Random Iteration of Euclidean Isometries

Markus Ådahl
Department of Mathematics
UmeåUniversity
SE-901 87 Umeå, Sweden

Ian Melbourne
Department of Maths and Stats
University of Surrey
Guildford GU2 7XH, UK

Matthew Nicol
Department of Maths and Stats
University of Surrey
Guildford GU2 7XH, UK

October 24th, 2002

Abstract

We consider the statistical behaviour of i.i.d. compositions of a finite set of Euclidean isometries of $\mathbb{R}^n$. We give a new proof of the central limit theorem and weak invariance principles, and we obtain the law of the iterated logarithm. Our results generalise immediately to Markov chains.

We also give simple geometric criteria for orbits to grow linearly or sublinearly with probability one and for nondegeneracy (nonsingular covariance matrix) in the statistical limit theorems.

Our proofs are based on dynamical systems theory rather than a purely probabilistic approach.

1 Introduction

In this short note we consider the behaviour of random iterations of a finite set of Euclidean isometries, elements of the group $\mathbf{E}(n)$, acting on Euclidean space. We were motivated by a paper of Ambroladze and Adahl [1] in which they proved that orbits are bounded with probability one if and only if the isometries have a common fixed point. Issues such as the growth rate in the unbounded case are not addressed in [1].

Gorostiza [7, 8] established the central limit theorem (CLT) and the weak invariance principle (WIP) for compositions of independent identically distributed (i.i.d.) Euclidean isometries with second moments, generalising results of [16, 17].

In this paper, we specialise to the case where the isometries are chosen from a finite subset $\{\gamma(1), \ldots, \gamma(d)\}$ of $\mathbf{E}(n)$. We give a new proof of the CLT and WIP and in addition we prove the law of the iterated logarithm (LIL). The idea is that random iteration of a finite set of isometries is equivalent to iterating a skew product map on $X \times \mathbf{E}(n)$ where X is a full shift on d symbols. We are thus in a position to apply results on the dynamical behaviour of Euclidean extensions of chaotic dynamical systems (references [13, 6, 12]).

Moreover, the setting in [13, 6, 12] works equally well for Euclidean extensions $X \times \mathbf{E}(n)$ for which X is a subshift of finite type. Hence our results for i.i.d.'s generalise immediately to finite Markov chains.

We also give simple geometric conditions for these statistical limit laws to be nondegenerate and we distinguish between linear and sublinear growth.

The remainder of this paper is structured as follows. In Section 2, we state the problem precisely and list our main results. The reformulation as a skew product $X \times \mathbf{E}(n)$ is given in Section 3. The proofs of our results are given in Section 4.

2 Random Euclidean Isometries

In this section, we describe our main results. For the most part, we concentrate on random compositions of i.i.d. isometries. The generalisation to Markov chains is given in Subsection (b) below.

The Euclidean group $\mathbf{E}(n)$ is the group of isometries of $\mathbb{R}^n$. We denote elements of $\mathbf{E}(n)$ as $\gamma = (h, k)$ where $h \in \mathbf{O}(n)$ and $k \in \mathbb{R}^n$. The action of $\mathbf{E}(n)$ on $\mathbb{R}^n$ is given by

$$\gamma \cdot z = hz + k.$$

Accordingly, the group multiplication is given by

$$(h, k)(h', k') = (hh', hk' + k),$$

and we have the semidirect product $\mathbf{E}(n) = \mathbf{O}(n) \ltimes \mathbb{R}^n$.

Fix a finite collection $\gamma(1), \ldots, \gamma(d) \in \mathbf{E}(n)$, and fix probabilities $p(1), \ldots, p(d) \in (0, 1)$ with $p(1) + \cdots + p(d) = 1$. Given a point $Z_0 \in \mathbb{R}^n$, we choose $\gamma(i)$ with probability $p(i)$ and map $Z_0 \mapsto Z_1 = \gamma(i)Z_0$. Then we iterate the process starting with the point Z_1, and so on.

More precisely, we consider a sequence of i.i.d. random variables $Y_1, Y_2, \ldots$ taking values in $\{\gamma(1), \ldots, \gamma(d)\}$ such that $Y_1 = \gamma(i)$ with probability $p(i)$. Given $Z_0 \in \mathbb{R}^n$, we then consider the trajectory $\{Z_N : N \geq 0\}$ in $\mathbb{R}^n$ given by

$$Z_0, \quad Z_1 = Y_1 Z_0, \quad Z_2 = Y_2 Z_1 = Y_2 Y_1 Z_0, \quad \ldots \quad Z_N = Y_N Z_{N-1} = Y_N Y_{N-1} \cdots Y_2 Y_1 Z_0$$

Note that $\{Z_N\}$ is a sequence of random variables with values in $\mathbb{R}^n$. We are interested in the rate of growth (linear, square-root growth, bounded, etc) of Z_N. (The rate of growth is independent of the initial condition Z_0.)

It is clear that if the isometries $\gamma(i)$ have a common fixed point, then Z_N is bounded. Indeed, if x_0 is the common fixed point, then $|Z_N - x_0| = |Z_0 - x_0|$ for all N. Conversely, Ambroladze and Adahl [1] show that if the $\gamma(i)$ do not have a common fixed point, then $\{Z_N\}$ is unbounded with probability one. Moreover, if K is any compact subset of $\mathbb{R}^n$, then $\Pr(Z_N \in K) \to 0$ as $N \to \infty$.

For each of the d isometries $\gamma(i) \in \mathbf{E}(n)$, we have the decomposition $\gamma(i) = (h(i), k(i)) \in \mathbf{O}(n) \ltimes \mathbb{R}^n$. It is convenient to introduce the subgroup $G \subset \mathbf{O}(n)$ defined to be the smallest closed subgroup containing $h(1), \ldots, h(d)$. The group G acts on $\mathbb{R}^n$ and we define $\mathrm{Fix}(G) = \{v \in \mathbb{R}^n : gv = v \text{ for all } g \in G\}$.

Our first result gives necessary and sufficient geometric conditions for the growth to be linear almost everywhere.

Theorem 2.1 *Define $\overline{k} = \sum_{i=1}^d p(i)k(i)$. Let $\pi : \mathbb{R}^n \to \mathbb{R}^n$ be orthogonal projection onto* $\mathrm{Fix}(G)$ *and define $\overline{Z} = \pi\overline{k}$. Then with probability one,*

$$\lim_{N\to\infty} \frac{1}{N} Z_N = \overline{Z}.$$

Thus, we conclude that the growth is almost surely linear if $\overline{k}$ has a nonzero component in $\mathrm{Fix}(G)$ and almost surely sublinear otherwise.

We now state our main results.

Theorem 2.2 (Central Limit Theorem (CLT)) *The sequence $\{\frac{1}{\sqrt{N}}(Z_N - N\overline{Z})\}$ converges in distribution to a normal distribution with mean zero and covariance matrix Σ. That is, for all rectangles $I \subset \mathbb{R}^n$,*

$$\Pr\{\tfrac{1}{\sqrt{N}}(Z_N - N\overline{Z}) \in I\} \to \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}} \int_I \exp\{-\tfrac{1}{2}\langle \Sigma^{-1}x, x\rangle\} dx_1 \cdots dx_n\,,$$

as $N \to \infty$. The covariance matrix Σ satisfies $g\Sigma = \Sigma g$ for all $g \in G$.

Our statistical limit theorems are *nondegenerate* if Σ is nonsingular, (equivalently, $\langle \Sigma x, x\rangle > 0$ for all nonzero $x \in \mathbb{R}^n$). We have the following conditions for nondegeneracy.

Theorem 2.3 (Nondegeneracy) *The following conditions are equivalent:*

(a) Σ *is singular.*

(b) $c \cdot (Z_N - N\overline{Z})$ *is bounded for some nonzero* $c \in \mathbb{R}^n$*.*

(c) There is a G-irreducible subspace $V \subset \mathbb{R}^n$ *such that* $(Z_N - N\overline{Z})|_V$ *is bounded.*

(d) There is a nontrivial G-irreducible subspace $V \subset \mathbb{R}^n$ *such that the restricted isometries* $\gamma(i)|_V$ *have a common fixed point, or there is a trivial (hence one-dimensional) G-irreducible subspace on which the restricted isometries (translations) are all equal.*

Set $W_N(0) = 0$, and

$$W_N(t) = \tfrac{1}{\sqrt{N}}(Z_{Nt} - Nt\overline{Z}), \quad t = 1/N, 2/N, \ldots$$

Linearly interpolating on each interval $[(r-1)/N, r/N]$, $r \geq 1$, we obtain a sequence of random elements $W_N \in C([0,\infty), \mathbb{R}^n)$.

The weak invariance principle (which is a refinement of the CLT) states that the sequence $\{W_N\}$ converges weakly to Brownian motion. Recall that a stochastic process $W : [0,\infty) \times \Omega \to \mathbb{R}^n$ is called an n-dimensional Brownian motion if (i) $W(0) = 0$ almost surely, (ii) there is an $n \times n$ covariance matrix Σ such that $W(t)$ has distribution $N(0, t\Sigma)$ for each $t \geq 0$, and (iii) for each $0 \leq t_1 < t_2 < \cdots < t_k$, the increments $W(t_1)$, $W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})$ are independent random vectors. It is a basic property of Brownian motion that the random element W lies almost surely in $C([0,\infty), \mathbb{R}^n)$.

Theorem 2.4 (Weak Invariance Principle (WIP)) *The sequence* $\{W_N\}$ *converges weakly in* $C([0,\infty), \mathbb{R}^n)$ *to an n-dimensional Brownian motion with covariance* Σ*. (In other words, the measures induced by* W_N *on* $C([0,\infty), \mathbb{R}^n)$ *converge weakly to an n-dimensional Wiener measure.)*

Given $c \in \mathbb{R}^n$, we define $\sigma_c^2 = c^T \Sigma c$. Theorem 2.2 is equivalent to the statement that $\frac{1}{\sqrt{N}} c \cdot (Z_N - N\overline{Z})$ converges in distribution to a one-dimensional normal distribution with mean zero and variance σ_c^2. The analogous comment applies to Theorem 2.4. We also have the following almost sure result.

Theorem 2.5 (Law of the Iterated Logarithm (LIL)) *With probability one, for all* $c \in \mathbb{R}^n$

$$\limsup_{N \to \infty} c \cdot (Z_N - N\overline{Z}) / \sqrt{2N \log\log N} = \sigma_c.$$

Recall that $\gamma(i) = (h(i), k(i))$ where $h(i) \in \mathbf{O}(n)$ and $k(i) \in \mathbb{R}^n$. Let E denote expectation with respect to the probability vector $(p(1), ..., p(d))$. So for example, $E(h) = \sum_{i=1}^d p(i)h(i)$.

Write $\mathbb{R}^n = \mathrm{Fix}(G) \oplus \mathrm{Fix}(G)^\perp$. Since Σ commutes with the action of G, we have the corresponding direct sum $\Sigma = \Sigma_1 \oplus \Sigma_2$. Write $h(i) = h(i)_1 \oplus h(i)_2$ and so on.

We define a projection Π on the space of $n \times n$ matrices by $\Pi(Y) = \frac{1}{2}(\int_G gYg^T d\nu + \int_G gY^Tg^T d\nu)$. Note that Π averages Y over G and symmetrises Y, so Π projects onto the space of symmetric matrices that commute with the action of G. In particular, $\Pi(\Sigma) = \Sigma$.

Theorem 2.6 (Formula for Σ)

$$\Sigma_1 = E\big((k_1 - \overline{k}_1)(k_1 - \overline{k}_1)^T\big)$$
$$\Sigma_2 = \Pi\{E(k_2k_2^T) + 2(I - E(h_2))^{-1}E(k_2)E(k_2^Th_2)\}.$$

(a) Irreducible actions of G

The statement of our main results is greatly simplified when G acts irreducibly on $\mathbb{R}^n$. This is the typical situation, in the sense that if $n \geq 2$ and d is large enough, then most d-tuples $\{h(1), \ldots, h(d)\}$ will generate the group $G = \mathbf{SO}(n)$ or the group $G = \mathbf{O}(n)$ both of which which certainly act irreducibly on $\mathbb{R}^n$.

More precisely, we claim that for sufficiently large d, a residual subset of d-tuples in $\mathbf{O}(n)^d$ will generate $\mathbf{SO}(n)$ or $\mathbf{O}(n)$. It is clear that $d = 2$ suffices when $n = 2$. It follows from [2] that $d = 3$ suffices for any $n \geq 3$. Moreover, when $n \geq 3$, the set of generating d-tuples is open and dense [10] and even Zariski open [5].

When G fails to act irreducibly on $\mathbb{R}^n$, we can decompose $\mathbb{R}^n = V_1 \oplus \cdots \oplus V_p$ into G-irreducible subspaces. Let π_j be the orthogonal projection onto V_j. Then $\{\pi_j Z_N\}$ is identical to the sequence $\{Z_N\}$ obtained by replacing $k(i)$ with $\pi_j k(i)$ and Z_0 with $\pi_j Z_0$. Hence, for many purposes, we may suppose from the outset that G acts irreducibly on $\mathbb{R}^n$.

If $n = 1$ and $G = \mathbf{1}$, then $\{Z_N\}$ is a simple random walk on $\mathbb{R}$. We exclude this special case and assume that G acts irreducibly and nontrivially on $\mathbb{R}^n$. Immediate consequences of this assumption are that $\mathrm{Fix}(G) = \{0\}$, $\overline{Z} = 0$ and $\Sigma = \sigma^2 I_n$. A summary of our main results for such actions of G is as follows.

Theorem 2.7 *Suppose that G acts irreducibly and nontrivially on $\mathbb{R}^n$. Then*

(a) $\lim_{N\to\infty} \frac{1}{N} Z_N = 0$ *with probability one.*

(b) $\{\frac{1}{\sqrt{N}} Z_N\}$ *converges in distribution to an n-dimensional normal distribution with mean zero and variance $\sigma^2 I_n$.*

(c) *$\sigma = 0$ if and only if the isometries $\gamma(i)$ have a common fixed point.*

(d) *The sequence $\{W_N\}$ converges weakly to an n-dimensional Brownian motion with covariance $\sigma^2 I_n$.*

(e) *With probability one, for all $c \in \mathbb{R}^n$, $\limsup_{N\to\infty} c \cdot Z_N/\sqrt{2N\log\log N} = \sigma$.*

(f) *$\sigma^2 = \frac{1}{n}\Big\{E|k|^2 + 2\,\mathrm{tr}\big((I - E(h))^{-1}E(k)E(k^T h)\big)\Big\}$.*

Remark 2.8 This result is an immediate consequence of our main results. In particular, to prove part (f), note that $\Sigma_2 = \Sigma = \sigma^2 I_n$ in Theorem 2.6. Taking traces on both sides and dividing by n yields the required result. We note that this formula for the variance in the irreducible case was obtained by Gorostiza [7].

(b) Markov chains

Fix a finite set of isometries $\{\gamma(1), \ldots, \gamma(d)\} \subset \mathbf{E}(n)$ as before. Let $P = \{P_{ij}\}$ be a $d \times d$ stochastic matrix consisting of entries $P_{ij} \geq 0$ with row sums $\sum_{j=1}^d P_{ij} = 1$ for all i. We assume that P is irreducible: for each (i,j) there exists an integer $n \geq 1$ such that the (i,j)'th entry of P^n is positive.

We suppose that the random sequence $\{Z_N\}$ is defined in an analogous way to the i.i.d. case, with $Z_N = Y_N Y_{N-1} \cdots Y_2 Y_1 Z_0$ except that if $Y_{N-1} = \gamma(i)$, then Y_N is chosen to be $\gamma(j)$ with probability P_{ij}. (Since we are interested in asymptotics, the method of choosing Y_1 is unimportant.)

If $P_{ij} > 0$ for all $1 \leq i, j \leq d$, then our main results go through without any change except for the formulas for the mean $\overline{Z}$ and the variance Σ. In fact $\overline{Z} = \pi \sum_{j=1}^d q(j)k(j)$ where $q = (q(1), \ldots, q(d))$ is the unique probability vector ($q(j) > 0$ and $\sum_{j=1}^d q(j) = 1$) satisfying $P^T q = q$. We do not attempt to give an explicit formula for Σ.

If some of the P_{ij} are zero, we need an additional hypothesis. A *word* W is a finite sequence of symbols $W = j_1 j_2 \cdots j_m$, where each $j_r \in \{1, \ldots, d\}$. The word is *admissible* if $P_{j_r j_{r+1}} > 0$ for all $r = 1, \ldots, m-1$. An admissible word $W = j_1 j_2 \cdots j_m$ is *periodic* if in addition $P_{j_m j_1} > 0$. If $W = j_1 j_2 \cdots j_m$ is a periodic word, then we define $h(W) = h(j_1)h(j_2)\cdots h(j_m)$.

Fix $i \in \{1, \ldots, d\}$ and let $\mathcal{W}_i$ denote the set of all periodic words $W = j_1 \cdots j_m$ (ranging over all possible $m \geq 1$) satisfying $j_1 = i$. Let $\mathcal{G}_i$ be the smallest closed subgroup of G containing $\{h(W) : W \in \mathcal{W}_i\}$. It is easily seen that the subgroups $\mathcal{G}_i$, $i = 1, \ldots, d$, are conjugate in G.

Theorem 2.9 *Suppose that P defines an irreducible Markov chain. If $\mathcal{G}_i = G$ for some (and hence all) i, then Theorems 2.1–2.5 are valid, except that*

$$\overline{Z} = \pi \sum_{j=1}^{d} p(j)k(j),$$

where $q = (q(1), \ldots, q(d))$ is the unique probability vector ($q(j) > 0$ and $\sum_{j=1}^{d} q(j) = 1$) satisfying $P^T q = q$.

3 Reformulation as a dynamical system

We continue to suppose that $\{\gamma(1), \ldots, \gamma(d)\} \subset \mathbf{E}(n)$ is a finite set of isometries chosen at random. Write $\gamma(i) = (h(i), k(i))$ where $h(i) \in \mathbf{O}(n)$ and $k(i) \in \mathbb{R}^n$. Let G be the closed group generated by $h(1), \ldots, h(d)$ and assume that G acts irreducibly and nontrivially on $\mathbb{R}^n$. Let $\Gamma = G \ltimes \mathbb{R}^n \subset \mathbf{E}(n)$.

Let $X = \{1, 2, \ldots, d\}^{\mathbb{N}}$ be the space of one-sided sequences whose entries lie in $\{1, \ldots, d\}$ and let $T : X \to X$ be the full shift on d symbols. In the i.i.d. case, we define μ to be the Bernoulli measure induced by the probability vector $(p(1), \ldots, p(d))$. In the Markov case, we restrict to the subshift of finite type $X = X_P$ corresponding to P (so we only consider sequences $(j_0, j_1, \cdots)$ that are admissible in the sense that $P_{j_r j_{r+1}} > 0$ for all $r \geq 0$), and we let μ be the Markov measure induced by the stochastic matrix $P = \{P_{ij}\}$ (see for example [11, Theorem I.10.1]).

Let $\xi : X \to \{\gamma(1), \ldots, \gamma(d)\}$ be defined by $\xi(x) = \gamma(i)$ if $x_0 = i$. In other words, ξ is a Γ-valued random variable that depends only on the 0'th coordinate of x.

Form the product $X \times \Gamma$ and define the Γ-extension $T_\xi : X \times \Gamma \to X \times \Gamma$ by

$$T_\xi(x, \gamma) = (Tx, \xi(x)\gamma).$$

Then $T_\xi^N(x, \gamma) = (T^N x, \xi_N(x)\gamma)$ where

$$\xi_N(x) = \xi(T^{N-1}x) \cdots \xi(Tx)\xi(x).$$

Note that $Y_j = \xi \circ T^j$ is a sequence of random variables taking values in $\{\gamma(1), \ldots, \gamma(d)\}$ just as in Section 2. Moreover, if $\gamma = (e, Z_0)$, then we can reconstruct the random sequence Z_N from the formula

$$T_\xi^N(x, (e, Z_0)) = (T^N x, (G_N, Z_N)),$$

where G_N is the product of the first N matrices selected from $\{h(0), \ldots, h(d)\}$.

Hence the statistical properties of the sequence $\{Z_N\}$ can be recovered from the statistical properties of the Γ-extension T_ξ.

An equivalent problem is to study skew-products of the form $T_\xi(x,\gamma) = (x, \gamma\xi(x))$. In [13], we began a systematic investigation of such skew-products. The difference is only notational, but we shall proceed with the cocycle ξ acting on the right to make the connection with [6, 12, 13].

The first step is to make use of the semidirect product structure of $\Gamma = G \ltimes \mathbb{R}^n$. Write $\xi(x) = (h(x), k(x))$. Then $h(x) = h(i)$ whenever $x_0 = i$ and similarly for k, so again h and k are functions that depend only the 0'th coordinate. Writing $\gamma = (g, v)$, (where $g \in \mathbf{O}(n)$, $v \in \mathbb{R}^n$) we have

$$T_\xi(x, g, v) = (Tx, gh(x), v + gk(x)).$$

Defining $T_h : X \times G \to X \times G$ by $T_h(x,g) = (Tx, gh(x))$ and $\phi : X \times G \to \mathbb{R}^n$ by $\phi(x,g) = gk(x)$, we can rewrite the skew product in the form

$$T_\xi(x, g, v) = (T_h(x, g), v + \phi(x, g))$$

and so

$$T_\xi^N(x, g, v) = (T_h^N(x, g), v + \phi_N(x, g)),$$

where $\phi_N = \sum_{j=0}^{N-1} \phi \circ T_h^j$. Moreover,

$$Z_N(x) = Z_0 + \phi_N(x, e).$$

In this way, questions about the rate of growth of Z_N reduce to statistical questions about the observation $\phi : X \times G \to \mathbb{R}^n$ defined on the compact G-extension $X \times G$ [13].

4 Proofs

In this section, we prove our main results. We restrict throughout to the i.i.d. case, until we reach the proof of Theorem 2.9.

Let ν denote Haar measure on G, and define the product measure $m = \mu \times \nu$ on $X \times G$. By the definition of G, it is easily seen that

Proposition 4.1 *The G-extension $T_h : X \times G \to X \times G$ is ergodic with respect to m.*

Proof By Livšic regularity, it suffices to show that T_h is topologically transitive (cf. [14, Corollary 4.5]). Let $U = U_X \times U_G$ and $V = V_X \times V_G$ denote open sets in $X \times G$. Choose $g_0 \in U_G$. We show that there exists $x \in U_X$ such that $\sigma_h^j(x, g_0) \in V$ for some $j \geq 1$.

Given a word $W = x_1 x_2 \cdots x_n$, we let C_W denote the corresponding cylinder set in X and we define $h(W) = h(x_1) \cdots h(x_n)$. Note that $h(W) = h_n(x)$ for all $x \in C_W$.

In the i.i.d. case, X is a full shift so all words are admissible. Fix a word W_1 of length n say such that $C_{W_1} \in U_X$. Let $g' = h(W_1)$ and choose $g'' = h(i_1)h(i_2)\cdots h(i_m)$ such that $g_0 g' g'' \in V_G$. (This can be done since the $h(i)$ generate a dense subset of G.) Define $W_2 = i_1 i_2 \cdots i_m$ to be the corresponding word in X so that $h(W_2) = g''$. Finally choose a word W_3 so that $C_{W_3} \in V_X$

Now let $x \in C_{W_1 W_2 W_3}$. It follows from the choice of W_1 and W_3 that $x \in U_X$ and $\sigma^{n+m} x \in V_X$. Moreover $\sigma_h^{n+m}(x, g_0) = (\sigma^{n+m} x, g_0 h_{n+m}(x)) = (\sigma^{n+m} x, g_0 h(W_1 W_2)) = (\sigma^{n+m} x, g_0 g' g'') \in V_X \times V_G$ as required. ∎

Proof of Theorem 2.1 Let $\overline{\phi} = \int_{X \times G} \phi \, dm$. By the ergodic theorem, $\phi_N / N \to \overline{\phi}$ as $N \to \infty$ for almost every $(x, g) \in X \times G$.

From the equivariance condition $\phi(x, g) = g k(x)$, it follows as in [13] that

$$\overline{\phi} = \int_G g \, d\nu \int_X k(x) \, d\mu(x) = \pi \int_X k \, d\mu.$$

But $k(x) = k(x_0)$ and so $\int_X k \, d\mu = \sum_{i=1}^d p(i) k(i)$. Hence $\overline{\phi} = \overline{Z}$. Moreover, it follows from equivariance that $\phi_N / N \to \overline{Z}$ for all $g \in G$, for almost every $x \in X$. In particular, $\phi_N(x, e)/N \to \overline{Z}$ for almost every $x \in X$.

Finally, $Z_N(x) = Z_0 + \phi_N(x, e)$ and the result follows. ∎

The proofs of Theorems 2.2–2.5 are now a straightforward application of results in [6, 12]. For completeness, we sketch the main ideas.

Standard techniques enable us to pass between G-extensions of one-sided and two-sided shifts on X. First suppose for simplicity that G is a connected compact Lie group and that the product measure m on $X \times G$ is not only ergodic (as guaranteed by Proposition 4.1) but also weak mixing. In the one-sided case, Field *et al.* [6] used spectral properties of the equivariant Ruelle operator [15] to obtain a martingale approximation for the sequence $\{\phi_N\}$ defined in Section 3. Results of Billingsley [3, 4] then yield the CLT and WIP for the sequence $\{\phi_N\}$ on $X \times G$ (with respect to the probability measure $\mu \times \nu$). Similarly, the LIL for $\{\phi_N\}$ is a consequence of the almost sure invariance principle (ASIP) established in [6]. Nicol *et al.* [13] proved the equivalence of conditions (a)–(c) in Theorem 2.3. The equivalence of condition (d) follows from [1].

It follows from Melbourne and Nicol [12] that the CLT, WIP and LIL for the sequence $\{\phi_N\}$ on $X \times G$ hold equally for the sequence $Z_N(x) = Z_0 + \phi_N(x, e)$ defined on X. It is also shown in [12] that the mixing hypothesis of [6] can be weakened to ergodicity and that G need not be connected. This completes the proof of Theorems 2.2–2.5.

Finally, we derive the expression for Σ.

Proof of Theorem 2.6 The proof generalises an argument of [7] in the irreducible case. Recall [6, 12] that

$$\Sigma = \lim_{N\to\infty} \frac{1}{N} \int_{X\times G} \phi_N \phi_N^T dm.$$

We can divide the proof up into the cases $G = \{\mathbf{1}\}$ and $\mathrm{Fix}(G) = \{0\}$.

Suppose first that $G = \{\mathbf{1}\}$ and define $\widehat{k} = k - \overline{k}$. Then $\Sigma = \lim_{N\to\infty} \frac{1}{N} \int_X \widehat{k}_N \widehat{k}_N^T dm$, where $\widehat{k}_N = \widehat{k} + \widehat{k}\circ T + \cdots + \widehat{k}\circ T^{N-1}$. It is well known (see for example [6]) that the full shift on d symbols exhibits exponential decay of correlations and that as a consequence

$$\Sigma = E(\widehat{k}\,\widehat{k}^T) + \sum_{j=1}^{\infty} E(\widehat{k}\circ T^j\,\widehat{k}^T) + \sum_{j=1}^{\infty} E(\widehat{k}\,\widehat{k}^T\circ T^j).$$

By independence, $E(\widehat{k}\circ T^j\,\widehat{k}^T) = E(\widehat{k}\circ T^j)E(\widehat{k})^T = 0$. Similarly, $E(\widehat{k}\,\widehat{k}^T\circ T^j) = 0$. Hence, $\Sigma = E(\widehat{k}\,\widehat{k}^T)$ as required.

Next, we consider the case $\mathrm{Fix}(G) = \{0\}$. We make the following observations

(a) $I - E(h)$ is invertible (cf. [7, Lemma 1]).

(b) If A is a nonsingular matrix and $\|A\| \le 1$, then

$$\sum_{r=1}^{N-1} \frac{N-r}{N} A^{r-1} = (I-A)^{-1} - \frac{1}{N}(I-A^N)(I-A)^{-2}.$$

To see that (a) is true, note that $E(h) = p_1h_1 + p_2h_2 + \cdots + p_dh_d$ where h_j is orthogonal. Assume that $I - E(h)$ is singular, that is that there exists $x \ne 0$ such that $(I - E(h))x = x$, or $E(h)x = x$. This is equivalent to $p_1h_1x + \cdots + p_dh_dx = x$. We have $|p_jh_jx| = p_j|x|$, so the individual lengths of the vectors p_jh_jx sum exactly to the length of x. So we can only have $E(h)x = x$ if all the h_jx have the same direction, which means that $h_jx = x$ for every j. In other words $x \in \mathrm{Fix}(G) = \{0\}$ contradicting the fact that $x \ne 0$. Statement (b) can be proved by induction.

We have $\Sigma = \lim_{N\to\infty} \dfrac{1}{N} \displaystyle\int \phi_N\phi_N^T$ and

$$\begin{aligned}
\frac{1}{N}\int \phi_N\phi_N^T &= \int \phi\phi^T + \frac{1}{N}\Big(\sum_{i>j}^{N-1} \int (U^i\phi)(U^j\phi)^T + \sum_{i<j}^{N-1} \int (U^i\phi)(U^j\phi)^T\Big) \\
&= \int \phi\phi^T + \frac{1}{N}\Big(\sum_{i>j}^{N-1} \int (U^{i-j}\phi)\phi^T + \sum_{i<j}^{N-1} \int \phi(U^{j-i}\phi)^T\Big) \\
&= \int \phi\phi^T + \sum_{r=1}^{N-1} \frac{N-r}{N} \int (U^r\phi)\phi^T + \sum_{r=1}^{N-1} \frac{N-r}{N} \int \phi(U^r\phi)^T,
\end{aligned}$$

where $U^j\phi = \phi\circ T_h^j$ and integrals are over $X\times G$ with respect to the product measure $m=\mu\times\nu$.

Note that $\int \phi\phi^T dm = \int_G g(\int_X k\,k^T d\mu)g^T d\nu = \Pi E(k\,k^T)$. A more complicated calculation gives

$$\begin{aligned}
\int (U^r\phi)\phi^T dm &= \int_{X\times G} g\,\phi\circ T_h^r\,\phi^T g^T dm \\
&= \int_X \Big(\int_G gh(x)h(Tx)\cdots h(T^{r-1}x)k(T^r x)\,k(x)^T g^T d\mu\Big)d\nu \\
&= \int_X \Big(\int_G gh(Tx)\cdots h(T^{r-1}x)k(T^r x)\,k(x)^T h(x) g^T d\mu\Big)d\nu \\
&= \widetilde{\Pi}\int_X h(Tx)\cdots h(T^{r-1}x)k(T^r x)\,k(x)^T h(x)d\mu \\
&= \widetilde{\Pi}\big(E(h)^{r-1}E(k)E(k^T h)\big),
\end{aligned}$$

where $\widetilde{\Pi}(Y) = \int_G gYg^T d\nu$ and we have used independence in the last line. By (a) and (b),

$$\sum_{r=1}^{N-1}\frac{N-r}{N}E(h)^{r-1}\to (I-E(h))^{-1},$$

and so

$$\sum_{r=1}^{N-1}\frac{N-r}{N}\int (U^r\phi)\phi^T \to \widetilde{\Pi}\big((I-E(h))^{-1}E(k)E(k^T h)\big).$$

Similarly

$$\sum_{r=1}^{N-1}\frac{N-r}{N}\int \phi(U^r\phi)^T \to \widetilde{\Pi}\big((I-E(h))^{-1}E(k)E(k^T h)\big)^T.$$

Thus we obtain the required formula for Σ. ∎

Proof of Theorem 2.9 The main step is to verify that Proposition 4.1 is still valid. Suppose that this is the case. The computation of $\overline{Z}$ differs in the Markov case only in the computation of $\int_X k\,d\mu$. To obtain the required formula for $\overline{Z}$, it suffices to observe that (by definition of the Markov measure μ [11]) the probability distribution of x_0 is given by the probability vector q associated to the defining stochastic matrix P.

The proofs of Theorem 2.2, 2.4 and 2.5 are identical to before, since the results in [6, 12] apply equally well to full shifts and to subshifts of finite type.

It remains to verify that Proposition 4.1 is still valid. Choose (admissible) words W_1 and W_3 as before. Without loss of generality, we can augment W_1 (which shrinks the cylinder C_{W_1}) so that W_1W_3 is admissible. Let i be the first symbol in W_3. If $U_1, \cdots, U_m \in \mathcal{W}_i$ and $k_1, \ldots, k_m \geq 1$, then $W_1 U_1^{k_1} \ldots U_m^{k_m} W_3$ is admissible. It suffices to show that $U_1, \cdots, U_m \in \mathcal{W}_i$ and $k_1, \ldots, k_m \geq 1$ can be chosen so that $h(W_1)h(U_1)^{k_1} \cdots h(U_m)^{k_m}$ generates G.

Since $\mathcal{G}_i$ generates G, we can arrange that $h(W_1)h(U_1)^{k_1} \cdots h(U_m)^{k_m}$ is as close to the identity as desired (with $k_1 = \cdots = k_m = 1$). Hence, we may as well suppose that $h(W_1) = e$. Now, choose U_r so that $h(U_1), \ldots, h(U_m)$ generates G. For any $r = 1, \ldots, m$, we can choose k_r so that $h(U_r)^{k_r}$ is close to the identity. Do this for $m-1$ of the k_r and choose the remaining $k_r = 1$, to obtain the group element $h(U_r)$. Choices of this type yield all the $h(U_r)$ generating G as required. ∎

Remark 4.2 (a) There is an alternative proof of Theorem 2.1 in the Markov chain case based on [9] where it is shown that $X \times G$ is not ergodic, if and only if there is a nontrivial irreducible representation R of G on $\mathbb{C}^m$ for some $m \geq 1$ and a measurable function $v : X \to \mathbb{C}^m$ with $|v| = 1$ a.e. such that

$$R(h(x))v(Tx) = v(x), \text{ a.e.}$$

Iterating, we have

$$R(h_n(x))v(T^n x) = v(x), \tag{4.1}$$

where (as usual) $h_n(x) = h(x)h(Tx) \cdots h(T^{n-1}x)$.

It follows from Livšic regularity [14, Theorem 3.1] that it is sufficient to exclude continuous solutions v to equation (4.1) in order to establish ergodicity. In particular, we can evaluate (4.1) on periodic solutions to conclude that if W is a periodic word, then

$$R(h(W))v(W^\infty) = v(W^\infty). \tag{4.2}$$

Suppose that y is a sequence, W is periodic of period k, and Wy is admissible. Then $y_r = W^r y$ is admissible. Taking $n = kr$ and $x = y_r$ in (4.1), we obtain $R(h(W))^r v(y) = v(y_r)$. But $y_r \to W^\infty$ and we can pass to a subsequence so that $R(h(W))^r \to I$, leading to the fact that $v(y) = v(W^\infty)$. Since the only restriction on y is that Wy is admissible, we conclude that (like h), $v(y)$ depends only on the zero'th symbol in y. Hence (4.2) implies that $R(h(W))v(i) = v(i)$ for all W periodic with Wi admissible. This certainly includes all $W \in \mathcal{W}_i$. Since $\mathcal{G}_i = G$, we conclude that $R(g)v(i) = v(i)$ for all $g \in G$. But R is a nontrivial irreducible representation of G so $v(i) = 0$ which is a contradiction. Hence $X \times G$ is ergodic as required.

(b) The method in part (a) of this remark can be used to show that our hypotheses on the $\mathcal{G}_i$ are necessary as well as sufficient for ergodicity of $X \times G$. To see this, take $i = 1$

say and suppose that $\mathcal{G}_1 \neq G$. We construct a nontrivial irreducible representation R of G and a function $v : X \to \mathbb{C}^m$, $|v| = 1$, depending only on zero'th coordinates, such that $R(h)v \circ T = v$. By [9], $X \times G$ is not ergodic.

Choose an irreducible representation of G in which $\mathrm{Fix}(\mathcal{G}_1) \neq \{0\}$ and set $v(1)$ to be any unit vector in $\mathrm{Fix}(\mathcal{G}_1)$. If $2 \leq i \leq d$, then by irreducibility of the Markov chain, we can choose a word U_i such that iU_i1 is admissible. Define

$$v(i) = h(iU_i)v(1).$$

This completes the construction of R and v. It remains to verify that $R(h)v \circ T = v$. Since v and h depend only on zero'th coordinates, it is sufficient to verify that

$$h(i_1)v(i_2) = v(i_1),$$

whenever i_1i_2 is admissible.

First, choose a word $\widetilde{U}_i$ such that $1\widetilde{U}_i i$ is admissible. Since $1\widetilde{U}_i iU_i \in \mathcal{W}_1$, we have $h(1\widetilde{U}_i)v(i) = h(1\widetilde{U}_i iU_i)v(1) = v(1)$. Hence,

$$v(i) = h(1\widetilde{U}_i)^{-1}v(1).$$

Now suppose that i_1i_2 is admissible. Then

$$h(i_1)v(i_2) = h(i_1i_2U_{i_2})v(1) = h(1\widetilde{U}_{i_1})^{-1}h(W)v(1),$$

where $W = 1\widetilde{U}_{i_1}i_1i_2U_{i_2} \in \mathcal{W}_1$. Hence

$$h(i_1)v(i_2) = h(1\widetilde{U}_{i_1})^{-1}v(1) = v(i_1),$$

as required.

(c) In the case that G is abelian, the hypothesis on $\mathcal{G}_i$ reduces to the usual "periodic data" condition, namely that $\{h(W)\}$ generates G where W ranges over all periodic words.

Acknowledgments MN would like to Umeä University and the Mittag Leffler Institute for their hospitality while this work was undertaken. IM was supported in part by NSF Grant DMS-0071735.

References

[1] A. Ambroladze and M. Adahl. Random iteration of isometries in unbounded metric spaces. In preparation.

[2] H. Auerbach. Sur les groupes linéaire bornés (III). *Studia Mat* **V** (1934) 43–49.

[3] P. Billingsley. The Lindeberg-Lévy Theorem for martingales. *Proc. Amer. Math. Soc.* **12** (1961) 788–792.

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

[5] M. J. Field. Generating sets for compact semisimple Lie groups. *Proc. Amer. Math. Soc.* **127** (1999) 3361–3365.

[6] M. Field, I. Melbourne, and A. Török. Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. To appear in *Ergod. Th. & Dynam. Sys.*

[7] L. Gorostiza. The central limit theorem for random motions of d-dimensional Euclidean space. *Ann. Probability* **1** (1973) 603–612.

[8] L. Gorostiza. Limit theorem for certain random motions of $\mathbb{R}^d$. *Bol. Soc. Mat. Mexicana* **21** (1976) 52–61.

[9] H. B. Keynes and D. Newton. Ergodic measures for nonabelian compact group extensions. *C*ompositio Mathematica **32** (1976) 53–70.

[10] M. Kuranishi. Two element generations on semi-simple Lie groups. *Kodai Math. Sem. Report* (1949) 9–10.

[11] R. Mañé. *Ergodic Theory and Differentiable Dynamics.* Springer, New York, 1987.

[12] I. Melbourne and M. Nicol. Statistical properties of endomorphisms and compact group extensions. Preprint.

[13] M. Nicol, I. Melbourne, and P. Ashwin. Euclidean extensions of dynamical systems. *Nonlinearity* **14** (2001) 275–300.

[14] W. Parry. Skew products of shifts with a compact Lie group. *J. London Math. Soc.* **56** (1997) 395–404.

[15] W. Parry and M. Pollicott. *Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics.* Astérique **187-188**, Société Mathématique de France, Montrouge, 1990.

[16] B. Roynette. Théorème central-limite pour le groupe des déplacements de R^d. *Ann. Inst. H. Poincar Sect. B* **10** (1975) 391–398.

[17] N. Tutubalin. The central limit theorem for random motions of Euclidean space. *Vestnik Moskov. Univ. Ser. I Mat. Meh.* **22** (1967) 100–108.