Chapter 8

Tensor Factorization with Application to Convolutive Blind Source Separation of Speech

Saeid Sanei
Cardiff University, UK

Bahador Makkiabadi
Cardiff University, UK

ABSTRACT

Tensor factorization (TF) is introduced as a powerful tool for solving multi-way problems. As an effective and major application of this technique, separation of sound particularly speech signal sources from their corresponding convolutive mixtures is described and the results are demonstrated. The method is flexible and can easily incorporate all possible parameters or factors into the separation formulation. As a consequence of that fewer assumptions (such as uncorrelatedness and independency) will be required. The new formulation allows further degree of freedom to the original parallel factor analysis (PARAFAC) problem in which the scaling and permutation problems of the frequency domain blind source separation (BSS) can be resolved. Based on the results of experiments using real data in a simulated medium, it has been concluded that compared to conventional frequency domain BSS methods, both objective and subjective results are improved when the proposed algorithm is used.

INTRODUCTION

Decomposition of mixed information into its constituent components has been very useful in many applications such as acoustics, communications, and biomedicine. Eigenvalue decomposition (EVD), singular value decomposition (SVD), and independent component analysis (ICA) based on various criteria such as uncorrelatedness, independency, minimizing mutual information, and differences in distributions, have been widely used for this purpose before. In the case of convolutive mixtures however, further processing to handle multiple time lags of the signals has to be undertaken. Although many researchers have worked on convolutive blind source separation, as comprehensively reported by Pederson et al. in (Pederson et al., 2007) no robust solution, as given for linear instantaneous cases, has been reported. Moreover, generally, the uncorrelated-
ness or independency assumption may not be true for many applications. In this chapter the fundamental techniques on convolutive BSS are reviewed and a new method has been developed for separation of convolutive sources. The method is based on tensor assumption (or tensorization) of the convolutive mixtures.

TF and multi-way array factorisation and decomposition have become very popular for multi-way signal processing recently. Using this approach generally, there won’t be any need for any strong assumptions about mutual statistical properties of the sources such as uncorrelatedness or independence but a mild assumption about the disjointedness in some domain often helps.

This revolutionary technique circumvents many problems in the way of traditional and current convolutive source separation techniques such as the above limiting assumptions, scaling, permutation, causality, circularity (of the convolution), and ill-posedness problems, and provides a unique solution directly.

In the following sections, first the problem of convolutive BSS is stated. Then, TF is discussed in detail and some examples for both overdetermined and underdetermined cases (where the number of sources are more than the number of sensors) are demonstrated.

**CONVOLUTIVE BLIND SOURCE SEPARATION**

The problem of convolutive BSS has been under research over the past two decades. A number of papers and reviews on convolutive BSS as addressed in (Pederson et al., 2007) have been published recently. In many practical situations the signals reach the sensors with different time delays. The corresponding delay between source \( j \) and sensor \( i \), in terms of number of samples, is directly proportional to the sampling frequency and conversely to the speed of sound in the medium, i.e. \( \delta_{ij} \propto d_{ij} \times f_s / c \), where \( d_{ij}, f_s \) and \( c \) are respectively, the distance between source \( j \) and sensor \( i \), the sampling frequency, and the speed of sound. For speech and music in the air as an example we may have \( d_{ij} \) in terms of meters, \( f_s \) between 8 to 44 KHz, and \( c=330 \text{ m/sec} \). Also, in an acoustic environment the sound signals can reach the sensors through multi-paths after reflections by obstacles (such as walls). A general matrix formulation of the CBSS for mixing and separating the source signals can be given as:

\[
\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t) + \mathbf{v}(t) \quad (1)
\]

and

\[
\mathbf{y}(t) = \mathbf{W}(t) * \mathbf{x}(t) \quad (2)
\]

where \( M \times 1 \mathbf{s}(t), \ N \times 1 \mathbf{x}(t), \) and \( N \times 1 \mathbf{v}(t) \) denote respectively the vector of source signals, observed signals, and noise at discrete time \( t \). \( \mathbf{H} \) is the mixing matrix of size \( N \times M \) and * denotes convolution operator. The separation is performed by means of a separating \( M \times N \) matrix, \( \mathbf{W} \), which uses only the information about \( \mathbf{x}(t) \) to reconstruct the original source signals denoted as \( \mathbf{y}(t) \).

Equation (1) and (2) are the general forms of both anechoic and echoic BSS models. In an anechoic model, however, the expansion of the mixing process may be given as:

\[
x_i(t) = \sum_{j=1}^{M} h_{ij} s_j(t - \delta_{ij}) + v_i(t), \quad \text{for } i = 1, \ldots, N
\]

(3)

where the attenuation, \( h_{ij} \), and delay, \( \delta_{ij} \), of source \( j \) to sensor \( i \) would be determined by the physical position of the source relative to the sensors. Then, the unmixing process is given as:

\[
y_j(t) = \sum_{i=1}^{N} w_{ij} x_i(t - \delta_{ij}), \quad \text{for } j = 1, \ldots, M
\]

(4)