MITSUBISHI ELECTRIC RESEARCH LABORATORIES
https://www.merl.com

From Subspaces to Submanifolds

Matthew Brand

TR2004-134 September 07, 2004

Abstract

This paper identifies a broad class of nonlinear dimensionality reduced (NLDR) problems where the exact local isometry between an extrinsically curved data manifold M and a low-dimensional parameterization space can be recovered from a finite set of high-dimensional point sampels. The method, Geodesic Nullsapce Analysis (GNA), rests on two results: First, the exact isometric parameterization of a local point clique on M haas an algebraic reduction to arc-length integrations when the ambient-space embedding of M is locally a product of planar quadrics. Second, the locally isometric global parameterization lies in the left invariant subspace of a linearizing operator that averages the nullspace projectors of the local parameterizations. We show how to use the GNA operator for denosing, dimensionality reduction, and resynthesis of both the original data and of new samples, making such s̈ubmanifoldm̈ethods an attractive alternative to subspace methods in data analysis.

British Machine Vision Conference (BMVC)

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

From subspace to submanifold methods

Matthew Brand
Mitsubishi Electric Research Labs
Cambridge, Massachusetts 02139, USA
www.merl.com/people/brand/

Abstract

This paper identifies a broad class of nonlinear dimensionality reduction (NLDR) problems where the *exact* local isometry between an extrinsically curved data manifold $\mathcal{M}$ and a low-dimensional parameterization space can be recovered from a finite set of high-dimensional point samples. The method, Geodesic Nullspace Analysis (GNA), rests on two results: First, the exact isometric parameterization of a local point clique on $\mathcal{M}$ has an algebraic reduction to arc-length integrations when the ambient-space embedding of $\mathcal{M}$ is locally a product of planar quadrics. Second, the locally isometric global parameterization lies in the left invariant subspace of a linearizing operator that averages the nullspace projectors of the local parameterizations. We show how to use the GNA operator for denoising, dimensionality reduction, and resynthesis of both the original data and of new samples, making such "submanifold" methods an attractive alternative to subspace methods in data analysis.

Nonlinear dimensionality reduction is the problem of constructing a low-dimensional representation of high-dimensional sample data. The data is presumed to sample a d-dimensional manifold $\mathcal{M}$ that is embedded in an ambient space $\mathbb{R}^D$, with $D > d$. The goal is to separate the extrinsic geometry of the embedding—how $\mathcal{M}$ is shaped in $\mathbb{R}^D$—from its intrinsic geometry—the native d-dimensional coordinate system on $\mathcal{M}$. For example, if we know how the manifold of faces is embedded in the space of images, the intrinsic geometric can be exploited to edit, compare, and classify faces, while the extrinsic geometry can be exploited to detect faces in images and synthesize new face images. In computer vision it is common to approximate this manifold with linear subspaces fitted to sample images via principal components analysis (PCA)—a gross approximation, yet enormously successful for data interpolation, extrapolation, compression, denoising, and visualization. Ultimately, NLDR "submanifold" methods will offer the same functionality, but with more fidelity to the true data distribution, because most of these operations are exclusive to the intrinsic or extrinsic geometry of $\mathcal{M}$.

Whereas PCA aims to preserve global structure (covariance about the data mean) NLDR methods aim to preserve local structure in small neighborhoods on the manifold. Differential geometry teaches that the local metric on infinitesimal neighborhoods plus information about the connectivity of those neighborhoods fully determines a manifold's intrinsic geometry [7]. This is approximated for finite data by imposing a neighborhood graph on the data and measuring relations between neighboring points in graph cliques, typically point-to-point distances [18, 21, 2], the coordinates of points projected into a local tangent space [4, 8, 22] or local barycentric coordinates [15]. A key assumption in this

literature is that local *linear* structure of point cliques in the ambient space can be used as a proxy for metric structure in corresponding neighborhoods on $\mathcal{M}$. E.g., distances in $\mathbb{R}^D$ stand for geodesic arc-lengths on $\mathcal{M}$. The graph then guides the combination of all local metric constraints into a quadratic form whose maximizing or minimizing eigenfunctions provide a minimum squared error basis for embedding the manifold in $\mathbb{R}^d$. For discrete data, the quadratic form is a gram matrix whose entries can be interpreted as inner products between points in an unknown space where the manifold is linearly embedded, therefore NLDR is a kernel method, albeit with unknown kernel function [12]. Of particular interest for signal processing and data modeling is the case where the sampled manifold patch $\mathcal{M}$ is locally isometric to a connected patch of $\mathbb{R}^d$, because in the continua limit of infinite sampling, the optimizing eigenfunctions yield a flat immersion that perfectly reproduces the local data density and intrinsic geometry of the manifold [8]. Thus most NLDR "embedding[1] algorithms" strive for isometry.

NLDR is built on mathematical foundations that were laid down over the last 40 years in the problem area of graph embeddings [19, 20, 9, 10, 6]; the current rapid advances in machine learning revolve around the insight that dimensionality reduction can be shoehorned into this framework by estimating a graph and local metric constraints to cover datasets of unorganized points [18, 15, 2, 4, 8]. This insight is also the Achilles' heel of NLDR methods, which suffer two mutually exacerbating maladies when presented with finite data embedding problems:

1. Local metric constraints are systematically distorted because data drawn from an extrinsically curved manifold is not locally linear at any finite scale. I.e. distances in $\mathbb{R}^D$ are a *biased* approximation of geodesic arc-lengths on $\mathcal{M}$.

2. If the local estimated metric constraints contain any errors, the global solution has minimum mean squared error (MMSE) with respect to an invented system of neighborhoods rather than w.r.t. the empirical data distribution.

Accordingly, NLDR methods have been noted, though somewhat unfairly, for the inconsistency and instability of their results [1], especially under small changes to the connectivity graph.

This paper contributes two results that address these problems:

1. There exists an interesting and useful class of extrinsically curved data manifolds—namely those that are are locally products of planar quadrics (PPQ)—whose local metric structure can be recovered *exactly* from finite sampling. Most developable surfaces are locally PPQ and all smooth manifolds have local PPQ approximations that are second-order accurate in directions of maximal curvature. Section 3 gives a least-squares procedure for computing geodesic parameterizations of point cliques under this model.

2. The coordination of this local metric structure into a globally consistent isometric immersion in $\mathbb{R}^d$ can be uniquely determined from a linearizing operator that averages the collected nullspaces of local clique parameterizations. The operator isolates the component of data that is a nonlinear function of the local metric structure

[1] Their names notwithstanding, NLDR "embedding algorithms" really only compute immersions; whether or not an immersion is an embedding and/or a local isometry is a property of the manifold, not of the algorithm [7].

(i.e., extrinsic curvature and noise) and defines a spectrum of data transformations that range from local denoising to global dimensionality reduction. Because this operator averages out artifacts due to uneven coverage of the data by the clique graph, the immersion has the MMSE property that error, if any, is distributed evenly over the data.

This paper will use the acronym GNA (*Geodesic Nullspace Analysis*) for the combination of these methods. GNA offers much of the functionality of subspace methods—data reduction, denoising, out-of-sample generalization, and sequential updating with new data points. It also offers significantly reduced error on benchmark NLDR problems, where ground truth is known, and also appears to perform well with real datasets. Finally, as a theoretical nicety, there are manifolds whose isometric parameterization can be recovered exactly from finite samples, an impossibility under locally linear models.

1 Preliminaries

Let $\mathcal{M}$ be a connected manifold patch that is locally isometric to an open subset of $\mathbb{R}^d$ and embedded in ambient Euclidean space $\mathbb{R}^{D>d}$ by an unknown C^2 function; $\mathcal{M}$ is a Riemannian submanifold of $\mathbb{R}^D$ with induced metric. $\mathcal{M}$ has extrinsic curvature in $\mathbb{R}^D$ but zero intrinsic curvature. However its isometric immersion in $\mathbb{R}^d$ may have nontrivial shape with concave boundary and nonzero genus. The data matrix $\mathbf{X} \doteq [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \mathbb{R}^{D\times N}$ records the location of N points sampled from $\mathcal{M}$ in $\mathbb{R}^D$. The desired isometric immersion $\mathbf{Y}_{\text{iso}} \doteq [\mathbf{y}_1, \cdots, \mathbf{y}_N] \in \mathbb{R}^{d\times N}$ must eliminate the extrinsic curvature to recover the isometry up to rigid motions in $\mathbb{R}^d$.

Let $\mathcal{G}$ be a bipartite clique graph that associates N samples to M cliques such that the data is covered by overlapping point cliques. $\mathcal{G}$ is specified by adjacency matrix $\mathbf{M} = [\mathbf{m}_1, \cdots, \mathbf{m}_M] \in \mathbb{R}^{N\times M}$ with $M_{nm} > 0$ iff the n^{th} point is in the m^{th} clique. Usually $\mathcal{G}$ is be constructed so that for every point forms a clique with its closest neighbors; fewer cliques are sufficient as long as the graph is rigid when embedded in the target space. For most of this paper it is assumed that $\mathbf{M}$ is an indicator matrix with $M_{nm} \in \{0,1\}$. Finally, let $\mathbf{X}_m \in \mathbb{R}^{d\times k}$ contain a locally isometric parameterization of the k points in the m^{th} clique; Euclidean pairwise distances in $\mathbf{X}_m$ are equal to geodesic distances on $\mathcal{M}$. One may visualize $\mathbf{X}_m$ as a rigid k-pointed star in $\mathbb{R}^d$ whose arms (graph edges) terminate with flexible joints (graph vertices). We will first address the *global coordination problem*, which seeks a minimal-distortion assembly of these cliques into a global parameterization that is locally isometric to $\mathcal{M}$, isomorphic to $\mathcal{G}$, and globally rigid in $\mathbb{R}^d$. We then examine the local parameterization problem, which seeks to compute the clique parameterizations $\mathbf{X}_m$ from samples in $\mathbb{R}^D$.

2 Global coordination via nullspaces

GNA coordination is a kernel method in two senses of the word: It constructs a row-space immersion error matrix that, after simple tranformation, yields a kernel matrix whose elements approximate inner-products between points in the target embedding. This matrix merges the nullspaces (also known as kernels) of local coordinate neighborhoods on the manifold. GNA takes a trivial property of subspace methods—that the data lives in the nullspace of its nullspace—and generalizes this global property to a local operator.

We are interested in two spaces associated with a clique's local parameterization $\mathbf{X}_m$: $\mathbf{P}_m \doteq \text{span}([\mathbf{1}, \mathbf{X}_m^\top]) \in \mathbb{R}^{k\times(d+1)}$ is an orthogonal basis of the rowspace of $\mathbf{X}_m$ and translations thereof; $\mathbf{Q}_m \doteq \text{null}(\mathbf{P}_m^\top) \in \mathbb{R}^{k\times(k-d-1)}$ is an orthogonal basis for the complementary nullspace. Together they satisfy $\mathbf{X}_m[\mathbf{P}_m, \mathbf{Q}_m] = [\mathbf{X}_m, \mathbf{0}]$ and $[\mathbf{P}_m, \mathbf{Q}_m]^\top[\mathbf{P}_m, \mathbf{Q}_m] = \mathbf{I}$, the identity matrix. $\mathbf{P}_m$ spans the range of affine (linear plus translation) functions of $\mathbf{X}_m$; $\mathbf{Q}_m$ spans the range of nonaffine functions of $\mathbf{X}_m$ (equivalently, nonlinear functions of $[\mathbf{X}_m^\top, \mathbf{1}]^\top$). The intrinsic coordinates of manifold $\mathcal{M}$ project exclusively onto $\mathbf{P}_m$; the extrinsic curvature of $\mathcal{M}$ in the ambient space projects exclusively onto $\mathbf{Q}_m$. Thus when the local parameterizations are consistent with isometry (i.e., they can be assembled into a globally consistent parameterization $\mathbf{Y}_{\text{iso}}$ using nothing but rigid transforms), then any clique taken from $\mathbf{Y}_{\text{iso}}$ has zero projection onto the corresponding local nullspace $\mathbf{Q}_m$. Equivalently, $\mathbf{Y}_{\text{iso}}$ lies in the nullspace of a union of nullspaces: $\mathbf{Y}_{\text{iso}}[\mathbf{F}_1\mathbf{Q}_1, \cdots, \mathbf{F}_M\mathbf{Q}_M] = \mathbf{0}$, where indicator matrix $\mathbf{F}_m \in \{0,1\}^{N\times(k+1)}$ has $(\mathbf{F}_m)_{ij} = 1$ if the i^{th} data point is represented by the j^{th} point the m^{th} clique. This nullspace constraint is the basis of the error measure in Charting [4] and Linear Tangent Space Alignment (LTSA) [22], and can yield perfect immersions given perfect local parameterizations. However, if there are local parameterization errors, the least-squares problem associated with the nullspace constraint distributes error according to graph coverage, rather than evenly across the points as we would desire from a minimum squared error solution. In GNA, this is accomplished with a row-space *linearizing operator*

$$\mathbf{K} \quad \doteq \quad (\sum_m \mathbf{F}_m(\mathbf{I} - \mathbf{P}_m\mathbf{P}_m^\top)\mathbf{F}_m^\top \text{diag}(\mathbf{m}_m))\text{diag}(\mathbf{M1})^{-1}$$

that has the following properties:

Proposition 1. *(Spectral radius)* $\mathbf{K}$ *and* $\mathbf{I} - \mathbf{K}$ *are positive semidefinite with eigenvalues bounded in* $[0,1]$.

Proof. $\mathbf{K}$ is a weighted average of idempotent orthogonal projectors. □

Proposition 2. *(Linearization) The product* $\mathbf{YK}$ *isolates the component of any global parameterization* $\mathbf{Y}$ *that is not affine to the local parameterizations of* $\mathcal{M}$*, averaged over cliques and assigning equal weight to all points.*

Proof. The projector $\mathbf{Q}_m\mathbf{Q}_m^\top = \mathbf{I} - \mathbf{P}_m\mathbf{P}_m^\top$ isolates the component of $\mathbf{YF}_m$ (points in the mth clique of $\mathbf{Y}$) that is not affine to the local coordinates $\mathbf{X}_m$. For each point, the diagonal matrices then make a weighted average of these errors, averaging over cliques. □

It is then natural to define the error of any global parameterization $\mathbf{Y}$ as the Euclidean norm of its non-affine component, $\|\mathbf{YK}\|_F = \text{trace}(\mathbf{YKK}^\top\mathbf{Y}^\top)^{1/2}$.

The transform $\mathbf{X} \to \mathbf{X}(\mathbf{I} - \mathbf{K})$ attenuates the component of $\mathbf{X}$ that is locally orthogonal to $\mathcal{M}$, i.e., noise and, on the clique scale, extrinsic curvature. It is analogous to noise suppression in PCA via projection in and out of the principal subspace. Repeated transforms $\mathbf{X} \to \mathbf{X}(\mathbf{I} - \mathbf{K})^n$ smooth and unroll the data by making the data closer and closer to rank-d in larger and larger neighborhoods. Ultimately we obtain a parameterization that lies in the error-minimizing affine subspace $\mathbf{Y} = \arg\min_{\mathbf{YY}^\top=\mathbf{I}} \text{trace}(\mathbf{YKK}^\top\mathbf{Y}^\top)$, whose rows are the $d+1$ left singular vectors of $\mathbf{K}$ associated with its minimal singular values. Under perfect local isometry, this is the exact nullspace of $\mathbf{K}$. Because $\mathbf{K}$ is invariant to global translations of the immersion, the nullspace contains a nuisance constant vector $\mathbf{1}$;

we may write w.l.o.g. $\mathbf{Y}^\top = [\mathbf{1}, \mathbf{Y}_{\text{aff}}^\top]$. One of many ways to isolate $\mathbf{Y}_{\text{aff}}$ is to flip the spectrum of $\mathbf{K}\mathbf{K}^\top$ and then remove any variance associated with translations: Using centering matrix $\mathbf{T} \doteq \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$, the GNA kernel is $\mathbf{T}(\mathbf{I} - \mathbf{K}\mathbf{K}^\top)\mathbf{T} = \mathbf{T} - \mathbf{K}\mathbf{K}^\top$ and its PCA is

$$\mathbf{Y}_{\text{aff}} \doteq \arg\max_{\mathbf{Y}\mathbf{Y}^\top = \mathbf{I}} \text{trace}(\mathbf{Y}(\mathbf{T} - \mathbf{K}\mathbf{K}^\top)\mathbf{Y}^\top)$$

$\mathbf{Y}_{\text{aff}}$ is maximally affine to $\mathcal{M}$ everywhere; moreover, under the following conditions, the affine map to $\mathcal{M}$ is the same for all cliques, modulo translation:

Theorem 1. *(Coordination) If $\mathcal{M}$ allows an isometric immersion $\mathbf{Y}_{iso}$ in $\mathbb{R}^d$ upon which clique graph $\mathcal{G}$ is globally rigid, and every local parameterization $\mathbf{X}_m$ is affine to its d-dimensional patch of $\mathcal{M}$, then $\mathbf{Y}_{\text{iso}}$ lies in the subspace spanned by $\mathbf{Y}_{aff}$.*

Proof. By definition, any immersion drawn from the row-space of $\mathbf{Y}_{\text{aff}}$ is minimally non-affine to all local parameterizations $\{\mathbf{X}_m\}_m^M$ and the corresponding patches of $\mathcal{M}$ (and, indeed, perfectly affine if all $\mathbf{X}_m$ are consistent with an isometric immersion). To show that $\mathbf{Y}_{\text{aff}}$ spans an isometric immersion, recall that global rigidity implies that the affine maps taking any two cliques in $\mathbf{Y}_{\text{aff}}$ to their corresponding cliques in $\mathbf{Y}_{\text{iso}}$ are fully constrained with respect to each other. Consider any two cliques in $\mathbf{Y}_{\text{aff}}$ that share points. If their affine maps to $\mathbf{Y}_{\text{iso}}$ differ, then the kernel $\mathbf{K}\mathbf{K}^\top$ must admit an immersion which is a nonlinear function of the shared points, which is a contradiction. Therefore all cliques must share the same affine map from $\mathbf{Y}_{\text{aff}}$ to $\mathbf{Y}_{\text{iso}}$, making $\mathbf{Y}_{\text{aff}}$ globally affine to $\mathbf{Y}_{\text{iso}}$. □

Note that a manifold that is locally isometric to $\mathbb{R}^d$ does not necessarily have an embedding (a topology-preserving map from $\mathcal{M}$) or locally isometric immersion in $\mathbb{R}^d$. It may have either, neither, or both. For example, lampshades, Moebius strips, and corkscrew ramps are bounded $\mathbb{R}^3$ submanifolds that are locally isometric to $\mathbb{R}^2$, but they cannot be flattened without distortion, catastrophes (a fold), and self-crossings, respectively. GNA will embed the lampshade with smooth distortion, immerse the Moebius strip with a fold, and isometrically immerse the ramp with self-crossings. In cases such as the lampshade, the nullspace averaging in $\mathbf{K}$ offers the following assurance:

Corollary 1. *(Minimal distortion) If $\mathcal{M}$ does not have a locally isometric immersion in $\mathbb{R}^d$, $\mathbf{Y}_{\text{aff}}$ is the immersion for which, on average, each point is minimally displaced from a neighborhood configuration whose parameterization is linear in that of the corresponding patch of $\mathcal{M}$.*

The coordination may also be due to errors in overlapping local parameterization that make them inconsistent. However, because the GNA kernel averages constraints on points rather than summing constraints over cliques, GNA immersions are faithful to the data in the following sense:

Corollary 2. *(*MMSE*)* $\mathbf{Y}_{\text{aff}}$ *has minimum squared error with respect to the* average *local parameterization of each point relative to its neighbors.*

In sum, the kernel $\mathbf{K}\mathbf{K}^\top$ is distinguished in that it does not allow the embedding to be a nonlinear function of the local parameterizations (as is possible in other local methods such as LLE, HLLE, and Laplacian Eigenmaps, which employ subsets or approximations of GNA's subspace constraints), nor does it distribute distortion errors preferentially to

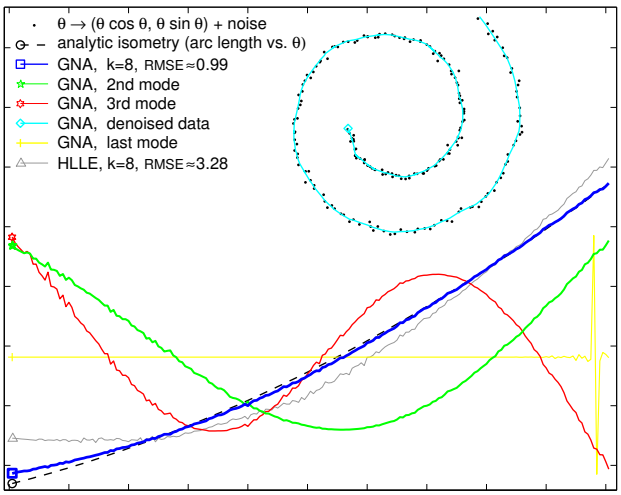

Figure 1: Denoising of samples from a 1D manifold embedded in 2D. The denoising projection $\mathbf{X} \to \mathbf{X}(\mathbf{I}-\mathbf{K})$ eliminates most of the data variance that is locally orthogonal to the manifold. The null mode (eigenvector) of $\mathbf{K}$ gives a basis for an isometric parameterization (blue); higher modes give a basis for the global shape of the ambient embedding (red and green); and the highest modes give a basis for local sample noise (yellow).

points that have low degree in the clique graph (a universal problem in the NLDR literature).

The actual isometry $\mathbf{Y}_{\text{iso}} = \mathbf{A}\mathbf{Y}_{\text{aff}}$ is found by solving for a linear shear $\mathbf{A} \in \mathbb{R}^{d\times d}$ that makes all clique parameterizations isometric to their local parameterizations on $\mathcal{M}$. $\mathbf{A}$ is a shear because isometry is invariant to rotation and translation. To factor these out, let $\overline{\mathbf{X}}_m$ be a centered local parameterization on $\mathcal{M}$ and let $\overline{\mathbf{Y}}_m$ be a set of coordinates from $\mathbf{Y}_{\text{aff}}$ for the same clique, centered and rotated into alignment with $\overline{\mathbf{X}}_m$ via Procrustes method [11, §12.4.1]. Then the least-squares estimate of the shear is (up to rotation) $\mathbf{A} = (\mathbf{X}_{\mathcal{S}}\mathbf{Y}_{\mathcal{S}}^\top)(\mathbf{Y}_{\mathcal{S}}\mathbf{Y}_{\mathcal{S}}^\top)^{-1}$ with $\mathbf{Y}_{\mathcal{S}} \doteq [\overline{\mathbf{Y}}_a\mathbf{D}_a, \overline{\mathbf{Y}}_b\mathbf{D}_b, \cdots]$, $\mathbf{X}_{\mathcal{S}} \doteq [\overline{\mathbf{X}}_a\mathbf{D}_a, \overline{\mathbf{X}}_b\mathbf{D}_b, \cdots]$ where $\mathbf{D}_m \doteq \operatorname{diag}(\mathbf{F}_m^\top\mathbf{m}_m)\operatorname{diag}(\mathbf{F}_m^\top\mathbf{M}\mathbf{1})^{-1}$ duplicates the error averaging of the kernel.

The left singular vectors of $\mathbf{K}$ are also the eigenvectors of an implied graph Laplacian[2] $\mathbf{L} = \mathbf{K}\mathbf{K}^\top$. In this view they provide a harmonic basis for deformations of the immersion: The low-frequency modes allow for large-scale curvature of the data while the high-frequency modes allow for local distortions such as noise. The left eigenvectors of $\mathbf{K}$ have essentially the same structure (see figure 1) thus thresholding the eigenvalues of $\mathbf{I}-\mathbf{K}$ to 0 and 1 produces an operator that suppresses high-frequency artifacts such as noise while preserving the ambient shape of the data.

2.1 Generalization

NLDR becomes useful to signal processing when new points can be mapped between the ambient and target spaces. Out-of-sample extensions [3] give the immersion of a new

[2] and of the matrix $\mathbf{TCT}$ of centered commute times on the associated continuous-time Markov chain.

point that is constrained by known points (but not vice versa). The basic idea is to compute a vector containing the kernel inner product of the new point with all original points, and project it onto the eigenvectors of the original immersion. The application of this idea to GNA leads to a simple formula: Let $\mathbf{K}'$ be a linearizing operator constructed as above, but for a data-subset matrix $[\mathbf{x}, \mathbf{x}_i, \mathbf{x}_j, \cdots]$, where $\mathbf{x} \in \mathbb{R}^D$ is the new point and $\{\mathbf{x}_i, \mathbf{x}_j, \cdots\} \subset \mathbf{X}$ are those original data points that belong to cliques that the new point $\mathbf{x}$ will join. $\mathbf{K}'$ is constructed using the same clique parameterization functions $f_m : \mathbb{R}^D \to \mathbb{R}^d$ as in the original $\mathbf{K}$, but these functions are also applied to $\mathbf{x}$. The immersion of $\mathbf{x}$ is

$$\mathbf{y} \in \mathbb{R}^d = -[\mathbf{0}, \mathbf{y}_i, \mathbf{y}_j, \cdots](\mathbf{K}'\mathbf{k}_1'^{\top}/\|\mathbf{k}_1'\|^2)$$

for $\{\mathbf{y}_i, \mathbf{y}_j, \cdots\} \subset \mathbf{Y}_{\text{iso}}$ and $\mathbf{k}_1'$ the first row of $\mathbf{K}'$. If $\mathbf{x}$ is assigned to just one clique $\mathbf{X}_m$, this reduces to affine regression; the row-space map taking $\mathbf{X}_m$ into the global isometric parameterization is applied to $\mathbf{x}$.

The same idea can be used for denoising novel points in the ambient space, assuming that $f_m(\cdot)$ is pseudo-invertible to yield points on a surface tangent to $\mathcal{M}$. The ambient Euclidean distance from a denoised point $\mathbf{x}'$ to $f_m^{-1}(f_m(\mathbf{x}))$ is

$$\|\mathbf{x}' - f_m^{-1}(-[\mathbf{0}, \mathbf{X}_m](\mathbf{Q}_m'\mathbf{q}_{m1}'^{\top}/\|\mathbf{q}_{m1}'^{\top}\|^2)\|$$

where $\mathbf{q}_{m1}'$ is the first row of $\mathbf{Q}_m'$, an orthogonal basis of $\text{null}([\mathbf{1}, [f_m(\mathbf{x}), \mathbf{X}_m]^{\top}])$. Just as $\mathbf{X}(\mathbf{I}-\mathbf{K})$ denoises the original data by averaging clique constraints on each point, we can average the backprojections $f_m^{-1}(\cdot)$ over all cliques containing $\mathbf{x}$ to obtain a least-squares estimate of denoised $\mathbf{x}'$. If $f_m(m)$ is an orthogonal projection then this scheme reduces to the familiar form

$$\mathbf{x}' \quad = \quad -[\mathbf{0}, \mathbf{x}_i, \mathbf{x}_j, \cdots](\mathbf{K}'\mathbf{k}_1'^{\top}/\|\mathbf{k}_1'\|^2),$$

which shows that the out-of-sample extensions for denoising and immersion are consistent under first-order assumptions. This scheme can also be used to map from the target space to the ambient space, but there is no reason to expect that it will generalize well off the boundary of the manifold—an important caution because datasets can have very complex boundaries that are hard to gauge from immersions (see figure 7).

It is also possible to treat a new point as information about the manifold that could change the immersion of all points. Taking advantage of near-linear-time updating schemes for eigenvalue decomposition (EVD) such as Lanczos rank-1 updating, points and cliques may be added or modified without requiring a complete recomputation:

Proposition 3. *Updating constraints on k points requires at most* $2k$ *rank-1 updates to the* EVD.

Proof. The new error matrix can be written $\mathbf{K}+\mathbf{J}$, where $\mathbf{J}$ has k nonzero columns corresponding to the affected points. Knowing the EVD of $\mathbf{T}-\mathbf{K}\mathbf{K}^{\top}$, we seek the EVD of $\mathbf{T}-(\mathbf{K}+\mathbf{J})(\mathbf{K}+\mathbf{J})^{\top} = \mathbf{T}-\mathbf{K}\mathbf{K}^{\top}-(\mathbf{K}\mathbf{J}^{\top}+\mathbf{J}\mathbf{K}^{\top})-\mathbf{J}\mathbf{J}^{\top}$. The span of the parenthesized summand is the combined span of the nonzero columns of $\mathbf{J}$ and the corresponding columns of $\mathbf{K}$. Therefore it has rank $2k$ at most and includes $\mathbf{J}\mathbf{J}^{\top}$ in its span.

To reduce the update to a series of rank-1 updates, the summands need to be decomposed into eigenpairs. This can be done in $O(Nk^3)$ time by orthogonalizing $[\mathbf{J}, \mathbf{K}]$ to get a subspace basis, projecting the summands into this subspace, performing an EVD of the resulting $2k \times 2k$ symmetric matrix, and using the eigenvectors to counter-rotate the basis. The resulting basis vectors and eigenvalues are then used for sequential rank-1 updates of the immersion EVD. □

2.2 Sample complexity

If $\mathbf{K}$ were constructed using just some subspace of the each local nullspace $\mathbf{Q}_m$, the immersion might not be fully determined, because $\mathbf{K}$ would also be invariant to local distortions. This leads to a simple but useful insight about clique size, which determines the dimension of the nullspace. For a d-dimensional immersion, one needs at least $k \geq d+2$ points to construct a nonempty local nullspace and a further $\binom{d+1}{2} - 1$ points for an estimator of the local Hessian to be contained in the span of the nullspace, for a total of $k \geq \binom{d+2}{2}$ points. Thus to totally eliminate nonlinearities up to second order, *any* local NLDR method that compares prospective immersions to local parameterizations will require clique sizes of $k = O(d^2)$ points.

This does not exclude the possibility that using fewer points or incomplete nullspace constraints will lead to an immersion with low distortion, because rigidly overlapping cliques are generally subject to the union of their constraints. GNA works quite well on data manifolds with dense clique coverage even when $k = d+2$. The quadratic forms defined by locally linear embedding (LLE) [15] and Hessian LLE (HLLE) [8] are locally invariant to $k-1$ and $k-\binom{d+2}{2}$ local distorting degrees of freedom, respectively, that provide room for errors to manifest in the final immersion, yet heavy overlap eliminates most of these errors. GNA immersions, however, are better constrained when the manifold is sparsely covered with fewer but larger cliques, a useful strategy for reducing complexity.

2.3 Point weighting

Error can creep into immersions through sampling noise, numerical error, and local parameterization errors. The GNA operator has a nice property in this respect: The error associated with perturbing a point declines quadratically with its distance from the center of each clique it participates in. In particular, consider the unweighted nullspace projection error $\|\mathbf{Y}\mathbf{F}_m\mathbf{Q}_m\mathbf{Q}_m^\top\|_F$ of global parameterization $\mathbf{Y}$ with respect to the m^{th} clique. Let immersion point $\mathbf{y}_i \in \mathbf{Y}$ correspond to clique point $\mathbf{x}_i \in \mathbf{X}_m$. The error associated with perturbing either point declines quadratically with the distance of $\mathbf{x}_i$ from the clique center, denoted $\overline{\mathbf{x}_m}$:

Proposition 4. *Perturbations of* $\mathbf{y}_i$ *or* $\mathbf{x}_i$ *cause* $\|\mathbf{Y}\mathbf{F}_m\mathbf{Q}_m\mathbf{Q}_m^\top\|_F$ *to vary as* $0 < a - b\|\mathbf{x}_i - \overline{\mathbf{x}_m}\|^2$ *for some constants* $a, b > 0$ *independent of* i.

Proof. Let $\mathbf{P}_m$ be an orthogonal basis of the columnspace of $[\mathbf{1}, \mathbf{X}_m^\top]$. W.l.o.g., let the first column of $\mathbf{P}_m$ be constant. By orthogonality, all other columns must sum to zero and are thus linear transforms of centered $\mathbf{X}_m$. Consequently values in the ith row of $\mathbf{P}_m$ are linear in $(\mathbf{x}_i - \overline{\mathbf{x}_m})$, and the ith element on the diagonal of projector $\mathbf{P}_m\mathbf{P}_m^\top$ is quadratic in $\|\mathbf{x}_i - \overline{\mathbf{x}_m}\|$. The norm of the inner product of $\mathbf{Y}$ with $\mathbf{Q}_m\mathbf{Q}_m^\top = \mathbf{I} - \mathbf{P}_m\mathbf{P}_m^\top$ therefore varies linearly with $-\|\mathbf{x}_i - \overline{\mathbf{x}_m}\|^2$. Small perturbations $\Delta\mathbf{x}_i$ of $\mathbf{x}_i$ cause the norm to vary with $-\|\mathbf{x}_i + \frac{k-1}{k}\Delta\mathbf{x}_i - \overline{\mathbf{x}_m}\|^2 \approx -\|\mathbf{x}_i - \overline{\mathbf{x}_m}\|^2$. □

Thus nullspace projectors are naturally more tolerant of error at the periphery of a clique, and immersions are determined more by central points than by peripheral points.

That said, there are conditions under which such tolerance is not enough. For example, if the manifold is locally a second-order algebraic surface (parabolic, hyperbolic, or elliptic), then the error in the parameterization by linear projection onto the tangent space estimate grows as $O(\|\widehat{\mathbf{x}_i - \overline{\mathbf{x}}}\|^3)$, with $\|\widehat{\mathbf{x}_i - \overline{\mathbf{x}}}\|$ being the distance of $\mathbf{x}_i$ to the clique

mean in the tangent space. Thus for locally linear models one may profitably adjust the point-clique weights M_{mn} to further discount errors associated with peripheral points.

3 Local isometric parameterizations

NLDR uses Euclidean distances in the ambient space as a proxy for geodesic distances on the manifold, but this is a biased approximation that always underestimates true distances. Where ever a manifolds curve away from their tangent spaces, this locally linear view of the manifold induces a "fish-eye" distortion that causes the global parameterization to contract in places and directions where the manifold has high extrinsic curvature. With finite data, it is impossible to define a clique size that eliminates this distortion, and many NLDR algorithms require large cliques (for rigidity or stability) that exacerbate the problem.

Local distortions can be substantially reduced and sometimes eliminated entirely by modelling clique curvature with algebraic surfaces of degree two. We show that these can be fitted to data to yield exact isometric parameterizations on a nontrivial class of quadric manifolds—those that are products of planar quadrics (PPQ), i.e., manifolds defined as products of parabolic, elliptic, hyperbolic, and straight plane curves. For example, generalized cylinders and minimal isometric immersions of d-torii are locally PPQ embeddings of intrinsically flat manifolds. When a manifold is not locally PPQ, the PPQ model is essentially a mixed first- and second-order approximation, the second-order terms being fitted in the directions of where the manifold exhibits greatest curvature. For example the benchmark NLDR "Swiss roll" problem [18, 15, 4, 8] is the product manifold of an Archimedes spiral and a line segment; to second order the spiral $f(\theta) = (\theta\cos\theta, \theta\sin\theta)$ is elliptic for $\theta < \pi/2$, parabolic at $\theta = \pi/2$, and hyperbolic for $\theta > \pi/2$.

Consider a local neighborhood around point $\mathbf{p} \in \mathcal{M} \subset \mathbb{R}^D$ in which the ambient embedding of d-dimensional $\mathcal{M}$ is locally quadric and has extent in a $2d$-dimensional affine subspace spanned by an orthogonal basis $\mathbf{T_p} \in \mathbb{R}^{D\times 2d}$. Clearly such a neighborhood exists and supersets the infinitesimal neighborhood in which $\mathcal{M}$ is locally linear around $\mathbf{p}$. In this neighborhood, $\mathcal{M}$ can be fitted by a quadric hypersurface $\mathcal{Q}_\mathbf{p} \subset \mathbb{R}^D$ of dimension $2d-1$ having matrix equation $\mathbf{x'}^\top \mathbf{F_p} \mathbf{x'} = 0$ for symmetric $\mathbf{F_p}$ and local homogeneous co-

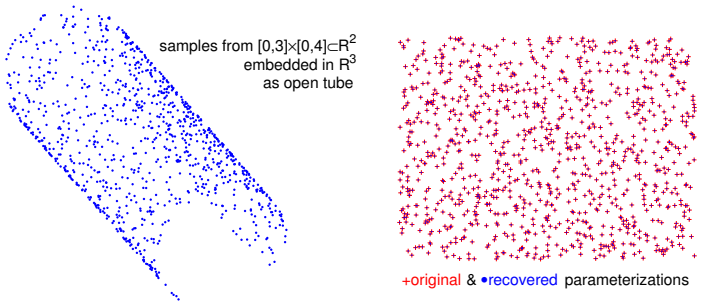

Figure 2: The GNA/PPQ parameterization of points sampled from this cylinder-shaped manifold is exact (up to numerical limits of floating-point square-root calculations) for any rigid clique graph having sufficient samples in each clique to determine a quadric. This includes sparse cliques, nonlocal cliques, and a graph consisting of a single clique.

ordinates $\mathbf{x}' \doteq [(\mathbf{x}-\mathbf{p})^\top \mathbf{T}_\mathbf{p}, 1]^\top \in \mathbb{R}^{2d+1}$. Strictly speaking, in this neighborhood $\mathcal{M}$ is a submanifold of $Q_\mathbf{p}$. Of practical import is the fact that the surface $Q_\mathbf{p}$ is a good local approximation of $\mathcal{M}$ over a much larger area than any linear model. The main result is that $\mathcal{M}$'s local PPQ decomposition and isometric parameterization can be determined from samples:

Theorem 2. *(PPQ) If d-dimensional $\mathcal{M}$ is locally a product of planar quadrics, then $\mathbf{F}_\mathbf{p}$ and its PPQ decomposition are recoverable with probability 1 from $O(d^2)$ random samples spanning $\mathbf{T}_\mathbf{p}$.*

The essence of the constructive proof is that $\mathbf{F}_\mathbf{p}$ has a canonical form that reveals the pairing of dimensions into planar quadrics. The planar quadrics can then be independently integrated for arc-length, giving a true geodesic parameterization.

Proof. (PPQ theorem 2) Empirically, a quadric is fitted to multivariate data via a matrix $\mathbf{B}$ containing the scatter of all the pairwise products of the ordinates of $\mathbf{x}'$: $B_{pq} = \sum_i x'_{pi} x'_{qi}$ where x'_{pi} is the p^{th} element of homogeneous point $\mathbf{x}'_i$. The minimizing eigenvector of symmetric nonnegative definite $\mathbf{B}$ contains the elements of quadric equation coefficient matrix $\mathbf{F}_\mathbf{p}$, while the associated eigenvalue gives the sum squared error of the fit, such that $\sum_i \mathbf{x}_i'^\top \mathbf{F}_\mathbf{p} \mathbf{x}'_i = \lambda_{\min}(\mathbf{B})$. Since this is a linear system of $\binom{2d+1}{2}$ unknowns, the sample complexity is $O(d^2)$. A partial EVD diagonalization $\begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}^\top \mathbf{F}_\mathbf{p} \begin{bmatrix} \mathbf{V} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \text{diag}(\mathbf{a}) & \mathbf{b}/2 \\ \mathbf{b}^\top/2 & c \end{bmatrix}$ gives a subspace rotation inside $\mathbf{T}_\mathbf{p}$ such that Q is expressed as a sum of quadratic functions $q_j(\cdot)$, one in each dimension:

$$\mathbf{x}'^\top \mathbf{F}_\mathbf{p} \mathbf{x}' = \sum_j^{2d} q_j(z_j) + c \doteq (\sum_j^{2d} a_j z_j^2 + b_j z_j) + c = 0, \tag{1}$$

where $\mathbf{z} \doteq [z_1, \cdots z_{2d}]^\top = \mathbf{V}^\top \mathbf{T}_\mathbf{p}^\top (\mathbf{x}-\mathbf{p})$ for $a_j \in \mathbf{a}$ and $b_j \in \mathbf{b}$. This diagonalization is unique up to possible multiplicity in the eigenvalues $a_j \in \mathbf{a}$. First consider the case where all eigenvalues are distinct. If $\mathcal{M}$ is locally PPQ, then any one of its constituent planar curves must be expressed as the linear combination of two of Q's quadratic summands, e.g., $r_i q_i(z_i) + q_j(z_j) + c_j = 0$ for $r_i \in \mathbb{R}\backslash 0$ and $c \in \mathbb{R}$ because the EVD gives the only orthogonal basis that eliminates all pairwise products $z_i z_j$ that couple the dimensions multiplicatively. In equation 1 all of the dimensions are coupled additively but if $\mathcal{M}$ is locally PPQ then the $2d$ quadratic summands in equation 1 must be paired off to form d orthogonal independent quadrics, each specifying a 1D curve in an $\mathbb{R}^2$ subspace spanned by a pair of the eigenvectors in $\mathbf{V}$. Therefore $\mathcal{M}$ is locally a submanifold of Q. The pairing can be determined from data because the vector of values of each summand q_i, iterated over all points, is a 1D affine transform of the value vector of its paired summand q_j. A linear regression to find satisfying pairs is fully determined when the summand vectors are linearly independent in $\mathbb{R}^{2d}$, requiring at least $2d$ points.

To handle eigenvalue multiplicities, recall that the m eigenvectors $\{\mathbf{v}_i, \mathbf{v}_j, \cdots\} \subset \mathbf{V}$ associated to a repeated eigenvalue $\{a_i = a_j = \cdots\} \subset \mathbf{a}$ are determined only up to an $\mathbb{R}^m$ mutual rotation. Local PPQ structure of $\mathcal{M}$ implies that there exists a rotation of these eigenvectors and the linear coefficients $\{b_i, b_j, \cdots\}$ such that the associated quadratic summands can be paired off with each other or with summands of other eigenvalues. There are four causes and resolutions of multiplicity:

1. Each parabola contributes an $a_j \neq 0$ for its abscissa and an $a_i = 0$ for its ordinate, thus ordinates of all parabolas are rotated together in the nullspace. The rotation can be recovered by regressing the corresponding summand value vectors onto those of nonzero a_j.

2. Each independent linear dimension contributes a summand with $a_i = b_i = 0$. These dimensions constitute the remainder of the nullspace after parabolic dimensions are removed via pairing.

3. Each perfect circle contributes a pair $a_i = a_j \neq 0$. Being a circle, the pairing is revealed by the multiplicity and invariant to rotation of the associated eigenvectors.

4. An accidental multiplicity, with $a_i = a_k$ being parameters from two independent planar curves, is possible because each planar pair can be arbitrarily weighted in the quadratic form $\mathbf{F}$. This implies that scatter matrix $\mathbf{B}$ has a d-dimensional null space of equivalent $\mathbf{F}$ parameterizations. In that space the subset of parameterizations having accidental multiplicities has dimension $< d$ and thus measure zero. Therefore the pairing is fully determined with probability 1. □

PPQ manifolds have straightforward isometric parameterizations: Because a locally product manifold is locally isometric to the product of its totally geodesic submanifolds, a PPQ manifold is isometric to the product space of arc-length parameterizations of each quadric (see [16] for stronger results). If $\mathcal{M}$ is globally PPQ (e.g., an open cylinder in $\mathbb{R}^3$), then this procedure can parameterize the entire manifold; see figure 2. Parabolas and circles offer straightforward analytic formulae for arc-length parameterizations; ellipses and hyperbolas do not but that problem is mooted in practice by robust and efficient numerical solutions for their arc-length integrals. In our setting, there is no ambiguity over which branch of a hyperbola or side of an ellipse to integrate for arc-length because $\mathcal{M}$ is locally connected and isometric to a Euclidean space.

Not all intrinsically flat manifolds are locally PPQ; e.g., a cone in $\mathbb{R}^3$ is not a product manifold. In fact the PPQ decomposition is trivially extended to cones by considering all triplets of quadratic summands from diagonalized $\mathbf{F_p}$; a cone is diagnosed as three summands that have a constant empirical sum and quadratic coefficients of varied sign. The decomposition remains well-defined because cones do not present any additional source of eigenvalue multiplicity. The remaining class of developable surfaces in $\mathbb{R}^3$—tangent surfaces of space curves—are not quadric but have piecewise conic approximations that are sufficiently accurate for industrial applications [5].

The PPQ parameterization is exact with data sampled from conforming manifolds. Although offered mainly in service of a theoretical argument, it also appears to work well with nonconforming manifolds and in low-noise settings. However one may expect that PPQ fitting will should be vulnerable to noisy data because it works by removing excess degrees of freedom from a flexible data model, and the minimized error is algebraic, not geometric. As the level of noise increases, it is prudent to retreat to more robust procedures that impose simpler models on the data, trading model expressivity for stability of estimated parameters to data perturbations. For example, if $\mathcal{M}$ is locally PPQ, then in a slightly smaller neighborhood around $\mathbf{p}$, each of its constituent planar curves has a good second-order (or exact) parabolic fit. I.e., there is a set of pairings $a_i z_i^2 + b_i z_i + c_i \approx b_j z_j + c_j$ with z_i an ordinate in some orthogonal basis of the tangent space, z_j an ordinate in some orthogonal basis in the normal space, and equality for linear

or parabolic submanifolds. This product of planar parabolics (PPP) model can be fitted to data by partitioning the local data space into an $\mathbb{R}^d$ tangent space and an $\mathbb{R}^d$ normal space (e.g., via local PCA), estimating Hessians in each of the normal dimensions, then solving for a rotation in the tangent space that diagonalizes a matrix whose columns are the diagonals of the Hessians. This is possible for all smooth Riemannian submanifolds of $\mathbb{R}^D$ having flat normal bundles [7, p. 137]. Solving for a rotation of the normal space that diagonalizes a matrix of all these diagonals gives the quadratic coefficients a_i. The remaining linear (b_i) and constant (c_i) terms can be recovered via least-squares linear fits. If the manifold is not locally PPP w.r.t. the tangent space, the second diagonalization will be approximate; any non-vanishing off-diagonal elements are the quadratic coefficients of curvatures that are ignored in the PPP fit.

Figure 3 compares this parameterization to a local PCA. The data randomly sampled from a patch of a $\mathbb{T}^3$ torus embedded isometrically in $\mathbb{R}^6$ and contaminated with 2% isotropic gaussian noise. The patch subtends $\pi/10$ radians, is *not* PPP, and is nearly flat in $\mathbb{R}^6$. The PPP geodesics have 1/3 the error of the PCA parameterization. One can see in the scatter plot that the PCA points are visibly offset from the true point locations all around the periphery of the patch.

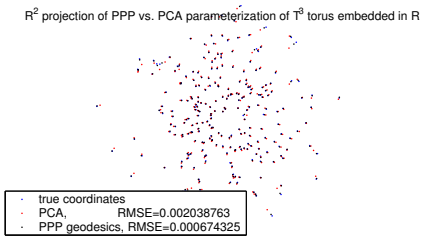

Figure 3: PPP & PCA parameterization errors.

4 Examples

Figure 4 compares the isometry errors of GNA, LTSA, HLLE, and SDE on a spiral problem taken from the SDE distribution demo. All methods compute an embedding that approximates the analytic arc-length parameterization, but GNA has an order of magnitude less error than LTSA and two orders of magnitude less error than HLLE and SDE, as measured in both RMSE and subspace angle between each method's embedding eigenvector and the perfect isometric embedding. The GNA is computed from local PPP parameterizations; if computed from full PPQ parameterizations (which require numerical integration) the error decreases by almost an order of magnitude, while if computed from linear tangent parameterizations the error increases to roughly half that of LTSA.

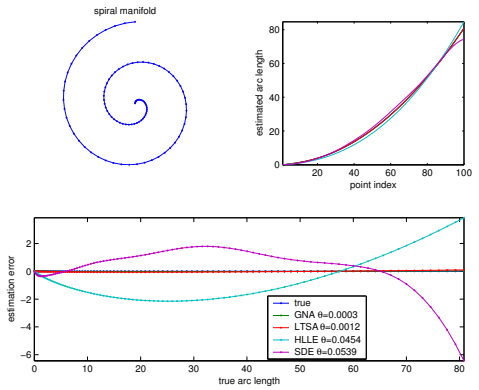

Figure 4: Spiral embedding errors of GNA, LTSA, HLLE, and SDE.

Figure 5 compares the performance of the GNA and HLLE on the "perforated swiss roll" demonstration problem generated by test codes distributed with HLLE. This is known to be problematic for Isomap and LLE. HLLE itself fails catastrophically on a substantial fraction of first 500 test examples generated; GNA performs robustly. It appears that in the HLLE failures the Hessian constraint admits enough local distortions to cause the manifold

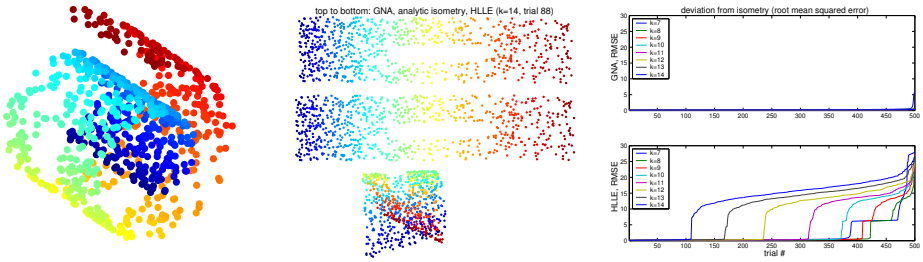

Figure 5: The perforated swiss roll. Left to right: The ambient embedding; parameterizations offered by GNA, analytic mapping, and HLLE (showing a catastrophic fold); and isometry error curves on 500 randomly generated examples for each clique size, sorted by HLLE error.

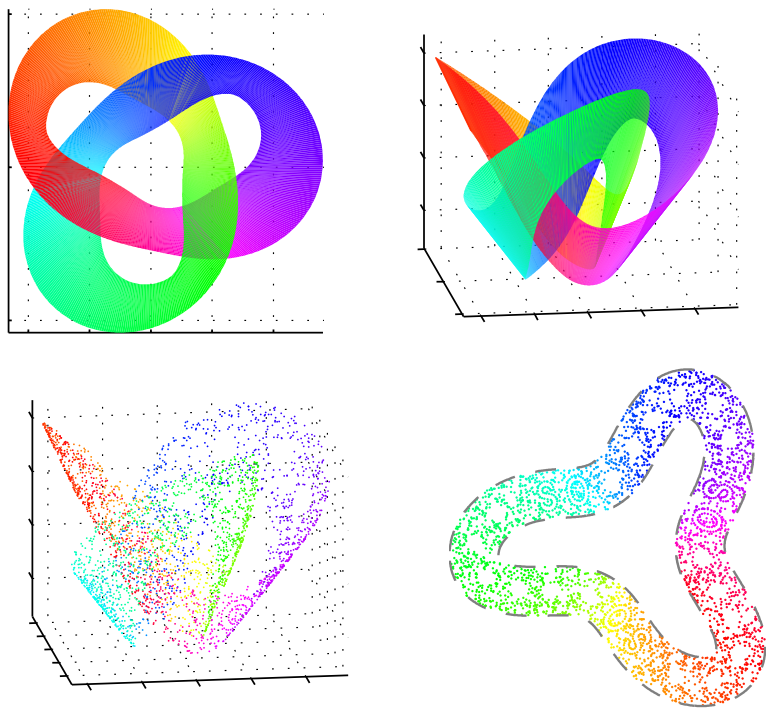

Figure 6: Parameterization of points randomly sampled from a trefoil-knotted ribbon manifold. At left, two views of the manifold's ambient embedding in $\mathbb{R}^3$. At right, points sampled from the manifold and their isometric GNA embedding in $\mathbb{R}^2$. Fine structure in the sampling pattern is preserved with very high fidelity. The outlines show the edges of the $\mathbb{R}^2$ pre-image of the ribbon manifold.

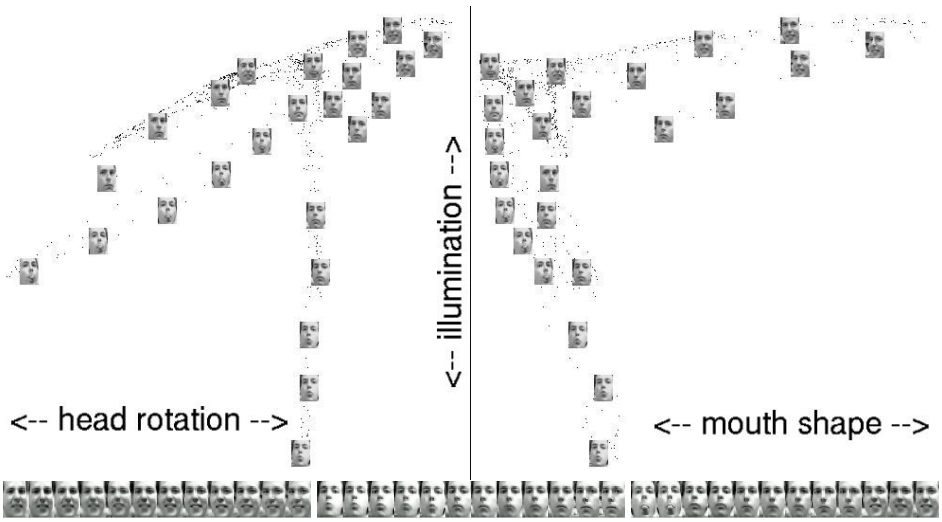

Figure 7: Front and side views of a GNA immersion of 1000 20×28 face images into $\mathbb{R}^3$ with $k = 30$. A random subset of points are rendered as images. The degrees of freedom appear to be head rotations, brightness-changing head tilts, and changes in expression. Below are images synthesized along those degrees of freedom.

to remain partly curved in the eigenbasis; the 2D projection then appears as a fold. Where HLLE succeeds, GNA reduces isometry errors by any average of 5% to 98%, depending on neighborhood size.

Figure 6 shows a manifold formed by sweeping a line segment along trefoil-knot trajectory in $\mathbb{R}^3$. The resulting manifold is intrinsically flat but not locally PPQ. Like all benchmark problems in NLDR, the geometry of this $\mathbb{R}^3$ embedding is known and an exact isometric pre-image in $\mathbb{R}^2$ can be deduced. We will compare this known isometry with the estimate made by GNA from points randomly and noisily sampled from the manifold, excluding some circular and letter-shaped regions. The GNA embedding reconstructs the isometric pre-image with errors averaging less than 10^{-4} times the diameter of the pre-image, i.e., if the pre-image were a meter across, errors would average $<$ .1mm. This reflects the precision limit of second-order methods under floating point calculations; i.e., the error is of the same magnitude as would be expected from numerical round-off after two square-root operations. The GNA solution is stable over a large range of clique sizes ($k = 7$ through 24), and shows gentle deformation outside this range as cliques become large enough that the manifold's non-PPQ curvature becomes significant, or small enough that sample noise becomes significant.

Figure 7 depicts a GNA immersion of 1000 frames of cropped facial images from a video. The data is almost certainly not manifold, but many researchers have noted that it immerses well in $\mathbb{R}^3$. The data has a fairly subtle boundary structure with what look like 1D tendrils, because the face exercises its extreme degrees of freedom separately. The data may be a fair sample of the density on the face manifold, but it not a fair sample of the geometry of the manifold itself. This is a chronic problem for NLDR, because other than

isometry constraints at their shared root, the data offers no information for coordinating the intrinsic degrees of freedom of two tendrils. Nonetheless, the GNA coordinates do organize the images roughly by head rotation, head illumination, and expression at the mouth. The figure shows synthetic image sequences corresponding to immersion-space trajectories along those axes. The images are synthesized by using the out-of-sample extension when the $\mathbb{R}^3$ control point is near immersed data points, and gaussian averaging when the control point is in open space.

5 Discussion

This paper developed a local principle for recovering maximally isometric parameterizations of curved manifolds from finite sampling. The local phase of GNA extends NLDR by opening up a significant class of curved data manifolds whose exact isometry can be recovered from samples. The global phase brings to NLDR a globally optimized piecewise generalization of row-space PCA, along with its guarantee of balanced error and its facility for data denoising, updating, and mappings of new data points. In this regard it offers most of the functionality of linear subspace methods, and it is easy to envision some subset of the methods described above being plugged directly into existing signal processing pipelines.

Global coordination takes $O(Nkd + Mkd^2)$ time using the thin eigenvalue decomposition (EVD). With GNA one can have fewer but larger cliques, such that $M \ll N$. The quadric parameterization phase takes $O(M(kDd^2 + d^4))$ time. In general, the dominating cost of graph-based NLDR is that of the graph itself, which takes $O(MN(D + \log N))$ time to construct using direct methods, approximations may offer substantial improvements.

Like all NLDR methods, GNA assumes a reasonable prior estimate of the dimensionality and clique size of a data set. The literature surrounding dimensionality estimation is quite sophisticated (see [13, 14, 17, 21]), but all methods depend on guesses about clique size or neighborhood scale. Of course, the intrinsic dimension may be smaller than the minimal dimension d allowing an embedding or isometric embedding in $\mathbb{R}^d$. For signal-processing applications, preserving local isometry is probably more important than preserving global topology, but for data visualizations, topologically correct embeddings may turn out to be more useful than isometric immersions. This points to an interesting trade-off between topological and metric constraints that has yet to be characterized.

References

[1] Mukund Balasubramanian and Eric L. Schwartz. The IsoMap algorithm and topological stability. *Science*, 295(5552):7, January 2002.

[2] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. volume 14 of *Advances in Neural Information Processing Systems*, 2002.

[3] Y. Bengio, J.F. Paiement, and P. Vincent. Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering. In *Advances in Neural Information Processing Systems*, volume 15, 2003.

[4] Matthew Brand. Charting a manifold. In *Advances in Neural Information Processing Systems*, volume 15, 2003.

[5] H.Y. Chen, I.K. Lee, S. Leopoldseder, H. Pottmann, T. Randrup, and J. Wallner. On surface approximation using developable surfaces. *Graphical Models and Image Processing*, 61:110–124, 1999.

[6] Fan R.K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. American Mathematical Society, 1997.

[7] Manfredo Perdigao do Carmo. *Riemannian Geometry*. Birkhauser, 1992.

[8] David L. Donoho and Carrie Grimes. Hessian eigenmaps. *Proceedings, National Academy of Sciences*, 2003.

[9] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematics Journal*, 23:298–305, 1973.

[10] Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czech. Math. Journal*, 25:619–633, 1975.

[11] Gene Golub and Charles van Loan. *Matrix Computations*. Johns Hopkins, third edition, 1996.

[12] J. Ham, D.D. Lee, S. Mika, and B. SchÃ¶lkopf. A kernel view of the dimensionality reduction of manifolds. In *Proc. ICML04*, 2004.

[13] D.R. Hundley and M.J. Kirby. Estimation of topological dimension. In *Proc. Intl. Conf. on Data Mining*. SIAM, 2003.

[14] B. Kegl. Intrinsic dimension estimation using packing numbers. In *Advances in Neural Information Processing Systems*, volume 15. MIT Press, 2003.

[15] Sam T. Roweis and Lawrence K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 22 2000.

[16] X. Senlin and N. Yilong. Submanifolds of product riemannian manifold. *Acta Mathematica Scientia*, 20(B):213–218, 2000.

[17] J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Review*, 2004. Submitted.

[18] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, December 22 2000.

[19] W.T. Tutte. Convex representations of graphs. *Proc. London Mathematical Society*, 10:304–320, 1960.

[20] W.T. Tutte. How to draw a graph. *Proc. London Mathematical Society*, 13:743–768, 1963.

[21] K. Q. Weinberger, F. Sha, and L. K. Saul. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proc. 21st ICML*, 2004.

[22] Z. Zhang and H. Zha. Nonlinear dimension reduction via local tangent space alignment. In *Proc., Conf. on Intelligent Data Engineering and Automated Learning*, number 2690 in Lecture Notes on Computer Science, pages 477–481. Springer-Verlag, 2003.