NUMERICAL METHODS FOR BIOMEMBRANES: 
CONFORMING SUBDIVISION METHODS VERSUS 
NON-CONFORMING PL METHODS

JINGMIN CHEN, THOMAS YU, PATRICK BROGAN, ROBERT KUSNER, YILIN YANG, 
AND ANDREW ZGERELLI

Abstract. The Canham-Helfrich-Evans models of biomembranes consist of a family of geometric constrained variational problems. In this article, we compare two classes of numerical methods for these variational problems based on piecewise linear (PL) and subdivision surfaces (SS). Since SS methods are based on spline approximation and can be viewed as higher order versions of PL methods, one may expect that the only difference between the two methods is in the accuracy order. In this paper, we prove that a numerical method based on minimizing any one of the 'PL Willmore energies' proposed in the literature would fail to converge to a solution of the continuous problem, whereas a method based on minimization of the bona fide Willmore energy, well-defined for SS but not PL surfaces, succeeds. Motivated by this analysis, we propose also a regularization method for the PL method based on techniques from conformal geometry. We address a number of implementation issues crucial for the efficiency of our solver. A software package called Wmincon accompanies this article, and provides parallel implementations of all the relevant geometric functionals. When combined with a standard constrained optimization solver, the geometric variational problems can then be solved numerically. To this end, we realize that some of the available optimization algorithms/solvers are capable of preserving symmetry, while others manage to break symmetry; we explore the consequences of this observation.

1. Introduction

Lipid bilayers are arguably the most elementary and indispensable structural components of biological membranes which form the boundary of all cells. It is known since the seminal work of Canham [14], Helfrich [39] and Evans [30] in the...
70's that bending elasticity, induced by curvature, plays the key role in driving the geometric configurations of such membranes.

The so-called spontaneous curvature model of Helfrich suggests that a biomembrane surface $S$ configures itself to minimize $\int_S H^2 dA$ subject to the area, volume and area difference (related to the bilayer characteristics) constraints, i.e. $S$ solves the variational Helfrich problem

\[
\min_S W(S) := \int_S H^2 dA \text{ s.t.} \left\{ \begin{array}{l}
(i) \quad A(S) := \int_S 1 \, dA = A_0, \\
(ii) \quad V(S) := \frac{1}{3} \int_S [x\hat{i} + y\hat{j} + z\hat{k}] \cdot \hat{n} dA = V_0, \\
(iii) \quad M(S) := -\int_S H \, dA = M_0.
\end{array} \right.
\]

Here $H = (\kappa_1 + \kappa_2)/2$ is the mean curvature. In (ii), $V(S)$ is the enclosed volume, expressed here as a surface integral of $S$ via the divergence theorem. The connection of (iii) to bilayer area difference comes from the relation $-\int_S H dA = \lim_{\varepsilon \to 0} \frac{1}{4\varepsilon} (\text{area}(S_+\varepsilon) - \text{area}(S_-\varepsilon))$, where $S_+\varepsilon$ and $S_-\varepsilon$ are the $\varepsilon$-offset surfaces and that the thickness of the lipid bilayer, $2\varepsilon$, is negligible compared to the size of the vesicle; see Figure 1. The constraint values $A_0, V_0$ and $M_0$ are determined by physical conditions (e.g. temperature, concentration). $W(S)$ is called the Willmore energy of the surface $S$. When the area-difference constraint (iii) is omitted, the variational problem is referred to as the Canham problem. When even the volume constraint (ii) is omitted, there is essentially no constraint as $W$ is scale-invariant; in this case the area constraint (i) only fixes the scale, and we refer to the variational problem as the Willmore problem.

Due to the scale-invariance of the Willmore energy, the solution, up to homothety, of any of the Willmore, Canham or Helfrich problems depends only on the reduced volume and reduced total mean curvature defined by:

\[
v_0 := V_0/[(4\pi/3)(A_0/4\pi)^{3/2}], \quad m_0 := M_0/[4\pi(A_0/4\pi)^{1/2}].
\]

This terminology is used by a group of biophysicists who have done many computational and physical experiments exploring the shapes of phospholipid vesicles, and we shall follow it. Note that $v_0$ is essentially what a geometer would call the isoperimetric ratio. By the isoperimetric inequality, we have $v_0 \in (0, 1]$ and $v_0 = 1$ is uniquely realized by a round sphere.

It is observed experimentally that no topological change occurs in any accessible time-scale, so we aim to solve any of the Helfrich, Canham or Willmore problems when $S$ is assumed to be an orientable closed surface with a fixed genus $g$. Spherical
(g = 0) vesicles are the most common among naturally occurring biomembranes, although higher genus ones have been synthesized in the laboratory [43,56,65]. The Canham, Helfrich and related models explain the large variety of shapes observed in even a closed vesicle with a spherical topology [50,51,65,74].

Several numerical treatments of these models have been proposed in the literature: [10,40], [15,32], [9,19], [21–23], [62]. Among these, the methods in [10,40] were used extensively by biophysicists to study real lipid bilayer membranes. While the key ingredients of these algorithms are implemented in Brakke’s well-known Surface Evolver software [10], the overall algorithms were not completely analyzed by the geometers who invented them [10,33,40] and even less so by the biophysicists who used them [43,56,65,74]. As such, there is little understanding of these methods, and the computational results claimed in the extensive biophysics literature are difficult to reproduce. Moreover, there is no systematic comparison of this method with the later ones, at both a theoretical or computational level.

These numerical methods continue to be used extensively in the study of phenomena in biomembranes, see, e.g. [2,4,44,45,77] and references therein. Similar geometric variational problems show up in other scientific areas. A notable example is found in the quasi-local mass problem of general relativity, in which maximizers of the Hawking mass – defined similarly as the Willmore energy – are sought.

The goal of this paper is to clarify and refine some of these numerical methods, and establish some theoretical understandings of them. Before we proceed, we mention further related work on discrete minimal surfaces and discrete elasticae (the 1-D counterpart to Willmore surface), see, e.g., [13,25,26,63] and the references therein.

1.1. PL and SS. A standard approach to represent surfaces of arbitrary topology is to use the piecewise linear (PL) approach. A PL surface can be specified by a mesh \(\mathcal{M} = (\mathcal{V}, \mathcal{F})\) where \(\mathcal{V} \in \mathbb{R}^{#\mathcal{V} \times 3}\) records the 3-D coordinates of the vertices of the control mesh, \(\#\mathcal{V}\) denotes the total number of vertices, and \(\mathcal{F} \in \mathcal{I}^{#\mathcal{F} \times 3}\) is a list of triplets of indices from \(\mathcal{I} := \{1, \ldots, \#\mathcal{V}\}\) which records the vertices of each of the \(\#\mathcal{F}\) triangle faces in the mesh \(\mathcal{M}\). We assume that the PL surfaces realized by the mesh are closed and orientable. The orientation can be conveniently encoded in a consistent ordering of the vertices in the face list \(\mathcal{F}\).

In a numerical method, \(\mathcal{F}\) is usually fixed and \(\mathcal{V}\) varies. This fits the framework of our variational problems well, as fixing \(\mathcal{F}\) also fixes the genus of the surface, and varying \(\mathcal{V}\) means we find the embedding of \(\mathcal{F}\) – viewed as an abstract simplicial complex – that optimizes the Willmore energy under the corresponding constraint(s).

A closed, oriented PL surface has a well-defined area \(A\) and enclosed volume \(V\), but no classically defined normals or mean curvatures, hence it also does not have a classically defined total mean curvature \(M\) or Willmore energy \(W\). As such, any numerical method for the Willmore, Canham, or Helfrich problems based on approximating the solution surfaces by PL surfaces may be classified as a nonconforming method in FEM parlance.

A subdivision surface (SS) is specified by the same data \(\mathcal{M} = (\mathcal{V}, \mathcal{F})\), except that the associated surface has enough regularity for a well-defined total mean curvature \(M\) or Willmore energy \(W\). We shall primarily be using the Loop and C2g0 subdivision surfaces introduced in [52] and [15], respectively. See Figure 2(a) for a genus 0 control mesh, and Figure 2(b) for the corresponding Loop subdivision surface. For each face \(f\) in \(\mathcal{M}\), there is a corresponding surface patch; see Figure 2(b).
The C2g0 scheme handles only genus 0 and 1 surfaces using control meshes with only valence 3 and 6 vertices; the resulting SS are $C^2$ everywhere, so there is no question about the well-definedness of $M$ and $W$. The Loop subdivision scheme handles surface of arbitrary genus and control meshes with arbitrary valences, but the resulting SS are not $C^2$ everywhere. They are however regular enough to have well-defined $M$ and $W$. (The subdivision functions are in $W^{2,2} \cap C^1$ when expressed in characteristic coordinates; see Section 3.1). Therefore, any numerical method for the Willmore, Canham, or Helfrich problems based on approximating the solution surfaces by SS may be classified as a conforming method.

1.2. Contributions of this paper. This paper contributes to the numerical study of biomembranes in the following ways:

(I) In Section 2 we give an exposition of numerical methods based on both PL and SS. In the PL case, we connect and compare several different ideas developed in the applied geometry literature [7,10,20,40,55]. In the SS case, we explain how all the functionals and their gradients with respect to control vertex coordinates can be efficiently computed based on a precomputation of basis functions. In both cases, we develop parallel implementations and explore new examples. A Matlab based toolbox named WMINCON, with CUDA and C++ implementations of all key functionals, is available online at

http://www.math.drexel.edu/~tyu/Software/Wmincon,

for reproducing the computational results. These algorithmic developments allow us to use the SS method to attack many instances of the Willmore, Canham and Helfrich problems that would otherwise be too slow, if not impossible, to solve on existing computers.

(II) In Section 3 we present an argument explaining why a conforming method would work. This result relies on the existence theory of Willmore minimizers pioneered by Simon [66]. In contrast, we show that a naive minimization of several PL Willmore energies would fail to solve the Willmore
problem. The analysis elucidates the difference between conforming and non-conforming methods.

(III) In Section 4, based on a well-founded principle (Section 4.1) exploiting the uniformization theorem and the theory of harmonic maps, we propose a regularization of the PL methods based on penalization by harmonic energy. Unlike the unregularized PL methods which are doomed to fail, the regularized method appears to yield solutions converging to those of the continuous problems.

(IV) In Section 5, we make the observation that certain optimization algorithms are capable of preserving symmetry, while others are capable of breaking symmetries. We carry out a number of experiments comparing different optimization algorithms in conjunction with our discretization methods. The experiments reveal subtle analytic properties of the optimization problems arising from the SS methods. We formulate a number of conjectures.

The kind of conforming and non-conforming methods we study in this article are those in the spirit of ‘minimizing a discretization’, i.e. the methods under study first discretize the variational problem, in either a conforming or non-conforming way, followed by solving the resulting finite-dimensional optimization problem. There are methods, such as those in [9,62], that are in the spirit of ‘discretizing a minimization’. These methods first consider a minimization process in the continuous setting, akin to a gradient flow, followed by strategies to discretize the flow. This last step can also be done in a conforming or non-conforming way. These methods are all based on explicit representations of surfaces; there are also methods based on implicit representations, such as the phase field methods of Du et al. [21–23].

2. Numerical methods based on PL and SS functionals

Recall that either a PL or SS is specified by a control mesh \( \mathcal{M} = (\mathcal{V}, \mathcal{F}) \). In our numerical method, we assume that \( \mathcal{F} \) is fixed and \( \mathcal{V} \) varies. For most \( \mathcal{V} \), an immersed surface, denoted by \( S[\mathcal{V}] \), is defined. The numerical methods considered here approximate the Helfrich problem (1.1) by a finite-dimensional analog:

\[
\min_{\mathcal{V}} W(\mathcal{V}) \quad \text{s.t.} \quad \begin{cases} 
(i) & A(\mathcal{V}) = A_0 \\
(ii) & V(\mathcal{V}) = V_0 \\
(iii) & M(\mathcal{V}) = M_0 
\end{cases}
\]

The numerical methods for the Canham and Willmore problems are similar: simply drop the corresponding constraints. Already mentioned in Section 1.1, the PL and SS methods have the following features and relative pros and cons:

- For SS (based on any regular enough scheme, such as Loop and C2g0), all four functionals are the exact, well-defined, values of the \( W, A, V \) and \( M \) of the corresponding subdivision surface. Their computations, however, have to be performed based on numerical integration.
- For PL, \( W \) and \( M \) are not well-defined for the corresponding PL surface. We will therefore replace \( W(\mathcal{V}) \) and \( M(\mathcal{V}) \) in (2.1) by a certain consistent discretization, to be reviewed below. These PL Willmore and total mean curvature energies are relatively simple to implement and no numerical integration is required.
The materials in this section are mostly not new, some of them are actually quite old. The intention is to unify them at one place in order to prepare us for the later sections.

At first glance, one may expect that the PL method is simply less accurate than the SS methods, i.e., a PL method would converge but at a lower rate compared to a SS method. Our analysis in Section 3 falsifies this speculation. A bulk of this section discusses the definition, properties and computation of these functionals and their derivatives. Efficient computation of these functionals and their gradients are necessary for the numerical solution of the Helfrich, Canham and Willmore problems using a standard nonlinear optimization solver; see Section 2.3.

2.1. $W$, $A$, $V$, $M$ for PL surfaces. The area $A$ and enclosed volume $V$ of the PL surface are of course part of the biomembrane problems. Their gradients are not only needed for our optimization solver but also are connected to the way $W$ and $M$ are defined and computed. For these reasons, we derive them for the convenience of the readers.

We aim to clarify some not so well-documented details in the literature, such as the sign issue of discrete mean curvature, which is irrelevant for $W$ but crucial for $M$, and the choice of local areas, which is irrelevant for $M$ but impact the behavior for $W$. Another goal is to elucidate the connections of a number of different discrete mean curvature operators and Willmore and total mean curvature energies.

2.1.1. $A$ and $V$. The area $A$ and enclosing volume $V$ of a closed oriented PL surface can be computed as:

$$A = \frac{1}{2} \sum_{f \in F} \| (\mathcal{V}_{f_2} - \mathcal{V}_{f_1}) \times (\mathcal{V}_{f_3} - \mathcal{V}_{f_1}) \|, \quad V = \frac{1}{6} \sum_{f \in F} \det((\mathcal{V}_{f_1}, \mathcal{V}_{f_2}, \mathcal{V}_{f_3})).$$

For any smooth functional $F : \text{domain}(F) \subset \mathbb{R}^{#V \times 3} \cong \prod_{v \in V} \mathbb{R}^3 \to \mathbb{R}$, we denote by $\nabla_v F$ ($\in \mathbb{R}^3$) its gradient with respect to the coordinates of the vertex indexed by $v \in V$. The volume gradient can be expressed as

$$\nabla_v V = \frac{1}{6} \sum_{i} \nabla_v \det((\mathcal{V}_v, \mathcal{V}_{w_i}, \mathcal{V}_{w_{i+1}})) = \frac{1}{6} \sum_{i} \mathcal{V}_{w_i} \times \mathcal{V}_{w_{i+1}} \in \mathbb{R}^3,$$

where $w_1, \ldots, w_i, w_{i+1}, \ldots \in V$ is a counterclockwise enumeration (viewed from the outside) of the vertices connected to $v$ (a.k.a. the ‘1-ring’ of $v$).

Note that the gradient formula can be used to show that the formula for $V$ is invariant under rigid motions when the PL surface is closed and consistently oriented. (Observe that $\sum_{v} (\nabla_v V, a) = 0$ for any constant vector $a$; the proof relies on both assumptions.) The volume gradient itself is invariant under translation and equivariant with respect to rotation. The latter is obvious from the formula; the former is obvious also as $V$ is translation-invariant, but it helps to see it directly from the formula: $\sum_{i} (\mathcal{V}_{w_i} + a) \times (\mathcal{V}_{w_{i+1}} + a) = \sum_{i} \mathcal{V}_{w_i} \times \mathcal{V}_{w_{i+1}} + \sum_{i} \mathcal{V}_{w_i} \times a + a \times \sum_{i} \mathcal{V}_{w_{i+1}} + \sum_{i} a \times a = \sum_{i} \mathcal{V}_{w_i} \times \mathcal{V}_{w_{i+1}} + \sum_{i} a \times a = \sum_{i} \mathcal{V}_{w_i} \times \mathcal{V}_{w_{i+1}}.$ Note that the two sums in the middle cancel only because $w_1, \ldots, w_{\text{val}(v)}$ form a closed-loop.

Remark 2.1. The vector $\nabla_v V$, in turn, has another geometric interpretation: if the “base of the pyramid around $v$” is coplanar, i.e., the vertices indexed by $w_1, \ldots, w_{\text{val}(v)}$ lie on the same plane and form a polygon, then $\nabla_v V$ is a vector orthogonal to the plane and its length is one-third the area of the polygon. (By translation-invariant, we can assume that the polygon is centered at the origin.) Clearly, it is independent of the coordinates of vertex $v$. In general, $3\|\nabla_v V\|$ can
be used to define a notion of the “area of a non-planar polygon.” This so-called ‘effective area’ is used to define one of the discrete mean curvatures.

Next, we have the following derivation for the area gradient:

\[
\nabla_v A = \frac{1}{2} \sum_i \nabla_v \| (\mathcal{V}_{w_i} - \mathcal{V}_v) \times (\mathcal{V}_{w_{i+1}} - \mathcal{V}_v) \|
\]

\[
= \frac{1}{2} \sum_i \left( (p_{i+1} - p_i) \cdot p_i + (p_{i+1} - p_i) \cdot p_i \right)
\]

\[
= \frac{1}{2} \sum_i \left( \cot \angle vw_{i+1} w_i p_i + (\cot \angle vw_{i+1} w_i) p_i \right),
\]  

(2.4)

where \(\alpha_i\) and \(\beta_i\) are the angles opposite the edge \(vw_i\) in the two incident triangles. In the above, the second equality can be seen from the chain rule, in which an intermediate map is of the form \(C(\mathbf{x}) = (\mathbf{a} - \mathbf{x}) \times (\mathbf{b} - \mathbf{x})\), which can be simplified to \(\mathbf{a} \times \mathbf{b} + (\mathbf{b} - \mathbf{a}) \times \mathbf{x}\) and hence has a constant derivative expressible by a cross product. The third equality follows from the vector triple product formula \((\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{c} \cdot \mathbf{b})\mathbf{a} + (\mathbf{c} \cdot \mathbf{a})\mathbf{b}\). The fifth equality follows from \(\mathbf{a} \cdot \mathbf{b} / \|\mathbf{a} \times \mathbf{b}\| = \cot(\text{angle between} \ \mathbf{a} \ \text{and} \ \mathbf{b})\).

Equation (2.4) is connected to the well-known cotangent formula for the Laplace-Beltrami operator; see Remark 2.3.

The following comment will be found useful when computing discrete mean curvature.

Remark 2.2. It is clear that \(A\), and hence also \(\nabla A\), has nothing do with the global orientation of the PL surface; in particular, they are well-defined even for a non-orientable PL surface. The direction of \(\nabla_v A\) tells ‘which way the PL surface is poking’ at the vertex \(v\). However, the enclosing volume \(V\) requires the surface to be both closed and orientable, and in this case the formula for \(V\) in (2.2) would only give the enclosing volume if all the faces are oriented in a counter-clockwise fashion when viewed from the outside. In particular, reversing the orientation of all the faces would flip the sign of \(V\) and reverse the direction of each \(\nabla_v V\).

2.1.2. \(W\) and \(M\). We review 5 discrete Willmore and two discrete total mean curvature energy functionals for PL surfaces which we learn from \([6, 11, 40, 55, 69]\). We label them as

\[
W_{\text{Centroid}}, W_{\text{Voronoi}}, W_{\text{EffArea}}, W_{\text{NormalCur}}, W_{\text{Bobenko}}, M_{\text{Cotan}}, M_{\text{Steiner}}
\]
in this paper and in the WMINCON package.
Recall that for any smooth orientable surface $S$ with continuous unit normals denoted by $n(x), x \in S$, we have

$$\frac{d}{dt} |_{t=0} \text{Area}(S_t) = -2 \int_S h(x)H(x)dA, \quad \forall h : S \to \mathbb{R},$$

where $S_t := \{ x + th(x)n(x) : x \in S \}$, and $H$ is the mean curvature defined relative to the choice of the normals $n$. The above functionals, except $W_{\text{Bobenko}}/M_{\text{Edge}}$, can be derived based on defining

‘discrete normals’ $n : \mathcal{I} \to S^2$, ‘discrete mean curvatures’ $H : \mathcal{I} \to \mathbb{R}$, and

‘local areas’ $a : \mathcal{I} \to \mathbb{R}^+$

at the vertices of a PL surface, indexed by $\mathcal{I} = \{1, \ldots, \#V\}$, that satisfy a discrete analog of (2.5), namely,

$$\lim_{t \to 0} \frac{1}{t} \left[ A(V + t[h(v)n(v)]_{v \in \mathcal{I}}) - A(V) \right] = -2 \sum_{v \in \mathcal{I}} h(v)H(v)a(v), \quad \forall h : \mathcal{I} \to \mathbb{R}.$$

The left-hand side is the directional derivative of $A$ at $V$ in the direction $[h(v)n(v)]_{v \in \mathcal{I}}$, which equals

$$\sum_v \langle \nabla_v A(V), h(v)n(v) \rangle_{\mathbb{R}^3}.$$

In order for it to equal the right-hand side of (2.6) for all scalar field $h$, it is necessary and sufficient, by setting $h(v') = \delta_{v,v'}$, for $n(v), H(v)$ and $a(v)$ to satisfy

$$\nabla_v A \cdot n(v) = -2H(v)a(v), \quad \forall v.$$

Once $a(v)$ is assigned, then $n(v)$ and $H(v)$ can be chosen so that the \textit{mean curvature vector} is

$$H(v) := H(v)n(v) = -\frac{\nabla_v A}{2a(v)}.$$

This only defines $H(v)$ and $n(v)$ up to a sign; the appropriate sign must be determined from the \textit{global orientation} of the PL surface; a natural way is to choose $n$ so that

$$\langle n(v), \nabla_v V \rangle > 0;$$

recall Remark 2.1 and Footnote 1. The sign of $H(v)$ can then be determined accordingly.

The local areas used in the various schemes are summarized in the following table.

<table>
<thead>
<tr>
<th>Scheme</th>
<th>local area $a(v)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centroid [40]</td>
<td>$\text{Area}(\text{star}(v))/3 := a_{\text{centroid}}(v)$</td>
</tr>
<tr>
<td>Voronoi [55]</td>
<td>$\text{Area}(\text{Voronoi cell around } v)$</td>
</tr>
<tr>
<td>EffAreaCur [69], [11] Page 223</td>
<td>$| \nabla_v V |$</td>
</tr>
<tr>
<td>NormalCur [11] Page 223</td>
<td>$\langle \nabla_v V, \nabla_v A \rangle/| \nabla_v A |$</td>
</tr>
</tbody>
</table>

Recall Remark 2.1 for the effective area. The local area used in ‘normal curvature’ is the length of the projection of $\nabla_v V$ onto the direction of $\nabla_v A$. The rationale for the use of these local areas are discussed in [11][12]; see also Section 3.2.2.
The discrete Willmore energy is then defined as
\begin{equation}
W_{\text{Centroid/Voronoi/EffArea/NormalCur}} = \sum_v H(v)^2 a(v),
\end{equation}
where the four choices of $a(v)$ in the table above correspond to the four discrete $W$-energies. Note that the sign of $H(v)$ is irrelevant to the definition of $W$.

Similarly, a discrete total mean curvature functional can be defined as
\begin{equation}
M_{\text{Cotan}} = \sum_v H(v) a(v) = \frac{1}{2} \sum_v \text{sign}(H(v)) \| \nabla_v A \|.
\end{equation}
Unlike $W$, the sign of $H(v)$, dependent on orientation, is crucial, but the choice of local area is irrelevant.

An alternative discrete total mean curvature, based on Steiner’s polynomial, is defined by
\begin{equation}
M_{\text{Steiner}} := \sum_e \text{length}(e) \theta(e),
\end{equation}
where $\theta(e) \in (-\pi, \pi)$ is the signed angle between the normals to the adjacent faces at $e$; see [69, Section 4.4], [6, Figure 6], [11, Page 227].

Remark 2.3. Besides the area-variation characterization (2.5), we also have the characterization of mean curvature based on the Laplace-Beltrami operator:
\begin{equation}
\Delta_S X(x) = 2H(x), \quad x \in S,
\end{equation}
where $X : S \to \mathbb{R}^3$ is the position function of the surface $S$. It is just a matter of taste to derive a discrete mean curvature based on a discrete Laplace operator or a discrete area variation. For our purpose here, we choose the latter simply because we need the area variation anyway for our solver. In fact, by combining (2.4), (2.7) and (2.11) one can retrieve the cotangent formula for the discrete Laplace-Beltrami operator. For yet another connection of the cotangent formula with area and Dirichlet energy, see Section 4.2. Also, see [70] for the convergence properties of the cotangent formula.

Bobenko’s Willmore energy is based on a rather different philosophy: it is designed to satisfy an exact Möbius invariant property and measures a ‘degree of sphericity’ [6, Proposition 2]. It is defined as
\begin{equation}
W_{\text{Bobenko}} := \frac{1}{2} \sum_v W(v) + 4\pi (1 - g),
\end{equation}
where $W(v) = \sum_{e \ni v} \beta(e) - 2\pi$ and $\beta(e)$ is an angle formed by the circumscribed circles of the two triangles sharing the edge $e$ [6, Definition 1].

A discrete Willmore energy $W_{\text{PL}}$ or total mean curvature $M_{\text{PL}}$ should have a consistency property in the sense that $W_{\text{PL}}(M^n) \to W(S)$ and $M_{\text{PL}}(M^n) \to M(S)$ for any sequence of PL surfaces $M^n$ converging to a smooth surface $S$ in an appropriate sense. $W_{\text{Bobenko}}$ is known to be consistent with the continuous

\[\text{This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.}\]
Willmore energy only in a very restrictive sense [7]. From our preliminary analysis, the other PL Willmore energies are better behaved in terms of consistency. We shall report on these in a separate report. Although such a result is not directly needed in this article, we believe that it will be necessary for the analysis of the PL method proposed in Section 4.

2.2. \( W, A, V, M \) for subdivision surfaces. The two specific subdivision schemes used in our solver are the Loop and \( C^2g0 \) schemes. Here, we present the details of Loop’s scheme [52]; the paper [15] contains similar details for the \( C^2g0 \) scheme. Our presentation will be brief, but contains the necessary implementation details when read in conjunction with the paper [68] by Stam.

2.2.1. Subdivision surfaces. Following the subdivision surface literature, a vertex is called ordinary if it has valence 6, otherwise it is called an extraordinary vertex. We assume that extraordinary vertices in \( M \) are isolated, i.e. no two extraordinary vertices can be neighbor of each other. If \( M \) lacks this property, one can simply apply a mid-point subdivision to \( M \) to resurrect that.

For each triangle face \( f \) in \( M \), we call \( f \) a regular face if all its three bounding vertices are ordinary, otherwise, under our assumption, exactly one of the three vertices is extraordinary and we call \( f \) an irregular face. The corresponding surface patches will be simply referred to as regular and irregular patches. See again Figure 2.

![Figure 3. Iteration of subdivision steps](image)

Although the parametric description is more important for us, it is useful to recall the popular algorithmic description of a Loop surface as the limit of an iteration of subdivision steps, see Figure 3; the subdivision step can be succinctly described by the diagrams in Figure 4.

**Parametrization of a regular patch.** The surface patch associated with a regular face \( f \) can be parameterized by a linear combination of 12 polynomials with coefficients \( \{c_{f,i}\}_{i=1}^{12} \) being the coordinates of the vertices in \( f \) and their immediate neighbors ordered as in Figure 5(b) [see Figure 5].

\[
\mathbb{R}^3 \leftarrow \Omega : s_f(v, w) = \sum_{i=1}^{12} c_{f,i}^T b_i(v, w) \tag{2.12}
\]

where \( \Omega := \{(v, w) : v \in [0, 1] \text{ and } w \in [0, 1-v]\} \), and \( b_1(v, w), \ldots, b_{12}(v, w) \) are the twelve degree 4 polynomials as shown in [68, Page 10-11]. (We do not copy these polynomials from Stam’s paper, but mention that they come from the so-called \( M_{222} \) box-spline.) For notational convenience, we organize each \( c_{f,i} \) as a row vector.

\(^4\)When \( f \) is close to an extraordinary vertex, it is possible that some of these 12 control vertices coalesce. In such a degenerate situation, one simply repeats the coalescing vertices when using the formula (2.12).
Figure 4. (a) edge rule; (b) vertex rule for an ordinary vertex; (c) vertex rule for an extraordinary vertex, where $k$ is the valence and $\beta = \frac{1}{k} \left( \frac{5}{8} - \left( \frac{3}{8} + \frac{1}{4} \cos \left( \frac{2\pi}{k} \right) \right)^2 \right)$. (In the text, we use ‘N’ to denote the valence of an extraordinary vertex, so as to be consistent with Stam’s paper.)

Figure 5. (a) A regular face and its neighboring vertices (b) ordering of the twelve control vertices (c) the parameter domain $\Omega$. Note: for any point $(v, w) \in \Omega$, its barycentric coordinates w.r.t. the bounding vertices of $\Omega$, listed in the order $(0, 0), (1, 0), (0, 1)$, are simply $(1 - v - w, v, w)$. (d) Ordering of the $N + 6$ control vertices around an irregular face (e) Partition of the parameter domain $\Omega$ of length 3, write $c_f := [c_{f,1}; \ldots; c_{f,12}] \in \mathbb{R}^{12 \times 3}$, and define $b := b^6 := [b_1; \ldots; b_{12}]$ as a column vector of functions of length 12. Then (2.12) simplifies to

\begin{equation}
    sf = c_f^T b^6.
\end{equation}

**Parametrization of an irregular patch.** The surface patch associated with an irregular face $f$ admits a parametrization $s_f : \Omega \to \mathbb{R}^3$ which is controlled by
$N+6$ control vertices around the face $f$, where $N$ is the valence of the extraordinary vertex of $f$. Following Stam’s convention, these $N+6$ vertices are ordered as in Figure 5(d). Like the regular case, $s_f$ is linearly related to the control vertices, so
\begin{equation}
    s_f(v, w) = \sum_{i=1}^{N+6} c_{f,i}^T b_i^N(v, w)
\end{equation}
for some basis functions $b_i^N$, $i = 1, \ldots, N + 6$, implicitly defined by the subdivision process. Unlike the regular ($N = 6$) case, none of these basis functions is a single polynomial anymore. Instead, it is an infinite piecewise polynomial, with pieces being the (recursively defined) sub-triangles $\Omega_{k,j}$, $j = 1, 2, \ldots, k = 1, 2, 3$, as shown in Figure 5(e). Note that in this figure the origin $(0,0)$ corresponds to the extraordinary vertex of $f$.

The parametrization (2.14) is tricky to compute; and this is where Stam’s idea comes in. As in the regular case, write $c_f := [c_{f,1}; \ldots; c_{f,N+6}] \in \mathbb{R}^{(N+6) \times 3}$, and $b^N := [b_1^N; \ldots; b_{N+6}^N]$. In a nutshell, Stam’s method transforms the control data $\{c_{f,i}\}$ into ‘eigen- control data’ $\hat{c}_f = V^{-1}c_f$, so
\begin{equation}
    s_f = c_f^T b^N = \hat{c}_f^T V^T b^N = \hat{c}_f^T \Phi.
\end{equation}
Here $V \in \mathbb{R}^{(N+6) \times (N+6)}$ is the matrix of (generalized) eigenvectors of the matrix $A$ (same notation as in Stam’s paper) that maps the $N + 6$ control vertices around $f$ to $N + 6$ control points in the next subdivision level as shown in Figure 6(b), so $AV = VA$ where $\Lambda$ is in a Jordan canonical form. For the Loop scheme, $\Lambda$ is diagonal when the valence $N$ is greater than 3, but has a Jordan block of size 2 when $N = 3$. Since the subdivision process is linear and stationary, i.e. the same linear subdivision rules are used across different scales, recall Figure 3 we have
\begin{equation}
    c_f^T b^N(v, w) = (Ac_f)^T b^N(2v, 2w), \quad (v, w) \in \frac{1}{2} \Omega.
\end{equation}
Putting these together, we have
\begin{equation}
    \Phi(v, w) = \Lambda^T \Phi(2v, 2w), \quad (v, w) \in \frac{1}{2} \Omega.
\end{equation}
The key point is that these eigenbasis functions $\Phi = [\phi_1; \ldots; \phi_{N+6}]$ are easier to evaluate compared to the original basis functions $b^N$: When $N > 3$, $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{N+6})$ is diagonal, and we have $\phi_i(v, w) = \lambda_i \phi(2v, 2w)$; applying this recursively we have
\begin{equation}
    \phi_i(v, w) = \lambda_i^{n-1} \phi(2^{n-1}v, 2^{n-1}w), \quad \text{when } (v, w) \in \Omega_k^n.
\end{equation}
(Recall Figure 5(e).) As a result, each $\phi_i$ is specified by three – not infinitely many – polynomials. Also, in virtue of (2.17), $\phi_i$ and its derivatives can be easily evaluated at arbitrary parameter values after the three polynomials are specified.

These three polynomials can be evaluated based on the same polynomial basis $b$ from the regular case (2.12). We write $V_i$ as the (generalized) eigen-vector associated to the eigen-basis function $\phi_i$. For $k = 1, 2, 3$, there are suitable linear maps $M_k$ (expressed as $P_k \bar{A}$ in Stam’s paper) so that $M_k V_i$ contains the data at the 12 control vertices that determine the polynomial $\phi_i|_{\Omega_k^n}$; see Figure 6(c)-(e). With an appropriate affine reparametrization of $\Omega^n_k$ by $\Omega$, denoted as $t_{1,k} : \Omega^n_k \to \Omega$ by Stam,
this polynomial can be expressed as

\[ \phi_i|_{\Omega^k_1}(v, w) = (M_k V_i)^T b(t_{1,k}(v, w)). \]  

(2.18)

See [68] for details, e.g. on how to exploit the circulant structure in the matrix \( A \) in order to facilitate the computation of the related matrices \( V, V^{-1}, M_k \).

Figure 6. (a) the \( N+6 \) control vertices around the extraordinary face \( f \) that determine \( \phi_i : \Omega \rightarrow \mathbb{R} \) (b) the \( N+6 \) control vertices at the next subdivision level that determine \( \phi_i|_{\Omega^2} \) (c)-(e) the 12 control vertices at the next subdivision level that determine \( \phi_i|_{\Omega^k_1}, k = 1, 2, 3 \)

2.2.2. Computation of \( W(\mathcal{V}), A(\mathcal{V}), V(\mathcal{V}), M(\mathcal{V}) \) and their gradients. To summarize the previous section, a regular patch of a Loop subdivision surface is parameterized by a single degree 4 polynomial on the reference triangle \( \Omega \) for each of the three spatial components, whereas an irregular patch admits a more complicated parametrization over \( \Omega \). In either case, an efficient algorithm exists for evaluating the parametrization and its derivatives at arbitrary parameter values \((v, w) \in \Omega\).

Armed with such evaluation algorithms, we now see how the various functionals and their gradient vectors in (2.1) can be computed.

**Formulas for \( A(\mathcal{V}) \) and \( \nabla A(\mathcal{V}) \).** We first discuss how to compute the area \( A(\mathcal{V}) \) of a Loop surface and the gradient \( \nabla A(\mathcal{V}) \) of \( A \) with respect to \( \mathcal{V} \). For each regular face \( f \) in a control mesh, we write \( A_6(\mathbf{c}_f) \) as the area of the surface patch associated to \( f \); and we view \( A_6 \) as a (real-valued) function of the variables in the array \( \mathbf{c}_f \). Similarly, we write \( A_N(\mathbf{c}_f) \) as the area of the surface patch associated to an irregular face \( f \) with an extraordinary vertex of valence \( N \neq 6 \); in this case, we also write \( \text{val}(f) := N \).

\[
A(\mathcal{V}) = \sum_{\text{area of regular patches}} A_6(\mathbf{c}_f) + \sum_{N \neq 6 : \text{val}(f) = N} \sum_{\text{area of irregular patches}} A_N(\mathbf{c}_f). 
\]

(2.19)

Recall from (2.13) and (2.15), the parametrization \( \mathbf{s}_f \) can be written as

\[
\mathbf{s}_f = \mathbf{c}_f^T \mathbf{b}^N, \text{ where } N = \text{val}(f).
\]

(2.20)

Note that \( \mathbf{c}_f \) is linearly related to \( \mathcal{V} \), whereas the basis functions \( \mathbf{b}^N \) are independent of \( \mathcal{V} \). These features allow for an accurate and efficient computation of \( A(\mathcal{V}) \) and \( \nabla A(\mathcal{V}) \), as we shall now see.

We define the map \( P_f \) by

\[
P_f \mathcal{V} = \mathbf{c}_f
\]

(2.21)
which picks out the local control data around the face \( f \) from the global control data \( \mathcal{V} \).

For convenience, drop the subscript and write \( s \) instead of \( s_f \). Also, we write \( s_1, s_2, s_3 \) for the components of \( s \) and \( s_{i,u}, s_{i,v} \) for their partial derivatives.

Note that

\[
A_N(c_f) = \int_{\Omega} \| n(v, w) \| \, dv \, dw, \quad \text{where}
\]

\[
n = \frac{\partial s}{\partial v} \times \frac{\partial s}{\partial w} = [s_{2,w}s_{3,v} - s_{2,v}s_{3,w}, \ s_{3,v}s_{1,w} - s_{3,w}s_{1,v}, \ s_{1,w}s_{2,v} - s_{1,v}s_{2,w}].
\]

Again, we drop the subscript \( f \) and write \( c_1, c_2, c_3 \) to refer to the columns of \( c_f \).

When \( f \) is a regular face, \( s_i = c_i^T b \), so

\[
s_{i,u} = c_i^T b_v, \quad s_{i,v} = c_i^T b_w, \quad i = 1, 2, 3.
\]

The gradient of \( n_1 \) with respect to \( c_f \), organized as a \( 12 \times 3 \) array (i.e. same dimension as \( c_f \)), can be expressed as\(^5\)

\[
\nabla_{c_f} n_1 = \nabla_{c_f} (s_{2,w}s_{3,v} - s_{2,v}s_{3,w})
= [0, b_v, 0]s_{3,v} + s_{2,v}[0, 0, b_w] - [0, b_w, 0]s_{3,w} - s_{2,w}[0, 0, b_v]
= [0, s_{3,w}b_v - s_{3,v}b_w, s_{2,v}b_w - s_{2,w}b_v].
\]

Similarly,

\[
\nabla_{c_f} n_2 = [-s_{3,w}b_v + s_{3,v}b_w, 0, -s_{1,v}b_w + s_{1,w}b_v],
\]

\[
\nabla_{c_f} n_3 = [s_{2,w}b_v - s_{2,v}b_w, s_{1,w}b_w - s_{1,v}b_v, 0].
\]

Next, we have

\[
\| n \| = \sqrt{\langle n, n \rangle},
\]

\[
\nabla_{c_f} \| n \| = \frac{1}{\langle n, n \rangle^{1/2}} \nabla_{c_f} \langle n, n \rangle
= \frac{1}{\langle n, n \rangle^{1/2}} (n_1 \nabla_{c_f} n_1 + n_2 \nabla_{c_f} n_2 + n_3 \nabla_{c_f} n_3).
\]

Therefore, the local area functional \( A_N \) and its gradient

\[
\nabla A_N(c_f) = \int_{\Omega} \nabla_{c_f} \| n(v, w) \| \, dv \, dw,
\]

can be computed based on the control data \( c_f = [c_1, c_2, c_3] \) and the basis function \( b \) via (2.23)-(2.26). Together with (2.21) and the chain rule, we can compute the total area and its gradient with respect to \( \mathcal{V} \) by

\[
A(\mathcal{V}) = \sum_{N} \sum_{f: \text{val}(f) = N} A_N(P_f \mathcal{V}), \quad \nabla A(\mathcal{V}) = \sum_{N} \sum_{f: \text{val}(f) = N} P_f^T \nabla A_N(c_f).
\]

\(^5\)Here and below, we have to deal with a number of scalar quantities \( S \) that vary with both the local control data \( c_f \) and parameter values \( (v, w) \) (e.g. \( \| n \|, n_i, E, F, G, e, f, g, \text{etc.} \)). In order to avoid confusion, we use the notation \( \nabla_{c_f} S \) to denote the gradient vector of \( S \) viewed as a function of \( c_f \); the gradient ‘vector’ is structured as an array of the same size as \( c_f \). Likewise, the gradient ‘vectors’ \( \nabla W(\mathcal{V}), \nabla A(\mathcal{V}), \nabla V(\mathcal{V}), \nabla M(\mathcal{V}) \) are structured as \#\( \mathcal{V} \times 3 \) arrays, i.e. the same dimensions as \( \mathcal{V} \).
Formulas for $W(V)$, $V(V)$, $M(V)$, and their gradients. Similar to $A(V)$, we aim to express the other three functionals in the C-H model in terms of $c_f$ and the basis functions $b$ and $Φ$. For $W$ and $M$, we need an expression for the mean curvature. Recall that

$$E = \langle s_{f,v}, s_{f,v} \rangle, \quad F = \langle s_{f,v}, s_{f,w} \rangle, \quad G = \langle s_{f,w}, s_{f,w} \rangle$$

represent the first fundamental form of the surface $s_f$, whereas

$$e = \langle s_{f,vv}, n \rangle / \|n\|, \quad f = \langle s_{f,vw}, n \rangle / \|n\|, \quad g = \langle s_{f,ww}, n \rangle / \|n\|$$

represent the second fundamental form. The mean curvature can be expressed as

$$H = \frac{eG - 2ff + gE}{2(EG - F^2)}.$$  

Therefore

$$M = - \iint H \, dA = - \sum_{f \in F} \iint_{\Omega} \frac{eG - 2ff + gE}{2(EG - F^2)} \|n\| \, dv \, dw$$

$$= - \sum_{f \in F} \iint_{\Omega} \frac{\bar{e}G - 2\bar{f}F + \bar{g}E}{2(EG - F^2)} \, dv \, dw,$$

where $\bar{e} := \langle s_{f,vv}, n \rangle$, $\bar{f} := \langle s_{f,vw}, n \rangle$, $\bar{g} := \langle s_{f,ww}, n \rangle$. Similarly,

$$W = \iint H^2 \, dA = \sum_{f \in F} \iint_{\Omega} \left[ \frac{eG - 2ff + gE}{2(EG - F^2)} \right]^2 \|n\| \, dv \, dw.$$

Remark 2.4. When $f$ is an irregular face, the mean curvature can potentially blow up when approaching the extraordinary vertex, however it is proved in [61] that the corresponding integrals (called $M_N$ and $W_N$ below) are always finite.

By the divergence theorem, with the choice of the vector field $\vec{X}(x, y, z) = xi + yj + zk$, the volume enclosed by a surface can be expressed as a surface integral:

$$V = \frac{1}{3} \iiint \text{div} \vec{X} \, dx \, dy \, dz = \frac{1}{3} \iiint \vec{X} \cdot n \|n\| \, dA = \frac{1}{3} \sum_{f \in F} \iint_{\Omega} \langle s_f, n \rangle \, dv \, dw.$$  

By (2.20), we can express the rightmost integral in (2.30)-(2.32) in terms of $c_f$ and $b_N$ where $f$ ranges over all faces and $N$ ranges over all valences existing in the control mesh; we denote the integral by $M_N(c_f)$, $W_N(c_f)$ and $V_N(c_f)$, respectively.

We explain how to compute the gradients of $M_N(c_f)$, $W_N(c_f)$ and $V_N(c_f)$.
The gradients of $E$, $F$, $G$, $\bar{e}$, $\bar{f}$, $\bar{g}$ can be computed as follows

\begin{equation}
\nabla_{c_i} E = \sum_{i=1,2,3} \nabla_{c_i} (c_i^T b_v) = 2 [s_{1,v} b_v, s_{2,v} b_v, s_{3,v} b_v],
\end{equation}

\begin{equation}
\nabla_{c_i} F = \sum_{i=1,2,3} \nabla_{c_i} (c_i^T b_w) = [s_{1,w} b_w + s_{1,w} b_v, s_{2,w} b_w + s_{2,w} b_v, s_{3,w} b_w + s_{3,w} b_v],
\end{equation}

\begin{equation}
\nabla_{c_i} G = \sum_{i=1,2,3} \nabla_{c_i} (c_i^T b_w) = 2 [s_{1,w} b_w, s_{2,w} b_w, s_{3,w} b_w],
\end{equation}

\begin{equation}
\nabla_{c_i} \bar{e} = \sum_{i} \nabla_{c_i} (c_i^T b_{uv}) n_i = [s_{1,u} \nabla_{c_i} n_1, s_{2,u} \nabla_{c_i} n_2, s_{3,u} \nabla_{c_i} n_3] + [n_1 b_{uv}, n_2 b_{uv}, n_3 b_{uv}],
\end{equation}

\begin{equation}
\nabla_{c_i} \bar{f} = [s_{1,u} \nabla_{c_i} n_1, s_{2,u} \nabla_{c_i} n_2, s_{3,u} \nabla_{c_i} n_3] + [n_1 b_{uv}, n_2 b_{uv}, n_3 b_{uv}],
\end{equation}

\begin{equation}
\nabla_{c_i} \bar{g} = [s_{1,u} \nabla_{c_i} n_1, s_{2,u} \nabla_{c_i} n_2, s_{3,u} \nabla_{c_i} n_3] + [n_1 b_{uv}, n_2 b_{uv}, n_3 b_{uv}].
\end{equation}

By \(2.30\),

\begin{equation}
\nabla M_N = \nabla_{c_i} \int_\Omega \left[ \frac{\bar{e} G - 2 \bar{f} F + \bar{g} E}{2 (EG - F^2)} \right] dv dw = \int_\Omega \nabla_{c_i} \bar{H} dv dw.
\end{equation}

Thanks to \(2.33\), the integrand $\nabla_{c_i} \bar{H}$ can be computed by the product and quotient rules. After $\nabla M_N$ is computed, $\nabla W_N$ can then be computed by

\begin{equation}
\nabla W_N = \nabla_{c_i} \int_\Omega \left[ \frac{\bar{e} G - 2 \bar{f} F + \bar{g} E}{2 (EG - F^2)} \right] dv dw = \int_\Omega \nabla_{c_i} \bar{H} dv dw.
\end{equation}

Note that every term in the integrand of \(2.35\) was computed previously.

Finally, for $V_N$ we have

\begin{equation}
\nabla V_N = \nabla_{c_i} \int_\Omega \langle s_f, n \rangle dv dw = \int_\Omega \sum_i \nabla_{c_i} (c_i^T b) n_i dv dw
\end{equation}

\begin{equation}
= \int_\Omega \{ [n_1 b, n_2 b, n_3 b] + [s_1 \nabla_{c_i} n_1 + s_2 \nabla_{c_i} n_2 + s_3 \nabla_{c_i} n_3] \} dv dw.
\end{equation}

With all the local functionals and their gradients with respect to the local control data computed, the global functionals and their gradients with respect to the global control data can be computed exactly as in \(2.28\):

\begin{equation}
M(V) = \sum_{N} \sum_{f: \text{val}(f) = N} M_N(P_f V),
\end{equation}

\begin{equation}
\nabla M(V) = \sum_{N} \sum_{f: \text{val}(f) = N} P_f^T \nabla M_N(c_f),
\end{equation}

and similarly for $W(V)$, $\nabla W(V)$, $V(V)$, $\nabla V(V)$. 
2.3. **Implementation details.** In the actual numerical computation of $A(V)$ and $\nabla A(V)$ based on (2.28), we use a symmetric 7-point Gauss quadrature rule on a triangle, with accuracy order 5 (see for example [13]), to approximate the integrals in $A_N(c_f), V_N(c_f), M_N(c_f), W_N(c_f)$, and their gradients (recall (2.27), (2.34)-(2.36)). The approximation is done using a composite quadrature on $\Omega$ using a uniform grid of size $1/n$. Typically we use $n = 8$ or $16$ at coarse subdivision levels, and $n = 1$ or $2$ at fine levels. We choose $n$ to be a dyadic integer in order to take advantage of the subdivision structure of $s_f$ when $N \neq 6$.

A key implementation detail is that quantities independent of $V$ are precomputed before entering the optimization loop. These quantities include $b^6(u, v)$ and $b^N(u, v) = V_N^{-T} \Phi_N(u, v)$ (for only those extraordinary valences $N$ that show up in $\mathcal{F}$) and their derivatives at the quadrature points, evaluated using Stam’s algorithm (Sections 2.2.1). A separate pre-processing step computes the maps $P_f$ for each face $f$ in $\mathcal{F}$, as $P_f$ depends only on the connectivity information in $\mathcal{F}$. Of course, one should not store $P_f$ as a $(N + 6) \times \#V$ (dense) matrix as suggested by (2.21). Instead, it suffices to store the information in $P_f$ by a list of $N + 6$ integer indices in $\{1, \ldots, \#V\}$ which keeps track of the indices of $c_f$ in the global vertex list $V$; we denote this list of vertex indices by $VFL(f)$. These preprocessing steps speed up our solver significantly already in a sequential implementation. See Figure 7 (top) for the basic organization of our SS solver.

---

**Figure 7.** (top) computation structure of the SS solver (bottom) architecture of the GPU enhanced solver
We note that both the PL and SS functionals are quite easily parallelized; we provide in WMINCON parallel CUDA implementations of all the functionals considered in this article. Together with the aforementioned precomputation trick, we are able to solve a lot of instances of Willmore, Canham and Helfrich problems not addressed in [32]. Figure 7 (bottom) illustrates the basic architecture of our GPU enhanced solver for both the PL and SS methods.

For our implementation of the solver, we primarily use SNOPT [34] and the ‘SQP’ option of fmincon in the Matlab optimization toolbox. These solvers are designed for smooth objectives and constraints, which is the case for the SS method and most of the PL methods. An interesting exception is the case of \( W_{Bobenko} \), which is non-smooth whenever the PL surface has two adjacent triangles of which the four vertices lie on a common circle. In this case, our use of the non-smooth optimization algorithm GRANSO [18] was instrumental to the discovery of Proposition 3.9 below.

All three solvers use a quasi-Newton (BFGS) algorithm, coupled with the sequential quadratic programming (SQP) method [58, Chapter 18] for handling constraints. Being all BFGS-SQP-based algorithms, they have decisively different properties even for smooth problems, see Section 5.

3. Analysis

In this section, we give a theoretical justification for why the SS methods succeed in solving the Willmore problem (Section 3.1), but that the PL methods presented in Section 2 are bound to fail (Section 3.2). For computational experiments illustrating the results in this section, see the companion conference paper [12].

3.1. Convergence of SS method. In this section, we establish Theorem 3.8. The argument is based on the existence result for Willmore surfaces established using the direct method in the calculus of variations (Theorem 3.2) [3, 47, 48, 66], and an observation (Proposition 3.1) connecting the existence result to our conforming subdivision methods; the connection requires a density result (Theorem 3.5) from the theory of subdivision surfaces.

A sequence \( (x^k)_{k=0}^{\infty} \) in a space \( \mathcal{X} \) is called a minimizing sequence for a functional \( \mathcal{E} : \mathcal{X} \rightarrow \mathbb{R} \) if

\[
\lim_{k \rightarrow \infty} \mathcal{E}(x^k) = \inf_{x \in \mathcal{X}} \mathcal{E}(x).
\]

Proposition 3.1. Let \( \mathcal{E} : \mathcal{X} \rightarrow \mathbb{R} \) be a continuous functional on a topological space \( \mathcal{X} \). Assume that we have a nested sequence of subspaces \( \{\mathcal{S}^j : j = 0, 1, 2, \ldots\} \) of \( \mathcal{X} \) such that \( \bigcup_j \mathcal{S}^j \) is dense in \( \mathcal{X} \). Assume that a minimizer \( x^* \in \text{argmin}_{x \in \mathcal{X}} \mathcal{E}(x) \) exists. Then any sequence of ‘approximate \( \mathcal{E} \)-minimizers’ \( x^j \in \mathcal{S}^j \), i.e.

\[
\mathcal{E}(x^j) = \inf_{x \in \mathcal{S}^j} \mathcal{E}(x) + o(1), \quad j \rightarrow \infty,
\]

is a minimizing sequence for \( \mathcal{E} \).

Proof. By the denseness assumption, there exists a sequence \( \tilde{x}^j \in \mathcal{S}^j \) converging to \( x^* \) in \( \mathcal{X} \). By continuity, \( \mathcal{E}(\tilde{x}^j) \rightarrow \mathcal{E}(x^*) \). But we also have

\[
\inf_{x \in \mathcal{X}} \mathcal{E}(x) \leq \mathcal{E}(x^j) = \inf_{x \in \mathcal{S}^j} \mathcal{E}(x) + o(1) \leq \mathcal{E}(\tilde{x}^j) + o(1),
\]

so \( \mathcal{E}(x^j) \rightarrow \mathcal{E}(x^*) = \inf_{x \in \mathcal{X}} \mathcal{E}(x) \). Thus, \( (x^j)_{j=0}^{\infty} \) is a minimizing sequence for \( \mathcal{E} \). \( \square \)
Note how this result relies on the conforming and dense nature of the spaces $S^j$. Notice also that the use of approximate minimizers frees us from the assumption that a minimizer of $E$ over each $S^j$ exists.

We now state the key existence result established in [3,48,66]. Since we are analyzing a parametric method, it is more convenient to state the result in terms of parametrizations. For this purpose, we assume $\Sigma$ is any (reference) genus $g$ surface with a smooth enough differentiable structure. A $C^{1,1}$ differentiable structure would suffice for our purpose, as it is enough to support the definition of not only $C^1(\Sigma,\mathbb{R}^3)$ and $C^{1,\alpha}(\Sigma,\mathbb{R}^3)$, $0 < \alpha \leq 1$, but also the whole range of Sobolev spaces $W^{2,p}(\Sigma,\mathbb{R}^3)$, $p \in [1,\infty)$. (See [1] and [31, Section 4.2.3].) We shall work with the Banach space

$$X := X(\Sigma) := C^1(\Sigma,\mathbb{R}^3) \cap W^{2,2}(\Sigma,\mathbb{R}^3)$$

(normed by $\|x\|_X = \|x\|_{C^1} + \|x\|_{W^{2,2}}$) and the nonlinear, open subspace

$$\text{Imm}_X := \text{Imm}_X(\Sigma) := \{f \in C^1(\Sigma,\mathbb{R}^3) \cap W^{2,2}(\Sigma,\mathbb{R}^3) \mid \text{rank}(df_x) = 2, \forall x \in \Sigma\},$$

on which the Willmore energy

$$W : \text{Imm}_X \to [0,\infty)$$

is well-defined and continuous.

The following is a reformulation of the well-known existence result pioneered by L. Simon [66] and completed in [3,48].

**Theorem 3.2** (Existence of $W$-minimizer of genus $g$, via the direct method). Let $\Sigma$ be a closed orientable surface of genus $g$ with a $C^{1,1}$ differentiable structure. For any minimizing sequence $x_j \in \text{Imm}_X$ for $W$, there exists a subsequence $x_{j'}$, and a sequence of Möbius transformations $G_{j'}$ in $\mathbb{R}^3$, such that the sequence of immersed surfaces $G_{j'}(x_{j'}(\Sigma))$ converges in Hausdorff distance to an immersed surface $x_*(\Sigma)$, $x_* \in \text{Imm}_X$. As such, $W(x_*) = \inf_{x \in \text{Imm}_X} W(x)$.

**Remark 3.3.** The Willmore minimizers are known to be embedded surfaces. We choose to work with general immersed surfaces because in our numerical method we do not have any mechanism built in to avoid self-intersections; another reason is that the solutions of the Helfrich problem for some values of $\nu_0$ and $m_0$ do have self-intersections.

**Remark 3.4.** While [3,48,66] prove the existence of a $W$-minimizer over all *infinitely smooth* genus $g$ immersed surfaces, our minimization space is taken to be the bigger $\mathcal{B}_r$. This causes no problem as $C^\infty(\Sigma,\mathbb{R}^3) \cap \text{Imm}_X$ is dense in $\text{Imm}_X$ and $W : \text{Imm}_X \to \mathbb{R}$ is continuous. (If $E : A \to \mathbb{R}$ is a continuous functional on a topological space $A$, and $B$ is a dense subspace of $A$, then any minimizer of $E : B \to \mathbb{R}$ must also be a minimizer of $E : A \to \mathbb{R}$.) Here, the definition of $C^\infty(\Sigma,\mathbb{R}^3)$, with $\Sigma$ endowed with a $C^{1,1}$ differentiable structure, requires the fact that a maximal $C^1$-atlas contains a $C^\infty$-atlas. In fact there is a real analytic sub-atlas[6] with respect to which the minimizer $x_*$ in Theorem 3.2 is real-analytic [66].

---

[6]The result that every $C^1$ manifold admits compatible $C^\infty$ and analytic ($C^\omega$) structures is due to Whitney [72].
The Loop subdivision scheme defines:
- a $C^2$-compatible atlas, via its characteristic maps, on any base complex $K$,
- a linear space of scalar-valued subdivision functions $S(K^j)$ at each subdivision level.

The atlas given by the characteristic maps of Loop’s scheme turns $K$ into a $C^2$-manifold. For details, see [1]. The subdivision functions are well-known to be in $C^1(K, \mathbb{R})$ [60,75,79]. By a trivial extension of [1, Theorem 42], we also have

\[ S(K^j) = S(K^j, \mathbb{R}) \subset W^{2,p}(K, \mathbb{R}), \quad \forall p \in [1, p^*), \]

where $p^* > 2$ is given by (3.2) below.

When we extend these subdivision functions componentwise to map into $\mathbb{R}^m$, we denote the corresponding linear space by $S(K^j, \mathbb{R}^m)$.

Arden’s thesis [1] essentially establishes the following result. While he focuses on the $p = 2$ case, the proof can be extended to a range of values of $p$ that is strictly bigger than the interval $[1,2]$.

**Theorem 3.5.** There exists a $p^* > 2$ such that for any $p \in [1, p^*)$, the space of Loop subdivision functions at all levels $\bigcup_{j=0}^{\infty} S(K^j, \mathbb{R})$ is dense in $W^{2,p}(K, \mathbb{R})$.

**Remark 3.6.** The value of $p^*$ above depends on the set of valences present in the underlying simplicial complex, denoted by $\text{val}(\mathcal{F})$, and the values of the sub- and sub-sub-dominant eigenvalues, $\lambda_n$ and $\mu_n$, of the subdivision matrix $A$ corresponding to the different valences $n \in \text{val}(\mathcal{F})$. (For the definition of $A$, recall the comments after (2.15) in Section 2.2.1.) We have

\[
\begin{align*}
p^* &= \min_{n \in \text{val}(\mathcal{F})} \frac{2 \log(\lambda_n)}{\log(\lambda_n^2/\mu_n)}, \quad \lambda_n = \frac{3}{8} + \frac{1}{4} \cos\left(\frac{2\pi}{n}\right) \quad \text{and} \\
\mu_n &= \begin{cases} 
\frac{1}{8} & \text{if } n = 3, \\
\lambda_n^2 & \text{if } n = 4, 5, 6, \\
\frac{3}{8} + \frac{1}{4} \cos\left(\frac{4\pi}{n}\right) & \text{if } n \geq 7.
\end{cases}
\end{align*}
\]

(3.2)

It is easy to check that $p^*$ approaches 2 from above as $\max(\text{val}(\mathcal{F})) \to \infty$. For details, see [75].

In particular, Theorem 3.5 implies, by Morrey’s inequality, that $\bigcup_{j=0}^{\infty} S(K^j, \mathbb{R}^3)$ is dense in $X = C^1(K, \mathbb{R}^3) \cap W^{2,2}(K, \mathbb{R}^3)$.

For notational simplicity, we write

\[ S^j = S(K^j, \mathbb{R}^3), \quad \text{Imm}_{S^j} := \text{Imm}_{S^j}(K) := S(K^j, \mathbb{R}^3) \cap \text{Imm}_X(K). \]

**Corollary 3.7.** $\bigcup_j \text{Imm}_{S^j}$ is dense in $\text{Imm}_X(K)$.

**Proof.** Since $\bigcup_j S^j$ is dense in $X$, and $\text{Imm}_X$ is open in $X$,

\[ \text{Imm}_X \cap \bigcup_j S^j = \bigcup_j \text{Imm}_{S^j} \]

is dense in $\text{Imm}_X$.

\[ \square \]

7In fact, $\text{Imm}_{S^j}$ is dense in $S^j$ for each $j$, so, together with the consequence of Arden’s result, we conclude that $\bigcup_j \text{Imm}_{S^j}$ is dense in $X(K)$. This stronger result requires extra arguments based on known results in the theory of subdivision surface and an application of Sard’s theorem.
Combining Corollary 3.7 and Proposition 3.1 above, we have the conclusion that
if \( x_j \) is an approximate \( W \)-minimizer over \( \text{Im}m_{g,j} \) in the sense of Proposition 3.1
then the sequence \((x_j)_j\) is a minimizing sequence in the setting of Theorem 3.2. So by Theorem 3.2, we have the following.

**Theorem 3.8.** Let \( \text{Im}m_{g,j} \) be the space of immersed Loop subdivision surfaces on
the \( j \)-times subdivided complex \( K^j \) of a genus \( g \) complex \( K = K^0 \) (as above). If
\( x_j \) is an approximate \( W \)-minimizer over \( \text{Im}m_{g,j} \) in the sense of Proposition 3.1
then there is a subsequence \( x_j' \), so that the surfaces \( x_j'(K) \), with suitable Möbius transformations applied, converge in Hausdorff distance to a surface \( x_*(K) \) with
\( x_* \in \text{Im}m_X \). This \( x_*(K) \) is a genus \( g \) Willmore minimizer.

This result should explain the empirical success of the SS method; see Section 6
for a possible improvement of this result. In contrast, we next illustrate why the PL \( W \)-minimizers typically have nothing to do with a continuous Willmore minimizer.

### 3.2. Failure of naive PL methods. We begin by establishing a negative result
for Bobenko’s Willmore energy, which may sound like an unworthy result given its
restrictive consistency property. But here we work in the setting most favorable for
\( W_{\text{Bobenko}} \) in terms of consistency, namely the setting for a regularly triangulated
torus. We know from [7] that by successively finer regular sampling and triangulation
of a toroidal Dupin cyclide at its curvature lines one obtains PL tori with
\( W_{\text{Bobenko}} \)-energy converging to the Willmore energy of the Dupin cyclide. This applies in particular to the Clifford tori (i.e. Dupin cyclides gotten from all possible
Möbius transformations of the torus of revolution with radii ratio \( 1 : \sqrt{2} \)), which are the only minimizers of the genus 1 Willmore problem [54]. However, we show
below that we will never get an approximation of a Clifford torus by minimizing
\( W_{\text{Bobenko}} \).

The construction used in this proof happens to be applicable, after a twist, to
proving a similar negative result for other PL \( W \)-energies.

3.2.1. \( W_{\text{Bobenko}} \). For every grid size \( m \) and \( n \), we define a family of triangulated
tori, denoted by \( T_{m,n,\varepsilon}, \varepsilon \in (0,\pi/2) \), with the following properties:

1. It has \( mn \) vertices, \( 2mn \) triangles and all vertices have valence 6.
2. All vertices lie on a sphere.
3. The diameter of the ‘tunnel’ of the torus goes to 0 as \( \varepsilon \downarrow 0 \).
4. The ‘tunnel’ of the torus is triangulated by \( 2n \) (long and skinny) triangles.

See the first four panels of Figure 8. In details, it is defined as follows. Let \( \Delta u = 2\pi/m, \Delta v = 2\pi/n \). For \( i = 0, \ldots, m - 1 \), \( j = 0, \ldots, n - 1 \), let

\[
(3.3) \quad V_{i,j} = \begin{bmatrix}
\sin(\tau(i\Delta u)) \cos(\tau(j\Delta v)) \\
\sin(\tau(i\Delta u)) \sin(\tau(j\Delta v)) \\
\cos(\tau(i\Delta u))
\end{bmatrix}, \quad \text{where } \tau(u) = \frac{(\pi - 2\varepsilon)u}{2\pi(1 - 1/m)} + \varepsilon.
\]

Note that \( \tau \) maps \([0, 2\pi(1 - 1/m)]\) to \([\varepsilon, \pi - \varepsilon]\). These \( m \cdot n \) points on the unit sphere are the vertices of \( T_{m,n,\varepsilon} \). For the triangulation, each \( V_{i,j} \) is connected to the six neighbors \( V_{i+1,j}, V_{i+1,j+1}, V_{i,j+1}, V_{i-1,j}, V_{i-1,j-1}, V_{i,j-1} \), forming also six triangular faces incident on the vertex. In above, the \(+/-\) in the first and second
indices are modulo \( m \) and modulo \( n \), respectively.

Note that the \( 2n \) vertices \( V_{0,j} \) and \( V_{m-1,j}, j = 0, \ldots, n - 1 \), are the vertices closest
to the north and south pole, respectively. While these two groups of vertices are
Figure 8. Various spherical tori ((a)-(d)) and flattened tori based on sphere inversions ((e)-(f))

the furthest apart geometrically, they are connected and form a vertical tunnel of the torus with $2n$ long skinny triangles.

Denote by $\mathcal{T}_{m,n}$ the set of all PL surfaces with connectivity of a $(m, n)$-regularly triangulated torus and, following [6], for any discrete Willmore energy $W_{\text{PL}}$ for PL surfaces, write

$$W_{\text{PL}}(\mathcal{T}_{m,n}) = \inf_{S \in \mathcal{T}_{m,n}} W_{\text{PL}}(S).$$

**Proposition 3.9.** For any grid size $m, n \geq 3$, $W_{\text{Bobenko}}(\mathcal{T}_{m,n,\varepsilon})$ decreases monotonically to $4\pi$ as $\varepsilon \downarrow 0$. Hence $W_{\text{Bobenko}}(\mathcal{T}_{m,n}) \leq 4\pi$ regardless of $m, n$.

**Proof.** We divide the proof into three steps.

1° For any triangulated torus $\mathcal{T}_{m,n,\varepsilon}$ defined above,

$$W_{\text{Bobenko}}(\mathcal{T}_{m,n,\varepsilon}) = \frac{1}{2} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} W_{i,j},$$

where $W_{i,j} = \sum_{(i',j') \sim (i,j)} \beta(i', j') - 2\pi$

and $\beta(i', j')$ is an angle formed by the circumscribed circles of the two triangles sharing the edge connecting $(i, j)$ to $(i', j')$ [6, Definition 1]. We then notice that

$$W_{i,j} = 0, \quad \forall i \neq 0, m - 1,$$

since the six vertices around $(i, j)$ form a convex neighborhood lying on a common sphere [6, Proposition 1]. By symmetry, $W_{0,j}$ and $W_{m-1,j}$ share the same value for all $j = 0, \ldots, n - 1$. So we have

$$W_{\text{Bobenko}}(\mathcal{T}_{m,n,\varepsilon}) = n W_{0,0}.$$

It remains to show that $W_{0,0}$ decreases monotonically to $4\pi/n$ as $\varepsilon \downarrow 0$.

2° To simplify computation, we take advantage of the Möbius invariance of $W_{\text{Bobenko}}$ by applying a sphere inversion that maps the unit sphere to the $z = 0$ plane. Specifically, we invert about the sphere with radius $\sqrt{2}$ and centered at the north pole $[0, 0, 1]^T$ of the unit sphere. This maps the south pole of the unit sphere
to the origin, and the north pole to infinity; and it turns our spherical torus \( T_{m,n,\varepsilon} \) to a flattened torus. See Figure 8.

Among the 7 vertices \( V_0, V_1, V_{-1}, V_{-1,1}, V_{0,1}, V_{1,1} \) contributing to \( W_0, V_0, V_1, V_{-1} \) are close to the north pole, they are sphere inverted to points far away from the origin; \( V_{-1}, V_{-1,1} \) are close to the south pole, they are mapped to points close to the origin. The last two neighbors \( V_0, V_{1,1} \) are at an approximately constant distance from the north pole: their common polar angle is uniformly larger than, and approaches, \( \pi/(m-1) \) as \( \varepsilon \to 0 \). Therefore, these 7 vertices are mapped to \( V_i' \in \mathbb{R}^2, i = 0, 1, \ldots, 6 \) (in the same cyclic order), with the form

\[
V_0' = \rho_1(\varepsilon)[1,0], \quad V_1' = \rho_1(\varepsilon)[c,s], \quad V_2' = \rho_2(\varepsilon)[1,0], \quad V_3' = \rho_2(\varepsilon)[c,-s],
\]

\[
V_4' = \rho_1(\varepsilon)[c,-s], \quad V_5' = \rho_3(\varepsilon)[1,0], \quad V_6' = \rho_3(\varepsilon)[c,s],
\]

where \( \rho_1(\varepsilon) \gg \rho_2(\varepsilon) \gg \rho_3(\varepsilon) \), as \( \varepsilon \to 0 \). By scale invariance, \( W_{0,0} \) depends on \( \varepsilon \)

through

\[
(3.5) \quad \epsilon_1 := \rho_3(\varepsilon)/\rho_1(\varepsilon), \quad \text{and} \quad \epsilon_2 := \rho_2(\varepsilon)/\rho_1(\varepsilon).
\]

Clearly, \( \rho_1 \) (resp. \( \rho_2 \)) increases (resp. decreases) as \( \varepsilon \) decreases. So \( \epsilon_2(\varepsilon) \) is monotonically increasing for \( \epsilon \in (0, \pi/2) \).

Below, we only need the facts that \( 1 > \epsilon_1 > \epsilon_2 > 0 \), \( \epsilon_2 \) is monotone in \( \varepsilon \) and \( \epsilon_1, \epsilon_2 \to 0 \) as \( \varepsilon \to 0 \).

\( 3^\circ \) To calculate \( W_{0,0} = \sum_{i=1}^6 \beta_i - 2\pi \) we analyze each angle \( \beta_i \) between the circumcircles of the two triangles sharing the edge \( e_i = V_i'V_i'' \), now thought of as functions of \( (\epsilon_1, \epsilon_2) \).

We first notice that the two triangles \( V_0'V_2''V_3, V_0'V_4''V_4' \) sharing \( e_3 \) are co-cyclic because \( V_0'V_2''V_4''V_4' \) form an isosceles quadrilateral. Therefore \( \beta_3 \) is either 0 or \( \pi \). A closer inspection based on orientation (or simply applying the formula below) tells us that \( \beta_3 = 0 \). Similarly, \( \beta_6 = 0 \). Next, we show that

\[
(3.6) \quad \lim_{\epsilon_1, \epsilon_2 \to 0} \beta_1 = \pi = \lim_{\epsilon_1, \epsilon_2 \to 0} \beta_4, \quad \lim_{\epsilon_1, \epsilon_2 \to 0} \beta_2 = \frac{2\pi}{n} = \lim_{\epsilon_1, \epsilon_2 \to 0} \beta_5.
\]

These limits are not hard to see geometrically based on Figure 9 but to get the finer monotonicity result we resort to algebra and use the formula

\[
\cos(\beta_i) = \frac{\langle A, C \rangle \langle B, D \rangle - \langle A, B \rangle \langle C, D \rangle - \langle B, C \rangle \langle D, A \rangle}{\|A\|\|B\|\|C\|\|D\|},
\]

where \( A = V_0' - V_{i+1}', B = V_{i-1}' - V_0', C = V_i' - V_{i-1}', D = V_{i+1}' - V_i' \). (Here the + and − are modulo 6 addition and subtraction operated on the indices \( 1, \ldots, 6 \).) By computation, we get \( \cos(\beta_1) = \cos(\beta_4) = -\frac{1-\epsilon_1 c - \epsilon_1 c^2 + \epsilon_2 c + \epsilon_2 c^2}{\sqrt{1-2\epsilon_1 c + \epsilon_1 c^2 + 1-2\epsilon_2 c + \epsilon_2 c^2}} \), \( \cos(\beta_2) = \frac{c-2\epsilon_1 c^2 + \epsilon_2 c^2}{1-2\epsilon_1 c + \epsilon_1 c^2} \), \( \cos(\beta_5) = \frac{c-2\epsilon_2 c^2 + \epsilon_2 c^2}{1-2\epsilon_2 c + \epsilon_2 c^2} \). The limits \( (3.6) \) then follow immediately and we have proved that \( \lim_{\epsilon_1, \epsilon_2 \to 0} W_{0,0}(\epsilon_1, \epsilon_2) = 4\pi/n \). To see that the convergence is
Figure 9. Computation of $W_{0,0}$ based on $V'_0 = [1,0]$, $V'_1 = [c,s]$, $V'_2 = \epsilon_2[1,0]$, $V'_3 = \epsilon_2[c,-s]$, $V'_4 = [c,-s]$, $V'_5 = \epsilon_1[1,0]$, $V'_6 = \epsilon_1[c,s]$. Among the six circumcircles, two pairs coincide due to the isosceles quadrilaterals $V'_0V'_2V'_3V'_4$ and $V'_0V'_5V'_6V'_1$, thus only 4 circles are seen; they are also divided into two groups of three (or two rather), displayed in solid and dashed line styles, which are oriented differently when viewed from the outside of the plane. This is caused by the ‘folding’ of the neighborhood of $V'_0$, and is also why $W_{0,0}$ does not vanish despite all the vertices are coplanar. In this figure, $(c,s) = (\cos(2\pi/5), \sin(2\pi/5))$, i.e. $n = 5$, $(\epsilon_1, \epsilon_2) = (.3, 1)$ and it does not come from an $\epsilon$ via (3.5). Steps $2^\circ$ and $3^\circ$ of the proof also do not rely explicitly on (3.5).

In the genus $g = 0$ case, as long as the connectivity is generic in some sense (see [6, Proposition 9]), the simplicial sphere is inscribable in a sphere and hence has a minimum $W_{Bobenko}$-energy $4\pi$ (0 in Bobenko’s definition of $W = our definition of $W_{Bobenko} - 4\pi(1 - g)$). This is of course the same energy level as what one expects from the smooth setting.

Proposition 3.9 shows that in the genus 1 case, the infimum is never what one expects from the smooth setting, even with the perfectly regular connectivity and
regardless of the grid size. Moreover, Proposition 3.9 and computational experiments suggest the following.

**Conjecture 3.10.** For any $m, n \geq 3$, $W_{\text{Bobenko}}(T_{m,n}) = 4\pi$. Moreover, the infimum is realized by degenerate spheres $(T_{m,n,\varepsilon}, \varepsilon \to 0)$ and not by any embedded genus 1 PL surface.

This conjecture suggests that a genus 1 $W_{\text{Bobenko}}$-minimizer does not exist, again in contrast to the smooth setting; cf. [66]. It also seems possible to formulate and prove a generalization of Proposition 3.9 to genus $g > 1$, as suggested by the computational result in Figure 10.

![Figure 10](image)

**Figure 10.** A $W_{\text{Bobenko}}$-minimization process rounds the 2- and 3-hole tori in (a) and (c) and closes up the holes, resulting in the empirical minimizers in (b) & (d), both with $W_{\text{Bobenko}} \approx 4\pi$.

Back to the genus 1 setting, the proof of Proposition 3.9 suggests the following generalization. Consider the family of planar tori, denoted by $T_{m,n,r}$, with the same connectivity as before and vertices

$$V_{i,j} = r_i \left[ \cos \left( \frac{2\pi j}{n} \right), \sin \left( \frac{2\pi j}{n} \right), 0 \right], \quad i = 0, \ldots, m-1, \quad j = 0, \ldots, n-1,$$

$$r = (r_0, r_1, \ldots, r_{m-1}), \quad r_0 > r_1 > \cdots > r_{m-1} > 0.$$

We have the following corollary of the proof.

**Corollary 3.11.** For any grid size $m, n \geq 3$, and fixed $r_0 > \cdots > r_{m-2} > 0$,

$$\lim_{r_{m-1} \to 0} W_{\text{Bobenko}}(T_{m,n,r}) = 4\pi$$

and the convergence is monotone.

**Proof.** All vertices in the intermediate layers ($0 < i < m - 1$) have zero energy. By the calculation in the previous proof, every vertex in the outermost ($i = 0$) layer has the same energy $2 \cos^{-1} \left( \frac{c-2\epsilon+c^2}{1-2c+c^2} \right)$, $c = r_{m-1}/r_0$. The fact that this energy is independent of $r_1$, together with the Möbius invariance of each $W_{i,j}$, actually imply that every vertex in the innermost layer ($i = m - 1$) also has the same energy $2 \cos^{-1} \left( \frac{c-2\epsilon+c^2}{1-2c+c^2} \right)$ (independent of $r_{m-2}$); this can be seen by turning $T_{m,n,r}$ inside out based on inverting about the sphere with radius $\sqrt{r_0r_{m-1}}$ centered at the origin.

Therefore $W_{\text{Bobenko}}(T_{m,n,r}) = \frac{1}{2} \cdot 2n \cdot 2 \cos^{-1} \left( \frac{c-2\epsilon+c^2}{1-2c+c^2} \right)$ and the result follows by taking $\epsilon \to 0$. □
3.2.2. $W_{\text{Centroid}}$ and $W_{\text{EffAreaCur}}$. For any mesh $T_{m,n,r}$ with $r_0$ normalized to 1, $r_0 = 1 > r_1 > \cdots > r_{m-1} > 0$, we have the following by direct calculation:

\[
\begin{align*}
    a_{\text{centroid}}(0, 0) &= \frac{4 - r_1 - r_1^2 - r_{m-1} - r_{m-1}^2}{6}, \\
    a_{\text{centroid}}(m - 1, 0) &= \frac{1 + r_{m-1} - 4r_{m-1}^2 + r_{m-2}^2 + r_{m-1}r_{m-2}}{6}, \\
\end{align*}
\]

\[
\begin{align*}
    \nabla_{(0,0)} A &= \begin{bmatrix} 2s, & 0, & 0 \end{bmatrix}, \\
    \nabla_{(m-1,0)} A &= \begin{bmatrix} -2r_{m-1}s, & 0, & 0 \end{bmatrix}, \\
    \nabla_{(0,0)} V &= \begin{bmatrix} 0, & 0, & r_1 - r_{m-1} + r_1^2 - r_{m-1}^2 \end{bmatrix}, \\
    \nabla_{(m-1,0)} V &= \begin{bmatrix} 0, & 0, & (r_{m-2} - 1)(r_{m-1} + r_{m-2} + 1) \end{bmatrix},
\end{align*}
\]

where $s = \sin(2\pi/n)$. Since the mesh is flat, it does not have an ‘inside’ or ‘outside’, we arbitrarily choose a consistent orientation in order to determine the direction of $\nabla_v V$. (Recall Remark 2.2.)

Next, we show the following.

**Proposition 3.12.** Let the grid sizes $m, n \geq 3$ be fixed.

(i) For any fixed $r_0 > 0$,

\[
\lim_{r_1 \to r_0} W_{\text{Centroid}}(T_{m,n,r}) = \frac{3}{2} \pi \sin \left( \frac{2\pi}{n} \right).
\]

Hence $W_{\text{Centroid}}(T_{m,n}) \leq 3\pi < 2\pi^2$ regardless of $m, n$.

(ii) For any fixed $r_0 > r_2 > \cdots > r_{m-2} > 0$,

\[
\lim_{r_{m-1} \to r_0} W_{\text{EffAreaCur}}(T_{m,n,r}) = 3n \sin \left( \frac{2\pi}{n} \right).
\]

Hence $W_{\text{EffAreaCur}}(T_{m,n}) \leq 6\pi < 2\pi^2$ regardless of $m, n$.

**Proof.** All vertices in the intermediate layers ($0 < i < m - 1$) have zero energy. By symmetry, every vertex in the outermost ($i = 0$) and innermost ($i = m - 1$) layer has the same energy. So $W_{\text{Centroid}}(T_{m,n,r}) = n(W_{0,0} + W_{m-1,0})$. By scale invariance of $W_{\text{Centroid}}$, we can assume $r_0 = 1$. By (3.7),

\[
\lim_{r_1, r_{m-1} \to 0} W_{0,0} = \lim_{r_1, r_{m-1} \to 0} \left[ \frac{\| \nabla_{(0,0)} A \|}{2a_{\text{centroid}}(0, 0)} \right]^2 a_{\text{centroid}}(0, 0) = \frac{3}{2} \pi \sin(2\pi/n),
\]

\[
\lim_{r_1, r_{m-1} \to 0} W_{m-1,0} = \lim_{r_1, r_{m-1} \to 0} \left[ \frac{\| \nabla_{(m-1,0)} A \|}{2a_{\text{centroid}}(m-1, 0)} \right]^2 a_{\text{centroid}}(m-1, 0) = 0.
\]

And the proof of (i) is completed.

Similarly, $W_{\text{EffAreaCur}}(T_{m,n,r}) = n(W_{0,0} + W_{m-1,0})$. Assume $r_0 = 1$. By (3.7),

\[
\lim_{r_{m-1} \to 0} W_{0,0} = \lim_{r_{m-1} \to 0} \left[ \frac{\| \nabla_{(0,0)} A \|}{2\| \nabla_{(0,0)} V \|} \right]^2 \| \nabla_{(0,0)} V \| = 3 \pi \sin(2\pi/n),
\]

\[
\lim_{r_{m-1} \to 0} W_{m-1,0} = \lim_{r_{m-1} \to 0} \left[ \frac{\| \nabla_{(m-1,0)} A \|}{2\| \nabla_{(m-1,0)} V \|} \right]^2 \| \nabla_{(m-1,0)} V \| = 0.
\]

And the proof is completed. □
In contrast to Conjecture 3.10 pertaining to $W_{\text{Bobenko}}$, we observe from computation and preliminary calculations that Proposition 3.12 is not sharp in the sense that the infimum values $W_{\text{Centroid}}(T_{m,n})$ and $W_{\text{EffAreaCur}}(T_{m,n})$ are strictly smaller than $3/2n \sin(2\pi/n)$ and $3n \sin(2\pi/n)$, respectively. It is observed that minimizers of both energies are some subtly ‘folded up’ planar meshes.

Although we do not pursue it here, a similar negative result holds for $W_{\text{Voronoi}}$. The situation for $W_{\text{NormalCur}}$, however, is less well-understood. Note that the outer- and inner-most vertices of $T_{m,n,r}$ satisfy

\[
\nabla_v V \bot \nabla_v A, \quad \nabla_v A \neq 0;
\]

recall (3.7). This means the corresponding local areas in $W_{\text{NormalCur}}$ vanish and $W_{\text{NormalCur}} = +\infty$. Therefore the planar meshes we considered cannot make $W_{\text{NormalCur}}$ small. This is, by design, a key difference between $W_{\text{EffAreaCur}}$ and $W_{\text{NormalCur}}$. However, this does not mean $W_{\text{NormalCur}}$ is rid of the type of negative results in Propositions 3.9 and 3.12. We observe from computations that minimizers of $W_{\text{NormalCur}}$ exhibit a similar behavior as those of $W_{\text{EffAreaCur}}$: they are highly ‘folded up’, but non-planar, meshes, with energies less than $2\pi^2$; see Figure 11(a)-(b).

![Figure 11](image)

(a) $W_{\text{NormalCur}} = 37.73$
(b) $W_{\text{NormalCur}} = 14.85$
(c) $W_{\text{NormalCur}} = 45.46$
(d) $W_{\text{NormalCur}} = 12.95$

\[\approx 2\pi^2\]
\[\approx 4\pi\]

3.2.3. Genus 0 case. In the genus 0 case, computational experiments suggest that $W_{\text{Centroid}}$, $W_{\text{Voronoi}}$ and $W_{\text{EffAreaCur}}$ fail in a way similar to the genus 1 case; we believe that a genus 0 version of Proposition 3.12 can be established.
However, the situation for $W_{\text{Bobenko}}$ and $W_{\text{NormalCur}}$ are different. Recall that the Willmore energy of any closed surface should always be greater than or equal to $4\pi$, with equality attained only by the round sphere [73]. $W_{\text{Bobenko}}$ is designed to satisfy this ‘ground-state’ property: $W_{\text{Bobenko}} \geq 4\pi$, with equality holds when and only when the PL surface is a convex polyhedron inscribed in a sphere [6, Theorem 5]. So there cannot be a genus 0 version of the negative result in Proposition 3.9. An empirical finding is that $W_{\text{NormalCur}}$ actually works well for the genus 0 Willmore problem. See Figure 11(c)-(d), which shows one of the many trials of the experiment with a randomized initial mesh; all trials of $W_{\text{NormalCur}}$-minimization applied to a genus 0 mesh result in a near round sphere with $W_{\text{NormalCur}}$ slightly larger than $4\pi$.

4. A regularized PL method

All failures we observed have one thing in common: triangles with bad aspect ratios develop. This is of course a familiar issue in FEM and mesh generation, but in the moving surface context here the issue seems different and understudied. In previous work by Hsu-Kusner-Sullivan [40], procedures such as ‘vertex averaging’, ‘edge notching’ and ‘equiangulation’, implemented in the Surface Evolver [10,11], are used to fix up triangles with bad aspect ratios along the way of the optimization process. While such ‘mesh smoothing’ procedures are probably well studied in the mesh generation community, when applied to the geometric variational problems here the approach seems difficult to analyze mathematically, because such an approach dynamically alters the connectivity of the mesh in the course of the optimization process. (The approach developed in this section does not require such a dynamic change in connectivity.)

Our numerical methods, based on either PL or SS, are parametric in nature. Yet, the variational problems are geometric and their solutions are independent of parametrization. This creates another problem instead of solving the existing one! Our observation is that it is possible to design a method in such a way that the two problems cancel each other.

Since the solution is independent of parametrization, we may request the parametrization to be (approximately) conformal, i.e. angle preserving, with respect to an appropriate conformal structure in the reference manifold $\Sigma$. If such a property can be built into a $W$-minimizing process, the method would only search over PL surfaces without long skinny triangles.

The approach adopted here is inspired by the classical uniformization theorem and results in harmonic maps [27,29], as well as the relatively recent developments in computational conformal geometry, see [17,24,35,38,53] and the references therein.

In a finite-dimensional approximation of the solution surface, we cannot expect to have exactly conformal parametrization. To fix the idea, let us first consider the problem of finding a conformally parameterized Willmore surface in the smooth setting.

4.1. Ideas in the smooth setting. We use the theory of harmonic maps. From now on we assume $\Sigma$ is a Riemann surface. For any map $x \in W^{1,2}(\Sigma, \mathbb{R}^3)$, we can

---

8Not to be confused with the conformality in finite-element methods.
define its Dirichlet energy by
\begin{equation}
\mathcal{E}(\mathbf{x}) = \frac{1}{2} \int_M \|d\mathbf{x}\|^2 d\Sigma,
\end{equation}
where \(\|d\mathbf{x}\|\) is the Hilbert-Schmidt norm first defined based on some Riemannian metric on \(\Sigma\) and the standard Riemannian metric in \(\mathbb{R}^3\), followed by the observation that \(\mathcal{E}\) is invariant under any conformal change of Riemannian metric in the domain; see [29] Pg. 126. This invariance is specific to \(\dim \Sigma = 2\), and it means that the Dirichlet energy depends only on the conformal structure of \(\Sigma\). Next, we have the following well-known result.

**Theorem 4.1** [29]. For any \(C^1\) immersion \(\mathbf{x} : \Sigma \to \mathbb{R}^n\), \(A(\mathbf{x}) \leq \mathcal{E}(\mathbf{x})\). Equality holds when and only when \(\mathbf{x}\) is conformal.

It is already evident from the theorem that the difference \(\mathcal{E}(\mathbf{x}) - A(\mathbf{x})\) measures the deviation of \(\mathbf{x}\) being conformal. In fact, the difference can be explicitly expressed by a conformal energy expressing the deviation of \(\mathbf{x}\) from satisfying the Cauchy-Riemann equations.

**Proposition 4.2.** Let \(\Sigma\) be a Riemann surface and \(S = \{\mathbf{x} \in \text{Imm}_{C^1}(\Sigma, \mathbb{R}^3) \cap W^{2,2}(\Sigma, \mathbb{R}^3) : A(\mathbf{x}) = A_0\}\) for a fixed \(A_0 > 0\). Assume that:

(i) a minimizer \(\mathbf{x}_\lambda \in \arg\min_{\mathbf{x} \in S} W(\mathbf{x}) + \lambda \mathcal{E}(\mathbf{x})\) exists for all small enough \(\lambda\),

(ii) \(\lim_{\lambda \downarrow 0} \mathbf{x}_\lambda =: \mathbf{x}_*\) exists. (Here convergence is in the topology of \(C^1(\Sigma, \mathbb{R}^3) \cap W^{2,2}(\Sigma, \mathbb{R}^3)\).)

Then

(I) \(\mathbf{x}_* : \Sigma \to \mathbb{R}^3\) is a Willmore minimizer, i.e. it is a solution of the Willmore problem \(\min_{\mathbf{x} \in S} W(\mathbf{x})\). Moreover, any other Willmore minimizer \(\mathbf{x}_{**}\) satisfies \(A_0 \leq \mathcal{E}(\mathbf{x}_*) \leq \mathcal{E}(\mathbf{x}_{**})\).

(II) If the conformal structure on \(\Sigma\) is such that there exists a conformal parametrization of the surface \(\mathbf{x}_*(\Sigma)\), then \(\mathbf{x}_*\), and in fact \(\mathbf{x}_\lambda\) for any \(\lambda > 0\), is a conformal parametrization of a Willmore minimizer, i.e. \(W(\mathbf{x}_\lambda) = W(\mathbf{x}_*)\) and \(A_0 = \mathcal{E}(\mathbf{x}_\lambda) = \mathcal{E}(\mathbf{x}_*)\) for any \(\lambda > 0\).

**Proof.** We first argue that \(\mathbf{x}_*\) must be a Willmore minimizer. If not, there exists an \(\mathbf{x}' \in S\) such that \(W(\mathbf{x}') < W(\mathbf{x}_*)\). But then for small enough \(\lambda > 0\), we must have \(W(\mathbf{x}') + \lambda \mathcal{E}(\mathbf{x}') < W(\mathbf{x}_*) + \lambda \mathcal{E}(\mathbf{x}_*)\). (If \(\mathcal{E}(\mathbf{x}') \leq \mathcal{E}(\mathbf{x}_*)\), this holds for any \(\lambda > 0\), otherwise choose \(\lambda < (\mathcal{E}(\mathbf{x}') - \mathcal{E}(\mathbf{x}_*)/(W(\mathbf{x}_*) - W(\mathbf{x}'))\).) By (ii) and the continuity of \(W : S \to \mathbb{R}\) and \(\mathcal{E} : S \to \mathbb{R}\), we have
\[W(\mathbf{x}') + \lambda \mathcal{E}(\mathbf{x}') < W(\mathbf{x}_\lambda) + \lambda \mathcal{E}(\mathbf{x}_\lambda), \quad \forall \text{ small enough } \lambda > 0.\]
This contradicts assumption (i), and we have proved the first claim in (I). We now argue the second claim again by contradiction. Let \(\mathbf{x}_{**}\) be any other Willmore minimizer, so \(W(\mathbf{x}_*) = W(\mathbf{x}_{**})\). Assume the contrary that \(\mathcal{E}(\mathbf{x}_*) > \mathcal{E}(\mathbf{x}_{**})\). Then
\[W(\mathbf{x}_*) + \lambda \mathcal{E}(\mathbf{x}_*) > W(\mathbf{x}_{**}) + \lambda \mathcal{E}(\mathbf{x}_{**})\]
for any \(\lambda\). Again by (ii) and the continuity of \(W\) and \(\mathcal{E}\),
\[W(\mathbf{x}_\lambda) + \lambda \mathcal{E}(\mathbf{x}_\lambda) > W(\mathbf{x}_{**}) + \lambda \mathcal{E}(\mathbf{x}_{**}), \quad \forall \text{ small enough } \lambda > 0.\]
This contradicts (i), so \(\mathcal{E}(\mathbf{x}_*) \leq \mathcal{E}(\mathbf{x}_{**})\). By Theorem 4.1 we also have \(A_0 \leq \mathcal{E}(\mathbf{x}_*) \leq \mathcal{E}(\mathbf{x}_{**})\).

By Theorem 4.1 and the assumption of the conformal structure on \(\Sigma\), the minimum values of \(W\) and \(\mathcal{E}\) over \(S\), namely \(W(\mathbf{x}_*)\) and \(A_0\), are attained simultaneously.
by some \( x \in S \). But any \( x_\lambda \), being a minimizer of \( W(\cdot) + \lambda A(\cdot) \), must also be such a joint minimizer. Then, by continuity, \( x_* \) is such a joint minimizer as well. \( \square \)

Remark 4.3. In the most ubiquitous spherical topology, there is only one conformal structure up to conformal equivalence. Together with the genus 0 case of the uniformization theorem, the assumption in (II) above is satisfied automatically when \( \Sigma \) is the Riemann sphere. In this case, solving the \( \mathcal{E} \)-penalized Willmore problem with any penalization parameter \( \lambda \) – not necessarily small – would deliver a conformally parameterized Willmore surface. In the higher genus cases, there is a 1 (when \( g = 1 \)) and \( 3g - 3 \) (when \( g \geq 2 \)) complex-dimensional space of conformal structures [42], and typically we do not know a priori the ideal conformal structure for the Willmore surface. Without the correct conformal structure, a conformally parametrized solution is impossible, but Proposition 4.2 says (in (I)) that the solution of the \( \mathcal{E} \)-penalized Willmore problem with a small penalization parameter \( \lambda \) would still deliver a Willmore surface parameterized as conformally as possible in the non-ideal conformal structure.

We shall use this result to guide us to develop an algorithm for fixing the naive PL method. Before we proceed, we present a version of Theorem 4.1 for PL surfaces.

4.2. A PL version of Theorem 4.1. Theorem 4.5 below is well-known to experts in computational conformal geometry; see, for example, [17,35] and references therein. We provide a proof for it not only for the sake of self-containedness but also to help us to understand its subtle relationship with the smooth counterpart Theorem 4.1.

Definition 4.4. Let \( K \) be a simplicial surface.

1. An angle assignment on \( K \) is an assignment of angles \( (\theta_1^\tau, \theta_2^\tau, \theta_3^\tau) \) to every triangle \( \tau \) in \( K \).
2. A PL immersion \( x: |K| \to \mathbb{R}^n \) is called angle preserving if every triangle \( \tau \) in \( K \) is similar to \( x(\tau) \in \mathbb{R}^n \), i.e. they share the same angles.

The notion of angle assignment in (1) above actually defines a conformal (or complex analytic) atlas in the traditional sense via a construction by L. Bers. The construction essentially exploits the fact that the map

\[ z \mapsto z^a \]

maps the interior – but not the boundary – of a cone in \( \mathbb{C} \) with apex angle \( \theta \) biholomorphically to the interior of a cone with apex angle \( a\theta \) (assume \( 0 < a < 2\pi/\theta \)). For details, see [24] Section 2 or [5] Lecture 2. In these references, the conformal structure is supposed to be induced by a PL embedding \( F: |K| \to \mathbb{R}^n \). However, it is obvious that the conformal atlas defined there depends only on the angle assignment induced by the embedding. We note that, however, many different angle assignments may give rise to equivalent conformal structures on \( K \). For example, in the genus 0 case, any two conformal structures are conformally equivalent.

Once such a conformal structure is defined, and noting also that piecewise linear maps have square integrable derivatives, the Dirichlet energy \( \mathcal{E}(x) \) of a PL immersion \( x \) is well-defined according to (4.1). Of course, the area \( A(x) \) is also well-defined in the standard sense. As such, the first half of Theorem 4.1, which requires only \( x \) to be in \( W^{1,2} \), holds verbatim, i.e. \( \mathcal{E}(x) \leq A(x) \). However, the second half of
Theorem 4.1 is less obvious, as the standard definition of conformality for smooth maps does not seem to apply directly to PL maps. Fortunately, if one simply defines conformality for PL immersions as in Definition 4.4, we get the following result.

**Theorem 4.5** (PL version of Theorem 4.1). Let $K$ be a simplicial surface with an angle assignment. For any PL immersion $x : |K| \to \mathbb{R}^n$, $A(x) \leq E(x)$. Equality holds when and only when $x$ is angle preserving.

**Proof.** Bers’ conformal structure and the fact that $E$ depends solely on the conformal structure guarantee that in the interior of every triangle $\tau$ of $K$, the Dirichlet energy (4.1) can be computed as

$$(4.2)\quad E(x) = \frac{1}{2} \sum_{\tau \in K} \int_{\tau} \|dx\|^2 d\text{Area}_\tau,$$

where “$d\text{Area}_\tau$” is the area element in the coordinates of $\tau$ after we identify $\tau$ with a triangle on the plane with the angles assigned to $\tau$. For details, consult [24, Section 2]. For any $\tau$, $x|_\tau$ is a linear map and we may represent it by a rank 2 matrix $A_\tau \in \mathbb{R}^{n \times 2}$, so

$$(4.3)\quad E(x) = \frac{1}{2} \sum_{\tau} \text{trace}(A^T_\tau A_\tau) \text{Area}(\tau) = \sum_{\tau} \frac{1}{2} \text{trace}(A^T_\tau A_\tau) \text{Area}(x(\tau)).$$

Since

$$\frac{1}{2} \text{trace}(A^T_\tau A_\tau) = \frac{1}{2} (\lambda^1_\tau + \lambda^2_\tau) \geq 1,$$

where $\lambda^1_\tau$ and $\lambda^2_\tau$, both positive, are the eigenvalues of $A^T_\tau A_\tau$, $E(x) \geq A(x)$.

Moreover, equality holds above if and only if $\lambda^1_\tau = \lambda^2_\tau$ for all triangles $\tau \in K$, which is exactly the case when $x$ is angle preserving. \[ Q.E.D. \]

In the PL setting above, the Dirichlet energy can also be expressed by the cotangent formula

$$(4.4)\quad E(x) = \frac{1}{4} \sum_{\tau} \sum_{i=1,2,3} \cot(\theta_i) \text{length}(x(e_i))^2$$

$$= \frac{1}{4} \sum_{e \in \text{Edge}} (\cot(\alpha_e) + \cot(\beta_e)) \text{length}(x(e))^2,$$

where $e_i$ is the edge of $\tau$ opposite the angle $\theta_i$, and $\alpha_e$ and $\beta_e$ are the angles opposite the edge $e$ in the two incident triangles. This follows immediately from a simple formula: if $f : \tau \to \mathbb{R}$ is a linear function on a triangle $\tau \subset \mathbb{R}^2$, then

$$\int_\tau \left| \frac{\partial f}{\partial x} \right|^2 + \left| \frac{\partial f}{\partial y} \right|^2 \, dx \, dy$$

$$= \frac{1}{2} \left\{ \cot(\theta_1) |f(v_2) - f(v_3)|^2 + \cot(\theta_2) |f(v_3) - f(v_1)|^2 + \cot(\theta_3) |f(v_1) - f(v_2)|^2 \right\},$$

where $v_1, v_2, v_3 \in \mathbb{R}^2$ are the vertices of $\tau$ and $\theta_i$ is the angle at vertex $v_i$.

**Remark 4.6.** A subtle point here is that if we simply define the Dirichlet energy of a PL immersion by either (4.3) or (4.4), then both the formulation and the proof of Theorem 4.5 completely bypass Bers’ construction. We may then view the ‘Bers-free version’ of Theorem 4.5 as a discrete analog of Theorem 4.1. This would be in line with the spirit of ‘discrete differential geometry’ [S]. We find it instructive to
know that Theorem 4.5 and Theorem 4.1 are not only analogous, but that one may view (at least the first part of) the former theorem as a special case of the latter classical result.

Finally, we note that when the angles \((\theta_1^\tau, \theta_2^\tau, \theta_3^\tau)\) assigned to \(K\) are exactly those of the triangles \(x(\tau)\) in \(\mathbb{R}^n\), then it is a tautology to say that \(x\) is angle preserving. In this case, \(A(x)\) is given by the cotangent formula (4.4). This, in turn, can be used to give an alternative proof for the cotangent formula of \(\nabla_v A\) in (2.4), explaining also why the two seemingly unrelated cotangent formulas look so alike.

4.3. A penalized PL method. Inspired by Proposition 4.2, we propose a numerical solution to the Willmore, Canham or Helfrich problem based on solving:

\[
\begin{align*}
\min_V W_{PL}(V) + \lambda \mathcal{E}(V), \\
\text{s.t. } A(V) = A_0 \text{ and relevant constraints on } V(V) \text{ and } M(V).
\end{align*}
\]

Here \(W_{PL}\) represents any of the PL Willmore energy discussed earlier, and \(\mathcal{E}\) is the Dirichlet energy of the piecewise linear immersion (determined by \(V\)) of the domain simplicial complex \(K\) (determined by the face list \(F\)) endowed with a suitable angle assignment.

Proposition 4.2 suggests that, while it is not necessary, it would be best if we use a conformal structure for which a conformal parametrization is possible, as it would free us from the need of making the penalization \(\lambda\) arbitrarily small, thus giving us a more efficient numerical method. In the genus 0 case, there is only one conformal structure up to conformal equivalence. So we may simply assume every triangle in \(K\) to be equilateral. Under this angle assignment, the Dirichlet energy is, by (4.4),

\[
\mathcal{E} = \frac{1}{2} \cot(60^\circ) \sum_{e \in \text{Edge}} \text{length}(e)^2 = \frac{1}{2\sqrt{3}} \sum_{e \in \text{Edge}} \text{length}(e)^2.
\]

As mentioned, we can still use (4.6) in the higher genus cases, as long as we use a small enough \(\lambda\); see the genus 1 results in the next section.

4.4. Computational results. Figure 12 shows how the \(\mathcal{E}\)-penalized method (4.5) performs on the genus 0 Canham problem with reduced volume/isoperimetric ratio \(v_0 := \frac{3\sqrt{6}}{4\pi} \left(\frac{A_0}{4\pi}\right)^{-\frac{3}{2}}\) in three different intervals known to give the shapes of a stomatocyte, discocyte (red-blood cell) and prolate. The results are consistent with those reported in the biophysics literature, and those produced by SS methods. See Section 5 for a more thorough comparison of different methods applied to these problems.

In each case, we found that the numerical solution changes little with the choice of \(\lambda\), which is consistent to the conclusion of Proposition 4.2. Also from the fact that \(\mathcal{E}\) is pretty close to the area \(A\), we conclude from Theorem 4.5 that the PL embeddings are close to being angle preserving.

In the higher genus case not all conformal structures are conformally equivalent. In the genus 1 case, the uniformization theorem implies that our solution surface can be mapped conformally to a flat torus. However, there is a one-complex dimensional family of non-equivalent conformal structures on a flat torus.

Here, we use \(W_{\text{Eff Area Cur}}\) to solve the genus 1 Willmore problem. And we choose the domain Riemann surface to be a regularly triangulated parallelogram with sides
1 and $\omega$ on the complex plane. We denote by $\mathcal{E}_\omega$ – easily computable according to (4.4) – the corresponding Dirichlet energy. In this case we know that the solution is the Clifford torus, with Willmore energy $2\pi^2$, and the ideal conformal structure is that of a square, i.e. $\omega = e^{i\pi/2}$. We compare the result with what we get by using the $60^\circ$ conformal structure, i.e. $\omega = e^{i\pi/3}$. See the results in Figure 13(b) & (c). We see that by using the ideal conformal structure the PL Willmore energy is closer to the expected $2\pi^2 \approx 19.73$, and $\mathcal{E} - A$ is closer to zero. (With the incorrect conformal structure, we cannot expect to have a conformal parametrization, and hence we cannot expect $\mathcal{E} - A \to 0$ as the grid sizes grow.) Also with the incorrect conformal structure the axis-symmetry is broken, the surface looks like a Möbius transformation of the surface of revolution Clifford torus (a Dupin cyclide), but with an indentation, indicated by the arrow in Figure 13(c). The indentation suggests that the method is not capturing the shape accurately. However, by reducing $\lambda$ from 1/2 to 1/10, the indentation disappears and the resulted surface becomes a more accurate approximation to a Clifford torus. This is completely in line with what Proposition 4.2 says.

In Figure 13(d), we show a numerical solution of the genus 1 Canham problem with reduced volume $v_0 = 0.8$. It is believed that the solution is a unique Möbius transformation of the surface of revolution Clifford torus; see [76]. And we also know that the ideal conformal structure to use is the square one. The solution is accurate even with a relatively big penalization factor. Again, it is in agreement with Proposition 4.2.
5. FURTHER COMPUTATIONAL RESULTS AND SYMMETRY PRESERVATION

A group of biophysicists obtained a plethora of computational results for the Canham and Helfrich problems for low genus $g$ and various constraint parameters $(v_0, m_0)$ \[41, 43, 51, 56, 57, 74\]. Many interesting observations are made about the uniqueness, non-uniqueness, symmetry and phase transition properties for these problems. So far few of these observations have been justified mathematically. In fact, the mere existence of a solution of the Canham problem, i.e. existence of a Willmore minimizer with a prescribed isoperimetric ratio and genus, is an unsolved problem in geometric analysis when the genus is larger than 0 \[46\]; for the genus 0 case, see \[64\]. The existence of solution of the Helfrich problem has not been addressed so far. Moreover, in the biophysics literature, the numerical behavior of the optimization method is never addressed; the only information we have is that Brakke’s Surface Evolver was used, as in the experiments done in \[40\].

In this final technical section, we give a few comparisons of the different implementations of the PL and SS methods presented earlier. We must begin with a confession:

The algorithms developed in Sections 2 and 4 only address how to discretize the variational problem into a standard finite-dimensional constrained optimization
problem of the form
\[ \min_{x \in \mathbb{R}^N} f(x) \quad \text{s.t.} \quad g_i(x) = 0, \]
and the theory in Section 3 sheds some light on what it means to the variational problem if we manage to solve the optimization problem. Needless to say, solving the latter problem itself — a high-dimensional, nonlinear, nonconvex, constrained optimization problem — is a major challenge and there are many algorithms and solvers available. In our experiments, we use the following three solvers: (i) fmincon in the Matlab optimization toolbox, (ii) SNOPT [34] and (iii) GRANSO [18].

Given the complexity of these solvers, there are countless issues to explore and compare in conjunction with our problems. We shall focus mainly on the approximation and symmetry properties of the optimization problems arising from the PL and SS methods, and will briefly touch upon the existence issue of these optimization problems. With regrets, we will not address the important question of the efficiency of solving these problems by various optimization algorithms/solvers.

5.1. Comparison I: PL vs SS. The original expectation is that a PL method is less accurate than the higher order SS counterparts, only that we found in Sections 3.2 and 4 that the non-conforming nature of PL methods require us to introduce regularization. We are finally in a position to compare the accuracy of our PL and SS methods.

It is known from Schygulla [64] that \( W \)-minimizer of genus 0 with any prescribed isoperimetric ratio \( v_0 \in (0, 1) \) exists, and that the minimum Willmore energy is strictly less than \( 8\pi \). In Figure 12 we solve for these minimizers using a regularized PL method for \( v_0 = 0.5, 0.62 \) and 0.85. The \( v_0 = 0.5 \) case is arguably the most difficult one as an ‘invagination’ develops in the vesicle; in this case the PL Willmore energy computed is bigger than \( 8\pi \), see the caption of Figure 12(b). We now solve the same three problems but using our Loop SS method. Without an explicit representation of the solution, we cannot directly compare the accuracy of the PL and SS methods. However, we see from our computation that in each case, the (true) Willmore energy of the SS approximation is lower than the (PL) Willmore energy of the PL approximation. In the \( v_0 = 0.5 \) case, the Willmore energy of the SS approximation is less than \( 8\pi \), as it should according to Schygulla’s result. These observations are consistent with the higher accuracy order of SS compared to PL approximations.

5.2. Comparison II: Symmetry preservation vs symmetry breaking. The results in Figures 12 and 14 are based on the fmincon and SNOPT solvers. With our third solver, GRANSO, and the same octahedral initial mesh in Figure 12(a), we got totally different, non-global local minimizers which inherit the octahedral symmetry from the initial mesh; see Figure 15.

Apparently, the underlying constrained optimization algorithm employed in the GRANSO solver has a symmetry preserving property. To the best of the authors’ knowledge, this kind of symmetry preservation and breaking properties is not well-addressed in the optimization literature. Here, we observe a fundamental difference among the three solvers: GRANSO is capable of preserving symmetry in the sense below\(^9\) while fmincon and SNOPT are capable of breaking symmetry.

\(^9\)The authors of GRANSO designed the solver with the intention that it can handle non-smooth problems, but without the intention for symmetry preservation. Our original use of it is due to
Figure 14. Genus 0 $v_0$-constrained $W$-minimizers computed based on our Loop SS method. (a) two initial meshes: a 3-times subdivided octahedron and its flattened counterpart. Both initial meshes, and both the SNOPT and fmincon optimization solvers give the stomatocyte in (b) and the prolate in (d). For $v_0 = 0.62$, the initial mesh with full octahedral symmetry happens to give the non-global local minimizer on the top of (c) (a prolate), whereas the flattened initial mesh (with $D_4$ symmetry, in Schoenflies notation) gives the typical bi-concave red-blood cell shape (an oblate) at the bottom. In (b)–(d), the black lines depict the control mesh of the SS surface, the red surface is the SS surface itself.

Symmetry preservation of gradient and Newton descent. To define ‘symmetry preservation’ precisely, we begin with the following fact about gradient flow which is well-known to geometers: Let $M$ be a Riemannian manifold, and $G$ be a group of isometries acting on $M$. If we have a smooth $G$-invariant functional $F : M \to \mathbb{R}$, i.e. $F(g \cdot x) = F(x)$ for all $x \in M$ and $g \in G$, then the gradient flow map $\Phi(x,t)$ is also $G$-invariant: $\Phi(g \cdot x,t) = g \cdot \Phi(x,t)$. In particular, if $x_0$ is a symmetric point (i.e. $g \cdot x_0 = x_0$ for all $g \in G$), then so is $\Phi(x_0,t)$ for any $t$.

Here we prove a version of this fact tailored for our setting; the proof easily extends to methods beyond gradient descent.

Any one of our geometric functionals $F = W, A, V$ or $M$ is an $O(3)$-invariant functional, i.e.

$$\forall g \in O(3), \quad F(g(V)) = F(V), \quad g(V) := (gv_1, \ldots, gv_N) \text{ for } V = (v_1, \ldots, v_N).$$

the non-smoothness of $W_{\text{Bobenko}}$; recall Section 3.2. The problems in this section are smooth and we are only exploiting its symmetry preserving property.
There is yet another invariance, namely invariance under simplicial isomorphisms. For a face list $F$ in a triangle mesh that specifies the simplicial complex structure of the mesh, there is a subgroup, denoted by $S(F)$, of the permutation group of $1,\ldots,N$ that corresponds to the group of simplicial isomorphisms. In other words, re-labelling the vertex indices according to the permutations in $S(F)$ gives the same triangulation (i.e. the geometric realization of the simplicial complex.) Our geometric functionals must satisfy
\begin{equation}
\forall \pi \in S(F), \quad F(\pi(V)) = F(V), \quad \pi(V) := (\nu_{\pi(1)}, \ldots, \nu_{\pi(N)}).
\end{equation}

**Definition 5.1.** Let $G$ be a finite subgroup of $O(3)$. A simplicial complex, specified by a face list $F$ of a mesh, is said to support (the symmetry group) $G$ if $G$ is isomorphic to some subgroup of $S(F)$. In this case, we denote the correspondence by
\[ G \ni g \longleftrightarrow \pi_g \in S(F). \]

A mesh $(V,F)$ in $\mathbb{R}^3$ is called $G$-symmetric if for every $g \in G$, $g(V) = \pi_g(V)$. Equivalently, if we define the group action of $G$ on the manifold $M = \mathbb{R}^{N \times 3}$ by
\[ g \cdot V := \pi_g^{-1}(g(V)), \quad g \in G, \]
then (with $F$ fixed) $V$ is $G$-symmetric if and only if $g \cdot V = V$ for all $g \in G$.

---

\footnote{The Schoenflies notation can be used to refer to any one of the possibilities of such a $G\triangleleft O(3)$.}
In this setting, the set of symmetric points is a linear submanifold of \( \mathbb{R}^{N \times 3} \). Note that whenever the underlying simplicial complex supports the symmetry group \( G \), then by (5.1) and (5.2) \( F \) is \( G \)-invariant in the sense that
\[
F(g \cdot \mathcal{V}) = F(\mathcal{V}), \quad \forall g \in G. \tag{5.3}
\]

A gradient flow \( \mathcal{V}(t) \) of \( F \) satisfies \( \dot{\mathcal{V}}(t) = -\nabla F(\mathcal{V}(t)) \) and the corresponding gradient descent algorithm satisfies
\[
\mathcal{V}_{k+1} = \mathcal{V}_k - \alpha_k \nabla F(\mathcal{V}_k).
\]

By the chain rule applied to (5.3), \( g^T \cdot \nabla F(g \cdot \mathcal{V}) = \nabla F(\mathcal{V}) \). Then, by orthogonality, \( \nabla F(g \cdot \mathcal{V}) = g \cdot \nabla F(\mathcal{V}) \). We can see that a gradient descent, which we formally denote by \( \text{GD}(\mathcal{V}, \alpha) := \mathcal{V} - \alpha \nabla F(\mathcal{V}) \), satisfies the transformation property:
\[
\text{GD}(g \cdot \mathcal{V}, \alpha) = g \cdot \mathcal{V} - \alpha \nabla F(g \cdot \mathcal{V}) = g \cdot \text{GD}(\mathcal{V}, \alpha).
\]

In particular, gradient descent preserves the symmetry of the previous iterate. It is also easy to check that the quasi-Newton BFGS method is symmetry preserving.

In our application, \( F \) can be chosen to be the penalty function of a Helfrich problem:
\[
F(\mathcal{V}; \mu) := W(\mathcal{V}) + \frac{\mu}{2} \left\{ (A(\mathcal{V}) - A_0)^2 + (V(\mathcal{V}) - V_0)^2 + (M(\mathcal{V}) - M_0)^2 \right\}, \quad \mu > 0.
\]

Since \( F \) inherits the \( O(3) \)-invariance from its constituent objective and constraint functions, any one of the gradient descent, Newton or BGFS methods applied to \( F \) would furnish a symmetry preserving algorithm for approximately solving the constrained optimization problem. Note that the symmetry preservation property holds regardless of the penalization parameter \( \mu \) or the line search parameters \( \alpha \), which vary from iteration to iteration. However, the penalization parameter \( \mu \) has to be big enough in order for the constraints to be approximately satisfied.

Sophisticated BFGS-SQP based solvers such as SNOPT, GRANSO and \texttt{fmincon} are not simply based on applying the BFGS method to the penalty function, and those who engineer these solvers do not have the symmetry preserving property in mind. It is therefore interesting to see from the experiments that the BFGS-SQP method used in GRANSO preserves symmetry almost perfectly, even in the presence of roundoff errors. We shall present a formal justification of this observation in a separate report.

In principle, we can optimize in a symmetry preserving way by optimizing only over the degrees of freedom that determine the control mesh up to the desired symmetry, as is done in Brakke’s Surface Evolver or \cite{B}. This has the added advantage of reducing the dimensionality of the problem, but requires an extra effort in coding and algorithmic development. GRANSO, or any solver with the same symmetry preserving property, frees us from the latter.

5.3. Comparison III: Symmetric vs ‘best’ minimizers. The example in the previous section says the obvious: applying a symmetry preserving optimization algorithm to an initial guess with the \textit{wrong} symmetry is not going to solve the problem. What if we have the correct symmetry? Our next experiment, based on comparing different solvers, shall reveal a rather subtle feature \textit{not} of the optimization algorithms, but of the optimization problems themselves.
We consider the genus 0 Helfrich problem with \((v_0, m_0) = (0.8, 1.2)\). It is believed that the minimizer is a surface of revolution with a pear shape. The two initial meshes in Figure 14(a) have extra symmetries (octahedral and \(D_{4h}\) symmetry) not possessed by a general surface of revolution. Applying the GRANSO solver, which presumably preserves symmetry, would fail to yield the correct solution; it is indeed what we observe from computation. If we use instead GRANSO with a \(D_n\)-symmetric initial mesh with no extra symmetry, we expect to approach the correct minimizer with a \(D_n\)-symmetric approximation, and it is again what we observe from computation; see Figure 16(i). In the case of the C2g0 and Loop SS methods, SNOPT and fmincon break the symmetry, as shown in Figure 16(ii) and (iii). If one zooms into the control meshes, one sees that the \(D_3\) symmetry is clearly broken around the neck area (indicated by the arrows in the relevant panels). Note that the neck area is also where the absolute Gauss curvature of the pear surface is the highest.

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{figure16.png}
\caption{Numerical solutions of the Helfrich problem with \((v_0, m_0) = (0.8, 1.2)\). For the PL method, we use \(W_{\text{EffAreaCur}}\) and \(M_{\text{Steiner}}\) and \(\lambda = 2\), our use of fmincon fails to give a feasible point, while GRANSO and SNOPT work well. GRANSO preserves the \(D_3\) symmetry of the initial mesh, while SNOPT and fmincon break the symmetry. The numerical values are the corresponding (true or PL) Willmore energies. The symmetry-breaking cases give slightly lower Willmore energies than the symmetry-preserving counterparts.}
\end{figure}

The asymmetric approximations to the (presumably symmetric) solution, produced by SNOPT and fmincon, have slightly smaller Willmore energies than those of the symmetric approximations produced by GRANSO. We believe that this is not caused by roundoff errors or truncation errors from numerical integration.
In (i), GRANSO terminates gracefully with a stationarity condition satisfied up to a tolerance. This suggests that GRANSO produces a $D_3$-symmetric critical point of the problem that is a local minimizer of the $(v_0, m_0) = (0.8, 1.2)$ Helfrich problem with the added $D_3$-symmetry constraint. By the principle of symmetric criticality, this critical point must also be a critical point of the full Helfrich problem (without any symmetry constraint.) Since (ii) and (iii) strongly indicate that this critical point is neither a local nor global minimizer, we are led to believe that it is a saddle point of the Helfrich problem.

We may contrast this experiment with that in Figure 11(c)-(d). We believe that the PL surface with an octahedral symmetry, as shown in Figure 11(d), is an absolute minimizer there.

5.4. $W$-minimizers of genus $g$ and Lawson’s $\xi_{g,1}$ surfaces. It is conjectured that the stereographic images of Lawson’s minimal surface $\xi_{g,1}$ in $S^3$ are the only $W$-minimizer of genus $g$ in $\mathbb{R}^3$. The term ‘Willmore conjecture’ – now the celebrated Marques-Neves theorem – refers to the $g = 1$ case of this more general conjecture. Here, we use our SS method to illustrate this conjecture for $g = 2$. There are two stereographic projections of $\xi_{g,1}$ from $S^3$ to $\mathbb{R}^3$ that give the resulting surfaces a $D_{2h}$ or $D_{3h}$ symmetry. We create initial control meshes with these two symmetries and solve the Willmore problem using our Loop SS method and the symmetry-preserving GRANSO solver. We also apply an iterative refinement, exploiting the underlying subdivision structure, to improve the accuracy of the solutions. The resulting surfaces are visually the same as the corresponding stereographic projections of Lawson’s $\xi_{2,1}$, and have the same Willmore energy of approximately 21.9. (The two surfaces are meant to be Möbius transformations of each other.) See Figure 17.

The same code for generating Figure 17, available in the Wmincon package, works for any genus $g \geq 1$. We have used it to empirically verify the generalized Willmore conjecture for up to genus $g = 6$.

6. Conclusion and future work

Admittedly, our analyses in Section 5 only address one fundamental aspect of what we may expect from the PL and SS methods. Even so, both the positive and negative results there have much room for improvements.

The positive result Theorem 3.8 asserts that some subsequence of a sequence of discrete minimizers converges to a minimizer of the continuous problem; the result does not say anything about the rest of the sequence. Does it mean that some elements of our hard-earned sequence $x_j$ may have nothing whatsoever to do with any minimizer of the continuous Willmore problem, even for arbitrarily large $j$? We do not believe so, as we may use Theorem 3.8 repeatedly in the following way:

- Apply the theorem to extract a convergent subsequence (convergent in the sense stated in the theorem) from $x_j$.
- Remove this subsequence from the original sequence, the remaining sequence is still a minimizing subsequence.
- Apply the theorem again to extract a convergent subsequence from the remaining sequence.
- ...

This is a free offprint provided to the author by the publisher. Copyright restrictions may apply.
Loosely speaking, this process should exhaust the whole sequence, and hence it seems plausible we can partition the original sequence $x_j$ into subsequences each of which converges to some Willmore minimizer. We therefore expect the following to hold true:

**Conjecture 6.1.** Let $x_j$ be any minimizing sequence in the setting of Theorem 3.2. (In particular, $x_j$ can be an approximate $W$-minimizer over $\text{Imm}_{S^j}$, as in Theorem 3.8.)

(I) There is a sequence of Möbius transformations $G_j$ in $\mathbb{R}^3$ such that the sequence $G_j \circ x_j(K)$ of surfaces can be partitioned into subsequences each of which converges in Hausdorff distance to some Möbius representative of a genus $g$ Willmore minimizer.

(II) If we further assume that the Willmore minimizer of genus $g$ is unique up to Möbius transformations, as would be implied by the generalized Willmore conjecture (see Section 5.4), then the whole sequence of surfaces $G_j \circ x_j(K)$ in (I) converges in Hausdorff distance to some Möbius representative of a genus $g$ Willmore minimizer.
The situation for the Canham and Helfrich problems is more challenging. On the geometric analysis side, even the existence problem for the Canham problem is not resolved for every positive genus and all isoperimetric ratios; see [64] and [46]. (In the genus zero case, Schygulla [64] solves the existence problem for all isoperimetric ratios.) On the numerical analysis side, a key difficulty is to prove the corresponding density result Corollary 3.7 with a fixed isoperimetric ratio constraint, i.e.

**Conjecture 6.2.** For any \( v \in (0,1) \), \( \bigcup_j \text{Imm}^v_{S_j} \) is dense in \( \text{Imm}^v_X(K) \), where \( \text{Imm}^v_{S_j} \) is the space of all immersed Loop subdivision surfaces over a closed oriented \( K \) with isoperimetric ratio \( v \) and \( \text{Imm}^v_{W^{2,2}\cap C^{1}}(K) \) is the space of elements in \( \text{Imm}_{W^{2,2}\cap C^{1}}(K) \) with isoperimetric ratio \( v \).

If this difficulty can be overcome, we expect that a result similar to Theorem 3.8 can be obtained for the Canham problem based on the existence results for the Canham problem established in [64] and [46]. The Helfrich problem is out of reach for now.

The negative results Propositions 3.9 and 3.12 are established on a case by case basis. Perhaps there is a universal negative result asserting that any consistent PL Willmore energy would fail in a similar manner. Besides the very special genus 0 cases for \( W_{\text{NormalCur}} \) and \( W_{\text{Bobenko}} \) (recall Figure 11), it appears that the space of PL surfaces (of a fixed combinatorial type) is ‘too big’ that minimizing any PL \( W \)-energy over it would always take us to some PL surface inconsistent with any smooth surface. It is an open question to formulate and prove this speculation.

A deeper mathematical study of these issues and the proposed regularization in Section 4 may lead to understandings of other non-conforming methods, such as the ones proposed in [9] [62]. However, as mentioned in the introduction, the methods in [9] [62] are in the spirit of ‘discretizing a minimization,’ whereas the methods studied in this article are in the spirit of ‘minimizing a discretization.’ It is an open question to see if these different methods can be analyzed in a coherent way.

It will be interesting to address rate of convergence issues in the future. Arden’s result [1] (Theorem 2) establishes a rate of convergence of Loop subdivision functions in \( W^{2,2} \), but we believe that the rates established there are suboptimal. Also, the approximation rates of other schemes (e.g. [15]) are yet to be explored. Regardless, further techniques are needed for transferring any approximation rate result in the pure approximation setting to an approximation rate result for the geometric variational problems. To this end, we expect knowledge on the second variation of the Willmore energy, such as the results by Weiner [71], will be helpful.

**Acknowledgments**

The second author is indebted to Tom Duchamp and Aaron Yip for extensive discussions and many of their insightful remarks. We also thank Tim Mitchell, Michael Overton, Justin Smith, and Shawn Walker for help.

**References**


CITIGROUP GLOBAL MARKETS INC., 390 GREENWICH STREET, NEW YORK, NEW YORK 10013
Email address: jingmchen@gmail.com

DEPARTMENT OF MATHEMATICS, DREXEL UNIVERSITY PHILADELPHIA, PENNSYLVANIA 19104
Email address: yut@drexel.edu

DEPARTMENT OF MATHEMATICS, DREXEL UNIVERSITY PHILADELPHIA, PENNSYLVANIA 19104
Email address: pbrogan12@gmail.com

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MASSACHUSETTS AT AMHERST, AMHERST, MASSACHUSETTS 01003
Email address: kusner@math.umass.edu

CENTER FOR COMPUTATIONAL ENGINEERING, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02142
Email address: yiliny@mit.edu

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING, UNIVERSITY OF PITTSBURGH, PITTSBURGH, PENNSYLVANIA 15261
Email address: anz37@pitt.edu