Dictionary Learning on Riemannian Manifolds

Yuchen Xie Baba C. Vemuri* Jeffrey Ho

Department of CISE, University of Florida, Gainesville FL, 32611, USA
{yxie,vemuri,jho}@cise.ufl.edu

Abstract. Existing dictionary learning algorithms rely heavily on the assumption that the data points are vectors in some Euclidean space $\mathbb{R}^d$, and the dictionary is learned from the input data using only the vector space structure of $\mathbb{R}^d$. However, in many applications, features and data points often belong to some Riemannian manifold with its intrinsic metric structure that is potentially important and critical to the application. The extrinsic viewpoint of existing dictionary learning methods becomes inappropriate and inadequate if the intrinsic geometric structure is required to be incorporated in the model. In this paper, we develop a very general dictionary learning framework for data lying on some known Riemannnian manifolds. Using the local linear structures furnished by the Riemannian geometry, we propose a novel dictionary learning algorithm that can be considered as data-specific, a feature that is not present in the existing methods. We show that both the dictionary and sparse coding can be effectively computed for Riemannian manifolds. We validate the proposed method using a classification problem in medical image analysis. The preliminary results demonstrate that the dictionary learned using the proposed method can and does indeed provide real improvements when compared with other direct approaches.

1 Introduction

Dictionary learning, which seeks to find a collection of atoms for sparse representation of the input data, has been widely used in image recognition, classification and restoration (e.g., [2]). Under this model, each data point is assumed to be generated *linearly* using only a small number of atoms, and this linear sparsity assumption is responsible for much of its generalization power and success. However, the underlying linear process requires that the data points as well as the atoms are treated as vectors in some vector space $\mathbb{R}^d$, and the dictionary is learned from the input data using only the vector space structure (and its associated inner product). For many applications in compute vision and medical image analysis that involve data points belonging to some known Riemannian manifolds such as the space of symmetric positive-definite matrices [7], hyperspheres for parameterizing square-root densities [15], Stiefel and Grassmann manifolds [3], etc., the existing extrinsic approaches that completely ignore the potentially important intrinsic structure implied by the data is clearly inadequate and unsound. To remedy this deficiency and inadequacy, our paper takes a step in this direction by investigating the problem of extending dictionary learning framework to incorporate intrinsic geometry implied by the input data.

* This research was in part supported by the NIH grant NS066340 to BCV.

The applicability and suitability of applying existing dictionary learning methods to solve problems that have to deal with manifold-valued data can be two thorny issues to consider. First, as a prerequisite, the data manifold must admit an embedding into some $\mathbb{R}^d$ in order to be able to apply the existing dictionary learning methods. However, for most manifolds, such as Grassmann and Stiefel manifolds, there simply does not exist known canonical embedding into $\mathbb{R}^d$ (or such embedding is difficult to compute). Second, even in the case when the existing method can be applied, due to their extrinsic viewpoint, important intrinsic properties of the data may not be represented in the dictionary. This can be illustrated by a simple example that it is possible that two points x, y on the manifold $\mathcal{M}$ have a large geodesic distance separating them but under the embedding $i : \mathcal{M} \to \mathbb{R}^d$, $i(x), i(y)$ has a small distance in $\mathbb{R}^d$. Therefore, sparse coding using dictionary learned in $\mathbb{R}^d$ is likely to code $i(x), i(y)$ (and hence x, y) using the same set of atoms with similar coefficients. Clearly, this will be undesirable and unsatisfactory if the applications require tasks such as classification and clustering, for which one would prefer the sparse coding to reflect some degree of actual similarity (i.e., geodesic distance) between the two samples x, y.

While the above example provides the motivation for seeking an extension to the existing dictionary learning framework to the more general Riemannain setting, it is by no means obvious how the extension should be correctly formulated. Let $\mathcal{M}$ denote the Riemannian manifold on which a collection of data points $x_1, \cdots, x_n$ are given. At the minimum, the goal of dictionary learning on $\mathcal{M}$ is to compute a collection of atoms $\{a_1, \cdots, a_m\} \subset \mathcal{M}$, also points on $\mathcal{M}$, such that each data point x_i can be *generated* using only a small number of atoms (sparsity). In the Euclidean setting, this is usually formulated as

$$\min_{D, w_1, \cdots, w_n} \sum_{i=1}^{n} \|x_i - D w_i\|^2 + \mathbf{Sp}(w_i), \tag{1}$$

where D is the matrix with columns composed of the atoms a_i, w_i the sparse coding coefficients and $\mathbf{Sp}(w_i)$ the sparsity promoting term. One immediate technical hurdle that any satisfactory generalization needs to overcome is the lack of a global linear structure that will allow the data to be generated from the atoms. Instead, the Riemannian geometry provides only local linear structures through the Riemannian exponential and logarithm maps, and by moving to the more general Riemannian setting, we essentially trade the unique global linear structure with infinitely many local linear structures, which is the main source of the various technical difficulties present in our generalization. However, this diversity of linear structures also provides us with an opportunity to formulate the dictionary learning using *data specific* approach.

Specifically, we will formally modify each summand in Equation 1 so that the sparse coding of a data x_i with respect to the atoms $\{a_1, \cdots, a_m\} \subset \mathcal{M}$ is now obtained by minimizing

$$\min_{w_i} \| \sum_{j=1}^{m} w_{ij} \log_{x_i} a_j \|_{x_i}^2 + \mathbf{Sp}(w_i), \tag{2}$$

with the important affine constraint that $\sum_{j=1}^{m} w_{ij} = 1$, where $w_i = (w_{i1}, \ldots, w_{im})^T$. That is, we are using the Riemannian exponential and logarithm maps at each data point x to define the generative process, and the sparse approximation of a given data point is first computed in its tangent space $T_x\mathcal{M}$ and then realized on $\mathcal{M}$ by applying the exponential map. We remark that this formulation is entirely coordinate-independent since each $\log_{x_i} a_j$ is coordinate-independent, and Equation 2 can be minimized using any local chart and its associated basis for $T_{x_i}\mathcal{M}$ (with a result that will be independent of these choices). Furthermore, a subspace S in a given coordinate system is represented as an affine subspace in a different coordinate system with a different origin. In short, Equation 2 is the direct generalization of linear sparsity condition with the exception that now the origin has been moved to the data point x_i. Computationally, the resulting optimization problem using Equation 2 can be effectively minimized. In particular, the gradient of the cost function in some cases admits closed-form formulas and in general, they can be evaluated numerically to provide the required inputs for the gradient-based optimization algorithm on manifolds.

To validate the proposed method, we have applied the dictionary learning framework to a classification problem in medical image analysis. The feature vectors specific for this problem belong to a hyper-sphere, and the preliminary results show that the dictionary learned using the proposed framework can and does indeed provide real improvements when compared with other direct approaches. Finally, before embarking on presenting the details of the proposed method, we mention that for technical reasons, we will assume that the Riemannian manifold $\mathcal{M}$ under consideration is complete and for each point $x \in \mathcal{M}$, the Riemannian log map $\log_x$ at x can be defined on $\mathcal{M}$ outside of a subset of codimension at least one. This technical condition is to ensure that for almost every arbitrary point $y \in \mathcal{M}$, $\log_x y$ can be uniquely defined.

2 Dictionary Learning on Riemannian Manifolds

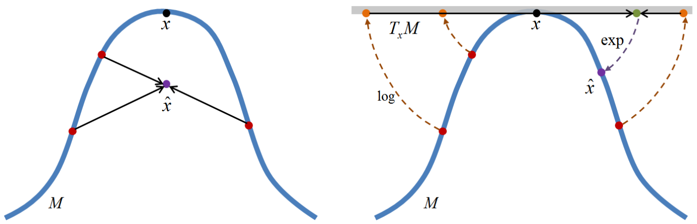

Fig. 1. The blue curve denotes a one-dimensional manifold $\mathcal{M}$. x is a data point on $\mathcal{M}$ and the red points represent the dictionary atoms. **Left:** the linear approximation $\hat{x}$ with the dictionary atoms may not be on the manifold $\mathcal{M}$. **Right:** our approach projects the atoms to the tangent space at x and performs the linear combination on the vector space $T_x\mathcal{M}$. The exponential map guarantees the approximation $\hat{x}$ is always on $\mathcal{M}$.

Given a collection of signals $x_1, \ldots, x_n \in \mathbb{R}^d$, classical dictionary learning methods [13, 2] try to find a dictionary $D \in \mathbb{R}^{d\times m}$ which includes m atoms such that each signal x_i can be represented as a sparse linear combination of these atoms $x_i \approx Dw_i$, where $w_i \in \mathbb{R}^m$. As others [10, 16], we formulate the dictionary learning problem using l_1 regularization on w_i

$$\min_{D,w_i} \sum_{i=1}^{n} \left(\|x_i - Dw_i\|_2^2 + \lambda \|w_i\|_1 \right) \tag{3}$$

where λ is a regularization parameter. Recent extensions of the classical dictionary learning include online dictionary learning [10] and using different regularization terms such as group-structured sparsity [8] and local coordinate constraint [17].

In this paper, we generalize the classical dictionary learning techniques in Euclidean space to Riemannian manifolds. Suppose $\mathcal{M}$ represents a Riemannian manifold. Let $x_1, \ldots, x_n \in \mathcal{M}$ be a collection of n data points on the manifold $\mathcal{M}$. Let $a_1, \ldots, a_m \in \mathcal{M}$ be atoms of the learned dictionary $\mathcal{D} = \{a_1, \ldots, a_m\}$. Because of the geometric structure of $\mathcal{M}$, we cannot use the linear combination of atoms $\hat{x}_i = \sum_{j=1}^{m} w_{ij} a_j$ to represent the data x_i, since $\hat{x}_i$ may not even be on the manifold $\mathcal{M}$. Thus we use the geodesic linear interpolation on Riemannian manifold, x_i can be represented by

$$\hat{x}_i = \exp_{x_i} \left(\sum_{j=1}^{m} w_{ij} \log_{x_i}(a_j) \right) \tag{4}$$

where $\exp_{x_i}$ and $\log_{x_i}$ are exponential and logarithmic maps at x_i, and $w_{ij} \in \mathbb{R}$ are weights. Note that in order to approximate data point x_i, we project all the atoms in the dictionary to the tangent space at x_i. Because $T_{x_i}\mathcal{M}$ is a vector space, we can perform linear combination $v_i = \sum_{j=1}^{m} w_{ij} \log_{x_i}(a_j)$ on $T_{x_i}\mathcal{M}$. Then the approximation $\hat{x}_i$ can be represented as the exponential map of v_i at x_i. A 1D case is illustrated in Figure 1. We hope to build a dictionary that minimizes the sum of reconstruction error for each data point. Here we define

$$\begin{aligned} E_{\text{data}} &= \sum_{i=1}^{n} \text{dist}(x_i, \hat{x}_i)^2 = \sum_{i=1}^{n} \| \log_{x_i}(\hat{x}_i) \|_{x_i}^2 \\ &= \sum_{i=1}^{n} \| \sum_{j=1}^{m} w_{ij} \log_{x_i}(a_j) \|_{x_i}^2 . \end{aligned} \tag{5}$$

By using the l_1 sparsity regularization, the dictionary learning on the manifold $\mathcal{M}$ can be written as the following optimization problem

$$\begin{aligned} &\min_{\mathbf{W},\mathcal{D}} \sum_{i=1}^{n} \| \sum_{j=1}^{m} w_{ij} \log_{x_i}(a_j) \|_{x_i}^2 + \lambda \|\mathbf{W}\|_1 \\ &s.t. \sum_{j=1}^{m} w_{ij} = 1, i = 1, \ldots, n \end{aligned} \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{n \times m}$ and the (i, j) entry of $\mathbf{W}$ is written as w_{ij}. The affine constraint $\sum_{j=1}^{m} w_{ij} = 1$ used in our formulation has also appeared in several recent papers on dictionary learning such as sparse subspace clustering [6] and local coordinate coding [17]. The affine constraint means that we are using affine subspaces to approximate the data instead of the usual subspaces, which are simply affine subspaces based at the origin. Moving away from vector spaces to general Riemannian manifolds, there is no corresponding notion of the origin that can be used to define subspaces, and this simple geometric fact requires the abandonment of the usual subspaces in favor of general affine subspaces. We can also introduce other regularizations to our dictionary learning framework instead of the l_1 norm. For example, the localization regularization used in local coordinate coding [17] can be generalized to $E_{\text{reg}} = \sum_{i=1}^{n} \sum_{j=1}^{m} |w_{ij}| \| \log_{x_i}(a_j) \|_{x_i}^2$. Similar to classical dictionary learning methods, we could use the iterative method to solve this optimization problem:

1. **Sparse coding step:** fix the dictionary $\mathcal{D}$ and optimize with respect to the coefficients $\mathbf{W}$.
2. **Codebook optimization step:** fix $\mathbf{W}$ and optimize with respect to the dictionary $\mathcal{D}$.

The first step becomes a regular sparse coding problem and many fast algorithms can be used to solve it. The second subproblem is much more challenging, because the dictionary atoms are on the manifold $\mathcal{M}$ and the optimization methods in Euclidean space are not appropriate any more.

In this paper, we present a line search based algorithm on Riemannian manifold to update the dictionary $\mathcal{D}$. Let $f(a_1, \ldots, a_m)$ be the cost function to be minimized. At first, we need to initialize the atoms in dictionary. One possible initialization is using m clusters of the data $x_1, \ldots, x_n$ generated by K-means algorithm on $\mathcal{M}$. Then we use the line search on manifold to optimize $f(a_1, \ldots, a_m)$. The basic idea is to find a descent direction v on the tangent space, and then walk a step along the geodesic γ whose initial velocity is v. The details are listed in Algorithm 1. The convergence analysis of the line search method on manifold is discussed in [1]. In the next section, we give a concrete example using our dictionary learning framework on square-root density function space.

3 Square-root Density Function Space

Probability density functions (pdfs) as a class of constrained non-negative functions have been widely used in many computer vision and medical imaging applications such as texture analysis and shape analysis. Without loss of generality, we restrict to the pdfs defined on the interval $[0, T]$ in this section: $\mathcal{P} = \{p : [0, T] \to \mathbb{R} | \forall s, p(s) \geq 0, \int_0^T p(s) ds = 1\}$. The Fisher-Rao metric has been introduced to study the Riemannian structure formed by the statistical manifold in [14]. For a pdf $p_i \in \mathcal{P}$, the Fisher-Rao metric is defined as $\langle v_j, v_k \rangle = \int_0^T v_j(s) v_k(s) \frac{1}{p_i(s)} ds$, where $v_j, v_k \in T_{p_i}(\mathcal{P})$. The Fisher-Rao metric is invariant to reparameterizations of the functions. In order to make the resulting manifold easy to compute on with Riemannian operations, the square root density representation $\psi = \sqrt{p}$ was used [15]. The space of square root density functions is defined as $\Psi = \{\psi : [0, T] \to \mathbb{R} | \forall s, \psi(s) \geq 0, \int_0^T \psi^2(s) ds = 1\}$. As we can see, Ψ forms a convex subset of the unit sphere in a Hilbert space. Then the Fisher-Rao metric can be obtained as $\langle v_j, v_k \rangle = \int_0^T v_j(s) v_k(s) ds$, where

Algorithm 1 Line search on Riemannian manifold

```
Input: A set of data $\mathcal{X} = \{x_1, \ldots, x_n\}$ on the manifold $\mathcal{M}$, coefficients $\mathbf{W} \in \mathbb{R}^{n \times m}$ and initial dictionary atoms $a_1^0, \ldots, a_m^0$.
Output: The optimal dictionary atoms $(a_1^*, \ldots, a_m^*)$ that minimize the cost function $f(a_1, \ldots, a_m)$.
  1. Set scalars $\alpha > 0, \beta, \sigma \in (0, 1)$ and initialize $k = 0$.
  2. Compute $\mathrm{grad} f(a_1^k, \ldots, a_m^k) = (\frac{\partial f(a_1^k)}{\partial a_1}, \ldots, \frac{\partial f(a_m^k)}{\partial a_m})$
  3. Pick $\eta^k = (\eta_1^k, \ldots, \eta_m^k) = -\mathrm{grad} f$, where $\eta_i^k \in T_{a_i^k}\mathcal{M}$.
  4. Find the smallest $s$ such that
         $f(\exp_{a_1^k}(\alpha\beta^s\eta_1^k), \ldots, \exp_{a_m^k}(\alpha\beta^s\eta_m^k)) \leq f(a_1^k, \ldots, a_m^k) - \sum_{i=1}^{m} \sigma\alpha\beta^s \|\eta_i^k\|_{a_i^k}.$
  5. Set $a_i^{k+1} = \exp_{a_i^k}(\alpha\beta^s\eta_i^k)$, $i = 1, \ldots, m$.
  6. Stop if $f$ does not change much, otherwise set $k = k + 1$ and go back to step 2.
```

$v_j, v_k \in T_{\psi_i}\Psi$ are tangent vectors. Given any two functions $\psi_i, \psi_j \in \Psi$, the geodesic distance between these two points is $\mathrm{dist}(\psi_i, \psi_j) = \cos^{-1}(\langle\psi_i, \psi_j\rangle)$, which is just the angle between ψ_i and ψ_j on the unit hypersphere. The geodesic at ψ_i with a direction $v \in T_{\psi_i}(\Psi)$ is defined as $\gamma(t) = \cos(t)\psi_i + \sin(t)\frac{v}{|v|}$. Then the exponential map can be represented as $\exp_{\psi_i}(v) = \cos(|v|)\psi_i + \sin(|v|)\frac{v}{|v|}$. To ensure the exponential map is a bijection, we restrict $|v| \in [0, \pi)$. The log map is then given by $\log_{\psi_i}(\psi_j) = u\cos^{-1}(\langle\psi_i, \psi_j\rangle)/\sqrt{\langle u, u\rangle}$, where $u = \psi_j - \langle\psi_i, \psi_j\rangle\psi_i$.

Using the geometric expressions of Ψ discussed above, we can perform the dictionary learning on square-root density functions. Let $x_1, \ldots, x_n \in \Psi$ be a collection of square-root density functions, and $a_i, \ldots, a_m \in \Psi$ be atoms in the dictionary $\mathcal{D}$. $\mathbf{W}$ is a $n \times m$ matrix. If we use l_1 regularization, our dictionary learning framework becomes

$$\begin{aligned} &\min_{\mathbf{W},\mathcal{D}} \sum_{i=1}^{n} \| \sum_{j=1}^{m} w_{ij} \cos^{-1}(\langle x_i, a_j\rangle) \frac{u_{ij}}{|u_{ij}|} \|_{x_i}^2 + \lambda\|\mathbf{W}\|_1 \\ &s.t. \sum_{j=1}^{m} w_{ij} = 1, i = 1, \ldots, n. \end{aligned} \tag{7}$$

where $u_{ij} = a_j - \langle x_i, a_j\rangle x_i$. This optimization problem can be efficiently solved using the algorithm presented in section 2.

4 Experiments

Let $x_1, \ldots, x_n \in \mathcal{M}$ be a collection of n training data. At first a codebook $\mathcal{D} = \{a_1, \ldots, a_m\}$ is learned from the training data by using the dictionary learning technique proposed in this paper, where each a_i denotes an atom in the dictionary $\mathcal{D}$. Then for each x_i, we generate the sparse code $\mathbf{w}_i \in \mathbb{R}^m$ based on the learned dictionary $\mathcal{D}$. Note that the sparse coding on Riemannian manifold is just one step in the dictionary learning algorithm with the codebook $\mathcal{D}$ fixed. Similar to [16], a linear SVM is trained for classification. In the testing stage, let $y_1, \ldots, y_k \in \mathcal{M}$ be a set of k testing data. The sparse coding can be performed on the testing data in a similar manner. Given each sparse code as a feature vector, the trained SVM model can be used to classify the testing data.

In our experiments, three methods are used for comparison purposes.

- Geodesic K-nearest neighbor (GKNN)
- Support Vector Machine (SVM) on vectorized data
- SVM on sparse codes with the codebook trained by KSVD

GKNN is the K-nearest neighbor classification using the Riemannian geodesic distance instead of the traditional Euclidean distance. It's commonly used in literature for classification on manifolds. K is set to 5 in our experiments. Since the feature vector for SVM should be in Euclidean space, we vectorize the data on manifolds and feed them to the SVM classifier. The LIBSVM [4] package is used in our experiment. We also apply the popular KSVD method [2] to train a dictionary for sparse coding, and then a linear SVM on sparse codes is employed in classification. By comparing the proposed method with the dictionary learning method in vector space, we try to discover whether keeping the Riemannian structure is important in the classification. We evaluate the dictionary learning algorithm for square root densities on the OASIS database [11]. OASIS contains T1 weighted MR brain images from a cross-sectional population of 416 subjects. Each MRI scan has a resolution of $176 \times 208 \times 176$ voxels. The ages of the subjects range from 18 to 96. We divide the OASIS population into three groups: young subjects (40 or younger), middle-aged subjects (between 40 and 60) and old subjects (60 or older). Because of the structural difference in the brain across different age groups, we can use the MR images to perform age group classification. At first, we align all the MR images in OASIS dataset using the nonrigid group-wise registration method described in [9]. For each image, we obtain a displacement field. Then the histogram of the displacement vectors is constructed in each image as the feature for classification [5]. In our experiment, the number of bins in each direction is set to $4 \times 4 \times 4$. Thus the 64 dimensional histogram is used as the feature vector for the KSVD and SVM classification, while the square root of the histogram is used in GKNN and our dictionary learning based method. The learned dictionary includes 100 atoms in our experiment. We use 5-fold cross validation and report the classification results in Table 1. Note that all four methods give pretty good classification accuracy for Young vs. Old and Middle-aged vs. Old, but our method outperforms the competing methods in the Young vs. Middle-aged experiment. Because the regional brain volume and cortical thickness of adults are relatively stable prior to reaching 60 years [12], the classification of young and middle-aged subjects is more challenging.

Table 1. The classification accuracy for different methods on the OASIS dataset.

	SVM	KSVD+SVM	GKNN	Proposed
Y vs. M	90.04	91.29	91.84	**98.34**
M vs. O	97.32	98.08	100	**100**
O vs. Y	98.12	99.46	100	**100**
Y vs. M vs. O	91.97	94.04	93.18	**98.62**

5 Conclusions

We have proposed a general dictionary learning framework for data on Riemannian manifolds. It's a novel generalization of the traditional dictionary learning methods on Euclidean space to manifolds. Specifically, we employ this dictionary learning framework on square-root density functions, which commonly appear in many computer vision and medical image analysis applications. We have provided experimental results on classification problems that validate the proposed algorithm, and in particular, comparisons with SVM, KSVD+SVM and geodesic K-nearest neighbor have shown that the proposed method using the learned dictionary on Riemannian manifolds provides real improvements.

References

1. Absil, P., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton Univ Pr (2008)
2. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. on Signal Processing 54(11), 4311–4322 (2006)
3. Cetingul, H., Vidal, R.: Intrinsic mean shift for clustering on stiefel and grassmann manifolds. In: CVPR. pp. 1896–1902 (2009)
4. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
5. Chen, T., Rangarajan, A., Vemuri, B.: Caviar: Classification via aggregated regression and its application in classifying oasis brain database. In: ISBI (2010)
6. Elhamifar, E., Vidal, R.: Sparse subspace clustering. In: CVPR. pp. 2790–2797 (2009)
7. Fletcher, P., Joshi, S.: Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Processing 87(2), 250–262 (2007)
8. Huang, J., Zhang, T., Metaxas, D.: Learning with structured sparsity. The Journal of Machine Learning Research 999888, 3371–3412 (2011)
9. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, S151–S160 (2004)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. The Journal of Machine Learning Research 11, 19–60 (2010)
11. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19(9), 1498–1507 (2007)
12. Mortamet, B., Zeng, D., Gerig, G., Prastawa, M., Bullitt, E.: Effects of healthy aging measured by intracranial compartment volumes using a designed mr brain database. MICCAI pp. 383–391 (2005)
13. Olshausen, B., Field, D.: Sparse coding with an overcomplete basis set: A strategy employed by v1? Vision research 37(23), 3311–3325 (1997)
14. Rao, C.: Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta Math. Soc 37(3), 81–91 (1945)
15. Srivastava, A., Jermyn, I., Joshi, S.: Riemannian analysis of probability density functions with applications in vision. In: CVPR (2007)
16. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: CVPR. pp. 1794–1801 (2009)
17. Yu, K., Zhang, T., Gong, Y.: Nonlinear learning using local coordinate coding. NIPS 22, 2223–2231 (2009)