ON THE FINITE SAMPLE BEHAVIOR OF THE CONSTANT MODULUS COST FUNCTION

Amir Leshem and Alle-Jan van der Veen

Delft University of Technology
Dept. Electrical Engineering
Mekelweg 4, 2628CD Delft, The Netherlands

ABSTRACT

We study the location of local minima of the finite sample approximation to the constant modulus cost function. This paper concentrates on source separation. The main result is a connection between the number of samples and the probability of obtaining a local minimum of the finite approximation within a given sphere around the local minimum of the CM cost function.

1. INTRODUCTION

Blind equalization and source separation is a wide field of research, initiated by the works of Sato [9] Godard [2] and Jutten and Herault [6]. One of the most widely used and analyzed among the many methods studied the constant modulus algorithm (MA) proposed by [2] and [12] for channel equalization or for cochannel interference suppression in radio communication systems [3]. The asymptotic behavior of the CM cost function is well understood, i.e. the location of the local minima of the CM cost function have been characterized, first in the noiseless case and then in the noisy case in [4], [14],[7]. We refer the reader to the excellent overview paper by Johnson et. al [5] for a detailed overview on the various algorithms using the CMA cost function, as well as many of its properties. For more on the CM array for multiple user separation see [10] and the references there in.

Although much research on the properties of the CM cost function as well as other cost functions has been done, we could not find any study of the performance of these methods with finitely many samples. One relevant result in the noiseless case is the uniqueness of the solutions to the CM problem for BPSK signals [11], which states that a sufficient condition for obtaining only the zero forcing solutions is having all 2^{d-1} many constellations with first element 1. This means that in the noiseless i.i.d. case with probability approaching 1 we will obtain the exact solution. Such a theorem can no longer hold in the noiseless case, and similar result for arbitrary CM signals still lacks a rigorous proof, although it is expected to be true.

Several algorithms which works well with finite samples in the noiseless case have appeared. One such algorithm is the ACMA [13] which analytically solves the non-linear equations involved. Another algorithm in different direction is the finite window CMA [16] which is based on the LS-CMA [1], but reuses the data.

In this paper we propose a general framework for analyzing the finite sample properties of various cost functions through the use of the central limit theorem, and Chebyshev's inequality. To simplify the mathematics involved we shall assume that the signal samples are also i.i.d. However more general treatment is possible. The ideas presented in this paper can be considered as an extension of Ljung's analysis [8] of the MSE cost function to a more general setup.

2. MAIN THEOREM

In this section we will study the relation of the local minima of a general cost function

$$J^{\infty}(\mathbf{w}) = E_{\mathbf{X}} J(\mathbf{w}; \mathbf{x}) \tag{1}$$

to the local minima of its finite sample approximation. At no point we specialize to the CM cost function, but at the next section we will try to estimate some of the constants involved.

The motivation for our study are two problems: Equalization of communication signals, and blind separation of a desired signal in multiuser environment. In order to maintain simplicity we will focus on the case of blind beamforming which is somewhat simpler to analyze. Note however that similar results can be obtained for the equalization problem.

In our case we are given a set of antennas outputs

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)$$

where $\mathbf{A}$ is the array response towards the received signals, $\mathbf{s}(t)$ is a vector of received signals assumed to be non-Gaussian, and $\mathbf{n}(t)$ is the receivers noise which is assumed to be Gaussian white with covariance $\sigma^2\mathbf{I}$.

Adaptive beam-former is basically a linear combiner of the antenna outputs, designed to separate one signal out of the mixtures, with maximal signal to interference plus noise ratio. Many adaptive methods try to separate users based on minimizing a stochastic cost function $E_{\mathbf{X}} J(\mathbf{w}, \mathbf{x})$. One of the most successful is the CMA trying to minimize the $CMA(2,2)$ cost function

$$J_{CM}(\mathbf{w}) = E_{\mathbf{X}} \left(|\mathbf{w}^H \mathbf{x}|^2 - r \right)^2.$$

where r is the dispersion constant.

We will assume that we are given a set of N realizations

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$$

of $\mathbf{x}$, and define $J^N(\mathbf{w};\mathbf{X})$ by:

$$J^N(\mathbf{w};\mathbf{X}) = \frac{1}{N}\sum_{k=1}^{N} J(\mathbf{w};\mathbf{x}_k). \qquad (2)$$

Let $\mathbf{w}_o$ be a local minimum of $J^\infty(\mathbf{w})$. We would like to bound the distance between $\mathbf{w}_o$ and the closest local minimum of $J^N(\mathbf{w};\mathbf{X})$, in terms of the statistical properties of the cost function and its derivatives. Obviously, since our definition is based on realizations of a random process, we cannot expect to obtain a deterministic result. Hence we would like that given a probability ε and a radius r to know whether the probability that a local minimum of $J^N(\mathbf{w};\mathbf{X})$ has a distance at most r from $\mathbf{w}_0$ is greater than $1-\varepsilon$.

We will assume that the Hessian of $J^\infty(\mathbf{w}_o)$ is positive definite. Especially it is non-singular. In the noiseless case this implies that the number of signals is equal to the number of sensors. This limitation is artificial and follows from the fact that in this case there are no true local minima, since adding a component in the direction of the noise subspace will not change the value of J^∞. To overcome it one should note that by the results of [7] the CM receiver are all in the signal subspace even in the noisy case. Hence a sensible first step would be to project the sensors onto the signal subspace. This exploits the available second order statistics, something desirable in the noisy case. A further step of whitening is possible but not necessary for our analysis.

Let $\mathbf{w}_N$ be the local minima of $J^N(\mathbf{w};\mathbf{X})$ closest to $\mathbf{w}_o$. First we shall prove the following theorem:

Theorem 2.1 *For every $r>0$ and any $\varepsilon>0$ there is $N=N(r,\varepsilon)$ such that $P(\|\mathbf{w}_N-\mathbf{w}_o\|<r)>1-\varepsilon$. Moreover we can choose $N(r,\varepsilon)=\frac{\|\nabla J^\infty(\mathbf{w}_o)\|^2)}{\lambda_{\min} r^2\varepsilon}$, where $\lambda_{\min}$ is the minimal eigenvalue of the Hessian of J^∞ evaluated at $\mathbf{w}_o$.*

The proof of theorem 2.1 is divided into two parts: First we study the location of local minima of the second order approximation to J^N around $\mathbf{w}_o$. Then we shaw that if r is chosen sufficiently small the existence of a local minimum of the second order approximation of J^N in a sphere of radius r ensures the existence of a local minimum of J^N in the same sphere. To simplify notation let $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_N]$. Let $J_2^N(\mathbf{w};\mathbf{X})$ be the second order Taylor approximation of J^N around $\mathbf{w}_o$:

$$\begin{aligned} J_2^N(\mathbf{w};\mathbf{X}) = {} & J^N(\mathbf{w}_0,\mathbf{X}) + \nabla J^N(\mathbf{w}_o,\mathbf{X})(\mathbf{w}-\mathbf{w}_o) \\ & + \tfrac{1}{2}(\mathbf{w}-\mathbf{w}_o)^H \mathbf{H}^N(\mathbf{w}_o,\mathbf{X})(\mathbf{w}-\mathbf{w}_o) \end{aligned} \qquad (3)$$

$J_2^N(\mathbf{w};\mathbf{X})$ has a local minimum at

$$\mathbf{w}_2^N = \mathbf{w}_o + \mathbf{H}^N(\mathbf{w}_o,\mathbf{X})^{-1}\nabla J^N(\mathbf{w}_o,\mathbf{X}) \qquad (4)$$

provided that $\mathbf{H}^N(\mathbf{w}_o,\mathbf{X})$ is positive definite. $\mathbf{H}^\infty(\mathbf{w}_o)$ is positive definite since $\mathbf{w}_o$ is a local minimum. We shall assume that it is positive definite. If this assumption is violated a higher order analysis is needed. Based on the central limit theorem we can deduce that $\mathbf{H}^N$ is positive definite provided that $N >> \lambda_{\min}^{-1}$, where $\lambda_{\min}$ is the minimal eigenvalue of $\mathbf{H}^\infty(\mathbf{w}_o)$. In this case the maximal eigenvalue of $\mathbf{H}^N(\mathbf{w}_0,\mathbf{X})^{-1}$ is (a.s.) $\frac{1}{\lambda_{\min}}(1+O(\sqrt{\frac{\log\log N}{N}})$ [15]. Hence we obtain:

$$\|\mathbf{w}_2^N-\mathbf{w}_o\|^2 < \frac{\|\nabla J^N(\mathbf{w}_o;\mathbf{X})\|^2}{\lambda_{\min}^2} \qquad (5)$$

Therefore

$$P\left(\|\mathbf{w}_2^N-\mathbf{w}_o\|>r\right) < P\left(\frac{\|\nabla J^N(\mathbf{w}_o;\mathbf{X})\|}{\lambda_{\min}}>r\right).$$

Assuming that $\mathbf{x}_k$ are i.i.d (and with some mild assumptions on ∇J^∞) we obtain:

$$\mathrm{Var}_{\mathbf{X}}\left(\|\nabla J^N(\mathbf{w}_o;\mathbf{X})\|^2\right) = \frac{1}{N}\mathrm{Var}_{\mathbf{x}}\left(\|\nabla J(\mathbf{w}_o;\mathbf{x})\|^2\right). \qquad (6)$$

Using Chebyshev's inequality and the positivity of $\|\nabla J^N(\mathbf{w}_o;\mathbf{x})\|$ we conclude that

$$P(\|\mathbf{w}_2^N-\mathbf{w}_o\|>r) < \frac{\mathrm{Var}_{\mathbf{x}}\|\nabla J(\mathbf{w}_o;\mathbf{x})\|}{\lambda_{\min} N r^2}. \qquad (7)$$

Setting N such that $\frac{\mathrm{Var}_{\mathbf{X}}\|\nabla J(\mathbf{w}_o;\mathbf{x})\|}{\lambda_{\min} N r^2}=\varepsilon$ yields the desired result for J_2^N. i.e.,

$$N = \frac{\mathrm{Var}_{\mathbf{x}}\|\nabla J(\mathbf{w}_o;\mathbf{x})\|}{\lambda_{\min}\varepsilon r^2}$$

Finally we comment on a possible strong improvement. We know that when N is large, $J^N(\mathbf{w};\mathbf{X})$ is the average of i.i.d variables. Hence we can apply the central limit theorem to obtain that

$$\frac{\sqrt{N}\left(\nabla J^N(\mathbf{w};\mathbf{X}) - E_{\mathbf{X}}\nabla J(\mathbf{w};\mathbf{X})\right)}{\sqrt{Var_{\mathbf{X}}\nabla J(\mathbf{w};\mathbf{X})}}$$

is asymptotically normally distributed. This enables to determine sharper bounds on the relation between r,ε and N. We shall demonstrate this by simulations, but a more thorough description will appear elsewhere due to lack of space.

The next step is to show that for r sufficiently small, the local minima of $J^N(\mathbf{w}_;\mathbf{X})$ are sufficiently close to these of $J_2^N(\mathbf{w}_;\mathbf{X})$. We know that

$$J^N(\mathbf{w};\mathbf{X}) = J_2^N(\mathbf{w};\mathbf{X}) + O(\|\mathbf{w}-\mathbf{w}_o\|^3).$$

Thus by definition of $J_2^N(\mathbf{w}_;\mathbf{X})$, there is an r_0 and a constant c such that for every $\mathbf{w}$ such that $\|\mathbf{w}_2^N-\mathbf{w}_o\|<r_0$, we have

$$|J_2^N(\mathbf{w}_;\mathbf{X}) - J^N(\mathbf{w}_;\mathbf{X})| < cr^3.$$

This is sufficient, by the following argument: Let ε be given. Let N be large enough such that

$$P(\|\mathbf{w}_2^N-\mathbf{w}_o\|>\frac{r}{2}) < \frac{4\mathrm{Var}_{\mathbf{x}}\|\nabla J(\mathbf{w}_o;\mathbf{x})\|^2}{\lambda_{\min} N r^2} = \varepsilon.$$

where $\mathbf{w}_2^N$ is the minimum of $J_2^N(\mathbf{w}_;\mathbf{X})$. We know that if $\|\mathbf{w}_2^N-\mathbf{w}_o\|<\frac{r}{2}$, then for any $\mathbf{w}$ such that $\|\mathbf{w}-\mathbf{w}_o\|=r$,

$$J_2^N(\mathbf{w}_;\mathbf{X}) - J_2^N(\mathbf{w}_2^N;\mathbf{X}) \geq \lambda_{min(H^N(\mathbf{w}_o)}\|\mathbf{w}-\mathbf{w}_2^N\|^2$$

and

$$\lambda_{min(H^N(\mathbf{w}_o))}\|\mathbf{w}-\mathbf{w}_2^N\|^2 \geq \lambda_{min(H^N(\mathbf{w}_o))}\left(\frac{r}{2}\right)^2.$$

Hence we obtain

$$J^N(\mathbf{w};\mathbf{X}) - J^N(\mathbf{w}_2^N;\mathbf{X}) > \frac{\lambda_{min(H^N(\mathbf{w}_o))}}{4}r^2 - 2cr^3.$$

This is positive for any r sufficiently small. Thus $J^N(\mathbf{w}_2^N;\mathbf{X})$ is smaller than $J^N(\mathbf{w};\mathbf{X})$ for any $\mathbf{w}$ such that $\|\mathbf{w}-\mathbf{w}_o\|=r$, so there must be a local minimum of $J^N(\mathbf{w};\mathbf{X})$ inside the sphere:

$$\{\mathbf{w} : \|\mathbf{w}-\mathbf{w}_o\| \leq r\}.$$

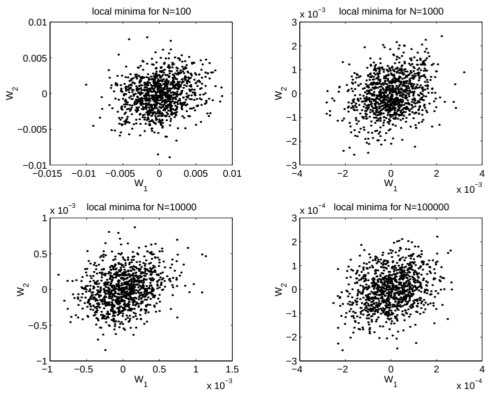

Figure 1: Distribution of local minima for various values of N. (a) $N = 100$, (b) $N = 1000$, (c) $N = 10^4$, (d) $N = 10^5$.

3. SIMULATIONS FOR THE CM COST FUNCTION

In this section we describe some simulations we have performed in order to test the behavior of the location of the N sample approximations to the CM cost function, to be denoted CMA(2,2,N).

In the first experiment we have mixed two random phase CM signals using the unitary matrix

$$\mathbf{A} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

and $\theta = 60^\circ$. We have added a random Gaussian noise with signal to noise ratio of 30dB. For logarithmically spaced number of samples between 100 and 10^5 we have initialized a minimization of the CM cost function with one of the zero forcing solutions, and converged to the nearest local minimum point using a BFGS Quasi-Newton algorithm. The justification for the initialization is that for good SNR the CM beamformer is expected to be near the ZF solution. For each value of N we have repeated the experiment 1000 times. Figure 1 (a-d) presents the deviations of the estimated CM beamformer from the CM beamformer (estimated by averaging the locations of all the experiments with 10^5 many samples), for various values of N. We can clearly see that the locations of the local minima tends to concentrate around the true solution, as expected from theorem 2.1. Next we have turned to estimate the distribution of the local minima as a function of N.

Using the asymptotic expression coming from the central limit theorem we have expected to find a linear connection between the logarithm of the radius which ensures probability $1 - \varepsilon$ of finding a local minimum to the CMA(2,2,N) cost function and the logarithm of the number of samples. To test that we have used the order statistics of the distance to the CM beamformer. Figure 2(a) presents the 10,30,50,70 and 90'th percentiles of the distance of local minima of the CMA(2,2,N) to the CM beamformer, as a function of $\log_{10} N$. We can clearly see that the lines are parallel and are linear. To verify the linearity we also present the least square fits, in figures 2(b)-(d).

In order to present the effect of the eigenvalue spread of the mixing matrix, which is expected to get into the probability expression through $\lambda_{\min}$ we have repeated the above experiment

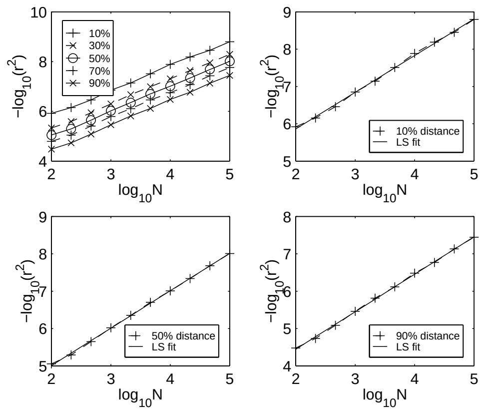

Figure 2: Order statistics of the log distance to the CM beamformer as a function of N. (a)10,30,50,70,90 percentiles as function of $\log_1 0(N)$. (b)-(d) LS line fitting to the 10, 50 and 90 percentiles.

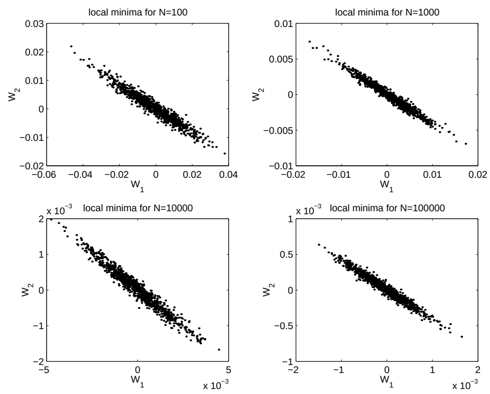

Figure 3: Distribution of local minima for various values of N. (a) $N = 100$, (b) $N = 1000$, (c) $N = 10^4$, (d) $N = 10^5$.

with a mixing matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

which has eigenvalues $-0.372, 5.372$. The eigenvalues of $\mathbf{A}^{-1}$ are $-2.6861, 0.1861$ which implies that the estimates are likely to be more spread in one direction. Figures 4(a)-(d) presents the locations of the local minima. We can clearly see the widening due to the smallest eigenvalue.

Finally in order to test the dependence between N and r for any given ε we have computed the coefficient a, bin the LS fitting of $-\log_{10}(r^2) = a \log_{10}(N) + b$, as a function of the percentile ε. Figures 5,6 describes the results for both simulations. We can see that the dependence is of the form $N = \frac{b(\varepsilon)}{r^2}$, i.e., the a coefficient is extremely close to 1 in all cases. This completely agrees with the analysis of the relation between r^2 and N presented in the previous section. The dependence on ε however can be improved using the central limit theorem.

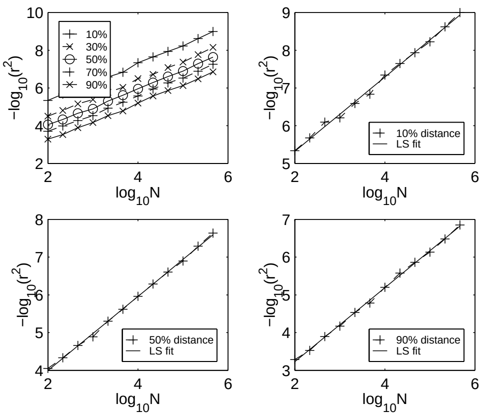

Figure 4: Order statistics of the log distance to the CM beamformer as a function of N. (a)10,30,50,70,90 percentiles as function of $\log_1 0(N)$. (b)-(d) LS line fitting to the 10, 50 and 90 percentiles.

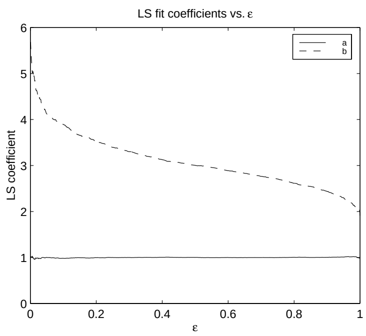

Figure 5: Regression coefficients as a function of percentile. Unitary mixing.

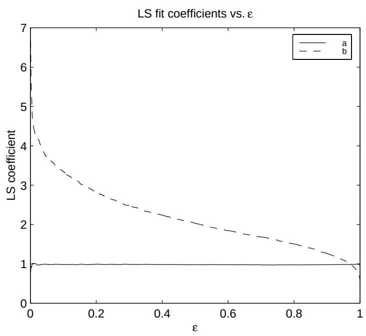

Figure 6: Regression coefficients as a function of percentile. Non-unitary mixing.

4. REFERENCES

[1] B.G. Agee. The least-squares CMA: A new technique for rapid correction of constant modulus signals. In *Proc. IEEE ICASSP*, pages 953–956, Tokyo, 1986.

[2] D.N. Godard. Self-recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Communications*, 28(11):1867–1875, November 1980.

[3] R. Gooch and J. Lundell. The CM array: An adaptive beamformer for constant modulus signals. In *Proc. IEEE ICASSP*, pages 2523–2526, Tokyo, 1986.

[4] Fijalkow I., A. Touzni, and J.R. Treichler. Fractionally spaced equalization using CMA:Robustness to channel noise and lack of disparity. *IEEE Trans. on Signal Processing*, 45(1):56–66, 1997.

[5] C.R. Johnson, P. Schniter, T.J. Enders, A.D. BEHM, D. Brown, and R.A. Casas. Blind equalization using the constant modulus criterion: A review. *Proceedings of the IEEE*, 86(10):1927–1950, 1998.

[6] C. Jutten and J. Herault. Blind separation of sources I. An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1):1–10, July 1991.

[7] D. Liu and L. Tong. An analysis of constant modulus algorithm for array signal processing. *Signal Processing*, 44(4):1523–1538, 1999.

[8] L. Ljung. *System identification: Theory for the user*. Prentice Hall information and system sciences. Prentice Hall, 1987.

[9] Y. Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *IEEE Trans. Communications*, 23:679–682, June 1975.

[10] J.J. Shynk and R.P. Gooch. The constant modulus array for cochannel signal copy and direction finding. *IEEE Trans. Signal Proc.*, 44(3):652–660, March 1996.

[11] S. Talwar, M. Viberg, and A. Paulraj. Blind estimation of multiple co-channel digital signals using an antenna array. *IEEE Signal Proc. Letters*, 1(2):29–31, February 1994.

[12] J.R. Treichler and B.G. Agee. A new approach to multipath correction of constant modulus signals. *IEEE Trans. Acoust., Speech, Signal Processing*, 31(2):459–471, April 1983.

[13] A.J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal Processing*, 44(5):1136–1155, May 1996.

[14] H.H. Zeng, L. Tong, and C.R. Johnson. Relationships between the constant modulus and Wiener receivers. *IEEE Trans. on Information Theory*, 44(4):1523–1538, 1998.

[15] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai. On the detection of the number of signals in the presence of white noise. *Journal of multivariate analysis*, 20:1–25, October 1986.

[16] X. Zhuang and A.L. Swindlehurst. Fixed window constant modulus algorithm. In *Proc. IEEE ICASSP*, volume 5, pages 2623–2626, 1999.