Penalized hyperbolic-polynomial splines

Original Citation:

Availability:
This version is available at: 2158/1240297 since: 2022-02-16T10:53:04Z

Published version:
DOI: 10.1016/j.aml.2021.107159

Terms of use:
Open Access
La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:
Penalized Hyperbolic-Polynomial Splines

Rosanna Campagna\textsuperscript{a}, Costanza Conti\textsuperscript{b,∗}

\textsuperscript{a}Department of Mathematics and Physics, University of Campania “Luigi Vanvitelli”, Italy
\textsuperscript{b}Department of Industrial Engineering, DIEF, University of Firenze, Italy

Abstract

With the aim of generalizing P-splines, we here define a special type of penalized splines, called HP-splines, where polynomial splines are replaced by the richer class of hyperbolic-polynomial splines and a suitably tailored discrete penalty term is used. Hyperbolic-polynomial splines, important in several applications, are a natural generalization of polynomial splines consisting of piecewise-defined functions with segments spanned by ‘atoms’ of type \( x^r e^{\alpha x} \) where \( r = 0, \ldots, \ell \) and \( \alpha \in \mathbb{R} \). HP-splines, that reduce to P-splines for \( \alpha = 0 \), are more suitable to data with an exponential trend which is frequent in applications.

Keywords: Penalized splines, Discrete penalty, P-splines, Hyperbolic-polynomial splines, B-splines

1. Introduction

P-splines appeared about twenty years ago with the aim of simplifying the work of O’Sullivan in the context of penalized splines [1]. Indeed, in 1986 O’Sullivan noticed that by using a cubic B-spline basis, the familiar measure of roughness given by the integrated squared second derivative, can be expressed as a quadratic function of the coefficients. P-splines, later introduced by Eilers and Marx [2], go one step further and, by the help of uniform knots B-splines, discard completely the derivatives and express the roughness as the sum of squares of differences of the coefficients. Differences are extremely easy to compute and their generalization to higher orders is straightforward. P-splines are known to be effective and appreciated in several applications (see, for example, the recent paper [3]). A comprehensive description of P-splines is given in [4].

The two ingredients of P-splines are polynomial B-splines and discrete difference penalties. B-splines are well known basis functions of polynomial spline spaces very attractive for modeling and regression. However, their use requires to deal with the optimal choice of the number and the positions of the knots which is not an easy task. One possible solution is to use equidistant knots but, in case they are too few, only a limited control over smoothness and fit is possibile. The key idea of P-splines is to use a relatively large number of knots and a finite difference penalty on the coefficients of adjacent B-splines. This penalty term has a connection to the familiar spline penalty based on the integral of the squared second derivative.

The aim of this short paper is to move beyond the P-splines and consider hyperbolic-polynomial penalized splines, HP-splines for short, where polynomial splines are replaced by the richer class of hyperbolic-polynomial ones. Hyperbolic-polynomial splines, special instance of \( L \)-splines, are a natural generalization of polynomial splines consisting of piecewise-defined functions with segments spanned by ‘atoms’ of type \( x^r e^{\alpha x} \) where \( r = 0, \ldots, \ell \) and \( \alpha \in \mathbb{R} \). They are important in several applications ranging from geometric modeling, to image analysis, passing trough isogeometric analysis and system theory (see, for example, [5–8]). The derivation of HP-splines is not more complicated than that of P-splines to which they reduce whenever \( \alpha = 0 \). By their nature, they are certainly more suitable to data showing an exponential trend. Multi-exponential decaying data are very frequent in applications and a continuous description of this type

∗Corresponding author
Email addresses: rosanna.campagna@unicampania.it (Rosanna Campagna), costanza.conti@unifi.it (Costanza Conti)
of data allows the use of mathematical tools for data analysis such as the Laplace Transform, see for example [9–11] and reference therein.

The paper is organized as follows: in Section 2, cubic P-splines are briefly reviewed and the idea leading to their generalization is discussed, based on the notion of penalized regression L-spline which is also presented. A particular instance of HP-splines and B-splines is considered in Section 3 together with the corresponding discrete penalty term used for the HP-splines definition. The closing Section 4 provides some numerical examples to show the effectiveness of HP-splines, particularly for data with an exponential trend.

2. Beyond P-splines

In this section, following Eilers and Marx (see [2, 4] for all details) we briefly review cubic P-splines as starting point of the new family of splines with penalty we propose in Section 3.

Consider the regression of 

\[ M(a_1, \ldots, a_n) = \sum_{i=1}^{m} w_i \left( y_i - \sum_{j=1}^{n} a_j B_j(x_i) \right)^2 + \lambda \int_a^b \left( \sum_{j=1}^{n} a_j B_j''(x) \right)^2 \, dx. \]

The minimization is done with respect to the splines coefficients \( a = (a_1, \ldots, a_n) \) while \((w_1, \ldots, w_m)\) are non zero weights and \( \lambda \) is a regularization parameter. By taking into consideration that the second derivative of a cubic B-spline is a linear combination of linear B-splines and combining the uniformity of the knots with the B-splines locality, it is easy to see that

\[ \sum_{j \in \mathbb{Z}} a_j B_j'(x) = \sum_{j \in \mathbb{Z}} (\Delta^2 a)_j b_j(x), \quad \text{where} \quad (\Delta^2 a)_j = a_j - 2a_{j-1} + a_{j-2}, \; j \in \mathbb{Z}, \]

While \( b_j \) denotes a linear B-spline. From above, applying a certain level of approximation (e.g. disregarding the integral contribution) one arrives at the penalized least squares objective function charactering the P-splines

\[ P(a_1, \ldots, a_n) = \sum_{i=1}^{m} w_i \left( y_i - \sum_{j=1}^{n} a_j B_j(x_i) \right)^2 + \lambda \sum_{j=1}^{n} \left( (\Delta^2 a)_j \right)^2, \]

where the minimization is done, again, with respect to the splines coefficients \( a = (a_1, \ldots, a_n) \). Of course, the problem in (1) is different from the problem in (2).

Our goal is to extend the P-splines idea to the L-splines, a generalization of polynomial splines related to a linear differential operator. Following [12], the definition of L-splines is recalled below.

Definition 2.1 (L-spline). For a given partition of \([a, b], \Xi := \{a = \xi_1 < \cdots < \xi_N = b\}\), the L-spline related to an order-\(\ell\) linear differential operator of type \(L_{\ell} := D^\ell + \sum_{j=0}^{\ell-1} c_j D^j\) and to the partition \(\Xi\), is a function \(s \in C^{\ell-2}[a, b]\) such that \(L_{\ell} s = 0\) in every interval \( (\xi_i, \xi_{i+1}) \), \(i = 1, \ldots, N-1\). The corresponding spline space is a linear space of dimension \(\ell + N - 2\) denoted as \(S_{L_{\ell}, \Xi}\).

We continue with the definition of penalized regression L-spline on the space

\[ H^d[a, b] := \{ u \in C^{d-1}[a, b], \; u^{(d-1)} \text{ absolutely continuous}, \; u^{(d)} \in L^2[a, b] \}. \]
**Definition 2.2** (Penalized regression $L$-spline). Given the data points $(x_i, y_i), i = 1, \ldots, m, x_1 < \cdots < x_m,$ let $a = x_1$ and $b = x_m$. Let $\Xi := \{a = \xi_1 < \xi_2 \cdots < \xi_N = b\}$ be a given knot partition of $[a, b]$, and $\mathcal{L}_d$ an order-$d$ linear differential operator with formally adjoint $\mathcal{L}_d^*$. The penalized regression $L$-spline, related to the differential operator $L_{2d} := \mathcal{L}_d^* \mathcal{L}_d$ of order $2d$, is the solution of the minimization problem

$$
\min_{u \in S_{L_{2d}, \Xi}^{\mathbb{R}}} \sum_{i=1}^{m} w_i (y_i - u(x_i))^2 + \lambda \int_{a}^{b} (\mathcal{L}_d u(x))^2 \, dx,
$$

with $(w_1, \ldots, w_m)$ non-zero weights and $\lambda$ a regularization parameter.

**Remark 2.1.** We observe that, if we search for the solution of (3) in $H^d[a, b]$, without any restriction to the spline space $S_{L_{2d}, \Xi}$, the minimum is obtained when the knots $\xi_k, k = 1, \ldots, N$ coincide with the data $x_i,$ $i = 1, \ldots, m$ (see [13], for example). But, since here we are interested in the regression of $m$ data, with $N < m$, the minimum needs to be searched in the splines space $S_{L_{2d}, \Xi}$ to which we restrict our attention. Nevertheless, a more general study of (3) (existence, uniqueness and solution characterization) is important and in fact under the authors’ attention.

Now, if $\{\phi_1, \ldots, \phi_n\}, n = 2d + N - 2,$ are bell-shaped compactly supported basis functions for the spline space $S_{L_{2d}, \Xi}$ defined in Definition 2.1, the problem in (3) can be written as

$$
\min_{a_1, \ldots, a_n} \sum_{i=1}^{m} w_i \left( y_i - \sum_{j=1}^{n} a_j \phi_j(x_i) \right)^2 + \lambda \int_{a}^{b} \left( \sum_{j=1}^{n} a_j \mathcal{L}_d \phi_j(x) \right)^2 \, dx.
$$

In order to get a simple and discrete penalty term, the idea is now to substitute the operator $\mathcal{L}_d$ with a discretization based on the grid points $\Xi$, through a difference operator denoted by $\Delta^{\Xi}_d$. By using the uniformity of the knots, we apply a summation by parts argument which interchanges the discrete difference operator $\Delta^{\Xi}_d$ from the basis function to the basis coefficients and arrive, still disregarding the integrals, at the simplified formulation

$$
\min_{a_1, \ldots, a_n} \sum_{i=1}^{m} w_i \left( y_i - \sum_{j=1}^{n} a_j \phi_j(x_i) \right)^2 + \lambda \sum_{j=1}^{n} \left( \Delta^{\Xi}_d a_j \right)^2.
$$

3. Hyperbolic-polynomial P-splines

*Exponential-polynomial splines* are a natural generalization of polynomial splines important in several applications ranging from geometric modelling to image analysis passing through isogeometric analysis and system theory. They are piecewise-defined functions consisting of segments belonging to the null space of a differential operator $L_{\ell} = (\mathcal{D} + \alpha I)^{\ell}$, where $\alpha \in \mathbb{C}, \mathcal{D}$ is the first derivative operator and $I$ is the identity operator. In case $\alpha \in \mathbb{R}$, they are called *hyperbolic-polynomial splines*. Certainly, further generalizations of the notion of spline are available. For example, splines with exponential-polynomial segments involving different $\alpha$ or splines with segments belonging to spaces of different dimension (see the classical book [14] or the more recent paper [15]). But, the study of this type of splines is out of scope of this short paper.

It is a well established fact that spline applications often require the use of spline bases with specific properties. For example, polynomial B-splines possess non-negativity, compactness of the support, minimal support with respect to degree and smoothness, stability (see, e.g., the celebrated book [16]). The latter properties, among others, make them essential building blocks in many contexts, including approximation theory, numerical differentiation and integration, signal and image processing, computer-aided design and computer graphics (see [8] and [17]). Similarly to the polynomial case, the exponential polynomial spline space is spanned by ‘bell-shaped’ compactly supported bases. These bell-shaped functions enjoy several properties of polynomial B-splines and are usually called B-spline-like functions or generalized B-splines (see [18] and references therein). Here, we are interested in a particular hyperbolic-polynomial spline model
defined, for \( \alpha \in \mathbb{R} \), by the composition of the differential operator \( \mathcal{L}_2 = (\mathcal{D} + \alpha \mathcal{I})^2 \) and its adjoint \( \mathcal{L}_2^* = (\mathcal{D} - \alpha \mathcal{I})^2 \), whose action on regular functions is

\[
\mathcal{L}_2 u := u'' + 2\alpha u' + \alpha^2 u, \quad \mathcal{L}_2^* u = u'' - 2\alpha u' + \alpha^2 u, \quad \alpha \in \mathbb{R}.
\]  

(6)

Their composition is identical with the 4th order differential operator \( \mathcal{L}_2^2 v = v^{(iv)} - 2\alpha^2 v'' + \alpha^4 v, \alpha \in \mathbb{R} \). The null spaces of \( \mathcal{L}_2 \) and \( \mathcal{L}_2^* \) are, respectively, the following two-dimensional and four-dimensional spaces

\[
\mathcal{E}_2 := \text{span}\{e^{-\alpha x}, e^{-\alpha x}\}, \quad \mathcal{E}_4 := \text{span}\{e^{\alpha x}, e^{\alpha x}, e^{-\alpha x}, e^{-\alpha x}\}, \quad \alpha \in \mathbb{R},
\]

(7)

that are Chebyshev spaces on the real line \([12]\). Note that for \( \alpha \to 0 \) the spaces \( \mathcal{E}_2, \mathcal{E}_4 \) reduce to \{1, x\} and \{1, x, x^2, x^3\}, respectively. For the corresponding spline space \( \mathcal{S}_{\bar{E}_2}^{E_4} \) with knots \( \Xi := \{a = \xi_1 < \cdots < \xi_n = b\} \) and dimension \( n + 2 \), generalized B-splines are constructed and investigated in \([9, \text{Section 3.1}]\). They have a compact support identified by 5 consecutive knots, are \( C^2 \)-regular, and with segments in the space \( \mathcal{E}_4 \). As detailed in \([9, \text{Section 3.1}]\), their construction is possible by expressing the segments in terms of proper Bernstein-like local bases and imposing regularity conditions at the knots in order to guarantee a global \( C^2 \)-regularity. An example of an hyperbolic-polynomial B-spline with uniform knots is given in Figure 1 (bottom) where the knots are denoted as ‘\( * \)’ and where the bell-shaped graph and the locality of the support are evident.

We continue by discussing the new family of hyperbolic-polynomial penalized splines, named HP-splines, we propose. The starting point is given by the data points \((x_i, y_i), i = 1, \ldots, m\). From them we set \( a = x_1, b = x_m \) and construct the uniform knots \( \Xi := \{a = \xi_1 < \cdots < \xi_n = b\} \), \( \xi_i = a + \frac{(b-a)}{n}(i-1), i = 1, \ldots, n \) with fixed grid size \( h = \frac{(b-a)}{n-1} \). Then, we consider the difference operator \( \Delta_{2}^{h,\alpha} \) defined as

\[
\Delta_{2}^{h,\alpha} u = e^{ah}u(x + h) - 2u(x) + e^{-ah}u(x - h), \quad x \in [a, b], \quad \alpha, h \in \mathbb{R},
\]

whose corresponding null space is, again, \( \mathcal{E}_2 \) in (7) while \( \mathcal{E}_4 \) is the null space of \( \Delta_{2}^{h,\alpha}(\Delta_{2}^{h,-\alpha}) \). As basis functions for \( \mathcal{S}_{\bar{E}_2}^{E_4} \) we consider \{\( B_0, \ldots, B_{n+1} \}\} as in \([9, \text{Section 3.1}]\) whose construction requires the uniform left and right extra knots \( \xi_0 = \xi_1 - \ell h, \ell = 1, 2, 3, \xi_{n+\ell} = \xi_n + \ell h, \ell = 1, 2, 3 \), respectively.

With these ingredients, we repeat the argument in Section 2 and substitute \( \sum_{j=0}^{n+1} a_j \mathcal{L}_2 B_j(x) \) in (4) with

\[
\sum_{j=0}^{n+1} a_j \Delta_{2}^{h,\alpha} B_j(x) = \sum_{j=0}^{n+1} a_j \left( e^{ah} B_j(x + h) - 2 B_j(x) + e^{-ah} B_j(x - h) \right).
\]

Due to the uniformity of the knots, it follows \( B_j(x - \ell h) = B_{j+\ell}(x) \) so that, with the convention \( B_{-1} = B_{n+2} \equiv 0 \), it is

\[
\sum_{j=0}^{n+1} a_j \Delta_{2}^{h,\alpha} B_j(x) = \sum_{j=0}^{n+1} a_j \left( e^{ah} B_{j-1}(x) - 2 B_j(x) + e^{-ah} B_{j+1}(x) \right).
\]

Therefore, applying the difference operator to the sequence of real values \( a = (a_j)_{j=0,\ldots,n+1} \) as

\[
(\Delta_{2}^{h,\alpha} a)_0 = e^{ah} a_1 - 2a_0, \quad (\Delta_{2}^{h,\alpha} a)_j = e^{ah} a_{j+1} - 2a_j + e^{-ah} a_{j-1}, \quad (\Delta_{2}^{h,\alpha} a)_{n+1} = -2a_{n+1} + e^{-ah} a_n,
\]

taking into consideration the local support of each basis function \( B_j \) and that

\[
\sum_{j=0}^{n+1} a_j \Delta_{2}^{h,\alpha} B_j(x) = \sum_{j=0}^{n+1} (\Delta_{2}^{h,\alpha} a)_j B_j(x),
\]

we end up with our new proposal of discrete penalty term analogue of the penalty term in (2), to which it reduces when \( \alpha = 0 \). To summarize, the HP-spline is defined as a solution to the penalized least square problem

\[
\min_{a_0, \ldots, a_{n+1}} E(a_0, \ldots, a_{n+1}) = \min_{a_0, \ldots, a_{n+1}} \sum_{i=1}^{m} w_i \left( y_i - \sum_{j=0}^{n+1} a_j B_j(x_i) \right)^2 + \lambda \sum_{j=0}^{n+1} \left( (\Delta_{2}^{h,\alpha} a)_j \right)^2,
\]

(8)
where the minimum is with respect to the exponential polynomial B-splines coefficients $a = (a_j)_{j=0}^{n+1}$, $(w_1, \ldots, w_m)$ are non-zero weights, and $\lambda$ is a regularization parameter. It is easy to see that the problem in (8) is equivalent to solve the linear system

$$B^T Wy = \left( B^T WB + \lambda (D_2^{h,\alpha})^T D_2^{h,\alpha} \right) a,$$

(9)

where $y \in \mathbb{R}^m$, $a \in \mathbb{R}^{(n+2)}$, $B \in \mathbb{R}^{m \times (n+2)}$, $W \in \mathbb{R}^{m \times m}$ and $D_2^{h,\alpha} \in \mathbb{R}^{n \times (n+2)}$, with

$$D_2^{h,\alpha} = \begin{bmatrix}
-2 & e^{\alpha h} & 0 & \cdots & 0 \\
e^{-\alpha h} & -2 & e^{\alpha h} & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & e^{-\alpha h} & -2 & e^{\alpha h} \\
\vdots & \ddots & \vdots & e^{-\alpha h} & -2
\end{bmatrix}, \quad B := (B_j(x_i))_{i=1,\ldots,m} \quad W := (\text{diag}(w_i))_{i=1,\ldots,m}.$$

The band structure of the matrix $B^T$, inherited by the B-spline locality, is shown in Figure 1 (top).

We conclude by observing that, as well-known, the linear system in (9) may suffer from numerical instability issues for relatively large values of the parameter $\alpha$. This fact is shortly discussed in the next section.

4. Numerical examples

This section is to show HP-splines in action on two different types of data sets. The tests are carried out with a MATLAB® R2020b software available to the authors. In all tests $W$ is the identity matrix of order $m$. The first test considers the benchmark Motorcycle Data of 94 data (see [19]). The associated HP-spline (black ‘-’), and the P-spline (magenta ‘:’), derived for $n = 40$, $\alpha = 0.3$ and $\lambda = 0.5$ (with $\alpha$ and $\lambda$ empirically driven by the data and selected in order to get a visually inspected agreement with them), are shown in Figure 2 with the data. Figure 3 show HP-splines with the same $\alpha$ and $\lambda$, for $n \in \{15, 20, 25, 30, 35, 40, 45, 50\}$. We see that only $n = 15$ and $n = 20$ grant a solution outside the 95% Bayesian confidence region, represented in the figure. As expected, without any prior information about the modelling problem, HP-spline behaves similarly to P-spline.

![Figure 1: $B^T$ structure and B-spline](image1.png)  
![Figure 2: HP-(-) vs P-(:) splines](image2.png)  
![Figure 3: 95% confidence band](image3.png)

The second test is related to a synthetic and noisy data set of $m = 40$ random points in $[-1.5, 1.5]$, and the corresponding evaluations of the exponential function

$$f(x) = 10^{-5}(e^{7x} - xe^{-7x}), \quad x \in [-1.5, 1.5],$$

affected by absolute Gaussian noise with zero mean and standard deviation $\sigma$ (specified in the figure captions). The associated HP-spline (black ‘-’), derived for $\alpha = 3$, and the P-spline (magenta ‘:’), are both shown in all Figures 4-12 together with the data (displayed as ‘*’) and the knots (displayed as ‘o’). In this
test $\alpha$ is computed in run time, by a best fitting algorithm based on two steps: first we consider a function in the space $E_4$ modelled by 5 free parameters, with $\alpha$ included; then the free parameters are defined by a nonlinear least-squares regression, through the MATLAB function \texttt{nlinfit}. Comparison of Figures 4–6 (where $n = 15$ and $\lambda = 0.1$) with Figures 7–9 (where $n = 20$ and $\lambda = 0.1$), is to stress the impact of the spline space dimension in capturing the ‘details’ while comparison of Figures 7–9 with Figures 10–12 (where $n = 20$ and $\lambda = 1$) is to stress the effects of the smoothing parameter. All tests confirm the validity of the HP-spline model that better captures the exponential data behavior.

We conclude this section with Table 1 showing the two-norm condition number of the matrix in (9), $\mu_2$. Our analysis considers $\alpha \leq 30$ and shows that $\mu_2$, though increasing with $m$ and $n$, is kept under control.
5. Conclusions

In this work, we propose a penalized hyperbolic-polynomial spline with a discrete penalty term suitable to model data showing an exponential trend, a frequent scenario in applications. The spline expansion is made through local hyperbolic-polynomial B-splines granting boundedness of the linear system to be solved. Both the B-splines and the discrete penalty term, reduce to the ones of P-spline for $\alpha = 0$. Theoretical investigation of HP-spline properties, including the issues mentioned in Remark 2.1, as well as dynamic selection strategy of the parameter $\alpha$, certainly deserve more attention and are presently under investigation.

References