

## Regularized Principal Component Analysis\*

Yonathan AFLALO<sup>1</sup>    Ron KIMMEL<sup>2</sup>

(Dedicated to Haim Brezis on the occasion of his 70th birthday)

**Abstract** Given a set of signals, a classical construction of an optimal truncatable basis for optimally representing the signals, is the principal component analysis (PCA for short) approach. When the information about the signals one would like to represent is a more general property, like smoothness, a different basis should be considered. One example is the Fourier basis which is optimal for representation smooth functions sampled on regular grid. It is derived as the eigenfunctions of the circulant Laplacian operator. In this paper, based on the optimality of the eigenfunctions of the Laplace-Beltrami operator (LBO for short), the construction of PCA for geometric structures is regularized. By assuming smoothness of a given data, one could exploit the intrinsic geometric structure to regularize the construction of a basis by which the observed data is represented. The LBO can be decomposed to provide a representation space optimized for both internal structure and external observations. The proposed model takes the best from both the intrinsic and the extrinsic structures of the data and provides an optimal smooth representation of shapes and forms.

**Keywords** Laplace-Beltrami operator, Principal component analysis, Isometry

**2000 MR Subject Classification** 62H25, 54C56

### 1 Introduction

One of the most popular procedures used for data representation is the principal component analysis (PCA for short). Based on recent results reported in [2] that proclaim the optimality of eigenfunctions of the Laplace-Beltrami operator (LBO for short), we revisit the construction of PCA for geometric structures and propose to redesign one of the fundamental tools in information analysis. Provided smoothness of a given data entity, we could exploit the intrinsic geometric structure to regularize the construction of a basis by which the observed data is represented. The same LBO that when it is applied to the coordinates of a surface produces the mean curvature vector, can be decomposed to provide a representation space optimized for both internal structure and external observations. As a motivational sketch, the reconstruction

of the horse model , given the training set  $\left\{ \begin{array}{c} \text{img alt="A green horse model" data-bbox="503 724 556 756"} \\ \text{img alt="A green horse model" data-bbox="568 724 621 756"} \end{array} \right\}$ , is demonstrated in

Figure 1. The reconstruction is computed by projecting the coordinates of the model onto the

---

Manuscript received November 12, 2015. Revised August 16, 2016.

<sup>1</sup>Samsung Research Labs, Ramat Gan, Israel.

<sup>2</sup>Computer Science Department, Technion, Israel Institute of Technology, Haifa, Israel.

E-mail: ron@cs.technion.ac.il

\*This work was supported by the ERC Advanced Grant 267414.

first 100 eigenvalues of the PCA (left), the LBO (middle), and the suggested hybrid measure (right). In this case, the proposed hybrid model best represents the shape of the desired model. It allows the horse to rotate its head, a pose that does not exist in the training set, and thus could not have been realized by a convex combination of its decomposition captured by the PCA basis.

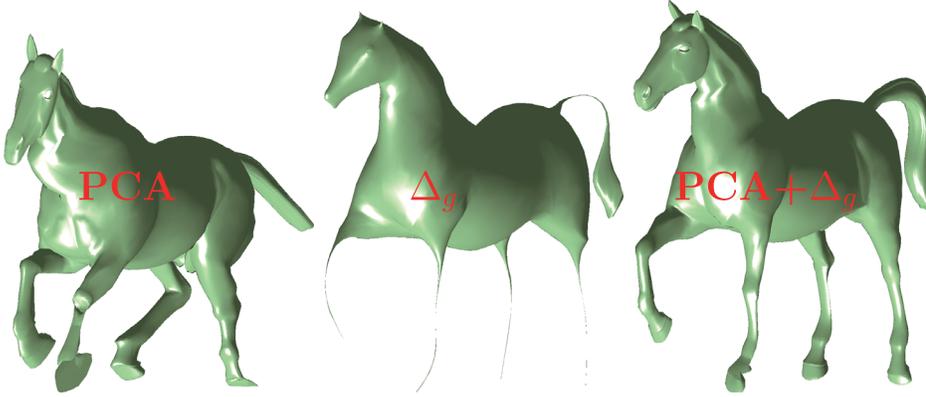


Figure 1 The reconstruction of a surface of a horse by projecting its coordinates onto the first 100 PCA eigenvectors (left) extracted from training on two horses, on the LBO eigenvectors (middle) extracted from one of the training horses, and the proposed hybrid (right).

## 2 Ubiquitous LBO

We start with the simple example of a scalar periodic function  $f$  defined on a closed curve  $C : [0, L) \rightarrow \mathbb{R}^2$  in the plane, such that one could write  $f : [0, L) \rightarrow \mathbb{R}^1$ . Here,  $L$  is the length of the planar curve, and  $s \in [0, L)$  denotes the arclength parametrization along the curve  $C(s) = \{x(s), y(s)\}$ . The second order derivative of  $f(s)$  can be written as  $f_{ss}(s) = \partial_{ss}f(s)$ . If instead of a general  $f$ , we differentiate the curve's coordinates, we obtain  $\partial_{ss}C(s) = C_{ss}(s) = \{x_{ss}(s), y_{ss}(s)\} = \kappa(s)\vec{n}(s)$ , where  $\kappa$  is the curvature of the curve, and  $\vec{n}$  its normal. Next, consider the operator  $\partial_{ss}$  in its discrete setting along the curve. That is, assume uniform sampling of  $n$  points along the closed curve, such that the operator is approximated by

$$\frac{\partial^2}{\partial s^2} \approx \left(\frac{n}{L}\right)^2 \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 1 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \quad (2.1)$$

The discrete operator is a circulant matrix which is diagonalized by the discrete Fourier transform. In other words, the eigenvectors of the matrix are the sine and cosine functions uniformly sampled along the closed curve. It appears that ordering these eigenvectors in an ascending order by their corresponding eigenvalues and picking only the first vectors as the representation

sub-space, one obtains an optimal subspace for representing smooth functions defined on the curve. In signal processing terminology, we say that a low pass filter applied to the family of functions with bounded gradient magnitude, is optimal in terms of representation error. In approximation theory terms, projecting bounded gradient functions to the leading eigenvectors is optimal in a worst case  $L_2$  sense.

The second derivative operator that generated the curvature and provided the Fourier transform as an efficient representation space, can be extended to more interesting domains. The second order derivative of a scalar function in a flat domain is known as the Laplacian. Its extension to curved domains is the Laplace-Beltrami operator or LBO.

The influence of the Laplace-Beltrami operator can be found in all modern fields of science. Its eigenspace is used to analyze and represent structures in graph theory, geometry, chemistry, biology, physics and machine learning. Its action on the coordinates of a surface produces the mean curvature vector, while its decomposition as an operator defined for regular sampled periodic scalar functions yields the vectors that compose the Fourier transform. Here, we exploit some of the LBO nice properties in order to accurately and efficiently represent signals in a way that goes beyond the classical convex hull of the given observation set. To that end, we revisit the notion of principal component analysis (PCA for short) in data understanding and couple it with a regularization term in an attempt to construct a natural orthonormal low dimensional representation space. We relate our construction to the Laplace-Beltrami operator that is known to provide an optimal basis for smooth functions on manifolds (see [2]). The efficiency of the proposed space is demonstrated to be a compact representation for articulated objects. The surface coordinates at one end, captured by the classical PCA, and the intrinsic geometric structure of the objects expressed by the metric at the other end, support one another in reconstructing the object from a small number of projections. Out-of-sample poses are encapsulated by the metric, while the fine details of the shapes are handled by the principal components. It is worth noting that a similar motivation led the authors of [17] to construct a different numerical solver to a dual regularized PCA problem, while different (non-geometric) smooth PCA methods can be found, for example, in [28]. For further reading on regularization methods using the LBO for other problems, see [19, 21–22, 31].

### 3 History in a Brief

A classical problem in machine learning is one in which the data is to be represented by a linear combination of a small set of vectors that belong to a specific orthonormal basis. The most common approach in data representation is probably the principal component analysis method known as PCA (see [18]). Related to the PCA, we have multidimensional scaling (MDS for short) methods (see [7]) and generalized MDS (GMDS for short) (see [9]). Observing the data from a different geometric perspective, local relations between data points can construct a manifold for which a Laplacian could be defined. The spectral domain of such a Laplacian could serve for data representation. We focus on a specific data representation application, namely shape reconstruction, that would allow us to demonstrate the regularized PCA procedure.

Reviewing all existing shape representation techniques is obviously beyond the scope of this paper yet, it is interesting to mention Gotsman and Karni [20] who used the LBO spectral domain to approximate coordinates of shapes, while Lévy [23] filtered the spectral domain to deform shapes. Eigenfunctions and eigenvalues of the LBO are often used to approximate diffusion distances on manifolds (see [6, 11–13, 27]). The spectral domain provides an efficient way to construct shape descriptors like the scale-space representation (see [35]), the heat kernel signature (HKS for short) (see [33, 15]), the wave kernel signature (WKS for short) (see [3]), and the global point signature (GPS for short) (see [30]). The eigenfunctions were also used for dimensionality reduction for object recognition, as demonstrated in [29].

Traditional signal representation techniques resort to low pass filters in the Fourier domain that are classically justified through statistical reasoning like the optimality of the Karhunen-Loève transform that coincides with the Fourier in some cases. Related stochastic geometry processing reasoning motivated Ben-Chen et al. [5] to support the eigenfunctions of the LBO as a convenient representation space. Recently, with Brezis, we have been able to prove that the LBO spectral domain is optimal for the representation of smooth functions on the manifold (see [2], where a deterministic optimality proof based on the Courant-Fischer min-max principle in [34], and Problems 37 and 49 in [8], was provided). The LBO eigen-space was also suggested by Belkin and Niyogi [4] to represent information about data manifolds. The number of required eigenfunctions was estimated in [36] (see [32, 16, 14] for related approaches).

The generalization of Fourier produced by the LBO handles optimally signals of bounded gradients. In some cases, more information about the data we would like to represent is provided, say, samples of typical signals. Assuming there is a low rank description to the information, we could use the observations to estimate a low dimensional representation space. Technically, instead of considering all possible periodic functions, assuming that we are given  $k$  observation  $\{f_i\}_{i=1}^k$ , where  $f_i : [0, L] \rightarrow \mathbb{R}^1$ , we represent the space of functions which we are interested in. Next, denote by  $x_i(j) = f_i(\frac{jL}{n})$  the sampled  $f_i(s)$  at equally spaced  $n$  points.

When dealing with observations of discrete data, the most popular tool for finding a low rank representation space is probably the PCA algorithm. It finds an orthonormal basis of  $m$  vectors  $P_j \in \mathbb{R}^n$  denoted by  $\mathbf{P}_{n \times m}$ , out of  $k$  sample vectors  $x_i \in \mathbb{R}^n$  denoted by  $\mathbf{X}_{n \times k}$  for  $k \geq m$ .

The PCA algorithm searches for a low rank space  $\mathbf{P}$  that would best describe these observations. In a flat domain, the PCA algorithm computes an orthonormal basis  $\mathbf{P}$  that minimizes the error of projecting  $x_i$  onto  $\mathbf{P}$ , namely,

$$\operatorname{argmin}_{\mathbf{P}} \sum_{i=1}^k \|\mathbf{P}\mathbf{P}^T x_i - x_i\|_2^2 \quad \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I}, \quad (3.1)$$

where the projection of  $x_i$  onto the basis  $\mathbf{P}$  is given by  $\mathbf{P}\mathbf{P}^T x_i$ . Here,  $\mathbf{I}$  denotes an identity matrix of size  $m \times m$ . Using these notations, one could show that

$$\begin{aligned} \|\mathbf{P}\mathbf{P}^T x_i - x_i\|_2^2 &\equiv (\mathbf{P}\mathbf{P}^T x_i - x_i)^T (\mathbf{P}\mathbf{P}^T x_i - x_i) \\ &= x_i^T x_i - \operatorname{trace}(\mathbf{P}\mathbf{P}^T x_i x_i^T). \end{aligned} \quad (3.2)$$

Recalling that  $\sum_{i=1}^m x_i x_i^T \equiv \mathbf{X}\mathbf{X}^T$ , the minimization (3.1) can be easily shown to be equivalent to the maximization problem

$$\operatorname{argmax}_{\mathbf{P}} \operatorname{trace}(\mathbf{P}\mathbf{P}^T\mathbf{X}\mathbf{X}^T) \quad \text{s.t. } \mathbf{P}^T\mathbf{P} = \mathbf{I}. \quad (3.3)$$

$\mathbf{P}$  that solves (3.3) can be extracted by a singular value decomposition (SVD for short) of  $\mathbf{X}$ .

## 4 Intrinsic LBO Versus Extrinsic PCA

A shape representation model that is invariant to poses of articulated objects and to gestures of most creatures in nature is the isometry model. As empirically shown in [10] for the case of facial expressions, the intrinsic geometry of a facial surface is more or less preserved for the same subject when smiling, posing a neutral expression or expressing sadness. Isometry or length preserving mappings were thus used for expression invariant face recognition. At the other end, the coordinates of the surface, that can be thought of as functions defined on the surface manifold, determine the pose itself.

We suggest to consider two geometric structures, the intrinsic one defined by the differential relations between surface points also known as the metric, and the extrinsic one defined by the coordinates themselves. As shown in [2], smooth functions defined on the surface are optimally approximated by the leading eigenfunctions of the Laplace-Beltrami operator. In some cases, more information about functions defined on the surfaces is provided, like examples of the surface coordinates that we would like to represent, given in different poses. In that case, the leading terms, or the principal components derived through a PCA method could provide the best representation space for realizations within the convex hull of the observed functions. The combination of both intrinsic structure and extrinsic observations could allow us to enjoy the best of both worlds. The result that we are looking for would be an orthonormal optimal representation space in terms of projection to the leading subspace. It could faithfully approximate realizations outside the convex hull of the observed functions. Here, we propose to combine the knowledge of both intrinsic geometry, that is captured by the decomposition of the LBO, and extrinsic examples apprehended by the PCA, to construct a naturally regularized PCA.

## 5 Regularized-PCA Formulation

Principal component analysis in [25], as a classical method in data representation and analysis, also plays a major role in modeling and studying surfaces.

Given the surface  $S$ , the principal components of  $k$  given scalar continuous functions,  $f_i : S \rightarrow \mathbb{R}$ , are defined by the orthonormal basis  $\{\psi_j\}$  that is obtained as a solution of

$$\operatorname{argmin}_{\{\psi_j\}} \sum_{i=1}^k \left\| f_i - \sum_{j=1}^m \langle f_i, \psi_j \rangle \psi_j \right\|_g^2 \quad (5.1)$$

for any given  $m$ . Up to this point the surface  $S$  and its metric  $g$  do not play a major role in the definition of  $\{\psi_j\}$ . The manifold  $S$  and its metric  $g$  should allow us to tune the representation

space better. In the above definition of continuous principal components, there is no smoothness assumption of the representation space  $\{\psi_j\}$  on the manifold.

Next, assume that the functions that we would like to represent, have a bounded gradient magnitude  $\|\nabla_g f\|_g^2 \leq 1$  on the surface  $S$ . Such smooth functions are best represented by the eigenfunctions of the LBO, and those can be obtained as a minimization of the Dirichlet energy

$$\operatorname{argmin}_{\{\psi_j\}} \sum_{j=1}^m \|\nabla_g \psi_j\|_g^2 \quad \text{s.t.} \quad \langle \psi_i, \psi_j \rangle_g = \delta_{ij}. \quad (5.2)$$

Next let us translate the general metric case into its discrete form. The metric  $(g_{ij})$  defines the mapping between an arc length on a given parametrization space to length on the given manifold. An area element can then be defined by  $\sqrt{\det(g)}$ , and in a discrete setting, one can define an area elements diagonal matrix  $\mathbf{A}$ . For example, for a triangulated surface, the area elements  $A_{ii}$  are proportional to the sum of areas of all triangles that share the vertex  $i$  (see Figure 2).

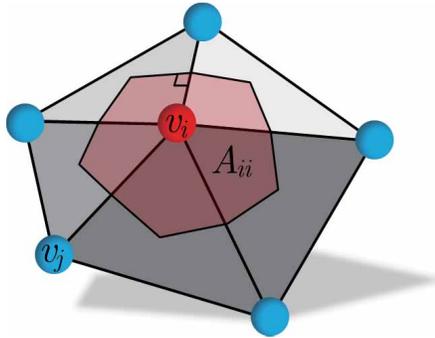


Figure 2 The area of the Voronoi cell about the vertex  $v_i$  (red region) defines the entry  $A_{ii}$  of the diagonal matrix  $\mathbf{A}$ .

On the surface  $S$ , the geometric projection of  $x_i$  onto  $\mathbf{P}$  reads  $\mathbf{P}\mathbf{P}^T \mathbf{A}x_i$ , and an orthonormal  $\mathbf{P}$  is defined as  $\mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{I}$ . The geometric PCA reads

$$\operatorname{argmin}_{\mathbf{P}} \sum_{i=1}^k \|\mathbf{P}\mathbf{P}^T \mathbf{A}x_i - x_i\|_g^2 \quad \text{s.t.} \quad \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{I}, \quad (5.3)$$

which can be shown to be equivalent to solving

$$\operatorname{argmax}_{\mathbf{P}} \operatorname{trace}(\mathbf{P}\mathbf{P}^T \mathbf{A} \mathbf{X} \mathbf{X}^T \mathbf{A}) \quad \text{s.t.} \quad \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{I}. \quad (5.4)$$

Next, we would also like to incorporate the natural smoothness of the signals that we would like to represent. It turns out that the geometric Dirichlet energy produces the Laplace-Beltrami operator that would have yielded the generalized Fourier basis. On a manifold, the discrete Laplacian can be defined, for example, by the general form  $\mathbf{L} = \mathbf{A}^{-1} \mathbf{W}$ . A specific realization

of these weights for two dimensional triangulated surfaces is given by the cotangent weights scheme (see for example [26, 24]). The Laplacian for a triangulated surface given by its vertices  $\{v_i\}$  and edges  $\{e_{ij}\}$  can be approximated by  $\mathbf{L} = \mathbf{A}^{-1}\mathbf{W}$ , where  $\mathbf{A}$  is the diagonal matrix of the areas of Voronoi cells about each vertex (see Figure 2).

The matrix  $\mathbf{W}$  is defined by the cotangent weights

$$W_{ij} = \begin{cases} \sum_{j:v_j \in \mathcal{N}_1(v_i)} (\cot \alpha_{ij} + \cot \beta_{ij}), & \text{if } i = j, \\ -\cot \alpha_{ij} - \cot \beta_{ij}, & \text{if } i \neq j, v_j \in \mathcal{N}_1(v_i), \\ 0, & \text{otherwise,} \end{cases} \quad (5.5)$$

where  $\mathcal{N}_1(v_i)$  is the set of 1-ring (neighboring) vertices about  $v_i$ , and  $\alpha_{ij}$  and  $\beta_{ij}$  are the angles opposite to the edge  $(v_i, v_j)$  (see Figure 3).

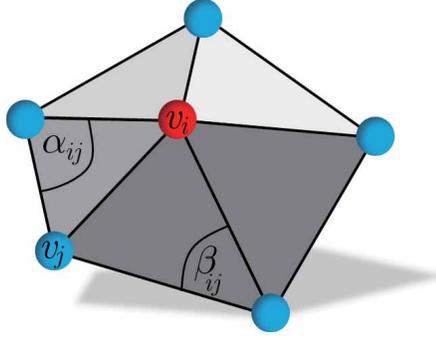


Figure 3 Cotangent weights as defined in the equation 5.5.

For the  $j$ -th column of  $\mathbf{P}$  we have

$$\begin{aligned} \|\nabla_g P_j\|_g^2 &\equiv \langle \nabla_g P_j, \nabla_g P_j \rangle_g = \langle \Delta_g P_j, P_j \rangle_g \\ &\approx (\mathbf{L}P_j)^T \mathbf{A}P_j = P_j^T \mathbf{W}^T \mathbf{A}^{-T} \mathbf{A}P_j \\ &= P_j^T \mathbf{W}P_j = \text{trace}(\mathbf{W}P_j P_j^T), \end{aligned} \quad (5.6)$$

and for the basis  $\mathbf{P}$  as a whole, optimizing for the geometric discrete Dirichlet energy reads

$$\underset{\mathbf{P}}{\text{argmin}} \text{trace}(\mathbf{P}\mathbf{P}^T \mathbf{W}) \quad \text{s.t. } \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{I}. \quad (5.7)$$

We could replace the above minimization problem with the following maximization one

$$\underset{\mathbf{P}}{\text{argmax}} \text{trace}(\mathbf{P}\mathbf{P}^T \mathbf{W}^+) \quad \text{s.t. } \mathbf{P}^T \mathbf{A} \mathbf{P} = \mathbf{I}, \quad (5.8)$$

where  $\mathbf{W}^+$  is the pseudo inverse, also known as Moore-Penrose matrix inverse, of  $\mathbf{W}$ .

## 6 Secrets of Marriage

In [1], it was proven that an optimal and unique solution to the regularized-PCA problem

$$\operatorname{argmin}_{\mathbf{P}} \sum_{i=1}^k (\|\mathbf{P}\mathbf{P}^T \mathbf{A}x_i - x_i\|_g^2 + \mu \operatorname{trace}(\mathbf{W}P_jP_j^T)) \quad \text{s.t. } \mathbf{P}^T \mathbf{A}\mathbf{P} = \mathbf{I}, \quad (6.1)$$

whose solution is equivalent to that of

$$\operatorname{argmax}_{\mathbf{P}} \operatorname{trace}(\mathbf{P}\mathbf{P}^T(\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A} - \mu\mathbf{W})) \quad \text{s.t. } \mathbf{P}^T \mathbf{A}\mathbf{P} = \mathbf{I}, \quad (6.2)$$

can be realized by decomposing the matrix  $\mathbf{M} = \mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A} - \mu\mathbf{W}$ . The leading eigenvectors of the decomposition of  $\mathbf{M}$  define  $\mathbf{P}$ . Another option is defining the intrinsic smoothness given by the equation (5.8) and the PCA given by the equation (5.4) as eigendecomposition maximization problems that again allows us to unite them into a single regularized-PCA problem

$$\operatorname{argmax}_{\mathbf{P}} \operatorname{trace}(\mathbf{P}\mathbf{P}^T(\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A} + \mu\mathbf{W}^+)) \quad \text{s.t. } \mathbf{P}^T \mathbf{A}\mathbf{P} = \mathbf{I}. \quad (6.3)$$

Here, the matrix that we have to decompose, namely  $\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A} + \mu\mathbf{W}^+$ , is positive definite and symmetric. Most numerical decomposition algorithms extract the eigenvectors of a given matrix one at a time, where the order is determined by the magnitude of the corresponding eigenvalues. The basic operation of such procedures, also known as Arnoldi iteration, involves in the multiplication of the matrix with a vector. It is obviously efficient for sparse matrices. Here, we are combining a low-rank (PCA) part, given by  $\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A}$ , with a sparse LBO, given by  $\mathbf{W}$ . This combination poses a numerical challenge.

As a remedy, we propose to apply a small size SVD to  $\mathbf{X}_{n \times k}$ , such that

$$\mathbf{X}_{n \times k} = \mathbf{U}_{n \times n} \mathbf{D}_{n \times k} \mathbf{V}_{k \times k}^T,$$

where  $\mathbf{D}$  is a diagonal, and  $\mathbf{U}$  and  $\mathbf{V}$  are both orthonormal matrices. Denoting  $\tilde{\mathbf{U}} = \mathbf{A}\mathbf{U}$ , we can write

$$\mathbf{A}\mathbf{X}\mathbf{X}^T \mathbf{A} = \mathbf{A}\mathbf{U}\mathbf{D}^2\mathbf{U}^T \mathbf{A} = \tilde{\mathbf{U}}\mathbf{D}^2\tilde{\mathbf{U}}^T. \quad (6.4)$$

Given a general vector  $y$ , the multiplication

$$(\tilde{\mathbf{U}}\mathbf{D}^2\tilde{\mathbf{U}}^T + \mathbf{W}^+)y = \tilde{\mathbf{U}}(\mathbf{D}^2(\tilde{\mathbf{U}}^T y)) + \mathbf{W}^+y \quad (6.5)$$

can be efficiently executed by the following procedure.

---

**Algorithm 1** Compute  $\tilde{\mathbf{U}}(\mathbf{D}^2(\tilde{\mathbf{U}}^T y)) + \mathbf{W}^+y$

---

**input**  $y$

$a \leftarrow \tilde{\mathbf{U}}^T y$

$a \leftarrow \mathbf{D}^2 a$

$a \leftarrow \tilde{\mathbf{U}} a$

Compute  $b$  by solving  $y = \mathbf{W}b$  in a least square sense.

**return**  $(a + b)$

---

Let us analyze the complexity of the above procedure. The first step takes  $O(kn)$ , after  $\tilde{\mathbf{U}}$  is truncated to an  $n \times k$  matrix, where  $k$  is the number of required PCA components, and  $k \ll n$ . The result, a  $k \times 1$  vector, is then multiplied by  $\mathbf{D}_{k \times k}^2$  which takes  $O(k^2)$ . The complexity of the third step is  $O(kn)$ , obtained by computing  $\tilde{\mathbf{U}}_{n \times k} \mathbf{a}_{k \times 1}$ . Next, numerically solving the sparse system  $y = \mathbf{W}b$ , with  $O(n)$  elements in  $\mathbf{W}$ , can be efficiently performed by classical procedures. The above marriage of a low rank term with a sparse structured part is computed very fast with very low requirements. For example, hundred eigenvectors for a 10,000 vertices surface can be computed in a few seconds by using Matlab on a modern laptop computer. Note also that the above procedure exploits both the sparsity of  $\mathbf{W}$  and the low rank property of  $\mathbf{X}$  within a unified procedure.

## 7 Experimenting with Shapes

In the following example, shown in Figure 4, we demonstrate the power of the regularized

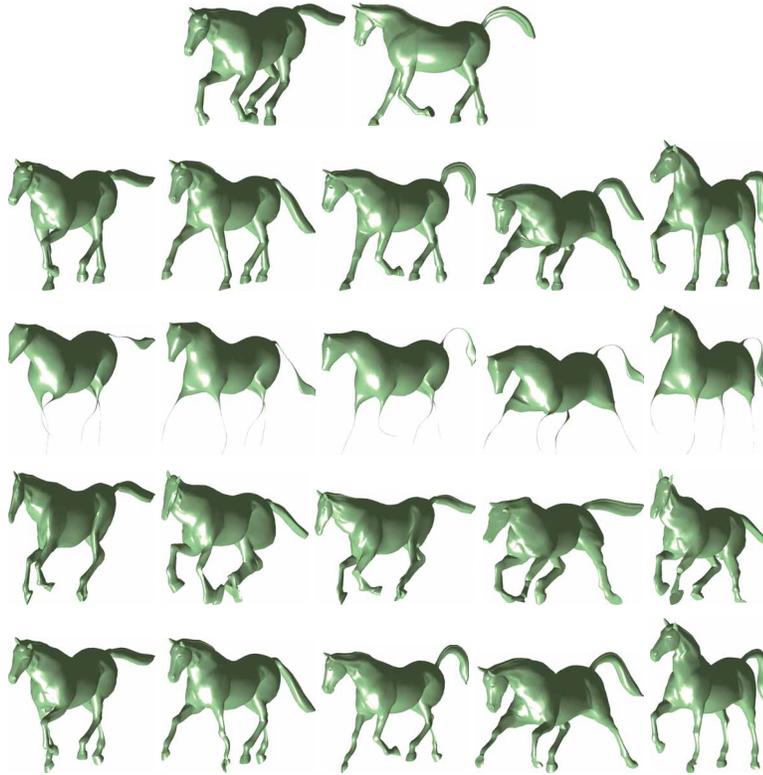


Figure 4 Top row: Training set. Second row: Test set. Third row: Projection to the 100 leading eigenfunctions of the LBO. Fourth row: Projection to the first 100 principal components vectors. Bottom row: Projection to the first 100 vectors computed by the regularized-PCA hybrid model.

PCA in modeling out-of-sample shapes. We treat the  $X, Y$  and  $Z$  coordinates of given shapes as training vectors for a PCA procedure. We also use the intrinsic geometry of one of the shapes, which is assumed to be similar for all poses, to define a Laplace-Beltrami operator for which we computed the leading eigenfunctions (corresponding to the smallest eigenvalues). The coordinates of five shapes, being more different than the training ones, are then projected onto the first 100 leading eigenvectors or the LBO (third row), the PCA (fourth row), and the regularized model (bottom row). From Figure 4, it appears that the intrinsic geometry provided by the leading eigenfunctions of the LBO captures the general pose while giving up the fine details. The PCA subspace, at the other end, captures the fine details and poses within the convex-hull of the given training set while it is failing to represent poses beyond the observed ones. The hybrid model captures both poses and details, and obviously enjoys the best of both representation spaces.

## 8 Happy Ending

The LBO  $\Delta_g$  applied to the coordinates of a surface  $S$  yields the mean curvature vector,  $H\vec{N} = \Delta_g S$ . At the spectral end, smooth functions on the manifold, like shape coordinates, are best represented by the leading eigenfunctions of the LBO,  $\Delta_g \psi_i = \lambda_i \psi_i$ . However, when a set of observations is provided, like many instances of the same objects in various poses  $\{S_i\}$ , classical statistical learning techniques suggest the PCA as a method of choice for finding a low rank matrix  $\mathbf{P}$  that would represent these observations, by minimizing  $\sum_i \|S_i - \mathbf{P}\mathbf{P}^T S_i\|$  or equivalently maximizing  $\text{trace}(\mathbf{P}\mathbf{P}^T \mathbf{S}\mathbf{S}^T)$ , where  $\mathbf{S}\mathbf{S}^T = \sum_i S_i S_i^T$ . We have shown that the best representation space could be provided by a model that unifies the two different sources of information about our data. By marrying the intrinsic structure with extrinsic observations, we have shown that even extreme out-of-sample configurations can be projected onto a new low dimensional space that preserves poses without sacrificing the fine details.

## References

- [1] Aflalo, Y., Brezis, H., Bruckstein, A. M., Kimmel, R. and Sochen, N., Best bases for signal spaces, *Comptes Rendus de l'Académie des Sciences, Series I*, 2016.
- [2] Aflalo, Y., Brezis, H. and Kimmel, R., On the optimality of shape and data representation in the spectral domain, *SIAM Journal on Scientific Computing*, **8**(2), 2015, 1141–1160.
- [3] Aubry, M., Schlickewei, U. and Cremers, D., The wave kernel signature: A quantum mechanical approach to shape analysis, *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference, IEEE, 2011, 1626–1633.
- [4] Belkin, M. and Niyogi, P., Convergence of Laplacian eigenmaps, *Proceedings of the Twentieth Annual Conference on NIPS*, 2006, 129–136.
- [5] Ben-Chen, M. and Gotsman, C., On the optimality of spectral compression of mesh data, *ACM Trans. Graph.*, **24**(1), 2005, 60–80.
- [6] Bérard, P., Besson, G. and Gallot, S., Embedding Riemannian manifolds by their heat kernel, *Geometric and Functional Analysis*, **4**(4), 1994, 373–398.
- [7] Borg, I. and Groenen, P., *Modern Multidimensional Scaling: Theory and Applications*, Springer-Verlag, New York, 1997.
- [8] Brezis, H., *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext Series, Springer-Verlag, New York, 2010.

- [9] Bronstein, A. M., Bronstein, M. M. and Kimmel, R., Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching, *Proceedings of the National Academy of Sciences*, **103**(5), 2006, 1168–1172.
- [10] Bronstein, A. M., Bronstein, M. M. and Kimmel, R., Rock, paper, and scissors: Extrinsic vs. intrinsic similarity of non-rigid shapes, IEEE 11th International Conference on Computer Vision, 2007.
- [11] Bronstein, A. M., Bronstein, M. M., Kimmel, R., et al., A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching, *International Journal of Computer Vision*, **89**(2–3), 2010, 266–286.
- [12] Coifman, R. R. and Lafon, S., Diffusion maps, *Applied and Computational Harmonic Analysis*, **21**(1), 2006, 5–30.
- [13] Coifman, R. R., Lafon, S., Lee, A. B., et al., Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *Proceedings of the National Academy of Sciences*, **102**(21), 2005, 7426–7431.
- [14] Coifman, R. R. and Maggioni, M., Diffusion wavelets, *Applied and Computational Harmonic Analysis*, **21**(1), 2006, 53–94.
- [15] Gėbal, K., Bærentzen, J. A., Aanaes, H. and Larsen, R., Shape analysis using the auto diffusion function, *Computer Graphics Forum*, **28**, 2009, 1405–1413.
- [16] Hammond, D. K., Vandergheynst, P. and Gribonval, R., Wavelets on graphs via spectral graph theory, *Applied and Computational Harmonic Analysis*, **30**(2), 2011, 129–150.
- [17] Jiang, B., Ding, C., Luo, B. and Tang, J., Graph-Laplacian PCA: Closed-form solution and robustness, IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, 3492–3498.
- [18] Jolliffe, I. T., *Principal Component Analysis*, Second Edition, Springer-Verlag, New York, 2002.
- [19] Kalofolias, V., Bresson, X., Bronstein, M. M. and Vandergheynst, P., Matrix completion on graphs, Proc. NIPS, Workshop Out of the Box: Robustness in High Dimension, 2014.
- [20] Karni, Z. and Gotsman, C., Spectral compression of mesh geometry, *ACM Transactions on Graphics*, 2000, 279–286.
- [21] Kovnatsky, A., Bronstein, M. M., Bresson, X. and Vandergheynst, P., Functional correspondence by matrix completion, Proc. Computer Vision and Pattern Recognition (CVPR), 2015.
- [22] Kovnatsky, A., Bronstein, M. M., Bronstein, A. M., et al., Coupled quasi-harmonic bases, *Computer Graphics Forum*, **32**, 2013, 439–448.
- [23] Lévy, B., Laplace-Beltrami eigenfunctions towards an algorithm that “understands” geometry, Shape Modeling and Applications, IEEE International Conference, IEEE, 2006, 13–13.
- [24] Meyer, M., Desbrun, M., Schröder, P. and Barr, A. H., Discrete differential-geometry operators for triangulated 2-manifolds, *Visualization and mathematics*, **3**(2), 2002, 52–58.
- [25] Pearson, K., On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, **2**(11), 1901, 559–572.
- [26] Pinkall, U. and Polthier, K., Computing discrete minimal surfaces and their conjugates, *Experiment. Math.*, **2**, 1993, 15–36.
- [27] Qiu, H. and Hancock, E. R., Clustering and embedding using commute times, *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**(11), 2007, 1873–1890.
- [28] Ramsay, J. O. and Silverman, B. W., *Functional data analysis*, Second Edition, Springer-Verlag, New York, 2005.
- [29] Rong, G., Cao, Y., and Guo, X., Spectral mesh deformation, *Vis. Comput.*, **24**(7), 2008, 787–796.
- [30] Rustamov, R., Ovsjanikov, M., Azencot, O., et al., Map-based exploration of intrinsic shape differences and variability, SIGGRAPH, ACM, 2013.
- [31] Shahid, N., Bresson, X., Bronstein, M. M. and Vandergheynst, P., Robust principal component analysis on graphs, Int. Conf. Comp. Vision (ICCV), 2015.
- [32] Shawe-Taylor, J. and Cristianini, N., *Kernel Methods For Pattern Analysis*, Cambridge University Press, Cambridge, 2004.
- [33] Sun, J., Ovsjanikov, M. and Guibas, L., A concise and provably informative multi-scale signature based on heat diffusion, Proceedings of the Symposium on Geometry Processing, Eurographics Association, SGP’09, Aire-la-Ville, Switzerland, 2009, 1383–1392.
- [34] Weinberger, H., *Variational Methods for Eigenvalue Approximation*, Society for Industrial and Applied Mathematics, 1974.

- [35] Zaharescu, A., Boyer, E., Varanasi, K. and Horaud, R., Surface feature detection and description with applications to mesh matching, *Computer Vision and Pattern Recognition, IEEE Conference on CVPR 2009, IEEE, 2009*, 373–380.
- [36] Zhou, X. and Srebro, N., Error analysis of laplacian eigenmaps for semi-supervised learning, *International Conference on Artificial Intelligence and Statistics, 2011*, 901–908.