

Partial Differential Equation Barriers to Identifiability in Infinite Mixture Models

Dung Le Nicola Barileto Alessandro Rinaldo Nhat Ho

The University of Texas at Austin

August 8, 2026

Abstract

We study identifiability of mixing measures in infinite mixture models. We show that, in many common cases, lack of identifiability can be characterized in terms of certain differential structures of the kernel family with respect to its parameters. In our main result we prove that, when the kernel is annihilated by a non-trivial differential or difference-differential operator over the parameter space, there exist infinitely many distinct mixing measures yielding the same mixture density. We give verifiable conditions for such operators to exist, covering many common cases, including the location-scale Gaussian, location-scale Student- t , gamma, beta, Dirichlet, negative binomial and non-central chi-squared families. Furthermore, our conditions apply to any exponential family whose parameter dimension exceeds the dimension of its sufficient statistic and, more generally, to kernels with polynomially-growing score functions. We complement our results with a minimax lower bound on the estimation error for the mixing measure in the Wasserstein distance under non-identifiability. On the flip side, we describe three classes of kernels for which identifiability is preserved and nonparametric statistical inference remains possible.

1 Introduction

Mixtures are among the most widely used models in statistics, offering a principled way to describe populations composed of unobserved subgroups and to build flexible distributions from simple parametric components (McLachlan and Peel, 2000; Fraley and Raftery, 2002). In their basic form, data $X_1, \dots, X_n$ taking values in a sample space $\mathcal{X} \subseteq \mathbb{R}^{d_1}$ are assumed to be independent draws from a density

$$p_{G_*}(x) = \int_{\Theta} f(x \mid \theta) G_*(d\theta), \quad x \in \mathcal{X}, \quad (1)$$

where $\{f(\cdot \mid \theta) : \theta \in \Theta\}$ is a family of probability density kernels (with respect to some common dominating measure on $\mathcal{X}$, say λ) indexed by the parameter space $\Theta \subseteq \mathbb{R}^{d_2}$, and the *mixing measure* G_* is a probability measure on Θ that weights the parameter values entering the mixture. Equivalently, the generative model implied by (1) admits a hierarchical representation: latent parameters $\theta_1, \dots, \theta_n$ are drawn independently from G_* , and each X_i is then drawn independently from $f(\cdot \mid \theta_i)$. For the purposes of statistical inference, two related objects are of interest: the density p_{G_*} and the mixing measure G_* . The latter, whose estimation motivates the present work, encodes the latent structure underlying tasks such as model-based clustering and the recovery of subgroup parameters.

In nonparametric settings, the unknown G_* is left unrestricted (or nearly so) and is modelled through mixing measures G with infinite support, which we call *infinite mixing measures*, giving rise to *infinite mixture models*. This flexible formulation is standard in Bayesian nonparametrics, through

priors on G such as the Dirichlet process (Ferguson, 1973, 1983; Lo, 1984) that model G_* as discrete and induce a latent clustering of the data according to the mixture component imputed to have generated each observation. While G_* governs this rich latent structure, the flexibility of the scheme comes at the cost that G_* is considerably harder to estimate than p_{G_*} . This is because, as the hierarchical form makes explicit, the data are directly informative about p_{G_*} , whereas information about G_* is only indirectly available through the additional noise layer contributed by the kernel $f(\cdot | \theta)$. This effectively makes p_{G_*} a noisy version of G_* , so that the statistical recovery of G_* may be framed as the problem of denoising p_{G_*} , or an estimate thereof, to extract the latent signal G_* . In light of this interpretation, the first fundamental question to address, for the task at hand to be at least well-posed, is about *identifiability*: even granting perfect estimation of the data-generating density—that is, $p_{\hat{G}} = p_{G_*}$ for some estimator $\hat{G}$ of G_* —does recovery of the mixing measure follow—that is, $\hat{G} = G_*$?

1.1 Contributions of the article

This work focuses precisely on the above-mentioned question of identifiability for infinite mixture models. To begin our discussion, write $\mathcal{A}$ for the linear operator $G \mapsto \int_{\Theta} f(\cdot | \theta) G(d\theta)$ that maps mixing measures to mixture densities. Using this formalism, the mixing measure is identifiable based on the mixture density whenever $\mathcal{A}$ is injective. In light of this, the key questions we address are: for which kernels f does injectivity of $\mathcal{A}$ hold? And when this property does not hold, what are the structural mechanisms responsible for the failure of identifiability?

Our starting point is the observation that many standard kernels belong to the null space of, or equivalently are *annihilated* by, a differential operator acting on the parameter θ . To build intuition, fix a probability measure G_* admitting a density g_* that is bounded away from zero on an open ball $B \subseteq \Theta^\circ$, where Θ° denotes the interior of Θ , and suppose that there exists a non-trivial linear differential operator $\mathcal{L}_\theta$ such that

$$\mathcal{L}_\theta f(x | \theta) = 0 \quad \text{for } \lambda\text{-almost every } x \in \mathcal{X} \text{ and every } \theta \in \Theta^\circ. \quad (2)$$

We refer to (2) as a *parameter partial differential equation*, or *parameter PDE*. For ease of exposition, we postpone the presentation of more general differential structures and of concrete examples. For the moment, let $\mathcal{L}_\theta^*$ denote the formal adjoint of $\mathcal{L}_\theta$ (Folland, 1999), the latter also being a differential operator that can be applied to any infinitely differentiable and compactly supported function $u : \mathbb{R}^{d_2} \rightarrow \mathbb{R}$. Hence, letting $g := \mathcal{L}_\theta^* u$ for an appropriately chosen u , the properties of adjoint operators and the parameter PDE lead to

$$\int_{\Theta} f(x | \theta) g(\theta) d\theta = \int_{\Theta} u(\theta) \mathcal{L}_\theta f(x | \theta) d\theta = 0 \quad \text{for } \lambda\text{-almost every } x \in \mathcal{X}. \quad (3)$$

Integrating (3) over x and using Fubini's theorem also gives $\int_{\Theta} g(\theta) d\theta = 0$. Therefore, defining $G(d\theta) := g(\theta) d\theta$ yields a signed measure such that $\mathcal{A}G = 0$. This, together with other properties of u , implies that the perturbed probability measure $G_* + \varepsilon G$, for $\varepsilon > 0$ small enough, is different from G_* while giving rise to the same mixture density: $p_{G_*} = p_{G_* + \varepsilon G}$ λ -almost everywhere.

This simple line of reasoning sheds new light on a direct connection between the presence of parameter PDEs and the non-identifiability of mixing measures in infinite mixture models. Importantly, the existence of parameter PDEs as in (2) is not a mere mathematical artifact, but will be shown to rather be a pervasive feature of commonly used kernel families. While a comprehensive

set of such families will be discussed later, it is useful to exemplify the discussion by means of a concrete and popular case, namely, *location-scale Gaussian mixtures*. With $d_1 = 1$ and $d_2 = 2$, let $\mathcal{N}(x \mid \mu, \nu) := (2\pi\nu)^{-1/2} \exp\{-(x - \mu)^2/(2\nu)\}$ denote the univariate Gaussian density with mean μ and variance ν , with kernel $f(x \mid \mu, \nu) = \mathcal{N}(x \mid \mu, \nu)$ indexed by $\theta := (\mu, \nu) \in \Theta := [-a, a] \times [\nu_{\min}, \nu_{\max}] \subset \mathbb{R}^2$, for $a > 0$ and $0 < \nu_{\min} < \nu_{\max} < \infty$. Denoting by ∂_y the partial derivative operator with respect to the scalar argument y , the Gaussian kernel famously solves the heat equation

$$\partial_\nu f(x \mid \mu, \nu) = \frac{1}{2} \partial_\mu^2 f(x \mid \mu, \nu) \quad \text{for every } x \in \mathbb{R} \text{ and } (\mu, \nu) \in \Theta^\circ, \quad (4)$$

which constitutes a parameter PDE in the sense of (2), with $\mathcal{L}_\theta = \partial_\nu - \frac{1}{2} \partial_\mu^2$. Hence, fixing an appropriate infinitely differentiable $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ and recognizing that $\mathcal{L}_\theta^* u = -\partial_\nu u(\theta, \nu) - \frac{1}{2} \partial_\theta^2 u(\theta, \nu)$, the same line of reasoning as in the previous paragraph implies that infinite location-scale Gaussian mixtures suffers from non-identifiability of the mixing measure.

Beyond these heuristic preliminaries, the contributions of the article can be summarized as follows. First, we formally show that the arguments outlined above hold in very general settings. In particular, Theorem 1 states that a kernel annihilated by a non-trivial differential operator generates a non-identifiable mixture model whenever the mixing measure has a density that is locally bounded away from zero. Theorem 2 extends this conclusion to difference-differential operators, in which the kernel is evaluated at shifted parameter values. Second, we give verifiable conditions under which such operators exist. Proposition 1 shows that any exponential family whose parameter dimension exceeds the dimension of its sufficient statistic admits a parameter PDE, while Proposition 2 generalizes this results also to kernels whose score is polynomial in a finite collection of statistics. Together with the examples of Section 4.3, these results identify common barriers to identifiability, which fails across the location-scale Gaussian, location-scale Student- t , gamma, beta, Dirichlet, negative binomial and non-central chi-squared families, and they indicate that over-parameterisation, in the sense $d_2 > q$ with q denoting the dimension of the sufficient statistic, is a systematic mechanism behind such barriers. As we document in Proposition 3, these instances of non-identifiability have crucial implications for statistical estimation of the mixing measure, which becomes impossible without further constraints on the model or G_* .

Third, we describe three classes of kernels for which no such barrier exists and identifiability is ensured. Proposition 4 introduces generalized translation families, for which identifiability reduces to the non-vanishing of the Fourier transform of a base kernel. Proposition 5 treats exponential families in which the sufficient statistic has dimension at least that of the parameter and its range is rich enough to determine measures on the natural parameter space, a condition of Müntz–Szász type. Proposition 6 considers a third class of kernels that both preserves identifiability and verifies the absence of any parameter PDE condition.

The remainder of the article is organised as follows. Section 2 fixes notation and preliminary concepts. Section 3 reviews the existing literature on identifiability and estimation in mixture models. Section 4 develops the theory of PDE barriers to identifiability, presents concrete examples, and discusses the consequences of non-identifiability on statistical estimation of the mixing measure. Section 5 treats kernel families for which identifiability is preserved. Section 6 concludes. The proofs of the results presented in the article are collected in the Supplementary Material.

2 Notation and preliminary concepts

We assume throughout that the sample space $\mathcal{X}$ is a Borel-measurable subset of $\mathbb{R}^{d_1}$ over which a σ -finite Borel measure λ is well defined; this is understood to be the Lebesgue measure for continuous kernels and the counting measure for discrete ones. Analogously, the parameter space $\Theta \subset \mathbb{R}^{d_2}$ is assumed to be regular compact¹ with a non-empty interior Θ° , and is equipped with the restriction of the Lebesgue measure in $\mathbb{R}^{d_2}$. For $k \in \mathbb{N}$, k -fold product spaces are denoted as $\mathcal{X}^k$ and Θ^k , and are endowed (when meaningful) with the product topology and the corresponding Borel σ -algebra. Similarly, P^k , where P is a probability measure on some measurable space such as $\mathcal{X}$ or Θ , denotes the k -fold product measure on the corresponding product space. A kernel $f : \mathcal{X} \times \Theta \rightarrow [0, \infty)$ is taken to be a jointly measurable function satisfying $\int_{\mathcal{X}} f(x | \theta) \lambda(dx) = 1$ for every $\theta \in \Theta$, so that $x \mapsto f(x | \theta)$ is a probability density function with respect to λ .

We write $\mathcal{M}(\Theta)$ for the Banach space of finite signed Radon measures on Θ endowed with the total variation norm $\|\cdot\|_{\text{TV}}$, $\mathcal{P}(\Theta) \subset \mathcal{M}(\Theta)$ for the set of Borel probability measures on Θ , and $\mathcal{D}(\Theta) \subset \mathcal{P}(\Theta)$ for the set of discrete probability measures (i.e., those with at most countable support). The mixture operator $\mathcal{A} : \mathcal{M}(\Theta) \rightarrow L^1(\mathcal{X}, \lambda)$ is defined by

$$G \mapsto (\mathcal{A}G)(x) := \int_{\Theta} f(x | \theta) G(d\theta), \quad x \in \mathcal{X},$$

and we write $p_G := \mathcal{A}G$ whenever $G \in \mathcal{P}(\Theta)$, so that p_G is a probability density with respect to λ .

The mixture model (1) is said *identifiable* if, for all $G_1, G_2 \in \mathcal{P}(\Theta)$, the equality $p_{G_1} = p_{G_2}$, holding λ -almost everywhere, implies $G_1 = G_2$. By linearity of the mixture operator, identifiability is implied by the injectivity of $\mathcal{A}$ on $\mathcal{M}(\Theta)$; that is, the model is identifiable if

$$\mathcal{A}G = 0 \text{ } \lambda\text{-almost everywhere} \implies G = 0$$

for all $G \in \mathcal{M}(\Theta)$, where $G = 0$ indicates that G is the identically zero signed measure. Also, following the seminal contribution of [Nguyen \(2013\)](#) we will, when discussing statistical estimation, measure the discrepancy between any two mixing measures $G_1, G_2 \in \mathcal{P}(\Theta)$ according to the *Wasserstein distance of order 1*,

$$W_1(G_1, G_2) := \inf_{\pi \in \Pi(G_1, G_2)} \int_{\Theta} \|\theta_1 - \theta_2\| \pi(d\theta_1, d\theta_2),$$

where $\Pi(G_1, G_2)$ is the set of probability measures π on Θ^2 with left- and right-marginal G_1 and G_2 , respectively ([Villani, 2008](#)).

For a multi-index $\alpha = (\alpha_1, \dots, \alpha_r) \in \mathbb{N}_0^r$, where $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$, we set $|\alpha| := \alpha_1 + \dots + \alpha_r$ and write $\partial_{\theta}^{\alpha} := \partial^{|\alpha|} / \partial \theta_1^{\alpha_1} \dots \partial \theta_r^{\alpha_r}$, with the convention that $\partial_{\theta}^{\alpha}$ is the identity when $|\alpha| = 0$. To ease notation, we will often omit $\alpha \in \mathbb{N}^r$ from subscripts and set definitions, and will instead directly state the required conditions on α (e.g., in terms of its size $|\alpha|$). For a non-empty open set $U \subseteq \mathbb{R}^{d_2}$ and $k \in \mathbb{N}_0 \cup \{\infty\}$, $C^k(U)$ denotes the space of k times continuously differentiable real functions on U , and $C_c^{\infty}(U)$ the space of infinitely differentiable functions whose support is a compact subset of U ; elements of $C_c^{\infty}(U)$ are referred to as test functions. For real functions w_1, w_2 on Θ with $w_1 w_2$ integrable, we write $\langle w_1, w_2 \rangle := \int_{\Theta} w_1(\theta) w_2(\theta) d\theta$ for the $L^2(\mathcal{X})$ inner product with respect to the Lebesgue measure on Θ . We denote by $B(\theta_0, \rho)$ the open Euclidean ball of centre θ_0 and radius ρ ,

¹A subset S of Euclidean space $\mathbb{R}^d$ is called *regular closed set* if S equals to the closure of its interior [Willard \(1970\)](#).

by $\bar{A}$ the closure of a set A , and by $e_1, \dots, e_d$ the standard basis of $\mathbb{R}^d$. For $h \in L^1(\mathbb{R}^d)$, the Fourier transform of h is

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

Analogously, for a finite measure μ on $\mathbb{R}^d$, its characteristic function is denoted with

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

Finally, we call a map between metric spaces a homeomorphism onto its image if it is a continuous bijection onto its image with continuous inverse. We refer to [Folland \(1999\)](#) for further background on the previous and other functional-analytic tools employed throughout the article.

3 Relation to existing work

Identifiability and estimation in mixture models have a long history, going back to the work of [Teicher \(1960, 1961, 1963\)](#). Much of the recent literature has dealt with estimation of the density p_{G_*} , with [Ghosal et al. \(1999\)](#); [Barron et al. \(1999\)](#); [Ghosal et al. \(2000\)](#); [Ghosal and van der Vaart \(2001\)](#) pioneering the study of L^1 consistency and contraction rates in the infinite-dimensional setting. The problem of quantifying convergence at the level of the mixing measure has been recently formalised by [Nguyen \(2013\)](#), who advocated for the Wasserstein distances as natural tools toward this aim on account of their geometrically meaningful behaviour between measures with arbitrary supports. This perspective has generated a growing body of work on estimation of the mixing measure ([Ho and Nguyen, 2016a,b](#); [Gao and van der Vaart, 2016](#); [Scricciolo, 2018](#); [Rousseau and Scricciolo, 2024](#); [Bariletto et al., 2026a](#); [Ascolani, 2026](#)). A central device in these analyses is the search for inequalities of the form $\int_{\mathcal{X}} |p_{G_1}(x) - p_{G_2}(x)| \lambda(dx) \geq \varphi(W_r(G_1, G_2))$, for some increasing φ and a Wasserstein distance W_r of order $r \geq 1$ (see [Villani, 2008](#), for a definition when $r > 1$), which convert density rates into mixing-measure rates ([Nguyen and Wei, 2026](#)). Such an inequality is possible in general only when $\mathcal{A}$ is injective, so that establishing it may be viewed as a quantitative refinement of identifiability itself; the attainable φ , and hence the estimation rates, have in turn been shown to be governed by analytic features of the kernel, notably the decay of its characteristic function and the smoothness of the parameterisation.

Work addressing identifiability directly, rather than through estimation rates, has largely been confined to finite mixing measures. In this setting, [Ho and Nguyen \(2016a,b\)](#); [Manole and Ho \(2022\)](#) showed that the identifiability and convergence behaviour of the mixing measure is controlled by some algebraic relations among the parameter derivatives of the kernel: when certain families of such derivatives of f are linearly dependent, distinct configurations of atoms become difficult or impossible to separate, degrading identifiability and estimation rates; see also [Nguyen and Wei \(2026\)](#) for a recent review and [Ho et al. \(2022\)](#); [Nguyen et al. \(2023\)](#); [Do et al. \(2025\)](#); [Bariletto et al. \(2026b\)](#) for extensions to the covariate-dependent case. Differential structure in the parameter has thus already been recognised as a key mechanism in the finite mixture setting, and the aim of this article is to develop these connections in the infinite-dimensional setting as well. It is notable that, in part also owing to applications in deconvolution problems ([Caillerie et al., 2011](#); [Dedecker and Michel, 2013](#); [Dedecker et al., 2015](#)) where the noise kernel $f(\cdot | \theta)$ is typically well understood and need not be very flexible, most of the literature on infinite mixing-measure estimation has focused on relatively simple kernel structures, such as location families or, only recently, location-scale families with a

scale parameter shared across components but estimated from the data (Bariletto et al., 2026a). As our analysis reveals, the scarcity of results on more general and widely used families, such as location-scale ones, is not accidental; rather, many of these flexible parameterisations give rise to differential structures that are incompatible with an identifiable model, rendering the recovery of G_* from a density estimate $p_{\hat{G}}$ impossible in general.

4 PDE barriers to mixing-measure identifiability

We are now in a position to formally initiate our study of identifiability of mixing measures in the context of infinite mixture models. In this section, we examine two types of PDE barriers to identifiability and provide examples that illustrate both.

4.1 Non-identifiability via parameter PDEs

The first type of PDE barrier to identifiability we examine is induced by the following kind of differential operator.

Definition 1. Let $U \subseteq \Theta^\circ$ be open and $r \in \mathbb{N}_0$. A parameter differential operator on U of order r is a map $\mathcal{L}_\theta$ acting on functions $w \in C^r(U)$ by the relation

$$\mathcal{L}_\theta w = \sum_{|\alpha| \leq r} c_\alpha(\theta) \partial_\theta^\alpha w, \quad \theta \in U,$$

with coefficients $c_\alpha \in C^{|\alpha|}(U)$. Its formal adjoint is $\mathcal{L}_\theta^* w = \sum_{|\alpha| \leq r} (-1)^{|\alpha|} \partial_\theta^\alpha (c_\alpha w)$, so that $\langle \mathcal{L}_\theta w_1, w_2 \rangle = \langle w_1, \mathcal{L}_\theta^* w_2 \rangle$ for all $w_1 \in C^r(U)$ and $w_2 \in C_c^\infty(U)$. The operator is non-trivial on a set $A \subseteq U$ if c_α is not identically zero on A for at least one α with $|\alpha| \geq 1$.

Accordingly, we say that the kernel f satisfies a *parameter PDE* on U if $\mathcal{L}_\theta f(x \mid \theta) = 0$ for λ -almost every $x \in \mathcal{X}$ and every $\theta \in U$, for some parameter differential operator $\mathcal{L}_\theta$ on U . As anticipated, the presence of such parameter PDE relations leads to non-identifiability of the mixing measure.

Theorem 1. Let $U \subseteq \Theta^\circ$ be open and let f satisfy a parameter PDE on U for a parameter differential operator $\mathcal{L}_\theta$ that is non-trivial on an open ball B with $\bar{B} \subset U$. Let $G_* \in \mathcal{P}(\Theta)$ admit a density g_* with respect to the Lebesgue measure on Θ such that $g_* \geq m$ Lebesgue-almost everywhere on B , for some $m > 0$. Then there exists $G \in \mathcal{P}(\Theta)$ with $G \neq G_*$ and $p_G = p_{G_*}$ λ -almost everywhere.

The proof of Theorem 1 details the construction of the probability density p_G obtained by perturbing the mixing measure g_* by an additive term of the form $\epsilon \bar{h}$, where $\bar{h}$ is a function on Θ such that $\int_\Theta \bar{h}(\theta) d\theta = 0$ and $\epsilon > 0$ is appropriately small. It is worth emphasizing that, since $\bar{h}$ and ϵ depends only on B and m , this construction is generic: there is a continuum of choices of $\bar{h}$ and ϵ , each leading to a mixing measure that is indistinguishable, based only on the induced mixture density, from G_* . Secondly, the conclusion is local in the parameter space: only the behavior of $\mathcal{L}_\theta$ and of g_* on the ball B matters, so that a kernel satisfying a parameter PDE on an arbitrarily small open region generates non-identifiability for every mixing measure whose density is bounded away from zero there. In particular, if g_* is continuous and bounded away from 0 on Θ° , the assumptions of the Theorem reduce to the non-triviality of $\mathcal{L}_\theta$.

Theorem 1 links non-identifiability to the existence of an annihilating differential operator. The next two results give sufficient conditions for this kind of PDE barriers to identifiability to exist, the first for exponential families and the second for a broader class in which the score function is a polynomial in a finite collection of statistics.

Definition 2. *The kernel f belongs to an exponential family with sufficient statistic of dimension $q \in \mathbb{N}$ if*

$$f(x | \theta) = h(x) \exp \left\{ \sum_{i=1}^q \eta_i(\theta) T_i(x) - A(\theta) \right\}, \quad x \in \mathcal{X}, \theta \in \Theta, \quad (5)$$

where $h : \mathcal{X} \rightarrow (0, \infty)$ is measurable, $T = (T_1, \dots, T_q) : \mathcal{X} \rightarrow \mathbb{R}^q$ is a vector of sufficient statistics, $\eta = (\eta_1, \dots, \eta_q) : \Theta \rightarrow \mathbb{R}^q$ is the natural parameter map, and $A : \Theta \rightarrow \mathbb{R}$ is the log-partition function.

Proposition 1. *Let f be an exponential family kernel as in (5), with η and A infinitely differentiable on Θ° , and suppose that $q < d_2$. Then there exist a non-empty open set $U \subseteq \Theta^\circ$ on which the kernel family satisfies a non-trivial parameter PDE.*

The proof of Proposition 1 hinges upon the rank deficiency of the natural parameter map. Since η takes values in $\mathbb{R}^q$ with $q < d_2$, its Jacobian $J_\eta(\theta)$, of dimension $q \times d_2$, has a non-trivial kernel at every θ . This makes it possible to construct a parameter differential operator that annihilates f on an open subset of Θ .

Outside the class of exponential families, the score need not be linear in a finite set of statistics. As the next result shows, a polynomial dependence suffices to deduce the presence of a parameter PDE from a similar dimension mismatch.

Definition 3. *The kernel f has a polynomial score of degree $p \geq 1$ relative to $S = (S_1, \dots, S_q) : \mathcal{X} \rightarrow \mathbb{R}^q$ if $\theta \mapsto f(x | \theta)$ is infinitely differentiable on Θ° for every $x \in \mathcal{X}$ and, for each $j \in \{1, \dots, d_2\}$,*

$$\frac{\partial_{\theta_j} f(x | \theta)}{f(x | \theta)} = P_j(S(x); \theta), \quad x \in \mathcal{X}, \theta \in \Theta^\circ,$$

where $P_j(\cdot; \theta)$ is a polynomial in q variables of total degree at most p whose coefficients are infinitely differentiable in θ for almost every $x \in \mathcal{X}$.

Proposition 2. *Let f have a polynomial score of degree p relative to $S : \mathcal{X} \rightarrow \mathbb{R}^q$, and suppose that $q < d_2$. Then there exist a non-empty open set $U \subseteq \Theta^\circ$ on which the kernel family satisfies a non-trivial parameter PDE.*

The proof of the previous result also relies on a comparison of dimensions, with the order of the annihilating operator governed by the mismatch between d_2 and q and by the degree p . The previous result is similar in spirit to the exponential family case of Proposition 1, where $p = 1$ and the dependence of $\log f$ on S is linear.

4.2 Non-identifiability via shift parameter PDEs

Several standard families are not subject to a parameter PDE in the sense of Definition 1, while they do satisfy a recurrence relation linking the kernel at θ with the kernel at translated parameter values, possibly combined with differentiation. While we postpone concrete examples, the following definition accommodates such relations.

Definition 4. Let $V = \{v_1, \dots, v_J\} \subset \mathbb{R}^{d_2}$ be a finite set of distinct shift vectors with $v_1 = 0$. Furthermore, let $r_1, \dots, r_J \in \mathbb{N}_0$, and $U \subseteq \Theta^\circ$ be open and such that $U + v_j \subseteq \Theta^\circ$ for all j . A difference-differential parameter operator $\mathcal{L}_\theta$ with shifts V is defined to act on functions $w \in C^{\max_j r_j}(\bigcup_j (U + v_j))$ by

$$\mathcal{L}_\theta w = \sum_{j=1}^J \sum_{|\alpha| \leq r_j} c_{j,\alpha}(\theta) (\partial_\theta^\alpha w)(\theta + v_j), \quad \theta \in U,$$

with coefficients $c_{j,\alpha} \in C^{|\alpha|}(U)$. The operator is non-trivial on a set $A \subseteq U$ if $c_{j,\alpha}$ is not identically zero on A for at least one pair (j, α) .

Accordingly, we say that f satisfies a *shift parameter PDE* on U if there exists a difference-differential parameter operator $\mathcal{L}_\theta$, defined on some open $U \subseteq \Theta^\circ$ in the sense of Definition 4, such that $\mathcal{L}_\theta f(x | \theta) = 0$ for λ -almost every $x \in \mathcal{X}$ and every $\theta \in U$.

Theorem 2. Let $\mathcal{L}_\theta$ be a difference-differential parameter operator with shifts $V = \{v_1, \dots, v_J\}$ on an open set $U \subseteq \Theta^\circ$, and let f satisfy the associated shift parameter PDE on U . Let $B = B(\theta_0, \rho) \subseteq U$ be an open ball such that $\overline{B} + v_j \subset \Theta^\circ$ for all $j = 1, \dots, J$ and whose radius ρ satisfies $\rho < \frac{1}{2} \min_{i \neq j} \|v_i - v_j\|$. Suppose that $\mathcal{L}_\theta$ is non-trivial on B . If the true mixing measure $G_* \in \mathcal{P}(\Theta)$ admits a density g_* with respect to the Lebesgue measure on Θ satisfying $g_* \geq m$ Lebesgue-almost everywhere on $\bigcup_{j=1}^J (B + v_j)$, for some $m > 0$, then there exists a $G \in \mathcal{P}(\Theta)$ with $G \neq G_*$ such that $p_G = p_{G_*}$ λ -almost everywhere.

The construction behind Theorem 2 parallels that of Theorem 1, whereby an appropriate test function is used to build a perturbation of g_* that, by exploiting the non-trivial null space of the difference-differential parameter operator, induces non-identifiability of the mixing measure.

4.3 Examples

We now discuss concrete kernel families for which the hypotheses of Theorems 1 or 2 are satisfied, and which therefore lead to non-identifiable infinite mixture models. The variety and popularity of these families are a testament to the pervasiveness of the PDE barriers to identifiability uncovered by our analysis.

Location-scale Gaussian kernel. With $\mathcal{X} = \mathbb{R}$ and $f(x | \mu, \nu) = \mathcal{N}(x | \mu, \nu)$, and for $\theta = (\mu, \nu)$ in a compact subset of $\mathbb{R} \times (0, \infty)$ with non-empty interior, the heat equation (4) gives a parameter PDE of order two with operator $\mathcal{L}_\theta = \partial_\nu - \frac{1}{2} \partial_\mu^2$, so Theorem 1 applies. This case was already discussed in Section 1.1.

Beta and Dirichlet kernels. Let $\mathcal{X} = (0, 1)$ and $f(x | \alpha, \beta) = [\Gamma(\alpha + \beta) / \Gamma(\alpha) \Gamma(\beta)] x^{\alpha-1} (1-x)^{\beta-1}$, the beta kernel, with (α, β) ranging in a compact subset Θ of $(0, \infty)^2$. Since $\alpha(\alpha + \beta)^{-1} f(x | \alpha + 1, \beta) = x f(x | \alpha, \beta)$ and $\beta(\alpha + \beta)^{-1} f(x | \alpha, \beta + 1) = (1-x) f(x | \alpha, \beta)$ for all $x \in \mathcal{X}$ and $(\alpha, \beta) \in \Theta$, we have

$$\frac{\alpha}{\alpha + \beta} f(x | \alpha + 1, \beta) + \frac{\beta}{\alpha + \beta} f(x | \alpha, \beta + 1) - f(x | \alpha, \beta) = 0,$$

a shift parameter PDE of order zero with shifts $(0, 0)$, $(1, 0)$ and $(0, 1)$. The same argument applies to the more general Dirichlet family on the open simplex $\Delta_{d-1} = \{x \in \mathbb{R}^d : x_i > 0, \sum_{i=1}^d x_i = 1\}$ with density $f(x \mid \alpha) = \Gamma(\sum_{i=1}^d \alpha_i) \prod_{i=1}^d (x_i^{\alpha_i-1} / \Gamma(\alpha_i))$, for which

$$\sum_{j=1}^d \frac{\alpha_j}{\sum_{m=1}^d \alpha_m} f(x \mid \alpha + e_j) - f(x \mid \alpha) = 0,$$

with shifts $0, e_1, \dots, e_d$.

Location-scale Student- t kernel with fixed degrees of freedom. Let $\mathcal{X} = \mathbb{R}$, let $\nu > 0$ be fixed, and let

$$f(x \mid \mu, \sigma) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)(\nu\pi)^{1/2}\sigma} \left[1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma} \right)^2 \right]^{-(\nu+1)/2},$$

with $\theta = (\mu, \sigma)$ in a compact subset of $\mathbb{R} \times (0, \infty)$. Direct computation gives the parameter PDE of order two

$$[\sigma \partial_\sigma^2 + \nu \sigma \partial_\mu^2 + (1-\nu) \partial_\sigma] f(x \mid \mu, \sigma) = 0,$$

so that Theorem 1 applies.

Gamma kernel. Let $\mathcal{X} = (0, \infty)$ and $f(x \mid \gamma, \beta) = \beta^\gamma x^{\gamma-1} e^{-\beta x} / \Gamma(\gamma)$, with $\theta = (\gamma, \beta)$ in a compact subset of $(0, \infty)^2$. Using the identities $xf(x \mid \gamma, \beta) = \gamma \beta^{-1} f(x \mid \gamma+1, \beta)$ and $\partial_\beta f = \gamma \beta^{-1} f - xf$ gives

$$\frac{\beta}{\gamma} \partial_\beta f(x \mid \gamma, \beta) - f(x \mid \gamma, \beta) + f(x \mid \gamma+1, \beta) = 0,$$

a shift parameter PDE with shifts $v_1 = (0, 0)$ and $v_2 = (1, 0)$ and orders $r_1 = 1, r_2 = 0$; Theorem 2 applies whenever the parameter space accommodates a ball together with its translate by v_2 .

Negative binomial kernel with varying number of successes. Let $\mathcal{X} = \mathbb{N}_0$, let λ be the counting measure, and let $f(x \mid r, p) = \frac{\Gamma(x+r)}{x! \Gamma(r)} p^r (1-p)^x$, with (r, p) in a compact subset of $(0, \infty) \times (0, 1)$. Then

$$f(x \mid r+1, p) + \frac{p(1-p)}{r} \partial_p f(x \mid r, p) - f(x \mid r, p) = 0,$$

a shift parameter PDE with shifts $(0, 0)$ and $(1, 0)$.

Non-central chi-squared kernel. Let $\mathcal{X} = (0, \infty)$ and

$$f(x \mid k, \lambda_{\text{nc}}) = \frac{1}{2} e^{-(x+\lambda_{\text{nc}})/2} \left(\frac{x}{\lambda_{\text{nc}}} \right)^{k/4-1/2} I_{k/2-1}((\lambda_{\text{nc}} x)^{1/2}),$$

where $I_\kappa(z) = \sum_{j \geq 0} (z/2)^{2j+\kappa} / [j! \Gamma(j+\kappa+1)]$ is the modified Bessel function of the first kind, the degrees of freedom k and the non-centrality parameter λ_{nc} ranging over a compact subset of $(0, \infty)^2$. The classical recurrence

$$\partial_{\lambda_{\text{nc}}} f(x \mid k, \lambda_{\text{nc}}) - \frac{1}{2} f(x \mid k+2, \lambda_{\text{nc}}) + \frac{1}{2} f(x \mid k, \lambda_{\text{nc}}) = 0$$

is a shift parameter PDE with shifts $(0, 0)$ and $(2, 0)$.

4.4 Consequences for statistical estimation

Having established that a wide range of kernel families induce non-identifiable infinite mixture models, we now briefly discuss the consequences of this phenomenon for statistical estimation. Since the large majority of mixtures used in practice model the mixing measure as discrete, the following result focuses on that case, although considering general members of $\mathcal{P}(\Theta)$ would not alter the conclusions.

Proposition 3. *Let $\{f(\cdot \mid \theta) : \theta \in \Theta\}$ be a kernel family such that the mapping $\theta \mapsto f(x \mid \theta)$ is continuous on Θ for λ -almost every x . If there exist two mixing measures $G_1, G_2 \in \mathcal{P}(\Theta)$ with $G_1 \neq G_2$ and $p_{G_1} = p_{G_2}$ λ -almost everywhere, then*

$$\inf_{n \in \mathbb{N}} \inf_{\hat{G}_n} \sup_{G_* \in \mathcal{D}(\Theta)} \mathbb{E}_{X_{1:n} \sim P_{G_*}^n} [W_1(\hat{G}_n(X_{1:n}), G_*)] \geq \frac{3c}{32},$$

where $c = W_1(G_1, G_2)$ and the inner infimum is taken over all estimators $\hat{G}_n : \mathcal{X}^n \rightarrow \mathcal{D}(\Theta)$.

The proof of Proposition 3 relies on standard tools from minimax theory together with the W_1 -density of $\mathcal{D}(\Theta)$ in $\mathcal{P}(\Theta)$. The above results shows that fully unconstrained estimation of the mixing measure is inconsistent, in the worst-case sense, whenever identifiability fails. In particular, it shows that our non-identifiability results, though stated for mixing measures with a locally positive density with respect to the Lebesgue measure on Θ , carry strong implications for estimation even in the seemingly more structured class of discrete mixtures. Hence, unless further knowledge of the structure of G_* is available and used to constrain estimation—how to do so being an open problem—recovery of G_* from a growing sample is made impossible, in the worst case, by the non-identifiability phenomena uncovered by our theory.

5 Identifiability-preserving kernel families

The results of Section 4 identify specific differential and difference-differential structures in the kernel parameterisation as mechanisms impeding the recovery of the underlying mixing measure from a given mixture density. As a complement to these negative results, we now describe three classes of kernels for which the mixture operator is injective, which ensures identifiability and restores the hope for consistent unconstrained estimation of G_* .

5.1 Generalized translation families

The first such class consists of kernels whose dependence on the parameter is, after a change of variables, that of a location family.

Definition 5. *The kernel f belongs to a generalized translation family if*

$$f(x \mid \theta) = h(x) \gamma(\theta) K(\tau(x) - \psi(\theta)), \quad x \in \mathcal{X}, \theta \in \Theta, \quad (6)$$

where $h : \mathcal{X} \rightarrow (0, \infty)$ and $\gamma : \Theta \rightarrow (0, \infty)$ are continuous, $\tau : \mathcal{X} \rightarrow \mathbb{R}^{d_1}$ is a homeomorphism onto $\mathbb{R}^{d_1}$, $\psi : \Theta \rightarrow \mathbb{R}^{d_1}$ is a homeomorphism onto its image, and $K \in L^1(\mathbb{R}^{d_1})$.

Proposition 4. *Let f belong to a generalized translation family and suppose that $\mathcal{F}[K](\xi) \neq 0$ for Lebesgue-almost every $\xi \in \mathbb{R}^{d_1}$. Then $\mathcal{A}$ is injective.*

The structure of generalized translation classes reduces the proof of Proposition 4 to a classical Fourier-analytic deconvolution argument. Moreover, the non-vanishing requirement on $\mathcal{F}[K]$ is mild. Indeed, suppose $d_1 = 1$ and $\int_{\mathbb{R}} |K(x)|e^{c|x|} dx < \infty$ for some $c > 0$. Then $\mathcal{F}[K]$ extends to a holomorphic function on the strip $\{z \in \mathbb{C} : |\operatorname{Im}(z)| < c\}$. A set of positive Lebesgue measure on $\mathbb{R}$ has an accumulation point in $\mathbb{R}$, which lies inside the strip; were $\mathcal{F}[K]$ to vanish on such a set, the identity theorem would force $\mathcal{F}[K] = 0$ on the strip and hence $K = 0$ almost everywhere. Since by construction $K \neq 0$ on a set of positive Lebesgue measure, its transform $\mathcal{F}[K]$ vanishes at most on a Lebesgue-null set, so any integrable kernel with exponentially decaying tails satisfies the hypothesis of Proposition 4. Analogous arguments apply in higher dimension.

The following standard kernels belong to generalized translation families, and the resulting infinite mixture models are therefore identifiable.

- (i) *Gaussian kernel with fixed variance.* With $\mathcal{X} = \mathbb{R}$ and $f(x | \theta) = \mathcal{N}(x | \theta, \sigma^2)$ for fixed $\sigma^2 > 0$, representation (6) holds with $h(x) = (2\pi\sigma^2)^{-1/2}$, $\gamma = 1$, $\tau(x) = x$, $\psi(\theta) = \theta$ and $K(u) = \exp\{-u^2/(2\sigma^2)\}$, whose transform $\mathcal{F}[K](\xi) = (2\pi)^{1/2}\sigma \exp(-\sigma^2\xi^2/2)$ is strictly positive.
- (ii) *Gamma kernel with fixed shape.* With $\mathcal{X} = (0, \infty)$ and $f(x | \theta) = [\Gamma(\alpha)\theta^\alpha]^{-1}x^{\alpha-1}e^{-x/\theta}$ for fixed $\alpha > 0$, take $h(x) = [\Gamma(\alpha)x]^{-1}$, $\gamma(\theta) = 1$, $\tau(x) = \log x$, $\psi(\theta) = \log \theta$ and $K(u) = e^{\alpha u} \exp(-e^u)$. Here τ maps $\mathcal{X}$ onto $\mathbb{R}$, and K decays exponentially at $-\infty$ and doubly exponentially at $+\infty$, so $K \in L^1(\mathbb{R})$ has exponentially decaying tails and the remarks after Proposition 4 apply.
- (iii) *Laplace kernel with fixed scale.* With $\mathcal{X} = \mathbb{R}$ and $f(x | \theta) = (2b)^{-1} \exp(-|x - \theta|/b)$ for fixed $b > 0$, take $h = 1$, $\gamma = 1$, $\tau(x) = x$, $\psi(\theta) = \theta$ and $K(u) = (2b)^{-1}e^{-|u|/b}$, whose transform $\mathcal{F}[K](\xi) = (1 + b^2\xi^2)^{-1}$ is everywhere positive.

The above examples show that fixing an appropriate subset of parameters, such as the scale or the shape, removes the PDE barriers documented in Section 4 for the analogous kernels with more free parameters. Moreover, as already noted, location families such as (i) and (iii) are precisely those on which mixing-measure estimation has largely concentrated so far (Nguyen, 2013; Gao and van der Vaart, 2016, among others). In hindsight, our results suggest that it is exactly their identifiability, absent in the more flexible (e.g., location-scale) alternatives, that has made the recent technical progress on estimation rates possible in the corresponding infinite mixture models.

5.2 Non-degenerate exponential family

Proposition 1 shows that an exponential family whose parameter dimension exceeds the dimension q of its sufficient statistic carries a parameter PDE and therefore yields non-identifiable infinite mixture models. We now consider the complementary regime, in which the sufficient statistic has dimension at least as high as the parameter vector.

Definition 6. Let $\Omega \subset (0, \infty)^q$ be compact. A set $S \subseteq \mathbb{R}^q$ is determining for Ω if the linear span of $\{y \mapsto y^s : s \in S\}$, where $y^s := y_1^{s_1} \cdots y_q^{s_q}$ for $y \in (0, \infty)^q$, is dense in the space $C(\Omega)$ of continuous real functions on Ω equipped with the supremum norm.

Proposition 5. Let f be an exponential family kernel as in (5) such that $h > 0$ λ -almost everywhere, A is continuous on the compact parameter space Θ , T is continuous on $\mathcal{X}$, and η is injective and continuous (hence a homeomorphism onto its image $\eta(\Theta)$). If $T(\mathcal{X})$ is determining for $\exp\{\eta(\Theta)\}$, where the exponential is applied coordinate-wise, then $\mathcal{A}$ is injective.

Definition 6 is a Müntz–Szász-type condition. For $q = 1$ and $S = \{s_1, s_2, \dots\}$, the classical theorem of Müntz (1914) characterizes density in terms of the divergence of $\sum_{n \in \mathbb{N}} s_n^{-1}$. Multivariate versions of this statement are available, though the picture is less complete (Kroó, 1994; Yang, 2014). The condition is easy to verify in many cases of interest. For instance, if S contains $\mathbb{N}_0^q$, as when $\mathcal{X} = \mathbb{N}_0^q$ and T is the identity, then the span is the algebra of polynomials, which separates points in Ω and contains the constants, so the density property follows from the Stone–Weierstrass theorem. The same holds when $T(\mathcal{X}) = \mathbb{R}^q$. Moreover, notice that injectivity of η on a compact set with non-empty interior in $\mathbb{R}^{d_2}$ requires $q \geq d_2$, so the proposition deals with the regime complementary to that of Proposition 1.

The following families satisfy the hypotheses of Proposition 5:

- (i) *Poisson kernel.* With $\mathcal{X} = \mathbb{N}_0$, λ the counting measure, $f(x | \theta) = \theta^x e^{-\theta} / x!$ and Θ a compact subset of $(0, \infty)$, take $h(x) = 1/x!$, $T(x) = x$, $\eta(\theta) = \log \theta$ and $A(\theta) = \theta$. Here $q = d_2 = 1$, η is a homeomorphism onto its image, and $T(\mathcal{X}) = \mathbb{N}_0$ is determining for $\exp(\eta(\Theta)) = \Theta$ by the previous remarks;
- (ii) *Negative binomial kernel with fixed number of successes.* With $\mathcal{X} = \mathbb{N}_0$, $f(x | \theta) = \binom{x+r-1}{x} (1-\theta)^x \theta^r$ for fixed $r > 0$ and Θ a compact subset of $(0, 1)$, take $h(x) = \binom{x+r-1}{x}$, $T(x) = x$, $\eta(\theta) = \log(1-\theta)$ and $A(\theta) = -r \log \theta$, so that $T(\mathcal{X}) = \mathbb{N}$ is determining for $\exp(\eta(\Theta)) = 1 - \Theta$ by the previous remarks;
- (iii) *Multivariate Gaussian kernel with fixed covariance.* With $\mathcal{X} = \mathbb{R}^d$,

$$f(x | \theta) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (x - \theta)^\top \Sigma^{-1} (x - \theta) \right\}$$

for a fixed positive definite Σ and Θ a compact subset of $\mathbb{R}^d$, take $\eta(\theta) = \theta$, $h(x) = \exp(-\frac{1}{2} x^\top \Sigma^{-1} x) (2\pi)^{-d/2} |\Sigma|^{-1/2}$, $T(x) = \Sigma^{-1} x$, and $A(\theta) = \frac{1}{2} \theta^\top \Sigma^{-1} \theta$. Here $q = d_2 = d$ and $T(\mathcal{X}) = \mathbb{R}^d$ is obviously determining for $\exp(\eta(\Theta)) = \exp(\Theta)$.

5.3 Beyond the determining set assumption

The determining set condition of Proposition 5 constrains the range of the sufficient statistic. We conclude by showing, in an example, that identifiability can hold without it, provided the sufficient statistics satisfy appropriate conditions. Specifically, let $\mathcal{X} = \mathbb{R}$, so $d_1 = 1$, and let $\Theta = [-a, a] \times [-b, b] \subset \mathbb{R}^2$ with $a, b > 0$, so $d_2 = 2$. Consider the kernel

$$f(x | \theta_1, \theta_2) = h(x) \exp \{ \theta_1 x + \theta_2 e^{x^2} - A(\theta) \}, \quad h(x) = \exp(-2be^{x^2}), \quad (7)$$

which we call a *super-exponential kernel*. Its sufficient statistics are $T_1(x) = x$ and $T_2(x) = e^{x^2}$, and the choice of base measure guarantees integrability: since $\theta_2 - 2b \leq -b < 0$ on Θ , the exponent $\theta_1 x + (\theta_2 - 2b)e^{x^2}$ is bounded above by $\theta_1 x - be^{x^2}$, so A is finite, continuous and bounded on the compact Θ .

Proposition 6. *Let f be the kernel (7). Then,*

- (a) *the only parameter differential operator of finite order, with continuous coefficients on Θ° , that annihilates f is the trivial one, so that f satisfies no non-trivial parameter PDE;*

(b) $\mathcal{A}$ is injective.

Both statements follow from a mismatch in growth between the two sufficient statistics. Moreover, we emphasize that the construction is not specific to the statistic e^{x^2} : the same argument applies to kernels of the form $f(x | \theta_1, \theta_2) = h(x) \exp\{\theta_1 x + \theta_2 e^{P(x)} - A(\theta)\}$ for any polynomial P of even degree with positive leading coefficient (with the base measure adjusted accordingly).

Remark. *Proposition 6 exhibits a kernel family for which (a) the absence of a parameter PDE and (b) identifiability of the mixing measure coexist. Whether the absence of these or similar differential structures is sufficient for identifiability in more general scenarios, and not just necessary for it as discussed in Section 4, remains an open question.*

6 Discussion

We have identified differential structure in the parameterisation of the kernel as a fundamental barrier to the identifiability of the mixing measure in infinite mixture models. When the kernel is annihilated by a non-trivial parameter differential operator, or by a difference-differential operator involving shifted parameter values, distinct mixing measures induce the same mixture density, and the mixing measure cannot be recovered in general. Such operators have been shown to exist for a wide range of standard families, where over-parameterisation emerges as a systematic route to non-identifiability. Complementing these barriers, we have exhibited three classes of kernels for which the mixture operator is injective, showing that the phenomenon is not universal and that identifiability can be recovered by suitably restricting the parameterisation.

From our analysis, two main questions emerge as worthy of future investigation. The first concerns the exact scope of the differential barrier mechanism. Our results establish that the presence of a parameter PDE is sufficient for non-identifiability, but not that its absence is sufficient for identifiability. The super-exponential family of Section 5.3 shows that the two properties can indeed coexist: there, no non-trivial parameter PDE exists and, at the same time, the mixture operator is injective. Whether the absence of differential structure is sufficient for identifiability in general, or whether further obstructions may arise, remains to be understood.

The second question concerns the implications of our results for statistical estimation of the mixing measure. Proposition 3 shows that, when identifiability fails, no estimator can recover the mixing measure in the worst case, so that our findings make the study of consistent and fully unconstrained nonparametric estimation hopeless. This worst-case analysis, however, leaves open the question of which constraints on the model class or on the true mixing measure may rule out the perturbations underlying our constructions and restore consistent estimation. For instance, the recent results of [Bariletto et al. \(2026a\)](#), which establish convergence rates for the mixing measure in location-scale mixtures with a scale parameter shared across components and learned from the data, may be read precisely as imposing one such constraint on the model class and the ground truth, i.e., that the scale coordinate be held common. A more systematic understanding of which constraints, of this or other kinds, suffice to restore identifiability would give a more nuanced picture of statistical estimation in infinite mixture models.

Supplementary Material

In this Supplementary Material, we collect the proofs of all the results presented in the main text of the article (Appendix A), and the statements and proofs of auxiliary results (Appendix B).

A Proofs of the results in the main text

Throughout this appendix we use the notation of Sections 2–5. For a parameter differential operator $\mathcal{L}_\theta = \sum_{|\alpha| \leq r} c_\alpha(\theta) \partial_\theta^\alpha$ on an open set $U \subseteq \Theta^\circ$, recall its formal adjoint $\mathcal{L}_\theta^* w = \sum_{|\alpha| \leq r} (-1)^{|\alpha|} \partial_\theta^\alpha (c_\alpha w)$, and the pairing $\langle w_1, w_2 \rangle = \int_\Theta w_1(\theta) w_2(\theta) d\theta$.

A.1 Proof of Theorem 1

Step 1. For $u \in C_c^\infty(B)$ and $v \in C^r(U)$, integration by parts on B , with no boundary contributions since u and its derivatives vanish near ∂B , gives

$$\langle \mathcal{L}_\theta v, u \rangle = \sum_{|\alpha| \leq r} \int_B c_\alpha(\theta) (\partial_\theta^\alpha v(\theta)) u(\theta) d\theta = \sum_{|\alpha| \leq r} (-1)^{|\alpha|} \int_B v(\theta) \partial_\theta^\alpha (c_\alpha(\theta) u(\theta)) d\theta = \langle v, \mathcal{L}_\theta^* u \rangle,$$

where $\bar{h} := \mathcal{L}_\theta^* u$ is continuous and compactly supported in B .

Step 2. We first show that $\mathcal{L}_\theta u = 0$ for all $u \in C_c^\infty(B)$ is impossible. Since $\mathcal{L}_\theta$ is non-trivial on B , there is $\theta_* \in B$ and a multi-index β with $|\beta| \geq 1$ and $c_\beta(\theta_*) \neq 0$. Choose $\varphi \in C_c^\infty(B)$ equal to 1 in a neighborhood of θ_* and set $u_\beta(\theta) = \frac{1}{\beta!} \varphi(\theta) (\theta - \theta_*)^\beta$, so that $\partial_\theta^\alpha u_\beta(\theta_*) = \mathbf{1}_{\{\alpha = \beta\}}$. Then, $\mathcal{L}_\theta u_\beta(\theta_*) = \sum_{|\alpha| \leq r} c_\alpha(\theta_*) \partial_\theta^\alpha u_\beta(\theta_*) = c_\beta(\theta_*) \neq 0$. Consequently, if $\mathcal{L}_\theta^* u = 0$ for every $u \in C_c^\infty(B)$, then $\langle \mathcal{L}_\theta v, u \rangle = \langle v, \mathcal{L}_\theta^* u \rangle = 0$ for all $u, v \in C_c^\infty(B)$, forcing $\mathcal{L}_\theta v \equiv 0$ on B for all v , which was shown to be impossible. Hence there exists $u \in C_c^\infty(B)$ with $\bar{h} := \mathcal{L}_\theta^* u \neq 0$.

Step 3. By Fubini's theorem and $\int_{\mathcal{X}} f(x | \theta) \lambda(dx) = 1$,

$$\int_\Theta \bar{h}(\theta) d\theta = \int_{\mathcal{X}} \left(\int_\Theta \bar{h}(\theta) f(x | \theta) d\theta \right) \lambda(dx),$$

while the parameter PDE and the adjoint identity give, for λ -almost every x ,

$$\int_\Theta f(x | \theta) \bar{h}(\theta) d\theta = \langle f(x | \cdot), \mathcal{L}_\theta^* u \rangle = \langle \mathcal{L}_\theta f(x | \cdot), u \rangle = 0.$$

Hence $\int_\Theta \bar{h} d\theta = 0$ and $\int_\Theta f(x | \theta) \bar{h}(\theta) d\theta = 0$ for λ -almost every x .

Step 4. Let $M := \sup_{\theta \in \Theta} |\bar{h}(\theta)| \in (0, \infty)$ and set $\varepsilon := m/(2M) > 0$. Define $g := g_* + \varepsilon \bar{h}$. On B we have $g \geq m - \varepsilon M = m/2 > 0$, and on $\Theta \setminus B$ we have $g = g_* \geq 0$, since $\bar{h}$ is supported in B ; thus $g \geq 0$ almost everywhere. Moreover $\int_\Theta g d\theta = \int_\Theta g_* d\theta + \varepsilon \int_\Theta \bar{h} d\theta = 1$, so g is the density of a probability measure $G \in \mathcal{P}(\Theta)$. Finally, for λ -almost every x ,

$$p_G(x) - p_{G_*}(x) = \varepsilon \int_\Theta f(x | \theta) \bar{h}(\theta) d\theta = 0.$$

At the same time, because $\bar{h}$ is non-zero at θ_* and continuous on B , it is non-zero on a set of positive Lebesgue measure. Therefore $G \neq G_*$, proving the claim.

A.2 Proof of Proposition 1

Write $f(x | \theta) = h(x) \exp\{\sum_{i=1}^q \eta_i(\theta) T_i(x) - A(\theta)\}$. Differentiating in θ_j , we obtain

$$\partial_{\theta_j} f(x | \theta) = f(x | \theta) \left(\sum_{i=1}^q \partial_{\theta_j} \eta_i(\theta) T_i(x) - \partial_{\theta_j} A(\theta) \right).$$

For a first-order operator $\mathcal{L}_\theta = \sum_{j=1}^{d_2} c_j(\theta) \partial_{\theta_j} + c_0(\theta)$, we get

$$\mathcal{L}_\theta f(x | \theta) = f(x | \theta) \left[\sum_{i=1}^q T_i(x) \sum_{j=1}^{d_2} c_j(\theta) \partial_{\theta_j} \eta_i(\theta) - \sum_{j=1}^{d_2} c_j(\theta) \partial_{\theta_j} A(\theta) + c_0(\theta) \right].$$

Since $f > 0$, the operator annihilates f on an open set $U \subseteq \Theta^\circ$ so long as

$$\sum_{j=1}^{d_2} c_j(\theta) \partial_{\theta_j} \eta_i(\theta) = 0 \quad (1 \leq i \leq q) \quad \text{and} \quad c_0(\theta) = \sum_{j=1}^{d_2} c_j(\theta) \partial_{\theta_j} A(\theta), \quad (8)$$

for all $\theta \in U$. Below, we will demonstrate a choice of U and of the functions $(c_0(\theta), c_1(\theta), \dots, c_{d_2}(\theta))$ for each $\theta \in U$ satisfying the above relationship.

First, notice that by writing $J_\eta(\theta)$ for the $q \times d_2$ Jacobian with $9i, j$ -th entry $[J_\eta(\theta)]_{ij} = \partial_{\theta_j} \eta_i(\theta)$ and letting $\mathbf{c}(\theta) = (c_1(\theta), \dots, c_{d_2}(\theta))^\top$, the first set of equations is equivalent to

$$J_\eta(\theta) \mathbf{c}(\theta) = 0. \quad (9)$$

Let $r := \max_{\theta \in \Theta^\circ} \text{rank}(J_\eta(\theta))$ and $R := \{\theta \in \Theta^\circ : \text{rank}(J_\eta(\theta)) = r\}$. By construction, for any $\theta' \in R$, there exists some $r \times r$ submatrix $M(\theta')$ of $J_\eta(\theta')$ such that $\det M(\theta') \neq 0$. Without loss of generality, we can assume that $M(\theta')$ is the principal submatrix (i.e. it occupies the first r rows and the first r columns) of the Jacobian $J_\eta(\theta')$. Since the entries $\partial_{\theta_j} \eta_i$ are continuous and the determinant is a multilinear polynomial in those entries (and therefore continuous), there exists a $\delta > 0$ such that $U := B(\theta', \delta) \subset R$ and, for each θ in U , $\text{rank}(J_\eta(\theta)) = \text{rank}(M(\theta)) = r$, where $M(\theta)$ is the leading principal submatrix of $J_\eta(\theta)$ of dimension $r \times r$.

Write $\mathbf{c}_A(\theta) = (c_1(\theta), \dots, c_r(\theta))^\top$ and $\mathbf{c}_B(\theta) = (c_{r+1}(\theta), \dots, c_{d_2}(\theta))^\top$. Because $d_2 - r \geq d_2 - q > 0$, $\mathbf{c}_B(\theta)$ contains at least one coordinate. Let $B(\theta)$ denote the $r \times (d_2 - r)$ submatrix of $J_\eta(\theta)$ adjacent to $M(\theta)$, and let $\mathbf{b}_1(\theta)$ be its first column. We now set $\mathbf{c}_B(\theta)$ as the one-shot vector $(1, 0, \dots, 0)^\top$ for every $\theta \in U$ and show that a solution to (9) exists for every $\theta \in U$. Indeed, the identity $J_\eta(\theta) \mathbf{c}(\theta) = 0$ becomes

$$M(\theta) \mathbf{c}_A(\theta) + B(\theta) \mathbf{c}_B(\theta) = 0,$$

which in turn is equivalent to

$$M(\theta) \mathbf{c}_A(\theta) = -B(\theta) \mathbf{c}_B(\theta) = -\mathbf{b}_1(\theta).$$

Since $\det(M(\theta)) \neq 0$ on U , the above system has a unique set of solutions $(c_j(\theta), 1 \leq j \leq r)$ for each $\theta \in U$, given by Cramer's rule as

$$c_j(\theta) = \frac{\det(M_j(\theta))}{\det(M(\theta))},$$

where $M_j(\theta)$ is formed by replacing the j -th column of $M(\theta)$ with the vector $-\mathbf{b}_1(\theta)$. From this choice of $\mathbf{c}(\theta)$ it follows that

$$\sum_{j=1}^{d_2} c_j(\theta) \frac{\partial \eta_i(\theta)}{\partial \theta_j} = 0 \quad \text{for all } 1 \leq i \leq r. \quad (10)$$

We now verify that equation (10) holds also for $i \geq r$. Let $\mathbf{r}_1(\theta), \dots, \mathbf{r}_q(\theta)$ be the row vectors of $J_\eta(\theta)$. For $i > r$, since $\text{rank}(J_\eta(\theta)) = r$ and $M(\theta)$ is invertible on U , for each $\theta \in U$ there exist real numbers $\lambda_1^i(\theta), \dots, \lambda_r^i(\theta)$ such that

$$\mathbf{r}_i(\theta) = \sum_{m=1}^r \lambda_m^i(\theta) \mathbf{r}_m(\theta).$$

As a result, for every $i > r$,

$$\sum_{j=1}^{d_2} c_j(\theta) \frac{\partial \eta_i(\theta)}{\partial \theta_j} = \langle \mathbf{c}(\theta), \mathbf{r}_i(\theta) \rangle = \sum_{m=1}^r \lambda_m^i(\theta) \langle \mathbf{c}(\theta), \mathbf{r}_m(\theta) \rangle = 0, \quad (11)$$

the final equality following from (10). Combining (10) and (11), we verify that (9) holds on U .

To complete the proof, it remains to show the second identity in (8). Towards that goal, we set, for each $\theta \in U$,

$$c_0(\theta) := \sum_{j=1}^{d_2} c_j(\theta) \frac{\partial A(\theta)}{\partial \theta_j},$$

where $(c_j(\theta), 1 \leq j \leq d_2)$ are the coefficient functions derived above. Since the entries of the Jacobian are infinitely continuously differentiable, both $\theta \in U \mapsto \det(M(\theta))$ and $\theta \in U \mapsto \det(M_j(\theta))$ are as well. As $\det(M(\theta))$ is bounded away from zero on U , we conclude that $c_j \in C^\infty(U)$ for all $j \geq 1$. This, along with the fact that $A \in C^\infty(\Theta^\circ)$, yields that $c_0 \in C^\infty(U)$, too. This completes the proof.

A.3 Proof of Proposition 2

Step 1. We prove by induction on $m = |\alpha|$ that, for every multi-index $\alpha \in \mathbb{N}_0^{d_2}$,

$$\partial_\theta^\alpha f(x \mid \theta) = Q_\alpha(S(x); \theta) f(x \mid \theta), \quad (12)$$

where $Q_\alpha(\cdot; \theta)$ is a polynomial in the q variables $S_1(x), \dots, S_q(x)$ with coefficients that lie in $C^\infty(\Theta^\circ)$ and of total degree $\deg Q_\alpha \leq p|\alpha|$.

For $|\alpha| = 0$, (12) holds with $Q_0 \equiv 1$. For $|\alpha| = 1$, say $\alpha = e_j$, Definition 3 gives $\partial_{\theta_j} f(x \mid \theta) = P_j(S(x); \theta) f(x \mid \theta)$ with $\deg P_j \leq p$, so (12) holds with $Q_{e_j} = P_j$; moreover, by assumption the coefficients of P_j are in $C^\infty(\Theta^\circ)$.

Assume (12) holds for all multi-indices of order m , and take $\beta = \alpha + e_j$, for some α with $|\alpha| = m$ and j , so that $|\beta| = m + 1$. Then

$$\partial_\theta^\beta f = \partial_{\theta_j} (\partial_\theta^\alpha f) = \partial_{\theta_j} [Q_\alpha(S(x); \theta) f(x \mid \theta)] = (\partial_{\theta_j} Q_\alpha) f + Q_\alpha (\partial_{\theta_j} f) = (\partial_{\theta_j} Q_\alpha + Q_\alpha P_j) f.$$

Define $Q_\beta := \partial_{\theta_j} Q_\alpha + Q_\alpha P_j$. Then Q_β is again a polynomial in $S_1, \dots, S_q$ with θ -dependent coefficients: differentiating Q_α in θ_j affects only its coefficients, not its degree in S , so $\deg(\partial_{\theta_j} Q_\alpha) \leq \deg Q_\alpha \leq pm$, while

$$\deg(Q_\alpha P_j) = \deg Q_\alpha + \deg P_j \leq pm + p = p(m+1),$$

whence $\deg Q_\beta \leq p(m+1) = p|\beta|$. Moreover the coefficients of Q_β are sums and products of the infinitely continuously differentiable coefficients of Q_α and P_j and of their θ -derivatives, hence themselves infinitely continuously differentiable. This completes the induction and establishes (12).

Step 2. Fix a maximal differential order $M \geq 1$. The number of parameter derivatives ∂_θ^α with $|\alpha| \leq M$ on a d_2 -dimensional parameter space equals the number of multi-indices $\alpha \in \mathbb{N}_0^{d_2}$ with $|\alpha| \leq M$, namely

$$N_{\text{der}}(M) := \binom{M+d_2}{d_2} = \frac{M^{d_2}}{d_2!} + \mathcal{O}(M^{d_2-1}) \quad (M \rightarrow \infty).$$

By (12), each such derivative, divided by f , is a polynomial in $S_1, \dots, S_q$ of total degree at most pM ; any linear combination $\sum_{|\alpha| \leq M} c_\alpha(\theta) \partial_\theta^\alpha f$ therefore equals f times a polynomial lying in the span of the monomials $S_1^{\gamma_1} \dots S_q^{\gamma_q}$ with $\gamma_1 + \dots + \gamma_q \leq pM$. The number of such monomials is at most

$$N_{\text{pol}}(M) := \binom{pM+q}{q} = \frac{p^q M^q}{q!} + \mathcal{O}(M^{q-1}) \quad (M \rightarrow \infty).$$

Since $q < d_2$,

$$\frac{N_{\text{der}}(M)}{N_{\text{pol}}(M)} = \frac{q!}{p^q d_2!} M^{d_2-q} (1 + o(1)) \xrightarrow{M \rightarrow \infty} \infty,$$

so there exists an integer $M^* \geq 1$ with $N_{\text{der}}(M^*) > N_{\text{pol}}(M^*)$. We fix such an M^* for the remainder of the proof.

Step 3. We seek coefficient functions $c_\alpha(\theta)$, indexed by α with $|\alpha| \leq M^*$, not all zero, such that

$$\mathcal{L}_\theta f(x | \theta) = \sum_{|\alpha| \leq M^*} c_\alpha(\theta) \partial_\theta^\alpha f(x | \theta) = 0$$

for λ -almost every $x \in \mathcal{X}$ and all θ in some open $U \subset \Theta^\circ$. By (12) and the strict positivity of f , this identity is equivalent to

$$\sum_{|\alpha| \leq M^*} c_\alpha(\theta) Q_\alpha(S(x); \theta) = 0, \quad \theta \in U. \quad (13)$$

Let $m_1, \dots, m_{N_r}$ enumerate the distinct monomials in $S_1, \dots, S_q$, of total degree at most pM^* , that appear across the polynomials $\{Q_\alpha : |\alpha| \leq M^*\}$, so that $N_r \leq N_{\text{pol}}(M^*)$; isolating these monomials as a basis removes any dependence coming from the sufficient-statistic map $S : \mathcal{X} \rightarrow \mathbb{R}^q$. Writing $N_c := N_{\text{der}}(M^*)$ for the number of unknown coefficients, for any fixed θ , define the matrix $\Phi(\theta) \in \mathbb{R}^{N_r \times N_c}$ whose (i, α) entry is the coefficient of the monomial m_i in $Q_\alpha(\cdot; \theta)$. Collecting the coefficients into $\mathbf{c}(\theta) = (c_{\alpha_1}(\theta), \dots, c_{\alpha_{N_c}}(\theta))^\top$, condition (13) is equivalent to

$$\Phi(\theta) \mathbf{c}(\theta) = 0, \quad \theta \in U. \quad (14)$$

Below we will prove that the above system holds on an open subset U of the parameter space. We will use similar arguments as in the proof of Theorem 1.

First, notice that, by the choice of M^* , we have $N_c > N_r$, so (14) is an underdetermined homogeneous system; in particular $\text{rank } \Phi(\theta) \leq N_r$ for every θ . Let $\rho := \max_{\theta \in \Theta^\circ} \text{rank } \Phi(\theta) \leq N_r$ and set $R := \{\theta \in \Theta^\circ : \text{rank } \Phi(\theta) = \rho\}$. If $\rho = 0$ then $\Phi \equiv 0$ on Θ° and (14) is solved by any non-zero constant $\mathbf{c}$; assume therefore $\rho \geq 1$. As R is a non-empty subset of Θ , let θ_0 be any point in R . Then, there exists a $\rho \times \rho$ submatrix $M(\theta_0)$ with $\det M(\theta_0) \neq 0$. After relabeling the monomials (rows) and the multi-indices (columns), we may assume that $M(\theta_0)$ is a principal submatrix of $\Phi(\theta)$. As the entries of Φ are C^∞ , there exists an open set $U \subseteq \Theta^\circ$ containing θ_0 such that $\det M(\theta) \neq 0$ on U , where $M(\theta)$ is the $\rho \times \rho$ principal submatrix of $\Phi(\theta)$.

Let $B(\theta)$ be the $\rho \times (N_c - \rho)$ submatrix of $\Phi(\theta)$ adjacent to $M(\theta)$, and $\mathbf{b}(\theta)$ be its first column. Split $\mathbf{c} = (\mathbf{c}_A, \mathbf{c}_B)$ conformably, with $\mathbf{c}_A = (c_{\alpha_1}, \dots, c_{\alpha_\rho})^\top$ and $\mathbf{c}_B = (c_{\alpha_{\rho+1}}, \dots, c_{\alpha_{N_c}})^\top$; since $N_c - \rho \geq N_c - N_r \geq 1$, $\mathbf{c}_B$ has at least one entry. Fix $\mathbf{c}_B(\theta) := (1, 0, \dots, 0)^\top$ for every $\theta \in U$. The top ρ rows of (14) then read

$$M(\theta)\mathbf{c}_A(\theta) + B(\theta)\mathbf{c}_B(\theta) = 0,$$

or, equivalently,

$$M(\theta)\mathbf{c}_A(\theta) = -\mathbf{b}(\theta).$$

Since $\det M(\theta) \neq 0$ on U , this determines $\mathbf{c}_A(\theta)$ uniquely given the above choice of $\mathbf{c}_B(\theta)$, by Cramer's rule,

$$c_{\alpha_j}(\theta) = \frac{\det M_{\alpha_j}(\theta)}{\det M(\theta)}, \quad 1 \leq j \leq r,$$

where $M_{\alpha_j}(\theta)$ is obtained from $M(\theta)$ by replacing its j -th column with $-\mathbf{b}(\theta)$. This choice annihilates the top ρ rows of $\Phi(\theta)\mathbf{c}(\theta)$ by construction. For the remaining rows, note that on U , since $\text{rank } \Phi(\theta) = \rho$ and $M(\theta)$ is invertible, each of bottom $N_r - \rho$ rows of $\Phi(\theta)$ is the same linear combination of the ρ top rows, using similar argument as in equality (11), the bottom $N_r - \rho$ entries of $\Phi(\theta)\mathbf{c}(\theta)$ are similarly annihilated. Hence $\Phi(\theta)\mathbf{c}(\theta) = 0$ on U , so (13), and with it $\mathcal{L}_\theta f(x | \theta) = 0$, holds for every $x \in \mathcal{X}$ and $\theta \in U$.

Finally, since the entries of Φ are C^∞ on U and $\det M$ is bounded away from zero there, each c_{α_j} is C^∞ on U , while the entries of $\mathbf{c}_B$ are constant; thus all coefficients of $\mathcal{L}_\theta$ lie in $C^\infty(U)$. The coefficient set to 1 has multi-index α_{r+1} , thus the operator $\mathcal{L}_\theta$ is non-trivial on U in the sense of Definition 1. This establishes the proposition.

A.4 Proof of Theorem 2

Step 1: Construction of measure discrepancy. The first step is to construct a non-vanishing everywhere function h such that

$$\int_{\Theta} g(\eta) f(x | \eta) d\eta = 0 \text{ for all } x \in \mathcal{X}.$$

Consider the linear difference-differential operator $\mathcal{L}_\theta$ defined as

$$\mathcal{L}_\theta[f(x | \cdot)] = \sum_{k=1}^K \sum_{|\alpha| \leq r_k} c_{k,\alpha}(\theta) \partial_\theta^\alpha f(x | \theta + v_k).$$

Then, for any function $u \in C_c^\infty(\Theta)$ such that $\text{supp}(u) \subset B$, we have

$$\int_{\Theta} u(\theta) \left(\sum_{k=1}^K \sum_{|\alpha| \leq r_k} c_{k,\alpha}(\theta) \partial_{\theta}^{\alpha} f(x | \theta + v_k) \right) d\theta = 0,$$

and thus as our function $\text{supp}(u) \subset B$, and as the summations are finite and the integrands are continuous on a compact set, we can write the above equation as

$$\sum_{k=1}^K \sum_{|\alpha| \leq r_k} \int_B \left(c_{k,\alpha}(\theta) u(\theta) \right) \partial_{\theta}^{\alpha} f(x | \theta + v_k) d\theta = 0. \quad (15)$$

As $u \in C_c^\infty(\Theta)$ and $\text{supp}(u) \subset B$, we have

$$\int_B c_{k,\alpha}(\theta) u(\theta) [\partial_{\theta}^{\alpha} f(x | \theta + v_k)] d\theta = (-1)^{|\alpha|} \int_B f(x | \theta + v_k) [\partial_{\theta}^{\alpha} (c_{k,\alpha}(\theta) u(\theta))] d\theta.$$

Plugging in this result into (15), we have

$$\sum_{k=1}^K \sum_{|\alpha| \leq r_k} (-1)^{|\alpha|} \int_B f(x | \theta + v_k) \partial_{\theta}^{\alpha} [c_{k,\alpha}(\theta) u(\theta)] d\theta = 0. \quad (16)$$

By applying the change of variables $\eta = \theta + v_k$, we have

$$\int_B f(x | \theta + v_k) \partial_{\theta}^{\alpha} [c_{k,\alpha}(\theta) u(\theta)] d\theta = \int_{B_k} f(x | \eta) \partial_{\eta}^{\alpha} [c_{k,\alpha}(\eta - v_k) u(\eta - v_k)] d\eta$$

where $B_k := B + v_k$. Plugging this equation into equation (16) leads to

$$\sum_{k=1}^K \sum_{|\alpha| \leq r_k} (-1)^{|\alpha|} \int_{B_k} f(x | \eta) \partial_{\eta}^{\alpha} [c_{k,\alpha}(\eta - v_k) u(\eta - v_k)] d\eta = 0. \quad (17)$$

By the properties of the bump function u , the multi-index partial derivatives $\partial_{\eta}^{\alpha} [c_{k,\alpha}(\eta - v_k) u(\eta - v_k)]$ are 0 on the complement $\Theta \setminus B_k$. By understanding $u = 0$ outside Θ , we can rewrite equation (17) as

$$\int_{\Theta} \left(\sum_{k=1}^K \sum_{|\alpha| \leq r_k} (-1)^{|\alpha|} \partial_{\eta}^{\alpha} [c_{k,\alpha}(\eta - v_k) u(\eta - v_k)] \right) f(x | \eta) d\eta = 0.$$

We define a function $h : \Theta \rightarrow \mathbb{R}$ as

$$h(\eta) := \sum_{k=1}^K \sum_{|\alpha| \leq r_k} (-1)^{|\alpha|} \partial_{\eta}^{\alpha} [c_{k,\alpha}(\eta - v_k) u(\eta - v_k)].$$

Next we will use the fact that the sets B_j 's are pairwise disjoint. Indeed, arguing by contradiction, if there exists a $\zeta \in B_i \cap B_j$ for some $i \neq j$, then $\|\zeta - \theta_0 - v_i\| < \rho$ and $\|\zeta - \theta_0 - v_j\| < \rho$, so by

using triangle inequality, $\|v_i - v_j\| < 2\rho$, contradicting $2\rho < \min_{i \neq j} \|v_i - v_j\|$. Now, from the disjointness of the sets B_k for $1 \leq k \leq K$, we have

$$h(\eta)\Big|_{B_k} = h_k(\eta) = \sum_{|\alpha| \leq r_k} (-1)^{|\alpha|} \partial_\eta^\alpha [c_{k,\alpha}(\eta - v_k)u(\eta - v_k)],$$

which indicates that $h(\eta) = \sum_{k=1}^K h_k(\eta)$. For any $1 \leq k \leq K$, we define linear partial differential operator $\mathcal{L}_k$ acting on a test function $u \in C_c^\infty(\Theta)$ as

$$(\mathcal{L}_k u)(\theta) = \sum_{|\alpha| \leq r} c_{k,\alpha}(\theta) \partial_\theta^\alpha u(\theta).$$

Using the similar argument as the proof of Theorem 1 in Section A.1, the adjoint operator $\mathcal{L}_k^*$ takes the following form

$$(\mathcal{L}_k^* u)(\theta) = \sum_{|\alpha| \leq r} (-1)^{|\alpha|} \partial_\theta^\alpha (c_{k,\alpha}(\theta)u(\theta)).$$

Furthermore, $\mathcal{L}_k^* u \in C_c^\infty(\Theta)$ for any $u \in C_c^\infty(\Theta)$. Putting these equations together, we have

$$h_k(\eta) = (\mathcal{L}_k^* u)(\eta - v_k).$$

As the operator $\mathcal{L}_\theta$ is non-trivial, it indicates that the coefficient function $c_{k_*,\alpha}$ is non-zero for some multi-index α and $1 \leq j_* \leq J$. Therefore, the PDE operator $\mathcal{L}_{j_*}$ is non-trivial. Using the argument similar to that of Section A.1, there exists a test function $u \in C_c^\infty(B)$ such that $\mathcal{L}_{j_*}^* u \neq 0$. Since $g(\eta)\Big|_{B_{k_*}} = g_{k_*}(\eta) = (\mathcal{L}_{k_*}^* u)(\eta - v_{k_*})$, it indicates that $g \neq 0$.

Step 2 - Construct probability measure G . By employing the same argument using Fubini's theorem as Step 3, Section A.1, noting that $\int_\Theta h(\eta)f(x | \eta) = 0$, we have

$$\int_\Theta h(\eta) d\eta = 0.$$

Let $\Omega_h := \text{supp}(h) \subset \bigcup_{k=1}^K B_k \subset \Theta$. Recall from the assumption that

$$g_*(\eta) \geq m > 0 \quad \text{for all } \eta \in \Omega_g.$$

Because $h \in C_c^\infty(\Theta)$, we define $M = \|h\|_{L^\infty(\Theta)}$, which is finite. Since $h \neq 0$, it indicates that $M > 0$. Let $\varepsilon := \frac{m}{2M} > 0$, then, it is clear that $\varepsilon \|h\|_{L^\infty(\Theta)} = m/2$. Now, we define

$$g(\theta) := g_*(\theta) + \varepsilon \cdot h(\theta).$$

From here, with the similar argument as that Step 4, Section A.1, we have $g \in L^1(\Theta)$ and non-negative almost everywhere and integrates to 1. Hence, it is a density function of a continuous probability measure $G \in \mathcal{M}(\Theta)$, and satisfies $p_G = p_{G_*}$ for λ -almost everywhere. In addition, as

$$d_{\text{TV}}(G, G_*) = \frac{1}{2} \int_\Theta |g(\theta) - g_*(\theta)| d\theta = \frac{1}{2} \int_\Theta |h(\theta)| d\theta > 0$$

we achieve that $G \neq G_*$, and thus, $W_1(G_c, G_*) \geq c > 0$.

A.5 Proof of Proposition 3

Let $G_1 \neq G_2$ be two measures in $\mathcal{P}(\Theta)$ satisfy $p_{G_1} = p_{G_2}$ for λ -almost everywhere, and set $c := W_1(G_1, G_2) > 0$. Since finitely supported measures are W_1 -dense in $\mathcal{P}(\Theta)$ when Θ is compact, we can construct a pair of discrete sequences, $\{G_1^{(m)}\}_{m=1}^\infty \subset \mathcal{D}(\Theta)$ and $\{G_2^{(m)}\}_{m=1}^\infty \subset \mathcal{D}(\Theta)$, such that $W_1(G_i^{(m)}, G_i) \rightarrow 0$ for $1 \leq i \leq 2$. Thus, for there exists a positive integer M_1 such that for all $m \geq M_1$, we have $W_1(G_i^{(m)}, G_i) < c/4$ for $m \geq M$. An application of the triangle inequality yields

$$W_1(G_1^{(m)}, G_2^{(m)}) \geq W_1(G_1, G_2) - W_1(G_1^{(m)}, G_1) - W_1(G_2^{(m)}, G_2) > c/2.$$

As the mapping $\theta \mapsto f(x | \theta)$ is continuous on the compact set Θ for λ -almost every x , it is also bounded for the x values. Therefore, we appeal to Lemma 2 and infer that $\|p_{G_1^{(m)}} - p_{G_i}\|_1 \rightarrow 0$. Using the fact that $p_{G_1} = p_{G_2}$ almost everywhere, we obtain that

$$\|p_{G_1^{(m)}} - p_{G_2^{(m)}}\|_1 \leq \|p_{G_1^{(m)}} - p_{G_1}\|_1 + \|p_{G_1} - p_{G_2}\|_1 + \|p_{G_2} - p_{G_2^{(m)}}\|_1 \xrightarrow{m \rightarrow \infty} 0.$$

Thus, there exists a positive integer M_2 such that $\|p_{G_1^{(m)}} - p_{G_2^{(m)}}\|_1 \leq 1/(2n)$ for all $m \geq M_2$. By sub-additivity of the total variation distance over product measures, we have

$$\|P_{G_1^{(m)}}^n - P_{G_2^{(m)}}^n\|_1 \leq n\|p_{G_1^{(m)}} - p_{G_2^{(m)}}\|_1 \leq 1/2.$$

Let $M = \max\{M_1, M_2\}$, Le Cam's two-point inequality (e.g., [Wainwright, 2019](#), Lemma 15.9) then gives, for any estimator $\hat{G}_n : \mathcal{X}^n \rightarrow \mathcal{D}(\Theta)$ and $m \geq M$,

$$\sup_{G_* \in \mathcal{D}(\Theta)} \mathbb{E}_{X_{1:n} \sim P_{G_*}^n} [W_1(\hat{G}_n, G_*)] \geq \frac{1}{2} \frac{W_1(G_1^{(m)}, G_2^{(m)})}{2} \left(1 - \frac{1}{2} \|P_{G_1^{(m)}}^n - P_{G_2^{(m)}}^n\|_1\right) \geq \frac{3c}{32}.$$

As the bound is uniform in n , we arrive at

$$\inf_{n \in \mathbb{N}} \inf_{\hat{G}_n} \sup_{G_* \in \mathcal{D}(\Theta)} \mathbb{E}_{X_{1:n} \sim P_{G_*}^n} [W_1(\hat{G}_n, G_*)] \geq \frac{3c}{32}.$$

As a consequence, we reach the conclusion of the proposition.

A.6 Proof of Proposition 4

Let $G \in \mathcal{M}(\Theta)$ satisfy $\mathcal{A}G = 0$ for λ -almost everywhere. Since $h > 0$, we have that

$$\int_{\Theta} \gamma(\theta) K(\tau(x) - \psi(\theta)) G(d\theta) = 0 \text{ for almost all } x.$$

Since the function γ is assumed to be continuous on the compact set Θ , it is also bounded. Thus, $\tilde{G}(d\theta) := \gamma(\theta) G(d\theta)$ defines a finite signed measure. Let $\mu := \tilde{G} \circ \psi^{-1}$ the push-forward measure of $\mu := \tilde{G}$ by ψ , by construction supported on the compact set $\psi(\Theta) \subset \mathbb{R}^{d_1}$. Because τ maps $\mathcal{X}$ onto $\mathbb{R}^{d_1}$, the above identity can be rewritten as

$$\int_{\psi(\Theta)} K(y - s) \mu(ds) = 0, \text{ for almost all } y \in \mathbb{R}^{d_1}.$$

Taking Fourier transforms and using the properties of convolutions, we obtain

$$\mathcal{F}[K](\xi) \mathcal{F}[\mu](\xi) = 0 \text{ for all } \xi \in \mathbb{R}^{d_1}$$

Because $\mathcal{F}[\mu]$ is continuous and $\mathcal{F}[K] \neq 0$ almost everywhere by hypothesis, we deduce that $\mathcal{F}[\mu] \equiv 0$ and hence $\mu = 0$. Since ψ is a homeomorphism onto its image, this implies $\tilde{G} = 0$, and since $\gamma > 0$, we conclude that $G = 0$. This proves injectivity of $\mathcal{A}$, as claimed.

A.7 Proof of Proposition 5

Let $G \in \mathcal{M}(\Theta)$ satisfy $\mathcal{A}G = 0$ for λ -almost everywhere. As $h > 0$,

$$\int_{\Theta} \exp(\langle \eta(\theta), T(x) \rangle) e^{-A(\theta)} G(d\theta) = 0, \quad \lambda\text{-almost every } x.$$

Because Θ is compact and A continuous, e^{-A} is bounded away from 0 and ∞ , so $\tilde{G}(d\theta) := e^{-A(\theta)} G(d\theta)$ is a finite signed measure with $\tilde{G}(S) = 0 \Leftrightarrow G(S) = 0$ for any measurable set $S \subseteq \Theta$. The map $\Psi := \exp \circ \eta : \Theta \rightarrow (0, \infty)^q$, applied coordinate-wise, is continuous and injective on the compact set Θ , hence a homeomorphism onto its image $\Omega := \Psi(\Theta) = \exp\{\eta(\Theta)\}$, which is therefore compact in $(0, \infty)^q$. Let $\mu := \tilde{G} \circ \Psi^{-1}$, so that μ is a finite signed measure on Ω , and since Ψ is a bijection between Θ and Ω , $\mu(S) = 0 \Leftrightarrow \tilde{G}(\Psi^{-1}(S)) = 0 \Leftrightarrow G(\Psi^{-1}(S)) = 0$ for any $S \subseteq \Omega$ measurable. That is $\mu = 0$ if and only if $\tilde{G} = 0$.

Fix $x \in \mathcal{X}$ and set $s := T(x) \in \mathbb{R}^q$. For $\theta \in \Theta$, writing $y = \Psi(\theta) = (e^{\eta_1(\theta)}, \dots, e^{\eta_q(\theta)})$, we have

$$\exp(\langle \eta(\theta), T(x) \rangle) = \exp\left(\sum_{i=1}^q s_i \eta_i(\theta)\right) = \prod_{i=1}^q (e^{\eta_i(\theta)})^{s_i} = \prod_{i=1}^q y_i^{s_i} = y^s.$$

Therefore, changing variables in the last display equation, we obtain

$$\int_{\Omega} y^s \mu(dy) = 0, \quad s \in T(\mathcal{X}). \quad (18)$$

Notice that this holds for every $s \in T(\mathcal{X})$ because the original identity holds for λ -almost every x and T is measurable. Note that each monomial $y \mapsto y^s$ is continuous and bounded on the compact set $\Omega \subset (0, \infty)^q$, so the integrals in (18) are well defined and finite.

Since $T(\mathcal{X})$ is determining for Ω , the linear span of $\{y \mapsto y^s : s \in T(\mathcal{X})\}$ is dense in $C(\Omega)$ under the supremum norm. Given $w \in C(\Omega)$ and $\varepsilon > 0$, choose a finite linear combination $w_\varepsilon = \sum_k a_k y^{s_k}$ with $s_k \in T(\mathcal{X})$ and $\|w - w_\varepsilon\|_\infty < \varepsilon$. By (18), $\int_{\Omega} w_\varepsilon d\mu = \sum_k a_k \int_{\Omega} y^{s_k} d\mu = 0$, so

$$\left| \int_{\Omega} w d\mu \right| = \left| \int_{\Omega} (w - w_\varepsilon) d\mu \right| \leq \|w - w_\varepsilon\|_\infty \|\mu\|_{\text{TV}} < \varepsilon \|\mu\|_{\text{TV}}.$$

Letting $\varepsilon \rightarrow 0$ gives $\int_{\Omega} w d\mu = 0$ for every $w \in C(\Omega)$. By the Riesz representation theorem, the only finite signed Radon measure on the compact set Ω that integrates every continuous function to zero is the zero measure, so $\mu = 0$, and therefore $G = 0$, proving the injectivity of $\mathcal{A}$.

A.8 Proof of Proposition 6

Recall the kernel $f(x | \theta) = h(x) \exp\{\theta_1 x + \theta_2 e^{x^2} - A(\theta)\}$ with base measure $h(x) = \exp(-2be^{x^2})$ and parameter space $\Theta = [-a, a] \times [-b, b]$.

(a) *Non-existence of a non-trivial parameter PDE.* Suppose, for contradiction, that some non-trivial parameter differential operator $\mathcal{L}_\theta = \sum_{|\alpha| \leq r} c_\alpha(\theta) \partial_\theta^\alpha$ with continuous coefficients on Θ° annihilates f ; that is, there is a multi-index α^* with $|\alpha^*| = r$ and $c_{\alpha^*} \neq 0$. Write

$$f(x | \theta) = h(x) e^{-A(\theta)} E(x | \theta), \quad E(x | \theta) := \exp(\theta_1 x + \theta_2 e^{x^2}).$$

Since $h(x) > 0$ is independent of θ , the identity $\mathcal{L}_\theta f = 0$ is equivalent to $\mathcal{L}_\theta [e^{-A(\theta)} E(x | \theta)] = 0$. Expanding the derivatives of the product $e^{-A(\theta)} E(x | \theta)$ by the Leibniz rule yields an operator $\tilde{\mathcal{L}}_\theta$ acting on E , i.e.

$$\tilde{\mathcal{L}}_\theta E(x | \theta) := e^{A(\theta)} \mathcal{L}_\theta [e^{-A(\theta)} E(x | \theta)] = \sum_{|\beta| \leq r} \tilde{c}_\beta(\theta) \partial_\theta^\beta E(x | \theta) = 0. \quad (19)$$

Note that the coefficients $\tilde{c}_\beta(\theta)$ are linear combinations of products of $c_\alpha(\theta)$ and higher order partial derivatives of $A(\theta)$. For the top-order terms $|\beta| = r$, the Leibniz rule gives $\tilde{c}_\beta(\theta) = c_\beta(\theta)$, so in particular $\tilde{c}_{\alpha^*}(\theta) = c_{\alpha^*}(\theta) \neq 0$; moreover $\mathcal{L}_\theta \equiv 0$ if and only if $\tilde{\mathcal{L}}_\theta \equiv 0$, so $\tilde{\mathcal{L}}_\theta$ is non-trivial whenever $\mathcal{L}_\theta$ is.

Since $E(x | \theta) = \exp(\theta_1 x + \theta_2 e^{x^2})$, differentiation gives

$$\partial_\theta^\beta E(x | \theta) = \frac{\partial^{|\beta|}}{\partial \theta_1^{\beta_1} \partial \theta_2^{\beta_2}} \exp(\theta_1 x + \theta_2 e^{x^2}) = x^{\beta_1} (e^{x^2})^{\beta_2} E(x | \theta).$$

Substituting into (19) and using $E(x | \theta) > 0$ for all $x \in \mathbb{R}$,

$$\sum_{|\beta| \leq r} \tilde{c}_\beta(\theta) x^{\beta_1} e^{\beta_2 x^2} = 0, \quad x \in \mathbb{R}. \quad (20)$$

We now show that the above relationship forces $\tilde{c}_\beta(\theta) = 0$ for every $|\beta| \leq r$, contradicting non-triviality. Indeed, for a non-negative integer $k \leq r$, grouping the terms in (20) by the value k of the second index β_2 yields the polynomial in x $P_k(x, \theta) := \sum_{\beta_1} \tilde{c}_{(\beta_1, k)}(\theta) x^{\beta_1}$. Thus, we can re-rewrite (20) as

$$\sum_{k=0}^M P_k(x, \theta) e^{kx^2} = 0, \quad x \in \mathbb{R}, \quad (21)$$

where $M \leq r$ is the largest index with $P_M(\cdot, \theta) \neq 0$ (such an M exists precisely because $\tilde{\mathcal{L}}_\theta$ is non-trivial). We will study the asymptotic behavior of the last identity as $x \rightarrow \infty$. Dividing both sides of (21) by e^{Mx^2} , we obtain

$$P_M(x, \theta) + \sum_{k=0}^{M-1} P_k(x, \theta) e^{(k-M)x^2} = 0. \quad (22)$$

For each $k < M$ one has $k - M < 0$, so $P_k(x, \theta) e^{(k-M)x^2} \rightarrow 0$ as $x \rightarrow \infty$; letting $x \rightarrow \infty$ in (22) therefore gives $\lim_{x \rightarrow \infty} P_M(x, \theta) = 0$. But a polynomial in x either is identically zero or diverges to

$\pm\infty$ or tends to a non-zero constant, so this limit is impossible unless $P_M(\cdot, \theta) \equiv 0$, contradicting the choice of M . Hence $P_k(\cdot, \theta) \equiv 0$ for every k , so $\tilde{c}_\beta(\theta) \equiv 0$ for all $|\beta| \leq r$. Since $\tilde{c}_\beta = c_\beta$ at the top order, the highest-order coefficients of $\mathcal{L}_\theta$ vanish. By descending induction on $|\alpha|$, we calculate that $c_\alpha \equiv 0$ for all $|\alpha| \leq r$. Equivalently, $\mathcal{L}_\theta$ is trivial, a contradiction. Thus f satisfies no non-trivial homogeneous parameter PDE; the first claim of the Proposition is proved.

(b) *Identifiability.* Suppose there exists a $\mu \in \mathcal{M}(\Theta)$ satisfying

$$\int_{\Theta} f(x \mid \theta) \mu(d\theta) = \int_{\Theta} h(x) \exp(\theta_1 x + \theta_2 e^{x^2} - A(\theta)) \mu(d\theta) = 0, \quad x \in \mathbb{R}. \quad (23)$$

We will show that this relationship implies that $\mu = 0$ (thus proving identifiability) in four steps.

Step 1: Reduction and holomorphic extension. As $h(x) = \exp(-2be^{x^2}) > 0$, (23) is equivalent to

$$\int_{\Theta} \exp(\theta_1 x + \theta_2 e^{x^2} - A(\theta)) \mu(d\theta) = 0, \text{ for all } x \in \mathbb{R}.$$

Define the signed measure $\tilde{\mu}$ by $\tilde{\mu}(d\theta) := e^{-A(\theta)} \mu(d\theta)$; since Θ is compact and A is continuous, e^{-A} is bounded away from 0 and ∞ , so $\tilde{\mu}$ is finite and $\tilde{\mu} = 0$ if and only if $\mu = 0$. Thus, identity (23) becomes

$$\int_{\Theta} \exp(\theta_1 x + \theta_2 e^{x^2}) \tilde{\mu}(d\theta) = 0, \quad x \in \mathbb{R}. \quad (24)$$

Let $F : \mathbb{C}^2 \rightarrow \mathbb{C}$ be the Laplace transform of $\tilde{\mu}$, i.e.

$$(w_1, w_2) \in \mathbb{C}^2 \mapsto F(w_1, w_2) := \int_{\Theta} \exp(w_1 \theta_1 + w_2 \theta_2) \tilde{\mu}(d\theta).$$

Because $\tilde{\mu}$ is finite with support in the compact box $[-a, a] \times [-b, b]$, the Paley–Wiener theorem implies that F is entire on $\mathbb{C}^2$ and satisfies

$$|F(w_1, w_2)| \leq \|\tilde{\mu}\|_{\text{TV}} \exp(a|\operatorname{Re} w_1| + b|\operatorname{Re} w_2|),$$

so F is of exponential type (see Definition 7 below).

Step 2: The diagonal slice and its Taylor expansion. Consider the entire function $M(z) := F(z, e^{z^2})$, a composition of entire functions. By (24), $M(x) = F(x, e^{x^2}) = 0$ for every $x \in \mathbb{R}$. The real line is a subset of $\mathbb{C}$ with an accumulation point lying in $\mathbb{C}$, so the identity theorem gives $M(z) = F(z, e^{z^2}) = 0$ for all $z \in \mathbb{C}$. Since F is entire on $\mathbb{C}^2$, it has a globally convergent Taylor expansion in its second argument,

$$F(w_1, w_2) = \sum_{n=0}^{\infty} c_n(w_1) w_2^n, \quad c_n(w_1) = \frac{1}{n!} \int_{\Theta} \theta_2^n \exp(w_1 \theta_1) \tilde{\mu}(d\theta).$$

Step 3: Fast decay of the leading coefficient. By the Paley–Wiener theorem each c_n is entire of exponential type. Evaluating the identity $M(z) = 0$ through the expansion gives $\sum_{n \geq 0} c_n(z) e^{nz^2} = 0$ for all $z \in \mathbb{C}$, and substituting $z = iy$ with $y \in \mathbb{R}$ yields that

$$\sum_{n=0}^{\infty} c_n(iy) e^{-ny^2} = 0 \quad \text{or, equivalently,} \quad c_0(iy) = - \sum_{n=1}^{\infty} c_n(iy) e^{-ny^2}. \quad (25)$$

From the integral formula for c_n , and using $|\tilde{\mu}| \leq \|\tilde{\mu}\|_{\text{TV}}$ and $|\theta_2| \leq b$,

$$|c_n(iy)| \leq \frac{1}{n!} \int_{\Theta} |\theta_2|^n |e^{iy\theta_1}| |\tilde{\mu}|(d\theta) \leq \frac{b^n}{n!} \|\tilde{\mu}\|_{\text{TV}}.$$

Using these bounds and (25),

$$|c_0(iy)| \leq \sum_{n=1}^{\infty} |c_n(iy)| e^{-ny^2} \leq \|\tilde{\mu}\|_{\text{TV}} \sum_{n=1}^{\infty} \frac{b^n}{n!} e^{-ny^2} = \|\tilde{\mu}\|_{\text{TV}} e^{-y^2} \sum_{n=1}^{\infty} \frac{b^n}{n!} e^{-(n-1)y^2}.$$

Noting that $e^{-(n-1)y^2} \leq 1$, we have

$$\sum_{n=1}^{\infty} \frac{b^n}{n!} e^{-(n-1)y^2} \leq \sum_{n=1}^{\infty} \frac{b^n}{n!} = e^b - 1.$$

As a result, with $C := \|\tilde{\mu}\|_{\text{TV}}(e^b - 1)$ we obtain

$$|c_0(iy)| \leq C e^{-y^2},$$

for all $y \in \mathbb{R}$.

Step 4: Logarithmic bound forces $c_0 \equiv 0$, then induction. Set $g(z) := c_0(iz)$, an entire function of exponential type with $|g(y)| \leq C e^{-y^2}$ on $\mathbb{R}$. Then, $\log |g(y)| \leq \log C - y^2$, leading to the following estimation for the logarithmic integral

$$\int_{-\infty}^{\infty} \frac{\log |g(y)|}{1+y^2} dy \leq \int_{-\infty}^{\infty} \frac{\log C - y^2}{1+y^2} dy = \log C \int_{-\infty}^{\infty} \frac{dy}{1+y^2} - \int_{-\infty}^{\infty} \frac{y^2}{1+y^2} dy = -\infty,$$

the last integral diverging because $y^2/(1+y^2) \rightarrow 1$. By Lemma 1, an entire function of exponential type whose logarithmic integral diverges to $-\infty$ must be identically equal to 0, hence $c_0 \equiv g \equiv 0$.

We now argue by induction. Suppose $c_k \equiv 0$ for all $0 \leq k < m$. Then (25) reduces to $\sum_{n \geq m} c_n(iy) e^{-ny^2} = 0$, so

$$c_m(iy) e^{-my^2} = - \sum_{n=m+1}^{\infty} c_n(iy) e^{-ny^2},$$

and the same estimate as for c_0 gives

$$|c_m(iy)| \leq \|\tilde{\mu}\|_{\text{TV}} e^{-y^2} \sum_{k=1}^{\infty} \frac{b^{m+k}}{(m+k)!} \leq C_m e^{-y^2}$$

for a constant C_m . As above, we have

$$\int_{-\infty}^{\infty} \log |c_m(iy)| / (1+y^2) dy = -\infty,$$

so Lemma 1 forces $c_m \equiv 0$. By induction $c_n \equiv 0$ for every $n \geq 0$.

Step 5: Conclusion. We have proved that, for all $n \geq 0$ and $y \in \mathbb{R}$,

$$c_n(iy) = \frac{1}{n!} \int_{\Theta} \theta_2^n \exp(iy\theta_1) \tilde{\mu}(d\theta) = 0.$$

Fix n and define the finite signed measure ν_n on $[-a, a]$ by $\nu_n(E) := \int_{E \times [-b, b]} \theta_2^n \tilde{\mu}(d\theta)$ for any measurable $E \subseteq [-a, a]$. The display states that $\mathcal{F}[\nu_n](y) = 0$ for all y , so by uniqueness of the Fourier transform, $\nu_n = 0$. Consequently, for every continuous ϕ on $[-a, a]$, $\int_{\Theta} \phi(\theta_1) \theta_2^n \tilde{\mu}(d\theta) = 0$, and therefore $\int_{\Theta} \theta_1^k \theta_2^n \tilde{\mu}(d\theta) = 0$ for all $k, n \geq 0$. Hence, $\int_{\Theta} P(\theta) \tilde{\mu}(d\theta) = 0$ for every polynomial P . Since polynomials are dense in $C(\Theta)$ by the Stone–Weierstrass theorem, $\int_{\Theta} w \tilde{\mu}(d\theta) = 0$ for all $w \in C(\Theta)$, which forces $\tilde{\mu} = 0$ and thus $\mu = 0$.

B Auxiliary results and proofs

Below, we collect some auxiliary results used in the proofs.

Definition 7. *An entire function $g : \mathbb{C} \rightarrow \mathbb{C}$ is of exponential type if there exist $C > 0$ and $\tau \geq 0$ with $|g(z)| \leq Ce^{\tau|z|}$ for all $z \in \mathbb{C}$.*

The following logarithmic-integral bound is the analytic device behind Proposition 6(b): an entire function of exponential type that is bounded on the real axis cannot have its log-modulus decay at a quadratic rate there unless it vanishes identically. This result is commonly referred to as Cartwright’s version of Levinson’s theorem

Lemma 1 (after [Koosis, 1988](#), Section III.G). *If g is entire of exponential type, bounded on $\mathbb{R}$, and $g \not\equiv 0$, then*

$$\int_{-\infty}^{\infty} \frac{\log |g(y)|}{1+y^2} dy > -\infty.$$

Proof. Step 1: Reduction to an even function with a controlled germ at 0. By Definition 7 there are $C \geq 1$, $\tau \geq 0$ with $|g(z)| \leq Ce^{\tau|z|}$, and $M \geq 1$ with $|g(x)| \leq M$ on $\mathbb{R}$. Suppose that g has a zero of order k at the origin, we can write $g(z) = cz^k \eta(z)$ with η entire, $\eta(0) = 1$, and $c \neq 0$. For $|z| \geq 1$, we have

$$|\eta(z)| \leq \frac{|g(z)|}{|c||z|^k} \leq \frac{C}{|c|} e^{\tau|z|}.$$

As η is continuous, it is bounded on $|z| \leq 1$. Thus, η retains the exponential type τ . Similarly, $|\eta(x)| \leq \frac{M}{|c|}$ for $|x| \geq 1$, and $\eta(x)$ is bounded on $[-1, 1]$. Thus, $\eta(x)$ is bounded on $\mathbb{R}$.

Now we construct $G(z) := \eta(z)\eta(-z)$, which is even, entire of exponential type 2τ , bounded on $\mathbb{R}$ by some B^2 , with $G(0) = 1$, $G'(0) = 0$, and $\log |G(z)| = O(|z|^2)$ near 0. Then, as

$$\log |g(z)| + \log |g(-z)| = 2\log(c) + 2k \log |z| + \log |G(z)|,$$

we achieve the following equality

$$\int_{-\infty}^{\infty} \frac{\log |g(y)|}{1+y^2} dy = \int_{-\infty}^{\infty} \frac{\log |c| + k \log |y|}{1+y^2} dy + \frac{1}{2} \int_{-\infty}^{\infty} \frac{\log |G(y)|}{1+y^2} dy.$$

Noting that function $\frac{\log |c| + k \log |y|}{1+y^2}$ is integrable in entire $\mathbb{R}$ for $c \neq 0$, our problem reduces to

$$\int_{-\infty}^{\infty} \frac{\log |G(y)|}{1+y^2} dy > -\infty \tag{26}$$

Step 2: Local weighted logarithmic integrability. We prove that for any positive R , we have

$$\int_{-R}^R |\log |G(x)|| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx < \infty \quad (27)$$

Indeed, as G is a non-zero entire function in $\mathbb{C}$, there are at most a finite number of zero of G at $[-R, R]$, let us say $\{x_j\}_{1 \leq j \leq k}$. Let $\delta > 0$ be a sufficient small real number such that the intervals $(x_j - \delta, x_j + \delta)$ and $(-\delta, \delta)$ are mutually disjoint. Then, the function $|G(x)|$ are positive in the compact set $\Omega := [-R, R] \setminus \bigcup_{j=1}^k (x_j - \delta, x_j + \delta)$, thus there exists a constant $C > 0$ such that $|\log |G(x)|| \leq C$, thus

$$\int_{\Omega} |\log |G(x)|| \left(\frac{1}{x^2} + \frac{1}{R^2} \right) dx \leq C \int_{-R}^R \left(\frac{1}{\delta^2} + \frac{1}{R^2} \right) dx = \frac{2RC}{\delta^2} + \frac{2C}{R}. \quad (28)$$

Now we prove the integrability of $|\log |G(x)|| \left(\frac{1}{x^2} - \frac{1}{R^2} \right)$ in the intervals $(-\delta, \delta)$ and each $(x_j - \delta, x_j + \delta)$. For interval $(-\delta, \delta)$, as $\log |G(x)| = O(|x|^2)$ and function $x^2 \left(\frac{1}{x^2} + \frac{1}{R^2} \right)$ is integrable in $(-\delta, \delta)$, we have

$$\int_{-\delta}^{\delta} |\log |G(x)|| \left(\frac{1}{x^2} + \frac{1}{R^2} \right) dx < \infty. \quad (29)$$

For each interval $(x_j - \delta, x_j + \delta)$, suppose that x_j a zero of G with finite multiplicity $m_j \geq 1$. We can express $G(x) = (x - x_j)^{m_j} \kappa(x)$, where $\kappa \neq 0$ in $(x_j - \delta, x_j + \delta)$. Thus, we have

$$\left| \frac{\log |G(x)|}{x^2} \right| \leq \frac{m_j |\log |x - x_j||}{x^2} + \frac{|\log |\kappa(x)||}{x^2}.$$

Since $(x_j - \delta, x_j + \delta)$ and $(-\delta, \delta)$ are disjoint, we have $|x| > \delta$ for each $x \in (x_j - \delta, x_j + \delta)$. The term $|\log |\eta(x)||$ is bounded because $\kappa(x_j) \neq 0$. For the logarithm term, we have

$$\int_{x_j - \delta}^{x_j + \delta} |\log |x - x_j|| dx = \int_{-\delta}^{\delta} |\log |t|| dt = 2\delta(1 - \log \delta) < \infty.$$

Thus, we achieve the estimation

$$\int_{x_j - \delta}^{x_j + \delta} |\log |G(x)|| \left(\frac{1}{x^2} + \frac{1}{R^2} \right) dx < \infty. \quad (30)$$

By combining the estimation from (28), (29) and (30), we achieve the estimation

$$\int_{-R}^R |\log |G(x)|| \left(\frac{1}{x^2} + \frac{1}{R^2} \right) dx < \infty.$$

In particular, this estimation implies (27). When $R = 1$, we similarly achieve that

$$\int_{-1}^1 \frac{|\log |G(x)||}{x^2} dx < \infty. \quad (31)$$

Step 3: Carleman's formula. In the rest of this section, we denote $u(z) = \log |G(z)|$ and

$$v(z) = \operatorname{Im} \left(-\frac{1}{z} - \frac{z}{R} \right) = \sin \theta \left(\frac{1}{r} - \frac{r}{R^2} \right) \text{ for } z = re^{i\theta}, r \geq 0, \theta \in [0, 2\pi).$$

In this step we give a demonstration for the result known as Carleman formula, which is an important part of our proof (for a systematic reference, see [Levin \(1980\)](#))

$$\int_{-R}^R \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx + \frac{2}{R} \int_0^\pi \log |G(Re^{i\theta})| \sin \theta d\theta = 2\pi \sum_{z_k \in \mathcal{Z}_+} m_k v(z_k), \quad (32)$$

where $\mathcal{Z}_+ := \{z_k \in \mathbb{C} : \text{Im}(z_k) > 0, |z_k| < R, G(z_k) = 0\}$. Let D_R^+ be the open upper half-disk $\{z = re^{i\theta} \in \mathbb{C} : r < R, \theta \in (0, \pi)\}$. Let $D_{R,\epsilon}^+$ be D_R^+ minus small disks of radius ϵ around each zero z_k of G , and minus small half-disks of radius ϵ around the origin and any real roots x_k , namely,

$$D_{R,\epsilon}^+ = \{z \in \mathbb{C} : |z| < R, \text{Im}(z) > 0\} \setminus \left(\overline{B_\epsilon(0)} \cup \bigcup_{x_k \in \mathcal{Z}_\mathbb{R}} \overline{B_\epsilon(x_k)} \cup \bigcup_{z_k \in \mathcal{Z}_+} \overline{B_\epsilon(z_k)} \right)$$

where $\mathcal{Z}_\mathbb{R} = \{x_k \in \mathbb{R} : |x_k| < R, G(x_k) = 0\}$ be the set of real zeros of G with absolute value less than R . As η is a non-zero function, G is a non-zero holomorphic function too, which implies that the number of zero point of G in D_R^+ is finite, in other words, $|\mathcal{Z}_\mathbb{R}|, |\mathcal{Z}_+| < \infty$. Here, we denote $\overline{B_\epsilon(x)} = \{z \in \mathbb{C} : |z - x| \leq \epsilon\}$ for any $x \in \mathbb{C}$ and $\epsilon > 0$.

Step 3.1: Green identity. By construction, all points where $G(z) = 0$, namely, the zeros $\mathcal{Z}_\mathbb{R}$ and $\mathcal{Z}_+$ have been removed along with an open neighborhood of radius ϵ surrounding them. Therefore, $G(z) \neq 0$ for all $z \in D_{R,\epsilon}^+$. Since $G(z)$ is a holomorphic function, $u(z) = \log |G(z)|$ is a real harmonic function on open set $D_{R,\epsilon}^+$. Thus, $u \in C^2(D_{R,\epsilon}^+)$ and whose Laplacian exactly 0 in $D_{R,\epsilon}^+$

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

Regarding the function v , since $-\frac{1}{z} - \frac{z}{R^2}$ is a complex analytic function everywhere except at its lone pole at the origin $z = 0$ and we remove the origin and its surrounding half-disk $\overline{B_\epsilon(0)}$, the point $z = 0$ is not in $D_{R,\epsilon}^+$. It indicates that $v \in C^2(\overline{D_{R,\epsilon}^+})$ and satisfies the Laplace's equation

$$\Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

Using Green's second identity, noting that $\Delta u = \Delta v = 0$ in $D_{R,\epsilon}^+$, we have

$$0 = \int_{D_{R,\epsilon}^+} (u\Delta v - v\Delta u) dA = - \oint_{\partial D_{R,\epsilon}^+} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds.$$

Since $\Delta u = 0$ and $\Delta v = 0$ inside $D_{R,\epsilon}^+$, this equation becomes

$$\oint_{\partial D_{R,\epsilon}^+} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = 0.$$

Let $\Xi(s) := u(s) \frac{\partial v}{\partial n}(s) - v(s) \frac{\partial u}{\partial n}(s)$, we now evaluate the integral of Ξ over the $\partial D_{R,\epsilon}^+$ boundary as $\epsilon \rightarrow 0$. From the construction of $D_{R,\epsilon}^+$, we have for ϵ sufficiently small,

$$\partial D_{R,\epsilon}^+ = \Gamma_{\text{Real}} \cup S_R \cup S_\epsilon \cup \left(\bigcup_{x_k \in \mathcal{Z}_\mathbb{R}} \gamma_{x_k} \right) \cup \left(\bigcup_{z_k \in \mathcal{Z}_+} \gamma_{z_k} \right) \quad (33)$$

where Γ_{Real} consists of the remaining flat line segments left behind on $[-R, R]$ after slicing out small intervals of width 2ϵ around the origin and any real roots x_k ; γ_{x_k} are the tiny clockwise semicircles of radius ϵ looping over the real roots x_k into the upper half-plane to keep them outside our domain; S_R is the counter-clockwise circular arc running from $(R, 0)$ to $(-R, 0)$ in the upper half-plane; S_ϵ is a small clockwise semicircle of radius ϵ centered at $z = 0$; γ_{z_k} is a boundary of the closed disk of radius ϵ around z_k .

Step 3.2: Integral of Ξ over Γ_{Real} . We first evaluate $\lim_{\epsilon \rightarrow 0^+} \int_{\Gamma_{\text{Real}}} \Xi(s) ds$. From the definition of Γ_{Real} , direct calculation leads to

$$\int_{\Gamma_{\text{Real}}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = \int_{-R}^R \phi(x) \chi_{I_{R,\epsilon}}(x) dx$$

where $\phi(x) = \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right)$ and $\chi_{I_{R,\epsilon}}(x)$ is the characteristic indicator function of the set $I_{R,\epsilon}$, defined as:

$$I_{R,\epsilon} = [-R, R] \setminus \left((-\epsilon, \epsilon) \cup \bigcup_{x_k \in \mathcal{Z}_{\mathbb{R}}} (x_k - \epsilon, x_k + \epsilon) \right).$$

It is clear that $\lim_{\epsilon \rightarrow 0^+} \chi_{I_{R,\epsilon}}(x) = 1$ for almost all $x \in [-R, R]$. Therefore, we have

$$\lim_{\epsilon \rightarrow 0^+} \phi(x) \chi_{I_{R,\epsilon}}(x) = \phi(x)$$

for almost surely $x \in [-R, R]$. Using the fact that $|\phi(x)|$ is integrable in $[-R, R]$ in Step 2 (see (27)), and putting these results together, an application of the dominated convergence theorem leads to

$$\begin{aligned} \lim_{\epsilon \rightarrow 0^+} \int_{\Gamma_{\text{Real}}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds &= \lim_{\epsilon \rightarrow 0^+} \int_{-R}^R \phi(x) \chi_{I_{R,\epsilon}}(x) dx = \int_{-R}^R \left(\lim_{\epsilon \rightarrow 0^+} \phi(x) \chi_{I_{R,\epsilon}}(x) \right) dx \\ &= \int_{-R}^R \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx. \end{aligned} \quad (34)$$

Step 3.3: Integral of Ξ over S_R . We now move to evaluate $\lim_{\epsilon \rightarrow 0^+} \int_{S_R} \Xi(s) ds$. Along the semicircle S_R , we have $v = 0$. Therefore,

$$\lim_{\epsilon \rightarrow 0^+} \int_{S_R} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = \lim_{\epsilon \rightarrow 0^+} \int_{S_R} u \frac{\partial v}{\partial n} ds.$$

The semicircle S_R is a 1-dimensional manifold parameterized smoothly by the map

$$\gamma_R : [0, \pi] \rightarrow \mathbb{C}, \quad \gamma_R(\theta) = Re^{i\theta} = (R \cos \theta, R \sin \theta).$$

To convert the line integral $\int_{S_R} u \frac{\partial v}{\partial n} ds$ into a standard Riemann integral over $[0, \pi]$, we determine the differential arc length element ds . The tangent vector is $\gamma'_R(\theta) = (-R \sin \theta, R \cos \theta)$. Therefore, we have $ds = \|\gamma'_R(\theta)\| d\theta = \sqrt{(-R \sin \theta)^2 + (R \cos \theta)^2} d\theta = R d\theta$. Direct computation yields that

$$\int_{S_R} u(z) \frac{\partial v}{\partial n}(z) ds = \int_0^\pi u(\gamma_R(\theta)) \cdot \frac{\partial v}{\partial n}(\gamma_R(\theta)) \cdot \|\gamma'_R(\theta)\| d\theta.$$

Plugging in $u(Re^{i\theta}) = \log |G(Re^{i\theta})|$, $\frac{\partial v}{\partial n}(\gamma_R(\theta)) = \frac{2}{R^2} \sin \theta$, and $ds = R d\theta$, we obtain that

$$\int_{S_R} u \frac{\partial v}{\partial n} ds = \int_0^\pi \log |G(Re^{i\theta})| \left(\frac{2}{R^2} \sin \theta \right) R d\theta = \frac{2}{R} \int_0^\pi \log |G(Re^{i\theta})| \sin \theta d\theta.$$

Putting these results together, we arrive at

$$\lim_{\epsilon \rightarrow 0^+} \int_{S_R} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = \frac{2}{R} \int_0^\pi \log |G(Re^{i\theta})| \sin \theta d\theta. \quad (35)$$

Step 3.4: Integral of Ξ over S_ϵ . We now move to evaluate $\lim_{\epsilon \rightarrow 0^+} \int_{S_\epsilon} \Xi(s) ds$. The semicircle S_ϵ can be parameterized by the map

$$\gamma_0 : [0, \pi] \rightarrow \mathbb{C}, \quad \gamma_0(\theta) = \epsilon e^{i(\pi-\theta)} = -\epsilon \cos \theta + i\epsilon \sin \theta.$$

The tangent vector of this path is $\gamma_0'(\theta) = \epsilon \sin \theta + i\epsilon \cos \theta$, which leads to $ds = \|\gamma_0'(\theta)\| d\theta = \sqrt{(\epsilon \sin \theta)^2 + (\epsilon \cos \theta)^2} d\theta = \epsilon d\theta$. Direct calculation yields that

$$\left| \int_{S_\epsilon} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds \right| = \left| \int_0^\pi \left(u(\gamma_0(\theta)) \frac{\partial v}{\partial n}(\gamma_0(\theta)) - v(\gamma_0(\theta)) \frac{\partial u}{\partial n}(\gamma_0(\theta)) \right) \epsilon d\theta \right|.$$

By construction, $G(z)$ is an entire function with $G(0) = 1$ and $G'(0) = 0$. Its Taylor series expansion centered at the origin is:

$$G(z) = 1 + z^2 \psi(z)$$

where $\psi(z) = \sum_{n=2}^\infty a_n z^{n-2}$ is an entire function. Because $\psi(z)$ is continuous on $\overline{B_1(0)}$, there exists $M_1 \geq 0$ such that for all $|z| = \epsilon \leq 1$ we have

$$|G(z) - 1| \leq \epsilon^2 M_1.$$

Invoking the inequality $|\log(1+x)| \leq 2|x|$ for all $|x| \leq \frac{1}{2}$, we choose $\epsilon \leq \min\left(1, \frac{1}{\sqrt{2M_1}}\right)$ to ensure $|G(z) - 1| \leq \frac{1}{2}$. This allows us to bound $u = \log |G(z)|$ as

$$|u(\gamma_0(\theta))| = |\operatorname{Re}(\log G(z))| \leq |\log G(z)| \leq 2|G(z) - 1| \leq 2M_1 \epsilon^2.$$

Furthermore, we can express $G'(z) = z\phi(z)$ where $\phi(z)$ is bounded by some $M_2 \geq 0$ on $\overline{B_1(0)}$. Then, we obtain that

$$\left| \frac{\partial u}{\partial n}(\gamma_0(\theta)) \right| = \left| \operatorname{Re} \left(\frac{G'(z)}{G(z)} e^{i\theta} \right) \right| \leq \frac{|G'(z)|}{|G(z)|} \leq \frac{\epsilon M_2}{1 - \epsilon^2 M_1} \leq 2M_2 \epsilon \quad \text{for } \epsilon^2 M_1 \leq \frac{1}{2}.$$

Similarly, since $|\sin \theta| \leq 1$, for any $\epsilon < R$ we have

$$|v(\gamma_0(\theta))| \leq \frac{1}{\epsilon} + \frac{\epsilon}{R^2} \leq \frac{2}{\epsilon}, \quad \left| \frac{\partial v}{\partial n}(\gamma_0(\theta)) \right| \leq \frac{1}{\epsilon^2} + \frac{1}{R^2} \leq \frac{2}{\epsilon^2}.$$

As ϵ is sufficiently small, putting these results together leads to

$$\left| \int_{S_\epsilon} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds \right| \leq \int_0^\pi \left[(2M_1 \epsilon^2) \cdot \left(\frac{2}{\epsilon^2} \right) + \left(\frac{2}{\epsilon} \right) \cdot (2M_2 \epsilon) \right] \epsilon d\theta \leq \pi(4M_1 + 4M_2)\epsilon.$$

Therefore, we obtain that

$$\lim_{\epsilon \rightarrow 0^+} \int_{S_\epsilon} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = 0. \quad (36)$$

Step 3.5: Integral of Ξ over γ_{z_k} . We now move to $\lim_{\epsilon \rightarrow 0^+} \int_{\gamma_{z_k}} \Xi(s) ds$ for any $z_k \in \mathcal{Z}_+$. Indeed, the path γ_{z_k} can be parameterized by

$$\gamma_{z_k}(\theta) = z_k + \epsilon e^{i(2\pi - \theta)} = z_k + \epsilon e^{-i\theta}, \quad \theta \in [0, 2\pi].$$

Therefore, the differential arc length element is $ds = \|\gamma'_{z_k}(\theta)\| d\theta = \epsilon d\theta$. Assume that $z_k \in \mathcal{Z}_+$ be a zero of $G(z)$ with multiplicity $m_k \geq 1$. Then, we can factor $G(z) = (z - z_k)^{m_k} \psi(z)$, where $\psi(z)$ is analytic and satisfies $\psi(z_k) \neq 0$. Now, we expand $u(\gamma_{z_k}(\theta))$ near z_k and obtain that

$$u(\gamma_{z_k}(\theta)) = \log |(\epsilon e^{-i\theta})^{m_k} \psi(z_k + \epsilon e^{-i\theta})| = m_k \log \epsilon + \log |\psi(z_k + \epsilon e^{-i\theta})| = m_k \log \epsilon + \log |\psi(z_k)| + O(\epsilon).$$

Similarly, we have

$$\frac{\partial u}{\partial n}(\gamma_{z_k}(\theta)) = \frac{m_k}{\epsilon} + O(1), \quad v(\gamma_{z_k}(\theta)) = v(z_k) + O(\epsilon), \quad \frac{\partial v}{\partial n}(\gamma_{z_k}(\theta)) = O(1).$$

Therefore, we obtain that

$$\begin{aligned} \int_{\gamma_{z_k}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds &= \int_0^{2\pi} \left[\left(m_k \log \epsilon + O(1) \right) O(1) - \left(v(z_k) + O(\epsilon) \right) \left(\frac{m_k}{\epsilon} + O(1) \right) \right] \epsilon d\theta \\ &= \int_0^{2\pi} [O(\epsilon \log \epsilon) - m_k v(z_k) - O(\epsilon)] d\theta. \end{aligned}$$

As a consequence, we arrive at

$$\lim_{\epsilon \rightarrow 0^+} \int_{\gamma_{z_k}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = \int_0^{2\pi} -m_k v(z_k) d\theta = -2\pi m_k v(z_k). \quad (37)$$

Step 3.6: Integral of Ξ over γ_{x_k} . We finally evaluate $\lim_{\epsilon \rightarrow 0^+} \int_{\gamma_{x_k}} \Xi(s) ds$ for any $x_k \in \mathcal{Z}_{\mathbb{R}}$. Indeed, we can parameterize γ_{x_k} as

$$\gamma_{x_k}(\theta) = x_k + \epsilon e^{-i\theta} \quad \text{for } \theta \in [\pi, 2\pi].$$

Direct calculation yields that

$$\begin{aligned} \int_{\gamma_{x_k}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds &= \int_{\pi}^{2\pi} \left[\left(m_k \log \epsilon + O(1) \right) O(1) - O(\epsilon) \left(\frac{m_k}{\epsilon} + O(1) \right) \right] \epsilon d\theta \\ &= \int_{\pi}^{2\pi} [O(\epsilon \log \epsilon) - O(\epsilon)] d\theta. \end{aligned}$$

Therefore, we arrive at

$$\lim_{\epsilon \rightarrow 0^+} \int_{\gamma_{x_k}} \left(u \frac{\partial v}{\partial n} - v \frac{\partial u}{\partial n} \right) ds = 0. \quad (38)$$

Step 3.7: Putting everything together. Putting the results from equations (33)-(38) together, we obtain that

$$\int_{-R}^R \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx + \frac{2}{R} \int_0^{\pi} \log |G(Re^{i\theta})| \sin \theta d\theta = 2\pi \sum_{z_k \in \mathcal{Z}_+} m_k v(z_k).$$

Step 4: Integrability of $|\log |G(x)||/x^2$ at infinity. We now prove that

$$\int_{|x| \geq 1} \frac{|\log |G(x)||}{x^2} dx < \infty.$$

Since $v(z) > 0$ for all $z \in D_R^+$, Carleman formula in equation (32) leads to

$$\int_{-R}^R \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx + \frac{2}{R} \int_0^\pi \log |G(Re^{i\theta})| \sin \theta d\theta \geq 0. \quad (39)$$

Furthermore, since $|G(z)| \leq \bar{C}e^{2\tau R}$ on S_R , we obtain that

$$\frac{2}{R} \int_0^\pi \log |G(Re^{i\theta})| \sin \theta d\theta \leq \frac{2}{R} \int_0^\pi (\log \bar{C} + 2\tau R) \sin \theta d\theta = \frac{4 \log \bar{C}}{R} + 8\tau. \quad (40)$$

Collecting these bounds in (39) and (40) leads to

$$\int_{-R}^R \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \geq -4 \frac{|\log \bar{C}|}{R} - 8\tau.$$

Since $\log |G(x)| = \log^+ |G(x)| - \log^- |G(x)|$ where $\log^- t := \min\{0, \log t\}$ for any $t > 0$, this bound can be written as

$$\int_{-R}^R \log^- |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \leq \int_{-R}^R \log^+ |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx + 4 \frac{|\log \bar{C}|}{R} + 8\tau. \quad (41)$$

When $1 \leq |x| \leq R$, from the hypothesis that $|G|$ is bounded in $\mathbb{R}$, there exists positive constant M' not depending on R such that $\log^+ |G(x)| \leq M'$. As a result, we obtain

$$\int_{1 \leq |x| \leq R} \log^+ |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \leq \int_{1 \leq |x| \leq R} \frac{M'}{x^2} dx = 2M' \left(1 - \frac{1}{R} \right) \leq 2M'. \quad (42)$$

Combining these bounds in (41) and (42), we arrive at

$$\int_{1 \leq |x| \leq R} \log^- |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \leq 2M' + \int_{-1}^1 \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx + 4|\log \bar{C}| + 8\tau. \quad (43)$$

Using the estimation in (31), we have

$$\int_{-1}^1 \frac{|\log |G(x)||}{x^2} dx = W < \infty.$$

Therefore, we obtain that when $R > 1$,

$$\int_{-1}^1 \log |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \leq 2 \int_{-1}^1 \frac{|\log |G(x)||}{x^2} dx = 2W. \quad (44)$$

Combining the results from equation (43) and (44) leads to

$$\int_{1 \leq |x| \leq R} \log^- |G(x)| \left(\frac{1}{x^2} - \frac{1}{R^2} \right) dx \leq 2M' + 2W + 4|\log \bar{C}| + 8\tau.$$

By taking $R \rightarrow \infty$ and using Monotone Convergence Theorem, this inequality eventually indicates that $\int_{|x| \geq 1} \frac{\log^- |G|}{x^2} dx < \infty$. As $\int_{|x| \geq 1} \frac{\log^+ |G|}{x^2} dx \leq 2M'$, these two results together leads to

$$\int_{|x| \geq 1} \frac{|\log |G(x)||}{x^2} dx < \infty.$$

Step 5: Conclusion. Since $\frac{1}{1+x^2} < \frac{1}{x^2}$ and $\frac{\log |G(x)|}{x^2}$ is in $L^1([-1, 1])$, we obtain that

$$\begin{aligned} \int_{|x| \geq 1} \frac{|\log |G(x)||}{1+x^2} dx &< \int_{|x| \geq 1} \frac{|\log |G(x)||}{x^2} dx < \infty, \\ \int_{|x| \leq 1} \frac{|\log |G(x)||}{1+x^2} dx &< \int_{|x| \leq 1} \frac{|\log |G(x)||}{x^2} dx < \infty. \end{aligned}$$

Consequently, we arrive at

$$\int_{-\infty}^{\infty} \frac{|\log |G(x)||}{1+x^2} dx < \infty,$$

which indicates that (26) is correct. We obtain the conclusion of the lemma. $\square$

The next lemma supplies the continuity used in Appendix A.5.

Lemma 2. *Suppose that, for λ -almost every x , the map $\theta \mapsto f(x | \theta)$ is bounded and continuous on Θ . If $G_k, G \in \mathcal{P}(\Theta)$ and $W_1(G_k, G) \rightarrow 0$, then $\|p_{G_k} - p_G\|_1 \rightarrow 0$.*

Proof. On the compact Θ , W_1 metrises weak convergence, so $G_k \rightarrow G$ weakly. For λ -almost everywhere x , $f(x | \cdot)$ is bounded and continuous, whence

$$p_{G_k}(x) = \int_{\Theta} f(x | \theta) dG_k(\theta) \rightarrow \int_{\Theta} f(x | \theta) dG(\theta) = p_G(x), \text{ a.e.}$$

Since each p_{G_k} and p_G is a probability density and $p_{G_k} \rightarrow p_G$ almost everywhere with respect to measure λ , Scheffé's lemma yields $\|p_{G_k} - p_G\|_1 \rightarrow 0$. This completes the proof of Lemma 2. $\square$

References

- Ascolani, F. (2026). Posterior concentration and adaptation of the mixing measure in Dirichlet process mixtures. *arXiv preprint arXiv:2606.29109*. (Cited on page 5.)
- Bariletto, N., Le, D., Rinaldo, A., and Ho, N. (2026a). Convergence Rates for Latent Mixing Measures in Infinite Homoscedastic Location-Scale Mixture Models. *arXiv preprint arXiv:2605.06845*. (Cited on pages 5, 6, and 13.)
- Bariletto, N., Nguyen, H., Ho, N., and Rinaldo, A. (2026b). On Bayesian Softmax-Gated Mixture-of-Experts Models. *arXiv preprint arXiv:2604.20551*. (Cited on page 5.)
- Barron, A., Schervish, M., and Wasserman, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.*, 27:536–561. (Cited on page 5.)
- 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. (Cited on page 5.)

- 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(1):234–265. (Cited on page 5.)
- 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. (Cited on page 5.)
- Do, D., Do, L., and Nguyen, X. (2025). Strong identifiability and parameter learning in regression with heterogeneous response. *Electronic Journal of Statistics*, 19(1):131–203. (Cited on page 5.)
- Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1:209–230. (Cited on page 2.)
- Ferguson, T. S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent advances in statistics*, pages 287–302. Elsevier. (Cited on page 2.)
- Folland, G. B. (1999). *Real analysis: modern techniques and their applications*. John Wiley & Sons. (Cited on pages 2 and 5.)
- Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631. (Cited on page 1.)
- Gao, F. and van der Vaart, A. (2016). Posterior contraction rates for deconvolution of Dirichlet-Laplace mixtures. *Electronic Journal of Statistics*, 10(1):608–627. (Cited on pages 5 and 11.)
- 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. (Cited on page 5.)
- Ghosal, S., Ghosh, J. K., and van der Vaart, A. (2000). Convergence rates of posterior distributions. *Annals of Statistics*, 28:500–531. (Cited on page 5.)
- Ghosal, S. and van der Vaart, A. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *The Annals of Statistics*, 29:1233–1263. (Cited on page 5.)
- Ho, N. and Nguyen, X. (2016a). Convergence rates of parameter estimation for some weakly identifiable finite mixtures. *Annals of Statistics*, 44:2726–2755. (Cited on page 5.)
- Ho, N. and Nguyen, X. (2016b). On strong identifiability and convergence rates of parameter estimation in finite mixtures. *Electronic Journal of Statistics*, 10:271–307. (Cited on page 5.)
- Ho, N., Yang, C.-Y., and Jordan, M. I. (2022). Convergence rates for Gaussian mixtures of experts. *Journal of Machine Learning Research*, 23(323):1–81. (Cited on page 5.)
- Koosis, P. (1988). *The Logarithmic Integral. Volume I*, volume 12 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. (Cited on page 26.)
- Kroó, A. (1994). A Geometric Approach to the Multivariate Müntz Problem. *Proceedings of the American Mathematical Society*, 121(1):199–208. (Cited on page 12.)

- Levin, B. Y. (1980). *Distribution of Zeros of Entire Functions*, volume 5 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, revised edition. (Cited on page 28.)
- Lo, A. Y. (1984). On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics*, 12(1):351–357. (Cited on page 2.)
- Manole, T. and Ho, N. (2022). Refined convergence rates for maximum likelihood estimation under finite mixture models. In *International Conference on Machine Learning*, pages 14979–15006. PMLR. (Cited on page 5.)
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons. (Cited on page 1.)
- Müntz, H. (1914). Über den Approximationssatz von Weierstrass. In *H. A. Schwarz’s Festschrift*, pages 303–312. Julius Springer, Berlin. (Cited on page 12.)
- Nguyen, H., Nguyen, T., and Ho, N. (2023). Demystifying softmax gating function in Gaussian mixture of experts. In *Advances in Neural Information Processing Systems*. (Cited on page 5.)
- Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Annals of Statistics*, 4(1):370–400. (Cited on pages 4, 5, and 11.)
- Nguyen, X. and Wei, Y. (2026). Optimal transport based theory for latent structured models. *arXiv preprint arXiv:2601.11465*. (Cited on page 5.)
- Rousseau, J. and Scricciolo, C. (2024). Wasserstein convergence in Bayesian and frequentist deconvolution models. *The Annals of Statistics*, 52(4):1691–1715. (Cited on page 5.)
- Scricciolo, C. (2018). Bayes and maximum likelihood for L_1 -Wasserstein deconvolution of Laplace mixtures. *Statistical Methods & Applications*, 27(2):333–362. (Cited on page 5.)
- Teicher, H. (1960). On the mixture of distributions. *The Annals of Mathematical Statistics*, 31(1):55–73. (Cited on page 5.)
- Teicher, H. (1961). Identifiability of mixtures. *The Annals of Mathematical Statistics*, 32(1):244–248. (Cited on page 5.)
- Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 34(4):1265–1269. (Cited on page 5.)
- Villani, C. (2008). *Optimal transport: Old and New*. Springer. (Cited on pages 4 and 5.)
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press. (Cited on page 21.)
- Willard, S. (1970). *General Topology*. Addison–Wesley Publishing Company, Reading, Massachusetts. (Cited on page 4.)
- Yang, X. (2014). The Multivariate Müntz–Szász Problem in Weighted Banach Space on $\mathbb{R}^n$. *Abstract and Applied Analysis*, 2014:1–8. (Cited on page 12.)