Unbiased adaptive equalization: the Gaussian Kernel contrast functions

Antoine Chevreuil and Christophe Vignat

Université de Marne-La-Vallée, Laboratoire Système de Communication
5, boulevard Descartes - 77454 Marne-La-Vallée Cedex 2 - France
Email {Antoine.Chevreuil,Christophe.Vignat}@univ-mlv.fr

Abstract

We address the issue of blind adaptive equalization; in this respect, we introduce a class of cost functions which basically exploit the finite alphabet property of the source (we restrict ourselves to the case of BPSK sources) . In an un-noisy context, we show that these cost functions are contrast functions in the sense that their maximizing arguments are delayed zero-forcing equalizers. When Gaussian noise is present, these Gaussian Kernel cost functions show the property that under a suitable norm constraint of the equalizer, their maximizing arguments remain zero-forcing solutions. The results extend to the multi-source case, where the above-mentioned functions allow to extract a source and remove the ISI.

I. Introduction

Many communication systems have to cope with the equalization of an unknown dispersive channel. So as to raise this difficulty, a standard solution consists in sending periodically a training sequence in order to identify the channel. However, the decrease of capacity this approach leads to is unaffordable in certain rapidly varying channel cases. In this respect, many contributions deal with the problem of retrieving the emitted symbols, resorting exclusively to the received signal. These approaches are called "blind".

A very popular method consists in adapting a linear equalizer according to a criterion so-called contrast function, the optimization of which leads to a "good channel inverse". Since the pioneering work by Sato [?], then Godard [?], Treichler [?], much interest has been focused on the Constant Modulus Algorithm that constraints the modulus of the reconstructed data to be constant. This algorithm has been widely studied: see [?] for an exhaustive review. In diversity contexts, the CMA allows to estimate a zero forcing solution for finite length equalizers, this result holding in the noise-free case. The nice theoretical properties of CMA however do not transpose to the noisy case. It is heuristically admitted that CMA performs well in high signal-to-noise (SNR) ratio environments, but the equalizers are biased, in the sense that CMA does not provide an exact zero-forcing solution: indeed, CMA achieves a nice trade-off between residual ISI and noise enhancement. More precisely, it is shown in [?] that CMA achieves approximatly the Wiener solution. As far as moderate to low SNR are concerned, the theoretical behavior of CMA remains, to our knowledge, unclear.

In general, the presence of Gaussian noise breaks the shape of the contrast function, this drawback involving bias on the solution. Recently, Hyvarinen introduced a new class of contrast functions relying on Gaussian mappings [?]. He noticed that the shape of these functions is preserved in the presence of Gaussian noise. However, this invariance property is not fully exploited, since the noise variance is assumed to be known in order to estimate the equalizer.

Due to their nice invariant property, we consider in this paper cost functions relying on such Gaussian mappings. These functions are chosen such that their maximizing arguments correspond to filters the output of which falls near desired values (actually the elements of the source alphabet). Actually, this is not a shot in the dark, since these functions can be cast into a strict statistical framework relying on a Kullback-Leibler philosophy, as remarked by Alvarez *et al.* [?]. These authors studied the stochastic behaviour of these functions. However, the algebraic counterpart is still unaddressed: we propose to derive new results in this direction.

Models, definitions and purposes are given in Section **??**. Under the noise free hypothesis (Section **??**), the GK functions are shown to be maximized at ZF filters, up to a constant. The inexistence of spurious local maxima is uneasy to prove; however, elementary arguments show how possible such points can be cancelled out.

Section **??** is devoted to the noisy case: under a suitable norm constraint on the equalizer (this notion is discussed), the maximization of this cost function is also shown to yield ZF solutions (within a constant). In Section **??**, numerical illustrations are provided. In particular, the behaviour of the algorithm is analyzed; comparisons with the standard CMA are proposed.

II. Problem Statement

Let s_n be an i.i.d. binary sequence. Suppose that this input sequence is conveyed through a channel characterized by a convolutive effect and an additive white Gaussian noise. We thus consider the following model for the received data

$$\begin{aligned} y_n &= \sum_{k=0}^{L_h} h_k s_{n-k} + w_n \\ &\triangleq [H(z)]s_n + w_n \end{aligned} \tag{1}$$

where $H(z)$ is a multi-variate $q \times 1$-dimensional FIR transfer function and w_n is a temporally and spatially white Gaussian noise, with variance σ_w^2. In order to recover the emitted data, the received data are fed into a multi-variate $1 \times q$ FIR equalizer denoted by $G(z) = \sum_{k=0}^{L_g} g_k z^{-k}$. The output z_n of the equalizer is written

$$\begin{aligned} z_n &= [G(z)]\, y_n \\ &= [G(z)]\,[H(z)]\, s_n + [G(z)]\, w_n \\ &= [F(z)]\, s_n + [G(z)]\, w_n \end{aligned} \tag{2}$$

The composite filter $F(z) = G(z)\, H(z)$ is a FIR scalar valued filter. We set $F(z) = \sum_{k=0}^{L_f} f_k z^{-k}$ with $L_f = L_h + L_g$ and we set :

$$\begin{aligned} \mathbf{f} &= \left(f_0, \cdots, f_{L_f}\right)^T \\ \mathbf{g} &= \left(g_0, \cdots, g_{L_g}\right)^T \\ \mathbf{s}_n &= \left(s_n, s_{n-1}, \cdots, s_{n-L_f}\right)^T \\ \mathbf{w}_n &= \left(w_n^T, \cdots, w_{n-L_g}^T\right)^T \\ \mathbf{y}_n &= \left(y_n^T, \cdots, y_{n-L_g}^T\right)^T \end{aligned}$$

In the following, we assume that there exist so-called zero forcing (ZF) solutions, i.e. there exist some $G(z)$'s such that

$$F(z) = z^{-K} \quad 0 \leq K \leq L_f - 1$$

This assumption is equivalent to the hypothesis that the entries of $H(z)$ have no common zeros. We explain in Section **??** how our approach extends to the case where this hypothesis is not valid, especially in single sensor and small excess bandwidth.

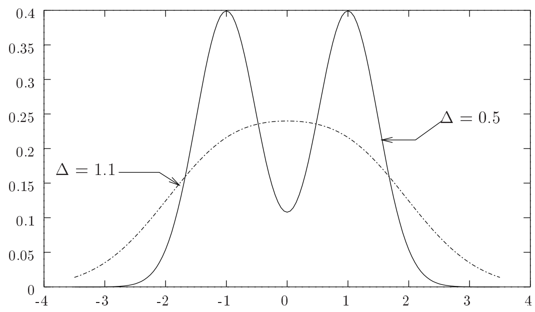

Figure 1: *Two possible shapes of function $\Phi_\sigma(x)$*

III. A Gaussian shaped cost function

Let us consider the following mapping

$$\Phi_\sigma(x) = \frac{1}{2\sigma\sqrt{2\pi}} \left(e^{-\frac{(x-1)^2}{2\sigma^2}} + e^{-\frac{(x+1)^2}{2\sigma^2}} \right) \tag{3}$$

the graph of which is depicted in Fig. **??** for two values of σ.

We consider the following cost function, that we call Gaussian kernel (GK) cost function, defined as

$$C(\mathbf{g}) = E\Phi_\sigma(z_n) \tag{4}$$

In the following, $||.||$ denotes the $\mathcal{L}^2$-norm. We have:

Theorem 1 *Under the previous assumptions, we have*

$$C(\mathbf{g}) = \frac{1}{2^N} \sum_{k=0}^{2^N-1} \Phi_{\Delta(\mathbf{g})}(\mathcal{M}_p \mathbf{f}) \tag{5}$$

where

$$\Delta(\mathbf{g}) = \sqrt{\sigma^2 + \sigma_w^2\, ||\mathbf{g}||^2}. \tag{6}$$

From (**??**), we have

$$C(\mathbf{g}) = E_{s,w} \Phi_\sigma \left(\mathbf{f}^T \mathbf{s}_n + \mathbf{g}^T \mathbf{w}_n\right)$$

where $E_{s,w}[.]$ denotes the mathematical expectation according to the joint probability density of the random variables $\left\{\mathbf{s}_n^T\ ,\ \mathbf{w}_n^T\right\}$. By Bayes' theorem , we have:

$$C(\mathbf{g}) = E_w E_{s|w} \Phi_\sigma \left(\mathbf{f}^T \mathbf{s}_n + \mathbf{g}^T \mathbf{w}_n\right)$$

The term $\mathbf{g}^T \mathbf{w}_n$ represents the noisy part after equalization. This process it Gaussian with zero mean and variance $||\mathbf{g}||^2\, \sigma_w^2$.

Denote by $\mathcal{M}$ the $2^{L_f} \times L_f$ dimensional matrix whose rows are all possible vectors with entries ± 1. The noise free part of z_n, namely, $\mathbf{f}^T \mathbf{s}_n$ is a discrete valued random variable, the values of which are $v_p = \mathcal{M}_p \mathbf{f}$ where $\mathcal{M}_p$

is the p-th row of matrix $\mathcal{M}$. The v_p's are equiprobable with probability $\frac{1}{2^N}$. It yields

$$E_{s|w}\Phi_\sigma\left(\mathbf{f}^T\mathbf{s}_n+\mathbf{g}^T\mathbf{w}_n\right)=\frac{1}{2^N}\sum_{p=0}^{2^N-1}\Phi_\sigma\left(\mathcal{M}_p\mathbf{f}+\mathbf{g}^T\mathbf{w}_n\right)$$

For a given p, the process $x=\mathcal{M}_p\mathbf{f}+\mathbf{g}^T\mathbf{w}_n$ is white Gaussian with mean $\mathcal{M}_p\mathbf{f}$ and variance $\|\mathbf{g}\|^2\sigma_w^2$ so that

$$C(\mathbf{g})=E_{\mathcal{N}\left(\mathcal{M}_p\mathbf{f},\|\mathbf{g}\|^2\sigma_w^2\right)}\Phi_\sigma(x)$$

As Φ_σ is the sum of the two Gaussian functions $\mathcal{N}\left(-1,\sigma^2\right)$ and $\mathcal{N}\left(1,\sigma^2\right)$, a straightforward result (see [?] for example) can be applied, namely

Lemma 1 *Let X be a Gaussian random variable $\mathcal{N}\left(m_x,\sigma_x^2\right)$. We have*

$$E[\frac{1}{\sqrt{2\pi}\sigma_B}e^{-\frac{(m-X)^2}{2\sigma^2}}]=\frac{1}{\sqrt{2\pi\left(\sigma_x^2+\sigma^2\right)}}e^{-\frac{(m-m_x)^2}{2\left(\sigma_x^2+\sigma^2\right)}}$$

In our particular case, take $\sigma_x^2=\|\mathbf{g}\|^2\sigma_w^2$, $m_x=\mathcal{M}_p\mathbf{f}$ and $m=-1$so that result (**??**) holds.

An important consequence of Theorem **??** is that the cost function $C(\mathbf{g})$ is written in a closed form as a function of the noise variance. Furthermore, the Gaussian character of $C(\mathbf{g})$ remains unchanged in the presence of noise. Within the next Sections, we prove that $C(\mathbf{g})$ allows to recover a ZF equalizer in the noise free case; then this property is extended to the noisy case.

IV. The noise free case

A. Theoretical analysis.

Along this section, we assume that $\sigma_w^2=0$. Hence $\Delta(\mathbf{g})=\Delta=\sigma$ does not depend on filter $\mathbf{g}$.

Theorem 2 *If $\Delta=\sigma<1$, $C(\mathbf{g})$ is a contrast function in the sense that it takes its maximum values if and only if $\mathbf{g}$ is proportional to a ZF equalizer.*

On the contrary, if $\Delta=\sigma\geq 1$, $C(\mathbf{g})$ is maximum if and only if $G(z)$ is such that $G(z)H(z)=0$ for all z.

Let us introduce the following auxiliary function:

$$J_\Delta(\mathbf{v})=\frac{1}{2^N}\sum_{p=0}^{2^N-1}\Phi_\Delta(v_p)$$

and denote by $\mathcal{V}$ the following set:

$$\mathcal{V}=\{\mathbf{v}\mid\exists\mathbf{f}\ \ v=\mathcal{M}\mathbf{f}\text{ and }F(z)=G(z)H(z)\}$$

In function $J_\Delta(\mathbf{v})$, the variables v_k are linearly independent whereas in $C(\mathbf{g})$ they are not since $\mathbf{v}$ is bound to belong to the set $\mathcal{V}$. That is, $J_\Delta(\mathbf{v})$ and $C(\mathbf{g})$ are related by

$$C(\mathbf{g})=J_\Delta(\mathbf{v})\quad\text{with }\mathbf{v}\in\mathcal{V}.$$

In the case $\Delta<1$, the numerous solutions of this problem are exhibited, among which $2N$ and only $2N$ ones can be shown to belong to set $\mathcal{V}$. The corresponding filters $\mathbf{f}$ are shown to be associated with all the possible ZF equalizers (up to a constant). The case $\Delta\geq 0$ is addressed the same way.

Lemma 2 *The maximizing arguments of function Φ_Δ are given by*

- *If $\Delta<1$, $\Phi_\Delta(x)$ admits the two maximum points $x=\pm\eta_\Delta$, where $0<\eta_\Delta<1$.*
- *If $\Delta\geq 1$, $\Phi_\Delta(x)$ admits the unique maximum $x=0$.*

The proof of Lemma **??** is given in Appendix **??**. Suppose $\Delta<1$. Consider the maximization of $J_\Delta(\mathbf{v})$ where the v_p's s are independent variables. $J_\Delta(\mathbf{v})$ is maximum if and only if each component $\phi_\Delta(v_p)$ is maximum. According to Lemma **??**, this occurs, as $\Delta=\sigma<1$, if and only if each v_p equals $\pm\eta_\Delta$. There are thus exactly 2^{2^N} such maximizing vectors $\mathbf{v}$. Going back to the original problem of maximizing $C(\mathbf{g})$, notice that a sufficient condition for achieving this goal is to exhibit a vector $\mathbf{v}$ that maximizes $J_\Delta(\mathbf{v})$ and belongs to $\mathcal{V}$.

In the following, we show

i if $\mathbf{v}$ maximizes $J_\Delta(\mathbf{v})$ and fulfills the condition $\mathbf{v}\in\mathcal{V}$, then necessarily, $\mathbf{v}$ is associated with an equalizer $\mathbf{g}$ that writes as $\mathbf{g}=\eta_\Delta\mathbf{g}_{ZF}$ where $\mathbf{g}_{ZF}$ is any ZF solution and η_Δ is the scalar defined in Lemma **??**;

ii such equalizers $\mathbf{g}=\eta_\Delta\mathbf{g}_{ZF}$ are conversely shown to maximize $J_\Delta(\mathbf{v})$ hence $C(\mathbf{g})$.

i Suppose that a solution $\mathbf{g}$ exists which maximizes both $J_\Delta(\mathbf{v})$ and such that $\mathbf{v}\in\mathcal{V}$. As $v\in\mathcal{V}$, we denote by $G(z)$ an associated equalizer and by $F(z)=G(z)H(z)$ the composite filter. Necessarily, as $\mathbf{v}$ maximizes $J_\Delta(\mathbf{v})$, all the v_k's equal $\pm\eta_\Delta$. Besides, recall that the v_k's represent the states of the output of the equalizer. This means that the output z_n of the equalizer is a discrete time series taking its values in the set $\{\pm\eta_\Delta\}$. Both signals s_n and z_n are thus distributed according to the same binomial law. As moreover z_n is a FIR filtered version of s_n, a classical theorem (see [?]) imposes that $F(z)$ is a trivial filter, i.e. $F(z)=\pm\eta_\Delta z^{-K}$ for some some delay K[1]. Thus, if $\mathbf{v}$ maximizes $J_\Delta(\mathbf{v})$ and belongs to $\mathcal{V}$, then it corresponds to a ZF equalizer up to the constant η_Δ.

[1]A direct entropy based argument also ensures the result.

ii Consider conversely $\mathbf{g} = \eta_\Delta \mathbf{g}_{ZF}$. Then $\mathbf{f} = \eta_\Delta [0, \cdots, 0, \pm 1, 0, \cdots 0]^T$ which means that $v_p = \pm \eta_\Delta$ for all p. Hence such a $\mathbf{v}$ obviously maximizes the unconstrained function $J_\Delta(\mathbf{v})$. Hence $C(g)$ and $J_\Delta(\mathbf{v})$ share the same maximum value, this latter being reached at the points $\mathbf{g} = \eta_\Delta \mathbf{g}_{ZF}$.
Suppose $\Delta < 1$. The maximization of $J_\Delta(\mathbf{v})$ is achieved if and only if $\mathbf{v}$ is the null vector. Consider now an equalizer $G(z)$ such that $G(z)H(z) = 0$. It corresponds to $\mathbf{v} = \mathbf{0}$, hence $C(\mathbf{g})$ is maximized if and only if $G(z)H(z) = 0$.

B. Implementation

We consider, as a consequence of Theorem **??**, a parameter $\sigma < 1$. In a practical context, the contrast function $C(\mathbf{g})$ defined in Eq. (**??**) has to be estimated from the data.

A block version of the gradient method is proposed. The true cost function $C(\mathbf{g})$ is estimated by using the estimate

$$\hat{C}_T(\mathbf{g}) = \frac{1}{T} \sum_{n=1}^{T} \Phi_\sigma(z_n) \tag{7}$$

where T represents the number of available data. The maximizing argument of this function is found using a gradient method. The algorithm is

- Take $\mathbf{g}^{(0)}$ an initial guess.
- At step number n, compute the values $z_n = \left[G^{(n)}(z)\right] y_n$
- compute an estimate of the gradient of $C(g)$ as $\hat{\nabla}^{(n)} = \frac{1}{T} \sum_{n=1}^{T} \Phi'_\sigma(z_n) \mathbf{y}_n$
- compute $\mathbf{g}^{(n+1)} = \mathbf{g}^{(n)} + \mu \hat{\nabla}^{(n)}$
- repeat until convergence

As compared to a stochastic gradient approach, this block version is computationally more demanding but shows better performance.

C. On the existence of local maxima

We did not manage to prove the (in)existence of local maxima. However, it is possible to customize the gradient method so as to avoid the drawbacks of such points. Indeed, the points maximizing $C(g)$ are ZF (up to a constant) whatever the parameter $0 < \sigma < 1$. On the contrary, if a local maximum point exists, it likely depends on σ. In order to avoid the convergence of the gradient method to a local maximum, an idea consists in changing σ as the index n runs. For instance, one could choose σ_n as a periodic series.

D. Discussion: on the existence of ZF solutions

So far, we have supposed that there exists a FIR ZF equalizer. Diversity models as in (**??**) can be met in multi-sensor or/and oversampling contexts. In general, the algebraic condition of irreducibility of the components of $H(z)$, which ensures the existence of an exact FIR ZF equalizer, is not fulfilled. For example, the bandwidth limitation inherent to all communication systems implies that $H(z)$, obtained after oversampling, admits components with close roots ([**?**]). Notice that Theorem **??** holds whatever the degree $L_\mathbf{g}$ of the equalizer. Thus we can choose $L_\mathbf{g}$ arbitrarily large, so as to invert $H(z)$ with arbitrary accuracy. For finite $L_\mathbf{g}$, the maximization of $C(\mathbf{g})$ leads to an approximation of an exact ZF, the approximation being all the more accurate as $L_\mathbf{g}$ is large.

For any fixed value of $\sigma < 1$, the maximization of the cost function $C(\mathbf{g})$ allows to extract a ZF equalizer. At this point, however, the extension of this result to the noisy case is not straightforward:

- In the noisy case, $\Delta = \Delta(\mathbf{g})$ depends on the parameter $\mathbf{g}$ (see Theorem **??**): do the maximum argument of $C(\mathbf{g})$ still coincide with ZF equalizers?
- In the noise free case, maximization of the contrast function $C(g)$ leads to a ZF solution (up to a constant); nevertheless, the norm of $\mathbf{g}$ is not bound to be minimal: does this drawback disappear in the noisy case?

V. Analysis in the Noisy Case

Denote by H_1 the following hypothesis:
H_1: the ZF equalizer $\mathbf{g}_{\min}$ of minimum norm is such that

$$\sigma_w \|\mathbf{g}_{\min}\| < 1$$

The following theorem gives the main result of the paper:

Theorem 3 *Under H_1, there exist a $\rho \geq \|\mathbf{g}_{\min}\|$ such that the maximizing argument of $C(\mathbf{g})$ on the sphere $\|g\| = \rho$ is precisely a ZF filter (within a scale) of norm ρ.*

The proof is simple. Indeed, under H_1, there exists σ^2 and $\rho \geq \|\mathbf{g}_{\min}\|$ such that $\sigma^2 + \sigma_w^2 \rho^2 < 1$. If the maximization of $C(\mathbf{g})$ when $\|g\|$ is kept constant is achieved, the maximizing argument is a ZF (within a scale) of norm ρ. Conversely, such a ZF exists since one can find ZF of arbitrary norm bigger than $\mathbf{g}_{\min}$.

A question follows: what is the significance of H_1? This latter means that the best ZF equalizers, i.e. the minimum norm ones, have output $z_n = s_{n-K} + [G(z)] w_n$

with a signal-to-noise ratio better than $0dB$. Hence, H_1 is not that restrictive, since the case $\sigma_w \|\mathbf{g}_{\min}\| \geq 1$ is likely to be hopeless as far a reliable transmission is needed.

Under this condition, Theorem **??** ensures that the maximization of $C(\mathbf{g})$ allows to estimate a ZF equalizer of arbitrary norm (greater than $\|\mathbf{g}_{\min}\|$. This is an original contribution at least for higher-order based methods.

VI. Extension to multi-source case

Due to length restriction, we cannot develop this part. A simple adaptation of Theorem **??** shows that, in the noise-free case, the maximizing argument of the GK cost function is, within a scale ambiguity, a ZF for one of the sources, provided the rank of the polynomial matrix $H(z)$ equals the number of sources for all z including ∞. However, this can, in no way, provide a method for extracting a desired source. Means of extracting source no k is currently under investiguation: one can think of associating a signature to a given user in order to help the separation of the sources (CDMA,...).

VII. Numerical Analysis

We consider a channel coming from a shping filter of oll-off 0.4 and from three multipaths with delays 0 (direct path), $1.9T_s$ and $3.2T_s$ where T_s is the symbol period. The analog signal is sampled at $T_s/2$ and the filter $H(z)$ is found to gather the two polyphase components of the channel sampled at $T_s/2$. The filter $G(z)$ is a degree 15 filter. For each experiment, we estimate $G(z)$ maximizing the cost function. The norm constraint is $\|\mathbf{g}\| = 1$. Moreover, $\mu_{GK} = 0.0001$. We then generate a length 10^6 sequence of symbols then calculate the number of errors at the output of the equalizers. The number of experiments is 100. The noise level varies from 30 to $10dB$. The number of data used for estimating the cost function is $T = 400$. The estimated BER is given in Figure **??**. It shows the good behavior of the method. We did not compare the BER obtained with the GK cost function to the BER obtained with CMA. The reason is that CMA does not provide good results in the BPSK case - its minimum arguments in the noise-free case do not coincide with ZF filters.

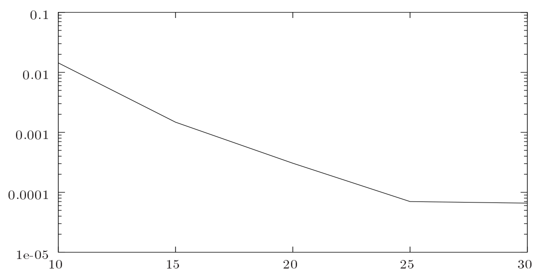

Figure 2: *BER as a function of SNR.*

References

[1] A. Benveniste and M. Goursat. Blind equalizers. *IEEE Trans. on Communications*, 32(8):871–883, August 1984.

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

[3] D.N. Godard. Self recovering equalization and carrier tracking in two dimensional data communications systems. *IEEE Tr. on Communications*, 28(11):1867–1875, November 1980.

[4] A. Hyvarinen. Noisy independant component analysis by gaussian moments. In *Proceedings of I.C.A.'99*, pages 473–478, 1999.

[5] C.R. Johnson Jr., P. Schniter, T.J. Endres, J.D. Behm, D.R. Brown, and R.A. Casas. Blind equalization using the constant modulus criterion: A review. *Proceedings of the IEEE special issue on Blind System Identification and Estimation*, October 1998.

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

[7] J.R. Treichler and B. G. Agee. a new approach to multipath correction of constant modulus signals. *IEEE Tran. On Acoustics, Speech and Signal Processing*, 31(2):459–472, April 1983.

[8] A.J. van der Veen. Resolution limits of blind multi-user channels identification schemes: the band-limited case. In *Proc. of ICASSP*, pages 2722–2725, 1996.