THEORETICAL ANALYSIS OF THE CONVERGENCE
PROPERTIES OF ESTIMATION SCHEMES WITH
APPLICATIONS TO ECHO CANCELLATION

Sasan Ardalan

CENTER FOR COMMUNICATIONS AND SIGNAL PROCESSING

North Carolina State University
Raleigh, NC 27650

ABSTRACT

The adaptive estimation algorithms used in echo cancellation are studied with respect to their cancellation properties.   In specific, the Fast Kalman estimation algorithm and the Gradient Search method are analyzed. Analysis of these algorithms is facilitated by associating their properties with a set of non-linear ordinary differential equations (ODE). The asymptotic solutions of these differential equations are then examined and the implications upon the adaptive algorithms' convergence properties discussed.

Table of Contents

THEORETICAL ANALYSIS OF ADAPTIVE ECHO CANCELLATION ALGORITHMS

1. Introduction

The adaptive estimation algorithms used in echo cancellation are studied with respect to their cancellation properties. In specific, the Fast Kalman estimation algorithm and the gradient search method are analyzed. Analysis of these algorithms is facilitated by associating their properties with a set of non-linear ordinary differential equations (ODE). The asymptotic solutions of these differential equations are then examined and the implications upon the adaptive algorithms' convergence properties discussed.

2. Theoretical Approach

The basis for the theoretical convergence analysis is that of Ljung [1]. Consider the estimate $x(t)$ generated by the general recursive algorithm of the form

$$x(t) = x(t-1) + \gamma(t) \ Q[t, \ x(t-1), \ \phi(t)], \qquad (2.1)$$

where $x(t)$ is the estimate Mx1 vector at discrete time t, $\phi(t)$ is an observation vector obtained at time t, and $Q\{\cdot\}$ is a Mx1 vector function to be specified. Additionally, the following dependence of $\phi(t)$ on $x(t-1)$ and an excitation $e(t)$ is assumed:

$$\phi(t) = A[x(t-1)] \ \phi(t-1) + B[x(t-1)] \ e(t), \qquad (2.2)$$

where $A[\cdot]$ and $B[\cdot]$ are, in general, matrix functions. Ljung has shown in [1] that the following differential equation can be associated with the algorithm in (2.1):

$$\frac{dx(\tau)}{d\tau} = f[x(\tau)] \qquad (2.3)$$

where

$$f(x) = E\{Q[x,\overline{\phi}(t;x)]\} \qquad (2.4)$$

and

$$\tau = \sum_{k=1}^{t} \gamma(k) \qquad (2.5)$$

The expectation in (2.4) is taken with respect to the excitation set $\{e(t)\}$, and evaluated as $t \to \infty$. The bar over $\phi(t;x)$ indicates that x is fixed in the evalaution.

In order to use the method of associated differential equations for convergence analysis, certain process and observation assumptions must be satisfied. These will be denoted in the following section. To finish the current section, we motivate the method of associated differential equations by outlining the following properties:

(i) The estimate x(t) in (2.1) can converge only to the stable stationary points of the ODE in (2.3), i.e., points for which $f(x^*) = 0$.

(ii) If $x(\cdot)$ belongs to the domain of attraction of a stable stationary point x^* of (2.3), then x(t) converges with probability one to x^* as $t \to \infty$.

(iii) The trajectories of (2.3) are the asymptotic paths of the estimates $x(\cdot)$ generated by (2.1) and (2.2).

Property (iii) is illustrated in figure 4.1. It is this property which is most useful in analyzing the convergence rate of the associated adaptive algorithms.
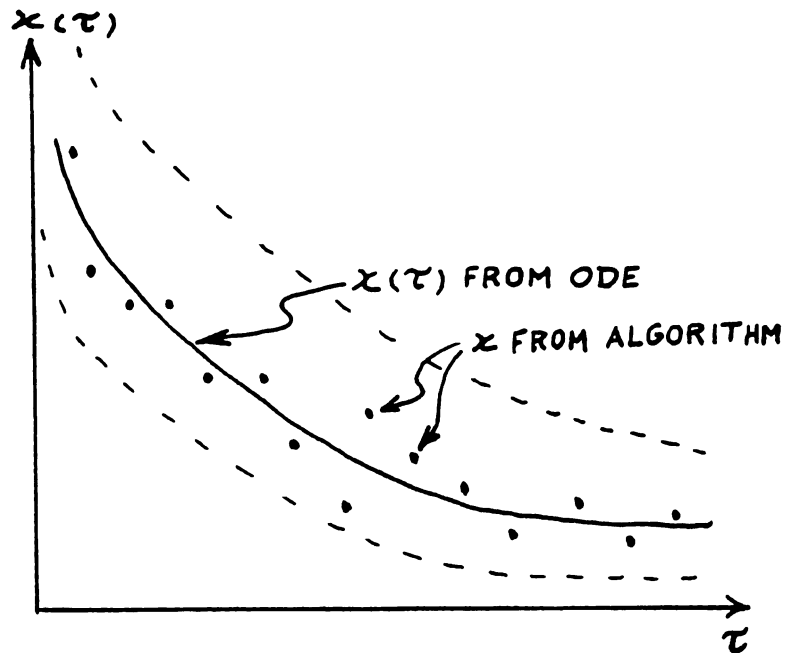
Figure 1. Illustration of Associated Differential Equation
Method, Property (iii).

In the following section, let D_S = $\{x | A(x)$ has all eigenvalues strictly

inside the unit circle $\}$.   Then for each x ε D_S, there exists a λ = $\lambda(x)$ such

that $|A(x)^k| < C\lambda(x)^k$; $\lambda(x) < 1$ where C is a constant.   Also, D_R is an open,

connected subset of D_S. The regularity conditions will be assumed to be valid

in D_R.

3. Assumptions Concerning the Method of Associated Differential Equations

In order for the method of associated differential equations to be valid, the following approximations and assumption must hold:

(a1) The sequence $\{e(t)\}$ from (2.2) is a sequence of independent random variables (but not necessarily stationary or zero mean).

(a2) $|e(t)| < C$ with probability one for all t.

(a3) The function $Q[t,x,\phi]$ from (2.1) is continuously differentiable with respect to x and ϕ for $x \in D_R$. These derivatives are, for fixed x and ϕ, bounded in t.

(a4) The matrix functions $A[\cdot]$ and $B[\cdot]$ from (2) are Lipschitz continuous in D_R.

(a5) The quantity

$$\lim_{t \to \infty} E\{Q[t,\bar{x},\bar{\phi}(t,\bar{x})]\}$$

exists for $\bar{x} \in D_R$ and is denoted by $f(\bar{x})$. This expectation is over the excitation sequence set $\{e(t)\}$.

(a6) $\displaystyle\sum_{t=1}^{\infty} \gamma(t) \to \infty$

(a7) $\displaystyle\sum_{t=1}^{\infty} [\gamma(t)]^p < \infty$, for some p.

(a8) $|\gamma(\cdot)|$ is a monotonically decreasing sequence,

(a9) $\displaystyle\lim \left\{ \sup \left[\frac{1}{\gamma(t)} - \frac{1}{\gamma(t-1)} \right] \right\} < \infty$

Concerning these assumptions, assumption (al) may be somewhat restrictive for studying the behavior of the adaptive algorithms applied to speech. However, note that the observations ϕ from (2.2) may have a "dummy" character, since it may have to be extended to fit both the structure of (2.2) and (al). As an illustration, suppose the sequence $\{e(t)\}$ is a stationary process with a spectral power density which is rational. This implies that $\{e(t)\}$ can be modeled as the output of a stable linear filter with a set of independent random variables $\{e'(t)\}$ as input. Therefore, we may simply extend the dimension of the vector ϕ and adjoin this new filter structure to the model of (2.2). Thus, (al) will hold. One should note that neither the complexity of the algorithm nor the analysis is appreciably altered, since the only change has been to increase the dimensionality of ϕ. The important computation, namely

$$E \{Q[t, \overline{x}, \overline{\phi} (t, \overline{x})\}$$

can still be performed, and often without obtaining explicit expressions for $A[\cdot]$, $B[\cdot]$, $\phi[\cdot]$, or $Q[\cdot]$.

Assumptions (a6) through (a9) are satisfied by a variety of functions. For example $\gamma(t) = ct^{-\alpha}$, $0 < \alpha \leqslant 1$, where c is constant. A special case is $\gamma(t) = 1/t$.

4. Application to Kalman Estimation

Consider the identification of a system S in terms of its impulse function $h(t)$. Let $\hat{\theta}(t)$ be the estimate of the impulse response at time t. Then the updating algorithm to be considered is

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\, \varepsilon(t) \tag{4.1}$$

where the prediction error $\varepsilon(t)$ is given by

$$\varepsilon(t) = y(t) - \hat{\theta}^T(t-1)\, \phi(t) \tag{4.2}$$

and

$$K(t) = P(t)\, \phi(t) \tag{4.3}$$

is the Kalman gain. Furthermore, $P(t)$ is the error covariance matrix and is given by the recursion

$$P(t) = P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{1+\phi^T(t)P(t-1)\phi(t)} \tag{4.4}$$

In the above equation, $\underline{\phi}(t)$ is the input vector

$$\phi(t) = [u(t)\ u(t-1)\ \ldots u(t-N+1)]^T \tag{4.5}$$

The vector $\theta(t)$ is also of dimension Nx1, $P(t)$ is NxN, and $K(t)$ is an Nx1 vector. By the application of the Matrix Inversion Lemma [22] we can write,

$$P^{-1}(t) = P^{-1}(t-1) + \phi(t)\phi^T(t) \tag{4.6}$$

Or,

$$P^{-1}(t) = \sum_{i=1}^{t} \phi(i)\phi^T(i) \tag{4.7}$$

Let,

$$R(t) = (1/t)[\sum_{i=1}^{t} \phi(i)\phi^T(i)] = (1/t)P^{-1}(t) \quad . \tag{4.8}$$

Then $R(t)$ converges to $E\{\phi(t)\phi^T(t)\}$ as $t\to\infty$. From (4.6) we can write,

$$R(t) = (1/t)[P^{-1}(t-1) + \phi(t)\phi^T(t)] \tag{4.9}$$

Using (4.8) we can write (4.9) as

$$R(t) = (1/t)[R(t-1)/(t-1) + \phi(t)\phi^T(t)] \qquad (4.10)$$

Let

$$\gamma(t) = 1/t \qquad (4.11)$$

then (4.10) becomes

$$R(t) = \gamma(t)[R(t-1)(1-\gamma(t))/\gamma(t)+\phi(t)\phi^T(t)]$$

Rearranging,

$$R(t) = R(t-1) + \gamma(t)[\phi(t)\phi^T(t) - R(t-1)] \qquad (4.12)$$

Now, define

$$Q(t;\phi) = \left|\begin{array}{l} R^{-1}(t)\phi(t)\epsilon(t) \\[2mm] col[\phi(t)\phi T(t) - R(t-1)] \end{array}\right| \qquad (4.13)$$

$$x(t) = \left|\begin{array}{l} \theta(t) \\[2mm] col[R(t)] \end{array}\right| \qquad (4.14)$$

Noting that instead of calculating P(t) we can recursively obtain R(t) and that the Kalman gain becomes from (4.3) and (4.8)

$$K(t) = \gamma(t)R^{-1}(t) \qquad (4.15)$$

we can write the recursive Kalman algorithm (4.1) and (4.12) in terms of (4.13) and (4.14) as,

$$x(t) = x(t-1) + \gamma(t)Q(t;\phi).$$

Thus, we have put the Kalman algorithm in the form of (2.1) with $\gamma(t)=1/t$. We may now develop the method of associated differential equations.

From (2.3) and (2.4) using (4.13) for $Q(\cdot)$,

$$\frac{dx}{d\tau} = \begin{vmatrix} d\theta/d\tau \\ \text{col } dR(\tau)/d\tau \end{vmatrix} = \begin{vmatrix} E\{R^{-1}(t)\phi(t)\varepsilon(t)\} \\ \text{col } E\{\phi(t)\phi^T(t)-R(t)\} \end{vmatrix} \quad (4.15)$$

Therefore, the following set of Ordinary Differential Equations can be associated with the Kalman algorithm,

$$\frac{d\theta}{d\tau} = f(\theta) = R^{-1}(\tau)E\{\phi(t)\varepsilon(t)\} \quad (4.16)$$

$$\frac{dR(\tau)}{d\tau} = G(\theta) - R(\tau) \quad (4.17)$$

where

$$G(\theta) = E\{\phi(t)\phi^T(t)\} . \quad (4.18)$$

Further, we have

$$\varepsilon(t) = y(t)-\theta^T\phi(t) \quad (4.19)$$

If we denote by θ^* the true (truncated) system impulse response such that

$$y(t) = (\theta^*)^T\phi(t) \quad (4.20)$$

then

$$\varepsilon(t) = \phi^T(t)\theta^* - \phi^T(t)\theta$$

$$= \phi^T(t)[\theta^*-\theta] . \quad (4.21)$$

Therefore,

$$f(\theta) = R^{-1}(\tau)E\{\phi(t+1)\phi^T(t+1)[\theta^*-\theta]\} . \quad (4.22)$$

Or,

$$f(\theta) = R^{-1}(\tau)G(\theta)[\theta^*-\theta] \quad (4.23)$$

where we have substituted $G(\theta)$ from (4.18).

We may complete the development of the associated differential equation method by considering a consequence of Theorem 2 in [1]. This consequence may be developed as follows:

In the differential equation from (2.3),

$$\frac{dx}{d\tau} = f(x) \ ,$$

let x* be a stable stationary point such that

$$\text{Prob } \{x(t) \to \beta(x^*,\rho)\} > 0$$

for all $\rho > 0$ where $\beta(x^*,\rho)$ denotes a ρ-neighbourhood of x^*, i.e., $\beta(x^*,\rho) = \{x| \ |x-x^*| < \rho\}$. Then for x^* a stable point, we have

$$\frac{dx}{d\tau} \bigg|_{x=x^*} = f(x^*) = 0 \qquad (4.24)$$

and

$$H(x^*) = \frac{d}{dx} f(x) \bigg|_{x=x^*} \qquad (4.25)$$

has all eigenvalues, λ_i, in the left half plane (LHP), i.e.,

$$\text{Real } [\lambda_i] < 0.$$

The matrix $H(x^*)$ represents a linearization around the stationary point x^*. Hence, for the linearized ODE around x^* to be stable the eigenvalues of $H(x^*)$ must be in the LHP.

In other words, if the parameter vector $\theta(t)$ converges to θ^* then we must have $f(\theta^*) = 0$ and

$$H(\theta^*) = \left. \frac{d}{d\theta} f(\theta) \right|_{\theta=\theta^*} \qquad (4.26)$$

must have all eigenvalues in the LHP. Clearly, $f(\theta^*) = 0$ in our case. Also $R^{-1}(\tau)G(\theta)$ is a nonsingular matrix. Thus, from (4.23) we see that $\theta=\theta^*$ is the only solution. Further, since we know that as $t \to \infty$, $R(\tau) \to G(\theta)$, from (4.23) and (4.17) we derive the important result

$$H(\theta^*) = \left. -G^{-1}(\theta) \, G(\theta) \right|_{\theta=\theta^*} = -I \qquad (4.27)$$

where I is the identity matrix. Clearly, the eigenvalues of $H(\theta^*)$ are all in the LHP (they are all equal to -1). Note that as the estimates θ converge to θ^* we imply that $\tau \to \infty$ or

$$R^{-1}(\tau) \to G^{-1}(\theta) \quad .$$

The above analysis shows that the estimates will converge to the true state θ^* and are stable. The stability of the estimates is further developed in a later section.

5. Convergence Properties

To analyze the convergence properties of the Kalman algorithm we make use of the previous results and consider the asymptotic trajectories of the associated differential equations:

$$\frac{d\theta}{d\tau} = R^{-1}(\tau) \; G(\theta) \; [\theta^*-\theta] \qquad (5.1)$$

$$\frac{dR(\tau)}{d\tau} = G(\theta)-R(\tau) \qquad (5.2)$$

First we introduce the following theorem [3]:

Let $A(t)$ be continuously differentiable and $B(t)$ be continuous for $t > t_0$ with

$$\int_{t_0}^{\infty} |A'(t)|dt < \infty, \qquad \int_{0}^{\infty} |B(t)|dt < \infty.$$

where $A'(t) = \frac{d}{dt} A(t)$.

Suppose that all characteristic roots $\lambda_1,\ldots, \lambda_n$ of

$$A_0 = \lim_{t \to \infty} A(t)$$

are simple, and let ζ_i be a characteristic vector of A_0 associated with the characteristic root λ_i. Also, let $\lambda_i(t)$ denote the

characteristic root of A(t) which converges to λ_1 as t→∞. If for some integer k none of the differences

$$\lambda_1(t) - \lambda_k(t) \; ; \; i = 1, \ldots, n$$

changes sign, then the equation

$$\frac{dx}{dt} = [\; A(t) + B(t)]x$$

has a solution $x_k(t)$ such that for t→∞,

$$x_k(t) = \exp \left\{ \int_{t_0}^{t} \lambda_k(s)ds \right\} \cdot [\zeta_k + O(1)]$$

Applying this theorem to the system of ordinary differential equations (5.1) and (5.2) we note that from (5.2)

$$\lim_{\tau \to \infty} R(\tau) = G(\theta) \tag{5.3}$$

and

$$\lim_{\tau \to \infty} R^{-1}(\tau)G(\theta) = I \; . \tag{5.4}$$

Hence, the asymptotic solutions of (5.1,5.2) are

$$\theta(\tau) - \theta^* = \sum_{k} [c_k \zeta_k + O(1)]e^{-(\tau - \tau_0)} \tag{5.5}$$

The solution converges exponentially to zero, i.e. $\theta \to \theta^*$. Note that the convergence rate also depends on the limit (5.3).

Some important results can be derived from the above analysis. The convergence rate does not depend on signal correlation as long as the covariance matrix $R(\tau)$ converges to $G(\theta)$. Hence, there is no distinction in convergence rate with regard to signal correlation as long as the signal is stationary and $R(\tau) \to G(\theta)$ as $\tau \to \infty$. This phenomena is observed in simulations in which the algorithm converges at approximately the same rate for speech signals and signals consisting of white, gaussian, inputs. However, as will be shown, this is not the case for another recursive estimation algorithm, the gradient search method. In the next section we will show that the convergence rate of the gradient search algorithm depends heavily upon the signal correlation. In Kalman estimation, it is important that the input signal have sufficient excitation modes and be quasi-stationary so that $G(\theta)$ exists and is nonsingular.

6. Application to Stochastic Gradient Algorithm

The state estimation update for the gradient search method may be written as

$$\theta(t) = \theta(t-1) + \Gamma\phi(t)\,\varepsilon(t) \tag{6.1}$$

where

$$\varepsilon(t) = y(t) - \phi^T(t)\theta(t-1) \ . \tag{6.2}$$

and Γ is a bounded symmetric positive definite matrix. Ljung [1] has shown that the general algorithm (2.1) and (2.2) could be analyzed by studying the difference equation

$$x^D(t) = x^D(t-1) + \gamma(t)f[x^D(t-1)] \tag{6.3}$$

instead of the differential equation (2.3). We will use both

approaches as necessary.

First the stability of the algorithm will be analyzed.   From (2.1)
we obtain

$$f(\theta) = E\{Q[\theta,\overline{\phi}(t,\theta)]\}$$

$$= \Gamma E\{\phi(t)\,\varepsilon(t)\} \quad . \qquad\qquad (6.4)$$

But from (4.2) and (4.21),

$$\varepsilon(t) = \phi^T(t)\theta^* - \phi^T(t)\theta \quad .$$

Whence,

$$f(\theta) = \Gamma E\{\phi(t+1)\,\phi^T(t+1)(\theta^*-\theta)\}, \qquad\qquad (6.5)$$

or,

$$f(\theta) = \Gamma G(\theta)[\theta^*-\theta], \qquad\qquad (6.6)$$

where we have used the property of (4.18). Now, substituting in (6.3),

$$\theta(t) = \theta(t-1) - \gamma(t)\Gamma G(\theta)[\theta-\theta^*] \qquad\qquad (6.7)$$

Subtracting θ^* from both sides and defining the coefficient error as
$\psi(t) = \theta(t) - \theta^*$, we obtain

$$\psi(t) = \psi(t-1) - \gamma(t)\Gamma G(\theta)\psi(t-1)$$

$$= [I-\gamma(t)\Gamma G(\theta)]\psi(t-1) \quad . \qquad\qquad (6.8)$$

The matrix A $=I-\gamma(t)\Gamma G(\theta)$ must have all eigenvalues in the unit

circle for the algorithm to be stable.   In order to generate this

condition it can easily be shown that the matrix $\Gamma G(\theta)$ must have all

eigenvalues less than 2.   If Γ is a diagonal matrix, and $G(\theta)$ is

diagonal with eigenvalues λ_i, then

$$\max \ \Gamma_k \cdot \lambda_k < 2 \qquad\qquad (6.9)$$

where Γ_k are the diagonal elements of Γ.

At this point we will use the differential equations to analyze the

convergence of the estimates to the true value θ^*. We have

$$\frac{d\theta}{d\tau} = f(\theta) \tag{6.10}$$

$$\frac{d\theta}{d\tau} = \Gamma G(\theta)[\theta^*-\theta] \tag{6.11}$$

Clearly, if $\Gamma G(\theta)$ is a nonsingular matrix with eigenvalues in the RHP then $\theta = \theta^*$ is a solution of $f(\theta) = 0$. Hence, θ^* is a stable stationary point and the estimates converge to θ^*.

7. Convergence Properties of the Gradient Search Algorithm

Based on property (iii) of the theoretical approach, the trajectories of the ODE (equation 6.11) are the "asymptotic paths" of the estimates $\theta(t)$ generated by (6.1). We can observe the asymptotic solutions of (6.11) by use of the following theorem from [3]:

Theorem

Let A be constant matrix whose characteristic roots $\lambda_1, \ldots,$ λ_n are simple and let ζ_i be a characteristic vector associated with λ_i. If $B(t)$ is a continuous matrix defined for $t \geqslant t_0$ such that

$$\int_{t_0}^{\infty} |B(t)| dt < \infty,$$

then the equation

$$\frac{dx}{dt} = [A + B(t)]x$$

has a fundamental system of solutions $x_1(t), \ldots, x_n(t)$ satisfying for $t \rightarrow \infty$,

$$x_k(t) \sim e^{\lambda_k t} \zeta_k \qquad k = 1,\ldots,n$$

Applying this theorem to (6.11) and associating A with $\Gamma G(\theta)$, then the solutions are

$$\theta(\tau) - \theta^* \sim \sum_{k=1}^{n} e^{-\lambda_k t} \zeta_k c_k \quad ; \quad k = 1,\ldots,n \qquad (7.1)$$

where λ_k are the eigenvalues of $\Gamma G(\theta)$. Hence the convergence of the algorithm depends heavily on the eigenvalues of $\Gamma G(\theta)$. If we let Γ be the identity matrix times a constant, i.e. $\Gamma = \Gamma_0 I$, then the λ_k may be written as $\lambda_k = \Gamma_0 \lambda_k'$, where λ_k' are the eigenvalues of $G(\theta)$. Hence by increasing Γ_0, the rate of convergence becomes faster. However, as noted earlier, Γ_0 must be bounded for stability.

If the input signal is white gaussian noise then

$$G(\theta) = \sigma^2 I$$

and

$$\lambda_k' = \sigma^2 \quad (k = 1,\ldots,n)$$

thus the algorithm converges at a uniform rate with $\lambda_k = \Gamma_0 \sigma^2$. For stability,

$$\Gamma_0 < 2/\sigma^2$$

and thus $\lambda_k < 2$. This phenomena serves to explain the slower convergence rate of the gradient search algorithm when compared to the Kalman estimation algorithm when the excitation signal is white and gaussian. For correlated input signals such as speech, the ratio of $\lambda_{max}/\lambda_{min}$ is large for the eigenvalues of $G(\theta)$. From (7.1) we observe that the smallest eigenvalue, λ_{min} dominates the convergence rate thus causing the gradient search method to converge very sluggishly. Moreover, its performance is substantially reduced when compared with the Kalman

method, where the convergence rate was independent of the eigenvalues of

$G(\theta)$ as long as the signal was stationary, that is, as long as

$$\lim_{\tau \to \infty} R^{-1}(\tau) \to G^{-1}(\theta)$$

8. Effects of Imperfect Modelling: Biased Estimates

In most cases the system model is not perfect, that is,

$$y(t) = \phi^T(t)\theta^* + v(t)$$

where $v(t)$ can be correlated with $\phi(t)$. This is the case for example,

in modelling the impulse response in a situation where it must be

truncated. In this case,

$$\varepsilon(t) = \phi^T(t)\theta^* + v(t) - \phi^T(t)\theta(t) \tag{8.1}$$

and

$$f(\theta) = E\{ \phi(t)[\phi^T(t)\theta^* + v(t) - \phi^T(t)\theta(t)] \} \tag{8.2}$$

$$= G(\theta)[\theta^* - \theta] + E\{\phi(t)v(t)\} \tag{8.3}$$

Clearly, if $v(t)$ is uncorrelated with the excitation $\phi(t)$, then

$$E\{\phi(t)v(t)\} = 0$$

and the algorithm converges with probability one to θ^* as shown before.

On the other hand, if $\phi(t)$ and $v(t)$ are correlated then

$$E\{ \phi(t)v(t) \} = \underline{b} \tag{8.4}$$

and for the gradient method

$$\frac{d\theta}{d\tau} = \Gamma G(\theta)[\theta^* - \theta] + \Gamma\underline{b} \tag{8.5}$$

For the Kalman method,

$$\frac{d\theta}{d\tau} = R^{-1}(\tau)G(\theta)[\theta^* - \theta] + R^{-1}(\tau)\underline{b} \tag{8.6}$$

Setting $f(\theta) = 0$ for both methods we find that as $\tau \to \infty$ the stable stationary point is

$$\theta \to \theta^* + G^{-1}(\theta)\underline{b} \tag{8.7}$$

which is clearly biased.

9. Convergence and Stability of the Estimates

We have indicated that the points satisfying $f(x^*) = 0$ in the ODE (2.3) are defined as the stable stationary points and $x \to x^*$ with probability one if $f(x^*) = 0$ and

$$H(x^*) = \frac{df(x)}{dx}\bigg|_{x=x^*}$$

has all eigenvalues in the left hand plane. To show the global stability of these points, i.e., that no matter what the initial condition, $x(0)$, is as long as $\|x(0)\| < \infty$ then the estimates converge to x^*, we make use of the following theorem.

Theorem (Lyaponov) Consider the continuous time free dynamic system

$$\frac{dx}{dt} = f(x,t)$$

where x and f are n-dimensional vectors and $f(\underline{0},t) = \underline{0}$ for all t. If a function $V(x,t)$ that is defined for all x and t is such that $V(\underline{0},t) = \underline{0}$ and satisfies

(i) $V(x,t)$ is positive definite

i.e., there exists a continuous nondecreasing function α

such that $\alpha(0) = 0$ and

$$0 < \alpha(\|x\|) \leqslant V(x,t) \quad x \neq \underline{0}$$

(ii) $\dot{V}(x,t) = \dfrac{\partial V}{\partial t} + \nabla V^T f(x,t) \leqslant -\gamma(\|x\|) < 0$

where γ is a continuous scalar function such that

$\gamma(0) = 0$.

(iii) $V(x,t) \leqslant \beta(\|x\|)$ (or $V(x,t)$ is "decrescent"), where

β is a continuous nondecreasing function and $\beta(0) = 0$.

(iv) $\alpha(\|x\|) \to \infty$ with $\|x\| \to \infty$.

(Or $V(x,t)$ is radially unbounded).

Then the equilibrium state $x = \underline{0}$ is uniformly asymptotically

stable in the large and $V(x,t)$ is called a Lyaponov function

for the system.

We will apply this theorem to the Kalman method.   First we make the

following change of variables in the ODE (4.23): $\psi = \theta^*-\theta$.   Thus,

we obtain the following ODE,

$$\frac{d\psi}{d\tau} = -R^{-1}(\tau)G(\theta)\psi \; . \tag{9.1}$$

Note that if the estimates are not fed back into the excitation vector ϕ

then $G(\theta)$ is independent of θ.   Clearly, if $\theta=\theta^*$ is a stationary point

of (4.23),   then $\psi^* = 0$ is a stationary point of (9.1).   We can

therefore apply the Lyaponov theorem to (9.1).   As a Lyaponov candidate

function   we try

$$V(\theta) = 1/2 \, E\{ \, \varepsilon^2(t;\theta) \, \} \tag{9.2}$$

that is the mean square error which is to be minimized. We must calculate

$$\dot{V}(\theta) = \frac{\partial V(\theta)}{\partial t} + \nabla V^T(\theta) f(\theta) \ . \tag{9.3}$$

Now,

$$\frac{\partial V(\theta)}{\partial t} = 0$$

also,

$$\dot{V} = \nabla V^T d\theta/dt = E\{ \ \varepsilon(t;\theta) \ \frac{d}{d\theta} \ \varepsilon(t;\theta) \}f(\theta) \tag{9.4}$$

where we have taken the gradient operation into the expectation. Now,

$$\frac{d}{d\theta} \ \varepsilon(t;\theta) = \frac{d}{d\theta} \ [\phi^T(t)\theta^* - \phi^T(t)\theta] = -\phi^T(t) \tag{9.5}$$

Substituting (9.5) into (9.4) we obtain

$$\dot{V} = -E\{ \ (\theta^*-\theta)^T \phi(t) \ \phi^T(t) \ \}R^{-1}(\tau)G(\theta)(\theta- \theta^*) \tag{9.6}$$

$$\dot{V} = -(\theta^*-\theta)G(\theta)R^{-1}(\tau)G(\theta)(\theta-\theta^*) \tag{9.7}$$

$$\dot{V} = -f^T(\theta)R^{-1}(\tau)f(\theta) \tag{9.8}$$

where we have made use of (4.23).

As long as $R^{-1}(\tau)$ is nonsingular positive definite, then $\dot{V}$ is negative definite. Furthermore, if we substitute $\psi = \theta-\theta^*$ into (9.8) then $f(\psi) = 0$ and $\dot{V} = 0$ also. Thus all conditions of the theorem are satisfied. Therefore, the stationary point θ^* is uniformly asymptotically stable in the large for all excitation signals for which $R^{-1}(\tau)$ is positive definite. Since $\lim_{\tau\to\infty} R^{-1}(\tau) = G(\theta)$ this implies that the eigenvalues of $G(\theta)$ must be nonzero and positive. This is the

case for white gaussian signals where $G(\theta) = \sigma^2 I$. For complex signals like speech the eigenvalues are positive and nonzero. However, in some cases $\lambda_{min}/\lambda_{max} \to 0$ as the dimension of $\phi(t)$ is increased.

For the stochastic gradient method we use the mean square error as the candidate Lyaponov function as before. In this case we have for $\dot{V}$,

$$\dot{V} = -(\theta-\theta^*)^T G(\theta) \Gamma G(\theta)(\theta-\theta^*) \qquad (9.9)$$

From (9.9) $\dot{V}$ is negative definite if $G(\theta)\Gamma G(\theta)$ is positive definite. If we write $G(\theta) = P\Lambda P^{-1}$ where Λ is a diagonal matrix of the eigenvalues of $G(\theta)$, and we let $\Gamma = \Gamma_0 I$ then,

$$G(\theta)\Gamma G(\theta) = \Gamma_0 P\Lambda P^{-1}P\Lambda P^{-1} = \Gamma_0 P\Lambda^2 P^{-1} \qquad (9.10)$$

Thus, if $\lambda_{max}/\lambda_{min}$ is large then Λ^2 moves towards singularity and the point θ^* cannot be garranteed to be uniformly asymptotically stable in the large. In the Kalman case ,

$$\dot{V} = -(\theta-\theta^*)G(\theta)(\theta-\theta^*) = -(\theta-\theta^*)P\Lambda P^{-1}(\theta-\theta^*) \qquad (9.11)$$

In this case if $\lambda_{max}/\lambda_{min}$ is large at least $\dot{V}$ does not depend on the square of the diagonal matrix as in the gradient method. Hence, global stability is more likely for the Kalman method in cases where $\lambda_{min}/\lambda_{max} \to 0$ than the gradient method if we can quantify in this manner.

References

1. L. Ljung, "Analysis of Recursive Stochastic Algorithms," IEEE Trans. on Automatic Control, vol. AC-22, No. 4, August 1977.

2. Soderstrom, Ljung, and Gustavsson, "A Theoretical Analysis of Recursive Identification Methods,", Automatica , vol. 14, 1978, pp. 231-244.

3. W. D. Coppel, Stability and Asymptotic Behavior of Differential Equations, D. C. Heath and Company, Boston, 1965.

4. Kumpati S. Narendra, "Stability of Nonlinear Systems," Nonlinear Systems Analysis and Synthesis: Volume 1. Fundamental Principles, The Amarican Society of Mechanical Engineers, New York, 1976.