2014

Masking Schemes for Image Manifolds

Hamid Dadkhahi, *University of Massachusetts - Amherst*
Marco Duarte, *University of Massachusetts - Amherst*

MASKING SCHEMES FOR IMAGE MANIFOLDS

Hamid Dadkhahi and Marco F. Duarte

Dept. of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003

ABSTRACT

We consider the problem of selecting an optimal mask for an image manifold, i.e., choosing a subset of the dimensions of the image space that preserves the manifold structure present in the original data. Such masking implements a form of compressed sensing that reduces power consumption in emerging imaging sensor platforms. Our goal is for the manifold learned from masked images to resemble the manifold learned from full images as closely as possible. We show that the process of finding the optimal masking pattern can be cast as a binary integer program, which is computationally expensive but can be approximated by a fast greedy algorithm. Numerical experiments show that the manifolds learned from masked images resemble those learned from full images for modest mask sizes. Furthermore, our greedy algorithm performs similarly to the exhaustive search from integer programming at a fraction of the computational cost.

Index Terms— Manifold Learning, Masking, Dimensionality Reduction

1. INTRODUCTION

Recent advances in sensing technology have enabled a massive increase in the dimensionality of data captured from digital sensing systems. Naturally, the high dimensionality of data affects various stages of the digital systems, from acquisition to processing and analysis of the data. To meet communication, computation, and storage constraints, in many applications one seeks a low-dimensional embedding of the high-dimensional data that shrinks the size of the data representation while retaining the information we are interested in capturing. This problem of *dimensionality reduction* has attracted significant attention in the signal processing and machine learning communities.

For high-dimensional data, the process of data acquisition followed by a dimensionality reduction method is inherently wasteful, since we are often not interested in obtaining the full-length representation of the data. This issue has been addressed by *compressed sensing*, a technique to simultaneously acquire and reduce the dimensionality of sparse signals in a randomized fashion [1]. Compressed sensing provides a good match to the requirements of cyber-physical systems, where power constraints are paramount. For instance, a fundamental challenge in the design of computational eyeglasses is addressing stringent resource constraints on data acquisition and processing that include sensing fidelity and energy budget, in order to meet lifetime and size design targets [2]. In such applications, one wishes to reduce the size of the representation of the data to be processed, often by applying standard compression algorithms. More recently, the design of linear embeddings that allow for data processing directly from lower-dimensional representation has been considered, with a particular emphasis in imaging [3–6]. However, while the aforementioned embeddings may reduce the computational and communication demands, they do not reduce the power consumption burden of data acquisition. This is because they require all image pixels to be sensed, and so they cannot be implemented more efficiently than standard acquisition.

Emerging imaging sensor architectures for embedded systems significantly increase the flexibility in power consumption by allowing pixel-level control of the acquisition process [2, 7]; the power consumption of imaging becomes proportional to the number of pixels to be acquired using the array. Thus, it is now possible to meet stringent power and communication requirements by designing *data-dependent image masking schemes* that reduce the number of pixels involved in acquisition while, like the aforementioned linear embeddings, preserving the information of interest. The selection of a masking pattern is ideally driven by knowledge of a model that captures the relevant information in the data, such as a nonlinear manifold model for parameter estimation.

In this paper, we consider the problem of designing masking patterns that preserve the geometric structure of a high-dimensional dataset modeled as a nonlinear manifold. Note that in terms of linear embeddings, masking schemes may be described as a restriction to embeddings where the projection directions are required to correspond to canonical vectors. Previous work on manifold-based dimensionality reduction does not address the highly constrained (masking) setting that is motivated by our application.

2. BACKGROUND

Manifolds and Linear Dimensionality Reduction: A set of data points $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ in a high-dimensional am-

Email: {hdadkhahi, mduarte}@ecs.umass.edu. This work was supported by NSF Grant IIS-1239341.

bient space $\mathbb{R}^d$ that have been generated by an ℓ-dimensional parameter correspond to a sampling of an ℓ-dimensional *manifold* $\mathcal{M} \subset \mathbb{R}^d$. Given the high-dimensional data set $\mathcal{X}$, we would like to find the parameterization that has generated the manifold. One way to discover this parametrization is to *embed* the high-dimensional data on the manifold to a low-dimensional space $\mathbb{R}^m$ so that the local geometry of the manifold is preserved. This process is known as *dimensionality reduction*, since $m \ll d$.

When the dimensionality reduction embedding is *linear*, it is defined by a matrix $\Phi \in \mathbb{R}^{m\times d}$ that maps the data in the ambient space $\mathbb{R}^d$ into a low-dimensional space $\mathbb{R}^m$. One such popular scheme is *principal component analysis* (PCA), defined as the orthogonal projection of the data onto a linear subspace of lower dimension m such that the variance of the projected data is maximized.

Nonlinear Manifolds and Manifold Learning: Unfortunately, PCA fails to preserve the geometric structure of a *nonlinear manifold*, i.e., a manifold where the map from the parameter space to the data space is nonlinear. Particularly, since PCA arbitrarily distorts individual pairwise distances, it can significantly change the local geometry of the manifold. Fortunately, *manifold learning methods (or nonlinear embedding methods)* can successfully embed the data into a low-dimensional space while preserving the local geometry of the manifold, measured by a neighborhood-preserving criteria that varies depending on the specific method, in order to simplify parameter estimation.

The *Isomap* method aims to preserve the pairwise *geodesic distances* between data points [8]. The geodesic distance $d_G(x_i, x_j)$ is defined as the length of the shortest path between two data points $x_i, x_j \in \mathcal{M}$ along the surface of the manifold $\mathcal{M}$. Isomap first finds an approximation to the geodesic distances between each pair of data points by constructing a neighborhood graph in which each point is connected only to its k nearest neighbors; the edge weights are equal to the corresponding pairwise distances. For neighboring pairs of data points, the Euclidean distance provides a good approximation for the geodesic distance, i.e., $d_G(x_i, x_j) \approx \|x_i - x_j\|_2 \quad \text{for} \quad x_j \in \mathcal{N}_k(x_i)$, where $\mathcal{N}_k(x_i)$ designates the set of k nearest neighbors to point $x_i \in \mathcal{X}$. For non-neighboring points, the length of the shortest path along the neighborhood graph is used to estimate the geodesic distance. Then, multidimensional scaling (MDS) [9] is applied to the resulting geodesic distance matrix to find a set of low-dimensional points that best match such distances.

Linear Dimensionality Reduction for Manifolds: An alternative linear embedding approach to PCA is the method of *random projections*, where the entries of the linear dimensionality reduction matrix are drawn independently following a standard probability distribution such as normal Gaussian or Rademacher. One can show that such random projections preserve the relevant pairwise distances with high probability [3, 4]. Unfortunately, random embeddings are independent of the geometric structure of the data, and thus cannot take advantage of training data.

Recently, a data-dependent linear embedding obtained via convex optimization (referred to as *NuMax*) has been proposed [5, 10]. The key concept in NuMax is to obtain an isometry on the set of pairwise data point differences, dubbed *secants*, after being normalized to lie on the unit sphere:

$$\mathcal{S} = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|_2} : x_i, x_j \in \mathcal{M} \right\}.$$

NuMax relies on a convex optimization problem that finds an embedding Φ into a space with minimum dimension such that the secants in $\mathcal{S}$ are preserved up to a norm distortion parameter δ. More precisely, finding the linear embedding is cast as the following rank-minimization problem:

$$\begin{aligned} P^* = \arg\min \ \ & \text{rank}(P) \\ \text{subject to} \ \ & |s^T P s - 1| \le \delta \quad \forall\, s \in \mathcal{S}, \quad P \succeq 0. \end{aligned} \tag{1}$$

After P^* is obtained, one can factorize $P^* = \Phi^T\Phi$ in order to obtain the desired low-dimensional embedding Φ. We note that the rank of the solution determines the dimensionality of the embedding, and is controlled by the choice of the distortion parameter $\delta \in [0, 1]$. Note also that $s^T P s = \|\Phi s\|_2^2$; thus, the first constraint essentially upper-bounds the distortion incurred by each secant $s \in \mathcal{S}$. The problem (1) is NP-hard, but one may instead solve its nuclear norm relaxation, where the rank of P is replaced by its nuclear norm $\|P\|_*$. Since P is a positive semidefinite symmetric matrix, its nuclear norm amounts to its trace, and thus (1) is equivalent to a semidefinite program and can be solved in polynomial time.

3. MANIFOLD MASKING

In this section, we emulate the criteria used in linear and nonlinear embedding algorithms from Section 2 to develop schemes to obtain structure-preserving masking patterns for manifold-modeled data. To unify notation, we are seeking a masking index set $\Omega = \{\omega_1, \ldots, \omega_m\}$ of cardinality m that is a subset of the dimensions $[d] := \{1, 2, \ldots, d\}$ of the high-dimensional space containing the original dataset.

Optimization-Based Mask Selection: Inspired by the optimization approach of NuMax and the neighborhood-preservation notion of Isomap, we formulate a method for manifold masking that aims at minimizing the distortion incurred by secants of neighboring data points.

Recall that Isomap attempts to preserve the geodesic distances rather than Euclidean distances of data points. Since only the Euclidean distances of neighboring data points match their geodesic counterparts, we are interested in devising a masking operator than preserves the pairwise distances of each data point with its k nearest neighbors. This gives rise to the reduced secant set

$$\mathcal{S}_k = \left\{ \frac{x_i - x_j}{\|x_i - x_j\|_2} : i \in [n], x_j \in \mathcal{N}_k(x_i) \right\} \subseteq \mathcal{S}.$$

To simplify notation, we define the masking linear operator $\Psi : x_i \mapsto \{x_i(j)\}_{j\in\Omega}$ corresponding to the masking index set Ω. We also denote the squared secants by the column vectors a_i with entries $a_i(j) = s_i^2(j)$ for all $j \in [d]$ and for each $i \in [|\mathcal{S}_k|]$. Since the secants are normalized, we have $\sum_{j=1}^{d} a_i(j) = 1$ for all $i \in [|\mathcal{S}_k|]$.

It can be shown [11] that the expectation of the masked secants over a uniform distribution for the masking index set Ω is given by $\mathbb{E}[\|\Psi s_i\|_2^2] = \frac{m}{d}$. Thus, the secants $s_i \in \mathcal{S}_k$ are inevitably subject to a compaction factor of $\sqrt{\frac{m}{d}}$ in expectation by the masking operator Ψ; this behavior bears out empirically with randomized maskings for the dataset considered in Section 4. Hence, we will aim to find a masking Ψ such that for all $s_i \in \mathcal{S}_k$, the squared norm of the masked secants $\|\Psi s_i\|_2^2$ matches $\frac{m}{d}$ as closely as possible. Note that $\|\Psi s_i\|_2^2 = \sum_{j\in\Omega} s_i^2(j) = \sum_{j=1}^{d} s_i^2(j) z(j) = a_i^T z$, where the d-dimensional indicator vector z is defined as $z(j) = 1$ if $j \in \Omega$, and zero otherwise. The average and maximum secant norm distortion caused by masking can be written in terms of the vector z and the squared secants matrix $A := [a_1\ a_2\ \ldots\ a_{|\mathcal{S}_k|}]^T$ as

$$\sum_{s_i\in\mathcal{S}_k} \left| \|\Psi s_i\|_2^2 - \frac{m}{d} \right| = \left\| Az - \frac{m}{d} 1_{|\mathcal{S}_k|} \right\|_1,$$

$$\max_{s_i\in\mathcal{S}_k} \left| \|\Psi s_i\|_2^2 - \frac{m}{d} \right| = \left\| Az - \frac{m}{d} 1_{|\mathcal{S}_k|} \right\|_\infty,$$

respectively, where $1_{|\mathcal{S}_k|}$ denotes the $|\mathcal{S}_k|$-dimensional all-ones vector. Thus, we find a masking pattern by casting the following integer program:

$$z^* = \arg\min_z \left\| Az - \frac{m}{d} 1_{|\mathcal{S}_k|} \right\|_p \quad (2)$$
$$\text{subject to} \quad 1_d^T z = m, z \in \{0,1\}^d,$$

where $p = 1$ and $p = \infty$ correspond to optimizing the average and maximum secant norm distortion caused by the masking, respectively. The equality constraint dictates that only m dimensions are to be retained in the masking process.

The integer program (2) is computationally intractable even for moderate-size datasets [12]. We note that the non-integer relaxation of (2) results in the trivial solution $z^* = \frac{m}{d} 1_d$. Note also that the matrix A depends on the dataset used; thus in general it does not hold necessary properties for relaxations of integer programs to be successful (e.g. being totally unimodular, having binary entries, etc.).

Greedy Algorithm for Mask Selection: We propose a heuristic greedy algorithm that can find an approximate solution for (2) in a drastically reduced time. The greedy approach in Algorithm 1 gives an approximate solution for the ℓ_p-norm minimization in (2). The algorithm iteratively selects elements of the masking index set Ω as a function of the squared secants matrix A. We initialize Ω as the empty set and denote $\Omega^c = [d] \setminus \Omega$. At iteration i of the algorithm, we find a new dimension that, when added to the existing dimensions in Ω, causes the squared norm of the masked secant to match the expected value of $\frac{i}{d}$ as closely as possible; thus, the greedy algorithm pursues the same objective function as the optimization approach of (2) for a mask of size i. More precisely, at step i of the algorithm, we find the column of A indexed by $\omega \in \Omega^c$ (which is indicated by A_ω), whose addition with the sum of previously chosen columns $\bar{A}_\Omega = \sum_{\omega\in\Omega} A_\omega$ has minimum distance (in ℓ_p-norm) to $\frac{i}{d} 1_{|\mathcal{S}_k|}$. Note that $\bar{A}_\Omega = Az$, where z again denotes the indicator vector for the masking index set $\Omega \subseteq [d]$.

Algorithm 1 Manifold-Aware Pixel Selection (MAPS)

Inputs: normalized squared secants matrix A, number of dimensions m
Outputs: masking index set Ω
Initialize: $\Omega \leftarrow \{\}$
for $i = 1 \rightarrow m$ **do**
  $\bar{A}_\Omega \leftarrow A_\Omega \cdot 1_{|\Omega|}$
  $\omega_i \leftarrow \arg\min_{\omega\in\Omega^c} \|A_\omega + \bar{A}_\Omega - \frac{i}{d} 1_{|\mathcal{S}_k|}\|_p$
  $\Omega \leftarrow \Omega \cup \{\omega_i\}$
end for

4. NUMERICAL EXPERIMENTS

In this section, we present a set of experiments that compare the performance of the proposed algorithms to those in the existing linear embedding literature, in terms of preservation of the low-dimensional structure of a nonlinear manifold. We evaluate the methods described in Section 3, together with two baseline methods: *random masking*, where we pick an m-subset of the d data dimensions uniformly at random, and *principal coordinate analysis (PCoA)*, where we select the indices of m dimensions with the highest variance across the dataset [11]. We use Mosek [13] to solve the binary integer program in (2), posed as an optimization problem with the CVX programming interface for Matlab [14].

For our experiments, we use a custom eye-tracking dataset from a computational eyeglass prototype.[1] The *Eyeglasses* dataset corresponds to captures from a prototype implementation of computational eyeglasses that use the imaging sensor array of [7]. These images are downsampled from their original size of 110×110 pixels to 40×40 pixels for the benefit of the integer program's computational complexity.

The algorithms are tested for linear embeddings of dimensions $m = 50, 100, 150, 200, 250$; for the masking algorithms of Section 3, m provides the size of the masking (number of dimensions preserved), while for the linear embedding algorithms of Section 2, m provides the dimensionality of the embedding. Note that since the linear embeddings employ all d dimensions of the original data, the latter algorithms have

[1] Additional simulations and datasets are presented in [11]. We would like to thank the authors of [2] for providing us with the eyeglasses dataset.

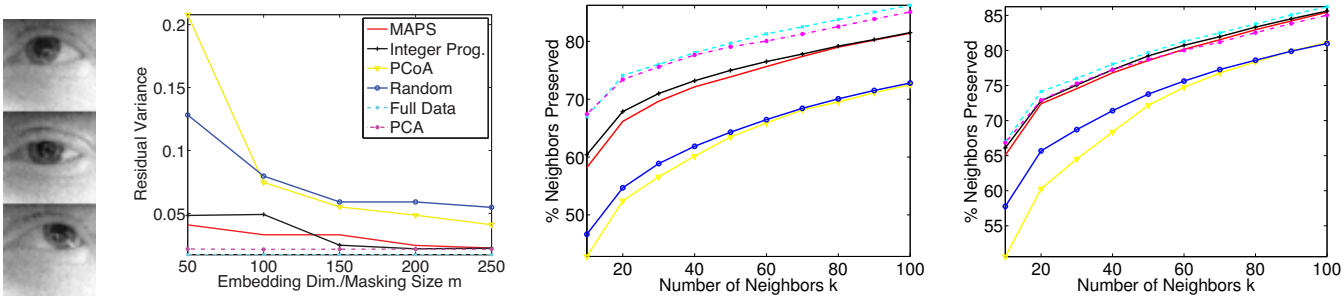

Fig. 1. Left: Example images from Eyeglasses dataset. Right: Performance comparison for linear embeddings (dashed lines) and masking algorithms (solid lines) with respect to original full-length data. Columns from left to right: residual variance as a function of m; percentage of preserved nearest neighbors for $m = 50$ and $m = 250$.

an intrinsic performance advantage against the former. The performance of random masking is averaged over 100 independent draws in each case.

The combinatorial nature of the integer program (2) renders it significantly expensive in computation. Thus, we choose to run the integer program for a limited time (24 hours) instead of running to completion. The remaining masking algorithms each take only up to 20 seconds (for $m = 250$) to complete using the same computing platform.

For each selection of masking/embedding algorithm, we apply Isomap directly on the masked images. We then check the performance of the manifold embedding obtained from the masked dataset to that of the manifold embedding from the full dataset using two different performance metrics.

First, we use *residual variance* as a global metric to measure how well the Euclidean distances in the embedded space match the geodesic distances in the ambient space. We pick the embedding dimensionality $\ell = 2$ to be the value after which the residual variance ceases to decrease substantially with added dimension. Note that the obtained value of ℓ agrees with the intuitive number of degrees of freedom for the Eyeglasses dataset (2-D gaze locations). Second, we use the percentage of preserved nearest neighbors, similar to [5]. More precisely, for a given neighborhood of size k, we obtain the fraction of the k-nearest neighbors in the full d-dimensional data that are among the k-nearest neighbors when the masked images are considered.

We display the results of manifold learning in Figure 1. MAPS and the integer program are shown only for the choice $p = 1$, as setting $p = \infty$ yields similar results. We observe that MAPS significantly outperforms random sampling and PCoA. Additionally, for small values of m the linear embedding algorithms of Section 2 can significantly outperform the masking algorithms of Section 3, which is to be expected since the latter approaches employ all d dimensions of the original data. More surprisingly, we see that for sufficiently large values of m the performance of MAPS approaches or matches that of the linear embedding algorithms, even though the embedding feasible set for masking methods is significantly reduced.

5. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] A. Mayberry, P. Hu, B. Marlin, C. Salthouse, and D. Ganesan, "ishadow: Design of a wearable, real-time mobile gaze tracker," in *MobiSys*, 2014.

[3] C. Hegde, M. Wakin, and R. Baraniuk, "Random projections for manifold learning," in *Neural Information Processing Systems (NIPS)*, Vancouver, BC, 2007, pp. 641–648.

[4] Y. Freund, S. Dasgupta, M. Kabra, and N. Verma, "Learning the structure of manifolds using random projections," in *Neural Information Processing Systems (NIPS)*, Vancouver, BC, Dec. 2007, pp. 473–480.

[5] C. Hegde, A. Sankaranarayanan, W. Yin, and R. Baraniuk, "A convex approach for learning near-isometric linear embeddings," *Submitted to J. Machine Learning Research*, 2012.

[6] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1395–1408, July 2009.

[7] Centeye, Inc., "Stonyman and Hawksbill vision chips," Available online at http://centeye.com/products/current-vision-chips-2, Nov. 2011.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[9] T. Cox and M. Cox, *Multidimensional Scaling*, Boca Raton : Chapman & Hall/CRC, 2001.

[10] C. Hegde, A. C. Sankaranarayanan, and R. G. Baraniuk, "Near-isometric linear embeddings of manifolds," in *Statistical Signal Processing Workshop (SSP)*, Aug. 2012.

[11] H. Dadkhahi and M. F. Duarte, "Masking strategies for image manifolds," *preprint available at http://www.ecs.umass.edu/ mduarte/images/MSIM2014.pdf*.

[12] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, NJ, 1982.

[13] MOSEK ApS, Denmark, "Mosek optimizer version 7," Available online at http://www.mosek.com, June 2013.

[14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Available online at http://cvxr.com/cvx, May 2013.