SCIENCE CHINA
Information Sciences

• HIGHLIGHT •

February 2017, Vol. 60 028102:1–028102:3
doi: 10.1007/s11432-015-0960-8

Principal basis analysis in sparse representation

Hong SUN[1,2*], Chengwei SANG[1,3] & Chenguang LIU[1,2]

[1]*School of electronic Information, Wuhan University, Wuhan* 430072*, China;*
[2]*Department Signal and Image Processing, Telecom ParisTech, Paris* 75013*, France;*
[3]*Department of Scientific Research, Air Force Early Warning Academy, Wuhan* 430019*, China*

Received March 14, 2016; accepted July 27, 2016; published online December 20, 2016

Citation Sun H, Sang C W, Liu C G. Principal basis analysis in sparse representation. Sci China Inf Sci, 2017, 60(2): 028102, doi: 10.1007/s11432-015-0960-8

We propose a principal basis analysis (PBA) method over a learned overcomplete basis set (called dictionary). The proposed PBA makes use of a novel criterion: the frequency of basis vector (atom of the dictionary) over a data set. This proposed criterion is well adapted to sparse signal representation. Moreover it adapts to an intrinsic characteristic of the regularity of signals.

In signal analysis, principal component analysis (PCA) [1] is a powerful tool. PCA is to identify the most meaningful basis over an orthonormal base. This basis will hopefully filter out the noise and reveal hidden structure of signal. PCA supposes that the components of the true signal behind the data have a maximum variance and the other components are mainly due to noise. However, in many practical cases, some components with low variance might actually be important relative to signal details, and some types of noise, such as non-Gaussian statistics noise, might actually have a significant variance.

In recent years, a more effective signal analysis method, sparse representation, based on adaptive overcomplete bases (dictionaries) [2], has attracted significant interest. The underlying assumption is that a meaningful signal could be represented by combining few principal components. We note that a dictionary is obtained by learning with a noisy data set. Therefore a principal signal base must be identified from the dictionary. However, it is impossible to exploit an energy-constrained signal bases since the base vectors of the dictionary (called atoms) are not necessarily orthogonal or independent. Under this circumstance, we propose a criterion of frequency of atom to identify signal principal bases.

Principal basis analysis based on frequency of sparse signal components. The sparse decomposition of data set $\{\boldsymbol{x}_m \in \mathbb{R}^N\}_{m=1}^M$ with a dictionary $\boldsymbol{D} = \left[\{\boldsymbol{d}_k \in \mathbb{R}^N\}_{k=1}^K\right]$ ($K > N$, overcomplete) can be given [2] as

$$\begin{aligned}\boldsymbol{X}_{N\times M} &\simeq \boldsymbol{D}_{N\times K}\boldsymbol{A}_{K\times M}\\ &= [\boldsymbol{d}_1,\ldots,\boldsymbol{d}_k,\ldots,\boldsymbol{d}_K]\\ &\quad\cdot[\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_m,\ldots,\boldsymbol{\alpha}_M].\end{aligned} \tag{1}$$

By learning from observed data matrix $\boldsymbol{X} = [\{\boldsymbol{x}_m\}_{m=1}^M]$, K optimal atoms $\boldsymbol{d}_k$s and M sparse coefficient vectors $\boldsymbol{\alpha}_m$s can be obtained by solving M forms as

$$\begin{aligned}\{\boldsymbol{D},\boldsymbol{\alpha}_m\} &= \underset{\boldsymbol{D},\boldsymbol{\alpha}_m}{\operatorname{argmin}} \parallel \boldsymbol{\alpha}_m \parallel_0\\ &\quad + \parallel \boldsymbol{D}\boldsymbol{\alpha}_m - \boldsymbol{x}_m \parallel_2^2\\ &\leqslant \varepsilon, \quad 1 \leqslant m \leqslant M,\end{aligned} \tag{2}$$

where ε is allowed error tolerance and the column vectors $\boldsymbol{\alpha}_m = [\{\alpha_m(k)\}_{k=1}^K]^{\mathrm{T}} \in \mathbb{R}^K$. K-SVD algorithm [3] is a typical way to solve (2).

* Corresponding author (email: hongsun@whu.edu.cn)

The authors declare that they have no conflict of interest.

The sparse decomposition (Eq. (1)) takes large number of patterns $\boldsymbol{d}_k$s as signal components with the fewest number of nonzero coefficients in each of sparse code vectors $\boldsymbol{\alpha}_m$s. We intend to choose the most meaningful atoms from the overcomplete dictionary to form a principal base.

Consider the sparse coefficient matrix $\boldsymbol{A}_{K\times M}$ in terms of row vectors $\boldsymbol{\lambda}_k = \left[\{\alpha_m(k)\}_{m=1}^M\right] \in \mathbb{R}^{1\times M}$.

$$\begin{aligned}\boldsymbol{A} &= [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_M] \\ &= \begin{bmatrix} \alpha_1(1) & \alpha_2(1) & \cdots & \alpha_M(1) \\ \alpha_1(2) & \alpha_2(2) & \cdots & \alpha_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1(K) & \alpha_2(K) & \cdots & \alpha_M(K) \end{bmatrix} \\ &= [\boldsymbol{\lambda}_1 \quad \boldsymbol{\lambda}_2 \quad \cdots \quad \boldsymbol{\lambda}_K]^{\mathrm{T}}.\end{aligned} \tag{3}$$

Thus Eq. (1) becomes

$$\begin{aligned}\boldsymbol{X} \simeq \boldsymbol{DA} = &[\boldsymbol{d}_1, \ldots, \boldsymbol{d}_k, \ldots, \boldsymbol{d}_K] \\ &\cdot \left[\boldsymbol{\lambda}_1^{\mathrm{T}}, \ldots, \boldsymbol{\lambda}_k^{\mathrm{T}}, \ldots, \boldsymbol{\lambda}_K^{\mathrm{T}}\right]^{\mathrm{T}}.\end{aligned} \tag{4}$$

The row vector $\boldsymbol{\lambda}_k$ is not necessarily sparse. Eq. (4) means that $\boldsymbol{\lambda}_k$ is the weight coefficient of the learned atoms $\boldsymbol{d}_k$. Its ℓ^0 zero pseudo-norm $\|\boldsymbol{\lambda}_k\|_0$ which counts the non-zero entries indicates the occurrence frequency of $\boldsymbol{d}_k$ over the data set $\{\boldsymbol{x}_m\}_{m=1}^M$. In fact, $\|\boldsymbol{\lambda}_k\|_0$ the frequency of atom $\boldsymbol{d}_k$ is an intrinsic feature of natural signal. A principal component in the sparse decomposition must occur in a meaningful signal with higher frequency even with a lower energy. By contrast, any noise pattern hardly repeats in observed data even with a higher energy. Therefore, we propose a principal basis analysis (PBA) method under the criterion of atom's frequency.

Sort vectors $\{\boldsymbol{\lambda}_k\}_{k=1}^K$ in an ascending order of their ℓ^0 -norms $\{\|\boldsymbol{\lambda}_k\|_0\}_{k=1}^K$ represented by

$$\begin{aligned}&\{\boldsymbol{\lambda}'_1, \ldots, \boldsymbol{\lambda}'_k, \ldots, \boldsymbol{\lambda}'_K\} \\ &\quad \xleftarrow{\text{sort}} \{\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_k, \ldots, \boldsymbol{\lambda}_K\},\end{aligned} \tag{5}$$

where $\|\boldsymbol{\lambda}'_1\|_0 \geqslant \|\boldsymbol{\lambda}'_2\|_0 \geqslant \cdots \geqslant \|\boldsymbol{\lambda}'_K\|_0$.

Corresponding to the order of $\{\boldsymbol{\lambda}'_k\}_{k=1}^K$, the reordered atom set is written as $\{\boldsymbol{d}'_1, \ldots, \boldsymbol{d}'_k, \ldots, \boldsymbol{d}'_K\}$. We take the first P atoms $\{\boldsymbol{d}'_1, \boldsymbol{d}'_2, \ldots, \boldsymbol{d}'_P\} (P \ll K)$ as principal bases of signal. An estimate of the underlying signal $\boldsymbol{S}$ embedded in the observed data $\boldsymbol{X}$ can be obtained simply by linear combination as

$$\hat{\boldsymbol{S}}_{N\times M} = [\boldsymbol{d}'_1, \boldsymbol{d}'_2, \ldots, \boldsymbol{d}'_P] \cdot [\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \ldots, \boldsymbol{\alpha}'_M], \tag{6}$$

where $[\boldsymbol{\alpha}'_1, \boldsymbol{\alpha}'_2, \ldots, \boldsymbol{\alpha}'_M]_{P\times M}$ comes from the coefficient matrix $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_M]_{K\times M}$ through reordering and truncating corresponding to $\{\boldsymbol{d}'_k\}_{k=1}^P$.

The dimension P of the signal principal base in Eq. (6) could be simply decided according to the histogram of $\{\|\boldsymbol{\lambda}'_k\|_0\}_{k=1}^K$. One can set the maximum point of the histogram to P defined as

$$P = \operatorname*{argmax}_k \operatorname{Hist}\left(\|\boldsymbol{\lambda}'_k\|_0\right). \tag{7}$$

Experiment. Apply the proposed PBA to image denosing with Gaussian noise and to SAR image despeckling with non-Gaussian noise model. Take noiseless images and noise images shown in Figure 1(a) to the experiments. We use the peak signal-to-noise ratio (PSNR) to assess the noise removal performance, and the structural similarity index metric (SSIM) to evaluate the preserving details performance. In our experiments, dictionaries used are of size 64×256 (256 atoms). Figure 1(b) illustrates that the proposed criterion of atom's frequency measured by ℓ^0-norm $\|\boldsymbol{\lambda}_k\|_0$ would be more robust for distinguishing signal details from noise under the threshold P (right of Figure 1(b)) than commonly used strength criterion [4] measured by ℓ^2-norm $\|\boldsymbol{\lambda}_k\|_2$ (left of Figure 1(b)). Figure 1(c) presents the results of denoising for noisy image with an additive white Gaussian noise. Compared with K-SVD method [2], which is one of the best denoising methods reported in the recent literatures, the PBA method outperforms by about 0.5 dB in PSNR and by about 35% in SSIM. We can see that from Figure 1(c) that the corner of Lena's mouth with weak energy (lower $\|\boldsymbol{\lambda}_k\|_2$) is better recovered by PBA method owing to the atom $\boldsymbol{d}_k$ in Figure 1(b) with higher frequency (higher $\|\boldsymbol{\lambda}_k\|_0$) for the region of corner of mouth. Figure 1(d) presents the results of despeckling for simulated 1-look SAR image with a multiplicative speckle noise. Compared with a probabilistic patch based (PPB) filter based on nonlocal means approach [5], which can cope with non-Gaussian noise, the PBA method shows huge advantages at preserving fine details and suppressing strong noise.

Conclusion. A central idea of the proposed principal basis analysis (PBA) is to identify a principal signal bases from an adaptive overcomplete dictionary based on the proposed criterion of the frequency of atom.

PBA has combined the variance criterion (used in PCA), the sparsity criterion (used in sparse representation) and the component's frequency criterion into a uniform frame (just the same frame of PCA).

PBA takes two steps: building a higher dimensional dictionary by learning, and identifying a lower dimensional principal base by truncating. It can clearly improve performances by many direct

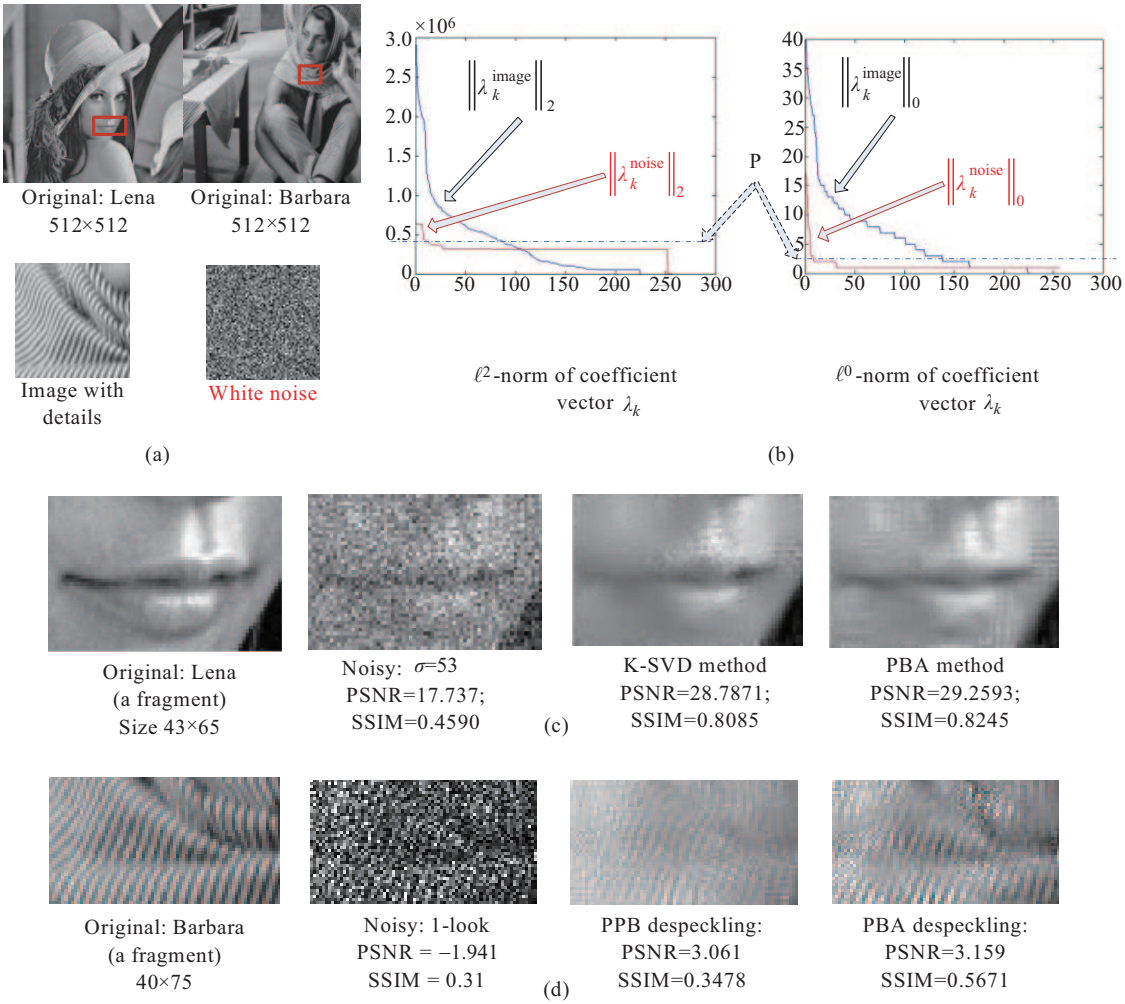

Figure 1 (Color online) (a) Experimental data; (b) strength criterion and atom's frequency criterion; (c) denoising results for additive noise; (d) despeckling results for multiplicative noise.

low-rank approximation methods [6] at preserving weak details of information and at suppressing strong noise. Furthermore, PBA method is so simple with a linear retrieval operation (Eq. (6)).

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 60872131). The idea of the principal basis analysis presented here arises through a lot of deep discussions with Professor Henri Maître at Telecom-ParisTech in France. We are also grateful to Prof. Didier Le Ruyet at CNAM in France for many fruitful discussions.

References

1 Pearson K. On lines and planes of closest fit to systems of points in space. Philos Mag, 1901, 2: 559–572
2 Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process, 2006, 15: 3736–3745
3 Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process, 2006, 54: 4311–4322
4 Zhao Q, Meng D Y, Xu Z B. Robust sparse principal component analysis. Sci China Inf Sci, 2014, 57: 092115
5 Deledalle C A, Denis L, Tupin F. Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. IEEE Trans Image Process, 2009, 18: 2661–2672
6 Zhang Z, Li F, Zhao M, et al. Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification. IEEE Trans Image Process, 2016, 25: 2429–2443