An Ordinary Differential Equation Technique for Continuous-Time Parameter Estimation

Douglas G. DeWolf and Donald M. Wiberg

Abstract—**An ordinary differential equation technique is developed via averaging theory and weak convergence theory to analyze the asymptotic behavior of continuous-time recursive stochastic parameter estimators. This technique is an extension of L. Ljung's work in discrete time. Using this technique, the following results are obtained for various continuous-time parameter estimators. The recursive prediction error method, with probability one, converges to a minimum of the likelihood function. The same is true of the gradient method. The extended Kalman filter fails, with probability one, to converge to the true values of the parameters in a system whose state noise covariance is unknown. An example of the extended least squares algorithm is analyzed in detail. Analytic bounds are obtained for the asymptotic rate of convergence of all these estimators applied to this example.**

I. Introduction

Recursive parameter estimation of linear stochastic systems has many practical applications, and additionally is worth mathematical analysis from a purely theoretical point of view. For example, the book by Ljung and Söderström [1] is devoted to analysis and applications of recursive parameter estimators in discrete time.

Here, only continuous-time parameter estimators are considered, and an ordinary differential equation technique is obtained that is similar to Ljung's [2] in discrete time. But the extension of the mathematical theory into continuous-time, which is formidable, is not the main purpose. The main purpose is to provide general conditions under which averaging theory can be applied to continuous-time parameter estimators. The main theorem is presented first, with the long proof relegated to Appendix A, so that the theorem can be applied without going into the intricate mathematics. The theorem is used to prove convergence of continuous-time versions of both gradient and recursive prediction error parameter estimators. The theorem is also used to analyze the asymptotic dynamics of the simplest possible nontrivial parameter estimation problem, and thus obtain an indication of the ordering of the dynamical behavior of the four most common parameter estimators.

Manuscript received October 10, 1991; revised May 26, 1992. This work was supported in part by Rockwell International Science Center.

The authors are with Electrical Engineering Department, University of California, Los Angeles, CA 90024-1594.

IEEE Log Number 9207225.

The resulting averaged equations are simpler in continuous-time than in discrete time, in the sense that the continuous-time Riccati equation is mathematically simpler than the so-called discrete-time Riccati equation. This simplicity permits the analytic computation of asymptotic trajectories of the averaged estimates for a simple example. Also, as the sampling time tends to zero, the limiting asymptotic dynamical behavior of discrete-time parameter estimators can be studied. Furthermore, the analysis can be used in continuous-time adaptive control [3], again giving simpler equations and limiting behavior of discrete-time adaptive control. The delta operator approach [4] has shown the value of considering the continuous-time version of discrete-time algorithms in many areas. In addition to all the foregoing reasons for considering continuous-time parameter estimators, a major reason is to study the relationship with optimal nonlinear filtering. This relationship must be considered in continuous-time, and is the starting point of the companion paper [5].

The book [6] by Chen, and papers by Van Schuppen [7], Moore [8], Chen and Moore [9], and Gevers, Goodwin, and Wertz [10] all find conditions for the convergence of various forms of the extended least squares (ELS) algorithm in continuous-time, and do not consider gradient, recursive prediction error (RPE), or extended Kalman filter (EKFPE) parameter estimators. Here, the ordinary differential equation (ODE) technique is developed, by which both convergence conditions and asymptotic trajectories of the averaged estimates can be found for all the aforementioned parameter estimators.

To begin, a simple introduction to averaging theory is presented. Then the main theorem is stated (an outline of the proof is given in Appendix A), resulting in an ODE suitable for analysis of the asymptotic dynamics of continuous-time recursive stochastic parameter estimation algorithms. Using the ODE technique, four different parameter estimators are analyzed. First, the EKFPE fails with probability one to converge to the true values of the parameters in a system whose state noise covariance is unknown. Second, the asymptotic behavior of an example of the ELS algorithm is analyzed in detail, illustrating how the ODE technique can give an analytic bound for the asymptotic rate of convergence. The filtered version of the ELS apparently is significantly slower than the RPE, which

0018-9286/93$03.00

is conjectured to be asymptotically optimal. Finally, both the gradient and RPE algorithms are proven to converge with probability one to a minimum of the likelihood function.

II. Averaging

A simple introduction to averaging of deterministic dynamic systems is given in Sanders and Verhulst [11]. They trace the origins of averaging theory back to Lagrange [12] in 1788. An introduction to averaging of stochastic systems is given in Friedlin and Wentzell [13], who credit Gikhman [14] in 1952 and Krasnoselskii and Krein [15] in 1955.

While L. Ljung's work [2] on the discrete-time ODE technique inspired the development of the continuous-time ODE technique contained here, the mathematics required is based on the work of Kushner [16] and his colleagues Clark [17], Schwartz [18], and Huang [19]. Although [16]–[19] do not consider a diffusion term in the slow equations, both Khazminskii [20] and Papanicolaou, Strook, and Varadhan [21] develop a weak convergence theory with the diffusion term. However, all of the authors [16]–[21] accelerate only the fast equations using t/ϵ scaling, whereas here both the fast and slow equations are accelerated with exponential time scaling. Also, here only ODE limits are considered, rather than the more general diffusion limits.

The authors [16]–[21] are mainly concerned with approximation of wideband noise systems by diffusions, for which the t/ϵ scaling is expedient and appropriate. Here, the diffusion equations are determined by the parameter estimator, and the exponential time scaling is implicit and is not an approximation. The exponential time scaling occurs naturally because of the gain variation as $1/t$. Minor adjustments to the proofs allow slightly more general gain variation than $1/t$.

To see how this exponential time scaling arises and to introduce the basic idea of averaging as applied to a parameter estimator, consider the following example in which appropriate limits are assumed to exist. Let $w(t)$ be a zero mean rapidly fluctuating noise signal and $x(t)$ an n-vector satisfying

$$dx/dt = t^{-1}f(x,w). \tag{1}$$

For simplicity assume that $w(t)$ is stationary with probability density $p(w)$ which does not depend upon the state x. The t^{-1} factor will force $x(t)$ to vary slowly compared to $w(t)$ when t is large. Therefore, x is called the slow variable, and w the fast variable. A change of time scale

$$x_a(t) = x(e^t) \tag{2}$$

changes (1) to

$$dx_a/dt = f(x_a(t), w(e^t)). \tag{3}$$

The rate of fluctuation (i.e., bandwidth) of $w(e^t)$ increases with time. In the limit the noise w averages the function f and asymptotically x_a follows the trajectories of $\bar{x}(t)$ given by

$$d\bar{x}/dt = \bar{f}(\bar{x}) \tag{4}$$

where $\bar{f}$ is given by

$$\bar{f}(x) = E_w f(x,w) = \int f(x,w)p(dw). \tag{5}$$

Thus, the asymptotic properties of $x(t)$ are determined by the deterministic ordinary differential equation (4). The differential equation (4) determines the asymptotic trajectories of the averaged estimates $\bar{x}(t)$, but the asymptotic trajectories of $\bar{x}(t)$ do not determine the rate of convergence of $x(t)$ because $x(\cdot)$ fluctuates around and converges to $\bar{x}(\cdot)$. However, the rate of convergence of $x(t)$ must be slower than that of $\bar{x}(t)$, and thus bounds for the rate of convergence are determined by (4).

Assumption A1) The slow variable remains bounded on $[t_0, T]$ for all $T > t_0$.

Assumption A1) is equivalent to no finite escape time for x in (1). Then under Assumption A1), $x(t)$ will, with probability one, converge only to the stable fixed (i.e., stationary) points of (4). The precise relation between solutions of (1) and solutions of (4) is described by Kushner in [16]. Kushner also treats the case where $w(\cdot)$ depends upon $x(\cdot)$.

The above averaging technique can be applied to parameter estimators, which obey a number of stochastic differential equations, only some of which are in a form similar to (1). Other stochastic differential equations are associated with a parameter estimation stochastic differential equation, in which the gain will be taken to decrease inversely proportional to time. Call θ the parameter estimate, z_1 the states, state estimates and other variables satisfying linear stochastic differential equations, and z_2 the variables of Riccati and other nonlinear ordinary differential equations. Then θ will be the slow variable, and $z = (z_1^T \; z_2^T)^T$ will be the fast variable. An averaging result suitable for analysis of the gradient and RPE parameter estimators is the following theorem.

Theorem 1: Given the system

$$d\theta = t^{-1}[F_1(\theta,z)dt + F_2(\theta,z)dw] \tag{6}$$

$$dz_1 = A_1(\theta,z_2)z_1dt + B_1(\theta,z_2)udt + S_1(\theta,z_2)dw \tag{7}$$

$$dz_2/dt = A_2(\theta,z_2) \tag{8}$$

where $w(\cdot)$ is a unit variance Wiener process. The input $u(\cdot)$ is a particular form, specifically, $u(\cdot)$ is any linear combination of output from an asymptotically stationary stable linear system excited by additive $w(\cdot)$ and of linear stabilizing feedback of z_1. More precisely, $u(\cdot)$ is generated by

$$u(t) = C_1(\theta,z_2)z_1 + C_3(\theta,z_2)z_3 \tag{9}$$

where

$$d\begin{pmatrix} z_1 \\ z_3 \end{pmatrix} = \begin{pmatrix} A_1(\theta, z_2) + B_1(\theta, z_2)C_1(\theta, z_2) & B_1(\theta, z_2)C_3(\theta, z_2) \\ 0 & A_3(\theta, z_2) \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_3 \end{pmatrix} dt + \begin{pmatrix} S_1(\theta, z_2) \\ S_3(\theta, z_2) \end{pmatrix} dw \tag{10}$$

in which $A_1 + B_1C_1$ and A_3 have eigenvalues with real parts negative and uniformly bounded away from zero. Note when θ is a function of t, then $u(t)$ is not stationary. Then assume

a) $F_1(\cdot,\cdot)$, $F_2(\cdot,\cdot)$, $A_1(\cdot,\cdot)$, $A_2(\cdot,\cdot)$, $B_1(\cdot,\cdot)$ and $S_1(\cdot,\cdot)$ are twice continuously differentiable. $S_1(\cdot,\cdot)$ is bounded.

b) $F_1(\cdot,\cdot)$, and $F_2(\cdot,\cdot)$ have compact θ-support, i.e., are equal to zero for $\|\theta\| > \rho$ where ρ is a large positive number.

c) $A_1(\cdot,\cdot)$, $A_2(\cdot,\cdot)$, $B_1(\cdot,\cdot)$, $F_1(\cdot,\cdot)$, and $F_2(\cdot,\cdot)$ satisfy

$$\max\Big(\theta^T F_1(\theta, z_1, z_2), z_1^T(A_1(\theta, z_2)z_1 + B_1(\theta, z_2)u(t)), z_2^T A_2(\theta, z_2), \sup_{\|x\|=1} \|x^T F_2(\theta, z_1, z_2)\|^2\Big) \le c\big(1 + \|\theta\|^2 + \|z_1\|^2 + \|z_2\|^2\big).$$

d) The real parts of the eigenvalues of $A_1(\theta, z_2)$ are negative and uniformly bounded away from zero.

e) For any fixed $\theta \in \boldsymbol{R}^{n_\theta}$, the z_2 equation (8) has a unique steady-state solution $z_{2\infty}(\theta)$, $z_2(t;\theta) \to z_{2\infty}(\theta)$ uniformly on compact θ-sets, and $z_{2\infty}(\cdot)$ is continuous.

f) For each initial condition (θ_0, z_0) and $c > 0$, if $\|\theta(\cdot)\| \le c$ then the z_2 equation (8) has a unique solution which is bounded for all time. For each c, the solution bound is uniform on compact z_0-sets.

Then if θ is held constant, the z system (7)–(10) has a unique asymptotic invariant measure. The projection of this asymptotic invariant measure on the (z_1, z_2) coordinates will be denoted $P^\theta(\cdot)$, a limit measure. Define the ordinary differential equation (ODE):

$$d\theta^0/dt = \boldsymbol{F}(\theta^0) \qquad \text{where } \boldsymbol{F}(\theta) = \int F_1(\theta, z)P^\theta(dz). \tag{11}$$

Furthermore, assume that for any $\theta^0(0)$, the ODE (11) has a unique solution for all $t \in [0, \infty)$. Then the asymptotics of $\theta^0(\cdot)$ and $\theta(\cdot)$ are related by

a) If for all initial conditions $\theta^0(t) \to R_c$, then $\theta(t) \to R_c$ wp1 as $t \to \infty$.

b) If for all initial conditions $\|\theta^0(\log(t+T)) - \theta^0(\log(t))\| \to 0$, then

$$\|\theta(t+T) - \theta(t)\| \to 0 \qquad wp1 \text{ as } t \to \infty.$$

c) If $\boldsymbol{F}(\cdot)$ is twice continuously differentiable, $\theta(t) \to \theta_\infty$ with probability strictly greater than 0 and $\theta(t)$ has positive definite covariance for each t, then θ_∞ is a stable fixed point for the $\theta^0(\cdot)$-ODE equation (i.e., $\boldsymbol{F}(\theta_\infty) = 0$ and $\boldsymbol{F}_\theta(\theta_\infty)$ has all its eigenvalues in the open left-half plane).

This theorem is an amalgam of [22, theorems II.1, II.2, and II.4] and proof can also be found in [32]. The proofs are lengthy and are outlined in Appendix A.

Remark: Hypotheses b) and c) imply Assumption A1). If Assumption A1) is assumed, then b) can be omitted, but c) is still needed to show the tightness of $\{x^\epsilon(t)\}$ in Theorem 6 of Appendix A used in the proof of Theorem 1.

Analysis of the extended Kalman filter and extended least squares estimators gives equations that are slightly different from (6)–(8), namely

$$d\theta = t^{-1}[F_1(\theta, z)dt + F_2(\theta, z)dw] + t^{-2}[F_3(\theta, z, t)dt + F_4(\theta, z, t)dw] \tag{12}$$

$$dz_1 = A_1(\theta, z_2)z_1dt + B_1(\theta, z_2)udt + S_1(\theta, z_2)dw + t^{-1}[G_1(\theta, z, t)dt + G_2(\theta, z, t)dw] \tag{13}$$

$$dz_2 = A_2(\theta, z_2)dt + t^{-1}[H_1(\theta, z, t)dt + H_2(\theta, z, t)dw]. \tag{14}$$

Mild generalizations of the proof in Appendix A and in [22] and [32] give the following theorem for estimators obeying (12)–(14).

Theorem 2: Assume the conditions of Theorem 1 and further assume F_3, F_4, G_1, G_2, H_1, and H_2 are bounded and twice differentiable. Additionally, if for each fixed θ, (8) and (14) have the same limit [i.e., the asymptotic measure of (14) is the atomic measure concentrated at the steady state solution of (8)], then the conclusions of Theorem 1 still hold and the ODE (11) for the system (12)–(14) is identical to the ODE (11) for the system (6)–(8).

Theorem 2 generalizes Theorem 1 to show that the extra terms in (12) and (13) compared to (6) and (7) do not affect the asymptotic behavior characterized by (11), although the extra terms may be important in transient response. In (12), the extra terms converge to zero, similar to the F_2 terms in Lemma 5 of Appendix A. In the linear equation (13), the extra terms can be ignored because of the linearity of (7) and the exponential convergence insured by hypothesis d) of Theorem 1. However, Baldi [23] has shown that asymptotic stability is not sufficient to establish the existence of an asymptotic measure of a stochastic differential equation. Therefore, to obtain the same ODE (11) for (6)–(8) and (12)–(14), it must further be assumed that the asymptotic measure of (14) is the atomic measure concentrated at the steady state solution of (8) for each fixed θ.

III. Systems Considered

Consider the multiple-input, multiple-output system in the state space form of Itô stochastic differential equations

$$dx = A_0xdt + B_0udt + \Sigma_0dv \tag{15}$$

$$dy = C_0xdt + Hdw \qquad R = HH^T > 0 \tag{16}$$

and a model for this system parameterized by the p-

vector θ

$$dz = A(\theta)zdt + B(\theta)udt + \Sigma(\theta)dv \tag{17}$$

$$dy = Czdt + Hdw \tag{18}$$

where u is as stated under Theorem 1 and also is y_t-adapted, which loosely means that $u(\tau)$ becomes known for times $\tau \le t$. The known matrix-valued functions $A(\cdot)$, $B(\cdot)$, and $\Sigma(\cdot)$ are assumed twice differentiable. C_0, C, and H are known matrices, and, for simplicity, are not functions of θ. Also for simplicity, the Wiener processes v and w are assumed uncorrelated, with unity incremental variance. These simplifications are for notational convenience only, because the theory developed here can be extended.

Assumption A2) The system matrix A_0 is asymptotically stable.

This assumption is needed to generate the invariant measure of the fast system, since the fast system includes (15).

The model (17), (18) is used to define the estimator equations. The estimate, denoted $\hat{x}$, must, therefore, have dimension equal to that of z. However, note that the dimension of the system and model are not necessarily assumed equal. When the system and model dimensions are equal, it is assumed that there exists a true value θ_0 of the parameter vector θ such that $A(\theta_0) = A_0$, $B(\theta_0) = B_0$, and $\Sigma(\theta_0) = \Sigma_0$. Also, then define $\tilde{x} = x - \hat{x}$.

The (asymptotic negative log) likelihood function $J(\theta)$ is defined by

$$J(\theta) = \lim_{t\to\infty} E\{[C_0 x(t) - C\hat{x}(t;\theta)]^T \cdot R^{-1}[C_0 x(t) - C\hat{x}(t;\theta)]\} \tag{19}$$

where $\hat{x}(t;\theta)$ is the state estimate of the Kalman filter derived from the model (17), (18) using a fixed θ value in (17), and $x(t)$ is generated by the system (15). Therefore $J(\theta)$ is a function of θ only. Although the general state space form (15)–(18) treated here may have $J(\theta)$ with many local minima, some specific forms (e.g., corresponding to autoregressive moving average in discrete time) have a unique minimum at the true parameter value [26].

Define a fixed, bounded, open set R_0 in θ space, where R_0 is usually taken very large.

Assumption A3) Assume all estimator initial conditions $\hat{\theta}(t_0)$ and all minima of $J(\theta)$ are in R_0. Assume all estimators are equipped with a projection facility to return $\hat{\theta}(t)$ generated by (6) back into R_0. Further assume R_0 is so large that, with probability one, all projected trajectories $\hat{\theta}(\cdot)$ are identical to the unprojected trajectories.

This projection causes difficulties in the convergence proofs. With probability one, the projected trajectories $\hat{\theta}(\cdot)$ cannot be identical to the unprojected trajectories, if the region is bounded. To be strictly mathematically correct, *if the region is large enough*, then the Lyapunov argument in Theorem 3 shows the averaged trajectories of $\hat{\theta}(\cdot)$, and, by further arguments, the trajectories of $\hat{\theta}(\cdot)$ themselves, will hit the boundary with probability less than ϵ for any $\epsilon > 0$. Then the convergence of the algorithms to a minimum of the likelihood function is with probability $1 - \epsilon$, not probability one. Therefore, for simplicity, Assumption A3) is made.

A projection facility is a standard device used in parameter estimation. Explanation is lengthy, so the interested reader is referred to [1, pp. 93, 162, and 366].

Under Assumption A3), in Theorem 1 then $F_1(\cdot,\cdot)$ and $F_2(\cdot,\cdot)$ can be taken to be zero outside of R_0. Then hypothesis b) of Theorem 1 is satisfied. Since Assumption A1) is implied by Assumption A3), then henceforth Assumption A1) is not specifically assumed.

IV. Extended Kalman Filter

L. Ljung [24] showed in discrete time that the extended Kalman filter [25] used as a parameter estimator (EKFPE) is not guaranteed to converge, and then fixed it by adding a term that guarantees convergence with probability one. Here a system is considered in which only the state noise covariance $\Sigma\Sigma^T$ is unknown. It has long been folk knowledge that the EKFPE will not necessarily converge to the true values of the parameters in such a system. This rather trivial result serves as a first example to apply the ODE technique, because it can easily be checked and because it illustrates a deficiency of the EKFPE.

First consider the case where all the system matrices A_0, B_0, C_0, and H are all known, the system dimension (dim x) of (15) equals the model dimension (dim $\hat{x}$) of (17), and only $\Sigma(\theta)$ is to be estimated. Define the extended state as $(x^T, \theta^T)^T$ and adjoin the parameter part of the extended state equation $d\theta = 0$.

Then the EKFPE [25] for the model (17) and (18) with the extended state equation $d\theta = 0$ is

$$d\hat{x} = (A_0\hat{x} + B_0 u)dt + P_x C^T R^{-1} d\epsilon \tag{20}$$

$$d\epsilon = dy - C\hat{x}dt \tag{21}$$

$$d\hat{\theta} = P_{\theta x} C^T R^{-1} d\epsilon \tag{22}$$

$$\hat{\mathscr{A}} = A_0 - P_x C^T R^{-1} C \tag{23}$$

$$dP_x/dt = \hat{\mathscr{A}} P_x + P_x \hat{\mathscr{A}}^T + P_x C^T R^{-1} C P_x + \Sigma(\hat{\theta})\Sigma^T(\hat{\theta}) \tag{24}$$

$$dP_{\theta x}/dt = P_{\theta x}\hat{\mathscr{A}}^T \tag{25}$$

$$dP_\theta/dt = -P_{\theta x} C^T R^{-1} C P_{\theta x}^T. \tag{26}$$

The fixed points of (20)–(26) can be found by using averaging theory. To put EKFPE in the form of (12)–(14), change variables as

$$V = t^{-1}P_\theta^{-1} \tag{27}$$

$$W^T = tVP_{\theta x}. \tag{28}$$

Then the EKFPE equations (20)–(26) become

$$d\hat{\theta} = t^{-1}V^{-1}W^T C^T R^{-1}(C\tilde{x}dt + Hdw) \tag{29}$$

$$dV/dt = t^{-1}(W^T C^T R^{-1} C W - V) \tag{30}$$

$$dx = (A_0 x + B_0 u)dt + \Sigma(\theta_0)dv \tag{31}$$

$$d\tilde{x} = \hat{\mathscr{A}}\tilde{x}dt + \Sigma(\theta_0)dv - P_xC^TR^{-1}Hdw \quad (32)$$

$$dW/dt = \hat{\mathscr{A}}W + t^{-1}WV^{-1}W^TC^TR^{-1}CW \quad (33)$$

$$dP_x/dt = \mathscr{A}P_x + P_x\hat{\mathscr{A}}^T + P_xC^TR^{-1}CP_x + \Sigma(\hat{\theta})\Sigma^T(\hat{\theta}). \quad (34)$$

In (12)–(14), identify θ, z_1, and z_2 with $(\hat{\theta}, \text{col}\, V)$, $(x, \tilde{x}, \text{col}\, W)$, and $\text{col}\, P_x$, respectively, from above. Because the EKFPE may diverge for some values of $\hat{\theta}(0)$, consider only those (unknown) regions of θ space in which the estimates $\hat{\theta}$ of (29) are entirely contained and in which the EKFPE is asymptotically stable. Those regions are not empty because for $\hat{\theta}(0) = \theta_0$, the EKFPE reduces to the Kalman filter which is asymptotically stable under Assumption A2). In those regions, the steady state of P_x is achieved and all conditions of Theorem 2 are satisfied. In those regions then the ordinary differential equation (11) describes the asymptotic behavior of $\hat{\theta}$ and V, and, defining $\tau = e^t$, from (29) and (30) gives

$$\begin{bmatrix} d\hat{\theta}^0/d\tau \\ dV^0/d\tau \end{bmatrix} = \begin{bmatrix} (V^0)^{-1}E(W^TC^TR^{-1}C\tilde{x}) \\ E(W^TC^TR^{-1}CW) - V^0 \end{bmatrix} \quad (35)$$

where the expectation E is taken with respect to the limit measure of the $[x, \tilde{x}, W, P_x]$ system.

But from (33) for fixed $\hat{\theta}$ and V, $W(t)$ is not a random variable when estimating process noise parameters only in the EKFPE. Since $E\tilde{x}$ vanishes, the top half of (35) is identically zero and every real p-vector is a fixed point θ_∞.

For this special case of estimating noise parameters only, the EKFPE (20)–(26) has an analytical solution with which to check the asymptotic averaging theory. Since $P_{\theta x}$ is initially zero, from (25) it remains zero. Then from (22) $\hat{\theta}$ remains constant at its initial value, as does P_θ from (26). Therefore direct analysis of the EKFPE parameter estimator agrees with the averaging theory.

Consider now the case where θ is partitioned as $(\theta_1^T, \theta_2^T)^T$ such that $A(\theta_1)$ and $B(\theta_1)$ are not functions of θ_2 and $\Sigma(\theta_2)$ is not a function of θ_1 in the model (17), (18). Using similar direct analysis, it can be shown that the resulting EKFPE algorithm is such that $\theta_2(t)$ remains constant at its initial value. Then sensitivity analysis can be used to assess the degree of incremental asymptotic error in $\hat{\theta}_1$ due to the incremental asymptotic error in $\hat{\theta}_2$. Consequently, the folk knowledge is verified that the EKFPE fails, with probability one, to converge to the true values of the parameters in a system whose state noise covariance Σ is unknown. This implies limited utility of the EKFPE in applications, unless accompanied by a sensitivity analysis of the effect of errors in the process noise covariance.

Next, parameter estimators that can estimate the process noise covariance will be considered.

V. Extended Least Squares

The conditions for convergence and the rate of convergence of various forms of the extended least squares (ELS) algorithm have been studied in [6]–[10]. Here, the ODE (11) will be used to study a simple example of the ELS. The simple example is a first-order system (15), (16) in which only the true state noise parameter Σ_0 is unknown and there is no input u. Equivalently, consider the innovations representation in which the innovations variance is

$$P = -a + (a^2 + \Sigma_0^2)^{1/2}. \quad (36)$$

So it is desired to estimate P in

$$dx = axdt + Pdw \quad (37)$$

$$dy = xdt + dw. \quad (38)$$

Under the conditions stated in [10], in the following version of the ELS algorithm, with probability one, $\hat{\theta}(t)$ converges to $\theta_0 = P - a - \kappa$, thus identifying Σ_0. Here, $\kappa^{-1} > 0$ is also known as the time constant of a prefilter on the regression vector (see [10])

$$d\hat{\theta} = \psi^e\left[dy - (a+\kappa)\psi^y dt - \hat{\theta}\psi^e dt\right]/(\text{tr}Q + \Psi) \quad (39)$$

$$dQ = \psi\psi^T dt \qquad Q(0) = Q_0 > 0 \quad (40)$$

$$d\psi^y = -\kappa\psi^y dt + dy \quad (41)$$

$$d\psi^e = -\kappa\psi^e dt + dy - (a+\kappa)\psi^y dt - \hat{\theta}\psi^e dt \quad (42)$$

where $\psi = (\psi^y\ \psi^e)^T$, and $\Psi = \sup_{0 \le \tau \le t} \psi^T(\tau)\psi(\tau)$.

To put the ELS (39)–(42) in a form suitable for the application of Theorem 2, change variables as

$$V = t^{-1}Q \quad (43)$$

and then (39) and (40) become

$$d\hat{\theta} = t^{-1}\psi^e\left[dy - (a+\kappa)\psi^y dt - \hat{\theta}\psi^e dt\right]/(\text{tr}V + t^{-1}\Psi) \quad (44)$$

$$dV/dt = t^{-1}(\psi\psi^T - V). \quad (45)$$

In (12)–(14), identify θ and z_1 with $(\hat{\theta}, \text{col}\, V)$ and (x, ψ), respectively, and there is no z_2 equation. For any fixed V and for fixed $\hat{\theta} > -\kappa$, (41) and (42) are asymptotically stable. Therefore, to apply Theorem 2, the estimate $\hat{\theta}$ must be restricted via projection [1, pp. 93, 162, and 366] to the region $\hat{\theta} > -\kappa$. (This is not necessary in [10].) Because the anticipated limit of $\hat{\theta}$ is $\theta_0 = P - a - \kappa$, where $a < 0$, $P > 0$, $\kappa > 0$, this restriction is natural. With Assumptions A2), A3), and the bound $-\kappa < \hat{\theta}(t) < \rho$, for very large ρ imposed on (39), then the conditions of Theorem 2 hold and the asymptotic behavior obeys the ODE (11), which for this example is

$$\frac{d}{d\tau}\begin{pmatrix} \hat{\theta}^0 \\ V^0 \end{pmatrix} = \begin{pmatrix} E\left\{\psi^e\left[x - (a+\kappa)\psi^y - \hat{\theta}^0\psi^e\right]\right\}/\text{tr}\, V^0 \\ E\{\psi\psi^T\} - V^0 \end{pmatrix} \quad (46)$$

where the expectation E is taken with respect to the asymptotic probability measure of the fixed θ fast system

$$d\begin{pmatrix} x \\ \psi^y \\ \psi^e \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 1 & -\kappa & 0 \\ 1 & -\kappa - a & -\kappa-\theta \end{pmatrix}\begin{pmatrix} x \\ \psi^y \\ \psi^e \end{pmatrix} dt + \begin{pmatrix} P \\ 1 \\ 1 \end{pmatrix} dw. \tag{47}$$

From the asymptotic covariance (algebraic Lyapunov) equation of the above fixed θ fast system, compute the functions $f(\theta)$ and $g(\theta)$ defined by

$$f(\theta) = E\{\psi^e x\} - (a+\kappa)E\{\psi^e\psi^y\} - E\{(\psi^e)^2\}\theta = -\frac{(\theta-\theta_0)(\theta+\theta_1)}{2(\theta+\kappa)(\theta+2\kappa)} \tag{48}$$

$$g(\theta) = E\{\psi^{y2}\} + E\{\psi^{e2}\} = (2 + f(\theta) - \Sigma_0^2/a(\kappa - a))/2\kappa \tag{49}$$

where the true value of θ is $\theta_0 = P - a - \kappa$ and $\theta_1 = P - a + \kappa > 0$. Then (46) becomes

$$\mathrm{tr}\, V^0 d\hat{\theta}^0/d\tau = f(\hat{\theta}^0) \tag{50}$$

$$d(\mathrm{tr}\, V^0)/d\tau = g(\hat{\theta}^0) - \mathrm{tr}\, V^0. \tag{51}$$

Transform back to real time t, and let $t = e^\tau$, $\rho(t) = t\,\mathrm{tr}\, V^0(\ln t)$ and $\theta(t) = \hat{\theta}^0(\ln t)$.

$$d\theta/dt = \rho^{-1}f(\theta) \tag{52}$$

$$d\rho/dt = g(\theta). \tag{53}$$

Dividing (52) by (53), separating variables, and integrating gives

$$\rho(\theta) = \beta e^{\theta/\kappa}(\theta+\theta_1)^{\gamma\theta_0}(\theta-\theta_0)^{-\gamma\theta_1} \tag{54}$$

where $\gamma = (1 - \Sigma_0^2/a(\kappa - a))/\kappa$ and β is a constant of integration which can be found by evaluating (54) at the initial time. Then from (52), (54), and (48),

$$d\theta/dt = (-2\beta)^{-1}e^{-\theta/\kappa}(\theta+\kappa)(\theta+2\kappa) \cdot(\theta+\theta_1)^{1-\gamma\theta_0}(\theta-\theta_0)^{\gamma\theta_1+1}. \tag{55}$$

Since θ is restricted to $\theta > -\kappa$, the only convergence point is θ_0, and the rate of convergence of the trajectories of the averaged estimates is fractional in t

$$|\theta - \theta_0| = O(t^{-\gamma^{-1}\theta_1^{-1}}) \qquad \text{as } t \text{ tends to infinity} \tag{56}$$

where the fraction $\gamma^{-1}\theta_1^{-1}$ is

$$\gamma^{-1}\theta_1^{-1} = \left[1 + (\Sigma_0^2 + a^2)^{1/2}/\kappa\right]^{-1} \cdot\left[1 + \Sigma_0^2/(\kappa|a| + a^2)\right]^{-1}. \tag{57}$$

Remark 1: For many values of a, Σ_0, and κ, this asymptotic rate of convergence of the ODE trajectories of the averaged estimates is slower than the $O(\sqrt{\ln t/t})$ asymptotic rate of convergence of the estimates found in [9].

Remark 2: If $\Sigma_0 \to 0$, the rate of convergence of the ODE trajectories associated with the averaged estimates of linear least squares $O(t^{-1})$ is not obtained because ψ^e still must go through its filter with time constant κ^{-1}.

Remark 3: If $\kappa \to \infty$, the rate of convergence $O(t^{-1})$ is apparently reached, but $\hat{\theta}(t)$ tends to $\theta_0 = P - a - \kappa \to -\infty$. Therefore, the unfiltered ELS could indeed be asymptotically optimal, but prefiltering to attain convergence apparently spoils the rate of convergence.

VI. Gradient Algorithm

Applying the arguments in [1, p. 45], to the continuous-time model (17), (18) gives the continuous-time gradient algorithm. It can also be derived by taking the continuous-time limit of the discrete-time gradient algorithm of [1, p. 97]. For the system (15), (16) and model (17), (18), a simplification occurs even when the system and model have different state dimensions. In continuous-time, for (15)–(18), the measurement noise covariance equals the innovations covariance. In (15), (16) the measurement noise covariance is $HH^T = R$, which is known. Therefore, the estimate of the innovations covariance $\hat{\Lambda}$ in [1] is $\hat{\Lambda} = R$, for (15)–(18). Consequently the gradient algorithm based on the model (17), (18) is

$$d\hat{\theta} = t^{-1}r^{-1}W^TC^TR^{-1}d\epsilon \tag{58}$$

$$t\,dr/dt = \mathrm{tr}\,(W^TC^TR^{-1}CW) - r \tag{59}$$

$$d\epsilon = dy - C\hat{x}dt \tag{60}$$

$$d\hat{x} = A_t\hat{x}dt + B_t u dt + PC^TR^{-1}d\epsilon \tag{61}$$

$$dP/dt = A_tP + PA_t^T + \Sigma_t\Sigma_t^T - PC^TR^{-1}CP \tag{62}$$

$$\mathscr{A}_t = A_t - PC^TR^{-1}C \tag{63}$$

$$dW/dt = \mathscr{A}_tWdt + [\partial(A\hat{x})/\partial\theta + \partial(Bu)/\partial\theta]_t dt + \left(\Pi_1C^TR^{-1}d\epsilon|\cdots|\Pi_{n_\theta}C^TR^{-1}d\epsilon\right) \tag{64}$$

$$d\Pi_i/dt = (\partial A/\partial\theta_i)_tP + P(\partial A/\partial\theta_i)_t^T + \mathscr{A}_t\Pi_i + \Pi_i\mathscr{A}_t^T + \Sigma_t(\partial\Sigma/\partial\theta_i)_t^T + (\partial\Sigma/\partial\theta_i)_t\Sigma_t^T \tag{65}$$

where the subscript t means evaluated at $\hat{\theta}(t)$, e.g., $A_t = A(\hat{\theta}(t))$, and where $W = \partial\hat{x}/\partial\theta$ and $\Pi_i = \partial P/\partial\theta_i$ and $r(t_0) = r_0 > 0$. Also, the vertical bars in the last term of (64) denote partitioning of the matrix into its column vectors.

Theorem 3: Assume that the Assumptions A2), A3), and

1) $P(t)$ is projected such that $P(t) \geq 0$.
2) $\hat{\theta}(t)$ is projected such that $A(\hat{\theta}(t))$ is exponentially stable.
3) $\Sigma(\hat{\theta}(t))$, $(\partial A/\partial\theta_i)(\hat{\theta}(t))$, $(\partial B/\partial\theta_i)(\hat{\theta}(t))$, and $(\partial\Sigma/\partial\theta_i)(\hat{\theta}(t))$ are truncated such that for some large constant κ, $\|\Sigma_t\| + \|(\partial A/\partial\theta_i)_t\| + \|(\partial B/\partial\theta_i)_t\| + \|(\partial\Sigma/\partial\theta_i)_t\| \leq \kappa$ for $i = 1, 2, \cdots, n_\theta$.
4) $A(\theta)$, $B(\theta)$, and $\Sigma(\theta)$ are truncated such that for some large constants κ_0 and κ_1, $\|A(\theta)\| + \|B(\theta)\| + \|\Sigma(\theta)\Sigma^T(\theta)\| \leq \kappa_0\|\theta\|^2 + \kappa_1$.

Then

a) $\hat{\theta}(t)$ can only converge (with positive probability) to local minima of the likelihood function $J(\theta)$.

b) If there is a true parameter θ_0 such that $A(\theta_0) = A_0$, $B(\theta_0) = B_0$ and $\Sigma(\theta_0) = \Sigma_0$ then θ_0 is the global minimum of $J(\cdot)$.

c) If

i) θ_0 is an isolated local minimum of $J(\cdot)$,

ii) R_c is a compact, convex, neighborhood of θ_0 with smooth boundary such that θ_0 is the only local minimum of $J(\cdot)$ in R_c,

iii) $(\partial J/\partial\theta)^T(\theta - \theta_0) < 0 \quad \forall\theta \in \partial R_c$

iv) Equations (58)–(65) have $\hat{\theta}(t)$ projected such that $\hat{\theta}(t) \in R_c$, $\forall t \geq 0$,

then $\hat{\theta}(t) \to \theta_0$ wp1.

Proof: Appendix B shows that the assumptions above guarantee that Theorem 1 applies to (58)–(65). From Theorem 1, the ODE (11) is

$$d\hat{\theta}^0/d\tau = (r^0)^{-1}E_\theta\{W^TC^TR^{-1}(C_0x - C\hat{x})\} \quad (66)$$

$$dr^0/d\tau = \operatorname{tr} E_\theta\{W^TC^TR^{-1}CW\} - r^0 \quad (67)$$

where E_θ is the expectation with respect to the limit probability measure of the fixed θ fast system (61)–(65). Since $W = \partial\hat{x}/\partial\theta$, then (66) is

$$d\hat{\theta}^0/d\tau = -2(r^0)^{-1}\partial J/\partial\theta \quad (68)$$

and J is a Lyapunov function since $r^0 > 0\ \forall t$. This proves a). To prove b), realize that the fixed θ fast system (61)–(65) is a linear estimator for $x(t)$. For $\theta = \theta_0$, that linear estimator is the Kalman filter for (15), (16). Then asymptotic optimality of the Kalman filter implies $P(\theta_0) \leq P(\theta)$ for any θ. This implies $J(\theta_0) \leq J(\theta)$. Finally, to prove c), if the gradient of $J(\cdot)$ points inward on the boundary of R_c, and $\theta(t)$ is projected to remain in R_c, then the Lyapunov argument in the proof of a) requires that $\hat{\theta}^0$ converge to the only isolated minimum of J in R_c.

Remark 1: This proof parallels that in the discrete-time case [1].

Remark 2: Theorem 1 assures existence and uniqueness of solutions of (58)–(65) globally.

Remark 3: Assumption A3) that the projected algorithm is identical to the unprojected algorithm is necessary to assure that $W = \partial\hat{x}/\partial\theta$ as the solution to the estimator equations evolves.

Example: Consider the example system (37), (38) of the previous section. Then the gradient algorithm (58)–(65) becomes

$$d\hat{\theta} = t^{-1}r^{-1}W\,d\epsilon$$
$$t\,dr/dt = W^2 - r \quad (69)$$

where the fast system x, $\tilde{x}$, and W obeys

$$d\begin{pmatrix} x \\ \tilde{x} \\ W \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a-\theta & 0 \\ 0 & 1 & a-\theta \end{pmatrix}\begin{pmatrix} x \\ \tilde{x} \\ W \end{pmatrix}dt + \begin{pmatrix} P \\ P-\theta \\ 1 \end{pmatrix}dw \quad (70)$$

where θ is projected such that $\theta > a$. Here the true value $\theta_0 = P$.

Analysis proceeds as in the preceding section, giving

$$\frac{d\theta}{dt} = \frac{[\theta + (P - 2a)]^2}{-\beta(\theta - a)^3}(\theta - P)^2 \quad (71)$$

which gives the rate of convergence of the trajectories of the ODE above as $|\hat{\theta}(t) - P| = O(t^{-1})$ as t tends to infinity.

VII. Recursive Prediction Error

The principles of [1, p. 47] are applied to the continuous-time model (17), (18) in [22] to obtain the continuous-time recursive prediction error (RPE) method. This is identical to the continuous-time limit of the discrete-time RPE [1]. Gerencsér *et al.* [27] show that this continuous-time RPE is identical to recursively computing the maximum likelihood estimate except for small perturbations that do not affect asymptotic behavior. The RPE is identical to the gradient method, (58)–(65), except that the scalar $r(t)$ is replaced by the $n_\theta \times n_\theta$ matrix $V(t)$, so (59) is replaced by

$$t\,dV/dt = W^TC^TR^{-1}CW - V \quad (72)$$

and V replaces r in (58).

Theorem 4: Assume the Assumptions A2), A3), and 1)–4) from Theorem 3. Then the conclusions of Theorem 3 are valid for the RPE (58), (60)–(65) and (72). Proof is as in Theorem 3, replacing r^0 by $V^0 > 0$ in (66)–(68).

Example: For $\dim\theta = n_\theta = 1$ then (59) and (72) are identical, implying that the gradient algorithm and RPE are identical. Therefore, the example of the previous section is also an example of the RPE, with a rate of convergence of the ODE as $|\hat{\theta}(t) - P| = O(t^{-1})$ as t tends to infinity.

Conclusions

An ODE technique has been developed, suitable for the asymptotic analysis of recursive parameter estimators in continuous-time. This extends L. Ljung's work [2] in discrete time, and is based on continuous-time averaging theory [16]–[21] of stochastic differential equations. The resulting equations are simpler than in discrete-time because the underlying fast equations give analytically simpler solutions, which permit analytic solutions of simple examples. Also, the technique permits analysis of discrete-time estimators as the sampling time tends to zero and analysis of continuous-time adaptive control algorithms.

The ODE technique gives only asymptotic results, necessitated by Assumptions A1) and A2). The Assumption A1) of no finite escape time of the slow variables can be obviated by the use of stochastic Lyapunov functions [6]. The Martingale convergence theorem gives sharper convergence results [10] without Assumption A1), but no rates of convergence.

The four most common parameter estimators for continuous-time systems in state space form have been analyzed. The extended Kalman filter fails, with probability one, to converge to the true values of the parameters in a system whose state noise covariance is unknown. The prefiltered ELS converges, but the ODE has an asymptotic rate fractional in inverse time, for a simple example. The gradient method is identical to the RPE if only one parameter is to be estimated, but must converge more slowly than the RPE in higher dimensions because of the well known fact that gradient optimization algorithms are slower than optimization algorithms based on Newton–Raphson methods. The simple example of RPE considered here has an ODE that converges at an asymptotic rate $O(1/t)$. It is conjectured that this holds in general for the RPE. The asymptotic rate $O(1/t)$ is optimal, in the sense that $O(1/t)$ is the asymptotic rate of convergence of the trajectories associated with the averaged estimates of the linear least squares estimator.

Appendix A
Proof of Theorem 1

This Appendix gives an outline of the proof of Theorem 1, which is given in detail in [22]. See also [32]. The following Theorems 5–8 combine to prove Theorem 1. Theorem 5 is similar to [16, theorem 6, ch. 6] except for the exponential time scaling and the presence of a diffusion term in the slow equation. As is customary in averaging theory, now $x(t) \in \boldsymbol{R}^{n_x}$ is the slow variable and $y(t) \in \boldsymbol{R}^{n_y}$ is the fast variable, which obey

$$dx = t^{-1}[F_1(x,y)dt + F_2(x,y)dw] \tag{73}$$

$$dy = H_1(x,y)dt + H_2(x,y)dw. \tag{74}$$

Let $\epsilon \in (0,1]$. Let $t_\epsilon \to \infty$ as $\epsilon \to 0$ (note that $t_{(\cdot)}$ is not required to be continuous). The accelerated system, $(x^\epsilon(\cdot), y^\epsilon(\cdot))$, is defined by:

$$x^\epsilon(t+s) - x^\epsilon(t) = \int_t^{t+s} F_1(x^\epsilon(\tau), y^\epsilon(\tau))\,d\tau + \int_{e^{t+t_\epsilon}}^{e^{t+s+t_\epsilon}} F_2(x^\epsilon(\log(\tau) - t_\epsilon), y^\epsilon(\log(\tau) - t_\epsilon))\frac{dw(\tau)}{\tau} \tag{75}$$

$$\begin{aligned} d\bar{y}^\epsilon(t) &= H_1(x^\epsilon(\log(t) - t_\epsilon), \bar{y}^\epsilon(t))dt \\ &\quad + H_2(x^\epsilon(\log(t) - t_\epsilon), \bar{y}^\epsilon(t))\,dw(t) \\ y^\epsilon(t) &= \bar{y}^\epsilon(e^{t+t_\epsilon}) \\ x^\epsilon(0) = x_\epsilon \qquad y_\epsilon(0) &= y_\epsilon \qquad (\therefore \bar{y}^\epsilon(e^{t_\epsilon}) = y_\epsilon) \end{aligned} \tag{76}$$

x_ϵ and y_ϵ will be random variables independent of $w(t)$ for $t > e^{t_\epsilon}$ and whose values may or may not be tied to the $(x(\cdot), y(\cdot))$ process. Loosely, $(x^\epsilon(\cdot), y^\epsilon(\cdot))$ is given by

$$\begin{aligned} x^\epsilon(t) &= x(e^{t+t_\epsilon}), & y^\epsilon(t) &= y(e^{t+t_\epsilon}) \\ x^\epsilon(0) &= x_\epsilon, & y^\epsilon(0) &= y_\epsilon. \end{aligned} \tag{77}$$

Also define the fixed x system as

$$d\bar{y}(t;x) = H_1(x, \bar{y}(t;x))dt + H_2(x, \bar{y}(t;x))dw(t) \tag{78}$$

and the input z system, for a bounded nonanticipative process $z(\cdot)$, as

$$d\bar{y}_z(t) = H_1(z(t), \bar{y}_z(t))dt + H_2(z(t), \bar{y}_z(t))dw(t). \tag{79}$$

The transition probability measures for these process are

$$\bar{P}^\epsilon(x,y,t,t_0,A) = P\{\bar{y}^\epsilon(t) \in A | x^\epsilon(\log(t_0) - t_\epsilon) = x, \bar{y}^\epsilon(t_0) = y\} \tag{80}$$

$$P^\epsilon(x,y,t,t_0,C) = P\{(x^\epsilon(t), y^\epsilon(t)) \in C | (x^\epsilon(t_0), y^\epsilon(t_0)) = (x,y)\} \tag{81}$$

$$\bar{P}(y,t,t_0,A|x) = P\{\bar{y}(t;x) \in A | \bar{y}(t_0;x) = y\} \tag{82}$$

$$\bar{P}_z(y,t,t_0,A) = P\{\bar{y}_z(t) \in A | \bar{y}_z(t_0) = y, z(t_0)\} \tag{83}$$

denoted for short as $P(y,t|x)$, $P^\epsilon(x^\epsilon, y^\epsilon, t)$, $\bar{P}(y,t|x)$, and $\bar{P}_z(y,t)$, respectively. Define $P^x(\cdot)$ as the stationary measure of the fixed x system (78), and

$$F(x) = \int F_1(x,y)P^x(dy). \tag{84}$$

Let $E_t\{\cdot\}$ denote expectation with respect to $w(s)$: $s \le t$, and define $\rho = e^{t+t_\epsilon}$, as used in $E_\rho\{\cdot\}$. Define $x^0(\cdot)$ as the solution to $\dot{x}^0 = F(x^0)$. The following ten assumptions are needed, in which alternatives in any assumption are independent of the other assumptions.

Assumption A4) Equation (75) has a unique solution for all time, with P^ϵ Borel measurable.

Assumption A5) F_1 is uniformly continuous in x.

Assumption A6) Either: a) F_1 is bounded, or b) for fixed x, $\sup_{0 \le s \le \delta} \|F_1(x, y^\epsilon(t+s))\|$ is uniformly integrable.

Assumption A7) Either: a) F_2 is bounded, or b) for fixed x, $\sup_{0 \le s \le \delta} E_\rho\|F_2(x, y^\epsilon(t+s))\|^2$ is bounded.

Assumption A8) Either: a) $\{y^\epsilon(0)\}$ is tight, or b) if $\{y^\epsilon(0)\}$ is tight then $\{y^\epsilon(t)\}$ is tight.

Assumption A9) A unique invariant measure $P^x(\cdot)$ exists for the fixed x system (78) and is tight for x in compact sets in $\boldsymbol{R}^{n_x}$.

Assumption A10) Either: a) $y^\epsilon(t)$ is bounded, or b) for $\epsilon > 0$ there exists a set function $\beta_\epsilon(\cdot)$ such that for any compact set $B \subseteq \boldsymbol{R}^{n_y}$ then $\epsilon\beta_\epsilon(B) \to 0$ and for fixed T, $K > 0$ the collection of measures

$$\Big\{\bar{P}_{z^\epsilon}(y,s)I_B(y)\colon \Big[\text{compact } B \subseteq \boldsymbol{R}^{n_y}, y \in \boldsymbol{R}^{n_y}, \epsilon \in (0,1], \exp(\epsilon\beta_\epsilon(B) + t_\epsilon) \le s \le \exp(T + t_\epsilon), \text{ and } z^\epsilon(\cdot) \ni \sup_s |z^\epsilon(s)| \le K\Big]\Big\} \quad \text{is tight.}$$

Assumption A11) a) $\bar{P}(y,t|x)$ is Borel measurable, *and* b) (78) has a unique solution for each x.

Assumption A12) For continuous $f(\cdot)$, and $u > 0$, define $i = \int f(\bar{y})\bar{P}(y, u, d\bar{y}|x)$. Assume: a) i is bounded and continuous in x and y, and b) if $z^\epsilon(\cdot)$ is a bounded continuous nonanticipative process where $\lim_{\epsilon \to 0} \sup_{t_0 \le s \le u} \|z^\epsilon(s) - x\| = 0$ wp1, then $\lim_{\epsilon \to 0} \int f(\bar{y}) P_{z^\epsilon}(y, u, d\bar{y}) = i$ uniformly on compact x-sets.

Assumption A13) Define $j = \int F_1(x, \bar{y})\bar{P}(y, t, d\bar{y}|x)$. Assume: a) j is continuous in x and y, and b) if $\lim_{\epsilon \to 0} \sup_{t_0 \le s \le u} \|z^\epsilon(s) - \bar{x}_\epsilon\| = 0$ wp1, and $\|x - \bar{x}_\epsilon\| \to 0$, then $\lim_{\epsilon \to 0} \int F_1(\bar{x}_\epsilon, \bar{y}) P_{z^\epsilon}(y, t_0, d\bar{y}) = j$ uniformly on compact sets.

Theorem 5: Let $\{t_\epsilon\}$ for $\epsilon \in (0, 1]$ be such that $t_\epsilon \to \infty$ as $\epsilon \to 0$. Assume $x^\epsilon(0) \Rightarrow x(0)$, that (11) has a unique solution on all $[0, \infty)$, and that Assumptions A4)–A13) hold. (The notation $\Rightarrow$ means "converges weakly".) Then $\boldsymbol{F}(\cdot)$ is continuous and $x^\epsilon(\cdot) \Rightarrow x^0(\cdot)$.

The proof is presented as a series of six lemmas.

Lemma 1: $\boldsymbol{F}(\cdot)$ is continuous.

Proof: As in [16, theorem 2, ch. 5]. Q.E.D.

Before the next lemma, define $\delta_0 = 1$, $\{\delta_\epsilon\} \to 0$ such that $\delta_\epsilon e^{t_\epsilon} \to \infty$. Let differentiable $f(\cdot)$ have compact x-support. Define the difference and differential operators A_d^ϵ, A_s^ϵ, A^ϵ, and A by

$$A_d^\epsilon f(x^\epsilon(t)) = E_\rho\left\{\delta_\epsilon^{-1} \int_t^{t+\delta_\epsilon} f_x^T(x^\epsilon(\tau)) F_1(x^\epsilon(\tau), y^\epsilon(\tau))\, d\tau\right\} \tag{85}$$

$$A_s^\epsilon f(x^\epsilon(t)) = E_\rho\left\{\delta_\epsilon^{-1} \int_{e^{t+t_\epsilon}}^{e^{t+t_\epsilon+\delta_\epsilon}} \tau^{-1} f_x^T(x^\epsilon(\log\tau - t_\epsilon)) \cdot F_2(x^\epsilon(\log\tau - t_\epsilon), y^\epsilon(\log\tau - t_\epsilon))\, dw(\tau)\right\} \tag{86}$$

$$A^\epsilon f(x^\epsilon(t)) = A_d^\epsilon f(x^\epsilon(t)) + A_s^\epsilon f(x^\epsilon(t)) \tag{87}$$

$$Af(x^0(t)) = f_x^T(x^0(t)) \boldsymbol{F}(x^0(t)) \tag{88}$$

where f_x is the gradient of $f(\cdot)$. Also denote the integrals of (85)–(87) as

$$I_d^\epsilon(t) = \int_0^t A_d^\epsilon f(x^\epsilon(\tau))\, dt \tag{89}$$

$$I_s^\epsilon(t) = \int_0^t A_s^\epsilon f(x^\epsilon(\tau))\, d\tau \tag{90}$$

$$I^\epsilon(t) = I_d^\epsilon(t) + I_s^\epsilon(t) \tag{91}$$

and define $x_d^\epsilon(\cdot)$ and $x_s^\epsilon(\cdot)$ such that $x_s^\epsilon(0) = 0$ and

$$\begin{aligned} x^\epsilon(t) - x^\epsilon(0) &= [x_d^\epsilon(t) - x_d^\epsilon(0)] + x_s^\epsilon(t) \\ &= \int_0^t F_1\, d\tau + \int_0^t F_2 \tau^{-1} dw(\tau). \end{aligned} \tag{92}$$

Lemma 2: $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ is tight in $C(\boldsymbol{R}, \boldsymbol{R}^n) \times C(\boldsymbol{R}, \boldsymbol{R}^n)$.

Proof: It suffices to show that each of $\{x_d^\epsilon(\cdot)\}$, $\{x_s^\epsilon(\cdot)\}$, $\{I_d^\epsilon(\cdot)\}$, and $\{I_s^\epsilon(\cdot)\}$ is tight. Recall that a family of random processes, $\{z^\epsilon(\cdot)\}$, in $C(\boldsymbol{R}, \boldsymbol{R}^n)$ is tight if

a) $\{z^\epsilon(0)\}$ is tight in $\boldsymbol{R}^n$,

b) $\forall \eta, \nu, T > 0$ $\exists \epsilon', \delta'$ such that for $\epsilon \le \epsilon'$, $\delta \le \delta'$ then

$$P\{w_{z^\epsilon}(\delta, T) \ge \nu\} \le \eta$$

$$w_{z^\epsilon}(\delta, T) = \sup_{\substack{t \le T \\ |s| \le \delta}} \|z^\epsilon(t+s) - z^\epsilon(t)\|$$

or

b′) $\forall \eta, \nu, T > 0$ $\exists \epsilon', \delta'$ such that for $\epsilon \le \epsilon'$, $\delta \le \delta'$, $t \le T$ then

$$P\left\{\sup_{0 \le s \le \delta} \|z^\epsilon(t+s) - z^\epsilon(t)\| \ge \nu\right\} \le \delta\eta.$$

a) is tightness of the initial conditions and b) or b′) are equicontinuity conditions. b′) implies b) and is sometimes easier to prove. This is essentially [16, theorem 4, p. 30].

$x^{\epsilon_d}(0) = x^\epsilon(0)$ and $\{x^\epsilon(0)\}$ is tight since, by assumption, it is weakly convergent. $x^{\epsilon_s}(0) = I^{\epsilon_d}(0) = I^{\epsilon_s}(0) = 0$ wp1, which trivially implies tightness. Thus, a) is satisfied by each of the families of processes.

Consider the following inequalities:

$$\begin{aligned} \sup_{0 \le s \le \delta} \|x_d^\epsilon(t+s) - x_d^\epsilon(t)\| &= \sup_{0 \le s \le \delta} \left\| \int_t^{t+s} F_1(x^\epsilon(\tau), y^\epsilon(\tau))\, d\tau \right\| \\ &\le \delta \sup_{\substack{0 \le s \le \delta \\ x \in \boldsymbol{R}^{n_x}}} \|F_1(x, y^\epsilon(t+s))\|. \end{aligned} \tag{93}$$

Now

$$\begin{aligned} P\left\{\sup_{0 \le s \le \delta} \|x_d^\epsilon(t+s) - x_d^\epsilon(t)\| \ge \nu\right\} &\le P\left\{\sup_{\substack{0 \le s \le \delta \\ x \in \boldsymbol{R}^{n_x}}} \|F_1(x, y^\epsilon(t+s))\| \ge \frac{\nu}{\delta}\right\} \\ &\le \frac{\delta}{\nu} E\left\{\sup_{\substack{x \in \boldsymbol{R}^{n_x} \\ 0 \le s \le \delta}} \|F_1(x, y^\epsilon(t+s))\| \cdot \left| \sup_{\substack{x \in \boldsymbol{R}^{n_x} \\ 0 \le s \le \delta}} \|F_1(x, y^\epsilon(t+s))\| \ge \frac{\nu}{\delta} \right.\right\} \\ &= \frac{\delta}{\nu} O(\delta). \end{aligned} \tag{94}$$

The first inequality in (94) follows from (93), while the second is the Chebyshev inequality. The last follows from the uniform integrability in Assumption A6). Thus, $\{x_d^\epsilon(\cdot)\}$ satisfies b′) and is tight. When $F_1(\cdot, \cdot)$ is bounded, the proof is simpler and will not be shown. Similar bounds prove $\{x_s^\epsilon(\cdot)\}$, $\{I_d^\epsilon(\cdot)\}$, $\{I_s^\epsilon(\cdot)\}$, are tight. Q.E.D.

Now, by Lemma 2, every ϵ-subsequence taken from $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ has a convergent (sub) subsequence as $\epsilon \to 0$. If it can be shown that all convergent subsequences have the same limit, then $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ converges to a unique limit process. Let $J_0 \subseteq (0, 1]$, with $\inf J_0 = 0$, denote the

ϵ-indexes of a convergent subsequence. Denote that subsequence $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ for $\epsilon \in J_0$. Denote the limit process by $(x^0(\cdot), I^0(\cdot))$. Skorohod embedding [16, p. 29] permits the assumption that $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ converges to $(x^0(\cdot), I^0(\cdot))$ uniformly wp1 on $[0, T]$ for any finite T, and $\{x^\epsilon(\cdot)\}$ to x^0 and $\{I^\epsilon(\cdot)\}$ to $I^0(\cdot)$, weakly.

For Lemma 3, the following definitions are needed. By A8), there exist compact sets $B_{\epsilon,T}$ such that for $\epsilon \in J_0$,

$$\lim_{\epsilon\to 0} \inf_{t\le T} P\{y^\epsilon(t) \in B_{\epsilon,T}\} = 1. \tag{95}$$

By A10)-b), there is a set function $\beta_\epsilon(\cdot)$ such that for $\epsilon \in J_0$,

$$\lim_{\epsilon\to 0} \epsilon\beta_\epsilon(B_{\epsilon,T}) = 0 \tag{96}$$

and that δ_ϵ satisfies

$$\lim_{\epsilon\to 0} \delta_\epsilon^{-1}\epsilon\beta_\epsilon(B_{\epsilon,T}) = 0 \qquad \text{with } \epsilon\beta_\epsilon(B_{\epsilon,T}) \le \delta_\epsilon. \tag{97}$$

This is consistent with the assumption on δ_ϵ in the above Lemma 2. Denote the indicator function of a set B by $I_B(\cdot)$. Define the averaged probability measures Q, Q', Q_B, and Q'_B by

$$Q(t) = \delta_\epsilon^{-1}\int_t^{t+\delta_\epsilon} P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau)\, d\tau,$$
$$Q_B(t) = I_B(y^\epsilon(t))Q(t) \tag{98}$$

$$Q'(t) = \delta_\epsilon^{-1}\int_{t+\epsilon\beta_\epsilon(B)}^{t+\delta_\epsilon} P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau)\, d\tau,$$
$$Q'_B(t) = I_B(y^\epsilon(t))Q'(t). \tag{99}$$

Note that if one of these sequences of measures has a weak limit as $\epsilon \to 0$, then each of the others will have the same limit.

For each t, define the set of measures

$$M(t) = \{Q'_{B_{\epsilon,T}}(t) : \epsilon \in J_0\}. \tag{100}$$

Lemma 3: For each t, $M(t)$ is tight.

Proof: It is to be shown that $\forall\eta$, $\exists$ compact K_η and ϵ_0 such that for $\epsilon \in J_0$,

$$\inf_{\epsilon\le\epsilon_0} Q'_{B_{\epsilon,T}}(t, K_\eta) \ge 1 - \eta. \tag{101}$$

Let $\eta > 0$. In A10)-a), if $y^\epsilon(\cdot)$ has compact closure, then set K_η equal to its closure to get

$$Q'_{B_{\epsilon,T}}(t, K_\eta) = I_{B_{\epsilon,T}}(y^\epsilon(t))\,\delta_\epsilon^{-1} \cdot\int_{t+\epsilon\beta_\epsilon}^{t+\delta_\epsilon} P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau)\, d\tau \quad \text{over } K_\eta$$
$$= I_{B_{\epsilon,T}}(y^\epsilon(t))(1 - \delta_\epsilon^{-1}\epsilon\beta_\epsilon). \tag{102}$$

Since for each fixed $T > 0$, the range of $y^\epsilon(\cdot) \subseteq \cup B_{\epsilon,T}$, then the RHS $\to 1$ as $\epsilon \to 0$, proving tightness.

If the alternative hypothesis A10)-b) holds, set $z^\epsilon(t) = x^\epsilon(\log t - t_\epsilon)$ in the hypothesis statement and in

$$P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau) = P_{z^\epsilon}(y^\epsilon(\log t - t_\epsilon), e^{\tau + t_\epsilon}) \tag{103}$$

so that for fixed $T > \delta_\epsilon > 0$ the tightness assumed in A10)-b) implies the tightness of

$$\{P_\epsilon(x^\epsilon(t), y^\epsilon(t), \tau) I_{B_{\epsilon,t}}(y^\epsilon(t)) : \epsilon \in J_0, t + \epsilon\beta_\epsilon \le \tau \le t + \delta_\epsilon\}. \tag{104}$$

But the measures in $M(t)$ are averages of the measures of this, so $M(t)$ is tight also. Q.E.D.

Lemma 4: $M(t)$, as an ϵ-sequence of measures, $\Rightarrow$ $P^{x^0(t)}(\cdot)$ as $\epsilon \to 0$, $\epsilon \in J_0$.

Proof: By the tightness, it suffices to show that any weakly converging subsequence of $M(t)$ converges to $P^{x^0(t)}(\cdot)$. Let $J_1 \subseteq J_0$ be a subsequence with $\inf J_1 = 0$, $\epsilon \in J_1$, and the weak limit $\lim_{\epsilon\to 0} Q'_{B_{\epsilon,T}}(t, \epsilon, \cdot) = Q_0(\cdot)$. For $g(\cdot)$ a bounded, continuous function on $\boldsymbol{R}^{n_y}$, then for $\epsilon \in J_1$, using the Chapman–Komolgorov equation,

$$\lim_{\epsilon\to 0} \int g(y) Q'_{B_{\epsilon,T}}(t, \epsilon, dy)$$
$$= \lim_{\epsilon\to 0} \delta_\epsilon^{-1} \int_y g(y) \left\{\int_{t+\epsilon\beta_\epsilon}^{t+\delta_\epsilon} I_{B_{\epsilon,T}}(y^\epsilon(t)) \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau, t, dy)\, d\tau\right.$$
$$= \lim_{\epsilon\to 0} \delta_\epsilon^{-1} \int_{(\bar{x},\bar{y})} \int_{t+\epsilon u+\epsilon\beta_\epsilon}^{t+\delta_\epsilon} \int_y g(y) \cdot P^\epsilon(\bar{x}, \bar{y}, \tau, \tau - \tau_\epsilon, dy) \cdot I_{B_{\epsilon,T}} P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \tag{105}$$

where $\tau_\epsilon = -\log(1 - e^{(\tau + t_\epsilon)}u)$, $\tau_\epsilon \to 0$ as $\epsilon \to 0$, but $u \ne 0$. Then for $z^\epsilon(t) = x^\epsilon(\log(t + \exp(\tau - \tau_\epsilon + t_\epsilon) - t_\epsilon)$,

$$\int g(y) P^\epsilon(\bar{x}, \bar{y}, \tau, dy) = \int g(y) \bar{P}^\epsilon(\bar{x}, \bar{y}, e^{\tau + t_\epsilon}, dy)$$
$$= \int g(y) \bar{P}_z^\epsilon(\bar{y}, u, dy). \tag{106}$$

Then, by A12)-b)

$$\lim_{\epsilon\to 0} \int g(y) P^\epsilon(\bar{x}, \bar{y}, \tau, dy) = \int g(y) \bar{P}(\bar{y}, u, dy \mid \bar{x}). \tag{107}$$

Combining (105)–(107) gives

$$\lim_{\epsilon\to 0} \int g(y) Q'_{B_{\epsilon,T}}(\epsilon, t, dy) = \lim_{\epsilon\to 0} \int \left[\int g(y) \bar{P}(\bar{y}, u, 0 \mid x^\epsilon(t))\right] Q'_{B_{\epsilon,T}}(\epsilon, t, d\bar{y}). \tag{108}$$

Since $Q' \Rightarrow Q'_0$, then the LHS is equal to $\int g(y) Q_0(dy)$. By A12)-a), the bracketed term is bounded and continuous. Since t is fixed, the weak convergence of $x^\epsilon(\cdot)$ and Skorohod embedding imply the wp1 convergence of $x^\epsilon(t)$ to $x^0(t)$ as $\epsilon \to 0$ in J^0. Then $x^\epsilon(t) \Rightarrow x^0(t)$, the continuity and weak convergence of Q' to Q'_0 imply

$$\int g(y)Q_0(dy) = \int \left[\int g(y)\bar{P}(\bar{y}, u, dy \mid \bar{x}) \right] Q_0(d\bar{y}). \tag{109}$$

Therefore, $Q_0(\cdot)$ is an invariant measure of (78), and by A9) $Q_0(\cdot) = P^{x_0(t)}(\cdot)$. Q.E.D.

Lemma 5: For $\epsilon \in J_0$, given $\{x^\epsilon(\cdot), I^\epsilon(\cdot)\}$ and its weak limit $(x^0(\cdot), I^0(\cdot))$ then

$$I^0(t) = \int_0^t Af(x^0(\tau))\, d\tau = \int_0^t f_x^T(x^0(\tau))F(x^0(\tau))\, d\tau. \tag{110}$$

Proof: Since $I^\epsilon = I_d^\epsilon + I_s^\epsilon$, from [29, theorem 4.1] it suffices to show that $I_s^\epsilon \to 0$ in probability uniformly on compact intervals and that $A_d^\epsilon f(\cdot)$ converges to $Af(x^0(\cdot))$ uniformly wp1 on compact intervals.

Define $G_2(t) = f_x^T(x^\epsilon(t))\; F_2(x^\epsilon(t), y^\epsilon(t))$. Let $f(\cdot)$ be continuous. Then (86) and (90) can be combined as

$$I_s^\epsilon f(t) = \int_0^t E_{\rho_0}\left\{ \delta_\epsilon^{-1} \int_{\rho_0}^{\rho_1} \sigma^{-1} G_2(\sigma)\, dw(\sigma) \right\} d\tau \tag{111}$$

where $\rho_1 = e^{\tau + \delta_\epsilon \tau + t_\epsilon}$ and $\rho_0 = e^{\tau + t_\epsilon}$. Then, from 2.2(g) of [22, ch. 1] and A7)

$$\begin{aligned} &P\{|I_s^\epsilon f(t)| \geq \nu\} \\ &\quad \leq P\left\{ t \sup_{\tau \leq t} E_{\rho_0}\left[\delta_\epsilon^{-1} \int_{\rho_0}^{\rho_1} \sigma^{-1} G_2(\sigma)\, dw(\sigma) \right] \geq \nu \right\} \\ &\quad \leq K\nu^{-2}\delta_\epsilon^{-2} e^{-t-t_\epsilon} t^2 (e^{\delta_\epsilon} - 1). \end{aligned} \tag{112}$$

Hence, $I_s^\epsilon \to 0$ in probability uniformly on compact sets.

Let

$$\tau_\epsilon = -\log(1 - e^{(t+t_\epsilon)}u),$$

$$z^\epsilon(t) = x^\epsilon\big(\log(t + e^{(t - \tau_\epsilon + t_\epsilon)}) - t_\epsilon\big),$$

and

$$G_1(x, y) = f_x^T(x) F_1(x, y).$$

Notice $\tau_\epsilon \to 0$ as $\epsilon \to 0$, but u is nonzero constant. As before, with $z^\epsilon(t) = x^\epsilon(\log(t + \exp(\tau - \tau_\epsilon + t_\epsilon)t) - t_\epsilon)$, then $z^\epsilon(0) = x^\epsilon(\tau - \tau_\epsilon)$ and $z^\epsilon(u) = x^\epsilon(\tau)$. Now it can be shown that $A_d^\epsilon f(\cdot)$ converges to $Af(x^0(\cdot))$ uniformly wp1 on compact intervals. Assumptions A5), A6)-b), A7) and the weak convergence imply, for $\epsilon \in J_0$,

$$\lim_{\epsilon \to 0} E\left\{ |G_1(x^\epsilon(\tau), y^\epsilon(\tau)) - G_1(x^\epsilon(\tau - \tau_\epsilon), y^\epsilon(\tau))| \,\middle|\, \begin{matrix} x^\epsilon(\tau - \tau_\epsilon) = \bar{x} \\ y^\epsilon(\tau - \tau_\epsilon) = \bar{y} \end{matrix} \right\} = 0. \tag{113}$$

By the Chapman–Komolgorov equation,

$$\begin{aligned} &\lim_{\epsilon \to 0} A_d^\epsilon f(x^\epsilon(t)) \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} E_{e^{t+t_\epsilon}}\left\{ \int_t^{t+\delta_\epsilon} G_1(x^\epsilon(\tau), y^\epsilon(\tau))\, d\tau \right\} \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} \int_{(\bar{x}, \bar{y})} \cdot E\left\{ G_1(x^\epsilon(\tau), y^\epsilon(\tau)) \,\middle|\, \begin{matrix} x^\epsilon(\tau - \tau_\epsilon) = \bar{x} \\ y^\epsilon(\tau - \tau_\epsilon) = \bar{y} \end{matrix} \right\} \\ &\qquad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} \int_{(\bar{x}, \bar{y})} \cdot E\left\{ G_1(x^\epsilon(\tau - \tau_\epsilon), y^\epsilon(\tau)) \,\middle|\, \begin{matrix} x^\epsilon(\tau - \tau_\epsilon) = \bar{x} \\ y^\epsilon(\tau - \tau_\epsilon) = \bar{y} \end{matrix} \right\} \\ &\qquad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} \int_{(\bar{x}, \bar{y})} \int_{(x, y)} \cdot G_1(\bar{x}, y) P^\epsilon(\bar{x}, \bar{y}, \tau, \tau - \tau_\epsilon, dx dy) \\ &\qquad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} \int_{(\bar{x}, \bar{y})} \int_y G_1(\bar{x}, y) \bar{P}_{z^\epsilon}(\bar{y}, u, 0, dy) \\ &\qquad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} \int_{(\bar{x}, \bar{y})} \int_y G_1(\bar{x}, y) \bar{P}(\bar{y}, u, 0, dy \mid \bar{x}) \\ &\qquad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau - \tau_\epsilon, t, d\bar{x}\, d\bar{y})\, d\tau \\ &\quad = \lim_{\epsilon \to 0} \frac{1}{\delta_\epsilon} \int_t^{t+\delta_\epsilon} E_{e^{t+t_\epsilon}}\{G_a(x^\epsilon(\tau), y^\epsilon(\tau))\}\, d\tau \end{aligned} \tag{114}$$

where $G_a(\cdot,\cdot)$ is defined as $G_a(\bar{x}, \bar{y}) = \int G_1(\bar{x}, y)\bar{P}(\bar{y}, u, 0, dy \mid \bar{x})$. By A13)-a), $G_a(\cdot,\cdot)$ is uniformly continuous (in both variables). Using a proof similar to that for (113), then

$$\lim_{\epsilon \to 0} E_{e^{t+t_\epsilon}}\{|G_a(x^\epsilon(\tau), y^\epsilon(\tau)) - G_a(x^\epsilon(t), y^\epsilon(\tau))|\} = 0 \tag{115}$$

holds uniformly on compact (τ, x, y)-sets. Continuing on from (114), then

$$\begin{aligned}
&\lim_{\epsilon\to 0} A_d^\epsilon f(x^\epsilon(t)) \\
&= \lim_{\epsilon\to 0} \frac{1}{\delta_\epsilon}\int_t^{t+\delta_\epsilon}\int_{(\bar{x},\bar{y})} G_a(\bar{x},\bar{y}) \\
&\quad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau-\tau_\epsilon, t, d\bar{x}\,d\bar{y})\,d\tau \\
&= \lim_{\epsilon\to 0} \frac{1}{\delta_\epsilon}\int_t^{t+\delta_\epsilon}\int_{(\bar{x},\bar{y})} G_a(x^\epsilon(t),\bar{y}) \\
&\quad \cdot P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau-\tau_\epsilon, t, d\bar{x}\,d\bar{y}) \\
&= \lim_{\epsilon\to 0}\int_{\bar{y}} G_a(x^\epsilon(t),\bar{y}) \\
&\quad \cdot\left(\frac{1}{\delta_e}\int_t^{t+\delta_\epsilon} P^\epsilon(x^\epsilon(t), y^\epsilon(t), \tau-\tau_e, \boldsymbol{R}^{n_x}\times d\bar{y})\,d\tau\right) \\
&= \lim_{\epsilon\to 0}\int G_a(x^\epsilon(t), y) Q(t,\epsilon,dy) \\
&= \int G_a(x^0(t),\bar{y}) P^{x_0(t)}(d\bar{y}) \\
&= \int_{\bar{y}}\int_y G_1(x^0(t), y)\bar{P}(\bar{y}, u, 0, dy \mid x^0(t)) P^{x_0(t)}(d\bar{y}) \\
&= \int G_1(x^0(t),\bar{y}) P^{x_0(t)}(d\bar{y}). \qquad (116)
\end{aligned}$$

Since $Q(t, \epsilon, dy) = P^{x_0(t)}(d\bar{y})$, use the definition of G_a and G_1 to get

$$\begin{aligned}
\lim_{\epsilon\to 0} A_d^\epsilon f(x^\epsilon(t)) &= \int f_x^T(x^0(t)) F_1(x^0(t),\bar{y}) P^{x^0(t)}(d\bar{y}) \\
&= f_x^T(x^0(t))\boldsymbol{F}(x^0(t)).\ \text{Q.E.D.} \qquad (117)
\end{aligned}$$

Having identified the limit of the convergent subsequence $\{I^\epsilon(\cdot)\}$ for $\epsilon\in J_0$, it remains to be shown that $x^0(\cdot)$ satisfies the ODE (11).

Lemma 6:

$$\begin{aligned}
f(x^0(t+s)) - f(x^0(t)) &= \int_t^{t+s} f_x^T(x^0(\tau))\boldsymbol{F}(x^0(\tau))\,d\tau \\
&= \int_t^{t+s} Af(x^0(\tau))\,d\tau. \qquad (118)
\end{aligned}$$

Proof: Using Itô differentiation of $f(x(\cdot))$ for $x(\cdot)$ obeying (73),

$$df = t^{-1}f_x^T F_1 dt + t^{-1} f_x^T F_2 dw + F_2^T f_{xx} F_2 dt/2t^2. \quad (119)$$

A proof similar to Lemma 2 shows the second term on the RHS converges to zero in probability uniformly on compact t-sets. Assumption A7) implies the third term does the same. The weak convergence as $\epsilon\to 0$ in J_0 implies

$$E_\rho f(x^\epsilon(t+s)) - f(x^\epsilon(t)) \Rightarrow f(x^0(t+s)) - f(x^0(t)). \quad (120)$$

It suffices to show that $E_\rho f_x^T F_1 dt \Rightarrow f_x^T \boldsymbol{F} dt$. Now define $D = E_\rho\{\delta_\epsilon^{-1}\int_t^{\tau+\delta_\epsilon} f_x^T F_1\, d\sigma\}dt$. From Lemma 5, $D \Rightarrow f_x^T \boldsymbol{F}\,dt$. Therefore, it suffices to show that $D - Ef_x^T F_1\,dt \to 0$ in probability uniformly on compact intervals. Since the Riemann sums converge uniformly in probability to their respective integrals, it suffices to note that $E_\rho f_x^T F_1 dt$ is actually a Riemann sum for D. Thus the weak convergence of D gives the proof. Q.E.D.

Equation (118) holds for any bounded continuous function $f(\cdot)$. Since the integrand is continuous, dividing both sides of (118) by s and taking limits as $s\to 0$ gives

$$df(x^0(t))/dt = f_x^T(x^0(t))\boldsymbol{F}(x^0(t)) \quad (121)$$

which implies the ODE (11). This proves Theorem 5.

Lemma 7: A unique asymptotic invariant measure exists for the z system of (7)–(10) under the hypothesis of Theorem 1, when θ is held constant.

Proof: In Theorem 1, $u(\cdot)$ is any linear combination of output from an asymptotically stationary stable linear system excited by additive $w(\cdot)$ and of linear stabilizing feedback of z_1 as given by (9) and (10).

Under the conditions of Lemma 7, θ is constant. Also, since (8) is not stochastic, assumption e) implies the asymptotic invariant measure is concentrated at $z_{2\infty}(\theta)$ in the z_2 variable. Hence, the asymptotic invariant measure is the product of an atomic measure in z_2 with the measure associated with z_1 and z_3. Consequently, (10) is asymptotically a stable, constant coefficient linear differential equation, additively excited by the Wiener process $w(\cdot)$. Linearity and stability assure existence and uniqueness of the resulting asymptotic invariant measure. Q.E.D.

Theorem 6: Let H_1 and H_2 of (76) be such that $y(t) = (z_1^T(t), z_2^T(t))^T$ and $x(t) = \theta(t)$ as in Theorem 1. Let $t_\epsilon\to\infty$ and define the accelerated system using $y_\epsilon(t) = \bar{y}(e^{t_\epsilon}, x_\epsilon)$, and $\{x_\epsilon\}$ a tight collection of random variables such that x_ϵ is independent of $\{w(s)\colon s\ge e^{t_\epsilon}\}$. With this correspondence, if a)–f) of Theorem 1 are satisfied, then Assumptions A3)–A12) are also satisfied and $\{x^\epsilon(t)\colon t\ge t_0, \epsilon>0\}$ is tight.

Proof: Assumptions A5), A9), and A11)-b) are explicit hypotheses of Theorem 6. A4) is a direct consequence of [30, theorems 6.34 and 10.22] using assumptions a) and c). A8) follows from f) and that $\{z_1(t)\}$ can be shown tight, and a similar argument can be used for A10). Also, a similar argument shows $\bar{y}(t,x)$ has finite moments of all order, independent of t. This and the quadratic bounds in c) imply A6)-b) and A7)-b). [30, theorems 6.34 and 10.22] imply (76) has a unique (in the sense of distributions) solution and the transition probabilities are measurable and Feller continuous as functions of (t, y). This and the continuity of $F_1(\cdot,\cdot)$ imply A11)-a), A12)-a), and A13)-a). Finally, A12)-b) and A13)-b) follow from the parametric continuity of solutions of an ODE. $\{x^\epsilon(t), \epsilon>0, t\ge 0\}$ is tight because $F_1(\cdot,\cdot)$ and $F_2(\cdot,\cdot)$ have compact x-support.

For Theorem 7, the following assumptions are needed:

B1) $x^0(\cdot)$ is the solution to the ODE $dx^0/dt = \boldsymbol{F}(x^0)$ which has a unique solution.

B2) a) There exists a set R_c such that $x^0(t)\to R_c$ for any $x^0(0)$, with uniform convergence on compact $(x^0(0))$ sets.

b) Let $\lambda(\cdot)$ be a differentiable monotone increasing function with $\lambda(0) = 0$. Then for all $x^0(0)$ and fixed $T > 0$, $\|x^0(\lambda(t + T)) - x^0(\lambda(t))\| \to 0$ as $t \to \infty$ with uniform convergence on compact $(x^0(0))$ sets.

B3) a) $F(\cdot)$ is twice continuously differentiable and $\exists \epsilon_0 > 0$ such that for all $\rho > 0$,

$$\lim_{T\to\infty} P\left\{\sup_{\epsilon<\epsilon_0} \sup_{t\geq T} \|x^\epsilon(t) - x_0\| < \rho\right\} \geq \pi_0 > 0.$$

b) $x^\epsilon(t)$ has positive definite covariance.

C1) If $x^\epsilon(0) \Rightarrow x^0(0)$, then $x^\epsilon(\cdot) \Rightarrow x^0(\cdot)$.

C2) If $s_\epsilon \to s_0 \leq \infty$ and $x^\epsilon(s_\epsilon) \Rightarrow x^0(0)$, then $x^\epsilon(s_\epsilon + \cdot) \Rightarrow x^0(\cdot)$.

C3) $\exists \epsilon_0 > 0$ such that $\{x^\epsilon(t): \epsilon \in (0, \epsilon_0]\}$ is tight.

Theorem 7: If $\{x^\epsilon(\cdot)\}$ satisfies C1)–C3), and $x^0(\cdot)$ satisfies B1), then

a) if $x(\cdot)$ satisfies B2)-a), then $\exists \epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$ then $x^\epsilon(t) \to R_c$ wp1 as $t \to \infty$, and if $x^0(\cdot)$ satisfies B2)-b), then $\exists \epsilon_0 > 0$ such that for all $\epsilon < \epsilon_0$ then $\|x^\epsilon(\lambda(t + T)) - x^\epsilon(\lambda(t))\| \to 0$ as $t \to \infty$, wp1.

b) if B3)-a) holds, then $F(x_0) = 0$. If, in addition, B3)-b) holds, then $F(x_0)$ has all its eigenvalues in the closed left-half plane.

Proof of Theorem 7-a): To show $x^\epsilon \to R_c$ wp1, it must be shown that if $\exists \epsilon_0$ such that if $\epsilon \leq \epsilon_0$, then $P\{\lim_{t\to\infty} d(x^\epsilon(t), R_c) = 0\} = 1$. Proceeding by contradiction, assume $\exists \delta > 0$, $\epsilon_k \to 0$, $t_k \to \infty$ such that $P\{d(x^{\epsilon_k}(t_k), R_c) \geq \delta\} \geq \pi_0 > 0$. Let $\gamma_k(\tau) = t_k - \min(\tau, t_k)$. For $\tau \geq 0$, C3) states that $\{x^{\epsilon_k}(\gamma_k(\tau))\}$ is tight, and so for each τ there is a weakly convergent subsequence. Denote the subsequential indexes as k, also, and the limit as $x_{0_\tau} \Leftarrow \{x^{\epsilon_k}(\gamma_k(\tau))\}$. Therefore, $\{x_{0_\tau}\}$ is tight by C3), and so $\exists M, \tau_0$ such that $P\{x_{0_\tau} \notin B_M\} \leq \pi_0/4$ for $\tau \geq \tau_0$, where B_M is the ball of radius M in R^{n_x}. For any t and $\tau \geq \tau_0$, then

$$\begin{aligned}
P\{d(x^{\epsilon_k}(\gamma_k(\tau) + t), R_c) \geq \delta\}&\\
\leq P\{x_{0_\tau} \notin B_M\}&\\
+P\left\{d(x^{\epsilon_k}(\gamma_k(\tau) = t), R_c) \geq \delta \middle| x_{0_\tau} \in B_M\right\}&\\
\leq P\left\{d(x^{\epsilon_k}(\gamma_k(\tau) + t), x_\tau^0(t)) \geq \delta/2 \middle| x_{0_\tau} \in B_M\right\}&\\
+P\left\{d(x_\tau^0(t), R_c) \geq \delta/2 \middle| x_{0_\tau} \in B_M\right\} + \frac{\pi}{4}& \quad (122)
\end{aligned}$$

where $x_\tau^0(\cdot)$ is the ODE (11) solution with initial condition x_{0_τ}. By B2), choose $\tau_1 \geq \tau_0$ such that $d(x^0(t), R_c) < \delta/2$. From C2), $x^{\epsilon_k}(\gamma_k(\tau) + \cdot) \Rightarrow x_\tau^0(\cdot)$. Skorohod embedding assures the convergence is uniform wp1 on compact t-intervals, namely, $[0, 2\tau_1]$. Set $t = \tau = \tau_1$ in (122) above to get

$$\begin{aligned}
\pi_0 \leq P\{d(x^{\epsilon_k}(\gamma_k(\tau) + t), R_c) \geq \delta\}&\\
\leq \frac{\pi_0}{4} + \frac{\pi_0}{4} + 0 = \frac{\pi_0}{2}.& \quad (123)
\end{aligned}$$

This contradiction proves the first part of Theorem 7-a). The second part can be proven in a somewhat similar manner. To prove Theorem 7-b), first a lemma is needed.

Lemma 8: Assume $0 < \eta_0 \leq \|F(x)\| \leq \eta_1$ and $\|F_x(x)\| \leq \eta_2$ for each $x \in B(x_c, \rho_0)$. Let $x^0(\cdot; x)$ be the solution to the ODE (11) with initial condition x. Then $\exists \rho_1$, $0 < \rho_1 < \rho_0$, and $T_1 > 0$ such that if $x \in B(x_c, \rho_1/2)$ then $x^0(\cdot; x)$ exits $B(x_c, \rho_1)$ within time T_1.

Proof: Since $F(x)$ is differentiable, $\|x^0(t; x) - x\| = \|F(x)t + f(t)\|$ where $\|f(t)\| \leq k\eta_1\eta_2 t^2$. Then $\|x^0(t; x) - x\| \geq |\eta_0 t - k\eta_1\eta_2 t^2|$. Setting $T_1 = \eta_0/2k\eta_1\eta_2$ and $\rho_1 = \eta_0^2/6k\eta_1\eta_2$ with $x \in B(x_c, \rho/2)$ assures that $x^0(\cdot; x)$ exits $B(x_c, \rho_1)$ within time T_1. Q.E.D.

Proof of Theorem 7-b): Proceeding by contradiction, assume $F(x_0) \neq 0$. Then $\exists \rho_0, \eta_0, \eta_1, \eta_2$ such that $0 < \eta_0 \leq \|F(x)\| \leq \eta_1$ and $\|F_x(x)\| \leq \eta_2$ for each $x \in B(x_0, \rho)$. By Lemma 8, $\exists \rho_1$, $0 < \rho_1 < \rho_0$, and $T_1 > 0$ such that if $x \in B(x_0, \rho_1/2)$ then $x^0(\cdot; x)$ exits $B(x_0, \rho_1)$ within T_1.

By hypothesis, $\exists T_2$ such that

$$P\left\{\sup_{t\geq T_2} \|x^\epsilon(t) - x_0\| \leq \rho_1/4\right\} \geq \pi_0.$$

By C3), $\{x^\epsilon(t)\}$ is tight, so $\exists z$ and $s_{\epsilon_\kappa} \to 2T_2$ such that $T_2 \leq s_{\epsilon_\kappa} \leq 2T_2$ and $x^{\epsilon_k}(s_c) \Rightarrow z$. Since $x_{\epsilon_k} \geq T_2$,

$$P\left\{\sup_k \left\|x^{\epsilon_k}(s_{\epsilon_k}) - x_0\right\| < \rho_1/4\right\} \geq \pi_0. \quad (124)$$

The weak convergence and C2) imply $x^{\epsilon_k}(s_{\epsilon_k} + \cdot) \Rightarrow x^0(\cdot)$ for $x^0(0) = z$. From the triangle inequality for $t \in [0, t_1]$,

$$\begin{aligned}
\sup_t \|x^0(t) - x_0\| \leq \sup_t \left\|x^0(t)\text{-}x^{\epsilon_k}(s_{e_k} + t)\right\|&\\
+ \sup_t \left\|x^{\epsilon_k}(s_{\epsilon_k} + t) - x_0\right\|.& \quad (125)
\end{aligned}$$

Refer to the three terms in the above (125) as A, B_k, and C_k. Then

$$P\{A \geq \rho_1/2\} \leq P\{B_k \geq \rho_1/4\} + P\{C_k \geq \rho_1/4\}. \quad (126)$$

Either $z \in B(x_c, \rho_1/2)$ and $x^0(\cdot; z)$ exits $B(x_c, \rho_1)$ within T_1 or $z \notin B(x_c, \rho_1/2)$, implying $A \geq \rho_1/2$ and $P\{A \geq \rho_1/2\} = 1$. By (122), $P\{C_k \geq \rho_1/4\} < 1 - \pi_0$. The weak convergence $x^{\epsilon_k}(s_{\epsilon_k} + \cdot) \Rightarrow x^0(\cdot)$ and Skorohod embedding imply $P\{B_k \geq \rho_1/4\} \to 0$. Then (124) gives a contradiction. q.e.d. for the first part of Theorem 7-b).

Next, assume $F_x(x_0)$ has an eigenvalue in the open right-half plane and denote the eigenpair by (λ, ν). Define $\zeta^0(t) = x^0(t) - x_0$ and $\zeta^\epsilon(t) = x^\epsilon(t) - x_0$ and linearize to get $d\zeta^0/dt = F_x(x_0)\zeta^0 + f(t)$, where $\|f(t)\| \leq K\|\zeta^0(t)\|^2$. Define $r^0(\cdot) = \nu^T\zeta^0(\cdot)$ and $r_\epsilon(\cdot) = \nu^T\zeta_\epsilon(\cdot)$, which satisfy

$$\lim_{T\to\infty} P\left\{\sup_{\epsilon<\epsilon_0} \sup_{t\leq T} \|\zeta^\epsilon(t)\| < \rho\right\} \geq \pi_0 > 0$$

and

$$\lim_{T\to\infty} P\left\{\sup_{\epsilon<\epsilon_0} \sup_{t\leq T} \|r^\epsilon(t)\| < \rho\right\} \geq \pi_0 > 0.$$

If $\zeta^0(t) \in B(x_0, \rho)$, the triangle inequality gives

$$\begin{aligned}
|r^0(0)| \exp(\operatorname{Re}\lambda - K\rho^2)t&\\
< |r^0(t)| \leq |r^0(t) - r^\epsilon(t)| - |r^\epsilon(t)|.& \quad (127)
\end{aligned}$$

Choose ρ small enough that $\operatorname{Re}\lambda - K\rho^2 > 0$ since $\lambda \in$ RHP. Note $r^0(0) \neq 0$ wp1, since the covariance of $x^\epsilon(t)$ is positive definite for each $t \geq 0$ and $\epsilon > 0$. Then reasoning as in the first part of this proof again leads to the contradiction $1 \leq 1 - \pi_0$. Q.E.D.

Theorem 8: Assume $x(\cdot)$ satisfies the hypotheses of Theorem 6 and that the ODE (11) has a unique solution defined for all time.

a) If for any initial conditions $x^0(t) \to R_c$, then $x(t) \to R_c$ wp1 as $t \to \infty$.

b) If for any initial conditions $\|x^0(\log(t+T)) - x^0(\log(t))\| \to 0$, then $\|x(t+T) - x(t)\| \to 0$ wp1 as $t \to \infty$.

c) If $F(x)$ is twice continuously differentiable, $x(t) \to x_\infty$ with probability strictly greater than zero, and $x(t)$ has positive definite covariance for each t, then x_∞ is a stable stationary point for the ODE (11).

Proof of Theorem 8-a): Let $t_\epsilon \to \infty$ and define the accelerated process for $x(\cdot)$ via (75). Theorem 6 implies the accelerated process satisfies the hypothesis of Theorem 5 and C3). Theorem 7 applies, so $\exists\epsilon_0$ such that $x^{\epsilon_0}(t) \to R_c$ wp1 for any choice of $x^{\epsilon_0}(0)$. Choosing $x^{\epsilon_0}(0) = x(e^{t_{\epsilon_0}})$, then $x(e^{t+t_{\epsilon_0}}) = x^{\epsilon_0}(t)$ and therefore $x(t) \to R_c$ wp1.

Proof of Theorems 8-b) and 8-c): Similar to the proof of Theorem 8-a).

Appendix B
Completion of the Proof of Theorem 3

This Appendix completes the proof of Theorem 3 by showing that A2), A3), and Assumptions 1)–4) of Theorem 3 guarantee that Theorem 1 applies to (57)–(64). From equations (57)–(64) let $(\theta^T, r)^T$, $(x^T, \hat{x}^T, \operatorname{col} W^T)^T$, and $(\operatorname{col} P^T, \operatorname{col}\Pi_i^T)$ be θ, z_1, and z_2 in Theorem 1. Then in (6)–(8) the system matrices are (where matrices are written instead of the column notation)

$$F_1(\theta, z) = \left(\left[r^{-1}W^TC^TR^{-1}(C_0x - C\hat{x})\right]^T \quad \cdot\operatorname{tr}(W^TC^TR^{-1}CW) - r\right)^T$$

$$F_2(\theta, z) = \left(\left[r^{-1}W^TC^TR^{-1}H\right]^T \quad 0\right)^T$$

$$A_1(\theta, z_2) = \begin{pmatrix} A_0 & 0 & 0 & \cdots & 0 \\ PC^TR^{-1}C_0 & \mathscr{A}(\theta, P) & 0 & \cdots & 0 \\ \Pi_1C^TR^{-1}C_0 & (\partial A/\partial\theta_1)(\theta) - \Pi_1C^TR^{-1}C & \mathscr{A}(\theta, P) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \Pi_{n_\theta}C^TR^{-1}C_0 & (\partial A/\partial\theta_{n_\theta})(\theta) - \Pi_{n_\theta}C^TR^{-1}C & 0 & \cdots & \mathscr{A}(\theta, P) \end{pmatrix}$$

$$B_1(\theta, z_2) = \begin{pmatrix} B_0 \\ B(\theta) \\ (\partial B/\partial\theta_1)(\theta) \\ \vdots \\ (\partial B/\partial\theta_{n_\theta})(\theta) \end{pmatrix} \quad \text{and} \quad S_1(\theta, z_2) = \begin{pmatrix} \Sigma_0 & 0 \\ 0 & PC^TR^{-1}H \\ 0 & \Pi_1C^TR^{-1}H \\ \vdots & \vdots \\ 0 & \Pi_{n_\theta}C^TR^{-1}H \end{pmatrix}$$

$$A_2(\theta, z_2) = \begin{pmatrix} A(\theta)P + PA^T(\theta) + \Sigma(\theta)\Sigma^T(\theta) - PC^TR^{-1}CP \\ \mathscr{A}\Pi_i + \Pi_i\mathscr{A}^T + (\partial A/\partial\theta_i)P + P(\partial A/\partial\theta_i)^T + (\partial\Sigma/\partial\theta_i)\Sigma^T + \Sigma^T(\partial\Sigma/\partial\theta_i) \end{pmatrix}. \tag{128}$$

From Theorem 6, the hypotheses a)–f) of Theorem 1 must be satisfied by the above. The functions (128) above are locally bounded and have locally bounded second derivatives, except F_1 and F_2 at $r = 0$. The solution of (58) guarantees $r(t) \geq r_0 > 0$ for all $t \geq t_0 > 0$, so the first part of hypothesis a) is satisfied. Call Φ the transition matrix A_t. By Assumption 2),

$$\|\Phi(t, \tau)\| \leq c_1e^{-c_2(t-\tau)} \qquad \text{for } c_1, c_2 > 0. \tag{129}$$

Then (61) can be written as

$$P(t) = \Phi(t, t_0)P(t_0)\Phi^T(t, t_0) + \int_{t_0}^t \Phi(t, \tau) \cdot\left[\Sigma_\tau\Sigma_\tau^T - P(\tau)C^TR^{-1}CP(\tau)\right]\phi^T(t, \tau)\, d\tau. \tag{130}$$

Using Assumptions 1) and 2) and taking norms gives

$$\|P(t)\| \leq c_1^2e^{-2c_2(t-t_0)}\|P(t_0)\| + \int_{t_0}^t c_1^2e^{-2c_2(t-\tau)}\|\Sigma_\tau\Sigma_\tau^T\|\, d\tau. \tag{131}$$

Then Assumption 3) assures that $P(t)$ is bounded from above.

Assumption 1) bounds $P(t)$ from below, and guarantees $\mathscr{A} = A_t - P(t)C^TR^{-1}C$ is asymptotically stable. This and Assumption A2) satisfy hypothesis d). Furthermore, $\Pi_i(t)$ is bounded as the solution of (64), a time-varying asymptotically stable linear system. Then hypothesis (f) is satisfied, and with continuity assured by the results of [31], so is hypothesis e). Also, S_1 is bounded, satisfying the last part of hypothesis a). Assumption A3) implies b) is satisfied. To verify the θ part of c) for F_1 and F_2, note that $r(t) \geq r_0 > 0$ so that $r_0^{-1} \geq r^{-1}(t) \geq 0$. Furthermore, since F_1 and F_2 are independent of z_2 and at worst quadratic in z_1, only the A_1, B_1, and A_2 parts of hypothesis c) remain to be investigated. Assumptions 3) and 4) assure the θ part of c) for A_1B_1, and A_2. Also, B_1 depends only on θ, leaving the z_1 and z_2 parts of c) for A_1 and A_2. Since hypothesis d) has already been shown for A_1, then $A_1(\theta, z_2)$ is negative and c) holds. Finally, $A_2(\theta, z_2)$ is linear in z_2 except for the $PC^TR^{-1}CP$ term. This is nonnegative, satisfying the last part of the last hypothesis c) to be verified.

ACKNOWLEDGMENT

The second author would like to acknowledge stimulating discussions with G. C. Goodwin, L. Ljung, H. J. Kushner, and I. M. Y. Mareels.

REFERENCES

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

[2] L. Ljung, "Analysis of Recursive Stochastic Algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, no. 4, pp. 551–575, 1977.

[3] G. C. Goodwin and D. Q. Mayne, "A parameter estimation perspective of continuous-time model reference adaptive control," *Automatica*, vol. 23, no. 1, pp. 57–70, 1987.

[4] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation.* Englewood Cliffs, NJ: Prentice-Hall, 1990.

[5] D. M. Wiberg and D. G. DeWolf, "A convergent approximation of the continuous-time optimal parameter estimator," *IEEE Trans. Automat. Contr.*, this issue.

[6] H. F. Chen, *Recursive Estimation and Control for Stochastic Systems*, Wiley, New York, 1985.

[7] J. H. van Schuppen, "Convergence results for continuous-time adaptive stochastic filtering algorithms," *J. Math. Anal. Appl.*, vol. 96, no. 1, pp. 209–225, 1983.

[8] J. B. Moore, "Convergence of continuous-time stochastic ELS parameter estimation," Int. Rep., Dept. Syst. Eng., Australian National Univ., Canberra, 1986.

[9] H. F. Chen and J. B. Moore, "Convergence rates of continuous-time stochastic ELS parameter estimation," *IEEE Trans. Automat. Contr.*, vol. AC-31, no. 3, pp. 267–269, 1987.

[10] M. Gevers, G. C. Goodwin, and V. Wertz, "Continuous-time stochastic adaptive control," *SIAM J. Contr. Optimiz.*, vol. 29, no. 2, p. 264, 1991.

[11] J. A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems.* New York: Springer-Verlag, 1985.

[12] J.-L. Lagrange, *Mécanique Analytique*, 2 vols. ed. Paris: Albert Blanchard, 1788.

[13] M. I. Friedlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems.* New York: Springer-Verlag, 1984 (orig. Russian 1979).

[14] I. I. Gikhman, "Po Povodu Odnoi Teoremy N. I. Bogolyubova," *Ukran. Mat. Zh.*, vol. 4, no. 2, pp. 215–218, 1952.

[15] M. A. Krasnoselskii and S. G. Krein, "O Printsipe Usredneniya v Nelineinoi Mekhanike," *Uspekhi Mat. Nauk.*, vol. 10, no. 3, pp. 147–152, 1955.

[16] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes.* Cambridge, MA: MIT Press, 1984.

[17] H. J. Kushner and D. S. Clark, *Stochastic Approximation for Constrained and Unconstrained Systems.* New York: Springer-Verlag, 1978.

[18] A. Schwartz, Convergence of Stochastic Approximations: An Invariant Measure Approach, Ph.D. dissertation, Division of Engineering, Brown Univ., 1982.

[19] H. J. Kushner and H. Huang, "Rates of convergence for stochastic approximation type algorithms," *SIAM Journal Contr. Optimiz.*, vol. 17, pp. 607–617, 1979.

[20] R. A. Khas'minskii, "O Printsipe Usredneniya Dlya Stokhasticheskikh Differentsialnykh Uravnenii Ito," *Kybernetika*, vol. 4, no. 3, pp. 260–279, 1968.

[21] G. C. Papanicolaou, D. Strook, and S. R. S. Varadhan, "Marting Approach to Some Limit Theorems," in *Proc. 1986 Duke U Conf. Turbulence*, Apr. 1986.

[22] D. G. DeWolf, "An ordinary differential equation limit technic for convergence analysis of continuous-time parameter estimato Ph.D. dissertation, UCLA Dept. Elec. Eng., Sept. 1989.

[23] P. Baldi, "Limit set of inhomogeneous Ornstein-Uhlenbeck processes, destabilization, and annealing," *Stochastic Processes Their Appl.*, vol. 23, pp. 153–167, 1986.

[24] L. Ljung, "Asymptotic Behavior of the extended Kalman filter as a parameter estimator for linear systems," *IEEE Trans. Auto. Control*, AC-24, no. 1, pp. 36–50, 1979.

[25] A. H. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Academic, ch. 9, 1970.

[26] L. Ljung, *System Identification Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, p. 292, 1987.

[27] L. Gerenscér, I. Gyöngy, and Gy. Micaletzky, "A continuous-time recursive maximum likelihood method," in *Proc. IFAC Ninth World Congress*, vol. X, pp. 75–77, Budapest, 1984.

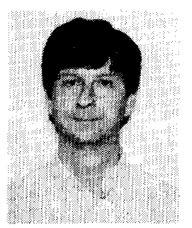
[28] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes.* New York: Springer-Verlag, 1977.

[29] P. Billingsley, *Convergence of Probability Measures.* New York: Wiley, 1968.

[30] D. Strook and S. R. S. Varadhan, *Multidimensional Diffusion Processes.* New York: Springer-Verlag, 1979.

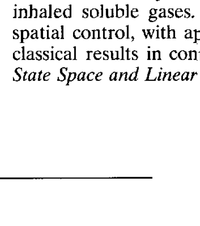
[13] A. C. M. Ran and L. Rodman, "On parameter dependence of solutions of algebraic Riccati equations," *Math. Contr. Sign. Syst.*, vol. 1, no. 3, pp. 269–284, 1988.

[32] D. G. DeWolf, "Continuous-time parameter estimation: Analysis via a limiting ordinary differential equation," *Control and Dynamic Systems*, vol. 38, C. T. Leondes, ed., San Diego: Academic, pp. 307–368, 1990.



Douglas DeWolf was born in Los Angeles, California on May 21, 1957. He received the B.A. and M.A. degrees in mathematics in 1979 and 1982, respectively, from the University of California, Los Angeles. He received his Ph.D. in electrical engineering in 1989.

From 1983 until 1989, he was a Howard Hughes Doctoral Fellow and is currently employed at Hughes Aircraft Company. His areas of interest include radar detection and tracking, Kalman and nonlinear filtering, numerical linear algebra, and adaptive signal processing (especially nonstationary signals).



Donald M. Wiberg received the B.S., M.S., and Ph.D. degrees in engineering from Caltech, Pasadena, CA, in 1959, 1960, and 1965, respectively.

He is presently a Professor in both Electrical Engineering and in Anesthesiology at University of California, Los Angeles. He has also consulted for Douglas Aircraft, R and D Associates, The Aerospace Corp., and Rockwell International. Previous research includes use of system identification for the noninvasive estimation of drug effect on the human respiratory control system. For healthy humans, he developed a noninvasive estimate of cardiac output using inhaled soluble gases. He originated the closed-loop mode method of spatial control, with application to nuclear reactors. He extended many classical results in control to state space, which appear in his textbook, *State Space and Linear Systems*.