Spatial Process Estimates as Smoothers: A Review

Soutir Bandyopadhyay

1 Basic Model

The observational model considered here has the form

$$Y_i = f(\mathbf{x}_i) + \epsilon_i, \text{ for } 1 \leq i \leq n. \tag{1.1}$$

where Y_i is the observed response at the ith combination of design variables (or ith location), $\mathbf{x}_i \in \mathcal{R}^d$, and f is the function of interest. The random components, $\{\epsilon_i\}$, usually associated with the measurement errors, are assumed to be uncorrelated, zero mean random variables with variances, $\{\sigma^2/W_i\}$. let us define, $W = \text{diag}(W_1, \cdots, W_n)$.

One parsimonious strategy for representing f is as the sum of a low order polynomial and a smooth function,

$$f(\mathbf{x}) = P(\mathbf{x}) + h(\mathbf{x}) \tag{1.2}$$

- In nonparametric regression and smoothing, the assumptions on h are based on higher order derivatives and in the case of thin-plate splines, the degree of P and the dimension of $\mathbf{x}$ imply a specific roughness penalty on h based on integrated squared (partial) derivatives.

- In spatial statistics models, the smoothness assumption is replaced by the assumption that h is a random field. Under the assumption that f is a realization of a spatial process, P can be identified as the **spatial trend** or **drift**, the fixed part of the model, and h is modeled as a mean zero Gaussian process. Let,

$$Cov(h(\mathbf{x}), h(\mathbf{x}')) = \rho k(\mathbf{x}, \mathbf{x}') \tag{1.3}$$

be a covariance function where k is a known function and ρ is an unknown multiplier. Conditions on the covariance function then control the smoothness properties of h.

2 The unifying theme

In this article we will construct the set of basis functions $\{g_\nu\}$ for $1 \leq \nu \leq n$, with the property that if $u_\nu = \sum_{k=1}^n g_\nu(\mathbf{x}_k) W_k Y_k$, then the function $\sum_{\nu=1}^n u_\nu g_\nu(\mathbf{x})$ will interpolate the data, Y_k. This is possible because of an orthogonality property for the basis. Moreover, the orthogonality will ensure that these interpolation coefficients, u_k, are uncorrelated with each other, simplifying the simulation of these estimators.

In addition to this basis, there is a sequence of increasing nonnegative weights (eigenvalues), $\{D_\nu\}$ so that

$$\hat{f}(\mathbf{x}) = \sum_{\nu=1}^{n} \frac{1}{1 + \lambda D_\nu} u_\nu g_\nu(\mathbf{x}) \tag{2.4}$$

where λ is the usual spline smoothing parameter or in terms of the spatial process estimate this is the "signal to noise" ratio, $\lambda = \sigma^2/\rho$.

Question: Where does this magic orthogonal basis come from?

- For spatial process models the usual kriging surface is a linear combination of low order polynomial functions and n functions, $\psi_j(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}_j)$, obtained by evaluating one argument of the covariance function at the observed locations. A standard matrix decomposition transforms this basis into the orthogonal one. (More later)

- For splines, one starts with polynomials and radial basis functions and then proceeds in the same way. (More later)

Note: In this overview, the case of smoothing has been emphasized and is appropriate when the function is observed with error ($\sigma > 0$). When measurement error is not present, this is the same as setting $\sigma = 0$ and thus $\lambda = 0$ in the tapering formula. The result is a linear combination of the basis functions that interpolates the observed data and can be extrapolated to provide estimates of the surface at unobserved points. Throughout this article the interpolation results can be inferred from the smoothing estimators by considering the limiting case as $\lambda \to 0$.

3 Thin-plate splines

A spline is defined implicitly as the solution to a variational (minimization problem). The form of the minimization criterion determines the kind of spline. A thin-plate spline estimator of f is the minimizer of the penalized sum of squares

$$S_\lambda(f) = n^{-1} \sum_{i=1}^{n} W_i (y_i - f(x_i))^2 + \lambda J_m(f) \tag{3.5}$$

for $\lambda > 0$ and where

$$J_m(f) = \int_{\mathcal{R}^d} \sum \frac{m!}{\alpha_1! \cdots \alpha_d!} \left(\frac{\partial^m f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \right)^2 d\mathbf{x}$$

The sum of the integrand is taken over all non-negative integer vectors, α, such that $\sum(\alpha_1 + \cdots + \alpha_d) = m$, and $2m > d$. For one dimension and $m = 2$, $J_2(f) = \int \{f''(x)\}^2 dx$, giving the standard cubic smoothing spline roughness penalty.

The set of functions where the roughness penalty is zero is termed the *null space*, and for J_m, consists of all polynomials with degree less than or equal to $(m - 1)$.

Bayesian viewpoint

Minimization of S_λ results in a function that tracks the data but is also constrained to be smooth. A Bayesian viewpoint is to interpret $-S_\lambda$, to within a constant, as the log of joint density function for f and Y. Let p be the conditional density of Y given f and π, the prior density for f.

$$-S_\lambda = \ln(p(Y|f)\pi(f)) = \ln(p(Y|f)) + \ln(\pi(f))$$

If one adopts equation 3.5 then the roughness penalty is associated with a log prior distribution for f. Furthermore, this prior can be interpreted as a distribution such that f is a realization from a smooth Gaussian process. From Bayes theorem we know that for fixed Y the posterior is proportional to the joint density. Thus maximizing the joint density or minimizing S_λ results in a mode of the posterior distribution.

3.1 Form of the spline estimate

The most important step in deriving the spline estimate is in identifying the solution as a finite linear combination of basis functions. In this way, the abstract minimization over a function space collapses into minimizing over the coefficients associated with the basis functions.

Thin-plate splines are expressed in terms of radial basis functions.

- Let $\psi_i(\mathbf{x}) = E_{md}(\mathbf{x} - \mathbf{x}_i)$ where E_{md} are the radial basis functions

$$E_{md}(\mathbf{r}) = \begin{cases} a_{md}\|\mathbf{r}\|^{(2m-d)} \log\left(\|\mathbf{r}\|\right) & : \quad d \text{ even} \\ a_{md}\|\mathbf{r}\|^{(2m-d)} & : \quad d \text{ odd} \end{cases}$$

  and a_{md} is a constant.

- Let $\{\phi_j\}$ be a set of t polynomial functions that span the space of all d-dimensional polynomials with degree $(m-1)$ or less.

Now consider a function of the form:

$$f(\mathbf{x}) \;=\; \sum_{j=1}^{t} \phi_j(\mathbf{x})\beta_j + \sum_{i=1}^{n} \psi_i(\mathbf{x})\delta_i$$

clearly parametrized by β and δ.

Three key results for the computational form for the spline:

Define T, a $n \times t$ matrix with $T_{k,j} = \phi_j(\mathbf{x}_k)$. Let,

$$\mathcal{V} \;=\; \{f : J_m(f) < \infty\} \text{ and } \mathcal{M} \;=\; \{f : f = \sum_{j=1}^{t} \phi_j\beta_j + \sum_{i=1}^{n} \psi_i\delta_i \text{ and } T^T\delta = 0\}$$

Then,

- $\mathcal{M} \subset \mathcal{V}$.
- $\min_{f \in \mathcal{V}} S_\lambda(f) = \min_{f \in \mathcal{M}} S_\lambda(f)$.
- If T has full rank then a unique solution exists.

Proof: See Minguet (1979), Duchon (1977), Wahba (1990), Kent and Mardia (1994) for more detail.

$\mathcal{M}$ is spanned by a finite number of basis functions. From the above properties it follows that the solution to the spline minimization problem can be reduced to a optimization problem over a finite dimensional space. With this result it is useful to rephrase the minimization problem in matrix/vector form, substituting the finite dimension form into $S_\lambda(f)$.

Let us define $K_{i,k} = \psi_i(\mathbf{x}_k)$, then if $T^T\delta = 0$, we get,

$$J_m\left(\sum_{j=1}^{t}\phi_j\beta_j + \sum_{i=1}^{n}\psi_i\delta_i\right) = \delta^T K\delta$$

Therefore,

$$S_\lambda\left(\sum_{j=1}^{t}\phi_j\beta_j + \sum_{i=1}^{n}\psi_i\delta_i\right) = (\mathbf{Y} - T\beta - K\delta)^T W (\mathbf{Y} - T\beta - K\delta) + \lambda\delta^T K\delta. \quad (3.6)$$

Now taking partial derivatives with respect to β and δ and noting that K is a symmetric matrix, the solution satisfies the system of equations

$$\begin{aligned} -2T^T W(\mathbf{Y} - T\beta - K\delta) &= 0 \\ -2KW(\mathbf{Y} - T\beta - K\delta) + 2\lambda K\delta &= 0. \end{aligned}$$

If $\{x_i\}$ are unique locations then K will have full rank. Multiply the second equation by K^{-1} and substitute this new equation into the first. This yields the system of equations

$$\begin{aligned} T^T\delta &= 0 \\ T\beta + K\delta + \lambda W^{-1}\delta &= \mathbf{Y} \end{aligned}$$

From the above two equations the solution is given by,

$$
\begin{aligned}
\hat{\beta} &= \{T^T(K+\lambda W^{-1})^{-1}T\}^{-1}T^T(K+\lambda W^{-1})^{-1}\mathbf{Y} \\
\hat{\delta} &= (K+\lambda W^{-1})^{-1}(\mathbf{Y}-T\hat{\beta})
\end{aligned}
$$

Note: $\hat{\beta}$ is a generalized least squares regression estimate if we identify $(K+\lambda W^{-1})^{-1}$ as a covariance matrix for $\mathbf{Y}$. This linear system also gives a simple form for the residuals from the splines fit. The predicted values are just $T\hat{\beta}+K\hat{\delta}$ and the residual vector must be equal to $\lambda W^{-1}\hat{\delta}$.

3.2 The QR parameterization

The previous derivation produces a solution that is readily interpreted but it is not a form that is suitable for computation. Also in terms of making comparison with Kriging estimates, a different parameterization is required.

Let $T = F_1 R$ denote the QR decomposition of T.

- $F = [F_1|F_2]$ is an orthogonal matrix.
- F_1 has columns that span the column space of T, F_2 has columns that span the space orthogonal to T and R is an upper triangular matrix.
- Reparameterize $\delta = F_2\omega_2$ for $\omega_2 \in \mathcal{R}^{n-t}$. This parameterization enforces the necessary number of linear constraints on δ.
- Setting $\beta = \omega_1$ and $\omega^T = (\omega_1^T, \omega_2^T)$, the system of equations reduces to the form:

$$
T\omega_1 + KF_2\omega_2 + \lambda W^{-1}F_2\omega_2 = \mathbf{Y}
$$

- Solving this system of equations we get,

$$
\begin{aligned}
\hat{\omega}_2 &= \left[F_2^T(K+\lambda W^{-1})F_2\right]^{-1}F_2^T\mathbf{Y} \\
\hat{\omega}_1 &= R^{-1}F_1^T\left[\mathbf{Y}-(K+\lambda W^{-1})F_2\hat{\omega}_2\right]
\end{aligned} \tag{3.7}
$$

3.3 The hat matrix, kernels and cross-validation

Examining the equations leading to $\hat{\omega}$, we see that all matrices involved are independent of $\mathbf{Y}$. Thus, for fixed λ and fixed locations, splines are linear functions of the data and there is an $n \times n$ hat matrix, $A(\lambda)$, such that,

$$[\hat{f}(\mathbf{x}_1), \cdots, \hat{f}(\mathbf{x}_n)]^T \quad = \quad A(\lambda)\,[\mathbf{Y}_1, \cdots, \mathbf{Y}_n]^T$$

- The ith row of the A matrix gives the weights that are applied to the data to give the function estimate at $\mathbf{x}_i$. When one plots these weights they look like higher order kernels, although an explicit formula between λ and bandwidth is not obvious. In one dimension the kernel approximation to a smoothing spline has been studied by many authors and it has been seen that for uniformly distributed locations the bandwidth is proportional to $\lambda^{1/2m}$.

- Let $\hat{\mathbf{Y}}_k$ denote the predicted value for the kth data point based on the spline estimate and let $\hat{\mathbf{Y}}_k^\star$ be the prediction for this point based on omitting the kth data point and refitting the spline to the remaining $(n-1)$ data values. Then

  a. *Cross-validation criterion:*

$$\begin{aligned} \text{CV residual } &= \mathbf{Y}_k - \hat{\mathbf{Y}}_k^\star &= \frac{(\mathbf{Y}_k - \hat{\mathbf{Y}}_k)}{1 - A_{kk}(\lambda)} \\ \text{CV criterion } &= CV(\lambda) &= \sum_{k=1}^{n} \frac{W_k(\mathbf{Y}_k - \hat{\mathbf{Y}}_k)^2}{(1 - A_{kk}(\lambda))^2} \end{aligned}$$

  b. *Generalized cross-validation criterion:*

$$\text{GCV criterion } = V(\lambda) = \frac{(1/n)\|W(I - A(\lambda))\mathbf{Y}\|^2}{(1 - tr\,A(\lambda)/n)^2}$$

  The principle is that good choices of λ will give small CV residuals. So one might choose λ by minimizing $V(\lambda)$.

4 Spatial process estimates

The derivation of the spline estimate in the previous section assumed that the unknown function is fixed. In contrast, a spatial process estimate is based on a model for f as a random function. Accordingly, we now assume f has the form from 1.2 and h is a mean zero gaussian process.

We will see that the covariance function for h now plays the same role for spatial process estimates as the radial basis functions do for splines.

4.1 The universal kriging estimator

Given observed data, $\{Y_k\}$, an estimate for $f(x)$ can be found as a linear combination of the observations that minimizes the variance and unbiased.

For clarity, let $\mathbf{x}_0$ be the new point for prediction and so the goal is an optimal estimate for $f(\mathbf{x}_0) = T_0^T\beta + h(\mathbf{x}_0)$. Any linear estimate has the form $\hat{Y}_0 = \mathbf{m}^T\mathbf{Y}$ with $\mathbf{m} \in \mathcal{R}^n$. So our goal is to identify $\mathbf{m}$. Now let F_1 and R be the QR decomposition of the T matrix, described in the previous section. Equivalently, we can express $\mathbf{m}$ as $\mathbf{m} = F\mathbf{a}$ for some $\mathbf{a} \in \mathcal{R}^n$. Thus the estimate is $\hat{Y}_0 = (F\mathbf{a})^T\mathbf{Y}$. Furthermore, $F\mathbf{a}$ can be partitioned as $F_1\mathbf{a}_1 + F_2\mathbf{a}_2$ where we have subdivided the vector $\mathbf{a}^T = (\mathbf{a}_1|\mathbf{a}_2)^T$.

Determining $\mathbf{a}_1$:

From the orthogonality of F we have,

$$E[(F\mathbf{a})^T\mathbf{Y}] \;=\; (F\mathbf{a})^T T\beta \;=\; \mathbf{a}_1^T R\beta.$$

Note that, $E(\hat{Y}_0) = T_0^T\beta$ and so in order for the estimate to be unbiased for all β it must follow that, $\mathbf{a}_1 = (R^T)^{-1}T_0$.

Determining $\mathbf{a}_2$:

Here we minimize the mean squared error of the estimate.

$$E\left[f(x_0) - \hat{Y}_0\right]^2 \;=\; var(f(x_0) - \hat{Y}_0)$$

because the estimate is unbiased.

$$
\begin{aligned}
var(f(x_0) - \hat{Y}_0) &= var(f(x_0) - \mathbf{m}^T\mathbf{Y}) \\
&= var(f(x_0)) - 2\rho\mathbf{m}^T\mathbf{k}_0 + \mathbf{m}^T(\rho K + \sigma^2 W^{-1})\mathbf{m}.
\end{aligned}
$$

where, $\rho K_{ij} = cov(h(\mathbf{x}_i), h(\mathbf{x}_j)) = \rho k(\mathbf{x}_i, \mathbf{x}_j)$ and $\rho k_{0,j} = cov(h(\mathbf{x}_0), h(\mathbf{x}_j)) = \rho k(\mathbf{x}_0, \mathbf{x}_j)$. The covariance matrix of the observations is $(\rho K + \sigma^2 W^{-1})$. We now switch to the notation where $\mathbf{m}$ has been partitioned.

$$
\begin{aligned}
E\left[f(x_0) - \hat{Y}_0\right]^2 &= E\left[Z_1 + Z_2\right]^2 \\
&= E(Z_1^2) + 2E(Z_1 Z_2) + E(Z_2^2) \\
&= var(Z_1) + 2cov(Z_1, Z_2) + var(Z_2).
\end{aligned}
\tag{4.8}
$$

where, $Z_1 = T_0^T\beta - \mathbf{a}_1^T F_1^T\mathbf{Y}$, $Z_2 = h(x_0) - \mathbf{a}_2^T F_2^T\mathbf{Y}$ and using the fact that, $E(Z_1) = E(Z_2) = 0$.

Next we will just justify 4.8 as a quadratic function of $\mathbf{a}_2$, and set the partial derivatives equal to zero. The solution to this linear system defines the minimizer.

Note: It is not necessary to work out the term for $var(Z_1)$ because it does not depend on $\mathbf{a}_2$.

$$
\begin{aligned}
cov(Z_1, Z_2) &= -\mathbf{a}_1^T F_1^T\rho\mathbf{k}_0 + \mathbf{a}_1^T F_1^T(\rho K + \sigma^2 W^{-1})F_2\mathbf{a}_2 \\
var(Z_2) &= var(h(x_0)) - 2\mathbf{a}_2^T F_2^T\rho\mathbf{k}_0 + \mathbf{a}_2^T F_2^T(\rho K + \sigma^2 W^{-1})F_2\mathbf{a}_2.
\end{aligned}
$$

Taking partial derivatives yields,

$$
\frac{\partial}{\partial\mathbf{a}_2}E\left[Z_1 + Z_2\right]^2 = 2F_2^T(\rho K + \sigma^2 W^{-1})F_1\mathbf{a}_1 - 2F_2^T\rho\mathbf{k}_0 + 2F_2^T(\rho K + \sigma^2 W^{-1})F_2\mathbf{a}_2.
$$

Now divide the equation by ρ and set $\lambda = \sigma^2/\rho$. Solving for the parameters when the partials are zero yields the solution

$$
\mathbf{a}_2 = \left(F_2^T(K + \lambda W^{-1})F_2\right)^{-1}\left(F_2^T\rho\mathbf{k}_0 - F_2^T(K + \lambda W^{-1})F_1\mathbf{a}_1\right)
$$

Remark:

Considering the equations for the two components of $\mathbf{a}$ it follows that,

$$\hat{f}(\mathbf{x}_0) \;=\; \mathbf{m}^T\mathbf{Y} \;=\; T_0^T\hat{\beta} + \hat{\mathbf{k}}_0^T\hat{\delta} \;=\; \sum_{j=1}^{t}\phi_j(\mathbf{x}_0)\hat{\beta}_j + \sum_{i=1}^{n}\psi_i(\mathbf{x}_0)\hat{\delta}_i.$$

where the basis functions, $\{\psi_i\}$ are defined in terms of the covariance function: $\psi_i = cov(h(\mathbf{x}), h(\mathbf{x}_i))$.

4.2 Kriging estimates are a type of spline

Because of similar notation, the reader may have already anticipated another key result for this estimate. Symbolically the form of the spatial process estimator is identical to the spline estimator presented in the previous section. The kriging estimator can be characterized as the solution to the minimization problem at 3.6 provided that the right identifications are made for the basis functions and the K matrix. Because of this equivalence the cross-validation formulas discussed earlier are also valid for the kriging estimate.

4.3 Thin-plate splines as spatial process estimators

If kriging estimates also solve a minimization problem, going in the other direction, perhaps a thin-plate spline is just a spatial process estimate(?). The matrix K derived from radial basis function is not itself a covariance matrix. In fact its diagonal elements are zero! However, following the discussion in Wahba (1990) one can construct a covariance function of the form

$$k(\mathbf{x}, \mathbf{x}^{'}) \;=\; E_m(\|\mathbf{x} - \mathbf{x}^{'}\|) + Q(\mathbf{x}, \mathbf{x}^{'})$$

The modification, Q is symmetric and involves combinations of the radial basis functions and polynomials up to degree $(m-1)$. It also has the following property: if $\mathbf{u}$ is any vector orthogonal to the columns of T, then $\sum_{j=1}^{n} Q(\mathbf{x}, \mathbf{x}_j)\mathbf{u}_j$ is a polynomial of degree $(m-1)$ or less in $\mathbf{x}$.

Based on this covariance there is of course a corresponding kriging estimate. Furthermore one can show that this estimate reduces to the usual thin-plate spline (Wahba 1990).

After going to the trouble to construct a valid covariance function from the radial basis function kernel, $E_m(\|\mathbf{x} - \mathbf{x}'\|)$, we see that the modification, (i.e., Q), has no impact on the estimator. One can use the radial basis functions alone, pretending that it is a valid covariance, and one will recover the same estimate from the kriging formulas. In summary, a spline can be identified as a spatial process estimate with a fairly peculiar covariance function.

5 Ridge regression estimates and shrinkage

Both the thin-plate spline and the spatial process estimators have a common form as ridge regressions. This general form is at the heart of the numerical procedures for both methods and leads to the construction of the orthogonal series and efficient formulas for determining the smoothing parameter, λ.

Assume the model

$$\mathbf{Y} = X\omega + \mathbf{e}$$

and let H be a nonnegative definite penalty matrix. The general form of the penalized (weighted) least squares problem is minimization of

$$(\mathbf{Y} - X\omega)^T W (\mathbf{Y} - X\omega) + \lambda\, \omega^T H \omega$$

over $\omega \in \mathcal{R}^n$. The solution is,

$$\omega = \left(X^T W X + \lambda H\right)^{-1} X^T W \mathbf{Y}.$$

5.1 A natural basis

Given the general framework, the specific applications to thin-plate splines and spatial process estimates involve identifying the relevant parameterization to find the X matrix and the penalty matrix, H. In both cases the constraint on the parameters, $T^T\delta = 0$, provides a route to reduce the $(n+t)$ basis functions to just n. Denote this basis as $\{\eta_i\}$ where the first t functions are equal to the polynomial terms, $\{\phi_i\}$ and the remaining $(n-t)$ terms are linear combinations of $\{\psi_j\}$ based on the columns of F_2. Explicitly we have $\eta_{i+t} = \sum_{k=1}^{n} F_{2k,i}\psi_i$ for these last members of the basis. We will refer to this as the **natural basis** because it automatically builds in the constraint necessary for a solution. To the general reader there is really nothing "natural" about this but it takes its name from the properties in the one-dimensional case. In the case of one-dimensional, mth order spline this constraint results in the higher derivatives $j = m, m+1, \cdots, 2m-1$ of the spline being zero at the boundaries.

Given a natural basis, the representation of the estimate is $f(\mathbf{x}) = \sum_{i=1}^{n} \omega_i \eta_i(\mathbf{x})$. Thus the relevant regression matrix is $X_{ij} = \eta_j(\mathbf{x}_i)$, and the associated penalty matrix is,

$$H = \begin{pmatrix} 0 & 0 \\ 0 & F_2^T K F_2 \end{pmatrix}.$$

5.2 Demmler-Reinsh basis

Here we will consider a linear transformation of the natural basis to produce an orthogonal one. This is accomplished by finding a matrix, G, that will diagonalize both X^TWX and H. Let B denote the inverse square root of X^TWX and let UDU^T be the singular value decomposition of BHB^T. Now set $G = U^TB$ and it is straightforward to verify that $G^T(X^TWX)G = I$ and $G^THG = D$ (use simultaneous diagonalization, see, Melzer, 2004), where D is diagonal. Finally we define the new orthogonal basis as

$$g_\nu(\mathbf{x}) = \sum_{i=1}^{n} G_{i\nu}\eta_i(\mathbf{x})$$

This basis, known as the **Demmler-Reinsch (DR) basis** in the context of splines. Assume that $f(\mathbf{x}) = \sum_{\nu=1}^{n} \alpha_\nu g_\nu(\mathbf{x})$. We have the following properties for the basis.

1. $\{g_\nu\}$ spans the same subspace of functions as the natural basis, $\{\eta_\nu\}$.
2. $$\sum_{i=1}^{n} g_\nu(\mathbf{x}_i) W_i g_\mu(\mathbf{x}_i) = \begin{cases} 0 & : \quad \mu \neq \nu \\ 1 & : \quad \mu = \nu \end{cases}$$

   and so

   $$\sum_{i=1}^{n} f(\mathbf{x}_i)^2 W_i \;=\; \sum_{\nu=1}^{n} \alpha_\nu^2$$
3. Setting $\omega = G\alpha$

   $$f(\mathbf{x}) \;=\; \sum_{\nu=1}^{n} \omega_\nu \eta_\nu(\mathbf{x})$$

   and

   $$\omega^T H \omega \;=\; \alpha^T D \alpha \;=\; \sum_{\nu=1}^{n} D_\nu \alpha_\nu^2.$$

These properties can be easily proved from the construction of the transformation matrix, G. An important consequence from the second property is a simple representation for the interpolating function. Given the sequences $(\mathbf{x}_i, Y_i)$, for $1 \leq i \leq n$, let $u_\nu = \sum_{i=1}^{n} g_\nu(\mathbf{x}_i) W_i Y_i$. The function $\sum_{\nu=1}^{n} u_\nu g_\nu(\mathbf{x})$ will interpolate these data.

5.3 Simplifications due to the Demmler-Reinsch basis

Using the parameterization with respect to the Demmler-Reinsch basis, assume that $f(\mathbf{x}) = \sum_{\nu=1}^{n} \alpha_\nu g_\nu(\mathbf{x})$ and let $\mathbf{u}$ be the coefficients that interpolate the data pairs $(\mathbf{x}_i, Y_i)$.

$$S_\lambda(f) \;=\; \sum_{\nu=1}^{n} (u_\nu - \alpha_\nu)^2 + \lambda \sum_{\nu=1}^{n} D_\nu \alpha_\nu^2$$

Note that now the parameters are decoupled and each term can just be minimized independently to give the solution

$$\hat{\alpha}_\nu \;=\; \frac{u_\nu}{1 + \lambda D_\nu}$$

leading to the function estimate presented in the introduction.

The DR basis also yields simple expressions for the residual sums of squares and the trace of the hat matrix. the ith residual is

$$Y_i - \hat{f}(\mathbf{x}_i) = \sum_{\nu=1}^{n} \left(u_\nu - \frac{u_\nu}{1+\lambda D_\nu} \right) g_\nu(\mathbf{x}_i) = \sum_{\nu=1}^{n} \frac{\lambda D_\nu u_\nu g_\nu(\mathbf{x}_i)}{1+\lambda D_\nu}$$

Using the orthogonal properties it follows that

$$\sum_{i=1}^{n} W_i (Y_i - \hat{f}(\mathbf{x}_i))^2 = \sum_{\nu=1}^{n} \left(\frac{\lambda D_\nu u_\nu}{1+\lambda D_\nu} \right)^2$$

The hat matrix has elements

$$A(\lambda)_{ij} = \sum_{\nu=1}^{n} \frac{g_\nu(\mathbf{x}_i) g_\nu(\mathbf{x}_j)}{1+\lambda D_\nu} W_j$$

Again, using the orthogonality property it follows that the trace of this matrix is

$$trace\,(A(\lambda)) = \sum_{\nu=1}^{n} \frac{1}{1+\lambda D_\nu}.$$

In each case, once the DR basis has been computed, the residual sum of squares and the trace can be evaluated rapidly in $O(n)$ operations.

References

[1] Duchon, J. (1977). "Splines minimizing rotation-invariant semi-norms in Sobolev spaces", *Constructive Theory of Several Variables*, Berlin, 85-100.

[2] Kent, J. T., Mardia, K. V. (1994). "The link between Kriging and Thin-plate splines", *Probability, Statistics and Optimization*, ed: F. P. Kelly, Wiley, New York, 324-339.

[3] Melzer, T. (2004)"SVD and its application to generalized eigenvalue problems"

[4] Minguet, J. (1979). "Multivariate interpolation at arbitrary points made simple", *J. Appl. Math. Phys.*, (ZAMP), 30, 292-304.

[5] Wahba, G.(1990). "Spline models for observational data", *Society for Industrial and Applied Mathematics*, Philadelphia.