New Results on the Performance of a Well-Known Class of Adaptive Filters

MAN MOHAN SONDHI* AND DEBASIS MITRA,* MEMBER, IEEE

Abstract—We derive a broad range of theoretical results concerning the performance and limitations of a class of analog adaptive filters. Applications of these filters have been proposed in many different engineering contexts which have in common the following idealized identification problem: A system has a vector input $x(t)$ and a scalar output $z(t) = h'x(t)$, where h is an unknown time-invariant coefficient vector. From a knowledge of $x(t)$ and $z(t)$ it is required to estimate h. The filter considered here adjusts an estimate vector $\hat{h}(t)$ in a control loop, thus

$$\frac{d}{dt}\hat{h} = KF[z(t) - \hat{z}(t)]\, x(t)$$

where $\hat{z}(t) = \hat{h}'x(t)$, F is a suitable, in general nonlinear, function, and K is the loop gain. The effectiveness of the filter is determined by the convergence properties of the misalignment vector, $r = h - \hat{h}$. With weak nondegeneracy requirements on $x(t)$ we prove the exponential convergence to zero of the norm $\|r(t)\|$. For all values of K, we give upper and lower bounds on the convergence rate which are tight in that both bounds have similar qualitative dependence on K. The dependence of these bounds on K is unexpected and important since it reveals basic limitations of the filters which are not predicted by the conventional approximate method of analysis, the "method of averaging." By analyzing the effects of an added forcing term $u(t)$ in the control equation we obtain upper bounds to the effects on the convergence process of various important departures from the idealized model as when noise is present as an additional component of $z(t)$, the coefficient vector h is time-varying, and the integrators in a hardware implementation have finite memory.

I. INTRODUCTION

A. The Adaptation Algorithms

IN THIS PAPER, we will derive a broad range of theoretical results concerning the performance and limitations of a certain class of analog adaptive filters. The results derived here are relevant to the wide variety of areas in which the filter has been proposed. In Section I-B we give a number of examples of communications-related applications.

To introduce the adaptation algorithms studied here, let us first consider the following idealized identification problem (see Fig. 1): An unknown system (black box) has a continuous vector input $x(t)$ of known dimension, and a scalar output $z(t)$ related through the equation

$$z(t) = h'x(t). \tag{1}$$

Here h is an unknown time-invariant coefficient vector and $'$ denotes transposition. From a knowledge of $x(t)$ and $z(t)$, it is desired to estimate the coefficient vector h.

The adaptive procedure consists of using an adjustable estimate $\hat{h}(t)$ of the coefficient vector to generate the corresponding estimate
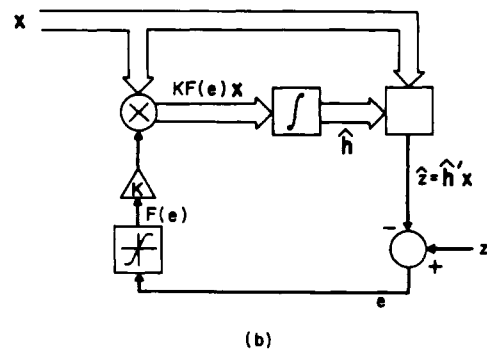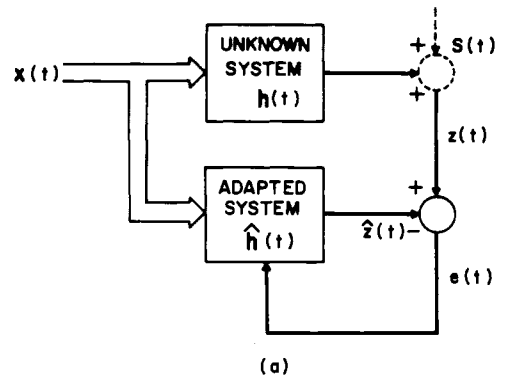
$$\hat{z}(t) = \hat{h}'(t)x(t) \tag{2}$$

Fig. 1. (a) Schematic of the identification problem: In the idealized problem $s(t) \equiv 0$ and $h(t)$ is constant. In Sections II-B-6 and II-C-3 these restrictions are removed. As stated in the text $s(t)$, shown here as observation noise, may in fact include terms due to plant noise.
(b) Schematic of the adaptation algorithm: In the idealized problem the integrators are perfect. In Sections II-B-6 and II-C-3 leakage in the integrators is allowed.

of the output. The difference

$$e(t) = z(t) - \hat{z}(t) \tag{3}$$

is in turn used to adjust the estimate $\hat{h}(t)$. Of the many adjustment algorithms that have been proposed, the ones we will discuss here are

$$\frac{d}{dt}\hat{h} = Ke(t)x(t) \tag{4a}$$

$$\frac{d}{dt}\hat{h} = KF[e(t)]\, x(t) \tag{4b}$$

where K is a constant that adjusts the gain of the control loop. In (4b) the scalar function F is assumed to be continuous and to satisfy both the Lipschitz condition and the following sector condition. For all σ

$$\gamma_1 \sigma^2 \leqslant \sigma F(\sigma) \leqslant \gamma_2 \sigma^2 \tag{5}$$

for some finite positive γ_1 and γ_2. Although (4a) is a special case of (4b), it will be convenient to treat it separately because it is linear in $\hat{h}$.

We have also considered the related algorithm

$$\frac{d}{dt}\,\hat{h} = K \,\text{sgn}\,[e(t)]\,x(t). \tag{4'}$$

For clarity of presentation, however, we will discuss our results for (4') in a sequel to this paper. (Due to the discontinuous right-hand side, this equation needs special treatment. The convergence process is more complicated and in some respects qualitatively different from that of (4a) and (4b). We are unable, for this reason, to give a unified account of the results).

It is useful to define a misalignment vector[1] r_t as

$$r_t \triangleq h - \hat{h}_t \tag{6}$$

and to rewrite (4) in terms of r_t. Noting[1] that $\dot{h} = 0$ and

$$e_t = r'_t x_t \tag{7}$$

equations (4a) and (4b) become

$$\frac{d}{dt}\,r_t = -K(r'_t x_t)x_t \tag{8a}$$

$$\frac{d}{dt}\,r_t = -KF(r'_t x_t)x_t. \tag{8b}$$

Our discussion in Sections I-B and I-C will indicate that due to their simplicity and versatility the algorithms (4) have found a variety of applications. However, in spite of this popularity there is little in the literature that sheds any light on the limitations of the algorithms, or on the dependence of convergence rates on the gain parameter K. And no theoretical results have been reported which elucidate the requirements on the input x_t that would guarantee a minimum convergence rate.

In the sequel, we will present a number of results which help remedy this situation. Our objective throughout will be to obtain results that are valid without imposing stringent requirements on the input, and which can be used without detailed knowledge about the input. This is of utmost importance for many applications where the input process x_t is derived from a complex signal such as speech.

In view of this objective, the key restriction we will place on x_t is a nondegeneracy requirement (Section II-A) which we call the "mixing condition." With this as the most restrictive requirement, we will be able to derive the following results:

i) We will prove uniform exponential convergence to zero of the norm[2] $\|r_t\|$ of the misalignment vector. It was not known previously that this type of convergence is guaranteed under conditions as weak as those assumed here.

ii) We will derive upper and lower bounds to the convergence rate. These bounds are valid for *all* K and for *all* t.

As we will show, these bounds provide valuable insight for deciding what the loop gain K should be. For this reason, we

have devoted considerable effort to obtain tight bounds. In particular, both the lower and upper bounds that we derive have the same qualitative dependence upon K, so that room for improvement is limited.

iii) The dependence of the lower and upper bounds on K, for large K, is rather unexpected and extremely interesting. In both bounds the *rate* of convergence is asymptotically proportional to $1/K$. This is important, as it illuminates a fundamental limitation of the filter which exists in addition to the known limitations such as that due to the presence of noise. Earlier investigators [9], [14] appear to have associated fast convergence with large values of K, in the absence of noise. The only evidence to the contrary appears to have been observed in simulations with sinusoidal inputs by Miller [31]. In view of our bounds, this issue is resolved for a rather broad class of inputs.

iv) Partly in order to provide convincing evidence that our bounds are quite realistic, and partly for its independent value, we have constructed a rather large class of well-behaved input processes x_t in which the components of the vector are restricted linear combinations of a fixed number of sinusoids with distinct arbitrary frequencies. For such inputs we will explicitly solve (8a) and show that the actual behavior of the filter is, at least qualitatively, identical to our bounds.

Finally we will study the effect of adding a vector forcing term u_t on the right-hand side of (8a) and (8b); i.e.,

$$\frac{d}{dt}\,r = -K(r'_t x_t)x_t + u_t \tag{9a}$$

$$\frac{d}{dt}\,r = -KF(r'_t x_t)x_t + u_t. \tag{9b}$$

v) For each of the equations (9a) and (9b) we will show that if a time averaged value of $\|u_t\|$ or $\|u_t\|$ itself is bounded, then so is $\|r_t\|$ and we will obtain explicit expressions for the bound on $\|r_t\|$.

vi) We will exploit this bounded-input bounded-output property by noting that the effect of many practical limitations which prevent implementations of the idealized algorithms (4) can be represented by the term u_t. Thus: a) the effect of an added component s_t in the observed output z_t; b) the effect of leakage in the integrators used in a hardware implementation; and c) the effect of slow variations in the unknown vector h, can all be lumped into the term u_t in (9).[3] We will obtain estimates of the effect of each of these perturbations on the convergence process.

It is desirable to analyze the effects of the added component s_t (case a) for a number of reasons. First, it will generally be the case that some form of noise, perhaps only measurement noise, will be present as an additional component of z_t. Noise in the measurement of x_t may be similarly represented. Secondly, the number of coefficients required to completely characterize the unknown system may be quite high and the coefficient vector h defined in (1) may represent only a subset of all the coefficients. In this case s_t may be taken to represent the effect of the coefficients not directly taken into account in (1). Thirdly, in certain applications as, for instance, during "double talk" in echo cancellation [14] a second signal is superimposed on the output of the unknown system.

---

[1] Whenever convenient, we will denote the independent variable as a subscript. We will also use a dot above a variable to denote its time derivative.

[2] By norm we will always mean the Euclidean norm. Thus $\|r\|^2 = r'r$.

[3] Note that in (9) u_t is quite unrestricted. In particular, it is allowed to depend on x_t and r_t.

For the last of the perturbations, namely c), our general result provides an upper bound on the capability of the filter to track variations in the unknown system's characteristic; this bound is not hard to calculate since the only information on the variation that is required is a bound on its derivative.

The rest of the paper is organized as follows: Following this paragraph, in Section I-B, we will discuss the motivation for considering the particular class of adaptive filters described by (4), and discuss some applications that have been proposed. In Section I-C we will give an overview of known theoretical results and discuss our contributions in the light of what is known. In Section II-A we will state the main assumption on the input process. Section II-B will be devoted to the linear algorithm (4a), (8a), and (9a), and Section II-C to the non-linear algorithm (4b), (8b), and (9b). In these sections we will summarize our results and try to convey the key ideas by informal discussion. The cumbersome details of a number of proofs will be deferred to Appendices I-V.

B. Motivation and Applications

An identification problem similar to the one outlined in Section I-A is encountered whenever one must infer the characteristics of a system without interrupting its normal operation. This situation arises frequently in practice, and consequently the problem has attracted considerable attention. To get an idea of the variety of approaches to this problem we refer the reader to the recent book by Eykhoff [1]. Here we only mention two types of solutions which have bearing on (and in a sense are precursors of) the adaptive algorithms (4).

One method consists of inferring the system impulse response (equivalent, in our case, to estimating h) from the correlation between input and output. The drawbacks of this approach are: i) a large amount of data must be processed in order to compute the requisite correlations; ii) the input process must be stationary so that time averages are meaningful; iii) the final matrix inversion can be rather difficult when the dimension of the vector h is large; and iv) if h is slowly varying, the whole procedure must be periodically repeated in blocks, with the assumption that h is constant within each block.

The second approach is a continuous multiparameter version of the classical Robbins–Monro [2] stochastic approximation. It eliminates the computation of correlations and yields a continuously adjusted estimate of h. Again, a large number of variations on this theme have appeared in the literature, and we refer the reader to a recent survey article by Tsypkin [3]. In its simplest form, the algorithm is exactly like (4) except that the gain parameter K is replaced by a scalar (or matrix) function $\Gamma(t)$. The difference is crucial, however, and worthy of comment. The function $\Gamma(t)$ in the stochastic approximation method is a predetermined function which must approach 0 as $t \to \infty$ in a manner such that $\int_0^\infty \Gamma dt \to \infty$ and $\int_0^\infty \Gamma^2 dt < \infty$. In such case, it can be shown that with stationary input and fixed h, the estimate $\hat{h}$ converges to h in the mean-square sense *even in the presence of additive noise* [e.g., as in (9)]. The fact that the estimate converges in spite of noise is, of course, the most remarkable aspect of stochastic approximation. However, convergence can be achieved only by a rigid demand that h be strictly time-invariant (or, as in some recent work [4], [5], that the form of the time-dependence of h be known). Also the input must be stationary.

For many applications, the inability to track slow fluctuations in h can be disastrous. Also the stringent stationarity requirement on the input can rarely be met in practice. By using a gain parameter which does not approach zero as $t \to \infty$, the adaptation (4) are able to cope with both these difficulties, as our results will show. This fact, plus simplicity of implementation, far outweigh the limitations of (4) in many applications. The basic limitations, as we will show later in this paper, are that perturbations (e.g., as in (9)) result in a residual error in the estimate, and that the convergence rate cannot be increased beyond a certain maximum, even for the ideal, noiseless condition (8). Of these, the first limitation is unavoidable in this type of algorithm. The second limitation can be relaxed, in principle, by making the right-hand sides of (4) more complicated. Such proposals have been made [6], [7]. They invariably require the continuous computation of matrices, resulting in a complexity that increases as the square of the dimension of h. For large dimensional systems (and we might mention here that such algorithms have been realized in hardware for a 200-dimensional vector h [17]) these improvements are not feasible.

In view of these advantages over the other available methods, it is not surprising that the adaptation scheme of (4) has been used for the identification of a wide variety of systems [1]. We mention here a few salient communications-related applications.

A version of (4) with time-discrete[4] adjustments was proposed by Widrow and Hoff [8] for adapting switching circuits. Narendra and McBride [9] proposed the algorithm for a general self-optimizing control system. Lion [6] showed how a system with rational transfer function may be adaptively identified using (4). Widrow *et al.* [10] have designed an antenna system adapted according to a digitized version of (4). A time-sampled version has also been proposed by Lucky [11] and by Gersho [12] for equalization of telephone lines for data communication and a number of equalizers [13] based upon these principles have been implemented. Sondhi [14] has described use of the algorithm for cancellation of echos on long distance telephone connections and both analog [15], [16], and digital [17], [18] versions of the canceller have been realized in hardware.[5] Finally, the algorithm has been used for identifying the linear predictor coefficients [20], as well as the fundamental pitch period [21] of a speech wave.

C. Known Theoretical Results

It is well known that in the case of (4), the error defined by (3) asymptotically approaches zero under very weak restrictions. This is easily shown by considering the equivalent formulation (8) in terms of the misalignment vector r. Premultiplying (8) by r' and recalling that $e_t = r'_t x_t$, we get

$$r'\dot{r} = \frac{1}{2}\frac{d}{dt}\|r\|^2 = -Ke^2 \qquad (10a)$$

and

$$r'\dot{r} = \frac{1}{2}\frac{d}{dt}\|r\|^2 = -KeF(e). \qquad (10b)$$

---

[4] All results derived in this paper will be for the continuous equations (8) and (9). The algorithms can be discretized in a variety of ways. Properties of the discretized version will differ from those presented here, in a manner dependent upon the particular quadrature formula used.

[5] In a recent paper [19], we considered a realization of the algorithm (4) in which a particular type of nonideal multiplier is used exclusively. However, the analysis of the present paper does not apply to such a realization.

Thus $\|r_t\|^2$ is nonincreasing in (8), which is, of course, the original motivation for these adaptation algorithms. Observe that this monotonicity property is true *for all values of K* and, as this is not true in discrete time versions of the adaptation algorithm, it constitutes an important distinction between the two versions. Integrating (10), it is clear that the quantities e^2 and $eF(e)$, respectively, have finite integrals from 0–∞. From this it follows that with appropriate weak restrictions (see e.g., [14]) $\lim_{t \to \infty} |e_t| = 0$.

The fact that $|e_t| \to 0$ is in itself useful in some situations [14], [21]. However, it is clear that without further restrictions, (10) *does not* imply that $\|r_t\| \to 0$, as $t \to \infty$.

In connection with the problem of convergence of $|e_t|$, we digress briefly to make some observations on a facet which might not have been properly appreciated in the past. First, we can show that in the noiseless situation, the control of $|e_t|$ is trivial in that $|e_t|$ can be made uniformly as small as desired by making the loop gain K sufficiently large. The complete statement of this result and the proof may be found in Appendix I. Secondly, we know from the bounds that are presented in this paper that the effect of increasing K, when K is sufficiently large, is to reduce the rate of convergence of r_t. Therefore we may conclude that, for K sufficiently large, in the generic situation e_t is uniformly small while the convergence of $\|r_t\|$ is effectively stalled at some, possibly high, value. We cannot envisage an application where such circumstances are anything but totally undesirable, because slight departures from the idealized assumptions will lead to large errors. Thus even in those cases where the control of e_t is desired, the formulation should be as an *identification* problem with the emphasis on the rapid convergence of the misalignment vector r_t.

For the time-discrete equivalent of (8a), a number of investigators have shown that the expectation $E(\|r\|) \to 0$ if the successive n vectors are independent and identically distributed (see, e.g., [10], [12]). Convergence (and in the case of (9a), boundedness) have also been proved for stationary inputs by Jones [22] and for M-dependent stationary inputs, by Kim and Davisson [23]. Daniell [24] has also proved convergence without requiring independence of the input vectors. Instead he requires a form of asymptotic independence. Among his other requirements are ergodicity and the uniform boundedness of a conditional expected value of the fourth power of the norm of the input vectors.

For continuous algorithms, the convergence of $\|r\|$ has been proved in the case of (4a) by Lion [6] for periodic inputs, and convergence of $E[\|r\|]$ by Kushner [25] for an appropriately restricted stochastic input.

The arguments used for proving convergence in the case when successive x's are independent are easily extended to apply to the nonlinear equation, (8b). However, for correlated inputs we have not come across a convergence proof for the nonlinear algorithms, continuous or discrete.

As far as convergence rate is concerned, all published results are essentially based upon averaging the right hand side of (8) and assuming r to be either constant or independent of x [14], [26]–[28]. The method of averaging in the deterministic form is originally due to Bogoliubov [29], while recent results on the extension to stochastic differential equations is due to Khasminskii [30]. For discretized algorithms the method is valid for independent inputs. For the continuous algorithms, it is valid [30] only over intervals of order $0(1/K)$. Thus for correlated inputs or for large K, no published estimates of convergence rate are to be found.

In certain applications, e.g., equalization for data transmission, it is possible to justify the assumption that $\{x\}$ is a stationary process (or even an uncorrelated sequence). In many other applications, e.g., when x is drived from a speech signal, this assumption is not justifiable. To the best of our knowledge, there are no results in the literature which apply once the stationarity requirement is relaxed.

II. SUMMARY AND DISCUSSION OF RESULTS

A. The Assumptions About the Input Process

As mentioned in Section I, for most of our results to hold we require that the input process be nondegenerate in a specific sense. Essentially, what we mean is that if x_t (and hence h and $\hat{h}$) is N-dimensional, then x_t should not stay confined to a subspace of dimension less than N for too long. There are several equivalent ways of specifying this property. The form in which we state the property will be referred to as the *mixing condition*.

The input function x_t, $t \geqslant 0$, will be said to have satisfied the mixing condition if there exist numbers $T > 0$ and $\alpha > 0$, such that

$$\frac{1}{T} \int_t^{t+T} [d'x_\tau]^2 \, d\tau \geqslant \alpha \|d\|^2 \qquad (11)$$

for all $t \geqslant 0$ and all fixed (i.e., time invariant) vectors d.

The mixing condition is equivalent to the requirement that the smallest eigenvalue of the scaled Gramian matrix

$$\frac{1}{T} \int_t^{t+T} x_\tau x_\tau' \, d\tau \qquad (12)$$

is bounded from below by a positive number. Thus, by the well-known equivalence of positive definite Gramians and linear independence of a set of functions, as well as from (11) directly, it follows that the mixing condition is equivalent to the requirement that x_t have components $x_t(i)$ which are linearly independent over all intervals of a finite length T. We emphasize that the condition makes no requirements of stationarity when the input is a stochastic process. Observe, however, that in the event that the input is a stationary ergodic process, then the mixing condition is satisfied by almost all paths if the covariance matrix is positive definite.

In addition to the mixing condition, we will require that there is a positive number L, such that for T as in the mixing condition and $t \geqslant 0$,

$$\frac{1}{T} \int_t^{t+T} \|x_\tau\|^2 \, d\tau \leqslant L^2. \qquad (13)$$

The bulk of our results will require no further properties of the process x_t. However, for the derivation of the lower bounds (Sections II-B-4 and II-C-2), we will not require the mixing condition. Instead, we will require that dx_t/dt exists and is uniformly bounded and that $\|x_t\|$ has a uniform lower bound.

B. Results for the Linear Algorithms

1) Upper Bound: We have already shown in Section I-C that the norm of the misalignment vector r_t is a nonincreasing

function of time for each of (8a) and (8b). Here we will show that if x_t satisfies (11) and (13), then for (8a), $\|r_t\|$ is bounded from above by a decaying exponential function. One way of proving this is as follows.

Integrating (8a) from t_0 to $(t_0 + T)$ gives

$$r_{t_0+T} = r_{t_0} - K \int_{t_0}^{t_0+T} x_t x_t' r_t \, dt. \tag{14}$$

On the right-hand side, replacing r_t by $r_{t_0+T} + r_t - r_{t_0+T}$ and rearranging terms gives[6]

$$\left[I + K \int_{t_0}^{t_0+T} x_t x_t' \, dt \right] r_{t_0+T} = r_{t_0} - K^2 v \tag{15}$$

where v is the vector

$$v = \int_{t_0}^{t_0+T} x_t x_t' \int_{t}^{t_0+T} x_\tau x_\tau' r_\tau \, d\tau \, dt. \tag{16}$$

If (15) is premultiplied by r_{t_0+T}', the left-hand side can be bounded from *below* by the mixing condition, and $r_{t_0+T}' v$ can be bounded from *above* by repeated use of the Schwarz inequality. This basically simple but somewhat cumbersome derivation is shown in Appendix II. The end result is

$$\|r_{t_0+T}\| \leqslant \sqrt{1 - s_o^2} \|r_{t_0}\| \tag{17}$$

where s_o is the unique positive root of the quadratic equation

$$(1 + KT\alpha)^2 (1 - s^2) = \left(1 + \frac{s}{2} (KTL^2)^{3/2}\right)^2. \tag{18}$$

In (18), L^2 is the average power bound of (13). The substitution leading to (15) can be used once more and, as shown in Appendix II, the bound of (17) is obtained again, with s_o as the unique positive root of the quadratic

$$\left(1 + KT\alpha + \frac{1}{2} K^2 T^2 \alpha^2\right)^2 (1 - s^2) = \left(1 + \frac{s}{4} (KTL^2)^{5/2}\right)^2. \tag{19}$$

By repeated use of inequality (17) we get for $n = 1, 2, \cdots$,

$$\|r_{t_0+nT}\| \leqslant (1 - s_o^2)^{n/2} \|r_{t_0}\|. \tag{20}$$

This geometrically decreasing bound can be converted directly to a continuous exponential bound when we use the fact that $\|r_t\|$ is nonincreasing. Thus for $t \geqslant t_0$

$$\|r_t\| \leqslant a e^{-b_1(t-t_0)} \|r_{t_0}\| \tag{21}$$

where

$$a = e^{b_1 T} \tag{22}$$

and

$$b_1 = -\frac{1}{2T} \ln (1 - s_o^2) > 0. \tag{23}$$

(For $t - t_0 \leqslant T$, of course, the bound might just as well be replaced by $\|r_{t_0}\|$.)

The dependence of b_1 on the gain parameter K as well on the other parameters α, L^2, and T, is interesting and somewhat

[6] The identify matrix is denoted by I.

unexpected. Before discussing this dependence, let us derive a bound similar to (21) (with b_1 replaced by a slightly different exponent b_2) by a rather different type of reasoning. The exponent b_2 will turn out to be significantly larger than b_1 for large K. Also, the second method gives, in general, a much better bound in the nonlinear case.

The derivation of this second bound starts with the representation

$$\int_{t_0}^{t_0+T} (x_t' r_{t_0})^2 \, dt = \int_{t_0}^{t_0+T} (x_t' r_t)^2 \, dt + \int_{t_0}^{t_0+T}$$
$$\cdot [x_t'(r_{t_0} - r_t)]^2 \, dt + 2 \int_{t_0}^{t_0+T} (x_t' r_t)[x_t'(r_{t_0} - r_t)] \, dt. \tag{24}$$

The left-hand side can again be bounded from *below* by using the mixing condition. The first term on the right is obtained by integrating (10a) from t_0 to $t_0 + T$. Thus

$$\int_{t_0}^{t_0+T} (x_t' r_t)^2 \, dt = \frac{1}{2K} [\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2]. \tag{25}$$

The other two terms are bounded from *above* through use of the system (8a) and repeated application of the Schwarz inequality. The details are given in Appendix III. The result is

$$\|r_{t_0+T}\| \leqslant \sqrt{(1 - \rho)} \|r_{t_0}\| \tag{26}$$

where

$$\rho = \frac{2\alpha KT}{1 + KL^2 T + \frac{1}{2} K^2 L^4 T^2}. \tag{27}$$

Thus the bound (22) follows, with b_1 replaced by

$$b_2 = -\frac{1}{2T} \ln (1 - \rho). \tag{28}$$

We summarize the above results in the following.

Theorem 1: Suppose x_t satisfies the mixing condition (11), and (13). Then the solution (8a) satisfies the following:

$$\|r_t\| \leqslant a e^{-b(t-t_0)} \|r_{t_0}\|, \quad t \geqslant t_0.$$

Here $a = e^{bT}$ and $b = \max (b_1, b_2) > 0$, with $b_1 = -\{\ln (1 - s_o^2)\}/2T$ and $b_2 = -\{\ln (1 - \rho)\}/2T; s_o$ is the unique positive root of the quadratic (19), and ρ is explicitly given in (27).

2) Dependence of Bound on K: Assuming α, L^2, and T fixed, the dependence on K of the exponential decay parameters b_1 and b_2 can be explicitly computed. Note first the behavior as $K \to 0$ and $K \to \infty$ respectively. From (27), as $K \to 0$, $\rho \to 2\alpha KT$, and as $K \to \infty$, $\rho \to 4\alpha/(KL^4 T)$. In these limiting cases b_2 is clearly just $\rho/2T$.

By inspection of the quadratic (20), it is easily established that b_1 has the same qualitative behavior. Plots of $bT = T \max (b_1, b_2)$ are shown in Fig. 2 as functions of KTL^2 for various values of α/L^2. To our knowledge these are the only known bounds valid for all values of K.

Fig. 2 can be used to adjust K so as to guarantee the upper bound with the fastest convergence. Note that α and L^2 are both, in general, functions of T. If these functions are unknown, then the optimum K is obtained from the curve corresponding to the given value of α/L^2. The value of KTL^2 where this curve peaks, gives the optimum K and the corresponding value of b gives the best exponential bound. If the
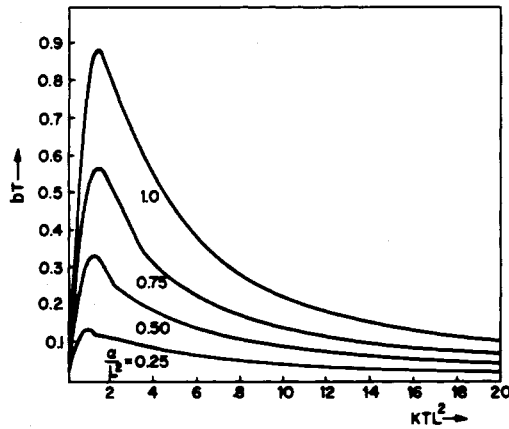
Fig. 2. Upper bound for convergence rate (Section II-B-2): Curves show bT versus KTL^2 for various values of α/L^2.

dependence of α and L^2 on T is known, this procedure must be repeated for various values of T till the optimum is found.

3) Comparison of Bounds to Actual Behavior: It is natural to ask at this point whether or not the worsening of the convergence rate of the bounds for large K is a technical limitation of our methods. Or do these bounds reflect the qualitative aspects of the actual convergence process? Further, even if such behavior cannot be ruled out, is it exceptional and, perhaps, dependent on some pathological property of the input process?

Here and in Section II-B-4, we provide some answers. We begin by synthesizing a realistic and well-behaved class of input vectors $\{x_t\}$ in which x_t satisfies the mixing condition and is such that for almost all initial conditions, $\|r_t\|$ decays asymptotically as exp $(-ct)$. Furthermore, c is proportional to K as $K \to 0$ and to $1/K$ as $K \to \infty$. Thus any bound which uses no more information about x, *must* have the same *qualitative* behavior as the ones derived here.

Our proof of this fact started with an observation by H. J. Landau, for which we are grateful to him. For a two dimensional system, Landau noted that if $x \equiv [a \cos (\omega t + \theta)$, $a \sin (\omega t + \theta)]$, and $\|r\| = $ esp $(-\varphi(t))$, then for certain initial conditions $\varphi(t) = ct$, where c has the above dependence on K. Although Landau's proof does not appear to generalize, it is nevertheless possible to generalize the input to N dimensions. Specifically, let $N = 2P$, and let $\omega_1 \cdots \omega_p$ be arbitrary distinct, nonzero frequencies and $a_1 \cdots a_p$ arbitrary nonzero constants. Then if the components of x_t are, with arbitrary θ_i's:

$$x_t(2i - 1) = a_i \sin (\omega_i t + \theta_i), \quad i = 1, \cdots, P$$

$$x_t(2i) = a_i \cos (\omega_i t + \theta_i), \quad i = 1, \cdots, P \quad (29)$$

we can show that for almost all initial conditions, $\|r_t\|$ asymptotically decays as exp $(-ct)$ with c as above. In fact, the same conclusion holds when the input vector is generalized to Mx_t, where x_t is defined in (29) and M is *any* fixed orthogonal matrix (i.e., $M'M = I$). This is because the solutions to (8a) for the two cases (corresponding to inputs x_t and Mx_t, respectively) are themselves related through M. All these results also hold for $N = 2P + 1$, with a constant component $x_t(2P + 1)$ appended to the ones in (29).

What makes the generalization from 2 to arbitrary dimension N possible, is the fact that when x_t is as defined by (29), the error e_t is obtained from the solution of a set of P *linear*

second-order constant-coefficient differential equations. Thus e_t (and hence from (10a), $\|r_t\|$) can be explicitly solved in terms of the roots, λ_i, of the characteristic equation of this system of differential equations. Appendix IV shows the derivation of the differential equation and the properties of the characteristic roots. In summary:

As $K \to 0$, the characteristic roots become $-Ka_i^2/2 \pm j\omega_i$, $i = 1, 2, \cdots, P$.

As $K \to \infty$, there are two real roots $\lambda_o = -K\Sigma a_i^2$ and $\lambda_1 = -(1/K)(1/\Sigma(a_i^2/\omega_i^2))$; the rest are complex conjugate, $-\mu_i/K \pm j\bar{\omega}_i$, $i = 1, \cdots, P - 1$. Here the μ's and $\bar{\omega}$'s are positive constants independent of K, and the $\bar{\omega}_i$'s interlace with the ω_i's.

Now $e(t)$ is a linear combination of the form $\Sigma d_i e^{\lambda_i t}$ where d_i depend on the initial conditions, and λ_i are the characteristic roots with the above behavior. Depending upon the initial conditions, some of the d_i's may be zero.

As $K \to 0$, therefore, the envelope of $|e_t|$ asymptotically decays as exp $(-\frac{1}{2} Ka_{\min}^2 t)$ where $a_{\min}^2$ is the smallest of the numbers $a_1^2, \cdots, a_p^2$.

As $K \to \infty$ the envelope of $|e_t|$ decays asymptotically as exp $(-\mu t/K)$, where μ is the smallest of the numbers μ_i and λ_1. The only exceptional case arises if the initial conditions are chosen so as to eliminate *all* but the exp $(\lambda_o t)$ mode.

From (10a) it follows that the asymptotic value of $\|r_t\|$ has the same envelope as $|e_t|$, which completes the proof.

In conclusion, we might mention that the existence of a value of K beyond which convergence becomes worse, has been observed in simulation studies [31]. A remark to this effect is also made in [32].

4) Lower Bound: A simple lower bound is obtained by using the Schwarz inequality to bound the right side of (10a). Thus

$$\frac{1}{2} \frac{d}{dt} \|r\|^2 = -K(r'x)^2$$

$$\geq -K \|r\|^2 \|x\|^2. \quad (30)$$

From (30) we get

$$\|r_t\| \geq \|r_{t_o}\| \exp \left(-K \int_{t_o}^t \|x_\tau\|^2 \, d\tau\right), \quad t \geq t_o. \quad (31)$$

As $K \to 0$, this bound has the same qualitative behavior as our upper bound, so we would not expect too much room for improvement. As $K \to \infty$, however, this lower bound goes to zero at an ever increasing rate, in sharp contrast to the behavior of the upper bound.

To see what improvement might be possible at large K, it is instructive to solve (8a) for the case when x is a constant vector. In that case if a_t is the component of r_t along x_t and b_t the component perpendicular to it, then it is easily seen that b_t is constant and $\|a_t\|$ goes to zero as $\|a_o\| \exp (-Kx'xt)$.

A constant vector x is, of course, disallowed by the mixing condition. However, it is clear that if x has a bounded derivative it cannot change appreciably during a time interval of order $1/K$, for sufficiently large K. Thus for large K, the component of r_t in the initial direction of x_t goes rapidly to zero, after which r converges to zero at a slower rate, which decreases as K is increased. This is just the behavior displayed by our improved lower bounds. The derivation proceeds as follows. With the definition

$$f_t \triangleq (x_t'r_t)/\|r_t\| \quad (32)$$

and assuming that x_t is continuously differentiable, a simple derivation using (8a) and (10a) gives the differential equation satisfied by f_t:

$$\frac{d}{dt} f_t = K f_t^3 - K(x_t' x_t) f_t + \dot{x}_t' r_t)/\|r_t\|. \tag{33}$$

Also, in terms of f_t we have

$$\frac{d}{dt} \|r_t\| = -K f_t^2 \|r_t\| \tag{34}$$

whence

$$\|r_t\| = \|r_{t_o}\| \exp\left[-K \int_{t_o}^{t} f_\tau^2 \, d\tau\right], \quad t \geqslant t_o. \tag{35}$$

We will now show that (33) implies an upper bound on f_t, which when substituted into (35) gives a lower bound for $\|r_t\|$. We will assume that for all t, there exist positive constants m and l, such that

$$\|\dot{x}_t\| \leqslant m \tag{36}$$

and

$$\|x_t\| \geqslant l. \tag{37}$$

From these assumptions and (33), it follows that

$$\frac{1}{2} \frac{d}{dt} f_t^2 \leqslant |f_t| \, [K|f_t|^3 - K l^2 |f_t| + m] \triangleq |f_t| \, G_K (|f_t|). \tag{38}$$

We have denoted the cubic polynomial with a parameter K by $G_K(\cdot)$.

The central idea in the derivation of our bound is contained in a boundedness property of f_t and certain simple properties of the cubic G_K which are summarized in Propositions 1 and 2, respectively. First, from (38) we have that whenever $G_K(|f_t|) \leqslant 0$ it follows that $df_t^2/dt \leqslant 0$. This together with the continuity of f_t allows us to bound f_t as follows.

Proposition 1: Suppose C is a positive constant such that $G_K(C) \leqslant 0$. Then $|f_{t_o}| \leqslant C$ implies $|f_t| \leqslant C$ for all $t \geqslant t_o$.

To proceed to our final bound we have to determine whether there are bounds on f_t, like C in the above proposition, which have the additional property of being proportional to $1/K$. The following shows that under certain conditions such bounds do exist.

Proposition 2: Let $\overline{K}$ be any constant satisfying $\overline{K} > (3\sqrt{3}/2) (m/l^3)$. Then: i) $G_{\overline{K}}(\cdot)$ has exactly two positive roots, $g_{\overline{K}}^{(1)}$ and $g_{\overline{K}}^{(2)}$, and the roots are distinct; say $0 < g_{\overline{K}}^{(1)} < g_{\overline{K}}^{(2)}$, and ii) define the positive constant a in terms of $\overline{K}$ thus

$$a \triangleq \overline{K} g_{\overline{K}}^{(1)}. \tag{39}$$

Then for all $K \geqslant \overline{K}$, $G_K (\frac{a}{K}) \leqslant 0$.

Proof: $G_K(\sigma)$, for $\sigma \geqslant 0$, always has a unique minimum (at $\sigma = l/\sqrt{3}$). Also, the value at the minimum is negative whenever $K > (3\sqrt{3}/2) [(m/l^3)$ in which case there exist two positive, real, distinct roots and i) follows. Further, $G_K(\sigma) > 0$ for $\sigma < g_k^{(1)}$ and $G_K(\sigma) < 0$ for $g_K^{(1)} < \sigma < g_K^{(2)}$. Also, for $K \geqslant \overline{K}$

$$G_K \left(\frac{a}{K}\right) = \frac{a^3}{K^2} - l^2 a + m \leqslant \frac{a^3}{\overline{K}^2} - l^2 a + m = G_{\overline{K}} \left(\frac{a}{\overline{K}}\right) = 0$$

so that ii) is proved. Q.E.D.

We summarize our lower bound in the following theorem.

Theorem 2: Suppose x_t has a continuous derivative for all t, $\|\dot{x}_t\|$ satisfies the uniform upper bound in (36), and $\|x_t\|$ satisfies the uniform lower bound in (37).

Let $\overline{K}$ be any constant satisfying $\overline{K} > (3\sqrt{3}/2)$ (m/l^3) and define the positive constant a in terms of $\overline{K}$ as in (39). Then for $t \geqslant t_o$

$$\|r_t\| \geqslant \|r_{t_o}\| \exp\left[-\frac{a^2}{K} (t - t_0)\right] \tag{40}$$

for all $K \geqslant \overline{K}$ and all initial conditions satisfying $|x_{t_o}' r_{t_o}|/ \|r_{t_o}\| \leqslant a/K$.

Proof: Identify a/K with C in Proposition 1 and note that we have shown in Proposition 2 that $G_K(C) \leqslant 0$ for all $K \geqslant \overline{K}$. From the first proposition, for $t \geqslant t_o$:

$$|f_t| = |x_t' r_t|/\|r_t\| \leqslant \frac{a}{K}, \tag{41}$$

so that from (35)

$$\|r_t\| \geqslant \|r_{t_o}\| \exp\left[-K \int_{t_o}^{t} f_\tau^2 d\tau\right] \geqslant \|r_{t_o}\| \exp\left[-\frac{a^2}{K} (t - t_o)\right],$$

which completes the proof. Q.E.D.

Using basically similar ideas and (33) and (35), we can derive another lower bound with a different emphasis, which provides additional insight into the actual convergence process. To derive this lower bound we construct an auxiliary continuous function $\overline{f}_t$ for $t \geqslant t_o$ which is the solution to the differential equation

$$\frac{d}{dt} \overline{f}_t = K \overline{f}_t^3 - K l^2 \overline{f}_t + m$$

$$= G_K(\overline{f}_t), \quad \text{with } \overline{f}_{t_o} = |f_{t_o}|. \tag{42}$$

The key observation is that

$$|f_t| \leqslant \overline{f}_t, \quad t \geqslant t_o. \tag{43}$$

This follows from inequality (38) and the fact that

$$|f_t| = \overline{f}_t \text{ implies that } \frac{d}{dt} f_t^2 \leqslant \frac{d}{dt} \overline{f}_t^2. \tag{44}$$

Inequalities (43) and (35) thus give a lower bound on $\|r_t\|$ which may be summarized in the following theorem.

Theorem 3: Suppose $\|x_t\|$ and $\|\dot{x}_t\|$ satisfy the uniform bounds (36) and (37). Then

$$\|r_t\| \geqslant \|r_{t_o}\| \exp\left(-K \int_{t_o}^{t} \overline{f}_\tau^2 \, d\tau\right), \quad t \geqslant t_o \tag{45}$$

where $\overline{f}_t$ is the solution of (42).

The bound of Theorem 3 is, of course, valid for all K and $|f_{t_o}|$. However, its qualitative properties are of greatest interest if K is restricted as in Theorem 2 (i.e., $K > 3\sqrt{3}$ $m/2l^3$) and

$$|f_{t_o}| < g_K^{(2)}. \tag{46}$$

If K and $|f_{t_o}|$ are restricted in this manner, then as t goes from t_o to ∞, $\overline{f}_t$ goes monotonically from $|f_{t_o}|$ to $g_K^{(1)}$. This assertion is a simple consequence of the properties of the cubic $G_K(\cdot)$ derived in Proposition 2. Thus for $|f_{t_o}| < g_K^{(1)}$, since $G_K(\sigma) > 0$ for $\sigma < g_K^{(1)}$, it follows from (42) that $\overline{f}_t$ *increases*
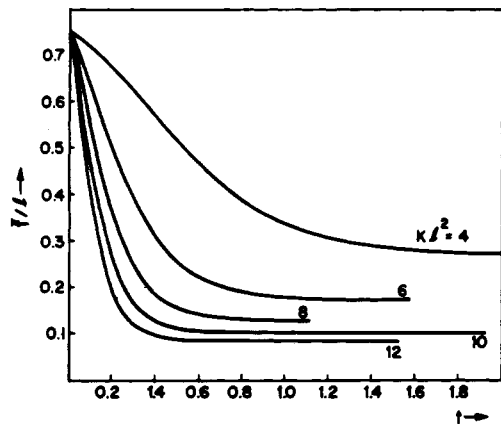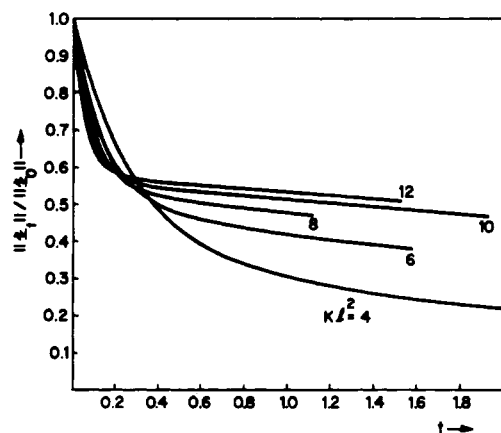
Fig. 3. (a) Lower bound for convergence rate (Section II-B-4): Curves show $\bar{f}_t/l$ versus t for $m/l = 1, \bar{f}_0 = 0.75$ and various values of Kl^2. (b) Lower bound (Section II-B-4): Curves show $\|r_t\|/\|r_0\|$ versus t for $m/l = 1, f_0 = 0.75$ and various values of Kl^2.

monotonically to $g_K^{(1)}$. Similarly $\bar{f}_t$ *decreases* monotonically to $g_K^{(1)}$ if $g_K^{(1)} \leqslant |f_{t_0}| < g_K^{(2)}$. Therefore, the bound of Theorem 3 is initially [i.e., for small $(t - t_0)$] proportional to $\exp[-Kf_{t_0}^2 (t - t_0)]$ and asymptotically [i.e., for large $(t - t_0)$] to $\exp[-K(g_K^{(1)})^2 (t - t_0)]$. As $g_K^{(1)} \sim m/Kl^2$ for large K, the asymptotic behavior is much like that of the upper bound of Section II-B-1. Sample plots computed from (42) and (45) are shown in Fig. 3.

5) Analysis With Forcing Term Included: In this paragraph we will estimate the effect of the inclusion of the forcing term u_t. Since (9a) is a linear time-varying equation, the general solution of the equation is [33],

$$r_t = \Phi(t, o)r_0 + \int_0^t \Phi(t, \tau)u_\tau \, d\tau, \quad t \geqslant 0 \quad (47)$$

where $\Phi(t, \tau)$ is the transition matrix of the equation (8a). Now the upper bound on the solution of the homogeneous equation, given in Theorem 1 (Section II-B-1) is equivalent to [7]

$$\|\Phi(t, \tau)\| \leqslant \exp[-b(t - \tau - T)], \quad t \geqslant \tau. \quad (48a)$$

(Here b is the largest of the decay constants of (23) and (28) and T is the constant associated with the mixing condition.)

[7] For a matrix M we denote by $\|M\|$ the matrix norm induced by the Euclidean vector norm.

The derivation of the results that we obtain here are actually simplified if we also use the earlier results in (17) and (26) which imply that

$$\|\Phi(t, \tau)\| \leqslant \exp[-b(i - 1)T], \quad \text{for } (i - 1)T \leqslant t - \tau < iT,$$
$$\text{and } i = 1, 2, \cdots. \quad (48b)$$

Thus (47) gives

$$\|r_t\| \leqslant \|\Phi(t, o)\| \, \|r_0\| + \int_0^t \|\Phi(t, \tau)\| \, \|u_\tau\| \, d\tau$$

and, on using (48a),

$$\|r_t\| \leqslant \|r_0\| e^{-b(t-T)} + v_t \quad (49)$$

where

$$v_t \triangleq \int_0^t \|\Phi(t, \tau)\| \, \|u_\tau\| \, d\tau.$$

As we shall see below, the specification of u_τ will generally depend on the application. In one of the cases considered below u_τ will be known as a function of τ. Most often, however, all that is known about u_τ is a bound in $\|u_\tau\|$, or on the time averaged value of $\|u_\tau\|$. Suppose then that either

$$\frac{1}{T} \int_t^{t+T} \|u_\tau\| \, d\tau \leqslant U, \quad t \geqslant 0$$

or

$$\|u_\tau\| \leqslant \bar{u}, \quad t \geqslant 0.$$

For these specifications v_t in (49) can be uniformly bounded by using (48b)

$$v_t \leqslant U \frac{T}{1 - e^{-bT}}, \quad t \geqslant 0 \quad (50a)$$

and

$$v_t \leqslant \bar{u} \frac{T}{1 - e^{-bT}}, \quad t \geqslant 0 \quad (50b)$$

respectively. The above, then, is a general bound on the norm of the ultimate $(t \to \infty)$ misalignment vector $(= h - \hat{h}_t)$, which is sometimes referred to as the *residual error*.

6) Applications of Bounded-Input Bounded-Output Result: As mentioned in Section I-A, a number of practical limitations to the implementation of the algorithm (4a) can be represented by the term u_t in (9a). We mention here three such limitations and use (49) and (50) to bound the perturbations due to these limitations.

a) Additional component in the observed signal: For a variety of reasons, some of which are mentioned in Section I-A, it is desirable to bound the effect of a signal s_t which appears as an additional component of the measured signal z_t. Equation (1) is replaced by

$$z_t = h'x_t + s_t.$$

Here we shall assume that

$$\frac{1}{T} \int_t^{t+T} s_\tau^2 \, d\tau \leqslant S^2, \quad t \geqslant 0.$$

Tracing the effect of the additional component through (2)–(4), we see that (9a) results with

$$u_t = -Ks_t x_t. \tag{51}$$

Note that in this case we have, by an application of Schwarz's inequality

$$\frac{1}{T} \int_t^{t+T} \| u_T \| \, d\tau \leqslant KSL$$

where L is the average power bound (13) on the input x_t. From (50a), therefore, the residual error is bounded thus:

$$v_t \leqslant KSL \frac{T}{1 - e^{-bT}}. \tag{52}$$

Recalling from Section II-B-1 that b is proportional to K as $K \to 0$, and to $1/K$ as $K \to \infty$, we see that the bound in (52) is independent of K for small K, but grows as K^2 for large K. The above behavior, with respect to K, is to be expected at least when the correlations between s_t and the components of x_t are large. When these correlations are large this behavior, for small K, is in agreement with results from the method of averaging (Section I-C).

b) Variations in the coefficient vector: As mentioned in the introduction, the algorithms considered here have an ability to track slow variations in the target vector h. An explicit bound on the error due to the variation in h may be obtained by noting that in going from (4a) to (8a) dh/dt was assumed to be zero. If this had not been done, we would have obtained (9a) with

$$u_t = \frac{d}{dt} h_t. \tag{53}$$

Again if $\| \dot{h} \|$ is bounded by $\bar{h}$, we get from (50) that

$$v(t) \leqslant \bar{h} \cdot \frac{T}{1 - e^{-bT}}. \tag{54}$$

In this case the bound increases both as $K \to 0$ and as $K \to \infty$. This is to be expected because we are bounding worst-case behavior. When K is very large or very small, the convergence is slow and h could change by large amounts before the system has a chance to converge.

c) Imperfect integrators: Any practical implementation of (4a) would, of course, use integrators. Thus the algorithm would be implemented as

$$\hat{h}_t = \hat{h}_o + K \int_0^t e_\tau x_\tau \, d\tau. \tag{55}$$

However, perfect integration is impossible in practice. If the integrators have exponential impulse response $e^{-\beta t}$ (i.e., the so-called RC integrators), then the effect of this imperfection is to change (4a) to

$$\frac{d}{dt} \hat{h} + \beta \hat{h} = K(r'x)x. \tag{56}$$

Subtracting βh from both sides gives

$$\frac{d}{dt} r + \beta r = -K(r'x)x + \beta h. \tag{57}$$

Equation (57) reduces to the form (9a), if we define $q_t \triangleq e^{\beta t} r_t$.

Substitution of this into (57) gives

$$\frac{d}{dt} q = -K(q_t' x_t)x_t + \beta h e^{\beta t}. \tag{58}$$

This now has the form we have studied above. Thus as in (49)

$$\| q_t \| \leqslant \| q_0 \| e^{-b(t-T)} + \beta \| h \| \int_0^t \| \Phi(t, \tau) \| e^{\beta \tau} \, d\tau \tag{59}$$

and, using (48b), for $NT \leqslant t < (N+1)T$ and $N = 0, 1, 2, \cdots$,

$$\| q_t \| \leqslant \| q_0 \| e^{-b(t-T)} + e^{\beta t} \| h \| (1 - e^{-\beta T}) \\ \cdot \frac{(1 - e^{-(N+1)(b+\beta)T})}{(1 - e^{-(b+\beta)T})}. \tag{60}$$

Finally, multiplying out by $e^{-\beta t}$ gives the bound:

$$\| r_t \| \leqslant \| r_0 \| e^{bT} e^{-(b+\beta)t} + \| h \| \\ \cdot \frac{(1 - e^{-\beta T})(1 - e^{-(N+1)(b+\beta)T})}{(1 - e^{-(b+\beta)T})}. \tag{61}$$

The residual error (as $t \to \infty$ and consequently $N \to \infty$) is

$$\| h \| (1 - e^{-\beta T})/(1 - e^{-(b+\beta)T}).$$

The residual error is thus proportional to $\| h \|$ and goes to zero as $\beta \to 0$ (i.e., as the integrators become more nearly ideal). At the other extreme, as $\beta \to \infty$, as expected, $\| r_t \| \to \| h \|$. Also for small T, the asymptotic bound is $\beta \| h \| / (b + \beta)$.

C. Results for the Nonlinear Algorithm

All but one of the results derived in Section II-B generalize to (8b) and (9b), respectively. The only exception is that we have not attempted to solve these equations explicitly for any general class of inputs such as the one defined in (29).

Qualitatively the results are very similar to those for the linear case. Most of the derivations are also similar. Therefore, we need only point out the few places where differences arise.

1) Upper Bound: When r_t is a solution of (8b), an upper bound for $\| r_t \|$ can be obtained by first transforming (8b) into an equation of the same form as (8a), and then using the results of Section II-B-1. Such a transformation is possible by the following simple observation. For $r_t' x_t \neq 0$, equation (8b) can be rewritten as

$$\frac{d}{dt} r_t = -K \frac{F(r_t' x_t)}{r_t' x_t} (r_t' x_t)x_t. \tag{62}$$

Defining $A(t)$ as

$$A(t) \triangleq F(r_t' x_t)/(r_t' x_t) \tag{63}$$

we may write (62) as

$$\frac{d}{dt} r_t = -KA(t)(r_t' x_t)x_t. \tag{64}$$

From the sector condition (5), at all points t such that $r_t' x_t \neq 0$, we have

$$0 < \gamma_1 \leqslant A(t) \leqslant \gamma_2. \tag{65}$$

At the time instants for which $r_t' x_t$ is zero, (64) is identical to (8b) with any finite value for $A(t)$, in particular the value

$\lim_{\sigma \to 0} F(\sigma)/\sigma$. With this specification, $A(t)$ satisfies (65) for all t. Defining the vector

$$y_t \triangleq \sqrt{A_t} x_t \qquad (66)$$

(64) becomes

$$\frac{d}{dt} r_t = -K(r_t' y_t) y_t \qquad (67)$$

which is the same as (8a) with x_t replaced by y_t. Further, if x_t satisfies the mixing condition then so does y_t. In fact, from (65), (11) and (13)

$$\frac{1}{T} \int_t^{t+T} (d' y_\tau)^2 \, d\tau \geqslant \gamma_1 \alpha \|d\|^2 \qquad (68)$$

and

$$\frac{1}{T} \int_t^{t+T} \|y_\tau\|^2 \, d\tau \leqslant \gamma_2 L^2 \qquad (69)$$

for all t and all fixed vectors d. Thus all the analysis of Section II-B-1 applies with α replaced by $\gamma_1 \alpha$ and L^2 by $\gamma_2 L^2$. In particular the bound of Theorem 1 follows immediately with the identification

$$b = \max (b_3, b_4) > 0 \qquad (70)$$

where

$$b_3 = -\{\ln (1 - s_1^2)\}/2T \qquad (71)$$

$$b_4 = -\{\ln (1 - \rho_1)\}/2T \qquad (72)$$

and s_1 and ρ_1 are obtained from s_0 and ρ, respectively, by replacing α by $\gamma_1 \alpha$ and L^2 by $\gamma_2 L^2$.

When $\gamma_1 = \gamma_2 = 1$, the value of b in (70) coincides with that for the linear case. Thus for $\gamma_2 \approx \gamma_1$, this is the best value we can attain, unless we can find improved bounds for the linear case. However, when γ_2 is much larger than γ_1, a much better bound can be obtained by a generalization of the method based on (24). Again the left-hand side is bounded from *below* by the mixing condition. The first and second terms on the right can be bounded from *above* by a procedure very similar to that used in the linear case. Finally, the third term is bounded from *above* in terms of the bounds on the first two terms, by using Schwarz's inequality. The details are listed in Appendix V, which has a pattern very similar to that of Appendix III. The result is again (26), with ρ replaced by

$$\rho_2 = \frac{4\alpha \gamma_1 KT}{[\sqrt{2} + \sqrt{\gamma_1 \gamma_2}\, KL^2 T]^2}. \qquad (73)$$

Thus finally we may replace (70) by

$$b = \max (b_3, b_4, b_5) > 0 \qquad (74)$$

where b_3 and b_4 are as in (71) and (72) and

$$b_5 = -\{\ln (1 - \rho_2)\}/2T. \qquad (75)$$

It is easily verified that for any given values of γ_1 and γ_2, $b_5 \geqslant b_4$ for sufficiently large K. Also, if $\gamma_2 > 2\gamma_1$, then $b_5 > b_4$ for all K.

2) Lower Bound: The derivation is again similar to that in the linear case. Thus applying the condition (5), and Schwarz's

inequality to the right side of (10b), we get

$$\frac{d}{dt} \|r_t\|^2 \geqslant -2\gamma_2 K e_t^2 \geqslant -2\gamma_2 K \|r_t\|^2 \|x_t\|^2 \qquad (76)$$

from which it follows that

$$\|r_t\| \geqslant \|r_{t_0}\| \exp \left(-\gamma_2 K \int_{t_0}^t \|x_\tau\|^2 \, d\tau\right). \qquad (77)$$

This reduces to the bound of (31) for the linear case, when $\gamma_2 = 1$.

Again as before, defining $f_t = (x_t' r_t)/\|r_t\|$ a differential equation satisfied by f_t is first derived and from this equation we obtain

$$\frac{1}{2} \frac{d}{dt} f_t^2 = K |f_t| \frac{|F(r_t' x_t)|}{\|r_t\|} (f_t^2 - x_t' x_t) + f_t \frac{(\dot{x}_t, r_t)}{\|r_t\|}. \qquad (78)$$

The corresponding equation for $\|r_t\|$ is

$$\frac{d}{dt} \|r_t\| = -K f_t F(r_t' x_t). \qquad (79)$$

The sector condition (5) then gives

$$\frac{d}{dt} \|r_t\| \geqslant -K\gamma_2 f_t^2 \|r_t\| \qquad (80)$$

from which

$$\|r_t\| \geqslant \|r_{t_0}\| \exp \left(-K\gamma_2 \int_{t_0}^t f_\tau^2 \, d\tau\right), \quad t \geqslant t_0. \qquad (81)$$

The lower bound for $\|r_t\|$ again follows from an upper bound for f_t.

As in the linear case, we will assume the uniform bounds on $(x_t' x_t)$ and $\|\dot{x}_t\|$ given by (36) and (37). Thus from (78)

$$\frac{1}{2} \frac{d}{dt} f_t^2 \leqslant K |f_t| \frac{|F(r_t' x_t)|}{\|r_t\|} (f_t^2 - l^2) + m |f_t|. \qquad (82)$$

From the sector condition (5) it, therefore, follows that

$$\frac{1}{2} \frac{d}{dt} f_t^2 \leqslant \left[K\gamma_2 |f_t| \left(f_t^2 - \frac{\gamma_1}{\gamma_2} l^2\right) + m\right] |f_t|. \qquad (83)$$

Comparing (81) and (83) to (35) and (38), respectively, it follows that the main results derived for the linear equations, (40) in Theorem 2 and (45), are also valid for the nonlinear equations with K replaced by $K\gamma_2$ and l^2 replaced by $\gamma_1 l^2/\gamma_2$.

3) Analysis With Forcing Term Included: The bounded-input-bounded-output property follows trivially from our analysis of the linear case. This is because the transformation used in deriving (67) from (8b) may obviously be used to transform (9b) to

$$\frac{d}{dt} r_t = -K(r_t' y_t) y_t + u_t. \qquad (84)$$

Here y is as defined in (66). Equation (84) is of the same form as (9a). In view of (68) and (69), it is thus obvious that the entire analysis of Sections II-B-5 and II-B-6 applies to the present case verbatim. The only difference is the value of the param-

eter b. For the nonlinear case it has the value given in (74), while for the linear case it has the value given in Theorem 1.

APPENDIX I
THE DEPENDENCE OF THE ERROR ON THE LOOP GAIN

Here we will prove the assertion made in Section I-C that in the ideal algorithms (4a) and (4b), the error e_t defined in (3) can be made uniformly as small as desired, by making the loop gain K sufficiently large. We will prove this for (4b). The proof for (4a) then follows as a special case.

In this Appendix we will assume that x_t is continuously differentiable and satisfies the following bounds. For $t \geq 0$

$$\|\dot{x}_t\| \leq m < \infty \qquad (A1\text{-}1)$$

$$0 < l \leq \|x_t\| \leq L < \infty. \qquad (A1\text{-}2)$$

From (7), which represents e_t in terms of the misalignment vector r_t given in (6), we get

$$\frac{d}{dt} e_t = \dot{r}'_t x_t + r'_t \dot{x}_t. \qquad (A1\text{-}3)$$

Substituting for $\dot{r}_t$ from (4b) and multiplying (A1-3) by e_t gives

$$\frac{d}{dt} e_t^2 = -2K e_t F(e_t) \|x_t\|^2 + 2(r'_t \dot{x}_t)(r'_t x_t). \qquad (A1\text{-}4)$$

The right-hand side can be bounded from above as follows. The sector condition (5) and the bound (A1-2) give

$$-2K e_t F(e_t) \|x_t\|^2 \leq -2K \gamma_1 l^2 e_t^2. \qquad (A1\text{-}5)$$

The bounds (A1-1) and (A1-2) give

$$2(r'_t x_t)(r'_t \dot{x}_t) \leq 2 \|r_t\|^2 \|x_t\| \|\dot{x}_t\|$$
$$\leq 2 \|r_t\|^2 mL$$
$$\leq 2 \|r_o\|^2 mL \qquad (A1\text{-}6)$$

where the last step follows from (10).

Thus from (A1-4), (A1-5), and (A1-6) we get

$$\frac{d}{dt} e_t^2 + 2K \gamma_1 l^2 e_t^2 \leq 2mL \|r_o\|^2. \qquad (A1\text{-}7)$$

Multiplying (A1-7) by $\exp(2K\gamma_1 l^2 t)$ and integrating gives

$$e_t^2 \leq e_o^2 \exp(-2Kl^2 \gamma_1 t) + \frac{mL}{K\gamma_1 l^2} \|r_o\|^2, \quad t > 0. \qquad (A1\text{-}8)$$

Thus for every initial condition e_o, r_o, and arbitrary positive numbers t', ϵ, there exists a K' which depends on the above quantities and such that for $K \geq K'$ and $t \geq t'$:

$$e_t^2 \leq \epsilon. \qquad (A1\text{-}9)$$

As mentioned above, the derivation specializes to the linear equation (4a), by setting $\gamma_1 = 1$.

APPENDIX II
UPPER BOUND FOR LINEAR EQUATIONS: DERIVATION OF EQUATIONS (17)–(19)

To derive (17) we premultiply (15) by r'_{t_o+T}, which gives

$$r'_{t_o+T} \left[I + K \int_{t_o}^{t_o+T} x_t x'_t \, dt \right] r_{t_o+T} = r'_{t_o+T} r_{t_o} - K^2 r'_{t_o+T} v \qquad (A2\text{-}1)$$

where the vector v is defined in (16). Now from the mixing condition, the left-hand side of (A2-1) is bounded as follows:

$$r'_{t_o+T} \left[I + K \int_{t_o}^{t_o+T} x_t x'_t \, dt \right] r_{t_o+T} \geq (1 + K\alpha T) \|r_{t_o+T}\|. \qquad (A2\text{-}2)$$

Thus (A2-1) becomes

$$(1 + K\alpha T) \|r_{t_o+T}\|^2 \leq \|r_{t_o+T}\| \|r_{t_o}\| + K^2 |r'_{t_o+T} v| \qquad (A2\text{-}3)$$

where we have bounded the first term on the right by Schwarz's inequality. It remains to bound the second term from above. From Schwarz's inequality

$$|r'_{t_o+T} v|^2 \leq \int_{t_o}^{t_o+T} (r'_{t_o+T} x_t)^2 \, dt \int_{t_o}^{t_o+T} g_t^2 \, dt \qquad (A2\text{-}4)$$

where

$$g_t = \int_t^{t_o+T} (x'_t x_\tau)(x'_\tau r_\tau) \, d\tau. \qquad (A2\text{-}5)$$

Again, by Schwarz's inequality

$$g_t^2 \leq \|x_t\|^2 \int_t^{t_o+T} \|x_\tau\|^2 \, d\tau \int_t^{t_o+T} (x'_s r_s)^2 \, ds$$
$$= -\frac{1}{2} \frac{d}{dt} \left[\int_t^{t_o+T} \|x_\tau\|^2 \, d\tau \right]^2 \cdot \frac{1}{2K} [\|r_t\|^2 - \|r_{t_o+T}\|^2] \qquad (A2\text{-}6)$$

where we have used (10a) to evaluate the final integral. From (A2-6), therefore,

$$\int_{t_o}^{t_o+T} g_t^2 \, dt \leq \frac{1}{4K} \left[\int_{t_o}^{t_o+T} \|x_t\|^2 \, dt \right]^2 [\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2]$$
$$\leq \frac{1}{4K} L^4 T^2 [\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2]. \qquad (A2\text{-}7)$$

Substituting into (A2-4), and bounding the first integral on the right side of (A2-4) also by Schwarz's inequality, we get

$$|r'_{t_o+T} v|^2 \leq \|r_{t_o+T}\|^2 \cdot \frac{1}{4K} L^6 T^3 [\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2]. \qquad (A2\text{-}8)$$

Substituting (A2-8) into (A2-3), dividing out by $\|r_{t_o}\| \|r_{t_o+T}\|$, and defining

$$\gamma = \|r_{t_o+T}\|/\|r_{t_o}\| \qquad (A2\text{-}9)$$

we get

$$(1 + K\alpha T)\gamma \leqslant 1 + \tfrac{1}{2} \sqrt{K^3 T^3 L^6} \sqrt{1 - \gamma^2}. \qquad (A2\text{-}10)$$

As γ goes from 0–1, the left-hand side of (A2-10) goes monotonically from 0 to some positive number, and the right-hand side from some positive number to 0. There is, therefore, a unique positive number γ_o such that

$$(1 + K\alpha T)\gamma_o = 1 + \tfrac{1}{2} \sqrt{K^3 T^3 L^6} \sqrt{1 - \gamma_o^2}. \qquad (A2\text{-}11)$$

Clearly (A2-10) implies $\gamma \leqslant \gamma_o$. The derivation of (17) is thus complete, since (18) follows trivially from (A2-11) by substituting $\gamma^2 = 1 - s^2$.

The derivation of (19) is a bit more cumbersome, but uses essentially the same technqiues. As mentioned in the text, the step of replacing r_t by $r_{t_o+T} + r_t - r_{t_o+T}$ can be used once more on (15) to give

$$\left[I + K \int_{t_o}^{t_o+T} x_t x_t' \, dt + K^2 \int_{t_o}^{t_o+T} \int_t^{t_o+T} x_t x_t' x_\tau x_\tau' \, d\tau dt \right]$$
$$r_{t_o+T} = r_{t_o} - K^3 w \qquad (A2\text{-}12)$$

where

$$w \triangleq \int_{t_o}^{t_o+T} dt \int_t^{t_o+T} d\tau \int_\tau^{T+t_o} d\lambda \, x_t x_t' x_\tau x_\tau' x_\lambda x_\lambda' r_\lambda. \qquad (A2\text{-}13)$$

When (A2-12) is premultiplied by r_{t_o+T}', the first two terms on the left are identical to those in (A2-2). The third term is bounded as follows:

$$K^2 \int_{t_o}^{t_o+T} (r_{t_o+T}' x_t) x_t' \, dt \int_t^{t_o+T} x_\tau (r_{t_o+T}' x_\tau) \, d\tau$$
$$= \frac{K^2}{2} r_{t_o+T}' \left[\int_{t_o}^{t_o+T} x_t x_t' \right]^2 r_{t_o+T}$$
$$\geqslant \frac{K^2 \alpha^2 T^2}{2} \|r_{t_o+T}\|^2. \qquad (A2\text{-}14)$$

Thus (A2-12) gives

$$A\|r_{t_o+T}\|^2 \leqslant \|r_{t_o}\| \|r_{t_o+T}\| + K^3 |r_{t_o+T}' w| \qquad (A2\text{-}15)$$

where

$$A = (1 + K\alpha T + \tfrac{1}{2} K^2 \alpha^2 T^2). \qquad (A2\text{-}16)$$

It remains to bound $|r_{t_o+T}' w|$. From Schwarz's inequality:

$$|r_{t_o+T}' w|^2 \leqslant B \cdot C \qquad (A2\text{-}17)$$

where

$$B = \int_{t_o}^{t_o+T} (r_{t_o+T}' x_t)^2 \, dt \leqslant \|r_{t_o+T}\|^2 L^2 T \qquad (A2\text{-}18)$$

$$C = \int_{t_o}^{t_o+T} dt \left[\int_t^{t_o+T} (x_t' x_\tau) g_\tau \, d\tau \right]^2 \qquad (A2\text{-}19)$$

with g_τ as defined in (A2-5). Thus

$$C \leqslant \int_{t_o}^{t_o+T} dt \int_t^{t_o+T} (x_t' x_\tau)^2 \, d\tau \int_t^{t_o+T} g_\tau^2 \, d\tau. \qquad (A2\text{-}20)$$

Analogously to the derivation of (A2-6) we see that

$$\int_t^{t_o+T} g_\tau^2 \, d\tau \leqslant \frac{1}{2K} \left[\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2 \right]$$
$$\cdot \int_t^{t_o+T} d\tau \int_\tau^{t_o+T} (x_\tau' x_\lambda)^2 \, d\lambda \qquad (A2\text{-}21)$$

$$\therefore \; C \leqslant \frac{1}{4K} \left[\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2 \right] \left[\int_{t_o}^{t_o+T} dt \int_t^T (x_t' x_\tau)^2 \, d\tau \right]^2. \qquad (A2\text{-}22)$$

Calling the term in square brackets in (A2-22) as D, we see that

$$D \leqslant \int_{t_o}^{t_o+T} \|x_t\|^2 \, dt \int_t^T \|x_\tau\|^2 \, d\tau$$
$$= \int_{t_o}^{t_o+T} -\frac{1}{2} \frac{d}{dt} \left[\int_t^{t_o+T} \|x_\tau\|^2 \, d\tau \right]^2 dt$$
$$\leqslant \tfrac{1}{2} L^4 T^2 \qquad (A2\text{-}23)$$

$$C \leqslant \frac{L^8 T^4}{16K} \left[\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2 \right]. \qquad (A2\text{-}24)$$

Substituting (A2-18) and (A2-24) into (A2-17) gives

$$|r_{t_o+T}' w|^2 \leqslant \frac{L^{10} T^5}{16K} \|r_{t_o+T}\|^2 \left[\|r_{t_o}\|^2 - \|r_{t_o+T}\|^2 \right]. \qquad (A2\text{-}25)$$

Substituting (A2-25) into (A2-15), and defining γ as in (A2-9) we get

$$A\gamma \leqslant 1 + \tfrac{1}{4} \sqrt{K^5 L^{10} T^5} \sqrt{1 - \gamma^2}. \qquad (A2\text{-}26)$$

This equation is of the same form as (A2-10). Thus as before, $\gamma \leqslant \gamma_o$, where γ_o is the unique positive root of

$$(1 + K\alpha T + \tfrac{1}{2} K^2 \alpha^2 T^2)\gamma \leqslant 1 + \tfrac{1}{4} \sqrt{K^5 T^5 L^{10}} \sqrt{1 - \gamma^2}. \qquad (A2\text{-}27)$$

Trivially, (19) is seen to be equivalent to (A2-27) by substituting $\gamma^2 = 1 - s^2$.

APPENDIX III
UPPER BOUND FOR LINEAR EQUATIONS: DERIVATION OF EQUATION (26)

From the mixing condition, the left-hand side of (24) is bounded as follows:

$$\int_{t_o}^{t_o+T} (x_t' r_{t_o})^2 \, dt \geqslant \alpha T \|r_{t_o}\|^2. \qquad (A3\text{-}1)$$

The first term on the right-hand side of (24) has already been evaluated in (25). The second term on the right is bounded

as follows:

$$\int_T^{t_0+T} [x_t'(r_{t_0} - r_t)]^2 \, dt$$

$$\leqslant \int_{t_0}^{t_0+T} \|x_t\|^2 \|r_{t_0} - r_t\|^2 \, dt$$

$$= K^2 \int_{t_0}^{t_0+T} \|x_t\|^2 \left\| \int_{t_0}^t (r_s' x_s) x_s \, ds \right\|^2 dt$$

$$\leqslant K^2 \int_{t_0}^{t_0+T} \|x_t\|^2 \left(\int_{t_0}^t \|x_s\|^2 \, ds \right) \left(\int_{t_0}^t (r_s' x_s)^2 \, ds \right) dt$$

$$= \frac{K}{2} \int_{t_0}^{t_0+T} \|x_t\|^2 \left(\int_{t_0}^t \|x_s\|^2 \, ds \right) (\|r_{t_0}\|^2 - \|r_t\|^2) \, dt$$

(A3-2)

$$\leqslant \frac{K}{2} (\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2) \frac{1}{2} \left[\int_{t_0}^{t_0+T} \|x_s\|^2 \, ds \right]^2$$ (A3-3)

$$\leqslant \frac{KL^4 T^2}{4} (\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2)$$ (A3-4)

where (A3-2) follows from (10a) and (A3-3) from the nonincreasing property of $\|r_t\|$ also implied by (10a).

The third term on the right-hand side of (24) is bounded as follows:

$$2 \int_{t_0}^{t_0+T} (x_t' r_t) [x_t'(r_{t_0} - r_t)] \, dt$$

$$= 2K \int_{t_0}^{t_0+T} (x_t' r_t) x_t' \int_{t_0}^t (x_s' r_s) x_s \, ds \, dt$$

$$= K \left\| \int_{t_0}^{t_0+T} (x_t' r_t) x_t \, dt \right\|^2$$

$$\leqslant K \left(\int_{t_0}^{t_0+T} \|x_t\|^2 \, dt \right) \left(\int_{t_0}^{t_0+T} (x_t' r_t)^2 \, dt \right)$$

$$\leqslant \frac{KL^2 T}{2K} [\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2]$$

$$= \frac{L^2 T}{2} [\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2].$$ (A3-5)

Substituting (A3-1), (A3-4), (A3-5), and (25) into (24) gives

$$\alpha T \|r_{t_0}\|^2 \leqslant \left(\frac{1}{2K} + \frac{L^2 T}{2} + \frac{KL^4 T^2}{4} \right) [\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2].$$

(A3-6)

Upon rearranging terms

$$\|r_{t_0+T}\|^2 \leqslant \left[1 - \frac{2\alpha KT}{1 + KL^2 T + \frac{1}{2}(KL^2 T)^2} \right] \|r_{t_0}\|^2.$$ (A3-7)

Equation (A3-7) is equivalent to (26), with ρ defined in (27).

APPENDIX IV
PROPERTIES OF THE LINEAR SYSTEM IN RESPONSE TO THE SPECIAL INPUT

In this Appendix we will derive an explicit solution of (4a), when the input is the special vector defined in (29). We will also discuss some of the qualitative aspects of this solution, in particular, its dependence on the loop gain K. It will be convenient[8] to split up the input vector x and the misalignment vector r into P two-dimensional vectors, and to define partial errors for each such pair. Thus let

$$x_i \triangleq \{a_i \sin(\omega_i t + \theta_i), \quad a_i \cos(\omega_i t + \theta_i)\}', \quad i = 1, \cdots, P$$

(A4-1)

$$r_i \triangleq \{r(2i-1), r(2i)\}', \quad i = 1, \cdots, P$$ (A4-2)

$$e_i \triangleq r_i' x_i, \quad i = 1, \cdots, P$$ (A4-3)

Then the total error

$$e = \sum_1^P e_i$$ (A4-4)

and (4a) can be written as

$$\frac{d}{dt} r_i = -Kex_i, \quad i = 1, \cdots, P$$ (A4-5)

Using these equations we get

$$\dot{e}_i = \dot{r}_i' x_i + r_i' \dot{x}_i$$
$$= -Kea_i^2 + r_i' \dot{x}_i, \quad i = 1, \cdots, P$$ (A4-6)

where we used the fact that $x_i' x_i = a_i^2$. Differentiating (A4-6) we get

$$\ddot{e}_i = -Ka_i^2 \dot{e} + r_i' \ddot{x}_i$$
$$= -Ka_i^2 \dot{e} - \omega_i^2 e_i, \quad i = 1, \cdots, P$$ (A4-7)

From (A4-7) e_i, $i = 1, \cdots, P$ can be solved in terms of e and the initial conditions. The easiest way is via the Laplace transform. Noting that

$$e_i(0) = r_i'(0) x_i(0), \quad i = 1, \cdots, P$$ (A4-8)

and

$$\dot{e}_i(0) = -Ka_i^2 e(0) + \dot{x}_i'(0) r_i(0), \quad i = 1, \cdots, P$$ (A4-9)

we get, from (A4-7)

$$s^2 E_i(s) + Ka_i^2 sE(s) + \omega_i^2 E_i(s) = (sx_i(0) + \dot{x}_i(0))' r_i(0)$$

(A4-10)

where the upper case symbols denote the Laplace transforms of the corresponding lower case symbols. Summing (A4-10) over i and using (A4-4), we therefore, get

$$\left[1 + Ks \sum_i \frac{a_i^2}{s^2 + \omega_i^2} \right] E(s) = \sum_i \frac{r_i'(0)[\dot{x}_i(0) + sx_i(0)]}{s^2 + \omega_i^2}.$$

(A4-11)

[8] To keep the notation reasonably simple in this Appendix, we will not use the subscript t.

Equation (A4-11) gives $E(s)$ as a ratio of two polynomials in s. Thus $e(t)$ can be explicitly obtained in terms of the poles of $E(s)$; i.e., the zeros of the polynomial

$$Q_K(s) = \prod(s) + Ks \sum a_i^2 \prod_i(s). \qquad \text{(A4-12)}$$

Here

$$\prod(s) = \prod_{i=1}^{P} (s^2 + \omega_i^2) \qquad \text{(A4-13)}$$

and

$$\prod_i(s) = \prod_{\substack{j=1 \\ j \neq i}}^{P} (s^2 + \omega_j^2). \qquad \text{(A4-14)}$$

If $\lambda_i, i = 1, \cdots, 2P$ are the roots of $Q_K(s)$, then

$$e(t) = \sum_{i=1}^{2P} d_i e^{\lambda_i t} \qquad \text{(A4-15)}$$

where d_i are constants dependent on the initial values $x_i(0)$, $\dot{x}_i(0)$ and $r_i(0)$, $i = 1, \cdots, P$. The computation of the d_i's is straightforward though cumbersome. However, as pointed out in the text, we need not compute the λ's and the d's explicitly in order to infer the qualitative properties of $e(t)$ and $\|r(t)\|$. All we need is the qualitative behavior of the $\lambda_i's$ as functions of K.

i) *Behavior of the roots as $K \rightarrow 0$.* When K is small, the roots of $Q_K(s)$ are close to the roots of $\Pi(s)$, i.e., to the values $s = \pm j\omega_i$, where $j = \sqrt{-1}$. Thus, for small K the roots are

$$\lambda_i = \pm j\omega_i + \epsilon_i, \quad i = 1, \cdots, P. \qquad \text{(A4-16)}$$

The perturbations ϵ_i are obtained by setting

$$\epsilon_i \frac{\partial Q_K(s)}{\partial s} + K \frac{\partial Q_K(s)}{\partial K} = 0 \qquad \text{(A4-17)}$$

where the derivatives are evaluated at the point $K = 0$, $s = \pm j\omega_i$. This gives $\epsilon_i = -(K a_i^2)/2$. Thus as $K \rightarrow 0$

$$\lambda_i = -\frac{K a_i^2}{2} \pm j\omega_i, \quad i = 1, \cdots, P. \qquad \text{(A4-18)}$$

ii) *Behavior of the roots as $K \rightarrow \infty$.* When K is large, it is more convenient to study the roots of

$$\overline{Q}_K(s) = \frac{1}{K} + s \sum \frac{a_i^2}{s^2 + \omega_i^2}$$

$$\triangleq \frac{1}{K} + sq(s^2) \qquad \text{(A4-19)}$$

which, of course, are identical to those of $Q_K(s)$ for all finite K. By setting $s = K\sigma$ it is easily established from (A4-19), that one root is $\sim -K \sum a_i^2$. Similarly, by setting $s = \sigma/K$ it is established that a second root is $\sim -1/(K \sum a_i^2/\omega_i^2)$. The remaining roots are close to the roots of $q(s^2)$. If we set $s^2 = z$, it is seen that $q(s)$ has poles at $z = -\omega_i^2$, and $dq(z)/dz < 0$ for all real z. Then clearly $q(z)$ is zero for $z = -\overline{\omega}_i^2$, where $\overline{\omega}_i^2$ are positive numbers interlacing with the ω_i^2, i.e.,

$$\omega_i^2 < \overline{\omega}_i^2 < \omega_{i+1}^2, \quad i = 1, \cdots, P - 1. \qquad \text{(A4-20)}$$

Thus $q(s^2)$ is zero for $s = \pm j\overline{\omega}_i$ for $i = 1, \cdots, P - 1$, and therefore the roots of $\overline{Q}_K(s)$ close to these are $\epsilon_i \pm j\overline{\omega}_i$. The perturbations ϵ_i are obtained by an equation analogous to (A4-17);

however, with the small parameter K replaced by $1/K$, and Q by $\overline{Q}$. The result is

$$\epsilon_i = \frac{1}{2K\overline{\omega}_i^2 q'(-\overline{\omega}_i^2)} \triangleq -\frac{\mu_i}{K} \qquad \text{(A4-21)}$$

where the μ's are independent of K. In summary, therefore, the roots of $\overline{Q}_K(s)$ are, for large K

$$\lambda_1 \simeq -K \sum a_i^2$$

$$\lambda_2 \simeq -\frac{1}{K} \left[1/\sum (a_i^2/\omega_i^2) \right]$$

$$\lambda_{i+2} \simeq -\frac{\mu_i}{K} + j\overline{\omega}_i, \quad i = 1, \cdots, P - 1$$

$$\lambda_{i+1+P} \simeq -\frac{\mu_i}{K} - j\overline{\omega}_i, \quad i = 1, \cdots, P - 1. \qquad \text{(A4-22)}$$

These values of the characteristic roots give the qualitative properties of $e(t)$ and $\|r(t)\|$ derived in the text.

We observe parenthetically that the above analysis does not explicate the particular behavior of $e(t)$ with respect to K that is proven in Appendix I, for it is necessary to also have the dependence on K of the residues associated with the poles of the Laplace transform $E(s)$, which for the sake of brevity we do not give here.

APPENDIX V

UPPER BOUND FOR THE NONLINEAR EQUATION: DERIVATION OF EQUATION (73)

In this Appendix we will prove for the nonlinear equation (8b) the upper bound on the rate of convergence given by (73) and (26), i.e.,

$$\|r_{t_o + T}\|^2 \leqslant (1 - \rho_2) \|r_{t_o}\|^2 \qquad \text{(26)}$$

where

$$\rho_2 = 4\alpha\gamma_1 KT/[\sqrt{2} + \sqrt{\gamma_1 \gamma_2} KL^2 T]^2 > 0 \qquad \text{(73)}$$

and γ_1, γ_2 are the positive constants associated with the sector condition (5) on the nonlinearity F. The other assumptions are the mixing condition (11) and the average power bound (13).

Note that, from the mixing condition,

$$\alpha T \|r_{t_o}\|^2 \leqslant \int_{t_o}^{t_o + T} (x_t' r_{t_o})^2 \, dt$$

$$= \int_{t_o}^{t_o + T} (x_t' r_t)^2 \, dt + \int_{t_o}^{t_o + T} [x_t'(r_{t_o} - r_t)]^2 \, dt$$

$$+ 2 \int_{t_o}^{t_o + T} (x_t' r_t)[x_t'(r_{t_o} - r_t)] \, dt. \qquad \text{(A5-1)}$$

Each of the three terms in (A5-1) is now bounded from above. First

$$\int_{t_o}^{t_o + T} (x_t' r_t)^2 \, dt \leqslant \frac{1}{\gamma_1} \int_{t_o}^{t_o + T} (x_t' r_t) F(x_t' r_t) \, dt$$

$$= \frac{1}{2\gamma_1 K} (\|r_{t_o}\|^2 - \|r_{t_o + T}\|^2). \qquad \text{(A5-2)}$$

For the second term,

$$\int_{t_0}^{t_0+T} [x_t'(r_{t_0} - r_t)]^2 \, dt \leqslant \int_{t_0}^{t_0+T} \|x_t\|^2 \|r_{t_0} - r_t\|^2 \, dt$$

$$= K^2 \int_{t_0}^{t_0+T} \|x_t\|^2 \left\| \int_{t_0}^{t} F(x_s'r_s)x_s \, ds \right\|^2 dt$$

$$\leqslant K^2 \int_{t_0}^{t_0+T} \|x_t\|^2 \left(\int_{t_0}^{t} \|x_s\|^2 \, ds \right) \left[\int_{t_0}^{t} F^2(x_s'r_s) \, ds \right] dt$$

$$\leqslant \gamma_2 K \int_{t_0}^{t_0+T} \|x_t\|^2 \left(\int_{t_0}^{t} \|x_s\|^2 \, ds \right)$$

$$\cdot \left[K \int_{t_0}^{t} F(x_s'r_s)(x_s'r_s) \, ds \right] dt$$

$$= \frac{1}{2} \gamma_2 K \int_{t_0}^{t_0+T} \|x_t\|^2 \left(\int_{t_0}^{t} \|x_s\|^2 \, ds \right) \cdot (\|r_{t_0}\|^2 - \|r_t\|^2) \, dt$$

$$\leqslant \frac{1}{2} \gamma_2 K (\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2) \int_{t_0}^{t_0+T} \|x_t\|^2$$

$$\cdot \left(\int_{t_0}^{t} \|x_s\|^2 \, ds \right) dt$$

$$\leqslant \frac{1}{4} \gamma_2 K L^4 T^2 (\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2). \tag{A5-3}$$

Finally, the third term in (A5-1) is bounded by using the bounds (A5-2) and (A5-3). Note that because of the nonlinearity F, we are not able to bound this term as tightly as was possible in Appendix III for the linear equation.

$$2 \int_{t_0}^{t_0+T} (x_t'r_t)[x_t'(r_{t_0} - r_{t_0+T})] \, dt$$

$$\leqslant 2 \left[\int_{t_0}^{t_0+T} (x_t'r_t)^2 \, dt \int_{t_0}^{t_0+T} [x_t'(r_{t_0} - r_t)]^2 \, dt \right]^{1/2}$$

$$\leqslant 2 \left[\frac{\gamma_2}{\gamma_1} \frac{L^4 T^2}{8} \right]^{1/2} (\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2). \tag{A5-4}$$

The bounds for the three terms of (A5-1) yield,

$$\alpha T \|r_{t_0}\|^2 \leqslant \left[\frac{1}{2\gamma_1 K} + \frac{\gamma_2 K L^4 T^2}{4} + \frac{L^2 T}{\sqrt{2}} \sqrt{\frac{\gamma_2}{\gamma_1}} \right]$$

$$(\|r_{t_0}\|^2 - \|r_{t_0+T}\|^2)$$

which upon rearrangement give (26) and (73).

ACKNOWLEDGMENT

REFERENCES

[1] P. Eykhoff, *System Identification*, New York: Wiley, 1974, ch. 7, p. 228; ch. 9, p. 322; sec. 5.3.
[2] J. R. Blum, "Multidimensional stochastic approximation methods," *Ann. Math. Stat.*, vol. 25, no. 4, pp. 737–744, 1954.
[3] Y. Z. Tsypkin, "Dynamic adaptation algorithms," *Automat. Remote Contr.*, (USSR), vol. 33, no. 1, pt. 1, pp. 59–67, 1972.
[4] V. Dupac, "A dynamic stochastic approximation method," *Ann. Math. Stat.*, vol. 36, no. 6, pp. 1695–1702, 1965.
[5] R. DeFigueiredo, "Convergent algorithms for pattern recognition in nonlinearly evolving nonstationary environments," *Proc. IEEE* (Lett.), vol. 56, pp. 188–189, Feb. 1968.
[6] P. M. Lion, "Rapid identification of linear and nonlinear systems," in *Proc. 1966 Joint Automatic Control Conf.* (Seattle, WA), pp. 605–615.
[7] J. S. Pazdera and H. J. Pottinger, "Linear system identification via Liapunov design techniques," in *Proc. 1966 Joint Automatic Control Conf.* (Seattle, WA), pp. 795–801.
[8] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," in *IRE Wescon Conv. Rec.*, pt. 4, pp. 96–104, 1960.
[9] K. S. Narendra and L. E. McBride, "Multiparameter self-optimizing systems using correlation techniques," *IEEE Trans. Automat. Contr.*, vol. AC-9, pp. 31–38, Jan. 1964.
[10] B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive antenna systems," *Proc. IEEE*, vol. 55, pp. 2143–2159, Dec. 1967.
[11] R. W. Lucky, "Automatic equalization for digital communication," *BSTJ*, vol. 44, pp. 547–588, Apr. 1965.
[12] A. Gersho, "Adaptive equalization of highly dispersive channels for data communication," *BSTJ*, vol. 48, pp. 55–70, Jan. 1969.
[13] R. W. Lucky and W. R. Rudin, "An automatic equalizer for general-purpose communication channels," *Bell Syst. Tech. J.*, vol. 46, pp. 2179–2208, Nov. 1967.
[14] M. M. Sondhi, "An adaptive echo canceller," *Bell Syst. Tech. J.*, vol. 46, no. 3, pp. 497–511, 1967.
[15] M. M. Sondhi and A. J. Presti, "A self-adaptive echo canceller," *BSTJ*, vol. 45, no. 10, pp. 1851–1854, 1966.
[16] F. K. Becker and H. R. Rudin, "Application of automatic transversal filters to the problem of echo suppression," *BSTJ*, vol. 45, no. 10, pp. 1847–1850, 1966.
[17] S. J. Campanella, H. G. Suyerhoud, and M. Onufry, "Analysis of an adaptive impulse response echo canceller," *COMSAT Technical Rev.*, vol. 2, no. 1, pp. 1–38, Spring 1972.
[18] N. Demytko and L. K. Mackechnie, "A high speed digital adaptive echo canceller," *Australian Telecommun. Res.*, vol. 7, no. 1, pp. 20–28, 1973.
[19] D. Mitra and M. M. Sondhi, "Adaptive filtering with non-ideal multipliers: Applications to echo cancellation," in *IEEE Conf. Rec., Int. Conf. on Communications*, vol. II, pp. 30-11 to 30-15, June 1975.
[20] J. D. Gibson, S. K. Jones, and J. L. Melsa, "Sequentially adaptive prediction and coding of speech signals," *IEEE Trans. Commun.*, vol. COM-22, pp. 1789–1796, Nov. 1974.
[21] J. N. Maksym, "Real-time pitch extraction by adaptive prediction of the speech waveform," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 149–153, June 1973.
[22] S. Jones, "Adaptive filtering with correlated training samples," *Internal Report*, Bell Laboratories, 1972.
[23] J-K. Kim and L. D. Davisson, "Adaptive linear estimation for stationary M-dependent processes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 23–31, Jan. 1975.
[24] T. P. Daniell, "Adaptive estimation with mutually correlated training sequences," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-6, pp. 12–19, Jan. 1970.
[25] H. J. Kushner, "On the convergence of Lion's identification method with random inputs," *IEEE Trans. Automat. Contr.*, pp. 652–654, Dec. 1970.
[26] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communication," *IBM J. Res. Develop.*, vol. 16, no. 6, pp. 546–555, Nov. 1972.
[27] J. R. Rosenberger and E. J. Thomas, "Performance of an adaptive echo canceller operating in a noisy, linear, time-invariant environment," *Bell Syst. Tech. J.*, vol. 50, no. 3, pp. 785–813, Mar. 1971.
[28] D. P. Derevitskii and A. L. Fradkov, "Two models for analyzing the dynamics of adaptation algorithms," *Automatika i Telemekh.*, no. 1, pp. 67–75 (Translated), Jan. 1974.
[29] N. N. Bogoliubov and J. A. Mitropolskii, *Asymptotic Methods in the Theory of Nonlinear Oscillations*, New York: Gordon and Breach, 1961.
[30] R. Z. Khasminskii, "On stochastic processes defined by differential equations with a small parameter," *Theory Prob. Appl.* (USSR), vol. XI, no. 2, pp. 211–228, 1966.
[31] B. J. Miller, "A General Method of Computing System Parameters With an Application to Adaptive Control," *Conf. Paper, AIEE Winter General Meeting* (New York), Jan. 1962.
[32] R. E. Butler and E. V. Bohn, "An Automatic Identification Technique for a Class of Nonlinear Systems," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 292–296, Apr. 1966.
[33] J. L. Willems, *Stability Theory of Dynamical Systems*, New York: Wiley 1970, ch. 4.