\|r\|_{L^2(\phi_d)}}_{O\left(e^{-\frac{n\log n}{2}}\right)}. \qquad (\because c_{n,d}\le 1)
\end{aligned}$$

Hence, we choose

$$n \approx \frac{2\log(1/\|g\|_{L^1(\phi_d)})}{\log\log(1/\|g\|_{L^1(\phi_d)})}.$$

Lower Bound on $c_{n,d}$

We now proceed to prove $c_{n,d} = e^{-O(n)}$:

Proposition (Asymptotic lower bound on $c_{n,d}$)

Suppose $\kappa_1 > 1$. Then, there exists a constant $A_1 = A_1(\kappa_1)$, depending only on κ_1, such that, if $n \geq A_1 d$, then we have $c_{n,d} \geq 3e^{-\kappa_1 n}$. That is, for all $P \in \Pi_n^d$,

$$\|P\|_{L^1(\phi_d)} \geq 3e^{-\kappa_1 n} \|P\|_{L^2(\phi_d)} .$$

Lower Bound on $c_{n,d}$ (Two Ingredients)

Proposition (Nikolskii-type inequality)

For all $P \in \Pi_n^d$, we have

$$\sup_{x \in \mathbb{R}^d} \left| P(x)\phi_d^{1/2}(x) \right| \le \left(\frac{(n+d)^{n+d}}{n^n (2\pi d)^d} \right)^{1/4} \|P\|_{L^2(\mathbb{R}^d, \phi_d)} .$$

Proposition (Restricted-range inequality)

Write $E_{n,d} := 2n + d$. Suppose $\kappa > 1$. Then, there exists a constant $A = A(\kappa)$, depending only on κ, such that, if $E_{n,d} \ge Ad$, then, for all $P \in \Pi_n^d$, we have

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \le \frac{1}{2} \|P\|^2_{L^2(\mathbb{R}^d, \phi_d)} .$$

Lower Bound on $c_{n,d}$ (Combining the Ingredients)

For large enough n, every $P \in \Pi_n^d$ satisfies

$$\begin{aligned}
&\|P\|^2_{L^2(\mathbb{R}^d,\phi_d)} \\
&\le 2 \int_{\|x\|_2 \le \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \qquad \text{(by Restricted-range inequality)} \\
&\le 2 \sup_{\|x\|_2 \le \sqrt{2\kappa E_{n,d}}} \left|\phi_d^{-1/2}(x)\right| \sup_{x\in\mathbb{R}^d} \left|P(x)\phi_d^{1/2}(x)\right| \int_{\mathbb{R}^d} |P\phi_d| \\
&\le 2\left((2\pi)^{d/4} e^{\kappa E_{n,d}/2}\right)\left(\frac{(n+d)^{n+d}}{n^n(2\pi d)^d}\right)^{1/4} \|P\|_{L^2(\mathbb{R}^d,\phi_d)} \|P\|_{L^1(\mathbb{R}^d,\phi_d)} \\
&\qquad \text{(by Nikolskii-type inequality)} \\
&\le \frac{1}{3} e^{\kappa_1 n} \|P\|_{L^2(\mathbb{R}^d,\phi_d)} \|P\|_{L^1(\mathbb{R}^d,\phi_d)}. \qquad (1 < \kappa < \kappa_1)
\end{aligned}$$

Lower Bound on $c_{n,d}$ (Discussion)

Let us consider the meaning of the restricted-range inequality.

Proposition (Restricted-range inequality (Stronger version))

Write $E_{n,d} := 2n + d$. Suppose $\kappa > 1$. Then, there exist positive constants $A = A(\kappa)$ and $c = c(\kappa)$, depending only on κ, such that, if $E_{n,d} \geq Ad$, then, for all $P \in \Pi_n^d$, we have

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \leq e^{-cE_{n,d}} \|P\|^2_{L^2(\mathbb{R}^d, \phi_d)}.$$

- Statisticians: The Gaussian tail is sufficiently light to bound any polynomial within a restricted range: $\|x\|_2 \leq \sqrt{2\kappa E_{n,d}}$.

Lower Bound on $c_{n,d}$ (Discussion)

Let us consider the meaning of the restricted-range inequality.

Proposition (Restricted-range inequality (Stronger version))

Write $E_{n,d} := 2n + d$. Suppose $\kappa > 1$. Then, there exist positive constants $A = A(\kappa)$ and $c = c(\kappa)$, depending only on κ, such that, if $E_{n,d} \geq Ad$, then, for all $P \in \Pi_n^d$, we have

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \leq e^{-cE_{n,d}} \|P\|^2_{L^2(\mathbb{R}^d,\phi_d)}.$$

- Statisticians: The Gaussian tail is sufficiently light to bound any polynomial within a restricted range: $\|x\|_2 \leq \sqrt{2\kappa E_{n,d}}$.
- Physicists: In the semi-classical regime ($E_{n,d} \to \infty$), the probability mass in the classically forbidden region becomes negligible.

Lower Bound on $c_{n,d}$ (Discussion)

Let us consider the meaning of the restricted-range inequality.

Proposition (Restricted-range inequality (Stronger version))

Write $E_{n,d} := 2n + d$. Suppose $\kappa > 1$. Then, there exist positive constants $A = A(\kappa)$ and $c = c(\kappa)$, depending only on κ, such that, if $E_{n,d} \geq Ad$, then, for all $P \in \Pi_n^d$, we have

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \leq e^{-cE_{n,d}} \|P\|^2_{L^2(\mathbb{R}^d,\phi_d)}.$$

- Statisticians: The Gaussian tail is sufficiently light to bound any polynomial within a restricted range: $\|x\|_2 \leq \sqrt{2\kappa E_{n,d}}$.
- Physicists: In the semi-classical regime ($E_{n,d} \to \infty$), the probability mass in the classically forbidden region becomes negligible.
- $E_{n,d}$ is the energy; $\{\|x\|_2 > \sqrt{2E_{n,d}}\}$ is the forbidden region.

Physical Interpretation

- A classical Hamiltonian of a particle in $\mathbb{R}^d$:

$$\mathcal{H}_{\text{cl}} = \frac{1}{2}\|\xi\|_2^2 + V(x)$$
$$= \underbrace{\frac{1}{2}\|\xi\|_2^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\|x\|_2^2}_{\text{potential energy}}.$$

- Given the energy level E, $\|x\|_2 > \sqrt{2E}$ is forbidden in classical physics. We do not allow the kinetic energy to be negative.

Quantum Harmonic Oscillator

- A classical Hamiltonian of a particle in $\mathbb{R}^d$:

$$\mathcal{H}_{\mathrm{cl}} = \frac{1}{2}\|\xi\|_2^2 + V(x)$$
$$= \underbrace{\frac{1}{2}\|\xi\|_2^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\|x\|_2^2}_{\text{potential energy}}.$$

Quantum Harmonic Oscillator

- A classical Hamiltonian of a particle in $\mathbb{R}^d$:

$$\begin{aligned}\mathcal{H}_{\mathrm{cl}} &= \frac{1}{2}\|\xi\|_2^2 + V(x) \\ &= \underbrace{\frac{1}{2}\|\xi\|_2^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\|x\|_2^2}_{\text{potential energy}} .\end{aligned}$$

- The quantum-mechanical analog:

$$\begin{aligned}\mathcal{H} &= -\frac{\hbar^2}{2}\nabla^2 + V \\ \psi &\mapsto -\frac{\hbar^2}{2}\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}\right)\psi + \frac{1}{2}\left(x_1^2 + \cdots + x_d^2\right)\psi.\end{aligned}$$

Quantum Harmonic Oscillator

- A classical Hamiltonian of a particle in $\mathbb{R}^d$:

$$\mathcal{H}_{\mathrm{cl}} = \frac{1}{2}\|\xi\|_2^2 + V(x) = \underbrace{\frac{1}{2}\|\xi\|_2^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\|x\|_2^2}_{\text{potential energy}} .$$

- The quantum-mechanical analog:

$$\mathcal{H} = -\frac{\hbar^2}{2}\nabla^2 + V$$
$$\psi \mapsto -\frac{\hbar^2}{2}\left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2}\right)\psi + \frac{1}{2}\left(x_1^2 + \cdots + x_d^2\right)\psi.$$

- $\hbar > 0$ is the Planck constant.

Schrödinger Equation

- Quantum harmonic oscillators ψ are characterized as the eigenvectors (eigenfunctions) of $\mathcal{H} = -\frac{\hbar^2}{2}\nabla^2 + V$:

$$\mathcal{H}\psi = E\psi.$$

Schrödinger Equation

- Quantum harmonic oscillators ψ are characterized as the eigenvectors (eigenfunctions) of $\mathcal{H} = -\frac{\hbar^2}{2}\nabla^2 + V$:

$$\mathcal{H}\psi = E\psi.$$

- $E > 0$, the energy level, is the corresponding eigenvalue.

Schrödinger Equation

- Quantum harmonic oscillators ψ are characterized as the eigenvectors (eigenfunctions) of $\mathcal{H} = -\frac{\hbar^2}{2}\nabla^2 + V$:

$$\mathcal{H}\psi = E\psi.$$

- $E > 0$, the energy level, is the corresponding eigenvalue.
- The above equation is known as (time-independent) Schrödinger equation.

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}}$.

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the **k**-th Hermite function,

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the $\mathbf{k}$-th Hermite function,
- where $h_{\mathbf{k}}$ is the $\mathbf{k}$-th Hermite polynomial in $\mathbb{R}^d$ and $E_{\mathbf{k}} = 2|\mathbf{k}| + d$ for state $\mathbf{k} \in \mathbb{N}_0^d$.

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}} \psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the $\mathbf{k}$-th Hermite function,
- where $h_{\mathbf{k}}$ is the $\mathbf{k}$-th Hermite polynomial in $\mathbb{R}^d$ and $E_{\mathbf{k}} = 2|\mathbf{k}| + d$ for state $\mathbf{k} \in \mathbb{N}_0^d$.
- $\|\psi_{\mathbf{k}}\|_{L^2(\mathbb{R}^d)} = \|h_{\mathbf{k}}\|_{L^2(\mathbb{R}^d, \phi_d)} = 1$.

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the $\mathbf{k}$-th Hermite function,
- where $h_{\mathbf{k}}$ is the $\mathbf{k}$-th Hermite polynomial in $\mathbb{R}^d$ and $E_{\mathbf{k}} = 2|\mathbf{k}| + d$ for state $\mathbf{k} \in \mathbb{N}_0^d$.
- $\|\psi_{\mathbf{k}}\|_{L^2(\mathbb{R}^d)} = \|h_{\mathbf{k}}\|_{L^2(\mathbb{R}^d, \phi_d)} = 1$.
- $\psi_{\mathbf{k}}^2(x)$ explains the spatial density of states (DOS) for $x \in \mathbb{R}^d$.

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}} \psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the **k**-th Hermite function,
- where $h_{\mathbf{k}}$ is the **k**-th Hermite polynomial in $\mathbb{R}^d$ and $E_{\mathbf{k}} = 2|\mathbf{k}| + d$ for state $\mathbf{k} \in \mathbb{N}_0^d$.
- $\|\psi_{\mathbf{k}}\|_{L^2(\mathbb{R}^d)} = \|h_{\mathbf{k}}\|_{L^2(\mathbb{R}^d, \phi_d)} = 1$.
- $\psi_{\mathbf{k}}^2(x)$ explains the spatial density of states (DOS) for $x \in \mathbb{R}^d$.
- Relation to the restricted-range inequality?

Schrödinger Equation (Solutions)

- If $\hbar = 2$, in particular, $\psi_{\mathbf{k}} = h_{\mathbf{k}} \phi_d^{1/2}$ satisfies $\mathcal{H}\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}}$.
- Note that $\psi_{\mathbf{k}}$ is called the **k**-th Hermite function,
- where $h_{\mathbf{k}}$ is the **k**-th Hermite polynomial in $\mathbb{R}^d$ and $E_{\mathbf{k}} = 2|\mathbf{k}| + d$ for state $\mathbf{k} \in \mathbb{N}_0^d$.
- $\|\psi_{\mathbf{k}}\|_{L^2(\mathbb{R}^d)} = \|h_{\mathbf{k}}\|_{L^2(\mathbb{R}^d, \phi_d)} = 1$.
- $\psi_{\mathbf{k}}^2(x)$ explains the spatial density of states (DOS) for $x \in \mathbb{R}^d$.
- Relation to the restricted-range inequality?

Proposition (Restricted-range inequality (Stronger version))

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) \leq e^{-cE_{n,d}} \|P\|_{L^2(\mathbb{R}^d, \phi_d)}^2 .$$

Lower Bound on $c_{n,d}$ (Proof 1/2)

Proof.

For $P \in \Pi_n^d$, write $P = \sum_{|\mathbf{k}|\le n} c_{\mathbf{k}} h_{\mathbf{k}}$ and

$$\int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} P^2(x)\phi_d(x) = \sum_{|\mathbf{k}|\le n}\sum_{|\mathbf{l}|\le n} c_{\mathbf{k}} M_{\mathbf{kl}} c_{\mathbf{l}} \le \operatorname{tr}(M)\sum_{|\mathbf{k}|\le n} c_{\mathbf{k}}^2 = \operatorname{tr}(M)\,\|P\|^2_{L^2(\mathbb{R}^d,\phi_d)},$$

where $M = (M_{\mathbf{kl}})_{\mathbf{k},\mathbf{l}}$ is a $(\dim \Pi_n^d) \times (\dim \Pi_n^d)$ p.s.d. matrix with

$$M_{\mathbf{kl}} := \int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} \underbrace{\psi_{\mathbf{k}}(x)\psi_{\mathbf{l}}(x)}_{h_{\mathbf{k}}(x)h_{\mathbf{l}}(x)\phi_d(x)}.$$

□

Lower Bound on $c_{n,d}$ (Proof 2/2)

Continued.

Recall that $E_{n,d} = 2n + d$ and that $E_{\mathbf{k}} = 2|\mathbf{k}| + d$. It hence suffices to prove the following inequality:

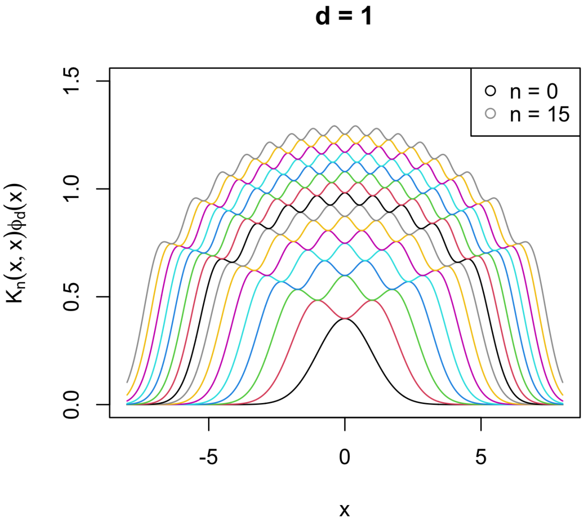
$$\mathrm{tr}(M) = \sum_{|\mathbf{k}| \le n} M_{\mathbf{kk}} = \int_{\|x\|_2 > \sqrt{2\kappa E_{n,d}}} \sum_{E_{\mathbf{k}} \le E_{n,d}} \psi_{\mathbf{k}}^2(x) \overset{?}{\le} e^{-cE_{n,d}}.$$

□

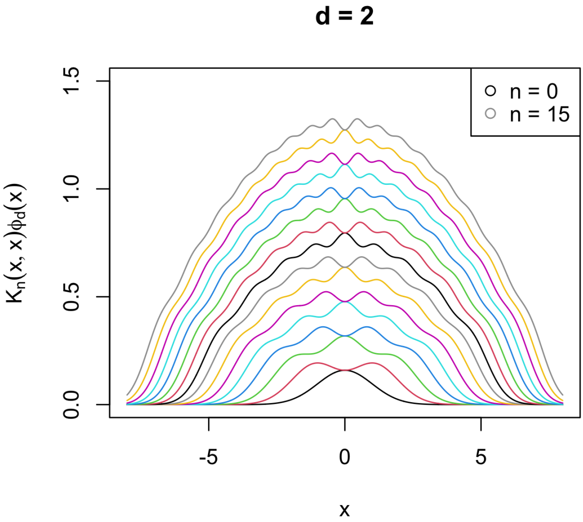
Proposition (Local Weyl Law [GS13])

$$\sum_{E_{\mathbf{k}} \le E_{n,d}} \psi_{\mathbf{k}}^2(x) \xrightarrow{E_{n,d} \to \infty} (4\pi)^{-d} \int_{\mathcal{H}_{\mathrm{cl}} \le E_{n,d}} d\xi$$
$$= (4\pi)^{-d} \omega_d \left(2E_{n,d} - \|x\|_2^2\right)^{d/2}.$$

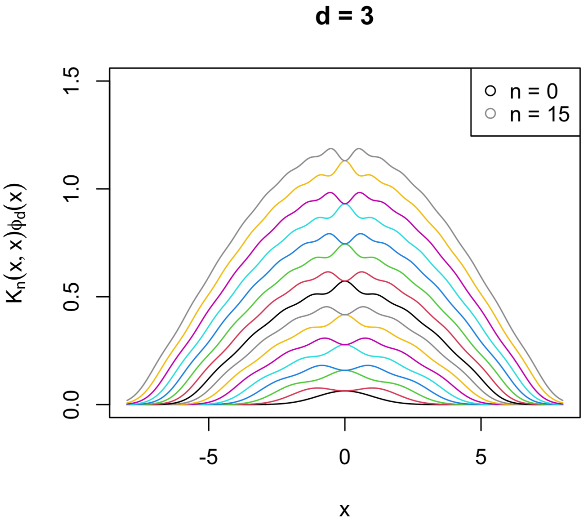
The Semicircle Law



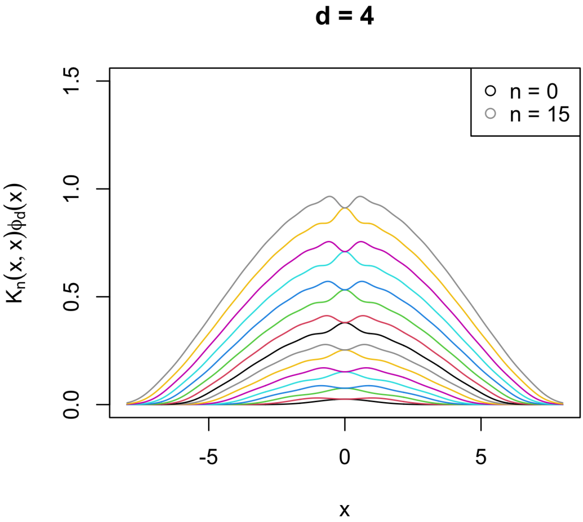
The Semicircle Law



The Semicircle Law



The Semicircle Law



Technical Details: Propagator and Density Matrix

- To find the explicit constant $c = c(\kappa)$, we consider:

Proposition (Mehler Kernel)

If we define the Mehler kernel by
$M(x, y; t) := \sum_{\mathbf{k}\in\mathbb{N}_0^d} e^{-tE_{\mathbf{k}}}\psi_{\mathbf{k}}(x)\psi_{\mathbf{k}}(y)$ *for* $t > 0$, *then*

$$M(x, y; t)$$
$$= (4\pi \sinh(2t))^{-d/2} \exp\left(-\frac{\|x\|_2^2 + \|y\|_2^2}{4\tanh(2t)} + \frac{\langle x, y\rangle_2}{2\sinh(2t)}\right).$$

- It suffices to prove this for $d = 1$ thanks to factorization.
- This is an exercise of bivariate Fourier transform (without any background knowlege in physics).
- Using bra-ket notation in physics, nonetheless, we may write

$$M(x, y; t) = \langle x|e^{-t\mathcal{H}}|y\rangle, \qquad e^{-t\mathcal{H}} = \sum_{\mathbf{k}} e^{-tE_k}|\psi_{\mathbf{k}}\rangle\langle\psi_{\mathbf{k}}|.$$

Sharpness

Theorem (Sharpness)

There exist two sequences of probability measures $\{\pi_n\}$ *and* $\{\eta_n\}$ *supported on* $[-M, M]$ *such that, if we define*

$$\mathrm{TV}_n := \mathrm{TV}(f_{\pi_n}, f_{\eta_n}), \qquad H_n := H(f_{\pi_n}, f_{\eta_n}),$$

then $\mathrm{TV}_n \downarrow 0$ *as* $n \to \infty$, *and moreover it holds for all* n *that* $\mathrm{TV}_n < e^{-e}$ *and that*

$$H_n \geq \mathrm{TV}_n^{1-\alpha^*(\mathrm{TV}_n)},$$

where we define

$$\alpha^*(t) := \frac{0.33}{\log\log(1/t)}, \qquad t > 0.$$

One-dimensional Result?

Suppose $\pi^\star$ and $\eta^\star$ are supported on $\mathbb{R}$. Then, we have $\mathrm{TV}(f_\pi, f_\eta) = \mathrm{TV}(f_{\pi^\star}, f_{\eta^\star})$ and $H(f_\pi, f_\eta) = H(f_{\pi^\star}, f_{\eta^\star})$ for

$$\pi = \pi^\star \otimes \delta_0^{\otimes(d-1)} = \pi^\star \otimes \delta_0 \otimes \cdots \otimes \delta_0,$$
$$\eta = \eta^\star \otimes \delta_0^{\otimes(d-1)} = \eta^\star \otimes \delta_0 \otimes \cdots \otimes \delta_0,$$

where δ_0 denotes the point mass at zero.

- Thus, the sharp example in one dimension immediately implies sharp examples in arbitrary dimensions.

Proof Sketch of the Sharpness

Recall the essential ingredients of the proof of the main theorem:

- The quantity $c_{n,d}$ can be bounded from below by $e^{-O(n)}$;

For $d = 1$, in particular, we note that the sequence of monomials $(x^n)_n$ is a sharp instance of the $c_n (= c_{n,1})$:

Proof Sketch of the Sharpness

Recall the essential ingredients of the proof of the main theorem:

- The quantity $c_{n,d}$ can be bounded from below by $e^{-O(n)}$;
- We can control the tail norm $\|r\|_{L^2(\phi_1)}$ by $e^{-\Omega(n \log n)}$.

For $d = 1$, in particular, we note that the sequence of monomials $(x^n)_n$ is a sharp instance of the $c_n (= c_{n,1})$:

Proof Sketch of the Sharpness

Recall the essential ingredients of the proof of the main theorem:

- The quantity $c_{n,d}$ can be bounded from below by $e^{-O(n)}$;
- We can control the tail norm $\|r\|_{L^2(\phi_1)}$ by $e^{-\Omega(n \log n)}$.

For $d = 1$, in particular, we note that the sequence of monomials $(x^n)_n$ is a sharp instance of the $c_n (= c_{n,1})$:

- The norm ratio $\|x^n\|_{L^1(\phi_1)} / \|x^n\|_{L^2(\phi_1)}$ is decaying exponentially in n.

Proof Sketch of the Sharpness

Recall the essential ingredients of the proof of the main theorem:

- The quantity $c_{n,d}$ can be bounded from below by $e^{-O(n)}$;
- We can control the tail norm $\|r\|_{L^2(\phi_1)}$ by $e^{-\Omega(n \log n)}$.

For $d = 1$, in particular, we note that the sequence of monomials $(x^n)_n$ is a sharp instance of the $c_n (= c_{n,1})$:

- The norm ratio $\|x^n\|_{L^1(\phi_1)} / \|x^n\|_{L^2(\phi_1)}$ is decaying exponentially in n.
- Knowing this fact, given n, we construct an example such that the $L^2(\phi_1)$ projection of $(f_{\pi_n} - f_{\eta_n})/\phi_1$ onto $\Pi_n (= \Pi_n^1)$ is proportional to x^n, i.e.,

$$\sum_{k=0}^{n} \frac{\Delta_k}{\sqrt{k!}} h_k(x) \propto x^n \iff \Delta_k \propto \begin{cases} \frac{1}{(n-k)!!}, & k \text{ is odd}, \\ 0, & k \text{ is even}, \end{cases}$$

provided that n is odd.

Robust Density Estimation

Definition

For a distribution class $\mathcal{P}$, its (global) Hellinger covering number is defined by

$$N_H(\mathcal{P}, \epsilon) := \min \{ N : \exists P_1, \ldots, P_N, \sup_{R \in \mathcal{P}} \inf_{1 \le i \le N} H(R, P_i) \le \epsilon \} .$$

The local Hellinger covering number of $\mathcal{P}$ is

$$N_{H,loc}(\mathcal{P}, \epsilon) := \sup_{P \in \mathcal{P}, \eta \ge \epsilon} N_H(B_H(P, \eta), \eta/2),$$

where $B_H(P, \eta) = \{R \in \mathcal{P} : H(P, R) \le \eta\}$.

Entropic Characterization of Learning in Hellinger

Let $\mathcal{P}_{M,d}$ be our distribution class (Gaussian mixtures with compactly supported mixing distributions).

Proposition (Learning Gaussian mixtures in Hellinger [JPW23])

Suppose $\mathcal{P}$ is a compact subset (in Hellinger) of $\mathcal{P}_{M,d}$. Let $\widehat{P} = \widehat{P}(X_1, \ldots, X_n)$ denote an estimator based on $X_1, \ldots, X_n$ drawn i.i.d. from $P \in \mathcal{P}$. Then,

$$\inf_{\widehat{P}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[H^2 \left(P, \widehat{P} \right) \right] \asymp \inf_{\widehat{P} \in \mathcal{P}} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[H^2 \left(P, \widehat{P} \right) \right] \asymp \epsilon_n^2,$$

where

$$\epsilon_n^2 \asymp \inf_{\epsilon > 0} \left(\epsilon^2 + \frac{1}{n} \log N_{H,loc}(\mathcal{P}, \epsilon) \right). \tag{6}$$

Entropic Characterization of Learning in Hellinger

- The upper bound follows from Le Cam-Birgé construction [LeC73, Bir83, Bir86].
- The lower bound follows from Fano's inequality and $H(f_\pi, f_\eta) \asymp \sqrt{\mathrm{KL}(f_\pi \| f_\eta)}$ under $\mathcal{P}_{M,d}$ [JPW23]:

$$\mathbb{P}\left[H\left(P, \widehat{P}\right) \geq \frac{\epsilon_n}{4}\right] \geq \frac{1}{2}.$$

- By triangular inequality and projection argument, we can restrict $\widehat{P}$ to be a proper estimator.

$$\mathbb{P}\left[\mathrm{TV}\left(P, \widehat{P}\right) \gtrsim \epsilon_n^{1+o(1)}\right] \geq \mathbb{P}\left[H\left(P, \widehat{P}\right) \geq \frac{\epsilon_n}{4}\right] \geq \frac{1}{2}.$$

Entropic Characterization of Learning in TV

Theorem (Learning Gaussian mixtures in total variation)

Under the same conditions as the above, we have

$$\begin{aligned}
\epsilon_n^{2\left(1+\frac{\Theta(1)}{\log(\log(1/\epsilon_n)\vee e)}\right)} &\lesssim \inf_{\widehat{P}} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\mathrm{TV}^2\left(P,\widehat{P}\right)\right] \\
&\asymp \inf_{\widehat{P}\in\mathcal{P}} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\mathrm{TV}^2\left(P,\widehat{P}\right)\right] \\
&\lesssim \epsilon_n^2,
\end{aligned}$$

where we define ϵ_n *as in* (6).

- The upper bound is trivial.

Robust Density Estimation

Proposition (Robust density estimation in TV)

Consider the data generating process as follows.

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} P := (1-\epsilon)P_{f_\pi} + \epsilon Q, \tag{7}$$

Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[\mathrm{TV}^2\left(f_\pi, \widehat{f}\right)\right] \lesssim \epsilon^2 + \frac{\log^{d+1}(n)}{n}, \tag{8}$$

where the expectation is under (7) and the supremum is taken over all Q and π such that $\mathrm{supp}(\pi) \subseteq [-M, M]^d$.

- By Yatracos' construction [Yat85].

Robust Density Estimation

Theorem (Minimax rate of robust density estimation)

Consider the data generating process as in (7). Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[H^2\left(f_\pi, \widehat{f}\right)\right] \lesssim \mathcal{E}^2(\epsilon, n), \tag{9}$$

where we define

$$\mathcal{E}^2(\epsilon, n) := \epsilon^{2\left(1-\frac{\Theta(1)}{\log(\log(1/\epsilon)\vee e)}\right)} + \left(\frac{1}{n}\right)^{1-o_d(1)}, \tag{10}$$

the expectation is under (7), and the supremum is taken over all Q and π such that $\mathrm{supp}(\pi) \subseteq [-M, M]^d$.

Robust Density Estimation

Theorem (Minimax rate of robust density estimation)

Consider the data generating process as in (7). Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[H^2\left(f_\pi, \widehat{f}\right)\right] \gtrsim \epsilon^{2\left(1-\frac{\Theta(1)}{\log(\log(1/\epsilon)\vee e)}\right)}, \tag{11}$$

where the expectation is under (7) and the supremum is taken over all Q and π such that $\mathrm{supp}(\pi) \subseteq [-M, M]^d$.

- The rate in (10) is minimax optimal in ϵ. It is, indeed, the exact minimax rate in the regime where $n \geq \mathrm{poly}(1/\epsilon)$.

Lemma ([CGR18])

Suppose P_1 and P_2 are probability measures such that $\mathrm{TV}(P_1, P_2) \leq \frac{\epsilon}{1-\epsilon}$. Then, there exist two probability measures Q_1 and Q_2 such that $(1-\epsilon)P_1 + \epsilon Q_1 = (1-\epsilon)P_2 + \epsilon Q_2$.

Robust Regret Bound

Theorem (Robust regret bound)

Consider the data generating process as follows.

$$X_i \sim (1-\epsilon)N(\theta_i, I_d) + \epsilon Q, \tag{12}$$
$$\theta_1, \ldots, \theta_n \overset{i.i.d.}{\sim} \pi.$$

Then, we have

$$\inf_{\widehat{f}} \sup_{\pi, Q} \mathbb{E}\left[\mathbb{E}_{X\sim f_\pi}\left\|\widehat{\theta}(X) - \widehat{\theta}^\star(X)\right\|^2\right] \lesssim \mathcal{E}^2(\epsilon, n),$$

where the outer expectation is under (12) *and the supremum is taken over all* Q *and* π *such that* $\mathrm{supp}(\pi) \subseteq [-M, M]^d$.

- Ispired by NPMLE papers [JZ09, SG20, SGS25].

Acknowledgements

- I thank Gemini 3 Pro for explaining about basics of quantum mechanics.

Acknowledgements

- I thank Gemini 3 Pro for explaining about basics of quantum mechanics.
- I thank Prof. Chao Gao for continued guidance.

Acknowledgements

- I thank Gemini 3 Pro for explaining about basics of quantum mechanics.
- I thank Prof. Chao Gao for continued guidance.
- I thank Prof. Nikolaos Ignatiadis for fruitful discussions on the implications of our results for empirical Bayes.

Acknowledgements

- I thank Gemini 3 Pro for explaining about basics of quantum mechanics.
- I thank Prof. Chao Gao for continued guidance.
- I thank Prof. Nikolaos Ignatiadis for fruitful discussions on the implications of our results for empirical Bayes.
- I thank you all for listening.

Acknowledgements

- I thank Gemini 3 Pro for explaining about basics of quantum mechanics.
- I thank Prof. Chao Gao for continued guidance.
- I thank Prof. Nikolaos Ignatiadis for fruitful discussions on the implications of our results for empirical Bayes.
- I thank you all for listening.
- I thank my wife.

References I

- Lucien Birgé, *Approximation dans les espaces métriques et théorie de l'estimation*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **65** (1983), no. 2, 181–237.
- ______, *On estimating a density using Hellinger distance and some other strange facts*, Probability theory and related fields **71** (1986), no. 2, 271–291.
- Mengjie Chen, Chao Gao, and Zhao Ren, *Robust covariance and scatter matrix estimation under Huber's contamination model*, The Annals of Statistics **46** (2018), no. 5, 1932–1960.
- Sanjoy Dasgupta, *Learning mixtures of Gaussians*, 40th annual symposium on foundations of computer science (Cat. No. 99CB37039), IEEE, 1999, pp. 634–644.

References II

Victor Guillemin and Shlomo Sternberg, *Semi-classical analysis*, International Press Boston, MA, 2013.

Subhashis Ghosal and Aad W Van Der Vaart, *Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities*, The Annals of Statistics (2001), 1233–1263.

Christopher R Genovese and Larry Wasserman, *Rates of convergence for the Gaussian mixture sieve*, The Annals of Statistics **28** (2000), no. 4, 1105–1127.

Zeyu Jia, Yury Polyanskiy, and Yihong Wu, *Entropic characterization of optimal rates for learning Gaussian mixtures*, The Thirty Sixth Annual Conference on Learning Theory, PMLR, 2023, pp. 4296–4335.

References III

Wenhua Jiang and Cun-Hui Zhang, *General maximum likelihood empirical Bayes estimation of normal means*, The Annals of Statistics (2009), 1647–1684.

Arlene KH Kim and Adityanand Guntuboyina, *Minimax bounds for estimating multivariate Gaussian location mixtures*, Electronic Journal of Statistics **16** (2022), no. 1, 1461–1484.

Lucien LeCam, *Convergence of estimates under dimensionality restrictions*, The Annals of Statistics (1973), 38–53.

Bruce G Lindsay, *Mixture models: Theory, geometry and applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, JSTOR, 1995, pp. i–163.

References IV

- Sujayam Saha and Adityanand Guntuboyina, *On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising*, The Annals of Statistics **48** (2020), no. 2, 738–762.
- Jake A Soloff, Adityanand Guntuboyina, and Bodhisattva Sen, *Multivariate, heteroscedastic empirical Bayes via nonparametric maximum likelihood*, Journal of the Royal Statistical Society Series B: Statistical Methodology **87** (2025), no. 1, 1–32.
- Wing Hung Wong and Xiaotong Shen, *Probability inequalities for likelihood ratios and convergence rates of sieve MLEs*, The Annals of Statistics (1995), 339–362.

References V

Yannis G Yatracos, *Rates of convergence of minimum distance estimators and Kolmogorov's entropy*, The Annals of Statistics **13** (1985), no. 2, 768–774.