International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.3, 2015

RESEARCH ARTICLE

ISSN: 2321-7758

LEARNING OF FINITE DIFFERENCE ANALYSIS OPERATOR THROUGH ANALYSIS K-SVD

Dr. INDUMATHI T S[1], ABHILASH T B[2]
[1]PROFESSOR & HEAD, [2]STUDENT ,
DEPARTMENT OF DIGITAL ELECTRONICS & COMMUNICATION SYSTEMS
VTU-CENTER FOR POST GRADUATION STUDIES, BANGALORE REGION

Article Received:12/05/2015 | Article Revised on:20/05/2015 | Article Accepted on:29/05/2015

ABSTRACT

Dictionary learning from the large set of training examples has drawn a considerable interest among many researchers. Out of two variants of sparse signal model, synthesis and analysis based, a much attention was given to synthesis signal model whereas analysis model has not been touched upon .In the analysis signal model, the product of analysis operator Ω and the signal x is assumed to be sparse. The analysis dictionary, also called as analysis operator, is much more redundant than synthesis signal model in the sense that the rows of dictionary is greater than the columns of dictionary.Here, in this work, we consider a dictionary developed by applying horizontal and vertical derivatives on a 2D signal of size$\sqrt{d}x\sqrt{d}$, called as finite difference analysis operator .The operator is made to contaminate with Gaussian noise and large set of training examples are created. The analysis KSVD is applied to learn the dictionary which exhibits strong linear dependencies between the rows and the learned dictionary is used for denoising application.

Keywords: dictionary, finite difference analysis operator, linear dependencies

I. INTRODUCTION

The main aim of this work is to develop an analysis dictionary from a large set of training examples. The task extends its hand to various branches such as optimisation techniques, linear algebra, approximation theory, signal processing, harmonic analysis and some concepts of statistical learning. The typical training of the analysis signal model consists of three major steps [1][2].They are Initialisation of dictionary, Sparse coding stage and Update dictionary stage.

At a very first stage , the dictionary Ω is initialised with the help of off-the-shelf dictionary Relaxation methods and Greedy algorithms are the two approaches employed in sparse coding stage to approximately estimate the recovered signal. In relaxation approach, tight frame on the constraints is made to relax so that the combinatorial problem is not NP-hard anymore [3] [4]. Interior point methods, Iterative shrinkage, Sequential shrinkage for union of orthobases are quite known good solvers in relaxation methods .In greedy algorithms, the inner product is obtained between the atoms of dictionary and the signal to find the best fit support in the dictionary. Once the best fit is obtained, the support is updated in both in sparse vector as well as in dictionary .The process is continued until the denoising error falls below the desired threshold.

II. Problem formulation

Usually, if the given signal x is not contaminated from any type of noise, one can readily compute the co-efficient for Ωx[2], however ,if the signals are assumed to be distorted with additive noises say v, the calculations stands no longer simple. Let v be a vector which denotes the additive white Gaussian noise with zero mean and standard deviation σ ,then signal y is the sum of original signal x contaminated with noise v

y=x+v (1.1)

Recovering the proper signal $\hat{x}$ can be viewed as denoising problems and it is no longer simple and can be formulated either based on co-rank measure or aiming for minimising the error between the true signals and their noisy versions. Based on the required co-rank for $\Omega_{\Lambda,}$ the formulation takes the form as[9]

$$\{\hat{x}, \hat{\Lambda}\} = \underset{\hat{x},\hat{\Lambda}}{\mathrm{Argmin}}\|x - y\|_2 \quad \text{Subject to}$$

$\Omega_\Lambda X=0$, Rank (Ω_Λ) =d-r (1.2)

Where Λ denotes the set of $|\Lambda|$ rows that are orthogonal to to analysis dictionary, termed as co-support of signal x and Ω_Λ denotes the sub matrix which contains only the rows that are indexed in Λ.

The above constraints when formulated based on required error tolerance ε, the equation (1.2) takes the form,

$$\{\hat{x}, \hat{\Lambda}\} = \underset{\hat{x},\hat{\Lambda}}{\mathrm{Argmin}}\,\mathrm{Rank}\,(\Omega_\Lambda) \quad \text{Subject to}$$

$\Omega_\Lambda X=0$, $\|x - y\|_2 \leq \varepsilon$ (1.3)

Based on availability of information, either equation (1.2) or equation (1.3) can be employed to approximate the clear versions of the original signal. At this context, we employ pursuit algorithms for estimation of co-support and the clear signal. The detailed explanation of pursuit algorithm that we employ is clearly explained in [9].

III. Analysis K-SVD: The Learning Algorithm

TASK: To develop an analysis operator Ω and to estimate signal x by minimising

$$\{\hat{\Omega}, \hat{X}\{\hat{\Lambda}_i\}_{i=1}^{K}\} =$$

$$\mathrm{Arg\,min}_{\Omega,X,\{\Lambda_i\}_{i=1}^{R}}\|X - Y\|_F^2 \text{ subject to}$$

$\Omega_{\Lambda_i}x_i$=0, Rank (Ω_{Λ_i}) =d-r, $\|w_j\|_2 = 1$

- **INPUT**: Training signals Y ϵ R^{dxR} ,initial dictionary Ω_0 ϵ R^{pxd} ,target co-rank d-r and stopping criteria(Number of iterations)
- **INITIALISATION**: Set the dictionary $\Omega:=\Omega_0$

Repeat for the k-iterations (stopping criteria)

Apply block co-ordinate descent algorithm

➢ **SPARSE CODING STAGE:**

Use either backward greedy algorithm or optimised backward greedy algorithm and formulate the optimisation task based on co-rank measure.

➢ **CODEBOOK UPDATE STAGE:**

-Extract relevant columns from the signal that are orthogonal to rows of analysis operator

-Compute $\hat{w}_j = \mathrm{Arg\,min}_{w_j}\|w_j^T Y_j\|_2^2$ subject to $\|w_j\|_2 = 1$

-update rows of dictionary Ω as Ω{j-th row}:=W_j^T

-repeat the above for p times

Algorithm 1;Analysis KSVD

Let Y=[y_1, y_2 ,y_3, y_4 ....y_R]ϵ R^{dxR} denotes the training set and every sample in a training is assumed to be noisy residing in the r-dimensional subspace related to dictionary Ω.Thus ,each sample y_i is contaminated with additive white Gaussian noise vector v_i which has zero mean i.e. $y_i = x_i + v_i$.

Each x_i satisfies a co-rank of d-r with respect to dictionary Ω and the optimisation task can be formulated as

$$\{\hat{\Omega}, \hat{X}\{\hat{\Lambda}_i\}_{i=1}^{K}\} =$$

$$\mathrm{Arg\,min}_{\Omega,X,\{\Lambda_i\}_{i=1}^{R}}\|X - Y\|_F^2 \text{ subject to}$$

$\Omega_{\Lambda_i}x_i$=0 ,Rank(Ω_{Λ_i})=d-r, $\|w_j\|_2 = 1$ (1.4)

Where x_i are the estimates of the noiseless signals which are arranged in the columns of matrix X and

Λ_i are their co-supports.The vectors w_j are the rows of Ω.

Block co-ordinate descent algorithm[6] is an iterative approximation technique which is simple to implement and provide faster convergence especially when the variables are loosely coupled, it is observed that the cost per iteration is very low when compared with gradient descents. They find their applications in areas such as low-rank matrix recovery, sparse dictionary learning, blind source separation, non-negative matrix factorisation non-negative tensor factorisation and many more.

As shown in algorithm1,at this set up, the analysis operator is learned in two phases by assuming the initial operator and continuing the two phases to a fixed number of iterations.

Phase I-Sparse coding stage (optimisation for x by keeping Ω fixed).

In this phase, the initial dictionary Ω_0 is assumed and optimisation is done individually by considering the each columns of X.The approximation problem is defined based on co-rank analysis for each signal y_i as equation below and any pursuit methods discussed in chapter 2 can be employed to compute noiseless signals and co-supports as well.

$\{\hat{x}_i, \hat{\Lambda}_i\} = \mathrm{Arg\ min}_{x_i,\Lambda_i} \|x_i - y_i\|_2$ Subject to $\Omega_{\Lambda_i} x_i$=0 ,Rank(Ω_{Λ_i})=d-r (1.5)

Phase II-codebook update stage(Update of Ω with computed $\hat{X}$ in phase I as input)

In this phase, the optimisation is carried out for each w_j rows of analysis operator Ω in a sequential manner. The update of each rows should be performed only those columns of $\hat{X}$ which are orthogonal to it with no influence to other signals.If $\hat{X}_j$ and Y_J denotes the sub matrix of X and Y (in columns),which are orthogonal to the rows of analysis operator w_j , then update for w_j is written as

$$\{\hat{w}_j, \hat{x}_J\} = \mathrm{Arg} \min_{w_j, X_J} \|X_J - Y_J\|_F^2$$

Subject to $\Omega_{\Lambda_i} x_i$=0 ,Rank(Ω_{Λ_i})=d-r,

$$\|w_j\|_2 = 1 \quad (1.6)$$

Here, the representation co-supports are also fixed and if Ω_i denotes the submatrix of Ω which contains only the rows from Ω that are orthogonal to x_i at that instant except the row w_j then the optimisation task can be formulated as

$$\{\hat{w}_j, \hat{x}_J\} = \mathrm{Arg} \min_{w_j, X_J} \|X_J - Y_J\|_F^2$$

Subject to $\Omega_{\Lambda_i} x_i$=0 ,$w_j^T X_J = 0$

$$\|w_j\|_2 = 1 \quad (1.7)$$

The equation (1.7)clearly indicates that the codebook update stage just involves only the co-supports and also the equation (1.7) is quite difficult to solve for the rows of analysis operator w_j. .The alternative approach is given by

$\hat{w}_j = \mathrm{Arg\ min}_{w_j} \|w_j^T Y_J\|_2^2$ Subject to

$$\|w_j\|_2 = 1 \quad (1.8)$$

The solution for the above problem is the singular vector corresponding to the smallest singular value of Y_J and can be efficiently obtained through singular value decomposition of Y_J.[5].

IV Results and discussions

The experiment consists of verifying the analysis K-SVD algorithm for the finite difference analysis operator Ω_{DIF}. The subspace dimension for each signal r is 4 and the signal dimension d=25 is given as input and the experiment is carried out for 5000 training signals. At the very first step, generate the $\Omega_{DIF} \epsilon\ R^{50x25}$ and generate analysis signal with a generated dictionary Ω_{DIF} with a given subspace dimension of the signal (d-r).Contaminate the signal with additive white Gaussian noise with zero mean and standard deviation $\sigma = .2/\sqrt{d}$.In this set up,the evaluation is performed for standard deviation equal to 04.Apply analysis- KSVD algorithm for fixed number of iterations. In this case, rank based optimised greedy algorithm is chosen and is experimented for 100 iterations. Find representation error and denoising error and plot the same with respect to number of iterations. If X denotes the noisy signal,X_{est} is the estimated signal T_{rue} is my original signal, then representation error per element is given by $\|X - Xest\|_F / \sqrt{dN}$ and

denoising error is calculated as

$$\|X - Xtrue\|_F / \sqrt{dN}. \quad [9]$$

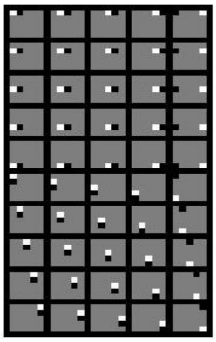


Figure 1 .The Ω_{DIF},of size 2dxd(d=25)

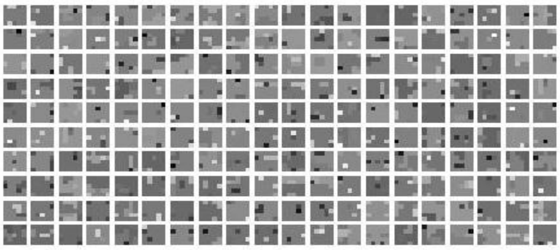


Figure 2 .Sparse Analysis Signals Each Of 5x5 Patches

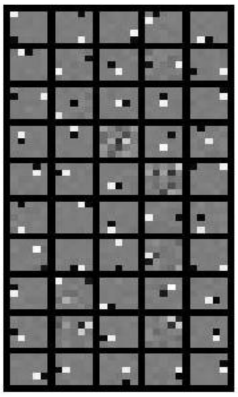


Figure 3 the analysis dictionary developed for a noisy condition for phase 1(75 iterations) for 5000 training samples

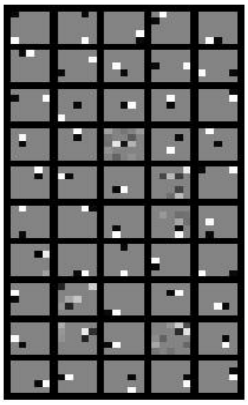


Figure 4 the analysis dictionary developed for a noisy condition for phase 2(25 iterations) for 5000 training samples

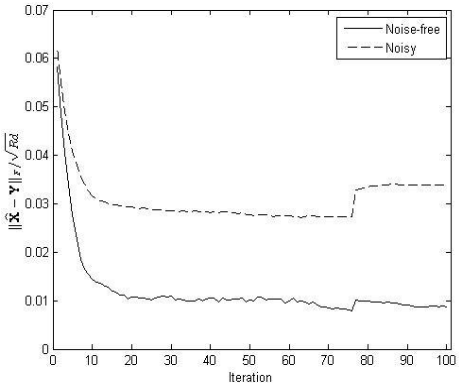
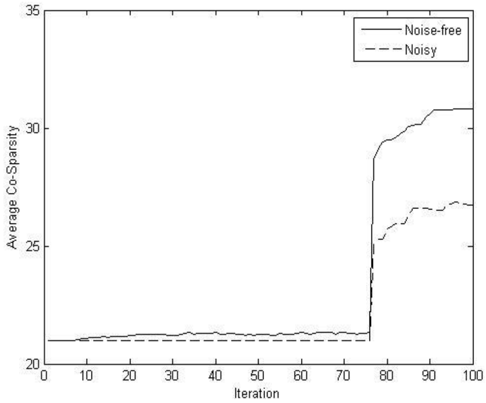


Figure 5 Average Representation Error per element for 5000 training samples



Fi
gure 6 Measure of Average Co-sparsity against the number of iterations for noisy and noise free versions of signals for 5000 training samples

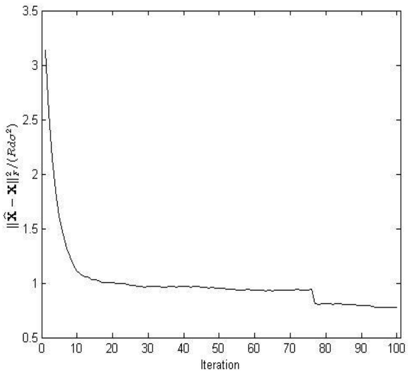


Figure 7 Representation Error obtained for noisy versions of signals with respect to number of iterations for 5000 training samples

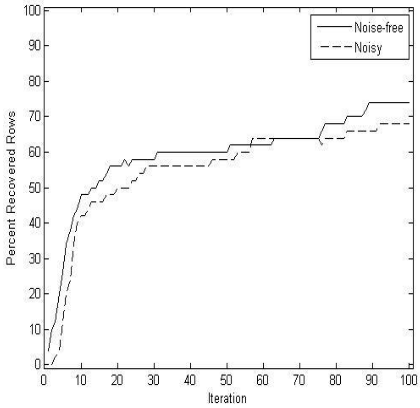


Figure.8 Percentage of recovered rows against the number of iterations for noisy and noiseless version of signals for 5000 training samples

The iterations were split into two phases where phase 1 constitutes 75 iterations and last 25 iterations are carried out in phase 2.Both the phases follows the atom update rule as in algorithm 1,and phase 2 iterations go with simple modification just to encourage linear dependence in the dictionary. In phase 2,the near zero entries in analysis dictionary are nullified and the mean value whose value is near zero is removed from the rows..These two post processing operations are performed to promote linear independencies in the dictionary and they can be recovered in a stable manner.

With the set up described above, the experiment is carried out for noisy versions of the signals. The value of standard deviation for the noisy version for Gaussian noise is .04. When noisy versions of the signals are considered, , at the first iteration, average co-sparsity so observed was 21.00, denoising error was .0708 and none of the rows were recovered and .3465 was the observed distance to true dictionary. At the end of first phase of the iterations i.e. at 75 the iteration, 64% of the rows were recovered with the distance of .1517 to true dictionary and .0388 was the denoising error. At the last iteration. The average co-sparsity was 26.7356 with .1268distance to true dictionary and at most 68% rows were recovered with denoising error of .0353. The results are plotted as shown in figure 5 through 8.

V. References

[1]. A.M. Buckstein, D.L. Donohue, and M. Elad, “From sparse solutions of systems of equations to sparse modelling of signals and images,” SIAM Review, vol. 51, no. 1, pp. 34–81, Feb. 2009

[2]. M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, 2010.

[3]. R. Rubinstein, A. Bruckstein, and M. Elad, “Dictionaries for sparse representation modelling,” IEEE Proceedings, vol. 98, no. 6, pp. 1045–1057, April 2010

[4]. M. Yaghoobi, S. Nam, R. Gribonval, and M.E. Davies, “Analysis operator learning for overcomplete cosparse representations,” in Proceedings of EUSIPCO, Sep. 2011.

[5]. "FOUNDATIONS of data science, “John Hopcraft, Ravindran Kannan, Version 21/8/2014

[6]. "The elements of statistical learning, Data mining, Inference & prediction” Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2nd editon, Springer series

[7]. S. Nam, M. Davies, M. Elad, and R. Gribonval, “The cosparse analysis model and algorithms,” to appear in Applied Computational Harmonic Analysis.

[8]. ”A WAVELET TOUR OF SIGNAL PROCESSING”3RD EDITION, The sparse way, Mallat Stephane ISBN:978-0-12-374370-1

[9]. Ron. Rubinstein,Member IEEE,Tomor Peleg,Student member,IEEE, A. Bruckstein, and M. Elad,Fellow IEEE, “Analysis K-SVD:A dictionary learning algorithm for the analysis sparse model