connecting two points in the multiuser constellation is a Delaunay segment, thus facilitating the construction of the Voronoi diagram. In addition, we have shown a direct link between the algebraic notion of "indecomposable error sequences" introduced by Verdú and the geometric notion of a Delaunay segment.

Based on the obtained results, two approximate ML detectors were introduced whose performance compared well to existing popular algorithms.

ACKNOWLEDGMENT

The authors wish to thank Jianer Chen, Ruiyu Wang, Predrag Spasojević, Tony Ottosson, Xiaodong Wang, and anonymous reviewers for helpful suggestions and discussions.

REFERENCES

[1] R. Lupas and S. Verdú, "Linear multiuser detector for synchronous code-division multiple-access communications," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
[2] ——, "Near-far resistance of multiuser detection in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, Apr. 1990.
[3] M. K. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.
[4] ——, "Near-optimum detection in synchronous code-division multiple-access systems," *IEEE Trans. Commun.*, vol. 39, pp. 725–736, May 1991.
[5] A. Duel-Hallen, "Decorrelating decision-feedback multiuser detection for synchronous code-division multiple-access channel," *IEEE Trans. Commun.*, vol. 41, pp. 285–290, Feb. 1993.
[6] ——, "A family of multiuser decision-feedback detectors for asynchronous code-division multiple-access channels," *IEEE Trans. Commun.*, vol. 43, pp. 421–433, Feb./Mar./Apr. 1995.
[7] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.
[8] C. Sankaran and A. Ephremides, "Solving a class of optimum multiuser detection problems with polynomial complexity," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1958–1961, Sept. 1998.
[9] E. Agrell and T. Ottosson, "ML optimal CDMA multiuser receiver," *Electron. Lett.*, vol. 31, no. 18, pp. 1554–1555, 1995.
[10] T. Ottosson, "Multirate schemes and multiuser decoding in DS/CDMA systems," Ph.D. dissertation, School Elec. Comp. Eng., Chalmers Univ. Technol., Göteborg, Sweden, 1995.
[11] E. Agrell, "Voronoi regions for binary linear block codes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 310–316, Jan. 1996.
[12] ——, "On the Voronoi neighbor ratio for binary linear block codes," *IEEE Trans. Inform. Theory*, vol. 44, pp. 3064–3072, Nov. 1998.
[13] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[14] J. Goodman and J. O'Rourke, *Handbook of Descrete and Computational Geometry*. Boca Raton, FL: CRC Press LLC, 1997.
[15] K. Mulmuley, *Computational Geometry, an Introduction Through Randomized Algorithms*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[16] J. Boissonnat and M. Yvinec, *Algoritmic Geometry*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[17] A. Okabe, B. Boots, and K. Sugihara, *Spatial Tesselations Concepts and Applications of Voronoi Diagrams*. New York: Wiley, 1992.
[18] W. Luo and A. Ephremides, "Indecomposable error sequences in multiuser detection," in *Proc. 33th Annu. Conf. Information Sciences and Systems*, Mar. 1999, pp. 275–280.
[19] L. K. Rasmussen, T. J. Lim, and T. M. Aulin, "Breadth-first maximum likelihood detection in multiuser CDMA," *IEEE Trans. Commun.*, vol. 45, pp. 1176–1178, Oct. 1997.
[20] M. Nasiri-Kenari, R. R. Sylvester, and C. K. Rushforth, "An efficient soft-in-soft-out multiuser detector for synchronous CDMA with error-control coding," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 947–953, Aug. 1998.

A Gradient Search Interpretation of the Super-Exponential Algorithm

Mamadou Mboup and Phillip A. Regalia, *Fellow, IEEE*

***Abstract*—This correspondence reviews the super-exponential algorithm proposed by Shalvi and Weinstein for blind channel equalization. The principle of this algorithm—*Hadamard exponentiation, projection over the set of attainable combined channel-equalizer impulse responses followed by a normalization*—is shown to coincide with a gradient search of an extremum of a cost function. The cost function belongs to the family of functions given as the ratio of the standard ℓ_{2p} and ℓ_2 sequence norms, where $p > 1$. This family is very relevant in blind channel equalization, tracing back to Donoho's work on minimum entropy deconvolution and also underlying the Godard (or Constant Modulus) and the earlier Shalvi–Weinstein algorithms. Using this gradient search interpretation, which is more tractable for analytical study, 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.**

***Index Terms*—Blind channel equalization, Hadamard power, stationary points, gradient method.**

I. INTRODUCTION

Most of the recent development of second-order statistics-based methods for blind channel identification/equalization are supported by the Bezout identity [1]. Among the more popular of the methods are subspace methods [1], including cyclostationarity-based methods [2], and least squares method [3]. The performance of these methods may, however, dregrade dramatically if the equalizer length is either larger (overmodeled) or smaller (undermodeled) than the length of the significant part of the true channel [4].

Direct equalization methods, based on the minimization or maximization of a higher order statistics cost function, seem to be more robust to the violation of so-called *length and zero conditions* [5]. The popular Godard [6] algorithm and the earlier Shalvi–Weinstein algorithm [7] are well-known examples of this class of method. As a matter of fact, these two algorithms are closer than they initially appear as they both amount to seeking a maximum of a same cost function (see, e.g., [8]), belonging to the family of Donoho's deconvolution objective functions [9].

In this correspondence, we show that the super-exponential algorithm proposed by Shalvi and Weinstein [10] also belongs to this same family of algorithms. Indeed, this already appeared in our previous work [11] where we have shown that the convergent points of this algorithm correspond to the maxima, over the set of attainable combined channel-equalizer impulse responses, of some member of the above family. 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 earlier Shalvi–Weinstein algorithm. The characterization of the stationary points of the super-exponential algorithm is not sufficient to deduce the

Manuscript received September 1, 1998; revised October 1, 1999. The material in this correspondence was presented in part at ICASSP'99, Phoenix, AZ, March 15–19, 1999.

M. Mboup is with the UFR de Mathématiques et Informatique-CRIP5, Université René Descartes-Paris V 45, 75270 Paris Cedex 06, France (e-mail: mboup@math-info.univ-paris5.fr).

P. A. Regalia is with the Département Signal et Image, Institut National des Télécommunications-9, 91011 Evry Cedex, France (e-mail: regalia@int-evry.fr).

Communicated by U. Madhow, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(00)07003-6.

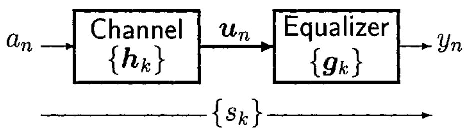

Fig. 1. The channel-equalizer combined response.

domain of attraction of each maximum of the corresponding cost function nor how these maxima are approached.

Section III is devoted to these questions. More explicitly, Proposition 1 establishes the equivalence between the super-exponential algorithm and a gradient-search algorithm whose associated cost function is precisely one member of Donoho's family of functions. The gradient search is constrained within the set of attainable combined channel-equalizer impulse responses. With this gradient search interpretation, the analysis of convergence reduces to an analytic description of the error surface. Indeed, this study becomes more classical although it may be a difficult task to obtain an analytic parametrization of the error surface [12]. From this follows a simple proof of convergence of the super-exponential algorithm. Finally, we show in Section IV that the gradient step-size choice giving rise to the super-exponential algorithm is optimal. Section II gives the notations and a brief review of the super-exponential algorithm.

II. Problem Setting and Preliminaries

We consider the cascade structure in Fig. 1 and we assume that the channel is real, stable, and causal. The combined channel-equalizer impulse response $\{s_k\}$, mapping the source sequence $\{a_n\}$ to the equalizer output sequence $\{y_n\}$ as in

$$y_n = \sum_k s_k a_{n-k} \tag{1}$$

is the convolution of the channel $\{\boldsymbol{h}_k\}_{k\geq 0}$ and the equalizer $\{\boldsymbol{g}_k\}_{k=0}^{L}$ impulse responses

$$s_k = \sum_i^L \boldsymbol{g}_i \boldsymbol{h}_{k-i}. \tag{2}$$

Here, the length of the equalizer is $L+1$. The received sequence $\{\boldsymbol{u}_n\}$ is N-dimensional, with $N \geq 1$.

This accommodates both the baud rate case $(N=1)$ and the multisensor and/or the fractionally spaced case $(N>1)$. Each coefficient $\boldsymbol{h}_k$ is, therefore, an N-column vector while each $\boldsymbol{g}_k$ is an N-row vector. We assume that the channel and equalizer are both bounded-input–bounded-output stable so that the convolution in (2) results in a square-summable sequence (in ℓ_2). In matrix form, this convolution reads as

$$\underbrace{\begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_L \\ s_{L+1} \\ \vdots \end{bmatrix}}_{\boldsymbol{s}} \underbrace{\begin{bmatrix} \boldsymbol{h}_0^t & \boldsymbol{0}^t & \dots & \boldsymbol{0}^t \\ \boldsymbol{h}_1^t & \boldsymbol{h}_0^t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}^t \\ \boldsymbol{h}_L^t & \dots & \boldsymbol{h}_1^t & \boldsymbol{h}_0^t \\ \boldsymbol{h}_{L+1}^t & \boldsymbol{h}_L^t & \dots & \boldsymbol{h}_1^t \\ \vdots & \ddots & \ddots & \vdots \end{bmatrix}}_{\mathcal{H}} \begin{bmatrix} \boldsymbol{g}_0^t \\ \boldsymbol{g}_1^t \\ \vdots \\ \boldsymbol{g}_L^t \end{bmatrix} \tag{3}$$

showing that the combined response $\boldsymbol{s}$ is restricted to the range space (in ℓ_2) of the convolution matrix $\mathcal{H}$. We denote this space by $\mathcal{S}_A$: the ℓ_2 subspace of attainable combined responses. The orthogonal projection operator onto $\mathcal{S}_A$ is given then by $\mathcal{P}_A = \mathcal{H}(\mathcal{H}^t\mathcal{H})^\sharp\mathcal{H}^t$, where the superscript $\sharp$ denotes the (pseudo-) inversion. If $\mathcal{P}_A = \mathbf{I}$, then any prescribed combined response is attainable; this is termed the *sufficient-order* case. If, on the other hand, $\mathcal{P}_A \neq \mathbf{I}$, then only a proper subset of ℓ_2 can be reached by varying the equalizer coefficients; this is termed the *undermodeled* case.

An ideal setting of the equalizer is one which brings the combined responses into a pinning vector form

$$\boldsymbol{s} = [\dots 0 \underbrace{1}_{n} 0 \dots]^t \triangleq \boldsymbol{e}_n$$

corresponding to an n sample pure delay.

Recall that for a given real vector $\boldsymbol{s}$, its mth Hadamard power, denoted as $\boldsymbol{s}^{\odot m}$, is defined componentwise by

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

Now, observe that if the dominant term of $\boldsymbol{s}$ is unique and in position n, then the ideal combined response $\boldsymbol{e}_n$ may be approached by the mth Hadamard power $(\boldsymbol{s}/s_n)^{\odot m}$ as m tends to infinity. This observation led to the so-called super-exponential algorithm proposed by Shalvi and Weinstein (see [10] and [13]) and given by

$$\boldsymbol{\nu} = \mathcal{P}_A \boldsymbol{s}_{(k)}^{\odot q} \tag{4a}$$

$$\boldsymbol{s}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|}. \tag{4b}$$

Here q is an odd number: $q = 2p-1$ with $p>1$ and the algorithm is initialized by some $\boldsymbol{s}_{(0)} \in \mathcal{S}_A$ such that $\|\boldsymbol{s}_{(0)}\| = 1$, using the same norm as in (4b). One may verify that the subspace angle between two successive updates $\boldsymbol{s}_{(k)}$ and $\boldsymbol{s}_{(k+1)}$ is insensitive to the choice of the norm in (4b); to simplify certain developments to follow, we may thus assume ℓ_2 normalization with no loss of generality. This algorithm has also been obtained in an independent and more constructive way in [11], as an iterative procedure for seeking a local maximum of the family of objective functions

$$f_{2p}(\boldsymbol{s}) = \left\{\frac{\|\boldsymbol{s}\|_{2p}}{\|\boldsymbol{s}\|_2}\right\}^{2p}, \qquad p = 2, \dots, \infty. \tag{5}$$

One may readily check that $f_{2p}(\beta\boldsymbol{s}) = f_{2p}(\boldsymbol{s})$ for any $\beta \neq 0$, showing that $f_{2p}(\boldsymbol{s})$ is insensitive to the radial parameter of $\boldsymbol{s}$. We may also verify the boundedness property: $0 < f_{2p}(\boldsymbol{s}) \leq 1$ for all $\boldsymbol{s} \in \ell_2$.

This family of functions traces back to Donoho [9] and is very relevant in blind deconvolution. For example, the popular Godard algorithm [6] as well as the earlier Shalvi-Weinstein algorithm [7] both amount to seeking a maximum of $f_4(\boldsymbol{s})$ although they use different assumptions on the source signal statistics (see, e.g., [7], [8]).

The next section goes further into the interpretation of the super-exponential algorithm (4) by showing that it may be viewed as a gradient descent procedure.

III. A Gradient Search Method

We recall that the directional derivative of a function $f: \mathbb{R}^m \to \mathbb{R}$ (locally Lipschitzian) at $\boldsymbol{s}$, in the direction $\boldsymbol{d} \in \mathbb{R}^m$ is [14]

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

If f is differentiable at $\boldsymbol{s}$, then this directional derivative becomes $f'(\boldsymbol{s}, \boldsymbol{d}) = \langle \nabla f(\boldsymbol{s}), \boldsymbol{d} \rangle$, where $\nabla f(\boldsymbol{s})$ is the gradient of f at $\boldsymbol{s}$. We proceed to introduce the concept of a projected gradient which, naturally, extends that of a directional derivative in a multidirectional setting. Let $\mathcal{S}$ be a vector subspace of $\mathbb{R}^m$ and let $\mathcal{P}$ be the orthogonal projection matrix from $\mathbb{R}^m$ into $\mathcal{S}$. Then, the projected gradient of f in $\mathcal{S}$, or equivalently, the gradient of f over $\mathcal{S}$, at a point $\boldsymbol{s} \in \mathbb{R}^m$ is defined as

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

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

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

and the gradient of f over $\mathcal{S}$, $\nabla^{\mathcal{S}} f(\boldsymbol{s})$, is simply the projection of the gradient of f into $\mathcal{S}$.

Now, we are ready to show the following equivalence.

Proposition 1: The super-exponential algorithm in (4) with ℓ_2 normalization is equivalent to the gradient search algorithm defined by

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

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

Remark 1: If this algorithm is initialized by some unit ℓ_2-norm vector $\boldsymbol{s}_{(0)} \in \mathcal{S}_A$, then $\boldsymbol{s}_{(k)}$ is also of unit ℓ_2-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 $\tilde{f}_{2p}(\boldsymbol{s}) = \ln\{f_{2p}(\boldsymbol{s})\}$. Because of this restriction, a stationary point of the algorithm need not coincide with a local maximum or a saddle point of $\tilde{f}_{2p}(\boldsymbol{s})$ in the strict sense (i.e., in the whole space $\mathbb{R}^L$).

Remark 3: Note also that the cost function may be identified as the restriction of $f_{2p}(\boldsymbol{s})$ in $\mathcal{S}_A$. In this case, (6) may be written in a more general form as

$$\boldsymbol{\nu} = \boldsymbol{s}_{(k)} + \mu_k \nabla^{\mathcal{S}_A} f_{2p}(\boldsymbol{s}_{(k)}) \tag{7}$$

with the variable step-size parameter given by

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

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

Proof: As the objective is to find a local maximum over $\mathcal{S}_A$ of the cost function $\tilde{f}_{2p}(\cdot)$, we consider the projected gradient $\nabla^{\mathcal{S}_A}\tilde{f}_{2p}(\cdot)$

$$\nabla^{\mathcal{S}_A}\tilde{f}_{2p}(\boldsymbol{s}) = \mathcal{P}_A \nabla \tilde{f}_{2p}(\boldsymbol{s}) \tag{8}$$

$$= 2p\mathcal{P}_A\left(\frac{\boldsymbol{s}^{\odot q}}{\|\boldsymbol{s}\|_{2p}^{2p}} - \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_2^2}\right). \tag{9}$$

Each iteration of the gradient algorithm in (6a) and (6b) yields a vector $\boldsymbol{s}_{(k)} \in \mathcal{S}_A$ satisfying $\|\boldsymbol{s}_{(k)}\|^2 = 1$. Therefore, by using the above expression of the gradient, one may rewrite (7) or equivalently (6a) as

$$\boldsymbol{\nu} = \boldsymbol{s}_{(k)} + \frac{\mathcal{P}_A\boldsymbol{s}_{(k)}^{\odot q}}{\|\boldsymbol{s}_{(k)}\|_{2p}^{2p}} - \frac{\boldsymbol{s}_{(k)}}{\|\boldsymbol{s}_{(k)}\|_2^2} = \frac{1}{f_{2p}(\boldsymbol{s}_{(k)})}\mathcal{P}_A(\boldsymbol{s}_{(k)}^{\odot q}) \tag{10}$$

and adding the normalization step, we recover the super-exponential algorithm (4). □

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

$$\lambda_k \triangleq f_{2p}(\boldsymbol{s}_{(k)}) \qquad \alpha_k \triangleq \|\mathcal{P}_A(\boldsymbol{s}_{(k)}^{\odot q})\|^2$$

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

$$\alpha_k \boldsymbol{s}_{(k+1)} = \lambda_k \boldsymbol{s}_{(k)} + \boldsymbol{x}_{(k)} \tag{11}$$

where

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

Remark 4: Observe that if we define θ_k as the angle between the successive updates $\boldsymbol{s}_{(k)}$ and $\boldsymbol{s}_{(k+1)}$, then we find that $\cos(\theta_k) = \frac{\lambda_k}{\alpha_k}$ and we may write

$$\boldsymbol{s}_{(k+1)} = \cos(\theta_k)\boldsymbol{s}_{(k)} + \sin(\theta_k)\boldsymbol{x}_{(k)} \tag{12}$$

which provides another interpretation for the algorithm.

Having established that the super-exponential algorithm is a gradient search method, we now use this interpretation to give a simple and short proof of convergence (see also [11]). We have

Proposition 2: Let the two sequences $\{\lambda_k\}_{k\geq 0}$ and $\{\alpha_k\}_{k\geq 0}$, where λ_k and α_k are defined as above, be obtained at the successive iterations of (6). Then, we have the strict interlacing property

$$0 < \lambda_k < \alpha_k < \lambda_{k+1} < 1, \tag{13}$$

and the inequality

$$\lambda_{k+1} - \lambda_k > 2p(\alpha_k - \lambda_k) \tag{14}$$

except when $\boldsymbol{s}_{(k)}$ 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 sequence $\{\alpha_k\}$ by 1 is a straightforward consequence of $\|\boldsymbol{s}_{(k)}\|^2 = 1$ while we already know that $\lambda_k \leq 1$, with equality if and only if $\boldsymbol{s}_{(k)}$ is a stationary point corresponding to an ideal combined response. Next, one may verify directly from the definition of the function f_{2p} that

$$\nabla f_{2p}(\boldsymbol{s}) = 2p(\boldsymbol{s}^{\odot q} - f_{2p}(\boldsymbol{s})\boldsymbol{s})$$

for any $\boldsymbol{s}$ such that $\|\boldsymbol{s}\|^2 = 1$. For any such $\boldsymbol{s}$, we have the orthogonality

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

This then implies $\langle \boldsymbol{s}_{(k)}, \boldsymbol{x}_{(k)}\rangle = 0$ and, applying the Pythagorean theorem to (11), we have

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

Since $\|\boldsymbol{s}_{(k+1)}\|_2 = \|\boldsymbol{s}_{(k)}\|_2 = 1$, we obtain $\alpha_k^2 = \lambda_k^2 + \|\boldsymbol{x}_{(k)}\|_2^2$, which shows that $\alpha_k \geq \lambda_k$ for all k, with equality if and only if $\boldsymbol{x}_{(k)} = \boldsymbol{0}$, i.e., at a stationary point.

We now establish the inequality (14) and next complete the proof of the interlacing property (13). To proceed, let us write $\boldsymbol{s}_{(k+1)}$ from (11) as

$$\boldsymbol{s}_{(k+1)} = \frac{\lambda_k}{\alpha_k}\boldsymbol{s}_{(k)} + \frac{1}{\alpha_k}\boldsymbol{x}_{(k)} = \boldsymbol{s}_{(k)} + \tilde{\boldsymbol{x}}_{(k)}$$

where we identify

$$\tilde{\boldsymbol{x}}_{(k)} = \frac{1}{\alpha_k}\boldsymbol{s}_{(k)} - \frac{\alpha_k - \lambda_k}{\alpha_k}\boldsymbol{x}_{(k)}.$$

Since the function $\|\cdot\|_{2p}^{2p}$ is convex, its graph lies above its tangent hyperplane at $\boldsymbol{x}_{(k)}$. Therefore, $\|\boldsymbol{s}_{(k)} + \tilde{\boldsymbol{x}}_{(k)}\|_{2p}^{2p}$ is minimized by

$$\lambda_{k+1} = \|\boldsymbol{s}_{(k)} + \tilde{\boldsymbol{x}}_{(k)}\|_{2p}^{2p} \geq \|\boldsymbol{s}_{(k)}\|_{2p}^{2p} + \langle \nabla\|\boldsymbol{s}_{(k)}\|_{2p}^{2p}, \tilde{\boldsymbol{x}}_{(k)}\rangle \tag{15}$$

$$\geq \lambda_k + \frac{2p}{\alpha_k}\langle \boldsymbol{s}_{(k)}^{\odot q}, \boldsymbol{x}_{(k)}\rangle - 2p\frac{\alpha_k - \lambda_k}{\alpha_k}\langle \boldsymbol{s}_{(k)}, \boldsymbol{s}_{(k)}^{\odot q}\rangle \tag{16}$$

$$\geq \lambda_k + 2p(\alpha_k - \lambda_k). \tag{17}$$

The strict inequality (14) then follows upon noting that the coincidence set between the graph of $\|\cdot\|_{2p}^{2p}$ and its tangent hyperplane at $\boldsymbol{s}_{(k)}$ does not contain the element $(\boldsymbol{s}_{(k+1)}, \|\boldsymbol{s}_{(k+1)}\|_{2p}^{2p})$ unless $\tilde{\boldsymbol{x}}_{(k)} = \boldsymbol{0}$.

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. □

The convergence of the algorithm is thus a direct consequence of the sequence $\{\lambda_k\}_{k\geq 0}$ being strictly increasing and bounded. In addition, observe that the interlacing property (13) shows that the two sequences $\{\lambda_k\}_{k\geq 0}$ and $\{\alpha_k\}_{k\geq 0}$ converge with the same rate, to the same limit. Unfortunately, this observation does not allow one to deduce the rate of convergence of the algorithm in the undermodeled case. We shall show in the next section, however, that the step-size associated in Remark 3 to the gradient algorithm is optimal for convergence speed.

IV. Stability Bound and Convergence Speed

As suggested in Remark 3, we consider the gradient algorithm of (6) in its general form

$$\boldsymbol{\nu} = \boldsymbol{s}_{(k)} + \mu_k \mathcal{P}_A \nabla f_{2p}(\boldsymbol{s}_{(k)}) \tag{18a}$$

$$\boldsymbol{s}_{(k+1)} = \frac{\boldsymbol{\nu}}{\|\boldsymbol{\nu}\|_2} \tag{18b}$$

where the (positive) variable step-size μ_k is now a free parameter. In this section, we derive a bound for the range of μ_k ensuring the stability of the algorithm. 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.

Theorem 1: Consider the algorithm of (18) with a variable step-size μ_k.

1) If for each iteration 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: To begin, set $\boldsymbol{x}_{(k)} = \mathcal{P}_A \nabla f_{2p}(\boldsymbol{s}_{(k)})$. We still have the orthogonality property $\langle \boldsymbol{s}_{(k)}, \boldsymbol{x}_{(k)} \rangle = 0$ so that we may write from (18a)

$$\|\boldsymbol{\nu}\|_2^2 = 1 + \mu_k^2 \|\boldsymbol{x}_{(k)}\|_2^2 = 1 + \mu_k^2(\alpha_k^2 - \lambda_k^2) \triangleq \gamma_k^2. \tag{19}$$

Setting

$$\tilde{\boldsymbol{x}}_{(k)} = \frac{1-\gamma_k}{\gamma_k}\, \boldsymbol{s}_{(k)} + \frac{\mu_k}{\gamma_k}\, \boldsymbol{x}_{(k)}$$

allows one to rewrite the algorithm in the compact form

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

Following the same steps as in the Proof of Proposition 2, we have

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

A sufficient condition for the algorithm to remain stable is given by

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

Moreover, the condition $h_k(\mu_k) > 0$ for all k, except at a stationary point, ensures a monotone convergence of (18). One can readily check that a closed-form expression for $h_k(\mu_k)$ is

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

Now, using the definition of γ_k from (19), we deduce that for $\mu_k > 0$

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

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 shows that the derivative $h_k'(\cdot)$ 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. □

V. 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 $(f_{2p}(\boldsymbol{s}_{(k)}))$ of a family of blind equalization criteria including the popular Godard and the Shalvi–Weinstein cost functions. This association of the super-exponential algorithm with the principle of maximizing a cost function has been made more explicit in this correspondence. We have shown that the algorithm is seeking a local maximum over the set of attainable combined response is of $f_{2p}(\cdot)$, using a gradient search method. This has allowed us to rewrite the super-exponential algorithm in a form of a classical gradient algorithm which is more tractable for convergence studies. We have given in Proposition 2 a simple proof of convergence for the algorithm. Some issues concerning the convergence rate have also been considered: we have shown that the variable step-size associated with the algorithm is optimal for convergence speed. We have also given for each iteration, the range value of the step-size that guarantees the stability. In the absence of an analytic description of the "attainable" error surface, deriving a convergence rate remains open in the undermodeled case.

References

[1] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 515–525, Feb. 1995.

[2] Y. Li and Z. Ding, "Blind channel identification based on second-order cyclostationary statistics," in *Int. Conf. Accoustics, Speech, and Signal Processing (ICASSP'93)*, vol. IV, 1993, pp. 81–84.

[3] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least squares approach to blind channel identication," *IEEE Trans. Signal Processing*, vol. 43, pp. 340–349, Dec. 1995.

[4] A. P. Liavas, P. A. Regalia, and J. P. Delmas, "Robustness of least-squares and subspace methods for blind channel identification/equalization with respeft to effective channel undermodeling/overmodeling," *IEEE Trans. Signal Processing*, vol. 47, pp. 1636–1645, June 1999.

[5] I. Fijalkow, A. Touzni, and J. R. Trichler, "Fractionally spaced equalization using cma: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, pp. 56–66, Jan. 1997.

[6] D. N. Godard, "Self-recovering equalization and carfrier tracking in two dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[7] O. S. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels," *IEEE Trans. Information Theory*, vol. 36, no. 3, pp. 312–321, Mar. 1990.

[8] P. A. Regalia, "On the equivalence between the Godard and Shalvi-Weinstein schemes of blind deconvolution," *Signal Processing*, vol. 73, pp. 185–190, Feb. 1999.

[9] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. F. Findley, Ed. New York: Academic, 1981.

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

[11] P. A. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria," *IEEE Trans. Signal Processing*, vol. 47, pp. 760–770, Mar. 1999.

[12] C. R. Johnson, Jr. and B. D. O. Anderson, "Godard blind equalizer error surface characteristics: White, zero-mean, binary source case," *Int. J. Adapt. Control and Signal Processing*, vol. 9, no. July-Aug., pp. 301–324, 1995.

[13] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," *IEEE Trans. Signal Processing*, vol. 45, pp. 650–657, Mar. 1997.

[14] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, 2nd ed. Berlin, Germany: Springer-Verlag, 1996.