Proceedings of the
American Control Conference
Seattle, Washington • June 1995

A NOTE ON A LYAPUNOV ARGUMENT FOR STOCHASTIC GRADIENT METHODS IN THE PRESENCE OF NOISE

ALI H. SAYED AND PETAR V. KOKOTOVIC

Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106–9560

Abstract—We employ energy-based arguments to establish a robustness and a convergence result for a gradient-descent adaptive law in the presence of both noisy measurements and uncertainty in the initial guess.

I. INTRODUCTION

Consider a collection of measurements $\{y(t) = \phi(t)\theta\}$ that are assumed to be linearly related to an unknown (column) vector of parameters θ via (row) regression vectors $\{\phi(t)\}$. The objective is to devise an estimator for θ. A traditional solution method is to seek a minimization of the quadratic cost function $J(\theta) \triangleq \|y(t) - \phi(t)\theta\|^2$. This is often achieved by employing the gradient descent solution (e.g., [1])

$$\dot{\hat{\theta}}(t) = \Gamma\phi^*(t)\epsilon(t), \quad \hat{\theta}(0) = \text{initial guess}, \qquad (1)$$

where

$$\epsilon(t) \triangleq y(t) - \phi(t)\hat{\theta}(t) \qquad (2)$$

is the estimation error, and Γ is positive-definite. The symbol $*$ stands for Hermitian conjugation. It is well-known that if one introduces the Lyapunov function

$$V(\theta) \triangleq \tilde{\theta}^*(t)\Gamma^{-1}\tilde{\theta}(t), \quad \tilde{\theta}(t) = \theta - \hat{\theta}(t), \qquad (3)$$

then $\dot{V} \leq 0$ and, under additional boundedness conditions on the data (which we shall invoke in the last section), the error signal $\epsilon(t)$ is guaranteed to converge to zero.

This analysis, while it establishes the convergence of the error $\epsilon(t)$ to zero, it leaves unanswered the behaviour of the algorithm in the presence of both measurement noise and uncertainty in the initial guess, $\hat{\theta}(0)$. Several studies of the noisy case in the literature have focused on modifications of the adaptation law (1) in order to guarantee that the convergence properties of the modified algorithms are as close as possible to the convergence properties of (1). In some cases, the proposed modified laws are considerably more complex than the gradient-descent solution (1).

In this note, we persist with (1) and show that a convergence analysis is possible in the presence of noisy measurements, where the noise signal is further required to satisfy certain conditions. In particular, we show that the analysis that is usually carried out by employing the Lyapunov function $V(\theta)$ is still applicable, and that it will allow us to conclude that an estimation error $e(t)$ tends to zero. This error, while it is different from the $\epsilon(t)$ above, is nevertheless the natural error signal to be defined in the noisy case. In the absence of noise, both error signals $e(t)$ and $\epsilon(t)$ will coincide.

As a result of the analysis provided ahead, we shall also conclude that the gradient-based estimator (1) is a robust solution, in the sense of H_∞-theory.

II. A PASSIVITY RELATION

We let the measurements $\{y(t)\}$ be noisy, say $y(t) = \phi(t)\theta + v(t)$. The definition of the estimation error $\epsilon(t)$ will now include a noise term, since

$$\epsilon(t) = y(t) - \phi(t)\hat{\theta}(t) = \phi(t)\tilde{\theta} + v(t) \triangleq e(t) + v(t),$$

where we have introduced $e(t) \triangleq \phi(t)\tilde{\theta}(t)$; this error is due to estimating the uncorrupted term $\phi(t)\theta$ by using $\phi(t)\hat{\theta}(t)$. In the absence of noise, both $\epsilon(t)$ and $e(t)$ are identical. Here, however, they now differ and $e(t)$ is the natural object of interest since we are interested in estimating the *uncorrupted* part of $y(t)$. We thus proceed with a closer study of the differential equation (1).

This material is based on work supported by the National Science Foundation under Award No. MIP-9409319.

Since $\dot{\tilde{\theta}}(t) = -\dot{\hat{\theta}}(t)$ we obtain

$$\dot{\tilde{\theta}}(t) = -\Gamma\phi^*(t)\epsilon(t) = -\Gamma\phi^*(t)\left[e(t) + v(t)\right]. \quad (4)$$

If we now consider the same function $V(\theta)$ as in (3), and use (4), we conclude that

$$-\frac{dV}{dt} \;=\; 2|e(t)|^2 + e^*(t)v(t) + v^*(t)e(t). \quad (5)$$

But, for any $e(t)$ and $v(t)$, the term $|e(t) + v(t)|^2$ is nonnegative, which implies that

$$e^*(t)v(t) + v^*(t)e(t) \geq -|e(t)|^2 \;-\; |v(t)|^2.$$

Substituting into (5) we readily conclude that, for every t, we have

$$\dot{V}(t) \;\leq\; |v(t)|^2 - |e(t)|^2. \quad (6)$$

In the special case $v(t) = 0$, we obtain $\dot{V}(t) \leq 0$ from which convergence of $\epsilon(t)$ to zero follows, as often argued in the literature. In general, however, we have the inequality in (6). This inequality highlights a uniform passivity property of the gradient estimator (1) (i.e., a passivity relation that holds for every time t - see also [3, 5]). It shows that, for every t, the energy of the estimation error, $|e(t)|^2$, never exceeds the sum of the noise energy, $|v(t)|^2$, and the (negative) rate of variation of the energy due to the parameter error, $-\dot{V}(t)$.

III. A Contraction Mapping

In fact, more can be concluded from (6) concerning the behaviour of the estimation procedure. For instance, if we integrate (6) over a finite interval $[0, T]$ we obtain the following revealing inequality

$$\left\{\tilde{\theta}^*(T)\Gamma^{-1}\tilde{\theta}(T) + \int_0^T |e(t)|^2 dt\right\} \leq$$

$$\left\{\tilde{\theta}^*(0)\Gamma^{-1}\tilde{\theta}(0) + \int_0^T |v(t)|^2 dt\right\}. \quad (7)$$

The second line consists of two terms: the total noise energy over $[0, T]$ and the (weighted) energy due to the error in the initial guess for θ. Likewise, the first line consists of two terms: the total estimation error energy over $[0, T]$ and the (weighted) energy due to the final error in the estimate for θ. The inequality then establishes the following robustness result: for the gradient estimator (1), the resulting estimation error energy (due to $e(t)$ and $\tilde{\theta}(T)$) is guaranteed to never exceed the disturbance energies (due to $v(t)$ and $\tilde{\theta}(0)$). In the language of H_∞-filtering (e.g., [6]), this means that the gradient estimator (1) is a robust filter (see also [2, 4]): the map from the disturbances $\{\Gamma^{-1/2}\tilde{\theta}(0), v(\cdot)\}$ to the resulting estimation errors $\{\Gamma^{-1/2}\tilde{\theta}(T), e(\cdot)\}$ is a contraction,

$$\frac{[\theta - \hat{\theta}(T)]^*\Gamma^{-1}[\theta - \hat{\theta}(T)] \;+\; \int_0^T |e(t)|^2 dt}{[\theta - \hat{\theta}(0)]^*\Gamma^{-1}[\theta - \hat{\theta}(0)] \;\;+\; \int_0^T |v(t)|^2 dt} \leq 1.$$

We may remark, however, that in H_∞ studies, the extra term $\tilde{\theta}^*(T)\Gamma^{-1}\tilde{\theta}(T)$ in the numerator is missing. This shows that the above ratio is in fact a stronger inequality, and the usefulness of the extra term in the numerator is demonstrated in the convergence analysis given below.

IV. Convergence of $e(t)$

To establish the convergence of $e(t)$ to zero, we shall assume that $v \in \mathcal{L}_2 \cap \mathcal{L}_\infty$ and $\{\phi, \dot{\phi}\} \in \mathcal{L}_\infty$. That is, v has finite energy over $[0, \infty)$ and $\{v, \phi, \dot{\phi}\}$ are all bounded over the same interval. This also implies $y \in \mathcal{L}_\infty$.

The finite energy assumption on v guarantees that the right-hand side of (7) remains bounded as $T \to \infty$. We thus conclude that $e \in \mathcal{L}_2$.

It further follows from (7) that, for all T, the term $\tilde{\theta}^*(T)\Gamma^{-1}\tilde{\theta}(T)$ is bounded. Hence, $\{\tilde{\theta}, \hat{\theta}\} \in \mathcal{L}_\infty$. Using (4) we further obtain $\dot{\tilde{\theta}} \in \mathcal{L}_\infty$. It then follows from the defining relation $e(t) = \phi(t)\tilde{\theta}(t)$, as well as from

$$\dot{e}(t) = \dot{\phi}(t)\tilde{\theta}(t) + \phi(t)\dot{\tilde{\theta}}(t),$$

that $\{e, \dot{e}\} \in \mathcal{L}_\infty$. In view of Barbalat's Lemma [7, p.186] we conclude that $\lim_{t\to\infty} e(t) = 0$.

References

[1] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[2] B. Hassibi, A. H. Sayed, and T. Kailath. LMS is H^∞ optimal. In *Proc. Conference on Decision and Control*, vol. 1, pp. 74–79, San Antonio, Texas, Dec. 1993.

[3] A. H. Sayed and T. Kailath, A state-space approach to adaptive RLS filtering, *IEEE Signal Processing Magazine*, vol. 11, no. 3, pp. 18–60, July 1994.

[4] G. Didinsky, Z. Pan, and T. Basar. Parameter identification for uncertain plants using H^∞–methods. *Submitted for publication.*

[5] A. H. Sayed and M. Rupp. On the robustness, convergence, and minimax performance of instantaneous-gradient adaptive filters. *Proc. Asilomar Conference on Signals, Systems, and Computers*, pp. 592–596, Oct. 1994.

[6] P. P. Khargonekar and K. M. Nagpal. Filtering and smoothing in an H_∞–setting, *IEEE Trans. on Automatic Control*, vo. 36, pp. 151–166, 1991.

[7] H. K. Khalil. *Nonlinear Systems.* MacMillan, 1992.