

## Construction of polygonal interpolants: a maximum entropy approach

N. Sukumar\*,<sup>†</sup>

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

### SUMMARY

In this paper, we establish a link between maximizing (information-theoretic) entropy and the construction of polygonal interpolants. The determination of shape functions on  $n$ -gons ( $n > 3$ ) leads to a non-unique under-determined system of linear equations. The barycentric co-ordinates  $\phi_i$ , which form a partition of unity, are associated with discrete probability measures, and the linear reproducing conditions are the counterpart of the expectations of a linear function. The  $\phi_i$  are computed by maximizing the uncertainty  $H(\phi_1, \phi_2, \dots, \phi_n) = -\sum_{i=1}^n \phi_i \log \phi_i$ , subject to the above constraints. The description is expository in nature, and the numerical results via the maximum entropy (MAXENT) formulation are compared to those obtained from a few distinct polygonal interpolants. The maximum entropy formulation leads to a feasible solution for  $\phi_i$  in any convex or non-convex polygon. This study is an instance of the application of the maximum entropy principle, wherein least-biased inference is made on the basis of incomplete information. Copyright © 2004 John Wiley & Sons, Ltd.

**KEY WORDS:** Shannon entropy; information theory; barycentric co-ordinates; natural neighbours; Laplace interpolant; meshfree interpolant; data interpolation

### 1. INTRODUCTION

In information theory, the notion of entropy as a measure of uncertainty or incomplete knowledge was introduced by Shannon [1]. The quantity  $H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^n p_i \log p_i$  was referred to as the informational entropy of an ensemble, where  $p_i$  is the probability associated with the occurrence of the event  $x_i$ , subject to the condition  $\sum_{i=1}^n p_i = 1$ . Building on Shannon's work, Jaynes [2, 3] proposed the principle of maximum entropy (MAXENT), in which through its bearing on well-known ensemble distributions in statistical mechanics, it was shown that maximizing  $H$  was the least-biased means to obtain rational (statistical) inference

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

<sup>†</sup>E-mail: nsukumar@ucdavis.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: contract CMS-0352654

*Received 23 January 2004*

*Revised 3 June 2004*

*Accepted 9 July 2004*

in the face of insufficient data. Jaynes [4] presents a concise and transparent explanation of the MAXENT formalism (Jaynes's principle). We point out that the meaning of *entropy* within the context of MAXENT is to be viewed in the information-theoretic sense, as opposed to the concept of entropy which arises in thermodynamics and statistical physics, where it is a measure of the equilibria and evolution of thermodynamic systems. The application of the principle of maximum entropy has touched many disciplines, ranging from economics to statistical mechanics, image enhancement to nuclear physics, and as diverse as urban planning and biology. The collection of Jaynes's papers in Reference [5] provide a chronological sequence of progress in the evolution of maximum entropy methods, whereas Khinchin [6] presents the mathematical roots and basis of information theory. For a description of the various applications of MAXENT in science and engineering, the interested reader can refer to Kapur [7].

In this paper, we view the data interpolation problem in a new light, and apply the maximum entropy principle to obtain least-biased interpolants on polygonal domains. To the author's knowledge, such an approach has not been pursued previously. The only study that does adopt the principle of maximum entropy with finite elements is the work of Beltzer [8]. In Reference [8], the MAXENT principle was used to characterize finite elements in terms of its complexity. Even though the use of the maximum entropy principle has primarily been applied to problems associated with *randomness*, such a notion is not necessarily required [4]. The MAXENT principle is equally applicable to any inferential problem with incomplete data, in which the minimally prejudiced (maximally non-committal) solution is sought. This is the viewpoint adopted in this study. The author's recent research on the development of a natural neighbour-based polygonal finite element method [9] and its use to construct a conforming  $h$ -adaptive finite element method on tree-based meshes [10] provided impetus to pursue the present investigation.

The outline of this paper is as follows. In Section 2, the maximum entropy principle for the polygonal interpolant problem is presented, and in Section 3 its solution is obtained by using the method of Lagrange multipliers combined with a numerical scheme in which the minimizer of a convex function is sought. In Section 4, numerical results using the MAXENT principle are presented for both, convex and non-convex polygons, and comparisons are made to results obtained from a few recently developed polygonal interpolants (see Reference [9] for details). The main results and conclusions from this study are indicated in Section 5.

## 2. PROBLEM STATEMENT AND FORMULATION

Consider a microscopic ensemble in which the sample space  $\Omega$  consists of mutually independent discrete events  $x_1, x_2, \dots, x_n$  that occur with probabilities  $p_1, p_2, \dots, p_n$ , respectively. Since in a random experiment,  $P(\Omega) = 1$ , it follows that the non-negative probabilities  $p_i$  must satisfy the condition  $\sum_i p_i = 1$ . Now, consider  $m$  ( $m < n - 1$ ) macroscopic observations, in which the expected value of a function  $g_r(x)$  (denoted by  $\langle g_r(x) \rangle$ ) is known. Then, the most likely probability distribution  $p_i$  is obtained by maximizing the informational entropy  $H(\cdot)$  [1, 2]

$$\text{Maximize} \left( H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i \right) \quad (1a)$$

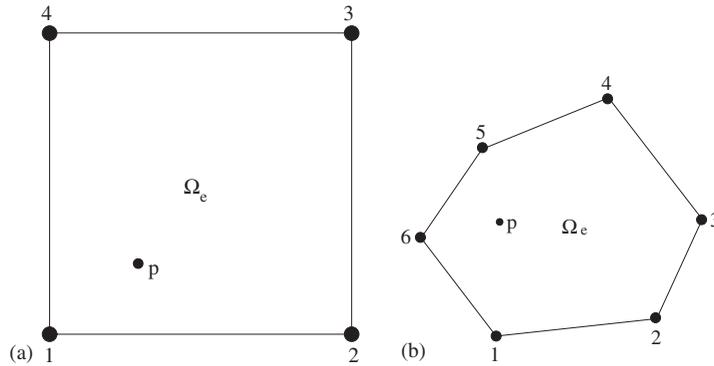


Figure 1. Approximation at point  $p$  in an  $n$ -gon: (a) square ( $n = 4$ ); and (b) hexagon ( $n = 6$ ).

subject to the constraints

$$\sum_{i=1}^n p_i = 1 \tag{1b}$$

$$\sum_{i=1}^n p_i g_r(x_i) = \langle g_r(x) \rangle \tag{1c}$$

where  $\phi \log \phi \doteq 0$  if  $\phi = 0$  and the specific form of the uncertainty measure  $H(\cdot)$  given in Equation (1a) is based on the satisfaction of three conditions, namely,  $H$  is a continuous function of the  $p_i$ , it attains its maximum when  $p_i = 1/n \forall i$  with the maximum entropy  $H_{\max} = H(1/n, \dots, 1/n)$  a monotonic function, and the composition law must hold [1, 2]. The above form of  $H$  guarantees that  $p_i \geq 0$ . The solution for  $p_i$  is obtained using the method of Lagrange multipliers, which we leave for a later stage when the specific polygonal interpolation problem is considered. The rationale for using MAXENT rests on the fact that the entropy of most of the admissible distributions that satisfy the constraints are concentrated in the neighbourhood of the maximum entropy [11], and hence the maximum entropy distribution is the least-biased and the one that has the maximum likelihood to occur.

We now draw an analogy between the above problem and the polygonal interpolation one, which forms the basis of the approach pursued in this study. For details on the properties associated with barycentric co-ordinates on polygons we refer the reader to Sukumar and Tabarraei [9]; here we only mention relations that are directly pertinent to our investigation. Consider the  $n$ -gons shown in Figure 1, where the nodes are located at the vertices of the polygons. To each node, we associate a shape function (barycentric co-ordinate)  $\phi_i \geq 0$  in  $\bar{\Omega}_e$ . Consider a point  $p$  in  $\bar{\Omega}_e$  with co-ordinate  $\mathbf{x} \equiv (x, y)$ . The interpolant  $u^h(\mathbf{x})$  for a scalar-valued function  $u(\mathbf{x}) : \bar{\Omega}_e \rightarrow \mathbb{R}$  can be written as

$$u^h(\mathbf{x}) = \sum_{i=1}^n \phi_i(\mathbf{x}) u_i \tag{2}$$

where  $u_i = u(\mathbf{x}_i)$  are the prescribed nodal values of  $u(\mathbf{x})$ . The property  $\phi_i \geq 0$  (convex combination) ensures that the interpolant satisfies a discrete maximum principle:  $\min_i u_i \leq u^h \leq \max_i u_i$ . We now impose the restriction that the  $\phi_i$  are such that  $u^h(\mathbf{x})$  can exactly reproduce constant

and linear functions (linear precision). In general,  $k$ th-order finite elements in one dimension satisfy  $k$ th-order completeness; however, in this study we only place linear completeness as a constraint on the interpolant. This requirement on trial and test approximations is sufficient to ensure optimal convergence of a Galerkin method for second-order partial differential equations. The linear reproducing conditions can be written in matrix form as

$$\mathbf{A}\boldsymbol{\phi} = \mathbf{b} \quad (3a)$$

or

$$\underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{bmatrix}}_{\mathbf{A} \quad (3 \times n)} \underbrace{\begin{Bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{Bmatrix}}_{\boldsymbol{\phi} \quad (n \times 1)} = \underbrace{\begin{Bmatrix} 1 \\ x \\ y \end{Bmatrix}}_{\mathbf{b} \quad (3 \times 1)} \quad (3b)$$

where  $(x_i, y_i)$  are the nodal co-ordinates. For  $n = 3$ , a unique interpolant (barycentric co-ordinates on a triangle) is obtained. However, if  $n > 3$ , the system is under-determined since there are more unknown shape functions  $\phi_i$  ( $i = 1, 2, \dots, n$ ) than available constraints. We note in passing that if we consider a hexagon with all its vertices on a unit circle and require  $u^h(\mathbf{x})$  to have quadratic precision, then there exists *no solution* for the shape functions  $\phi_i$  ( $i = 1, 2, \dots, 6$ ). This follows since the set  $\{1, x, y, xy, x^2, y^2\}$  consists of linearly dependent basis functions ( $x_i^2 + y_i^2 = 1$ ), and therefore the matrix  $\mathbf{A}$  ( $6 \times 6$ ) is singular. Returning to the problem on-hand, the solution of the linear system  $\mathbf{A}\boldsymbol{\phi} = \mathbf{b}$  can be obtained using the method of generalized (Morse–Penrose) inverse [12]. The generalized inverse  $\mathbf{A}^+$  satisfies the following properties:  $\mathbf{A}^+\mathbf{A} = \mathbf{A}\mathbf{A}^+$ ,  $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$ , and  $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$ . The general solution is:  $\boldsymbol{\phi} = \mathbf{A}^+\mathbf{b} + (\mathbf{I} - \mathbf{A}^+\mathbf{A})\mathbf{c}$ , where  $\mathbf{c}$  is an arbitrary constant vector. If  $\mathbf{c} = \mathbf{0}$ , we obtain the minimum norm solution  $\boldsymbol{\phi} = \mathbf{A}^+\mathbf{b}$  (Pinv  $m$ -file in Matlab™). The minimum norm solution is equivalent to that obtained via the constrained optimization problem: Minimize  $\|\boldsymbol{\phi}\|$  subject to the linear system in Equation (3). The caveat here is that solutions  $\phi_i < 0$  are also admissible (see Section 3), and hence  $\|\boldsymbol{\phi}\|$  is not suitable as an uncertainty measure [2].

From Equations (1) and (3), we immediately note the correspondence: the weights  $\phi_i$  can be associated with probability measures  $p_i$ , nodal co-ordinate  $\mathbf{x}_i$  is the event  $x_i$ , and the averages  $\langle g_1 \rangle \equiv \langle x \rangle = x$  and  $\langle g_2 \rangle \equiv \langle y \rangle = y$ . In light of the MAXENT formulation, the shape functions  $\phi_i$  can be viewed as least-biased barycentric co-ordinates. If  $\boldsymbol{\phi} = \{\phi_1, \phi_2, \dots, \phi_n\}$  is a point in  $n$ -dimensional space, which is constrained by  $\phi_i \geq 0$ ,  $\sum_i \phi_i = 1$ ,  $\sum_i \phi_i x_i = x$  and  $\sum_i \phi_i y_i = y$ , then we obtain a surface (domain  $D$ ) in an  $(n-3)$ -dimensional hyperplane. In  $D$ ,  $0 \leq H \leq \log n$ , with the maximum occurring at the centroid of the polygon (maximum uncertainty). There are multiple local minima for the entropy which occur at the nodal locations: at  $\mathbf{x} = \mathbf{x}_j$ ,  $\phi_i(\mathbf{x}_j) = \delta_{ij}$  (Kronecker-delta) holds with minimum entropy  $H = 0$  (maximum certainty or deterministic). In general, increase in  $H$  implies less information is available, whereas decrease in  $H$  indicates more information (closer to being deterministic) is known. If a feasible solution exists, then the maximum entropy solution is unique and maximal. Along the boundary of the polygon (Figure 1), the interpolant is linear (only two  $\phi_i$  are non-zero at a point along the

interior of the edges of the polygon). The above inferences that we have drawn on the MAXENT interpolant were also true for the polygonal interpolants considered in Reference [9], which in part provided motivation to pursue the present investigation.

### 2.1. Method of Lagrange multipliers

We can now state the polygonal interpolation problem using the MAXENT formulation. We seek to find  $\phi_i$  ( $i = 1, 2, \dots, n$ ) that are the solution of the following problem:

$$\text{Maximize} \left( H(\phi_1, \phi_2, \dots, \phi_n) = - \sum_{i=1}^n \phi_i \log \phi_i \right) \quad (4a)$$

subject to the *linear constraints*

$$\sum_{i=1}^n \phi_i = 1 \quad (4b)$$

$$\sum_{i=1}^n \phi_i x_i = x \quad (4c)$$

$$\sum_{i=1}^n \phi_i y_i = y \quad (4d)$$

where  $(x, y) \in \Omega_e$  (interior of the polygon). On the boundary of the convex hull, the MAXENT solution for  $\phi_i$  is known. We use the method of Lagrange multipliers to solve the constrained optimization problem. Let  $\lambda_r$  ( $r = 0, 1, 2$ ) be the Lagrange multipliers associated with the three constraints, and we set the variation of the augmented Lagrangian to zero ( $\delta L = 0$ ) [2]

$$\delta \left[ \sum_{i=1}^n -\phi_i \log \phi_i + \lambda_0 \left( 1 - \sum_{i=1}^n \phi_i \right) + \lambda_1 \left( x - \sum_{i=1}^n \phi_i x_i \right) + \lambda_2 \left( y - \sum_{i=1}^n \phi_i y_i \right) \right] = 0 \quad (5)$$

or

$$[-1 - \log \phi_i - \lambda_0 - \lambda_1 x_i - \lambda_2 y_i] \delta \phi_i = 0 \quad \forall \delta \phi_i \quad (6)$$

and since the variations  $\delta \phi_i$  are arbitrary, we obtain

$$-1 - \log \phi_i - \lambda_0 - \lambda_1 x_i - \lambda_2 y_i = 0 \quad (i = 1, 2, \dots, n) \quad (7)$$

In keeping with convention, we let  $\lambda_0 = \log Z - 1$  [2]. The function  $Z$  is known as the partition function, and is used to represent the canonical distribution in statistical mechanics. Then, the above equation can be written as

$$\log \phi_i + \log Z = -\lambda_1 x_i - \lambda_2 y_i \quad (8)$$

or

$$\phi_i = \frac{e^{-\lambda_1 x_i - \lambda_2 y_i}}{Z} \quad (9)$$

and since  $\sum_i \phi_i = 1$ , we have

$$Z(\lambda_1, \lambda_2) = \sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} \quad (10)$$

and therefore the shape functions  $\phi_i$  can be written as

$$\phi_i = \frac{e^{-\lambda_1 x_i - \lambda_2 y_i}}{\sum_{j=1}^n e^{-\lambda_1 x_j - \lambda_2 y_j}} \equiv \frac{Z_i}{\sum_{j=1}^n Z_j} \quad (11)$$

where  $Z_i$  can be viewed as a weight function, and the last expression is a common form in which most polygonal interpolation schemes appear [9]. Now, to complete the formulation, the  $\phi_i$  must satisfy the constraints given in Equations (4c) and (4d). The constraints can also be written as [2]

$$-\frac{\partial \lambda_0}{\partial \lambda_r} = -\frac{\partial(\log Z)}{\partial \lambda_r} = \langle g_r(\mathbf{x}) \rangle, \quad (r = 1, 2) \quad (12)$$

where  $\langle g_1(\mathbf{x}) \rangle = x$  and  $\langle g_2(\mathbf{x}) \rangle = y$ . On substituting Equation (11) in Equations (4c) and (4d), we obtain the following non-linear equations for  $\lambda_1$  and  $\lambda_2$ :

$$f_1(\lambda_1, \lambda_2) = 0, \quad f_2(\lambda_1, \lambda_2) = 0 \quad (13a)$$

where

$$f_1(\lambda_1, \lambda_2) = \frac{\sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} x_i}{Z} - x \quad (13b)$$

$$f_2(\lambda_1, \lambda_2) = \frac{\sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} y_i}{Z} - y \quad (13c)$$

and on solving for  $\lambda_1$  and  $\lambda_2$ , the shape functions  $\phi_i$  are obtained from Equation (11). On using Equations (11) and (4a), the maximum entropy is:  $H_{\max} = \log Z(\lambda_1, \lambda_2) + \lambda_1 x + \lambda_2 y$ . Note that Equation (13) defines a non-linear mapping:  $x = x(\lambda_1, \lambda_2)$ ,  $y = y(\lambda_1, \lambda_2)$ , which is used in Section 4.3 to derive the derivatives of the shape functions.

*2.1.1. Nearest-neighbour interpolation.* If the only constraint is  $\sum_i \phi_i = 1$  (constant precision), then from Equation (11) with  $\lambda_r$  ( $r = 1, 2$ ) set to zero, we obtain

$$\phi_i = \frac{1}{n} \quad \forall i \quad (14)$$

which is the nearest-neighbour interpolant. The shape functions are a constant over the polygon.

*2.1.2. Bilinear interpolation.* Consider the unit square shown in Figure 1(a). Let the domain  $\Omega_e = (0, 1)^2$  with the origin at node 1. For a point  $p$  in  $\Omega_e$  with co-ordinate  $(x, y)$ , we use

Equation (13) to obtain the following equations for  $\lambda_1$  and  $\lambda_2$ :

$$\frac{e^{-\lambda_1} + e^{-\lambda_1 - \lambda_2}}{1 + e^{-\lambda_1} + e^{-\lambda_2} + e^{-\lambda_1 - \lambda_2}} = x \quad (15a)$$

$$\frac{e^{-\lambda_2} + e^{-\lambda_1 - \lambda_2}}{1 + e^{-\lambda_1} + e^{-\lambda_2} + e^{-\lambda_1 - \lambda_2}} = y \quad (15b)$$

which simplifies to

$$\frac{e^{-\lambda_1}}{1 + e^{-\lambda_1}} = x \quad (16a)$$

$$\frac{e^{-\lambda_2}}{1 + e^{-\lambda_2}} = y \quad (16b)$$

and therefore

$$e^{-\lambda_1} = \frac{x}{1-x}, \quad e^{-\lambda_2} = \frac{y}{1-y} \quad (17)$$

From Equations (10) and (17), we have  $Z^{-1} = (1-x)(1-y)$  and on using Equation (11), we obtain the MAXENT shape functions

$$\phi_1(x, y) = (1-x)(1-y), \quad \phi_2(x, y) = x(1-y), \quad \phi_3(x, y) = xy, \quad \phi_4(x, y) = y(1-x) \quad (18)$$

which are identical to bilinear finite element shape functions.

### 3. NUMERICAL ALGORITHM

The non-linear equations in Equation (13) immediately suggest the use of an iterative solution scheme such as Newton's method. However, to solve Equation (13) without cognizance of its special structure is fraught with numerical difficulties. As is well-known, Newton's method for root finding is sensitive (especially so in multi-dimensions) to the initial guess. Numerical tests have revealed that convergence is realized if and only if the initial guess is proximal to the exact solution, and as the point  $p(x, y)$  approaches the boundary of the domain, the situation is exacerbated and non-convergence or even convergence to the wrong solution results.

The fact that Equation (13) arises from a maximum entropy formalism suggests that there is more information in  $f_r$  that one can use. This fact was used by Agmon and co-workers [13, 14] towards the development of a suitable algorithm to find the distribution of maximum entropy. In Reference [15], the maximum entropy (primal) problem is recast into one (dual problem) in which the Lagrange multipliers are determined as the set that minimizes a convex *potential function*  $F(\lambda_1, \lambda_2)$ . A brief description of the algorithm follows. On noting that  $\sum_i \phi_i = 1$  and letting  $\tilde{x}_i = x_i - x$  and  $\tilde{y}_i = y_i - y$ , we can re-write the linear reproducing conditions in

Equations (4c) and (4d) as

$$\sum_{i=1}^n \phi_i \tilde{x}_i = 0 \quad (19a)$$

$$\sum_{i=1}^n \phi_i \tilde{y}_i = 0 \quad (19b)$$

which can be viewed as using a shifted co-ordinate system with the origin placed at  $(x, y)$ . This choice is preferred in numerical computations since it is more stable than the form that appears in Equation (4). Now, the expressions for  $Z$  and  $\phi_i$  take the form

$$\tilde{Z}(\lambda_1, \lambda_2) = \sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i}, \quad \phi_i = \frac{e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i}}{\tilde{Z}} \quad (20)$$

where the Lagrange multipliers  $\lambda_1$  and  $\lambda_2$  remain unchanged, and there is a change only in  $\lambda_0$  with the new value  $\tilde{\lambda}_0 = \lambda_0 + \lambda_1 x + \lambda_2 y$ . We note that  $e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} = e^{-\lambda_1 x_i - \lambda_2 y_i} \cdot e^{\lambda_1 x + \lambda_2 y}$ , and therefore the equality between the expressions for  $\phi_i$  in Equations (11) and (20) is apparent. In addition, the maximum entropy is now given by  $H_{\max} = \log \tilde{Z}(\lambda_1, \lambda_2)$ . Instead of Equation (13), we now have [14]

$$f_r(\lambda_1, \lambda_2) = \frac{\partial(\log \tilde{Z})}{\partial \lambda_r} = 0, \quad (r = 1, 2) \quad (21a)$$

$$f_1(\lambda_1, \lambda_2) = -\frac{\sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{x}_i}{\tilde{Z}} \quad (21b)$$

$$f_2(\lambda_1, \lambda_2) = -\frac{\sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{y}_i}{\tilde{Z}} \quad (21c)$$

Since  $f_1(\lambda_1, \lambda_2) \equiv \partial F / \partial \lambda_1 = 0$  and  $f_2(\lambda_1, \lambda_2) \equiv \partial F / \partial \lambda_2 = 0$ , we note that  $\mathbf{f}$  is a conservative vector field with scalar potential  $\mathbf{f} = \nabla \log \tilde{Z}(\lambda_1, \lambda_2) \equiv \nabla F(\lambda_1, \lambda_2)$  [14]. In addition, as  $\tilde{Z} > 0$ , the above equations can also be written as

$$f_r(\lambda_1, \lambda_2) = \frac{\partial \tilde{Z}}{\partial \lambda_r} = 0, \quad (r = 1, 2) \quad (22a)$$

$$f_1(\lambda_1, \lambda_2) = -\sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{x}_i \quad (22b)$$

$$f_2(\lambda_1, \lambda_2) = -\sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{y}_i \quad (22c)$$

where now  $\mathbf{f} = \nabla \tilde{Z}(\lambda_1, \lambda_2) \equiv \nabla \tilde{F}(\lambda_1, \lambda_2)$ . Let  $\lambda^t \equiv (\lambda_1^t, \lambda_2^t)$  denote any trial set of Lagrange multipliers and the set  $\lambda \equiv (\lambda_1, \lambda_2)$  be the one that maximizes the entropy (since  $F(\lambda) = H_{\max}$ ).

In the dual problem, the point  $\mathbf{x}$  is fixed, and the Lagrange multipliers are varied, as opposed to the primal formulation in which the constraint equations define a mapping from the point  $\mathbf{x}$  to the solution set  $\lambda$  and vice versa. The positive-definiteness of  $F$  was established in Reference [15] (see Section 4.3 for the derivation of the Hessian of  $\log Z$ ), and consequently  $F$  attains a global minimum at  $\lambda^t = \lambda$ . Since  $\lambda$  is the minimizer of both  $F$  and  $\tilde{F}$ , we now show that  $\tilde{F} = \tilde{Z}(\lambda^t)$  attains a global minimum at  $\lambda^t = \lambda$  for the polygonal interpolation problem.

*Claim 1*

The function  $\tilde{F} = \tilde{Z}(\lambda^t)$  attains a global minimum at  $\lambda^t = \lambda$ .

*Proof*

For the polygonal interpolation problem, the Hessian matrix of  $\tilde{F}(\lambda_1^t, \lambda_2^t) = Z(\lambda_1^t, \lambda_2^t)$ , which is denoted by  $\tilde{G}_{rs} = \partial f_r / \partial \lambda_s^t = \partial^2 \tilde{F} / \partial \lambda_r^t \partial \lambda_s^t$ , is

$$\tilde{\mathbf{G}} = \begin{bmatrix} \sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{x}_i^2 & \sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{x}_i \tilde{y}_i \\ \sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{x}_i \tilde{y}_i & \sum_{i=1}^n e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i} \tilde{y}_i^2 \end{bmatrix} \tag{23}$$

Letting  $a_i = e^{-\lambda_1 \tilde{x}_i - \lambda_2 \tilde{y}_i}$ ,  $\mathbf{u} = (\sqrt{a_1} \tilde{x}_1, \dots, \sqrt{a_n} \tilde{x}_n)$ , and  $\mathbf{v} = (\sqrt{a_1} \tilde{y}_1, \dots, \sqrt{a_n} \tilde{y}_n)$ , we can write  $\det(\tilde{\mathbf{G}}) = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - |\mathbf{u} \cdot \mathbf{v}|^2$ . On using the Cauchy–Schwartz inequality,  $\|\mathbf{u}\| \|\mathbf{v}\| - |\mathbf{u} \cdot \mathbf{v}| \geq 0$ , we obtain  $\det(\tilde{\mathbf{G}}) \geq 0$ , with equality iff  $\mathbf{u}$  and  $\mathbf{v}$  are collinear. Collinearity occurs only on the boundary of the convex hull ( $\lambda_r \rightarrow \infty$ ,  $r = 1$  or  $2$ ), and hence in the interior of the domain,  $\det(\tilde{\mathbf{G}}) > 0$ . The matrix  $\tilde{\mathbf{G}}$  is also symmetric and diagonally dominant with positive diagonal entries, which implies that  $\tilde{\mathbf{G}}$  is a positive-definite matrix. Since  $\nabla \tilde{F} = \mathbf{0}$  at  $\lambda^t = \lambda$  and the Hessian matrix is positive-definite in the interior of the polygon, it follows that  $\tilde{F}(\lambda^t)$  attains a global minimum at  $\lambda^t = \lambda$ .  $\square$

We can now recast the maximum entropy (primal) problem as the following dual problem [14]: find  $(\lambda_1^t, \lambda_2^t)$  such that

$$F = \log \tilde{Z}(\lambda_1^t, \lambda_2^t) \text{ is minimized} \tag{24}$$

or equivalently,  $\tilde{F} = \tilde{Z}(\lambda_1^t, \lambda_2^t)$  is minimized. The problem as posed in Equations (13) or (21) involve a system of  $m$  non-linear equations which is, in general, difficult to solve numerically. On the other hand, by reformulating the MAXENT problem as a convex minimization problem, a robust numerical solution procedure can be readily implemented. The solution of Equation (24) can be obtained using a one-dimensional line search algorithm, where it is always possible to explicitly bracket the solution and to proceed in a downhill direction towards the minimizing solution set [16]. If we had used the constraints in Equation (4), then  $F = \log Z(\lambda_1^t, \lambda_2^t) + \lambda_1^t x + \lambda_2^t y$ , and hence the use of Equation (19) leads to fewer arithmetic operations to compute  $F$  and its derivatives. At  $\lambda$ ,  $\log \tilde{Z}$  (minimum) and the entropy (maximum) assume the same value. In general,  $F = \log \tilde{Z}$  is an upper bound for the entropy [15].

The algorithm to solve Equation (24) consists of an initial guess, a suitable search direction, and a scalar  $\alpha$  that provides a measure of the distance to be traversed along the search direction at each iteration step. We denote  $\lambda^k \equiv (\lambda_1^k, \lambda_2^k)^T$  to be the solution at the  $k$ th iteration; the initial guess is  $\lambda^0$ . A second-order Taylor series expansion of  $F$  is:

$$F(\lambda^k + \Delta \lambda^k) = F(\lambda^k) + [\nabla F]_{(\lambda_1^k, \lambda_2^k)} \cdot \Delta \lambda^k + \frac{1}{2} (\Delta \lambda^k)^T \mathbf{G}_{(\lambda_1^k, \lambda_2^k)} (\Delta \lambda^k) \tag{25}$$

where  $\mathbf{G}$  is the Hessian matrix of  $F$ , and from the above equation we can infer that  $\Delta\lambda^k = -\mathbf{G}^{-1}\nabla F$  (Newton's method) is the direction that minimizes  $F(\lambda^k + \Delta\lambda^k)$ . Given the solution at iteration  $k$ , the update for the Lagrange multipliers is [13]

$$\lambda_r^{k+1} = \lambda_r^k + \alpha\Delta\lambda_r^k \tag{26}$$

where  $\alpha$  is determined through the condition that  $F(\lambda_1^{k+1}, \lambda_2^{k+1})$  attains a minimum along the search direction [13]. A line search (bisection or using the golden section search [16]) is suitable to find the  $\alpha$  that leads to the minimum  $F$  along the indicated direction. The choice of selecting  $\alpha$ , which is not present in a standard Newton method, provides the improved performance and the fast convergence in the present algorithm. In the vicinity of the boundary (at least one  $\lambda_r \rightarrow \infty$ ),  $F$  is asymptotically linear [15] and therefore the Hessian matrix becomes nearly singular, and hence should not be used to determine the search direction [14]. For convenience, we have used  $\Delta\lambda^k = -\nabla F$  (direction of steepest descent) with starting guess  $\lambda^0 = \mathbf{0}$  for all points in the domain. Either  $\log \tilde{Z}$  or  $\tilde{Z}$  can be used (difference of a few iterations in general) if just the gradient is used in the update; however, since the former has a smaller range, it is preferred. Prior to the update for  $\lambda$  through Equation (26), the stopping (convergence) criterion is invoked:  $\|\nabla F\|^{(k)} < \varepsilon$ , where  $\varepsilon$  is a suitably small real number.

For given data  $\langle g_r \rangle$ , the feasibility of a solution was explored in Reference [14]. The functions  $\langle g_r \rangle$  must be linearly independent, which is trivially met in the present case since  $\langle g_1 \rangle = x$  and  $\langle g_2 \rangle = y$  are independent. The above condition alone is not sufficient, since even if matrix  $\mathbf{A}$  in Equation (3) has full rank of 3 ( $A^+$  exists), a solution that meets  $\phi_i \geq 0$  might not exist. This fact was mentioned earlier when discussing the minimum norm solution which need not adhere to the  $\phi_i \geq 0$  restriction. To demonstrate this, we use  $\sum_{i=1}^n \phi_i^2$  as the objective function to be minimized, subject to the constraints in Equation (3). On using the method of Lagrange multipliers, we obtain

$$\phi_i^{mn} = \frac{1}{2} (\lambda_0 + \lambda_1 x_i + \lambda_2 y_i), \quad \lambda_0 = \lambda_0(\lambda_1, \lambda_2) \tag{27}$$

where  $\lambda_0$  is related to  $\lambda_1$  and  $\lambda_2$  through Equation (4a). Now consider a unit square (four-node element), and proceeding as in Section 2.1.2, we can write the minimum norm shape functions as

$$\phi_1^{mn}(\mathbf{x}) = \frac{3 - 2x - 2y}{4}, \quad \phi_2^{mn}(\mathbf{x}) = \frac{1 + 2x - 2y}{4} \tag{28a}$$

$$\phi_3^{mn}(\mathbf{x}) = \frac{2x + 2y - 1}{4}, \quad \phi_4^{mn}(\mathbf{x}) = \frac{-2x + 2y + 1}{4} \tag{28b}$$

and it is now apparent that for certain  $\mathbf{x}$ ,  $\phi_i^{mn}$  are negative. For instance, if  $\mathbf{x} = \mathbf{0}$ , then  $\phi^{mn} = \{0.75, 0.25, -0.25, 0.25\}$ .

In addition to  $\sum_i \phi_i = 1$ , if we have only one additional constraint ( $m = 1$ ), for e.g.  $\langle g_1(x) \rangle = x$ , then for a feasible solution [14]

$$\min_{i=1,2,\dots,n} x_i < x < \max_{i=1,2,\dots,n} x_i \tag{29}$$

which indicates that  $x$  must be between its lower and upper bounds. The above inequality was generalized to more than one constraint [14], and for the problem on-hand it takes

the form

$$\min_{i=1,2,\dots,n} \{c_1x_i + c_2y_i\} < c_1x + c_2y < \max_{i=1,2,\dots,n} \{c_1x_i + c_2y_i\} \tag{30}$$

which should hold for any non-zero vector  $\mathbf{c} \equiv (c_1, c_2)$ . Letting  $\tilde{x}_i = x_i - x$ ,  $\tilde{y}_i = y_i - y$ , and  $\tilde{\mathbf{x}}_i \equiv (\tilde{x}_i, \tilde{y}_i)$ , we can re-write the above inequalities as

$$\min_{i=1,2,\dots,n} \mathbf{c} \cdot \tilde{\mathbf{x}}_i < 0, \quad \max_{i=1,2,\dots,n} \mathbf{c} \cdot \tilde{\mathbf{x}}_i > 0 \tag{31a}$$

or in matrix form ( $\mathbf{Bc} = \mathbf{d}$ ) as

$$\begin{bmatrix} \tilde{x}_1 & \tilde{y}_1 \\ \tilde{x}_2 & \tilde{y}_2 \\ \dots & \dots \\ \tilde{x}_n & \tilde{y}_n \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \end{Bmatrix} = \begin{Bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{Bmatrix} \tag{31b}$$

where  $d_i d_j < 0$  for at least one  $(i, j)$  pair. In Equation (31), equality on one side occurs when  $\lambda_r \rightarrow \infty$  ( $r = 1$  or  $2$ ), i.e. the point  $p$  lies on a polygonal edge, whereas equality on both sides implies that the constraints are linearly dependent ( $d_i = 0 \forall i$ ) [15]. Clearly, the columns in  $\mathbf{B}$  are always linearly independent for a polygon.

*Claim 2*

A feasible solution for the MAXENT problem posed in Equation (4) exists for any point  $p(x, y)$  within the convex hull of a set of nodes.

*Proof*

We establish the satisfaction of the feasibility condition for arbitrary  $n$ -gons (convex and non-convex) through a simple geometric inference. Consider the non-convex  $n$ -gon ( $n = 15$ ) shown in Figure 2. The boundary of the convex hull,  $\partial\Omega_{ch}$ , is defined by connecting the

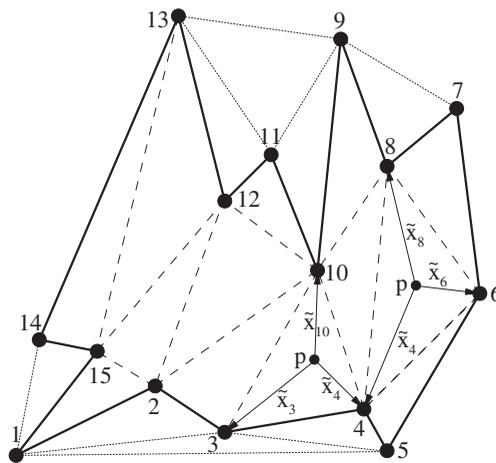


Figure 2. Feasibility of the MAXENT solution in a non-convex polygon.

nodes 1–5–6–7–9–13–14–1. Any polygon can be partitioned into a collection of triangles (trivially seen for a convex polygon) using its vertices alone; such a partition is shown in Figure 2. For a point  $p$  within an arbitrary triangle  $t$ , the feasibility conditions are automatically satisfied since the barycentric co-ordinates in the interior of a triangle are strictly positive. So, there always exist a  $r$  and  $s$  such that  $\mathbf{c} \cdot \tilde{\mathbf{x}}_r > 0$ ,  $\mathbf{c} \cdot \tilde{\mathbf{x}}_s < 0$ . The vectors  $\tilde{\mathbf{x}}_i$  for two particular triangles (obtained after a simplex-partition of the polygon) are shown in Figure 7. Since any point  $p(x, y)$  within the convex hull also lies within one of the partitioned triangles, it follows that Equation (31) is satisfied, as required.  $\square$

A detailed discussion of entropy optimization algorithms is presented in Reference [17]. The use of these algorithms and considerations of alternative entropy measures are topics for future research.

#### 4. NUMERICAL EXAMPLES

We first present numerical results for the MAXENT shape functions on convex and non-convex polygons. The shape function and entropy of the polygonal interpolants (Laplace and mean-value) used in Reference [9] are compared to those obtained by the MAXENT interpolant. Then we derive the derivatives of the MAXENT shape functions and use the patch test to verify its accuracy. Lastly, the  $L^2$  error norm estimates of the different polygonal interpolants are presented for a surface approximation problem. The MAXENT shape function for node  $i$  will be denoted by  $\phi_i^{\text{mxt}}$ , whereas  $\phi_i^{\text{lap}}$  and  $\phi_i^{\text{mvc}}$  will be used for the Laplace and mean value shape functions, respectively. In the numerical computations, a tolerance  $\varepsilon = 10^{-7}$  is used, and unless indicated otherwise,  $F = \log \tilde{Z}$  is adopted to compute the MAXENT shape functions.

##### 4.1. Convex polygons

In Figure 3, the MAXENT shape function and entropy are shown on a unit square. It is observed that the shape function is identical to bilinear finite element shape function, thereby providing a numerical verification of the proof outlined in Section 2.1.2. The stringent tolerance for convergence does lead to some increase in the number of iterations for convergence—away from the boundary of the square, convergence is attained in less than five iterations, close to the boundary, 5–15 iterations are required, and if the sampling point is within  $\delta = 0.01$  of the boundary, 50–100 iterations are required. In Figure 4, the contour plots and convergence path using  $F = \log \tilde{Z}$  is shown, whereas in Figure 5, the results for  $\tilde{Z}$  are illustrated. In Figure 4(a), the function  $\log \tilde{Z}$  is plotted for  $\mathbf{x} = (0.56, 0.42)$ , and a 2D contour plot is illustrated in Figure 4(b). In this case, three iterations are required for convergence, and the converged solution (4-digit precision) for the Lagrange multipliers is:  $\lambda = (-0.2411, 0.3227)$ . In Figures 4(c) and 4(d) the contour plot of  $\log \tilde{Z}$  is shown for the points  $(0.9, 0.12)$  and  $(0.99, 0.12)$ , respectively. In Figure 4(c), six iterations are required for convergence and  $\lambda = (-2.1972, 1.9924)$  is the converged solution, whereas in Figure 4(d), 60 iterations are required and  $\lambda = (-4.5951, 1.9924)$  is the final solution. The number of iterations and the converged solution  $\lambda$  using  $\tilde{Z}$  (Figures 5(a)–5(d)) are identical to those obtained using  $\log \tilde{Z}$ , with the only difference arising in the last case (Figure 5(d)) where convergence is attained in 62 iterations.

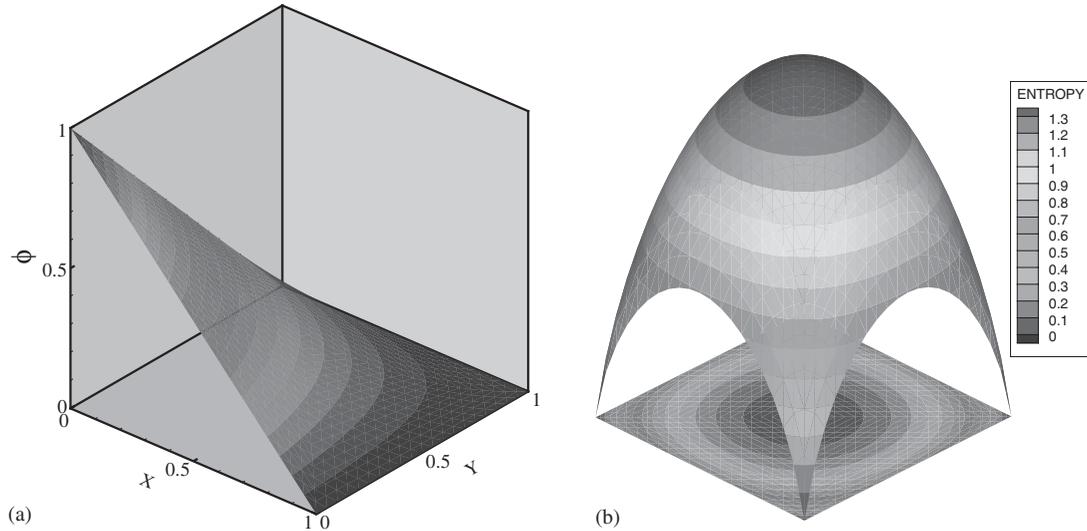


Figure 3. Numerical computations using the MAXENT formulation on a unit square. (a) shape function; and (b) entropy.

In Figures 6(a)–6(c), the MAXENT, Laplace, and mean-value shape functions for node 1 in a regular hexagon are shown. The plots for the Laplace and MAXENT shape functions are similar, whereas  $\phi_1^{\text{mvc}}$  is distinct. To further reveal the relationship between these interpolants, we plotted the ratio of the entropy of the Laplace and mean value shape functions to that obtained (maximum entropy) using the MAXENT shape functions. These results are presented in Figures 6(d) and 6(e). It is clearly observed that using the entropy as a measure, the Laplace interpolant is proximal to the MAXENT interpolant.

#### 4.2. Non-convex polygons

We had earlier proved that the MAXENT formulation leads to a feasible solution for the shape functions on non-convex polygons. Referring once again to the  $n$ -gon shown in Figure 2, we now conduct numerical computations in support of the proof. In Figures 7(a) and 7(b), we show the shape function variation of node 6 ( $\phi_6^{\text{mxt}}$ ) and the maximum entropy distribution for the case when the shape functions are assumed to be known along the boundary of the convex hull,  $\partial\Omega_{\text{ch}}$ . In this case, nodal interpolation is not realized at nodes that are located in the interior of the convex hull. In Figures 7(c) and 7(d), the plots are depicted for the case when the shape functions are known along  $\partial\Omega_e$ . We note that the variation of  $\phi_6^{\text{mxt}}$  using either approach is very similar, with some difference arising along the edges that do not include node 6. The latter case is presented since even though it involves a modification to the maximum entropy distribution, it would lead to inter-element continuity; however continuity inside the element and close to the edge would not be rigorously met. The maximum value of the entropy ( $\log 15$ ) is attained at the centroid.

We present shape functions on two meshes that contain non-convex elements. In both examples, we assume linear interpolation along the boundary  $\partial\Omega_e$ . In Figure 8(a), the domain is a bi-unit square and node 7 is located at  $(0.4, 0)$ . The MAXENT shape function for

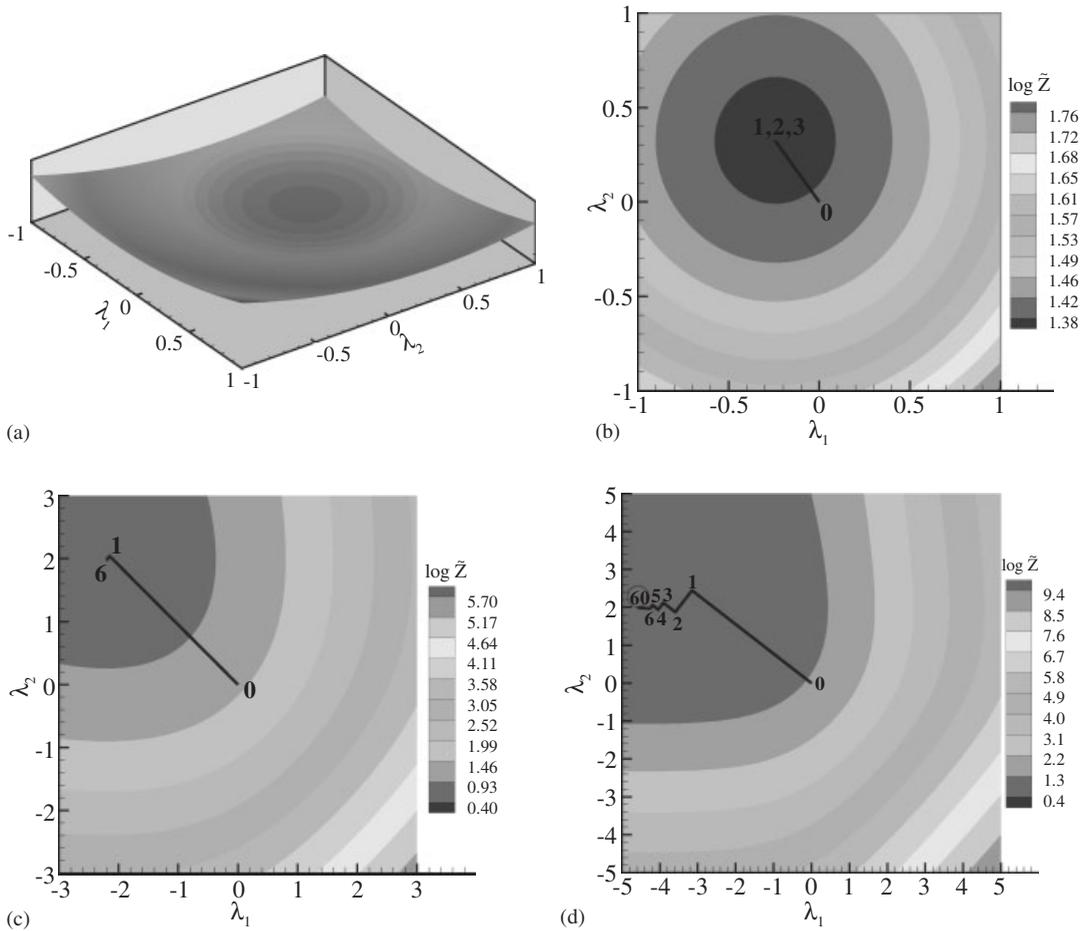


Figure 4. Contour plots of  $\log \tilde{Z}$  and convergence of the algorithm: (a),(b)  $\mathbf{x} = (0.56, 0.42)$ ; (c)  $\mathbf{x} = (0.9, 0.12)$ ; and (d)  $\mathbf{x} = (0.99, 0.12)$ .

node 7 is plotted in Figure 8(b), with  $\phi_7^{\text{mxt}} = 1$  at  $\mathbf{x} = \mathbf{x}_7$  and it decreases monotonically with increasing distance from  $\mathbf{x}_7$ . As indicated earlier, conformity along the edge is ensured since the shape function is linear along the edges that include node 7 as a vertex. In Figure 8(c), a four-node non-convex element is shown and in Figure 8(d) the shape function plot for node 4 is illustrated. The above numerical results substantiate the existence of the MAXENT solution for non-convex polygons. However, an open topic for future research-work is the construction of conforming (barycentric) approximations on meshes with non-convex polygons.

4.3. Computation of shape function derivatives

On using Equation (11), we can write  $(\phi_i \equiv \phi_i^{\text{mxt}})$

$$\log \phi_i = \log (e^{-\lambda_1 x_i - \lambda_2 y_i}) - \log Z(\lambda_1, \lambda_2) \tag{32}$$

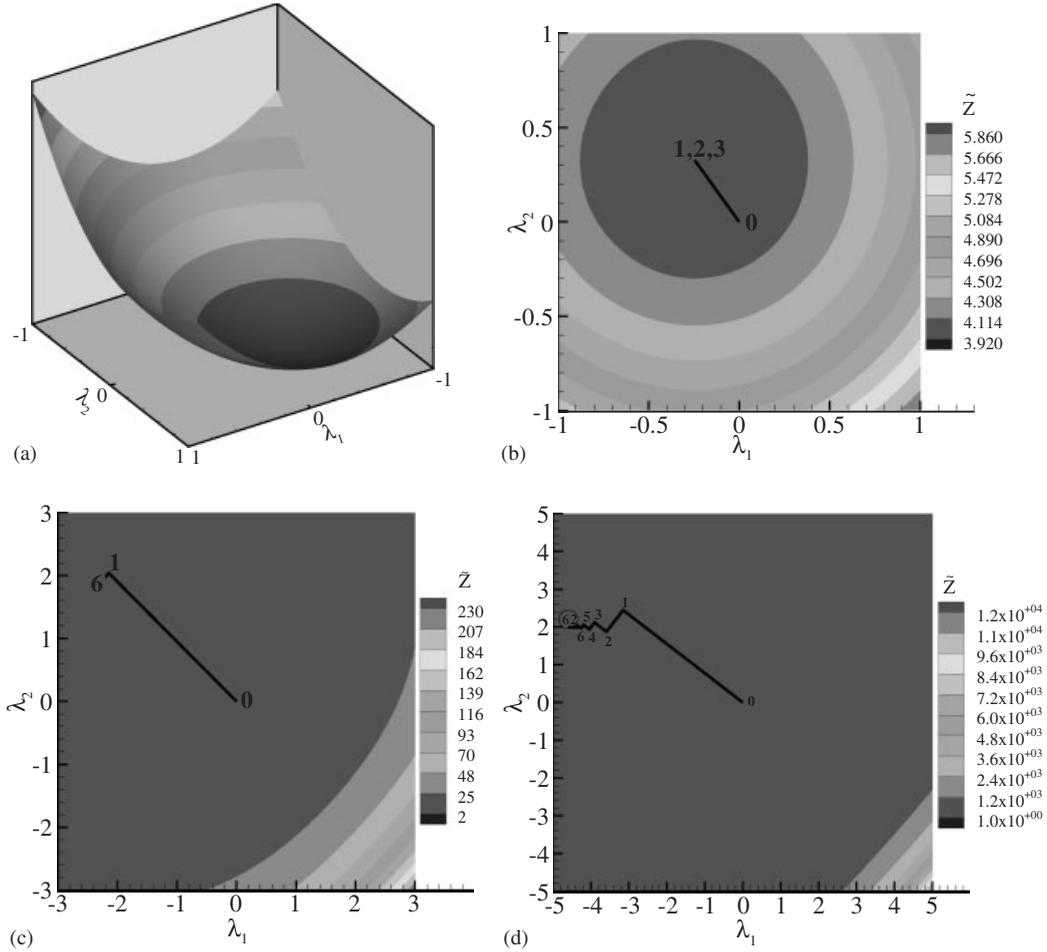


Figure 5. Contour plots of  $\tilde{Z}$  and convergence of the algorithm: (a),(b)  $\mathbf{x} = (0.56, 0.42)$ ; (c)  $\mathbf{x} = (0.9, 0.12)$ ; and (d)  $\mathbf{x} = (0.99, 0.12)$ .

and therefore

$$\frac{\delta\phi_i}{\phi_i} = -\delta\lambda_1 x_i - \delta\lambda_2 y_i - \delta(\log Z) \tag{33}$$

But since  $-\delta(\log Z) = \delta\lambda_r \langle g_r(\mathbf{x}) \rangle$  (sum implied on  $r$ ) from Equation (12), the above reduces to

$$\frac{\delta\phi_i}{\phi_i} = (x - x_i)\delta\lambda_1 + (y - y_i)\delta\lambda_2 \tag{34}$$

and the derivatives of the MAXENT shape functions are

$$\phi_{i,\alpha}^{\text{mxt}} = \phi_i^{\text{mxt}} \{ (x - x_i)\lambda_{1,\alpha} + (y - y_i)\lambda_{2,\alpha} \} \quad (\alpha = x, y) \tag{35}$$

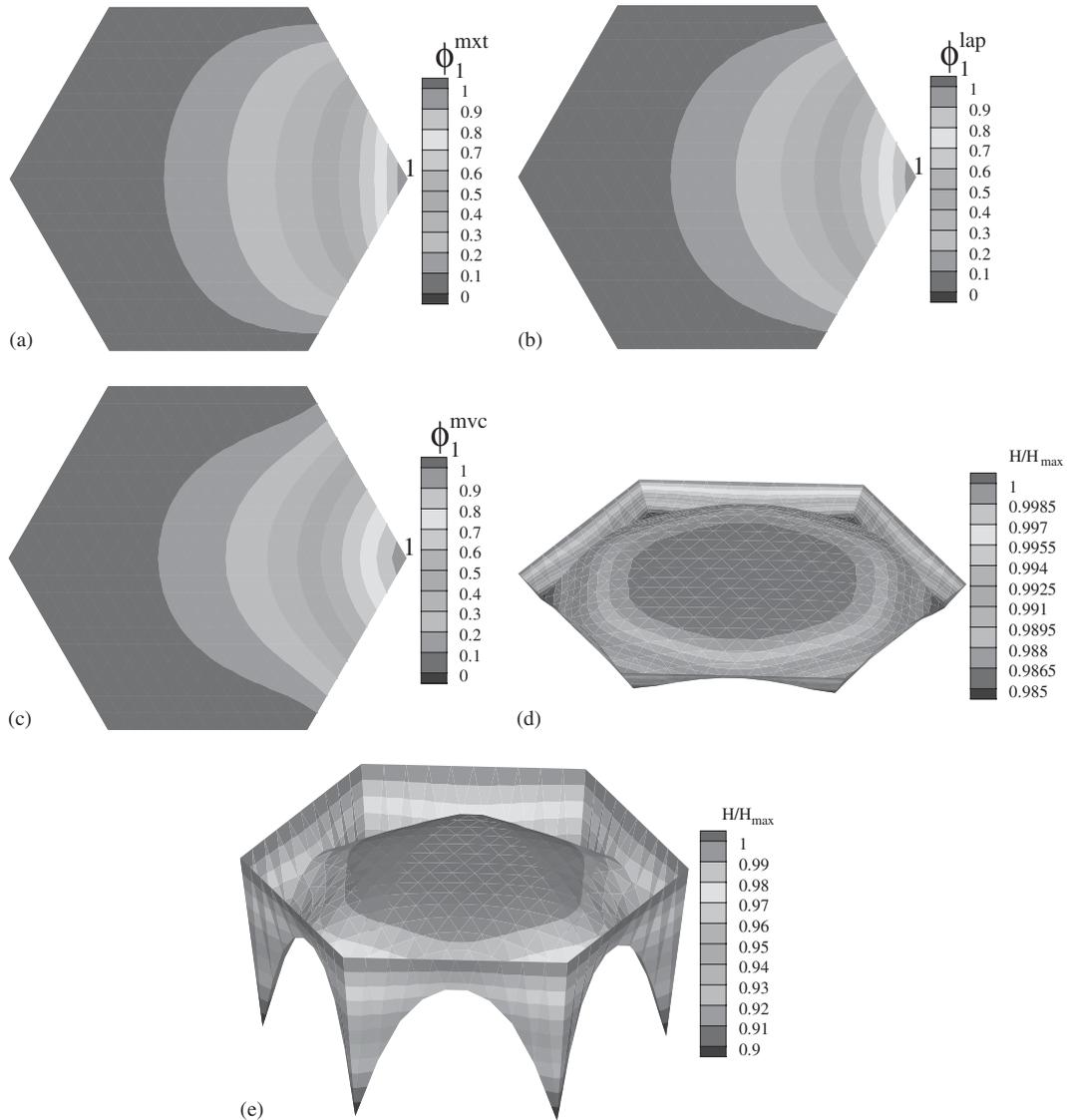


Figure 6. Shape function and entropy plots on a regular hexagon: (a)  $\phi_1^{\text{mxt}}$ ; (b)  $\phi_1^{\text{lap}}$ ; (c)  $\phi_1^{\text{mvc}}$ ; (d)  $H(\phi_1^{\text{lap}})/H_{\text{max}}$ ; and (e)  $H(\phi_1^{\text{mvc}})/H_{\text{max}}$ .

where  $(\cdot)_{,x} = \partial(\cdot)/\partial x$  and  $(\cdot)_{,y} = \partial(\cdot)/\partial y$ . Now, we need to determine  $\lambda_{r,x} \equiv \partial\lambda_r/\partial\langle g_1(\mathbf{x}) \rangle$  and  $\lambda_{r,y} \equiv \partial\lambda_r/\partial\langle g_2(\mathbf{x}) \rangle$ . To this end, we once again consider Equation (12)

$$-\frac{\partial(\log Z)}{\partial\lambda_r} = \langle g_r(\mathbf{x}) \rangle \quad (r = 1, 2) \tag{36}$$

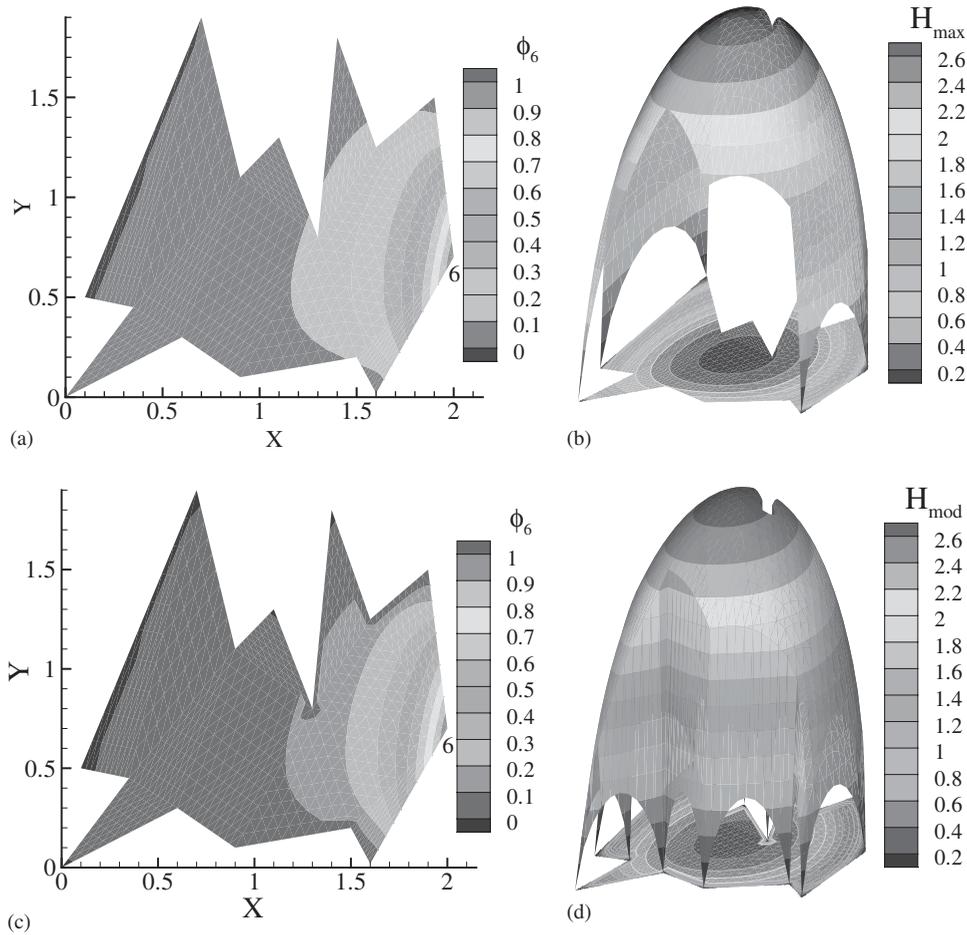


Figure 7. MAXENT solution in a non-convex polygon. Shape function  $\phi_6^{\text{mxt}}$  and distribution of entropy assuming: (a) and (b) interpolation along  $\partial\Omega_{\text{ch}}$ ; and (c) and (d) interpolation along  $\partial\Omega_e$ .

and on taking the derivative of the above equation with respect to  $\lambda_s$ , we obtain

$$\frac{\partial \langle g_r(\mathbf{x}) \rangle}{\partial \lambda_s} = \frac{\partial \langle g_s(\mathbf{x}) \rangle}{\partial \lambda_r} = -\frac{\partial^2 (\log Z)}{\partial \lambda_r \lambda_s} \equiv -\frac{\partial^2 F}{\partial \lambda_r \lambda_s} = -G_{rs} \tag{37}$$

where for notational convenience we have used  $F = \log Z$  (earlier  $F$  was set to  $\log \tilde{Z}$ ) and  $\mathbf{G}$  is the Hessian of  $\log Z(\lambda_1, \lambda_2)$ . On using the chain rule of differentiation, we can write

$$\frac{\partial \lambda_r}{\partial \langle g_s(\mathbf{x}) \rangle} = -[\mathbf{G}^{-1}]_{rs} \tag{38}$$

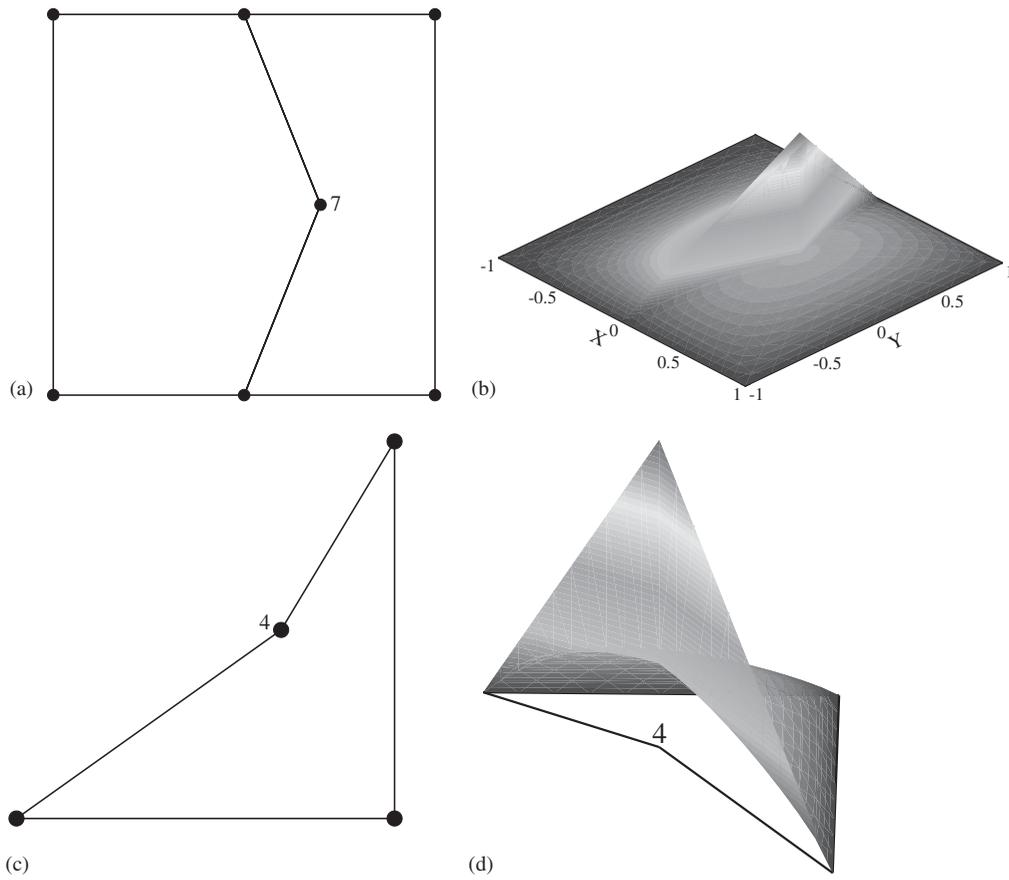


Figure 8. Modified MAXENT shape function in a non-convex polygonal mesh: (a) seven-node mesh; (b)  $\phi_7^{\text{mxt}}$ ; (c) four-node mesh; and (d)  $\phi_4^{\text{mxt}}$ .

and since  $\langle g_1(\mathbf{x}) \rangle = x$  and  $\langle g_2(\mathbf{x}) \rangle = y$ , we obtain

$$\begin{bmatrix} \frac{\partial \lambda_1}{\partial x} & \frac{\partial \lambda_1}{\partial y} \\ \frac{\partial \lambda_2}{\partial x} & \frac{\partial \lambda_2}{\partial y} \end{bmatrix} = - \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}^{-1} \tag{39}$$

Since  $F(\lambda_1, \lambda_2) = \log Z(\lambda_1, \lambda_2) = \log \left( \sum_i^n e^{-\lambda_1 x_i - \lambda_2 y_i} \right)$ , we can write its derivative with respect to  $\lambda_1$  as:

$$\frac{\partial F}{\partial \lambda_1} = \frac{\sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} (-x_i)}{Z} \tag{40}$$

or

$$G_{11} = \frac{\partial^2 F}{\partial \lambda_1^2} = \frac{Z \sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} x_i^2 + \left[ \sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} (x_i) \right] \left[ \sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} (-x_i) \right]}{Z^2} \tag{41}$$

which simplifies to

$$G_{11} = \frac{\sum_{i=1}^n e^{-\lambda_1 x_i - \lambda_2 y_i} x_i^2}{Z} - x^2 \tag{42}$$

and therefore

$$G_{11} = \sum_{i=1}^n \phi_i^{\text{mxt}} x_i^2 - x^2 \equiv \langle x^2 \rangle - x^2 \quad (\text{variance of } x) \tag{43}$$

Proceeding likewise for the other components of  $\mathbf{G}$ , we have

$$G_{12} = G_{21} = \langle xy \rangle - xy, \quad G_{22} = \langle y^2 \rangle - xy \tag{44}$$

Now we can write Equation (39) as

$$\begin{bmatrix} \frac{\partial \lambda_1}{\partial x} & \frac{\partial \lambda_1}{\partial y} \\ \frac{\partial \lambda_2}{\partial x} & \frac{\partial \lambda_2}{\partial y} \end{bmatrix} = - \begin{bmatrix} \langle x^2 \rangle - x^2 & \langle xy \rangle - xy \\ \langle xy \rangle - xy & \langle y^2 \rangle - y^2 \end{bmatrix}^{-1} \tag{45}$$

and on using the above relations in Equation (35), the derivatives of the MAXENT shape functions are computed.

To test the shape function derivatives, we use the MAXENT shape functions as trial and test approximations within a Galerkin method. The model problem (patch test) is:  $\nabla^2 u = 0$  in  $\Omega = (0, 1)^2$  with essential boundary conditions  $u(x, y) = x + y$  imposed on  $\partial\Omega$ . The two meshes shown in Figure 9 are considered, and  $2 \times 2$  quadrature rule is used for the numerical integration. From Section 2.1.2, the MAXENT solution should be identical to bilinear finite elements for the mesh depicted in Figure 9(a). The relative error in the  $L^2(\Omega)$  and energy norms for the mesh shown in Figure 9(a) were  $1.9 \times 10^{-8}$  and  $6.7 \times 10^{-8}$ , respectively, whereas those for the mesh in Figure 9(b) were  $1.3 \times 10^{-8}$  and  $4.5 \times 10^{-8}$ , respectively. Since the linear reproducing conditions are met with  $\mathcal{O}(10^{-7})$ – $\mathcal{O}(10^{-8})$  accuracy, the patch test is passed to within an expected precision, and hence we have numerically verified the earlier derivation of the shape function derivatives. Note that along  $a$ – $b$  and  $b$ – $c$  in Figure 9(b), both the MAXENT as well as the Laplace interpolant are linear, whereas quadratic interpolation is realized along  $a$ – $b$ – $c$  with finite elements. The linearity property of the Laplace shape functions has enabled the construction of conforming approximations along edges with hanging nodes on quadtree meshes [10].

#### 4.4. Surface approximation

Consider a domain  $\Omega \in \mathbf{R}^2$  that is described by  $N$  nodes with location  $\mathbf{x}_i$  ( $i = 1, 2, \dots, N$ ). Given a function  $u(\mathbf{x})$ , with its value  $u(\mathbf{x}_i)$  specified at the  $N$  nodes, we seek to find an interpolant  $u^h$  for  $u$  that fits the data. To this end, it suffices if we can construct an interpolant

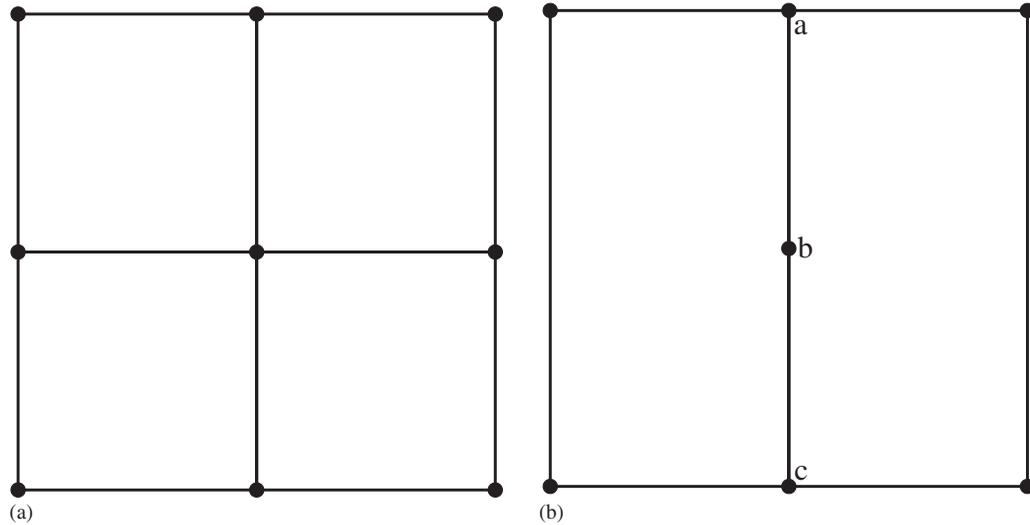


Figure 9. Meshes for the patch test: (a) regular; and (b) edge with three nodes.

within each polygon  $V_i$ , where  $\Omega = \cup V_i$ , and the global interpolant is then the additive sum of the local interpolant on each polygon. From approximation theory, we know that given independent elements  $\{x_i\}_{i=1}^n$  with  $a_i x_i \in X$  ( $X$  is a normed linear space), then the *best approximation* to an element  $y \in X$  is determined by  $\min_{a_i} \|y - \sum_{i=1}^n a_i x_i\|$ , i.e. the weights  $a_i$  are obtained by minimizing the  $L^2$  norm of the error [18].

If one considers the interpolation of random fields, then the above notion of best-approximation also applies to the MAXENT interpolant. Let  $\hat{u}$  be the approximation for  $u$  that minimizes the square of the expected error. Then, on using the MAXENT shape function in Equation (2), we have

$$\langle (u - \hat{u})^2 \rangle = \sum_{i=1}^n \phi_i^{\text{mxt}} (u_i - \hat{u})^2 \quad (46)$$

or

$$\langle (u - \hat{u})^2 \rangle = \langle u^2 \rangle - 2\hat{u}\langle u \rangle + \hat{u}^2 \quad (47)$$

and on adding and subtracting  $\langle u \rangle^2$  to the right-hand side, we obtain

$$\langle (u - \hat{u})^2 \rangle = \underbrace{\langle u^2 \rangle - \langle u \rangle^2}_{\text{variance of } u} + (\hat{u} - \langle u \rangle)^2 \quad (48)$$

and since the variance of  $u$  is a fixed quantity, the expected error is minimized if  $\hat{u} = \langle u \rangle$ , and therefore

$$\hat{u}(\mathbf{x}) = u^h(\mathbf{x}) = \sum_{i=1}^n \phi_i^{\text{mxt}} u_i \quad (49)$$

is the *least-biased* interpolant with respect to the location of the nodal co-ordinates  $\mathbf{x}_i$  and it also minimizes the mean-square expected error [2].

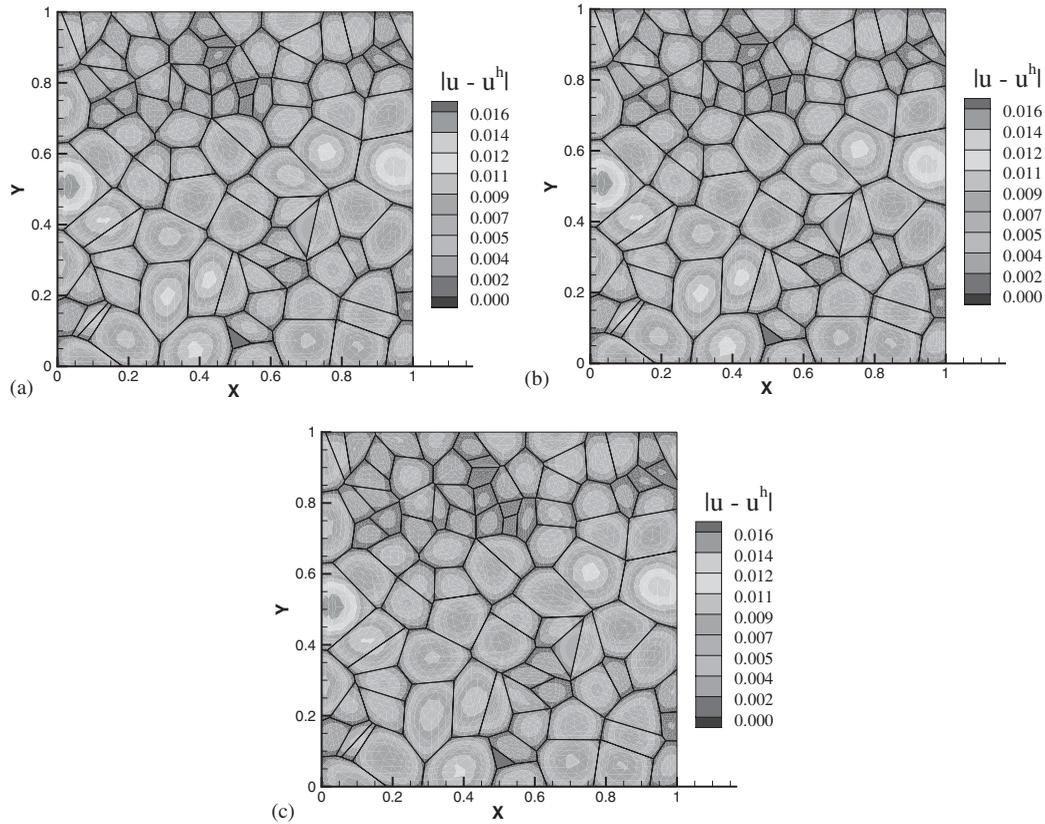


Figure 10. Point-wise absolute error for the interpolation of  $u(x, y) = x^2 + y^2$ : (a) MAXENT; (b) Laplace; and (c) mean-value interpolants.

We consider the problem of data interpolation to examine the accuracy of the MAXENT interpolant vis-à-vis some of the other polygonal interpolants. In the examples that follow, the quadratic function  $u(x) = x^2 + y^2$  is interpolated within the domain  $\Omega = (0, 1)^2$ . To construct a mesh, the seed points that define the Voronoi diagram were inserted using a random number generator that was based on a uniform probability distribution in  $(0, 1)$  [16]. If  $m$  points are inserted, a Voronoi mesh with  $m$  elements is obtained; the vertices of the Voronoi cells are the nodes. For comparison purposes, we measure the absolute accuracy using the discrete  $L^2(\Omega)$  norm, where  $\|u - u^h\|_{L^2}^2 = \sum_{i=1}^M (u(\mathbf{x}_i) - u^h(\mathbf{x}_i))^2$  and  $M$  is the total number of points chosen in the domain. The maximum error norm  $\max_i |u(\mathbf{x}_i) - u^h(\mathbf{x}_i)|_\infty$  is also computed. Points that are located on  $\partial\Omega_e$  are not used in the computations since the interpolants under consideration are identical along  $\partial\Omega_e$ . In Figure 10, the point-wise absolute error is presented for the MAXENT, Laplace, and mean-value interpolants; the mesh consists of 100 elements (202 nodes). Each  $n$ -gon is partitioned into  $16n$  triangles, and the interpolant is evaluated at the vertices of these triangles. The discrete  $L^2$  ( $L^\infty$ ) error norms that were obtained are: 0.4486 ( $1.556 \times 10^{-2}$ ), 0.4457 ( $1.559 \times 10^{-2}$ ), and 0.4124 ( $1.562 \times 10^{-2}$ ), respectively. The above trends were also

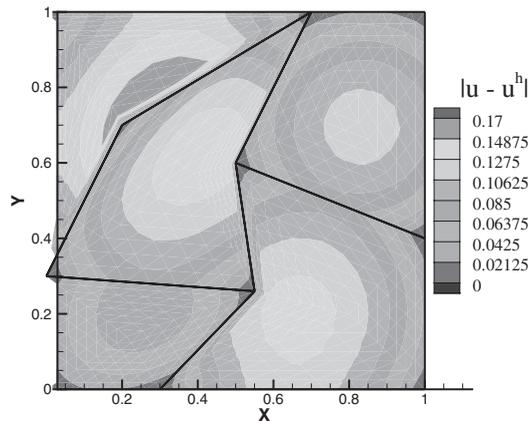


Figure 11. Point-wise absolute error for the modified MAXENT interpolant on a mesh containing non-convex elements.

consistent on other meshes that we considered. From the contour plots shown in Figure 10, the close correspondence between the results for the Laplace and MAXENT interpolants is once again noted.

Lastly, we study data interpolation using the MAXENT interpolant on a domain that contains non-convex polygons. In Figure 11, the contour plot of the absolute error is superimposed on the mesh. Of the five elements in the mesh, three are non-convex. Each  $n$ -gon is partitioned into  $64n$  triangles and the error computed at the vertices of these triangles. The maximum error norm (sup-norm) is 0.17 for the mesh shown in Figure 11.

## 5. CONCLUDING REMARKS

In this paper, we have presented the construction of polygonal interpolants using the principle of maximum entropy (MAXENT). On an  $n$ -gon ( $n > 3$ ), the shape functions  $\phi_i$  ( $i = 1, 2, \dots, n$ ) were the unknowns, and the satisfaction of the linear reproducing conditions was imposed on the interpolant. The non-negative shape functions  $\phi_i$  were determined by maximizing the information-theoretic entropy  $H(\phi_1, \phi_2, \dots, \phi_n) = -\sum_{i=1}^n \phi_i \log \phi_i$  subject to the above constraints. The maximum entropy (primal) formulation using the method of Lagrange multipliers was recast (dual problem) into one in which the minimum of a convex function was sought [14]. A line search algorithm was used in the numerical computations. On a rectangle, MAXENT interpolation and bilinear finite element interpolation were identical; however for arbitrarily shaped  $n$ -gons, the MAXENT formulation lead to different shape functions. The MAXENT formulation was proven to lead to a feasible solution for  $\phi_i$  in any convex or non-convex polygon. Among the interpolants that we considered, the entropy of the Laplace shape functions was found to be closest to  $H_{\max}$ , and this proximity was also observed in a data interpolation problem. This study has revealed a novel approach to construct polygonal interpolants via the application of the principle of maximum entropy. Its implications in the interpolation of random fields and the solution of partial differential equations, the use of MAXENT in inverse problems, topology

optimization and material design, and the natural link towards the development of multiscale numerical methods are deemed to be noteworthy.

## ACKNOWLEDGEMENTS

The author is grateful for the research support of the National Science Foundation through contract CMS-0352654 to the University of California, Davis. It is a pleasure to also acknowledge Professor Yueyue Fan for many helpful discussions on constrained optimization and generalized inverse.

## REFERENCES

1. Shannon CE. A mathematical theory of communication. *The Bell Systems Technical Journal* 1948; **27**:379–423.
2. Jaynes ET. Information theory and statistical mechanics. *Physical Review* 1957; **106**(4):620–630.
3. Jaynes ET. Information theory and statistical mechanics. II. *Physical Review* 1957; **108**(2):171–190.
4. Jaynes ET. On the rationale of maximum-entropy methods. *Proceedings of the IEEE* 1982; **70**(9):939–952.
5. Rosenkrantz RD. (ed.). *E.T. Jaynes, Paper on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
6. Khinchin A. *Mathematical Foundations of Information Theory*. Dover: New York, NY, 1957.
7. Kapur JN. *Maximum-Entropy Models in Science and Engineering* (1st (revised) edn). Wiley: New Delhi, India, 1993.
8. Beltzer AI. Entropy characterization of finite elements. *International Journal of Solids and Structures* 1996; **33**(24):3549–3560.
9. Sukumar N, Tabarraei A. Conforming polygonal finite elements. *International Journal for Numerical Methods in Engineering* 2004; **61**:2045–2066.
10. Tabarraei A, Sukumar N. Adaptive computations on conforming quadtree meshes. *Finite Elements in Analysis and Design*, 2005, to appear (Special Issue on Papers Presented at the Robert J. Melosh Finite Element Medal Competition.)
11. Jaynes ET. Concentration of distributions at entropy maxima. In *E.T. Jaynes, Paper on Probability, Statistics and Statistical Physics*, Rosenkrantz RD (ed.). Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989; 317–336.
12. Penrose R. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society* 1955; **51**:406–413.
13. Agmon N, Alhassid Y, Levine RD. An algorithm for determining the Lagrange parameters in the maximal entropy formalism. In *The Maximum Entropy Formalism*, Tribus M, Levine RD (eds). MIT Press: Cambridge, MA, 1978; 206–209.
14. Agmon N, Alhassid Y, Levine RD. An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics* 1979; **30**:250–258.
15. Alhassid Y, Agmon N, Levine RD. An upper bound for the entropy and its application to the maximal entropy problem. *Chemical Physics Letters* 1978; **53**(1):22–26.
16. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical Recipes in Fortran: The Art of Scientific Computing* (2nd edn). Cambridge University Press: New York, NY, 1992.
17. Fang S-C, Rajasekera JR, Tsao HSJ. *Entropy Optimization and Mathematical Programming*. Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997.
18. Davis PJ. *Interpolation and Approximation*. Dover: New York, NY, 1975.