ASYMPTOTIC ANALYSIS OF SCALE-INVARIANT COST FUNCTIONS FOR BLIND ADAPTIVE PROCESSING

Edgar H. Satorius and James J. Mulligan

Edgar H. Satorius, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
James J. Mulligan, U.S. Army CECOM IEWD, Vint Hill Farms, Warrenton, VA

Abstract In this paper, we provide an asymptotic precision analysis of blind adaptive filter coefficients derived from a wide class of scale-invariant cost functions implicitly implemented in a batch processing mode. This analysis is based on a first-order Taylor expansion of the cost functions in the vicinity of their maxima and represents an extension of Donoho's classic asymptotic precision analysis [1] to the complex case. Through this analysis we have a means of discriminating among different nonlinear cost functions in the sense of yielding more precise estimates with N finite samples. We also find that cost functions based on very large order statistics tend to have highly desirable convergence properties over a wide range of constellations.

1. Introduction

Blind adaptive processing provides for the recovery of unknown signals via a finite dimensional, linear projection of a channel output data vector. In component form we have

$$z_n = \sum_{i=0}^{L-1} w_i \, y_{n-i} \ , \tag{1}$$

where w_i denote the L projection (blind adaptive filter) coefficients and the z_n are the complex output samples from the blind adaptive filter (equalizer). The y_n denote complex samples from the unknown channel which can in turn be expressed as a convolution of the sampled channel impulse response, f_i, with an unknown sequence of independent and identically distributed (iid) source symbols, a_n

$$y_n = \sum_i f_i \, a_{n-i} \ . \tag{2}$$

The blind adaptive filter coefficients are derived using only the available channel output data, without knowledge of either the transmitted signal waveform or the linear channel.

Approaches to blind adaptive processing can be broadly categorized into four different classes: (a) property restoral techniques wherein the projection coefficients are derived via the maximization (or minimization) of various cost criteria [1-5]; (b) direct channel coefficient estimation techniques that utilize higher-order statistics (i.e., cumulants) [6]; (c) maximum likelihood estimation techniques wherein both the channel and the unknown signal are estimated simultaneously [7] and (d) direct channel coefficient estimation techniques that exploit cyclostationary properties of the transmitted waveform [8]. The first class of techniques, which is the focus of this paper, is perhaps the broadest since it can incorporate a wide range of cost criteria. Furthermore, this class of techniques is the oldest and consequently has been studied extensively. The majority of this analysis has dealt with the convergence of time-recursive, gradient-based algorithms for achieving the maximum/minimum of the cost functions, such as the Godard (or constant modulus) algorithm [3-4,9].

In this paper, we provide a much different type of analysis of the first class of techniques. In particular, we provide an asymptotic precision analysis of the blind adaptive filter coefficients derived from a wide class of scale-invariant cost functions implicitly implemented in a batch processing mode. In doing this, we

1058-6393/95 $4.00 © 1995 IEEE

examine the behavior of the cost functions in the vicinity of their maxima. These cost functions, O_N, map N time samples of the blind adaptive filter output, z_n $(1 \le n \le N)$, into a real number which is maximized (or minimized) over the space of projection coefficients, w_ℓ $(1 \le \ell \le L)$. Since we generally don't care about arbitrary gains introduced by the blind adaptive filter, these cost functions are scale-invariant, i.e., $O_N(\alpha z_n) = O_N(z_n)$, where α is an arbitrary scale factor. Also, as N increases, $O_N \to O$ which is a function of z's probability distribution. For example, consider sample kurtosis: $O_N(z_n) = \left[\frac{1}{N}\sum_{n=1}^{N} z_n^{\,4}\right] / \left[\frac{1}{N}\sum_{n=1}^{N} z_n^{\,2}\right]^2$, which depends on all N samples of z_n. However, as $N \to \infty$, $O_N \to O \equiv E[z^4] / \{E[z^2]\}^2$, which only depends on the distribution of z (through its moments). Such cost functions encompass Godard's algorithm [3] as well as the cumulant-based cost functions that have recently been proposed for property restoral [2].

In the remainder of this paper, we present a summary of our asymptotic analysis in Section 2 followed by some examples in Section 3.

2. Asymptotic Performance Analysis

We begin by noting that the complex gradient of $O_N(z_n = \sum_{i=0}^{L-1} w_i\, y_{n-i})$ with respect to $\underline{w} \equiv \underline{w}_x + j\underline{w}_y$ (the complex vector of adaptive filter coefficients) can be written as

$$\frac{\partial O_N}{\partial w_k} \equiv \frac{\partial O_N}{\partial w_{xk}} + j\frac{\partial O_N}{\partial w_{yk}} = \sum_i \frac{\partial O_N}{\partial z_i} y_{i-k}^* , \quad (3)$$

where $*$ denotes complex conjugate and: $\frac{\partial O_N}{\partial z_i} \equiv \frac{\partial O_N}{\partial z_{xi}} + j\frac{\partial O_N}{\partial z_{yi}}$. Denoting $\Psi_N \equiv N\frac{\partial O_N}{\partial z_i}$, we assume for the cost functions of interest that as $N \to \infty$, $\Psi_N \to \Psi(z_i)$ which is only a function of z_i's probability distribution. As it turns out most candidate cost functions have this property. Examples of Ψ will be provided below.

Having derived an expression for the gradient (3) and established a simple condition on its asymptotic limit ($\Psi_N \to \Psi$), we now develop a first-order Taylor expansion of O_N in the vicinity of its maximum. Specifically, we denote the maximizer of O_N by $\hat{\underline{w}} \equiv \hat{\underline{w}}_x + j\hat{\underline{w}}_y$ and the exact equalizer solution by $\underline{w}^o$, i.e., $\sum_{i=0}^{L-1} w_i^o\, y_{n-i} = a_{n-n_0}$ where n_0 is an arbitrary integer (and so we are ignoring truncation effects). Further, we assume N is sufficiently large so that $\hat{\underline{w}} \approx \underline{w}^o$, in which case we have

$$\begin{aligned} 0 &\approx \underline{g}_x(\underline{w}_x^o, \underline{w}_y^o) + \left\{\nabla_{\underline{w}_x} \underline{g}_x\right\} \bullet (\hat{\underline{w}}_x - \underline{w}_x^o) \\ &\quad + \left\{\nabla_{\underline{w}_y} \underline{g}_x\right\} \bullet (\hat{\underline{w}}_y - \underline{w}_y^o) \\ 0 &\approx \underline{g}_y(\underline{w}_x^o, \underline{w}_y^o) + \left\{\nabla_{\underline{w}_x} \underline{g}_y\right\} \bullet (\hat{\underline{w}}_x - \underline{w}_x^o) \\ &\quad + \left\{\nabla_{\underline{w}_y} \underline{g}_y\right\} \bullet (\hat{\underline{w}}_y - \underline{w}_y^o) , \end{aligned} \quad (4)$$

where $\underline{g}_x \equiv \nabla_{\underline{w}_x} O_N$ and $\underline{g}_y \equiv \nabla_{\underline{w}_y} O_N$. From (4) we can derive a large N expression for the desired complex coefficient covariance matrix

$$\underline{\underline{S}} \equiv E\left[(\hat{\underline{w}} - \underline{w}^o)(\hat{\underline{w}} - \underline{w}^o)^H\right] , \quad (5)$$

in terms of the Hessian matrix

$$\underline{\underline{H}} \equiv \begin{bmatrix} \underline{\underline{H}}_{xx} & \underline{\underline{H}}_{xy} \\ \underline{\underline{H}}_{yx} & \underline{\underline{H}}_{yy} \end{bmatrix} , \quad (6)$$

where $\underline{\underline{H}}_{xx} \equiv \nabla_{\underline{w}_x} \underline{g}_x$, $\underline{\underline{H}}_{xy} \equiv \nabla_{\underline{w}_y} \underline{g}_x$, $\underline{\underline{H}}_{yx} \equiv \nabla_{\underline{w}_x} \underline{g}_y$ and $\underline{\underline{H}}_{yy} \equiv \nabla_{\underline{w}_y} \underline{g}_y$.

In [1], a set of conditions are provided on Ψ which insure that $\hat{\underline{w}} \to \underline{w}^o$ as $N \to \infty$. Most notably is the orthogonality condition, which for the complex case is given by

$$E\left[\Psi(a)a^*\right]=0, \tag{7a}$$

where again, a_n denote the input iid symbols. In addition, we impose the conditions (in analogy with [1])

$$E[\Psi(a)]=0 \text{ and } E[\Psi(a)a]=0. \tag{7b}$$

These latter conditions are valid for a wide class of symmetric constellations and cost functions.

Before presenting a large N expression for $\underline{\underline{S}}$, we first note that the asymptotic analysis of the Hessian matrix, $\underline{\underline{H}}$, simplifies considerably under a certain large L approximation which relates to the covariance matrix of the channel output, y_n, as well as Ψ. An example is: $Trace\left(\underline{\underline{R}}_{y_x y_x}\right) >> Trace\left(\underline{\underline{R}}_1\right)$ where

$$\begin{aligned}
\underline{\underline{R}}_{y_x y_x}\Big|_{kj} &\equiv E[y_{x,i-k}y_{x,i-j}] \\
\underline{\underline{R}}_1\Big|_{kj} &\equiv E[\Re(f_{-k}a_i)\Re(f_{-j}a_i)] \\
&\quad -\left\{E[\frac{\partial\Psi_x(a_i)}{\partial a_x}]\right\}^{-1} \bullet \\
&\quad E[\frac{\partial\Psi_x(a_i)}{\partial a_x}\Re(f_{-k}a_i)\Re(f_{-j}a_i)]
\end{aligned} \tag{8}$$

($\Re(\bullet)$ denotes the real part and Ψ_x denotes the real part of Ψ). Similar conditions arise for the other parts of the channel output covariance matrix all of which are analogous to the large L approximation established in [1] for the real case. Yet another simplification arises, which does not have a counterpart in the real case, if

$$\begin{aligned}
&E\left[\frac{\partial\Psi_x(a_i)}{\partial a_x}\right]=E\left[\frac{\partial\Psi_y(a_i)}{\partial a_y}\right] \\
&and\ E\left[\frac{\partial\Psi_x(a_i)}{\partial a_y}\right]=-E\left[\frac{\partial\Psi_y(a_i)}{\partial a_x}\right],
\end{aligned} \tag{9}$$

which can be thought of as averaged Cauchy-Riemann equations, i.e., we are requiring that Ψ be, on the average, analytic. This assumption too is valid over a wide class of symmetric constellations and cost functions.

Under the various conditions embodied in (7)-(9), we have the following simple expression for $\underline{\underline{S}}$, which is completely analogous to that derived in [1] (details provided in [10], to be published)

$$\begin{aligned}
\underline{\underline{S}} &= E\left[(\underline{\hat{w}}-\underline{w})(\underline{\hat{w}}-\underline{w})^H\right] \\
&\approx \frac{1}{N}\frac{E\left[|\Psi|^2\right]}{\left|E\left[\frac{\partial\Psi_x(a_i)}{\partial a}\right]\right|^2 E\left[|a|^2\right]}\underline{\underline{R}}_{f^*f}^{-1} \\
&\equiv \frac{A(\Psi,a)}{N}\underline{\underline{R}}_{f^*f}^{-1},
\end{aligned} \tag{10}$$

where $\underline{\underline{R}}_{f^*f}\Big|_{kj} \equiv \sum_\ell f_\ell^* f_{\ell+k-j}$. From (10) and under a large L approximation, we can derive a corresponding asymptotic expression for the expected intersymbol interference (ISI) near convergence (again ignoring finite truncation effects)

$$\overline{ISI} \equiv E\left\{\frac{\sum_{n\neq 0}|\hat{c}_n|^2}{|\hat{c}_0|^2}\right\} \approx \frac{L-1}{N}A(\Psi,a), \tag{11}$$

where $\hat{c}_n \equiv \sum_{i=0}^{L-1}\hat{w}_i f_{n-i}$ is the combined channel/equalizer impulse response.

The asymptotic complex coefficient covariance matrix (10) goes to zero as N^{-1} and furthermore factors into two parts, one of which depends only on the channel (i.e., $\underline{\underline{R}}_{f^*f}$) and the other only upon the cost function as well as the underlying random process (i.e., $A(\Psi,a)$). Thus $A(\Psi,a)$ can be used as an asymptotic figure-of-merit for comparing the performance of different cost functions.

3. Examples

As an interesting example, consider first the magnitude of the 4-th order standardized cumulant [2]

$$O_N(z) \equiv \left| \frac{\frac{1}{N}\sum_{k=1}^{N}|z_k|^4}{\left\{\frac{1}{N}\sum_{k=1}^{N}|z_k|^2\right\}^2} - 2 - \frac{\left|\frac{1}{N}\sum_{k=1}^{N}z_k^2\right|^2}{\left\{\frac{1}{N}\sum_{k=1}^{N}|z_k|^2\right\}^2} \right| \quad (12)$$

for which

$$\Psi(a) = A_0\{E\left[|a|^2\right](a^2 - E\left[a^2\right])a^* - (E\left[|a|^4\right] - \left|E\left[a^2\right]\right|^2)a\}, \quad (13)$$

and A_0 is an irrelevant constant. Assuming symmetric circular or square constellations, a simple calculation reveals that (7) and (9) are satisfied and we have

$$A(\Psi,a) = \frac{E\left[\left\{|a|^2 - \gamma\right\}^2 |a|^2\right]}{\left\{E\left[2a_x^2 + |a|^2 - \gamma\right]\right\}^2 E\left[|a|^2\right]}, \quad (14)$$

where $\gamma \equiv E\left[|a|^4\right] / E\left[|a|^2\right]$. Note that $A(\Psi,a)$ vanishes for constant modulus constellations ($|a| =$ constant) which implies that $\underline{\underline{S}}$ (as well as $\overline{ISI}$) must converge faster than N^{-1} as $N \to \infty$, a truly remarkable result.

A second interesting example arises from a uniformly most powerful (UMP) scale invariant hypothesis test between factored generalized Gaussian and circular Gaussian distributions [5]

$$O_N(z) = \frac{\left\{\frac{1}{N}\sum_{k=1}^{N}|z_k|^2\right\}^{1/2}}{\left\{\frac{1}{N}\sum_{k=1}^{N}\left(|z_{xk}|^s + |z_{yk}|^s\right)\right\}^{1/s}}, \quad (15)$$

where s is a real parameter, $s > 2$. For this example

$$\Psi_x(a) = A_0\left\{1 - K|a_x|^{s-2}\right\}a_x$$
$$and\ \Psi_y(a) = A_0\left\{1 - K|a_y|^{s-2}\right\}a_y, \quad (16)$$

where again A_0 is an irrelevant constant and $K = E\left[|a|^2\right] / E\left[|a_x|^s + |a_y|^s\right]$. As in the previous example, a simple calculation based on symmetric circular or square constellations shows that (7) and (9) are satisfied and we have

$$A(\Psi,a) = \frac{\gamma_1 E\left\{|a_x|^{2s-2} + |a_y|^{2s-2}\right\} - 1}{\left\{1 - \gamma_2(s-1)E\left[|a_x|^{s-2}\right]\right\}^2}, \quad (17)$$

where $\gamma_1 \equiv E\left[|a|^2\right] / \left\{E\left[|a_x|^s + |a_y|^s\right]\right\}^2$ and $\gamma_2 \equiv E\left[|a|^2\right] / E\left[|a_x|^s + |a_y|^s\right]$.

Plots of $A(\Psi,a)$ versus s, corresponding to 16-, 64-, 256-state QAM and V.29 symbol constellations, are presented in Figure 1. Note that in all cases, $A(\Psi,a)$ goes to zero as $s \to \infty$, with the square QAM constellations generally yielding smaller A values than V.29 (which is not surprising since $O_N(z)$ arises from a hypothesis test between rectangular and circular distributions as noted above). This implies that $\underline{\underline{S}}$ (as well as $\overline{ISI}$) must converge faster than N^{-1} as $N \to \infty$, in the limit as $s \to \infty$, i.e., in the limit of very large order statistics. This is consistent with the observation in [5] that the large s cost function (15) results in very rapid convergence in terms of the number of samples, N, to achieve a given expected ISI level.

4. Conclusions

In this paper, we have presented an asymptotic performance analysis for scale-invariant cost functions that arise in blind equalization. As examples, we considered the standardized, 4-th order cumulant as well as a cost function specifically designed for recovering rectangular constellations. Our asymptotic performance analysis suggests that the former is particularly effective in recovering constant modulus constellations whereas the latter is indeed effective in recovering large-order QAM constellations (simulation results presented in [5] support this conclusion). To paraphrase Donoho [1], although asymptotic analysis, under the noiseless and zero equalizer filter truncation assumptions, does not tell the whole story it may be the only analytic result available for comparing the performance of highly nonlinear blind equalization techniques. Furthermore, it may be readily extended for analyzing blind array processing techniques wherein the inputs are not necessarily iid.

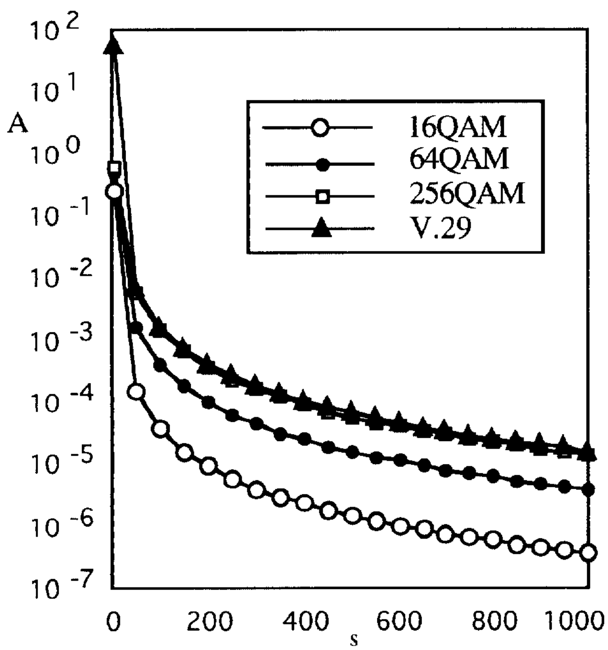

Figure 1. Asymptotic figure-of-merit for different symbol constellations.

References

1. Donoho, D., "On Minimum Entropy Deconvolution," Proc. of the 2-nd Applied Time Series Symposium, Tulsa, OK, 3-5 March 1980; also published in: **Applied Time Series Analysis II**, D. Finley (Ed.), (Academic Press, New York, 1981), pp. 565-608.

2. Shalvi, O., Weinstein, E., "Universal Methods for Blind Deconvolution," chap. 4 in **Blind Deconvolution**, S. Haykin, ed., Prentice Hall, 1994, pp. 121-180.

3. D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," **IEEE Trans. Comm.**, 1980, 28, pp. 1867-1875.

4. Treichler, J., Agee, B., "A New Approach to Multipath Correction of Constant Modulus Signals," **IEEE Trans. Acoust., Speech and Sig. Proc.**, 1983, 31, pp. 459-472.

5. Satorius, E., Mulligan, J., "An Alternative Methodology for Blind Equalization," **Digital Signal Processing: A Review Journal**, vol. 3, No. 3, pp. 199-209, July 1993.

6. D. Hatzinakos, C. L. Nikias, "Blind equalization using a tricepstrum-based algorithm," **IEEE Trans. Comm.**, 1991, 39, pp. 669-682.

7. Raheli, R., Polydoros, J., Tzou, C-K., "Per Survivor Processing," **Digital Signal Processing: A Review Journal**, vol. 3, No. 3, pp. 175-187, July 1993.

8. Tong, L., Xu, G., Kailath, T., "Blind Identification and Equalization Based on Second-Order Statistics – A Time Domain Approach," **IEEE Trans. Infor. Theory**, 1994, v.40, pp. 340-349.

9. Z. Ding, C. R. Johnson, R. A. Kennedy, "On the (non)existence of undesireable equilibria of Godard blind equalizers," **IEEE Trans. Sig. Proc.**, 1992, 40, pp. 2425-2432.

10. Satorius, E., Mulligan, J., "Notes on Calculating the Complex, Asymptotic Figure-of-Merit," Internal memorandum, 1994.