Maximum Entropy Approximation

N. Sukumar
University of California, Davis
http://dilbert.engr.ucdavis.edu/~suku
USNCCM8, Austin, TX
July 26, 2005
Collaborators and Acknowledgements

• **Professor Yueyue Fan** for helpful discussions on generalized inverse

• **JAVA** implementation of meshfree shape functions *(Roy Wright, UCD)*

• Research support from **NSF (CMS Division)** is acknowledged
Motivation

Inputs (Parameters)  Simulator (Function)  Output (Response)

\[ x \equiv (x_1, x_2, \ldots, x_d) \]

\[ u : \mathbb{R}^d \rightarrow \mathbb{R} \]

\[ y \]

Seek \( u^h \) s.t.

\[ u^h(x_i) = y_i \]

- Polynomial interpolants
- Splines
- Kriging
- Radial basis functions
Motivation: Applications

- Geometric Modeling, Computer Graphics, and Visualization
- Finite Element/Meshfree Galerkin Methods
- Numerical Estimation and Prediction
- Design and Analysis of Computer Experiments
Motivation: Data Approximation

\[ u^h(x) = \sum_{i=1}^{M} \phi_i(x) u_i \]
Objectives

• Merits of constructing data approximants via a constrained optimization problem

• Introduce the Maximum Entropy Principle, and to present its derivation and implementation for one-dimensional and polygonal approximation

• The promise and potential of MAXENT to solve problems with epistemic (ignorance) uncertainty
Meshfree Approximations

- DEM (Nayroles et al, 1992)
- EFG (Belytschko et al, 1994)
- RKPM (Liu et al, 1994)
- PUM (Babuska and Melenk, 1996)
- Hp-Clouds (Duarte and Oden, 1996)
- MLPG (Atluri et al, 1997)
- BNM (Mukherjee et al, 1997)
- Finite Spheres (De and Bathe, 2000)

- NEM (Braun and Sambridge, 1996)
- NEM [Laplace] (Sukumar et al, 2000)
Construction of Basis Functions

- Finite Elements
- MLS/RBFs \((L^2\) metric)
- Natural Neighbors
- MAXENT

**ISSUES**

- Defining a good neighborhood: pattern recognition, clustering, learning theory
- EBCs: Interpolants are desirable
- Numerical integration (Galerkin method)
Voronoi Neighbors
Delaunay Circumcircle and Natural Neighbors

\( \mathbf{p} \) lies outside the circumcircles in green

Convex hull
Sibson Interpolant

\[ \phi_i(p) = \frac{A_i(p)}{A(p)} \]

(Sibson, 1980)
Laplace Interpolant

\[ \alpha_i(p) = \frac{s_i(p)}{h_i(p)} \]

\[ \phi_i(p) = \frac{\alpha_i(p)}{\sum_j \alpha_j(p)} \]

(Christ et al, 1982; Belikov et al, 1997; Hiyoshi and Sugihara, 1999)
Properties

- Non-negative and PU: \( 0 \leq \phi_i \leq 1, \sum_i \phi_i(x) = 1 \)
- Interpolate data: \( \phi_i(x_j) = \delta_{ij} \)
- Linear completeness/precision: \( \sum_i \phi_i x_i = x \)
- Smoothness: \( \phi_i^{\text{LAP}} \in C^0(\Omega), \phi_i^S \in C^1(\Omega \setminus x_j) \)
- Linear essential boundary conditions can be exactly imposed
Surface Interpolation (Sibson)

Courtesy of Sung Park, CS@UCD

Bathymetry and topography data (~10,000 points) near NW Australia (Courtesy of Malcolm Sambridge)
Volume Reconstruction (Sibson): Human Head

(CT scan courtesy of NC Memorial Hospital)

256^3

10^4

5 \times 10^5

Courtesy of Sung Park, CS@UCD
Construction of Polygonal Interpolants


- **Mean value coordinates** (Floater, CAGD, 2003)

- **Laplace shape functions** (Sukumar and Tabarraei, IJNME, 2004)
Construction of Polygonal Interpolants (Cont’d)

- Maximum entropy (MAXENT) shape functions (Sukumar, IJNME, 2004)

  ✓ Imposing linear reproducibility leads to an under-determined system of linear equations for \( \{\phi_i\} \)

  ✓ Use Shannon entropy (Shannon, 1948) and max entropy principle (Jaynes, 1957) to find \( \{\phi_i\} \)

  ✓ Constrained optimization problem is solved
Wachspress Basis Functions

\[ \phi_i(x) = \frac{w_i(x)}{\sum_{j=1}^{n} w_j(x)}, \quad w_i(x) = \frac{\cot \gamma_i + \cot \delta_i}{\|x - x_i\|^2} \]

(Meyer et al., JGT, 2002)
Mean Value Coordinates

\[ \phi_i(x) = \frac{w_i(x)}{\sum_{j=1}^{n} w_j(x)}, \quad w_i(x) = \frac{\tan(\alpha_{i-1}/2) + \tan(\alpha_i/2)}{\|x - x_i\|} \]

(Floater, CAGD, 2003)
Laplace Shape Functions

Canonical Elements
Polygon Interpolant Using Affine Mapping

Laplace Shape Function
Polygonal (Laplace) Basis Function
Principle of Maximum Uncertainty/Entropy

(Shannon, 1948; Jaynes, 1957)

• Provides the least-biased solution when incomplete/insufficient information is available.

• For a discrete probability distribution \( \{p_i\}, i = 1, 2, \ldots, n \), with \( \sum p_i = 1 \), let the average (expected) value of property \( E^r \) be known: \( \sum_i p_i E_i^r = \langle E^r \rangle \).

• Maximizing the information-entropy \( H(p_i) = -\sum_{i=1}^{n} p_i \log p_i \) subject to the constraints leads to the most probable solution (Gibbs-Boltzmann distribution in statistical mechanics).
Principle of Minimum Relative Entropy

(Kullback, 1959)

- Given a prior distribution \( q \), the Kullback-Leibler distance (mutual information) between \( p \) and \( q \) is

\[
D(p | q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}, \quad I(X; Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}
\]

- Minimizing the relative-entropy with a uniform prior, \( q_i = 1/n \), is equivalent to maximizing Shannon entropy

- Other measures: Renyi and Tsallis entropies
MAXENT at Work!

• Coin toss: $p_1 + p_2 = 1$ and MAXENT gives $p_1 = p_2 = 1/2$
  $\equiv$ Principle of indifference or insufficient reason

• Wallis provided a combinatorial justification for the choice of the specific form of $H(p)$

• Suppose a die has been tossed $N$ times and we are told that the average number of spots is 4.5 and not 3.5 (honest die). Then, MAXENT gives

$$\{p_1, p_2, \ldots, p_6\} = \{0.054, 0.079, 0.114, 0.165, 0.240, 0.347\}$$
MAXENT Applications

• Statistical mechanics and physics
• Communication and natural language modeling
• Image reconstruction and biology (protein folding)
• Economics and urban planning
• Materials science (crystallography/microstructure)
• . . . and many more where uncertainty resides
MAXENT in Computational Mechanics

- Elegant and least-biased solution for scattered data approximation by associating shape functions with discrete probability measures

- Broader implications in computational mechanics:
  - Numerical estimation/prediction
  - Tailored approximants for meshfree methods
  - Microstructural design and optimization
  - Ill-posed (non-unique) inverse problems
  - Multiscale modeling
Problem Statement: MAXENT Shape Functions

\[ \text{Max } H(\phi_i) = -\sum_{i=1}^{n} \phi_i \log \phi_i \quad \text{s.t.} \]

\[ \sum_{i=1}^{n} \phi_i = 1 \]

\[ \sum_{i=1}^{n} \phi_i x_i = x \]

\[ \sum_{i=1}^{n} \phi_i y_i = y \]

\[ \phi_i : \text{`Probability of influence' of node } i \text{ at } x \]

Constraints

\[ \mathbf{P} \quad \mathbf{\phi} = \mathbf{p} \]
Minimum Norm Solution

General Solution of \( P\varphi = p \):

\[
\varphi = P^+ p + (I - P^+ P)c
\]

and if \( c = 0 \) we obtain the min-norm solution:

\[
\varphi = P^+ p, \quad P^+ \equiv \text{Generalized Inverse}
\]

which is the solution of

\[
\text{Min} \left( H(\varphi) = \|\varphi\|_2 \right)
\]

s.t. \( P\varphi = p \)

Since \( \phi_i < 0 \) is possible, \( H(\bullet) \) is not suitable as an uncertainty measure.
MLS and Weighted Minimum-Norm Solution

**MLS:**  \[ \text{Min} \; \| \mathbf{W}^{1/2} (\mathbf{P}^T \mathbf{a} - \mathbf{u}) \|_2^2 \quad \Rightarrow \quad \mathbf{Aa} = \mathbf{Bu} \]

\[ \mathbf{A} = \mathbf{PWP}^T, \quad \mathbf{B} = \mathbf{PW} \]

\[ \phi = \mathbf{B}^T \mathbf{A}^{-1} \mathbf{p} = \mathbf{WP}^T \gamma \]

**Primal Problem:**  \[ \text{Min} \; \left( \phi^T \mathbf{W}^{-1} \phi \right) \quad s.t. \quad \mathbf{P} \phi = \mathbf{p} \]

Let \[ \psi = \mathbf{W}^{-1/2} \phi, \quad \mathbf{Q} = \mathbf{PW}^{1/2} \]

then \[ \text{Min} \; \| \psi \|_2^2 \quad s.t. \quad \mathbf{Q} \psi = \mathbf{p} \]

\[ \phi = \mathbf{W}^{1/2} \mathbf{Q}^+ \mathbf{p} \] \((\mathbf{Q}^+: \text{Matlab function pinv})\)
MAXENT Solution Using Lagrange Multipliers

• First variation of augmented Lagrangian is zero \((\delta L = 0)\)

\[
L = -\sum_{i=1}^{n} \phi_i \log \phi_i + \lambda_0 \left( 1 - \sum_i \phi_i \right) + \lambda_1 \left( x - \sum_i \phi_i x_i \right) \\
+ \lambda_2 \left( y - \sum_i \phi_i y_i \right)
\]

\[
\delta L = (-1 - \log \phi_i - \lambda_0 - \lambda_1 x_i - \lambda_2 y_i) \delta \phi_i = 0 \quad \forall \delta \phi_i
\]

and since the variations \(\delta \phi_i\) are arbitrary

\[
-1 - \log \phi_i - \lambda_0 - \lambda_1 x_i - \lambda_2 y_i = 0 \quad (i = 1, 2, \ldots, n)
\]
• Letting $\lambda_0 = \log Z - 1$ ($Z$ is the partition function), we get

$$\log \phi_i + \log Z = -\lambda_1 x_i - \lambda_2 y_i$$

• Since $\sum_i \phi_i = 1$,

$$\phi_i = \frac{e^{-\lambda_1 x_i - \lambda_2 y_i}}{Z}, \quad Z = \sum_{j=1}^{n} e^{-\lambda_1 x_j - \lambda_2 y_j}$$
MAXENT Solution (Cont’d)

• If only one constraint exists ($\lambda_1 = \lambda_2 = 0$), then $Z = n$

$$\phi_i = \frac{1}{n} \forall i \quad \text{(nearest-neighbor interpolant)}$$

• In general, $\lambda_1$ and $\lambda_2$ satisfy two non-linear equations:

$$- \frac{\partial (\log Z)}{\partial \lambda_1} = x \iff \sum_{i=1}^{n} e^{-\lambda_1 x_i - \lambda_2 y_i} x_i - x = 0$$

$$- \frac{\partial (\log Z)}{\partial \lambda_2} = y \iff \sum_{i=1}^{n} e^{-\lambda_1 x_i - \lambda_2 y_i} y_i - y = 0$$
Numerical Algorithm for MAXENT Shape Functions

• Let $\widetilde{x}_i = x_i - x$, $\widetilde{y}_i = y_i - y$. Then,

$$f_1(\lambda_1, \lambda_2) = \frac{\partial (\log \tilde{Z})}{\partial \lambda_1} = 0 \quad \iff \quad -\sum_{i=1}^{n} e^{-\lambda_1 \widetilde{x}_i - \lambda_2 \widetilde{y}_i} \widetilde{x}_i \tilde{Z} = 0$$

$$f_2(\lambda_1, \lambda_2) = \frac{\partial (\log \tilde{Z})}{\partial \lambda_2} = 0 \quad \iff \quad -\sum_{i=1}^{n} e^{-\lambda_1 \widetilde{x}_i - \lambda_2 \widetilde{y}_i} \widetilde{y}_i \tilde{Z} = 0$$

• The vector field $\mathbf{f}$ is the gradient of a scalar potential:

$$F = \log \tilde{Z}(\lambda_1, \lambda_2), \quad \mathbf{f} = \nabla F$$
Numerical Algorithm (Cont’d)

• Recast the MAXENT formulation as a convex minimizer (dual) problem (Agmon et al., JCP, 1979):

Find \( (\lambda_1, \lambda_2) \) s.t. \( F = \log \tilde{Z}(\lambda_1, \lambda_2) \) is minimized

• Initial guess
  \[
  \lambda^0 = 0
  \]
  \[
  \lambda_{r+1}^k = \lambda_r^k + \alpha \Delta \lambda_r^k, \quad \Delta \lambda^k = -\nabla F
  \]

• \( \alpha \) is determined by the condition that \( F(\lambda_1^{k+1}, \lambda_2^{k+1}) \) is minimized along the search direction

• Convergence criterion:
  \[
  \|\nabla F\|^k < 10^{-7}
  \]
MAXENT Shape Functions in 1D

\[ \phi_i = \frac{e^{-\lambda_1 x_i}}{Z}, \quad Z = \sum_{j=1}^{3} e^{-\lambda_1 x_j} \quad x_1 = 0, \quad x_2 = 1/2, \quad x_3 = 1 \]

\[ Z = 1 + e^{-\lambda_1/2} + e^{-\lambda_1}, \quad \phi_1 = \frac{1}{Z}, \quad \phi_2 = \frac{e^{-\lambda_1/2}}{Z}, \quad \phi_3 = \frac{e^{-\lambda_1}}{Z} \]

\[ \sum_{i=1}^{3} \phi_i x_i = x : \quad \frac{\eta}{2} + \eta^2 = x(1 + \eta + \eta^2), \quad \eta = e^{-\lambda_1/2} \]
MAXENT Shape Functions in 1D (Cont’d)

\[
\eta = \frac{2x - 1 + \sqrt{12x(1-x)} + 1}{4(1-x)}
\]

\[
\phi_1 = \frac{1}{1 + \eta + \eta^2}
\]

\[
\phi_2 = \frac{\eta}{1 + \eta + \eta^2}
\]

\[
\phi_3 = \frac{\eta^2}{1 + \eta + \eta^2}
\]
MAXENT Shape Functions in 1D (Cont’d)

n = 11

n = 6
Square: MAXENT Shape Functions

\[ \Omega = (0,1)^2 \]

\[ Z = \sum_{j=1}^{4} e^{-\lambda_1 x_j - \lambda_2 y_j} \]

which simplifies to

\[ \frac{e^{-\lambda_1}}{1 + e^{-\lambda_1}} = x, \quad \frac{e^{-\lambda_2}}{1 + e^{-\lambda_2}} = y \Rightarrow e^{-\lambda_1} = \frac{x}{1 - x}, \quad e^{-\lambda_2} = \frac{y}{1 - y} \]
Square (Cont’d)

Since \( \phi_i = \frac{e^{-\lambda_1 x_i - \lambda_2 y_i}}{Z} \), \( Z = \sum_{j=1}^{n} e^{-\lambda_1 x_j - \lambda_2 y_j} \),

we obtain \( Z^{-1} = (1 - x)(1 - y) \) and therefore

\[
\begin{align*}
\phi_1(x, y) &= (1 - x)(1 - y), \\
\phi_2(x, y) &= x(1 - y), \\
\phi_3(x, y) &= xy, \\
\phi_4(x, y) &= y(1 - x)
\end{align*}
\]

which are the same as bilinear finite element shape functions
Square (Cont’d)

MAXENT ≡ Bilinear FE Interpolation

Shape Function

Entropy = log \( \hat{Z} \)
Square: Convergence

\[ F = \log \tilde{Z} @ x = (0.56, 0.42) \]
Square: Convergence (Cont’d)

\[ F = \log \tilde{Z} @ x = (0.9,0.12) \]

\[ F = \log \tilde{Z} @ x = (0.99,0.12) \]

Use of nonlinear CG leads to faster convergence
Hexagon: Shape Functions

MAXENT

Laplace

Mean-Value Coordinates
Hexagon: Normalized Entropy

Mean-value coordinates

Laplace
Bubble (Shape) Function

Contour plot

3D
Mid-Side Node: Shape Function

Five-node element

Shape function of node a
Mid-Side Node: Maximum Entropy Distribution
Shape Function (MAXENT) Derivatives

\[ \frac{\partial \phi_i}{\partial \alpha} \phi_i \left( (x - x_i) \frac{\partial \lambda_1}{\partial \alpha} + (y - y_i) \frac{\partial \lambda_2}{\partial \alpha} \right), \quad \alpha = x, y \]

\[ \begin{bmatrix}
\frac{\partial \lambda_1}{\partial x} & \frac{\partial \lambda_1}{\partial y} \\
\frac{\partial \lambda_2}{\partial x} & \frac{\partial \lambda_2}{\partial y}
\end{bmatrix}
= -\begin{bmatrix}
\langle x^2 \rangle - x^2 & \langle xy \rangle - xy \\
\langle xy \rangle - xy & \langle y^2 \rangle - y^2
\end{bmatrix}^{-1},
\]

where \( \langle f \rangle = \sum_{i=1}^{n} \phi_i f_i \)
Galerkin Method (Patch Test)

Error norms:

\[ \frac{\| u - u^h \|_2}{\| u \|_2} \approx 10^{-8}, \quad \frac{\| u - u^h \|_E}{\| u \|_E} \approx 10^{-7} \]
Shape Function Visualization: JAVA Applet

(Developed by Roy Wright, UCD)
JAVA Applet (Cont’d)
JAVA Applet (Cont’d)

3D plot
JAVA Applet (Cont’d)

Mouse-click to insert
Right-click to delete
JAVA Applet (Cont’d)

MVC (Non-Convex)

Enabled
Visualization of Shape Functions

Contour plot
Side Node

Contour plot
Side Node (Cont’d)

3D Plot
Interior Node
Interior Node (Cont’d)

3D Plot
Related Applications: A. Supervised Learning


**Objective:** Estimation of unknown quantities based on observed samples (numerical estimation), for e.g., pollutants in a city, spam e-mail, speech recognition

Feature Random Vector $X \in \mathbb{R}^d$ relationship Observation $Y \in \mathbb{R}$ /Classification RV

For $d \uparrow$, the curse of dimensionality!

Given $P_{X,Y}$ and iid data $\{(X_1, Y_1), (X_2, Y_2), \ldots, (X_2, Y_2)\}$

ESTIMATE $P_{Y|X}$
A. Supervised Learning (LIME Algorithm)

**Distortion function** $D: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$ is the mean squared error

$$D(r, s) = \frac{1}{d} \sum_{k=1}^{d} (r_k - s_k)^2$$

Compute weights $w_i(X)$ (partition of unity) by solving

$$\text{Minimize} \left[ D\left( \sum_{i=1}^{k} w_j X_j, X \right) - \alpha H(w) \right],$$

where $\alpha$ is chosen and $k$ training samples are picked

$$\hat{P}_{Y|X}(g | x) = \sum_{i=1}^{k} w_i(x) I_{Y_i(x) = g}$$
B. Local MAXENT Meshfree Method

(Arroyo and Ortiz, 2005)

Minimize \( \left[ \beta U(\phi) - H(\phi) \right] \)

\[
U(\phi) = \sum_{i=1}^{n} \phi_i \| x - x_i \|^2 \quad \text{(second-order moment)}
\]

\[
H(\phi) = -\sum_{i=1}^{n} \phi_i \log \phi_i \quad \text{(Shannon entropy)}
\]

subject to the three linear reproducing conditions

Presentation by Marino Arroyo forthcoming on Wednesday, July 27, 2005 (USNCCM8)
C. Nodal Refinement

``sprinkle nodes``

integration cell

node

``sprinkle nodes``

crack
Linked the use of the maximum entropy principle to data approximation; use of extremum principles to compute shape functions have well-established roots (Kriging, Delaunay, thin-plate splines, MLS, Laplace).

Numerical formulation to solve the MAXENT problem in 1D and 2D was described, which readily extends to \( \mathbb{R}^d \quad (d \in \mathbb{N}) \). A JAVA applet to plot meshfree shape functions has been developed.

The use of information-theoretic principles in materials and mechanics computations holds promise.