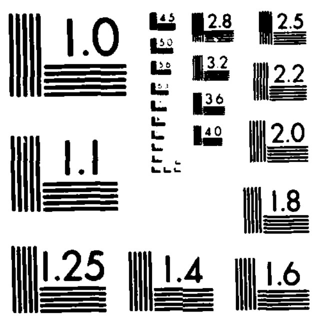
AD-A129 209 ASYMPTOTIC BEHAVIOR OF STOCHASTIC APPROXIMATION AND LARGE DEVIATIONS..(U) BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMICAL SYSTE.. H J KUSHNER

1/1

UNCLASSIFIED JAN 83 LCDS-83-1 AFOSR-TR-83-0483 F/G 12/1 NL

END
DATE
FILMED
7-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics

Brown University Providence RI 02912

ASYMPTOTIC BEHAVIOR OF STOCHASTIC APPROXIMATION and LARGE DEVIATIONS

by

Harold J. Kushner

January 1983 LCDS #83-1

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 83-0483	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ASYMPTOTIC BEHAVIOR OF STOCHASTIC APPROXIMATION AND LARGE DEVIATIONS	5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER: LCDS Report #83-1
7. AUTHOR(s): Harold J. Kushner	8. CONTRACT OR GRANT NUMBER(s): AFOSR-81-0116
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University, Providence RI 02912	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: PE61102F; 2304/A4
11. CONTROLLING OFFICE NAME AND ADDRESS: Mathematical & Information Sciences Directorate, Air Force Office of Scientific Research, Bolling AFB DC 20332	12. REPORT DATE: JAN 83
	13. NUMBER OF PAGES: 27
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

SEE REVERSE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ITEM #20, CONTINUED:

The theory of large deviations is applied to the study of the asymptotic properties of the stochastic approximation algorithms (1.1) and (1.2). The method provides a useful alternative to the currently used technique of obtaining rate of convergence results by studying the sequence $\{(X_n-\theta)/\sqrt{a_n}\}$ (for (1.1)), where θ is a 'stable' point of the algorithm. Let G be a bounded neighborhood of θ, which is in the domain of attraction of θ for the 'limit ODE'. The process $x^n(\cdot)$ is defined as a 'natural interpolation' of $\{X_j, j \geq n\}$ with $x^n(0) = X_n$, and interpolation intervals $\{a_j, j \geq n\}$. Define $\tau_G^n = \min\{t: x^n(t) \notin G\}$. Then it is shown (among other things) that $P_x\{\tau_G^n \leq T\} \sim \exp -n^q V$, where q depends on $\{a_n, c_n\}$, and V depends on the $b(\cdot)$, cov ξ_n, and G. Such estimates imply that the asymptotic behavior is much better than suggested by the 'local linearization methods', and they yield much new insight into the asymptotic behavior. The technique is applicable to related problems in the asymptotic analysis of recursive algorithms, and requires weaker conditions on the dynamics than do the 'linearization methods'. The necessary basic background is provided and the optimal control problems associated with getting the V above are derived.

ASYMPTOTIC BEHAVIOR OF STOCHASTIC APPROXIMATION and LARGE DEVIATIONS†

by

Harold J. Kushner

Divisions of Applied Mathematics and Engineering
Lefschetz Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912

January 1983

†Work supported in part by the Air Force Office of Scientific Research under AFOSR 81-0116, by the National Science Foundation under NSF-Eng. 77-12946-03 and in part by the Office of Naval Research under N00014-76-C-0279-P6.

Abstract

The theory of large deviations is applied to the study of the asymptotic properties of the stochastic approximation algorithms (1.1) and (1.2). The method provides a useful alternative to the currently used technique of obtaining rate of convergence results by studying the sequence $\{(X_n-\theta)/\sqrt{a_n}\}$ (for (1.1)), where θ is a 'stable' point of the algorithm. Let G be a bounded neighborhood of θ, which is in the domain of attraction of θ for the 'limit ODE'. The process $x^n(\cdot)$ is defined as a 'natural interpolation' of $\{X_j, j \geq n\}$ with $x^n(0) = X_n$, and interpolation intervals $\{a_j, j \geq n\}$. Define $\tau_G^n = \min\{t : x^n(t) \notin G\}$. Then it is shown (among other things) that $P_x\{\tau_G^n \leq T\} \sim \exp -n^q V$, where q depends on $\{a_n, c_n\}$, and V depends on the $b(\cdot)$, cov ξ_n, and G. Such estimates imply that the asymptotic behavior is much better than suggested by the 'local linearization methods', and they yield much new insight into the asymptotic behavior. The technique is applicable to related problems in the asymptotic analysis of recursive algorithms, and requires weaker conditions on the dynamics than do the 'linearization methods'. The necessary basic background is provided and the optimal control problems associated with getting the V above are derived.

1. Introduction

The paper deals with a useful and heretofore unexploited approach to the asymptotic behavior of stochastic approximation (SA) like algorithms of the form

$$X_{n+1} = X_n + a_n b(X_n) + a_n \xi_n \quad , \quad a_n = (n+1)^{-\rho} , \tag{1.1}$$

or of the 'Kiefer-Wolfowitz' form

$$X_{n+1} = X_n + a_n b(X_n) + a_n \xi_n / c_n \quad , \quad a_n = (n+1)^{-\rho} \quad , \quad c_n = (n+1)^{-\gamma}, X_n \in R^r, \tag{1.2}$$

where $0 < \rho \leq 1$ and $0 < \gamma < \rho/2$. To avoid excess notation, let $\{\xi_n\}$ be mutually independent and identically distributed. The noise sequence $\{\xi_n\}$ is mean zero and Gaussian, with covariance matrix $R \geq 0$. As seen below, it is hard to do some of the required calculations in the non-Gaussian case, although the basic theory is much more widely applicable. Despite the restriction to the Gaussian case, the results shed considerable new light on the asymptotic behavior. One would expect that the order of the obtained estimates would hold under much weaker conditions. Of particular interest are estimates (as a function of n) of the probability that the 'tail' of the SA sequence $\{X_m, j \geq n\}$ escapes from a neighborhood of a 'stable' point of the algorithm. By a 'stable point' we mean a point θ at which $\dot{x} = b(x)$ is asymptotically stable. Under our conditions, if X_n is in a small neighborhood of θ

often enough, then it converges to θ w.p.1. We are not interested in the w.p.1. convergence, only in the 'rate of convergence' or in the behavior of $\{X_n\}$ in a neighborhood of θ So we simply assume that $X_n \to \theta$ w.p.1. The estimates in the sequel imply that the asymptotic behavior is much better than one would expect from using the usual limit theory, which is based on the asymptotic normality of the sequence of suitably normalized errors (say of $(X_n-\theta)/\sqrt{a_n}$) for (1.1). The classical theory is much more 'local' about θ, and does not exploit as fully as possible the stabilizing properties of the ODE $\dot{x} = b(x)$ in a neighborhood of θ.

An additional advantage of our approach is that $b(\cdot)$ is not required to have continuous derivatives, as the classical theory requires. It need only be Lipschitz continuous. Thus one can treat problems where (e.g.) $b(\cdot)$ is obtained from a min-max operation, or where (scalar case) the slope of $b(\cdot)$ is discontinuous at θ. E.g. $b(x) = -k_1(x-\theta)$ for $x > \theta$ and $b(x) = -k_2(x-\theta)$ for $x < \theta$, where $k_1 \neq k_2$, and $k_i > 0$.

Results of simulations support the idea that the iterates spend (asymptotically) almost all the time on the part with the smaller slope, and this behavior is implied by our results. Also, simple constraints can readily be handled. For example let $\{X_n\}$ be confined to $[a,b]$, where $-\infty < a < b < \infty$, and $b(x) > 0$ on $[a,b]$. Then $X_n \to b$, and we can obtain estimates of the behavior of the sequence near b (e.g., probability of escape from a small neighborhood of b). This cannot be done with the classical rate of convergence theory for SA's.

The particular problem of interest will now be described. Let G denote a bounded open set which is in the domain of attraction of θ (for $\dot{x} = b(x)$) and whose boundary is piecewise differentiable. Roughly, we are interested in estimates of the type $P\{X_{n+m} \notin G, \text{ some } m > 1 \mid X_n \in \text{neighborhood of } \theta\}$ and we now make this precise. Define $t_n = \sum_0^{n-1} a_i$ and $m(t) = \max\{n: t_n \leq t\}$. Then $m(t_n) = n$ and $\sum_n^{m(t_n+t)} a_i/t \to 1$ for each $t > 0$. Both t_n and $m(t)$ depend on ρ. For each n, define the process $x^n(\cdot)$ on $[0,\infty)$ as follows. It is piecewise linear, with initial condition $X_n = x^n(0)$ and break points $\{0, t_{n+1}-t_n, t_{n+2}-t_n, \ldots\} = \{0, a_n, a_{n+1}+a_n, \ldots\}$, and $X^n(t_m-t_n) = X_m$. Thus $x^n(\cdot)$ 'starts' at the n^{th} iteration. Such an interpolation has been very useful in the analysis of the asymptotic properties of $\{X_n\}$, and is the key to the so-called 'ODE method' [1],[2]. Define $\tau_G^n = \min\{t: x^n(t) \notin G\}$. If $X_n \to 0$ w.p.1 (or even 'weakly'), then $E\tau_G^n$ is not necessarily defined. But $P_x\{\tau_G^n \leq T\}$ is of considerable interest as a criterion of performance and stability of the algorithm, where T is any positive number. Here P_x denotes the probability, conditioned on the event that $X_n = x \in G$. The dependence on ρ and γ and the structure of $b(\cdot)$ is of particular interest.

Since the probability $P_x\{\tau_G^n \leq T\}$ tends to zero as $n \to \infty$, it is natural to look for a normalizing sequence. In particular, we seek a sequence $\lambda_n \to 0$ such that the limit in (1.3) exists, where $0 < V < \infty$.

$$\lim_n \lambda_n \log P_x\{\tau_G^n \leq T\} = -V \quad . \tag{1.3}$$

Under quite broad conditions, (1.3) is continuous in x in a neighborhood of θ.

Let $C_x[0,T]$ denote the space of R^r valued continuous functions on $[0,T]$, with initial value x, and with the topology of uniform convergence. Let $A \subset C_x[0,T]$. Then estimates for $\lim_n \lambda_n \log P_x\{x^n(\cdot) \in A\}$ are also provided. We restrict attention to the Gaussian case, since it is hard to obtain the proper normalizing sequences $\{\lambda_n\}$ in general, and the Gaussian case is quite interesting in itself. (The results in the sequel also indicate what is needed in the more general cases.) Very similiar reasons require the use of the 'small white noise model' in singular perturbation studies. But, despite this restriction, singular perturbation theory has achieved some significant results [3],[4]. Results on the robustness of the estimates with respect to the noise statistics appear in [11].

Estimates such as (1.3) cannot be obtained from the classical rate of convergence theory for SA's. In order to put our results in perspective, some of the classical theory is outlined briefly in Section 2. The theory of large deviations is the appropriate vehicle for getting (1.3). The necessary background is provided in Section 3. Our results involve a modification of a basic theorem of Freidlin (Theorem 2.1 in [5]), and in Section 3, his result is stated, together with a rough idea of the proof, in order to facilitate its modification for our needs. In Section 4, the basic large deviations theorem for SA's is stated, as are the modifications to Freidlins proof which are needed to get the extensions for our cases. In Section 5, the basic theorem is specialized to the 'escape time problem', and the $\{\lambda_n\}$ are calculated in Section 6. The V_0 are obtained from the solution to a variational problem, and this is discussed in Sections 4 and 5.

The basic result is that $\lambda_n = 0(a_n)$ for (1.1) and $\lambda_n = 0(a_n/c_n^2)$ for (1.2). Also (for x near θ)

$$P_x\{\tau_G^n \le T\} \sim \exp - V_\rho n^\rho \qquad \text{(for (1.1))}$$

(1.4)

$$P_x\{\tau_G^n \le T\} \sim \exp - V_\rho n^{\rho-2\gamma} . \qquad \text{(for (1.2))}$$

The V_ρ is constant for $\rho \in (0,1)$, and their values appear in Section 6. The estimates (1.4) imply that the asymptotic behavior is much better than one would expect from the classical rate of convergence theory. Solving for the V_ρ involves solving a variational or optimal control problem, as will be seen. But, the qualitative results such as (1.4) are of interest even if the exact values of the V_ρ are not known.

The theory of large deviations is of considerable potential use in the study of the asymptotic behavior of recursive algorithms. It is of potential use, where one wants to avoid the 'local linearization' methods otherwise used to study the asymptotic behavior, or to take greater advantage of the stability of the 'limit ODE'. Also, see [6],[7] where it is used to obtain estimates of the probability of breakdown of an ALOAH type communications network. The application of the Theory of large deviations to the SA problem involves some new considerations. The norming sequences are not standard in the large deviations literature, and the 'Lagrangians' $L(x,\beta,s)$ can depend on time here. The distinct differences between the cases $\rho = 1$ and $\rho < 1$ are not at all obvious.

2. Classical Rates of Convergence for (1.1)

In order to put the results of this paper into perspective with the other main method of studying the behavior of $\{X_n\}$ near θ, some classical results are reviewed here. Our attention is confined to (1.1).

Let $X_n \to \theta$ w.p.1 and define $U_n = (n+1)^{\rho/2}(X_n - \theta)$, and let $b(\cdot)$ be continuously differentiable, with $b(\theta) = 0$. Drop the i.i.d. assumption on $\{\xi_n\}$, but let it be stationary and define $R = \sum_{-\infty}^{\infty} E\,\xi_j \xi_0'$, where the sum is assumed to be absolutely convergent. Then for (1.1),

$$U_{n+1} = [I + a_n(b_x(\theta) + \frac{\rho}{2(n+1)^{1-\rho}}) + 0(1/n)]U_n \tag{2.1}$$

$$+ (n+1)^{-\rho/2}\xi_n + 0(1/n)\xi_n .$$

Define $U^n(\cdot)$ as $x^n(\cdot)$ was defined, but using $\{U_j, j \geq n\}$ instead of $\{X_j, j \geq n\}$. For $\rho = 1$ ($\rho < 1$, resp.) let $I/2 + b_x(\theta)$ ($b_x(\theta)$, resp.) have its eigenvalues in the open left half plane. (The matrix is then said to be <u>stable</u>.) Then, under quite broad conditions [8], $\{U^n(\cdot)\}$ converges weakly to the stationary solution of the Itô equations

$$dU = (I/2 + b_x(\theta))Udt + R^{1/2}dw, \quad \rho = 1, \tag{2.2a}$$

$$dU = b_x(\theta)Udt + R^{1/2}dw, \quad \rho < 1, \tag{2.2b}$$

where $w(\cdot)$ is a standard Wiener process.

In particular, the sequence $\{U_n\}$ converges in distribution to the stationary random variable U^ρ of (2.2), where $U^\rho \sim N(0,\Sigma_\rho)$ and

$$\Sigma_1 = \int_0^\infty [\exp t(I/2 + b_x(\theta))]\ R\ [\exp t(I/2 + b_x'(\theta))]dt,$$

$$\Sigma_\rho = \int_0^\infty [\exp tb_x(\theta)]\ R \exp tb_x'(\theta)dt \quad , \quad \rho < 1.$$

Note the differences between the cases $\rho = 1$ and $\rho < 1$. In particular, the more stringent stability requirement on $b_x(\theta)$, when $\rho = 1$. The limit (1.3) holds only under stability of $\dot{x} = b(x)$, so the more stringent requirement on $b_x(\theta)$ is not needed. In fact, (1.3) can be obtained even if $b_x(\theta)$ has a zero eigenvalue, provided that $\dot{x} = b(x)$ is stable at θ.

An analysis of (2.2) can provide much useful information as the asymptotic behavior of $\{X_n\}$. But it cannot help us with the large deviation estimate (1.3), where the set G is fixed. This is partly because $(X_n-\theta) \sim (n+1)^{-\rho/2}U_n$, which goes to zero in probability as $n \to \infty$. Also the validity of (2.2) requires continuity of $b_x(\cdot)$ at $x = \theta$. Eqn (2.2) also gives us a somewhat more pessimistic idea of the asymptotic behavior than (1.3) does.

3. The Theory of Large Deviations

As mentioned in the previous section, 'central limit' type ideas cannot be used to obtain estimates such as (1.3). The theory of large deviations is set up for just this purpose. It has proven to be a rather powerful tool for handling realted problems in probability and statistics [9]. Our basic background ideas come from Freidlin [5], although they must be modified to suit our needs. Freidlin obtains large deviations estimates related to (1.3) for the system $\dot{x}^{\varepsilon} = b(x^{\varepsilon}, \xi(t/\varepsilon))$, $x^{\varepsilon} \in R^r$, where $b(\cdot,\xi)$ is uniformly Lipschitz and bounded and $\xi(\cdot)$ is a bounded stochastic process. We start by recapitulating the main ideas, and then adjusting them to suit our needs.

Suppose that there is a function $H(\cdot,\cdot)$ such that for each x and piecewise constant function $\alpha(\cdot)$, the limit in (3.1) exists.

$$\int_0^T H(x,\alpha(u))du = \lim_{\varepsilon} \varepsilon \log E \exp \int_0^{T/\varepsilon} \alpha'(\varepsilon u) b(x,\xi(u))du. \tag{3.1}$$

(An example will be given before the lemma below). Define the dual functional (called the Cramer or Legendre transform)

$$L(x,\beta) = \sup_{\beta} [\alpha'\beta - H(x,\alpha)].$$

For $\phi(\cdot)$ absolutely continuous, define $S(T,\phi)$ by

$$S(T,\phi) = \int_0^T L(\phi(u),\dot{\phi}(u))du,$$

and set $S(T,\phi)$ equal to ∞ if $\phi(\cdot)$ is not absolutely continuous.

Let $A \subset C_x[0,T]$ and let A^o and $\bar{A}$ denote the interior and closure of A, resp.

Assume

(A3.1) $H(\cdot,\cdot)$ is continuous and $H(x,\cdot)$ is continuously differentiable for each x (this will be true for our problem).

Let $P_x\{\ \}$ denote expectation conditioned on $x^\varepsilon(0) = x$. Then by [5], Theorem 2.1, we have the large deviations estimate (3.2).

$$(3.2)\qquad -\inf_{\varphi\in A^\circ} S(T,\varphi) \le \underline{\lim}_{\varepsilon}\, \varepsilon \log P_x\{x^\varepsilon(\cdot)\in A\} \le \overline{\lim}_{\varepsilon}\, \varepsilon \log P_x\{x^\varepsilon(\cdot)\in A\}$$
$$\le -\inf_{\varphi\in \bar{A}} S(T,\varphi).$$

Thus, obtaining the estimates requires solving a variational problem.

For the SA problems of interest, a sequence $\lambda_n \to 0$ replaces $\varepsilon \to 0$. Also $L(x,\beta)$ can be written explicitly, and the variational problem is equivalent to an optional control problem (see Section 5).

Example. Let $b(x,\xi) = b(x) + \xi$, where $\xi(\cdot)$ is mean zero, stationary and Gaussian with an integrable correlation function. Define $\bar{R} = \int_{-\infty}^{\infty} E\xi(u)\xi'(0)\,du$. If Ψ is scalar valued, Gaussian and $E\Psi = 0$, then $E\exp\Psi = \exp E\Psi^2/2$. Let $\alpha(\cdot)$ be piecewise constant on $[0,T]$. Then

$$\int_0^T H(x,\alpha(u))du = \frac{1}{2}\int_0^T \alpha'(u)\bar{R}\alpha(u)du + \int_0^T \alpha'(u)b(x)du.$$

Thus $H(x,\alpha) = \alpha'\bar{R}\alpha/2 + \alpha' b(x)$.

Freidlin's proof can be modified to suit our needs. Since his proof is not short, but the modifications few, we only indicate the required modifications. This will be done in the next section. But, to get a better idea of what is needed, we backtrack and briefly discuss Freidlin's technique of proof. He used the following result (of Gartner [10], Lemmas 1.1 and 1.2) concerning large deviations estimates for a sequence of random vectors, in order to first obtain a 'finite dimensional' form of (3.2). Then via a sequence of bounds and approximations, he takes the 'finite dimensional' result into (3.2). Let $d(x,y)$ denote either the Euclidean distance, or the norm $\sup_{t\leq T}|x(t)-y(t)|$ if x and y are functions.

Lemma 1. ([10], Lemma 1.1 and 1.2) Let $\{\eta^\varepsilon\}$ denote a sequence of R^k-valued random vectors and let there be a sequence of positive numbers $\delta_\varepsilon \to 0$ such that the limit $H_0(\alpha)$ exists for each $\alpha \in R^k$:

$$H_0(\alpha) = \lim_{\varepsilon\to 0} \delta_\varepsilon \log E \exp \alpha'\eta^\varepsilon/\delta_\varepsilon$$

Let $H_0(\cdot)$ be continuously differentiable. Define the dual function $L_0(\beta) = \sup_\alpha[\alpha'\beta - H_0(\alpha)]$. Define $\Phi_0(s) = \{\beta : L_0(\beta) \leq s\}$. Then for each vector β, and each $s \geq 0$, $h > 0$ and $c > 0$, there is an $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$.

(3.3a) $$\delta_\varepsilon \log P\{d(\eta^\varepsilon,\Phi_0(s)) > c\} \leq -(s-h)$$

(3.3b) $$\delta_\varepsilon \log P\{d(\eta^\varepsilon,\beta) < c\} \geq -(L_0(\beta)+h)\ .$$

Let $\beta \in R^k$. Then from (3.3), we readily obtain (3.4)(which is the finite dimensional version of (3.2))

$$(3.4)\qquad -\inf_{\beta\in B^0} L_0(\beta) \le \underline{\lim}_{\varepsilon}\ \delta_\varepsilon \log P\{\eta^\varepsilon \in B\} \le \overline{\lim}_{\varepsilon}\ \delta_\varepsilon \log P\{\eta^\varepsilon \in B\} \le -\inf_{\beta\in \bar{B}} L_0(\beta).$$

The derivation of (3.4) from (3.3) is quite straightforward and goes roughly as follows. Let $\beta \in B^0$, and define $N_c(\beta)$ = c-neighborhood of β. Choose c such that $N_c(\beta) \in B^0$. Fix small $h > 0$. Then by (3.3b)

$$\delta_\varepsilon \log P\{\eta^\varepsilon \in B\} \ge \delta_\varepsilon \log P\{d(\eta^\varepsilon, \beta) < c\} \ge -(L_0(\beta) + h/2),$$

for small ε. Now choose c and β such that the right side is within h of $-\inf_{\beta\in B} L_0(\beta)$. Owing to the arbitrariness of h, this yields the left side of (3.4). Next, for any s such that the (compact) set $\Phi_0(s)$ is disjoint (distance $> c > 0$) from $\bar{B}$, we have

$$\delta_\varepsilon \log P\{\eta^\varepsilon \in \bar{B}\} \le \delta_\varepsilon \log P\{d(\eta^\varepsilon, \Phi_0(s)) > c\}.$$

Now use the (3.3a) and the largest possible $\Phi_0(s)$ (this requires that $s \le \inf_{\beta\in\bar{B}} L_0(\beta)$). The details of obtaining (3.4) from (3.3) are readily completed.

A rough outline of the argument of Freidlins Theorem 2.1 [5].

Now that we have the basic lemma used by Freidlin, we comment on his derivation of (3.2). In the next section the proof is extended to cover the SA case. Starting with the above Lemma 1, Freidlin proved (3.2) by an argument along the following lines. Fix x and $\Delta > 0$ and let $N = T/\Delta$ be an integer. Let $\psi(\cdot)$ denote a function that is constant on each interval $[i\Delta, i\Delta + \Delta)$. Define the function $x^{\psi,\varepsilon}(\cdot)$ by

$$x^{\psi,\varepsilon}(t) = x + \int_0^t b(\psi(s),\xi(s/\varepsilon))ds, \quad x^{\psi,\varepsilon}(t) \in R^r. \tag{3.5}$$

Let $\phi(\cdot)$ denote a continuous function and let ϕ_Δ denote the vector $\{\phi(i\Delta),\ i < N\}$. Define the vector $x_\Delta^{\psi,\varepsilon} = \{x^{\psi,\varepsilon}(i\Delta),\ i < N\}$. Define the functional and set, resp.,

$$S^\psi(T,\phi) = \int_0^T L(\psi(s),\dot{\phi}(s))ds \quad , \quad \Phi^\psi(s) = \{\phi(\cdot) : \phi(0) = x,\ S^\psi(T,\phi) \le s\}.$$

Now, using the fact that the limit in (3.1) exists, Lemma 1 can be applied to the vectors $\eta^\varepsilon = x_\Delta^{\psi,\varepsilon}$, $\beta = \phi_\Delta$ and with $\delta_\varepsilon = \varepsilon$. To see this and to see how to obtain the $H_0(\cdot)$ and $L_0(\cdot)$ used in Lemma 1, for a set of r-vectors $\{\bar{\alpha}_i\}$, let $\alpha(\cdot)$ in (3.1) take value $(\bar{\alpha}_0 + \ldots + \bar{\alpha}_{N-1}) = \alpha_0$ on $[0,\Delta)$, $(\bar{\alpha}_1 + \ldots + \bar{\alpha}_{N-1}) = \alpha_1$ on $[\Delta,2\Delta)$, ... , and $\bar{\alpha}_{N-1} = \alpha_{N-1}$ on $[N\Delta-\Delta,T]$. Define the vector $\bar{\alpha}$ by $\bar{\alpha} = (\bar{\alpha}_0, \ldots, \bar{\alpha}_{N-1})$. Then

$$\bar{\alpha}' x_\Delta^{\psi,\varepsilon} = \sum_{i=0}^{N-1} \bar{\alpha}_i' x^{\psi,\varepsilon}(i\Delta) =$$

$$\alpha_0' \int_0^\Delta b(\psi(u), \xi(u/\varepsilon))du + \ldots + \alpha_{N-1}' \int_{N\Delta-\Delta}^T b(\psi(u), \xi(u/\varepsilon))du.$$

Thus, since the limit (3.1) exists, $H_0(\bar{\alpha})$ is well defined for each $\bar{\alpha}$, and so is $L_0(\beta)$.

Applying Lemma 1 in this way yields a large deviation estimate of the type (3.3) for the 'samples' of $x^{\psi,\varepsilon}(\cdot)$ and $\phi(\cdot)$, with sampling interval Δ.

Via a sequence of approximations based on this Δ-approximation, Freidlin proves the analog (3.6) of (3.3) for the sequence† $\{x^{\psi,\varepsilon}(\cdot), \varepsilon > 0\}$; namely that for each fixed $\psi(\cdot)$ and each $\phi(\cdot)$, $s \geq 0$, $h > 0$ and $c > 0$, there is an $\varepsilon_0 > 0$ such that for $\varepsilon \leq \varepsilon_0$,

$$\varepsilon \log P\{d(x^{\psi,\varepsilon}, \Phi^\psi(s)) > c\} \leq -(s-h).$$

(3.6)

$$\varepsilon \log P\{d(x^{\psi,\varepsilon}, \phi) < c\} \geq -(S^\psi(T,\phi) + h)$$

Inequality (3.2) (with $S^\psi(T,\phi)$ replacing $S(T,\phi)$) for $\{x^{\psi,\varepsilon}(\cdot)\}$ follows from (3.6), just as (3.4) followed from (3.3) [5, Lemma 3.1]. The sequence of approximations alluded to above use the fact that the behavior of

---

† For the identification of our terms with those in [5], our $(x^{\psi,\varepsilon}, \Phi^\psi, \varepsilon, S^\psi(T,\phi), d, s)$ replace $(\tilde{x}^{\varepsilon,\psi}, \Phi_\psi, \varepsilon, S_{0,T}^\psi(\phi), \rho, s)$ in [5].

$\Phi(\cdot)$ and $x^{\psi,\varepsilon}(\cdot)$ between the Δ-sample points is 'regular' enough so that if the large deviations estimates hold for the samples for small enough $\Delta > 0$, then (3.6) holds. These approximations depend heavily on the Lipschitz continuity and boundedness of $b(\cdot,\xi)$ in order to show that the path excursions between the $i\Delta$-sampling times can be made as small as desired by making Δ small enough. Freidlin then proved (3.2) by using (3.6) and a sequence of approximations with suitably chosen $\psi(\cdot)$ and $\Phi(\cdot)$. These approximations also use the boundedness of $b(\cdot,\cdot)$ and the Lipschitz condition to show that the excursions of $x^{\varepsilon}(\cdot)$ between the $i\Delta$-sampling points can be made (uniformly) as small as desired, by making Δ small enough. We use these comments in the next section. Next, we obtain an estimate which will be needed to extend the result to the SA case.

<u>A bound on the sample excursions and sums of the noise terms for</u> (1.1) <u>and</u> (1.2). Since ξ_n are not bounded in the SA case, we need an estimate of the excursions of the paths of (the SA interpolations) $x^n(\cdot)$ between the $i\Delta$-sampling points, when $x^n(t) \in G$ for $t \in [0,T]$. This is provided by the following theorem.

<u>Theorem 2</u>. <u>Let</u> $\{\xi_n\}$ <u>be mutually independent, mean zero and Gaussian with</u> $\operatorname{var} \xi_n \leq \sigma^2 < \infty$. <u>Define</u> $\lambda_n = \sum_{n}^{m(t_n+T)} a_j^2$. <u>For each</u> $a > 0$ <u>and</u> $M < \infty$ <u>there is a</u> $\Delta_0 > 0$ <u>such that for</u> $\Delta < \Delta_0$

$$(3.7)\qquad P_\Delta^n \equiv P\{\sup_{i\Delta\le T}\ \sup_{s<\Delta} \left|\sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+s)} a_j\xi_j\right| \ge a\} \le \exp -M/\lambda_n$$

<u>If</u> a_j <u>is replaced by</u> a_j/c_j and $\lambda_n = \sum_n^{m(t_n+T)} a_j^2/c_j^2$, <u>then</u> (3.7) <u>still holds if</u> $\rho - 2\gamma > 0$.

<u>Proof</u>. We do only the first case. The second is treated in the same way. Suppose that the ξ_j are scalar valued; otherwise work with one component at a time. For any $\gamma > 0$, Chebychev's inequality and the Gaussian property yields (using $E \exp \alpha\xi_j \le \exp \alpha^2\sigma^2/2$)

$$P\{\sup_{s\le\Delta} \sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+s)} a_j\xi_j \ge a\} = P\{\sup_{s\le\Delta}\left[\exp \gamma \sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+s)} a_j\xi_j\right] \ge \exp \gamma a\}$$

$$\le (\exp -\gamma a)\exp\left[\gamma^2\sigma^2 \sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)} a_j^2/2\right]$$

$$\le \exp - \left[\frac{2\sigma^2}{a^2} \sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)} a_j^2\right]^{-1},$$

where the last inequality is obtained by minimizing the next to last term over $\gamma > 0$. Repeating for $-\xi_j$ replacing ξ_j, we get

$$(3.8)\qquad P_\Delta^n \le 2\sum_{i\Delta<T} \exp - C/(\lambda_n C_{in}), \quad C = a^2/2\sigma^2,$$

$$C_{in} = \sum_{m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)} a_j^2/\lambda_n .$$

Now, letting Δ be small enough yields the Theorem, since $\lim_n \sup_{i\le T/\Delta} C_{in} = 0$.

Q.E.D.

4. Large Deviations Estimates for (1.1) and (1.2).

The key to the extension of Freidlins theorem to the SA case is in getting the proper norming sequence $\{\lambda_n\}$ and the proper analog of the H-functional introduced in (3.1). Our guide is the method of proof, via Lemma 1.

The form of $H(\cdot,\cdot)$ and $\{\lambda_n\}$ for the systems (1.1) and (1.2). Let $\alpha(\cdot)$ and $\psi(\cdot)$ denote R^r-valued functions which are constant on the intervals $[i\Delta, i\Delta+\Delta)$, and where $|\psi(t)|$ is bounded by $\sup_{x\in G}|x|$. For each $x \in G$ and each n define the piecewise linear function $x^{\psi,n}(\cdot)$ as follows. The break points are $\{0, t_{n+1}-t_n, t_{n+2}-t_n, \ldots\}$ and for $j \geq n$, set

$$(4.1) \qquad x^{\psi,n}(t_k) = x + \sum_{n}^{k-1} a_j b(\psi(t_j - t_n)) + \sum_{n}^{j-1} a_j \xi_j \quad ,$$

for the case (1.1). For the case (1.2), replace ξ_j by ξ_j/c_j. The $x^{\psi,n}(\cdot)$ replace the $x^{\psi,\varepsilon}(\cdot)$ of (3.5), and they have the same interpolation intervals as do the $x^n(\cdot)$. Until further notice, work with case (1.1).

A natural analog of the H-functional of (3.1) is that defined by the limit in (4.2) (if it exists, for a suitable normalizing sequence $\{\lambda_n\}$). Recall that $T = N\Delta$.

$$(4.2) \qquad \int_0^T H(x,\alpha(s),s)\,ds = \lim_n \lambda_n \log E \exp \sum_{i=0}^{N-1} \alpha'(i\Delta) \sum_{j=m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)-1} a_j (b(x)+\xi_j)/\lambda_n .$$

Since

$$(4.3) \qquad \lambda_n \log E \exp \sum_{i=0}^{N-1} \alpha'(i\Delta) \sum_{j=m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)-1} a_j \xi_j$$

$$= \frac{\lambda_n}{2} \sum_{i=0}^{N-1} \alpha'(i\Delta) R \alpha(i\Delta) \sum_{j=m(t_n+i\Delta)}^{m(t_n+i\Delta+\Delta)-1} a_j^2/\lambda_n^2 ,$$

a natural candidate for λ_n (and the one we use) is

$$(4.4) \qquad \lambda_n = \sum_{n}^{m(t_n+T)} a_j^2 .$$

By the same reasoning, for case (1.2), the natural candidate for λ_n is

$$(4.4') \qquad \lambda_n = \sum_{n}^{m(t_n+T)} a_j^2/c_j^2$$

In order to get the correct form of $H(x,\alpha,s)$, we need to check the limit of (4.3) as $n \to \infty$. If there is a function $h(\cdot)$ such that

$$(4.5) \qquad (4.3) \to \int_0^T \frac{\alpha'(s)R\alpha(s)}{2} h(s)ds, \quad \text{as} \quad n \to \infty ,$$

then

$$(4.6) \qquad H(x,\alpha,s) = \alpha' R\alpha h(s)/2 + \alpha' b(x).$$

From the results of Section 6, we get that the limit exists and that $h(s) = 1/T$ when $\rho < 1$ and $h(s) = (1-e^{-T})^{-1} e^{-s}$ when $\rho = 1$. Note the peculiarity that the H-functional depends on s as well as on α and x, when $\rho = 1$. For (1.2), one gets $h(s) = 1/T$ when $\rho < 1$ and $h(s) = (1-2\gamma)\, e^{-s(1-2\gamma)}(1-e^{-T(1-2\gamma)})^{-1}$ when $\rho = 1$.

The next Theorem yields an analog to (3.2) for the SA case. The specialization to the escape time problem appears in the next section.

Theorem 3. If $b(\cdot)$ is Lipschitz continuous and $\{\xi_n\}$ is i.i.d. Gaussian with zero mean, then for each set $A \subset C_x[0,T]$,

(4.7)
$$-\inf_{\phi\in A^0} S(T,\phi) \le \underline{\lim}_n\, \lambda_n \log P_x\{x^n(\cdot)\in A\} \le \overline{\lim}_n\, \lambda_n \log P_x\{x^n(\cdot)\in A\}$$
$$\le -\inf_{\phi\in \bar{A}} S(T,\phi),$$

where

$$S(T,\phi) = \int_0^T L(\phi(s),\dot\phi(s),s)ds,$$

(4.8a)
$$L(x,\beta,s) = \sup_\alpha\, [\alpha'\beta - H(x,\alpha,s)].$$

In particular $H(\cdot,\cdot,\cdot)$ is given by (4.6) and if R is positive definite, then

(4.8b)
$$L(x,\beta,s) = h^{-1}(s)(\beta-\bar{b}(x))'R^{-1}(\beta-\bar{b}(x))/2 \equiv L_0(x,\beta)h^{-1}(s).$$

Later, we treat the degenerate case where $R = \begin{bmatrix} 0 & 0 \\ 0 & R_{22} \end{bmatrix}$, and where R_{22} is positive definite. For this case, define R^{-1} by $R^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & R_{22}^{-1} \end{bmatrix}$.

Comments on the proof. In Section 6, we show that the limits (as $n \to \infty$) of (4.3) exist as stated for the given $h(\cdot)$. We now discuss (4.7). Let $\phi(\cdot)$ be a continuous function. By the smoothness of $H(\cdot,\cdot,\cdot)$, Gartners result, Lemma 1, can be applied to the vectors $x_\Delta^{\psi,n} = \{x^{\psi,n}(i\Delta), i\Delta < T\}$, $\beta = \phi_\Delta = \{\phi(i\Delta), i < N\}$, just as it was applied to $x_\Delta^{\psi,\varepsilon}$ and ϕ_Δ below (3.5). From here on one follows Freidlin [5], Theorem 2.1, almost word for word. Only the differences will be noted here.

First, one proves the analog of (3.6) for λ_n replacing ε and $x^{\psi,n}(\cdot)$ replacing $x^{\psi,\varepsilon}(\cdot)$. Let $(\cdot)_F$ denote citations to Freidlin [5]. His proof uses auxiliary (Lemmas 3.1 and 3.2$)_F$. Inequality (3.6) is derived in (Lemma 3.1$)_F$. The proof of this lemma carries over, except for the inequality 3 lines below $(3.2)_F$ and the set inclusion 3 lines below $(3.5)_F$. But, by our Theorem 2, the inequality below $(3.2)_F$ can be replaced by the following (in our terminology[†]): <u>For each</u> $\eta > 0$, $c > 0$ and $M < \infty$, <u>there are</u> $\Delta_0 > 0$, $n_0 < \infty$ <u>and</u> $c_0 > 0$ <u>such that for</u> $\Delta \le \Delta_0$ <u>and</u> $n > n_0$,

$$P\{d(x^{\psi,n},\phi) < c\} \ge P\{d(x_\Delta^{\psi,n},\phi_\Delta) < c_0\}$$
$$- P\{\sup_{i\Delta<T, s<\Delta} |x^{\psi,n}(i\Delta+s)-x^{\psi,n}(i\Delta)| > \eta\},$$

<u>where the last probability on the r.h.s. is</u> $\le \exp -M/\lambda_n$. The proof uses the finiteness of $\sup_{x\in G}|b(x)|$ and Theorem 2. Similarly, the set inclusion

[†]Friedlins symbols $(\rho_{0,T}, \eta^{\varepsilon}, \delta, \delta', \tilde{\psi}^{\varepsilon,\psi}, \bar{\phi}^{\Delta}, \varepsilon)$ are replaced by our $(d, x_\Delta^{\psi,n}, c, c_0, x^{\psi,n}, \phi_\Delta, \lambda_n)$. Also his $\bar{\rho}$ is simply the Euclidean distance d.

below $(3.5)_F$ holds modulo a set whose probability is $\leq \exp -M/\lambda_n$ for small enough Δ, and large enough n. Lemma $(3.2)_F$ carries over with no change. The counterpart to our $x^{\psi,n}(\cdot)$ in Freidlins proof is his $\tilde{x}^{\varepsilon,\psi}(\cdot)$ (called $x^{\psi,\varepsilon}(\cdot)$ in (3.5).

Only one change is required in the main part of the proof of $(\text{Theorem } 2.1)_F$. In the paragraph below $(3.13)_F$, the fact that $\{\tilde{x}^{\varepsilon,\psi}(\cdot); \varepsilon > 0, t \leq T, |\psi(t)| \leq B = \sup_{x\in G} |x|\} = Q_0$ and $\{x^{\varepsilon}(\cdot); \varepsilon > 0, t \leq T\} = Q_1$ belong to a compact set in $C_x[0,T]$ is used. ($\tilde{x}^{\varepsilon,\psi}$ is the $x^{\psi,\varepsilon}$ in (3.5)). This compactness is used to guarantee that for each $\delta > 0$, there is an $N_\delta < \infty$ and functions $\phi_1,\ldots,\phi_{N_\delta}$ (not depending on ε or $\psi(\cdot)$) such that the union of the δ-neighborhoods of the ϕ_i cover $Q_0 \cup Q_1$. Our trajectories $\{x^{\psi,n}(\cdot), x^n(\cdot), t \leq T, \varepsilon > 0, |\psi(t)| \leq B\} = R_n$ do not belong to a compact set, since the Gaussian noise is unbounded. But, by Theorem 2 we obtain the following: For each $M < \infty$ and $\delta > 0$ there are $N_\delta < \infty$ and $\phi_1,\ldots,\phi_{N_\delta}$ such that the union of the δ-neighborhoods of the ϕ_i covers R_n except for a set of paths whose probability is $\leq k \exp -M/\lambda_n$, where k does not depend on n. Since number M in the estimates of the probabilities of the exceptional sets can be made arbitrarily large, we can carry through all the details (essentially as done in [5]) to obtain (4.7). The inequalities (4.7) are specialized to the escape time problem in the next section.

5. Escape Time Formulas

Let $A = \{\phi(\cdot):\phi(0) = x,\ \phi(t) \notin G,\ \text{some}\ t \le T\}$. Then by (4.7) and the fact that A is closed,

$$(5.1) \qquad -\inf_{\phi\in A^{o}} S(T,\phi) \le \underline{\lim_{n}}\ \lambda_n \log P_x\{\tau_G^n \le T\} \le$$

$$\overline{\lim_{n}}\ \lambda_n \log P_x\{\tau_G^n \le T\} \le -\inf_{\phi\in A} S(T,\phi).$$

In a sense, for 'almost all G', there is equality in (5.1). To see this, define $G_\delta = \{y: d(y,G) < \delta\}$, and set $A_\delta = \{\phi(\cdot): \phi(0) = x, \phi(t) \in \partial G_\delta \text{ for some } t \leq T\}$, and $A_0 = A$. Define $S_x(G_\delta) = \inf_{\phi \in A_\delta} S(T,\phi)$. As $\delta \downarrow 0$, $S_x(G_\delta)$ decreases and it is continuous at all $\delta \geq 0$ except for a countable number. Assume

(A5.1) $S_x(G_\delta) \downarrow S_x(G)$ <u>as</u> $\delta \downarrow 0$.

The condition always holds in the <u>non-degenerate</u> case described below.

<u>Theorem 4</u>. <u>Under</u> (A5.1) <u>and the given properties of</u> $\{\xi_n\}$ <u>and</u> $b(\cdot)$, <u>equality holds in</u> (5.1).

The proof follows from the facts that (a): $\lim_\delta S_x(G_\delta) = \inf_{\phi \in A^0} S(T,\phi)$, and (b): by (A5.1), $\lim_\delta S_x(G_\delta) = \inf_{\phi \in A} S(T,\phi)$.

Condition (A5.1) also implies that $S_x(G)$ is continuous at $x = \theta$, if it holds at $x = \theta$. Thus, it is not much of a restriction to assume equality in (5.1) and to let $\theta = x$, as we do henceforth. Now, consider the variational problem of getting the $\inf_{\phi \in A} S(T,\phi)$. If $R > 0$, we say that the problem is <u>nondegenerate</u>. If R is <u>singular</u>, suppose for convenience that R takes the form $R = \begin{bmatrix} 0 & 0 \\ 0 & R_{22} \end{bmatrix}$, where $R_{22} > 0$ and partition the vectors as follows: $x = (x_1, x_2)$, $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$, $\phi = (\phi_1, \phi_2)$, where x_2, α_2, β_2 and ϕ_2 have the dimension of R_{22}. As noted below, in the non-degenerate case, we have

$$L(x,\beta,s) = T(\beta-b(x))'R^{-1}(\beta-b(x))/2 = L_0(x,\beta)T, \quad \rho < 1,$$

(5.2)

$$L(x,\beta,s) = (1-e^{-T})(\beta-b(x))'R^{-1}(\beta-b(x))e^{s}/2$$
$$= L_0(x,\beta)\, e^{s}(1-e^{-T}) \quad , \rho = 1.$$

Define $u = R^{-1/2}(\beta-b(x))$. Then the variational problem of calculating $\inf_{\phi \in A} S(T,\phi)$ is equivalent to minimizing

(5.3) $$S(T,\phi) = \frac{T}{2}\int_0^T u'u\, ds \quad \text{or} \quad \left(\frac{1-e^{-T}}{2}\right)\int_0^T e^{s}u'u\, ds,$$

(depending on whether $\rho < 1$ or $\rho = 1$) subject to

(5.4) $$\dot\phi = b(\phi) + R^{1/2}u, \quad \phi(0) = x, \quad \phi(t) \in \partial G \text{ for some } t \le T.$$

In the degenerate case, the variational problem is (5.3'), (5.4').

(5.3') $$S(T,\phi) = \frac{T}{2}\int_0^T u_2'u_2 ds \quad \text{or} \quad \left(\frac{1-e^{-T}}{2}\right)\int_0^T e^{s}u_2'u_2 ds,$$

subject to

(5.4') $$\dot\phi_1 = b_1(\phi)$$
$$\dot\phi_2 = b_2(\phi) + R_{22}^{1/2}u_2, \quad \phi(0) = x \quad , \quad \phi(t) \in \partial G \quad \text{for some} \quad t \le T.$$

6. The Values of λ_n And $H(\cdot,\cdot,\cdot)$.

In this section, we evaluate λ_n and obtain the $h(\cdot)$ in (4.6) and and the V of (1.3). Until further notice, we work with case (1.1). By the discussion leading to (4.6), to obtain $H(\cdot,\cdot,\cdot)$ we need only find $h(\cdot)$ such that

(6.1) $$\sum_{m(t_n+t)}^{m(t_n+t+\Delta)} a_j^2/\lambda_n \to h(t)\Delta + o(\Delta).$$

We use integrals in place of sums henceforth, since the ratios of the integrals to the sums converge to unity in all cases, as $n \to \infty$. Also $\alpha_n \sim \beta_n$ means that $\alpha_n/\beta_n \to 1$ as $n \to \infty$. By the definition of $m(t_n + t)$, for $t \geq 0$,

$$\left|t - \sum_{n}^{m(t_n+t)-1} a_j\right| \leq a_n.$$

6.1 The case $\rho = 1$ and (1.1). By definition, $m(t_n + t)$ and λ_n satisfy

$$T \sim \int_n^{m(t_n+T)} s^{-1}\, ds \quad , \quad \lambda_n \sim \int_n^{m(t_n+T)} s^{-2}\, ds .$$

Thus $m(t_n+T) \sim ne^T$ and

(6.2) $$\lambda_n \sim n^{-1}(1-e^{-T}).$$

Now, in order to evaluate (6.1). We need only evaluate (6.3) for $\rho = 1$

(6.3) $$\int_{m(t_n+t)}^{m(t_n+t+\Delta)} s^{-2\rho}/\lambda_n = h(t)\Delta + o(\Delta)$$

Since this equals $(1-e^{-T})^{-1} e^{-t}(1-e^{-\Delta})$, we have $h(t) = (1-e^{-T})^{-1} e^{-t}$.

6.2 The case $\rho < 1$ and (1.1). Here

$$T \sim \int_n^{m(t_n+T)} s^{-\rho} ds \quad , \quad \lambda_n \sim \int_n^{m(t_n+T)} s^{-2\rho} ds .$$

We have $m(t_n+T)^{1-\rho} = n^{1-\rho} + (1-\rho)T$ or $m(t_n+T) \sim n[1+T/n^{1-\rho}]$, and $m(t_n+T)-n \sim Tn^{\rho}$. This yields $\lambda_n \sim Tn^{-\rho}$. Now, evaluating (6.3) yields (6.3) $= \Delta/T + o(\Delta)$. Thus $h(t) = 1/T$.

6.3 <u>The case</u> $\rho = 1$ <u>and</u> (1.2), <u>with</u> $1 > 2\gamma$. Here λ_n is defined by (4.4'). Thus $\lambda_n \sim (1-2\gamma)^{-1} n^{2\gamma-1}(1-e^{-T(1-2\gamma)})$. To obtain the proper weighing function $h(\cdot)$, we need to evaluate

$$\int_{m(t_n+t)}^{m(t_n+t+\Delta)} s^{-2\rho+2\gamma} ds/\lambda_n = h(t)\Delta + o(\Delta) \tag{6.4}$$

The left side of (6.4) is asymptotically equivalent to

$$\Delta(1-2\gamma)e^{-t(1-2\gamma)}(1-e^{-T(1-2\gamma)})^{-1} + o(\Delta) = \Delta h(t) + o(\Delta).$$

6.4 <u>The case</u> $\rho < 1$ <u>and</u> (1.2) <u>with</u> $\rho > 2\gamma$. Here $\lambda_n \sim Tn^{-\rho+2\gamma}$ and $h(t) = 1/T$.

6.5 <u>The Lagrangian</u> $L(\cdot,\cdot,\cdot)$ <u>and</u> V <u>of</u> (1.3). Since $L(\cdot,\cdot,\cdot)$ and λ_n have common factors, we define some new terms in order to obtain the simplest form of the asymptotic estimates. Define the set (replaces A in the degenerate case)

$$A_D = \{\phi : \phi(0) = x,\ \dot{\phi}_1 = b_1(\phi),\ \phi(t) \in \partial G, \text{ some } t \le T\}.$$

For (1.1) the $L(\cdot,\cdot,\cdot)$ are given by (5.2). Define

$$\bar{S}_1 = \inf_{\phi \in A} \int_0^T L_0(\phi(s),\dot{\phi}(s))e^s\,ds$$

$$\bar{S}_0 = \inf_{\phi \in A} \int_0^T L_0(\phi(s),\dot{\phi}(s))\,ds$$

$$\bar{S}_1^D = \inf_{\phi \in A_D} \int_0^T L_0(\phi(s),\dot{\phi}(s))e^s\,ds$$

$$\bar{S}_0^D = \inf_{\phi \in A_D} \int_0^T L_0(\phi(s),\dot{\phi}(s))\,ds.$$

For the Kiefer-Wolfowitz procedure (1.2), we need the definition

$$\bar{S}_1(kw) = \inf_{\phi \in A} \int_0^T L_0(\phi(s),\dot{\phi}(s))e^{(1-2\gamma)s}\,ds.$$

Define $\bar{S}_1^D(kw)$ similarly, with A_D replacing A.

For (1.1), we finally obtain for the <u>non-degenerate case</u>

$$\lim_n n^{-1} \log P_x\{\tau_G^n \le T\} = -\bar{S}_1 \quad (\text{for } \rho = 1)$$

(6.5)

$$\lim_n n^{-\rho} \log P_x\{\tau_G^n \le T\} = -\bar{S}_0 \quad (\text{for } \rho < 1)$$

For the <u>degenerate case</u>, replace $\bar{S}_0$ and $\bar{S}_1$ by $\bar{S}_0^D$ and $\bar{S}_1^D$, resp., in (6.5). All $\bar{S}_\gamma^\beta$ are continuous in x in a neighborhood of 0. For (1.2), we have for the <u>non-degenerate case</u>

$$\lim_n n^{-(1-2\gamma)} \log P_x \{\tau_G^n \leq T\} = -\bar{S}_1(kw), \quad (\rho=1)$$

(6.6)

$$\lim_n n^{-(\rho-2\gamma)} \log P_x \{\tau_G^n \leq T\} = -\bar{S}_0 , \quad (\rho<1)$$

with the obvious alterations of the right side for the degenerate case.

Note that $\bar{S}_1 > \bar{S}_0$. This, together with the relationship $n > n'$ for $\rho < 1$, implies the 'considerable' superiority of the coefficient sequence $a_n = n^{-1}$.

References

[1] H. J. Kushner, D. S. Clark, "Stochastic Approximation Methods for Constrained and Unconstrained Systems", Appl. Math. Sciences Ser. #26, Springer, Berlin.

[2] L. Ljung, "Analysis of recursive stochastic algorithms", IEEE Trans. on Automatic Control, AC-22, 1977, 551-575.

[3] D. Ludwig, "Persistence of dynamical systems under random perturbations", SIAM Rev, 17, 1975, 605-640.

[4] B. Z. Bobrovsky, Z. Schuss, "A singular perturbation method for the computation of the first passage time in a non-linear filter", SIAM J. on Appl. Math, 42, 1982, 174-187.

[5] M. I. Friedlin, "The averaging principle and theorems on large deviations", Russian Math Surveys, 33, July-Dec., 1978.

[6] G. Ruget, Some tools for the study of channel-sharing algorithms CISM publication ed. by Longo, Springer, 1980.

[7] M. Cottrel, J. C. Fort, G. Malgouyres, "Evénments rare pour l'étude de certain algorithms stochastiques", Rept. 80.T.35, Univ. Paris Sud, Dept. of Mathematics, 1980.

[8] H. J. Kushner, Huang, H., "Rates of convergence for stochastic approximation type algorithms", SIAM J. on Control and Optimiz., 19, 1981,87-105.

[9] Grandes Deviations et Applications Statistiques, Astérisque, 68, 1979, Société Mathématique de France.

[10] J. Gartner, "On large deviations from the invariant measure", Theory of Probability and its Applications, 22, 1977, 24-39.

[11] H. J. Kushner, "Robustness and approximation of escape times and large deviations estimates for systems with small noise effects", LCDS Report 82-5, Brown Univ., to appear in SIAM J. on Appl. Math.

ATE

MED