

Overview and construction of meshfree basis functions: From moving least squares to entropy approximants

N. Sukumar^{1,*},[†] and R. W. Wright²

¹*Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, U.S.A.*

²*Department of Mathematics, University of California, Davis, CA 95616, U.S.A.*

SUMMARY

In this paper, an overview of the construction of meshfree basis functions is presented, with particular emphasis on moving least-squares approximants, natural neighbour-based polygonal interpolants, and entropy approximants. The use of information-theoretic variational principles to derive approximation schemes is a recent development. In this setting, data approximation is viewed as an inductive inference problem, with the basis functions being synonymous with a discrete probability distribution and the polynomial reproducing conditions acting as the linear constraints. The maximization (minimization) of the Shannon–Jaynes entropy functional (relative entropy functional) is used to unify the construction of globally and locally supported convex approximation schemes. A JAVA applet is used to visualize the meshfree basis functions, and comparisons and links between different meshfree approximation schemes are presented. Copyright © 2006 John Wiley & Sons, Ltd.

Received 10 August 2006; Accepted 10 August 2006

KEY WORDS: information theory; relative entropy; radial basis functions; natural neighbours; convex approximation schemes; JAVA programming

1. INTRODUCTION

Data approximation from a scattered set of points (nodes) in \mathbb{R}^d is required in many applications: computer graphics and visualization, image processing, regression models, supervised learning, and finite element and meshfree methods to name a few. In this paper, we focus on the construction of approximation schemes within the convex hull of a set of nodes—convex polygons (n -gons) and scattered set of nodes are considered. Barycentric co-ordinates on irregular polygons are of

*Correspondence to: N. Sukumar, Department of Civil and Environmental Engineering, University of California, One Shields Avenue, Davis, CA 95616, U.S.A.

[†]E-mail: nsukumar@ucdavis.edu

Contract/grant sponsor: NSF; contract/grant number: CMS-0626481, DMS-0135345

interest in computer graphics and geometric modelling [1–3], and they are also used in polygonal finite element methods [4]. The scattered data approximation problem that we consider has been studied using moving least-squares (MLS) approximants [5], natural neighbour-based interpolants [6, 7], and radial basis functions (RBFs) [8–10]. A recent advance in this direction has been the use of information-theoretic variational principles to construct meshfree basis functions [11–13]. We elaborate on the rationale of this approach, and provide a unifying framework to view entropy approximants. A JAVA applet is developed to visualize meshfree basis functions, with an aim to readily discern the similarities and distinctions between different meshfree approximants.

The outline of this paper follows. We first present some of the essential properties of data approximations schemes, and then finite element and meshfree Galerkin methods are touched upon. In Section 3, the main functionalities and capabilities of the JAVA applet are presented, and basis function plots that are created using the applet appear in Sections 4 and 5. The construction of MLS approximants and barycentric co-ordinates on irregular polygons are described in Section 4. To motivate the adoption of entropy-based approximants, the key ingredients of Bayesian theory of probability are outlined in Section 5. In Section 5.1, we present the derivation of basis functions using Jaynes's principle of maximum-entropy (MAXENT) [14, 15] as well as through its generalization, the principle of minimum relative entropy (Shannon–Jaynes entropy functional) [16–18]. Entropy-based higher-order approximation schemes are proposed in Section 5.2, and we close with a few concluding remarks in Section 6.

2. PRELIMINARIES

Consider a set of distinct nodes $\{\mathbf{x}_i\}_{i=1}^n$ in \mathbb{R}^d . The convex hull, $\mathcal{C} \subset \mathbb{R}^d$, is the smallest convex set that contains all the nodes. Let us introduce a point p with co-ordinate \mathbf{x} within $\mathcal{C} \subset \mathbb{R}^2$ (Figure 1). An approximation scheme for a scalar-valued function $u(\mathbf{x}) : \mathcal{C} \rightarrow \mathbb{R}$ can be written as

$$u^h(\mathbf{x}) = \sum_{i=1}^n \phi_i(\mathbf{x}) u_i \quad (1)$$

where u_i are coefficients (nodal values if u^h is an interpolant) and $\phi_i(\mathbf{x})$ is the basis function of node i . If \mathbf{x}_i are the vertices of a polyhedron, then $\{\phi_i\}_{i=1}^n$ are the barycentric co-ordinates of p . On convex polygons, the MAXENT basis functions are a barycentric co-ordinate, but they are non-interpolatory on a set of scattered nodes (Figure 1(d)). Maximum-entropy basis functions were

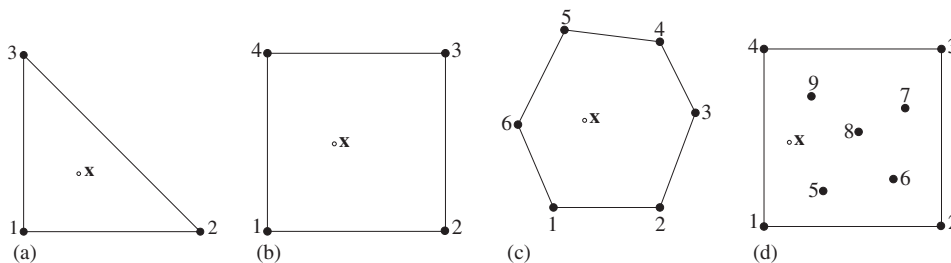


Figure 1. (a) Triangle; (b) square; (c) hexagon; and (d) scattered nodes.

introduced by Sukumar [11] and Arroyo and Ortiz [12]. In Reference [12], its link to convex analysis [19] were noted, and these approximants were referred to as convex approximation schemes. For Equation (1) to be a linearly precise barycentric co-ordinate or convex approximation scheme [12], the basis functions must satisfy the following properties:

1. Ability to reproduce constant and linear functions exactly:

$$\forall \mathbf{x}, \quad \sum_{i=1}^n \phi_i(\mathbf{x}) = 1 \quad \text{and} \quad \sum_{i=1}^n \phi_i(\mathbf{x}) \mathbf{x}_i = \mathbf{x} \tag{2}$$

For second-order partial differential equations (PDEs), approximants that possess constant and linear precision are sufficient for convergence in a Galerkin method [20].

2. The $\{\phi_i(\mathbf{x})\}_{i=1}^n$ are sufficiently smooth (typically C^∞) in the interior of \mathcal{C} .
3. Convex combination:

$$\phi_i(\mathbf{x}) \geq 0 \quad \forall i, \mathbf{x} \tag{3}$$

which in conjunction with Equation (2) indicates that the $\phi_i(\mathbf{x})$ are bounded between zero and unity and satisfy the convex hull property. The non-negative condition leads to the variation diminishing property and to positive-definite mass matrices [12]. Convex approximation schemes are not prone to the Runge phenomena [21], which occurs when using higher-order one-dimensional Lagrange interpolation on uniform grids. In addition, optimal conditioning can be established for non-negative basis functions [22–24]. The adoption of NURBS-based convex basis functions to ensure geometric exactness in finite element analysis has been recently introduced by Hughes and co-workers [25]. MLS approximants [5], which are widely used in meshfree Galerkin methods, are not convex approximants since MLS basis functions can be negative.

Consider the Poisson equation with homogeneous Dirichlet (essential) boundary conditions:

$$-\nabla^2 u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega \tag{4}$$

where $\Omega \subset \mathbb{R}^2$ is the problem domain and $\partial\Omega$ is its boundary. In the finite element method, the domain is partitioned into M non-overlapping subdomains (elements), $\Omega = \bigcup_{k=1}^M \mathcal{T}_k$, where \mathcal{T}_k is either a triangle or a quadrilateral (isoparametric transformation of the elements shown in Figures 1(a) and (b)). In finite elements, basis functions are associated with each vertex (node) of \mathcal{T}_k , and within an element, the local restriction of a basis function is known as a shape function.

In a Galerkin method, a weak or variational statement of the strong form, Equation (4), is used: find $u \in H_0^1(\Omega)$ such that

$$\int_{\Omega} \nabla(\delta u) \cdot \nabla u \, d\Omega = \int_{\Omega} f \delta u \, d\Omega \quad \forall \delta u \in H_0^1(\Omega) \tag{5}$$

where $H_0^1(\Omega)$ is the Sobolev space of functions in $L^2(\Omega)$ whose derivatives are also square-integrable, and with vanishing function values on $\partial\Omega$, and δu denotes the first variation of u . In finite element and meshfree Galerkin methods, $u(\mathbf{x})$ is approximated by $u^h(\mathbf{x})$ (belongs to a finite-dimensional subspace of $H_0^1(\Omega)$) of the form given in Equation (1). A similar approximation is used for the test function δu . Since the nodal coefficients that appear in δu^h are arbitrary,

a linear system of equations is obtained:

$$\mathbf{K}\mathbf{u} = \mathbf{f}, \quad \mathbf{K}_{ij} = \int_{\Omega} \nabla \phi_i \cdot \nabla \phi_j \, d\Omega, \quad \mathbf{f}_i = \int_{\Omega} \phi_i f \, d\Omega \quad (6)$$

In the finite element method, the shape functions $\phi_i(\mathbf{x})$ are defined on reference elements (Figures 1(a) and (b)), and through an isoparametric transformation, they are obtained on any triangle or convex quadrilateral. Finite element basis functions satisfy three key properties, which facilitates the exact imposition of linear Dirichlet boundary conditions:

1. Basis functions satisfy linear reproducing conditions (linearly complete) on meshes with either triangular or quadrilateral elements.
2. Interior nodal basis functions do not contribute at a point p that lies on the boundary of the domain.
3. The basis functions possess the Kronecker-delta property (cardinal basis): $\phi_i(\mathbf{x}_j) = \delta_{ij}$.

On meshes with convex n -gons, MAXENT basis functions are a barycentric co-ordinate and they satisfy all the above properties [11]. Maximum-entropy basis functions meet the first and second property on convex domains, but for three or more nodes that are collinear on the boundary of the domain, a weaker Kronecker-delta property is satisfied. Extensive research in meshfree methods has focused on modifications to MLS (e.g. use of singular weight functions or transformations) and to other meshfree approximants with the objective of constructing basis functions with the Kronecker-delta property. The lack of the Kronecker-delta property on the boundary by itself does not pose any limitation; the first two properties that are indicated above are the critical ones and they suffice to enable the imposition of essential boundary conditions in MAXENT meshfree methods as in finite elements [12].

3. JAVA APPLLET FOR VISUALIZATION OF MESHFREE BASIS FUNCTIONS

There have been significant advances in the use of meshfree approximants for the solution of partial differential equations [26–34]. However, in spite of the maturity of meshfree methods, there are currently no tools in the public-domain to visualize meshfree basis functions. We have developed JAVA applets to visualize basis functions in one, two, and three dimensions; the two-dimensional JAVA applet menu is shown in Figure 2. The applets serve as a suitable aid to readily discern the similarities and distinctions between the different meshfree approximation schemes.

The creation of a web-accessible JAVA programme allows users to create an arbitrary nodal set (convex or non-convex) in one, two or three dimensions by inputting the co-ordinates of its nodes, with the aid of a direct data entry form or a point-and-click interface. Options such as P (polygon), Q (quadtree grid), and R (random nodes in a unit square) are indicated in Figure 2. This programme displays a visualization of the basis function associated with a node and specific formulation, both picked by the user. Available formulations in two dimensions include Delaunay and polygonal interpolation schemes, MLS, and maximum-entropy approximations using a uniform prior distribution (lead to global basis functions) and a compactly supported prior distribution (lead to compactly supported basis functions). In one dimension, individual basis functions or all the basis functions on a grid can be displayed. Two-dimensional basis functions may be displayed as contour plots or surface plots. The plots are generated by dividing the polygonal element into triangles and then dividing the triangles recursively. The user can calculate the value of the basis

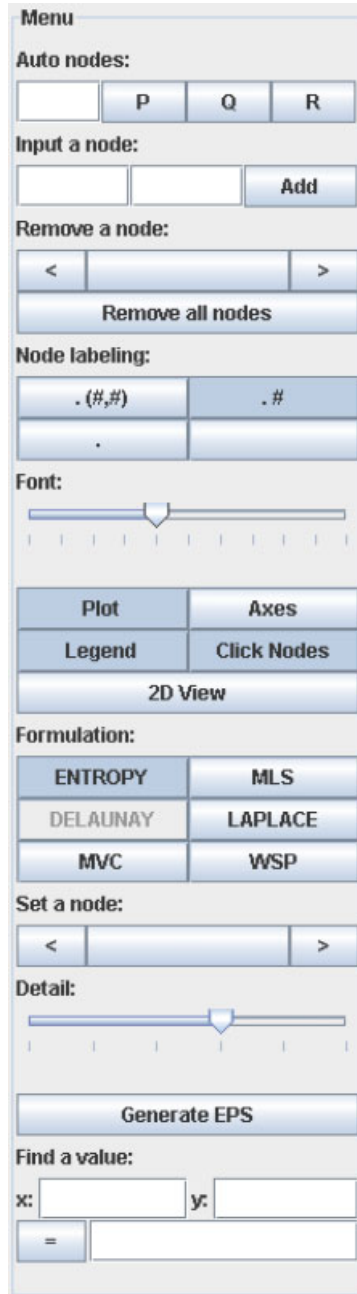


Figure 2. JAVA applet menu.

functions at an arbitrary point within the convex hull. In 3D, basis function values on planes that cut the convex hull are computed.

The visualization package is coded as a Java 1.4 applet and embedded in an HTML page that detects screen resolution and adjusts the applet's behaviour accordingly. The code is object-oriented and hence modular, making it fairly easily extended. Visualizations resemble as closely as possible the desirable styles of texts and journals, and software such as Mathematica™ and Matlab™. The generated visualizations are suitable for publication (EPS option in Figure 2). The capabilities of the applet are demonstrated through basis function plots that appear in the ensuing sections.

4. MESHFREE APPROXIMANTS AND POLYGONAL INTERPOLANTS

Currently, most meshfree Galerkin methods are based on approximants that can be classified into three distinct types: RBFs [8, 9], MLS approximants [5], and natural neighbour interpolation schemes [6, 7]. In meshfree methods, the approximation scheme is of the form Equation (1), but the construction of the nodal basis functions $\{\phi_i\}_{i=1}^n$ is not tied to a background element structure. A brief description of MLS, RBFs, and polygonal interpolants follows.

4.1. Moving least squares approximants

In the MLS approximation, each node is associated with a compactly supported weight function, $w : [0, \infty) \rightarrow \mathbb{R}$:

$$w_i(\mathbf{x}) \equiv w(q_i), \quad q_i = \frac{\|\mathbf{x} - \mathbf{x}_i\|}{r_i^{\max}} \quad (7)$$

where $\|\cdot\|$ is the L^2 norm of its argument, r_i^{\max} is the radius of support for the nodal weight function, and $w(q)$ is a smooth, non-increasing weight function that is maximal at $q=0$ and vanishes for $q \geq 1$. The global MLS approximation is [5]:

$$u^h(\mathbf{x}) = \sum_{j=1}^m p_j(\mathbf{x}) a_j(\mathbf{x}) \equiv \sum_{i=1}^n \phi_i(\mathbf{x}) u_i \quad (8)$$

where \mathbf{p} is a basis vector (for example, $\mathbf{p} = \{1 \ x \ y\}^T$ is a linear basis in 2D) and $a_j(\mathbf{x})$ are unknown parameters that are found by solving a quadratic weighted least squares minimization problem [5]:

$$\min_{\mathbf{a}} \frac{1}{2} \sum_{i=1}^n w_i(\mathbf{x}) [\mathbf{p}^T(\mathbf{x}_i) \mathbf{a}(\mathbf{x}) - u_i]^2 \quad \text{or} \quad \min_{\mathbf{a}} \frac{1}{2} (\mathbf{P}\mathbf{a} - \mathbf{u})^T \mathbf{W} (\mathbf{P}\mathbf{a} - \mathbf{u}) \quad (9)$$

On carrying out the minimization, the solution for the MLS basis functions is given by [28]

$$\phi_i(\mathbf{x}) = \mathbf{p}^T(\mathbf{x}) \mathbf{A}(\mathbf{x})^{-1} \mathbf{B}_i(\mathbf{x}) \quad (10a)$$

where the matrices $\mathbf{A}(\mathbf{x})$ and $\mathbf{B}(\mathbf{x})$ are

$$\mathbf{A}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) \mathbf{p}(\mathbf{x}_i) \mathbf{p}^T(\mathbf{x}_i) \quad (10b)$$

$$\mathbf{B}(\mathbf{x}) = [w_1(\mathbf{x}) \mathbf{p}(\mathbf{x}_1), w_2(\mathbf{x}) \mathbf{p}(\mathbf{x}_2), \dots, w_n(\mathbf{x}) \mathbf{p}(\mathbf{x}_n)] \quad (10c)$$

In the numerical implementation, a cubic spline weight function, $w(q) \in C^2(\mathbf{R}_+)$, is used [35]:

$$w(q) = \begin{cases} \frac{2}{3} - 4q^2 + 4q^3 & \text{if } 0 \leq q \leq \frac{1}{2} \\ \frac{4}{3} - 4q + 4q^2 - \frac{4q^3}{3} & \text{if } \frac{1}{2} \leq q \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

The nodal weight function support radius $r_i^{\max} = \alpha h_i$, where h_i is chosen for each node by a procedure very similar to Algorithm 2 in Reference [36]:

For the set of nodes $\{\mathbf{x}_i\}_{i=1}^n$ and their convex hull \mathcal{C} ,

1. Choose positive integers m_1 and m_2 and a constant $\alpha \geq \alpha_0 > 1$ and set $h_i = 0$ for $i = 1, 2, \dots, n$.
2. Assemble a set of points \mathcal{P} in three steps:
 - (a) Create a $[m_1]^d$ uniform grid of points over \mathcal{C} and discard those falling outside.
 - (b) Add m_2 random, uniformly distributed points in \mathcal{C} .
 - (c) Add the nodes $\{\mathbf{x}_i\}_{i=1}^n$.
3. For each $\mathbf{p} \in \mathcal{P}$,
 - (a) Find the $d+1$ nodes $\{\mathbf{x}_{i^*}\}_{i^*=1}^{d+1}$ in $\{\mathbf{x}_i\}_{i=1}^n$ that are closest to \mathbf{p} , and compute their Euclidean distance $d_{i^*} = \|\mathbf{p} - \mathbf{x}_{i^*}\|$.
 - (b) If $h_{i^*} < d_{i^*}$, set $h_{i^*} = d_{i^*}$.
4. Set $r_i^{\max} = \alpha h_i$ for $i = 1, 2, \dots, n$.

For the 2D applet, $m_1 = 100$, $m_2 = 10\,000$, $\alpha_0 = 1.01$, and the constant α is chosen by the user via a slider. This procedure is necessary to ensure that every point in \mathcal{C} is covered by at least $d + 1$ nodal weight function supports, so that the matrix $\mathbf{A}(\mathbf{x})$, defined in Equation (10b), has full rank.

4.2. Radial basis functions

Consider the approximation of a function $u(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$ using the set of scattered nodes $\{\mathbf{x}_i\}_{i=1}^n$. In the RBF approximation, a fixed radial function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is chosen, i.e. $\Phi(\mathbf{x}) \equiv \varphi(\|\mathbf{x}\|)$ with $\varphi : [0, \infty) \rightarrow \mathbb{R}$. On using translates of this radial function with centres at \mathbf{x}_i , the following ansatz is made [10]:

$$u^h(\mathbf{x}) = \sum_{i=1}^n \varphi(\|\mathbf{x} - \mathbf{x}_i\|) a_i \quad (12)$$

where a_i are unknown coefficients. Often, a polynomial term is also included in the above approximation if global polynomial reproducibility is desired. For certain choices of $\varphi(\cdot)$, for example, Gaussian, multiquadrics, or thin-plate splines, the matrix $K_{ij} = \varphi(\|\mathbf{x}_j - \mathbf{x}_i\|)$ is positive-definite and invertible, and hence the data interpolation problem, $u^h(\mathbf{x}_j) = u(\mathbf{x}_j)$ ($j = 1, 2, \dots, n$), results in a unique solution for \mathbf{a} . The use of RBFs in collocation-based meshfree methods was initiated by Kansa [37, 38], and new developments and advances continue to emerge in this topical research area. In this paper, RBFs are adopted as prior distributions (weights) within the Shannon–Jaynes maximum-entropy formalism.

4.3. Polygonal interpolants

Using elements of projective geometry, Wachspress [39] proposed rational polynomial interpolants for convex polygons. Recently, there have been additional contributions on the construction of barycentric co-ordinates on irregular polygons [1, 2, 4, 11]. A review on the construction of polygonal interpolants is presented by Sukumar and Malsch [40].

In Reference [1], a simple expression is obtained for Wachspress’s basis functions:

$$\phi_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{j=1}^n w_j(\mathbf{x})}, \quad w_i(\mathbf{x}) = \frac{A(p_{i-1}, p_i, p_{i+1})}{A(p_{i-1}, p_i, p)A(p_i, p_{i+1}, p)} = \frac{\cot \gamma_i + \cot \delta_i}{\|\mathbf{x} - \mathbf{x}_i\|^2} \tag{13}$$

where the last expression is used in our numerical implementation. In the above equation, $A(a, b, c)$ is the signed area of triangle $[a, b, c]$, and γ_i and δ_i are shown in Figure 3(a).

Floater [2] used the mean value theorem for harmonic functions to develop barycentric co-ordinates on polygons. The linearly precise mean value co-ordinate is [2]:

$$\phi_i(\mathbf{x}) = \frac{w_i(\mathbf{x})}{\sum_{j=1}^n w_j(\mathbf{x})}, \quad w_i(\mathbf{x}) = \frac{\tan(\alpha_{i-1}/2) + \tan(\alpha_i/2)}{\|\mathbf{x} - \mathbf{x}_i\|} \tag{14}$$

where the angle α_i is shown in Figure 3(b). In the JAVA applet, the algorithm for mean value co-ordinates proposed by Hormann (Figure 6 in Reference [41]) is used. The implementation is valid for convex and non-convex polygons.

Natural neighbour interpolation methods are Voronoi-based convex approximation schemes that interpolate nodal data and share many common properties with the finite element interpolant. Cueto and co-workers [42] provide an overview of the construction of natural neighbour-based interpolants. In Reference [4], Laplace basis functions [7] are constructed on regular polygons, and through an isoparametric mapping, the basis functions are defined on irregular polygons. The Wachspress basis functions and mean value co-ordinates are directly computed on irregular polygons, which is also the case in a recently proposed non-conforming finite element method on polyhedral meshes [43]. The interested reader can refer to Reference [40] and the references therein for further details on the construction and implementation of polygonal interpolants.

In Figure 4, Wachspress, mean value and Laplace basis functions for the hexagon in Figure 1(c) are plotted. These basis functions share the properties of polygonal barycentric co-ordinates. In Figure 5, the capabilities of the applet are further illustrated by presenting Laplace basis

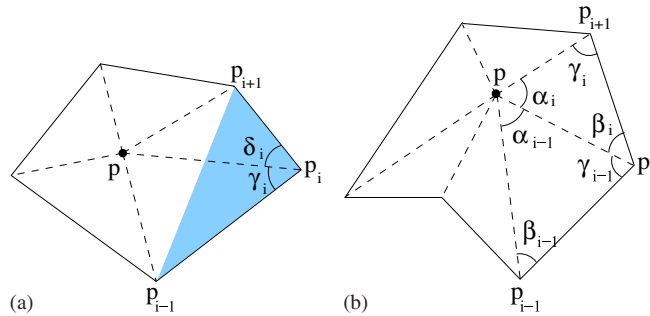


Figure 3. Barycentric co-ordinates: (a) Wachspress [1]; and (b) mean value co-ordinates [2].

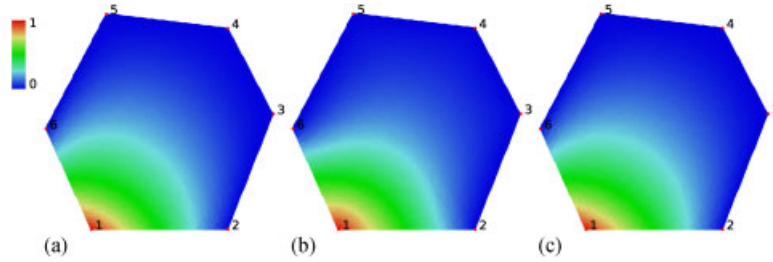


Figure 4. Hexagonal basis functions (node 1): (a) Wachspress; (b) mean value co-ordinates; and (c) Laplace.

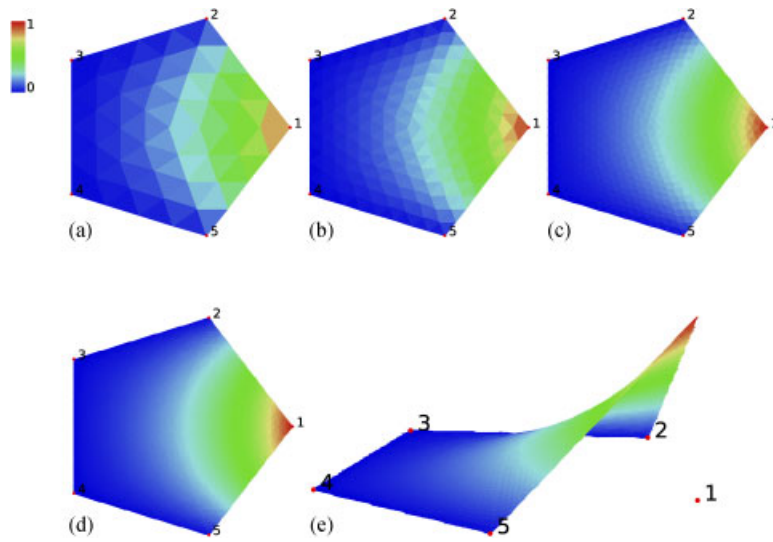


Figure 5. Laplace basis functions (node 1) for a regular pentagon at varying resolutions (a)–(d) and (e) 3D plot.

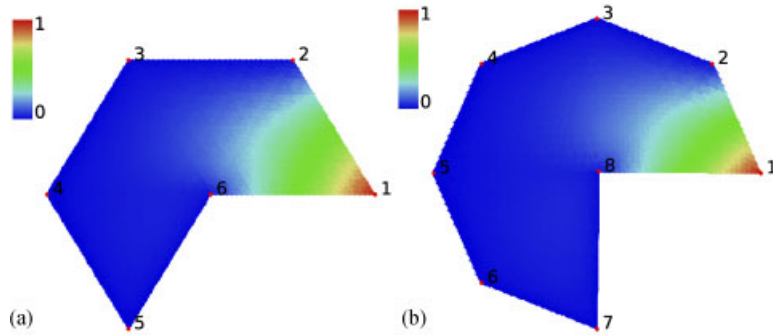


Figure 6. Mean value co-ordinates on concave polygons (node 1): (a) hexagon; and (b) octagon.

function contour plots on a regular pentagon for varying resolutions; a 3D perspective is shown in Figure 5(e). Mean value co-ordinates are also linearly precise on concave (non-convex) polygons. In Figure 6, contour plots of mean value co-ordinates are shown for a concave hexagon and a concave octagon.

5. BAYESIAN THEORY OF PROBABILITY AND ENTROPIC MEASURES

A recent development in the construction of meshfree approximants has been the use of information-theoretic variational principles [11–13]. To provide greater details and insights on the rationale for this approach, we present some of the essential ingredients of Bayesian theory of probability and its ties to inductive inference. In References [11, 12], data approximation is viewed as a problem in inductive inference. Pure mathematics follows the principle of deductive logic—given a cause, many logical consequences can be readily inferred (Figure 7(a)). However, in scientific problems, the reverse is more common: given certain effects or observations, the most likely underlying causes are desired. This requires inductive logic (Figure 7(b)), as in ill-posed inverse problems (e.g. heat conduction, scattering, image reconstruction) that arise in science and engineering [45].

Probability theory as a rational inductive inference procedure was initiated by Bayes and Laplace, and subsequently formalized by Jeffreys [46] and Cox [47]. In information theory [48], the notion of entropy as a measure of uncertainty or incomplete knowledge was introduced by Shannon [14]. Building on these previous contributions, Jaynes [15, 49] proposed the principle of maximum-entropy (MAXENT), in which it was shown that maximizing entropy provides the least-biased statistical inference when insufficient information is available. In References [11, 12], the basis functions $\{\phi_i\}_{i=1}^n$ are viewed as a discrete probability distribution $\{p_i\}_{i=1}^n$, and the polynomial reproducing conditions are the under-determined constraints. To regularize the ill-posed problem, the maximum-entropy principle was used. In this paper, as a generalization, the Shannon–Jaynes entropy functional and the MAXENT or minimum relative entropy principle [16–18] is invoked to obtain the basis functions. Sivia [44] presents an excellent introduction to Bayesian inference and maximum-entropy methods, whereas Jaynes [50] provides a more rigorous and in-depth look at probability theory from the Bayesian perspective.

In Bayesian theory, probability is a subjective measure that represents a degree-of-belief and is always ‘conditional,’ which is contrary to the (objective) frequentist definition. The Bayesian view consists of three stages that are essential to the process of inductive inference [50–52]:

1. *Bayes’s theorem*: If h stands for a hypothesis, d for a set of data, and I for background (testable) information, then Bayes’s theorem states that:

$$\underbrace{p(h | d, I)}_{\text{posterior pdf}} = \underbrace{p(h | I)}_{\text{prior}} \times \underbrace{p(d | h, I)}_{\text{likelihood}} / \underbrace{p(d | I)}_{\text{evidence}} \quad (15)$$

where $p(\cdot)$ is used to denote either the probability (discrete) or the probability density function, pdf (continuous), and in a parameter-estimation problem, the denominator is just a normalizing factor since the posterior pdf must integrate to unity. In essence, Bayes’s theorem is a rule for manipulating probabilities and not for their assignment—the prior probability of h gets updated to the posterior probability as a result of acquiring the data.

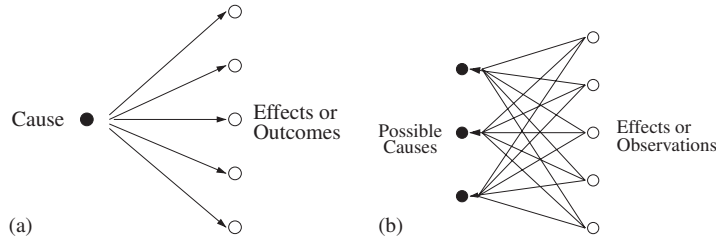


Figure 7. (a) Deductive logic; and (b) inductive logic [44].

2. *Maximum-entropy principle*: In information theory, Shannon introduced the notion of entropy as a measure of uncertainty [14]. The Shannon entropy of a discrete probability distribution is:

$$H(\mathbf{p}) = \mathcal{E}[-\log p] = - \sum_{i=1}^n p_i \ln p_i \tag{16}$$

where $p_i \equiv p(x_i)$ is the probability of the occurrence of the event x_i , $p \ln p \doteq 0$ if $p = 0$, $\mathcal{E}[\cdot]$ is the expectation operator, and the above form of the entropy $H(\cdot)$ satisfies the axiomatic requirements of an uncertainty measure, with (1) $H(\mathbf{p}) \geq 0$; (2) $H(\mathbf{p})$ attains its maximum value when $p_1 = p_2 = \dots = p_n = 1/n$ and it's a monotonic function; and (3) $H(p_1, p_2, \dots, p_n) = H(p_1, p_2, \dots, p_n, 0)$ being the most important properties [53].

Entropy maximization was proposed by Jaynes [15] as a means for least-biased statistical inference when insufficient information is available, and was shown to reproduce equilibrium (Gibbs–Boltzmann) and non-equilibrium distributions in statistical mechanics [16, 50, 54]. It is the only consistent variational principle for the *assignment* of probabilities under a set of constraints (testable information) [16, 18]. Let the available data pertaining to a random variable X consist of the expected value of functions $g_r(x)$ ($r = 0, 1, \dots, m$), with $g_0(x) = 1$ being the normalizing condition. Then, the discrete probabilities are found by solving [15]

$$\max_{\mathbf{p} \in \mathbb{R}_+^n} \left(H(\mathbf{p}) = - \sum_{i=1}^n p_i \ln p_i \right) \tag{17a}$$

$$\sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i g_r(x_i) = \mathcal{E}[g_r(x)] \quad (r = 1, 2, \dots, m) \tag{17b}$$

where \mathbb{R}_+^n is the non-negative orthant. Often, the first $m + 1$ moments of the random variable are available, which leads to the classical maximum-entropy problem of moments [55]. For instance, if the mean μ of a random variable X is known, then the discrete problem is posed as [44, p. 121]

$$\max_{\mathbf{p} \in \mathbb{R}_+^n} \left(H(\mathbf{p}) = - \sum_{i=1}^n p_i \ln p_i \right) \tag{18a}$$

$$\sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i x_i = \mu \tag{18b}$$

which is solved using the method of Lagrange multipliers, and in the continuous case with limits 0 to ∞ , we obtain the exponential distribution [44, p. 121]

$$p(x | \mu) = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right), \quad x \geq 0 \quad (19)$$

If the first moment μ and variance σ^2 of X are known, then in the continuous case with limits $\pm\infty$, we obtain the solution [44, p. 122]

$$p(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (20)$$

which is the Gaussian distribution—a consequence that follows if only the mean (first moment) and variance of data are known. It was recognized that for $H(\cdot)$ to be invariant under the transformation $y = f(x)$ in the continuous case, the general form of the entropy should be [16–18]

$$H(\mathbf{p}, \mathbf{m}) = -\sum_{i=1}^n p_i \ln\left(\frac{p_i}{m_i}\right) \quad \text{or} \quad H(p, m) = -\int p(x) \ln\left(\frac{p(x)}{m(x)}\right) dx \quad (21)$$

where $m(x)(m_i)$ is a prior distribution that estimates $p(x)(p_i)$. In the literature, the quantity $D(\mathbf{p}||\mathbf{m}) = -H(\mathbf{p}, \mathbf{m})$ is also referred to as the Kullback–Leibler distance (directed divergence) [56], and the variational principle is known as the principle of minimum relative (cross) entropy [18]. The relative entropy, $D(\mathbf{p}||\mathbf{m}) \geq 0$, which is proven below.

Proof

If f is a concave function and X a random variable, then by Jensen's inequality (see Reference [48, p. 25])

$$\mathcal{E}[f(X)] \leq f(\mathcal{E}[X]) \quad (22)$$

The Shannon–Jaynes entropy (negative of the relative entropy) functional is

$$-D(\mathbf{p}||\mathbf{m}) = -\sum_{i=1}^n p_i \ln\left(\frac{p_i}{m_i}\right) = \sum_{i=1}^n p_i \ln\left(\frac{m_i}{p_i}\right) \quad (23)$$

On considering the concave function $f(x) = \ln x$ and invoking Jensen's inequality, we can write

$$-D(\mathbf{p}||\mathbf{m}) = \sum_{i=1}^n p_i \ln\left(\frac{m_i}{p_i}\right) \leq \ln\left(\sum_{i=1}^n p_i \frac{m_i}{p_i}\right) = \ln \sum_{i=1}^n m_i = \ln 1 = 0 \quad (24)$$

which completes the proof. \square

Since \ln is a strictly concave function, $D(\mathbf{p}||\mathbf{m})$ attains its minimum value of zero if and only if $\mathbf{p} = \mathbf{m}$. If a uniform prior, $m_i = 1/n$, is used, the principle of minimum relative entropy is identical to the MAXENT principle using Shannon entropy.

3. *Hypothesis space*: The choice of the hypothesis space is the key in any inductive inference problem—this refers to the measure space to define $m(x)$ or m_i when using the maximum-entropy principle or the prior probability $p(h | I)$ in Bayes's theorem. The selection of the prior distribution, $m(x)$, is a key element in the construction of MAXENT approximation schemes, which is discussed in the next section.

5.1. Maximum-entropy approximation schemes

In References [11, 12], the maximum-entropy principle using Shannon entropy and a modified entropy functional, respectively, were used. In this paper, as a unifying framework and generalization, we adopt the Shannon–Jaynes entropy measure, Equation (21), and for consistency, the variational problem is posed as the maximization of the entropy functional, and therefore the dual (unconstrained) problem becomes a convex minimization problem. The parallels between the conditions on ϕ_i in Equations (2) and (3) and those on p_i in a MAXENT formulation are evident. Referring to the nodal sets shown in Figure 1, the basis function value $\phi_i(\mathbf{x})$ is viewed as the ‘probability of influence of a node i at \mathbf{x} ’. The maximum-entropy formulation is: find $\phi(\mathbf{x}) \in \mathbb{R}_+^n$ as the solution of the constrained optimization problem:

$$\max_{\phi \in \mathbb{R}_+^n} \left(H(\phi, \mathbf{m}) = - \sum_{i=1}^n \phi_i(\mathbf{x}) \ln \left(\frac{\phi_i(\mathbf{x})}{m_i(\mathbf{x})} \right) \right) \tag{25a}$$

subject to the linear reproducing conditions given in Equation (2):

$$\sum_{i=1}^n \phi_i(\mathbf{x}) = 1, \quad \sum_{i=1}^n \phi_i(\mathbf{x}) \mathbf{x}_i = \mathbf{x} \tag{25b}$$

where $m_i(\mathbf{x})$ is a prior estimate, and the constraints form an under-determined linear system. Let λ_s ($s = 0, 1, \dots, d$) be the Lagrange multipliers associated with the $d + 1$ constraints. The solution of the variational problem can be written as

$$\phi_i(\mathbf{x}) = \frac{Z_i(\mathbf{x})}{Z(\mathbf{x})}, \quad Z_i(\mathbf{x}) = m_i(\mathbf{x}) \exp(-\mathbf{x}_i^T \boldsymbol{\lambda}(\mathbf{x})) \tag{26}$$

where the maximum-entropy basis functions naturally assume an exponential form, and $Z(\mathbf{x}) = \sum_j Z_j(\mathbf{x})$ is known as the partition function in statistical mechanics. In addition, $\mathbf{x}_i^T = [x_i \ y_i \ z_i]$ and $\boldsymbol{\lambda}(\mathbf{x}) = [\lambda_1(\mathbf{x}) \ \lambda_2(\mathbf{x}) \ \lambda_3(\mathbf{x})]^T$ in three dimensions. We mention in passing that such exponential (Darmois–Koopman–Pitman) family of distributions are well known and widely studied in statistical theory [57] and information geometry [58].

The $\phi_i(\mathbf{x})$ in Equation (26) must satisfy the d linear constraints given in Equation (25b), which yields d nonlinear equations. On considering the dual formulation, a simple unconstrained convex minimization problem is obtained. To this end, we let $\tilde{x}_i = x_i - x$, $\tilde{y}_i = y_i - y$, and $\tilde{z}_i = z_i - z$ (shifted nodal co-ordinates in \mathbb{R}^3), and then redefine Z appropriately. Now, the dual problem is: find $\boldsymbol{\lambda}$ such that [59, 60]

$$\boldsymbol{\lambda} = \arg \min \ln Z(\boldsymbol{\lambda}^t) \tag{27}$$

On using convex duality [19, 60], a detailed mathematical treatment of the primal and dual optimization problems is presented in Reference [12]. Numerical algorithms such as steepest descent, Newton’s method, quasi-Newton (variable metric) methods, and interior-point methods are used to solve such unconstrained optimizations problems [60]. Interior-point methods are attractive for large systems with equality and inequality constraints [61]; for entropy maximization, a `Matlab`TM code, which is based on primal-dual interior method is available in the public-domain [62]. For the JAVA applet, the convex minimization problem is solved using a variable step size gradient descent algorithm [63], with a convergence tolerance $\varepsilon = 10^{-3}$.

In Reference [11], a uniform prior was used, whereas in Reference [12], a variational principle using a modified entropy functional (pareto optima of two objectives) was proposed:

$$\min_{\phi \in \mathbb{R}_+^n} M(\phi, \mathbf{x}) \quad \text{or} \quad \max_{\phi \in \mathbb{R}_+^n} -M(\phi, \mathbf{x}), \quad M(\phi, \mathbf{x}) = \beta U(\phi, \mathbf{x}) - H(\phi) \quad (28a)$$

where $\beta \equiv \beta(\mathbf{x})$ is non-negative, $H(\phi)$ is the Shannon entropy, and $U(\phi, \mathbf{x})$ is the objective function introduced by Rajan [64]:

$$U(\phi, \mathbf{x}) = \sum_{i=1}^n \phi_i(\mathbf{x}) \|\mathbf{x}_i - \mathbf{x}\|^2 \quad (28b)$$

The above functional form draws the connection to statistical mechanics, with the free energy $\mathcal{G} = U - TH$, where U is the internal energy, H is the entropy, and T is the temperature [12, 61]. Rajan's linear programming problem is [64]:

$$\min_{\phi} U(\phi, \mathbf{x}), \quad \phi_i(\mathbf{x}) \geq 0, \quad \sum_{i=1}^n \phi_i(\mathbf{x}) = 1, \quad \sum_{i=1}^n \phi_i(\mathbf{x}) \mathbf{x}_i = \mathbf{x} \quad (29)$$

whose solution is the finite element (Delaunay) interpolant. At $\mathbf{x} = \mathbf{x}_j$, nodal interpolation is realized since the minimum value $U = 0$ is attained if $\phi_i(\mathbf{x}_j) = \delta_{ij}$. When $\beta \rightarrow \infty$ in Equation (28a), Rajan's problem is obtained, and $\beta = 0$ recovers Equation (25a) with a uniform prior. In Sukumar [13], the minimum relative entropy principle was used to unify the above developments. The entropy functional considered by Arroyo and Ortiz [12] is obtained if a Gaussian (RBF) prior, $m_i(\mathbf{x}) = \exp(-\beta \|\mathbf{x}_i - \mathbf{x}\|^2)$, is used in Equation (25a):

$$\begin{aligned} H(\phi, \mathbf{m}) &= - \sum_{i=1}^n \phi_i(\mathbf{x}) \ln \left(\frac{\phi_i(\mathbf{x})}{\exp(-\beta \|\mathbf{x}_i - \mathbf{x}\|^2)} \right) \\ &= - \sum_{i=1}^n \phi_i(\mathbf{x}) \ln \phi_i(\mathbf{x}) - \beta \sum_{i=1}^n \phi_i(\mathbf{x}) \|\mathbf{x}_i - \mathbf{x}\|^2 \\ &= -\beta U(\phi, \mathbf{x}) + H(\phi) \end{aligned} \quad (30)$$

which is identical to Equation (28a). The parameter β in the Gaussian distribution is inversely proportional to the variance; it determines the support-width of the basis function [12].

The choice of the prior, $m_i(\mathbf{x})$, gives us greater flexibility in the construction of new approximants, and provides a simple and appealing means to construct globally or compactly supported convex approximation schemes. In the spirit of previous research on meshfree methods [26, 27, 65] and partition of unity methods [66], given a prior $m_i(\mathbf{x})$ and a set of linear constraints (reproducing conditions), entropy maximization can be viewed as a 'correction' to obtain an approximation with polynomial and/or non-polynomial reproducibility. The use of the Shannon or relative entropy functional provides a means to obtain the least-biased statistical inference solution. With Shannon entropy, the *flattest possible distribution* that is consistent with the constraints is realized. The maximum-entropy formulation leads to a convex optimization problem, with the approximant possessing many desirable properties for the Galerkin solution of PDEs [12]. The continuity of maximum-entropy basis functions with a Gaussian prior is established in Arroyo and Ortiz [12], and in Sukumar and Wets [67], variational analysis and the theory of epi-convergence [68] is used

to prove the same for any prior distribution. The above properties are lost if other functionals are adopted—for example, in Sukumar [11] it is shown that if the minimum-norm objective functional (leads to the generalized- or pseudo-inverse [69]) is used, then $\phi_i(\mathbf{x}) < 0$ is also admissible and interpolation on the boundary is not realized. If the minimum-norm objective is adopted with the non-negative condition, $\phi_i(\mathbf{x}) \geq 0$, as additional constraints, then a convex approximant is obtained; however, numerical tests reveal that the basis functions are continuous but not continuously differentiable in the interior of \mathcal{C} .

1. *Uniform prior:* For a uniform prior, $m_i(\mathbf{x}) = 1/n$, and as indicated earlier, the Shannon–Jaynes entropy functional is the same as the Shannon entropy (modulo a constant). For this case, the maximum-entropy basis functions are identical to bilinear finite element basis functions on a square, and are smooth and bounded in \mathcal{C} [11]. To illustrate a simple closed-form computation, consider one-dimensional approximation in $\mathcal{C} = [0, 1]$ with three nodes located at $x_1 = 0$, $x_2 = \frac{1}{2}$, and $x_3 = 1$. On using Equation (25), the solution for $\phi_i(x)$ is readily derived [40]:

$$\phi_1(x) = \frac{1}{Z}, \quad \phi_2(x) = \frac{\eta}{Z}, \quad \phi_3(x) = \frac{\eta^2}{Z}, \quad \eta \equiv \eta(x) = \frac{2x - 1 + \sqrt{12x(1-x) + 1}}{4(1-x)} \quad (31)$$

where $Z = 1 + \eta + \eta^2$.

2. *Non-uniform prior:* Instead of a uniform prior, a non-uniform prior for node i gives more weight to \mathbf{x}_i than to other nodal locations. Now, different choices of the prior $m_i(\mathbf{x})$ can be used in the Shannon–Jaynes entropy functional:

- The prior can be selected to be global radial basis functions such as the Gaussian, $m_i(\mathbf{x}) = \exp(-\|\mathbf{x}_i - \mathbf{x}\|^2/c^2)$, inverse multiquadrics, $m_i(\mathbf{x}) = (\|\mathbf{x}_i - \mathbf{x}\|^2 + c^2)^{-1/2}$, etc.
- On choosing a weight function, $w(\mathbf{x})$, with compact support, we set $m_i(\mathbf{x}) = w_i(\mathbf{x})$ as the prior for node i , where $w_i(\mathbf{x})$ is a translation and scaling of $w(\mathbf{x})$. If the only constraint is: $\sum_i \phi_i(\mathbf{x}) = 1$, then the maximum-entropy basis functions are: $\phi_i(\mathbf{x}) = w_i(\mathbf{x}) / \sum_j w_j(\mathbf{x})$, which is the well-known Shepard function [70]. If $w_i(\mathbf{x})$ is constructed using the C^2 cubic spline weight function given in Equation (11), then unlike the MLS approximant, a convex approximant with desirable properties on the boundary is obtained. Other choices for $m_i(\mathbf{x})$ include compactly supported RBFs, for example the C^2 function $m(r) = (1-r)_+^4(4r+1)$ [10], where $(\cdot)_+ = (\cdot)$ if the argument is non-negative and zero otherwise.
- As alternative compactly supported priors, R -functions [71, 72] or implicit (level set) functions that are defined on a graph are also suitable.

Maximum-entropy basis functions with a uniform prior in $\mathcal{C} = [0, 1]$ are depicted in Figure 8. For the plots in Figure 8(a), the closed-form expressions for $\phi_i(x)$ are given in Equation (31). Nodal interpolation is met on bdry \mathcal{C} but not at the interior nodes. In Figure 9, the maximum-entropy basis functions for nodes 1 and 6 in Figure 1(d) are illustrated. We note that $\phi_1(\mathbf{x})$ is unity at \mathbf{x}_1 and is piece-wise linear on the boundary, whereas $\phi_6(\mathbf{x}_6) \neq 1$ and $\phi_6(\mathbf{x})$ vanishes on the boundary of the square. Laplace and MAXENT basis functions on a weakly convex polygon are shown in Figure 10. Along the edge 1–2, Laplace basis functions satisfy the Kronecker-delta property but the maximum-entropy basis functions do not; however, $\phi_i(\mathbf{x}) = 0$ ($i = 3-5$) along edge 1–2 for both Laplace and MAXENT basis functions.

To demonstrate the properties of convex approximants with a non-uniform prior, we first consider a one-dimensional grid. In Figure 11, basis function plots using MLS, and MAXENT with a compactly supported prior are presented. The cubic spline weight function given in Equation (11)

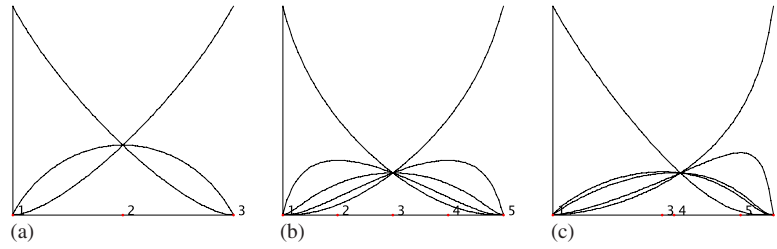


Figure 8. One-dimensional maximum-entropy basis functions with a uniform prior: (a) regular grid ($n = 3$); (b) regular grid ($n = 5$); and (c) random grid ($n = 5$).

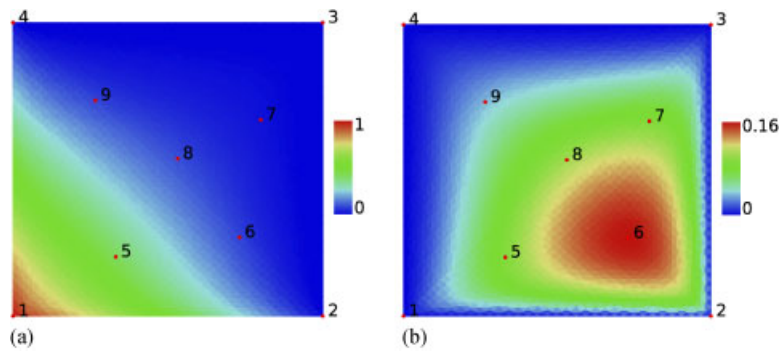


Figure 9. Two-dimensional maximum-entropy basis functions with a uniform prior: (a) $\phi_1(\mathbf{x})$; and (b) $\phi_6(\mathbf{x})$.

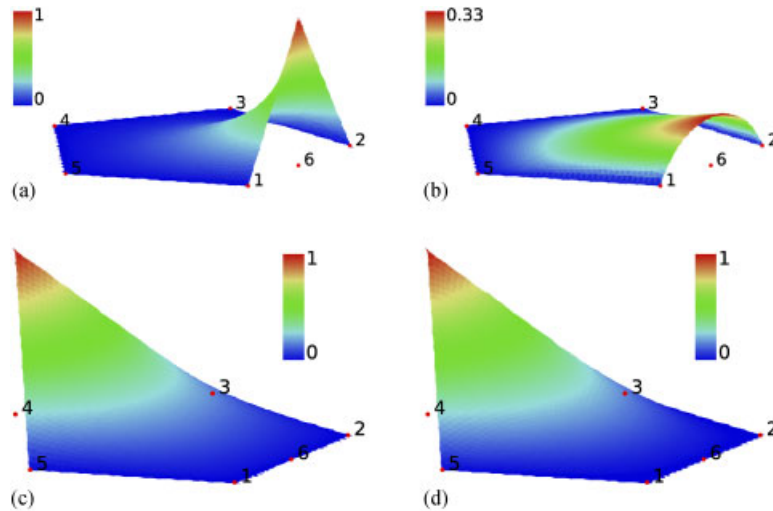


Figure 10. Basis functions on a weakly convex pentagon: (a) Laplace (node 6); (b) maximum-entropy (node 6); (c) Laplace (node 4); and (d) maximum-entropy (node 4).

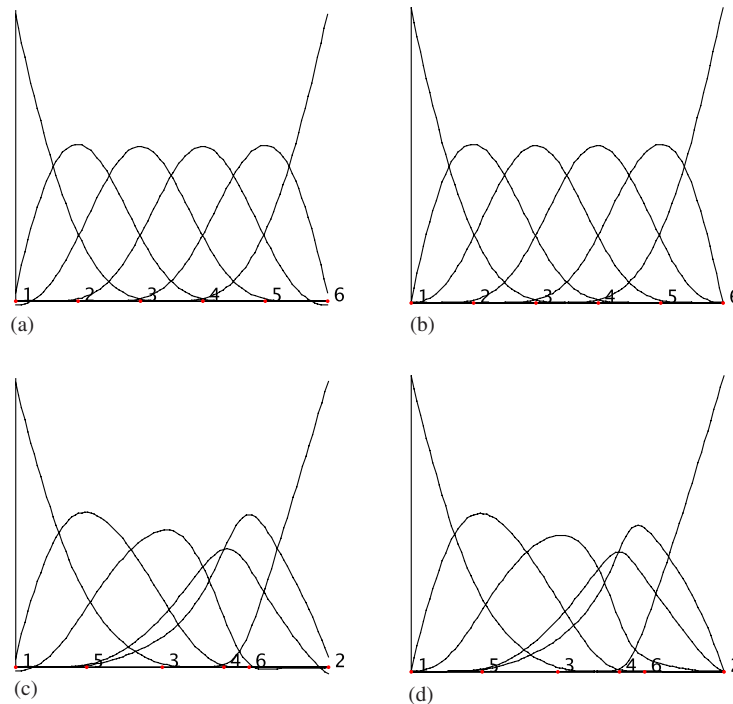


Figure 11. One-dimensional basis functions: (a,b) uniform grid with MLS ($\alpha=2.5$) and maximum-entropy with a cubic spline prior ($\alpha=2.5$); and (c,d) random grid with MLS ($\alpha=2.5$) and maximum-entropy with a cubic spline prior ($\alpha=2.5$).

is used as the compactly supported prior. We observe that interior MLS basis functions have a non-zero contribution on the boundary (Figures 11(a) and (c)), whereas boundary MAXENT basis functions with a cubic spline prior (Figures 11(b) and (d)) satisfy the Kronecker-delta property. Next, we consider the two-dimensional grid shown in Figure 1(d), and study the MAXENT plots using a Gaussian prior, when β is varied; see Reference [12] for its applications in nonlinear solid mechanics. For $\beta=0, 1, 10, 100$, the MAXENT basis function plots for node 8 are presented in Figure 12. The value $\beta=0$ corresponds to a uniform prior. It's observed that as β is increased the nodal basis function support shrinks, and when $\beta=100$ (theoretically when $\beta \rightarrow \infty$) the basis function support is proximal to the triangular (Delaunay) basis function (Figure 12(d)). In Figure 13, comparisons between the MLS basis function and the MAXENT basis function using the compactly supported cubic spline prior are presented. The interior MLS basis function is non-zero on bndry \mathcal{C} (Figure 13(c)), whereas the interior MAXENT basis function vanishes on the boundary of the square (Figure 13(d)).

As of this writing, the three-dimensional applet is somewhat less developed than the other two and is restricted to maximum-entropy basis function plots with a uniform prior. For Figure 14, a regular tetrahedron is created, along with one or two interior nodes. Basis functions are plotted along planes that cut the convex hull.

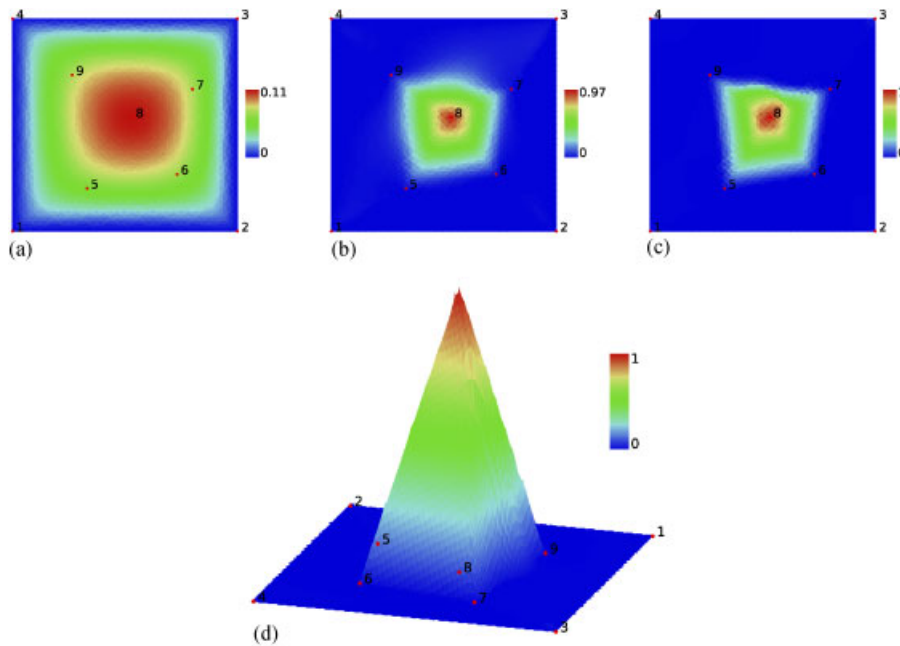


Figure 12. Maximum-entropy basis function, $\phi_8(\mathbf{x})$, with a Gaussian prior: (a) $\beta = 0$; (b) $\beta = 1$; (c) $\beta = 10$; and (d) $\beta = 100$.

5.2. Higher-order approximation schemes

In References [11, 12], linearly complete approximations were constructed using the maximum-entropy principle. Furthermore, in Reference [12], it was shown that the additional constraint $\sum_i \phi_i(x)x_i^2 = x^2 + c$ in one-dimension with $c = 0$ does not yield a feasible solution if $\phi_i \geq 0$. On choosing $c \neq 0$, non-negative $\phi_i(x)$ can be obtained, which bear resemblance to univariate *B*-splines [12]. Alternatively, the non-negative condition, $\phi_i(\mathbf{x}) \geq 0$, can be relaxed to obtain an appropriate ‘entropy functional’ that can be maximized. To this end, we start with the generalization of the Shannon–Jaynes entropy that was proposed by Skilling [73]:

$$H(\phi, \mathbf{m}) = \sum_{i=1}^n \left[\phi_i - m_i - \phi_i \ln \left(\frac{\phi_i}{m_i} \right) \right] \tag{32}$$

where $\phi_i \equiv \phi_i(\mathbf{x})$ and the prior estimate $m_i \equiv m_i(\mathbf{x})$ need not be normalized so that applicability is extended to physical distributions other than probabilities. In the absence of any constraints, $H(\cdot, \cdot)$ is maximized when $\phi_i = m_i$ ($i = 1, 2, \dots, n$). On using the above expression, distributions with positive and negative values (signed basis functions) are obtained.

Let $\phi_i = v_i - w_i$, where $v_i \in \mathbb{R}_+$ and $w_i \in \mathbb{R}_+$, so that $\phi_i \in \mathbb{R}$. Also, let m_i^v and m_i^w be the prior estimate for v_i and w_i , respectively. The total entropy is

$$H(\mathbf{v}, \mathbf{w}, \mathbf{m}^v, \mathbf{m}^w) = \sum_{i=1}^n \left[v_i - m_i^v - v_i \ln \left(\frac{v_i}{m_i^v} \right) \right] + \sum_{i=1}^n \left[w_i - m_i^w - w_i \ln \left(\frac{w_i}{m_i^w} \right) \right] \tag{33}$$

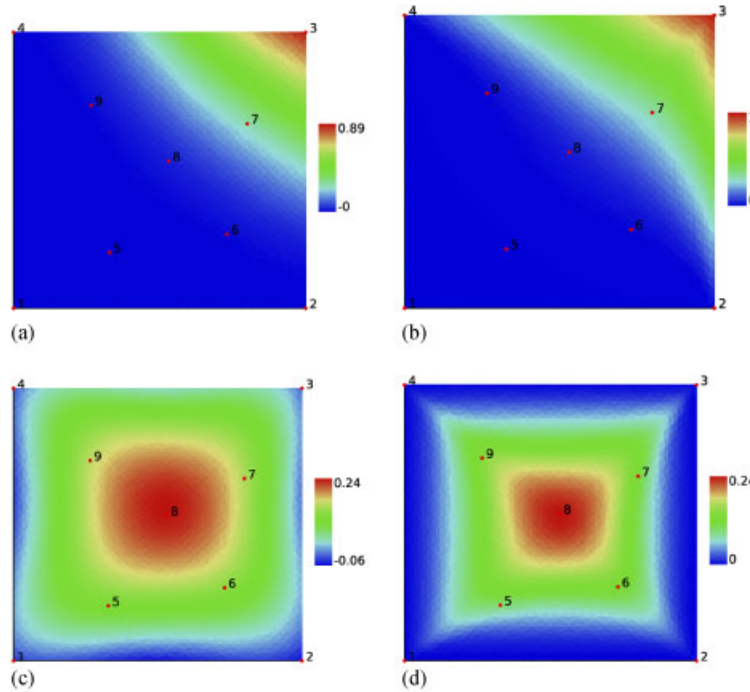


Figure 13. Basis functions: $\phi_3(\mathbf{x})$ with: (a) MLS; (b) maximum-entropy with the cubic spline prior; and $\phi_8(\mathbf{x})$ with (c) MLS; and (d) maximum-entropy with the cubic spline prior.

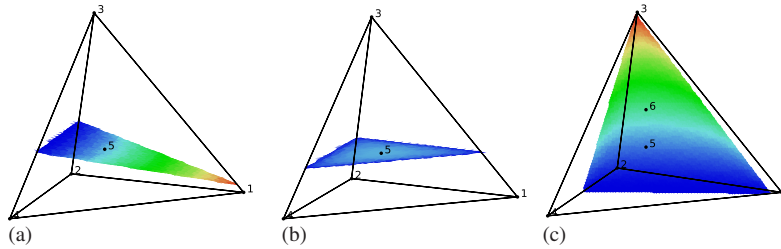


Figure 14. Three-dimensional maximum-entropy basis functions plots: (a) $\phi_1(\mathbf{x})$ along a plane near nodes 1 and 5; (b) $\phi_5(\mathbf{x})$ along a plane containing node 5; and (c) $\phi_3(\mathbf{x})$ along a plane near nodes 1, 3, 5, and 6.

An expression for H in which only ϕ and \mathbf{m} appears is desired. Since $\phi_i = v_i - w_i$, we have

$$\frac{\partial H}{\partial v_i} = \frac{\partial H}{\partial \phi_i} \frac{\partial \phi_i}{\partial v_i} = \frac{\partial H}{\partial \phi_i}, \quad \frac{\partial H}{\partial w_i} = \frac{\partial H}{\partial \phi_i} \frac{\partial \phi_i}{\partial w_i} = -\frac{\partial H}{\partial \phi_i} \tag{34}$$

and therefore

$$\frac{\partial H}{\partial v_i} + \frac{\partial H}{\partial w_i} = 0 \tag{35}$$

On using Equation (33) and the above relation, we obtain

$$v_i w_i = m_i^v m_i^w \quad (36)$$

If $v_i = (\psi_i + \phi_i)/2$ and $w_i = (\psi_i - \phi_i)/2$, then $\psi_i = \sqrt{\phi_i^2 + 4m_i^v m_i^w}$. Finally, the entropy expression for the positive/negative distribution ϕ is [74, 75]:

$$H(\phi, \mathbf{m}^v, \mathbf{m}^w) = \sum_{i=1}^n \left[\psi_i - m_i^v - m_i^w - \phi_i \ln \left(\frac{\psi_i + \phi_i}{2m_i^v} \right) \right] \quad (37)$$

Now, when no constraints are imposed, H is maximized when $\phi_i = m_i^v - m_i^w$. For the data approximation problem, we choose $m_i^v = 2m_i$ and $m_i^w = m_i$, where $m_i \equiv m_i(\mathbf{x})$ is a non-negative weight function. The expression for the entropy becomes

$$H(\phi, \mathbf{m}) = \sum_{i=1}^n \left[\psi_i - 3m_i - \phi_i \ln \left(\frac{\psi_i + \phi_i}{4m_i} \right) \right] \quad (38)$$

where $\psi_i = \sqrt{\phi_i^2 + 8m_i^2}$. On using the above form of H within the maximum-entropy variational principle, signed basis functions with higher-order completeness are constructed.

The implementation of the signed maximum-entropy approximant has been carried in `Matlab`TM [62]. In Figure 15, the MAXENT basis functions using a Gaussian prior weight function are shown. The domain is $\mathcal{C} = [0, 1]$, which is discretized by five equi-spaced nodes. In Figure 15(a)–(c), quadratically complete basis functions are depicted for varying values of β , whereas in Figure 15(d)–(f), basis functions with cubic complete basis functions are plotted for different β . The plots in Figure 15(g)–(i) are for basis functions that can reproduce $\{1, x, f(x)\}$, where $f(x) = \exp(-(x - 0.5)^2)$. In all cases, as β is increased, the basis functions are less negative and are also closer to being an interpolant on the boundary.

6. CONCLUDING REMARKS

In this paper, we presented an overview and recent advances in the construction of meshfree approximation schemes. Meshfree basis functions such as MLS approximants, natural neighbour-based polygonal interpolants, and maximum-entropy (MAXENT) approximants were considered. The construction and applications of MAXENT approximants have recently come to the forefront [11–13], and hence greater emphasis was placed on the theoretical underpinnings of Bayesian theory of probability, maximum-entropy principle [14, 15], and its numerical solution. We used the Shannon–Jaynes entropy functional or relative entropy [17, 18] within the variational formulation to generalize the construction of MAXENT approximants. The merits of constructing basis functions using the maximum-entropy variational principle were examined, and the extension of Shannon–Jaynes entropy to physical distributions other than probabilities [73] was used to construct higher-order maximum-entropy basis functions. The use of maximum-entropy approximation schemes in higher-dimensional parameter spaces is also promising [61]. A JAVA applet was developed,[‡] and

[‡]Access to the applet will be made available through the first author's web page.

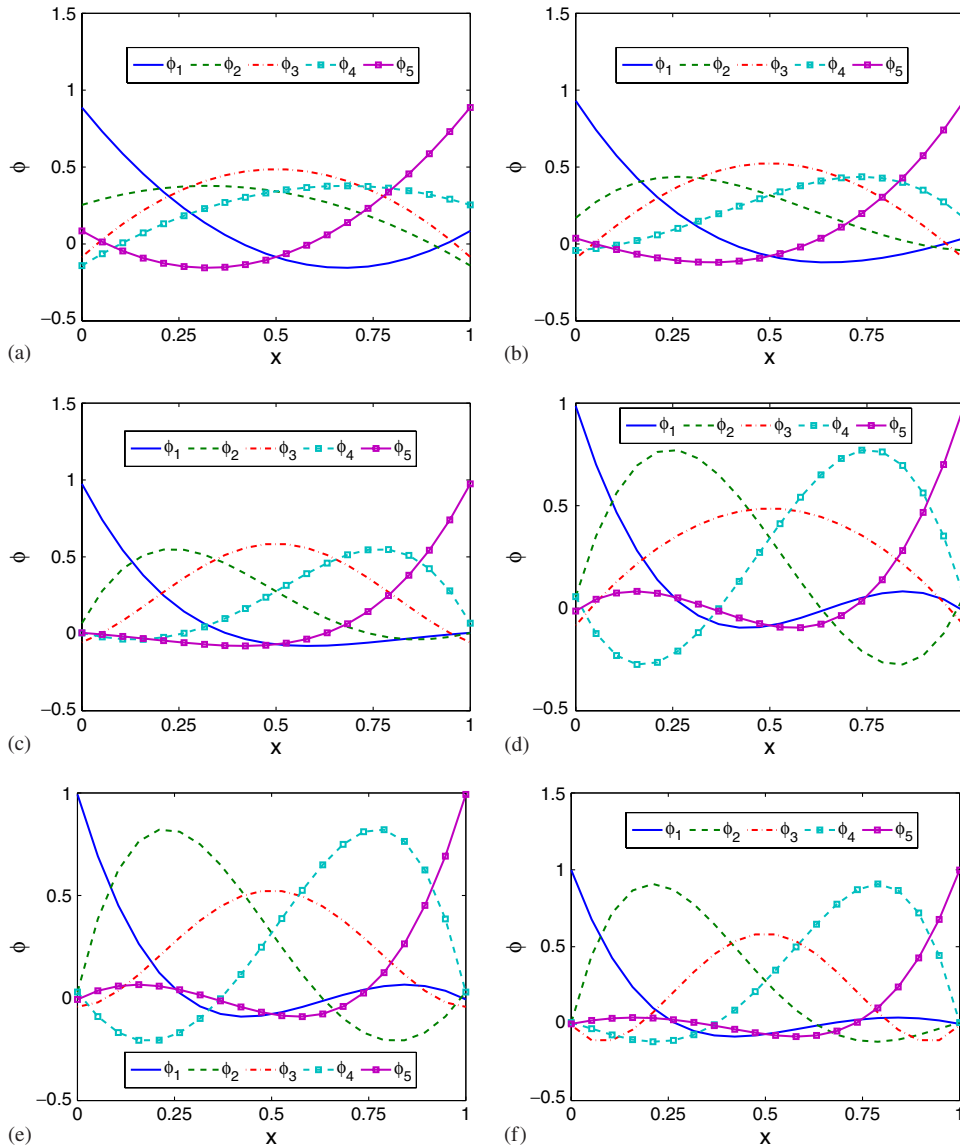
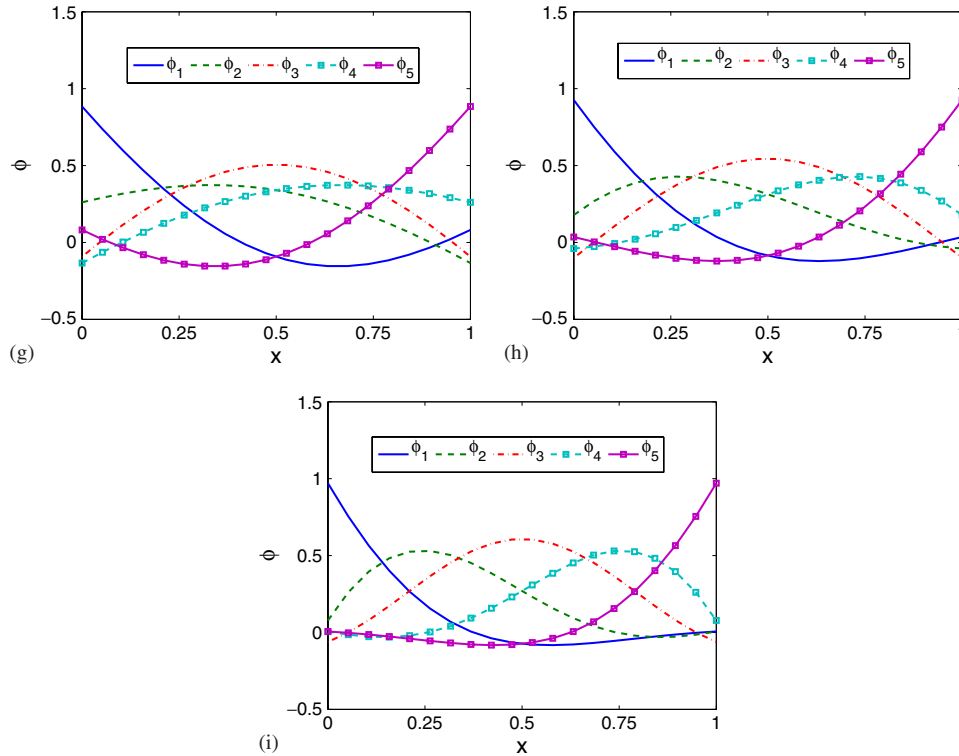


Figure 15. Higher-order basis functions: (a)–(c) quadratic completeness using a Gaussian prior with $\beta = 0, 2, 5$; (d)–(f) cubic completeness using a Gaussian prior with $\beta = 0, 2, 5$; and (g)–(i) reproducing the functions $\{1, x, \exp(-(x - 0.5)^2)\}$ using a Gaussian prior with $\beta = 0, 2, 5$.

basis function plots were presented to reveal the similarities and distinctions between different meshfree approximants. The maximum-entropy formulation with a non-uniform prior provides a simple and elegant means to directly impose linear essential boundary conditions in meshfree methods. With the development of stable nodal integration schemes for meshfree Galerkin methods,

Figure 15. *Continued.*

background cells would no longer be needed for numerical integration. This advance would pave the way towards the conception of stable meshfree particle methods, which are particularly attractive for the solution of problems that arise in nonlinear solid mechanics.

ACKNOWLEDGEMENTS

The authors are grateful for the research support of the National Science Foundation through contract CMS-0626481 to the University of California, Davis. The financial support to RWW through a NSF VIGRE graduate trainee award (NSF Grant DMS-0135345) is also acknowledged. NS thanks Roger Wets and Michael Saunders for many helpful discussions.

REFERENCES

1. Meyer M, Lee H, Barr A, Desbrun M. Generalized barycentric coordinates on irregular polygons. *Journal of Graphics Tools* 2002; **7**(1):13–22.
2. Floater MS. Mean value coordinates. *Computer Aided Geometric Design* 2003; **20**(1):19–27.
3. Floater MS, Hormann K. Surface parameterization: a tutorial and survey. In *Advances in Multiresolution for Geometric Modelling*, Dodgson NA, Floater MS, Sabin MA (eds), Mathematics and Visualization. Springer: Berlin, Heidelberg, 2005; 157–186.
4. Sukumar N, Tabarraei A. Conforming polygonal finite elements. *International Journal for Numerical Methods in Engineering* 2004; **61**(12):2045–2066.

5. Lancaster P, Salkauskas K. Surfaces generated by moving least squares methods. *Mathematics of Computation* 1981; **37**:141–158.
6. Sibson R. A vector identity for the Dirichlet tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society* 1980; **87**:151–155.
7. Christ NH, Friedberg R, Lee TD. Weights of links and plaquettes in a random lattice. *Nuclear Physics B* 1982; **210**(3):337–346.
8. Hardy RL. Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research* 1971; **76**:1905–1915.
9. Buhmann MD. *Radial Basis Functions: Theory and Implementations*. Cambridge University Press: Cambridge, U.K., 2003.
10. Wendland H. *Scattered Data Approximation*. Cambridge University Press: Cambridge, U.K., 2005.
11. Sukumar N. Construction of polygonal interpolants: a maximum entropy approach. *International Journal for Numerical Methods in Engineering* 2004; **61**(12):2159–2181.
12. Arroyo M, Ortiz M. Local maximum-entropy approximation schemes: a seamless bridge between finite elements and meshfree methods. *International Journal for Numerical Methods in Engineering* 2006; **65**(13):2167–2202.
13. Sukumar N. Maximum entropy approximation. *AIP Conference Proceedings* 2005; **803**(1):337–344.
14. Shannon CE. A mathematical theory of communication. *The Bell Systems Technical Journal* 1948; **27**:379–423.
15. Jaynes ET. Information theory and statistical mechanics. *Physical Review* 1957; **106**(4):620–630.
16. Jaynes ET. Information theory and statistical mechanics. In *Statistical Physics: The 1962 Brandeis Lectures*, Ford K (ed.). W. A. Benjamin: New York, 1963; 181–218.
17. Kullback S. *Information Theory and Statistics*. Wiley: New York, NY, 1959.
18. Shore JE, Johnson RW. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory* 1980; **26**(1):26–36.
19. Rockafellar RT. *Convex Analysis*. Princeton University Press: Princeton, NJ, 1970.
20. Strang G, Fix G. *An Analysis of the Finite Element Method*. Prentice-Hall: Englewood Cliffs, NJ, 1973.
21. Epperson JF. On the Runge example. *The American Mathematical Monthly* 1987; **94**(4):329–341.
22. Farouki RT, Goodman TNT. On the optimal stability of the Bernstein basis. *Mathematics of Computation* 1996; **65**(216):1553–1566.
23. Farouki RT. Private communication, 2006.
24. Peña JM. B-splines and optimal stability. *Mathematics of Computation* 1997; **66**(220):1555–1560.
25. Hughes TJR, Cottrell JA, Bazilevs Y. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering* 2005; **193**(39–41):4135–4195.
26. Belytschko T, Lu YY, Gu L. Element-free Galerkin methods. *International Journal for Numerical Methods in Engineering* 1994; **37**:229–256.
27. Liu WK, Jun S, Zhang YF. Reproducing kernel particle methods. *International Journal for Numerical Methods in Engineering* 1995; **20**:1081–1106.
28. Belytschko T, Krongauz Y, Organ D, Fleming M, Krysl P. Meshless methods: an overview and recent developments. *Computer Methods in Applied Mechanics and Engineering* 1996; **139**:3–47.
29. Li S, Liu WK. Meshfree and particle methods and their applications. *Applied Mechanics Review* 2002; **55**(1):1–34.
30. Huerta A, Belytschko T, Fernández-Méndez S, Rabczuk T. Meshfree methods. In *Encyclopedia of Computational Mechanics*, Stein E, de Borst R, Hughes TJR (eds), vol. 1. Wiley: Chichester, 2004; 279–309 (Chapter 10).
31. Fries TP, Matthies HG. Classification and overview of meshfree methods. *Technical Report Informatikbericht-Nr. 2003-03*, Institute of Scientific Computing, Technical University Braunschweig, Braunschweig, Germany, 2004.
32. Atluri SN, Shen S. *The Meshless Local Petrov–Galerkin (MLPG) Method*. Tech Science Press: Encino, CA, 2002.
33. Liu GR. *Mesh Free Methods: Moving Beyond the Finite Element Method*. CRC Press: Boca Raton, FL, 2003.
34. Li S, Liu WK. *Meshfree Particle Methods*. Springer: New York, NY, 2004.
35. Dolbow J, Belytschko T. An introduction to programming the meshless Element Free Galerkin method. *Archives of Computational Methods in Engineering* 1998; **5**(3):207–241.
36. Du Q, Gunzburger M, Ju L. Meshfree, probabilistic determination of point sets and support regions for meshless computing. *Computer Methods in Applied Mechanics and Engineering* 2002; **191**:1349–1366.
37. Kansa EJ. Multiquadrics—a scattered data approximation scheme for applications to computational fluid-dynamics. 1. Surface approximations and partial derivative estimates. *Computers and Mathematics with Applications* 1990; **19**(8/9):127–145.

38. Kansa EJ. Multiquadrics—a scattered data approximation scheme for applications to computational fluid-dynamics. 2. Solutions to parabolic, hyperbolic and elliptic partial-differential equations. *Computers and Mathematics with Applications* 1990; **19**(8/9):147–161.
39. Wachspress EL. *A Rational Finite Element Basis*. Academic Press: New York, NY, 1975.
40. Sukumar N, Malsch EA. Recent advances in the construction of polygonal finite element interpolants. *Archives of Computational Methods in Engineering* 2006; **13**(1):129–163.
41. Hormann K. Barycentric coordinates for arbitrary polygons in the plane. *Technical Report IfI-05-05*, Department of Informatics, Clausthal University of Technology, February 2005.
42. Cueto E, Sukumar N, Calvo B, Martínez MA, Goñiño J, Doblaré M. Overview and recent advances in natural neighbour Galerkin methods. *Archives of Computational Methods in Engineering* 2003; **10**(4):307–384.
43. Rashid MM, Selimotic M. A three-dimensional finite element method with arbitrary polyhedral elements. *International Journal for Numerical Methods in Engineering* 2006; **67**(2):226–252.
44. Sivia DS. *Data Analysis: A Bayesian Tutorial*. Oxford University Press: Oxford, 1996.
45. Kapur JN. *Maximum-Entropy Models in Science and Engineering*. (1st rev. edn). Wiley: New Delhi, India, 1993.
46. Jeffreys H. *Theory of Probability*. Clarendon Press: Oxford, 1939.
47. Cox RT. Probability, frequency and reasonable expectation. *American Journal of Physics* 1946; **14**:1–13.
48. Cover TM, Thomas JA. *Elements of Information Theory*. Wiley: New York, NY, 1991.
49. Jaynes ET. Information theory and statistical mechanics. II. *Physical Review* 1957; **108**(2):171–190.
50. Jaynes ET. *Probability Theory: The Logic of Science*. Cambridge University Press: Cambridge, U.K., 2003.
51. Jaynes ET. The relation of Bayesian and maximum entropy methods. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Erickson GJ, Smith CR (eds), Foundations, vol. 1. Dordrecht, The Netherlands, 1988; 25–29.
52. Gull SF. Bayesian inductive inference and maximum entropy. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Erickson GJ, Smith CR (eds), Foundations, vol. 1. Dordrecht, The Netherlands, 1988; 53–74.
53. Khinchin A. *Mathematical Foundations of Information Theory*. Dover: New York, NY, 1957.
54. Rosenkrantz RD (ed.). *E. T. Jaynes: Paper on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
55. Mead LR, Papanicolaou N. Maximum-entropy in the problem of moments. *Journal of Mathematical Physics* 1984; **25**(8):2404–2417.
56. Kullback S, Leibler RA. On information and sufficiency. *Annals of Mathematical Statistics* 1951; **22**(1):79–86.
57. Barndorff-Nielsen O. *Information and Exponential Families in Statistical Theory*. Wiley: New York, NY, 1978.
58. Amari S, Nagaoka H. *Methods of Information Geometry*. Oxford University Press: New York, NY, 2000.
59. Agmon N, Alhassid Y, Levine RD. An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics* 1979; **30**:250–258.
60. Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge University Press: Cambridge, U.K., 2004.
61. Gupta MR. An information theory approach to supervised learning. *Ph.D. Thesis*, Department of Electrical Engineering, Stanford University, Palo Alto, CA, U.S.A., March 2003.
62. Saunders MA. PDCO: primal-dual method for convex objectives. Available at <http://www.stanford.edu/group/SOL/software/pdco.html>, Department of Management Science and Engineering, Stanford University, Stanford, CA, 2002.
63. Burden RL, Faires JD. *Numerical Analysis* (10th edn). Thomson/Brooks/Cole: Belmont, CA, 2005.
64. Rajan VT. Optimality by the Delaunay triangulation in R^d . *Discrete and Computational Geometry* 1994; **12**(2):189–202.
65. Krongauz Y, Belytschko T. Consistent pseudo-derivatives in meshless methods. *Computer Methods in Applied Mechanics and Engineering* 1997; **146**(3–4):371–396.
66. Babuška I, Melenk JM. The partition of unity method. *International Journal for Numerical Methods in Engineering* 1997; **40**:727–758.
67. Sukumar N, Wets RJ-B. Deriving the continuity of maximum-entropy basis functions via variational analysis. 2006, submitted.
68. Rockafellar RT, Wets RJ-B. *Variational Analysis* (2nd edn). Springer: Berlin, 2004.
69. Penrose R. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society* 1955; **51**:406–413.
70. Shepard D. A two-dimensional interpolation function for irregularly spaced points. *ACM National Conference*, 1968; 517–524.

71. Shapiro V. Theory of R -functions and applications: a primer. *Technical Report CPA88-3*, Cornell Programmable Automation, Sibley School of Mechanical Engineering, Ithaca, NY 14853, 1991.
72. Rvachev VL, Sheiko TI, Shapiro V, Tsukanov I. On completeness of RFM solution structures. *Computational Mechanics* 2000; **25**(2–3):305–316.
73. Skilling J. The axioms of maximum entropy. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Erickson GJ, Smith CR (eds), Foundations, vol. 1. Dordrecht, The Netherlands, 1988; 173–187.
74. Gull SF, Skilling J. The MEMSYS5 users' manual. *Technical Report*, Maximum Entropy Data Consultants Ltd., Suffolk, U.K., 1990.
75. Hobson MP, Lasenby AN. The entropic prior for distributions with positive and negative values. *Monthly Notices of the Royal Astronomical Society* 1998; **298**(3):905–908.