DIMENSIONALITY REDUCTION WITH ADAPTIVE APPROXIMATION

Effrosyni Kokiopoulou and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute - ITS
CH- 1015 Lausanne, Switzerland
{effrosyni.kokiopoulou,pascal.frossard}@epfl.ch

ABSTRACT

In this paper, we propose the use of (adaptive) nonlinear approximation for dimensionality reduction. In particular, we propose a dimensionality reduction method for learning a parts based representation of signals using redundant dictionaries. A redundant dictionary is an overcomplete set of basis vectors that spans the signal space. The signals are jointly represented in a common subspace extracted from the redundant dictionary, using greedy pursuit algorithms for simultaneous sparse approximation. The design of the dictionary is flexible and enables the direct control on the shape and properties of the basis functions. Moreover, it allows to incorporate a priori and application-driven knowledge into the basis vectors, during the learning process. We apply our dimensionality reduction method to images and compare it with Principal Component Analysis (PCA) and Non-negative Matrix Factorization (NMF) and its variants, in the context of handwritten digit image recognition and face recognition. The experimental results suggest that the proposed dimensionality reduction algorithm is competitive to PCA and NMF and that it results into meaningful features with high discriminant value.

1. INTRODUCTION

Recent years have witnessed a large volume of high dimensional multimedia data. Generally, a pattern of interest is observed in a high dimensional ambient space but it is of much lower intrinsic dimension. For instance, all the possible appearances of a facial image span only a small part of the high dimensional image space. The purpose of feature extraction and dimensionality reduction techniques is to discover the intrinsic dimension of the data and to extract the low dimensional meaningful information from them.

Subspace analysis helps to reveal the latent low dimensional structures from the observed high dimensional data. One popular subspace method for dimensionality reduction is the Non-negative Matrix Factorization (NMF) [1], which has been proposed for learning a parts-based representation.

In contrast to the basis vectors of Principal Component Analysis (PCA) that are of global support, the basis vectors obtained from NMF are sparser, spatially localized and therefore more interpretable. A lot of variants of NMF (see e.g., [2, 4] and references therein) have been proposed, which attempt to provide more control over the properties of the basis vectors and/or coefficients vectors, by introducing additional (possibly non convex) constraints into the NMF optimization problem. However, in some cases, this results in sophisticated non-convex optimization problems that are hard to solve.

We propose a subspace method that formulates the dimensionality reduction problem as a matrix factorization problem, where the basis vectors are extracted from a redundant dictionary of localized basis functions. The flexibility in the design of the dictionary provides direct control on the shape and the properties of the basis functions, such as spatial locality and sparse support. Moreover, it provides naturally the potential to incorporate a priori and application-driven knowledge into the learning process, without resorting to sophisticated constraints, as is usually the case with NMF. During learning, the matrix factorization is obtained using greedy algorithms from simultaneous sparse approximation of signals. To the best of our knowledge, our paper is the first one to highlight the potential advantages of nonlinear approximation[1] to dimensionality reduction. Thus, it opens the way to the development of a new family of dimensionality reduction algorithms based on sparse nonlinear approximations. An additional advantage of the method is that compression and dimensionality reduction are jointly performed; hence, classification can be directly performed in the compressed domain. We provide experimental results on image recognition, which demonstrate that the proposed method is competitive with NMF and PCA.

2. DIMENSIONALITY REDUCTION USING SIMULTANEOUS APPROXIMATION

We assume the existence of a redundant dictionary $\mathcal{D}$ which spans the Hilbert space $\mathcal{H}$ of the signals of interest. A redun-

---

[1]Nonlinear approximation refers to the type of subspace approximation where the subspace is adapted to the signal that is approximated.

Algorithm: **SOMP**

Input: Signal matrix $S \in R^{m \times n}$ and `tol`: approximation error tolerance.

Output: Set of selected atoms Ψ, approximation A and residual matrix R.

1. Initialize the residual $R_0 = S$, $\Psi = []$, $t = 1$.
2. Find index γ_t by solving the optimization problem
   $$\max_{\gamma \in \Gamma} \|R_t^\top \phi_\gamma\|_1$$
3. Augment $\Psi = [\Psi, \ \phi_{\gamma_t}]$
4. Compute an orthonormal basis $V = [v_1, \ldots, v_t]$ of the span$\{\Psi\}$.
5. Compute the orthogonal projector $P_t = V_t V_t^\top$ on the span$\{\Psi\}$.
6. Compute the new approximation and residual
   $$A_t = P_t S$$
   $$R_t = (I - P_t)S$$
7. If $\|R\|_F \leq$ `tol`, then stop. Otherwise, increment iteration $t = t + 1$, and go to step (2).

Table 1. The SOMP algorithm.

dant dictionary

$$\mathcal{D} = \{\phi_\gamma, \ \gamma \in \Gamma\}$$

is an overcomplete basis i.e., it includes more vectors than the dimension of the subspace. Redundancy provides flexibility in the construction and in general boosts the approximation rate, especially for multidimensional signals [5]. The elements of the dictionary ϕ_γ are usually called atoms and are of unity norm. When the atoms are constructed by applying geometric transformations (e.g., translations, dilations and rotations) on a generating mother function ϕ, the dictionary is called *structured* and in this case, the index Γ indicates the parameters of the transformation. The following analysis holds for any redundant dictionary (i.e., overcomplete basis) spanning the signal space $\mathcal{H}$.

Consider a signal as an element of $\mathcal{H} \subseteq R^m$. We collect the training samples and form a signal matrix

$$S = [s_1, s_2, \ldots, s_n] \in \mathbb{R}^{m \times n}, \tag{1}$$

where s_i is the i-th column of S. For dimensionality reduction, our goal is to decompose S in the following form

$$S \approx \Psi C, \ \Psi \in \mathbb{R}^{m \times r}, \ C \in \mathbb{R}^{r \times n}, \tag{2}$$

where Ψ are the basis vectors drawn from the redundant dictionary and C are the corresponding coefficients. In other words, every column of S is represented in the same set of basis functions Ψ using different coefficients. This is a dimensionality reduction step where each signal (column of S) is represented in the subspace spanned by the columns of Ψ, using only $r \ll m$ coefficients.

If the columns of Ψ are spatially localized basis functions then the decomposition (2) results in a parts-based representation. Note that the design of the dictionary determines the properties of Ψ. Therefore, one has direct control on the shape and properties of the basis functions, thanks to the flexible design of the dictionary. Note that in NMF and its variants, one has only implicit control on the properties of the basis functions, accomplished via additional constraints introduced in the optimization problem.

If we denote by $\|\cdot\|_F$ the Frobenius norm, then we formulate the above problem as the following optimization problem.

> Optimization problem: **OPT1**
> $\min_{\Psi, C} \|S - \Psi C\|_F^2$
> subject to
> $\quad \Psi \subseteq \mathcal{D}$.

In order to solve OPT1, one may employ suboptimal algorithms that have been proposed in the context of joint signal compression by simultaneous sparse approximations [3]. We have chosen to use Simultaneous Orthogonal Matching Pursuit (SOMP) [3], since it lends itself as a computationally attractive algorithm for solving OPT1 in practice and enables a fast FFT-based implementation in the case of structured dictionaries. Note that other methods for simultaneous approximations could be potentially used.

SOMP is a greedy algorithm that extracts a subset Ψ of the dictionary, such that all the columns of S are simultaneously approximated. In each step, SOMP greedily selects the atom from the dictionary, which best matches all the residual signals at each iteration. Initially, SOMP sets the residual matrix $R = S$. Once the best matching atom ϕ_{γ_1} has been selected, the algorithm updates the residual matrix by projecting it on the orthogonal complement of ϕ_{γ_1}, and hence it removes its component from R.

In the next steps, the algorithm repeats the same procedure on the updated residual matrix. Thus, it greedily selects in step t, the best matching atom ϕ_{γ_t} by solving the simple optimization problem

$$\gamma_t = \max \arg_{\gamma \in \Gamma} \|R_t^\top \phi_\gamma\|_1,$$

and then it includes the selected ϕ_{γ_t} in Ψ. The residual matrix is updated by $R_{t+1} = (I - P)S$, where P is the orthogonal projector on the span$\{\Psi\}$. The main steps of the SOMP algorithm are summarized in Table 1. Note that the Orthogonal Matching Pursuit (OMP) converges in a finite number of iterations, since the norm of the residual decreases monotonically. This can be generalized to the case of SOMP [3]. Therefore, unlike the NMF algorithms, SOMP is not prone to local minima and it is not sensitive to initializations.

3. EXPERIMENTAL RESULTS

We apply our dimensionality reduction method to images and we build two different structured dictionaries where ϕ is (a)

Gaussian: $\phi(x, y) = \frac{1}{\sqrt{\pi}} \exp(-(x^2+y^2))$ and (b) *Anisotropic refinement* (AR) function $\phi(x, y) = \frac{2}{\sqrt{3\pi}}(4x^2-2)\exp(-(x^2+y^2))$ [5]. The AR function has an edge-like form and has been successfully used for image coding. For the construction of the atoms in each dictionary, we sample uniformly 10 orientation angles in $[0, \pi]$ and 5 logarithmically equi-distributed scales in $[1, N/4]$, where N is the image size. The translation parameters are discretized by pixel locations.

We compare experimentally the proposed method with NMF [1] and PCA. Given a data matrix $S \in \mathbb{R}^{m \times n}$ with non-negative entries, NMF seeks two non-negative factors $W \in R^{m \times r}$ and $H \in R^{r \times n}$ such that $S \approx WH$. The columns of the matrix W contain the basis vectors and H contains the corresponding coefficient vectors for the approximation of the columns of S. We use the MATLAB software package `nmfpack` [4] developed by P. Hoyer, which provides implementations for standard NMF as well as for Local NMF (LNMF) [2].

We experiment with the dimension of the reduced space $r = [10 : 10 : 50]$ (in MATLAB notation) and for each value of r, we report the classification performance in terms of average error rate across 50 random realizations of the training/test set. For classification, each training signal s_i is projected using the transpose of the matrix holding the basis vectors. The classification is accomplished in the reduced space by simple nearest neighbor (NN) classification.

Handwritten digit image recognition. We use the handwritten digit collection that is publicly available at S. Roweis web page[2]. This collection contains 20×16 bit binary images of "0" through "9", and each class contains 39 samples. We form the training set by a random subset of 10 samples per class and the remaining 29 samples are assigned in the test set. In the SOMP algorithm we use the Gaussian dictionary after preliminary experiments which show that, for this data set, it outperforms the AR dictionary. However, due to lack of space, we do not report these experiments here. Note also that this is one way to incorporate a priori and application-driven knowledge into the learning process. Figure 1(a) depicts the average classification error rate for various values of the dimension r of the reduced space. Observe that SOMP seems to be superior to its competitors and results in more discriminant features. The poor performance of PCA may be due the shortcoming of its global features to capture the spatial local characteristics of the digits.

We also examine the basis functions extracted from all methods. Figure 2 shows the recovered basis vectors from the digits data set, where dark pixels correspond to high values. The basis functions in sub-figures 2(a) and 2(b) are obtained from SOMP using the Gaussian and AR dictionary respectively. The figure also depicts the recovered basis functions from NMF and LNMF. Note that the features obtained from NMF resemble parts of digits but do not seem to be spatially

localized. On the contrary, the features of LNMF are spatially localized and they seem quite similar to the Gaussian atoms.
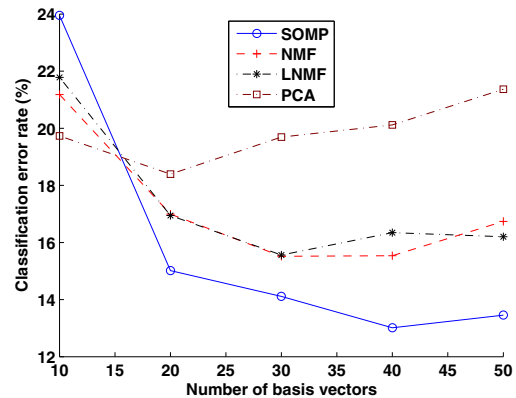
(a) Handwritten digits
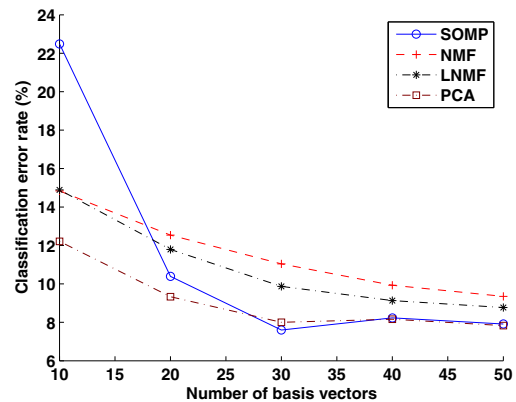
(b) Faces

Fig. 1. Image recognition experiments

Face recognition. We use the ORL (formerly Olivetti) database [6], which contains 40 individuals and 10 different images for each individual, including variation in facial expression (smiling/non smiling) and pose. We form the training set by a random subset of 5 different facial expressions per subject and use the remaining 5 as a test set.

Figure 1(b) depicts the average classification error rate across 50 random realizations of the training/test set. For this data set we use the AR dictionary, since the AR functions are more effective in representing the fine facial characteristics. Notice that SOMP is superior to the NMF algorithms and competitive to PCA. Also observe that for small dimensions r of the reduced space, the SOMP algorithm seems to be slightly inferior to its competitors, but as r increases, SOMP becomes more effective. This may occur due to the greedy nature of SOMP. In the first steps the SOMP algorithm typically selects atoms of large scale, in order to reduce quickly the approximation error. However, the large scale atoms usu-
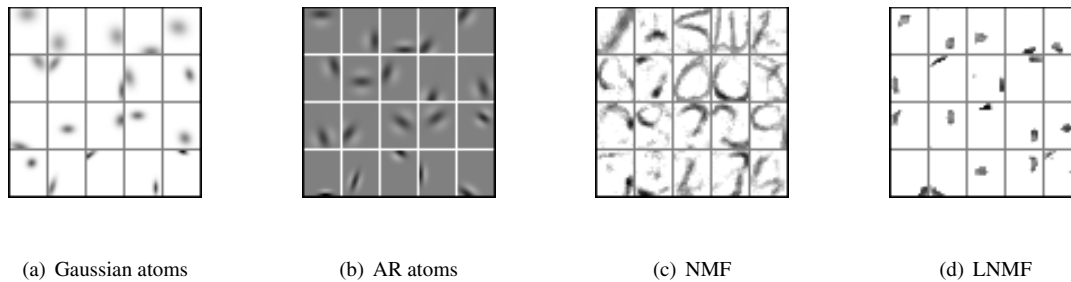
| (a) Gaussian atoms | (b) AR atoms | (c) NMF | (d) LNMF |

Fig. 2. Recovered basis vectors from the handwritten digit collection.

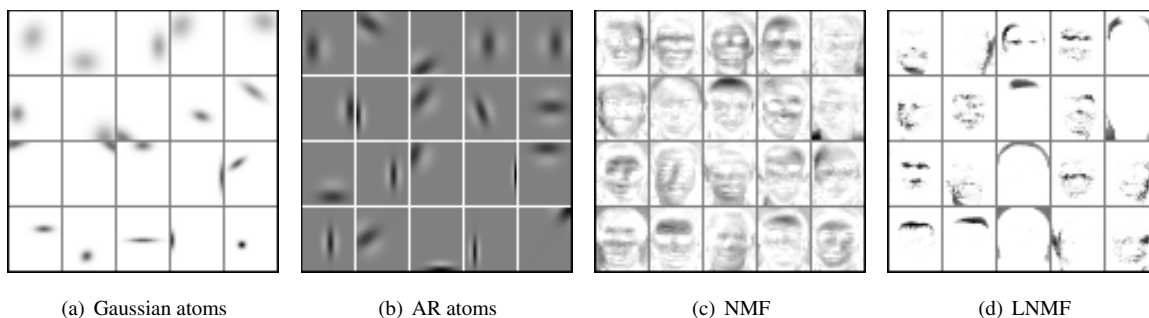| (a) Gaussian atoms | (b) AR atoms | (c) NMF | (d) LNMF |

Fig. 3. Recovered basis vectors from the ORL face data set.

ally carry low frequency information, which in general is not effective for the classification task.

Finally, Figure 3 shows the recovered basis vectors from the face data set. The basis functions in sub-figures 3(a) and 3(b) are obtained from SOMP using the Gaussian and AR dictionary respectively. Notice that the features obtained from NMF look like faces and seem to be of global support. On the contrary, the features of LNMF are more localized.

4. CONCLUSION AND FUTURE WORK

We have proposed the use of (adaptive) nonlinear approximation using redundant dictionaries, for dimensionality reduction. We formulate dimensionality reduction as a matrix factorization problem and use greedy pursuit algorithms from simultaneous sparse approximation, in order to solve it. The experimental results indicate the effectiveness of the proposed scheme and suggest that the extracted features are meaningful and of high discriminant value. We are currently working on designing improved (e.g., supervised) dimensionality reduction methods.

5. REFERENCES

[1] D.D. Lee and H.S. Seung, "Algorithms for Non-negative Matrix Factorization", *Advances in Neural Information Processing Systems*, vol. 13, pp. 556-562, 2001.

[2] S.Z. Li, X. Hou, H. Zhang and Q. Cheng, "Learning Spatially Localized, Parts-Based Representation", *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 1-6, 2001.

[3] J. Tropp, A. Gilbert and M. Strauss, "Algorithms for Simultaneous Sparse Approximation. Part I: Greedy pursuit", *Signal Processing*, special issue "Sparse approximations in signal and image processing" , vol. 86, April 2006, pp. 572-588.

[4] P. O Hoyer, "Non-negative Matrix Factorization with Sparseness Constraints", *Journal of Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.

[5] R. Figueras i Ventura, P. Vandergheynst and P. Frossard, "Low Rate and Flexible Image Coding with Redundant Representations", *IEEE Transactions on Image Processing*, vol. 15(3), pp. 726-739, March 2006.

[6] F. Samaria and A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification", *2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, Dec. 1994.