IV. Conclusions

This correspondence addresses the problem of locating the directions of arrival of a set of coherent and correlated signals using multiple sensor elements in arbitrary noise field. It is shown here that by making use of a symmetric array, it is possible to perform a preprocessing on the array output correlations to generate a set of new quantities that are functionally similar to the actual correlations in an incoherent environment. A Hermitian Toeplitz matrix generated from these "correlation-type" quantities is then used to estimate all actual arrival angles. Finally, a resolution threshold for two arbitrarily correlated, equipowered sources is derived in terms of their angular separation and this is compared to similar results in uncorrelated and coherent scenes [7], [8].

References

[1] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.

[2] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, 1018-1028, Aug. 1981.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectral Est. Workshop*, 1979, pp. 243-258.

[4] J. E. Evans, J. R. Johnson, and D. F. Sun, "Application of advanced signal processing techniques to angle of arrival estimation in ATC navigation and surveillance system," M.I.T. Lincoln Lab., Lexington, MA, Rep. 582, 1982.

[5] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 806-811, Aug. 1985.

[6] Y. Lee, "Direction finding from first order statistics and spatial spectrum estimation," Ph.D. dissertation, Polytechnic Univ., Brooklyn, NY, Jan. 1989.

[7] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331-341, Apr. 1985.

[8] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," submitted to *IEEE Trans. Acoust., Speech, Signal Processing*.

A Comparison of Two Quantized State Adaptive Algorithms

W. A. SETHARES and C. RICHARD JOHNSON, JR.

Abstract—**Quantized State (QS) adaptive algorithms reduce the numerical complexity and dynamic range requirements of Least Mean Squares (LMS) adaptation by replacing multiplications with shifts, bit comparisons, or table lookups. This correspondence provides a theoretical foundation with which to distinguish two primary QS algorithm forms and to predict which algorithm is most appropriate in a given context. An extended Lyapunov approach is used to derive a persistence of excitation (PE) condition which guarantees linear stability of the Quantized Error (QE) form. Averaging theory is then used to derive a persistence of excitation condition which guarantees exponential stability of the Quantized Regressor (QReg) form. Failure to meet this latter condition (which is not equivalent to the spectral richness PE condition for LMS) can result in exponential instability. The QE and QReg algorithms are then compared in terms of conditions for stability, convergence properties of the prediction and parameter errors, convergence rates, and steady state errors. The major results are gathered together in a table for easy reference.**

Manuscript received July 29, 1987; revised May 25, 1988. This work was supported by NSF Grant MIP-8608787.

W. A. Sethares was with the School of Electrical Engineering, Cornell University, Ithaca, NY. He is now with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706.

C. R. Johnson, Jr., is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

IEEE Log Number 8824486.

I. Introduction

Recent interest in Quantized State adaptive algorithms [1]-[8] has been sparked by the use of adaptive filters in high speed data communications [9] and in speech processing [10] where the high data rate requires a computationally simple algorithm with large dynamic range. Two candidate algorithms are the Quantized Regressor (QReg) and Quantized Error (QE) algorithms [11]. Since the numerical requirements of these two Quantized State algorithm forms are identical, they are direct competitors in any given adaptive application that requires fast processing of the data stream. The choice of which algorithm to use will depend on other factors such as the characteristics of the expected input sequence, the desired steady state performance, and the desired convergence properties. This correspondence compares these two algorithm forms extensively, and provides guidelines to help choose the proper algorithm for a given application.

In order to formally introduce the QS algorithms, it is helpful to have the LMS adaptive algorithm [12] firmly in mind. Consider the following adaptive filtering task. The output $y(k)$ of an FIR (tapped delay line) filter defined by

$$y(k+1) = \sum_{i=1}^{n} b_i u(k-i+1) = X_k^T \theta^* \qquad (1.1)$$

with $\theta^* = (b_1, b_2, \cdots, b_n)^T$ and with the regressor sequence $X_k = (u(k), u(k-1), \cdots, u(k-n+1))^T$ is to be estimated by

$$\hat{y}(k+1) = \sum_{i=1}^{n} \hat{b}_i(k)\, u(k-i+1) = X_k^T \hat{\theta}_k \qquad (1.2)$$

where $\hat{\theta}_k = (\hat{b}_1(k), \hat{b}_2(k), \cdots, \hat{b}_n(k))^T$, and each $u(k)$ is bounded by some value β. The error between the output $y(k)$ and the estimated output $\hat{y}(k)$, called the prediction error, is used to update the parameter estimates $\hat{\theta}_k$. The traditional LMS algorithm is

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu X_k [y(k) - \hat{y}(k)] \qquad (1.3)$$

where $\mu \ll 2/\beta^2$ is a small positive stepsize. Note that the prediction error can be expressed in terms of the parameter error $\tilde{\theta}_k = \theta^* - \hat{\theta}_k$ as

$$e_k = y(k) - \hat{y}(k) = X_k^T \theta^* - X_k^T \hat{\theta}_k = X_k^T \tilde{\theta}_k \qquad (1.4)$$

so that (1.3) can be rewritten

$$\tilde{\theta}_{k+1} = (I - \mu X_k X_k^T)\tilde{\theta}_k. \qquad (1.5)$$

The Quantized Regressor (QReg) algorithm quantizes each entry of the regressor vector, replacing (1.3) with

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu Q(X_k)[y(k) - \hat{y}(k)] \qquad (1.6)$$

while the Quantized Error (QE) algorithm quantizes the scalar error sequence and updates the parameters by

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu X_k Q([y(k) - \hat{y}(k)]). \qquad (1.7)$$

In (1.6) and (1.7), $Q(\cdot)$ represents some quantization function: a bounded, discrete valued, element by element, monotonic nondecreasing function which does not change the sign of the argument. The maximum quantized value is called the *range* of the quantizer, and is typically determined by the number of bits used in the quantization. Some typical quantization functions are rounding or truncating to the nearest $1/q$, rounding or truncating to the nearest power of two, and the sign (or signum) operator. For arbitrary quantizers, the QReg and QE updates are no simpler to implement than LMS, but for certain quantizers (such as the sign and power of two quantizers), the products $\mu Q(X_k)e_k$ and $\mu X_k Q(e_k)$ can be computed using simple bit shifts instead of floating point multiplications, offering a substantial simplification in numerical complexity.

Section II shows that the QE algorithm, like LMS, has an interpretation as a gradient procedure on an appropriate error surface in the parameter space. The QE algorithm minimizes a simple function of the error that is *not* the mean squared error. QReg does not appear to have such an interpretation. Neither algorithm is Lyapunov stable. Section III shows, instead, that the prediction error of QE is mean absolute convergent for any input sequence, while the behavior of the prediction error of QReg is heavily dependent on the characteristics of the input sequence.

Persistence of excitation conditions (conditions on the input sequence that cause robust convergence) are stated for both algorithms in Section IV. These conditions are *not* identical. The most striking observation is that the class of input signals for which QReg converges is *strictly smaller* than the class of input signals which cause QE and LMS to converge. This observation has implications beyond the immediate algorithms of interest, showing that persistence of excitation conditions are an algorithm dependent concept. Section V compares the convergence rates, which give a rough idea of the relative "degree of robustness" of the two algorithms.

II. Comparison: Gradient Procedure

LMS can be interpreted as a gradient procedure in the parameter error space with the instantaneous error function $J_{\text{LMS}}(e_k) = 1/2\, e_k^2$ [13]. At each timestep k, the parameter update for LMS moves in the direction of the negative gradient $dJ_{\text{LMS}}(e_k)/d\hat{\theta}_k = -X_k e_k$ a distance proportional to the stepsize μ. The LMS algorithm is one special case of the general gradient algorithm

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \mu \frac{dJ(e_k)}{d\hat{\theta}_k} \tag{2.1}$$

where J represents some nonnegative function that defines the instantaneous error surface.

Given a particular quantization function Q, the first step is to determine the appropriate function $J(\cdot)$ of the error which, when differentiated, gives rise to that quantizer. That is, find a function $J(\cdot)$ such that

$$\frac{dJ(e_k)}{de_k} = Q(e_k). \tag{2.2}$$

This is solved by integrating $Q(e_k)$ with respect to e_k. For example, suppose that $Q(\cdot)$ is the dual sign (DS) quantizer of [5]

$$Q(x) = \begin{cases} m \operatorname{sgn}(x) & 1 < |x| \\ \operatorname{sgn}(x) & |x| \le 1 \end{cases} \tag{2.3}$$

where $m > 1$. Then

$$J_{DS}(x) = \begin{cases} m|x| - (m-1) & |x| > 1 \\ |x| & |x| \le 1 \end{cases} \tag{2.4}$$

fulfills (2.2) where the constant $(m - 1)$ was added to make $J(\cdot)$ continuous. The gradient of J_{DS} with respect to the parameter estimate $\hat{\theta}_k$ is

$$\frac{dJ_{DS}(e_k)}{d\hat{\theta}_k} = \frac{dJ_{DS}(e_k)}{de_k}\frac{de_k}{d\hat{\theta}_k} = Q(e_k)\frac{de_k}{d\hat{\theta}_k} = -Q(e_k)X_k \tag{2.5}$$

except at the points of discontinuity of $Q(\cdot)$ where J_{DS} is nondifferentiable. The parameter update form corresponding to this approximation of the negative gradient direction of J_{DS} is [insert (2.5) into (2.1)] precisely the QE parameter update (1.7). This derivation could have been carried out for almost any quantizer. The error function for the special case $Q(\cdot) = \operatorname{sgn}(\cdot)$ was derived in [14].

The error function for the QE algorithm was easily derived, since the update term is in the form $Q(e)\,de$, which is easy to integrate. A similar analysis for the QReg algorithm, on the other hand, would require that the update term $Q(de)e$ be "integrated." Since it is not clear what such an integration means, the algorithm does not appear to be any type of gradient procedure.

III. Comparison: Lyapunov Stability and Convergence of the Prediction Error

Think of Lyapunov stability this way. There is a balloon inflated about the desired value θ^*. Some nonnegative function of the parameter error defines the radius of the balloon. Lyapunov stability means that air can leak out of the balloon (implying that the parameter error decreases) but none can ever enter. The sum squared parameter error $\tilde{\theta}_k^T\tilde{\theta}_k$ is a Lyapunov function for the LMS algorithm. It is a standard result [13] that $\tilde{\theta}_{k+1}^T\tilde{\theta}_{k+1} \le \tilde{\theta}_k^T\tilde{\theta}_k$ if $\mu X_k^T X_k < 2$, which can be guaranteed by choosing the stepsize μ sufficiently small.

Iterating the Lyapunov function t times shows that the prediction error is mean square convergent, i.e.,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} e_i^2 \to 0. \tag{3.1}$$

This does not show that e_i itself converges, but it does show that e_i must be small on the average. Does this analysis extend to the QS algorithms? Unfortunately, not always.

The QReg algorithm is not Lyapunov stable. The example of the next section shows that certain inputs inflate the error balloon indefinitely. Since $e_i = X_i^T\tilde{\theta}_i$ and $\tilde{\theta}_i^T\tilde{\theta}_i \to \infty$, e_i cannot be guaranteed convergent in any reasonable sense.

The QE algorithms are somewhat better behaved. The change in the sum squared parameter error $V_k = \tilde{\theta}_k^T\tilde{\theta}_k$ is

$$V_{k+1} - V_k = \tilde{\theta}_{k+1}^T\tilde{\theta}_{k+1} - \tilde{\theta}_k^T\tilde{\theta}_k. \tag{3.2}$$

Using (1.7) and the definition of $\tilde{\theta}_k$ as the difference between θ^* and $\hat{\theta}_k$, this can be rewritten

$$V_{k+1} - V_k = -2\mu X_k^T\tilde{\theta}_k Q(e_k) + \mu^2 X_k^T X_k Q^2(e_k). \tag{3.3}$$

Since $Q(\cdot)$ never reverses the sign of its argument, $X_k^T\tilde{\theta}_k Q(e_k)$ is nonnegative. Thus, $V_{k+1} \le V_k$ whenever

$$\mu X_k^T X_k Q^2(X_k^T\tilde{\theta}_k) \le 2X_k^T\tilde{\theta}_k Q(X_k^T\tilde{\theta}_k). \tag{3.4}$$

Paradoxically, when the algorithm functions well (by driving $\tilde{\theta}_k$ toward zero), it violates (3.4) and can increase the summed squared parameter error. Thus, convergence to zero is generically impossible. Metaphorically, when the balloon deflates smaller than some radius δ, air enters as easily as it leaves. Even when $\tilde{\theta}_k$ is large, however, (3.4) can be violated if $\tilde{\theta}_k$ happens to be nearly orthogonal to X_k, showing that the QE algorithm is not Lyapunov stable (at least with this candidate Lyapunov function).

Even if the behavior of $\tilde{\theta}_k$ cannot be explicitly guaranteed, the prediction error can be guaranteed to converge in a certain average sense.

Theorem 1: The prediction error e_k of the QE algorithm (1.7) is mean absolute convergent to a δ ball, that is,

$$\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} |e_i| \le \delta. \tag{3.5}$$

Proof: See the Appendix.

In (3.5), the size of δ (given in the proof) is dependent on the stepsize μ, the maximum value of the input, and the maximum and minimum (nonzero) values assumed by the quantization function. The theorem shows that, *on the average*, the prediction error of the QE algorithm remains smaller than δ.

IV. Comparison: Persistence of Excitation Conditions

The Lyapunov stability and prediction error convergence results of the previous section (which placed no constraints on the input/

regressor sequence) are not necessarily robust to measurement noises, slight mismodeling errors, or slow time variation of the desired parameterization. This was illustrated by the parameter drift examples of [15] and the numerical instabilities demonstrated in [19]. QE and QReg are susceptible to both problems. In the imagery of the previous section, disturbances pump small puffs of air into the balloon over several steps. If the algorithm does not have a "safety valve" to release this extra air, then the balloon may continue to inflate forever. What does such a safety valve look like?

For nonzero inputs, the LMS transition matrix $I - \mu X_k X_k^T$ from (1.5) is a contraction in one direction of $\mathbb{R}^n$, the direction of the nonzero eigenvector. If there is a fixed time span t such that the transition matrix contracts at least a fixed amount α in every direction every t steps, then the "average" transition matrix $[I - (\mu/t) \Sigma_{i=k}^{k+t-1} X_i X_i^T]$ is uniformly a contraction, and $\|\tilde{\theta}_k\|^2 = \tilde{\theta}_k^T \tilde{\theta}_k$ will converge exponentially to zero. Recall that a system with state Z_k is said to be exponentially stable about its equilibrium Z^* if there is a $\gamma \in (0, 1)$ and a finite N such that $\|Z^* - Z_k\| \leq N \|Z_0\| \gamma^k$ for every k. Equivalently, Z_k is said to converge exponentially to Z^*. Thus, a uniform contraction condition on the average transition matrix provides an exponential safety valve. With no disturbances, the balloon contracts to a single point as time evolves. With disturbances, the contraction is eventually halted by the continuous addition of tiny puffs of disturbance.

More formally, the regressor sequence X_k is said to be *persistently exciting for LMS* (PE for LMS) if there are positive α, β, and t such that

$$\beta I > M_t = \frac{1}{t} \sum_{i=k}^{k+t-1} X_i X_i^T > \alpha I \quad \text{for every } k. \tag{4.1}$$

The PE theorem for LMS is then as follows.

Theorem 2: Consider the LMS algorithm (1.3) and the associated error equation (1.5). If X_k is PE for LMS, then there is a μ^* such that $\|\tilde{\theta}_k\| \to 0$ exponentially for every $\mu \in (0, \mu^*)$.

Proof: This is proven in slightly modified form in [13], [16], and [19]. △△△

Similarly, the evolution of the parameter estimate error for the QReg algorithm is described by

$$\tilde{\theta}_{k+1} = \left(I - \mu Q(X_k) X_k^T\right) \tilde{\theta}_k. \tag{4.2}$$

While the one step transition matrix for LMS is $I - \mu X X^T$, the one step transition matrix for QReg is $I - \mu Q(X) X^T$. Since the stability properties of LMS depend on the average of the outer product of XX^T, it is reasonable to conjecture that the behavior of the Quantized Regressor algorithm should depend on the average value of $Q(X) X^T$.

The most straightforward results are obtained when the X_k sequence is t periodic. Let

$$M_t = \frac{1}{t} \sum_{i=1}^{t} Q(X_i) X_i^T. \tag{4.3}$$

If Re $\lambda_i(M_t) > 0$ for $i = 1, 2, \cdots, n$, then the regressor vector X_k will be said to be *persistently exciting for the QReg algorithm* (PE for QReg). The PE theorem for the QReg algorithm is as follows.

Theorem 3: Consider the quantized regressor algorithm (1.6) and the associated error equation (4.2). If X_k is a t-periodic sequence that is PE for QReg, then there exists a μ^* such that (4.2) is exponentially stable for every $\mu \in (0, \mu^*)$. If X_k is not persistently exciting, and Re $\lambda_i(M_t) < 0$ for some i, then there exists a μ^* such that for every $\mu \in (0, \mu^*)$, (4.2) is unstable.

Proof: A modest modification of theorem 1 of [17], replacing the sgn operator by the more general quantization operator Q. △△△

Theorem 3 can be used to find examples of input/regressor sequences that ensure exponential stability of the QReg algorithm, providing an exponential safety valve against small disturbances analogous to the exponential stability of LMS. The second part of the theorem shows that certain other regressor sequences cause unbounded parameter estimates.

Example: Let $Q(\cdot)$ round to the nearest integer with range $M > 2\eta + 1$, and consider the three periodic input $2\eta + 0.6$, $-\eta - 0.4$, $-\eta - 0.4$, $\cdots$, where $\eta > 1$ is an integer. The excitation matrix is

$$M_3 = \tfrac{1}{3} \sum_{i=1}^{3} Q(X_i) X_i^T = (3\eta^2 + 2\eta + \tfrac{1}{3}) I - \tfrac{1}{3}(5\eta^2 + 2\eta + 0.4) 1_{3\times 3}$$

where $1_{3\times 3}$ is the 3×3 matrix of all 1's. This has an eigenvalue at -0.066 and two eigenvalues at $3\eta^2 + 2\eta + 0.33$. Since the negative eigenvalue is independent of η, this input causes instability of the QReg algorithm. This same input, when applied to the LMS algorithm, is persistently exciting since M_3 calculated from (4.1) has all positive eigenvalues. Inputs that stabilize LMS may therefore destabilize QReg.

This example uses an integer valued quantizer. Is the divergence due to the coarseness of the quantizer? The good news is that given any regressor sequence X_k, it is always possible to quantize fine enough to stabilize the QReg algorithm. The bad news is that given any quantizer, *no matter how fine*, it is always possible to find a regressor X_k that causes M_t of (4.3) to have an eigenvalue with negative real part. This regressor will cause instability of the QReg algorithm. Details and caveats may be found in [18].

There is no way to write a simple transition matrix for the QE algorithm due to the discontinuity of $Q(\cdot)$. Hence, the averaging approach of theorems 2 and 3 is inapplicable. Instead, an extended Lyapunov approach is fruitful. Consider (3.4), and imagine that $\tilde{\theta}_k$ is some fixed vector ψ. Then, providing that the X_k regularly span $\mathbb{R}^n$, all the X_k cannot be orthogonal to ψ. Thus, some steps will deflate the balloon even if others tend to inflate it.

A system with state Z_k is said to be linearly convergent with window t to a ball δ about Z^* if there exist ω, δ, and $t > 0$ such that $\|Z^* - Z_{k+t}\| \leq \|Z^* - Z_k\| - \omega$ whenever $\|Z^* - Z_k\| > \delta$. This requires that the norm decrease over every t steps, but does not require that the norm decrease at every timestep.

Theorem 4: Consider the QE algorithm (1.7). If X_k is persistently excited as in (4.1) (PE for LMS), then there exist $\delta > 0$ and a μ^* such that $\hat{\theta}_k$ is linearly convergent to a δ ball about θ^* of magnitude $[0, \delta]$ whenever $\mu \in (0, \mu^*)$.

Proof: See the Appendix. △△△

Essentially, the PE condition for a given algorithm guarantees that the parameter estimates will converge to a small region about the desired parameterization, and that this convergence is "robust" to small nonidealities such as measurement noise, small nonlinearities, mismodeling, and slow parameter variation. The "robustness" of LMS and QReg is a consequence of the exponential stability, while the robustness of QE is a consequence of the linear stability to a ball.

Both LMS and QE are persistently excited by the same condition (4.1), while QReg is persistently excited when the inputs fulfill (4.3). It is not hard to show that any sequence that is PE for QReg is also PE for LMS (and hence for QE) but that the reverse implication is false. It is thus strictly more difficult to persistently excite the QReg algorithm than to persistently excite the QE algorithm. Said yet another way, the class of input/regressor sequences that make QE "work" is strictly larger than the class of signals that cause QReg to "work."

This is probably the most striking difference between the QE and the QReg algorithms. Successful application of the QReg algorithm requires that the inputs satisfy the PE for QReg condition, while successful application of the QE algorithm requires *no more knowledge* than that the LMS algorithm will "work." There is thus no loss in terms of stability for quantizing the error, while quantization of the regressor introduces the danger of instability.

In the ideal setting with no disturbances, the error systems for LMS and QReg can converge to zero, while the QE algorithm converges (at best) to a small ball about the desired parameterization. Thus, in "steady state," the parameter estimates of LMS and QReg actually achieve θ^* while the parameter estimates of QE rattle in a

δ neighborhood about θ^*. Thus, assuming that both are persistently excited, the ultimate performance of the QReg algorithm in terms of prediction and parameter errors is superior to the ultimate performance of the QE in the noise free case. This advantage is not as clear when nonidealities are introduced, but is likely retained for sufficiently small nonidealities.

The PE conditions for QReg and QE force a reevaluation of what is meant by persistence of excitation. No longer does PE mean just the summed outer product of the regressor sequence (since PE for QReg involves the quantization function), nor is it simply "the condition" that implies exponential asymptotic stability of the homogeneous error system (since PE for QE converges linearly to a nonzero ball). PE is thus an algorithm dependent concept and not an intrinsic property of a vector sequence.

V. Comparison: Convergence Rates

When the algorithms are persistently excited, the parameter estimates of LMS and QReg converge exponentially, while the parameter estimates of QE converge linearly. Is there a way to quantify the rates at which the algorithms converge?

First, consider LMS. From (1.5), the t step transition is

$$\tilde{\theta}_{k+t} = (I - \mu t M_t)\tilde{\theta}_k + o(\mu) \tag{5.1}$$

where M_t is defined as in (4.1). The largest eigenvalue of $(I - \mu t M_t)$ is $1 - \mu t\alpha$, and the smallest eigenvalue is $1 - \mu t\beta$, where α and β are defined in (4.1). When X_k is t periodic, there are unique eigenvectors V_α and V_β corresponding to the maximum and minimum eigenvalues. In the direction of V_α (the "slow" direction), the parameter estimate errors contract by a factor of approximately $1 - \mu t\alpha$ every t iterations. In the direction of V_β (the "fast" direction), the parameter estimate errors contract by a factor of approximately $1 - \mu t\beta$ every t iterations. When X_k is not periodic, the directions of maximum and minimum contraction vary with time. Since $(1 - \mu t\alpha)$ can be approximated by $(1 - \mu\alpha)^t + o(\mu)$, the actual convergence of the parameter estimate error vector is approximately bounded between the rates of the fastest and slowest eigendirections

$$\|\tilde{\theta}_0\|(1 - \mu\beta)^k + o(\mu) \leq \|\tilde{\theta}_k\| \leq \|\tilde{\theta}_0\|(1 - \mu\alpha)^k + o(\mu) \tag{5.2}$$

which becomes more accurate for smaller μ. In simulations (for instance, in [13]), the faster rate tends to dominate the initial iterations while the convergence of the tail is better approximated by the slower rate.

An analogous derivation can be carried out for the exponential convergence of the QReg algorithm. Iterating (4.2) t steps and averaging as in theorem 3 shows that (5.1) is valid where M_t is now defined as in (4.3). Using the appropriate α and β bounds on the real parts of the eigenvalues of M_t, the motion of the parameter estimate error can be approximated exactly as in (5.2). Again, the exponential contraction of the parameter estimate errors lies approximately in the range $(1 - \mu\alpha)$ and $(1 - \mu\beta)$. This is roughly equivalent to a range $e^{-\mu\alpha}$ to $e^{-\mu\beta}$, which shows that the *rate* is approximately proportional to $\mu\alpha$ in the slow direction, and proportional to $\mu\beta$ in the fast direction.

In contrast to the LMS and QReg algorithms, the QE algorithm converges linearly. For the simple case $Q(\cdot) = \text{sgn}(\cdot)$ (the signed error algorithm), the t step transition can be bounded by

$$\tilde{\theta}_{k+t}^T\tilde{\theta}_{k+t} - \tilde{\theta}_k^T\tilde{\theta}_k \leq -2\mu\delta\sqrt{\alpha t} + \mu^2 C = -\omega \tag{5.3}$$

as in the proof of theorem 4, equation (A.9), which is valid as long as $\|\tilde{\theta}_k\| > \delta$. In (5.3), δ is the size of the final convergent ball, μ is the stepsize, C is a constant proportional to $t^2\beta^2$, and α and β are the degrees of excitation from (4.1). Thus, the convergence rate over t iterations is proportional to $\mu\sqrt{\alpha t}$, and the average decrease of the squared norm at each timestep is $\mu\sqrt{\alpha}/\sqrt{t}$. These approximations assume that μ is small enough so that $\mu^2 C$ is inconsequential. Thus, doubling the degree of excitation increases the linear convergence by only $\sqrt{2}$. Doubling the stepsize doubles the rate, although it also doubles the size of the final convergent ball. An input that causes convergence at a rate ω over t steps converges at a rate 2ω if the same input is simply grouped differently and called a $2t$ periodic sequence. Yet the rate (5.3) implies that the $2t$ grouping would converge at a rate of $\sqrt{2}\omega$, showing that (5.3) is a conservative estimate of the convergence rate.

For large parameter estimate errors, the exponential rate of LMS and QReg offers rapid convergence while for small errors, the linear rate of QE is faster than the exponential. This can be seen graphically from the gradient error function. The slope of the $J_{\text{LMS}}(e_k) = e_k^2$ curve, for instance, is steeper than $J_{DS}(e_k)$ for large e_k, but is shallower for small e_k.

The convergence of both QReg and QE is robust to small disturbances when they are persistently excited. What is the effect of such disturbances? The introduction of an additive disturbance in the prediction error (which corresponds to measurement noise, small nonlinearities, or slight mismodeling) implies that the prediction error and parameter errors cannot converge to zero. Instead, they converge to some ball about zero. There is thus no longer a qualitative difference in the steady state behavior between QReg and QE since neither of the algorithms converges to a single point in the nonideal (or noisy) case.

The size of this ball about the origin is dependent on the rate of contraction. The "degree of robustness" of LMS and QReg are easy to compare since they both contract at an exponential rate proportional to the smallest eigenvalue of the appropriate excitation matrix. The QE algorithm, on the other hand, contracts linearly at a rate proportional to $\sqrt{\alpha}$. Although fast convergence is clearly desirable initially (for large errors), it is often advantageous to move slowly when near the desired parameterization, to "average out" spurious signals. Fast convergence for small errors implies that the algorithm reacts quickly to small disturbances, which tends to increase the variance of the parameter estimates. Thus, QReg will tend to reject small noises better than QE, while QE will tend to react better in high noise situations. The QReg algorithm might therefore be preferable to the QE algorithm (assuming PE for QReg) when it is expected to operate in an environment with a high signal-to-noise ratio, while the QE might be preferable in a high noise environment.

VI. Conclusion: More Comparisons

LMS and two Quantized State algorithm forms were compared in terms of numerical complexity; the QReg and QE algorithms were found to be of equivalent complexity, and both were simpler (for appropriate quantizers) than LMS since they replace the multiplications of LMS with bit shifts. The algorithms were then compared in terms of conditions for stability, properties of the convergence of the prediction errors, and convergence rates of the parameter estimates.

LMS and QE can be interpreted in terms of a gradient descent procedure while QReg cannot. The LMS prediction error is mean square convergent while the QE prediction error is mean absolute convergent. Persistence of excitation conditions were then found for all three algorithms, and it was shown that it is strictly more difficult to persistently excite QReg than LMS or QE, which have identical conditions. When the algorithms are persistently excited, the convergence rate of LMS and QReg is exponential to the desired parameterization θ^*, while the QE is linearly convergent to a δ ball about θ^*. Approximate convergence rates were then derived. The exponential rates for LMS and QReg are proportional to $\mu\alpha$ while the linear rate of QE is proportional to $\mu\sqrt{\alpha}$. When the QReg algorithm is not persistently excited, there is the danger of divergence. More *a priori* knowledge of the characteristics of the input sequence is therefore *required* in order to successfully implement the QReg algorithm than to successfully implement QE. The tradeoff is that the convergence of QE may be slower for large errors, and that the parameter variance may be larger for small disturbances. These comparisons are summed up in the accompanying table.

Several issues are raised.

1) In most applications (for instance, [9] and [10]), both QReg

	Least Mean Squares—LMS	Quantized Regressor—QReg	Quantized Error—QE
Algorithm form	$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu X_k e_k$	$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu Q(X_k) e_k$	$\hat{\theta}_{k+1} = \hat{\theta}_k + \mu X_k Q(e_k)$
Numerical Complexity	n multiplies n adds 1 shift	Q = power of two $n + 1$ shifts n adds	Q = power of two $n + 1$ shifts n adds
Lyapunov Stable?	Yes	No	?
Gradient Procedure?	Yes	No	Yes
Convergence of Prediction Error	mean square convergent	not guaranteed	mean absolute convergent
PE Condition	$\lambda_{\min}(\Sigma XX^T) > \alpha$	Re $\lambda_i(\Sigma Q(X)X^T) > \alpha \; \forall i$	$\lambda_{\min}(\Sigma XX^T) > \alpha$
if PE $\cdots$ parameter convergence to	θ^*	θ^*	a δ ball about θ^*
convergence rate	exponential rate proportional to $\mu\alpha$	exponential rate proportional to $\mu\alpha$	linear rate ω proportional to μ and $\sqrt{\alpha}$
notes		possible divergence if PE condition not met	formulas relating δ, ω, μ, and α

and QE tend to behave well, although occasional failures have been noted, especially during quiescent periods when signal power is low. Is there a way to translate the system theoretic warnings of this correspondence into practical guidelines for the use of quantized adaptive algorithms? We believe so, especially in the choice of stepsize, quantization fineness, and degree of PE.

2) In applications of QReg, one would not expect a t-periodic destabilizing sequence to occur regularly. More likely, such sequences would appear amidst other, stabilizing inputs, and cause "glitches" of misbehavior. Is there a way of detecting or guarding against such input dependent misbehavior?

3) Is there a fixed quantizer for which divergence of the QReg algorithm is impossible? We suspect not. Is there, then, a way to quantize dynamically that will ensure stability?

4) The PE for QReg condition is not equivalent to the standard PE for LMS condition. The implications (in terms of rate of convergence, and convergence to a ball rather than a point in the ideal case) for PE for QE are not equivalent to the implications of the standard PE for LMS condition. Consolidating our understanding of the "true nature" of persistence of excitation in light new PE conditions such as these is likely to be a nontrivial task.

Appendix

Proof of Theorem 1: Iterating (3.3) from $k = 0$ to $k = t$ yields

$$V_t - V_0 = -2\mu \sum_{\substack{i=1 \\ i \notin [j: Q(e_j)=0]}}^{t-1} e_i Q(e_i) + \sum_{\substack{i=1 \\ i \notin [j: Q(e_j)=0]}}^{t-1} X_i^T X_i Q^2(e_i) \tag{A.1}$$

where the sums are taken over all i such that $Q(e_i) \neq 0$. These i can be excluded since $V_{k+1} = V_k$ if $Q(e_i) = 0$. The range of the quantizer is M and the maximum value of $X_i^T X_i$ is $n\beta^2$ (recall that $\beta \geq |u(i)|$ for every i). Let q be the smallest nonzero value assumed by the discrete valued quantizer. Since $e_i Q(e_i) > 0$ for every i in the sum, and since $|Q(e_i)| > q > 0$ for these i, V_t can be bounded above by

$$V_t \leq \sum_{\substack{i=1 \\ i \notin [j: Q(e_j)=0]}}^{t-1} [n\mu^2\beta^2 M^2 - 2\mu q |e_i|] + V_0. \tag{A.2}$$

But V_t is a sum of squares and hence must be nonnegative. Therefore, (3.5) holds with $\delta = n\mu\beta^2 M^2/2q$. △△△

Proof of Theorem 4: Suppose first that the quantizer does not have a dead zone, that is, for every $z \neq 0$, $M > Q(z) > q$. Equation (3.3) shows that the one step difference can be bounded

$$V_{k+1} - V_k \leq \mu^2 M^2 X_k^T X_k - 2\mu q |X_k^T \tilde{\theta}_k|. \tag{A.3}$$

The t-step difference can be bounded similarly as

$$\begin{aligned} V_{k+1} - V_k \leq\; & \mu^2 M^2 (X_k^T X_k + X_{k+1}^T X_{k+1} + \cdots + X_{k+t-1}^T X_{k+t-1}) \\ & - 2\mu q [|X_k^T \tilde{\theta}_k| + |X_{k+1}^T \tilde{\theta}_{k+1}| + \cdots \\ & + |X_{k+t-1}^T \tilde{\theta}_{k+t-1}|] \end{aligned} \tag{A.4}$$

where t is chosen to be the t of (4.1) over which the X_k's span $\mathbb{R}^n$. Since $|a - b| \geq |a| - |b|$,

$$|X_{k+1}^T \tilde{\theta}_{k+1}| = |X_{k+1}^T(\tilde{\theta}_k - \mu X_k \operatorname{sgn}(e_k))|. \tag{A.5}$$

Similarly, each of the terms $|X_{k+j}^T\tilde{\theta}_{k+j}|$ can be underbounded by the difference between $|X_{k+j}^T\tilde{\theta}_k|$ and $\mu\ \Sigma_{l=0}^{j-1} X_{k+j}^T X_{k+l}$. Hence, the t-step difference can be *overbounded* by

$$V_{k+t} - V_k \leq 2\mu q[|X_k^T\tilde{\theta}_k| + |X_{k+1}^T\tilde{\theta}_k| + \cdots + |X_{k+t-1}^T\tilde{\theta}_k|] + \mu^2 M^2 C \quad (A.6)$$

where $C = \Sigma_{j=1}^{t-1}\Sigma_{l=0}^{j-1} X_{k+j}^T X_{k+l}$. Note that C is not dependent on $\tilde{\theta}$, and can be overbounded (liberally) by $t^2 \sup_k \|X_k\|^2$. Let $s_k = \Sigma_{l=k}^{k+t-1} |X_l^T\tilde{\theta}_k|$, and suppose that

$$\|\tilde{\theta}_k\| > \delta \quad \text{for some } \delta > 0. \quad (A.7)$$

Then

$$s_k^2 = \left\{\sum_{l=k}^{k+t-1} |X_l^T\tilde{\theta}_k|\right\}^2 \geq \sum_{l=k}^{k+t-1} (X_l^T\tilde{\theta}_k)^2 = \tilde{\theta}_k^T\left\{\sum_{l=k}^{k+t-1} X_l X_l^T\right\}\tilde{\theta}_k \geq \alpha t\|\tilde{\theta}_k\|^2 \geq \alpha t\delta^2, \quad (A.8)$$

and hence, $s_k \geq \delta\sqrt{\alpha t}$, where α is the degree of excitation. Thus, (A.6) may be rewritten

$$V_{k+t} - V_k \leq -2\mu q\delta\sqrt{\alpha t} + \mu^2 M^2 C. \quad (A.9)$$

For small enough μ, the right-hand side of (A.9) can be made negative, and there is an $\omega > 0$ such that $V_{k+t} - V_k \leq -\omega$, which holds for every positive μ less than $\mu^* = (2\delta q\sqrt{\alpha t} - \omega)/M^2C$. This implies that $\|\tilde{\theta}_k\|$ decreases linearly until (A.7) is violated. From (1.6), the maximum motion of the parameter estimates in one timestep is $\mu M\|X\|$. Thus, once $\|\tilde{\theta}_k\|$ enters the δ ball, it can never leave the ball of radius $\delta + \mu M\|X\|$. The restriction on the presence of a dead zone can now be removed by deleting the terms in (A.3)–(A.6) that map to zero, since they have no influence on the value of V_k. Details are carried out in [18]. △△△

References

[1] R. Kumar and J. B. Moore, "Detection techniques in least squares identification," *Automatica*, vol. 17, no. 6, pp. 805–819, 1981.

[2] ——, "Adaptive equalization via fast quantized-state methods," *IEEE Trans. Commun.*, vol. COM-29, pp. 1492–1501, Oct. 1981.

[3] M. A. Aizerman, E. M. Braverman, and L. I. Rozonoer, "The method of potential functions for the restoring the characteristic of a function converter from randomly observed points," *Automat. Remote Contr.*, vol. 25, no. 12, Dec. 1964.

[4] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Trans. Inform. Theory*, vol. IT-30, Mar. 1984.

[5] C. P. Kwong, "Dual sign algorithm for adaptive filtering," *IEEE Trans. Commun.*, vol. COM-34, Dec. 1986.

[6] D. Hirsch and W. J. Wolf, "A simple adaptive equalizer for efficient data transmission," *IEEE Trans. Commun. Technol.*, vol. COM-18, Feb. 1970.

[7] N. J. Bershad, "Comments on 'Comparison of the convergence of two algorithms for adaptive FIR digital filters,'" *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 1604–1606, Dec. 1985.

[8] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "Comparison of the convergence of two algorithms for adaptive FIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 670–678, June 1981.

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*. Englewood Cliffs, NJ: Prentice-Hall, 1984, Section 6.5.

[10] C. F. N. Cowan and P. M. Grant, *Adaptive Filters*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[11] W. A. Sethares and C. R. Johnson, Jr., "Exciting conditions for quantized state adaptive algorithms," in *Proc. ICASSP*, Dallas, TX, Apr. 1987.

[12] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[13] C. R. Johnson, Jr., *Lectures on Adaptive Parameter Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[14] Z. Tsypkin, OCHOBbI NHΦOPMAUNOHHON TEOPNN NΔEHTNΦKAUNN, Moscow, 1984.

[15] W. A. Sethares, D. A. Lawrence, C. R. Johnson, Jr., and R. R. Bitmead, "Parameter drift in LMS adaptive filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, Aug. 1986.

[16] R. R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 183–191, Mar. 1984.

[17] W. A. Sethares, I. M. Y. Mareels, B. D. O. Anderson, C. R. Johnson, Jr., and R. R. Bitmead, "Excitation conditions for signed regressor least mean square adaptation," *IEEE Trans. Circuits Syst.*, vol. 35, June 1988.

[18] W. A. Sethares, "Quantized state adaptive algorithms," Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1987.

[19] A. Weiss and D. Mitra, "Digital adaptive filters: Conditions for convergence, rates of convergence, effects of noise and errors arising from the implementation," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 637–652, Nov. 1979.

[20] W. A. Sethares and C. R. Johnson, Jr., "The LMS adaptive algorithm in a quantized environment," in *Proc. CISS*, Baltimore, MD, Apr. 1987.

Source Location Bias in the Coherently Focused High-Resolution Broad-Band Beamformer

D. N. SWINGLER AND J. KROLIK

***Abstract*—A simple expression for the source location (angle of arrival) bias is developed for Wang and Kaveh's focused broad-band beamformer. It is shown to depend on the source temporal frequency spectrum only through its centroidal frequency. The bias is zero if the angle of arrival is aligned with the primary steering angle (i.e., focusing angle) or, more interestingly, if the source centroidal frequency equals the focusing frequency.**

Manuscript received January 27, 1988. This work was supported by the Defence Research Establishment Atlantic, Dartmouth, N.S., Canada.

D. N. Swingler is with the Division of Engineering, Saint Mary's University, Halifax, N.S., Canada, B3H 3C3.

J. Krolik is with the Department of Electrical Engineering, Concordia University, Montreal, P.Q., Canada, H3G 1M8.

IEEE Log Number 8824493.

0096-3518/89/0100-0143$01.00 © 1989 IEEE