SEPARATION OF POLYNOMIAL POST NON-LINEAR MIXTURES OF DISCRETE SOURCES

Boaz Lachover and *Arie Yeredor*

School of Electrical Engineering, Tel-Aviv University

ABSTRACT

We consider the problem of blind estimation of the parameters of noisy non-linear mixtures of sources with unknown discrete alphabets. The nonlinear mixtures are modeled using the "post non-linear" model, in which the source signal undergo a linear mixture first, and then each mixed signal undergoes an unknown nonlinear transformation. The individual nonlinear transformations are modeled in this paper as second-order polynomials, whose parameters are unknown. Using the Estimate-Maximize algorithm, we derive estimators for all the unknown parameters. We also computed the Cramér-Rao Lower Bound for the estimation, to which the obtained mean squared estimation error is empirically compared.

1. INTRODUCTION

Blind source separation (BSS) concerns the separation of mixed sources when nearly no information about mixing, nor about the sources, is available. Assume M statistically independent sources $s_1(t), \ldots, s_M(t)$, and M observations $x_1(t), \ldots, x_M(t)$. In the linear BSS model each observation consist of a linear mixture of the sources, that is $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, where $\mathbf{A}$ is an unknown mixing matrix. A sufficient condition for linear separation in this model is to achieve M statistically independent outputs $\mathbf{y} = \mathbf{B}\mathbf{x}(t)$ as estimates of the sources, where $\mathbf{B}$ serves as a separating matrix.

Though the basic linear mixture BSS serves as a reasonable model in some problems, in many practical applications the observed mixtures do not fit a linear model, thus the use of non-linear models may be required. In those cases, still assuming a static, memoryless mixture, one can define the observations as follows:

$$x_m(t) = F_m(\mathbf{s}(t)) \qquad m = 1, 2, \ldots, M \tag{1}$$

where $F_m(\mathbf{s}(t))$ is some nonlinear function involving the sources vector $\mathbf{s}(t)$.

In [1], Taleb shows that using the statistical independence as the sole criterion for separation in non-linear mixtures is not a sufficient condition, thus the discussion needs to be restricted to certain sets of possible non-linear functions, with associated sources' distributions in order to attain separation.

An interesting special case of a non linear model is the post non linear (PNL) mixture in which sources are first linearly mixed and then each of the mixtures is distorted by an unknown individual invertible non linear function. that is:

$$x_m(t) = f_m(\mathbf{a}_m\mathbf{s}(t)) \qquad m = 1, 2, \ldots, M \tag{2}$$

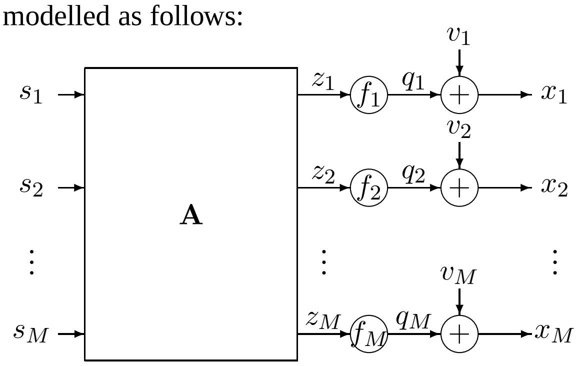
where, $\mathbf{a_m}$ is the m^{th} row of $\mathbf{A}$, and $f_m(\cdot)$ is the m^{th} (unknown) non linear function.

In order to achieve separation in the PNL problem, one should find a separating matrix as well as some inverse nonlinear functions, $g_1(\cdot), \ldots, g_M(\cdot)$, in order to mitigate the nonlinearity. In [2] Taleb and Jutten propose a parameterization of the inverse functions. Then by minimizing the outputs' mutual information with respect to the PNL mixture parameters, the estimates of the sources are constructed. In [3] and in [4] Ziehe *et al.* resolve the non-linearity by either maximizing the outputs' correlation or by bringing the outputs to be as Gaussian as possible. A similar idea of trying to "Gaussianize" the sources using non-linear functions, is also used in [5] as an initialization procedure for PNL BSS algorithms.

A geometrical approach to PNL BSS mixture is suggested by Babaie-Zadeh *et al.* in [6], for cases where the sources' probability distributions have a bounded support, with "sufficient" probability to attain the boundary value. This approach can generally be expanded for discrete sources. As many "man made" sources are indeed discrete (a.k.a. digital), the associated BSS problem is of considerable interest. However, although linear BSS of discrete sources has been treated quite extensively (e.g., [7], [8], [9] and [10]), the PNL model does not seem to have been addressed in that context, especially under noisy conditions - hence the motivation for this paper.

2. THE NOISY PNL MODEL FOR DISCRETE SOURCES

We consider the noisy PNL model for discrete sources, modelled as follows:



which can be written as:

$$\mathbf{x}[t] = \mathcal{F}(\mathbf{A}\mathbf{s}[t]) + \mathbf{v}[t] \qquad t = 1, 2, \ldots, T \qquad (3)$$

where $\mathbf{v}[t]$ is an $M \times 1$ zero mean spatially white Gaussian noise vector with independent, identically distributed (i.i.d.) elements and $\mathcal{F}$ denotes the *individual* non-linear functions, that is $\mathcal{F}(\mathbf{z}) = (f_1(z_1), \ldots, f_M(z_M))^T$. Each discrete source can take a value from a finite set of values (alphabet), that is: $s_m[t] \in \{S_m^{(n_m)}\}_{n_m=1}^{N_M}$. The values of $S_m^{(n_m)}$ will be assumed unknown as part of the blind framework, though we shall assume that the alphabet sizes N_m are known.

For succinct parameterization, the scaling ambiguity is resolved by assuming that the diagonal entries of $\mathbf{A}$ are all ones[1].

[1]This is a valid "working assumption" as long as the true diagonal values are nonzero - which is part of our model assumption.

Following Taleb's result in [1], regarding the need to restrict the possible set of non-linear functions, in the discrete sources PNL model the non-linear functions can map the finite set of possible observations vectors to nearly any possible sets of vectors. Therefore, to ensure identifiability, the set of allowable non-linear functions needs to be restricted. In this work the non-linear distortions are modelled as quadratic polynomials (higher orders could be considered as well, as long as they follow the above restriction, but for simplicity reasons we restricted the discussion in this work to second order polynomials). In order to resolve inherent scaling commutation and offset ambiguities between the polynomial parameters, the mixing parameters and the sources' alphabets, each non-linear operation is parameterized as

$$f_m(x) = b_m x^2 + x \quad , \quad m = 1, 2, ...M. \qquad (4)$$

This assumption further restricts the non-linear functions to pass through the origin - yet this is often a reasonable restriction.

A linear mixture of M discrete sources creates a discrete parallelogram in the M dimensional distributions space. The further non-linear operations cause the internal and external parallels of the parallelogram to bend.

In the noiseless case the unknown parameters can be calculated (algebraically) with absolute accuracy from a small subset of the points in this parallelogram (which can be obtained from just a few noiseless observations, assuming they exhibit the required "richness"). However, in the noisy case no true points of the parallelogram are available, and the parameters have to be estimated from the noisy data, with estimation accuracy that improves with the observation length.

Thus, in the case of additive zero-mean Gaussian noise (independent of the sources), the noisy observations' distribution can be considered a multivariate Gaussian mixture, with the bent parallelogram's internal and external points as the associated means. As all observations are assumed i.i.d. in time, the likelihood function of this mixture is given by:

$$\mathbf{L}(\boldsymbol{\theta}) = \log(f_{\mathbf{x}[T],\ldots,\mathbf{x}[1]}(\mathbf{x}[T],\ldots,\mathbf{x}[1];\boldsymbol{\theta})) = \sum_{t=1}^{T}(\log(f_{\mathbf{x}[t]}(\mathbf{x}[t];\boldsymbol{\theta}))) = \sum_{t=1}^{T}\log\sum_{n_1=1}^{N_1}\cdots\sum_{n_M=1}^{N_M}\frac{\tilde{p}(\mathbf{n})e^{-\sum_{j=1}^{M}\frac{\left(\Delta_j(\mathbf{n},t)\right)^2}{2\sigma_j^2}}}{(2\pi)^{\frac{M}{2}}\prod_{m=1}^{M}\sigma_m}, \quad (5)$$

where

$$\Delta_j(\mathbf{n},t) = x_j[t] - b_j\left(\sum_{i=1}^{M}a_{ji}S_i^{(n_i)}\right)^2 - \sum_{i=1}^{M}a_{ji}S_i^{(n_i)} \quad (6)$$

and $\tilde{p}(\mathbf{n})$ denotes the a-priori probability that the indices of the vectors are given by $\mathbf{n} \triangleq (n_1, n_2, \ldots, n_M)$. The parameters vector $\boldsymbol{\theta}$ includes the sources ensemble of alphabet values, the mixing matrix elements, the non-linear distortions parameters $\{b_m\}_{m=1}^{M}$, the a-priori probabilities and the noises variances $\{\sigma_m^2\}_{m=1}^{M}$. However, for simplification of the derivations, we adopt the practically reasonable assumption, that for each source all alphabet values are equiprobable (and the sources are statistically independent), and that the variances of the additive noise sources are known to be equal, and take a nominal unit value; Namely,

$$\tilde{p}(\mathbf{n}) = \prod_{m=1}^{M}(N_m)^{-1}\ \forall\mathbf{n} \quad , \quad \sigma_m^2 = 1\ \forall m \quad (7)$$

Maximizing $\mathbf{L}(\boldsymbol{\theta})$ with respect to the entire set of remaining unknown parameters: The off-diagonal mixing coefficients $\{A_{nm}\}_{n\neq m}$, the polynomial coefficients $\{b_m\}$ and the sources' alphabets $\{S_n^{(n_m)}\}$, is a mathematically demanding task. Therefore, the Estimate-Maximize (EM) algorithm (e.g., [11], [12]) is employed in this specially constrained (and specially parameterized) Gaussian Mixture Model (GMM). Using auxiliary "index-vectors" $\mathbf{z}[t]$ (indicating the true indices of the sources' symbols at each sample) as the unavailable data to form the "complete data", the EM auxiliary function is given by:

$$Q(\boldsymbol{\theta},\boldsymbol{\theta}^k) = C - T\log(N) - \frac{1}{2}\sum_{t=1}^{T}\sum_{d=1}^{N}\sum_{j=1}^{M}\left(\Delta_j(\tilde{\mathbf{n}}^d,t)\right)^2 p(\mathbf{z}[t]=\tilde{\mathbf{n}}^d|\mathbf{x}[t],\boldsymbol{\theta}^k) \quad (8)$$

where $p(\mathbf{z}[t]=\tilde{\mathbf{n}}^d|\mathbf{x}[t],\boldsymbol{\theta}^k)$ is the probability that the t^{th} observation vector was attained due to the $\tilde{n}^d$ vector of sources' indices (out of a possible $\prod_{m=1}^{M}N_m$ possible vectors of sources' indices), given the observation vector and an estimate $\boldsymbol{\theta}^k$ to the parameters. Those probabilities are given by:

$$p(\mathbf{z}[t]=\tilde{\mathbf{n}}^d|\mathbf{x}[t];\boldsymbol{\theta}^k) = \frac{exp\left(-\sum_{j=1}^{M}\left(\Delta_j(\tilde{\mathbf{n}}^d,t)\right)^2\right)}{\sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2}\cdots\sum_{n_M=1}^{N_M}exp\left(-\sum_{j=1}^{M}\left(\Delta_j(\mathbf{n},t)\right)^2\right)} \quad (9)$$

Using those probabilities, the minimization of the last term in (8) with respect to the $\mathrm{n_i}$-th value of the i-th source is attained by rooting the 3$^{\text{rd}}$ order polynomial:

$$c_3(S_i^{(\bar{n}_i)})^3 + c_2(S_i^{(\bar{n}_i)})^2 + c_1(S_i^{(\bar{n}_i)}) + c_0 = 0 \quad (10)$$

where:

$$c_3 \triangleq \sum_{t=1}^{T}\sum_{[n_i]}\sum_{j=1}^{M}\left(-2\alpha_{j,i}^2\right)p_{(n_i=\bar{n}_i)}[t]$$

$$c_2 \triangleq \sum_{t=1}^{T}\sum_{[n_i]}\sum_{j=1}^{M}\left(-3\alpha_{j,i}\beta_{j,i}\right)p_{(n_i=\bar{n}_i)}[t]$$

$$c_1 \triangleq \sum_{t=1}^{T}\sum_{[n_i]}\sum_{j=1}^{M}\left(2\gamma_{j,i}[t]\alpha_{j,i} - \beta_{j,i}^2\right)p_{(n_i=\bar{n}_i)}[t]$$

$$c_0 \triangleq \sum_{t=1}^{T}\sum_{[n_i]}\sum_{j=1}^{M}\left(\gamma_{j,i}[t]\beta_{j,i}\right)p_{(n_i=\bar{n}_i)}[t]$$

and:

$$\alpha_{j,i} \triangleq b_j a_{ji}^2$$

$$\beta_{j,i} \triangleq a_{ji}(2b_j\sum_{l=1,l\neq i}^{M}(a_{jl}S_l^{(n_l)}) + 1)$$

$$\gamma_{j,i}[t] \triangleq x_j[t] - (b_j(\sum_{l=1 l\neq i}^{M}a_{jl}S_l^{(n_l)})^2 + \sum_{l=1,l\neq i}^{M}a_{jl}S_l^{(n_l)})$$

$$\sum_{[n_i]} \triangleq \sum_{n_1=1}^{N_1}\cdots\sum_{n_{i-1}=1}^{N_{i-1}}\sum_{n_{i+1}=1}^{N_{i+1}}\cdots\sum_{n_M=1}^{N_M}$$

$p_{(n_i=\bar{n}_i)}[t] \triangleq p(\mathbf{z}[t] = (n_1, \ldots, n_M)|\mathbf{x}[t]; \boldsymbol{\theta}^k)$ where $n_i = \bar{n}_i$

Minimization of the last term with respect to the a_{ji} element in the mixing matrix also requires the rooting of a 3rd order polynomial:

$$c_3 a_{ji}^3 + c_2 a_{ji}^2 + c_1 a_{ji} + c_0 = 0 \qquad (11)$$

where:

$$c_3 \triangleq \sum_{t=1}^{T} \sum_{c_1=1}^{N_1} \ldots \sum_{c_M=1}^{N_M} \sum_{j=1}^{M} \left(-2\alpha_{j,i}^2\right) p_{\mathbf{n}}[t]$$
$$c_2 \triangleq \sum_{t=1}^{T} \sum_{c_1=1}^{N_1} \ldots \sum_{c_M=1}^{N_M} \sum_{j=1}^{M} \left(-3\alpha_{j,i}\beta_{j,i}\right) p_{\mathbf{n}}[t]$$
$$c_1 \triangleq \sum_{t=1}^{T} \sum_{c_1=1}^{N_1} \ldots \sum_{c_M=1}^{N_M} \sum_{j=1}^{M} \left(2\gamma_{j,i}[t]\alpha_{j,i} - \beta_{j,i}^2\right) p_{\mathbf{n}}[t]$$
$$c_0 \triangleq \sum_{t=1}^{T} \sum_{c_1=1}^{N_1} \ldots \sum_{c_M=1}^{N_M} \sum_{j=1}^{M} \left(\gamma_{j,i}[t]\beta_{j,i}\right) p_{\mathbf{n}}[t]$$

and:

$$\alpha_{j,i} \triangleq b_j (S_i^{(n_i)})^2$$
$$\beta_{j,i} \triangleq S_i^{(n_i)} (2b_j \sum_{l=1, l\neq i}^{M} (a_{jl} S_l^{(n_l)}) + 1)$$
$$\gamma_{j,i}[t] \triangleq (x_j[t] - (b_j (\sum_{l=1, l\neq i}^{M} (a_{jl} S_l^{(n_l)}))^2 + \sum_{l=1, l\neq i}^{M} (a_{jl} S_l^{(n_l)})))$$
$$p_{\mathbf{n}}[t] \triangleq p(\mathbf{z}[t] = (n_1, n_2, \ldots, n_M)|\mathbf{x}[t]; \boldsymbol{\theta}^k)$$

Minimization of this term with respect to the non-linear function parameter b_j attained by:

$$b_j = \frac{\sum_{t=1}^{T} \sum_{n_1=1}^{N_1} \cdots \sum_{n_M=1}^{N_M} \left(x_j[t] - \sum_{l=1}^{M} a_{jl} S_l^{(n_l)}\right) \left(\sum_{l=1}^{M} a_{jl} S_l^{(n_l)}\right)^2 p_{\mathbf{n}}[t]}{\sum_{t=1}^{T} \sum_{n_1=1}^{N_1} \cdots \sum_{n_M=1}^{N_M} \left(\sum_{l=1}^{M} a_{jl} S_l^{(n_l)}\right)^4 p_{\mathbf{n}}[t]}$$

3. SIMULATION RESULTS

We generated three discrete sources, each with equiprobable and uniform alphabet values of $0, 3, 6, 9$ for $s_1[t]$, $0, 3, 6, 9, 12$ for $s_2[t]$ and $0, 3, 6, 9, 12, 15$ for $s_3[t]$. The linear mixing was generated using

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 & 0.2 \\ 0.25 & 1 & 0.4 \\ 0.3 & 0.5 & 1 \end{bmatrix}$$

Each of the linear mixture outputs was independently distorted by a second degree polynomial (see (4)) using $b_1 = 0.008$, $b_2 = 0.01$ and $b_3 = 0.007$. Zero-mean unit-variance Gaussian white noise was independently added to each observation signal. Applying EM to observations sets of variable lengths T and averaging over 400 trials, we obtained the Mean Square Errors (MSEs) of all model parameters estimates. MSEs obtained for observation lengths $T = 500, 1000, 2000, 4000$ and 8000 are compared to the numerically calculated Cramér-Rao Lower Bound (CRLB) (see [13] for detailed derivation of the bound) in Fig. 1.

The MSE is seen to attain the bound, and both are, as expected, inversely proportional to T.

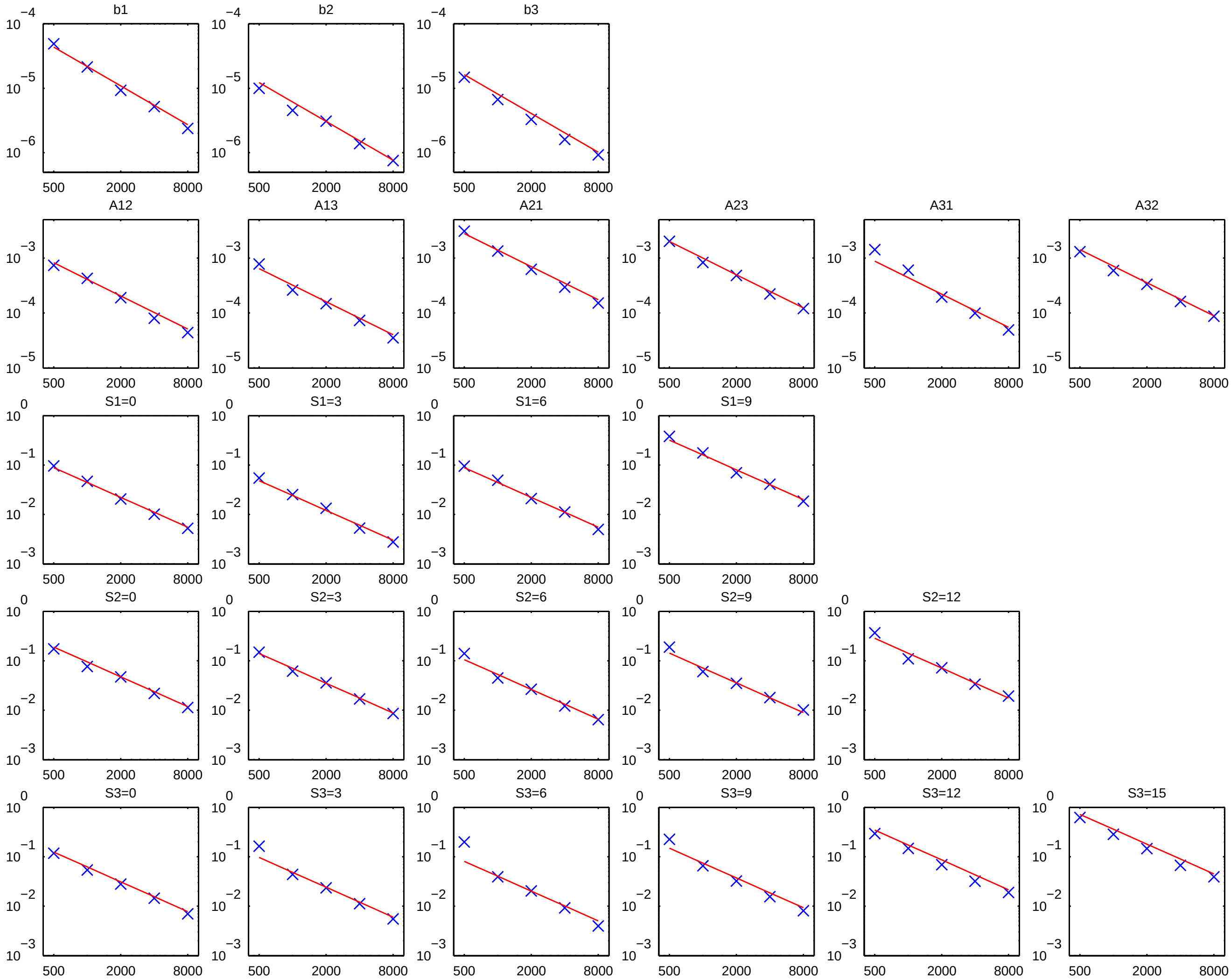


Figure 1: Estimates of the mixing parameters, polynomial parameters and sources' alphabets: MSE (blue X) compared to the CRLB (solid red) vs. the observations lengths T

4. REFERENCES

[1] A. Taleb, "A generic framework for blind source separation in structured nonlinear models," *IEEE Trans. Signal Processing*, vol. 50, pp. 1819–1830, Aug. 2002.

[2] A. Taleb and C. Jutten, "Source separation in post nonlinear mixtures," *IEEE Trans. Signal Processing*, vol. 47, pp. 2807–2820, Oct. 1999.

[3] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Muller, "Separation of post nonlinear mixtures using ace and temporal decorrelation," *proceedings ICA, San-Diego, California*, 2001.

[4] A. Ziehe, M. Kawanabe, S. Harmeling, and K.-R. Muller, "Blind separation of post nonlinear mixtures using gaussianizing transformations and temporal decorrelation," *proceedings ICA, Nara, Japan*, 2003.

[5] J.-S. Casals, M. Babaie-Zadeh, C. Jutten, and D.-T. Pham, "Improving algorithm speed in PNL mixture separation and wiener system inversion," *proceedings ICA, Nara, Japan*, 2003.

[6] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "A geometric approach for separating post non linear mixtures," *proceedings EUSIPCO, Toulouse, France*, 2004.

[7] F. Gamboa and E. Gassiat, "Source separation when the input sources are discrete or have constant modulus," *IEEE Trans. Signal Processing*, vol. 45, pp. 3062–3072, Dec. 1997.

[8] O. Grellier and P. Comon, "Blind separation of discrete sources," *IEEE Signal Processing Letters*, vol. 5, pp. 212–214, Aug. 1998.

[9] A. Belouchrani and J.-F. Cardoso, "Maximum likelihood source separation for discrete sources," *EUSIPCO, Edinburgh, Great Britain*, 1994.

[10] A. Belouchrani and J.-F. Cardoso, "Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation," *EUSIPCO, Edinburgh, Great Britain*, 1994.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistics society*, pp. 1–38, 1977.

[12] J. Bilmes., "A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," 1997.

[13] B. Lachover, "Blind separation of discrete sources in a noisy environment," Master's thesis, Department of EE-Systems, Tel-Aviv University, Israel, 2005.