**Katholieke Universiteit Leuven**

Departement Elektrotechniek ESAT-SISTA/TR 2004-228

# A class of template splines[1]

Diana M. Sima[2] Sabine Van Huffel[2]

July 2005
Accepted for publication to
*Computational Statistics & Data Analysis.*

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/dsima/abstracts/04-228.html*

[2]ESAT-SISTA, K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium, (diana.sima,sabine.vanhuffel@esat.kuleuven.ac.be). This work was supported by **Research Council KUL:** GOA-Mefisto 666, IDO/99/003 and IDO/02/009 (Predictive computer models for medical classification problems using patient data and expert knowledge), several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0078.01 (structured matrices), G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), research communities (ICCoS, ANMMM); **IWT:** PhD Grants, **Belgian Federal Science Policy Office:** IUAP P5-22 ("Dynamical Systems and Control: Computation, Identification & Modelling"); **EU:** PDT-COIL, BIOPATTERN, eTUMOUR.

**Abstract**

A common frame of *template splines* that unifies the definitions of various spline families, such as smoothing, regression or penalized splines, is considered. The nonlinear nonparametric regression problem that defines the template splines can be reduced, for a large class of Hilbert spaces, to a parameterized regularized linear least squares problem, which leads to an important computational advantage. Particular applications of template splines include the commonly used types of splines, as well as other atypical formulations. In particular, this extension allows an easy incorporation of additional constraints, which is generally not possible in the context of classical spline families.

*Keywords*: nonparametric regression, reproducing kernel Hilbert space, splines.

# A class of template splines

Diana M. Sima *, Sabine Van Huffel

*ESAT-SCD, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.*

**Abstract**

A common frame of *template splines* that unifies the definitions of various spline families, such as smoothing, regression or penalized splines, is considered. The nonlinear nonparametric regression problem that defines the template splines can be reduced, for a large class of Hilbert spaces, to a parameterized regularized linear least squares problem, which leads to an important computational advantage. Particular applications of template splines include the commonly used types of splines, as well as other atypical formulations. In particular, this extension allows an easy incorporation of additional constraints, which is generally not possible in the context of classical spline families.

*Key words:* nonparametric regression, reproducing kernel Hilbert space, splines
*1991 MSC:* 62G08, 65D07

## 1 Introduction

We consider nonlinear nonparametric models of the form $y_i = f(t_i) + \epsilon_i$, where $y_i$ denotes a real-valued measured observation, $t_i$ is a regression abscissa from a certain given bounded real interval $\mathcal{I}$, $\epsilon_i$ is a zero-mean additive noise term, and the function $f : \mathcal{I} \to \mathbb{R}$ is an unknown nonlinearity to be determined.

The problem of estimating the unknown nonlinearity $f$, given measured data $(t_1, y_1), \ldots, (t_m, y_m)$, is a nonparametric regression problem. The range of applications for this type of problems is very wide: from density function estimations to econometric predictions, from biology to sociology. The literature on

* Corresponding author. Tel.: +32-16-32-11-43; fax: +32-16-32-19-70
*Email addresses:* diana.sima@esat.kuleuven.ac.be (Diana M. Sima), sabine.vanhuffel@esat.kuleuven.ac.be (Sabine Van Huffel).

nonparametric regression is also quite rich (see the monographs authored by Eubank (1999); Green and Silverman (1994); Härdle (1990); Wahba (1990)). But one tool that is almost always mentioned in connection with nonparametric regression problems is *splines*.

In this paper, we design a new and more general formulation, which encompasses many of the commonly used types of splines. We partly motivate this work by the need to link the family of penalized splines (Eilers and Marx, 1997) to the smoothing splines family (Wahba, 1990), whose theory is much more rigorous. In a recent article, Hall and Opsomer (2005) prove some basic statistical properties of consistency and expressions for mean squared errors for penalized splines. The extension that we propose is a super-family for both the smoothing splines and for penalized splines, among others. Good statistical properties that hold for smoothing splines, such as the optimality of model selection via generalized cross validation, still hold for the template spline extension. Moreover, with this general formulation we are able to solve not only smoothing problems, but also various constrained smoothing problems.

In section 2, we define template splines by generalizing the approach of Wahba (1990), which embedded the smoothing splines into the theory of reproducing kernel Hilbert spaces. We allow more freedom in choosing some essential elements of the template splines (compared to the smoothing splines definition), and this fact will play a role in widening the splines family.

In section 3 we show that by allowing these generalizations, we don't overly restrict the computational properties, since the template spline solution will also be computable via a linear parameterization by solving a regularized linear least squares minimization.

Derivations of the well-known types of splines from the more general template spline formulation and some illustrative examples are detailed in section 4.

A word about notation: we shall use bold-face symbols for vectors, for example: $\mathbf{y}$ is the vector $(y_1, \ldots, y_m)^\top$, and $\mathbf{t}$ is the vector $(t_1, \ldots, t_m)^\top$.

## 2 Template splines on reproducing kernel Hilbert spaces

Given the measured data $(t_1, y_1), \ldots, (t_m, y_m)$, the goal is to fit this data using a simple nonparametric model $y_i = f(t_i) + \epsilon_i$. Depending on the application, we have to define more specifically some properties to which a good $f$ should comply. The first step is to choose a function space where our search for a good $f$ can be restricted to. A general framework for function spaces is described next; it is a framework that has attractive theoretical properties, but which

also proves to be advantageous from a computational point of view, as it will be shown in Section 3.

### *2.1 Reproducing kernel Hilbert spaces*

The theory of reproducing kernel Hilbert spaces is a functional analysis tool that gives a sound foundation to the use of splines and other concepts in curve fitting, function estimation, model description or model building applications (Wahba, 2003).

A reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ is a linear function space, with an embedded inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and an induced norm $\| \cdot \|_{\mathcal{H}}$. It is a Hilbert space on some domain $\mathcal{I}$, thus it is complete with respect to the induced distance topology, but the most important property of a RKHS is that all pointwise evaluations can be written as bounded linear functionals. That is, for every $t \in \mathcal{I}$ the linear functional $L_t$ defined by the relation $L_t f = f(t)$ is bounded ($\|L_t f\|_{\mathcal{H}} \leq \text{const}_t \|f\|_{\mathcal{H}}$). As a consequence of Riesz's representation theorem (which says that every bounded linear functional has a *representer* in the Hilbert space), for each $t \in \mathcal{I}$ there exists an element $\eta_t \in \mathcal{H}$ such that

$$f(t) = \langle \eta_t, f \rangle_{\mathcal{H}}, \qquad \forall\, f \in \mathcal{H}.$$

In other words, each $t \in \mathcal{I}$ has a corresponding $\eta_t \in \mathcal{H}$ that "describes" all possible values of the functions in $\mathcal{H}$ at $t$ through the inner product of $\mathcal{H}$.

The function $K : \mathcal{I} \otimes \mathcal{I} \to \mathbb{R}$ defined as $K(s,t) := \langle \eta_s, \eta_t \rangle_{\mathcal{H}}$ for all $s, t \in \mathcal{I}$ is the *reproducing kernel* for $\mathcal{H}$. It is symmetric, positive definite and satisfies $\langle K(s,\cdot), K(t,\cdot) \rangle_{\mathcal{H}} = K(s,t)$.

There is a one-to-one relationship between the set of reproducing kernel Hilbert function spaces defined on an interval $\mathcal{I}$ and the set of positive definite functions defined on $\mathcal{I} \times \mathcal{I}$. Moreover, under general circumstances (*e.g.*, continuity and square-integrability), any positive definite function $K$ admits an eigenfunction-eigenvalue decomposition:

$$K(s,t) = \sum_{\nu=1}^{\infty} \sigma_\nu \Phi_\nu(s) \Phi_\nu(t),$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$, with $\sum_{\nu=1}^{\infty} \sigma_\nu^2 < \infty$, are the eigenvalues of $K$, and $\Phi_1, \Phi_2, \ldots$, an orthonormal sequence of square-integrable functions on $\mathcal{I}$, are called the eigenfunctions of $K$. $\Phi_1, \Phi_2, \ldots$, form a Hilbert basis for the RKHS that corresponds to the kernel $K$.

These properties imply good news from the practical point of view. Choosing the RKHS $\mathcal{H}$ can either be replaced by choosing a desired kernel $K$ or by

choosing a sequence of orthonormal functions that act as a basis or a generator family for $\mathcal{H}$. For this reason, in practical applications, the theoretical formulation of $\mathcal{H}$ is often omitted; typically, it is replaced by the definition of a certain sequence of generators for the "splines", such as truncated polynomials, piecewise polynomials or B-spline bases (de Boor, 1978).

### *2.2 Unconstrained smoothing*

When a space $\mathcal{H}$ is specified, we can already formulate the following least squares problem:

**Definition 1 (The $\mathcal{H}$-smoothing problem)** *Solving the least squares problem*

$$\min_{f \in \mathcal{H}} \sum_{i=1}^{m} (y_i - f(t_i))^2 \tag{1}$$

*for the function $f$ is called the $\mathcal{H}$-smoothing problem.*

Via $\mathcal{H}$-smoothing, the measured data $\mathbf{y} \in \mathbb{R}^m$ is *orthogonally projected* onto the space $\mathcal{H}(\mathbf{t})$, which is a vector space of the discretizations of all the functions in $\mathcal{H}$ at the abscissas $t_1, \ldots, t_m$. In other words, $\mathcal{H}$-smoothing finds the function $\hat{f}$ in $\mathcal{H}$ that best approximates the data in $\mathbf{y}$ at given abscissas $\mathbf{t}$. For the values of $t \in \mathcal{I}$ that are not among the elements of $\mathbf{t}$, $\hat{f}$ can be used as predictor.

### *2.3 Constrained smoothing and template splines*

To impose additional constraints such as smoothness, monotonicity, convexity, equality or inequality relations, on the estimated function $f$, it is possible in some cases to revise the definition of the space $\mathcal{H}$, such that any function in $\mathcal{H}$ satisfies the extra constraints. However, the design of $\mathcal{H}$ in this case can become tedious. An alternative method is based on the projection framework for constrained smoothing developed by Mammen et al. (2001); basically, the additional constraints are imposed in a sequential manner: "smooth then constrain."

Here, we propose a one-step solution that is feasible and efficient for a wide class of constraints that can be written as (semi)norm conditions.

Let $\mathcal{P}$ denote a linear operator from $\mathcal{H}$ to $\mathcal{H}'$, where $\mathcal{H}'$ is a certain normed space.

**Definition 2 (The $\mathcal{P}$-constrained $\mathcal{H}$-smoothing problem)** *Solving the con-*

*strained least squares problem*

$$\min_{f\in\mathcal{H}} \sum_{i=1}^{m} (y_i - f(t_i))^2, \qquad \textit{such that} \quad \|\mathcal{P}f\|_{\mathcal{H}'} \leq \delta \quad \textit{for a fixed } \delta \geq 0, \tag{2}$$

*for the function $f$ is called the $\mathcal{P}$-constrained $\mathcal{H}$-smoothing problem.*

The value $\|\mathcal{P}f\|_{\mathcal{H}'}$ is generally a measure of the quality of an arbitrary $f$ with respect to the desired constraints, or to the prior knowledge on properties that good estimators $f$ should satisfy. Sometimes, if no such prior knowledge is available, $\|\mathcal{P}f\|_{\mathcal{H}'}$ could also be a measure of parsimony for the model $f$.

Using a Lagrange multiplier argument, we arrive at the following:

**Definition 3 (The template spline problem)** *The following minimization is the* template spline *minimization problem:*

$$\min_{f\in\mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} (y_i - f(t_i))^2 + \lambda \|\mathcal{P}f\|^2_{\mathcal{H}'}. \tag{3}$$

This formulation is equivalent to the $\mathcal{P}$-constrained $\mathcal{H}$-smoothing problem, for a certain relation between $\lambda \geq 0$ and $\delta \geq 0$. However, $\lambda$ or $\delta$ are usually not exactly known in practice. The penalty formulation of the template spline problem (3) shows that a trade-off between closeness of the model $f$ to the observations and "quality" of the model – via the $\mathcal{P}$-constraint – can be ensured by choosing an appropriate $\lambda \geq 0$. In Section 3.2 we focus on the generalized cross validation (GCV) criterion that can be used for choosing the so-called *smoothing* (or *penalty* or *regularization*) parameter $\lambda$, given the measured data.

**Remark 4 (More general formulations)** *Sometimes in spline literature, the function evaluation operator that appears in the sum of squares of problems (1, 2, 3), $t_i \to f(t_i)$, is replaced by a more general bounded linear operator $L_i$,* i.e., *$t_i \to L_i f$. All results in this paper are straightforwardly generalizable to this setting.*

*Moreover, if the zero-mean noise term $\epsilon_i$ is assumed to have a known covariance matrix other than the identity matrix, the least squares term in all minimization problems (1, 2, 3) can be easily transformed into* weighted least squares *formulations, using the inverse of the noise covariance matrix as weight matrix.*

## 3 Computing template splines

### *3.1 Transformation to a linear least squares problem*

The following theorem says that when $\mathcal{H}$ is a RKHS and $\mathcal{P}$ has some general properties (see below), the solution of the template problem (3) has a representation in terms of the reproducing kernel and/or its eigenfunctions, and it can be easily computed via linear least squares or linear equations. In particular, problem (1) can also be efficiently solved as a linear least squares problem.

Let $\mathcal{H}_\mathcal{P}$ denote the null space of $\mathcal{P}$ in $\mathcal{H}$ (that is, $\mathcal{H}_\mathcal{P} := \ker \mathcal{P} := \{f \in \mathcal{H} : \mathcal{P}f = 0\}$), and assume it is finite dimensional, of dimension $d$. For instance, if $\mathcal{P}$ is the $d^{\text{th}}$ order derivative operator on some space $\mathcal{H}$ of (infinitely) continuously differentiable functions, then $\ker \mathcal{P}$ is spanned by the finite basis $\{1, t, t^2, \ldots, t^{d-1}\}$.

Assume that $\mathcal{P}$ preserves orthogonality relations from $\mathcal{H}$ to $\mathcal{H}'$: thus, if $\langle f, g\rangle_\mathcal{H} = 0$, then $\langle \mathcal{P}f, \mathcal{P}g\rangle_{\mathcal{H}'} = 0$.

**Theorem 5 (generalized from Wahba (1990), Th. 1.3.1)** *The solution of the template spline problem (3), for a fixed $\lambda \geq 0$, has a closed-form expression given by*

$$\hat{f}_\lambda := \sum_{k=1}^{d} a_k \phi_k + \sum_{i=1}^{m} b_i \mu_i, \tag{4}$$

*where*

- *$\phi_1, \ldots, \phi_d$ are a basis for $\mathcal{H}_\mathcal{P}$, ($d := \dim \mathcal{H}_\mathcal{P}$),*
- *$\mu_i = K(t_i, \cdot)$, for $i = 1, \ldots, m$,*
- *the coefficients $\mathbf{a} := (a_1, \ldots, a_d)^\top$ and $\mathbf{b} := (b_1, \ldots, b_m)^\top$ are given by*

$$\begin{aligned} \mathbf{a} &= \left(A^\top A - A^\top B(B^2 + m\lambda C)^{-1} BA\right)^{-1} A^\top \left(I_m - B(B^2 + m\lambda C)^{-1} B\right) \mathbf{y}, \\ \mathbf{b} &= (B^2 + m\lambda C)^{-1} B(\mathbf{y} - A\mathbf{a}), \end{aligned} \tag{5}$$

*where the matrices $A$ ($m \times d$), $B$ ($m \times m$) and $C$ ($m \times m$) are computed as*

$$\begin{aligned} A &= \{(\phi_k(t_l)\}_{(l=1,\ldots,m;\ k=1,\ldots,d)}, \\ B &= \{K(t_i, t_k)\}_{(i,j=1,\ldots,m)} = \{\langle \mu_i, \mu_j \rangle_\mathcal{H}\}_{(i,j=1,\ldots,m)}, \\ C &= \{\langle \mathcal{P}\mu_i, \mathcal{P}\mu_j \rangle_{\mathcal{H}'}\}_{(i,j=1,\ldots,m)}. \end{aligned} \tag{6}$$

**PROOF.**

Since the set $\{\phi_1, \ldots, \phi_d\}$ is a basis for $\mathcal{H}_\mathcal{P}$, any $f \in \mathcal{H}$ can be written as $f = \sum_{k=1}^d a_k \phi_k + f^\perp$, for some $a_1, \ldots, a_d \in \mathbb{R}$ and an $f^\perp \in \mathcal{H}_\mathcal{P}^\perp$. It will be shown that $f^\perp$ can be further decomposed such that the optimal solution of the minimization (3) has a finite decomposition in $\mathcal{H}$. To this end, write $f^\perp = \sum_{i=1}^m b_i \mu_i + \rho$, where $\mu_i := K(t_i, \cdot)$ and $\rho \in \mathcal{H}$ is orthogonal to $\phi_1, \ldots, \phi_d$ and to $\mu_1, \ldots, \mu_m$.

The least squares objective to be minimized becomes:

$$\begin{aligned}
\mathcal{F}(f) :=& \frac{1}{m} \sum_{i=1}^m (y_i - f(t_i))^2 + \lambda \|\mathcal{P} f\|_{\mathcal{H}'}^2 \\
=& \frac{1}{m} \sum_{i=1}^m \left( y_i - \sum_{k=1}^d a_k \phi_k(t_i) - \sum_{j=1}^m b_j \mu_j(t_i) - \rho(t_i) \right)^2 \\
&+ \lambda \left\| \mathcal{P} \left( \sum_{k=1}^d a_k \phi_k + \sum_{j=1}^m b_j \mu_j + \rho \right) \right\|_{\mathcal{H}'}^2 \\
=& \frac{1}{m} \|\mathbf{y} - A\mathbf{a} - B\mathbf{b}\|_2^2 + \lambda \mathbf{b}^\top C \mathbf{b} + \lambda \left\| \mathcal{P} \rho \right\|_{\mathcal{H}'}^2 ,
\end{aligned}$$

where $A$, $B$ and $C$ are given in (6). We used the facts that $\rho(t_i) = \langle \rho, K(t_i, \cdot) \rangle_\mathcal{H} = \langle \rho, \mu_i \rangle_\mathcal{H} = 0$ from the choice of $\rho$ orthogonal to $\mu_1, \ldots, \mu_m$, that $\phi_1, \ldots, \phi_d$ are in the null space of $\mathcal{P}$, and that $\rho$ being orthogonal to $\sum_{j=1}^m b_j \mu_j$ in $\mathcal{H}$ implies $\langle \sum_{j=1}^m \mathcal{P} \mu_j, \mathcal{P} \rho \rangle_{\mathcal{H}'} = 0$, by the "preservation of orthogonality" property of $\mathcal{P}$.

For $f$ to be a minimizer of $\mathcal{F}$, we must have $\|\mathcal{P}\rho\|_{\mathcal{H}'} = 0$. This implies $\rho \in \ker \mathcal{P}$, which contradicts the choice of $\rho$ orthogonal to the basis of $\ker \mathcal{P}$; thus, the only possibility is $\rho = 0$.

The normal system of equations for the linear least squares problem that we obtained is:

$$\begin{cases} A^\top A \mathbf{a} + A^\top B \mathbf{b} = A^\top \mathbf{y}, \\ B^\top A \mathbf{a} + (B^\top B + m\lambda C) \mathbf{b} = B^\top \mathbf{y}. \end{cases} \tag{7}$$

Since $B$ is symmetric positive definite, the closed-form expressions (5) are easily obtained. □

In summary, the theorem gives us a simple way to express the minimization (3) as a problem that is linearly parameterized by the vectors $\mathbf{a}$ and $\mathbf{b}$, more precisely, as a *regularized linear least squares problem*:

$$\min_{\mathbf{a} \in \mathbb{R}^d, \, \mathbf{b} \in \mathbb{R}^m} \frac{1}{m} \|\mathbf{y} - A\mathbf{a} - B\mathbf{b}\|_2^2 + \lambda \mathbf{b}^\top C \mathbf{b}. \tag{8}$$

In practice, the smoothing parameter $\lambda$ will not be known *a priori*; the methods for choosing a good $\lambda$ often rely on optimizing some criterion and involves many evaluations of spline solutions (from their coefficients $\mathbf{a}$, $\mathbf{b}$), for various values of $\lambda$. For this reason, it is required to have efficient algorithms for the computation of coefficients $\mathbf{a}$ and $\mathbf{b}$, instead of the expressions in (5).

The regularized least squares problem of minimizing

$$\frac{1}{m}\|\mathbf{y} - A\mathbf{a} - B\mathbf{b}\|_2^2 + \lambda \mathbf{b}^\top C\mathbf{b} = \frac{1}{m}\left\| \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} - \begin{bmatrix} A & B \\ 0 & \sqrt{m\lambda}D \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|_2^2,$$

where $D$ is a Cholesky factor of the symmetric positive definite matrix $C$ ($D^\top D = C$), can be efficiently solved if we compute in advance the generalized singular value decomposition (GSVD) (Golub and Van Loan, 1996) of the matrix pair $\left( \begin{bmatrix} A & B \end{bmatrix}, \begin{bmatrix} 0 & D \end{bmatrix} \right)$.

Many methods for choosing "optimal" regularization parameter $\lambda$ can be employed in the context of template splines. A recent simulation study (Lee, 2003) compares several parameter selection procedures (leave-one-out cross validation, generalized cross validation, Mallows' $C_p$, Akaike's information criterion, and risk estimation methods); the conclusion is that none of the methods can be overall the best.

Here we present the application of generalized cross validation (Craven and Wahba, 1979) to template splines. The GCV criterion aims at minimizing the function $G(\lambda) := \frac{\|\mathbf{y} - A\mathbf{a}_\lambda - B\mathbf{b}_\lambda\|_2^2}{[\mathrm{Tr}(I_m - H(\lambda))]^2}$ where $(\mathbf{a}_\lambda, \mathbf{b}_\lambda)$ denote the optimal solution vectors that are obtained from solving problem (8) holding the regularization parameter at the fixed value $\lambda$, and $H(\lambda)$ is the *influence matrix* (or *hat* matrix, or *smoother* matrix) defined as the $m \times m$ matrix that makes the equality $H(\lambda)\mathbf{y} = A\mathbf{a}_\lambda + B\mathbf{b}_\lambda$ hold. Its formula reads:

$$H(\lambda) = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} A^\top A & A^\top B \\ B^\top A & B^\top B + m\lambda C \end{bmatrix}^{-1} \begin{bmatrix} A^\top \\ B^\top \end{bmatrix}.$$

GCV was originally introduced for problems such as ridge regression (Golub et al., 1979) and smoothing splines (Craven and Wahba, 1979), which belong to the same category of regularized linear least squares problems as the template splines do. For these problems, it is possible to show that the GCV criterion is a *predictive mean squared error* criterion. In the context of template splines this will mean that, under reasonable assumptions and in the asymptotic situation,

the values chosen for $\lambda$ by minimizing the GCV function, when the sample size grows to infinity, converge to the optimal $\lambda$ that corresponds to the spline model that is closest to the *true* underlying model.

For efficient computations, the GSVD of the matrix pair $\left(\begin{bmatrix} A & B \end{bmatrix}, \begin{bmatrix} 0 & D \end{bmatrix}\right)$ can be used, because in this case, the matrix that must be inverted in the formula of $H(\lambda)$ becomes a diagonal matrix, and, thus, computing this inverse is very fast, for any $\lambda$. (See Hansen (1997).) Interesting techniques for making the optimization of the GCV function fast for large-scale problems are described by Golub and von Matt (1997), making use of a randomization method to compute the trace of a large matrix.

## 4 Examples

In this section, we exemplify the described theory, by first showing that some classical spline families belong to our generalization. Afterwards, we design some examples that are atypical for the classical splines, but which can however be successfully tackled by the template splines.

### *4.1 Smoothing, regression and penalized splines as template splines*

Smoothing splines and penalized splines start from the same idea of using a weighted penalty term in order to obtain a desired degree of smoothness of the reconstructed nonparametric model. Regression splines rely on a different technique to obtain a smoother or a rougher approximation: tuning the number of basis splines, as explained in §4.1.2.

Despite their similarities, the penalized splines cannot be seen as a subset of smoothing splines (or *vice-versa*), mainly because of the fact that their penalty terms are of totally different nature.

But with the template splines, we generalize these approaches by allowing the operator $\mathcal{P}$ to be more general than just an orthogonal projection (as for smoothing splines) and to take value in other normed spaces (such as $\mathbb{R}^n$, for penalized splines). In the meantime, the good properties of smoothing splines, such as the possibility of using GCV for choosing the penalty parameter, are still kept.

### *4.1.1 Smoothing splines as template splines*

Smoothing splines (Wahba, 1990) are a subset of the template splines family, obtained when the space $\mathcal{H}'$ is the same as the space $\mathcal{H}$, and the operator $\mathcal{P}$ is an orthogonal projection.

A classical example of a smoothing spline is the *natural spline*. The space $\mathcal{H}$ is defined as the Sobolev space,

$$\mathcal{H} := \{f : [0,1] \to \mathbb{R} : f, f', \ldots, f^{(d-1)} \text{ are absolutely continuous and } f^{(d)} \in \mathcal{L}_2\},$$

where $f^{(k)}$ denotes the $k^{\text{th}}$ derivative and $\mathcal{L}_2$ denotes the space of square integrable functions. $\mathcal{H}$ is endowed with the square seminorm:

$$\|f\|_{\mathcal{H}}^2 := \sum_{k=1}^{d} \left[f^{(k)}(0)\right]^2 + \int_0^1 \left[f^{(d)}(u)\right]^2 du.$$

The corresponding reproducing kernel is

$$K(s,t) := \sum_{k=1}^{d} \phi_k(s)\phi_k(t) + \int_0^1 G_d(s,u)G_d(t,u)dt,$$

where each $\phi_k$ (for $k = 1, \ldots, d$) denotes the monomial $\phi_k(t) = \frac{t^{k-1}}{(k-1)!}$ and $G_d$ is the Green function $G_d(t,u) = \frac{(t-u)_+^{d-1}}{(d-1)!}$, with $(x)_+ = x$ for $x \geq 0$ and 0 otherwise.

The operator $\mathcal{P}$ is simply defined as the orthogonal projection onto the subspace

$$\mathcal{H}_{\mathcal{P}} := \{f \in \mathcal{H} : f(0) = f(1) = f'(0) = f'(1) = \cdots = f^{(d-1)}(0) = f^{(d-1)}(1) = 0\}.$$

The optimal natural spline solution solves the problem:

$$\min_{f \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} (y_i - f(t_i))^2 + \lambda \int_0^1 \left[f^{(d)}(u)\right]^2 du.$$

Thus, the natural spline reconstruction from noisy measurements aims at $d$-order smoothness, with no boundary conditions.

The restriction of having $\mathcal{H}'$ equal to $\mathcal{H}$ can be exploited in the smoothing spline solution derivation (Wahba, 1990) to yield a computational advantage. The smoothing spline solution can still be obtained by solving a regularized linear least squares problem (details in (Wahba, 1990, Chapter 1)). But $\mathcal{H}' = \mathcal{H}$ implies that for smoothing splines we can redefine the matrices that appear in our Theorem 5 and have $B = C$, which gives a simpler numerical solution for the smoothing spline, since a QR decomposition can be used instead of a GSVD.

### *4.1.2 Regression splines as template splines*

A regression spline (Friedman, 1991) is a piecewise polynomial function that is smooth at the jointure points (knots) up to the degree of the polynomials minus 1. Typical polynomial functions that are used with regression splines are truncated polynomials or B-splines (de Boor, 1978).

The space $\mathcal{H}$ can be defined as the span of the considered polynomial set. An example is the span of the truncated polynomials of degree $q$, span $\{t, t^2, \ldots, t^q, (t-\tau_1)_+^q, (t-\tau_2)_+^q, \ldots, (t-\tau_N)_+^q\}$, where $\tau_1, \ldots, \tau_N$ are $N$ given knots, and $(x)_+$ denotes, as before, $x$ or 0, depending whether $x$ is nonnegative or negative. Another example is the span of B-spline basis. The widely used B-splines are defined recursively such that, at the given knots $\tau_1, \ldots, \tau_N$, each of the functions in the B-spline basis set is $q$-times continuously differentiable. The recursion formula reads:

$$v_i^{\{1\}}(t) = \begin{cases} 1, \text{ if } \tau_i \leq t \leq \tau_{i+1} \\ 0, \text{ otherwise,} \end{cases}; \; v_i^{\{k\}}(t) = \frac{v_i^{\{k-1\}}(t) \cdot (t - \tau_i)}{\tau_{i+k-1} - \tau_i} + \frac{v_{i+1}^{\{k-1\}}(t) \cdot (\tau_{i+k} - t)}{\tau_{i+k} - \tau_{i+1}}, \tag{9}$$

and the $q$-order basis that defines $\mathcal{H}$ as span $\{v_1^{\{q\}}, \ldots, v_{N-q}^{\{q\}}\}$, contains thus $N - q$ functions.

For fitting with regression splines, we can set the operator $\mathcal{P}$ as the identically zero mapping (and $\mathcal{H}' = \{0\}$). Thus the penalty part disappears from the problem (3), and we are left with a simple least squares problem.

Sometimes, the number of knots $N$ and the knot positions must be optimized before performing regression. In the situation when we consider equally-spaced knots (*i.e.*, uniform regression splines), only the integer parameter $N$ has to be chosen. Therefore, we consider as hyperparameter $\lambda$ the number of knots $N$ and the generalized cross validation (or other typical method for model order selection) can then be used in order to choose an appropriate value for $N$. The parameter $N$ controls in this case the bias-variance trade-off. A large $N$ means that the measured $\mathbf{y}$ can be modeled with more fidelity, and a smaller $N$ means that the reconstructed spline is smooth.

### *4.1.3 Penalized splines as template splines*

The simple smoothing method of *penalized splines* was introduced by Eilers and Marx (1997) (and in very similar terms, by Ruppert and Carroll (1996)). It is based on (uniform) regression splines, but with an added penalty term, weighted by a certain regularization parameter $\lambda$. An interesting detail is found in (Ruppert, 2002), namely that the number of knots can be set to an arbitrary value (a number large enough, but always less than the number of regression

points); the technique of penalized splines controls the degree of smoothness only through the parameter $\lambda$.

In formal notation, let $v_1, \ldots, v_n$ denote a basis family of truncated polynomials or of B-splines of degree $q$. Then, given the measurements $y_1, \ldots, y_m$, a function $f$ can be modeled as a linear combination of $v_1, \ldots, v_n$, *i.e.*, $f = \sum_{k=1}^{n} c_k v_k$, by minimizing the penalized least squares function

$$\min_{\mathbf{c}\in\mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} \left( y_i - \sum_{k=1}^{n} c_k v_k(t_i) \right)^2 + \lambda \|\Delta \mathbf{c}\|^2, \tag{10}$$

where $\Delta$ is a finite difference operator (of a certain chosen order) and $\lambda$ is a positive scalar parameter that controls the degree of smoothness.

From the formulation (10) it is clear that the penalized splines are also a particular example of template splines. We define the space $\mathcal{H}$ as the span of the function family $\{v_1, \ldots, v_n\}$, and the space $\mathcal{H}'$ as the $n$-dimensional Euclidean space $\mathbb{R}^n$. The operator $\mathcal{P}$ can be easily set using the canonical decomposition of any element $f$ in $\mathcal{H}$, as $\mathcal{P}(f) = \mathcal{P}\left(\sum_{k=1}^{n} c_k v_k\right) = \Delta \mathbf{c}$, where $\Delta$ is the same approximation matrix for a finite difference operator that is used in the penalized splines problem (10).

### *4.2 Other applications of template splines*

#### *4.2.1 Smoothing in transformed spaces*

We consider the abscissa interval $\mathcal{I}$ as the "time axis" (and therefore the measurements $y_1, \ldots, y_m$ are given in the "time-domain"), but we are interested in imposing smoothness to the *Fourier transform* of the solution $f$.

Let $\mathcal{F}$ be the Fourier transform and $\mathcal{F}^{-1}$ denote the inverse Fourier transform. Consider a RKHS $\mathcal{H}$ of functions defined on the time-domain $\mathcal{I}$. Then the image space is $\mathcal{H}' := \mathcal{F}\mathcal{H} = \{\mathcal{F}(f) \ : \ f \in \mathcal{H}\}$. We can measure smoothness in $\mathcal{H}'$ with the aid of a projection $\mathcal{Q}$ (see, for instance, the natural spline type of smoothing in section 4.1.1), *i.e.*, the norm $\|\mathcal{Q}g\|_{\mathcal{H}'}$ (for any $g$ in $\mathcal{H}'$) is the measure that we use for smoothness. It is then possible to define the mapping $\mathcal{P} : \mathcal{H} \to \mathcal{H}'$ by the composition

$$\mathcal{P}(f) = \mathcal{Q}(\mathcal{F}(f)).$$

Solving the template spline problem (3) in this case is equivalent to solving

$$\min_{g\in\mathcal{H}'} \frac{1}{m} \sum_{i=1}^{m} \left( y_i - (\mathcal{F}^{-1} \circ g)(t_i) \right)^2 + \lambda \left\| \mathcal{Q}g \right\|_{\mathcal{H}'}^2,$$

which is a template spline problem in the space $\mathcal{H}'$.

### *4.2.2 Constrained smoothing*

Classical smoothing splines or regression splines are designed to smooth scattered data in order to obtain meaningful approximations. However, additional constraints are usually hard to impose. (See for instance (Kuijt and van Damme, 2001), where linear programming methods are used for several constrained smoothing problems.) With templates splines, constraints can be imposed through the penalty term. Tuning the penalty coefficient allows to slightly break the constraints, which is advantageous for identifying situations when the data does not obey the constraints.

In this paragraph, we give two numerical examples from an application where template splines are designed for constrained smoothing problems. In the first example, we show that imposing a symmetry constraint helps finding a reliable model for noisy data that comes from a symmetric function. The second example illustrates that when using a wrong constraint, the GCV value for $\lambda$ gives a model that clearly breaks the constraint.

**Example 6** *We consider several scalar functions on an interval* $[-\alpha, \alpha]$ *that are symmetric with respect to the vertical axis through zero. We generate discretizations and add Gaussian noise. We then compute from the noisy data two models, both based on a B-spline basis: a penalized spline where a second order derivative operator is used to obtain smoothness, and a template spline where the penalty contains the second order derivative for smoothness, but also a term for symmetry, which is governed by a "V-shaped" matrix* $[I \ -J]$*, where* $I$ *is the identity matrix and* $J$ *is its mirror reflection. We choose the penalty coefficient by GCV in both models, but we observe that using the symmetry constraint gives models that have smaller mean squared error from the original simulated functions, than when no constraint is imposed. See Figure 1 for illustration.*

**Example 7** *We consider now as original function a function that is not symmetric with respect to the vertical axis through zero,* e.g.*, the arctangent function. We compute however a template spline with symmetry penalty. As illustrated in Figure 2, the fit is symmetric, but does not reflect the data, when we choose a penalty parameter* $\lambda$ *which is quite big. On the other hand, the GCV criterion chooses a* $\lambda$ *that corresponds to a spline reconstruction that is reasonable for the given data, and ignores the inappropriate symmetry constraint.*

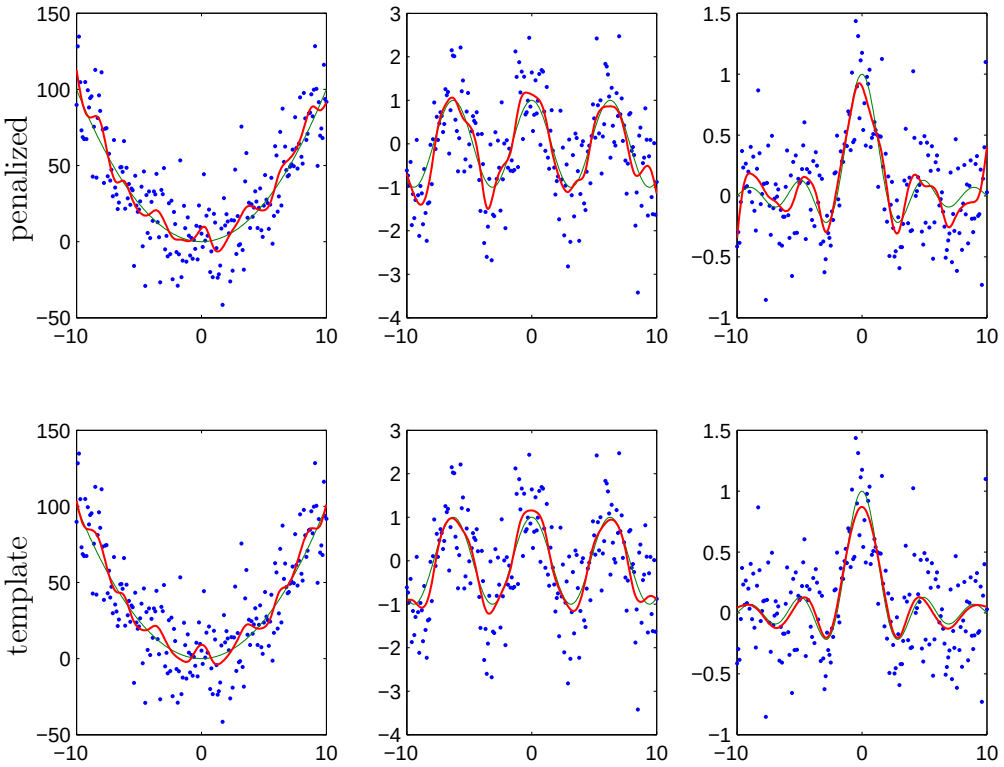



Fig. 1. In all plots, the thin solid line (green) is the simulated symmetric function, respectively: $x^2$, $\cos(x)$ and $\mathrm{sinc}(x)$; the dots are the noise corrupted data. The top row of plots shows as thick (red) lines the penalized spline reconstructions, while in the bottom row these thick lines represent the template splines with symmetry constraint.

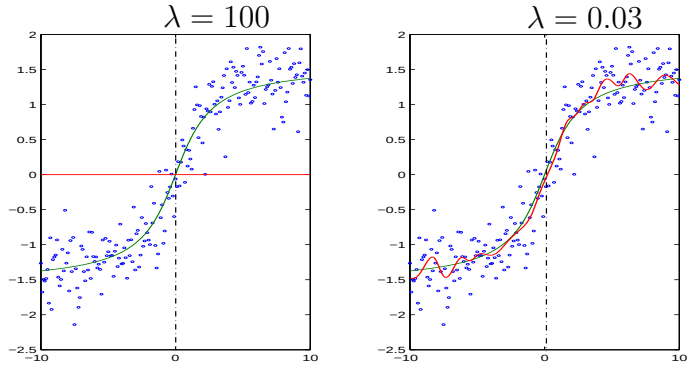



Fig. 2. On the left, the spline corresponds to a very large value of the penalty parameter, thus the symmetry is strongly imposed, yielding an almost zero solution. On the right, the penalty parameter is computed with GCV, and the symmetry constraint is completely ignored, but the data is reasonably fitted.

## 5 Conclusions

The template spline introduced in this paper generalizes some of the most used families of splines from literature. An issue that we find important is

that, due to this common framework, the family of penalized splines can be put in its own right beside the classical family of smoothing splines.

Template splines solutions are quite easily computable, since a (regularized) linear least squares optimization can be solved instead.

Template splines can constitute a solid basis for nonparametric regression problems where the "data acquisition space" is different from the space where interesting properties appear, or where strong or weak constraints should be imposed.

## Acknowledgements

Diana Sima is a research assistant and Dr. Sabine Van Huffel is a full professor at the Katholieke Universiteit Leuven, Belgium. Research supported by **Research Council KUL:** GOA-AMBioRICS, several PhD/postdoc & fellow grants; **Flemish Government: FWO:** PhD/postdoc grants, projects, G.0407.02 (support vector machines), G.0269.02 (magnetic resonance spectroscopic imaging), G.0270.02 (nonlinear Lp approximation), G.0360.05 (EEG, Epileptic), research communities (ICCoS, ANMMM); **IWT:** PhD Grants; **Belgian Federal Science Policy Office:** IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'); **EU:** PDT-COIL, BIOPATTERN, ETUMOUR.

## References

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik 31, 377–403.

de Boor, C., 1978. A Practical Guide to Splines. Springer, Berlin.

Eilers, P. H. C., Marx, B. D., 1997. Flexible smoothing with B-splines and penalties. Statistical Science 11 (2), 89–121.

Eubank, R., 1999. Nonparametric Regression and Spline Smoothing. Marcel Dekker, New York.

Friedman, J. H., 1991. Multivariate adaptive regression splines (with discussion). Annals of Statistics 19 (1).

Golub, G. H., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215–223.

Golub, G. H., Van Loan, C. F., 1996. Matrix Computations, 3rd Edition. Johns Hopkins University Press, Baltimore, Maryland.

Golub, G. H., von Matt, U., 1997. Generalized cross-validation for large scale problems. In: Van Huffel, S. (Ed.), Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling. SIAM, pp. 139–148.

Green, P. J., Silverman, B. W., 1994. Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman and Hall, London.

Hall, P., Opsomer, J., 2005. Theory for penalised spline regression. Biometrika 92, 105–118.

Hansen, P. C., 1997. Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion. SIAM, Philadelphia.

Härdle, W., 1990. Applied Nonparametric Regression. Econometric Society Monographs. Cambridge University Press.

Kuijt, F., van Damme, R., 2001. A linear approach to shape preserving spline approximation. Advances in Computational Mathematics 14, 25–48.

Lee, T. C. M., 2003. Smoothing parameter selection for smoothing splines: A simulation study. Computational Statistics and Data Analysis 42 (1-2), 139–148.

Mammen, E., Marron, J., Turlach, B., Wand, M., 2001. A general projection framework for constrained smoothing. Statistical Science 16 (3), 232–248.

Ruppert, D., 2002. Selecting the number of knots for penalized splines. Journal of Computational and Graphical Statistics 11, 735–757.

Ruppert, D., Carroll, R., August 1996. A simple roughness penalty approach to regression spline estimation. Tech. Rep. 1167, School of OR&IE, Cornell University.

Wahba, G., 1990. Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics 59. SIAM.

Wahba, G., August 2003. An introduction to reproducing kernel Hilbert spaces and why are they so useful. In: Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003). Rotterdam.