Variational principles in grid generation and geometric modelling: Theoretical justifications and open problems

Article *in* Numerical Linear Algebra with Applications · June 2004

DOI: 10.1002/nla.363

1 author:

Vladimir A. Garanzha

Russian Academy of Sciences

NUMERICAL LINEAR ALGEBRA WITH APPLICATIONS
Numer. Linear Algebra Appl. 2004; **11**:535–563 (DOI: 10.1002/nla.363)

Variational principles in grid generation and geometric modelling: theoretical justifications and open problems

V. A. Garanzha*,†

Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 119991, Russia

SUMMARY

The paper is devoted to analysis of variational principles for construction of mappings with prescribed properties in grid generation and geometric modelling. An attempt is made to formulate general requirements which should be satisfied by the variational principle. Theoretical justification is considered along with review of unsolved problems and fundamental mathematical difficulties. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: grid generation; variational principle; polyconvexity; quasi-isometric mapping; manifold of bounded curvature

1. INTRODUCTION

Variational grid generation methods [1–12] have become important tools in many real-life applications. Despite considerable progress in this area, lots of unsolved problems still exist. Some of them are the subject of this paper.

With onset of unstructured grid generation methods the widespread opinion was that grid generation methods in a classical sense, namely mapped 'elliptic' grid generators and variational grid generators, are not necessary any more. Now ready methods to split computational domains into simple standard subdomains are available. Or meshless methods can kill the idea of grid generation itself. However this initial euphoria was over very fast.

First of all it was found that 'meshless' methods require grid generation. Another example is that best unstructured mesh improvement can be achieved using variational methods, which in fact are exactly those suggested for mapped grid generation but in different guise. Another reason is that lots of new applications for variational mapping construction methods were found, such as sensitivity analysis and topological optimization, geometric design, design of

*Correspondence to: V. A. Garanzha, Computing Center of Russian Academy of Sciences, Vavilova 40, Moscow 119991, Russia.
†E-mail: garan@ccas.ru

Contract/grant sponsor: Grant NWO-NL-RFBR; contract/grant number: 047-008-007
Contract/grant sponsor: Fundamental research programs of Russian Academy of Sciences; contract/grant numbers: 03, 16, 17

Received 22 January 2003
Revised 9 September 2003
Accepted 6 October 2003
Published online 5 May 2004

machining tools, animation, morphing, texture mappings, computational anatomy problems, including brain mapping, colonoscopy, and many others.

As a result, engineers started to understand that the common denominator of a successful technique is that (a) a variational principle should be used, (b) variational problem should be well posed.

One of the basic motivations of the research outlined in this paper is that the drawbacks in rigorous foundations can adversely affect engineering properties of the resulting practical methods. In particular, one has to understand clearly the valid domain of application of a certain technique as well as its intrinsic 'breakdown threshold'. This notion is hard to define formally, but given that in the state-of-the-art engineering development manual intervention is supposed to be very small, the natural question for engineers is 'will this black box method solve my problem and at what cost?', or 'is it safe to use this tool in the production cycle?'.

Theoretically it is possible to 'break down' any technique. This is particularly simple for mapped finite difference elliptic grid generation methods [13]. In Reference [8] examples of simple domains were presented where finite difference methods were not able to construct non-degenerate grids, while discrete variational methods, suggested in Reference [8], handled the same problems successfully. Such breakdowns can be attributed to the lack of rigorous existence and unisolvability results for generating equations in the practically interesting cases. The convergence of finite difference (or any other) approximations to the solutions of continuous problems is also not proven.

For discrete variational methods the 'breakdown threshold' is much higher, but it still exists and simply has a different expression. Non-degenerate grids can always be constructed. However the grid properties can be very far from expected. One can encounter various instabilities, such as non-unique grids, lack of convergence of sequence of grids, grid instabilities with respect to small variations of input data and 'non-smooth' structured grids despite the fact that the functional of interest is elliptic [14]. The main problem is that there is no simple criteria to predict the appearance of such phenomena. In most (but not all) cases it goes virtually undetected and as a result little attention is paid to these problems.

Another problem is that the variational methods for construction of mappings are derived based on the assumption that the underlying spatial mappings are smooth and one-to-one [7]. Unfortunately, such properties are not provable for minimizers of variational problems. Even in the case of harmonic mappings technique from a practical point of view one has to consider two different solutions. The first one is obtained using minimization of the Dirichlet-type functional (or solving the Beltrami equations) and is smooth and unique. However it is well known that in order to obtain the computational grid one has to exchange dependent and independent variables. As a result it is necessary to minimize highly non-linear non-convex functionals. Contrary to the suggestions from Reference [7], these two problems are not provably equivalent, the solutions are not provably the same and the conclusions derived from the analysis of Beltrami equations can serve only as heuristics.

It is also obvious that the classes of geometrical objects of interest to engineers should include non-smooth objects. Hence classical results from differential geometry and geometry of Riemannian manifolds are of limited use in this case. In particular, the theory of manifolds of bounded curvature (MBC) which is reviewed in the current paper is based on concepts of distances and angles. The Christoffel symbols found in differential geometry are simply not used! The classical grid generation methods may exhibit various singularities when non-smooth boundaries and metrics are present. As a result the behaviour of these methods is unpredictable in the range of high curvature objects to non-smooth objects which are very important for engineers.

Another example is related to sensitivity analysis. Here the natural assumption is that variations of input data are small, but generally one cannot assume that they are smooth! The conclusion is that the variational method should not exhibit such pathological sensitivity with respect to curvature and thus should not make excessive appeal to the smoothness of input data.

The variational principles in grid generation and geometric modelling in principle have artificial nature. In particular they are not introduced as approximations of physical models. This ambiguity allows one to construct a good variational formulation as soon as clear understanding of its underlying principles is available. The main idea which is adopted in this work is that the continuous variational principle should be considered. This continuous variational principle should provide a spatial mapping as a minimizer of the variational problem. This spatial mapping in turn should provide a provably optimal parameterization for the problem of interest. The target finite element grid should be constructed as the finite element solution of the continuous variational problem.

The problems related to validation of the above procedure are considered.

2. GENERAL REQUIREMENTS FOR VARIATIONAL PRINCIPLE

1. The variational problem should be well posed, its solution should exist and should be stable with respect to input data.
2. The variational principle should not admit singular mappings as minimizers, hence the class of admissible mappings consists of locally invertible quasi-isometric mappings [15]. By definition, under quasi-isometric mapping the ratio of the distance between any two close enough points and the distance between their images is uniformly bounded from below and from above.
3. The solutions of the variational problem should be as smooth as possible.
4. The ability to construct quasi-uniform mappings is a key property in grid generation. Quasi-uniformity means minimal or at least bounded distortion of distance between any close enough points under the mapping.
5. The deviation of the solutions of the variational problem from the target ones should be bounded in maximum norm. The target solution as a rule is unknown mapping with known properties. This principle also naturally leads to the quasi-isometry concept.
6. The variational principle should make sense for quite general spatial domains. In particular, the boundary smoothness requirements should not be very restrictive.
7. The solutions of the variational problem should be orientation-preserving and with proper boundary conditions should be globally one-to-one.
8. The discretized variational principle should make sense, its solution should converge to the minimizer of the continuous variational problem.
9. The discrete solutions (say piecewise-affine mappings) should be locally invertible and with additional constraints should be globally invertible.

Assumption that the variational principle itself is completely artificial implies that it can be as simple as possible. Previous experience shows that most of the variational principles successfully used in grid generation are very similar to the stored energy functionals in hyperelasticity theory [16], but generally free from hyperelasticity axiomatics.

Let us denote by Ω_0 an open bounded domain in co-ordinates $\eta=\{\eta_1,\ldots,\eta_n\}$. The spatial mapping of interest $y(\eta):\Omega_0\to\mathbb{R}^n$ is constructed as a minimizer of the functional depending on the gradient of the mapping:

$$\int_{\Omega_0} f(\nabla_\eta y)\,\mathrm{d}\eta, \quad f:\mathbb{R}^{n\times n}\to\mathbb{R} \tag{1}$$

where $\nabla_\eta y$ denotes matrix with entries $\partial y_i/\partial\eta_j$. In fact this is Jacobi matrix of the mapping $y(\eta)$.

Comprehensive analysis of the mathematical formulation of minimization problems for such class of functionals, including analysis of conditions of well posedness, regularity of solutions, restrictions on domains and boundary conditions can be found in Reference [16].

The *quasi-isometry* means that for any two sufficiently close points in η co-ordinates, say α and β, the following inequality holds:

$$\frac{L}{C}|\beta-\alpha|\leqslant|y(\beta)-y(\alpha)|\leqslant LC|\beta-\alpha| \tag{2}$$

where $C>0$ is a constant, $|\cdot|$ is the Euclidean length and L is the length scale. The quasi-isometry implies that the function $y(\eta)$ is not just continuous, but also bi-Lipschitz continuous.

Let

$$\mathbb{W}^{k,p}(\Omega,\mathbb{R}^n)=\{v_i\in\mathbb{L}^p(\Omega);\,\partial^\alpha v_i\in\mathbb{L}^p \text{ for any multi-index } |\alpha|\leqslant k;\ i=1,\ldots,n\},\quad 1\leqslant p\leqslant\infty$$

be the standard Sobolev space.

We will use another definition of quasi-isometry. The mapping $y(\eta)\in\mathbb{W}^{1,1}(\Omega_0,\mathbb{R}^n)$, is called quasi-isometric if the inequality

$$\frac{L}{C_1}\leqslant\sigma_i(\nabla_\eta y)\leqslant LC_1 \tag{3}$$

is satisfied almost everywhere in Ω_0. Here σ_i are the singular values of $\nabla_\eta y$ (or the square roots of the eigenvalues of the matrix $\nabla_\eta y^{\mathrm{T}}\nabla_\eta y$) and $C_1>0$ is a constant.

From (2) it follows that the gradient of $y(\eta)$ belongs to $\mathbb{L}_\infty$ and inequality (3) holds. On the other hand, any function with a gradient belonging to $\mathbb{L}_\infty$ is locally Lipschitz continuous, in a sense that there exists a continuous representative of the equivalence class [17]. In what follows we will use (3) as a definition of a quasi-isometric mapping. In fact equivalence of these two definitions is not that obvious but we will use both of them rather loosely anyway keeping in mind that a precise mathematical statement of equivalence can be formulated.

Instead of quasi-isometry, it is possible to consider a weaker constraint

$$\frac{1}{C_2}\leqslant\frac{\sigma_i(\nabla_\eta y)}{\sigma_j(\nabla_\eta y)}\leqslant C_2$$

almost everywhere in Ω_0. Here $C_2 \geqslant 1$ is a constant. The above condition is called the *bounded distortion* inequality. It also means that the pointwise distortion measure (which is sometimes called linear dilatation)

$$\mu(\nabla_\eta y) = \frac{\sigma_{\max}(\nabla_\eta y)}{\sigma_{\min}(\nabla_\eta y)} \tag{4}$$

is bounded from above almost everywhere. Obviously mappings with bounded distortion can violate quasi-isometry constraints.

There are several well-established approaches to construct spatial mappings. The *harmonic mappings* approach is relatively simple and is based on convex functionals [18]. In many practically important 2-D cases it guarantees the construction of one-to-one mappings [19–22]. However in the presence of non-smooth boundaries the dilatation of a harmonic mapping (i.e. the ratio of maximal to minimal singular values of the Jacobi matrix) near boundaries can be unbounded. This phenomenon is known in mechanics as stress concentration. In the case of a harmonic map with a non-smooth metric the unbounded dilatation can be present far away from the boundaries [22]. When singularities are present, the harmonic maps are not stable with respect to input data. So despite its considerable success in the field of grid generation, the harmonic mappings approach can be considered only as a solution technique for some particular problems.

The *conformal mappings* technique provides a rather good answer to the above formulated requirements in some simple cases.

In Reference [15] it was shown how to construct mapping of curvilinear quadrangle with smooth sides onto a square being both conformal with respect to special metric and quasi-isometric. The existence and uniqueness of the solution of the variational problem was proved as well. Since the minimizer of the variational problem is smooth enough, it is guaranteed that the discrete solution converges to the exact solution, and when the mesh size is small enough this discrete solution is a globally invertible piecewise-bilinear mapping, i.e. it is an unfolded mesh. However in this approach the total metric distortion, i.e. the constant C_1 is far from minimal. This approach cannot be applied for domains with non-smooth boundaries and cannot be generalized to the 3-D case.

There were also some more or less successful attempts to apply variational principles from *hyperelasticity* to grid generation, see for example References [6, 23]. It is unclear whether the above requirements are satisfied in these works.

Some new theoretical results related to variational principles for *crystallography* in 3-D [24] also can be relevant for grid generation. In Reference [24] the following results are presented: for a certain variational principle in 3-D the minimizing spatial mapping exists, is unique and globally invertible and is smooth in a classical sense. The applicability of these results for grid generation and geometric modelling requires further research.

3. HOW TO OBTAIN WELL-POSED VARIATIONAL PROBLEM

The basic principles of spatial mappings construction via solution of well-posed variational problems are formulated in the context of the mathematical theory of hyperelasticity with finite deformations. Let us briefly review the basic mathematical ideas for variational problems of the type (1) to be well posed as formulated by Ball [25]:

1. The function $f(\nabla_\eta y)$ should be *polyconvex*, namely it can be written as a convex function of minors of $\nabla_\eta y$. In 2-D case it means that there exists a convex function $g(\cdot,\cdot)$, such that $f(\nabla_\eta y) = g(\det \nabla_\eta y, \nabla_\eta y)$. In the 3-D case the existence of a convex function $g(\cdot,\cdot,\cdot)$ is assumed, such that $f(\nabla_\eta y) = g(\det \nabla_\eta y, \nabla_\eta y, \operatorname{adj} \nabla_\eta y)$, where $\operatorname{adj} Q = Q^{-\mathrm{T}} \det Q$ denotes the adjugate matrix.
2. $f(\nabla_\eta y)$ should possess the so-called *barrier property*, namely it should be bounded from below and should tend to $+\infty$ when feasible $y(\eta)$ tends to the boundary of the feasible set, for example when $\det \nabla_\eta y \to +0$.
3. $f(\nabla_\eta y)$ should satisfy certain *growth conditions* [25].
4. The set of admissible mappings should be defined by a *polyconvex inequality* [26].
   The above conditions allow to prove an existence theorem for the functional of interest. In particular, for the 3-D case Ball introduces [25] the following set of admissible mappings Φ:

$$\Phi = \{u \in \mathbb{W}^{1,p}(\Omega_0); \operatorname{adj} \nabla u \in \mathbb{L}^q(\Omega_0), \det \nabla u \in \mathbb{L}^r(\Omega_0), u = u_0 \text{ on } \partial\Omega_0, \det \nabla u > 0 \text{ a.e. in } \Omega_0\} \tag{5}$$

where 'a.e.' means almost everywhere. The growth conditions, which are consistent with the definition of admissible deformations, are defined as follows:

$$f(A) \geqslant \alpha(\|A\|^p + \|\operatorname{adj} A\|^q + (\det A)^r) + \beta$$

for all $A \in \mathbb{M}^3_+$, where the constants are as follows:

$$\alpha > 0, \quad p \geqslant 2, \quad q \geqslant p/(p-1), \quad r > 1$$

and $\mathbb{M}^3_+$ is the set of 3×3 matrices with positive determinant. Here $\|\cdot\|$ means the Frobenius norm of the matrix: $\|A\| = \sqrt{\operatorname{tr}(A^{\mathrm{T}} A)}$.

The existence theorems are based on the assumption that there exists at least one element of Φ providing a finite value of the functional. In this case the minimizer of the functional exists and belongs to Φ. The above assumption is in fact a quite non-trivial hidden constraint on Ω_0 and on the boundary condition $u_0(\partial\Omega_0)$. Unfortunately the existence theory itself does not provide a simple criteria allowing to establish the validity of this assumption. So from a practical point of view this existence theory is not closed.

Since the original paper of Ball [25] lots of variations of this existence theory were considered. It was applied to many different problems, which required modifications in, say, the definition of growth conditions and definition of the set of admissible deformations. But the most fundamental property—polyconvexity cannot be relaxed. Any polyconvex functional is also rank one convex, namely

$$f(\lambda S_1 + (1-\lambda)S_2) \leqslant \lambda f(S_1) + (1-\lambda) f(S_2) \quad \text{where rank } (S_1 - S_2) \leqslant 1, \ 0 \leqslant \lambda \leqslant 1$$

When $f(\cdot)$ is smooth enough, the rank one convexity is equivalent to the Legendre–Hadamard (ellipticity) condition. Thus we come to the conclusion that 'elliptic grid generation' methods should be the ones based on the polyconvex variational principles.

4. POLYCONVEXITY OF THE SET OF ADMISSIBLE MAPPINGS

The formal definition of the polyconvex set of admissible spatial mappings can be written as follows [16]:

$$L(\nabla_\eta y) \leqslant 0, \quad L : \mathbb{R}^{n\times n} \to \mathbb{R}$$

where L is the polyconvex function of minors of $\nabla_\eta y$.

One can consider the minors of $\nabla_\eta y$ as independent variables and consider the Cartesian space where these minors define the co-ordinates. In particular, we will use the n^2+1-D space of matrix $\nabla_\eta y$ entries plus its determinant. Some important examples of polyconvex constraints are presented below.

1. Orientation-preserving mappings are defined by the inequality

$$\det \nabla_\eta y > 0 \quad \text{a.e. in } \Omega_0 \tag{6}$$

This inequality defines a half-space, i.e. a convex domain in extended co-ordinates, which is shown in Figure 1(left)

2. 'Stiffening' function by Ciarlet and Necas [26]:

$$\mathrm{tr}(E^{d^{\mathrm{T}}} E^d) \leqslant \alpha, \quad E^d = E - \tfrac{1}{3}(\mathrm{tr}\, E)I, \quad E = \tfrac{1}{2}(\nabla_\eta y^{\mathrm{T}} \nabla_\eta y - I) \quad \text{a.e. in } \Omega_0 \tag{7}$$

where $\alpha > 0$ is a constant. This constraint does not allow the singular values of $\nabla_\eta y$ to be much different from each other, meaning that the shape distortion measure (4) is bounded. In Reference [16] it was stated that this constraint allows 'polyconvexification', meaning that it can be replaced by a equivalent polyconvex relation. Inequality (7) does not define the sign of the determinant. The definition of the set of admissible deformations (5) can be augmented by constraint (7) and the variational problem with functional (1) will still be well posed [16, 26].

3. Another convenient polyconvex definition of mappings with bounded distortion was suggested in Reference [27].

$$\det \nabla_\eta y > K\left(\frac{1}{n}\,\mathrm{tr}(\nabla_\eta y^{\mathrm{T}} \nabla_\eta y)\right)^{n/2} \quad \text{a.e. in } \Omega_0,\ 0 < K \leqslant 1 \tag{8}$$

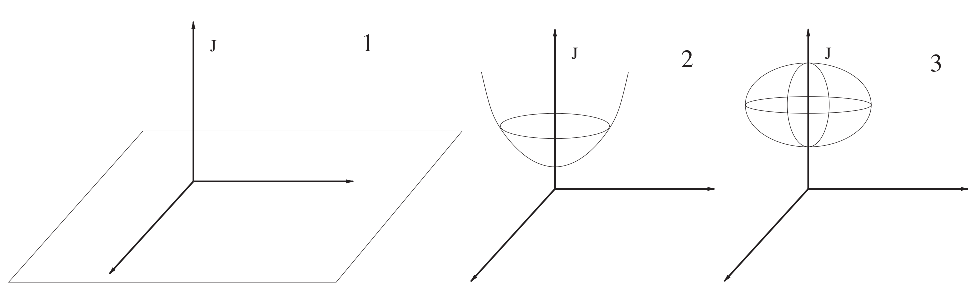

Figure 1. 3-D cross-section of feasible sets in extended co-ordinates. 1—the set of orientation-preserving mappings, 2—the set of mappings with bounded distortion [27], 3—the polyconvex set of quasi-isometric mappings [11].

In extended co-ordinates this inequality defines a paraboloid of high degree, see Figure 1(centre). This inequality simultaneously constrains the shape distortion of the mapping and guarantees that the mapping is orientation-preserving. Thus (8) can replace simultaneously both (6) and (7).

4. In Reference [11] the following polyconvex definition of the set of quasi-isometric mappings was suggested:

$$\det \nabla_\eta y > t\phi_\theta(\det \nabla_\eta y, \nabla_\eta y) \quad \text{a.e. in } \Omega_0, \ 0 < t \leqslant 1$$

where $y(\eta) \in \mathbb{W}^{1,1}(\Omega_0, \mathbb{R}^n)$ and

$$\phi_\theta(J, T) = \theta \left(\frac{1}{n} \operatorname{tr}(T^{\mathrm{T}} T)\right)^{n/2} + \frac{(1-\theta)}{2}\left(v + \frac{J^2}{v}\right), \quad 0 < \theta < 1$$

$v > 0$ is a constant which defines the target average value of $\det \nabla_\eta y$. This inequality defines an ellipsoid of high degree in extended co-ordinates, see Figure 1(right).

5. LOCALLY AND GLOBALLY INVERTIBLE SPATIAL MAPPINGS

In order to be applicable for geometric modelling, the minimizers of the variational problems should be globally invertible mappings, or in some particular cases at least locally invertible. From a practical point of view it means that relatively easily checked and/or guaranteed algebraic properties of the solutions should lead to the required global topological properties.

Important examples of such relations are presented in Reference [27]. There it was shown that a mapping $y(\eta) \in \mathbb{W}^{1,n}(\Omega_0, \mathbb{R}^n)$ satisfying the bounded distortion inequality (8) almost everywhere is either constant or open and discrete. Namely, the image of any open set under this mapping is an open set, and any point can have only finite number of preimages.

In Reference [28], an inverse function theorem was proved, which is fully reproduced below due to its importance.

Theorem (*Ball* [28])
Let $\Omega \subset \mathbb{R}^n$ be a non-empty bounded connected strongly Lipschitz open set. Let $u_0 : \bar{\Omega} \to \mathbb{R}^n$ be continuous in $\bar{\Omega}$ and one-to-one in Ω. Let $p > n$ and let $u(x) \in \mathbb{W}^{1,p}(\Omega; \mathbb{R}^n)$ satisfy $u|_{\partial\Omega} = u_0|_{\partial\Omega}$, $\det \nabla u(x) > 0$ almost everywhere in Ω. Let $u_0(\Omega)$ satisfy the cone condition, and suppose that for some $q > n$

$$\int_\Omega |\nabla u^{-1}(x)|^q \det \nabla u(x) \, \mathrm{d}x < \infty$$

Then u is a homeomorphism of Ω onto $u_0(\Omega)$, and the inverse function $x(u)$ belongs to $\mathbb{W}^{1,q}(u_0(\Omega); \mathbb{R}^n)$. The matrix of weak derivatives of $x(\cdot)$ is given by

$$\nabla x(v) = \nabla u^{-1}(x(v)) \text{ almost everywhere in } u_0(\Omega)$$

If, further, $u_0(\Omega)$ is strongly Lipschitz, then u is a homeomorphism of $\bar{\Omega}$ onto $u_0(\bar{\Omega})$.

Here $|\cdot|$ is the spectral matrix norm. The *cone condition* means that there exist a fixed cone, such that any boundary point of $u_0(\Omega)$ is the vertex of a cone which belongs to $u_0(\Omega)$ and is congruent to this fixed cone.

This theorem is of key importance for the variational principles in geometric modelling since the minimizers of variational problems are orientation-preserving but, unfortunately, are not probably regular enough in order to be locally invertible. A quite disappointing example of such behaviour is suggested in Reference [28]. The mapping u of the unit disc $D=\{|x|<1\}$ in $\mathbb{R}^2$ is given in polar co-ordinates r,φ by

$$u:(r,\varphi)\to\left(\frac{1}{\sqrt{2}}r,2\varphi\right)$$

Obviously the mapping u is quasi-isometric, $\det\nabla u(x)=1$ if $x\neq 0$ and the singular values of $\nabla u(x)$ are equal to $\frac{1}{\sqrt{2}},\sqrt{2}$ almost everywhere. However, u is not locally invertible at the origin. In fact the mapping $u(x)$ does not define a manifold parameterization.

Thus, in practically useful cases the global invertibility of orientation preserving minimizers is assured directly from regularity properties and boundary conditions avoiding the proof of local invertibility.

The above theorem can be used to analyse whether a computational grid is a globally invertible mapping. In fact this theorem is too general for that problem. For global invertibility it is enough to assert that the grid (i.e. the piecewise-smooth mapping) coincides on the boundary with the continuous homeomorphism, is orientation preserving and is Lipschitz continuous. Obviously mappings defined on finite elements grids with non-degenerate low-order elements are Lipschitz continuous so global invertibility results from References [29, 30], can be applied directly.

Nevertheless, Ball's inverse function theorem has another important application in grid generation: it provides rigorous grounds for exchanging dependent and independent variables. It is very well known since the pioneering works by Crowley, then Winslow [2] and Godunov, Prokopov [3] that the grid generation as a separate discipline started from this 'simple trick' of interchanging variables in functionals. Not too much attention was paid to the validation of this procedure. The most striking example is related to 2-D harmonic co-ordinates $x(u)$ which are constructed as minimizers of the Dirichlet functional

$$\int_\Omega \frac{1}{2}\,\mathrm{tr}(\nabla x^{\mathrm{T}}\nabla x)\,\mathrm{d}u \tag{9}$$

subject to boundary conditions of the first kind. Since in practice $x(\Omega)$ is a convex domain, say a square, the useful solution is the inverse function $u(x)$ which provides the non-degenerate co-ordinates in Ω. Thus one can exchange variables in (9) resulting in 2-D in the following highly non-linear functional:

$$\int_{x(\Omega)} \frac{1}{2}\,\frac{\mathrm{tr}(\nabla u^{\mathrm{T}}\nabla u)}{\det\nabla u}\,\mathrm{d}x \tag{10}$$

The integrand of this functional is polyconvex, but the growth conditions are too weak for the Ball's existence and inverse function theory to be applied. In fact recent results about mappings with finite distortion [31] are that the mapping $u(x)\in\mathbb{W}^{1,2}(\Omega,\mathbb{R}^2)$ which provides a finite value of integral (10) is open and discrete—so it is close to mappings of bounded distortion. Anyway, it is not known to the author whether the existence theorem holds for (10) and when precisely the extremal mapping of (10) is the inverse harmonic mapping. It seems that recent results from Reference [22] can be of help here. Numerical evidences suggest that the finite element solutions of (9) and (10) can be completely different [8, 14].

Besides the case when the boundary conditions are specified on all of the boundary of domain it is important to know when the solutions of the variational problem with 'free' boundary conditions are globally invertible mappings. This very important situation was considered in References [16, 26]. The set of 3-D admissible deformations (5) is replaced by the following one:

$$\Phi_1 = \Big\{ u(x) \in \mathbb{W}^{1,p}(\Omega_0);\, \mathrm{adj}\, \nabla u \in \mathbb{L}^q(\Omega_0), \det \nabla u \in \mathbb{L}^r(\Omega_0),$$

$$u = u_0 \text{ on } \Gamma_0,\ \det \nabla u > 0 \text{ a.e. in } \Omega_0, \int_{\Omega_0} \det \nabla u \,\mathrm{d}x \leqslant \mathrm{vol}\, u(\Omega_0) \Big\} \tag{11}$$

where vol means the volume of the domain. The area of Γ_0 is supposed to be positive, but it is just an open subset of $\partial\Omega_0$. If the other assumptions of Ball's existence theory hold, then the minimizer of the variational problem in the modified set of admissible deformations Φ_1 also exists and is globally invertible almost everywhere. More precisely, it means that for every $y \in u(\bar{\Omega}_0)$ the number of preimages is equal to 1 almost everywhere.

6. VARIATIONAL PRINCIPLE FOR QUASI-ISOMETRIC MAPPINGS

The mapping $y(\eta) \in \mathbb{W}^{1,\infty}(\Omega_0, \mathbb{R}^n)$ is quasi-isometric if there exist constants $0 < t \leqslant 1$, $v > 0$, such that

$$\det \nabla_\eta y > t\phi_\theta(\det \nabla_\eta y, \nabla_\eta y) \quad \text{a.e. in } \Omega_0 \tag{12}$$

where

$$\phi_\theta(J, T) = \theta \left(\frac{1}{n} \mathrm{tr}(T^{\mathrm{T}} T) \right)^{n/2} + \frac{(1-\theta)}{2} \left(v + \frac{J^2}{v} \right), \quad 0 < \theta < 1$$

$v > 0$ is a constant which has a sense of the target average value of $\det \nabla_\eta y$.

The relation between (12) and (3) was discussed in Reference [11], where it was shown that if (12) holds, then the constant C_1 from (3) can be evaluated, namely

$$C_1 < \left(c_2 + \sqrt{c_2^2 - 1} \right)^{1/n} \left(c_1 + \sqrt{c_1^2 - 1} \right)^{(n-1)/n} \qquad c_1 = \frac{1 - (1-\theta)t}{\theta t}$$

$$c_2 = \frac{1 - \theta t}{(1-\theta)t} \qquad L = v^{1/n}$$

We see that the value $1/t$ has the meaning of the total metric distortion and any function $y(\eta) \in \mathbb{W}^{1,1}(\Omega_0, \mathbb{R}^n)$ satisfying (12) in fact belongs to $\mathbb{W}^{1,\infty}(\Omega_0, \mathbb{R}^n)$. Since in our case Ω_0 is strongly Lipschitz bounded domain then $y(\eta)$ is Lipschitz continuous [17].

Now the variational principle for construction of quasi-isometric mappings takes the following form [11]:

$$\int_{\Omega_0} f(\nabla_\eta y)\,\mathrm{d}\,\eta,$$

$$f = \begin{cases} (1-t)\dfrac{\phi_\theta(\det \nabla_\eta y, \nabla_\eta y)}{\det \nabla_\eta y - t\phi_\theta(\det \nabla_\eta y, \nabla_\eta y)} & \det \nabla_\eta y - t\phi_\theta(\det \nabla_\eta y, \nabla_\eta y) > 0 \\ +\infty & \text{otherwise} \end{cases} \tag{13}$$

This variational principle satisfies most of the requirements formulated in the previous section. It is well posed for arbitrary $n \geqslant 2$ [32] in the following sense. If the feasible set (12) is not empty then a minimizer exists and is a quasi-isometric mapping [32]. In Reference [32] the existence theorem was proved in the case when $y(\partial\Omega_0)$ is prescribed on all the boundary. When $y(\partial\Omega_0)$ coincides with the boundary values of the continuous homeomorphism the global invertibility of the minimizers of (13) follows directly from Ball's inverse function theorem.

The generalization of existence results to the case of 'free' boundary conditions, when $y(\Gamma_0)$ is given where Γ_0 is an open subset of $\partial\Omega_0$ seems to be straightforward, provided that the constraint

$$\int_{\Omega_0} \det \nabla_\eta y \,\mathrm{d}\eta \leqslant \mathrm{vol}\, y(\Omega_0)$$

is added to guarantee the global invertibility of the minimizers almost everywhere.

Due to similarity with elasticity problems it is natural to expect that global uniqueness of the solution and convergence of finite element approximations is provable for problems with very smooth boundary conditions and a solution close to the identity mapping when the implicit function theorem can be applied [16, 33, 34].

However numerical experiments suggest that convergence of the finite element approximations is observed in quite general cases so further theoretical analysis is necessary.

The existence theory for hyperelasticity described in the previous sections in principle does not exclude Lavrentiev phenomenon. The Lavrentiev phenomenon [35] generally means that we have different minimizers in different functional spaces. It is just known that the definition of the set of admissible deformations (5) excludes the cavitation, which is a particular case of Lavrentiev phenomenon.

Unlike hyperelasticity problems, no Lavrentiev phenomenon is possible for (13) since minimally regular solutions are still Lipschitz continuous. However if we set $t = 0$ in (13) then it can be shown that the cavitation can appear which is consistent with Ball analysis. Hence one cannot relax the requirement $t > 0$.

The unsolved theoretical problems for (13) include the stability conditions for the minimizers and the proof (if any) that the minimizer lies strictly inside the feasible set. To 'lie strictly inside the feasible set' means that for the minimizer of the variational problem (13) a stronger inequality

$$\det \nabla_\eta y - t_1 \phi_\theta(\det \nabla_\eta y, \nabla_\eta y) > 0 \tag{14}$$

is satisfied almost everywhere with $t_1 \geqslant t + \delta$, where δ is a strictly positive constant.

One of the results of Ball's existence theory for hyperelasticity is that when the minimizer belongs to $\mathbb{W}^{1,\infty}(\Omega_0, \mathbb{R}^n)$ and lies strictly inside the feasible set, meaning that the determinant of the Jacobi matrix is larger than some positive constant almost everywhere, then the weak variational formulation of Euler–Lagrange equations of the functional makes sense. However it is not known how to guarantee such properties of minimizers [36]. For (13) the regularity property is guaranteed, so it remains to prove (14) for the minimizer, if it is possible. In fact, if it is true then one also can expect higher regularity of the minimizer, in particular Hölder continuity of $\nabla_\eta y$.

7. CONTROL OF THE MAPPINGS PROPERTIES VIA COMPOSITION OF MAPPINGS

Up to this point we have considered the variational principle for construction of quasi-uniform mappings. However in practice it is necessary to construct mappings and grids with controlled variations of the local shape, size and orientation of the elements. We do not consider here orientation/alignment control which is a very hard problem. A more general polyconvex functional allowing for construction of quasi-isometric mappings with alignment control was suggested in Reference [37]. Here only shape and size control are considered. The general idea is to use composition of mappings in order to modify the properties of the solutions. It is illustrated in Figure 2.

The mapping $y(\eta)$ is represented as a composition of mappings $\xi(\eta)$, $x(\xi)$ and $y(x)$. It is assumed that ξ and x are co-ordinates in n-dimensional space, while more general case of y co-ordinates in m-dimensional space, $m \geqslant n$ is considered as well. In particular, $y(x)$ may define a parameterization of the surface. The mappings $y(x)$ and $\eta(\xi)$ are specified while the function $x(\xi)$ is the new unknown solution. We assume again that $y(x)$ and $\eta(\xi)$ are quasi-isometric mappings, but possibly with large constants C_1. Using the notations

$$H = \nabla_\xi \eta, \quad S = \nabla_\xi x, \quad Q = \nabla_x y, \quad T = \nabla_\eta y, \quad J = \det T$$

we get

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

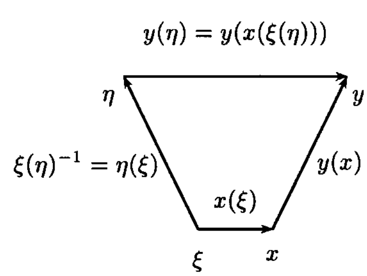

Figure 2. Composition of mappings. $y(x)$ and $\eta(\xi)$ are prescribed mappings, $^{-1}$ means inverse mapping.

and functional (13) is simply rewritten as

$$\int_{\Omega} f(Q\nabla_{\xi}xH^{-1})\det H \, \mathrm{d}\xi \tag{15}$$

where $\eta(\Omega)=\Omega_0$. Note that the function f depends only on the orthogonal invariants of the matrix $T^{\mathrm{T}}T$, and using equalities

$$T^{\mathrm{T}}T = H^{-\mathrm{T}}S^{\mathrm{T}}GSH^{-1}, \quad G(x)=Q^{\mathrm{T}}Q, \quad \tilde{H}(\xi)=H^{\mathrm{T}}H$$

and the fact that $\mathrm{tr}(AB)=\mathrm{tr}(BA)$, we get that f can be written via the orthogonal invariants of the matrix [11]

$$S^{\mathrm{T}}GS\tilde{H}^{-1}$$

This formulation is quite general since the matrices Q and H now can be non-square and metric tensors $\tilde{H}(\xi), G(x)$ are not assumed to be smooth. The only restriction is that their eigenvalues are positive and should have uniform lower and upper bounds. We will show later that even when input control data are just two metric tensors, the discrete approximation to the functional on each simplex element naturally employs a factorized representation (15).

The existence results in Reference [32] in fact were obtained in the presence of the matrix $\tilde{H}(\xi)\in\mathbb{L}^{\infty}$. The presence of $\tilde{H}(\xi)$ means that the integrand f in (15) is a function of $\xi, \nabla_{\xi}x$. When $G(x)$ is present as well, then f is the function of $\xi, x, \nabla_{\xi}x$. However when f is not a continuous function of x, the standard direct method for proving existence cannot be applied. But from a practical point of view it is natural to assume more or less the same regularity of the functions $\tilde{H}(\xi)$ and $G(x)$. Ball's inverse function theorem can be of help here. It is established in Reference [32] that any minimizer $y(\eta)$ of the variational problem (13) which coincides on the boundary of Ω_0 with continuous one-to-one mapping does satisfy the Ball inverse function theorem conditions. In fact it is possible to exchange dependent and independent variables in (13) and the resulting problem with respect to $\eta(y)$ will also be well posed and the minimizer will satisfy the inverse function theorem conditions.

Now introducing the metric $G(x)$ into this variational problem for the inverse mapping also results in a well-posed formulation, but the inverse function theorem now allows to establish the existence result also for the original problem (13) when $\tilde{H}(\xi)=I$ and $G(x)$ given, but not smooth. The details of this statement will be published elsewhere.

8. FUNDAMENTAL DISCRETIZATION PROBLEMS

Unlike other 'standard' problems, say related to finite element solution of linear elliptic problems, the discretization problems in the mappings theory are fundamental and mostly unsolved.

1. Surprisingly the problem to approximate locally invertible mapping by a converging sequence of locally invertible piecewise affine mappings is not solved even for quasi-isometric mappings, not mentioning general Sobolev mappings [36].
2. A more difficult problem: given a mapping with a metric distortion below a certain threshold $1/t$ (12). Is it possible to construct a sequence of converging piecewise affine

mappings, each being below the same threshold $1/t$? If it is possible, how to do it in practice?
3. The variational principle for a surface grid generation should be invariant with respect to surface parameterization, including the case of non-smooth parameterizations. How can this property be implemented on the discrete level? What is the counterpart of the finite element patch test condition [38] in the case of spatial mappings?

Our hypothesis is that the solutions to the above discretization problems in the 2-D case should be based on the theory of manifolds of bounded curvature.

9. THEORY OF MANIFOLDS WITH BOUNDED CURVATURE AND FOUNDATIONS OF ADAPTIVE GRID GENERATION

9.1. Adaptive grids and convergence of metric

In many numerical simulation methods based on adaptive grids the same 'standard' approach is applied and works reasonably well.

Given a discrete solution, one computes a certain metric, a new grid is constructed which is optimal in some sense with respect to this metric, the solution is recomputed on this grid so a new metric is in turn available [39]. This cycle is repeated until some convergence criteria are satisfied.

In fact the underlying assumption for the above procedure is that the solution of the continuous problem defines a certain continuous metric and the iterative solution technique creates a sequence of 'discrete' metrics. The convergence of the 'discrete' metric to the exact one implies convergence of the discrete solutions to the solution of the continuous problem.

A natural question then arises: what class of metrics should one consider and what is the meaning of convergence of the sequence of metrics?

In fact in the 2-D case these questions are correctly formulated and answered in the framework of the theory of manifolds of bounded curvature (MBC). The MBC theory was developed by A.D. Alexandrov and his school [40]. Some results of this theory which are most relevant to adaptive grid generation and geometric modelling are presented below based on the review [41]. We will use notations from [41] in the current section.

9.2. 2-D manifolds of bounded curvature (MBC)

A metric space (M,ρ) is called *a space with intrinsic metric* if it is linearly connected (there exists a path connecting any two points) and for any two of its points X,Y the quantity $\rho(X,Y)$ is equal to the lowest bound of the lengths of the arcs connecting these points. An example of the intrinsic metric is the sphere where the distance between points is the length of the shortest arc connecting them. On the other hand, the Cartesian distance between the points on the sphere is not an intrinsic metric [41]. Any Hilbert space and any normed vector space are spaces with intrinsic metric.

Let us consider 2-D manifold with intrinsic metric. The definition of curvature is based on the notion of triangle excess. Consider a triangle T where the three vertices are connected by shortest curves. The excess $\delta(T)$ is defined as

$$\delta(T)=\alpha+\beta+\gamma-\pi$$

where α, β, γ are the angles between the sides at the vertices. In fact the formal axiomatics is based on the concept of upper angles which are always defined [41]. For the Riemannian manifolds, $\delta(T)$ coincides with the integral Gauss curvature of the geodesic triangle T which follows from the Gauss–Bonnet formula.

A manifold has a bounded curvature when every point X has a neighbourhood U such that the sum of excesses of pairwise non-overlapping triangles lying inside U is uniformly bounded from above

$$\sum \delta(T) < N$$

where N does not depend on the choice of triangles and depends only on U.

The curvature is defined as follows. The positive and negative parts of curvature of the open set G are defined as the exact upper and lower bounds of the sums of excesses of pairwise non-overlapping triangles lying in G:

$$\omega^+(G) = \sup \sum \delta(T), \quad \omega^-(G) = \inf \sum \delta(T)$$

The curvature itself is defined as

$$\omega(G) = \omega^+(G) + \omega^-(G)$$

while the absolute curvature is defined by

$$|\omega|(G) = \omega^+(G) - \omega^-(G)$$

The curvature at a point is understood as the integral characteristics of the point set.

9.3. *Definition of curvature and peak points via isothermal co-ordinates*

Another definition of curvature is based on the analytic representation of 2-D MBC by means of a line element. Namely, any point X on MBC has a neighbourhood U which is isometric to some flat domain G with a metric defined by a line element

$$\mathrm{d}s^2 = \lambda(x, y)(\mathrm{d}x^2 + \mathrm{d}y^2)$$

where logarithm of λ is the difference of two subharmonic functions [41]. Here x, y are Cartesian co-ordinates on the plane. $\lambda(x, y)$ may vanish or have points of discontinuity, which is not possible for Riemannian manifolds.

Let us denote by φ the one-to-one mapping which maps U onto G and let $z = \varphi(X)$ be the point on the plane. In the domain $G = \varphi(U)$, a set function $\omega(z)$ is defined which does not depend on the choice of the particular isothermal co-ordinate system.

Let $\Gamma(X, r)$ be a circle, i.e. the union of points whose distance from X is equal to r. Then the following important statement holds (Lemma 8.1.1 from Reference [41]):

Let $X \in U$, $z = \varphi(X)$. Then there is a number $\delta_1 > 0$ such that if $0 < r < \delta_1$, then $\Gamma(X, r)$ is a simple closed curve. Let $\sigma(X, r)$ be the length of the curve $\Gamma(X, r)$. Then

$$\frac{\sigma(X, r)}{r} \to \theta(X) = 2\pi - \omega(z) \quad \text{as } r \to 0$$

The quantity $\omega(X) = \omega(z)$ is called the curvature of the manifold at the point X. $\theta(X) = 2\pi - \omega(X)$ is the total angle at the point X.

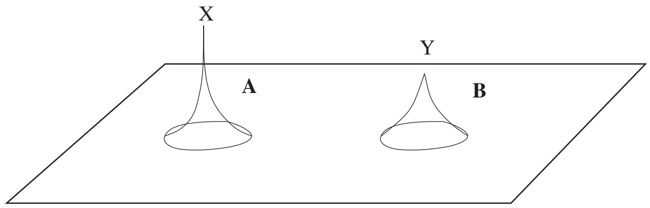

Figure 3. A—a peak point, curvature at the vertex X is equal to 2π, B—the curvature at the vertex Y is less than 2π and the local tangential cone at the vertex is well defined.

When

$$\omega(X) = 2\pi$$

then X is called *a peak point*, which is illustrated in Figure 3.

Another important notion is the so-called *tangent cone* condition.

The 2-D manifold of bounded curvature is called a cone when it has a non-zero curvature only at one point—the vertex, and has zero curvature everywhere. Let us denote by $Q(\theta, h)$ the set of all points of the cone at a distance less than h from its vertex. Here θ denotes the total angle at the vertex of the cone. In fact the cone can be defined directly, not using the concept of manifolds of bounded curvature.

Let M be a 2-D manifold with intrinsic metric and A be an arbitrary point of M. M has a tangent cone at A if there is a cone $Q(\theta, r)$ that admits a one-to-one map φ onto a neighbourhood of A such that A corresponds to the vertex O of the given cone and for any $X, Y \in Q(\theta, r)$ such that $X \neq Y$

$$\frac{\rho_Q(X, Y)}{\rho_M(\varphi(X), \varphi(Y))} \to 1 \quad \textit{as } X \to O, \ Y \to O$$

The 2-D MBC has a tangent cone at any point X which is not a peak point.

9.4. Chebyshev parameterization of MBC

Besides isothermal co-ordinates, the manifolds with bounded curvature admit other types of parameterizations. In particular, it is possible to introduce quasi-isometric co-ordinates.

In order to present these results, one has to introduce the *cut* and *paste* operations over MBC. In particular one can cut from a 2-D MBC another manifold, which will also have bounded curvature, provided that the cut itself is not pathological. On the other hand several MBC's can be pasted or glued together along the boundaries, provided that the length of the glued fragments of the boundaries is the same. The formal definition of cut and paste operations is quite complicated and is omitted here. The only definition we need is that of a polygon:

A polygon is a domain whose boundary consist of finitely many shortest curves and each connected component of the boundary is a homeomorphic image of either a circle or an open interval.

Let $\mathbb{R}^2_+$ be the positive quadrant $\{(x, y) \in \mathbb{R}^2 \,|\, x > 0, \ y > 0\}$ on the plane. Let us assume that from $\mathbb{R}^2_+$ a polygon P_0 homeomorphic to a disc is cut out and then a polygon P, homeomorphic

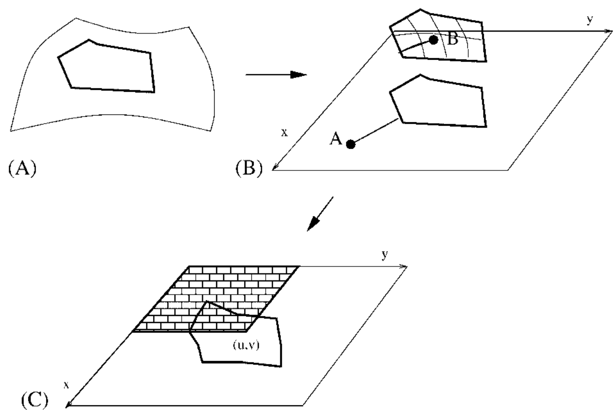

Figure 4. A—A polygon is cut from the initial manifold. B—A polygon with the same length of the sides is cut from the positive quadrant on the plane. The polygon from the manifold is pasted into the hole. C—The manifold M from B is one-to-one mapped onto the positive quadrant creating Chebyshev co-ordinates in M, D_{uv} is shown as a bricked area.

to a disc is pasted in its place, which is cut out from a 2-D MBC. The manifold $M=(\mathbb{R}^2_+/P_0)\cup P$ formed under this pasting is in turn a 2-D MBC.

Then the following theorem holds:

Theorem (*Bakelman* [*42, 43*])
Let M be the manifold defined as above and $|\omega|(M)<\pi/2$. Then there is a one-to-one map φ of M onto $\mathbb{R}^2_+$ such that if one introduces in $\mathbb{R}^2_+$ the line element

$$\mathrm{d}s^2=\mathrm{d}u^2+2\cos\tau(u,v)\,\mathrm{d}u\,\mathrm{d}v+\mathrm{d}v^2 \tag{16}$$

where $\tau(u,v)=\pi/2-\omega(\varphi^{-1}(D_{uv}))$ and D_{uv} is the rectangle $\{(x,y)\in\mathbb{R}^2_+\mid 0<x\leqslant u,\ 0<y\leqslant v\}$, then φ is an isometric map of M onto the square $\mathbb{R}^2_+$ endowed with the metric generated by the line element (16).

From this theorem it follows that for each point X of a 2-D MBC with small enough absolute curvature one can find a circle in which the metric of the manifold is defined by (16). This co-ordinate system is called a *Chebyshev co-ordinate system.*

The theorem is illustrated in Figure 4.

In other words this theorem states that when the absolute curvature of a manifold is small enough then it admits a global quasi-isometric parameterization via flattening.

The above theorem has a converse, which essentially says that the metric on the plane defined by the line element (16) with $\tau(u,v)=\pi/2-\omega(D_{uv})$ defines a manifold of bounded curvature. It is only assumed that the total variation of ω is less than $\pi/2$.

9.5. Quasi-isometric parameterization of MBC

It is interesting that the practical flattening methods for surfaces with complicated boundaries are based essentially on the same cut-and-paste procedure. In Reference [37] quasi-

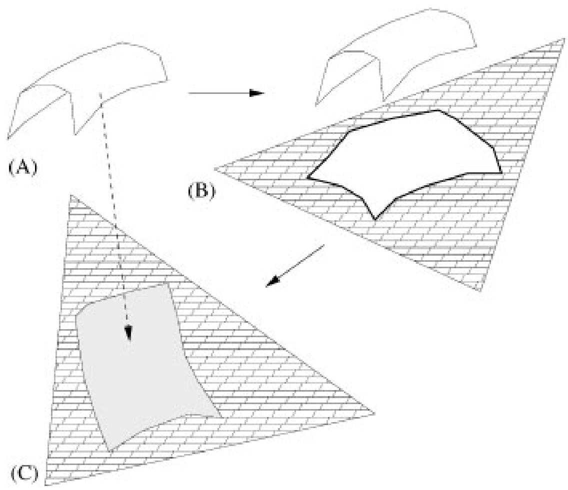

Figure 5. A—The initial surface with its boundary. B—A polygon with the same length of the sides is cut from a triangle on the plane. The surface is pasted into the hole. C—The resulting manifold B is one-to-one mapped onto another triangle creating quasi-isometric co-ordinates on the initial surface.

isometric flattening procedure was suggested, based on cut and paste operations illustrated in Figure 5.

The details of this algorithm are explained in Reference [37] and the motivation to choose a triangle as a domain to paste manifold into was based on graph theory arguments.

The relations between the curvature of manifold and the distortion of flattening are very important for practice. Finding optimal distortion bounds, or quasi-isometry constants, or in other words bi-Lipschitz constants is currently a hot topic of research. The results from Reference [43] are not optimal for quasi-isometric flattening since the Chebyshev parameterization is very restrictive. However the proof presented in Reference [43] can be easily generalized for the case of quasi-isometric flattening leading to much better estimates.

Recently a new promising generalization of the above theorem was obtained in Reference [44], where the following result was proved.

Theorem (*Bonk and Lang* [*44*])
Suppose that Z is a complete MBC homeomorphic to a plane. Let ρ denote the metric of Z. If $\omega^+(Z) < 2\pi$ and $\omega^-(Z) < \infty$, then there exists a one-to-one mapping $\phi : Z \to \mathbb{R}^2$ satisfying

$$\frac{1}{L}|\phi(X) - \phi(Y)| \leqslant |\rho(X, Y)| \leqslant L|\phi(X) - \phi(Y)|$$

with

$$L = \left(\frac{2\pi + \omega^-(Z)}{2\pi - \omega^+(Z)}\right)^{1/2}$$

In the notations of Reference [44] Z is L-bi-Lipschitz equivalent to $\mathbb{R}^2$. The inequality is sharp when Z is a cone.

This theorem seems to provide the best known theoretical distortion estimate for quasi-isometric parameterization. It can be applied to manifolds with boundary using the cut and paste operations.

Let us formulate the following conjecture:

Conjecture
Suppose Z is a complete MBC homeomorphic to a plane. Let G denote a domain in Z homeomorphic to a closed disc such that the boundary of G is a rectifiable simple closed arc. If $\omega^+ < 2\pi$ for each point of Z and $\omega^-(Z) < \infty$, then Z is L-bi-Lipschitz equivalent to $\mathbb{R}^2$ with

$$L = 0((q(2\pi + \omega^{-1}(Z)))^{1/2})$$

where

$$q = \sup_G \frac{2\sigma(G)}{p^2}$$

and $\sigma(G)$, p denote the area and the perimeter of G, respectively.

The idea of this conjecture is suggested by the inequality (see Theorem 8.5.2 in Reference [27])

$$\sigma(G) \leqslant \frac{p^2}{2(2\pi - \omega_0)}, \quad \omega_0 = \omega^+(G) < 2\pi$$

which is attained for a cone. From this inequality it follows that $(q(2\pi + \omega^-(Z)))^{1/2} \leqslant L$. The intuitive meaning in the case of surfaces is that q can be considered as a measure of the 'pockets' and the distortion of flattening should be a function of these 'depths of pockets'. It is very important to get rid of the very restrictive upper bound on the positive part of the curvature. The constraint $\omega^+(Z) < 2\pi$ is overly restrictive since in practice a quasi-isometric flattening with relatively small distortion can be constructed for surfaces with a large absolute curvature. A simple example is the plane with a regular net of small bumps which has an unbounded ω^+ but can be flattened with a small distortion.

9.6. Polyhedral metric

The key concept in the theory of MBC is the manifold with a polyhedral metric.

It is said that the manifold has a polyhedral metric if

(A) for any internal point X of the manifold one can find a neighbourhood that admits an isometric map onto some circular cone;
(B) for any boundary point X of the manifold one can find a neighbourhood that admits an isometric map onto some circular sector.

There is an important theorem related to the triangulation of a manifold with a polyhedral metric.

Theorem (*Reshetnyak* [*41*])
Let M be a 2-D manifold with boundary. We assume that M is connected and is endowed with a polyhedral metric. Then M admits a triangulation such that any triangle is isometric to a triangle on the plane.

It is important to see the difference between a polyhedron and a manifold with a polyhedral metric. In particular, a finite circular cone does define a manifold with a polyhedral metric but it is not a polyhedron.

9.7. Metric convergence

One of the key facts of the MBC theory is the possibility to approximate any 2-D MBC by a manifold with a polyhedral metric. In order to present these results one has to introduce the definitions of convergence.

Let (M,ρ) and (M_k,ρ_k), $k=1,2,\dots$ be the metric spaces with intrinsic metric. The spaces (M_k,ρ_k) converge to the space (M,ρ) if for each k a homeomorphism φ_k of M onto M_k is specified, such that $\rho_k(\varphi_k(X),\varphi_k(Y))\to\rho(X,Y)$ *uniformly* as $k\to\infty$. Denoting by $\tilde{\rho}_k(X,Y)=\rho_k(\varphi_k(X),\varphi_k(Y))$, one obtains a sequence of metrics $\tilde{\rho}_k$ on M.

Another important concept is the *local convergence* of metric spaces. Let in the manifold M a sequence of domains M_k, $k=1,2,\dots$ be specified, such that for each k the closure of M_k is compact, it is contained in M_{k+1} and

$$\bigcup_{k=1}^{\infty} M_k = M$$

The sequence of metrics ρ_k, $k=1,2,\dots$ converges locally to the metric ρ if for any k the sequence of metrics ρ_m, $m=k,k+1,\dots$ converges to ρ on the set M_k, that is $\rho_m(X,Y)\to\rho(X,Y)$ as $m\to\infty$ uniformly when $X,Y\in M_k$ irrespective of the value of k.

The local convergence is useful when, for example, the boundary of the manifold should be approximated by the boundary of a polygon.

One of the main convergence results is that any manifold of bounded curvature is the limit of manifolds with a polyhedral metric provided that the absolute curvature of a polyhedral metrics is bounded. The metric convergence is uniform. Moreover, the positive and negative parts of the curvature of the polyhedral metrics converge to the positive and negative parts of the curvature of the MBC in a weak sense [40].

9.8. Tangent cone condition and approximation of MBC by manifolds with polyhedral metrics

Now we are in a position to formulate one of the mains results on approximations of MBC.

Let us use the following definition of the *proportional convergence*:

Assume that on the set M metrics ρ, ρ_k, $k=1,2,\dots$ are specified. The metrics ρ_k converge proportionally to the metric ρ if for any $\varepsilon>0$ there exists a number k_0 such that for any $k\geqslant k_0$

$$\frac{1}{1+\varepsilon}\,\rho(X,Y)\leqslant\rho_k(X,Y)\leqslant(1+\varepsilon)\rho(X,Y)$$

for any $X,Y\in M$.

The theorem which can be applied in the case of quasi-isometric parameterizations is formulated as follows.

Theorem (*Reshetnyak* [*41*])
Let M be a compact 2-D manifold with intrinsic metric ρ. We assume that M has a tangent cone at each point of it. Then there is a sequence of polyhedral metrics ρ_k, $k=1,2,\ldots$ defined in M that converges proportionally to the metric of the manifold.

In fact a 2-D MBC has a tangent cone at any point of it which is not a peak point [45].

When peak points are present, MBC can still be a limit of a uniformly convergent sequence of manifolds with polyhedral metrics.

10. PROPORTIONAL CONVERGENCE AND DISCRETIZATION PROBLEMS

The metric convergence theory does not provide ready solutions to the questions posed in Section 8, it rather suggests that these problems should be somewhat reformulated.

Consider for example problem 8.1. Let $y(\eta):\Omega_0\to\mathbb{R}^2$ be the locally invertible quasi-isometric spatial mapping. If the domain Ω_0 is regular enough then $y(\eta)$ is Lipschitz continuous. Using as an intrinsic metric the lowest bound of the length of the curves lying on the graph of the function $y(\eta)$ and connecting any two points, we obtain a 2-D MBC. If $y(\eta)$ is regular enough then there exists a sequence of manifolds with polyhedral metric which converge proportionally to this manifold. Each manifold with a polyhedral metric can be triangulated. From a first glance it looks like the problem is solved. Unfortunately it is not true. First of all the curvature of a Lipschitz mapping can be unbounded, so it is not necessarily MBC. Another problem is that the existence of a manifold with a polyhedral metric does not necessarily imply the existence of a polyhedron where this metric is realized. In our case existence of a 'good' polyhedral metric does not mean that a piecewise-affine mapping exists, which realizes exactly this metric. A manifold which is intrinsically flat in general is a ruled hypersurface from the point of view of extrinsic geometry.

Only manifolds with polyhedral metric homeomorphic to a sphere with a non-negative curvature at every vertex can always be realized as a convex polyhedron. This statement is based on the famous folding theorem by Alexandrov [46].

It is possible to formulate the following hypotheses about the relations between variational problems for spatial mappings and theory of manifolds:

1. It is possible to define a functional (13) using the concept of manifolds with polyhedral metric.
2. The extrema of (13) should be sought in the class of bounded curvature mappings.
3. Proportional convergence of polyhedral metric to metric satisfying the tangent cone condition implies convergence of the value of the functional (13).
4. Proportional convergence of a polyhedral metric to a metric satisfying the tangent cone condition implies convergence of the gradient of discrete functional to the gradient of the exact functional in a weak sense, i.e. the weak variational formulation for Euler–Lagrange equations is well defined.

Hypothesis 4 is based on the results about weak convergence of positive and negative parts of a curvature.

It seems that in many particular cases it is possible to construct polyhedrons with metric converging to the given polyhedral metric. However, a serious question is whether this step is absolutely necessary. It seems that manifolds with polyhedral metric are good numerical analysis objects by themselves and numerical simulation can be done using them. This thesis is illustrated below using an optimal surface parameterization as a test case. It is shown also that one does not need the polyhedron itself in order to compute the value of the functional and to run the minimization solver. The polyhedral metric is just enough!

11. GEODESIC TRIANGLES AND PATCH TEST FOR GRID FUNCTIONALS

Let us return to the notations introduced in Section 7. and consider the problem of finding optimal parameterizations of surfaces. Thus let $\eta(\xi):\mathbb{R}^2\to\mathbb{R}^2$ and $y(x):\mathbb{R}^2\to\mathbb{R}^3$ be given. In particular, let $y(x)$ define an existing surface parameterization, while $\eta(\xi)$ be a specified mapping.

We require that the discrete approximation to the functional of interest should satisfy a simple compatibility condition. Let us consider a mapping between two developable surfaces, or more precisely when $y(x)$ and $\eta(\xi)$ are both isometric.

Then with proper boundary conditions functional (15) should attain its absolute minimum on the isometric mapping $x(\xi)$. We require that the discrete functional attains its absolute minimum on the same isometric mapping.

Since we consider here the intrinsic geometry problem, the particular shape of the surface is not relevant.

Since developable surfaces are intrinsically flat, they are locally indistinguishable which is illustrated in Figure 6.

In order to build a discrete approximation to (15) fully employing all introduced controls, we assume that a certain triangulation is available in ξ-co-ordinates and construct a composition of mappings, similar to those in (15).

Our unknowns are images of the nodes of this triangulation in x-co-ordinates. It is already assumed that each triangle mapping $y(\eta)$ is affine. However the mappings $y(x)$ and $\eta(\xi)$ are not affine ones and should be somehow approximated on each triangle. The idea of the approximation is very simple and is attributed to Alexandrov as well (Figures 7 and 8).

Let us take three points A,B,C in x-co-ordinates and compute their images under the mapping $y(x)$. Connecting these three points via the shortest arcs we obtain a geodesic triangle on the surface $y(x)$. And finally replacing this triangle by a flat one with edge lengths coinciding

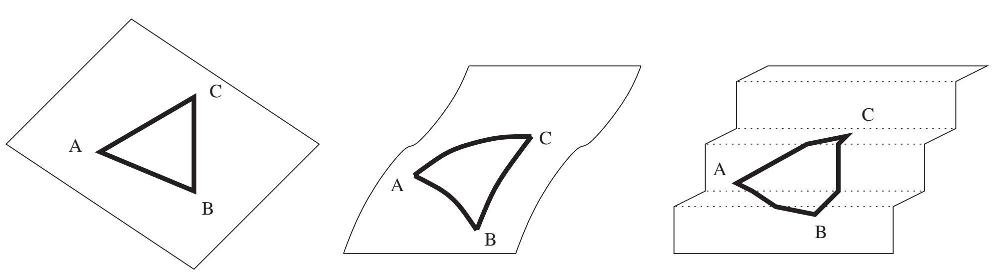

Figure 6. Indistinguishable geodesic triangles on developable surfaces.

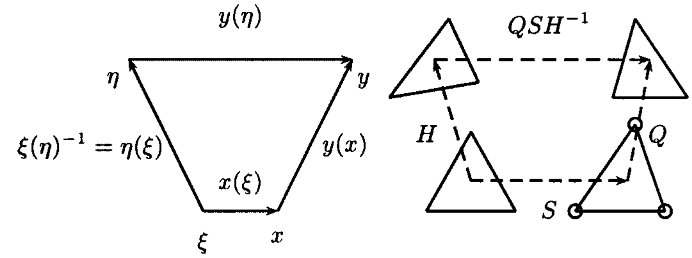

Figure 7. Composition of mappings and their approximation by composition of affine mappings. The matrix QSH^{-1} is supposed to approximate the gradient $\nabla_\eta y$. $\circ$ symbols mark the unknowns of the variational problem.

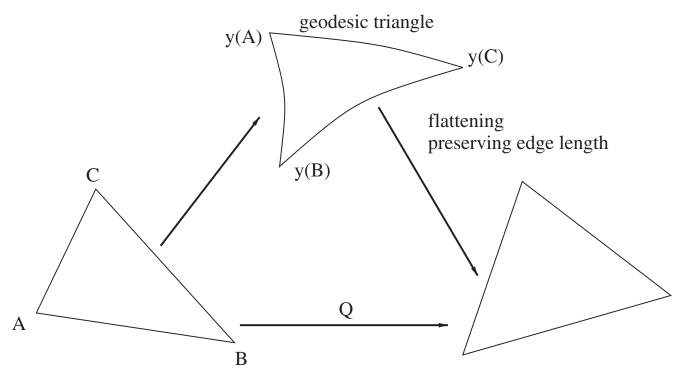

Figure 8. Construction of an affine approximation to $\nabla_x y(x)$.

with those of the geodesic triangle, we obtain a local affine approximation to $y(x)$ and natural formula for the matrix Q which can be easily derived and is omitted here. We assume that our triangles are homeomorphic to a disc. This operation is always well defined since the distance along geodesics satisfies the triangle rule, provided that one vertex is not lying on the shortest path connecting the other two vertices. Note that the spatial orientation of the last flat triangle is irrelevant since the functional depends only on the matrix $Q^{\mathrm{T}}Q$.

The discrete functional is just the sum of elementary terms over all triangles, where each term is just $f(QSH^{-1})\det H$ multiplied by the area of the triangles in ξ co-ordinates, which are constants.

Suppose that we have a certain developable surface in 3-D space y and two different flattenings of this surface. The first one is isometric (see Figure 9 (left)) and the second one is distorted and is quasi-isometric (see Figure 9 (right)).

Since in the first case $G=I$, the discrete functional will attain its absolute minimum provided that the target shapes for each plane triangle are correctly chosen. It can be easily shown that in the second case the solution via composition of mappings can be reduced to the first one and the functional also attains its absolute minimum, so independently of the surface parameterization, the optimal geodesic grid on the surface is the same (of course provided that all geodesics are unique, which is not always the case).

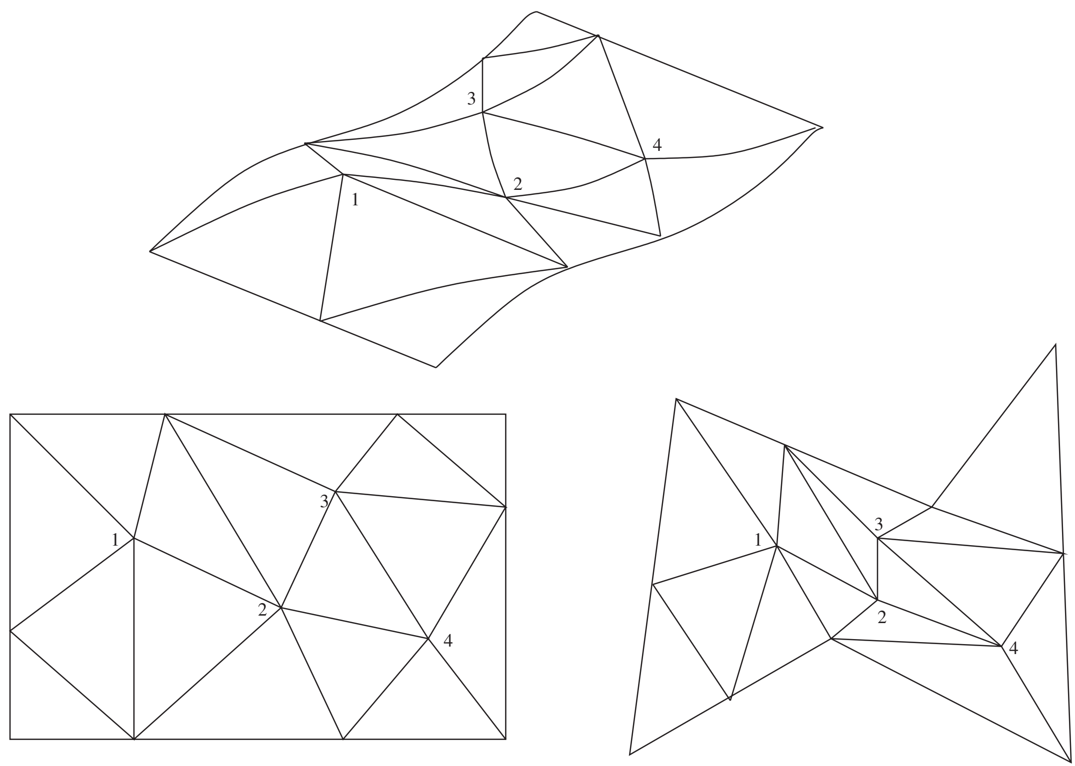

Figure 9. Patch test: different flattenings and parameterizations of the developable surface and the same optimal mesh on the surface.

One has a good reason to claim that the discrete functional has a manifold with a polyhedral metric as an argument.

12. GLOBAL PARAMETERIZATION OF SURFACES VIA UNFOLDING AND FLATTENING

In the example shown in Figure 9, the surface parameterization is constructed via a flattening operation. This is very powerful tool in geometric modelling [47]. In particular it can provide a global parameterization for surfaces defined by multiple patches (in real life up to 4–5 thousands). Any triangulated surface homeomorphic to a disk with holes can be flattened [47]. A more general statement can be formulated as a hypothesis: every manifold of bounded curvature without peak points, homeomorphic to a disk with holes, admits quasi-isometric flattening. Of course this statement is inexact because it does not take into account the ratio of perimeter squared/area of the surface which greatly influences the distortion bounds of flattening.

More precise hypothesis can be formulated as follows: every compact manifold of bounded curvature without peak points admits a finite quasi-isometric chart. Above statement means

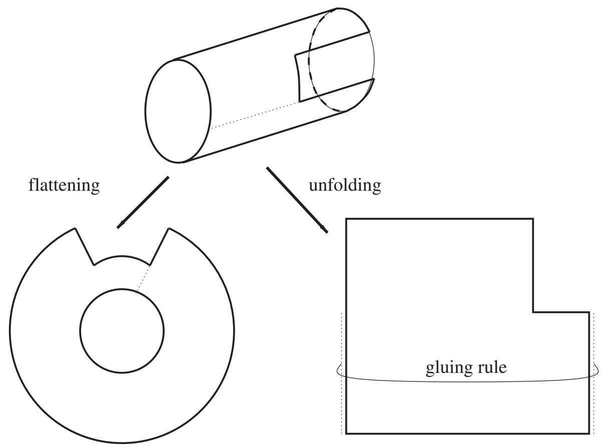

Figure 10. Flattening vs unfolding for surface parameterization.

that in this manifold one can choose a finite number of domains, homeomorphic to disks, which admit quasi-isometric flattening thus creating local quasi-isometric co-ordinates in each domain. The transformation rules for the local co-ordinates in the overlapping domains are also quasi-isometric. The union of these domains coincides with the manifold itself. In fact a similar statement was already proved in Reference [43] under more restrictive conditions. Obviously the theorem [44] allows to relax these conditions.

The quasi-isometric parameterization can be constructed via cut and paste operation.

It seems that the optimal strategy is to create a finite number of cuts on the manifold and perform a quasi-isometric flattening which is either global combined with unfolding, or can be applied for polygons cut from a manifold.

As a rule of thumb the cuts should decrease the depth of pockets in manifold. Finding optimal cuts is a very hard and unsolved problem.

In practice, we have two basic operations: flattening of a surface and improvement of plane grids—which means optimization of local co-ordinate systems (Figure 10).

Both the flattening operation and the variational improvement of plane grids can be done using the composition of mapping framework. In the flattening operation the target shape for each plane triangle is given via the matrix H which in turn is computed via the flattening of the geodesic triangle of the surface. In a plane mesh improvement one already has a parameterization of the surface $y(x)$ provided by the flattening procedure, so it is necessary to recompute the matrix Q for each plane triangle during the optimization procedure. If additional adaptation is necessary, say adaptation to curvature of the surface, then we just have a different definition of metrics $G(x)$, but the same rules for construction of the matrix Q, since the length of the geodesics is defined only by $G(x)$.

13. NUMERICAL EXAMPLE

The practical minimization method for the suggested functional is based on the idea of frozen metrics G. Obviously the matrix H for each triangle is computed only once and 'accompanies'

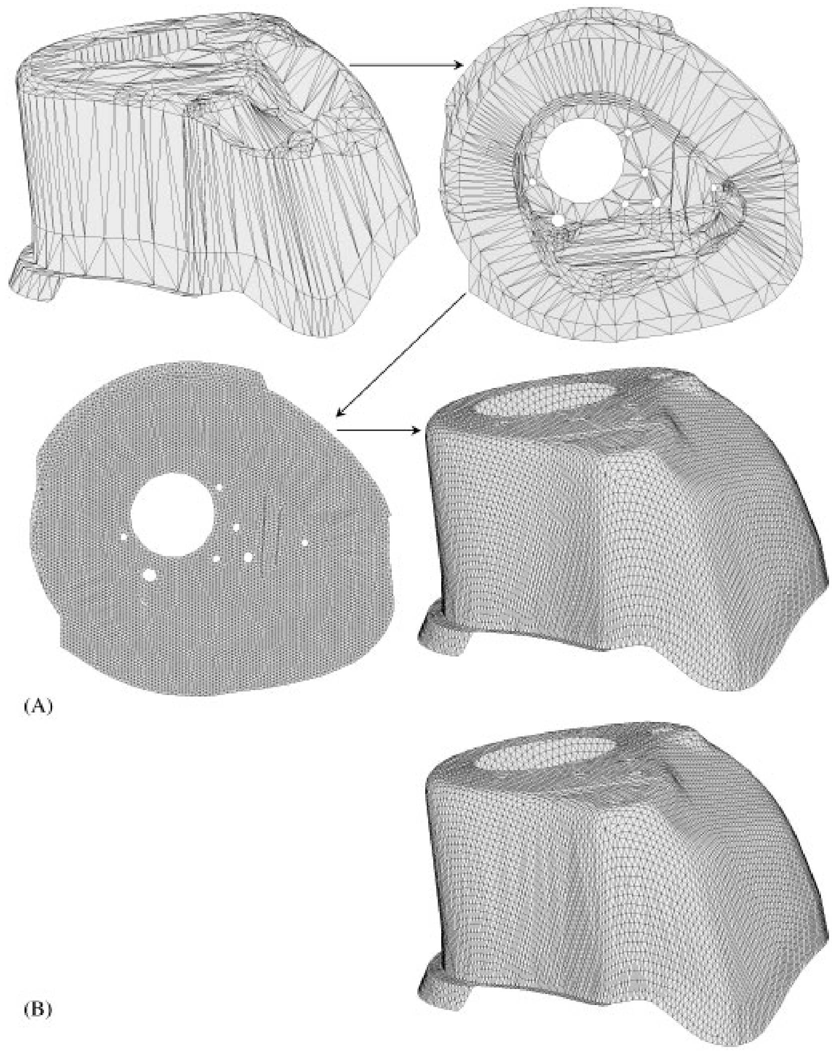

Figure 11. Surface meshing via quasi-isometric flattening, plane grid generation and variational improvement. A—a mesh constructed using a discrete functional satisfying the patch test; B—the patch test is violated.

each triangle during the minimization process. We can interpret it as a shape/size definition in 'Lagrangian' co-ordinates. On the other hand the matrix Q changes each time when the triangle vertices are updated, it behaves more like a shape/size definition in 'Eulerian' co-ordinates. So the idea of the iterative solution is to solve a partial minimization problem for the functional when for each triangle $f = f(Q^{k-1}S^kH^{-1})$, k being the iteration number. Generally this partial minimization problem involves just 1–2 iterations of the preconditioned gradient method [11]. This approach was found very stable and converged quite fast.

In practice, one can have a surface description given by a tesselation which approximates the geometry with a prescribed chordal error. In this case one starts from the manifold with a polyhedral metric which is very bad and should be approximated by a polyhedral metric which is good—namely it provides surface triangles which are almost equilateral. Then the basic problem is to compute a quasi-isometric flattening of an extremely ill-conditioned triangulation [37].

An example of a meshing procedure based on global flattening is shown in Figure 11. Note the difference between the meshes in subfigures A and B. In both cases the mesh quality is quite good, but in case B the discrete functional violates the patch test and problems with mesh convergence can be expected.

Contrary to what was said in previous sections, the above figure shows the polyhedron as a solution of the variational problem. In fact it is not true. The solution of the variational problem is some plane triangular grid where a constant metric tensor is defined on each triangle. Thus the solution is the manifold with a polyhedral metric. However the meshing procedure takes into account the sharp edges in the surface and puts the triangle edges onto these sharp ‘feature’ lines. So when plane mesh vertices are mapped onto the surface and connected by the straight edges, the length of these edges is very close to the length of geodesics. But the polyhedral metric on the final surface by itself is only an approximation of the polyhedral metric computed via the optimization procedure. In our case these two metrics converge to each other in a very nice way.

14. CONCLUSIONS

The main conclusion is that the validations of variational principles for grid generation should use mathematical methods developed in finite hyperelasticity and in the theory of manifolds of bounded curvature, in particular those related to proportional approximations by manifolds with polyhedral metrics. It is well known that the research to construct generalizations of theory of manifolds of bounded curvature to 3-D case is under way. These results should be used for validations of spatial variational principles.

REFERENCES

1. Godunov SK, Prokopov GP. Calculation of conformal mappings and grid generation. *Zhurnal Vychislitel’noy Matematiki i Matematicheskoy Fyziki* 1967; **7**(6):902–914.
2. Winslow AM. Numerical solution of the quasilinear Poisson equation in a non-uniform triangle mesh. *Journal of Computational Physics* 1966; **1**:149–172.
3. Godunov SK, Prokopov GP. The use of moving meshes in gas-dynamic calculations. *USSR Computational Mathematics Mathematical Physics* 1972; **12**(2):182–191.
4. Winslow AM. Adaptive mesh zoning by the equipotential method. Lawrence Livermore Laboratory, *UCID-19062*, 1981.
5. Brackbill JU, Saltzman JS. Adaptive zoning for singular problems in two dimensions. *Journal of Computational Physics* 1982; **46**(3):342–368.
6. 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.
7. Liseikin VD. Construction of regular grids on *n*-dimensional manifolds. *Zhurnal Vychislitel’noy Matematiki i Matematicheskoy Fyziki* 1991; **31**(11):1670–1683.
8. Ivanenko SA, Charakch‘yan AA. Curvilinear grids consisting of convex quadrilaterals. *Zhurnal Vychislitel’noy Matematiki i Matematicheskoy Fyziki* 1988; **28**(10):1498–1506.
9. Hormann K, Greiner G. MIPS: an efficient global parameterization method. In *Curve and Surface Design.* Laurent P-J, Sablonniere P, Schumacher LL (eds). Vanderbilt University Press: Nashville, 2000; 163–172.

10. Liseikin VD. *Grid Generation Methods*. Springer: Berlin, 1999.
11. Garanzha VA. Barrier variational generation of quasi-isometric grids. *Computational Mathematics and Mathematical Physics* 2000; **40**(11):1617–1637.
12. Huang W. Variational mesh adaptation: isotropy and equidistribution. *Journal of Computational Physics* 2001; **174**:903–924.
13. Thompson JF, Warsi ZUA, Mastin CW. *Numerical Grid Generation. Foundations and Applications*. North-Holland: Amsterdam, 1985.
14. Garanzha VA, Kaporin IE. Regularization of the barrier variational grid generation method. *Computational Mathematics and Mathematical Physics* 1999; **39**(9):1426–1440.
15. Godunov SK, Gordienko VM, Chumakov GA. Quasi-isometric parametrization of a curvilinear quadrangle and a metric of constant curvature. *Siberian Advances in Mathematics* 1995; **5**(2):1–20.
16. Ciarlet PG. *Mathematical Elasticity*, *V.1*: *Three Dimensional Elasticity. Studies in Mathematics and Applications*, vol. 20. Elsevier: New York, 1988.
17. Morrey CB. *Multiple Integrals in the Calculus of Variations.* Springer: Berlin, 1966.
18. Eells JE, Lemair L. Another report on harmonic maps. *Bulletin of the London Mathematical Society* 1988, **20**(86):387–524.
19. Rado T. Aufgabe 41. *Jahresbericht der Deutschen Mathematiker Vereinigung* 1926; **35**(49):49.
20. Kneser H. Losung der Aufgabe 41. *Jahresbericht der Deutschen Mathematiker Vereinigung* 1926; **35**:123–124.
21. Choquet G. Sur un type de transformation analytique generalizant le representation conforme et definie au moyen de fonctions harmoniques. *Bulletin des Sciences Mathematiques* 1945; **69**:156–165.
22. Alessandrini G, Nesi V. Univalent σ-harmonic mappings. *Archives for Rational Mechanics and Analysis* 2001; **158**:155–171.
23. de Almeida VF. Domain deformation mapping: application to variational mesh generation. *SIAM Journal on Scientific Computing* 1999; **20**(4):1252–1275.
24. Gangbo W, van der Putten R. Uniqueness of equilibrium configurations in solid crystals. *SIAM Journal of Mathematical Analysis* 2000; **32**(3):465–492.
25. Ball JM. Convexity conditions and existence theorems in non-linear elasticity. *Archives for Rational Mechanics and Analysis* 1977; **63**:337–403.
26. Ciarlet PG, Necas J. Unilateral problems in non-linear, three-dimensional elasticity. *Archives for Rational Mechanics and Analysis* 1985; **87**:319–338.
27. Reshetnyak YG. Mappings with bounded deformation as extremals of Dirichlet type integrals. *Siberian Mathematical Journal* 1968; **9**:487–498.
28. Ball JM. Global invertibility of Sobolev functions and the interpenetration of matter. *Proceedings of Royal Society of Edinburgh* 1981; **88A**:315–328.
29. Pourciau BH. Global properties of proper Lipschitzian maps. *SIAM Journal of Mathematical Analysis* 1983; **14**(4):796–799.
30. Pourciau BH. Homeomorphisms and generalized derivatives. *Journal of Mathematical Analysis with Applications* 1983; **93**(2):338–343.
31. Iwaniec T, Sverak V. On mappings with integrable dilatation. *Proceedings of American Mathematical Society* 1993; **118**:181–188.
32. Garanzha VA, Zamarashkin NL. Spatial quasi-isometric mappings as minimizers of a polyconvex functional. *Computational Mathematics and Mathematical Physics* 2003; **43**(6):815–826.
33. Zhang K. Energy minimizers in nonlinear elastostatics and the implicit function theorem. *Archives for Rational Mechanics and Analysis* 1991; **114**:95–117.
34. Carstensen C, Dolzmann G. An *a priori* error estimate for finite element discretizations in nonlinear elasticity for polyconvex materials under small loads. *Max Planck Institute Research Report*, 1999.
35. Lavrentiev M. Sur quelques problemes du calcul des variations. *Annali di Matematica Pura ed Applicata* 1925; **4**:7–28.
36. Ball JM. *Singularities and Computation of Minimizers for Variational Problems*. Mathematical Institute of University of Oxford: Oxford, 1999 (preprint).
37. Garanzha VA. Maximum norm optimization of quasi-isometric mappings. *Numerical Linear Algebra with Applications* 2002; **9**(6–7):493–510.
38. Strang G, Fix J. *Analysis of Finite Element Method.* Prentice-Hall: Englewood Cliffs, NJ, 1973.
39. George PL, Borouchaki H. *Delaunay Triangulation and Meshing. Application to Finite Elements.* Hermes: Paris, 1998.
40. Alexandrov AD, Zalgaller VA. Two-dimensional manifolds of bounded curvature. *Trudy Mathematicheskogo Instituta Steklova* 1962; **63**. (English translation: Intrinsic geometry of surfaces. *Translated Mathematical Monographs*, vol. 15. American Mathematical Society: Providence, RI, 1967; Zbl.122,170).
41. Reshetnyak YuG. Two-dimensional manifolds of bounded curvature. In *Geometry IV* (*Non-regular Riemannian Geometry*), Reshetnyak Yu (ed.). Springer-Verlag: Berlin, 1991; 3–165.
42. Bakelman IYa, Generalized Chebyshev nets and manifolds of bounded curvature. *Doklady Akademii Nauk SSSR* 1961; **138**:506–507 (English translation: *Soviet Mathematical Doklads*; **2**:631–632, Zbl. 104,168).

43. Bakelman IYa. Chebyshev nets in manifolds of bounded curvature. *Trudy Mathematicheskogo Instituta Steklova* 1965; **76**:124–129 (English translation: *Proceedings Steklov Institute of Mathematics* 1967; **76**: 154–160, Zbl104,168).
44. Bonk M, Lang U. Bi-Lipschitz parameterization of surfaces. *Mathematische Annalen* 2003; **327**:135–169. DOI 10.1007/s00208-003-0443-8.
45. Burago YuD. On the proportional approximation of metric. *Trudy Mathematicheskogo Instituta Steklova* 1965; **76**:120–123 (English translation: Proceedings Steklov Institute of Mathematics 1967; **76**:149–153, Zbl104,168).
46. Alexandrov AD, *Intrinsic Geometry of Convex Surfaces.* Gostekhizdat: Leningrad, 1948 (in Russian).
47. Floater MS. Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design* 1997; **14**:231–250.
48. Rado T. Uber den Begriff der Riemannschen Flachen. *Acta Szeged.* 1925; **2**(Jbuch 51,273):101–121.

