

Repulsive Surfaces

CHRIS YU, Carnegie Mellon University
CALEB BRAKENSIEK, Independent Researcher
HENRIK SCHUMACHER, RWTH Aachen University
KEENAN CRANE, Carnegie Mellon University

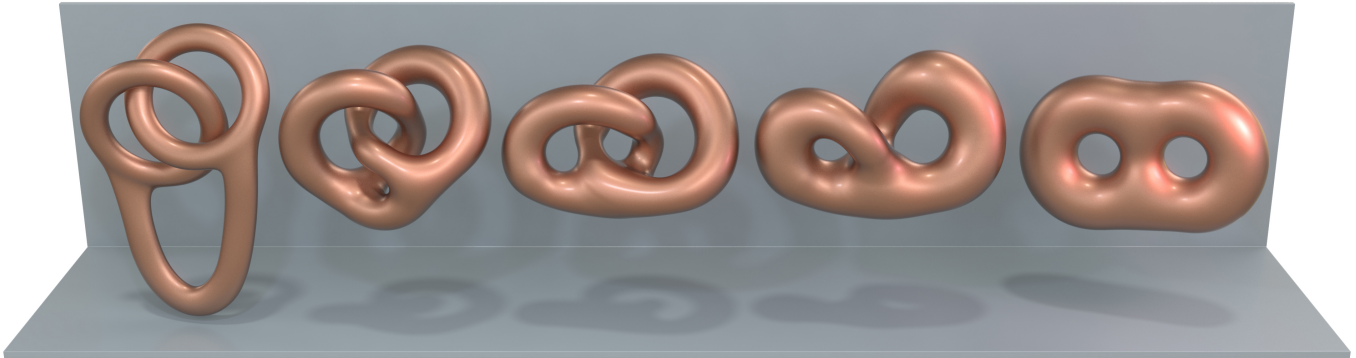


Fig. 1. We develop a numerical scheme for optimizing surface geometry while avoiding self-intersections. Here, we automatically find an unexpected transition between linked and unlinked states of a pair of “handcuffs” by simply minimizing a repulsive energy.

Functionals that penalize bending or stretching of a surface play a key role in geometric and scientific computing, but to date have ignored a very basic requirement: in many situations, surfaces must not pass through themselves or each other. This paper develops a numerical framework for optimization of surface geometry while avoiding (self-)collision. The starting point is the *tangent-point energy*, which effectively pushes apart pairs of points that are close in space but distant along the surface. We develop a discretization of this energy for triangle meshes, and introduce a novel acceleration scheme based on a fractional Sobolev inner product. In contrast to similar schemes developed for curves, we avoid the complexity of building a multiresolution mesh hierarchy by decomposing our preconditioner into two ordinary Poisson equations, plus forward application of a fractional differential operator. We further accelerate this scheme via hierarchical approximation, and describe how to incorporate a variety of constraints (on area, volume, *etc.*). Finally, we explore how this machinery might be applied to problems in mathematical visualization, geometric modeling, and geometry processing.

CCS Concepts: • **Computing methodologies** → **Shape modeling**; • **Mathematics of computing** → **Continuous optimization**.

Additional Key Words and Phrases: Computational design, shape optimization, surfaces

Authors’ addresses: Chris Yu, Carnegie Mellon University; Caleb Brakensiek, Independent Researcher; Henrik Schumacher, RWTH Aachen University; Keenan Crane, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM. 0730-0301/2021/12-ART268 \$15.00
<https://doi.org/XX>

ACM Reference Format:

Chris Yu, Caleb Brakensiek, Henrik Schumacher, and Keenan Crane. 2021. Repulsive Surfaces. *ACM Trans. Graph.* 40, 6, Article 268 (December 2021), 19 pages. <https://doi.org/XX>

1 INTRODUCTION AND RELATED WORK

A geometric functional assigns a real-valued score $\mathcal{E}(f)$ to each immersion $f : M \rightarrow \mathbb{R}^m$ of a surface M . Such functionals serve as regularizers in many geometric problems, helping to define a unique solution, or simply making the geometry “nicer” in some sense. For instance, in geometric modeling they are used to smoothly interpolate given boundary data [Bucur and Butazzo 2006], in mathematical visualization they can be used to endow an abstract surface with a concrete geometry [Chern et al. 2018], and in digital geometry processing they are used for, *e.g.*, hole filling [Clarenz et al. 2004] or denoising of measured data [Elsey and Esedoğlu 2009]. However, classic functionals ignore a basic requirement of many applications—namely, that surfaces should not exhibit (self-)intersections. This condition is critical when surfaces represent physical membranes (*e.g.*, in biological simulation), boundaries of solid objects (*e.g.*, for digital manufacturing), or certain mathematical objects (*e.g.*, isotopy classes of embeddings). It is therefore surprising that, to date, there has been little focus on interpenetration in variational surface modeling. We build on the recent framework of Yu et al. [2021], extending their machinery for repulsive curves to the more computationally demanding case of surfaces.

Curvature Functionals. A basic functional for surfaces is total surface area; gradient descent on area leads to *mean curvature flow*, which has been used for surface denoising [Desbrun et al. 1999] but can develop non-smooth singularities or pinch-off artifacts. Though

efforts have been made to desingularize this flow [Kazhdan et al. 2012], sharp peaks and cusps are ultimately impossible to detect from area alone. For this reason, functionals used in geometric modeling typically incorporate curvature information—most prominently the *Willmore energy* $\mathcal{E}_W(f) := \int_M (H^2 - K) dA$, where H and K are the mean and Gaussian curvatures, *resp.* Significant work has focused on numerical optimization of Willmore energy [Droske and Rumpf 2004; Bobenko and Schröder 2005; Crane et al. 2013b; Soliman et al. 2021], but since this energy is Möbius invariant, it effectively provides a notion of regularity for surfaces in the 3-sphere S^3 , rather than the Euclidean \mathbb{R}^3 . In the context of geometric modeling, this means that even minimizers of Willmore energy can have poor distributions of curvature—see for example Fig. 4, *bottom left*. Though further energies have been developed to address such issues [Moreton and Séquin 1992; Joshi and Séquin 2007], none of these energies avoid intersections.

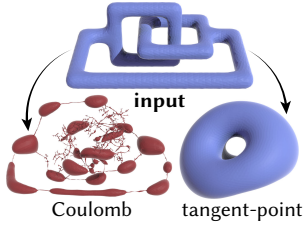


Fig. 2. Ad-hoc schemes such as vertex-vertex Coulomb forces do not correspond to a meaningful smooth energy and can be numerically unstable. Here we minimize Coulomb and tangent-point energies subject to a fixed area constraint.

Repulsive Forces. Collision response from physical simulation [Bridson et al. 2002] and contact mechanics [Wriggers and Zavarise 2004; Harmon et al. 2009] can be used to locally *resolve* contact, but does not help guide shape optimization toward a state that is far from interpenetration. Moreover, whereas level set methods ensure (by construction) that surfaces have no self-intersections, the *raison d’être* of such methods is to allow the surface topology to *change*, rather than to preserve it [Osher and Fedkiw 2006]. We instead consider “all-pairs” energies of the form

$$\mathcal{E}(f) = \int_{M \times M} k(x, y) dx_f dy_f,$$

where dx_f denotes the area element induced by f , and the kernel $k : M \times M \rightarrow \mathbb{R}$ is designed to discourage self-contact.

A tempting choice is a Coulomb-like potential

$$k_{\text{Coulomb}}(x, y) = \frac{1}{|f(x) - f(y)|^\alpha}$$

for some falloff parameter $\alpha > 0$. On a triangle mesh, this amounts to penalizing the distance between all pairs of vertices. However, as noted by Yu et al. [2021, Section 3.1], this energy is too weak to prevent collision for $\alpha < 2$ (*i.e.*, intersecting surfaces have finite energy), and yet ill-defined in the continuum limit for $\alpha \geq 1$. The issue is that there are *always* points y within an arbitrarily small geodesic distance $d(x, y) \geq |f(x) - f(y)|$ of any point $x \in M$, causing the energy to blow up. Numerically, ad-hoc vertex-vertex penalties are hence unstable and highly unpredictable (Fig. 2).

For curves, [O’Hara 1991] regularizes the Coulomb-like potential by subtracting the contribution of points that are nearby on the curve:

$$k_{\text{Möbius}}(x, y) = \frac{1}{|f(x) - f(y)|^2} - \frac{1}{d(x, y)^2}.$$

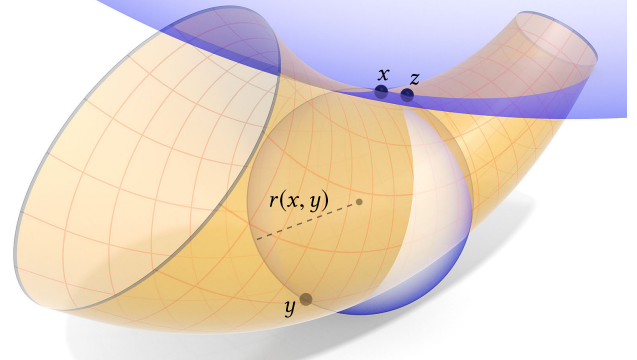


Fig. 3. For each pair of points x, y on the surface, the tangent-point energy considers the radius $r(x, y)$ of the smallest sphere tangent to x and passing through y , penalizing $1/r(x, y)$. Hence, the contribution will be very large for points y close in space but distant along the surface—and small for points z nearby along the surface, where the radius is huge.

The resulting energy is well-defined and strong enough to prevent collisions; it is referred as *Möbius energy* because it is invariant under Möbius transformations [Freedman et al. 1994]. However, Möbius invariance is actually a significant drawback for geometric modeling: it leads not only to uneven curvature ([Kusner and Sullivan 1998, Figure 5]), but also “tight spots” where points distant in S^3 become arbitrarily close when projected into \mathbb{R}^3 (see [Yu et al. 2021, Figure 3]). Moreover, the geodesic distance $d(x, y)$, though easy to compute for curves, is prohibitively expensive to compute for all pairs of points on a surface—much less to differentiate with respect to motions of the surface.

Tangent-Point Energy. For all these reasons, we instead consider the *tangent-point energy* introduced for curves by [Buck and Orloff 1995], and only recently extended to higher dimensions by [Strzelecki and von der Mosel 2013]. For each pair of points $x, y \in M$, consider the radius $r(x, y)$ of the smallest sphere tangent to $f(x)$ and passing through $f(y)$ (Fig. 3). The tangent-point kernel k is then proportional to $1/r(x, y)$; see Sec. 2.1 and [Strzelecki and von der Mosel 2018] for further discussion. Hence, points that are close in space only because they are also close along the surface are automatically ignored. This energy has several features that make it a prime candidate for repulsive surface optimization, namely:

- It provides an infinite barrier to self-intersection [Strzelecki and von der Mosel 2013].
- Like Willmore energy it penalizes bending [Yu et al. 2021, Section 3.2], preventing singularities and cusps.
- Unlike Willmore and Möbius energy, it is neither Möbius nor scale invariant, helping to evenly distribute curvature and avoid tight spots.
- Unlike Möbius energy, it does not require geodesic distances and instead depends only on quantities like surface normals N and extrinsic distances $|f(x) - f(y)|$ that are cheap to compute and easy to differentiate.

However, two significant challenges remain, namely, (i) deriving an inner product that accelerates optimization and (ii) efficiently inverting this inner product.

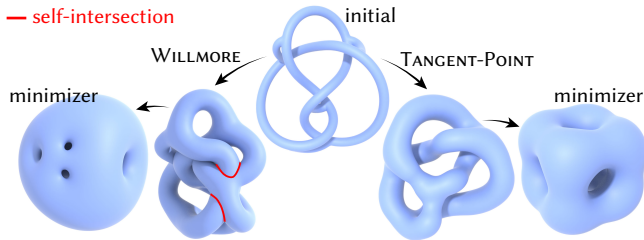


Fig. 4. Willmore energy does nothing to prevent intersections (in red), and can have minimizers that asymmetrically distribute curvature over the surface. *Right*: tangent-point energy avoids intersections and tends to provide a more uniform curvature distribution.

Accelerating Optimization. To integrate a parabolic gradient flow of order k with average node spacing h , one must typically take time steps of size around $O(1/h^k)$, which is prohibitively expensive for fine meshes. However, redefining gradient descent with respect to a different inner product can effectively turn it into a 0th-order equation—mitigating the time step restriction. This idea of *Sobolev gradients* has long been applied to surface flows [Pinkall and Polthier 1993; Renka and Neuberger 1995; Eckstein et al. 2007; Martin et al. 2013; Schumacher 2017; Soliman et al. 2021], and more recently to elastic energies in geometry processing [Kovalsky et al. 2016; Claici et al. 2017; Zhu et al. 2018]. However, all these works consider energies with integer-order differentials, whereas tangent point energy has a differential of *fractional* order. As recently shown by Yu et al. [2021], a fractional inner product performs far better than even integer Sobolev schemes, especially for finely-tessellated or highly-knotted curves. We adopt the same basic strategy, adapting it to surfaces.

Efficient Evaluation. A second challenge is the dramatic increase in problem size going from curves to surfaces: rather than $O(n^2)$ pairs of elements on a curve, we now have $O(n^4)$ element pairs on a surface (where $n \approx 1/h$). Hierarchical Barnes-Hut approximation is still sufficient to approximate the energy and its differential (Sec. 4), but we must now invert a dense inner product matrix with $O(n^4)$ entries. Yu et al. [2021] use a multigrid solver based on a simple multiresolution curve hierarchy, but rebuilding a multiresolution *surface* mesh hierarchy at each step is far more difficult and expensive. Our key insight is that the inverse of our fractional operator can be approximated by the inverse of two ordinary (integer-order) Laplace operators, together with *forward* application of a lower-order fractional derivative (Sec. 5.2). Since this decomposition is only approximate in the discrete setting, we use it to precondition an iterative linear solver (GMRES) that does not require a mesh hierarchy. Note that simply *omitting* distant terms, or using a compactly-supported kernel (*à la* Li et al. [2020, Section 4.2]) is not a viable option, since global interactions play a key role in many applications (Sec. 9). Moreover, due to the logarithmic nature of our hierarchical approximation scheme, dropping far-field terms reduces cost by only about 20%—at the cost of a lower-quality descent direction.

Overall, then, we apply preconditioning in two distinct ways: first to precondition the flow, then to precondition the linear solver used to compute the flow direction (see Sec. 5.2 for further comments). The result is a highly effective scheme for the challenging surface case, with performance suitable for iterative design exploration (see Sec. 8.4 for detailed statistics).

1.1 Contributions

Overall, in this paper we develop

- the first discretization of tangent-point energy for surfaces,
- a novel preconditioner that avoids a multigrid hierarchy,
- a hierarchical solver that scales to large meshes, and
- a framework for handling auxiliary constraints and penalties.

We also perform a preliminary investigation of applications in geometric modeling, mathematical visualization, and geometry processing. Notably, although one can prove that minimizers of the tangent-point energy exist [Kolasiński et al. 2015, Theorem 2], these proofs are non-constructive. Since we provide the first discretization and optimization procedure for the tangent-point energy on surfaces, we obtain the very first glimpse (experimentally) at what some of these surfaces might actually look like.

We begin by defining our problem in the smooth setting (Sec. 2), followed by a novel discretization of the tangent-point energy and a basic numerical strategy for minimizing it subject to constraints (Sec. 3). We then significantly accelerate this strategy in two distinct ways. First, we choose an inner product in the smooth setting that vastly improves the convergence of the gradient flow (Sec. 2.4). Second, in the discrete setting, we propose a preconditioner that dramatically reduces the cost of solving for the descent step (Sec. 5). We also accelerate evaluation of the energy and its derivatives, as well as dense matrix-vector products, using hierarchical acceleration (Sections 4 and 5). We then consider dynamic remeshing (Sec. 6) and auxiliary penalties and constraints (Sec. 7), which enable a variety of potential applications (Sec. 9); Sec. 8 provides numerical validation.

2 SMOOTH FORMULATION

In this section we define the smooth tangent-point energy \mathcal{E}^P , and give some remarks on the order of derivatives appearing in its differential $d\mathcal{E}^P$. Knowing the order of the differential is essential for accelerating the gradient flow $\frac{d}{dt}f = -d\mathcal{E}^P(f)$, since it enables us to define a new inner product (in Sec. 2.4) with respect to which gradient flow effectively becomes a 0th-order equation. (Readers may wish to consult Yu et al. [2021, Section 4.1] for a slower-paced, didactic introduction to this approach.) Hence, the integrator developed in Sec. 3 will be able to take dramatically larger time steps whose size does not depend strongly on mesh resolution (Fig. 5).

2.1 Energy

As discussed in Sec. 1, we can define a repulsive energy by considering the *tangent-point radius* $r_f(x, y)$, defined as the radius of the smallest sphere tangent to $f(x)$ and passing through $f(y)$ (Fig. 3). Letting $N_f(x)$ be the unit normal at x , this radius can be computed as

$$r_f(x, y) = \frac{|f(x) - f(y)|^2}{2|P_f(x)(f(x) - f(y))|}, \quad (1)$$

where $P_f(x) = N_f(x)N_f(x)^\top$ denotes the orthogonal projector onto the normal space at x . Note that expressing r_f via the projector avoids picking a sign for the normal, which will be useful in Sec. 5.1 (it is also valid for submanifolds of arbitrary dimension and codimension). Omitting the constant factor 2, the tangent-point kernel

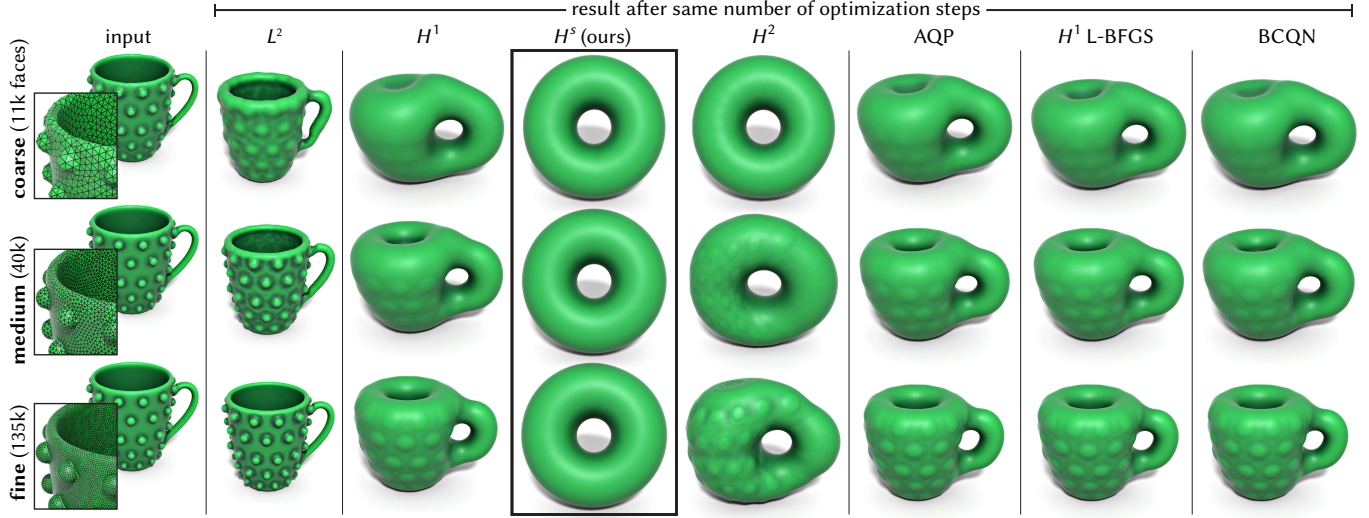


Fig. 5. Unlike other schemes, our fractional preconditioner does not suffer from a mesh-dependent time step restriction. Here for example we take 300 optimization steps of maximum size (determined by line search) for each scheme. As resolution increases, all methods but H^s make slower and slower progress. Note also that schemes based on H^1 preconditioning (H^1 , H^1 L-BFGS, AQP, BCQN) quickly eliminate high-frequency details but are slower to smooth the bulk shape; conversely, H^2 quickly smooths out the bulk shape but fine details remain. Using H^s for $1 < s < 2$ nicely handles both local and global features.

(due to Buck and Orloff [1995]) is then given by

$$k_{f,p}(x,y) := \frac{2^p}{r_f(x,y)^p} = \frac{|P_f(x)(f(x) - f(y))|^p}{|f(x) - f(y)|^{2p}} \quad (2)$$

for some $p > 0$, and hence the energy itself is

$$\mathcal{E}^p(f) := \iint_{M^2} k_{f,p}(x,y) dx_f dy_f. \quad (3)$$

While in principle it is possible to allow the exponents in the numerator and denominator to vary independently [Blatt and Reiter 2015], we use exponents $p, 2p$ (as above), which simplifies analysis. Note that, because $k_{f,p}$ has units m^{-p} (in meters) and \mathcal{E}^p is a double integral over an n -dimensional manifold, \mathcal{E}^p has units m^{2n-p} . Therefore, $p > 2n$ is required for the energy to be truly repulsive (i.e., to have units corresponding to inverse meters); otherwise, the energy could be reduced to 0 by simply shrinking the domain to a single point. As we deal with surfaces here ($n = 2$), $p > 4$ is sufficient. Unless otherwise noted, we use $p = 6$ for all examples in this paper.

2.2 Gradient Flow

Attempting to perform standard L^2 gradient descent on the tangent-point energy yields a flow

$$\frac{d}{dt}f = -d\mathcal{E}^p(f).$$

This flow exhibits poor convergence due to the presence of high-order spatial derivatives on the right-hand side, which even aggressive line search or general-purpose preconditioning (e.g., L-BFGS) cannot alleviate; see Fig. 5. However, we can obtain a different descent strategy by defining the gradient with respect to a different inner product. In particular, if A is the linear operator defining the

inner product, the descent equation becomes

$$\frac{d}{dt}f = -A^{-1}d\mathcal{E}^p(f). \quad (4)$$

An optimal choice of A will match the order of the differential, so that the right hand side no longer involves any spatial derivatives (hence avoiding a mesh-based time step restriction). We first establish the order of the differential $d\mathcal{E}^p$ in the surface case (Sec. 2.3), then define a *fractional Sobolev inner product* that matches this order (Sec. 2.2).

2.3 Order of the Differential

Though originally defined for curves, the tangent-point energy \mathcal{E}^p can be formulated for a quite broad class of n -dimensional sets $\Sigma \subset \mathbb{R}^m$ “with tangent planes,” that need not even be manifolds [Strzelecki and von der Mosel 2013]. In the case of 2-dimensional surfaces, one can argue (as discussed below) that $d\mathcal{E}^p$ is a nonlocal, nonlinear differential operator of *fractional* order $2(2 - 2/p) \in]3, 4[$, rather than integer order. This distinguishes the tangent-point energy from standard geometric energies like Willmore, and it is why we have to develop special tools for it.

In more detail: Strzelecki and von der Mosel [2013] show that if tangent-point energy is finite for some n -dimensional $\Sigma \subset \mathbb{R}^m$, then Σ must be an *embedded* submanifold of Hölder class $C^{1,\alpha}$, where $\alpha = 2 - 2n/p$. Intuitively: it must be free of self-intersections, and also fairly regular. This result is improved by [Blatt 2013], who establishes that $\mathcal{E}^p(\Sigma)$ is finite if and only if Σ is an embedded submanifold of *fractional Sobolev class* $W^{s,p}$, where $s = 2 - n/p$. In particular, this implies that Σ can be expressed as an embedding $f \in W^{s,p}(M; \mathbb{R}^m)$ for some smooth manifold M . For $n = 2$ we have $s \in]3/2, 2[$, so we inevitably have to deal with fractional Sobolev spaces. Knowing the natural habitat of \mathcal{E}^p is key because it allows for the following observation: the differential $d\mathcal{E}^p$ is a mapping from

$W^{s,p}$ to the dual space $(W^{s,p})^* = W^{-s,p}$. Hence it is plausible that $d\mathcal{E}^p$ reduces the differentiability of its argument by $2s = 2(2 - 2/p)$, as claimed above.

2.4 Inner Product

Standard (integer) Sobolev inner products are expressed via the Laplacian Δ . We likewise consider the *fractional Laplacian* of order $0 < 2\sigma < 2$ on \mathbb{R}^n , which can be expressed in integral form up to a constant factor as

$$\langle (-\Delta)^\sigma u, v \rangle_{L^2} = \iint_{\mathbb{R}^n} \frac{(u(x) - u(y))(v(x) - v(y))}{|x - y|^{2\sigma+n}} dx dy \quad (5)$$

for sufficiently smooth functions $u, v : \mathbb{R}^n \rightarrow \mathbb{R}$ [Kwaśnicki 2017]. While this formula only relates to \mathbb{R}^n , we can obtain an analogous operator L^σ of fractional order 2σ on functions $u, v : M \rightarrow \mathbb{R}$ by mimicking this expression on the n -dimensional manifold M :

$$\langle L^\sigma u, v \rangle_{L^2} = \iint_{M^2} \frac{(u(x) - u(y))(v(x) - v(y))}{|f(x) - f(y)|^{2\sigma+n}} dx_f dy_f. \quad (6)$$

Note that, for $p > 2n$, the order of $d\mathcal{E}^p$ is $2s = 2(2 - n/p) > 3$, which is outside the bounds $0 < 2\sigma < 2$. We can increase the order of this operator by introducing a first order derivative operator \mathcal{D}_f in the numerator, yielding a higher-order operator

$$\langle Bu, v \rangle_{L^2} = \iint_{M^2} \frac{\langle \mathcal{D}_f u(x) - \mathcal{D}_f u(y), \mathcal{D}_f v(x) - \mathcal{D}_f v(y) \rangle}{|f(x) - f(y)|^{2\sigma+n}} dx_f dy_f. \quad (7)$$

More precisely, we use $\mathcal{D}_f u(x) := du(x) df(x)^\dagger \in \text{End}(\mathbb{R}^m)$, where $df(x)^\dagger \in \text{Hom}(\mathbb{R}^m; T_x M)$ denotes the Moore-Penrose pseudoinverse of $df(x)$. If we now let $\sigma = s - 1$, then the operator B achieves the desired order $2s$.

Low order term. As proposed by Yu et al. [2021], we can get even better preconditioning in situations with close contact by adding an additional term of lower order, which in our case translates to

$$\langle B_0 u, v \rangle_{L^2} = \iint_{M^2} \frac{(u(x) - u(y))(v(x) - v(y))}{|f(x) - f(y)|^{2\sigma+n}} k_{f,2}(x, y) dx_f dy_f. \quad (8)$$

The inclusion of the tangent-point kernel $k_{f,2}(x, y)$ effectively distorts lengths in regions of high energy: as the local energy increases, so too does the apparent length induced by the inner product. As a result, self-intersecting configurations, having infinite energy, are so distant (if not infinitely so) that they are unlikely to be reached within a finite time. The kernel $k_{f,2}(x, y)$ is chosen here so that B and B_0 have the same units and thus behave similarly under scaling.

The overall operator $A = B + B_0$ then defines a positive-semidefinite inner product that consider throughout this work. The order of this inner product matches that of the Sobolev space $W^{s,2} = H^s$, hence we use the term H^s to refer to our fractional preconditioner.

3 DISCRETIZATION

We next discretize all components needed for our optimization scheme. The basic idea is to minimize tangent-point energy via gradient descent, preconditioned by our fractional inner product. In practice we also want to incorporate constraints, which we do by projecting the descent direction onto the tangent space of the constraint manifold, then projecting the surface itself onto this manifold. The overall algorithm for each descent step is:

ALGORITHM 1: Assembly of the exact discrete fractional operator L^σ

initialize $L^\sigma \leftarrow 0$

forall distinct pairs of faces S, T **do**

forall vertices i in S or T **do**

forall vertices j in S or T **do**

$L_{ij}^\sigma \leftarrow L_{ij}^\sigma + \frac{(\bar{\phi}_i(S) - \bar{\phi}_i(T))(\bar{\phi}_j(S) - \bar{\phi}_j(T))}{|X_f(S) - X_f(T)|^{2\sigma+2}} a_f(S) a_f(T)$

end

end

end

return L^σ

- (1) Assemble the derivative $dEP(f)$ of the energy (Sec. 3.1).
- (2) Construct the fractional operator $A = B + B_0$ (Sec. 3.2).
- (3) Solve Eq. 14 to obtain the descent direction \mathbf{x} .
- (4) Take a step in the direction \mathbf{x} using Armijo line search.
- (5) Project onto the constraint manifold of Φ (Sec. 3.3.2).

This initial algorithm is quite inefficient, and is accelerated in Sections 4 and 5. Sec. 5.4 gives the final accelerated algorithm.

3.1 Discrete Energy

On a triangle mesh $M = (V, E, F)$ with vertex coordinates $f : V \rightarrow \mathbb{R}^3$, we approximate the energy \mathcal{E}^p (Eq. 3) using midpoint quadrature. In particular, we define the *discrete tangent-point kernel* on a pair of faces $S, T \in F$ as

$$K_{f,p}(S, T) = \frac{|P_f(S)(X_f(S) - X_f(T))|^p}{|X_f(S) - X_f(T)|^{2p}}, \quad (9)$$

where $X_f(S)$ denotes the barycenter of face S for an embedding f . The full energy is then defined as a double sum

$$E^p(f) = \sum_{S \in F} \sum_{T \in F} K_{f,p}(S, T) a_f(S) a_f(T), \quad (10)$$

where $a_f(S)$ denotes the area of face S under embedding f . The differential $dEP(f)$ is then just the vector of partial derivatives of the discrete energy E^p with respect to $f \in \mathbb{R}^{3|V|}$, which can be evaluated via any standard technique.

3.2 Discrete Inner Product

The fractional operator L^σ is discretized as a $|V| \times |V|$ matrix with entries obtained from the right-hand side of Eq. 6. The rows and columns of L^σ are indexed by vertices, with entries

$$L_{ij}^\sigma = \sum_{S \in F} \sum_{T \in F, S \neq T} \frac{(\bar{\phi}_i(S) - \bar{\phi}_i(T))(\bar{\phi}_j(S) - \bar{\phi}_j(T))}{|X_f(S) - X_f(T)|^{2\sigma+2}} a_f(S) a_f(T), \quad (11)$$

where ϕ_i denotes the piecewise linear hat-function centered at vertex i , and where $\bar{\phi}_i(S)$ denotes its evaluation on the barycenter of S , i.e., $\bar{\phi}_i(S)$ is $1/3$ if vertex i is contained in face S and 0 otherwise. A naïve entry-by-entry assembly of L^σ via Eq. 11 would take time quartic in the number of mesh elements; however, since the summand vanishes for most pairs S, T , assembly can be done in quadratic time by iterating over mesh elements rather than matrix entries—see Alg. 1.

3.2.1 High- and Low-Order Terms. The high-order term of our discrete inner product $B \in \mathbb{R}^{|V| \times |V|}$ (corresponding to Eq. 7) is also assembled via Alg. 1, but using the summand

$$\frac{(\mathcal{D}_f \phi_i(S) - \mathcal{D}_f \phi_i(T), \mathcal{D}_f \phi_j(S) - \mathcal{D}_f \phi_j(T))}{|X_f(S) - X_f(T)|^{2\sigma+2}} a_f(S) a_f(T) \quad (12)$$

in place of the one in Eq. 11. Here \mathcal{D}_f is a discretization of \mathcal{D}_f . Intuitively, $\mathcal{D}_f u(S)$ is the derivative of the function $u = \sum_{i \in V} u_i \phi_i$ within the triangle S , and can be evaluated as

$$\frac{1}{2a_f(S)} (N_f(S) \times (u_i e_{jk} + u_j e_{ki} + u_k e_{ij}))^\top,$$

where i, j, k are the vertices of triangle S , and e_{ij}, e_{jk} , and e_{ki} are the unit edge vectors of S . The low-order matrix $B_0 \in \mathbb{R}^{|V| \times |V|}$ is likewise assembled using the summand

$$\frac{(\bar{\phi}_i(S) - \bar{\phi}_i(T))(\bar{\phi}_j(S) - \bar{\phi}_j(T))}{|X_f(S) - X_f(T)|^{2\sigma+2}} K_{f,2}(S, T) a_f(S) a_f(T) \quad (13)$$

using the discrete kernel $K_{f,2}(S, T)$ from Eq. 9. We then get a $|V| \times |V|$ matrix $A = B + B_0$ which acts on scalar functions. A corresponding operator A_3 for vector-valued functions $u, v : V \rightarrow \mathbb{R}^3$ is obtained by replacing each entry A_{ij} with the 3×3 block $A_{ij} \mathbf{I}_{3 \times 3}$ (where \mathbf{I} is the identity), giving a matrix of size $3|V| \times 3|V|$.

3.3 Constraints

3.3.1 Gradient Projection. Suppose we want our surface to satisfy a constraint $\Phi(f) = 0$ for some constraint function $\Phi : \mathbb{R}^{3|V|} \rightarrow \mathbb{R}^k$ (e.g., to control area or volume). The constrained descent direction \mathbf{x} can be obtained by solving the saddle point problem

$$\begin{bmatrix} A_3 & d\Phi(f)^\top \\ d\Phi(f) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} -dE^P(f) \\ 0 \end{bmatrix}, \quad (14)$$

where $d\Phi(f)$ denotes the Jacobian of Φ and $\lambda \in \mathbb{R}^k$ are Lagrange multipliers. The resulting descent direction \mathbf{x} will then be tangent to the constraint manifold $\{f \mid \Phi(f) = 0\}$.

Importantly, we must *always* have some kind of constraint: the matrix A_3 has a null space corresponding to translations of the surface, which means the unconstrained system $A_3 \mathbf{x} = -dE^P(f)$ does not have a unique solution. One simple solution is to fix the barycenter (Sec. 7.1.1), though other constraints such as pinned vertices (Sec. 7.2) can serve the same purpose.

3.3.2 Corrective Projection. Though our descent direction \mathbf{x} is now tangent to the constraint manifold, the embedding f may still drift away from the constraint manifold. Hence, after finding a feasible step size $\tau > 0$ via line search, we project the new state $f + \tau \mathbf{x}$ back onto the constraint manifold of Φ via Newton's method. In particular, we reuse the matrix from Eq. 14 and solve

$$\begin{bmatrix} A_3 & d\Phi(f)^\top \\ d\Phi(f) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ -\Phi(f + \tau \mathbf{x}) \end{bmatrix} \quad (15)$$

to obtain a Newton step \mathbf{h} , which we add to the updated embedding $f + \tau \mathbf{x}$. The step \mathbf{h} is the least-norm solution of the linear equation $d\Phi(f) \mathbf{h} = -\Phi(f)$, with respect to the metric encoded by A_3 . This correction can be repeated several times if the constraint violation is not sufficiently small—though for the constraints we explored a single step was always sufficient.

4 ENERGY AND DERIVATIVE EVALUATION

The algorithm of Sec. 3 is bottlenecked by several operations of at least quadratic complexity. The first bottleneck is evaluation of the energy E^P and its derivatives dE^P , which requires iteration over all pairs of elements. We thus use hierarchical Barnes-Hut approximation [Barnes and Hut 1986] to accelerate computation.

Note that the kernel $K_{f,p}(S, T)$ (Eq. 9) involves only the barycenters of S and T , and the normal projector of S . To clarify this dependence, we here write it as $K_p(X_f(S), P_f(S); X_f(T))$, with

$$K_p(X, P; Y) := \frac{|P(X - Y)|^p}{|X - Y|^{2p}}. \quad (16)$$

4.1 Approximate Energy

At each step we construct a bounding-volume hierarchy (BVH) for the current face set, which is used to evaluate the Barnes-Hut approximation

$$\tilde{E}^P(f) = \sum_{S \in F} \sum_{\mathcal{I} \in \text{adm}(S)} K_p(X_f(S), P_f(S); X_{\mathcal{I}}) a_f(S) a_{\mathcal{I}}. \quad (17)$$

Here $\text{adm}(S)$ is the set of BVH nodes \mathcal{I} that are *admissible* with respect to face S , but have no admissible ancestors. In particular, a Taylor series analysis of Eq. 16 indicates that to keep error bounded, \mathcal{I} must either (1) contain a single face, or (2) satisfy

$$\max(r(S), r(\mathcal{I})) < \theta \text{dist}(S, \text{conv}(\mathcal{I})),$$

where $r(S), r(\mathcal{I})$ are the radii of the triangle S and the node \mathcal{I} , *resp.* (measured from their barycenters), $\theta \geq 0$ is a fixed parameter, dist is the minimum distance between sets, and conv is the convex hull.

In practice, we approximate the convex hull by the axis-aligned bounding box, leading to a stricter admissibility condition; the radius $r(\mathcal{I})$ is then the bounding box radius relative to the barycenter of all faces in \mathcal{I} . Likewise, rather than store individual faces in BVH leaves, we stop when a node contains no more than l faces ($l = 8$ in our experiments). Hence, if a leaf is not admissible, we directly sum over all its faces. Overall, each node \mathcal{I} stores the total area $a_{\mathcal{I}}$, barycenter $X_{\mathcal{I}}$, and axis-aligned bounding box $B_{\mathcal{I}}$ of its faces.

The separation parameter θ controls the approximation quality; the higher θ is, the faster the computation, but the less accurate the result. For $\theta = 0$, the sum degenerates to an all-pairs exact computation. Unless otherwise noted, we use $\theta = 0.5$ for all experiments. Note that we do not update the energy approximation during line search, but rather just fix the admissibility of clusters at the beginning of the search.

4.2 Approximate Derivative

Computing an approximate derivative with Barnes-Hut is not entirely analogous to computing the energy. For each vertex $v \in V$, we evaluate the sum

$$\tilde{\partial}_v E^P(f) = \sum_{S \in F(v)} \sum_{\mathcal{I} \in \text{adm}(S)} \partial_v (K_p(X_f(S), P_f(S); X_{\mathcal{I}}) a_f(S) a_{\mathcal{I}}),$$

where $F(v)$ denotes the set of faces containing v . This approximates both the forward and reverse terms that would be differentiated by v in an exact computation. Note that the outer sum over all $S \in F(v)$ for both energy and derivative evaluations can be evaluated as a parallel reduction without modification.

5 ITERATIVE LINEAR SOLVER

An even more significant bottleneck than the energy is the dense saddle point problem of Eq. 14. Rather than solving this problem via dense matrix inversion, we will solve it instead using GMRES, an iterative method. In general, efficient iterative methods require two key ingredients: fast matrix-vector products, and effective preconditioners. Here, we will describe methods for both.

5.1 Hierarchical Matrices

We use hierarchical matrices [Hackbusch 2015] to perform fast multiplication with A without explicitly assembling the matrix. In this section, we present the special case of rank-1 compression of kernel matrices, while noting that the original method can also perform higher-rank approximations. In our setting, a *kernel matrix* H is a matrix of size $|F| \times |F|$ whose entries are defined by

$$H_{ST} = (1 - \delta_{ST}) h(X_f(S), P_f(S); X_f(T), P_f(S)),$$

where $h : (\mathbb{R}^m \times \text{End}(\mathbb{R}^m)) \times (\mathbb{R}^m \times \text{End}(\mathbb{R}^m)) \rightarrow \mathbb{R}$ is a suitable kernel function. To motivate this approach, we first reduce the actions of the operators L^σ , B , and B_0 to the multiplication with certain kernel matrices.

5.1.1 Applying the operator L^σ . An elementary computation shows (see App. A) that the action of the discrete linear operator L^σ on a vector $\mathbf{v} \in \mathbb{R}^{|V|}$ can be written as

$$L^\sigma \mathbf{v} = 2 \mathbf{U}^\top [\text{diag}(\text{diag}(a_f)^{-1} H a_f) - H] \mathbf{U} \mathbf{v}.$$

Here a_f is the $|F|$ -vector of face areas; \mathbf{U} is the $|F| \times |V|$ -matrix that averages values on vertices onto faces and multiplies with the face areas; and H is the kernel matrix of size $|F| \times |F|$ to the singular kernel $h(X, P; Y, Q) = |X - Y|^{-(2\sigma+2)}$. \mathbf{U} is sparse, so we just need an efficient product with H to evaluate the full product with L^σ .

5.1.2 Applying the High-Order Term. To evaluate a matrix-vector product with $A = B + B_0$, it suffices to evaluate B and B_0 separately. This can be done in a similar fashion as for L^σ . For the higher order term B , we have the identity

$$B \mathbf{v} = 2 \mathbf{V}^\top [\text{diag}(\text{diag}(a_f)^{-1} H a_f) - H] \mathbf{V} \mathbf{v},$$

where $\mathbf{V} = \text{diag}(a_f) D_f$ with the discrete derivative operator D_f described in Sec. 3.2.1 and where the kernel h of the kernel matrix H is given by $h(X, P; Y, Q) = |X - Y|^{-(2(s-1)+2)}$.

5.1.3 Applying the Low-Order Term. Likewise, we can write the action of B_0 as

$$B_0 \mathbf{v} = 2 \mathbf{V}^\top [\text{diag}(\text{diag}(a_f)^{-1} H a_f) - H] \mathbf{V} \mathbf{v},$$

where the kernel h of the kernel matrix H is given by

$$h(X, P; Y, Q) = \frac{k_2(X, P; Y) + k_2(Y, Q; X)}{2|X - Y|^{2(s-1)+2}}.$$

5.1.4 Block Cluster Tree. EnergyAndDerivativeEvaluation In order to compress these kernel matrices, we reuse the BVH from Sec. 4, but additionally compute the average projector $P_I := a_I^{-1} \sum_{S \in I} a_f(S) P_f(S)$ for each node I . From this, we construct a *block cluster tree*, whose nodes (termed *block clusters*) consist of pairs of BVH nodes (termed

clusters in the following). For a given separation parameter $\chi \geq 0$, we say that two BVH clusters I and J are a *separated pair* if

$$\max(r(I), r(J)) \leq \chi \text{dist}(\text{conv}(I), \text{conv}(J)).$$

Here again, $r(I), r(J)$ are the radii of the nodes I, J as measured from their barycenters. The parameter χ controls the accuracy of the approximation; it will be discussed further in the next section. Then, denoting the BVH root by \mathcal{R} , we construct the block cluster tree by starting with the single pair $(\mathcal{R}, \mathcal{R})$, and iteratively splitting nonseparated nodes (I, J) into the Cartesian products of their constituents' children until all leaf nodes are either separated or cannot be split any further. In practice, the tree structure is not important to maintain; only the lists of leaf nodes matter. We refer to the separated leaf nodes of the block cluster tree as *admissible blocks* and to the others as *inadmissible blocks*; Fig. 6 illustrates the decomposition of the full matrix into these blocks.

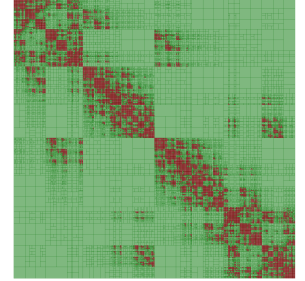


Fig. 6. A block decomposition of a kernel matrix H induced by a block cluster tree. Admissible blocks are shown in green, while inadmissible blocks are in red.

5.1.5 Hierarchical Multiplication. The block cluster tree allows us to perform approximate multiplication with a *kernel matrix* H as follows. Every pair of BVH clusters (I, J) corresponds to a block of H with rows indexed by I and columns by J . Let H_{IJ} denote this matrix block and let \mathbf{x}_J and $\mathbf{1}_J$ denote the slices of $\mathbf{x} \in \mathbb{R}^{|F|}$ and of the all-ones vector indexed by J , respectively. Then, for all leaf blocks (I, J) , we compute the product $\mathbf{y} = H \mathbf{x}$ in two steps:

(1) If (I, J) is inadmissible, then we multiply exactly:

$$\mathbf{y}_I \leftarrow \mathbf{y}_I + H_{IJ} \mathbf{x}_J.$$

(2) If (I, J) is admissible, we employ rank-one approximation:

$$\mathbf{y}_I \leftarrow \mathbf{y}_I + \mathbf{1}_I h(X_I, P_I; X_J, P_J) \mathbf{1}_J^\top \mathbf{x}_J.$$

Here, we can see more clearly the effect of χ . For $\chi = 0$, all blocks are considered inadmissible, and the action of H is evaluated exactly. For $\chi > 0$, the larger the value, the more blocks will be considered admissible and thus multiplied using the fast approximation in Step 2, leading to faster evaluation time – but also higher error, analogous to the θ parameter for Barnes-Hut. For our experiments, we found $\chi = 0.5$ to be a broadly acceptable value. Note that, while a straightforward implementation of these two steps is sufficient to evaluate the product, a much faster implementation can be obtained by employing multipole methods; see App. B for details.

5.2 Preconditioner

We can now efficiently evaluate matrix-vector products with A , but to efficiently solve $A \mathbf{x} = \mathbf{b}$ we still need a good *preconditioner* for A , i.e., an efficient approximation of A^{-1} . Since we do not explicitly build A , classical preconditioners (e.g., diagonal or incomplete Cholesky) cannot be used.

Instead, consider the Laplace-Beltrami operator Δ_f of the embedding $f : M \rightarrow \mathbb{R}^3$, which for a triangle mesh is given by the sparse cotan-Laplace matrix [Crane et al. 2013a, Section 6.4]. Our inner product A is closely related to the fractional Laplacian $(-\Delta_M)^s$, and has the same order $2s$. Evaluating the fractional power of a large sparse matrix is prohibitively expensive—but we can still obtain a cheap approximation of the inverse $(-\Delta_M)^{-s}$ via the factorization

$$(-\Delta_M)^{-s} = (-\Delta_M)^{-1}(-\Delta_M)^{2-s}(-\Delta_M)^{-1}.$$

What remains is a forward application of the fractional Laplacian $(-\Delta_M)^{2-s}$, to which we do not have direct access. Fortunately, since $0 < 2-s < 2$, we can replace $(-\Delta_M)^{2-s}$ with L^{2-s} (as per Sec. 2.4), whose action can efficiently be approximated *à la* Sec. 5.1.1. Thus, if we prefactor $(-\Delta_M)$, we can apply the preconditioner

$$\tilde{A}^{-1} := (-\Delta_M)^{-1}L^{2-s}(-\Delta_M)^{-1}$$

via just two back-substitutions and one hierarchical matrix-vector product—all of which are efficiently evaluated.

Note that despite having the same order as our operator A , \tilde{A}^{-1} is not directly suitable as an inner product: it approximates only the inverse of the high-order term, which means we cannot easily incorporate other inner product terms such as Eq. 8 or Sec. 7.3.5. Empirically, it provides poor preconditioning when applied to the gradient flow equation directly. However, it is highly effective as a preconditioner for GMRES, enabling efficient inversion of A (plus any auxiliary terms).

5.3 Schur Complement

Though we can now iteratively solve the *unconstrained* problem $Ax = b$, we must still be able to handle constraints. Empirically, we find that applying an iterative method directly to the saddle-point system (Eq. 14) exhibits poor convergence. Hence, we instead use the *Schur complement* [Zhang 2005] to handle the additional rows. In particular, let M be the saddle point matrix

$$M := \begin{bmatrix} A_3 & d\Phi^T \\ d\Phi & 0 \end{bmatrix}. \quad (18)$$

The Schur complement of M with respect to A_3 is

$$(M/A) = -d\Phi (A_3^{-1}d\Phi^T), \quad (19)$$

and the blocks of M^{-1} can be expressed as

$$\begin{bmatrix} A_3^{-1} + (A_3^{-1}d\Phi^T)(M/A_3)^{-1}d\Phi A_3^{-1} & -(A_3^{-1}d\Phi^T)(M/A)^{-1} \\ -(M/A)^{-1}(A_3^{-1}d\Phi^T)^T & (M/A)^{-1} \end{bmatrix}.$$

The operator A_3^{-1} can be applied using the iterative method just outlined; a product with A_3 is equivalent to three separate products with A . The complement M/A is dense, but has dimensions $k \times k$, corresponding to the number of scalar constraints. So, as long as k is small, $(M/A)^{-1}$ can be computed quickly. Thus, all blocks of M^{-1} are computed without having to invert a large matrix.

Moreover, to get the constrained descent direction \mathbf{x} , we need only the top-left block: letting $\mathbf{g} := dE^P(f)$, we compute the descent direction by directly applying the top-left block to \mathbf{x} , yielding

$$\mathbf{x} = A_3^{-1}\mathbf{g} + (A_3^{-1}d\Phi^T)(M/A)^{-1}d\Phi (A_3^{-1}\mathbf{g}). \quad (20)$$

Eq. 19 requires one application of A_3^{-1} per row of $d\Phi$. Eq. 20 contains three occurrences of A_3^{-1} , but $A_3^{-1}\mathbf{g}$ can be reused in both places where it appears, and $A_3^{-1}d\Phi^T$ can be reused from its earlier computation in Eq. 19. Thus, the method requires just $k + 1$ iterative solves, where k is the number of constraints.

5.3.1 Corrective Projection. We similarly use the Schur complement to solve Eq. 15 for the corrective step \mathbf{h} . Only the top-right block of the Schur complement is needed, giving the expression

$$\mathbf{h} = -(A_3^{-1}d\Phi^T)(M/A)^{-1}(-\Phi(f)). \quad (21)$$

(M/A) does not need to be recomputed, and $A_3^{-1}d\Phi^T$ can again be reused. Thus, constraint projection incurs no significant costs.

5.4 Accelerated Algorithm Overview

The accelerated algorithm is as follows:

- (1) Assemble the (approximate) derivative $dE^P(f)$ of the energy using Barnes-Hut (Sec. 4).
- (2) Construct a BVH that partitions the faces of the mesh, and use it to create a block cluster tree (Sec. 5.1.4).
- (3) Use the Schur complement to solve the constrained saddle point problem (Eq. 14).
 - (a) Evaluate products with $(A_3)^{-1}$ by using a matrix-free iterative method (e.g., GMRES), with the preconditioner from Sec. 5.2, and an initial guess of 0.
 - (b) Within the iterative method, evaluate products with A using the block cluster tree (Sec. 5.1.2, Sec. 5.1.3).
- (4) Take a step in the direction of \mathbf{x} using standard line search.
- (5) Reuse the Schur complement to project the resulting embedding onto the constraint manifold of Φ (Sec. 5.3.1).

If no constraints are imposed, then the algorithm can be simplified: step 3 can be replaced by a single iterative solve $A_3x = b$, and step 5 can be omitted entirely.

6 DYNAMIC REMESHING

Minimizing the tangent-point energy often induces large surface deformations that degrade triangle quality. We therefore use a dynamic remeshing scheme similar to the approach of Chen and Holst [2011]. The exact algorithm we use is as follows:

- (1) Edges with length greater than $3L_0/2$ are split and edges with length smaller than $L_0/2$ are collapsed, unless this operation would result in triangle foldover.
- (2) For N iterations:
 - (a) All edges that violate the Delaunay condition are flipped until no such flippable edges can be found.
 - (b) Vertex positions are smoothed by computing a displacement vector from neighboring triangles

$$u_i = \rho \frac{\sum_{S \in F(i)} a_f(S)(c_f(S) - f(i))}{\sum_{S \in F(i)} a_f(S)}.$$

Here $F(i)$ denotes the set of faces containing vertex i , $c_f(S)$ is the circumcenter of the triangle S , and $\rho < 1$ is a constant. This displacement is projected onto the tangent space of the vertex and added to the original position.

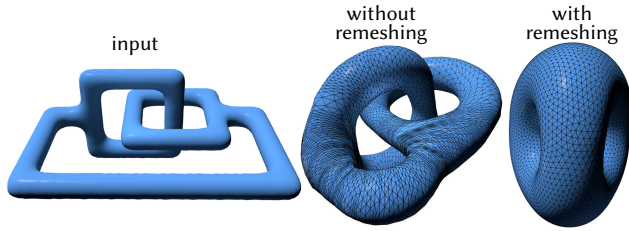


Fig. 7. Adaptive remeshing not only improves element quality—it also helps to avoid local minima where the surface gets “stuck.”

Our implementation uses $\rho = 0.5$ and $N = 5$; L_0 is set to the average edge length of the initial mesh and remains constant throughout. We apply this remeshing procedure at the end of each iteration, after the final step of Sec. 5.4. Remeshing is crucial to reaching minimizers of the tangent-point energy; without it, degrading triangle quality can impede or even halt progress, as seen in Fig. 7.

7 CONSTRAINTS AND PENALTIES

A variety of constraints and penalties can be imposed on the tangent-point energy, both for regularization of minimizers and for specific design purposes. In this section, we discuss the constraints and penalties that we have investigated; more are certainly possible, and in particular, combining the tangent-point energy with other classical surface energies could make for interesting future work.

7.1 Constraints

We consider four types of constraints: fixed barycenter, vertex pins, total area, and total volume.

7.1.1 Fixed Barycenter Constraint. A fixed barycenter constraint can be defined as

$$\Phi_C(f) = \frac{\sum_{i \in V} f(i) a_f(i)}{\sum_{i \in V} a_f(i)} - X_0,$$

where X_0 is the target barycenter location and $a_f(i)$ denotes the area associated to vertex i . Its Jacobian $d\Phi_C$ is a $3 \times 3|V|$ matrix consisting of $|V|$ copies of the 3×3 identity matrix appended horizontally. This constraint primarily serves to eliminate the nullspace of the fractional Laplacian (Sec. 3.3); either a barycenter constraint or at least one pin constraint must be added to every problem to be well-posed. For domains with multiple components, barycenters are constrained separately for each component.

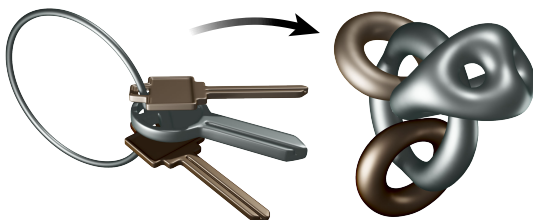


Fig. 8. To handle multiple components (as shown here), we fix the barycenter of each one during preconditioning, then add back in the mean motion of each component from the original L^2 -gradient after the solve.

Barycenter Motions. In some cases, it might be desirable to allow the barycenter to float freely, e.g., when a scene contains fixed obstacles for the surface to avoid. A simple modification enables this motion: compute the weighted average over all vertices of the L^2 gradient before projection, and then add the constant translation by that vector back to the descent direction after projection. For domains with multiple components, the average motion is computed separately for each component (Fig. 8).

7.1.2 Vertex Pin Constraints. A vertex pin constraint simply fixes a vertex to a position. Every pinned vertex i produces a constraint function $\Phi_{P_i}(f) = f(i) - f_0(i)$, where $f_0(i)$ is the pinned position. The Jacobian $d\Phi_{P_i}$ is a $3 \times 3|V|$ matrix, but the only nonzero entries consist of a single copy of the identity matrix in the block indexed by i . A pin also eliminates the nullspace of the Laplacian, so if any pins are used, then a barycenter constraint is unneeded.

7.1.3 Total Area Constraint. A total area constraint preserves the total surface area of the mesh, and can be written as

$$\Phi_A(f) = (\sum_{T \in \mathcal{F}} a_f(T)) - A_0,$$

where A_0 is the target area. The Jacobian $d\Phi_A$ is a $3|V|$ row vector with the area gradient at each vertex, which is equivalent to twice the mean curvature normal.

7.1.4 Total Volume Constraint. Likewise, a total (signed) volume constraint can be written as

$$\Phi_A(f) = \frac{1}{6} (\sum_{(ijk) \in \mathcal{F}} f(i) \cdot (f(j) \times f(k))) - V_0,$$

where V_0 is the target volume. For each vertex, the Jacobian $d\Phi_A$ is proportional to the area-weighted vertex normal.

7.2 Fast Positional Constraints

As discussed in Sec. 5.3, computing the Schur complement requires one iterative solve per constraint. For constraints such as fixed barycenters (3 rows per mesh component) or pinned vertices (3 rows per vertex), these solves incur significant cost. Hence, rather than use the full-blown Schur complement to handle these simpler linear constraints, we incorporate them into the matrix A . More precisely, if Φ_Q describes all constraints on position, then we replace A with another saddle-point matrix

$$A' := \begin{bmatrix} A & d\Phi_Q^T \\ d\Phi_Q & 0 \end{bmatrix},$$

Conceptually, the corresponding matrix A'_3 then becomes the upper-left block of the original saddle-point matrix M (Eq. 18), though in practice we work with the smaller matrix A' , since constraints on position can be separated into three independent components. The same constraint rows and columns are also appended to the integer Laplacians in the preconditioner (Sec. 5.2). A forward matrix-vector product with A' now entails sparse products with $d\Phi_C$ and $d\Phi_{P_i}$, in addition to the usual hierarchical product with A from Sec. 5.1.

Overall, this scheme enables efficient handling of Dirichlet boundary conditions (Fig. 9). Note that fast convergence in this scenario requires that orthogonality to these constraints be sufficiently similar under the two inner products defined by the integer Laplacian Δ and the fractional operator A . Empirically, this is the case for linear positional constraints, but is *not* the case for constraints such as



Fig. 9. Fast positional constraints enable Dirichlet boundary conditions. Here, minimizing tangent-point energy on a cylinder yields a nearly spherical geometry (left). Using tangent-point energy to regularize area minimization avoids the usual pinch-off singularities in minimal surfaces (center). We can also compute (near-)minimal surfaces with obstacles, in the spirit of [Giusti 1973] (right).

total area and volume. Thus, we reserve the Schur complement for these more difficult constraints.

7.3 Penalties

In addition to hard constraints, a number of soft penalty potentials can be added to regularize the flow in some way. These potentials are added directly to the objective function with some weighting coefficient alongside the tangent-point energy, and their gradients are accumulated in the same step.

7.3.1 Total Area and Volume Potentials. Soft penalties for total area and volume can be used in place of hard constraints, encouraging these quantities to stay close to their initial values without enforcing this exactly. For total area, the potential is defined as

$$\mathcal{E}_{\text{area}}(f) = ((\sum_{T \in \mathcal{F}} a_f(T))/A_0 - 1)^2.$$

The raw deviation is normalized by the initial area A_0 to make the penalty scale invariant. The total volume potential is defined analogously.

7.3.2 Static Obstacles. For practical modeling purposes, it may be desirable not to design an object in isolation, but instead to design it within its intended environment. To that end, we provide the ability to place “obstacles”, which are static meshes that exert a repulsive force on the optimization surface. These obstacles can be used to model surrounding environments such as rooms and the objects within them, which must be avoided by the object under design. From an obstacle O with embedding f_O , each point x in the domain experiences a repulsive potential equal to

$$\mathcal{E}_{\text{obs}}(x) = \sum_{S \in \mathcal{F}_O} |f_O(S) - x|^{-p} a_{f_O}(S)$$

with p matching the exponent of the tangent-point energy. Naïvely, this requires iteration over all faces of O , but Barnes-Hut can be used as in Sec. 4 to approximate the obstacle potential.

7.3.3 Implicit Obstacles and Attractors. Similarly to static mesh obstacles, one can also use implicit surfaces defined by signed distance fields as obstacles or attractors. Given a signed distance field $d : \mathbb{R}^3 \rightarrow \mathbb{R}$, the repulsive potential experienced at any point x due

to the implicit obstacle defined by $d(x) = 0$ is simply

$$\mathcal{E}_i(x) = d(x)^{-p}.$$

An implicit attractor, rather than repelling other objects away from it, pulls objects towards it. The attractive potential experienced at any point x is simply the reciprocal of the above, or

$$\mathcal{E}_a(x) = d(x)^p.$$

7.3.4 Boundary Length and Curvature. For meshes with boundary (e.g. Fig. 20), it may be beneficial to regularize the shape of the boundary curves. We support two potentials for this purpose. One is a regularizer on the length of the boundary, defined as

$$\mathcal{E}_b = (L - \sum_{e \in \partial M} l(e))^2,$$

where L is a target boundary length, and $l(e)$ is the length of boundary edge e . The other regularizes the curvature, and is defined as

$$\mathcal{E}_c = \sum_{v \in \partial M} \theta(v)^2 / \ell(v),$$

where $\theta(v)$ is the turning angle at vertex v , and $\ell(v)$ is the dual length (i.e., half the length of the two incident edges).

7.3.5 Willmore Energy. One can also add surface fairing energies such as the Willmore energy. For example, we use the following discrete variant of the squared mean curvature integral:

$$\mathcal{E}_{\text{Willmore}}(f) = f^T \mathbf{A} \mathbf{M}^{-1} \mathbf{A} f.$$

Here \mathbf{A} is the stiffness matrix of the cotan Laplacian and \mathbf{M} is the lumped mass matrix. Up to mass lumping, this is the discrete Willmore energy from [Dziuk 2008]. As suggested in [Eckstein et al. 2007; Schumacher 2017], we add an H^2 inner product term $\mathbf{A} \mathbf{M}^{-1} \mathbf{A}$ to the matrix that we invert in Sec. 5.

8 EVALUATION AND COMPARISONS

8.1 Consistency

Evaluating convergence of our discretization and approximation schemes is not straightforward, since to date there are only conjectures about what minimizers might look like (Sec. 9.1.1). Instead, we numerically study the *consistency* of our discretization: we generate several smooth surfaces, compute their true tangent-point energies, and compare to our discrete energy and its Barnes-Hut approximation.

The exact energy can be computed directly only for very simple shapes, like a sphere or torus of revolution. To get a more generic picture, we took the parameterized torus of revolution $f_0(\phi, \theta) = ((1 + \frac{1}{3} \cos(\phi)) \cos(\phi), (1 + \frac{1}{3} \cos(\theta)) \sin(\phi), \frac{1}{3} \sin(\theta))$ and perturbed it by a random trigonometric polynomial $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ of small magnitude (to ensure embeddedness) and small order (to obtain moderate curvature), obtaining a final smooth surface $f := f_0 + \Phi \circ f_0$. We computed $\mathcal{E}^p(f)$ up to 6 digits of precision via numerical integration with *Mathematica*'s `NIntegrate` command using the “LocalAdaptive” strategy. We then computed an affinely squeezed Delaunay triangulation of $[0, 2\pi] \times [0, 2\pi]$ and used it to sample the surface f . Nonuniformities in triangle shape and size were repaired by our remeshing routine (Sec. 6) followed by projection back onto the surface f . Using the resulting discrete surface f_h we computed the Barnes-Hut energy $\tilde{E}^p(f_h)$ (Eq. 17) for varying

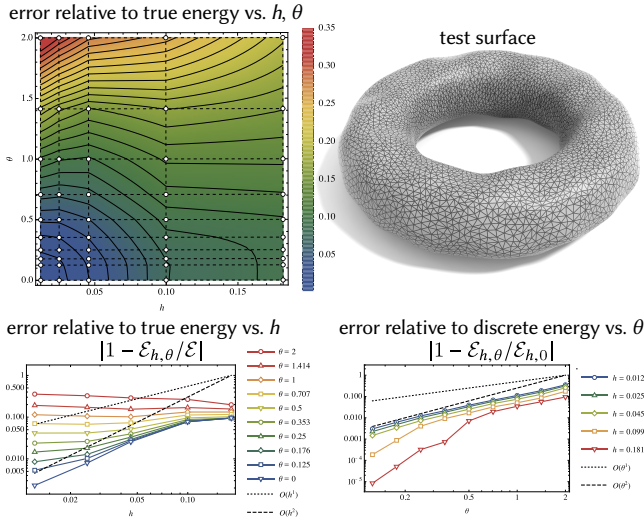


Fig. 10. Empirically, our discrete tangent-point energy appears to converge to the true smooth energy at a rate somewhere between $O(h)$ and $O(h^2)$; as expected, our Barnes-Hut approximation also converges to the discrete energy as $\theta \rightarrow 0$. Reference values are obtained by applying highly accurate numerical integration to the tangent-point energy on a smooth parameterized surface (triangulated in *top right*). See supplemental for additional examples.

separation parameter θ ; for $\theta = 0$, this gives the all-pairs energy $E(f_h)$ from Eq. 10. The resulting relative errors are shown in Fig. 10.

The discrete energy $E^P(f_h)$ uses face normals, which are known to exhibit $O(h)$ error, where $h > 0$ is the longest edge length. So it is expected that the discretization error $e_h := |E^P(f_h) - \mathcal{E}^P(f)|$ is no better than $O(h)$. Surprisingly, our experiments show a numerical rate that is considerably better (Fig. 10, *bottom left* and for $\theta = 0$). Moreover, we use center of mass data on BVH nodes; so the deviation $e_{h,\theta} := |\tilde{E}^P(f_h) - E^P(f_h)|$ of the Barnes-Hut approximation from the discrete energy should be dominated by the midpoint rule’s consistency error which is $O(\theta^2)$. Indeed our experiments seem to confirm this (see Fig. 10, *bottom right*).

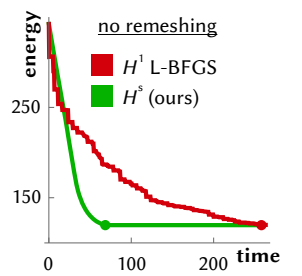
8.2 Comparisons

We next compare to other accelerated descent strategies from geometric computing. Overall, our fractional Sobolev scheme converges to local minimizers much faster than past schemes (dramatically so, in the case of highly knotted configurations). This outcome should not come as a surprise: the all-pairs energy we seek to minimize behaves very differently from those arising in, e.g., curvature flows or elasticity, which are based on discrete differential operators with small local stencils.

8.2.1 Comparison Methods. Our comparisons are guided by the extensive comparisons from Yu et al. [2021, Section 7]; we compare with the best of those methods. As a baseline we consider ordinary L^2 gradient descent, which replaces A in Eq. 14 with the mass matrix. Likewise, replacing A with the weak Laplacian Δ (encoded by the cotan matrix) yields standard H^1 Sobolev preconditioning; H^2

Sobolev preconditioning is achieved by solving Eq. 14 with the weak bi-Laplacian Δ^2 in place of A . (This latter preconditioner is essentially an ideal choice for Willmore flow [Schumacher 2017].) Like H^1 preconditioning, the *accelerated quadratic proxy (AQP)* method uses the weak Laplacian Δ as the inner product, but also computes a Nesterov acceleration step from the previous two configurations; this strategy is compatible only with linear constraints [Kovalsky et al. 2016, Section 2]. Another common strategy, which we refer to as H^1 L-BFGS, is to initialize L-BFGS with the weak Laplacian rather than the identity matrix, and likewise use the Laplacian to evaluate inner products. Finally, *Blended cured quasi-Newton (BCQN)* essentially interpolates between ordinary H^1 Sobolev preconditioning and H^1 L-BFGS, and uses barrier penalties to prevent triangle inversion. Since our flow is almost H^s orthogonal with tangential motions (and since we remesh) we do not experience inversions, and hence omit inversion barriers/inversion-aware line search.

To make a fair comparison, all methods use identical code for energy/differential evaluations (Sec. 4) and remeshing (Sec. 6). Since edge splits and collapses invalidate the history of BFGS-based methods, we reset memory vectors whenever such operations occur. However, even without remeshing H^1 L-BFGS is not competitive—e.g., it takes about 4x as



long as our method just to reach the state in Fig. 7, *center* (see inset). All comparisons use barycenter and total area constraints. Since AQP and L-BFGS do not support nonlinear constraints, we instead use a penalty function for area (Sec. 7.3.1). Augmenting AQP with a Schur complement would simply H^1 projected gradient descent minus Nesterov acceleration—which empirically does not significantly improve performance for our problem.

See Yu et al. [2021] and its supplement for further discussion and comparison of other popular schemes (such as implicit time stepping) with our fractional Sobolev approach.

8.3 Time Step Restriction

Fig. 5 verifies that matching the order of the inner product to that of the energy differential essentially lifts the mesh-dependent time step restriction. Here, we sampled the same surface at three resolutions, and ran each method for the same number of iterations. Our H^s scheme makes more progress for an equal number of iterations—but more importantly, the per-iteration progress of H^s is largely unaffected by mesh resolution, whereas all other methods slow down as resolution increases. Hence, even if some of these methods could be further accelerated by a constant factor (e.g., via code-level optimization), asymptotic behavior would ultimately dominate.

8.4 Wall-Clock Performance

We timed real-world performance of each method on several challenge meshes, using an AMD Ryzen Threadripper 3990X with 32 GB of RAM. Though our solver benefits from multiple threads (see Sec. 1), we ran this benchmark single-threaded to ensure fair comparison. Fig. 12 plots energy as a function of time; we ran each method

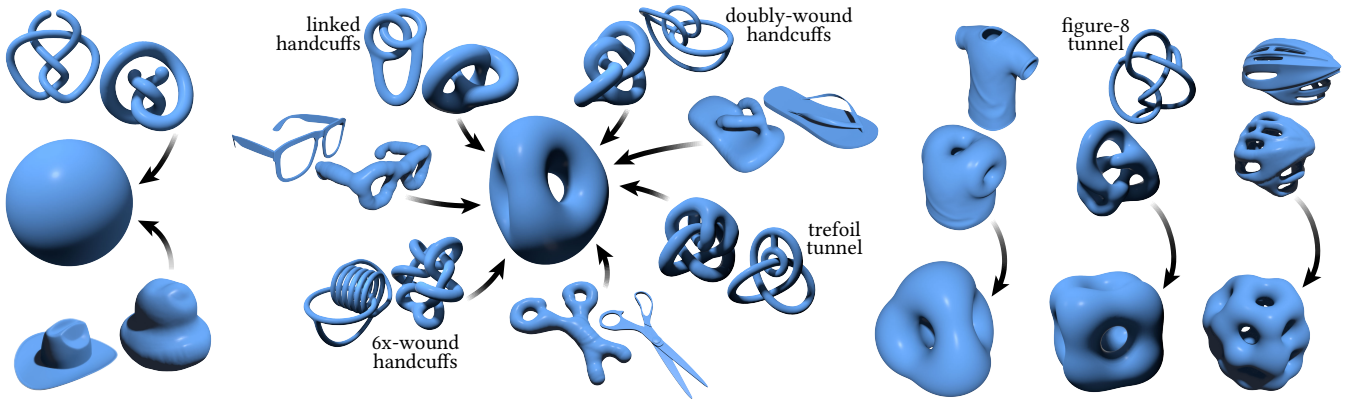


Fig. 11. Gallery of isotopies obtained by minimizing tangent-point energy—notice that highly knotted surfaces, as well as surfaces with thin sheets and handles, successfully flow to their canonical embeddings. Surfaces are grouped by their isotopy equivalence classes, which are extremely difficult to determine via visual inspection (and also not simply determined by Euler characteristic—see Fig. 16). Labeled meshes are used for performance comparisons in Fig. 12.

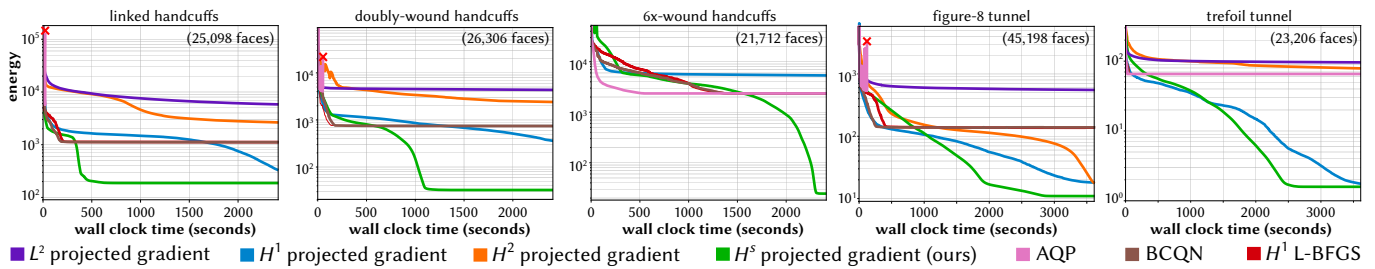


Fig. 12. Energy plots showing the effectiveness of a suite of methods at minimizing the tangent-point energy. Our H^s method (in green) reaches minimizers more quickly and consistently than the alternatives. Iterations where energy blows up (above 10^{10}) causing the AQP method to fail are marked with an X. Renderings of the meshes used and their minimizers can be seen in Fig. 11.

for 3600 seconds on the figure-8 and trefoil tunnels, and 2400 seconds on all others. Reference energy values were computed by evaluating the exact energy, without Barnes-Hut. Our H^s projected gradient method gave the best performance in all cases, reliably reaching a minimum within the allotted time. In some cases the initial decrease

is faster for other methods, likely because there are initially many small local features to be smoothed out. Subsequently, however, these methods make much slower progress at evolving the global shape. Though AQP and BQN are also based on H^1 preconditioning, they do not do as well here as the “vanilla” H^1 preconditioner. One possible reason is that they do not support hard nonlinear constraints—hence penalties may fight with the main objective. See Yu et al. [2021] for a much more extensive discussion and analysis of fractional methods versus a similar set of alternatives.

Fig. 13 gives timings for a variety of models using four threads—note the roughly linear scaling with mesh size. Here, the larger meshes are in a sense over-tessellated: for interactive design exploration one can typically start with a much coarser mesh (say, 10k faces or less), then refine and optimize the final geometry if desired.

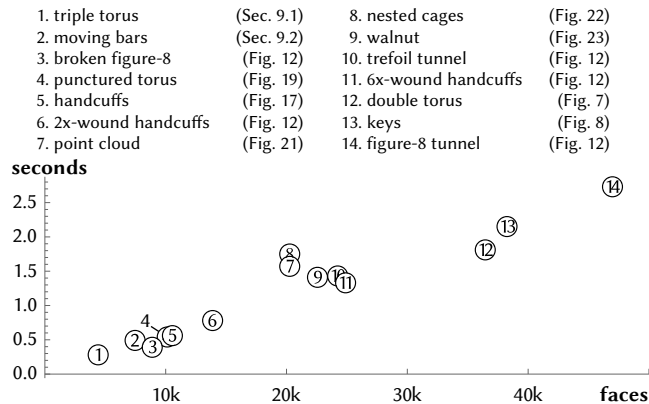


Fig. 13. Per-step timings for a representative set of examples—notice scaling is roughly linear with mesh size.

9 EXAMPLES AND APPLICATIONS

We here explore a variety of applications that help to further evaluate our method, show how it can be used in context, and also identify issues that might be improved in future work. These applications are also illustrated in the accompanying video—note that for many of these examples we take time steps far smaller than the optimal step determined by line search, in order to produce smooth animation.

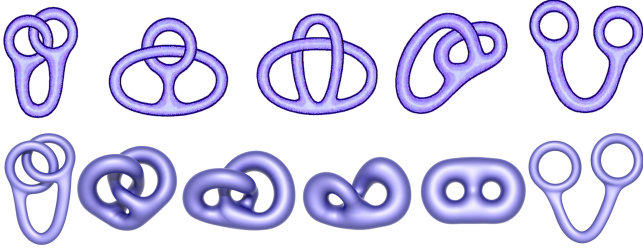


Fig. 14. *Top*: even careful illustrations of topological phenomena (here drawn by mathematician Peter Lynch) can be difficult to understand without a good visual imagination. *Bottom*: our method automatically generates continuous motions that are easier to interpret (see video), enabling exploration by students and researchers who do not have significant artistic training.

9.1 Mathematical Visualization and Exploration

A basic question in geometric topology is whether two embedded manifolds are the same up to *ambient isotopy*—intuitively, whether one surface can be deformed into another by smoothly warping the space around it. More formally, an ambient isotopy between two embeddings $f_0, f_1 : M \rightarrow \mathbb{R}^3$ is a continuous map $F : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ such that for all $x \in M$, $F(x, 0) = x$, $F(f_0(x), 1) = f_1(x)$, and $F(x, t)$ is a homeomorphism from \mathbb{R}^3 to \mathbb{R}^3 for every time $0 \leq t \leq 1$.

In general this equivalence question can be quite hard to answer—for instance, even detecting whether a given embedding of the circle in \mathbb{R}^3 is equivalent to the unit circle (or “unknot”) has not yet admitted a polynomial time algorithm [Lackenby 2016]. Hence, computational tools are used to explore such questions experimentally; a notable example for curves is *KnotPlot* [Scharein 1998]. The software developed for our project effectively provides the first “KnotPlot for surfaces.” Especially the fact that our solver exhibits rapid convergence and excellent scaling enables us to investigate questions that would be impossible with naïve numerical methods.

9.1.1 Canonical Embeddings. Global minimizers of geometric energies provide the “simplest” possible geometric representative of a given topological space. Such minimizers also play a critical role in geometric algorithms since they provide a canonical domain for, e.g., surface correspondence and data transfer—see for instance recent algorithms in both the intrinsic [Schmidt et al. 2020; Gillespie et al. 2021] and extrinsic [Kazhdan et al. 2012; Ye et al. 2018] settings. Formally proving that a given surface is a global minimizer is quite challenging. For instance, even the classic *Willmore conjecture* (which says that the *Clifford torus* minimizes Willmore energy for genus-1 surfaces) was resolved only very recently, after about 50 years of sustained effort [Marques and Neves 2014]. Hence, numerical tools are essential for formulating hypotheses about the behavior of minimizers and other critical points. To date, there are no clear conjectures about tangent-point minimizers for surfaces of genus $g \geq 2$. For reasons discussed in Sec. 1, these minimizers likely exhibit symmetries in \mathbb{R}^3 rather than \mathbb{S}^3 , making them potentially useful as a base domain for algorithms in extrinsic shape processing. To do so, one would simply need to track the parametric correspondence (e.g., via UV-coordinates), and perhaps minimize tangential distortion after flowing to a geometric minimizer (*à la* Schmidt et al. [2020]).

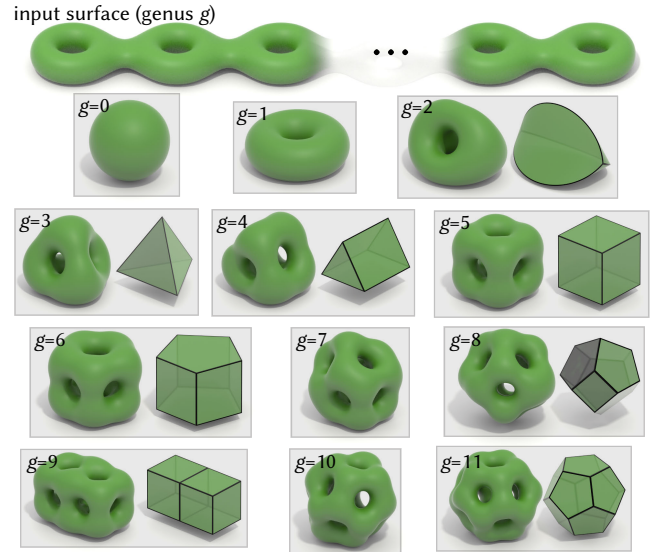


Fig. 15. Global minimizers of geometric energies provide canonical domains that can be used to map between surfaces of the same topology, or simply help visualize a topological space. Here we show conjectured minimizers of the tangent-point energy for unknotted surfaces of genus g ; adjacent figures illustrate symmetries (when present).

Unknotted Minimizers. Fig. 15 shows a numerical study of unknotted surfaces of increasing genus g , initialized by a linear arrangement of handles. For $g = 0, 1$, and 2 we get a round sphere, a torus of revolution, and a surface with symmetries of a triangular prism, respectively. Other surfaces appear to exhibit symmetries of a highly regular polyhedron—for instance, for $g = 3, 4, 5, 6, 8, 9$, and 11 we get symmetries of the tetrahedron, triangular prism, cube, pentagonal prism, truncated bipyramid, rectangular prism, and dodecahedron, respectively. Symmetries (if any) for $g = 7$ and 10 are less clear—or we may have simply failed to reach a global minimum. Interestingly, an octahedral configuration appears not to be a minimizer for genus 7 , even if we start with a symmetric configuration (and similarly for the icosahedron, not shown). In general it seems triangular “faces” are not preferred in higher-genus configurations due to the small angle between “edges”—much as electron repulsion maximizes bond angles in molecular geometries (e.g., stable compounds like graphite prefer bond angles near 120° , whereas only unstable compounds like white phosphorus exhibit tetrahedral symmetry).

Knotted Minimizers. A key feature of tangent-point energy (versus, say, Willmore energy) is that it enables us to find minimizers within a given isotopy class. Hence, just as it is quite common to make tables of canonical knot embeddings, we can now make tables of canonical embeddings for knotted surfaces. For instance, Fig. 16 shows the first-ever visualization of the different ways a genus-2 surface can be embedded in space. In the past, these isotopy classes have been depicted only as trivalent graphs—we take each such



Fig. 16. Geometric functionals provide a bridge between topology and geometry by enabling one to construct canonical geometric representatives of a given topological space. Here, minimizers of tangent point energy are used to visualize nontrivial isotopy classes of a genus-2 surface. (Numbers indicate number of crossings; subscripts index trivalent graphs from [Ishii et al. 2012, Table 1]).

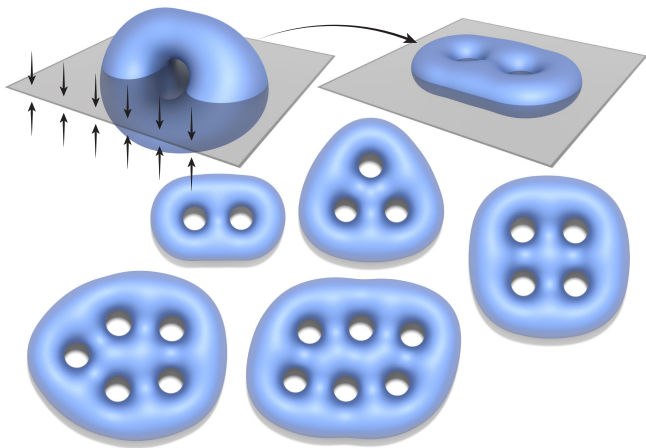


Fig. 17. *Top*: minimizers of tangent-point energy often exhibit three-dimensional symmetries which can be difficult to understand from a single view—by adding an attracting plane, we get embeddings that can be nicely displayed in two-dimensional illustrations. *Bottom*: constrained minimizers for genus 2 through 6.

graph from [Ishii et al. 2012, Table 1], and construct a topologically equivalent initial mesh that is optimized by our approach (see inset). As with knots most of these minimizers do not exhibit much extrinsic symmetry, except for, e.g., 6_7 and 5_3 which exhibit bilateral and 3-fold symmetry, *resp.*

Planar Representatives. Although minimizers exhibit a high degree of symmetry in \mathbb{R}^3 , it can be hard to determine even the genus

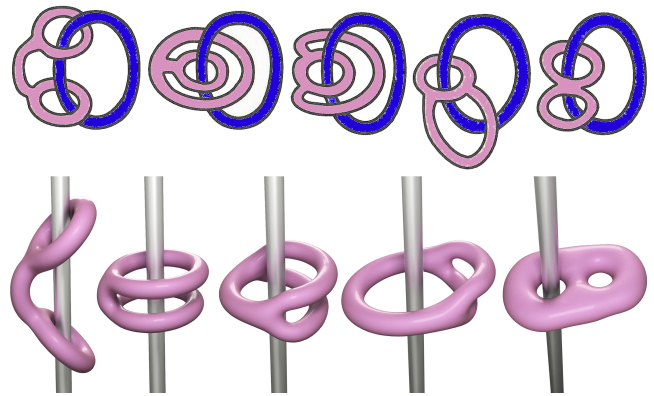
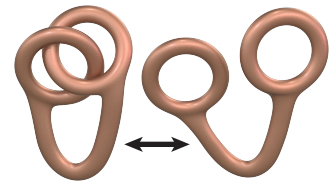


Fig. 18. Surprisingly, one can remove a handle of a double torus from a loop or pole without cutting or pinching the surface. *Top*: hand-drawn illustration by Wells [1997]. *Bottom*: isotopy computed automatically by our method (see video); no keyframing or boundary conditions were used.

of a minimizer when viewed from just a single viewpoint. In contrast, topological figures depicted by expert illustrators tend to be somewhat “2.5-dimensional” so that they can be better understood when projected onto the image plane. We can replicate this behavior by adding a simple attractive plane potential, as depicted in Fig. 17, yielding minimizers that are much easier to recognize (contrast with Fig. 15). An additional plane constraint yields a linear arrangement of handles, as commonly drawn by hand (see inset).

9.1.2 *Illustrating Isotopies.*

Our method also provides significant utility for mathematical visualization and illustration. Traditionally, interesting homotopies and isotopies are depicted by a sequence of drawings (or perhaps physical models) highlighting key moments of transition—a practice that has developed over time into a true art form [Francis and Francis 1987]. However, even the best drawings can be difficult to understand without significant thought and visual imagination. To obtain continuous motions (that are more easily understood), a small number of carefully “hand-crafted” computer animations have been produced over the years by either artist keyframing, or explicit programming of meticulously derived parametric formulas [Levy and Thurston 1995; Bednorz and Bednorz 2019]. More recently, automatic optimization-based tools have been used to produce animations, such as the *minimax sphere eversion* [Francis et al. 1997], as well as recent work in computer graphics on metric embedding [Chern et al. 2018] and conformally-constrained Willmore surfaces [Soliman et al. 2021]. Since these optimization-based tools are largely automatic, they help to democratize the creation of topological animations—our scheme extends such tools to the important and difficult case of ambient isotopies.



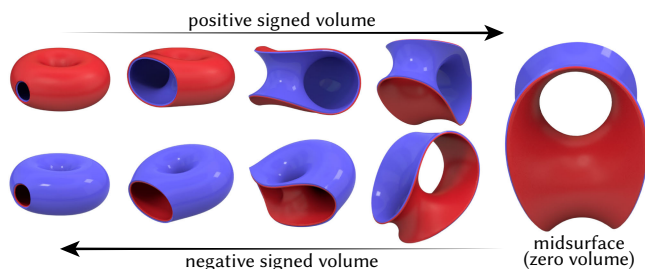


Fig. 20. Minimizing the tangent-point energy of a punctured torus while pushing signed volume toward zero yields a surface with reflection symmetry. Applying a reflection and reversing the flow hence yields an eversion that turns the surface “inside-out” while avoiding self-intersections.

One classic example is “unlinking” a pair of handcuffs (as shown in the inset above), though mathematically speaking these handcuffs are not actually linked: surprisingly, they belong to the same isotopy class. Fig. 14 compares a hand drawing of this isotopy with a different isotopy automatically computed via our method—and which is much better depicted in the accompanying video. To create this animation we simply minimize tangent-point energy from both start and end configurations, together with a potential that encourages the surface to lay parallel to the view plane. Since we reach the same minimizer in both cases (seen in Fig. 1, *far right*), we can compose these two sequences (one in reverse) to depict the complete motion. Other similar examples are shown in Fig. 11, and in the video.

Fig. 18 shows another classic example: removing one handle of a pair of handcuffs from a rigid pole or ring. The hand-drawn illustration helps to indicate several stages of this isotopy, which are also captured in our animation. However, the remarkable fact about our version is that it is driven *purely* by energy minimization—we did not perform any keyframing, nor impose any boundary conditions, yet it still constructs an isotopy in several “stages”: flatten out the two handles, perform a so-called *IH-move* (see Fig. 19 and [Ishii 2008]), and then optimize the geometry of the untangled surface.

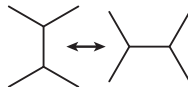


Fig. 19. An IH-move.

Our specific setup here is to minimize tangent-point energy while fixing surface area, and incorporating an infinite repulsive cylinder (modeled by an implicit surface). As in the previous example we use an attractive plane orthogonal to the pole to obtain a more canonical-looking minimizer. The only hand-tuning was reducing the repulsive strength of the cylinder near the end of the animation, to give the handles of the final surface a similar size. Importantly, allowing the barycenter to float freely (*à la* Sec. 7.1.1) is essential here, since the center of mass must ultimately move away from the pole.

Punctured Torus Eversion. Our discrete tangent-point energy can also be evaluated on surfaces with boundary, since we simply take a sum over pairs of triangles. Since we did not develop a careful treatment of boundary conditions, we simply penalize the total length and total squared curvature to ensure the boundary at least remains regular. In Fig. 20 we use this setup to compute an isotopic

eversion between the two orientations of a punctured torus. Unlike the classical sphere eversion, where one typically starts with a symmetric midsurface and flows toward the round sphere, we start with the punctured torus and use our flow to find the midsurface. The key observation is that the *oriented* volume of the surface $\int_M \langle f(x), N_f(x) \rangle dx_f$ will be zero for a symmetric configuration; fixing the area ensures that our zero-volume penalty does not cause the surface to collapse to a point. Once we reach zero volume we transform the midsurface by a reflection and 90-degree rotation, and run the same flow in reverse (with opposite colors) to obtain the eversion.

9.2 Geometry Processing and Shape Modeling

The no-collision condition is also natural in geometry processing and shape modeling, especially when a surface is meant to represent the boundary of a solid object (*e.g.*, for computational fabrication). Tangent-point energy can be used to augment existing geometric modeling and processing tasks with short- and long-range collision avoidance—we here explore several aspirational examples. In general, there has been relatively little work on collision-aware geometric modeling. For instance, Harmon et al. [2011] and Fang et al. [2021] explore modeling systems based on local collision penalties [Li et al. 2020], similar in spirit to classic *log barrier methods* [Boyd et al. 2004, Section 11.2]. However, since these penalties do not discretize a smooth energy, their behavior will depend on surface tessellation; moreover, descent strategies developed for these methods are tailored to local collisions, rather than long-range interactions and global untangling (*à la* Sec. 9.1). Other methods explicitly mesh the free space between objects [Müller et al. 2015; Jiang et al. 2017], which is costly for 3D problems, and cumbersome for problems with large deformations.

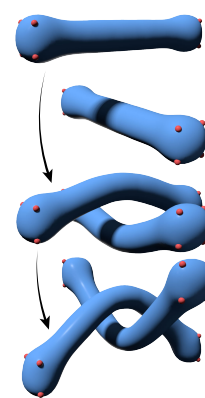


Fig. 21. Modeling with long-range influence.

Proximity-Aware Variational Modeling. As a basic example, Fig. 21 shows a simple example of interactive surface editing, where surface geometry is guided by point constraints, and nearby geometry is moved out of the way by the tangent-point energy. To better preserve the details of an initial mesh one might also combine tangent-point energy with a discrete shell energy [Grinspun et al. 2003], which would entail transferring the material configuration across meshing operations (a question which is beyond the scope of this work). Fig. 22 shows another example where pinned points and edges are interpolated while optimizing the rest of the geometry. (Here we disable remeshing, but could easily modify remeshing to ignore pinned vertices). Unlike harmonic interpolation or area minimization, for which point constraints are ill-posed, we get nice curvature behavior even near the pins; unlike Willmore flow (which provides good curvature behavior), we avoid self-intersection. Tangent-point energy could also in principle be used as a regularizer to discourage collision in other common modeling paradigms, such as *as rigid as possible* (ARAP) modeling [Sorkine and Alexa 2007].

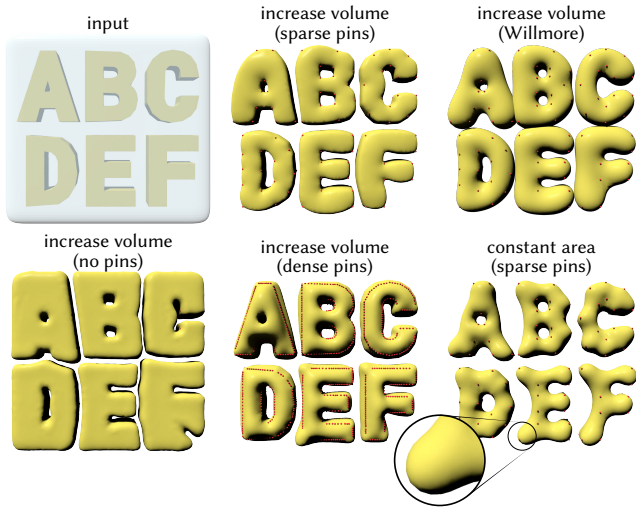


Fig. 22. The tangent-point energy can be used to make variational surface modeling responsive to proximity, rather than just collisions. Here for instance we pin a sparse or dense set of points and modify volume and surface area to adjust the appearance of some text (in some cases enclosed in a box). Like Willmore energy (top right), we get smooth behavior near point constraints (see magnified portion), but avoid overlap.

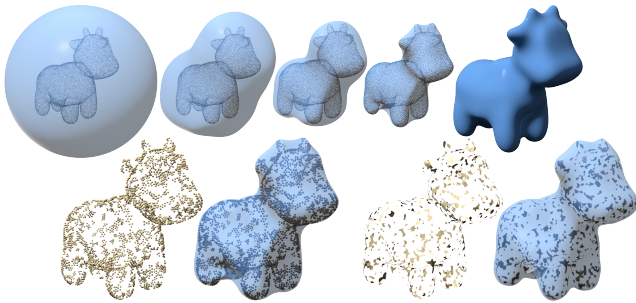


Fig. 23. Here we perform a simple “shrink wrapping” to obtain a manifold, intersection-free reconstruction (top), which works well even for points or polygon soup with severe holes and missing data.

9.2.1 Shrink Wrapping. One class of methods for reconstructing a surface from a collection of points is to “shrink-wrap” them with a triangle mesh [Kobbelt et al. 1999; Hanocka et al. 2020]; such methods are especially suitable in problems where one wishes to fit a high-quality template mesh to a known class of shapes (e.g., head or body scans). A basic problem, however, is that the mesh can get “tangled” during wrapping, inhibiting progress or requiring intricate remeshing to resolve self-intersections. Tangent-point energy may prove useful as a regularizer for such methods—Fig. 23 shows a basic shrink wrapping example on a point cloud, and on a polygon soup with severe holes. Here we minimize tangent-point energy with a gradually decreasing volume constraint.

9.2.2 Nested Envelopes. In a similar vein, nested sequences of solids $U_1 \subset \dots \subset U_k \subset \mathbb{R}^n$ represented by progressively coarser meshes have applications in multiresolution solvers, cage-based editing,

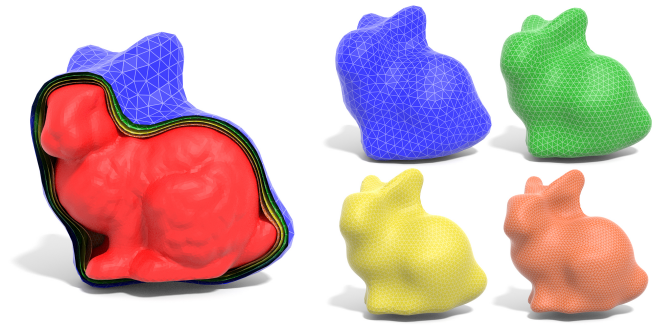


Fig. 24. We can also “shrink wrap” a model to get a sequence of progressively coarser approximating envelopes that exhibit a strict containment property, and are free of self-intersection. Here we aim for a 1.15x increase in volume at each level.

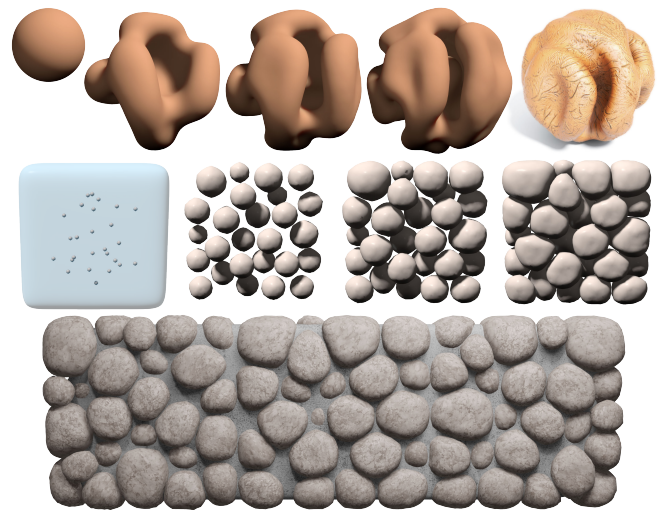


Fig. 25. We can also use tangent-point energy for generative modeling by “growing” a surface subject to constraints. Top: confining to a sphere while increasing area leads to a wrinkled shape reminiscent of a walnut. Bottom: growing many small spheres inside a slab yields a tileable cobblestone pattern.

and physical simulation [Sacht et al. 2015]. In Fig. 24 we construct each surface ∂U_k by minimizing tangent point energy plus a volume constraint, and gradually adjusting the constrained volume to achieve a fixed constant factor (here, 1.15x) of the volume of ∂U_{k-1} . This variational approach may offer interesting generalizations of ordinary nested cages, since it can easily incorporate constraints and objectives beyond just collision avoidance.

9.2.3 Generative Modeling. Rather than using the tangent-point energy to edit or process existing data, we can also use it to generate new geometry. In nature, the growth of organic shapes is often governed by simple combinations of objectives, e.g., a balance between area and volume while avoiding self-collision. We can likewise use such forces to drive the growth of organic-looking objects, such as the “walnut” depicted in Fig. 25, top. The same technique is used in

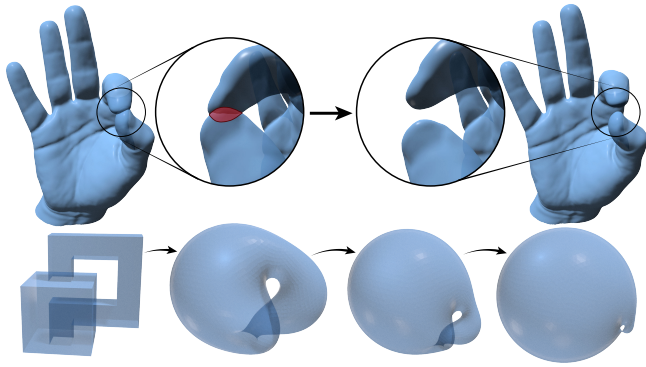


Fig. 26. For exponents $p < 4$, the tangent-point energy \mathcal{E}^p is no longer infinite for self-intersecting surfaces, but still discourages overlap. Here we try using this “subcritical” energy to resolve intersections, which works for small intersections (*top*), but fails for an unembeddable surface like the Klein bottle (*bottom*).

Fig. 25, *bottom*, where multiple objects are packed into a volume to create a repeating organic pattern.

9.2.4 Resolving Intersections. In many geometry processing tasks, input data is not free of self-intersections. For exponents $p > 4$, the tangent-point energy \mathcal{E}^p of a non-embedded surface is infinite; to *resolve* intersections in the input, we can try reducing the exponent to a value $p < 4$, at which point \mathcal{E}^p becomes finite but still discourages collision. Here we find that it also helps to disable the low-order term from Eq. 8. Empirically, the same system framework now appears capable of eliminating small self-intersections (Fig. 26, *top*), though struggles in more difficult scenarios like the Klein bottle depicted in Fig. 26, *bottom*, which cannot be globally embedded without self-intersection. Further analysis of the energy for these “subcritical” values may help to provide more robust tools for global collision resolution.

10 LIMITATIONS AND FUTURE WORK

The experiments from Sec. 9 suggest many opportunities for improvement. For instance, major performance gains could be achieved

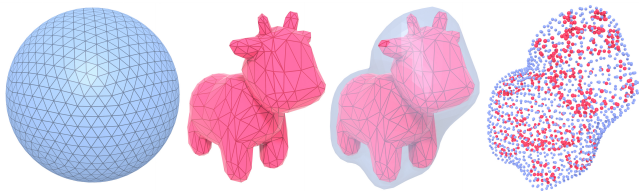


Fig. 27. When using a coarse mesh for our surface (*far left*) and/or obstacles (*center left*) small intersections can occur (*center right*), since our energy discretization approximates each surface by a collection of quadrature points (*far right*). Such artifacts could easily be avoided by, *e.g.*, checking for collisions during line search.

purely through better software engineering, *e.g.*, better parallel implementation of hierarchical matrix multiplication (which is currently bottlenecked around 4–8 threads), or curvature-adapted remeshing (*à la* [Dunyach et al. 2013]) to reduce mesh size. It would also be useful to track attributes across remeshing operations, to enable (for instance) mapping of data from one shape to another through the canonical minimizer. Since we discretize the tangent-point energy, we provide no hard guarantee that collisions will not occur (Fig. 27). One pragmatic solution is to simply perform continuous collision detection to limit the time step, as done for many years in computer animation [Moore and Wilhelms 1988] and knot drawing [Scharein 1998] (and more recently in geometry processing [Smith and Schaefer 2015; Jiang et al. 2017; Li et al. 2020]). Better might be to approximate tangent-point energy via a hard upper bound on each element pair, akin to the *MD energy* in Scharein [1998, Section 3.5.1], or element-element penalties from Li et al. [2020].

Several issues require deeper investigation. For one, unlike Yu et al. [2021], our preconditioner cannot accommodate dense constraints (*e.g.*, preservation of each triangle area), which would require $O(|F|)$ iterative solves. Here one can instead use a stiff penalty; revisiting the multigrid approach via hierarchical coarsening [Botsch and Kobbelt 2004; Shi et al. 2006] may also prove fruitful. Our energy approximation becomes inaccurate in situations of very tight contact (*à la* Sections 9.2.1 and 9.2.2), since we have few quadrature points per unit surface area; adding additional quadrature points (or adaptive refinement) in regions of near-contact may yield tighter fits. A related issue is that some amount of bending regularization is inherent in the energy itself, making it hard to approximate, *e.g.*, sharp edges.

For shape interpolation and mathematical visualization, it would be quite useful to find trajectories that minimize total tangent-point energy (over time), rather than just flowing to a common minimizer—here work on shell-space geodesics may prove valuable [Heeren et al. 2012]. Likewise, integrating repulsive regularization into a thin shell model may help retain a “memory” of the initial shape. Finally, we do not treat Neumann boundary conditions, nor more general arrangements of repulsive curves *and* surfaces that may have interesting modeling applications.

ACKNOWLEDGMENTS

The authors thank Saul Schleimer and Henry Segerman for helpful discussions about topological examples, and Nick Stadie for perspective on molecular symmetries. This work was supported by a Packard Fellowship, NSF Award 1943123, and gifts from Autodesk, Activision Blizzard, Adobe, Disney, and Facebook. The third author was supported by DFG-Project 282535003: *Geometric curvature functionals: energy landscape and discrete methods*.

REFERENCES

- J. Barnes and P. Hut. 1986. A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature* 324, 6096 (1986), 446–449.
- Adam Bednorz and Witold Bednorz. 2019. Analytic sphere eversion using ruled surfaces. *Differential Geometry and its Applications* 64 (2019), 59–79.
- S. Blatt. 2013. The Energy Spaces of the Tangent Point Energies. *Journal of Topology and Analysis* 5, 3 (2013), 261–270.
- Simon Blatt and Philipp Reiter. 2015. Regularity theory for tangent-point energies: the non-degenerate sub-critical case. *Adv. Calc. Var.* 8, 2 (2015), 93–116.

- Alexander I. Bobenko and Peter Schröder. 2005. Discrete Willmore Flow. In *Proceedings of the Third Eurographics Symposium on Geometry Processing* (Vienna, Austria) (SGP '05). Eurographics Association, Goslar, DEU, 101–es.
- Mario Botsch and Leif Kobbelt. 2004. A Remeshing Approach to Multiresolution Modeling. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (Nice, France) (SGP '04). Association for Computing Machinery, New York, NY, USA, 185–192.
- Stephen Boyd, Stephen P Boyd, and Lieven Vandenbergh. 2004. *Convex optimization*. Cambridge university press.
- Robert Bridson, Ronald Fedkiw, and John Anderson. 2002. Robust treatment of collisions, contact and friction for cloth animation. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 594–603.
- G. Buck and J. Orloff. 1995. A simple energy function for knots. *Top. Appl.* 61, 3 (1995).
- Dorin Bucur and Giuseppe Butazzo. 2006. VARIATIONAL METHODS IN SHAPE OPTIMIZATION PROBLEMS.
- Long Chen and Michael Holst. 2011. Efficient mesh optimization schemes based on Optimal Delaunay Triangulations. *Computer Methods in Applied Mechanics and Engineering* 200, 9 (2011), 967 – 984.
- Albert Chern, Felix Knöppel, Ulrich Pinkall, and Peter Schröder. 2018. Shape from metric. *ACM Transactions on Graphics* (TOG) 37, 4 (2018), 1–17.
- Sebastian Clatici, Mikhail Bessmeltsev, Scott Schaefer, and Justin Solomon. 2017. Isometry-aware preconditioning for mesh parameterization. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 37–47.
- Ulrich Clarenz, Udo Diewald, Gerhard Dziuk, Martin Rumpf, and R Rusu. 2004. A finite element method for surface restoration with smooth boundary conditions. *Computer Aided Geometric Design* 21, 5 (2004), 427–445.
- Keenan Crane, Fernando de Goes, Mathieu Desbrun, and Peter Schröder. 2013a. Digital Geometry Processing with Discrete Exterior Calculus. In *ACM SIGGRAPH 2013 courses* (Anaheim, California) (SIGGRAPH '13). ACM, New York, NY, USA, 126.
- Keenan Crane, Ulrich Pinkall, and Peter Schröder. 2013b. Robust Fairing via Conformal Curvature Flow. *ACM Trans. Graph.* 32, 4 (2013).
- Mathieu Desbrun, Mark Meyer, Peter Schröder, and Alan H Barr. 1999. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 317–324.
- Marc Droske and Martin Rumpf. 2004. A level set formulation for Willmore flow. *Interfaces and free boundaries* 6, 3 (2004), 361–378.
- Marion Duniach, David Vanderhaeghe, Loïc Barthe, and Mario Botsch. 2013. Adaptive remeshing for real-time mesh deformation. In *Eurographics 2013*. The Eurographics Association.
- Gerhard Dziuk. 2008. Computational parametric Willmore flow. *Numer. Math.* 111, 1 (2008), 55–80.
- Ilya Eckstein, Jean-Philippe Pons, Yiyang Tong, C.-C. Jay Kuo, and Mathieu Desbrun. 2007. Generalized Surface Flows for Mesh Processing. In *Proceedings of the Fifth Eurographics Symposium on Geometry Processing* (Barcelona, Spain) (SGP '07). Eurographics Association, 183–192.
- Matthew Elsey and Selim Esedoglu. 2009. Analogue of the total variation denoising model in the context of geometry processing. *Multiscale Modeling & Simulation* 7, 4 (2009), 1549–1573.
- Yu Fang, Minchen Li, Chenfanfu Jiang, and Danny M. Kaufman. 2021. Guaranteed Globally Injective 3D Deformation Processing. *ACM Trans. Graph.* (SIGGRAPH) 40, 4, Article 75 (2021).
- George Francis, John M Sullivan, Rob B Kusner, Ken A Brakke, Chris Hartman, and Glenn Chappell. 1997. The minimax sphere eversion. In *Visualization and mathematics*. Springer, 3–20.
- George K Francis and GK Francis. 1987. *A topological picturebook*. Vol. 2. Springer.
- M. Freedman, Z. He, and Z. Wang. 1994. Mobius Energy of Knots and Unknots. *Annals of Mathematics* 139, 1 (1994), 1–50.
- Mark Gillespie, Boris Springborn, and Keenan Crane. 2021. Discrete Conformal Equivalence of Polyhedral Surfaces. *ACM Trans. Graph.* 40, 4 (2021).
- E. Giusti. 1973. *Minimal Surfaces with Obstacles*. Springer Berlin Heidelberg, Berlin, Heidelberg, 121–153.
- Eitan Grinspun, Anil N Hirani, Mathieu Desbrun, and Peter Schröder. 2003. Discrete shells. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Citeseer, 62–67.
- W. Hackbusch. 2015. *Hierarchical matrices: algorithms and analysis*. Vol. 49. Springer.
- Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. 2020. Point2Mesh: A Self-Prior for Deformable Meshes. *ACM Trans. Graph.* 39, 4, Article 126 (July 2020), 12 pages.
- David Harmon, Daniele Panozzo, Olga Sorkine, and Denis Zorin. 2011. Interference-aware geometric modeling. *ACM Transactions on Graphics* (TOG) 30, 6 (2011), 1–10.
- David Harmon, Etienne Vouga, Breannan Smith, Rasmus Tamstorf, and Eitan Grinspun. 2009. Asynchronous contact mechanics. In *ACM SIGGRAPH 2009 papers*. 1–12.
- Behrend Heeren, Martin Rumpf, Max Wardetzky, and Benedikt Wirth. 2012. Time-discrete geodesics in the space of shells. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1755–1764.
- Atsushi Ishii. 2008. Moves and invariants for knotted handlebodies. *Algebraic & Geometric Topology* 8, 3 (2008), 1403–1418.
- Atsushi Ishii, Kengo Kishimoto, Hiromasa Moriuchi, and Masaaki Suzuki. 2012. A table of genus two handlebody-knots up to six crossings. *Journal of Knot Theory and Its Ramifications* 21, 04 (2012), 1250035.
- Zhongshi Jiang, Scott Schaefer, and Daniele Panozzo. 2017. Simplicial complex augmentation framework for bijective maps. *ACM Transactions on Graphics* 36, 6 (2017).
- Pushkar Joshi and Carlo Séquin. 2007. Energy minimizers for curvature-based surface functionals. *Computer-Aided Design and Applications* 4, 5 (2007), 607–617.
- Michael Kazhdan, Jake Solomon, and Mirela Ben-Chen. 2012. Can mean-curvature flow be modified to be non-singular?. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1745–1754.
- Leif P. Kobbelt, Jens Vorsatz, and Ulf Labsik. 1999. A Shrink Wrapping Approach to Remeshing Polygonal Surfaces. *Computer Graphics Forum* 18, 3 (1999), 119–130.
- Slawomir Kolasiński, Paweł Strzelecki, and Heiko von der Mosel. 2015. Compactness and isotopy finiteness for submanifolds with uniformly bounded geometric curvature energies. arXiv:arXiv:1504.04538
- Shahar Z. Kovalsky, Meirav Galun, and Yaron Lipman. 2016. Accelerated Quadratic Proxy for Geometric Optimization. *ACM Trans. Graph.* 35, 4, Article 134 (July 2016), 11 pages.
- Robert B Kusner and John M Sullivan. 1998. Möbius-invariant knot energies. *Ideal Knots* 19 (1998), 315–352.
- Mateusz Kwaśnicki. 2017. Ten equivalent definitions of the fractional laplace operator. *Fractional Calculus and Applied Analysis* 20, 1 (Jan 2017).
- Marc Lackenby. 2016. Elementary knot theory. arXiv preprint arXiv:1604.03778 (2016).
- Silvio Levy and William P Thurston. 1995. *Making waves: A guide to the ideas behind Outside In*. Geometry Center.
- Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M. Kaufman. 2020. Incremental Potential Contact: Intersection- and Inversion-free Large Deformation Dynamics. *ACM Trans. Graph.* (SIGGRAPH) 39, 4, Article 49 (2020).
- Fernando C Marques and André Neves. 2014. Min-max theory and the Willmore conjecture. *Annals of mathematics* (2014), 683–782.
- Tobias Martin, Pushkar Joshi, Miklós Bergou, and Nathan Carr. 2013. Efficient Non-linear Optimization via Multi-scale Gradient Filtering. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 89–100.
- Matthew Moore and Jane Wilhelms. 1988. Collision detection and response for computer animation. In *Proceedings of the 15th annual conference on Computer graphics and interactive techniques*. 289–298.
- Henry P Moreton and Carlo H Séquin. 1992. Functional optimization for fair surface design. *ACM SIGGRAPH Computer Graphics* 26, 2 (1992), 167–176.
- Matthias Müller, Nuttapon Chentanez, Tae-Yong Kim, and Miles Macklin. 2015. Air meshes for robust collision handling. *ACM Transactions on Graphics* (TOG) 34, 4 (2015), 1–9.
- Jun O'Hara. 1991. Energy of a knot. *Topology* 30, 2 (1991), 241–247.
- Stanley Osher and Ronald Fedkiw. 2006. *Level set methods and dynamic implicit surfaces*. Vol. 153. Springer Science & Business Media.
- U. Pinkall and K. Polthier. 1993. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics* 2, 1 (1993), 15–36.
- Robert J Renka and JW Neuberger. 1995. Minimal surfaces and Sobolev gradients. *SIAM Journal on Scientific Computing* 16, 6 (1995), 1412–1427.
- Leonardo Sacht, Etienne Vouga, and Alec Jacobson. 2015. Nested cages. *ACM Transactions on Graphics* (TOG) 34, 6 (2015), 1–14.
- Robert Glenn Scharein. 1998. *Interactive topological drawing*. Ph.D. Dissertation. University of British Columbia.
- Patrick Schmidt, Marcel Campen, Janis Born, and Leif Kobbelt. 2020. Inter-surface maps via constant-curvature metrics. *ACM Transactions on Graphics* (TOG) 39, 4 (2020), 119–1.
- Henrik Schumacher. 2017. On H^2 -gradient Flows for the Willmore Energy. arXiv preprint arXiv:1703.06469 (2017).
- Lin Shi, Yizhou Yu, Nathan Bell, and Wei-Wen Feng. 2006. A Fast Multigrid Algorithm for Mesh Deformation. *ACM Trans. Graph.* 25, 3 (2006), 1108–1117.
- Jason Smith and Scott Schaefer. 2015. Bijective parameterization with free boundaries. *ACM Transactions on Graphics* (TOG) 34, 4 (2015), 1–9.
- Yousuf Soliman, Albert Chern, Olga Diamanti, Felix Knöppel, Ulrich Pinkall, and Peter Schröder. 2021. Constrained Willmore Surfaces. *ACM Trans. Graph.* 40, 4 (2021).
- Olga Sorkine and Marc Alexa. 2007. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, Vol. 4. 109–116.
- Paweł Strzelecki and Heiko von der Mosel. 2013. Tangent-point repulsive potentials for a class of non-smooth m -dimensional sets in \mathbb{R}^n . Part I: Smoothing and self-avoidance effects. *J. Geom. Anal.* 23, 3 (2013), 1085–1139.
- Paweł Strzelecki and Heiko von der Mosel. 2018. Geometric curvature energies: facts, trends, and open problems. In *New directions in geometric and applied knot theory*. De Gruyter, Berlin, 8–35.
- David Wells. 1997. *The Penguin dictionary of curious and interesting numbers*. Penguin.

- Peter Wriggers and Giorgio Zavarise. 2004. Computational contact mechanics. *Encyclopedia of computational mechanics* (2004).
- Zi Ye, Olga Diamanti, Chengcheng Tang, Leonidas Guibas, and Tim Hoffmann. 2018. A unified discrete framework for intrinsic and extrinsic Dirac operators for geometry processing. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 93–106.
- Chris Yu, Henrik Schumacher, and Keenan Crane. 2021. Repulsive Curves. *ACM Trans. Graph.* 40, 2, Article 10 (May 2021), 21 pages.
- Fuzhen Zhang. 2005. *The Schur Complement and its Applications*. Numerical Methods and Algorithms, Vol. 4. Springer, New York.
- Yufeng Zhu, Robert Bridson, and Danny M. Kaufman. 2018. Blended Cured Quasi-Newton for Distortion Optimization. *ACM Trans. Graph.* 37, 4, Article 40 (2018), 14 pages.

A ACTION OF THE FRACTIONAL OPERATORS

In section Sec. 5.1 we claimed that the actions of the fractional operators L^σ , B , and B_0 can be expressed by suitable kernel matrices that we then compress by hierarchical methods. This is not obvious, so we include a brief derivation here. Consider the kernel matrix

$$H_{ST} := (1 - \delta_{ST}) |X_f(S) - X_f(T)|^{-(2\sigma+2)}.$$

Rewriting Eq. 11 for general \mathbf{u} and $\mathbf{v} \in \mathbb{R}^{|V|}$ in terms of this kernel yields

$$\mathbf{u}^\top L^\sigma \mathbf{v} = \sum_{S \in F} \sum_{T \in F} (\bar{u}(S) - \bar{u}(T)) (\bar{v}(S) - \bar{v}(T)) a_f(S) H_{ST} a_f(T).$$

Multiplying the product inside the sum gives

$$(\bar{u}(S) \bar{v}(S) + \bar{u}(T) \bar{v}(T) - \bar{u}(T) \bar{v}(S) - \bar{u}(S) \bar{v}(T)) a_f(S) H_{ST} a_f(T)$$

for the pair (S, T) . Because $H_{ST} = H_{TS}$, we can move some terms between the summands for (S, T) and (T, S) , and thus reorganize the sum into

$$\begin{aligned} \mathbf{u}^\top L^\sigma \mathbf{v} &= 2 \sum_{S \in F} \sum_{T \in F} (\bar{u}(S) \bar{v}(S) - \bar{u}(S) \bar{v}(T)) a_f(S) H_{ST} a_f(T) \\ &= 2 \sum_{S \in F} \bar{u}(S) a_f(S) \left(a_f(S)^{-1} \sum_{T \in F} H_{ST} a_f(T) \right) a_f(S) \bar{v}(S) \\ &\quad - 2 \sum_{S \in F} \sum_{T \in F} \bar{u}(S) a_f(S) H_{ST} a_f(T) \bar{v}(T). \end{aligned}$$

Recall that $\mathbf{U} \in \text{Hom}(\mathbb{R}^{|V|}; \mathbb{R}^{|F|})$ is defined by $(\mathbf{U}\mathbf{u})(S) = a_f(S)\bar{u}(S)$. Thus the above collapses to

$$\begin{aligned} \mathbf{u}^\top L^\sigma \mathbf{v} &= 2 \mathbf{u}^\top \mathbf{U}^\top \text{diag}(a_f)^{-1} \text{diag}(H a_f) \mathbf{U} \mathbf{v} - 2 \mathbf{u}^\top \mathbf{U}^\top H \mathbf{U} \mathbf{v} \\ &= 2 \mathbf{u}^\top \mathbf{U}^\top \left[\text{diag}(\text{diag}(a_f)^{-1} H a_f) - H \right] \mathbf{U} \mathbf{v}. \end{aligned}$$

The derivation follows analogously for the high- and low-order matrices B and B_0 , with the substitution of the operator $\mathbf{V} = \text{diag}(a_f)D_f$ for \mathbf{U} in the case of B .

B FAST MATRIX-VECTOR MULTIPLICATION

Step 1 of Sec. 5.1.5 corresponds to thinning out the matrix shown in Fig. 6 by removing all the green parts. The remainder is a sparse block matrix with variable block size. We store this sparse matrix in CSR format and evaluate products using sparse BLAS routines.

In Step 2 the kernel matrix $H_{\mathcal{I}\mathcal{J}}$ is compressed into the rank-one-matrix $\mathbf{1}_{\mathcal{I}} h(X_{\mathcal{I}}, P_{\mathcal{I}}; X_{\mathcal{J}}, P_{\mathcal{J}}) \mathbf{1}_{\mathcal{J}}^\top$. In this step, we are cautious not to move the input data $\mathbf{x}_{\mathcal{J}}$ and output data $\mathbf{y}_{\mathcal{I}}$ directly to and from the clusters \mathcal{I} and \mathcal{J} . Instead, as is common for multipole-type methods, we store data on BVH clusters. For each cluster \mathcal{I}, \mathcal{J} , we allocate scalars $\tilde{\mathbf{x}}_{\mathcal{I}}$ and $\tilde{\mathbf{y}}_{\mathcal{I}}$. Starting with leaf clusters, we set

$$\tilde{\mathbf{x}}_{\mathcal{I}} \leftarrow \sum_{T \in \mathcal{I}} \mathbf{x}(T) \quad \text{for each leaf cluster } \mathcal{I}.$$

Then, during a parallel traversal of the BVH in post-order, for each cluster \mathcal{J} , we add the $\tilde{\mathbf{x}}$ -values of its children into $\tilde{\mathbf{x}}_{\mathcal{J}}$. After this *upward pass* is finished, we loop over all clusters \mathcal{I} and set

$$\tilde{\mathbf{y}}_{\mathcal{I}} \leftarrow \sum_{\mathcal{J}} h(X_{\mathcal{I}}, P_{\mathcal{I}}; X_{\mathcal{J}}, P_{\mathcal{J}}) \tilde{\mathbf{x}}_{\mathcal{J}}, \quad (22)$$

with the sum over all \mathcal{J} such that $(\mathcal{I}, \mathcal{J})$ is admissible. This can be evaluated via sparse matrix multiplication: first, we fix an ordering of the BVH clusters, e.g., depth-first ordering. Then we assemble a sparse matrix \tilde{H} with the nonzero value $h(X_{\mathcal{I}}, P_{\mathcal{I}}; X_{\mathcal{J}}, P_{\mathcal{J}})$ at the position that correspond to the admissible block cluster $(\mathcal{I}, \mathcal{J})$. Storing $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ as vectors, Eq. 22 amounts to

$$\tilde{\mathbf{y}} \leftarrow \tilde{H} \tilde{\mathbf{x}}.$$

Afterwards, we use a *downward pass* through the BVH to distribute the $\tilde{\mathbf{y}}$ -values back into the vector \mathbf{y} : We traverse the BVH in pre-order and let each cluster \mathcal{I} add its $\tilde{\mathbf{y}}$ -value into each of its children's $\tilde{\mathbf{y}}$ -values. Finally each leaf cluster adds its value into each of its member's \mathbf{y} -entry, i.e.,

$$\mathbf{y}(S) \leftarrow \mathbf{y}(S) + \tilde{\mathbf{y}}_{\mathcal{I}} \quad \text{for each leaf } \mathcal{I} \text{ and each } S \in \mathcal{I}.$$

The structure of the kernel matrices of L^σ , B , and B_0 is very similar. This allows us to use a single block cluster tree to compress all of them. Moreover, the sparsity patterns for the two sparse matrices used to perform Steps 1 and 2 can be shared and the corresponding nonzero values can be computed in a single parallelized loop over the admissible and inadmissible blocks, *resp.*

For the application of A_3 to a vector \mathbf{v} of size $3|V|$, we could apply $A = B + B_0$ separately on three vectors $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 of size $|V|$ that each store only one spatial component of the vertex positions. However, it turns out to be more efficient to store $\mathbf{v}_1, \mathbf{v}_2$, and \mathbf{v}_3 as columns of a matrix of size $|V| \times 3$ and to replace the sparse matrix-vector products by sparse matrix-dense matrix products.