UNDERMODELED EQUALIZATION IN NOISY MULTI-USER CHANNELS

Phillip A. Regalia

Département Signal et Image
Institut National des Télécommunications
9 rue Charles Fourier
F-91011 Evry cedex France
regalia@int-evry.fr

Mamadou Mboup

UFR Mathématiques et Informatique
Université René Descartes
45, rue des Saints-Pères
75270 Paris cedex 06
mboup@math-info.univ-paris5.fr

ABSTRACT

Blind equalization in noisy multi-user channels has met with increasing attention with the advent of multi-access digital communication systems. We develop a unified formulation which combines the desired sources and the backgound noise into a common convolutional model. We then obtain a characterization of stationary points for a family of blind criteria in undermodeled cases, which incorporates the influence of differening source statistics and background noise correlation properties. We derive also a *global* step-size bound which ensures convergence of a gradient search procedure, and confirm that the super-exponential algorithm results from an optimal choice of this step-size parameter.

1. INTRODUCTION

Many methods in blind equalization can be understood as minimum entropy methods, first developed by Donoho [1], and subsequently rediscovered [2] and refined [3] by Shalvi and Weinstein. Extensions of this technique to multiple source deconvolution problems may be found in [5], [9], and equivalences with the constant modulus (or Godard) algorithm have also been placed in evidence [4].

A key result from [4] (mono-source case) and [5] (multi-source case) asserts that each extremum of a particular minimum entropy criterion yields an ideal equalizer, i.e., giving a combined (channel-equalizer) impulse response having a sole nonzero term. The validity of this result, however, hinges strongly on the assumption that an arbitrary configuration of the combined (channel-equalizer) impulse response can be attained, including thus any ideal solution which would restore perfectly the transmitted sequence.

In practice, the presence of channel noise, as well as co-channel interference due to multiple users, will prohibit any equalizer setting from restoring perfectly the transmitted sequence, and the reconstruction error tends to worsen with shorter equalizer lengths. The behavior of blind equalization algorithms in these “undermodeled” cases is not as clearly understood, as minimum entropy criteria are usually non convex, exhibiting numerous local minima.

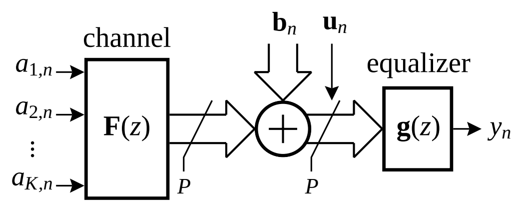

Figure 1: The channel-equalizer cascade, including noise.

The intent of this work is characterize the stationary points of a family of minimum entropy criteria, for the multi-source noisy channel setting depicted in Figure 1, and using a finite (and generally insufficient) length equalizer. We shall also examine a gradient search procedure and obtain a step-size bound ensuring convergence regardless of the initialization point, and verify that the super-exponential algorithm results from a certain “optimal” choice of the step-size.

2. PROBLEM STRUCTURE

We consider a multichannel noisy deconvolution problem, depicted in Figure 1, in which the observed vector process $\{\mathbf{u}_n\}$, having P components, is of the form

$$(P \updownarrow) \quad \mathbf{u}_n = \mathbf{F}(z)\,\mathbf{a}_n + \mathbf{b}_n$$

where:

- The source signals $\{a_{i,n}\}$ which comprise the vector $\mathbf{a}_n$ are each independent, identically distributed (i.i.d.) random sequences, presumably non-Gaussian, and mutually independent as well. We assume, with no loss of generality, that each is scaled to unit variance: $E[a_{i,n}^2] = 1$.

- The transfer matrix $\mathbf{F}(z)$ is stable and causal:

$$\mathbf{F}(z) = \sum_{k=0}^{\infty} \mathbf{F}_k z^k, \qquad \text{with} \quad \sum_{k=0}^{\infty} \|\mathbf{F}_k\| < \infty.$$

  Here z (rather than z^{-1}) denotes the unit delay operator: $z\,\mathbf{a}_n = \mathbf{a}_{n-1}$. The product $\mathbf{F}(z)\,\mathbf{a}_n$ may then be interpreted as the convolution sum

$$\mathbf{F}(z)\,\mathbf{a}_n = \sum_{k=0}^{\infty} \mathbf{F}_k z^k \mathbf{a}_n = \sum_{k=0}^{\infty} \mathbf{F}_k\,\mathbf{a}_{n-k}$$

- The background noise vector $\mathbf{b}_n$ is Gaussian, and independent of the source signals.

If the noise term is indeed Gaussian, then we may write an innovations model of the form

$$\mathbf{b}_n = \mathbf{L}(z)\,\alpha_n = \sum_{k=0}^{\infty} \mathbf{L}_k\,\alpha_{n-k}$$

where the vector process $\{\alpha_n\}$ is normalized white noise:

$$E[\alpha_n\,\alpha_m^T] = \begin{cases} \mathbf{I}, & n = m; \\ \bigcirc, & n \neq m. \end{cases}$$

Since $\{\alpha_n\}$ is white and Gaussian, each sample is independent and identically distributed (i.i.d.).

We can now combine the signal and noise terms into a common convolutional model as

$$\begin{aligned} \mathbf{u}_n &= \mathbf{F}(z)\,\mathbf{a}_n + \mathbf{L}(z)\,\alpha_n \\ &= \underbrace{[\mathbf{F}(z) \quad \mathbf{L}(z)]}_{\mathcal{F}(z)} \begin{bmatrix} \mathbf{a}_n \\ \alpha_n \end{bmatrix} \\ &= \sum_{k=0}^{\infty} \underbrace{[\mathbf{F}_k \quad \mathbf{L}_k]}_{\mathcal{F}_k} \begin{bmatrix} \mathbf{a}_{n-k} \\ \alpha_{n-k} \end{bmatrix} \end{aligned}$$

Some of the sources (namely, the $\{\alpha_{i,n}\}$) may be Gaussian, while the others (namely, the $\{a_{i,n}\}$) may be non-Gaussian.

The equalizer in Figure 1 is a multi-input/single-output transversal filter:

$$y_n = \sum_{k=0}^{M} \mathbf{g}_k \mathbf{u}_{n-k} = \sum_{k=0}^{M} \mathbf{g}_k z^k \mathbf{u}_n = \mathbf{g}(z)\,\mathbf{u}_n.$$

Here each impulse response term $\mathbf{g}_k$ is a row vector of P elements. We may write this in terms of the original source vector $\left[\begin{smallmatrix} \mathbf{a}_n \\ \alpha_n \end{smallmatrix}\right]$ as

$$\begin{aligned} y_n &= \mathbf{g}(z)\,\mathbf{u}_n \\ &= \mathbf{g}(z)\,\mathcal{F}(z) \begin{bmatrix} \mathbf{a}_n \\ \alpha_n \end{bmatrix} \\ &= \mathbf{s}(z) \begin{bmatrix} \mathbf{a}_n \\ \alpha_n \end{bmatrix} \\ &= \sum_{k=0}^{\infty} \mathbf{s}_k \begin{bmatrix} \mathbf{a}_{n-k} \\ \alpha_{n-k} \end{bmatrix} \end{aligned}$$

Each term $\mathbf{s}_k$ is a row vector having as many entries as there are sources (both Gaussian and non-Gaussian). The impulse response sequence $\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$, …, is the *combined* (channel-equalizer) impulse response. It may be expressed as the convolution of the channel and equalizer impulse response sequences $\{\mathcal{F}_k\}_{k=0}^{\infty}$ and $\{\mathbf{g}_k\}_{k=0}^{M}$, respectively:

$$\underbrace{\begin{bmatrix} \mathbf{s}_0^T \\ \mathbf{s}_1^T \\ \mathbf{s}_2^T \\ \vdots \\ \mathbf{s}_M^T \\ \mathbf{s}_{M+1}^T \\ \vdots \end{bmatrix}}_{\triangleq\,\mathbf{s}} = \underbrace{\begin{bmatrix} \mathcal{F}_0^T & \bigcirc & \bigcirc & \cdots & \bigcirc \\ \mathcal{F}_1^T & \mathcal{F}_0^T & \bigcirc & \cdots & \bigcirc \\ \mathcal{F}_0^T & \mathcal{F}_1^T & \mathcal{F}_0^T & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \bigcirc \\ \mathcal{F}_M^T & \mathcal{F}_{M-1}^T & \cdots & \mathcal{F}_1^T & \mathcal{F}_0^T \\ \mathcal{F}_{M+1}^T & \mathcal{F}_M^T & \cdots & \mathcal{F}_2^T & \mathcal{F}_1^T \\ \vdots & \ddots & \ddots & \ddots & \vdots \end{bmatrix}}_{\triangleq\,\mathcal{F}} \underbrace{\begin{bmatrix} \mathbf{g}_0^T \\ \mathbf{g}_1^T \\ \mathbf{g}_2^T \\ \vdots \\ \mathbf{g}_M^T \end{bmatrix}}_{\triangleq\,\mathbf{G}}$$

We observe that irrespective of how the equalizer coefficients $\{\mathbf{g}_k\}$ are chosen, the combined response vector $\mathbf{s}$ is restricted to the column space (in ℓ_2) of the channel convolution matrix $\mathcal{F}$; as in [4], we call this linear subspace the set of attainable combined responses, denoted $\mathcal{S}_A$:

$$\mathcal{S}_A = \Big\{ \mathbf{s} : \mathbf{s} = \mathcal{F}\mathbf{G}, \text{ for some equalizer setting } \{\mathbf{g}_k\} \Big\}.$$

The projection operator from ℓ_2 to $\mathcal{S}_A$ is denoted

$$\mathcal{P}_A = \mathcal{F}(\mathcal{F}^T \mathcal{F})^{\sharp} \mathcal{F}^T,$$

where the superscript $\sharp$ denotes (pseudo-) inversion. A given $\mathbf{s}$ is then attainable ($\mathbf{s} \in \mathcal{S}_A$) if and only if $\mathcal{P}_A\,\mathbf{s} = \mathbf{s}$. If $\mathcal{P}_A = \mathbf{I}$ (the identity), then an arbitrary configuration of the combined response vector $\mathbf{s}$ is attainable; this is called the *sufficient order* case [6]. 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 called the *undermodeled* case. With nonzero noise, the transfer matrix $\mathcal{F}(z)$ has more inputs than outputs, so that undermodeling will prevail.

3. EQUALIZATION CRITERION

We consider the equalization criterion [1], [2], [5], [6]

$$J_{2p} = \frac{\text{cum}_{2p}(y_n)}{[\text{cum}_2(y_n)]^p}, \qquad p = 2, 3, 4, \ldots,$$

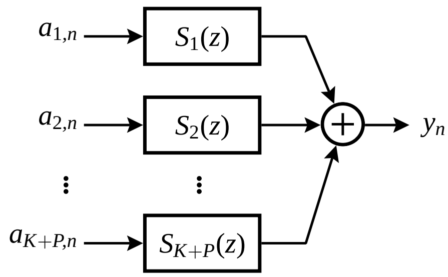

Figure 2: Equivalent multi-input/single-output representation, in terms of the combined transfer functions $S_i(z)$, one for each source.

in which $\text{cum}_{2p}(\cdot)$ denotes the cumulant of order $2p$ of the argument [7]; one seeks to maximize $|J_{2p}|$.

To simplify some developments to follow, let us invoke some notational "tweaking". Suppose there are K sources $\{a_{i,n}\}$, $i = 1, 2, \ldots, K$, and rename the Gaussian innovation $\{\alpha_{1,n}, \ldots, , \alpha_{P,n}\}$ as $\{a_{K+1,n}, \ldots, a_{K+P,n}\}$. This allows us to write the equalizer output as

$$y_n = \sum_{i=1}^{K+P} \sum_{k=0}^{\infty} s_{i,k}\, a_{i,n-k}, \qquad \begin{cases} i &= \text{source index;} \\ k &= \text{time delay index;} \end{cases}$$
$$= \sum_{i=1}^{K+P} S_i(z)\, a_{i,n}$$

in which $S_i(z) = \sum_{k=0}^{\infty} s_{i,k} z^k$ is the transfer function mapping the i^{th} source a_i to the equalizer output, as in Fig. 2.

As y_n is a weighted sum of independent random variables, and one may show that [7]

$$\text{cum}_{2p}(y_n) = \sum_{i=1}^{K+P} \underbrace{\text{cum}_{2p}(a_{i,n})}_{\triangleq\, \gamma_i} \sum_{k=0}^{\infty} s_{i,k}^{2p}.$$

Note that if the noise innovation is indeed Gaussian, then the corresponding cumulants vanish ($\gamma_{K+1} = \cdots = \gamma_{K+P} = 0$). The development to follow, however, will allow these cumulants to differ from zero.

Writing the equalization criterion in terms of the combined response $\mathbf{s}$ and the source cumulants $\{\gamma_i\}$ gives

$$J_{2p}(\mathbf{s}, \{\gamma_i\}) = \frac{\text{cum}_{2p}(y_n)}{[\text{cum}_2(y_n)]^p} = \frac{\displaystyle\sum_{i=1}^{K+P} \gamma_i \sum_{k=0}^{\infty} s_{i,k}^{2p}}{\left(\displaystyle\sum_{i=1}^{K+P} \sum_{k=0}^{\infty} s_{i,k}^2\right)^p} \tag{1}$$

If γ^- and γ^+ denote the most negative and most positive source cumulants, then $\gamma^- \le J_{2p} \le \gamma^+$. Moreover, J_{2p} is radially invariant:

$$J_{2p}(\beta \mathbf{s}, \{\gamma_i\}) = J_{2p}(\mathbf{s}, \{\gamma_i\}), \qquad \text{for all scalars } \beta \ne 0.$$

We may thus scale $\mathbf{s}$ to unit ℓ_2 norm: $\|\mathbf{s}\|_2 = 1$.

We now allow $\mathbf{s}$ to vary over $\mathcal{S}_A$, and we seek the stationary points of J_{2p} over $\mathcal{S}_A$. To simplify some notations to follow, we will write $J_{2p}(\mathbf{s})$ for $J_{2p}(\mathbf{s}, \{\gamma_i\})$.

The directional derivative of the function J_{2p} at $\mathbf{s}$, with respect to a directional vector $\mathbf{r}$, is given by [8]

$$J'_{2p}(\mathbf{s}; \mathbf{r}) \triangleq \lim_{t \to 0} \frac{J_{2p}(\mathbf{s} + t\mathbf{r}) - J_{2p}(\mathbf{s})}{t}$$

where t is a positive real scalar that tends to zero. If the function J_{2p} is differentiable at $\mathbf{s}$, then one has [8]

$$J'_{2p}(\mathbf{s}; \mathbf{r}) = \langle \nabla J_{2p}(\mathbf{s}), \mathbf{r} \rangle$$

where $\nabla J_{2p}(\mathbf{s})$ is the gradient of J_{2p} at $\mathbf{s}$, and where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in ℓ_2: $\langle \mathbf{s}, \mathbf{r} \rangle = \sum_k s_k r_k$.

Now, since the vector $\mathbf{s} + t\mathbf{r}$ spans $\mathcal{S}_A$ as $\mathbf{r}$ spans $\mathcal{S}_A$, a given vector $\mathbf{s} \in \mathcal{S}_A$ is a stationary point of J_{2p} over $\mathcal{S}_A$ if and only if the directional derivative of J_{2p} at $\mathbf{s}$ vanishes for all directional vectors $\mathbf{r}$ in $\mathcal{S}_A$:

$$J'_{2p}(\mathbf{s}; \mathbf{r}) = \langle \nabla J_{2p}(\mathbf{s}), \mathbf{r} \rangle = 0, \qquad \text{for all } \mathbf{r} \in \mathcal{S}_A. \tag{2}$$

Let now $\mathbf{v}$ be an arbitrary vector in ℓ_2. Its projection $\mathcal{P}_A \mathbf{v}$ lies in $\mathcal{S}_A$, and as $\mathbf{v}$ spans ℓ_2, the projection $\mathcal{P}_A \mathbf{v}$ spans $\mathcal{S}_A$. As such, the orthogonality relation (2), characterizing any stationary point $\mathbf{s}$, is equivalent to

$$\text{for all } \mathbf{v} \in \ell_2, \quad 0 = \langle \nabla J_{2p}(\mathbf{s}), \mathcal{P}_A \mathbf{v} \rangle = \langle \mathcal{P}_A \nabla J_{2p}(\mathbf{s}), \mathbf{v} \rangle,$$

in which the second equality follows from symmetry of $\mathcal{P}_A$. This latter expression is equivalent to

$$\mathcal{P}_A \nabla J_{2p}(\mathbf{s}) = \mathbf{0}.$$

It now suffices to calculate the gradient using (1), but recalling our scaling assumption $\|\mathbf{s}\|_2 = 1$:

$$[\nabla J_{2p}(\mathbf{s})]_{i,k} = \frac{\partial J_{2p}(\mathbf{s})}{\partial s_{i,k}} = 2p \left(\gamma_i s_{i,k}^{2p-1} - J_{2p}(\mathbf{s})\, s_{i,k} \right).$$

We may stack these scalars into a vector according to

$$\begin{aligned} \nabla J_{2p}(\mathbf{s}) &= 2p \left(\begin{bmatrix} \gamma_1 s_{1,0}^{2p-1} \\ \vdots \\ \gamma_{K+P} s_{K+P,0}^{2p-1} \\ \gamma_1 s_{1,1}^{2p-1} \\ \vdots \\ \gamma_{K+P} s_{K+P,1}^{2p-1} \\ \vdots \end{bmatrix} - J_{2p}(\mathbf{s}) \begin{bmatrix} s_{1,0} \\ \vdots \\ s_{K+P,0} \\ s_{1,1} \\ \vdots \\ s_{K+P,1} \\ \vdots \end{bmatrix} \right) \\ &= 2p \left(\mathcal{C} \mathbf{s}^{\odot(2p-1)} - J_{2p}(\mathbf{s})\, \mathbf{s} \right) \end{aligned}$$

in which:

- the vector $\mathbf{s}^{\odot(2p-1)}$ denotes the Hadamard power of order $2p-1$:

$$[\mathbf{s}^{\odot(2p-1)}]_{i,k} = (s_{i,k})^{2p-1};$$

- $\mathcal{C}$ is a diagonal matrix containing copies of the source cumulants of order $2p$:

$$\mathcal{C} = \begin{bmatrix} \mathbf{C} & & & \bigcirc \\ & \mathbf{C} & & \\ & & \mathbf{C} & \\ \bigcirc & & & \ddots \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \gamma_1 & & \bigcirc \\ & \ddots & \\ \bigcirc & & \gamma_{K+P} \end{bmatrix}.$$

Now, since $\mathbf{s} \in \mathcal{S}_A$, we have $\mathcal{P}_A \mathbf{s} = \mathbf{s}$, and the condition $\mathcal{P}_A \nabla J_{2p}(\mathbf{s}) = \mathbf{0}$ can be rephrased as:

Theorem 1 *A candidate $\mathbf{s} \in \mathcal{S}_A$ (scaled to unit ℓ_2 norm) is a stationary point of $J_{2p}(\mathbf{s})$ over $\mathcal{S}_A$ if and only if*

$$\mathcal{P}_A(\mathcal{C}\,\mathbf{s}^{\odot(2p-1)}) = J_{2p}(\mathbf{s})\,\mathbf{s}.$$

This generalizes the characterization of stationary points obtained previously in the mono-source case [6].

If we set $\mathbf{r} = \beta\mathbf{s}$, where β is any nonzero scalar, then we find that

$$\mathcal{P}_A(\mathcal{C}\mathbf{r}^{\odot(2p-1)}) = J_{2p}(\mathbf{s})\,\beta^{2p-1}\mathbf{s} = [J_{2p}(\mathbf{r})\,\beta^{2p-2}]\,\mathbf{r},$$

so that $\mathbf{r} = \beta\mathbf{s}$ satisfies an equation of the same structure as in Theorem 1, and hence is also a stationary point. This is as expected, since $J_{2p}(\mathbf{r}) = J_{2p}(\beta\mathbf{s}) = J_{2p}(\mathbf{s})$.

If we return to the sufficient order case ($\mathcal{P}_A = \mathbf{I}$), the characterization of stationary points then simplifies to $\mathcal{C}\mathbf{s}^{\odot(2p-1)} = J_{2p}(\mathbf{s})\,\mathbf{s}$, which reads componentwise as

$$s_{i,k}\Big(\gamma_i\,(s_{i,k})^{2p-2} - J_{2p}(\mathbf{s})\Big) = 0, \qquad \text{for all } i,k.$$

This says that all nonzero terms, once scaled according to the source cumulants, share a common amplitude. If two or more terms are nonzero, then the stationary point may be shown to be a saddle point (as in [5]), and hence not a candidate convergent point for an algorithm which seeks an extremum of J_{2p}.

4. A GRADIENT SEARCH PROCEDURE

Let $\mathbf{s}_{(0)}$ be an initial attainable setting in the combined response space, scaled to unit ℓ_2 norm. We consider a gradient search procedure of the form

$$\begin{aligned} \mathbf{v}_{(k+1)} &= \mathbf{s}_{(k)} \pm \mu_k \mathcal{P}_A \nabla J_{2p}(\mathbf{s}_{(k)}) \\ &= \mathbf{s}_{(k)} \pm \mu_k\, 2p\,\mathcal{P}_A\Big(\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)} - J_{2p}(\mathbf{s}_{(k)})\,\mathbf{s}_{(k)}\Big) \\ \mathbf{s}_{(k+1)} &= \mathbf{v}_{(k+1)} \big/ \|\mathbf{v}_{(k+1)}\|_2 \end{aligned}$$

in which the ℓ_2 normalization of the second line is introduced because J_{2p} is radially invariant; the sign in front of μ_k is chosen according to whether the algorithm is to ascend ($\pm\mu_k \to +\mu_k$) or descend ($\pm\mu_k \to -\mu_k$). We observe that with the particular step-size choice

$$\mu_k = \frac{1}{2p\,|J_{2p}(\mathbf{s}_{(k)})|} \tag{3}$$

the gradient algorithm simplifies to

$$\begin{aligned} \mathbf{v}_{(k+1)} &= \pm\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)} \\ \mathbf{s}_{(k+1)} &= \mathbf{v}_{(k+1)} \big/ \|\mathbf{v}_{(k+1)}\|_2 \end{aligned}$$

which is recognized as the super-exponential algorithm; see [3, 9] for implementation aspects.

The problem considered is to deduce the range for the step-size parameter μ_k which ensures convergence of the sequence $\{\mathbf{s}_{(k)}\}$ to an extremum of J_{2p}. Two cases may be distinguished. For the chosen order $2p$, the source cumulants may all be nonpositive: $\gamma_i \leq 0$ for all i (resp., nonnegative: $\gamma_i \geq 0$ for all i), or they may be mixed, i.e., some positive and others negative. The former case will be assumed in the developments to follow, as it is of greater practical interest. For example, with order $2p = 4$, most usable source constellations yield negative fourth-order cumulants. The following result extends that from [10] to the multi-source setting; for notational convenience we write $J_{2p}(k)$ for $J_{2p}(\mathbf{s}_{(k)})$.

Theorem 2 *Suppose all cumulants are nonnegative: $\gamma_i \geq 0$ (resp., all cumulants nonpositive: $\gamma_i \leq 0$). If $\mathbf{s}_{(k)}$ is not a stationary point of J_{2p}, the inequality $|J_{2p}(k+1)| > |J_{2p}(k)|$ holds whenever μ_k lies in the range*

$$0 < \mu_k < \frac{1}{p}\,\frac{|J_{2p}(k)|}{2|J_{2p}(k)|^2 - \|\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\|_2^2}.$$

Remark 1: We can also check that $|J_{2p}(k)| \leq \|\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\|_2$, since from the Cauchy-Schwarz inequality we have

$$|J_{2p}(k)| = \Big|\langle \mathbf{s}_{(k)}, \mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\rangle\Big| \leq \underbrace{\|\mathbf{s}_{(k)}\|_2}_{1} \cdot \|\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\|_2$$

with equality iff $\mathbf{s}_{(k)}$ and $\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}$ are colinear. As such, the upper bound for μ_k can be lower bounded as

$$\frac{1}{p}\,\frac{|J_{2p}(k)|}{2|J_{2p}(k)|^2 - \|\mathcal{P}_A\mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\|_2^2} \geq \frac{1}{p\,|J_{2p}(k)|} \geq \frac{1}{p\,\max_i |\gamma_i|}$$

since $|J_{2p}(k)| \leq \max_i |\gamma_i|$ at each iteration. Any fixed step-size in the range $0 < \mu < 1/(p \max_i |\gamma_i|)$ will thus ensure convergence of $|J_{2p}|$ to a local maximum. ⋄

The proof treats the nonnegative cumulant case, for which $0 \leq J_{2p} \leq \gamma^+$, since the nonpositive cumulant case is quite similar. Let $\mathbf{x}$ be a free vector in ℓ_2, and introduce the scalar-valued function

$$M_{2p}(\mathbf{x}) = \sum_{i=1}^{K+P} \gamma_i \sum_{k=0}^{\infty} x_{i,k}^{2p} \geq 0,$$

which assumes the same form as the numerator of J_{2p}. One can verify that this function is convex, i.e., for two choices $\mathbf{x}_1$ and $\mathbf{x}_2$ in ℓ_2,

$$M_{2p}\Big(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}_2\Big) \;\leq\; \lambda M_{2p}(\mathbf{x}_1) + (1-\lambda) M_{2p}(\mathbf{x}_2), \quad \text{for } 0 \leq \lambda \leq 1.$$

As $M_{2p}(\mathbf{x})$ is continuous throughout ℓ_2, we can introduce its gradient $\nabla M_{2p}(\mathbf{x})$ as

$$\nabla M_{2p}(\mathbf{x}) = \begin{bmatrix} \partial M_{2p}(\mathbf{x})/\partial x_{1,0} \\ \vdots \\ \partial M_{2p}(\mathbf{x})/\partial x_{K+P,0} \\ \partial M_{2p}(\mathbf{x})/\partial x_{1,1} \\ \vdots \\ \partial M_{2p}(\mathbf{x})/\partial x_{K+P,1} \\ \vdots \end{bmatrix} = 2p\mathcal{C}\mathbf{x}^{\odot(2p-1)}$$

Convexity of the function $M_{2p}(\mathbf{x})$ can be shown [8] to induce the inequality

$$M_{2p}(\mathbf{x} + \Delta\mathbf{x}) \geq M_{2p}(\mathbf{x}) + \langle \nabla M_{2p}(\mathbf{x}), \Delta\mathbf{x} \rangle,$$

for all $\mathbf{x}$ and $\Delta\mathbf{x}$ in ℓ_2 (and not just for $\Delta\mathbf{x}$ chosen "small").

This inequality thus applies to the particular choices $\mathbf{x} = \mathbf{s}_{(k)}$, and $\mathbf{x} + \Delta\mathbf{x} = \mathbf{s}_{(k+1)}$, giving

$$M_{2p}(\mathbf{s}_{(k+1)}) - M_{2p}(\mathbf{s}_{(k)}) \geq 2p \langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}, \mathbf{s}_{(k+1)} - \mathbf{s}_{(k)} \rangle.$$

Now since both $\mathbf{s}_{(k)}$ and $\mathbf{s}_{(k+1)}$ have unit ℓ_2 norm, we can check that

$$M_{2p}(\mathbf{s}_{(k)}) = \frac{M_{2p}(\mathbf{s}_{(k)})}{\|\mathbf{s}\|_2^{2p}} = J_{2p}(k),$$

and similarly, $M_{2p}(\mathbf{s}_{(k+1)}) = J_{2p}(k+1)$ whenever $\|\mathbf{s}_{(k+1)}\|_2 = 1$. Our inequality now reads as

$$J_{2p}(k+1) - J_{2p}(k) \geq 2p \langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}, \mathbf{s}_{(k+1)} - \mathbf{s}_{(k)} \rangle, \qquad (4)$$

and it suffices to deduce which values of μ_k render the right-hand side positive. This takes the form

$$\Big\langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}, \underbrace{\frac{\mathbf{s}_{(k)} + \mu_k \mathcal{P}_A \nabla J_{2p}(\mathbf{s}_{(k)})}{\|\mathbf{s}_{(k)} + \mu_k \mathcal{P}_A \nabla J_{2p}(\mathbf{s}_{(k)})\|}}_{\mathbf{s}_{(k+1)}} \Big\rangle > \underbrace{\langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}, \mathbf{s}_{(k)} \rangle}_{J_{2p}(k)}$$

Solving for μ_k compatible with this inequality leads to

$$\begin{aligned} 0 < \mu_k &< \frac{2J_{2p}(k)\langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)} - J_{2p}(k)\,\mathbf{s}_{(k)}, \mathcal{P}_A \nabla J_{2p}(k)\rangle}{J_{2p}^2(k)\|\mathcal{P}_A \nabla J_{2p}(k)\|_2^2 - \langle \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}, \mathcal{P}_A \nabla J_{2p}(k)\rangle^2} \\ &= \frac{1}{p} \frac{J_{2p}(k)}{2[J_{2p}(k)]^2 - \|\mathcal{P}_A \mathcal{C}\mathbf{s}_{(k)}^{\odot(2p-1)}\|_2^2} \end{aligned}$$

as claimed. ◇

Remark 2: The right-hand side of (4), when positive, represents the worst-case increase in the function J_{2p} at each iteration. The value of μ_k which maximizes this worst-case increase is found as $\mu_k^{\text{opt}} = 1/[2p\,J_{2p}(k)]$. This is the step-size value giving rise to the super-exponential algorithm; cf. (3).

Remark 3: We have observed monotonic convergence using this step-size range for the mixed cumulant case as well, although a proof for this case remains open.

5. REFERENCES

[1] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. F. Findley, ed., Academic, New York, 1981; pp. 565–608.

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

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

[4] Y. Li and Z. Ding, "Global convergence of fractionally-spaced Godard (CMA) adaptive equalizers," *IEEE Trans. Signal Processing*, vol. 44, pp. 818–826, April 1996.

[5] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filtering criteria," *IEEE Trans. Signal Processing*, vol. 45, pp. 658–672, March 1997.

[6] P. A. Regalia and M. Mboup, "Undermodeled equalization: A characterization of stationary points for a family of blind criteria," *IEEE Trans. on Sig. Proc.*, vol. 47, March 1999.

[7] J. M. Mendel, "Tutorial on higher order statistics (spectra) in signal processing and system theory," *Proc. IEEE*, vol. 79, pp. 278–305, March 1991.

[8] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[9] K. L. Yeung and S. F. Yau, "A cumulant-based super-exponential algorithm for blind deconvolution of multi-input multi-output systems," *Signal Processing*, vol. 67, pp. 141–162, June 1998.

[10] M. Mboup and P. A. Regalia, "On the equivalence between the super-exponential algorithm and a gradient search method," *Proc. ICASSP-99*, Phoenix, AZ, March 1999.