Discrete Explanation of RBF Kernels

j.p.lewis
CGIT/IMSC/USC
june 03 (bugfix apr04)

Scattered two-dimensional data can be smoothly interpolated using the thin-plate spline, which minimizes the squared second-derivative (approximate curvature) over the surface. The computation can be done by directly solving the PDE $Lu = r$ with relaxation or multigrid. It can also be done (if there are not too many points to interpolate) by a weighted sum of radial basis functions centered at each data point:

$$f(x, y) = \sum w_k f(\sqrt{x^2 + y^2})$$

Radial basis functions $r, r^3, r^2 \log r$, etc. have all been used. Where do these functions come from? The answer requires variational calculus, differential equations, Green's functions, and preferably Fourier transforms. I'll outline this below, but also give a nearly parallel discrete formulation that needs only linear algebra and (unfortunately) Fourier transforms – this second formulation should be more comfortable for computer graphics people.

' = transpose, wrt = with respect to, rhs = right hand side

Continuous

Fit an unknown function f to the data y_k, regularized by minimizing a smoothness term.

$$E[f] = \sum (f_k - y_k)^2 + \lambda \int ||Pf||^2$$

$$\text{e.g.} \quad ||Pf||^2 = \int \left(\frac{d^2 f}{dx^2} \right)^2 dx$$

Discrete

A similar discrete version.

$$E[f] = (f - y)' S' S (f - y) + \lambda f' P' P f$$

- To simplify things, here the data points to interpolate are required to be at discrete sample locations in the vector y, so the length of this vector defines a "sample rate" (reasonable).
- S is a "selection matrix" with 1s and 0s on the diagonal (zeros elsewhere). It has 1s corresponding to the locations of data in y. y can be zero (or any other value) where there is no data.
- P is a diagonal-constant matrix that encodes the discrete form of the regularization operator. E.g. to minimize the integrated curvature, rows of P will contain:

$$\begin{bmatrix} -2, 1, 0, 0, \ldots \\ 1, -2, 1, 0, \ldots \\ 0, 1, -2, 1, \ldots \end{bmatrix}$$

Continuous

The variational derivative of E wrt f leads to a differential equation

$$P'Pf(x) = \frac{1}{\lambda} \sum (f(x) - y_k) \delta(x - x_k)$$

Here P' is the "adjoint" of P. If P is symmetric (true for the second derivative operator) then the adjoint is the same, otherwise it is time-reversed.

Discrete

Take the derivative of E with respect to the vector f, obtaining

$$2S(f - y) + \lambda 2P'Pf = 0$$

The differential equation can be solved by finding the Green's function of the differential operator and convolving with the right hand side (r.h.s) (works only for a linear operator).
Schematically,

$$
\begin{aligned}
Lf &= rhs && \text{L is the operator P'P,} \\
& && \text{rhs is the data fidelity} \\
f &= g \star rhs && \text{f obtained by convolving g} \star \text{rhs} \\
Lg &= \delta && \text{choosing rhs} = \delta \text{ gives this eqn}
\end{aligned}
$$

g is the "convolutional inverse" of L. This is easy to solve in the Fourier domain, where convolution becomes multplication. The transform of δ is a constant, so in the Fourier domain g is the reciprocal of $L = P'P$.
In summary, the kernel g is the inverse Fourier transform of the reciprocal of the Fourier transform of the "adjoint-squared" smoothing operator P.
The Fourier transform of the derivative of a function is

$$F[\frac{d}{dx}f(x)] = |\omega|F[f(x)]$$

i.e. it has linearly more energy at higher frequencies; the Fourier transform of the second derivative should go up with the square of the distance from the frequency origin, etc., in particular the operator $P'P$ boosts the transform proportionally to ω^4. And the inverse Fourier transform of $1/\omega^4$ is $|x|^3$ ignoring scale. This is the 1-D kernel corresponding to the cubic spline.
(In more detail, the integrals of ω^4 and $|x|^3$ diverge, so instead find the transform of them windowed by $\exp(-a|x|)$ and then take the limit as $a \to 0$:

$$
\begin{aligned}
& \int e^{-a|x|}|x|^3 e^{-i2\pi\omega x}dx && \text{transform of } e^{-a|x|}|x|^3 \\
&= \int_0^\infty e^{-ax}x^3 e^{-i2\pi\omega x}dx \\
&+ \int_{-\infty}^0 e^{ax}(-x)^3 e^{-i2\pi\omega x}dx \\
&= \frac{6}{(a+2i\pi\omega)^4} + \frac{6}{(a+2i\pi\omega)^4} \\
& \lim_{a->0} \frac{12}{(a+2i\pi\omega)^4} && \text{now take limit as a} \to 0 \\
&= \frac{3}{4\pi^4\omega^4} && \text{i.e.} \sim 1/\omega^4
\end{aligned}
$$

)

$$P'Pf = \frac{1}{\lambda}S(f-y)$$

Multiply by G, being the inverse of $P'P$:

$$f = GP'Pf = \frac{1}{\lambda}GS(f-y)$$

So the RBF kernel related to $G = (P'P)^{-1}$.
As usual, taking the svd helps.

$$P = UDV' \Rightarrow P'P = VD^2V'$$

The inverse of VD^2V' is $VD^{-2}V'$.
Next, the eigenvectors of a circulant matrix are sinusoids, and P is diagonal-constant (toeplitz?), or nearly circulant. So $VD^{-2}V'$ is approximately the same as taking the Fourier transform and then the reciprocal, remembering that D are the singular values of P, which have to be squared to get the eigenvalues of $P'P$ ($\approx$ Fourier coefficients).