

CONVERGENCE RATES FOR LATENT MIXING MEASURES IN INFINITE HOMOSCEDASTIC LOCATION-SCALE MIXTURE MODELS

BY NICOLA BARILETTO^{1,*[a](mailto:nicola.bariletto@utexas.edu)}, DUNG LE^{1,*[b](mailto:quangdung0110@utexas.edu)}, ALESSANDRO RINALDO^{1,[c](mailto:alessandro.rinaldo@austin.utexas.edu)} AND NHAT HO^{1,[d](mailto:minhnhat@utexas.edu)}

¹Department of Statistics and Data Sciences, The University of Texas at Austin, ^anicola.bariletto@utexas.edu; ^bquangdung0110@utexas.edu; ^calessandro.rinaldo@austin.utexas.edu; ^dminhnhat@utexas.edu

We study posterior contraction rates for mixing measures in homoscedastic location-scale mixture models with infinitely many components. While posterior convergence at the level of densities is well understood, ensuring convergence of the latent mixing measure is more challenging and has remained an open problem in settings where both location and scale parameters are unknown. We address this by deriving novel lower-bounds that connect the L^1 distance between mixture densities to discrepancies, based on the Wasserstein distances and the operator norm, between the underlying mixing measures and scale matrices. Our approach combines the dual formulation of the W_1 distance with functional-analytic approximation techniques. This leads to general inequalities, whose strength is determined (i) by the smoothness of the mixture kernel via the rate of decay of its characteristic function, and (ii) by a key lower-bound on the L^1 metric involving the operator norm discrepancy between scale parameters. Moreover, a novel PDE inversion condition yields a sharper inequality for important ordinary-smooth cases. We specialize these bounds to popular mixtures based on multivariate Gaussian, Cauchy, and Laplace kernels. As a consequence, we obtain first-of-their-kind contraction rates in the context of Dirichlet process mixtures with an unknown scale parameter shared across components. As a byproduct of our inequalities, we can distinguish the convergence behavior of the location mixing measure from that of the scale parameter across a range of kernel choices, leading to nuanced insights into their respective rates.

1. Introduction. Mixture modeling is a central tool in modern statistical analysis, due to its flexibility and the interpretability of the implied generative process. Concretely, one models data $X_1, \dots, X_n \in \mathbb{R}^d$ as arising from a density of the form

$$p_G(x) := \int_{\Theta} f(x | \theta) G(d\theta),$$

where G , a probability measure on the parameter space Θ , is known as the *mixing measure*, while $\{f(\cdot | \theta) : \theta \in \Theta\}$ is a family of probability density kernels indexed by Θ . The statistical problem, then, consists in estimating G and the resulting density p_G . A popular methodological choice involves modeling G as discrete, i.e., $G = \sum_{j=1}^k w_j \delta_{\theta_j}$ with $k \in \mathbb{N} \cup \{\infty\}$ and $(w_j)_{j=1}^k$ a sequence of positive weights summing to one. This formulation is appealing for two reasons. First, it is highly expressive: for instance, any density can be approximated in the L^1 sense by discrete mixtures with suitable kernels. Second, from a generative perspective, it induces a latent partition of the data into at most k groups $j = 1, \dots, k$, according to which component $f(\cdot | \theta_j)$ generates each observation, which provides a natural basis for clustering.

*Equal contribution.

MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62C10, 62G07, 60B10.

Keywords and phrases. Bayesian nonparametrics, Dirichlet process mixtures, Jackson kernel, Kantorovich–Rubinstein duality, McShane–Whitney extension, Optimal transport, PDE inversion, Wasserstein metrics.

Mixture models have received substantial attention in statistics and machine learning over the last decades, especially within the Bayesian community [21, 47] following the introduction of Bayesian nonparametric models that place flexible priors, such as the Dirichlet process [18], on discrete mixing measures with $k = \infty$ [19, 44, 51]. The appeal of the nonparametric approach is that it allows for flexible density estimation and clustering schemes, often making both the data partition and the underlying mixing measure the two primary objects of interest [52, 71]. This flexibility has also led to computational methods for handling the resulting infinite-dimensional structure [17, 45, 73], as well as to extensive theoretical analysis. This article offers several new contributions to such line of work, which we now briefly review.

1.1. *Related literature.* Much of the existing theory has focused on the asymptotic behavior of Bayesian mixtures for density estimation. To introduce the problem, define the L^1 distance $\|p - p'\|_1 := \int_{\mathbb{R}^d} |p(x) - p'(x)| dx$ between densities p and p' with respect to the Lebesgue measure on \mathbb{R}^d , and suppose that $X_i \stackrel{\text{i.i.d.}}{\sim} p_{G_0}$ for some unknown mixing measure G_0 . A natural question is whether, given a prior Π supported on a suitable space of mixing measures \mathcal{G} with $G_0 \in \mathcal{G}$, the posterior distribution

$$\Pi(A \mid X_1, \dots, X_n) := \frac{\int_A \prod_{i=1}^n p_G(X_i) \Pi(dG)}{\int_{\mathcal{G}} \prod_{i=1}^n p_{G'}(X_i) \Pi(dG')}$$

concentrates in L^1 neighborhoods of p_{G_0} as $n \rightarrow \infty$. This question has been studied extensively, both for fixed but arbitrarily small neighborhoods [leading to L^1 posterior consistency, 2, 4, 12, 25, 29, 43, 68, 72, 74, 77], and for shrinking neighborhoods, yielding L^1 posterior contraction rates ω_n satisfying

$$(1) \quad \Pi(G : \|p_G - p_{G_0}\|_1 \geq K\omega_n \mid X_1, \dots, X_n) \rightarrow 0$$

in $P_{G_0}^n$ -probability for all large enough $K > 0$ as $n \rightarrow \infty$; here, P_G^n denotes the n -fold product measure induced by $P_G(dx) := p_G(x) dx$, while $(\omega_n)_{n \in \mathbb{N}}$ is a positive vanishing sequence. Investigating sufficient conditions for posterior behavior like in Equation (1), as done by [7, 8, 24, 26–31, 62, 66, 69, 75], provides useful information on the asymptotic rate at which the posterior concentrates around the ground truth p_{G_0} , and yields guidance on model choices that support efficient density estimation.

However, L^1 contraction rates do not fully characterize posterior behavior in terms of clustering and contraction to the true mixing measure, since the evolution of the mixture components is a more refined phenomenon than L^1 convergence of the density. Several works have addressed this problem, focusing for instance on the recovery of k_0 , the true number of components [1, 49, 50], and on the allocation of data to redundant components in overfitted mixture models [60]. A prominent strand of literature, pioneered by [52], studies posterior contraction towards G_0 using the Wasserstein distances to measure discrepancy between mixing measures [see also 33, 35, 36, 61]. Concretely, if ρ is a metric on Θ , the corresponding Wasserstein distance of order $r \in [1, \infty)$ is defined as

$$W_r(G, G') := \left(\inf_{\pi \in \mathcal{P}(G, G')} \int_{\Theta \times \Theta} \rho(\theta, \theta')^r \pi(d\theta, d\theta') \right)^{1/r},$$

where $\mathcal{P}(G, G')$ is the set of probability measures π on $\Theta \times \Theta$ with marginals G and G' . Besides their solid grounding in the rich field of optimal transport [70], the appeal of the W_r metrics in the study of mixture models lies in their ability to capture discrepancies between mixing measures that may be discrete and have disjoint supports, a feature not shared by many classical probability metrics. These features have motivated significant recent interest

in optimal transport metrics within the Bayesian nonparametric community, for instance to measure dependence between random measures [9–11].

Convergence analyses for mixing measures with infinitely many atoms, such as those arising from a Dirichlet process prior, have been presented only for a restricted class of mixtures, namely *location mixtures*, where $f(x | \theta) = g(x - \theta)$ for $\theta \in \Theta \subseteq \mathbb{R}^d$ and some density kernel g . Location mixtures have important applications, for instance in deconvolution problems with additive noise of known scale, as studied in a Bayesian nonparametric framework by [23, 52, 61, 64]; see also [6, 14, 15] for non-Bayesian analyses using Wasserstein distances. However, restricting variation to a location parameter severely limits flexibility in learning the underlying density and mixing structure, often motivating the adoption of more complex model designs.

1.2. *Contributions of the article.* A more flexible and popular alternative to location mixtures are *location-scale mixtures* associated to families of kernels of the form

$$f(x | \theta, \Sigma) = |\Sigma|^{-1/2} g \left((x - \theta)^\top \Sigma^{-1} (x - \theta) \right),$$

where $\theta \in \mathcal{X} \subseteq \mathbb{R}^d$ is a location parameter, $\Sigma \in \mathbb{R}^{d \times d}$ is a positive-definite symmetric scale matrix (with $|\Sigma|$ denoting its determinant), and g is a typically symmetric kernel. Despite their prominence in applications, the posterior behavior of nonparametric mixing measures with a prior on both the location and scale parameters has, to the best of our knowledge, never been studied before. Our goal is to provide, for the first time, such an analysis for priors supported on mixing measures of the form

$$G(d\theta, dS) = P(d\theta) \times \delta_\Sigma(dS),$$

that is, *homoscedastic location-scale mixtures* in which the matrix Σ is shared across components and, unlike in previous work, estimated from data rather than fixed at the unknown value Σ_0 . Concretely, we aim to establish posterior contraction rates of the form

$$(2) \quad \Pi(G : \mathcal{D}(G, G_0) \geq K\varepsilon_n \mid X_1, \dots, X_n) \rightarrow 0$$

in $P_{G_0}^n$ -probability for all large enough $K > 0$ as $n \rightarrow \infty$, where $G_0 = P_0 \times \delta_{\Sigma_0}$ and \mathcal{D} is a meaningful discrepancy between mixing measures G and G_0 . To achieve this, after verifying an L^1 contraction rate ω_n as in Equation (1), a key step is to derive a lower-bound of the form $\|p_G - p_{G_0}\|_1 \gtrsim \mathcal{D}(G, G_0)$.¹ Following standard practice, we will design \mathcal{D} to incorporate W_1 as a discrepancy between P and P_0 (using the Euclidean ground metric on $x, y \in \mathcal{X}$) and the operator norm

$$\|S\|_{\text{op}} := \max_{v \in \mathbb{R}^d : \|v\| \leq 1} |v^\top S v|$$

to measure the distance between Σ and Σ_0 .

As mentioned above, virtually all existing work is limited to location mixtures only. In our notation, this is equivalent to fixing $\Sigma = \Sigma_0$ —hence assuming it known—and restricting attention to the location mixing measure P , so that $\mathcal{D}(G, G_0) \equiv W_1(P, P_0)$ and the inequality one seeks is $\|p_G - p_{G_0}\|_1 \gtrsim \varphi(W_1(P, P_0))$ for some increasing function $\varphi : [0, \infty) \rightarrow [0, \infty)$. Existing approaches, pioneered by [52], establish this kind of inequality by first convolving P and P_0 with a suitable smoothing kernel K_δ , and then exploiting the triangle inequality

$$(3) \quad W_1(P, P_0) \leq W_1(P, P * K_\delta) + W_1(P * K_\delta, P_0 * K_\delta) + W_1(P_0 * K_\delta, P_0)$$

¹Here and in the rest of the paper, $a(x) \lesssim b(x)$ and $a(x) \gtrsim b(x)$ denote inequalities up to a multiplicative constant, while $a(x) \asymp b(x)$ means $a(x) \lesssim b(x) \lesssim a(x)$. The variables x that these constants are independent of will be clear from the context or explicitly spelled out case by case.

to bound $W_1(P, P_0)$ by handling each term separately [23, 52, 61, 64].² While this strategy represents the state of the art in the location mixture setting, it fails in the more general location-scale case, as the notion of convolution has no obvious extension to bound discrepancies of the form $W_1(P \times \delta_\Sigma, P_0 \times \delta_{\Sigma_0})$ where different scales Σ and Σ_0 are involved.

To fill this gap, our first contribution is to develop new proof techniques that approach the problem from a functional analytic perspective. Our starting point is the observation that, for many location-scale mixtures of interest, an inequality of the form

$$(4) \quad \|p_G - p_{G_0}\|_1 \gtrsim \Psi(\|\Sigma - \Sigma_0\|_{\text{op}}),$$

for some strictly increasing function $\Psi : [0, \infty) \rightarrow [0, \infty)$, is available under mild parameter boundedness conditions. Moreover, the well-known Kantorovich–Rubinstein duality theorem [70] allows to rewrite

$$W_1(P, P_0) = \inf_{h \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} h(\theta) (P(d\theta) - P_0(d\theta)),$$

where $\text{Lip}_1(\mathcal{X}) := \{h \in \mathbb{R}^{\mathcal{X}} : \forall \theta, \theta' \in \mathcal{X}, |h(\theta) - h(\theta')| \leq \|\theta - \theta'\|\}$ and the infimum is attained at some $h^* \in \text{Lip}_1(\mathcal{X})$. Our strategy, then, is to develop a functional approximation of h^* that enables the use of Fourier transforms and the inversion theorem to analyze $W_1(P, P_0)$. In fact, since \mathcal{X} will be assumed to be a compact subset of \mathbb{R}^d , leveraging Fourier inversion requires to first extend h^* to a function $\tilde{h}_{\mathcal{X}}^*$ defined on the whole \mathbb{R}^d , and then to approximate $\tilde{h}_{\mathcal{X}}^*$ with another function $\tilde{h}_{\mathcal{X}, \Lambda}^*$, where $\Lambda > 0$ controls the approximation accuracy, that guarantees sufficient regularity. This constitutes a key technical step in our analysis.

With these approximations in place, we rewrite

$$\begin{aligned} W_1(P, P_0) &= \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta) (P(d\theta) - P_0(d\theta)) \\ &\quad + \int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta)) (P(d\theta) - P_0(d\theta)). \end{aligned}$$

The rest of our argument then proceeds in two steps:

1. we show that the difference $\int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta)) (P(d\theta) - P_0(d\theta))$ represents a controlled approximation error of order Λ^{-1} , then
2. we bound the integral $\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta) (P(d\theta) - P_0(d\theta))$, which is amenable to a Fourier-analytic treatment owing to the regularity of $\tilde{h}_{\mathcal{X}, \Lambda}^*$. In particular, a careful application of the Fourier inversion theorem, together with Equation (4), yields a general inequality of the form

$$(5) \quad \|p_G - p_{G_0}\|_1 \gtrsim \varphi(W_1(P, P_0)) + \Psi(\|\Sigma - \Sigma_0\|_{\text{op}}),$$

for a certain increasing function φ .

While the specific form of the function Ψ is a direct consequence of the requirement in Equation (4), φ can be characterized in terms of the properties of the density kernel f . In particular, φ is related to the smoothness of f measured by the rate of decay of its characteristic function. Our findings on this point, which we present in Section 3 and summarize in Table 1, are as follows:

²Here $*$ denotes the convolution operator, that is $P * K := \int_{\mathcal{X}} K(x - y)P(dy)$.

TABLE 1
Lower-bounds on the L^1 distance under different kernel smoothness settings. See Assumption 2 for a definition of Ξ and p .

Kernel smoothness	$\ p_G - p_{G_0}\ _1 \gtrsim$
Super-smooth ($\alpha > 0$) (Theorem 3.2)	$\min \{ \exp(-C/W_1(P, P_0)^\alpha), (\Psi \circ \Xi^{-1})(\exp(-C/W_1(P, P_0)^\alpha)) \}$ $+ \Psi(\ \Sigma - \Sigma_0\ _{\text{op}})$
Ordinary-smooth ($\beta > 0$) (Theorem 3.4)	$\min \{ W_1(P, P_0)^{d+\beta+1}, (\Psi \circ \Xi^{-1})(CW_1(P, P_0)^{d+p+1}) \}$ $+ \Psi(\ \Sigma - \Sigma_0\ _{\text{op}})$
PDE inversion ($\beta \in \mathbb{N}$) (Theorem 3.7)	$\min \{ W_1(P, P_0)^\beta, \Psi(CW_1(P, P_0)^\beta) \}$ $+ \Psi(\ \Sigma - \Sigma_0\ _{\text{op}})$

- (i) In the *super-smooth case of order* $\alpha > 0$, that is, when the characteristic function $\Phi_\Sigma(\xi)$ of $f(\cdot | 0, \Sigma)$ satisfies $\log \Phi_\Sigma(\xi) \asymp -\|\xi\|^\alpha$, we obtain

$$\varphi(t) = \min \{ \exp(-C/t^\alpha), (\Psi \circ \Xi^{-1})(\exp(-C/t^\alpha)) \}$$

for some constant $C > 0$, where $\Xi : [0, \infty) \rightarrow [0, \infty)$ is another function (introduced later) determined by f . This result is dimension-free and applies to important kernel families such as the Gaussian and Cauchy ones.

- (ii) In the *ordinary-smooth case of order* $\beta > 0$, where $\log \Phi_\Sigma(\xi) \asymp -\log(1 + \|\xi\|^\beta)$, we get that

$$\varphi(t) = \min \{ t^{d+\beta+1}, (\Psi \circ \Xi^{-1})(Ct^{d+p+1}) \},$$

for some constant $C > 0$, where $p \geq 1$ is determined by f and introduced later. In this case, the dependence on d and p reflects the mild regularity imposed by the ordinary-smoothness assumption and may be suboptimal for certain kernels such as the Laplace one, motivating the next result.

- (iii) In the presence of a *partial differential equation (PDE) inversion condition of order* $\beta \in \mathbb{N}$ (involving the existence of an appropriate differential operator \mathcal{T}_Σ such that $\mathcal{T}_\Sigma p_G = P$ for all $G = P \times \delta_\Sigma \in \mathcal{G}$), we obtain that

$$\varphi(t) = \min \{ t^\beta, \Psi(Ct^\beta) \},$$

for some constant $C > 0$. This novel PDE-based condition is notable because (a) it implies that the kernel f is ordinary-smooth of order β provided that β is even (Proposition 3.6), and (b) it yields a dimension-free inequality, replacing the $\beta + d + 1$ and $\beta + p + 1$ exponents of scenario (ii) with the more favorable exponent β . We exploit this condition explicitly in the analysis of Laplace mixtures, while noting that it also applies to other important mixture classes, such as those arising from shifted exponential and fixed-shape Gamma kernels.

The different forms of smoothness indicated above are verified in important and popular choices of the kernel; in Section 4, we verify these conditions for the Gaussian, Cauchy, and Laplace cases, with the resulting bounds summarized in Table 2. We also show that, in the Gaussian mixture setting, a sharper inequality is obtained with an isotropic scale $\Sigma = \sigma^2 I_d$ (where I_d denotes the $d \times d$ identity matrix), deriving from the unique semi-group structure of Gaussian kernel families. In turn, when combined with L^1 contraction results of the form in

TABLE 2
Lower-bounds on the L^1 distance for different kernel types.

Kernel type	$\ p_G - p_{G_0}\ _1 \gtrsim$
Gaussian (general scale, Theorem 4.1)	$\exp\left(-\frac{C \exp(\tilde{C}/W_1(P, P_0)^2)}{W_1(P, P_0)^2}\right) + \exp\left(-M \frac{\log(1/\ \Sigma - \Sigma_0\ _{\text{op}})}{\ \Sigma - \Sigma_0\ _{\text{op}}}\right)$
Gaussian (isotropic scale, Theorem 4.2)	$\exp\left(-\frac{C}{W_1(P, P_0)^2}\right) + \exp\left(-M \frac{\log(1/\ \Sigma - \Sigma_0\ _{\text{op}})}{\ \Sigma - \Sigma_0\ _{\text{op}}}\right)$
Cauchy (Theorem 4.3)	$\exp\left(-\frac{C}{W_1(P, P_0)}\right) + \ \Sigma - \Sigma_0\ _{\text{op}}^2$
Laplace (Theorem 4.4)	$W_1(P, P_0)^2 + \ \Sigma - \Sigma_0\ _{\text{op}}$

Equation (1), these bounds lead to posterior contraction rates for mixing measures and scale parameters in the corresponding infinite mixture models. A key feature of our approach is that, once an L^1 rate is available, the structure of the inequality in Equation (5) allows one to derive a W_1 rate for the mixing measure P and a separate $\|\cdot\|_{\text{op}}$ rate for the scale parameter Σ . This is made explicit in Propositions 5.1 and 5.2 in Section 5, which treat multivariate Gaussian and univariate Laplace mixtures under suitable Dirichlet process priors. For the Cauchy and multivariate Laplace kernels, instead, no explicit L^1 rate ω_n is currently available to the best of our knowledge, so that our inequalities yield rates for recovering P_0 and Σ_0 that are defined implicitly in terms of the unknown sequence ω_n (assuming the latter exists). We refer the reader to Section 5 for a detailed discussion of this issue for infinite Cauchy mixtures.

Our posterior contraction results, summarized in Table 3, shed new light on two important points. First, across all of the kernel types we consider, there is significant heterogeneity in posterior contraction behavior between the location mixing measure P and the scale parameter Σ , with the latter converging at substantially faster rates. Second, focusing on the kernel classes for which explicit L^1 rates are available, the comparison between Gaussian and Laplace mixtures shows that faster density estimation does not necessarily translate into faster recovery of the underlying parameters, as smoother kernels tend to blur more strongly the signal carried by P_0 and the variation in the scale parameter. This points to a practically relevant tradeoff: the choice of kernel should reflect whether the main objective is accurate density estimation or more structured tasks, such as clustering or parameter estimation, which depend directly on P and Σ .

To summarize, the main contributions of the article are as follows.

- Using the dual formulation of the W_1 metric, together with functional-analytic approximation techniques, we develop new tools to lower-bound the L^1 distance between homoscedastic location-scale mixture densities in terms of the W_1 distance between the underlying location mixing measures and the operator norm distance between the underlying scales. The resulting inequalities depend on the smoothness of the mixture kernel and, in the ordinary-smooth case, can be significantly sharpened by verifying a novel and widely applicable PDE inversion condition;
- We specialize these abstract inequalities to common kernel choices, obtaining explicit bounds for multivariate Gaussian, Cauchy, and Laplace mixtures;
- Combining these results with existing L^1 contraction theory, for the first time in the literature we derive posterior convergence rates for mixing measures and scale parameters in infinite homoscedastic location-scale mixture models with Dirichlet process priors, focusing on the Gaussian and Laplace cases.

TABLE 3

L^1 , W_1 and operator norm posterior contraction rates for Dirichlet process mixtures with different kernel types. For the Cauchy and multivariate Laplace cases, we express rates relative to an implicit L^1 rate ω_n .

Kernel type	$\ p_G - p_{G_0}\ _1$	$W_1(P, P_0)$	$\ \Sigma - \Sigma_0\ _{\text{op}}$
Gaussian (general scale) (Proposition 5.1)	$n^{-1/2}(\log n)^{(d+1)/2}$	$(\log \log n)^{-1/2}$	$(\log n)^{-1} \log \log n$
Gaussian (isotropic scale) (Proposition 5.1)	$n^{-1/2}(\log n)^{(d+1)/2}$	$(\log n)^{-1/2}$	$(\log n)^{-1} \log \log n$
Cauchy (implicit in Theorem 4.3)	ω_n	$(\log(1/\omega_n))^{-1}$	$\omega_n^{1/2}$
Laplace ($d = 1$) (Proposition 5.2)	$n^{-1/4}(\log n)^{5/4}$	$n^{-1/8}(\log n)^{5/8}$	$n^{-1/4}(\log n)^{5/4}$
Laplace ($d \geq 2$) (implicit in Theorem 4.4)	ω_n	$\omega_n^{1/2}$	ω_n

The rest of the article is organized as follows. Section 2 introduces notation and collects a few preliminary results. Section 3 develops abstract inequalities of the form in Equation (5) under different smoothness assumptions on the mixture kernel. Section 4 specializes these inequalities to commonly used kernels. Section 5 builds on the previous sections to derive posterior contraction rates for homoscedastic location-scale infinite mixture models with Dirichlet process priors. Section 6 concludes. Complete proofs are provided in the supplementary material [3].

2. Notation and preliminaries. Throughout the article, we assume that a sample of \mathbb{R}^d -valued observations X_1, \dots, X_n is available and drawn i.i.d. from an underlying mixture distribution P_{G_0} (for $G_0 = P_0 \times \delta_{\Sigma_0}$) with density

$$p_{G_0}(x) = \int_{\mathcal{X}} f(x | \theta, \Sigma_0) P_0(d\theta), \quad x \in \mathbb{R}^d,$$

with respect to the Lebesgue measure on \mathbb{R}^d (endowed with its natural Borel σ -algebra $\mathcal{B}(\mathbb{R}^d)$). The specific form of the kernel f will vary throughout, with

$$\Phi_{\Sigma}(\xi) := \int_{\mathbb{R}^d} e^{i\xi^\top x} f(x | 0, \Sigma) dx, \quad \xi \in \mathbb{R}^d,$$

representing the associated characteristic function; with a slight abuse of notation, we also write $\Phi_{\mu}(\xi) := \int_{\mathbb{R}^d} e^{i\xi^\top x} \mu(dx)$ for any probability measure μ on \mathbb{R}^d . Relatedly, for any integrable function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, we denote its Fourier transform by

$$\mathcal{F}[h](\xi) := \int_{\mathbb{R}^d} e^{-i\xi^\top x} h(x) dx, \quad \xi \in \mathbb{R}^d.$$

Prior and posterior distributions, which we have already defined and denoted respectively by $\Pi(\cdot)$ and $\Pi(\cdot | X_1, \dots, X_n)$, are formally understood as probability measures on the space

of absolutely continuous probabilities on \mathbb{R}^d , endowed with the topology of weak convergence and the corresponding Borel σ -algebra. However, since the priors we consider are supported on specific subclasses of mixture distributions, it is more convenient to view both the prior and the posterior as measures on the space of mixing measures generating such mixtures.

In particular, the priors Π we consider concentrate on homoscedastic location-scale mixture densities arising from mixing measures of the form $G = P \times \delta_\Sigma$. For all such G in the support of the prior (including the unknown G_0), we work under the standing assumption that there exist constants $R, \lambda_{\min}, \lambda_{\max} > 0$, with $\lambda_{\min} < \lambda_{\max}$, and a measurable set $\mathcal{X} \subseteq \bar{B}(0, R) := \{x \in \mathbb{R}^d : \|x\| \leq R\}$,³ such that

$$\text{supp}(P) \subseteq \mathcal{X}, \quad \Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max}),$$

where $\text{supp}(P)$ denotes the support of P , and $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ is the set of symmetric positive-definite matrices with eigenvalues bounded below by λ_{\min} and above by λ_{\max} . Therefore, the parameter space of interest is

$$\mathcal{G} := \{G = P \times \delta_\Sigma : \text{supp}(P) \subseteq \mathcal{X}, \Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})\},$$

and we write $\mathcal{P} := \{p_G : G \in \mathcal{G}\}$. Throughout the text, we assume that the prior on \mathcal{G} is of product form across its coordinates P and Σ .

Our main interest is in the popular modeling strategy in which P is assigned a Dirichlet process prior [18], written $P \sim \text{DP}(a, H)$, where $a > 0$ is a concentration parameter and H is a diffuse base probability measure on \mathbb{R}^d ; the scale Σ , instead, is modeled independently of P with a prior having full support on $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$. In particular, for any finite measurable partition A_1, \dots, A_k of \mathbb{R}^d , this prior structure implies that the vector $(P(A_1), \dots, P(A_k))$ has a Dirichlet distribution with parameters $(aH(A_1), \dots, aH(A_k))$, and if $\text{supp}(H) = \mathcal{X}$, then the support of Π coincides with the set of mixing measures $G = P \times \delta_\Sigma$ with $\Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ and $\text{supp}(P) \subseteq \mathcal{X}$ [46].

A key feature of the Dirichlet process is the implied stick-breaking representation [65] for $P \sim \text{DP}(a, H)$:

$$P = \sum_{j \in \mathbb{N}} w_j \delta_{\theta_j}, \quad w_1 = B_1, \quad w_j = B_j \prod_{\ell=1}^{j-1} (1 - B_\ell), \quad j \geq 2,$$

where $B_j \stackrel{\text{i.i.d.}}{\sim} \text{Beta}(1, a)$ and $\theta_j \stackrel{\text{i.i.d.}}{\sim} H$. This yields the generative mechanism

$$X_1, \dots, X_n \mid G \stackrel{\text{i.i.d.}}{\sim} p_G, \quad p_G = \sum_{j \in \mathbb{N}} w_j f(\cdot \mid \theta_j, \Sigma),$$

which can be interpreted as a two-stage procedure (once G is fixed): first, latent mixture component assignments $c_i \in \mathbb{N}$ are drawn according to the discrete measure $\sum_{j \in \mathbb{N}} w_j \delta_j$ derived from P , and then X_i is drawn from $f(\cdot \mid \theta_{c_i}, \Sigma)$. This induces a natural partition of the data, where X_i and X_j are assigned to the same cluster whenever $c_i = c_j$. Clearly, this rich generative structure is governed by the mixing measure G , which is the main motivation driving the study of posterior contraction towards the unknown G_0 .

As already discussed, a natural strategy is to deduce contraction around G_0 from L^1 contraction around p_{G_0} , leveraging the extensive literature on Bayesian nonparametric density

³Denote by $B(0, R)$ the corresponding open ball.

estimation. To make this precise, we recall a general result (an immediate adaptation of Theorem 2.1 in [26]) giving sufficient conditions for L^1 contraction. Define

$$V_1(p_{G_0}, p_G) := \int_{\mathbb{R}^d} \log \left(\frac{p_{G_0}(x)}{p_G(x)} \right) p_{G_0}(x) dx,$$

$$V_2(p_{G_0}, p_G) := \int_{\mathbb{R}^d} \left(\log \frac{p_{G_0}(x)}{p_G(x)} \right)^2 p_{G_0}(x) dx,$$

and, for any $\omega > 0$ and $\mathcal{A} \subseteq \mathcal{P}$, let $D(\omega, \mathcal{A}, \|\cdot\|_1)$ denote the L^1 ω -packing number of \mathcal{A} , i.e., the largest number of densities in \mathcal{A} that are at least ω apart from each other in L^1 distance.

THEOREM 2.1. *Assume that, for a positive sequence $(\omega_n)_{n \in \mathbb{N}}$ with $\lim_{n \rightarrow \infty} \omega_n = 0$ and $\lim_{n \rightarrow \infty} n\omega_n^2 = \infty$, there exist sets $\mathcal{G}_n \subseteq \mathcal{G}$ and a constant $C > 0$ such that*

$$\log D(\omega_n, \{p_G : G \in \mathcal{G}_n\}, \|\cdot\|_1) \leq n\omega_n^2,$$

$$\Pi(\mathcal{G} \setminus \mathcal{G}_n) \leq \exp(-(C+4)n\omega_n^2),$$

$$\Pi(G : V_1(p_{G_0}, p_G) \leq \omega_n^2, V_2(p_{G_0}, p_G) \leq \omega_n^2) \geq \exp(-Cn\omega_n^2)$$

for all large enough $n \in \mathbb{N}$. Then

$$\Pi(G : \|p_G - p_{G_0}\|_1 \geq K\omega_n \mid X_1, \dots, X_n) \rightarrow 0$$

in $P_{G_0}^n$ -probability for all large enough $K > 0$ as $n \rightarrow \infty$.

The conditions of the theorem are standard: they ensure an L^1 rate by requiring that (i) complex densities receive negligible prior mass, and (ii) enough mass is placed near the true density in the Kullback–Leibler sense. On the other hand, as implied by Lemma 1 in [52], under mild continuity assumptions on $(\theta, \Sigma) \mapsto f(\cdot \mid \theta, \Sigma)$ one has

$$\|p_G - p_{G'}\|_1 \lesssim W_1(G, G')$$

for discrete mixing measures $G, G' \in \mathcal{G}$.⁴ Thus, convergence of mixing measures implies convergence of densities, but the converse is not necessarily true. Establishing a reverse inequality

$$\|p_G - p_{G'}\|_1 \gtrsim \mathcal{D}(G, G'),$$

of the form in Equation (5) is however essential to deduce posterior contraction around G_0 from L^1 results such as Theorem 2.1, and identifying conditions under which such inverse bounds hold has been a central question since [52]. The main goal of this article is to address this problem in the homoscedastic location-scale setting where $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$. We do so by first deriving general inverse bounds under different smoothness assumptions on f (in Section 3), while in Sections 4 and 5 we specialize these results to Gaussian, Cauchy, and Laplace kernels, and combine them with existing L^1 contraction results (typically obtained via arguments related to Theorem 2.1) to derive corresponding W_1 posterior contraction rates.

⁴Using $\rho((x, \Sigma), (x', \Sigma')) = \|x - x'\| + \|\Sigma - \Sigma'\|_{\text{op}}$ as a ground metric on the product space $\Theta = \mathcal{X} \times \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$.

3. Inverse Bounds for Mixing Measure Discrepancies. In this section, we derive lower-bounds for the L^1 distance between mixture densities p_G and $p_{G'}$, where $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$, in terms of suitable functions of the W_1 distance between the mixing measures P and P' , and the operator norm distance between the scale matrices Σ and Σ' .

We begin by stating two assumptions that will be used throughout.

ASSUMPTION 1. There exists a strictly increasing function $\Psi : [0, \infty) \rightarrow [0, \infty)$ such that, for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} ,

$$\|p_G - p_{G'}\|_1 \geq C\Psi(\|\Sigma - \Sigma'\|_{\text{op}}),$$

for a positive quantity C that may depend on R , λ_{\min} , λ_{\max} , and d .

The specific form of Ψ depends on the choice of kernel, but the requirement itself is mild: since the location space \mathcal{X} is compact, a small L^1 distance between densities is expected to imply closeness of the corresponding scale parameters. This will be verified for several important kernel families in Section 4.

ASSUMPTION 2. There exist an exponent $p \geq 1$ and a strictly increasing function $\Xi : [0, \infty) \rightarrow [0, \infty)$ such that, for all $\xi \in \mathbb{R}^d$ and $\Sigma, \Sigma' \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$,

$$|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| \leq C\Xi(\|\Sigma - \Sigma'\|_{\text{op}})\|\xi\|^p \max\{|\Phi_{\Sigma'}(\xi)|, |\Phi_\Sigma(\xi)|\},$$

for a positive quantity C that may depend on R , λ_{\min} , λ_{\max} , and d .

Assumption 2 requires the characteristic function associated with the kernel $f(\cdot | 0, \Sigma)$ to vary in a controlled manner with respect to Σ . In particular, the dependence is governed by Ξ , with a degree- p polynomial modulation in $\|\xi\|$ and a linear one in the magnitude of the characteristic function itself.

We are now in a position to state our main results. Towards that goal, we will consider three settings separately: when the kernel f is super-smooth, when it is ordinary smooth, and when the mixture satisfies a certain PDE inversion condition.

3.1. Super-smooth kernel. We first consider the case in which the mixture kernel f is super-smooth. For the next definition, recall that $\Phi_\Sigma(\xi)$ denotes the characteristic function, evaluated at ξ , associated with the density $f(\cdot | 0, \Sigma)$.

DEFINITION 3.1. The density kernel f is *super-smooth of order* $\alpha > 0$ if

$$\exp(-c_1\|\xi\|^\alpha) \leq |\Phi_\Sigma(\xi)| \leq \exp(-c_2\|\xi\|^\alpha)$$

for all $\xi \in \mathbb{R}^d$ and $\Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$, where c_1 and c_2 are constants depending only on λ_{\min} , λ_{\max} , and d .

That is, super-smoothness requires an exponential decay of $\Phi_\Sigma(\xi)$ as a function of $\|\xi\|$. This condition is satisfied by several commonly used kernels, including the Gaussian and Cauchy kernels considered in Section 4. Combined with Assumptions 1 and 2, super-smoothness yields the following inverse bound.

THEOREM 3.2. Under Assumptions 1 and 2, if the kernel f is super-smooth of order α , there exist positive quantities C, \tilde{C}, C' , depending on R , λ_{\min} , λ_{\max} , p , and d , such that

$$(6) \quad \|p_G - p_{G'}\|_1 \geq C' \left[\Psi(\|\Sigma - \Sigma'\|_{\text{op}}) + \min \left\{ \exp \left(-\frac{C}{W_1(P, P')^\alpha} \right), (\Psi \circ \Xi^{-1}) \left(\tilde{C} \exp \left(-\frac{C}{W_1(P, P')^\alpha} \right) \right) \right\} \right]$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} .

The bound in Theorem 3.2 depends on $\|\Sigma - \Sigma'\|_{\text{op}}$ only through the function Ψ , which naturally aligns with Assumption 1. In contrast, the dependence on $W_1(P, P') \equiv w$ is more complex, entering through terms of the form $\exp(-C/w^\alpha)$ that are possibly mediated by the composition $\Psi \circ \Xi^{-1}$. Furthermore, the lower-bound deteriorates as the smoothness level α increases. From a deconvolution perspective, this is intuitive: a smoother kernel makes the recovery of the underlying mixing measure statistically more challenging, as reflected by a looser bound.

Beyond the lower-bound results and their implications for statistical estimation of mixture models, a central contribution of this article is the development of new techniques to derive such bounds, based on functional approximations in the dual formulation of W_1 and tools from Fourier analysis and PDEs. Therefore, we now outline the main steps of the argument, deferring the full proof to the supplementary material.

PROOF SKETCH OF THEOREM 3.2. The proof consists of five steps. We first use the Kantorovich–Rubinstein duality theorem to represent $W_1(P, P')$ as an integral of an optimal 1-Lipschitz function h^* , which we extend to all of \mathbb{R}^d via a McShane–Whitney argument. We then construct a spectrally truncated approximation of h^* using a Jackson kernel, controlling the resulting approximation error. Fourier inversion and Fubini’s theorem allow us to express the integral in terms of characteristic functions, which we split into two contributions via the triangle inequality. Each contribution is then bounded using the super-smoothness assumption and Assumptions 1 and 2. Optimizing over the truncation parameter and rearranging finally yields the claimed inequality. Below is a more detailed discussion of each step.

Step 1: McShane–Whitney extension. By the Kantorovich–Rubinstein duality theorem [70], for every P and P' there exists $h^* \in \text{Lip}_1(\mathcal{X})$ attaining

$$W_1(P, P') = \int_{\mathcal{X}} h^*(\theta) (P(d\theta) - P'(d\theta)).$$

Replacing h^* with $h^*(\cdot) - h^*(0)$ leaves the integral unchanged and ensures $|h^*(x)| \leq \|x\|$ for all $x \in \mathcal{X}$. Our goal is to analyze $W_1(P, P')$ via Fourier inversion, which requires extending h^* to all of \mathbb{R}^d and working with a sufficiently regular approximation thereof. We first extend h^* to a globally 1-Lipschitz continuous function on \mathbb{R}^d via the McShane–Whitney infimal convolution [34, 48, 76]

$$\bar{h}_{\mathcal{X}}^*(x) := \inf_{y \in \mathcal{X}} \{h^*(y) + \|x - y\|\},$$

and then multiply it by a compactly supported and infinitely differentiable cutoff $\eta \in C_c^\infty(\mathbb{R}^d)$ satisfying $\eta(x) = 1$ for $x \in \bar{B}(0, R)$ and $\eta(x) = 0$ for $x \notin B(0, 2R)$ [39]. The resulting function $\tilde{h}_{\mathcal{X}}^*$ belongs to $L^1(\mathbb{R}^d)$, is globally $(1 + 2RM_\eta)$ -Lipschitz (where $M_\eta := \sup_{x \in \mathbb{R}^d} \|\nabla \eta(x)\| < \infty$), and preserves the property

$$W_1(P, P') = \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta) (P(d\theta) - P'(d\theta)).$$

Step 2: Spectral truncation. To apply Fourier inversion, we convolve $\tilde{h}_{\mathcal{X}}^*$ with a rescaled Jackson kernel $K_\Lambda(x) := \Lambda^d K(\Lambda x)$, whose Fourier transform is supported on $\bar{B}(0, \Lambda)$, to obtain $\tilde{h}_{\mathcal{X}, \Lambda}^* := \tilde{h}_{\mathcal{X}}^* * K_\Lambda$ [5, 16]. The Lipschitz continuity of $\tilde{h}_{\mathcal{X}}^*$ implies the approximation bound

$$\|\tilde{h}_{\mathcal{X}, \Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty := \sup_{x \in \mathbb{R}^d} |\tilde{h}_{\mathcal{X}, \Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x)| \leq (1 + 2RM_\eta)M_K/\Lambda,$$

where $M_K := \int_{\mathbb{R}^d} \|x\| K(x) dx < \infty$. Combined with $\sup_{A \in \mathcal{B}(\mathbb{R}^d)} |P(A) - P'(A)| \leq 2$, this gives

$$(7) \quad W_1(P, P') \leq \frac{C_0}{\Lambda} + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta) (P(d\theta) - P'(d\theta))$$

for a constant $C_0 > 0$ depending only on R and f .

Step 3: Fourier inversion. Since $\tilde{h}_{\mathcal{X}, \Lambda}^* \in L^1(\mathbb{R}^d)$ and $\mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*]$ is supported in $\bar{B}(0, \Lambda)$ and bounded, the Fourier inversion and Fubini's theorems yield

$$\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta) (P(d\theta) - P'(d\theta)) = \frac{1}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*](\xi) (\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi.$$

Writing $\Phi_P(\xi) - \Phi_{P'}(\xi) = \Phi_{p_G}(\xi)/\Phi_\Sigma(\xi) - \Phi_{p_{G'}}(\xi)/\Phi_{\Sigma'}(\xi)$ and applying the triangle inequality, the right-hand side is bounded above by C_R times

$$(8) \quad \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi + \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi,$$

where $C_R > 0$ depends only on R and d .

Step 4: Bounding via super-smoothness. For the first term in Equation (8), we use $|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)| \leq \|p_G - p_{G'}\|_1$ and the lower-bound $|\Phi_\Sigma(\xi)| \geq \exp(-c_1 \|\xi\|^\alpha)$ from the super-smoothness assumption to get

$$\int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \|p_G - p_{G'}\|_1 \cdot \Lambda^d \exp(c_1 \Lambda^\alpha).$$

For the second term, Assumption 2 and the super-smoothness bounds together yield

$$\int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \cdot \Lambda^{d+p} \exp((c_1 - c_2) \Lambda^\alpha).$$

Step 5: Conclusion. Substituting these bounds into Equation (7) and using Assumption 1 to replace $\Xi(\|\Sigma - \Sigma'\|_{\text{op}})$ with $(\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1)$, we obtain

$$W_1(P, P') \lesssim \frac{C_0}{\Lambda} + \|p_G - p_{G'}\|_1 \cdot \Lambda^d e^{c_1 \Lambda^\alpha} + (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1) \cdot \Lambda^{d+p} e^{(c_1 - c_2) \Lambda^\alpha}.$$

Setting $\Lambda \asymp W_1(P, P')^{-1}$ and applying the elementary bound $t^{-k} \lesssim \exp(t^{-\alpha})$ for any $k, \alpha > 0$ and $t \in (0, 1]$, we rearrange to obtain that either

$$\|p_G - p_{G'}\|_1 \gtrsim \exp\left(-\frac{C}{W_1(P, P')^\alpha}\right),$$

or

$$\|p_G - p_{G'}\|_1 \gtrsim (\Psi \circ \Xi^{-1})\left(\exp\left(-\frac{C}{W_1(P, P')^\alpha}\right)\right),$$

which together with Assumption 1 give Equation (6). \square

3.2. Ordinary-smooth kernel. The second setting we consider involves an ordinary-smooth kernel f , which is formalized as follows.

DEFINITION 3.3. The density kernel f is *ordinary-smooth of order* $\beta > 0$ if

$$c_1(1 + \|\xi\|^\beta)^{-1} \leq |\Phi_\Sigma(\xi)| \leq c_2(1 + \|\xi\|^\beta)^{-1}$$

for all $\xi \in \mathbb{R}^d$ and $\Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$, where c_1 and c_2 are constants depending on λ_{\min} , λ_{\max} , and d .

Unlike in the super-smooth case, ordinary-smooth kernels are characterized by a polynomial decay of the characteristic function, corresponding to a substantially lower degree of smoothness. This leads to the following inverse bound.

THEOREM 3.4. *Under Assumptions 1 and 2, if the kernel f is ordinary-smooth of order β , there exist positive constants C and C' , depending on R , λ_{\min} , λ_{\max} , p , and d , such that*

$$(9) \quad \begin{aligned} \|p_G - p_{G'}\|_1 &\geq C' \left[\Psi(\|\Sigma - \Sigma'\|_{\text{op}}) \right. \\ &\left. + \min \left(W_1(P, P')^{d+\beta+1}, (\Psi \circ \Xi^{-1})(CW_1(P, P')^{d+p+1}) \right) \right] \end{aligned}$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} .

The structure of the bound in Theorem 3.4 is similar to its super-smooth counterpart in Theorem 3.2, in terms of the role played by the functions Ψ and Ξ in mediating the dependence on $\|\Sigma - \Sigma'\|_{\text{op}}$ and $W_1(P, P')$. However, the W_1 component of the bound is significantly improved under the ordinary-smooth assumption: the inverse-exponential dependence in the super-smooth case is replaced by a polynomial one, with β worsening the rate as it increases. A drawback of Theorem 3.4, however, is the explicit dependence on the dimension d , which may inflate the polynomial degree. Intuitively, this reflects the fact that the ordinary-smooth assumption imposes only mild conditions on the tail behavior of the characteristic function of f , whereas in many cases of interest (e.g., Laplace mixtures, as shown in Section 4) the mixture possesses useful additional structure. This motivates, in the next subsection, turning to a more refined condition to obtain dimension-free bounds.

PROOF SKETCH OF THEOREM 3.4. The proof follows the same structure as that of Theorem 3.2, with Steps 1–3 carrying over verbatim. The key differences arise in Step 4, where the polynomial decay of the characteristic function under ordinary smoothness leads to polynomial rather than exponential bounds, and in Step 5, where the choice of truncation parameter and the subsequent rearrangement differ accordingly. We therefore focus on these two steps.

Step 4': Bounding via ordinary smoothness. For the first term in Equation (8), we use $|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)| \leq \|p_G - p_{G'}\|_1$ and the lower-bound $|\Phi_\Sigma(\xi)| \geq (1 + c_1 \|\xi\|^\beta)^{-1}$ from the ordinary-smoothness assumption to get

$$\int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \|p_G - p_{G'}\|_1 \cdot \Lambda^{d+\beta}.$$

For the second term, Assumption 2 gives

$$|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \|\xi\|^p \max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|).$$

Under ordinary smoothness, the ratio $\max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|)/|\Phi_\Sigma(\xi)|$ is bounded above by c_1/c_2 , uniformly in ξ , so that

$$\int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \cdot \Lambda^{d+p}.$$

Note that, in contrast to the super-smooth case, both bounds are now purely polynomial in Λ .

Step 5': Conclusion. Substituting these bounds into Equation (7) and using Assumption 1 to replace $\Xi(\|\Sigma - \Sigma'\|_{\text{op}})$ with $(\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1)$, we obtain

$$W_1(P, P') \lesssim \frac{C_0}{\Lambda} + \|p_G - p_{G'}\|_1 \cdot \Lambda^{d+\beta} + (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1) \cdot \Lambda^{d+p}.$$

Setting $\Lambda \asymp W_1(P, P')^{-1}$ and rearranging, we obtain that either

$$\|p_G - p_{G'}\|_1 \gtrsim W_1(P, P')^{d+\beta+1},$$

or

$$\|p_G - p_{G'}\|_1 \gtrsim (\Psi \circ \Xi^{-1})\left(CW_1(P, P')^{d+p+1}\right).$$

These two cases, combined with Assumption 1, give Equation (9). \square

3.3. Kernel with PDE inversion. To remedy the dependence on d in the inverse bound based on ordinary-smooth kernels, we now introduce the following PDE inversion assumption. Before proceeding, as we are about to work with differential operators defined in the weak functional-analytic sense of distributions, we refer the reader to, e.g., Chapter 9 of [20] for a concise introduction to the topic.

DEFINITION 3.5. For any $\nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}^d$, set $|\nu| := \sum_{j=1}^d \nu_j$ and define the distributional derivative operator $\partial^\nu := \frac{\partial^{|\nu|}}{\partial x_1^{\nu_1} \dots \partial x_d^{\nu_d}}$. Then the family \mathcal{P} satisfies a *PDE inversion condition of order $\beta \in \mathbb{N}$* if there exists a linear differential operator

$$\mathcal{T}_\Sigma = \sum_{\nu \in \mathbb{N}^d : 0 \leq |\nu| \leq \beta} c_\nu(\Sigma) \partial^\nu$$

such that

1. there exists $\nu \in \mathbb{N}^d$ with $|\nu| = \beta$ such that $c_\nu(\Sigma) \neq 0$,
2. there exists a constant $L > 0$, independent of ν , such that the coefficients $\Sigma \mapsto c_\nu(\Sigma)$ are L -Lipschitz continuous from $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ to \mathbb{R} , and
3. the equation $\mathcal{T}_\Sigma p_G = P$ holds in the sense of distributions on \mathbb{R}^d . That is, for every test function $\phi \in C_c^\infty(\mathbb{R}^d)$, we have

$$\langle p_G, \mathcal{T}_\Sigma^* \phi \rangle = \int_{\mathbb{R}^d} \phi(x) P(dx),$$

where $\mathcal{T}_\Sigma^* := \sum_{\nu \in \mathbb{N}^d : 0 \leq |\nu| \leq \beta} (-1)^{|\nu|} c_\nu(\Sigma) \partial^\nu$ is the formal adjoint of \mathcal{T}_Σ , and $\langle \cdot, \cdot \rangle$ denotes the natural pairing between distributions and test functions.

To the best of our knowledge, the formulation of PDE inversion conditions of the above kind in the context of mixture model theory is new. As it turns out, conditions of this form are especially well-suited to our problem, as they provide a convenient way to invert the convolution $p_G(x) = P * f(x | \cdot, \Sigma)$ by means of a differential operator of finite order $\beta \in \mathbb{N}$, whose coefficients depend smoothly on the scale matrix Σ . Exploiting this relationship—which, as shown below, holds for Laplace mixtures⁵—allows one to relate differences between the mixture densities p_G and $p_{G'}$ directly to differences between the corresponding mixing measures P and P' and the scale parameters Σ and Σ' . Finally, we emphasize that the use of distribution theory is necessary here, since the mixing measure P may be purely atomic, as is the case almost surely under a Dirichlet process prior, and differentiation must be interpreted in a weak sense via test functions $\phi \in C_c^\infty(\mathbb{R}^d)$.

Before presenting an inverse bound based on the PDE inversion condition, we relate it to the ordinary smoothness of the kernel, thereby justifying its use as a substitute for obtaining sharper inequalities.

⁵Though we do not pursue this here, the PDE inversion condition of Definition 3.5 is applicable to other kinds of mixtures, such as those driven by shifted exponential and fixed-shape Gamma kernels.

PROPOSITION 3.6. *Suppose that $f(x | 0, \Sigma) = |\Sigma|^{-1/2} g(x^\top \Sigma^{-1} x)$ for some $g : [0, \infty) \rightarrow [0, \infty)$. If \mathcal{P} satisfies the PDE inversion condition of even order $\beta \in \mathbb{N}$, then f is ordinary-smooth of order β .*

That is, when the index β is even (as it is the case for Laplace mixtures, cf. Theorem 4.4 below), the PDE condition is a proper strengthening of ordinary-smoothness of the kernel f . This, as we show next, plays a crucial role in sharpening the inverse bound obtained in Theorem 3.4.

THEOREM 3.7. *Under Assumption 1, if \mathcal{P} satisfies a PDE inversion condition of order β , there exist positive quantities C and C' , depending only on R , λ_{\min} , λ_{\max} , and d , such that*

$$(10) \quad \|p_G - p_{G'}\|_1 \geq C' \left[\Psi(\|\Sigma - \Sigma'\|_{\text{op}}) + \min \left\{ W_1(P, P')^\beta, \Psi(CW_1(P, P')^\beta) \right\} \right],$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} .

Theorem 3.7 improves upon Theorem 3.4 by removing the explicit dependence on the dimension d in the exponent. That is, while the bound in Theorem 3.4 describes a dimension-dependent relationship under simple ordinary-smoothness, the PDE inversion condition yields, remarkably, a dimension-free rate governed solely by the smoothness order β . This shows that, for mixtures satisfying this condition, the efficiency in recovering the mixing measure is determined by the smoothness properties of the kernel (with higher smoothness leading to weaker bounds) rather than by the ambient dimension.

PROOF SKETCH OF THEOREM 3.7. The proof shares Step 1 with Theorems 3.2 and 3.4, producing a compactly supported, globally Lipschitz continuous function $\tilde{h}_\chi^* \in L^1(\mathbb{R}^d)$ satisfying $W_1(P, P') = \int_{\mathbb{R}^d} \tilde{h}_\chi^*(\theta) (P(d\theta) - P'(d\theta))$. The key departure begins in Step 2'', where we use a smooth mollifier rather than a Jackson kernel to construct the approximation $\tilde{h}_{\chi, \Lambda}^*$, which grants additional regularities needed to invoke the PDE inversion condition. Steps 3''–5'' then exploit this condition directly in the spatial domain, bypassing Fourier inversion entirely and leading to dimension-free bounds.

Step 2'': Mollifier approximation. Rather than convolving with a Jackson kernel as in the previous proofs, we convolve \tilde{h}_χ^* with a scaled mollifier $\psi_\Lambda(x) := \Lambda^d \psi(\Lambda x)$, where $\psi(x) = C_\psi^{-1} \varrho(x)$ is a normalized bump function with $C_\psi := \int_{\mathbb{R}^d} \varrho(x) dx$ and

$$\varrho(x) = \begin{cases} \exp\left(-\frac{1}{1 - \|x\|^2}\right) & \text{if } \|x\| < 1, \\ 0 & \text{if } \|x\| \geq 1, \end{cases}$$

so that $\psi \in C_c^\infty(\mathbb{R}^d)$ and $\int_{\mathbb{R}^d} \psi(x) dx = 1$. We then define

$$\tilde{h}_{\chi, \Lambda}^*(x) := (\tilde{h}_\chi^* * \psi_\Lambda)(x) = \int_{\mathbb{R}^d} \tilde{h}_\chi^*(x - y) \psi_\Lambda(y) dy.$$

The resulting function $\tilde{h}_{\chi, \Lambda}^*$ belongs to $C_c^\infty(\mathbb{R}^d)$, which makes it an admissible test function for the distributional PDE inversion condition. The Lipschitz continuity of \tilde{h}_χ^* again implies $\|\tilde{h}_{\chi, \Lambda}^* - \tilde{h}_\chi^*\|_\infty \leq (1 + 2RM_\eta)M_\psi/\Lambda$, where $M_\psi := \int_{\mathbb{R}^d} \|x\| \psi(x) dx < \infty$, and combined with $\sup_{A \in \mathcal{B}(\mathbb{R}^d)} |P(A) - P'(A)| \leq 2$ this yields

$$(11) \quad W_1(P, P') \leq \frac{C_0}{\Lambda} + \int_{\mathbb{R}^d} \tilde{h}_{\chi, \Lambda}^*(\theta) (P(d\theta) - P'(d\theta))$$

for a constant $C_0 > 0$ depending only on R and f .

Step 3'': PDE inversion. Since $\mathcal{T}_\Sigma p_G = P$ and $\mathcal{T}_{\Sigma'} p_{G'} = P'$ hold in the sense of distributions, and $\tilde{h}_{\mathcal{X},\Lambda}^* \in C_c^\infty(\mathbb{R}^d)$ is an admissible test function, we may write

$$\begin{aligned} \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta) (P(d\theta) - P'(d\theta)) &= \langle \tilde{h}_{\mathcal{X},\Lambda}^*, \mathcal{T}_\Sigma p_G - \mathcal{T}_{\Sigma'} p_{G'} \rangle \\ &= \langle \mathcal{T}_\Sigma^* \tilde{h}_{\mathcal{X},\Lambda}^*, p_G - p_{G'} \rangle + \langle (\mathcal{T}_\Sigma^* - \mathcal{T}_{\Sigma'}^*) \tilde{h}_{\mathcal{X},\Lambda}^*, p_{G'} \rangle, \end{aligned}$$

where $\mathcal{T}_\Sigma^* := \sum_{0 \leq |\nu| \leq \beta} (-1)^{|\nu|} c_\nu(\Sigma) \partial^\nu$ is the formal adjoint of \mathcal{T}_Σ , and we added and subtracted $\mathcal{T}_{\Sigma'} p_{G'}$ to isolate the two sources of discrepancy. We label the two resulting terms I_1 (observation penalty) and I_2 (scale mismatch).

Step 4'': Bounding I_1 and I_2 . For I_1 , Hölder's inequality gives $|I_1| \leq \|p_G - p_{G'}\|_1 \|\mathcal{T}_\Sigma^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty$. To bound $\|\mathcal{T}_\Sigma^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty$, we use the fact that $\tilde{h}_{\mathcal{X},\Lambda}^* = \tilde{h}_{\mathcal{X}}^* * \psi_\Lambda$ to transfer one derivative onto $\tilde{h}_{\mathcal{X}}^*$ and the remaining ones onto ψ_Λ . Young's inequality and the scaling identity $\|\partial^\nu \psi_\Lambda\|_1 = \Lambda^{|\nu|-1} \|\partial^\nu \psi\|_1$ then yield $\|\partial^\nu \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty \lesssim \Lambda^{|\nu|-1}$ for all $\nu \in \mathbb{N}^d$ with $1 \leq |\nu| \leq \beta$. Since the coefficients $c_\nu(\Sigma)$ are uniformly bounded over $\Sigma \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$, we conclude

$$|I_1| \lesssim (1 + \Lambda^{\beta-1}) \|p_G - p_{G'}\|_1.$$

For I_2 , we use $\|p_{G'}\|_1 = 1$ together with the L -Lipschitz continuity of $\Sigma \mapsto c_\nu(\Sigma)$ from the PDE inversion condition, which gives $|c_\nu(\Sigma) - c_\nu(\Sigma')| \leq L \|\Sigma - \Sigma'\|_{\text{op}}$. Applying the same derivative bounds as for I_1 , we obtain

$$|I_2| \lesssim (1 + \Lambda^{\beta-1}) \|\Sigma - \Sigma'\|_{\text{op}}.$$

Note that both bounds are now dimension-free, in contrast to those obtained in the proof of Theorem 3.4.

Step 5'': Conclusion. Substituting the bounds on I_1 and I_2 into Equation (11) and using Assumption 1 to replace $\|\Sigma - \Sigma'\|_{\text{op}}$ with $\Psi^{-1}(\|p_G - p_{G'}\|_1)$, we obtain

$$W_1(P, P') \lesssim \frac{C_0}{\Lambda} + (1 + \Lambda^{\beta-1}) \|p_G - p_{G'}\|_1 + (1 + \Lambda^{\beta-1}) \Psi^{-1}(\|p_G - p_{G'}\|_1).$$

Two cases arise depending on the size of $W_1(P, P')$. If $W_1(P, P')$ is bounded away from zero, setting $\Lambda = 1$ and rearranging yields $\|p_G - p_{G'}\|_1 \gtrsim \min\{W_1(P, P')^\beta, \Psi(CW_1(P, P')^\beta)\}$ directly. If instead $W_1(P, P')$ is small, setting $\Lambda \asymp W_1(P, P')^{-1}$ and rearranging leads to the same bound. In both cases, combining with Assumption 1 gives Equation (10). \square

4. Examples. We now specialize the general inequalities from Section 3 to specific choices of the mixture kernel f . In particular, we focus on three canonical cases, namely Gaussian, Cauchy, and Laplace kernels, which exemplify the different assumptions considered in our theory.

4.1. *Gaussian kernel.* The multivariate Gaussian kernel, which is arguably the most widely adopted choice in applied mixture modeling, is given by

$$f(x | \theta, \Sigma) = \frac{\exp\left(-\frac{1}{2}(x - \theta)^\top \Sigma^{-1}(x - \theta)\right)}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}}, \quad x \in \mathbb{R}^d,$$

for some $(\theta, \Sigma) \in \mathcal{X} \times \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$. The following theorem shows that it belongs to the class of super-smooth kernels and provides the corresponding explicit inverse bound.

THEOREM 4.1. *The multivariate Gaussian kernel is super-smooth of order 2. Moreover, if f is Gaussian, there exists a constant $M > 0$, depending only on R , λ_{\min} , and λ_{\max} , such that Assumptions 1 and 2 hold with*

$$\Psi(t) = \exp\left(-M \frac{\log(1/t)}{t}\right), \quad \Xi(t) = t, \quad \text{and} \quad p = 2.$$

Consequently,

$$\|p_G - p_{G'}\|_1 \geq C' \left[\exp\left(-\frac{C \exp(\tilde{C}/W_1(P, P')^2)}{W_1(P, P')^2}\right) + \exp\left(-M \frac{\log(1/\|\Sigma - \Sigma'\|_{\text{op}})}{\|\Sigma - \Sigma'\|_{\text{op}}}\right) \right],$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} , where $C, C', \tilde{C} > 0$ depend on $R, \lambda_{\min}, \lambda_{\max}$, and d .

As a preview of the posterior contraction results in Section 5, Theorem 4.1 shows how an L^1 estimation rate ω_n for p_{G_0} propagates to P_0 and Σ_0 . In particular, one obtains

$$W_1(P, P_0) \lesssim (\log \log(1/\omega_n))^{-1/2}, \quad \|\Sigma - \Sigma_0\|_{\text{op}} \lesssim \frac{\log \log(1/\omega_n)}{\log(1/\omega_n)}.$$

These rates are quite slow, reflecting the strong smoothing effect induced by the Gaussian kernel. This will serve as a benchmark when comparing with other kernel choices.

Before moving on to other kernel types, we briefly focus on the Gaussian setting where the scale matrices under consideration are of the form $\Sigma = \sigma^2 I_d$, so that \mathcal{G} consists of mixing measures $G = P \times \delta_{\sigma^2 I_d}$ with $\sigma^2 \in [\lambda_{\min}, \lambda_{\max}]$. We refer to this as an *isotropic Gaussian mixture*. In this case, the next result improves the W_1 rate dependence from double-logarithmic to logarithmic. From a technical point of view, the proof is not significantly different from the general super-smooth case and only exploits the additional semi-group structure of the Gaussian kernel family.

THEOREM 4.2. *For an isotropic Gaussian mixture, there exist positive constants M, C , and C' , depending only on R, λ_{\min} , and λ_{\max} , and d , such that*

$$\|p_G - p_{G'}\|_1 \geq C' \left[\exp\left(-\frac{C}{W_1(P, P')^2}\right) + \exp\left(-M \frac{\log(1/|\sigma^2 - \sigma'^2|)}{|\sigma^2 - \sigma'^2|}\right) \right],$$

for any $G = P \times \delta_{\sigma^2 I_d}$ and $G' = P' \times \delta_{\sigma'^2 I_d}$ in \mathcal{G} .

While Theorem 4.2 leaves the estimation rate for $\Sigma_0 = \sigma_0^2 I_d$ unchanged relative to the fully general counterpart in Theorem 4.1, it significantly improves the implied W_1 bound,

$$W_1(P, P_0) \lesssim (\log(1/\omega_n))^{-1/2},$$

which will be reflected in our analysis of posterior contraction rates in Section 5. We also trivially note that, for $d = 1$, Theorem 4.2 fully replaces Theorem 4.1 as a strictly tighter result.

4.2. Cauchy kernel. We next consider the multivariate Cauchy kernel,

$$f(x | \theta, \Sigma) = \frac{\Gamma\left(\frac{1+d}{2}\right)}{\pi^{\frac{d+1}{2}} |\Sigma|^{\frac{1}{2}} [1 + (x - \theta)^\top \Sigma^{-1} (x - \theta)]^{\frac{d+1}{2}}},$$

which leads to the following inverse bound.

THEOREM 4.3. *The multivariate Cauchy kernel is super-smooth of order 1. Moreover, if f is Cauchy, Assumptions 1 and 2 hold with*

$$\Psi(t) = t^2, \quad \Xi(t) = \sqrt{t}, \quad \text{and} \quad p = 1.$$

Consequently,

$$\|p_G - p_{G'}\|_1 \geq C' \left[\exp\left(-\frac{C}{W_1(P, P')}\right) + \|\Sigma - \Sigma'\|_{\text{op}}^2 \right],$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} , where $C, C' > 0$ may depend on $R, \lambda_{\min}, \lambda_{\max}$, and d .

In this case, an L^1 rate ω_n for p_{G_0} translates into

$$W_1(P, P_0) \lesssim (\log(1/\omega_n))^{-1}, \quad \|\Sigma - \Sigma_0\|_{\text{op}} \lesssim \omega_n^{1/2}.$$

Compared to the Gaussian setting, both rates are faster. This is consistent with the lower degree of smoothness of the Cauchy kernel ($\alpha = 1$), which reduces the loss of information induced by convolving G with f and makes the inversion problem statistically easier.

4.3. Laplace kernel. Finally, we consider the multivariate Laplace kernel

$$f(x \mid \theta, \Sigma) = \frac{2 \left(\frac{(x-\theta)^\top \Sigma^{-1} (x-\theta)}{2} \right)^{\frac{2-d}{2}} B_{\frac{d-2}{2}} \left(\sqrt{2(x-\theta)^\top \Sigma^{-1} (x-\theta)} \right)}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}},$$

where B_κ denotes the modified Bessel function of the second kind. As previewed in Section 3, Laplace mixtures are ordinary-smooth and satisfy the PDE inversion condition introduced in Definition 3.5, which allows one to obtain the following inverse bound.

THEOREM 4.4. *The multivariate Laplace kernel is ordinary-smooth of order 2. Moreover, if f is Laplace, Assumptions 1 and 2 hold with*

$$\Psi(t) = \Xi(t) = t \quad \text{and} \quad p = 2,$$

and \mathcal{P} satisfies a PDE inversion condition of order 2 for the operator

$$\mathcal{T}_\Sigma = \text{Id} - \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \Sigma_{jk} \frac{\partial^2}{\partial x_j \partial x_k},$$

where Id denotes the identity operator. Consequently,

$$\|p_G - p_{G'}\|_1 \geq C' [W_1(P, P')^2 + \|\Sigma - \Sigma'\|_{\text{op}}],$$

for any $G = P \times \delta_\Sigma$ and $G' = P' \times \delta_{\Sigma'}$ in \mathcal{G} , where $C' > 0$ may depend on $R, \lambda_{\min}, \lambda_{\max}$, and d .

In this setting, an L^1 rate ω_n for p_{G_0} yields

$$W_1(P, P_0) \lesssim \omega_n^{1/2}, \quad \|\Sigma - \Sigma_0\|_{\text{op}} \lesssim \omega_n.$$

These rates improve upon both the Gaussian and Cauchy cases. This reflects the weaker smoothness of the Laplace kernel (which is only ordinary-smooth) together with the additional structure provided by the PDE inversion property allowing for a more direct recovery of the underlying mixing measure.

5. Posterior asymptotics for Dirichlet process mixtures. Equipped with the general inequalities obtained in Section 3 and their specializations in Section 4, we now derive posterior contraction results for mixing measures in Bayesian nonparametric homoscedastic location-scale mixture models. As discussed earlier, our approach proceeds in two steps:

1. We verify an L^1 posterior contraction rate ω_n for density estimation, namely such that

$$\Pi(G : \|p_G - p_{G_0}\|_1 \geq K\omega_n \mid X_1, \dots, X_n) \rightarrow 0$$

in $P_{G_0}^n$ -probability for all large $K > 0$ as $n \rightarrow \infty$, for a given kernel f defining the density $p_G(x) = P * f(x \mid \cdot, \Sigma)$ (as usual, with $G = P \times \delta_\Sigma$). Such results are well studied in the literature on Bayesian asymptotics and will be borrowed accordingly;

2. We use the inverse inequalities developed in the previous sections to translate the L^1 rate ω_n into contraction rates for $W_1(P, P_0)$ and $\|\Sigma - \Sigma_0\|_{\text{op}}$.

We provide explicit results for multivariate Gaussian mixtures and univariate Laplace mixtures under Dirichlet process priors, as these are settings where suitable L^1 contraction rates are available. In contrast, we do not include any explicit result for Cauchy mixtures, despite the bounds obtained in Theorem 4.3. This is because, to the best of our knowledge, no L^1 posterior contraction results are currently available for infinite Cauchy mixtures. This gap appears to stem from limitations in the existing techniques used to estimate the entropy numbers of such model classes, which are in turn essential to establish posterior contraction rates (cf. the discussion in Section 2). In particular, standard approaches, originating from [30], rely on approximating infinite mixtures by finitely supported ones via moment-matching strategies, which fail in their current form with heavy-tailed kernels. As a consequence, entropy bounds and posterior contraction rates for infinite Cauchy mixtures constitute an open problem which is beyond the scope of this article.

In the Gaussian case, combining our inequalities with Theorem 9.9 in [29] [see also 69], which establishes the L^1 contraction rate $\omega_n = n^{-1/2}(\log n)^{(d+1)/2}$, yields the following result.

PROPOSITION 5.1. *Assume that $X_1, X_2, \dots \stackrel{\text{i.i.d.}}{\sim} p_{G_0} = \int_{\mathbb{R}} f(\cdot \mid \theta, \Sigma_0) P_0(d\theta)$, where f is the d -variate Gaussian kernel, $P_0(\bar{B}(0, R_0)) = 1$ for some $R_0 \in (0, \infty)$, and $\Sigma_0 \in \mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ for some $0 < \lambda_{\min} < \lambda_{\max} < \infty$. Consider a homoscedastic Gaussian location-scale mixture prior Π on $G = P \times \delta_\Sigma$, where $P \sim \text{DP}(a, H)$ (for some $a > 0$) with base measure H having a strictly positive and continuous density supported on $[-R, R]^d$ for some $R \in [R_0, \infty)$, and Σ has a prior with positive continuous density supported on $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$. Then the following statements hold true.*

- (a) For $K > 0$ sufficiently large,

$$\begin{aligned} & \Pi \left(G : \exp \left(- \frac{C \exp(\tilde{C}/W_1(P, P_0)^2)}{W_1(P, P_0)^2} \right) \right. \\ & \quad \left. + \exp \left(- M \frac{\log(1/\|\Sigma - \Sigma_0\|_{\text{op}})}{\|\Sigma - \Sigma_0\|_{\text{op}}} \right) \geq K n^{-1/2} (\log n)^{\frac{d+1}{2}} \mid X_1, \dots, X_n \right) \rightarrow 0 \end{aligned}$$

in $P_{G_0}^n$ -probability as $n \rightarrow \infty$, where the positive constants M , C , and \tilde{C} may depend on R , λ_{\min} , λ_{\max} , and d . Consequently,

$$\begin{aligned} & \Pi \left(G : W_1(P, P_0) \geq K \frac{1}{\sqrt{\log \log n}} \mid X_1, \dots, X_n \right) \rightarrow 0, \\ & \Pi \left(G : \|\Sigma - \Sigma_0\|_{\text{op}} \geq K \frac{\log \log n}{\log n} \mid X_1, \dots, X_n \right) \rightarrow 0 \end{aligned}$$

in $P_{G_0}^n$ -probability for all large $K > 0$ as $n \rightarrow \infty$.

(b) In the isotropic Gaussian mixture setting, the above convergence statements may be replaced by

$$\begin{aligned} & \Pi \left(G : \exp \left(- \frac{C}{W_1(P, P_0)^2} \right) \right. \\ & \quad \left. + \exp \left(- M \frac{\log(1/|\sigma^2 - \sigma_0^2|)}{|\sigma^2 - \sigma_0^2|} \right) \geq K n^{-1/2} (\log n)^{\frac{d+1}{2}} \mid X_1, \dots, X_n \right) \rightarrow 0 \end{aligned}$$

and

$$\begin{aligned} & \Pi \left(G : W_1(P, P_0) \geq K \frac{1}{\sqrt{\log n}} \mid X_1, \dots, X_n \right) \rightarrow 0, \\ & \Pi \left(G : |\sigma^2 - \sigma_0^2| \geq K \frac{\log \log n}{\log n} \mid X_1, \dots, X_n \right) \rightarrow 0. \end{aligned}$$

in $P_{G_0}^n$ -probability as $n \rightarrow \infty$.

Proposition 5.1 highlights a key feature of the Gaussian setting. Despite the essentially parametric L^1 rate for density estimation, the super-smooth nature of the Gaussian kernel in general leads to slow, double logarithmic rates for recovering the mixing measure P_0 and logarithmic rates for the scale matrix Σ_0 . The double logarithmic rate is however turned into a milder (but nonetheless slow) logarithmic one in the isotropic case. On the other hand, the dependence on the dimension d , which appears at the exponent of the logarithmic factor in the L^1 rate, only affects lower-order terms and does not influence the leading behavior of the W_1 and $\|\cdot\|_{\text{op}}$ rates (in the fixed- d regime considered here). Finally, Proposition 5.1 highlights that the rate of recovery for Σ_0 is much faster than the rates for P_0 , both in the general and isotropic cases, uncovering significant heterogeneity in parameter estimation efficiency with this kind of infinite mixture model.

We now turn to the Laplace case. Using the L^1 contraction rate $\omega_n = n^{-1/4}(\log n)^{5/4}$ available for univariate Dirichlet process mixtures of Laplace kernels,⁶ and combining it with Theorem 4.4, we obtain the following result.

PROPOSITION 5.2. *Assume that $X_1, X_2, \dots \stackrel{\text{i.i.d.}}{\sim} p_{G_0} = \int_{\mathbb{R}} f(\cdot \mid \theta, \sigma_0) P_0(d\theta)$, where f is the univariate Laplace kernel and $P_0([-R_0, R_0]) = 1$ for some $R_0 \in (0, \infty)$. Consider a homoscedastic Laplace location-scale mixture prior Π on $G = P \times \delta_\sigma$, where $P \sim \text{DP}(a, H)$ (for some $a > 0$) with base measure H having a strictly positive continuous density supported on $[-R, R] \subset \mathbb{R}$ for some $R \in [R_0, \infty)$, and σ has a prior with positive continuous density supported on $[\underline{\sigma}, \bar{\sigma}] \subset (0, \infty)$ for some $0 < \underline{\sigma} < \sigma_0 < \bar{\sigma} < \infty$. Then, for $K > 0$ sufficiently large,*

$$\Pi \left(G : W_1(P, P_0)^2 + |\sigma - \sigma_0| \geq K \frac{(\log n)^{5/4}}{n^{1/4}} \mid X_1, \dots, X_n \right) \rightarrow 0$$

in $P_{G_0}^n$ -probability as $n \rightarrow \infty$. Consequently,

$$\Pi \left(G : W_1(P, P_0) \geq K \frac{(\log n)^{5/8}}{n^{1/8}} \mid X_1, \dots, X_n \right) \rightarrow 0,$$

⁶Adapted from Theorem 4.4 in [62] and setting, in the author's notation, $p = 1$.

$$\Pi \left(G : |\sigma - \sigma_0| \geq K \frac{(\log n)^{5/4}}{n^{1/4}} \mid X_1, \dots, X_n \right) \rightarrow 0$$

in $P_{G_0}^n$ -probability for all large $K > 0$ as $n \rightarrow \infty$.

Laplace mixtures display a very different behavior compared to the Gaussian case. Although their L^1 rate is slower (of order $n^{-1/4}$, up to poly-logarithmic factors, compared to the Gaussian $n^{-1/2}$), the resulting rates of recovery of P_0 and Σ_0 are polynomial rather than logarithmic. This reflects the substantially weaker smoothing induced by the Laplace kernel combined with the PDE inversion structure, which yield a much sharper inverse bound (compare Theorems 4.4 and 4.2) and compensate for the slower density estimation rate. However, Laplace mixtures exhibit the same qualitative separation in contraction rates as Gaussian mixtures: the location mixing measure P_0 is recovered at rate $n^{1/8}(\log n)^{5/8}$, whereas the scale parameter Σ_0 is recovered at the faster rate $n^{1/4}(\log n)^{5/4}$. Ignoring poly-logarithmic factors, this reflects significant heterogeneity in the form of a polynomial gap in the estimation rates for these two parameters of interest.

5.1. Mixing measure priors beyond the Dirichlet process. We conclude this section by noting that, although the results above are stated for Dirichlet process mixtures, the underlying methodology is not restricted to this specific prior choice. The arguments rely on two ingredients: (i) the inverse inequalities developed in Sections 3 and 4, and (ii) available L^1 posterior contraction rates for the corresponding mixture models. Since the inequalities are general, they can be combined with contraction results for a wide range of alternative priors on the mixing measure.

Examples include Pitman–Yor processes [54, 57], normalized inverse-Gaussian processes [40], more general normalized random measures with independent increments [38, 42, 53, 58], Gibbs-type priors [13, 32], and species sampling models [37, 56]. For instance, the inequalities developed here could be combined with the contraction rates established in [63] for Pitman–Yor and normalized inverse-Gaussian mixtures. We do not pursue these extensions further, as they do not introduce additional conceptual difficulties given the generality of our framework.

6. Discussion. We studied posterior contraction for mixing measures in homoscedastic location-scale mixture models with infinitely many components. We developed a general approach to derive lower-bounds for the L^1 distance between mixture densities in terms of discrepancies between the corresponding mixing measures. Our method relies on the dual formulation of the W_1 distance in terms of 1-Lipschitz functions, combined with functional approximation and Fourier inversion techniques, and yields inequalities whose strength reflects the smoothness of the mixture kernel and, in the ordinary-smooth case, the applicability of a novel PDE inversion condition. These results provide a unified route from L^1 convergence to contraction at the level of both the mixing measure and the shared scale parameter, and apply to a range of commonly used kernels in the context of Dirichlet process mixtures.

In particular, we point out a connection with deconvolution problems, which have motivated much of the previous work on mixing measure convergence, as they can be framed, in the mixture setting, as the recovery of a signal distributed according to the unknown mixing measure when observations are contaminated by additive noise drawn from the mixture kernel [23, 61, 64]. As noted earlier, existing contributions focus on the location-only mixture case, which in deconvolution terminology corresponds to the setting where the noise scale is known. Our results can be interpreted as characterizing recovery of the mixing measure in the more challenging scenario where the noise scale is also unknown. Moreover, in contrast to

other multivariate deconvolution analyses [61], our approach does not rely on independence assumptions across noise coordinates and instead accommodates general covariance structures. New results on L^1 density estimation for infinite Cauchy mixtures would also make it possible, when combined with our inequality in Theorem 4.3, to fully characterize the currently unexplored problem of deconvolution with heavy-tailed Cauchy error distribution.

The functional-analytic techniques we developed to handle the dual formulation of the W_1 metric are quite general and open several directions for future work. A natural extension is to heteroscedastic location-scale mixtures, where the scale parameter varies across components. In addition, our results for different kernel choices reveal significant heterogeneity, with Laplace mixtures achieving substantially faster rates than Gaussian mixtures. This raises the question of whether, when the primary objects of interest are the mixing measure and the scale (as in the deconvolution problems discussed above), it may be advantageous to use a misspecified kernel with lower smoothness, potentially improving recovery rates—perhaps up to an approximation error—when the true kernel is very smooth. A natural setting to investigate this question is the class of exponential power kernels considered for density estimation by [62], where the smoothness can be tuned continuously and misspecification can be studied in a systematic way. Finally, it would be of interest to extend our results to models with more complex prior structures on the mixing measure, such as those inducing dependence across multiple mixing measures [22, 41, 59, 67, among many others] or repulsion across atoms to achieve well-separated components [55, 78].

SUPPLEMENTARY MATERIAL

Supplement to “Convergence Rates for Latent Mixing Measures in Infinite Homoscedastic Location-Scale Mixture Models”

This document collects the detailed proofs of all the results presented in the article.

REFERENCES

- [1] ASCOLANI, F., LIJOI, A., REBAUDO, G. and ZANELLA, G. (2023). Clustering consistency with Dirichlet process mixtures. *Biometrika* **110** 551–558.
- [2] BARILETTO, N., FLORES, B. and WALKER, S. G. (2025). Posterior Consistency in Parametric Models via a Tighter Notion of Identifiability. *arXiv preprint arXiv:2504.11360*.
- [3] BARILETTO, N., LE, D., RINALDO, A. and HO, N. (2026). Supplement to “Convergence Rates for Latent Mixing Measures in Infinite Homoscedastic Location-Scale Mixture Models”.
- [4] BARRON, A., SCHERVISH, M. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics* **27** 536–561.
- [5] BUTZER, P. L. and NESSEL, R. J. (1971). *Fourier analysis and approximation, Vol. 1* 7. Birkhäuser Basel.
- [6] CAILLERIE, C., CHAZAL, F., DEDECKER, J. and MICHEL, B. (2011). Deconvolution for the Wasserstein metric and geometric inference. *Electronic Journal of Statistics* **5** 1394–1423.
- [7] CANALE, A. and DE BLASI, P. (2017). Posterior asymptotics of nonparametric location-scale mixtures for multivariate density estimation. *Bernoulli* **23** 379–404.
- [8] CASTILLO, I. (2024). *Bayesian Nonparametric Statistics. École d’Été de Probabilités de Saint-Flour LI - 2023*. Springer Nature.
- [9] CATALANO, M. and LAVENANT, H. (2025). Measures of dependence based on Wasserstein distances. *arXiv preprint arXiv:2510.06034*.
- [10] CATALANO, M., LAVENANT, H., LIJOI, A. and PRÜNSTER, I. (2024). A Wasserstein index of dependence for random measures. *Journal of the American Statistical Association* **119** 2396–2406.
- [11] CATALANO, M., LIJOI, A. and PRÜNSTER, I. (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *The Annals of Statistics* **49** 2916–2947.
- [12] CHAE, M. and WALKER, S. G. (2017). A novel approach to Bayesian consistency. *Electronic Journal of Statistics* **11** 4723–4745.
- [13] DE BLASI, P., FAVARO, S., LIJOI, A., MENA, R. H., PRÜNSTER, I. and RUGGIERO, M. (2015). Are Gibbs-type priors the most natural generalization of the Dirichlet process? *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37** 212–229.

- [14] DEDECKER, J., FISCHER, A. and MICHEL, B. (2015). Improved rates for Wasserstein deconvolution with ordinary smooth error in dimension one. *Electronic Journal of Statistics* **9** 234–265.
- [15] DEDECKER, J. and MICHEL, B. (2013). Minimax rates of convergence for Wasserstein deconvolution with supersmooth errors in any dimension. *Journal of Multivariate Analysis* **122** 278–291.
- [16] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive approximation* **303**. Springer Science & Business Media.
- [17] ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90** 577–588.
- [18] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1** 209–230.
- [19] FERGUSON, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics* 287–302. Elsevier.
- [20] FOLLAND, G. B. (1999). *Real analysis: modern techniques and their applications*. John Wiley & Sons.
- [21] FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97** 611–631.
- [22] FRANZOLINI, B., LIJOI, A., PRÜNSTER, I. and REBAUDO, G. (2025). Multivariate species sampling models. *arXiv preprint arXiv:2503.24004*.
- [23] GAO, F. and VAN DER VAART, A. (2016). Posterior contraction rates for deconvolution of Dirichlet-Laplace mixtures. *Electronic Journal of Statistics* **10** 608–627.
- [24] GENOVESE, C. R. and WASSERMAN, L. (2000). Rates of convergence for the Gaussian mixture sieve. *The Annals of Statistics* **28** 1105–1127.
- [25] GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics* **27** 143–158.
- [26] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. (2000). Convergence rates of posterior distributions. *The Annals of Statistics* **28** 500–531.
- [27] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics* **35** 192 – 223.
- [28] GHOSAL, S. and VAN DER VAART, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *The Annals of Statistics* **35** 697–723.
- [29] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference* **44**. Cambridge University Press.
- [30] GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *The Annals of Statistics* **29** 1233–1263.
- [31] GINÉ, E. and NICKL, R. (2011). Rates of contraction for posterior distributions in L^r -metrics, $1 \leq r \leq \infty$. *The Annals of Statistics* **39** 2883 – 2911.
- [32] GNEDIN, A. and PITMAN, J. (2006). Exchangeable Gibbs partitions and Stirling triangles. *Journal of Mathematical Sciences* **138** 5674–5685.
- [33] GUHA, A., HO, N. and NGUYEN, X. (2021). On posterior contraction of parameters and interpretability in Bayesian mixture modeling. *Bernoulli* **27** 2159–2188.
- [34] HEINONEN, J. (2001). *Lectures on analysis on metric spaces*. Springer Science & Business Media.
- [35] HO, N. and NGUYEN, X. (2016). Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *The Annals of Statistics* **44** 2726–2755.
- [36] HO, N. and NGUYEN, X. (2016). On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electronic Journal of Statistics* **10** 271–307.
- [37] KINGMAN, J. F. (1978). The representation of partition structures. *Journal of the London Mathematical Society* **2** 374–380.
- [38] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics* **21** 59–78.
- [39] LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2 ed. *Graduate Texts in Mathematics* **218**. Springer, New York.
- [40] LIJOI, A., MENA, R. H. and PRÜNSTER, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association* **100** 1278–1291.
- [41] LIJOI, A., NIPOTI, B. and PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* 1260–1291.
- [42] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N. L. Hjort, C. Holmes, P. Müller and S. G. Walker, eds.) 80–136. Cambridge University Press.
- [43] LIJOI, A., PRÜNSTER, I. and WALKER, S. G. (2005). On consistency of nonparametric normal mixtures for Bayesian density estimation. *Journal of the American Statistical Association* **100** 1292–1296.
- [44] LO, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics* **12** 351–357.

- [45] MACEACHERN, S. N. and MÜLLER, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* **7** 223–238.
- [46] MAJUMDAR, S. (1992). On topological support of Dirichlet prior. *Statistics & Probability Letters* **15** 381–384.
- [47] MCLACHLAN, G. J. and PEEL, D. (2000). *Finite Mixture Models*. John Wiley & Sons.
- [48] MCSHANE, E. J. (1934). Extension of range of functions. *Bulletin of the American Mathematical Society* **40** 837–842.
- [49] MILLER, J. W. and HARRISON, M. T. (2013). A Simple Example of Dirichlet Process Mixture Inconsistency for the Number of Components. In *Advances in Neural Information Processing Systems* **26**.
- [50] MILLER, J. W. and HARRISON, M. T. (2014). Inconsistency of Pitman–Yor Process Mixtures for the Number of Components. *Journal of Machine Learning Research* **15** 3333–3370.
- [51] NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9** 249–265.
- [52] NGUYEN, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *The Annals of Statistics* **4** 370–400.
- [53] NIETO-BARAJAS, L. E., PRÜNSTER, I. and WALKER, S. G. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* **32** 2343–2360.
- [54] PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields* **92** 21–39.
- [55] PETRALIA, F., RAO, V. and DUNSON, D. (2012). Repulsive mixtures. *Advances in neural information processing systems* **25**.
- [56] PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*, **30** 245–267. Institute of Mathematical Statistics.
- [57] PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* **25** 855–900.
- [58] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics* **31** 560–585.
- [59] RODRIGUEZ, A., DUNSON, D. B. and GELFAND, A. E. (2008). The nested Dirichlet process. *Journal of the American Statistical Association* **103** 1131–1154.
- [60] ROUSSEAU, J. and MENGENSEN, K. (2011). Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **73** 689–710.
- [61] ROUSSEAU, J. and SCRICCILOLO, C. (2024). Wasserstein convergence in Bayesian and frequentist deconvolution models. *The Annals of Statistics* **52** 1691–1715.
- [62] SCRICCILOLO, C. (2011). Posterior rates of convergence for Dirichlet mixtures of exponential power densities. *Electronic Journal of Statistics* **5** 270–308.
- [63] SCRICCILOLO, C. (2014). Adaptive Bayesian density estimation in L_p -metrics with Pitman–Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Analysis* **9** 475–520.
- [64] SCRICCILOLO, C. (2018). Bayes and maximum likelihood for L_1 -Wasserstein deconvolution of Laplace mixtures. *Statistical Methods & Applications* **27** 333–362.
- [65] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4** 639–650.
- [66] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics* **29** 687–714.
- [67] TEH, Y. W., JORDAN, M. I., BEAL, M. J. and BLEI, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association* **101** 1566–1581.
- [68] TOKDAR, S. T. (2006). Posterior consistency of Dirichlet location-scale mixture of normals in density estimation and regression. *Sankhyā: The Indian Journal of Statistics* **67** 90–110.
- [69] TOKDAR, S. T. (2013). Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. *Biometrika* **100** 623–640.
- [70] VILLANI, C. (2008). *Optimal transport: Old and New*. Springer.
- [71] WADE, S. (2023). Bayesian cluster analysis. *Philosophical Transactions of the Royal Society A* **381** 20220149.
- [72] WALKER, S. G. (2004). New approaches to Bayesian consistency. *The Annals of Statistics* **32** 2028–2043.
- [73] WALKER, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation* **36** 45–54.
- [74] WALKER, S. G. and HJORT, N. L. (2001). On Bayesian consistency. *Journal of the Royal Statistical Society, Series B* **63** 811–821.
- [75] WALKER, S. G., LIJOI, A. and PRÜNSTER, I. (2007). On rates of convergence for posterior distributions in infinite-dimensional models. *The Annals of Statistics* **35** 738–746.

- [76] WHITNEY, H. (1934). Analytic Extensions of Differentiable Functions Defined in Closed Sets. *Transactions of the American Mathematical Society* **36** 63.
- [77] WU, Y. and GHOSAL, S. (2010). The L_1 -consistency of Dirichlet mixtures in multivariate Bayesian density estimation. *Journal of Multivariate Analysis* **101** 2411–2419.
- [78] XU, Y., MÜLLER, P. and TELESKA, D. (2016). Bayesian inference for latent biologic structure with determinantal point processes (DPP). *Biometrics* **72** 955–964.

SUPPLEMENT TO “CONVERGENCE RATES FOR LATENT MIXING MEASURES IN INFINITE HOMOSCEDASTIC LOCATION-SCALE MIXTURE MODELS.”

BY NICOLA BARILETTO^{1,a}, DUNG LE^{1,b}, ALESSANDRO RINALDO^{1,c} AND NHAT HO^{1,d}

¹Department of Statistics and Data Sciences, University of Texas at Austin, ^anicola.bariletto@utexas.edu;
^bquangdung0110@utexas.edu; ^calessandro.rinaldo@austin.utexas.edu; ^dminhnhat@utexas.edu

This document collects the detailed proofs of all the results presented in the article “Convergence Rates for Latent Mixing Measures in Infinite Homoscedastic Location-Scale Mixture Models,” together with auxiliary lemmas and their proofs.

1. Proofs of theorems and propositions presented in the main text.

Proof of Theorem 3.2. Recall that the probability measures P and P' are supported on $\mathcal{X} = \bar{B}(0, R)$. By the fundamental Kantorovich-Rubinstein Duality Theorem [12] for optimal transport on the compact metric space \mathcal{X} , the W_1 distance can be expressed as

$$W_1(P, P') = \sup_{h \in \text{Lip}_1(\mathcal{X})} \int_{\mathcal{X}} h(\theta)(P(d\theta) - P'(d\theta)) = \int_{\mathcal{X}} h^*(\theta)(P(d\theta) - P'(d\theta)),$$

for a function $h^* : \mathcal{X} \rightarrow \mathbb{R}$ with $h^* \in \text{Lip}_1(\mathcal{X})$.

We define $h_{\mathcal{X}}^*(x) = h^*(x) - h^*(0)$ for all $x \in \mathcal{X}$ to guarantee that $h_{\mathcal{X}}^*(0) = 0$. Furthermore, $|h_{\mathcal{X}}^*(x) - h_{\mathcal{X}}^*(y)| = |h^*(x) - h^*(y)| \leq \|x - y\|$. Thus, $h_{\mathcal{X}}^* \in \text{Lip}_1(\mathcal{X})$. The shift also leaves the optimal transport integral unchanged, which implies that

$$\int_{\mathcal{X}} h_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) = \int_{\mathcal{X}} h^*(\theta)(P(d\theta) - P'(d\theta)) = W_1(P, P').$$

Extension from compact domain \mathcal{X} to \mathbb{R}^d . We now extend the domain of the function $h_{\mathcal{X}}^*$ from \mathcal{X} to \mathbb{R}^d , while preserving its exact 1-Lipschitz property. This can be done via the McShane-Whitney extension, which is recalled in Lemma 2.1 (see [5], [13], or [8] for reference). Given the extension of the function $h_{\mathcal{X}}^*$ to $\bar{h}_{\mathcal{X}}^*$, we can rewrite the integral from the compact subset \mathcal{X} to the full Euclidean space \mathbb{R}^d as follows:

$$\begin{aligned} \int_{\mathbb{R}^d} \bar{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) &= \int_{\mathcal{X}} \bar{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d \setminus \mathcal{X}} \bar{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) \\ &= \int_{\mathcal{X}} h_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) \\ &= W_1(P, P'). \end{aligned}$$

Spatial truncation via smooth cutoff. Since $\bar{h}_{\mathcal{X}}^* \in \text{Lip}_1(\mathbb{R}^d)$, we have the global geometric growth bound:

$$\forall x \in \mathbb{R}^d, \quad |\bar{h}_{\mathcal{X}}^*(x)| = |\bar{h}_{\mathcal{X}}^*(x) - \bar{h}_{\mathcal{X}}^*(0)| \leq \|x\|.$$

Our next goal is to further modify $\bar{h}_{\mathcal{X}}^*(x)$ to obtain an absolutely integrable function that is amenable to Fourier inversion, and we do so via spatial truncation as follows.

Concretely, we aim to construct a spatial cutoff function $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$ in $C_c^\infty(\mathbb{R}^d)$ that takes values exactly in $[0, 1]$, evaluates identically to 1 on the closed ball $\bar{B}(0, R)$ (where P and P'

are supported), and vanishes completely outside the open ball $B(0, 2R)$. Lemma 2.2 provides the detailed construction of such a cutoff function (see also [6]). Hence, given η , we denote

$$\tilde{h}_{\mathcal{X}}^*(x) := \bar{h}_{\mathcal{X}}^*(x)\eta(x)$$

for all $x \in \mathbb{R}^d$. As we show in Lemma 2.3, $\tilde{h}_{\mathcal{X}}^* \in L_1(\mathbb{R}^d) := \{h \in \mathbb{R}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |h(x)| dx < \infty\}$ and it is globally Lipschitz on \mathbb{R}^d . By the properties of η , we also clearly have that

$$\begin{aligned} \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) &= \int_{\mathcal{X}} \tilde{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d \setminus \mathcal{X}} \tilde{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) \\ &= \int_{\mathcal{X}} h_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) \\ (1) \qquad \qquad \qquad &= W_1(P, P'). \end{aligned}$$

Spectral bandlimiting via the Jackson kernel. Since $\tilde{h}_{\mathcal{X}}^* \in L_1(\mathbb{R}^d)$, its Fourier transform $\mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)$ is well-defined. Our goal is to analyze $\int_{\mathcal{X}} \tilde{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta))$, which by the Fourier inversion theorem can be rewritten as $\int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)(\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi$. However, the integrand $\mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)(\Phi_P(\xi) - \Phi_{P'}(\xi))$ may decay slowly and fail to be absolutely integrable. To prevent this, we truncate the frequency domain within a neighborhood $\{\xi \in \mathbb{R}^d : \|\xi\| \leq \Lambda\}$ for some $\Lambda > 0$. This can be done via convolving the function $\tilde{h}_{\mathcal{X}}^*$ with the Jackson kernel.

To that end, consider the multivariate Jackson kernel (see [2], [1])

$$K(x) := \left(\frac{3}{8\pi\sqrt{d}}\right)^d \prod_{i=1}^d \text{sinc}^4\left(\frac{x_i}{4\sqrt{d}}\right), \quad x \in \mathbb{R}^d,$$

where $\text{sinc}(u) := \sin u/u$, which by construction satisfies $\int_{\mathbb{R}^d} K(x) dx = 1$. Lemma 2.4 establishes several properties of the Fourier transform of K and its moments, which we use in what follows. For any $\Lambda > 0$, define the function $K_{\Lambda}(x) := \Lambda^d K(\Lambda x)$. By Lemma 2.4, $\text{supp}(\mathcal{F}[K]) \subseteq \bar{B}(0, 1)$, which implies that $\text{supp}(\mathcal{F}[K_{\Lambda}]) \subseteq \bar{B}(0, \Lambda)$.¹ Consider the function

$$\tilde{h}_{\mathcal{X}, \Lambda}^*(x) := (\tilde{h}_{\mathcal{X}}^* * K_{\Lambda})(x) = \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x-y) K_{\Lambda}(y) dy.$$

Direct computation yields that

$$\begin{aligned} \tilde{h}_{\mathcal{X}, \Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x) &= \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x-y) K_{\Lambda}(y) dy - \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x) K_{\Lambda}(y) dy \\ &= \int_{\mathbb{R}^d} [\tilde{h}_{\mathcal{X}}^*(x-y) - \tilde{h}_{\mathcal{X}}^*(x)] K_{\Lambda}(y) dy. \end{aligned}$$

Since the function $\tilde{h}_{\mathcal{X}}^*$ is a globally $(1 + 2RM_{\eta})$ -Lipschitz function (which follows from Lemma 2.3), we further obtain that, for any $x \in \mathbb{R}^d$,

$$\begin{aligned} |\tilde{h}_{\mathcal{X}, \Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x)| &\leq \int_{\mathbb{R}^d} |\tilde{h}_{\mathcal{X}}^*(x-y) - \tilde{h}_{\mathcal{X}}^*(x)| \Lambda^d K(\Lambda y) dy \\ &\leq (1 + 2RM_{\eta}) \int_{\mathbb{R}^d} \|y\| \Lambda^d K(\Lambda y) dy \\ &= \frac{(1 + 2RM_{\eta})M_K}{\Lambda}, \end{aligned}$$

¹For any function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, $\text{supp}(h)$ denotes the closure of the set $\{x \in \mathbb{R}^d : h(x) \neq 0\}$.

where the last inequality follows from Lemma 2.4. As a consequence, we arrive at

$$(2) \quad \|\tilde{h}_{\mathcal{X},\Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty \leq \frac{(1 + 2RM_\eta)M_K}{\Lambda}.$$

From the dual formulation of $W_1(P, P')$ in Equation (1), we have

$$\begin{aligned} W_1(P, P') &= \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(P(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X},\Lambda}^*(\theta))(P(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta)d(P(d\theta) - P'(d\theta)). \end{aligned}$$

A direct application of Hölder's inequality leads to

$$\int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X},\Lambda}^*(\theta))(P(d\theta) - P'(d\theta)) \leq \|\tilde{h}_{\mathcal{X},\Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty \|P - P'\|_1 \leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda},$$

where the final inequality is due to $\|P - P'\|_1 \leq 2$ and the bound of $\|\tilde{h}_{\mathcal{X},\Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty$ in Equation (2).² Hence, we obtain the following bound

$$(3) \quad W_1(P, P') \leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda} + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta)(P(d\theta) - P'(d\theta)).$$

To apply the Fourier inversion theorem to $\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta)(P(d\theta) - P'(d\theta))$, we first need to demonstrate that $\tilde{h}_{\mathcal{X},\Lambda}^* \in L_1(\mathbb{R}^d)$ and $\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*] \in L_1(\mathbb{R}^d)$. Because $\tilde{h}_{\mathcal{X},\Lambda}^*(x) = (\tilde{h}_{\mathcal{X}}^* * K_\Lambda)(x)$, an application of Young's inequality leads to $\|\tilde{h}_{\mathcal{X},\Lambda}^*\|_1 \leq \|\tilde{h}_{\mathcal{X}}^*\|_1 \|K_\Lambda\|_1 = \|\tilde{h}_{\mathcal{X}}^*\|_1 < \infty$ (thanks to Lemma 2.3). As for $\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]$, by the properties of convolutions we obtain that $\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) = \mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)\mathcal{F}[K_\Lambda](\xi)$. As $\text{supp}(\mathcal{F}[K_\Lambda]) \subseteq \bar{B}(0, \Lambda)$, we have $\text{supp}(\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]) \subseteq \bar{B}(0, \Lambda)$. Furthermore, $\|\mathcal{F}[K_\Lambda]\|_\infty \leq 1$ and $\|\mathcal{F}[\tilde{h}_{\mathcal{X}}^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$ (see Lemma 2.3). Therefore, $\|\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]\|_\infty \leq \|\mathcal{F}[\tilde{h}_{\mathcal{X}}^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$. Putting these results together, we obtain that $\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*] \in L_1(\mathbb{R}^d)$.

Now, from the Fourier inversion theorem, as $\tilde{h}_{\mathcal{X},\Lambda}^* \in L_1(\mathbb{R}^d)$ and $\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*] \in L_1(\mathbb{R}^d)$, we can write

$$\tilde{h}_{\mathcal{X},\Lambda}^*(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) e^{i\xi^\top x} d\xi,$$

leading to

$$\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta)(P(d\theta) - P'(d\theta)) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) e^{i\xi^\top \theta} d\xi \right) (P(d\theta) - P'(d\theta)).$$

Since $|\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]|$ and $P - P'$ have compact support, and $|\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]|$ is bounded, we have

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) e^{i\xi^\top \theta}| d\xi d\theta |P - P'| = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi)| d\xi d\theta |P - P'| < \infty.$$

Thus, by Fubini's theorem, we have that

$$\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta)(P(d\theta) - P'(d\theta)) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{i\xi^\top \theta} (P(d\theta) - P'(d\theta)) \right) \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) d\xi$$

²With a slight abuse of notation, $\|P - P'\|_1 := \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |P(A) - P'(A)|$ denotes the total variation distance between probability measures P and P' .

$$\begin{aligned}
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) (\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi \\
&= \frac{1}{(2\pi)^d} \int_{\bar{B}(0,\Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) (\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi,
\end{aligned}$$

where the final identity is due to the fact that $\text{supp}(\mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*]) \subseteq \bar{B}(0, \Lambda)$.

Putting all of these results together, we arrive at

$$(4) \quad W_1(P, P') \leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda} + \frac{1}{(2\pi)^d} \int_{\bar{B}(0,\Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) (\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi.$$

Next, since $p_G = P * f(\cdot | 0, \Sigma)$ and, therefore, $\Phi_{p_G}(\xi) = \Phi_P(\xi)\Phi_\Sigma(\xi)$ for all $\xi \in \mathbb{R}^d$, we obtain that

$$\begin{aligned}
\Phi_P(\xi) - \Phi_{P'}(\xi) &= \frac{\Phi_{p_G}(\xi)}{\Phi_\Sigma(\xi)} - \frac{\Phi_{p_{G'}}(\xi)}{\Phi_{\Sigma'}(\xi)} \\
&= \frac{\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)}{\Phi_\Sigma(\xi)} + \Phi_{P'}(\xi) \left(\frac{\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)}{\Phi_\Sigma(\xi)\Phi_{\Sigma'}(\xi)} \right) \Phi_{\Sigma'}(\xi),
\end{aligned}$$

where we used the fact that both $\Phi_{\Sigma'}(\xi)$ and $\Phi_\Sigma(\xi)$ are non-zero for all $\xi \in \mathbb{R}^d$ by assumption. Given that $|\Phi_{P'}(\xi)| \leq 1$ for all $\xi \in \mathbb{R}^d$ and $\|\mathcal{F}[\tilde{h}_{\mathcal{X}}^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)} (2R)^{d+1} =: C(R, d)$, an application of the triangle inequality leads to

$$\begin{aligned}
&\left| \int_{\bar{B}(0,\Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X},\Lambda}^*](\xi) (\Phi_P(\xi) - \Phi_{P'}(\xi)) d\xi \right| \\
(5) \quad &\leq \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0,\Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi + \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0,\Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi.
\end{aligned}$$

By construction, $|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)| \leq \|p_G - p_{G'}\|_1$ for any $\xi \in \mathbb{R}^d$. Furthermore, applying the super-smooth assumption, we have $\exp(-c_1\|\xi\|^\alpha) \leq |\Phi_\Sigma(\xi)|$. Therefore, we obtain that

$$\begin{aligned}
&\frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0,\Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \leq \frac{C(R, d)}{(2\pi)^d} \|p_G - p_{G'}\|_1 \int_{\bar{B}(0,\Lambda)} \exp(c_1\|\xi\|^\alpha) d\xi \\
(6) \quad &\leq C_1 \|p_G - p_{G'}\|_1 \Lambda^d \exp(c_1\Lambda^\alpha),
\end{aligned}$$

where $C_1 = \frac{C(R, d)}{(4\pi)^{d/2}\Gamma(d/2+1)}$ is a constant depending on R and d .

By Assumption 2, we get

$$(7) \quad |\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \|\xi\|^p \max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|)$$

and thus, by super-smoothness,

$$\begin{aligned}
&\int_{\bar{B}(0,\Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \int_{\bar{B}(0,\Lambda)} \|\xi\|^p \left[\frac{\max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|)}{|\Phi_\Sigma(\xi)|} \right] d\xi \\
&\leq \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \int_{\bar{B}(0,\Lambda)} \|\xi\|^p \exp((c_1 - c_2)\|\xi\|^\alpha) d\xi \\
&\leq \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \exp((c_1 - c_2)\Lambda^\alpha) \int_{\bar{B}(0,\Lambda)} \|\xi\|^p d\xi \\
(8) \quad &= C' \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \Lambda^{d+p} \exp((c_1 - c_2)\Lambda^\alpha)
\end{aligned}$$

for some C' depending on d . Plugging the bounds in Equations (6) and (8) into the bound in Equation (5) yields that

$$(9) \quad \left| \int_{\tilde{B}(0,\Lambda)} \mathcal{F}[\tilde{h}_{\chi^*,\Lambda}^*(\xi)] \left(\Phi_P(\xi) - \Phi_{P'}(\xi) \right) d\xi \right| \leq C_2 \|p_G - p_{G'}\|_1 \Lambda^d \exp(c_1 \Lambda^\alpha) \\ + C_3 \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \Lambda^{d+p} \exp((c_1 - c_2) \Lambda^\alpha),$$

where C_2 and C_3 are constants depending on R , λ_{\min} , λ_{\max} , and d . Since $\|p_G - p_{G'}\|_1 \gtrsim \Psi(\|\Sigma - \Sigma'\|_{\text{op}})$ and Ψ is strictly increasing, we have that $\Psi^{-1}(\|p_G - p_{G'}\|_1/C_0) \geq \|\Sigma - \Sigma'\|_{\text{op}}$ for some $C_0 > 0$.

Combining Equations (3) and (9) and noting that Ξ is also a strictly increasing function, we obtain that

$$W_1(P, P') \leq \frac{C_1}{\Lambda} + C_2 \|p_G - p_{G'}\|_1 \Lambda^d \exp(c_1 \Lambda^\alpha) \\ + C_3 (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1/C_0) \Lambda^{d+p} \exp((c_1 - c_2) \Lambda^\alpha),$$

where $C_1 := 2(1 + 2RM_\eta)M_K$. Choosing now $\Lambda = 2C_1/W_1(P, P')$ we arrive at the inequality

$$(10) \quad 1 \leq \left[C_2' V W_1(P, P')^{-(d+1)} + C_3' (\Xi \circ \Psi^{-1})(V/C_0) W_1(P, P')^{-(d+p+1)} \right] \\ \times \exp\left(C_4' W_1(P, P')^{-\alpha}\right),$$

where we set $V = \|p_G - p_{G'}\|_1$, $C_2' = 2C_2(2C_1)^d$, $C_3' = 2C_3(2C_1)^{d+p}$, and $C_4' = c_1(2C_1)^\alpha$. Using Lemma 2.9 for $p_1 = (d+1)/\alpha$ and $p_2 = (d+p+1)/\alpha$, there exist $M_{p_1,1}$ and $M_{p_2,1}$ depending only on p_1 and p_2 such that

$$W_1(P, P')^{-(d+1)} \leq M_{p_1,1} \exp(W_1(P, P')^{-\alpha}), \\ W_1(P, P')^{-(d+p+1)} \leq M_{p_2,1} \exp(W_1(P, P')^{-\alpha}).$$

Combining these results with (10), we have

$$1 \leq C_2'' M_{p_1,1} V \exp((1 + C_4') W_1(P, P')^{-\alpha}) \\ + C_3'' M_{p_2,1} (\Xi \circ \Psi^{-1})(V/C_0) \exp((1 + C_4') W_1(P, P')^{-\alpha}) \\ \leq C_2'' V \exp(C_{\text{new}} W_1(P, P')^{-\alpha}) + C_3'' (\Xi \circ \Psi^{-1})(V/C_0) \exp(C_{\text{new}} W_1(P, P')^{-\alpha}),$$

where $C_2'' = C_2' M_{p_1,1}$, $C_3'' = C_3' M_{p_2,1}$, and $C_{\text{new}} = 1 + C_4'$. This inequality implies either $1/2 \leq C_2'' V \exp(C_{\text{new}} W_1(P, P')^{-\alpha})$ or $1/2 \leq C_3'' (\Xi \circ \Psi^{-1})(V/C_0) \exp(C_{\text{new}} W_1(P, P')^{-\alpha})$. The first case implies

$$V \geq \frac{1}{2C_2''} \exp\left(-C_{\text{new}} \frac{1}{W_1(P, P')^\alpha}\right),$$

while the second case implies

$$V \geq C_0 (\Psi \circ \Xi^{-1}) \left(\frac{1}{2C_3''} \exp\left(-C_{\text{new}} \frac{1}{W_1(P, P')^\alpha}\right) \right).$$

As a result, we have

$$V \geq C_1 \min \left(\exp\left(-\frac{C}{W_1(P, P')^\alpha}\right), (\Psi \circ \Xi^{-1}) \left(\tilde{C} \exp\left(-\frac{C}{W_1(P, P')^\alpha}\right) \right) \right),$$

where $C = C_{\text{new}}$, $C_1 = \min\{1, 1/2C_2'', C_0\}$, and $\tilde{C} = 1/2C_3''$. Combining with Assumption 1, this completes the proof.

Proof of Theorem 3.4. Following the same duality, extension, cutoff, and spectral band-limiting arguments as in the proof of Theorem 3.2, we arrive at

$$(11) \quad W_1(P, P') \leq \frac{C_1}{\Lambda} + \frac{1}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*](\xi) \left(\Phi_P(\xi) - \Phi_{P'}(\xi) \right) d\xi,$$

where $\Lambda > 0$ is to be chosen later, $\tilde{h}_{\mathcal{X}, \Lambda}^*$ is defined in the proof of Theorem 3.2, and $C_1 := 2(1 + 2RM_\eta)M_K$. Using the same decomposition as in Equation (5), recalling that $\Phi_{p_G}(\xi) = \Phi_P(\xi)\Phi_\Sigma(\xi)$ and setting $C(R, d) := \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$, we have

$$(12) \quad \left| \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*](\xi) \left(\Phi_P(\xi) - \Phi_{P'}(\xi) \right) d\xi \right| \\ \leq \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi + \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi.$$

We now bound each term separately.

As for the first term, since $|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)| \leq \|p_G - p_{G'}\|_1$ for all $\xi \in \mathbb{R}^d$, and the ordinary-smoothness assumption gives $|\Phi_\Sigma(\xi)| \geq c_1(1 + \|\xi\|^\beta)^{-1}$, an application of Lemma 2.5 yields

$$(13) \quad \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \leq \frac{C(R, d)}{c_1(2\pi)^d} \|p_G - p_{G'}\|_1 \int_{\bar{B}(0, \Lambda)} (1 + \|\xi\|^\beta) d\xi \\ \leq C'_2 \|p_G - p_{G'}\|_1 \max\{\Lambda^{d+\beta}, \Lambda^d\},$$

where C'_2 is a constant depending on R , d , and c_1 .

As for the second term, from Assumption 2,

$$|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| \lesssim \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \|\xi\|^p \max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|).$$

Moreover, by ordinary-smoothness, $\max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|) \leq c_2(1 + \|\xi\|^\beta)^{-1}$, so that

$$\frac{\max(|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|)}{|\Phi_\Sigma(\xi)|} \leq \frac{c_2(1 + \|\xi\|^\beta)^{-1}}{c_1(1 + \|\xi\|^\beta)^{-1}} = \frac{c_2}{c_1}.$$

Thus, another application of Lemma 2.5 yields

$$(14) \quad \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)|}{|\Phi_\Sigma(\xi)|} d\xi \lesssim \frac{c_2}{c_1} \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \int_{\bar{B}(0, \Lambda)} \|\xi\|^p d\xi \\ = C'_3 \Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \Lambda^{d+p},$$

where C'_3 is a constant depending on R , d , λ_{\min} , λ_{\max} , and p . Substituting Equations (13) and (14) into Equation (12), and in turn into (11), we arrive at the bound

$$(15) \quad W_1(P, P') \leq \frac{C_1}{\Lambda} + C'_2 \|p_G - p_{G'}\|_1 \max\{\Lambda^{d+\beta}, \Lambda^d\} \\ + C'_3 (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1 / C_0) \Lambda^{d+p},$$

where we used $\Xi(\|\Sigma - \Sigma'\|_{\text{op}}) \leq (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1 / C_0)$, which follows from $\|p_G - p_{G'}\|_1 \geq C_0 \Psi(\|\Sigma - \Sigma'\|_{\text{op}})$ for some C_0 (by Assumption 1) and the fact that both Ψ and Ξ are strictly increasing. We now distinguish two cases, based on the size of $W_1(P, P')$.

Case 1: $W_1(P, P') \geq 2C_1$. Setting $\Lambda = 1$ in Equation (15) and using $W_1(P, P') - C_1 \geq W_1(P, P')/2$, we get

$$W_1(P, P') \leq 2C_2' \|p_G - p_{G'}\|_1 + 2C_3' (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1 / C_0).$$

Hence either $W_1(P, P') \leq 4C_2' \|p_G - p_{G'}\|_1$ or $W_1(P, P') \leq 4C_3' (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1 / C_0)$. Since $\text{supp}(P), \text{supp}(P') \subseteq \bar{B}(0, R)$ implies $W_1(P, P') \leq 2R$, we have $W_1(P, P') \geq (2R)^{-(d+\beta)} W_1(P, P')^{d+\beta+1}$ and similarly with $d+p$ in place of $d+\beta$. Combining these observations, there exist constants C_1' and c_1' depending on $R, d, \lambda_{\min}, \lambda_{\max}$, and p such that

$$(16) \quad \|p_G - p_{G'}\|_1 \geq C_1' \min \left\{ W_1(P, P')^{d+\beta+1}, (\Psi \circ \Xi^{-1})(c_1' W_1(P, P')^{d+p+1}) \right\}.$$

Case 2: $W_1(P, P') < 2C_1$. Setting $\Lambda = 2C_1/W_1(P, P')$ in Equation (15) gives

$$1 \leq C_{2,\text{II}}' \|p_G - p_{G'}\|_1 W_1(P, P')^{-(d+\beta+1)} \\ + C_{3,\text{II}}' (\Xi \circ \Psi^{-1})(\|p_G - p_{G'}\|_1 / C_0) W_1(P, P')^{-(d+p+1)},$$

where $C_{2,\text{II}}' = 2C_2'(2C_1)^{d+\beta}$ and $C_{3,\text{II}}' = 2C_3'(2C_1)^{d+p}$. Hence either the first or second summand is at least $1/2$, yielding respectively $\|p_G - p_{G'}\|_1 \geq W_1(P, P')^{d+\beta+1}/2C_{2,\text{II}}'$ or $\|p_G - p_{G'}\|_1 \geq C_0 (\Psi \circ \Xi^{-1})(W_1(P, P')^{d+p+1}/2C_{3,\text{II}}')$. Combining, there exists a constant $C > 0$ such that

$$(17) \quad \|p_G - p_{G'}\|_1 \gtrsim \min \left\{ W_1(P, P')^{d+\beta+1}, (\Psi \circ \Xi^{-1})(C W_1(P, P')^{d+p+1}) \right\}.$$

Combining (16) and (17) with Assumption 1, we conclude that

$$\|p_G - p_{G'}\|_1 \gtrsim \min \left\{ W_1(P, P')^{d+\beta+1}, (\Psi \circ \Xi^{-1})(C W_1(P, P')^{d+p+1}) \right\} + \Psi(\|\Sigma - \Sigma'\|_{\text{op}}),$$

where the implicit constant and C depend on $R, \lambda_{\min}, \lambda_{\max}$, and d . This completes the proof.

Proof of Proposition 3.6. Let $P = \delta_0$, so that $G = \delta_0 \times \delta_\Sigma \in \mathcal{G}$, and the mixture density p_G corresponds to

$$p_G(x) = \int_{\mathbb{R}^d} f(x | \theta, \Sigma) \delta_0(d\theta) = f(x | 0, \Sigma), \quad x \in \mathbb{R}^d.$$

By the PDE inversion condition, we have $\mathcal{T}_\Sigma f(\cdot | 0, \Sigma) = \delta_0$. Applying the Fourier transform to both sides yields $\mathcal{F}[\mathcal{T}_\Sigma f(\cdot | 0, \Sigma)](\xi) = 1$, for all $\xi \in \mathbb{R}^d$. Consequently, we have, for all $\xi \in \mathbb{R}^d$, that

$$1 = \mathcal{F}[\mathcal{T}_\Sigma f(\cdot | 0, \Sigma)](\xi) \\ = \left(\sum_{\nu \in \mathbb{N}^d: 0 \leq |\nu| \leq \beta} c_\nu(\Sigma) (-i\xi)^\nu \right) \mathcal{F}[f(\cdot | 0, \Sigma)](\xi) \\ = \left(\sum_{\nu \in \mathbb{N}^d: 0 \leq |\nu| \leq \beta} c_\nu(\Sigma) (-i\xi)^\nu \right) \Phi_\Sigma(-\xi).$$

Let Q_Σ denote the characteristic polynomial of the differential operator \mathcal{T}_Σ , namely,

$$(18) \quad Q_\Sigma(\xi) := \sum_{\nu \in \mathbb{N}^d: 0 \leq |\nu| \leq \beta} c_\nu(\Sigma) (-i\xi)^\nu, \quad \xi \in \mathbb{R}^d.$$

Noting that $\Phi_\Sigma(\cdot)$ is an even function because, by assumption, $f(\cdot | 0, \Sigma)$ is even, we conclude that $Q_\Sigma(\xi)\Phi_\Sigma(\xi) = 1$ for all $\xi \in \mathbb{R}^d$, which in turn implies that $|\Phi_\Sigma(\xi)| = |Q_\Sigma(\xi)|^{-1}$ for all $\xi \in \mathbb{R}^d$.

We first prove the lower bound $|\Phi_\Sigma(\xi)| \geq c_1(1 + \|\xi\|^\beta)^{-1}$ for some $c_1 > 0$, which is equivalent to showing $|Q_\Sigma(\xi)| \leq C_{\text{upper}}(1 + \|\xi\|^\beta)$ for some $C_{\text{upper}} > 0$. An application of the triangle inequality to the polynomial $Q_\Sigma(\xi)$ leads to

$$|Q_\Sigma(\xi)| \leq \sum_{\nu \in \mathbb{N}^d: 0 \leq |\nu| \leq \beta} |c_\nu(\Sigma)| \cdot \|\xi\|^{|\nu|}.$$

Given the PDE inversion condition, the coefficients $c_\nu(\Sigma)$ are continuous mappings from the set of positive-definite matrices $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ to \mathbb{R} . Because $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$ is a closed and bounded compact metric space, any continuous function defined on it is globally bounded. Thus, there exists a constant $M > 0$, depending only on λ_{\min} and λ_{\max} , such that $\sup_\Sigma |c_\nu(\Sigma)| \leq M$ for all $\nu \in \mathbb{N}^d$.

Furthermore, for any integer degree $0 \leq |\nu| \leq \beta$, we have

$$\|\xi\|^{|\nu|} \leq 1 + \|\xi\|^\beta \quad \text{for all } \xi \in \mathbb{R}^d.$$

Therefore, we obtain that

$$|Q_\Sigma(\xi)| \leq M \sum_{\nu \in \mathbb{N}^d: 0 \leq |\nu| \leq \beta} (1 + \|\xi\|^\beta) = M \binom{\beta + d}{d} (1 + \|\xi\|^\beta) =: C_{\text{upper}}(1 + \|\xi\|^\beta).$$

Next, we prove the upper bound $|\Phi_\Sigma(\xi)| \leq c_2(1 + \|\xi\|^\beta)^{-1}$ for some $c_2 > 0$, which is equivalent to proving $|Q_\Sigma(\xi)| \geq C_{\text{lower}}(1 + \|\xi\|^\beta)$ for some $C_{\text{lower}} > 0$. As $|\Phi_\Sigma(\xi)| \leq 1$ for all $\xi \in \mathbb{R}^d$, it follows that

$$|Q_\Sigma(\xi)| = |\Phi_\Sigma(\xi)|^{-1} \geq 1.$$

Recall that $f(x | 0, \Sigma) = |\Sigma|^{-1/2} g(x^\top \Sigma^{-1} x)$. Using the change of variables $y = \Sigma^{-1/2} x$, we have that

$$\Phi_\Sigma(\xi) = \int_{\mathbb{R}^d} e^{i\xi^\top (\Sigma^{1/2} y)} g(y^\top y) dy = \int_{\mathbb{R}^d} e^{i(\Sigma^{1/2} \xi)^\top y} g(\|y\|^2) dy.$$

Define $I(v) := \int_{\mathbb{R}^d} e^{iv^\top y} g(\|y\|^2) dy$, where $v = \Sigma^{1/2} \xi$. For any orthogonal matrix O , we have

$$I(Ov) = \int_{\mathbb{R}^d} e^{i(Ov)^\top y} g(\|y\|^2) dy = \int_{\mathbb{R}^d} e^{iv^\top (O^\top y)} g(\|y\|^2) dy = I(v).$$

This rotational invariance means there exists a function ψ such that $I(v) = \psi(\|v\|^2)$. Equivalently,

$$\Phi_\Sigma(\xi) = \psi(\|\Sigma^{1/2} \xi\|^2) = \psi(\xi^\top \Sigma \xi).$$

Therefore, we obtain $Q_\Sigma(\xi) = 1/\psi(\xi^\top \Sigma \xi)$. By setting $y = \Sigma^{1/2} \xi$, it is clear that the function $\tilde{Q}(y) := Q_\Sigma(\Sigma^{-1/2} y) = 1/\psi(\|y\|^2)$ is a multivariate polynomial in y that is invariant under any orthogonal rotation matrix O , since $\|Oy\|^2 = \|y\|^2$. Hence, it is a polynomial strictly in $\|y\|^2$. Denoting $\beta = 2m$, we can represent the polynomial $Q_\Sigma(\xi)$ as follows:

$$Q_\Sigma(\xi) = \sum_{k=0}^m a_k(\Sigma) (\xi^\top \Sigma \xi)^k,$$

where all odd-degree coefficients disappear. From the formulation of Q_Σ in Equation (18), we have $a_k(\Sigma) \in \mathbb{R}$ for $0 \leq k \leq m$. By the first assumption in the PDE inversion condition,

$a_m(\Sigma) \neq 0$. Furthermore, the fact that $|Q_\Sigma(\xi)| \geq 1$ for all ξ implies that the leading coefficient must be strictly positive, $a_m(\Sigma) > 0$.

Since the original coefficients $c_\nu(\Sigma)$ are continuous on the compact metric space $\mathbb{S}_d^+(\lambda_{\min}, \lambda_{\max})$, the derived polynomial coefficients $a_k(\Sigma)$ are also continuous on this compact set. Thus, the leading coefficient attains a strictly positive minimum $a_m^* := \min_\Sigma a_m(\Sigma) > 0$, and the remaining coefficients attain finite maximum magnitudes $A_k^* := \max_\Sigma |a_k(\Sigma)| < \infty$. We now establish a uniform lower bound for $|Q_\Sigma|$. An application of the reverse triangle inequality leads to

$$|Q_\Sigma(\xi)| \geq a_m(\Sigma) \left| \xi^\top \Sigma \xi \right|^m - \sum_{k=0}^{m-1} |a_k(\Sigma)| |\xi^\top \Sigma \xi|^k.$$

Since $\lambda_{\min} \|\xi\|^2 \leq \xi^\top \Sigma \xi \leq \lambda_{\max} \|\xi\|^2$, it holds that

$$|Q_\Sigma(\xi)| \geq a_m^* \lambda_{\min}^m \|\xi\|^{2m} - \sum_{k=0}^{m-1} A_k^* \lambda_{\max}^k \|\xi\|^{2k}.$$

Hence, for $\|\xi\| \geq 1$, we have

$$|Q_\Sigma(\xi)| \geq M_1 \|\xi\|^\beta - M_2 \|\xi\|^{\beta-2} = \|\xi\|^\beta \left(M_1 - \frac{M_2}{\|\xi\|^2} \right),$$

where $M_1 := a_m^* \lambda_{\min}^m$ and $M_2 := \sum_{k=0}^{m-1} A_k^* \lambda_{\max}^k$ are strictly positive constants independent of Σ . By choosing $\|\xi\| \geq \max \left\{ \sqrt{2M_2/M_1}, 1 \right\}$, we obtain that

$$|Q_\Sigma(\xi)| \geq \frac{M_1}{2} \|\xi\|^\beta.$$

Combining this bound with the fact that $|Q_\Sigma(\xi)| \geq 1$ for all ξ , we arrive at

$$|Q_\Sigma(\xi)| \geq C_{\text{lower}}(1 + \|\xi\|^\beta),$$

for a sufficiently small constant $C_{\text{lower}} > 0$ depending only on λ_{\min} , λ_{\max} , and the operator \mathcal{T}_Σ . Putting these results together leads to the conclusion that the kernel f is an ordinary-smooth function of order β .

Proof of Theorem 3.7. Let \bar{h}_χ^* be the 1-Lipschitz continuous function achieving the Kantorovich–Rubinstein infimum, constructed as in the proof of Theorem 3.2, and set $\tilde{h}_\chi^*(x) := \bar{h}_\chi^*(x)\eta(x)$, with η denoting the cutoff function constructed in Lemma 2.2. Rather than convolving with the Jackson kernel, we convolve \tilde{h}_χ^* with a scaled normalized mollifier in order to obtain a smooth proxy amenable to differentiation under the integral. Concretely, let ψ be the mollifier defined in Lemma 2.2, and set $\psi_\Lambda(x) := \Lambda^d \psi(\Lambda x)$ for $\Lambda > 0$.³ Define

$$\tilde{h}_{\chi, \Lambda}^*(x) := (\tilde{h}_\chi^* * \psi_\Lambda)(x) = \int_{\mathbb{R}^d} \tilde{h}_\chi^*(x-y) \psi_\Lambda(y) dy.$$

Since \tilde{h}_χ^* has support in $\bar{B}(0, 2R)$ and ψ_Λ has support in $\bar{B}(0, 1/\Lambda)$, we have $\text{supp}(\tilde{h}_{\chi, \Lambda}^*) \subseteq \bar{B}(0, 2R + 1/\Lambda)$. We split the rest of the proof into several steps.

³Compared to Lemma 2.2, this is a slight abuse of notation, as in the Lemma $\psi_\epsilon(x)$ is defined as $\epsilon^{-d} \psi(x/\epsilon)$.

Approximation bound. Direct computation gives

$$\tilde{h}_{\mathcal{X},\Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x) = \int_{\mathbb{R}^d} [\tilde{h}_{\mathcal{X}}^*(x-y) - \tilde{h}_{\mathcal{X}}^*(x)] \psi_{\Lambda}(y) dy, \quad x \in \mathbb{R}^d.$$

Since $\tilde{h}_{\mathcal{X}}^*$ is globally $(1 + 2RM_{\eta})$ -Lipschitz continuous (Lemma 2.3), for any $x \in \mathbb{R}^d$ we have

$$|\tilde{h}_{\mathcal{X},\Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x)| \leq (1 + 2RM_{\eta}) \int_{\mathbb{R}^d} \|y\| \Lambda^d \psi(\Lambda y) dy = \frac{(1 + 2RM_{\eta})M'_{\psi}}{\Lambda},$$

where $M'_{\psi} := \int_{\mathbb{R}^d} \|y\| \psi(y) dy < \infty$. Hence,

$$(19) \quad \|\tilde{h}_{\mathcal{X},\Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_{\infty} \leq \frac{(1 + 2RM_{\eta})M'_{\psi}}{\Lambda}.$$

Applying the same Hölder argument as in the proof of Theorem 3.2 and setting $C'_1 := 2(1 + 2RM_{\eta})M'_{\psi}$, we obtain

$$(20) \quad W_1(P, P') \leq \frac{C'_1}{\Lambda} + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta) (P(d\theta) - P'(d\theta)).$$

PDE inversion. Since f satisfies the PDE inversion condition of order β , we have $\mathcal{T}_{\Sigma} p_G = P$ and $\mathcal{T}_{\Sigma'} p_{G'} = P'$ in the sense of distributions.⁴ Therefore,

$$\begin{aligned} \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X},\Lambda}^*(\theta) (P(d\theta) - P'(d\theta)) &= \langle \tilde{h}_{\mathcal{X},\Lambda}^*, \mathcal{T}_{\Sigma} p_G - \mathcal{T}_{\Sigma'} p_{G'} \rangle \\ &= \langle \tilde{h}_{\mathcal{X},\Lambda}^*, \mathcal{T}_{\Sigma} (p_G - p_{G'}) \rangle + \langle \tilde{h}_{\mathcal{X},\Lambda}^*, (\mathcal{T}_{\Sigma} - \mathcal{T}_{\Sigma'}) p_{G'} \rangle \\ &= \langle \mathcal{T}_{\Sigma}^* \tilde{h}_{\mathcal{X},\Lambda}^*, p_G - p_{G'} \rangle + \langle (\mathcal{T}_{\Sigma}^* - \mathcal{T}_{\Sigma'}^*) \tilde{h}_{\mathcal{X},\Lambda}^*, p_{G'} \rangle \\ &=: I_1 + I_2, \end{aligned}$$

where the use of the adjoint operators is justified by the fact that $\tilde{h}_{\mathcal{X},\Lambda}^* \in C_c^{\infty}(\mathbb{R}^d)$.

Bound on I_1 . By Hölder's inequality,

$$(21) \quad |I_1| \leq \|p_G - p_{G'}\|_1 \|\mathcal{T}_{\Sigma}^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty}.$$

We bound $\|\mathcal{T}_{\Sigma}^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty}$ by the triangle inequality:

$$\|\mathcal{T}_{\Sigma}^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty} \leq |c_0(\Sigma)| \|\tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty} + \sum_{\nu \in \mathbb{N}^d: 1 \leq |\nu| \leq \beta} |c_{\nu}(\Sigma)| \|\partial^{\nu} \tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty}.$$

Applying the Fourier transform to both sides of $\mathcal{T}_{\Sigma} p_G = P$ and evaluating at $\xi = 0$ (where $\Phi_{p_G}(0) = \Phi_P(0) = 1$) gives $c_0(\Sigma) \equiv 1$. By Young's inequality and $\|\tilde{h}_{\mathcal{X}}^*\|_{\infty} \leq 2R$ (see the proof of Lemma 2.3), we have $\|\tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty} \leq 2R$. For the higher-order terms, we claim that $\|\partial^{\nu} \tilde{h}_{\mathcal{X},\Lambda}^*\|_{\infty} \leq C_{\text{der}} \Lambda^{|\nu|-1}$ for all $\nu \in \mathbb{N}^d$ such that $1 \leq |\nu| \leq \beta$ and some $C_{\text{der}} < \infty$. Indeed, fixing any canonical direction e_m with $\nu_m \geq 1$ and writing $\nu' = \nu - e_m$, the convolution structure gives

$$\partial^{\nu} \tilde{h}_{\mathcal{X},\Lambda}^* = (\partial_{x_m} \tilde{h}_{\mathcal{X}}^*) * (\partial^{\nu'} \psi_{\Lambda}).$$

⁴For standard references on distribution theory, see [3, 9, 10].

By Young's inequality and $\|\nabla \tilde{h}_{\mathcal{X}}^*\|_\infty \leq 1 + 2RM_\eta$ (Lemma 2.3),

$$\|\partial^{\nu'} \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty \leq (1 + 2RM_\eta) \|\partial^{\nu'} \psi_\Lambda\|_1.$$

A change of variables and the chain rule give $\|\partial^{\nu'} \psi_\Lambda\|_1 = \Lambda^{|\nu'|} \|\partial^{\nu'} \psi\|_1 \leq C_{\psi,\beta} \Lambda^{|\nu'|-1}$, where $C_{\psi,\beta} := \max_{\nu \in \mathbb{N}^d: 1 \leq |\nu| \leq \beta} \|\partial^{\nu'} \psi\|_1 < \infty$. Hence $C_{\text{der}} := C_{\psi,\beta}(1 + 2RM_\eta)$.

Since the eigenvalues of Σ lie in $[\lambda_{\min}, \lambda_{\max}]$ and the coefficients $c_\nu(\Sigma)$ depend smoothly on Σ , we have $\max_{\nu \in \mathbb{N}^d: 1 \leq |\nu| \leq \beta} |c_\nu(\Sigma)| \leq M_\beta < \infty$ (where M_β is independent of Σ). Setting $C_1'' := C_{\text{der}} M_\beta \binom{d+\beta}{d} + 2R$, for $\Lambda \geq 1$ we obtain

$$\|\mathcal{T}_\Sigma^* \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty \leq C_1'' (1 + \Lambda^{\beta-1}),$$

and therefore

$$(22) \quad |I_1| \leq C_1'' (1 + \Lambda^{\beta-1}) \|p_G - p_{G'}\|_1.$$

Bound on I_2 . By Hölder's inequality,

$$(23) \quad |I_2| \leq \|(\mathcal{T}_\Sigma^* - \mathcal{T}_{\Sigma'}^*) \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty.$$

The Lipschitz continuity of the coefficients $\Sigma \mapsto c_\nu(\Sigma)$ gives $|c_\nu(\Sigma) - c_\nu(\Sigma')| \leq L \|\Sigma - \Sigma'\|_{\text{op}}$ uniformly in ν . Combining with the derivative bound obtained before and setting $C_2'' := CL \binom{d+\beta}{d}$, we get

$$(24) \quad \begin{aligned} \|(\mathcal{T}_\Sigma^* - \mathcal{T}_{\Sigma'}^*) \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty &\leq \sum_{\nu \in \mathbb{N}^d: 1 \leq |\nu| \leq \beta} |c_\nu(\Sigma) - c_\nu(\Sigma')| \|\partial^{\nu'} \tilde{h}_{\mathcal{X},\Lambda}^*\|_\infty \\ &\leq C_2'' \|\Sigma - \Sigma'\|_{\text{op}} (1 + \Lambda^{\beta-1}), \end{aligned}$$

and therefore $|I_2| \leq C_2'' \|\Sigma - \Sigma'\|_{\text{op}} (1 + \Lambda^{\beta-1})$.

Conclusion. Combining Equations (20), (22), and (24), and using $\|\Sigma - \Sigma'\|_{\text{op}} \leq \Psi^{-1}(\|p_G - p_{G'}\|_1/C_0)$ for some $C_0 > 0$ (by Assumption 1), we obtain, for $\Lambda \geq 1$,

$$(25) \quad W_1(P, P') \leq \frac{C_1'}{\Lambda} + C_1'' (1 + \Lambda^{\beta-1}) \|p_G - p_{G'}\|_1 + C_2'' \Psi^{-1}(\|p_G - p_{G'}\|_1/C_0) (1 + \Lambda^{\beta-1}).$$

We now distinguish two cases based on the size of $W_1(P, P')$.

Case 1: $W_1(P, P') \geq 2C_1'$. Setting $\Lambda = 1$ in Equation (25) and using $W_1(P, P') - C_1' \geq W_1(P, P')/2$, we get

$$W_1(P, P') \leq 4C_1'' \|p_G - p_{G'}\|_1 + 4C_2'' \Psi^{-1}(\|p_G - p_{G'}\|_1/C_0).$$

Hence either $W_1(P, P') \leq 8C_1'' \|p_G - p_{G'}\|_1$ or $W_1(P, P') \leq 8C_2'' \Psi^{-1}(\|p_G - p_{G'}\|_1/C_0)$. Since $\text{supp}(P), \text{supp}(P') \subseteq \bar{B}(0, R)$ implies $W_1(P, P') \leq 2R$, we have $W_1(P, P') \leq (2R)^{1-\beta} W_1(P, P')^\beta$. Combining, there exist constants C_1'' and c_1'' such that

$$(26) \quad \|p_G - p_{G'}\|_1 \geq C_1'' \min \left\{ W_1(P, P')^\beta, \Psi(c_1'' W_1(P, P')^\beta) \right\}.$$

Case 2: $W_1(P, P') < 2C'_1$. Setting $\Lambda = 2C'_1/W_1(P, P') > 1$ in Equation (25) and letting $C''_{1,\text{II}} := 4C''_1(2C'_1)^{\beta-1}$ and $C''_{2,\text{II}} := 4C''_2(2C'_1)^{\beta-1}$, we obtain

$$1 \leq C''_{1,\text{II}} W_1(P, P')^{-\beta} \|p_G - p_{G'}\|_1 + C''_{2,\text{II}} W_1(P, P')^{-\beta} \Psi^{-1}(\|p_G - p_{G'}\|_1/C_0).$$

Hence either the first or the second summand is at least $1/2$, yielding respectively $\|p_G - p_{G'}\|_1 \geq W_1(P, P')^\beta / (2C''_{1,\text{II}})$ or $\|p_G - p_{G'}\|_1 \geq C_0 \Psi(W_1(P, P')^\beta / (2C''_{2,\text{II}}))$. Combining, there exist constants C''_{II} and c''_{II} such that

$$(27) \quad \|p_G - p_{G'}\|_1 \geq C''_{\text{II}} \min \left\{ W_1(P, P')^\beta, \Psi(c''_{\text{II}} W_1(P, P')^\beta) \right\}.$$

Combining Equations (26) and (27) with Assumption 1, we conclude that

$$\|p_G - p_{G'}\|_1 \gtrsim \min \left\{ W_1(P, P')^\beta, \Psi(C W_1(P, P')^\beta) \right\} + \Psi(\|\Sigma - \Sigma'\|_{\text{op}}),$$

where the implicit constant and C depend on R , λ_{\min} , λ_{\max} , and d . This completes the proof.

Proof of Theorem 4.1. We split the proof in four parts, as follows.

Verification of Assumption 2. The characteristic function of the Gaussian kernel is $\Phi_\Sigma(\xi) = \exp(-\xi^\top \Sigma \xi / 2)$. Setting $\psi(t) := e^{-t/2}$, so that $\psi'(t) = -\psi(t)/2$, the mean value theorem gives a $\tau \in (0, 1)$ such that

$$\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi) = \psi(\xi^\top \Sigma' \xi) - \psi(\xi^\top \Sigma \xi) = \psi'(\xi^\top \Sigma(\tau) \xi) \cdot \xi^\top (\Sigma' - \Sigma) \xi,$$

where $\Sigma(\tau) = (1 - \tau)\Sigma' + \tau\Sigma$. Hence,

$$\begin{aligned} |\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| &\leq \frac{1}{2} \psi(\xi^\top \Sigma(\tau) \xi) \|\Sigma' - \Sigma\|_{\text{op}} \|\xi\|^2 \\ &\leq \frac{1}{2} \max \{ |\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)| \} \|\Sigma' - \Sigma\|_{\text{op}} \|\xi\|^2, \end{aligned}$$

so Assumption 2 holds with $p = 2$ and $\Xi(t) = t$.

Verification of Assumption 1: univariate case. Consider two univariate mixture densities

$$p_G(x) = \int f(x | \theta, \sigma^2) P(d\theta) \quad \text{and} \quad p_{G'}(x) = \int f(x | \theta, \sigma'^2) P'(d\theta),$$

where $f(\cdot | \theta, \sigma^2)$ denotes the univariate Gaussian kernel with mean θ and variance σ^2 . Without loss of generality, suppose $\sigma^2 \geq \sigma'^2$ and set $\gamma^2 := \sigma^2 - \sigma'^2$. By the semigroup property of Gaussian convolutions, $f(\cdot | 0, \sigma^2) = f(\cdot | 0, \gamma^2) * f(\cdot | 0, \sigma'^2)$, so that

$$p_G = P * f(\cdot | 0, \sigma^2) = (P * f(\cdot | 0, \gamma^2)) * f(\cdot | 0, \sigma'^2) =: \tilde{P} * f(\cdot | 0, \sigma'^2),$$

where \tilde{P} is the distribution of $X + \gamma Z$ with $X \sim P$ and $Z \sim f(\cdot | 0, 1)$, independent of each other.

For a positive integer k to be specified later, consider the Hermite polynomial

$$H_{2k}(x; \sigma'^2) := \mathbb{E}_{Z \sim f(\cdot | 0, 1)} [(x + i\sigma' Z)^{2k}].$$

By Lemma 2.6, $\int_{\mathbb{R}} H_{2k}(x; \sigma'^2) f(x | \theta, \sigma'^2) dx = \theta^{2k}$, so that

$$\Delta M_{2k} := \int \theta^{2k} \tilde{P}(d\theta) - \int \theta^{2k} P'(d\theta) = \int_{\mathbb{R}} H_{2k}(x; \sigma'^2) (p_G(x) - p_{G'}(x)) dx.$$

We now lower-bound ΔM_{2k} . Since \tilde{P} is the law of $X + \gamma Z$,

$$\int \theta^{2k} \tilde{P}(d\theta) = \sum_{j=0}^k \binom{2k}{2j} \mathbb{E}_P[X^{2k-2j}] \gamma^{2j} \mathbb{E}[Z^{2j}] \geq \gamma^{2k} (2k-1)!!.$$

Using $(2k)! \geq (2k/e)^{2k}$ and $k! \leq k^k$, we get $(2k-1)!! = \frac{(2k)!}{2^k k!} \geq (2k/e^2)^k$. Since P' is supported on $[-R, R]$, we have $\int \theta^{2k} P'(d\theta) \leq R^{2k}$, and therefore

$$(28) \quad \Delta M_{2k} \geq \left(\frac{2k}{e^2}\right)^k \gamma^{2k} - R^{2k}.$$

To obtain an upper-bound on ΔM_{2k} , set $V := \|p_G - p_{G'}\|_1$ and $\Lambda := R + L$, for a $L > 0$ to be specified. From the definition of H_{2k} ,

$$|H_{2k}(x; \sigma^2)| \leq 2^{k-1} (x^{2k} + \sigma^{2k} (2k-1)!!) \leq 2^{k-1} (x^{2k} + \lambda_{\max}^k 2^k k^k).$$

For any choice of $|x| > \Lambda$, $\theta \in [-R, R]$ and $\sigma \in [\lambda_{\min}, \lambda_{\max}]$, the Gaussian kernel satisfies $f(x | \theta, \sigma^2) \leq C_0 \exp(-(|x| - R)^2 / 2\lambda_{\max}^2)$ with $C_0 = (2\pi\lambda_{\min}^2)^{-1/2}$, so that $|p_G(x)|, |p_{G'}(x)| \leq C_0 \exp(-(|x| - R)^2 / 2\lambda_{\max}^2)$. Splitting the integral at $|x| = \Lambda$ and applying Lemma 2.7 to bound the tail, we obtain

$$(29) \quad \Delta M_{2k} \leq C_2^k (V + e^{-L^2/2\lambda_{\max}^2}) (R^{2k} + L^{2k} + k^k),$$

where $C_2 > 0$ depends only on λ_{\min} , λ_{\max} , and R . Combining Equations (28) and (29) gives

$$(30) \quad k^k \gamma^{2k} \leq R_1^k + C_3^k (V + e^{-L^2/2\lambda_{\max}^2}) (L^{2k} + k^k + R^{2k}),$$

where $R_1 := (Re)^2/2$ and $C_3 := C_2 e^2/2$.

We now tune L and k . Set $L := \lambda_{\max} (2 \log(1/V))^{1/2}$, so that $e^{-L^2/2\lambda_{\max}^2} = V$ and

$$(31) \quad k^k \gamma^{2k} \leq R_1^k + 2C_3^k V (R^{2k} + (2\lambda_{\max}^2)^k (\log(1/V))^k + k^k).$$

Set $k := \lfloor \log(1/V) / \log \log(1/V) \rfloor$. If $V < e^{-e}$, one has $k \leq \log(1/V)$, so that $V(\log(1/V))^k \leq 1$ and $Vk^k \leq 1$. Substituting into (31) and absorbing all terms into a geometric bound yields

$$\gamma^2 \leq C_4 \frac{\log \log(1/V)}{\log(1/V)},$$

where $C_4 := R_1 + 4C_3 R^2 + 2C_3 + 4C_3 \lambda_{\max}^2$. If $V \leq e^{-e^{8C_4}}$ (so that $\log \log(1/V) \geq 8C_4$) then, using $e^t \geq t^2/2$, we have that $C_4 \log \log(1/V) \leq (\log(1/V))^{1/2}$, and thus $\gamma^2 \leq (\log(1/V))^{-1/2}$. Hence and

$$(32) \quad V \geq \exp\left(-2C_4 \frac{\log(1/\gamma^2)}{\gamma^2}\right),$$

whenever $V \leq C_5 := \min\{e^{-e}, e^{-e^{8C_4}}\}$.

On the other hand, if $V \geq C_5$, then, noting that the function $t \mapsto t \log t$ attains its minimum $-e^{-1}$ at $t = e^{-1}$, we have that

$$(33) \quad V \geq C_5 \geq C_6 \exp\left(-2C_4 \frac{\log(1/\gamma^2)}{\gamma^2}\right),$$

where $C_6 := C_5 e^{2C_4/e}$. Combining Equations (32) and (33) yields, with $M := 2C_4$ and $C' := \min\{1, C_6\}$,

$$(34) \quad \|p_G - p_{G'}\|_1 \geq C' \exp\left(-M \frac{\log(1/\gamma^2)}{\gamma^2}\right).$$

Thus, Assumption 1 is verified in the univariate case with $\Psi(t) = \exp(-M \log(1/t)/t)$.

Verification of Assumption 1: multivariate case. Let $u \in \mathbb{R}^d$ with $\|u\| = 1$ be the principal unit eigenvector of $\Sigma - \Sigma'$, so that $u^\top(\Sigma - \Sigma')u = \|\Sigma - \Sigma'\|_{\text{op}}$. Define the projected mixture densities

$$p_{G,u}(y) := \int f(y | u^\top \theta, u^\top \Sigma u) P(d\theta), \quad p_{G',u}(y) := \int f(y | u^\top \theta, u^\top \Sigma' u) P'(d\theta),$$

which represent the marginal densities of $u^\top X$ under p_G and $p_{G'}$ respectively. By Lemma 2.8,

$$\|p_G - p_{G'}\|_1 \geq \|p_{G,u} - p_{G',u}\|_1.$$

Setting $\sigma^2 := u^\top \Sigma u$, $\sigma'^2 := u^\top \Sigma' u$, and $\gamma^2 := |\sigma^2 - \sigma'^2| = \|\Sigma - \Sigma'\|_{\text{op}}$, the univariate bound (34) gives

$$\|p_{G,u} - p_{G',u}\|_1 \geq C' \exp\left(-M \frac{\log(1/\gamma^2)}{\gamma^2}\right) = C' \exp\left(-M \frac{\log(1/\|\Sigma - \Sigma'\|_{\text{op}})}{\|\Sigma - \Sigma'\|_{\text{op}}}\right),$$

which establishes Assumption 1 with $\Psi(t) = \exp(-M \log(1/t)/t)$.

Conclusion. With Assumptions 1 and 2 verified for $\Psi(t) = \exp(-M \log(1/t)/t)$, $\Xi(t) = t$, $p = 2$, and super-smoothness of order $\alpha = 2$ being obvious given the form of the Gaussian characteristic function, an application of Theorem 3.2 completes the proof.

Proof of Theorem 4.2. Before delving into the details of the proof, we remark that the key challenge in generalizing the argument below to the case of general Σ and Σ' matrices is due to the fact that, if we were to move from Σ to Σ' using the semi-group property, we would need to guarantee that $\Sigma - \Sigma'$ is a positive definite matrix, which is not always true unless we have further knowledge on the structure of Σ . For instance, as we do below, $\Sigma = \sigma^2 I_d$ works to this end, as positive-definiteness of $\Sigma - \Sigma'$ can be guaranteed by assuming, without loss of generality, that $\sigma > \sigma'$.

Therefore, without loss of generality, assume that $\sigma > \sigma'$. Define the variance gap $\Delta := \sigma^2 - \sigma'^2 > 0$ and $\tilde{P} = P * f(\cdot | 0, \Delta \cdot I_d)$, so that \tilde{P} is a continuous probability measure with a smooth density given by

$$\tilde{P}(dy) = \left(\int_{\mathbb{R}^d} \frac{1}{(2\pi\Delta)^{d/2}} \exp\left(-\frac{\|y - \theta\|^2}{2\Delta}\right) P(d\theta) \right) dy.$$

Due to the Gaussian semigroup property, we have $p_G(x) \equiv \int f(x | \theta, \sigma'^2 I_d) \tilde{P}(d\theta)$. To estimate $W_1(P, \tilde{P})$, we consider the coupling (X, Y) such that $X \sim P$, $Y = X + Z$ where $Z \sim \mathcal{N}(0, \Delta \cdot I_d)$ and independent from X , which means that

$$W_1(P, \tilde{P}) \leq \mathbb{E}[\|Y - X\|_2] \leq \mathbb{E}[\|Z\|_2] \leq \sqrt{\mathbb{E}[\|Z\|_2^2]} = \sqrt{\text{Tr}(\Delta I_d)} = \sqrt{d\Delta}.$$

Therefore, an application of the triangle inequality leads to

$$(35) \quad W_1(P, P') \leq W_1(P, \tilde{P}) + W_1(\tilde{P}, P') \leq \sqrt{d\Delta} + W_1(\tilde{P}, P').$$

For $X \sim P'$, as P' has compact support, we have $\mathbb{E}[\|X\|] < \infty$. For $X \sim P$, $Z \sim \mathcal{N}(0, \Delta \cdot I_d)$ independent, we have $Y = X + Z \sim \tilde{P}$, thus $\mathbb{E}[\|Y\|] \leq \mathbb{E}[\|X\|] + \mathbb{E}[\|Z\|] < \infty$ as P has compact support. Thus, both two measures P and \tilde{P} has finite first moment. Using Lemma 2.10, there exists a function $h_{\mathcal{X}}^* \in \text{Lip}_1(\mathbb{R}^d)$, with $h_{\mathcal{X}}^*(0) = 0$, which is a solution of Kantorovich-Rubinstein dual problem:

$$W_1(\tilde{P}, P') = \int_{\mathbb{R}^d} h_{\mathcal{X}}^*(\theta) (\tilde{P}(d\theta) - P'(d\theta)).$$

Lemma 2.2 guarantees the existence of a cutoff function $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$ in $C_c^\infty(\mathbb{R}^d)$ which takes values in $[0, 1]$, evaluates identically to 1 on the closed ball $\bar{B}(0, R)$ (where P' is supported), and vanishes outside the open ball $B(0, 2R)$. Hence, given η , we denote

$$\tilde{h}_{\mathcal{X}}^*(x) := h_{\mathcal{X}}^*(x)\eta(x)$$

for all $x \in \mathbb{R}^d$. As we show in Lemma 2.3, $\tilde{h}_{\mathcal{X}}^* \in L_1(\mathbb{R}^d)$ and it is globally $(1 + 2RM_\eta)$ -Lipschitz continuous on \mathbb{R}^d , where $M_\eta = \|\nabla\eta\|_\infty$. By the properties of η and noting that the support of P and P' are in $B(0, R)$, we also clearly have that

$$\begin{aligned} & \int_{\mathbb{R}^d} h_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))(\tilde{P}(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d \setminus B(0, R)} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))(\tilde{P}(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d \setminus B(0, R)} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))\tilde{P}(d\theta) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d \setminus B(0, R)} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))(\tilde{P}(d\theta) - P(d\theta)) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)). \end{aligned}$$

It is clear that $h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta) = h_{\mathcal{X}}^*(\theta)(1 - \eta(\theta))$. Therefore, $\nabla(h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta)) = (1 - \eta(\theta))\nabla h_{\mathcal{X}}^*(\theta) - h_{\mathcal{X}}^*(\theta)\nabla\eta(\theta)$, so that

$$\|\nabla(h_{\mathcal{X}}^* - \tilde{h}_{\mathcal{X}}^*)\|_\infty \leq \|(1 - \eta)\nabla h_{\mathcal{X}}^*\|_\infty + \|h_{\mathcal{X}}^*\nabla\eta\|_\infty.$$

Since $h_{\mathcal{X}}^* \in \text{Lip}_1(\mathbb{R}^d)$, we have $|h_{\mathcal{X}}^*(\theta)| \leq \|\theta\|_2$. Furthermore, as η is strictly supported on $B(0, 2R)$, we have

$$\|h_{\mathcal{X}}^*\nabla\eta\|_\infty \leq 2RM_\eta.$$

Furthermore, as $\|\nabla h_{\mathcal{X}}^*\|_\infty \leq 1$ and $\eta \in [0, 1]$, we have

$$\|(1 - \eta)\nabla h_{\mathcal{X}}^*\|_\infty \leq 1.$$

Putting these bounds together, we arrive at

$$\|\nabla(h_{\mathcal{X}}^* - \tilde{h}_{\mathcal{X}}^*)\|_\infty \leq 1 + 2RM_\eta.$$

Therefore,

$$\begin{aligned} & \int_{\mathbb{R}^d \setminus B(0, R)} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))(\tilde{P}(d\theta) - P(d\theta)) = \int_{\mathbb{R}^d} (h_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}}^*(\theta))(\tilde{P}(d\theta) - P(d\theta)) \\ (36) \quad & \leq (1 + 2RM_\eta)W_1(\tilde{P}, P), \end{aligned}$$

where the final inequality is due to the Kantorovich-Rubinstein duality theorem while the equality is due to $h_{\mathcal{X}}^*(\theta) = \tilde{h}_{\mathcal{X}}^*(\theta)$ when $\theta \in B(0, R)$.

Moving to $\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta))$, we can bound this term using the spectral bandlimiting via the Jackson kernel as in the proof of Theorem 3.2. In particular, we consider the multivariate Jackson kernel [1, 2].

$$K(x) := \left(\frac{3}{8\pi\sqrt{d}}\right)^d \prod_{i=1}^d \text{sinc}^4\left(\frac{x_i}{4\sqrt{d}}\right), \quad x \in \mathbb{R}^d,$$

where $\text{sinc}(u) := \sin u/u$, which by construction satisfies $\int_{\mathbb{R}^d} K(x)dx = 1$. For any $\Lambda > 0$, define the function $K_\Lambda(x) := \Lambda^d K(\Lambda x)$. By Lemma 2.4, $\text{supp}(\mathcal{F}[K]) \subseteq \bar{B}(0, 1)$, which implies that $\text{supp}(\mathcal{F}[K_\Lambda]) \subseteq \bar{B}(0, \Lambda)$. Consider the function

$$\tilde{h}_{\mathcal{X}, \Lambda}^*(x) := (\tilde{h}_{\mathcal{X}}^* * K_\Lambda)(x) = \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x-y)K_\Lambda(y)dy.$$

Direct computation yields that

$$\begin{aligned} \tilde{h}_{\mathcal{X}, \Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x) &= \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x-y)K_\Lambda(y)dy - \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(x)K_\Lambda(y)dy \\ &= \int_{\mathbb{R}^d} [\tilde{h}_{\mathcal{X}}^*(x-y) - \tilde{h}_{\mathcal{X}}^*(x)]K_\Lambda(y)dy. \end{aligned}$$

Since $\tilde{h}_{\mathcal{X}}^*$ is a globally $(1 + 2RM_\eta)$ -Lipschitz continuous function, we further obtain for any $x \in \mathbb{R}^d$ that

$$\begin{aligned} |\tilde{h}_{\mathcal{X}, \Lambda}^*(x) - \tilde{h}_{\mathcal{X}}^*(x)| &\leq \int_{\mathbb{R}^d} |\tilde{h}_{\mathcal{X}}^*(x-y) - \tilde{h}_{\mathcal{X}}^*(x)|\Lambda^d K(\Lambda y)dy \\ &\leq (1 + 2RM_\eta) \int_{\mathbb{R}^d} \|y\|\Lambda^d K(\Lambda y)dy \\ &= \frac{(1 + 2RM_\eta)M_K}{\Lambda}, \end{aligned}$$

where the last inequality follows from Lemma 2.4. As a consequence, we arrive at

$$(37) \quad \|\tilde{h}_{\mathcal{X}, \Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty \leq \frac{(1 + 2RM_\eta)M_K}{\Lambda}.$$

Now, we have

$$\begin{aligned} &\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}}^*(\theta)(\tilde{P}(d\theta) - P'(d\theta)) \\ &= \int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta))(\tilde{P}(d\theta) - P'(d\theta)) + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta)d(\tilde{P}(d\theta) - P'(d\theta)). \end{aligned}$$

A direct application of Hölder's inequality leads to

$$\int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta))(\tilde{P}(d\theta) - P'(d\theta)) \leq \|\tilde{h}_{\mathcal{X}, \Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty \|\tilde{P} - P'\|_1 \leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda},$$

where the final inequality is due to $\|\tilde{P} - P'\|_1 \leq 2$ and the bound of $\|\tilde{h}_{\mathcal{X}, \Lambda}^* - \tilde{h}_{\mathcal{X}}^*\|_\infty$ in Equation (37). Hence, we obtain the following bound

$$\begin{aligned} &\int_{\mathbb{R}^d} (\tilde{h}_{\mathcal{X}}^*(\theta) - \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta))(\tilde{P}(d\theta) - P'(d\theta)) \\ &\leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda} + \int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta)d(\tilde{P}(d\theta) - P'(d\theta)). \end{aligned}$$

To apply the Fourier inversion theorem to $\int_{\mathbb{R}^d} \tilde{h}_{\mathcal{X}, \Lambda}^*(\theta)(P(d\theta) - P'(d\theta))$, we first need to demonstrate that $\tilde{h}_{\mathcal{X}, \Lambda}^* \in L_1(\mathbb{R}^d)$ and $\mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*] \in L_1(\mathbb{R}^d)$. Because $\tilde{h}_{\mathcal{X}, \Lambda}^*(x) = (\tilde{h}_{\mathcal{X}}^* * K_\Lambda)(x)$, an application of Young's inequality leads to $\|\tilde{h}_{\mathcal{X}, \Lambda}^*\|_1 \leq \|\tilde{h}_{\mathcal{X}}^*\|_1 \|K_\Lambda\|_1 = \|\tilde{h}_{\mathcal{X}}^*\|_1 < \infty$ (thanks to Lemma 2.3). As for $\mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*]$, by the properties of convolutions we obtain that $\mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*](\xi) = \mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)\mathcal{F}[K_\Lambda](\xi)$. As $\text{supp}(\mathcal{F}[K_\Lambda]) \subseteq \bar{B}(0, \Lambda)$, we have $\text{supp}(\mathcal{F}[\tilde{h}_{\mathcal{X}, \Lambda}^*]) \subseteq$

$\bar{B}(0, \Lambda)$. Furthermore, $\|\mathcal{F}[K_\Lambda]\|_\infty \leq 1$ and $\|\mathcal{F}[\tilde{h}_\chi^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$ (see Lemma 2.3). Therefore, $\|\mathcal{F}[\tilde{h}_{\chi, \Lambda}^*]\|_\infty \leq \|\mathcal{F}[\tilde{h}_\chi^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$. Putting these results together, we obtain that $\mathcal{F}[\tilde{h}_{\chi, \Lambda}^*] \in L_1(\mathbb{R}^d)$. Now, from the Fourier inversion theorem, as $\tilde{h}_{\chi, \Lambda}^* \in L_1(\mathbb{R}^d)$ and $\mathcal{F}[\tilde{h}_{\chi, \Lambda}^*] \in L_1(\mathbb{R}^d)$, we can write

$$\tilde{h}_{\chi, \Lambda}^*(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) e^{i\xi^\top x} d\xi,$$

leading to

$$\int_{\mathbb{R}^d} \tilde{h}_{\chi, \Lambda}^*(\theta) (\tilde{P}(d\theta) - P'(d\theta)) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) e^{i\xi^\top \theta} d\xi \right) (\tilde{P}(d\theta) - P'(d\theta)).$$

By Fubini's theorem, we have that

$$\begin{aligned} \int_{\mathbb{R}^d} \tilde{h}_{\chi, \Lambda}^*(\theta) (\tilde{P}(d\theta) - P'(d\theta)) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{i\xi^\top \theta} (\tilde{P}(d\theta) - P'(d\theta)) \right) \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) d\xi \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) (\Phi_{\tilde{P}}(\xi) - \Phi_{P'}(\xi)) d\xi \\ &= \frac{1}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) (\Phi_{\tilde{P}}(\xi) - \Phi_{P'}(\xi)) d\xi, \end{aligned}$$

where the final identity is due to the fact that $\text{supp}(\mathcal{F}[\tilde{h}_{\chi, \Lambda}^*]) \subseteq \bar{B}(0, \Lambda)$.

Putting all of these results together, we arrive at

$$\begin{aligned} &\int_{\mathbb{R}^d} (\tilde{h}_\chi^*(\theta) - \tilde{h}_{\chi, \Lambda}^*(\theta)) (\tilde{P}(d\theta) - P'(d\theta)) \\ (38) \quad &\leq \frac{2(1 + 2RM_\eta)M_K}{\Lambda} + \frac{1}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) (\Phi_{\tilde{P}}(\xi) - \Phi_{P'}(\xi)) d\xi. \end{aligned}$$

Next, since $p_G = P * f(\cdot | 0, \sigma^2 I_d)$, it is clear that $p_G = \tilde{P} * f(\cdot | 0, \sigma'^2 I_d)$. Therefore, for all $\xi \in \mathbb{R}^d$ we obtain that

$$\begin{aligned} \Phi_{\tilde{P}}(\xi) - \Phi_{P'}(\xi) &= \frac{\Phi_{p_G}(\xi)}{\Phi_{\sigma'^2 I_d}(\xi)} - \frac{\Phi_{p_{G'}}(\xi)}{\Phi_{\sigma'^2 I_d}(\xi)} \\ &= \frac{\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)}{\Phi_{\sigma'^2 I_d}(\xi)}. \end{aligned}$$

Since $\|\mathcal{F}[\tilde{h}_{\chi, \Lambda}^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1} =: C(R, d)$, we obtain that

$$(39) \quad \left| \int_{\bar{B}(0, \Lambda)} \mathcal{F}[\tilde{h}_{\chi, \Lambda}^*](\xi) (\Phi_{\tilde{P}}(\xi) - \Phi_{P'}(\xi)) d\xi \right| \leq \frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_{\sigma'^2 I_d}(\xi)|} d\xi.$$

By construction, $|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)| \leq \|p_G - p_{G'}\|_1$ for any $\xi \in \mathbb{R}^d$. Furthermore, $\Phi_{\sigma'^2 I_d}(\xi) = \exp\left(-\frac{\sigma'^2}{2}\|\xi\|^2\right)$. Therefore, we obtain that

$$\begin{aligned} &\frac{C(R, d)}{(2\pi)^d} \int_{\bar{B}(0, \Lambda)} \frac{|\Phi_{p_G}(\xi) - \Phi_{p_{G'}}(\xi)|}{|\Phi_{\sigma'^2 I_d}(\xi)|} d\xi \leq \frac{C(R, d)}{(2\pi)^d} \|p_G - p_{G'}\|_1 \int_{\bar{B}(0, \Lambda)} \exp\left(\frac{\sigma'^2}{2}\|\xi\|^2\right) d\xi \\ (40) \quad &\leq C_1 \|p_G - p_{G'}\|_1 \Lambda^d \exp\left(\frac{\sigma'^2}{2}\Lambda^2\right), \end{aligned}$$

where $C_1 = \frac{C(R,d)}{(4\pi)^{d/2}\Gamma(d/2+1)}$ is a constant depending on R and d .

Collecting the results from equations (35), (36), (38), (39), and (40), we arrive at

$$(41) \quad W_1(P, P') \leq (2 + 2RM_\eta)\sqrt{d}\sqrt{\Delta} + \frac{2(1 + 2RM_\eta)M_K}{\Lambda} + C_1\|p_G - p_{G'}\|_1\Lambda^d \exp\left(\frac{\sigma'^2}{2}\Lambda^2\right).$$

Let $V := \|p_G - p_{G'}\|_1$, $\bar{C}_2 = (2 + 2RM_\eta)\sqrt{d}$, $C_3 = 2(1 + 2RM_\eta)M_K$. Here we consider two situations.

Case 1: small V . Recall the following inequality for $\Delta = \sigma^2 - \sigma'^2$ from Theorem 4.1:

$$\|p_G - p_{G'}\|_1 \geq \exp\left(-M\frac{\log(1/\Delta)}{\Delta}\right),$$

which implies

$$\log(1/V) \leq M\frac{\log(1/\Delta)}{\Delta}.$$

For $V \leq \exp(-e^{1/M})$, using Lemma 2.11, we have

$$\frac{1}{\Delta} \geq \frac{1}{M} \frac{\log(1/V)}{\log \log(1/V)}.$$

Replacing this inequality into the inequality (41), we obtain that

$$W_1(P, P') \leq C_2\sqrt{\frac{\log \log(1/V)}{\log(1/V)}} + \frac{C_3}{\Lambda} + C_4V\Lambda^d \exp\left(\frac{\sigma'^2}{2}\Lambda^2\right),$$

where $C_2 = \sqrt{M} \cdot \bar{C}_2$. Let $N = \log(1/V)$. When $V \leq e^{-(d+1)^2}$, then we have $N > (d+1)\log N$, we can choose

$$\Lambda = \frac{\sqrt{2}}{\sigma'} \sqrt{N - \frac{d+1}{2} \log N}.$$

This choice of Λ gives us

$$\frac{1}{\sigma'}\sqrt{N} \leq \Lambda \leq \frac{\sqrt{2}}{\sigma'}\sqrt{N}.$$

Under this scenario,

$$\frac{C_3}{\Lambda} \leq \frac{C_3\sigma'}{\sqrt{N}} = \frac{C_3\sigma'}{\sqrt{\log(1/V)}}.$$

Furthermore, we have

$$\exp\left(\frac{\sigma'^2}{2}\Lambda^2\right) = \exp\left(N - \frac{d+1}{2} \log N\right) = \frac{e^N}{N^{(d+1)/2}}.$$

Therefore,

$$C_4V\Lambda^d \exp\left(\frac{\sigma'^2}{2}\Lambda^2\right) = C_4e^{-N} \left(N^{d/2}\right) \left(\frac{e^N}{N^{(d+1)/2}}\right) = C_4N^{-1/2} = \frac{C_4}{\sqrt{\log(1/V)}}.$$

As a result, we have

$$W_1(P, P') \leq C_2\sqrt{\frac{\log \log(1/V)}{\log(1/V)}} + \frac{C_6}{\sqrt{\log(1/V)}} \leq C_7\sqrt{\frac{\log \log(1/V)}{\log(1/V)}},$$

where $C_6 = C_4 + C_3\sigma'$, $C_7 = C_2 + C_6$, or

$$W_1^2(P, P') \leq C_7^2 \frac{\log \log(1/V)}{\log(1/V)}.$$

If $V \leq e^{-e^{8C_7^2}}$ (so that $\log \log(1/V) \geq 8C_7^2$) then, using $e^t \geq t^2/2$, we have that $C_7^2 \log \log(1/V) \leq (\log(1/V))^{1/2}$, and thus $W_1^2(P, P') \leq (\log(1/V))^{-1/2}$, which implies $2 \log(1/W_1^2(P, P')) \geq \log \log(1/V)$. Thus, for

$$V \leq C_5 := \min\{\exp(-e^{1/M}), e^{-(d+1)^2}, e^{-e^{8C_7^2}}\}$$

we achieve

$$(42) \quad V \geq \exp\left(-2C_7^2 \frac{\log(1/W_1^2(P, P'))}{W_1^2(P, P')}\right) = \exp\left(-4C_7^2 \frac{\log(1/W_1(P, P'))}{W_1^2(P, P')}\right)$$

Case 2: $V \geq C_5$. then noting that the function $t \mapsto t \log t$ attains its minimum $-e^{-1}$ at $t = e^{-1}$, we have that

$$(43) \quad V \geq C_5 \geq C_8 \exp\left(-4C_7^2 \frac{\log(1/W_1(P, P'))}{W_1^2(P, P')}\right),$$

where $C_8 := C_5 e^{-2C_7^2/e}$. Combining Equations (42) and (43) yields, with $M := 4C_7^2$ and $C' := \min\{1, C_8\}$,

$$(44) \quad \|p_G - p_{G'}\|_1 \geq C' \exp\left(-M \frac{\log(1/W_1(P, P'))}{W_1^2(P, P')}\right).$$

uniformly for all P and P' . This concludes the proof.

Proof of Theorem 4.3. We split the proof in four parts, as follows.

Verification of Assumption 2. The characteristic function of the Cauchy kernel is $\Phi_\Sigma(\xi) = \exp(-\sqrt{\xi^\top \Sigma \xi})$. Setting $\psi(t) := e^{-t}$, so that $\psi'(t) = -\psi(t)$, the mean value theorem gives a $\tau \in (0, 1)$ such that

$$\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi) = \psi'\left((\xi^\top \Sigma(\tau)\xi)^{1/2}\right) \left((\xi^\top \Sigma'\xi)^{1/2} - (\xi^\top \Sigma\xi)^{1/2}\right),$$

where $\Sigma(\tau) := (1 - \tau)\Sigma' + \tau\Sigma$. Using the inequality $|\sqrt{a} - \sqrt{b}| \leq \sqrt{|a - b|}$, we get

$$\begin{aligned} |\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| &\leq \psi\left((\xi^\top \Sigma(\tau)\xi)^{1/2}\right) |\xi^\top (\Sigma' - \Sigma)\xi|^{1/2} \\ &\leq \max\{|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|\} \|\Sigma' - \Sigma\|_{\text{op}}^{1/2} \|\xi\|, \end{aligned}$$

so Assumption 2 holds with $C = 1$, $p = 1$, and $\Xi(t) = \sqrt{t}$.

Verification of Assumption 1: univariate case. Consider two univariate mixture densities

$$p_G(x) = \int f(x | \theta, \sigma) P(d\theta) \quad \text{and} \quad p_{G'}(x) = \int f(x | \theta, \sigma') P'(d\theta),$$

where $f(\cdot | \theta, \sigma)$ is the univariate Cauchy kernel with location θ and scale σ . Without loss of generality, suppose $\sigma' > \sigma$. Setting $V := \|p_G - p_{G'}\|_1$, since $\mathcal{F}[p_G](\xi) = \Phi_P(\xi) \exp(-\sigma|\xi|)$,

$$\begin{aligned} V &\geq \text{Re}(\mathcal{F}[p_G](\xi)) - \text{Re}(\mathcal{F}[p_{G'}](\xi)) \\ &= \exp(-\sigma|\xi|) \int \cos(\xi\theta) P(d\theta) - \exp(-\sigma'|\xi|) \int \cos(\xi\theta) P'(d\theta). \end{aligned}$$

Using $1 - \xi^2\theta^2/2 \leq \cos(\xi\theta) \leq 1$ and $|\theta| \leq R$ for $\theta \in \text{supp}(P) \cup \text{supp}(P')$,

$$(45) \quad 1 - \frac{1}{2}R^2\xi^2 \leq \int \cos(\xi\theta) P(d\theta) \leq 1.$$

Hence, for all $\xi \in \mathbb{R}$,

$$(46) \quad \begin{aligned} V &\geq \exp(-\sigma|\xi|) \left(1 - \frac{1}{2}R^2\xi^2\right) - \exp(-\sigma'|\xi|) \\ &= \exp(-\sigma|\xi|) \left(1 - \exp(-(\sigma' - \sigma)|\xi|)\right) - \frac{1}{2}R^2\xi^2 \exp(-\sigma|\xi|). \end{aligned}$$

Applying $1 - e^{-x} \geq x - x^2/2$ for $x \geq 0$ and $e^{-\sigma|\xi|} \geq 1 - \sigma|\xi|$ to the first term, and $e^{-\sigma|\xi|} \leq 1$ to the second, we obtain for $|\xi| \leq 2/(\sigma' - \sigma)$,

$$V \geq |\xi|(\sigma' - \sigma) - \xi^2 \left(\sigma(\sigma' - \sigma) + \frac{1}{2}(\sigma' - \sigma)^2 + R^2\right).$$

The right-hand side is maximized at

$$|\xi| = \xi_0 := \frac{\sigma' - \sigma}{2\sigma(\sigma' - \sigma) + (\sigma' - \sigma)^2 + 2R^2},$$

which satisfies $\xi_0 \leq 2/(\sigma' - \sigma)$. Substituting and simplifying,

$$V \geq \frac{(\sigma' - \sigma)^2}{4\sigma(\sigma' - \sigma) + 2(\sigma' - \sigma)^2 + 4R^2} \geq \frac{(\sigma' - \sigma)^2}{6\lambda_{\max}^2 + 4R^2} =: C_0(\sigma' - \sigma)^2.$$

This confirms Assumption 1 in the univariate case with $\Psi(t) = t^2$.

Verification of Assumption 1: multivariate case. Let $u \in \mathbb{R}^d$ be the principal eigenvector (thus, of unit norm) of $\Sigma - \Sigma'$, so that $u^\top(\Sigma - \Sigma')u = \|\Sigma - \Sigma'\|_{\text{op}}$. Define the projected mixture densities

$$\begin{aligned} p_{G,u}(y) &:= \int f(y | u^\top \theta, (u^\top \Sigma u)^{1/2}) P(d\theta), \\ p_{G',u}(y) &:= \int f(y | u^\top \theta, (u^\top \Sigma' u)^{1/2}) P'(d\theta), \end{aligned}$$

which correspond to the marginal densities of $u^\top X$ under p_G and $p_{G'}$ respectively. By Lemma 2.8,

$$\|p_G - p_{G'}\|_1 \geq \|p_{G,u} - p_{G',u}\|_1.$$

Setting $\sigma := (u^\top \Sigma u)^{1/2}$ and $\sigma' := (u^\top \Sigma' u)^{1/2}$, the univariate bound gives $\|p_{G,u} - p_{G',u}\|_1 \geq C_0(\sigma' - \sigma)^2$. Then,

$$|\sigma' - \sigma| = \frac{|\sigma' - \sigma|(\sigma' + \sigma)}{\sigma' + \sigma} = \frac{|u^\top(\Sigma' - \Sigma)u|}{\sigma' + \sigma} \geq \frac{\|\Sigma' - \Sigma\|_{\text{op}}}{2\lambda_{\max}^{1/2}},$$

from which it follows that

$$\|p_G - p_{G'}\|_1 \geq \frac{C_0}{4\lambda_{\max}} \|\Sigma - \Sigma'\|_{\text{op}}^2,$$

thus establishing Assumption 1 with $\Psi(t) = t^2$.

Conclusion. With Assumptions 1 and 2 verified for $\Psi(t) = t^2$, $\Xi(t) = \sqrt{t}$ and $p = 1$, and since the super-smoothness condition of order $\alpha = 1$ is automatically verified given the form of the Cauchy characteristic function, an application of Theorem 3.2 completes the proof.

Proof of Theorem 4.4. We split the proof into five parts, as follows.

PDE inversion condition. The characteristic function of the multivariate Laplace kernel is $\xi \in \mathbb{R}^d \mapsto \Phi_\Sigma(\xi) = (1 + \frac{1}{2}\xi^\top \Sigma \xi)^{-1}$, so $\Phi_\Sigma(\xi)^{-1} = 1 + \frac{1}{2}\xi^\top \Sigma \xi$. By the linearity of the Fourier transform and the identity $\mathcal{F}[\partial^\nu p_G](\xi) = (-i)^{|\nu|} \xi^\nu \mathcal{F}[p_G](\xi)$,

$$\begin{aligned} \mathcal{F}[\mathcal{T}_\Sigma p_G](\xi) &= \mathcal{F}[p_G](\xi) - \frac{1}{2} \sum_{j=1}^d \sum_{k=1}^d \Sigma_{jk} (-\xi_j \xi_k) \mathcal{F}[p_G](\xi) = \left(1 + \frac{1}{2}\xi^\top \Sigma \xi\right) \mathcal{F}[p_G](\xi) \\ &= \Phi_\Sigma(\xi)^{-1} \mathcal{F}[p_G](\xi) = \mathcal{F}[P](\xi), \end{aligned}$$

where the last equality uses $\mathcal{F}[p_G](\xi) = \Phi_P(\xi) \Phi_\Sigma(\xi)$. By uniqueness of the Fourier transform, $\mathcal{T}_\Sigma p_G = P$ in the sense of distributions, and similarly $\mathcal{T}_{\Sigma'} p_{G'} = P'$. The Lipschitz continuity condition on the coefficients is immediate, since the only non-constant coefficients are $c_\nu(\Sigma) = -\Sigma_{jk}/2$ for $\nu \in \mathbb{N}^d$ with $|\nu| = 2$, which depend linearly on Σ .

Verification of Assumption 2. Setting $\psi(t) := (1 + t/2)^{-1}$, so that $|\psi'(t)| = (1/2)|\psi(t)|^2 \leq (1/2)|\psi(t)|$, the mean value theorem implies that

$$\begin{aligned} |\Phi_{\Sigma'}(\xi) - \Phi_\Sigma(\xi)| &= |\psi'(\xi^\top \Sigma(\tau)\xi)| \cdot |\xi^\top (\Sigma' - \Sigma)\xi| \leq \frac{1}{2} \psi(\xi^\top \Sigma(\tau)\xi) \|\Sigma' - \Sigma\|_{\text{op}} \|\xi\|^2 \\ &\leq \frac{1}{2} \max\{|\Phi_\Sigma(\xi)|, |\Phi_{\Sigma'}(\xi)|\} \|\Sigma' - \Sigma\|_{\text{op}} \|\xi\|^2, \end{aligned}$$

for some $\tau \in (0, 1)$, where we set $\Sigma(\tau) := (1 - \tau)\Sigma' + \tau\Sigma$. Hence Assumption 2 holds with $p = 2$ and $\Xi(t) = t$.

Verification of Assumption 1: univariate case. Consider two univariate mixtures

$$p_G(x) = \int f(x | \theta, \sigma) P(d\theta) \quad \text{and} \quad p_{G'}(x) = \int f(x | \theta, \sigma') P'(d\theta),$$

where $f(\cdot | \theta, \sigma)$ is the univariate Laplace kernel with location θ and scale σ . The PDE inversion condition gives $\mathcal{T}_\sigma p_G = P$ and $\mathcal{T}_{\sigma'} p_{G'} = P'$, where $\mathcal{T}_\sigma := \text{Id} - \frac{\sigma}{2} \frac{d^2}{dx^2}$.

Let $h \in C_c^\infty(\mathbb{R})$ be a non-negative bump function supported on $(R+1, R+2)$ with $\|h\|_1 = 1$; the key property here is that $\text{supp}(h)$ is disjoint from $[-R, R] \supseteq \text{supp}(P) \cup \text{supp}(P')$. Therefore $\langle P, h \rangle = \langle P', h \rangle = 0$, and the PDE inversion gives

$$\int_{\mathbb{R}} p_G(x) (\mathcal{T}_\sigma h(x)) dx = \langle \mathcal{T}_\sigma p_G, h \rangle = \langle P, h \rangle = 0.$$

On the other hand, decomposing $\mathcal{T}_\sigma = \mathcal{T}_{\sigma'} + (\mathcal{T}_\sigma - \mathcal{T}_{\sigma'})$ and using $\langle p_{G'}, \mathcal{T}_{\sigma'} h \rangle = \langle P', h \rangle = 0$,

$$\int_{\mathbb{R}} p_{G'}(x) (\mathcal{T}_\sigma h(x)) dx = \langle p_{G'}, (\mathcal{T}_\sigma - \mathcal{T}_{\sigma'}) h \rangle = -\frac{1}{2}(\sigma - \sigma') \int_{\mathbb{R}} p_{G'}(x) h''(x) dx,$$

since $\mathcal{T}_\sigma - \mathcal{T}_{\sigma'} = -\frac{\sigma - \sigma'}{2} \frac{d^2}{dx^2}$. Subtracting and applying Hölder's inequality,

$$(47) \quad V \|\mathcal{T}_\sigma h\|_\infty \geq \frac{1}{2} |\sigma - \sigma'| \left| \int_{\mathbb{R}} h(x) p_{G'}''(x) dx \right|.$$

For $x \in (R+1, R+2)$ and $\theta' \in [-R, R]$, we have $x - \theta' \geq 1 > 0$, so the Laplace kernel simplifies to $f(x | \theta', \sigma') = (2\sigma')^{-1/2} \exp(-(x - \theta')/\sqrt{\sigma'/2})$, giving

$$f''(x | \theta', \sigma') = \frac{2}{\sigma'} f(x | \theta', \sigma').$$

Hence $p_{G'}''(x) = \frac{2}{\sigma'} p_{G'}(x)$ for $x \in (R+1, R+2)$, and since $\text{supp}(h) \subseteq (R+1, R+2)$,

$$\left| \int_{\mathbb{R}} h(x) p_{G'}''(x) dx \right| = \frac{2}{\sigma'} \int_{\mathbb{R}} h(x) p_{G'}(x) dx.$$

For $x \in [R + 1, R + 2]$ and $\theta' \in [-R, R]$, the Laplace kernel is bounded below by

$$f(x | \theta', \sigma') \geq \frac{1}{\sqrt{2\lambda_{\max}}} \exp\left(-\frac{2R+2}{\sqrt{\lambda_{\min}/2}}\right) =: c_{\min} > 0,$$

so $\int_{\mathbb{R}} h(x) p_{G'}(x) dx \geq c_{\min} \|h\|_1 = c_{\min}$. For the upper bound,

$$\|\mathcal{T}_{\sigma} h\|_{\infty} \leq \|h\|_{\infty} + \frac{\lambda_{\max}}{2} \|h''\|_{\infty} =: C_h < \infty.$$

Substituting into Equation (47),

$$V \geq \frac{c_{\min}}{C_h \lambda_{\max}} |\sigma - \sigma'|,$$

which confirms Assumption 1 in the univariate case with $\Psi(t) = t$.

Verification of Assumption 1: multivariate case. Let $u \in \mathbb{R}^d$ with $\|u\| = 1$ be the principal unit eigenvector of $\Sigma - \Sigma'$. The characteristic function of $u^{\top} X$ for $X \sim p_G$ at $t \in \mathbb{R}$ is

$$\mathbb{E}[e^{it u^{\top} X}] = \Phi_{p_G}(tu) = \Phi_P(tu) \Phi_{\Sigma}(tu) = \Phi_{P_u}(t) \cdot \frac{1}{1 + \frac{1}{2} t^2 u^{\top} \Sigma u},$$

where P_u is the pushforward of P under $\theta \mapsto u^{\top} \theta$. By uniqueness of characteristic functions, this identifies the projection $u^{\top} X$ as a univariate Laplace mixture with scale $\sigma := u^{\top} \Sigma u$ and location mixing measure P_u , and similarly for $u^{\top} X'$ with scale $\sigma' := u^{\top} \Sigma' u$. By Lemma 2.8,

$$\|p_G - p_{G'}\|_1 \geq \|p_{G,u} - p_{G',u}\|_1 \geq \frac{c_{\min}}{C_h \lambda_{\max}} |\sigma - \sigma'| = \frac{c_{\min}}{C_h \lambda_{\max}} \|\Sigma - \Sigma'\|_{\text{op}},$$

where the last equality uses $|\sigma - \sigma'| = |u^{\top} (\Sigma - \Sigma') u| = \|\Sigma - \Sigma'\|_{\text{op}}$ by the choice of u . This establishes Assumption 1 with $\Psi(t) = t$.

Conclusion. With the PDE inversion condition verified, Assumptions 1 and 2 established for $\Psi(t) = \Xi(t) = t$, $p = 2$, and ordinary smoothness of order $\beta = 2$ being obvious given the form of the Laplace characteristic function, an application of Theorem 3.7 completes the proof.

Proof of Proposition 5.1. The validity of both parts of the statement is a consequence of the inequalities established in our Theorems 4.1 and 4.2, applied in conjunction with Theorem 9.9 of [4] (specifically, their results for the super-smooth setting, which yield a $n^{-1/2}(\log n)^{(d+1)/2}$ Hellinger⁵ contraction rate). In particular, we rely on the fact if the base measure H is supported on $[-R, R]^d$, then the posterior is supported on mixing measures $G = P \times \delta_{\Sigma}$ with P having support included in $[-R, R]^d$ [7]. Moreover, assumptions (9.5), (9.6), and (9.7) adopted in Theorem 9.9 of [4] are trivially satisfied by our model due to the assumed compact support conditions. Furthermore, their assumption (9.8) is irrelevant in our context, as it is not required in their proof for the super-smooth case.

⁵The Hellinger distance h between densities f and f' is defined as $h^2(f, f') := \int (\sqrt{f} - \sqrt{f'})^2 dx$. Because $\|f - f'\|_1 \lesssim h(f, f')$, any Hellinger contraction rate is also an L^1 contraction rate. In particular, a general reverse inequality also exists, so the two metrics are topologically equivalent.

Proof of Proposition 5.2. The validity of the statement follows directly from Theorem 4.4 in [11], which under our assumptions establishes a $n^{-1/4}(\log n)^{5/4}$ Hellinger contraction rate, together with the inequality established in our Theorem 4.4. In particular, we remark that the lower-bound

$$\frac{dH}{dx}(\theta) \gtrsim e^{-b|\theta|^\delta} \quad \text{for some } c, \delta > 0,$$

as required in the statement of Theorem 4.4 of [11],⁶ is irrelevant in our case where P_0 is compactly supported, as it is only employed in the portion of the proof of the result dealing with the more general case where only a weaker exponential-power upper-bound on the tails of P_0 is assumed.

2. Auxiliary lemmas with proofs.

LEMMA 2.1. *Define the extended function $\bar{h}_\mathcal{X}^* : \mathbb{R}^d \rightarrow \mathbb{R}$ via the infimal convolution operator over the domain \mathcal{X} :*

$$\bar{h}_\mathcal{X}^*(x) := \inf_{y \in \mathcal{X}} \left\{ h_\mathcal{X}^*(y) + \|x - y\| \right\}, \quad x \in \mathbb{R}^d.$$

The following properties hold:

- (a) $\bar{h}_\mathcal{X}^*(x) = h_\mathcal{X}^*(x)$ for all $x \in \mathcal{X}$;
- (b) $\bar{h}_\mathcal{X}^*$ is a globally 1-Lipschitz continuous function on \mathbb{R}^d , meaning $\bar{h}_\mathcal{X}^* \in \text{Lip}_1(\mathbb{R}^d)$.

PROOF. Since $h_\mathcal{X}^* \in \text{Lip}_1(\mathcal{X})$ and $h_\mathcal{X}^*(0) = 0$, we have $|h_\mathcal{X}^*(y)| \leq \|y\| \leq R$ for all $y \in \mathcal{X}$, so $h_\mathcal{X}^*(y) + \|x - y\| \geq -R$ uniformly in y . Hence the infimum defining $\bar{h}_\mathcal{X}^*(x)$ is finite for every $x \in \mathbb{R}^d$.

Part (a). For any $x \in \mathcal{X}$, taking $y = x$ in the infimum gives $\bar{h}_\mathcal{X}^*(x) \leq h_\mathcal{X}^*(x)$. Conversely, the 1-Lipschitz property of $h_\mathcal{X}^*$ on \mathcal{X} gives $h_\mathcal{X}^*(x) - h_\mathcal{X}^*(y) \leq \|x - y\|$ for all $y \in \mathcal{X}$, so $h_\mathcal{X}^*(x) \leq h_\mathcal{X}^*(y) + \|x - y\|$ for all $y \in \mathcal{X}$, and taking the infimum over y yields $h_\mathcal{X}^*(x) \leq \bar{h}_\mathcal{X}^*(x)$. Combining both inequalities, $\bar{h}_\mathcal{X}^*(x) = h_\mathcal{X}^*(x)$ for all $x \in \mathcal{X}$.

Part (b). For any $u, v \in \mathbb{R}^d$ and any $y \in \mathcal{X}$, the triangle inequality gives $\|u - y\| \leq \|u - v\| + \|v - y\|$. Adding $h_\mathcal{X}^*(y)$ to both sides and taking the infimum over $y \in \mathcal{X}$,

$$\bar{h}_\mathcal{X}^*(u) \leq \|u - v\| + \bar{h}_\mathcal{X}^*(v).$$

Swapping the roles of u and v and combining yields $|\bar{h}_\mathcal{X}^*(u) - \bar{h}_\mathcal{X}^*(v)| \leq \|u - v\|$ for all $u, v \in \mathbb{R}^d$, so $\bar{h}_\mathcal{X}^* \in \text{Lip}_1(\mathbb{R}^d)$. \square

LEMMA 2.2. *Define the isotropic bump function $\varrho : \mathbb{R}^d \rightarrow \mathbb{R}$ as*

$$\varrho(x) := \begin{cases} \exp\left(-\frac{1}{1-\|x\|^2}\right) & \text{if } \|x\| < 1 \\ 0 & \text{if } \|x\| \geq 1. \end{cases}$$

Then define the normalized mollifier $\psi(x) = \frac{1}{C_\varrho} \varrho(x)$, where $C_\varrho := \int_{\mathbb{R}^d} \varrho(x) dx \in (0, \infty)$, and its $\epsilon = \frac{R}{2}$ -scaled mollifier as

$$\psi_\epsilon(x) = \frac{1}{\epsilon^d} \psi\left(\frac{x}{\epsilon}\right).$$

⁶The bound itself is stated in Equation (4.12) of that same paper.

Then define the cutoff function $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$ as

$$\eta(x) := (\mathbf{1}_A * \psi_\epsilon)(x) = \int_{\mathbb{R}^d} \mathbf{1}_A(x-y)\psi_\epsilon(y) dy,$$

where $A = \bar{B}(0, R + \epsilon) = \bar{B}(0, \frac{3R}{2})$. Then the following properties hold:

- (a) $\eta \in C_c^\infty(\mathbb{R}^d)$;
- (b) $\eta(x) = 1$ for all $x \in \bar{B}(0, R)$, $\eta(x) = 0$ for all $x \notin B(0, 2R)$, and $0 \leq \eta(x) \leq 1$ for all $x \in \mathbb{R}^d$;
- (c) $M_\eta := \|\nabla\eta\|_\infty < \infty$.

PROOF. The function $g : \mathbb{R} \rightarrow \mathbb{R}$ defined by $g(t) = e^{-1/t}$ for $t > 0$ and $g(t) = 0$ for $t \leq 0$ belongs to $C^\infty(\mathbb{R})$ by Lemma 2.20 in [6]. Since $\varrho(x) = g(1 - \|x\|^2)$ is a composition of g with a smooth polynomial, $\varrho \in C^\infty(\mathbb{R}^d)$. It is strictly positive on $B(0, 1)$ and vanishes on $\|x\| \geq 1$, so $\text{supp}(\varrho) = \bar{B}(0, 1)$ is compact and $\varrho \in C_c^\infty(\mathbb{R}^d)$. Hence $C_\varrho \in (0, \infty)$ and ψ is well-defined with $\int_{\mathbb{R}^d} \psi(x) dx = 1$. The scaled mollifier satisfies $\psi_\epsilon \in C_c^\infty(\mathbb{R}^d)$, $\text{supp}(\psi_\epsilon) = \bar{B}(0, \epsilon)$, and $\int_{\mathbb{R}^d} \psi_\epsilon(x) dx = 1$ by the substitution $y = x/\epsilon$.

Part (a). For any multi-index ν , $\partial^\nu \eta(x) = \int_{\mathbb{R}^d} \mathbf{1}_A(y) (\partial^\nu \psi_\epsilon)(x-y) dy$, which is continuous since $\partial^\nu \psi_\epsilon$ is smooth, bounded, and compactly supported. Hence $\eta \in C^\infty(\mathbb{R}^d)$. The support satisfies

$$\text{supp}(\eta) \subseteq \text{supp}(\mathbf{1}_A) + \text{supp}(\psi_\epsilon) = \bar{B}(0, \frac{3R}{2}) + \bar{B}(0, \frac{R}{2}) = \bar{B}(0, 2R),$$

so $\eta \in C_c^\infty(\mathbb{R}^d)$.

Part (b). Since $\mathbf{1}_A \geq 0$ and $\psi_\epsilon \geq 0$, we have $\eta(x) \geq 0$. Since $\mathbf{1}_A \leq 1$, we have $\eta(x) \leq \int_{\mathbb{R}^d} \psi_\epsilon(y) dy = 1$.

For $x \in \bar{B}(0, R)$ and any $y \in \text{supp}(\psi_\epsilon)$, the triangle inequality gives $\|x-y\| \leq \|x\| + \|y\| \leq R + R/2 = 3R/2$, so $x-y \in A$ and $\mathbf{1}_A(x-y) = 1$. Hence $\eta(x) = \int_{\mathbb{R}^d} \psi_\epsilon(y) dy = 1$.

For $x \notin B(0, 2R)$ and any $y \in \text{supp}(\psi_\epsilon)$, $\|x-y\| \geq \|x\| - \|y\| \geq 2R - R/2 = 3R/2$, so $x-y \notin A$ and $\mathbf{1}_A(x-y) = 0$. Hence $\eta(x) = 0$. By continuity, $\eta(x) = 0$ on $\|x\| = 2R$ as well.

Part (c). Since $\eta(x) = 1$ for $\|x\| < R$ and $\eta(x) = 0$ for $\|x\| > 2R$, we have $\nabla\eta(x) = 0$ outside the compact annulus $\bar{B}(0, 2R) \setminus B(0, R)$. As $\eta \in C^\infty(\mathbb{R}^d)$, the function $x \mapsto \|\nabla\eta(x)\|$ is continuous and attains its maximum on this compact set, giving $M_\eta := \|\nabla\eta\|_\infty = \sup_{x \in \mathbb{R}^d} \|\nabla\eta(x)\| < \infty$. \square

LEMMA 2.3. For $\tilde{h}_\chi^*(x) := \bar{h}_\chi^*(x)\eta(x)$, the following properties hold:

- (a) $\tilde{h}_\chi^* \in L_1(\mathbb{R}^d)$;
- (b) \tilde{h}_χ^* is globally $(1 + 2RM_\eta)$ -Lipschitz continuous, namely

$$\forall x, y \in \mathbb{R}^d, \quad |\tilde{h}_\chi^*(x) - \tilde{h}_\chi^*(y)| \leq (1 + 2RM_\eta)\|x - y\|;$$

- (c) $\|\mathcal{F}[\tilde{h}_\chi^*]\|_\infty \leq \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1}$.

PROOF. Part (a). Since $\text{supp}(\tilde{h}_\chi^*) \subseteq \text{supp}(\eta) \subseteq \bar{B}(0, 2R)$, and \bar{h}_χ^* is 1-Lipschitz with $\bar{h}_\chi^*(0) = 0$, we have $|\tilde{h}_\chi^*(x)| \leq |\bar{h}_\chi^*(x)| \cdot \eta(x) \leq \|x\| \leq 2R$ for all $x \in \mathbb{R}^d$. Therefore,

$$\int_{\mathbb{R}^d} |\tilde{h}_\chi^*(x)| dx \leq \int_{\bar{B}(0, 2R)} 2R dx = \frac{\pi^{d/2}}{\Gamma(d/2+1)}(2R)^{d+1} < \infty,$$

so $\tilde{h}_{\mathcal{X}}^* \in L_1(\mathbb{R}^d)$.

Part (b). Since $\bar{h}_{\mathcal{X}}^* \in \text{Lip}_1(\mathbb{R}^d)$, Rademacher's theorem gives $\|\nabla \bar{h}_{\mathcal{X}}^*(x)\| \leq 1$ for almost every $x \in \mathbb{R}^d$. Since $\eta \in C_c^\infty(\mathbb{R}^d)$, the product rule gives $\nabla \tilde{h}_{\mathcal{X}}^*(x) = \eta(x)\nabla \bar{h}_{\mathcal{X}}^*(x) + \bar{h}_{\mathcal{X}}^*(x)\nabla \eta(x)$ almost everywhere. By the triangle inequality,

$$\|\nabla \tilde{h}_{\mathcal{X}}^*(x)\| \leq |\eta(x)|\|\nabla \bar{h}_{\mathcal{X}}^*(x)\| + |\bar{h}_{\mathcal{X}}^*(x)|\|\nabla \eta(x)\| \leq 1 + 2RM_\eta,$$

where we used $\eta(x) \leq 1$, $\|\nabla \bar{h}_{\mathcal{X}}^*(x)\| \leq 1$, $|\bar{h}_{\mathcal{X}}^*(x)| \leq \|x\| \leq 2R$ on $\text{supp}(\nabla \eta) \subseteq \bar{B}(0, 2R)$, and $\|\nabla \eta\|_\infty = M_\eta$. Taking the essential supremum, $\tilde{h}_{\mathcal{X}}^*$ is globally $(1 + 2RM_\eta)$ -Lipschitz.

Part (c). By the triangle inequality and part (a),

$$|\mathcal{F}[\tilde{h}_{\mathcal{X}}^*](\xi)| \leq \int_{\mathbb{R}^d} |\tilde{h}_{\mathcal{X}}^*(x)| dx \leq \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} (2R)^{d+1}.$$

□

LEMMA 2.4. *Let $M_K := \int_{\mathbb{R}^d} \|x\| K(x) dx$. Then the following properties hold:*

- (a) $\mathcal{F}[K](0) = 1$ and $\text{supp}(\mathcal{F}[K]) \subseteq \bar{B}(0, 1)$;
- (b) $M_K \leq 3\sqrt{d}(1 + 256d^2)/(8\pi)$.

PROOF. *Part (a).* Since $\int_{\mathbb{R}^d} K(x) dx = 1$, we have $\mathcal{F}[K](0) = 1$. For the support bound, the one-dimensional Fourier transform satisfies $\mathcal{F}[\text{sinc}](\xi) = \pi \mathbf{1}_{[-1,1]}(\xi)$, so

$$\mathcal{F}[\text{sinc}^4](\xi) = \frac{1}{(2\pi)^3} (\mathcal{F}[\text{sinc}] * \mathcal{F}[\text{sinc}] * \mathcal{F}[\text{sinc}] * \mathcal{F}[\text{sinc}])(\xi),$$

and $\text{supp}(\mathcal{F}[\text{sinc}^4]) \subseteq [-4, 4]$ by the Minkowski sum property of convolution supports. Setting $k(x) := \frac{3}{8\pi\sqrt{d}} \text{sinc}^4\left(\frac{x}{4\sqrt{d}}\right)$, we have $\mathcal{F}[k](\xi) = 4\sqrt{d} \mathcal{F}[\text{sinc}^4](4\sqrt{d}\xi)$, giving $\text{supp}(\mathcal{F}[k]) \subseteq [-1/\sqrt{d}, 1/\sqrt{d}]$. Since $K(x) = \prod_{i=1}^d k(x_i)$, the Fourier transform factorizes as $\mathcal{F}[K](\xi) = \prod_{i=1}^d \mathcal{F}[k](\xi_i)$, so

$$\text{supp}(\mathcal{F}[K]) \subseteq \prod_{i=1}^d \left[-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right].$$

For any ξ in this box, $\|\xi\| \leq \sqrt{d} \cdot \frac{1}{\sqrt{d}} = 1$, hence $\text{supp}(\mathcal{F}[K]) \subseteq \bar{B}(0, 1)$.

Part (b). Using $\|x\| \leq \sum_{j=1}^d |x_j|$ and the product structure of K ,

$$M_K = \int_{\mathbb{R}^d} \|x\| K(x) dx \leq \sum_{i=1}^d \frac{3}{8\pi\sqrt{d}} \int_{-\infty}^{\infty} |x_i| \text{sinc}^4\left(\frac{x_i}{4\sqrt{d}}\right) dx_i.$$

Splitting each integral at $|x_i| = 1$ and using $|\text{sinc}(x_i/4\sqrt{d})| \leq 1$ on $[0, 1]$ and $|\sin(x_i/4\sqrt{d})| \leq 1$ on $[1, \infty)$,

$$\int_{-\infty}^{\infty} |x_i| \text{sinc}^4\left(\frac{x_i}{4\sqrt{d}}\right) dx_i \leq 2 \int_0^1 x_i dx_i + 2(4\sqrt{d})^4 \int_1^{\infty} x_i^{-3} dx_i = 1 + 256d^2.$$

Therefore $M_K \leq \frac{3\sqrt{d}(1+256d^2)}{8\pi} < \infty$. □

LEMMA 2.5. In \mathbb{R}^d , the following identities and inequalities hold:

- (a) $\int_{\bar{B}(0,R)} (1 + \|\xi\|^\beta) d\xi \leq \frac{2(2\pi)^{d/2}}{d\Gamma(d/2)} \max\{R^d, R^{d+\beta}\};$
 (b) $\int_{\bar{B}(0,R)} \|\xi\|^p d\xi = \frac{(2\pi)^{d/2}}{(d+p)\Gamma(d/2)} R^{d+p}.$

PROOF. Moving to polar coordinates gives

$$(48) \quad \int_{\bar{B}(0,R)} (a + b\|\xi\|^\beta) d\xi = \frac{(2\pi)^{d/2}}{\Gamma(d/2)} \int_0^R r^{d-1} (a + br^\beta) dr = \frac{(2\pi)^{d/2}}{\Gamma(d/2)} \left(\frac{aR^d}{d} + \frac{bR^{d+\beta}}{d+\beta} \right).$$

Setting $a = b = 1$ yields part (a):

$$\int_{\bar{B}(0,R)} (1 + \|\xi\|^\beta) d\xi = \frac{(2\pi)^{d/2}}{\Gamma(d/2)} \left(\frac{R^d}{d} + \frac{R^{d+\beta}}{d+\beta} \right) \leq \frac{2(2\pi)^{d/2}}{d\Gamma(d/2)} \max\{R^d, R^{d+\beta}\}.$$

Part (b) follows from Equation (48) by setting $a = 0$, $b = 1$, and $\beta = p$. \square

LEMMA 2.6. Let $X, Y \sim f(\cdot | 0, 1)$ be independent standard Gaussian random variables. Then $\mathbb{E}[(X + iY)^n] = 0$ for all $n \geq 1$.

PROOF. From the moment generating function and characteristic function of the standard Gaussian,

$$\mathbb{E}[e^{t(X+iY)}] = \mathbb{E}[e^{tX}] \mathbb{E}[e^{itY}] = e^{t^2/2} \cdot e^{-t^2/2} = 1, \quad t \in \mathbb{R}.$$

Expanding the left-hand side as a power series and matching coefficients of $t^n/n!$ for each $n \geq 1$ yields the result. \square

LEMMA 2.7. For $\sigma, R, L > 0$ and k a positive integer, the following bounds hold:

- (a) $\int_{x \geq R+L} e^{-(x-R)^2/2\sigma^2} dx \leq (\pi/2)^{1/2} \sigma e^{-L^2/2\sigma^2};$
 (b) $\int_{x \geq R+L} x^{2k} e^{-(x-R)^2/2\sigma^2} dx \leq 3^{2k-1} e^{-L^2/2\sigma^2} (\pi/2)^{1/2} \sigma \cdot (R^{2k} + L^{2k} + \sigma^{2k}(2k-1)!).$

PROOF. For part (a), observe that for $x \geq R+L$ we have $(x-R)^2 = ((x-R-L)+L)^2 \geq (x-R-L)^2 + L^2$, so

$$\int_{x \geq R+L} e^{-(x-R)^2/2\sigma^2} dx \leq e^{-L^2/2\sigma^2} \int_0^\infty e^{-t^2/2\sigma^2} dt = (\pi/2)^{1/2} \sigma e^{-L^2/2\sigma^2}.$$

For part (b), use $x^{2k} \leq 3^{2k-1}((x-R-L)^{2k} + R^{2k} + L^{2k})$ for $x \geq R+L$. Setting $t = x - R - L$ and applying the same exponential bound as in part (a),

$$\begin{aligned} \int_{x \geq R+L} x^{2k} e^{-(x-R)^2/2\sigma^2} dx &\leq 3^{2k-1} e^{-L^2/2\sigma^2} \int_0^\infty (t^{2k} + R^{2k} + L^{2k}) e^{-t^2/2\sigma^2} dt \\ &= 3^{2k-1} e^{-L^2/2\sigma^2} (\pi/2)^{1/2} \sigma \cdot (R^{2k} + L^{2k} + \sigma^{2k}(2k-1)!), \end{aligned}$$

where the last equality uses $\int_0^\infty t^{2k} e^{-t^2/2\sigma^2} dt = (\pi/2)^{1/2} \sigma^{2k+1} (2k-1)!!$. \square

LEMMA 2.8. Let X and Y be continuous random vectors in \mathbb{R}^d with densities p_X and p_Y , and let $u \in \mathbb{R}^d$ with $\|u\| = 1$. Denoting by $X_u := u^\top X$ and $Y_u := u^\top Y$ the projections onto u ,

$$\|p_X - p_Y\|_1 \geq \|p_{X_u} - p_{Y_u}\|_1.$$

PROOF. For any Borel set $A \subseteq \mathbb{R}$, the preimage $A^u := \{x \in \mathbb{R}^d : u^\top x \in A\}$ is a Borel set in \mathbb{R}^d . Therefore,

$$\begin{aligned} \frac{1}{2} \|p_{X_u} - p_{Y_u}\|_1 &= \sup_{A \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(X_u \in A) - \mathbb{P}(Y_u \in A)| = \sup_{A \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(X \in A^u) - \mathbb{P}(Y \in A^u)| \\ &\leq \sup_{B \in \mathcal{B}(\mathbb{R}^d)} |\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| = \frac{1}{2} \|p_X - p_Y\|_1, \end{aligned}$$

which completes the proof. \square

LEMMA 2.9. *For any $p \geq 0$ and $\varepsilon > 0$, there exists a constant $M_{p,\varepsilon} > 0$, depending only on p and ε , such that $u^p \leq M_{p,\varepsilon} e^{\varepsilon u}$ for all $u > 0$.*

PROOF. Let $k := \lceil p \rceil$. From the Taylor expansion of $e^{\varepsilon u}$, we have

$$e^{\varepsilon u} \geq 1 + \frac{\varepsilon^k u^k}{k!} \geq \min\left(1, \frac{\varepsilon^k}{k!}\right) \max\{1, u^k\} \geq \min\left(1, \frac{\varepsilon^k}{k!}\right) u^p.$$

Dividing both sides by $\min\left(1, \frac{\varepsilon^k}{k!}\right)$ yields

$$u^p \leq \max\left(1, \frac{k!}{\varepsilon^k}\right) e^{\varepsilon u}.$$

The result follows by setting $M_{p,\varepsilon} := \max\left(1, \frac{k!}{\varepsilon^k}\right)$. \square

LEMMA 2.10. *Let μ and ν be two probability measures on \mathbb{R}^d with finite first moments, i.e.,*

$$\int_{\mathbb{R}^d} \|x\| \mu(dx) < \infty, \quad \int_{\mathbb{R}^d} \|x\| \nu(dx) < \infty.$$

Then there exists a function $h^ \in \text{Lip}_1(\mathbb{R}^d)$ with $h^*(0) = 0$ that attains the Kantorovich-Rubinstein dual problem:*

$$\sup_{h \in \text{Lip}_1(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} h(x) \mu(dx) - \int_{\mathbb{R}^d} h(x) \nu(dx) \right) = \int_{\mathbb{R}^d} h^*(x) \mu(dx) - \int_{\mathbb{R}^d} h^*(x) \nu(dx).$$

PROOF. Let $(h_n)_{n \in \mathbb{N}}$ be a maximizing sequence for the Kantorovich-Rubinstein dual problem. Since adding a constant to any h_n does not affect $\int h_n d\mu - \int h_n d\nu$ (as μ and ν are both probability measures), we may assume without loss of generality that $h_n(0) = 0$ for all n . Consequently, $|h_n(x)| = |h_n(x) - h_n(0)| \leq \|x\|$ for all $x \in \mathbb{R}^d$, so the family $(h_n)_{n \in \mathbb{N}}$ is uniformly bounded and equicontinuous on every compact subset of \mathbb{R}^d .

We now apply a diagonal argument with the Arzelà-Ascoli theorem. Applying Arzelà-Ascoli on $\bar{B}(0, 1)$, we extract a subsequence $(h_n^1)_{n \in \mathbb{N}} \subset (h_n)_{n \in \mathbb{N}}$ converging uniformly on $\bar{B}(0, 1)$. Iteratively, for each $k \geq 1$, by Arzelà-Ascoli on $\bar{B}(0, k+1)$, we extract a subsequence $(h_n^{k+1})_{n \in \mathbb{N}} \subset (h_n^k)_{n \in \mathbb{N}}$ converging uniformly on $\bar{B}(0, k+1)$. The diagonal sequence $(h_n^n)_{n \in \mathbb{N}}$ then converges uniformly on every compact subset of \mathbb{R}^d to some pointwise limit h^* . Since uniform-on-compacts limits of 1-Lipschitz functions are 1-Lipschitz, $h^* \in \text{Lip}_1(\mathbb{R}^d)$ and clearly $h^*(0) = 0$.

It remains to show that h^* attains the supremum. We have $h_n^n(x) \rightarrow h^*(x)$ pointwise, and the bound $|h_n^n(x)| \leq \|x\|$ holds uniformly in n . Since

$$\int_{\mathbb{R}^d} \|x\| \mu(dx) + \int_{\mathbb{R}^d} \|x\| \nu(dx) < \infty,$$

the Dominated Convergence Theorem (applied separately to μ and ν) yields

$$\int_{\mathbb{R}^d} h_n^n(x) \mu(dx) - \int_{\mathbb{R}^d} h_n^n(x) \nu(dx) \longrightarrow \int_{\mathbb{R}^d} h^*(x) \mu(dx) - \int_{\mathbb{R}^d} h^*(x) \nu(dx).$$

Since $(h_n)_{n \in \mathbb{N}}$ is a maximizing sequence, the left-hand side converges to $W_1(\mu, \nu)$, so h^* attains the supremum. This completes the proof. \square

LEMMA 2.11. *Let $x, y, M > 0$ satisfy $x \leq My \log y$ and $x \geq e^{1/M}$. Then $y \geq x/(M \log x)$.*

PROOF. Define $f(t) := t \log t$ for $t > 0$. Since $f'(t) = \log t + 1$, the function f is decreasing on $(0, e^{-1})$ and increasing on (e^{-1}, ∞) , with global minimum $f(e^{-1}) = -e^{-1}$. From $x > 0$ and $x \leq My \log y$, we have $y \log y > 0$, hence $y > 1$.

The hypothesis $x \geq e^{1/M}$ gives $\log x \geq 1/M$, hence $M \log x \geq 1$, which implies $x/(M \log x) \leq x$. Therefore

$$f\left(\frac{x}{M \log x}\right) = \frac{x}{M \log x} \log\left(\frac{x}{M \log x}\right) \leq \frac{x}{M \log x} \log x = \frac{x}{M}.$$

Combined with $x/M \leq y \log y = f(y)$, this gives $f(x/(M \log x)) \leq f(y)$.

We conclude with a brief case analysis. If $x/(M \log x) \leq 1$, then since $y > 1$, the conclusion $y \geq x/(M \log x)$ holds trivially. Otherwise, $x/(M \log x) > 1$, and both $x/(M \log x)$ and y lie in $(1, \infty) \subset (e^{-1}, \infty)$, where f is strictly increasing; hence $f(x/(M \log x)) \leq f(y)$ implies $x/(M \log x) \leq y$. \square

REFERENCES

- [1] BUTZER, P. L. and NESSEL, R. J. (1971). *Fourier analysis and approximation, Vol. 1* 7. Birkhäuser Basel.
- [2] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive approximation* **303**. Springer Science & Business Media.
- [3] FOLLAND, G. B. (1999). *Real analysis: modern techniques and their applications*. John Wiley & Sons.
- [4] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of nonparametric Bayesian inference* **44**. Cambridge University Press.
- [5] HEINONEN, J. (2001). *Lectures on analysis on metric spaces*. Springer Science & Business Media.
- [6] LEE, J. M. (2013). *Introduction to Smooth Manifolds*, 2 ed. *Graduate Texts in Mathematics* **218**. Springer, New York.
- [7] MAJUMDAR, S. (1992). On topological support of Dirichlet prior. *Statistics & Probability Letters* **15** 381–384.
- [8] MCSHANE, E. J. (1934). Extension of range of functions. *Bulletin of the American Mathematical Society* **40** 837–842.
- [9] RUDIN, W. (1991). *Functional Analysis*, Second ed. McGraw-Hill, New York.
- [10] SCHWARTZ, L. (1966). *Théorie des distributions*. Hermann, Paris.
- [11] SCRICCILOLO, C. (2011). Posterior rates of convergence for Dirichlet mixtures of exponential power densities. *Electronic Journal of Statistics* **5** 270–308.
- [12] VILLANI, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society.
- [13] WHITNEY, H. (1934). Analytic Extensions of Differentiable Functions Defined in Closed Sets. *Transactions of the American Mathematical Society* **36** 63.