INTRODUCTION TO AVERAGING.

1. Averaging in Markov Chains.

1.1. **Statement.** Consider a recurrence

$$x^\varepsilon_{n+1}-x^\varepsilon_n=\varepsilon F(x^\varepsilon_n,\omega_n),\quad x_0=a \tag{1}$$

where $x\in\mathbb{R}^d$, $F=(F_{(1)},F_{(2)},\dots F_{(d)})$. We assume that $\{F(\cdot,\omega_n)\}$ are independent identically distributed functions, that F takes values in a ball of fixed radius (independent of x) and that for each $k_1,k_2\dots k_d$ the map $x\to\mathbb{E}(\prod_{l=1}^d F_{(l)}^{k_l}(x,\omega))$ is smooth and has two bounded derivatives. Denote

$$\bar F(x)=\mathbb{E}(F(x,\omega)),$$

$$D_{\alpha\beta}(x)(x)=\mathbb{E}\left(\left(F_{(\alpha)}(x,\omega)-\bar F_{(\alpha)}(x)\right)\left(F_{(\beta)}(x,\omega)-\bar F_{(\beta)}(x)\right)\right).$$

Let $y(t)$ denote the solution of

$$\dot y=\bar F(y),\quad y(0)=a.$$

Fix $T>0$.

Theorem 1. [16, 12] *(a) For each $\delta>0$ we have*

$$\mathbb{P}\left(\max_{[0,T]}\left|x_{[t/\varepsilon]}-y(t)\right|>\delta\right)\to 0 \text{ as } \varepsilon\to 0.$$

(b) Moreover define $Z^\varepsilon(t)$ by letting $Z^\varepsilon(\varepsilon n)=\frac{1}{\sqrt{\varepsilon}}[x_n-y(n\varepsilon)]$ and interpolating linearly in between. Then as $\varepsilon\to 0$ $Z^\varepsilon(t)$ converges to a Gaussian random process $Z(t)$. Z(t) satisfies the following equation

$$Z(t)=W(t)+\int_0^t D\bar F(y(s))Z(s)ds,\quad Z(0)=0 \tag{2}$$

where $W(t)$ is a Gaussian Markov process starting at 0 with independent increments, zero mean and covariance

$$\mathbb{E}(W_\alpha(t)W_\beta(t))=\int_0^t D_{\alpha\beta}(y(s))ds. \tag{3}$$

(c) For each $1\le R\le c(\sqrt{\varepsilon})^{-1}$ there exist constants C_1,C_2 such that

$$\mathbb{P}\left(Z^\varepsilon(t)>R\sqrt{\varepsilon}\right)\le C_1\exp(-C_2R^2).$$

1.2. **Proof of Theorem 1.** The proof consists of several steps. Let y_n be a sequence satisfying the following recursion

$$y_{n+1} - y_n = \varepsilon \bar{F}(y_n), \quad y_0 = a.$$

Then y_n differs from $y(n\varepsilon)$ by terms of order ε, so it suffices to study $z_n = x_n - y_n$. We have

$$z_{n+1} - z_n = \varepsilon[F_n(x, \omega_n) - \bar{F}_n(x)] + \varepsilon[\bar{F}(x_n) - \bar{F}(y_n)].$$

Denote

$$\xi_n = F(x_n, \omega_n) - \bar{F}(x_n), \quad w_n = \sum_{j=1}^{n} \xi_j, \quad w_0 = 0.$$

Lemma 1.1.

$$(a) \quad |z_n| \leq \text{Const} \max_{k \leq n} |w_k| \varepsilon.$$

$$(b) \quad |z_{n_2} - z_{n_1}| \leq \text{Const} \max_{k \leq n_2 - n_1} |w_{n_1+k} - w_{n_1}| \, \varepsilon.$$

Proof. We prove (a). (b) is similar except for notational complications. Conditions of Theorem 1 and Hadamard Lemma imply that

$$\bar{F}(x_n) - \bar{F}(y_n) = A(y_n, z_n) z_n$$

where $A(y, z)$ is a bounded matrix valued function. Denote $A_n = A(y_n, z_n)$. Then

$$z_{n+1} - z_n = \varepsilon \xi_n + \varepsilon A_n z_n.$$

This equation is linear so it can be solved. Namely let B_n satisfy

$$B_{n+1} - B_n = \varepsilon A_n B_n, \quad B_0 = 1,$$

then B_n and B_n^{-1} are uniformly bounded for $n \leq T/\varepsilon$. Substitute $z_n = B_n r_n$, then

$$r_n = \varepsilon \sum_{j=0}^{n-1} B_{j+1}^{-1} \xi_j = \tag{4}$$

$$\varepsilon \sum_{j=0}^{n-1} B_{j+1}^{-1}(w_j - w_{j-1}) = \varepsilon \sum_{j=0}^{n-1} (B_{j+1}^{-1} - B_{j+2}^{-1}) w_j + \varepsilon B_{n+1} w_{n-1} = I + I\!I.$$

Since

$$||B_{j+1}^{-1} - B_{j+2}^{-1}|| \leq \text{Const} ||B_{j+2} - B_{j+1}|| \leq \text{Const}\varepsilon$$

we get

$$|I| \leq \text{Const}\varepsilon^2 \left| \sum_{j=0}^{n} w_j \right| \leq \text{Const}(\varepsilon n) \left(\max_{0 \leq k \leq n-1} |w_k| \right) \varepsilon$$

Since $\varepsilon n \le T$ we have the required bound on the first term. On the other hand $|I\!I| \le \text{Const}\varepsilon \max_{0\le k\le n-1}|w_k|$. Since $|z_n| \le \text{Const}|r_n|$ the lemma follows. □

We now give some a priori bounds for z_n. Define the process $W^\varepsilon(t)$ by $W^\varepsilon(n\varepsilon) = \sqrt{\varepsilon}w_n$ with linear interpolation in between.

Lemma 1.2. *$\{W^\varepsilon(t)\}$ is tight.*

Proof. Since $W^\varepsilon(0) = 0$ it is enough to show that for all $[t_1, t_2]$ we have

$$\mathbb{E}(|W^\varepsilon(t_1) - W^\varepsilon(t_2)|^4) \le \text{Const}|t_1 - t_2|^2.$$

In terms of w_n we have to show that

$$\mathbb{E}(|w_{n_1} - w_{n_2}|^4) \le \text{Const}|n_1 - n_2|^2.$$

Therefore lemma 1.2 follows from Lemma 1.3 below. □

Lemma 1.3. *Let $n_2 > n_1$. Denote*

$$S = S_{n_1,n_2} = w_{n_2} - w_{n_1} = \sum_{j=n_1+1}^{n_2} \xi_j.$$

Then

$$(a) \quad \mathbb{E}(S) = 0.$$

$$(b) \quad \mathbb{E}(S_{(\alpha)}S_{(\beta)}) = \sum_{j=n_1+1}^{n_2} \mathbb{E}(D_{\alpha\beta}).$$

In particular,

(c) $\mathbb{E}(|S|^2) \le \text{Const}(n_2 - n_1)$.

(d) $\mathbb{E}(|S|^4) \le \text{Const}(n_2 - n_1)^2$.

Proof. (a) is clear from the definition of $\bar{F}$. Next,

$$\mathbb{E}(S_{(\alpha)}S_{(\beta)}) = \sum_{j_1,j_2=n_1+1}^{n_2} \mathbb{E}\left(\left[F_{(\alpha)}(x_{j_1},\omega_{j_1}) - \bar{F}_{(\alpha)}(x_{j_1})\right]\left[F_{(\beta)}(x_{j_2},\omega_{j_2}) - \bar{F}_{(\beta)}(x_{j_2})\right]\right).$$

Now by Markov property the terms with $j_2 \ne j_1$ give 0. This proves (b). (c) follows from (b). To prove (d) consider $\mathbb{E}(|S_{(1)}|^4)$, the other components are similar. We have

$$\mathbb{E}(|S_{(1)}|^4) = \sum_{j_1j_2j_3j_4} \mathbb{E}\left([F_{(1)}(x_{j_1},\omega_{j_1}) - \bar{F}_{(1)}(x_{j_1})][F_{(1)}(x_{j_2},\omega_{j_2}) - \bar{F}_{(1)}(x_{j_2})]\times\right.$$
$$\left.[F_{(1)}(x_{j_3},\omega_{j_3}) - \bar{F}_{(1)}(x_{j_3})][F_{(1)}(x_{j_4},\omega_{j_4}) - \bar{F}_{(1)}(x_{j_4})]\right).$$

Order the indices so that $j_1 \leq j_2 \leq j_3 \leq j_4$. As before all terms where $j_3 \neq j_4$ vanish, so we only need to consider contribution of the terms with $j_4 = j_3$. Among those there are at most $(n_2 - n_1)^2$ terms with $j_2 = j_3 = j_4$. So we have

$$\mathbb{E}(S^4_{(1)}) =\leq \text{Const}\left[(n_2 - n_1)^2 + \right.$$

$$\left. \sum_{n_1 < j_1 \leq j_2 < m \leq n_2} \mathbb{E}\left([F_{(1)}(x_{j_1}, \omega_{j_1}) - \bar{F}_{(1)}(x_{j_1})][F_{(1)}(x_{j_2}, \omega_{j_2}) - \bar{F}_{(1)}(x_{j_2})][F_{(1)}(x_m, \omega_m) - \bar{F}_{(1)}(x_m)]^2\right)\right]$$

The second term here equals

$$\sum_m (\mathbb{E}\left([F_{(1)}(x_m, \omega_m) - \bar{F}_{(1)}(x_m)]^2 S^2_{n_1,m}\right) \leq \sum_m (\mathbb{E}\left(S^2_{n_1,m}\right)$$

By part (c) the last expression is bounded by

$$\text{Const} \sum_{m=n_1+1}^{n_2} (m - n_1) \leq \text{Const}(n_2 - n_1)^2. \quad \square$$

Lemma 1.2 implies Theorem 1(a). To get (b) we need to refine it.

Lemma 1.4. *As $\varepsilon \to 0$ $W^\varepsilon(t)$ converges to $W(t)$–the process given by (3).*

Proof. By Lemma 1.2 it is enough to establish convergence of finite dimensional distributions. We begin with one dimensional ones. So fix $0 \leq t \leq T$. Fix $\delta > 0$. Divide interval $[0, t]$ into subintervals of length Δt. Let G_m be a sequence of independent Gaussian random variables with zero mean and unit covariance. Denote $n_m = m\Delta t/\varepsilon$ and let σ_m be a matrix such that

$$\sigma_m^2 = \int_{m\Delta/\varepsilon}^{(m+1)\Delta/\varepsilon} D(y(s))ds.$$

Define $\Sigma^\varepsilon_m = \sum_{j=n_m+1}^{n_{m+1}} z_j$, and

$$\tilde{\Sigma}^\varepsilon_m = \begin{cases} \Sigma^\varepsilon_m & \text{if } z_{n_m} < \varepsilon^{-1/2+\delta} \\ \frac{1}{\sqrt{\varepsilon}}\sigma_m G_m & \text{otherwise} \end{cases}.$$

Then by Lemma 1.2

$$\mathbb{P}\left(W^\varepsilon(t) \neq \sum_m \tilde{\Sigma}^\varepsilon_m\right) \to 0 \text{ as } \varepsilon \to 0.$$

Take some $\eta \in \mathbb{R}^d$. Lemma 1.3 gives $\mathbb{E}(\tilde{\Sigma}^\varepsilon_m) = 0$ and $\mathbb{E}((\eta\tilde{\Sigma}^\varepsilon_m)^2)$ is either $\frac{||\sigma_m \eta||^2}{\varepsilon}$ or

$$\sum_{j=n_m+1}^{n_{m+1}} < D(x_j)\eta, \eta >= \sum_{j=n_m+1}^{n_{m+1}} < D(y_j)\eta, \eta > +o(\Delta/\varepsilon) = \frac{||\sigma_m \eta||^2}{\varepsilon} + o(\Delta/\varepsilon)$$

depending on the value of x_{n_m}. This gives

$$\mathbb{E}(\exp(i < \eta, \tilde{\Sigma}^\varepsilon_m > /\sqrt{\varepsilon})|\mathcal{F}_{n_m}) = 1 - \frac{||\sigma_m \eta||^2 \Delta}{2} + \zeta_m \Delta \tag{5}$$

where $\zeta_m \to 0$ uniformly in m as $\varepsilon \to 0, \Delta \to 0$. Iterating (5) we get

$$\ln \mathbb{E}(\exp(i < \eta, W^\varepsilon(t) > /\sqrt{\varepsilon})) \sim -\frac{1}{2} \sum_m ||\sigma_m \eta||^2 + o_{\varepsilon\to 0, \Delta \to 0}(1) =$$

$$-\frac{1}{2} \int_0^t < D(y(s))\eta, \eta > ds + o_{\varepsilon\to 0, \Delta \to 0}(1).$$

Since Δ is arbitrary we can make $\Delta \to 0$ obtaining

$$\mathbb{E}(\exp(i < \eta, W(t) >) = -\frac{1}{2} \int_0^t < D(y(s))\eta, \eta > ds.$$

A similar computation shows that for all $t_1 \leq t_2 \leq \dots \leq t_k$, $\eta_1, \eta_2 \dots \eta_k$ we have

$$\mathbb{E}(\exp(i \sum_{l=1}^k < \eta_k W(t_k) >)) = \exp\left(-\frac{1}{2} \int_0^{t_k} < D(y(s))\tilde{\eta}(s), \tilde{\eta}(s) > ds\right),$$

where $\tilde{\eta}(s) = \sum_{t_l > s} \eta_l$. The lemma follows. □

We now return to the equation for z_n. Using again Hadamard lemma we get

$$z^\varepsilon_{n+1} - z^\varepsilon_n = \varepsilon \xi_n + \varepsilon D\bar{F}(y_n) z^\varepsilon_n + \varepsilon Q(y_n, z_n)(z_n, z_n)$$

where $Q(y, z)$ is a bounded quadratic form. Thus

$$z^\varepsilon_{n+1} - z^\varepsilon_n = \varepsilon \xi_n + \varepsilon D\bar{F}(y_n) z^\varepsilon_n + \varepsilon \rho_n$$

where

$$|\rho_n| \leq \text{Const}|z_n|^2 \leq \text{Const}\varepsilon \left| \frac{z_n}{\sqrt{\varepsilon}} \right|^2.$$

Therefore

$$z_n = \varepsilon \sum_{j=0}^n \xi_j + \varepsilon \sum_{j=0}^n D\bar{F}(y_j) z_j + \varepsilon \sum_{j=0}^n \rho_j \tag{6}$$

Define $Z^\varepsilon(t)$ by the condition $Z^\varepsilon(n\varepsilon)=\frac{1}{\sqrt{\varepsilon}}z_n$ with linear interpolation in between. By Lemmas 1.1 and 1.2, $Z^\varepsilon(t)$ is tight. Also by Lemma 1.2

$$\varepsilon \sup_{n\leq T/\varepsilon}\sum_{j=0}^{n}\rho_j \to 0$$

in probability. Choose a subsequence $\varepsilon_k\to 0$ such that $Z^\varepsilon_k(t)$ converges to some random process $Z(t)$. Then (6) implies that

$$Z(t)=W(t)+\int_0^t D\bar{F}(y(s))Z(s)ds.$$

This implies (b).

To prove (c) it is enough to estimate $\mathbb{P}(w_n>R\sqrt{n})$. Take $k=\sqrt{n}/R$ then

$$\mathbb{E}\left(\exp\left(\frac{\delta}{k}[w_{k(1)}-1]\right)\right)\approx 1-\frac{\delta}{k}+O\left(\frac{\delta^2}{k}\right).$$

This implies that there is a constant c such that

$$\mathbb{E}\left(\exp\left(\frac{\delta}{k}[w_{k(1)}-1]\right)\right)\leq\left(1-\frac{c}{k}\right).$$

By induction

$$\mathbb{E}\left(\exp\left(\frac{\delta}{k}[w_{km(1)}-m]\right)\right)\leq\left(1-\frac{c}{k}\right)^m.$$

Taking $m=n/k$ we obtain

$$\mathbb{E}\left(\exp\left(\frac{\delta}{k}[w_{n(1)}-n^{1/2+\gamma}]\right)\right)\leq C_1\exp(-C_2\frac{m}{k})=C_1\exp(-C_2R^2).$$

Similar estimates are valid for other coordinates. This proves (c). Theorem 1 is established. □

Exercise 1. *(a) Suppose that z^ε_n satisfies the relation*

$$z_{n+1}-z_n=\varepsilon D\bar{F}(y_n)z_n+\sqrt{\varepsilon}\zeta_n,\quad z_0=0$$

where ζ_n are independent Gaussians with variance $D(y_n)$. Show that as $\varepsilon\to 0$ $z_{[t\varepsilon]}\to Z(t)$ satisfying (2).

(b) Use (a), (4) and formulas (39) and (40) of Appendix A to show that if $\Gamma(s,t)$ satisfies

$$\frac{d\Gamma}{dt}=D\bar{F}(y(t))\Gamma,\quad \Gamma(s,s)=\mathrm{id}$$

then for $t_1\leq t_2$

$$\mathbb{E}(Z_{(\alpha)}(t_1)Z_\beta(t_2))=\int_0^{t_1}\Gamma(s,t_1)D(y(s))\Gamma(s,t_1)^*ds.$$

Exercise 2. *(a) Generalize Theorem 1 for the recurrence*

$$x^\varepsilon_{n+1} - x^\varepsilon_n = \varepsilon F(x^\varepsilon_n, \omega_n) + \varepsilon^2 G(x^\varepsilon_n, \omega_n, \varepsilon), \quad x_0 = a \tag{7}$$

if G satisfies the same assumaptions as F.

(b) Show that under coordinate changes (1) is transformed to (7).

(c) Let M be a compact manifold. Consider equation $\dot{x^\varepsilon} = \varepsilon X(x, \omega_t)$, where $X(\dot{;}\omega_t)$ is constant on $[n, n+1)$ and the values at different intervals are independent and identically distributed. Show that it $x \to X(x, \omega)$ is uniformly bounded togather with the first three derivatives then $x^\varepsilon(t/\varepsilon)$ converges weakly to $y(t)$ satisfying the equation $\dot{y} = \mathbb{E}(X(y, \omega)$. Show moreover that $(\exp^{-1}_{y(t)}(x(t)))/\sqrt{\varepsilon}$ converges to a Gaussian process and that for $R \geq 1$

$$\mathbb{P}(\text{dist}(x(t), y(t)) \leq \sqrt{\varepsilon} R) \leq C_1 e^{-C_2 R^2}.$$

Exercise 3. *Consider the recurence*

$$x^\varepsilon_{n+1} - x^\varepsilon_n = \varepsilon F(x^\varepsilon_n, \omega_n)$$

$$v^\varepsilon_{n+1} - v^\varepsilon_n = \varepsilon DF(x_n, \omega_n) v_n$$

so that

$$v_n = \frac{\partial x_n}{\partial x_0} v_0.$$

Show that the averaged equation takes form

$$\dot{x} = \bar{F}(x) \quad \dot{v} = D\bar{F}(x) v.$$

1.3. **Urn models.** Consider the following process. An urn initially contains b_0 black and w_0 white balls. Fix some k, $m > k/2$. At each step k balls are drawn and returned to the urn. Also, if there are at least m balls of a particular color among the draw then a new ball of this color is added to the urn. Let b_n and w_n be the numbers of black and white balls after the nth ball is added to the urn. We want to understand what happens when $n \to \infty$. To reduce to the setting of Theorem 1 let consider who the number of balls change between N and $2N$. Denote $\varepsilon = 1/N$, $x_n = b_{n+N}/N$, $y_n = w_{n+N}/N$. Then $x_{n+1} - x_n = \varepsilon$ with probability

$$\frac{P(x,y)}{P(x,y) + P(y,x)} + O\left(\frac{1}{N}\right)$$

where

$$P(x,y) = \sum_{j=m}^{k} \left(\begin{array}{c} k \\ j \end{array} \right) x^j y^{k-j}$$

and zero otherwise with similar expression for the change of y. This gives the averaged equation

$$\dot{x} = \frac{P(x,y)}{P(x,y)+P(y,x)} \quad \dot{y} = \frac{P(y,x)}{P(x,y)+P(y,x)} \tag{8}$$

Eliminating time we get

$$\frac{dy}{dx} = \frac{P(x,y)}{P(y,x)} \tag{9}$$

Depending on the parameters (9) exhibits two different behaviors

(I) For $k = 1$ (and so $m = 1$) the solutions have the form $x = cy$.

(II) For $k \geq 2$ the solutions blow up in finite time (this can be seen by comparison with $dy/dx = y^k/x^k$.) In fact if $x(0) > y(0)$ then x reaches infinity while y remains finite and if $y(0) > x(0)$ then the opposite happens.

This suggests the similar behavior for the urn process. The theorem below gives a partial justification for this reasoning.

Theorem 2. *(a) If $k = 1$ then with probability 1 b_n and w_n increase to infinity and there exists the limit*

$$\lim_{n\to\infty} \frac{w_n}{n}.$$

(b) Let $k > 1$. Fix $c > 1/2$. Then

$$\mathbb{P}\left(b_n \text{ stays finite} | w_N \geq cN\right) \to 1, \quad N \to \infty.$$

Similarly, for $c < 1/2$

$$\mathbb{P}\left(w_n \text{ stays finite} | w_N \leq cN\right) \to 1, \quad N \to \infty.$$

Theorem 2(b) says that in case $k > 1$ either b_n stays finite, w_n stays finite or $w_n/b_n \to 1$. We will eliminate the third possibility in Section 4.

Proof. (a) Fix $0 < c < 1$ and take $\delta > 0$ so that $c - \delta > 0$, $c + \delta < 1$. Choose $\gamma < 1/2$. Theorem 1 implies that if $w_N/N \in (c-\delta, c+\delta)$ then

$$\mathbb{P}\left(\left|\frac{w_{2N}}{2N} - \frac{w_N}{N}\right| > N^{-\gamma}\right) \leq C_1 \exp(-C_2 N^{1-2\gamma}),$$

In other words if $W_N/N \approx c$ then with large probability $w_{2N}2N \approx c$. Thus

$$\left|\frac{w_{4N}}{4N} - \frac{w_{2N}}{2N}\right| \leq \frac{1}{2^\gamma N^\gamma}, \quad \left|\frac{w_{8N}}{8N} - \frac{w_{4N}}{4N}\right| \leq \frac{1}{4^\gamma N^\gamma}, \text{etc.}$$

More precisely the statement below follows by an easy induction.

Lemma 1.5. *Suppose that*

$$\left|\frac{w_N}{N}-c\right|<\frac{\delta}{2}, \quad \frac{1}{N^\gamma}\sum_{k=1}^{\infty}\frac{1}{2^{k\gamma}}<\frac{\delta}{2}.$$

Then

$$\mathbb{P}\left(\frac{w_j}{j}\in[c-\delta,c+\delta] \text{ for all } j\geq N\right)\geq 1-\varepsilon,$$

where

$$\varepsilon=\sum_{k=1}^{\infty}C_1\exp\left[-C_2(2^kN)^{2\gamma-1}\right].$$

This lemma implies that if $0<c<1$ is a limit point of w_N/N then $w_N/N\to c$. Now since $\left|\frac{w_{N+1}}{N+1}-\frac{w_N}{N}\right|<\frac{1}{N}$ the set of limit points is either one point or an interval. Lemma 1.5 rules out the interval case so part (a) follows.

(b) We consider only the case $c>1/2$, the opposite case is similar. Since all solutions to (9) with $x(0)>y(0)$ have $x(t)$ going to infinity in finite time while $y(t)$ stays bounded for arbitrary $d<1$ there exists a number $t_0=t_0(c,d)$ such that if $x(0)>c-\delta$, $y(0)=1-x(0)$ then

$$\frac{x(t_0)}{x(t_0)+y(t_0)}>d+\delta.$$

Thus Theorem 1 implies that if $W_N>cN$ then $w_{\bar{N}}>d\bar{N}$ where $\bar{N}=[1+t_0]N$ except on the set of probability $C_1e^{-C_2N^{-\gamma}}$. This shows that if $c>1/2$ is a limit point of w_n/n then any $c<d<1$ is also a limit point. Next, if d is sufficiently large then $w_n>dn$ implies

$$\mathbb{P}\left(w_{12n}\geq 10w_n, \quad b_{12n}\leq 2b_n\right)\geq 1-C_1\exp(-C_2n^\gamma).$$

Let now $n_0=n$ $n_{j+1}=12n_j$. Then by induction

$$\mathbb{P}\left(w_{n_{j+1}}\geq 10w_{n_j}, \quad b_{n_{j+1}}\leq 2b_{n_j}\right)\geq 1-\sum_j C_1\exp(-C_2(12^jn)^\gamma). \tag{10}$$

But if (10) holds then

$$w_{n_j}\geq 10^jw_{n_0}, \quad b_{n_j}^3\leq 8^jb_{n_0}$$

so for large m we have $w_m>b_m^3$. Now if $w_m>b_m^3$ the the probability that we will see at least one black ball is at most $e^{-\xi(w_n,b_n)/\mathrm{Const}}$ where

$$\xi(w,b)=\sum_{k=0}^{\infty}\frac{P(b,w+k)}{P(w+k,b)+P(b,w+k)}.$$

Now

$$\xi(b,w) \leq \text{Const}\frac{b^m}{w^{m-1}} \leq \frac{\text{Const}}{w} \tag{11}$$

since $m \geq 2$. This proves (b). □

Exercise 4. *Prove property (II) of equation (9).*

Exercise 5. *Prove (11).*

Exercise 6. *(a) Generalize Theorem 2 to the situation when we add several balls, in particular, when the number of balls added depends on the ratio of the balls drawn.*

(b) Generalize Theorem 2 to three or more colors.

The problems below refer to the case $k = m = 1$. Let $\xi = \lim_{n\to\infty} w_n/n$.

Exercise 7. *([10, 22]) Suppose we begin with w_0 white and b_0 black balls. Show that*

$$\mathbb{P}\left((w_0,b_0) \to (w_1,b_1) \to (w_2,b_2) \to \cdots \to (w_n,b_n)\right) = \frac{(w_n-1)!(b_n-1)!(w_0+b_0-1)!}{(w_n+b_n-1)!(w_0-1)!(b_0-1)!}$$

Deduce that

$$\mathbb{P}\left(w_n = A, b_n = B\right) = \frac{(A-1)!(B-1)!(w_0+b_0-1)!(A+B-w_0-b_0)!}{(A+B-1)!(w_0-1)!(b_0-1)!(A-w_0)!(B-b_0)!}.$$

Conclude that ξ has density

$$p(x) = \frac{(w_0+b_0-1)!}{(w_0-1)!(b_0-1)!}x^{w_0-1}(1-x)^{b_0-1}.$$

Exercise 8. *Show that*

$$\mathbb{P}\left(\left|\frac{w_n}{n} - \xi\right| > \frac{C}{\sqrt{n}}\right) \to 0, \quad \textit{as } C \to \infty.$$

Exercise 9. [11] *(a) Suppose that at each step we add r_w balls of winning color and r_l balls of loosing color. Show that if $r_l > 0$ then $\frac{w_n}{b_n} \to 1$.*

(b) Consider the following random growth model on $\mathbb{Z}$. Let $a_1 = 1, b_1 = -1$. To describe the evolution at time n start a simple random walk $\omega(j)$ at 0 and stop it at the moment τ_n when it reaches either a_n or b_n. In the first case increase a_n by 1, in the second decrease b_n by 1. Show that almost surely $\lim_{n\to\infty} a_n/n = 1/2$.

Exercise 10. *([6]) (Rubin coupling) In this problem we take some sequence $\alpha(n)$ and add white ball with probability $\frac{\alpha(w_n)}{\alpha(w_n)+\alpha(b_n)}$ and black ball with probability $\frac{\alpha(b_n)}{\alpha(w_n)+\alpha(b_n)}$. Let $\{\tau_j\}$ and $\{\sigma_j\}$ be independent processes with independent increments such that $\tau_1 = \tau_2 = \cdots = \tau_{w_0} =$*

$\sigma_1 = \sigma_2 = \cdots = \sigma_{b_0}$ *and afterwards both* $\tau_j - \tau_{j-1}$ *and* $\sigma_j - \sigma_{j-1}$ *have exponential distribution with parameter* $\alpha(j)$. *Let* ρ *denote union of* σ *and* τ. *Enumerate points in* ρ *by* $\rho_1 < \rho_2 < \cdots < \rho_n$. *Let* $\tilde{w}_n$ $(\tilde{b}_n)$ *be the number of points from* τ (σ) *up to the moment* ρ_n. *Show that* (w_n, b_n) *and* $(\tilde{w}_n, \tilde{b}_n)$ *have the same distribution. Deduce that both* w_n *and* b_n *go to infinity iff* $\sum_n 1/\alpha(n) = \infty$.

Exercise 11. [20, 21] *(Street gang shooting). Suppose that instead of adding balls we remove a ball of the opposite color. We stop then there is only one color left. Let* $\hat{w}$ *and* $\hat{b}$ *be the numbers of white and black balls at the end (so* $\hat{w}\hat{b} = 0$*). Suppose at the beginning* $w_n = cn$, $b_n = (1-c)n$. *Show that as* $n \to \infty$ *the distribution of*

$$\sqrt{n}\left(\left[c^2 - (1-c)^2\right] - \left[(\hat{w}/n)^2 - \left(\hat{b}/n\right)^2\right]\right)$$

approaches a limit. Deduce that in case $c = 1/2$ *the winer has about* $n^{3/4}$ *balls.*

Vertex reinforced random walk (VRRW) on a graph G is defined as follows. If the walker is at vertex v which is connected to vertices $v_1, v_2 \ldots v_l$ then he chooses v_j with probability $w_n(v_j)/(\sum_i w_n(v_i))$, where $w_n(v_i)$ equals 1+the number of visits to v_i up to time n.

Exercise 12. *For VRRW on a complete graph on* r *vertices show that* $w_n(v_i)/n \to 1/r$.

Hint. *Show that* $(\max_i w_n(v_i) - \min_i w_n(v_i))/n$ *is decreasing. Deduce that* $(1/r, \ldots, 1/r)$ *is the only possible limit point of* $(w_n(v_1)/n, \ldots, w_n(v_r)/n)$.

Exercise 13. *In case* $G = \{1, 2, 3\}$ *show that there exist limits* $\xi_i = \lim_{n\to\infty} w_n(i)/n$ *and that* $\xi_2 = 1/2$.

Hint. *Reduce to* $k = m = 1$ *urn model.*

2. Anosov Theorem.

Here we present a general result about averaging. Consider a family of equations on some manifold M

$$\dot{z}_\varepsilon = Z(z, \varepsilon) \tag{12}$$

for ε in some interval $[-\varepsilon_0, \varepsilon_0]$. We assume that $Z \in C^2(M \times [-\varepsilon_0, \varepsilon_0])$. We further suppose that for $\varepsilon = 0$, Z has several first integrals $h_1, h_2 \dots h_m$ which are independent in the sense that

$$\mathrm{rk}\left(\frac{\partial h}{\partial z}\right) = m \tag{13}$$

at every point. This implies that level sets $M_c = \{\vec{h} = c\}$ are smooth submanifolds. Let $\phi_\varepsilon(t)$ denote the flow generated by $Z(\cdot, \varepsilon)$. We impose two additional conditions

(COMP) M_h are compact and ϕ_0 restricted to M_h preserves a measure μ_h which is smooth and depends smoothly on h.

(ERG) (ϕ_0, μ_h) is ergodic for almost every h.

Since h are first integrals the Hadamard Lemma allows us to write

$$\dot{h} = \varepsilon Y(z, \varepsilon). \tag{14}$$

Observe that $Y(z, 0) = \mathcal{L}_{\frac{dZ}{d\varepsilon}} h$, where $\mathcal{L}$ denote Lie derivative. Consider the averaged equation

$$\dot{\bar{h}} = \bar{Y}(\bar{h}) \tag{15}$$

where

$$\bar{Y}(\bar{h}) = \int Y(z, 0) d\mu_{\bar{h}}(z).$$

Fix some ball $V \subset \mathbb{R}^m$. Let $d\nu = dh d\mu_h$ where dh is the uniform measure on V. For $z \in M$ let $\tau_\varepsilon(z)$ be the first moment either the solution of (15) with initial condition $\bar{h}(0) = h(z)$ leaves V or $\phi_\varepsilon(z)$ leaves $h^{-1}V$.

Theorem 3. *([1, 24]) Fix $T > 0$. Then as $\varepsilon \to 0$*

$$\sup_{t \in [0, \min(T, \tau_\varepsilon(z))]} \left| h_\varepsilon(t/\varepsilon) - \bar{h}(t) \right| \to 0$$

in probability where $\bar{h}(t)$ denotes the solution of (15) with initial condition $\bar{h}(0) = h(z)$.

Proof. By multiplying Z by a function $\alpha(h)$ where $\alpha \equiv 1$ on V and $\alpha \equiv 0$ outside of a larger ball we may assume that all solutions remain in V for all times. Observe that $\mathrm{div}_\nu(Z(\cdot, 0)) = 0$ since $Z(\cdot, 0)$ preserves

h and μ_h. Since $h^{-1}V$ is compact, it follows that there exists a constant K_1 such that

$$|\mathrm{div}_\nu(Z)| \leq K_1\varepsilon. \tag{16}$$

Thus we have

Corollary 2.1. *There exists a constant K_2 such that for any set $\Omega \subset h^{-1}(V)$ for $0 \leq t \leq T/\varepsilon$ we have*

$$\nu(\phi_\varepsilon(t)\Omega) \leq e^{K_2 T}\nu(\Omega)$$

Next, there is a constant K_3 such that for all $z \in h^{-1}V$ we have

$$|Y(z,\varepsilon)| < K_3.$$

It follows that all functions $\hat{h}_\varepsilon(s) = h_\varepsilon(s/\varepsilon)$ are Lipshitz with Lipschitz constant K_2. In particular the family $\{\hat{h}_\varepsilon\}$ is tight. Let $\tilde{h}(s)$ denote a weak limit point as $\varepsilon \to 0$. Theorem 3 is equivalent to the statement that $\tilde{h}$ satisfy

$$\tilde{h}(s) = \tilde{h}(0) + \int_0^s \bar{Y}(\tilde{h}(u))du.$$

We have

$$h_\varepsilon(s/\varepsilon) - h_\varepsilon(0) = \varepsilon \int_0^{s/\varepsilon} Y(z_\varepsilon(v),\varepsilon)dv =$$

$$\varepsilon \int_0^{s/\varepsilon} Y(z_\varepsilon(v),0)dv + O(\varepsilon) =$$

$$= \varepsilon \int_0^{s/\varepsilon} \bar{Y}(h(z_\varepsilon(v)))dv + O(\varepsilon) + \varepsilon \int_0^{s/\varepsilon} [Y(z_\varepsilon(v),0) - \bar{Y}(h(z_\varepsilon(v)))]dv.$$

Denote the last term by $\gamma_\varepsilon(s)$. We need to show that $\gamma_\varepsilon(s) \to 0$ in probability as $\varepsilon \to 0$. Since $\gamma_\varepsilon(s)$ is uniformly Lipschitz it suffices to show that for any fixed s_0,

$$\gamma_\varepsilon(s_0) \to 0 \text{ in probability.} \tag{17}$$

Indeed, if (17) holds then given δ let $s_k = \frac{k\delta}{2K_3}$. Then

$$\sup_{s\in[0,T]} |\gamma_\varepsilon(s)| \leq \frac{\delta}{2} + \sup_k |\gamma_\varepsilon(s_k)|.$$

By (17), for small ε the second term is less than $\delta/2$ except for the set of vanishing probability, so $\sup_{s\in[0,T]} |\gamma_\varepsilon(s)| \to 0$ in probability as claimed.

Choose a function $L(\varepsilon)$ such that $L(\varepsilon) \to \infty$ yet $e^{L^2(\varepsilon)}\varepsilon \to 0$ as $\varepsilon \to 0$. Denote $t_k = L(\varepsilon)k$. We have

$$\gamma_\varepsilon(s_0) = \varepsilon \sum_{k=0}^{\frac{s_0}{\varepsilon L(\varepsilon)}-1} \int_{t_k}^{t_{k+1}} \left[Y(z_\varepsilon(v), 0) - \bar{Y}(h(z_\varepsilon(v)))\right] dv + O(\varepsilon L(\varepsilon)).$$

Fix $\delta > 0$. We need to show that

$$\nu\left(|\gamma_\varepsilon(s_0)| > \delta\right) \to 0 \text{ as } \varepsilon \to 0 \tag{18}$$

Let $z_{k,\varepsilon}(t)$ denote the solution of

$$\dot{z}_{k,\varepsilon}(t) = Z(z_{k,\varepsilon}, 0), \quad z_{k,\varepsilon}(t_k) = z_\varepsilon(t_k).$$

Let

$$\mathcal{G}_\varepsilon(k) = \left\{ z : \left| \frac{1}{L(\varepsilon)} \int_{t_k}^{t_{k+1}} \left[Y(z_{k,\varepsilon}(v), 0) - \bar{Y}(h(z_\varepsilon(t_k)))\right] dv \right| \leq \frac{\delta}{2T} \right\}.$$

$\mathcal{G}_\varepsilon(z) = \{k : z \in \mathcal{G}_\varepsilon(k)\}, \quad \mathcal{B}_\varepsilon(k) = h^{-1}(V) - \mathcal{G}_\varepsilon(k), \quad \mathcal{B}_\varepsilon(z) = \{k : z \notin \mathcal{G}_\varepsilon(k)\}.$

For $z \in \mathcal{G}_\varepsilon(k)$ we have

$$\varepsilon \int_{t_k}^{t_{k+1}} \left[Y(z_\varepsilon(v), 0) - \bar{Y}(h(z_\varepsilon(v)))\right] dv =$$
$$\varepsilon \int_{t_k}^{t_{k+1}} \left[Y(z_\varepsilon(v), 0) - Y(z_{k,\varepsilon}(v), 0)\right] dv +$$
$$\varepsilon \int_{t_k}^{t_{k+1}} \left[Y(z_{k,\varepsilon}(v), 0) - \bar{Y}(h(z_{k,\varepsilon}(v)))\right] dv +$$
$$\varepsilon \int_{t_k}^{t_{k+1}} \left[\bar{Y}(h(z_{k,\varepsilon}(v))) - \bar{Y}(h(z_\varepsilon(v)))\right] dv =$$
$$I + II + III.$$

Lemma 2.2. *For $t_k \leq t \leq t_{k+1}$*

$$|z_{k,\varepsilon}(t) - z_\varepsilon(t)| \leq C_1 \varepsilon L(\varepsilon) e^{C_2 L(\varepsilon)}.$$

Proof. Let $r(t) = z_{k,\varepsilon} - z_\varepsilon$, then by Hadamard Lemma

$$\dot{r}(t) = \varepsilon\alpha(z_\varepsilon, r, \varepsilon) + \beta(z_\varepsilon, r, \varepsilon) r,$$

for some bounded α, β. Let $\Gamma(t)$ denote the solution of

$$\dot{\Gamma} = \beta\Gamma, \quad \Gamma(0) = 1.$$

Observe that $\Gamma, \Gamma^{-1} = O(e^{CL(\varepsilon)})$ for some C. Now

$$r = \varepsilon\Gamma(t) \int_0^t \alpha^{-1}(s)\Gamma(s) ds = O(\varepsilon e^{2CL(\varepsilon)} L(\varepsilon))$$

as claimed. □

Lemma 2.2 implies that $I = O(\varepsilon^2 L^2(\varepsilon)e^{C_2 L(\varepsilon)})$, $\mathit{III} = O(\varepsilon^2 L^2(\varepsilon)e^{C_2 L(\varepsilon)})$. By the definition of $\mathcal{G}_\varepsilon(k)$ we have $\mathit{II} \leq \frac{\delta \varepsilon L(\varepsilon)}{2T}$.

Now summation over k gives

$$|\gamma_\varepsilon(s_0)| \leq \frac{\delta}{2} + O(s_0 \varepsilon L(\varepsilon) e^{C_2 L(\varepsilon)}) + O(\varepsilon L(\varepsilon) \mathrm{Card}(\mathcal{B}_\varepsilon(z))).$$

Hence to establish (18) it suffices to show that

$$\nu\left(z : \mathrm{Card}(\mathcal{B}_\varepsilon(z)) > \frac{\delta}{4\varepsilon L(\varepsilon)}\right) \to 0.$$

By Chebyshev inequality it suffices to show that

$$\int \varepsilon L(\varepsilon) \mathrm{Card}(\mathcal{B}_\varepsilon(z)) d\nu(z) \to 0.$$

But this integral equals

$$\sum_k \varepsilon L(\varepsilon) \nu(\mathcal{B}_\varepsilon(k)).$$

Observe that $z \in \mathcal{B}_\varepsilon(k)$ iff $\phi_\varepsilon(t_k) z \in \mathcal{B}_\varepsilon(0)$, so by Corollary 2.1

$$\nu(\mathcal{B}_\varepsilon(k)) \leq e^{K_2 T} \nu(B_\varepsilon(0)).$$

Summation over k gives

$$\sum_k \varepsilon L(\varepsilon) \nu(\mathcal{B}_\varepsilon(k)) \leq T e^{K_2 T} \nu(B_\varepsilon(0)).$$

Now

$$\mathcal{B}_\varepsilon(0) = \left\{ z : \frac{1}{L(\varepsilon)} \left| \int_0^{L(\varepsilon)} \left[Y(\phi_0(v) z, 0) - \bar{Y}(h(z)) \right] dv \right| > \frac{\delta}{2T} \right\}.$$

By (ERG) for almost all h

$$\mu_h(\mathcal{B}_\varepsilon(0)) \to 0 \text{ as } \varepsilon \to 0$$

Hence $\nu(\mathcal{B}_\varepsilon(0)) \to 0$ as $\varepsilon \to 0$ by Dominated Convergence Theorem. This completes the proof of Theorem 3. □

Exercise 14. *Consider Hamiltonian system*

$$\dot{q} = p \quad \dot{p} = \nabla U(q).$$

Let $E = \frac{p^2}{2} - U(q)$ *be the energy of this system. Suppose that level sets are compact and that the system is ergodic on almost every level. Let* μ_E *be the restriction of* $dpdq$ *on the level set (that is* $dpdq = d\mu_E dE$*). Consider a perturbation*

$$\dot{q} = p \quad \dot{p} = \nabla U + \varepsilon F(q) - \varepsilon \sigma p.$$

Show that effective equation for E is

$$\dot{\bar{E}} = -\sigma \int p^2 d\mu_{\bar{E}}.$$

We call $I(z)$ *almost adiabatic invariant* for system (12) if for each $\delta > 0$

$$\text{mes}(max_{t\in[0,T/\varepsilon]}|I(z(t)) - I(z(0))| > \delta) \to 0 \text{ as } \varepsilon \to 0.$$

Exercise 15. *Consider a product system*

$$\dot{q_1} = p_1 \quad \dot{p_1} = \nabla U_1(q_1) \quad \dot{q_2} = p_2 \quad \dot{p_2} = \nabla U_2(q_2).$$

Consider a perturbation with Hamiltonian

$$H(p_1, p_2, q_1, q_2) = \frac{p_1^2 + p_2^2}{2} - U_1(q_1) - U_2(q_2) - \varepsilon V(q_1 - q_2).$$

Assume that for almost E_1, E_2 the system is ergodic on

$$p_1^2 - U_1(q_1) = E_1, \quad p_2^2 - U_2(q_2) = E_2.$$

Prove that energies of the first and the second system are adiabatic invariants.

Exercise 16. *Derive a discrete analogue of Theorem 3. Namely consider a system*

$$x_{n,\varepsilon} = f_\varepsilon x_n,$$

such that f_0 has first integrals $h_1, h_2 \dots h_m$ such that M_h is compact, f_0 preserves a smooth measure on M_h and (f, μ_h) is ergodic for almost all h. Obtain effective equation

$$\dot{\bar{h}} = \int (\mathcal{L}_{\frac{df}{d\varepsilon}} \bar{h}) d\mu_h.$$

Exercise 17. *Generalize Exercise 16 to the situations when f_ε can have discontinuities on some compact submanifold K which is transversal to $\{h = \tilde{h}, \varepsilon = \tilde{\varepsilon}\}$ assuming that the first two derivatives of f are uniformly bounded outside K.*

Hint. Add to $\mathcal{B}$ orbits passing near singularities.

Exercise 18. *(*[26, 25]*) Consider a system on segment [0,1] consisting of two particles with masses m_1 and m_2 respectively separated by a heavy piston of mass $M \gg 1$. Let initial velocities of particles equal $v_1(0)$ and $v_2(0)$ respectively and let $V(0)/\sqrt{M}$ denote piston initial velocity. Let $\varepsilon = 1/\sqrt{M}$.*

(a) Show that a collision of the piston with the left particle changes velocities as follows

$$V^+ = V^- + 2m_1\varepsilon v_1^- + O(\varepsilon^2), \quad |v_1^+| = |v_1^-| - 2\varepsilon V^- + O(\varepsilon^2)$$

and a collision of the piston with the right particle changes velocities as follows

$$V^+ = V^- - 2m_2\varepsilon|v_2^-| + O(\varepsilon^2), \quad |v_2^+| = |v_2^-| + 2\varepsilon V^- + O(\varepsilon^2).$$

(b) Show that as $\varepsilon \to 0$ (i.e $M \to \infty$) the system approaches the limit where the piston is fixed and particle bounce off elastically. Deduce that the frequency of collisions of the piston with left (right) particle is $2x/|v_1|$ $(2(1-x)/|v_2|)$.

(c) Let x denote the position of the piston. Obtain effective equation

$$\dot{x} = V, \quad \dot{V} = \frac{m_1 v_1^2}{x} - \frac{m_2 v_2^2}{1-x}.$$

$$\frac{\dot{|v_1|}}{|v_1|} = -\frac{\dot{V}}{x} \quad \frac{\dot{|v_2|}}{|v_2|} = \frac{\dot{V}}{1-x}.$$

(d) Deduce from (c) that

$$v_1^2(s) = \frac{v_1^2(0)x^2(0)}{x^2(s)}, \quad v_2^2(s) = \frac{v_2^2(0)(1-x)^2(0)}{(1-x)^2(s)}$$

and that the motion of x is periodic with Hamiltonian

$$\frac{1}{2}\left[V^2 - \frac{m_1 x^2(0) v_1^2(0)}{x^2} - \frac{m_2(1-x)^2(0) v_2^2(0)}{(1-x)^2}\right].$$

(e) Suppose there are several particles of equal mass on each side. Then if two particles collide they exchange their velocities. Since the particles are identical we obtain the same piston dynamics if we allow the particles to pass through each other. Let $v_1^l, v_2^l \dots v_{k_l}^l$ denote velocities on the left and $v_1^r, v_2^r \dots v_{k_r}^r$ denote velocities on the right. Obtain effective Hamiltonian

$$H = \frac{1}{2}\left[V^2 - \frac{m_1 x^2(0)}{x^2}\sum_k (v_k^l)^2 - \frac{m_2(1-x)^2(0)}{(1-x)^2}\sum_k (v_k^r)^2\right].$$

Exercise 19. *Consider a system of particles and the piston inside a planar domain D. Assume that billiards in*

$$D_l = D \bigcap \{x < a\} \text{ and } D_r = D \bigcap \{x > a\}$$

are ergodic. Let $V_l(a)$, $V_r(a)$ denote the areas of $D_l(a)$ and $V_l(a)$ respectively and let $P_l(a) =$ and $P_r(a)$ be their perimeters. Let $l(a) = D_l'(a)$ be the length of $\{x = a\} \bigcap D$. Show that if the piston is at position x then frequencies of collisions with left (right) particles are

$$\frac{|v_k^{l(r)}| l(x)}{\pi V_l(r)(x)}$$

and the average velocity transfered to piston is $\varepsilon m_{1(2)}|v_k^{l(r)}|/4$. *Show further that the average gain of kinetic energy of a light particle per collision is*

$$\mp\varepsilon\frac{K_j l(x)V}{2V_{l(r)}(x)} = \varepsilon\frac{K_j}{2}\frac{\frac{d}{dt}V_{l(r)}(x)}{V_{l(r)}(x)}.$$

Deduce that the motion of the piston has effective Hamiltonian

$$H = \frac{V^2}{2} + \frac{l(x)}{2}\left[\frac{\sqrt{V_l(x(0))}K_{jl}(0)}{V_l^{3/2}(x)} - \frac{\sqrt{V_r(x(0))}K_{jr}(0)}{V_r^{3/2}(x)}\right].$$

3. Hyperbolic systems.

Let M be a compact manifold and $f : M \to M$ be a diffeomorphism. We suppose that f is hyperbolic and mixing in the following sense.

(I) There exists a foliation $\mathcal{W}$ with smooth leaves such that $f(\mathcal{W}(x)) = \mathcal{W}(f(x))$.

A collection $\mathcal{P}$ of sets each of which belong to a single leaf of $\mathcal{W}$ is called *an almost Markov family* if there are constants r_1, r_2, v, C, γ such that $\forall P \in \mathcal{P}$

(a) $\mathrm{diam}(P) \leq r_1$;

(b) $\mathrm{Vol}(P) \geq v$;

(c) $P = \overline{\mathrm{Int}(P)}$, moreover $\mathrm{Vol}\{p : d(p, \partial P) \leq \varepsilon\} \leq C\varepsilon^\gamma$;

(d) for any u-set F there are disjoint sets $P_i \in \mathcal{P}$ such that $\bigcup_i P_i \subset F$ and $F \setminus \bigcup_i P_i \subset \{p : d(p, \partial F) \leq r_2\}$;

(e) $\bigcup_{\mathcal{P}} P = M$.

Almost Markov families exist. For example, if $r_1$and C are large and v is small then the collection of all sets satisfying (a)–(c) is an almost Markov family.

To define mixing we describe the measures we consider. Choose an almost Markov family $\mathcal{P}$. Fix some constants R, α. Let $E_1(\mathcal{P}, R, \alpha)$ be the set of the measures given by the following expression: for $A \in C(M)$

$$\ell(A) = \int_P A(x) e^{G(x)} dx,$$

where $P \in \mathcal{P}$,

$$|G(x_1) - G(x_2)| \leq R d(x_1, x_2)^\alpha \tag{19}$$

and $\ell(1) = 1$. We will refer to the above functional as $\ell(P, G)$ and write $\ell(P)$ for $\ell(P, 0)$. Let $E_2(\mathcal{P}, R, \alpha)$ be the convex hall of $E_1(\mathcal{P}, R, \alpha)$ and $E(\mathcal{P}, R, \alpha) = \overline{E_2(\mathcal{P}, R, \alpha)}$. Usually we will drop some of the parameters $\mathcal{P}, R, \alpha$ if it does not cause a confusion. We assume the following.

(II) There exists a measure $\nu : \forall \ell \in E \ \forall A \in C^\beta$

$$|\ell(A \circ f^n) - \nu(A)| \leq a(n) ||A||_{C^\beta} \tag{20}$$

where $a(n) < \frac{C}{n^2}$. (20) implies that

$$\sigma(A) = \sum_{n=-\infty}^{\infty} \nu(A(x) A(f^n x))$$

is finite.

Assumptions (I) and (II) are satisfied e.g. by Anosov diffeomorphisms (see Appendix C). Another example is presented below.

Exercise 20. *([15]) (a) Let $R \in SL_d(\mathbb{Z})$ be such that* $\mathrm{Sp}(R)$ *does not contain roots of unity. Consider $f : \mathbb{T}^d \to \mathbb{T}^d$ given by $f(x) = Rx$* mod 1. *Let Γ_u be the sum of expanding eigenspaces of R and Γ_{cs} be the sum of complimentary eigenspaces. Let $\pi_* : \mathbb{R}^d \to \Gamma_*$ denote the corresponding projections. Prove that $\forall \lambda \in \mathbb{Z}^d$*

$$||\pi_u(\lambda)|| \geq \frac{\mathrm{Const}}{||\lambda||^d}.$$

(b) Use (a) to prove that f satisfy the assumptions (I) and (II) with $\mathcal{W}(x) = x + \Gamma_u$, $a(n) = \mathrm{Const}\theta^n$.

Hints. For (a) Let $P(x) = x^k + \sum_j a_j x^j$ be the minimal polynomial of $R|_{V_{cs}}$. $\forall Q$ $\exists r_1 \dots r_{k-1}$, and $q < Q^k$ such that $|\frac{r_j}{q} - a_j| \leq \frac{1}{qQ}$. Let $P_Q(x) = x^k + \sum_j \frac{r_j}{q} x^j$, then $||P_Q(R)\lambda|| \geq \frac{1}{Q}$. Let $v = \pi_{cs}\lambda$ then

$$P_Q(R)\lambda = P_Q(R)(\lambda - v) + P_Q(R)(v).$$

Take $Q \sim \mathrm{Const}||\lambda|| \dots$

For (b) use Fourier decomposition.

Consider the sequence $z_n \in \mathbb{R}^d$ given by

$$z_{n+1} - z_n = \varepsilon A(z_n, f^n x), \quad z_0 = a \tag{21}$$

where function $A(z, x)$ is three times differentiable with respect to z and the norms $||\frac{\partial^\alpha A(z,\cdot)}{\partial^\alpha z}||_{(C^\beta)^d}$, are uniformly bounded for $0 \leq |\alpha| \leq 3$. Let q_n be the solution of the averaged equation

$$q_{n+1} - q_n = \varepsilon \bar{A}(q_n), \quad q_0 = a.$$

where

$$\bar{A}(q) = \int A(q, x) d\nu(x).$$

Let $DA(z, x)$ denote the partial derivative of A with respect to z. Let $\Delta_n = z_n - q_n$. Denote $\Delta_t^\varepsilon = \frac{\Delta_{[\frac{t}{\varepsilon}]}}{\sqrt{\varepsilon}}$.

Theorem 4. *([19, 7]) $\forall \mathcal{P}, R, \alpha$ $\forall \ell \in E(\mathcal{P}, R, \alpha)$ the following holds. If x is distributed according to some $\ell(P.G)$ then as $\varepsilon \to 0$ Δ_t^ε converges weakly to the solution of*

$$d\Delta(t) = D\bar{A}(q(t))\Delta dt + dB$$

where B is a Gaussian process with independent increments, zero mean and covariance matrix

$$< B, B > (t) = \int_0^t \sigma(A(q(s), \cdot)) ds. \tag{22}$$

The proof will consist of several steps.

3.1. **Multiple Mixing.**

Lemma 3.1. *Fix k. There are constants C_1 and C_2 $\forall A_1, A_2 \dots A_k \in C^\beta$ $\forall \ell \in E$*

$$\left| \ell\left(\prod_{j=1}^{k} A(f^{n_j}x)\right) - \prod_{j=1}^{k} \nu(A_j)\right| \leq C_1 \left[a\left(\frac{m}{C_2}\right) + \theta^m\right] \prod_{j=1}^{k} ||A_j||_{C^\beta}$$

where $m = \min(n_j - n_{j-1}), n_0 = 0$.

Proof. We make induction on k. We can assume that $||A_j|| \leq 1$. For $k = 1$ the result holds by Assumption (II).

Let us show how to pass from k to $k+1$. Denote $N = \frac{n_1+n_2}{2}$. Consider an almost Markov decomposition $f^N P = (\bigcup P_j) \bigcup Z$. Choose $y_j \in P_j$. We have

$$\int_P e^{G(x)} \rho_P(x) \prod_{j=1}^{k+1} A(f^{n_j}x)dx =$$

$$\sum_j c_j A_1(f^{-(N-n_1)}y_j) \int_{P_j} e^{\tilde{G}(y)} \prod_{j=2}^{k+1} A(f^{n_j-N}y)dy + O(\theta^m)$$

for some $\tilde{G}$ satisfying (19). The first term is

$$\sum_j c_j A_1(f^{-(N-n_1)}y_j) =$$

$$\int_P e^{G(x)} A_1(f^{n_1}x)dx + O(\theta^m) = \nu(A_1) + O(\theta^m)$$

and the second one equals

$$\prod_{j=2}^{k+2} \nu(A_j) + O\left(a\left(\frac{m}{C_2(k)}\right) + \theta^m\right)$$

by induction. □

3.2. **Moment estimates.** Let $A_j \in C^\beta$ be a sequence of functions such that $||A_j||_{C^\beta} \leq K$, $\nu(A_j) = 0$. Let $S_n = \sum_{j=0}^{n-1} A_j(f^j x)$.

Lemma 3.2.

$$(a)|\ell(S_n) \leq \text{Const};$$
$$(b)\ell(S_n^2) \leq \text{Const}n;$$
$$(c) \quad \ell(S_n^4) \leq \text{Const}n^2,$$

where the constants in (a)–(c) depend only on K but not on sequence A_j.

(d) Let $A(t,x)$ be a function defined on $[0,\mathbf{T}]\times M$ such that for all $t\in[0,\mathbf{T}]$ $A(t,\cdot)\in C^\beta$, $||A(t,\cdot)||_{C^\beta}\leq K$ and $\int A(t,x)d\mu(x)=0$. Let

$$S_\varepsilon(t)=\sum_{j=0}^{[\frac{t}{\varepsilon}]}A(\varepsilon j,f^jx) \tag{23}$$

then as $\varepsilon\to 0$

$$\varepsilon\ell(S_\varepsilon(t))\to\int_0^t\sigma(A(s,\cdot))ds,$$

where

$$\sigma(A)=\sum_{j=-\infty}^{\infty}\nu(A(A\circ f^j));$$

(e) The family $\{\sqrt{\varepsilon}S_\varepsilon(t)\}$ is tight.

Proof.

$$(a)\quad |\ell(S_n)|=|\sum_{j=0}^{n-1}\ell(A_j(f^jx))|\leq \text{Const}\sum_j a(j)\leq\text{Const}.$$

$$(b)\quad \ell(S_n^2)=\sum_{j,k}\ell(A_j(f^jx)A_k(f^kx))\leq\text{Const}\sum_{j,k}a\left(\frac{|j-k|}{C}\right).$$

Now for fixed m there are less than $2n$ pairs (j,k) with $|j-k|=m$. So

$$\ell(S_n^2)\leq\text{Const}n\sum_m a\left(\frac{m}{C}\right)\leq\text{Const}.$$

(e) Fix some large M. We have

$$\ell(S_\varepsilon(t)^2)=\sum_{j,k=0}^{n-1}\ell(A(\varepsilon j,f^jx)A(\varepsilon k,f^kx)=$$

$$\sum_{|j-k|<M}\ell(A(\varepsilon j,f^jx)A(\varepsilon k,f^kx)+$$

$$\sum_{|j-k|\geq M}\ell(A(\varepsilon j,f^jx)A(\varepsilon k,f^kx)=I+I\!I.$$

By the argument of (b) $|\varepsilon I\!I_\varepsilon|\leq\text{Const}\sum_{m>M}a(m)\to 0$ as $M\to\infty$. On the other hand for fixed M the following holds. Let $\varepsilon j\to s$, then

$$\sum_{|k-j|<M}\ell(A(\varepsilon j,f^jx)A(\varepsilon k,f^kx))\to\sum_{|q|<M}\nu(A(s,x)A(s,f^qx))=\sigma(A(s,\cdot))+o_{M\to\infty}(1).$$

Thus

$$\varepsilon\ell(S_\varepsilon^2(t))\to\int_0^t\sigma(A(s,\cdot))ds+o(1).$$

Letting $M \to \infty$ we obtain (d).

$$(c) \quad \ell(S_n^4) = \sum_{j_1,j_2,j_3,j_4} \ell((A_{j_1}(f^{j_1}x)A_{j_2}(f^{j_2}x)A_{j_3}(f^{j_3}x)A_{j_4}(f^{j_4}x)).$$

First, let us estimate the terms where not all indices j_p are different. The sum over terms with at most two different indices is bounded by Const$\times$(the number of terms), hence by Constn^2. Also

$$J = \sum \ell(A_{j_1}(f^{j_1}x)A_{j_2}(f^{j_2}x)A^2_{j_3}(f^{j_3}x)) \leq \text{Const} \sum a\left(\frac{\min j_p - j_{p-1}}{C}\right).$$

For fixed m the number of terms with $\min(n_j - n_{j-1}) = m$ equals Constn^2. Thus

$$J \leq \text{Const} n^2 \sum_m a(m).$$

Now up to the terms of order n^2

$$\ell(S_n^4) = 12 \sum_{j_3} \sum_{j_1,j_2=1}^{j_3} \sum_{j_4=j_3}^{n} \ell(A_{j_1}(f^{j_1}x)A_{j_2}(f^{j_2}x)A_{j_3}(f^{j_3}x)A_{j_4}(f^{j_4}x)) + O(n^2)$$

$$12 \sum_{j_3} \sum_{j_4=j_3}^{n} \ell(S^2_{j_3}A_{j_3}(f^{j_3}x)A_{j_4}(f^{j_4}x)) + O(n^2).$$

Proposition 3.3. $\forall l \; \forall j_3$

$$\ell\left(\sum_{j_4=j_3}^{n} S^2_{j_3}A_{j_3}(f^{j_3}x)A_{j_4}(f^{j_4}x)\right) \leq \text{Const} j_3.$$

Proof. Again it suffices to verify this for $l \in E_1$, say $\ell = \ell(P, G)$. Consider an almost Markov decomposition $f^{j_3}P = (\bigcup_q P_q) \bigcup Z$. Choose $y_q \in P_q$ then

$$\int_P e^{G(x)} S^2_{n_3}(x)A_{j_3}(f_{j_3}x)A_{j_4}(f_{n_4}x)dx =$$

$$O(\theta^{j_3}) + \sum_q c_q S^2_{n_3}(y_q) \sum_{j_4=j_3}^{n} \int_{P_q} e^{\tilde{G}(y)} A_{j_3}(y)A_{j_4}(f^{j_4-j_3}y)dy +$$

$$\sum_q c_q \sum_{j_4=j_3}^{n} \int_{P_q} e^{\tilde{G}(y)} \rho_{P_q}(y)[S^2_{n_3}(f^{-j_3}y) - S^2_{n_3}(y_q)]A_{j_3}(y)A_{j_4}(f^{j_4-j_3}y)dy =$$

$$I + II + O(\theta^{j_3}).$$

By Theorem 3.1 $I \leq \text{Const} \sum_q c_q S_{j_3}^2(y_q)$. Now $\underset{f^{-j_3}P_q}{\text{Osc}} S_{j_3}^2 \leq \text{Const} j_3$, so

$$\sum_q c_q S_{j_3}^2(y_q) \leq \text{Const} j_3 + \ell(S_{j_3}^2) \leq \text{Const} j_3.$$

$$I\!I = \sum_q c_q \sum_{j_4=j_3}^{n} \int_P e^{\tilde{G}(y)} [S_{j_3}(f^{-j_3}y) - S_{j_3}(y_q)][S_{j_3}(f^{-j_3}y) + S_{j_3}(y_q)] \times$$
$$A_{j_3}(y) A_{j_4}(f^{j_4-j_3}y) dy =$$
$$\sum_q c_q \sum_{k=0}^{j_3-1} \sum_{j_4=j_3}^{n} \int_P \left\{ e^{\tilde{G}(y)} [S_{j_3}(f^{-j_3}y) - S_{j_3}(y_q)][A_k(f^{k-j_3}y) + A_k(f^k y_q)] \right\} \times$$
$$A_{j_3}(y) A_{j_4}(f^{j_4-j_3}y) dy$$

The part in brackets is uniformly bounded and uniformly Holder continuous. Thus by Theorem 3.1 the sum over j_4 is uniformly bounded for any q, k. Hence

$$I\!I \leq \text{Const} \sum_q c_q \sum_k 1 = \text{Const} j_3 \sum_q c_q \leq \text{Const} j_3.$$

□

Now

$$\ell(S_n^4) \leq \text{Const} \sum_{j<n} j + O(n^2) = O(n^2).$$

(e) follows from (c) and (37). This concludes the proof of Lemma 3.2. □

3.3. **Convergence to the Gaussian process.**

Lemma 3.4. *Let $S_\varepsilon(t)$ be defined as in (23), then as $\varepsilon \to 0$ the process $\sqrt{\varepsilon} S_\varepsilon(t)$ converges weakly to a Gaussian random process $\mathbf{S}(t)$ with zero mean and covariance matrix*

$$< \mathbf{S}(t), \mathbf{S}(t) >= \int_0^t \sigma(A(s, \cdot)) ds.$$

Proof. By Lemma 3.2(e) $\{S_\varepsilon(t)\}$ is a tight family so we need only to verify convergence of finite dimensional distributions. Let us start with one dimensional distributions. Denote $n = \frac{1}{\varepsilon}$. Define

$$\hat{S}_k = \sum_{j=(k-1)n^{\frac{3}{5}}}^{kn^{\frac{3}{5}} - n^{\frac{1}{10}}} A(\varepsilon j, f^j x), \quad \bar{S}_k = \sum_{j=kn^{\frac{3}{5}} - n^{\frac{1}{10}}}^{kn^{\frac{3}{5}} - 1} A(\varepsilon j, f^j x),$$

$$S^*(t) = \sum_{k=0}^{\left[\frac{t}{n^{\frac{3}{5}}}\right]-1} \hat{S}_k, \quad S^{**}(t) = \sum_{k=0}^{\left[\frac{t}{n^{\frac{3}{5}}}\right]-1} \bar{S}_k.$$

Then by Lemma 3.2 $S^{**}(t) \to 0$ in $L^2(l)$ and, in particular $S^{**}(t) \to 0$ in probability. Let $\psi_k(\xi) = \ell(e^{i\sqrt{\varepsilon}\hat{S}_k\xi})$.

Proposition 3.5.

$$\psi_k(\xi) = 1 - \varepsilon^{\frac{2}{5}}\sigma(A(k\varepsilon^{\frac{2}{5}}, \cdot))(1 + o(1)).$$

Proof. We have

$$\psi_k(\xi) = \mathbb{E}_l\left(1 + i\sqrt{\varepsilon}\hat{S}_k\xi - \frac{\varepsilon\hat{S}_k^2}{2}\xi^2 - i\varepsilon^{\frac{3}{2}}\frac{\hat{S}_k^3}{6}\xi^3 + O\left(\varepsilon^2 h S_k^2\xi^4\right)\right).$$

Using Lemma 3.2 we get

$$\psi_k(\xi) = 1 - \varepsilon^{\frac{2}{5}}\sigma(A(s, \cdot))(1 + o(1)) + O(\varepsilon^{\frac{1}{2}} + \varepsilon^{\frac{3}{5}} + \varepsilon^{\frac{4}{5}}),$$

where the main term comes from $\varepsilon\frac{\hat{S}_k^2}{2}\xi^2$. This proves the proposition. □

Let $\phi_k(\xi) = \ell(e^{i\sqrt{\varepsilon}S_k^*\xi})$.

Proposition 3.6.

$$\ln\phi_{k+1}(\xi) = \ln\phi_k(\xi) - \varepsilon^{\frac{2}{5}}\sigma\left(A(k\varepsilon^{\frac{2}{5}}, \cdot)\right)\frac{\xi^2}{2} + o\left(\varepsilon^{\frac{2}{5}}\right). \tag{24}$$

Proof. It suffices to verify this for $\ell \in E_1$.

(I) Case $k = 0$ constitutes Proposition 3.5.

(II) $k > 0$. Decompose $f^{kn^{\frac{3}{5}}}P = (\bigcup_j P_j) \bigcup Z$. Let $q = kn^{\frac{3}{5}}$. Choose $y_j \in P_j$. Then

$$\ell\left(\exp(i\sqrt{\varepsilon}S_{k+1}^*\xi)\right) =$$
$$\sum_j c_j \exp(i\sqrt{\varepsilon}S_k^*(f^{-q}y_j)\xi)\int_{P_j} e^{i\sqrt{\varepsilon}S_1^*(y)\xi}e^{\tilde{G}(y)}dy + O(\theta^{n^{\frac{1}{10}}}).$$

By Proposition 3.5

$$\int_{P_j} e^{i\sqrt{\varepsilon}S_1^*(y)\xi}e^{\tilde{G}(y)}dy = (1 - \varepsilon^{\frac{2}{5}}\sigma(A(k\varepsilon^{\frac{2}{5}}, \cdot))(1 + o(1)).$$

Hence

$$\phi_{k+1}(\xi) = \sum_j c_j \exp(i\sqrt{\varepsilon}S_k^*(f^{-q}y_j))(1 - \varepsilon^{\frac{2}{5}}\sigma(A(k\varepsilon^{\frac{2}{5}}, \cdot))(1 + o(1))) =$$
$$\phi_k(\xi)(1 - \varepsilon^{\frac{2}{5}}\sigma(A(k\varepsilon^{\frac{2}{5}}, \cdot))(1 + o(1))) + O(\theta^{-n^{\frac{1}{10}}}).$$

Taking logarithms of both sides we obtain the statement required. □

Now summing (24) for $k = 0 \dots [tn^{\frac{2}{5}}]$ we get

$$\ln \ell(e^{i\sqrt{\varepsilon}S^*(t)\xi}) \sim -\frac{\xi^2}{2}\int_0^t \sigma(A(s,\cdot))ds.$$

Since $\sqrt{\varepsilon}[S_\varepsilon(t) - S^*_\varepsilon(t)] \to 0$ in probability we see that one dimensional distributions of $\sqrt{\varepsilon}S_\varepsilon(t)$ converge to those of $\mathbf{S}(t)$. To consider the general case let $t_1 \dots t_r$, $\xi_1 \dots \xi_r$ be some numbers. Denote $\eta_j = \sum_{m=1}^j \xi_m$. We have

$$\sum_j \xi_j S_\varepsilon(t_j) = \sum_j \eta_j [S_\varepsilon(t_j) - S_\varepsilon(t_{j-1})].$$

By the same argument as in the proof of Proposition 24 we obtain

$$\ln \ell\left(\exp[i\sqrt{\varepsilon}\sum_j \xi_j S_\varepsilon(t_j)]\right) \sim -\frac{1}{2}\sum_j \eta_j^2 \int_{t_{j-1}}^{t_j} \sigma(A(s,\cdot))ds.$$

This implies convergence of multidimensional distributions and so proves Theorem 3.4. □

3.4. **End of the proof.** We have

$$\Delta_{n+1} - \Delta_n = \varepsilon\left[A(z_n, f^n x) - \bar{A}(q_n)\right] =$$

$$\varepsilon\left[A(q_n, f^n x) - \bar{A}(q_n)\right] + \varepsilon\left[A(z_n, f^n x) - \bar{A}(q_n, f^n x)\right].$$

Using Hadamard Lemma we rewrite the second term as

$$A(z_n, f^n x) - \bar{A}(q_n, f^n x) = [DA(q_n, f^n x) + \zeta(q_n, f^n x, \Delta_n)]\,\Delta_n$$

where ζ is a smooth function of its arguments, $\zeta(q, x, 0) = 0$. Denote

$$\mathcal{Q}_n = D\bar{A}(q_n) + \bar{\zeta}(q_n, \Delta_n),$$

$$\beta_n = [DA(q_n, f^n x) + \zeta(q_n, f^n x, \Delta_n) - \mathcal{Q}_n]\,\Delta_n,$$

$$\gamma_n = A(q_n, f^n x) - \bar{A}(q_n).$$

Then our equation can be rewritten as

$$\Delta_{n+1} - \Delta_n = \varepsilon\left[\mathcal{Q}_n \Delta_n + \beta_n + \gamma_n\right].$$

Define

$$\Theta_t^\varepsilon = \sqrt{\varepsilon}\sum_{j=1}^{t/\varepsilon} \beta_j, \quad \Gamma_t^\varepsilon = \sqrt{\varepsilon}\sum_{j=1}^{t/\varepsilon} \gamma_j.$$

Proposition 3.7. *(a) As $\varepsilon \to 0$, Γ_t^ε converges to B–the Gaussian process defined by (22);*

(b) As $\varepsilon \to 0$, $\Theta_t^\varepsilon \to 0$.

Proof. (a) is a special case of Lemma 3.4.

To prove (b) denote $\Theta_N=\sum_{j=n}^N \beta_n$, $L=\varepsilon^{-1/4}$. Then

$$\Theta_N=\left(\sum_{m=0}^{N/L}\sum_{n=mL}^{(m+1)L-1}\left[DA(q_n,f^nx)+\zeta(q_n,f^nx,\Delta_{mL})-\mathcal{Q}_n\right]\Delta_{mL}\right)+O(\varepsilon LN)$$

$$=\left(\sum_{m=0}^{N/L}\eta_m\Delta_{mL}\right)+O(\varepsilon LN).$$

Thus

$$\Theta_N\leq \text{Const}\left[L+\max_{0\leq n\leq N}|\Delta_n|\sum_m|\eta_m|\right].$$

Arguing as in the proof of Lemma 3.2 we obtain

$$\mathbb{E}_\ell(|\eta_m|)\leq \text{Const}\sqrt{L}$$

so $\mathbb{E}(\sum_m|\eta_m|)\leq \text{Const}N/\sqrt{L}$. On the other hand part (a) suffices to conclude via Lemma 1.1 that given $\delta>0$ there exists $K>0$ such that $\ell(\sup_{0\leq n\leq N}|\Delta_n|>K\sqrt{\varepsilon})<\delta$. (b) follows. □

Now the proof of Theorem 4 is concluded as the proof of Theorem 1.

Exercise 21. *Prove part (c) of Theorem 1 in present setting.*

3.5. **Fully coupled dynamics.** Now we want to extend the results of the previous subsection to the system

$$x_{n+1}=f(x_n,z_n) \tag{25}$$

$$z_{n+1}-z_n=\varepsilon A(z_n,f^nx),\quad z_0=a \tag{26}$$

where for each z the map $x\to f_z(x):=f(x,z)$ is Anosov. So in case $f_z\equiv f$ we get the results of the previous subsection. The results of the previous subsection depended on the mixing of Anosov systems given by Corollary C.2. Now since points with different z coordinate have different x images we can not work with one f any longer and have to consider map $F_\varepsilon: M\times\mathbb{R}^d\to M\times\mathbb{R}^d$ given by

$$F_\varepsilon(x,z)=(f_z(x),z+\varepsilon A(z,x)).$$

Our first goal is to define standard pairs for F. Observe that $F(x,z):=F_0(x,z)=(f_z(x),z)$. This map is partially hyperbolic. Namely we have the splitting

$$T(M\times\mathbb{R}^d)=E_u\oplus E_s\oplus E_c$$

where $E_u(x,z)$ and $E_s(x,z)$ are Anosov subspaces for (x,z) and $E_c = (0, \mathbb{R}^d)$. Thus for small ε F_ε is also partially hyperbolic and it preserves cones

$$\mathcal{K}_u = \{v_u + v_s + v_c : ||v_s|| \leq \delta ||v_0||, ||v_c|| \leq ||v_0||\}. \tag{27}$$

Define standard pairs as in Appendix C with cone family given by (27). Let E be the set of measures corresponding to the standard pairs, E_1 be the convex hall of E and E_2 be the closure of E_1. Let

$$\bar{A} = \int A(z,x) d\nu^z_{SRB}(x),$$

where $\nu^z_{SRB}(x)$ is the SRB measure for f_z. Fix some z^* and let $q(t)$ be the solution of $\frac{dq}{dt} = \bar{A}(q)$, $q(0) = z^*$ and $\Delta^\varepsilon_t = \frac{z^\varepsilon_{t/\varepsilon} - q(t)}{\sqrt{\varepsilon}}$. Let

$$\sigma_{\alpha\beta}(z) = \sum_{n=-\infty}^{\infty} \int (A_\alpha(x,z) - \bar{A}_\alpha)(A_\beta(f^n_z x, z) - \bar{A}_\beta) d\nu^z_{SRB}(x).$$

Theorem 5. *([2, 9]) If x is distributed according to some measure from E_2 and $|z - z^*| \leq \text{Const}\varepsilon$. As $\varepsilon \to 0$ Δ^ε_t converges weakly to the solution of*

$$Delta(t) = \int_0^t D\bar{A}(q(s))\Delta(s) ds + B(t)$$

where B is a Gaussian process with independent increments, zero mean and covariance matrix

$$< B, B > (t) = \int_0^t \sigma(q(s)) ds. \tag{28}$$

We now state an extension of Corollary C.2 needed to prove Theorem 5.

Corollary 3.8. *If $B(z,x)$ is such that for each z*

$$\int B(z,x) d\nu^z_{SRB}(x) = 0$$

Then for any standard pair ℓ

$$|\mathbb{E}_\ell(B \circ F^n_\varepsilon)| \leq \text{Const}\,(\theta^n + \varepsilon)\, ||A||_{C^2}. \tag{29}$$

Corollary 3.8 is proven in the next subsections. Here we explain why this corollary implies the theorem. The point is that our proof of Theorem 4 was based on moment estimates of Lemma 3.2 and the fact that standard pairs form a Markov family. Now the almost Markov property still holds for arbitrary partially hyperbolic systems, in particulr for F_ε. On the other hand in the proof of Lemma 3.2 we bound error terms by $\sum_{n=1}^{1/\varepsilon} a(n)$ and we used that this sum is uniformly bounded by $\sum_{n=1}^{\infty} a(n)$. Now we have instead $\sum_{n=1}^{1/\varepsilon} (\theta^n + \varepsilon)$ which is also uniformly bounded. The rest of the proof of Theorem 5 proceeds as the proof of Theorem 4.

3.6. **Proof of Corollary 3.8.** Here we explain key ideas of the proof of Corollary 3.8. Details can be found in [8, 9]. We first present the proof for the uncoupled case then $f_z \equiv f$. If (D, ρ) is a standard pair let $\bar{D} = \pi_M(D)$. Then $D = \{(x, \bar{z} + \varepsilon Q(x))\ x \in \bar{D}\}$ where $(\bar{x}, \bar{z})$ is some point in D. Then $F_\varepsilon^n(x, z) = (f^n x, \bar{z} + \varepsilon Q(x) + \sum_{k=0}^{n-1} A(\bar{z}, f^k x) + O(\varepsilon^2 n^2))$. Now consider two cases.

(I) $n \leq C|\ln \varepsilon|$. Then $B(F_\varepsilon^n(x, z))$

$$= B(f^n x, \bar{z}) + \varepsilon D_z B(f^n x, \bar{z}) Q(x) + \varepsilon \sum_{k=0}^{n-1} D_z B(f^n x, \bar{z}) A(f^k x, \bar{z}) + O(\varepsilon^2 n^2).$$

To check (29) we need to bound three integrals. By Corollary C.2

$$\int_{\bar{D}} B(f^n x) \rho(x) dx = O(\theta^n)$$

since $\nu_{SRB}(B) = 0$.

$$\int_{\bar{D}} D_z B(f^n x, z) Q(x) \rho(x) = \int D_z B(x, z) d\nu_{SRB} \int_{\bar{D}} Q(x) \rho(x) dx + O(\theta^n) = O(\theta^n)$$

since

$$\int D_z B d\nu_{SRB} = D_z \int B d\nu_{SRB} = 0.$$

Also Lemma 3.1 imply that

$$\int_{\bar{D}} D_z B(f^n x, \bar{z}) A(\bar{z}, f^k x) d\nu_{SRB} \rho(x) dx = O(\theta^{n-k}).$$

Hence $\sum_k \dots$ converges uniformly and the main contribution comes from the terms where $m := n - k$ is of order 1. Now for fixed m

$$\int_{\bar{D}} D_z B(f^n x, \bar{z}) A(\bar{z}, f^{n-m} x) d\nu_{SRB} \rho(x) dx = \int D_z B(x, \bar{z}) A(\bar{z}, f^{-m} x) dx + O(\theta^n).$$

Thus the LHS of (29) is asymptotic to

$$\varepsilon \sum_{m=1}^{\infty} \int D_z B(x, \bar{z}) A(\bar{z}, f^{-m} x) d\nu_{SRB} x$$

with error term $O(\varepsilon^2 |\ln \varepsilon|^2 + \theta^n)$.

Now in case (II) $n > C|\ln \varepsilon|$ we need to show that the RHS of (29) is $O(\varepsilon)$ so we let $n_1 = n - C|\ln \varepsilon|/2$ use the almost Markov decomposition

$$\mathbb{E}_\ell(H(F_\varepsilon^{n_1} x)) = \sum_j c_j \ell_j(H) + O(\theta^{n_1} ||H||_{C^0}).$$

Applying part (I) to each ℓ_j we get (29) in the uncoupled case.

The proof in the coupled case follows the same strategy (shadowing of the F_ε-orbits by F-orbits) but it there are additional difficulties which we now describe. The point is that now it is no longer true that for $n \sim |\ln \varepsilon|$ $F_\varepsilon^n q$ is close to $F^n q$. Indeed the x-distance between $F_\varepsilon q$ and Fq is $O(\varepsilon)$ and even if there were no extra perturbations at the following iterations still exponential expansion of Anosov diffeos would imply that the distance between $F^{n-1}(Fq)$ and $F^{n-1}(F_\varepsilon q)$ grows as Const^n. However it is still true that $F_\varepsilon^n D$ is close to $F^n D$. To be precise for each $q \in D$ we find $p_n \in D$ such that $F_\varepsilon^n p_n = \exp_{F^n q}(Z_n)$ where $Z_n \in E_c \oplus E_s$. Then an inductive argument shows that

$$Z_{n+1} = \pi_{n+1}(DF(Z_n + \varepsilon Y))$$

where

$$Y = \frac{dF}{d\varepsilon}|_{\varepsilon=0} \circ F^{-1} := (X, A)$$

and π_n is the projection to $E_c \oplus E_s$ along $T(F^n D)$. Now for large n $T(F^n D)$ is very close to E^u so we have

$$Z_n \sim \varepsilon(V(F^n q), \sum_{k=0}^{n-1} A(F^n q) \tag{30}$$

where $V = \sum_{m=1}^{\infty} df_z^m(X_s)$. The second component in (30) is the same as before but the first contributes another term

$$\int \partial_V B d\nu_{SRB}^z.$$

However there is another term which gets out of control. Namely we want to estimate

$$\int_D B(F_\varepsilon^n p_n)\rho(p_n) dp_n$$

rather than

$$\int_D B(F_\varepsilon^n p_n)\rho(q) dq.$$

Now replacing $\rho(q)$ by $\rho(p_n)$ causes little difficulty since if we consider q as a function of p_n than we see that the map $p_n \to \rho(q(p_n))$ is uniformly Holder continuous and our mixing result only use the Holder norm of ρ. Next we have $F^{n+1} p_{n+1} = \exp_{F^{n+1} p_n} R_n$ where

$$R_n \sim -(1 - \pi_{n+1})(\varepsilon Y).$$

We would like to use the fact that

$$\frac{dp_{j+1}}{dp_j} = 1 - \varepsilon \text{div}(1 - \pi_{n+1} Y) + O(\varepsilon^2 |\ln \varepsilon|)$$

and so

$$dp_n = dq \prod_j \frac{dp_{j+1}}{dp_j} = dq \left[(1 - (\sum_j \varepsilon \mathrm{div}(1 - \pi_{j+1} Y)) + O(\varepsilon^2 |\ln^2 \varepsilon|) \right]$$

which would allow to reduce the contribution of $\frac{dq}{dp_n}$ to a correlation sum as above. However E_s and hence π_{n+1} is not smooth so the divergence term can blow up. So we modify our strategy as follows. Let E_{as} be a smooth distribution near E_s. We search for p_n such that $F_\varepsilon^n p_n = \exp_{F^n q}(\bar{Z}_n)$ where $\bar{Z}_n \in E_c \oplus E_{as}$. Then we get

$$Z_{n+1} = \bar{\pi}_{n+1}(DF(\bar{Z}_n + \varepsilon Y))$$

where $\bar{\pi}_n$ is the projection to $E_c \oplus E_{as}$ along $T(F^n V)$. Thus

$$Z_n \sim \varepsilon \bar{V} + \varepsilon \sum_k (0, A(F^k q))$$

where $\bar{V} = \sum_m \Gamma_m X_{as}$, $\Gamma_m = \bar{\pi} dF \bar{\pi} dF \dots \bar{\pi} dF$ and $\bar{\pi}$ is the projection to $E_c \oplus E_{as}$ along E_u. On the other hand we can now show that in a weak sense

$$\frac{dp_{n+1}}{d\pi_n} \sim -\varepsilon \mathrm{div} \left[(1 - \bar{\pi}) \bar{V} + X_u \right].$$

This allows to show that for $n \sim C|\ln \varepsilon|$ the RHS of (29) is asymptotic to

$$\varepsilon \left[\sum_m D_z B(q) A(F^{-m} q) d\nu^z_{SRB}(q) + \int (\partial_{\bar{V}} B)(q) d\nu^z_{SRB}(q) - \right.$$

$$\left. \sum_m \mathrm{div}_u ((1 - \bar{\pi}) \bar{V} + X_u)(F^{-m} q) B(q) d\nu^z_{SRB}(q) \right].$$

This completes the proof of (29) and so Theorem 5 is proven.

4. Fixed points.

4.1. **Ornstein-Uhlembeck process.** If $W(t)$ is $BM(\sigma)$ and A is a constant matrix the *Ornstein-Uhlembeck process* is the solution of the following equation

$$Z(t) - Z(0) = \int_0^t AZ(s)ds + W(t).$$

We shall denote this process by $OU(A, \sigma)$. By Exercise 1 $OU(A, \sigma)$ is Gaussian with zero mean and varaince

$$\int_0^t e^{(t-s)A}\sigma(e^{(t-s)A})^* ds.$$

To describe its asymptotic properties we distinguish three cases.

(I) $\mathrm{Sp}(A)$ has negative real part. As $t \to \infty$ $Z(t)$ converges weakly to

$$\mathcal{N}\left(0, \int_0^\infty e^{tA}\sigma(e^{tA})^* ds\right). \tag{31}$$

(II) $\mathrm{Sp}(A)$ has positive real part. Let $Y(t) = e^{-tA}Z(t)$, then as $t \to \infty$ $Y(t)$ converges weakly to

$$\mathcal{N}\left(0, \int_0^\infty e^{-tA}\sigma(e^{-tA})^* ds\right).$$

(III) A has some eigenvalues with positive real part. Split $\mathbb{R}^d = V_1 \oplus V_2$ where V_1 corresponds to positive and V_2 to non-positive parts of the spectrum. Let $A_1 = A|_{V_i}$. If r is a smallest real part of a positive eigenvalue we can write

$$Z(t) = e^{rt}\left[e^{t(A_1-r)}Y_1(t) + O(e^{-rt})\right]$$

where $Y_1(t)$ converges weakly to

$$\mathcal{N}\left(0, \int_0^\infty e^{-tA_1}\sigma|_{V_1}(e^{-tA_1})^* ds\right).$$

For the rest of this section we shall deal with the setting of Section 1 where $x_0 = 0$ and 0 is a fixed point of $\bar{F}$. Let

$$A = D\bar{F}(0), \quad \sigma_{\alpha,\beta} = \mathbb{E}(F_\alpha(0,\omega)F_\beta(0,\omega)).$$

We restate Theorem 1 in this setting.

Proposition 4.1. *As $t \to \infty$ $x^\varepsilon_{[t/\varepsilon]}$ converges weakly to $OU(A, \sigma)$.*

Exercise 22. *Consider a particle in a small time random field in the presence of friction.*

$$\frac{d^2x}{dt^2} = \varepsilon F(t,\omega) - k\frac{dx}{dt}.$$

Suppose that $|F| <$ Const, F *is constant on* $[n, n+1)$, *the values of* F *at different intervals (with integer ends) are independent and identically distributed and*

$$\mathbb{E}(F) = 0, \quad \mathbb{E}(F^2) = \sigma.$$

Show that as $\varepsilon \to 0$ $\sqrt{\varepsilon}x^\varepsilon(t/\varepsilon)$ *converges to the integral of* $OU(-k, \sigma)$.

4.2. **Hyperbolic fixed point.** In this subsection we suppose that 0 is hyperbolic fixed point for $\bar{F}$.

Theorem 6. *([18]) Let* U *be a small neighbourhood of* 0. *Then there exists* $C > 0$ *such that for all* $\sigma > 0$ *for all* $\delta > 0$ *there is* ε_0 *such that for* $\varepsilon \leq \varepsilon_0$ *we have*

$\mathbb{P}(x_n$ *exits* U *before time* $C\varepsilon|\ln\varepsilon|$ *and the point of exit is within distance* σ *from* $W^u(0)) \geq 1 - \delta$.

Proof. We can choose coordinates $(p, q) \in \mathbb{R}^u \times \mathbb{R}^s$ such that $W^u(x)$ is given by $q = 0$ and $W^s(0)$ is given by $p = 0$. (This coordinate change changes (1) to (7) but this will not distrupt our argument).

We write $x_n = (p_n, q_n)$. The analysis of Subsection 4.1 implies that for each R there exists T_0 such that

$$\mathbb{P}(|p_n| > R\sqrt{\varepsilon}) \geq 1 - \delta/10.$$

Now for this T_0 there is $\bar{R}$ such that

$$\mathbb{P}(|q_n| \leq \bar{R}\sqrt{\varepsilon}) \geq 1 - \delta/10.$$

Next, if U is small enough then there exists $\lambda > 0$ such that the averaged system has the following property

if $(\bar{p}(0), \bar{q}(0)) \in U$ then

$$|\bar{p}(1)| \geq (1+\lambda)|\bar{p}(0)| \quad |q(1)| \leq (1+\lambda)^{-1}|\bar{q}(0)|.$$

It follows that for the actual system there are constants $\tilde{C}_1, \tilde{C}_2, \sigma_0$

$$\mathbb{P}(|p_{1/\sqrt{\varepsilon}}| \leq (1+\frac{\lambda}{2})|p_0|) \leq \tilde{C}_1 \exp(-tC_2(|p_0|/\sqrt{\varepsilon})^2),$$

$$\mathbb{P}(|q_{1/\sqrt{\varepsilon}}| \geq (1+\frac{\lambda}{2})^{-1}|q_0| + \varepsilon^{1/2-\sigma_0}) \leq \tilde{C}_1 \exp(-tC_2\varepsilon^{-2\sigma_0}).$$

Let $n_k = \frac{T_0+k}{\varepsilon}$. Denote

$$\mathcal{A}_0 = \{|p_{T_0/\varepsilon}| > R\sqrt{\varepsilon}, |q_{T_0/\varepsilon}| < \bar{R}\sqrt{\varepsilon}\}$$

$\mathcal{A}_k = \{(p_j, q_j)$ has exited U by the time n_k or $|p_{n_{k+1}}| > (1+\frac{\lambda}{2})|p_{n_k}|$, $|q_{n_{k+1}}| < (1+\frac{\lambda}{2})|q_{n_k}| + \varepsilon^{1/2-\sigma_0}\}$. for $k \geq 1$. Then by induction

$$\mathbb{P}\left(\bigcap_{j=1}^{k} |\mathcal{A}_0\right) \leq 1 - \sum_{j=1}^{k} \tilde{C}_1 \exp(-\tilde{C}_2(1+\frac{\lambda}{2})^j) + k\exp(-\varepsilon^{-2\sigma_0}).$$

Now if all $\mathcal{A}_j$ take place for $j = 0, 1, 2 \ldots m := (C|\ln\varepsilon|/2)$ and C is large enough then the the first alternative of $\mathcal{A}_m$ must prevail. □

Exercise 23. *Let $W^u(0)$ be one dimensional, so that $W^u(0) - \{0\}$ consists of two halves which we call W^u_1 and W^u_2. Show that*

$$\lim_{\varepsilon\to 0} \mathbb{P}\,(\textit{The exit point is near } W^u_1) = \frac{1}{2}.$$

We now revisit the urn model of Subsection 1.3 in case $k \geq 2$. We use the notation of Subsection 1.3. Let $\rho_n = \frac{x_n}{y_n}$. Denote

$$a = \frac{d}{d\rho}|_{\rho=1}\left(\frac{P(\rho,1) - \rho P(1,\rho)}{P(\rho,1) + P(1,\rho)}\right).$$

Exercise 24. *(a) Show that $a > 0$.*

(b) Show that $(\rho^{\varepsilon}_{[t/\varepsilon]} - 1)/\sqrt{\varepsilon} \to R(t)$ where $R(t) = \int_0^t aR(s)ds + W(t)$ where

$$\mathbb{E}W(t) = 0, \quad \mathbb{E}W^2(t) = \int_0^t \frac{2e^{2a(t-s)}ds}{(1+2s)^2}.$$

(c) Show that for large t

$$\frac{(1+2t)}{e^{at}}R(t) \to \mathcal{N}\left(0, \frac{2}{a}\right).$$

(d) Show that 1 is not a limit point of ρ_n.

Exercise 25. *Let N be a compact normally hyperbolic manifold for $\bar{F}$. Let U be a small neighbourhood of N. Prove that there exists $C > 0$ such that for all $\sigma > 0$ for all $\delta > 0$ there is ε_0 such that for $\varepsilon \leq \varepsilon_0$ we have*

$\mathbb{P}(x_n$ exits U before time $C\varepsilon|\ln\varepsilon|$ and the point of exit is within distance σ from $W^u(N)) \geq 1 - \delta$.

Exercise 26. *Let $\bar{x}(t)$ be a hyperbolic trajectory for $\bar{F}$. Let N_δ is the first time $\mathrm{dist}(x_N, \bar{x}(N\varepsilon)) \geq \delta$. Show that if δ is small enough then there is $C > 0$ such that for all $\sigma > 0$*

$\mathbb{P}(N_\delta > C|\ln\varepsilon|$ or $\mathrm{dist}(x_{N_\delta}, \bar{x}(N_\delta\varepsilon)) \geq \sigma) \to 0$ as $\varepsilon \to 0$.

4.3. **Attracting fixed point.** In this section we suppose that 0 is an attracting fixed point for $\bar{F}$. Proposition 4.1 and (31) imply that if $T(\varepsilon)\varepsilon$ tends to infinity not too fast then $x_{T(\varepsilon)}/\sqrt{\varepsilon} \to \mathcal{N}(0, \hat{\sigma})$, where

$$\hat{\sigma} = \int_0^\infty e^{tA} \sigma (e^{tA})^* dt,$$

$A = D\bar{F}(0)$, $\sigma_{\alpha\beta} = \mathbb{E}(F_\alpha(0,\omega)F_\beta(0,\omega))$. Below we give a precise estimate on a possible growth of $T(\varepsilon)$.

Theorem 7. *([12]) There is $\delta > 0$ such that if $\varepsilon T(\varepsilon) \to \infty$, $T(\varepsilon)/e^{\delta/\varepsilon} \to 0$ then $x_{T(\varepsilon)}/\sqrt{\varepsilon} \to \mathcal{N}(0, \hat{\sigma})$.*

Proof. Let C be a large constant. By taking a subsequance we can assume that for small ε either we always have $T(\varepsilon) \geq C|\ln \varepsilon|/\varepsilon$ or for small ε either we always have $T(\varepsilon) < C|\ln \varepsilon|/\varepsilon$. Consider the first case first. Divide the interval $[0, T(\varepsilon)]$ into subintervals of length C/ε. Let $x_{(j)} = x_{Cj/\varepsilon}$. Let Φ be a time C map of the vectorfield $\bar{F}$. Then by Theorem 1

$$x_{(j+1)} = \Phi(x_{(j)}) + \xi_j$$

where $\xi_j/\sqrt{\varepsilon}$ is asymptotically Gaussian and

$$\mathbb{P}(|\xi_j| > \varepsilon^{1/2-\delta}) \leq C_1 e^{-C_2 \varepsilon^{-2\delta}}.$$

Hence

$$\mathbb{P}(\max_j |\xi_j| > \varepsilon^{1/2-\delta}) \leq C_1 |\ln \varepsilon| e^{-C_2 \varepsilon^{-2\delta}}.$$

Now if $\max_j |\xi_j| \leq \varepsilon^{1/2-\delta}$ then also $\max_j |x_{(j)}| \leq \text{Const}\varepsilon^{1/2-\delta}$ and so $\Phi(x_{(j)} = Bx_{(j)} +$
$beta_j$ where $B = e^{CA}$ and $|\beta_j| \leq \text{Const}\varepsilon^{1-2\delta}$. Thus

$$x_{(n)} = \sum_j B^j \xi_{n-j} + O(\varepsilon^{1-\delta}).$$

and the result follows.

In the second case we apply the above reasoning to the last orbit segment of length $C|\ln \varepsilon|/\varepsilon$. Let $m = C|\ln \varepsilon|$. We get

$$x_{(n)} = \sum_{j=0}^{m-1} B^j \xi_{n-j} + B^m x_{(n-m)} + O(\varepsilon^{1-\delta})$$

and the the result follows as before provided that we know that $\mathbb{P}(|x_{(n-m)}| < \varepsilon^{1/2-\delta})$ is close to 1. Thus the theorem follows from the next estimate. □

Lemma 4.2. *There exist $\bar{C}_1, \bar{C}_2, \bar{C}_3, \bar{C}_4$ such that*

$$\mathbb{P}(|x_{(j)}| > R\sqrt{\varepsilon}) \leq \bar{C}_1 e^{-\bar{C}_2 R^2} + \bar{C}_3 j e^{-\bar{C}_4/\varepsilon}.$$

Proof. Let U be a neighbourhood of 0 such that if $x \in U$ then $|\Phi(x)| \leq |x|/2$. Thus there exists $c > 0$ such that $x_{(k)} \in U$ if $\max_{j \leq k} |\xi_j| \leq c$. Let y_j be the process defines similar to $x_{(j)}$ except that if y_j exist U we set $y_j = 0$. Then $\mathbb{P}(y_j \neq x_j) \leq \bar{C}_3 j e^{-\bar{C}_4/\varepsilon}$ so it is enouth to show that

$$\mathbb{P}(|y_j| > R\sqrt{\varepsilon}) \leq \bar{C}_1 e^{-\bar{C}_2/\varepsilon R^2}.$$

But this can be easily proven by induction using the estimate

$$\mathbb{P}(|y_{j+1}| > R\sqrt{\varepsilon}) \leq \mathbb{P}\left(|y_j| > \frac{3R\sqrt{\varepsilon}}{2}\right) + \mathbb{P}\left(|\xi_j| > \frac{R\sqrt{\varepsilon}}{4}\right).$$

□

Exercise 27. *Consider a product of random marices*

$$M_{n+1} = (1 + \varepsilon Q_n) M_n \tag{32}$$

where Q_n are independent and of bounded support. Let $A = \mathbb{E}(Q)$. Suppose that A has simple real spectrum. Let λ be the leading eigenvalue and e be a leading eigenvector.

(a) Let $u_n = M_n u$ for a fixed vector u and $v_n = u_n/||u_n||$. Show that if $n \gg |\ln \varepsilon|/\varepsilon$ then v_n is close to $\pm e$ with probability close to 1. Deduce that if

$$\Lambda_\varepsilon = \lim_{n\to\infty} \frac{\ln ||M_n||}{n}$$

is the Lyapunov exponent of (32) then $\frac{\Lambda_\varepsilon}{\varepsilon} \to \lambda$.

(b) Show that moreover $\frac{v_n \mp e}{\sqrt{\varepsilon}}$ converges to Gaussian random vector with zero mean. Deduce that there exists the limit

$$\lim_{\varepsilon\to 0} \frac{\Lambda_\varepsilon - \varepsilon\lambda}{\varepsilon^2}.$$

Hint. Consider the flow on the unit sphere corresponding to v_n.

5. Diffusive regime.

5.1. **The result.** In this section we consider recurrence (1) in case $\bar{F} \equiv 0$. We shall use the following statement.

Theorem 8. *([27]) Let σ, b satisfy the conditions of Proposition B.1. Then for each x_0 there exists unique process $X(t)$ such that $X(0) = x_0$ and for all functions $\phi \in BC^2(\mathbb{R}^d)$ we have*

$$\phi(X(t)) - \phi(X(0)) - \int_0^t (\mathcal{L}\phi)(X(s))ds$$

is a martingale. That is for any $s_1 \leq s_2 \leq s_k$ for any $\psi_j \in BC^0(\mathbb{R}^d)$ for any $s_k \leq t_1 \leq t_2$ we have

(33)
$$\mathbb{E}\left(\left[\phi(X(t_2)) - \phi(X(t_1)) - \int_0^t (\mathcal{L}\phi)(X(s))ds\right] \prod_j \psi_j(X(s_j))\right) = 0.$$

We call the processes satisfying (33) diffusion processes with diffusion matrix σ and drift b or simply the diffusion with characteristics (b, σ).

We postpone the proof of Theorem 8 till subsection 5.3 and first derive the following corollary.

Consider the recurrence

$$(34)\quad x^\varepsilon_{n+1} - x^\varepsilon_n = \varepsilon F(x^\varepsilon_n, \omega_n) + \varepsilon^2 G(x_n, \omega_n) + \varepsilon^3 H(x_n, \omega_n, \varepsilon), \quad x_0 = a$$

where $\{F(\cdot, \omega_n), G(\cdot, \omega)\}$ are independent identically distributed functions that F, G, H take values in a ball of fixed radius (independent of x) and that for each $k_1, k_2 \dots k_d$, $m_1, m_2 \dots m_d$ the map

$$x \to \mathbb{E}\left(\prod_{l=1}^d F^{k_l}_{(l)}(x, \omega) \prod_{l=1}^d G^{m_l}_{(l)}(x, \omega)\right)$$

is smooth and has two bounded derivatives. Assume that for all x $\mathbb{E}_x(F(x, \omega)) = 0$ and

$$\sigma_{\alpha\beta}(x)(x) = \mathbb{E}\left(\left(F_{(\alpha)}(x, \omega) - \bar{F}_{(\alpha)}(x)\right)\left(F_{(\beta)}(x, \omega) - \bar{F}_{(\beta)}(x)\right)\right).$$

$b_\alpha(x) = \mathbb{E}_x(G_\alpha)$ satisfy the condition of Proposition B.1.

Theorem 9. *([17]) Define X^ε_t by setting $X^\varepsilon_{n/\varepsilon} = x^\varepsilon_{n/\varepsilon}$ and interpolating linearly in between. Then as ε_0 X^ε_t converges to $X(t)$ the diffusion process with characteristics (b, σ).*

5.2. **Averaging.** The proof of Theorem 9 relies on the following a priori estimate.

Lemma 5.1. *There exists a constant C such that*

$$\mathbb{E}\,|X_{n_2}-X_{n_1}|^4 \le \text{Const}\varepsilon^4(n_2-n_1)^2.$$

Consequently $\{x_t^\varepsilon\}$ is tight.

Exercise 28. *Prove Lemma 5.1 by the argument of Lemma 1.3.*

Proof of Theorem 9. Let ϕ be a smooth bounded function. We have

$$\phi(x_{n+1})-\phi(x_n)=\varepsilon D\phi(F)+\varepsilon^2\left[D\phi(G)+\frac{1}{2}D\phi(F,F)\right]+O(\varepsilon^3).$$

Using that $\mathbb{E}(\cdot)=\mathbb{E}(\mathbb{E}(\cdot|\mathcal{F}_n))$ we get

$$\mathbb{E}_x(\phi(x_{n+1})-\phi(x_n))=\varepsilon^2\mathbb{E}_x\left(\left(\mathcal{L}\phi\right)(x_n)\right)+O(\varepsilon^3).$$

Summing over and passing to a weak limit we obtain

$$\mathbb{E}(\phi(X(t))-\phi(X(0))=\int_0^t(\mathcal{L}\phi)(X_s)ds.$$

Likewise by the Markov property

$$\mathbb{E}_x\left(\left[\phi(x_{n_2})-\phi(x_{n_1})-\varepsilon^2\left(\sum_{n=n_1}^{n_2-1}\mathbb{E}_x\left(\left(\mathcal{L}\phi\right)(x_n)\right)\right)\right]\prod_j\psi_j(x_{[s_j\varepsilon^{-2}]})\right)=O(\varepsilon)$$

which implies (33). □

5.3. **Solvability of the martingale problem.**

Proof. Existence follows by Theorem 9. To get uniqueness we begin by analyzing the relation

$$\mathbb{E}\left(\phi(X(t))-\phi(X(0))\right)=\mathbb{E}\left(\int_0^t(\mathcal{L}\phi)(X(s))ds\right). \tag{35}$$

Approximating smooth bounded functions of (t,x) by sums of products $\phi_1(x)\phi_2(t)$ we see that for any function $u(t,x)$ with two bounded x-derivatives and one bounded t derivative we have

$$\mathbb{E}(u(t,X(t))-u(0,X(0)))=\mathbb{E}\left(\int_0^t\left(\partial_t+\mathcal{L}u\right)(X(s))ds\right). \tag{36}$$

Now let u satisfy

$$\partial_t u+\mathcal{L}u=0,\quad u(t,x)=\phi(x)$$

then the last equation becomes $u(0,x)=\mathbb{E}_x(\phi(X(t)))$. In other words (36) completely describes the one dimensional distributions of $X(t)$. Likewise (33) shows that for any $s_1,s_2\dots s_k$ the distribution of X_{t_2}

conditioned on $X(s_1), X(s_2)\dots X(s_k), X(t_1)$ is the same as the distribution of $X(t_2-t_1)$ starting from $X(t_1)$. In other words our process is Markov and (36) describes its transition density. □

5.4. **Extensions.** The assumptions of Proposition B.1 are often too restrictive.

Exercise 29. *Prove the uniqueness of the limit in Theorem 9 without the assumption (41).*

Hint. By continuity $< D(x)v, v >\geq c||v||^2$ in any compact domain D. Modify F outside a large ball to get (41) globally. Use Lemma 5.1 to show that if the ball is large enough then the new process differs little from the old one.

Exercise 30. *Prove Theorem 9 with boundness assumption on G replaced by*

$$||G||_{C^2(B(0,R))}\leq \text{Const}R, \quad E_x G(x,\omega)=Ax.$$

Hint. Get *a priori* bounds on $\ln||x_n||^2$.

Exercise 31. *Let x be a diffusion process on $\mathbb{R}^1$ with characteristics $(a(x),\sigma^2(x))$. Let $\phi:\mathbb{R}^1\to\mathbb{R}^1$ be a diffeomorphism with inverse ψ. Show that under the appropriate assumptions on ϕ and its derivatives $y_t=\phi(x_t)$ is a diffusion process with characteristics*

$$(\phi' a+\frac{1}{2}\phi'' a^2, (\phi'\sigma)^2)\circ\psi.$$

Exercise 32. *(a) Show that $OU(A,\sigma)$ is a diffusion process with characteristics (Ax,σ).*

(b) Use (a) and Theorem 9 to give an alternative proof of Proposition 4.1 (without using Theorem 1).

Appendix A. Random processes.

A.1. **Weak convergence.** Here we give some background about random processes. More information can be found in [3]. By a continuous random process X we mean a measure $\mathbb{P}^X$ on $C([0,T],\mathbb{R}^d)$ for some T, d. We say that a family $\{X^\varepsilon(t)\}$ converges to $X(t)$ if $\mathbb{P}^{X^\varepsilon}$ converge to $\mathbb{P}^X$. We say that a family $\{X^\varepsilon(t)\}$ is tight if given $\eta>0$ there is a compact subset $K_\eta \subset C([0,T],\mathbb{R}^d)$ such that for all ε $\mathbb{P}^{X^\varepsilon}(K) > 1-\eta$. Given a random process X and $t_1, t_2 \dots t_m$ let $\mu^X_{t_1,t_2\dots t_m}$ denote the measure

$$\mu^X_{t_1,t_2\dots t_m}(\Omega) = \mathbb{P}^X((X(t_1), X(t_2)\dots X(t_m)) \in \Omega).$$

We call these measures finite dimensional distributions of X. Then $\{X^\varepsilon(t)\}$ converge to $X(t)$ if $\{X^\varepsilon(t)\}$ is tight and the finite dimensional distribution converge. Finally if $X^\varepsilon(0)$ are uniformly bounded then in order to check that the tightness it suffices to show that there exists a constant C such that for all ε and for all $0 \leq t_1 \leq t_2 \leq T$ we have

$$\mathbb{P}\left([X(t_2)-X(t_1)]^4\right) \leq C(t_2-t_1)^2. \tag{37}$$

A.2. **Gaussian processes.** A random vector in $\mathbb{R}^d$ is called Gaussian with mean μ and variance σ if

$$\mathbb{E}(\exp(is < X, v >)) = \exp\left(is < \mu, v > - \frac{s^2 < \sigma v, v >}{2}\right). \tag{38}$$

We shall write in this case $X \sim \mathcal{N}(\mu,\sigma)$.

(38) imply the following.

(1) If X_1 and X_2 are independent, $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ then

$$X_1 + X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1+\sigma_2). \tag{39}$$

(2) If $X \sim \mathcal{N}(\mu,\sigma)$ and A is a deterministic matrix then

$$AX \sim \mathcal{N}(A\mu, A\sigma A^*) \tag{40}$$

A *Brownian Motion* is a Gaussian process started at 0 with zero mean, independent increments and the variance

$$\mathbb{E}(Z_\alpha(t) Z_\beta(t)) = t\sigma_{\alpha\beta}.$$

We shall denote such Brownian Motion $BM(\sigma)$.

Appendix B. Parabolic equations.

Let $BC^k(\mathbb{R}^d)$ denote the space of functions $\mathbb{R}^d \to \mathbb{R}$ such that the function and its first k derivatives are bounded and uniformly continuous.

We use the following result (see [23]).

Proposition B.1. *Consider the following operator on $BC^2(\mathbb{R}^d)$.*

$$(\mathcal{L}\phi)(x) = \sum_{\alpha\beta} \sigma_{\alpha\beta}(x)(\partial_\alpha \partial_\beta \phi)(x) + \sum_\alpha b_\alpha(x)(\partial_\alpha \phi)(x)$$

where σ and b are bounded and uniformly Holder continuous. Assume also that σ is positive selfadjoint and

$$< \sigma(x)v, v >\geq c||v||^2. \tag{41}$$

Then for any bounded uniformly Holder continuous function $\phi(x)$ the Cauchy problem

$$\partial_t u + \mathcal{L}u = 0, \quad u(t, x) = \phi(x)$$

is well posed on $[0, t]$.

Appendix C. Anosov diffeomorphisms.

A diffeomorphism $f : M \to M$ is called Anosov if there exists an f-invariant splitting $TM = E_u \oplus E_s$ and constants $C > 0, \theta < 1$ such that

$$\forall v \in E_s \quad ||df^n(v)|| \leq C\theta^n ||v||,$$

$$\forall v \in E_u \quad ||df^{-n}(v)|| \leq C\theta^n ||v||,$$

We shall use the properties of Anosov diffeomorphisms summarized below. See e.g. [5, 14].

Proposition C.1. *(a) E_* are uniquely integrable. Thus for any x there is a smooth submanifold $W^*(x)$ such that $TW^*(x) = E_*$. Moreover*

$$\text{if } x_2 \in W^s(x_1) \text{ then } d(f^n x_1, f^n x_2) \leq C\theta^n$$

$$\text{if } x_2 \in W^u(x_1) \text{ then } d(f^{-n} x_1, f^{-n} x_2) \leq C\theta^n.$$

W^* are smooth but the foliations W^s are not. However W^s are Holder and absolutely continuous. In particular the enjoy properties (b) and (c) below.

(b) Let V_1, V_2 be transversal to E_s and $p_s : V_1 \to V_2$ be the projection along W^s leaves then

$$d(p_s x_1, p_s x_2) \leq C d^\alpha(x_1, x_2).$$

(c) The Jacobian $J(p_s)$ is uniformly bounded away from 0 and ∞. Moreover if $d(x, p_s x) \leq \delta$ and $d(TV_1(x), TV_2(p_s x)) \leq \delta$ then

$$|J(p_s) - 1| \leq C\delta^\alpha.$$

(d) Let $K_u = v_u + v_s : \quad ||v_s|| \leq \delta ||v_u||$ then $df(K_u(x)) \subset K_u(fx)$.

Call $\ell = (D, \rho)$ a standard pair if D is a submanifold, $\dim(D) = \dim(E_u)$, $TD \in K_u$, $\text{diam}(D) < R$,

$$\text{mes}(\partial_\varepsilon D) \leq C\varepsilon^\gamma, \tag{42}$$

$\int_D \rho dx = 1$ and $|\ln \rho(x_1) - \ln \rho(x_2)| \leq C d(x_1, x_2)$. Denote $\mathbb{E}_\ell = \int_D A(x)\rho(x)dx$.

We assume that f is topologically mixing.

The main result of this section is the following.

Theorem 10. *If ℓ_1 and ℓ_2 are standard pairs then*

$$|\mathbb{E}_{\ell_1}(A \circ f^n) - \mathbb{E}_{\ell_2}(A \circ f^n)| \leq \text{Const}\theta^n ||A||_{C^\alpha}.$$

Corollary C.2. *There exists an (invariant) measure ν such that for any standard pair ℓ*

$$|\mathbb{E}_\ell(A \circ f^n) - \nu(A)| \leq \text{Const}\theta^n ||A||_{C^\alpha}.$$

Proof. Call a probability measure μ admissible if it has a decomposition

$$\mu(\Omega) = \int_\Lambda \mathbb{E}_{\ell_\alpha}(\Omega) d\zeta(\alpha)$$

where ℓ_α satisfy all of the conditions of the standard pairs except that (42) is replaced by $\mu(\partial_\varepsilon D) \leq C\varepsilon^\gamma$, for $\varepsilon \leq \varepsilon_0$. We claim that Theorem 10 implies that for any admissible μ_1, μ_2

$$|\mu_1(A \circ f^n) - \mu_2(A \circ f^n)| \leq \text{Const}\theta^n ||A||_{C^\alpha}. \tag{43}$$

Let us prove (43).

Observe that, the set admissible measures is invariant. Indeed

$$\mathbb{E}_\ell(A \circ f) = \int_{fD} A(y)\tilde{\rho}(y)dy$$

where $\tilde{\rho}(y) = \rho(f^{-1}y) \det(df|TD)(f^{-1}y)$, so

$$|\ln \tilde{\rho}(y_1) - \ln \tilde{\rho}(y_2)|$$

$$\leq \left|\ln \rho(f^{-1}y_1) - \ln \rho(f^{-2}y_2)\right| + \left|\ln \det(df|TD)(f^{-1}y_1) - \ln \det(df|TD)(f^{-1}y_2)\right|$$

preserving the density estimate. Also $\partial_\varepsilon(fD) \subset f(\partial_{\theta\varepsilon}(D))$ so boundary regularity improves. However we necessary have diameter bound. To get it we must cut fD into N pieces $N = N(f, R)$ which worsens our bound by Const$N\varepsilon$. So if $\tilde{\mu} = f\mu$ then

$$\tilde{\mu}(\partial_\varepsilon D_\alpha < \varepsilon) \leq C\theta^\gamma\varepsilon^\gamma + \bar{C}N\varepsilon = \varepsilon^\gamma \left(C\theta^\gamma + \bar{C}N\varepsilon^{1-\gamma}\right) \leq \varepsilon^\gamma \left(C\theta^\gamma + \bar{C}N\varepsilon_0^{1-\gamma}\right)$$

and the last factor is less than C if ε_0 is small enough.

Moreover the above argument actually implies that $f^n\mu = \bar{\mu} + \bar{\bar{\mu}}$ where $\bar{\mu}$ is admissible with all l_α standard and $\bar{\bar{\mu}}(1) \leq C\tilde{\theta}^n$. Fix an arbitrary standard pair ℓ_0. Then if μ_1, μ_2 are admissible then $f^{n/2}\mu_j = \bar{\mu}_1 + \bar{\bar{\mu}}_j$ and

$$\left|\bar{\mu}_j(A \circ f^{n/2} - \bar{\mu}_2(A \circ f^{n/2})\right| \leq C\theta^{n/2}||A||_{C^\alpha}$$

verifying (43). By invariance for any admissible μ the sequence $\{\mu(A \circ f^n)\}$ is Cauchy with exponentially small difference. This gives Corollary C.2. □

Also since any smooth foliation with leaves in K_u can be cut into standard pieces Corollary C.2 implies

Corollary C.3.

$$\left|\int_M A(f^n x)B(x)dx - \nu(A)\int_M B(x)dx\right| \leq \text{Const}\theta^n.$$

Proof of Theorem 10. We use coupling approach (see [28, 4]). We want to divide $f^n D_1$ and $f^n D_2$ into pieces which are close to each other and compare the integrals over nearby pieces using Proposition C.1. However since distortion is nonuniform some pieces of $f^n D_j$ are heavier than others, so it makes sense to couple each heavy piece of D_1 to several light pieces of D_2. In fact it is more convenient to spit each point into infinitely many pieces. To this end we introduce an auxiliary object. Let $Y_j = D_j \times [0,1]$. Equip it with the measure m_j given by $dm_j = \rho_j(x)dxdt$. The technical core of the coupling argument is the following.

Lemma C.4. *There exists a measure preserving map $\tau : Y_1 \to Y_2$ and a function $R : Y_1 \to \mathbb{N}$ such that*
(a) If $\tau(x,t) = (y,s)$ then $\forall n > R(x,t)$

$$d(f^n x, f^n y) \leq \text{Const}\theta^{n-R}.$$

$$(b) \quad m_1(R > n) \leq C\theta^n.$$

Write $f^n(x,t) = (f^n x, t)$. Lemma C.4 implies Theorem 10 since

$$\int_{D_1} A(f^n x)\rho_1(x)dx - \int_{D_2} A(f^n x)\rho_2(x)dx$$

$$= \iint_{Y_1} A(f^n x) m_1(x,t)dx - \iint_{Y_2} A(f^n x) dm_2(x,t)dx$$

$$= \iint_{Y_1} \left[A(f^n x) - A(f^n \tau x)\right] dm_1(x,t) = I + I\!I$$

where I denotes the integral over $\{R \leq n/2\}$ and $I\!I$ denotes the integral over $\{R > n/2\}$. Then $I = O(\theta^{n\alpha/2}$ by (a) and $I\!I = O(\theta^n)$ by (b). □

Proof of Lemma C.4. We need to construct τ and R as above. Let $\mathcal{Y}$ be the set of rectangles corresponding to the standard pairs. We describe the algorithm to couple $Y_1, Y_2 \in \mathcal{Y}$. It works recursively. During the first run we define the coupling map between subsets $P_j^\infty \subset Y_j$. For points where τ is not defined we define recovery time $S(\omega)$ such that $P_j^n = \{\omega \in Y_j : S(\omega) = n\}$ will be of the form $\bigcup_k f^{-n} Y_{jkn}$, $Y_{jkn} \in \mathcal{Y}$, $m_1(\bigcup_k Y_{1kn}) = m_2(\bigcup_k Y_{2kn})$. We then use our algorithm recursively to couple $\bigcup_k Y_{1kn}$ to $\bigcup_k Y_{2kn}$. More precisely we can further subdivide Y_{jkn} so that the new subdivision (which we still call $\{Y_{jkn}\}$) satisfies $m_1(Y_{1kn}) = m_2(Y_{2kn})$. Let $Y_{jkn} = D_{jkn} \times I_{jkn}$. Let r_{jkn} be the orientation preserving affine map between I_{jkn} and $[0,1]$ and set $\bar{Y}_{jkn} = D_{jkn} \times [0,1]$. We equip $\bar{Y}_{jnk}$ with the measure $d\bar{m}_{jkn} = \bar{\rho}_{jkn} dxdt$ where $\bar{\rho}_{jkn} =$

$|I_{jkn}|\rho_{jkn}$. Denote $\Delta_{jkn}(x,t)=(f^n x, r_{jkn}t)$. Let tau_{jkn} and R_{jkn} be the coupling map and coupling time for the pair (Y_{1kn}, Y_{2kn}). Set

$$\tau=\begin{cases}\tau_{\text{first run}} & \text{on } P_1^\infty\\ \Delta_{2nk}^{-1}\tau_{1kn}\Delta_{1kn} & \text{on } f^{-n}(Y_{1kn})\end{cases},$$

$$R=\begin{cases}0 & \text{on } P_1^\infty\\ n+\tau_{1kn}\circ\Delta_{1kn} & \text{on } f^{-n}(Y_{1kn})\end{cases}.$$

We now describe the first run of our algorithm. Observe that for any there is a constant ε_1 such that for any (D,ρ) forming a standard pair there is $x\in D$ such that $D\supset B(x,\varepsilon_1)$. Now by the topological mixing assumption there are n_0, r such that for any D_1, D_2 there is $x_1\in D_1$ such that for all $z\in B(f^{n_0}x_1, r)$ the intersection $W^s_\delta(z)\bigcap f^{n_0}D_2\neq\emptyset$. We let

$$c_1=\int_{f^{-n_0}B(x_1,r)}\rho_1 dx,\quad c_2=\int_{f^{-n_0}p_sB(x_1,r)}\rho_2 dx,$$

$$(\bar t_1,\bar t_2)=\begin{cases}(1,c_1/c_2) & \text{if } c_2>c_1\\ (c_2/c_1,1) & \text{if } c_1\geq c_2\end{cases}.$$

We define $Q_1^{n_0}=f^{-n_0}B(x_1,r)\times[0,t_1]$, $Q_2^{n_0}=f^{-n_0}p_sB(x_1,r)$ and set $\tilde S=n_0$ on $\tilde P_j^{n_0}=Y_j-Q_j^{n_0}$. We now proceed to define $\tilde P_j^n$ inductively for $n>n_0$. Let $Q_j^{n-1}=Y_j-\bigcup_j\tilde P_j^n$. We assume by induction that $f^{n-1}Q_j^{n-1}=\bigcup_k Z_{jk(n-1)}$ where $Z_{jk(n-1)}=D_{jk(n-1)}\times[0,t_{jk(n-1)}]$ and $m_1(Z_{1k(n-1)})=m_2(Z_{2k(n-1)}$, $D_{2k(n-1)}=p_sD_{1k(n-1)}$. Divide $f(D_{jk(n-1)}$ into standard pieces $f(D_1kn-1)=\bigcap_l D_{1kln}$ and let $D_{2kln}=p_s(D_{1kln})$. In general $m_1(D_{1kln})\neq m_2(D_{2kln})$, so we cut off the top of the larger rectangle. Call the trimmed rectangles $\tilde Z_{jkln}$. Add $f^{-n}(Z_{jkln}-\tilde Z_{jkln})$ to $\tilde P_j^n$. Let $P_j^\infty=\bigcap_n Q_j^n$, let $P_j^n=\tilde P_j^{n-N}$. Then P_j^n consists of rectangles over standard pairs by the argument of Corollary C.2. To complete the proof we need to show four things.

- (i) τ is defined almost everywhere;
- (ii) τ is measure preserving;
- (iii) τ satisfies condition (a) of Lemma C.4;
- (iv) τ satisfies condition (b) of Lemma C.4.

To prove (i) it is enough to demonstrate that $\tau|_{P_1^\infty}$ is measure preserving. But $P_1^\infty=\bigcap_n Q_1^n$. Let $\mathcal{A}_n$ be the algebra generated by Z_{jkn}. Then $A_n\subset A_{n+1}$ A_n restricted to P_1^∞ increases to the Borel sigma-algebra on P_1^∞. But $m_1(Z_{1kln})=m_2(Z_{2kln})$. So (i) follows from martingale convergence Theorem. Likewise it suffices to establish (iii) on P_1^∞ but where it is obvious since if $\tau(x,t)=(y,s)$ and $(x,t)\in P_1^\infty$

then $f^{n_0}x \in W^s_\delta(f^{n_0}y)$. It remains to establish (iv) which implies (i). To this end we claim that

$$m_1(s > n) \le C\theta^n. \tag{44}$$

This follows from the estimate

$$\left|\frac{m_1(Z_{1kln})}{m_2(Z_{2kln})} - 1\right| \le C\theta^n \tag{45}$$

which we now prove. Observe that $m_j|_{Z_{jk(n-1)}} = c_{jk(n-1)}\rho_{jk(n-1)}$ where

$$\frac{\rho_{jk(n-1)}(y_1)}{\rho_{jk(n-1)}(y_2)} = \left(\prod_{m=1}^{r} \frac{\det(df|Tf^{n-m-1}V_j)(f^{-m}y_1)}{\det(df|Tf^{n-m-1}V_j)(f^{-m}y_2)}\right)(1 + O(\theta^r)).$$

Let p_n be the projection along the stable leaves between f^nD_1 and f^nD_2. Then by Proposition C.1(c) $|J(p_{n-1}) - 1| \le C\theta^n$. The last two inequalities and the fact that $m_1(Z_{1k(n-1)}) = m_2(Z_{2k(n-1)})$ imply that $c_1/c_2 = 1 + O(\theta^n)$. Combining this with $J(p_n) = 1 + O(\theta^n)$ we obtain (45).

(44) provides a bound on $m_1(s > n)$ uniform over all pairs (Y_1, Y_2). Increasing s and slightly decreasing P_1^∞ if necessary we can assume that $m_1(s = n) = q_n$ where q_n is independent of Y_1, Y_2.

Next if τ is not defined after m runs let $s_m(\omega)$ be the sum of recovery times for first m runs. Let $p_m(n) = m_1(s = n)$. Then

$$p_m(n) = \sum_l q_l p_{m-1}(n - l).$$

For the generating functions

$$P_m(z) = \sum_n p_m(n)z^n, \quad Q(z) = \sum_n q_n z^n$$

we get $P_m(z) = P_{m-1}(z)Q(z)$. Thus $P_m(z) = (Q(z))^m$. Let

$$r_n = m_1(\omega : \exists m : s_m(\omega) = n), \quad R_n(z) = \sum r_n z^n.$$

Then $R(z) = \frac{Q(z)}{1-Q(z)}$ converges in some disc $\{|z| > 1/\tilde{\theta}\}$. Thus $|r_n| \le C(\tilde{\theta} + \varepsilon)^n$. and the result follows. $\square$

References

[1] Anosov D. V. *Oscillations of Systems of ODE with rapidly oscillating solutions,* Izvestia Math. **24** (1960) 721–742.

[2] Bakhtin V. I. *Cramer asymptotics in the averaging method for systems with fast hyperbolic motions,* Tr. Mat. Inst. Steklova **244** (2004) 65–86.

[3] Billingsley P. *Convergence of Probability Measures,* 1979, Wiley and Sons, New York.

[4] Bressaud X. & Liverani C. *Anosov diffeomorphisms and coupling,* Erg. Th. & Dynam. Sys. **22** (2002) 129–152.

[5] Brin M. & Stuck G. *Introduction to dynamical systems,* Cambridge University Press, Cambridge, 2002.

[6] Davis B. *Weak limits of perturbed random walks and the equation* $Y_t = B_t + \alpha \sup\{Y_s \colon s \leq t\} + \beta \inf\{Y_s \colon s \leq t\}$, Ann. Probab. **24** (1996) 2007–2023.

[7] Dolgopyat D. *Limit Theorems for partially hyperbolic systems,* Trans. AMS **356** (2004) 1637-1689.

[8] Dolgopyat D. *On differentiability of SRB states for partially hyperbolic systems,* Invent. Math. **155** (2004) 389–449.

[9] Dolgopyat D. *Averaging and invariant measures,* preprint.

[10] Eggenberger F. & Polya G *Uber die Statistik verketter vorgage* Z. Angew. Math. Mech **1** (1923) 279–289.

[11] Freedman D. *Bernard Fieidman's urn,* Ann. Math. Stat **36** (1965) 956–970.

[12] Freidlin M. I. & Wentzell A. D. *Random perturbations of dynamical systems,* Translated from the 1979 Russian original by Joseph Szucs. Second edition. Grundlehren der Mathematischen Wissenschaften **260** Springer-Verlag, New York, (1998)

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

[14] Katok A. & Hasselblatt B. *Introduction to the modern theory of dynamical systems,* Encyclopedia Math., Appl. **54** (1995) Cambridge University Press, Cambridge.

[15] Katznelson Y. *Ergodic automorphisms of* $\mathbb{T}^n$ *are Bernoulli* Israel J. Math. **10** (1971) 186-195.

[16] Khasminskii R. Z. *Stochastic processes defined by differential equations with a small parameter,* Theor. Probability Appl. **11** (1966), 211–228]

[17] Khasminskii R. Z. *A limit theorem for solutions of differential equations with a random right hand part,* Theor. Probability Appl. **11** (1966) 444–462.

[18] Kifer Yu. *The exit problem for small random perturbations of dynamical systems with a hyperbolic fixed point,* Israel J. Math. **40** (1981) 74–96.

[19] Kifer Yu. *Limit theorems in averaging for dynamical systems,* Erg. Th. & Dyn. Sys. **15** (1995) 1143–1172.

[20] Kingman J. F. C. *Martingales in the OK Corral,* Bull. London Math. Soc. **31** (1999) 601–606.

[21] Kingman J. F. C. & Volkov S. E. *Solution to the OK Corral model via decoupling of Friedman's urn,* J. Theoret. Probab. **16** (2003), 267–276.

[22] Kotz S. & Johnson N. *Urn models and their applications,* Wiley and Sons, New York, 1977.

[23] Krylov N. V. *Lectures on elliptic and parabolic equations in Holder spaces,* Grad Studies in Math. **12**, (1996) AMS, Providence.

[24] Lochak P. & Meunier C. *Multiphase Averaging for Classical Systems,* Springer Appl. Math. Sci. **72** 1988.

[25] Neishtadt A. I. & Sinai Ya. G. *Adiabatic piston as a dynamical system,* J. Stat. Phys. **116** (2004), 815–820.

[26] Sinai Ya. G. *Dynamics of a massive particle surrounded by a finite number of light particles,* Th. & Math. Phys. **121** (1999) 1351–1357.

[27] Stroock D. W. & Varadhan S. R. S. *Multidimensional diffusion processes,* Grundlehren der Mathematischen Wissenschaften **233** (1979) Springer-Verlag, Berlin-New York.

[28] Young L.-S. *Recurrence times and rates of mixing,* Israel J. Math. **110** (1999) 153–188.