The Annals of Statistics
2023, Vol. 51, No. 3, 1376–1400
https://doi.org/10.1214/23-AOS2298

COVERAGE OF CREDIBLE INTERVALS IN BAYESIAN MULTIVARIATE ISOTONIC REGRESSION

BY KANG WANG[a] AND SUBHASHIS GHOSAL[b]

Department of Statistics, North Carolina State University, [a]*kwang22@ncsu.edu,* [b]*sghosal@stat.ncsu.edu*

We consider the nonparametric multivariate isotonic regression problem, where the regression function is assumed to be nondecreasing with respect to each predictor. Our goal is to construct a Bayesian credible interval for the function value at a given interior point with assured limiting frequentist coverage. A natural prior on the regression function is given by a random step function with a suitable prior on increasing step-heights, but the resulting posterior distribution is hard to analyze theoretically due to the complicated order restriction on the coefficients. We instead put a prior on unrestricted step-functions, but make inference using the induced posterior measure by an "immersion map" from the space of unrestricted functions to that of multivariate monotone functions. This allows for maintaining the natural conjugacy for posterior sampling. A natural immersion map to use is a projection with respect to a distance function, but in the present context, a block isotonization map is found to be more useful. The approach of using the induced "immersion posterior" measure instead of the original posterior to make inference provides a useful extension of the Bayesian paradigm, particularly helpful when the model space is restricted by some complex relations. We establish a key weak convergence result for the posterior distribution of the function at a point in terms of some functional of a multiindexed Gaussian process that leads to an expression for the limiting coverage of the Bayesian credible interval. Analogous to a recent result for univariate monotone functions, we find that the limiting coverage is slightly higher than the credibility, the opposite of a phenomenon observed in smoothing problems. Interestingly, the relation between credibility and limiting coverage does not involve any unknown parameter. Hence, by a recalibration procedure, we can get a predetermined asymptotic coverage by choosing a suitable credibility level smaller than the targeted coverage, and thus also shorten the credible intervals.

1. Introduction. Nonparametric inference often involves a regression function or a density function in modeling. Commonly, a smoothness assumption on a function of interest is imposed, but in some applications, qualitative information, such as monotonicity, unimodality and convexity, on the shape of the function may be available. This leads to a control on the complexity of the function space analogous to what a smoothness assumption does, allowing convergence without requiring the latter. Monotonicity is the simplest and the most extensively studied shape restriction, especially in the univariate case. In regression analysis, this problem is commonly referred to as isotonic regression when the conditional mean function of the response variable is assumed to be nondecreasing. Starting from the early works on monotone shape-restricted problems, such as [1, 10], research on non-Bayesian approaches, mainly on the least squares estimator (LSE) and the nonparametric maximum likelihood estimator (MLE), has been fruitful; see [4, 31, 32, 46]. Assuming a nonzero derivative, the pointwise asymptotic distribution of the MLE or the LSE turns out to be the rescaled Chernoff distribution, that is, the minimizer of a quadratically drifted standard two-sided

Received February 2022; revised May 2023.
MSC2020 subject classifications. Primary 62G08; secondary 62F15, 62G05, 62G20.
Key words and phrases. Isotonic regression, credible interval, limiting coverage, Gaussian process, Block isotonization, immersion posterior.

Brownian motion [11, 32, 45, 56]. The same limiting distribution can also be found in other problems where monotonicity is implied, such as the monotone hazard rate estimation with randomly right-censored observations in survival analysis [39, 40], and many statistical inverse problems, including the current status model and deconvolution problems [33]. Global properties of shape-restricted estimators were also studied extensively [32, 42]. The convergence rates and limiting distributional behaviors of $\mathbb{L}_p$- and $\mathbb{L}_\infty$-distance between monotone shape-restricted estimators and the true function were investigated by [26, 27]. Nonasymptotic risk bounds for the LSE under a monotone shape restriction were derived by [5, 18, 57]. Testing for monotonicity was addressed in [25, 29, 30, 35].

The Bayesian approach to shape-restricted problems was also explored, albeit to a lesser extent. Neelon and Dunson [44] applied piecewise linear structures to the regression function, and put monotone restrictions on the priors for the slope values. Cai and Dunson [12] proposed a linear spline model and added an initial Markov random field prior to the coefficients, and then the monotone constraint was incorporated by considering the relation between the nonnegative slopes and coefficients. Wang [55] adopted the free-knot cubic regression spline model, converted the shape restriction to the coefficients and then projected the unconstrained coefficients with conventional priors to the target set, inducing the constrained priors. Shively et al. [52] also used Bayesian splines with constrained normal priors on the coefficients to comply with the monotone shape restriction. Lin and Dunson [43] addressed this problem by using a Gaussian process prior and projected unconstrained posterior samples to the monotone function class by a min-max formula. Chakraborty and Ghosal [14, 15, 17] also used the idea of projection-posterior, making the investigation of frequentist limiting coverage of credible sets possible. Salomond [48] used a mixture representation of a nonincreasing density on $[0, \infty)$ and obtained the nearly minimax posterior contraction rate for both a Dirichlet process and a finite mixture prior on the mixing distribution. Bayesian tests for monotonicity were developed by [15, 17, 49]. A Bayesian credible interval with assured frequentist coverage for a monotone regression quantile, and an accelerated rate of contraction for it using a two-stage sampling were obtained by [16].

Multivariate monotone function estimation was also studied in the literature. Non-Bayesian works focused on the construction of the LSE with respect to various partial orderings on the domain; see [4, 46]. Only the consistency of the isotonic estimator was known until a recent rise in interest in multivariate shape-restricted problems. In a multivariate isotonic regression model, the $\mathbb{L}_2$-risk of the LSE, respectively for $d = 2$ and for a general dimension d was studied by [19, 37]. They found that the LSE achieved the optimal minimax rate up to logarithmic factors, and adapted to the parametric rate for a piecewise constant true regression function only when $d \leq 2$. Han [36] confirmed that the global empirical risk minimizer is indeed rate-optimal in some set structured models even with rapidly diverging entropy integral, and thus gave a simpler proof for the optimal convergence rate of the LSE in the multivariate isotonic regression. Deng and Zhang [22] investigated a block-estimator proposed by [28] and obtained an $\mathbb{L}_q$-risk bound. This is minimax rate optimal and adapts to the parametric rate up to a logarithmic factor when the true regression function is piecewise constant. Pointwise distributional limits for the block-estimator were obtained by [38], which lays the foundation for subsequent inference.

The Bayesian approach to multivariate isotonic regression is much less developed. Saarele and Arjas [47] proposed a Bayesian approach to this problem based on marked point processes and resulted in piecewise constant realizations of the regression function. Lin and Dunson [43] mentioned that the method of projecting the Gaussian process posterior can also be applied in the regression surface case. Nonetheless, the theoretical studies presented in those works are either lacking or inadequate.

The construction of confidence regions in function estimation problems was studied by many authors, mostly in smoothness regimes. For shape-restricted problems, confidence regions using limit theory were constructed by [13, 23, 24, 50]. The natural bootstrap method does not lead to a valid confidence interval for the function value at a point, but a modified bootstrap method works [41, 51]. Confidence intervals by inverting the acceptance region of a likelihood ratio test for the value of a monotone function at a point were obtained by [2, 3, 34]. This approach has the advantage that no additional nuisance parameters need to be estimated and plugged into the limit distribution. Deng et al. [21] constructed a confidence interval for multivariate monotone regression from a pivotal limit result for the block-estimator of [22] relying on the limiting distributional theory by [38].

In this paper, we consider a Bayesian approach to the multivariate monotone regression problem. Our objective is to construct a Bayesian credible interval for the function value at an interior point and study its frequentist coverage. As in the univariate problem studied by [14], we ignore the shape restriction at the prior stage and put a prior on random step functions without an order restriction, retaining the posterior conjugacy. We then apply corrections to posterior samples using monotone mapping. This induces a posterior distribution supported on the space of multivariate monotone functions, by which we can obtain credible intervals. However, contrasting with the univariate case, the projection-posterior acquired through the minimization of the empirical $\mathbb{L}_2$-metric does not possess a limiting distribution. This is due to the fact that the partial sum process, which characterizes the empirical $\mathbb{L}_2$-projection, is not tight in the limit; see [38]. We also note that the non-Bayesian confidence interval constructed in [21] is also not obtained by distance minimization but by a block max-min procedure. We instead use a map related to the block max-min operation to enforce multivariate monotonicity on posterior samples. As the map immerses a general function into the space of monotone functions, such a map will be referred to as an immersion map, and the induced posterior will be termed an immersion posterior.

The rest of the paper is organized as follows. In the next section, we introduce the notion of an immersion posterior distribution, which will be used to address the problem under study, and is very helpful for similar Bayesian problems with complicated restrictions on the parameter space. In Section 3, we introduce the model and assumptions, construct a prior distribution, and obtain the immersion posterior used to make an inference. Our main results are presented in Section 4. We derive the weak limit of the scaled and centered pointwise immersion posterior distribution. Based on the limit theory, we compute the asymptotic frequentist coverage of credible intervals. Numerical results are present in Section 5. We include all the proofs of the main theorems in Section 6. Proofs of all auxiliary lemmas and propositions are provided in the Supplementary Material.

2. Immersion posterior. Consider a general statistical model with observation $X \sim P_\theta$, where $\theta \in \Theta_0$. Suppose that the parameter space Θ_0 is a complicated subset of a larger, but simpler to represent, set Θ. This is often the case for shape-restricted inference, where structural constraints, such as monotonicity, convexity and log-concavity, are imposed on a regression function or a density function. In differential equation models, the parameter space is implicitly described as the set of solutions of a system of ordinary or partial differential equations involving some unknown parameters. In a vector autoregressive process, the set of autoregression coefficients leading to stationary processes may be the parameter space of interest, but it is described by many complicated constraints. Because of the complicated restrictions on Θ_0, a prior for θ with support on Θ_0 may be hard to construct, and the corresponding posterior may be difficult to compute. More importantly, the corresponding posterior may be hard to analyze from a frequentist perspective. This may be particularly important for studying delicate properties such as the limiting coverage of a Bayesian credible region.

Often, the distribution P_θ makes sense for any $\theta \in \Theta$, so that Θ_0 can be embedded in Θ, keeping the statistical problem meaningful. For shape-restricted models, this becomes the standard nonparametric regression or the density estimation problem. A differential equation model also embeds in a nonparametric regression model. A prior distribution Π may be specified on Θ, initially disregarding the restriction of θ to Θ_0. This is typically a standard problem, and often a conjugate prior distribution can be identified. The resulting posterior distribution $\Pi(\cdot|X)$ thus resides in the whole of Θ, and hence is not appropriate to make an inference about θ, which is known to live in Θ_0. The requirement can be met by considering the random measure induced by a mapping ι from Θ to Θ_0, in that, we consider the random measure $\Pi^*(B|X) = \Pi(\iota(\theta) \in B|X)$ to make an inference on θ. The map ι immerses θ into the desirable space Θ_0, and hence will be referred to as the *immersion map*. The induced posterior Π^* will be referred to as the *immersion posterior*. This provides an extension of the Bayesian paradigm since the identity map as the immersion map for the situation $\Theta_0 = \Theta$ reduces the immersion posterior to the classical Bayesian posterior.

The approach has been successfully used in several works including [14–17, 43] for shape-restricted problems, and by [6–9] for differential equation models. These authors used a projection map $\mathfrak{p}$ obtained by minimizing a certain distance from the posterior sample to the restricted space, and the resulting induced random measure is called the projection posterior distribution. The projection map $\mathfrak{p}$ satisfies the appealing property $\mathfrak{p}(\theta) = \theta$ for all $\theta \in \Theta_0$.

While a projection map with respect to an appropriate distance is a natural choice for an immersion map, the restriction to a projection map is unnecessary for the concept to be used. Depending on the aspect to be studied, there may not be a natural distance associated with it. This happens, for instance, if we are interested in studying the posterior distribution of the function value at a given point. It is also not necessary for the immersion map ι to satisfy $\iota(\theta) = \theta$ for all $\theta \in \Theta_0$. Neither Θ_0 needs to be a subset of Θ, nor the immersion map needs to be defined all over Θ. All that is needed is that an alternative parameter space Θ exists where the model distribution P_θ makes sense, a prior Π can be put on Θ such that the posterior distribution can be computed relatively easily, and the random distribution induced by a map ι from the support of the posterior distribution $\Pi(\cdot|X)$ to Θ_0 can be analyzed theoretically to establish some desirable properties. In most situations, the family of measures $\{P_\theta : \theta \in \Theta\}$ is dominated, so the support of the posterior distribution $\Pi(\cdot|X)$ is contained in the support of the prior distribution Π. The immersion map may be allowed to depend on the sample size like a prior distribution may be allowed to depend on the sample size. Even dependence of ι on the data X may be allowed. Although there is no uniqueness in the choice of the immersion map, the main purpose is to increase flexibility in the posterior measure to achieve a targeted asymptotic frequentist property, such as coverage of a credible region. A choice of an immersion map is therefore guided by a desirable frequentist property. Even if Θ_0 and Θ coincide, the flexibility of the immersion posterior may be helpful to satisfy a desirable convergence property of the immersion posterior that the classical Bayesian posterior may lack.

In many applications, the prior distribution may be actually a sequence of prior distributions specified through a sieve indexed by a discrete variable $J = J_n$ depending on the sample size n. Let Θ_J stand for the sieve (typically a finite-dimensional subset of Θ) and Π_J stand for the prior at that stage concentrated on Θ_J. Then the computation of the posterior in the unrestricted space reduces to a finite-dimensional computation, often also aided by posterior conjugacy. It is then typical that the immersion map ι on Θ_J has the range in $\Theta_0 \cap \Theta_J$ so that the computation of the immersion posterior involves finite-dimensional computations only. Most examples from the existing literature, as well as the method used in this paper, fall in this setting.

3. Notation, model, prior and posterior distribution.

3.1. *Notation*. We summarize the notation we shall use in this paper. The notation $\mathbb{R}$, $\mathbb{N}$ and $\mathbb{Z}$ will stand for the real line, the set of natural numbers and the set of all integers, respectively. The positive half-line with and without 0, and the set of nonnegative integers are respectively denoted by $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{>0}$ and $\mathbb{Z}_{\geq 0}$. Bold Latin or Greek letters will be used to indicate column vectors and the nonbold letter with a subscript will denote a coordinate of the corresponding vector. For example, a_i is the ith coordinate of $\boldsymbol{a} \in \mathbb{R}^d$. Let $\mathbf{1}$ denote the d-dimensional all-one column vector and $\mathbf{0}$ the all-zero column vector. Let $\boldsymbol{A}^{\mathrm{T}}$ denote the transpose of a matrix or a vector $\boldsymbol{A}$. For an arbitrary set A, the indicator function will be denoted by $\mathbb{1}_A(\cdot)$, and $\#A$ will denote the cardinality of a finite set A. Let $\lceil a \rceil$ stand for the smallest integer greater than or equal to a real number a. The symbol $\lesssim$ will stand for an inequality up to an unimportant constant multiple. For two positive real sequences a_n and b_n, we also use $a_n \ll b_n$, or equivalently, $b_n \gg a_n$ if $a_n = o(b_n)$. For $a, b \in \mathbb{R}$, let $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$, let $\boldsymbol{a} \wedge \boldsymbol{b} = (a_1 \wedge b_1, \ldots, a_d \wedge b_d)^{\mathrm{T}}$, $\boldsymbol{a} \vee \boldsymbol{b} = (a_1 \vee b_1, \ldots, a_d \vee b_d)^{\mathrm{T}}$ and the pointwise product $\boldsymbol{a} \circ \boldsymbol{b} = (a_1 b_1, \ldots, a_d b_d)^{\mathrm{T}}$. For a vector $\boldsymbol{a} \in \mathbb{R}^d$, the Euclidean and the maximum norms are respectively denoted by $\|\boldsymbol{a}\|$ and $\|\boldsymbol{a}\|_\infty = \max\{|a_k| : 1 \leq k \leq d\}$. Let $[\boldsymbol{j}_1 : \boldsymbol{j}_2] = \{\boldsymbol{j} \in \mathbb{Z}^d : j_{1,k} \leq j_k \leq j_{2,k}, \text{ for all } 1 \leq k \leq d\}$ stand for the lattice with boundaries $\boldsymbol{j}_1, \boldsymbol{j}_2 \in \mathbb{Z}^d$.

For a multivariate function $f : \mathbb{R}^d \to \mathbb{R}$, let $\partial_k^l f(\boldsymbol{x}) = \partial^l f(\boldsymbol{x}) / \partial x_k^l$ for $k \in \{1, \ldots, d\}$ and $l \in \mathbb{Z}_{\geq 0}$ at a suitable point $\boldsymbol{x} \in \mathbb{R}^d$. For a multiple index $\boldsymbol{l} = (l_1, \ldots, l_d)^{\mathrm{T}} \in \mathbb{Z}_{\geq 0}^d$, we use $\partial^{\boldsymbol{l}} = \partial_1^{l_1} \cdots \partial_d^{l_d}$, $\boldsymbol{l}! = l_1! \cdots l_d!$ and $\boldsymbol{x}^{\boldsymbol{l}} = x_1^{l_1} \cdots x_d^{l_d}$. We adopt the coordinatewise partial ordering on $\mathbb{R}^d$, that is, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, $\boldsymbol{x} \preceq \boldsymbol{y}$ if and only if $x_k \leq y_k$ for all $1 \leq k \leq d$. We say that a function f on $\mathbb{R}^d$ is multivariate monotone if $f(\boldsymbol{x}) \leq f(\boldsymbol{y})$ for all $\boldsymbol{x} \preceq \boldsymbol{y}$. The class of all multivariate monotone functions on $[0, 1]^d$ will be denoted by $\mathcal{M}$. Let $\mathbb{L}_p[\boldsymbol{a}, \boldsymbol{b}]$, $1 \leq p \leq \infty$, stand for the Lebesgue $\mathbb{L}_p$-space on a multivariate interval $[\boldsymbol{a}, \boldsymbol{b}]$. Convergence in probability under a measure P is denoted by $\to_{\mathrm{P}}$. Distributional equality will be denoted by $=_d$ and weak convergence by $\rightsquigarrow$.

3.2. *Model*. We observe n independent and identically distributed random samples $\mathbb{D}_n = ((\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n))$ from the nonparametric multiple regression model,

$$Y = f(\boldsymbol{X}) + \varepsilon, \tag{3.1}$$

where Y is the response variable, $\boldsymbol{X}$ is a d-dimensional predictor and ε is a random error with mean 0 and finite variance σ^2, independent of $\boldsymbol{X}$. Instead of assuming any global smoothness condition on f, we assume that f is a multivariate monotone function. To construct the likelihood function, we assume that ε is normally distributed, but the actual data-generating process need not be so.

The first assumption is about the local regularity of the true regression function f_0 near a point of interest $\boldsymbol{x}_0$. This assumption, as in [38], is an essential ingredient to establish the limiting distribution.

ASSUMPTION 1. Let $f_0 \in \mathcal{M}$. For $\boldsymbol{x}_0 \in (0, 1)^d$ and $1 \leq k \leq d$, let β_k be the order of the first nonzero derivative of f at $\boldsymbol{x}_0$ along the kth coordinate, that is, $\beta_k = \min_{l \geq 1}\{l : \partial_k^l f_0(\boldsymbol{x}_0) \neq 0\}$ and $\beta_k = \infty$ if $\partial_k^l f_0(\boldsymbol{x}_0) = 0$ for all $l \geq 1$. Without loss of generality, we may assume that f_0 depends on its first s arguments locally at $\boldsymbol{x}_0$, that is, $1 \leq \beta_1, \ldots, \beta_s < \infty$, and that $\beta_{s+1} = \cdots = \beta_d = \infty$ for some $0 \leq s \leq d$. Define an index set $L = \{\boldsymbol{l} : 0 < \sum_{k=1}^s l_k / \beta_k \leq 1 \text{ and } l_k = 0, \text{ for } k = s + 1, \ldots, d\}$. For a positive sequence

$\omega_n \downarrow 0$, set $\boldsymbol{r}_n = (\omega_n^{1/\beta_1}, \ldots, \omega_n^{1/\beta_s}, 1, \ldots, 1)^{\mathrm{T}}$. For any $t > 0$,

$$\lim_{\omega_n \downarrow 0} \omega_n^{-1} \sup_{\substack{\boldsymbol{x} \in [0,1]^d, \\ |x_k - x_{0,k}| \le t r_{n,k}, \\ 1 \le k \le d}} \left| f_0(\boldsymbol{x}) - f_0(\boldsymbol{x}_0) - \sum_{\boldsymbol{l} \in L} \frac{\partial^{\boldsymbol{l}} f_0(\boldsymbol{x}_0)}{\boldsymbol{l}!} (\boldsymbol{x} - \boldsymbol{x}_0)^{\boldsymbol{l}} \right| = 0. \tag{3.2}$$

Assumption 1 takes into account varying convergence rates across different coordinates, according to their respective smoothness levels. Each term in the expansion contributes toward approximation rates larger than or equal to ω_n. Let

$$L_0 = \left\{ \boldsymbol{l} : 0 < \sum_{k=1}^{s} l_k/\beta_k < 1 \text{ and } l_k = 0 \text{ for } k = s+1, \ldots, d \right\}, \tag{3.3}$$

$$L^* = \left\{ \boldsymbol{l} : \sum_{k=1}^{s} l_k/\beta_k = 1 \text{ and } l_k = 0 \text{ for } k = s+1, \ldots, d \right\}. \tag{3.4}$$

Under Assumption 1, a unique feature for functions in $\mathcal{M}$ is that the derivatives of order $\boldsymbol{l} \in L_0$ are zero (see Lemma 1 of [38]). Only those derivatives corresponding to the index set L^* can be nonzero. Thus, the nonzero terms in the expansion of (3.2) contribute the same approximation rate ω_n. However, Assumption 1 cannot exclude the nonzero mixed derivatives. Additional assumptions will be needed when we want to eliminate the mixed derivative terms.

Next, we make the following assumption on the distributions of the covariate $\boldsymbol{X}$ and the error ε from the data generating process (3.1).

ASSUMPTION 2. The covariate $\boldsymbol{X}$ has a density g such that $a_1 \le g(\boldsymbol{x}) \le a_2$ for all $\boldsymbol{x} \in [0,1]^d$ and $0 < a_1 \le a_2 < \infty$. Suppose g is continuous in a neighborhood of the set $\{(x_{0,1}, \ldots, x_{0,s}, x_{s+1}, \ldots, x_d) : x_k \in [0,1], \text{ for } s+1 \le k \le d\}$. The random error ε, with mean 0 and variance σ_0^2, has a finite $2(\sum_{k=1}^{s} \beta_k^{-1} + 1)$th moment.

3.3. *Prior.* We put a prior distribution on f through a sieve of piecewise constant functions with gradually refining intervals of constancy, forming a partition of $[0,1]^d$. For $\boldsymbol{J} \in \mathbb{Z}_{>0}^d$, let $I_{\boldsymbol{j}} = \prod_{k=1}^{d} ((j_k - 1)/J_k, j_k/J_k]$ be a hyperrectangle in $[0,1]^d$, indexed by a d-dimensional vector $\boldsymbol{j}$, for $\boldsymbol{j} \in [\boldsymbol{1} : \boldsymbol{J}] \backslash \{\boldsymbol{1}\}$ and $I_{\boldsymbol{1}} = \prod_{k=1}^{d} [0, 1/J_k]$. Then $\{I_{\boldsymbol{j}}\}_{\boldsymbol{j} \in [\boldsymbol{1}:\boldsymbol{J}]}$ forms a partition of $[0,1]^d$. We define a class of piecewise constant functions $\mathcal{K}_{\boldsymbol{J}} := \{f = \sum_{\boldsymbol{j} \in [\boldsymbol{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}} : \theta_{\boldsymbol{j}} \in \mathbb{R}\}$. As we follow the immersion posterior approach, we do not initially impose the order restriction. A prior is imposed on $f = \sum_{\boldsymbol{j} \in [\boldsymbol{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$ in $\mathcal{K}_{\boldsymbol{J}}$ by giving independent Gaussian priors to $\theta_{\boldsymbol{j}}$, namely,

$$\theta_{\boldsymbol{j}} \sim \mathrm{N}(\zeta_{\boldsymbol{j}}, \sigma^2 \lambda_{\boldsymbol{j}}^2) \quad \text{independently for all } \boldsymbol{j} \in [\boldsymbol{1} : \boldsymbol{J}], \tag{3.5}$$

where $\max_{\boldsymbol{j}} |\zeta_{\boldsymbol{j}}| < \infty$ and $\min_{\boldsymbol{j}} \lambda_{\boldsymbol{j}}^2 \ge b > 0$.

The values of the prior parameters, $\zeta_{\boldsymbol{j}}$ and $\lambda_{\boldsymbol{j}}$, will not affect our asymptotic results. However, in practice, when very little prior information is available, it is sensible to choose $\zeta_{\boldsymbol{j}} = 0$ and $\lambda_{\boldsymbol{j}}$ large for all $\boldsymbol{j}$.

3.4. *Posterior distribution.* We use the Gaussian distribution

$$Y_i \sim \mathrm{N}\Big(\sum_{\boldsymbol{j} \in [\boldsymbol{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}\{\boldsymbol{X}_i \in I_{\boldsymbol{j}}\}, \sigma^2 \Big), \tag{3.6}$$

which leads to, in the unrestricted parameter space, a Gaussian joint likelihood for $(\theta_{\boldsymbol{j}} : \boldsymbol{j} \in [\mathbf{1} : \boldsymbol{J}])$ without any cross-product terms in the exponent. This gives independent Gaussian posterior distribution for each $\theta_{\boldsymbol{j}}$, given σ, such that by conjugacy,

$$\theta_{\boldsymbol{j}} | \mathbb{D}_n, \sigma \sim \mathrm{N}\big((N_{\boldsymbol{j}} \bar{Y}|_{I_{\boldsymbol{j}}} + \zeta_{\boldsymbol{j}} \lambda_{\boldsymbol{j}}^{-2})/(N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2}), \sigma^2/(N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2})\big), \tag{3.7}$$

where $N_{\boldsymbol{j}} = \#\{i : \boldsymbol{X}_i \in I_{\boldsymbol{j}}\}$ and $\bar{Y}|_{I_{\boldsymbol{j}}} = \sum_{i=1}^n Y_i \mathbb{1}\{\boldsymbol{X}_i \in I_{\boldsymbol{j}}\}/N_{\boldsymbol{j}}$.

The parameter σ^2 can be estimated by maximizing the marginal likelihood function given by

$$\begin{aligned}&(2\pi\sigma^2)^{-n/2} \prod_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} (1 + \lambda_{\boldsymbol{j}}^2 N_{\boldsymbol{j}})^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n \Big(Y_i - \sum_{\boldsymbol{j}: \boldsymbol{X}_i \in I_{\boldsymbol{j}}} \zeta_{\boldsymbol{j}} \Big)^2 - \sum_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} \frac{N_{\boldsymbol{j}}^2 (\bar{Y}|_{I_{\boldsymbol{j}}} - \zeta_{\boldsymbol{j}})^2}{N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2}} \right\} \right],\end{aligned}$$

and the resulting estimator

$$\hat{\sigma}_n^2 = \frac{1}{n} \left[\sum_{i=1}^n \Big(Y_i - \sum_{\boldsymbol{j}: \boldsymbol{X}_i \in I_{\boldsymbol{j}}} \zeta_{\boldsymbol{j}} \Big)^2 - \sum_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} \frac{N_{\boldsymbol{j}}^2 (\bar{Y}|_{I_{\boldsymbol{j}}} - \zeta_{\boldsymbol{j}})^2}{N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2}} \right], \tag{3.8}$$

may be plugged into the expression (3.7). Alternatively, in a fully Bayesian framework, we can give σ^2 an inverse-Gamma prior $\mathrm{IG}(b_1, b_2)$ with parameters $b_1 > 0$, $b_2 > 0$, and obtain that the posterior distribution of σ^2 is given by $\mathrm{IG}(b_1 + n/2, b_2 + n\hat{\sigma}_n^2/2)$. It will be shown in Lemma B.5 of [54] that the marginal maximum likelihood estimator of σ^2 as well as the posterior for σ^2 concentrate in a shrinking neighborhood of its true value σ_0^2. Then it easily follows that the asymptotic behavior of the posterior distribution of f is identical with that when σ is known to be σ_0. Hence, it suffices to study the asymptotic behavior of the posterior distribution given σ.

The unrestricted posterior distribution of f given σ is induced from (3.7) by the representation $f = \sum_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}}$. To obtain the immersion posterior distribution to make an inference, we consider three possible immersion maps.

Define

$$\mathcal{M}_{\boldsymbol{J}} = \left\{ f = \sum_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}} : \theta_{\boldsymbol{j}} \in \mathbb{R} \text{ and } \theta_{\boldsymbol{j}_1} \le \theta_{\boldsymbol{j}_2} \text{ if } \boldsymbol{j}_1 \preceq \boldsymbol{j}_2 \right\}, \tag{3.9}$$

consisting of the coordinatewise nondecreasing functions taking constant value on every $I_{\boldsymbol{j}}$.

Based on the isotonization procedure introduced in [28], consider transformations $\underline{\iota}$ and $\bar{\iota}$ acting on $f = \sum_{\boldsymbol{j} \in [\mathbf{1}:\boldsymbol{J}]} \theta_{\boldsymbol{j}} \mathbb{1}_{I_{\boldsymbol{j}}} \in \mathcal{K}_{\boldsymbol{J}}$ mapping to an element of $\mathcal{M}_{\boldsymbol{J}}$ defined by

$$\underline{\iota}(f)(\boldsymbol{x}) = \max_{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x})} \min_{\substack{\boldsymbol{j}_0(\boldsymbol{x}) \preceq \boldsymbol{j}_2 \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}, \tag{3.10}$$

$$\bar{\iota}(f)(\boldsymbol{x}) = \min_{\boldsymbol{j}_0(\boldsymbol{x}) \preceq \boldsymbol{j}_2} \max_{\substack{\boldsymbol{j}_1 \preceq \boldsymbol{j}_0(\boldsymbol{x}) \\ N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} > 0}} \frac{\sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]}}, \tag{3.11}$$

where $\boldsymbol{j}_0(\boldsymbol{x}) = \lceil \boldsymbol{x} \circ \boldsymbol{J} \rceil$, $N_{[\boldsymbol{j}_1:\boldsymbol{j}_2]} = \sum_{\boldsymbol{j} \in [\boldsymbol{j}_1:\boldsymbol{j}_2]} N_{\boldsymbol{j}}$, and $\boldsymbol{x} \in [0,1]^d$, for $\boldsymbol{j}_1$, $\boldsymbol{j}_2$ in $\mathbb{Z}^d$. The immersion posterior can be derived through the immersion map, ι, which is chosen to be either $\underline{\iota}$ or $\bar{\iota}$. This is determined by examining the resulting induced distribution of

$$f_* = \underline{\iota}(f), \tag{3.12}$$

$$f^* = \bar{\iota}(f). \tag{3.13}$$

It is obvious that $\iota(f) \in \mathcal{M}_{\boldsymbol{J}}$ and $\iota(f) = f$ if $f \in \mathcal{M}_{\boldsymbol{J}}$ and $N_{\boldsymbol{j}} > 0$ for all $\boldsymbol{j} \in [\mathbf{1} : \boldsymbol{J}]$. Generally, $\underline{\iota}(f)(\boldsymbol{x}) \le \overline{\iota}(f)(\boldsymbol{x})$ for any $\boldsymbol{x} \in [0,1]^d$, but this may fail to hold if $N_{\boldsymbol{j}} = 0$ for some $\boldsymbol{j}$, see [21]. To neutralize the effect stemming from the order of minimization and maximization, we propose using the average of $\underline{\iota}$ and $\overline{\iota}$, leading to another immersion map $\iota = (\underline{\iota} + \overline{\iota})/2$, by which f is mapped to

$$\tilde{f} = (f_* + f^*)/2. \tag{3.14}$$

The projection map for the univariate case is typically computed by the pool adjacent violator algorithm (see Section 2.3 of [4]), which requires $O(J)$ computations for a function with J steps. The computation of f_* or f^* requires no more than $(\prod_{k=1}^d J_k)^3$ operations by the brute-force search utilizing the block max-min or min-max formulas.

3.5. *Effect of the immersion map.* To see the effect of the immersion map on the posterior distribution of the function value at a point $\boldsymbol{x}_0 = (0.5, 0.5) \in [0,1]^2$, we conduct a small simulation study and compare the unrestricted and immersion posterior density for a randomly generated sample of three different sizes $n = 100, 200$ and 500, and three different regression functions: (i) $f_0(x_1, x_2) = x_1 + x_2$; (ii) $f_0(x_1, x_2) = \sqrt{x_1 + x_2}$; (iii) $f_0(x_1, x_2) = \mathbb{1}\{x_1 < 1/3\} + 2\mathbb{1}\{1/3 \le x_1 < 2/3\} + 3\mathbb{1}\{x_1 \ge 2/3\}$. Let X_1 and X_2 be distributed independently and uniformly on $[0,1]$ and error $\varepsilon \sim \mathrm{N}(0, \sigma^2)$ with true value of σ to be 0.1. We choose the number of grid points $J_1 = J_2 = J = \lceil n^{1/4} \log_{10} n \rceil$. The random heights, $\{\theta_{(j_1, j_2)} : j_1, j_2 \le J\}$, are endowed with the independent Gaussian prior $\mathrm{N}(0, 1000\sigma^2)$. The variance σ^2 is estimated using the maximum marginal likelihood method. We plot both the unrestricted posterior density and the estimated immersion posterior density in the same figure. The latter is based on 2000 posterior samples transformed by the immersion map $(\overline{\iota} + \underline{\iota})/2$.

As evident from Figure 1, the immersion posterior density functions exhibit lower variance across all instances, albeit to varying degrees depending on the true regression functions and sample sizes. Furthermore, the modes of the immersion posterior are nearer to the true value. The impacts of the other immersion maps, $\overline{\iota}$ and $\underline{\iota}$, on the posterior were found to be similar.

4. Coverage of credible intervals. Let $\boldsymbol{x}_0 \in (0,1)^d$ be fixed. Suppose that we want to make an inference on $f(\boldsymbol{x}_0)$. For a given $0 < \gamma < 1$, consider a $(1-\gamma)$-credible interval with endpoints the $\gamma/2$ and $(1-\gamma/2)$ quantiles of $f_*(\boldsymbol{x}_0)$, $f^*(\boldsymbol{x}_0)$, or $\tilde{f}(\boldsymbol{x}_0)$ defined in (3.12)–(3.14). To obtain the limiting frequentist coverage of these credible intervals, we obtain the weak limit of the immersion posterior distributions of f for all three immersion maps $\underline{\iota}$, $\overline{\iota}$ and $(\underline{\iota} + \overline{\iota})/2$ at $\boldsymbol{x} = \boldsymbol{x}_0$.

Let H_1 and H_2 be two independent centered Gaussian processes indexed by $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^d_{\ge 0} \times \mathbb{R}^d_{\ge 0}$ with the covariance kernel

$$\prod_{k=1}^{s} (u_k \wedge u_k' + v_k \wedge v_k') D_s(\boldsymbol{u} \wedge \boldsymbol{u}', \boldsymbol{v} \wedge \boldsymbol{v}'), \tag{4.1}$$

where $D_d(\boldsymbol{u}, \boldsymbol{v}) = g(\boldsymbol{x}_0)$, where g is the probability density function of $\boldsymbol{X}$, and for $s = 0, \ldots, d-1$, and $D_s(\boldsymbol{u}, \boldsymbol{v})$ is given by

$$\int_{\substack{x_k \in [(\boldsymbol{x}_0 - \boldsymbol{u})_k, (\boldsymbol{x}_0 + \boldsymbol{v})_k] \cap [0,1] \\ s+1 \le k \le d}} g(x_{0,1}, \ldots, x_{0,s}, x_{s+1}, \ldots, x_d)\, \mathrm{d}x_{s+1} \cdots \mathrm{d}x_d. \tag{4.2}$$

Additionally, we define a Gaussian process

$$\begin{aligned} U(\boldsymbol{u}, \boldsymbol{v}) = {} & \frac{\sigma_0 H_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} + \frac{\sigma_0 H_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} \\ & + \sum_{\boldsymbol{l} \in L^*} \frac{\partial^{\boldsymbol{l}} f_0(x_0)}{(\boldsymbol{l} + \mathbf{1})!} \prod_{k=1}^{s} \frac{v_k^{l_k+1} - (-u_k)^{l_k+1}}{u_k + v_k} \end{aligned} \tag{4.3}$$

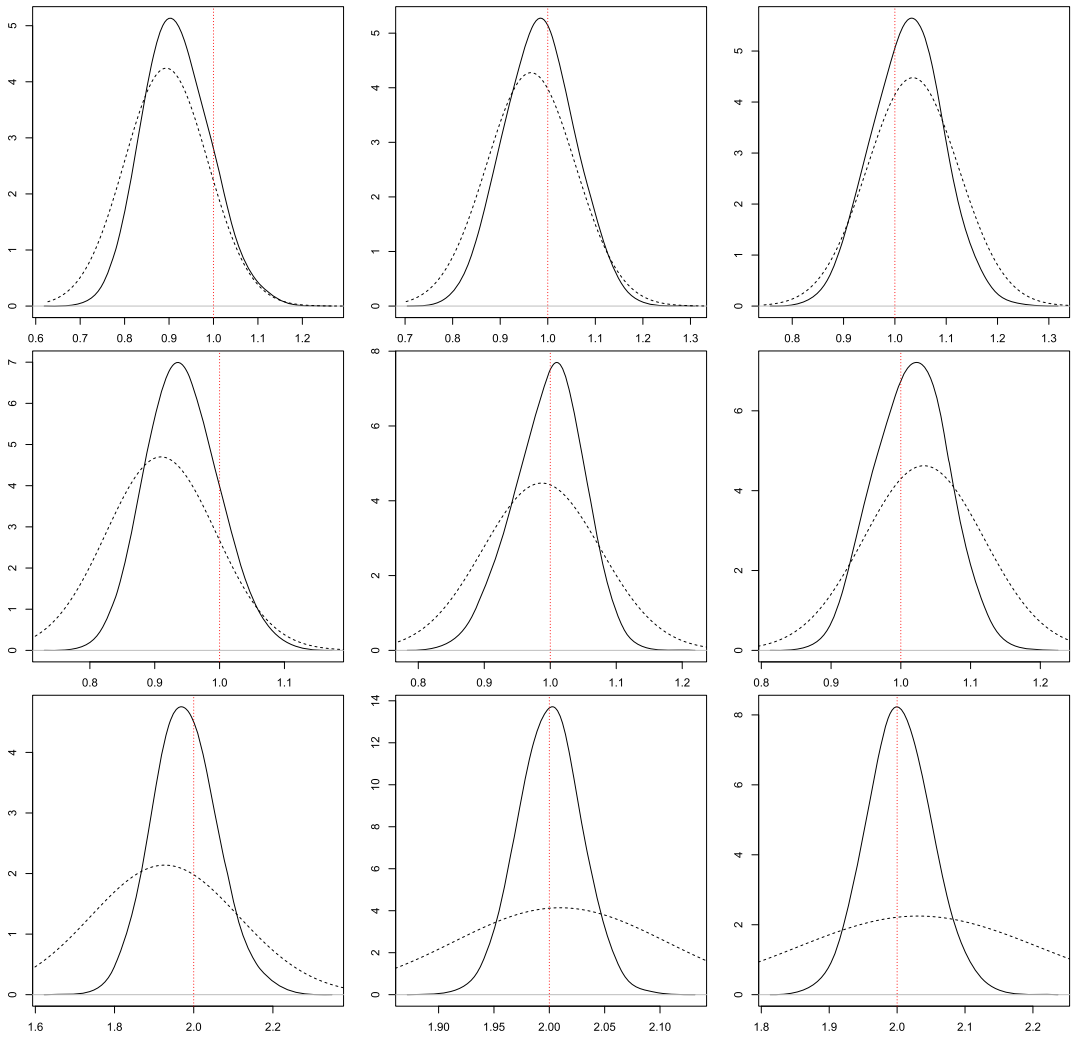


FIG. 1. *Unrestricted and immersion posterior density functions of* $f(\boldsymbol{x}_0)$. *The solid black line stands for the immersion posterior density, while the black dashed line represents the unrestricted posterior density. The true function value is indicated by the red dotted vertical line. The rows correspond to functions* (*i*), (*ii*) *and* (*iii*), *respectively, while the columns represent sample sizes of* $n = 100, 200$ *and* 500.

indexed by $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^d_{\geq 0} \times \mathbb{R}^d_{\geq 0}$, and its functionals

$$Z_* = \sup_{\substack{\boldsymbol{u}\succeq \boldsymbol{0} \\ u_k \leq x_{0,k} \\ s+1\leq k\leq d}} \inf_{\substack{\boldsymbol{v}\succeq \boldsymbol{0} \\ v_k \leq 1-x_{0,k} \\ s+1\leq k\leq d}} U(\boldsymbol{u}, \boldsymbol{v}), \qquad Z^* = \inf_{\substack{\boldsymbol{v}\succeq \boldsymbol{0} \\ v_k \leq 1-x_{0,k} \\ s+1\leq k\leq d}} \sup_{\substack{\boldsymbol{u}\succeq \boldsymbol{0} \\ u_k \leq x_{0,k} \\ s+1\leq k\leq d}} U(\boldsymbol{u}, \boldsymbol{v}). \tag{4.4}$$

The following result describes the asymptotic behavior of the normalized immersion posterior distributions of $f(\boldsymbol{x}_0)$. Recall that $\mathbb{D}_n$ represents the data and $r_{n,k}$ in Assumption 1 is the convergence rate along the kth direction through adjusting the overall rate ω_n according to the local smoothness levels. The weak limit of the normalized immersion posterior distribution function plays a central role in the study of the limiting coverage of the credible intervals based on the immersion posterior quantiles.

THEOREM 4.1. *Let* $\omega_n = n^{-1/(2+\sum_{k=1}^s \beta_k^{-1})}$ *and let* $\boldsymbol{r}_n = (\omega_n^{1/\beta_1}, \ldots, \omega_s^{1/\beta_s}, 1, \ldots, 1)^{\mathrm{T}}$. *Suppose that* $\boldsymbol{J}$ *satisfies* $J_k \gg r_{n,k}^{-1}$, *for each* $k = 1, \ldots, d$, *and* $\prod_{k=1}^d J_k \ll n\omega_n$. *Under As-*

sumptions 1 and 2, for any $z \in \mathbb{R}$, we have

$$\Pi\big(\omega_n^{-1}\big(f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le z | \mathbb{D}_n\big) \rightsquigarrow \mathrm{P}(Z_* \le z | H_1); \tag{4.5}$$

$$\Pi\big(\omega_n^{-1}\big(f^*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le z | \mathbb{D}_n\big) \rightsquigarrow \mathrm{P}\big(Z^* \le z | H_1\big); \tag{4.6}$$

$$\Pi\big(\omega_n^{-1}\big(\tilde{f}(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le z | \mathbb{D}_n\big) \rightsquigarrow \mathrm{P}\big((Z_* + Z^*)/2 \le z | H_1\big). \tag{4.7}$$

Furthermore, for any $(z_1, z_2) \in \mathbb{R}^2$,

$$\begin{aligned}\Pi\big(\omega_n^{-1}\big(f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le z_1, \omega_n^{-1}\big(f^*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le z_2 | \mathbb{D}_n\big) \\ \rightsquigarrow \mathrm{P}\big(Z_* \le z_1, Z^* \le z_2 | H_1\big).\end{aligned} \tag{4.8}$$

REMARK 1. We make some remarks on Theorem 4.1.

1. The weak limit is understood in the usual sense for random variables since we consider the limiting behavior of the random probability measure of a fixed set $(-\infty, z]$. We refer to the proof technique of [38], which provides the distributional theory for the block-estimator in general multivariate isotonic regression, especially the small and large deviation arguments therein.

2. For the choice of J_k, the lower bound $r_{n,k}^{-1}$ is essential for Theorem 4.1. That eliminates the effect of the roughness of piecewise constant function approximation in view of the local contraction rate. But the upper bound, $n\omega_n$ in Theorem 4.1, is not fundamentally necessary for the validity of the weak limit. Instead, we can set the hyperparameters λ_j large enough, specifically, $\min \lambda_j^2 \gg \omega_n^{-1}\sqrt{n}$, to obtain the limiting theory. The rest of the proof of Theorem 4.1 will not be affected much without such an upper bound except for the treatment of σ^2. The estimation of σ^2 is not a hard problem and any consistent procedure will work. For any $\beta_k \ge 1$ and any $0 \le s \le d$, we observe that $r_k \le n^{-1/3}$ for all $1 \le k \le d$. Without the local smoothness information, we can choose $J_k \gg n^{1/3}$. On the other side, if we admit that $\beta_k = 1$, $1 \le k \le d$, is the leading case for the multivariate regression function, we can then choose $J_k \gg n^{1/(2+d)}$. One may raise concerns that selecting J in this manner seems not optimal when the true regression function is less smooth. However, this concern typically does not pose a big issue in practice. When lacking prior information about the regression function, an appropriately large J_k can be chosen and an uninformative prior should be applied to θ_j, such as a normal prior with a large variance. Our empirical study indicates that opting for a larger J can indeed enhance the performance of our method. For practical applications, we recommend selecting J_k no less than 15, particularly when dealing with a smaller sample size. Notably, $\boldsymbol{J}$ is not a tuning parameter in this context; the immersion posterior is governed by shape restrictions rather than a tuning process like bandwidth selection in kernel smoothing. The choice of $\boldsymbol{J}$ does not influence the contraction rate or the distributional theory, distinguishing it from typical tuning parameters.

3. It is also important to note that we employ a working normal model to derive the posterior distribution. The validity of this method remains intact even when the model is misspecified. The finite-moment condition for the random error ε can be relaxed to the second order, as in [38], by selecting a sufficiently large λ_j^2 as in the last point.

The covariance kernels of the processes H_1 and H_2 depend on g, and the limiting Gaussian process also involves the derivative values of f_0 at $\boldsymbol{x}_0$. A considerable simplification happens in some special cases where the parameters appear through a scale parameter in the kernel. It will be seen shortly that this fact has a far-reaching implication in that the limiting coverage of a credible interval constructed from the immersion posterior is free of the unknown parameters of the model. If L^* defined by (3.4) only contains $\beta_k \boldsymbol{e}_k$ for $k = 1, \ldots, s$, where

$\boldsymbol{e}_k$ denotes the standard unit vector in $\mathbb{R}^d$ with one in the kth component and zero elsewhere, then the limiting processes in Theorem 4.1 can be further simplified by self-similarity. A factor depending on f_0 comes out as a multiplicative constant, and the remaining factor is only a known functional of H_1 and H_2. The case $s = d$ stands for the regular case that all directional derivatives of f_0 at $\boldsymbol{x}_0$ are positive at a certain order. Then the covariance kernel further simplifies as a completely known function and a factor involving derivatives of the regression function and predictor density g. The result is precisely formulated in the result below.

PROPOSITION 4.1. *If* $L^* = \{\beta_k \boldsymbol{e}_k : 1 \le k \le s\}$, *then*

$$\begin{aligned}
&\sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \Bigg\{ \frac{\sigma_0 H_1(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^s (u_k+v_k) D_s(\boldsymbol{u},\boldsymbol{v})} + \frac{\sigma_0 H_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^s (u_k+v_k) D_s(\boldsymbol{u},\boldsymbol{v})} \\
&\qquad + \sum_{k=1}^s \Bigg[\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!} \cdot \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k+v_k} \Bigg] \Bigg\} \\
&=_d A_{\boldsymbol{\beta}} \cdot \sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \Bigg\{ \frac{H_1(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^s (u_k+v_k) D_s(\boldsymbol{u},\boldsymbol{v})} + \frac{H_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^s (u_k+v_k) D_s(\boldsymbol{u},\boldsymbol{v})} \\
&\qquad + \sum_{k=1}^s \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k+v_k} \Bigg\},
\end{aligned}$$

where $A_{\boldsymbol{\beta}} = (\sigma_0^2 \prod_{k=1}^s (\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!})^{1/\beta_k})^{1/(2+\sum_{k=1}^s \beta_k^{-1})}$.

Furthermore, if $s = d$, *then the above expression further simplifies to*

$$\tilde{A}_{\boldsymbol{\beta}} \sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \Bigg\{ \frac{\tilde{H}_1(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d (u_k+v_k)} + \frac{\tilde{H}_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d (u_k+v_k)} + \sum_{k=1}^d \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k+v_k} \Bigg\},$$

where $\tilde{A}_{\boldsymbol{\beta}} = (\frac{\sigma_0^2}{g(\boldsymbol{x}_0)} \prod_{k=1}^d (\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!})^{1/\beta_k})^{1/(2+\sum_{k=1}^d \beta_k^{-1})}$, *and* $\tilde{H}_1$ *and* $\tilde{H}_2$ *are two independent centered Gaussian processes with covariance kernel given by* $\prod_{k=1}^d (u_k \wedge u_k' + v_k \wedge v_k')$, $(\boldsymbol{u},\boldsymbol{v}), (\boldsymbol{u}',\boldsymbol{v}') \in \mathbb{R}^d_{\ge 0} \times \mathbb{R}^d_{\ge 0}$.

The same conclusion also applies to the inf sup*-functional obtained by switching the positions of the supremum and the infimum.*

REMARK 2 (Univariate case). We specialize to the univariate case $s = d = 1$, with a general β, expanding from the case $\beta = 1$ studied by [14]. Then

$$\tilde{H}_i(u,v) =_d W_i(v) + W_i(-u) =_d W_i(v) - W_i(-u), \quad (u,v) \in \mathbb{R}^2_{\ge 0}, \tag{4.9}$$

where W_1, W_2 are two independent standard two-sided Brownian motions starting from 0. Observe that the sup-inf functional

$$\begin{aligned}
&\sup_{u>0} \inf_{v>0} \Bigg\{ \frac{\tilde{H}_1(u,v)}{u+v} + \frac{\tilde{H}_2(u,v)}{u+v} + \frac{v^{\beta_1+1} - (-u)^{\beta_1+1}}{u+v} \Bigg\} \\
&=_d \sup_{u>0} \inf_{v>0} \Bigg\{ \frac{(W_1(v) + W_2(v) + v^{\beta_1+1}) - (W_1(-u) + W_2(-u) + u^{\beta_1+1})}{v-(-u)} \Bigg\},
\end{aligned}$$

coincides with the slope of the greatest convex minorant of the process $W_1(t) + W_2(t) + t^{\beta+1}$. By the switching relation (cf. [33], page 56), for any $z \in \mathbb{R}$,

$$\begin{aligned}
&\mathrm{P}\Bigg(\tilde{A}_\beta \sup_{u>0} \inf_{v>0} \Bigg\{ \frac{\tilde{H}_1(u,v)}{u+v} + \frac{\tilde{H}_2(u,v)}{u+v} + \frac{v^{\beta+1} - (-u)^{\beta+1}}{u+v} \Bigg\} \le z \Bigg) \\
&\quad = \mathrm{P}\big(\arg\min\{ W_1(t) + W_2(t) + t^{\beta+1} - \tilde{A}_\beta^{-1} z t : t \in \mathbb{R} \} \ge 0 \big).
\end{aligned}$$

If $\beta = 1$, the last display can be further simplified by applying the change of variable, $t = s + z/(2\tilde{A}_1)$, and is equal to

$$\mathrm{P}\big(2\tilde{A}_1 \arg\min\{W_1(s) + W_2(s) + s^2 : s \in \mathbb{R}\} \le z\big),$$

with $\tilde{A}_1 = (\sigma_0^2 f'(x_0)/(2g(x_0)))^{1/3}$. This reproduces the main result of [14].

Now we are ready for the evaluation of the limiting coverage of an immersion posterior credible interval for $f(\boldsymbol{x}_0)$. Let

$$Q_{n,\gamma}^{(1)} = \inf\{z : \Pi\big(f_*(\boldsymbol{x}_0) \le z|\mathbb{D}_n\big) \ge 1 - \gamma\} \tag{4.10}$$

stand for the $(1-\gamma)$-quantile of $f_*(\boldsymbol{x}_0)$. Similarly, let $Q_{n,\gamma}^{(2)}$ and $Q_{n,\gamma}^{(3)}$ stand for that of $f^*(\boldsymbol{x}_0)$ and $\tilde{f}(\boldsymbol{x}_0)$, respectively. Let $\tilde{U}(\boldsymbol{u}, \boldsymbol{v})$ stand for the Gaussian process

$$\frac{\tilde{H}_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \frac{\tilde{H}_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^d (u_k + v_k)} + \sum_{k=1}^d \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k} \tag{4.11}$$

indexed by $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}_{\ge 0}^d \times \mathbb{R}_{\ge 0}^d$.

The following result gives the ultimate conclusion of the paper about asymptotic coverage of credible intervals for the regression function value at an interior point.

THEOREM 4.2. *Under the assumed setup, Assumptions 1 and 2, and the condition that* $L^* = \{\beta_k \boldsymbol{e}_k : 1 \le k \le s\}$, *the asymptotic coverage of the quantile-based one-sided credible interval* $(-\infty, Q_{n,\gamma}^{(1)}]$ *is given by*

$$\mathrm{P}\Bigg(\mathrm{P}\Bigg(\sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \Bigg\{ \frac{H_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} + \frac{H_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} + \sum_{k=1}^s \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k} \Bigg\} \le 0 \Big| H_1 \Bigg) \le 1 - \gamma \Bigg).$$

If $Q_{n,\gamma}^{(1)}$ *is replaced by* $Q_{n,\gamma}^{(2)}$, *the above limit is changed by swapping the order of the supremum and infimum operations. If* $Q_{n,\gamma}^{(1)}$ *is replaced by* $Q_{n,\gamma}^{(3)}$, *the above limit is changed by replacing the expression on the right-hand side with the average of the* sup inf *and* inf sup *operations.*

Moreover, if $s = d$:

(i) $\mathrm{P}_0(f_0(\boldsymbol{x}_0) \le Q_{n,\gamma}^{(1)}) \to \mathrm{P}(Z_B^{(1)} \le 1 - \gamma)$;
(ii) $\mathrm{P}_0(f_0(\boldsymbol{x}_0) \le Q_{n,\gamma}^{(2)}) \to \mathrm{P}(Z_B^{(2)} \le 1 - \gamma)$;
(iii) $\mathrm{P}_0(f_0(\boldsymbol{x}_0) \le Q_{n,\gamma}^{(3)}) \to \mathrm{P}(Z_B^{(3)} \le 1 - \gamma)$,

where $Z_B^{(1)} = \mathrm{P}(\sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \tilde{U}(\boldsymbol{u}, \boldsymbol{v}) \le 0|\tilde{H}_1)$, $Z_B^{(2)} = \mathrm{P}(\inf_{\boldsymbol{v}\ge\boldsymbol{0}} \sup_{\boldsymbol{u}\ge\boldsymbol{0}} \tilde{U}(\boldsymbol{u}, \boldsymbol{v}) \le 0|\tilde{H}_1)$ *and* $Z_B^{(3)} = \mathrm{P}(\frac{1}{2}\{\sup_{\boldsymbol{u}\ge\boldsymbol{0}} \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \tilde{U}(\boldsymbol{u}, \boldsymbol{v}) + \inf_{\boldsymbol{v}\ge\boldsymbol{0}} \sup_{\boldsymbol{u}\ge\boldsymbol{0}} \tilde{U}(\boldsymbol{u}, \boldsymbol{v})\} \le 0|\tilde{H}_1)$.

PROOF. We observe that $f_0(\boldsymbol{x}_0) \le Q_{n,\gamma}^{(1)}$ if and only if

$$\Pi\big(f_*(\boldsymbol{x}_0) \le f_0(\boldsymbol{x}_0)|\mathbb{D}_n\big) = \Pi\big(\omega_n^{-1}\big(f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)\big) \le 0|\mathbb{D}_n\big) \le 1 - \gamma.$$

Hence, by Theorem 4.1 and Proposition 4.1, as the multiplicative positive constant in the limiting process can be dropped because the interval $(-\infty, 0]$ remains invariant under a scale-change, the first conclusion follows immediately. The special cases follow from the second part of Proposition 4.1. □

REMARK 3. For $d = 1$, $Z_B^{(1)}$, $Z_B^{(2)}$ and $Z_B^{(3)}$ all coincide, and may be simply denoted by Z_B as in [14].

The distributions of $Z_B^{(1)}$ and $Z_B^{(2)}$ are related, as shown next.

PROPOSITION 4.2. *For any* $z \in [0, 1]$, *we have* $\mathrm{P}(Z_B^{(1)} \le z) = \mathrm{P}(Z_B^{(2)} \ge 1 - z)$, *and* $Z_B^{(3)}$ *is symmetrically distributed about* $1/2$.

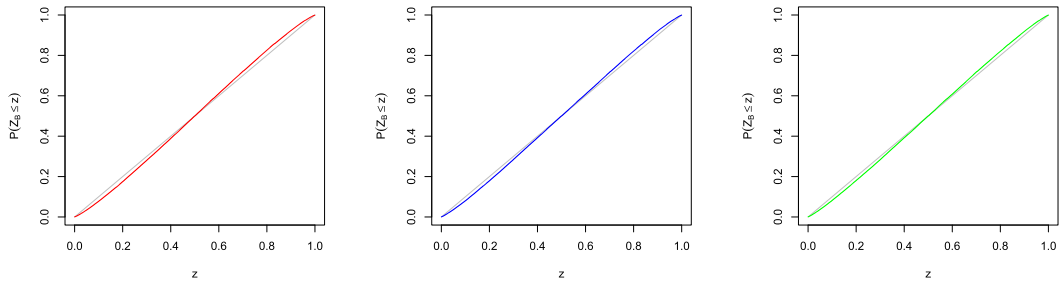
From Theorem 4.2 and Proposition 4.2, it follows that the limiting coverage of a one-sided Bayesian credible interval for $f(\boldsymbol{x}_0)$ using one of the three proposed immersion posteriors can be evaluated, is free of the true regression function (and also is free of the density g of the predictor if $s = d$, and hence depends only on the credibility level), but in general, need not be equal to the credibility. Nevertheless, a targeted limiting coverage can be obtained by starting with a certain credibility level that can be explicitly computed by back-calculation. As in the univariate monotone problems studied by [14, 15], numerical calculations show that the required credibility to obtain a specific limiting coverage is less than the targeted coverage, the opposite of the phenomenon [20] observed for smoothing problems. However, unlike in the univariate case where the limiting Bayes–Chernoff distribution determining the asymptotic coverage of the credible interval is symmetric, the corresponding random variables $Z_B^{(1)}$ and $Z_B^{(2)}$ for the posterior based on the immersion maps $\underline{\iota}$ and $\bar{\iota}$ appearing in the multivariate case are not symmetric. This has implications for the limiting coverage of a two-sided credible interval, which is more commonly used in practice. For instance, for $0 < \gamma < 1/2$, a two-sided $(1-\gamma)$-credible interval $[Q_{n,1-\gamma/2}, Q_{n,\gamma/2}]$ based on the immersion posterior using the map $\underline{\iota}$, the limiting coverage is given by $\mathrm{P}(Z_B^{(1)} \le 1 - \gamma/2) - \mathrm{P}(Z_B^{(1)} \le \gamma/2)$. The corresponding limit for the immersion posterior using the map $\bar{\iota}$ is $\mathrm{P}(Z_B^{(2)} \le 1 - \gamma/2) - \mathrm{P}(Z_B^{(2)} \le \gamma/2)$. Interestingly, a separate table for the distribution function of $Z_B^{(2)}$ is not needed, as it can be obtained from that of $Z_B^{(1)}$ in view of Proposition 4.2. The symmetry of $Z_B^{(3)}$, however, implies that the credibility level $1 - \gamma$ needed to make the asymptotic coverage of an equal-tailed $(1-\gamma)$-credible interval $1-\alpha$ is obtained by choosing $1 - \gamma = 1 - 2F^{-1}_{Z_B^{(3)}}(\alpha/2)$, which is readily obtained once the cumulative distribution function $F_{Z_B^{(3)}}$ of $Z_B^{(3)}$ is tabulated.

5. Numerical results.

5.1. *Distribution of* Z_B. In this section, we present tables detailing the distribution and quantiles of Z_B for the case $d = 1$ when $\beta = 1, 3, 5$, as well as those for $Z_B^{(1)}$, $Z_B^{(2)}$, $Z_B^{(3)}$ for the case $d = 2$ when $\boldsymbol{\beta} = (1, 1), (1, 3), (3, 3)$. The distributions of these variables are simulated using the Monte Carlo method, with the Gaussian processes concerned being generated by discrete approximation. The quantile table can function as a recalibration reference to achieve the exact frequentist asymptotic coverage.

5.1.1. *Case* $d = 1$. First, we generate approximations to the Gaussian processes $\tilde{H}_1$ and $\tilde{H}_2$. Let $\tilde{H}$ denote either $\tilde{H}_1$ or $\tilde{H}_2$. To approximate $\tilde{H}$, we generate $14m$ independent standard Gaussian random variables, specifically $\{\zeta_j : j = 1, \ldots, 7m\}$ and $\{\zeta'_j : j = 1, \ldots, 7m\}$, where $m = 50$. Then $\tilde{H}$ can be approximated as follows:

$$\tilde{H}(u, v) \approx \frac{1}{\sqrt{m}} \left[\sum_{j=1}^{\lceil mu \rceil} \zeta_j + \sum_{j=1}^{\lceil mv \rceil} \zeta'_j \right], \tag{5.1}$$



(a) $\beta = 1$ (b) $\beta = 3$ (c) $\beta = 5$

FIG. 2. *Distribution functions of* Z_B.

for $u, v \in [0, 7]$. Given each instance of $\tilde{H}_1$, we generate 500 realizations of $\tilde{H}_2$. For each realization, we compute the sup-inf functional. The proportion of nonpositive outcomes then serves as a sample value of Z_B. We repeat the generation process 50,000 times to obtain the approximate distribution function of Z_B.

In Figure 2, we draw the simulated distribution functions of Z_B with $\beta = 1, 3$ and 5. We give the values of $\mathrm{P}(Z_B \le z)$ with different smoothness levels for selected z values in Table 1 and the values of the quantiles of Z_B's distribution in Table 2.

5.1.2. *Case* $d = 2$. To approximate $\tilde{H}(\boldsymbol{u}, \boldsymbol{v})$, for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2$, we generate 4 random matrices $\zeta^{(1)}$, $\zeta^{(2)}$, $\zeta^{(3)}$ and $\zeta^{(4)}$ with independent standard Gaussian random variables. The dimensions of these 4 matrices are $\lceil mt_1 \rceil \times \lceil mt_2 \rceil$, $\lceil ms_1 \rceil \times \lceil mt_2 \rceil$, $\lceil ms_1 \rceil \times \lceil ms_2 \rceil$ and $\lceil mt_1 \rceil \times \lceil ms_2 \rceil$, for $m = 5$ and $t_1 = t_2 = s_1 = s_2 = 5$. $\tilde{H}(\boldsymbol{u}, \boldsymbol{v})$ is then approximated by

$$\frac{1}{m}\left(\sum_{i=1}^{\lceil mv_1 \rceil}\sum_{j=1}^{\lceil mv_2 \rceil}\zeta_{ij}^{(1)} + \sum_{i=1}^{\lceil mu_1 \rceil}\sum_{j=1}^{\lceil mv_2 \rceil}\zeta_{ij}^{(2)} + \sum_{i=1}^{\lceil mu_1 \rceil}\sum_{j=1}^{\lceil mu_2 \rceil}\zeta_{ij}^{(3)} + \sum_{i=1}^{\lceil mv_1 \rceil}\sum_{j=1}^{\lceil mu_2 \rceil}\zeta_{ij}^{(4)}\right),$$

for $u_1, u_2, v_1, v_2 \in [0, 5]$.

To get a sample of any one of $Z_B^{(1)}$, $Z_B^{(2)}$ or $Z_B^{(3)}$, we first generate a sample of $\tilde{H}_1$. Given this sample, we generate 500 realizations of $\tilde{H}_2$. We then compute the three functionals that define $Z_B^{(1)}$, $Z_B^{(2)}$ and $Z_B^{(3)}$. The conditional probabilities are approximated by the frequency

TABLE 1
Values of $\mathrm{P}(Z_B \le z)$

z	0.700	0.750	0.800	0.850	0.900	0.950	0.975	0.990	0.995
$\beta = 1$	0.719	0.772	0.826	0.875	0.923	0.965	0.985	0.994	0.997
$\beta = 3$	0.715	0.768	0.821	0.870	0.921	0.963	0.983	0.994	0.997
$\beta = 5$	0.716	0.768	0.820	0.869	0.919	0.962	0.983	0.993	0.997

TABLE 2
Values of $q = \inf\{z : \mathrm{P}(Z_B \le z) \ge p\}$

p	0.700	0.750	0.800	0.850	0.900	0.950	0.975	0.990	0.995
$\beta = 1$	0.683	0.730	0.777	0.825	0.878	0.932	0.964	0.994	0.997
$\beta = 3$	0.687	0.734	0.781	0.829	0.882	0.935	0.966	0.986	0.992
$\beta = 5$	0.686	0.734	0.782	0.831	0.882	0.936	0.966	0.986	0.994

TABLE 3
Values of $\mathrm{P}(Z_B \le z)$ *for various* z *and* $\boldsymbol{\beta}$, *and* $Z_B = Z_B^{(1)}, Z_B^{(2)}, Z_B^{(3)}$

	$\boldsymbol{\beta} = (1,1)$			$\boldsymbol{\beta} = (3,1)$			$\boldsymbol{\beta} = (3,3)$		
z	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$
0.700	0.705	0.752	0.725	0.704	0.741	0.721	0.708	0.735	0.718
0.750	0.762	0.803	0.778	0.760	0.791	0.773	0.762	0.787	0.771
0.800	0.817	0.851	0.832	0.814	0.842	0.827	0.817	0.838	0.825
0.850	0.871	0.898	0.880	0.868	0.889	0.877	0.868	0.885	0.874
0.900	0.921	0.939	0.927	0.917	0.932	0.924	0.918	0.930	0.922
0.950	0.966	0.975	0.968	0.964	0.971	0.966	0.964	0.970	0.965
0.975	0.985	0.989	0.987	0.983	0.987	0.986	0.984	0.986	0.985
0.990	0.995	0.997	0.995	0.995	0.997	0.995	0.995	0.996	0.994
0.995	0.997	0.998	0.998	0.998	0.998	0.998	0.997	0.998	0.997

of nonpositive functional values. This process is repeated 50,000 times for $\boldsymbol{\beta} = (1,1), (3,1)$ and $(3,3)$ to estimate the distribution of $Z_B^{(1)}$, $Z_B^{(2)}$ or $Z_B^{(3)}$.

Since in higher-dimensional cases, $Z_B^{(1)}$ and $Z_B^{(2)}$ are not equal in distribution and their distribution functions are not symmetric about 0.5, we give both the values of $\mathrm{P}(Z_B^{(1)} \le z)$ and $\mathrm{P}(Z_B^{(2)} \le z)$ for some selected z values in Table 3. The corresponding distribution functions are plotted in Figure 3. We present the quantiles of $Z_B^{(3)}$ with different smoothness levels in Table 4.

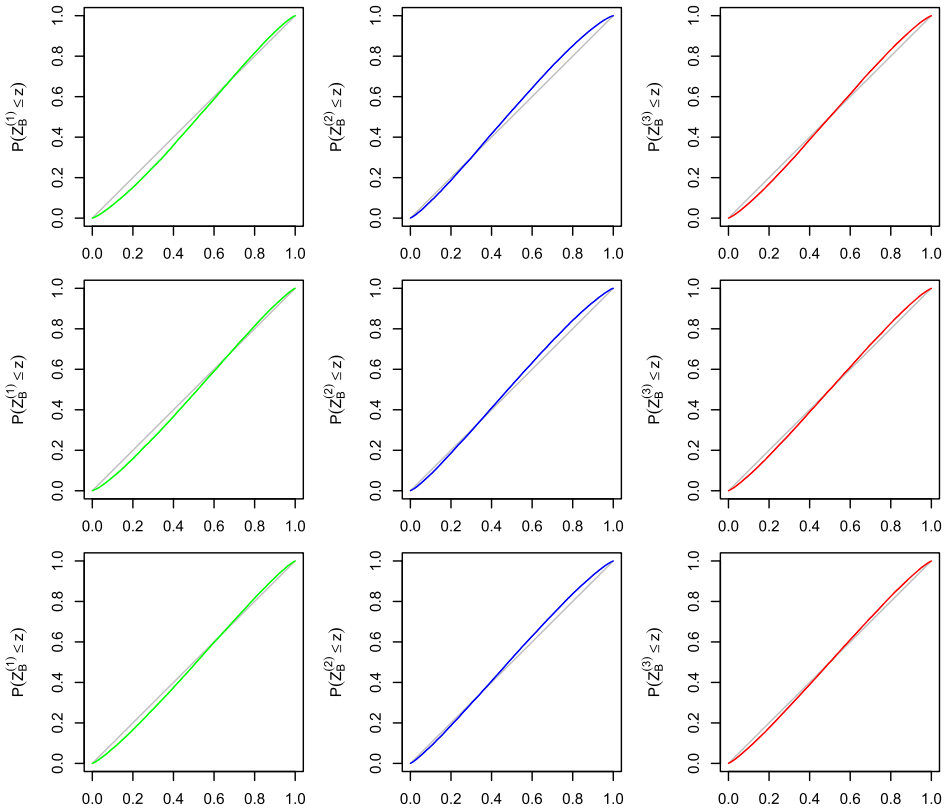


FIG. 3. *Distribution functions of* $Z_B^{(1)}$, $Z_B^{(2)}$ *or* $Z_B^{(3)}$. *The three plots in the first row are for* $\boldsymbol{\beta} = (1,1)$; *the second row is for* $\boldsymbol{\beta} = (3,1)$; *the last row is for* $\boldsymbol{\beta} = (3,3)$.

TABLE 4
Values of $q = \inf\{z : \mathrm{P}(Z_B \le z) \ge p\}$ *for various* p, *and* $Z_B = Z_B^{(1)}, Z_B^{(2)}, Z_B^{(3)}$

	$\boldsymbol{\beta} = (1,1)$			$\boldsymbol{\beta} = (3,1)$			$\boldsymbol{\beta} = (3,3)$		
p	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$	$Z_B^{(1)}$	$Z_B^{(2)}$	$Z_B^{(3)}$
0.700	0.697	0.653	0.677	0.699	0.665	0.681	0.695	0.669	0.684
0.750	0.741	0.699	0.724	0.743	0.711	0.728	0.741	0.715	0.732
0.800	0.787	0.749	0.771	0.789	0.759	0.776	0.787	0.763	0.778
0.850	0.833	0.801	0.819	0.835	0.811	0.823	0.833	0.815	0.825
0.900	0.881	0.855	0.872	0.883	0.865	0.876	0.883	0.869	0.878
0.950	0.933	0.917	0.928	0.937	0.925	0.931	0.935	0.927	0.933
0.975	0.963	0.951	0.959	0.965	0.957	0.962	0.965	0.959	0.964
0.990	0.983	0.977	0.982	0.985	0.981	0.984	0.985	0.983	0.984
0.995	0.991	0.987	0.990	0.991	0.989	0.992	0.993	0.991	0.992

5.2. *Comparison with Deng, Han and Zhang's method.* For pointwise inference in multivariate isotonic regression, Deng et al. [21] constructed the confidence interval by the asymptotic distribution of pivotal statistics. Their method will be referred to as DHZ in the following. Let $\hat{\boldsymbol{u}}(\boldsymbol{x}_0)$ and $\hat{\boldsymbol{v}}(\boldsymbol{x}_0)$ be such that

$$\hat{f}^-(\boldsymbol{x}_0) = \max_{\boldsymbol{u} \preceq \boldsymbol{x}_0} \min_{\substack{\boldsymbol{v} \succeq \boldsymbol{x}_0 \\ \#\{i : \boldsymbol{X}_i \in [\boldsymbol{u}:\boldsymbol{v}]\} > 0}} \bar{Y}|_{[\boldsymbol{u}:\boldsymbol{v}]} = \min_{\substack{\boldsymbol{v} \succeq \boldsymbol{x}_0 \\ \#\{i : \boldsymbol{X}_i \in [\hat{\boldsymbol{u}}(\boldsymbol{x}_0):\boldsymbol{v}]\} > 0}} \bar{Y}|_{[\hat{\boldsymbol{u}}(\boldsymbol{x}_0):\boldsymbol{v}]},$$

$$\hat{f}^+(\boldsymbol{x}_0) = \min_{\boldsymbol{v} \succeq \boldsymbol{x}_0} \max_{\substack{\boldsymbol{u} \preceq \boldsymbol{x}_0 \\ \#\{i : \boldsymbol{X}_i \in [\boldsymbol{u}:\boldsymbol{v}]\} > 0}} \bar{Y}|_{[\boldsymbol{u}:\boldsymbol{v}]} = \max_{\substack{\boldsymbol{u} \preceq \boldsymbol{x}_0 \\ \#\{i : \boldsymbol{X}_i \in [\boldsymbol{u}:\hat{\boldsymbol{v}}(\boldsymbol{x}_0)]\} > 0}} \bar{Y}|_{[\boldsymbol{u}:\hat{\boldsymbol{v}}(\boldsymbol{x}_0)]},$$

and $\hat{f}(\boldsymbol{x}_0) = (\hat{f}^-(\boldsymbol{x}_0) + \hat{f}^+(\boldsymbol{x}_0))/2$. Under the same data generating conditions as in Theorem 4.2 and additionally assuming $\boldsymbol{X}$ is uniform distributed, Deng et al. [21] showed that

$$\frac{\sqrt{\#\{i : \boldsymbol{X}_i \in [\hat{\boldsymbol{u}}(\boldsymbol{x}_0) : \hat{\boldsymbol{v}}(\boldsymbol{x}_0)]\}}}{\sigma} \big(\hat{f}(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)\big) \rightsquigarrow K_{\boldsymbol{\beta}},$$

where $K_{\boldsymbol{\beta}}$ is a universal distribution that depends solely on the local regularity $\boldsymbol{\beta}$. Let $1 - \gamma \in (0.5, 1)$ be the confidence level. They proposed the following confidence interval for $f_0(\boldsymbol{x}_0)$:

$$\left[\hat{f}(\boldsymbol{x}_0) - \frac{c_\gamma \hat{\sigma}}{\sqrt{\#\{i : \boldsymbol{X}_i \in [\hat{\boldsymbol{u}}(\boldsymbol{x}_0) : \hat{\boldsymbol{v}}(\boldsymbol{x}_0)]\}}}, \hat{f}(\boldsymbol{x}_0) + \frac{c_\gamma \hat{\sigma}}{\sqrt{\#\{i : \boldsymbol{X}_i \in [\hat{\boldsymbol{u}}(\boldsymbol{x}_0) : \hat{\boldsymbol{v}}(\boldsymbol{x}_0)]\}}}\right], \tag{5.2}$$

where c_γ is the critical value obtained by simulating the limiting distribution of $K_{\boldsymbol{\beta}}$ and $\hat{\sigma}$ is a consistent estimator of σ.

We consider five regression functions: (1) $f_1(x_1, x_2) = (x_1 + x_2)^2$; (2) $f_2(x_1, x_2) = \sqrt{x_1 + x_2}$; (3) $f_3(x_1, x_2) = x_1 x_2$; (4) $f_4(x_1, x_2) = e^{x_1 + x_2}$; (5) $f_5(x_1, x_2) = e^{x_1 x_2}$. Set $\varepsilon_i \sim \mathrm{N}(0, 1)$ and $X_1, X_2 \sim \mathrm{Unif}(0, 1)$, mutually independent, for $i = 1, \ldots, n$. We consider sample sizes $n = 200, 500, 1000$ and 2000. To construct credible intervals, we choose $J = \lceil n^{1/3} \log(\log n) \rceil$. We compare the coverage and length of our immersion credible interval (IB), the recalibrated credible interval (IB(adj)) and the DHZ's confidence interval under two credible/confidence levels 0.95 and 0.90. The coverage percentage and the average length are calculated over 2000 replications. The result is summarized in Table 5.

The unadjusted credible intervals generally overcover the true function value for larger sample sizes, whereas the recalibrated credible intervals provide more accurate coverage to different extents for different functions. DHZ's method yields more precise coverage at the given confidence level when the sample sizes are relatively smaller. However, our credible

TABLE 5
Coverage percentage (C) and length (L) comparison

			IB		IB(adj)		DHZ	
f	Level	n	C	L	C	L	C	L
f_1	0.05	200	93.6	0.903 (0.145)	90.0	0.805 (0.132)	92.4	1.138 (0.600)
		500	98.8	0.777 (0.111)	97.7	0.692 (0.101)	95.0	0.959 (0.490)
		1000	97.0	0.630 (0.086)	94.4	0.562 (0.078)	94.8	0.827 (0.435)
		2000	97.4	0.535 (0.072)	95.4	0.476 (0.066)	94.9	0.686 (0.324)
	0.10	200	88.1	0.761 (0.126)	81.5	0.668 (0.112)	86.1	0.898 (0.473)
		500	96.9	0.656 (0.097)	94.4	0.576 (0.087)	90.0	0.757 (0.387)
		1000	92.8	0.532 (0.075)	88.8	0.467 (0.068)	88.8	0.652 (0.343)
		2000	94.3	0.451 (0.063)	90.4	0.397 (0.057)	89.7	0.541 (0.256)
f_2	0.05	200	91.0	0.503 (0.089)	87.0	0.447 (0.081)	95.2	0.722 (0.339)
		500	96.6	0.380 (0.061)	93.8	0.338 (0.055)	95.4	0.546 (0.303)
		1000	94.9	0.308 (0.047)	91.7	0.274 (0.043)	94.8	0.439 (0.252)
		2000	95.9	0.253 (0.039)	93.3	0.225 (0.035)	95.3	0.357 (0.175)
	0.10	200	85.0	0.423 (0.077)	79.0	0.371 (0.068)	89.8	0.570 (0.268)
		500	92.7	0.320 (0.052)	88.0	0.280 (0.047)	90.3	0.431 (0.239)
		1000	90.0	0.259 (0.040)	84.7	0.227 (0.036)	88.9	0.346 (0.199)
		2000	91.8	0.213 (0.033)	87.0	0.186 (0.030)	90.6	0.281 (0.138)
f_3	0.05	200	91.8	0.476 (0.084)	87.2	0.423 (0.076)	94.7	0.740 (0.410)
		500	96.0	0.371 (0.061)	93.4	0.329 (0.055)	95.0	0.532 (0.246)
		1000	95.0	0.293 (0.046)	91.6	0.260 (0.042)	95.4	0.433 (0.207)
		2000	95.6	0.242 (0.037)	93.0	0.215 (0.034)	94.8	0.353 (0.165)
	0.10	200	84.9	0.400 (0.072)	79.6	0.350 (0.064)	89.4	0.584 (0.323)
		500	92.2	0.311 (0.053)	87.8	0.273 (0.047)	89.7	0.419 (0.194)
		1000	90.0	0.246 (0.040)	84.7	0.216 (0.036)	90.3	0.341 (0.163)
		2000	91.6	0.204 (0.033)	86.9	0.178 (0.029)	89.8	0.279 (0.131)
f_4	0.05	200	97.0	1.260 (0.188)	94.4	1.122 (0.170)	89.2	1.234 (0.621)
		500	99.8	1.086 (0.133)	99.5	0.968 (0.120)	92.8	1.077 (0.538)
		1000	99.0	0.869 (0.100)	97.9	0.774 (0.092)	94.4	0.927 (0.437)
		2000	99.7	0.728 (0.083)	98.2	0.649 (0.075)	94.8	0.800 (0.405)
	0.10	200	92.8	1.063 (0.162)	87.9	0.932 (0.144)	83.4	0.974 (0.490)
		500	99.3	0.917 (0.115)	98.4	0.805 (0.103)	87.0	0.850 (0.424)
		1000	97.0	0.733 (0.088)	93.8	0.644 (0.079)	89.1	0.731 (0.345)
		2000	97.2	0.615 (0.072)	94.9	0.540 (0.064)	89.3	0.631 (0.320)
f_5	0.05	200	94.0	0.540 (0.093)	90.2	0.480 (0.083)	95.4	0.798 (0.398)
		500	97.4	0.432 (0.068)	95.6	0.384 (0.062)	94.2	0.597 (0.316)
		1000	96.5	0.338 (0.051)	93.4	0.301 (0.047)	95.1	0.491 (0.265)
		2000	96.9	0.280 (0.042)	94.4	0.249 (0.038)	95.5	0.401 (0.195)
	0.10	200	88.2	0.454 (0.079)	82.6	0.398 (0.070)	90.3	0.629 (0.314)
		500	94.4	0.364 (0.059)	90.6	0.319 (0.053)	89.1	0.471 (0.249)
		1000	92.2	0.285 (0.045)	87.8	0.249 (0.040)	91.1	0.387 (0.209)
		2000	93.4	0.236 (0.036)	89.2	0.207 (0.033)	90.1	0.317 (0.154)

intervals are generally shorter and exhibit less variation compared to DHZ's confidence intervals. The variation observed in our method across different regression functions may be attributed to the roughness of the partition used. In practical applications, a slightly larger J can be set, provided that the credible intervals can be computed within a reasonable time.

6. Proofs.

6.1. *Proof of Theorem 4.1.* For $\boldsymbol{t} \in \mathbb{R}^d$, let $j(\boldsymbol{t}) = \lceil (\boldsymbol{x}_0 + \boldsymbol{t} \circ \boldsymbol{r}_n) \circ \boldsymbol{J} \rceil$. Let

$$f_{*,c}(\boldsymbol{x}_0) = \max_{\substack{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{u} \preceq c\mathbf{1}, \\ u_k \le x_{0,k}, \\ s+1 \le k \le d}} \min_{\substack{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{v} \preceq c\mathbf{1}, \\ v_k \le 1 - x_{0,k}, \\ s+1 \le k \le d}} \frac{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{N_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]}}, \tag{6.1}$$

where γ is a positive constant to be determined later. We also introduce the notation $W_n^* = \omega_n^{-1}(f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0))$, $W_{n,c}^* = \omega_n^{-1}(f_{*,c}(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0))$,

$$\begin{aligned} W_c = \sup_{\substack{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{u} \preceq c\mathbf{1}, \\ u_k \le x_{0,k}, \\ s+1 \le k \le d}} \inf_{\substack{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{v} \preceq c\mathbf{1}, \\ v_k \le 1 - x_{0,k}, \\ s+1 \le k \le d}} \Bigg\{ & \frac{\sigma_0 H_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} \\ & + \frac{\sigma_0 H_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} + \sum_{\boldsymbol{l} \in L^*} \frac{\partial^{\boldsymbol{l}} f_0(x_0)}{(\boldsymbol{l}+\mathbf{1})!} \prod_{k=1}^s \frac{v_k^{l_k+1} - (-u_k)^{l_k+1}}{u_k + v_k} \Bigg\}, \\ W = \sup_{\substack{\boldsymbol{u} \succeq \mathbf{0}, \\ u_k \le x_{0,k}, \\ s+1 \le k \le d}} \inf_{\substack{\boldsymbol{v} \succeq \mathbf{0}, \\ v_k \le 1 - x_{0,k}, \\ s+1 \le k \le d}} \Bigg\{ & \frac{\sigma_0 H_1(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} \\ & + \frac{\sigma_0 H_2(\boldsymbol{u}, \boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u}, \boldsymbol{v})} + \sum_{\boldsymbol{l} \in L^*} \frac{\partial^{\boldsymbol{l}} f_0(x_0)}{(\boldsymbol{l}+\mathbf{1})!} \prod_{k=1}^s \frac{v_k^{l_k+1} - (-u_k)^{l_k+1}}{u_k + v_k} \Bigg\}. \end{aligned}$$

The proof of the theorem is carried out in several steps using Lemma B.1 of [54], presented as lemmas below.

LEMMA 6.1. *Under the conditions of Theorem 4.1, for every $c > 0$ and $\gamma > 0$, $\mathcal{L}(W_{n,c}^* | \mathbb{D}_n)$ converges weakly to $\mathcal{L}(W_c | H_1)$ as random probability measures.*

PROOF. For every $\boldsymbol{u}, \boldsymbol{v} \succeq \mathbf{0}$, we can write

$$\frac{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}} \theta_{\boldsymbol{j}}}{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}} - f_0(\boldsymbol{x}_0) = A_n(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\theta}) + A_n'(\boldsymbol{u}, \boldsymbol{v}) + B_n(\boldsymbol{u}, \boldsymbol{v}), \tag{6.2}$$

and then $W_{n,c}^* = \max_{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{u} \preceq c\mathbf{1}} \min_{c^{-\gamma}\mathbf{1} \preceq \boldsymbol{u} \preceq c\mathbf{1}} \{A_n(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\theta}) + A_n'(\boldsymbol{u}, \boldsymbol{v}) + B_n(\boldsymbol{u}, \boldsymbol{v})\}$, where

$$A_n(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\theta}) = \omega_n^{-1} \frac{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}} (\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}} | \mathbb{D}_n])}{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}}, \tag{6.3}$$

$$A_n'(\boldsymbol{u}, \boldsymbol{v}) = \omega_n^{-1} \frac{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}} (\mathrm{E}[\theta_{\boldsymbol{j}} | \mathbb{D}_n] - \bar{Y}|_{I_{\boldsymbol{j}}})}{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}}, \tag{6.4}$$

$$B_n(\boldsymbol{u}, \boldsymbol{v}) = \omega_n^{-1} \big(\overline{Y}|_{I_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]}} - f_0(\boldsymbol{x}_0) \big). \tag{6.5}$$

Since the max-min functional is continuous on the space $\mathbb{L}_\infty([c^{-\gamma}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma}\mathbf{1}, c\mathbf{1}])$, it suffices to show that $A_n + A_n' + B_n$ converges weakly in $\mathbb{L}_\infty([c^{-\gamma}\mathbf{1}, c\mathbf{1}] \times [c^{-\gamma}\mathbf{1}, c\mathbf{1}])$, conditional on the data $\mathbb{D}_n$. By Lemma B.2 of [54] and Lemma 6.2, we prove the weak convergence of A_n. We show that A_n' converges to zero uniformly in Lemma 6.3. The convergence of B_n is completed by combining Lemma B.2 of [54], Lemma 6.4 and Lemma 6.5. □

LEMMA 6.2. *Under the conditions of Theorem 4.1, for every $c > 0$, let $\mathbb{H}_{2,n}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\theta}) = \omega_n \sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n])$. Then $\mathbb{H}_{2,n}$ converges weakly to a centered Gaussian process H_2 in $\mathbb{L}_\infty([\boldsymbol{0}, c\boldsymbol{1}] \times [\boldsymbol{0}, c\boldsymbol{1}])$ for every $c > 0$ in P_0-probability.*

PROOF. By (3.7), Lemmas B.2, B.4 and B.5 of [54], the covariance kernel of $\mathbb{H}_{2,n}$ given $(\mathbb{D}_n, \sigma_n^2)$, is given by $\omega_n^2 \sigma_n^2 \sum_{\boldsymbol{j} \in [j(-\boldsymbol{u} \wedge \boldsymbol{u}'):j(\boldsymbol{v} \wedge \boldsymbol{v}')]} N_{\boldsymbol{j}}^2/(N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2})$, which converges in P_0-probability to $\sigma_0^2 \prod_{k=1}^s (u_k \wedge u_k' + v_k \wedge v_k') D_s(\boldsymbol{u} \wedge \boldsymbol{u}', \boldsymbol{v} \wedge \boldsymbol{v}')$. Thus finite-dimensional distributions of $\mathbb{H}_{2,n}$ converge weakly to those of a centered Gaussian process $\sigma_0 H_2$ in P_0-probability.

Next, we need to show that $\mathcal{L}(\mathbb{H}_{2,n}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{\theta}) : (\boldsymbol{u}, \boldsymbol{v}) \in [\boldsymbol{0}, c\boldsymbol{1}] \times [\boldsymbol{0}, c\boldsymbol{1}])$ is tight on $\mathbb{L}_\infty([\boldsymbol{0}, c\boldsymbol{1}] \times [\boldsymbol{0}, c\boldsymbol{1}])$ for any $c > 0$ in P_0-probability. In view of Theorem 18.14 of [53], we need to verify that, for every $\epsilon > 0$ and $\eta > 0$, there exists a finite partition $\{T_p : p \le K\}$ of $[\boldsymbol{0}, c\boldsymbol{1}] \times [\boldsymbol{0}, c\boldsymbol{1}]$ with K depending only on ϵ and η such that

$$\mathrm{P}\Big(\sup_{(\boldsymbol{u}_1, \boldsymbol{v}_1), (\boldsymbol{u}_2, \boldsymbol{v}_2) \in T_p} \big\{ \big| \mathbb{H}_{2,n}(\boldsymbol{u}_1, \boldsymbol{v}_1) - \mathbb{H}_{2,n}(\boldsymbol{u}_2, \boldsymbol{v}_2) \big| : 1 \le p \le K \big\} > \epsilon \big| \mathbb{D}_n \Big) < \eta$$

with P_0-probability tending to 1. Let $\delta > 0$, to be determined later, which depends only on ϵ and η. Let $0 = s_0 < s_1 < \cdots < s_l = c$ with $(s_{t-1}, s_t]$ of equal length at least δ and $l \le 2c/\delta$. We choose a partition $\{T_p : p \le K\}$ of $[\boldsymbol{0}, c\boldsymbol{1}] \times [\boldsymbol{0}, c\boldsymbol{1}]$ to be

$$\mathcal{P}(\delta) = \Bigg\{ \prod_{k=1}^d (s_{t_k - 1}, s_{t_k}] \times \prod_{k=1}^d (s_{r_k - 1}, s_{r_k}] : t_k, r_k \in \{1, \ldots, l\} \Bigg\}, \tag{6.6}$$

with cardinality $K = \#\mathcal{P}(\delta) = l^{2d}$. It suffices to verify that, for any $p \le K$,

$$\mathrm{P}\Big(\sup_{(\boldsymbol{u}_1, \boldsymbol{v}_1), (\boldsymbol{u}_2, \boldsymbol{v}_2) \in T_p} \big\{ \big| \mathbb{H}_{2,n}(\boldsymbol{u}_1, \boldsymbol{v}_1) - \mathbb{H}_{2,n}(\boldsymbol{u}_2, \boldsymbol{v}_2) \big| \big\} > \epsilon \big| \mathbb{D}_n \Big) < \eta \Big(\frac{\delta}{2c} \Big)^{2d}.$$

Let $\mathcal{J}(\boldsymbol{u}, \boldsymbol{v}) = [j(-\boldsymbol{u}) : j(\boldsymbol{v})]$. For $(\boldsymbol{u}_1, \boldsymbol{v}_1)$, $(\boldsymbol{u}_2, \boldsymbol{v}_2)$, we write $\mathbb{H}_{2,n}(\boldsymbol{u}_1, \boldsymbol{v}_1) - \mathbb{H}_{2,n}(\boldsymbol{u}_2, \boldsymbol{v}_2)$ as the difference of the sums of $\omega_n N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n])$ over the sets $\mathcal{J}(\boldsymbol{u}_1, \boldsymbol{v}_1) \setminus \mathcal{J}(\boldsymbol{u}_1 \wedge \boldsymbol{u}_2, \boldsymbol{v}_1 \wedge \boldsymbol{v}_2)$ and $\mathcal{J}(\boldsymbol{u}_2, \boldsymbol{v}_2) \setminus \mathcal{J}(\boldsymbol{u}_1 \wedge \boldsymbol{u}_2, \boldsymbol{v}_1 \wedge \boldsymbol{v}_2)$, after canceling out the common terms. Thus, its absolute value can be bounded by the sum of the corresponding absolute values over these two index sets. To verify tightness, it then suffices to show that

$$\mathrm{P}\Bigg(\max \Bigg\{ \omega_n \Bigg| \sum_{\mathcal{J}(\boldsymbol{u}, \boldsymbol{v}) \setminus \mathcal{J}(\boldsymbol{s_{t-1}}, \boldsymbol{s_{r-1}})} N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n]) \Bigg| : (\boldsymbol{u}, \boldsymbol{v}) \in T_p \Bigg\} > \frac{\epsilon}{4} \Big| \mathbb{D}_n \Bigg)$$

is bounded by $\eta(\delta/(2c))^{2d}/4$, with $T_p = \prod_{k=1}^d (s_{t_k - 1}, s_{t_k}] \times \prod_{k=1}^d (s_{r_k - 1}, s_{r_k}]$, for any $\boldsymbol{s_t} = (s_{t_1}, \ldots, s_{t_d})$ and $\boldsymbol{s_r} = (s_{r_1}, \ldots, s_{r_d})$.

Let $S_{(-j(-\boldsymbol{u}), j(\boldsymbol{v}))} = \sum_{\mathcal{J}(\boldsymbol{u}, \boldsymbol{v}) \setminus \mathcal{J}(\boldsymbol{s_{t-1}}, \boldsymbol{s_{r-1}})} N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n])$, a collection of random variables indexed by a $2d$-dimensional vector in a finite-index set. The negative sign in front of $j(-\boldsymbol{u})$ in the subscript of S is to make the σ-fields,

$$\mathscr{F}_{\boldsymbol{j}}^{(k)} = \begin{cases} \sigma\big\langle N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n]) : -\big(j(\boldsymbol{s_{t-1}})\big)_k < -\big(j(-\boldsymbol{u})\big)_k \le j \big\rangle & \text{if } k \le d, \\ \sigma\big\langle N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n]) : \big(j(\boldsymbol{s_{r-1}})\big)_{k-d} < \big(j(\boldsymbol{v})\big)_{k-d} \le j \big\rangle & \text{if } k > d. \end{cases}$$

increase with respect to each of the first d components in the subscript. In the sum above, all $\boldsymbol{j}$ are in $\mathcal{J}(\boldsymbol{s_t}, \boldsymbol{s_r}) \setminus \mathcal{J}(\boldsymbol{s_{t-1}}, \boldsymbol{s_{r-1}})$. We note that for every $k \le 2d$, the random sequence $\{S_{(j_1, \ldots, j_{k-1}, j, j_{k+1}, \ldots, j_{2d})}, \mathscr{F}_j^{(k)}\}$ is a martingale. Applying Lemma B.6 of [54] with $p = 4d + 2$, we can get an upper bound of the probability of the maximal deviation needed to verify tightness to be a constant multiple of

$$(\omega_n/\epsilon)^{(4d+2)} \mathrm{E}\Bigg(\Bigg| \sum_{\mathcal{J}(\boldsymbol{s_t}, \boldsymbol{s_r}) \setminus \mathcal{J}(\boldsymbol{s_t - 1}, \boldsymbol{s_r - 1})} N_{\boldsymbol{j}}(\theta_{\boldsymbol{j}} - \mathrm{E}[\theta_{\boldsymbol{j}}|\mathbb{D}_n]) \Bigg|^{4d+2} \Big| \mathbb{D}_n \Bigg). \tag{6.7}$$

Observe that $\#\mathcal{J}(\boldsymbol{s_t}, \boldsymbol{s_r}) \le \prod_k (r_{n,k} J_k (s_{t_k} + s_{r_k}) + 2)$, $\#\mathcal{J}(\boldsymbol{s_{t-1}}, \boldsymbol{s_{r-1}}) \ge \prod_k r_{n,k} J_k (s_{t_k} + s_{r_k} - 2\delta)$. As $\delta \le s_{t_k}, s_{r_k} \le c$ and $J_k \gg r_{n,k}^{-1}$, it follows that the cardinality of the index set $\mathcal{J}(\boldsymbol{s_t}, \boldsymbol{s_r}) \setminus \mathcal{J}(\boldsymbol{s_{t-1}}, \boldsymbol{s_{r-1}})$ is bounded by a multiple of

$$\prod_{k=1}^{d} r_{n,k} J_k \left(\prod_{k=1}^{d} (s_{t_k} + s_{r_k}) - \prod_{k=1}^{d} (s_{t_k} + s_{r_k} - 2\delta) \right) \le (2d\delta)(2c)^{d-1} \prod_{k=1}^{d} r_{n,k} J_k,$$

where the last inequality follows from Lemma B.7 of [54].

The variance $\sigma_n^2 N_{\boldsymbol{j}}^2 / (N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2}) \lesssim n (\prod_{k=1}^{d} J_k)^{-1}$ with P_0-probability tending to 1 by Lemma B.4 of [54]. Hence, (6.7) is bounded by a constant multiple of

$$\begin{aligned} &\epsilon^{-(4d+2)} \omega_n^{4d+2} \left(\frac{n \#(\mathcal{J}(s_{\boldsymbol{t}}, s_{\boldsymbol{r}}) \setminus \mathcal{J}(s_{\boldsymbol{t-1}}, s_{\boldsymbol{r-1}}))}{\prod_{k=1}^{d} J_k} \right)^{2d+1} \\ &\lesssim \epsilon^{-(4d+2)} \omega_n^{4d+2} \left(\prod_{k=1}^{d} r_{n,k} \right)^{2d+1} n^{2d+1} \delta^{2d+1}, \end{aligned}$$

which simplifies to $\epsilon^{-(4d+2)} \delta^{2d+1}$. With δ chosen a sufficiently small constant multiple of $\eta \epsilon^{4d+2}$, the tightness condition is verified. □

LEMMA 6.3. *Under the conditions of Theorem 4.1, $A_n'(\boldsymbol{u}, \boldsymbol{v})$ converges to 0 in P_0-probability uniformly in $(\boldsymbol{u}, \boldsymbol{v})$.*

PROOF. Let $E_n = \{a_1 n / (2 \prod_{k=1}^{d} J_k) \le N_{\boldsymbol{j}} \le 2a_2 n / (\prod_{k=1}^{d} J_k)\}$ for some $a_1, a_2 > 0$ and $\bar{\varepsilon}|_{I_{\boldsymbol{j}}} = \sum_{i \in I_{\boldsymbol{j}}} \varepsilon_i / N_{\boldsymbol{j}}$. By Lemma B.4 of [54], we have for every $T > 0$,

$$\mathrm{P}_0 \Big(\max_{\boldsymbol{j}} |\bar{\varepsilon}|_{I_{\boldsymbol{j}}}| > T \Big) \le \sum_{\boldsymbol{j}} \mathrm{P}_0 \big(|\bar{\varepsilon}|_{I_{\boldsymbol{j}}}| > T \,|\, E_n \big) + \mathrm{P}_0(E_n^c). \tag{6.8}$$

By Assumption 2 and the Marcinkiewicz—-Zygmund inequality,

$$\mathrm{E} \big(|\bar{\varepsilon}|_{I_{\boldsymbol{j}}}|^{2(\sum_{k=1}^{s} \beta_k^{-1} + 1)} \,|\, E_n \big) \lesssim \left(a_1 n \Big/ \left(2 \prod_{k=1}^{d} J_k \right) \right)^{-(\sum_{k=1}^{s} \beta_k^{-1} + 1)}.$$

Then (6.8) is bounded by a constant multiple of $(\prod_{k=1}^{d} J_k)^{\sum_{k=1}^{s} \beta_k^{-1} + 2} n^{-(\sum_{k=1}^{s} \beta_k^{-1} + 1)} + o(1)$, which tends to zero because $\prod_{k=1}^{d} J_k \ll n \omega_n = n^{(\sum_{k=1}^{s} \beta_k^{-1} + 1)/(\sum_{k=1}^{s} \beta_k^{-1} + 2)}$.

On the other hand, $\max_{\boldsymbol{j}} |\overline{f_0(\boldsymbol{X}_i)}|_{I_{\boldsymbol{j}}}| \le f_0(\mathbf{1})$. Thus, $\max_{\boldsymbol{j}} |\bar{Y}|_{I_{\boldsymbol{j}}}| = O_{\mathrm{P}_0}(1)$. Because $\mathrm{E}[\theta_{\boldsymbol{j}} | \mathbb{D}_n] = (N_{\boldsymbol{j}} \bar{Y}|_{I_{\boldsymbol{j}}} + \zeta_{\boldsymbol{j}} \lambda_{\boldsymbol{j}}^{-2}) / (N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2})$, on the event E_n,

$$\begin{aligned} |A_n'(\boldsymbol{u}, \boldsymbol{v})| &= \omega_n^{-1} \left| \frac{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}) : j(\boldsymbol{v})]} \lambda_{\boldsymbol{j}}^{-2} N_{\boldsymbol{j}} (N_{\boldsymbol{j}} + \lambda_{\boldsymbol{j}}^{-2})^{-1} (\zeta_{\boldsymbol{j}} - \bar{Y}|_{I_{\boldsymbol{j}}})}{\sum_{\boldsymbol{j} \in [j(-\boldsymbol{u}) : j(\boldsymbol{v})]} N_{\boldsymbol{j}}} \right| \\ &\lesssim \omega_n^{-1} \Big(\max_{\boldsymbol{j}} |\bar{Y}|_{I_{\boldsymbol{j}}}| + \zeta_{\boldsymbol{j}} \Big) \Big(\min_{\boldsymbol{j}} N_{\boldsymbol{j}} \Big)^{-1}, \end{aligned} \tag{6.9}$$

which is of the order of $(n\omega_n)^{-1} \prod_{k=1}^{d} J_k$ in P_0-probability. As $\prod_{k=1}^{d} J_k \ll n\omega_n$ and $\mathrm{P}_0(E_n) \to 1$, we can conclude $A'(\boldsymbol{u}, \boldsymbol{v}) \to_{\mathrm{P}_0} 0$ uniformly for any $\boldsymbol{u} \succeq \mathbf{0}$ and $\boldsymbol{v} \succeq \mathbf{0}$ provided that $\boldsymbol{x}_0 - \boldsymbol{u} \circ \boldsymbol{r}_n$ and $\boldsymbol{x}_0 + \boldsymbol{v} \circ \boldsymbol{r}_n$ in $[0, 1]^d$. □

To establish the weak convergence of B_n in $\mathbb{L}_\infty([\mathbf{0}, c\mathbf{1}] \times [\mathbf{0}, c\mathbf{1}])$, write

$$B_n(\boldsymbol{u}, \boldsymbol{v}) = \omega_n^{-1} \big(\bar{\varepsilon}|_{I_{[j(-\boldsymbol{u}) : j(\boldsymbol{v})]}} + \overline{f_0(\boldsymbol{X})}|_{I_{[j(-\boldsymbol{u}) : j(\boldsymbol{v})]}} - f_0(\boldsymbol{x}_0) \big). \tag{6.10}$$

LEMMA 6.4. *Let* $Z_{ni}(\boldsymbol{u},\boldsymbol{v}) = \omega_n \varepsilon_i \mathbb{1}_{\{\boldsymbol{X}_i \in I_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]}\}}$ *and* $\mathbb{H}_{1,n}(\boldsymbol{u},\boldsymbol{v}) = \sum_{i=1}^n Z_{ni}(\boldsymbol{u},\boldsymbol{v})$. *Under the conditions of Theorem* 4.1, $\mathbb{H}_{1,n}(\boldsymbol{u},\boldsymbol{v}) \rightsquigarrow \sigma_0 H_1(\boldsymbol{u},\boldsymbol{v})$ *in* $\mathbb{L}_\infty([\mathbf{0}, c\mathbf{1}] \times [\mathbf{0}, c\mathbf{1}])$.

LEMMA 6.5. *Under the conditions of Theorem* 4.1, *for any* $c > 0$, *uniformly in* $(\boldsymbol{u},\boldsymbol{v}) \in [\mathbf{0}, c\mathbf{1}] \times [\mathbf{0}, c\mathbf{1}]$, *we have*

$$\omega_n^{-1}\big(\overline{f_0(\boldsymbol{X})}|_{I_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]}} - f_0(\boldsymbol{x}_0)\big) \to_{\mathrm{P}_0} \sum_{\boldsymbol{l}\in L^*} \frac{\partial^{\boldsymbol{l}} f_0(\boldsymbol{x}_0)}{(\boldsymbol{l}+\mathbf{1})!} \prod_{k=1}^s \frac{v_k^{l_k+1} - (-u_k)^{l_k+1}}{u_k + v_k}.$$

LEMMA 6.6. *Under the conditions of Theorem* 4.1, *for any* $M_n \uparrow \infty$, $\Pi(|f_*(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)| > M_n\omega_n|\mathbb{D}_n) \to 0$ *in* P_0*-probability.*

With the aid of Lemma 6.6, the second condition of Proposition B.1 of [54] is verified by Lemma 6.7 in the following.

LEMMA 6.7. *Let* $\boldsymbol{u}^*$ *and* $\boldsymbol{v}^*$ *be any pair indexes such that*

$$f_*(\boldsymbol{x}_0) = \max_{\boldsymbol{u}\succeq\mathbf{0}} \min_{\boldsymbol{v}\succeq\mathbf{0}} \frac{\sum_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}\theta_{\boldsymbol{j}}}{\sum_{[j(-\boldsymbol{u}):j(\boldsymbol{v})]} N_{\boldsymbol{j}}} = \frac{\sum_{[j(-\boldsymbol{u}^*):j(\boldsymbol{v}^*)]} N_{\boldsymbol{j}}\theta_{\boldsymbol{j}}}{\sum_{[j(-\boldsymbol{u}^*):j(\boldsymbol{v}^*)]} N_{\boldsymbol{j}}}. \tag{6.11}$$

Let $\omega_n = n^{-1/(2+\sum_{k=1}^s \beta_k^{-1})}$ *and let* $\boldsymbol{r}_n = (\omega_n^{1/\beta_1}, \ldots, \omega_s^{1/\beta_s}, 1, \ldots, 1)^{\mathrm{T}}$. *Suppose that* $\boldsymbol{J}$ *satisfies* $J_k \gg r_{n,k}^{-1}$, *for each* $k = 1, \ldots, d$, *and* $\prod_{k=1}^d J_k \ll n\omega_n$. *Under Assumptions* 1 *and* 2, *there exists* $\gamma > 0$ *such that*

$$\lim_{c\to\infty} \limsup_{n\to\infty} \Pi\Big(c^{-\gamma} \le \min_{1\le k\le d}\{v_k^*\} \le \max_{1\le k\le d}\{v_k^*\} \le c|\mathbb{D}_n\Big) = 1,$$

in P_0*-probability.*

The proofs of Lemmas 6.4–6.7 are provided in [54].

The proof of Theorem 4.1 can now be completed. Using arguments similar to Proposition 7 of [38], it can be verified that $\mathrm{P}(W_c \ne W) \to 0$ as $c \to \infty$. Hence, the proof follows by an application of Lemma B.1 of [54].

6.2. *Proof of Proposition* 4.1. This can be shown by the self-similarity property of Gaussian processes H_1 and H_2: for $\boldsymbol{t} \in \mathbb{R}^d_{>0}$ such that $t_{s+1} = \cdots = t_d = 1$, we have that $H_i(\boldsymbol{t}\circ\boldsymbol{u}, \boldsymbol{t}\circ\boldsymbol{v}) =_d (\prod_{j=1}^s t_j)^{1/2} H_i(\boldsymbol{u},\boldsymbol{v})$, $i = 1, 2$. By the choice of $\boldsymbol{t}$, multiplying a vector coordinatewise by $\boldsymbol{t}$ does not change the last $d - s$ coordinates, and thus $D_s(\boldsymbol{t}\circ\boldsymbol{u}, \boldsymbol{t}\circ\boldsymbol{v}) = D_s(\boldsymbol{u},\boldsymbol{v})$. Then, since a scaling of the domain does not alter suprema and infima, the expression in the limiting distribution is equal to

$$\begin{aligned}
&\sup_{\boldsymbol{u}\succeq\mathbf{0}} \inf_{\boldsymbol{v}\succeq\mathbf{0}} \left\{ \frac{\sigma_0 H_1(\boldsymbol{t}\circ\boldsymbol{u}, \boldsymbol{t}\circ\boldsymbol{v}) + \sigma_0 H_2(\boldsymbol{t}\circ\boldsymbol{u}, \boldsymbol{t}\circ\boldsymbol{v})}{\prod_{k=1}^s (t_k u_k + t_k v_k) D_s(\boldsymbol{u},\boldsymbol{v})} \right. \\
&\qquad \left. + \sum_{k=1}^s \left[\frac{\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!} \cdot \frac{(t_k v_k)^{\beta_k+1} - (-t_k u_k)^{\beta_k+1}}{t_k u_k + t_k v_k} \right] \right\} \\
&=_d \sup_{\boldsymbol{u}\succeq\mathbf{0}} \inf_{\boldsymbol{v}\succeq\mathbf{0}} \left\{ \left(\sigma_0^{-2} \prod_{j=1}^s t_j \right)^{-1/2} \frac{H_1(\boldsymbol{u},\boldsymbol{v}) + H_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^s (u_k + v_k) D_s(\boldsymbol{u},\boldsymbol{v})} \right. \\
&\qquad \left. + \sum_{k=1}^s \left[\frac{t_k^{\beta_k} \partial_k^{\beta_k} f_0(\boldsymbol{x}_0)}{(\beta_k+1)!} \cdot \frac{v_k^{\beta_k+1} - (-u_k)^{\beta_k+1}}{u_k + v_k} \right] \right\}.
\end{aligned}$$

By equating $(\sigma_0^{-2}\prod_{j=1}^s t_j)^{-1/2}$ to $t_k^{\beta_k}\partial_k^{\beta_k} f_0(\boldsymbol{x}_0)/(\beta_k+1)!$ for each $k=1,\ldots,s$, we can find the solution t_k to the system of equations, and also the common factor $A_{\boldsymbol{\beta}}$ as stated in the proposition.

If $s=d$, then $D_d(\boldsymbol{u},\boldsymbol{v}) = g(\boldsymbol{x}_0)$ and $H_i(\boldsymbol{u},\boldsymbol{v}) =_d \sqrt{g(\boldsymbol{x}_0)}\tilde{H}_i(\boldsymbol{u},\boldsymbol{v})$. For $\boldsymbol{t}\in\mathbb{R}^d_{>0}$, $\tilde{H}_i(\boldsymbol{t}\circ\boldsymbol{u},\boldsymbol{t}\circ\boldsymbol{v}) =_d (\prod_{j=1}^d t_j)^{1/2}\tilde{H}_i(\boldsymbol{u},\boldsymbol{v})$, $i=1,2$. Hence by self-similarity, the last expression reduces to

$$\begin{aligned}
&\sup_{\boldsymbol{u}\succeq\boldsymbol{0}}\inf_{\boldsymbol{v}\succeq\boldsymbol{0}}\Bigg\{\frac{\sigma_0}{\sqrt{g(\boldsymbol{x}_0)}}\Bigg(\frac{\tilde{H}_1(\boldsymbol{t}\circ\boldsymbol{u},\boldsymbol{t}\circ\boldsymbol{v})}{\prod_{k=1}^d(t_ku_k+t_kv_k)}+\frac{\tilde{H}_2(\boldsymbol{t}\circ\boldsymbol{u},\boldsymbol{t}\circ\boldsymbol{v})}{\prod_{k=1}^d(t_ku_k+t_kv_k)}\Bigg)\\
&\qquad+\sum_{k=1}^d\Bigg[\frac{\partial_k^{\beta_k}f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\cdot\frac{(t_kv_k)^{\beta_k+1}-(-t_ku_k)^{\beta_k+1}}{t_ku_k+t_kv_k}\Bigg]\Bigg\}\\
&=_d\sup_{\boldsymbol{u}\succeq\boldsymbol{0}}\inf_{\boldsymbol{v}\succeq\boldsymbol{0}}\Bigg\{\sqrt{\frac{\sigma_0^2}{g(\boldsymbol{x}_0)\prod_{j=1}^d t_j}}\Bigg(\frac{\tilde{H}_1(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}+\frac{\tilde{H}_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}\Bigg)\\
&\qquad+\sum_{k=1}^d\Bigg[\frac{t_k^{\beta_k}\partial_k^{\beta_k}f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\cdot\frac{v_k^{\beta_k+1}-(-u_k)^{\beta_k+1}}{u_k+v_k}\Bigg]\Bigg\}.
\end{aligned}$$

By exploring the equation system for t_k as follows:

$$\sqrt{\frac{\sigma_0^2}{g(\boldsymbol{x}_0)\prod_{j=1}^d t_j}}=\frac{t_k^{\beta_k}\partial_k^{\beta_k}f_0(\boldsymbol{x}_0)}{(\beta_k+1)!}\quad\text{for } k=1,\ldots,d,$$

we can find the common factor $\tilde{A}_{\boldsymbol{\beta}}$ in a similar way of solving a set of equations.

PROOF OF PROPOSITION 4.2. For $0\le z\le 1$,

$$\begin{aligned}
\mathrm{P}(Z_B^{(1)}\le z)&=\mathrm{P}(1-Z_B^{(1)}\ge 1-z)\\
&=\mathrm{P}\Big(\mathrm{P}\Big(-\sup_{\boldsymbol{u}\succeq\boldsymbol{0}}\inf_{\boldsymbol{v}\succeq\boldsymbol{0}}\tilde{U}(\boldsymbol{u},\boldsymbol{v})\le 0\big|\tilde{H}_1\Big)\ge 1-z\Big)\\
&=\mathrm{P}\Big(\mathrm{P}\Big(\inf_{\boldsymbol{u}\succeq\boldsymbol{0}}\sup_{\boldsymbol{v}\succeq\boldsymbol{0}}[-\tilde{U}(\boldsymbol{u},\boldsymbol{v})]\le 0\big|\tilde{H}_1\Big)\ge 1-z\Big).
\end{aligned}$$

Note that $\tilde{H}_i(\boldsymbol{u},\boldsymbol{v}) =_d \tilde{H}_i(\boldsymbol{v},\boldsymbol{u})$ and $\tilde{H}_i =_d -\tilde{H}_i$ for $i=1,2$. Denote $\tilde{H}_1^* = -\tilde{H}_1$. Then we have

$$\begin{aligned}
&\mathrm{P}\Big(\inf_{\boldsymbol{u}\succeq\boldsymbol{0}}\sup_{\boldsymbol{v}\succeq\boldsymbol{0}}[-\tilde{U}(\boldsymbol{u},\boldsymbol{v})]\le 0\big|\tilde{H}_1\Big)\\
&=_d\mathrm{P}\Bigg(\inf_{\boldsymbol{u}\succeq\boldsymbol{0}}\sup_{\boldsymbol{v}\succeq\boldsymbol{0}}\Bigg\{\frac{\tilde{H}_1^*(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}+\frac{-\tilde{H}_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}\\
&\qquad+\sum_{k=1}^d\frac{-v_k^{\beta_k+1}+(-u_k)^{\beta_k+1}}{u_k+v_k}\Bigg\}\le 0\Big|-\tilde{H}_1^*\Bigg)\\
&=_d\mathrm{P}\Bigg(\inf_{\boldsymbol{u}\succeq\boldsymbol{0}}\sup_{\boldsymbol{v}\succeq\boldsymbol{0}}\Bigg\{\frac{\tilde{H}_1^*(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}+\frac{\tilde{H}_2(\boldsymbol{u},\boldsymbol{v})}{\prod_{k=1}^d(u_k+v_k)}+\sum_{k=1}^d\frac{u_k^{\beta_k+1}-v_k^{\beta_k+1}}{u_k+v_k}\Bigg\}\le 0\Big|\tilde{H}_1^*\Bigg)
\end{aligned}$$

$$=_d \mathrm{P}\left(\inf_{\boldsymbol{u}\succeq\mathbf{0}}\sup_{\boldsymbol{v}\succeq\mathbf{0}}\left\{\frac{\tilde{H}_1^*(\boldsymbol{v},\boldsymbol{u})}{\prod_{k=1}^d(u_k+v_k)}+\frac{\tilde{H}_2(\boldsymbol{v},\boldsymbol{u})}{\prod_{k=1}^d(u_k+v_k)}+\sum_{k=1}^d\frac{u_k^{\beta_k+1}-v_k^{\beta_k+1}}{u_k+v_k}\right\}\le 0\Big|\tilde{H}_1^*\right)$$

$$=\mathrm{P}\Big(\inf_{\boldsymbol{u}\succeq\mathbf{0}}\sup_{\boldsymbol{v}\succeq\mathbf{0}}\tilde{U}(\boldsymbol{v},\boldsymbol{u})\le 0\Big|\tilde{H}_1^*\Big).$$

Hence, $\mathrm{P}(Z_B^{(1)}\le z)=\mathrm{P}(Z_B^{(2)}\ge 1-z)$. The symmetry of the distribution of $Z_B^{(3)}$ holds by similar arguments. □

Acknowledgments. The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding. The second author was supported in part by NSF Grant number DMS-1916419.

SUPPLEMENTARY MATERIAL

Supplement: Additional proofs and supporting lemmas (DOI: 10.1214/23-AOS2298SUPP; .pdf). We provide remaining proofs and all the supporting lemmas and propositions in the Supplementary Material [54].

REFERENCES

[1] AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Stat.* **26** 641–647. MR0073895 https://doi.org/10.1214/aoms/1177728423

[2] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693 https://doi.org/10.1214/009053606000001578

[3] BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743 https://doi.org/10.1214/aos/1015345959

[4] BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions. The Theory and Application of Isotonic Regression. Wiley Series in Probability and Mathematical Statistics*. Wiley, London–Sydney. MR0326887

[5] BELLEC, P. C. (2018). Sharp oracle inequalities for least squares estimators in shape restricted regression. *Ann. Statist.* **46** 745–780. MR3782383 https://doi.org/10.1214/17-AOS1566

[6] BHAUMIK, P. and GHOSAL, S. (2015). Bayesian two-step estimation in differential equation models. *Electron. J. Stat.* **9** 3124–3154. MR3453972 https://doi.org/10.1214/15-EJS1099

[7] BHAUMIK, P. and GHOSAL, S. (2017). Efficient Bayesian estimation and uncertainty quantification in ordinary differential equation models. *Bernoulli* **23** 3537–3570. MR3654815 https://doi.org/10.3150/16-BEJ856

[8] BHAUMIK, P. and GHOSAL, S. (2017). Bayesian inference for higher-order ordinary differential equation models. *J. Multivariate Anal.* **157** 103–114. MR3641739 https://doi.org/10.1016/j.jmva.2017.03.003

[9] BHAUMIK, P., SHI, W. and GHOSAL, S. (2022). Two-step Bayesian methods for generalized regression driven by partial differential equations. *Bernoulli* **28** 1625–1647. MR4411505 https://doi.org/10.3150/21-bej1363

[10] BRUNK, H. D. (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Stat.* **26** 607–616. MR0073894 https://doi.org/10.1214/aoms/1177728420

[11] BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos., Indiana Univ., Bloomington, Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

[12] CAI, B. and DUNSON, D. B. (2007). Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies. *J. Amer. Statist. Assoc.* **102** 1158–1171. MR2412540 https://doi.org/10.1198/016214506000000942

[13] CAI, T. T., LOW, M. G. and XIA, Y. (2013). Adaptive confidence intervals for regression functions under shape constraints. *Ann. Statist.* **41** 722–750. MR3099119 https://doi.org/10.1214/12-AOS1068

[14] CHAKRABORTY, M. and GHOSAL, S. (2021). Coverage of credible intervals in nonparametric monotone regression. *Ann. Statist.* **49** 1011–1028. MR4255117 https://doi.org/10.1214/20-aos1989
[15] CHAKRABORTY, M. and GHOSAL, S. (2021). Convergence rates for Bayesian estimation and testing in monotone regression. *Electron. J. Stat.* **15** 3478–3503. MR4280172 https://doi.org/10.1214/21-ejs1861
[16] CHAKRABORTY, M. and GHOSAL, S. (2021). Bayesian inference on monotone regression quantile: Coverage and rate acceleration. Preprint.
[17] CHAKRABORTY, M. and GHOSAL, S. (2022). Rates and coverage for monotone densities using projection-posterior. *Bernoulli* **28** 1093–1119. MR4388931 https://doi.org/10.3150/21-bej1379
[18] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2015). On risk bounds in isotonic and other shape restricted regression problems. *Ann. Statist.* **43** 1774–1800. MR3357878 https://doi.org/10.1214/15-AOS1324
[19] CHATTERJEE, S., GUNTUBOYINA, A. and SEN, B. (2018). On matrix estimation under monotonicity constraints. *Bernoulli* **24** 1072–1100. MR3706788 https://doi.org/10.3150/16-BEJ865
[20] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525 https://doi.org/10.1214/aos/1176349157
[21] DENG, H., HAN, Q. and ZHANG, C.-H. (2021). Confidence intervals for multiple isotonic regression and other monotone models. *Ann. Statist.* **49** 2021–2052. MR4319240 https://doi.org/10.1214/20-aos2025
[22] DENG, H. and ZHANG, C.-H. (2020). Isotonic regression in multi-dimensional spaces and graphs. *Ann. Statist.* **48** 3672–3698. MR4185824 https://doi.org/10.1214/20-AOS1947
[23] DÜMBGEN, L. (2003). Optimal confidence bands for shape-restricted curves. *Bernoulli* **9** 423–449. MR1997491 https://doi.org/10.3150/bj/1065444812
[24] DÜMBGEN, L. and JOHNS, R. B. (2004). Confidence bands for isotonic median curves using sign tests. *J. Comput. Graph. Statist.* **13** 519–533. MR2063998 https://doi.org/10.1198/1061860043506
[25] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961 https://doi.org/10.1214/aos/996986504
[26] DUROT, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *Ann. Statist.* **35** 1080–1104. MR2341699 https://doi.org/10.1214/009053606000001497
[27] DUROT, C., KULIKOV, V. N. and LOPUHAÄ, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *Ann. Statist.* **40** 1578–1608. MR3015036 https://doi.org/10.1214/12-AOS1015
[28] FOKIANOS, K., LEUCHT, A. and NEUMANN, M. H. (2020). On integrated L^1 convergence rate of an isotonic regression estimator for multivariate observations. *IEEE Trans. Inf. Theory* **66** 6389–6402. MR4173546 https://doi.org/10.1109/TIT.2020.3013390
[29] GHOSAL, S., SEN, A. and VAN DER VAART, A. W. (2000). Testing monotonicity of regression. *Ann. Statist.* **28** 1054–1082. MR1810919 https://doi.org/10.1214/aos/1015956707
[30] GIJBELS, I., HALL, P., JONES, M. C. and KOCH, I. (2000). Tests for monotonicity of a regression mean with guaranteed level. *Biometrika* **87** 663–673. MR1789816 https://doi.org/10.1093/biomet/87.3.663
[31] GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. MR0093415 https://doi.org/10.1080/03461238.1956.10414944
[32] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (*Berkeley, Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052
[33] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints*: *Estimators, Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. MR3445293 https://doi.org/10.1017/CBO9781139020893
[34] GROENEBOOM, P. and JONGBLOED, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875 https://doi.org/10.1214/15-AOS1335
[35] HALL, P. and HECKMAN, N. E. (2000). Testing for monotonicity of a regression mean by calibrating for linear functions. *Ann. Statist.* **28** 20–39. MR1762902 https://doi.org/10.1214/aos/1016120363
[36] HAN, Q. (2021). Set structured global empirical risk minimizers are rate optimal in general dimensions. *Ann. Statist.* **49** 2642–2671. MR4338378 https://doi.org/10.1214/21-aos2049
[37] HAN, Q., WANG, T., CHATTERJEE, S. and SAMWORTH, R. J. (2019). Isotonic regression in general dimensions. *Ann. Statist.* **47** 2440–2471. MR3988762 https://doi.org/10.1214/18-AOS1753
[38] HAN, Q. and ZHANG, C.-H. (2020). Limit distribution theory for block estimators in multiple isotonic regression. *Ann. Statist.* **48** 3251–3282. MR4185808 https://doi.org/10.1214/19-AOS1928
[39] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. MR1334065
[40] HUANG, Y. and ZHANG, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274. MR1311975 https://doi.org/10.1214/aos/1176325628

[41] KOSOROK, M. R. (2008). Bootstrapping in Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research*: *Festschrift in Honor of Professor Pranab K. Sen*. *Inst. Math. Stat.* (*IMS*) *Collect.* **1** 282–292. IMS, Beachwood, OH. MR2462212 https://doi.org/10.1214/193940307000000202

[42] KULIKOV, V. N. and LOPUHAÄ, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *Ann. Statist.* **33** 2228–2255. MR2211085 https://doi.org/10.1214/009053605000000462

[43] LIN, L. and DUNSON, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika* **101** 303–317. MR3215349 https://doi.org/10.1093/biomet/ast063

[44] NEELON, B. and DUNSON, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60** 398–406. MR2066274 https://doi.org/10.1111/j.0006-341X.2004.00184.x

[45] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

[46] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR0961262

[47] SAARELA, O. and ARJAS, E. (2011). A method for Bayesian monotonic multiple regression. *Scand. J. Stat.* **38** 499–513. MR2833843 https://doi.org/10.1111/j.1467-9469.2010.00716.x

[48] SALOMOND, J.-B. (2014). Concentration rate and consistency of the posterior distribution for selected priors under monotonicity constraints. *Electron. J. Stat.* **8** 1380–1404. MR3263126 https://doi.org/10.1214/14-EJS929

[49] SALOMOND, J.-B. (2014). Adaptive Bayes test for monotonicity. In *The Contribution of Young Researchers to Bayesian Statistics*. *Springer Proc. Math. Stat.* **63** 29–33. Springer, Cham. MR3133254 https://doi.org/10.1007/978-3-319-02084-6_7

[50] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812 https://doi.org/10.1214/13-AOS1089

[51] SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 https://doi.org/10.1214/09-AOS777

[52] SHIVELY, T. S., SAGER, T. W. and WALKER, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 159–175. MR2655528 https://doi.org/10.1111/j.1467-9868.2008.00677.x

[53] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 https://doi.org/10.1017/CBO9780511802256

[54] WANG, K. and GHOSAL, S. (2023). Supplement to "Coverage of credible intervals in Bayesian multivariate isotonic regression." https://doi.org/10.1214/23-AOS2298SUPP

[55] WANG, X. (2008). Bayesian free-knot monotone cubic spline regression. *J. Comput. Graph. Statist.* **17** 373–387. MR2421700 https://doi.org/10.1198/106186008X321077

[56] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

[57] ZHANG, C.-H. (2002). Risk bounds in isotonic regression. *Ann. Statist.* **30** 528–555. MR1902898 https://doi.org/10.1214/aos/1021379864