A FEASIBLE BLIND EQUALIZATION SCHEME IN LARGE CONSTELLATION MIMO SYSTEMS

Xu Zhao and Mike Davies

Institute for Digital Communications
Joint Research Institute for Signal & Image Processing
School of Engineering & Electronics, the University of Edinburgh
Email: X.Zhao@ed.ac.uk , Mike.Davies@ed.ac.uk

ABSTRACT

In real time communications, system performance and computational complexities play key roles. Reducing the computational load and providing accurate performances are the main challenges in present systems. In this paper, a Blind Equalization (BE) with affordable complexity and good performance in large constellation MIMO systems is proposed. Saving computational cost happens both in the signal separation part and in signal detection part. First, based on Binary Phase Shift Keying (BPSK) or Quadrature amplitude modulation (QAM) signal characteristics, an efficient and simple nonlinear function for the Independent Component Analysis (ICA) is introduced. Second, using the idea of the sphere decoding (SD), we choose the soft information of channels smartly and overcome the so-called curse of dimensionality of the Expectation Maximization (EM) algorithm to enhance the final results.

Index Terms— Noisy ICA, BE, MIMO, EM

1. INTRODUCTION

In digital wireless communications, distortions introduced by fading and multipath propagation cause intersymbol interference (ISI) in received signals, producing errors in signal detection. Many equalizers are designed to compensate for the channel effects. As opposed to traditional techniques, blind equalization (BE) methods do not require training or pilot sequences, can utilise the bandwidth resources efficiently and can perform in a wider range of communication environments. So blind equalization has attracted a great deal of interest in the recent years.

Independent component analysis (ICA) [1] as a statistical technique, has received a lot of attention in the signal processing community. Using only the independence of the original signals, ICA identifies an unknown channel or mixing matrix first and then estimates source signals. It can usually estimate the source signals up to certain indeterminacies: arbitrary scaling and permutation. It is suitable for multiple input multiple output systems (MIMO) systems. Our work is concentrated on general complex ICA models with additive Gaussian noise.

$$Y = HS + N \tag{1}$$

Where $Y \in C^{n \times k}$ is the matrix containing observed signals from the sensors, and $S \in C^{n \times k}$ is the complex discrete source signals. $N \in C^{n \times k}$ is the matrix of the noise with covariance, Σ, which is uncorrelated with the source signals. $H \in C^{n \times n}$ is an unknown linear square matrix whose elements are drawn independently from a Rayleigh distribution and we assume that it is invertible. Note that, H is instantaneous narrow band model but we can not guarantee it is orthogonal.

ICA can blindly equalize MIMO systems [2]; however, the performance may not be good enough for communication applications when we consider common BER demands in wireless communication systems. In this paper, we propose a quasi-maximum likelihood (ML) method which combines the ICA and EM algorithms to implement blind equalization in digital MIMO systems. Especially we emphasize the case of large and dense constellation modulations. Section II introduces a simple nonlinear function for BE. To further refine the BER performance, a simple SD-EM solution is presented in section III and IV. Section V shows the simulation results and conclusions are given in section VI.

2. THRESHOLD NONLINEAR FUNCTIONS IN QAM MODULATION

In blind signal separation systems, nonlinear functions can reveal the high order correlation among the signals. Such high order correlations indicate mutual dependence, which then forms an error signal to drive the output signals to a state of higher independence. These high order statistics (HOS) are produced by the nonlinear functions implicitly. An important point is that good nonlinear functions are essentially defined by the pdf of the original source signals. So, when using entropy/nongaussianity as the cost function, under digital QAM modulation schemes, the optimal nonlinearity for complex valued data is

$$J(W) = E\{log \;\; p_{QAM}(W^H X)\} \tag{2}$$

where W is the unmixing matrix and X is the whitened received signals, which satisfies $E\{XX^H\} = I$. $P_{QAM} : R \times R \to R$ is the joint pdf of the QAM source with added noise. The signal model assumes the existence of noise, hence with the addition of complex white Gaussian noise to the sources, a mixture of Gaussian (MoG) kernels is appropriate. Then the model for the pdf in equation (2) for an M-QAM source with the Gaussian mixture model is

$$p_{QAM}(y) = \frac{1}{M 2\pi\sigma^2} \sum_{i=1}^{M} e^{\left(\frac{-1}{2\pi\sigma^2}((y^R - A_i^R)^2 + (y^I - A_i^I)^2)\right)} \tag{3}$$

where σ^2 is the variance of the Gaussian mixture. $A_i \in A$ is the set of complex points in the QAM constellation. Note that for convenience we are treating the noise as being isotropic in the source domain, as was also done in [3]. We note, however the better model is isotropic noise in the sensing domain, as in (1).

When the size of A increases, the computation load becomes very large and it is prohibitive for real time operation. To reduce this unaffordable complexity, we use a very simple nonlinear function to approximate this pdf [4]. For QAM signals, the real parts and the imaginary parts of the signal are statistically independent. So we can apply the same nonlinearity independently to the real and the imaginary parts of the signal. Figure 1 illustrates real part of MoG QAM16 pdf and its approximation and the corresponding nonlinear function based on this approximation.

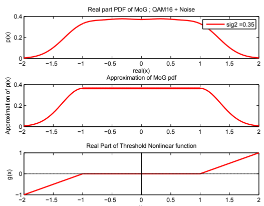

Fig. 1: Nonlinear approximation of MoG

The *Split Nonlinear Threshold Function* is given by

$$g(y) = \begin{cases} 0, & |y_R| < \nu; |y_I| < \nu \\ \alpha[(y_R - sign(y_R)\nu) & \\ +j(y_I - sign(y_I)\nu)], & |y_R| \geq \nu; |y_I| \geq \nu \\ \alpha(y_R - sign(y_R)\nu), & |y_R| \geq \nu; |y_I| < \nu \\ j\alpha(y_I - sign(y_I)\nu), & |y_R| < \nu; |y_I| \geq \nu \end{cases}$$

where ν is the threshold and α is the slope of nonlinearity which effects the convergence ability and stability. We set $\nu = 1$ and $\alpha = 0.5$ in the following simulations. The key advantages of this nonlinearity over the a direct use of (3) are:

- Simplicity - The nonlinearity only requires a small number of simple bit-level sign operations which can take the place of a large amount of multiply and add operations.
- Unimodel density model - It avoids potential problems of introducing spurious local mimima.
- Flexible - The single nonlinear activate function is applicable to all QAM modulation schemes irrespective of constellation size.

Applying this nonlinearity into a fast complex fixed-point algorithm [3], we get the following update,

$$w^{n+1} = -\frac{1}{2}E\{xg^*(y)\} + E\{g_a^{'}(y)\}w^n + E\{xx^T g_b^{'}(y)\}(w^n)^*. \tag{4}$$

The *split threshold nonlinear* update equations are given by

$$g(y) = \frac{1}{2}\alpha\Big\{g_\nu(y_R)[y_R - sign(y_R)\nu] + \\ jg_\nu(y_I)[y_I - sign(y_I)\nu]\Big\} \tag{5}$$

$$g_a^{'}(y) = \frac{1}{4}\alpha[g_\nu(y_R) + g_\nu(y_I)] \tag{6}$$

$$g_b^{'}(y) = \frac{1}{4}\alpha[g_\nu(y_R) - g_\nu(y_I)]. \tag{7}$$

Here, we define the *nonlinear select function*, $g_\nu(x)$, as

$$g_\nu(x) = \begin{cases} 0, & |x| < \nu \\ 1, & |x| \geq \nu. \end{cases}$$

Excellent simulation results based on this nonlinearity are shown in section V. Note that, since the nonlinearity above correctly treats the complex QAM signals as I-Q independent, there is no the phase ambiguity. Further improvements based on this property can be carried out for advanced refinements.

3. EM ALGORITHM AND ITS LIMITATIONS

The EM algorithm [5], can be used as a maximum likelihood estimator in the situation of incomplete data. The EM updates are analytically simple and numerically stable for distributions that belong to the exponential family, such as Gaussian. Considering model (1) again, we get a Gaussian observation model:

$$p(Y|S, H) = \frac{1}{(2\pi|\Sigma|)^N} \exp[-Tr(Y - HS)^H \Sigma^{-1}(Y - HS)]. \tag{8}$$

We assume that the noise covariance Σ can be estimated, and focus on estimating the channel parameter H. We can write the $p(Y)$ in terms of the hidden variables S. The parameters are estimated by maximizing the log likelihood. EMs principle is quite simple; perform two steps until convergence: the E-step is the computation of the conditional expectation of the complete likelihood, and the M-step is the maximization of this function.

$$E - step : Compute \qquad Q(H, H_k) = E\{\log p(S|Y, H)\} \tag{9}$$

$$M - step : Find \qquad H_{k+1} = Arg \max_H Q(H, H_k) \tag{10}$$

In this case, the solution is given [2],

$$H = \langle YS^H \rangle \langle SS^H \rangle^{-1} \tag{11}$$

where $\langle \cdot \rangle$ denotes average with respect to the source posterior $p(S|Y)$. The E-step comes from formula (9), and formula (11) is the solution of M-step. Here we emphasize an important advantage of the EM algorithm for digital communications. Generally speaking, the EM updating equations can be expressed in a gradient form with a specified step size controlled by the noise covariance. However, in contrast to the convergence property in the continuous source domains, for discrete sources, the EM algorithm exhibits approximate Newton behavior and enjoys fast, typically super-linear convergence in the neighborhood of the local minimum [2]. It guarantees fast convergence speed in discrete source domains. Although the EM algorithm has gained popularity in mixture analysis, it has several limitations:

Limitation 1: The EM algorithm is very sensitive to its initial values.
Limitation 2: The computation cost is exponential in the number of sources.

The limitation 1 can be explained by the fact that the EM is a local optimization method only [6]. We will clarify this difficulty in detail later.

The limitation 2 is the curse of dimensionality which is a well known problem with the EM algorithm. EM-MoG has computational cost that grows exponentially with the number of sources. When the number of sources increase, both E-step and M-step are difficult to solve. In digital MIMO systems, since the M-step is obtained from formula (11), the problem of high dimensions is due to the E-step, especially for large constellations. For example, in the case of 4 transmitters and QAM16 signals, there are $16^4 = 65536$ different configurations for the source symbols at each sampling. Then, each iteration of the E-step requires the computation of $65536 \times N$ conditional probabilities . Where, N is the frame length of each transmitted block. Such problem is known to be NP-hard. Hence, for high-rate systems with large number of antennas, direct calculation proves to be infeasible. One solution is to use a stochastic algorithm and replace the summations over all the possible hidden source states by Monte-Carlo integrations [7]. However, this is still not efficient enough for real time operations. We consider a deterministic approximation which we describe next.

4. SD-EM ALGORITHM

The following proposed algorithm overcomes the difficulty above by using soft sphere decoding to search over only the most probable hidden source points $s \in A_q^m$ that lie in a certain sphere of radius d around the received vector, where A_q^m denotes the n-dimensional hidden source points spanned by a q-QAM constellation in each dimension. Thereby reducing the search space and hence the required computational effort (see Figure 2). Clearly, the closest point inside the sphere assuming there is one will also be the closest point for the whole hidden source space. The summation over the points inside the dashed circle. Figure (2) takes the place of the summation over the points lying in the entire source space. The points inside the sphere are good in a likelihood sense and their collection builds up a set called as the *admissible set* and its number is called the *size of the admissible set*. We emphasize that closer scrutiny of this fundamental idea leads to two questions.

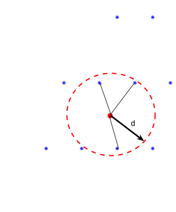

Fig. 2: Idea of choosing the admissible set.

How should we choose the size of the admissible set? Obviously, it connects with the radius d directly. If d is too large, we may obtain too many points and the algorithm will remain exponential in size, whereas if d is too small, we may not obtain enough points inside the sphere for correct information to the next iteration. Then reasonable choice of d plays an important role in the final performance and the reduction of the complexity. Considering the possibility of points inside the sphere is correlated to the SNR, we suggest when SNR is large, the size of the admissible set could be small. Otherwise, it is selected with a large value.

How can we determine which points are inside the sphere? Sphere decoding (SD) [8] provides a constructive answer this question. Figure (2) shows the basic principle of the SD algorithm. The star points represent the noiseless received constellation HS and the centre of the circle represents the actual received signal contaminated with noise. SD calculates the closest point by:

$$\hat{S}_{SD} = Arg \min_{S \in A_q^m} \parallel Y - HS \parallel^2 \leq d^2. \tag{12}$$

SD also can be explained as a tree search. The ML search solves this tree search successively, its complexity depends on the size of the tree. The SD only searches within the bracket, in the right picture in figure 3. With the introduction above, we can specify our algorithm:

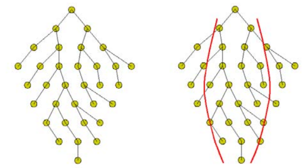

Fig. 3: Tree illustration of SD search

Algorithm 1 SD-EM algorithm

Input: received signals Y; The size of the admissible set; Iteration number of the EM algorithm.

1. Get the initial value of the channel state information, $\hat{H}$, by our *Threshold Nonlinear* ICA algorithm or any standard ICA.
2. Using sphere decoding approach with $\hat{H}$ to construct the admissible set A.
3. Updating the $\hat{H}$ by the EM algorithm only within the admissible set A.
4. Go back to step 3 with new $\hat{H}$ until the iteration ends.

Output: Estimate signals $\hat{S}$; $\hat{H}$.

Unlike the Monte Carlo EM (MCEM) algorithm, which approximates the conditional expectation in (9) by the Monte Carlo average. The essential idea behind this algorithm is that the SD-EM approximates the conditional expectation by the important samples which lie around the received signal vector.

5. SIMULATIONS

We test the separability performance of the *split threshold nonlinear* function in this paper by measuring the distance of estimation from the true value. We define it as: $P = W^H H_{real}$

$$ICI(P) = \frac{1}{n} \sum_i \sum_j \left[\left(\frac{|P_{ij}|}{max|P_{ij}|} \right)^2 - 1 \right]. \tag{13}$$

First, we compare it with the MoG pdf proposed in [3]. For QAM16. 4 transmitters and 4 receivers MIMO system. The channel H is a 4×4 complex instantaneous matrix, which is constant for each block interval (512 symbols), and it follows a Rayleigh fading distribution. N follows the complex additive white Gaussian distribution with diagonal covariance matrix and 1000 Monte Carlo runs are

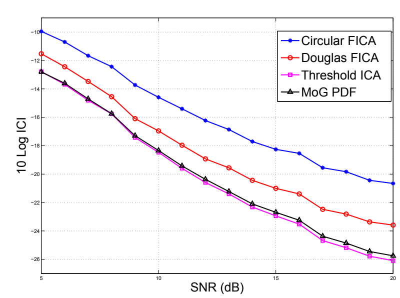

Fig. 4: Compararison of separation quality

taken. In this simulation, the performance of our algorithm (Threshold ICA) is tested against the complex FastICA (Circular FICA) [9] with nonlinearity $G(y) = log(0.1 + |y|^2)$. And a recent fixed-point algorithm (Douglas FICA) proposed in [10] that showed excellent performance for this kind of problem. Figure 4 shows that the performance of the *threshold nonlinear* ICA is always better than other methods and it has the same capability of the true MoG pdf over various SNR. Sometimes, it is even better. The reason is that the MoG pdf with small variance can introduce spurious mimima into the cost function due to the oscillatory nature of the pdf. Artificially increasing variance σ removes this problem. However the large variance can reduce the discrimination of the score function. Here we emphasize that compared with the nonlinear function based on the MoG pdf, the *Hyperbolic Nonlinearities* such as $tanh, sinh$, and the *Inverse circular Nonlinearities* such as $arctan, arccos$, the *nonlinear Threshold Function* requires only a small amount of bit-level sign operations. These always exist in real time processors such as DSP, FPGA, and it avoids many multiply and add operations. Such a property is important in real time systems.

In our final simulation a 4×4 system is set up: QAM16 with 512 symbols in each block, Rayleigh channel, 1000 runs. First, we use the threshold nonlinear ICA to obtain rough estimations. Then, using these as the initial values, we apply the SD-EM algorithm to futher improve the estimations. Here only 3 iterations of the SD-EM are used. The size of admissible set D is fixed at 17 which is much smaller than the entire 65536 configuration. Figure (5) shows substantial improvements by the SD-EM algorithm in terms of ICI.

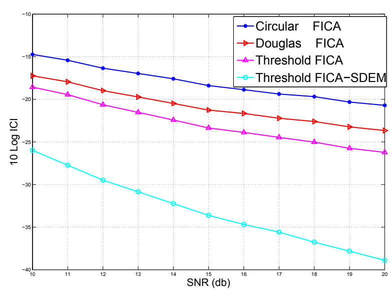

Fig. 5: Compararison of separation quality

In communiations, the final performance is BER. In figure (6), we compare our method with zero forcing (ZF) detector with exact channel state information (CSI) and the *threshold nonlinear* ICA. Also, we present the bound of the SD-EM method in which we initialize the SD-EM algorithm with the exact CSI. Using the same system configuration of the last simulation, we observe that the *threshold nonlinearity* can reach similar performance of the ZF detector and the SD-EM algorithm improves it significantly over all SNR.

However, there still exist a performance gap between the SD-EM algorithm and the optimal (with known CSI) ML solution in figure (6). We present the ML solution through the bar line associated with integer SNR. Note that the ML solution and the bound of the SD-EM almost are coincident with each other, which indicates to us that the performance gap is not introduced by the SD-EM algorithm. We speculate that the gap is due to the fact that the EM gets trapped in a local rather than global minimum, and that this occurs when the channel is close to singular or SNR is low. The number of local minima depends on the number of mixtures, the size and the dimension of the data.

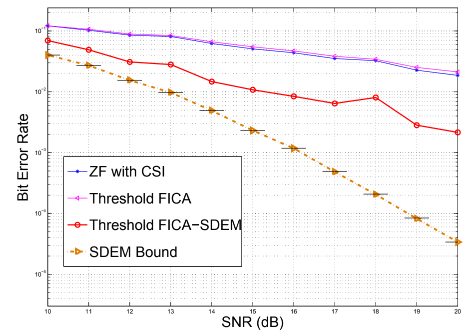

Fig. 6: Compararison of BER

6. CONCLUSION

We present a complete scheme for blind equalization of large constellation MIMO systems. We design simple approaches for the signal separation and the signal detection to reduce the computational complexity while maintain the acceptable performance. Such an efficient combination makes this feasible for a real time communication system.

7. REFERENCES

[1] P. Comon, *Independent component analysis, a new concept?*, Signal Processing, 36(3),1994.

[2] X. Zhao, M.E. Davies, *ICA with the EM Algorithm in the Low Noise Case*. ICA Research Network International Workshop, 2006.

[3] M. Novey and T. Adal, *Complex fixed-point ICA algorithm for separation of QAM sources using Gaussian mixture model*, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), Honolulu, Hawaii, April 2007.

[4] X. Zhao, M.E. Davies. *A Simple Nonlinearity for Blind Equalization in Large Constellation MIMO Systems*. Submited to ICC2008.

[5] A.P. Dempster, N.M. Laird, D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm*. J. R. Statist. Soc. B 391-38, 1977.

[6] Wu,C. *On the Convergence Properties of the EM Algorithm*. The Annals of Statistics, 11(1):95103, 1983.

[7] Adel Belochrani, J.F.cardoso, *Maximum Likelihood Source Seperation by the EM technique: Deterministic and Stochastic implementation*. In Proc. NOTLTA pages49-53, 1995.

[8] U. Fincke and M. Pohst. *Improved methods for calculating vectors of short length in a lattice, including a complexity analysis*, Mathematics of Computation, 44 (170):463-471, April,1985.

[9] E. Bingham and A. Hyvarinen, *A fast fixed-point algorithm for independent component analysis of complex valued signals*, Int. J. Neural Systems, vol. 10, pp. 18, 2000.

[10] S. Douglas, *FastICA algorithms for the blind separation of complex-valued signal mixtures*, in Proc. Asilomar Conf. Signals, Systems and Computers, October 2005, pp. 1320 - 1325.