The Annals of Probability
1990, Vol. 18, No. 2, 698–712

NONCONVERGENCE TO UNSTABLE POINTS IN URN MODELS AND STOCHASTIC APPROXIMATIONS[1]

By Robin Pemantle

Cornell University

A particle in $\mathbf{R}^d$ moves in discrete time. The size of the nth step is of order $1/n$ and when the particle is at a position $\mathbf{v}$ the expectation of the next step is in the direction $\mathbf{F}(\mathbf{v})$ for some fixed vector function $\mathbf{F}$ of class C^2. It is well known that the only possible points $\mathbf{p}$ where $\mathbf{v}(n)$ may converge are those satisfying $\mathbf{F}(\mathbf{p}) = \mathbf{0}$. This paper proves that convergence to some of these points is in fact impossible as long as the "noise"—the difference between each step and its expectation—is sufficiently omnidirectional. The points where convergence is impossible are the unstable critical points for the autonomous flow $(d/dt)\mathbf{v}(t) = \mathbf{F}(\mathbf{v}(t))$. This generalizes several known results that say convergence is impossible at a repelling node of the flow.

1. Introduction. Let $\mathbf{F}$: $\mathbf{R}^d \to \mathbf{R}^d$ be a vector field, and consider a particle moving in discrete time whose position $\mathbf{v}(n)$ obeys the law

$$\mathbf{v}(n+1) = \mathbf{v}(n) + a_n \mathbf{F}(\mathbf{v}(n)) + \xi_n. \tag{1}$$

To explain the symbols in the foregoing equation, let $\mathscr{F}_n$ be the σ-algebra of events up to time n, assume that a_n is an $\mathscr{F}_n$-measurable random variable and that ξ_n is a random vector with $\mathbf{E}(\xi_n | \mathscr{F}_n) = \mathbf{0}$. In other words, from time n to time $n+1$ the particle moves a fraction a_n of the way along the "arrow" $\mathbf{F}(\mathbf{v}(n))$, where a_n is determined by the past, and then its position is further altered by adding some mean-zero noise, ξ_n. This setup is fairly general and encompasses several urn models and stochastic approximations that will be detailed below. In all of these examples, the noise has bounded size, but no further assumptions are made as to independence, or to the noise having a classical distribution, such as Gaussian. In fact such assumptions do not make the problems any easier.

The literature on these processes contains results of two types. The first type are convergence results. These say that $\mathbf{v}(n)$ converges almost surely to some possibly random limit $\mathbf{v}$ as $n \to \infty$. For these results, a_n and ξ_n are almost always taken to be on the order of $1/n$. In urn models or reinforcement models (see Section 2) this means that all times in the past have equal effects on the future. Also assumptions are needed on $\mathbf{F}$ ranging from mild continuity assumptions in the one-dimensional case to strong geometric conditions in the n-dimensional case to prevent the particle from running around in circles.

Received September 1988; revised April 1989.
[1]This research was supported by an NSF graduate fellowship and by an NSF postdoctoral fellowship.
AMS 1980 *subject classifications*. Primary 60G99; secondary 62L20.
Key words and phrases. Reinforced random walk, stochastic approximation, unstable equilibrium, urn model.

REMARK. The reason $1/n$ is a natural choice for the step size is that

$$\sum 1/n = \infty \quad \text{and} \quad \sum 1/n^2 < \infty. \tag{2}$$

This means that $\mathbf{v}(n)$ does not converge in a trivial manner by having a path of bounded variation, but that the total variance of the increments is finite so that at points where the drift $\mathbf{F}$ disappears, $\mathbf{v}(n)$ may converge in the manner of a diffusion whose clock converges. Furthermore, all of the motivating examples for this study have step size $1/n$, arising naturally from the interpretation. Replacing $1/n$ by $1/n^{\gamma}$ for $1/2 < \gamma < 1$ leaves the finite and infinite sums in (2) intact and the results of this paper are valid in that case although I know of no application where $\gamma < 1$.

The second type of result characterizes the law of the random limit $\mathbf{v}$. These results are qualitative, saying for example that the probability of finding $\mathbf{v}$ in a certain set is 1 or perhaps only nonzero. The aim of this paper is to prove a theorem that generalizes several existing results of the second type. In the next section the existing results will be summarized and the main theorem, Theorem 1, will be stated formally. Here is a brief description.

It is generally easy to show that the limit $\mathbf{v}$ if it exists satisfies $\mathbf{F}(\mathbf{v}) = \mathbf{0}$ almost surely. Since the stochastic process of (1) is in some sense a discrete version of the differential flow

$$\frac{d}{dt}\mathbf{v}(t) = \mathbf{F}(\mathbf{v}(t)), \qquad \mathbf{v}(0) = \mathbf{x}_0, \tag{3}$$

it is natural to classify the points where $\mathbf{F}(\mathbf{v}) = \mathbf{0}$ according to what kind of critical points they are for the flow. To do so, let T be the linear approximation to $\mathbf{F}$ near a critical point $\mathbf{p}$ so that $\mathbf{F}(\mathbf{p} + \mathbf{v}) = T(\mathbf{v}) + O(|\mathbf{v}|^2)$. Then $\mathbf{p}$ is an attracting point if all the eigenvalues of T have negative real part, in which case the flow (3) always converges to $\mathbf{p}$ as $t \to \infty$ for $\mathbf{x}_0$ in some neighborhood of $\mathbf{p}$. If some eigenvalue of T has positive real part then $\mathbf{p}$ is an unstable critical point and the flow exits a neighborhood of $\mathbf{p}$ for all choices of $\mathbf{x}_0$ that are not on a stable manifold of smaller dimension. Since the word *unstable* is ambiguous in common usage, call $\mathbf{p}$ *linearly unstable* in this case. If all the real parts of the eigenvalues of T are positive then $\mathbf{p}$ is a repelling node and the flow can never converge to $\mathbf{x}_0$ unless it begins there.

The results quoted from the literature in Section 2 give various conditions implying that $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$ when $\mathbf{p}$ is a repelling point. Theorem 1 gives a stronger result: $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$ if $\mathbf{p}$ is a linearly unstable critical point.

There are two inherent limitations on the usefulness of this theorem. First, it is not very interesting unless the corresponding convergence theorem is known. That is, no one cares that $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$ unless $\mathbf{v}(n)$ is known to converge somewhere. The results quoted in the next section all have their corresponding convergence theorems, so this generalization is meaningful at least in those cases. Second, the theorem is rather weak unless the set of points $\{\mathbf{x}: \mathbf{F}(\mathbf{x}) = 0\}$ is discrete. When this set is discrete, Theorem 1 says that unstable critical points are not in the support of the random limit $\mathbf{v}$. But if there are uncountably many unstable critical points then $\mathbf{v}(n)$ can still converge to one of them even though each point has probability 0 of being hit. In this case, all one can conclude is that the law of the limit has no point masses at unstable critical points. This information may be nontrivial; see Pemantle (1989).

As mentioned before, Section 2 states the main theorem and fits it to known results. The proof is then given in Sections 3 and 4, with Section 3 containing the geometry and Section 4 containing the probability.

2. Statement of Theorem 1 and motivating examples. Because the motion of the particle in some of the examples is restricted to a subset of $\mathbf{R}^d$, the theorem will be stated in sufficient generality to allow for this. Let $\Delta \subseteq \mathbf{R}^d$ be an open subset of an affine subspace in $\mathbf{R}^d$. Let $\mathbf{F}\colon \Delta \to T\Delta$ be of class C^1, where $T\Delta$ is the translation of the affine subspace that contains the origin. (Note: The tangent space to Δ at each point will be identified with $T\Delta$. Thus $\mathbf{F}$ is viewed as a tangent vector field as well as a vector-valued function.) Define a stochastic process according to (1):

$$\mathbf{v}(n+1) = \mathbf{v}(n) + a_n \mathbf{F}(\mathbf{v}(n)) + \xi_n,$$

with a_n being $\mathscr{F}_n$-measurable and $\mathbf{E}(\xi_n | \mathscr{F}_n) = \mathbf{0}$ and such that $\mathbf{v}(n)$ always remains in Δ.

THEOREM 1. *Let the stochastic process $\{\mathbf{v}(n)\colon n \geq 0\}$ be defined so that it satisfies* (1) *for some sequence of random variables $\{a_n\}$ and $\{\xi_n\}$ as described after* (1). *Let $\mathbf{p}$ be any point of Δ with $\mathbf{F}(\mathbf{p}) = 0$, let $\mathscr{N}$ be a neighborhood of $\mathbf{p}$ and assume that there are constants $\gamma \in (1/2, 1]$ and $c_1, c_2, c_3, c_4 > 0$ for which the following conditions are satisfied whenever $\mathbf{v}(n) \in \mathscr{N}$ and n is sufficiently large*:

(4) (i) $\mathbf{p}$ *is a linearly unstable critical point*,

(5) (ii) $c_1/n^\gamma \leq a_n \leq c_2/n^\gamma$,

(6) (iii) $\mathbf{E}\big((\xi_n \cdot \theta)^+ | \mathscr{F}_n\big) \geq c_3/n^\gamma$ *for every unit vector* $\theta \in T\Delta$,

(7) (iv) $|\xi_n| \leq c_4/n^\gamma$,

where $(\xi \cdot \theta)^+ = \max\{\xi \cdot \theta, 0\}$ *is the positive part of* $\xi \cdot \theta$.

Assume $\mathbf{F}$ *is smooth enough to apply the stable manifold theorem*: *at least* C^2. *Then* $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$.

Condition (iii) looks strange, but it just says that the increment ξ_n is sometimes on the order of $1/n^\gamma$ in any direction. To see why this is necessary, suppose that $\Delta = \mathbf{R}^2$, that ξ_n is $1/n^\gamma$ times a standard normal in the x-direction and that $\mathbf{F}(x, y) = (-x, y)$. Then the origin is unstable in the y-direction but the process starting at a point on the x-axis can converge to the origin because it never gets jiggled in the y-direction. Some of the conditions can be weakened. For instance both (ii) and (iv) can be allowed to fail on sets whose probabilities decrease fast enough with time. And condition (i) may be replaced by the condition that $\mathbf{F}(\mathbf{v})$ is identically 0, which is useful for the two applications mentioned in the last paragraph of this section. See Pemantle (1988), Section 3.2, for these variations.

Examples. In all the examples, γ is taken to be 1. The first example is an urn model. Hill, Lane and Sudderth (1980) consider the following process. An urn containing red and black balls has a given initial composition. At each new time a ball is drawn from the urn and replaced along with another ball of the same color. The draws are not exactly representative of the contents of the urn but are determined by the contents in the following manner. Let the number of red and black balls at time n be R_n and B_n, respectively, and let $v(n) = R_n/(R_n + B_n)$. Instead of drawing a red ball with probability $v(n)$ draw a red ball with probability $f(v(n))$, where f is any function mapping $[0, 1]$ into itself.

Hill, Lane and Sudderth show that under a condition on the discontinuities of f, $v(n)$ converges almost surely to a random variable v for which $f(v) = v$. They also give the following nonconvergence theorem.

THEOREM [Hill, Lane and Sudderth (1980)]. *Suppose $f(p) = p$ and that p is an upcrossing for f, i.e., $f(x) < x$ for $x < p$ in some neighborhood of p and $f(x) > x$ for $x > p$ in some neighborhood of p. Then $\mathbf{P}(v(n) \to p) = 0$.*

To see how this fits into the framework of the previous section, let $F(x) = f(x) - x$ and note that

$$\begin{aligned}\mathbf{E}(v(n+1) \mid \mathscr{F}_n) &\\ &= f(v(n)) \frac{R_n + 1}{R_n + B_n + 1} + (1 - f(v(n))) \frac{R_n}{R_n + B_n + 1} \\ &= v(n) + \frac{1}{R_n + B_n + 1} F(v(n)).\end{aligned}$$

So if $a_n = 1/(R_n + B_n + 1)$ and ξ_n is defined by the equation

$$v(n+1) = v(n) + a_n F(v(n)) + \xi_n,$$

then $\mathbf{E}(\xi_n \mid \mathscr{F}_n) = \mathbf{0}$ and the process $\mathbf{v}(n)$ obeys (1). When F is smooth and the upcrossing is not tangential, $\mathbf{p}$ is linearly unstable and condition (i) of Theorem 1 is satisfied. The other hypotheses are evidently satisfied, so the result in this case is included in Theorem 1.

The next example is a generalization of this model to urns of more than two colors. Arthur, Ermol'ev and Kaniovskii (1983) consider the following multicolor version. Let there be balls of colors $1, \dots, d$ in the urn initially and let the urn evolve as before by drawing a ball and replacing it along with another ball of the same color. This time the probability of drawing a ball of color i at time n is allowed to depend on the fractions of all colors present. The dependence is via the vector quantity $\mathbf{v}(n)$, whose coordinates $v_j(n)$ are the fractions of the balls in the urn having color j for $j = 1, \dots, d$. Writing $f_i(\mathbf{v})$

for the probability of drawing color i when the vector of fractions is $\mathbf{v}$, the function $\mathbf{f}$ with coordinates f_i is a vector function mapping the unit simplex $\Delta \subseteq \mathbf{R}^d$ into itself. Let $\mathbf{F}$: $\Delta \to T\Delta$ be defined by $\mathbf{F}(\mathbf{v}) = \mathbf{f}(\mathbf{v}) - \mathbf{v}$. As before, it is easily verified that the stochastic process $\{\mathbf{v}(n)\}$ obeys the law

$$\mathbf{v}(n+1) = \mathbf{v}(n) + \frac{1}{n+1+v_1(0)+\cdots+v_d(0)}\mathbf{F}(\mathbf{v}(n)) + \xi_n,$$

where the noise ξ_n has mean 0, conditioned on the past.

Arthur, Ermol'ev and Kaniovskii generalize Hill, Lane and Sudderth's definition of an upcrossing by defining a point $\mathbf{p}$ with $\mathbf{F}(\mathbf{p}) = \mathbf{0}$ to have the *s-property* if there is some positive definite matrix C such that the dot product of $C(\mathbf{F}(\mathbf{v}))$ with $\mathbf{v} - \mathbf{p}$ is always positive whenever $\mathbf{F}(\mathbf{v}) \neq 0$. This is slightly stronger than requiring $\mathbf{p}$ to be a repelling point since a repelling point only needs to satisfy this positivity property for $\mathbf{v}$ in some neighborhood of $\mathbf{p}$. The reason they assume the global property is that it is necessary for their convergence theorem.

THEOREM [Arthur, Ermol'ev and Kaniovskii (1983), Theorem 2]. *With an urn process defined as above, suppose* $\mathscr{F}(\mathbf{p}) = 0$, $\mathbf{p}$ *has the s-property, and also suppose that* $\mathbf{F}$ *is Hölder with some exponent* $\mu > 0$ *in a neighborhood of* $\mathbf{p}$. *Then* $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$.

Essentially the same theorem is proved in Nevel'son and Hasminskii's (1973) book, but for stochastic approximations instead of for urn models. They consider processes such as the Robbins–Monro (1951) and the Kiefer–Wolfowitzapproximation procedures, which can be written in the form of (1). They have a theorem (Theorem 4.1 of Chapter 5) saying that the probability of convergence to a point $\mathbf{p}$ is 0 if $\mathbf{F}$ satisfies the s-condition, under some additional hypotheses on the magnitudes of $\{a_n\}$ and $\{\xi_n\}$. The additional hypotheses are too complicated to reproduce here, but it is worth noting that they are satisfied when a_n and $|\xi_n|$ are on the order of $1/n$. It is also worth mentioning how they get their convergence theorems. They find a Liapounov function V which always increases in the direction $\mathbf{F}$, i.e., $\nabla V(\mathbf{v}) \cdot \mathbf{F}(\mathbf{v}) \geq 0$ with equality only when $\mathbf{F}(\mathbf{v}) = \mathbf{0}$. This Liapounov function can always be found in the one-dimensional case, but its existence needs to be assumed in higher dimensions.

The next example is a process called vertex reinforced random walk defined by Diaconis (personal communication). A particle moves in discrete time among the states $1, \dots, d$. There is a positive symmetric matrix R that governs the motion of the particle by letting the transition probabilities from state i to state j at time n be proportional to $R_{ij}\chi_n(j)$, where $\chi_n(j)$ is 1 + the number of visits to state j up to time n. Letting $v_i(n) = \chi_{n-d}(i)/n$, the vector $\mathbf{v}(n)$ lies on the unit simplex and obeys (1) for suitably defined $\mathbf{F}$, a_n and ξ_n. There is a Liapounov function V which implies the almost sure convergence of $\mathbf{v}(n)$ under some nonsingularity conditions on R. The Liapounov function has only one critical point $\mathbf{p}$ in the interior of the simplex and it can be shown that (5)–(7) are always satisfied, whereas (4) is true if and only if $\mathbf{p}$ is not a weak maximum for the Liapounov function V. The following theorem is then a somewhat lengthy application of Theorem 1.

THEOREM [Pemantle (1988), Theorem 5.12]. *Let* $\mathbf{p}$ *be the unique critical point of* V *in the interior of the simplex. Then* $\mathbf{P}(\mathbf{v}(n) \to \mathbf{p}) = 0$ *if* $\mathbf{p}$ *is not a (weak) maximum for* V. *This happens if and only if the matrix* R *has more than one positive eigenvalue.*

Theorem 1 subsumes other variants on urn schemes. The number of balls added of the color drawn may depend on time as in Pemantle (1989) or may be a random function of the color drawn as in Athreya (1969). Since these two papers are concerned only with linear reinforcement [i.e., $\mathbf{F}(\mathbf{v}) \equiv 0$], there are no linearly unstable equilibria and Theorem 1 cannot be applied. However, as mentioned above, the instability assumption may be replaced by the assumption that $\mathbf{F}(\mathbf{v}) \equiv \mathbf{0}$ with the same conclusion [see Pemantle (1988), Chapter 3]. When these variants are added to a nonlinear urn with some general urn function $\mathbf{F}$ as in Arthur, Ermol'ev and Kaniovskii (1983), Theorem 1 may again be applied to give previously unknown results.

3. Proof of Theorem 1: Geometric part. The proof of Theorem 1 presented here follows the methods of Pemantle (1988), Theorem 3.5, and Pemantle (1989), Theorem 3. The basic idea is as follows. Because the matrix T has eigenvalues with positive real part, there are some directions in which any perturbation of the particle ought to be positively reinforced. The first step is to find a function $\eta\colon \Delta \to \mathbf{R}$ that somehow measures the distance of the particle from $\mathbf{p}$ in an unstable direction. The trick here is to make sure that $\mathbf{F}$ always points in a direction that η increases and that η is smooth enough to work with. The next step is to prove that the noise ξ_n will always cause the particle to stray a distance on the order of $n^{1/2-\gamma}$ from $\mathbf{p}$. The last step is to prove that once the particle has strayed that far, it may never return to a smaller neighborhood of $\mathbf{p}$. By a tail law, if it keeps straying away with a probability of returning that is bounded away from 1, it eventually must stay away.

The construction of the function η involves some technical geometric details which may be skipped; the relevant properties are summarized in Proposition 3 below. Whether the process $\mathbf{v}(n)$ can converge to $\mathbf{p}$ with nonzero probability is clearly a local property, i.e., it depends only on the behavior of $\mathbf{F}$ in a neighborhood of $\mathbf{p}$. So Theorem 1 is first proved under global hypotheses which are weakened to local hypotheses in the last paragraph of the proof by a coupling argument.

One of the tools necessary for the construction of η is the stable manifold theorem stated below, for whose proof the reader is referred to Hirsch, Pugh and Schub (1977), Theorem 5A.1. Let $\mathbf{F}\colon \Delta \to T\Delta$ be a C^2 vector field with $\mathbf{F}(\mathbf{p}) = \mathbf{0}$ and let $T\colon T\Delta \to T\Delta$ be the linear approximation to $\mathbf{F}(\mathbf{p} + \mathbf{v})$ so that $\mathbf{F}(\mathbf{p} + \mathbf{v}) = T(\mathbf{v}) + O(|\mathbf{v}|^2)$. If T has no purely imaginary eigenvalues then $\mathbf{F}$

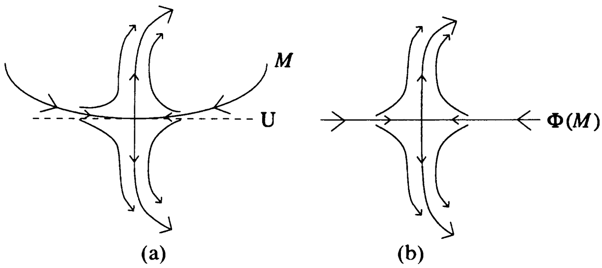

FIG. 1. *(a) The flow given by* $\mathscr{F}(x,y) = (-x + k(x^2 - y)^4, y - 3x^2)$, *with stable manifold* $\mathscr{M}$. *(b) The map* Φ *translates each vertical by the appropriate amount so that* $\mathscr{M}$ *is mapped to U.*

may be taken to be only C^1. Then $T\Delta$ decomposes as $U \oplus W$, where $T(U) \subseteq U$, $T(W) \subseteq W$, $T|_W$ has only eigenvalues with positive real part and $T|_U$ has only eigenvalues with nonpositive real part. (Hereafter, T, U and W will be defined as in the previous sentence.) The stable manifold theorem gives a C^2 invariant manifold for the flow $(d/dt)\mathbf{v}(t) = \mathbf{F}(\mathbf{v}(t))$ which is tangent to U and is the graph of a C^2 function g: $U \to W$. The notation for multidimensional derivatives employed below uses $\mathbf{D}\Phi$ or $\mathbf{D}_\mathbf{v}(\Phi)$ for the differential of Φ at $\mathbf{v}$, so for example if Φ: $\Delta \to \mathbf{R}$ is a scalar function then $\mathbf{D}_\mathbf{v}(\Phi)$ is a function mapping tangent vectors at $\mathbf{v}$ into directional derivates of Φ at $\mathbf{v}$.

THEOREM (Stable manifold theorem). *Let* $\mathbf{F}$, $\mathbf{p}$, T, U *and* W *be defined as above. Then there is a neighborhood* $\mathscr{N}$ *of* $\mathbf{0}$ *in* U *and a* C^2 *function* g: $\mathscr{N} \to W$ *such that* $g(\mathbf{0}) = \mathbf{0}$, $\mathbf{D}_\mathbf{0}(g) = 0$ *and the set of points* $\mathscr{M} = \{\mathbf{p} + \mathbf{u} + g(\mathbf{u})\colon \mathbf{u} \in U\}$ *is an invariant manifold for the flow* $(d/dt)\mathbf{v}(t) = \mathbf{F}(\mathbf{v}(t))$. *In other words,* $\mathbf{F}$ *is always tangent to* $\mathscr{M}$.

To illustrate this construction, suppose the dimension is 2 and $\mathbf{F}(x,y) = (-x + k(x^2 - y)^4, y - 3x^2)$ [see Figure 1(a).] Then the origin is a critical point whose stable manifold $\mathscr{M}$ is the set $y = x^2$. The linearization of $\mathbf{F}$ at the origin is given by $T(x,y) = (-x, y)$ so the subspace U corresponding to the negative eigenvalues is just the x-axis, which is tangent to $\mathscr{M}$ at the origin.

The program for constructing η will be to assume first that the stable manifold $\mathscr{M}$ is actually equal to $\mathbf{p} + U$ and construct η in this case. [This happens when $\mathbf{F}(\mathbf{p} + \mathbf{u})$ is in U for $\mathbf{u} \in U$.] The general case can then be reduced to this case by a C^2 change of coordinates involving the function g from the stable manifold theorem. The following proposition constructs η in this special case, by constructing a quadratic form D which will be the square of η. It is a well-known construction whose proof can be found in Hirsch and Smale (1974), page 145 and following.

PROPOSITION 1. *Let* T: $\mathbf{R}^d \to \mathbf{R}^d$ *be a linear operator all of whose eigenvalues have real part* $\geq \lambda_{\min} > 0$. *Then there is a* $k > 0$ *and a quadratic form* D

such that

$$D(\mathbf{v} + \varepsilon T\mathbf{v}) \geq (1 + k\varepsilon) D(\mathbf{v}) \quad \textit{for all } \varepsilon \in [0,1]. \tag{8}$$

(*In fact, k can be chosen arbitrarily close to $\lambda_{\min}$.*)

In the special case $\mathscr{M} = U$, the definition of η is now completed by letting $\eta(\mathbf{u} + \mathbf{w}) = (D(\mathbf{w}))^{1/2}$ for $\mathbf{u} \in U$ and $\mathbf{w} \in W$.

Proposition 2. *Let $\mathbf{F}$, $\mathbf{p}$, T, U and W be as above and assume $\mathscr{M} = U$. Then the quadratic form D constructed above satisfies*

$$D(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) \geq (1 + k_1\varepsilon) D(\mathbf{v})$$

in some neighborhood of $\mathbf{p}$ for any $k_1 < k$.

Proof. The proposition will be proved by examining the power series expansion of $\mathbf{F}$. Since $\mathbf{F}(\mathbf{p} + U) \subseteq U$, the key approximation will require only the W component of the argument to be small. Pick a basis so that the first r elements are in U and the remaining $d - r$ are in W. All the calculations will take place in this basis, so let the coordinates of $\mathbf{v}$ in this basis be denoted $a_1, \dots, a_r$ and $b_{r+1}, \dots, b_d$. Let the quadratic form D be represented in this basis by (D_{ij}); since D only pays attention to the part of $\mathbf{v}$ in W, this means $D(\mathbf{v}) = \sum D_{ij} b_i b_j$. Each coordinate $F_i(\mathbf{p} + \mathbf{v})$ of the vector function $\mathbf{F}(\mathbf{p} + \mathbf{v})$ has a power series expansion in the variables $\{a_i\}$ and $\{b_i\}$, but since $\mathbf{F}(\mathbf{p} + U) \subseteq U$, $F_i(\mathbf{p} + \mathbf{v}) = 0$ whenever $i > r$ and $\mathbf{v} \in U$. This means that for $i > r$ the power series expansion of $F_i(\mathbf{p} + \mathbf{v})$ looks like

$$F_i(\mathbf{p} + \mathbf{v}) = (T\mathbf{v})_i + \sum c_{jk} a_j b_k + \sum c_{jk} b_j b_k + o(a_j b_k \text{ and } b_j b_k \text{ terms}).$$

The important fact is that all the higher-order monomials have some b_k in them. Then

$$D(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) = \sum D_{ij}\big(b_i + \varepsilon(T\mathbf{v})_i + \varepsilon \cdot \text{stuff}\big)\big(b_j + \varepsilon(T\mathbf{v})_j + \varepsilon \cdot \text{stuff}\big),$$

where the "stuff" represents monomials of quadratic and higher order in the a_i's and b_i's having at least one b term. Since $(T\mathbf{v})_i$ is linear in the variables b_i, all the cross terms from the above equation involving "stuff" are at least cubic in the a_i's and b_i's with at least two factors that are b_i's. So

$$D(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) = D(\mathbf{v} + \varepsilon T\mathbf{v}) + \varepsilon O(a_i b_j b_k \text{ and } b_i b_j b_k \text{ terms}).$$

Then, using the fact that $b_j b_k$ is always at most a constant times $D(\mathbf{v})$, a neighborhood can be chosen to make $\sup_i\{a_i, b_i\}$ small enough so that all the $\varepsilon O(a_i b_j b_k)$ and $\varepsilon O(b_i b_j b_k)$ terms sum to less than $(k - k_1)\varepsilon D(\mathbf{v})$. Together with the fact that $D(\mathbf{v} + \varepsilon T\mathbf{v}) \geq (1 + k\varepsilon) D(\mathbf{v})$, this implies $D(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) \geq (1 + k_1\varepsilon) D(\mathbf{v})$. □

Finally, η can be defined in the general case. Forget the requirement that $\mathscr{M} = U$ and let $\mathbf{F}$ be any C^2 vector field with $\mathbf{F}(\mathbf{p}) = \mathbf{0}$ and T, U and W as above. Use the stable manifold theorem to get a neighborhood $\mathscr{N}$ of $\mathbf{0}$ in U and a C^2 function g: $\mathscr{N} \to W$ with $g(\mathbf{0}) = \mathbf{0}$, $\mathbf{D_0}(g) = 0$ and the stable manifold $\mathscr{M}$ given by $\{\mathbf{p} + \mathbf{u} + g(\mathbf{u}) \colon u \in U\}$. Define Φ: $\Delta \to \Delta$ by writing $\mathbf{v} \in T\Delta$ as $\mathbf{u} + \mathbf{w}$ for $\mathbf{u} \in U$ and $\mathbf{w} \in W$ and letting

$$\Phi(\mathbf{p} + \mathbf{v}) = \mathbf{v} - g(\mathbf{u}). \tag{9}$$

Then $\Phi(\mathscr{M}) \subseteq \mathbf{p} + U$ and therefore $\mathbf{D}\Phi$ maps tangent vectors to $\mathscr{M}$ into U. In particular, $\mathbf{D}\Phi \circ \mathbf{F} \circ \Phi^{-1}$ maps U into itself. Also $\mathbf{D_p}(\Phi) \circ \mathbf{F} \circ \Phi^{-1}(\mathbf{p}) = \mathbf{0}$ and since $\mathbf{D_0}(g) = 0$, the linear approximation to $\mathbf{D}_{\Phi-1(\mathbf{p}+\mathbf{v})}(\Phi) \circ \mathbf{F} \circ \Phi^{-1}(\mathbf{p} + \mathbf{v})$ is just T. Then the hypotheses of the preceding proposition are satisfied with $\mathbf{D}\Phi \circ \mathbf{F} \circ \Phi^{-1}$ in place of $\mathbf{F}$. Apply the proposition to get a quadratic form D and let

$$\eta(\mathbf{v}) = \left[D(\Phi(\mathbf{v}) - \mathbf{p})\right]^{1/2}. \tag{10}$$

Figure 1(b) illustrates how Φ straightens out $\mathscr{M}$ by mapping it to U. Note that Φ is not a linearization, i.e., Φ does not straighten out any of the other invariant manifolds. Since $\mathscr{M}$ has codimension 1 in this example, η is (up to a constant multiple) just the absolute value of the vertical distance to the stable manifold, so $\eta(x, y) = |y - x^2|$. In higher codimensions, η measures the distance to the stable manifold in a direction parallel to W, where distance is given by the appropriate quadratic form so as to be nondecreasing in the $\mathbf{F}$ direction.

The next proposition summarizes the properties of Φ and states all the facts about D and η that are needed in Section 4.

Proposition 3. *Let $\mathbf{F}$ be a C^2 vector field with a linearly unstable point $\mathbf{p}$. There is a C^2 map Φ from a neighborhood of $\mathbf{p}$ in Δ to Δ such that $\Phi(\mathbf{p}) = \mathbf{p}$, $\mathbf{D_p}(\Phi)$ is the identity, and $\Phi(\mathscr{M}) \subseteq \mathbf{p} + U$, where $\mathscr{M}$ is the stable manifold associated with the flow $\mathbf{F}$ and U is its tangent space at $\mathbf{p}$. Furthermore, there are a quadratic form D and a function η defined in terms of D by* (10) *such that the following are true*:

(i) $D^{1/2}(c\mathbf{v}) = cD(\mathbf{v})$ *for* $c > 0$.

(ii) $D^{1/2}$ *is convex, i.e.,* $D^{1/2}(c\mathbf{v}_1 + (1 - c)\mathbf{v}_2) \le cD^{1/2}(\mathbf{v}_1) + (1 - c)D^{1/2}(\mathbf{v}_2)$ *for* $c \in [0, 1]$.

(iii) $D^{1/2}$ *is Lipschitz.*

(iv) *There exist* $k_2, k_3 > 0$ *such that* $\eta(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) \ge (1 + k_2\varepsilon)\eta(\mathbf{v}) - k_3\varepsilon^2$ *for* $\varepsilon \in [0, 1]$ *and* $\mathbf{v}$ *in some neighborhood of* $\mathbf{p}$.

(v) *For* $\eta(\mathbf{v}) \ne 0$ *in some neighborhood of* $\mathbf{p}$, $\mathbf{D_v}(\eta)(\mathbf{F}(\mathbf{v})) > 0$.

Proof. (i) is obvious. To prove (ii), recall from the construction that there is a basis in which $D^{1/2}$ is just the standard norm of a projection onto a subspace. Then $D^{1/2}$ is a seminorm and hence convex. To prove (iii), note that $D^{1/2}$ is Lipschitz at the origin because it is the square root of a quadratic. By (i) and (ii) $D^{1/2}$ satisfies the triangle inequality so Lipschitz at the origin implies Lipschitz everywhere. To show (iv), calculate

$$\begin{aligned}
&\eta(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) \\
&\quad = \left[D(\Phi(\mathbf{v} + \varepsilon\mathbf{F}(\mathbf{v})) - \mathbf{p})\right]^{1/2} \\
&\quad = \left[D(\Phi(\mathbf{v}) + \varepsilon\mathbf{D}\Phi \circ \mathbf{F}(\mathbf{v}) + O(\varepsilon^2) - \mathbf{p})\right]^{1/2} \\
&\quad = \left[D(\mathbf{y} + \varepsilon\mathbf{D}\Phi \circ \mathbf{F} \circ \Phi^{-1}(\mathbf{y}) - \mathbf{p})\right]^{1/2} + O(\varepsilon^2) \\
&\qquad\quad \text{where } \mathbf{y} = \Phi(\mathbf{v}) \text{ and the } O(\varepsilon^2) \text{ term can be pulled out by (iii)} \\
&\quad \le \left[(1 + k_1\varepsilon)D(\mathbf{y} - \mathbf{p})\right]^{1/2} + O(\varepsilon^2) \\
&\qquad\qquad\qquad\qquad \text{for } \mathbf{v} \text{ in some neighborhood of } \mathbf{p} \\
&\quad = (1 + k_1\varepsilon)^{1/2}\eta(\mathbf{v}) + O(\varepsilon^2).
\end{aligned}$$

Choosing k_2 so that $1 + k_2\varepsilon \le (1 + k_1\varepsilon)^{1/2}$ for $\varepsilon \in [0, 1]$ and k_3 so that $k_3\varepsilon^2$ is a bound for the $O(\varepsilon^2)$ term, (iv) is shown. Finally, (v) follows from (iv) by letting $\varepsilon \to 0$. □

4. Proof of Theorem 1: Probabilistic part. As mentioned at the beginning of the previous section, the probabilistic part of the argument breaks into two lemmas. The first lemma shows that the particle always strays far enough from $\mathbf{p}$. To make the notation briefer, let S_n denote $\eta(\mathbf{v}(n))$ and let $X_n = S_n - S_{n-1}$ be the difference sequence.

LEMMA 1. *Assume the hypotheses of Theorem* 1 *hold for all* $\mathbf{v}(n)$, *not just for* $\mathbf{v}(n) \in \mathscr{N}$ *and n large. Then there is a constant* c_5 *determined by* $\mathbf{F}$ *and* c_1, c_2, c_3 *and* c_4 *such that for sufficiently large n*,

$$\mathbf{P}\Big(\sup_{k \ge n} S_k > c_5 n^{1/2-\gamma} \mid \mathscr{F}_n\Big) \ge 1/2.$$

PROOF. Fix n and let $\tau = \inf\{k \ge n: S_k > c_5 n^{1/2-\gamma}\}$. We will calculate the variance of S stopped at τ. On the one hand, this is limited by the fact that conditions (ii) and (iv) of Theorem 1 prohibit large jumps, so S_k is never much more than $c_5 n^{1/2-\gamma}$ for $k \le \tau$. On the other hand, condition (iii) forces the variance to keep up a certain minimum growth on the order of $n^{-2\gamma}$ until the stopping time is reached. The accumulated variance is the tail sum of these order $n^{-2\gamma}$ terms, so it is of the order $n^{1-2\gamma}$. When c_5 is small enough, these two facts together imply that the stopping time is reached often enough for Lemma 1 to hold. To calculate how $\mathbf{E}(S^2_{\tau \wedge M} \mid \mathscr{F}_n)$ increases with M, fix any $M \ge n$ and calculate

$$\begin{aligned}
&\mathbf{E}\big(S^2_{\tau \wedge (M+1)} \mid \mathscr{F}_n\big) - \mathbf{E}\big(S^2_{\tau \wedge M} \mid \mathscr{F}_n\big) \\
&\quad = \mathbf{E}\big(\mathbf{1}_{\tau > M}\big(2X_{M+1}S_M + X^2_{M+1}\big) \mid \mathscr{F}_n\big) \\
&\quad = \mathbf{E}(\mathbf{E}(\mathbf{1}_{\tau > M} 2X_{M+1}S_M \mid \mathscr{F}_M) \mid \mathscr{F}_n) \\
&\qquad + \mathbf{E}\big(\mathbf{E}\big(\mathbf{1}_{\tau > M} X^2_{M+1} \mid \mathscr{F}_M\big) \mid \mathscr{F}_n\big).
\end{aligned} \tag{11}$$

Each piece of the quantity in (11) can now be estimated. Since $\mathbf{1}_{\tau > M}$ is $\mathscr{F}_M$-measurable, it can be pulled out. To estimate the piece $\mathbf{E}(X_{M+1}S_M | \mathscr{F}_M)$, calculate

$$\begin{aligned}
\mathbf{E}(X_{M+1} | \mathscr{F}_M)& \\
&= \mathbf{E}(\eta(\mathbf{v}(M+1)) - \eta(\mathbf{v}(M)) | \mathscr{F}_M) \\
&= \mathbf{E}\big(D^{1/2}(\Phi(\mathbf{v}(M+1)) - \mathbf{p}) | \mathscr{F}_M\big) - S_M \\
&\geq D^{1/2}\big[\mathbf{E}(\Phi(\mathbf{v}(M+1)) - \mathbf{p} | \mathscr{F}_M)\big] - S_M \quad \text{by convexity of } D^{1/2} \\
&= D^{1/2}\Big[\mathbf{E}\big(\Phi(\mathbf{v}(M)) - \mathbf{p} + \mathbf{D}_{\mathbf{v}(M)}\Phi(\mathbf{v}(M+1) - \mathbf{v}(M)) \\
&\qquad\qquad + O(|\mathbf{v}(M+1) - \mathbf{v}(M)|^2) | \mathscr{F}_M\big)\Big] - S_M \\
&= D^{1/2}\big[\Phi(\mathbf{v}(M)) - \mathbf{p} + \mathbf{D}_{\mathbf{v}(M)}(\Phi)\mathbf{E}(\mathbf{v}(M+1) - \mathbf{v}(M) | \mathscr{F}_M)\big] \\
&\quad + O\big(\mathbf{E}(|\mathbf{v}(M+1) - \mathbf{v}(M)|^2 | \mathscr{F}_M)\big) - S_M \\
&\qquad\qquad \text{by linearity of } \mathbf{D}_{\mathbf{v}(M)}(\Phi) \text{ and because } D^{1/2} \text{ is Lipschitz} \\
&= D^{1/2}\big[\Phi(\mathbf{v}(M)) - \mathbf{p} + \mathbf{D}_{\mathbf{v}(M)}(\Phi)(a_M \mathbf{F}(\mathbf{v}(M)))\big] + O(M^{-2\gamma}) - S_M \\
&= D^{1/2}\big[\Phi(\mathbf{v}(M) + a_M\mathbf{F}(\mathbf{v}(M))) - \mathbf{p} + O(a_M\mathbf{F}(\mathbf{v}(M)))^2\big] + O(M^{-2\gamma}) \\
&\qquad\qquad \text{because } \Phi(\mathbf{v}) + \mathbf{D}_{\mathbf{v}}\Phi(\mathbf{w}) = \Phi(\mathbf{v} + \mathbf{w}) + O(|\mathbf{w}|^2) \\
&= D^{1/2}\big[\Phi(\mathbf{v}(M) + a_M\mathbf{F}(\mathbf{v}(M))) - \mathbf{p}\big] + O(M^{-2\gamma}) - S_M \\
&= \eta(\mathbf{v}(M) + a_M\mathbf{F}(\mathbf{v}(M))) + O(M^{-2\gamma}) - S_M \\
&\geq k_2 a_M S_M + O(M^{-2\gamma}) \quad \text{by part (iv) of Proposition 3.}
\end{aligned}$$

Therefore the estimate for this piece of (11) is

$$\mathbf{E}(2X_{M+1}S_M | \mathscr{F}_M) \geq k_2 a_M S_M^2 + O(M^{-2\gamma}S_M). \tag{12}$$

The last piece, $\mathbf{E}(X_{M+1}^2 | \mathscr{F}_M)$, is on the order of $M^{-2\gamma}$. The calculation begins in the same way as for the first piece: Convexity of $D^{1/2}$ and smoothness of Φ imply

$$\begin{aligned}
X_{M+1} &= \eta(\mathbf{v}(M+1)) - \eta(\mathbf{v}(M)) \\
&\geq \mathbf{D}_{\mathbf{v}(M)}(\eta)(\mathbf{v}(M+1) - \mathbf{v}(M)) + O(|\mathbf{v}(M+1) - \mathbf{v}(M)|^2) \qquad (13) \\
&= \mathbf{D}_{\mathbf{v}(M)}(\eta)(a_M\mathbf{F}(\mathbf{v}(M)) + \xi_M) + O(|\mathbf{v}(M+1) - \mathbf{v}(M)|^2),
\end{aligned}$$

where the gradient of η at a point on the stable manifold $\mathscr{M}$, where $\eta = D^{1/2} = 0$ is singular, is taken to be the limit of $\mathbf{D}\eta$ along any approaching path. (Geometrically, this means taking a support hyperplane at a place where the tangent plane is not unique, but the inequality holds by convexity.) Let θ be a unit vector in the direction of the gradient of η; since the quadratic form D is nondegenerate at $\mathbf{0}$, $\mathbf{D}\eta(\mathbf{v})$ is at least some constant c_6 times $\mathbf{v} \cdot \theta$. Applying hypothesis (iii) of Theorem 1 together with (v) of Proposition 3 and (13) gives

$$\begin{aligned}
&\mathbf{E}(X_{M+1}^{+} \mid \mathscr{F}_M) \\
&\quad \geq \mathbf{E}\Big(\big(\mathbf{D}_{\mathbf{v}(M)}(\eta)(a_M \mathbf{F}(\mathbf{v}(M)) + \xi_M) + O(M^{-2\gamma})\big)^{+} \mid \mathscr{F}_M\Big) \\
&\quad \geq \mathbf{E}\Big(\big(\mathbf{D}_{\mathbf{v}(M)}(\eta)(\xi_M) + O(M^{-2})\big)^{+} \mid \mathscr{F}_M\Big) \\
&\quad = \mathbf{E}\big((c_6 \theta \cdot \xi)^{+} \mid \mathscr{F}_M\big) \\
&\quad \geq c_3 c_6 / M^{\gamma} + O(M^{-2\gamma}).
\end{aligned}$$

It follows that for const. $< (c_3 c_6)^2$ and large enough M,

$$\mathbf{E}\big(X_{M+1}^2 \mid \mathscr{F}_M\big) \text{ is at least const. } M^{-2\gamma}. \tag{14}$$

Now both summands of the quantity in (11) have been calculated in (12) and (14). For any constant c, if $S_M \leq c/M^{\gamma}$ then for large M the summand in (14) dominates and hence

$$\mathbf{E}\big(2X_{M+1}S_M + X_{M+1}^2 \mid \mathscr{F}_M\big) \geq \text{const.}/M^{2\gamma}.$$

On the other hand, if $S_M > c/M^{\gamma}$ then for large M the order $M^{-2\gamma}$ term in (14), which is positive, has at least the magnitude of the order $M^{-2\gamma}S_M$ term in (12) so the sum is at least as big as the first term in (12) and again

$$\mathbf{E}\big(2X_{M+1}S_M + X_{M+1}^2 \mid \mathscr{F}_M\big) \geq \text{const.}/M^{2\gamma}.$$

Substituting this back in (11) gives

$$\begin{aligned}
&\mathbf{E}\big(S_{\tau \wedge (M+1)}^2 \mid \mathscr{F}_n\big) - \mathbf{E}\big(S_{\tau \wedge M}^2 \mid \mathscr{F}_n\big) \\
&\quad \geq \mathbf{E}\left(\mathbf{1}_{\tau > M} \frac{\text{const.}}{M^{2\gamma}} \,\middle|\, \mathscr{F}_n\right) \\
&\quad \geq \frac{\text{const.}}{M^{2\gamma}} \mathbf{P}(\tau = \infty \mid \mathscr{F}_n).
\end{aligned}$$

Then by induction

$$\begin{aligned}
&\mathbf{E}\big(S_{M \wedge \tau}^2 \mid \mathscr{F}_n\big) \\
&\quad \geq S_n^2 + \text{const.} \cdot \mathbf{P}(\tau = \infty \mid \mathscr{F}_n) \sum_{i=n}^{M-1} \frac{1}{i^{2\gamma}} \\
&\quad \geq \text{const.} \cdot \mathbf{P}(\tau = \infty \mid \mathscr{F}_n) \left(\frac{1}{n^{2\gamma-1}} - \frac{1}{M^{2\gamma-1}}\right).
\end{aligned}$$

But by hypotheses (ii) and (iv) of Theorem 1 and the fact that η is Lipschitz, $S_{\tau \wedge M} < c_5 n^{1/2-\gamma} + \text{const.}/n^{\gamma}$, so for n sufficiently large $c_5 n^{1/2-\gamma} > \text{const.}/n^{\gamma}$, so $2c_5 n^{1/2-\gamma} > S_{\tau \wedge M}$ and therefore

$$\frac{4c_5^2}{n^{2\gamma-1}} \geq \mathbf{E}\big(S_{M \wedge \tau}^2 | \mathscr{F}_n\big) \geq \text{const.} \cdot \mathbf{P}(\tau = \infty | \mathscr{F}_n)\left(\frac{1}{n^{2\gamma-1}} - \frac{1}{M^{2\gamma-1}}\right)$$

and letting $M \to \infty$, $\mathbf{P}(\tau = \infty | \mathscr{F}_n)$ is bounded by a constant times c_5^2. This can be made smaller than $\frac{1}{2}$ by choosing c_5 small enough, in which case

$$\mathbf{P}\big(\sup_{k \geq n} |S_k - p| > c_3 n^{1/2-\gamma} | \mathscr{F}_n\big) = 1 - \mathbf{P}(\tau = \infty | \mathscr{F}_n) \geq 1/2. \qquad \square$$

The second lemma shows that the particle fails to come back near $\mathbf{p}$ with probability bounded away from 0.

Lemma 2. *Again assume the hypotheses of Theorem* 1 *hold for all* $\mathbf{v}(n)$, *not just for* $\mathbf{v}(n) \in \mathscr{N}$ *and n large. Then there is a constant* $a > 0$ *for which*

$$\mathbf{P}\left(\inf_{k \geq n} S_k \geq \frac{c_5 n^{1/2-\gamma}}{2} \,\middle|\, \mathscr{F}_n, S_n \geq c_5 n^{1/2-\gamma}\right) \geq a.$$

Proof. The idea this time is that the variance of the variables $\{X_k\colon k \geq n\}$ is not enough to give a high probability of getting back under $c_5 n^{1/2-\gamma}/2$. The inequality used is a one-sided Tchebycheff inequality relying on the fact that the expectation of S_n is increasing, so if S_n has a probability of $1 - \varepsilon$ of coming back under $c_5 n^{1/2-\gamma}/2$, then ε of the time it must grow to the order of ε^{-1}, contributing to the variance on the order of ε^{-2}.

So assume that $S_n > c_5 n^{1/2-\gamma}$. Let $\tau = \inf\{k \geq n\colon S_k \leq c_5 n^{1/2-\gamma}/2\}$. The claim is that $\mathbf{P}(\tau = \infty) \geq a$. Define a sequence of variables $Y_{n+1}, Y_{n+2}, \dots$ by $Y_k = 0$ for $k > \tau$ and $Y_k = X_k - \mu_k$ for $\tau \geq k > n$, where $\mu_k = \mathbf{E}(X_k | \mathscr{F}_{k-1})$. What is important is that μ_k is positive. This is an easy consequence of the estimate (12), plugging in the fact that $S_k > c_5 n^{1/2-\gamma}/2$. Then the sequence $\{Z_k\colon k > n\}$, where $Z_k = S_n + \sum_{j=n+1}^{k} Y_j$, defines a martingale such that $Z_k \leq S_{k \wedge \tau}$ and hence S_k never gets below $c_5 n^{1/2-\gamma}/2$ for $k > n$ as long as Z_k never does. Since its increments are $O(k^{-\gamma})$, $\{Z_k\}$ is L^2-bounded and the variable Z_∞ appearing in the calculations below is well defined almost surely. It suffices to find a constant $a > 0$ for which

$$\mathbf{P}\left(\sum_{k=n+1}^{\tau} Y_k \leq -c_5 n^{1/2-\gamma}/2\right) \leq 1 - a. \tag{15}$$

Again using the fact that the increments Y_k are bounded by c/k^{γ} for some constant c,

$$\operatorname{Var}\left(\sum_{k=n+1}^{\tau} Y_k\right) \leq \sum_{k=n+1}^{\infty} \frac{c}{n^{2\gamma}} \leq \frac{c}{n^{2\gamma-1}}. \tag{16}$$

Also

$$
\begin{aligned}
&\operatorname{Var}\left(\sum_{k=n+1}^{\tau} Y_k\right)\\
&\quad\geq \mathbf{P}(\tau<\infty)\left(-c_5 n^{1/2-\gamma}/2\right)^2+\mathbf{P}(\tau=\infty)\mathbf{E}\left((Z_\infty-S_n)^2|\tau=\infty\right)\\
(17)\qquad&\quad\geq \mathbf{P}(\tau=\infty)(\mathbf{E}(Z_\infty-S_n|\tau<\infty))^2\\
&\quad\geq \mathbf{P}(\tau=\infty)\left(\frac{c_5 n^{1/2-\gamma}}{2}\frac{\mathbf{P}(\tau<\infty)}{\mathbf{P}(\tau=\infty)}\right)^2\\
&\quad=\frac{c_5^2}{4n^{2\gamma-1}}\frac{\mathbf{P}(\tau<\infty)^2}{\mathbf{P}(\tau=\infty)},
\end{aligned}
$$

where the penultimate term is calculated from the fact that $\mathbf{E}(Z_\infty|\tau<\infty) < -c_5 n^{1/2-\gamma}/2$ while $\mathbf{E}(Z_\infty)$ must be 0. Combining inequalities (16) and (17) gives

$$
\frac{\mathbf{P}(\tau=\infty)}{\mathbf{P}(\tau<\infty)^2}\geq\frac{c_5^2}{4c}
$$

so either $\mathbf{P}(\tau=\infty)\geq 1/2$ or $\mathbf{P}(\tau<\infty)\leq (c_5^2/4c)(1/2)^2 = c_5^2/16c$. Letting $a = \min\{1/2, c_5^2/16c\}$ finishes the proof. □

Now an easy tail argument finishes the proof of Theorem 1. Suppose that $\mathbf{P}(\mathbf{v}(n)\to\mathbf{p})>0$. Then there is some neighborhood $\mathscr{N}$ of $\mathbf{p}$, some n and some event $\mathscr{A}\in\mathscr{F}_n$ for which the probability that $\mathbf{v}(n)$ converges to $\mathbf{p}$ and never leaves $\mathscr{N}$ after time n is greater than $1-a/2$. In fact n can be chosen arbitrarily large and $\mathscr{N}$ arbitrarily small, and in particular so that $\mathscr{N}_1\subseteq\mathscr{N}$, where $\mathscr{N}$ is as is the hypothesis of Theorem 1. Couple the process $\{\mathbf{v}(k)\}$ to a process $\{\mathbf{v}'(k)\colon k\geq n\}$ so that $\mathbf{P}(\mathbf{v}(k)=\mathbf{v}'(k)|\mathscr{A})=1$ for $n\leq k\leq$ the exit time from $\mathscr{N}_1$ after n, and such that $\mathbf{v}'(n)$ satisfies (i)–(iv) of Theorem 1 after time n for all $\mathbf{v}'(n)$, not just locally. By the coupling, $\mathbf{v}'(n)$ also converges to $\mathbf{p}$ with probability at least $1-a/2$ given $\mathscr{A}$. But Lemmas 1 and 2 together imply that the probability of failing to converge to $\mathbf{p}$ is at least $a/2$ conditioned on any set in $\mathscr{F}_n$. By contradiction, the probability of convergence to $\mathbf{p}$ must be 0.

Acknowledgments. I would like to thank Nancy Kopell for substantial help with the geometry, and Rich Montgomery and Charles Pugh for further help with the geometry.

REFERENCES

Arthur, B., Ermol'ev, Y. and Kaniovskii, Y. (1983). A generalized urn problem and its applications. *Cybernetics* **19** 61–71.

Athreya, K. (1969). On a characteristic property of Pólya's urn. *Studia Sci. Math. Hungar.* **4** 31–35.

Hill, B., Lane, D. and Sudderth, W. (1980). A strong law for some generalized urn processes. *Ann. Probab.* **8** 214–226.

HIRSCH, M., PUGH, C. and SHUB, M. (1977). *Invariant Manifolds. Lecture Notes in Math.* **583**. Springer, New York.

HIRSCH, M. and SMALE, S. (1974). *Differential Equations, Dynamical Systems and Linear Algebra*. Academic, New York.

NEVEL'SON, M. and HASMINSKII, R. (1973). Stochastic approximation and recursive estimation. *Trans. Math. Monographs* **47**. Amer. Math. Soc., Providence, R.I.

PEMANTLE, R. (1988). Random processes with reinforcement. Ph.D. dissertation, Massachusetts Inst. Technology.

PEMANTLE, R. (1989). A time-dependent version of Pólya's urn. Technical Report 192, Dept. Statistics, Univ. California, Berkeley.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853