Design for smooth models over complex regions

Peter Curtis and Hugo Maruri-Aguilar

Abstract Smooth supersaturated models are a modelling alternative for computer experiments. They are polynomial models that behave like splines and allow fast computations. In this contribution we use the Gram-Schmidt orthogonal decomposition to build smooth supersaturated models over complex regions and then perform a two stage modelling and design strategy. We apply our methodology in a complex example taken from the literature of soap film smoothing.

1 Introduction

Interpolating splines were first proposed by Schoenberg (1964) and they are defined as the solution $y(x)$ that minimizes a measure of roughness $\Psi_m(y)$ given by

$$\Psi_m(y) = \int_a^b (y^{(m)})^2 dx \tag{1}$$

when searching among all interpolating functions for a given data set, see [10]. This problem is solved with interpolating polynomial splines of degree $2m-1$. We are interested in the case involving second derivatives, i.e. $m = 2$ and the interpolating spline is a cubic spline [5].

Thin plate splines extend the theory of splines for multivariate x. They minimize an extension of the criterion Ψ_2 above, which in the bivariate case is $\Psi_2(y) = \int\int \left(\left(\frac{\partial^2 y}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 y}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 y}{\partial x_2^2}\right)^2 \right) dx_1 dx_2$, where the integral is computed over

Peter Curtis
Queen Mary, University of London, Mile End Road E1 4NS London, e-mail: P.R.Curtis@qmul.ac.uk

Hugo Maruri-Aguilar
Queen Mary, University of London, Mile End Road E1 4NS London, e-mail: H.Maruri-Aguilar@qmul.ac.uk

all $\mathbb{R}^2$, see [8, 7]. If interpolation is not required, spline smoothing may be used to model and the criterion to minimize is a linear combination of the residual sum of squares and the roughness measure Ψ_2, see [11]. A polynomial regression alternative under Ψ_2 constraints is described in [2]. The literature for spline modelling and smoothing is vast, we refer the reader to [11, 5].

This paper is concerned with design for smooth supersaturated models (SSM). These models are polynomial interpolators that minimize the same measure of curvature used for splines and are thus competitive alternatives to splines. SSM are linear polynomials of high degree with more terms than the number of observations, and those extra degrees of freedom are used to achieve a spline-like behavior over a specified smoothing region, see [3]. Although the convergence to splines is guaranteed asymptotically [1], in practice a good approximation to splines is achieved after adding a few extra terms thus the computational cost of SSM is not particularly burdensome. The smoothing region for SSM can be arbitrarily defined to allow modelling over difficult, non-standard regions.

In this contribution we study two specific aspects involving SSM. The first is to build orthogonal bases for our model. This is achieved by a Gram-Schmidt procedure. Having an orthogonal SSM enables us to efficiently model and tackle design for smoothness criteria in non-standard regions and this is the second aspect we study. This paper is a first attempt to use SSM for non-standard regions, where spline and smoothing methodology is still largely under development, [9, 12, 13]. To the best of our knowledge no attempts on design have been produced for non-standard regions, and we differentiate from the literature on designs for spline models [14, 6].

The order of the paper is as follows. We first review the SSM method, with emphasis on one dimensional computations for the sake of brevity. We present an orthogonalization procedure for the SSM bases. We then briefly state the general multidimensional SSM and establish our design approach and illustrate with a non-standard example and the roughness measure Ψ_2. This same example is reworked to allow direct comparison with soap film smoothing [13] where instead of using Ψ_2 we average a distortion measure over a region of interest:

$$J(y) = \int\int \left(\frac{\partial^2 y}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2} \right)^2 dx_1 dx_2. \tag{2}$$

2 Smooth supersaturated model

Consider the problem of finding an interpolator to n data points while simultaneously minimizing a measure of roughness. Note than rather than searching over the space of functions with second derivatives, we are using polynomial functions and thus the existence of second derivatives is guaranteed in the development that follows. Indeed the vector spaces implied by our models are Sobolev spaces [11]. They admit a seminorm induced in our case by Ψ_2 as shown below.

2.1 Polynomial formulation and roughness computation

The available data is $(x_1,y_1),\ldots,(x_n,y_n)$, where no two values of factor x are repeated. Let the interpolator be the following polynomial in x with real coefficients

$$y(x) = f(x)^T\theta = \sum_{i=0}^{N-1} \theta_i x^i \tag{3}$$

with $f(x)^T = (1,x,x^2,\ldots,x^{N-1})$ and $\theta^T = (\theta_0,\ldots,\theta_{N-1})$. We assume that $N > n$, i.e. the model has more terms than data points. We compute the design-model matrix X which is of size $n \times N$. To compute Ψ_2 we need the following definition.

Definition 1. Let $\mathscr{X} \subset \mathbb{R}$ be a closed bounded region; let $s(x)$ and $t(x)$ be univariate polynomial functions. We define $\langle s(x),t(x)\rangle := \int_{\mathscr{X}} s''(x)t''(x)dx$.

The roughness of a polynomial $y(x)$ is $\Psi_2(y) = \langle y,y\rangle$. The following Lemma establishes the roughness of $y(x)$. We omit the proof and give an example of it.

Lemma 1. *Let $y(x),\theta$ and $f(x)$ be as in Equation (3). Let K be the matrix of inner products $\langle x^i,x^j\rangle$ for $i,j = 0,\ldots,N-1$. Then*

1. *The roughness of $y(x)$ is $\Psi_2(y) = \theta^T K\theta$, and*
2. *The matrix K is of the form $K = DDMD^TD^T$ where M is the usual moment matrix of terms in $f(x)$, computed over the region $\mathscr{X}$. The matrix D is a square matrix of size N with zero entries apart from entry $D_{i+1,i}$ that equals i for $i = 1,\ldots,N-1$.*

Example 1. Consider $\mathscr{X} = [0,1]$ and the model of Equation (3) with $N = 4$, i.e. $f(x) = (1,x,x^2,x^3)^T$ and $\theta^T = (\theta_0,\theta_1,\theta_2,\theta_3)$. The matrices K,D and M are

$$K = \begin{pmatrix} 0&0&0&0\\ 0&0&0&0\\ 0&0&4&6\\ 0&0&6&12 \end{pmatrix},\ D = \begin{pmatrix} 0&0&0&0\\ 1&0&0&0\\ 0&2&0&0\\ 0&0&3&0 \end{pmatrix} \text{ and } M = \begin{pmatrix} 1&1/2&1/3&1/4\\ 1/2&1/3&1/4&1/5\\ 1/3&1/4&1/5&1/6\\ 1/4&1/5&1/6&1/7 \end{pmatrix},$$

so that $\Psi_2(y) = \theta^T K\theta = 4\theta_2^2 + 12\theta_2\theta_3 + 12\theta_3^2$. We also verify the result $K = DDMD^TD^T$. Note that $Df(x) = (0,1,2x,3x^3)^T$ and $DDf(x) = (0,0,2,6x)^T$ which are the vectors of first and second derivatives of elements of $f(x)$, respectively.

In general, the absence of coefficients θ_0 and θ_1 in Ψ_2 reflects the fact that the constant and linear term do not contribute to roughness of the model. Consequently, the matrix K has two rows and columns of zeroes.

Furthermore, the integral $\langle\cdot,\cdot\rangle$ of Definition 1 is a linear operator that satisfies symmetry and non-negativity properties. However it is not a full scalar product over the space of polynomials as it is possible that this integral equals zero despite using non-zero polynomial functions, e.g. $\langle 1,x\rangle = 0$. For our purposes, this fact does not mean that we cannot include terms with zero roughness, as the identifiability of such terms is guaranteed by the design matrix X under mild conditions.

2.2 The constrained quadratic optimization

We now establish the SSM problem as a quadratic problem in the parameters [3]. Minimization of Ψ_2 subject to interpolation conditions is equivalent to minimize $\theta^T K\theta$ subject to $X\theta = Y$, where Y is a vector collecting all response values $Y^T = (y_1,\ldots,y_n)$ and thus $X\theta = Y$ represents all interpolation conditions. Using standard constrained optimization techniques, the solution is obtained from the linear system:

$$\begin{pmatrix} 0 & X \\ X^T & K \end{pmatrix} \begin{pmatrix} \lambda \\ \theta \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix}, \tag{4}$$

where λ is a vector of Lagrange multipliers for the interpolation constraints. Existence of the solution of the above system is straightforward when K and X are full rank. For the case when K is not full rank, a condition is imposed on the rank of X. This technical condition is mild and does not restrict the applicability of SSM.

The inverse of the left hand side matrix in (4) can be written in block form as

$$\begin{pmatrix} 0 & X \\ X^T & K \end{pmatrix}^{-1} = \begin{pmatrix} -Q & H^T \\ H & P \end{pmatrix}$$

with matrices Q,H,P that satisfy conditions $XH = I_n$, $XP = 0_{n,N}$ $X^TQ = KH$ and $X^TH^T + KP = I_N$ and I_n, I_N are identity matrices of sizes n and N. The parameter estimates are $\theta^* = HY$, and the minimum value of roughness achieved by the polynomial model is $\Psi_2^* = Y^TQY$, see [1].

2.3 Gram-Schmidt ortogonalization of SSM bases

The linear system at the core of SSM methodology can be improved in several ways depending on the choice of polynomial bases. The objective is to linearly transform the bases to achieve a more stable or sparse version of Equation (4).

A starting point is to use monomials in $f(x)$, as we have done. Monomials have a simple interpretation yet the system in (4) is not sparse. A second option are orthogonal polynomials over $\mathscr{X}$, e.g. if $\mathscr{X} = [-1,1]$ we would use Legendre polynomials. Use of Legendre polynomials $L_i(x)$ as building elements of $f(x)$, so it becomes $f(x)^T = (L_0(x),\ldots,L_{N-1}(x))$, implies a more sparse system, particularly on the matrix K but there is no guarantee that X will be more sparse. In a multivariate setting Legendre polynomials have the aditional advantage of simplifying sensitivity analyses. Another option is to perform Gram-Schmidt orthogonalization over the columns of X which will not necessarily guarantee simplification of the matrix K.

An alternative we pursue here is orthogonalization of the basis, using the inner product $\langle\cdot,\cdot\rangle$ of Definition 1. We give below a Gram-Schmidt orthogonalization algorithm for a set of terms such that the matrix K of inner dot products is full rank. For simplicity we write e.g. f_1 instead of $f_1(x)$.

Input A list of polynomials $g_1,\ldots,g_r$ such that the K matrix of inner products $\langle g_i,g_j\rangle$ is full rank.
Output A list of $\langle\cdot,\cdot\rangle$-orthonormal polynomials $f_1,\ldots,f_r$.
Initialization Set $i:=1$ and standardize $f_1:=g_1/\sqrt{\langle g_1,g_1\rangle}$.
Step 1 Project $h_i:=g_i-\sum_{j=1}^{i-1}\langle g_i,f_j\rangle f_j$ and standardize $f_i:=h_i/\sqrt{\langle h_i,h_i\rangle}$.
Step 2 Update $i:=i+1$. If $i\leq r$ repeat from Step 1.

Example 2. Consider the monomial terms x^2,x^3,x^4,x^5 and the region $\mathscr{X}=[0,1]$. The matrix K is full rank and after applying the above algorithm we retrieve the list of polynomials $\frac{x^2}{2}$, $\frac{1}{\sqrt{3}}(x^3-\frac{3}{2}x^2)$, $\frac{\sqrt{5}}{2}(x^4-2x^3+x^2)$, $\sqrt{7}\left(x^5-\frac{5}{2}x^4+2x^3-\frac{1}{2}x^2\right)$. The K matrix for this new list of polynomial model terms is the identity of size four.

Starting with full rank K, the K matrix for the updated basis is an identity matrix. We can still accomodate terms that have zero inner product as long as they are linearly independent, and the algorithm above works by simply avoiding the standardization step. For the list of terms $1,x,\ldots,x^5$, we would only append the terms $1,x$ to the list of Example 2.

3 Multivariate SSM

For multiple input variables, SSM extend easily by using hierarchical polynomials, i.e. if a term is included in the model, all its divisors are also included. Consider k input variables $x_1,\ldots,x_k$. A monomial is the power product $x^\alpha=x_1^{\alpha_1}\cdots x_k^{\alpha_k}$ where $\alpha_1,\ldots,\alpha_k$ are non negative integer exponents collected in the exponent vector α. The starting point is a hierarchical multivariate polynomial, written as

$$y(x)=\sum_{\alpha\in L}\theta_\alpha x^\alpha,$$

where L is a list of N exponents satisfying the hierarchy restriction and θ_α is the coefficient of monomial term x^α. For example, the bivariate polynomial $\theta_{00}+\theta_{10}x_1+\theta_{01}x_2+\theta_{11}x_1x_2$ has list of exponents $L=\{(0,0),(1,0),(0,1),(1,1)\}$ and is written as $f(x)^T\theta$ with $f(x)^T=(1,x_1,x_2,x_1x_2)$ and $\theta^T=(\theta_{00},\theta_{10},\theta_{01},\theta_{11})$.

Over a closed, bounded region $\mathscr{X}\subset\mathbb{R}^k$, the measure of roughness extends to

$$\Psi_2(y)=\int_{\mathscr{X}}tr(H(y)^2)dx \tag{5}$$

where $H(y)$ is the Hessian matrix of y. The bivariate thin plate splines criterion is precisely an instance of this definition of Ψ_2, and the main difference is that the integration region $\mathscr{X}$ we consider is finite.

This multivariate roughness measure extends the inner product of Definition 1 to $\langle s(x),t(x)\rangle:=\int_{\mathscr{X}}\sum_{i,j}\frac{\partial^2 s(x)}{\partial x_i\partial x_j}\frac{\partial^2 t(x)}{\partial x_i\partial x_j}dx$, where the sum is performed over all pairs i,j taking values from $1,\ldots,k$. Model roughness is the familiar $\Psi_2(y)=\langle y,y\rangle$. The

computation of the K matrix and the Gram-Schmidt algorithm of Section 2.3 follows with little change. We next give a bivariate example.

Example 3. Consider the bivariate region $\mathscr{X} = [0,1]^2$ and the list of terms x_1^2, x_1x_2, x_2^2, $x_1^3, x_1^2x_2$ and $x_1x_2^2$. The K matrix of inner products between elements in the list is

$$K = \begin{pmatrix} 4 & 0 & 0 & 6 & 2 & 0 \\ 0 & 2 & 0 & 0 & 2 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 \\ 6 & 0 & 0 & 12 & 3 & 0 \\ 2 & 2 & 0 & 3 & 4 & 2 \\ 0 & 2 & 2 & 0 & 2 & 4 \end{pmatrix}.$$

The first three terms are already pairwise orthogonal and only need standardisation so we have $x_1^2/2, x_1x_2/\sqrt{2}$, $x_2^2/2$. Completing the algorithm we obtain $\left(x_1^3 - 3x_1^2/2\right)/\sqrt{3}$, $x_1^2x_2 - x_1x_2 - x_1^2/2$ and $x_1x_2^2 - x_2^2/2 - x_1x_2$. The matrix K for this new list of model terms is the identity of size six. Mirroring our previous results, constant and linear terms $1, x_1, x_2$ have all zero roughness and if included would produce three rows and columns of zeroes in K.

The construction of SSM over a complex region $\mathscr{X}$ gravitates around efficient computation of the K matrix. By a multivariate version of Lemma 1, this construction relates to the matrix of moments over $\mathscr{X}$. A simple proposal that has worked well is to divide the complex region $\mathscr{X}$ into non-overlapping boxes so moments are computed with simple formulas for each box and then added. After this initial step we compute the matrix K. We use the Gram-Schmidt algorithm to obtain an orthogonal polynomial basis so that the new K is a diagonal matrix with a few diagonal zeroes (for constant and linear terms) and the rest of diagonal entries take value one. Computation of θ^* and Ψ_2^* uses the same formulæ as the univariate case.

4 Designing for smooth models in complex regions

Recall that the observed roughness of the SSM is $\Psi_2^* = Y^T QY$. A simple proposal to design for smoothness is to minimize a function of the eigenvalues of the matrix Q. Here we use $\phi(Q) = tr(Q)$ which equals the sum of eigenvalues of Q and minimizes the expected roughness Ψ_2 under general conditions, see [1].

Computation of matrix Q and its eigenvalues does not depend on the actual data values, see the development in Section 2.2. At least two design alternatives are possible here. One is to select an initial design with points that minimize $\phi(Q)$. The alternative we pursued is that given an initial design th at is kept fixed, select a number of additional points that minimize $\phi(Q)$.

In the examples below a synthetic function was used. See the left panel of Figure 1 for a depiction of the Ramsay horseshoe test function [9, 13], shown as contours over the non-standard region $\mathscr{X}$. Validation errors were computed to establish an empirical measure of model fit.

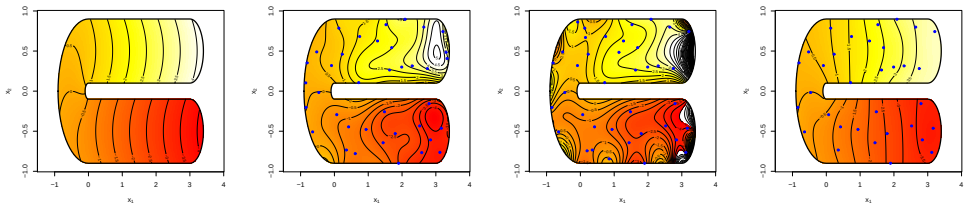

Fig. 1 Original function, sequential fits Ψ_2, J and non-sequential soap film fit (from left).

4.1 Design using roughness Ψ_2

Using the roughness measure Ψ_2 and an initial set of $n = 33$ random points over the region $\mathscr{X}$, a SSM with $N = 70$ terms was fit to the data. To compute the K matrix, the region $\mathscr{X}$ was partitioned in boxes. At this stage, validation error using 30 further random points produced RMSE $= 0.377$. The variance of predicted values was 0.145.

From a a set of 1000 random candidate points, five extra points were selected at random. The value of criterion $\phi(Q)$ was recorded and the procedure repeated 10000 times. The best set of extra points was kept. The augmented design places points in sensible places and with the same 30 extra random validation points produced a better model with RMSE $= 0.322$ and variance of predicted values of 0.106 which shows an improved consistency across the region. Furthermore, the RMSE is 3.22% of the data range, well within an informal rule of thumb of 5%. The centre left panel of Figure 1 gives both initial and second stage designs together with the final fit.

4.2 Design for distortion J

The distortion criterion for soap film smoothing $J(y)$ of Equation (2) can be written as a seminorm, different than that for Ψ_2. We applied SSM techniques using this criterion under similar conditions as the previous study. The initial fit with validation RMSE $= 2.034$ and variance of predictions 3.983. Using a similar procedure as in Section 4.1, five extra design points were generated to produce an updated model with similar validation errors RMSE $= 2.091$ and variance of predictions 4.061. The performance of SSM under distorsion J suffers from the fact that the matrix K for J has very low rank compared with the K matrix for roughness Ψ_2. Indeed for the same data and settings, the rank of K under J is 32 while under Ψ_2 is 67. This phenomenon creates aliasing of terms and very few terms are being used to minimize distortion thus the method is both inefficient and prone to instability.

A non-sequential soap film smoother [13] is clearly superior with RMSE $=$ 0.0868 and variance of predictions 0.0077, see Figure 1. In addition to the inesta-

bility and alisaing described, our SSM-*J* procedure does not address the boundary conditions of soap film so this underperformance is to be expected.

5 Discussion and future work

The use of orthonormal bases enables the SSM-Ψ_2 method to be applied in non-standard regions with reasonable fit. We provided a simple, response independent sequential design for smoothness in such case. Our method is versatile and works well in a variety of situations, being particularly suited to sensitivity analyses [1]. In the soap film example, poor performance of SSM-*J* can be attributed in part to the fact that we did not incorpore boundary conditions. A potential solution to this is to use Hermite interpolators [4], which is possible when the boundary is an algebraic variety.

References

1. Ron A. Bates, Peter R. Curtis, Hugo Maruri-Aguilar, and Henry P. Wynn. Optimal design for smooth supersaturated models. *J. Statist. Plann. Inference*, 154:3–11, 2014.
2. Ron A. Bates, Beatrice Giglio, and Henry P. Wynn. A global selection procedure for polynomial interpolators. *Technometrics*, 45(3):246–255, 2003.
3. Ron A. Bates, Hugo Maruri-Aguilar, and Henry P. Wynn. Smooth supersaturated models. *J. Stat. Comput. Simul.*, 84(11):2453–2464, 2014.
4. T. Becker and V. Weispfenning. The chinese remainder problem, multivariate interpolation, and Gröbner bases. In *ISSAC '91: Proceedings of the 1991 international symposium on Symbolic and algebraic computation*, pages 64–69, 1991.
5. Carl de Boor. *A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*. Springer-Verlag, New York, revised edition, 2001.
6. Holger Dette, Viatcheslav B. Melas, and Andrey Pepelyshev. Optimal design for smoothing splines. *Ann. Inst. Statist. Math.*, 63(5):981–1003, 2011.
7. Jean Duchon. Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *Rev. Francaise Automat. Informat. Recherche Operationnelle Ser. RAIRO Analyse Numerique*, 10(R-3):5–12, 1976.
8. R. L. Harder and R. N. Desmarais. Interpolation using surface splines. *Journal of Aircraft*, 9(2):189–191, 1972.
9. Tim Ramsay. Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(2):307–319, 2002.
10. I. J. Schoenberg. Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.*, 52:947–950, 1964.
11. Grace Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
12. Haonan Wang and M. Giovanna Ranalli. Low-rank smoothing splines on complicated domains. *Biometrics*, 63(1):209–217, 314, 2007.
13. Simon N. Wood, Mark V. Bravington, and Sharon L. Hedley. Soap film smoothing. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70(5):931–955, 2008.
14. David Woods and Susan Lewis. All-bias designs for polynomial spline regression models. *Aust. N. Z. J. Stat.*, 48(1):49–58, 2006.