Quasi-Bayes in Latent Variable Models *

Sid Kankanala

Department of Economics, Yale University

January 20, 2024

Click here for the latest version

Abstract

Latent variable models are widely used to account for unobserved determinants of economic behavior. Traditional nonparametric methods to estimate latent heterogeneity do not scale well into multidimensional settings. Distributional restrictions alleviate tractability concerns but may impart non-trivial misspecification bias. Motivated by these concerns, this paper introduces a quasi-Bayes approach to estimate a large class of multidimensional latent variable models. Our approach to quasi-Bayes is novel in that we center it around relating the characteristic function of observables to the distribution of unobservables. We propose a computationally attractive class of priors that are supported on Gaussian mixtures and derive contraction rates for a variety of latent variable models. As a first application, we use data from the India Young Lives Survey to estimate production functions for cognition and health for children aged 1-12 in India. As a secondary application, we model individual log earnings from the Panel Study of Income Dynamics (PSID) as the sum of permanent and transitory components. Simulations illustrate the performance of quasi-Bayes estimators relative to common alternatives.

*Email address: sid.kankanala@yale.edu. I thank Xiaohong Chen for continuous guidance and support throughout this project. I thank Donald Andrews and Yuichi Kitamura for valuable comments and suggestions that improved the quality of this paper. I have benefited from discussions with Orazio Attanasio, Tim Christensen, Max Cytrynbaum, Lihua Lei, Costas Meghir, Emily Nix, Edward Vytlacil, Stefan Wager, Qianyao Ye, Victoria Zinde-Walsh, seminar participants at Yale and the JHU Workshop on Measurement Errors and Latent Variables. I thank Stéphane Bonhomme and Irene Botosaru for graciously providing the PSID earnings data. All errors are my own.

1 Introduction

Many questions in economics involve inferring information on unobserved determinants of economic behavior. For instance, Cunha, Heckman, and Schennach (2010) and Attanasio, Meghir, and Nix (2020b) study the importance of human capital development on child outcomes. In this setting, observed variables such as test scores and height serve as imperfect proxies for latent ability and health. Similarly, models on search unemployment (Heckman and Singer, 1984) and labor income dynamics (Abowd and Card, 1989; Altonji, Smith Jr, and Vidangos, 2013) explicitly incorporate unobserved heterogeneity and measurement error in observables (e.g. observed wages, hours and earnings).

To account for the informational uncertainty that arises from unobservables, models that incorporate latent variables are widely used. In multidimensional settings, a common strategy to maintain tractability involves imposing distributional restrictions on observables and unobservables.[1] These approaches are attractive in that they provide a researcher with a tractable distribution (e.g. Gaussian) to sample from for counterfactual analysis. On the downside, they may impart non-trivial misspecification bias through the distributional restrictions.[2] Motivated by these concerns, this paper introduces a novel quasi-Bayes approach to estimate a large class of multidimensional latent variable models.

Traditional quasi-Bayes (Kim, 2002; Chernozhukov and Hong, 2003) combines a GMM objective function $Q_n(\theta)$ and a prior $\nu(\theta)$ to create a quasi-posterior distribution. Intuitively, the GMM objective function reweights the prior distribution to assign greater weight to areas of the parameter space where the objective function is small. As a consequence, posterior samples concentrate around minimizers of the GMM objective function. The usual quasi-posterior takes the form

$$\nu(\theta|\mathcal{Z}) \propto e^{-Q_n(\theta)}\nu(\theta). \tag{1}$$

In this paper, we generalize the representation in (1) to settings where θ represents a latent distribution of interest. While traditional quasi-Bayesian approaches in econometrics focus on moment restrictions that identify a finite dimensional parameter, our approach instead centers around identifying restrictions which relate the characteristic functions of observables to the distribution of unobservables. Our analysis begins with the observation that a large class of latent variable models are characterized by a collection of simple identifying restrictions on the characteristic function of the observables. We use these restrictions to build a pseudo-likelihood for the model. When combined with a prior, this provides a researcher with quasi-Bayes decision rules such as point estimators (e.g. posterior means) and credible sets.

[1] For instance, in their empirical specification, Cunha et al. (2010) and Attanasio et al. (2020a,b) fix the distribution of unobservables to lie in a Gaussian family. This induces a Gaussian likelihood on observables, which facilitates the use of parametric maximum likelihood.

[2] In low dimensional settings, deconvolution based approaches (e.g. Horowitz and Markatou, 1996; Li and Vuong, 1998; Bonhomme and Robin, 2010) are commonly used to estimate the distribution of unobservables. These methods require minimal distributional restrictions on unobservables but are computationally challenging and highly sensitive to user chosen tuning parameters, especially so in multidimensional setups.

We propose a class of priors that are supported on Gaussian mixtures. Our focus on this class is partially motivated from the observation that finite Gaussian mixture models are widely used in a variety of empirical settings to model rich forms of heterogeneity. As Gaussian mixtures can be efficiently sampled from, they provide researchers with a convenient framework to perform counterfactual analysis. One consequence of building a framework around such a structure is that we obtain decision rules with particularly desirable theoretical guarantees in settings where the true data generating process closely resembles a (possibly infinite) Gaussian mixture, while at the same time remaining viable in the more general case. A further desirable property of focusing on Gaussian mixture based priors is that they greatly facilitate the interplay analysis between the prior and pseudo-likelihood. Importantly, Gaussian mixtures admit a simple characteristic function which in turn leads to a tractable pseudo-likelihood. Indeed, one of the main ideas behind our analysis is that priors on Gaussian mixtures translate effortlessly to priors on characteristic functions and vice versa.

The main contributions of this paper are as follows. We provide a unified treatment for three classes of latent variable models. These are (*i*) models with classical measurement error, (*ii*) models with repeated measurements and (*iii*) linear multi-factor models. In each case, we derive characteristic function based identifying restrictions and use them to build a pseudo-likelihood for the model. We then combine this likelihood with a theoretically motivated class of priors and derive L^2 rates of convergence (posterior contraction rates) for the induced quasi-Bayes posterior. As our analysis is the first quasi-Bayesian approach to these classes of models, we expect that the general analysis may be of independent interest towards related extensions. As a by product of our analysis, we also contribute to a general understanding of characteristic function based pseudo-likelihoods. Importantly, certain likelihoods may only lead to partial identification but can nonetheless be modified in a suitable way to achieve point identification.

To the best of our knowledge, our paper is the first nonparametric Bayesian approach to utilize a pseudo-likelihood based on characteristic functions. In our analysis, we use a Dirichlet process prior to induce a prior on Gaussian mixtures which in turn leads to an induced prior on characteristic functions. One consequence of this analysis is that we obtain novel L^∞ posterior contraction results even for the special case of density estimation. Intuitively, convergence of characteristic functions implies risk bounds in stronger metrics.[3] This approach is of independent interest and may be useful in a variety of related extensions.

As a first application, we use data collected from the India Young Lives Survey to estimate production functions for cognition and health for children aged 1-12 in India, revisiting Attanasio, Meghir, and Nix (2020b). As in the original analysis, we use the latent factor model of Cunha, Heckman, and Schennach (2010) to estimate the joint distribution of unobservables and observables. Our analysis deviates from the literature in that we estimate this joint distribution nonparametrically and do so without imposing a specific distribution on the measurement

[3] For example, if f_X and f_Y have characteristic functions φ_X and φ_Y, $\|f_X - f_Y\|_{L^\infty} \leq \|\varphi_X - \varphi_Y\|_{L^1}$.

errors.[4] In line with previous results in the literature, our results show that cognition displays persistence across the development period and that parental investment is effective at developing cognition at all ages, with a higher return for younger children. Our results differ from the literature in that, while previous studies attribute cognition at later ages primarily to persistence, our results suggest a greater influence of alternative factors. Specifically, we find that investment continues to maintain a significant impact during adolescence, albeit still lower than in early childhood. As our sample size is relatively modest, and we do not consider estimation by relaxing only a single restriction, it is difficult to ascertain whether our differences are due to nonparametrically modeling the latent distribution and/or leaving free the distribution of the measurement errors. To that end, we view our general methodology as complementary to the existing literature in that it may serve as a robustness check to possible violations of these restrictions.

As a secondary application, we apply our methodology to model individual earnings data from the Panel Study of Income Dynamics (PSID), revisiting Bonhomme and Robin (2010). Our results in this setting are consistent with the literature (e.g. Geweke and Keane, 2000; Bonhomme and Robin, 2010) in that the distribution of wage shocks in U.S data appear to be non-Gaussian and leptokurtic. Additionally, as our priors are supported on infinite Gaussian mixtures, our results also complement previous approaches (e.g. Geweke and Keane, 2000) that advocate for Gaussian mixtures to model earnings dynamics. To assess the performance of our method, we compare the implied fit of the wage growth moments to alternative estimators and the observed data. Overall, we find that our approach is capable of reproducing the higher moments observed in U.S wage growth data.

The paper is organized as follows. Section 1.2 briefly summarizes notation that frequently appears throughout the text. Section 2 provides a brief review of Gaussian mixtures, the Dirichlet process and the prior it induces on Gaussian mixtures. Section 3 introduces the quasi-Bayes framework for all the latent variable models considered in this paper. Section 4 develops the quasi-Bayes limit theory and main results. Section 5 provides simulation evidence on quasi-Bayes estimators relative to common alternatives. In Section 6, we apply our methodology to study human capital development in India, using data from the Young Lives Survey. In Section 7, we apply our methodology to study earnings dynamics in U.S wages, using data from the PSID. Section 8 concludes. Section 9 contains proofs and auxiliary results for all the statements in the main text.

1.1 Related Literature

Latent variable models have a long history in econometrics and statistics. For a comprehensive overview, we refer to Chen, Hong, and Nekipelov (2011); Schennach (2016, 2022). Heckman and Singer (1984) establish the consistency of a sieve maximum likelihood estimator for the

[4] This refers to estimation methods that explicitly depend on the specific form of the measurement error distribution in their implementation. For example, Gaussian measurement errors are frequently used to induce a Gaussian likelihood on the observables. If the latent distribution is also assumed to be Gaussian, its mean and covariance structure can be recovered through minimum distance least squares.

distribution of unobservables in a single spell duration model. For classical measurement error with a known error distribution, Fan (1991a,b); Lounici and Nickl (2011) establish optimal convergence rates of kernel deconvolution estimators. For early work on measurement error in panel frameworks, see Horowitz and Markatou (1996). For models with repeated measurements, Li and Vuong (1998) use Kotlarski's lemma (Kotlarski, 1967) to build a deconvolution based estimator. For extensions and related results on models with repeated measurements or auxiliary information, see Li (2002); Schennach (2004a,b); Chen, Hong, and Tamer (2005); Delaigle, Hall, and Meister (2008); Cunha, Heckman, and Schennach (2010); Kato, Sasaki, and Ura (2021); Kurisu and Otsu (2022). For general multi-factor models with mutually independent factors, Bonhomme and Robin (2010) propose a kernel deconvolution based estimator. For instrumental variable based approaches to models with measurement error, see Newey (2001); Schennach (2007); Hu (2008); Hu and Schennach (2008); Wang and Hsiao (2011).

Our paper is also related to a growing literature on Bayesian inference in density estimation and deconvolution. For pure Bayesian approaches to density estimation, see Ghosal, Ghosh, and Van Der Vaart (2000); Ghosal and Van Der Vaart (2001, 2007); Kruijer, Rousseau, and van der Vaart (2010); Shen, Tokdar, and Ghosal (2013). For pure Bayesian approaches to deconvolution with a known error distribution, see Donnet, Rivoirard, Rousseau, and Scricciolo (2018); Rousseau and Scricciolo (2023). To the best of our knowledge, our paper is the first to propose a quasi-Bayes framework for these models. As such, our paper also contributes to a growing literature on quasi-Bayes in nonparametric econometric models (e.g. Liao and Jiang, 2011; Kato, 2013).

Quasi-Bayesian methods have a long history in econometric models identified through moment restrictions. Traditional approaches (e.g. Chernozhukov and Hong, 2003; Gallant, Hong, Leung, and Li, 2022) have typically focused on strongly identified parametric models. Recent work (e.g. Chen, Christensen, and Tamer, 2018; Andrews and Mikusheva, 2022) in the literature also demonstrates that certain quasi-Bayesian decision rules have desirable properties in settings where identification is weak or non standard. Although our focus in this paper is not on settings with non standard identification, we do note that (in our nonparametric ill-posed inverse setting) finite sample identification strength is typically user determined by tuning parameters that determine the structure of the parameter space. A complex parameter space leads to objective function that are near-flat in substantial portions of the parameter space. As traditional estimators (e.g. deconvolution based) are typically based on directly inverting these objective functions, they are highly sensitive to the underlying structure.[5] By contrast, quasi-Bayes provides additional regularization to solve the ill-posed inverse problem through a prior and flexible choice of pseudo-likelihood.

[5]Intuitively, an estimator that is based on inverting an objective function that is near-flat in substantial portions of a parameter space typically displays large variability.

1.2 Notation

Denote the standard Euclidean real coordinate space of dimension d by $\mathbb{R}^d$. For complex coordinates, we use $\mathbb{C}^d$. Let $\mathbf{i}$ denote the imaginary unit. The magnitude of a complex valued function $f : \mathbb{R}^d \to \mathbb{C}$ is defined as $|f(t)|^2 = (\Re[f(t)])^2 + (\Im[f(t)])^2$, where $\Re$ and $\Im$ denote the real and imaginary parts, respectively. Given a random vector $Z \in \mathbb{R}^d$, we denote its characteristic function by $\varphi_Z(t) = \mathbb{E}[e^{\mathbf{i}t'Z}]$. Given functions $f, g : \mathbb{R}^d \to \mathbb{R}$, we denote their convolution by $f \star g(x) = \int_{\mathbb{R}^d} f(x-z)g(z)dz$. Convolution of a function $g : \mathbb{R}^d \to \mathbb{R}$ with a Borel measure μ on $\mathbb{R}^d$ is denoted in a similar manner by $g \star \mu(x) = \int_{\mathbb{R}^d} g(x-z)d\mu(z)$. The Fourier Transform of a function $f : \mathbb{R}^d \to \mathbb{R}$ and a Borel measure μ on $\mathbb{R}^d$ is denoted by $\mathcal{F}[f](t) = \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} f(x)dx$ and $\mathcal{F}[\mu](t) = \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} d\mu(x)$, respectively.

Let $\mathbb{E}$ and $\mathbb{P}$ denote the usual expectation and probability operators. Let $\mathbb{E}_n$ denote the empirical analog of $\mathbb{E}$. We use $\|.\|$ and $\|.\|_\infty$ to denote the Euclidean and infinity norm on $\mathbb{R}^d$, respectively. We use $\mathbf{S}_+^d$ to denote the set of positive definite matrices on $\mathbb{R}^{d\times d}$. Given two measures (λ, μ), we denote the product measure by $\nu = \lambda \otimes \mu$. Let $\mathbf{H}^p = (\mathbf{H}^p, \|.\|_{\mathbf{H}^p})$ denote the usual p-Sobolev space of functions on $\mathbb{R}^d$. Given a symmetric positive definite matrix $\Sigma \in \mathbf{S}_+^d$, we denote its ordered eigenvalues by $\lambda_1(\Sigma) \leq \lambda_2(\Sigma) \leq \ldots \leq \lambda_d(\Sigma)$. The L^2 space with respect to the Lebesgue measure on $\mathbb{R}^d$ is denoted by $L^2(\mathbb{R}^d)$, with associated norm given by $\|f\|_{L^2}^2 = \int_{\mathbb{R}^d} |f(t)|^2 \, dt$.

We introduce more specific notation for multivariate Gaussian distributions to denote various dependencies that we make use of. Let $\mathcal{N}(0, \Sigma)$ denote the multivariate Gaussian distribution with mean zero and covariance matrix Σ. We use ϕ_Σ and φ_Σ to denote the $\mathcal{N}(0, \Sigma)$ density and characteristic function, respectively.

2 Review

We begin in Section 2.1 with a brief review of Gaussian mixtures. In Section 2.2, we review the Dirichlet process and the prior it induces on Gaussian mixtures.

2.1 Gaussian Mixtures

Given a probability distribution P on $\mathbb{R}^d$, the Gaussian mixture with mixing distribution P and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ is denoted by

$$\phi_{P,\Sigma}(x) = \phi_\Sigma \star P(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x-z)dP(z). \tag{2}$$

If P is a discrete distribution, i.e $P = \sum_{j=1}^\infty p_j \delta_{\mu_j}$ which assigns probability mass p_j to a point $\mu_j \in \mathbb{R}^d$, the preceding definition reduces to

$$\phi_{P,\Sigma}(x) = \sum_{j=1}^\infty p_j \phi_\Sigma(x - \mu_j).$$

If $\mathcal{F}$ denotes the Fourier transform operator, the characteristic function (CF) of a Gaussian mixture is given by

$$\varphi_{P,\Sigma}(t) = \mathcal{F}[\phi_\Sigma \star P](t) = \mathcal{F}[\phi_\Sigma](t) \times \mathcal{F}[P](t) = e^{-t'\Sigma t/2}\mathcal{F}[P](t). \tag{3}$$

For a discrete distribution $P = \sum_{j=1}^{\infty} p_j \delta_{\mu_j}$, the preceding expression simplifies to

$$\varphi_{P,\Sigma}(t) = e^{-t'\Sigma t/2} \sum_{j=1}^{\infty} p_j e^{\mathbf{i}t'\mu_j}.$$

From the representation in (2), it follows that a Gaussian mixture is completely characterized by its mixing distribution P and covariance matrix Σ.

In practice, a researcher may choose to model the distribution of a random vector $X \in \mathbb{R}^d$ as an exact Gaussian mixture of the form in (2). That is, there exists a positive definite matrix $\Sigma_0 \in \mathbb{R}^{d\times d}$ and a mixing distribution F_0 such that the density of X can be expressed as

$$f_X(x) = \phi_{\Sigma_0} \star F_0(x) = \int_{\mathbb{R}^d} \phi_{\Sigma_0}(x-z)dF_0(z). \tag{4}$$

As the mixing distribution F_0 may be continuously distributed and/or possess unbounded support, the representation in (4) offers a flexible parametrization that lies somewhere between a parametric and fully nonparametric model. Furthermore, while the representation in (4) may be misspecified, there always exists a mixing distribution and covariance matrix (Σ_0, F_0) such that the misspecification bias can be made arbitrarily small.[6]

2.2 Dirichlet Process Priors

The Dirichlet process is a distribution on the space of probability distributions on $\mathbb{R}^d$. That is, a random draw from the Dirichlet process is a distribution P on $\mathbb{R}^d$. As the name suggests, the precise definition of the process is closely related to that of the finite dimensional Dirichlet distribution on the simplex. The following definition clarifies this.

Definition 1. A random distribution P on $\mathbb{R}^d$ is a Dirichlet process distribution with base measure α if for every finite measurable partition of $\mathbb{R}^d = \bigcup_{i=1}^k A_i$, we have

$$(P(A_1), P(A_2), \ldots, P(A_k)) \sim \mathrm{Dir}(k, \alpha(A_1), \ldots, \alpha(A_k)), \tag{5}$$

where Dir(.) is the Dirichlet distribution with parameters k and $\alpha_1(A_1), \ldots, \alpha_k(A_k) > 0$ on the

[6]To be specific, suppose $X \in \mathbb{R}^d$ admits a density $f_X \in L^2(\mathbb{R}^d)$. A simple way to achieve a (not necessarily optimal) Gaussian mixture approximation is to let $dP = f_X$ and $\Sigma_n = \sigma_n^2 I$ for some $\sigma_n^2 \downarrow 0$. As the Fourier transform preserves distance (up to a constant), it follows that

$$\|\phi_{P,\Sigma_n}(x) - f_X(x)\|_{L^2} = \|f_X \star \phi_{\sigma_n^2 I} - f_X\|_{L^2} \asymp \|(e^{-\|t\|^2\sigma_n^2/2} - 1)\varphi_X(t)\|_{L^2} \xrightarrow[n\to\infty]{} 0.$$

k dimensional unit simplex $\Delta = \{x \in \mathbb{R}^k : x_i \geq 0\ , \sum_{i=1}^{k} x_i = 1\}$.[7]

An alternative definition that is especially useful in simulating the process is the following.

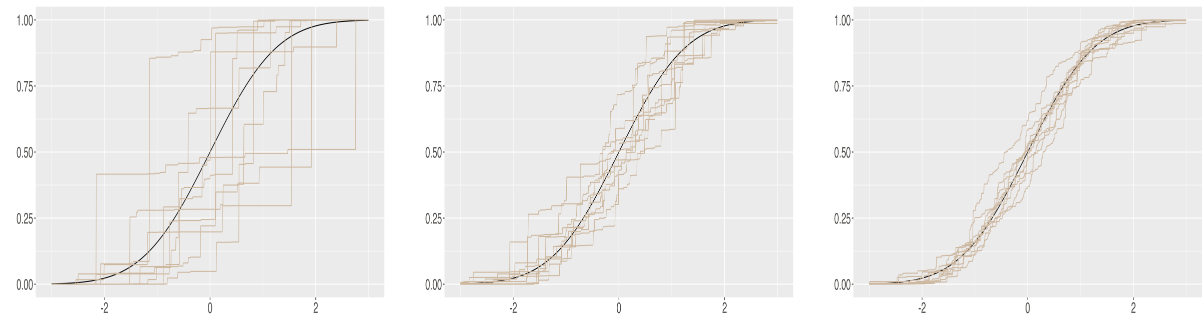
Definition 2 (Dirichlet process)**.** Fix $\beta > 0$ and a base probability distribution α on $\mathbb{R}^d$. Let $V_1, V_2, \ldots \overset{i.i.d}{\sim} \text{Beta}(1, \beta)$ and $\mu_1, \mu_2, \ldots \overset{i.i.d}{\sim} \alpha$. Define the weights

$$p_1 = V_1 \ \ , \ p_j = V_j \prod_{i=1}^{j-1}(1 - V_i) \qquad j \geq 2.$$

It can be shown (Ghosal and Van der Vaart, 2017, Lemma 3.4) that $\sum_{i=1}^{\infty} p_i = 1$ almost surely. The Dirichlet process with concentration parameter β and base measure α is the law of the random discrete distribution

$$P = \sum_{i=1}^{\infty} p_i \delta_{\mu_i}\ ,$$

where δ_{μ_i} is the point mass at μ_i. That is, P takes value μ_i with probability p_i. Notationally, we suppress the dependence on β and express this as $P \sim \text{DP}_\alpha$. As Figure 1 illustrates, the parameter β controls the spread around the base measure α, with larger values of β leading to tighter concentration.



(a) $\beta = 2$ (b) $\beta = 10$ (c) $\beta = 50$

Figure 1: CDF Draws of $P \sim \text{DP}_\alpha$. Black solid line indicates CDF of base measure $\alpha = N(0, 1)$.

A Dirichlet process prior can be used to build a prior on Gaussian mixtures in (2). Specifically, given a Dirichlet process prior DP_α and an independent prior G on the set $\mathbf{S}_+^d$ of positive definite matrices, the induced prior on Gaussian mixtures is given by

$$\phi_{P,\Sigma}(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x - z) dP(z)\ , \quad (P, \Sigma) \sim \text{DP}_\alpha \otimes G. \tag{6}$$

In the remaining sections, we will frequently refer to the prior in (6). For ease of notation, we denote this product prior by $\nu_{\alpha,G} = \text{DP}_\alpha \otimes G$.

[7] The probability density of $\text{Dir}(k, \alpha_1, \ldots, \alpha_k)$, with respect to the $k - 1$ dimensional Lebesgue measure on the unit simplex Δ, is proportional to $\prod_{i=1}^{k} x_i^{\alpha_i - 1}$.

3 Models and Procedures

In this section, we introduce three commonly used latent variable models: (i) classical measurement error[8], (ii) models with repeated measurements and (iii) linear independent multi-factor models. For each model, we consider identifying restrictions that relate the distribution of the latent variable to the characteristic function of observables. These restrictions are used to construct a quasi-likelihood. When combined with a prior as in (6), we obtain a quasi-Bayes posterior that is supported on infinite Gaussian mixtures.

3.1 Classical Measurement Error

Consider a classical measurement error model where we observe a random sample from

$$Y = X + \epsilon \quad , \quad \mathbb{E}[\epsilon] = 0. \tag{7}$$

Here, $Y \in \mathbb{R}^d$ is an observed vector, $X \in \mathbb{R}^d$ is an unobserved vector whose distribution is of interest and ϵ is an unobserved error that is independent of X. We may view Y as an imperfect proxy or error contaminated version of X.

We are interested in recovering the latent distribution of X. As the individual contributions of X and ϵ cannot be separately identified from an observation of Y, identification of the latent distribution of X generally requires some further auxiliary information on the distribution F_ϵ of the unobserved error ϵ. In the deconvolution literature (e.g. Lounici and Nickl, 2011; Rousseau and Scricciolo, 2023), it is common to assume that the distribution F_ϵ is completely known. As in Dattner et al. (2016); Kato and Sasaki (2018), we consider a slightly weaker requirement where a researcher has auxiliary information in the form of a random sample $\epsilon_1, \dots, \epsilon_m \overset{i.i.d}{\sim} F_\epsilon$ of size m.[9] As we discuss in Section 3.2, one way to generate such auxiliary information is through repeated measurements of Y.

We start by observing that, as the unobserved error ϵ is independent from the latent vector X, the characteristic functions of (Y, X, ϵ) factor as

$$\varphi_Y(t) = \varphi_X(t)\varphi_\epsilon(t) \qquad \forall\, t \in \mathbb{R}^d. \tag{8}$$

Our starting point towards constructing a quasi-likelihood is to observe that (8) can be viewed as a collection of identifying restrictions on the model. As the characteristic functions $\varphi_Y(t) = \mathbb{E}[e^{\mathbf{i}t'Y}]$ and $\varphi_\epsilon(t) = \mathbb{E}[e^{\mathbf{i}t'\epsilon}]$ depend on population expectations, it is infeasible to evaluate either side of (8) directly. Towards that goal, given the observed data, we estimate the population

[8] This is also known as the multivariate deconvolution model (Fan, 1991a,b; Lounici and Nickl, 2011)

[9] There are no restrictions on the dependence of the auxiliary sample with the original sample.

expectations using their empirical analogs:

$$\widehat{\varphi}_Y(t) = \mathbb{E}_n[e^{\mathbf{i}t'Y}] = \frac{1}{n}\sum_{j=1}^{n} e^{\mathbf{i}t'Y_j} \tag{9}$$

$$\widehat{\varphi}_\epsilon(t) = \mathbb{E}_m[e^{\mathbf{i}t'\epsilon}] = \frac{1}{m}\sum_{j=1}^{m} e^{\mathbf{i}t'\epsilon_j}. \tag{10}$$

As estimated characteristic functions do not have uniformly (over $t \in \mathbb{R}^d$) valid statistical guarantees, it is necessary to consider a bounded set of restrictions that grows with the sample size.[10] To that end, given a fixed $T > 0$, we denote the ball of radius T and the L^2 magnitude (norm) of a function $f : \mathbb{R}^d \to \mathbb{C}$ over the ball by

$$\mathbb{B}(T) = \{t \in \mathbb{R}^d : \|t\|_\infty \leq T\} \ , \quad \|f\|^2_{\mathbb{B}(T)} = \int_{\mathbb{B}(T)} |f(t)|^2 \, dt. \tag{11}$$

If θ represents a distribution with characteristic function φ_θ, we aim to construct a quasi-likelihood for the latent distribution of X using the quantity

$$L(\theta) = -\|\widehat{\varphi}_Y - \varphi_\theta \widehat{\varphi}_\epsilon\|^2_{\mathbb{B}(T)}. \tag{12}$$

While in principle $L(.)$ can be evaluated at any distribution θ, in the interest of computational tractability, it will be convenient to restrict our attention to distributions that admit a simple characteristic function. At a minimum, this ensures that the quasi-likelihood in (12) can be easily evaluated at such distributions. Towards this goal, we consider candidate solutions in the class of infinite Gaussian mixtures:

$$\phi_{P,\Sigma}(x) = \phi_\Sigma \star P(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x - z) dP(z). \tag{13}$$

As discussed in Section 2.1, a Gaussian mixture is completely characterized by its mixing distribution P and covariance matrix Σ. Therefore, to induce a quasi-Bayes posterior we endow (P, Σ) with a prior. As discussed in Section 2.2, we use a Dirichlet process DP_α prior for the mixing distribution P. We then choose an independent prior G for the covariance matrix Σ. The choice of prior G is flexible, with some leading to more straightforward computation in higher dimensions. General conditions on G are specified in Section 4.1. In particular, an Inverse-Wishart prior is always permitted.

We define, as discussed in Section 2.2, a prior on Gaussian mixtures given by

$$\phi_{P,\Sigma}(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x - z) dP(z) \ , \quad (P, \Sigma) \sim \nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G.$$

Let $\mathscr{Z}_n$ denote the observed data. If $\varphi_{P,\Sigma}$ denotes the characteristic function of a Gaussian

[10] Under standard regularity conditions, the usual statistical guarantee (see Lemma 1) is

$$\sup_{\|t\|_\infty \leq T} |\widehat{\varphi}_Y(t) - \varphi_Y(t)| = O_{\mathbb{P}}(n^{-1/2}\sqrt{\log T}).$$

mixture $\phi_{P,\Sigma}$, we define the deconvolution quasi-Bayes posterior by

$$\nu_{\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{14}$$

As the quasi-Bayes posterior is constructed using $\widehat{\varphi}_Y$ and $\widehat{\varphi}_\epsilon$, the conditioning on $\mathcal{Z}_n$ is used to denote that the quasi-posterior is defined conditionally on the observed data. Our construction defines the posterior directly on characteristic space. However, as there is a one-to-one relation between $\phi_{P,\Sigma}$ and $\varphi_{P,\Sigma}$, without loss of generality we view $\nu_{\alpha,G}(.\,|\,\mathcal{Z}_n)$ as a quasi-Bayes posterior over distributions and their characteristic functions.

The quasi-Bayes posterior uses the identifying restrictions in (8) as a quasi-likelihood for the model. Intuitively, we expect samples from the posterior to concentrate on Gaussian mixtures that approximately satisfy the model's identifying restrictions. As the latent distribution X uniquely satisfies the restrictions, these Gaussian mixture samples are expected to concentrate around the distribution of X. This intuition is formalized in Section 4.2.

As a formal estimator of the latent distribution, we use the expected density under the posterior distribution in (14). Specifically, we denote the posterior mean by

$$\mathbb{E}_{\nu_{\alpha,G}|\mathcal{Z}_n}(f) = \int f d\nu_{\alpha,G}(f|\mathcal{Z}_n). \tag{15}$$

Remark 1 (On Alternative Priors)**.** Our priors are, by construction, supported on infinite Gaussian mixtures. In theory, alternative priors such as logistic Gaussian process priors (Lenk, 1991) or log-sieve based priors (Ghosal et al., 2000) could be used as well. Unfortunately, realizations from these priors do not admit a closed form characteristic function. This considerably complicates evaluation of the likelihood $L(.)$ in (12). Motivated by computational tractability, we focus on Gaussian mixtures as they admit a simple closed form characteristic function.

Remark 2 (On Quasi vs Pure Bayes)**.** Pure Bayesian approaches to deconvolution have been studied in the literature (Donnet et al., 2018; Rousseau and Scricciolo, 2023). In these frameworks, it is assumed that the density f_ϵ is known. From the model in (7), the density of Y is given by the convolution $f_Y(y) = f_X \star f_\epsilon(y) = \int_{\mathbb{R}^d} f_X(y-t)f_\epsilon(t)dt$. It follows that the pure-Bayes posterior distribution under a $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ prior is given by

$$\nu(\phi_{P,\Sigma}|\mathcal{Z}_n) = \frac{\prod_{i=1}^n \int_{\mathbb{R}^d} \phi_{P,\Sigma}(Y_i - t)f_\epsilon(t)dt}{\int \prod_{i=1}^n \int_{\mathbb{R}^d} \phi_{P,\Sigma}(Y_i - t)f_\epsilon(t)dt d\nu_{\alpha,G}(P,\Sigma)}. \tag{16}$$

In particular, even if the density f_ϵ is known, the convolution structure in the posterior likelihood makes it very challenging to compute in higher dimensions. By contrast, quasi-Bayes uses a loss function that is based on transforming the convolution structure $f_X \star f_\epsilon(t)$ into simple multiplication $\varphi_X(t) \times \varphi_\epsilon(t)$ in the Fourier domain.[11] This remains true in the more complicated setups studied in Section 3.2 and 3.3.[12]

Remark 3 (Implementation)**.** As our quasi-likelihoods are smooth functions of (P, Σ), gradient based Markov chain Monte Carlo (MCMC) methods can be used to construct a random sample from the quasi-Bayes posterior. In our implementation, we use the No-U-Turn Sampler (NUTS) version of Hamiltonian Monte Carlo (Hoffman, Gelman et al., 2014).

3.2 Models with Repeated Measurements

In this section, we continue our investigation into quasi-Bayes representations for latent variable models. In Section 3.1, we considered the setting where a researcher has auxiliary information on the error distribution that contaminates the latent variable. Specifically, we assumed the availability of a random sample from the error distribution. In practice, this specific form of auxiliary information may not always be available. Here we examine the more common situation where a researcher observes several noisy measurements of the latent variable X. To fix ideas, suppose we observe (Y_1, Y_2) from the model:

$$\begin{aligned} Y_1 &= X + \epsilon_1 \ , \ \mathbb{E}[\epsilon_1] = 0, \\ Y_2 &= X + \epsilon_2 \ , \ \mathbb{E}[\epsilon_2] = 0. \end{aligned} \tag{17}$$

Here, $X \in \mathbb{R}^d$ is a latent random vector whose distribution is of interest and (ϵ_1, ϵ_2) are unobserved errors. As a baseline example, X may denote latent child ability and (Y_1, Y_2) measurements of test scores in different subjects (e.g. math, english). Similar to the preceding section, it is not necessary for Y_1 and Y_2 to have the same number of observations n. Nonetheless, for simplicity and notational ease, we continue under this setting.

In the special case where $(\epsilon_1, \epsilon_2, X)$ are mutually independent and ϵ_2 is a symmetric distribution around zero, the model in (17) can be reduced to the classical measurement error model in (7) by differencing out the error. Specifically, consider the observed quantities

$$\left(\frac{Y_1 + Y_2}{2}\right) = X + \left(\frac{\epsilon_1 + \epsilon_2}{2}\right), \tag{18}$$

$$\left(\frac{Y_1 - Y_2}{2}\right) = \left(\frac{\epsilon_1 - \epsilon_2}{2}\right). \tag{19}$$

The main idea being that, under the appropriate symmetry and independence conditions, the observations from (19) can be used as an auxiliary sample for the error in (18). As such, the analysis in the preceding section and the quasi-Bayes posterior in (14) can be directly applied to this setting. This analysis, while convenient to implement, may not always be applicable.

[11] In a sense, this is similar to the idea behind Fast Fourier Transform (FFT) methods for convolution (Cooley et al., 1967). There, convolution is replaced with simple multiplication in the Fourier domain and the convolution is obtained by inversion.

[12] We could not find any references that study nonparametric pure Bayesian approaches to repeated measurements (Section 3.2) and multi-factor models (Section 3.3).

In particular, weaker restrictions on the joint distribution of $(X, \epsilon_1, \epsilon_2)$ may violate either the mutual independence of $(X_1, \epsilon_1, \epsilon_2)$ or the requirement of symmetrically distributed errors.

Under the weaker restriction that the errors satisfy $\mathbb{E}[\epsilon_1|X, \epsilon_2] = 0$ and $\epsilon_2 \perp X$, it is known (Li and Vuong, 1998; Cunha et al., 2010) that the latent distribution of X can be identified through the representation[13]

$$f_X(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-\mathbf{i}t'x} \exp\left(\int_{\mathcal{P}_t} \frac{\mathbb{E}[\mathbf{i}Y_1 e^{\mathbf{i}\zeta' Y_2}]}{\mathbb{E}[e^{\mathbf{i}\zeta' Y_2}]} d\zeta \right) dt. \tag{20}$$

Here, $\int_{\mathcal{P}_t}$ is the path integral over the straight line segment joining the origin to $t \in \mathbb{R}^d$. The closed form representation in (20) is convenient for theoretical analysis, but feasible analogs of it are very challenging to implement. This is especially true in higher dimensions where both the path and outer integral are difficult to evaluate precisely. This is further complicated by the fact that the empirical analog of the ratio appearing inside the path integral is volatile and highly sensitive to the choice of $t \in \mathbb{R}^d$. Intuitively, bad choices of t lead to situations where $|\mathbb{E}[e^{\mathbf{i}t'Y_2}]|$ is significantly smaller than sampling uncertainty $|(\mathbb{E}_n - \mathbb{E})[e^{\mathbf{i}t'Y_2}]|$. This makes the empirical analog of the ratio a highly volatile function of $t \in \mathbb{R}^d$.

Our starting point towards a general quasi-Bayes representation of model (17) is to consider characteristic function based implications of the representation in (20). By applying the Fourier transform to both sides of (20) and taking the logarithm, we obtain

$$\log \varphi_X(t) = \int_{\mathcal{P}_t} \frac{\mathbb{E}[\mathbf{i}Y_1 e^{\mathbf{i}\zeta' Y_2}]}{\mathbb{E}[e^{\mathbf{i}\zeta' Y_2}]} d\zeta \qquad \forall\, t \in \mathbb{R}^d. \tag{21}$$

By differentiating both sides of (21) and rearranging, we obtain

$$\varphi_{Y_2}(t) \nabla \log \varphi_X(t) = \mathbb{E}[\mathbf{i}Y_1 e^{\mathbf{i}\zeta' Y_2}] \qquad \forall\, t \in \mathbb{R}^d. \tag{22}$$

We view (22) as a collection of identifying restrictions on the model. Given the observed data $\{(Y_{1,i}, Y_{2,i})\}_{i=1}^n$, we estimate the empirical counterpart to (22) using

$$\widehat{\varphi}_{Y_2}(t) = \mathbb{E}_n[e^{\mathbf{i}t'Y_2}] = \frac{1}{n}\sum_{j=1}^n e^{\mathbf{i}t'Y_{2,j}} \ ,$$

$$\widehat{\varphi}_{Y_1,Y_2}(t) = \mathbb{E}_n[\mathbf{i}Y_1 e^{\mathbf{i}\zeta' Y_2}] = \frac{1}{n}\sum_{j=1}^n \mathbf{i}Y_{1,j} e^{\mathbf{i}t'Y_{2,j}}.$$

If θ represents a distribution with characteristic function φ_θ, the identifying restrictions in (22) suggest a quasi-likelihood of the form

$$L(\theta) = -\|\widehat{\varphi}_{Y_2} \nabla \log \varphi_\theta - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)}. \tag{23}$$

[13]The representation in (20) belongs to a family of Koltarski type (Kotlarski, 1967; Evdokimov and White, 2012) identities that often appear in the identification of econometric models with measurement error. For examples in auction based frameworks, see (Krasnokutskaya, 2011; Haile and Kitamura, 2019).

Up to sampling uncertainty, the quasi-likelihood in (23) is based on the objective function

$$Q(\theta) = -\|\varphi_{Y_2}(\nabla \log \varphi_\theta - \nabla \log \varphi_X)\|^2_{\mathbb{B}(T)}. \tag{24}$$

A natural question, which we study in detail in Section 4.3, is whether the quasi-likelihood in (23) is sufficient to identify the true latent distribution. From the objective function in (24), this can be rephrased as to whether gradient based information is sufficient to determine the latent characteristic function. We show that this is indeed the case in dimension $d = 1$ but that for higher dimensions, the quasi-likelihood in (23) must be augmented with further information that accounts for the behavior of f on the boundary $\partial\mathbb{B}(T)$ of $\mathbb{B}(T)$.[14] Intuitively, the magnitude $\|f\|_{\mathbb{B}(T)}$ of a function $f : \mathbb{B}(T) \to \mathbb{C}$ is determined by the magnitude of its gradient $\|\nabla f\|_{\mathbb{B}(T)}$ and the behavior of $f|_{\partial\mathbb{B}(T)}$, the function obtained by restricting f to $\partial\mathbb{B}(T)$.[15]

Given a DP_α prior and an independent prior G on covariance matrices, we consider (as defined in Section 2.2) a prior on Gaussian mixtures given by

$$\phi_{P,\Sigma}(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x - z)dP(z)\ , \quad (P, \Sigma) \sim \nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G.$$

If $\varphi_{P,\Sigma}$ denotes the characteristic function of a Gaussian mixture $\phi_{P,\Sigma}$, we define the (preliminary) repeated measurements quasi-Bayes posterior by

$$\nu_{\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla \log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla \log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{25}$$

We show in Section 4.2 that the quasi-Bayes posterior in (25) is valid for dimension $d = 1$. For higher dimensions, as discussed above, we introduce an adjustment to account for the boundary $\partial\mathbb{B}(T)$. For a function $g : \mathbb{B}(T) \to \mathbb{C}^d$, define the metric

$$\|g\|^2_{\partial,\,\mathbb{B}(T)} = \int_{\mathbb{B}(T)} \|g(t)\|^2 dt + \int_{\partial\mathbb{B}(T)} \int_0^1 \|g(tz)\|^2 dt d\mathcal{H}^{d-1}(z). \tag{26}$$

Here, $\mathcal{H}^{d-1}(.)$ denotes the $d-1$ dimensional Hausdorff measure.[16] Intuitively, for a function $h : \mathbb{B}(T) \to \mathbb{R}$ with $h(0) = 0$, we can write $h(z) = \int_0^1 \langle \nabla h(tz), z\rangle dt$ for every $z \in \partial\mathbb{B}(T)$. The second term of (26) therefore accounts for the boundary by specifically controlling the gradient along the straight line segment from the origin.

[14]The boundary $\partial\Omega$ of a domain $\Omega \subset \mathbb{R}^d$ is the remainder of the closure of Ω after removing its interior. In the case of a ball $\mathbb{B}(T) = \{t \in \mathbb{R}^d : \|t\| \leq T\}$, this is just $\partial\mathbb{B}(T) = \{t \in \mathbb{R}^d : \|t\| = T\}$.

[15]The necessity of considering more than just the gradient is clear as adding a large constant to f modifies $\|f\|_{\mathbb{B}(T)}$ but leaves $\|\nabla f\|_{\mathbb{B}(T)}$ unchanged. In dimension $d = 1$, the condition $f(0) = 0$ and knowledge of $\|\nabla f\|_{\mathbb{B}(T)}$ is sufficient to determine the behavior of $f|_{\partial\mathbb{B}(T)}$. Unfortunately, this does not hold in higher dimensions.

[16]Up to a known normalization constant c, integration w.r.t the Hausdorff measure is given by

$$\int_{\partial\mathbb{B}(T)} h(z)d\mathcal{H}^{d-1}(z) = c\mathbb{E}[h(V)] \quad \text{where} \quad V = T\frac{Z}{\|Z\|}\ , \quad Z \sim \mathcal{N}(0, I_d)$$

for every Borel function $h : \partial\mathbb{B}(T) \to \mathbb{R}$.

For $d > 1$, we define the repeated measurements quasi-Bayes posterior by

$$\nu_{\partial,\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{27}$$

Remark 4 (On Multiple Identifying Restrictions)**.** If $(\epsilon_1, \epsilon_2, X)$ are mutually independent in model (17), the roles of Y_1 and Y_2 can be interchanged in (22). In particular, this provides us with two sets of identifying restrictions. It is straightforward to modify the quasi-Bayes posteriors in (25) and (27) to include both sets of restrictions. One way to achieve this is to use the quasi-likelihood

$$L^*(\theta) = -(\|\widehat{\varphi}_{Y_2}\nabla\log\varphi_\theta - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)} + \|\widehat{\varphi}_{Y_1}\nabla\log\varphi_\theta - \widehat{\varphi}_{Y_2,Y_1}\|^2_{\mathbb{B}(T)}).$$

Remark 5 (On the choice of Quasi-Likelihood)**.** In practice, the choice of quasi-likelihood used to induce a valid quasi-Bayes posterior is not unique. For example, one could base a quasi-likelihood on the identity (21) directly. While such a choice is more straightforward for theoretical analysis, implementation requires evaluation of the path integral. As discussed previously, this is further complicated by the fact that the empirical analog of the ratio appearing inside the path integral is volatile and highly sensitive to the choice of $t \in \mathbb{R}^d$.

3.3 Multi-Factor Models

In this section, we study the linear multi-factor model

$$\mathbf{Y} = \mathbf{A}\mathbf{X} \tag{28}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_L)' \in \mathbb{R}^L$ is a vector of L measurements, $\mathbf{X} = (X_1, \ldots, X_K)'$ is a vector of K latent and mutually independent factors and $\mathbf{A}$ is a known $L \times K$ matrix of factor loadings. We take for granted that measurements are demeaned so that $\mathbb{E}[\mathbf{X}] = \mathbf{0}$. The model in (28) was initially introduced by Bonhomme and Robin (2010).[17]

Remark 6 (Factor Loadings)**.** In most cases, $\mathbf{A}$ is not fully known but can be consistently estimated through means and covariances of the observed $\mathbf{Y} = (Y_1, \ldots, Y_L)$. In particular, the discrepancy from replacing $\mathbf{A}$ with an estimate $\widehat{\mathbf{A}}$ has stochastic order $O_{\mathbb{P}}(n^{-1/2})$. As the nonparametric rates of convergence are not faster than $O(n^{-1/2})$, all our main results go through with $\widehat{\mathbf{A}}$ replacing $\mathbf{A}$. For expositional simplicity, we continue with the case of known $\mathbf{A}$.

Remark 7 (Repeated Measurements)**.** The repeated measurements model in (17) can be realized as a special case of the multi-factor model in (28). Indeed, consider univariate repeated measurements (Y_1, Y_2) of a latent signal X with measurement errors (ϵ_1, ϵ_2). Then (28) holds

[17] To ease the exposition, this section closely follows the notation introduced in Bonhomme and Robin (2010).

with

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}}_{\mathbf{A}} \underbrace{\begin{pmatrix} X \\ \epsilon_1 \\ \epsilon_2 \end{pmatrix}}_{\mathbf{X}} \tag{29}$$

We are interested in recovering the distribution of $\mathbf{X}$ from a random sample of $\mathbf{Y}$. Denote the columns of $\mathbf{A}$ by $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_K)$. In Bonhomme and Robin (2010), it is shown that the identifying restrictions of the model in (28) may be expressed as

$$\nabla\nabla' \log \varphi_{\mathbf{Y}}(t) = \sum_{k=1}^{K} \mathbf{A}_k \mathbf{A}_k' (\log \varphi_{X_k})''(t'\mathbf{A}_k) \qquad \forall\, t \in \mathbb{R}^L \;, \tag{30}$$

where $\nabla\nabla' \log \varphi_{\mathbf{Y}}(t)$ denotes the Hessian of the map $t \to \log \varphi_{\mathbf{Y}}(t)$.

To simplify the exposition further, we introduce some additional notation. Let $\mathcal{V}_{\mathbf{Y}}(t)$ denote the vector of upper triangular elements of $\nabla\nabla' \log \varphi_{\mathbf{Y}}(t)$. Let $\mathcal{V}_{\mathbf{X}}(t)$ denote the vector with elements $\{(\log \varphi_{X_i})''(t'\mathbf{A}_i)\}_{i=1}^{K}$. Let $V(\mathbf{A}_k)$ denote the vector of upper triangular elements of $\mathbf{A}_k \mathbf{A}_k'$ and $\mathbf{Q} = [V(\mathbf{A}_1), \ldots, V(\mathbf{A}_K)]$ the matrix with columns composed of those vectors.

As the matrices in (30) are symmetric, the identifying restrictions may be expressed as

$$\mathcal{V}_{\mathbf{Y}}(t) = \mathbf{Q}\mathcal{V}_{\mathbf{X}}(t) \quad \forall\, t \in \mathbb{R}^L.$$

Equivalently, if $\mathbf{Q}$ has full column rank, the identifying restrictions are

$$\mathbf{Q}^*\mathcal{V}_{\mathbf{Y}}(t) - \mathcal{V}_{\mathbf{X}}(t) = \mathbf{0} \quad \forall\, t \in \mathbb{R}^L \tag{31}$$

where $\mathbf{Q}^* = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$. The empirical analog to $\mathcal{V}_{\mathbf{Y}}(t)$ is given below in (33). As it involves the squared reciprocal of the characteristic function, the estimand is volatile and highly sensitive to the choice of $t \in \mathbb{R}^L$.[18] Similar to our strategy in the preceding section, we focus instead on a smoothed version of the identifying restriction. If the characteristic function of $\mathbf{Y}$ is non vanishing, the identifying restrictions can equivalently be expressed as

$$\varphi_{\mathbf{Y}}^2(t)[\mathbf{Q}^*\mathcal{V}_{\mathbf{Y}}(t) - \mathcal{V}_{\mathbf{X}}(t)] = \mathbf{0} \quad \forall\, t \in \mathbb{R}^L$$

Denote by $\mathbf{Q}_k^*$, the k^{th} row of $\mathbf{Q}_k^*$. From the definition of $\mathcal{V}_{\mathbf{X}}(t)$, it follows that the identifying restrictions for the k^{th} latent factor X_k reduce to

$$\varphi_{\mathbf{Y}}^2(t)[\mathbf{Q}_k^*\mathcal{V}_{\mathbf{Y}}(t) - (\log \varphi_{X_k})''(t'\mathbf{A}_k)] = 0 \quad \forall\, t \in \mathbb{R}^L. \tag{32}$$

[18]Intuitively, bad choices of t lead to situations where $|\varphi_{\mathbf{Y}}(t)|$ is significantly smaller than sampling uncertainty $|(\mathbb{E}_n - \mathbb{E})e^{\mathbf{i}\mathbf{t}'\mathbf{Y}}|$. This makes the function in (33) a highly volatile function of $t \in \mathbb{R}^d$. Reweighting by $\widehat{\varphi}_{\mathbf{Y}}^2$ leads to a more stable quasi-likelihood.

Our starting point for a quasi-Bayes framework is to view (32) as a collection of complex valued restrictions on the model. Given the observed data $\{\mathbf{Y}_i\}_{i=1}^n$, we estimate the empirical counterpart to (32) using

$$\widehat{\varphi}_{\mathbf{Y}} = \mathbb{E}_n[e^{\mathbf{i}t'\mathbf{Y}}] = \frac{1}{n}\sum_{j=1}^{n} e^{\mathbf{i}t'\mathbf{Y}_j} \ ,$$
$$[\nabla\nabla' \log \widehat{\varphi}_{\mathbf{Y}}(t)]_{l,k} = -\frac{\mathbb{E}_n[Y_l Y_k e^{\mathbf{i}t'\mathbf{Y}}]}{\widehat{\varphi}_{\mathbf{Y}}(t)} + \frac{\mathbb{E}_n[Y_l e^{\mathbf{i}t'\mathbf{Y}}]\mathbb{E}_n[Y_k e^{\mathbf{i}t'\mathbf{Y}}]}{\widehat{\varphi}^2_{\mathbf{Y}}(t)}. \tag{33}$$

Using these estimands, we obtain empirical analogs $\widehat{\varphi}_{\mathbf{Y}}$ and $\widehat{\mathcal{V}}_{\mathbf{Y}}(t)$.

If θ represents a distribution with characteristic function φ_θ, the identifying restrictions in (32) suggest a quasi-likelihood of the form

$$L(\theta) = -\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^* \widehat{\mathcal{V}}_{\mathbf{Y}} - (\log \varphi_\theta)''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T)} \tag{34}$$

Up to sampling uncertainty, the quasi-likelihood in (34) is based on the objective function

$$Q(\theta) = -\|\varphi^2_{\mathbf{Y}}(t)[(\log\varphi_\theta)''(t'\mathbf{A}_k) - (\log\varphi_{X_k})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T)}. \tag{35}$$

A natural question is whether the quasi-likelihood in (34) is sufficient to identify the true latent distribution. From the objective function in (35), this can be rephrased as to whether gradient based information is sufficient to determine the latent characteristic function. Intuitively, this follows if $\varphi_\theta = \varphi_{X_k}$ is the unique solution to the second order differential equation

$$(\log\varphi_\theta)''(t) = (\log\varphi_{X_k})''(t). \tag{36}$$

As all characteristic functions satisfy the boundary condition $\log\varphi_\theta(0) = \log(1) = 0$, (36) has a unique solution if θ satisfies a second boundary condition. As the latent factor X_k is demeaned, a natural boundary condition is given by $(\log\varphi_\theta)'(0) = (\log\varphi_{X_k})'(0) = \mathbf{i}\mathbb{E}[X_k] = 0$. This can be achieved by restricting θ to be a mean zero distribution.

Given a DP_α prior and an independent prior G on covariance matrices, we consider (as defined in Section 2.2) a prior on Gaussian mixtures given by

$$\phi_{P,\Sigma}(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x-z)dP(z) \ , \quad (P,\Sigma) \sim \nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G.$$

If $\varphi_{P,\Sigma}$ denotes the characteristic function of a Gaussian mixture $\phi_{P,\Sigma}$, we define the multi-factor quasi-Bayes posterior for latent factor X_k by

$$\nu_{\alpha,G}(\varphi_{P,\sigma^2} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\sigma^2)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\sigma^2)}. \tag{37}$$

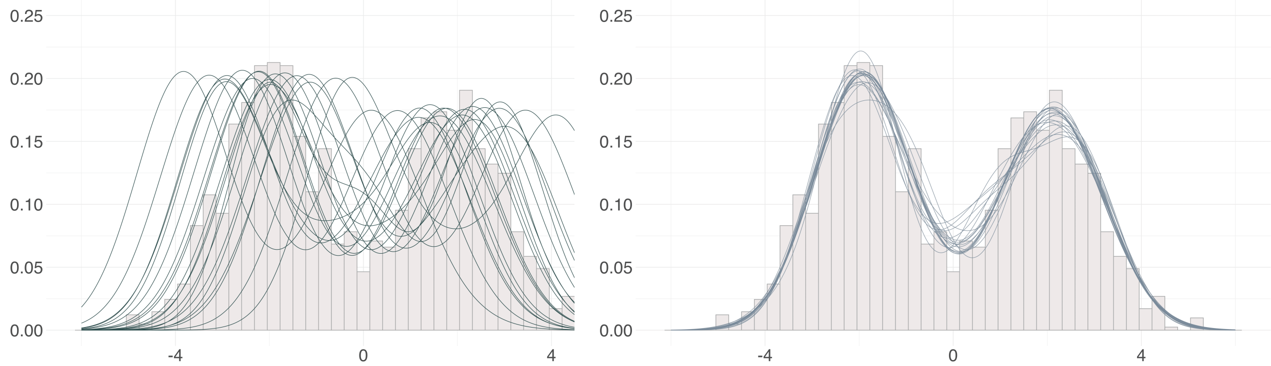
Without further modifications, the quasi-Bayes posterior in (37) does not account for the mean zero boundary condition discussed above. To enforce this, we propose the demeaned quasi-posterior distribution obtained from demeaning samples from (37). That is,

$$\overline{\nu}_{\alpha,G}(.\mid \mathcal{Z}_n) \sim Z - \mathbb{E}[Z] \qquad \text{where} \qquad Z \sim \nu_{\alpha,G}(.|\mathcal{Z}_n). \tag{38}$$

Remark 8 (On Joint Posteriors)**.** From (32), the identifying restrictions can be separated for each latent factor and this leads to a simple univariate quasi-Bayes posterior. However, as the latent factors $(X_1, \dots, X_K)$ are mutually independent, the analysis is unchanged if we model them jointly using independent priors. Let $\mathbf{P} = (P_1, \dots, P_K)$ denote a vector of mixing distributions and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_K^2)$ a vector of variance parameters. Let $\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2}(t)$ denote the vector with elements $\{(\log \varphi_{P_i,\sigma_i^2})''(t'\mathbf{A}_i)\}_{i=1}^K$. With some abuse of notation, we denote the vector of individual characteristic functions by $\varphi_{\mathbf{P},\boldsymbol{\sigma}^2} = (\varphi_{P_1,\sigma_1^2}, \dots, \varphi_{P_K,\sigma_K^2})$. The joint quasi-Bayes posterior induced from the multi-factor model in (28) is then given by

$$\begin{aligned} \nu_{\alpha,G}(\varphi_{\mathbf{P},\boldsymbol{\sigma}^2} \mid \mathcal{Z}_n) &= \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\widehat{\mathcal{V}}_{\mathbf{Y}} - \mathbf{Q}\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2})\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(\mathbf{P},\boldsymbol{\sigma}^2)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\widehat{\mathcal{V}}_{\mathbf{Y}} - \mathbf{Q}\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2})\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(\mathbf{P},\boldsymbol{\sigma}^2)} \\ \overline{\nu}_{\alpha,G}(.\mid \mathcal{Z}_n) &\sim Z - \mathbb{E}[Z] \ , \ Z \sim \nu_{\alpha,G}(.|\mathcal{Z}_n). \end{aligned} \tag{39}$$

Here G is the common univariate prior on each variance parameter $\{\sigma_i^2\}_{i=1}^K$. The prior $\nu_{\alpha,G}$ is the one obtained by placing mutually independent $(\mathrm{DP}_\alpha \otimes G)$ priors on each of $(P_i, \sigma_i^2)_{i=1}^K$



(a) Posterior Samples, $\nu_{\alpha,G}$ (b) Posterior samples, $\overline{\nu}_{\alpha,G}$

Figure 2: Repeated Measurements (see Remark 7) multi-factor posterior with $X \sim 0.5\mathcal{N}(-2,1) + 0.5\mathcal{N}(2,1)$ and $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0,1)$. Posterior draws for the latent distribution of X relative to true latent histogram.

As Figure 2 illustrates, in this specific example, the quasi-Bayes posterior $\nu_{\alpha,G}$ appears to concentrate on bimodal Gaussians that differ only by shifts in central tendency (mean). This is consistent with our discussion above in that they concentrate around the identified set, i.e solutions to the second order differential equation in (36). These solutions only identify the distribution up to the location parameter. The demeaned quasi-Bayes posterior $\overline{\nu}_{\alpha,G}$ produces posterior samples with a fixed mean at zero. The boundary conditions for the differential equation are satisfied and posterior samples concentrate around the true mean-zero latent distribution.

4 Theory

In this section, we develop the quasi-Bayes limit theory for the models in Section 3. First, we outline conditions on the prior in Section 4.1. In Sections 4.2-4.4, we present the main results for each model.

4.1 Priors

The models and the associated quasi-Bayes posteriors in Section 3 are all based on a common family of prior distributions. In all cases, we model latent distributions using a prior on Gaussian mixtures given by

$$\phi_{P,\Sigma}(x) = \int_{\mathbb{R}^d} \phi_\Sigma(x-z)dP(z) \ , \qquad (P,\Sigma) \sim \mathrm{DP}_\alpha \otimes G. \tag{40}$$

We first state and then discuss the main assumptions that we impose on the Dirichlet process DP_α and covariance prior G.

Assumption 1 (Dirichlet Prior)**.** The Dirichlet process DP_α, in the sense of Definition 2, arises from a Gaussian base measure α. That is, $\alpha = \mathcal{N}(\mu^*, \Sigma^*)$ for some mean vector $\mu^* \in \mathbb{R}^d$ and positive definite matrix $\Sigma^* \in \mathbf{S}_+^d$.

Assumption 2 (Covariance Prior)**.** (i) G is a probability measure with support contained in the space of positive semi-definite matrices on $\mathbb{R}^{d\times d}$. (ii) There exists $\kappa \in (0,1], v_1 \geq 0, v_2 \geq 0$ and universal constants $C, C', D, D' > 0$ that satisfy

$$\begin{aligned} &(a) \quad G(\Sigma : \lambda_d(\Sigma^{-1}) > t_d) \leq C \exp\bigl(-C' t_d^\kappa\bigr) \ , \\ &(b) \quad G\biggl(\Sigma : \bigcap_{1\leq j\leq d} \Bigl\{t_j \leq \lambda_j(\Sigma^{-1}) \leq t_j(1+\delta)\Bigr\}\biggr) \geq D t_1^{v_1} \delta^{v_2} \exp\bigl(-D' t_d^\kappa\bigr) \end{aligned}$$

for every $\delta \in (0,1)$ and $0 < t_1 \leq t_2 \leq \ldots \leq t_d < \infty$.

We note that, as the researcher chooses the prior, Assumption 1 and 2 can always be satisfied. For the Dirichlet process prior, by varying (μ^*, Σ^*), a researcher can impose varying degrees of prior information. In practice, setting $\mu^* = 0$ and $\Sigma^* = CI_d$ for a large $C > 0$ leads to a diffuse noninformative prior.

Assumption 2 is frequently used in the literature on density estimation (Shen et al., 2013; Ghosal and Van der Vaart, 2017). It allows for a general class of covariance priors. In dimension $d = 1$, it holds with $\kappa = 1/2$ if G is the distribution of the square of an inverse-gamma distribution. In dimension $d > 1$, it holds with $\kappa = 1$ if G is the Inverse-Wishart distribution. Various other choices may be used as well, with some leading to more straightforward computation in higher dimensions. For instance, we can always write $\Sigma = DCD$ where C is a correlation matrix and $D = \sqrt{\mathrm{Diag}(\Sigma)}$ is the scale matrix. In higher dimensions, it is often computationally simpler to place priors on C and D separately. For D, the diagonal can consist of independent univariate priors (e.g. inverse-gamma). For the correlation matrix C, a common choice is the class of Lewandowski-Kurowicka-Joe (LKJ) priors (Lewandowski et al., 2009).

Remark 9 (On Empirical Bayes Priors)**.** For the base measure α in Assumption 1, one may choose (μ^*, Σ^*) using an empirical Bayes approach. All our main results go through with a data dependent prior $\widehat{\alpha} = \mathcal{N}(\widehat{\mu}, \widehat{\Sigma})$, provided that asymptotically with probability approaching one, $\|\widehat{\mu}\|$ is bounded and the eigenvalues of $\widehat{\Sigma}$ are bounded away from zero and infinity. For example, in the classical measurement error model (7), one can choose (μ^*, Σ^*) to be the empirical mean and covariance matrix of the observed vector Y.

4.2 Main Results (Multivariate Deconvolution)

In this section, we present the main results for the deconvolution-based quasi-Bayes posterior of Section 3.1. Recall that in this case, we observe a sample $\{Y_i\}_{i=1}^n$ and an auxiliary sample $\{\epsilon_i\}_{i=1}^m$, from the model in (7). The main conditions that we impose on the observations is summarized in the following condition.

Condition 4.1 (Data)**.** *(i)* $(Y_i)_{i=1}^n \in \mathbb{R}^d$ is a sequence of independent and identically distributed random vectors. *(ii)* $(\epsilon_1, \ldots, \epsilon_m) \overset{i.i.d}{\sim} F_\epsilon$ where F_ϵ is the distribution of the error ϵ in model (7). *(iii)* The sample size of the auxiliary sample grows at rate $m = m_n \asymp n$. *(iv)* The distributions have finite second moments: $\mathbb{E}(\|Y\|^2) < \infty$ and $\mathbb{E}(\|\epsilon\|^2) < \infty$.

The rate requirement on the auxiliary sample size in Condition 4.1*(iii)* is made for simplicity and could be relaxed further.[19] As is common in the literature, we characterize the ill-posedness in the model through the decay rate of the characteristic function of the error $\varphi_\epsilon(.)$. This is summarized in the following definition.

Definition 3 (Inverse Regime)**.** We say a model is either mildly or severely ill-posed if

$$\inf_{t \in \mathbb{R}^d : \|t\|_\infty \leq T} |\varphi_\epsilon(t)| \asymp \begin{cases} T^{-\zeta} & \text{mildly ill-posed} \\ \exp\big(-RT^\zeta\big) & \text{severely ill-posed} \end{cases}$$

for some $R, \zeta \geq 0$.

Recall that the deconvolution-based quasi-Bayes posterior is given by

$$\nu_{\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{41}$$

Intuitively, our general strategy to study (41) proceeds as follows. The quantity $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T)}$ represents a smoothed version of the distance $\|\phi_{P,\Sigma} - f_X\|^2_{L^2}$, where f_X denotes the latent density of X. We first obtain contraction rates with respect to this smoothed, weaker metric. Its

[19] Lower sample sizes only influence the analysis through weaker statistical guarantees on the estimated characteristic function $\widehat{\varphi}_\epsilon(.)$. It is straightforward to incorporate this into the analysis.

implications towards distance in the stronger metric $\|\phi_{P,\Sigma} - f_X\|_{L^2}^2$ is then determined by the complexity of $\mathbb{B}(T)$ and the model ill-posedness.

Before considering the general case, we first study the case where the true latent density f_X admits a representation as a (possibly infinite) Gaussian mixture:

$$f_X(x) = \phi_{\Sigma_0} \star F_0(x) = \int_{\mathbb{R}^d} \phi_{\Sigma_0}(x-z)dF_0(z). \tag{42}$$

Here, $\Sigma_0 \in \mathbf{S}_+^d$ is a positive definite matrix and F_0 is a mixing distribution. Note that, we do not assume knowledge of (Σ_0, F_0), only that the representation in (42) is true for some (Σ_0, F_0). Importantly, if F_0 is continuously distributed or has unbounded support, the specification in (42) allows for a Gaussian mixture with infinite components. The model specification in (42) is commonly used to study Gaussian mixture based approaches to density estimation.[20]

Finite Gaussian mixture representations are widely used in empirical settings.[21] Hence, the model in (42) serves as a natural starting point. Intuitively, as the prior concentrates on Gaussian mixtures, the expectation is that the model should exhibit negligible bias in approximating a latent density with Gaussian mixture structure as in (42). We formalize this intuition below by showing that nearly parametric rates of convergence can be obtained. To that end, our main assumption on the Gaussian mixture specification is the following.

Condition 4.2 (Exact Gaussian Mixtures)**.** $f_X = \phi_{\Sigma_0} \star F_0$ where $\Sigma_0 \in \mathbb{R}^{d\times d}$ is a positive definite matrix and F_0 is a mixing distribution that satisfies $F_0(t \in \mathbb{R}^d : \|t\| > z) \leq C\exp(-C'z^\chi)$ for all sufficiently large z, where $C, C', \chi \geq 0$ are universal constants.

Condition (4.2) imposes an exponential tail on the mixing distribution F_0. Importantly, this covers all Gaussian mixtures that have modes contained in a compact subset of $\mathbb{R}^d$. As the mixing distribution may be continuously distributed, the corresponding Gaussian mixture may be infinite as well. The following result shows that, under suitable scaling of the covariance prior, the quasi-Bayes posterior contracts towards the true latent density f_X at a nearly parametric rate.

Theorem 1 (Rates with Exact Gaussian Mixtures)**.** *Suppose Conditions 4.1, 4.2 hold and the covariance prior is taken as $G_n \sim G/\sigma_n^2$ where G satisfies Assumption 2 and σ_n^2 is as specified below. Furthermore, suppose Σ_0 is in the support of G. Let $\kappa, \chi > 0$ be as in Assumption 2 and Condition 4.2, respectively.*

(a) Let $\lambda = \max\{\chi^{-1}(d+2) + d/2, d+1\}$. Suppose the model is mildly ill-posed with $\zeta \geq 0$ as in Definition 3. If $T_n \asymp \sqrt{\log n}\sqrt{\log\log n}$ and $\sigma_n^2 \asymp (\log n)^{-\lambda/\kappa}$, then there exists a

[20] Convergence rates (for univariate density estimation) for the model in (42), using a Gaussian mixture sieve maximum likelihood estimator, were initially studied by Genovese and Wasserman (2000). Contraction rates (and improved rates on the MLE) for pure Bayesian density estimation were subsequently obtained in Ghosal and Van Der Vaart (2001).

[21] As in, for example, Geweke and Keane (2000); Cunha et al. (2010); Attanasio et al. (2020a,b).

universal constant $C > 0$ such that

$$\nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > C\frac{(\log n)^{(\lambda+\zeta)/2}}{\sqrt{n}}(\log\log n)^{\zeta/2}\,\middle|\,\mathcal{Z}_n\right) = o_{\mathbb{P}}(1).$$

(*b*) *Let $\lambda = \max\{\chi^{-1}(d+2) + d/2, d+1, d/\zeta + 1/2\}$. Suppose the model is severely ill-posed with $\zeta \in (0,2]$ and $R > 0$ as in Definition 3. If $T_n = (c_0 \log n)^{1/\zeta}$ for any c_0 satisfying $c_0 R = \gamma < 1/2$ and σ_n^2 is defined by*

$$\sigma_n^2 \asymp \begin{cases} (\log n)^{2/\zeta - 1 - \lambda/\kappa}(\log\log n)^{-2/\kappa} & \zeta \in (0,2) \\ (\log n)^{-\lambda/\kappa} & \zeta = 2. \end{cases}$$

Then, there exists a universal constant $C > 0$ and $V \in (0, 1/2)$ such that

$$\begin{cases} \nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-1/2+\gamma}(\log n)^{\lambda/2}\,|\mathcal{Z}_n\right) = o_{\mathbb{P}}(1) & \zeta \in (0,2) \\ \nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-V}\,|\mathcal{Z}_n\right) = o_{\mathbb{P}}(1) & \zeta = 2. \end{cases}$$

Under a mildly ill-posed regime, the quasi-Bayes posterior attains nearly (up to a logarithmic factors) parametric rates of convergence. If the model is severely ill-posed at a rate ζ that is lower than Gaussian type ill-posedness, nearly parametric rates in the form $n^{-1/2+\epsilon}$ for any $\epsilon > 0$ can be obtained. In the extreme case where the model exhibits Gaussian ill-posedness, the model bias (which is determined by the decay rate of a Gaussian mixture characteristic function) has similar order to the variance. In this case, the rate is still polynomial but the exponent V depends on second order factors such as the constant R in Definition 3 and the eigenvalues of the matrix Σ_0 in representation (42). In the case of univariate frequentist kernel deconvolution, a similar finding can be found in Butucea and Tsybakov (2008a,b).

Remark 10 (On Covariance Scaling)**.** The strategy of scaling the covariance prior can be traced back to Ghosal and Van Der Vaart (2007) where it was employed for univariate pure Bayesian density estimation. This requirement was eventually removed in Shen et al. (2013). We conjecture that the scaling may be an artifact of the proof and can possibly be relaxed. As our quasi-Bayesian setup is quite different to a pure Bayesian model, the analysis in Shen et al. (2013) is not directly applicable here. We leave this investigation for future work. In our simulations and empirical illustrations, we do not apply any scaling to the covariance prior.

Remark 11 (Convergence in stronger metrics)**.** An interesting aspect of the quasi-Bayes framework is that all formal analysis takes place directly in the Fourier domain. As the Fourier transform $\mathcal{F}$ is a distance preserving isomorphism between L^2 spaces, $\|.\|_{L^2}$ contraction rates in the Fourier domain translate directly to $\|.\|_{L^2}$ rates for the density. However, this can also be used to derive contraction rates in stronger metrics such as $\|.\|_{L^\infty}$. In particular, as the Fourier transform satisfies $\|f\|_\infty \leq \|\mathcal{F}[f]\|_{L^1}$, we have the following variant of Theorem 1.

Corollary 1 (L^∞ Rates with Exact Gaussian Mixtures)**.** *Suppose Conditions 4.1, 4.2 hold and the covariance prior is taken as* $G_n \sim G/\sigma_n^2$ *where* G *satisfies Assumption 2 and* σ_n^2 *is as specified below. Furthermore, suppose* Σ_0 *is in the support of* G*. Let* $\kappa, \chi > 0$ *be as in Assumption 2 and Condition 4.2, respectively.*

(a) *Let* $\lambda = \max\{\chi^{-1}(d+2) + d/2, d+1\}$*. Suppose the model is mildly ill-posed with* $\zeta \geq 0$ *as in Definition 3. If* $T_n \asymp \sqrt{\log n}\sqrt{\log\log n}$ *and* $\sigma_n^2 \asymp (\log n)^{-\lambda/\kappa}$*, then there exists a universal constant* $C > 0$ *such that*

$$\nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^\infty} > C\frac{(\log n)^{(\lambda+d/2+\zeta)/2}}{\sqrt{n}}(\log\log n)^{(\zeta+d/2)/2}\,\middle|\,\mathcal{Z}_n\right) = o_{\mathbb{P}}(1).$$

(b) *Let* $\lambda = \max\{\chi^{-1}(d+2) + d/2, d+1, d/\zeta\}$*. Suppose the model is severely ill-posed with* $\zeta \in (0,2]$ *and* $R > 0$ *as in Definition 3. If* $T_n = (c_0 \log n)^{1/\zeta}$ *for any* c_0 *satisfying* $c_0 R = \gamma < 1/2$ *and* σ_n^2 *is defined by*

$$\sigma_n^2 \asymp \begin{cases} (\log n)^{2/\zeta - 1 - \lambda/\kappa}(\log\log n)^{-2/\kappa} & \zeta \in (0,2) \\ (\log n)^{-\lambda/\kappa} & \zeta = 2. \end{cases}$$

Then, there exists a universal constant $C > 0$ *and* $V \in (0, 1/2)$ *such that*

$$\begin{cases} \nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^\infty} > Cn^{-1/2+\gamma}(\log n)^{(\lambda+d/\zeta)/2}\,|\mathcal{Z}_n\right) = o_{\mathbb{P}}(1) & \zeta \in (0,2) \\ \nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^\infty} > Cn^{-V}\,|\mathcal{Z}_n\right) = o_{\mathbb{P}}(1) & \zeta = 2. \end{cases}$$

The L^2 and L^∞ rates in Theorem 1 and Corollary 1 appear to be novel even for the special case of density estimation.[22] As discussed in Remark 11, this is an interesting consequence of our characteristic function based quasi-Bayesian approach. While it may be possible for pure-Bayesian approaches to achieve similar results, the analysis to verify this appears to be significantly more complicated. For ease of exposition, all the remaining results in Section 4 are stated only for the $\|.\|_{L^2}$ metric. By analogous arguments to that of Corollary 1 and Remark 11, it is straightforward to extend these results to stronger metrics such as $\|.\|_{L^\infty}$.

As Gaussian mixtures can approximate any density $f_X \in L^2$ arbitrarily well, it is straightforward to modify the preceding results to obtain consistency for a general density.[23] We focus instead on the more difficult task of establishing a rate of convergence. To that end, consider the case where the true latent density f_X is not an exact Gaussian mixture and thus, the prior model exhibits non-negligible bias. In this model, the Gaussian mixture bias is closely related to the

[22] Rates of convergence with Dirichlet Process priors in L^1 or Hellinger risk can be found in Ghosal and Van Der Vaart (2001)

[23] Here, consistency means that

$$\nu_{\alpha,G_n}(\|f_X - \phi_{P,\Sigma}\|_{L^2} > \epsilon|\mathcal{Z}_n) = o_{\mathbb{P}}(1) \qquad \forall\ \epsilon > 0.$$

quantity

$$\mathcal{B}(\sigma) = \min_P \|f_X \star f_\epsilon - \phi_{P,\sigma^2 I} \star f_\epsilon\|_{L^2} \asymp \min_P \|\varphi_\epsilon(\varphi_X - \varphi_{P,\sigma^2 I})\|_{L^2}, \tag{43}$$

where $\sigma^2 I$ denotes the identity matrix with each diagonal entry σ^2 and the minimum is over all mixing distributions P.

Remark 12 (Bias Bounds)**.** For a conservative upper bound on the rate at which $\mathcal{B}(\sigma)$ tends to 0 as $\sigma \to 0$, observe that by picking $P = f_X$, we obtain

$$\|\varphi_\epsilon(\varphi_X - \varphi_{P,\sigma^2 I})\|_{L^2} = \|(e^{-\|t\|^2\sigma^2/2} - 1)\varphi_X(t)\varphi_\epsilon(t)\|_{L^2} \lessapprox \sigma^2.$$

The preceding bound follows from $\left|e^{-\|t\|^2\sigma^2/2} - 1\right| \leq \|t\|^2\sigma^2/2$. As such, the concluding estimate is valid provided that $\int_{\mathbb{R}^d} \|t\|^4 |\varphi_X(t)|^2 \, |\varphi_\epsilon(t)|^2 < \infty$ or equivalently $f_X \star f_\epsilon \in \mathbf{H}^2$. For $f_X \star f_\epsilon$ with smoothness order greater than 2, this argument does not provide better rates. In particular, with higher smoothness, the choice $P = f_X$ is not optimal in (43): one can typically do better by allowing $P = P_{X,\epsilon,\sigma}$ to depend implicitly on all the features $(f_\epsilon, f_X, \sigma)$.

We consider the following characterization of the Gaussian mixture bias.

Condition 4.3 (Gaussian Mixture Bias)**.** *(i)* $f_X \in \mathbf{H}^p(R)$ for some $p > 1/2$ and $R < \infty$. *(ii)* There exists universal constants $C, M < \infty$ and $\chi > 0$ such that for all $\sigma > 0$ sufficiently small, there exists a mixing distribution $F_\sigma = F_{X,\epsilon,\sigma}$ supported on the cube $I_\sigma = [-C(\log \sigma^{-1})^{1/\chi}, C(\log \sigma^{-1})^{1/\chi}]^d$ that satisfies $\|f_X \star f_\epsilon - \phi_{F_\sigma,\sigma^2 I} \star f_\epsilon\|_{L^2} \leq M\sigma^{p+\zeta}$.

Variants of Condition 4.3 are commonly used in density estimation (Shen et al., 2013; Ghosal and Van der Vaart, 2017). A stronger version of Condition 4.3 is also used in Donnet et al. (2018) within the context of univariate pure Bayes density deconvolution. For examples of explicit constructions of F_σ in Condition 4.3, see Shen et al. (2013) and Ghosal and Van der Vaart (2017) for density estimation and Rousseau and Scricciolo (2023) for deconvolution. We note that, at least in our quasi-Bayes setup, Condition 4.3 can be weakened further. In particular, it suffices that the approximation holds with respect to the weaker norm $\|.\|_{\mathbb{B}(T_n)}$ and for a sequence σ_n, where (σ_n^{-1}, T_n) grow at a suitable rate. Finally, it is worth noting that Bayesian procedures do not require knowledge of the optimal mixing distribution F_σ in Condition 4.3 for implementation. The existence is purely a proof device towards obtaining theoretical guarantees.

The following result derives the quasi-Bayes posterior contraction rate when the Gaussian mixture bias is as in Condition 4.3.

Theorem 2 (Rates with Gaussian Mixture Bias)**.** *Suppose the model is mildly ill-posed with* $\zeta \geq 0$ *as in Definition 3. Suppose Conditions 4.1, 4.3 hold and the covariance prior is taken as* $G_n \sim G/\sigma_n^2$ *where* G *satisfies Assumption 2. Let* $\kappa, \chi > 0$ *be as in Assumption 2 and Condition 4.3, respectively. Define* $\lambda = \max\{\chi^{-1}(d+2), \chi^{-1}d + 1\}$. *If* $T_n \asymp n^{1/[2(p+\zeta)+d]}\sqrt{\log n}$ *and*

$\sigma_n^2 \asymp T_n^{2-d/\kappa}/(\log n)^{\lambda/\kappa+1} \log\log n$, *then there exists a universal constant* $C > 0$ *such that*

$$\nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-p/[2(p+\zeta)+d]}(\log n)^{(\lambda+\zeta)/2+d/4} \,\middle|\, \mathcal{Z}_n\right) = o_{\mathbb{P}}(1).$$

Theorem 2 provides contraction rates for the mildly ill-posed case. As the theorem illustrates, rates are determined by the Gaussian mixture bias. From our discussion above, this implicitly depends on the underlying smoothness of the density. Rates for the severely ill-posed case can be obtained in a similar manner. In this case, the Gaussian mixture bias decays exponentially in σ (see e.g. Donnet et al., 2018).

4.3 Main Results (Repeated Measurements)

In this section, we present the main results for the repeated measurements quasi-Bayes posterior of Section 3.2. Recall that in this case, we observe a sample $\{Y_{1,i}, Y_{2,i}\}_{i=1}^n$ from the model in (17). The main conditions that we impose on the observations is summarized in the following condition.

Condition 4.4 (Data)**.** *(i)* $(Y_{1,i}, Y_{2,i})_{i=1}^n$ is a sequence of independent and identically distributed random vectors. *(ii)* Finite second moments: $\mathbb{E}(\|Y_1\|^2) < \infty$ and $\mathbb{E}(\|Y_2\|^2) < \infty$.

To begin, we consider the repeated measurements quasi-Bayes posterior without explicit boundary correction. From Section 3.2, this takes the form

$$\nu_{\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int \exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{44}$$

Up to sampling uncertainty, the quasi-Bayes posterior in (44) is based on a transformation of the objective function

$$Q(\varphi_{P,\Sigma}) = \|\varphi_{Y_2}(\nabla\log\varphi_{P,\Sigma} - \nabla\log\varphi_X)\|^2_{\mathbb{B}(T)}. \tag{45}$$

From this formulation, some natural questions arise. What are the statistical guarantees of such a posterior? Does L^2 gradient behavior over a domain like $\mathbb{B}(T)$ provide enough identifying information to recover the true latent density? What is the difference between this posterior and one where boundary behavior is explicitly accounted for? In the remainder of this section, we investigate these questions in detail.

As in Section 4.2 (see (42) and the discussion following it), first we consider the case where the true latent density f_X admits a representation as a (possibly infinite) Gaussian mixture:

$$f_X(x) = \phi_{\Sigma_0} \star F_0(x) = \int_{\mathbb{R}^d} \phi_{\Sigma_0}(x-z)dF_0(z). \tag{46}$$

Our main assumption on the Gaussian mixture specification is as follows.

Condition 4.5 (Exact Gaussian Mixtures in Repeated Measurements)**.** *(i)* $f_X = \phi_{\Sigma_0} \star F_0$ where $\Sigma_0 \in \mathbb{R}^{d\times d}$ is a positive definite matrix. *(ii)* The mixing distribution F_0 satisfies $F_0(t \in \mathbb{R}^d : \|t\|> z) \leq C\exp(-C'z^\chi)$, $|\varphi_{F_0}(t)| \geq c\exp\big(-c'\|t\|^2\big)$ and $\|\nabla \log \varphi_{F_0}(t)\|\leq C\|t\|$ for all sufficiently large z and $\|t\|$, where $\chi > 0$ and $c, c', C, C' > 0$ are universal constants.

By integrating the differential inequality, the Condition $\|\nabla \log \varphi_{F_0}(t)\|\leq C\|t\|$ essentially states that $\varphi_{F_0}(.)$ has at most Gaussian decay. As a first result, we show that quasi-Bayes posterior in (44) contracts rapidly around the logarithmic gradient of the true characteristic function.

Theorem 3 (Preliminary Contraction of Gradients)**.** *Suppose the characteristic function of Y_2 has at most Gaussian decay, that is $\inf_{\|t\|_\infty\leq T} |\varphi_{Y_2}(t)| \geq \exp\big(-RT^2\big)$ for some $R > 0$. Suppose Conditions 4.4, 4.5 hold and the covariance prior is taken as G where G satisfies Assumption 2. Furthermore, suppose Σ_0 is in the support of G. Let $\chi > 0$ be as in Condition 4.5 and $\lambda = \max\{\chi^{-1}(d + 2) + d/2, d + 1\}$. If $T_n = (c_0 \log n)^{1/2}$ for any c_0 satisfying $c_0R = \gamma < 1/2$, then there exists a universal constant $C > 0$ such that*

$$\nu_{\alpha,G}\left(\|\nabla \log \varphi_X - \nabla \log \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} > Cn^{-1/2+\gamma}(\log n)^{\lambda/2} \,\middle|\, \mathcal{Z}_n\right) = o_{\mathbb{P}}(1).$$

By picking c_0 sufficiently small, Theorem 3 shows that the quasi-Bayes posterior contracts around the gradient of $\log \varphi_X$ at nearly parametric rates of the form $n^{-1/2+\epsilon}$ for any $\epsilon > 0$.

While the results of Theorem 3 are encouraging, the question remains as to whether this can be improved to convergence in a stronger metric such as $\|f_X - \phi_{P,\Sigma}\|_{L^2}$. As φ_X and $\varphi_{P,\Sigma}$ are the characteristic functions of random variables, they necessarily satisfy the initial value condition $\nabla \log \varphi_X(0) = \nabla \log \varphi_{P,\Sigma}(0) = 0$. From this initial value condition, in dimension $d = 1$, the fundemental theorem of calculus and Cauchy-Schwarz imply

$$\begin{aligned}\sup_{t\in\mathbb{B}(T)} |\log \varphi_X(t) - \log \varphi_{P,\Sigma}(t)| &= \sup_{t\in\mathbb{B}(T)} \left|\int_0^t [\nabla \log \varphi_X(s) - \nabla \log \varphi_{P,\Sigma}(s)]ds\right| \\ &\leq \sqrt{T}\|\nabla \log \varphi_X - \nabla \log \varphi_{P,\Sigma}\|_{\mathbb{B}(T)}.\end{aligned} \tag{47}$$

In particular, at least in dimension $d = 1$, contraction of gradients is informative towards recovering the true latent distribution. This is formalized in the next result.

Theorem 4 (Rates with Exact Gaussian Mixtures, $d = 1$)**.** *Suppose $d = 1$ and the characteristic function of Y_2 has at most Gaussian decay, that is $\inf_{\|t\|_\infty\leq T} |\varphi_{Y_2}(t)| \geq \exp\big(-RT^2\big)$ for some $R > 0$. Suppose Conditions 4.4, 4.5 hold and the covariance prior is taken as $G_n \sim G/\sigma_n^2$ where G satisfies Assumption 2. Furthermore, suppose Σ_0 is in the support of G. Let $\kappa, \chi > 0$ be as in Assumption 2 and Condition 4.5 respectively and define $\lambda = \max\{1/2 + 3/\chi, 2\}$. If $T_n = (c_0 \log n)^{1/2}$ for any c_0 satisfying $c_0R = \gamma < 1/2$ and $\sigma_n^2 \asymp (\log n)^{-\lambda/\kappa}$, then there exists a* *universal constant $C > 0$ and $V \in (0, 1/2)$ such that*

$$\nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-V} \,|\mathcal{Z}_n\right) = o_{\mathbb{P}}(1).$$

With Gaussian decay on φ_{Y_2}, the model bias (which is determined by the decay rate of a Gaussian mixture characteristic function) has similar order to the variance. In this case, as in the discussion following Theorem 1, the precise value of the exponent V depends on second order factors such as the constant R and the eigenvalues of the matrix Σ_0 in the representation (46).

The strategy employed to prove Theorem 4 is based on (47). Unfortunately, for dimension $d > 1$ and a function $f : \mathbb{B}(T_n) \to \mathbb{C}$, there is no direct relationship between $\|f\|_{\mathbb{B}(T_n)}$ and $\|\nabla f\|_{\mathbb{B}(T_n)}$, even with initial value condition $f(0) = 0$. However, for sufficiently nice domains like $\mathbb{B}(T_n)$, it can be shown that[24]

$$\|f\|_{\mathbb{B}(T_n)} \leq C_n(\|\nabla f\|_{\mathbb{B}(T_n)} + \|f\|_{\partial\mathbb{B}(T_n)}) \,, \tag{48}$$

holds, where C_n is a constant that depends only on T_n and $\|.\|_{\partial\mathbb{B}(T_n)}$ is the L^2 norm on the boundary $\partial\mathbb{B}(T_n)$ of $\mathbb{B}(T_n)$.[25] In the theory of Sobolev spaces, inequality (48) is commonly referred to as a Poincaré inequality (Evans, 2022).

The discussion above suggests that a suitable modification of the quasi-Bayes posterior to account for the second term in (48) may resolve the problem. The starting point in our analysis towards this goal is to observe that a function $f : \mathbb{B}(T_n) \to \mathbb{C}$ with initial value condition $f(0) = 0$ satisfies

$$f(z) = \int_0^1 \langle \nabla f(tz), z\rangle dt \qquad \forall\, z \in \partial\mathbb{B}(T_n). \tag{49}$$

Specifically, (49) says that the value of $f(.)$ at a boundary point $z \in \partial\mathbb{B}(T_n)$ can be determined by a scaled average of ∇f on the line segment connecting the origin to z. In particular, by accounting for values of ∇f on this line segment, we can induce a boundary corrected quasi-Bayes posterior. This motivates the following distance metric. For a function $g : \mathbb{B}(T_n) \to \mathbb{C}^d$, define the metric

$$\|g\|^2_{\partial,\,\mathbb{B}(T)} = \int_{\mathbb{B}(T_n)} \|g(t)\|^2 dt + \int_{\partial\mathbb{B}(T_n)} \int_0^1 \|g(tz)\|^2 dt d\mathcal{H}^{d-1}(z). \tag{50}$$

As in Section 3.2, $d\mathcal{H}^{d-1}$ denotes the $d-1$ dimensional Hausdorff measure on $\partial\mathbb{B}(T_n)$. For $d > 1$,

[24] We show this formally in Lemma 13. This makes use of the general Poincaré trace results in Maggi and Villani (2005, 2008)

[25] The L^2 norm on the boundary is taken with respect to the $d-1$ dimensional Hausdorff measure $\mathcal{H}^{d-1}$. Specifically,

$$\|f\|^2_{\partial\mathbb{B}(T_n)} = \int_{\partial\mathbb{B}(T_n)} |f(t)|^2 \, d\mathcal{H}^{d-1}(t).$$

we define the (boundary corrected) repeated measurements quasi-Bayes posterior by

$$\nu_{\partial,\alpha,G}(\varphi_{P,\Sigma} \mid \mathcal{Z}_n) = \frac{\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\Sigma)}{\int\exp\big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\Sigma)}. \tag{51}$$

The following result verifies that quasi-Bayes posterior in (51) contracts towards the true latent distribution, for all dimensions.

Theorem 5 (Rates with Exact Gaussian Mixtures, $d > 1$)**.** *Suppose the characteristic function of Y_2 has at most Gaussian decay, that is $\inf_{\|t\|_\infty \le T}|\varphi_{Y_2}(t)| \ge \exp\big(-RT^2\big)$ for some $R > 0$. Suppose Conditions 4.4, 4.5 hold and the covariance prior is taken as $G_n \sim G/\sigma_n^2$ where G satisfies Assumption 2. Furthermore, suppose Σ_0 is in the support of G. Let $\kappa, \chi > 0$ be as in Assumption 2 and Condition 4.5 respectively and define $\lambda = \max\{\chi^{-1}(d+2) + d/2, d+1\}$. If $T_n = (c_0\log n)^{1/2}$ for any c_0 satisfying $c_0 R = \gamma < 1/2$ and $\sigma_n^2 \asymp (\log n)^{-\lambda/\kappa}$, then there exists a universal constant $C > 0$ and $V \in (0, 1/2)$ such that*

$$\nu_{\partial,\alpha,G_n}\bigg(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-V}\,|\mathcal{Z}_n\bigg) = o_{\mathbb{P}}(1).$$

Next, analogous to our discussion in Section 4.2, we consider the case where the true latent density f_X is not an exact Gaussian mixture and thus, the prior model exhibits non-negligible bias. In this model, the Gaussian mixture bias is closely related to the quantity

$$\mathcal{B}(\sigma, T) = \min_P \|\varphi_{Y_2}(\nabla\log\varphi_{P,\sigma^2 I} - \nabla\log\varphi_X)\|_{\partial,\,\mathbb{B}(T)} \tag{52}$$

where $\sigma^2 I$ denotes the identity matrix with each diagonal entry σ^2 and the minimum is over all mixing distributions P. For a conservative upper bound on the rate at which $\mathcal{B}(\sigma)$ tends to 0 as $\sigma \to 0$, observe that by picking $P = f_X$ we obtain

$$\begin{aligned}\|\varphi_{Y_2}(\nabla\log\varphi_{P,\sigma^2 I} - \nabla\log\varphi_X)\|_{\partial,\,\mathbb{B}(T)} &= \|\varphi_{Y_2}(\nabla\log(\varphi_X\varphi_{\sigma^2 I}) - \nabla\log\varphi_X)\|_{\partial,\,\mathbb{B}(T)}\\ &= \|\varphi_{Y_2}\nabla\log\varphi_{\sigma^2 I}\|_{\partial,\,\mathbb{B}(T)}.\end{aligned}$$

Since the logarithmic gradient of the Gaussian characteristic function is $\nabla\log\varphi_{\sigma^2 I}(t) = -\sigma^2 t$, the preceding equality reduces to

$$\|\varphi_{Y_2}(\nabla\log\varphi_{P,\sigma^2 I} - \nabla\log\varphi_X)\|_{\partial,\,\mathbb{B}(T)} \le \sigma^2\|\varphi_{Y_2}(t)t\|_{\partial,\,\mathbb{B}(T)} \lessapprox \sigma^2(1 + T^{d/2-1}). \tag{53}$$

The preceding bound is valid provided that[26] $\int_{\mathbb{R}^d}\|t\|^2|\varphi_{Y_2}(t)|^2\,dt < \infty$ or equivalently that $f_{Y_2} = f_X \star f_{\epsilon_2} \in \mathbf{H}^1$. For $f_X \star f_{\epsilon_2}$ with smoothness order greater than 1, this argument does not provide better rates. Analogous to the discussion in Section 3.1, the choice $P = f_X$ in (52)

[26] By change of variables on $(\partial\mathbb{B}(T), d\mathcal{H}^{d-1})$ and the coarea formula (Evans, 2022, C.3)

$$\int_0^1\int_{\partial\mathbb{B}(T)} t^2\|z\|^2|\varphi_{Y_2}(tz)|^2\,d\mathcal{H}^{d-1}(z)dt \lessapprox T^{d-1}\int_0^1\int_{\partial\mathbb{B}(t\times T)}\frac{|\varphi_{Y_2}(y)|^2}{\|y\|^{d-3}}d\mathcal{H}^{d-1}(y)dt = T^{d-2}\int_{\mathbb{B}(T)}\frac{|\varphi_{Y_2}(y)|^2}{\|y\|^{d-3}}dy \lessapprox T^{d-2}.$$

may not be optimal. One way to see this directly is to express $\varphi_{Y_2} = \varphi_X \varphi_{\epsilon_2}$ and write

$$\varphi_{Y_2}(\nabla \log \varphi_{P,\sigma^2 I} - \nabla \log \varphi_X) = \nabla \log \varphi_{P,\sigma^2 I}[\varphi_{\epsilon_2}(\varphi_X - \varphi_{P,\sigma^2 I})] + \varphi_{\epsilon_2}(\nabla \varphi_{P,\sigma^2 I} - \nabla \varphi_X). \quad (54)$$

If we consider mixing distributions P with characteristic function having at most Gaussian decay, then $\sup_{t \in \mathbb{B}(T)} \|\nabla \log \varphi_{P,\sigma^2 I}(t)\| \lessapprox T$. In particular, this implies that the Gaussian mixture bias is bounded above by

$$\mathcal{B}(\sigma, T) \lessapprox \min_P \left[T \|\varphi_{\epsilon_2}(\varphi_X - \varphi_{P,\sigma^2 I})\|_{\partial,\, \mathbb{B}(T)} + \|\varphi_{\epsilon_2}(\nabla \varphi_X - \nabla \varphi_{P,\sigma^2 I})\|_{\partial,\, \mathbb{B}(T)} \right]. \quad (55)$$

The quantity on the right closely resembles the Gaussian mixture bias that arises from a deconvolution model as in (43). While this illustrates some of factors that determine the bias, the preceding argument is however quite loose as it treats the characteristic function and its gradient separately. We generally expect the bias in (52) to have smaller order than the right side of (55). Indeed, the simple choice $P = f_X$ on the right side of (55) leads to a worse bound than (53) and requires more stringent[27] smoothness conditions imposed on (f_X, f_ϵ). Motivated by this discussion, we consider the following characterization of the Gaussian mixture bias.

Condition 4.6 (Gaussian Mixture Bias). *(i)* $f_X \in \mathbf{H}^p$ for some $p > 0$ and $\sup_{t \in \mathbb{R}^d} \|\nabla \log \varphi_X(t)\| \leq D$ for some $D < \infty$. *(ii)* There exists universal constants $C, M < \infty$ and $\chi > 0$ such that for all $\sigma > 0$ sufficiently small and $T > 0$ sufficiently large, there exists a mixing distribution $F_\sigma = F_{X,\epsilon,\sigma}$ supported on the cube $I_\sigma = [-C(\log \sigma^{-1})^{1/\chi}, C(\log \sigma^{-1})^{1/\chi}]^d$ that satisfies *(a)* $\|\varphi_{Y_2}(\nabla \log \varphi_{F_\sigma,\sigma^2 I} - \nabla \log \varphi_X)\|^2_{\partial,\, \mathbb{B}(T)} \leq M(1 + T^{d-2})\sigma^{2(s+\zeta)}$ for some $s > \max\{3/2, d/2 + 1/2\}$, *(b)* $\inf_{\|t\| \leq T} |\varphi_{F_\sigma}(t)| \geq T^{-\gamma_1}\sigma^{-\gamma_2}$ for some $\gamma_1, \gamma_2 < \infty$ and *(c)* $\|\nabla \log \varphi_{F_\sigma}\| \leq M$.

The following result derives the quasi-Bayes posterior contraction rate when the Gaussian mixture bias is as in Condition 4.6. This condition is stated to cover the mildly ill-posed setting. The requirement in the severely ill-posed case is similar except that the Gaussian mixture bias is expected to be exponential in σ.

Theorem 6 (Rates with Gaussian Mixture Bias). *Suppose Conditions 4.4, 4.6 hold and the covariance prior is taken as* $G_n \sim G/\sigma_n^2$ *where* G *satisfies Assumption 2. Let* $\kappa, \chi > 0$ *be as in Assumption 2 and 4.2, respectively. Define* $\lambda = \max\{\chi^{-1}(d+2), \chi^{-1}d + 1\}$ *and* $\beta = \max\{0, d-2\}$. *Suppose there exists* $\zeta > 0$ *such that the characteristic function of* Y_2 *satisfies* $\inf_{\|t\| \leq T} |\varphi_{Y_2}(t)| \geq RT^{-\zeta}$ *for some* $R > 0$. *If* $T_n \asymp n^{1/[2(s+\zeta)+(d-\beta)]}\sqrt{\log n}$ *and* $\sigma_n^2 \asymp T_n^{2-d/\kappa}/(\log n)^{\lambda/\kappa+1} \log\log n$, *then there exists a universal constant* $C > 0$ *such that*

$$\nu_{\alpha,G_n}\left(\|f_X - \phi_{P,\Sigma}\|_{L^2} > Cn^{-\min\{p, s-\beta/2-1.5\}/[2(s+\zeta)+(d-\beta)]}(\log n)^{\lambda/2} \,\middle|\, \mathcal{Z}_n \right) = o_{\mathbb{P}}(1).$$

As in Theorem 2, rates are determined by the Gaussian mixture bias. An interesting avenue

[27] To be specific, it requires $f_{\epsilon_2} \in \mathbf{H}^2, f_X \in \mathbf{H}^1, \int_{\mathbb{R}^d} \|t\|^4 \|\nabla \varphi_X(t)\|^2 |\varphi_\epsilon(t)|^2 \, dt < \infty$ instead of the much weaker requirement that $f_X \star f_\epsilon \in \mathbf{H}^1$ in (53).

for future research would be explicitly characterize the Gaussian mixture bias based on lower level smoothness assumptions on the latent density and unobserved errors. As the analysis in Rousseau (2010) and Shen et al. (2013) shows, the precise construction of the optimal mixing distribution can be quite involved. We leave this investigation for future work.

4.4 Main Results (Multi-Factor Models)

In this section, we present the main results for the multi-factor quasi-Bayes posterior of Section 3.3. Recall that in this case, we observe a sample $\{\mathbf{Y}_i\}_{i=1}^n$ from the model in (28). The main conditions that we impose on the observations are summarized in the following condition.

Condition 4.7 (Data). (i) $(\mathbf{Y}_i)_{i=1}^n$ is a sequence of independent and identically distributed random vectors. (ii) Finite second moment: $\mathbb{E}(\|\mathbf{Y}\|^2) < \infty$. (iii) The factors $X_1, \ldots, X_K$ are mutually independent and demeaned, i.e $\mathbb{E}[X_k] = 0$ for $k = 1, \ldots, K$.

The multi-factor quasi-Bayes posterior for latent factor X_k is given by

$$\nu_{\alpha,G}(\varphi_{P,\sigma^2} \mid \mathcal{Z}_n) = \frac{\exp\big(- \frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T)}\big)\nu_{\alpha,G}(P,\sigma^2)}{\int \exp\big(- \frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T)}\big)d\nu_{\alpha,G}(P,\sigma^2)}. \tag{56}$$

Up to sampling uncertainty, the quasi-Bayes posterior in (56) is based on a transformation of the objective function

$$Q(\varphi_{P,\sigma^2}) = \|\varphi^2_{\mathbf{Y}}(t)[(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k) - (\log\varphi_{X_k})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T)}. \tag{57}$$

As a first observation, we note that near minimizers of (57) are not uniquely identified without further normalization restrictions. Intuitively, such minimizers are (approximately) solutions to the second order differential equation

$$(\log\varphi_{P,\sigma^2})''(t) = (\log\varphi_{X_k})''(t). \tag{58}$$

For a unique solution of (58) to exist, we require two boundary conditions. As all characteristic functions equal 1 at zero, all solutions must necessarily satisfy the first boundary condition $(\log\varphi_{P,\sigma^2})(0) = 0$. For the second condition, observe that by Condition 4.7(iii), the true factors are demeaned and hence their characteristic functions satisfy $(\log\varphi_{X_k})'(0) = \mathbf{i}\mathbb{E}[X_k] = 0$. Unfortunately, solutions (or near solutions) to (58) do not necessarily satisfy this boundary condition. One way to enforce this constraint is to force the normalization $(\log\varphi_{P,\sigma^2})'(0) = 0$ directly on the prior. In essence, this would entail replacing the prior with a restricted prior that is supported on mean zero Gaussian mixtures. While this is theoretically possible, in practice a restricted prior entails significant computational challenges. A second option, that we use in the remainder of this section, is to use the original quasi-Bayes posterior in (56) but to demean the posterior samples. To that end, consider the demeaned posterior measure

$$\overline{\nu}_{\alpha,G}(.|\mathcal{Z}_n) \sim Z - \mathbb{E}[Z] \qquad \text{where} \quad Z \sim \nu_{\alpha,G}(.|\mathcal{Z}_n). \tag{59}$$

Each realization of (59) is a demeaned Gaussian mixture. If (P, Σ) is the mixing distribution and covariance matrix associated to the original Gaussian mixture sample from (56), we denote the density and characteristic function of the demeaned mixture by $\overline{\phi}_{P,\Sigma}$ and $\overline{\varphi}_{P,\Sigma}$, respectively.

First, we consider the case where the latent density f_{X_k} admits a representation as a (possibly infinite) Gaussian mixture:

$$f_{X_k}(x) = \phi_{\sigma_0^2} \star F_0(x) = \int_{\mathbb{R}} \phi_{\sigma_0^2}(x - z) dF_0(z). \tag{60}$$

Our main assumption on the Gaussian mixture specification is as follows.

Condition 4.8 (Exact Gaussian Mixtures in Multi-Factor Models)**.** (i) $f_{X_k} = \phi_{\sigma_0^2} \star F_0$ for some constant $\sigma_0^2 > 0$. (ii) The mixing distributions $F_{0,k}$ satisfy $F_0(t \in \mathbb{R} : |t| > z) \leq C \exp(-C' z^{\chi})$, $|\varphi_{F_0}(t)| \geq c \exp\big(-c' t^2\big)$ and $\big|\partial_t^2 \log \varphi_{F_0}(t)\big| \leq C$ for all sufficiently large z and $|t|$, where $\chi > 0$ and $c, c', C, C' > 0$ are universal constants.

By integrating the differential inequality, the Condition $\big|\partial_t^2 \log \varphi_{F_0}(t)\big| \leq C$ essentially states that $\varphi_{F_0}(.)$ has at most Gaussian decay. As a first result, we show that quasi-Bayes posterior in (59) contracts rapidly around the logarithmic gradient of the true characteristic function.

Theorem 7 (Rates with Exact Gaussian Mixtures)**.** *Suppose the characteristic function of* $\mathbf{Y}$ *has at most Gaussian decay, that is* $\inf_{\|t\|_\infty \leq T} |\varphi_{\mathbf{Y}}(t)| \geq \exp\big(-RT^2\big)$ *for some* $R > 0$*. Suppose Conditions 4.7, 4.8 hold and the covariance prior is taken as* $G_n \sim G/\sigma_n^2$ *where* G *satisfies Assumption 2. Furthermore, suppose* σ_0^2 *is in the support of* G*. Let* $\kappa, \chi > 0$ *be as in Assumption 2 and Condition 4.8 respectively and define* $\lambda = \max\{\chi^{-1}(d+2)+d/2, d+1\}$*. If* $T_n = (c_0 \log n)^{1/2}$ *for any* c_0 *satisfying* $2c_0 R = \gamma < 1/2$ *and* $\sigma_n^2 \asymp (\log n)^{-\lambda/\kappa}$*, then there exists a universal constant* $C > 0$ *and* $V \in (0, 1/2)$ *such that*

$$\overline{\nu}_{\alpha,G_n}\left(\|f_{X_k} - \phi_{P,\sigma^2}\|_{L^2} > Cn^{-V} \; | \mathcal{Z}_n \right) = o_{\mathbb{P}}(1). \tag{61}$$

Similar to the preceding sections, with Gaussian decay on $\varphi_{\mathbf{Y}}$, the model bias (which is determined by the decay rate of a Gaussian mixture characteristic function) has similar order to the variance. In this case, as in the discussion following Theorem 1, the precise value of the exponent V depends on second order factors such as the constant R and σ_0^2 in representation (60).

Next, analogous to our discussion in Section 4.2 and 4.3, we consider the case where the true latent density f_{X_k} is not an exact Gaussian mixture and thus, the prior model exhibits non-negligible bias. In this model, the Gaussian mixture bias is closely related to the quantity

$$\mathcal{B}(\sigma, T) = \min_P \|\varphi_{\mathbf{Y}}^2(t)[\big(\log \varphi_{P,\sigma^2}\big)''(t'\mathbf{A}_k) - \big(\log \varphi_{X_k}\big)''(t'\mathbf{A}_k)]\|_{\mathbb{B}(T)}. \tag{62}$$

where the minimum is over all mixing distributions P.[28]

By similar reasoning to the preceding sections, we consider the following characterization of the Gaussian mixture bias.

Condition 4.9 (Gaussian Mixture Bias)**.** *(i)* $f_{X_k} \in \mathbf{H}^p$ for some $p > 0$ and $\sup_{t\in\mathbb{R}} |(\log \varphi_{X_k})''(t)| \leq D$ for some $D < \infty$. *(ii)* There exists universal constants $C, M < \infty$ and $\chi > 0$ such that for all $\sigma > 0$ sufficiently small and $T > 0$ sufficiently large, there exists a mixing distribution $F_\sigma = F_{X_k,\mathbf{Y},\sigma}$ supported on the cube $I_\sigma = [-C(\log \sigma^{-1})^{1/\chi}, C(\log \sigma^{-1})^{1/\chi}]$ that satisfies *(a)* $\|\varphi^2_{\mathbf{Y}}(t)[(\log \varphi_{F_\sigma,\sigma^2})''(t'\mathbf{A}_k) - (\log \varphi_{X_k})''(t'\mathbf{A}_k)]\|_{\mathbb{B}(T)} \leq M\sigma^{2\zeta+s}$ for some $s > \max\{(5-d)/2, 0\}$, *(b)* $\inf_{\|t\|\leq T} |\varphi_{F_\sigma}(t)| \geq T^{-\gamma_1}\sigma^{-\gamma_2}$ for some $\gamma_1, \gamma_2 < \infty$, *(c)* $|(\log \varphi_{F_\sigma})''(t)| \leq M$ and $\sup_{r=1,2} |\partial_t^r \varphi_{F_\sigma}(t)| \leq M$.

The following result derives the quasi-Bayes posterior contraction rate when the Gaussian mixture bias is as in Condition 4.9.

Theorem 8 (Rates with Gaussian Mixture Bias)**.** *Suppose Conditions 4.7, 4.9 hold and the covariance prior is taken as $G_n \sim G/\sigma_n^2$ where G satisfies Assumption 2. Let $\kappa, \chi > 0$ be as in Assumption 2 and Condition 4.9, respectively. Define $\lambda = \max\{\chi^{-1}(d+2), \chi^{-1}d+1\}$. Suppose there exists $\zeta > 0$ such that the characteristic function of $\mathbf{Y}$ satisfies $\inf_{\|t\|_\infty\leq T} |\varphi_{\mathbf{Y}}(t)| \geq RT^{-\zeta}$ for some $R > 0$. If $T_n \asymp n^{1/[2(s+2\zeta)+d]}\sqrt{\log n}$ and $\sigma_n^2 \asymp T_n^{2-d/\kappa}/(\log n)^{\lambda/\kappa+1} \log\log n$, then there exists a universal constant $C > 0$ such that*

$$\nu_{\alpha,G_n}\left(\|f_{X_k} - \phi_{P,\sigma^2}\|_{L^2} > Cn^{-\min\{p,s+(d-5)/2\}/[2(s+2\zeta)+d]}(\log n)^{\lambda/2+\zeta+5/4} \,\middle|\, \mathcal{Z}_n \right) = o_{\mathbb{P}}(1).$$

Remark 13 (Rates for Joint Posteriors)**.** Let $\mathbf{P} = (P_1, \dots, P_K)$ denote a vector of mixing distributions and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_K^2)$ a vector of variance parameters. Let $\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2}(t)$ denote the vector with elements $\{(\log \varphi_{P_i,\sigma_i^2})''(t'\mathbf{A}_i)\}_{i=1}^K$. With some abuse of notation, we denote the vector of individual characteristic functions by $\varphi_{\mathbf{P},\boldsymbol{\sigma}^2} = (\varphi_{P_1,\sigma_1^2}, \dots, \varphi_{P_K,\sigma_K^2})$. The joint multi-factor quasi-Bayes posterior is given by

$$\begin{aligned} \nu_{\alpha,G}(\varphi_{\mathbf{P},\boldsymbol{\sigma}^2} \mid \mathcal{Z}_n) &= \frac{\exp\left(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\widehat{\mathcal{V}}_{\mathbf{Y}} - \mathbf{Q}\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2})\|^2_{\mathbb{B}(T)}\right)\nu_{\alpha,G}(\mathbf{P}, \boldsymbol{\sigma}^2)}{\int \exp\left(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\widehat{\mathcal{V}}_{\mathbf{Y}} - \mathbf{Q}\mathcal{V}_{\mathbf{P},\boldsymbol{\sigma}^2})\|^2_{\mathbb{B}(T)}\right)d\nu_{\alpha,G}(\mathbf{P}, \boldsymbol{\sigma}^2)} \\ \overline{\nu}_{\alpha,G}(. \mid \mathcal{Z}_n) &\sim Z - \mathbb{E}[Z] \ , \ Z \sim \nu_{\alpha,G}(.|\mathcal{Z}_n). \end{aligned} \quad (63)$$

Here G is the common univariate prior on each variance parameter $\{\sigma_i^2\}_{i=1}^K$. The prior $\nu_{\alpha,G}$ is the one obtained by placing mutually independent $(\mathrm{DP}_\alpha \otimes G)$ priors on each of $(P_i, \sigma_i^2)_{i=1}^K$. An analogous argument to the preceding cases provides contraction rates for the joint quasi-Bayes posterior. The main idea being that, as the priors are independent among the factors and

[28] For a conservative upper bound on the rate at which $\mathcal{B}(\sigma, T)$ tends to 0 as $\sigma \to 0$, observe that by picking $P = f_X$, we obtain $\mathcal{B}(\sigma, T) \leq \sigma^2 \|\varphi^2_{\mathbf{Y}}(t)\|_{\mathbb{B}(T)} \lessapprox \sigma^2$. The final bound is valid if $\|\varphi^2_{\mathbf{Y}}(t)\|_{L^2} < \infty$ which ,by the Hausdorff–Young inequality, is true whenever $\int_{\mathbb{R}^L} |f_{\mathbf{Y}}(y)|^{4/3}\, dy < \infty$. Analogous to our discussions in the preceding sections, one can typically do better by allowing $P = P_{X,\epsilon,\sigma}$ to depend implicitly on all the features $(f_{\mathbf{Y}}, f_{X_k}, \sigma)$.

the factors are mutually independent, analysis of the quasi-Bayes posterior in (63) reduces to studying univariate quasi-posteriors. In principle, one could obtain rates where certain factors are modelled as exact (infinite) Gaussian mixtures and others with non-neglegible Gaussian mixture bias.

5 Simulations

In this section, we provide simulation evidence on the finite sample performance of our quasi-Bayes posteriors. In all cases, we use $m = 1000$ Monte Carlo replications to estimate the population expectation of interest.

5.1 Deconvolution

In this section, we examine the finite sample properties of the deconvolution quasi-Bayes posterior, introduced in Section 3.1. The setup is as follows. Two samples (Y_1, Y_2) are generated according to the specification

$$
\begin{aligned}
Y_1 &= X + \epsilon_1 \qquad & \epsilon_1 &\sim N(0,1) \\
Y_2 &= X + \epsilon_2 \qquad & \epsilon_2 &\sim N(0,1)
\end{aligned}
$$

where ϵ_1, ϵ_2 are independent.[29] To nest this into the setup of Section 3.1, we write

$$
\left(\frac{Y_1 + Y_2}{2}\right) = X + \left(\frac{\epsilon_1 + \epsilon_2}{2}\right), \tag{64}
$$

$$
\left(\frac{Y_1 - Y_2}{2}\right) = \left(\frac{\epsilon_1 - \epsilon_2}{2}\right). \tag{65}
$$

In particular, we can use the observations in (65) as an auxiliary sample for the error in (64). This provides us with a sample to construct the quasi-Bayes posterior in Section 3.1.

The implementation details are as follows. For the base measure of the Dirichlet process prior DP_α we take $\alpha = \mathcal{N}(\widehat{\mu}, \widehat{\sigma}^2)$ where $(\widehat{\mu}, \widehat{\sigma}^2)$ are the sample mean and variance of the observations in (64). The standard deviation prior is taken to be Inv-Gamma$(2, 2)$. In all cases, we use $\mathbb{B}(T) = [-2, 2]$ as the set of identifying restrictions.

Table 1 compares the mean integrated squared error (MISE) of the deconvolution quasi-Bayes posterior mean against common alternatives in the literature. The alternative estimators use the bandwidth selection rule proposed in (Delaigle and Gijbels, 2004) to select the tuning parameter T. By contrast, we use $T = 2$, equivalently $\mathbb{B}(T) = [-2, 2]$, in all designs.[30] As Table 1 illustrates, the MISE efficiency of quasi-Bayes over traditional estimators is substantial. One reason for this, among other factors, is that even a very modestly informative prior can be

[29] In all our simulations, the results were similar with standard Laplace errors.

[30] We found that it was possible to improve performance further by selecting T based on the specific design or Monte Carlo realization. However, as the validity of traditional deconvolution selection procedures such as (Delaigle and Gijbels, 2004) is unclear in this context, we opted to implement everything using the same fixed choice.

beneficial towards variance reduction by down weighting implausible regions of the parameter space. In a nonparametric ill-posed inverse setup such as ours, this variance reduction can be substantial.

$\sqrt{n} \times$ MISE					
	HM	**LV**	**BR**	**Deconvolution**	**quasi-Bayes (D)**
$\mathcal{N}(0,1)$	0.285	0.250	0.155	0.164	0.037
Laplace$(0,1)$	1.107	1.171	0.822	0.791	0.202
Gamma$(2,1)$	0.981	1.107	0.759	0.822	0.290
Gamma$(5,1)$	0.506	0.411	0.348	0.348	0.032
Log-Normal$(0,1)$	7.595	7.271	11.067	9.170	1.446
$0.5\mathcal{N}(-2,1)+0.5\mathcal{N}(2,1)$	1.962	1.993	1.519	1.297	0.042

Table 1: $n = 1000$. **HM**, **LV**, **BR** refer to the estimators proposed in Horowitz and Markatou (1996), Li and Vuong (1998) and Bonhomme and Robin (2010), respectively. **Deconvolution** refers to the traditional deconvolution estimator (Delaigle and Gijbels, 2004). The MISE for these estimators are taken from Table 1 and 2 in Bonhomme and Robin (2010). **quasi-Bayes (D)** refers to the deconvolution quasi-Bayes (Section 3.1) posterior mean.

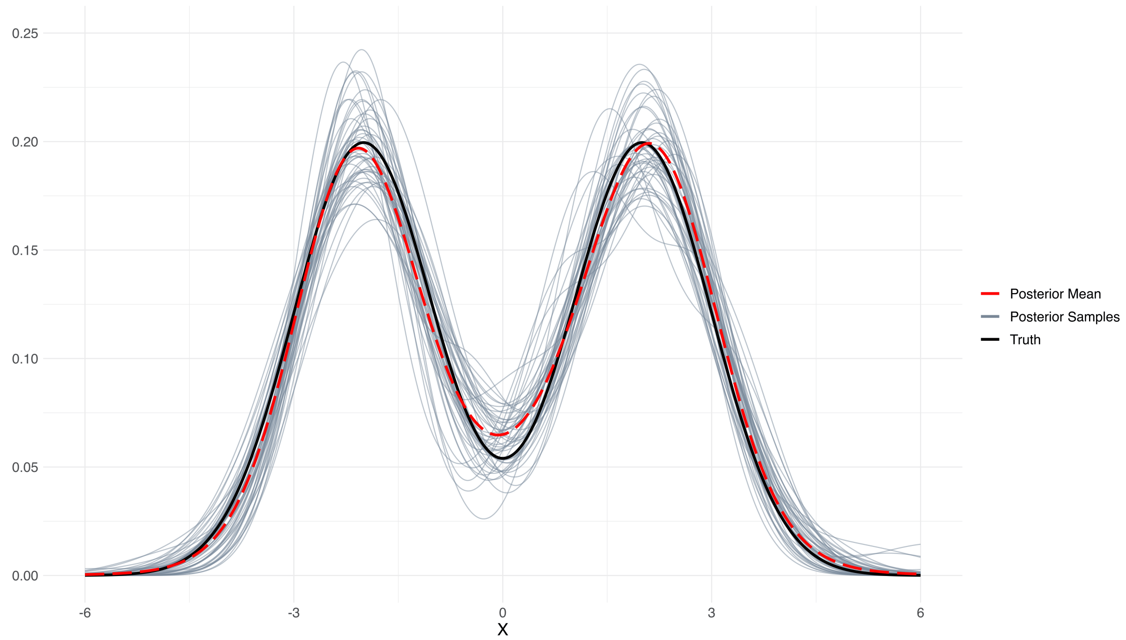


Figure 3: Deconvolution, $n = 1000$. Model : $X \sim 0.5\mathcal{N}(-2,1) + 0.5\mathcal{N}(2,1)$.

As Table 1 illustrates, the quasi-Bayes estimator uniformly outperforms alternatives. As a final thought, we note that the alternative estimators **HM**, **LV** and **BR** in Table 1 do not rely on a symmetric error distribution in their implementation, whereas traditional deconvolution and quasi-Bayes deconvolution use this fact to write the repeated measurements problem as a deconvolution problem through (64) and (65). In some situations, the assumption of symmetric errors may be too strong. In the interest of a more direct comparison, we also provide simulation results on the repeated measurements quasi-Bayes posterior, introduced in Section 3.2. This is provided below in Table 2.

5.2 Repeated Measurements

In this section, we examine the finite sample properties of the repeated measurements quasi-Bayes posterior, introduced in Section 3.2. To start with, we repeat the simulation study done in Table 1. There, the deconvolution quasi-Bayes posterior was constructed by transforming the repeated measurements (Y_1, Y_2) into a deconvolution problem through (64) and (65). Instead, in this section, we use the observations (Y_1, Y_2) directly to construct the repeated measurements quasi-Bayes posterior. The implementation details are as follows. For the base measure of the Dirichlet process prior DP_α we take $\alpha = \mathcal{N}(\widehat{\mu}, \widehat{\sigma}^2)$ where $(\widehat{\mu}, \widehat{\sigma}^2)$ are the sample mean and variance of the observations in (64). The standard deviation prior is taken to be Inv-Gamma$(2, 2)$. In all cases, we use $\mathbb{B}(T) = [-1, 1]$ as the set of identifying restrictions.

$\sqrt{n} \times$ MISE				
	HM	**LV**	**BR**	**quasi-Bayes (R)**
$\mathcal{N}(0, 1)$	0.285	0.250	0.155	0.033
Laplace$(0, 1)$	1.107	1.171	0.822	0.151
Gamma$(2, 1)$	0.981	1.107	0.759	0.307
Gamma$(5, 1)$	0.506	0.411	0.348	0.041
Log-Normal$(0, 1)$	7.595	7.271	11.067	1.851
$0.5\mathcal{N}(-2, 1) + 0.5\mathcal{N}(2, 1)$	1.962	1.993	1.519	0.980

Table 2: $n = 1000$. **HM**, **LV**, **BR** refer to the estimators proposed in Horowitz and Markatou (1996), Li and Vuong (1998) and Bonhomme and Robin (2010), respectively. The MISE for these estimators are taken from Table 2 in Bonhomme and Robin (2010). **quasi-Bayes (R)** refers to the repeated measurements quasi-Bayes (Section 3.2) posterior mean.

As Table 2 illustrates, the repeated measurements quasi-Bayes estimator uniformly outperforms alternatives. We found that it was possible to improve performance further by selecting T based on the specific design or Monte Carlo realization. We discuss the possibility of empirical (or hierarchical) Bayes selection of T in the conclusion (see Section 8).

Next, we consider a multivariate setup that closely resembles our empirical application in Section 6. The setup is as follows. We have two baseline non-negative inputs: child cognition C_1 and a covariate X at time $t = 1$. Cognition evolves over time as

$$\log C_2 = \alpha_1 + \delta_1 \log(C_1) + \delta_2 \log(X) + u_1 \,, \tag{66}$$

$$\log C_3 = \alpha_2 + \delta_3 \log(C_2) + \delta_4 \log(X) + u_2. \tag{67}$$

where (u_1, u_2) are unobserved errors and $\theta = (\alpha_1, \alpha_2, \delta_1, \delta_2, \delta_3, \delta_4)$ is an unknown parameter of interest. We do not observe (X, C_1, C_2, C_3) directly. Instead, for each variable, we observe three noisy measurements specified as follows.

$$m_{l,j} = \lambda_{i,j} \log C_i + \epsilon_{i,j} \qquad l, j = 1, 2, 3 \,, \tag{68}$$

$$\overline{m}_j = \gamma_j \log X + \varepsilon_j \qquad j = 1, 2, 3 \,, \tag{69}$$

where $\{\epsilon_{l,j}\}$ and $\{\varepsilon_j\}$ are mutually independent measurement errors. Here, the $\{\lambda_{l,j}\}$ and $\{\gamma_j\}$ are unknown factor loadings with the normalization that $\lambda_{i,1} = 1$ and $\gamma_1 = 1$. Intuitively, the remaining factor loadings determine the scale of measurement relative to the base measurement that corresponds to $j = 1$. Observe that, with three measurements, the remaining factor loadings can be estimated using the data as follows:

$$\widehat{\lambda}_{l,j} = \frac{\widehat{\mathrm{Cov}}(m_{l,j}, m_{l,j'})}{\widehat{\mathrm{Cov}}(m_{l,1}, m_{l,j'})} \quad , \quad \widehat{\gamma}_j = \frac{\widehat{\mathrm{Cov}}(\overline{m}_j, \overline{m}_{j'})}{\widehat{\mathrm{Cov}}(\overline{m}_1, \overline{m}_{j'})}. \tag{70}$$

for any j' satisfying $j' \neq j$ and $j' \neq 1$, where $\widehat{\mathrm{Cov}}(.)$ is the empirical covariance using the data. As such, up to the stochastic error from estimating the covariances, we observe noisy measurements $\widehat{Z}_{l,j} = \widehat{\lambda}_{l,j}^{-1} m_{l,j}$ and $\widetilde{Z}_j = \widehat{\gamma}_j^{-1} \overline{m}_j$ that satisfy the identity

$$\widehat{Z}_{l,j} = \log(C_l) + \bar{\epsilon}_{l,j} \;\; , \;\; \widetilde{Z}_j = \log(X) + \bar{\varepsilon}_j \; , \tag{71}$$

with unobserved measurement errors $\bar{\epsilon}_{l,j} = \lambda_{l,j}^{-1}\epsilon_{i,j}$ and $\bar{\varepsilon}_j = \gamma_j^{-1}\varepsilon_j$.

To estimate the parameter $\theta = (\alpha_1, \alpha_2, \delta_1, \delta_2, \delta_3, \delta_4)$ in (66) and (67), we proceed as follows.

1. Use the observations in (71) to compute the repeated measurements quasi-Bayes posterior, introduced in Section 3.2).
2. A sample of size m from this quasi-Bayes posterior consists of a sequence of Gaussian mixtures $\{\phi_{P_k,\Sigma_k}\}_{k=1}^m$. Use each Gaussian mixture ϕ_{P_k,Σ_k} to generate a synthetic data set $\mathcal{D}_k$.
3. Estimate the parameter θ using the synthetic data set $\mathcal{D}_k$ and the regression specifications in (66) and (67). This provides us with a (conditional on data set $\mathcal{D}_k$) estimate $\widehat{\theta}_k$.
4. As a formal estimator of θ, we use the posterior mean $\widehat{\theta} = m^{-1}\sum_{k=1}^m \widehat{\theta}_k$.

As a Monte Carlo exercise, we consider the following initial conditions:

$$(\log C_1, \log X) \sim 0.65\mathcal{N}(\mu_1, \Sigma_0) + 0.35\mathcal{N}(\mu_2, \Sigma_0) \; , \tag{72}$$

$$\mu_1 = (-4, 2) \;\; , \;\; \mu_2 = (6, 3) \; , \;\; \Sigma_0 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The evolution over time is specified by

$$\log C_2 = 1 + 0.5\log(C_1) + 0.5\log(X) + u_1 \;\; , \; u_1 \sim \mathcal{N}(0,1) \tag{73}$$

$$\log C_3 = 1 + 0.5\log(C_2) + 0.5\log(X) + u_2 \;\; , \; u_2 \sim \mathcal{N}(0,1). \tag{74}$$

This corresponds to a true parameter $\theta = (1, 1, 0.5, 0.5, 0.5, 0.5, 0.5)$.

For (68) and (69), we take the factor loadings to be $\lambda_{j'} = \gamma_{j'} = 0.5$ for $j' \neq 1$. All the measurement errors in (68) and (69) are taken to be independent $\mathcal{N}(0, 1)$ noise. Observe that, with such a choice, the measurement errors of the observed proxies in (71) are $\mathcal{N}(0, 4)$ for $j' \neq 1$. Table 3 provides the mean and standard deviation of the quasi-Bayes posterior mean for each of the coefficients appearing in (73) and (74).

	$t = 2$ Coefficients			$t = 3$ Coefficients		
	α_1	δ_1	δ_2	α_2	δ_3	δ_4
True	1	0.5	0.5	1	0.5	0.5
Mean	1.022	0.502	0.491	1.002	0.499	0.497
Standard Deviation	0.140	0.011	0.059	0.136	0.021	0.072

Table 3: Mean and Standard Deviation of quasi-Bayes posterior mean.

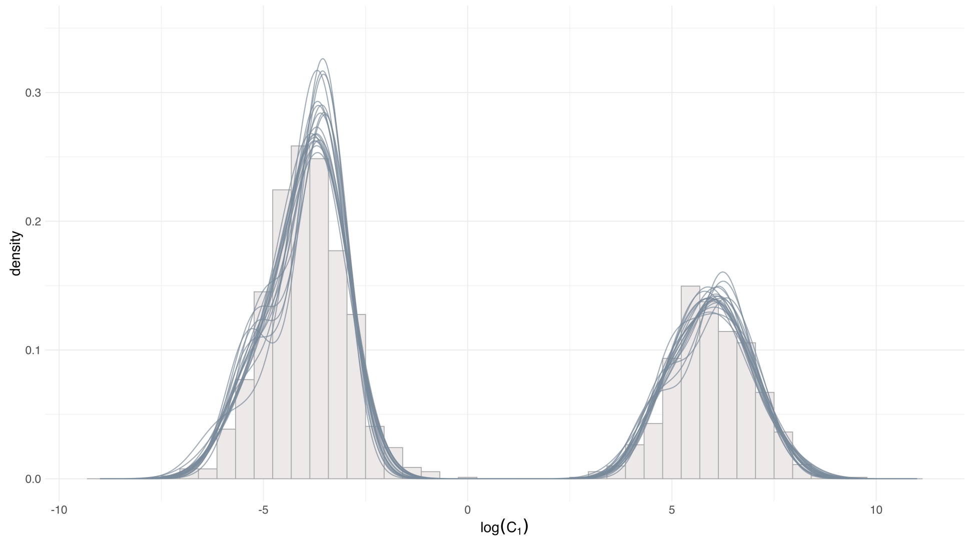


Figure 4: Sample realization of quasi-posterior samples for the latent distribution of $\log(C_1)$, relative to true latent histogram.

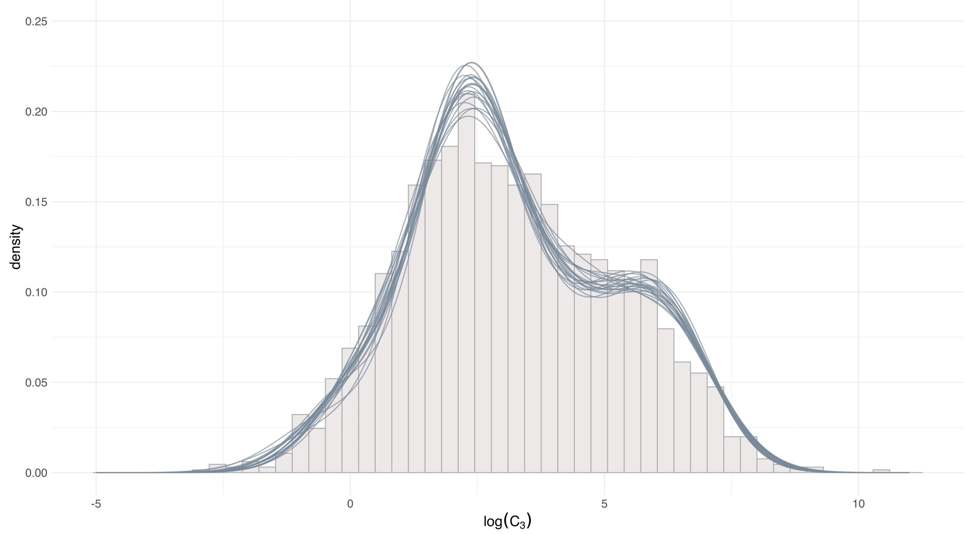


Figure 5: Sample realization of quasi-posterior samples for the latent distribution of $\log(C_3)$, relative to true latent histogram.

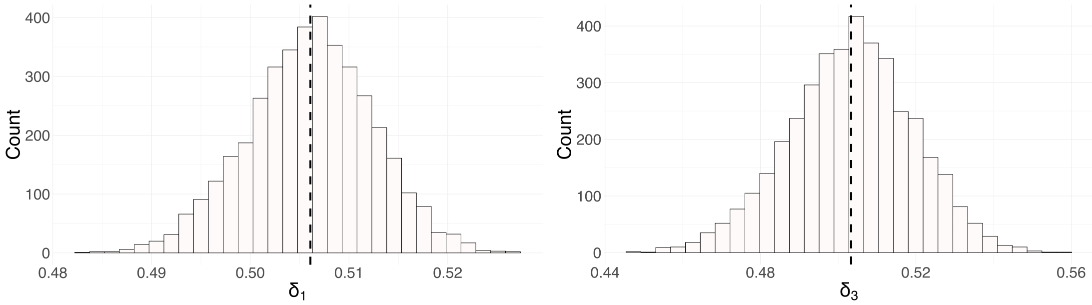


Figure 6: Sample realization of the quasi-posterior distribution of δ_1, δ_3. Black dashed line indicates posterior mean.

6 Empirical Illustration I (Human Capital Development)

In this section, we apply our methodology to study human capital development for young children in India. Our analysis closely follows Attanasio et al. (2020b). Using data collected from the Young Lives Project, Attanasio et al. (2020b) estimate production functions for cognition and health for young children in India. Their analysis makes use of the latent factor approach initially introduced in Cunha et al. (2010).[31] As measurements on production function inputs and child outcomes are not directly observed, a first step in their analysis specifies and estimates[32] a parametric form for the joint distribution of unobserved and observed variables. We apply our methodology to nonparametrically estimate this joint distribution. The joint distribution is then used to estimate production functions for cognition and health.

6.1 Data and Model

We begin with a brief summary of the data and empirical framework used in Attanasio et al. (2020b). The sample consists of $2,011$ children from seven districts[33] and 98 separate communities. Their families were surveyed at child ages $1, 5, 8$ and 12. The data consists of information obtained from household, child and community questionnaires. The households in the sample are relatively poor, with over half of all respondants living on less than 2$ per day. We restrict the sample to children observed in all rounds, which leads to a sample size of $n = 1910$ children.

We observe multiple measurements on child cognition, child health, parental investment, parental cognition, parental health and parental resources.[34] If $m_{j,k,t}$ denotes the j^{th} measurement of

[31] They also make use of additional identification results in Agostinelli and Wiswall (2016)

[32] To estimate the model, they assume the joint distribution follows a two component Gaussian mixture and impose normality restrictions on the measurement errors. These restrictions induce a Gaussian mixture likelihood on the observed measurements. Estimation then follows from applying the Expected Maximization (EM) algorithm.

[33] These are Hyderabad and a "poor" and "nonpoor" district in Coastal Andhra, Rayalaseema, and Telangana.

[34] The precise list of measurements can be found in Table 4 of Attanasio et al. (2020b).

latent factor $\theta_{k,t}$ at time t, we assume a log linear relationship of the form

$$m_{j,k,t} = a_{j,k,t} + \lambda_{j,k,t} \log(\theta_{k,t}) + \epsilon_{j,k,t}. \tag{75}$$

Here, the $a_{j,k,t}$ and $\lambda_{j,k,t}$ are fixed constants and factor loadings, respectively. The $\epsilon_{j,k,t}$ are unobserved mean zero errors that are independent of each other and the latent factors. At each age, we also observe (without measurement error) prices for food, clothing, notebooks and worm medication (Mebendazol).[35]

As the units of measurement typically differ across the measurements, the factor loadings determine how values in one measurement relate to another. To achieve identification, each latent factor is chosen to have a reference measurement that has unit factor loading. For factors θ_k that do not vary over time (such as parental cognition and health), we work under the normalization that $\mathbb{E}[\log \theta_k] = 0$. For dynamic factors that vary over time, we normalize the mean only at the initial age of observation. The mean in future periods in identified relative to the initial period by assuming that the growth in measurements is due only to the growth of the latent factor.[36] Under these restrictions, the constants $a_{j,k,t}$ and $\lambda_{j,k,t}$ can be recovered through means and covariances of the observed measurements.[37] By using empirical analogs of these means and covariances, we obtain consistent estimates $\widehat{a}_{j,k,t}, \widehat{\lambda}_{j,k,t}$. In particular, up to a negligible stochastic error of order $n^{-1/2}$ that arises from estimating the means and covariances, we can view

$$Z_{j,k,t} = \widehat{\lambda}_{j,k,t}^{-1}(m_{j,k,t} - \widehat{a}_{j,k,t}) \tag{76}$$

as repeated measurements of the latent factors $\log(\theta_{k,t})$.

The parental investment specification is given by

$$\begin{aligned} \log \theta_{I,t} = \gamma_0 + \gamma_{c,t} \log \theta_{c,t} + \gamma_{h,t} \log \theta_{h,t} + \gamma_{cp} \log \theta_{cp} \\ + \gamma_{hp} \log \theta_{hp} + \gamma_{pt} \log(p_t) + \gamma_{I,t} \log(\theta_{Y,t}) + \epsilon_t. \end{aligned} \tag{77}$$

Here, ϵ_t represents an unobserved mean zero shock. $\theta_{c,t}$ and $\theta_{h,t}$ represent child cognition and health at time t, respectively, while θ_{cp} and θ_{hp} denote parental cognitive and health attributes, both assumed to be invariant with time. The variable p_t stands for the prices of goods as described above, and $\theta_{Y,t}$ represents parental resources at time t.

If $\theta_{I,t}$ represents parental investment at time t and the variables are as described above, we model the production functions as

$$\log \theta_{k,t+1} = \beta_t + \frac{1}{\rho_t} \log \left(\delta_{c,t} \theta_{c,t}^{\rho_t} + \delta_{h,t} \theta_{h,t}^{\rho_t} + \delta_{cp,t} \theta_{cp}^{\rho_t} + \delta_{hp,t} \theta_{hp}^{\rho_t} + \delta_{I,t} \theta_{I,t}^{\rho_t} \right) + \nu_{k,t} \tag{78}$$

[35] For variables without measurement error, the associated factor loading is set to 1 and the measurement error to zero, so that the observed variable serves as its own repeated measurement.

[36] For further discussion on these normalization restrictions, see Attanasio et al. (2020b).

[37] In cases with more measurements than needed for identification, multiple combinations of covariances can identify the same factor loading. In this case, as in Agostinelli and Wiswall (2016), we average across them.

where $k \in \{c, h\}$, $\{\delta_{l,t}\} \in [0, 1]$, $\rho_t \in (-\infty, 1]$ and $\sum_l \delta_{l,t} = 1$. Here, $\nu_{k,t}$ is an unobserved mean zero shock. The parameter ρ_t determines the elasticity of substitution between inputs to the production functions. At $\rho_t = 0$, the production function reduces to Cobb-Douglas (unit elasticity) while $\rho_t = 1$ corresponds to perfectly substitutable inputs.

We aim to estimate the child production function for each time period t. We proceed as follows. First, we use the repeated measurements in (76) to estimate the quasi-Bayes posterior in (27). Given a sample distribution from this posterior, we simulate a synthetic data set and use it to estimate the parameters appearing on the right side of (78). This leads to an induced posterior distribution on the parameters. We then use the posterior mean to obtain point estimates of the parameters.

A concern towards identification of the production function parameters is that parents may choose their investment in part based on the evolution of human capital. That is, parental investment $\theta_{I,t}$ may be endogenous. Following (Attanasio et al., 2020b), we correct for this through a control function approach.[38] In our context, given a sample distribution from the posterior and a corresponding synthetic data set, this is achieved by using the residuals from (77) as an extra regressor when estimating the production functions in (78).

6.2 Implementation

We use the measurements in (76) to construct the gradient-based quasi-Bayes posterior introduced in Section 3.2. In this case, the joint distribution that we estimate is $d = 27$ dimensional. The posterior is defined using $T = 0.5$ so that the grid of restrictions is given by $\mathbb{B}(T) = \{t \in \mathbb{R}^d : \|t\|_\infty \leq 0.5\}$. As a base measure for the Dirichlet process prior DP_α, we take $\alpha = \mathcal{N}(\mathbf{0}, 5I_d)$. We place a prior on covariance matrices through a correlation and scale matrix individually. That is, $\Sigma = DCD$ where D is a diagonal scale matrix and C is a correlation matrix. We use an LKJ prior (Lewandowski et al., 2009) with shape parameter $\eta = 2$ for C and independent Inv-Gamma$(2, 2)$ priors for each element of the diagonal of D.

6.3 Results

We start by examining some aspects of the marginal structure of the posterior distribution. Specifically, we look at the distributional evolution of child cognition and parental investment over the development period.

[38] This approach assumes that $\mathbb{E}[\nu_{k,t}|\mathcal{X}_t, Z_t] = \kappa_{k,t}\epsilon_t$ for some constants $\kappa_{k,t}$, where $\mathcal{X}_t$ is the variables in the production functions and Z_t are the instruments (prices and household resources) that only appear in the investment equation.

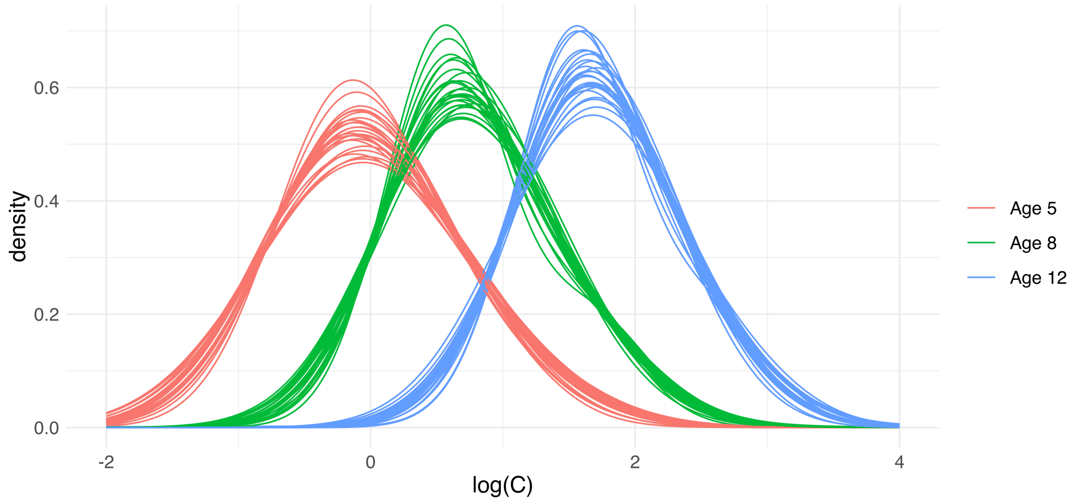


Figure 7: Posterior samples for latent log cognition across the development period.

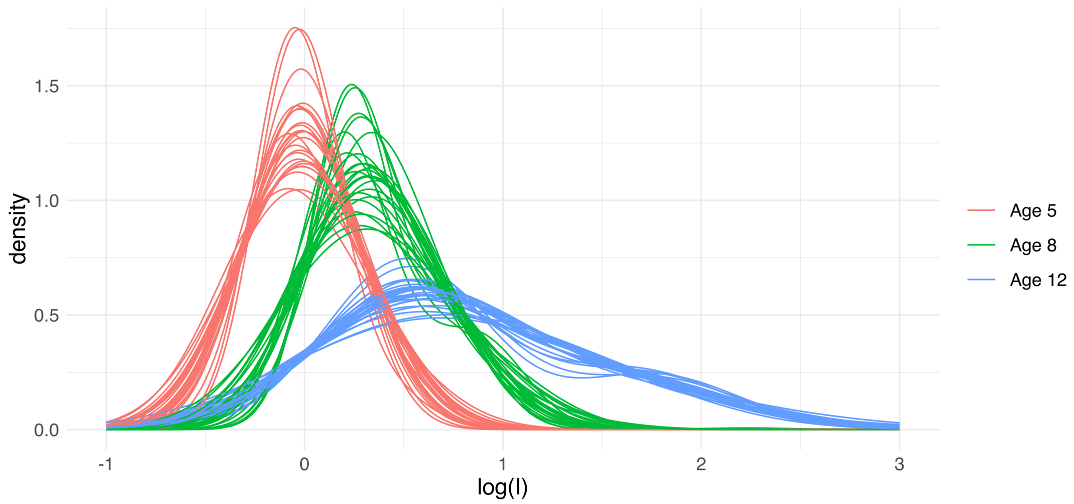


Figure 8: Posterior samples for latent log investment across the development period.

As expected, mean cognitive skills and investment increases over the development period. Cognitive dispersion does seem to vary significantly over the development period. This is in contrast to similar studies using U.S data (e.g. Agostinelli and Wiswall, 2016) where dispersion increases significantly at later ages. In U.S based studies with a representative sample, dispersion is linked to income inequality and available opportunities for skill growth. By contrast, households in this sample are relatively poor and the children have limited access to such opportunities. It is interesting that investment displays significant right skewness at age 12. This was also the case with the observed measurements in the data. Some parents may choose to invest more heavily in their children during adolescence as opposed to early childhood. This can be influenced by various factors, including cultural beliefs and the perceived needs of the child at different stages of development.

	Age 5	Age 8	Age 12
Cognition (Lagged)	-	0.243 $[0.065, 0.426]$	0.435 $[0.184, 0.691]$
Investment	0.572 $[0.429, 0.710]$	0.406 $[0.230, 0.578]$	0.421 $[0.125, 0.713]$
Parental Cognition	0.304 $[0.184, 0.428]$	0.201 $[0.070, 0.331]$	0.062 $[0, 0.193]$
Health (Lagged)	0.102 $[0, 0.228]$	0.122 $[0, 0.291]$	0.072 $[0, 0.225]$
Parental Health	0.022 $[0, 0.099]$	0.029 $[0, 0.109]$	0.011 $[0, 0.064]$
Elasticity ρ_t	0.168 $[-0.225, 0.624]$	0.015 $[-0.359, 0.388]$	-0.284 $[-0.834, 0.150]$

Table 4: Production function for cognitive skills, with 90% Bayesian Credible Band.

Table 4 provides posterior mean estimates for the child production function with the corresponding 90% pointwise Bayesian credible bands. We note that the credible bands in this Table and elsewhere are provided only to convey a general sense of Bayesian uncertainty. In particular, they are not meant to be interpreted in the usual sense of frequentist standard errors.[39]

In line with previous results in the literature (e.g. Attanasio et al. 2020b), our results suggest that (*i*) cognition displays persistence across the development period, (*ii*) parental investment affects cognitive development at all ages, with a higher return for younger children and (*ii*) a link between parental cognition and child cognition, although the effect fades out over time. Overall, our results share many similarities. The main difference in our empirical findings relative to the literature is that investment continues to maintain a significant impact during adolescence, albeit lower than in early childhood. As our sample size in this setting is relatively modest, and we do not consider estimation by relaxing only a single restriction, it is difficult to ascertain whether our differences are due to nonparametrically modeling the latent distribution and/or leaving free the distribution of the measurement errors. We hope to examine this more closely in future work. To that end, we view our general methodology as complementary to the existing literature in that it may serve as a robustness check to possible violations of these restrictions.

[39]More formally, our paper only investigates estimation. Inferential results which establish the frequentist validity of certain credible intervals would at the very least require the quasi-Bayes objective function to be optimally-weighted. As the quasi-Bayes objective functions is an L^2 norm over characteristic function based moment restrictions, optimal weighting in this context closely resembles the definition proposed in (Carrasco and Florens, 2000).

	Age 5	**Age 8**	**Age 12**
Resources	0.096	0.209	0.567
	$[-0.101, 0.297]$	$[-0.067, 0.490]$	$[0.189, 0.955]$
Parental Cognition	0.05	0.102	0.148
	$[-0.070, 0.169]$	$[-0.026, 0.232]$	$[-0.056, 0.355]$
Cognition (Lagged)	-	0.038	0.101
		$[-0.104, 0.183]$	$[-0.153, 0.359]$
Health (Lagged)	0.007	0.012	0.028
	$[-0.075, 0.085]$	$[-0.146, 0.168]$	$[-0.212, 0.276]$
Price Clothes	0.011	-0.024	-0.006
	$[-0.073, 0.092]$	$[-0.120, 0.071]$	$[-0.153, 0.138]$
Price NoteBook	-0.006	-0.001	0.055
	$[-0.080, 0.069]$	$[-0.113, 0.111]$	$[-0.076, 0.186]$
Price Mebendazol	0.005	-0.041	-0.022
	$[-0.058, 0.069]$	$[-0.145, 0.065]$	$[-0.123, 0.079]$
Price Food	0.008	0.025	-0.012
	$[-0.080, 0.093]$	$[-0.095, 0.148]$	$[-0.110, 0.084]$

Table 5: Posterior means of the reduced form Investment Regression (77) Coefficients, with 90% Bayesian Credible Band.

The results from the reduced form investment regression suggest that investment is largely linked to parental resources and to a lesser extent, parental cognition. While in principle one would expect to observe possible links to child cognition, it is important to note that investment is measured 3-4 years after lagged cognition. In this case, the effect may be substantially weaker. Unfortunately, our data set does not contain information on shorter time spans to be able to adequately discern the immediate effect.

Next, we use the estimated quasi-posterior to perform two counterfactuals. The first is a one time income transfer of a fraction of the observed empirical mean income at age 5. The second is a one time increase in health of a fraction of one standard deviation of health at age 1. Note that the income transfer counterfactual uses a fixed transfer that does not vary based on the sample drawn from the quasi-posterior. The health intervention is specific to each individual sample of the quasi-posterior as one standard deviation of health depends on the distribution of health for that sample. We analyze the effects separately for the poorest 25%, the middle 50% and the richest 25%. The counterfactual outcomes are shown in Figures 9 and 10.

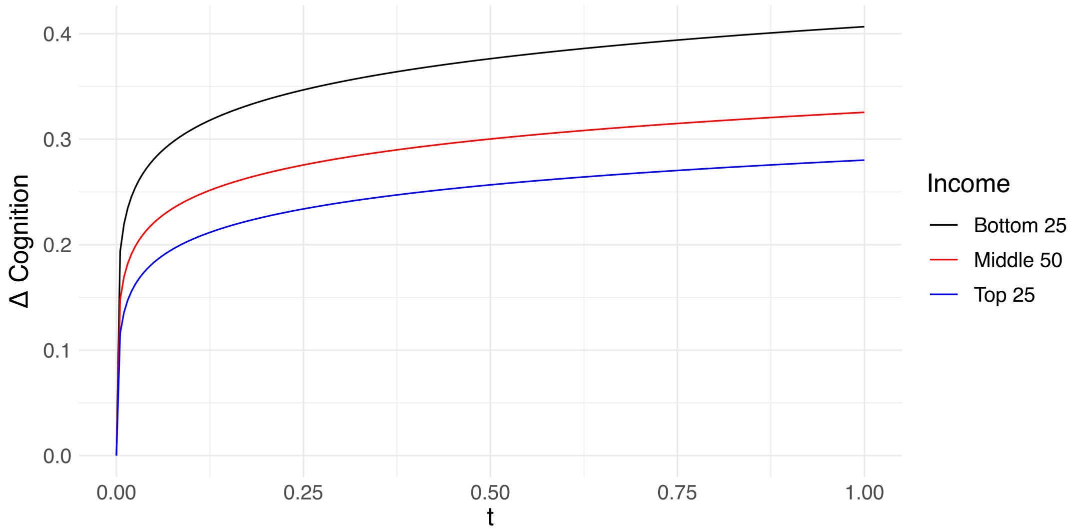


Figure 9: Posterior mean of change in cognition at age 12 (in units of standard deviation). Counterfactual: age 5 income transfer of $t \times \mu$ where μ is the observed empirical mean income at age 5.

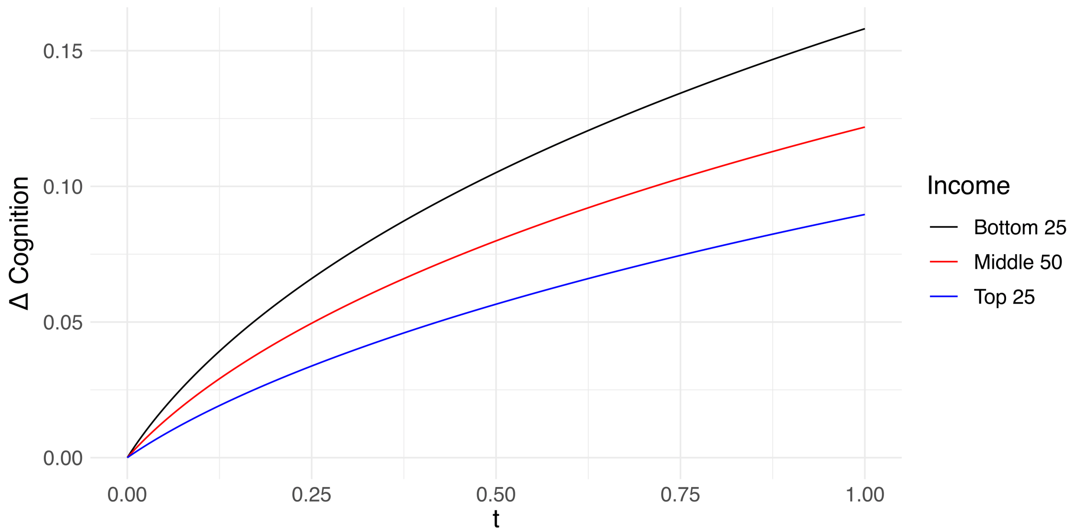


Figure 10: Posterior mean of change in cognition at age 12 (in units of standard deviation). Counterfactual: Age 1 Health intervention of $t \times \mu$ where μ is the standard deviation of health at age 1.

Observe that, as our entire sample is relatively poor, income transfers are expected to matter at all percentiles of the income distribution. Moreover, as illustrated in Figures 9 and 10, the poorer segments of society benefit the greatest from an income transfer or health intervention. Our findings suggest that early health interventions or parental subsidies may be an important element in policies designed to target human capital development in developing countries.

7 Empirical Illustration II (Earnings Dynamics)

In this section, we apply our methodology to study the latent structure of permanent and transitory components in the Panel Study of Income Dynamics (PSID) data. Our analysis closely follows Bonhomme and Robin (2010) in which deconvolution-based methods were used to nonparametrically estimate the latent structure.

7.1 Data and Model

We begin with a brief summary of the data and model used in Bonhomme and Robin (2010). The data is from the PSID, between 1978 and 1987. Let $y_{i,t}$ denotes annual log earnings for individual i at time period t and $x_{i,t}$ an associated set of regressors. The regressors are a quadratic polynomial in age and indicators for education, race, geography and year. The OLS residuals of $y_{i,t}$ on $x_{i,t}$ are denoted by $w_{i,t}$. The residual differences are denoted by $\Delta w_{i,t} = w_{i,t} - w_{i,t-1}$. After restricting the sample to male workers with no missing observations for $\Delta w_{i,t}$ and a wage growth that does not exceed 150% in absolute value, our sample size consists of $n = 624$ individuals. For each individual, we observe wages between 1978 and 1987, for a total of $M = 10$ time periods.

The model in (Bonhomme and Robin, 2010) is given by[40]

$$
\begin{aligned}
w_{i,t} &= f_i + w^P_{i,t} + w^T_{i,t} \; , \qquad i = 1, \dots, n \; , \; t = 1, \dots, M \\
w^P_{i,t} &= w^P_{i,t-1} + \epsilon_{i,t} \; , \\
w^T_{i,t} &= \eta_{i,t} \\
\eta_{i,1} &= \eta_{i,M} = 0.
\end{aligned} \tag{79}
$$

Here, f_i is an individual level fixed effect and $\{\epsilon_{i,t}\}_{t=1}^M$ and $\{\eta_{i,t}\}_{t=2}^{M-1}$ are mean zero errors. We think of $w^P_{i,t}$ as the permanent component and $\eta_{i,t}$ as the transitory component. Thus, the distribution of the permanent component of income is determined by $\epsilon_{i,t}$ and the distribution of the transitory component by $\eta_{i,t}$.

From first differencing the model in (79), we can write

$$
\Delta w_{i,t} = \epsilon_{i,t} + \eta_{i,t} - \eta_{i,t-1} \; , \qquad i = 1, \dots, n \; , \; t = 2, \dots, M. \tag{80}
$$

We view $\mathbf{Y}_i = (\Delta w_{i,2}, \dots, \Delta w_{i,M})$ as the observations for $i = 1, \dots, n$. The model in (80) is a special case of the multi-factor model $\mathbf{Y}_i = \mathbf{A}\mathbf{X}_i$ in Section 3.3. Here, $\mathbf{A}$ is a known matrix with elements in $\{-1, 0, 1\}$ and the latent factors are $\mathbf{X}_i = \{\eta_{i,2}, \dots, \eta_{i,M-1}, \epsilon_{i,2}, \dots, \epsilon_{i,M}\}$.

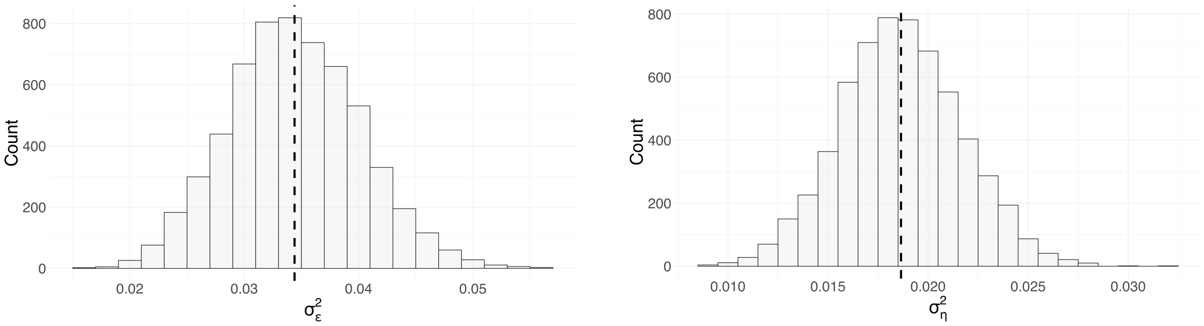
7.2 Implementation

We use the measurements in (80) to construct the joint multi-factor quasi-Bayes posterior introduced in Section 3.3. Over the $M = 10$ time periods, there are 17 (the dimension of $\mathbf{X}_i$)

[40] Similar models also appear in Abowd and Card (1989); Geweke and Keane (2000); Hall and Mishkin (1982); Horowitz and Markatou (1996).

latent distributions that must be estimated. The posterior is defined using $T = 0.5$ so that the grid of restrictions is given by $\mathbb{B}(T) = \{t \in \mathbb{R}^d : \|t\|_\infty \leq 0.5\}$. As the observed measurements in (80) are themselves residuals from an OLS regression, they are demeaned and exhibit low finite sample variance (see Table 6 below). For the base measure α of the Dirichlet process prior DP_α, we take $\alpha = \mathcal{N}(0, 1)$. Given the dispersion in the data, we view this as a relatively weak prior. For the standard deviation, we use Inv-Gamma$(0.01, 0.01)$ priors. This is a commonly used as non informative prior in settings with low dispersion.[41]

7.3 Results



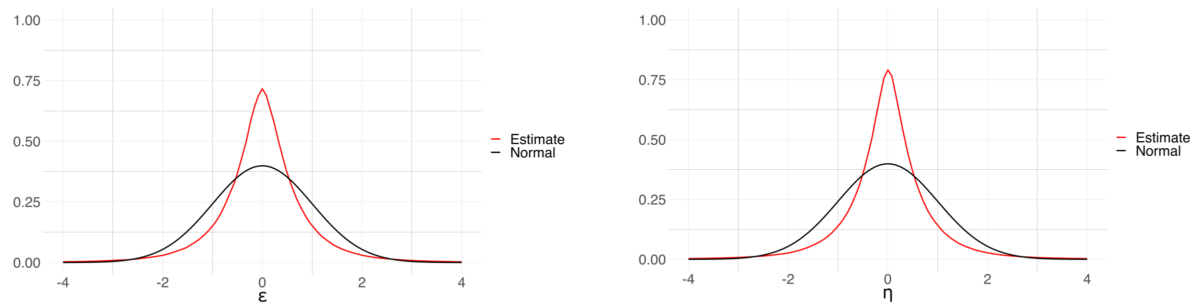
(a) Permanent shock ϵ

(b) Transitory shock η

Figure 11: Posterior Histograms for the average σ^2_ϵ and σ^2_η across the time period. Black dashed line indicates posterior mean.

The posterior mean of the average variance is $\widehat{\sigma}^2_\epsilon = 0.034$ and $\widehat{\sigma}^2_\eta = 0.019$ for the permanent and transitory components, respectively. Interestingly, our estimate for the transitory component variance matches the estimate obtained in (Bonhomme and Robin, 2010). However, our estimate for the permanent component variance is larger (they obtained $\widehat{\sigma}^2_{\epsilon,BR} = 0.0208$). As a consequence, we find a larger variance share being attributed to permanent shocks. To be specific, our estimates suggest that permanent shocks account for 47% of the total variance of wage growth residuals. Next, we examine the distribution of the permanent and transitory components.

[41]The general idea is that Inverse-Gamma priors place exponentially small mass near zero and so choosing Inv-Gamma(ϵ, ϵ) for a sufficiently small ϵ places the mode of the distribution near zero to accomodate low dispersion settings. Gelman (2006) argues that a more robust choice is to use $\sigma \sim$ Half-Cauchy$(0, V)$ for a large V. The intuition being that Half-Cauchy priors have polynomial tails near zero (hence they assign enough mass around zero) and a large V provides robustness against large dispersion. As a robustness check, we also tried Half-Cauchy$(0, 100)$ priors and did not find any significant differences in the results.



(a) Permanent shock ϵ

(b) Transitory shock η

Figure 12: Posterior mean of the average standardized f_η and f_ϵ across the time period, with standard Gaussian overlay.

As Figure 12 illustrates, neither of the two components are Gaussian. As in Bonhomme and Robin (2010), the estimated distributions appear to be leptokurtic and symmetric. We note that our estimates do not have the tail wigglyness that usually accompanies deconvolution based estimators. As a consequence, even though the estimated variance for the transitory component is similar, the generalized deconvolution estimator in Bonhomme and Robin (2010) has a taller mode at zero (to balance the wigglyness at the tails). Next, we compare the difference in fit across various estimators.

	Wage growth $\Delta w_{i,t}$				
	Data	**BR**	**Normal**	**Normal Mixture**	**quasi-Bayes**
Variance	0.055	0.037	0.057	0.058	0.067
Skewness	0.001	-0.02	0.00	0.00	0.00
Kurtosis	10.158	5.600	3.000	6.30	8.223

Table 6: Average moments of $\Delta w_{i,t}$ across the time period. **BR** denotes the generalized deconvolution estimator in (Bonhomme and Robin, 2010). **Normal** and **Normal Mixture** denote maximum likelihood estimates when the permanent and transitory components follow a normal or two-component normal mixture. **quasi-Bayes** denotes the joint multi-factor quasi-Bayes (Section 3.3) posterior mean.

Table 6 compares the observed moments of the wage growth residuals with the implied moments under various estimators. Observe that, as wage data typically exhibits many outliers, the kurtosis in the observed data is quite large (relative to standard normality). The two component normal mixture maximum likelihood estimate and quasi-Bayes appear to be the closest fits to the data. The maximum likelihood estimate fits the variance exactly but performs worse at higher moments. By contrast, the quasi-Bayes approach returns a larger variance but obtains a better fit for the observed kurtosis in the data.

8 Conclusion

This paper developed a quasi-Bayes framework for three classes of latent variable models. In each case, we used the models identifying restrictions to construct a characteristic function based quasi-likelihood. We then combined this with a prior to induce a quasi-Bayes posterior. As the prior is supported on infinite Gaussian mixtures, our modeling framework connects with a large empirical literature (e.g. Geweke and Keane 2000; Cunha et al. 2010; Attanasio et al. 2020a,b) that utilizes finite Gaussian mixtures to model rich forms of heterogeneity. Simulation and empirical exercises demonstrate that our quasi-Bayes procedures are viable and perform favorably relative to existing alternatives. We end this section by highlighting some future directions for research.

In this paper, we provided first steps towards a characteristic function based quasi-Bayes framework for latent variable models. The results could be extended to several strands of literature that make use of identification arguments in characteristic space. Possible extensions include nonlinear regression with measurement error (Hausman, Newey, Ichimura, and Powell, 1991; Schennach, 2004a), instrumental variable based approaches to measurement error (e.g. Schennach, 2007), random coefficient models (e.g. Hoderlein, Klemelä, and Mammen, 2010; Gautier and Kitamura, 2013) and nonlinear earnings dynamics (Arellano, Blundell, and Bonhomme, 2017). These extensions are currently in progress.

The theory and implementation of our quasi-Bayes procedures could be expanded upon in a variety of ways. In practice, one may choose a varying radius grid $\mathbf{T} = (T_1, \ldots, T_d)$ and consider the set of restrictions $\mathbb{B}(\mathbf{T}) = \{t \in \mathbb{R}^d : |t_i| \leq T_i \quad i = 1, \ldots, d\}$. Empirical Bayes selection of priors and tuning parameters that our procedures depend on is an important topic to be investigated. One possibility is to choose the parameters to maximize the marginal quasi-likelihood. Another possible avenue is to consider a hierarchical Bayes setup where the prior hyperparameters are themselves modelled using a prior. A more challenging task would be to establish the frequentist validity of quasi-Bayes credible intervals. This would, at the very least, require the quasi-Bayes objective function to be optimally-weighted. As the quasi-Bayes objective functions is an L^2 norm over characteristic function based moment restrictions, optimal weighting in this context closely resembles the definition proposed in Carrasco and Florens (2000). We hope to address all these issues in future work.

9 Appendix : Proofs

The following notation is frequently referred to in the proofs and so is listed here for convenience. The imaginary unit is denoted by $\mathbf{i}$. Let λ denote the Lebesgue measure on $\mathbb{R}^d$. We denote the diameter of a set $A \subset \mathbb{R}^d$ by $\mathcal{D}(A) = \sup_{x,y\in A}\|x-y\|$. Given a positive definite matrix $\Sigma \in \mathbf{S}_+^d$, we denote the ordered eigenvalues by $\lambda_1(\Sigma) \le \ldots \le \lambda_d(\Sigma)$. To indicate the dependence of the quasi-posterior on the data $\mathcal{Z}_n$ and T (the radius of the ball $\mathbb{B}(T)$), we will often use the notation $\nu(.|\mathcal{Z}_n, T)$.

Lemma 1. *Given a random vector $Z \in \mathbb{R}^d$ with $\mathbb{E}(\|Z\|^2) < \infty$ and $t \in \mathbb{R}^d$, define*

$$\chi_t(Z) = e^{\mathbf{i}t'Z} = \cos\big(t'Z\big) + \mathbf{i}\sin\big(t'Z\big).$$

Then, there exists a universal constant $D > 0$ such that for every $T > 0$ we have

$$\mathbb{E}\bigg(\sup_{\|t\|_\infty \le T} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]|\bigg) \le D\frac{\max\{\sqrt{\log T}, 1\}}{\sqrt{n}}.$$

Proof of Lemma 1. It suffices to show the result for the real and imaginary part separately. We verify it for the real part, the imaginary part is completely analogous. For the remainder of this proof, we continue assuming $\chi_t(Z) = \cos(t'Z)$. Let $\mathcal{F} = \{\chi_t(Z) : |t| \le T\}$. By an application of (Giné and Nickl, 2021, Remark 3.5.14), there exists a universal constant $L > 0$ such that

$$\mathbb{E}\bigg(\sup_{\|t\|_\infty \le T} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]|\bigg) \le \frac{L}{\sqrt{n}}\int_0^8 \sqrt{\log N_{[]}(\mathcal{F}, \|.\|_{L^2(\mathbb{P})}, \epsilon)}d\epsilon.$$

Let $\{t_i\}_{i=1}^M$ denote a minimal $\delta > 0$ covering of $[-T, T]^d$. Define the functions

$$e_i(Z) = \sup_{t\in\mathbb{R}^d : \|t\|_\infty \le T, \|t-t_i\|_\infty < \delta} |\ \chi_t(Z) - \chi_{t_i}(Z)\ | \qquad i = 1, \ldots, M.$$

It follows that $\{\chi_{t_i}(Z) - e_i\ ,\ \chi_{t_i}(Z) + e_i\}_{i=1}^M$ is a bracket covering for $\mathcal{F}$. Since the mapping $t \to \chi_t(Z)$ has Lipschitz constant bounded by $\|Z\|$, we obtain $\|e_i\|^2_{L^2(\mathbb{P})} \le \mathbb{E}(\|Z\|^2)\delta^2$. Since $M \le (3T\delta^{-1})^d$, it follows that there exists a universal constant $L > 0$ such that

$$\int_0^8 \sqrt{\log N_{[]}(\mathcal{F}, \|.\|_{L^2(\mathbb{P})}, \epsilon)}d\epsilon \le L\max\{\sqrt{\log T}, 1\}.$$

□

Lemma 2. *Given a random variable $Z \in \mathbb{R}^d$ with $\mathbb{E}(\|Z\|^2) < \infty$ and $t \in \mathbb{R}$, define*

$$\chi_t(Z) = e^{\mathbf{i}t'Z} = \cos\big(t'Z\big) + \mathbf{i}\sin\big(t'Z\big).$$

Then, there exists a universal constant $L > 0$ such that for any sequence $T = T_n \uparrow \infty$ with

$\log(T_n) \lessapprox n$*, we have that*

$$\mathbb{P}\bigg(\sup_{|t|\le T_n} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]| \le D\frac{\sqrt{\log T_n}}{\sqrt{n}} \bigg) \to 1.$$

Proof of Lemma 2. It suffices to show the result for the real and imaginary part separately. We verify it for the real part, the imaginary part is completely analogous. For the remainder of this proof, we continue assuming $\chi_t(Z) = \cos(t'Z)$. Let $\mathcal{F} = \{\chi_t(Z) : |t| \le T_n\}$. Since $\mathcal{F}$ is uniformly bounded, (Giné and Nickl, 2021, Theorem 3.3.9) and Lemma 1 imply that there exists universal constants $C, D > 0$ which satisfy

$$\mathbb{P}\bigg(\sup_{|t|\le T_n} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]| > D\frac{\sqrt{\log T_n}}{\sqrt{n}} + x \bigg) \le \exp\bigg(- \frac{x^2}{C[1+\sqrt{\log T_n}/\sqrt{n}\,]n^{-1}} \bigg)$$

for all $x > 0$. With $x = \sqrt{\log T_n}/\sqrt{n}$ and observing that $\sqrt{\log T_n} \lessapprox \sqrt{n}$, we obtain

$$\mathbb{P}\bigg(\sup_{|t|\le T_n} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]| \le L\frac{\sqrt{\log T_n}}{\sqrt{n}} \bigg) \to 1$$

for some universal constant $L > 0$. □

Lemma 3. *Consider a measurable partition* $\mathbb{R}^d = \bigcup_{j=0}^N V_j$ *and points* $z_j \in V_j$ *for* $j = 1, \dots, N$*. Let* $F^* = \sum_{j=1}^N w_j \delta_{z_j}$ *denote the discrete probability measure with weight* w_j *at* z_j*. Then, for any probability measure* F *on* $\mathbb{R}^d$ *with* $\int_{\mathbb{R}^d} \|x\|^2 dF(x) < \infty$*, we have that*

1.
$$|\mathcal{F}[F](t) - \mathcal{F}[F^*](t)| \le \|t\| \sup_{j=1,\dots,N} \mathcal{D}(V_j) + 2\sum_{j=1}^N |F(V_j) - w_j|\,.$$

2.
$$\begin{aligned}\|\nabla\mathcal{F}[F](t) - \nabla\mathcal{F}[F^*](t)\| \le \bigg(\int_{\mathbb{R}^d} \|x\|^2 dF(x) \bigg)^{1/2} \bigg\{ \sum_{j=1}^N |F(V_j) - w_j| \bigg\}^{1/2} \\ + \bigg(1 + \sup_{v\in\bigcup_{j=1}^N V_j} \|v\|\|t\| \bigg) \sup_{j=1,\dots,N} \mathcal{D}(V_j) + \sup_{i=1,\dots,N} \|z_i\| \sum_{j=1}^N |F(V_j) - w_j|\,.\end{aligned}$$

3.
$$\begin{aligned}\sup_{k=1,\dots,d} \left|\partial_{t_k}^2 \mathcal{F}[F](t) - \partial_{t_k}^2 \mathcal{F}[F^*](t)\right| \le \bigg(\int_{\mathbb{R}^d} \|x\|_\infty^4 dF(x) \bigg)^{1/2} \bigg\{ \sum_{j=1}^N |F(V_j) - w_j| \bigg\}^{1/2} \\ + \bigg(2 \sup_{v\in\bigcup_{j=1}^N V_j} \|v\| + \sup_{v\in\bigcup_{j=1}^N V_j} \|v\|^2\|t\| \bigg) \sup_{j=1,\dots,N} \mathcal{D}(V_j) \\ + \sup_{i=1,\dots,N} \|z_i\|^2 \sum_{j=1}^N |F(V_j) - w_j|\,.\end{aligned}$$

Here, $\mathcal{D}(A)$ denotes the diameter of a set A.

Proof of Lemma 3. Observe that

$$\begin{aligned}
&\mathcal{F}[F](t) - \mathcal{F}[F^*](t) \\
&= \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} dF(x) - \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} dF^*(x) \\
&= \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} dF(x) - \sum_{j=1}^{N} w_j e^{\mathbf{i}t'z_j} \\
&= \int_{V_0} e^{\mathbf{i}t'x} dF(x) + \sum_{j=1}^{N} \int_{V_j} (e^{\mathbf{i}t'x} - e^{\mathbf{i}t'z_j}) dF(x) + \sum_{j=1}^{N} e^{\mathbf{i}t'z_j} [F(V_j) - w_j].
\end{aligned}$$

Since the mapping $\mu \to e^{\mathbf{i}t'\mu}$ has Lipschitz constant at most $\|t\|$ and $\sum_{j=1}^N F(V_j) = 1$, we obtain

$$\begin{aligned}
|\mathcal{F}[F](t) - \mathcal{F}[F^*](t)| &\leq F(V_0) + \|t\| \sup_{j=1,\ldots,N} \mathcal{D}(V_j) \sum_{j=1}^{N} F(V_j) + \sum_{j=1}^{N} |F(V_j) - w_j| \\
&\leq F(V_0) + \|t\| \sup_{j=1,\ldots,N} \mathcal{D}(V_j) + \sum_{j=1}^{N} |F(V_j) - w_j| \,.
\end{aligned}$$

Since $V_0 = \mathbb{R}^d \setminus \cup_{j \geq 1} V_j$ and $F(\mathbb{R}^d) = 1 = \sum_{j=1}^N w_j$, we obtain

$$F(V_0) = \sum_{j=1}^{N} w_j - \sum_{j=1}^{N} F(V_j) \leq \sum_{j=1}^{N} |F(V_j) - w_j| \,.$$

For the second claim, first observe that the moment condition on F ensures that the gradient exists. By differentiating the preceding expression for $\mathcal{F}[F](t) - \mathcal{F}[F^*](t)$, we obtain

$$\begin{aligned}
&\nabla\mathcal{F}[F](t) - \nabla\mathcal{F}[F^*](t) \\
&= \mathbf{i} \int_{V_0} x e^{\mathbf{i}t'x} dF(x) + \mathbf{i} \sum_{j=1}^{N} \int_{V_j} (x e^{\mathbf{i}t'x} - z_j e^{\mathbf{i}t'z_j}) dF(x) + \mathbf{i} \sum_{j=1}^{N} z_j e^{\mathbf{i}t'z_j} [F(V_j) - w_j].
\end{aligned}$$

For $\mu \in V_j$, the mapping $\mu \to \mu e^{\mathbf{i}t'\mu}$ has Lipschitz constant at most $1 + \sup_{v \in V_j} \|v\| \|t\|$ and by Cauchy-Schwarz, $\int_{V_0} \|x\| dF(x) \leq (\int_{\mathbb{R}^d} \|x\|^2 dF(x))^{1/2} \{F(V_0)\}^{1/2}$. By using the same bound for $F(V_0)$ as above, we obtain

$$\begin{aligned}
&\|\nabla\mathcal{F}[F](t) - \nabla\mathcal{F}[F^*](t)\| \\
&\leq \left(\int_{\mathbb{R}^d} \|x\|^2 dF(x) \right)^{1/2} \left\{ \sum_{j=1}^{N} |F(V_j) - w_j| \right\}^{1/2} + \left(1 + \sup_{v \in \cup_{j=1}^N V_j} \|v\| \|t\| \right) \sup_{j=1,\ldots,N} \mathcal{D}(V_j) \\
&\quad + \sup_{i=1,\ldots,N} \|z_i\| \sum_{j=1}^{N} |F(V_j) - w_j| \,.
\end{aligned}$$

For the third claim, let $z_{j,k}$ denote the k^{th} element of z_j. Twice differentiating (with respect to t_k) the expression for $\mathcal{F}[F](t) - \mathcal{F}[F^*](t)$ yields

$$\begin{aligned}
&\partial^2_{t_k}\mathcal{F}[F](t) - \partial^2_{t_k}\mathcal{F}[F^*](t) \\
&= -\int_{V_0} x_k^2 e^{\mathbf{i}t'x} dF(x) - \sum_{j=1}^N \int_{V_j} (x_k^2 e^{\mathbf{i}t'x} - z_{j,k}^2 e^{\mathbf{i}t'z_j}) dF(x) - \sum_{j=1}^N z_{j,k}^2 e^{\mathbf{i}t'z_j}[F(V_j) - w_j].
\end{aligned}$$

For $\mu = (\mu_1, \dots, \mu_d) \in V_j$ and a fixed component μ_k the mapping $\mu \to \mu_k^2 e^{\mathbf{i}t'\mu}$ has Lipschitz constant at most $2\sup_{v\in V_j}\|v\| + \sup_{v\in V_j}\|v\|^2\|t\|$ and by Cauchy-Schwarz, we have the bound $\int_{V_0} x_k^2 dF(x) \leq (\int_{\mathbb{R}^d}\|x\|_\infty^4 dF(x))^{1/2}\{F(V_0)\}^{1/2}$. The claim follows from an analogous bound to the preceding case.

□

Lemma 4. *Consider a measurable partition $\mathbb{R}^d = \bigcup_{j=0}^N V_j$ and points $z_j \in V_j$ for $j = 1, \dots, N$. Let $F^* = \sum_{j=1}^N w_j \delta_{z_j}$ denote the discrete probability measure with weight w_j at z_j. Then, there exists a universal constant $D > 0$ such that for any positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$ with minimum eigenvalue $\underline{\sigma}^2 > 0$ and probability measure F on $\mathbb{R}^d$, we have that*

$$\|\varphi_{F,\Sigma} - \varphi_{F^*,\Sigma}\|_{L^2} \leq D\left[\underline{\sigma}^{-(d+2)/2} \sup_{j=1,\dots,N} \mathcal{D}(V_j) + \underline{\sigma}^{-d/2}\sum_{j=1}^N |F(V_j) - w_j|\right],$$

$$\|\varphi_{F,\Sigma} - \varphi_{F^*,\Sigma}\|_{\mathbb{B}(T)} \leq D\left[T^{(d+2)/2} \sup_{j=1,\dots,N} \mathcal{D}(V_j) + T^{d/2}\sum_{j=1}^N |F(V_j) - w_j|\right].$$

Proof of Lemma 4. By Lemma 3, we have

$$|\mathcal{F}[F](t) - \mathcal{F}[F^*](t)| \leq \|t\| \sup_{j=1,\dots,N} \mathcal{D}(V_j) + 2\sum_{j=1}^N |F(V_j) - w_j|.$$

Since $e^{-t'\Sigma t} \leq e^{-\|t\|^2\underline{\sigma}^2}$, it follows that

$$\begin{aligned}
&\|\varphi_{F,\Sigma} - \varphi_{F^*,\sigma}\|_{L^2}^2 \\
&= \int_{\mathbb{R}^d} |\varphi_\Sigma(t)|^2 |\mathcal{F}[F](t) - \mathcal{F}[F^*](t)|^2 dt \\
&= \int_{\mathbb{R}^d} e^{-t'\Sigma t} |\mathcal{F}[F](t) - \mathcal{F}[F^*](t)|^2 dt \\
&\leq \int_{\mathbb{R}^d} e^{-\|t\|^2\underline{\sigma}^2} |\mathcal{F}[F](t) - \mathcal{F}[F^*](t)|^2 dt \\
&\leq 2 \sup_{j=1,\dots,N} \{\mathcal{D}(V_j)\}^2 \int_{\mathbb{R}^d} e^{-\|t\|^2\underline{\sigma}^2}\|t\|^2 dt + 8\left\{\sum_{j=1}^N |F(V_j) - w_j|\right\}^2 \int_{\mathbb{R}^d} e^{-\|t\|^2\underline{\sigma}^2} dt.
\end{aligned}$$

The first claim follows from observing that the two integrals scale with rate at most $\underline{\sigma}^{-(d+2)}$ and $\underline{\sigma}^{-d}$, respectively. The second claim follows by an analogous argument from truncating the integral to the set $\{\|t\|_\infty \leq T\}$ and using the trivial bound $e^{-t'\Sigma t} \leq 1$.

□

Lemma 5. *Suppose F is a probability measure supported on $[-R,R]^d$ for some $R>0$. Then, given any $k\geq 1$, there exists a discrete probability measure F' with at most $(k+1)^d+1$ support points in $[-R,R]^d$ such that for every $t\in\mathbb{R}^d$, we have*

$$
\begin{aligned}
&(i) \quad \left|\mathcal{F}[F](t)-\mathcal{F}[F'](t)\right| \leq 2\frac{\|t\|^{k+1}(e\sqrt{d}R)^{k+1}}{(k+1)^{k+1}}\ ,\\
&(ii) \quad \|\nabla\mathcal{F}[F](t)-\nabla\mathcal{F}[F'](t)\| \leq 2\sqrt{d}R\frac{\|t\|^{k}(e\sqrt{d}R)^{k}}{k^{k}}\ ,\\
&(iii) \quad \sup_{l=1,\ldots,d}\left|\partial^2_{t_l}\mathcal{F}[F](t)-\partial^2_{t_l}\mathcal{F}[F'](t)\right| \leq 2dR^2\frac{\|t\|^{k-1}(e\sqrt{d}R)^{k-1}}{(k-1)^{k-1}}.
\end{aligned}
$$

In particular, there exists universal constants $C,D>0$ such that for all T,R sufficiently large and $\epsilon\in(0,1)$, the choice $k=\lceil C\max\{\log(\epsilon^{-1}),RT\}\rceil$ satisfies

$$
\begin{aligned}
&(i) \quad \sup_{\|t\|_\infty\leq T}\left|\mathcal{F}[F](t)-\mathcal{F}[F'](t)\right| \leq D\epsilon\ ,\\
&(ii) \quad \sup_{\|t\|_\infty\leq T}\|\nabla\mathcal{F}[F](t)-\nabla\mathcal{F}[F'](t)\| \leq D\epsilon\ ,\\
&(iii) \quad \sup_{\|t\|_\infty\leq T}\sup_{l=1,\ldots,d}\left|\partial^2_{t_l}\mathcal{F}[F](t)-\partial^2_{t_l}\mathcal{F}[F'](t)\right| \leq D\epsilon.
\end{aligned}
$$

Furthermore, the support points of F' can be chosen on the grid $\mathcal{Z}=\{T^{-1}\epsilon(z_1,\ldots,z_d): z_i\in\mathbb{Z}\ ,\ |z_i|\leq\lceil R/(T^{-1}\epsilon)\rceil\}$, with a multiplictive penalty of at most R and R^2 in cases (ii) and (iii), respectively.

Proof of Lemma 5. Given any $k\geq 1$, by (Ghosal and Van Der Vaart, 2001, Lemma A.1), there exists a discrete measure F' with at most $(k+1)^d+1$ support points on $[-R,R]^d$ such that

$$
\int_{\mathbb{R}^d} z_1^{l_1}\ldots z_d^{l_d}\,dF(z)=\int_{\mathbb{R}^d} z_1^{l_1}\ldots z_d^{l_d}\,dF'(z) \qquad \forall\ \ 0\leq l_1,\ldots,l_d\leq k.
$$

For any $t,z\in\mathbb{R}^d$ and $j\in\mathbb{Z}$, a multinomial expansion yields

$$
(t'z)^j=\sum_{k_1+\cdots+k_d=j,k_1\geq 0,\ldots,k_d\geq 0}\frac{j!}{k_1!\ldots k_d!}\prod_{i=1}^d t_i^{k_i}z_i^{k_i}.
$$

It follows that F and F' assign the same expectation to $(t'z)^j$ for every $t\in\mathbb{R}^d$, provided that $j\leq k$. This yields

$$
\begin{aligned}
\mathcal{F}[F](t)-\mathcal{F}[F'](t)=\int_{\mathbb{R}^d} e^{\mathbf{i}t'z}d(F-F')(z)&=\int_{\mathbb{R}^d}\sum_{j=0}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z)\\
&=\int_{\mathbb{R}^d}\sum_{j=k+1}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z).
\end{aligned}
$$

Observe that for every $x \in \mathbb{R}$, we have the bound $\left|\sum_{j=k+1}^{\infty}(\mathbf{i}x)^j/j!\right| = \left|e^{\mathbf{i}x} - \sum_{j=0}^{k}(\mathbf{i}x)^j/j!\right| \leq |x|^{k+1}/(k+1)! \leq |x|^{k+1} e^{k+1}/(k+1)^{k+1}$. For any $z \in [-R,R]^d$, we have that $|t'z| \leq \|t\|\|z\| \leq \|t\|\sqrt{d}R$. Since F, F' are probability measures and have support contained in $[-R,R]^d$, it follows that

$$\begin{aligned}
\left|\mathcal{F}[F](t) - \mathcal{F}[F'](t)\right| &\leq \int_{\mathbb{R}^d}\left|\sum_{j=k+1}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}\right|dF + \int_{\mathbb{R}^d}\left|\sum_{j=k+1}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}\right|dF' \\
&\leq 2\frac{\|t\|^{k+1}(e\sqrt{d}R)^{k+1}}{(k+1)^{k+1}}.
\end{aligned}$$

For the second claim, the same reasoning as above implies that F and F' assign the same expectation to the vector $z(t'z)^j$ for every $t \in \mathbb{R}^d$, provided that $j \leq k-1$. This yields

$$\begin{aligned}
\nabla\mathcal{F}[F](t) - \nabla\mathcal{F}[F'](t) = \mathbf{i}\int_{\mathbb{R}^d} z e^{\mathbf{i}t'z} d(F-F')(z) &= \mathbf{i}\int_{\mathbb{R}^d} z\sum_{j=0}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z) \\
&= \mathbf{i}\int_{\mathbb{R}^d} z\sum_{j=k}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z).
\end{aligned}$$

We have $\|z\| \leq \sqrt{d}R$ for every z in the support of F, F'. From using the bound in the preceding case with k replacing $k+1$, it follows that

$$\begin{aligned}
\|\nabla\mathcal{F}[F](t) - \nabla\mathcal{F}[F'](t)\| &\leq \sqrt{d}R\Bigg(\int_{\mathbb{R}^d}\left|\sum_{j=k}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}\right|dF + \int_{\mathbb{R}^d}\left|\sum_{j=k}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}\right|dF'\Bigg) \\
&\leq 2\sqrt{d}R\frac{\|t\|^k(e\sqrt{d}R)^k}{k^k}.
\end{aligned}$$

For the third claim, let z_l denote the l^{th} coordinate of $z = (z_1, \dots, z_d)$. The same reasoning as above implies that F and F' assign the same expectation to the vector $z_l^2(t'z)^j$ for every $t \in \mathbb{R}^d$, provided that $j \leq k-2$. This yields

$$\begin{aligned}
\partial_{t_l}^2\mathcal{F}[F](t) - \partial_{t_l}^2\mathcal{F}[F'](t) = -\int_{\mathbb{R}^d} z_l^2 e^{\mathbf{i}t'z}d(F-F')(z) &= -\int_{\mathbb{R}^d} z_l^2\sum_{j=0}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z) \\
&= -\int_{\mathbb{R}^d} z_l^2\sum_{j=k-1}^{\infty}\frac{(\mathbf{i}t'z)^j}{j!}d(F-F')(z).
\end{aligned}$$

Since $z_l^2 \leq \|z\|^2 \leq dR^2$, the claim follows from an analogous bound to the preceding case.

For the final claim regarding the location of the support points, suppose $F' = \sum_{i=1}^{N} p_i\delta_{\mu_i}$ is a discrete probability measure that satisfies the requirements of the first part of the Lemma. Let $F^* = \sum_{i=1}^{N} p_i\delta_{\mu_i^*}$ denote the probability measure obtained by replacing each μ_i with $\mu_i^* \in \arg\min_{t\in\mathcal{Z}}\|\mu_i - t\|$. From the definition of $\mathcal{Z}$, it follows that $\|\mu_i - \mu_i^*\| \leq DT^{-1}\epsilon$. We claim F'

satisfies all the same bounds. For the first bound, observe that

$$\sup_{\|t\|_\infty \le T} \left|\mathcal{F}[F'] - \mathcal{F}[F^*]\right| = \sup_{\|t\|_\infty \le T} \left|\sum_{j=1}^N p_j [e^{\mathbf{i}t\mu_j} - e^{\mathbf{i}t\mu_j^*}]\right| \le \sup_{\|t\|_\infty \le T} \sup_{j=1,\dots,N} \left|e^{\mathbf{i}t\mu_j} - e^{\mathbf{i}t\mu_j^*}\right|.$$

Since the mapping $\mu \to e^{\mathbf{i}t\mu}$ has Lipschitz constant bounded by $\|t\|$, it follows that

$$\sup_{\|t\|_\infty \le T} \left|\mathcal{F}[F'] - \mathcal{F}[F^*]\right| \le \sqrt{d}T \sup_{j=1,\dots,N} \|\mu_j - \mu_j^*\| \le D\epsilon.$$

For the second bound, observe that for $\mu \in [-R, R]^d$, the mapping $\mu \to \mu e^{\mathbf{i}t\mu}$ has Lipschitz constant at most $1 + \sqrt{d}R\|t\|$. Hence

$$\begin{aligned}
\sup_{\|t\|_\infty \le T} \|\nabla\mathcal{F}[F'](t) - \nabla\mathcal{F}[F^*](t)\| &\le \sup_{\|t\|_\infty \le T} \sup_{j=1,\dots,N} \|\mu_j e^{\mathbf{i}t'\mu_j} - \mu_j^* e^{\mathbf{i}t'\mu_j^*}\| \\
&\le D(RT+1) \sup_{j=1,\dots,N} \|\mu_j - \mu_j^*\| \\
&\le DR\epsilon.
\end{aligned}$$

For the third bound, let $\mu_{j,l}$ denote the l^{th} component of μ_j. On $[-R, R]^d$, the mapping $\mu_j \to \mu_{j,l}^2 e^{\mathbf{i}t'\mu_j}$ has Lipschitz constant at most $2\sqrt{d}R + dR^2\|t\|$. Hence

$$\begin{aligned}
\sup_{\|t\|_\infty \le T} \sup_{l=1,\dots,d} \left|\partial_{t_l}^2 \mathcal{F}[F](t) - \partial_{t_l}^2 \mathcal{F}[F'](t)\right| &\le \sup_{\|t\|_\infty \le T} \sup_{l=1,\dots,d} \sup_{j=1,\dots,N} \left|\mu_{j,l}^2 e^{\mathbf{i}t'\mu_j} - (\mu_{j,l}^*)^2 e^{\mathbf{i}t'\mu_j^*}\right| \\
&\le D(R + R^2T) \sup_{j=1,\dots,N} \|\mu_j - \mu_j^*\| \\
&\le DR^2\epsilon.
\end{aligned}$$

□

Lemma 6. *Suppose F is a probability measure supported on $[-L, L]^d$ for some $L > 0$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a positive-definite matrix with smallest eigenvalue $\underline{\sigma}^2 > 0$. Then, for all $\epsilon \in (0, 1)$, there exists a discrete probability measure F' with at most $D \max\{(\log(\epsilon^{-1}))^d, (L/\underline{\sigma})^d(\log(\epsilon^{-1}))^{d/2}\}$ support points on $[-L, L]^d$ such that $\|\varphi_{F,\Sigma} - \varphi_{F',\Sigma}\|_{L^2} \le D'\underline{\sigma}^{-d/2}\epsilon$, where $D, D' > 0$ are universal constants. Furthermore, the support points can be chosen such that $\inf_{i\neq j}\|\mu_i - \mu_j\| \ge \underline{\sigma}\epsilon$.*

Proof of Lemma 6. Let D denote a generic universal constant that may change from line to line. By Lemma 5, there exists a discrete measure F' with at most $k^d + 1$ support points on $[-L, L]^d$ such that

$$\left|\mathcal{F}[F](t) - \mathcal{F}[F'](t)\right| \le 2\frac{\|t\|^k (e\sqrt{d}L)^k}{k^k} \quad \forall\, t \in \mathbb{R}^d.$$

Observe that $|\varphi_\Sigma| = e^{-t'\Sigma t/2} \le e^{-\|t\|^2\underline{\sigma}^2/2}$. From using the preceding bound, we obtain for every

$M > 0$, the estimate

$$\begin{aligned}\|\varphi_{F,\Sigma} - \varphi_{F',\Sigma}\|_{L^2}^2 &= \int_{\mathbb{R}^d} |\varphi_\Sigma(t)|^2 \left|\mathcal{F}[F](t) - \mathcal{F}[F'](t)\right|^2 dt \\ &\leq D\Bigg[\int_{\|t\|>M} |\varphi_\Sigma(t)|^2 \, dt + \left(\frac{Le\sqrt{d}}{k}\right)^{2k} \int_{\|t\|\leq M} \|t\|^{2k} dt\Bigg] \\ &\leq D\Bigg[\int_{\|t\|>M} e^{-\|t\|^2\underline{\sigma}^2} dt + \left(\frac{Le\sqrt{d}}{k}\right)^{2k} \int_{\|t\|\leq M} \|t\|^{2k} dt\Bigg].\end{aligned}$$

By change of variables (to spherical coordinates) the first integral scales at rate $DM^{d-2}e^{-M^2\underline{\sigma}^2}\underline{\sigma}^{-2}$ and the second integral scales at rate $DM^{2k+d}/(2k+d)$. The choice $M = \sqrt{4\log(\epsilon^{-1})}/\underline{\sigma}$ leads to

$$\int_{\|t\|>M} e^{-\|t\|^2\underline{\sigma}^2} dt \leq DM^{d-2}e^{-M^2\underline{\sigma}^2}\underline{\sigma}^{-2} \leq D\epsilon^2\underline{\sigma}^{-d}.$$

Furthermore, if $k \geq e^3\sqrt{d}\max\{LM, \log\big(\epsilon^{-1}\big)\}$, we have that

$$\left(\frac{Le\sqrt{d}}{k}\right)^{2k} \int_{\|t\|\leq M} \|t\|^{2k} dt \leq D\bigg(\frac{Le\sqrt{d}M}{k}\bigg)^{2k} \frac{M^d}{2k+d} \leq D\epsilon^2.$$

The claim follows from observing that the number of support points in F' is $N = k^d + 1$.

For the final claim regarding the separation of the support points, suppose $F' = \sum_{i=1}^N p_i\delta_{\mu_i}$ is a discrete probability measure that satisfies the requirements of the first part of the Lemma. Let $\mathcal{Z}$ denote a maximal $\underline{\sigma}\epsilon$ separated subset of $[-L, L]^d$. For each μ_i, select $\mu_i^* \in \arg\min_{t\in\mathcal{Z}}\|\mu_i - t\|$ and let $F^* = \sum_{i=1}^N p_i\delta_{\mu_i^*}$. From the definition of $\mathcal{Z}$, it follows that $\sup_{i=1}^N\|\mu_i - \mu_i^*\| \leq \underline{\sigma}\epsilon$. Observe that

$$\begin{aligned}\|\varphi_{F',\Sigma} - \varphi_{F^*,\Sigma}\|_{L^2} &= \Bigg\|\sum_{j=1}^N p_j[e^{it\mu_j} - e^{it\mu_j^*}]e^{-t'\Sigma t/2}\Bigg\|_{L^2} \\ &\leq D\sum_{j=1}^N p_j\|(e^{it\mu_j} - e^{it\mu_j^*})e^{-t'\Sigma t/2}\|_{L^2} \\ &\leq D \sup_{j=1,\ldots,N}\|(e^{it\mu_j} - e^{it\mu_j^*})e^{-t'\Sigma t/2}\|_{L^2}.\end{aligned}$$

Since the mapping $\mu \to e^{it\mu}$ has Lipschitz constant bounded by $\|t\|$, we obtain

$$\begin{aligned}\|(e^{it\mu_j} - e^{it\mu_j^*})e^{-t'\Sigma t/2}\|_{L^2}^2 &= \int_{\mathbb{R}^d} \left|e^{it\mu_j} - e^{it\mu_j^*}\right|^2 e^{-t'\Sigma t} dt \\ &\leq D\|\mu_j - \mu_j^*\|^2 \int_{\mathbb{R}^d} \|t\|^2 e^{-t'\Sigma t} dt \\ &\leq D\underline{\sigma}^2\epsilon^2 \int_{\mathbb{R}^d} \|t\|^2 e^{-t'\Sigma t} dt.\end{aligned}$$

Since $e^{-t'\Sigma t} \leq e^{-\|t\|^2\underline{\sigma}^2}$, the integral on the right scales with rate at most $D\underline{\sigma}^{-(d+2)}$. It follows that

$$\|(e^{it\mu_j} - e^{it\mu_j^*})e^{-t'\Sigma t/2}\|_{L^2} \leq D\underline{\sigma}\epsilon\underline{\sigma}^{-(d+2)/2} = D\underline{\sigma}^{-d/2}\epsilon.$$

In particular, we obtain

$$\|\varphi_{F,\Sigma} - \varphi_{F^*,\Sigma}\|_{L^2} \leq \|\varphi_{F,\Sigma} - \varphi_{F',\Sigma}\|_{L^2} + \|\varphi_{F^*,\Sigma} - \varphi_{F',\Sigma}\|_{L^2} \leq D\underline{\sigma}^{-d/2}\epsilon.$$

Finally, note that if $\mu_i^* = \mu_j^*$ for some $i \neq j$, F^* can be reduced to a discrete measure with $N^* \leq N$ unique support points.

□

Lemma 7. *Fix any positive definite matrix $\Sigma_0 \in \mathbb{R}^{d\times d}$ and denote by σ_0^2, the smallest eigenvalue of Σ_0. Then, there exists a universal constant $D > 0$ (only depending on d) such that for any distribution P and positive definite matrix Σ satisfying $\|\Sigma - \Sigma_0\| \leq \sigma_0^2/2$, we have*

$$\|\varphi_{P,\Sigma} - \varphi_{P,\Sigma_0}\|_{L^2} \leq D\sigma_0^{-d/2-1}\|\Sigma - \Sigma_0\|.$$

Proof of Lemma 7. For any distribution P and positive definite matrix Σ, we have that

$$\begin{aligned}\|\varphi_{P,\Sigma} - \varphi_{P,\Sigma_0}\|_{L^2}^2 = \int_{\mathbb{R}^d} |\mathcal{F}[P](t)|^2 \left|e^{-t'\Sigma t/2} - e^{-t'\Sigma_0 t}\right|^2 dt &\leq \int_{\mathbb{R}^d} \left|e^{-t'\Sigma t/2} - e^{-t'\Sigma_0 t/2}\right|^2 dt \\ &= \int_{\mathbb{R}^d} e^{-t'\Sigma_0 t} \left|1 - e^{t'(\Sigma_0-\Sigma)t/2}\right|^2 dt.\end{aligned}$$

The mapping $t \to e^{t'(\Sigma_0-\Sigma)t/2}$ has gradient norm at most $e^{\|t\|^2\|\Sigma-\Sigma_0\|/2}\|\Sigma - \Sigma_0\|\|t\|$. If the bound $\|\Sigma - \Sigma_0\| \leq \sigma_0^2/2$ holds, it follows that

$$\begin{aligned}\int_{\mathbb{R}^d} e^{-t'\Sigma_0 t} \left|1 - e^{t'(\Sigma_0-\Sigma)t/2}\right|^2 dt &\leq \int_{\mathbb{R}^d} e^{-\|t\|^2\sigma_0^2} \left|1 - e^{t'(\Sigma_0-\Sigma)t/2}\right|^2 dt \\ &\leq \|\Sigma - \Sigma_0\|^2 \int_{\mathbb{R}^d} e^{-\|t\|^2\sigma_0^2/2}\|t\|^2 dt.\end{aligned}$$

The claim follows from observing that the integral on the right scales with rate at most $D\sigma_0^{-(d+2)}$.

□

Lemma 8. *Suppose $(W, Z) \in \mathbb{R}^{d^*} \times \mathbb{R}^d$ for some $d^*, d \in \mathbb{N}$ and $\mathbb{E}(\|Z\|^2) < \infty, \mathbb{E}(\|W\|^2) < \infty$. For every $t \in \mathbb{R}^d$, define*

$$\chi_t(W, Z) = We^{\mathbf{i}t'Z} = W\cos\big(t'Z\big) + \mathbf{i}W\sin\big(t'Z\big).$$

Then, there exists a universal constant $D > 0$ such that for every $T > 0$ we have

$$\mathbb{E}\bigg(\sup_{\|t\|_\infty \leq T} \|\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]\|\bigg) \leq D\frac{\max\{\sqrt{\log T}, 1\}}{\sqrt{n}}.$$

Proof of Lemma 8. It suffices to show the result for the real and imaginary part separately. We verify it for the real part, the imaginary part is completely analogous. If $W = (W_1, W_2, \dots, W_{d_1})$, it suffices to verify the result for each vector sub component $W_k e^{\mathbf{i}t'Z}$ where $k \in \{1, \dots, d^*\}$. Fix any $k \in \{1, \dots, d^*\}$. For the remainder of this proof, we continue assuming $\chi_t(W, Z) = W_k \cos(t'Z)$. Define the class of functions $\mathcal{F} = \{\chi_t(W, Z) : \|t\|_\infty \leq T\}$. Since $\|W\|$ is an envelope of $\mathcal{F}$, an application of (Giné and Nickl, 2021, Remark 3.5.14) implies that there exists a universal constant $L > 0$ such that

$$\mathbb{E}\bigg(\sup_{\|t\|_\infty \leq T} |\mathbb{E}_n[\chi_t(W, Z)] - \mathbb{E}[\chi_t(W, Z)]|\bigg) \leq \frac{L}{\sqrt{n}} \int_0^{8\|W\|_{L^2(\mathbb{P})}} \sqrt{\log N_{[]}(\mathcal{F}, \|.\|_{L^2(\mathbb{P})}, \epsilon)} d\epsilon.$$

Let $\{t_i\}_{i=1}^M$ denote a minimal $\delta > 0$ covering of $[-T, T]^d$. Define the functions

$$e_i(W, Z) = \sup_{t \in \mathbb{R}^d : \|t\|_\infty \leq T, \|t - t_i\|_\infty < \delta} |\, \chi_t(W, Z) - \chi_{t_i}(W, Z)\,| \qquad i = 1, \dots, M.$$

It follows that $\{\chi_{t_i}(W, Z) - e_i\ ,\ \chi_{t_i}(W, Z) + e_i\}_{i=1}^M$ is a bracket covering for $\mathcal{F}$. Since the mapping $t \to \chi_t(W, Z)$ has Lipschitz constant bounded by $\|W\|\|Z\|$, Cauchy-Schwarz implies that $\|e_i\|^2_{L^2(\mathbb{P})} \leq \|W\|^2_{L^2(\mathbb{P})} \|Z\|^2_{L^2(\mathbb{P})} \delta^2$. Since $M \leq (3T\delta^{-1})^d$, it follows that there exists a universal constant $L > 0$ such that

$$\int_0^{8\|W\|_{L^2(\mathbb{P})}} \sqrt{\log N_{[]}(\mathcal{F}, \|.\|_{L^2(\mathbb{P})}, \epsilon)} d\epsilon \leq L \max\{\sqrt{\log T}, 1\}.$$

□

Lemma 9. *Suppose* $(W, Z) \in \mathbb{R}^{d^*} \times \mathbb{R}^d$ *for some* $d^*, d \in \mathbb{N}$ *and* $\mathbb{E}(\|Z\|^2) < \infty, \mathbb{E}(\|W\|^2) < \infty$. *For every* $t \in \mathbb{R}^d$, *define*

$$\chi_t(W, Z) = W e^{\mathbf{i}t'Z} = W \cos\big(t'Z\big) + \mathbf{i} W \sin\big(t'Z\big).$$

Then, there exists a universal constant $L > 0$ *such that for any sequence* $T_n \uparrow \infty$ *with* $\log(T_n) = o(n)$, *we have that*

$$\mathbb{P}\bigg(\sup_{\|t\|_\infty \leq T_n} |\mathbb{E}_n[\chi_t(W, Z)] - \mathbb{E}[\chi_t(W, Z)]| \leq D \frac{\sqrt{\log T_n}}{\sqrt{n}}\bigg) \to 1.$$

Proof of Lemma 9. It suffices to show the result for the real and imaginary part separately. We verify it for the real part, the imaginary part is completely analogous. If $W = (W_1, W_2, \dots, W_{d^*})$, it suffices to verify the result for each vector sub component $W_k e^{\mathbf{i}t'Z}$ where $k \in \{1, \dots, d^*\}$. Fix any $k \in \{1, \dots, d^*\}$. For the remainder of this proof, we continue assuming $\chi_t(W, Z) = W_k \cos(t'Z)$. For a given sequence of deterministic constants $L_n \uparrow \infty$, define

$$\chi_{1,t}(W, Z) = \chi_t(W, Z) \mathbb{1}\{\|W\| \leq L_n\}\ ,$$
$$\chi_{2,t}(W, Z) = \chi_t(W, Z) \mathbb{1}\{\|W\| > L_n\}\ .$$

Observe that

$$(\mathbb{E}_n - \mathbb{E})[\chi_t(W,Z)] = \sum_{i=1}^{n} \Xi_{1,t}(W_i, Z_i) + \sum_{i=1}^{n} \Xi_{2,t}(W_i, Z_i) \ ,$$

where $\Xi_{1,t}(W,Z) = n^{-1}[\chi_{1,t}(W,Z) - \mathbb{E}\chi_{1,t}(W,Z)]$ and $\Xi_{2,t}(W,Z) = n^{-1}[\chi_{2,t}(W,Z) - \mathbb{E}\chi_{2,t}(W,Z)]$.

First, we derive a bound for $\sum_{i=1}^{n} \Xi_{2,t}(W_i, Z_i)$. Observe that

$$\begin{aligned}
\mathbb{P}\bigg(\sup_{\|t\|_\infty \le T_n} \bigg| \sum_{i=1}^{n} \Xi_{2,t}(W_i, Z_i) \bigg| > \frac{\sqrt{\log(T_n)}}{\sqrt{n}} \bigg) &\le \frac{\sqrt{n}}{\sqrt{\log(T_n)}} \mathbb{E}\bigg(\sup_{\|t\|_\infty \le T_n} \sum_{i=1}^{n} |\Xi_{2,t}(W_i, Z_i)| \bigg) \\
&\le \frac{2\sqrt{n}}{\sqrt{\log(T_n)}} \mathbb{E}(\|W\| \mathbb{1}\{\|W\| > L_n\}) \\
&\le \frac{2\sqrt{n}}{\sqrt{\log T_n} L_n} \mathbb{E}(\|W\|^2 \mathbb{1}\{\|W\| > L_n\}).
\end{aligned}$$

Since $\mathbb{E}(\|W\|^2) < \infty$, the term on the right is $o(1)$ when $L_n = \sqrt{n}/\sqrt{\log T_n}$.

It remains to bound the first sum $\sum_{i=1}^{n} \Xi_{1,t}(W_i, Z_i)$ when $L_n = \sqrt{n}/\sqrt{\log T_n}$. Observe that

$$\begin{aligned}
\sup_{\|t\|_\infty \le T_n} \mathbb{E}[\, |\Xi_{1,t}(W,Z)|^2 \,] &\le n^{-2} \|W\|_{L^2(\mathbb{P})}^2 \\
\sup_{\|t\|_\infty \le T_n} |\Xi_{1,t}(W,Z)| &\le 2n^{-1} L_n.
\end{aligned}$$

The preceding bounds, Lemma 8 and (Giné and Nickl, 2021, Theorem 3.3.9) imply that there exists universal constants $C, D > 0$ which satisfy

$$\begin{aligned}
&\mathbb{P}\bigg(\sup_{\|t\|_\infty \le T_n} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]| > D\frac{\sqrt{\log T_n}}{\sqrt{n}} + x \bigg) \\
&\le \exp\bigg(- \frac{x^2}{Cn^{-1}[L_n \sqrt{\log T_n}/\sqrt{n} + \|W\|_{L^2(\mathbb{P})}^2 + xL_n]} \bigg)
\end{aligned}$$

for all $x > 0$. The choice $x = \sqrt{\log T_n}/\sqrt{n}$ with $L_n = \sqrt{n}/\sqrt{\log T_n}$ yields

$$\mathbb{P}\bigg(\sup_{\|t\|_\infty \le T_n} |\mathbb{E}_n[\chi_t(Z)] - \mathbb{E}[\chi_t(Z)]| > D\frac{\sqrt{\log T_n}}{\sqrt{n}} + \frac{\sqrt{\log T_n}}{\sqrt{n}} \bigg) \le \exp\big(- E \log(T_n) \big)$$

for some universal constant $E > 0$. Since $T_n \uparrow \infty$, the claim follows.

□

Lemma 10. *Suppose F is a probability measure supported on $[-L, L]^d$ for some $L > 0$ and $\Sigma \in \mathbb{R}^{d \times d}$ is a positive-definite matrix. Then, for all $\epsilon \in (0,1)$, there exists a discrete probability measure F' with at most $D \max\{(\log(\epsilon^{-1}))^d, L^d T^d\}$ support points on $[-L, L]^d$ such that $\|\varphi_{F,\Sigma} - \varphi_{F',\Sigma}\|_{\mathbb{B}(T)} \le D'\epsilon$, where $D, D' > 0$ are universal constants. Furthermore, the support points can be chosen such that $\inf_{i \ne j} \|\mu_i - \mu_j\| \ge T^{-(d+2)/2}\epsilon$.*

Proof of Lemma 10. Let D denote a generic universal constant that may change from line to line. By Lemma 5, there exists a discrete measure F' with at most k^d+1 support points on $[-L,L]^d$ such that

$$\big|\mathcal{F}[F](t)-\mathcal{F}[F'](t)\big|\le 2\frac{\|t\|^k(e\sqrt{d}L)^k}{k^k}\quad\forall\, t\in\mathbb{R}^d.$$

From using the preceding bound and noting that the eigenvalues of Σ are non-negative, we obtain for every $M>0$, the estimate

$$\begin{aligned}\|\varphi_{F,\Sigma}-\varphi_{F',\Sigma}\|^2_{\mathbb{B}(T)} &\le \int_{\|t\|_\infty\le T}|\varphi_\Sigma(t)|^2\left|\mathcal{F}[F](t)-\mathcal{F}[F'](t)\right|^2dt\\ &\le 2\left(\frac{Le\sqrt{d}}{k}\right)^{2k}\int_{\|t\|_\infty\le T}e^{-t'\Sigma t}\|t\|^{2k}dt\\ &\le 2\left(\frac{Le\sqrt{d}}{k}\right)^{2k}\int_{\|t\|_\infty\le T}\|t\|^{2k}dt\\ &\le D\left(\frac{Le\sqrt{d}T}{k}\right)^{2k}T^d.\end{aligned}$$

The quantity on the right is bounded above by $D\epsilon^2$ if $k\ge e^3\sqrt{d}\max\{LT,\log\big(\epsilon^{-1}\big)\}$. The claim follows from observing that the number of support points in F' is $N=k^d+1$.

For the final claim regarding the separation of the support points, suppose $F'=\sum_{i=1}^N p_i\delta_{\mu_i}$ is a discrete probability measure that satisfies the requirements of the first part of the Lemma. Let $\mathcal{Z}$ denote a maximal $T^{-(d+2)/2}\epsilon$ separated subset of $[-L,L]^d$. For each μ_i, select $\mu_i^*\in\arg\min_{t\in\mathcal{Z}}\|\mu_i-t\|$ and let $F^*=\sum_{i=1}^N p_i\delta_{\mu_i^*}$. From the definition of $\mathcal{Z}$, it follows that $\sup_{i=1}^N\|\mu_i-\mu_i^*\|\le\underline{\sigma}\epsilon$. Observe that

$$\begin{aligned}\|\varphi_{F',\Sigma}-\varphi_{F^*,\Sigma}\|_{\mathbb{B}(T)} &= \left\|\sum_{j=1}^N p_j[e^{\mathbf{i}t\mu_j}-e^{\mathbf{i}t\mu_j^*}]e^{-t'\Sigma t/2}\right\|_{\mathbb{B}(T)}\\ &\le D\sum_{j=1}^N p_j\|(e^{\mathbf{i}t\mu_j}-e^{\mathbf{i}t\mu_j^*})e^{-t'\Sigma t/2}\|_{\mathbb{B}(T)}\\ &\le D\sup_{j=1,\dots,N}\|(e^{\mathbf{i}t\mu_j}-e^{\mathbf{i}t\mu_j^*})e^{-t'\Sigma t/2}\|_{\mathbb{B}(T)}.\end{aligned}$$

Since the mapping $\mu\to e^{\mathbf{i}t\mu}$ has Lipschitz constant bounded by $\|t\|$, we obtain

$$\begin{aligned}\|(e^{\mathbf{i}t\mu_j}-e^{\mathbf{i}t\mu_j^*})e^{-t'\Sigma t/2}\|^2_{\mathbb{B}(T)} &\le \int_{\|t\|_\infty\le T}\left|e^{\mathbf{i}t\mu_j}-e^{\mathbf{i}t\mu_j^*}\right|^2e^{-t'\Sigma t}dt\\ &\le D\|\mu_j-\mu_j^*\|^2\int_{\|t\|_\infty\le T}\|t\|^2e^{-t'\Sigma t}dt\\ &\le DT^{-(d+2)}\epsilon^2\int_{\|t\|_\infty\le T}\|t\|^2dt.\end{aligned}$$

The integral on the right scales with rate at most DT^{d+2}. It follows that

$$\|\varphi_{F,\Sigma} - \varphi_{F^*,\Sigma}\|_{\mathbb{B}(T)} \leq \|\varphi_{F,\Sigma} - \varphi_{F',\Sigma}\|_{\mathbb{B}(T)} + \|\varphi_{F^*,\Sigma} - \varphi_{F',\Sigma}\|_{\mathbb{B}(T)} \leq D\epsilon.$$

Finally, note that if $\mu_i^* = \mu_j^*$ for some $i \neq j$, F^* can be reduced to a discrete measure with $N^* \leq N$ unique support points. ☐

Lemma 11. *Fix any positive definite matrix $\Sigma_0 \in \mathbb{R}^{d\times d}$ and denote by σ_0^2, the smallest eigenvalue of Σ_0. Then, there exists a universal constant $D > 0$ (only depending on d) such that for any distribution P and positive definite matrix Σ satisfying $\|\Sigma - \Sigma_0\| \leq \sigma_0^2/2$, we have*

$$\|\varphi_{P,\Sigma} - \varphi_{P,\Sigma_0}\|_{\mathbb{B}(T)} \leq DT^{(d+2)/2}\|\Sigma - \Sigma_0\|.$$

Proof of Lemma 11. For any distribution P and positive definite matrix Σ, we have that

$$\begin{aligned}\|\varphi_{P,\Sigma} - \varphi_{P,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \leq \int_{\|t\|_\infty \leq T} |\mathcal{F}[P](t)|^2 \left|e^{-t'\Sigma t/2} - e^{-t'\Sigma_0 t}\right|^2 dt &\leq \int_{\|t\|_\infty \leq T} \left|e^{-t'\Sigma t/2} - e^{-t'\Sigma_0 t/2}\right|^2 dt \\ &= \int_{\|t\|_\infty \leq T} e^{-t'\Sigma_0 t} \left|1 - e^{t'(\Sigma_0 - \Sigma)t/2}\right|^2 dt.\end{aligned}$$

The mapping $t \to e^{t'(\Sigma_0 - \Sigma)t/2}$ has gradient norm at most $e^{\|t\|^2\|\Sigma-\Sigma_0\|/2}\|\Sigma - \Sigma_0\|\|t\|$. If the bound $\|\Sigma - \Sigma_0\| \leq \sigma_0^2/2$ holds, it follows that

$$\begin{aligned}\int_{\|t\|_\infty \leq T} e^{-t'\Sigma_0 t} \left|1 - e^{t'(\Sigma_0 - \Sigma)t/2}\right|^2 dt &\leq \int_{\|t\|_\infty \leq T} e^{-\|t\|^2\sigma_0^2} \left|1 - e^{t'(\Sigma_0 - \Sigma)t/2}\right|^2 dt \\ &\leq \|\Sigma - \Sigma_0\|^2 \int_{\|t\|_\infty \leq T} e^{-\|t\|^2\sigma_0^2/2}\|t\|^2 dt \\ &\leq D\|\Sigma - \Sigma_0\|^2 T^{d+2}.\end{aligned}$$

☐

Lemma 12. *Suppose $F \sim DP_\alpha$ where the base measure α is a Gaussian measure on $\mathbb{R}^d$. Fix any $q \in \mathbb{N}$. Then, there exists a universal constant $C > 0$ such that for any sequence $u_n \uparrow \infty$,*

$$\mathbb{P}\left(\left\{\int \|x\|^q dF(x)\right\}^{1/q} > u_n\right) \leq 2e^{-Cu_n^2}$$

holds for all sufficiently large n.

Proof of Lemma 12. Suppose $\alpha = N(\mu, \Sigma)$ for some mean vector $\mu \in \mathbb{R}^d$ and positive definite covariance matrix $\Sigma \in \mathbf{S}_+^d$. Without loss of generality, it suffices to verify the result with $\mu = 0$. By the stick breaking representation of DP_α we can write

$$F \overset{d}{=} \sum_{i=1}^\infty p_i \delta_{Z_i} \ , \quad \int \|x\|^q dF(x) \overset{d}{=} \sum_{i=1}^\infty p_i \|Z_i\|^q,$$

where $(Z_1, Z_2, \dots) \overset{i.i.d}{\sim} \alpha$ is independent of $(p_1, p_2, \dots)$ and $(p_i)_{i=1}^{\infty}$ are non-negative random variables with $\sum_{i=1}^{\infty} p_i = 1$. For an element $w = (w_1, w_2, \dots)$ with $w_i \in \mathbb{R}^d$, the $\ell^q(\mathbb{R}^d)$ norm is given by $\|w\|^q = \sum_{i=1}^{\infty}\|w_i\|^q$. Conditional on $(p_i)_{i=1}^{\infty}$, let $\mathcal{Z} = (p_i^{1/q} Z_1, p_2^{1/q} Z_2, \dots)$. Define $\gamma = (\mathbb{E}\|Z_1\|^q)^{1/q}$ and note that $\mathbb{E}\|\mathcal{Z}\|^q = \sum_{i=1}^{\infty} p_i \mathbb{E}\|Z_i\|^q = \gamma^q < \infty$. In particular, we can view $\mathcal{Z}$ (defined conditional on $(p_i)_{i=1}^{\infty}$) as a mean-zero Gaussian random element on $\ell^q(\mathbb{R}^d)$. From an application of (Giné and Nickl, 2021, Theorem 2.1.20), it follows that

$$\mathbb{P}\bigg(\|\mathcal{Z}\| > u + \gamma\bigg) \leq 2e^{-\frac{u^2}{2\gamma^2}} \quad \forall\, u > 0.$$

The bound holds conditionally. However, as the term on the right is independent of $(p_i)_{i=1}^{\infty}$, it also holds unconditionally and we obtain

$$\mathbb{P}\Bigg(\bigg\{\int \|x\|^q dF(x)\bigg\}^{1/q} > u + \gamma\Bigg) \leq 2e^{-\frac{u^2}{2\gamma^2}} \quad \forall\, u > 0.$$

The claim follows by letting $u = u_n$ and noting that $u_n > 2\gamma$ for all sufficiently large n.

□

Lemma 13. *For $d \geq 2$, let $\Omega \subset \mathbb{R}^d$ denote a ball (with respect to any norm) of radius R. Let $f : \Omega \to \mathbb{C}$ be such that $\nabla f \in C(\overline{\Omega})$. Then, there exists a universal constant $C > 0$ such that*

$$\|f\|_{L^2(\Omega)} \leq C\bigg(R\|\nabla f\|_{L^2(\Omega)} + \sqrt{R}\|f\|_{L^2(\partial\Omega)}\bigg),$$

where $L^2(\partial\Omega)$ is the L^2 norm on $\partial\Omega$ with respect to the $d-1$ dimensional Hausdorff measure.

Proof of Lemma 13. Let $\lambda(.)$ denote the Lebesgue measure on $\mathbb{R}^d$. Denote the $d-1$ dimensional Hausdorff measure by$\mathcal{H}^{d-1}$. Define the exponents

$$p = \frac{2d}{d+2}\ ,\ q = \frac{2(d-1)}{d}.$$

An application of (Maggi and Villani, 2005, Theorem 1.2) on the real and complex parts of f separately implies that there exists a universal constant $D > 0$ (that depends only on p) such that

$$\|f\|_{L^2(\Omega)} \leq D(\|\nabla f\|_{L^p(\Omega)} + \|f\|_{L^q(\partial\Omega)}).$$

Since Ω is a ball of radius R, there exists a universal constant $C > 0$ such that

$$\mathcal{H}^{d-1}(\partial\Omega) \leq CR^{d-1}\ ,\ \lambda(\Omega) \leq CR^d.$$

By Hölder's inequality, it follows that

$$\|\nabla f\|_{L^p(\Omega)}^p = \int_\Omega \|\nabla f\|^p d\lambda = \int_\Omega \|\nabla f\|^{\frac{2d}{d+2}} d\lambda \leq \left(\int_\Omega \|\nabla f\|^2 d\lambda\right)^{\frac{d}{d+2}} [\lambda(\Omega)]^{\frac{2}{d+2}}.$$

Taking the $p = 2d/(d+2)$ root on both sides of the preceding inequality and using the estimate on $\lambda(\Omega)$ implies that there exists a universal constant $C > 0$ such that

$$\|\nabla f\|_{L^p(\Omega)} \leq C\|\nabla f\|_{L^2(\Omega)} R.$$

Similarly, for the boundary term, Hölder's inequality yields

$$\|f\|_{L^q(\partial\Omega)}^q = \int_{\partial\Omega} |f|^q \, d\mathcal{H}^{d-1} = \int_{\partial\Omega} |f|^{\frac{2(d-1)}{d}} \, d\mathcal{H}^{d-1} \leq \left(\int_{\partial\Omega} |f|^2 \, d\mathcal{H}^{d-1}\right)^{\frac{d-1}{d}} [\mathcal{H}^{d-1}(\partial\Omega)]^{1/d}.$$

Taking the $q = 2(d-1)/d$ root on both sides and using the estimate on $\mathcal{H}^{d-1}(\partial\Omega)$ implies that there exists a universal constant $C > 0$ such that

$$\|f\|_{L^q(\partial\Omega)} \leq C\|f\|_{L^2(\partial\Omega)}\sqrt{R}.$$

□

Proof of Theorem 1. The proof proceeds through several steps which we outline below. We use $D > 0$ as a generic universal constant that may change from line to line. For ease of notation, we suppress the dependence of $m = m_n$ and $G = G_n$ on n

Depending on whether the model is mildly or severely ill-posed, define λ as follows.

$$\lambda = \begin{cases} \max\{\chi^{-1}(d+2) + d/2, d+1\} & \text{mildly ill-posed} \\ \max\{\chi^{-1}(d+2) + d/2, d+1, d/\zeta + 1/2\} & \text{severely ill-posed.} \end{cases}$$

Let $\epsilon_n^2 = n^{-1}(\log n)^\lambda$.

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2\right) d\nu_{\alpha,G}(P,\Sigma) \geq \exp\left(-Cn\epsilon_n^2\right) \tag{81}$$

holds with $\mathbb{P}$ probability approaching 1.

Since $m = m_n \asymp n$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}_Y - \varphi_Y\|_{\mathbb{B}(T_n)}^2 = \int_{\mathbb{B}(T_n)} |\widehat{\varphi}_Y(t) - \varphi_Y(t)|^2 \, dt \leq D\frac{T_n^d \log(T_n)}{n},$$

$$\|\widehat{\varphi}_\epsilon - \varphi_\epsilon\|_{\mathbb{B}(T_n)}^2 = \int_{\mathbb{B}(T_n)} |\widehat{\varphi}_\epsilon(t) - \varphi_\epsilon(t)|^2 \, dt \leq D\frac{T_n^d \log(T_n)}{n}.$$

with $\mathbb{P}$ probability approaching 1. On this set (and using that $|\varphi_{P,\Sigma}| \leq 1$) we obtain

$$
\begin{aligned}
\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} &\leq D\Big(\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} + \frac{T_n^d \log(T_n)}{n} \Big) \\
&\leq D\Big(\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} + \epsilon_n^2 \Big) ,
\end{aligned}
$$

where the last inequality follows from the definition of T_n and ϵ_n^2.

It follows that

$$
\begin{aligned}
&\int \exp\Big(- nD\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \Big) d\nu_{\alpha,G}(P,\Sigma) \\
&\geq \exp(-nD\epsilon_n^2) \int \exp\Big(- nD\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \Big) d\nu_{\alpha,G}(P,\Sigma).
\end{aligned}
$$

By Condition 4.2, the mixing distribution F_0 satisfies $F_0(t \in \mathbb{R}^d : \|t\| > z) \leq C \exp(-C' z^\chi)$. There exists a universal constant $R > 0$ such that the cube $I_n = [-R(\log \epsilon_n^{-1})^{1/\chi}, R(\log \epsilon_n^{-1})^{1/\chi}]^d$ satisfies $1 - F_0(I_n) \leq D\epsilon_n$. Denote the probability measure induced from the restriction of F_0 to I_n by

$$
\overline{F}_0(A) = \frac{F(A \cap I_n)}{F(I_n)} \qquad \forall \text{ Borel } A \subseteq \mathbb{R}^d.
$$

Note that the restricted probability measure satisfies

$$
\sup_{t\in\mathbb{R}^d} \Big|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\Big| = \sup_{t\in\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} e^{\mathbf{i}t'x} d(F - \overline{F}_0)(x) \Big| \leq \|F_0 - \overline{F}_0\|_{TV} \leq 1 - F_0(I_n) \leq D\epsilon_n.
$$

By Condition 4.2, $f_X = \phi_{\Sigma_0} \star F_0$. As the eigenvalues of Σ_0 are bounded away from zero, it follows that

$$
\begin{aligned}
\|\varphi_Y - \varphi_\epsilon \varphi_{\overline{F}_0,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \leq D\|\varphi_Y - \varphi_\epsilon \varphi_{\overline{F}_0,\Sigma_0}\|^2_{L^2} &= \|\varphi_\epsilon \varphi_{F_0,\Sigma_0} - \varphi_\epsilon \varphi_{\overline{F}_0,\Sigma_0}\|^2_{L^2} \\
&= \int_{\mathbb{R}^d} |\varphi_{\Sigma_0}(t)|^2 \, |\varphi_\epsilon(t)|^2 \, \Big|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\Big|^2 dt \\
&\leq \int_{\mathbb{R}^d} |\varphi_{\Sigma_0}(t)|^2 \, \Big|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\Big|^2 dt \\
&\leq D\epsilon_n^2 \int_{\mathbb{R}^d} e^{-t'\Sigma_0 t} dt \\
&\leq D\epsilon_n^2.
\end{aligned}
$$

Let $\iota = \max\{d, d/\chi + d/2\}$. By Lemma 6, there exists a discrete probability measure $F_0^* = \sum_{i=1}^N p_i \delta_{\mu_i}$ where $N = D\big(\log(\epsilon_n^{-1})\big)^\iota$ and $\mu_i \in I_n$, that satisfies

$$
\|\varphi_\epsilon \varphi_{\overline{F}_0,\Sigma_0} - \varphi_\epsilon \varphi_{F_0^*,\Sigma_0}\|_{\mathbb{B}(T_n)} \leq D\|\varphi_{\overline{F}_0,\Sigma_0} - \varphi_{F_0^*,\Sigma_0}\|_{L^2} \leq D\epsilon_n.
$$

From the second claim of Lemma 6, we can also assume without loss of generality that the support points separation satisfies $\inf_{k\neq j}\|\mu_k - \mu_j\| \geq c_0\epsilon_n$ for some constant $c_0 > 0$

(depending only on Σ_0). From combining the preceding bounds, it follows that

$$\begin{aligned}
&\int \exp\Big(- nD\|\varphi_Y - \varphi_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \\
&\geq \exp\big(-nD\epsilon_n^2\big) \int \exp\Big(- nD\|\varphi_\epsilon\varphi_{\overline{F}_0,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \\
&\geq \exp\big(-nD\epsilon_n^2\big) \int \exp\Big(- nD\|\varphi_\epsilon\varphi_{F_0^*,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma).
\end{aligned}$$

Observe that if $\|\Sigma - \Sigma_0\|$ is sufficiently small, Lemma 7 implies that for any distribution P we have

$$\|\varphi_\epsilon\varphi_{P,\Sigma} - \varphi_\epsilon\varphi_{P,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \leq D\|\varphi_{P,\Sigma} - \varphi_{P,\Sigma_0}\|^2_{L^2} \leq D\|\Sigma - \Sigma_0\|^2.$$

In particular, for all such (P,Σ), this implies

$$\|\varphi_\epsilon\varphi_{F_0^*,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} \leq D\|\varphi_\epsilon\varphi_{F_0^*,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma_0}\|_{L^2} + D\|\Sigma - \Sigma_0\|.$$

From the preceding bounds, it follows that

$$\begin{aligned}
&\int \exp\Big(- nD\|\varphi_\epsilon\varphi_{F_0^*,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \\
&\geq \int_{(P,\Sigma):\|\Sigma-\Sigma_0\|\leq D\epsilon_n, \|\varphi_{F_0^*,\Sigma_0} - \varphi_{P,\Sigma_0}\|_{L^2}\leq D\epsilon_n} \exp\Big(- nD\|\varphi_\epsilon\varphi_{F_0^*,\Sigma_0} - \varphi_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \\
&\geq \exp\big(-nD\epsilon_n^2\big) \int_{(P,\Sigma):\|\Sigma-\Sigma_0\|\leq D\epsilon_n, \|\varphi_{F_0^*,\Sigma_0} - \varphi_{P,\Sigma_0}\|_{L^2}\leq D\epsilon_n} d\nu_{\alpha,G}(P,\Sigma).
\end{aligned}$$

Let $V_i = \{t \in \mathbb{R}^d : \|t - \mu_i\| \leq c_0\epsilon_n/2\}$ for $i = 1, \ldots, N$ and $V_0 = \mathbb{R}^d \setminus \bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \ldots, V_N\}$ is a disjoint partition of $\mathbb{R}^d$. For any fixed (P,Σ), an application of Lemma 4 yields

$$\|\varphi_{P,\Sigma_0} - \varphi_{F_0^*,\Sigma_0}\|_{L^2} \leq D\bigg(\epsilon_n^d + \sum_{j=1}^N |P(V_j) - p_j|\bigg).$$

Define $\mathcal{G}_n = \{(P,\Sigma) : \sum_{j=1}^N |P(V_j) - p_j| \leq \epsilon_n,\ \|\Sigma - \Sigma_0\| \leq \epsilon_n\}$. The preceding bounds imply that

$$\int_{(P,\Sigma):\|\Sigma-\Sigma_0\|\leq D\epsilon_n, \|\varphi_{F_0^*,\Sigma_0} - \varphi_{P,\Sigma_0}\|_{L^2}\leq D\epsilon_n} d\nu_{\alpha,G}(P,\Sigma) \geq \int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\Sigma).$$

Since the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$, the integral appearing on the right can be expressed as

$$\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\Sigma) = \int_{\Sigma:\|\Sigma-\Sigma_0\|\leq\epsilon_n} \int_{P:\sum_{j=1}^N |P(V_j)-p_j|\leq\epsilon_n} d\mathrm{DP}_\alpha(P) dG(\Sigma).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^N \alpha(V_j) \geq C\epsilon_n^d \exp\bigl(-C'(\log \epsilon_n^{-1})^{2/\chi}\bigr)$ for universal constants $C, C' > 0$. By definition of DP_α, $(P(V_1), \ldots, P(V_N)) \sim \mathrm{Dir}(N, \alpha(V_1), \ldots, \alpha(V_N))$. As $N = D\{\log\bigl(\epsilon_n^{-1}\bigr)\}^\iota$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) implies

$$\begin{aligned}\int_{P:\sum_{j=1}^N |P(V_j)-p_j| \leq \epsilon_n} d\mathrm{DP}_\alpha(P) \geq C\exp\bigl(-C'(\log \epsilon_n^{-1})^{\iota+\max\{2/\chi,1\}}\bigr) &= C\exp\bigl(-C'(\log \epsilon_n^{-1})^{\lambda}\bigr) \\ &\geq C\exp\bigl(-C'' n\epsilon_n^2\bigr)\end{aligned}$$

for universal constants $C, C', C'' > 0$.

It remains to bound the outer integral. The law of $G = G_n$ is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2 and the definition of σ_n^2, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\Sigma:\|\Sigma-\Sigma_0\| \leq \epsilon_n} dG(\Sigma) = \int_{\Sigma:\|\Sigma-\sigma_n^2\Sigma_0\| \leq \sigma_n^2\epsilon_n} dL(\Sigma) \geq C\exp\bigl(-C'\sigma_n^{-2\kappa}\bigr) \geq C\exp\bigl(-C'' n\epsilon_n^2\bigr)$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(ii) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} > 2E\epsilon_n^2} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\right) d\nu_{\alpha,G}(P,\Sigma) \leq \exp\bigl(-nE\epsilon_n^2\bigr).$$

The law of $G = G_n$ is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\Sigma:\|\Sigma^{-1}\| > E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) = \int_{\Sigma:\|\Sigma^{-1}\| > E(n\epsilon_n^2)^{1/\kappa}} dL(\Sigma) \leq \exp\bigl(-E' n\epsilon_n^2\bigr).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \geq 0$, the preceding bound implies

$$\begin{aligned}&\int_{\Sigma:\|\Sigma^{-1}\| > E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\right) d\nu_{\alpha,G}(P,\Sigma) \\ &\leq \int_{\Sigma:\|\Sigma^{-1}\| > E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) \\ &\leq \exp\bigl(-E' n\epsilon_n^2\bigr).\end{aligned}$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$

sufficiently large such that

$$\int_{(P,\Sigma):\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\le E\epsilon_n^2,\|\Sigma^{-1}\|\le E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\right)d\nu_{\alpha,G}(P,\Sigma)$$

$$\ge \exp\left(-E'n\epsilon_n^2\right).$$

(*iii*) We prove the main statement of the theorem. From the bounds derived in steps (*i*) and (*ii*), it follows that for any $C'>0$, there exists a $M>0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\le M^2\epsilon_n^2, \|\Sigma^{-1}\|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n,T_n\right)\ge 1-\exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

For any choice of (P,Σ) satisfying $\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\le M\epsilon_n$, an application of Lemma 2 (and noting that $|\varphi_{P,\Sigma}|\le 1$) yields

$$\begin{aligned}
\|\varphi_Y-\varphi_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} &\le \|\varphi_Y-\widehat{\varphi_Y}\|_{\mathbb{B}(T_n)}+\|(\varphi_\epsilon-\widehat{\varphi_\epsilon})\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}+\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\
&\le \|\varphi_Y-\widehat{\varphi_Y}\|_{\mathbb{B}(T_n)}+\|\varphi_\epsilon-\widehat{\varphi_\epsilon}\|_{\mathbb{B}(T_n)}+\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\
&\le D\frac{\sqrt{T_n^d\log T_n}}{\sqrt{n}}+\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\
&\le D\epsilon_n+\|\widehat{\varphi}_Y-\widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\
&\le D\epsilon_n.
\end{aligned}$$

Let $\tau_T=\sup_{\|t\|_\infty\le T}|\varphi_\epsilon(t)|^{-1}$. Since $\varphi_Y=\varphi_X\varphi_\epsilon$, the preceding bound implies that

$$\begin{aligned}
D\epsilon_n\ge\|\varphi_Y-\varphi_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} &= \|(\varphi_X-\varphi_{P,\Sigma})\varphi_\epsilon\|_{\mathbb{B}(T_n)}\\
&\ge \tau_{T_n}^{-1}\|(\varphi_X-\varphi_{P,\Sigma})\mathbb{1}\{t\in\mathbb{B}(T_n)\}\|_{L^2}.
\end{aligned}$$

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c>0$ for which $\lambda_1(\Sigma)\ge c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}\asymp(\log n)(\log\log n)$ in the mildly ill-posed case. In the severely ill-posed case, we have

$$T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}\asymp\begin{cases}(\log n)(\log\log n) & \zeta\in(0,2)\\ \log n & \zeta=2.\end{cases}$$

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_X - \varphi_{P,\Sigma})\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 &\leq 2\|\varphi_X \mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 + 2\|\varphi_{P,\Sigma}\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 \\ &\leq 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma t} dt \\ &\leq 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2\int_{\|t\|_\infty > T_n} e^{-c\|t\|^2 \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} dt \\ &\leq D\left[e^{-CT_n^2} T_n^{d-2} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa} e^{-CT_n^2 \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} T_n^{d-2}\right].\end{aligned}$$

For all mildly ill-posed models and severely ill-posed with $\zeta \in (0, 2)$, the preceding bound reduces to Dn^{-1}. For severely ill-posed models with $\zeta = 2$ it reduces to Dn^{-2K} for some constant $K \in (0, 1/2]$ (that depends on, among other factors, the smallest eigenvalue of Σ_0).

We verify the conclusion of the theorem. Suppose that the model is mildly ill-posed or severely ill-posed with $\zeta \in (0, 2)$. From combining the preceding bounds (and noting that $n^{-1} \lessapprox \tau_{T_n}\epsilon_n$), it follows that for every $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\varphi_X - \varphi_{P,\Sigma}\|_{L^2} \leq M\tau_{T_n}\epsilon_n \,|\mathcal{Z}_n, T_n\right) \geq 1 - \exp\left(-C'n\epsilon_n^2\right).$$

holds with $\mathbb{P}$ probability approaching 1. If the model is severely ill-posed with constant K as specified above, we have

$$\nu_{\alpha,G}\left(\|\varphi_X - \varphi_{P,\Sigma}\|_{L^2} \leq M\tau_{T_n}\epsilon_n + n^{-K} \,|\mathcal{Z}_n, T_n\right) \geq 1 - \exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

For mildly ill-posed models, the claim follows from observing that

$$\tau_{T_n}\epsilon_n \asymp \frac{(\log n)^{(\lambda+\zeta)/2}}{\sqrt{n}}(\log\log n)^{\zeta/2}.$$

Similarly, for severely ill-posed models with $\zeta \in (0, 2]$ and $T_n = (c_0 \log n)^{1/\zeta}$ for some c_0 satisfying $c_0 R = \gamma < 1/2$, we have

$$\tau_{T_n}\epsilon_n \asymp n^{-1/2}(\log n)^{\lambda/2} e^{\gamma \log n} = n^{\gamma - 1/2}(\log n)^{\lambda/2}.$$

□

Proof of Corollary 1. The analysis is analogous to that of Theorem 1. Let $\tau_T = \sup_{\|t\|_\infty \leq T} |\varphi_\epsilon(t)|^{-1}$. From the proof of that result, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\varphi_X - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2 \leq M^2\tau_{T_n}^2\epsilon_n^2, \|\Sigma^{-1}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n\right) \geq 1 - \exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1. For any (P, Σ) an application of Cauchy-Schwarz yields

$$\int_{\mathbb{B}(T_n)} |\varphi_X(t) - \varphi_{P,\Sigma}(t)|\, dt \leq D T_n^{d/2} \|\varphi_X - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}.$$

In particular, convergence rates for the quantity on the right imply convergence rates for the left up to a multiplicative factor of $T_n^{d/2}$. It remains to examine the bias from truncating the L^1 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c > 0$ for which $\lambda_1(\Sigma) \geq c(n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2 (n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2} \asymp (\log n)(\log\log n)$ in the mildly ill-posed case. In the severely ill-posed case, we have

$$T_n^2 (n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2} \asymp \begin{cases} (\log n)(\log\log n) & \zeta \in (0,2) \\ \log n & \zeta = 2. \end{cases}$$

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned} \|(\varphi_X - \varphi_{P,\Sigma}) \mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^1} &\leq \|\varphi_X \mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^1} + \|\varphi_{P,\Sigma} \mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^1} \\ &\leq \int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t/2} dt + \int_{\|t\|_\infty > T_n} e^{-t'\Sigma t/2} dt \\ &\leq \int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t/2} dt + \int_{\|t\|_\infty > T_n} e^{-c\|t\|^2 \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa}/2} dt \\ &\leq D\left[e^{-CT_n^2} T_n^{d-2} + \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa} e^{-CT_n^2 \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa}} T_n^{d-2} \right]. \end{aligned}$$

For all mildly ill-posed models and severely ill-posed with $\zeta \in (0, 2)$, the preceding bound reduces to $Dn^{-1/2}$. For severely ill-posed models with $\zeta = 2$ it reduces to Dn^{-K} for some constant $K \in (0, 1/2]$ (that depends on, among other factors, the smallest eigenvalue of Σ_0).

From combining the preceding bounds, we obtain (similarly to the concluding remarks in the proof of Theorem 1) contraction rates for $\|\varphi_X - \varphi_{P,\Sigma}\|_{L^1}$. The claim then follows from observing that $\|f_X - \phi_{P,\Sigma}\|_{L^\infty} \leq \|\varphi_X - \varphi_{P,\Sigma}\|_{L^1}$.

□

Proof of Theorem 2. The proof proceeds through several steps which we outline below. We use $D > 0$ as a generic universal constant that may change from line to line. For ease of notation, we suppress the dependence of $m = m_n$ and $G = G_n$ on n.

Define

$$\alpha_n = \epsilon_n^{1/(p+\zeta)} \quad , \quad \epsilon_n^2 = \frac{(\log n)^{\lambda + d/2}}{n^{\frac{2(p+\zeta)}{2(p+\zeta)+d}}} \quad , \quad \lambda = \begin{cases} \chi^{-1}(d+2) & \chi < 2 \\ d/\chi + 1 & \chi \geq 2. \end{cases}$$

(i) First, we derive a lower bound for the normalizing constant of the posterior measure.

Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \geq \exp\big(-Cn\epsilon_n^2\big) \tag{82}$$

holds with $\mathbb{P}$ probability approaching 1.

Since $m = m_n \asymp n$, an application of Lemma 2 implies that

$$\begin{aligned}\|\widehat{\varphi}_Y - \varphi_Y\|^2_{\mathbb{B}(T_n)} &= \int_{\mathbb{B}(T_n)} |\widehat{\varphi}_Y(t) - \varphi_Y(t)|^2\, dt \leq D\frac{T_n^d \log(T_n)}{n}\ ,\\ \|\widehat{\varphi}_\epsilon - \varphi_\epsilon\|^2_{\mathbb{B}(T_n)} &= \int_{\mathbb{B}(T_n)} |\widehat{\varphi}_\epsilon(t) - \varphi_\epsilon(t)|^2\, dt \leq D\frac{T_n^d \log(T_n)}{n}.\end{aligned}$$

with $\mathbb{P}$ probability approaching 1. On this set (and using that $|\varphi_{P,\Sigma}| \leq 1$) we obtain

$$\begin{aligned}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} &\leq D\Big(\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} + \frac{T_n^d \log(T_n)}{n}\Big)\\ &\leq D\Big(\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} + \epsilon_n^2\Big)\ ,\end{aligned}$$

where the last inequality follows from the definition of T_n and ϵ_n. It follows that

$$\begin{aligned}&\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)\\ &\geq \exp(-nD\epsilon_n^2) \int \exp\Big(-nD\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma).\end{aligned}$$

By Condition 4.3, there exists universal constants $\chi, C, M < \infty$ and a mixing distribution S_{α_n} supported on the cube $I_n = [-C(\log \epsilon_n^{-1})^{1/\chi}, C(\log \epsilon_n^{-1})^{1/\chi}]^d$ that satisfies

$$\|f_X \star f_\epsilon - \phi_{S_{\alpha_n},\alpha_n^2 I} \star f_\epsilon\|_{L^2} \leq D\epsilon_n^2.$$

Note that $\mathcal{F}[f_X \star f_\epsilon] = \varphi_X \varphi_\epsilon$ and $\mathcal{F}[\phi_{S_{\alpha_n},\alpha_n^2 I} \star f_\epsilon] = \varphi_\epsilon \varphi_{S_{\alpha_n},\alpha_n^2 I}$. As the Fourier transform preserves L^2 distance (up to a constant), it follows that

$$\begin{aligned}\|\varphi_Y - \varphi_\epsilon \varphi_{S_{\alpha_n},\alpha_n^2 I}\|^2_{\mathbb{B}(T_n)} \leq D\|\varphi_Y - \varphi_\epsilon \varphi_{S_{\alpha_n},\alpha_n^2 I}\|^2_{L^2} &= D\|\varphi_X \varphi_\epsilon - \varphi_\epsilon \varphi_{S_{\alpha_n},\alpha_n^2 I}\|^2_{L^2}\\ &\leq D\|f_X \star f_\epsilon - \phi_{S_{\alpha_n},\alpha_n^2 I} \star f_\epsilon\|^2_{L^2}\\ &\leq D\alpha_n^{2(p+\zeta)}\\ &\leq D\epsilon_n^2.\end{aligned}$$

From combining the preceding bounds, we obtain

$$\begin{aligned}&\int \exp\Big(-nD\|\varphi_Y - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)\\ &\geq \exp\big(-nD\epsilon_n^2\big) \int \exp\Big(-nD\|\varphi_\epsilon \varphi_{S_{\alpha_n},\alpha_n^2 I} - \varphi_\epsilon \varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma).\end{aligned}$$

Fix any (P, Σ) in the support of $\nu_{\alpha,G}$. Since $|\varphi_\epsilon| \leq 1$, we have that

$$\|\varphi_\epsilon \varphi_{S_{\alpha_n}, \alpha_n^2 I} - \varphi_\epsilon \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} \leq \|\varphi_{S_{\alpha_n}, \alpha_n^2 I} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}.$$

By an application of Lemma 10, there exists a discrete probability measure $F_{\alpha_n} = \sum_{i=1}^N p_i \delta_{\mu_i}$ with $N \leq D T_n^d \{\log(\epsilon_n^{-1})\}^{d/\chi}$ and $\mu_i \in I_n$ that satisfies

$$\|\varphi_{S_{\alpha_n}, \alpha_n^2 I} - \varphi_{F_{\alpha_n}, \alpha_n^2 I}\|_{\mathbb{B}(T_n)} \leq D\epsilon_n.$$

From the second claim of Lemma 10, we can also assume without loss of generality that the support points have separation satisfying $\inf_{k \neq j} \|\mu_k - \mu_j\| \geq \epsilon_n T_n^{-(d+2)/2}$. From combining the preceding bounds, it follows that

$$\begin{aligned}
&\int \exp\left(- nD \|\varphi_\epsilon \varphi_{S_{\alpha_n}, \alpha_n^2 I} - \varphi_\epsilon \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \exp(-nD\epsilon_n^2) \int \exp\left(- nD \|\varphi_{F_{\alpha_n}, \alpha_n^2 I} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \Sigma).
\end{aligned}$$

Define the set

$$\Omega_n = \left\{ \Sigma \in \mathbf{S}_+^d : \lambda_j(\Sigma) \in \left[\frac{\alpha_n^2}{1 + \epsilon_n T_n^{-(d+2)/2}}, \alpha_n^2 \right] \ \forall\, j = 1, \ldots, d. \right\}$$

Observe that for any distribution P and $\Sigma \in \Omega_n$, an application of Lemma 11 yields

$$\begin{aligned}
\|\varphi_{P, \alpha_n^2 I} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} \leq D T_n^{(d+2)/2} \|\Sigma - \alpha_n^2 I\| &= D T_n^{(d+2)/2} \max_{j=1,\ldots,d} \left| \lambda_j(\Sigma) - \alpha_n^2 \right| \\
&\leq D\epsilon_n.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\int \exp\left(- nD \|\varphi_{F_{\alpha_n}, \alpha_n^2 I} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \int_{\Sigma \in \Omega_n} \exp\left(- nD \|\varphi_{F_{\alpha_n}, \alpha_n^2 I} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \exp(-nD\epsilon_n^2) \int_{\Sigma \in \Omega_n} \exp\left(- nD \|\varphi_{F_{\alpha_n}, \alpha_n^2 I} - \varphi_{P,\alpha_n^2 I}\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \Sigma).
\end{aligned}$$

Define $V_i = \{t \in \mathbb{R}^d : \|t - \mu_i\| \leq \epsilon_n^2 T_n^{-(d+2)/2}\}$ for $i = 1, \ldots, N$ and $V_0 = \mathbb{R}^d \setminus \bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \ldots, V_N\}$ is a disjoint partition of $\mathbb{R}^d$. For any fixed distribution P an application of Lemma 4 yields

$$\|\varphi_{P, \alpha_n^2 I} - \varphi_{F_{\alpha_n}, \alpha_n^2 I}\|_{\mathbb{B}(T_n)} \leq D \left[\epsilon_n + T_n^{d/2} \sum_{j=1}^N |P(V_j) - p_j| \right].$$

The preceding bound implies that

$$\int_{\Sigma\in\Omega_n} \exp\Big(- nD\|\varphi_{F_{\alpha_n},\alpha_n^2 I} - \varphi_{P,\alpha_n^2 I}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq \int_{(P,\Sigma):\Sigma\in\Omega_n,\sum_{j=1}^N |P(V_j)-p_j|\leq \epsilon_n T_n^{-d/2}} \exp\Big(- nD\|\varphi_{F_{\alpha_n},\alpha_n^2 I} - \varphi_{P,\alpha_n^2 I}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq \exp\big(-nD\epsilon_n^2\big) \int_{(P,\Sigma):\Sigma\in\Omega_n,\sum_{j=1}^N |P(V_j)-p_j|\leq \epsilon_n T_n^{-d/2}} d\nu_{\alpha,G}(P,\Sigma).$$

Since the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$, the integral appearing on the right can be expressed as

$$\int_{\Sigma\in\Omega_n}\int_{P:\sum_{j=1}^N |P(V_j)-p_j|\leq \epsilon_n T_n^{-d/2}} d\mathrm{DP}_\alpha(P) dG(\Sigma).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^N \alpha(V_j) \geq C\epsilon_n^{2d} T_n^{-d(d+2)/2} \exp\big(-C'(\log\epsilon_n^{-1})^{2/\chi}\big)$ for universal constants $C, C' > 0$. By definition of DP_α, $(P(V_1),\ldots,P(V_N)) \sim \mathrm{Dir}(N,\alpha(V_1),\ldots,\alpha(V_N))$. As $N = DT_n^d\{\log\big(\epsilon_n^{-1}\big)\}^{d/\chi}$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) and the definition of (T_n,ϵ_n^2) implies

$$\int_{P:\sum_{j=1}^N |P(V_j)-p_j|\leq \epsilon_n T_n^{-d/2}} d\mathrm{DP}_\alpha(P) \geq C\exp\big(- C'T_n^d\{\log\epsilon_n^{-1}\}^{d/\chi+\max\{2/\chi,1\}}\big)$$
$$\geq C\exp\big(-C'n\epsilon_n^2\big).$$

It remains to bound the outer integral. The law of $G = G_n$ is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2 and the definition of $(\alpha_n^2,\sigma_n^2,\epsilon_n^2)$, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\Sigma\in\Omega_n} dG(\Sigma) = \int_{\Sigma\in\sigma_n^2\Omega_n} dL(\Sigma) \geq C\exp\big(- C'\sigma_n^{-2\kappa}\alpha_n^{-2\kappa}\big) \geq C\exp\big(-C''n\epsilon_n^2\big).$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(ii) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} > 2E\epsilon_n^2} \exp\Big(- \frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \leq \exp\big(- nE\epsilon_n^2\big).$$

The law of $G = G_n$ is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) = \int_{\Sigma:\|\Sigma^{-1}\|>E(n\epsilon_n^2)^{1/\kappa}} dL(\Sigma) \leq \exp\big(-E'n\epsilon_n^2\big).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \geq 0$, the preceding bound implies

$$\begin{aligned}&\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\right)d\nu_{\alpha,G}(P,\Sigma)\\ &\leq \int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma)\\ &\leq \exp\big(-E'n\epsilon_n^2\big).\end{aligned}$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$ sufficiently large such that

$$\int_{(P,\Sigma):\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\leq E\epsilon_n^2,\|\Sigma^{-1}\|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\right)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq \exp\big(-E'n\epsilon_n^2\big).$$

(*iii*) We prove the main statement of the theorem. From the bounds derived in steps (*i*) and (*ii*), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)}\leq M^2\epsilon_n^2, \|\Sigma^{-1}\|\leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n, T_n\right) \geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

For any choice of (P,Σ) satisfying $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\leq M\epsilon_n$, an application of Lemma 2 (and noting that $|\varphi_{P,\Sigma}| \leq 1$) yields

$$\begin{aligned}\|\varphi_Y - \varphi_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} &\leq \|\varphi_Y - \widehat{\varphi_Y}\|_{\mathbb{B}(T_n)} + \|(\varphi_\epsilon - \widehat{\varphi_\epsilon})\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} + \|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\ &\leq \|\varphi_Y - \widehat{\varphi_Y}\|_{\mathbb{B}(T_n)} + \|\varphi_\epsilon - \widehat{\varphi_\epsilon}\|_{\mathbb{B}(T_n)} + \|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\ &\leq D\frac{\sqrt{T_n^d\log T_n}}{\sqrt{n}} + \|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\ &\leq D\epsilon_n + \|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\\ &\leq D\epsilon_n.\end{aligned}$$

Since $\varphi_Y = \varphi_X\varphi_\epsilon$ and $\inf_{\|t\|_\infty\leq T_n}|\varphi_\epsilon(t)| \geq \tau_{T_n}^{-1}$, the preceding bound implies that

$$\begin{aligned}D\epsilon_n \geq \|\varphi_Y - \varphi_\epsilon\varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} &= \|(\varphi_X - \varphi_{P,\Sigma})\varphi_\epsilon\|_{\mathbb{B}(T_n)}\\ &\geq \tau_{T_n}^{-1}\|(\varphi_X - \varphi_{P,\Sigma})\mathbb{1}\{t \in \mathbb{B}(T_n)\}\|_{L^2}.\end{aligned}$$

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c > 0$ for which $\lambda_1(\Sigma) \geq c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp (\log n)(\log\log n)$.

Since $f_X \in \mathbf{H}^p(M)$, we have that

$$\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2\, dt \leq DT_n^{-2p}.$$

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}
\|(\varphi_X - \varphi_{P,\Sigma})\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 &\leq 2\|\varphi_X\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 + 2\|\varphi_{P,\Sigma}\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 \\
&\leq 2\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2\, dt + 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma t} dt \\
&\leq 2\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2\, dt + 2\int_{\|t\|_\infty > T_n} e^{-c\|t\|^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} dt \\
&\leq D\Big[T_n^{-2p} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}e^{-CT_n^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}}T_n^{d-2}\Big].
\end{aligned}$$

Since $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp (\log n)(\log\log n)$, the preceding bound reduces to DT_n^{-2p}. From combining the preceding bounds (and noting that $T_n^{-p} \lessapprox \tau_{T_n}\epsilon_n$), it follows that for every $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\varphi_X - \varphi_{P,\Sigma}\|_{L^2} \leq M\tau_{T_n}\epsilon_n \,|\mathcal{Z}_n, T_n\Big) \geq 1 - \exp\big(-C'n\epsilon_n^2\big).$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$\tau_{T_n}\epsilon_n \asymp T_n^{\zeta}\epsilon_n \asymp n^{-p/[2(p+\zeta)+d]}(\log n)^{(\lambda+\zeta)/2+d/4}.$$

□

Proof of Theorem 3. The proof proceeds through several steps which we outline below. We use $D > 0$ as a generic universal constant that may change from line to line. Define

$$\epsilon_n^2 = n^{-1}(\log n)^{\lambda} \quad , \quad \lambda = \begin{cases} \chi^{-1}(d+2) + d/2 & \chi < 2 \\ d+1 & \chi \geq 2. \end{cases}$$

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|_{\mathbb{B}(T_n)}^2\Big) d\nu_{\alpha,G}(P,\Sigma) \geq \exp\big(-Cn\epsilon_n^2\big) \tag{83}$$

holds with $\mathbb{P}$ probability approaching 1.

Fix $\epsilon > 0$ sufficiently small. By Condition 4.5, the mixing distribution F_0 satisfies $F_0(t \in$

$\mathbb{R}^d : \|t\|> z) \leq C\exp(-C'z^\chi)$ for some $C, C' > 0$. Fix a universal constant $R > 0$ such that $\exp(-C'z^\chi) \leq \epsilon$ for every $z > R(\log \epsilon^{-1})^{1/\chi}$. In particular, note that the cube $I = [-R(\log \epsilon^{-1})^{1/\chi}, R(\log \epsilon^{-1})^{1/\chi}]^d$ satisfies $1 - F_0(I) \leq D\epsilon$. Denote the probability measure induced from the restriction of F_0 to I by

$$\overline{F}_0(A) = \frac{F_0(A \cap I)}{F_0(I)} \qquad \forall \text{ Borel } A \subseteq \mathbb{R}^d.$$

Observe that

$$\sup_{t\in\mathbb{R}^d} \left|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\right| = \sup_{t\in\mathbb{R}^d} \left|\int_{\mathbb{R}^d} e^{\mathbf{i}t'x} d(F - \overline{F}_0)(x)\right| \leq \|F_0 - \overline{F}_0\|_{TV} \leq 1 - F_0(I) \leq D\epsilon.$$

For all sufficiently large $M > 0$, the tail bound on the mixing distribution F_0 implies

$$\begin{aligned}
\mathbb{E}[\|X\|\mathbb{1}\{\|X\|> M\}] &= \int_0^\infty \mathbb{P}(\|X\|> M, \|X\|> t)dt \\
&\leq M\mathbb{P}(\|X\|> M) + \int_M^\infty \mathbb{P}(\|X\|> t)dt \\
&\leq D\left[M\exp\bigl(-C'M^\chi\bigr) + \int_M^\infty \exp\bigl(-C't^\chi\bigr)dt\right] \\
&\leq D\left[M\exp\bigl(-C'M^\chi\bigr) + \exp\bigl(-C'M^\chi\bigr)M^{1-\chi}\right].
\end{aligned}$$

It follows that

$$\begin{aligned}
\sup_{t\in\mathbb{R}^d}\|\nabla\varphi_{F_0}(t) - \nabla\varphi_{\overline{F}_0}(t)\| &= \sup_{t\in\mathbb{R}^d}\left\|\int_{\mathbb{R}^d} xe^{\mathbf{i}t'x} d(F_0 - \overline{F}_0)(x)\right\| \\
&\leq \sup_{t\in\mathbb{R}^d}\left\|\int_{x\in I} xe^{\mathbf{i}t'x} d(F_0 - \overline{F}_0)\right\| + \sup_{t\in\mathbb{R}^d}\left\|\int_{x\notin I} xe^{\mathbf{i}t'x} dF_0\right\| \\
&\leq D\left[(\log \epsilon^{-1})^{1/\chi}\|F_0 - \overline{F}_0\|_{TV} + (\log \epsilon^{-1})^{1/\chi}\epsilon\right] \\
&\leq D(\log \epsilon^{-1})^{1/\chi}\epsilon.
\end{aligned}$$

By Condition 4.5, $\inf_{\|t\|_\infty\leq T_n}|\varphi_{F_0}(t)| \geq c\exp\bigl(-c'T_n^2\bigr)$ for some $c, c' > 0$. Since $T_n^2 \lessapprox \log(n)$, it follows that there exists a $C_1 > 0$ such that

$$\inf_{\|t\|_\infty\leq T_n}|\varphi_{F_0}(t)| \geq n^{-C_1}.$$

In particular, we can choose $L > 1$ large enough such that the choice $\epsilon = \epsilon_n^L$ implies $\overline{F}_0$ has support contained in the cube $I_n = [-E(\log \epsilon_n^{-1})^{1/\chi}, E(\log \epsilon_n^{-1})^{1/\chi}]^d$ for some universal constant $E > 0$ and $T_n n^{C_1}(\log \epsilon_n^{-L})^{1/\chi}\epsilon_n^L \leq \epsilon_n$. Since $\sup_{\|t\|_\infty\leq T_n}\|\nabla \log\varphi_{F_0}(t)\| \lessapprox T_n$, it follows that

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F_0}(t)\| \frac{\left|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\right|}{|\varphi_{F_0}(t)|} \le \sup_{\|t\|_\infty \le T_n} \frac{\|\nabla \log \varphi_{F_0}(t)\|}{|\varphi_{F_0}(t)|} \sup_{t \in \mathbb{R}^d} \left|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\right| \le D\epsilon_n$$

$$\sup_{\|t\|_\infty \le T_n} \frac{\|\nabla \varphi_{F_0}(t) - \nabla \varphi_{\overline{F}_0}(t)\|}{|\varphi_{F_0}(t)|} \le \sup_{\|t\|_\infty \le T_n} |\varphi_{F_0}(t)|^{-1} \sup_{t \in \mathbb{R}^d} \|\nabla \varphi_{F_0}(t) - \nabla \varphi_{\overline{F}_0}(t)\| \le D\epsilon_n.$$

Since $\epsilon_n \downarrow 0$, observe that the preceding bound also implies

$$\left|\varphi_{\overline{F}_0}(t)\right| \ge |\varphi_{F_0}(t)| - \left|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\right| \ge |\varphi_{F_0}(t)| - D\epsilon_n |\varphi_{F_0}(t)| \ge \frac{1}{2} |\varphi_{F_0}(t)|$$

for all sufficiently large n.

It follows that

$$\begin{aligned}
&\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F_0}(t) - \nabla \log \varphi_{\overline{F}_0}(t)\| \\
&= \sup_{\|t\|_\infty \le T_n} \left\| \nabla \log \varphi_{F_0}(t) \frac{\varphi_{\overline{F}_0}(t) - \varphi_{F_0}(t)}{\varphi_{\overline{F}_0}(t)} + \frac{\nabla \varphi_{F_0}(t) - \nabla \varphi_{\overline{F}_0}(t)}{\varphi_{\overline{F}_0}(t)} \right\| \\
&\le \sup_{\|t\|_\infty \le T_n} \left(\left| \frac{\varphi_{\overline{F}_0}(t) - \varphi_{F_0}(t)}{\varphi_{\overline{F}_0}(t)} \right| \|\nabla \log \varphi_{F_0}(t)\| + \left\| \frac{\nabla \varphi_{F_0}(t) - \nabla \varphi_{\overline{F}_0}(t)}{\varphi_{\overline{F}_0}(t)} \right\| \right) \\
&\le 2 \sup_{\|t\|_\infty \le T_n} \left(\left| \frac{\varphi_{\overline{F}_0}(t) - \varphi_{F_0}(t)}{\varphi_{F_0}(t)} \right| \|\nabla \log \varphi_{F_0}(t)\| + \left\| \frac{\nabla \varphi_{F_0}(t) - \nabla \varphi_{\overline{F}_0}(t)}{\varphi_{F_0}(t)} \right\| \right) \\
&\le D\epsilon_n.
\end{aligned}$$

Next, we show that $\overline{F}_0$ can be suitably approximated by a discrete measure. Let $\iota = \max\{d, d/\chi + d/2\}$. By Lemma 5, there exists a discrete measure $F' = \sum_{i=1}^N p_i \delta_{\mu_i}$ with at most $N = D(\log \epsilon_n^{-1})^\iota$ support points on I_n such that

$$\sup_{\|t\|_\infty \le T_n} \left|\varphi_{\overline{F}_0}(t) - \varphi_{F'}(t)\right| \le T_n^{-1} n^{-C_1} \epsilon_n$$

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \varphi_{\overline{F}_0}(t) - \nabla \varphi_{F'}(t)\| \le n^{-C_1} \epsilon_n.$$

From the final claim of Lemma 5, we can also assume without loss of generality that the support points satisfy $\inf_{k \ne j} \|\mu_k - \mu_j\| \ge \epsilon_n^{L_2}$ for some $L_2 > 1$. Observe that $2 \inf_{\|t\|_\infty \le T_n} \left|\varphi_{\overline{F}_0}(t)\right| \ge \inf_{\|t\|_\infty \le T_n} |\varphi_{F_0}(t)| \ge n^{-C_1}$ and

$$\begin{aligned}
\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{\overline{F}_0}(t)\| &\le \sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F_0}(t) - \nabla \log \varphi_{\overline{F}_0}(t)\| + \sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F_0}(t)\| \\
&\le D(\epsilon_n + T_n) \\
&\le DT_n.
\end{aligned}$$

The preceding bounds imply

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{\overline{F}_0}(t)\| \frac{\left|\varphi_{\overline{F}_0}(t) - \varphi_{F'}(t)\right|}{\left|\varphi_{\overline{F}_0}(t)\right|} \le D\epsilon_n \ , \quad \sup_{\|t\|_\infty \le T_n} \frac{\|\nabla \varphi_{\overline{F}_0}(t) - \nabla \varphi_{F'}(t)\|}{\left|\varphi_{\overline{F}_0}(t)\right|} \le D\epsilon_n.$$

From an analogous argument to the bound for F_0 and $\overline{F}_0$, it follows that

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{\overline{F}_0}(t) - \nabla \log \varphi_{F'}(t)\| \le D\epsilon_n.$$

Fix any $L_3 > L_2$ sufficiently large such that

$$T_n^2 n^{C_1} \epsilon_n^{L_3} \le \epsilon_n \ , \ \ n^{C_1} \sqrt{n} \epsilon_n \sqrt{\log n} \epsilon_n^{L_3/2} \le \epsilon_n \ , \ T_n n^{C_1} (\log \epsilon_n^{-1})^{1/\chi} \epsilon_n^{L_3} \le \epsilon_n.$$

Define $V_i = \{t \in I_n : \|t - \mu_i\| \le \epsilon_n^{L_3}\}$ for $i = 1, \dots, N$ and set $V_0 = \mathbb{R}^d \setminus \bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \dots, V_N\}$ is a disjoint partition of $\mathbb{R}^d$. By Lemma 3, for any distribution P that satisfies $\int_{\mathbb{R}^d} \|x\|^2 dP(x) \le n\epsilon_n^2 \log n$ and $\sum_{j=1}^N |P(V_j) - p_j| \le \epsilon_n^{L_3}$, we have that

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F'}(t)\| \frac{|\varphi_P(t) - \varphi_{F'}(t)|}{|\varphi_{F'}(t)|} \le D\epsilon_n \ , \quad \sup_{\|t\|_\infty \le T_n} \frac{\|\nabla \varphi_P(t) - \nabla \varphi_{F'}(t)\|}{|\varphi_{F'}(t)|} \le D\epsilon_n.$$

For all such P, an analogous argument to the bound for F_0 and $\overline{F}_0$ implies that

$$\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_P(t) - \nabla \log \varphi_{F'}(t)\| \le D\epsilon_n.$$

From combining all the preceding bounds, observe that such all such P also satisfy

$$\begin{aligned} &\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_P(t) - \nabla \log \varphi_{F_0}(t)\| \le D\epsilon_n \\ &\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_P(t)\| \le D\epsilon_n + \sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{F_0}(t)\| \le DT_n. \end{aligned}$$

Given any positive definite Σ, the preceding bound also implies that

$$\begin{aligned} \sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_{P,\Sigma}(t)\| &\le \sup_{\|t\|_\infty \le T_n} \Big[\|\nabla \log \varphi_P(t)\| + \|\nabla \log \varphi_\Sigma(t)\| \Big] \\ &\le D\Big(T_n + T_n \|\Sigma\| \Big). \end{aligned}$$

By Lemma 2 and 9, we have

$$\begin{aligned} &\int_{\mathbb{B}(T_n)} |\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)|^2 \, dt \le D \frac{T_n^d \log(T_n)}{n} \ , \\ &\int_{\mathbb{B}(T_n)} \|\widehat{\varphi}_{Y_1,Y_2}(t) - \varphi_{Y_1,Y_2}(t)\|^2 dt \le D \frac{T_n^d \log(T_n)}{n}. \end{aligned}$$

with $\mathbb{P}$ probability approaching 1. On this set, when P satisfies the conditions specified above and $\|\Sigma\| \leq D$, we have that

$$\|\widehat{\varphi}_{Y_2}\nabla\log\varphi_{P,\Sigma}-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} \leq D\Bigg[\|\varphi_{Y_2}\nabla\log\varphi_{P,\Sigma}-\varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}+\frac{T_n^{d+1}\log(T_n)}{n}\Bigg]$$
$$\leq D\Bigg[\|\varphi_{Y_2}\nabla\log\varphi_{P,\Sigma}-\varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}+\epsilon_n^2\Bigg].$$

Define the set

$$\mathcal{G}_n=\left\{(P,\Sigma):\|\Sigma-\Sigma_0\|\leq\epsilon_n^2,\int_{\mathbb{R}^d}\|x\|^2dP(x)\leq n\epsilon_n^2\log(n)\,,\sum_{j=1}^N|P(V_j)-p_j|\leq\epsilon_n^{L_3}\right\}.$$

From combining all the preceding bounds, it follows that

$$\int\exp\bigg(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq\int_{\mathcal{G}_n}\exp\bigg(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq\exp\big(-nD\epsilon_n^2\big)\int_{\mathcal{G}_n}\exp\bigg(-nD\|\varphi_{Y_2}\nabla\log\varphi_{P,\Sigma}-\varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma).$$

Since $\varphi_{Y_1,Y_2}=\varphi_{Y_2}\nabla\log\varphi_X$ and $\varphi_X=\varphi_{F_0,\Sigma_0}$, the preceding integral can be expressed as

$$\int_{\mathcal{G}_n}\exp\bigg(-nD\|\varphi_{Y_2}\nabla\log\varphi_{F,\Sigma}-\varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma)$$
$$=\int_{\mathcal{G}_n}\exp\bigg(-nD\|\varphi_{Y_2}(\nabla\log\varphi_{P,\Sigma}-\nabla\log\varphi_{F_0,\Sigma_0})\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma).$$

For every $(P,\Sigma)\in\mathcal{G}_n$, the preceding bounds imply that

$$\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_{P,\Sigma}-\nabla\log\varphi_{F_0,\Sigma_0}\|$$
$$=\sup_{\|t\|_\infty\leq T_n}\|(\nabla\log\varphi_P-\nabla\log\varphi_{F_0})+(\nabla\log\varphi_\Sigma-\nabla\log\varphi_{\Sigma_0})\|$$
$$\leq\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_P-\nabla\log\varphi_{F_0}\|+\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_\Sigma-\nabla\log\varphi_{\Sigma_0}\|$$
$$\leq D\epsilon_n+D\|\Sigma-\Sigma_0\|T_n$$
$$\leq D\epsilon_n\,,$$

where we used that $\|\Sigma-\Sigma_0\|\leq\epsilon_n^2$ and $\epsilon_n^2T_n\leq\epsilon_n$. Since $\|\varphi_{Y_2}\|_{L^2}<\infty$, it follows that

$$\int_{\mathcal{G}_n}\exp\bigg(-nD\|\varphi_{Y_2}(\nabla\log\varphi_{P,\Sigma}-\nabla\log\varphi_{F_0,\Sigma_0})\|^2_{\mathbb{B}(T_n)}\bigg)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq\exp\big(-nD\epsilon_n^2\big)\int_{\mathcal{G}_n}d\nu_{\alpha,G}(P,\Sigma).$$

Define the sets

$$\mathcal{G}_{n,1} = \left\{(P,\Sigma) : \|\Sigma - \Sigma_0\| \leq \epsilon_n^2 \,, \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3}\right\} \,,$$

$$\mathcal{G}_{n,2} = \left\{(P,\Sigma) : \|\Sigma - \Sigma_0\| \leq \epsilon_n^2, \int_{\mathbb{R}^d} \|x\|^2 dP(x) > n\epsilon_n^2 \log(n) \,, \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3}\right\}.$$

Observe that $\mathcal{G}_n = \mathcal{G}_{n,1} \setminus \mathcal{G}_{n,2}$ and hence

$$\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\Sigma) = \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\Sigma) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\Sigma). \tag{84}$$

For the second term in (84), we have

$$\begin{aligned}\int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\Sigma) &\leq \int_{P:\int_{\mathbb{R}^d}\|x\|^2 dP(x) > n\epsilon_n^2 \log(n)} d\nu_{\alpha,G}(P,\Sigma) \\ &\leq \int_{P:\int_{\mathbb{R}^d}\|x\|^2 dP(x) > n\epsilon_n^2 \log(n)} d\mathrm{DP}_\alpha(P) \\ &\leq \exp\bigl(-Dn\epsilon_n^2 \log n\bigr) \,,\end{aligned}$$

where the second inequality is due to $\nu_{\alpha,G}$ being a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and the third inequality follows from an application of Lemma 12.

For the first term in (84), we have that

$$\int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\Sigma) = \int_{\Sigma:\|\Sigma-\Sigma_0\|\leq\epsilon_n^2} \int_{P:\sum_{j=1}^N |P(V_j)-p_j| \leq \epsilon_n^{L_3}} d\mathrm{DP}_\alpha(P) dG(\Sigma).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^N \alpha(V_j) \geq C\epsilon_n^{L_3 d} \exp\bigl(-C'(\log \epsilon_n^{-1})^{2/\chi}\bigr)$ for universal constants $C, C' > 0$. By definition of DP_α, $(P(V_1), \ldots, P(V_N)) \sim \mathrm{Dir}(N, \alpha(V_1), \ldots, \alpha(V_N))$. As $N = D\{\log\bigl(\epsilon_n^{-1}\bigr)\}^\iota$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) implies

$$\begin{aligned}\int_{P:\sum_{j=1}^N |P(V_j)-p_j| \leq \epsilon_n} d\mathrm{DP}_\alpha(P) \geq C \exp\bigl(-C'(\log \epsilon_n^{-1})^{\iota+\max\{2/\chi,1\}}\bigr) &= C \exp\bigl(-C'(\log \epsilon_n^{-1})^{\lambda}\bigr) \\ &\geq C \exp\bigl(-C'' n\epsilon_n^2\bigr).\end{aligned}$$

It remains to bound the outer integral. By Assumption 2, there exists a universal constant $C > 0$ and $q > 0$ such that

$$\int_{\Sigma:\|\Sigma-\Sigma_0\|\leq\epsilon_n^2} dG(\Sigma) \geq C\epsilon_n^q. \tag{85}$$

From combining the preceding bounds, it follows from (84) that

$$\begin{aligned}\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\Sigma) &= \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\Sigma) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\Sigma)\\ &\geq C\exp\big(-C''n\epsilon_n^2\big) - \exp\big(-Dn\epsilon_n^2\log n\big)\\ &\geq \exp\big(-Dn\epsilon_n^2\big).\end{aligned}$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(ii) We prove the main statement of the theorem. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}>2E\epsilon_n^2} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\Big)d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \exp\big(-nE\epsilon_n^2\big).$$

From combining the preceding bound with the one derived in step (i), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\leq M^2\epsilon_n^2\,\Big|\,\mathcal{Z}_n, T_n\Big) \geq 1-\exp\big(-C'n\epsilon_n^2\big).$$

holds with $\mathbb{P}$ probability approaching 1.

By an application of Lemma 9, we have that

$$\|\widehat{\varphi}_{Y_1,Y_2} - \varphi_{Y_1,Y_2}\|_{\mathbb{B}(T_n)}\leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

Since $\varphi_{Y_1,Y_2} = \varphi_{Y_2}\nabla\log\varphi_{F_0,\Sigma_0}$ and $\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_{F_0,\Sigma_0}\|\leq DT_n$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}_{Y_2}\nabla\log\varphi_{F_0,\Sigma_0} - \varphi_{Y_1,Y_2}\|_{\mathbb{B}(T_n)}\leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\begin{aligned}&\nu_{\alpha,G}\Big(\|\widehat{\varphi}_{Y_2}(\nabla\log\varphi_{P,\Sigma} - \nabla\log\varphi_{F_0,\Sigma_0})\|^2_{\mathbb{B}(T_n)}\leq M^2\epsilon_n^2\,\Big|\,\mathcal{Z}_n, T_n\Big)\\ &\geq 1-\exp\big(-C'n\epsilon_n^2\big)\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1.

Since $T_n = (c_0\log n)^{1/2}$ for some c_0 satisfying $c_0R = \gamma < 1/2$, an application of Lemma 2 implies that

$$\sup_{\|t\|_\infty \le T_n} \frac{|\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)|}{|\varphi_{Y_2}(t)|} \le D n^{-1/2+\gamma}\sqrt{\log\log n}$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, it follows that

$$|\widehat{\varphi}_{Y_2}(t)| \ge |\varphi_{Y_2}(t)| - |\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)| \ge \frac{1}{2}|\varphi_{Y_2}(t)| \ge \frac{1}{2}n^{-\gamma}$$

uniformly over the set $\{\|t\|_\infty \le T_n\}$, with $\mathbb{P}$ probability approaching 1. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\begin{aligned}&\nu_{\alpha,G}\Big(\|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{F_0,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \le M^2 n^{2\gamma}\epsilon_n^2 \,\Big|\, \mathcal{Z}_n, T_n\Big)\\ &\ge 1 - \exp\big(-C' n \epsilon_n^2\big)\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$\epsilon_n n^\gamma \asymp n^{-1/2+\gamma}(\log n)^{\lambda/2}.$$

□

Proof of Theorem 4. We use $D > 0$ as a generic universal constant that may change from line to line. Define

$$\epsilon_n^2 = n^{-1}(\log n)^\lambda \quad , \quad \lambda = \begin{cases} \frac{3}{\chi} + \frac{1}{2} & \chi < 2 \\ 2 & \chi \ge 2. \end{cases}$$

The proof continues from the conclusion of Theorem 3 with a few modifications to the preceding bounds to account for the change in covariance prior. For ease of notation, we suppress the dependence of $G = G_n$ on n.

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\Big(- \frac{n}{2}\|\widehat{\varphi}_{Y_2} \nabla \log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma) \ge \exp\big(- Cn\epsilon_n^2\big) \qquad (86)$$

holds with $\mathbb{P}$ probability approaching 1.

The argument is identical to part (*i*) of Theorem 3 except that Equation (85) is replaced with the following argument. The law of G is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2 and the definition of $(\sigma_n^2, \epsilon_n^2)$, there exists universal constant $C, C', C'' > 0$ such that

$$\int_{\Sigma:\|\Sigma-\Sigma_0\|\leq \epsilon_n^2} dG(\Sigma) = \int_{\Sigma:\|\Sigma-\sigma_n^2\Sigma_0\|\leq \sigma_n^2\epsilon_n^2} dL(\Sigma) \geq C\exp\big(-C'\sigma_n^{-2\kappa}\big) \geq C\exp\big(-C''n\epsilon_n^2\big).$$

The lower bound for the normalizing constant then follows from the conclusion of part (i) of Theorem 3.

(ii) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}>2E\epsilon_n^2} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \exp\big(-nE\epsilon_n^2\big).$$

The law of G is given by Σ/σ_n^2 where $\Sigma \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) = \int_{\Sigma:\|\Sigma^{-1}\|>E(n\epsilon_n^2)^{1/\kappa}} dL(\Sigma) \leq \exp\big(-E'n\epsilon_n^2\big).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and $\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} \geq 0$, the preceding bound implies

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma)$$
$$\leq \exp\big(-E'n\epsilon_n^2\big).$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$ sufficiently large such that

$$\int_{\substack{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\leq E\epsilon_n^2,\\ \|\Sigma^{-1}\|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq 1 - \exp\big(-E'n\epsilon_n^2\big).$$

(iii) We prove the main statement of the theorem. From the bounds derived in steps (i) and (ii), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} \leq M^2\epsilon_n^2, \|\Sigma^{-1}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\Big|\, \mathcal{Z}_n, T_n\Big)$$
$$\geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

By an application of Lemma 9, we have that

$$\|\widehat{\varphi}_{Y_1,Y_2} - \varphi_{Y_1,Y_2}\|_{\mathbb{B}(T_n)} \leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

Since $\varphi_{Y_1,Y_2} = \varphi_{Y_2} \nabla \log \varphi_{F_0,\Sigma_0}$ and $\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_{F_0,\Sigma_0}\| \leq DT_n$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}_{Y_2} \nabla \log \varphi_{F_0,\Sigma_0} - \varphi_{Y_1,Y_2}\|_{\mathbb{B}(T_n)} \leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}_{Y_2}(\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{F_0,\Sigma_0})\|^2_{\mathbb{B}(T_n)} \leq M^2 \epsilon_n^2, \|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\left(-C' n \epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

Since $T_n = (c_0 \log n)^{1/2}$ for some c_0 satisfying $c_0 R = \gamma < 1/2$, an application of Lemma 2 implies that

$$\sup_{\|t\|_\infty \leq T_n} \frac{|\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)|}{|\varphi_{Y_2}(t)|} \leq D n^{-1/2+\gamma} \sqrt{\log \log n}$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, it follows that

$$|\widehat{\varphi}_{Y_2}(t)| \geq |\varphi_{Y_2}(t)| - |\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)| \geq \frac{1}{2} |\varphi_{Y_2}(t)| \geq \frac{1}{2} n^{-\gamma}.$$

uniformly over the set $\{\|t\|_\infty \leq T_n\}$, with $\mathbb{P}$ probability approaching 1. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{F_0,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \leq M^2 n^{2\gamma} \epsilon_n^2, \|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\left(-C' n \epsilon_n^2\right).$$

holds with $\mathbb{P}$ probability approaching 1.

In dimension $d = 1$, the fundemental theorem of calculus, Cauchy-Schwarz and the initial value condition $\nabla \log \varphi_X(0) = \nabla \log \varphi_{P,\Sigma}(0) = 0$ imply that

$$\begin{aligned}|\log \varphi_{F_0,\Sigma_0}(t) - \log \varphi_{P,\Sigma}(t)| &= \left|\int_0^t [\nabla \log \varphi_{F_0,\Sigma_0}(s) - \nabla \log \varphi_{P,\Sigma}(s)]ds\right| \\ &\leq \sqrt{T_n}\|\nabla \log \varphi_{F_0,\Sigma_0} - \nabla \log \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}\end{aligned}$$

holds for every $t \in \mathbb{B}(T_n)$. Furthermore, for every fixed $t \in \mathbb{R}^d$, the mean value theorem implies that

$$|\varphi_{F_0,\Sigma_0}(t) - \varphi_{P,\Sigma}(t)| \leq \sup_{s_t \in [0,1]} \left| e^{s_t \log \varphi_{F_0,\Sigma_0}(t) + (1-s_t)\log \varphi_{P,\Sigma}(t)} \right| |\log \varphi_{F_0,\Sigma_0}(t) - \log \varphi_{P,\Sigma}(t)| \,.$$

Since $|\varphi_{F_0,\Sigma_0}| \leq 1$ and $|\varphi_{P,\Sigma}| \leq 1$ (as they are characteristic function of random variables), the preceding bound reduces to

$$|\varphi_{F_0,\Sigma_0}(t) - \varphi_{P,\Sigma}(t)| \leq |\log \varphi_{F_0,\Sigma_0}(t) - \log \varphi_{P,\Sigma}(t)| \,.$$

From combining the preceding bounds and noting that the Lebesgue measure of $\mathbb{B}(T_n)$ is of order T_n, it follows that there exists a universal constant $D > 0$ such that

$$\|\varphi_{F_0,\Sigma_0} - \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)} \leq DT_n \|\nabla \log \varphi_{F_0,\Sigma_0} - \nabla \log \varphi_{P,\Sigma}\|_{\mathbb{B}(T_n)}.$$

It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\begin{aligned}&\nu_{\alpha,G}\left(\|\varphi_{P,\Sigma} - \varphi_{F_0,\Sigma_0}\|_{\mathbb{B}(T_n)} \leq MT_n n^\gamma \epsilon_n, \|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right) \\ &\geq 1 - \exp\bigl(-C' n\epsilon_n^2\bigr).\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1.

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c > 0$ for which $\lambda_1(\Sigma) \geq c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2 (n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp \log n$. It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_{F_0,\Sigma_0} - \varphi_{P,\Sigma})\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 &\leq 2\|\varphi_{F_0,\Sigma_0}\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 + 2\|\varphi_{P,\Sigma}\mathbb{1}\{\|t\|_\infty > T_n\}\|_{L^2}^2 \\ &\leq 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma t} dt \\ &\leq 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2\int_{\|t\|_\infty > T_n} e^{-c\|t\|^2 \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} dt \\ &\leq D\left[e^{-CT_n^2} T_n^{d-2} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa} e^{-CT_n^2 \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} T_n^{d-2} \right].\end{aligned}$$

From substituting $T_n^2 (n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp \log n$, the preceding bound reduces to Dn^{-2K} for some constant $K \in (0, 1/2]$ (that depends on, among other factors, the smallest eigenvalue of Σ_0).

With constant K as specified above, it follows that for any $C' > 0$, there exists $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\varphi_{F_0,\Sigma_0} - \varphi_{P,\Sigma}\|_{L^2} \leq M T_n n^{\gamma}\epsilon_n + n^{-K} \,|\mathcal{Z}_n, T_n\Big) \geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$T_n n^{\gamma}\epsilon_n \asymp n^{-1/2+\gamma}(\log n)^{\lambda/2+1/2}.$$

□

Proof of Theorem 5. The proof proceeds through several steps which we outline below. We use $D > 0$ as a generic universal constant that may change from line to line. Define

$$\epsilon_n^2 = n^{-1}(\log n)^{\lambda} \quad , \quad \lambda = \begin{cases} \chi^{-1}(d+2) + d/2 & \chi < 2 \\ d+1 & \chi \geq 2. \end{cases}$$

For ease of notation, we supress the dependence of $G = G_n$ on n.

(i) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma}) - \widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big) d\nu_{\alpha,G}(P,\Sigma) \geq \exp\big(-Cn\epsilon_n^2\big) \tag{87}$$

holds with $\mathbb{P}$ probability approaching 1.

The argument to verify this is completely analogous to part (i) of Theorem 3. To be specific, all the estimates in that proof hold uniformly over $\|t\|_\infty \leq T_n$ and so they also hold over the set $\{tz : \|z\|_\infty \leq T_n \ , \ t \in [0,1]\}$. Furthermore, as $\mathcal{H}^{d-1}(\partial\mathbb{B}(T_n)) \leq DT_n^{d-1}$, Lemma 2 and 9 imply that

$$\int_{\partial\mathbb{B}(T_n)} |\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)|^2 \, d\mathcal{H}^{d-1}(t) \leq D\frac{T_n^{d-1}\log(T_n)}{n} \ ,$$
$$\int_{\partial\mathbb{B}(T_n)} \|\widehat{\varphi}_{Y_1,Y_2}(t) - \varphi_{Y_1,Y_2}(t)\|^2 d\mathcal{H}^{d-1}(t) \leq D\frac{T_n^{d-1}\log(T_n)}{n}$$

holds with $\mathbb{P}$ probability approaching 1. The only other significant change is that Equation (85) is replaced with the following argument. The law of G is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2 and the definition of σ_n^2, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\Sigma:\|\Sigma-\Sigma_0\|\leq\epsilon_n^2} dG(\Sigma) = \int_{\Sigma:\|\Sigma-\sigma_n^2\Sigma_0\|\leq\sigma_n^2\epsilon_n^2} dL(\Sigma) \geq C\exp\big(-C'\sigma_n^{-2\kappa}\big) \geq C\exp\big(-C''n\epsilon_n^2\big).$$

(*ii*) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log\left(\varphi_{P,\Sigma}\right)-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}>2E\epsilon_n^2} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\right)d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \exp\left(-nE\epsilon_n^2\right).$$

The law of G is given by Σ/σ_n^2 where $\Sigma \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) = \int_{\Sigma:\|\Sigma^{-1}\|>E(n\epsilon_n^2)^{1/\kappa}} dL(\Sigma) \leq \exp\left(-E'n\epsilon_n^2\right).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \geq 0$, the preceding bound implies

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\right)d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma)$$
$$\leq \exp\left(-E'n\epsilon_n^2\right).$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$ sufficiently large such that

$$\int_{\substack{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log\left(\varphi_{P,\Sigma}\right)-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\leq E\epsilon_n^2,\\ \|\Sigma^{-1}\|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}}} \exp\left(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\right)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq 1-\exp\left(-E'n\epsilon_n^2\right).$$

(*iii*) We prove the main statement of the theorem. From the bounds derived in steps (*i*) and (*ii*), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\leq M^2\epsilon_n^2,\, \|\Sigma^{-1}\|\leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n, T_n\right)$$
$$\geq 1-\exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

By an application of Lemma 9, we have that

$$\|\widehat{\varphi}_{Y_1,Y_2}-\varphi_{Y_1,Y_2}\|_{\partial,\,\mathbb{B}(T)}\leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

Since $\varphi_{Y_1,Y_2} = \varphi_{Y_2}\nabla\log\varphi_{F_0,\Sigma_0}$ and $\sup_{\|t\|_\infty\le T_n}\|\nabla\log\varphi_{F_0,\Sigma_0}\|\le DT_n$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}_{Y_2}\nabla\log\varphi_{F_0,\Sigma_0} - \varphi_{Y_1,Y_2}\|_{\partial,\,\mathbb{B}(T)}\le D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C'>0$, there exists a $M>0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}_{Y_2}(\nabla\log\varphi_{P,\Sigma}-\nabla\log\varphi_{F_0,\Sigma_0})\|^2_{\partial,\,\mathbb{B}(T)}\le M^2\epsilon_n^2,\,\|\Sigma^{-1}\|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n,T_n\right)$$
$$\ge 1-\exp\bigl(-C'n\epsilon_n^2\bigr)$$

holds with $\mathbb{P}$ probability approaching 1.

Since $T_n=(c_0\log n)^{1/2}$ for some c_0 satisfying $c_0R=\gamma<1/2$, an application of Lemma 2 implies that

$$\sup_{\|t\|_\infty\le T_n}\frac{|\widehat{\varphi}_{Y_2}(t)-\varphi_{Y_2}(t)|}{|\varphi_{Y_2}(t)|}\le Dn^{-1/2+\gamma}\sqrt{\log\log n}$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, it follows that

$$|\widehat{\varphi}_{Y_2}(t)|\ge|\varphi_{Y_2}(t)|-|\widehat{\varphi}_{Y_2}(t)-\varphi_{Y_2}(t)|\ge\frac{1}{2}\,|\varphi_{Y_2}(t)|\ge\frac{1}{2}n^{-\gamma}.$$

uniformly over the set $\{\|t\|_\infty\le T_n\}$, with $\mathbb{P}$ probability approaching 1. It follows that for any $C'>0$, there exists a $M>0$ such that

$$\nu_{\alpha,G}\left(\|\nabla\log\varphi_{P,\Sigma}-\nabla\log\varphi_{F_0,\Sigma_0}\|^2_{\partial,\,\mathbb{B}(T)}\le M^2n^{2\gamma}\epsilon_n^2,\,\|\Sigma^{-1}\|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n,T_n\right)$$
$$\ge 1-\exp\bigl(-C'n\epsilon_n^2\bigr).$$

holds with $\mathbb{P}$ probability approaching 1.

As $\varphi_{P,\Sigma}$ and φ_{F_0,Σ_0} are characteristic functions, they satisfy the initial value condition $\log\varphi_{P,\Sigma}(0)=\log\varphi_{F_0,\Sigma_0}(0)=0$. In particular, every $z\in\partial\mathbb{B}(T_n)$ can be expressed as

$$\log\varphi_{P,\Sigma}(z)-\log\varphi_{F_0,\Sigma_0}(z)=\int_0^1\langle\nabla\log\varphi_{P,\Sigma}(tz)-\nabla\log\varphi_{F_0,\Sigma_0}(tz),z\rangle dt.$$

By Cauchy-Schwarz and Jensen's inequality, this implies that

$$|\log\varphi_{P,\Sigma}(z)-\log\varphi_{F_0,\Sigma_0}(z)|^2\le\|z\|^2\int_0^1\|\nabla\log\varphi_{P,\Sigma}(tz)-\nabla\log\varphi_{F_0,\Sigma_0}(tz)\|^2dt.$$

In particular since $\|z\| \leq DT_n$ for every $z \in \partial\mathbb{B}(T_n)$, we obtain the bound

$$\begin{aligned}&\int_{\partial\mathbb{B}(T_n)} |\log \varphi_{P,\Sigma}(z) - \log \varphi_{F_0,\Sigma_0}(z)|^2 \, d\mathcal{H}^{d-1}(z) \\ &\leq T_n^2 \int_{\partial\mathbb{B}(T_n)} \int_0^1 \|\nabla \log \varphi_{P,\Sigma}(tz) - \nabla \log \varphi_{F_0,\Sigma_0}(tz)\|^2 dt d\mathcal{H}^{d-1}(z).\end{aligned}$$

Suppose (P, Σ) is such that $\|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{F_0,\Sigma_0}\|^2_{\partial,\, \mathbb{B}(T)} \leq M^2 n^{2\gamma} \epsilon_n^2$. By definition of the metric, this means that

$$\begin{aligned}&\int_{\mathbb{B}(T_n)} \|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{F_0,\Sigma_0}\|^2 dt \leq M^2 n^{2\gamma} \epsilon_n^2 \;, \\ &\int_{\partial\mathbb{B}(T_n)} \int_0^1 \|\nabla \log \varphi_{P,\Sigma}(tz) - \nabla \log \varphi_{F_0,\Sigma_0}(tz)\|^2 dt d\mathcal{H}^{d-1}(z) \leq M^2 n^{2\gamma} \epsilon_n^2.\end{aligned}$$

From the preceding bounds and the Poincaré inequality (Lemma 13), it follows that there exists a universal constant $D > 0$ such that

$$\|\log \varphi_{P,\Sigma} - \log \varphi_{F_0,\Sigma_0}\|^2_{\mathbb{B}(T_n)} \leq DM^2 T_n^3 n^{2\gamma} \epsilon_n^2.$$

Since $|\varphi_X(t) - \varphi_{P,\Sigma}(t)| \leq |\log \varphi_X(t) - \log \varphi_{P,\Sigma}(t)|$ for every $t \in \mathbb{R}^d$, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\begin{aligned}&\nu_{\alpha,G}\Big(\|\varphi_{P,\Sigma} - \varphi_{F_0,\Sigma_0}\|_{\mathbb{B}(T_n)} \leq M T_n^{1.5} n^{\gamma} \epsilon_n, \|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\Big|\, \mathcal{Z}_n, T_n \Big) \\ &\geq 1 - \exp\big(-C' n \epsilon_n^2\big).\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1.

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\| \leq M^2 \sigma_n^2 (n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c > 0$ for which $\lambda_1(\Sigma) \geq c(n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2 (n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2} \asymp \log n$. It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_{F_0,\Sigma_0} - \varphi_{P,\Sigma}) \mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} &\leq 2\|\varphi_{F_0,\Sigma_0} \mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} + 2\|\varphi_{P,\Sigma} \mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} \\ &\leq 2 \int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2 \int_{\|t\|_\infty > T_n} e^{-t'\Sigma t} dt \\ &\leq 2 \int_{\|t\|_\infty > T_n} e^{-t'\Sigma_0 t} dt + 2 \int_{\|t\|_\infty > T_n} e^{-c\|t\|^2 \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa}} dt \\ &\leq D\Big[e^{-CT_n^2} T_n^{d-2} + \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa} e^{-CT_n^2 \sigma_n^{-2} (n\epsilon_n^2)^{-1/\kappa}} T_n^{d-2} \Big].\end{aligned}$$

From substituting $T_n^2 (n\epsilon_n^2)^{-1/\kappa} \sigma_n^{-2} \asymp \log n$, the preceding bound reduces to Dn^{-2K} for some constant $K \in (0, 1/2]$ (that depends on, among other factors, the smallest eigenvalue of Σ_0).

With constant K as specified above, it follows that for any $C' > 0$, there exists $M > 0$

such that

$$\nu_{\alpha,G}\Big(\|\varphi_{F_0,\Sigma_0}-\varphi_{P,\Sigma}\|_{L^2}\leq MT_n^{1.5}n^{\gamma}\epsilon_n+n^{-K}\,|\mathcal{Z}_n,T_n\Big)\geq 1-\exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$T_n^{1.5}n^{\gamma}\epsilon_n\asymp n^{-1/2+\gamma}(\log n)^{\lambda/2+3/4}.$$

□

Proof of Theorem 6. The proof proceeds through several steps which we outline below. We use $D>0$ as a generic universal constant that may change from line to line. For ease of notation, we suppress the dependence of $G=G_n$ on n.

Let $\beta=\max\{0,d-2\}$ and define

$$\alpha_n^2=(\epsilon_n^2T_n^{-\beta})^{\frac{1}{\zeta+s}}\ ,\ \epsilon_n^2=\frac{(\log n)^{\lambda+d/2}}{n^{\frac{2(s+\zeta)-\beta}{2(s+\zeta)+d-\beta}}}\ ,\ \lambda=\begin{cases}\chi^{-1}(d+2) & \chi<2\\ d/\chi+1 & \chi\geq 2.\end{cases}$$

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C>0$ such that

$$\int\exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma)\geq\exp\big(-Cn\epsilon_n^2\big)\qquad(88)$$

holds with $\mathbb{P}$ probability approaching 1.

As $\mathcal{H}^{d-1}(\partial\mathbb{B}(T_n))\leq DT_n^{d-1}$, Lemma 9 implies that

$$\int_{\mathbb{B}(T_n)}\|\widehat{\varphi}_{Y_1,Y_2}(t)-\varphi_{Y_1,Y_2}(t)\|^2dt\leq D\frac{T_n^d\log(T_n)}{n}\ ,$$
$$\int_{\partial\mathbb{B}(T_n)}\|\widehat{\varphi}_{Y_1,Y_2}(t)-\varphi_{Y_1,Y_2}(t)\|^2d\mathcal{H}^{d-1}(t)\leq D\frac{T_n^{d-1}\log(T_n)}{n}.$$

holds with $\mathbb{P}$ probability approaching 1. On this set, it follows that

$$\int\exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq\exp\big(-nD\epsilon_n^2\big)\int\exp\Big(-\frac{nD}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\varphi_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma).$$

Observe that $\varphi_{Y_1,Y_2}=\varphi_{Y_2}\nabla\log\varphi_X$. By Condition 4.6, there exists universal constants $\chi,C,M<\infty$ and a mixing distribution S_{α_n} supported on the cube $I_n=[-C(\log\epsilon_n^{-1})^{1/\chi},C(\log\epsilon_n^{-1})^{1/\chi}]^d$ that satisfies

$$\|\varphi_{Y_2}(\nabla\log\varphi_X-\nabla\log\varphi_{S_{\alpha_n},\alpha_n^2I})\|^2_{\partial,\,\mathbb{B}(T_n)}\leq DT_n^{\beta}\alpha_n^{2(\zeta+s)}\leq D\epsilon_n^2.$$

It follows that

$$\int \exp\bigg(-\frac{nD}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\varphi_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\bigg)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq \exp(-nD\epsilon_n^2)\int \exp\bigg(-\frac{nD}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\varphi_{Y_2}\nabla\log\big(\varphi_{S_{\alpha_n},\alpha_n^2 I}\big)\|^2_{\partial,\,\mathbb{B}(T)}\bigg)d\nu_{\alpha,G}(P,\Sigma).$$

By Condition 4.6 and the definition of (α_n, T_n), there exists a $C_1 > 0$ such that

$$\inf_{\|t\|_\infty\leq T_n}\big|\varphi_{S_{\alpha_n}}(t)\big| \geq n^{-C_1}.$$

Next, we show that S_{α_n} can be suitably approximated by a discrete measure. Fix any $\gamma > 1$ such that $\epsilon_n^{2\gamma}T_n^d \leq D\epsilon_n^2$. By Lemma 5, there exists a discrete measure $F' = \sum_{i=1}^N p_i\delta_{\mu_i}$ with at most $N = D(\log\epsilon_n^{-1})^{d/\chi}T_n^d$ support points on I_n such that

$$\sup_{\|t\|_\infty\leq T_n}\big|\varphi_{S_{\alpha_n}}(t)-\varphi_{F'}(t)\big| \leq n^{-C_1}\epsilon_n^\gamma$$
$$\sup_{\|t\|_\infty\leq T_n}\|\nabla\varphi_{S_{\alpha_n}}(t)-\nabla\varphi_{F'}(t)\|\leq n^{-C_1}\epsilon_n^\gamma.$$

From the final claim of Lemma 5, we can also assume without loss of generality that the support points satisfy $\inf_{k\neq j}\|\mu_k-\mu_j\|\geq \epsilon_n^{L_2}$ for some $L_2 > 0$. By Condition 4.6, $\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_{S_{\alpha_n}}(t)\|\leq D$. From the preceding bounds, it follows that

$$\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_{S_{\alpha_n}}(t)\|\frac{\big|\varphi_{S_{\alpha_n}}(t)-\varphi_{F'}(t)\big|}{\big|\varphi_{S_{\alpha_n}}(t)\big|}\leq D\epsilon_n^\gamma$$
$$\sup_{\|t\|_\infty\leq T_n}\frac{\|\nabla\varphi_{S_{\alpha_n}}(t)-\nabla\varphi_{F'}(t)\|}{\big|\varphi_{S_{\alpha_n}}(t)\big|}\leq D\epsilon_n^\gamma.$$

Since $\epsilon_n \downarrow 0$, observe that the preceding bound also implies

$$|\varphi_{F'}(t)| \geq \big|\varphi_{S_{\alpha_n}}(t)\big| - \big|\varphi_{S_{\alpha_n}}(t)-\varphi_{F'}(t)\big| \geq \big|\varphi_{S_{\alpha_n}}(t)\big| - D\epsilon_n\big|\varphi_{S_{\alpha_n}}(t)\big| \geq \frac{1}{2}\big|\varphi_{S_{\alpha_n}}(t)\big|$$

for all sufficiently large n and $\|t\|\leq T_n$. It follows that

$$\sup_{\|t\|_\infty\leq T_n}\|\nabla\log\varphi_{S_{\alpha_n}}(t)-\nabla\log\varphi_{F'}(t)\|$$
$$=\sup_{\|t\|_\infty\leq T_n}\left\|\nabla\log\varphi_{S_{\alpha_n}}(t)\frac{\varphi_{F'}(t)-\varphi_{S_{\alpha_n}}(t)}{\varphi_{F'}(t)}+\frac{\nabla\varphi_{S_{\alpha_n}}(t)-\nabla\varphi_{F'}(t)}{\varphi_{F'}(t)}\right\|$$
$$\leq \sup_{\|t\|_\infty\leq T_n}\left(\left|\frac{\varphi_{F'}(t)-\varphi_{S_{\alpha_n}}(t)}{\varphi_{F'}(t)}\right|\|\nabla\log\varphi_{S_{\alpha_n}}(t)\|+\left\|\frac{\nabla\varphi_{S_{\alpha_n}}(t)-\nabla\varphi_{F'}(t)}{\varphi_{F'}(t)}\right\|\right)$$
$$\leq 2\sup_{\|t\|_\infty\leq T_n}\left(\left|\frac{\varphi_{F'}(t)-\varphi_{S_{\alpha_n}}(t)}{\varphi_{S_{\alpha_n}}(t)}\right|\|\nabla\log\varphi_{S_{\alpha_n}}(t)\|+\left\|\frac{\nabla\varphi_{S_{\alpha_n}}(t)-\nabla\varphi_{F'}(t)}{\varphi_{S_{\alpha_n}}(t)}\right\|\right)$$
$$\leq D\epsilon_n^\gamma.$$

Fix any $L_3 > L_2$ sufficiently large such that

$$n^{C_1}T_n\epsilon_n^{L_3} \leq \epsilon_n^{\gamma}\ ,\ \ n^{C_1}\sqrt{n}\epsilon_n\sqrt{\log n}\epsilon_n^{L_3/2} \leq \epsilon_n^{\gamma}\ ,\ n^{C_1}T_n(\log \epsilon_n^{-1})^{1/\chi}\epsilon_n^{L_3} \leq \epsilon_n^{\gamma}.$$

Define $V_i = \{t \in I_n : \|t-\mu_i\| \leq \epsilon_n^{L_3}\}$ for $i = 1, \ldots, N$ and set $V_0 = \mathbb{R}^d \setminus \bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \ldots, V_N\}$ is a disjoint partition of $\mathbb{R}^d$. By Lemma 3, for any distribution P that satisfies $\int_{\mathbb{R}^d}\|x\|^2 dP(x) \leq n\epsilon_n^2 \log n$ and $\sum_{j=1}^N |P(V_j) - p_j| \leq \epsilon_n^{L_3}$, we have that

$$\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_{F'}(t)\| \frac{|\varphi_P(t) - \varphi_{F'}(t)|}{|\varphi_{F'}(t)|} \leq D\epsilon_n^{\gamma}\ ,\ \sup_{\|t\|_\infty \leq T_n} \frac{\|\nabla\varphi_P(t) - \nabla\varphi_{F'}(t)\|}{|\varphi_{F'}(t)|} \leq D\epsilon_n^{\gamma}.$$

For all such P, an analogous argument to the bound for S_{α_n} and F' implies that

$$\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_P(t) - \nabla \log \varphi_{F'}(t)\| \leq D\epsilon_n^{\gamma}.$$

From combining all the preceding bounds, observe that all such P also satisfy

$$\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_P(t) - \nabla \log \varphi_{S_{\alpha_n}}(t)\| \leq D\epsilon_n^{\gamma}$$
$$\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_P(t)\| \leq D\epsilon_n^{\gamma} + \sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_{S_{\alpha_n}}(t)\| \leq D.$$

Given any positive definite Σ with $\|\Sigma\| \leq D\alpha_n^2$ the preceding bound also implies that

$$\begin{aligned}\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_{P,\Sigma}(t)\| &\leq \sup_{\|t\|_\infty \leq T_n} \left[\|\nabla \log \varphi_P(t)\| + \|\nabla \log \varphi_\Sigma(t)\|\right] \\ &\leq D(1 + T_n\alpha_n^2) \\ &\leq D\ ,\end{aligned}$$

where the last inequality follows from $T_n\alpha_n^2 \lessapprox 1$. By Lemma 2, we have that

$$\int_{\mathbb{B}(T_n)} \|\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)\|^2 dt \leq D\frac{T_n^d \log(T_n)}{n}\ ,$$
$$\int_{\partial\mathbb{B}(T_n)} \|\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)\|^2 d\mathcal{H}^{d-1}(t) \leq D\frac{T_n^{d-1} \log(T_n)}{n}$$

holds with $\mathbb{P}$ probability approaching 1. On this set, when (P, Σ) satisfy the preceding requirements, we have that

$$\begin{aligned}\|\widehat{\varphi}_{Y_2}\nabla \log \varphi_{P,\Sigma} - \varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} &\leq D\left[\|\varphi_{Y_2}\nabla \log \varphi_{P,\Sigma} - \varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} + \frac{T_n^d \log(T_n)}{n}\right] \\ &\leq D\left[\|\varphi_{Y_2}\nabla \log \varphi_{P,\Sigma} - \varphi_{Y_1,Y_2}\|^2_{\mathbb{B}(T_n)} + \epsilon_n^2\right].\end{aligned}$$

Define the sets

$$\Omega_n = \left\{ \Sigma \in \mathbf{S}_+^d : \lambda_j(\Sigma) \in \left[\frac{\alpha_n^2}{1+\epsilon_n^\gamma T_n^{-1}}, \alpha_n^2 \right] \ \forall\, j = 1, \ldots, d. \right\}$$

$$\mathcal{G}_n = \left\{ (P, \Sigma) : \Sigma \in \Omega_n, \int_{\mathbb{R}^d} \|x\|^2 dP(x) \leq n\epsilon_n^2 \log(n) \,, \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\}.$$

From combining the all the preceding bounds, it follows that

$$\begin{aligned}
&\int \exp\Big(- nD\|\widehat{\varphi}_{Y_2}\nabla \log(\varphi_{P,\Sigma}) - \varphi_{Y_2}\nabla \log\big(\varphi_{S_{\alpha_n}, \alpha_n^2 I}\big)\|^2_{\partial,\, \mathbb{B}(T)}\Big) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \int_{\mathcal{G}_n} \exp\Big(- nD\|\widehat{\varphi}_{Y_2}\nabla \log(\varphi_{P,\Sigma}) - \varphi_{Y_2}\nabla \log\big(\varphi_{S_{\alpha_n}, \alpha_n^2 I}\big)\|^2_{\partial,\, \mathbb{B}(T)}\Big) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \exp\big(- nD\epsilon_n^2\big) \int_{\mathcal{G}_n} \exp\Big(- nD\|\varphi_{Y_2}[\nabla \log(\varphi_{P,\Sigma}) - \nabla \log\big(\varphi_{S_{\alpha_n}, \alpha_n^2 I}\big)]\|^2_{\partial,\, \mathbb{B}(T)}\Big) d\nu_{\alpha,G}(P, \Sigma).
\end{aligned}$$

For every $(P, \Sigma) \in \mathcal{G}_n$, the preceding bounds imply that

$$\begin{aligned}
&\sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{S_{\alpha_n}, \alpha_n^2 I}\| \\
&= \sup_{\|t\|_\infty \leq T_n} \|(\nabla \log \varphi_P - \nabla \log \varphi_{S_{\alpha_n}}) + (\nabla \log \varphi_\Sigma - \nabla \log \varphi_{\alpha_n^2 I})\| \\
&\leq \sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_P - \nabla \log \varphi_{S_{\alpha_n}}\| + \sup_{\|t\|_\infty \leq T_n} \|\nabla \log \varphi_\Sigma - \nabla \log \varphi_{\alpha_n^2 I}\| \\
&\leq D\epsilon_n^\gamma + D\|\Sigma - \alpha_n^2 I\| T_n \\
&\leq D\epsilon_n^\gamma.
\end{aligned}$$

Since $\|\varphi_{Y_2}\|_{L^2} < \infty$ and $\gamma > 1$ is such that $T_n^d \epsilon_n^{2\gamma} \leq D\epsilon_n^2$, it follows (from the change of variables and coarea argument in [26]) that

$$\begin{aligned}
&\int_{\mathcal{G}_n} \exp\Big(- nD\|\varphi_{Y_2}(\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_{S_{\alpha_n}, \alpha_n^2 I})\|^2_{\partial,\, \mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P, \Sigma) \\
&\geq \exp\big(- nD\epsilon_n^2\big) \int_{\mathcal{G}_n} d\nu_{\alpha,G}(P, \Sigma).
\end{aligned}$$

Define the sets

$$\begin{aligned}
\mathcal{G}_{n,1} &= \left\{ (P, \Sigma) : \Sigma \in \Omega_n \,, \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\}, \\
\mathcal{G}_{n,2} &= \left\{ (P, \Sigma) : \Sigma \in \Omega_n, \int_{\mathbb{R}^d} \|x\|^2 dP(x) > n\epsilon_n^2 \log(n) \,, \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\}.
\end{aligned}$$

Observe that $\mathcal{G}_n = \mathcal{G}_{n,1} \setminus \mathcal{G}_{n,2}$. It follows that

$$\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P, \Sigma) = \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P, \Sigma) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P, \Sigma).$$

For the second term, observe that

$$\begin{aligned}\int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\Sigma) &\le \int_{P:\int_{\mathbb{R}^d}\|x\|^2 dP(x)>n\epsilon_n^2\log(n)} d\nu_{\alpha,G}(P,\Sigma)\\ &\le \int_{P:\int_{\mathbb{R}^d}\|x\|^2 dP(x)>n\epsilon_n^2\log(n)} d\mathrm{DP}_\alpha(P)\\ &\le \exp\big(-Dn\epsilon_n^2\log n\big)\ ,\end{aligned}$$

where the second inequality is due to $\nu_{\alpha,G}$ being a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and the third inequality follows from an application of Lemma 12.

For the first term, we have that

$$\int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\Sigma) = \int_{\Sigma:\Sigma\in\Omega_n}\int_{P:\sum_{j=1}^N |P(V_j)-p_j|\le\epsilon_n^{L_3}} d\mathrm{DP}_\alpha(P)dG(\Sigma).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^N \alpha(V_j) \ge C\epsilon_n^{L_3 d}\exp\big(-C'(\log\epsilon_n^{-1})^{2/\chi}\big)$ for universal constants $C, C' > 0$. By definition of DP_α, $(P(V_1),\dots,P(V_N)) \sim \mathrm{Dir}(N, \alpha(V_1),\dots,\alpha(V_N))$. As $N = D\{\log\big(\epsilon_n^{-1}\big)\}^{d/\chi}T_n^d$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) implies

$$\begin{aligned}\int_{P:\sum_{j=1}^N |P(V_j)-p_j|\le\epsilon_n^{L_3}} d\mathrm{DP}_\alpha(P) \ge C\exp\big(-C'T_n^d(\log\epsilon_n^{-1})^{d/\chi+\max\{2/\chi,1\}}\big) &= C\exp\big(-C'T_n^d(\log\epsilon_n^{-1})^{\lambda}\big)\\ &\ge C\exp\big(-C''n\epsilon_n^2\big).\end{aligned}$$

It remains to bound the outer integral. The law of $G = G_n$ is given by Ω/σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2 and the definition of $(\alpha_n^2, \sigma_n^2, \epsilon_n^2)$, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\Sigma\in\Omega_n} dG(\Sigma) = \int_{\Sigma\in\sigma_n^2\Omega_n} dL(\Sigma) \ge C\exp\big(-C'\sigma_n^{-2\kappa}\alpha_n^{-2\kappa}\big) \ge C\exp\big(-C''n\epsilon_n^2\big).$$

It follows that

$$\begin{aligned}\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\Sigma) &= \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\Sigma) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\Sigma)\\ &\ge C\exp\big(-C''n\epsilon_n^2\big) - \exp\big(-Dn\epsilon_n^2\log n\big)\\ &\ge \exp\big(-Dn\epsilon_n^2\big).\end{aligned}$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(*ii*) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}>2E\epsilon_n^2} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \exp\big(-nE\epsilon_n^2\big).$$

The law of G is given by Σ/σ_n^2 where $\Sigma \sim L$ and L is a probability measure on $\mathbb{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma) = \int_{\Sigma:\|\Sigma^{-1}\|>E(n\epsilon_n^2)^{1/\kappa}} dL(\Sigma) \leq \exp\big(-E'n\epsilon_n^2\big).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and $\|\widehat{\varphi}_Y - \widehat{\varphi}_\epsilon\varphi_{P,\Sigma}\|^2_{\mathbb{B}(T_n)} \geq 0$, the preceding bound implies

$$\int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma)$$
$$\leq \int_{\Sigma:\|\Sigma^{-1}\|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\Sigma)$$
$$\leq \exp\big(-E'n\epsilon_n^2\big).$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$ sufficiently large such that

$$\int_{\substack{(P,\Sigma):\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\leq E\epsilon_n^2,\\ \|\Sigma^{-1}\|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\Big)d\nu_{\alpha,G}(P,\Sigma)$$
$$\geq 1-\exp\big(-E'n\epsilon_n^2\big).$$

(*iii*) We prove the main statement of the theorem. From the bounds derived in steps (*i*) and (*ii*), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\widehat{\varphi}_{Y_2}\nabla\log(\varphi_{P,\Sigma})-\widehat{\varphi}_{Y_1,Y_2}\|^2_{\partial,\,\mathbb{B}(T)}\leq M^2\epsilon_n^2, \|\Sigma^{-1}\|\leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\Big|\,\mathcal{Z}_n, T_n\Big)$$
$$\geq 1-\exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

By an application of Lemma 9, we have that

$$\|\widehat{\varphi}_{Y_1,Y_2}-\varphi_{Y_1,Y_2}\|_{\partial,\,\mathbb{B}(T)}\leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

Since $\varphi_{Y_1,Y_2} = \varphi_{Y_2}\nabla \log \varphi_X$ and $\sup_{\|t\|_\infty \le T_n} \|\nabla \log \varphi_X\| \le D$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}_{Y_2}\nabla \log \varphi_X - \varphi_{Y_1,Y_2}\|_{\partial,\, \mathbb{B}(T)} \le D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Bigg(\|\widehat{\varphi}_{Y_2}(\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_X)\|^2_{\partial,\, \mathbb{B}(T)} \le M^2\epsilon_n^2, \|\Sigma^{-1}\| \le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\Bigg|\, \mathcal{Z}_n, T_n\Bigg)$$
$$\ge 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

An application of Lemma 2 implies that

$$\sup_{\|t\|_\infty \le T_n} \frac{|\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)|}{|\varphi_{Y_2}(t)|} \le DT_n^\zeta \frac{\sqrt{\log n}}{\sqrt{n}}$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, we also have

$$|\widehat{\varphi}_{Y_2}(t)| \ge |\varphi_{Y_2}(t)| - |\widehat{\varphi}_{Y_2}(t) - \varphi_{Y_2}(t)| \ge \frac{1}{2}\,|\varphi_{Y_2}(t)|$$

uniformly over the set $\{\|t\|_\infty \le T_n\}$. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Bigg(\|\nabla \log \varphi_{P,\Sigma} - \nabla \log \varphi_X\|^2_{\partial,\, \mathbb{B}(T)} \le M^2T_n^{2\zeta}\epsilon_n^2, \|\Sigma^{-1}\| \le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\Bigg|\, \mathcal{Z}_n, T_n\Bigg)$$
$$\ge 1 - \exp\big(-C'n\epsilon_n^2\big).$$

holds with $\mathbb{P}$ probability approaching 1.

As $\varphi_{P,\Sigma}$ and φ_X are characteristic functions, they satisfy the initial value condition $\log \varphi_{P,\Sigma}(0) = \log \varphi_X(0) = 0$. In particular, every $z \in \partial\mathbb{B}(T_n)$ can be expressed as

$$\log \varphi_{P,\Sigma}(z) - \log \varphi_X(z) = \int_0^1 \langle \nabla \log \varphi_{P,\Sigma}(tz) - \nabla \log \varphi_X(tz), z\rangle dt.$$

By Cauchy-Schwarz and Jensen's inequality, this implies that

$$|\log \varphi_{P,\Sigma}(z) - \log \varphi_X(z)|^2 \le \|z\|^2 \int_0^1 \|\nabla \log \varphi_{P,\Sigma}(tz) - \nabla \log \varphi_X(tz)\|^2 dt.$$

In particular since $\|z\| \le DT_n$ for every $z \in \partial\mathbb{B}(T_n)$, we obtain the bound

$$\begin{aligned}&\int_{\partial\mathbb{B}(T_n)} |\log\varphi_{P,\Sigma}(z) - \log\varphi_X(z)|^2 \, d\mathcal{H}^{d-1}(z) \\ &\le T_n^2 \int_{\partial\mathbb{B}(T_n)} \int_0^1 \|\nabla \log\varphi_{P,\Sigma}(tz) - \nabla\log\varphi_X(tz)\|^2 dt d\mathcal{H}^{d-1}(z).\end{aligned}$$

Suppose (P,Σ) is such that $\|\nabla\log\varphi_{P,\Sigma} - \nabla\log\varphi_X\|^2_{\partial,\,\mathbb{B}(T)} \le M^2 T_n^{2\zeta}\epsilon_n^2$. By definition of the metric, this means that

$$\begin{aligned}&\int_{\mathbb{B}(T_n)} \|\nabla\log\varphi_{P,\Sigma} - \nabla\log\varphi_X\|^2 dt \le M^2 T_n^{2\zeta}\epsilon_n^2 \;, \\ &\int_{\partial\mathbb{B}(T_n)} \int_0^1 \|\nabla\log\varphi_{P,\Sigma}(tz) - \nabla\log\varphi_X(tz)\|^2 dt d\mathcal{H}^{d-1}(z) \le M^2 T_n^{2\zeta}\epsilon_n^2.\end{aligned}$$

From the preceding bounds and the Poincaré inequality (Lemma 13), it follows that there exists a universal constant $D > 0$ such that

$$\|\log\varphi_{P,\Sigma} - \log\varphi_X\|^2_{\mathbb{B}(T_n)} \le DM^2 T_n^{3+2\zeta}\epsilon_n^2.$$

Since $|\varphi_X(t) - \varphi_{P,\Sigma}(t)| \le |\log\varphi_X(t) - \log\varphi_{P,\Sigma}(t)|$ for every $t \in \mathbb{R}^d$, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\begin{aligned}&\nu_{\alpha,G}\Bigg(\|\varphi_{P,\Sigma} - \varphi_X\|_{\mathbb{B}(T_n)} \le M T_n^{1.5+\zeta}\epsilon_n, \|\Sigma^{-1}\| \le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \;\Bigg|\; \mathcal{Z}_n, T_n \Bigg) \\ &\ge 1 - \exp\big(-C' n\epsilon_n^2\big).\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1.

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(T_n)$. Suppose $\|\Sigma^{-1}\| \le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c > 0$ for which $\lambda_1(\Sigma) \ge c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp \log n \log\log n$.

Since $f_X \in \mathbf{H}^p(M)$, we have that

$$\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2 \, dt \le DT_n^{-2p}.$$

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_X - \varphi_{P,\Sigma})\mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} &\le 2\|\varphi_X \mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} + 2\|\varphi_{P,\Sigma}\mathbb{1}\{\|t\|_\infty > T_n\}\|^2_{L^2} \\ &\le 2\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2 \, dt + 2\int_{\|t\|_\infty > T_n} e^{-t'\Sigma t} dt \\ &\le 2\int_{\|t\|_\infty > T_n} |\varphi_X(t)|^2 \, dt + 2\int_{\|t\|_\infty > T_n} e^{-c\|t\|^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} dt \\ &\le D\Bigg[T_n^{-2p} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa} e^{-CT_n^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}} T_n^{d-2}\Bigg].\end{aligned}$$

Since $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp (\log n)(\log\log n)$, the preceding bound reduces to DT_n^{-2p}. From combining the preceding bounds, it follows that for every $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\bigg(\|\varphi_X - \varphi_{P,\Sigma}\|_{L^2} \leq M(T_n^{1.5+\zeta}\epsilon_n + T_n^{-p})\,\bigg|\,\mathcal{Z}_n, T_n\bigg) \geq 1 - \exp\big(-C'n\epsilon_n^2\big).$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$T_n^{3+2\zeta}\epsilon_n^2 \asymp \frac{(\log n)^{\lambda+d/2+\zeta+3/2}}{n^{\frac{2s-\beta-3}{2(s+\zeta)+d-\beta}}}.$$

□

Proof of Theorem 7. The proof proceeds through several steps which we outline below. We use $D > 0$ as a generic universal constant that may change from line to line. Define

$$\epsilon_n^2 = n^{-1}(\log n)^{\lambda} \quad , \quad \lambda = \begin{cases} \chi^{-1}(d+2) + d/2 & \chi < 2 \\ d+1 & \chi \geq 2. \end{cases}$$

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\bigg(- \frac{n}{2}\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T)}^2\bigg) d\nu_{\alpha,G}(P,\sigma^2) \geq \exp\big(- Cn\epsilon_n^2\big) \quad (89)$$

holds with $\mathbb{P}$ probability approaching 1.

Note that $|t'\mathbf{A}_k| \leq \sqrt{d}\|t\|_\infty\|\mathbf{A}_k\| \leq \sqrt{d}T_n\|\mathbf{A}_k\|$ uniformly over $t \in \mathbb{B}(T)$. Since the preceding constant is finite, it suffices to work under the setting where $|t'\mathbf{A}_k| \leq DT_n$. Fix $\epsilon > 0$ sufficiently small. By Condition 4.8, the mixing distribution F_0 satisfies $F_0(t \in \mathbb{R} : |t| > z) \leq C\exp(-C'z^{\chi})$. Hence, there exists a universal constant $R > 0$ such that the cube $I = [-R(\log\epsilon^{-1})^{1/\chi}, R(\log\epsilon^{-1})^{1/\chi}]$ satisfies $1 - F_0(I) \leq D\epsilon$. Denote the probability measure induced from the restriction of F_0 to I by

$$\overline{F}_0(A) = \frac{F_0(A\cap I)}{F_0(I)} \qquad \forall \text{ Borel } A \subseteq \mathbb{R}.$$

Observe that

$$\sup_{t\in\mathbb{R}}\Big|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\Big| = \sup_{t\in\mathbb{R}}\bigg|\int_{\mathbb{R}^d} e^{\mathbf{i}t'x}d(F - \overline{F}_0)(x)\bigg| \leq \|F_0 - \overline{F}_0\|_{TV} \leq 1 - F_0(I) \leq D\epsilon.$$

For $r = 1, 2$ and all sufficiently large $M > 0$, the tail bound on F_0 implies

$$\begin{aligned}
\mathbb{E}[|X|^r \,\mathbb{1}\{|X| > M\}] &= \int_0^\infty \mathbb{P}(\,|X|^r > M^r, |X|^r > t)dt \\
&\le M^r \mathbb{P}(\,|X| > M) + \int_{M^r}^\infty \mathbb{P}(X^r > t)dt \\
&\le D\left[M^r \exp\bigl(-C' M^\chi\bigr) + \int_{M^r}^\infty \exp\Bigl(-C' t^{\chi/r}\Bigr)dt\right] \\
&\le D[M^r \exp\bigl(-C' M^\chi\bigr) + M^{1-\chi} \exp\bigl(-C' M^\chi\bigr)] \\
&\le D M^r \exp\bigl(-C' M^\chi\bigr).
\end{aligned}$$

For $r = 1, 2$ it follows that

$$\begin{aligned}
\sup_{t\in\mathbb{R}} \left|\partial_t^r \varphi_{F_0}(t) - \partial_t^r \varphi_{\overline{F}_0}(t)\right| &= \sup_{t\in\mathbb{R}} \left|\int_{\mathbb{R}} x^r e^{\mathbf{i}t'x} d(F_0 - \overline{F}_0)(x)\right| \\
&\le \sup_{t\in\mathbb{R}} \left|\int_{x\in I} x^r e^{\mathbf{i}t'x} d(F_0 - \overline{F}_0)\right| + \sup_{t\in\mathbb{R}} \left|\int_{x\notin I} x^r e^{\mathbf{i}t'x} dF_0\right| \\
&\le D\left[(\log \epsilon^{-1})^{r/\chi} \|F_0 - \overline{F}_0\|_{TV} + (\log \epsilon^{-1})^{r/\chi} \epsilon\right] \\
&\le D(\log \epsilon^{-1})^{r/\chi} \epsilon.
\end{aligned}$$

We can write $(\log \varphi_{F_0})'' - (\log \varphi_{\overline{F}_0})''$ as

$$\begin{aligned}
&(\log \varphi_{F_0})'' - (\log \varphi_{\overline{F}_0})'' \\
&= (\log \varphi_{F_0})'' \frac{(\varphi_{\overline{F}_0}^2 - \varphi_{F_0}^2)}{\varphi_{\overline{F}_0}^2} + \frac{\varphi_{F_0}(\varphi_{F_0}'' - \varphi_{\overline{F}_0}'')}{\varphi_{\overline{F}_0}^2} + \frac{\varphi_{\overline{F}_0}''(\varphi_{F_0} - \varphi_{\overline{F}_0})}{\varphi_{\overline{F}_0}^2} + \frac{(\varphi_{\overline{F}_0}')^2 - (\varphi_{F_0}')^2}{\varphi_{\overline{F}_0}^2}.
\end{aligned}$$

Since $\inf_{|t|\le DT_n} |\varphi_{F_0}(t)| \ge C \exp\bigl(-C' T_n^2\bigr)$ for some $C, C' > 0$ and $T_n^2 \lessapprox \log(n)$, there exists a $C_1 > 0$ such that

$$\inf_{|t|\le DT_n} |\varphi_{F_0}(t)| \ge n^{-C_1}.$$

Let $L > 1$ be such that $n^{2C_1}(\log \epsilon_n^{-L})^{2/\chi} \epsilon_n^L \le \epsilon_n$. The choice $\epsilon = \epsilon_n^L$ implies $\overline{F}_0$ has support contained in the cube $I_n = [-E(\log \epsilon_n^{-1})^{1/\chi}, E(\log \epsilon_n^{-1})^{1/\chi}]$ for some universal constant $E > 0$. Since $\epsilon_n \downarrow 0$, we also have that

$$\left|\varphi_{\overline{F}_0}(t)\right| \ge |\varphi_{F_0}(t)| - \left|\varphi_{F_0}(t) - \varphi_{\overline{F}_0}(t)\right| \ge |\varphi_{F_0}(t)| - D\epsilon_n |\varphi_{F_0}(t)| \ge \frac{1}{2} |\varphi_{F_0}(t)| \ge \frac{1}{2} n^{-C_1}$$

for all sufficiently large n and $|t| \le DT_n$. Since $|(\log \varphi_{F_0})''|, \left|\varphi_{F_0}'\right|, \left|\varphi_{F_0}''\right|$ are bounded by a universal constant D and $n^{2C_1}(\log \epsilon_n^{-L})^{2/\chi} \epsilon_n^L \le \epsilon_n$, the preceding expression for $(\log \varphi_{F_0})'' - (\log \varphi_{\overline{F}_0})''$ implies that

$$\sup_{|t|\le DT_n} \left|(\log \varphi_{F_0})''(t) - (\log \varphi_{\overline{F}_0})''(t)\right| \le D n^{2C_1} (\log \epsilon_n^{-L})^{2/\chi} \epsilon_n^L \le D\epsilon_n.$$

Next, we show that $\overline{F}_0$ can be suitably approximated by a discrete measure. Let $\iota =$

$\max\{d, d/\chi + d/2\}$. By Lemma 5, there exists a discrete measure $F' = \sum_{i=1}^N p_i \delta_{\mu_i}$ with at most $N = D(\log \epsilon_n^{-1})^\iota$ support points on I_n such that

$$\sup_{|t|\leq DT_n} \left|\partial_t^r \varphi_{\overline{F}_0}(t) - \partial_t^r \varphi_{F'}(t)\right| \leq n^{-2C_1}\epsilon_n \qquad r = 0, 1, 2.$$

From the final claim of Lemma 5, we can also assume without loss of generality that the support points satisfy $\inf_{k\neq j}\|\mu_k - \mu_j\| \geq \epsilon_n^{L_2}$ for some $L_2 > 0$. Observe that $2\inf_{|t|\leq DT_n}\left|\varphi_{\overline{F}_0}(t)\right| \geq \inf_{|t|\leq DT_n}|\varphi_{F_0}(t)| \geq n^{-C_1}$ and

$$\begin{aligned}
\sup_{|t|\leq DT_n}\left|(\log\varphi_{\overline{F}_0})''(t)\right| &\leq \sup_{|t|\leq DT_n}\left|(\log\varphi_{F_0})''(t) - (\log\varphi_{\overline{F}_0})''(t)\right| + \sup_{|t|\leq DT_n}\left|(\log\varphi_{F_0})''(t)\right| \\
&\leq D(\epsilon_n + D) \\
&\leq D.
\end{aligned}$$

From the preceding bounds and an analogous expression to that of $(\log\varphi_{F_0})'' - (\log\varphi_{\overline{F}_0})''$, it follows that

$$\sup_{|t|\leq DT_n}\left|(\log\varphi_{\overline{F}_0})''(t) - (\log\varphi_{F'})''(t)\right| \leq D\epsilon_n.$$

Fix any $L_3 > L_2$ sufficiently large such that

$$n^{2C_1+1}\epsilon_n^2\log(n)\epsilon_n^{L_3/2} \leq \epsilon_n\ ,\ n^{2C_1}(\log\epsilon_n)^{2/\chi}T_n\epsilon_n^{L_3} \leq \epsilon_n.$$

Define $V_i = \{t \in I_n : |t - \mu_i| \leq \epsilon_n^{L_3}\}$ for $i = 1, \ldots, N$ and set $V_0 = \mathbb{R}\setminus\bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \ldots, V_N\}$ is a disjoint partition of $\mathbb{R}$. By Lemma 3, for any distribution P that satisfies $(\int_{\mathbb{R}}|x|^4\,dP(x))^{1/4} \leq \sqrt{n}\epsilon_n\sqrt{\log n}$ and $\sum_{j=1}^N|P(V_j) - p_j| \leq \epsilon_n^{L_3}$, we have that

$$\sup_{|t|\leq T_n}|\partial_t^r\varphi_{F'}(t) - \partial_t^r\varphi_P(t)| \leq n^{-2C_1}\epsilon_n \qquad r = 0, 1, 2.$$

For all such P, the preceding bound and an analogous expression to that of $(\log\varphi_{F_0})'' - (\log\varphi_{\overline{F}_0})''$, yields

$$\sup_{|t|\leq DT_n}\left|(\log\varphi_{F'})''(t) - (\log\varphi_P)''(t)\right| \leq D\epsilon_n.$$

From combining all the preceding bounds, observe that all such P also satisfy

$$\begin{aligned}
&\sup_{|t|\leq DT_n}\left|(\log\varphi_P)''(t) - (\log\varphi_{F_0})''(t)\right| \leq D\epsilon_n \\
&\sup_{|t|\leq DT_n}\left|(\log\varphi_P)''(t)\right| \leq D\epsilon_n + \sup_{|t|\leq DT_n}\left|(\log\varphi_{F_0})''(t)\right| \leq D.
\end{aligned}$$

Given any $\sigma^2 \geq 0$, the preceding bound also implies that

$$\sup_{|t|\leq DT_n} \left|(\log \varphi_{P,\sigma^2})''(t)\right| \leq \sup_{|t|\leq DT_n} \left[\left|(\log \varphi_P)''(t)\right| + \left|(\log \varphi_{\sigma^2})''(t)\right| \right]$$
$$\leq D(1+\sigma^2).$$

Observe that the (l,k) element of $\widehat{\varphi}^2_{\mathbf{Y}}(t)[\nabla\nabla' \log \widehat{\varphi}_{\mathbf{Y}}(t)]$ is given by

$$-\widehat{\varphi}_{\mathbf{Y}}(t)\mathbb{E}_n[Y_l Y_k e^{\mathbf{i}t'\mathbf{Y}}] + \mathbb{E}_n[Y_l e^{\mathbf{i}t'\mathbf{Y}}]\mathbb{E}_n[Y_k e^{\mathbf{i}t'\mathbf{Y}}].$$

From this representation and an application of Lemma 2 and 9, we have that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|^2_{\mathbb{B}(T)} \leq D\frac{T_n^d \log(T_n)}{n} \leq D\epsilon_n^2.$$

with $\mathbb{P}$ probability approaching 1. On this set, when P is specified as above and $\sigma^2 \leq D$, we have that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}$$
$$\leq D\left[\|\varphi^2_{\mathbf{Y}}(\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)} + \epsilon_n^2\right].$$

Define the set

$$\mathcal{G}_n = \left\{(P,\sigma^2) : \left|\sigma^2 - \sigma_0^2\right| \leq \epsilon_n^2, \int_{\mathbb{R}} |x|^4\, dP(x) \leq n^2\epsilon_n^4(\log n)^2\,, \sum_{j=1}^N |P(V_j) - p_j| \leq \epsilon_n^{L_3}\right\}.$$

From combining the all the preceding bounds, it follows that

$$\int \exp\bigg(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\bigg) d\nu_{\alpha,G}(P,\sigma^2)$$
$$\geq \int_{\mathcal{G}_n} \exp\bigg(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\bigg) d\nu_{\alpha,G}(P,\sigma^2)$$
$$\geq \exp\big(-nD\epsilon_n^2\big) \int_{\mathcal{G}_n} \exp\bigg(-nD\|\varphi^2_{\mathbf{Y}}(\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\bigg) d\nu_{\alpha,G}(P,\sigma^2).$$

Since $\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}(t) = (\log \varphi_X)''(t'\mathbf{A}_k)$ and $\varphi_X = \varphi_{F_0,\sigma_0^2}$, the preceding integral can be expressed as

$$\int_{\mathcal{G}_n} \exp\bigg(-nD\|\varphi^2_{\mathbf{Y}}(\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}} - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\bigg) d\nu_{\alpha,G}(P,\sigma^2)$$
$$= \int_{\mathcal{G}_n} \exp\bigg(-nD\|\varphi^2_{\mathbf{Y}}[(\log \varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T_n)}\bigg) d\nu_{\alpha,G}(P,\sigma^2).$$

For every $(P,\Sigma) \in \mathcal{G}_n$, the preceding bounds imply that

$$\sup_{\|t\|_\infty \leq T_n} \left|(\log \varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\right| \leq D\epsilon_n.$$

Since $\|\varphi^2_{\mathbf{Y}}\|_{L^2}< \infty$, the preceding bound implies that

$$\int_{\mathcal{G}_n} \exp\Big(- nD\|\varphi^2_{\mathbf{Y}}[(\log\varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\sigma^2)$$
$$\geq \exp\big(-nD\epsilon_n^2\big)\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\sigma^2).$$

Define the sets

$$\mathcal{G}_{n,1} = \left\{(P,\sigma^2) : \left|\sigma^2-\sigma_0^2\right| \leq \epsilon_n^2, \sum_{j=1}^N |P(V_j)-p_j| \leq \epsilon_n^{L_3}\right\},$$
$$\mathcal{G}_{n,2} = \left\{(P,\sigma^2) : \left|\sigma^2-\sigma_0^2\right| \leq \epsilon_n^2, \int_{\mathbb{R}} |x|^4\, dP(x) > n^2\epsilon_n^4(\log n)^2 , \sum_{j=1}^N |P(V_j)-p_j| \leq \epsilon_n^{L_3}\right\}.$$

Observe that $\mathcal{G}_n = \mathcal{G}_{n,1}\setminus\mathcal{G}_{n,2}$. Hence

$$\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\sigma^2) = \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\sigma^2) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\sigma^2).$$

For the second term, observe that

$$\begin{aligned}\int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\sigma^2) &\leq \int_{P:\int_{\mathbb{R}}|x|^4 dP(x)>n^2\epsilon_n^4(\log n)^2} d\nu_{\alpha,G}(P,\sigma^2)\\ &\leq \int_{P:\int_{\mathbb{R}}|x|^4 dP(x)>n^2\epsilon_n^4(\log n)^2} d\mathrm{DP}_\alpha(P)\\ &\leq \exp\big(-Dn\epsilon_n^2\log n\big)\ ,\end{aligned}$$

where the second inequality is due to $\nu_{\alpha,G}$ being a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha\otimes G$ and the third inequality follows from an application of Lemma 12.

For the first term, we have that

$$\int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\sigma^2) = \int_{\sigma^2:\left|\sigma^2-\sigma_0^2\right|\leq\epsilon_n^2}\int_{P:\sum_{j=1}^N|P(V_j)-p_j|\leq\epsilon_n^{L3}} d\mathrm{DP}_\alpha(P)dG(\sigma^2).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^N\alpha(V_j)\geq C\epsilon_n^{L_3 d}\exp\big(-C'(\log\epsilon_n^{-1})^{2/\chi}\big)$ for universal constants $C,C'>0$. By definition of DP_α, $(P(V_1),\dots,P(V_N))\sim \mathrm{Dir}(N,\alpha(V_1),\dots,\alpha(V_N))$. As $N = D\{\log\big(\epsilon_n^{-1}\big)\}^\iota$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) implies

$$\begin{aligned}\int_{P:\sum_{j=1}^N|P(V_j)-p_j|\leq\epsilon_n^{L3}} d\mathrm{DP}_\alpha(P) \geq C\exp\big(- C'(\log\epsilon_n^{-1})^{\iota+\max\{2/\chi,1\}}\big) &= C\exp\big(- C'(\log\epsilon_n^{-1})^{\lambda}\big)\\ &\geq C\exp\big(- C''n\epsilon_n^2\big)\end{aligned}$$

It remains to bound the outer integral. The law of G is given by Ω/σ_n^2 where $\Omega\sim L$ and L is a probability measure on $\mathbb{R}_+$ that satisfies Assumption 2. By Assumption 2 and the definition of σ_n^2, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\sigma^2:|\sigma^2-\sigma_0^2|\leq\epsilon_n^2} dG(\sigma^2) = \int_{\sigma^2:|\sigma^2-\sigma_n^2\sigma_0^2|\leq\sigma_n^2\epsilon_n^2} dL(\sigma^2) \geq C\exp\big(-C'\sigma_n^{-2\kappa}\big) \geq C\exp\big(-C''n\epsilon_n^2\big).$$

It follows that

$$\begin{aligned}\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\sigma^2) &= \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\sigma^2) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\sigma^2)\\ &\geq C\exp\big(-C''n\epsilon_n^2\big) - \exp\big(-Dn\epsilon_n^2\log n\big)\\ &\geq \exp\big(-Dn\epsilon_n^2\big).\end{aligned}$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(*ii*) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E > 0$, we have

$$\int_{(P,\sigma^2):\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2>2E\epsilon_n^2} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2\Big)d\nu_{\alpha,G}(P,\Sigma$$
$$\leq \exp\big(-nE\epsilon_n^2\big).$$

The law of G is given by Σ/σ_n^2 where $\Sigma \sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E' > 0$, there exists $E > 0$ such that

$$\int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\sigma^2) = \int_{\sigma^2:|\sigma^{-2}|>E(n\epsilon_n^2)^{1/\kappa}} dL(\sigma^2) \leq \exp\big(-E'n\epsilon_n^2\big).$$

As the prior is a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha\otimes G$ and $\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2\geq 0$, the preceding bound implies

$$\begin{aligned}&\int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2\Big)d\nu_{\alpha,G}(P,\sigma^2)\\ &\leq \int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\sigma^2)\\ &\leq \exp\big(-E'n\epsilon_n^2\big).\end{aligned}$$

From combining the preceding bounds, it follows that for any $E' > 0$ we can pick $E > 0$ sufficiently large such that

$$\int_{\substack{(P,\sigma^2):\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2\leq E\epsilon_n^2,\\ |\sigma^{-2}|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T_n)}^2\Big)d\nu_{\alpha,G}(P,\sigma^2$$
$$\geq 1 - \exp\big(-E'n\epsilon_n^2\big).$$

(*iii*) We prove the main statement of the theorem. From the bounds derived in steps (*i*) and (*ii*), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\bigg(\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)} \leq M^2\epsilon_n^2, \|\Sigma^{-1}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\bigg|\,\mathcal{Z}_n, T_n\bigg) \geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

The (l,k) element of $\widehat{\varphi}^2_{\mathbf{Y}}(t)\widehat{\mathcal{V}}_{\mathbf{Y}}(t)$ is given by

$$-\widehat{\varphi}_{\mathbf{Y}}(t)\mathbb{E}_n[Y_lY_ke^{\mathbf{i}t'\mathbf{Y}}] + \mathbb{E}_n[Y_le^{\mathbf{i}t'\mathbf{Y}}]\mathbb{E}_n[Y_ke^{\mathbf{i}t'\mathbf{Y}}].$$

From this representation and an application of Lemma 2 and 9, we have that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|_{\mathbb{B}(T)} \leq D\epsilon_n.$$

Since $\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}(t) = (\log\varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k)$ and $\Big|(\log\varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k)\Big| \leq D$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}(\log\varphi_{F_0,\Sigma_0})''(t'\mathbf{A}_k) - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|_{\mathbb{B}(T_n)} \leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\bigg(\|\widehat{\varphi}^2_{\mathbf{Y}}[(\log\varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T_n)} \leq M^2\epsilon_n^2, \big|\sigma^{-2}\big| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\bigg|\,\mathcal{Z}_n, T_n\bigg) \geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1. Since $T_n = (c_0\log n)^{1/2}$ for some c_0 satisfying $2c_0R = \gamma < 1/2$, an application of Lemma 2 implies that

$$\sup_{\|t\|_\infty \leq T_n} \frac{\big|\widehat{\varphi}^2_{\mathbf{Y}}(t) - \varphi^2_{\mathbf{Y}}(t)\big|}{\big|\varphi^2_{\mathbf{Y}}(t)\big|} \leq Dn^{-1/2+\gamma}\sqrt{\log\log n}$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, it follows that

$$\big|\widehat{\varphi}^2_{\mathbf{Y}}(t)\big| \geq \big|\varphi^2_{\mathbf{Y}}(t)\big| - \big|\widehat{\varphi}^2_{\mathbf{Y}}(t) - \varphi^2_{\mathbf{Y}}(t)\big| \geq \frac{1}{2}\big|\varphi^2_{\mathbf{Y}}(t)\big| \geq \frac{1}{2}n^{-\gamma}.$$

uniformly over the set $\{\|t\|_\infty \leq T_n\}$, with $\mathbb{P}$ probability approaching 1. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|(\log \varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log \varphi_{P,\Sigma})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)} \leq M^2 n^{2\gamma}\epsilon_n^2, \|\sigma^{-2}\| \leq M^2\sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\bigl(-C' n\epsilon_n^2\bigr)$$

holds with $\mathbb{P}$ probability approaching 1.

Recall that we use the posterior measure

$$\overline{\nu}_{\alpha,G}(. \mid T, k, \mathcal{Z}_n) \sim Z - \mathbb{E}[Z] \qquad \text{where} \qquad Z \sim \nu_{\alpha,G}(.|T, k, \mathcal{Z}_n). \tag{90}$$

Denote the characteristic function of a demeaned Gaussian mixture φ_{P,σ^2} by $\overline{\varphi}_{P,\sigma^2}$. For any distribution Z, we have $(\log \varphi_Z)'' = (\log \varphi_{Z-\mathbb{E}[Z]})''$. From this observation and the preceding inequalities for $\nu_{\alpha,G}$, it follows that

$$\overline{\nu}_{\alpha,G}\left(\|(\log \varphi_{F_0,\sigma_0^2})''(t'\mathbf{A}_k) - (\log \overline{\varphi}_{P,\sigma^2})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)} \leq M^2 n^{2\gamma}\epsilon_n^2, \left|\sigma^{-2}\right| \leq M^2\sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\bigl(-C' n\epsilon_n^2\bigr)$$

holds with $\mathbb{P}$ probability approaching 1.

Denote the elements of $\mathbf{A}_k$ by $\mathbf{A}_k = (a_1, \ldots, a_L)$. Fix any i such that $a_i \neq 0$. Without loss of generality, let $i = 1$ and $a_i > 0$. Consider the change of variables

$$z_1 = t'\mathbf{A}_k \ , \ z_2 = t_2 \ , \ldots, z_L = t_L.$$

The Jacobian of the change of variables $(t_1, \ldots, t_L) \to (z_1, \ldots, z_L)$ is given by $J(z_1, \ldots, z_L) = a_1^{-1}$. Let $c_L = \inf_{t\in\mathbb{B}(T_n)} t'\mathbf{A}_k$ and $c_U = \sup_{t\in\mathbb{B}(T_n)} t'\mathbf{A}_k$. It follows that for any non-negative Borel function $f : \mathbb{R} \to \mathbb{R}_+$, we have that

$$\int_{\mathbb{B}(T_n)} f(t'\mathbf{A}_k)dt = |a_1|^{-1} (2T_n)^{d-1} \int_{c_L}^{c_U} f(z_1)dz_1.$$

In particular, since $c_U \geq a_1 T_n$ and $c_L \leq -a_1 T_n$, we have $\|f\|^2_{\mathbb{B}(a_1 T_n)} \leq DT_n^{1-d}\|f(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)}$ for some universal constant $D > 0$. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\overline{\nu}_{\alpha,G}\left(\|(\log \varphi_{F_0,\sigma_0^2})''(.) - (\log \overline{\varphi}_{P,\sigma^2})''(.)\|^2_{\mathbb{B}(a_1 T_n)} \leq M^2 n^{2\gamma} T_n^{1-d}\epsilon_n^2, \left|\sigma^{-2}\right| \leq M^2\sigma_n^2 (n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\bigl(-C' n\epsilon_n^2\bigr)$$

holds with $\mathbb{P}$ probability approaching 1. Since the Gaussian mixture and the true latent distribution are demeaned, we have $(\log \varphi_{F_0,\sigma_0^2})'(0) = (\log \overline{\varphi}_{P,\sigma^2})'(0) = 0$. From the

Fundemental theorem of calculus and Cauchy-Schwarz, we obtain

$$\Big|(\log\varphi_{F_0,\sigma_0^2})'(t)-(\log\overline{\varphi}_{P,\sigma^2})'(t)\Big| = \left|\int_0^t [(\log\varphi_{F_0,\sigma_0^2})''(s)-(\log\overline{\varphi}_{P,\sigma^2})''(s)]ds\right|$$
$$\leq \sqrt{a_1}\sqrt{T_n}\|(\log\varphi_{F_0,\sigma_0^2})''(.)-(\log\overline{\varphi}_{P,\sigma^2})''(.)\|_{\mathbb{B}(a_1T_n)}$$

for every $t\in\mathbb{B}(a_1T_n)$. As all characteristic functions satisfy $\log\varphi(0)=0$, we similarly obtain

$$\Big|\log\varphi_{F_0,\sigma_0^2}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\Big| = \left|\int_0^t [(\log\varphi_{F_0,\sigma_0^2})'(s)-(\log\overline{\varphi}_{P,\Sigma})'(s)]ds\right|$$
$$\leq \sqrt{a_1}\sqrt{T_n}\|(\log\varphi_{F_0,\sigma_0^2})'(.)-(\log\overline{\varphi}_{P,\sigma^2})'(.)\|_{\mathbb{B}(a_1T_n)}$$

for every $t\in\mathbb{B}(a_1T_n)$. Furthermore, for every fixed $t\in\mathbb{R}$, the mean value theorem implies that

$$\Big|\varphi_{F_0,\sigma_0^2}(t)-\overline{\varphi}_{P,\Sigma}(t)\Big| \leq \sup_{s_t\in[0,1]}\Big|e^{s_t\log\varphi_{F_0,\sigma_0^2}(t)+(1-s_t)\log\overline{\varphi}_{P,\sigma^2}(t)}\Big|\,\Big|\log\varphi_{F_0,\sigma_0^2}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\Big|.$$

Since $|\varphi_X|\leq 1$ and $\left|\varphi_{P,\sigma^2}\right|\leq 1$ (as they are characteristic function of random variables), the preceding bound reduces to

$$\Big|\varphi_{F_0,\sigma_0^2}(t)-\overline{\varphi}_{P,\sigma^2}(t)\Big| \leq \Big|\log\varphi_{F_0,\sigma_0^2}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\Big|.$$

From combining all the preceding bounds, it follows that there exists a universal constant $D>0$ such that

$$\|\varphi_{F_0,\sigma_0^2}-\overline{\varphi}_{P,\sigma^2}\|_{\mathbb{B}(a_1T_n)}\leq DT_n^2\|(\log\varphi_{F_0,\sigma_0^2})''(.)-(\log\overline{\varphi}_{P,\sigma^2})''(.)\|_{\mathbb{B}(a_1T_n)}.$$

It follows that for any $C'>0$, there exists a $M>0$ such that

$$\overline{\nu}_{\alpha,G}\left(\|\varphi_{F_0,\sigma_0^2}-\overline{\varphi}_{P,\sigma^2}\|^2_{\mathbb{B}(a_1T_n)}\leq M^2n^{2\gamma}T_n^{5-d}\epsilon_n^2, \left|\sigma^{-2}\right|\leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\Big|\,\mathcal{Z}_n,T_n\right)$$
$$\geq 1-\exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(a_1T_n)$. Suppose $\left|\sigma^{-2}\right|\leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c>0$ for which $\sigma^2\geq c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}\asymp\log n$.

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_{F_0,\sigma_0^2} - \overline{\varphi}_{P,\sigma^2})\mathbb{1}\{|t| > a_1T_n\}\|_{L^2}^2 &\leq 2\|\varphi_{F_0,\sigma_0^2}\mathbb{1}\{|t| > a_1T_n\}\|_{L^2}^2 + 2\|\overline{\varphi}_{P,\sigma^2}\mathbb{1}\{\|t\|_\infty > a_1T_n\}\|_{L^2}^2 \\ &\leq 2\int_{|t|>a_1T_n} e^{-t^2\sigma_0^2}dt + 2\int_{|t|>a_1T_n} e^{-t^2\sigma^2}dt \\ &\leq 2\int_{|t|>a_1T_n} e^{-t^2\sigma_0^2}dt + 2\int_{|t|>a_1T_n} e^{-ct^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}}dt \\ &\leq D\Big[e^{-CT_n^2}T_n^{d-2} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}e^{-CT_n^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}}T_n^{d-2}\Big].\end{aligned}$$

From substituting $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp \log n$, the preceding bound reduces to Dn^{-2K} for some constant $K \in (0, 1/2]$ (that depends on, among other factors, σ_0^2).

With constant K as specified above, it follows that for any $C' > 0$, there exists $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\varphi_{F_0,\sigma_0^2} - \varphi_{P,\sigma^2}\|_{L^2} \leq MT_n^{(5-d)/2}n^\gamma\epsilon_n + n^{-K}\ |\mathcal{Z}_n, T_n\Big) \geq 1 - \exp\big(-C'n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$T_n^{(5-d)/2}n^\gamma\epsilon_n \asymp n^{-1/2+\gamma}(\log n)^{\lambda/2+(5-d)/4}.$$

□

Proof of Theorem 8. The proof proceeds through several steps which we outline below. We use $D, D' > 0$ as a generic universal constant that may change from line to line. Define

$$\alpha_n^2 = \epsilon_n^{\frac{2}{s+\zeta}} \quad , \quad \epsilon_n^2 = \frac{(\log n)^{\lambda+d/2}}{n^{\frac{2(s+2\zeta)}{2(s+2\zeta)+d}}} \quad , \quad \lambda = \begin{cases}\chi^{-1}(d+2) & \chi < 2 \\ d/\chi + 1 & \chi \geq 2.\end{cases}$$

(*i*) First, we derive a lower bound for the normalizing constant of the posterior measure. Specifically, we aim to show that there exists a $C > 0$ such that

$$\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}_{\mathbf{Y}}^2(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|_{\mathbb{B}(T)}^2\Big)d\nu_{\alpha,G}(P,\sigma^2) \geq \exp\big(-Cn\epsilon_n^2\big) \quad (91)$$

holds with $\mathbb{P}$ probability approaching 1.

Note that $|t'\mathbf{A}_k| \leq \sqrt{d}\|t\|_\infty\|\mathbf{A}_k\| \leq \sqrt{d}T_n\|\mathbf{A}_k\|$ uniformly over $t \in \mathbb{B}(T_n)$. Since the preceding constant is finite, it suffices to work under the setting where $|t'\mathbf{A}_k| \leq D'T_n$.

Observe that the (l, k) element of $\widehat{\varphi}_{\mathbf{Y}}^2(t)[\nabla\nabla'\log\widehat{\varphi}_{\mathbf{Y}}(t)]$ is given by

$$-\widehat{\varphi}_{\mathbf{Y}}(t)\mathbb{E}_n[Y_lY_ke^{\mathbf{i}t'\mathbf{Y}}] + \mathbb{E}_n[Y_le^{\mathbf{i}t'\mathbf{Y}}]\mathbb{E}_n[Y_ke^{\mathbf{i}t'\mathbf{Y}}].$$

From this representation and an application of Lemma 2 and 9, it follows that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|^2_{\mathbb{B}(T)} \leq D\frac{T_n^d\log(T_n)}{n} \leq D\epsilon_n^2$$

with $\mathbb{P}$ probability approaching 1. On this set, it follows that

$$\int \exp\Big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\sigma^2)$$
$$\geq \exp\big(-nD\epsilon_n^2\big)\int \exp\Big(-nD\|\varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}} - \widehat{\varphi}^2_{\mathbf{Y}}(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\sigma^2).$$

From the model, we have that $\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}(t) = (\log\varphi_{X_k})''(t'\mathbf{A}_k)$. By Condition 4.9, there exists universal constants $\chi, C, M < \infty$ and a mixing distribution S_{α_n} supported on $I_n = [-C(\log\epsilon_n^{-1})^{1/\chi}, C(\log\epsilon_n^{-1})^{1/\chi}]$ that satisfies

$$\|\varphi^2_{\mathbf{Y}}(t)[(\log\varphi_{S_{\alpha_n},\alpha_n^2})''(t'\mathbf{A}_k) - (\log\varphi_{X_k})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T)} \leq D\alpha_n^{4\zeta+2s} \leq D\epsilon_n^2.$$

It follows that

$$\int \exp\Big(-nD\|\varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}} - \widehat{\varphi}^2_{\mathbf{Y}}(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\sigma^2)$$
$$\geq \exp\big(-nD\epsilon_n^2\big)\int \exp\Big(-nD\|\varphi^2_{\mathbf{Y}}(\log\varphi_{S_{\alpha_n},\alpha_n^2})''(t'\mathbf{A}_k) - \widehat{\varphi}^2_{\mathbf{Y}}(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)}\Big) d\nu_{\alpha,G}(P,\sigma^2).$$

By Condition 4.9 and the definition of (α_n, T_n), there exists a $C_1 > 0$ such that

$$\inf_{|t|\leq D'T_n}\big|\varphi_{S_{\alpha_n}}(t)\big| \geq n^{-C_1}.$$

Next, we show that S_{α_n} can be suitably approximated by a discrete measure. By Lemma 5, there exists a discrete measure $F' = \sum_{i=1}^N p_i\delta_{\mu_i}$ with at most $N = D(\log\epsilon_n^{-1})^{d/\chi}T_n^d$ support points on I_n such that

$$\sup_{|t|\leq D'T_n}\big|\partial_t^r\varphi_{S_{\alpha_n}}(t) - \partial_t^r\varphi_{F'}(t)\big| \leq n^{-2C_1}\epsilon_n \qquad r = 0,1,2.$$

From the final claim of Lemma 5, we can also assume without loss of generality that the support points satisfy $\inf_{k\neq j}\|\mu_k - \mu_j\| \geq \epsilon_n^{L_2}$ for some $L_2 > 0$. From the preceding bounds, we obtain

$$\sup_{|t|\leq D'T_n}\sup_{r=0,1,2}\frac{\big|\partial_t^r\varphi_{S_{\alpha_n}}(t) - \partial_t^r\varphi_{F'}(t)\big|}{\big|\varphi_{S_{\alpha_n}}(t)\big|^2} \leq D\epsilon_n.$$

Since $\epsilon_n \downarrow 0$, observe that this also implies

$$|\varphi_{F'}(t)| \geq \big|\varphi_{S_{\alpha_n}}(t)\big| - \big|\varphi_{S_{\alpha_n}}(t) - \varphi_{F'}(t)\big| \geq \big|\varphi_{S_{\alpha_n}}(t)\big| - D\epsilon_n\big|\varphi_{S_{\alpha_n}}(t)\big| \geq \frac{1}{2}\big|\varphi_{S_{\alpha_n}}(t)\big|$$

for all sufficiently large n and $|t| \leq D'T_n$. We write the difference $(\log\varphi_{S_{\alpha_n}})'' - (\log\varphi_{F'})''$

as

$$
\begin{aligned}
&(\log \varphi_{S_{\alpha_n}})'' - (\log \varphi_{F'})'' \\
&= (\log \varphi_{S_{\alpha_n}})'' \frac{(\varphi_{F'}^2 - \varphi_{S_{\alpha_n}}^2)}{\varphi_{F'}^2} + \frac{\varphi_{S_{\alpha_n}}(\varphi_{S_{\alpha_n}}'' - \varphi_{F'}'')}{\varphi_{F'}^2} + \frac{\varphi_{F'}''(\varphi_{S_{\alpha_n}} - \varphi_{F'})}{\varphi_{F'}^2} + \frac{(\varphi_{F'}')^2 - (\varphi_{S_{\alpha_n}}')^2}{\varphi_{F'}^2}.
\end{aligned}
$$

By Condition 4.9, we have $\left|(\log \varphi_{S_{\alpha_n}})''\right| \leq D$ and $\sup_{r=1,2} \left|\partial_t^r \varphi_{S_{\alpha_n}}\right| \leq D$. From the preceding representation and bounds, we then obtain

$$
\sup_{|t| \leq D' T_n} \left|(\log \varphi_{S_{\alpha_n}})''(t) - (\log \varphi_{F'})''(t)\right| \leq D\epsilon_n.
$$

Fix any $L_3 > L_2$ sufficiently large such that

$$
n^{2C_1+1} \epsilon_n^2 \log(n) \epsilon_n^{L_3/2} \leq \epsilon_n \text{ , } n^{2C_1} (\log \epsilon_n)^{2/\chi} T_n \epsilon_n^{L_3} \leq \epsilon_n.
$$

Define $V_i = \{t \in I_n : |t - \mu_i| \leq \epsilon_n^{L_3}\}$ for $i = 1, \ldots, N$ and set $V_0 = \mathbb{R} \setminus \bigcup_{i=1}^N V_i$. From the definition of the $\{\mu_i\}_{i=1}^N$, it follows that $\{V_0, V_1, \ldots, V_N\}$ is a disjoint partition of $\mathbb{R}$. By Lemma 3, for any distribution P that satisfies $(\int_{\mathbb{R}} |x|^4 \, dP(x))^{1/4} \leq \sqrt{n}\epsilon_n \sqrt{\log n}$ and $\sum_{j=1}^N |P(V_j) - p_j| \leq \epsilon_n^{L_3}$, we have that

$$
\sup_{|t| \leq D' T_n} |\partial_t^r \varphi_{F'}(t) - \partial_t^r \varphi_P(t)| \leq n^{-2C_1} \epsilon_n \qquad r = 0, 1, 2.
$$

For all such P, the preceding bound and an analogous expression to that of $(\log \varphi_{S_{\alpha_n}})'' - (\log \varphi_{F'})''$, yields

$$
\sup_{|t| \leq D' T_n} \left|(\log \varphi_{F'})''(t) - (\log \varphi_P)''(t)\right| \leq D\epsilon_n.
$$

From combining all the preceding bounds, observe that all such P also satisfy

$$
\begin{aligned}
&\sup_{|t| \leq D' T_n} \left|(\log \varphi_P)''(t) - (\log \varphi_{S_{\alpha_n}})''(t)\right| \leq D\epsilon_n \\
&\sup_{|t| \leq D' T_n} \left|(\log \varphi_P)''(t)\right| \leq D\epsilon_n + \sup_{|t| \leq D' T_n} \left|(\log \varphi_{S_{\alpha_n}})''(t)\right| \leq D.
\end{aligned}
$$

Given any $\sigma^2 \leq D$, the preceding bound also implies that

$$
\begin{aligned}
\sup_{|t| \leq D' T_n} \left|(\log \varphi_{P,\sigma^2})''(t)\right| &\leq \sup_{|t| \leq D' T_n} \left[\left|(\log \varphi_P)''(t)\right| + \left|(\log \varphi_{\sigma^2})''(t)\right| \right] \\
&\leq D(1 + \sigma^2) \\
&\leq D.
\end{aligned}
$$

By Lemma 2, we have that

$$\int_{\mathbb{B}(T_n)} |\widehat{\varphi}_{\mathbf{Y}}(t) - \varphi_{\mathbf{Y}}(t)|^2 \, dt \leq D \frac{T_n^d \log(T_n)}{n} \leq D\epsilon_n^2$$

holds with $\mathbb{P}$ probability approaching 1. On this set, when (P, σ^2) satisfy the preceding requirements, we have that

$$\|\widehat{\varphi}_{\mathbf{Y}}^2 (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k) - \varphi_{\mathbf{Y}}^2 (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|_{\mathbb{B}(T_n)}^2 \leq D\epsilon_n^2.$$

Define the sets

$$\Omega_n = \left\{ \sigma^2 \in \mathbb{R}_+ : \sigma^2 \in \left[\frac{\alpha_n^2}{1+\epsilon_n}, \alpha_n^2 \right] \right\}.$$

$$\mathcal{G}_n = \left\{ (P, \sigma^2) : \sigma^2 \in \Omega_n, \int_{\mathbb{R}} |x|^4 \, dP(x) \leq n^2 \epsilon_n^4 (\log n)^2 \, , \sum_{j=1}^N |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\}.$$

From combining the preceding bounds, it follows that

$$\begin{aligned}
&\int \exp\left(- nD\|\varphi_{\mathbf{Y}}^2 (\log \varphi_{S_{\alpha_n}, \alpha_n^2})''(t'\mathbf{A}_k) - \widehat{\varphi}_{\mathbf{Y}}^2 (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \sigma^2) \\
&\geq \int_{\mathcal{G}_n} \exp\left(- nD\|\varphi_{\mathbf{Y}}^2 (\log \varphi_{S_{\alpha_n}, \alpha_n^2})''(t'\mathbf{A}_k) - \widehat{\varphi}_{\mathbf{Y}}^2 (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \sigma^2) \\
&\geq \exp\left(- nD\epsilon_n^2 \right) \int_{\mathcal{G}_n} \exp\left(- nD\|\varphi_{\mathbf{Y}}^2 [(\log \varphi_{S_{\alpha_n}, \alpha_n^2})''(t'\mathbf{A}_k) - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \sigma^2).
\end{aligned}$$

For every $(P, \sigma^2) \in \mathcal{G}_n$, the preceding bounds also imply that

$$\begin{aligned}
&\sup_{|t| \leq D'T_n} \left| (\log \varphi_{P,\sigma^2})''(t) - (\log \varphi_{S_{\alpha_n}, \alpha_n^2})''(t) \right| \\
&\leq \sup_{|t| \leq D'T_n} \left| (\log \varphi_P)''(t) - (\log \varphi_{S_{\alpha_n}})''(t) \right| + \sup_{|t| \leq D'T_n} \left| (\log \varphi_{\sigma^2})''(t) - (\log \varphi_{\alpha_n^2})''(t) \right| \\
&\leq D\epsilon_n + D\left| \sigma^2 - \alpha_n^2 \right| \\
&\leq D\epsilon_n.
\end{aligned}$$

Since $\|\varphi_{\mathbf{Y}}^2\|_{L^2} < \infty$, we obtain that

$$\begin{aligned}
&\int_{\mathcal{G}_n} \exp\left(- nD\|\varphi_{\mathbf{Y}}^2 [(\log \varphi_{S_{\alpha_n}, \alpha_n^2})''(t'\mathbf{A}_k) - (\log \varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|_{\mathbb{B}(T_n)}^2 \right) d\nu_{\alpha,G}(P, \sigma^2) \\
&\geq \exp\left(- nD\epsilon_n^2 \right) \int_{\mathcal{G}_n} d\nu_{\alpha,G}(P, \sigma^2).
\end{aligned}$$

Define the sets

$$\mathcal{G}_{n,1} = \left\{ (P, \sigma^2) : \sigma^2 \in \Omega_n,\ \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\},$$

$$\mathcal{G}_{n,2} = \left\{ (P, \sigma^2) : \sigma^2 \in \Omega_n,\ \int_{\mathbb{R}} |x|^4 \, dP(x) > n^2 \epsilon_n^4 (\log n)^2 ,\ \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3} \right\}.$$

Observe that $\mathcal{G}_n = \mathcal{G}_{n,1} \setminus \mathcal{G}_{n,2}$. Hence

$$\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P, \sigma^2) = \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P, \sigma^2) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P, \sigma^2).$$

For the second term, observe that

$$\begin{aligned}
\int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P, \sigma^2) &\leq \int_{P: \int_{\mathbb{R}} |x|^4 dP(x) > n^2 \epsilon_n^4 (\log n)^2} d\nu_{\alpha,G}(P, \sigma^2) \\
&\leq \int_{P: \int_{\mathbb{R}} |x|^4 dP(x) > n^2 \epsilon_n^4 (\log n)^2} d\mathrm{DP}_\alpha(P) \\
&\leq \exp\big(-D n \epsilon_n^2 \log n \big) ,
\end{aligned}$$

where the second inequality is due to $\nu_{\alpha,G}$ being a product measure $\nu_{\alpha,G} = \mathrm{DP}_\alpha \otimes G$ and the third inequality follows from an application of Lemma 12.

For the first term, we have that

$$\int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P, \sigma^2) = \int_{\sigma^2 \in \Omega_n} \int_{P: \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3}} d\mathrm{DP}_\alpha(P) dG(\sigma^2).$$

As DP_α is constructed using a Gaussian base measure α, it is straightforward to verify that $\inf_{j=1}^{N} \alpha(V_j) \geq C \epsilon_n^{L_3 d} \exp\big(-C' (\log \epsilon_n^{-1})^{2/\chi} \big)$ for universal constants $C, C' > 0$. By definition of DP_α, $(P(V_1), \ldots, P(V_N)) \sim \mathrm{Dir}(N, \alpha(V_1), \ldots, \alpha(V_N))$. As $N = D\{\log\big(\epsilon_n^{-1}\big)\}^{\iota}$, an application of (Ghosal and Van der Vaart, 2017, Lemma G.13) implies

$$\begin{aligned}
\int_{P: \sum_{j=1}^{N} |P(V_j) - p_j| \leq \epsilon_n^{L_3}} d\mathrm{DP}_\alpha(P) \geq C \exp\big(- C' (\log \epsilon_n^{-1})^{\iota + \max\{2/\chi, 1\}} \big) &= C \exp\big(- C' (\log \epsilon_n^{-1})^{\lambda} \big) \\
&\geq C \exp\big(- C'' n \epsilon_n^2 \big).
\end{aligned}$$

It remains to bound the outer integral. The law of $G = G_n$ is given by Ω / σ_n^2 where $\Omega \sim L$ and L is a probability measure on $\mathbb{R}_+$ that satisfies Assumption 2. By Assumption 2 and the definition of $(\alpha_n^2, \sigma_n^2, \epsilon_n^2)$, there exists a universal constant $C, C', C'' > 0$ such that

$$\int_{\sigma^2 \in \Omega_n} dG(\sigma^2) = \int_{\sigma^2 \in \sigma_n^2 \Omega_n} dL(\sigma^2) \geq C \exp\big(- C' \sigma_n^{-2\kappa} \alpha_n^{-2\kappa} \big) \geq C \exp\big(-C'' n \epsilon_n^2 \big).$$

It follows that

$$\begin{aligned}\int_{\mathcal{G}_n} d\nu_{\alpha,G}(P,\sigma^2) &= \int_{\mathcal{G}_{n,1}} d\nu_{\alpha,G}(P,\sigma^2) - \int_{\mathcal{G}_{n,2}} d\nu_{\alpha,G}(P,\sigma^2)\\ &\geq C\exp\big(-C''n\epsilon_n^2\big) - \exp\big(-Dn\epsilon_n^2\log n\big)\\ &\geq \exp\big(-Dn\epsilon_n^2\big).\end{aligned}$$

The estimate for the lower bound of the normalizing constant follows from combining all the preceding bounds.

(ii) Next, we establish a preliminary local concentration bound under the prior. Observe that for any $E>0$, we have

$$\int_{(P,\sigma^2):\|\widehat{\varphi}^2_{\mathbf{Y}}\big(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\big)\|^2_{\mathbb{B}(T_n)}>2E\epsilon_n^2} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\Big)d\nu_{\alpha,G}(P,\Sigma$$
$$\leq \exp\big(-nE\epsilon_n^2\big).$$

The law of G is given by Σ/σ_n^2 where $\Sigma\sim L$ and L is a probability measure on $\mathbf{S}_+^d$ that satisfies Assumption 2. By Assumption 2, it follows that for every $E'>0$, there exists $E>0$ such that

$$\int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\sigma^2) = \int_{\sigma^2:|\sigma^{-2}|>E(n\epsilon_n^2)^{1/\kappa}} dL(\sigma^2)\leq \exp\big(-E'n\epsilon_n^2\big).$$

As the prior is a product measure $\nu_{\alpha,G}=\mathrm{DP}_\alpha\otimes G$ and $\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\geq 0$, the preceding bound implies

$$\begin{aligned}&\int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\Big)d\nu_{\alpha,G}(P,\sigma^2)\\ &\leq \int_{\sigma^2:|\sigma^{-2}|>E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}} dG(\sigma^2)\\ &\leq \exp\big(-E'n\epsilon_n^2\big).\end{aligned}$$

From combining the preceding bounds, it follows that for any $E'>0$ we can pick $E>0$ sufficiently large such that

$$\int_{\substack{(P,\sigma^2):\|\widehat{\varphi}^2_{\mathbf{Y}}\big(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)\big)\|^2_{\mathbb{B}(T_n)}\leq E\epsilon_n^2,\\ |\sigma^{-2}|\leq E\sigma_n^2(n\epsilon_n^2)^{1/\kappa}}} \exp\Big(-\frac{n}{2}\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}_k^*\widehat{\mathcal{V}}_{\mathbf{Y}}-(\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)}\Big)d\nu_{\alpha,G}(P,\sigma^2$$
$$\geq 1-\exp\big(-E'n\epsilon_n^2\big).$$

(iii) We prove the main statement of the theorem. From the bounds derived in steps *(i)* and

(ii), it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}^2_{\mathbf{Y}}(\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k))\|^2_{\mathbb{B}(T_n)} \leq M^2\epsilon_n^2, \|\Sigma^{-1}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \middle| \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

The (l,k) element of $\widehat{\varphi}^2_{\mathbf{Y}}(t)\widehat{\mathcal{V}}_{\mathbf{Y}}(t)$ is given by

$$-\widehat{\varphi}_{\mathbf{Y}}(t)\mathbb{E}_n[Y_lY_ke^{\mathbf{i}t'\mathbf{Y}}] + \mathbb{E}_n[Y_le^{\mathbf{i}t'\mathbf{Y}}]\mathbb{E}_n[Y_ke^{\mathbf{i}t'\mathbf{Y}}].$$

From this representation and an application of Lemma 2 and 9, we have that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}\mathbf{Q}^*_k\widehat{\mathcal{V}}_{\mathbf{Y}} - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|_{\mathbb{B}(T)} \leq D\epsilon_n.$$

Since $\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}(t) = (\log\varphi_{X_k})''(t'\mathbf{A}_k)$ and $\left|(\log\varphi_{X_k})''(t'\mathbf{A}_k)\right| \leq D$, an application of Lemma 2 implies that

$$\|\widehat{\varphi}^2_{\mathbf{Y}}(\log\varphi_X)''(t'\mathbf{A}_k) - \varphi^2_{\mathbf{Y}}\mathbf{Q}^*_k\mathcal{V}_{\mathbf{Y}}\|_{\mathbb{B}(T_n)} \leq D\epsilon_n$$

holds with $\mathbb{P}$ probability approaching 1.

From combining the preceding bounds, it follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|\widehat{\varphi}^2_{\mathbf{Y}}[(\log\varphi_{X_k})''(t'\mathbf{A}_k) - (\log\varphi_{P,\sigma^2})''(t'\mathbf{A}_k)]\|^2_{\mathbb{B}(T_n)} \leq M^2\epsilon_n^2, \left|\sigma^{-2}\right| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \middle| \mathcal{Z}_n, T_n \right)$$
$$\geq 1 - \exp\left(-C'n\epsilon_n^2\right)$$

holds with $\mathbb{P}$ probability approaching 1.

An application of Lemma 2 implies that

$$\sup_{\|t\|_\infty \leq T_n} \frac{\left|\widehat{\varphi}^2_{\mathbf{Y}}(t) - \varphi^2_{\mathbf{Y}}(t)\right|}{\left|\varphi^2_{\mathbf{Y}}(t)\right|} \leq DT_n^{2\zeta}\frac{\sqrt{\log n}}{\sqrt{n}}.$$

with $\mathbb{P}$ probability approaching 1. As the quantity on the right converges to zero, it follows that

$$\left|\widehat{\varphi}^2_{\mathbf{Y}}(t)\right| \geq \left|\varphi^2_{\mathbf{Y}}(t)\right| - \left|\widehat{\varphi}^2_{\mathbf{Y}}(t) - \varphi^2_{\mathbf{Y}}(t)\right| \geq \frac{1}{2}\left|\varphi^2_{\mathbf{Y}}(t)\right|$$

uniformly over the set $\{\|t\|_\infty \leq T_n\}$, with $\mathbb{P}$ probability approaching 1. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\left(\|(\log\varphi_{X_k})''(t'\mathbf{A}_k) - (\log\varphi_{P,\Sigma})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)} \leq M^2 T_n^{4\zeta}\epsilon_n^2, \|\sigma^{-2}\| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n\right)$$
$$\geq 1 - \exp\big(-C' n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

Recall that we use the posterior measure

$$\overline{\nu}_{\alpha,G}(.\mid T,k,\mathcal{Z}_n) \sim Z - \mathbb{E}[Z] \qquad \text{where} \qquad Z \sim \nu_{\alpha,G}(.|T,k,\mathcal{Z}_n). \tag{92}$$

Denote the characteristic function of a demeaned Gaussian mixture φ_{P,σ^2} by $\overline{\varphi}_{P,\sigma^2}$. For any distribution Z, we have $(\log\varphi_Z)'' = (\log\varphi_{Z-\mathbb{E}[Z]})''$. From this observation and the preceding inequalities for $\nu_{\alpha,G}$, it follows that

$$\overline{\nu}_{\alpha,G}\left(\|(\log\varphi_{X_k})''(t'\mathbf{A}_k) - (\log\overline{\varphi}_{P,\sigma^2})''(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)} \leq M^2 T_n^{4\zeta}\epsilon_n^2, \left|\sigma^{-2}\right| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n\right)$$
$$\geq 1 - \exp\big(-C' n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1.

Denote the elements of $\mathbf{A}_k$ by $\mathbf{A}_k = (a_1,\ldots,a_L)$. Fix any i such that $a_i \neq 0$. Without loss of generality, let $i = 1$ and $a_i > 0$. Consider the change of variables

$$z_1 = t'\mathbf{A}_k\ ,\ z_2 = t_2\ ,\ldots, z_L = t_L.$$

The Jacobian of the change of variables $(t_1,\ldots,t_L) \to (z_1,\ldots,z_L)$ is given by $J(z_1,\ldots,z_L) = a_1^{-1}$. Let $c_L = \inf_{t\in\mathbb{B}(T_n)} t'\mathbf{A}_k$ and $c_U = \sup_{t\in\mathbb{B}(T_n)} t'\mathbf{A}_k$. It follows that for any non-negative Borel function $f : \mathbb{R} \to \mathbb{R}_+$, we have that

$$\int_{\mathbb{B}(T_n)} f(t'\mathbf{A}_k)dt = |a_1|^{-1}(2T_n)^{d-1}\int_{c_L}^{c_U} f(z_1)dz_1.$$

In particular, since $c_U \geq a_1 T_n$ and $c_L \leq -a_1 T_n$, we have $\|f\|^2_{\mathbb{B}(a_1T_n)} \leq DT_n^{1-d}\|f(t'\mathbf{A}_k)\|^2_{\mathbb{B}(T_n)}$ for some universal constant $D > 0$. It follows that for any $C' > 0$, there exists a $M > 0$ such that

$$\overline{\nu}_{\alpha,G}\left(\|(\log\varphi_{X_k})''(.) - (\log\overline{\varphi}_{P,\sigma^2})''(.)\|^2_{\mathbb{B}(a_1T_n)} \leq M^2 T_n^{4\zeta}T_n^{1-d}\epsilon_n^2, \left|\sigma^{-2}\right| \leq M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa} \,\middle|\, \mathcal{Z}_n, T_n\right)$$
$$\geq 1 - \exp\big(-C' n\epsilon_n^2\big)$$

holds with $\mathbb{P}$ probability approaching 1. Since the Gaussian mixture and the true latent distribution are demeaned, we have $(\log\varphi_{X_k})'(0) = (\log\overline{\varphi}_{P,\sigma^2})'(0) = 0$. From the

Fundemental theorem of calculus and Cauchy-Schwarz, we obtain

$$\begin{aligned}\left|(\log\varphi_{X_k})'(t)-(\log\overline{\varphi}_{P,\sigma^2})'(t)\right| &= \left|\int_0^t [(\log\varphi_{X_k})''(s)-(\log\overline{\varphi}_{P,\sigma^2})''(s)]ds\right| \\ &\le \sqrt{a_1}\sqrt{T_n}\|(\log\varphi_{X_k})''(.)-(\log\overline{\varphi}_{P,\sigma^2})''(.)\|_{\mathbb{B}(a_1T_n)}\end{aligned}$$

for every $t\in\mathbb{B}(a_1T_n)$. As all characteristic functions satisfy $\log\varphi(0)=0$, we similarly obtain

$$\begin{aligned}\left|\log\varphi_{X_k}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\right| &= \left|\int_0^t [(\log\varphi_{X_k})'(s)-(\log\overline{\varphi}_{P,\Sigma})'(s)]ds\right| \\ &\le \sqrt{a_1}\sqrt{T_n}\|(\log\varphi_{X_k})'(.)-(\log\overline{\varphi}_{P,\sigma^2})'(.)\|_{\mathbb{B}(a_1T_n)}\end{aligned}$$

for every $t\in\mathbb{B}(a_1T_n)$. Furthermore, for every fixed $t\in\mathbb{R}$, the mean value theorem implies that

$$\left|\varphi_{X_k}(t)-\overline{\varphi}_{P,\Sigma}(t)\right| \le \sup_{s_t\in[0,1]}\left|e^{s_t\log\varphi_{X_k}(t)+(1-s_t)\log\overline{\varphi}_{P,\sigma^2}(t)}\right|\left|\log\varphi_{X_k}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\right|.$$

Since $|\varphi_{X_k}|\le 1$ and $\left|\varphi_{P,\sigma^2}\right|\le 1$ (as they are characteristic function of random variables), the preceding bound reduces to

$$\left|\varphi_{X_k}(t)-\overline{\varphi}_{P,\sigma^2}(t)\right| \le \left|\log\varphi_{X_k}(t)-\log\overline{\varphi}_{P,\sigma^2}(t)\right|.$$

From combining all the preceding bounds, it follows that there exists a universal constant $D>0$ such that

$$\|\varphi_{X_k}-\overline{\varphi}_{P,\sigma^2}\|_{\mathbb{B}(a_1T_n)}\le DT_n^2\|(\log\varphi_{X_k})''(.)-(\log\overline{\varphi}_{P,\sigma^2})''(.)\|_{\mathbb{B}(a_1T_n)}.$$

It follows that for any $C'>0$, there exists a $M>0$ such that

$$\begin{aligned}&\overline{\nu}_{\alpha,G}\left(\|\varphi_{X_k}-\overline{\varphi}_{P,\sigma^2}\|^2_{\mathbb{B}(a_1T_n)}\le M^2T_n^{4\zeta+5-d}\epsilon_n^2, \left|\sigma^{-2}\right|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}\,\middle|\,\mathcal{Z}_n, T_n\right) \\ &\ge 1-\exp\left(-C'n\epsilon_n^2\right)\end{aligned}$$

holds with $\mathbb{P}$ probability approaching 1.

It remains to examine the bias from truncating the L^2 norm to the set $\mathbb{B}(a_1T_n)$. Suppose $\left|\sigma^{-2}\right|\le M^2\sigma_n^2(n\epsilon_n^2)^{1/\kappa}$ holds. It follows that there exists a $c>0$ for which $\sigma^2\ge c(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}$ holds. From the definition of σ_n^2, we have $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2}\asymp\log n$.

It follows that there exists a universal constant $C > 0$ such that

$$\begin{aligned}\|(\varphi_{X_k} - \overline{\varphi}_{P,\sigma^2})\mathbb{1}\{|t| > a_1T_n\}\|_{L^2}^2 &\leq 2\|\varphi_{X_k}\mathbb{1}\{|t| > a_1T_n\}\|_{L^2}^2 + 2\|\overline{\varphi}_{P,\sigma^2}\mathbb{1}\{\|t\|_\infty > a_1T_n\}\|_{L^2}^2 \\ &\leq 2\int_{|t|>a_1T_n} |\varphi_{X_k}(t)|^2\,dt + 2\int_{|t|>a_1T_n} e^{-t^2\sigma^2}dt \\ &\leq 2\int_{|t|>a_1T_n} |\varphi_{X_k}(t)|^2\,dt + 2\int_{|t|>a_1T_n} e^{-ct^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}}dt \\ &\leq D\Big[T_n^{-2p} + \sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}e^{-CT_n^2\sigma_n^{-2}(n\epsilon_n^2)^{-1/\kappa}}T_n^{d-2}\Big].\end{aligned}$$

Since $T_n^2(n\epsilon_n^2)^{-1/\kappa}\sigma_n^{-2} \asymp (\log n)(\log\log n)$, the preceding bound reduces to DT_n^{-2p}. From combining the preceding bounds, it follows that for every $C' > 0$, there exists a $M > 0$ such that

$$\nu_{\alpha,G}\Big(\|\varphi_X - \varphi_{P,\Sigma}\|_{L^2} \leq M(T_n^{2\zeta+(5-d)/2}\epsilon_n + T_n^{-p})\,\Big|\,\mathcal{Z}_n, T_n\Big) \geq 1 - \exp\big(-C'n\epsilon_n^2\big).$$

holds with $\mathbb{P}$ probability approaching 1. The claim follows from observing that

$$T_n^{4\zeta+(5-d)}\epsilon_n^2 \asymp \frac{(\log n)^{\lambda+2\zeta+5/2}}{n^{\frac{2s+d-5}{2(s+2\zeta)+d}}}.$$

□

Bibliography

Abowd, John M and David Card (1989): "On the covariance structure of earnings and hours changes," *Econometrica*, 57 (2), 411–445.

Agostinelli, Francesco and Matthew Wiswall (2016): "Estimating the technology of children's skill formation," Tech. rep., National Bureau of Economic Research.

Altonji, Joseph G, Anthony A Smith Jr, and Ivan Vidangos (2013): "Modeling earnings dynamics," *Econometrica*, 81 (4), 1395–1454.

Andrews, Isaiah and Anna Mikusheva (2022): "Optimal decision rules for weak GMM," *Econometrica*, 90 (2), 715–748.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme (2017): "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85 (3), 693–734.

Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina (2020a): "Estimating the production function for human capital: results from a randomized controlled trial in Colombia," *American Economic Review*, 110 (1), 48–85.

Attanasio, Orazio, Costas Meghir, and Emily Nix (2020b): "Human capital development and parental investment in India," *The Review of Economic Studies*, 87 (6), 2511–2541.

Bonhomme, Stéphane and Jean-Marc Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics," *The Review of Economic Studies*, 77 (2), 491–533.

Butucea, C and AB Tsybakov (2008a): "Sharp optimality for density deconvolution with dominating bias, II," *SIAM Theory of Probability and its Applications*, 52 (2), 237–249.

Butucea, Cristina and Alexandre B Tsybakov (2008b): "Sharp optimality in density deconvolution with dominating bias. I," *Theory of Probability & Its Applications*, 52 (1), 24–39.

Carrasco, Marine and Jean-Pierre Florens (2000): "Generalization of GMM to a continuum of moment conditions," *Econometric Theory*, 16 (6), 797–834.

Chen, Xiaohong, Timothy M Christensen, and Elie Tamer (2018): "Monte Carlo confidence sets for identified sets," *Econometrica*, 86 (6), 1965–2018.

Chen, Xiaohong, Han Hong, and Denis Nekipelov (2011): "Nonlinear models of measurement errors," *Journal of Economic Literature*, 49 (4), 901–937.

Chen, Xiaohong, Han Hong, and Elie Tamer (2005): "Measurement error models with auxiliary data," *The Review of Economic Studies*, 72 (2), 343–366.

Chernozhukov, Victor and Han Hong (2003): "An MCMC approach to classical estimation," *Journal of econometrics*, 115 (2), 293–346.

Cooley, James, Peter Lewis, and Peter Welch (1967): "Application of the fast Fourier transform to computation of Fourier integrals, Fourier series, and convolution integrals," *IEEE Transactions on Audio and Electroacoustics*, 15 (2), 79–84.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78 (3), 883–931.

Dattner, Itai, Markus Reiss, and Mathias Trabs (2016): "Adaptive quantile estimation in deconvolution with unknown error distribution," *Bernoulli*, 22 (1), 143.

Delaigle, Aurore and Irène Gijbels (2004): "Practical bandwidth selection in deconvolution kernel density estimation," *Computational statistics & data analysis*, 45 (2), 249–267.

Delaigle, Aurore, Peter Hall, and Alexander Meister (2008): "On Deconvolution with Repeated Measurements," *The Annals of Statistics*, 665–685.

Donnet, Sophie, Vincent Rivoirard, Judith Rousseau, and Catia Scricciolo (2018): "Posterior concentration rates for empirical Bayes procedures with applications to Dirichlet process mixtures," *Bernoulli*, 24 (1), 231–256.

Evans, Lawrence C (2022): *Partial differential equations*, vol. 19, American Mathematical Society.

Evdokimov, Kirill and Halbert White (2012): "Some extensions of a lemma of Kotlarski," *Econometric Theory*, 28 (4), 925–932.

Fan, Jianqing (1991a): "Global behavior of deconvolution kernel estimates," *Statistica Sinica*, 541–551.

——— (1991b): "On the optimal rates of convergence for nonparametric deconvolution problems," *The Annals of Statistics*, 1257–1272.

Gallant, A Ronald, Han Hong, Michael P Leung, and Jessie Li (2022): "Constrained estimation using penalization and MCMC," *Journal of Econometrics*, 228 (1), 85–106.

Gautier, Eric and Yuichi Kitamura (2013): "Nonparametric estimation in random coefficients binary choice models," *Econometrica*, 81 (2), 581–607.

Gelman, Andrew (2006): "Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper)," *Bayesian Analysis*, 1 (3), 515–534.

Genovese, Christopher R and Larry Wasserman (2000): "Rates of convergence for the Gaussian mixture sieve," *The Annals of Statistics*, 28 (4), 1105–1127.

Geweke, John and Michael Keane (2000): "An empirical analysis of earnings dynamics among men in the PSID: 1968–1989," *Journal of econometrics*, 96 (2), 293–356.

Ghosal, Subhashis, Jayanta K Ghosh, and Aad W Van Der Vaart (2000): "Convergence rates of posterior distributions," *Annals of Statistics*, 500–531.

Ghosal, Subhashis and Aad Van Der Vaart (2001): "Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities," *The Annals of Statistics*, 29 (5), 1233–1263.

——— (2007): "Posterior convergence rates of Dirichlet mixtures at smooth densities," .

Ghosal, Subhashis and Aad Van der Vaart (2017): *Fundamentals of nonparametric Bayesian Inference*, Cambridge University Press.

Giné, Evarist and Richard Nickl (2021): *Mathematical foundations of infinite-dimensional statistical models*, Cambridge university press.

Haile, Philip A and Yuichi Kitamura (2019): "Unobserved heterogeneity in auctions," *The Econometrics Journal*, 22 (1), C1–C19.

Hall, Robert E and Frederic S Mishkin (1982): "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households," *Econometrica*, 50 (2), 461–481.

Hausman, Jerry A, Whitney K Newey, Hidehiko Ichimura, and James L Powell (1991): "Identification and estimation of polynomial errors-in-variables models," *Journal of Econometrics*, 50 (3), 273–295.

Heckman, James and Burton Singer (1984): "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica: Journal of the Econometric Society*, 271–320.

Hoderlein, Stefan, Jussi Klemelä, and Enno Mammen (2010): "Analyzing the random coefficient model nonparametrically," *Econometric Theory*, 26 (3), 804–837.

Hoffman, Matthew D, Andrew Gelman, et al. (2014): "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo." *J. Mach. Learn. Res.*, 15 (1), 1593–1623.

Horowitz, Joel L and Marianthi Markatou (1996): "Semiparametric estimation of regression models for panel data," *The Review of Economic Studies*, 63 (1), 145–168.

Hu, Yingyao (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144 (1), 27–61.

Hu, Yingyao and Susanne M Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76 (1), 195–216.

Kato, Kengo (2013): "Quasi-Bayesian analysis of nonparametric instrumental variables models," *The Annals of Statistics*, 41 (5), 2359–2390.

Kato, Kengo and Yuya Sasaki (2018): "Uniform confidence bands in deconvolution with unknown error distribution," *Journal of Econometrics*, 207 (1), 129–161.

Kato, Kengo, Yuya Sasaki, and Takuya Ura (2021): "Robust inference in deconvolution," *Quantitative Economics*, 12 (1), 109–142.

Kim, Jae-Young (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107 (1-2), 175–193.

Kotlarski, Ignacy (1967): "On characterizing the gamma and the normal distribution," *Pacific Journal of Mathematics*, 20 (1), 69–76.

Krasnokutskaya, Elena (2011): "Identification and estimation of auction models with unobserved heterogeneity," *The Review of Economic Studies*, 78 (1), 293–327.

Kruijer, Willem, Judith Rousseau, and Aad van der Vaart (2010): "Adaptive Bayesian density estimation with location-scale mixtures," *Electronic Journal of Statistics*, 4, 1225–1257.

Kurisu, Daisuke and Taisuke Otsu (2022): "On the uniform convergence of deconvolution estimators from repeated measurements," *Econometric Theory*, 38 (1), 172–193.

Lenk, Peter J (1991): "Towards a practicable Bayesian nonparametric density estimator," *Biometrika*, 78 (3), 531–543.

Lewandowski, Daniel, Dorota Kurowicka, and Harry Joe (2009): "Generating random correlation matrices based on vines and extended onion method," *Journal of multivariate analysis*, 100 (9), 1989–2001.

Li, Tong (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110 (1), 1–26.

Li, Tong and Quang Vuong (1998): "Nonparametric estimation of the measurement error model using multiple indicators," *Journal of Multivariate Analysis*, 65 (2), 139–165.

Liao, Yuan and Wenxin Jiang (2011): "Posterior consistency of nonparametric conditional moment restricted models," *Annals of statistics*, 39 (6), 3003–3031.

Lounici, Karim and Richard Nickl (2011): "Global uniform risk bounds for wavelet deconvolution estimators," *The Annals of Statistics*, 39 (1), 201–231.

Maggi, Francesco and Cédric Villani (2005): "Balls have the worst best Sobolev inequalities," *The Journal of Geometric Analysis*, 15, 83–121.

Maggi, F and Cedric Villani (2008): "Balls have the worst best Sobolev inequalities. Part II: variants and extensions," *Calculus of Variations and Partial Differential Equations*, 31, 47–74.

Newey, Whitney K (2001): "Flexible simulated moment estimation of nonlinear errors-in-variables models," *Review of Economics and statistics*, 83 (4), 616–627.

Rousseau, Judith (2010): "Rates of convergence for the posterior distributions of mixtures of betas and adaptive nonparamatric estimation of the density," *Annals of Statistics*, 38 (1), 146–180.

Rousseau, Judith and Catia Scricciolo (2023): "Wasserstein convergence in Bayesian and frequentist deconvolution models," *arXiv preprint arXiv:2309.15300*.

Schennach, Susanne (2016): "Recent advances in the measurement error literature," *Annual Review of Economics*, 8, 341–377.

——— (2022): "Measurement systems," *Journal of Economic Literature*, 60 (4), 1223–1263.

Schennach, Susanne M (2004a): "Estimation of nonlinear models with measurement error," *Econometrica*, 72 (1), 33–75.

——— (2004b): "Nonparametric regression in the presence of measurement error," *Econometric Theory*, 20 (6), 1046–1093.

——— (2007): "Instrumental variable estimation of nonlinear errors-in-variables models," *Econometrica*, 75 (1), 201–239.

Shen, Weining, Surya T Tokdar, and Subhashis Ghosal (2013): "Adaptive Bayesian multivariate density estimation with Dirichlet mixtures," *Biometrika*, 100 (3), 623–640.

Wang, Liqun and Cheng Hsiao (2011): "Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models," *Journal of Econometrics*, 165 (1), 30–44.