Surrogate Models
Hans Bruun Nielsen
IMM, Numerical Analysis Section

- Alternatives to physically based mathematical models and local Taylor expansions
- Metamodels, Surface Response, Neural Networks, ...
- Space Mapping
- Radial Basis Functions (RBF)
- Kriging, “Design and Analysis of Computer Experiments” (DACE)

Approximation tools. Interested in applications in

- Data representation (fitting)
- Optimization

Data Fitting

Given \( \{(x_i, y_i)\}_{i=1}^{m} \), \( y_i = Y(x_i) + e_i \)

Seek (an approximation to) \( Y(x) \)

May have a mathematical model

\( Y(x) \approx M(p,x) \)

Parameters \( p \) eg determined by minimizing

\( g(p) = \sum w_i^2 (y_i - M(p,x_i))^2 \)

In lack of a proper model we may use a polynomial as “surrogate model”.

Poor approximation.
Polynomials have too “long memory”.
The Taylor expansion

\[
P_n(x+h) = P_n(x) + \sum_{k=1}^{n} \frac{1}{k!} h^k P^{(k)}(x)
\]

is exact for any \( h \)

Cubic Splines

Information that should be carried by 3rd and higher derivatives is lost.
Local nature. Put knots where they are needed.

M.J.D. Powell: *Curve fitting by splines in one variable*

Knots \( \kappa_0, \kappa_1, \ldots, \kappa_n \)

\[
s(x) = \sum_{j=1}^{n+1} c_j B_j(x)
\]

Basis spline \( B_j \) is nonzero only in four consecutive knot intervals. Local support.
\( c_j \) has influence only in \([\kappa_{j-1}, \kappa_j]\)
Example. Apnea. Measurements of pressure in throat as function of distance (22 values in \([0, 10]\) cm) and time (every 0.1 second).

$$x \in \mathbb{R}^d.$$ Polynomials and splines generalize.

Curse of dimensionality

Interpolation or fitting,

$$Bc \simeq f$$

Serious risk of rank deficient \(B\).

Bicubic splines \((x = (u, v) \in \mathbb{R}^2)\).

$$s(x) = \sum_{i,j} c_{ij} B_i(\xi) B_j(\eta)$$

\(\xi\) and \(\eta\) : knots in \(u\) and \(v\)-directions, resp.

Level curves of a function that we want to approximate show \(eg\) that we need close knots in both directions at \((0.6, 0.5)\).

Also close where it is not needed. The system,

$$Ac \simeq f$$

is either rank deficient or needs many “superfluous” data points.

Alternative approximating function.

Given data points \((x_i, y_i), \quad i = 1, 2, \ldots, m\)

with distinct \(x_i \in \mathbb{R}^d\) and \(y_i \in \mathbb{R}\)

Surrogate model

$$s(x) = c^T \phi(x) + \beta^T \psi(x) = \sum_{i=1}^m c_i \phi_i(x) + \sum_{j=1}^n \beta_j \psi_j(x)$$

where the \(\phi_i\) are basis functions \(eg\) for a low order polynomial that models a “global trend”, and

$$\phi_i(x) = \phi(\|x - x_i\|_2)$$

\(\psi_j\) a Gaussian

$$\psi(r) = e^{-\theta r^2}$$

The figures show \(x \in [-1, 1]^2\)
Surrogate models based on Kriging and Radial Basis Functions (RBF) both have the form
\[ s(x) = c^T \phi(x) + \beta^T \psi(x). \]

Different derivation, but (under certain conditions on \( \phi \)) same model.

We consider interpolation, i.e. \( s(x_i) = y_i, \ i = 1, \ldots, m. \)

Let \( \Phi \in \mathbb{R}^{m \times n}, \ \Psi \in \mathbb{R}^{m \times n} \) be the matrices defined by
\[\Phi_{ij} = \phi(||x_i - x_j||_2), \quad \Psi_{ij} = \psi(x_i)\]
The interpolation condition can be expressed as
\[ \Phi \psi = y. \]

In the case of RBF this is combined with the condition that \( \psi \) should be orthogonal to the range of \( \Psi \),
and we get the linear system of equations
\[
\begin{pmatrix}
\Phi^T & \Psi^T \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
\phi \\
\beta
\end{pmatrix}
= \begin{pmatrix}
y \\
0
\end{pmatrix}
\Leftrightarrow
\begin{pmatrix}
\Phi & \Psi \\
0 & -\Psi \Phi^{-1} \Psi
\end{pmatrix}
\begin{pmatrix}
\phi \\
\beta
\end{pmatrix}
= \begin{pmatrix}
y \\
-\Psi \Phi^{-1} y
\end{pmatrix}.
\]

Solution: \( \beta = (\Psi^T \Phi^{-1} \Psi)^{-1} \) \( \Phi^T \Phi^{-1} y, \quad c = \Phi^{-1} (y - \Psi \beta). \)

**Example.** Rosenbrock’s function. \( n = 1, \ \psi(x) = 1 \)

Start with 9 points. Best \( \theta = [0.1, 100]. \) \( \text{RMS} = \sqrt{\Omega} \)

Successively insert new data points where \( \Omega(x) = \text{minimal}. \)
RBF. Gaussian. Choose $\theta$

RBF. Inverse multiquadric: $\phi_j(x) = (\theta \| x - x_j \|^2 + 1)^{-1/2}$
Plans for future work

- Better error estimation for Kriging
- Extend DACE to cope with errors in data
- Strategy for use in optimization
- Choice of $\theta$ in RBF
- Extend DACE to cope with other RBFs
- ...

With Kristine Fissenfeldt Thuesen