



**HAL**  
open science

## Towards Distributed Component Analysis

Reza Sameni

► **To cite this version:**

| Reza Sameni. Towards Distributed Component Analysis. 2015. hal-01382076

**HAL Id: hal-01382076**

**<https://hal.science/hal-01382076>**

Preprint submitted on 15 Oct 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Towards Distributed Component Analysis

(working paper)

Reza Sameni\*

October 2015

**Abstract**—The notion of spatially distributed sources arises in many fields of applied signal processing. Following previous empirical findings in this area, in this work, we attempt to validate some of these findings within a simplified mathematical framework for a finite electrostatic distributed source within a homogeneous volume conductor. The model is very accurate for signals such as the electrocardiogram and electroencephalogram. The method is used to analyze second-order statistics of distributed sources. The findings of this research are expected to be helpful for better interpretation of PCA and ICA results, specifically for biomedical applications.

## I. INTRODUCTION

In the past two decades, blind source separation (BSS) has been extensively used in various applied domains including audio, video, telecommunications and biomedical signal processing. While, the classical and most common formulation of BSS is based on the assumption of linear time-invariant mixtures of punctual sources, nonlinear and time-variant models have also found great attention in recent years [Comon and Jutten, 2010]. However, to the author’s knowledge, in all these developments the sources of interest are assumed to be (or approximated by) finite number of spatially *punctual sources*. On the other hand, there are many applications, in which the punctuality of the sources does not hold or is not a good approximation. For example, in biomedical applications, sensors are typically rather close to the signal sources (heart, brain, or muscles), which violates the far-field assumptions. In these cases, the sources estimated by classical algorithms, such as *principal component analysis* (PCA) and *independent component analysis* (ICA) are only approximations of the true distributed activity. In [Sameni, 2008], we introduced the notion of spatially distributed sources for electrocardiogram (ECG) signals and their impact on the signals extracted by applying conventional linear time-invariant PCA and ICA [Sameni et al., 2006]. It was empirically shown that components extracted from multichannel ECG (1) have infinite number of dimensions, i.e., always appear in full-rank regardless of the number of recordings<sup>1</sup>; (2) the extracted components are not necessarily invariant under the change of electrode configuration. In this work, we attempt to justify these findings and extend them using a rather rigorous electrostatic framework for a spatially distributed stochastic electrostatic density function (used to model a distributed source signal). The current study is merely

\*R. Sameni is with the School of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran. Email: [reza.sameni@gmail.com](mailto:reza.sameni@gmail.com), [rsameni@shirazu.ac.ir](mailto:rsameni@shirazu.ac.ir) Web: [www.sameni.info](http://www.sameni.info).

<sup>1</sup>Although in practice, the weaker components can be masked by other interference and noise.

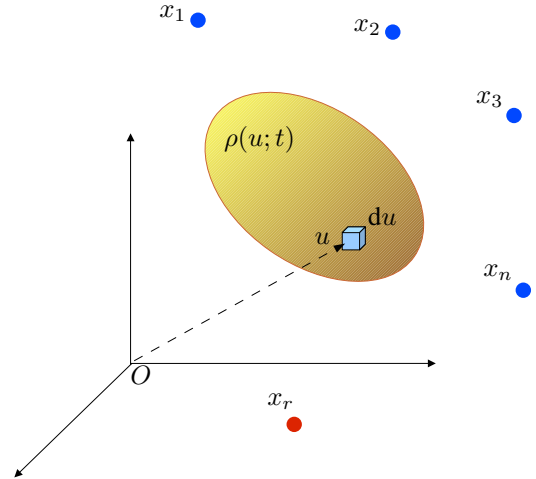


Fig. 1. Distributed source model and the electrode positions

based on the second-order statistics of distributed sources within a homogeneous volume conductor. It is shown that besides our previous findings, a non-homogeneous volume conductor can also result in the extraction of ‘fake’ sources that somehow correspond to reflections of the source distribution onto surfaces with conductivity discontinuity.

In this report, a summary of the proposed framework is presented, which will be accompanied by further details and experimental results in the future versions of the manuscript.

## II. DISTRIBUTED SOURCES

We consider a distributed source with a time varying stochastic density  $\rho(u; t)$  within a finite homogeneous volume (Fig. 1). The *field potentials* due to this distribution can be represented as follows:

$$\phi(r; t) = \int \frac{\rho(u; t)}{|r - u|^p} du \quad (1)$$

where  $r = [x, y, z]^T$  is the Cartesian coordinates of the observation point<sup>2</sup>,  $p$  is the decay exponent which depends on the nature of  $\rho(u; t)$  and the field potentials,  $du$  is the differential volume unit, and the integral is taken over  $V$ , the volume containing the distributed source. Therefore, the potential difference of each measurement (observation) point

<sup>2</sup>For simplicity, we have normalized the equation by the conductivity of the medium, as it does not affect our calculations.

$(x_i)$ , with respect to the reference can be written as follows:

$$\begin{aligned}\Delta\phi(x_i; t) &\doteq \phi(x_i; t) - \phi(x_r; t) \\ &= \int \rho(u; t) \left[ \frac{1}{|x_i - u|^p} - \frac{1}{|x_r - u|^p} \right] du \\ &\doteq \int \rho(u; t) g(x_i, u) du\end{aligned}\quad (2)$$

where  $g(x_i, u)$  is the Green's function for the observation point  $x_i$ , which acts as a distance-dependent weighting factor for  $\rho(u; t)$ <sup>3</sup>.

Next, suppose that we make  $n$  observations of the field potentials from different locations, each time with respect to the same reference point  $x_r$ . In this case, (2) should be rewritten in vector form. Defining  $\mathbf{x} \doteq [x_1, x_2, \dots, x_n]^T$ ,  $\boldsymbol{\phi}(\mathbf{x}; t) \doteq [\Delta\phi(x_1; t), \dots, \Delta\phi(x_n; t)]^T$  and  $\mathbf{G}(\mathbf{x}, u) \doteq [g(x_1, u), \dots, g(x_n, u)]^T$ , (2) can be represented in the following form:

$$\boldsymbol{\phi}(\mathbf{x}; t) = \int \rho(u; t) \mathbf{G}(\mathbf{x}, u) du \quad (3)$$

where  $\boldsymbol{\phi}(\mathbf{x}; t)$  can be considered as a vector of differential observations corresponding to, e.g., the signals recorded from the body or brain surface. Without loss of generality, these observations can be considered to be zero-mean (in time). The covariance matrix of the observations is therefore found as follows:

$$\begin{aligned}R_\phi(\mathbf{x}) &= E_t\{\boldsymbol{\phi}(\mathbf{x}; t)\boldsymbol{\phi}(\mathbf{x}; t)^T\} \\ &= \iint \mathbf{G}(\mathbf{x}, u) E_t\{\rho(u; t)\rho(v; t)\} \mathbf{G}(\mathbf{x}, v)^T du dv \\ &= \iint \mathbf{G}(\mathbf{x}, u) r_\rho(u, v) \mathbf{G}(\mathbf{x}, v)^T du dv\end{aligned}\quad (4)$$

where  $r_\rho(u, v) \doteq E_t\{\rho(u; t)\rho(v; t)\}$  is the cross-correlation function of the source density for different points inside the volume  $V$ , and  $E_t\{\cdot\}$  represents averaging over time.

Next, using the Karhunen-Loève Transform (KLT) [Van Trees, 2004, Ch. 3],  $r_\rho(\cdot, \cdot)$  may be expanded in terms of a unique and (possibly) infinite set of orthogonal eigenfunctions  $\{f_i(u)\}_{i=1}^\infty$ , i.e.:

$$\lambda_i f_i(u) = \int r_\rho(u, v) f_i(v) dv, \quad (5)$$

$$\int f_i(v) f_j(v) dv = \delta_{i-j} \quad (6)$$

where  $\{\lambda_i\}_{i=1}^\infty$  are the eigenvalues and  $\delta_{i-j}$  is the Kronecker delta function. Following Mercer's theorem [Van Trees, 2004, Ch. 3], these eigenvalues and eigenfunctions can be used to expand  $r_\rho(\cdot, \cdot)$  as follows:

$$r_\rho(u, v) = \sum_{i=1}^\infty \lambda_i f_i(u) f_i(v) \quad (7)$$

<sup>3</sup>Note that  $g(x_i, u)$  is also a function of the reference point  $x_r$ , which we do not show for notational simplicity, as the reference is a fixed point for all the observation points.

Combining (4) and (7), we find:

$$\begin{aligned}R_\phi(\mathbf{x}) &= \sum_{i=1}^\infty \lambda_i \iint \mathbf{G}(\mathbf{x}, u) f_i(u) f_i(v) \mathbf{G}(\mathbf{x}, v)^T du dv \\ &= \sum_{i=1}^\infty \lambda_i \int \mathbf{G}(\mathbf{x}, u) f_i(u) du \int \mathbf{G}(\mathbf{x}, v)^T f_i(v) dv \\ &= \sum_{i=1}^\infty \lambda_i \boldsymbol{\alpha}_i(\mathbf{x}) \boldsymbol{\alpha}_i(\mathbf{x})^T\end{aligned}\quad (8)$$

where

$$\begin{aligned}\mathbf{G}(\mathbf{x}, s) &= \sum_{i=1}^\infty \boldsymbol{\alpha}_i(\mathbf{x}) f_i(s) \\ \boldsymbol{\alpha}_i(\mathbf{x}) &\doteq \int \mathbf{G}(\mathbf{x}, s) f_i(s) ds = \langle \mathbf{G}(\mathbf{x}, \cdot), f_i(\cdot) \rangle\end{aligned}$$

i.e.,  $\boldsymbol{\alpha}_i(\mathbf{x})$  is the coefficient vector for the orthogonal expansion of the Green's function  $\mathbf{G}(\mathbf{x}, u)$  in terms of the eigenfunction  $f_i(u)$ . Apparently,  $\boldsymbol{\alpha}_i(\mathbf{x}) \boldsymbol{\alpha}_i(\mathbf{x})^T \in \mathbb{R}^{n \times n}$  are symmetric rank-1 matrices with eigenvalues equal to  $\|\boldsymbol{\alpha}_i(\mathbf{x})\|^2$ . Therefore,  $R_\phi(\mathbf{x})$  is an infinite weighted sum of a set of positive definite rank-1 matrices. Note that although the eigenfunctions  $f_i(\cdot)$  are orthogonal, the  $\boldsymbol{\alpha}_i(\mathbf{x})$  are not necessarily orthogonal. Therefore, (8) is a redundant decomposition of  $R_\phi(\mathbf{x})$  in the  $n$ -dimensional space.

On the other hand, the matrix  $R_\phi(\mathbf{x})$ , being the covariance matrix of the observed signals, can be decomposed by PCA

$$R_\phi(\mathbf{x}) = Q D Q^T = \sum_{i=1}^n d_i \mathbf{q}_i \mathbf{q}_i^T = \sum_{i=1}^\infty \lambda_i \boldsymbol{\alpha}_i(\mathbf{x}) \boldsymbol{\alpha}_i(\mathbf{x})^T \quad (9)$$

where  $Q = [\mathbf{q}_1, \dots, \mathbf{q}_n]$  is the orthogonal eigenmatrix,  $D = \text{diag}(d_1, \dots, d_n)$  contains the eigenvalues on its diagonal, and the  $\mathbf{q}_i \mathbf{q}_i^T$  are orthogonal rank-1 projectors onto  $\text{range}(Q)$ . From (9), we can see how the eigenvectors and eigenvalues of a set of observations  $\boldsymbol{\phi}(r; t)$  are related to the source distribution and electrode positioning. In fact, the set of eigenvectors  $\{\mathbf{q}_i\}_{i=1}^n$  can be interpreted as an orthogonal basis for the subspace spanned by the column space of (the generally) nonorthogonal and infinite set of  $\{\boldsymbol{\alpha}_i(\mathbf{x})\}_{i=1}^\infty$ .

Equation (9) shows how the stochastic properties of the source density and the sensor configurations, both, influence the eigenstructure of the covariance matrix of the observations. From (9) it is also seen that the covariance matrix of measurements from distributed sources is generally full-rank, no matter how many observations are made; unless if the KLT expansion of  $r_\rho(\cdot, \cdot)$  be finite (i.e.,  $\exists l, \forall j > l : \lambda_j = 0$ ) or due to a specific sensor configuration the expansion of  $\mathbf{G}(\mathbf{x}, \cdot)$  in terms of  $f_i(\cdot)$  be finite (i.e.,  $\exists l, \forall j > l : \boldsymbol{\alpha}_j(\mathbf{x}) = 0$ ). Either of these cases, can be due to special sensor configurations or special source densities. This explains many observations made from practical distributed sources, such as cardiac, muscular, or brain signals. In these cases it has been previously reported that the decomposition of the signals recorded from the body surface (the ECG, EEG, or the EMG), using PCA or ICA can lead into multiple components corresponding to the sources of interest [Sameni, 2008].

### III. SPECIAL CASES

Some special cases of the proposed framework are explained in the following.

#### A. White Random Field

If  $\rho(u; t)$  is a white random field, we have:

$$r_\rho(u, v) = \sigma(u)^2 \delta(u - v)$$

where  $\delta(\cdot)$  is the Dirac delta function. In this case, (4) reduces to:

$$R_\phi(\mathbf{x}) = \int \sigma(u)^2 \mathbf{G}(\mathbf{x}, u) \mathbf{G}(\mathbf{x}, u)^T du \quad (10)$$

which is generally a full-rank  $n \times n$  matrix. A special case that reduces (10) to a finite summation of rank-1 matrices is:

$$\sigma(u)^2 = \sum_{i=1}^m \sigma_i^2 \delta(u - u_i) \quad (11)$$

which corresponds to  $m$  uncorrelated punctual sources inside  $\mathbf{V}$ . In this case, we have:

$$R_\phi(\mathbf{x}) = \sum_{i=1}^m \sigma_i^2 \mathbf{G}(\mathbf{x}, u_i) \mathbf{G}(\mathbf{x}, u_i)^T \quad (12)$$

In this case,  $R_\phi(\mathbf{x})$  has a rank of  $\min(m, n)$ .

#### B. Synchronous (Coherent) Field

If the random field  $\rho(u; t)$  is synchronous throughout the entire volume  $\mathbf{V}$ , we have  $r_\rho(u, v) = \sigma^2$ . In this case the double integral in (4) is separable and  $R_\phi(\mathbf{x})$  may be written as follows:

$$R_\phi(\mathbf{x}) = \left[ \int \sigma \mathbf{G}(\mathbf{x}, u) du \right] \left[ \int \sigma \mathbf{G}(\mathbf{x}, v) dv \right]^T \quad (13)$$

which is a rank-1 matrix (the product of a column vector by its transpose). We therefore conclude that a fully synchronous distributed source behaves like a single punctual source, which is a well known result in the electromagnetics context. This result can also be extended to isolated groups of synchronous sources.

#### C. Dipolar Random Field

Dipolar random fields are another interesting special case that we study. For these sources we assume:

$$\rho(u; t) = f(t) \delta(u - u_0) + g(t) \delta(u + u_0) \quad (14)$$

where  $f(t)$  and  $g(t)$  are zero-mean random processes. Equation (14) leads to:

$$\begin{aligned} r_\rho(u, v) &= E_t \{ \rho(u; t) \rho(v; t) \} \\ &= \sigma_f^2 \delta(u - u_0) \delta(v - u_0) + \sigma_g^2 \delta(u + u_0) \delta(v + u_0) \\ &\quad + \sigma_{fg} \delta(u - u_0) \delta(v + u_0) + \sigma_{fg} \delta(u + u_0) \delta(v - u_0) \end{aligned} \quad (15)$$

where  $\sigma_f^2$ ,  $\sigma_g^2$ , and  $\sigma_{fg}$ , correspond to the variances of  $f(t)$ ,  $g(t)$ , and their cross-correlations, respectively. Combining (15) and (4) we find:

$$\begin{aligned} R_\phi(\mathbf{x}) &= \\ &\sigma_f^2 \mathbf{G}(\mathbf{x}, u_0) \mathbf{G}(\mathbf{x}, u_0)^T + \sigma_g^2 \mathbf{G}(\mathbf{x}, -u_0) \mathbf{G}(\mathbf{x}, -u_0)^T \\ &\quad + \sigma_{fg} \mathbf{G}(\mathbf{x}, u_0) \mathbf{G}(\mathbf{x}, -u_0)^T + \sigma_{fg} \mathbf{G}(\mathbf{x}, -u_0) \mathbf{G}(\mathbf{x}, u_0)^T \end{aligned} \quad (16)$$

This equation can be rearranged as the summation of three rank-1 matrices as follows:

$$\begin{aligned} R_\phi(\mathbf{x}) &= \\ &(\sigma_f^2 + \sigma_{fg}) \mathbf{G}(\mathbf{x}, u_0) \mathbf{G}(\mathbf{x}, u_0)^T \\ &\quad + (\sigma_g^2 + \sigma_{fg}) \mathbf{G}(\mathbf{x}, -u_0) \mathbf{G}(\mathbf{x}, -u_0)^T \\ &\quad - \sigma_{fg} [\mathbf{G}(\mathbf{x}, u_0) - \mathbf{G}(\mathbf{x}, -u_0)] [\mathbf{G}(\mathbf{x}, u_0) - \mathbf{G}(\mathbf{x}, -u_0)]^T \end{aligned} \quad (17)$$

which has a maximum rank of three. Apparently, if the two punctual sources are uncorrelated, i.e.,  $\sigma_{fg} = 0$ ,  $R_\phi(\mathbf{x})$  will have a maximum rank of two, which reduces to the special case studied in the previous subsection.

#### D. Stationary Random Field

Another special case that is rather practical is a stationary random field. For a *covariance stationary* random field we have [Van Trees, 2004, Ch. 3]:

$$r_\rho(u, v) = r_\rho(u - v) \doteq r_\rho(\tau) \quad (18)$$

In this case, (5) can be solved using classical methods of *spectral decomposition* [Van Trees, 2004, Ch. 3], and the  $\lambda_i$  in (5) are related to the contribution of the different frequency components in the spectrum of  $r_\rho(\tau)$ . Examples of practical covariance stationary random fields include exponentially decaying cross-correlation functions of the form:

$$r_\rho(\tau) = A \exp\left(-\frac{|\tau|^q}{\sigma}\right), \quad q = 1, 2 \quad (19)$$

which can be used to model correlation functions that decay with the increase of distance between points. The eigenvalues and eigenfunctions of such correlation functions can be found in [Van Trees, 2004, Ghanem and Spanos, 2003, Holmes et al., 1998].

### IV. EXTENSION TO NON-HOMOGENEOUS MEDIA

The basic equation (1) is the solution of Poisson's equation in homogeneous media:

$$\nabla \cdot [-\nabla \sigma(r) \phi(r; t)] = \rho(r; t) \quad (20)$$

The general problem can be studied using this equation, which enables us to consider conductivity mismatch between boundaries of brain layers. Refer to [Nunez and Srinivasan, 2006, P. 60] for the background. One can also use Green's function and related identities to convert surface integrals to volume integrals.

### V. GENERALIZATION AND FUTURE WORK

In our recent research, the hereby presented framework has been extended in various aspects, including spatially stationary versus nonstationary sources and extension to nonhomogeneous media. We have developed a simulation platform for generating synthetic signals with desired distributions and a given  $R_\phi(\mathbf{x})$  for EEG signals [Tavakkol-Shoorijeh, 2014]. Conventional PCA and ICA have been applied to these synthetic distributed signals and we show that the results are consistent with the ones obtained from real data (specifically ECG and EEG).

The findings of this research are expected to be helpful for better interpretation of ICA results, specifically for biomedical applications. We can show how reliable the components extracted by PCA and ICA be, when the sources are not punctual. The presented approach is hoped to be extended to higher order statistics (used in some ICA algorithms), and lead to the development of a general framework for studying spatially distributed sources, which we coin *distributed component analysis* (DCA).

Following this work, some of the major questions that should be answered are as follows:

- 1) An interesting question is what the sources extracted by conventional ICA and PCA correspond to. Our formulations show that they are projections of the distributed source onto specific basis functions; but the exact basis functions need to be derived.
- 2) Another important item that should be verified, is to show that for distributed sources, the components extracted by conventional PCA and ICA generally change with the location of the sensors (unlike conventional ICA). Do you know any good reference that shows this for conventional ICA (except for the variance and order of sources)? Perhaps, one approach can be to show that the *mutual information* of the data changes with the change of electrode positions for distributed sources.
- 3) Following the previous item, an interesting question is whether we can find *sensor position-invariant* sources corresponding to distributed sources. Note that in this approach we are not interested in independent or uncorrelated components. We just want to find similar components using any electrode configuration. This is very similar to the idea that we had in [Sameni, 2008] for canonical representations of the ECG.
- 4) It is believed that similar, but more complicated, equations can be derived for *cumulant tensors* [Hyvärinen et al., 2001, Ch. 11], of  $\phi(t)$ . These tensors can be used to relate ICA solutions based on higher-order statistics to the aforementioned physical properties of the distributed source.
- 5) In the derivation of the hereby presented equations, we made no specific assumptions on the nature of the source density  $\rho(u; t)$ . For the case of ECG signals, taking  $p = 1$  (in (1)), this density function can represent  $\nabla \cdot \mathbf{J}$ , the divergence of the cardiac impressed current dipole moment density defined in [Sameni, 2008, Ch. 3].
- 6) In future studies, a more realistic formulation is to consider non-homogeneous propagation media, which as mentioned in [Sameni, 2008, Ch. 3], leads to some sort of ‘reflection’ sources due to the surfaces of conductivity discontinuity. By using Green’s function and analytic identities to convert surface integrals to volume integrals it should be possible to study non-homogeneous distributions with similar formulations.
- 7) The hereby adopted framework was based on the integral form of Poisson’s equation. In future works, one can analyse the problem from a convolution (filtering) perspective, which may lead to a filter bank formulation

of the problem.

- 8) Overall, what is the interpretation of the components extracted by conventional ICA over distributed sources?

## VI. FURTHER READING

- 1) An excellent book: [Ghanem and Spanos, 2003].
- 2) [Van Trees, 2004, Ch. 3] has some related formulations.
- 3) See [Arfken et al., 2005] for the nonhomogeneous case using Green’s functions expansions.
- 4) The idea of *clouds of sources*: <http://sccn.ucsd.edu/~arno/indexica.html>
- 5) [De Lathauwer et al., 2000] has referred to the model; but does not study components higher than 3 and has not given physiological interpretations for the extracted components.
- 6) See [Bronzino, 2000, Ch. 16] for a propagation model.

## REFERENCES

- G.B. Arfken, H.J. Weber, and F.E. Harris. *Mathematical Methods For Physicists International Student Edition*. Elsevier Science, 2005. ISBN 9780080470696.
- J. Bronzino, editor. *The Biomedical Engineering Handbook*. CRC Press, Inc., second edition, 2000.
- P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Independent Component Analysis and Applications Series. Elsevier Science, 2010. ISBN 9780080884943.
- L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by blind source subspace separation. *Biomedical Engineering, IEEE Transactions on*, 47: 567–572, May 2000.
- Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Dover Publications, 2003.
- Philip Holmes, John L. Lumley, and Gal Berkooz. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press, 1998.
- A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, 2001.
- Paul L Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2006.
- R. Sameni, C. Jutten, and M. B. Shamsollahi. What ICA Provides for ECG Processing: Application to Noninvasive Fetal ECG Extraction. In *Proc. of the International Symposium on Signal Processing and Information Technology (ISSPIT’06)*, pages 656–661, Vancouver, Canada, August 2006.
- Reza Sameni. *Extraction of Fetal Cardiac Signals from an Array of Maternal Abdominal Recordings*. PhD thesis, Sharif University of Technology – Institut National Polytechnique de Grenoble, July 2008. URL <http://www.sameni.info/Publications/Thesis/PhDThesis.pdf>.
- Behnam Tavakkol-Shoorijeh. *Distributed Component Analysis and its Applications in Biosignal Processing*. Master’s thesis, Biomedical Engineering, School of Electrical & Computer Engineering, Shiraz University, September 2014. Supervised by: Dr. Reza Sameni.

H.L. Van Trees. *Detection, Estimation, and Modulation Theory*. Detection, Estimation, and Modulation Theory. Wiley, 2004. ISBN 9780471463825.