ROBUSTNESS TO QUANTIZATION ERRORS IN LMS ADAPTATION VIA DEGREE OF EXCITATION

Geoffrey A. Williamson

Dept. of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago IL 60616

William A. Sethares

Dept. of Electrical and Computer Engineering
University of Wisconsin
Madison WI 53706

ABSTRACT

The Quantized Regressor (QReg) algorithm is a variant of the Least Mean Square (LMS) algorithm. It is attractive due to its computational simplicity, and it serves as a model of LMS under quantization errors due to digital implementation. The quantization present in the QReg algorithm makes its convergence properties distinct from those of LMS. In this paper, we describe excitation conditions which guarantee convergence of QReg assuming that the quantization in the algorithm is fine enough. In addition, given a fixed fineness of quantization, we develop excitation conditions such that QReg with that quantization fineness converges. The excitation conditions take the form of a *degree* of excitation, and one may interpret this as a measure of robustness of LMS to quantization errors. We provide examples which demonstrate the theory.

I. INTRODUCTION

By far the most widely used adaptive algorithm for FIR adaptive filters is the Least Mean Square (LMS) update scheme [1], and a good deal is known about its convergence properties. However, implementations of the LMS adaptive filter invariably involve quantization of signals within the adaptive mechanism, which can change the characteristics of the adaptive filter behavior. A particularly pervasive source of quantization arises from digital implementations of the adaptive filter [2]. Finite precision arithmetic by nature uses quantized representations of real signals. Additionally, quantization in the LMS adaptive filter may arise by *design*. For instance, in order to reduce the algorithm computational complexity, one may introduce coarse quantization of signals, which replaces costly multiplications by bit shift operations, as in, *e.g.*, [3]. In terms of the convergence properties of the (now quantized) LMS update scheme, the presence of the quantization errors may have adverse effects, as demonstrated in [4], even when the standard Persistent Excitation (PE) condition is fulfilled.

We consider in this paper a modified version of the LMS algorithm, in which the data in the regressor vector are quantized for use in the adaptive algorithm. Previous studies of this *Quantized Regressor* (QReg) algorithm appeared in [4] and [5]. Here, we formulate conditions which guarantee that the QReg algorithm rejects the destabilizing influence of the errors introduced by the quantization. The conditions hinge on satisfaction of a "degree of persistent excitation" for the (unquantized) LMS algorithm. The result demonstrates a tradeoff between the fineness of quantization in QReg and the degree of excitation required to guarantee adaptive filter convergence. Thus, given a maximum quantization error Δ^*, a degree of persistent excitation α (which depends on Δ^*) will ensure exponential convergence of the adaptive filter parameters. Conversely, if the regressor satisfies a PE condition of degree α, then quantization errors bounded by Δ^* (depending on α) may be tolerated. In the course of developing these results, we provide quantitative relationships between the degree of persistent excitation α and the quantization fineness Δ^* such that the convergence results hold. We also show that previous examples of misbehavior due to quantization effects (such as those in [4]) violate these convergence criteria.

Excitation which provides a strong level of stability for the LMS algorithm will yield stability for QReg, if the quantization in QReg is fine enough. This effect is yet another example of the robustness provided to adaptive systems by persistent excitation [6]. In this case, we have satisfaction of a degree of persistent excitation condition providing robustness to errors due to quantization.

We approach the effects of quantization from a deterministic viewpoint. A wide range of related work has used stochastic models for quantization in LMS; two examples are [7] and [8]. Note also the relevance of studies of *signed* regressor algorithms, *e.g.* [9] and [10], to this work, though the signum function quantizer is not a member of the class of quantizers we consider here.

In Section II, we review the LMS algorithm and recall its stability under Persistent Excitation. We then introduce the QReg algorithm, as implemented with a quantizer taken from a class for which quantization errors are bounded. This class includes uniform quantizers, which model round-off errors in fixed-point arithmetic [11]. We note the persistent excitation condition for the QReg algorithm. (We say an excitation sequence is *persistently exciting* for a given algorithm if that algorithm is *exponentially stable* under that excitation.) Section III presents our main results, which, as discussed above, relate quantization fineness in QReg to a degree of excitation guaranteeing stability. Section IV contains examples which illustrate the theory developed in the paper, and Section V provides concluding remarks.

II. THE LMS AND QUANTIZED REGRESSOR LMS ALGORITHMS

The well-known LMS algorithm [1] attempts to model a time sequence $\{y_k\}$ as $X_k^T\theta$, where the regressor vector $X_k = [x_k \; x_{k-1} \; \cdots \; x_{k-n+1}]$ is derived from a data sequence $\{x_k\}$, and where θ_k is an n-vector of parameter estimates. The LMS algorithm is

$$\theta_{k+1} = \theta_k + \mu X_k e_k, \tag{2.1}$$

with $e_k = y_k - X_k^T\theta_k$ termed the prediction error. If we assume that y_k is generated as $X_k^T\theta^*$, for some "true" parameter vector θ^*, then we may speak of parameter errors $\tilde{\theta}_k = \theta^* - \theta_k$. The prediction error is then $e_k = X_k^T\tilde{\theta}_k$, and (2.1) becomes

$$\tilde{\theta}_{k+1} = [I - \mu X_k X_k^T]\tilde{\theta}_k. \tag{2.2}$$

The *error system equation* (2.2) is well-studied [12]. When the sequence $\{x_k\}$ is periodic, one may describe a sharp stability/instability boundary for the exponential convergence of $\tilde{\theta}_k$ to 0 in (2.2) [13], given in terms of the following *persistent excitation (PE) condition.* Although we confine our attention throughout this paper to periodic excitation, the analysis may be extended to encompass more general excitation sequences, using the framework in [6].

PE for LMS: An N-periodic n-vector sequence $\{X_k\}$ is PE for LMS if there exist $\beta \geq \alpha > 0$ such that

$$\beta I \geq \frac{1}{N}\sum_{k=1}^{N} X_k X_k^T \geq \alpha I. \tag{2.3}$$

$\nabla\nabla\nabla$

If $\{X_k\}$ satisfies (2.3) for $\beta \geq \alpha > 0$, then there is a μ^* such that for all $0 < \mu < \mu^*$, (2.2) is exponentially stable at the origin [13]. We think of α and β in (2.3) as specifying the *degree* of PE for LMS. In order to make this explicit, we define

PE(α,β) for LMS: If $\{X_k\}$ satisfies (2.3) for a given $\beta \geq \alpha > 0$, then we say that $\{X_k\}$ is *PE(α,β) for LMS.* $\nabla\nabla\nabla$

Note that as α increases, the contractive term in (2.2) is strengthened. Thus, a larger degree of PE corresponds to enhanced exponential stability for a given, fixed, μ.

Now suppose that a quantized version of X_k is used in place of X_k in the parameter update algorithm (2.1). We restrict our attention to quantizers of fineness Δ, that is, functions $Q_\Delta(\cdot)$ which are monotonic non-decreasing, and for which $|Q_\Delta(x) - x| \leq \Delta$. The uniform quantizer depicted in Figure 1 is of fineness Δ; this quantizer is a common model for one type of fixed point roundoff. If $Q_\Delta(X_k)$ replaces X_k in the LMS algorithm (2.1) (with $Q_\Delta(\cdot)$ acting on a vector indicating a term-by-term quantization), we

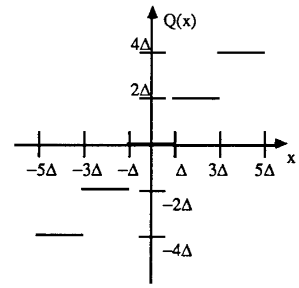

Figure 1: Uniform quantizer of fineness Δ.

have the Quantized Regressor LMS algorithm of quantization fineness Δ (QReg(Δ)):

$$\theta_{k+1} = \theta_k + \mu Q_\Delta(X_k)e_k. \tag{2.4}$$

In error system form, (2.4) is

$$\tilde{\theta}_{k+1} = [I - \mu Q_\Delta(X_k)X_k^T]\tilde{\theta}_k. \tag{2.5}$$

The corresponding condition on $\{X_k\}$ which yields exponential stability for (2.5) (given small enough μ) is [4]

PE for QReg(Δ): An N-periodic n-vector sequence $\{X_k\}$ is PE for QReg(Δ) if there exists $\gamma > 0$ such that

$$\min_i \mathrm{Re}\left\{\lambda_i\left(\frac{1}{N}\sum_{k=1}^{N} Q_\Delta(X_k)X_k^T\right)\right\} \geq \gamma. \tag{2.6}$$

$\nabla\nabla\nabla$

In (2.6), we appeal to the real part of the eigenvalues of the excitation matrix due to its lack of symmetry (the LMS excitation matrix in (2.3) is symmetric non-negative definite, with all its eigenvalues real and non-negative). Conversely, if for $\{X_k\}$, satisfaction of (2.6) requires a *negative* γ, then for μ small (2.5) will be *unstable* at the origin [13].

If $\{X_k\}$ is PE for QReg(Δ), then $\{X_k\}$ is PE for LMS, but *not* the other way around [4]. In other words, excitation conditions which yield good performance for LMS are different from conditions yielding good performance for QReg(Δ). However, in Section III we draw relationships between PE for LMS and PE for QReg(Δ) depending on α, β, and the quantization fineness Δ.

III. PE AND QUANTIZATION

We now state quantitative relationships between *degrees of persistent excitation* and *tolerable levels of quantization.*

Theorem 1: Suppose an N-periodic, n-vector sequence $\{X(k)\}$ is PE(α,β) for LMS. Then for $\Delta < \Delta^* = \alpha/(\sqrt{n\beta})$, $\{X(k)\}$ is PE for QReg(Δ).

Proof: For notational convenience, set

$$R = \frac{1}{N}\sum_{k=1}^{N} Q_\Delta(X_k)X_k^T. \tag{3.1}$$

A sufficient condition for $\{X(k)\}$ to be PE for QReg(Δ) is that the symmetric part of R is positive definite. Setting $\delta_k = Q_\Delta(X_k) - X_k$, we have

$$\text{sym}(R) = \frac{1}{2}(R + R^T) = \frac{1}{N}\sum_{k=1}^{N} X_k X_k^T + S, \qquad (3.2)$$

with $S = (1/2N)\sum_{k=1}^{N}[\delta_k X_k^T + X_k \delta_k^T]$ defining S in (3.2). Since $\{X(k)\}$ is PE(α, β), sym(R) is positive definite if $|\lambda|_{\max}(S) < \alpha$. Noting that S is symmetric, and using $\|\delta_k\| \leq \Delta\sqrt{n}$,

$$\begin{aligned} |\lambda|_{\max}(S) &= \max_{\|v\|=1} |v^T S v| \\ &= \max_{\|v\|=1} \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}(v^T \delta_k X_k^T v + v^T X_k \delta_k^T v) \qquad (3.3) \\ &\leq \Delta\sqrt{n}\max_k \|X_k\|. \end{aligned}$$

Now, $\lambda_{\max}\left(\sum X_k X_k^T\right) \geq \max_k \|X_k\|^2$, so that $\{X(k)\}$ being PE(α, β) for LMS implies $\max_k \|X_k\| \leq \sqrt{\beta}$. Eqn. (3.3) then shows that $|\lambda|_{\max}(S) \leq \Delta\sqrt{n\beta}$. So, if $\Delta < \alpha/\sqrt{n\beta} = \Delta^*$, then $\lambda_{\min}(\text{sym}(R)) > 0$. In other words, $\{X(k)\}$ is PE for QReg(Δ). $\nabla\nabla\nabla$

Thm. 1 shows that a given degree of PE for LMS guarantees PE for QReg(Δ) if Δ is small enough. The bound Δ^* given by Thm. 1 may be conservative, because only magnitude information is used in the development of (3.3), while the basic instability mechanism for QReg derives from *misalignment* between $Q_\Delta(X_k)$ and X_k [5]. However, the bound is tight for the first order case ($n = 1$). When $n = 1$, directional information is not distorted by sign-preserving quantizers, and only magnitude information is important. Thus, Thm. 1 yields an exact bound in that situation.

Thm. 2 provides a converse to Thm. 1, by showing that given a quantization fineness Δ^*, a corresponding degree of PE(α, β) may be chosen which, if satisfied by a sequence $\{X(k)\}$, implies $\{X(k)\}$ is PE for QReg(Δ^*).

Theorem 2: Given $\Delta^* > 0$, dimension n, and condition number $r = \beta/\alpha$, set $\alpha^* = nr(\Delta^*)^2$. If $\alpha \geq \alpha^*$, then having $\{X(k)\}$ be PE($\alpha, r\alpha$) implies that $\{X(k)\}$ is PE for QReg(Δ) when $\Delta < \Delta^*$.

Proof: By Thm. 1, having $\{X(k)\}$ be PE($\alpha, r\alpha$) implies that $\{X(k)\}$ is PE for QReg(Δ) if $\Delta < \alpha/\sqrt{nr\alpha} = \sqrt{\alpha/nr}$. If $\alpha \geq \alpha^*$, then $\Delta < \Delta^* = \sqrt{\alpha^*/nr}$ implies $\Delta < \sqrt{\alpha/nr}$, proving the theorem. $\nabla\nabla\nabla$

Thm. 2 specifies α^*, the degree of PE which is needed in order to tolerate quantization errors up to Δ^*. Notice that this degree depends in part on the eigenvalue disparity r of the LMS excitation matrix (this excitation matrix is the summed outer product in (2.3)). Putting the two theorems together, we have the

Main Result: Given that $\{X(k)\}$ is PE(α, β) for LMS, then there is a fineness of quantization $\Delta^*(\alpha, \beta)$ under which $\{X(k)\}$ is PE for QReg(Δ), for every $\Delta < \Delta^*$. Conversely, if we have a given quantization fineness Δ and condition number r, there are $\alpha^*(\Delta)$ and $\beta^* = r\alpha^*$ such that having $\{X(k)\}$ be PE(α^*, β^*) for LMS implies $\{X(k)\}$ is PE for QReg(Δ).

Proof: A simple application of Thms. 1 and 2. $\nabla\nabla\nabla$

Figure 2 portrays a diagrammatic view of the Main Result. Essentially, we have described a subset of the class of excitation which is PE for QReg(Δ): that which is PE(α^*, β^*) for LMS. A useful interpretation of this result is that persistent excitation for LMS gives robustness to quantization error. On the one hand, if we expect a certain degree of PE from the input signal, then we may determine a sufficient level of quantization fineness for which QReg stability is retained. On the other hand, for quantization of a given coarseness, we may specify a degree of PE under which the QReg algorithm is stable.

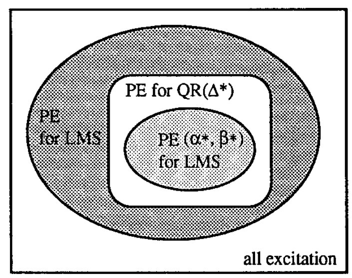

Figure 2: Relationships between PE for LMS and PE for QReg(Δ).

How much PE is needed for QReg(Δ^*)? A glance at Thm. 2 shows that the smallest possible value of α obtained is $(\Delta^*)^2 n$, which occurs when $\beta = \alpha$, or $r = 1$. Thus, this smallest degree of PE requires that the eigenvalue disparity of the excitation matrix be unity. In order to allow for a non-unity eigenvalue disparity, one must increase the needed lower bound on the PE condition.

IV. EXAMPLES

We now present some simple examples. These examples illustrate the application of the theory in Section III, provide insight into the potential restriction of the quantization bounds, and relate this work to the examples of QReg instability appearing in [4] and [5].

Example 1 (from [4] and [5]): Set $n = 3$, and let $\{x_k\}$ be the 3-periodic sequence $\ldots, 2.6, -1.4, -1.4, \ldots$. The quantizer $Q_\Delta(\cdot)$ is a uniform round-off quantizer of fineness 1/2 (essentially, $Q_{1/2}(\cdot)$ rounds to the nearest integer). The LMS excitation matrix in (2.3) is PE, with $\alpha = 0.0133$ and $\beta = 5.333$; $r = \beta/\alpha = 400$ in this case.

The minimum real part of the eigenvalues of R from (3.1) is -0.0667, implying that QReg(1/2) is locally unstable at $\tilde{\theta} = 0$. Applying Thm. 1 gives $\Delta^*(\alpha, \beta) = \alpha/\sqrt{n\beta} = 1/300 \ll 1/2$, which indicates that the quantization fineness $\Delta = 1/2$ is much coarser than the level required in Thm. 1. Conversely, Thm. 2 determines the minimum excitation level α^*, given $\Delta^* = 1/2$ and $r = 400$, as $\alpha^* = (\Delta^*)^2 nr = 300 \gg 0.0133$. Thus, the level of PE for LMS provided by the regressor sequence is far below that specified by Thm. 2, and we are not guaranteed robustness to quantization errors with $\Delta = 1/2$.

The misbehavior of QReg(1/2) for this example is in agreement with the main result, although the bounds of the theorems are rather unrealistic. In fact, with a quantization fineness $\Delta' = 0.465$, QReg(0.465) is *stable* under this excitation sequence. The theoretically required quantization fineness of $\Delta^* = 1/300$ is much smaller than the level of fineness which actually suffices. $\nabla\nabla\nabla$

Example 2: (similar to Ex. 3 in [5]): Let $n = 3$, and consider $\{x_k\} = \{\ldots, 1.9, 0.9, -1.1, \ldots\}$ as a 3-periodic data sequence. Let $Q_\Delta(\cdot)$ be a quantizer which *truncates* to the nearest integer multiple of 2Δ. (The characteristic of this quantizer is the graph of Fig. 1 *shifted right by* $\Delta/2$.) Thus, for $\Delta = 1/2$, $Q_{1/2}(\cdot)$ truncates to the nearest integer, and $\{Q_{1/2}(x_k)\} = \{\ldots, 1.0, 0.0, -2.0, \ldots\}$.

The sequence is PE(α, β) with $\alpha = 0.9633$, $\beta = 2.33$, and $r = \beta/\alpha = 2.42$. However, QReg(1/2) is *divergent*, with the minimum real part of the eigenvalues of R in (3.1) (calculated for this case) being $-0.5667 < 0$. Thm. 2 gives $\alpha^* = 1.81$, which is the degree of PE for which QReg(1/2) is convergent, given r above. The actual excitation has degree of PE $0.9633 = \alpha < \alpha^* = 1.81$, violating the conditions of Thm. 2. Applying Thm. 1 shows that quantizing finer than $\Delta^* = 0.364$ will enable convergence of QReg(Δ), $\Delta < \Delta^*$. Indeed, if we set $\Delta = 0.36$, then the minimum real part of the eigenvalues of R from (3.1) is $0.4080 > 0$, implying such convergence for small enough μ in (2.4).

This example demonstrates a non-trivial situation in which the theoretical bounds on quantization fineness and degree of PE are close to the true limits. $\nabla\nabla\nabla$

V. CONCLUSION

In this paper we have addressed issues regarding quantization in the popular LMS algorithm. We focused on the QReg algorithm (defined in (2.4)), which is a possible alternative to standard LMS when computational simplicity is paramount [4]. With regard to this algorithm, conditions for parameter convergence were developed by relating *degrees* of persistent excitation (provided to standard LMS by an input sequence) to quantization fineness sufficient for convergence of QReg. For a class of excitation which satisfies a degree of PE condition for LMS, we may find a quantization fineness Δ^* such that this class of excitation is PE for QReg(Δ^*). Conversely, given a quantization fineness Δ^* and a condition number for the excitation matrix, we may specify an excitation class which is PE for QReg(Δ^*) in terms of a degree of PE condition for LMS.

An interpretation of this result is that *persistent excitation will provide a degree of robustness to quantization errors.* This paper has quantified the level of robustness attained, and has provided examples exhibiting the flavor of the result in application.

REFERENCES

[1] B. Widrow *et al.*, "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. IEEE*, vol. 64, no. 8, pp. 1151-1162, August 1976.

[2] C. Carioscos and B. Liu, "A roundoff error analysis of the LMS adaptive algorithm," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-32, no. 1, pp. 34-41, February 1984.

[3] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall, 1984, Section 6.5.

[4] W. A. Sethares and C. R. Johnson, Jr., "A comparison of two quantized state adaptive algorithms," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-37, no. 1, pp. 138-143, January 1989.

[5] W. A. Sethares and C. R. Johnson, Jr., "Exciting conditions for quantized state adaptive algorithm," *Proc. ICASSP*, pp. 133-136, Dallas TX, April 1987.

[6] B. D. O. Anderson *et al.*, *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press, 1986.

[7] N. J. Bershad, "Nonlinear quantization effects in the LMS and block LMS adaptive algorithms – a comparison," *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-37, no. 10, pp. 1504-1512, October 1989.

[8] A. Weiss and D. Mitra, "Digital adaptive filters: conditions for convergence, rates of convergence, effects of noise and errors arising from implementation," *IEEE Trans. on Info. Thy.*, vol. IT-25, pp. 637-652, November 1979.

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

[10] W. A. Sethares *et al.*, "Excitation conditions for signed-regressor least mean squares adaptation," *IEEE Trans. Circuits and Sys.*, vol. CAS-35, pp. 613-624.

[11] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, 1975, Chapter 5.

[12] R. R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," *IEEE Trans. on Info. Thy.*, vol. IT-30, no. 2, pp. 183-191, March 1984.

[13] R. R. Bitmead and C. R. Johnson, Jr, "Discrete averaging principles and robust adaptive identification," in *Control and Dynamic Systems: Advances in Theory and Applications* (vol. XXV), ed. C. T. Leondes, Academic Press, pp. 237-271, 1987.