Maximum norm optimization of quasi-isometric mappings

Article *in* Numerical Linear Algebra with Applications · September 2002

DOI: 10.1002/nla.301

1 author:

Vladimir A. Garanzha
Russian Academy of Sciences
34 PUBLICATIONS **108** CITATIONS

SEE PROFILE

NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2002; **9**:493–510 (DOI: 10.1002/nla.301)

Maximum norm optimization of quasi-isometric mappings

V. A. Garanzha*,†

Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 117967, Russia

SUMMARY

A reliable method for maximum norm optimization of spatial mappings is suggested. It is applied to the problem of optimal flattening of surfaces and to precisely controlled surface morphing. Robustness and grid independence of the method are demonstrated on real-life tests. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: surface grid generation; quasi-isometric mappings; optimization of mappings; surface flattening; graph drawing; morphing; grid sensitivity

1. INTRODUCTION

Variational methods for grid generation and for construction of spatial mappings with prescribed properties [1–7] have become important tools in many real-life applications. Despite considerable progress in this area, lots of unsolved problems still exist.

Since the introduction of de Boor equidistribution principle [8], it became clear that the ability to construct spatial mappings which provide best maximum norm solutions to various interpolation, approximation and adaptive simulation problems should be one of the basic tools in the numerical analysis.

However, it was not clear how equidistribution principle can be generalized to many spatial dimensions.

The idea that optimal spatial mappings should be sought in the class of quasi-isometric mappings was suggested by Godunov [9]. Mapping with minimal deviation from isometry in fact is the natural definition of the equidistribution principle in multidimensional case.

One of the aims of this paper is to show how the control over properties of local mappings in the finite element method based on the metric tensors allows one to minimize the deviation of size, shape and orientation of grid cells from the prescribed values in the maximum norm.

The potential applications of the optimality principle include grid adaptation to flow features, grid orthogonality near boundary, surface grid generation, quasi-isometric projections

* Correspondence to: V. A. Garanzha, Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 117967, Russia.
† E-mail: garan@ccas.ru

Contract/grant sponsor: NWO-NL-RFBR; contract/grant number: 047-008-007

Received 4 October 2001
Revised 10 June 2002

of complicated surfaces, shape optimization and various structural optimization problems with most advantages in the case of large deformations.

There is also a separate class of 2.5-D applications where this principle plays a crucial role: construction of surfaces from wireframes, morphing, optimization of surface triangulations, repair of CAD data such as elimination of wrinkles and many other.

Moreover, applications from other fields such as straight-line planar graph drawing with good angular resolution can be tackled by the presented methods as well, along with 3-D generalizations of this graph-theoretical problem.

This optimality principle is implemented via the minimization of the functional which is elliptic and in the discrete case has an infinite barrier on the boundary of the feasible set consisting of grids with quasi-isometric local mappings [10]. This feasible set is contracted using the continuation technique which forces the global lower and upper spectral bounds of metric tensors of local mappings in the finite element grid to increase and to decrease, respectively.

The particular topic of interest in this paper is the construction of quasi-isometric flattenings of complicated surfaces. This problem is very important for many applications in mechanics, computational geometry and in surface grid generation [11, 7]. Moreover, it can be used as a building block in the construction of globally optimal adaptive grids and as a tool for sharply decreasing the sensitivity of grids and solutions on these grids with respect to variations of control parameters, which is extremely important in applications coupling fluid flow simulations with optimization/optimal control.

The important issue related to construction of quasi-isometric flattening of surfaces is that initial guesses are generally infeasible, hence the method should handle the following problems: (1) creation of non-degenerate local mappings or in other words grid untangling; (2) stabilization and untangling of free boundaries in the small; usual pattern when creating the non-degenerate mapping from infeasible initial guess is the creation of knots and spiral-like structures on the free boundaries such that the resulting mapping is non-degenerate but multivalued; such behaviour must be prohibited; (3) global non-overlapping of free boundaries. The simple solution of the above problems presented in this work was found to be quite reliable and due to existence of some rigorous foundations has been successfully tested on hard industrial problems.

Numerical results related to surface grid generation, construction of surface flattenings and morphing are presented.

2. OPTIMALITY PRINCIPLE FOR SPATIAL MAPPINGS

We will use here a simple and intuitive concept of optimality. The spatial mapping is called optimal if it is smooth enough and is as close as possible to a uniform mapping. The uniformity means that the distance between any two sufficiently close points is the same as the distance between their images. In fact this is the definition of isometric mapping. The mapping resembles the uniform one when the distance function is not much distorted under this mapping, meaning that the ratio of the distance between any two close enough points and the distance between their images has uniform lower and upper bounds. This is essentially the definition of the quasi-isometric mapping.

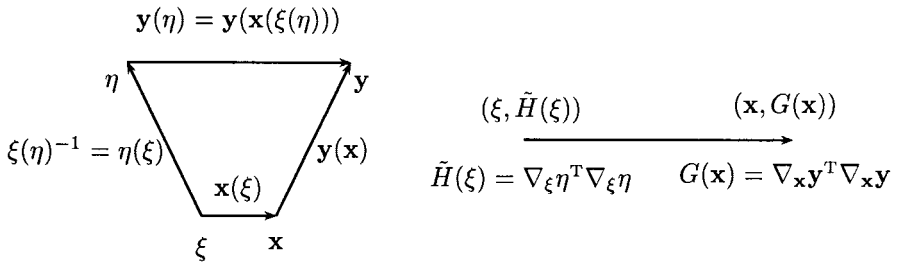

Figure 1. Composition of mappings. $\mathbf{y}(\mathbf{x})$ and $\eta(\xi)$ are prescribed mappings, $^{-1}$ means inverse mapping.

It is convenient to use the following invariant definition of quasi-uniformity. Let the mapping of interest be $\mathbf{y}(\eta): \mathbb{R}^n \to \mathbb{R}^n$. Denote by $\nabla_\eta \mathbf{y}$ the gradient of the mapping (the Jacobi matrix) with the entries $\partial y_i / \partial \eta_j$. The mapping $\mathbf{y}(\eta)$ is quasi-uniform when

$$\bar{v}^{1/n} \frac{1}{\Gamma} \leqslant \sigma_i(\nabla_\eta \mathbf{y}) \leqslant \Gamma \bar{v}^{1/n} \tag{1}$$

where σ_i is any singular value of matrix $\nabla_\eta \mathbf{y}$ and $\Gamma \geqslant 1$ is the constant. The value $\bar{v}$ is just the constant volume scaling factor. If the co-ordinates are properly prescaled, then it is possible to set $\bar{v} = 1$. Now we can define an optimal mapping as the one having minimal possible value Γ provided that this mapping also satisfies certain additional constraints, such as boundary conditions, smoothness, etc. Obviously each quasi-uniform mapping is also quasi-isometric.

In practice quasi-uniformity is just not enough. If it is necessary to construct computational grids with full control over size, shape and orientation of elements it is necessary to have means of such control.

The mapping properties can be controlled via composition of mappings which is illustrated on Figure 1.

The mapping $\mathbf{y}(\eta)$ is represented as a composition of mappings $\xi(\eta), \mathbf{x}(\xi)$ and $\mathbf{y}(\mathbf{x})$. The mappings $\mathbf{y}(\mathbf{x})$ and $\eta(\xi)$ are specified while the function $\mathbf{x}(\xi)$ is the new unknown solution. We assume again that $\mathbf{y}(\mathbf{x})$ and $\eta(\xi)$ are quasi-isometric mappings, but possibly with large constants Γ. Using the notations

$$H = \nabla_\xi \eta, \quad S = \nabla_\xi \mathbf{x}, \quad W = \nabla_{\mathbf{x}} \mathbf{y}, \quad T = \nabla_\eta \mathbf{y}, \quad J = \det T$$

we get

$$T = WSH^{-1}, \quad J = \frac{\det W \det S}{\det H}, \quad \mathrm{d}\eta = \det H \, \mathrm{d}\xi$$

Inequalities (1) can be rewritten as follows:

$$\bar{v}^{2/n} \frac{1}{\Gamma^2} I \leqslant T^{\mathrm{T}} T \leqslant \bar{v}^{2/n} \Gamma^2 I$$

or

$$\bar{v}^{2/n} \frac{1}{\Gamma^2} \tilde{H} \leqslant S^{\mathrm{T}} G S \leqslant \bar{v}^{2/n} \Gamma^2 \tilde{H} \tag{2}$$

where

$$\tilde{H}(\xi)=\nabla_\xi\eta^{\mathrm{T}}\nabla_\xi\eta,\quad G(\mathbf{x})=\nabla_{\mathbf{x}}\mathbf{y}^{\mathrm{T}}\nabla_{\mathbf{x}}\mathbf{y}$$

are symmetric positive metric tensors which define the distance in ξ and $\mathbf{x}$ co-ordinates respectively. Formulation (2) is quite general since now it is based just on metric tensors related to prescribed mappings. However the metric tensors can be specified directly as it happens in many methods for adaptive grid generation. Note different nature of controls imposed by $\tilde{H}(\xi)$ and $G(\mathbf{x})$, since latter depends on the unknown solution $\mathbf{x}(\xi)$. Suppose that $\mathbf{x}(\xi)$ is computed via some iterative method. Then on each iteration $G(\mathbf{x})$ should be recomputed. This process can be considered as adaptation in 'Eulerian' co-ordinates. On the other hand, $\tilde{H}(\xi)$ is fixed once and for all. It looks like *a priori* control, control in 'Lagrangian' co-ordinates. Using mechanical analogy, during the iterative solution $\tilde{H}(\xi)$ is 'frozen' into each material point and 'accompanies' it.

The idea of using composition of mappings in order to control the properties of the computational grid was suggested in Reference [1].

Now we are in a position to formulate the final optimality principle: the spatial mapping

$$\mathbf{x}=\mathbf{x}(\xi_1,\ldots,\xi_m),\ \mathbf{x}=(x_1,\ldots,x_n)^{\mathrm{T}},\quad n\geqslant m \tag{3}$$

is optimal if it preserves orientation and is quasi-isometric with respect to prescribed control metrics, i.e.

$$\frac{1}{\Gamma^2}\,\bar{v}^{2/n}H^{\mathrm{T}}H\leqslant S_e^{\mathrm{T}}GS_e\leqslant\Gamma^2\bar{v}^{2/n}H^{\mathrm{T}}H \tag{4}$$

with minimal possible Γ. Here $G=G^{\mathrm{T}}>0$, $G=G(\mathbf{x})\in\mathbb{R}^{n\times n}$ is the metrics in the 'physical' co-ordinates x_i, $H=H(\xi)\in\mathbb{R}^{n\times n}$, $\det H>0$ and $H^{\mathrm{T}}H$ is the 'accompanying' metrics defined in the logical co-ordinates $\xi_1,\ldots,\xi_m$, while $\bar{v}$ is the constant volumetric factor. The square matrix S_e is defined as follows:

$$S_e=(S,Q),\quad S=(\mathbf{g}_1,\ldots,\mathbf{g}_m),\quad Q=(\mathbf{q}_1,\ldots,\mathbf{q}_{n-m}),\quad \mathbf{g}_i=\frac{\partial\mathbf{x}}{\partial\xi_i} \tag{5}$$

Here $S=\nabla_\xi\mathbf{x}$ is the Jacobi matrix of mapping (3). If $n=m$ then $S_e=S$, otherwise some given $n-m$ vectors $\mathbf{q}_i=\mathbf{q}_i(\xi_1,\ldots,\xi_m)$ are used to complement the Jacobi matrix of (3) to square matrix. It is assumed that $\mathrm{rank}(\mathbf{q}_1,\ldots,\mathbf{q}_{n-m})=n-m$. Definition (5) means that when $m<n$ the set of auxiliary vector fields is added to the local tangential basis on the m-dimensional hypersurface in order to form a complete basis in the n-dimensional space.

This optimality principle is more complicated compared to (2) since it allows to take into account the orientation of the tangent basis vectors of the mapping.

Using the notation

$$A=H^{-\mathrm{T}}S_e^{\mathrm{T}},\quad A=(\mathbf{a}_1,\ldots,\mathbf{a}_n),\quad a_{ij}=(\mathbf{a}_i)_j\ \text{ or }\ A=A(\xi,\nabla_\xi\mathbf{x}) \tag{6}$$

one can rewrite (4) as follows:

$$\det A>0,\ \frac{1}{\Gamma^2}\,\bar{v}^{2/n}I<AGA^{\mathrm{T}}<\Gamma^2\bar{v}^{2/n}I,\quad \Gamma\geqslant 1 \tag{7}$$

Mapping (3) which satisfies (7) with minimal possible Γ can be considered as the maximum norm solution to the problem of finding best approximation to an isometric mapping where

$$AGA^{\mathrm{T}} = \bar{v}^{2/n} I \tag{8}$$

Formulation (7) is mathematically elegant but it is extremely difficult to use in practical optimization algorithms since the condition number Γ^2 as a function of the entries of the matrix A has some unfavourable algebraic properties (cf. Reference [12]).

So this inequality is in turn approximated by the following optimality principle [10] which is finally adopted in this work:

$$\det A > t\phi_\theta(A), \quad 0 < t \leqslant 1 \tag{9}$$

where

$$\phi_\theta(A) = \theta \frac{((1/n)\,\mathrm{tr}(GA^{\mathrm{T}}A))^{n/2}}{(\det G)^{1/2}} + (1-\theta)\left(\frac{\bar{v}}{(\det G)^{1/2}} + \frac{(\det G)^{1/2}}{\bar{v}}(\det A)^2\right) \tag{10}$$

and $0 < \theta < 1$ is some constant. The optimization in this case just reduces to the maximization of parameter t in inequality (9). When $\theta = 1$, $m = n$ inequality (9) is exactly the definition of a class of mappings with bounded distortion introduced by Reshetnyak [13] immediately providing rigorous mathematical grounds for the problem formulation, in terms of regularity of mappings, constraints on domains and boundary conditions, see for details References [13, 14]. Choosing $0 < \theta < 1$ means that the subclass of quasi-isometric mappings is selected from the mappings with bounded distortion. In Reference [10] it was shown that (7) follows from (9) with

$$\Gamma \leqslant (c_1 + \sqrt{c_1^2 - 1})^{1/n}(c_2 + \sqrt{c_2^2 - 1})^{(n-1)/n}, \quad c_1 = \frac{1-\theta t}{(1-\theta)t}, \quad c_2 = \frac{1-(1-\theta)t}{\theta t} \tag{11}$$

Estimates (11) become tighter when t increases hence (9) also can be considered as the approximate solution to (8) in the maximum norm.

For practical applications it is useful to employ also slightly different optimality principle, namely

$$\det A > tq(\xi)\psi(H)\phi_\theta(A) \tag{12}$$

where $0 \leqslant q(\xi_1, \ldots, \xi_m) \leqslant 1$ and $\psi(H) > 0$ are piecewise-smooth functions. This modification can be used to allow larger distortions in some subdomains of the computational domains and enforce more strict control on other parts of the domains. In the most practically important case of $q(\xi)$ taking only two values—0 and 1, the mapping is optimized only in a part of the domain and the non-degeneracy in the rest of the domain serves as the constraint. Later it will be demonstrated that this ability is crucial for the problems with free boundaries.

3. VARIATIONAL PRINCIPLE

The variational principle implementing optimality principle (9) is written as follows: given the domain $\mathscr{D}$ in co-ordinates $\xi_1, \ldots, \xi_m$, find a mapping $\mathbf{x}(\xi)$ as the extremal of the following

functional:

$$\int_{\mathscr{D}} f(A)\,\mathrm{d}\xi \tag{13}$$

where $A=A(\xi,\nabla_\xi \mathbf{x})$ is defined in (6) and

$$f(A)=\begin{cases}(1-t)\det H\,\dfrac{\phi_\theta(A)}{\det A-t\phi_\theta(A)}, & \det A-t\phi_\theta(A)>0\\ +\infty & \text{otherwise}\end{cases} \tag{14}$$

For the modified optimality principle (12), a slightly different functional is suggested, namely

$$f(A)=f_2(A)=\begin{cases}(1-q(\xi)t)\,\dfrac{\psi(H)\phi_\theta(A)}{\det A-tq(\xi)\psi(H)\phi_\theta(A)}, & \det A-tq(\xi)\psi(H)\phi_\theta(A)>0\\ +\infty & \text{otherwise}\end{cases} \tag{15}$$

We will use the same notation for $f(A)$ and for $f(\xi,\nabla_\xi \mathbf{x})=f(A(\xi,\nabla_\xi \mathbf{x}))$.

When the volume of the image of $\mathscr{D}$ under mapping (3) is known, the volumetric factor is computed as follows

$$\bar{v}=\int_{\mathscr{D}}(\det G)^{1/2}\det S\,\mathrm{d}\xi \Big/ \int_{\mathscr{D}}\det H\,\mathrm{d}\xi$$

otherwise it is prescribed *a priori.*

Let us summarize the properties of the functional (13), (14)

- **3.1.** The stationary point of $f(A)$ is given by (8);
- **3.2.** $f(A)$ tends to $+\infty$ when feasible A tends to the boundary of the feasible set;
- **3.3.** f is polyconvex function of $\nabla_\xi \mathbf{x}$ provided that $A(\xi,\nabla_\xi \mathbf{x})$ satisfies the feasible set inequality (9);
- **3.4.** f satisfies the strong Legendre–Hadamard condition (is strongly rank one convex) inside the feasible set meaning that the functional possesses the strong ellipticity property.

Property **3.1** was proven in Reference [10].

The barrier property **3.2** follows from the fact that the numerator in $f(A)$ is bounded from below by a positive constant. So when denominator tends to $+0$, f tends to $+\infty$.

The polyconvexity property **3.3** is a crucial notion in non-linear calculus of variations [15]. The function $f(A)$ is called polyconvex if it can be expressed as a convex function of minors of A. Let us define the function $\tilde{f}(A,J)$ as follows:

$$\tilde{f}(A,J)=\begin{cases}(1-t)\det H\,\dfrac{\tilde{\phi}_\theta(A,J)}{J-t\tilde{\phi}_\theta(A,J)}, & J-t\tilde{\phi}_\theta(A)>0\\ +\infty & \text{otherwise}\end{cases}$$

where

$$\tilde{\phi}_\theta(A,J)=\theta\frac{((1/n)\,\mathrm{tr}(GA^{\mathrm{T}}A))^{3/2}}{(\det G)^{1/2}}+(1-\theta)\left(\frac{\bar{v}}{(\det G)^{1/2}}+\frac{(\det G)^{1/2}}{\bar{v}}J^2\right)$$

When $J=\det A$, f and $\tilde{f}$ coincide. So in order to show the polyconvexity of $f(A)$ it is enough to demonstrate the convexity of $\tilde{f}(A,J)$.

Since A is a linear function of $\nabla_\xi \mathbf{x}, \det A$ can be expressed as a linear combination of minors of $\nabla_\xi \mathbf{x}$ and convexity of $\tilde{f}$ leads to polyconvexity of $f(\cdot, \nabla_\xi \mathbf{x})$.

The first step is to prove the following auxiliary statement: the *function* $\tilde{\phi}_\theta(A,J)/J$ *is convex on* $\mathbb{R}^{n\times n}\times\mathbb{R}_+$.

The function $\tilde{\phi}_\theta(A,J)/J$ can be written as the sum of two terms

$$\frac{\tilde{\phi}_\theta(A,J)}{J}=\theta\,\frac{((1/n)\operatorname{tr}(GA^{\mathrm{T}}A))^{n/2}}{J(\det G)^{1/2}}+(1-\theta)\left(\frac{\bar{v}}{J(\det G)^{1/2}}+\frac{(\det G)^{1/2}}{\bar{v}}J\right)$$

The second term is obviously convex provided that $J>0$. It remains only to prove the convexity of the first term. Keeping in mind that the function $x^{n/2}$ is convex, to this end it is enough to show the convexity of the function $g(A,J)=((1/n)\operatorname{tr}(A^{\mathrm{T}}A))^{n/2}/J$. Matrix G does not change the convexity since it can be eliminated by an affine transformation of A. Let us consider the second derivatives of the function $g(A,J)$:

$$\frac{\partial^2 g}{\partial a_{ij}\partial a_{k\ell}}=\delta_{ik}\delta_{j\ell}\,\frac{1}{nJ^{2/n}}$$

$$\frac{\partial^2 g}{\partial J\partial a_{ij}}=-\frac{4}{n^2}\,a_{ij}J^{-((2+n)/n)}$$

$$\frac{\partial^2 g}{\partial J^2}=\frac{2(2+n)}{n^3}\sum_{i,j=1}^{n}a_{ij}^2J^{-((2+2n)/n)}$$

If we establish that the Hessian matrix of $g(A,J)$ is non-negative definite on the convex set $\mathbb{R}^{n\times n}\times\mathbb{R}_+$, it will guarantee the convexity of $g(A,J)$.

Let us write the Hessian matrix of $g(A,J)$ using a block partitioning corresponding to A and J

$$\mathscr{H}=\left[\begin{array}{c|c}\mathscr{H}_{11}&\mathscr{H}_{12}\\\hline\mathscr{H}_{21}&\mathscr{H}_{22}\end{array}\right]=\left[\begin{array}{c|c}(2/nJ^{2/n})I&-(4/n^2)J^{-((2+n)/n)}\mathbf{A}\\\hline-(4/n^2)J^{-((2+n)/n)}\mathbf{A}^{\mathrm{T}}&(2(2+n)/n^3)\operatorname{tr}A^{\mathrm{T}}A(J^{-(2+2n)/n})\end{array}\right]$$

where $\mathbf{A}$ is the vector of all entries of the matrix A. Obviously $\mathbf{A}^{\mathrm{T}}\mathbf{A}=\operatorname{tr}(A^{\mathrm{T}}A)$. The diagonal blocks of matrix $\mathscr{H}$ are non-negative definite and in order to show that $\mathscr{H}$ is non-negative definite it is enough to demonstrate that the Schur complement $S_{22}=\mathscr{H}_{22}-\mathscr{H}_{21}\mathscr{H}_{11}^{-1}\mathscr{H}_{12}$ is non-negative definite. Since

$$S_{22}=\frac{2(n-2)}{n^3}\operatorname{tr}A^{\mathrm{T}}A(J^{-(2+2n)/n})\geqslant 0$$

the proof is complete.

The next step is the following: *let* $\mathbf{m}\in\mathbb{R}^n$ *and* $\varphi(\mathbf{m})$ *be a convex function in* $\mathbb{R}^n$, *then the function* $\varphi(\mathbf{m})/(1-\varphi(\mathbf{m}))$ *is a convex function on the subset of* $\mathbb{R}^n$ *defined by inequality*

$1-\varphi(\mathbf{m})>0$. This statement is well known [16]. Choosing φ in the above statement as

$$t\,\frac{\tilde{\phi}_\theta(A,J)}{J}$$

and using the fact that the set

$$J-t\tilde{\phi}_\theta(A,J)>0$$

is convex we see that $\tilde{f}(A,J)$ is the convex function everywhere which in turn means that f is polyconvex function of $\nabla_\xi \mathbf{x}$ everywhere.

The Legendre–Hadamard condition **3.4** for f can be written as follows:

$$\sum_{i,j=1}^{m}\sum_{k,l=1}^{n}\frac{\partial^2 f}{\partial s_{ik}\partial s_{j\ell}}\,p_i p_j z_l z_k>0 \tag{16}$$

where $\mathbf{z}\in\mathbb{R}^n$, $\mathbf{p}\in\mathbb{R}^m$ are arbitrary non-zero vectors and s_{ij} are entries of $S=\nabla_\xi \mathbf{x}$.

Inequality (16) in the case $m=n$ was proved in Reference [17].

The strong rank one convexity means that

$$f(\lambda Q_1+(1-\lambda)Q_2)<\lambda f(Q_1)+(1-\lambda)f(Q_2) \tag{17}$$

for every $0<\lambda<1$, $Q_1,Q_2\in\mathbb{R}^{n\times m}$ with $\operatorname{rank}(Q_1-Q_2)=1$. When $f(A)$ is smooth enough properties (16) and (17) are equivalent. Since any polyconvex function is rank one convex [15], we see that (16) is valid in the case $m<n$ as well.

The properties of the function $f_2(A)$ are similar.

A similar set of properties was used to prove the existence of minimizers for the variational problems in hyperelasticity [15]. However, the formulations of the existence theorems by Ball [15] cannot be directly applied to problem (13).

4. GEOMETRIC INTERPRETATION OF THE OPTIMALITY PRINCIPLE

Let us consider mapping (3) in the case $n=m=2$ in a small neighbourhood of the point $\mathbf{x}_0$ in physical co-ordinates. The size of the neighbourhood is small enough to consider the mapping to be affine and metrics $G(\mathbf{x}_0)$ to be constant. The local properties of the metrics G can be described by the ellipse e_1

$$(\mathbf{x}-\mathbf{x}_0)^{\mathrm{T}}G(\mathbf{x}-\mathbf{x}_0)=\varepsilon$$

which is shown in Figure 2 (left). The orientation of this ellipse is given by the eigenvectors of $G(\mathbf{x}_0)$, while its aspect ratio is equal to the square root of the condition number of $G(\mathbf{x}_0)$.

Let us introduce an affine mapping $\mathbf{y}=\mathbf{y}(\mathbf{x})$, such that the image of the ellipse e_1 in co-ordinates y_1, y_2 is a circle with a specified radius. In transformed co-ordinates we can consider a cell of any given shape, say a triangle or a quadrilateral. The exact shape of this cell is defined by the matrix $H(\xi_1,\xi_2)$ in such a way that the Jacobi matrix of the mapping of some ideal cell in co-ordinates ξ_1,ξ_2, say of the unit square or unilateral triangle onto the cell in co-ordinates y_1, y_2 coincides with H. If this cell can be identified only up to an arbitrary translation and rotation, and $G(\mathbf{x}_0)$ is ill-conditioned or 'anisotropic' then the

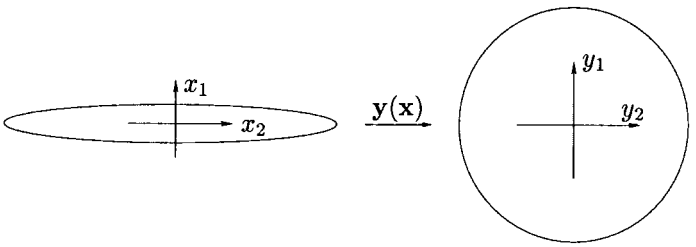

Figure 2.

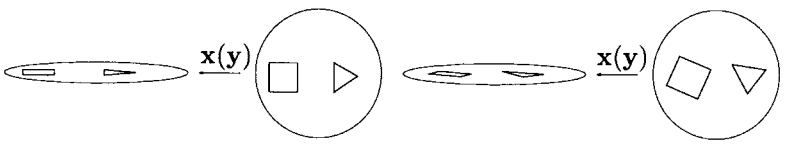

Figure 3.

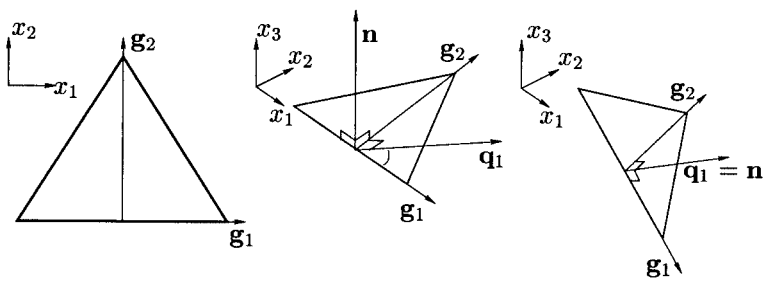

Figure 4.

inverse mapping $\mathbf{x}=\mathbf{x}(\mathbf{y})$ can produce drastically different shapes from the same cell in y_1, y_2 co-ordinates depending on its orientation, which is illustrated in Figure 3.

The cells shown in Figure 3 on the left are preferable to those on the right according to numerous well-established criteria, first of all in terms of orthogonality and alignment.

This simple example clearly demonstrates that the control of the cell shape alone is not enough to obtain high-quality adaptive grids and it should be complemented by the cell alignment control.

It also provides the motivation to use the suggested optimality principle in the case $m<n$ with the basic goal being the simultaneous shape and orientation control. In what follows we will concentrate on the shape and orientation control via a shape control matrix H and auxiliary vectors $\mathbf{q}_i$, and assume that $G=I$.

Figure 4 illustrates three important particular cases of the optimality principle as applied to triangles. For simplicity here we will assume that the mapping (3) is affine, i.e. it is just a linear mapping of one triangle onto another.

The covariant basis vectors $\mathbf{g}_i$ of this affine mapping can be computed as follows. Let the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be the triangle vertices, while $\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3 \in \mathbb{R}^2$ are the vertices of an ideal plane triangle. Using barycentric co-ordinates inside each triangle and assuming that these co-ordinates are the same for both triangles we obtain the $(n+1) \times 3$ auxiliary matrix

$$\tilde{S} = \begin{pmatrix} 1 & 1 & 1 \\ \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ \mathbf{l}_1 & \mathbf{l}_2 & \mathbf{l}_3 \end{pmatrix}^{-1} \tag{18}$$

such that

$$\mathbf{g}_i^{\mathrm{T}} = (\tilde{S}_{2,i+1} \ \tilde{S}_{3,i+1} \ \cdots \ \tilde{S}_{n+1,i+1})$$

Hence the n-dimensional vectors $\mathbf{g}_1, \mathbf{g}_2$ represent a local tangential (or covariant) basis for the triangle in the n-dimensional space.

In Figure 4 (left) $m = n = 2$ and optimality in the sense of (9) is obviously insensitive to triangle orientation. In the centre is shown the case $m = 2$, $n = 3$ with additional constraint $x_3 = 0$. In this case the covariant basis vectors of the mapping (3) $\mathbf{g}_1, \mathbf{g}_2$ are the usual 2-D covariant basis vectors of the mapping of an ideal triangle in logical co-ordinates onto the true triangle with zero x_3 components. The matrix H is specified as follows:

$$H = (\mathbf{h}_1, \mathbf{h}_2, \mathbf{q}_1)$$

where $\mathbf{h}_1$, $\mathbf{h}_2$ are the 2-D covariant basis vectors with zero x_3 components of the mapping of an ideal triangle in logical co-ordinates onto the target triangle and $\mathbf{q}_1$ is a given 3-D vector with prescribed length such that $\det H > 0$, $\mathbf{q}_1^{\mathrm{T}} \mathbf{h}_1 > 0$, $\mathbf{q}_1^{\mathrm{T}} \mathbf{h}_2 = 0$. Since vector $\mathbf{q}_1$ is not orthogonal to the plane $x_3 = 0$ then for the optimal triangle

$$S_e = H \tag{19}$$

so not only the target shape is recovered but also the triangle is aligned in such a way that $\mathbf{q}_1^{\mathrm{T}} \mathbf{g}_1 > 0$, $\mathbf{q}_1^{\mathrm{T}} \mathbf{g}_2 = 0$.

In Figure 4 (right) the case where $m = 2$, $n = 3$ is illustrated, where $\mathbf{g}_1$, $\mathbf{g}_2$ are now the tangent basis vectors on the triangle, $\mathbf{h}_1$, $\mathbf{h}_2$ are the target tangent basis vectors and $\mathbf{q}_1$ is the target normal to the plane of the triangle. In this case (19) does not necessarily hold for the optimal triangle since it is defined up to rotation around the $\mathbf{q}_1$ axis. However this optimal triangle should be orthogonal to the target normal vector $\mathbf{q}_1$.

5. DISCRETIZATION AND PROPERTIES OF THE DISCRETE FUNCTIONAL

We consider only the case when the domain $\mathscr{D}$ is covered by the simplicial grid, in particular in the 2-D case the triangular grid in $\mathscr{D}$ is already present. Then the mapping (3) in each simplex is replaced by the affine mapping and the vectors $\mathbf{q}_i$, if present, are constant on each triangle. Let us denote by $\mathbf{R}$ the vector of all co-ordinates of all grid vertices. In this case the discrete counterpart of the functional (13) is simply written as the following sum:

$$\mathscr{I}^h(\mathbf{R}) = \sum_{k=1}^{N_e} f(A_k) \tag{20}$$

where A_k is the matrix A computed on the kth simplex and is a linear function of these simplex co-ordinates, while N_e is the total number of elements in the grid. $f(A)$ is given by (14).

The following auxiliary functional is used as well:

$$\mathscr{I}_\varepsilon^h(\mathbf{R}) = \sum_k f_\varepsilon(A_k) \tag{21}$$

where

$$f_\varepsilon(A) = \det(H)\frac{\phi_\theta(A) + C\varepsilon\,\mathrm{tr}(AA^{\mathrm{T}})}{\chi_\varepsilon(\det A)}, \quad \chi_\varepsilon(q) = \frac{q}{2} + \frac{1}{2}\sqrt{\varepsilon^2 + q^2}$$

where C is a positive constant when $n \geqslant 3$ and $C = 0$ when $n = 2$.

The properties of the discrete functionals can be summarized as follows:

(1) $\mathscr{I}^h(\mathbf{R})$ tends to $+\infty$ when the feasible grid $\mathbf{R}$ tends to the boundary of the feasible set or when $|\mathbf{R}| \to \infty$.
(2) In the case $t = 0$ when no boundary conditions are imposed, the norm of the gradient $\partial\mathscr{I}^h(\mathbf{R})/\partial\mathbf{R}$ tends to ∞ when the feasible grid tends to the boundary of the feasible set.
(3) When the grid $\mathbf{R}$ consists of one simplex without any boundary conditions then $\mathscr{I}^h(\mathbf{R})$ has a stationary point which is unique up to the rigid body degrees of freedom.
(4) Suppose that in the grid $\mathbf{R}$ all nodes but one are fixed, then $\mathscr{I}^h$ is the convex function of the co-ordinates of this node.
(5) In the case $t = 0$ when no boundary conditions are imposed and the feasible set is not empty, then for sufficiently small ε all stationary points of the regularized functional (21) are feasible.

Property (1) is obvious, while the proof of statements (2) and (5) is more complicated. They are based on the analysis of scalar values

$$\rho(A) = \sum_{i=1}^{n} \mathbf{a}_i^{\mathrm{T}} \frac{\partial f(A)}{\partial \mathbf{a}_i}, \qquad \rho_\varepsilon(A) = \sum_{i=1}^{n} \mathbf{a}_i^{\mathrm{T}} \frac{\partial f_\varepsilon(A)}{\partial \mathbf{a}_i}$$

and the fact that when no boundary conditions are imposed

$$\mathbf{R}^{\mathrm{T}} \frac{\partial \mathscr{I}^h(\mathbf{R})}{\partial \mathbf{R}} = \sum_{k=1}^{N_e} \rho(A_k), \qquad \mathbf{R}^{\mathrm{T}} \frac{\partial \mathscr{I}_\varepsilon^h(\mathbf{R})}{\partial \mathbf{R}} = \sum_{k=1}^{N_e} \rho_\varepsilon(A_k) \tag{22}$$

When $t = 0$ and $\det A \to +0$ then

$$\rho(A) = -(1-\theta)\bar{v}\,\frac{\det H}{\det A(\det G)^{1/2}} + \text{bounded terms}$$

for $f(A)$ defined by (14) which in turn means that the inner product (22) cannot be bounded.

When $t = 0$ then inequality $\rho_\varepsilon(A) > 0$ is valid for any A. Moreover, when $\det A \leqslant 0$ then

$$\lim_{\varepsilon \to +0} \rho_\varepsilon(A) = +\infty$$

so the functional (21) cannot have limiting stationary points outside the feasible set.

In fact (2) is not yet proved for the more general case when some of the boundary nodes are fixed. Moreover, statement (5) is not necessarily true for problems where some grid nodes are fixed.

6. PRACTICAL NON-LINEAR SOLVER FOR THE VARIATIONAL PROBLEM

The non-linear solver includes two main phases. The first one is the continuation technique for functional (21) in order to construct feasible solutions [18], when it is necessary.

The second phase is the contraction of the feasible set in order to create an optimal grid also based on the continuation technique.

The minimization method is the implicit gradient-type method based on the reduced Hessian matrix of the functional. The positive definite part of the Hessian is created analytically using the ellipticity of the target functional. The sparse linear systems are solved using robust preconditioned CG method [19]. The solution technique is described in more details in [20].

7. DEFORMATION OF TRIANGULATIONS

Since in the discrete model mapping (3) is replaced by the piecewise-affine mapping, we need to investigate the problems related to the deformations of triangular planar grids. Each valid plane triangulation (i.e. triangulation without self-intersections and non-positive areas of the triangles) with given connectivity of triangles defines the connected set consisting of all triangulations which can be obtained from the initial one via node movement, i.e. via continuous deformation such that all intermediate triangulations arc valid as well. Since any triangulation with a correct planar connectivity is a particular case of a planar graph with a prescribed order of edges around each vertex, each triangulation can be made valid, i.e. it can be drawn on the plane without self-intersections [21, 22]. In other words, the feasible set in this case is not empty. However it is not clear whether such a feasible set is globally connected.

When some nodes of the triangulation are fixed, counterexamples can be constructed since for multiply connected plane domains it is easy to construct such boundary polylines that different triangulations with the same connectivity inside the domain cannot be morphed into each other.

Our main problems of interest are (1) construction of a valid triangulation from the tangled one via node movement and (2) morphing of a valid triangulation into an optimal one, subject to target shape and size specification.

The definition of a feasible set via local properties of triangles is not enough to guarantee that the solution of the minimization problem will not have self-intersections. In the plane case at least two basic pathologies occur. The first one is the locking on the boundary of triangulation shown in Figure 5 (left).

In this case the local minimum of the discrete functional is the grid with multivalued behaviour around some boundary vertex and it can be present even for very regular connectivity patterns of triangulations. And the gradient method cannot ‘jump away’ from this parasite solution.

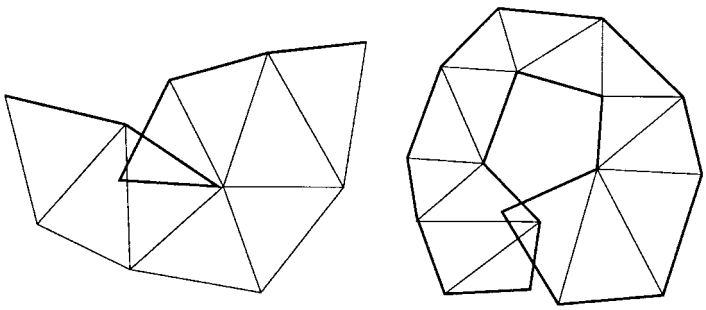

Figure 5. Locking phenomenon on the boundary (left) and globally overlapping triangulation (right).

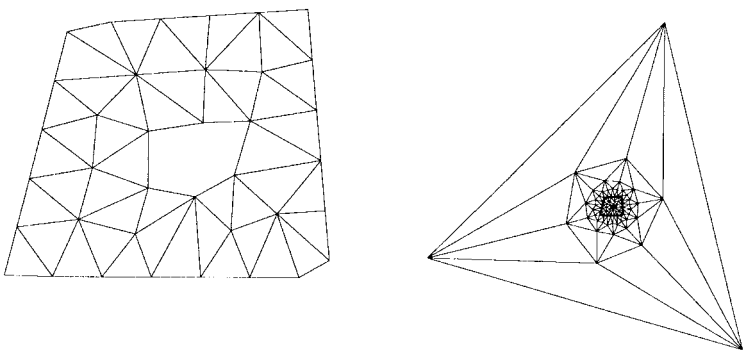

Figure 6. Sample planar graph (left) and scaled triangulated planar graph with original graph inside (right).

The global overlap is more natural phenomena since it can provide global optimal triangulation where triangles are as close to target shapes and sizes as possible.

There is no way to control both phenomena locally. However, these problems can be solved quite easily using the fact that any planar graph can be triangulated, i.e. edges can be added in such a way that all faces, including the external one, become triangles (Figure 6).

For such a triangulation when all signed areas of triangles are positive any internal self-intersections are impossible since at least one such triangulation without self-intersections always exists and the current triangulation can be represented as the image of valid triangulation under one-to-one mapping.

Moreover, for such triangulations the feasible set is simply connected. The proof is based on results from Reference [23]. The nodes of any valid triangulation of any convex domain on the plane can be found as the unique solution of a linear system with a square non-singular M-matrix. Consider two different triangulations $\mathscr{T}_1$ and $\mathscr{T}_2$ such that the external boundary in each case is a single triangle and the connectivity and numbering of vertices is the same, only the vectors of the co-ordinates of the triangulation nodes R_1, R_2 are different. Without loss of generality we can assume that the boundary triangles are the same, since an affine

mapping can be applied to one triangulation to make the boundaries coincide. Hence we get the problem from Reference [23]—morphing of triangulations with fixed boundary nodes lying on closed convex polyline.

Then R_1, R_2 are the solutions of linear systems

$$S_1 R_1 = F_1, \quad S_2 R_2 = F_2 \tag{23}$$

where the right-hand side is due to the presence of fixed boundary nodes.

The details on how to construct such linear systems can be found in Reference [11]. The solution to the linear system

$$S_t R_t = F_t, \qquad S_t = tS_1 + (1-t)S_2, \quad F_t = tF_1 + (1-t)F_2$$

also exists and is unique since S_t is a non-singular M-matrix as well. The vector R_t makes up the nodes of the valid triangulation for any $0 \leqslant t \leqslant 1$ which means that the deformation path connecting two triangulations always exists. Moreover, if for two triangulations $\mathscr{T}_1$ and $\mathscr{T}_2$ one is able to construct two different deformation paths defined by parametric matrices S_t and $\tilde{S}_t$ then the linear combination $\theta S_t + (1-\theta)\tilde{S}_t$ also defines a valid triangulation for any $0 \leqslant \theta \leqslant 1$, which means that any deformation path can be morphed into another path and the feasible set is simply connected. In fact the results from Reference [23] are more general and can be applied to a mixed grid consisting of triangles and quadrilaterals.

Also method [11] by itself was developed in order to compute one-to-one projections of triangulated surfaces on the plane. The main advantage of this method is its simplicity and low computational complexity. It allows to compute non-degenerate projections very fast provided that a robust linear solver is available. However our goal is the construction of quasi-isometric mappings hence the non-degenerate projection of a surface triangulation is only an initial guess to other morphing procedure.

The modified optimality principle (12) is applied in this case with $q = 1$ for all triangles present in the original triangulation and $q = 0$ for all newly added triangles. After untangling phase, all self-intersections are eliminated so during the contraction of the feasible set the global upper bound on the distortion measure is minimized only for the original triangulation, while the auxiliary one only prevents reappearance of global overlaps.

This solution was found to be extremely reliable and survived extensive testing on hard industrial problems.

It was found that initial non-degenerate triangulation computed using the method [11] generally proved to be a poor initial guess for the mapping optimization technique. It is well known that the stiffness of the non-linear minimization problem may depend on the current guess for the solution. In this case the problem was extremely stiff during the initial steps of the solution. On the other hand, the non-degenerate triangulation computed via the continuation technique for the functional (21), while taking more time, generally provides much better guess to the quasi-isometric mapping and as a result the non-linear minimization problem is not so stiff. So the total computation time was generally more favourable when using untangling technique based on continuation.

8. NUMERICAL EXPERIMENTS

The presented numerical experiments cover only two applications: projections of triangulated surfaces and morphing based on rotation of surface triangles.

In the first test case, the surface is modified via minimization of the functional (13) with $n=3$, $m=2$. For each triangle vector $\mathbf{q}_1$ is prescribed in the following way:

$$\mathbf{q}_1 = \begin{cases} \mathbf{n} & \text{when } n_3/\sqrt{n_1^2+n_2^2} \geqslant c \\ \mathbf{b}/|\mathbf{b}|,\ \mathbf{b}=(n_1,n_2,c\sqrt{n_1^2+n_2^2})^{\mathrm{T}} & \text{when } n_3/\sqrt{n_1^2+n_2^2} < c \end{cases}$$

where $c=1$ and $\mathbf{n}$ is the unit normal vector to the initial triangle plane.

The surface before morphing is shown in Figure 7 (left), while the result of minimization is shown on the right. It can be seen that such a morphing respects small local features on the surface despite large displacements and deformations. The total number of triangles in this surface is about 30 000 and only small fragment is shown.

The same surface was developed on the plane without any additional triangulation and using a planar graph triangulation approach.

In the first case shown in Figure 8 (left) the projection is self-intersecting, while with additional constraints there are no overlaps which is shown on the right.

The robustness of the projection method is illustrated in another example shown in Figure 9.

Here the surface is described by two different triangulations. The first one is based on the repair of the surface initially consisting of unconnected facets and has triangles with extremely low quality. The ratio of maximal-to-minimal edge lengths exceeds 10^5. The second one has a high quality.

The non-linear solver handled this extremely stiff problem and computed the quasi-isometric projection. In both cases the final global condition number of the mappings was of the order of unity and was almost insensitive to the surface triangulation quality. Meanwhile the projection technique from Reference [11] generally resulted in global condition numbers of mappings in the range 10^4–10^9 for low-quality surface triangulations. The reason for such a behaviour is that the method from Reference [11] is close in spirit to the methods based on harmonic

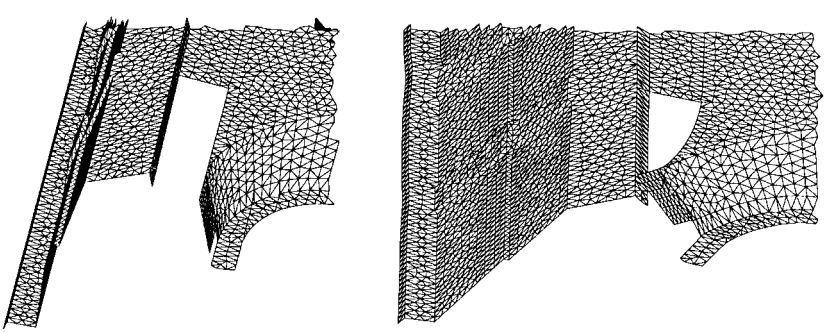

Figure 7. Fragment of the initial surface (left) and the surface after morphing (right).

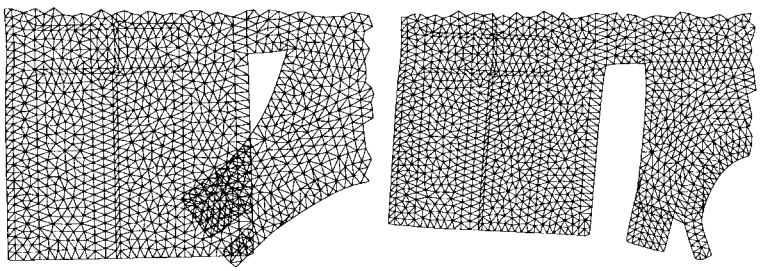

Figure 8. Developed surface without self-intersection control (left) and with self-intersection control based on planar graph triangulation (right).

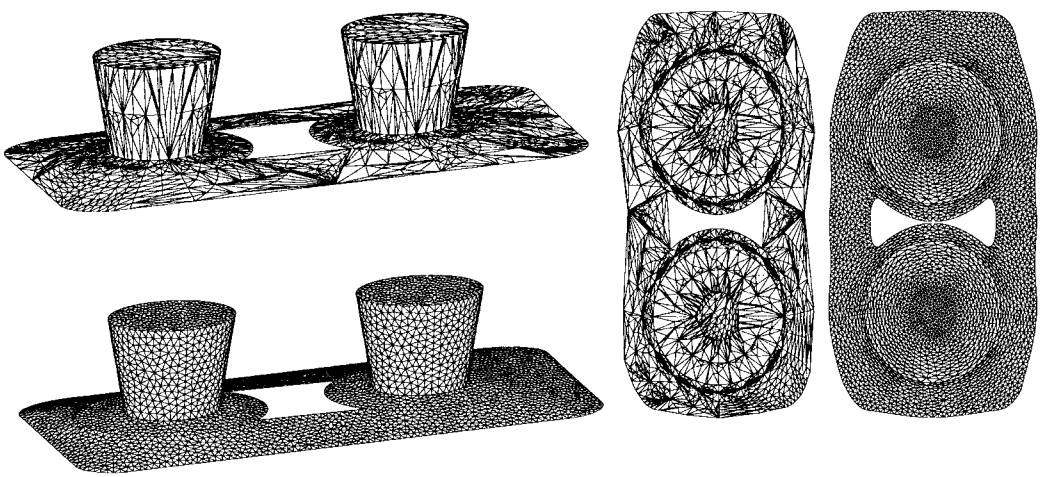

Figure 9. An initial surface with extremely low-quality triangulation (upper left), a surface with quality triangulation (lower left) and flattened surfaces (right).

maps which may lead to very large global condition numbers. Detailed direct comparison of methods based on harmonic and quasi-isometric mapping which was done in Reference [17] in the case of structured grids led exactly to the same conclusions. A heuristic method for decreasing metric distortion of flattening suggested in Reference [7] allows for smaller distortions; however, it still creates large condition numbers for bad surface triangulations and does not provide bounded condition numbers when the grid is refined.

The application of surface morphing to these surfaces is shown in Figure 10.

The numerical results demonstrate that the projection and morphing tool are able to provide grid-independent results and can be applied in the case of extremely ill-conditioned triangulations. In fact, all quality grids shown above were constructed via projection of initial bad triangulation on the plane, by meshing the resulting plane domains and by mapping resulting triangulations back onto real surfaces.

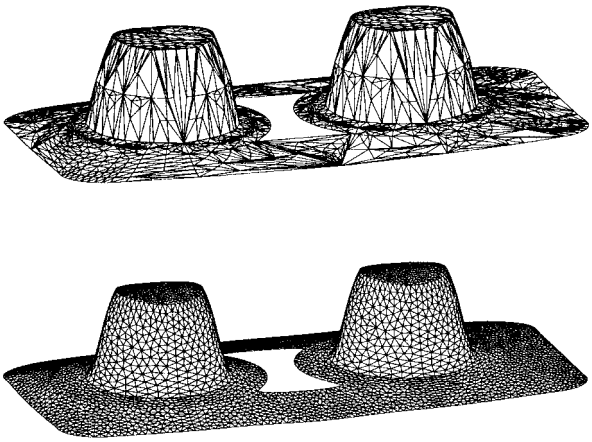

Figure 10. Morphing of the surface with two different triangulations.

9. CONCLUSIONS

The mapping optimality principle (7) can be considered as a general principle applicable to most grid generation/adaptation problems and to many optimization problems based on spatial mappings. The presented method is the first consistent and reliable implementation of this optimality principle. Industrial-level robustness is achieved by the method. Numerical evidence suggests that semi-discrete mappings created by the method tend to be smooth, when this notion makes sense, and grid convergence is observed when increasing the number of grid cells. A rigorous analysis of the method requires, however, further research.

ACKNOWLEDGEMENTS

This work has been partially supported by the NWO-NL-RFBR Grant dossier No. 047-008-007.

REFERENCES

1. Godunov SK, Prokopov GP. Calculation of conformal mappings and grid generation. *Zhurnal Vychislitel'noy Matematiki i Matematicheskoy Fiziki* 1967; **7**(6):902–914.
2. Brackbill JU, Saltzman JS. Adaptive zoning for singular problems in two dimensions. *Journal of Computational Physics* 1982; **46**(3):342–368.
3. Jacquotte OP. A mechanical model for a new grid generation method in computational fluid dynamics. *Computer Methods in Applied Mechanics and Engineering* 1988; **66**:323–338.
4. Liseikin VD. Construction of regular grids on *n*-dimensional manifolds. *Zhurnal Vychislitel'noy Matematiki i Matematicheskoy Fiziki* 1991; **31**(11):1670–1683.
5. Ivanenko SA. Adaptive-harmonic grids. Communications in Applied Mathematics, Computing Center RAS, 1997.
6. de Almeida VF. Domain deformation mapping: application to variational mesh generation. *SIAM Journal on Scientific Computing* 1999; **20**(4):1252–1275.
7. Hormann K, Greiner G. MIPS: an efficient global parameterization method. In *Curve and Surface Design*, Laurent P-J, Sablonnière P, Schumacher LL (eds). Vanderbilt University Press: Nashville, 2000; 163–172.
8. de Boor C. Good approximation by splines with variable knots. II. *Conference on numerical solution of differential equations. Lecture Notes in Mathematics*, vol. 363. Springer: Berlin, 1973; 12–20.
9. Godunov SK, Gordienko VM, Chumakov GA. Quasi-isometric parameterization of a curvilinear quadrangle and a metric of constant curvature. *Siberian Advances in Mathematics* 1995; **5**(2):1–20.

10. Garanzha VA. Barrier method for construction of quasi-isometric grids. *Computational Mathematics and Mathematical Physics* 2000; **40**:1617–1637.
11. Floater MS. Parameterization and smooth approximation of surface triangulations. *Computer Aided Geometrical Design* 1997; **14**:231–250.
12. Iwaniec T. The failure of lower semicontinuity for the linear dilatation. *Bulletin of the London Mathematical Society* 1998; **30**:55–61.
13. Reshetnyak YG. Mappings with bounded deformation as extremals of Dirichlet type integrals. *Siberian Mathematical Journal* 1968; **9**:487–498.
14. Vodopyanov SK, Goldshtein VM. Quasiconformal mappings and spaces of functions with generalized first derivatives. *Siberian Mathematical Journal* 1976; **17**:399–411.
15. Ball JM. Convexity conditions and existence theorems in nonlinear elasticity. *Archives for Rational Mechanics and Analysis* 1977; **63**:337–403.
16. Rockafellar RT. *Convex Analysis*. Princeton University Press: Princeton, NJ, 1970.
17. Garanzha VA. Barrier variational generation of quasi-isometric grids. *Numerical Linear Algebra with Applications* 2001; **8**:329–353.
18. Garanzha VA, Kaporin IE. Regularization of the barrier variational grid generation method. *Computational Mathematics and Mathematical Physics* 1999; **39**(9):1426–1440.
19. Kaporin IE. On preconditioning conjugate gradient method when solving discrete counterparts of differential problems. *Differentsialnye Uravneniya* 1990; **26**(7):1225–1236.
20. Branets LV, Garanzha VA. Global condition number of trilinear mapping. Application to 3-D grid generation. In Ivanenko SA, Garanzha VA (eds). *Grid Generation: New Trends and Applications in Real-Life Simulations*, Communications on Applied Mathematics, Computing Center Russian Academy of Sciences: Moscow, 2001; 45–60.
21. Fary I. On straight line representation of planar graphs. *Acta Scientiarum Mathematicarum Szeged* 1948; **11**:229–233.
22. Tutte WT. How to draw a graph. *Proceedings of London Mathematical Society* 1960; **10**:304–320.
23. Floater MS, Gotsman C. How to morph tilings injectively. *Journal of Computational and Applied Mathematics* 1999; **101**:117–129.

