SAND2008-8219P

Analysis of 2D, Rotation-Invariant, Non-Barrier Metrics in the Target-Matrix Paradigm

Patrick M. Knupp *
Applied Mathematics & Applications Dept., MS 1414
The Sandia National Laboratories
Albuquerque, NM USA
pknupp@sandia.gov

December 1, 2008

Abstract

The Target-Matrix Paradigm is a method for mesh optimization that optimizes Finite Element mesh quality in terms of user-defined target matrices. There are two goals in this Third of a series of papers describing the paradigm. The first goal is to develop a trial list of mathematical properties that a well-posed local quality metric would necessarily satisfy. The list is used to perform a detailed analysis of certain 2D, rotation-invariant, non-barrier metrics that can control the shape and size of local mesh elements. It turns out to be difficult to construct 'shape and size' metrics that can satisfy all eight of the properties. Only one such metric was found; it is new and non-obvious. The second goal is to distinguish between the trial properties that are necessary and those that are sufficient. The approach was to perform numerical experiments in order to observe the practical consequences of failing to satisfy one or more of the Eight Properties when optimizing a mesh. The numerical study revealed that some of the Eight Properties are more important than others in terms of detrimental effects on real meshes. Based on the results, a revised list of Eight *necessary* properties is given.

*This is SAND2008-XXXX. This work was funded by the Department of Energy's Mathematics, Information, and Computational Science Program (SC-31) and was performed at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy under contract DE-ACO4-94AL85000.

1 Introduction

Finite Element computations on complex geometries require appropriate meshes in order to control accuracy and efficiency. Mesh generation provides an initial mesh that may or may not meet the needs of the application. Even if it does so initially, the mesh may need to be updated in order to adapt to the solution or to changes in the domain geometry in time-evolving situations. Often the mesh is updated though a combination of topological and geometric changes that can control element shape, size, and orientation. Mesh optimization is one technique for performing geometric changes to the mesh; this is accomplished by fixing the mesh topology and changing the coordinates of mesh vertexes so that a particular multi-variable objective function that measures the fitness (or 'quality') of the mesh relative to the application is optimized. Mesh optimization has two major parts: (1) determining which objective function should be used, and (2) finding the most efficient method for rapidly computing the optimal mesh. The present work concentrates on the former issue. Many others have studied this problem, but not within the context of the Target-Matrix paradigm. For structured meshes, variational methods which seek out the optimal map have been studied [1], [2], [3]. Methods for unstructured meshes have tended to focus more on the discrete entities in a mesh: edge-lengths, element areas or volumes, and angles between edges or element faces [4], [5], [6], [7]. Barrier methods have been devised which can ensure the optimal mesh is valid, i.e., has positive Jacobian (in the case of variational methods [8]) or is untangled (in the case of the discrete optimization methods [9], [10]).

In order that the optimal meshes are appropriate to the application, mesh optimization objective functions have incorporated various weighting functions and parameters that can be adjusted in response to the solution or to other requirements. Unfortunately, it has proved difficult to automate the process of constructing weights and parameter values so that the most appropriate mesh is obtained. Part of the difficulty stems from the fact that the relationship between the weighting functions and the resulting mesh is imprecise. For example, mesh generators that solve Poisson's equation for the optimal mesh cannot easily guarantee that the map has a positive Jacobian determinant everywhere. Other methods, such as those using Harmonic maps, have provided clearer relations between the weights and the resulting mesh [11], [12]. Never-the-less, the problem remains an active research area in finite element meshing. One promising new method that attacks the question in the discrete mesh setting is the Target-Matrix paradigm [13], [14].

Basic components of the Target-Matrix Paradigm (TMP) include: *sample points*, which give the locations in the element where quality is to be measured, *active matrices*, which correspond to the Jacobian matrix of the local map from the master element to the physical element, *target matrices*, which correspond to the

Jacobian matrix of the local map from the target (or reference) element to the physical element, *local metrics*, which measure shape, size, and orientation at a sample point with respect to the target, *barrier metrics*, which prevent mesh tangling, and *objective functions*, which combine local qualities into a global measure of mesh quality in terms of the coordinates of mesh vertexes. Specific local metrics for creating planar meshes with the requisite local shape, size, and orientation were given in [14]. Both barrier and non-barrier forms of the metrics were given so that, if the initial mesh is untangled, a barrier metric will keep it so, while if the initial mesh is tangled, a non-barrier metric might untangle it.

The workhorse local metrics of the paradigm thus far have been the *shape*-metrics [15], [16], and the *shape-size-orientation*-metric [17]. *Shape-size*-metrics were also proposed in [14]; these metrics are important because they permit the optimal mesh to be as close as possible to the shape and size of the target or reference mesh, while being invariant to target orientation. It was found that the *non-barrier* forms of the *shape-size* metrics were particularly challenging to construct while satisfying the Eight Properties of a well-formulated TMP-metric (to be described in the next section). This paper focuses upon the non-barrier, shape-size metrics in order to investigate both their theoretical properties and the practical results of using them. Only the 2D metrics are considered in this paper. This work lays the groundwork for the much harder 3D case, which has also been studied, and for which another paper is forth-coming.

2 Properties of Well-Formulated Local Target Metrics

In [14], certain matrix sets that are important in describing the local metrics within the TMP were described. A short summary is given here so that the Eight Properties of a well-formulated metric can be described.

2.1 Canonical Matrix Sets

Let d be a positive integer and M_d be the set of real $d \times d$ matrices. Let I_d be the Identity Matrix in M_d[1] Let $B \in M_d$ and define $\beta = det(B)$ to be the determinant of B. Let $tr(B)$ be the trace of B, $|B|^2 = tr(B^t B)$ be the Frobenius norm-squared of B.[2] The adjoint of B is denoted by $adj(B)$. The elements of the adjoint B_{ij} are $(-1)^{i+j}$ times the determinant of the cofactor matrix obtained by deleting the j^{th} row and i^{th} column of B.

In TMP, the matrix B is defined as the product of the active Jacobian matrix A and the inverse of the reference Jacobian (or target) matrix, i.e., $B = AW^{-1}$. If $B = I$, then $A = W$, i.e., when B is the identity, the active Jacobian matrix equals the target-matrix. Thus, for example, the local metric $|B - I|^2$, when used in an objective function that is to be minimized, creates an optimal mesh that is 'closer' to the mesh suggested by the set of target matrices than was the initial mesh. The $|B - I|^2$ metric tends to create meshes whose Jacobians have the same size, shape, and orientation as the target Jacobians, because the global minimum of the metric is $A = W$. Suppose instead, there was a metric whose global minimum was $A = RW$, with R an arbitrary rotation. Such a metric would tend to create meshes having the same local size and shape as the mesh implied by the set of targets. However, the local orientation implied by the targets would not necessarily be present in the optimal mesh due to the rotation-invariance of the metric. This property could be useful, for example, in simplifying the target construction process or in reducing the requirements on the optimal mesh.

Motivated by the previous considerations, four canonical matrix sets $M_d^{(i)}$, $M_d^{(si+)}$, $M_d^{(o+)}$, and $M_d^{(so+)}$ are defined by describing the form of the matrices belonging to them. In particular, the members of the four sets have the following forms: I_d, sI_d, R, and sR, respectively, where s is an arbitrary scalar and R is an arbitrary rotation. Thus, for example, $M_d^{(o+)}$ is the set of $d \times d$ rotation matrices. For $d = 2$, the scalar s can be positive or negative, while for $d = 3$, one must insist that $s \geq 0$ because when $d = 3$, sR is a flip when s is

[1] For brevity, the symbol I will also be used.

[2] If B and C belong to M_d, their inner product is defined to be $B \cdot C = tr(B^t C)$.

negative. In this paper, local metrics are sought such that the global minimizers belong to either $M_2^{(o+)}$ (the 2D rotations) or $M_2^{(so+)}$ (the 2D scaled rotations). A critical issue is that the metrics must not specify a particular s or R; rather, it is required these be arbitrary so that the target can be constructed without paying attention to local size (s) or orientation (R). Before the Eight Properties of a well-formulated local target-metric are given, we briefly discuss some properties of rotation and flip matrices.

2.2 Rotations and Flips

Matrices belonging to M_d^{o+} are rotations. Recall that $B \in M_d$ is a rotation provided (i) $B^tB = I$ and (ii) $det(B) = 1$. In contrast, matrices that satisfy (i) and (iii) $det(B) = -1$ are called *flips*. Matrices that satisfy (i) only are *orthogonal*, thus both rotations and flips are orthogonal. For convenience, we denote arbitrary rotations, flips, or orthogonal matrices by R, F, and U, respectively. It is essential in mesh optimization to make the distinction between rotations and flips because the latter lead to inverted mesh elements. For example, if $B = AW^{-1}$ is a flip, then $\beta = -1$ and thus $det(A) = -det(W)$. Since a basic assumption of the paradigm is that the target matrices are always constructed so that $det(W) > 0$, the result of having B be a flip is that $det(A) < 0$, i.e., the local Jacobian determinant is negative.[3]

Definition
Let s be a scalar and R a rotation. Then a *scaled-rotation* is a matrix of the form sR.

Note that $det(sR) = s^d$. Thus for $d = 2$, $det(sR) \geq 0$ for any s, while for $d = 3$, $det(sR) \geq 0$ provided $s \geq 0$. In particular, $-R$ is a rotation when $d = 2$, but is a flip when $d = 3$.

Definition
Let s be a scalar and F a flip. Then a *scaled-flip* is a matrix of the form sF.

Note that $det(sF) = -s^d$. Thus for $d = 2$, $det(sF) \leq 0$ for any s, while for $d = 3$, $det(sF) \leq 0$ provided $s \geq 0$. In particular, $-F$ is a flip when $d = 2$, but is a rotation when $d = 3$.

For $d = 2$, rotation and flip matrices have the forms

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

[3]The assumption is reasonable since one would rarely want the target element to be inverted.

$$F = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

with $0 \leq \theta < 2\pi$. Define $P = diag(1, -1)$[4] Then $FP = R$, $PF = R^t$, $RP = F$, and $PR = F$. The set of rotations is closed under matrix multiplication, but the set of flips is not.

Finally, note that for any d, $(adjR)^t = R$, while $(adjF)^t = -F$. Thus, the adjoint of a rotation is a rotation for all d, but the adjoint of a flip is a flip for $d = 2$, but is a rotation when $d = 3$. For reasons such as this, 3D o+ metrics must necessarily be different than 2D o+ metrics and are thus reserved for a separate paper.

The following Proposition is used often in Section 3.
Proposition 1.
The quantities $\psi^2_{\pm}(B) = |B|^2 \pm 2\beta$ are non-negative for all $B \in M_2$. Further, ψ_+ is zero if and only if $B = tF$, and ψ_- is zero if and only if $B = tR$, with t any real number.
Proof.
For any $B \in M_2$, one has $(B_{11} \pm B_{22})^2 + (B_{12} \mp B_{21})^2 \geq 0$. Expanding this relation and gathering terms proves the first part of the Proposition. To prove the second part, let $t = \sqrt{B^2_{11} + B^2_{21}}$, and if $t \neq 0$, let $\cos\theta = B_{11}/t$, and $\sin\theta = B_{21}/t$. If $\psi_+ = 0$, then $B_{22} = -B_{11}$ and $B_{12} = B_{21}$. Substitution shows one can write $B = tF$. Similarly, if $\psi_- = 0$, then $B_{22} = B_{11}$ and $B_{12} = -B_{21}$. Substitution shows one can write $B = tR$. If $t = 0$, then $\psi_{\pm} = 0$ forces $B = 0$, which is both a scaled rotation and a scaled flip. §

2.3 Derivatives of Local Metrics

A local metric is a multi-variable function $\mu(B)$ from $B \in M_d$ to the real numbers. For example, if $d = 2$, one can write $\mu(B) = \mu(B_{11}, B_{12}, B_{21}, B_{22})$. Thus, the usual definitions of continuity and differentiability in multi-variable calculus apply to local metrics.

Given a point $B \in M_d$ and any non-zero $Z \in M_d$, let the set of points in an ϵ-neighborhood of B have the form $B + \epsilon Z$, with $\epsilon > 0$. The metric $\mu(B)$ is *continuous* at B if and only if $\lim_{\epsilon \to 0} \mu(B + \epsilon Z) = \mu(B)$ for all Z. The metric is discontinuous at B if (i) the limit does not exist for some Z, (ii) $\mu(B)$ is undefined, or (iii) the limit differs depending on the choice of Z.

The difference $\Delta\mu(B; \epsilon Z) = \mu(B + \epsilon Z) - \mu(B)$ is used in the definition of the

[4]The notation $B = diag(p, q)$ means that the matrix is diagonal, with $B_{11} = p$ and $B_{22} = q$.

directional derivative at the point B:

$$\begin{aligned}\delta\mu(B;Z) &= \lim_{\epsilon\to 0}\frac{\Delta\mu(B;\epsilon Z)}{\epsilon}\\ &= Z\cdot\frac{d\mu}{dB}\end{aligned}$$

where

$$\left(\frac{d\mu}{dB}\right)_{ij}=\frac{\partial\mu}{\partial(B_{ij})}$$

is a matrix in M_d. The stationary point equation (SPE) of a metric is the equation $\frac{d\mu}{dB}=0$. Points which satisfy this equation are called stationary points (SP's). The directional derivative is zero at a stationary point.

For future reference, some results of the definition of the derivative of $\mu(B)$ are useful in deriving the SPE's for the various metrics that will be discussed in the Section 3:

$$\begin{aligned}\frac{1}{2}\frac{d}{dB}|B|^2 &= B\\ \frac{d}{dB}\beta &= (adjB)^t\\ \frac{1}{2}\frac{d}{dB}|B^tB|^2 &= 2BB^tB\end{aligned}$$

Given a matrix $B\in M_d$, define the vector $\mathbf{v}(B)\in R^{d^2}$ by

$$\mathbf{v}(B)=(B_{11},B_{12},\ldots,B_{1d},B_{21},B_{22},\ldots,B_{2d},\ldots,B_{d1},B_{d2},\ldots,B_{dd})$$

Let $i=1,2,\ldots,d^2$. Then one can write $v_i(B)=B_{rs}$ where $r=\left[\frac{i-1}{2}\right]+1$ and $s=i-(r-1)d$. Define the outer product of two matrices X and Y in M_d in terms of the outer product of their corresponding vectors:

$$X\otimes Y=\mathbf{v}(X)\otimes\mathbf{v}(Y)$$

The outer product of two $d\times d$ matrices is a $d^2\times d^2$ matrix. For example, $(X\otimes Y)_{ij}=v_i(X)v_j(Y)$ and, if $d=2$, $(B\otimes B)_{23}=B_{12}B_{21}$.

Let $\partial v_i(B)=\partial B_{rs}$. Then the gradient $\nabla\mu$ of $\mu(B)$ is

$$\nabla\mu = \left(\frac{\partial\mu}{\partial B_{11}},\frac{\partial\mu}{\partial B_{12}},\ldots,\frac{\partial\mu}{\partial B_{dd}}\right)$$

or $(\nabla\mu)_i=\frac{\partial\mu}{\partial v_i}$.

If μ is sufficiently differentiable, the Hessian $\mathcal{H}\mu$ of $\mu(B)$ is a $d^2 \times d^2$ matrix with elements

$$(\mathcal{H}\mu)_{ij} = \frac{\partial^2 \mu}{\partial v_i \partial v_j}$$

The Taylor Series, to order ϵ^2, is

$$\begin{aligned} \mu(B + \epsilon Z) &= \mu(B) + \epsilon\, Z \cdot \nabla\mu + \frac{\epsilon^2}{2} \mathbf{v}^t(Z)\, (\mathcal{H}\mu)\, \mathbf{v}(Z) \\ &= \mu(B) + \epsilon \left(Z \cdot \frac{d\mu}{dB} \right) + \frac{\epsilon^2}{2}\, (\mathcal{H}\mu) \cdot (Z \otimes Z) \end{aligned}$$

The second term in the series is ϵ times the first directional derivative. If B is a stationary point, the second term is zero.

As an example, let $\mu(B) = |B|^2$. Then $\frac{d\mu}{dB} = 2\,B$ and $\mathcal{H}\mu = 2\,I$ so that

$$\mu(B + \epsilon Z) = |B|^2 + 2\,\epsilon\, (Z \cdot B) + \epsilon^2 |Z|^2$$

This agrees with the straightforward calculation $\mu(B + \epsilon Z) = |B + \epsilon Z|^2 = B \cdot B + 2\epsilon(B \cdot Z) + (\epsilon Z \cdot \epsilon Z)$.

From the Taylor Series one can write the difference as

$$\Delta\mu(B; Z) = \epsilon \left(Z \cdot \frac{d\mu}{dB} \right) + \frac{\epsilon^2}{2}\, (\mathcal{H}\mu) \cdot (Z \otimes Z)$$

Now define the second difference

$$\begin{aligned} \Delta^2\mu(B; Z) &= \mu(B + \epsilon Z) - 2\mu(B) + \mu(B - \epsilon Z) \\ &= \Delta\mu(B; Z) + \Delta\mu(B; -Z) \end{aligned}$$

and the second directional derivative

$$\begin{aligned} \delta^2\mu(B; Z) &= \lim_{\epsilon \to 0} \frac{\Delta^2\mu(B; Z)}{\epsilon^2} \\ &= (\mathcal{H}\mu) \cdot (Z \otimes Z) \end{aligned}$$

Example: if $\mu(B) = |B|^2$, then

$$\begin{aligned} \delta^2\mu(Y; Z) &= \lim_{\epsilon \to 0} \frac{2\epsilon^2 |Z|^2}{\epsilon^2} \\ &= 2\,|Z|^2 \end{aligned}$$

Local minima, maxima, and saddles of the function $\mu(B)$ are defined as follows. Let the metric be defined on the neighborhood of points around B of the form

$B + \epsilon Z$ with $Z \in M_d$. The point B is a *minimum* of μ if for all Z, there exists $\epsilon > 0$ sufficiently small such that $\mu(B+\epsilon Z) > \mu(B)$. Note that this can be written as $\Delta\mu(B; Z) > 0$. If $\Delta\mu(B; Z) \geq 0$, then the point B is a *semi-minimum*. Similarly, the point B is a *maximum* of μ if for all Z, there exists $\epsilon > 0$ sufficiently small such that $\mu(B + \epsilon Z) < \mu(B)$. Note that this can be written as $\Delta\mu(B; Z) < 0$. If $\Delta\mu(B; Z) \leq 0$, then the point B is a *semi-maximum*.

If $\mu(B)$ is sufficiently differentiable with respect to B, then to order ϵ^2

$$\Delta\mu(B; Z) \quad = \quad \epsilon \left(Z \cdot \frac{d\mu}{dB} \right) + \frac{\epsilon^2}{2} \, (\mathcal{H}\mu) \cdot (Z \otimes Z)$$

If B is a stationary point of μ, then B is a minimum provided $(\mathcal{H}\mu) \cdot (Z \otimes Z) > 0$ for all Z.[5] If B is a stationary point of μ, then B is a maximum provided $(\mathcal{H}\mu) \cdot (Z \otimes Z) < 0$ for all Z. If B is a stationary point, and neither of the previous hold, then it is a saddle point.

Finally, note that if $\mu(B) = \mu_1(B) + \gamma \, \mu_2(B)$ is a composite metric, then the Taylor Series can be used to show that

$$\frac{d\mu}{dB} \quad = \quad \frac{d\mu_1}{dB} + \gamma \, \frac{d\mu_2}{dB}$$

and

$$(\mathcal{H}\mu) \quad = \quad (\mathcal{H}\mu_1) + \gamma \, (\mathcal{H}\mu_2)$$

provided each of the metrics is sufficiently differentiable. Thus one can find the derivative and Hessian of a composite metric by first finding the derivatives and Hessians of the individual metrics which it contains. The first formula shows that if B_s is a stationary point of both μ_1 and μ_2, then it is a stationary point of μ. If B_s is a stationary point of μ_1, but not of μ_2, then it is not a stationary point of μ.

2.4 The Eight Properties

The paradigm requires the construction of local metrics $\mu(B)$ from $\mathcal{D} \subseteq M_d$ to the real numbers. It would be helpful in constructing such metrics to possess a list of mathematical properties that a well-formulated metric must satisfy. Certain ideal properties would seem to ensure that a TMP metric is well-posed; these are used to create a trial list of properties that will be investigated to determine whether they are necessary, sufficient, or both. The trial list of properties is stated with the assumption that the local metric is to be *minimized*[6]

[5] i.e., the Hessian is positive definite.

[6] These properties were first presented in [14] and are refined in this paper.

1. The metric is a continuous function of B on a domain $\mathcal{D}$. For non-barrier metrics, the domain $\mathcal{D}$ on which the metric is defined is equal to M_d. For barrier metrics, the domain $\mathcal{D}$ on which the metric is defined is equal to M_d^+.[7]

2. The metric does not attain a local or global minimum when $|B| \to \infty$.

3. There exists a constant $c \geq 0$ such that $\mu(B) \geq c$ for all $B \in \mathcal{D}$.

4. There exists a global minimizer $B_m \in \mathcal{D}$ such that $\mu(B_m) = c$ and $|B_m| < \infty$.

5. B_m is a member of one of the four canonical sets M_d^i, M_d^{si+}, M_d^{o+}, or M_d^{so+}.

6. If there is more than one global minimizer, they all belong to the same canonical set as B_m.

7. The metric is differentiable with respect to B on $\mathcal{D}^* \subseteq \mathcal{D}$. If the metric attains an extrema at any non-differentiable point in $\mathcal{D}$, then that point is a global minimizer.

8. The set of stationary points of the metric on $\mathcal{D}^*$ coincides with the set of global minimizers on $\mathcal{D}^*$.

Property 1 ensures that points in $\mathcal{D}$ which are 'close' to one another have local metric values that are close. When satisfied, the local mesh quality varies continuously with the geometric properties of the local mesh. For non-barrier metrics, it is desirable that $\mathcal{D} = M_d$ because then the value of the metric exists for any matrix B that is derived at a sample point in the mesh. When optimizing meshes, one does not know in advance what these matrices will be, so it is best that the metric be able to deal with anything that comes along. However, barrier-metrics such as condition number or inverse mean ratio must necessarily have their domain restricted to the set of matrices $M_d^{(+)}$ that have positive determinants because it is not possible to have both a barrier and be defined on all of M_d in such a manner as to succeed when optimizing a tangled mesh.

Property 2 ensures that the metric does not have a minimum point at infinity. To be more precise, let $\lim_{t\to\infty} \mu(tZ) = \mu_\infty(Z)$ for $Z \neq 0$. Then Property 2 is satisfied provided this limit does not exist for any Z (sufficient condition). For example, if $\mu(B) = |B|^2$, then $\mu(tZ) = t^2|Z|^2$, and so the limit does not exist for any $Z \neq 0$. Alternatively, if the limit exists for some Z, then Property 2 is satisfied provided there exists t^* such that $\mu(tZ) - \mu_\infty(Z) \leq 0$ for all $t > t^*$ (necessary condition). For example, if $\mu(B) = -\frac{1}{|B|^2}$, then $\mu(tZ) = -\frac{1}{t^2|Z|^2}$.

[7] M_d^+ is the set of $d \times d$ matrices having a positive determinant.

Thus $\mu_\infty(Z) = 0$, and $\mu(tZ) - \mu_\infty(Z) < 0$ for $t > 0$.

Because the local metric is to be minimized, it is necessary that it be bounded below, hence Property 3. Property 4 ensures the existence of a finite minimizing matrix, while 5 says the B_m will be of the form I_d, sI_d, R, or sR. Property 6 says that if there is more than one global minimizer, then they all have the same form (for example, they will all be rotations). Property 7 says that the set of points at which the metric is differentiable belongs to a stated set $\mathcal{D}^*$; ideally, this set is the same as M_2, or at least not much smaller. Metrics having a few non-differentiable points in $\mathcal{D}$ are permitted, but can only be global minimizers. In particular, metrics with non-differentiable points that are local maxima or minima should be avoided so that the global objective function can be more readily minimized and not give multiple solutions.[8] Property 8 says that the metric will not have unwanted stationary points (local minima, maxima, or saddles).

Metrics are designed at first to satisfy Property Five, i.e., the global minimum of the metric is to belong to a particular set of matrices. As will be seen, it is not always possible to design a metric satisfying all Eight Properties. A practical question then, is to determine the consequences of using a metric that does not satisfy all the properties. In this document we highlight the difficulties encountered in designing rotation-invariant, non-barrier metrics such that they satisfy all Eight Properties and numerically investigate the consequences of not satisfying all of them. Before doing so, additional motivation for this list of properties is provided by connecting it to the value of the metric as a function of its local vertex coordinates.

2.5 Local Metrics as a Function of Vertex Coordinates

The objective functions used in mesh optimization are functions of the coordinates $\mathbf{x}$ of the free vertexes in the mesh. For the most part, the notation in this paper hides this fact by concentrating on the matrix B. However, since the active (or Jacobian) matrix A is a function of the vertex coordinates within an element, so is B. This can be expressed as $B = B(\mathbf{x})$. The actual function depends on the form of the map from the master element to the physical element (e.g. quadratic vs. linear), however, both A and B are linear in each of the local vertex coordinates. Given a sample point in the master element and the map, the matrix $B(\mathbf{x})$ is uniquely determined. Because B can be considered a function of the mesh coordinates, one can consider the local metric to also be a function of the coordinates. The primary topic in this section is to consider the

[8] This requirement partly depends on the numerical optimization solver. If the solver is designed to take non-differentiable points into account, then the metric can be non-differentiable. Otherwise, the metric should be differentiable on all of $\mathcal{D}$.

properties of the composition function $\tilde{\mu}(\mathbf{x}) = \mu(B(\mathbf{x}))$ when $\mu = \mu(B)$ satisfies the Eight Properties above.

Domain of $\tilde{\mu}$: If μ is defined on $\mathcal{D}$, then $\tilde{\mu}$ is defined on $\mathcal{D}_x$ where

$$\mathcal{D}_x = \{\mathbf{x} \,|\, B(\mathbf{x}) \in \mathcal{D}\}$$

Property 1. The function $\tilde{\mu}$ is continuous on $\mathcal{D}_x$ because the elements of B are continuous functions of $\mathbf{x}$ and μ is a continuous function of B.

Property 2. Suppose $\lim_{t\to\infty} \mu(tZ)$ does not exist for any Z. But $\mu(tZ) = \mu(tZ(\mathbf{x})) = \mu(Z(t\mathbf{x})) = \tilde{\mu}(t\mathbf{x})$, so the limit of the latter as $t \to \infty$ does not exist either. Similarly, if there exists t^* such that $\mu(tZ) - \mu_\infty(Z) \le 0$ for $t > t^*$, then $\tilde{\mu}(t\mathbf{x}) - \tilde{\mu}_\infty(\mathbf{x}) \le 0$ for $t > t^*$, where $\tilde{\mu}_\infty(\mathbf{x}) = \lim_{t\to\infty} \tilde{\mu}(t\mathbf{x})$.

Property 3. If $\mu(B) \ge c \ge 0$ on $\mathcal{D}$, then $\tilde{\mu}(B) \ge c \ge 0$ on $\mathcal{D}_x$.

Property 4. If $\mu(B_m) = c$, then $\tilde{\mu}(x_m) = c$ *provided* x_m exists, i.e., given B_m, one can find a solution $\mathbf{x}_m$ to the equation $B_m = B(\mathbf{x}_m)$. If all the mesh vertexes are free, then $\mathbf{x}_m$ exists, but if too many of them are fixed, then it may not exist. If $\mathbf{x}_m$ does not exist, the most that can be said is that, since $\tilde{\mu}$ is continuous and bounded below, and $|B_m| < \infty$, then there there must exist a finite global minimum of the function $\tilde{\mu}(\mathbf{x})$. The analysis of the 2D metrics in the next section assumes that all of the vertexes are free; in that case, $\tilde{\mu}$ attains the value c when $B = B_m$.[9]

Property 6. If Property 6 holds, then $\tilde{\mu}$ may have more than one global minimizer.

Property 7. If μ is differentiable with respect to B on $\mathcal{D}^*$, then

$$\begin{aligned} \frac{d\tilde{\mu}}{dx} &= \frac{d\mu}{dB} \cdot \frac{dB}{dx} \\ \frac{d\tilde{\mu}}{dy} &= \frac{d\mu}{dB} \cdot \frac{dB}{dy} \end{aligned}$$

Thus $\tilde{\mu}$ is differentiable with respect to $\mathbf{x}$ on $\mathcal{D}_x^* = \{\mathbf{x} \,|\, B(\mathbf{x}) \in \mathcal{D}^*\}$ because B is differentiable with respect to $\mathbf{x}$. If μ is non-differentiable at a point B, then $\tilde{\mu}$ is non-differentiable at the corresponding point $\mathbf{x}$.

[9] This assumption is justified on the grounds that the goal is to analyze the ideal behavior of the local metrics first, in order to construct the best possible metric. When applied to real meshes having fixed vertexes, the assumption will not apply, which means that the optimal mesh Jacobians will not exactly match the set of target matrices. This is nothing new because that has always been the basic understanding of optimization methods.

Property 8. If μ satisfies Property 8, with B_m a global minimizer, then $\mu(B_m) = c$ if and only if $\frac{d\mu}{dB}|_{B=B_m} = 0$. Given B_m, let $\mathbf{x}_m$ be a solution to $B_m = B(\mathbf{x})$. Then (i) $\tilde{\mu}(\mathbf{x}_m) = c$ and, from the equations in the previous paragraph, (ii) $\frac{d\tilde{\mu}}{d\mathbf{x}}|_{\mathbf{x}=\mathbf{x}_m} = 0$. Thus, the global minimizers of $\tilde{\mu}$ on $\mathcal{D}_x^*$ coincide with its stationary points on the same set.

Proposition 2.
If μ satisfies Properties 1-8, then the function $\tilde{\mu}$ has no saddle points, nor any extrema except for global minima.
Proof.
Suppose there exists a maximum point $\mathbf{x}_e \in \mathcal{D}_x$ such that $\tilde{\mu}(\mathbf{x}_e) \geq \tilde{\mu}(\mathbf{x}_e + \epsilon\,\mathbf{x})$ for every point $\mathbf{x}$ in an ϵ-neighborhood about $\mathbf{x}_e$. Then $\mu(B(\mathbf{x}_e)) \geq \mu(B(\mathbf{x}_e + \epsilon\,\mathbf{x})) = \mu(B(\mathbf{x}_e) + \epsilon B(\mathbf{x})) = \mu(B_e + \epsilon B)$. Hence $B_e \in \mathcal{D}$ is a maximum point of $\mu(B)$. This is a contradiction since then Property 7 or Property 8 would not be satisfied. A similar argument can be made for the case $\mathbf{x}_e$ being a saddle point. If $\mathbf{x}_e$ is a minimum, then a similar argument shows that $\mathbf{x}_e$ is a global minimum. §

To summarize, if the metric $\mu(B)$ satisfies the Eight Properties, and it is assumed all the mesh vertexes are free, then the function $\tilde{\mu}(\mathbf{x})$ is continuous on $\mathcal{D}_x$, differentiable on $\mathcal{D}_x^* \subseteq \mathcal{D}_x$, bounded below, attains a global minimum on $\mathcal{D}_x$, has coincident stationary points and global minimizers, and contains no local extrema.

The statements in Section 2 apply, for the most part, to both the cases $d = 2$ and $d = 3$. In the next section on constructing suitable o+ and so+ metrics, only the case $d = 2$ is considered.

3 Analysis of various so+ and o+ Metrics

This section has four major parts, focusing on (3.1) an analysis of two so+ metrics, (3.2) an analysis of several o+ metrics, (3.3) an analysis of a particularly interesting o+ metric that satisfies all Eight mathematical properties, and (3.4) additional observations on the latter. The metrics presented are analyzed in order to ascertain whether or not each satisfies the Eight Properties described in the previous section. One of the more challenging aspects of this is to find *all* of the solutions to the stationary point equation (SPE), which in general consists of a non-linear system of four equations in four unknowns. Various mathematical techniques are applied to illustrate the possible approaches to finding solutions to the SPE. Other challenges are to analyze the behavior of the metrics at non-differentiable points and to classify the stationary points in terms of being local or global minima, maxima, or saddle points.

To begin, a few notations and observations will be useful. First, let

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

Then

$$adj(B) = \begin{pmatrix} B_{22} & -B_{12} \\ -B_{21} & B_{11} \end{pmatrix}$$

For any vector $\mathbf{x} = (x_1, x_2) \in R^2$, define $\mathbf{x}^\perp = (-x_2, x_1)$. Then $\mathbf{x} \cdot \mathbf{x}^\perp = 0$, $|\mathbf{x}| = |\mathbf{x}^\perp|$, and $(\mathbf{x}^\perp)^\perp = -\mathbf{x}$. Now define $\mathbf{b}_1 = (B_{11}, B_{21})$ and $\mathbf{b}_2 = (B_{12}, B_{22})$. Then we can write $B = [\mathbf{b}_1, \mathbf{b}_2]$ and $(adjB)^t = [-\mathbf{b}_2^\perp, \mathbf{b}_1^\perp]$. Let $R = [\mathbf{r}_1, \mathbf{r}_1^\perp]$, with $\mathbf{r}_1$ a unit vector, be a rotation matrix.

3.1 Two so+ Metrics

Metrics of the type so+ must have global minimizers belonging to $M_d^{(so+)}$. For $d = 2$, the global minimizers have the form $B = sR$, with s any real number.

3.1.1 The first so+ metric

Proposition 3.
The matrix $B \in M_2$ is a scaled rotation if and only if $B = (adjB)^t$.
Proof.
Suppose B is a scaled rotation, i.e., $B = sR$ with s any real scalar. Then

$$\begin{aligned} adj(B) &= adj(sR) \\ &= s\,R^t \\ &= B^t \end{aligned}$$

from which the first part of the result follows. Now suppose $B = (adjB)^t$. Then $\mathbf{b}_1 = -\mathbf{b}_2^{\perp}$ and $\mathbf{b}_2 = \mathbf{b}_1^{\perp}$. If $B \neq 0$, we can write the first as $\mathbf{b}_1 = s\mathbf{r}_1$ with $s = |-\mathbf{b}_2^{\perp}|$ and $\mathbf{r}_1 = -\mathbf{b}_2^{\perp}/s$. Then $\mathbf{b}_2 = s\mathbf{r}_1^{\perp}$, so that $B = sR$. If $B = 0$, we can choose $s = 0$ to claim that $B = sR$. §

Proposition 3 suggests the so+ metric

$$\begin{aligned} \mu_2^{(so+)}(B) &= \frac{1}{2}|B - (adjB)^t|^2 \quad (1) \\ &= |B|^2 - 2\beta \quad (2) \end{aligned}$$

The greatest lower bound of the metric is zero. The metric is zero if and only if $B = (adjB)^t$. From Proposition 3, the global minimizers are scaled rotations and thus belong to $M_2^{(so+)}$. The metric has no barrier since it's value is finite when $\beta = 0$.

This metric clearly satisfies Properties 1-6 in Section 2.4, with $\mathcal{D} = M_2$.

The metric in equation (2) is differentiable with respect to B on the set $\mathcal{D}^* = M_2$, thus satisfying Property 7. Furthermore, the stationary point equation is $B - (adjB)^t = 0$ and so the stationary points satisfy $B = (adjB)^t$. Therefore the stationary points of the metric in equation (1) or (2) coincide with the global minimizers and Property 8 is satisfied. In conclusion, the so+ metric in (1)-(2) satisfies the Eight Properties on M_2.

3.1.2 The second so+ metric

Although the previous so+ metric is fully satisfactory, another so+ metric is analyzed in this section because it does not satisfy Property 8 and thus can be used to investigate the numerical consequences of failing that property. Let

$$\begin{aligned} \mu_2^{(so+)} &= |B^tB - \beta I|^2 \\ &= |B^tB|^2 - 2\beta|B|^2 + 2\beta^2 \end{aligned}$$

Clearly, this metric is zero when $B = sR$. On the other hand, if $\mu = 0$, B must satisfy $B^tB = \beta I$. Solutions to this relation have the form $B = tU$; substitution of this form into the relation yields $t^2 = t^2 det(U)$. Either $t = 0$ or $1 = det(U)$. Therefore, the global minimizers have the form $B = tR$, with t arbitrary; this is the set of scaled rotations.

The metric is defined on all of M_2 and is a non-barrier metric because its value is finite on the set of singular matrices. Properties 1-6 are satisfied. The metric is also differentiable on M_2 (Property 7), and the SPE is

$$\beta B = \{\frac{1}{2}|B|^2 - \beta\}(adjB)^t + BB^tB$$

Note that this set of four equations is non-linear and thus finding all the solutions is likely to be difficult. Looking first for solutions of the form $B = tU$, one finds by substitution, that both $B = tR$ and $B = tF$, with t arbitrary, are stationary points. Thus, the metric does not satisfy Property 8 because the set of stationary points does not coincide with the set of global minimizers.

To find *all* the solutions of the SPE, recall that for any B, there exists orthogonal matrices U and B, and a diagonal matrix Δ, with $\Delta_{ii} \geq 0$, such that $B = U\Delta V^t$ (the singular value decomposition). Then $\beta = det(UV)det(\Delta) = \pm det(\Delta)$, $BB^tB = U\Delta^3V^t$, and $(adjB)^t = (adjU)^t(adj\Delta)(adjV)$. Substituting these relations into the SPE, pre-multiplying by U^t, and post-multiplying by V yields the following equation

$$det(UV)\, det(\Delta)\, \Delta = \{\frac{1}{2}|\Delta|^2 - det(UV)\, det(\Delta)\}[det(UV)](adj\Delta) + \Delta^3$$

In terms of the entries of Δ, this is the same as the pair

$$\begin{aligned} [det(UV)]\, \Delta_{11}^2\Delta_{22} &= \{\frac{1}{2}|\Delta|^2 - [det(UV)]\, \Delta_{11}\Delta_{22}\}[det(UV)]\Delta_{22} + \Delta_{11}^3 \\ [det(UV)]\ \Delta_{11}\Delta_{22}^2 &= \{\frac{1}{2}|\Delta|^2 - [det(UV)]\, \Delta_{11}\Delta_{22}\}[det(UV)]\Delta_{11} + \Delta_{22}^3 \end{aligned}$$

which can be factored as

$$\begin{aligned} \{\Delta_{11} - [det(UV)]\, \Delta_{22}\}\, \{\Delta_{11} + [det(UV)]\, \Delta_{22}\} \left\{\Delta_{11} - \frac{1}{2}\, [det(UV)]\, \Delta_{22}\right\} &= 0 \\ \{\Delta_{22} - [det(UV)]\, \Delta_{11}\}\, \{\Delta_{22} + [det(UV)]\, \Delta_{11}\} \left\{\Delta_{22} - \frac{1}{2}\, [det(UV)]\, \Delta_{11}\right\} &= 0 \end{aligned}$$

There are two sets of solutions to the above: (a) $\Delta_{22} = [det(UV)]\, \Delta_{11}$, or (b) $\Delta_{22} = -[det(UV)]\, \Delta_{11}$. Thus, if $det(UV) = 1$, solution (a) gives $\Delta = tI$ for any $t \geq 0$, and solution (b) gives $\Delta = 0$ because $\Delta_{ii} \geq 0$. So, B has the form tUV^t, which is a scaled rotation because $det(UV^t) = 1$.

If, on the other hand, $det(UV) = -1$, solution (a) gives $\Delta = 0$ because $\Delta_{ii} \geq 0$, while (b) gives $\Delta = tI$ for any $t \geq 0$. So, B has the form tUV^t, which is a scaled flip because $det(UV^t) = -1$.

Thus, no new solutions were found by the SVD technique. For other metrics, however, the SVD can yield additional solutions.

To complete the analysis of this metric, the nature of the stationary points of the form tR and tF must be determined. First, note that $\mu_2^{(so+)}(tF) = 8t^4$. Thus, these stationary points cannot be global minima, except when $t = 0$. But $t = 0$ gives $B = 0$, which is also a scaled rotation. Let $\epsilon > 0$ and $Y, Z \in M_2$. Let

$B = Y + \epsilon Z$; then B lies in an ϵ-neighborhood of the point Y. The difference in the value of the metric at the points B and Y is, to order ϵ^2,

$$\Delta\mu_2^{(so+)}(Y;Z) \quad = \quad 2\epsilon\, P(Y,Z) + \epsilon^2 Q(Y,Z)$$

with

$$\begin{aligned} P(Y,Z) &= (M \cdot Y^t Y) - \left[Z \cdot adj(Y^t)\right] |Y|^2 + 2det(Y) \left[Z \cdot \left((adjY)^t - Y\right)\right] \\ Q(Y,Z) &= |M|^2 + 2(Z^t Z) \cdot (Y^t Y) - 2det(Y) \left[|Z|^2 - 2det(Z)\right] \\ &+ 2\left[Z \cdot (adjY)^t\right]^2 - 4(Z \cdot Y)(Z \cdot adj(Y^t)) - 2|Y|^2 det(Z) \end{aligned}$$

with $M = (Z^t Y) + (Z^t Y)^t$. For $Y = tR$, the difference becomes

$$\Delta\mu_2^{(so+)}(tR;Z) \quad = \quad 2t^2\epsilon^2\{|Z|^2 - 2det(Z) + \left(tr(Z^t R)\right)^2\}$$

This is non-negative, so the scaled rotations are local minima. For $Y = tF$, the difference becomes

$$\Delta\mu_2^{(so+)}(tF;Z) \quad = \quad 16t^3(Z \cdot F)\epsilon$$

The difference is non-negative for $Z = F$ and non-positive for $Z = -F$, and thus $Y = tF$ is a saddle point.

In summary, the second so+ metric satisfies Properties 1-7 on M_2. Property 8 is not satisfied since there exists stationary points in M_2 of the form $B = tF$ with $t \neq 0$ that are not global minimizers; these points are saddle points.

3.2 Metrics for o+

Metrics of this type require that the global minimizers are rotations. Achieving this via a *non-barrier* metric that satisfies the Eight Properties is surprisingly challenging. This is due, in part, to the fact that $B = 0$ is a scaled rotation and thus can be a global minimizer of a non-barrier so+ metric, but $B = 0$ is not a rotation and therefore is not permitted to be a global minimizer of an o+ metric (with or without barrier).

3.2.1 The Metric $\mu_2^{(o)}$

The following metric is not an o+ metric, but because it appears as a term in some of the subsequent o+ metrics, it is worth analyzing separately here. Let

$$\mu_2^{(o)} = |B^t B - I|^2$$

The global minimizers are $B = R$ and $B = F$. Properties 1-4 and 7 are satisfied by this metric on the set M_2.

The SPE for this metric is

$$B = BB^tB$$

There are *five* solutions: (i) $B = 0$, (ii) $B = R$, (iii) $B = F$, (iv) $B = \mathbf{u}_1 \otimes \mathbf{v}_1$, and (v) $B = \mathbf{u}_2 \otimes \mathbf{v}_2$, where $B = U\Delta V^t$, $U = [\mathbf{u}_1, \mathbf{u}_2]$, and $V = [\mathbf{v}_1, \mathbf{v}_2]$.[10] The latter two stationary points are non-orthogonal.

To classify the stationary points, let $B = B_s + \epsilon Z$, with B_s any stationary point and $Z \in M_2$. Then B lies in an ϵ-neighborhood of B_s. If $B = B_s + \epsilon Z$, then

$$\begin{aligned} B^tB &= B_s^tB_s + 2\epsilon\, M + \epsilon^2 Z^tZ \\ |B|^2 &= |B_s|^2 + 2\epsilon\, tr(Z^tB_s) + \epsilon^2|Z|^2 \end{aligned}$$

with $M = (Z^tB_s) + (Z^tB_s)^t$ and

$$\begin{aligned} (B^tB)^2 &= (B_s^tB_s)^2 + \epsilon\{B_s^tB_sM + MB_s^tB_s\} \\ &+ \epsilon^2\{B_s^tB_sZ^tZ + Z^tZB_s^tB_s + M^2\} + \Theta(\epsilon^3) \end{aligned}$$

The previous gives

$$\begin{aligned} |B^tB|^2 &= |B_s^tB_s|^2 + 2\epsilon\,(M \cdot B_s^tB_s) \\ &+ \epsilon^2\{2\,(Z^tZ \cdot B_s^tB_s) + |M|^2\} + \Theta(\epsilon^3) \end{aligned}$$

Noting that $\mu_2^{(o)}(B) = |B^tB|^2 - 2|B|^2 + 2$, we have

$$\begin{aligned} \Delta\mu_2^{(o)}(B_s; Z) &= |B^tB|^2 - |B_s^tB_s|^2 - 2\,\{|B|^2 - |B_s|^2\} \\ &= 2\epsilon\,\{M \cdot B_s^tB_s - 2(Z \cdot B_s)\} \\ &+ \epsilon^2\{2Z^tZ \cdot B_s^tB_s + |M|^2 - 2|Z|^2\} \end{aligned}$$

However,

$$\begin{aligned} M \cdot B_s^tB_s &= tr(MB_s^tB_s) \\ &= tr(Z^tB_s + B_s^tZ)B_s^tB_s \\ &= tr(Z^tB_sB_s^tB_s + B_s^tZB_s^tB_s) \\ &= tr(Z^tB_s + B_s^tZ) \end{aligned}$$

where the last line uses the stationary point equation and the fact that for any two matrices $tr(PQ) = tr(QP)$. Hence $M \cdot B_s^tB_s = tr(M) = 2Z \cdot B_s$. Therefore, the order ϵ term in the difference above is zero and so

$$\Delta\mu_2^{(o)}(B_s; Z) = \epsilon^2\{2\,Z^tZ \cdot B_s^tB_s + |M|^2 - 2|Z|^2\}$$

[10] The matrix $\mathbf{u} \otimes \mathbf{v}$ is the outer product of the vectors $\mathbf{u}$ and $\mathbf{v}$.

Using the difference result just derived, the stationary point $B_s = 0$ is a local maximum because $\mu_2^{(o)}(0) \neq 0$ and, to second order, $\Delta\mu_2^{(o)}(0; Z) = -2\epsilon^2|Z|^2 \leq 0$ for all Z.

The stationary points (ii) and (iii) are global minima because $\mu_2^{(o)}(R) = 0$ and because the difference when $B_s = R$ is $\epsilon^2|M|^2$ and is thus non-negative. The reasoning for $B_s = F$ is very similar.

The non-orthogonal stationary point (iv) requires the following facts:

$$\begin{aligned}(\mathbf{u}_1 \otimes \mathbf{v}_1)^t(\mathbf{u}_1 \otimes \mathbf{v}_1) &= (\mathbf{v}_1 \otimes \mathbf{v}_1) \\ Z^tZ \cdot (\mathbf{v}_1 \otimes \mathbf{v}_1) &= |Z\mathbf{v}_1|^2 \\ Z^t(\mathbf{u}_1 \otimes \mathbf{v}_1) &= (Z^t\mathbf{u}_1) \otimes \mathbf{v}_1 \\ |M|^2 &= 2\{\left(Z^t\mathbf{u}_1 \cdot \mathbf{v}_1\right)^2 + |Z^t\mathbf{u}_1|^2\}\end{aligned}$$

For $B_s = (\mathbf{u}_1 \otimes \mathbf{v}_1)$, the difference becomes

$$\Delta\mu_2^{(o)}(B_s; Z) = 2\epsilon^2\{|Z\mathbf{v}_1|^2 + \left(Z^t\mathbf{u}_1 \cdot \mathbf{v}_1\right)^2 + |Z^t\mathbf{u}_1|^2 - |Z|^2\}$$

For $Z = I$, the difference is $2\epsilon^2(\mathbf{u}_1 \cdot \mathbf{v}_1)^2$, which is non-negative. On the other hand, if we choose $Z = [\mathbf{u}_1^\perp, \mathbf{u}_1^\perp]$, then $Z^t\mathbf{u}_1 = 0$, so the difference is $2\epsilon^2\{|Z\mathbf{v}_1|^2 - |Z|^2\}$. But $|Z\mathbf{v}_1| \leq |Z||\mathbf{v}_1| = |Z|$, and thus the difference is non-positive for any U and V. Therefore, this stationary point is a saddle.

Similar reasoning shows the stationary point (v) is also a saddle.

In summary, the metric $\mu_2^{(o)}$ has two distinct sets of global minimizers, namely, the set of rotations and the set of flips. Moreover, it has five stationary points: the two global minimizers, two saddle points, and one local maximum.

3.2.2 An o+ Metric With Non-orthogonal Stationary Points

Orthogonal matrices satisfy $B^tB = I$, so a straightforward o+ metric to consider is

$$\mu_2^{(o+)}(B) = |B^tB - I|^2 + \gamma(\beta - 1)^2 \tag{3}$$

with the parameter $\gamma > 0$ to be determined. The first term of the metric is minimized by $B = R$ or $B = F$, and the second by matrices with unit determinant. Thus $\mu_2^{(o+)} = 0$ if and only if B is a rotation. The domain of the metric is $\mathcal{D} = M_2$. Properties 1-6 are satisfied.

The metric is differentiable on M_2 (Property 7), and the stationary point equation is

$$B = \frac{\gamma}{2}(\beta - 1)(adjB)^t + BB^tB \tag{4}$$

Note that if B_s is a solution to (4), then so is -B_s. Substituting $B = U\Delta V^t$ into (4) gives

$$\Delta \;=\; \frac{\gamma}{2}\left\{det(\Delta) - det(UV)\right\}(adj\Delta) + \Delta^3$$

The permissible solutions to this equation are

a. $\Delta = 0$,

b. $\Delta = tI$ with $t = \sqrt{\frac{2+\gamma det(UV)}{2+\gamma}}$,

c. $\Delta_{11} = \sqrt{\frac{1\pm t}{2}}$ and $\Delta_{22} = \sqrt{\frac{1\mp t}{2}}$ with $t = \sqrt{1 - \left(\frac{2\gamma}{\gamma-2}\right)^2}$, where $\gamma \neq 2$.

The following are thus solutions to (4)

a'. $B = 0$,

b'. $B = \pm R$ (from $det(UV) = +1$),

b''. $B = \pm tF$ with $t = \sqrt{\frac{2-\gamma}{2+\gamma}}$ (from $det(UV) = -1$),

c'. $B = \pm U diag(\sqrt{\frac{1\pm t}{2}}, \sqrt{\frac{1\mp t}{2}})V^t$ with $t = \sqrt{1 - \left(\frac{2\gamma}{\gamma-2}\right)^2}$ and $\gamma \neq 2$.

Property 8 is thus not satisfied because solutions a', b'', and c' are not global minimizers.[11] Some of the unwanted stationary points can be excluded by a judicious choice of γ. For example, if $\gamma > 2$, then t in solution b'' is imaginary. Similarly, if $\gamma > \frac{2}{3}$, then t in solution c' is imaginary. Summarizing, if $0 < \gamma \leq \frac{2}{3}$, then all four solutions are real; if $\frac{2}{3} < \gamma \leq 2$, then solutions a', b', and b'' are real; and, if $2 < \gamma$, then only solutions a' and b' are real. So, even when $2 < \gamma$, Property 8 is not satisfied. This completes the analysis of the SPE.

The nature of the metric at its stationary points is examined next. Let B_s be a stationary point of $\mu_2^{(o+)}$ in (3), $\epsilon > 0$, and let $Z \in M_2$, so that $B = B_s + \epsilon Z$ lies in an ϵ-neighborhood of B_s. Also let $M = B^tB - I$, $M_s = B_s^tB_s - I$, $Y = B_s^tZ + Z^tB_s$, and $\beta_s = det(B_s)$. Then,

$$\begin{aligned} M &= M_s + \epsilon Y + \epsilon^2 Z^tZ \\ \beta - 1 &= (\beta_s - 1) + \epsilon\, tr[(adjZ)B_s] + \epsilon^2 det(Z) \end{aligned}$$

[11] Note that solution c' is not only not a rotation, it is not even orthogonal!

and thus, to order ϵ^2,

$$\begin{aligned}\Delta\mu_2^{(o+)}(B_s;Z) &= 2\epsilon\left\{tr(M_sY)+\gamma(\beta_s-1)tr[(adjZ)B_s]\right\}\\ &+ \epsilon^2\left\{tr(Y^2+M_sZ^tZ+Z^tZM_s)\right.\\ &+ \left.\gamma[2(\beta_s-1)\,det(Z)+(tr[(adjZ)B_s])^2]\right\} \qquad (5)\end{aligned}$$

First, consider the stationary point in a', i.e., $B_s = 0$. Then the difference (5) becomes

$$\Delta\mu_2^{(o+)}(0;Z) = -2\epsilon^2\{|Z|^2+\gamma det(Z)\}$$

If $Z = I$, the difference is $-2(2+\gamma)\epsilon^2$. Because $0 < \gamma$, the difference is negative, showing that there exists a point in every neighborhood of $B_s = 0$ where the value of the metric at that point is less than the value at $B_s = 0$. Thus, $B_s = 0$ can never be a local minimum.

If instead $Z = diag(1,-1)$, then the difference is $-2(2-\gamma)\epsilon^2$, which is positive provided $2 < \gamma$. Therefore, $B_s = 0$ is a saddle point when $2 < \gamma$ because there exists points in every ϵ-neighborhood for which the difference is positive ($Z = P$) and points for which the difference is negative ($Z = I$).

Before considering the case $\gamma \leq 2$, note that $|Z|^2 + 2\,det(Z) = (Z_{11}+Z_{22})^2 + (Z_{12}-Z_{21})^2 \geq 0$. Also, $|Z|^2 - 2\,det(Z) = (Z_{11}-Z_{22})^2 + (Z_{12}+Z_{21})^2 \geq 0$, and therefore, $|Z|^2 - 2\,|det(Z)| \geq 0$ for all Z.

Now, if $\gamma \leq 2$, then, for all Z, $|Z|^2 + \gamma det(Z) \geq |Z|^2 - \gamma|det(Z)| \geq |Z|^2 - 2|det(Z)| \geq 0$. Thus, the difference (5) is less than or equal to zero. It has already been shown that for $Z = I$, the difference is strictly negative. Therefore, $B_s = 0$ is a local maximum when $\gamma \leq 2$.

Second, consider the stationary point in b', i.e., $B_s = R$. Then the difference (5) becomes

$$\Delta\mu_2^{(o+)}(R;Z) = \epsilon^2\{|Y|^2 + (tr[(adjZ)R])^2\}$$

and is therefore non-negative for all Z. Hence, $B = R$ is a minimum, as expected.

Third, consider the *real* stationary point in b'', i.e., $B_s = tF$ with $t = \sqrt{\frac{2-\gamma}{2+\gamma}}$ and $\gamma \leq 2$. Then the difference (5) becomes

$$\Delta\mu_2^{(o+)}(tF;Z) = -\epsilon t\left(\frac{4\gamma}{2+\gamma}\right) tr[(adjZ)F]$$

For $Z = F$ and $\gamma \neq 2$, the difference (5) is positive, while for $Z = -F$, the difference is negative. Thus, $B_s = tF$ is a *real* saddle point for $\gamma < 2$.

Fourth, consider the non-orthogonal *real* stationary points in c', with $0 < \gamma \leq \frac{2}{3}$. Then the difference (5) becomes

$$\begin{aligned} \Delta\mu_2^{(o+)}(UDV^t; Z) &= 2\epsilon\{tr[(H^tH - I)((Z^tUH) + (Z^tUH)^t)] \\ &+ \gamma(\beta_s - 1)tr[(adjZ)UH]\} \end{aligned}$$

where $H = DV^t$. These stationary points appear to be a saddle since if the above difference is positive when $Z = Z_1$, then choosing $Z = -Z_1$ makes the difference negative.

In summary, the o+ metric considered in the section satisfies Properties 1-7 on M_2. Property 8 is not satisfied since $B = 0$ is a stationary point for all values of γ. When $2 < \gamma$, the point $B = 0$ is a saddle, but when $0 < \gamma \leq 2$, the point is a local maximum. Moreover, when $\gamma < 2$, there exist other stationary points that are saddles. When $0 < \gamma \leq \frac{2}{3}$, some of the stationary points are non-orthogonal.

3.2.3 Three other o+ metrics, with parameter γ.

3.2.2.1
The analysis of the following o+ metric

$$\mu_2^{(o+)}(B) = |B - (adjB)^t|^2 + \gamma(\beta - 1)^2 \tag{6}$$

with $\gamma > 0$ is much simpler than the previous. The global minimizers are $B = \pm R$, i.e., the rotations. Properties 1-7 are satisfied on M_2. The stationary point equation is

$$B = \left[1 - \frac{\gamma}{2}(\beta - 1)\right](adjB)^t \tag{7}$$

Using techniques similar to those in the previous sections, one finds that the only real solutions to the SPE are $B = 0$ or $B = \pm R$. Property 8 is not satisfied since $B = 0$ is not a global minimizer. Note that the stationary points do not depend on the value chosen for γ in this metric.

The behavior of the metric at the stationary point $B = 0$ is examined next. Letting $B = B_s + \epsilon Z$, one finds for $B_s = 0$, the difference to order ϵ^2 is

$$\Delta\mu_2^{(o+)}(0; Z) = 2\epsilon^2\{|Z|^2 - (2 + \gamma)det(Z)\}$$

When $Z = I$, the difference is negative, while for $Z = F$, the difference is positive. Thus, $B = 0$ is a saddle point for $0 < \gamma$.

3.2.2.2
Another o+ metric having $B = 0$ as a stationary point, but not a global minimizer, is

$$\mu_2^{(o+)}(B) \quad = \quad |B - (adjB)^t|^2 + \gamma|B^tB - I|^2 \tag{8}$$

with $\gamma > 0$. Properties 1-7 are satisfied on M_3. The SPE is

$$(1-\gamma)B = (adjB)^t - \gamma BB^tB \tag{9}$$

Clearly, $B = 0$ is a stationary point. To find others, let $B = U\Delta V^t$; then the SPE becomes

$$(1-\gamma)\Delta = (det\, UV)(adj\Delta) - \gamma\Delta^3$$

The *seven* algebraic solutions to this equation are

a. $\Delta = 0$,

b. $\Delta = \pm I$ with $det(UV) = 1$,

c. $\Delta = \pm tI$ with $t = \sqrt{\frac{\gamma-2}{\gamma}}$ and $det(UV) = -1$,

d. $\Delta = \pm tP$ with $t = \sqrt{\frac{\gamma-2}{\gamma}}$ and $det(UV) = +1$,

e. $\Delta = \pm P$ with $det(UV) = -1$,

f. $\Delta = diag\left(\sqrt{\frac{p+t}{2}}, -\sqrt{\frac{p-t}{2}}\right)$ with $p = \frac{\gamma-1}{\gamma}$, $t = \sqrt{p^2 - \frac{4}{\gamma^2}}$, and $det(UV) = 1$,

g. $\Delta = diag\left(\sqrt{\frac{p+t}{2}}, \sqrt{\frac{p-t}{2}}\right)$ with $p = \frac{\gamma-1}{\gamma}$, $t = \sqrt{p^2 - \frac{4}{\gamma^2}}$, and $det(UV) = -1$,

However, since it is required that $\Delta_{ii} \geq 0$ for $i = 1, 2$, a number of these can be ruled out, leaving just four potential solutions

a'. $\Delta = 0$,

b'. $\Delta = I$ with $det(UV) = 1$,

c'. $\Delta = tI$ with $t = \sqrt{\frac{\gamma-2}{\gamma}}$ and $det(UV) = -1$,

g'. $\Delta = diag\left(\sqrt{\frac{p+t}{2}}, \sqrt{\frac{p-t}{2}}\right)$ with $p = \frac{\gamma-1}{\gamma}$, $t = \sqrt{p^2 - \frac{4}{\gamma^2}}$, and $det(UV) = -1$,

Solution c' can be eliminated by choosing $\gamma < 2$ because then t is imaginary. Solution g' can be eliminated by choosing $\gamma < 3$ because then t is imaginary. Choosing $\gamma < 2$ eliminates both.

In terms of B, the *real* stationary points of the metric (8) are thus[12]

a''. $B = 0$, for all γ,

b''. $B = R$, for all γ,

c''. $B = tF$ with $t = \sqrt{\frac{\gamma-2}{\gamma}}$, $\gamma > 2$, and $det(UV) = -1$,

g''. $B = U\, diag\left(\sqrt{\frac{p+t}{2}}, \sqrt{\frac{p-t}{2}}\right) V^t$ with $p = \frac{\gamma-1}{\gamma}$, $t = \sqrt{p^2 - \frac{4}{\gamma^2}}$, $\gamma > 3$, and $det(UV) = -1$,

The Stationary Point a''.
Since $\mu_2^{(o+)}(0) = 2\gamma > 0$, the point $B = 0$ is not a global minimizer of the metric, even though it is a stationary point. To determine the nature of this point, let $B = B_s + \epsilon Z = \epsilon Z$. To order ϵ^2, the difference between the metric value at the stationary point and a nearby point is

$$\Delta\mu_2^{(o+)}(0 : Z) = 2\epsilon^2 \left[(1-\gamma)|Z|^2 - 2det(Z)\right]$$

If $Z = I$, the difference is $-4\gamma\epsilon^2$, which is negative for $0 < \gamma$. Therefore, the point $B = 0$ is not a local minimum.

Proposition 4.
If $\gamma < 2$, then $B = 0$ is a saddle point.
Proof.
If $Z = F$, the difference above becomes $4\epsilon^2(2-\gamma)$, which is positive when $\gamma < 2$. But the difference is negative for $Z = I$. Thus, for $\gamma < 2$, there exists points in every ϵ-neighborhood of $B = 0$ for which the difference is positive and for which the difference is negative, and therefore the point $B = 0$ is a saddle. §

Proposition 5.
If $2 \leq \gamma$, then $B = 0$ is a local maximum of the metric.
Proof.
Suppose $2 \leq \gamma$. Then

$$\begin{aligned} 1 - \gamma &\leq -1 \\ (1-\gamma)|Z|^2 &\leq -|Z|^2 \\ (1-\gamma)|Z|^2 - 2det(Z) &\leq -(|Z|^2 + 2det(Z)) \end{aligned}$$

[12]Note that if B_s is a solution to the SPE (9), then so is -B_s.

But the right-hand-side of the last line is non-positive for all Z. Thus, the difference is non-positive, and $\mu_2^{(o+)}(\epsilon Z) \leq \mu_2^{(o+)}(0)$. Since the difference is strictly negative when $Z = I$, $B = 0$ is a local maximum. §

The Stationary Point b''.
For $B_s = R$, the difference is

$$\Delta\mu_2^{(o+)}(R; Z) \quad = \quad 2\epsilon^2\{(1+\gamma)\,[\,|Z|^2 - 2det(Z)\,] + \gamma\,[\,tr(Z^tR)\,]^2\}$$

Thus, the difference is non-negative, and (as was stated earlier), the stationary point is a global minimum.

The Stationary Point c''.
Using the techniques described herein, one can show that the stationary point c'' gives

$$\Delta\mu_2^{(o+)}(tF; Z) = 8t\epsilon\, tr(Z^tF) + \Theta(\epsilon^2)$$

Suppose the difference is non-negative for a particular Z; then it will be non-positive for -Z. Thus, the stationary point is a saddle.

The Stationary Point g''.
Using the techniques described herein, one can show that for any stationary point of this metric,

$$\Delta\mu_2^{(o+)}(B_s; Z) = 4\epsilon\, tr\{Z^t[B_s - (adjB_s)^t]\} + \Theta(\epsilon^2)$$

For the stationary point in g'', $B_s \neq (adjB_s)^t$ holds unless $t = 0$ and $det(UV) = 1$. But $det(UV) = -1$ for this stationary point and thus the order ϵ term in the difference above is non-zero. Suppose the difference is non-negative for a particular Z; then it will be non-positive for -Z. Thus, the stationary point g'' is a saddle.

In summary, the o+ metric (8) has $B = R$ as the global minimizer and satisfies Properties 1-7 on M_2. It has four stationary points, $B = 0$, $B = R$, $B = tF$, and $B = U\Delta V^t$. The point $B = 0$ is a saddle when $\gamma > 2$ and is a local maximum when $\gamma \leq 2$. In addition, if $\gamma \leq 2$, then the stationary point $B = tF$ does not exist; otherwise it is a saddle point. If $\gamma \leq 3$, then the stationary point $B = UDV^t$ does not exist; otherwise it is a saddle point. For this metric, the best choice of the parameter γ is $\gamma \leq 2$ since that eliminates two of the saddle points.

3.2.2.3
Yet one more o+ metric is considered, namely

$$\mu_2^{(o+)}(B) = |B^tB - \beta I|^2 + \gamma(\beta - 1)^2 \tag{10}$$

with $\gamma > 0$. This metric is zero if and only if $B = \pm R$, making it an o+ metric. The metric satisfies Properties 1-7 on M_2. The Stationary Point Equation is

$$\beta B = \left\{ \left(1 + \frac{\gamma}{2}\right) \beta - \frac{1}{2} \left(\gamma + |B|^2\right) \right\} (adjB)^t + BB^t B \tag{11}$$

The only solutions of the form $B = tU$ are $B = 0$ and $B = \pm R$. Trying $B = U\Delta V^t$ reveals that these are the only *real* stationary points.

The nature of the stationary point $B = 0$ is investigated using

$$\Delta\mu_2^{(o+)}(0; Z) \quad = \quad -2\gamma(detZ)\,\epsilon^2 + \Theta(\epsilon^4)$$

For $Z = I$, the difference is negative, while for $Z = P$, the difference is positive. Therefore, $B = 0$ is a saddle point for all γ.

All of the o+ metrics in Sections 3.2.2 and 3.2.3 have $B = 0$ as a stationary point. The point is not a global minimizer and thus Property 8 is not satisfied. In section 3.2.4 an o+ metric is devised which does not have $B = 0$ as a stationary point.

3.2.4 An o+ Metric with $B = 0$ not a stationary point

Let $B \neq 0$ and define

$$\mu_2^{(o+)}(B) \quad = \quad \frac{|B - (adjB)^t|^2}{2|B|^2} + \gamma(\beta - 1)^2 \tag{12}$$

$$= \quad \left(1 - \frac{2\beta}{|B|^2}\right) + \gamma(\beta - 1)^2 \tag{13}$$

The global minimizers are $B = R$. Suppose $B = \epsilon Z$ with $Z \neq 0$. Then

$$\lim_{\epsilon \to 0} \frac{2\beta}{|B|^2} \quad = \quad \frac{2\,det(Z)}{|Z|^2}$$

The metric is therefore not continuous at $B = 0$ since the limit depends on the choice of Z. The metric thus fails to satisfy Property 2 if one were to choose M_2 as the domain.

The metric is differentiable everywhere except $B = 0$, and the SPE is

$$\frac{2\beta}{|B|^4} B \quad = \quad \left\{ \frac{1}{|B|^2} - \gamma(\beta - 1) \right\} (adjB)^t \tag{14}$$

For $0 < \gamma$, the only solutions are $B = R$.

3.2.5 A parameter-free o+ Metric

The goal in this section is to devise a parameter-free o+ metric that prevents $B = 0$ from being a stationary point. Doing this requires that the metric contain a barrier, not against inversion ($\beta = 0$), but against zero column lengths.

Define the diagonal matrix $D = diag(|\mathbf{b}_1|, |\mathbf{b}_2|)$. Then D^{-1} exists provided $|\mathbf{b}_1| \neq 0$ and $|\mathbf{b}_2| \neq 0$, i.e., if $|B| \neq 0$. On the other hand, if either $|\mathbf{b}_1| = 0$ or $|\mathbf{b}_2| = 0$, then D^{-1} does not exist. Let $\mathcal{D}'$ be the set of matrices for which D^{-1} does not exist, i.e., those having one or more column vector lengths equal to zero. Note that $\mathcal{D}'$ is a subset of the set of singular matrices on M_2.

Proposition 6.
Let D^{-1} exist. Then the matrix $B \in M_2$ is a rotation if and only if $B = (adj[BD^{-1}])^t$.
Proof.
If $B = R$ is a rotation, then $D = I$ and $(adj[BD^{-1}])^t = (adjR)^t = R$, thus proving the first part of the assertion. Now suppose that $B = (adj[BD^{-1}])^t$. These relations can be expressed in terms of the column vectors of B

$$\begin{aligned} \mathbf{b}_1 &= -\mathbf{b}_2^{\perp}/|\mathbf{b}_2| \\ \mathbf{b}_2 &= +\mathbf{b}_1^{\perp}/|\mathbf{b}_1| \end{aligned}$$

Therefore, $|\mathbf{b}_1| = |\mathbf{b}_2| = 1$, and $\mathbf{b}_1 \cdot \mathbf{b}_2 = 0$. Thus $B^t B = I$. Next, $\beta = \mathbf{b}_2 \cdot \mathbf{b}_1^{\perp} = 1$. This proves the second part of the assertion. §

The Proposition suggests the following o+ metric, defined on $\mathcal{D} = M_2 - \mathcal{D}'$.

$$\begin{aligned} \mu_2^{(o+)}(B) &= |B - [adj(BD^{-1})]^t|^2 \qquad (15) \\ &= |B|^2 - 2\beta \left(\frac{1}{|\mathbf{b}_1|} + \frac{1}{|\mathbf{b}_2|} \right) + 2 \qquad (16) \end{aligned}$$

The metric is undefined at points in $\mathcal{D}'$; moreover, the metric is discontinuous there because if $B^* \in \mathcal{D}'$, $\lim_{B \to B^*} \mu_2^{(o+)}$ can differ, depending on how B^* is approached.[13] The discontinuity in this metric is more serious than the discontinuity in the metric (12) because here it occurs at more than one point.

Properties 1-7 are satisfied on $\mathcal{D}$.

To calculate the stationary point equation, we use the following facts:

$$\frac{d}{dB}|\mathbf{b}_1| = [\frac{\mathbf{b}_1}{|\mathbf{b}_1|}, \mathbf{0}]$$

[13] For example, Take $B^* = diag(1, 0) \in \mathcal{D}'$ and $B = diag(1, y) \in \mathcal{D}$. Then $\mu_2^{(o+)}(B) = 3 + y^2 - 2y - 2y/|y|$. When $y > 0$, the limit of the metric as $y \to 0$ is 1, but when $y < 0$, the limit is 5.

$$\frac{d}{dB}|\mathbf{b}_2| \quad = \quad [\mathbf{0}, \frac{\mathbf{b}_2}{|\mathbf{b}_2|}]$$

Then the stationary point equation (SPE) on $\mathcal{D}$ is readily calculated to be:

$$B + \beta B D^{-3} - tr(D^{-1})(adjB)^t \quad = \quad 0 \tag{17}$$

Solutions such as $B = 0$ are automatically excluded because D^{-1} does not exist there.

Proposition 7.
The matrix B is a solution to the SPE on $\mathcal{D}$ if and only if B is a rotation.
Proof.
If $B = R$, then $(adjB)^t = R$ and $D = I$. The left-hand-side of the SPE is then zero. On the other hand, suppose (17) holds. In column form, this is

$$\begin{aligned}\left(1 + \frac{\beta}{|\mathbf{b}_1|^3}\right)\mathbf{b}_1 &= -tr(D^{-1})\mathbf{b}_2^\perp \\ \left(1 + \frac{\beta}{|\mathbf{b}_2|^3}\right)\mathbf{b}_2 &= +tr(D^{-1})\mathbf{b}_1^\perp\end{aligned}$$

Therefore, $\mathbf{b}_1 \cdot \mathbf{b}_2 = 0$, and $\beta = |\mathbf{b}_1||\mathbf{b}_2|^2 = |\mathbf{b}_1|^2|\mathbf{b}_2|$. Therefore $|\mathbf{b}_1| = |\mathbf{b}_2|$ and $\beta = |\mathbf{b}_1|^3 = |\mathbf{b}_2|^3$. Substituting these results back into the column relations gives $\beta = 1$. §

Proposition 7 shows that Property 8 is satisfied by the metric in (15) when defined on $\mathcal{D}$.

From a practical point of view, restricting the domain to exclude matrices in $\mathcal{D}'$ may not too much of a problem because if the mesh contains a point such that $B \in \mathcal{D}'$, it can always be randomly perturbed slightly so that the perturbed matrix is in $\mathcal{D}$. However, the lack of continuity on M_2 may turn out to be important in practice. So, even though this metric and the one in 3.2.4 technically satisfy the Eight Properties on a restricted domain, something more satisfying is sought in the next section.

3.3 The Rotation-based o+ Metric

To improve upon the previous metric, we seek a metric of the form $|B - R|^2$ so that one immediately obtains rotations as the only global minimizers. Such a metric would seem to require that one specify R to be a particular rotation; but that would defeat the purpose, which is to have B be an arbitrary rotation. To avoid this problem, it is required that the rotation is a matrix function $R = R(B)$.

3.3.1 Derivation and analysis

Fortunately, such a function exists. First, define the scalar function

$$\psi(B) \quad = \quad \sqrt{|B|^2 + 2\beta} \tag{18}$$

From Proposition 1, it is clear that ψ is a real, non-negative number for any $B \in M_2$. Further, $\psi = 0$ if and only if $B = tF$. Note that ψ is a continuous function of B.

Proposition 8.
For $\psi \neq 0$, the matrix

$$R(B) \quad = \quad \frac{B + (adjB)^t}{\psi(B)} \tag{19}$$

is a rotation.
Proof.
The components of the matrix are $R_{11} = \frac{B_{11}+B_{22}}{\psi(B)} = R_{22}$ and $R_{21} = \frac{B_{21}-B_{12}}{\psi(B)} = -R_{21}$. Thus, $R_{11}^2 + R_{21}^2 = 1$, $R_{12}^2 + R_{22}^2 = 1$, and $R_{11}R_{12} + R_{21}R_{22} = 0$. So $R^tR = I$. Furthermore, $det(R) = R_{11}R_{22} - R_{12}R_{21} = 1$. §

When $\psi(B) = 0$, we have $B = tF$. Consider the matrix $R(B)$ in an ϵ-neighborhood of tF.

$$\begin{aligned}
\lim_{B \to tF} R(B) &= \lim_{\epsilon \to 0} R\left(tF + \epsilon Z\right) \\
&= \lim_{\epsilon \to 0} \frac{[tF + \epsilon Z] + [adj\,(tF + \epsilon Z)]^t}{\psi(tF + \epsilon Z)} \\
&= \lim_{\epsilon \to 0} \frac{[tF + (adj\,tF)^t] + \epsilon\,[Z + (adjZ)^t]}{\epsilon\psi(Z)} \\
&= \lim_{\epsilon \to 0} \frac{\epsilon\,[Z + (adjZ)^t]}{\epsilon\psi(Z)} \\
&= \lim_{\epsilon \to 0} \frac{[Z + (adjZ)^t]}{\psi(Z)} \\
&= R(Z)
\end{aligned}$$

Although the limit exists, it depends on Z. Thus $R(B)$ is undefined for $B = tF$ because it is discontinuous there.

A new non-barrier o+ metric is defined in terms of $\psi(B)$

$$\mu_2^{(o+)}(B) \quad = \quad |B|^2 - 2\,\psi(B) + 2 \tag{20}$$

The metric is defined on $\mathcal{D} = M_2$. Properties 1-2 are satisfied.

Proposition 9.
The metric above is non-negative.
Proof.
First, we note that $|B|^2 - 2\beta = (B_{11} - B_{22})^2 + (B_{12} + B_{21})^2 \geq 0$ for any B. Thus $8\beta \leq 4|B|^2$. Second, from $(2 - |B|^2)^2 \geq 0$ we obtain $4|B|^2 \leq 4 + |B|^4$. Putting the two together,

$$\begin{aligned} 8\beta &\leq 4 + |B|^4 \\ 4|B|^2 + 8\beta &\leq 4 + 4|B|^2 + |B|^4 \\ 4\,\psi^2 &\leq (2 + |B|^2)^2 \\ 2\,\psi &\leq 2 + |B|^2 \\ 0 &\leq 2 - 2\,\psi + |B|^2 \\ 0 &\leq \mu_2^{(o+)} \end{aligned}$$

§

Proposition 9 shows that Property 3 is satisfied. The next chore is to find the global minima of the metric.

Proposition 10.
The global minima of the metric is the set of rotations.
Proof.
First, suppose $B = R$. Then $|R|^2 = 2$, $\psi(R) = 2$, and thus $\mu_2^{(o+)}(R) = 0$. Second, suppose $|B|^2 - 2\psi + 2 = 0$. The following relation can be verified

$$\psi^2 \mu_2^{(o+)} = |(\psi - 1)B - (adjB)^t|^2 \tag{21}$$

Therefore, when the metric value is zero, we must have $(\psi - 1)B = (adjB)^t$, i.e., $\psi B = (B + adjB)^t$. If $\psi = 0$, then $B + (adjB)^t = 0$, which requires that B be a scaled-flip. But the value of the metric for scaled-flips is greater than zero, so scaled-flips cannot be global minimizers. If $\psi \neq 0$, the relation becomes $B = (B + adjB^t)/\psi$. The right-hand side has already been shown to be a rotation. Thus we have shown $\mu_2^{(o+)}(B) = 0$ if and only if $B = R$. §

Proposition 10 shows that Properties 4-6 are satisfied by this metric. Note that the proof of the proposition shows that we can write

$$\mu_2^{(o+)}(B) \quad = \quad |B - R(B)|^2 \tag{22}$$

provided $\psi(B) \neq 0$. Recall that $M_2^{(so-)}$ is the set of scaled flips. Let $\mathcal{D}^* = M_2 - M_2^{(so-)}$. Then the form of the metric in (22) holds on $\mathcal{D}^*$.

Lastly, we ask whether Properties 7-8 are satisfied.

Proposition 11.
The metric (20) is differentiable on $\mathcal{D}^*$. Moreover, the stationary points of the metric are rotations.

Proof.
From $\psi^2 = |B|^2 + 2\beta$ we obtain $\psi \frac{d\psi}{dB} = B + (adjB)^t$. Then from the discussion after Proposition 8, $d\psi/dB$ exists provided $\psi \neq 0$, i.e., ψ is differentiable on $\mathcal{D}^*$ (and is not differentiable on $M_2^{(so-)}$). On $\mathcal{D}^*$, $\psi \neq 0$, and thus from (19),

$$\frac{d\psi}{dB} = R(B) \tag{23}$$

and thus

$$\frac{1}{2}\frac{d}{dB}\mu_2^{(o+)} \quad = \quad B - R(B)$$

Therefore, the stationary points of the metric (20) are rotations. §

This set of propositions shows that the stationary points and global minima of the metric coincide, so Property 8 is satisfied.

Classification of the Critical Points
Let $B = Y + \epsilon Z$ with $Y, Z \in M_2$ and $\epsilon > 0$. Then, for any point $Y \neq tF$,

$$\psi(B) \quad = \quad \psi(Y) + \epsilon \left\{Z \cdot R(Y)\right\} + \epsilon^2 \left\{ \frac{\psi^2(Z) - \left[Z \cdot R(Y)\right]^2}{2\psi(Y)} \right\} + \Theta(\epsilon^3)$$

Using the previous, the values of the metric in an ϵ-neighborhood of Y are

$$\begin{aligned} \mu_2^{(o+)}(B) - \mu_2^{(o+)}(Y) &= \left\{|B|^2 - |Y|^2\right\} - 2\left\{\psi(B) - \psi(Y)\right\} \\ &= \epsilon^2 \left\{ |Z|^2 - \frac{\psi^2(Z) - (Z \cdot Y)^2}{\psi(Y)} \right\} + \Theta(\epsilon^3) \end{aligned}$$

Then, at the stationary points $Y = R(Y)$, one has $\psi(Y) = 2$, so that

$$\begin{aligned} \Delta\mu_2^{(o+)}(Y; Z) &= \frac{1}{2}\epsilon^2 \left\{2|Z|^2 - \psi^2(Z) + (Z \cdot Y)^2\right\} \\ &= \frac{1}{2}\epsilon^2 \left\{2|Z|^2 - \left(|Z|^2 + 2det(Z)\right) + (Z \cdot Y)^2\right\} \\ &= \frac{1}{2}\epsilon^2 \left\{\left(|Z|^2 - 2det(Z)\right) + (Z \cdot Y)^2\right\} \end{aligned}$$

Each of the two terms on the right-hand-side of the above expression are non-negative for all Z and thus the value of the metric at the stationary points is a minimum. This agrees with previous observation that rotations are global minimizers.

Now consider the difference at the non-differentiable point $Y = tF$. Then the formulas above do not apply; instead, $\psi(tF + \epsilon Z) = \epsilon\,\psi(Z)$, and

$$\Delta\mu_2^{(o+)}(tF; Z) \quad = \quad 2\epsilon\,[t\,(Z \cdot F) - \psi(Z)] + \Theta(\epsilon^2)$$

If $t = 0$, then the difference is negative for all Z; thus the point $B = 0$ is a local maximum. Property Seven is therefore *not* satisfied. If $t \neq 0$, then for $Z = I$, the right-hand-side is -4ϵ, which is negative. For $Z = tF$, the right-hand-side is $2\epsilon|tF|^2$, which is non-negative. Therefore, excluding $B = 0$, the points of non-differentiability are neither maxima nor minima.

3.4 Additional Observations on the 2D Rotation Form

For added insight into the $d = 2$ non-barrier metrics, we present more results related to the functions ψ and $R(B)$.

3.4.1 The function $\psi(B)$ is a matrix inner product.

Note that (22) can be expanded to give

$$\begin{aligned} \mu_2^{(o+)}(B) &= |B - R(B)|^2 \\ &= |B|^2 - 2B \cdot R(B) + 2 \end{aligned}$$

Comparing this to (20) suggests that

$$\psi(B) = B \cdot R(B) \tag{24}$$

When $B \in \mathcal{D}^*$, this fact can be derived directly from (19).

3.4.2 The so+ metric in terms of B minus a scaled rotation.

For $B \in \mathcal{D}^*$, the matrix $\frac{\psi}{2}R(B)$ is a scaled-rotation. Then

$$\begin{aligned} B - \frac{\psi}{2}R(B) &= B - \frac{1}{2}\left(B + [adjB]^t\right) \\ &= \frac{1}{2}\left(B - [adjB]^t\right) \end{aligned}$$

Therefore, when $B \in \mathcal{D}^*$, the metric $\mu_2^{(so+)}$ in (1) can be written as

$$\frac{1}{2}\mu_2^{(so+)}(B) \quad = \quad |B - \frac{\psi}{2}R(B)|^2 \tag{25}$$

With this form, one directly sees that the so+ metric has scaled-rotations as the global minimizers.

3.4.3 A derivation of $\psi(B)$ based on rotations.

An interesting derivation of the expression (18) for ψ is given, starting with the fact that $\psi = B \cdot R(B)$. Recall that every rotation in M_2 can be written as

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{26}$$

with $0 \leq \theta < 2\pi$. Then

$$\begin{aligned} \psi(B) &= B \cdot R \\ &= (trB)\cos\theta + (B_{21} - B_{12})\sin\theta \end{aligned} \tag{27}$$

Thus, θ is implicitly a function of B. To find $\theta = \theta(B)$, we impose the requirement that the derivative of $\psi = \psi(B, \theta)$ with respect to B is equal to R. Using the chain rule, we have

$$\frac{d\psi}{dB} = \frac{\partial\psi}{\partial B} + \frac{\partial\psi}{\partial\cos\theta}\frac{\partial\cos\theta}{\partial B} + \frac{\partial\psi}{\partial\sin\theta}\frac{\partial\sin\theta}{\partial B}$$

An explicit calculation of these terms shows that,

$$\frac{d\psi}{dB} = R + (trB)\frac{\partial}{\partial B}\cos\theta + (B_{21} - B_{12})\frac{\partial}{\partial B}\sin\theta$$

But since the left-hand-side is required to equal R. $\theta(B)$ must satisfy

$$(trB)\frac{\partial}{\partial B}\cos\theta + (B_{21} - B_{12})\frac{\partial}{\partial B}\sin\theta = 0 \tag{28}$$

for all $B \in \mathcal{D}^*$. Because $\sin^2\theta = 1 - \cos^2\theta$, the following holds identically

$$\sin\theta\frac{\partial}{\partial B}\sin\theta + \cos\theta\frac{\partial}{\partial B}\cos\theta = 0$$

Multiply (28) by $\cos\theta$ to find

$$\begin{aligned} (trB)\cos\theta\frac{\partial}{\partial B}\cos\theta + (B_{21} - B_{12})\cos\theta\frac{\partial}{\partial B}\sin\theta &= 0 \\ -(trB)\sin\theta\frac{\partial}{\partial B}\sin\theta + (B_{21} - B_{12})\cos\theta\frac{\partial}{\partial B}\sin\theta &= 0 \\ [(B_{21} - B_{12})\cos\theta - (trB)\sin\theta]\frac{\partial}{\partial B}\sin\theta &= 0 \end{aligned}$$

We are not interested in the case where θ (and thus $\sin\theta$) are independent of B, so θ must satisfy

$$(B_{21} - B_{12})\cos\theta - (trB)\sin\theta = 0$$

The solution to this equation takes the form $\cos\theta = (trB)/\Delta$, $\sin\theta = (B_{21} - B_{12})/\Delta$, where

$$\begin{aligned} \Delta &= \sqrt{(trB)^2 + (B_{21} - B_{12})^2} \\ &= \sqrt{|B|^2 + 2\beta} \end{aligned}$$

is non-zero.

Substitution of the expressions for $\cos\theta(B)$ and $\sin\theta(B)$ into the expression (27) gives

$$\begin{aligned} \psi(B) &= (trB)\cos\theta + (B_{21} - B_{12})\sin\theta \\ &= (trB)^2/\Delta + (B_{21} - B_{12})^2/\Delta \\ &= \left[(trB)^2 + (B_{21} - B_{12})^2\right]/\Delta \\ &= \Delta^2/\Delta \\ &= \Delta \end{aligned}$$

This result is in agreement with (18) and explains why the derivative of $\psi(B)$ is a rotation.

3.4.4 A metric to produce flips.

As a closing comment, note that one can also devise metrics whose global minima are flips. Let $\psi_-(B) = \sqrt{|B|^2 - 2\beta}$ and

$$\begin{aligned} \mu_2^{(o-)} &= |B|^2 - 2\psi_- + 2 \\ &= |B - F(B)|^2 \end{aligned} \tag{29}$$

with $F(B) = \frac{B - (adjB)^t}{\psi_-(B)}$. Analysis of this metric is similar to that in Section 3.3.

3.5 Summary of the Metrics and Their Properties

For convenience, Table 1 summarizes the metrics described in this section, along with their mathematical properties.

Table 1: Summary of Metrics and Their Mathematical Properties

Metric	Formula	Comments
S1	$\frac{1}{2}\|B-(adjB)^t\|^2$	Satisfies all Eight Properties on M_2.
S2	$\|B^tB-\beta I\|^2$	Unwanted saddle points $B=tF$.
SS1	$\|B^tB-I\|^2$	Spurious global minima. Unwanted local maximum.
SS2	$\|B^tB-I\|^2+\gamma(\beta-1)^2$	Unwanted saddle point $B=0$ when $\gamma>2$. Unwanted local maximum $B=0$ when $\gamma\leq 2$. Other non-orthogonal SP's when $\gamma\leq\frac{2}{3}$.
SS3	$\|B-(adjB)^t\|^2+\gamma(\beta-1)^2$	Unwanted saddle at $B=0$.
SS4	$\|B-(adjB)^t\|^2+\gamma\|B^tB-I\|^2$	Local maximum $B=0$ when $\gamma\leq 2$. Unwanted saddles $B=tF$ when $2\leq\gamma\leq 3$. Additional saddle points when $\gamma>3$.
SS6	$\|B^tB-\beta I\|^2+\gamma(\beta-1)^2$	Unwanted saddle point $B=0$ for all γ.
SS7	$\|B^tB-(adjB)^t\|^2/2\|B\|^2+\gamma(\beta-1)^2$	Discontinuous at $B=0$.
SS5	$\|B\|^2-2\beta tr(D^{-1})+2$	Discontinuous on $\mathcal{D}'$.
SS0	$\|B\|^2-2\psi+2$	Non-differentiable at $B=tF$, with $B=0$ a local maximum.

4 Numerical Results

In this section optimal meshes are computed using the so+ and o+ metrics discussed in the previous section. The main objective is to determine the practical effect of a metric failing to satisfy one or more of the Eight Properties investigated in this work. A secondary objective is to illustrate that some of the metrics do indeed perform well on a realistic problem.

The methodology for these experiments is to numerically compute the optimal mesh for a given local metric. The optimal mesh is examined for defects such as tangling. Several outcomes of each experiment are possible:

1. No defect is observed and
    a. The metric satisfies all Eight Properties, or
    b. The metric satisfies all but one of the Eight Properties, or
    c. The metric fails to satisfy more than one of the Eight Properties.
2. A defect is observed and
    a. The metric satisfies all Eight Properties, or
    b. The metric satisfies all but one of the Eight Properties, or
    c. The metric fails to satisfy more than one of the Eight Properties.

The conclusion for each of the cases given above is expressed in terms of whether or not the properties are sufficient or necessary to avoid a mesh defect.

1. No defect is observed
    a. *Collectively, the Eight Properties may be sufficient.*
    b. *The property failed may be sufficient, but not necessary.*
    c. *None of the failed properties may be necessary.*
2. A defect is observed and
    a. *Collectively, the Eight Properties may not be sufficient.*
    b. *The property failed may be necessary.*
    c. *At least one of the failed properties may be necessary.*

We use the words '*may be*' in these conclusions because perhaps the same metric on a different mesh and domain would reveal a defect where none was found in these experiments. Further, if a defect is found in these experiments, it could possibly be attributed to something else besides a failure to satisfy a mathematical property. For example, a solver designed for a smooth objective function

could perhaps create a mesh defect if it was applied to a non-differentiable metric. In such a case, the correct conclusion of the experiment might be that the wrong solver has been used, and not that the mathematical condition is necessary. Therefore, conclusions of the above types may only be valid for numerical optimization solvers that are similar to the one employed in these experiments. These experiments used Mesquite's Quasi-Newton solver, which is designed to find local minima of *differentiable* functions.

Finally, mesh defects might arise from a poor implementation of a mathematically sound metric. To minimize this possibility, the metrics were implemented with some care to avoid numerical roundoff, divisions by zero, and the like. In spite of the limitations noted, the experiments have proved to be fairly illuminating. See Table 1 for a summary of the metrics used in the experiments described in this section.

4.1 Experiments using the Deforming Airfoil

In this problem there is first an un-deformed airfoil having a high quality mesh upon it (see top row of Figure 1). The airfoil then deforms via motion of the inner domain boundary, and the mesh vertexes on the boundary move with it. This produces the tangled meshes shown in the bottom row of Figure 1. The optimization improves the quality of the tangled mesh on the deformed airfoil using a reference mesh consisting of the high-quality mesh on the undeformed airfoil to calculate the target matrices. The initial mesh has 13725 elements, 54 of which are inverted. This set up is run with each of the different so+ and o+ quality metrics discussed in Section 3.

In Figure 2, the results of optimizing with the S1 (left) and S2 (right) metrics are shown on the leading (top) and trailing (bottom) edges of the airfoil. The results from S1 and S2 appear to be nearly identical, even though the latter metric fails to satisfy Property 8, having unwanted saddle points at $B = tF$. Because the S1 metric has no unwanted saddle points, the result of this comparison suggests that the absence of saddle points is only a sufficient, but not necessary condition for a well-performing metric. As a side point, note that the boundary layer thickness of the meshes in this figure are thicker than those in the reference mesh in the top row of Figure 1. This is attributed to the fact that the metrics S1 and S2 are designed to preserve the shape of the elements in the reference mesh, but not the size. To preserve both shape and size, one needs the o+ metrics.

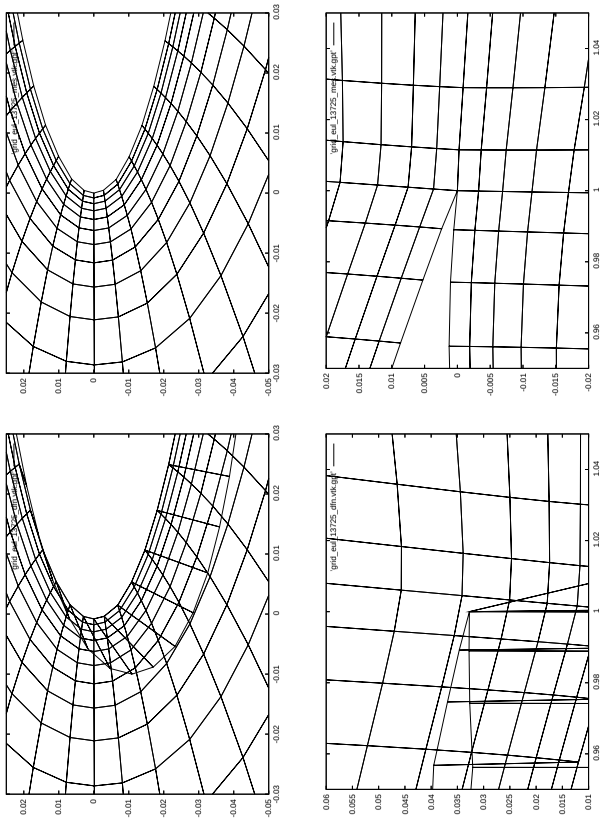
In Figures 3 and 4, the *leading* edge of the optimized meshes are shown for the o+ metrics. Figure 3 shows the results of optimizing using the metrics SS1 (top left), SS2 with $\gamma = \frac{1}{2}$ (top right), SS2 with $\gamma = 1$ (middle left), SS2 with $\gamma = 3$

(middle right), SS3 with $\gamma = 2$ (bottom left), and SS4 with $\gamma = \frac{1}{2}$ (bottom right). Figure 4 shows the results of optimizing using the metrics SS4 with $\gamma = 2$ (top left), SS4 with $\gamma = 4$ (top right), SS6 with $\gamma = \frac{1}{2}$ (middle left), SS7 with $\gamma = 1$ (middle right), SS5 (bottom left), and SS0 (bottom right). Figures 5 and 6 show the *trailing* edge of the optimized meshes for the o+ metrics, with the metrics in the same order as the previous pair of figures.

Table 2 gives a description of the results of optimizing with each of the metrics in terms of the mesh defects observed. Many of the defects can be plainly seen in the figures as well. The table also gives conclusions for each of the metrics in terms of what the results suggest for the necessity or sufficiency of the Eight Properties. As predicted, the SS1 mesh corresponding to the $\mu_2^{(o)}$ metric is tangled, most likely due to the unwanted global minimizers that are flips. Thus Properties 5 and 6 appear necessary. Moreover, the composite metrics SS2 and SS4, which contain the $|B^tB - I|^2$ term in them, exhibit a tendency to invert the mesh even though there are no unwanted global minimizers in these composite metrics. *It is clearly important to avoid the use of metrics with unwanted global minimizers, even if they only appear as one term in a composite metric.* Both the optimized SS5 and SS7 meshes are tangled, and the likely cause is the lack of continuity in each metric. Unwanted stationary points, whether they be saddles or local maxima, do not appear to cause mesh defects, thus their absence may not be a necessary property of a well-posed metric. Optimization with the composite metrics SS3 and SS6 did not lead to mesh defects, but they still have a minor drawback, namely, that they contain the parameter γ which must be selected. It was observed that, in particular, if γ is 'large', then the meshes tend toward being less smooth, which is a well-known result of optimization with an area-like term such as $(\beta - 1)^2$. This effect is illustrated in Figure 7. In that light, the metric SS0 appears the best of all since there is no parameter γ and because it produced no mesh defects even though it has a non-differentiable point which is a local maximum. Moreover, the optimal meshes from SS0 are reasonably smooth.

Figure 7 shows the leading and trailing optimal meshes for the SS3 metric, with $\gamma = 2.0$ on the left and $\gamma = 9.0$ on the right. The mesh for $\gamma = 9.0$ (leading edge) has radial lines that are more curved on the upper edge than SS3 with $\gamma = 2.0$, but otherwise there are no obvious flaws with respect to the reference mesh. On the trailing edge, the SS3 optimal mesh with $\gamma = 9.0$ is noticeably less smooth and with higher aspect ratios that the $\gamma = 2.0$ mesh. SS0 is closer to the $\gamma = 2.0$ mesh.

Figure 1: *Reference (top) & Initial (bottom) Airfoil Meshes*

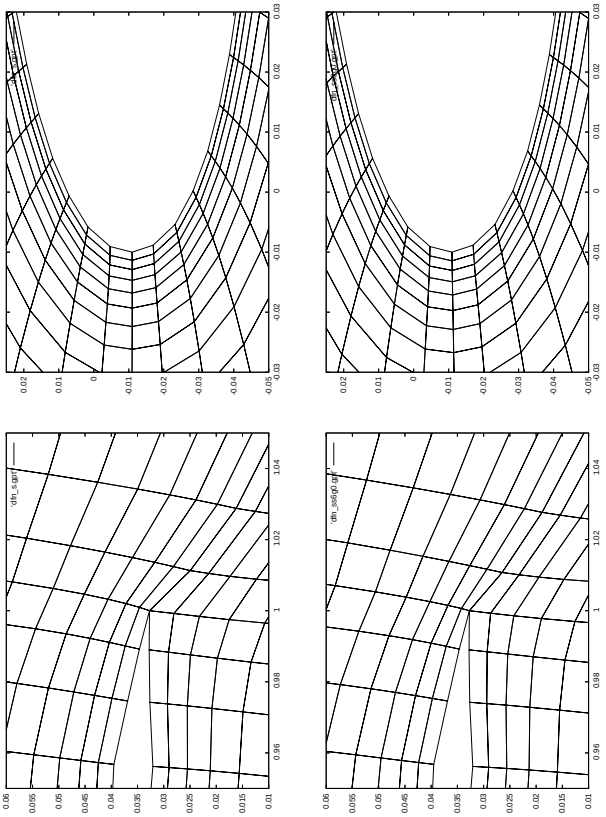


4.2 The Role of Solver Differences

All of the preceding experiments used Mesquite's quasi-Newton solver. As mentioned earlier, the appearance or absence of a defect in the optimal mesh might depend on the choice of numerical optimization solver. For example, Mesquite's Hessian-based Feasible Newton solver uses the full Hessian while quasi-Newton only uses the diagonal blocks of the Hessian, and the Steepest Descent solver only uses gradient information. In particular, the Hessian-based solvers require metrics which are twice-differentiable, while the steepest descent solver only requires once-differentiable metrics.

To investigate, the SS0 problem was re-run with the Feasible Newton and Steep-

Figure 2: *Leading & Trailing Edge of Optimized S1 and S2 Meshes*



est Descent solvers. The stopping criterion was 1.e-05 for the maximum change in vertex position per iteration. Recall that SS0 is continuous, but not differentiable on the set of scaled flips. A comparison of the three optimal leading-edge meshes resulting from the three solvers shows that, although there are minor differences, none of them contains a major mesh defect (see Figure 8). The same was true for the three optimal trailing-edge meshes, where the most noticeable difference was in comparing the steepest descent optimal mesh to the two others. To understand why the former has mesh lines beyond the airfoil that lean considerably more to the right, the stopping criterion was changed to 1.e-06 in the steepest descent calculation. With that change, the optimal mesh was closer in appearance to the Quasi- and Feasible-Newton optimal meshes. Thus, the difference in the Steepest Descent mesh is attributed to tightness of the

Table 2: Results & Conclusions of the Experiments Using the Quasi-Newton Solver

Metric	Result	Conclusion
S1	No mesh defects.	The Eight Properties may be sufficient.
S2	Same as S1.	Absence of Saddle Points may be sufficient, but not necessary.
SS1	Tangled optimal mesh.	Absence of Unwanted Global Minimizers may be necessary.
SS2	Tangled optimal mesh when $\gamma \leq \frac{2}{3}$	Spurious Global Minimizers in first term encourage mesh tangling.
SS2	Leading edge cells tending toward collapse when $\frac{2}{3} < \gamma \leq 2$	Defect due to lack of barrier and not because of the extra stationary points.
SS2	Same as previous, when $2 < \gamma$	Same as previous.
SS3	No mesh defects.	Absence of Saddle Points may be sufficient, but not necessary.
SS4	No mesh defects when $\gamma < 2$.	Absence of local maximum may be sufficient, but not necessary.
SS4	Tangled trailing edge cells when $2 \leq \gamma \leq 3$.	Saddles at $B = tF$ may be the cause or perhaps the lack of a barrier.
SS4	Tangled optimal mesh. $3 < \gamma$.	Spurious Global Minimizers in second term encourage mesh tangling.
SS6	No mesh defects.	Absence of a Saddle Point may be sufficient, but not necessary.
SS7	Tangled leading edge mesh.	Absence of discontinuity may be necessary.
SS5	Tangled optimal mesh when $\gamma < 2$.	Absence of discontinuity may be necessary.
SS0	No mesh defects.	Absence of non-differentiable points may be sufficient, but not necessary.

stopping criterion, and not to being more sensitive to the deficiencies of the SS0 metric. It is somewhat surprising that the lack of differentiability in SS0 was not a problem for any of the solvers, even though they all require the existence of at least first derivatives. One possible explanation perhaps is that Mesquite used numerical derivatives to calculate the gradient and Hessian so, unless the mesh is sufficiently fine, the non-differentiability may not be detected. Analytic derivatives for these metrics have not been implemented in Mesquite yet, so this explanation could not be tested.

Continuing, $SS3$ with $\gamma = 2.0$ was run next with the three solvers. A comparison of the three optimal leading-edge meshes shows that none has a major mesh defect (see Figure 9). The three optimal trailing-edge meshes also had no major defects. Tightening the stopping criterion in Steepest Descent cause the optimal mesh to more closely resemble the other two optimal meshes by lessening the rightward lean of the mesh lines. Thus, it appears that the spurious saddle point for this metric does not cause a problem for any of the solvers. Metric

S2 was also tried using the three solvers. Again, none of the solvers generated a mesh defect, even though S2 has many extraneous saddle points (see Figure 10). There is a fairly strong difference between the two Newton vs. the Steepest Descent meshes on the trailing edge using the 1.e-05 stopping tolerance; this difference vanished when the tolerance was changed to 1.e-06 for Steepest Descent.

These results suggest that the necessity/sufficiency of the list of Properties is not strongly dependent on the choice between these three solvers.

5 The Eight *Necessary* Properties

The Experiments reported were designed to investigate the necessity/sufficiency of the Eight Properties suggested in Section 2 for a well posed metric. The experiments did not test Properties 2, 3 and 4 because they seem clearly necessary if one is to numerically minimize an objective function.

The necessity of Property 1 was tested via the discontinuous metrics SS5 and SS7. The results suggest that the property is necessary since even a single undefined point ($B = 0$ in SS7) seems to have caused a defect. Properties 5 and 6 were tested via the metric SS1 which has global minimizers which do not belong to one of the four canonical sets and others which do; these properties also appear necessary. Property 7 was tested via the metric SS0, which is continuous, but non-differentiable at $B = tF$. Since SS0 produces meshes with no observable defects, Property 7 appears to be not absolutely necessary. The following wording is suggested to make Property 7 a necessary condition:

Property Seven as a Necessary Condition:

7. The metric may be non-differentiable with respect to B on $M_2 - \mathcal{D}^*$, a set much smaller than M_2. However, the points at which the metric is non-differentiable may not be local minima.

To test the necessity of this revised Property 7 would require construction of a metric that is non-differentiable at a local minimum. We have not discovered any so+ or o+ metrics with this property, so until contrary evidence is produced, we shall accept the necessity of the revised Property 7.

Finally, Property 8 was tested via metrics S2, SS3, and SS6. It appears that Property 8 is not necessary. The following wording is suggested to make Property 8 a necessary condition:

Property Eight as a Necessary Condition:

8. The set of stationary points of the metric (on $\mathcal{D}^*$) need not coincide with the set of global minimizers, but there cannot be any local minima.

To test the necessity of this revised Property 8 would require construction of a metric that has a stationary point that is a local minimum. We have not discovered any so+ or o+ metrics with this property, so until contrary evidence is produced, we shall accept the necessity of the revised Property 8.

Collectively, the original Eight Properties given in Section 2.4 appear to be sufficient for a well-posed metric.

With the necessary properties in hand, we return to Table 1 to see which of the metrics studied satisfy the Eight Necessary Properties. They are: S1, S2, SS3, SS6, and SS0, with S1 being the best so+ metric and SS0 being the best o+ metric.

6 Summary & Future Work

The Target-Matrix Paradigm includes certain local metrics having global minimizers that are either scaled rotations (to control local shape) or strict rotations (to control both shape and size); either type is orientation-invariant. Non-barrier forms of these metrics are important when the initial mesh to be optimized is tangled. Local metrics for two-dimensional meshes can differ radically from local metrics for three-dimensional meshes due to differences in certain properties of 2×2 vs. 3×3 matrices. To limit the scope of this work, this paper focused on 2D, non-barrier, so+ and o+ metrics. The primary goal was to develop a set of mathematical properties which are *necessary* for a well-posed TMP metric. A trial list of such properties was given in Section 2.4. Section 3 introduced two so+ metrics and six o+ metrics having global minimizers that are, respectively, scaled-rotations or simple-rotations. Table 1 summarizes, for each of the metrics, the properties from the trial list that are not satisfied. With these determined, a set of numerical optimization experiments was performed using each of the so+ and o+ metrics on a deforming airfoil problem. The presence of a defect, such as tangling, in the optimal mesh may indicate that one of the Eight trial properties is necessary for a well-posed metric. The lack of any defects in the optimal mesh may indicate that one or more of the Eight trial properties is non-essential. Muddying the waters, however, is the issue of the optimization solver: some solvers are more robust than others in terms of the types of mathematical deficiencies they are capable of addressing. To investigate the interaction of the solvers and the mathematical weaknesses of

the different so+ and o+ metrics, the experiments were repeated on a limited set of the better-behaved metrics using Quasi-Newton, Feasible-Newton, and Steepest Descent solvers. Keeping in mind the possibility that another mesh besides the deforming airfoil might alter some of the conclusions, it appears from the present results on the airfoil that properties such as continuity and non-extraneous global minimizers are essential (necessary), while absence of extraneous saddle points and local maxima may be non-essential (sufficient). Even lack of differentiability seems non-essential for the solvers that were used, provided numerical derivatives are used. Of course, the better-behaved a metric is in terms of the original Eight Properties, the more likely it is that the optimal meshes will be free of defects. Additionally, it is concluded from the comparison of results using the three optimization solvers that they are equally robust in terms of coping with the lack of first derivatives and with extraneous local saddles and maxima.

In pursuing the primary goal of the paper, a lot of new ground was covered. For example, in Section 2.3, definitions of continuity and differentiability of TMP metrics were given. The Eight Properties described in Section 2.4 consider the properties of the local metric $\mu(B)$, while Section 2.5 discussed the consequences of satisfying these Eight properties as they impact the properties of the metric as a function of its vertex coordinates. It was shown that consideration of the properties of the metric $\mu = \mu(B)$ is very useful in understanding the properties of the corresponding metric $\tilde{\mu} = \tilde{\mu}(\mathbf{x})$. Mathematical techniques introduced in Section 3 permitted a detailed analysis of the various so+ and o+ metrics in terms of properties such as continuity, differentiability, and critical point classification. Some of the metrics are much better behaved than others in terms of their mathematical properties. For example, metric (1) is an excellent so+ metric since it satisfies all Eight of the trial properties. The composite o+ metrics containing the term $B^tB - I$ were less satisfactory in that even though the global minimizers did not include the set of flips, the tendency to tangle was stronger in these metrics compared to o+ metrics that did not contain this term. Metrics that are undefined or discontinuous at certain points proved to be incapable of generating satisfactory optimal meshes so, for example, the idea of dividing the metric by $|B|$ to avoid the $B = 0$ stationary point did pan out. Composite metrics containing the term $(\beta - 1)^2$ did reasonably well, but if γ is chosen too large, the optimal meshes tended to be less smooth. Metric (20) is a new, non-obvious o+ metric that satisfies all the trial properties except Seven, because it contains a non-differentiable point that was a local maximum. Never-the-less, it performed well in the experiments and is more attractive than metrics (6) or (10) in that it contains no user-parameter γ.

Finally, this work shows that a careful investigation of the mathematical properties of candidate local metrics can be productive in terms of weeding out competing metrics, in determining the best values of parameters such as γ in

composite metrics, in choosing appropriate solvers, and in explaining or avoiding defects in the optimal mesh.

In future work, we shall study the problem of devising well-posed metrics for $d = 3$, using the Eight *Necessary* Properties. As one can imagine, this may be considerably more difficult than for $d = 2$, for several reasons including (when $d = 3$), (1) the stationary point equations become nine non-linear equations in nine unknowns instead of just four, (2) $|adjB|^2$ is not equal to $|B|^2$, (3) rotations do not have the simple form given in Section 2.2, and (4) $-R$ is a flip, not a rotation.

Other avenues for future work include (a) performing a study of the 2D *barrier* metrics to see whether they satisfy the Eight Necessary Properties, and (b) broadening the investigation on how the necessary properties might change if a different numerical optimization solver is used (or the same solvers, but with analytic derivative calculations), or if a different mesh optimization problem (different domain and mesh topology) were studied.

Acknowledgments
Thanks to Jason Kraftcheck for implementing the basic infrastructure in Mesquite, along with supplying a Mesquite driver code, so that these metrics could be tested.

Figure 3: *Leading Edge of Optimized Meshes - I*

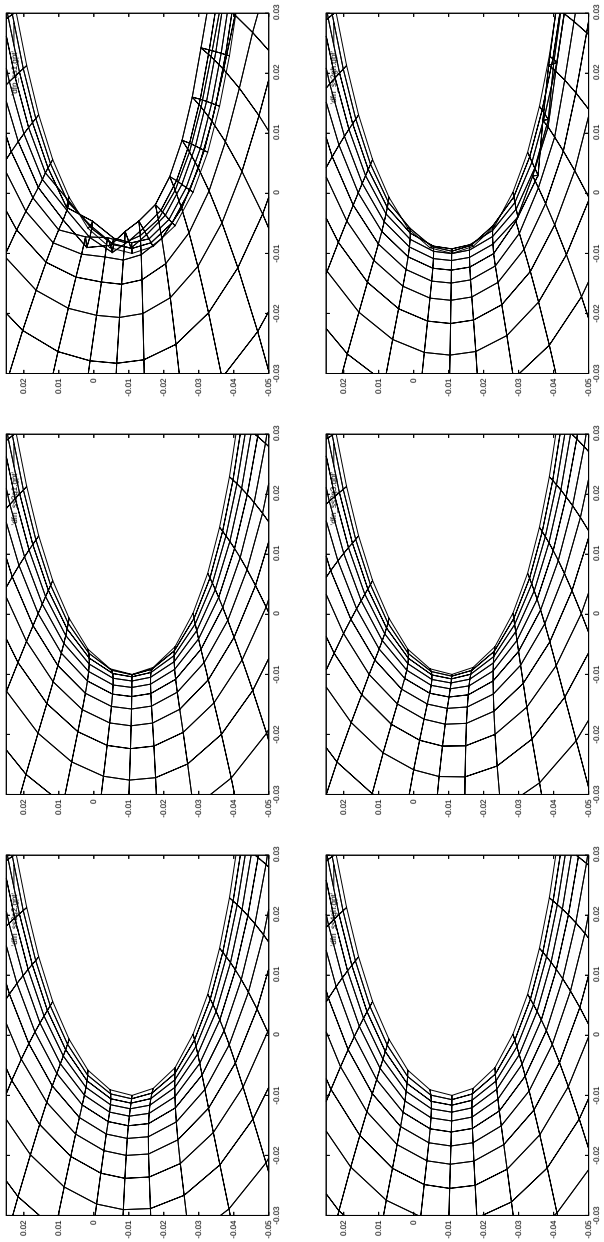


Figure 4: *Leading Edge of Optimized Meshes - II*

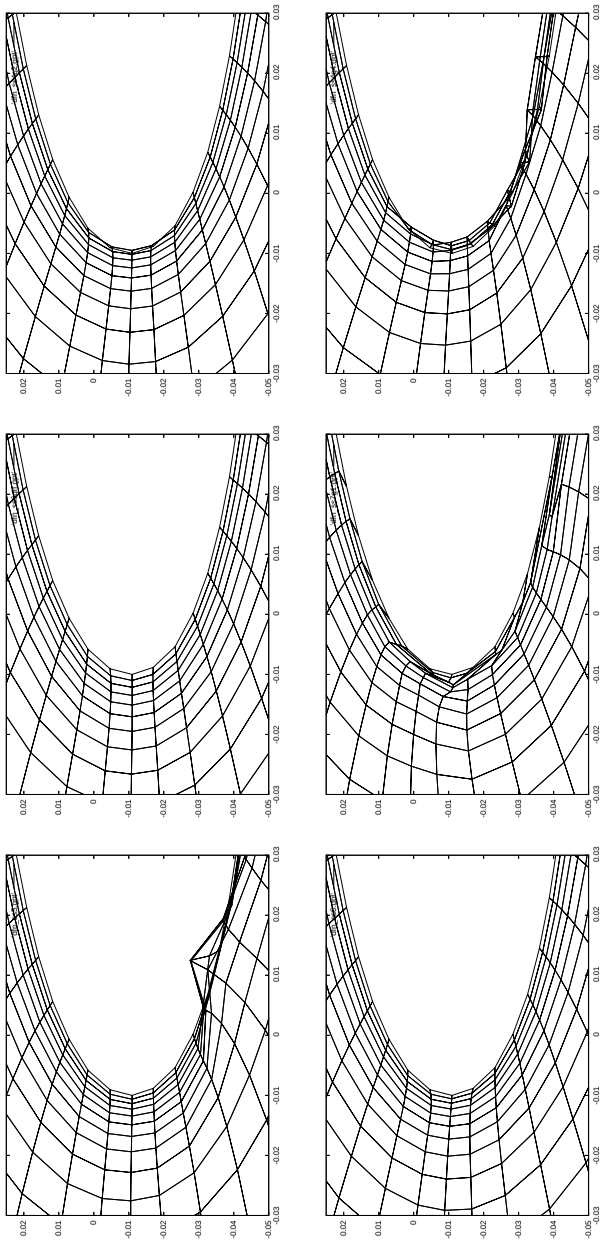


Figure 5: *Trailing Edge of Optimized Meshes - I*

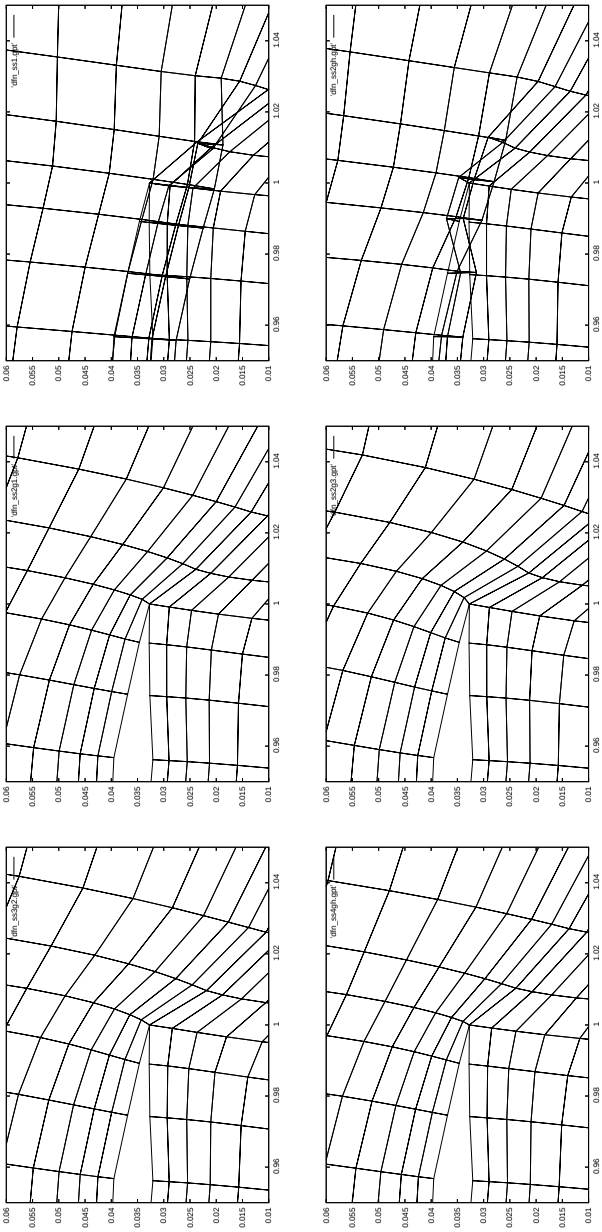


Figure 6: *Trailing Edge of Optimized Meshes - II*

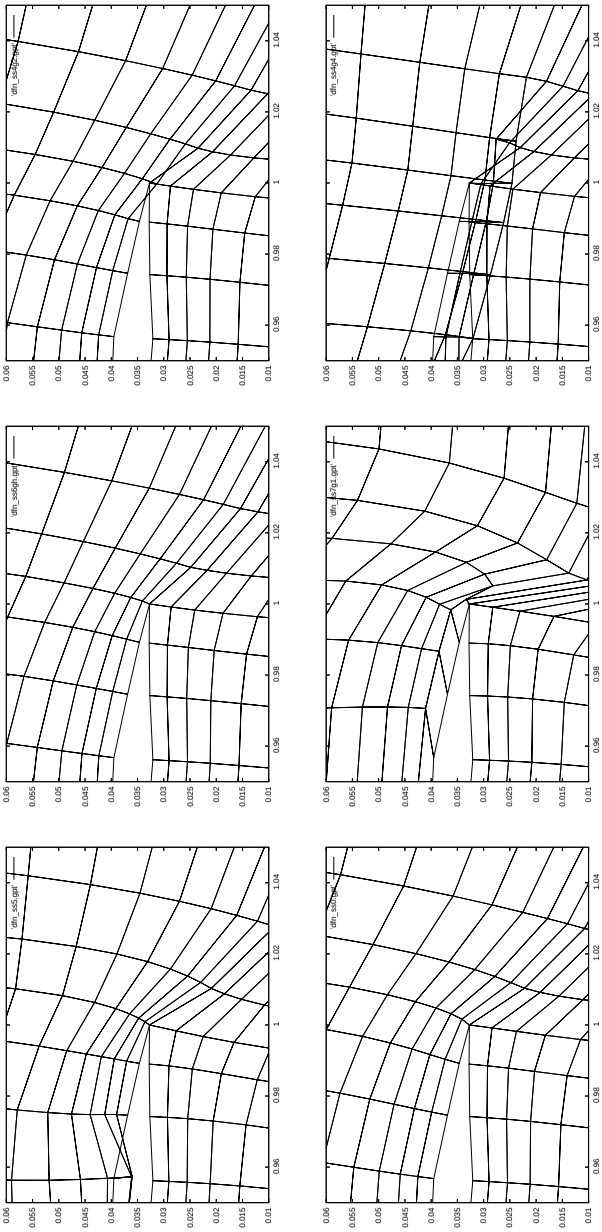


Figure 7: *Comparing $\gamma = 2$ with $\gamma = 9$ in SS3 Optimized Meshes*

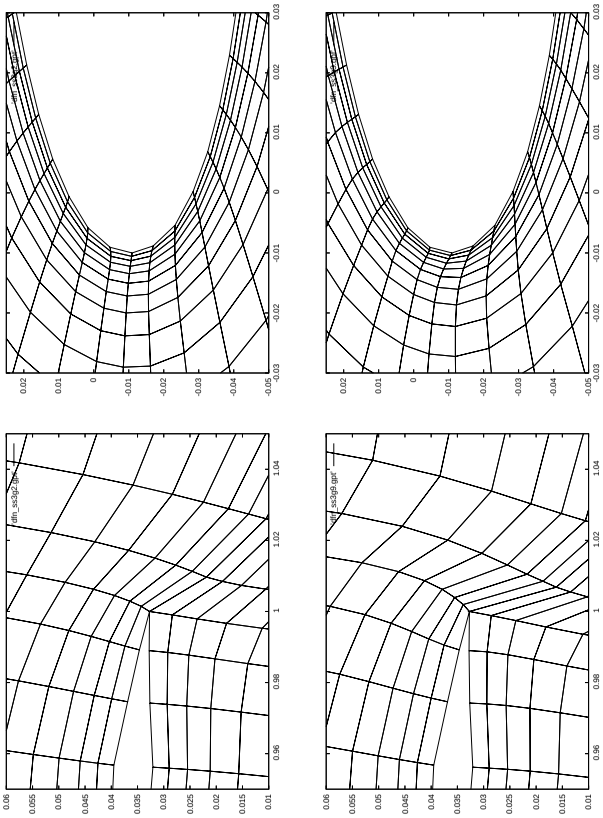


Figure 8: *Comparing Optimal Meshes from the Quasi-Newton (left), Feasible-Newton (Middle), and Steepest Descent (Right) Solvers on SS0*

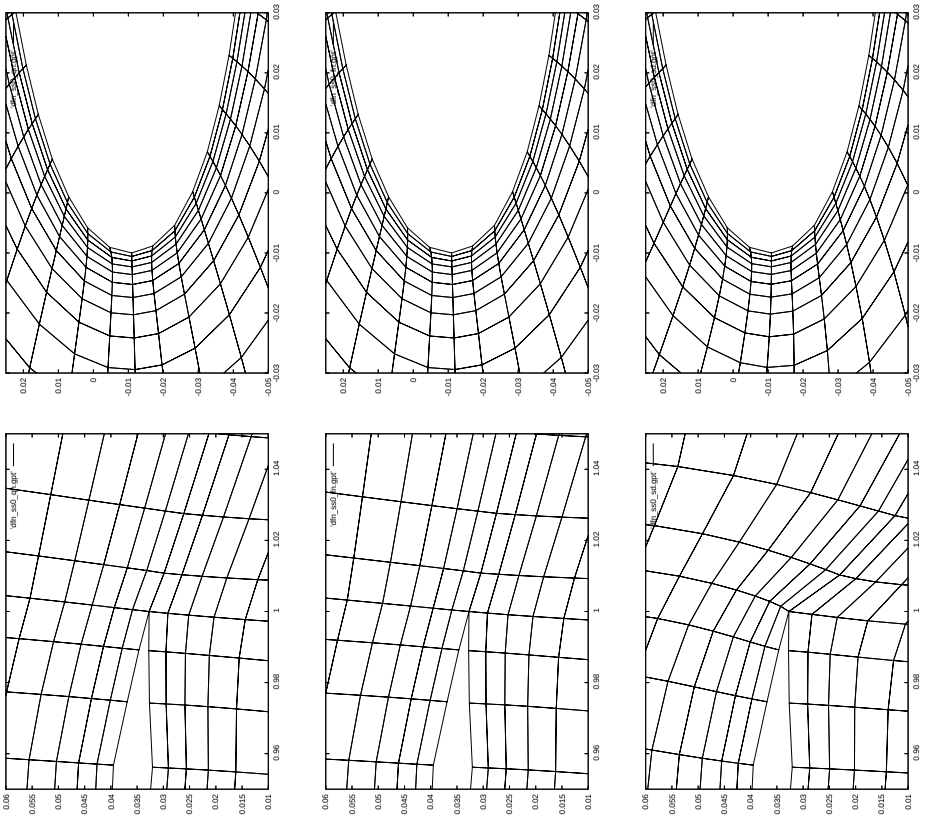


Figure 9: *Comparing Optimal Meshes from the Quasi-Newton (left), Feasible-Newton (Middle), and Steepest Descent (Right) Solvers on SS3 with* $\gamma = 2$

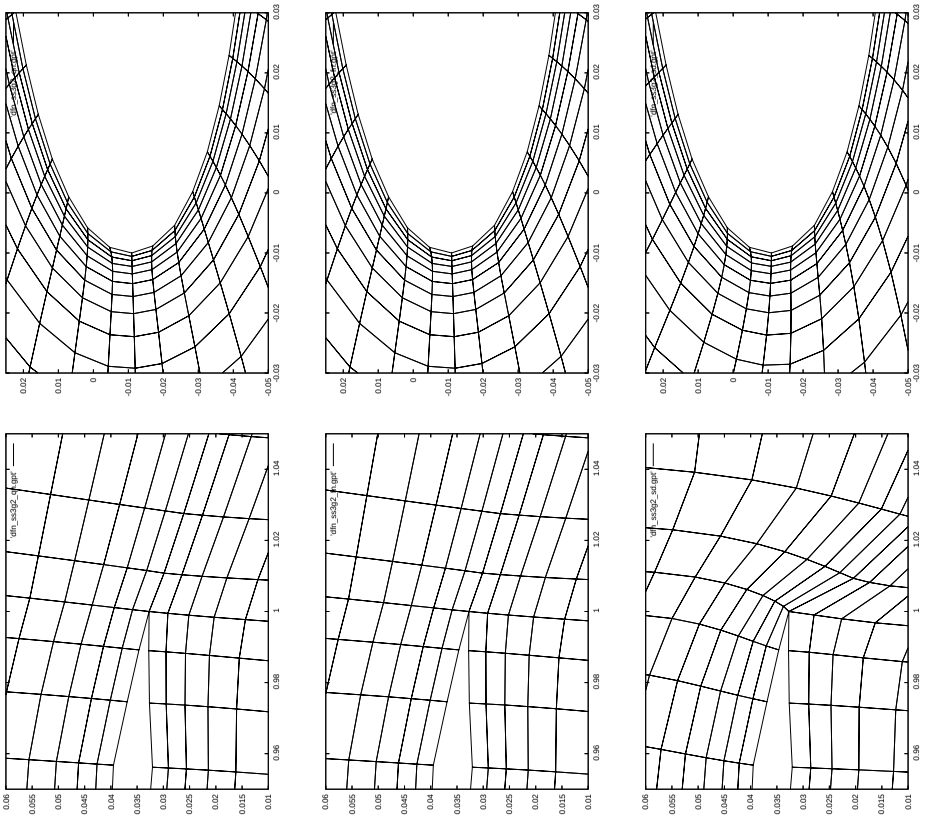
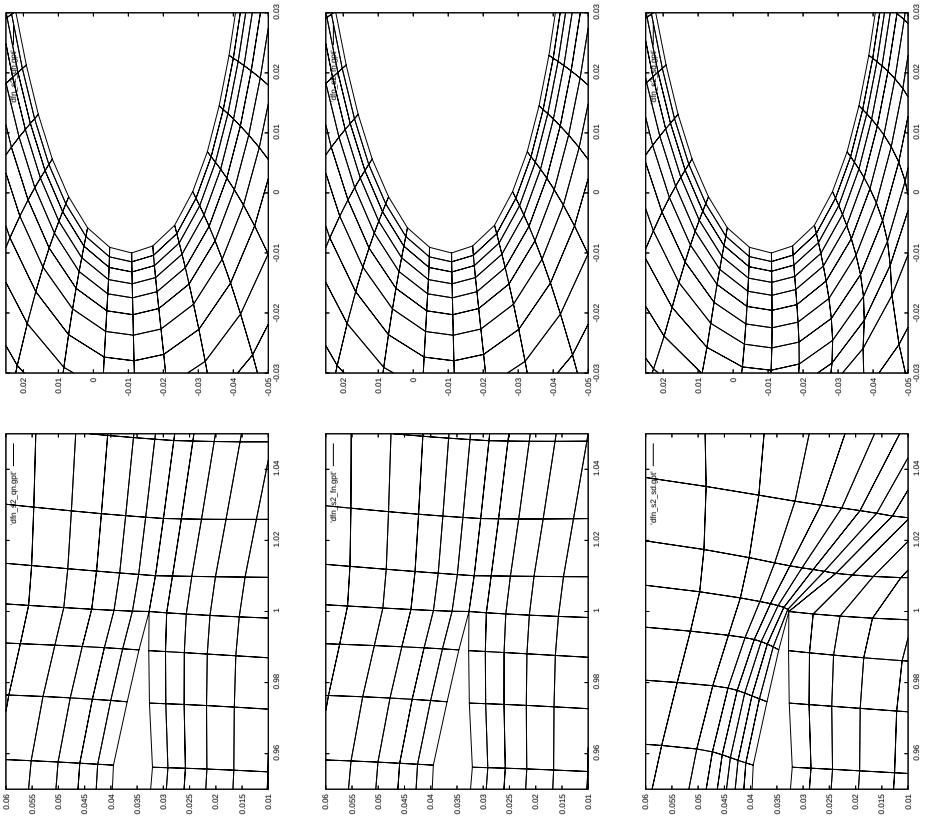


Figure 10: *Comparing Optimal Meshes from the Quasi-Newton (left), Feasible-Newton (Middle), and Steepest Descent (Right) Solvers on S2*



References

[1] J. Brackbill and J. Salzman,, *Adaptive zoning for singular problems in two-dimensions*, J. Comp. Phys., **46**. pp. 342-368, 1982.

[2] S. Steinberg and P. Roache, *Variational grid generation*, Num. Meth. PDE's, **2**, pp. 71-96,, 1986.

[3] P. Knupp, *Jacobian-weighted Elliptic Grid Generation*, pp. 1475-1490, SIAM J. Sci. Comp., Vol. 17, No. 6, 1996.

[4] J. Castillo, *A discrete variational method*, SIAM J. Sci. Stat. Comput., Vol. 12, No. 2, pp. 454-468, 1991.

[5] A. Pardhannani and G. Carey, *Optimization of Computational Grids*, Num. Meth. PDE's, **4**, pp. 95-117, 1988.

[6] P. Zavatierri, *Optimization Strategies in Unstructured Mesh Generation*, Int. J. Num. Meth. Engr., **39**, pp. 2055-2071, 1996.

[7] J.Tinoco-Ruiz, and P. Barrera-Sanchez, *Area functionals in Plane Grid Generation*, pp. 293-302, in Numerical Grid Generation in Computational Field Simulations, M. Cross et. al. eds., Greenwich UK, 1998.

[8] J. Thompson, J. Thames, and C. Mastin, *Automatic numerical generation of body-fitted curvilinear coordinate system for field containing any number of arbitrary two-dimensional bodies*, J. Comp. Phys., **15**, pp. 299-319, 1974.

[9] J.Tinoco-Ruiz, and P. Barrera-Sanchez, *Smooth and convex grid generation over general planar regions*, Math. and Comp in Sim., 1998.

[10] L. Freitag and P. Knupp, *Tetrahedral mesh improvement via optimization of the element condition number*, Int. J. Num. Meth. Engr., Vol. 53, No. 6, pp. 1377-1391, 2002.

[11] A. Dvinsky, *Adaptive grid generation from harmonic maps on Riemannian manifolds*, J. Comp. Phys., **95**, pp. 450-476, 1991.

[12] V. Liseikin, *A Computational Differential Geometry Approach to Grid Generation*, Springer-Verlag, 2004.

[13] P. Knupp, *Formulation of a Target-Matrix Paradigm for Mesh Optimization*, SAND2006-2730J, Sandia National Laboratories, Albuquerque NM, 2006.

[14] P. Knupp and H. Hetmaniuk, *Local 2D Metrics for Mesh Optimization in the Target-Matrix Paradigm*, SAND2006-7382J, Sandia National Laboratories, Albuquerque NM, 2006.

[15] P. Knupp, *Algebraic Mesh Quality Metrics*, SIAM J. Sci. Comput., **23**, pp. 193-218, 2001.

[16] P. Knupp, *Algebraic Mesh Quality Metrics for Unstructured Initial Meshes*, Finite Elements in Design and Analysis, **39**, pp. 217-241, 2002.

[17] P. Knupp, *Updating Meshes on Deforming Domains*, Communications in Numerical Methods in Engineering, 24:467-476, 2008.