

Blended Cured Quasi-Newton for Distortion Optimization

YUFENG ZHU, University of British Columbia & Adobe Research

ROBERT BRIDSON, Autodesk & University of British Columbia

DANNY M. KAUFMAN, Adobe Research

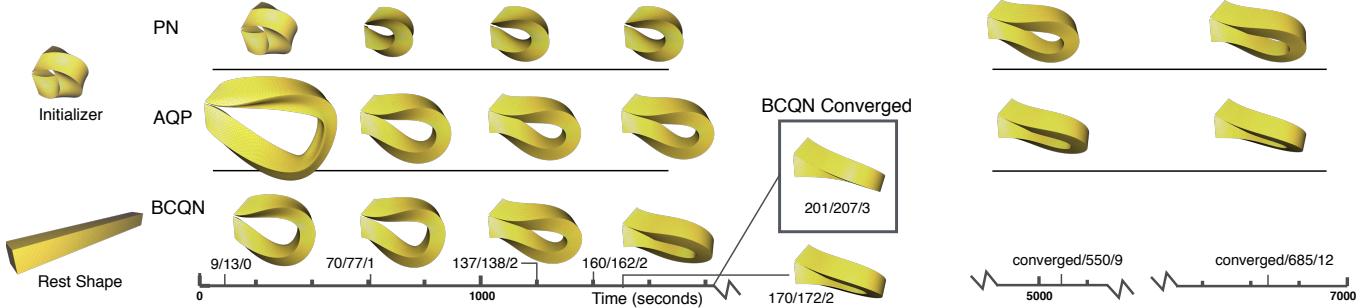


Fig. 1. **Twisting.** A stress-test 3D deformation problem. **Left:** we initialize a 1.5M element tetrahedral mesh bar with a straight rest shape into a tightly twisted coil, constraining both ends to stay fixed. **Right:** minimizing the ISO deformation energy to find a constrained equilibrium with (top to bottom) Projected Newton (PN), Accelerated Quadratic Proxy (AQP) and our Blended Cured quasi-Newton (BCQN) method, we show intermediate shapes at reported wall-clock time (seconds) and iteration counts at those times (BCQN/AQP/PN). AQP, much slower than BCQN, requires many more iterations to converge while PN, despite requiring fewer iterations, is well over 25X slower due to per-iteration costs dominated by factorization.

Optimizing distortion energies over a mesh, in two or three dimensions, is a common and critical problem in physical simulation and geometry processing. We present three new improvements to the state of the art: a barrier-aware line-search filter that cures blocked descent steps due to element barrier terms and so enables rapid progress; an energy proxy model that adaptively blends the Sobolev (inverse-Laplacian-processed) gradient and L-BFGS descent to gain the advantages of both, while avoiding L-BFGS's current limitations in distortion optimization tasks; and a characteristic gradient norm providing a robust and largely mesh- and energy-independent convergence criterion that avoids wrongful termination when algorithms temporarily slow their progress. Together these improvements form the basis for Blended Cured Quasi-Newton (BCQN), a new distortion optimization algorithm. Over a wide range of problems over all scales we show that BCQN is generally the fastest and most robust method available, making some previously intractable problems practical while offering up to an order of magnitude improvement in others.

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

Additional Key Words and Phrases: numerical optimization, geometry optimization, distortion, deformation, elasticity, simulation, preconditioning, fast solvers

Authors' addresses: Yufeng Zhu, University of British Columbia & Adobe Research; Robert Bridson, Autodesk & University of British Columbia ; Danny M. Kaufman, Adobe Research.

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 ACM 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.

© 2018 Association for Computing Machinery.

0730-0301/2018/8-ART40 \$15.00

<https://doi.org/10.1145/3197517.3201359>

ACM Reference Format:

Yufeng Zhu, Robert Bridson, and Danny M. Kaufman. 2018. Blended Cured Quasi-Newton for Distortion Optimization. *ACM Trans. Graph.* 37, 4, Article 40 (August 2018), 14 pages. <https://doi.org/10.1145/3197517.3201359>

1 INTRODUCTION

Many fundamental *physical* and *geometric* modeling tasks reduce to minimizing nonlinear measures of distortion over meshes. Simulating elastic bodies, parametrization, deformation, shape interpolation, deformable inverse kinematics, and animation all require robust, efficient, and easily automated *distortion optimization*. By *robust* we mean the algorithm should solve every reasonable problem to any accuracy given commensurate time, and only report success when the accuracy has truly been achieved. By *efficient* we mean rapid convergence in wall-clock time, even if that may mean more (but cheaper) iterations. By *automated* we mean the user needn't adjust algorithm parameters or tolerances at all to get good results when going between different problems. With these three attributes, a distortion optimization algorithm can be reliably used in production software.

We propose a new algorithm, Blended Cured Quasi-Newton (BCQN), with three core contributions based on analysis of where prior methods faced difficulties:

- an **adaptively blended quadratic energy proxy** for distortion energies that iteratively combines the Sobolev gradient and a quasi-Newton secant approximation; this allows low cost per iterate with second-order acceleration while avoiding secant artifacts where the Laplacian is more robust;
- a **barrier-aware filter** on search directions, that gains larger step sizes and so improved convergence progress in line search for the common case of iterates where individual elements degenerate towards collapse; and

- a **characteristic gradient norm convergence criterion**, immune to terminating prematurely due to algorithm stagnation, that is consistent across mesh sizes, scales, and choice of energy so that per-problem adjustment is unnecessary.

Over a wide range of test cases we show that BCQN achieves our goals for production software: BCQN makes the solution of some previously intractable problems practical, offers up to an order of magnitude speed-up in other cases and, in all cases investigated so far, robustly converges while improving on or closely matching the performance of the best-in-class optimizers available.

2 PROBLEM STATEMENT AND OVERVIEW

Distortion optimization seeks local minimizers

$$x^* = \underset{x \in \mathbb{R}^{dn}}{\operatorname{argmin}} E(x), \quad (1)$$

for n vertex locations in d -dimensional space stored in vector x , where the energy $E(x)$ is a measure of distortion, and x is subject to constraints.¹ The energy is expressed as a sum over elements t in a triangulation T (triangles or tetrahedra depending on dimension),

$$E(x) = \sum_{t \in T} v_t W(F_t(x)), \quad (2)$$

where $v_t > 0$ is the area or volume of the rest shape of element t , W is an energy density function taking the deformation gradient as its argument, and F_t computes the deformation gradient for element t . This problem may be given as is, or may be the result of a discretization of a continuum problem with, for example, linear finite elements.

2.1 Iterative solvers for nonlinear minimization

Solution methods for distortion optimization generally apply an algorithmic strategy of iterated approximation and stepping [Bertsekas 2016], built from three primary ingredients: an energy approximation, a line search, and a termination criteria.²

2.1.1 Energy Approximation. At the current iterate x_i we form a local quadratic approximation of the energy, or *proxy*:

$$E_i(x) = E(x_i) + (x - x_i)^T \nabla E(x_i) + \frac{1}{2}(x - x_i)^T H_i(x - x_i) \quad (3)$$

where H_i is a symmetric matrix. Near the solution, if H_i accurately approximates the Hessian we can achieve fast convergence optimizing this proxy, but it is also critical that it be stable – symmetric positive definite (SPD) – to ensure the proxy optimization is well-posed everywhere; we also want H_i to be cheap to solve with, preferring sparser matrices and ideally not having to refactor at each iteration.

2.1.2 Line Search. Quadratic models allow us to apply linear solvers to find stationary points $x_i^* = \underset{x}{\operatorname{argmin}} E_i(x)$ of the local energy approximation. A step

$$p_i = x_i^* - x_i = -H_i^{-1} \nabla E(x_i) \quad (4)$$

¹For simplicity of presentation we restrict our attention here to constraining a subset of vertex positions to given values, i.e. Dirichlet boundary conditions.

²Alternatively, trust-region methods are available, though not considered in the current work nor as popular within the field.

towards this stationary point then forms a direction for probable energy descent. However, quadratic models are only locally accurate for nonlinear energies in general, thus line-search is used to find an improved length $\alpha_i > 0$ along p_i to update to a new iterate

$$x_{i+1} \leftarrow x_i + \alpha_i p_i, \quad (5)$$

that provides adequate decrease in nonlinear energy E . Of particular concern for the distortion problems we face are energies that blow up to infinity for degenerate (flattened) elements: in a given step, the elements where this may come close to happening rapidly depart from the proxy, and the step size α_i may have to be very small indeed, see Figure 2, impeding progress globally.

2.1.3 Termination. Iteration continues until we are able to stop with a “good enough” solution – but this requires a precise computational definition. Typically we monitor some quantity that approaches zero if and *only if* the iterates are approaching a stationary point. The standard in unconstrained optimization is to check the norm of the gradient of the energy, which is zero only at a stationary point and otherwise positive; however, the raw gradient norm depends on the mesh size, scaling, and choice of energy, which makes finding an appropriate tolerance to compare against highly problem-dependent and difficult to automate.

3 RELATED WORK

3.1 Energies and applications

A wide range of physical simulation and geometry processing computations are cast as *variational* tasks to minimize measures of distortion over domains.

To simulate elastic solids with large deformations we typically minimize hyper-elastic potentials formed by integrating strain-energy densities over the body. These material models date back to Mooney [1940] and Rivlin [1948]. Their Mooney-Rivlin and Neo-Hookean materials, and many subsequent hyperelastic materials, e.g. St. Venant-Kirchoff, Ogden, Fung [Bonet and Burton 1998], are constructed from empirical observation and analysis of deforming real-world materials. Unfortunately, all but a few such energy densities are nonconvex w.r.t. strain.³ This makes their minimization highly challenging. Constants in these models are determined by experiment for scientific computing applications [Ogden 1972], or alternately are directly set by users in other cases [Xu et al. 2015], e.g., to meet artistic needs.

In geometry processing a diverse range of energies have been proposed to minimize various mapping distortions, generally focused on minimizing either measures of isometric [Aigerman et al. 2015; Chao et al. 2010; Liu et al. 2008; Smith and Schaefer 2015; Sorkine and Alexa 2007] or conformal [Ben-chen et al. 2008; Desbrun et al. 2002; Hormann and Greiner 2000; Lévy et al. 2002; Mullen et al. 2008; Weber et al. 2012] distortion. While some of these energies do not prohibit inversion [Chao et al. 2010; Desbrun et al. 2002; Lévy et al. 2002; Sorkine and Alexa 2007] many others have been explicitly constructed

³We observe this is distinct from the global energy being nonconvex w.r.t. vertex positions: ARAP [2007] is an example of a convex energy density that nonetheless can lead to a nonconvex global energy.

with nonconvex terms that guarantee preservation of local injectivity [Aigerman et al. 2015; Hormann and Greiner 2000; Smith and Schaefer 2015]. Other authors have also added constraints to strictly bound distortion [Bommes et al. 2013; Lipman 2012] for example, but we restrict attention to unconstrained minimization — and note constrained optimization often relies on unconstrained algorithms as an inner kernel.

Our goal here is to provide a tool to minimize arbitrary energy density functions as-is. We take as input energy functions provided by the user, irrespective of whether these energies are custom-constructed for geometry tasks, physical energies extracted from experiment, or energies hand-crafted by artists. Our work focuses on the better optimization of the important *nonconvex* energies whose minimization remains the primary challenging bottleneck in many modern geometry and simulation pipelines. In the following sections, to evaluate and compare algorithms, we consider a range of challenging nonconvex deformation energies currently critical in physical simulation and geometry processing: Mooney-Rivlin (MR) [Bower 2009], Neo-Hookean (NH) [Bower 2009], Symmetric Dirichlet (ISO) [Smith and Schaefer 2015], Conformal Distortion (CONF) [Aigerman et al. 2015], and Most-Isometric Parameterizations (MIPS) [Hormann and Greiner 2000].

3.2 Energy approximations

Broadly, existing models for the local energy approximation in (3) fall into four rough categories that vary in the construction of the *proxy*⁴ matrix H_i . *Newton-type* methods exploit expensive second-order derivative information; *first-order* methods use only first derivatives and apply lightweight fixed proxies; *quasi-Newton* methods iteratively update proxies to approximate second derivatives using just differences in gradients; *Geometric Approximation* methods use more domain knowledge to directly construct proxies which relate to key aspects of the energy, resembling Newton-type methods but not necessarily taking second derivatives.

Newton-type methods generally can achieve the most rapid convergence but require the costly assembly, factorization and backsolve of new linear systems per step. At each iterate Newton’s method uses the energy Hessian, $\nabla^2 E(x_i)$, to form a proxy matrix. This works well for convex energies but requires modification for nonconvex energies [Nocedal and Wright 2006] to ensure that the proxy is at least positive semi-definite (PSD). Composite Majorization (CM), a tight convex majorizer, was recently proposed as an analytic PSD approximation of the Hessian [Shtengel et al. 2017]. The CM proxy is efficient to assemble but is limited to two-dimensional problems and just a trio of energies: ISO, NH and symmetric ARAP. More general-purpose solutions include adding small multiples of the identity, and projection of the Hessian to the PSD cone but these generally damp convergence too much [Liu et al. 2017; Nocedal and Wright 2006; Shtengel et al. 2017]. More effective is the Projected Newton (PN) method that projects per-element Hessians to the PSD cone prior to assembly [Teran et al. 2005]; also of interest is Chen and Weber’s analytic Hessian projection in a reduced basis setting [2017]. Both CM and PN generally converge rapidly in the nonconvex setting with CM often outperforming PN in

⁴Names and notations for H_i vary across the literature depending on method and application. For consistency, here, across all methods we will refer to H_i as the *proxy* matrix — inclusive of cases where it is the actual Hessian or direct modification thereof.

the subset of 2D cases where CM can be applied [Shtengel et al. 2017], while PN is more general purpose for 3D and 2D problems. Both PN and CM have identical per-element stencils and so identical proxy structures. Despite low iteration counts they both scale prohibitively due to per-iteration cost and storage as we attempt increasingly large optimization problems.

First-order methods build descent steps by preconditioning the gradient with a fixed proxy matrix. These proxies are generally inexpensive to solve and sparse so that cost and storage remain tractable as we scale to larger systems. However, they often suffer from slower convergence as we lack higher-order information. Direct gradient descent, $H_i \leftarrow Id$, and Jacobi-preconditioned gradient descent, $H_i \leftarrow \text{diag}(\nabla^2 E(x_i))$ offer attractive opportunities for parallelization [Fu et al. 2015; Wang and Yang 2016] but suffer from especially slow convergence due to poor scaling. The Laplacian matrix, L , forms an excellent preconditioner, that both smooths and rescales the gradient [Kovalsky et al. 2016; Martin et al. 2013; Neuberger 1985]. Unlike the Hessian, the Laplacian is a constant PSD proxy that can be pre-factorized once and backsolved separately per-coordinate. Iterating descent with $H_i \leftarrow L$, is the Sobolev-preconditioned gradient descent (SGD) method. SGD was first introduced, to our knowledge, by Neuberger [1985], but has since been rediscovered in graphics as the local-global method for minimizing ARAP [Sorkine and Alexa 2007]. As noted by Kovalsky et al. [2016] local-global for ARAP is exactly SGD. More recently Kovalsky et al. [2016] developed the highly effective Accelerated Quadratic Proxy (AQP) method by adding a Nesterov-like acceleration [Nesterov 1983] step to SGD. This improves AQP’s convergence over SGD. However, as this acceleration is applied after line search, steps do not guarantee energy decrease and can even contain collapsed or inverted elements — preventing further progress; see e.g. Figure 4. More generally, the Laplacian is constant and so ignores valuable local curvature information — we see this issue in a number of examples in Section 8 where AQP stagnates and is prohibitively slow to converge. Curvature can make the critical difference to enable progress.

Quasi-Newton methods lie in between these two extremes. They successively, per descent iterate, update approximations of the system Hessian using a variety of strategies. Quasi-Newton methods employing sequential gradients to update proxies, i.e. L-BFGS and variants, have traditionally been highly successful in scaling up to large systems [Bertsekas 2016]. Their updates can be performed in a compute and memory efficient manner and can guarantee the proxy is SPD even where the exact Hessian is not. While not fully second-order, they achieve superlinear convergence, regaining much of the advantage of Newton-type methods. L-BFGS convergence can be improved with the choice of initializer. Initializing with the diagonal of the Hessian [Nocedal and Wright 2006], application-specific structure [Jiang et al. 2004] or even the Laplacian [Liu et al. 2017] can help. However, so far, for distortion optimization problems, L-BFGS has consistently and surprisingly failed to perform competitively *irrespective* of choice of initializer [Kovalsky et al. 2016; Rabinovich et al. 2017]. Nocedal and Wright point out that the secant approximation can implicitly create a *dense* proxy, unlike the sparse true Hessian, directly and incorrectly coupling distant vertices. This is visible as swelling artifacts for intermediate iterations of the L-BFGS methods in Figure 5.

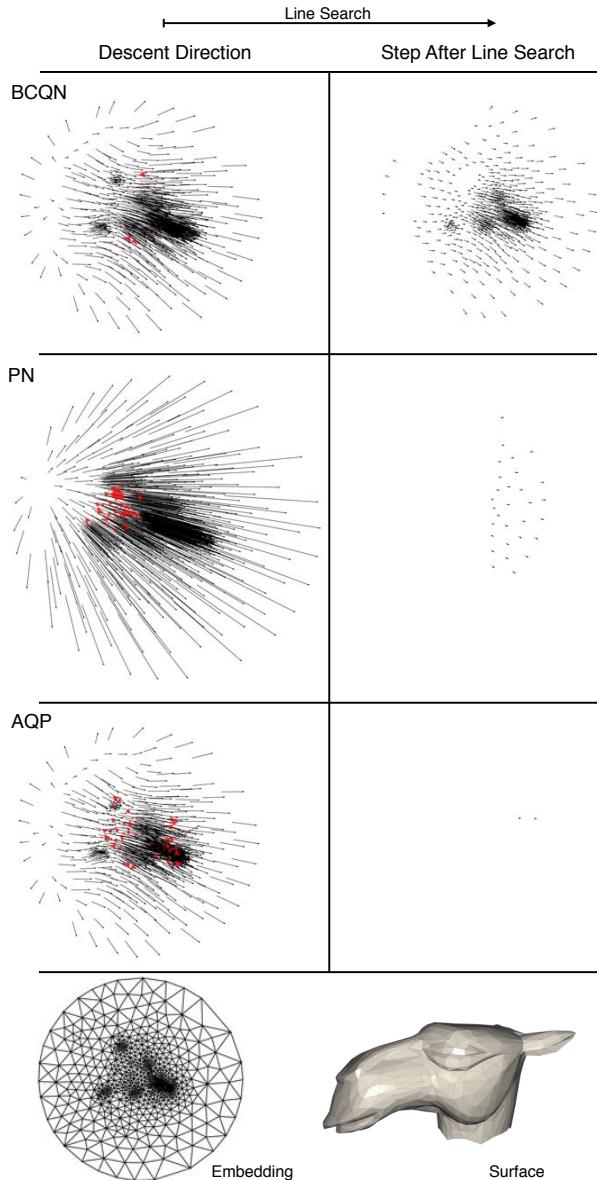


Fig. 2. Line-search filtering. Barrier terms in nonconvex energies (here we use ISO) of the form $1/g(\sigma)$ can severely restrict step sizes in line searches even when expensive, high-quality methods such as Newton-type methods are applied. **Left column:** descent-direction vector fields, per vertex, in a descent step generated by BCQN, PN and AQP with potential blocking triangles rendered in red. **Right column, bottom rows:** after line-search, close to collapsing elements have restricted the global step size for AQP and PN that block progress. **Right column, top row:** BCQN’s barrier-aware line-search filtering cures the descent direction enabling progress with significant descent.

Geometric Approximation methods for distortion optimization have also been developed recently: SLIM [Rabinovich et al. 2017] and the AKAP preconditioner [Claici et al. 2017]. SLIM extends the local-global strategy to a wide range of distortion energies while AKAP applies an approximate Killing vector field operator as the proxy matrix. Both

require re-assembly and factorization of their proxies for each iterate. SLIM and AKAP convergence are generally well improved over SGD and AQP [Claici et al. 2017; Rabinovich et al. 2017]. However, they do not match the convergence quality of the second-order, Newton-type methods, CM and PN [Shtengel et al. 2017]. SLIM falls well short of both CM and PN [Shtengel et al. 2017]. AKAP is more competitive than SLIM but remains generally slower to converge than PN in our testing, and is much slower than CM. At the same time SLIM and AKAP stencils, and so their fill-in, match CM’s and PN’s; see Figure 8. SLIM and AKAP thus require the same per-iteration compute cost and storage for linear solutions as PN and CM without the same degree of convergence benefit [Shtengel et al. 2017].

In summary, for smaller systems Newton-type methods have been, till now, our likely best choice for distortion optimization, while as we scale we have inevitably needed to move to first-order methods to remain tractable, while accepting reduced convergence rates and even the possibility of nonconvergence altogether. We develop a new quasi-Newton method, BCQN, that adaptively blends gradient information with the matrix Laplacian at each iterate to regain improved and robust convergence with efficient per-iterate storage and computation across scales while avoiding the current pitfalls of L-BFGS methods.

3.3 Line search

Once we have applied the effort to compute a search direction we would like to maximize its effectiveness by taking as large a step along it as possible. Because the energies we treat are nonlinear, too large a step size will actually make things worse by accidentally increasing energy. A wide range of line-search methods are thus employed that search along the step direction for *sufficient* decrease [Nocedal and Wright 2006]. However, when we seek to minimize non-convex energies on meshes the situation is more challenging. Most (although not all) popular and important nonlinear energies, both in geometry processing and physics, are composed by the sum of rational fractions of singular values, σ , of the deformation gradient $W(F) = W(\sigma) = f(\sigma)/g(\sigma)$ where the denominator $g(\sigma) \rightarrow 0$ as $\sigma_i \rightarrow 0, \forall i \in [1, d]$. Notice that these $1/g(\sigma)$ barrier functions block element inversion. Irrespective of their source, these blocking nonconvex energies rapidly increase energy along any search direction that would collapse elements. To prevent this (and likewise the possibility of getting stuck in an inverted state) search directions are capped to prevent inversion of every element in the mesh. This is codified by Smith and Schaeffer’s [2015] line-search filter, applied before traditional line search, that computes the maximal step size that guarantees no inversions anywhere.

Unfortunately, this has some serious consequences for progress. Notice that if even a single element is close to inversion this can amputate the full descent step so much that almost no progress can be made at all; see Figure 2, bottom and middle rows. This, in many senses seems unfair as we should expect to be able to make progress in other regions where elements may be both far from inversion and yet also far from optimality. To address these barrier issues we develop an efficient barrier-aware filter that allows us to avoid blocking contributions from individual elements close to collapse while still taking large steps elsewhere in the mesh, see Figure 2, top row.

3.4 Termination

Naturally we want to take as few iterates as possible while being sure that when we stop, we have arrived at an accurate solution according to some easily specified tolerance. The gold-standard in optimization is to iterate until the gradient is small $\|\nabla E\| < \epsilon$, for a specified tolerance $\epsilon > 0$. This is robust as ∇E is zero only at stationary points, and with a bound on Hessian conditioning near the solution can even provide an estimate on the distance of x to the solution.

However, an appropriate value of ϵ for a given application is highly dependent on the mesh, its dimensions, degree of refinement, energy, etc. A common engineering rule of thumb to deal with refinement consistency is to instead divide the L_2 -norm of ∇E by the number of mesh vertices. However, as we see in the inset figure, this normalization does not help significantly, for example here across changes in mesh resolution for the 2D swirl test; see Section 8.2 for more experiments.

To avoid problem dependence, recent distortion optimization codes generally either take a fixed (small) number of iterations [Rabinovich et al. 2017] or iterate until an absolute or relative error in energy $|E_{i+1} - E_i|$ and/or position $\|x_{i+1} - x_i\|$ are small [Kovalsky et al. 2016; Shtengel et al. 2017]. However, experiments underscore there is not yet any method which always converges satisfactorily in the same fixed number of iterations across varying boundary conditions, shape difficulty, mesh resolution, and choice of energy. Measuring the change in energy or position, absolutely or in relative terms, unfortunately cannot distinguish between an algorithm converging and simply stagnating in its progress far from the solution; again, there is not yet any method which can provably guarantee any degree of progress at every iterate before true convergence. Figure 4 illustrates on the swirl example how a reference AQP implementation declares convergence well before it reaches a satisfactory solution, when early on it hits a difficult configuration where it makes little local progress.

To provide reassuring termination criteria in practice and to enable fair comparisons of current and future distortion optimization problems we develop a gradient-based stopping criterion that remains consistent for optimization problems even as we vary scale, mesh resolution and energy type. This allows us, and future users, to set a default convergence tolerance in our solver once and leave it unchanged, independent of scale, mesh and energy. This likewise enables us to compare algorithms without the false positives given by non-converged algorithms that have halted due to lack of progress.

4 BLENDED QUASI-NEWTON

In this section we construct a new quadratic energy proxy which adaptively blends the Sobolev gradient with L-BFGS-style updates to capture curvature information, avoiding the troubles previous quasi-Newton methods have encountered in distortion optimization. Apart

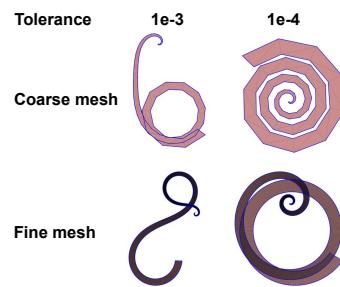


Fig. 3. Standard termination measures, e.g. the vertex-scaled gradient norm above, are inconsistent across mesh, energy and scale changes.

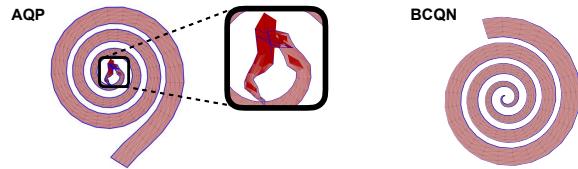


Fig. 4. In the 2D swirl example, BCQN with our reliable termination criterion (right) only stops once it has actually reached a satisfactory solution. The reference AQP implementation (left) erroneously declares success early on when it finds two iterates have barely changed, but this is due only to hitting a difficult configuration where AQP struggles to make progress.

from the aforementioned issue of a dense proxy incorrectly coupling distant vertices in L-BFGS, we also find that the gradients for non-convex energies with barriers can have highly disparate scales, causing further trouble for L-BFGS. The much smoother Sobolev gradient diffuses large entries from highly distorted elements to the neighborhood, giving a much better scaling. The Laplacian also provides essentially the correct structure for the proxy, only directly coupling neighboring elements in the mesh, and is well-behaved initially when far from the solution, thus we seek to stay close to the Sobolev gradient, as much as possible, while still capturing valuable curvature information from gradient history.

The standard (L-)BFGS approach exploits the secant approximation from the difference in successive gradients, $y_i = \nabla E(x_{i+1}) - \nabla E(x_i)$ compared to the difference in positions $s_i = x_{i+1} - x_i$,

$$\begin{aligned} \nabla^2 E(x_{i+1}) s_i &\approx y_i \\ \Rightarrow \quad \nabla^2 E(x_{i+1})^{-1} y_i &\approx s_i, \end{aligned} \quad (6)$$

to update an inverse proxy matrix $D_i = H_i^{-1}$ (approximating $\nabla^2 E^{-1}$ in some sense) so that $D_{i+1}y_i = s_i$. The BFGS quasi-Newton update is generically

$$QN_i(z, D) = V_i(z)^T D V_i(z) + \frac{s_i s_i^T}{s_i^T z}, \quad V_i(z) = I - \frac{z s_i^T}{s_i^T z}. \quad (7)$$

We can understand this as using a projection matrix V_i to annihilate the old D 's action on z , then adding a positive semi-definite symmetric rank-one matrix to enforce $QN_i(z, D)z = s_i$. Classic BFGS uses $D_{i+1} = QN_i(y_i, D_i)$, whereas L-BFGS uses

$$D_{i+1} = QN_i(y_i, \tilde{D}_i), \quad (8)$$

where \tilde{D}_i has the oldest QN update removed, and crucially represents each D as a product of linear operators, rather than an explicit full matrix. Only the last m $\{s, y\}$ vector pairs (we use $m = 5$) along with the initial D_1 (we use the inverse Laplacian, storing only its Cholesky factor) are stored; application of D is then just a few vector dot-products and updates along with backsolves for the Laplacian.

4.1 Greedy Laplacian blending

Experiments show that far from the solution, the Laplacian is often (although not always) a much more effective proxy than the L-BFGS secant version: see AQP/SGD vs. L-BFGS in Figure 5. In particular, the difference in energies y may introduce spurious coupling or have badly scaled entries near distorted triangles. In this case if the energy were based on the Laplacian itself (the *Dirichlet* energy), the difference

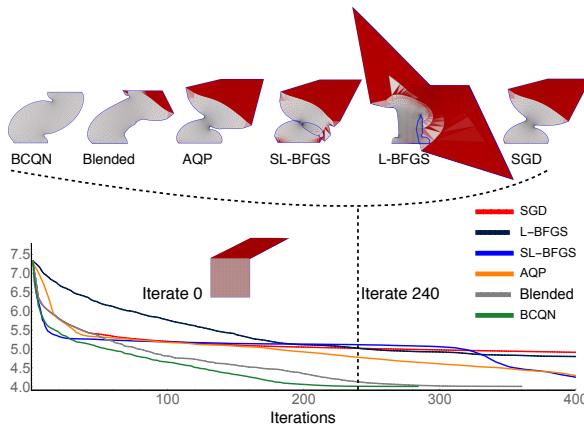


Fig. 5. A 2D shearing deformation stress test with MIPS energy, comparing methods by plotting iteration vs. energy. Both L-BFGS as well as Laplacian initialized L-BFGS (SL-BFGS) have slow convergence as previously reported [Kovalsky et al. 2016] — especially when compared to SGD and AQP which use just the Laplacian. At iteration 240 the visualized deformations show both L-BFGS-based methods suffering from swelling due to inaccurate coupling of distant elements. Applying our blending model alone (Blended) is highly effective, while our full BCQN method gives the best results overall.

in gradients would be the better behaved L_s . This motivates trying to update with L_s instead of y ,

$$D_{i+1} = QN_i(Ls_i, \tilde{D}_i). \quad (9)$$

This would keep us consistent with Sobolev preconditioning, which is often very effective in initial iterations. However, to achieve the superlinear convergence L-BFGS offers, near solutions we wish to come closer to satisfying the secant equation, and so aim to move towards using y instead. More generally we seek to balance between gradient information and the action of the Laplacian.

We thus propose an adaptive blending strategy that updates the proxy per iterate using

$$z_i = (1 - \beta_i)y_i + \beta_i Ls_i \quad (10)$$

in $QN(z_i, \tilde{D}_i)$, with a blending parameter $\beta_i \in [0, 1]$ constructed below. Towards our goals we first derive here a greedy blending that is “as Laplacian as possible” to *approximate* the energy gradient. Then, below in Section 4.2, we present our final, empirically derived variant that we have experimentally determined to so far work best. As we will soon see this blend is highly effective; however, further work remains to find and confirm optimal blending terms.

We begin by greedily selecting a β_i to scale Ls_i to be as close to y_i as possible,

$$\beta_i = \underset{\beta \in [0, 1]}{\operatorname{argmin}} \|y_i - \beta Ls_i\|^2, \quad (11)$$

in other words using the projection of y_i onto Ls_i . This comes as close as possible to satisfying the secant equation with Ls_i and then makes up the remainder with y_i . Solving (11) gives

$$\beta_i = \operatorname{proj}_{[0, 1]} \left(\frac{y_i^T Ls_i}{\|Ls_i\|^2} \right). \quad (12)$$

Observe that when Ls is roughly aligned with the gradient jump y , but y is as large or larger, β grows and Laplacian smoothing increases

— as we might hope for initially when far from the solution, where the Sobolev gradient is most effective. When the energy Hessian diverges strongly from the Laplacian approximation, perhaps when the cross-terms between coordinates missing from the scalar Laplacian are important, then β will decrease, so that contributions from y_i again grow. Finally, as the gradient magnitudes decreases close to the solution, β will similarly decay, ideally regaining the superlinear convergence of L-BFGS near local minima.

4.2 Blended quasi-Newton

With the blending projection (12) in place we experimented with a range of rescalings in hopes of further improving efficiency and robustness. After extensive testing we have so far found the following scaling to offer the best performance:

$$\begin{aligned} \beta_i &= \operatorname{proj}_{[0, 1]} \left(\frac{\operatorname{normest}(L) y_i^T L s_i}{B(T)} \right), \\ \text{with } B(T) &= \left(\sum_{t \in T} v_t \right)^{\frac{2(d-1)}{d}}. \end{aligned} \quad (13)$$

Here $\operatorname{normest}(L)$ is an efficient estimate of the matrix 2-norm using power iteration [Higham 1992], and $B(T)$ is a *constant* normalizing term with appropriate dimensions and so no longer has the same potential concern for sensitivity in the denominator when Ls is small but s is not. Both terms are computed just once before iterations begin and reused throughout.

As mentioned, we initialize the inverse proxy with $D_1 = L^{-1}$, thus starting with Laplacian preconditioning. With line search satisfying Wolfe conditions our proxy remains SPD across all steps [Nocedal and Wright 2006] and so delivers descent directions. Each step jointly updates D_i using the standard two-loop recursion and finds the next descent direction $s_i = -D_i \nabla E(x_i)$. Figure 5 illustrates the gains possible from blended quasi-Newton compared to both standard L-BFGS and Sobolev gradient algorithms, while then applying our barrier-aware filter, derived in the next section gives best results with our full BCQN algorithm.

5 BARRIER-AWARE LINE SEARCH FILTERING

As discussed in Section 3.3 and shown in Figure 2, barrier factors $1/g(\sigma)$ in nonconvex energies typically dominate step size in line search. Even a single element that is brought close to collapse by the descent direction, p_i , can restrict the line search step size severely. The computed step size α_i then scales p_i *globally* so that all elements, not just those that are going to collapse along p_i , are prevented from making progress. To avoid this, a natural strategy suggests itself: when the descent direction would cause elements to degenerate towards collapse along the full step, rather than simply truncating line search as in Smith and Schaefer [2015], we filter collapsing contributions from the search direction prior to line search. We call this strategy *barrier-aware line search filtering*.

5.1 Curing line search

Figure 6 illustrates how simple, direct filters that zero out contributions from nearly-inverted elements in either the search direction (6a) or the gradient before Laplacian smoothing (6b) fail. We must be able to make progress in nearly-inverted elements when the search direction

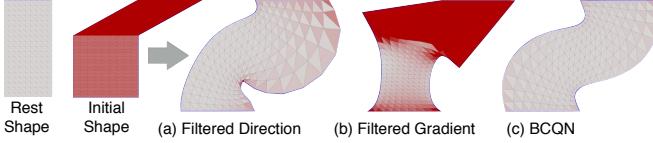


Fig. 6. Direct filtering does not work. Zeroing out inverting components of descent directions or gradients makes the search direction inconsistent with the objective and so prevents convergence, leading to termination at poor solutions (a) and (b). **Left:** we initialize a 2D shear deformation, constraining the top of a bar to slide rightwards. **Middle:** direct filtering of the descent direction (a) and the gradient (b) allow large descent steps forward unblocked from the contributions of close-to-collapsed elements. However, this results in termination at shapes that do not satisfy optimality of the original minimization. **Right:** compare to a locally optimal solution for this problem (c) obtained with BCQN.

can help, or there is no hope for reaching the actual solution; simple zeroing fails to converge, which is no surprise as it in essence is arbitrarily manipulating the target energy, changing the problem being solved. We instead seek to *augment* the original optimization problem in a way that does not change the solution, but gives us a tool to safely deal with problem elements so that the resulting search direction p_i does not cause them to collapse or invert, ideally with a small fixed cost per iteration.

5.2 One-sided barriers in distortion optimization

Element $t \in T$ is inverted at positions x precisely when the orientation function $a_t(x) = \det(F_t(x))$ is negative. Concatenating over T , the global vector-valued function for element orientations is then

$$a(\cdot) = (a_1(\cdot), \dots, a_m(\cdot))^T. \quad (14)$$

As long as $a(x) > 0$, no element is collapsed or inverted, and the energy remains finite. Note, however, many energies are also finite for inverted elements $a_t(x) < 0$, only blowing up at collapse $a_t(x) = 0$, so technically there may exist local minima where $\nabla E(x^*) = 0$ yet some elements are inverted. Generally, practitioners wish to rule these potential solutions out. However, this is done with two implicit, but so far informal assumptions of locality: the initial guess is not inverted, i.e. $a(x_1) > 0$, and that the solver employed must follow a path which never jumps through the barrier to inversion.

We formalize these requirements in distortion optimization as

$$\min_x \{E(x) : a(x) \geq 0\}. \quad (15)$$

Adding the constraint $a(x) \geq 0$ now explicitly restricts our optimization to noninverting deformations but otherwise leaves the desired solution unchanged. (See our Supplement for proof.)

5.3 Iterating away from collapse

With problem statement (15) in place, we now exploit it in curing the search direction from collapsing elements. At each iterate i , form the projection

$$\min_p \left\{ \|p + D_i \nabla E(x_i)\|_2^2 : a(x_i) + \nabla a(x_i)^T p \geq 0 \right\} \quad (16)$$

of the “uncured” predicted descent direction $-D_i \nabla E(x_i)$ onto the subset satisfying a linearization of the no-collapse condition. Satisfying (16) exactly would ensure that projected directions would not violate

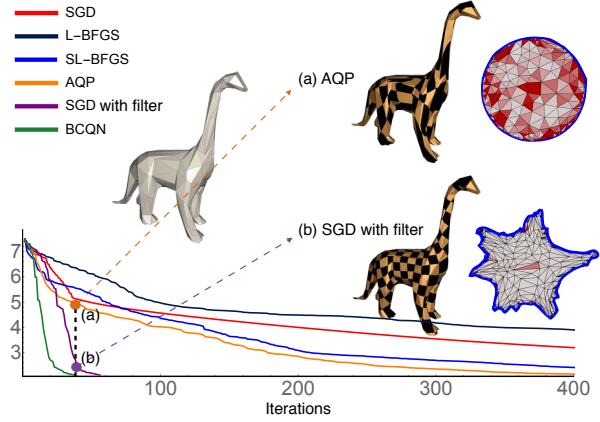


Fig. 7. Line search filtering. **Bottom:** We optimize a UV parameterization for a range of methods with the MIPS energy to consider line search filtering behavior, plotting energy (y-axis) against iteration counts. Just adding our barrier-aware line search filtering alone to SGD improves its convergence by well over an order of magnitude, and almost an order of magnitude over AQP as well as plain L-BFGS and SL-BFGS. BCQN with blending and line search filtering improves convergence even further. **Top:** a comparison between the embeddings and texture-maps gained by AQP and barrier-aware filtered SGD at the 40th iterate.

linear collapse conditions and likewise preserve symmetry [Smith et al. 2012]. However, its exact solution is neither necessary nor efficient. Instead, we construct an approximate solution to (16) as a filter that *helps* avoid collapse, preserves symmetry, and guarantees a low cost for computation for all descent steps.

Strict convexity of the projection guarantees that a minimizer p^* of (16) is given by the KKT⁵ conditions [Bertsekas 2016]

$$p^* + D_i \nabla E(x_i) - \nabla a(x_i) \lambda^* = 0, \quad (17)$$

$$0 \leq \lambda^* \perp a(x_i) + \nabla a(x_i)^T p^* \geq 0. \quad (18)$$

Setting $C_i = \nabla a(x_i)$, $M_i = \nabla a(x_i)^T \nabla a(x_i)$, and $b_i = a(x_i)$, we form the Schur complement of the above system to arrive at the equivalent Linear Complementarity Problem (LCP) [Cottle et al. 2009]

$$0 \leq \lambda^* \perp M_i \lambda^* + C_i^T p_i + b_i \geq 0. \quad (19)$$

We then form a damped Jacobi splitting $M_i = \omega^{-1} S_i + (M_i - \omega^{-1} S_i)$, with diagonal $S_i = \text{diag}(M_i)$ and damping parameter $\omega \in (0, 1)$. This gives us an iterated LCP process ranging over iteration superscripts j ,

$$0 \leq \lambda^{j+1} \perp \omega^{-1} S_i \lambda^{j+1} + M_i \lambda^j - \omega^{-1} S_i \lambda^j + C_i^T p_i + b_i \geq 0. \quad (20)$$

5.4 Line search filtering

Each iteration of the splitting (20) simplifies to the damped projected Jacobi (DPJ) update⁶

$$\lambda^{j+1} \leftarrow \left[\lambda^j - \omega S_i^{-1} (C_i^T (C_i \lambda^j) + c_i) \right]^+, \quad (21)$$

with constant $c_i = C_i^T p_i + b_i$. Here each of the m entries in λ^{j+1} can be updated in parallel (unlike Gauss-Seidel iterations). As M_i is PSD this

⁵Here and in the following $\lambda = (\lambda_1, \dots, \lambda_m)^T \in \mathbb{R}^m$ is the vector of Lagrange multipliers and $x \perp y$ is the *complementarity condition* $x_t y_t = 0, \forall t$.

⁶We use the convention $[\cdot]^+ = \max[0, \cdot]$.

iteration process converges to (19) [Cottle et al. 2009] and so to (16). We do not seek a tight solution. Instead, we just want to be sure the worst blocks to line search are filtered away. Thus we initialize with $\lambda^1 = 0$ to avoid unnecessary perturbation, use a coarse termination tolerance for DPJ (see below), and never apply more than a maximum of 20 DPJ iterations.

At each DPJ iteration j we check for termination with an LCP specialized measure, the Fischer-Burmeister function [Fischer 1992] $\text{FB}(\lambda^j, M_i \lambda^j + c_i)$ evaluated as

$$\text{FB}(a, b) = \sqrt{\sum_{k \in [1, m]} \left(a_k + b_k - \sqrt{a_k^2 + b_k^2} \right)^2}. \quad (22)$$

As we initialize with $\lambda^1 = 0$, when p_i is non-collapsing $\text{FB} = 0$, and thus no line search filtering iterations will be applied. Likewise, we stop iterations whenever the FB measure is roughly satisfied by either a relative error of $< 10^{-3}$ or an absolute error $< 10^{-6}$.

Filtering thus applies a fixed maximum upper limit on computation and performs no iterations when not necessary. Upon termination of DPJ iterations, plugging our final λ into (17) we obtain our update to form the line search filtered descent direction

$$p_i^\ell = p_i + C_i \lambda. \quad (23)$$

As illustrated in Figures 2 and 7 the filter's small number of applied iterations can make a dramatic difference in line search and so convergence.

6 TERMINATION CRITERIA

Every iterative method for minimizing an objective function $E(x)$ must incorporate stopping criteria: when should an approximate solution be considered good enough to stop and claim success? Clearly, in the usual case, where the actual local minimum value of $E(x)$ is unknown, basing the test on the current value of $E(x_i)$ is futile. As noted in Section 3.4, stopping when successive iterates are closer than some tolerance is vulnerable to false positives (halting far from a solution), as is using a fixed number of iterations. Although monitoring $\|\nabla E\|$ is robust, each individual problem may need a different tolerance to define a satisfactory solution even when normalized by number of vertices or total volume/area: see Figures 3 and 11. We thus propose a new way to derive and construct an appropriate, roughly problem-independent, relative scale for a gradient-based measure for a stopping criterion.

6.1 Characteristic gradient norm

All energies we consider are summations of per-element energy densities $W(\cdot)$ computed from the deformation gradient $F_t(x)$ and weights v_t , in each element t , as per equation (2). To simplify the following we can then evaluate energy densities on the vectorized deformation gradient as $W(\text{vec}(F_t)) = W(G_t x)$, where G_t is the linear gradient operator for element t . The full energy gradient is then

$$\nabla E(x) = \sum_{t \in T} v_t G_t^T \nabla W(G_t x). \quad (24)$$

We wish to generate a characteristic value we can compare this gradient to meaningfully, with the same dimensions; we will do this with each component of the above summation separately.

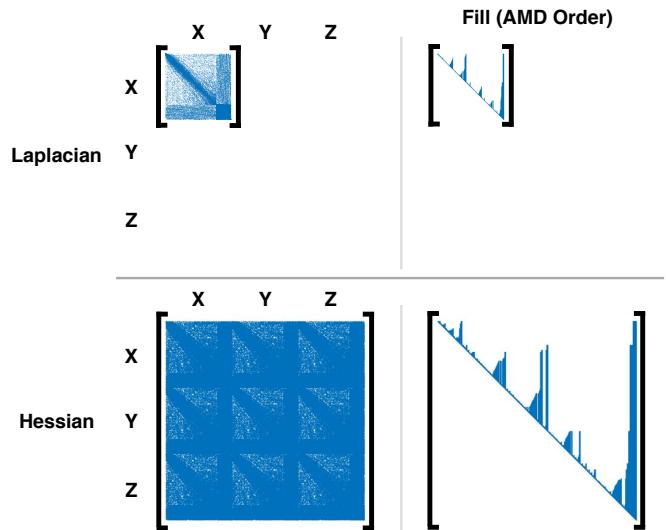


Fig. 8. **Sparsity Differences in Proxies.** Left: The scalar Laplacian (top) is smaller and sparser than the Hessian and its approximations (bottom) used in CM, PN, SLIM and AKAP. Right: This results in a much cheaper factorization and solve for the Laplacian; it is applied in both BCQN and AQP independently to each coordinate where only a one-time factorization precompute is required; CM, PN, SLIM and AKAP require factorization at each iterate.

First observe that the deformation gradient, F_t , the argument to W , is dimensionless and therefore ∇W has the same dimensions as W , and even as the element Hessian $\nabla^2 W$. For the simplest quadratic energy densities, this Hessian has the attractive property of being constant; we thus choose to use the 2-norm of the Hessian, evaluated about the deformation gradient at rest ($F_t = I$), to get a representative value for ∇W :

$$\langle W \rangle = \|\nabla^2 W(I)\|_2. \quad (25)$$

Second, note that the i^{th} part of G_t for a triangle (respectively tetrahedra) t containing vertex i will attain its maximum value for fields which are constant along the opposing edge (triangle) and that value will be the reciprocal of the altitude. Up to a factor of 2 (3), this is the length (area) of the opposing edge (triangle) divided by the rest area (volume), of the element, i.e. v_t . Summing over all incident elements, weighted by v_t , we arrive at a characteristic value for vertex i of ℓ_i equalling the perimeter (surface area) of the one-ring of vertex i . We compute this value ℓ_i , for all vertices, giving us the vector $\ell(T) = (\ell_1, \dots, \ell_n)^T \in \mathbb{R}^n$, with one scalar entry per vertex.

The product of our energy and mesh values together form the characteristic value for the norm of the gradient

$$\langle W \rangle \|\ell(T)\|, \quad (26)$$

where we take the same vector norm as that with which we evaluate $\|\nabla E(x)\|$; we use the 2-norm in all our experiments. For all methods we stop iterating when

$$\|\nabla E(x)\| \leq \epsilon \langle W \rangle \|\ell(T)\|, \quad (27)$$

given a dimensionless tolerance ϵ from the user, which is now essentially mesh- and energy-independent. See Figures 3, 10 and 11 as well as our experimental analysis in Section 8 for evaluation.

7 THE BCQN ALGORITHM

Algorithm 1 contains our full BCQN algorithm in pseudocode. The dominant cost, for both memory and runtime, is the Laplacian solve embedded in the application of D , which again is not stored as a single matrix, but rather is a linear transformation involving a few sparse triangular solves with the Laplacian's Cholesky factor and outer-product updates with a small fixed number of L-BFGS history vectors. Recall that we separately solve for each coordinate with a scalar Laplacian, not using a larger vector Laplacian on all coordinates simultaneously; this also exposes some trivial parallelism. Apart from the Laplacian, all steps are either linear (dot-products, vector updates, gradient evaluations, etc.) or typically sublinear (DPJ assembly and iterations, which only operate on the small number of collapsing triangles, and again are easily parallelized).

As Lipton et al. proved [1979], the lower bounds for Cholesky factorization on a two-dimensional mesh problem with n degrees of freedom are $O(n \log n)$ space and $O(n^{3/2})$ sequential time, and in three-dimensional problems where vertex separators are at least $O(n^{2/3})$, their Theorem 10 shows the lower bounds are $O(n^{4/3})$ space and $O(n^2)$ sequential time. On moderate size problems running on current computers, the cost to transfer memory tends to dominate arithmetic, so the space bound is more critical until very large problem sizes are reached.

7.1 Comparison with other algorithms

The per-iterate performance profile of AQP is most similar to BCQN: it too is dominated by a Laplacian solve. The only difference is the extra linear and sublinear work which BCQN does for the quasi-Newton update and the barrier-aware filtering; even on small problems, this overhead is usually well under half the time BCQN spends, and as the next section will show, the improved convergence properties of BCQN render it faster.

The second-order methods we compare against, PN and CM, as well as the more approximate proxy methods, SLIM and AKAP, all use a fuller stencil which couples coordinates. The same asymptotics for Cholesky apply, but whereas AQP and BCQN can solve a scalar $n \times n$ Laplacian d times (once for each coordinate, independently), these other methods must solve a single denser $nd \times nd$ matrix, with d^2 times more nonzeros: see Figure 8. Moreover, for all these methods the proxy matrix changes at each iteration and must be refactored, adding substantially to the cost: factorizations are significantly slower than backsolves.

8 EVALUATION

8.1 Implementation

We implemented a common test-harness code to enable the consistent evaluation of the comparative performance and convergence behavior of SGD, PN, CM, AQP, L-BFGS and BCQN across a range of energies and distortion optimization tasks including parameterization as well as 2D and 3D deformations, where these methods allow. For AQP this extends the number of energies it can be tested with, while more generally providing a consistent environment for evaluating all methods. We hope that this code will also help support the future evaluation and development of new methods for distortion optimization.

ALGORITHM 1: Blended Cured Quasi-Newton (BCQN)

```

Given:  $x_1, E, \epsilon, T$ 
Initialize and Precompute:
   $r = \epsilon \langle W \rangle \|\ell(T)\|$  // Characteristic termination value (§6)
   $L, D \leftarrow L^{-1}$  // Initialize blend model (§4)
   $g_1 = \nabla E(x_1), i = 1$ 
while  $\|g_i\| > r$  do // Termination criteria (§6)
   $p \leftarrow -Dg_i$  // Precondition gradient (§4)
  // Assemble for DPJ iterations (§5):
   $C \leftarrow \nabla a(x_i)$ 
   $M \leftarrow C^T C, c \leftarrow C^T p + a(x_i)$ 
   $S \leftarrow \text{diag}(M)^{-1}, \lambda \leftarrow 0$ 
   $fb \leftarrow \text{FB}(\lambda, M\lambda + c)$  // LCP residual (Equation (22) in §5)
  for  $j = 1$  to 20 // Line-search filtering (§5)
    if  $fb < 10^{-6}$  then break end if
     $fb \leftarrow fb_{\text{next}}$ 
     $\lambda \leftarrow [\lambda - \frac{1}{2}S(C^T(C\lambda) + c)]^+$  // Parallel project (§5)
     $fb_{\text{next}} \leftarrow \text{FB}(\lambda, M\lambda + c)$ 
    if  $|fb - fb_{\text{next}}|/fb < 10^{-3}$  then break end if
  end for
   $p^\ell \leftarrow p + C\lambda$  // Line-search filtered search direction (§5)
   $\alpha \leftarrow \text{LineSearch}(x_i, p^\ell, E)$  // Line search (§4)
   $x_{i+1} = x_i + \alpha p^\ell$  // Descent step (§4)
   $g_{i+1} = \nabla E(x_{i+1})$ 
   $D \leftarrow \text{Blend}(D, L, x_{i+1}, x_i, g_{i+1}, g_i)$  // BCQN blending update (§4)
   $i \leftarrow i + 1$ 
end while

```

The main body of the test code is in MATLAB to support rapid prototyping. All linear system solves are performed with MATLAB's native calls to CHOLMOD [Chen et al. 2008] with additional computational-heavy modules, primarily common energy, gradient and iterative LCP evaluations, implemented in C++. As linear solves are the bottleneck in all methods covered here, an additional speed-up to all methods is possible with Pardiso [Petra et al. 2014a,b] in place of CHOLMOD; however, as discussed in Section 8.4 this does not change the relative merits of the methods, and would add an additional external dependency to the test code. For verification we also confirm that iterations in the test-harness AQP and CM implementations match the official AQP [Kovalsky et al. 2016] and CM [Shtengel et al. 2017] codes.

All experiments were timed on a four-core Intel 3.50GHz CPU. We have parallelized the damped Jacobi LCP iterations with Intel TBB; with more cores the overhead reported below for LCP iterations is expected to diminish rapidly. For all UV parameterization problems we follow Kovalsky et al. [2016] and compute an initial, locally injective embedding via a single linear solve with the cotan Laplacian; if it fails (is not locally injective), we then fall back to plain Tutte, so that robustness is maintained. For all constrained deformation examples, with the exception of the Armadillo in Figure 14, we begin with an initially injective mapping. For the Armadillo deformation example only, we apply the LBD method [Kovalsky et al. 2015] to create a rough, locally injective initialization from the initial constrained non-injective deformation and pass this to both compared methods. To enforce Dirichlet boundary conditions, i.e. positional constraints, we use a standard subspace projection [Nocedal and Wright 2006], i.e. removing those degrees of freedom from the problem. For constrained problems we apply the standard approach of projecting gradients

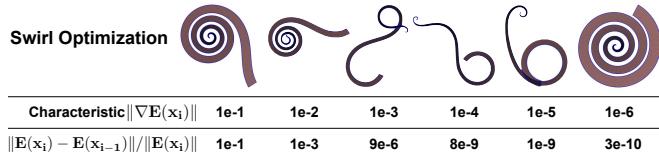


Fig. 9. **Termination criteria comparison.** Left to right: We find key points in the sequential progress of the optimized mesh in the Swirl optimization (ISO energy) example at regular intervals of $10\times$ decrease in our characteristic norm. We compare with the relative error measures at these same points.

to the null-space of constraints: a stationary point is then reached only if the (projected) gradient vanishes, just as in the unconstrained setting. When line search is employed we first find a maximal non-inverting step size with Smith and Schaefer’s method [2015], followed by standard line search with Armijo and curvature conditions.

8.2 Termination

To evaluate termination criteria behavior we first instrumented two geometry optimization stress-test examples: the *Swirl* deformation [Chen et al. 2013] and the *Hilbert curve* UV parametrization [Smith and Schaefer 2015]. We run both examples to convergence ($\epsilon = 10^{-6}$ using our characteristic gradient) reaching the final target shapes for each. Within these optimizations we record the 2-norm of gradient, the vertex-normalized 2-norm of gradient, the relative error measure [Kovalsky et al. 2016; Shtengel et al. 2017] and our characteristic gradient norm for all iterations.

Figure 9 shows the Swirl mesh obtained during BCQN iteration at regular intervals of $10\times$ decrease in our characteristic norm. Observe that they correspond to natural points of progress; see our supplemental video of the entire optimization sequence for reference. For comparison we also provide the corresponding relative error measures, which varies much less steadily.

In Figure 10 we compare termination criteria more closely for a UV parametrization problem, the Hilbert curve example. We plot our characteristic gradient norm (blue) and the relative energy error [Kovalsky et al. 2016; Shtengel et al. 2017] (orange) as BCQN proceeds. Note that the characteristic gradient norm provides consistent decrease corresponding to improved shapes and so provides a practical measure of improvement. The local error in energy, on the other hand, varies greatly, making it impossible to judge how much global progress has been made towards the optimum.

Figure 11 illustrates consistency across changing tolerance values, mesh resolutions, and scales. We show the iterates at measures 10^{-3} , 10^{-4} and 10^{-5} for both our characteristic gradient norm and the raw gradient norm, for meshes with varying refinement and varying dimension (rescaling coordinates by a large factor). Similar to Figure 3 comparing the vertex-normalized gradient norm, there are large disparities for the raw gradient norm, but our characteristic gradient norm is consistent.

Tolerances. The Swirl and Hilbert curve examples are both extreme stress tests that require passing through low curvature regions to transition from unfolding to folding; see e.g., Figure 9 above and our videos. For these extreme tests we used a tolerance of 10^{-6} for our characteristic gradient norm to consistently reach the final target shape. However,

for most practical distortion optimization tasks such a tolerance is excessively precise. In experiments across a wide range of energies and UV parametrization, 2D and 3D deformation tasks, including those detailed below, we found that $\|\nabla E(x)\| \leq 10^{-3} \langle W \rangle \|\ell(V, T)\|$ consistently obtained good-looking solutions with essentially no further visible (or energy value) improvement possible. We argue this is a sensible default except in pathological examples. For all examples discussed here and below, with the exception of the Swirl and the Hilbert curve tests, we thus use $\epsilon = 10^{-3}$ for testing termination.

8.3 Newton-type methods

While Newton’s method, on its own, handles energies like ARAP reasonably well [Chao et al. 2010] it is insufficient for strongly nonconvex energies: modification of the Hessian is required [Nocedal and Wright 2006; Shtengel et al. 2017]. Here we examine the convergence, performance and scalability of PN [Teran et al. 2005], a general-purpose modification for nonconvex energies, and CM [Shtengel et al. 2017], a more recent convex majorizer currently restricted to 2D problems and a trio of energies (ISO, Symmetric ARAP and NH), and compare them with AQP and BCQN. For the 2D parameterization problems in Figure 12 we can compare all four methods while for the 3D deformation problems in Figures 13 and 14 CM is not applicable.

Figures 12, 13, and 14 examine the scaling behavior of the various methods under mesh refinement, for 2D parameterization and 3D deformation. The Newton-type methods PN and CM (when applicable) maintain low iteration counts that only grow slowly with increasing mesh size; from the outset BCQN and AQP require more iterations, though the iteration count also grows slowly for BCQN. Nonetheless, BCQN is the fastest across all scales in each test as its overall cost per iteration remains much lower. BCQN iterations require no refactorizations (which scales poorly, particularly in 3D, as discussed in Section 7) and only solves smaller and sparser scalar Laplacian problems per coordinate compared to the larger and denser system of

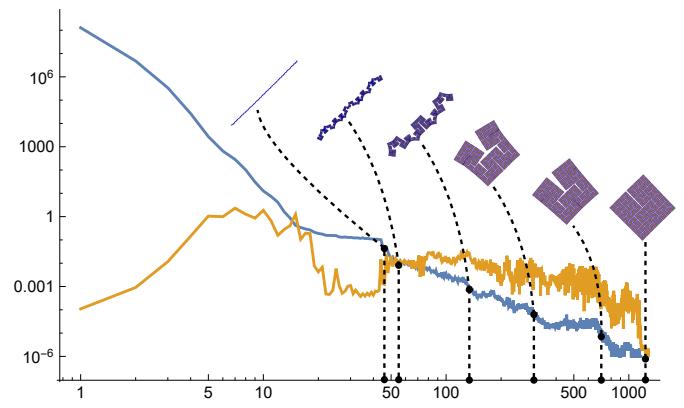


Fig. 10. **Measuring improvement.** Solving a UV parametrization of the Hilbert curve with BCQN, we plot our characteristic gradient norm in blue and the relative energy error in orange as the method proceeds, on a logarithmic scale. Iterates are shown at decreases in the characteristic gradient norm by factors of 10, illustrating its efficacy as a global measure of progress, while the relative energy error measures only local changes with little overall trend.

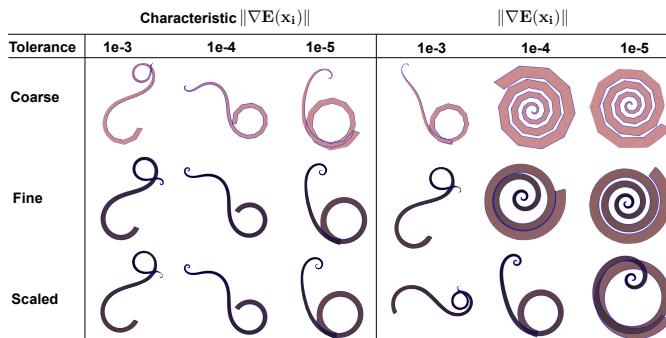


Fig. 11. **Termination criteria comparison across mesh refinement and scale.** Left and right: we show the Swirl optimization when our characteristic norm (left) and the standard gradient norm (right) reach 10^{-3} , 10^{-4} and 10^{-5} . Top to bottom: the rows show optimization with a coarse mesh, a fine mesh, and the same fine mesh uniformly scaled in dimension by 100x. Note the consistency across mesh resolution and scaling for our characteristic norm and the disparity across the standard gradient norm.

CM and PN. This advantage for BCQN only increases as problem size grows; indeed, for the largest problems BCQN succeeded where CM and PN ran out of memory for factorization.

8.4 A note on solving proxies and Pardiso

Recent methods including CM have taken advantage of the efficiencies and optimizations provided by the Pardiso solver. While this can improve runtime of the factorization and backsolves by a constant factor, it cannot change the asymptotic lower bounds on complexity; the sparse matrix orderings in both CHOLMOD and Pardiso already appear to achieve the bound on typical mesh problems. In tests on our computer, across a large range of scales in two and three dimensions, we found Pardiso was occasionally slower than CHOLMOD but usually 1.4 to 3 times faster, and at most 8.1 times faster (for backsolving with a 3D scalar Laplacian).

Individual iterates of AQP have the same overall efficiency as BCQN (dominated by the linear solves); switching to Pardiso leaves the relative performance of the two methods unchanged. While CM and PN are even more dependent on the efficiency of the linear solver, due to more costly refactorization each step, the same speed-ups possible with Pardiso also apply to BCQN, so again there is no significant change in relative performance between the methods.

8.5 First-order methods

Among existing first-order methods for distortion optimization AQP has so far shown best efficiency [Kovalsky et al. 2016] with improved convergence over SGD as well as standard L-BFGS. Likewise, as we see in Figures 12, 13 and 14, when we scale to increasingly larger problems AQP will dominate over Newton-type methods and so potentially offers the promise of reliability across applications. Finally, BCQN performs a small, fixed amount of extra work per-iteration in the line-search filter and quasi-Newton update. Thus in Figures 12, 13, 15, 16 and 17 we compare AQP and BCQN over a range of practical distortion optimization applications, UV-parameterization, 2D deformation, and 3D deformation, with nonconvex energies from geometry processing

and physics. Throughout we note three key features distinguishing BCQN as follows.

Reliability and robustness: AQP will fail to converge in some cases, see e.g. Figure 4, while BCQN reliably converges. In our testing AQP fails to converge in over 40% of our tests with nonconvex energies; see e.g. Figures 15 and 16. This behavior is duplicated in our test-harness code and AQP’s reference implementation.

Convergence speed: when AQP is able to converge, BCQN consistently provides faster convergence rates for nonconvex energies. In our experiments convergence rates range up to well over 10X with respect to AQP.

Performance: BCQN is efficient. When AQP is able to converge, BCQN remains fast with up to a well over 7X speedup over AQP on nonconvex energies.

8.6 Across-the-board comparisons

Here we compare the performance and memory usage of BCQN with best-in-class distortion optimization methods across the board: AQP, PN and CM for both 2D parameterization and 3D deformation tasks. Results are summarized in Figures 12, 13 and 14. Note that CM does not extend to 3D.

In Figures 12 and 13 we examine the scaling of AQP, PN, CM and BCQN to larger meshes and thus to larger problem sizes in both 2D parameterization (up to 23.9M triangles) and 3D deformation (up to 7.8M tetrahedra). As noted above: from the outset, BCQN requires more iterations than CM and PN; however, BCQN’s overall low cost per iteration makes it faster in performance across problem sizes when compared to both CM and PN. We then note that AQP, on the other hand, has slower convergence and so, at smaller sizes it often does not compete with CM and PN. However, once we reach larger mesh problems, e.g. ~ 6 M triangles in Figure 12, the cost of factorization and backsolve of the denser linear systems of CM and PN becomes significant so that even AQP’s slower convergence results in improvement. This is the intended domain for which first-order methods are designed but here too, as we see in Figure 12, BCQN continues to outperform both AQP as well as CM and PN across all scales. Please see our supplemental video for visual comparisons of the relative progress of PN, CM, AQP and BCQN.

9 CONCLUSION

In this work we have taken new steps to both advance the state of the art for optimizing challenging nonconvex deformation energies and to better evaluate new and improved methods as they are subsequently developed. Looking forward these minimization tasks are likely to remain fundamental bottlenecks in practical codes and so advancement here is critical. Our three primary contributions together form the BCQN algorithm which pushes current limits in optimizing distortion forward in terms of speed, reliability, and automatability. At the same time, looking ahead we also expect that each contribution individually should lead to even more improvements in the near future.

9.1 Limitations and future work

While our focus is on recent challenging nonconvex energies not addressed by the popular local-global framework, similar to AQP we have observed significant speedup for energies that are convex in

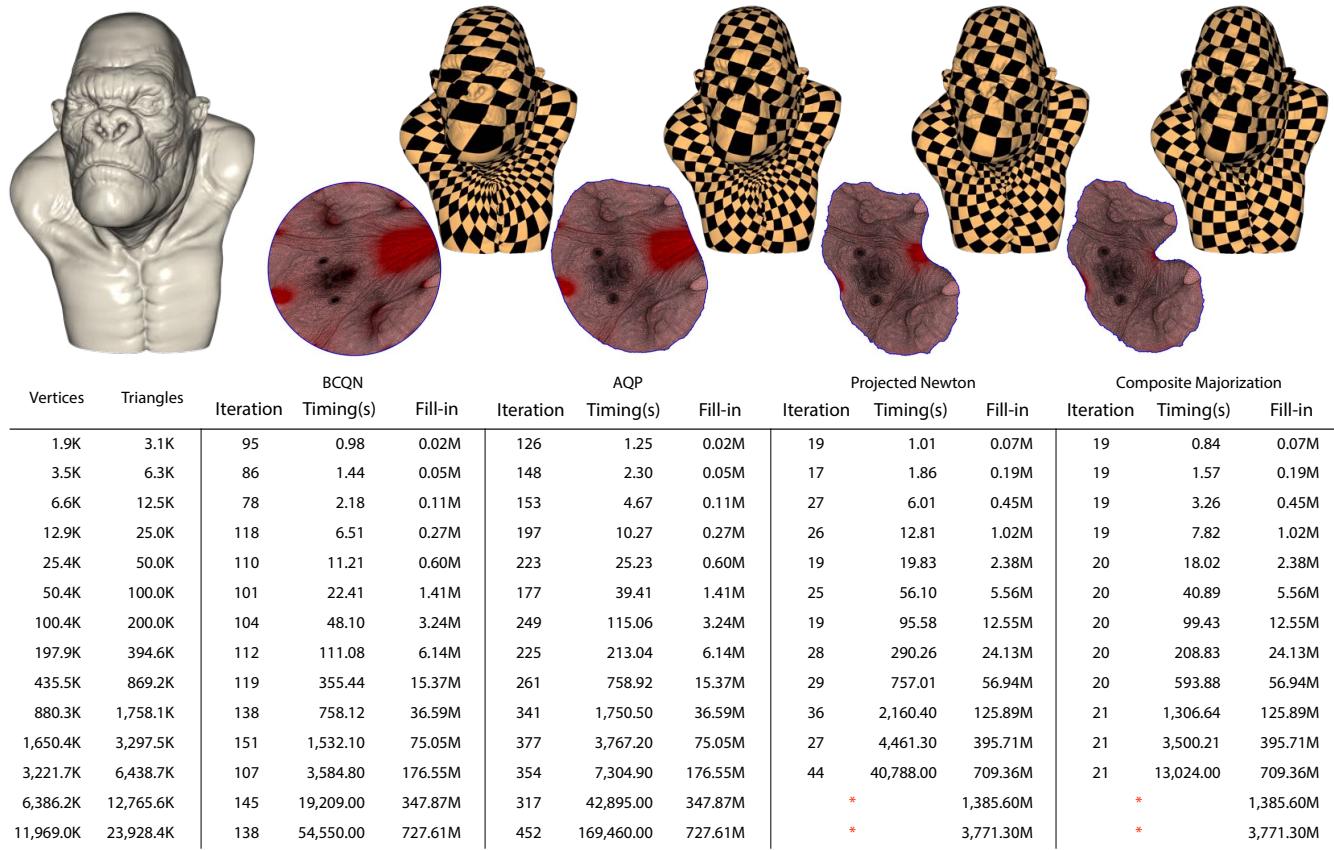


Fig. 12. **UV Parameterization Scaling, Timing and Sparsity.** Performance statistics and memory use for increasing mesh sizes up to 23.9M triangles, comparing BCQN with AQP, PN and CM. For the Gorilla UV parametrization with ISO energy we repeatedly double the mesh resolution and, for each method, report number of iterations to convergence (characteristic norm $< 10^{-3}$), wall-clock time (seconds) to convergence, and the nonzero fill-in for the linear systems solved by each method. We use * to indicate out-of-memory failure for matrix factorization; see §8 for discussion. Also note that stencils for CM and PN are identical (differing only in actual entries) while AQP and BCQN both solve with the same smaller scalar Laplacian.

singular values, e.g. ARAP, as well. Currently, in comparing AQP and BCQN on the same set of 2D and 3D tasks with the ARAP energy we observe a generally modest improvement in convergence, up to a little over 4x, which is somewhat balanced by the small additional overhead of BCQN iterations. Note for energies like ARAP there is no barrier, hence no need for our line search filtering, but other opportunities for improvement may be available in future research.

While our current blending model works well in our extensive testing, it is empirically constructed; it is in no sense proven optimal. We believe that further analysis, better understanding and additional improvements in quasi-Newton blending are all exciting and promising avenues of future investigation.

Our focus has been on unconstrained optimization, such as problems where linear constraints (e.g. pinned vertices) have already been removed from the degrees of freedom. As stated earlier, unconstrained optimization is also often a crucial inner step in constrained optimization methods such as Augmented Lagrangian or Interior Point. However, we caution that while our filtering step, based on linearized constraints to prevent element inversion, is a sound and effective

means to improve the subsequent line search, it is not a general-purpose tool to handle even just linear constraints: it relies crucially on the fact that the no-inversion constraints we add should already be satisfied by the unconstrained solution, i.e. ultimately are inactive. We leave the question of incorporating active constraints directly into filtering and blending BFGS for future work and, as a starting point, observe that linear constraints can be directly added with a KKT formulation.

Finally, we note that while we have focused here on optimizing distortion energies defined on meshes, there is a wide range of critical optimization problems that take a similar general structure: minimizing separable nonlinear energies on graphs. Further extensions are thus exciting directions of ongoing investigation.

ACKNOWLEDGEMENTS

We thank Kevin Wampler and Vova Kim for valuable discussion, and NSERC of Canada for partly supporting this research.

REFERENCES

- Noam Aigerman, Roi Poranne, and Yaron Lipman. 2015. Seamless surface mappings. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 72.

Vertices	Tetrahedra	BCQN			AQP			Projected Newton			Laplacian Fill-in
		Iteration	Timing(s)	Iteration	Timing(s)	Iteration	Timing(s)	Iteration	Timing(s)	Fill-in	
1.8K	6.3K	62	4.22	262	15.97	41	22.01	0.61M	0.03M		
3.0K	10.8K	63	7.02	311	30.56	42	39.61	1.35M	0.07M		
6.6K	25.6K	76	18.27	342	79.51	43	97.21	4.88M	0.27M		
16.0K	66.6K	130	74.90	408	226.02	44	282.64	21.04M	1.09M		
31.7K	137.2K	151	168.59	475	517.32	44	644.97	65.06M	3.29M		
65.3K	291.6K	178	429.48	546	1,346.30	44	1,673.90	214.44M	10.44M		
133.1K	608.4K	274	1,391.70	538	2,640.00	44	15,650.00	695.33M	31.25M		
261.9K	1,219.5K	291	2,805.10	685	76,711.00	45	1,135.20M	87.26M			
535.7K	2,532.7K	352	7,930.60	825	28,500.00	45	1,104,480.00	3,078.80M	270.73M		
1,633.5K	7,873.2K	392	20,908.00	919	88,215.00	*	9,672.17M	689.42M			

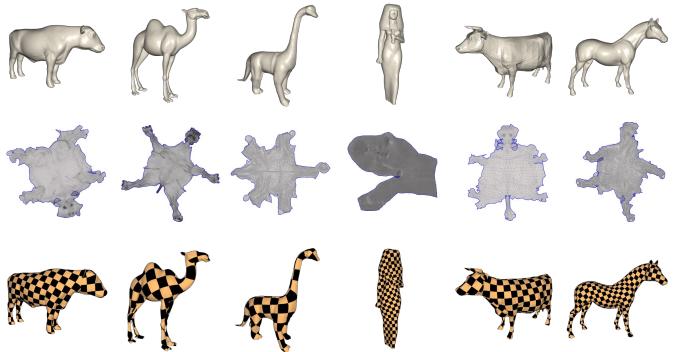


Fig. 13. **Three-Dimensional Deformation Scaling, Timing and Sparsity.** Performance statistics and memory use for increasing mesh sizes up to 7.8M tetrahedra, comparing BCQN with AQP and PN. (CM does not extend to 3D.) We initialize a bar with a straight rest shape to start in a tightly twisted shape, constraining both ends to stay fixed and then optimize over increasing resolutions. For each method we report number of iterations to convergence (characteristic norm $< 10^{-3}$) and wall-clock time (seconds) to convergence, as well as the nonzero fill-in for the linear system solved by CM as compared with the Laplacian employed by both BCQN and AQP. We use * to indicate out of memory for the computation on our test system; see §8 for discussion.

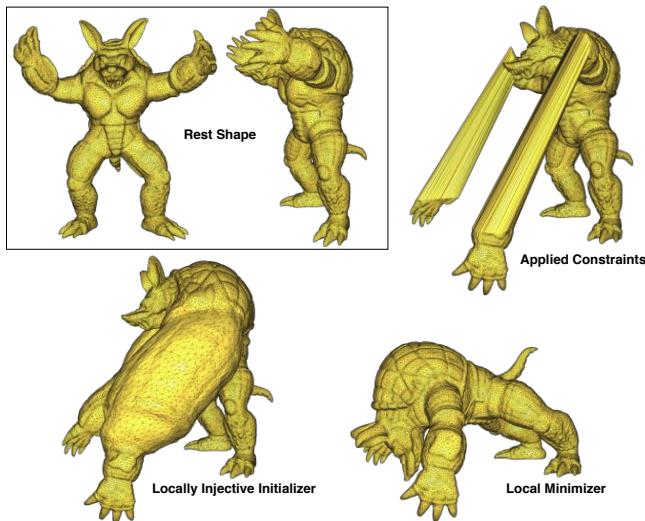


Fig. 14. **Armadillo Deformation Test.** We compare three-dimensional deformation optimizations of a 1.5M element tetrahedral mesh of the T-pose armadillo with BCQN and PN. We constrain the armadillo’s feet to rest position, its hands to touch the ground and use the LBD method [Kovalsky et al. 2015] to create a locally injective initialization for the solvers from the starting constrained configuration. Here BCQN requires 393 iterations to converge while PN converges in just 9. However, as BCQN remains cheaper and more scalable per iterate it takes only 4,148 seconds, while PN spends 13,447 seconds in our test system. See Figure 8 for the comparison of fill-in and size.

- Mirela Ben-chen, Craig Gotsman, and Guy Bunin. 2008. Conformal flattening by curvature prescription and metric scaling. *Computer Graphics Forum* 27 (2008).
- Dimitri P. Bertsekas. 2016. *Nonlinear Programming*. Athena Scientific.
- David Bommes, Marcel Campen, Hans-Christian Ebke, Pierre Alliez, and Leif Kobbelt. 2013. Integer-grid Maps for Reliable Quad Meshing. *ACM Trans. Graph.* 32, 4 (July 2013), 98:1–98:12.

Example	Vertices	Triangles	Energy	BCQN		AQP	
				Iteration	Timing	Iteration	Timing
Bull	17.9K	34.5K	ISO	169	11.82	†	†
			MIPS	200	24.08	†	†
			CONF	425	50.06	1,085	74.45
Camel	40.2K	78.1K	ISO	412	92.60	593	114.89
			MIPS	3,344	821.57	66,479	12,281.00
			CONF	9,196	2,067.60	81,100	15,712.00
Dino	24.6K	47.9K	ISO	162	30.43	950	98.66
			MIPS	283	54.13	†	†
			CONF	156	29.53	883	98.02
Isis	188.1K	374.3K	ISO	347	404.63	353	366.01
			MIPS	204	298.55	†	†
			CONF	104	166.65	611	612.98
Cow	3.1K	5.8K	ISO	78	1.14	†	†
			MIPS	239	3.29	323	3.98
			CONF	59	1.05	108	1.23
Horse	20.6K	39.6K	ISO	104	12.93	643	57.40
			MIPS	387	47.73	†	†
			CONF	519	68.22	5,552	523.97

Fig. 15. **UV parameterization.** Top row: 3D meshes for UV parametrization with ISO, MIPS, and CONF distortion energies. Middle two rows: converged maps and texturing from BCQN on ISO examples. Bottom: for each method / problem pair we report number of iterations to convergence (characteristic norm $< 10^{-3}$) and wall-clock time (seconds) to convergence. We use † to indicate when AQP does not converge; see §8.5.

- J Bonet and AJ Burton. 1998. A simple orthotropic, transversely isotropic hyperelastic constitutive equation for large strain computations. *Computer methods in applied mechanics and engineering* 162, 1 (1998), 151–164.
- Allan F Bower. 2009. *Applied mechanics of solids*. CRC press.
- Isaac Chao, Ulrich Pinkall, Patrick Sanan, and Peter Schröder. 2010. A simple geometric model for elastic deformations. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 38.
- Renjie Chen and Ofir Weber. 2017. GPU-accelerated Locally Injective Shape Deformation. *ACM Trans. Graph.* 36, 6 (Nov. 2017), 214:1–214:13.
- Renjie Chen, Ofir Weber, Daniel Keren, and Mirela Ben-Chen. 2013. Planar Shape Interpolation with Bounded Distortion. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 108:1–108:12.
- Yanqing Chen, Timothy A. Davis, William W. Hager, and Sivasankaran Rajamanickam. 2008. Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate. *ACM Transactions on Mathematical Software (TOMS)* 35, 3 (2008), 14.
- Sebastian Claici, Mikhail Bessmeltsev, Scott Schaefer, and Justin Solomon. 2017. Isometry-Aware Preconditioning for Mesh Parameterization. In *Proceedings of the Symposium on Geometry Processing*.
- Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. 2009. *The Linear Complementarity Problem*. Society for Industrial & Applied Mathematics (SIAM).
- Mathieu Desbrun, Mark Meyer, and Pierre Alliez. 2002. Intrinsic parameterizations of surface meshes. In *Computer Graphics Forum*, Vol. 21. Wiley Online Library, 209–218.
- A Fischer. 1992. A special Newton-type optimization method. 24 (1992), 269–284.
- Xiao-Ming Fu, Yang Liu, and Baining Guo. 2015. Computing locally injective mappings by advanced MIPS. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 71.
- N.J. Higham. 1992. Estimating the matrix p-norm. *Numer. Math.* 62 (1992).

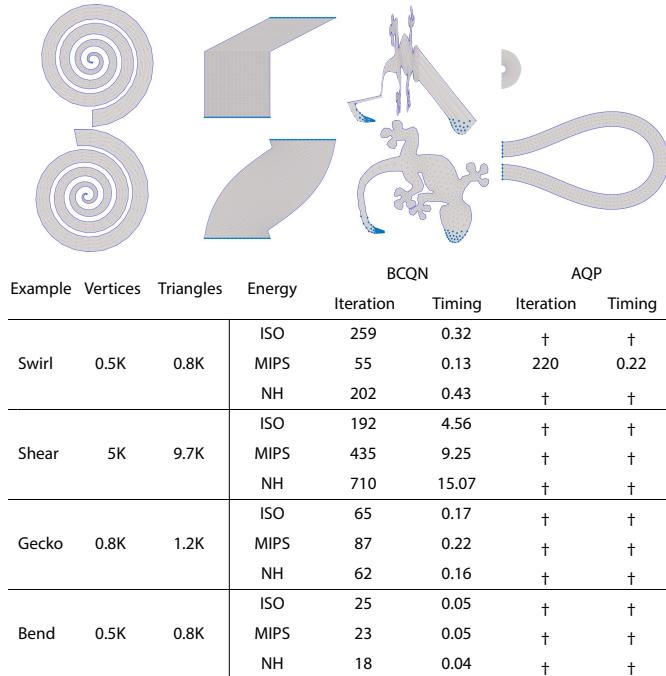


Fig. 16. Two-Dimensional Deformation. **Top:** initial conditions and vertex constraints (blue points) for deformation problems minimizing ISO, MIPS, and NH deformation energies. **Middle:** converged solutions from BCQN on ISO examples. **Bottom:** for each method / problem pair we report number of iterations to convergence (characteristic norm $< 10^{-3}$) and wall-clock time (seconds) to convergence. We use † to indicate when AQP does not converge; see §8.5.

- Kai Hormann and Günther Greiner. 2000. *MIPS: An efficient global parametrization method*. Technical Report. DTIC Document.
- Lianjun Jiang, Richard H Byrd, Elizabeth Eskow, and Robert B Schnabel. 2004. *A preconditioned L-BFGS algorithm with application to molecular energy minimization*. Technical Report. DTIC Document.
- Shahar Z Kovalsky, Noam Aigerman, Ronen Basri, and Yaron Lipman. 2015. Large-scale bounded distortion mappings. *ACM Trans. Graph.* 34, 6 (2015), 191.
- Shahar Z Kovalsky, Meirav Galun, and Yaron Lipman. 2016. Accelerated Quadratic Proxy for Geometric Optimization. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)* 35, 4 (2016).
- Bruno Lévy, Sylvain Petitjean, Nicolas Ray, and Jérôme Maillot. 2002. Least squares conformal maps for automatic texture atlas generation. In *ACM Transactions on Graphics (TOG)*, Vol. 21. ACM, 362–371.
- Yaron Lipman. 2012. Bounded Distortion Mapping Spaces for Triangular Meshes. *ACM Trans. Graph.* 31, 4 (July 2012), 108:1–108:13. <https://doi.org/10.1145/2185520.2185604>
- Richard Lipton, Donald Rose, and Robert Targan. 1979. Generalized Nested Dissection. *SIAM J. Numer. Anal.* 16, 2 (1979), 346–358.
- Ligang Liu, Lei Zhang, Yin Xu, Craig Gotsman, and Steven J Gortler. 2008. A local/global approach to mesh parameterization. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 1495–1504.
- Tiantian Liu, Sofien Bouaziz, and Ladislav Kavan. 2017. Quasi-Newton Methods for Real-Time Simulation of Hyperelastic Materials. *ACM Transactions on Graphics (TOG)* 36, 3 (2017), 23.
- 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.
- Melvin Mooney. 1940. A Theory of Large Elastic Deformation. *Journal of Applied Physics* 11, 9 (1940), 582–592.
- Patrick Mullen, Yiyi Tong, Pierre Alliez, and Mathieu Desbrun. 2008. Spectral Conformal Parameterization. In *Proceedings of the Symposium on Geometry Processing (SGP '08)*. Eurographics Association, 1487–1494.
- Yuriii Nesterov. 1983. A method of solving a convex programming problem with convergence rate $O(1/\sqrt{k})$. *Soviet Mathematics Doklady* 27 (1983), 372–376.

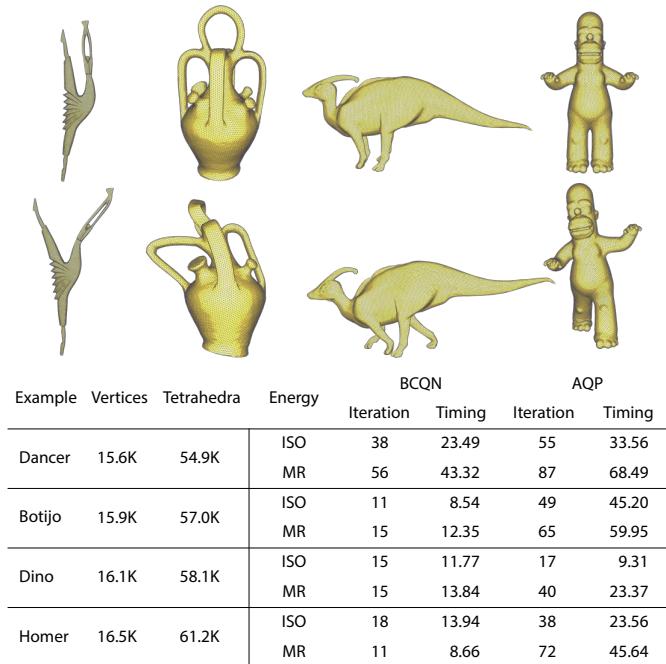


Fig. 17. Three-Dimensional Deformation. **Top:** initial conditions for vertex-constrained deformation problems minimizing ISO and MR deformation energies. **Middle:** converged solutions satisfying constraints from BCQN on MR examples. **Bottom:** for each method / problem pair we report number of iterations to convergence (characteristic norm $< 10^{-3}$) and wall-clock time (seconds) to convergence.

- JW Neuberger. 1985. Steepest descent and differential equations. *Journal of the Mathematical Society of Japan* 37, 2 (1985), 187–195.
- J. Nocedal and S. Wright. 2006. *Numerical Optimization*. Springer New York.
- Raymond Ogden. 1972. Large Deformation Isotropic Elasticity – On the Correlation of Theory and Experiment for Incompressible Rubberlike Solids. In *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 565–584.
- Cosmin G. Petra, Olaf Schenk, and Mihai Anitescu. 2014a. Real-time stochastic optimization of complex energy systems on high-performance computers. *IEEE Computing in Science & Engineering* 16, 5 (2014).
- Cosmin G. Petra, Olaf Schenk, Miles Lubin, and Klaus Gärtner. 2014b. An augmented incomplete factorization approach for computing the Schur complement in stochastic optimization. *SIAM Journal on Scientific Computing* 36, 2 (2014).
- Michael Rabinovich, Roi Poranne, Daniele Panozzo, and Olga Sorkine-Hornung. 2017. Scalable Locally Injective Mappings. *ACM Trans. Graph.* 36, 2 (April 2017), 16:1–16:16.
- Ronald S. Rivlin. 1948. Some Applications of Elasticity Theory to Rubber Engineering. In *Proc. Rubber Technology Conference*, 1–8.
- Anna Shtengel, Roi Poranne, Olga Sorkine-Hornung, Shahar Z. Kovalsky, and Yaron Lipman. 2017. Geometric Optimization via Composite Majorization. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 11.
- Breannan Smith, Danny M. Kaufman, Etienne Vouga, Rasmus Tamstorf, and Eitan Grinspun. 2012. Reflections on Simultaneous Impact. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2012)* 31, 4 (2012), 106:1–106:12.
- Jason Smith and Scott Schaefer. 2015. Bijective parameterization with free boundaries. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 70.
- Olga Sorkine and Marc Alexa. 2007. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, Vol. 4.
- Joseph Teran, Eftychios Sifakis, Geoffrey Irving, and Ronald Fedkiw. 2005. Robust quasistatic finite elements and flesh simulation. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*. ACM, 181–190.
- Huamin Wang and Yin Yang. 2016. Descent methods for elastic body simulation on the GPU. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 212.
- Ofir Weber, Ashish Myles, and Denis Zorin. 2012. Computing extremal quasiconformal maps. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 1679–1689.
- Hongyi Xu, Fusheng Sin, Yufeng Zhu, and Jernej Barbic. 2015. Nonlinear material design using principal stretches. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 75.