www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume - 3 Issue - 8 August, 2014 Page No. 7555-7560

A survey and comparative analysis on different algorithms for Blind Source Separation

Purnima Mittal
Electronics and Communication Department
BPRCE, Gohana, Haryana-131301
Purnima.mittal@gmail.com

Abstract: ***Blind separation for speech signal is the original purpose of BSS problem, and becomes the research attention of signal processing in recent years. Separation of speeches has an important theoretical importance in voice communications, acoustic target detection, etc. Blind source separation is a well-established signal processing problem. The sources to be estimated present some diversity in order to be efficiently retrieved. Assuming the transmitted signals to be mutually independent in a linear multiple-input-multiple-output (MIMO) memory-less system, the transmitted signal is subjected to Additive white Gaussian noise. The received signals are, hence, corrupted by inter-user interference (IUI), and we can model them as the outputs of a linear multiple-input-multiple-output (MIMO) memory-less system. In this paper, we have surveyed different techniques of blind signal separation.***

Keywords : Blind signal separation, Principal Component Analysis, MIMO, Singular value decomposition

1. Introduction

Blind signal separation, also known as blind source separation, is the separation of a set of source signals from a set of mixed signals, without the aid of information (or with very little information) about the source signals or the mixing process. This problem is in general highly underdetermined, but useful solutions can be derived under a surprising variety of conditions. Much of the early literature in this field focuses on the separation of temporal signals such as audio [11]. However, blind signal separation is now routinely performed on multidimensional data, such as images and tensors, which may involve no time dimension whatsoever. Since the chief difficulty is the problem of it's under determination, methods for blind source separation generally seek to narrow the set of possible solutions in a way that is unlikely to exclude the desired solution [12]. In one approach, exemplified by principal and independent component analysis, one seeks source signals that are minimally correlated or maximally independent in a probabilistic or information-theoretic sense. A second approach, exemplified by nonnegative matrix factorization, is to impose structural constraints on the source signals [13]. These structural constraints may be derived from a generative model of the signal, but are more commonly heuristics justified by good empirical performance. A common theme in the second approach is to impose some kind of low-complexity constraint on the signal, such as sparsity in some basis for the signal space. This approach can be particularly effective if one requires not the whole signal, but merely its most salient features.

2. Techniques of Blind Signal Separation

There are different methods of blind signal separation:

i) Principal components analysis
ii)Singular value decomposition
iii)Independent component analysis
iv)Dependent component analysis
v) Non-negative matrix factorization
vi)Low-complexity coding and decoding
vii) Stationary subspace analysis
viii) Common spatial pattern

We have elaborated few of them:

2.1.1 Principal component analysis

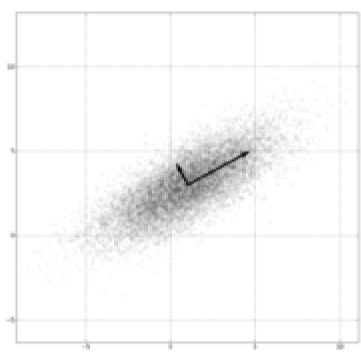

Fig. 1 PCA of a multivariate Gaussian distribution

Principal component analysis (PCA) has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it. Figure 1 shows PCA of a multivariate Gaussian distribution centred at (1,3) with a standard deviation of 3 in roughly the (0.878, 0.478) direction and of 1 in the orthogonal direction. The vectors shown are the eigenvectors of the covariance matrix scaled by the square root of the corresponding eigenvalue, and shifted so their tails are at the mean.

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables [1]. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. PCA was invented in 1901 by Karl Pearson,[2] as an analogue of the principal axes theorem in mechanics; it was later independently developed (and named) by Harold Hotelling in the 1930s.[3] The method is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute.[4] The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the component score).[5] PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a projection or "shadow" of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced. PCA is closely related to factor analysis. Factor analysis typically incorporates more domain specific assumptions about the underlying structure and solves eigenvectors of a slightly different matrix. PCA is also related to canonical correlation analysis (CCA). CCA defines coordinate systems that optimally describe the cross-covariance between two datasets while PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset [6],[7].

2.1.2 Singular Value Decomposition (SVD)

The singular value decomposition of a matrix A is the factorization of A into the product of three matrices A = UDV T where the columns of U and V are orthonormal and the matrix D is diagonal with positive real entries. The SVD is useful in many tasks. Here we mention two examples. First, the rank of a matrix A can be read off from its SVD. This is useful when the elements of the matrix are real numbers that have been rounded to some finite precision. Before the entries were rounded the matrix may have been of low rank but the rounding converted the matrix to full rank [8]. The original rank can be determined by the number of diagonal elements of D not exceedingly close to zero. Second, for a square and invertible matrix A, the inverse of A is VD−1UT. To gain insight into the SVD, treat the rows of an n × d matrix A as n points in a d-dimensional space and consider the problem of finding the best k-dimensional subspace with respect to the set of points. Here best means minimize the sum of the squares of the perpendicular distances of the points to the subspace [8]. The problem is called the best least squares fit. In the best least squares fit, one is minimizing the distance to a subspace. An alternative problem is to find the function that best fits some data. Here one variable y is a function of the variables x1, x2, . . . ,xd and one wishes to minimize the vertical distance, i.e., distance in the y direction, to the subspace of the xi rather than minimize the perpendicular distance to the subspace being fit to the data.

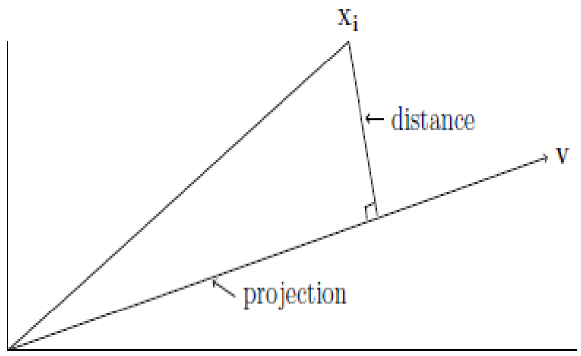

Figure 2: The projection of the point xi onto the line through the origin in the direction of v [8]

Returning to the best least squares fit problem, consider projecting a point xi onto a line through the origin. Then

$x^2_{i1} + x^2_{i2} + \cdot\cdot\cdot + ^2_{id}$ = (length of projection)^2 + (distance of point to line)^2 .

See Figure 2. Thus (distance of point to line)^2 = $x^2_{i1} + x^2_{i2} + \cdot\cdot\cdot + ^2_{id}$ −(length of projection)^2 .

To minimize the sum of the squares of the distances to the line, one could minimize

$\sum_{i=1}^{n}$ $(x^2_{i1} + x^2_{i2} + \cdot\cdot\cdot + ^2_{id})$ minus the sum of the squares of the lengths of the projections of the points to the line. However, $\sum_{i=1}^{n}$ $(x^2_{i1} + x^2_{i2} + \cdot\cdot\cdot + ^2_{id})$ is a constant (independent of the line), so minimizing the sum of the squares of the distances is equivalent to maximizing the sum of the squares of the lengths of the projections onto the line. Similarly for best-fit subspaces, we could maximize the sum of the squared lengths of the projections onto the subspace instead of minimizing the sum of squared distances to the subspace [8].

2.1.3 Independent component analysis

Independent Component Analysis (ICA) is a statistical technique, perhaps the most widely used, for solving the blind source separation problem [10, 11]. In this section, we present the basic Independent Component Analysis model and show under which conditions its parameters can be estimated.

ICA model

The general model for ICA is that the sources are generated through a linear basis transformation, where additive noise can be present. Suppose we have N statistically independent signals, si(t), i = 1, ...,N. We assume that the sources themselves cannot be directly observed and that each signal, si(t), is a realization of some fixed probability distribution at each time point t. Also, suppose we observe these signals using N sensors, then we obtain a set of N observation signals xi(t), i = 1, ...,N that are mixtures of the sources. A fundamental aspect of the mixing process is that the sensors must be spatially separated (e.g. microphones that are spatially distributed around a room) so that each sensor records a different mixture of the sources [9]. With this spatial separation assumption in mind, we can model the mixing process with matrix multiplication as follows:

$$x(t) = As(t) \qquad (1)$$

where A is an unknown matrix called the mixing matrix and x(t), s(t) are the two vectors representing the observed signals and source signals respectively. Incidentally, the justification for the description of this signal processing technique as blind is that we have no information on the mixing matrix, or even on the sources themselves. The objective is to recover the original signals, $s_i(t)$, from only the observed vector $x_i(t)$. We obtain estimates for the sources by first obtaining the "un-mixing matrix" W, where, $W = A^{-1}$.

This enables an estimate, s^(t), of the independent sources to be obtained:

$$\hat{s}(t) = Wx(t) \qquad (2)$$

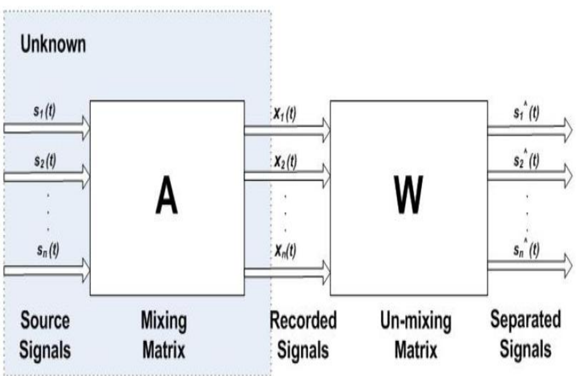

Figure 3: Blind source separation (BSS) block diagram. s(t) are the sources. x(t) are the recordings, ˆ s(t) are the estimated sources A is mixing matrix and W is un-mixing matrix [9].

The diagram in Figure 3 illustrates both the mixing and un-mixing process involved in BSS. The independent sources are mixed by the matrix A (which is unknown in this case). We seek to obtain a vector y that approximates s by estimating the un-mixing matrix W. If the estimate of the un-mixing matrix is accurate, we obtain a good approximation of the sources. The above described ICA model is the simple model since it ignores all noise components and any time delay in the recordings [9].

2.2 Eigenvalues and Eigenvectors

Now, consider a small example showing the characteristics of the eigenvectors. Some artificial data has been generated, which is illustrated in the Figure 4. The small dots are the points in the data set [18].

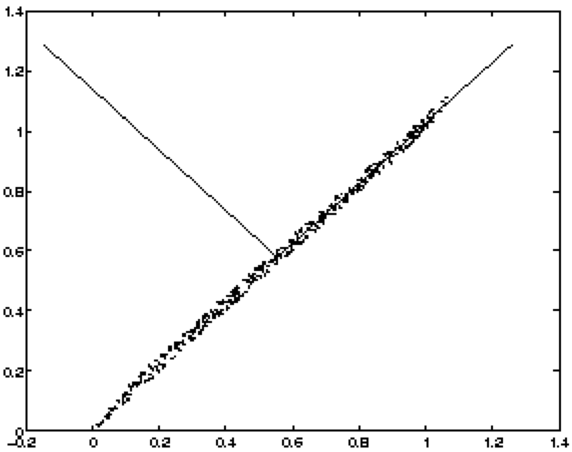

Figure 4: Eigenvectors of the artificially created data [18]

Sample mean and sample covariance matrix can easily be calculated from the data. Eigenvectors and eigenvalues can be calculated from the covariance matrix. The directions of eigenvectors are drawn in the Figure 4 as lines. The first eigenvector having the largest eigenvalue points to the direction of largest variance (right and upwards) whereas the second eigenvector is orthogonal to the first one (pointing to left and upwards). In this example the first eigenvalue corresponding to the first eigenvector is λ_1= 0.1737 while the other eigenvalue is λ_2= 0.0001. By comparing the values of eigenvalues to the total sum of eigenvalues, we can get an idea how much of the energy is concentrated along the particular eigenvector. In this case, the first eigenvector contains almost all the energy. The data could be well approximated with a one-dimensional representation. Sometimes it is desirable to investigate the behavior of the system under small changes. Assume that this system or phenomenon is constrained to a *n*-dimensional manifold and can be approximated with a linear manifold. Suppose one has a small change along one of the coordinate axes in the original coordinate system [18]. If the data from the phenomenon is concentrated in a subspace, we can project this small change δ_x to the approximate subspace built with PCA by projecting δx on all the basis vectors in the linear subspace by

$$\delta_y = A_K \, \delta_x \qquad (3)$$

Where, the matrix A_K has the K first eigenvectors as rows. Subspace has then a dimension of K. δ_y represents the change caused by the original small change [18]. This can be transformed back with a change of basis by taking a linear combination of the basis vectors by

$$\delta_x = A_K^T \delta_y \tag{4}$$

Then, we get the typical change in the real-world coordinate system caused by a small change δ_x by assuming that the phenomenon constrains the system to have values in the limited subspace only. The eigenvectors of a square matrix are the non-zero vectors which, after being multiplied by the matrix, remain proportional to the original vector, i.e. any vector $\mathbf{x}$ that satisfies the equation:

$$\mathbf{Ax} = \lambda\mathbf{x} \tag{5}$$

Where $\mathbf{A}$ is the matrix in question, $\mathbf{x}$ is the eigenvector and λ is the associated eigenvalue. As will become clear later on, eigenvectors are not unique in the sense that any eigenvector can be multiplied by a constant to form another eigenvector. For each eigenvector there is only one associated eigenvalue, however. If you consider a 2×2 matrix as a stretching, shearing or reflection transformation of the plane, you can see that the eigenvalues are the lines passing through the origin that are left unchanged by the transformation [18]. Note that square matrices of any size, not just 2×2 matrices, can have eigenvectors and eigenvalues. In order to find the eigenvectors of a matrix we must start by finding the *eigenvalues*. To do this we take everything over to the LHS of the equation:

$$\mathbf{Ax} - \lambda\mathbf{x} = \mathbf{0} \tag{6}$$

Then we pull the vector $\mathbf{x}$ outside of a set of brackets:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{7}$$

The only way this can be solved is if $\mathbf{A}-\lambda\mathbf{I}$ does not have an inverse[1], therefore we find values of λ such that the determinant of $\mathbf{A}-\lambda\mathbf{I}$ is zero:

$$|\mathbf{A} - \lambda\mathbf{I}| = \mathbf{0} \tag{8}$$

Once we have a set of eigenvalues we can substitute them back into the original equation to find the eigenvectors [18].

2.3 Gram-Schmidt Orthogonalization:

To replace the linearly independent vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v}_n$ one by one with mutually orthogonal vectors $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$ that span the same subspace, begin with

$$\mathbf{u_1} = \mathbf{v_1} \tag{9}$$

For $k = 1,2,\ldots n-1$ in turn, take

$$\mathbf{u_{k+1}} = \mathbf{v_{k+1}} - \frac{\mathbf{u_1}\cdot\mathbf{v_{k+1}}}{\mathbf{u_1}\cdot\mathbf{u_1}}\mathbf{u_1} - \frac{\mathbf{u_2}\cdot\mathbf{v_{k+1}}}{\mathbf{u_2}\cdot\mathbf{u_2}}\mathbf{u_2} - \cdots - \frac{\mathbf{u_k}\cdot\mathbf{v_{k+1}}}{\mathbf{u_k}\cdot\mathbf{u_k}}\mathbf{u_k} \tag{10}$$

---

3. Related Work

Chen Meng et. Al. [12] describes that Blind equalization and signal separation are two well-established signal processing problems. In this paper, the author presents a quadratic programming algorithm for fast blind equalization and signal separation. By introducing a special non-mean-square error (MSE) objective function. The author reformulates fractionally spaced blind equalization into an equivalent quadratic programming problem. Based on a clear geometric interpretation and a formal proof, the author shows that a perfect equalization solution is obtained at every local optimum of the quadratic program. Because blind source separation is, by nature and mathematically, a closely related problem, the authors also generalize the algorithm for blind signal separation. The authors show that by enforcing source orthogonalization through successive processing, the quadratic programming approach can be applied effectively. Moreover, the quadratic program is easily extendible to incorporate additional practical conditions, such as jamming suppression constraints. The author also provides evidence of good performance through computer simulations.

Qiao Li-yan et. Al. [13] presents a novel framework for separating and reconstructing multichannel speech sources from compressively sensed linear mixtures simultaneously. The conventional approaches for blind speech separation are almost based on the Nyquist sampling theory. The author proposed an approach which uses the multichannel compressive sensing theory for blind speech separation. The linear programming and gradient-based methods are used to separate the sources. Compared with the conventional blind speech separation, the proposed approach can reduce the requirements of sampling speed and operating rate of the devices. Moreover, our approach has lower computational complexity. The main contribution of this paper lies in proposing a novel procedure to estimate the sources from the measurements without reconstructing the mixed signals. Simulation results demonstrate the proposed algorithm can separate multichannel speech sources successfully.

Jérémy Rapin et. Al. [14] describes that Non-negative blind source separation (BSS) has raised interest in various fields of research, as testified by the wide literature on the topic of non-negative matrix factorization (NMF). In this context, it is fundamental that the sources to be estimated present some diversity in order to be efficiently retrieved. Sparsity is known to enhance such contrast between the sources while producing very robust approaches, especially to noise. In this paper, the author introduces a new algorithm in order to tackle the blind separation of non-negative sparse sources from noisy measurements. The author first shows that sparsity and non-negativity constraints have to be carefully applied on the sought-after solution. In fact, improperly constrained solutions are unlikely to be stable and are therefore sub-optimal. The proposed algorithm, named nGMCA (non-negative Generalized Morphological Component Analysis), makes use of proximal calculus techniques to provide properly constrained solutions. The performance of nGMCA compared to other state-of-the-art algorithms is demonstrated by numerical experiments encompassing a wide variety of settings, with negligible parameter tuning. In particular, nGMCA is shown to provide robustness to noise and performs well on synthetic mixtures of real NMR spectra.

Frédéric Krüger et. Al [15] states that Electrical distribution companies struggle to find precise energy demand for their networks. They have at their disposal statistical tools such as power load profiles, which are however usually not precise enough and do not take into account factors such as the presence of electrical heating devices or the type of housing of the end users. In this paper, author shows how the determination of accurate load profiles can be considered as a noisy blind source separation problem solved by an evolutionary algorithm. The power load profiles obtained demonstrate considerable improvement in the load curve forecasts of 20kV/400V substations.

Jie Yang et. Al. [16] provides a convex model based subspace projection method with enhanced functionality, which can be used for Underdetermined Blind Source Separation (UBSS). The model takes into account both projection and size of the signal's subspace, without estimating the source number at Time- Frequency (TF) point. Simulation results show that it overcomes the shortage of conventional subspace method and achieves high separation performance. The proposed method can be employed as a preprocessing technology for audio separation and enhancement, as well as biomedical images, etc.

Gilles Chabriel et. Al. [17] Matrix decompositions such as the eigenvalue decomposition (EVD) or the singular value decomposition (SVD) have a long history in signal processing. They have been used in spectral analysis, signal/noise subspace estimation, principal component analysis (PCA), dimensionality reduction, and whitening in independent component analysis (ICA). Very often, the matrix under consideration is the covariance matrix of some observation signals. However, many other kinds of matrices can be encountered in signal processing problems, such as time-lagged covariance matrices, quadratic spatial time-frequency matrices, and matrices of higher-order statistics. In concert with this diversity, the joint diagonalization (JD) or approximate JD (AJD) of a set of matrices has been recently recognized to be instrumental in signal processing, mainly because of its importance in practical signal processing problems such as source separation, blind beamforming, image denoising, blind channel identification for multiple-input, multiple-output (MIMO) telecommunication system, Doppler-shifted echo extraction in radar, and ICA. Perhaps one of the first such algorithms is the joint approximate diagonalization of eigenmatrices (JADE) algorithm. In this algorithm, the matrices under consideration are Hermitian and the considered joint diagonalizer is a unitary matrix. More recently, generalizations and/or new decompositions were found to be of considerable interest. They concern new sets of matrices, a nonunitary joint diagonalizer, and new decompositions.

4. Conclusion and Future Work

It can be concluded that the existing method have done much efficient work in the field of source signal separation in a MIMO system. But, a lot of scope is there for improvement in terms of number of input tuning parameters, time for computation, quality of extracted signals. Achieving good separation quality from the existing methods, is highly dependent on tuning its many parameters. This makes the use of the above algorithms problematic in a real-world system as the changing environment mean its parameters would need to be continuously updated in a sophisticated manner. Also, training of input data is required for quality separation according to existing methods which makes the separation process much more complex and time consuming So, there is a need of an algorithm which does not require any training of data with lesser number of tuning parameters.

References

[1] Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4

[2] Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF). *Philosophical Magazine* 2 (11): 559572.doi:10.1080/14786440109462720.

[3] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441, and 498-520.

[4] Abdi. H., & Williams, L.J. (2010). "Principal component analysis.". *Wiley Interdisciplinary Reviews: Computational Statistics*, 2: 433–459.doi:10.1002/wics.101.

[5] Shaw P.J.A. (2003) *Multivariate statistics for the Environmental Sciences*, Hodder-Arnold. ISBN 0-340-80763-6.

[6] Barnett, T. P., and R. Preisendorfer. (1987). "Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis.". *Monthly Weather Review 115*.

[7] Hsu, Daniel, Sham M. Kakade, and Tong Zhang (2008). "A spectral algorithm for learning hidden markov models.". *arXiv preprint arXiv:0811.4413*.

[8] http://www.cs.princeton.edu/courses/ archive/ spring12/cos598C/svdchapter.pdf.

[9] Ganesh R. Naik and Dinesh K Kumar, "An Overview of Independent Component Analysis and Its Applications ", *Informatica* 35 (2011) 63–81.

[10] A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis. *Wiley-Interscience*, May 2001.

[11] J. V. Stone, Independent Component Analysis: A Tutorial Introduction (Bradford Books). *The MIT Press*, September 2004.

[12] Chen Meng, Student Member, IEEE, Jamal Tuqan, "A Quadratic Programming Approach to Blind Equalization and Signal Separation", *IEEE Transactions on Signal Processing*, VOL. 57, NO. 6, JUNE 2009.

[13] Qiao Li-yan, Congru Yin, Hongwei Xu, Hongpeng Li, Ning Fu, "Blind Separation of Speech Sources in Multichannel Compressed Sensing", IEEE 2012.

[14] Jérémy Rapin, Jérôme Bobin, Anthony Larue, and Jean-Luc Starck, "Sparse and Non-Negative BSS for Noisy Data", *IEEE Transactions On Signal Processing*, VOL. 61, NO. 22, Nov. 15, 2013.

[15] Frédéric Krüger, "Blind Signal Separation applied to Power Load Profiles Computation", IEEE 2013.

[16] Jie Yang, Liang-Jun Zhang, Kai-Wang Lu and Qing-Nian Zhang," Underdetermined Blind Source Separation Using Modified Subspace Method Based on Convex Model", IEEE International Conference on Consumer Electronics (ICCE) 2014.

[17] Gilles Chabriel, Martin Kleinsteuber, Eric Moreau, Hao Shen, Petr Tichavsky, and Arie Yeredor Joint Matrices, " Decompositions and Blind Source Separation", *IEEE* 2014.
[18] http://users.ics.aalto.fi/jhollmen/dippa/
node30.html

Author Profile

Purnima Mittal is pursuing M.tech in Electronics and Communication engineering from Deenbandhu Chhotu Ram Univerity of Science and Technology, Murthal, Sonepat, Haryana. She had completed her B.tech in ECE from Hindu College of Engineering, Sonepat in 2005. She had 4 years of teaching experience in various colleges.