

Posterior Concentration Rates for Bayesian Penalized Splines

Paul Bach^{*} and Nadja Klein[†]

Abstract. Despite their widespread use in practice, the asymptotic properties of Bayesian penalized splines have not been investigated so far. We close this gap and study posterior concentration rates for Bayesian penalized splines in a Gaussian nonparametric regression model. A key feature of the approach is the hyperprior on the smoothing variance, which allows for a data-driven amount of smoothing but complicates the theoretical analysis considerably as it destroys conjugacy and precludes analytic expressions for the posterior moments. To derive our theoretical results, we rely on several new concepts including a carefully defined proper version of the partially improper penalized splines prior as well as an innovative spline estimator that projects the observations onto the first basis functions of a Demmler-Reinsch basis. Our results show that posterior concentration at near optimal rate can be achieved if the order of the penalty matches the regularity of the unknown function and if the hyperprior on the smoothing variance strikes a fine balance between oversmoothing and undersmoothing, which can for instance be met by a Weibull hyperprior with shape parameter $1/2$. We complement our theoretical results with empirical evidence demonstrating the overall competitive performance of Bayesian penalized splines in practice.

Keywords: Bayesian smoothing, Demmler-Reinsch, Kullback-Leibler and testing, nonparametric regression, posterior contraction.

1 Introduction

Bayesian penalized splines are highly popular in applied regression modeling: Since their introduction in the early 2000's by Berry et al. (2002); Lang and Brezger (2004), they have been applied in numerous settings ranging from nonparametric and additive regression over generalized additive models, additive survival analysis and multivariate response models to structured additive distributional and quantile regression (e.g., Lang and Brezger, 2004; Brezger and Lang, 2006; Hennerfeind et al., 2006; Lang et al., 2003; Klein et al., 2015; Waldmann et al., 2013, amongst many others). Further applications include measurement error models and transformation models (Berry et al., 2002; Song and Lu, 2012); see also Harezlak et al. (2018); Fahrmeir et al. (2021) for an overview and real data examples.

The basic idea of the approach is to expand the unknown function to be estimated in a rich spline basis with a large number of knots and to place a smoothness prior on the coefficients to prevent overfitting. The overall amount of regularization is controlled in a data-driven manner through a hyperprior $p(\tau)$ on the smoothing variance τ .

^{*}Faculty of Statistics, TU Dortmund University, Germany, paul.bach@tu-dortmund.de

[†]Scientific Computing Center, Karlsruhe Institute of Technology, Germany, nadja.klein@kit.edu

Despite its overwhelming success in applied statistics, the theoretical understanding of the approach is lagging far behind. In particular, the asymptotic properties of Bayesian penalized splines have not been investigated so far. In this paper, we make an important step to close this gap in the literature and study posterior concentration rates for Bayesian penalized splines in a Gaussian nonparametric regression model.

The derivation of posterior concentration rates for Bayesian penalized splines is already for the Gaussian response model quite challenging. This is because of the hyperprior $p(\tau)$ on the smoothing variance, which is a key feature of the approach and both a blessing and a curse: From a practical perspective, the hyperprior is a blessing as it allows for a data-driven amount of smoothing. From a theoretical perspective, however, the hyperprior is a curse as it destroys conjugacy and precludes analytic expressions for the posterior moments, which considerably complicates the asymptotic analysis.

The derivation of posterior concentration rates is generally much more involved if analytic expressions for the posterior moments are unavailable (cf. Ghosal and van der Vaart, 2017, Chapter 8). In this case one cannot use a “direct approach”, where one shows that the posterior mean converges to the truth and that the posterior variance vanishes sufficiently fast. To overcome this difficulty, we use the Kullback-Leibler (KL) and testing strategy for the proof of our main posterior concentration result. This is a general strategy to prove posterior concentration results that does not require analytic expressions for the posterior moments (Schwartz, 1965; Ghosal and van der Vaart, 2007). While the KL and testing strategy itself is well-established as a general proof technique, we need to introduce several completely new tools when applying it in the present context of Bayesian penalized splines.

First, the KL and testing strategy requires a proper prior but the penalized splines prior is partially improper. Previous definitions of a proper prior in the literature are based on a spectral decomposition of the roughness penalty matrix (Wand and Ormerod, 2008), but these are not well suited for the derivation of posterior concentration rates. Therefore, we introduce a new definition of a proper prior that includes an additional penalty matrix based on empirical projections, which allows us to use the KL and testing strategy in the present setting.

Second, the KL and testing strategy requires the existence of a suitable test. To this end, one can either verify a metric entropy condition or explicitly construct a test, which is sometimes more convenient (Ghosal and van der Vaart, 2007). Given the close relation of Bayesian penalized splines and frequentist penalized splines, it seems natural to construct a test on the basis of the frequentist penalized splines estimator. This estimator was introduced by O’Sullivan (1986) and its asymptotic behavior has been studied extensively (Claeskens et al., 2009; Kauermann et al., 2009; Xiao, 2019; Huang and Su, 2021). However, the test based on the frequentist penalized splines estimator is not well suited for the asymptotic analysis of Bayesian penalized splines because the type II error probability of this test is very difficult to control. Therefore, our test relies on a new spline estimator, which we refer to as *truncated DR estimator*. This estimator projects the observations onto the first basis functions of a *Demmler-Reinsch (DR) basis*, which is a special orthonormal basis of the splines. As a byproduct of our posterior concentration results, we establish that

the truncated DR estimator is able to achieve the optimal rate of convergence (Stone, 1982).

1.1 Related Literature

Several papers have derived posterior concentration rates for related Bayesian spline regression approaches: In Section 7.7 of their seminal paper on posterior concentration rates for non-iid observations, Ghosal and van der Vaart (2007) consider a Gaussian nonparametric regression model and expand the unknown function in a relatively small B-spline basis with iid standard Gaussian priors on the coefficients. The work of de Jonge and van Zanten (2012) extends this approach to multivariate functions using tensor product B-splines. A similar approach is used by Yoo and Ghosal (2016), who additionally consider contraction with respect to the uniform norm. Yoo and Ghosal (2016) use a Normal-Inverse-Gamma prior, which allows them to exploit conjugacy for their derivation of posterior concentration rates. Similarly, Serra and Krivobokova (2017) exploit conjugacy for their investigation of empirical Bayesian smoothing splines.

Due to the hyperprior $p(\tau)$ on the smoothing variance, we cannot exploit conjugacy for our derivation of posterior concentration rates and the formal structure of the problem that we face shows parallels to that of shrinkage priors in high-dimensional linear regression models. Asymptotic results for shrinkage priors have been established by Armagan et al. (2013); Song and Liang (2023), both papers using the KL and testing strategy. Wei et al. (2020) and Bai et al. (2022) build upon this work and establish posterior concentration results for splines in sparse additive models. However, none of the cited papers has studied Bayesian penalized splines with a roughness penalty and a hyperprior on the smoothing variance.

The main practical advantage of Bayesian penalized splines is that we do not need to choose the “right” number of spline knots to obtain good results in practice. We simply choose a sufficiently large number of knots and let the penalized splines prior with roughness penalty and hyperprior on the smoothing variance prevent overfitting.

A closely related Bayesian strategy is to place a hyperprior on the number of spline knots and optionally also on their locations. This approach has been investigated theoretically by de Jonge and van Zanten (2012); Belitser and Serra (2014) where it was shown that the approach even yields rate-adaptive posterior concentration. A very general rate-adaptive posterior concentration result was established by Shen and Ghosal (2015). These authors study random series priors with a random number of basis functions and their results cover in particular Gaussian nonparametric regression with B-splines and a hyperprior on the number of knots.

However, Bayesian penalized splines have been shown to be competitive to splines with a hyperprior on the knots in simulations (Lang and Brezger, 2004) and they are also more tractable from a computational point of view, especially for non-Gaussian response models. Thanks to the continuous nature of all parameters, Bayesian penalized splines are also compatible with **Stan** (Carpenter et al., 2017) making the approach very accessible to practitioners.

1.2 Main Contributions

- We close a gap in the literature and study posterior concentration rates for Bayesian penalized splines. This contribution is of high practical relevance as Bayesian penalized splines are widely used in modern statistical analyses. In our main theorem, Theorem 3.1, we establish sufficient conditions for posterior concentration at near optimal rate.
- Based on our theoretical results we can identify specific hyperpriors $p(\tau)$ for the smoothing variance that lead to posterior concentration at near optimal rate. To the best of our knowledge, these are the first theoretical guarantees in the literature ensuring optimal asymptotic performance for Bayesian penalized splines.
- Our theoretical results rely on several new concepts, namely a novel proper version of the partially improper penalized splines prior based on empirical projections as well as an innovative spline estimator. We prove that this truncated DR estimator achieves the optimal rate of convergence.
- To define the truncated DR estimator, we establish the concept of DR bases for penalized splines in a rigorous manner. Similar concepts are scattered throughout the literature (Nychka and Cummins, 1996; Ruppert et al., 2003; Claeskens et al., 2009), but a comprehensive introduction has been missing so far.
- We complement our theoretical results with empirical evidence demonstrating the competitive performance of Bayesian penalized splines in practice. Thereby, we use the “scale-dependent” Weibull hyperprior with shape 1/2 suggested by Klein and Kneib (2016a).

1.3 Outline

This paper is structured as follows: In Section 2 we formally introduce Bayesian penalized splines. Our main posterior concentration result, Theorem 3.1, is contained in the subsequent Section 3. Section 4 describes how our results can be extended to cover an unknown residual variance and multivariate function estimation. Section 5 presents the results of a simulation study, which demonstrates the competitive performance of Bayesian penalized splines in practice. The final Section 6 concludes with a discussion. The Supplement (Bach and Klein, 2025) contains proofs for our theoretical results and several auxiliary results.

2 Bayesian Penalized Spline Regression

Throughout, we consider the Gaussian nonparametric regression model

$$Y_i = f_0(x_i) + \varepsilon_i, \quad \varepsilon_i \stackrel{iid}{\sim} N_1(0, \sigma_0^2), \quad i = 1, \dots, n, \quad (1)$$

with deterministic design points $x_1, \dots, x_n \in [0, 1]$, where the aim is to estimate the unknown regression function $f_0 : [0, 1] \rightarrow \mathbb{R}$.

In the frequentist context, Xiao (2019) distinguishes three types of penalized splines: *T-splines* (Ruppert, 2002) use the truncated power series basis and a ridge penalty, while *O-splines* (O’Sullivan, 1986; Wand and Ormerod, 2008) use B-splines and the integrated squared q th derivative as penalty. *P-splines* (Eilers and Marx, 1996) are closely related to O-splines and use the sum of squared finite differences of adjacent B-spline coefficients as roughness penalty. This distinction is also valid in the Bayesian context, where we have *Bayesian O-splines* (Wand and Ormerod, 2008; Harezlak et al., 2018), *Bayesian P-splines* (Lang and Brezger, 2004) and *Bayesian T-splines* (Berry et al., 2002). We focus on Bayesian O-splines throughout, but our results can easily be carried over to Bayesian P-splines and Bayesian T-splines.

2.1 Preliminaries

To estimate f_0 , we use the splines $\mathcal{S}(m, \xi)$, where $m \in \mathbb{N}$ is the order and $\xi \in [0, 1]^{k+2}$ is the vector of knots with entries $0 = \xi_0 < \xi_1 < \dots < \xi_k < \xi_{k+1} = 1$. Splines are piecewise polynomials that are smoothly connected at the knots, i.e., for order $m \geq 2$, which we assume throughout, the splines are defined as

$$\mathcal{S}(m, \xi) = \{f \in C^{m-2}[0, 1] : \forall j = 0, \dots, k : f|_{[\xi_j, \xi_{j+1}]} \in \mathcal{P}_{m-1}\},$$

where $\mathcal{P}_{m-1}$ is the space of polynomials of degree $m-1$ or less. It is well-known that the splines form a vector space of dimension $d = m + k$ and a convenient basis are the B-splines, which we denote by $\{B_1, \dots, B_d\}$.¹ The B-splines allow us to expand any spline $f \in \mathcal{S}(m, \xi)$ uniquely in the form $f = \sum_{j=1}^d B_j \beta_j$, where $\beta \in \mathbb{R}^d$ is the B-spline coefficient vector.

The splines $\mathcal{S}(m, \xi)$ form a d -dimensional subspace of the Banach space of continuous functions $(C[0, 1], \|\cdot\|_\infty)$ whose associated Borel σ -algebra we denote by $\mathcal{B}_C$. We denote the *vector of function evaluations* of a continuous function $f \in C[0, 1]$ by $f^n = (f(x_1), \dots, f(x_n))^T \in \mathbb{R}^n$ and the *empirical bilinear form* of two continuous functions $f, g \in C[0, 1]$ by $\langle f, g \rangle_n = n^{-1} \sum_{i=1}^n f(x_i)g(x_i) = \langle f^n, g^n \rangle / n$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product. The empirical bilinear form induces the *empirical semi-norm* $\|f\|_n = (\langle f, f \rangle_n)^{1/2} = \|f^n\|_2 / \sqrt{n}$, where $\|\cdot\|_2$ is the Euclidean norm. The restriction of the empirical bilinear form to the splines $\mathcal{S}(m, \xi)$ is an inner product if and only if the *Gramian matrix* $G_B = (\langle B_j, B_k \rangle_n)_{j,k=1,\dots,d}$ in terms of the B-splines $\{B_1, \dots, B_d\}$ is positive definite. With the B-spline design matrix $B = (B_1^n, \dots, B_d^n)$ of size $n \times d$, the Gramian matrix is equal to $G_B = B^T B / n$.

For all $q \in \{1, \dots, m-1\}$ we define the *differential bilinear form of order q* on the space of splines $\mathcal{S}(m, \xi)$ by $\langle D^q f, D^q g \rangle_{\mathcal{L}^2} = \int_0^1 D^q f(x) D^q g(x) dx$, where $D^q f$ denotes the q th derivative of a spline $f \in \mathcal{S}(m, \xi)$. The differential bilinear forms induce the *differential semi-norms* $\|D^q f\|_{\mathcal{L}^2} = (\langle D^q f, D^q f \rangle_{\mathcal{L}^2})^{1/2}$, $q \in \{1, \dots, m-1\}$, which we use to quantify the roughness of a spline.

¹To avoid ambiguity, we always refer to the normalized B-splines that arise by repeating the boundary knots m times (see, e.g., Xiao, 2019, Section 2.1, for a precise definition).

We write $R_B^{(q)} = (\langle D^q B_j, D^q B_k \rangle_{\mathcal{L}^2})_{j,k=1,\dots,d}$ for the *roughness penalty matrix* of order q in terms of the B-splines $\{B_1, \dots, B_d\}$. The matrix $R_B^{(q)}$ plays an important role as it allows us to express the roughness of a spline $f = \sum_{j=1}^d B_j \beta_j$ in the form

$$\|D^q f\|_{\mathcal{L}^2}^2 = \int_0^1 (D^q f(x))^2 dx = \beta^\top R_B^{(q)} \beta.$$

Finally, we write $Y^n = (Y_1, \dots, Y_n)^\top$ for the vector of observations from model (1), $x^n = (x_1, \dots, x_n)^\top$ for the vector of design points and X_q for the $n \times q$ design matrix of the monomials $\{1, x, \dots, x^{q-1}\}$ evaluated at the design points. We assume that X_q has full rank and write $H^{(q)} = X_q(X_q^\top X_q)^{-1} X_q^\top$ for the corresponding projection matrix.

2.2 The Bayesian Penalized Splines Prior

The key idea of Bayesian penalized splines is to combine a relatively large number of spline knots k (in practice often 20–40) with a smoothness prior on the coefficients in order to prevent overfitting. In what follows, we provide a rigorous definition of the Bayesian penalized splines prior Π .

Definition 2.1. To define Π , we first fix a spline space $\mathcal{S}(m, \xi)$ and the corresponding B-spline basis $\{B_1, \dots, B_d\}$. With this, Π is defined as the distribution of the random spline function $f = \sum_{j=1}^d B_j \beta_j$, where the B-spline coefficient vector $\beta \in \mathbb{R}^d$ is endowed with the hierarchical smoothness prior

$$\beta \mid \tau \sim N_d \left(0, \left(R_B^{(q)} / \tau + B^\top H^{(q)} B / (n \tau_{poly}) \right)^{-1} \right), \quad (2)$$

$$\tau \sim p(\tau). \quad (3)$$

Thereby, $\tau > 0$ is the smoothing variance and $p(\tau)$ is a hyperprior yet to be specified. Moreover, $\tau_{poly} > 0$ is a positive constant.

Remark 2.2. Some remarks are in order.

- The Bayesian penalized splines prior Π depends on many different quantities which becomes evident by writing $\Pi = \Pi(m, \xi, q, p(\tau), x^n, \tau_{poly})$. Formally, we regard Π as a probability measure on the Borel sets $\mathcal{B}_C$ of the continuous functions $C[0, 1]$. This is justified as the B-spline expansion $\beta \mapsto \sum_{j=1}^d B_j \beta_j$ is Borel measurable.
- In practice, often a partially improper conditionally Gaussian prior of the form $p(\beta \mid \tau) \propto \exp(-\beta^\top R_B^{(q)} \beta / (2\tau))$, $\beta \in \mathbb{R}^d$, is used instead of (2). We include the additional penalty matrix $B^\top H^{(q)} B / (n \tau_{poly})$ for the polynomials $\mathcal{P}_{q-1}$ as this yields a proper prior, which is required for the KL and testing strategy.
- The distributional family of the hyperprior $p(\tau)$ is a controversial issue and different families have been suggested, such as an Inverse Gamma family $\tau \sim IG(a, b)$, a Gamma family $\tau \sim Ga(a, b)$, a Weibull family $\tau \sim Weibull(k, \lambda)$, a Uniform family $\tau \sim U(0, b)$ or a Scaled Beta Prime family $\tau \sim SBP(a, b, c)$ (Lang and Brezger, 2004; Klein and Kneib, 2016a; Harezlak et al., 2018). Figure 1 illustrates Π through prior function draws using a Weibull hyperprior for τ .

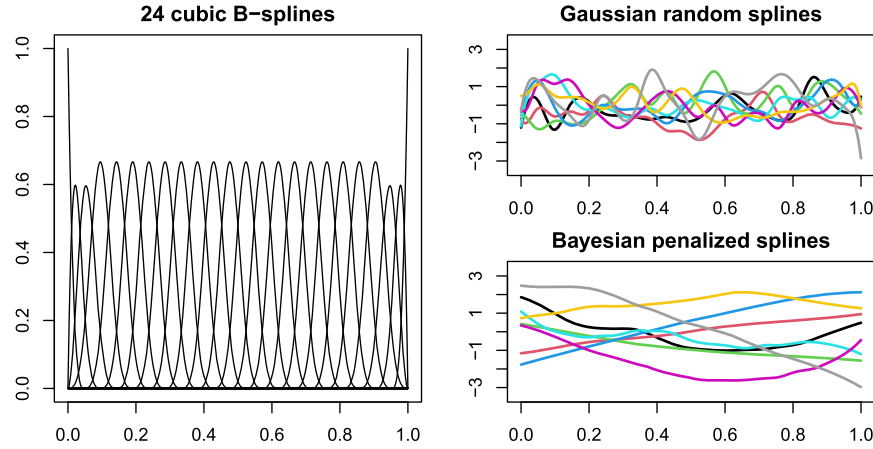


Figure 1: Illustration of the Bayesian penalized splines prior Π through prior function draws. Shown are eight prior function draws from Π (bottom right). For comparison, we also show eight prior function draws from Gaussian random splines where iid standard normal priors $\beta_j \stackrel{iid}{\sim} N_1(0, 1)$ are placed on the B-spline coefficients (top right). In both cases, the same cubic B-spline basis with 20 equidistant interior knots was used (shown on the left).

3 Posterior Concentration Rates

To investigate the asymptotic behavior of Bayesian penalized splines, we next construct a sequence of priors Π_n , $n \in \mathbb{N}$, which is supported on a sequence of spline spaces $\mathcal{S}(m, \xi_n)$, $n \in \mathbb{N}$, of increasing dimension $d_n \rightarrow \infty$. This is necessary to reduce approximation bias as we will not assume that the unknown regression function f_0 is a spline function.

Given observations $Y^n \in \mathbb{R}^n$ from model (1), we then obtain a sequence of posteriors $\Pi_n(\cdot | Y^n)$, $n \in \mathbb{N}$, via Bayes' rule

$$\Pi_n(A | Y^n) = \frac{\int_A N_n(Y^n; f^n, \sigma_0^2 I_n) \Pi_n(df)}{\int_{C[0,1]} N_n(Y^n; f^n, \sigma_0^2 I_n) \Pi_n(df)}, \quad A \in \mathcal{B}_C,$$

where $N_n(Y^n; f^n, \sigma_0^2 I_n)$ denotes the density of an n -variate Gaussian distribution with mean vector f^n and covariance matrix $\sigma_0^2 I_n$ evaluated at $Y^n \in \mathbb{R}^n$.

Ideally, this sequence of posteriors $\Pi_n(\cdot | Y^n)$, $n \in \mathbb{N}$, concentrates around f_0 in the sense that shrinking neighborhoods of f_0 receive more and more posterior mass as the sample size n grows. This idea is formalized by the notion of posterior concentration rates (Ghosal and van der Vaart, 2007; Rousseau, 2016; Ghosal and van der Vaart, 2017). Whether the posterior concentrates around f_0 depends crucially on our assumptions, which we state next.

3.1 Assumptions

We make the following assumptions for the true model (1) and our sequence of priors $\Pi_n = \Pi_n(m, \xi_n, q, p_n(\tau), x^n, \tau_{poly})$, $n \in \mathbb{N}$:

- A1) The true regression function f_0 is in $C^{m_0}[0, 1]$ for some $m_0 \in \mathbb{N}$ and the residual variance σ_0^2 is equal to some positive constant (we treat the unknown residual variance case in Section 4.1).
- A2) The order m of the spline spaces $\mathcal{S}(m, \xi_n)$ is a constant integer and exceeds the regularity of the unknown regression function f_0 , i.e., $m > m_0$.
- A3) The number of spline knots k_n diverges at rate $k_n \sim c_k n^\delta$, where $c_k > 0$ is some positive constant and $1/(2m_0 + 1) < \delta < 1$. The spline knots ξ_n are equidistant, i.e., $\xi_{n,j} = jh_n$ for $j = 0, \dots, k_n + 1$, with mesh width $h_n = 1/(k_n + 1)$.
- A4) The order of the derivative q in the roughness penalty is a constant integer and matches the regularity of the unknown regression function f_0 , i.e., $q = m_0$.
- A5) The smoothing variance τ is equipped with a Lebesgue probability density function $p_n(\tau)$, $\tau > 0$, that satisfies the following two conditions for $\tau_n^* = n^{-1/(2m_0+1)}$ and $\epsilon_n = n^{-m_0/(2m_0+1)}(\log n)^{1/2}$:
 - a) $\exists c_1 > 0$ such that $\mathbb{P}(\tau \in [c_1\tau_n^*/2, c_1\tau_n^*]) = \int_{c_1\tau_n^*/2}^{c_1\tau_n^*} p_n(\tau) d\tau \geq \exp(-n\epsilon_n^2)$ for all sufficiently large n .
 - b) $\exists c_2 > 0$ such that $\mathbb{P}(\tau \geq c_2\tau_n^*) = \int_{c_2\tau_n^*}^{\infty} p_n(\tau) d\tau \leq \exp(-5 n\epsilon_n^2)$ for all sufficiently large n .
- A6) The parameter τ_{poly} that controls the strength of the penalty for the polynomial part is equal to some positive constant or diverges to $+\infty$ at polynomial speed, i.e., $\tau_{poly} = c_{poly}n^r$ for some $c_{poly} > 0$ and $r \geq 0$.
- A7) The design points $x_1, \dots, x_n \in [0, 1]$ are distributed in a regular manner, i.e., there exists a fixed distribution function Q that is associated with a positive and continuous design density ρ such that $\|Q_n - Q\|_\infty = o(k_n^{-1})$, where $Q_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}}$, $x \in [0, 1]$, is the empirical cumulative distribution function of the design points.

Similar assumptions as A1)–A4) and A7) have been used by Claeskens et al. (2009) and Xiao (2019) for their analyses of the frequentist penalized splines estimator. Assumptions A5) and A6) are unique to the present Bayesian setting. Assumption A5) refers to the hyperprior on the smoothing variance and is particularly interesting. In Section 3.3 we investigate which of the hyperpriors $p(\tau)$ suggested in the literature satisfy Assumption A5). Before that, we state our main posterior concentration result for Bayesian penalized splines, Theorem 3.1.

3.2 Main Theorem

Theorem 3.1. *Under Assumptions A1)–A7), the posterior $\Pi_n(\cdot | Y^n)$ concentrates around f_0 at rate $\epsilon_n = n^{-m_0/(2m_0+1)}(\log n)^{1/2}$, i.e., there exists a constant $M > 0$ such that*

$$\mathbb{E}_0 \Pi_n(f \in C[0, 1] : \|f - f_0\|_n \geq M\epsilon_n | Y^n) = o(1),$$

where $\mathbb{E}_0$ is the expectation under the true model (1).

A proof of Theorem 3.1 is provided in Appendix D of the Supplement.

Remark 3.2. Under Assumption A1), the contraction rate $\epsilon_n = n^{-m_0/(2m_0+1)}(\log n)^{1/2}$ in Theorem 3.1 is *near optimal* in the sense that it is the minimax rate (Stone, 1982) times a logarithmic factor.

3.3 Hyperpriors Leading to Near Optimal Rate

Theorem 3.1 shows that we can achieve posterior concentration at near optimal rate for Bayesian penalized splines under certain conditions on the hyperprior $p(\tau)$. In the following Corollary 3.3 we investigate which of the hyperpriors $p(\tau)$ for the smoothing variance that were suggested in the literature (cf. Remark 2.2) satisfy Assumption A5) and therefore lead to posterior concentration at near optimal rate.

Corollary 3.3. *The following hyperpriors $p_n(\tau)$ satisfy Assumption A5).*

1. *Uniform:* $\tau \sim U(0, c\tau_n^*)$, where $c > 0$ is some positive constant.
2. *Gamma:* $\tau \sim \text{Ga}(a, b_n)$ with constant shape parameter $a \in (0, 1]$ and increasing rate parameter $b_n = c_b n\epsilon_n^2/\tau_n^*$, where $c_b > 0$ is some positive constant.
3. *Weibull:* $\tau \sim \text{Weibull}(k, \lambda_n)$ with constant shape parameter $k \in (0, 1]$ and increasing rate parameter $\lambda_n = c_\lambda (n\epsilon_n^2)^{1/k}/\tau_n^*$, where $c_\lambda > 0$ is a positive constant.

The following hyperpriors $p_n(\tau)$ do not satisfy Assumption A5).

1. *Inverse Gamma:* $\tau \sim \text{IG}(1, b_n)$ with scale parameter $b_n > 0$.
2. *Scaled Beta Prime:* $\tau \sim \text{SBP}(1, 1, 1/\lambda_n)$ with rate parameter $\lambda_n > 0$.

A proof of Corollary 3.3 is provided in Appendix D of the Supplement.

Remark 3.4. The results of this section prove that we can achieve posterior concentration at near optimal rate for Bayesian penalized splines. Assumption A3) shows that we can use a faster rate for the number of spline knots k_n than the slow regression spline rate $k_n \sim c_k n^{1/(2m_0+1)}$ which is commonly used for posterior concentration results in Bayesian spline regression (e.g. Ghosal and van der Vaart, 2007, Section 7.7).

Intuitively, the roughness penalty $\|D^q f\|_{\mathcal{L}^2}^2$ in the prior Π drags the posterior mass away from overly wiggly splines so that overfitting can be avoided.

From Assumption A5) it becomes clear that the hyperprior $p_n(\tau)$ on the smoothing variance τ needs to strike a fine balance to achieve posterior concentration at near optimal rate. It needs to be sufficiently thick on $[c_1\tau_n^*/2, c_1\tau_n^*]$ to avoid underfitting (oversmoothing) but at the same time it needs to be sufficiently thin on $[c_2\tau_n^*, \infty)$ to avoid overfitting (undersmoothing).

Corollary 3.3 shows that the Uniform, Gamma and Weibull hyperpriors satisfy Assumption A5) and therefore lead to posterior concentration at near optimal rate according to Theorem 3.1. For those hyperpriors that do not satisfy Assumption A5), we cannot make this conclusion. As the above conditions are only sufficient, we can, of course, not completely rule the latter hyperpriors out. Nevertheless, it seems reasonable to rather use a hyperprior for which optimal asymptotic performance is guaranteed. Armagan et al. (2013) argue similarly for the screening of shrinkage priors in a high-dimensional linear regression context.

Our results cover in particular a Weibull hyperprior with shape parameter $1/2$. Klein and Kneib (2016a) compared various different hyperprior families $p(\tau)$ for the smoothing variance in simulations and identified a Weibull hyperprior with shape parameter $1/2$ as optimal default, which they referred to as *scale-dependent hyperprior*. Thus, our theoretical results complement their recommendation based on simulations.

Finally, we conjecture that Assumption A1), which refers to the regularity of the unknown regression function f_0 , can be relaxed to Sobolev regularity of integer order, i.e., $f_0 \in W^{m_0}[0, 1] = \{f \in C^{m_0-1}[0, 1] : D^{m_0-1}f \text{ is absolutely continuous and } \int_0^1 (D^{m_0}f(x))^2 dx < \infty\}$. This is plausible in the light of the results of Huang and Su (2021) for frequentist penalized splines. Extending our results to Hölder classes with a continuous regularity parameter, however, does not seem to be straightforward. The main technical obstacle in this context seems to be the discrete nature of the roughness penalty order $q \in \{1, \dots, m-1\}$.

4 Extensions of Posterior Concentration Results

In this section we explain how the posterior concentration results of the previous section can be extended to cover an unknown residual variance and multivariate function estimation.

4.1 Unknown Residual Variance

In the case where σ_0^2 is unknown, we endow the B-spline coefficients $\beta \in \mathbb{R}^d$ with the same prior as before and the residual variance σ^2 with an independent Inverse Gamma prior $\sigma^2 \sim IG(a_\epsilon, b_\epsilon)$.

To establish posterior concentration results in the unknown σ_0^2 case, we first need to define a suitable notion of *joint posterior concentration*. We say that the joint posterior

concentrates at rate ϵ_n around the truth (f_0, σ_0^2) , if the expected posterior probability of neighborhoods of the form

$$\{(f, \sigma^2) \in C[0, 1] \times (0, \infty) : \|f - f_0\|_n < M\epsilon_n \wedge |\sigma^2/\sigma_0^2 - 1| < M\epsilon_n\} \quad (4)$$

converges to 1. Joint posterior concentration in this or a comparable sense has been studied by Choi and Schervish (2007); Wei et al. (2020); Bai et al. (2022). However, none of these papers has considered Bayesian penalized splines with a roughness penalty.

To prove joint posterior concentration in the sense of (4), we can again use the KL and testing strategy. The KL condition can be verified by introducing rectangles that are contained in the resulting KL type neighborhoods. Furthermore, the test

$$\phi_n = \mathbb{1} \left\{ \|\hat{f}_t - f_0\|_n \geq M/2 \epsilon_n \vee \left| \widehat{\sigma}_t^2 / \sigma_0^2 - 1 \right| \geq M/2 \epsilon_n \right\},$$

can be used for the testing condition. Thereby, $\hat{f}_t$ is a truncated DR estimator as in the proof of Theorem 3.1 and $\widehat{\sigma}_t^2 = \|Y^n - \hat{f}_t^n\|_2^2 / (n - t)$ is the corresponding estimator of the residual variance.

With this, we find that the joint posterior concentrates at near optimal rate around the truth (f_0, σ_0^2) in the sense of (4) under almost the same assumptions as before if we use an independent Inverse Gamma prior $\sigma^2 \sim IG(a_\epsilon, b_\epsilon)$ with constant shape $a_\epsilon \in (0, 1]$ and constant scale $b_\epsilon > 0$ on the residual variance σ^2 . We only get the restriction $m_0 \geq 2$ for the regularity of the unknown regression function $f_0 \in C^{m_0}[0, 1]$. To cover the case $m_0 = 1$, the prior on σ^2 needs to have even less prior mass close to the origin compared to an Inverse Gamma distribution. The square root of an Inverse Gamma distribution, for instance, works well.

4.2 Multivariate Function Estimation

By combining our results with those from Huang and Su (2021) in the frequentist context, we can also extend our posterior concentration results to the multivariate setting where a multivariate function $f_0 : [0, 1]^p \rightarrow \mathbb{R}$ with $p \geq 2$ is to be estimated. To this end, we first introduce a suitable prior $\tilde{\Pi}$ for multivariate functions.

A multivariate prior To construct a suitable prior $\tilde{\Pi}$ for multivariate functions, we combine a relatively large number of tensor product B-splines with a smoothness prior on the basis coefficients. We assume that the $p \geq 2$ marginal spline spaces have the same order $m \geq 2$ and the same vector of knots $\xi \in [0, 1]^{k+2}$. We thus consider the tensor product spline space $\mathcal{S}(m, \xi)^{\otimes p}$ of dimension $\tilde{d} = (m + k)^p$. A convenient basis of this function space are the tensor product B-splines $\{\tilde{B}_1, \dots, \tilde{B}_{\tilde{d}}\}$, which are simply products of the marginal B-spline basis functions (see, e.g., de Boor, 2001, for details).

Following Huang and Su (2021), a suitable measure of the roughness of a tensor product spline $f \in \mathcal{S}(m, \xi)^{\otimes p}$ is the thin plate penalty functional. For $p = 2$ covariates and penalty order $q = 2$, we obtain the expression

$$\int_{[0, 1]^2} \{\partial_1^2 f(x)\}^2 dx + 2 \int_{[0, 1]^2} \{\partial_1 \partial_2 f(x)\}^2 dx + \int_{[0, 1]^2} \{\partial_2^2 f(x)\}^2 dx,$$

where ∂_1 and ∂_2 are the partial derivatives with respect to x_1 and x_2 . For general $p \geq 2$ and $q \in \{1, \dots, m-1\}$, the corresponding differential bilinear form on the tensor product spline space is

$$\sum_{\alpha \in \{1, \dots, p\}^q} \int_{[0,1]^p} (D^\alpha f(x))(D^\alpha g(x)) dx, \quad f, g \in S(m, \xi)^{\otimes p},$$

where D^α is the q -tuple of partial derivatives $\partial_{\alpha_1} \dots \partial_{\alpha_q}$. Denoting the resulting $\tilde{d} \times \tilde{d}$ roughness penalty matrix in terms of the tensor product B-splines by $\tilde{R}_B^{(q)}$, a suitable prior $\tilde{\Pi}$ for multivariate functions is the distribution of the random tensor product spline $f = \sum_{j=1}^{\tilde{d}} \tilde{B}_j \tilde{\beta}_j$, where the tensor product B-spline coefficient vector $\tilde{\beta} \in \mathbb{R}^{\tilde{d}}$ is endowed with the hierarchical smoothness prior

$$\tilde{\beta} \mid \tau \sim N_{\tilde{d}} \left(0, \left(\tilde{R}_B^{(q)} / \tau + \tilde{B}^\top \tilde{H}^{(q)} \tilde{B} / (n\tau_{poly}) \right)^{-1} \right) \quad \text{with} \quad \tau \sim p(\tau).$$

Thereby, $\tilde{B}$ is the $n \times \tilde{d}$ tensor product B-spline design matrix and $\tilde{H}^{(q)}$ is the projection matrix onto the p -variate polynomials of degree $q-1$ or less. The prior $\tilde{\Pi}$ can be regarded as a natural multivariate extension of the prior Π introduced in Section 2.2.

Posterior concentration rates To achieve posterior concentration in the multivariate setting, we make the following assumptions.

- B1) Similar to de Jonge and van Zanten (2012) we assume that $f_0 \in C^{m_0}[0, 1]^p$, which means that all partial derivatives of order m_0 exist and are continuous, i.e., $D^\alpha f_0 \in C[0, 1]^p$ for $\alpha \in \{1, \dots, p\}^{m_0}$.
- B2) We use splines of order $m > m_0$ for the p marginal spline spaces.
- B3) We use k_n equidistant interior knots for the marginal spline spaces and for $k_n \rightarrow \infty$ we assume $k_n \sim c_k n^\delta$, where $c_k > 0$ and $1/(2m_0 + p) < \delta < 1/p$.
- B4) For the order of the thin plate roughness penalty we assume $q = m_0$.
- B5) For the hyperprior $p_n(\tau)$ on the smoothing variance τ we assume with $\tau_n^* = n^{-p/(2m_0+p)}$ and $\epsilon_n = n^{-m_0/(2m_0+p)}(\log n)^{1/2}$:
 - a) $\exists c_1 > 0 : \int_{c_1 \tau_n^*}^{\tau_n^*} p_n(\tau) d\tau \geq \exp(-n\epsilon_n^2)$ for all sufficiently large n .
 - b) $\exists c_2 > 0 : \int_{c_2 \tau_n^*}^{\infty} p_n(\tau) d\tau \leq \exp(-5 n\epsilon_n^2)$ for all sufficiently large n .
- B6) The additional penalty parameter τ_{poly} is either positive or diverges to $+\infty$ at polynomial speed as in Assumption A6).

- B7) The design points $x_i \in [0, 1]^p$, $i = 1, \dots, n$ are sufficiently regular, i.e., $\|Q_n - Q\|_\infty = o(k_n^{-p})$, where Q_n is the p -variate empirical cumulative distribution function of the x_i and Q is a fixed p -variate distribution function associated with a positive continuous design density on $[0, 1]^p$.

Theorem 4.1. *Under Assumptions B1)–B7), the posterior $\tilde{\Pi}_n(\cdot \mid Y^n)$ concentrates around f_0 at near optimal rate, i.e., there exists a constant $M > 0$ such that*

$$\mathbb{E}_0 \tilde{\Pi}_n(f \in C[0, 1]^p : \|f - f_0\|_n \geq M\epsilon_n \mid Y^n) = o(1),$$

where $\epsilon_n = n^{-m_0/(2m_0+p)}(\log n)^{1/2}$.

The proof of Theorem 4.1 follows by very similar arguments as the proof of Theorem 3.1 and is omitted for brevity.

Remark 4.2. The rate ϵ_n in Theorem 4.1 is again near optimal in the sense that it is the minimax rate times a logarithmic factor. Furthermore, Corollary 3.3 directly carries over for the corresponding choices of τ_n^* and ϵ_n . In conclusion, Bayesian penalized splines can also achieve posterior concentration at near optimal rate for multivariate functions.

5 Simulation Study

In this section we conduct a simulation where we benchmark Bayesian penalized splines against competing Bayesian methods from the literature. We focus on the univariate setting and the comparison with the closely related Bayesian spline regression approach where one places a hyperprior on the spline knots, which we refer to as “Bayesian model selection” following Kang and Jeong (2024). However, we also include Gaussian process regression to obtain a more complete picture.

Simulation design We follow Smith and Kohn (1996) and consider three test functions: (i) the linear function $f_1 = 2x - 1$, (ii) the sine function $f_2 = \sin(10\pi x)$ and (iii) the mixture of Gaussian densities

$$f_3(x) = \frac{1}{4}N_1(x; 0.15, 0.05^2) + \frac{1}{4}N_1(x; 0.6, 0.2^2), \quad x \in [0, 1].$$

We use $n \in \{100, 250, 500\}$ observations with regular design points $x_i = i/n$, $i = 1, \dots, n$, and set $\sigma_0 \in \{1/8, 1/4, 1/2\}$ times the observed range of the test function f . This yields signal-to-noise ratios (SNRs) of 8, 4 and 2, respectively, which we refer to as high, medium and low.

Estimation methods We consider the following three methods for estimation.

1. Bayesian penalized splines (BPS): We use cubic Bayesian penalized splines with $k = 20$ equidistant interior knots, a second order derivative penalty ($q = 2$)

and a Weibull hyperprior with shape parameter $1/2$ for the smoothing variance τ . For the rate parameter we use $\lambda = 1/500$, which ensures that prior draws have an appropriate scale. For σ^2 we use an independent Inverse Gamma prior $\sigma^2 \sim IG(1/1000, 1/1000)$.

For posterior sampling we use a Metropolis-within-Gibbs sampler. More specifically, we use Gibbs steps for β and σ^2 and Metropolis-Hastings steps with iteratively weighted least squares proposals as in Klein and Kneib (2016a) for τ . To reduce the computational cost of the Gibbs steps for β , we solve the generalized eigenvalue problem (Parlett, 1998, Chapter 15) for the pair of matrices $(R_B^{(q)}, R_B^{(q)} + B^\top B)$ before posterior sampling. Thereby, we obtain a fixed regular $d \times d$ matrix A that simultaneously diagonalizes $R_B^{(q)}$ and $B^\top B$ via congruence. Exploiting this matrix A , we then only need to invert a diagonal matrix in every iteration of the sampler.

2. Bayesian model selection (BMS): As a first competitor to Bayesian penalized splines we consider Bayesian model selection. This is a Bayesian spline regression approach that uses a hyperprior on the number of spline knots and optionally also on their location. We use the recent R package `gambms` (Kang and Jeong, 2024), which offers a flexible and numerically stable implementation. We use a maximum of 20 equidistant knots and an intrinsic prior, which is a special mixture of g -priors, on the spline coefficients (see Kang and Jeong, 2024, for details).
3. Gaussian processes (GPs): As a second competitor to Bayesian penalized splines we consider Gaussian process regression (Rasmussen and Williams, 2006). We use the R package `kernlab` (Karatzoglou et al., 2004), which offers a fast and stable implementation based on incomplete Cholesky decompositions of the kernel matrix, for which we use the popular squared exponential kernel.

Results Figure 2 shows boxplots of the root mean squared error (RMSE), defined as $(1/n \sum_{i=1}^n (\hat{f}(x_i) - f(x_i))^2)^{1/2}$, across $R = 50$ replicate data sets. Altogether, Bayesian penalized splines perform slightly worse than Bayesian model selection but better than Gaussian processes in terms of RMSE. However, Bayesian penalized splines perform significantly worse than both competitors for the sine function f_2 and a low SNR. In this case, Bayesian penalized splines lead to severe oversmoothing (see Figure 3). The oversmoothing disappears for the larger sample sizes $n = 250$ and $n = 500$ for which Bayesian penalized splines lead to comparable performance (see Figures 2–3). Interestingly, Bayesian model selection occasionally seems to oversmooth for the larger sample sizes (see Figure 3).

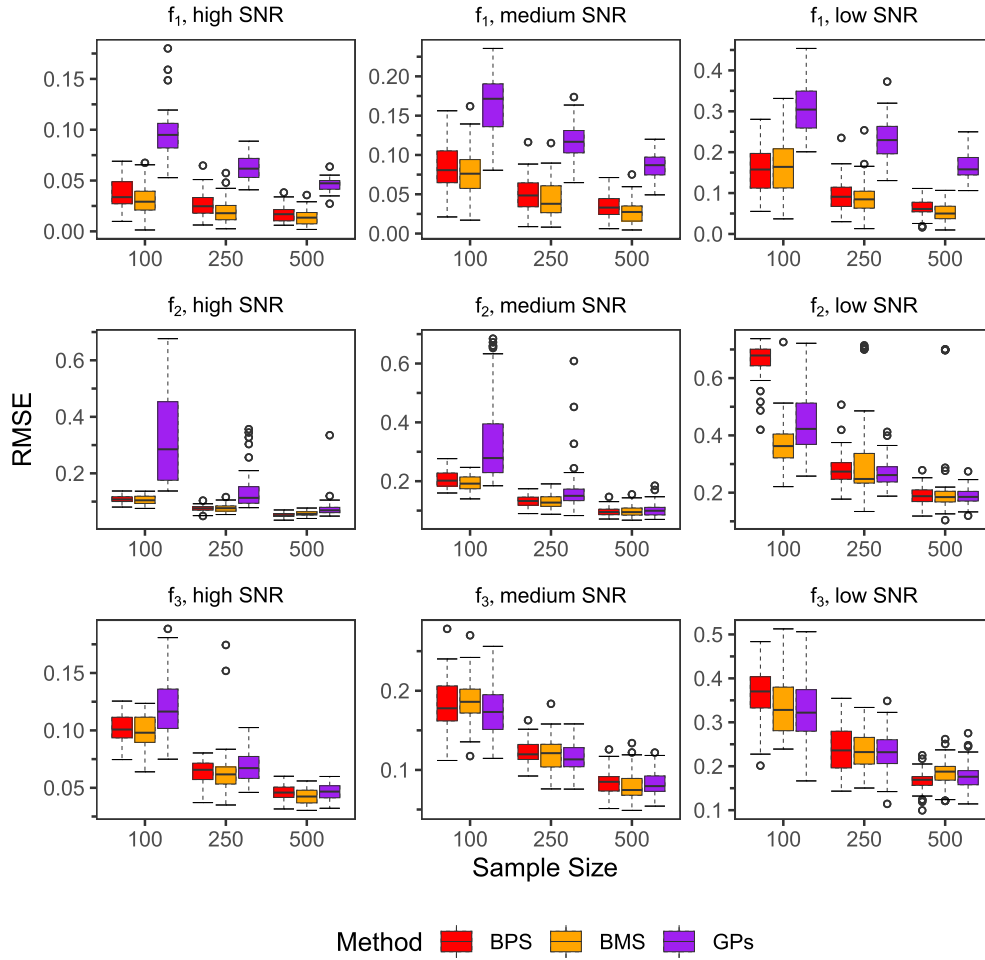


Figure 2: RMSEs. Shown are boxplots of the root mean squared error across $R = 50$ replicate data sets of sizes $n \in \{100, 250, 500\}$. The rows show the results for the different test functions ($f_1/f_2/f_3$), while the columns show the results for different SNRs (high/medium/low). Thereby, BPS refers to Bayesian penalized splines, BMS refers to Bayesian model selection and GPs refers to Gaussian processes.

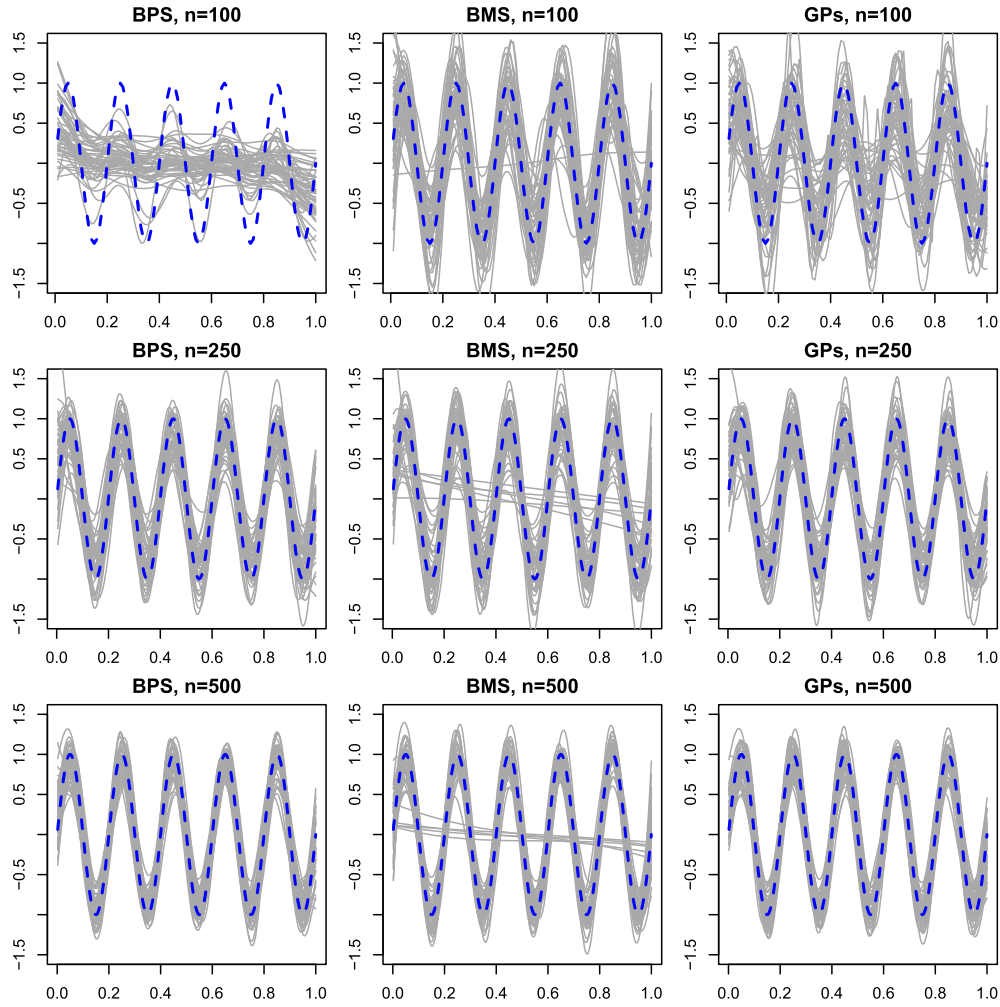


Figure 3: Function estimates for f_2 and low SNR. Shown are function estimates $\hat{f}$ (gray) of the sine function f_2 (dashed blue) for a low SNR across $R = 50$ replicate data sets. The rows show the function estimates for the different sample sizes ($n = 100/250/500$), while the columns show the estimates for the different methods (BPS/BMS/GPs). Thereby, BPS refers to Bayesian penalized splines, BMS refers to Bayesian model selection and GPs refers to Gaussian processes.

6 Discussion

Bayesian penalized splines are widely used in applied regression modeling but their theoretical understanding is lagging far behind. In this paper, we make an important step to close this gap in the literature and study posterior concentration rates for Bayesian penalized splines in a Gaussian nonparametric regression model.

We first consider univariate function estimation with known residual variance before explaining how our posterior concentration results can be extended to cover an unknown residual variance as well as multivariate function estimation using tensor product splines. Our results rely on innovative tools such as the truncated Demmler-Reinsch basis estimator and we show that Bayesian penalized splines can achieve the optimal rate of concentration up to a logarithmic factor, justifying their use as a default prior from a Bayesian nonparametric perspective (Ghosal and van der Vaart, 2017).

However, our results also reveal a theoretical limitation of Bayesian penalized splines. In contrast to Bayesian model selection, where one places a hyperprior on the spline knots, Bayesian penalized splines do not yield rate-adaptive posterior concentration. This is because the order of the roughness penalty q needs to match the regularity m_0 of the unknown regression function f_0 exactly to achieve the optimal rate of concentration. Hence, one needs to know m_0 to achieve the optimal rate for Bayesian penalized splines, which seems rather restrictive from a theoretical perspective.

To assess the practical consequences of the lack of rate-adaptivity for Bayesian penalized splines compared to Bayesian model selection, we have conducted a simulation study. Our results indicate that Bayesian penalized splines are actually slightly inferior to Bayesian model selection in terms of mean squared error, but overall the differences are minor. In addition, we find that Bayesian penalized splines still perform better than the popular Gaussian processes with a squared exponential kernel. This is in line with previous simulation studies showing that Bayesian penalized splines generally yield satisfactory estimation performance.

A practical advantage of Bayesian penalized splines compared to Bayesian model selection is that it is easier to set up a Markov chain Monte Carlo sampler for Bayesian penalized splines, in particular for non-Gaussian response models. The relative ease of posterior sampling has led Bayesian penalized splines to be used also for more sophisticated regression models, such as distributional or multivariate copula regression models (Klein et al., 2015; Klein and Kneib, 2016b).

In conclusion, we see three promising directions for future research: First, given the wide range of models where Bayesian penalized splines are routinely used in practice, it seems desirable to extend our concentration results to further models such as generalized additive or multivariate response regression models. A second, more computationally oriented direction is to implement Bayesian model selection for distributional or copula regression models where Bayesian penalized splines are already well-established. A third direction for future research is to devise a rate-adaptive version of Bayesian penalized splines, which ideally maintains the computational simplicity of the commonly used non-adaptive approach.

In recent work, Lim and Jeong (2023) suggest a hybrid spline that combines features from Bayesian penalized splines and Bayesian model selection, and they show that this approach leads to rate-adaptive posterior concentration. However, this hybrid spline is computationally more intensive and further strategies to achieve rate-adaptivity for Bayesian penalized splines are conceivable.

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

Paul Bach and Nadja Klein gratefully acknowledge support by the German research foundation (DFG) through the Emmy Noether grant KL 3037/1-1.

Supplementary Material

Supplement to “Posterior Concentration Rates for Bayesian Penalized Splines” (DOI: [10.1214/25-BA1523SUPP](https://doi.org/10.1214/25-BA1523SUPP); .pdf). The supplementary material contains proofs of our theoretical results and several auxiliary results.

References

- Armagan, A., Dunson, D. B., Lee, J., Bajwa, W. U., and Strawn, N. (2013). “Posterior consistency in linear models under shrinkage priors.” *Biometrika*, 100(4): 1011–1018. MR3142348. doi: <https://doi.org/10.1093/biomet/ast028>. 1473, 1480
- Bach, P. and Klein, N. (2025). “Supplement to “Posterior Concentration Rates for Bayesian Penalized Splines”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/25-BA1523SUPP>. 1474
- Bai, R., Moran, G. E., Antonelli, J. L., Chen, Y., and Boland, M. R. (2022). “Spike-and-slab group lassos for grouped regression and sparse generalized additive models.” *Journal of the American Statistical Association*, 117(537): 184–197. MR4399078. doi: <https://doi.org/10.1080/01621459.2020.1765784>. 1473, 1481
- Belitser, E. and Serra, P. (2014). “Adaptive priors based on splines with random knots.” *Bayesian Analysis*, 9(4): 859–882, 24. MR3293959. doi: <https://doi.org/10.1214/14-BA879>. 1473
- Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). “Bayesian smoothing and regression splines for measurement error problems.” *Journal of the American Statistical Association*, 97(457): 160–169. MR1947277. doi: <https://doi.org/10.1198/016214502753479301>. 1471, 1475
- Brezger, A. and Lang, S. (2006). “Generalized structured additive regression based on Bayesian P-splines.” *Computational Statistics & Data Analysis*, 50(4): 967–991. MR2210741. doi: <https://doi.org/10.1016/j.csda.2004.10.011>. 1471

- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). “Stan: A probabilistic programming language.” *Journal of Statistical Software*, 76(1). 1473
- Choi, T. and Schervish, M. J. (2007). “On posterior consistency in nonparametric regression problems.” *Journal of Multivariate Analysis*, 98(10): 1969–1987. MR2396949. doi: <https://doi.org/10.1016/j.jmva.2007.01.004>. 1481
- Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009). “Asymptotic properties of penalized spline estimators.” *Biometrika*, 96(3): 529–544. MR2538755. doi: <https://doi.org/10.1093/biomet/asp035>. 1472, 1474, 1478
- De Boor, C. (2001). *A Practical Guide to Splines*. New York, NY: Springer. MR1900298. 1481
- De Jonge, R. and van Zanten, J. H. (2012). “Adaptive estimation of multivariate functions using conditionally Gaussian tensor-product spline priors.” *Electronic Journal of Statistics*, 6: 1984–2001. MR3020254. doi: <https://doi.org/10.1214/12-EJS735>. 1473, 1482
- Eilers, P. H. C. and Marx, B. D. (1996). “Flexible smoothing with B-splines and penalties.” *Statistical Science*, 11(2): 89–121. MR1435485. doi: <https://doi.org/10.1214/ss/1038425655>. 1475
- Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. D. (2021). *Regression: Models, Methods and Applications*. Berlin: Springer. MR4404768. doi: <https://doi.org/10.1007/978-3-662-63882-8>. 1471
- Ghosal, S. and van der Vaart, A. W. (2007). “Convergence rates of posterior distributions for noniid observations.” *The Annals of Statistics*, 35(1): 192–223. MR2332274. doi: <https://doi.org/10.1214/009053606000001172>. 1472, 1473, 1477, 1479
- Ghosal, S. and van der Vaart, A. W. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge, UK: Cambridge University Press. MR3587782. doi: <https://doi.org/10.1017/9781139029834>. 1472, 1477, 1487
- Harezlak, J., Ruppert, D., and Wand, M. P. (2018). *Semiparametric Regression with R*. New York, NY: Springer. MR3889001. doi: <https://doi.org/10.1007/978-1-4939-8853-2>. 1471, 1475, 1476
- Hennerfeind, A., Brezger, A., and Fahrmeir, L. (2006). “Geoadditive survival models.” *Journal of the American Statistical Association*, 101(475): 1065–1075. MR2324146. doi: <https://doi.org/10.1198/016214506000000348>. 1471
- Huang, J. Z. and Su, Y. (2021). “Asymptotic properties of penalized spline estimators in concave extended linear models: Rates of convergence.” *The Annals of Statistics*, 49(6): 3383–3407. MR4352534. doi: <https://doi.org/10.1214/21-aos2088>. 1472, 1480, 1481
- Kang, G. and Jeong, S. (2024). “Model selection-based estimation for generalized additive models using mixtures of g-priors: Towards systematization.” *Bayesian Analysis*, 1(1): 1–30. 1483, 1484

- Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). “kernlab – An S4 package for kernel methods in R.” *Journal of Statistical Software*, 11(9): 1–20. 1484
- Kauermann, G., Krivobokova, T., and Fahrmeir, L. (2009). “Some asymptotic results on generalized penalized spline smoothing.” *Journal of the Royal Statistical Society: Series B*, 71: 487–503. MR2649606. doi: <https://doi.org/10.1111/j.1467-9868.2008.00691.x>. 1472
- Klein, N. and Kneib, T. (2016a). “Scale-dependent priors for variance parameters in structured additive distributional regression.” *Bayesian Analysis*, 11(4): 1071–1106. MR3545474. doi: <https://doi.org/10.1214/15-BA983>. 1474, 1476, 1480, 1484
- Klein, N. and Kneib, T. (2016b). “Simultaneous inference in structured additive conditional copula regression models: A unifying Bayesian approach.” *Statistics and Computing*, 26(4): 841–860. MR3515025. doi: <https://doi.org/10.1007/s11222-015-9573-6>. 1487
- Klein, N., Kneib, T., Lang, S., and Sohn, A. (2015). “Bayesian structured additive distributional regression with an application to regional income inequality in Germany.” *Annals of Applied Statistics*, 9(2): 1024–1052. MR3371346. doi: <https://doi.org/10.1214/15-AOAS823>. 1471, 1487
- Lang, S., Adebayo, S. B., Fahrmeir, L., and Steiner, W. J. (2003). “Bayesian geoadditive seemingly unrelated regression.” *Computational Statistics*, 18: 263–292. MR1994652. doi: <https://doi.org/10.1007/s001800300144>. 1471
- Lang, S. and Brezger, A. (2004). “Bayesian P-Splines.” *Journal of Computational and Graphical Statistics*, 13(1): 183–212. MR2044877. doi: <https://doi.org/10.1198/1061860043010>. 1471, 1473, 1475, 1476
- Lim, S. and Jeong, S. (2023). “Synergizing roughness penalization and basis selection in Bayesian spline regression.” *arXiv preprint arXiv:2311.13481*. 1488
- Nychka, D. and Cummins, D. (1996). “Comment on: Eilers, P., Marx, B., Flexible smoothing with B-splines and penalties.” *Statistical Science*, 11: 89–121. MR1435485. doi: <https://doi.org/10.1214/ss/1038425655>. 1474
- O’Sullivan, F. (1986). “A statistical perspective on ill-posed inverse problems.” *Statistical Science*, 1(4): 502–518. MR0874480. 1472, 1475
- Parlett, B. N. (1998). *The Symmetric Eigenvalue Problem*. Philadelphia, PA: SIAM. MR1490034. doi: <https://doi.org/10.1137/1.9781611971163>. 1484
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press. MR2514435. 1484
- Rousseau, J. (2016). “On the frequentist properties of Bayesian nonparametric methods.” *Annual Review of Statistics and Its Application*, 3: 211–231. 1477
- Ruppert, D. (2002). “Selecting the number of knots for penalized splines.” *Journal of Computational and Graphical Statistics*, 11(4): 735–757. MR1944261. doi: <https://doi.org/10.1198/106186002321018768>. 1475

- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press. MR1998720. doi: <https://doi.org/10.1017/CBO9780511755453>. 1474
- Schwartz, L. (1965). “On Bayes procedures.” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(1): 10–26. MR0184378. doi: <https://doi.org/10.1007/BF00535479>. 1472
- Serra, P. and Krivobokova, T. (2017). “Adaptive empirical Bayesian smoothing splines.” *Bayesian Analysis*, 12(1): 219–238. MR3597573. doi: <https://doi.org/10.1214/16-BA997>. 1473
- Shen, W. and Ghosal, S. (2015). “Adaptive Bayesian procedures using random series priors.” *Scandinavian Journal of Statistics*, 42(4): 1194–1213. MR3426318. doi: <https://doi.org/10.1111/sjos.12159>. 1473
- Smith, M. and Kohn, R. (1996). “Nonparametric regression using Bayesian variable selection.” *Journal of Econometrics*, 75(2): 317–343. 1483
- Song, Q. and Liang, F. (2023). “Nearly optimal Bayesian shrinkage for high-dimensional regression.” *Science China Mathematics*, 66(2): 409–442. MR4535982. doi: <https://doi.org/10.1007/s11425-020-1912-6>. 1473
- Song, X.-Y. and Lu, Z.-H. (2012). “Semiparametric transformation models with Bayesian P-splines.” *Statistics and Computing*, 22(5): 1085–1098. MR2950087. doi: <https://doi.org/10.1007/s11222-011-9280-x>. 1471
- Stone, C. J. (1982). “Optimal global rates of convergence for nonparametric regression.” *The Annals of Statistics*, 10(4): 1040–1053. MR0673642. 1473, 1479
- Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., and Flexeder, C. (2013). “Bayesian semiparametric additive quantile regression.” *Statistical Modelling*, 13(3): 223–252. MR3179525. doi: <https://doi.org/10.1177/1471082X13480650>. 1471
- Wand, M. P. and Ormerod, J. T. (2008). “On semiparametric regression with O’Sullivan penalized splines.” *Australian & New Zealand Journal of Statistics*, 50(2): 179–198. MR2431193. doi: <https://doi.org/10.1111/j.1467-842X.2008.00507.x>. 1472, 1475
- Wei, R., Reich, B. J., Hoppin, J. A., and Ghosal, S. (2020). “Sparse Bayesian additive nonparametric regression with application to health effects of pesticides mixtures.” *Statistica Sinica*, 30: 55–79. MR4285485. doi: <https://doi.org/10.5705/ss.202017.0315>. 1473, 1481
- Xiao, L. (2019). “Asymptotic theory of penalized splines.” *Electronic Journal of Statistics*, 13(1): 747–794. MR3925516. doi: <https://doi.org/10.1214/19-ejs1541>. 1472, 1475, 1478
- Yoo, W. W. and Ghosal, S. (2016). “Supremum norm posterior contraction and credible sets for nonparametric multivariate regression.” *The Annals of Statistics*, 44(3): 1069–1102. MR3485954. doi: <https://doi.org/10.1214/15-AOS1398>. 1473