Gaussian Process Implicit Surfaces

Oliver Williams Andrew Fitzgibbon
Microsoft Research, Cambridge, UK

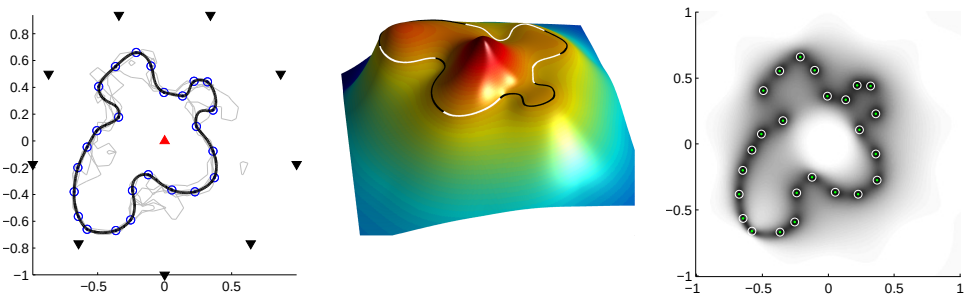
Many applications in computer vision and computer graphics require the definition of curves and surfaces. *Implicit surfaces* [7] are a popular choice for this because they are smooth, can be appropriately constrained by known geometry, and require no special treatment for topology changes. Given a scalar function $f : \mathbb{R}^d \mapsto \mathbb{R}$, one can define a manifold $\mathcal{S}$ of dimension $d-1$ wherever $f(x)$ passes through a certain value (e.g., 0)

$$\mathcal{S}_0 \triangleq \{x \in \mathbb{R}^d | f(x) = 0\}. \tag{1}$$

In this paper we introduce Gaussian processes (GPs) to this area by deriving a covariance function equivalent to the *thin plate spline* regularizer [2] in which smoothness of a function $f(x)$ is encouraged by the energy

$$E(f) = \int_\Omega \left(\nabla^{\mathrm{T}} \nabla f(x)\right)^2 \, dx \tag{2}$$

where Ω is a region of interest. This regularizer induces functions with many desirable properties for implicit surfaces [7]. Our GP approach has equal modelling power to earlier variational and radial basis function approaches [7] and similar results to a recent SVM approach [6] but with the additional strength of a meaningful probabilistic interpretation (Fig. 1ac). There is also no additional computational overhead in using a GP for this: both the GP and variational approaches require inversion of an $n \times n$ matrix, where n is the number of "constraint" or data points the function is fitted to.



(a) (b) (c)

Figure 1: **2D contour** (see text for details). (a) *Initial 2D point set* $\mathcal{X}$*:* $\{\odot = 0, \blacktriangle = +1, \blacktriangledown = -1\}$ *with contour defined by* $\mu(x) = 0$ *(solid line, also shown on (b)) and 5 random samples from the GP posterior;* (b) *Mean* $\mu(x)$ *of GP conditioned on points in (a) using thin plate covariance;* (c) *Greyscale plot illustrating marginal curve likelihood* $P(f(x) = 0)$.

1 Spline regularization is a Gaussian process

For clarity, we consider the 1D problem in this section; the generalization to 2D and 3D is demonstrated below. It is known that a regularizer of the form in equation (2) can be thought of as a Gaussian process prior (see e.g., [3]). To show this, we consider the energy as a probability and define D as the linear differential operator

$$E(f) = -\log P(f) + \text{const} = \int_\Omega \left(D^2 f(x)\right)^2 \, dx. \tag{3}$$

Using $f(\Omega)$ to denote the vector of function values for all points in the region of interest, the prior may be written (ignoring the constant) as follows:

$$-\log P(f(\Omega)) = f(\Omega)^{\mathrm{T}} [D^2]^{\mathrm{T}} D^2 f(\Omega), \tag{4}$$

which corresponds to a multivariate Gaussian distribution with zero mean and covariance $C = \left([D^2]^{\mathrm{T}} D^2\right)^{-1} = \left(D^4\right)^{-1}$.

1.1 The covariance function

We now identify the *covariance function* $c(u, v)$ giving the entries of C without recourse to online matrix inversion. Rewriting $D^4 C = I$ reveals $c(u, v)$ to be the Green's function of the fourth derivative operator [8]

$$\int_\Omega D^4(u, w) c(w, v) \, dw = \delta(u - v) \quad \Rightarrow \quad \frac{\partial^4}{\partial r^4} c(r) = \delta(r) \tag{5}$$

where $r \triangleq u - v$ to impose stationarity on the covariance. The solution of this is

$$c(r) = \tfrac{1}{6}|r|^3 + a_3 r^3 + a_2 r^2 + a_1 r + a_0. \tag{6}$$

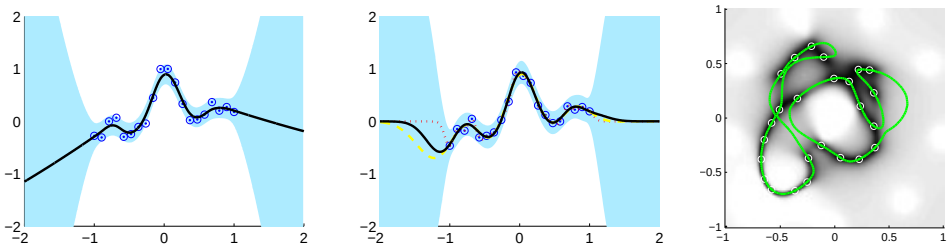
Any odd-powered polynomial terms must be equal to zero in order for the covariance function to be symmetric (i.e., $a_3 = a_1 = 0$). The other requirement is that c be positive semi-definite (psd) [5] and this may be used to set the constants a_0, a_2. Since the cubic term will "overpower" the other terms as $r \to \infty$, this is not possible for all $r \in \mathbb{R}$ (psd functions tend to zero at infinity [2]). However, if we restrict ourselves to the domain $\Omega \subset \mathbb{R}$, it is possible to set a_2, a_0 such that the covariance tends to zero at its perimeter: if R is the largest magnitude of r within Ω, stipulating that $c(R) = 0$ and $\frac{\partial}{\partial r} c(R) = 0$ gives the *thin plate covariance*

$$c(r) = \tfrac{1}{12}\left(2|r|^3 - 3Rr^2 + R^3\right). \tag{7}$$

1.2 1D regression demonstration

Fig. 2 shows a set of $n = 20$ points and their (noisy) function values $\mathcal{X} \equiv \{x_i, t_i = f(x_i) + \epsilon_i\}_{i=1}^n$ where $\epsilon_i \sim \text{Normal}\left(\epsilon_i \mid 0, \, \sigma^2\right)$. Assuming a Gaussian process prior with zero mean and covariance as equation (7), a prediction for a set of points $\mathcal{U} = \{u_j\}_{j=1}^m \subseteq \Omega$ is given by [5]

$$P(f(\mathcal{U})|\mathcal{X}) = \text{Normal}\left(f(\mathcal{U}) \mid \mu, \, Q\right) \tag{8}$$



(a) (b) (c)

Figure 2: **Thin plate vs. squared exponential covariance.** (a) *Mean (solid line) and 3 s.d. error bars (filled region) for GP regression with covariance given by equation (7);* (b) *Prediction with covariance function* $c(u_i, u_j) = e^{-\alpha\|u_i - u_j\|^2}$ *with* $\alpha = 2$ *(dashed line),* 10 *(solid) and* 100 *(dotted); error bars correspond to* $\alpha = 10$*;* (c) *Curve/likelihood obtained for points of Fig. 1 using squared exponential covariance ("best"* $\alpha = 8$ *setting shown).*

where

$$\mu = C_{ux}^{\mathrm{T}}(C_{xx} + \sigma^2 I)^{-1} t \quad \text{and} \quad Q = C_{uu} - C_{ux}^{\mathrm{T}}(C_{xx} + \sigma^2 I)^{-1} C_{ux}. \tag{9}$$

The matrices are formed by evaluating $c(\cdot, \cdot)$ between sets of points: i.e., $C_{xx} = [c(x_i, x_j)]$, $C_{ux} = [c(u_i, x_j)]$, and $C_{uu} = [c(u_i, u_j)]$.

Fig. 2a shows the mean function and 3 standard deviation error bars for this data set when $\Omega = [-2, 2]$ and $\sigma^2 = 0.01$ ($\mathrm{sd}(u_i) = \sqrt{Q_{ii}}$). Observe how the mean attempts to minimize the second derivative as dictated by equation (2). In comparison, Fig. 2b shows the predictions if the *squared exponential* [5] covariance function is employed: the mean interpolant is smooth, but the "compactness" of this covariance leads the mean function to tend towards the GP mean (here zero) away from the data. This will cause undesirable geometric effects in higher dimensions (Fig. 2c). This covariance also suffers from a nuisance length-scale parameter (although it does offer smoothness control not available with our thin plate covariance).

2 2D curves, 3D surfaces

In order to model 2D curves, given some known points on the curve, we will find $f(x)$ as in equation (1) with $d = 2$. The training data comprise $\{x_i, t_i\}$ pairs with $t_i = 0$ for points on the curve, plus additional points at ± 1 as illustrated in Fig. 1a. The equivalent of equation (5) in $\mathbb{R}^d$ is

$$\left(\nabla^{\mathrm{T}}\nabla\right)^2 c(r) = \delta(r) \tag{10}$$

where now $c(u, v) = c(r)$ with $r \triangleq \|u - v\|$. Solutions to this (with constraints at the boundary of Ω similar to those in 1D) are given by

$$\text{2D} \qquad c(r) = 2r^2 \log|r| - (1 + 2\log(R))r^2 + R^2 \tag{11a}$$

$$\text{3D} \qquad c(r) = 2|r|^3 + 3Rr^2 + R^3 \tag{11b}$$

Fig. 1 shows a GP fitted to a curve using this covariance; the prediction formulae remain the same as in equation (8), however in this example the supplied points are assumed to be noiseless ($\sigma^2 = 0$). Likewise, Fig. 3 shows some results for modelling 3D objects.

The power of using a GP for this task is that it makes probabilistic predictions and Fig. 1a shows the curves generated by some random samples from the full, joint GP posterior over Ω; note that some of these samples have different topology to the mean curve. Similarly, it is possible to provide a marginal likelihood that one or more points lie on the (unknown) surface $\mathcal{S}_0$

$$P\big(x \in \mathcal{S}_0\big) \equiv P\big(f(x) = 0\big). \tag{12}$$

A visualization of this is shown in Fig. 1c. The probabilistic nature of these surfaces is a very useful property for many applications e.g., visual tracking [1].

Further work will build on these results by (i) incorporating surface normal data to the model; and (ii) using sparse GP approximations [4] to achieve efficiency.

References

[1] A. Blake and M. Isard. *Active Contours*. Springer-Verlag, 1998.

[2] F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural network architectures. *Neural Comp.*, 7:219–269, 1995.

[3] D.J.C. MacKay. Gaussian processes - a replacement for supervised neural networks? In *Advances in Neural Information Processing Systems*, volume 9, 1997.

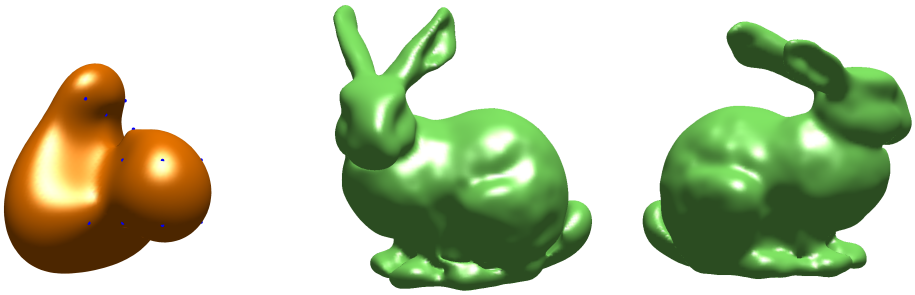
[4] J. Quiñonero-Candela and C.E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *J. Machine Learning Research*, 6:1939–1959, 2005.

[5] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[6] B. Schölkopf, J. Giesen, and S. Spalinger. Kernel methods for implicit surface modelling. In *Advances in Neural Information Processing Systems*, volume 17, pages 1193–1200, 2005.

[7] G. Turk and J.F. O'Brien. Variational implicit surfaces. Technical report, Georgia Institute of Technology, 1998.

[8] G. Wahba. *Spline Models for Observational Data*. SIAM, 1990.



(a) (b)

Figure 3: **3D surfaces.** Mean surfaces $\mu(x) = 0$ when $x \in \mathbb{R}^3$, rendered as an high resolution polygonal mesh generated by the marching cubes algorithm. (a) *A simple "blob" defined by 15 points on the surface, one interior +1 point and 8 exterior -1 points arranged as a cube;* (b) *Two views of the Stanford bunny defined by 800 surface points, one interior +1 point, and a sphere of 80 exterior -1 points.*