ON THE EQUIVALENCE BETWEEN THE SUPER-EXPONENTIAL ALGORITHM AND A GRADIENT SEARCH METHOD

Mamadou Mboup

UFR de Math-Info,
Université René Descartes-Paris V,
45, rue des Saints-Pères
75270 Paris cedex 06 France,
e-mail: mboup@math-info.univ-paris5.fr

Phillip A. Regalia

Département Signal et Image
Inst. National des Télécom.
9 rue Charles Fourier
91011 Evry cedex, France
e-mail: regalia@sim.int-evry.fr

ABSTRACT

This paper reviews the Super-exponential algorithm proposed by Shalvi and Weinstein for blind channel equalization. We show that the algorithm coincides with a gradient search of a maximum of a cost function, which belongs to a family of functions very relevant in blind channel equalization. This family traces back to Donoho's work on minimum entropy deconvolution, and also underlies the Godard (or Constant Modulus) and the Shalvi-Weinstein algorithms. Using this gradient search interpretation, we give a simple proof of convergence for the Super-exponential algorithm. Finally, we show that the gradient step-size choice giving rise to the super-exponential algorithm is optimal.

1. INTRODUCTION

Most of the recent development of second order statistics based methods for blind channel identification/equalization are supported by the Bezout identity [1]. This identity requires a multiple output channel setting, corresponding to an array of sensors at the receiver or induced by fractional sampling of the receiver's output [2]. Among the more popular of these methods are subspace methods [1] and least-squares methods [3]. All these methods can claim perfect identification in academic situations: noise-free case and *length and zero conditions* [1]. These conditions include the exact-order case, *i.e.*, the length of the identifier exactly matches that of the true channel. Now a recent study [4] shows that the performance of these methods may degrade dramatically if the equalizer length is either bigger (*overmodeled*) or smaller (*undermodeled*) than the length of the significant part of the true channel. In these more realistic conditions, the direct equalization methods based on the minimization or maximization of a higher order statistics cost function seem to be more robust. The family of functions

$$f_{2p}(\boldsymbol{x}) = \left\{ \frac{\|\boldsymbol{x}\|_{2p}}{\|\boldsymbol{x}\|_2} \right\}^{2p}, \quad p = 2, \ldots, \infty, \tag{1}$$

which traces back to Donoho's work [5], provides the most popular cost functions for blind channel equalization. For example, the popular Godard algorithm [6] as well as the Shalvi-Weinstein algorithm in [7] amount to seeking a maximum of $f_4(\boldsymbol{x})$. This then shows that these two popular algorithms belong to a same family although they use different assumptions on the source signal statistics [7], [8].

In this paper, we show that the Super-exponential algorithm proposed by Shalvi and Weinstein [9] also belongs to this same family of algorithms. Indeed, this already appeared in our previous work [10] where we have shown that the convergent points of this algorithm correspond to the maxima of $f_{2p}(\cdot)$ over the set of attainable combined channel-equalizer impulse responses.

One important remaining question is whether, given an initial condition, there exists any dependency between the trajectory of the super-exponential algorithm and that of the Godard or the Shalvi-Weinstein algorithm. That the super-exponential algorithm seeks a maximum of $f_{2p}(\cdot)$ does not allow one to deduce the domain of attraction of each maximum neither does it tell us how these maxima are approached.

Section 4 is devoted to these questions. More explicitly, Property 1 establishes the equivalence between the Super-exponential algorithm and a gradient search method, within the set of attainable combined channel-equalizer impulse responses. With this gradient search interpretation, the analysis of convergence reduces to an analytical description of the error surface. Indeed, this study becomes more classical although it may be a difficult task to obtain an analytical parametrization of this error surface [11]. Incidentally, we give a simple proof of the convergence of the Super-exponential algorithm. Finally, we show in section 5 that the gradient step-size choice giving rise to the super-exponential algorithm is optimal. Section 2 fixes the notations and presents the problem statement. Section 3 reviews the Super-exponential algorithm.

2. PROBLEM STATEMENT

The problem under investigation is illustrated in figure 1 where we assume that the channel impulse response is finite[1] in duration and that the noise is absent. We also assume the real case to simplify further the expressions.

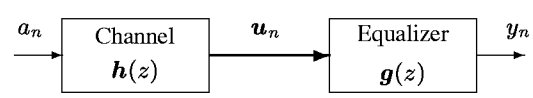

Figure 1: Blind channel equalization scheme

[1]For the sake of simplicity, we assume that the channel impulse response is finite. However, all the subsequent developments remain valid in the infinite impulse response case provided the channel-equalizer combined response be in ℓ_2.

An unknown sequence of independent and identically distributed signal, $\{a_k\}$, is sent through a single-input-N-output channel

$$\boldsymbol{h}(z) = \sum_{k=0}^{M} \boldsymbol{h}_k z^k.$$

The channel is assumed to be stable but also unknown and possibly nonminimum phase. The received signal, observed at time n after demodulation and sampling, is the N element vector $\boldsymbol{u}_n = \sum_k \boldsymbol{h}_k a_{n-k}$. This signal is next applied to the input of an N-input-single output equalizer of the form $\boldsymbol{g}(z) = \sum_{k=0}^{L} \boldsymbol{g}_k z^k$ to reduce the ISI. The output of the equalizer, y_n, then reads as $y_n = \sum_{k=0}^{L} \boldsymbol{g}_k \boldsymbol{u}_{n-k}$, which can also be expressed in terms of the source signal a_n

$$y_n = \sum_{k=0}^{L} x_k a_{n-k},$$

where $\{x_k\}$ denotes the channel-equalizer combined response. This combined response is obtained as the convolution $\boldsymbol{x} = \boldsymbol{g} \star \boldsymbol{h}$ which translate in matrix form as

$$\underbrace{\begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_L \\ x_{L+1} \\ \vdots \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} \boldsymbol{h}_0^T & \boldsymbol{0}^T & \cdots & \boldsymbol{0}^T \\ \boldsymbol{h}_1^T & \boldsymbol{h}_0^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}^T \\ \boldsymbol{h}_L^T & \cdots & \boldsymbol{h}_1^T & \boldsymbol{h}_0^T \\ \boldsymbol{h}_{L+1}^T & \boldsymbol{h}_L^T & \cdots & \boldsymbol{h}_1^T \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix}}_{\mathcal{H}} \begin{bmatrix} \boldsymbol{g}_0^T \\ \boldsymbol{g}_1^T \\ \vdots \\ \boldsymbol{g}_L^T \end{bmatrix} \tag{2}$$

involving the convolution matrix $\mathcal{H}$ associated with $\boldsymbol{h}(z)$.

3. THE SUPER EXPONENTIAL ALGORITHM

An ideal setting of the equalizer is one which brings the combined response into a pinning vector form $\boldsymbol{x} = [\cdots 0 \underbrace{1}_{n} 0 \cdots]^T \triangleq \boldsymbol{e}_n$, corresponding to an n samples pure delay, where n is any integer. Recall that for a given vector $\boldsymbol{s}$, its mth Hadamard exponent, denoted as $\boldsymbol{s}^{\odot m}$, is defined componentwise by

$$\left(\boldsymbol{s}^{\odot m}\right)_k = s_k^m.$$

Now, observe that if the dominant term of $\boldsymbol{x}$ is unique and in position n, then the ideal combined response $\boldsymbol{e}_n$ may be approached by the mth Hadamard exponent $(\boldsymbol{x}/x_n)^{\odot m}$ as m tends to infinity. With this observation in mind, it is clear that if $q > 1$ is any integer, then the iterative procedure

$$\boldsymbol{\nu} = \boldsymbol{x}_{(k)}^{\odot q} \tag{3a}$$

$$\boldsymbol{x}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|}, \tag{3b}$$

where the subscript (k) denotes the iteration number, converges asymptotically in k to the ideal response $\boldsymbol{e}_n$, when initialized by a unit-norm response $\boldsymbol{x}_{(0)}$ having a unique dominant term x_n. One may check that the subspace angle between $\boldsymbol{x}_{(k)}$ and $\boldsymbol{x}_{(k+1)}$ is insensitive to the choice of the norm in (3b); To simplify certain developments to follow, we may thus assume ℓ_2 normalization with no loss of generality. These heuristic considerations led to the so-called super-exponential algorithm proposed by Shalvi and Weinstein in the complex case [9].

Of course, since the combined response depends on the unavailable channel's impulse response $\boldsymbol{h}$, the iterative procedure in (3) is, in the form thus presented, only of academic interest. Nonetheless, one can find in [9] how to implement the algorithm, in terms of the adjustable equalizer's parameters and some statistical cumulants of the observed data.

As the algorithm in (3) is design to converge to the ideal response $\boldsymbol{e}_n$, for some n, it tacitly assumes that this ideal response is attainable. Otherwise, the algorithm would have no value. Now, it is clear from (2) that the combined response $\boldsymbol{x}$ is restricted to the range space of the convolution matrix $\mathcal{H}$. In the sequel, we will denoted this range space by $\mathcal{S}_A$: the vector subspace of ℓ_2, of attainable combined responses. The orthogonal projection operator onto $\mathcal{S}_A$ is given by

$$\mathcal{P}_A = \mathcal{H}(\mathcal{H}^T\mathcal{H})^{\#}\mathcal{H}^T, \tag{4}$$

where the symbol $\#$ denotes the pseudo inversion operator.

The case $\mathcal{P}_A = \boldsymbol{I}$ means that it is always possible to find an equalizer setting yielding any prescribed combined response. This case will be termed the *sufficient order* setting. Such a configuration is rarely met in practice, even in the fractionally-spaced case ($N \geq 2$), because of the real channel structure, which typically exhibits long tails of small leading and trailing impulse response terms (see *e.g.* [4]). Therefore, the situation where $\mathcal{P}_A \neq \boldsymbol{I}$ is more realistic. In this situation, the equalizer will be termed *undermodeled.* The undermodeled case will, in general, preclude the possibility of a given ideal combined response $\boldsymbol{e}_n$ being attainable.

To be consistent with this latter remark, the algorithm in (3) must be modified as

$$\boldsymbol{\nu} = \mathcal{P}_A\left(\boldsymbol{x}_{(k)}^{\odot q}\right) \tag{5a}$$

$$\boldsymbol{x}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|}, \tag{5b}$$

initialized by $\boldsymbol{x}_{(0)} \in \mathcal{S}_A$, such that $\|\boldsymbol{x}_{(0)}\| = 1$. This latter version of the algorithm, which also appears in [9], has been obtained in a more constructive way in [10], as an iterative procedure for seeking a local maximum of the family of objective functions $f_{2p}(\boldsymbol{x})$.

In the next section, we will give a more direct interpretation of the super-exponential algorithm in (5), by showing that it may be written in a form of a gradient search algorithm.

4. A GRADIENT SEARCH METHOD

We recall that the directional derivative of a function $f : \mathbb{R}^m \to \mathbb{R}$ (locally Lipschitzian) at $\boldsymbol{x}$, in the direction $\boldsymbol{d} \in \mathbb{R}^m$ is

$$f'(\boldsymbol{x}, \boldsymbol{d}) = \lim_{t \downarrow 0} \frac{f(\boldsymbol{x} + t\boldsymbol{d}) - f(\boldsymbol{x})}{t}$$

If f is differentiable at $\boldsymbol{x}$, then this directional derivative becomes $f'(\boldsymbol{x}, \boldsymbol{d}) = \langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle$, where $\nabla f(\boldsymbol{x})$ denotes the gradient of f at the point $\boldsymbol{x}$. We proceed to introduce the concept of a projected gradient which, naturally, extends that of a directional derivative in a multi-directional setting. Let $\mathcal{S}$ be a vector subspace of $\mathbb{R}^m$ and let $\mathcal{P}$ be the orthogonal projection operator onto $\mathcal{S}$. Then, the

gradient of the function f in $\mathcal{S}$, at a point $\boldsymbol{x} \in \mathbb{R}^m$ is defined as

$$\nabla^{\mathcal{S}} f(\boldsymbol{x}) = \begin{bmatrix} f'(\boldsymbol{x}, \boldsymbol{p}_1) \\ \vdots \\ f'(\boldsymbol{x}, \boldsymbol{p}_m) \end{bmatrix},$$

where $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_m$ are the columns of $\mathcal{P}$. If f is differentiable at $\boldsymbol{x}$, then

$$\nabla^{\mathcal{S}} f(\boldsymbol{x}) = \mathcal{P} \nabla f(\boldsymbol{x}),$$

and the gradient $\nabla^{\mathcal{S}} f(\boldsymbol{x})$ is simply the projection of the gradient of f onto $\mathcal{S}$.

Now, we may show the following equivalence:

Property 1 *The super-exponential algorithm in (5) with ℓ_2 normalization is equivalent to the gradient search algorithm*

$$\boldsymbol{\nu} = \boldsymbol{x}_{(k)} + \frac{1}{2p f_{2p}(\boldsymbol{x}_{(k)})} \mathcal{P}_A \nabla f_{2p}(\boldsymbol{x}_{(k)}), \tag{6a}$$

$$\boldsymbol{x}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|_2}. \tag{6b}$$

Remark 1 *If this algorithm is initialized by some unit ℓ_2-norm vector $\boldsymbol{x}_{(0)} \in \mathcal{S}_A$, then $\boldsymbol{x}_{(k)}$ is also of unit-norm and belongs to $\mathcal{S}_A$ for all iterations k.*

Remark 2 *Note that the cost function associated with the gradient algorithm (6) corresponds to the restriction in $\mathcal{S}_A$ of the function $\tilde{f}_{2p}(\boldsymbol{x}) = \log\{f_{2p}(\boldsymbol{x})\}$. Because of this restriction, a stationary point of the algorithm need not correspond to a local maximum or a saddle point of $\tilde{f}_{2p}(\boldsymbol{x})$ in the whole space $\mathbb{R}^L$.*

Remark 3 *Note also that the cost function may be identified as the restriction of $f_{2p}(\boldsymbol{x})$ in $\mathcal{S}_A$. In this case, a more general form of the gradient algorithm reads as*

$$\boldsymbol{\nu} = \boldsymbol{x}_{(k)} + \mu_k \nabla^{\mathcal{S}_A} f_{2p}(\boldsymbol{x}_{(k)}); \quad \boldsymbol{x}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|_2}, \tag{7}$$

where μ_k is a variable step-size parameter. The super-exponential algorithm corresponds to the choice

$$\mu_k = \frac{1}{2p f_{2p}(\boldsymbol{x}_{(k)})}.$$

In the sequel, we set $q = 2p - 1$.

Proof: Since the function $\tilde{f}_{2p}(\cdot)$ has to be maximized in $\mathcal{S}_A$, we consider the gradient $\nabla^{\mathcal{S}_A} \tilde{f}_p(\cdot)$:

$$\nabla^{\mathcal{S}_A} \tilde{f}_p(\boldsymbol{x}) = \mathcal{P}_A \nabla \tilde{f}_p(\boldsymbol{x}) = 2p \mathcal{P}_A \left(\frac{\boldsymbol{x}^{\odot q}}{\|\boldsymbol{x}\|_{2p}^{2p}} - \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2} \right)$$

Each iteration of the gradient algorithm in equations (6a-6b), yields a vector $\boldsymbol{x}_{(k)} \in \mathcal{S}_A$ satisfying $\|\boldsymbol{x}_{(k)}\|_2 = 1$. Therefore, by using the above expression of the gradient, one may rewrite (7)

$$\boldsymbol{\nu} = \boldsymbol{x}_{(k)} + \frac{\mathcal{P}_A \boldsymbol{x}_{(k)}^{\odot q}}{\|\boldsymbol{x}_{(k)}\|_{2p}^{2p}} - \frac{\boldsymbol{x}_{(k)}}{\|\boldsymbol{x}_{(k)}\|_2^2} = \frac{\mathcal{P}_A \left(\boldsymbol{x}_{(k)}^{\odot q} \right)}{f_{2p}(\boldsymbol{x}_{(k)})} \tag{8a}$$

$$\boldsymbol{x}_{(k+1)} = \boldsymbol{\nu} / \|\boldsymbol{\nu}\|_2. \tag{8b}$$

We recover the super-exponential algorithm given in (5a-5b). ◇

To simplify further the expressions, let us fix the following notations:

$$\lambda_k \triangleq f_{2p}(\boldsymbol{x}_{(k)}); \quad \alpha_k \triangleq \left\| \mathcal{P}_A \left(\boldsymbol{x}_{(k)}^{\odot q} \right) \right\|_2,$$

and rewrite the gradient search algorithm (6) in a more compact form

$$\lambda_k \boldsymbol{\nu} = \alpha_k \boldsymbol{x}_{(k+1)} = \lambda_k \boldsymbol{x}_{(k)} + \boldsymbol{s}_k, \tag{9}$$

where $\boldsymbol{s}_{(k)} = 1/2p \mathcal{P}_A \nabla f_{2p}(\boldsymbol{x}_{(k)})$.

Having established that the super-exponential algorithm is a gradient search method, we now use this interpretation to give a simple proof of convergence. We have the following

Proposition 1 *Let the two sequences $(\lambda_k)_{k \geq 0}$ and $(\alpha_k)_{k \geq 0}$ defined as above, be obtained at the successive iterations of (6). Then we have the strict interlacing property*

$$0 < \lambda_k < \alpha_k < \lambda_{k+1} < \alpha_{k+1} \leq 1, \tag{10}$$

and the inequality

$$\lambda_{k+1} - \lambda_k > 2p(\alpha_k - \lambda_k), \tag{11}$$

except when $\boldsymbol{x}_{(k)}$ is a stationary point, in which case we have the chain of equalities $\lambda_k = \alpha_k = \lambda_{k+1} = \alpha_{k+1}$.

Proof: Observe first that the boundedness of the two sequences by 1 is a straightforward consequence of $\|\boldsymbol{x}_{(k)}\|_2 = 1$. Next, one may verify directly from (1) that $\nabla f_{2p}(\boldsymbol{x}) = 2p \left(\boldsymbol{x}^{\odot q} - f_{2p}(\boldsymbol{x}) \boldsymbol{x} \right)$, for any $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_2 = 1$. For any such $\boldsymbol{x}$, we then have the orthogonality

$$\langle \boldsymbol{x}, \nabla f_{2p}(\boldsymbol{x}) \rangle = 2p \underbrace{\langle \boldsymbol{x}, \boldsymbol{x}^{\odot q} \rangle}_{f_{2p}(\boldsymbol{x})} - 2p f_{2p}(\boldsymbol{x}) \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0.$$

This then implies $\langle \boldsymbol{x}_{(k)}, \boldsymbol{s}_{(k)} \rangle = 0$ and, applying the Pythagorian theorem on eq.(9), we have

$$\alpha_k^2 \|\boldsymbol{x}_{(k+1)}\|_2^2 = \lambda_k^2 \|\boldsymbol{x}_{(k)}\|_2^2 + \|\boldsymbol{s}_{(k)}\|_2^2 \iff \alpha_k^2 = \lambda_k^2 + \|\boldsymbol{s}_{(k)}\|_2^2,$$

which shows that $\alpha_k > \lambda_k$ for all k unless $\boldsymbol{s}_{(k)} = \boldsymbol{0}$.

We now establish the inequality (11) and next complete the proof of the interlacing property (10). To proceed, let us write $\boldsymbol{x}_{(k+1)}$ from (9) as $\boldsymbol{x}_{(k+1)} = \boldsymbol{x}_{(k)} + \tilde{\boldsymbol{s}}_{(k)}$, where we identify $\tilde{\boldsymbol{s}}_{(k)} = \frac{1}{\alpha_k} \boldsymbol{s}_{(k)} - \frac{\alpha_k - \lambda_k}{\alpha_k} \boldsymbol{x}_{(k)}$. Since the function $\|\cdot\|_{2p}^{2p}$ is convex, its graph lies above its tangent hyperplane. Therefore, $\|\boldsymbol{x}_{(k)} + \tilde{\boldsymbol{s}}_{(k)}\|_{2p}^{2p}$ is minorized by

$$\begin{aligned} \lambda_{k+1} = \|\boldsymbol{x}_{(k)} + \tilde{\boldsymbol{s}}_{(k)}\|_{2p}^{2p} &\geq \|\boldsymbol{x}_{(k)}\|_{2p}^{2p} + \langle \nabla \|\boldsymbol{x}_{(k)}\|_{2p}^{2p}, \tilde{\boldsymbol{s}}_{(k)} \rangle \\ &\geq \|\boldsymbol{x}_{(k)}\|_{2p}^{2p} + 2p \langle \boldsymbol{x}_{(k)}^{\odot q}, \tilde{\boldsymbol{s}}_{(k)} \rangle \\ &\geq \lambda_k + 2p(\alpha_k - \lambda_k). \end{aligned}$$

The strict inequality (11) then follows upon noting that the coincidence set between the graph of $\|\cdot\|_{2p}^{2p}$ and its tangent hyperplane at $\boldsymbol{x}_{(k)}$ does not contain $(\boldsymbol{x}_{(k+1)}, \|\boldsymbol{x}_{(k+1)}\|_{2p}^{2p})$ unless $\tilde{\boldsymbol{s}}_{(k)} = 0$.

Finally, subtracting α_k from both sides of that inequality yields

$$\lambda_{k+1} > \alpha_k + (2p - 1)(\alpha_k - \lambda_k) > \alpha_k$$

which completes the proof. ◇

As the sequence $(\lambda_k)_{k \geq 0}$ is strictly increasing and bounded, we deduce the convergence of the algorithm. Besides, observe

that the interlacing property (10) shows that the two sequences $(\lambda_k)_{k\geq 0}$ and $(\alpha_k)_{k\geq 0}$ converge with the same rate at the same limit. Unfortunately, this observation does not allow one to deduce the rate of convergence of the algorithm. Nonetheless, we shall show in the next section that the step-size associated in Remark 3 to the gradient algorithm in eqs.(6) is optimal for the convergence speed.

5. STABILITY BOUND AND CONVERGENCE SPEED

As suggested in Remark 3, we consider the variable step-size gradient algorithm given in eqs.(7) and derive a bound for the range value of μ_k ensuring the stability. We also show that the selection of the step-size quoted in Remark 3 is optimal for the convergence speed. These results are established in the following

Theorem 1 *Consider the variable step-size algorithm of eqs.(7).*

1. *If for each iteration* $\mathbf{k}$ *the step-size* μ_k *is chosen in the range*

$$0 < \mu_k \leq \frac{2\lambda_k}{2\lambda_k^2 - \alpha_k^2}$$

then the algorithm remains stable.

2. *The choice*

$$\mu_k = \frac{1}{\lambda_k},$$

giving rise to the super-exponential algorithm, is optimal for the convergence speed.

Proof: Let $\boldsymbol{s}_{(k)} = \mathcal{P}_A \nabla f_{2p}(\boldsymbol{x}_{(k)})$. We still have the orthogonality property $\langle \boldsymbol{s}_{(k)}, \boldsymbol{x}_{(k)} \rangle = 0$ so that we may write

$$\|\boldsymbol{\nu}\|_2^2 = 1 + \mu_k^2 \|\boldsymbol{s}_{(k)}\|_2^2 = 1 + \mu_k^2(\alpha_k^2 - \lambda_k^2) \triangleq \gamma_k^2. \qquad (12)$$

Setting $\tilde{\boldsymbol{s}}_{(k)} = \frac{1-\gamma_k}{\gamma_k}\boldsymbol{x}_{(k)} + \frac{\mu_k}{\gamma_k}\boldsymbol{s}_{(k)}$, allows one to re-write the algorithm in the compact form

$$\boldsymbol{x}_{(k+1)} = \boldsymbol{x}_{(k)} + \tilde{\boldsymbol{s}}_{(k)}.$$

Following the same steps as before, we have

$$\lambda_{k+1} - \lambda_k \geq 2p\langle \tilde{\boldsymbol{s}}_{(k)}, \boldsymbol{x}_{(k)}^{\odot q} \rangle.$$

A sufficient condition for the algorithm to remain stable is

$$h_k(\mu_k) \triangleq \langle \tilde{\boldsymbol{s}}_{(k)}, \boldsymbol{x}_{(k)}^{\odot q} \rangle \geq 0.$$

Now, one can verify easily that the expression of the function $h_k(\mu_k)$ is given by

$$h_k(\mu_k) = \frac{\gamma_k - 1}{\mu_k \gamma_k}(\gamma_k + 1 - \lambda_k \mu_k).$$

Using the definition of γ_k from (12), we deduce that for $\mu_k > 0$,

$$h_k(\mu_k) \geq 0 \iff \mu_k \leq \frac{2\lambda_k}{2\lambda_k^2 - \alpha_k^2}.$$

This completes the proof of part 1 of the theorem.

Part 2 of the theorem will be deduced from the simple observation that at each iteration k, the optimal value for μ_k is the one which maximizes $h_k(\mu_k)$. Some straightforward algebra show that the derivative, $h_k'(\mu_k)$, of $h_k(\cdot)$ reads as

$$h_k'(\mu_k) = \frac{\alpha_k^2 - \lambda_k^2}{\gamma_k^3}(1 - \lambda_k \mu_k).$$

Thus the choice $\mu_k = \frac{1}{\lambda_k}$ is optimal, and this completes the proof.

6. CONCLUDING REMARKS

The super-exponential algorithm for blind channel equalization has been reviewed in the more realistic undermodeled case. In a previous work, we have characterized the possible convergence points of this algorithm as the maxima of a member, $f_{2p}(\boldsymbol{x})$, of a family of blind channel equalization criteria, including the popular Godard and Shalvi-Weinstein cost functions. More explicitly, we have shown here, that this algorithm is equivalent to a gradient search method. Using this interpretation, a simple proof of convergence has been given. Some issues concerning the convergence rate have also been considered. In particular, we have shown that the variable step-size associated with the algorithm is optimal. We have also given, for each iteration, a range value of the step-size that guarantees the stability.

7. REFERENCES

[1] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. on Sig. Proc.*, vol. 43, pp. 515–525, Febrary 1995.

[2] J. Treichler, I. Fijalkow, and C. Johnson, "Fractionally spaced equalizers. How long should they be?," *Signal Processing Mag.*, pp. 65–81, May 1996.

[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least squares approach to blind channel identification," *IEEE Trans. on Sig. Proc.*, vol. 43, pp. 340–349, December 1995.

[4] A. Liavas, P. Regalia, and J. Delmas, "Robustness of least-squares and subspace methods for blind channel identification/equalization with respect to effective channel undermodeling/overmodeling." submitted to IEEE Trans. on SP.

[5] D. Donoho, "On minimum entropie deconvolution," in *Applied time series analysis II* (D. F. Findley, ed.), New York: Academic, 1981.

[6] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Comm.*, vol. 44, no. 11, pp. 1867–1875, 1980.

[7] O. S. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. on Info. Theory*, vol. 36, pp. 312–321, March 1990.

[8] P. A. Regalia, "On the equivalence between the Godart and Shalvi-Weinstein schemes of blind equalization." Accepted in Signal Processing, 1998.

[9] O. S. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. on Info. Theory*, vol. 39, pp. 504–519, March 1993.

[10] P. A. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria." Accepted IEEE Trans. on Sig. Proc., 1998.

[11] J. C.R. Johnson and B. D. O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary source case," *International Journal of Adaptive Control and Signal Processing*, vol. 9, pp. 301–324, July-August 1995.