



ELSEVIER

Computational Geometry 14 (1999) 49–65

Computational
Geometry

Theory and Applications

www.elsevier.nl/locate/comgeo

Optimal triangulation and quadric-based surface simplification

Paul S. Heckbert*, Michael Garland¹

Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3891, USA

Abstract

Many algorithms for reducing the number of triangles in a surface model have been proposed, but to date there has been little theoretical analysis of the approximations they produce. Previously we described an algorithm that simplifies polygonal models using a quadric error metric. This method is fast and produces high quality approximations in practice. Here we provide some theory to explain why the algorithm works as well as it does. Using methods from differential geometry and approximation theory, we show that in the limit as triangle area goes to zero on a differentiable surface, the quadric error is directly related to surface curvature. Also, in this limit, a triangulation that minimizes the quadric error metric achieves the optimal triangle aspect ratio in that it minimizes the L_2 geometric error. This work represents a new theoretical approach for the analysis of simplification algorithms. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Triangle aspect ratio; Curvature; Approximation theory; Anisotropic mesh generation; Quadric error metric

1. Introduction

The simplification of detailed geometric surface models is important for a number of applications. A typical simplification algorithms—the type we will focus on in this paper—takes a polygonal model as input and produces an approximation composed of fewer triangles that preserves surface shape. Simplification algorithms are an important component in the creation of multiresolution models, models that represent the geometry and other attributes of an object at multiple levels of detail. By reducing a model's size, we can accelerate programs that subsequently process the data, cut storage space and network bandwidth requirements, and decrease the time required to display the model. Natural application areas include computer-aided design, architectural walkthroughs, finite element methods, scientific visualization, shape acquisition, graphics on the Web, movie special effects, virtual reality, and video games.

* Corresponding author. E-mail: ph@cs.cmu.edu

¹ E-mail: garland@uiuc.edu

1.1. Optimal approximation

Ideally, we would like to find the *optimal approximation*: the triangulated surface with a given number of triangles that has the least error relative to the original model. We will use the L_2 measure of geometric error as our ideal error metric. While perceptual error metrics might be more desirable for display applications, they are also significantly harder to analyze. Even with a purely geometric error metric, optimal approximation is not feasible, in general. Computing the optimal approximation of a surface with respect to the L_∞ metric is NP-hard [1]; finding such an optimal approximation requires time exponential in the number of vertices.

To date, little has been proven about the optimality of existing surface approximation algorithms. The existing results are narrow in scope. Some of the few are polynomial-time algorithms to find approximations to height fields or convex polytopes that are within a factor of optimal [1]. For surfaces more general than height fields and convex polytopes, there are simplification algorithms with bounded error (e.g., [4]), but the number of triangles in their approximations are not bounded, so such algorithms are not optimal in a strong sense. The authors are not aware of any polynomial time algorithms that generate approximations to general surfaces that are provably good in both error and number of triangles.

Even simpler versions of the surface approximation problem have only limited theoretical results to date. For example, optimal approximation of a sphere by a triangulated surface is related to optimal packing of n equal circles on a sphere. Although the latter problem has been studied for decades, solutions are known only for small n ($n < 200$ or so) [5]. Little has been proven about optimality for arbitrary surfaces.

1.2. Overview

Previously we described an algorithm for surface simplification based on iterative edge contraction and quadric error metrics [10,11]. This algorithm is fast and achieves good quality results in practice. Simplifying a manifold surface model with n vertices has an $O(n \log n)$ running time [9]. In this paper, we analyze its approximation errors.

Our principal tools in this analysis are differential geometry, which provides techniques for analyzing surface curvature, and approximation theory, which provides methods for analyzing approximation errors. Approximation theorists have determined the L_2 -optimal shape of triangles when approximating bivariate functions. They found that the aspect ratio (length/width) of an optimal triangle is related to the second derivatives of the function, i.e., its curvature [16] (see Section 3.2 for more detail). As is typical in approximation theory, this analysis is done in the limit as triangle area goes to zero.

In this paper we attempt to interrelate practical surface simplification methods from computer graphics with theoretical, asymptotic results from approximation theory. Specifically, we pose the questions: why does the quadric-based algorithm work as well as it does? And how close to optimal are the approximations generated by this algorithm?

We show that minimization of the quadric error metric computes curvature information indirectly, and that minimization of this metric yields, in the limit of small triangles, for differentiable surfaces, a triangulation with optimal triangle shape. Note that this does not imply that the quadric-based algorithm yields optimal approximations for finite problems (practical problems employing a finite number of triangles). Nevertheless, it validates that the algorithm is theoretically “on the right track” in the sense

that as the original mesh becomes finer and finer, the resulting approximation will become more nearly optimal, subject to suitable assumptions.

The remaining sections of the paper are organized as follows. First, we review the quadric-based simplification algorithm, along with relevant concepts from differential geometry and approximation theory. Next, we derive the quadric error metric for a differentiable manifold, and we prove that minimization of the quadric error metric generates a triangulation with optimal triangle shape, in the limit. Finally, we check this empirically, and present conclusions.

2. Quadric-based simplification

Given an initial triangulated surface, we want to automatically generate an approximation with fewer triangles that is faithful to the original geometry. Our simplification algorithm [10,11] is based on iterative edge contraction, a framework used by several others as well [19,14,12]. Every edge is assigned a “cost” that is typically meant to reflect the geometric error introduced into the model as a result of contracting the edge. A greedy approach is used: on each iteration, the lowest-cost edge is contracted, and the costs of neighboring edges are updated. An edge contraction, which we denote $(v_i, v_j) \rightarrow \bar{v}$, modifies the surface by unifying two vertices into one, thereby removing one vertex and two faces (see Fig. 1). The primary difference between the various contraction-based methods is the error metric used to assign costs to edges.

While our algorithm is designed to accommodate non-manifold surfaces, for our purpose of analyzing its theoretical properties, we will assume that the input surface is a closed manifold. In other words, every point on the surface has a neighborhood that is homeomorphic to a disk. We will also assume in this paper that the topology of the surface is preserved during simplification.

2.1. Quadric error metric

Suppose that we are given a plane determined by a point \mathbf{p} and a unit normal \mathbf{n} . The squared distance of any point \mathbf{v} to this plane is given by

$$((\mathbf{v} - \mathbf{p}) \cdot \mathbf{n})^2 = \mathbf{v}^T \mathbf{n} \mathbf{n}^T \mathbf{v} - 2(\mathbf{n} \mathbf{n}^T \mathbf{p})^T \mathbf{v} + \mathbf{p}^T \mathbf{n} \mathbf{n}^T \mathbf{p}. \quad (1)$$

This is a quadratic function of \mathbf{v} . More generally, to efficiently compute the weighted sum of squared distances to a set of planes, we use a *quadric error metric* of the form

$$Q(\mathbf{v}) = \mathbf{v}^T \mathbf{A} \mathbf{v} + 2\mathbf{b}^T \mathbf{v} + c, \quad (2)$$

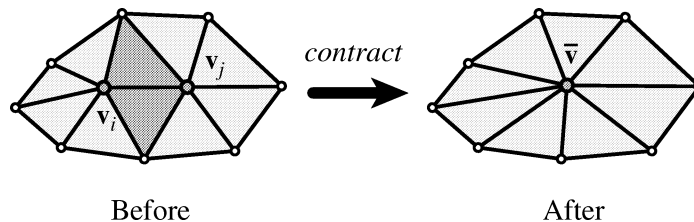


Fig. 1. Edge (v_i, v_j) is contracted. The darker triangles become degenerate and are removed.

where the *quadric* Q is given by

$$Q = (\mathbf{A}, \mathbf{b}, c). \quad (3)$$

We call it a quadric error metric because the isosurfaces of $Q(\mathbf{v})$ are quadric surfaces. This requires 10 coefficients to store the symmetric 3×3 matrix \mathbf{A} , the 3-vector \mathbf{b} and the scalar c .

Every vertex \mathbf{v} in the original model has a set of adjacent faces $\{f_1, \dots, f_k\}$. Each face f_i has a unit normal \mathbf{n}_i that, together with any point \mathbf{p}_i in the plane of that face, determines a *fundamental quadric*

$$Q_i = (\mathbf{A}_i, \mathbf{b}_i, c_i) = (\mathbf{n}_i \mathbf{n}_i^T, -\mathbf{A}_i \mathbf{p}_i, \mathbf{p}_i^T \mathbf{A}_i \mathbf{p}_i). \quad (4)$$

We define the initial quadric Q associated with the vertex \mathbf{v} to be the weighted sum of these fundamental quadrics:

$$Q = \sum_{i=1}^k w_i Q_i. \quad (5)$$

In this paper we use area-weighting, where w_i is the area of face f_i , but in other contexts other weighting schemes may be preferred. The value $Q(\mathbf{v})$ is the area-weighted sum of squared distances of \mathbf{v} to the planes of its neighboring triangles. Note that, since the vertex \mathbf{v} necessarily lies at the intersection of all these planes, the error associated with every vertex on the original model is 0.

We define the cost of the contraction $(\mathbf{v}_i, \mathbf{v}_j) \rightarrow \bar{\mathbf{v}}$ to be $Q_i(\bar{\mathbf{v}}) + Q_j(\bar{\mathbf{v}}) = (Q_i + Q_j)(\bar{\mathbf{v}})$. The minimum of this function occurs where $\nabla Q(\mathbf{v}) = 2\mathbf{A}\mathbf{v} + 2\mathbf{b} = 0$. Solving this equation, we find that the optimal position is

$$\bar{\mathbf{v}} = -\mathbf{A}^{-1}\mathbf{b} \quad (6)$$

and its error is

$$Q(\bar{\mathbf{v}}) = \mathbf{b}^T \bar{\mathbf{v}} + c = -\mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} + c. \quad (7)$$

2.2. Quadric properties

For a given quadric Q , the level surface $Q(\mathbf{v}) = \varepsilon$ is the set of all points whose error with respect to Q is ε . That is, it is the locus of points to which a vertex can be relocated with constant error. This isosurface is a (potentially degenerate) ellipsoid whose principal axes are defined by the eigenvalues and eigenvectors of the matrix \mathbf{A} [9]. The ellipsoids are degenerate, or “open,” when some eigenvalues of \mathbf{A} are 0; in other words, when \mathbf{A} is singular. The equation for finding the optimal position $\bar{\mathbf{v}}$ corresponds to finding the center of the ellipsoid isosurfaces. When the ellipsoids are degenerate (i.e., \mathbf{A} is non-invertible), the isosurfaces are either infinite cylinders (one 0 eigenvalue) or pairs of parallel planes (two 0 eigenvalues).

Fig. 2 illustrates the quadric isosurfaces produced by the simplification of a bunny model. Notice that the quadrics characterize the local shape of the surface. For vertices on creases, such as on the neck and ears, the ellipsoids are cigar shaped. They are elongated in the direction of the crease. In contrast, where the surface is less curved, such as on the forehead, the quadrics are thin and roughly circular, like pancakes. Intuitively, we might conclude that the quadrics will be elongated in directions of low curvature and thin in directions of high curvature. In Section 4, we quantify this hypothesis.

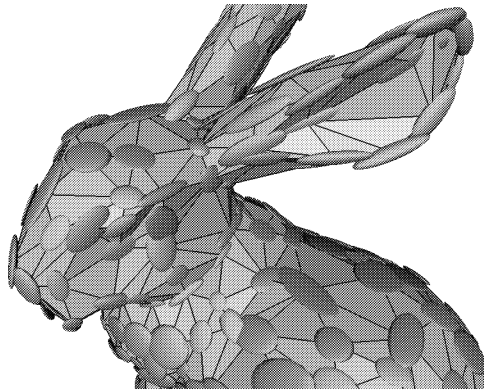


Fig. 2. Simplified bunny model with a visualization of the quadrics used for its construction. Only 1.4% of the original 70,000 faces remain. Centered around each vertex is an isosurface of the corresponding quadric (a.k.a. Riemannian metric tensor).

2.3. Neighborhoods

After repeated edge contractions, the quadric associated with each vertex of the approximate model is the sum of the fundamental quadrics from a connected neighborhood of nearby vertices from the original model (Fig. 7). On smooth surfaces, these neighborhoods are fairly regular in shape (roughly elliptical, typically elongated in the direction of lower curvature), but on more complex surfaces they can be gerrymandered.

Summation of the quadric matrices during edge contractions is equivalent to a neighborhood merge, causing some fundamental quadrics to be multiply-counted. The number of times a given triangle's fundamental quadric is counted in a given neighborhood is equal to the number of that triangle's vertices that are inside the neighborhood. Thus, perimeter faces are counted once or twice, while faces interior to the neighborhood are counted thrice. No vertices are counted more than three times. In all cases they are area-weighted. Although it may appear undesirable, multiple counting has not been found to be a problem in practice.

3. Background

In this section we review results from differential geometry, approximation theory, and mesh generation that we make use of in Section 4.

3.1. Differential geometry

We will employ the theory of local differential geometry [15,2,13] to analyze the mathematical properties of the quadric error metric.

A smooth surface patch is defined by

$$\mathbf{x} = \mathbf{x}(u, v) = [f_1(u, v) \ f_2(u, v) \ f_3(u, v)]^T, \quad (8)$$

where $(u, v) \in \mathbb{R}^2$ and the functions f_i are of class C^2 . We shall be concerned with the surface in the neighborhood of a point $\mathbf{p} = \mathbf{x}(u_0, v_0)$.

3.1.1. Tangents

The partial derivatives of \mathbf{x} ,

$$\mathbf{x}_1 = \mathbf{x}_u = \frac{\partial \mathbf{x}}{\partial u} \quad \text{and} \quad \mathbf{x}_2 = \mathbf{x}_v = \frac{\partial \mathbf{x}}{\partial v}, \quad (9)$$

evaluated at \mathbf{p} span the tangent plane of the surface at \mathbf{p} , so any tangent vector \mathbf{t} at \mathbf{p} can be written as $\mathbf{t} = \mathbf{x}_1 \delta u + \mathbf{x}_2 \delta v$. Consequently, we can parameterize this tangent vector by a direction vector in the 2-D parameter space: $\mathbf{u} = [\delta u \ \delta v]^T$. The unit surface normal \mathbf{n} at the point \mathbf{p} is given by

$$\mathbf{n} = \frac{\mathbf{x}_1 \times \mathbf{x}_2}{\|\mathbf{x}_1 \times \mathbf{x}_2\|} \quad (10)$$

provided that $\mathbf{x}_1 \times \mathbf{x}_2 \neq 0$. Note that, by convention, all functions such as \mathbf{x}_1 are implicitly evaluated at the point \mathbf{p} under consideration.

The length of a tangent vector can be defined in terms of the matrix

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, \quad \text{where } g_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j, \quad (11)$$

with determinant $g = g_{11}g_{22} - g_{12}^2$. The squared length of a tangent vector in unit direction \mathbf{u} is given by the *first fundamental form* $\mathbf{u}^T \mathbf{G} \mathbf{u}$. Such a measure, a second degree function of direction, defined for each point on a manifold, is called a *Riemannian metric tensor* [15].

3.1.2. Curvature

Geometrically, surface curvature is defined in terms of the intersection curve of the surface and a plane passing through the normal and the tangent vector in the direction \mathbf{u} at that point. The *normal curvature* of the surface in the direction \mathbf{u} is then the reciprocal of the radius of the osculating circle at that point. Curvatures can be positive or negative depending on the sign of the normal vector. Zero curvature means the surface is flat (in a particular direction).

Algebraically, curvature can be quantified in terms of the matrix

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad \text{where } b_{ij} = \mathbf{n} \cdot \mathbf{x}_{ij} = -\mathbf{n}_i \cdot \mathbf{x}_j. \quad (12)$$

The change in the normal vector \mathbf{n} in the unit direction \mathbf{u} , also known as the *second fundamental form*, is $\mathbf{u}^T \mathbf{B} \mathbf{u}$.

Together, the two fundamental forms allow one to express the normal curvature κ_n in the direction \mathbf{u} as

$$\kappa_n = \frac{\mathbf{u}^T \mathbf{B} \mathbf{u}}{\mathbf{u}^T \mathbf{G} \mathbf{u}}. \quad (13)$$

Unless the curvature is equal in all directions, there must be a direction \mathbf{e}_1 in which the normal curvature reaches a minimum and a direction \mathbf{e}_2 in which it reaches a maximum. These are called *principal directions*. The corresponding *principal curvatures* κ_1, κ_2 at point \mathbf{p} are the eigenvalues of the Weingarten map $\mathbf{G}^{-1} \mathbf{B}$.

3.2. Approximation theory

Approximation theory analyzes the errors of function approximation. When working with surfaces, one generally studies the limit as the areas of the approximating elements (in our case, triangles) vanish, since it is much easier to prove properties of approximations for the limit than for finite approximations. To ensure that these limits are defined, we assume that the function is twice differentiable.

Researchers have studied the effect of triangle size and shape on approximations to a bivariate function or height field $f(u, v)$. We will quantify error using the L_2 metric, which is the square root of the integral of the squared difference between two functions. Under this metric, one can ask: what triangulation with a given number of triangles minimizes the error of piecewise linear approximation? The asymptotic answer, discovered by Nadler, is that as the number of triangles goes to infinity, an optimal triangles' orientation is given by the eigenvectors of the Hessian of the function at each point, and their size in each principal direction is given by the reciprocal square root of the absolute value of the corresponding eigenvalue [16].

3.2.1. Aspect ratio

The *aspect ratio* of a rectangle is simply its width divided by its height. The aspect ratio of a triangle is a bit more complex. It can be defined in various ways, most nearly equivalent. We define the aspect ratio of a triangle by finding the ellipse of least area through the three vertices, and take the ratio of major to minor axes. The aspect ratio of an equilateral triangle is thus 1.

Nadler thus found that an optimal triangle's aspect ratio is

$$\rho = \left| \frac{\lambda_2}{\lambda_1} \right|^{1/2}, \quad (14)$$

where $\{\lambda_i\}$ are the eigenvalues of the Hessian. The Hessian of a function $f(u, v)$ is the matrix

$$\mathbf{H} = \begin{bmatrix} f_{uu} & f_{uv} \\ f_{vu} & f_{vv} \end{bmatrix}. \quad (15)$$

When $\det \mathbf{H} > 0$, the aspect ratio of (14) is the unique optimum for all L_p norms with $p \geq 1$ [7, 18]. This case coincides with a positive Gaussian curvature, if we regard f as a surface in 3-D. When $\det \mathbf{H} < 0$, the L_2 -optimal aspect ratio is not unique; there is a one-parameter family of solutions generated by stretching (14) along one of the directions of zero curvature [16, Eq. (3)]. The L_∞ -optimal aspect ratios differ from (14) by a small factor [7].

Long, thin “sliver” triangles can be bad in certain contexts; for instance, they can lead to large condition numbers in the matrices used for certain finite element simulations. Equilateral triangles are desirable in such contexts. But for our goal, deriving an approximation with minimal geometric error, slivers can be optimal.

We define *optimal triangulation* to be a triangulation that conforms to the above law, in the limit as the number of triangles goes to infinity and their areas go to zero.

3.3. Mesh generation

Two dimensional mesh generation is the subdivision of a 2-D domain into triangles or quadrilaterals. In many cases, meshes are used for finite element analysis, as in the solution of partial differential

equations. Adaptive meshing techniques alternate solution of a system of equations with remeshing of the domain. Some of these methods strive for optimal triangulations during remeshing using the Hessian of an approximate solution function to control triangle size and shape [17,3].

The intentional generation of stretched triangles is called *anisotropic mesh generation* [20]. This is often done using the Hessian to construct a Riemannian metric tensor that gives the desired edge length as a function of direction. A mesh generation algorithm yielding asymptotically optimally stretched triangles in this manner was given by D’Azevedo [6], but his method is restricted to structured meshes and a very small space of surfaces (vertex degree 6 and zero Riemann–Christoffel tensor everywhere).

Mesh generation methods have been employed to create simplification algorithms by appropriate definition of the desired edge length function. Frey used numerical estimates of surface curvature to construct an isotropic Riemannian metric tensor, and then used this to control a mesh generator [8]. This method did not generate anisotropic meshes, however.

Our quadric error metric can be regarded as an anisotropic Riemannian metric tensor, and it is applicable to unstructured meshes and general surfaces.

4. Analysis of quadric metric

We now relate our quadric error metric to the optimal triangulation results by analyzing its properties in the limit as the areas of the triangles go to zero. More precisely, we imagine a case of a twice-differentiable manifold from which original models can be constructed by tessellating it with specified edge lengths. There will be two limit processes. The first limit will drive the number of triangles of the original model to infinity while driving their areas to zero. We then sum the fundamental quadrics within a neighborhood around a surface point. In the limit as original triangle size goes to zero, the sum becomes an integral. This yields a formula for the infinitesimal error quadric as a function of surface curvature and neighborhood shape. The second limit will drive the area of these neighborhoods to zero.

We prove that, in these limits, the quadric error is minimized by triangulations with optimal aspect ratio. We also derive a quantitative relationship between the error quadrics and surface curvature.

4.1. Theoretical quadric error metric

In order to analyze the quadric error metric, we consider its behavior on a differentiable manifold M defined by a patch \mathbf{x} . Suppose that we are given a point of interest \mathbf{p}_0 on M with surface normal \mathbf{n}_0 (Fig. 3). Let $\mathbf{e}_1, \mathbf{e}_2$ be the principal directions at \mathbf{p}_0 , and let κ_1, κ_2 be the corresponding principal curvatures. If \mathbf{p}_0 is an umbilic point (i.e., κ_n is equal in all directions), it is sufficient to pick two arbitrary, orthogonal “principal” directions. In the coordinate frame $\mathbf{e}_1, \mathbf{e}_2, \mathbf{n}_0$, we can approximate the neighborhood of M around \mathbf{p}_0 to second degree by a surface patch [15] of the form

$$\mathbf{p}(u, v) = \left[u \ v \ \frac{1}{2}(\kappa_1 u^2 + \kappa_2 v^2) \right]^T. \quad (16)$$

This can be either an elliptical or hyperbolic paraboloid. Here $\mathbf{p}_0 = \mathbf{p}(0, 0)$ and the axes (u, v) coincide with the principal axes $\mathbf{e}_1, \mathbf{e}_2$. Such a coordinate frame exists for any point on our manifold. Use of this frame simplifies the derivation substantially.

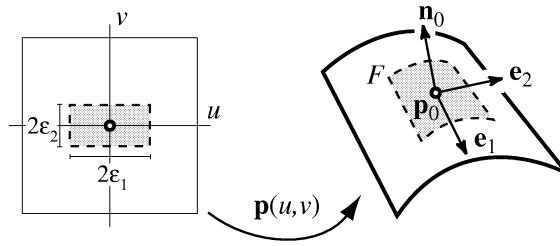


Fig. 3. Local parameterization of the surface about \mathbf{p}_0 . The neighborhood F is the projection of a rectangular region of the parameter domain onto the surface. This patch of surface is approximated by $\mathbf{p}(u, v)$.

For this surface, the matrix of the first fundamental form at \mathbf{p} is

$$\mathbf{G} = \begin{bmatrix} 1 + \kappa_1^2 u^2 & \kappa_1 \kappa_2 uv \\ \kappa_1 \kappa_2 uv & 1 + \kappa_2^2 v^2 \end{bmatrix}, \quad g = 1 + \kappa_1^2 u^2 + \kappa_2^2 v^2, \quad (17)$$

and the unit surface normal is $\mathbf{n} = \mathbf{m} / \sqrt{g}$, written in terms of the non-unit normal $\mathbf{m} = \mathbf{p}_1 \times \mathbf{p}_2 = [-\kappa_1 u \ -\kappa_2 v \ 1]^T$. It is easy to verify that at $\mathbf{p} = \mathbf{p}_0$, the matrix $\mathbf{G}^{-1} \mathbf{B}$ has eigenvalues κ_1 and κ_2 .

For the sake of simplicity, let us assume that the small neighborhood F around the point \mathbf{p}_0 (Fig. 3) has the rectangular parameter domain $-\varepsilon_1 \leq u \leq \varepsilon_1, -\varepsilon_2 \leq v \leq \varepsilon_2$. An elliptical domain could also be used, and it would yield identical results to first order. We will leave the size and aspect ratio of this rectangle unspecified for now; later we will determine the values that minimize the quadric error metric.

Every point \mathbf{p} in the vicinity of \mathbf{p}_0 has a unique tangent plane from which we can construct a quadric. Just as a vertex accumulates a sum of quadsrics during simplification, we shall consider the result of the point \mathbf{p}_0 accumulating the fundamental quadsrics of all infinitesimal triangles in F . We will not attempt to simulate multiple-counting; its effect on this limit process is negligible.

In the limit as the triangles of the original model go to zero area, the sum of area-weighted fundamental quadsrics of the infinitesimal triangles given by (4) and (5) becomes a surface integral over F . The quadric at point \mathbf{p}_0 will therefore have components

$$\mathbf{A} = \iint_F \mathbf{nn}^T dA, \quad (18)$$

$$\mathbf{b} = \iint_F -\mathbf{nn}^T \mathbf{p} dA, \quad (19)$$

$$c = \iint_F \mathbf{p}^T \mathbf{nn}^T \mathbf{p} dA, \quad (20)$$

where integration of matrices and vectors is defined by integrating each scalar component separately.

Let us focus on the matrix \mathbf{A} . Making the substitutions $dA = \sqrt{g} du dv$ and $\mathbf{n} = \mathbf{m} / \sqrt{g}$, it simplifies to

$$\mathbf{A} = \iint \frac{\mathbf{mm}^T}{\sqrt{g}} du dv. \quad (21)$$

The matrix

$$\mathbf{m}\mathbf{m}^T = \begin{bmatrix} \kappa_1^2 u^2 & \kappa_1 \kappa_2 uv & -\kappa_1 u \\ \kappa_1 \kappa_2 uv & \kappa_2^2 v^2 & -\kappa_2 v \\ -\kappa_1 u & -\kappa_2 v & 1 \end{bmatrix} \quad (22)$$

is easy to integrate by itself, but not with \sqrt{g} in the denominator. To tackle this problem, we use the Taylor series approximation

$$\frac{1}{\sqrt{g}} = 1 - \frac{1}{2}\kappa_1^2 u^2 - \frac{1}{2}\kappa_2^2 v^2 + O(u^4 + u^2 v^2 + v^4). \quad (23)$$

In the limit of infinitesimal neighborhoods, the fourth and higher degree terms become negligible. Using this approximation,

$$\mathbf{A} = \int_{-\varepsilon_2}^{\varepsilon_2} \int_{-\varepsilon_1}^{\varepsilon_1} \mathbf{m}\mathbf{m}^T (1 - \frac{1}{2}\kappa_1^2 u^2 - \frac{1}{2}\kappa_2^2 v^2) du dv. \quad (24)$$

If this integral is evaluated, dropping terms of degree six or higher in ε_1 and ε_2 , a diagonal matrix results, with entries

$$a_{11} = \frac{4}{3}\varepsilon_1^3 \varepsilon_2 \kappa_1^2, \quad (25)$$

$$a_{22} = \frac{4}{3}\varepsilon_1 \varepsilon_2^3 \kappa_2^2, \quad (26)$$

$$a_{33} = 4\varepsilon_1 \varepsilon_2 - \frac{2}{3}\varepsilon_1 \varepsilon_2 (\varepsilon_1^2 \kappa_1^2 + \varepsilon_2^2 \kappa_2^2). \quad (27)$$

Since \mathbf{A} is diagonal, these are also its eigenvalues, and the eigenvectors are the two principal directions and the surface normal. These formulas are approximate for finite neighborhoods, and become exact in the limit as the neighborhood size parameters ε_1 and ε_2 go to zero.

Following a similar procedure, we can evaluate the integrals for \mathbf{b} and \mathbf{c} :

$$\mathbf{b} = [0 \quad 0 \quad \frac{2}{3}\varepsilon_1 \varepsilon_2 (\varepsilon_1^2 \kappa_1 + \varepsilon_2^2 \kappa_2)]^T, \quad (28)$$

$$\mathbf{c} = [\frac{1}{5}\kappa_1^2 \varepsilon_1^5 \varepsilon_2 + \frac{2}{9}\kappa_1 \kappa_2 \varepsilon_1^3 \varepsilon_2^3 + \frac{1}{5}\kappa_2^2 \varepsilon_1 \varepsilon_2^5]. \quad (29)$$

We now have a complete quadric Q . Applying the formula for the optimal vertex position $\bar{\mathbf{v}} = -\mathbf{A}^{-1}\mathbf{b}$, we find that

$$\bar{\mathbf{v}} = [0 \quad 0 \quad -\frac{1}{6}(\kappa_1 \varepsilon_1^2 + \kappa_2 \varepsilon_2^2)]^T, \quad (30)$$

and its error with respect to Q is

$$Q(\bar{\mathbf{v}}) = \frac{4}{45}(\kappa_1^2 \varepsilon_1^5 \varepsilon_2 + \kappa_2^2 \varepsilon_1 \varepsilon_2^5). \quad (31)$$

4.2. Theoretical aspect ratio

We now know the parameters of the quadric at any point as a function of the principal curvatures κ_1 and κ_2 and the neighborhood size $2\varepsilon_1 \times 2\varepsilon_2$. The former are determined by the original surface, but the

latter are properties of the neighborhoods. We must eliminate these latter variables to make a complete analysis.

So we push further and ask: what neighborhood shape is optimal, and what does this tell us about the quadric error metric and the shape of triangles in the approximation? We restrict ourselves to rectangular neighborhoods oriented parallel to the principal directions. We show that, in the limit as neighborhood area goes to zero, minimizing the quadric error metric generates triangles with optimal aspect ratio.

To find the neighborhood aspect ratio that minimizes error, we take the expression for minimum quadric error (31) and reparameterize it in terms of *aspect ratio* $\rho = \varepsilon_1/\varepsilon_2$ and mean size $\varepsilon = \sqrt{\varepsilon_1\varepsilon_2}$. Substituting $\varepsilon_1 = \varepsilon\rho^{1/2}$ and $\varepsilon_2 = \varepsilon\rho^{-1/2}$ yields

$$Q(\bar{v}) = \frac{4}{45}\varepsilon^6(\kappa_1^2\rho^2 + \kappa_2^2\rho^{-2}). \quad (32)$$

Now, let us fix the area by holding the size parameter ε constant, and find the aspect ratio ρ that minimizes $Q(\bar{v})$. This occurs when

$$\frac{\partial Q}{\partial \rho}(\bar{v}) = \frac{4}{45}\varepsilon^6(2\kappa_1^2\rho - 2\kappa_2^2\rho^{-3}) = 0. \quad (33)$$

Solving for ρ , we find that minimization of the quadric error metric yields neighborhoods with limiting aspect ratio

$$\rho = \left| \frac{\kappa_2}{\kappa_1} \right|^{1/2}. \quad (34)$$

We can show that the aspect ratio (34) that results from minimizing the quadric error metric agrees with the optimum determined by Nadler. Because $\mathbf{p}(u, v)$ has the simple form (16), the Hessian of its third coordinate at \mathbf{p}_0 is a diagonal matrix with eigenvalues $\lambda_1 = \kappa_1$ and $\lambda_2 = \kappa_2$. Therefore ε_i should be proportional to $|\kappa_i|^{-1/2}$, and the optimal aspect ratio is $\varepsilon_1/\varepsilon_2 = \sqrt{|\kappa_2/\kappa_1|}$. At points of positive Gaussian curvature, the aspect ratio “preferred” by the quadric error metric is the unique optimum; at points of negative curvature, it is one of the optima.

Since the vertex for each neighborhood is centered within its neighborhood, the aspect ratio of the approximating triangles is identical to the aspect ratio of the neighborhood. We have thus shown that, in the limit, minimization of the quadric error metric achieves an optimal triangle aspect ratio. This is our main result.

Note that in approximation theory analysis of bivariate functions, an L_2 metric uses distance measured vertically, while for optimal surface approximation, the L_2 error metric is generally defined using perpendicular distance to the surface. The two could thus disagree when applied to finite neighborhoods, but for infinitesimal neighborhoods and parabolic patches such as $\mathbf{p}(u, v)$, these distance vectors converge. This is what allows us to apply the bivariate optimality criteria of Nadler to smooth manifolds.

There are two special cases worth noting. Where the surface is locally flat, both principal curvatures are zero, and the above formula is undefined, but in this case, any aspect ratio is optimal. And where one principal curvature is zero and the other is nonzero, the aspect ratio of triangles will be infinite. (This does not happen in practice, since such a triangulation would result in an infinite number of triangles.)

4.2.1. Properties of minimized quadric

We can determine the properties of the quadrics in more detail using the derived neighborhood aspect ratio. We rewrite the dimensions of the parameter space of F as $\varepsilon_1 = \varepsilon|\kappa_2/\kappa_1|^{1/4}$ and $\varepsilon_2 = \varepsilon|\kappa_1/\kappa_2|^{1/4}$.

Substituting these values into (25) we find that the components of A for this neighborhood are

$$a_{11} = \frac{4}{3}\varepsilon^4|\kappa_1|^{3/2}|\kappa_2|^{1/2}, \quad (35)$$

$$a_{22} = \frac{4}{3}\varepsilon^4|\kappa_1|^{1/2}|\kappa_2|^{3/2}, \quad (36)$$

$$a_{33} = 4\varepsilon^2 + O(\varepsilon^4). \quad (37)$$

This confirms our intuition from Section 2: the eigenvalues of A are indeed related to the curvature of the surface, and the quadrics are elongated in the direction of minimum curvature.

Similarly, we can compute the optimal position

$$\bar{\mathbf{v}} = [0 \quad 0 \quad -\frac{1}{6}\varepsilon^2|\kappa_1\kappa_2|^{1/2}(s_1 + s_2)]^T, \quad (38)$$

where s_i is the signum function

$$s_i = \begin{cases} -1 & \text{if } \kappa_i < 0, \\ 0 & \text{if } \kappa_i = 0, \\ 1 & \text{if } \kappa_i > 0 \end{cases}$$

and its corresponding error is

$$Q(\bar{\mathbf{v}}) = \frac{8}{45}\varepsilon^6|\kappa_1\kappa_2|. \quad (39)$$

Note that, in this case, $Q(\bar{\mathbf{v}})$ is purely a function of the Gaussian curvature K . Hence, the minimal error is an intrinsic property of the surface; it depends only on the metric tensor G .

4.3. Relation to Dupin indicatrix

Nadler's optimal triangle aspect ratio is also predicted by a simple geometric construction. At a point on the original surface, take the tangent plane and offset it in the normal direction inward or outward. For a smooth surface and a small offset, the curve of intersection of the plane with the original surface will be an ellipse or a hyperbola, and the aspect ratio of these curves will be the optimal ratio given in (14). Thus, slicing a surface with a plane parallel to the tangent gives an approximate indication of the optimal triangle shape.

More formally, this intersection curve is called the *Dupin indicatrix* [15]. The indicatrix for the surface $\mathbf{p}(u, v)$ is a conic in the tangent plane of \mathbf{p}_0 which is $1/\sqrt{|\kappa_n|}$ away from \mathbf{p}_0 in any tangent direction. Thus, its principal axes are $r_i = 1/\sqrt{|\kappa_i|}$, and its aspect ratio is $\sqrt{|\kappa_2/\kappa_1|}$. The conic is an ellipse if the principal curvatures have the same sign (Fig. 4), and it is a pair of hyperbolas if they have opposite sign.

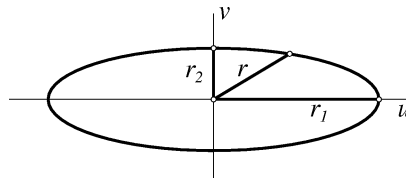


Fig. 4. The Dupin indicatrix about a point with positive Gaussian curvature.

5. Empirical results

The theory above tells us the aspect ratio of triangles in the infinitesimal limit as we minimize the quadric error metric over rectangular neighborhoods oriented parallel to the principal directions. The real quadric-based simplification algorithm is not this idealized, however. It works with sums over finite sets of triangles, not integrals; and somewhat irregular neighborhoods (Fig. 7), not perfect rectangles.

We know from experience that our algorithm is not optimal for most real simplification tasks, but we suspect that as the original models become more finely tessellated, for surfaces with slowly changing curvature, far from the boundary, the algorithm will approach this theoretical limiting behavior. For example, we find that the eigenvectors of our quadrics point approximately in the principal directions determined by surface curvature. The theoretical prediction is least accurate where the surface curvature is rapidly changing (e.g., near a crease) or near a boundary. Generally, the eigenvectors for the two smallest eigenvalues of the quadric matrix A correspond to the principal directions, and the eigenvector for the largest eigenvalue corresponds to the normal.

In practice, the neighborhoods are sometimes irregular in shape. This is due, in part, to the greedy nature of our algorithm. On each iteration it contracts the edge of least cost. Adjacent neighborhoods “compete” for edges to contract. Thus, the neighborhood of n triangles that forms around a given vertex is not necessarily the same as the set of n triangles whose quadric error at that point is smallest. Finding global minima in this manner would probably be much slower than the present algorithm.

Nevertheless, the empirical neighborhoods conform roughly to theory. Since our algorithm at each iteration contracts the edge of least error, we would predict that edges along directions of low curvature will tend to be contracted first, and neighborhoods will become elongated in the direction of low curvature. This tends to orient the neighborhoods parallel to the principal directions. On a smooth surface, neighborhoods are roughly centered around p_0 because, for such surfaces, curvature changes slowly, and the optimal vertex location for a neighborhood is near the center of that neighborhood. It is only when curvature changes within a neighborhood that the optimal vertex location moves far off-center.

A good check of our theoretical results is to test on a smooth, closed surface with fine tessellation, such as the ellipsoid in Fig. 5. A simplified version of the ellipsoid model is shown in Fig. 6, with neighborhoods shown in Fig. 7. Consistent with our prediction, neighborhoods shown in the figure are

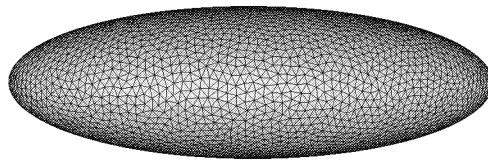


Fig. 5. Original ellipsoid model with 11,272 faces.

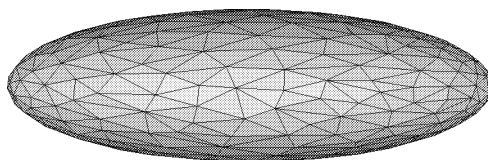


Fig. 6. Approximation of Fig. 5 using 800 faces.

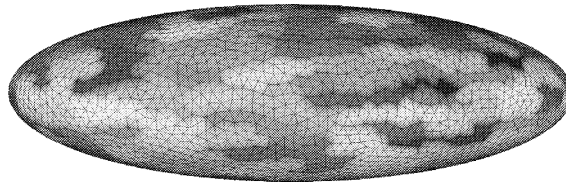


Fig. 7. Neighborhoods of original surface corresponding to vertices on the approximation.

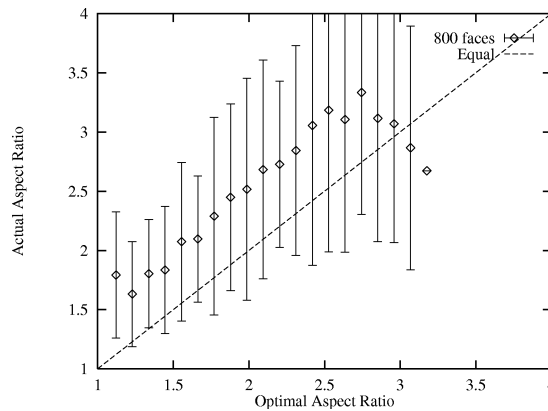


Fig. 8. Graph of theoretically optimal aspect ratios versus actual aspect ratios. An aspect ratio of 1 means equilateral, and larger values correspond to more stretched triangles. Vertical bars show mean and plus or minus one standard deviation for each bucket.

typically elongated in the direction of least curvature. We check the aspect ratios of the triangles of Fig. 6 in Fig. 8. This shows the optimal aspect ratios (computed from the principal curvatures of the underlying ellipsoid at the center of each triangle) versus the actual aspect ratios (computed by fitting a tight ellipse to the triangle, as described). Although we have not proven convergence of the greedy, quadric-based algorithm to optimal aspect ratios, we see that in practice the actual values track the theoretical values closely for the full range of aspect ratios. The slight bias toward higher aspect ratios may be an artifact of our empirical aspect ratio formulas or of pairwise contraction.

The algorithm is further demonstrated in Fig. 9, which shows a model simplified to 1% of its original size, and the appropriately stretched triangles that result. This shows that the algorithm behaves well in regions of both positive and negative curvature.

6. Conclusions

We have taken the quadric error metric from our previously published quadric-based surface simplification algorithm and analyzed its asymptotic behavior. Using methods from differential geometry and approximation theory, we have shown that the quadric error metric is directly related to surface curvature, and that its minimization yields triangulations with optimal aspect ratio in the limit.

More precisely, we have proven that when used on a differentiable manifold, in the limit as the areas of the triangles in the original model go to zero and the area of a rectangular neighborhood goes to zero,

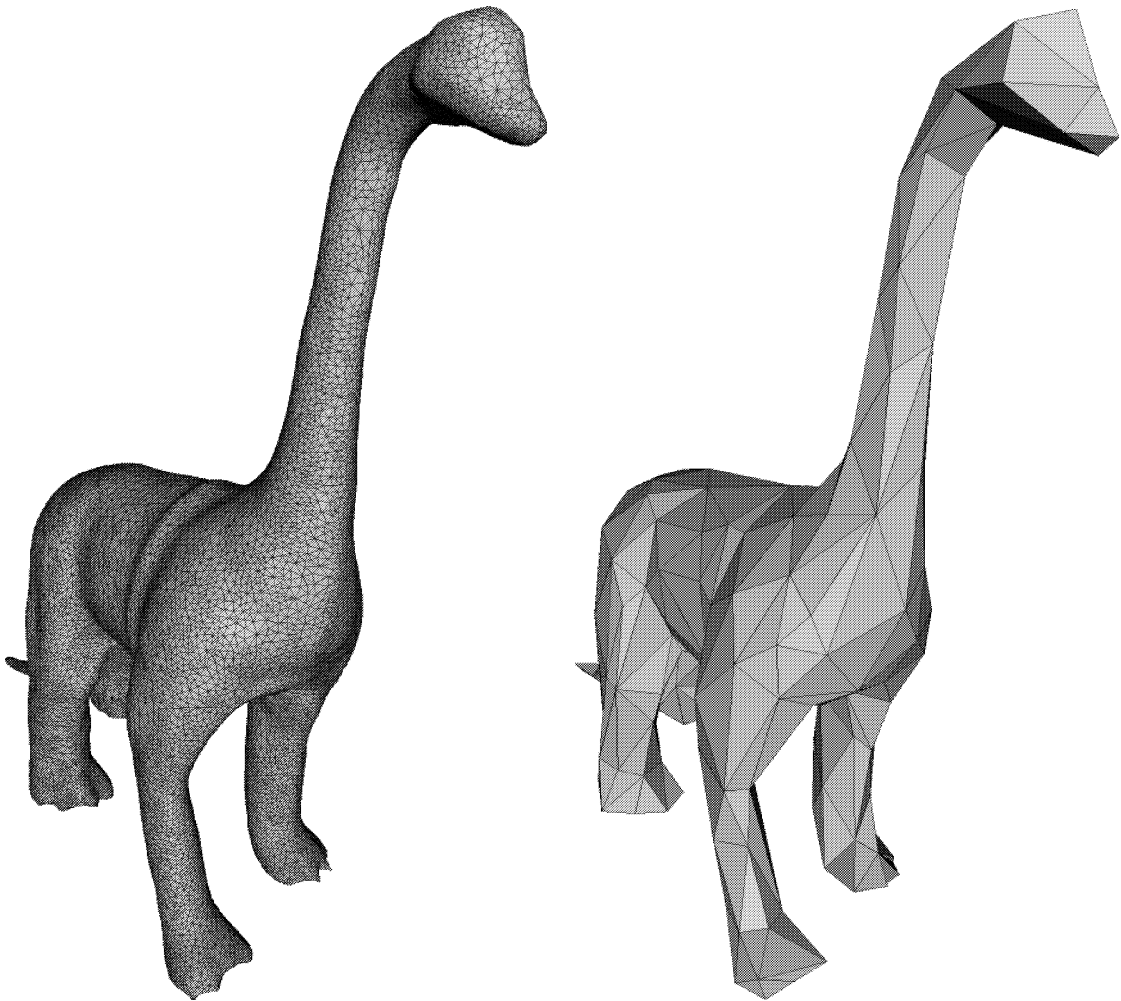


Fig. 9. A 47,904 face brontosaurus model (a) along with a 500 face approximation (b), the latter generated with quadric-based simplification. Note how the triangles stretch along the neck.

minimization of the quadric error metric generates triangles that have optimal aspect ratio in the sense of L_2 geometric error. An optimal aspect ratio is the square root of the ratio of the absolute values of principal curvatures of the surface at the point in question.

While we have not proven that our simplification algorithm yields optimal approximations for real, finite-size problems, we have shown empirically that our algorithm follows this theoretical ideal, for smooth, detailed models.

Although we have used differential geometry and approximation theory to validate our error metric, our simplification algorithm is not limited to differentiable surfaces, as those theories generally are. While curvature in differential geometry is determined by an infinitesimal neighborhood, with our error metric, as in the real world, curvature is scale-dependent.

Several areas for future work suggest themselves:

- Use the approach demonstrated in this paper to test the asymptotic optimality of other simplification algorithms.
- Rigorously prove (or disprove) that quadric-based simplification yields well-shaped neighborhoods in the limit, and that the triangles' size, in addition to their aspect ratio, is optimal.
- Modify the quadric-based algorithm to bring its empirical behavior closer to the optimal orientation, size, and aspect ratio. Perhaps greedy edge selection should be replaced by a more brute-force approach akin to simulated annealing, when quality is more important than speed.
- Verify empirically that the results are invariant to the size and orientation of triangles in the original triangulation.
- Extract curvature information from the error quadrics and use it in other ways. One could, for example, extract quadrics that locally fit the surface (up to the sign of κ_1).

C++ code for our algorithm is available at <http://www.cs.cmu.edu/~garland/quadrics/>.

Acknowledgements

We thank Konrad Polthier for conversations regarding differential geometry, the reviewers for their helpful comments, and the Schlumberger Foundation and NSF grants CCR-9357763 and CCR-9505472 for funding.

References

- [1] P.K. Agarwal, P.K. Desikan, An efficient algorithm for terrain simplifications, in: Proc. ACM–SIAM Sympos. Discrete Algorithms, 1997, pp. 139–147.
- [2] P.J. Besl, R.C. Jain, Invariant surface characteristics for 3D object recognition in range images, *Comput. Vision, Graphics, Image Process.* 33 (1986) 33–80.
- [3] F.J. Bossen, P.S. Heckbert, A pliant method for anisotropic mesh generation, in: 5th Internat. Meshing Roundtable, October 1996, pp. 63–74, <http://www.cs.cmu.edu/~ph>.
- [4] J. Cohen, A. Varshney, D. Manocha, G. Turk, H. Weber, P. Agarwal, F. Brooks, W. Wright, Simplification envelopes, in: SIGGRAPH '96 Proc., August 1996, pp. 119–128, <http://www.cs.unc.edu/~geom/envelope.html>.
- [5] H.T. Croft, K.J. Falconer, R.K. Guy, *Unsolved Problems in Geometry*, Springer, Berlin, 1991.
- [6] E.F. D'Azevedo, Optimal triangular mesh generation by coordinate transformation, *SIAM J. Sci. Statist. Comput.* 12 (4) (1991) 755–786.
- [7] E.F. D'Azevedo, R.B. Simpson, On optimal interpolation triangle incidences, *SIAM J. Sci. Statist. Comput.* 10 (6) (1989) 1063–1075.
- [8] P.J. Frey, H. Borouchaki, Unit surface mesh simplification, in: Trends in Unstructured Mesh Generation, Vol. AMD-220, ASME, July 1997, pp. 51–64.
- [9] M. Garland, Quadric-based polygonal surface simplification, Ph.D. Thesis, Technical Report CMU-CS-99-105, Computer Science Department, Carnegie Mellon University, 1999, <http://www.cs.cmu.edu/~garland/thesis/>.
- [10] M. Garland, P.S. Heckbert, Surface simplification using quadric error metrics, in: SIGGRAPH 97 Proc., August 1997, pp. 209–216, <http://www.cs.cmu.edu/~garland/quadrics/>.
- [11] M. Garland, P.S. Heckbert, Simplifying surfaces with color and texture using quadric error metrics, in: IEEE Visualization 98 Conference Proceedings, October 1998, pp. 263–269, 542, <http://www.cs.cmu.edu/~garland/quadrics/>.

- [12] A. Guéziec, Surface simplification inside a tolerance volume, Technical Report, IBM Research Report RC 20440, Yorktown Heights, NY, May 1997, <http://www.research.ibm.com/resources/>.
- [13] D. Hilbert, S. Cohn-Vossen, *Geometry and the Imagination*, Chelsea, New York, 1952.
- [14] H. Hoppe, Progressive meshes, in: SIGGRAPH '96 Proc., August 1996, pp. 99–108, <http://research.microsoft.com/~hoppe/>.
- [15] E. Kreyszig, *Introduction to Differential Geometry and Riemannian Geometry*, University of Toronto Press, Toronto, 1968.
- [16] E. Nadler, Piecewise linear best L_2 approximation on triangulations, in: C.K. Chui et al. (Eds.), *Approximation Theory V*, Academic Press, Boston, 1986, pp. 499–502.
- [17] J. Peraire, J. Peiró, Adaptive remeshing for three-dimensional compressible flow computations, *J. Comput. Phys.* 103 (1992) 269–285.
- [18] S. Rippa, Long and thin triangles can be good for linear interpolation, *SIAM J. Numer. Anal.* 29 (1) (1992) 257–270.
- [19] R. Ronfard, J. Rossignac, Full-range approximation of triangulated polyhedra, *Comput. Graphics Forum* 15 (3) (1996).
- [20] R.B. Simpson, Anisotropic mesh transformations and optimal error control, *Appl. Numer. Math.* 14 (1–3) (1994) 183–198.