LOW COMPLEXITY ITERATIVE MAXIMUM LIKELIHOOD DETECTION FOR MULTIUSER MIMO SYSTEMS

Ejaz Khan

Department of Electronic and Electrical Engg.
University College Dublin
Belfield, Dublin 4, IRELAND.
email: ejaz.khan@ee.ucd.ie

ABSTRACT

The information capacity of wireless communication systems may be increased dramatically by employing multiple transmit and receive antennas. In this paper, we consider multiuser wireless communication system, employing multiple transmit and receive antennas. We assume that channel has been estimated reasonably well with training sequence or some blind method. We estimate symbols user-wise by Maximum Likelihood approach (ML) considering other users as interferers. Two models are considered for the symbols of the interferers, corresponding to Gaussian and discrete priors. In the latter case, in which the finite alphabet gets exploited for the Multiple Access Interference (MAI) symbols, a simplification for the posterior MAI symbol probabilities is introduced based on Mean Field Theory.

1. INTRODUCTION

Multiple Input Multiple Output (MIMO) system has gained much interest recently [3,6]. Deploying multiple antennas at both, the base station and the remote stations increase capacity of the wireless channels. The gain in capacity is because of diversity, spatial multiplexing, interference rejection and array gain. In order to fully exploit the advantages of an antenna array, one must know the channel that will distort the signal as well as well as interfering noise.

Most commonly used receivers in communications are follows:
Zero forcing (ZF) receiver separates cochannel signals but at the expense of increase in signal-to-noise (SNR) at the output of the receiver and because of noise enhancement the performance of the ZF receiver degrades at low SNR. In order to improve the performance of the receiver a natural choice is to minimize the overall error, which results in minimum mean square error (MMSE) receiver. Better results can be obtained if some valid constraints are used for detection. The third type of receiver is ML receiver but unfortunately the computational complexity grows exponentially in the number of users, in case of CDMA and in the number of antennas in case of MIMO systems. An alternative to ML (by enumeration) technique is to increase the likelihood function iteratively until local/global maximum is reached. This iterative technique is called expectation maximization (EM) algorithm. The EM algorithm is a broadly applicable approach to the iterative computation of ML estimates, useful in variety of incomplete-data problems, where algorithms such as the Newton-Raphson method may turn out to be more complicated. On each iteration of the EM algorithm, there are two steps- called expectation step or E-step and maximization step or M-step. The basic idea of the EM algorithm is to associate with the given incomplete-data problem, a complete-data problem for which ML estimation is computationally more tractable.

In [8] joint channel estimation and decoding for linear MIMO systems has been carried out assuming short training sequence for channel estimation. In [9] the authors consider an algorithm based on Expectation-Maximization (EM) for the problem of separating superimposed digitally modulated signals impinging on an antenna array. They found that their algorithm closely resembles previously proposed methods based on Iterative Least Squares (ILS) techniques [10]. They also used SAGE algorithm [11] to improve the performance of their system. In this paper, we consider the problem of estimation of symbols user-wise (i.e., considering other users as interferers). We use two approaches for its estimation. In the Gaussian prior case [4], only the Multiple Access Interference (MAI) are modeled as stationary (white) sequences. We use ML formulation that gets implemented via Expectation Maximization (EM) algorithm. Alternatively, we consider exploiting the finite alphabet for the MAI symbols, leading to significant MAI reduction capability. To simplify and to reduce the complexity of the resulting EM algorithm in the second ap-

Research funded by Science Foundation Ireland (under principle investigator grant to Dr. Conor Heneghan)

proach, we consider the introduction of Mean Field methods for the approximation of the posterior MAI symbol probabilities. The paper is organized as follows: In section 2, we define the communication model. In section 3 we describe EM algorithm. Section 4 describes user-wise symbol estimation with Gaussian prior on MAI. In section 5, we describe user-wise symbol estimation procedure using discrete prior on MAI symbols. Conclusions are drawn in the last section.

2. COMMUNICATION MODEL

We model a wireless communication system with K users. Each user is equipped with N transmit antennas. The base station has M receive antennas. We assume flat fading between each transmit-receive pair. We denote $\alpha_{m,n}$ as complex fading gain from the n^{th} transmitter antenna to the m^{th} receive antenna, where $\alpha_{m,n} \sim N_c(0,1)$ is assumed to be zero mean circularly symmetric complex Gaussian random variable with unit variance. This is equivalent to the assumption that signals transmitted from different antennas undergo independent Rayleigh fades. It is also assumed that the fading gains remain constant over the entire signal frame, but they may vary from one frame to another. The received discrete time signal at instant t can be written as

$$x_t = Hd_t + n_t, \qquad (1)$$

where $d_t = [d_{1t}^T d_{2t}^T \cdots d_{Kt}^T]^T$, is the symbol vector. $x_t = [x_{1t} x_{2t} \cdots x_{Mt}]^T$, is the received signal, $n_t = [n_{1t} n_{2t} \cdots n_{Mt}]^T$ is a Gaussian noise vector. $d_{it} = [d_{it}^1 d_{it}^2 \cdots d_{it}^N]^T$ is a vector consisting of symbols transmitted from N transmit antennas at an instant t. $d_{it}^i \in \{-1, 1\}$. $(.)^T$ is the transpose operator. Channel matrix H is given by

$$H = [H_1 H_2 \cdots H_K], \qquad (2)$$

where H_i is as follows

$$H_i = \begin{pmatrix} \alpha_{1,1}^i & \alpha_{2,1}^i & \cdots & \alpha_{N,1}^i \\ \alpha_{1,2}^i & \alpha_{2,2}^i & \cdots & \alpha_{N,2}^i \\ \vdots & \vdots & \ddots & \\ \alpha_{1,M}^i & \alpha_{2,M}^i & \cdots & \alpha_{N,M}^i \end{pmatrix} \qquad (3)$$

3. EM FRAMEWORK FOR MAXIMUM LIKELIHOOD ESTIMATION

First of all, we briefly describe EM algorithm. EM algorithm [2,5] is an iterative approach to Maximum Likelihood Estimation (MLE), originally formalized in (Demster, Laird and Rubin). Each iteration is composed of two steps: an expectation (E) step and a maximization (M) step. The aim is to maximize the loglikelihood $l(\theta; D) = logL(\theta; D)$, where θ are parameters of the model and D are the data. Suppose that this optimization problem would be simplified by the knowledge of the additional variable χ, known as missing or hidden data. The set $D_c = D \cup \chi$ is refered to as the complete data set (in the same context D is refered to as incomplete data set). Correspondingly, the loglikelihood function $l_c(\theta; D_c)$ is refered to as complete data likelihood. χ is chosen such that the function $l_c(\theta; D_c)$ would be easily maximized if χ were known. However, since χ is not obsevable, l_c is a random variable and cannot be maximized directly. Thus, the EM algorithm relies on integrating over the distribution of χ, with the auxiliary function $Q(\theta, \hat{\theta}) = E_\chi[l_c(\theta; D_c|D, \hat{\theta}]$, which is the expected value of the complete data likelihood, given the observed data D and the parameter $\hat{\theta}$ computed at the previous iteration. Intuitively, computing Q corresponds to filling the missing data using the knowledge of the observed data and previous parameters. The auxiliary function is deterministic and can be maximized. An EM algorithm iterates the following two steps, for k=1,2,...., until local or global maximum of the likelihood is found.

Expectation: Compute

$$Q(\theta; \theta^{(k)}) = E_\chi[l_c(\theta; D_c|D, \theta^{(k)}] \qquad (4)$$

Maximization: Update the parameters as

$$\theta^{(k+1)} = arg_{max_\theta} Q(\theta; \theta^{(k)}), \qquad (5)$$

In some cases, it is difficult to analytically maximize $Q(\theta; \theta^{(k)})$, as required by the M-step of the above algorithm, and we are only able to compute a new value $\theta^{(k+1)}$ that produces an increase of Q at each iteration. In this case we have so called generalized EM (GEM) algorithm.

4. USER-WISE CHANNEL-SYMBOLS ESTIMATION WITH GAUSSIAN MAI PRIOR

The received signal is given by the Eq.1. We assume that channel has already been estimated. Each user channel is modeled as Gaussian vector which might be correlated in space, i.e., between antennas, but are assumed independent between users. The channel vector for user i can be written as $h_j^i \in N(0, R_{h_j h_j})$. In the first approach we assume that the interfering symbols as Gaussian i.i.d. random variables with known variance σ_r^2. Given T snapshot, i.e., $\{x_t\}_1^T$, we are now ready to define the complete data set. The complete data set is chosen as $\{x, d_r, H\}$, where d_r is the group of the interfering users' information bits transmitted at all time instants and H is the channel matrix and x is composed of the received vector from time instant 1 to time instant T . Without loss of generality, we will detect user 1 first. The pdf of the complete data set is given by

$$f(x, d_r, H; d_1) = f(x|H, d) f(d_r, ; d_1), \qquad (6)$$

d_1 vector is composed of user 1 transmitted data at all time instants, $f(x|H, d)$, and $f(d_r; d_1)$ are given by

$$f(x|H,d) = K_1 exp(\frac{-1}{\sigma^2}(x - Hd)^H(x - Hd)), \quad (7)$$

where K_1 is constant not depending on parameters to be estimated, $(.)^H$ is the Hermitian transpose and

$$f(d_r; d_1) = K_2 exp(\frac{-1}{2\sigma_r^2} d_r^T d_r), \quad (8)$$

where K_2 is another constant. In the above equation we have assumed without loss of generality that the prior mean for the interfering users' symbols is zero and the variance σ_r^2 of the symbols is known.

Having the above equations we are now ready to evaluate the E-step of the algorithm. Since we are conditioning on the received data, we take expectations with respect to d_r (interfering users' symbols).

$$Q(d_1; d_1^{(k)}) =$$
$$E\{\log f(x, d_r, H; d_1|x, H; d_1^{(k)})\}, \quad (9)$$

where $(.)^{(k)}$ is the iteration index and E is the expectation operator.

Evaluating the expectations and dropping the terms that do not depend on the parameters the above equation can be written as

$$Q(d_1; d_1^{(k)}) = \sum_{i=1}^{T} E\{-(x_i - H_1 d_{1i} - H_r d_{ri})^H$$
$$(x_i - H_1 d_{1i} - H_r d_{ri}) - d_r^T d_r | x, H; d_1^{(k)}\}, \quad (10)$$

d_{ri} are the symbols transmitted by interfering users (with N transmit antennas each) at time instant i, x_i is the received signal at instant i, d_{1i} is the transmitted data vector of user 1 at instant i, H_1 is the channel matrix for user 1 and H_r is the channel matrix for the interfering users. $H = [H_1|H_r]$. The symbols d_{1i} are obtained by maximizing Eq.10 over BPSK. It is clear from Eq.10 that we need

$$\hat{d_{ri}} = E\{d_{ri}|x, H; d_1^{(k)}\}, \quad (11)$$

in addition to $\hat{d_{ri}}$ we also need second order moment of d_{ri}, which can be easily evaluated once the conditional means are available.

Now the problem is to derive the expressions for $\hat{d_{ri}}$, i.e., the conditional mean of the interfering users bit. R_{xx} is given by

$$R_{xx} = E\{H_1 d_1 d_1^H H_1^H + H_r d_r d_r^H H_r^H\} + \sigma^2 I, \quad (12)$$

where d_1 and d_r are the data vector composed of transmitted symbols at all time instants of user 1 and the rest of the users respectively. In deriving the above equation, we used the fact that $E\{d_r\} = 0$. From now we will omit the EM iteration index, i.e., k. Using Bayes formula we can write the conditional pdf of d_{ri} as a function of known pdfs is follows, after neglecting irrelavant terms (using the fact that transmitted symbols at instant i results in received vector at the same instant),

$$f(d_{ri}|x, H; d_1) = f(x_i|H, d_i) f(d_{ri})/Z_1, \quad (13)$$

where d_i is the vector of symbols of all the users at instant i, x_i is the received vector at instant i, d_{ri} are the interfering users data vector transmitted at instant i, and H is the channel matrix. Z_1 is independent of d_{ri}, hence neglected. Substituting the corresponding expressions and rearranging gives

$$f(d_{ri}|x; H, d_1) = K_1 K_2 exp(-\frac{1}{\sigma^2}(x_i - Hd_i)^H(x_i - Hd_i)$$
$$-\frac{1}{2\sigma_r^2} d_{ri}^T d_{ri}). \quad (14)$$

Since the conditional pdf of $\hat{d}_r$ will be Gaussian, it is easy to show that

$$\hat{d_{ri}} = \frac{R_{dd}}{\sigma^2}(H_r^H x_i - H_r^H H_1 d_{1i}), \quad (15)$$

where

$$R_{dd}^{-1} = \frac{1}{\sigma^2} H_r^H H_r + \frac{I}{2\sigma_r^2}, \quad (16)$$

where I is identity matrix.

The algorithm detects user-wise symbols. First, user 1 symbols are estimated from the above procedure. Then the contribution of that user is subtracted from the received signal to get more clean signal. Then the user second is detected. The same procedure is repeated for the other users. After convergence of the EM algorithm (for detecting the symbols of user 1), the solution of d_{1i} from Eq.10 is projected on finite alphabet to get the symbols estimate. The same process is done for the other users too. The overall algorithm works as follows:

1) First we initialize d_1,
2) We evaluate $\hat{d_{ri}}$ from Eq.15
3) These values are plugged into Eq.10 to get symbol update. These steps are repeated until convergence.

5. USER-WISE SYMBOL ESTIMATION USING DISCRETE MAI PRIOR

The EM algorithm for discrete MAI prior is

$$Q(d_1; d_1^{(k)}) = \sum_{i=1}^{T} E\{-(x_i - H_1 d_{1i} - H_r d_{ri})^H$$

$$(x_i - H_1 d_{1i} - H_r d_{ri}) | x, H; d_1^{(k)}\},$$

The steps for deriving the algorithm are essentially the same except that the conditional mean of d_{ri} will be different than previously discussed, i.e., Gaussian random variable for the priors, which will result in different symbols estimates. The conditional mean for d_{ri} is given by

$$\hat{d_{ri}} = E\{d_{ri}|x; H, d_1^{(k)}\} = \sum_{d_{ri}} d_{ri} f(d_{ri}|x, H; d_1^{(k)}). \quad (17)$$

From now for the sake of simplicity we will omit the EM iteration index, i.e., k. In order to calculate the conditional mean we have to evaluate the above expression, which is summation of all interfering users' symbols at instant i multiplied by their corresponding pdfs, which is computationally very expensive. Mean Field (MF) methods [1,7], provide tractable approximations for the computation of high dimensional sums and integrals in the probabilistic models. By neglecting certain dependencies between the random variables, a closed set of equations for the expected values of these variables are derived which often can be solved in a time that grows polynomially in the number of variables [1, chapter.2]. The MF approximation is obtained by taking the approximating family of probability distribution by all product distribution, i.e.,

$$Q(d_{ri}) = \Pi_j Q_j(d_{ri}^j).$$

We now choose a distribution which is close to the true distribution, i.e., $f(d_{ri}|x, H; d_1)$. The parameter of the distribution is chosen so as to minimize Kullback-Leibler (KL) distance, i.e.,

$$KL(Q \| f(d_{ri}|x, H; d_1) = \sum_{d_{ri}} Q(d_{ri}) \ln \frac{Q(d_{ri})}{f(d_{ri}|x; H, d_1)}, \quad (18)$$

where

$$Q(d_{ri}) = \Pi_{j=1}^{(K-1)N} Q_j(d_{ri}^j),$$

and $d_{ri}^j \in \{-1, 1\}$. $f(d_{ri}|x, H; d_1)$ can also be expressed as (after neglecting constant terms)

$$f(d_{ri}|x, H; d_1) = \frac{f(x_i|H, d_i)}{\sum_{d_{ri}} f(x_i|H, d_i)} = \frac{exp(-H(d_i))}{Z}, \quad (19)$$

where Z is independent of d_{ri}, $f(x_i|H, d_i)$ has the Gaussian distribution and the d_i is the vector of symbols of all users at instant i, i.e., $d_i = [d_{1i} \ \ d_{ri}]$. After some simplification $H(d_i)$ can be written as

$$H(d_i) = \frac{1}{\sigma^2}(-x_i^H H d_i - d_i^T H^H x_i + d_i^T H^H H d_i). \quad (20)$$

The above equation has the form

$$H(d_i) = \sum_{j,n} d_{ri}^j J_{jn} d_{ri}^n - 2\sum_j d_{ri}^j \theta_j + C, \quad (21)$$

where C is a term independent of d_{ri}, $J_{jn} = \frac{1}{\sigma^2}(H_r^H H_r)_{j,n}$, and θ_j can be defined in a similar fashion and $H = [H_1 | H_r]$. The KL distance between $Q(d_{ri})$ and $f(d_{ri}|x, H; d_1)$ can be written as

$$KL(Q \| f(d_{ri}|x, H; d_1) = \ln Z + V[Q] - S[Q], \quad (22)$$

where

$$S[Q] = -\sum_{d_{ri}} Q(d_{ri}) \ln Q(d_{ri}), \quad (23)$$

is the entropy and

$$V[Q] = \sum_{d_{ri}} Q(d_{ri}) H(d_i), \quad (24)$$

is the variational energy. The most general form of probability distribution for our problem (BPSK case) is

$$Q_j(d_{ri}^j; m_j) = \frac{1 + d_{ri}^j m_j}{2}, \quad (25)$$

where m_j is the variational parameter which corresponds to the mean, i.e., $m_j = E\{d_{ri}^j\}$. The entropy can be written as

$$S[Q] = -\sum_j \frac{1+m_j}{2} \ln \frac{1+m_j}{2} + \frac{1-m_j}{2} \ln \frac{1-m_j}{2}, \quad (26)$$

and similarly the variational energy can be written as

$$V[Q] = \sum_{j,n} J_{jn} m_j m_n - 2\sum_j m_j \theta_j. \quad (27)$$

In order to evaluate m_j we have to minimize the variational free energy, i.e.,

$$F[Q] = V[Q] - S[Q] \quad (28)$$

Differentiating this equation with respect to $m_j^{'}s$ gives non-linear fixed point equations, i.e.,

$$m_j = tanh(-\sum_n J_{jn} m_n + \beta_j), j = 1, 2 \cdots (K-1)N \quad (29)$$

In the matrix form we can write the above equation as

$$\mathbf{m} = \mathbf{tanh}(-\mathbf{Jm} + \beta), \quad (30)$$

where $\beta_j = 2\theta_j$. The above Mean Field theory (MFT) is called Naive MFT (NMFT) as it does not take correlations into account while approximating posteriori distribution. Below we describe a method to some how take correlations into account.

5.1. Linear response theory

In approximating the posteriori probability $f(d_r|x;H,d_1)$, the correlations were neglected, when $Q(d_r)$ is chosen to factorize, i.e.,

$$E_{exact}\{d_{ri}d_{ri}\} \simeq E_Q\{d_{ri}d_{ri}\} = E_Q\{d_{ri}\}E_Q\{d_{ri}\}, \tag{31}$$

where $E_Q\{.\}$ stands for expectation with respect to distribution Q. A correction to the estimate is found by differentiating the following equation

$$E\{d_{ri}\} = Z^{-1}\sum_{d_{ri}} d_{ri}e^{H(d_i)} \tag{32}$$

with respect to β_i to obtain linear response relation [1,12], i.e.,

$$\frac{\partial E\{d_{ri}\}}{\partial \beta_j} = E\{d_{ri}d_{rj}\} - E\{d_{ri}\}E\{d_{rj}\}. \tag{33}$$

The above relation is exact when expectation is taken according to exact probability distribution. However, if $E\{d_{ri}\}$ is reasonably well approximated with the mean field method, we can get the right hand side of the above equation by differentiating the left side of the equation with respect to β_j. In this way, we can improve the covariance and hence the second moment of the interfering users' bits which will result in improved symbol detection as compared to Naive Mean Field Theory (NMFT). NMFT does not take into account correlations between random variables. This improvement is gained at the expense of very little additional complexity. In [13] there is nice explaination about working of linear reponse theory. The huge computational task (complexity grows exponentially with the number of interfering users multiplied by the transmitted symbols per user) of exact averages over $f(d_{ri}|x,H;d_1)$ has been replaced by solving the above set of $(K-1)N$ nonlinear equations [see Eq. 30], which can be done in time that grows only polynomially. As the above equation is nonlinear there may be local minima or saddle points. In order to avoid it, the solution must be compared by their value of variational free energy $F[Q]$.

6. SIMULATIONS AND CONCLUSIONS

In this paper, we derived two receivers for user-wise symbols estimate. In the first approach, the Gaussian prior on the interfering users' symbols is assumed and the EM algorithm is used for user-wise symbols estimation. In the second proposed receiver a discrete prior is assumed on the interfering users' bits. In the later case, the complexity of computing the posteriori probabilities grows exponentially in the number of interfering users multiplied by the symbols per user. We derived low complexity method to circumvent this problem. The exact posteriori probabilities are replaced by the approximate separable distributions. The distributions are calculated by MFT (variational approach). Simulation results are shown in figure.1 for discrete prior case on MAI. The simulations were performed by considering two transmit and four receive antennas. The number of the users were two in the system. The solid line represents ML by enumeration, the star-solid line is ML using naive mean field approximation and the dashed line represents ML using linear response theory. It is clear from the figure that we get very close performance in the terms of the BER to the exact ML (ML by enumeration) approach. The linear response theory performs better than NMFT approach.

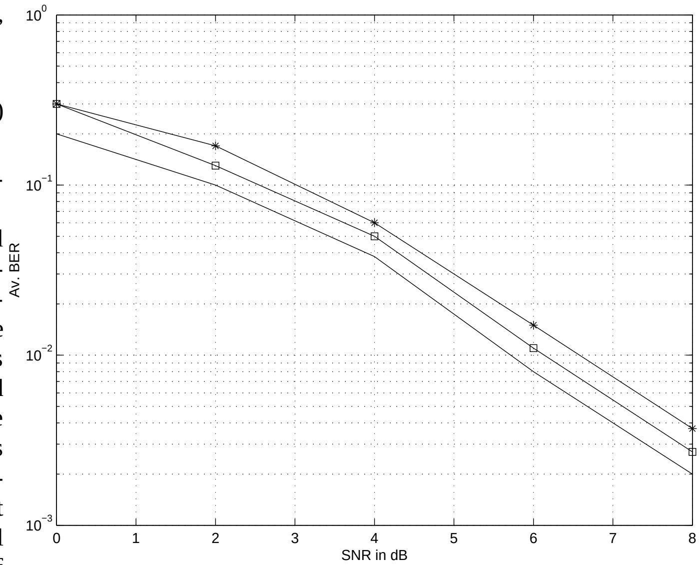

Fig. 1. Av. BER of K=2 N=2, M=4 vs $SNR(dB)$. Solid line is ML by enumeration, Star-solid line is ML using naive mean field theory, and dashed line is ML using linear response theory

7. REFERENCES

[1] M. Opper and D. Saad, eds, "Advanced mean field methods: Theory and practice". MIT press, 2001.

[2] A. B. Dempster, N. B. Liard, and D. B. Rubin, " Maximum likelihood from incomplete data via EM algorithm", Journal of Royal Statistical society B, vol.39, no.1, pp. 1-38, 1977.

[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," IEEE JSAC vol. 17, No. 3, March 1999.

[4] E.de. Carvalho and D.T.M. Slock, "Maximum likelihood FIR multi-channel estimation with Gaussian prior for the symbols," Proc. ICASSP, Munich, Germany, April 1997.

[5] G. McLachlan and T. Krishnan, "The EM algorithm and extensions", John Wiley ad sons Inc. 1997.

[6] V. Tarokh, A. Naguib, N. Seshadri, and A. R. calberbank, "Combined array processing and space-time coding", IEEE Tran. Information Theory, vol. 45, No. 4, May 1999.

[7] G. Parisi, ”Statistical field theory”, Addison Wesley, Redwood, CA, 1988.

[8] A. Grant, ”Joint decoding and channel estimation for linear MIMO channels”, Wireless comms. and networking conf. , Chicago, Sep. 2000.

[9] P. Bohlin, ”Iterative least square techniques with application to adaptive antennas and CDMA-systems”, Licentiate thesis, Chalmers Univ. of Tech., Goteborg, Sweden, 2001. .

[10] S. Talwar, M. Viberg, and A. Paulraj, ”Blind separation of synchronous Co-channel digital signals using an antenna array - Part I: Algorithms”, IEEE Trans. on Signal Processing, vol. 44(5), pp. 1184-1197, May, 1996.

[11] J. A. Fessler, and A. O. Hero, ”Space-Alternating Generalized Expectation Maximization Algorithm ”, IEEE Trans. on Signal Processing, vol.42(10). pp. 2664-2677, Oct. 1994.

[12] P. F. R. Sorensen, O. Winther, and L. K. Hansen, ”Mean field approaches to independent component analysis”, Neural Computation, pp.889-918, 2002.

[13] M. Opper, and O. Winther, ”Variational linear response ”, In advances in Neural Information Processing (NIPS), 2003.