Thin-Plate Splines

David Eberly
Geometric Tools, LLC
http://www.geometrictools.com/

Created: March 1, 1996
Last Modified: May 22, 2011

Contents

1 Introduction

Recall that natural cubic splines are piecewise cubic polynomial and exact interpolating functions for tabulated data $(x_i, f(x_i))$. The globally constructed spline has continuous second-order derivatives. The second derivatives at the endpoints are zero (no bending at end points). It is also possible to clamp the endpoints by specifying zero first derivatives there. The spline curve represents a thin metal rod that is constrained not to move at the grid points.

The concept applies equally as well in two dimensions. A *thin-plate spline* is a physically motivated 2D interpolation scheme for arbitrarily spaced tabulated data $(x_i, y_i, f(x_i, y_i))$. These splines are the generalization of the natural cubic splines in 1D. The spline surface represents a thin metal sheet that is constrained not to move at the grid points. The construction is based on choosing a function that minimizes an integral that represents the bending energy of a surface. The origins of thin-plate splines in 2D appears to be [1, 2].

In fact, the concept applies to any dimension for arbitrarily spaced tabulated data $(\mathbf{x}_i, f(\mathbf{x}_i))$. The method of construction for all dimensions is presented in [3] and is based on functional analysis[1].

In n dimensions, the idea of thin-plate splines is to choose a function $f(\mathbf{x})$ that exactly interpolates the data points $(\mathbf{x}_i, y_i)$, say, $y_i = f(\mathbf{x}_i)$, and that minimizes the bending energy,

$$E[f] = \int_{\mathbb{R}^n} |D^2 f|^2 \, dX \tag{1}$$

where $D^2 f$ is the matrix of second-order partial derivatives of f and $|D^2 f|^2$ is the sum of squares of the matrix entries. The infinitesimal element of hypervolume is $dX = dx_1 \cdots dx_n$, where x_i are the components of $\mathbf{x}$.

It is also possible to formulate the problem with a smoothing parameter for regularization [4]. A function f is chosen that does not necessarily exactly interpolate all the data points but that does minimize

$$E[f] = \sum_{i=1}^{m} |f(\mathbf{x}_i) - y_i|^2 + \lambda \int_{\mathbb{R}^n} |D^2 f|^2 \, dX \tag{2}$$

The smoothing parameter is $\lambda > 0$ and is chosen *a priori*. The summation makes it clear that there are m data points.

2 The Calculus of Variations

The ideas are presented in a mathematically informal manner.

2.1 Functionals of f and f'

To motivate the minimization, consider a functional that is an integral involving a function F that depends on independent variable x, on a function f, and on the derivative function f',

$$E[f] = \int_a^b F(x, f, f') \, dx \tag{3}$$

[1] Given the promise of the title, I purchased the paper from SpringerLink for 34 USD and was disappointed. It is simple only to other research mathematicians with extensive training in functional analysis.

For example, $F(x, f, f') = f$, in which case $E[f] = \int_a^b f(x)\,dx$ is just the definite integral of f for the interval $[a, b]$. Another example is $F(x, f, f') = \sqrt{1 + (f')^2}$, in which case $E[f] = \int_a^b \sqrt{1 + (f'(x))^2}\,dx$ is the arclength of the graph of f for the interval.

We wish to construct f for which $E[f]$ of Equation (3) is a minimum. The calculus of variations allows us to do this, a process that is the extension of directional derivatives for multivariate functions to directional derivatives of functions whose independent inputs are themselves functions. We may example E as f varies in the direction of another function g,

$$\phi(t) = E[f + tg] = \int_a^b F(x, f + tg, f' + tg')\,dx \tag{4}$$

We assume that g does not change f at the interval endpoints, so $g(a) = 0$ and $g(b) = 0$. For each scalar t we obtain the real number $\phi(t)$ from the integration. If f is a function that minimizes E, then we expect $\phi(t) = E[f + tg] \geq E[f] = \phi(0)$ for t near zero. From standard calculus, for $\phi(0)$ to be a minimum, we expect that its derivative is zero: $\phi'(0) = 0$. If we formally differentiate Equation (4), we obtain

$$\phi'(t) = \int_a^b \frac{\partial F(x, f + tg, f' + tg')}{\partial f} g + \frac{\partial F(x, f + tg, f' + tg')}{\partial f'} g'\,dx \tag{5}$$

The integrand is an application of the chain rule to differentiate $F(x, f + tg, f + tg')$. Setting t to zero, we have at a minimum,

$$0 = \phi'(0) = \int_a^b \frac{\partial F(x, f, f')}{\partial f} g + \frac{\partial F(x, f, f')}{\partial f'} g'\,dx \tag{6}$$

The second term in the integrand involves $g'(x)$. We can use integration by parts, $\int u\,dv = uv - \int v\,du$, with $u = \partial F/\partial f'$ and $dv = g' dx$,

$$\int_a^b \frac{\partial F}{\partial f'} g'\,dx = \left.\frac{\partial F}{\partial f'} g\right|_a^b - \int_a^b \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) g\,dx = -\int_a^b \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) g\,dx \tag{7}$$

where the last equality follows from $g(a) = g(b) = 0$. Combining this with Equation (6), we have

$$0 = \int_a^b \left[\frac{\partial F}{\partial f} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right)\right] g\,dx \tag{8}$$

This equation is true no matter which "direction" g we choose. The only way for this to happen is if

$$\frac{\partial F}{\partial f} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) = 0 \tag{9}$$

This is referred to as the *Euler-Lagrange differential equation*.

As an example, let us construct the function $f(x)$ for which the arclength integral is a minimum on the interval $[x_0, x_1]$. Let the function values at the endpoints be y_0 and y_1. The integrand is $F(x, f, f') = \sqrt{1 + (f')^2}$. The Euler-Lagrange differential equation is

$$0 = \frac{\partial F}{\partial f} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) = 0 - \frac{d}{dx}\left(\frac{f'}{[1 + (f')^2]^{1/2}}\right) = \frac{-f''}{[1 + (f')^2]^{3/2}} \tag{10}$$

The equation is satisfied when $f''(x) = 0$ for all x, which means $f(x) = y_0 + (y_1 - y_0)(x - x_0)/(x_1 - x_0)$. This is exactly what we expect—the shortest-length curve connecting two points is a line segment.

2.2 Functionals of f, f', and f''

The same idea of a directional derivative applies when the integrand depends on the function and its first- and second-order derivatives, $F(x, f, f', f'')$. When computing the directional derivative, we use a function $g(x)$ for which $g(a) = g(b) = 0$ and $g'(a) = g'(b) = 0$. The equivalent of Equation (6) is

$$0 = \int_a^b \frac{\partial F}{\partial f} g + \frac{\partial F}{\partial f'} g' + \frac{\partial F}{\partial f''} g''\, dx \tag{11}$$

The second term in the integrand is integrated by parts once. The third term is integrated by parts twice, and uses $g(a) = g(b) = g'(a) = g'(b) = 0$ to eliminate the nonintegral terms that occur. The result is

$$0 = \int_a^b \left[\frac{\partial F}{\partial f} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial f''}\right) \right] g\, dx \tag{12}$$

Once again, this equation is true no matter the choice of g, and the only way that can happen is if

$$\frac{\partial F}{\partial f} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial f''}\right) = 0 \tag{13}$$

The introduction of f'' allows us to handle the bending energy integral.

2.3 Cubic Splines and Green's Functions

The data points are (x_i, y_i) for $1 \le i \le m$. We require that $f(x_i) = y_i$ for all i. The bending energy is

$$E[f] = \int_{-\infty}^{\infty} [f''(x)]^2\, dx \tag{14}$$

A complicating factor is that the integral is over the entire real line, so the calculus of variations argument must be extended to handle this. Effectively, we need to work with *distributions*. In this case, think of this as introducing *Dirac delta functions* into the problem. Recall that the Dirac delta function has the substitution property $\phi(a) = \int_{-\infty}^{\infty} \phi(x)\delta(x - a)\, dx$.

In the notation for the calculus of variations, the integrand is $F(x, f, f', f'') = (f'')^2$; that is, F depends only on the second derivative of f. Equation (9) must be satisfied, $f^{(4)}(x) = 0$, the fourth-order derivative of f. If it were the case that f has a continuous fourth-order derivative, then f would have to be a cubic polynomial. However, it is then not possible to satisfy all the conditions $f(x_i) = y_i$ (unless the data points do all lie on the same cubic graph). This requires us to treat $f^{(4)}(x) = 0$ in a distributional sense—the fourth derivative is zero for all x except at the points x_i where the fourth derivative is discontinuous.

We can construct a *Green's function* $G(x, s)$ that is the solution to $\partial^4 G/\partial x^4 = \delta(x - s)$, where $\delta(x)$ is the Dirac delta function. The classical solution is

$$G(x, s) = |x - s|^3 \tag{15}$$

although technically there is a normalizing factor $1/12$. Observe that it has a derivative discontinuity at $x = s$. The function f that minimizes Equation (14) is a linear combination of the $G(x, s)$ with the s-values set to the x_i where the derivative discontinuities must occur. Also notice that any linear polynomial is in the kernel of $E[f]$ (the set of functions for which $E[f] = 0$), so we need to account for this. The form of f is

$$f(x) = \sum_{i=1}^{m} a_i G(x, x_i) + b_0 + b_1 x = \sum_{i=1}^{m} a_i |x - x_i|^3 + b_0 + b_1 x \tag{16}$$

This equation has $m + 2$ unknown values, the a_i and b_j, but we have only m constraints $f(x_i) = y_i$. The remaining two come from an orthogonality condition that is mentioned in [3]. Specifically, the linear polynomial $b_0 + b_1 x$ is in the orthogonal complement of the function space that contains the Green's functions. This manifests itself as $\sum_{i=1}^{m} a_i = 0$ and $\sum_{i=1}^{m} a_i x_i = 0$.

2.4 Euler-Lagrange Equations for Multivariate f

Consider functions of the form $f(x_1, \ldots, x_n)$. Then our F function is of the form $F(x_1, \ldots, x_n, f, f_{x_1}, \ldots, f_{x_n})$ where $f_{x_i} = \partial f / \partial x_i$. The equivalent of Equation (9) is

$$\frac{\partial F}{\partial f} - \sum_{i=1}^{n} \frac{d}{dx_i} \left(\frac{\partial F}{\partial f_{x_i}} \right) = 0 \tag{17}$$

Second-order derivatives $f_{x_i x_j} = \partial^2 f / \partial x_i \partial x_j$ may also be included in F. The equivalent of Equation (13) is

$$\frac{\partial F}{\partial f} - \sum_{i=1}^{n} \frac{d}{dx_i} \left(\frac{\partial F}{\partial f_{x_i}} \right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{d}{dx_i} \frac{d}{dx_j} \left(\frac{\partial F}{\partial f_{x_i x_j}} \right) = 0 \tag{18}$$

3 Thin-Plate Splines in n Dimensions

For the bending energy integrand $F = |D^2 f|^2$, the sum of squares of second-order partial derivatives of f, Equation (18) is the *biharmonic equation*

$$0 = \Delta^2 f = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{x_i x_j}^2 \tag{19}$$

where Δ is the Laplacian operator, which is applied twice.

Just as for the cubic spline, f need only have fourth-order partial derivatives that are continuous almost everywhere (except at the data points). We need a Green's function $G(\mathbf{x}, \mathbf{s})$ that is a solution to the biharmonic equation. The classical solution is the following, written as a radially symmetric function with $r = |\mathbf{x} - \mathbf{x}|$,

$$G(r) = \begin{cases} c_0 r^{4-n} \ln r, & n = 2 \text{ or } n = 4 \\ c_1 r^{4-n}, & \text{otherwise} \end{cases} \tag{20}$$

where

$$c_0 = \frac{(-1)^{n/2+1}}{8\sqrt{\pi}(2 - n/2)!}, \quad c_1 = \frac{\Gamma(n/2 - 2)}{16\pi^{n/2}} \tag{21}$$

As in the cubic-spline case, the constants c_0 and c_1 are omitted and absorbed into the computations of the unknown coefficients of the thin-plate spline.

The minimizer f is a linear combination of the Green's function with the argument $\mathbf{s}$ set to the $\mathbf{x}_i$ of the data points. There is also a linear polynomial term (this term is in the kernel of $E[f]$). The function is

$$f(\mathbf{x}) = \sum_{i=1}^{m} a_i G(\mathbf{x}, \mathbf{x}_i) + \left(b_0 + \sum_{j=1}^{n} b_j x_j \right) \tag{22}$$

where x_j is the jth component of variable $\mathbf{x}$.

Define $\mathbf{y}$ to be the $m \times 1$ vector whose components are the data point y_i values. Define $\mathbf{a}$ to be the $m \times 1$ vector whose components are the coefficients a_i. Define $\mathbf{b}$ to be the $(n+1) \times 1$ vector whose components are the b_j. The constraints $y_i = f(\mathbf{x}_i)$ lead to the system of equations

$$\mathbf{y} = M\mathbf{a} + N\mathbf{b} \tag{23}$$

where M is the $m \times m$ matrix whose entries are $M_{ij} = G(\mathbf{x}_i, \mathbf{x}_j)$ and where N is the $m \times (n+1)$ matrix whose rows are $(1, \mathbf{x}_i)$. An orthogonality condition that comes from the functional analysis in [3] is $N^{\mathrm{T}}\mathbf{a} = \mathbf{0}$. The equations have solution

$$\mathbf{a} = M^{-1}(\mathbf{y} - N\mathbf{b}), \quad \mathbf{b} = (N^{\mathrm{T}} M^{-1} N)^{-1} N^{\mathrm{T}} M^{-1} \mathbf{y} \tag{24}$$

Of course, $\mathbf{b}$ is computed first. The minimum bending energy is $\mathbf{a}^{\mathrm{T}} M \mathbf{a}$. When $\mathbf{a}$ is zero, this quadratic form is zero—this is the case when f is a linear function whose graph is a hyperplane (no bending of the surface).

4 Smoothed Thin-Plate Splines

The smoothed functional is mentioned in Equation (2), which mahy be rewritten as

$$E[f] = \int_{\mathbb{R}^n} \left(\sum_{i=1}^{m} |f(\mathbf{x}) - y_i|^2 \delta(\mathbf{x} - \mathbf{x}_i) + \lambda |D^2 f|^2 \right) dX \tag{25}$$

where $\delta(\mathbf{x})$ is the Dirac delta function of a multivariate input.

The Euler-Lagrange differential equation for the integrand is

$$0 = \sum_{i=1}^{m} (f(\mathbf{x}) - y_i) \delta(\mathbf{x} - \mathbf{x}_i) + \lambda \Delta^2 f \tag{26}$$

where Δ^2 is the biharmonic operator. Using the Green's functions mentioned previously, the solution to the differential equation is of the form

$$f(\mathbf{x}) = \sum_{i=1}^{m} \frac{y_i - f(\mathbf{x}_i)}{\lambda} G(\mathbf{x}, \mathbf{x}_i) + \left(b_0 + \sum_{j=1}^{n} b_j x_j \right) = \sum_{i=1}^{m} w_i G(\mathbf{x}, \mathbf{x}_i) + \left(b_0 + \sum_{j=1}^{n} b_j x_j \right) \tag{27}$$

where x_j is the jth component of the variable $\mathbf{x}$ and where $w_k = (y_k - f(\mathbf{x}_k))/\lambda$, in which case $y_k = f(\mathbf{x}_k) + \lambda w_k$. Evaluate Equation (27) at $\mathbf{x}_k$ to obtain

$$y_k = f(\mathbf{x}_k) + \lambda w_k = \sum_{i=1}^{m} w_i G(\mathbf{x}_k, \mathbf{x}_i) + \left(b_0 + \sum_{j=1}^{n} b_j x_j^{(k)} \right) + \lambda w_k \tag{28}$$

where $x_j^{(k)}$ is the jth component of $\mathbf{x}_k$. Writing this in vector and matrix form, we have the matrix system

$$\mathbf{y} = (M + \lambda I)\mathbf{w} + N\mathbf{b}, \quad N^{\mathrm{T}}\mathbf{w} = \mathbf{0} \tag{29}$$

where I is the $m \times m$ identity matrix. The second equation is the same orthogonality condition mentioned previously. The solution is the same as in Equation (24) except that M is replaced by $M + \lambda I$ and $\mathbf{w}$ is the variable name rather than $\mathbf{a}$. The minimum of the functional is $\lambda \mathbf{w}^{\mathrm{T}}(M + \lambda I)\mathbf{w}$. As λ increases, the value is asymptotic to the discrete summation (first term) of the functional.

5 Source Code

The files `Wm5IntpThinPlateSpline2.*` and `Wm5IntpThinPlateSpline3.*` contains implementations for 2D and 3D. A sample application that illustrates use of the code is

```
WildMagic5/SampleMathematics/ThinPlateSplines
```

In particular, the 2D sample shows how increasing λ causes the functional value to become asymptotic to the discrete summation of the functional. In the example, it is simple to show that the discrete summation is $16/3 = 5.\bar{3}$. The functional value for $\lambda = 10000$ is nearly equal to the summation.

References

[1] J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, RAIRO Analyse Numérique, vol. 10, pp. 5-12, 1976.

[2] J. Duchon, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, Lecture Notes in Mathematics, vol. 57, pp. 85-100, 1977.

[3] Jean Meinguet, *Multivariate interpolation at arbitrary points made simple*, Journal of Applied Mathematics and Physics (ZAMP), vol. 30, pp. 292-304, 1979.

[4] G. Wahba, *Spline models for observational data*, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), 180 pages, 1990.