Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 2, 726–734
DOI: 10.1214/14-AIHP649

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

A three-series theorem on Lie groups

Ming Liao

Department of Mathematics, Auburn University, Auburn, AL 36849, USA. E-mail: liaomin@auburn.edu

Received 10 June 2013; revised 20 September 2014; accepted 30 September 2014

Abstract. We obtain a necessary and sufficient condition for the convergence of independent products on Lie groups, as a natural extension of Kolmogorov's three-series theorem. Application to independent random matrices is discussed.

Résumé. Nous obtenons une condition nécessaire et suffisante pour la convergence de produits indépendants sur des groupes de Lie, comme extension naturelle du théorème des trois séries de Kolmogorov. Une application à des matrices aléatoires indépendantes est discutée.

MSC: 60B15

Keywords: Lie groups; Three-series theorem

1. Introduction and main results

Let x_n be a sequence of independent real-valued random variables. Fix any constant $r > 0$. Kolmogorov's three-series theorem (see for example [1, Theorem 22.8]) states that the series $\sum_{n=1}^{\infty} x_n$ converges almost surely if and only if the following three conditions hold.

(K1) $\sum_{n=1}^{\infty} P(|x_n| > r) < \infty$;
(K2) $\sum_{n=1}^{\infty} E(x_n 1_{[|x_n|\le r]})$ converges, where 1_A is the indicator of a set A; and
(K3) $\sum_{n=1}^{\infty} E[(x_n 1_{[|x_n|\le r]} - b_n)^2] < \infty$, where $b_n = E(x_n 1_{[|x_n|\le r]})$.

Extensions of the three-series theorem to more general spaces have been explored in literature. In particular, Maksimov [6] obtained a one-sided extension of the three-series theorem to Lie groups, providing a set of sufficient conditions for the convergence of products of independent random variables in a Lie group, with some partial result toward the more difficult necessity part.

The purpose of this paper is to present a complete extension of the three-series theorem to a general Lie group. Our result is a simpler form of a conjecture proposed in [6], and is in more close analogy with the classical result. We not only establish the more difficult necessity part, the proof of sufficiency is also much shorter than [6]. The result will be applied to study the convergence of products of independent random matrices.

Let G be a Lie group of dimension d with identity element e. There are a relatively compact neighborhood U of e and a smooth function $\phi = (\phi_1, \phi_2, \dots, \phi_d): U \to \mathbb{R}^d$ which maps U diffeomorphically onto a convex neighborhood $\phi(U)$ of the origin 0 in $\mathbb{R}^d$, with $\phi(e) = 0$. The U is not assumed to be open and ϕ is assumed extendable to be a smooth function on an open set containing the closure $\overline{U}$ of U. In the rest of the paper, U and ϕ are fixed, but they may be chosen arbitrarily as long as the above properties are satisfied.

Let x be a random variable in G. Its U-truncated mean b is defined by

$$\phi(b) = E\big[\phi(x) 1_{[x\in U]}\big]. \tag{1}$$

Note that because $\phi(U)$ is convex, $E[\phi(x)1_{[x\in U]}]\in\phi(U)$ and $b=\phi^{-1}\{E[\phi(x)1_{[x\in U]}]\}$.

Theorem 1. *Let x_n be a sequence of independent G-valued random variables with U-truncated means b_n. Then $\hat{x}_n = x_1x_2\cdots x_n$ converges almost surely in G as $n\to\infty$ if and only if the following three conditions hold.*

(G1) $\sum_{n=1}^{\infty} P(x_n\in U^c)<\infty$, *where U^c is the complement of U in G*;
(G2) $\hat{b}_n = b_1b_2\cdots b_n$ *converges in G as $n\to\infty$; and*
(G3) $\sum_{n=1}^{\infty} E[\|\phi(x_n)1_{[x_n\in U]}-\phi(b_n)\|^2]<\infty$, *where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$.*

Note that under (G1), (G3) is equivalent to $\sum_{n=1}^{\infty} E[\|\phi(x_n)-\phi(b_n)\|^2 1_{[x_n\in U]}]<\infty$.

The proof of Theorem 1 will begin in the next section. Note that by Kolmogorov's 0–1 law, the independent product $\hat{x}_n$ either converges almost surely or diverges almost surely.

When $G=\mathbb{R}^d$ as an additive group, one may take ϕ to be the identity map on $\mathbb{R}^d$ and U to be the ball of radius $r>0$ centered at 0, then Theorem 1 becomes precisely Kolmogorov's three-series theorem on $\mathbb{R}^d$.

We briefly comment on the relation between the almost sure convergence and the convergence in distribution. On Euclidean spaces, it is well known that the two convergences are equivalent for a series of independent random variables. This is not true for an independent product on a Lie group G. Because if G has a compact subgroup $H\neq\{e\}$, then for any sequence of independent random variables x_n, each is distributed according to the normalized Haar measure on H, the product $x_1x_2\cdots x_n$ converge in distribution to x_1, but it is clearly not convergent almost surely. By Theorem 2.2.16(ii) in Heyer [4], if the only compact subgroup of G is $\{e\}$, then the convergence in distribution and the almost sure convergence are equivalent for an infinite product of independent random variables in G.

For $k\geq 1$, let $\mathcal{M}_k$ be the space of $k\times k$ real matrices, which may be identified with $\mathbb{R}^d$, where $d=k^2$. The Euclidean norm of $x=\{x_{ij}\}\in\mathcal{M}_k$ is $\|x\|=\sqrt{\sum_{i,j}x_{ij}^2}$, and it satisfies $\|xy\|\leq\|x\|\|y\|$ for $x,y\in\mathcal{M}_k$.

Let G be the group of $k\times k$ real matrices of nonzero determinants under matrix product. Its identity element e is the identity matrix I. Its Lie algebra is $\mathcal{M}_k$ with the Lie group exponential map $\exp(x)$ being the usual matrix exponential $\mathrm{e}^x=I+\sum_{n=1}^{\infty}x^n/n!$.

Theorem 2. *Let G be the matrix group as above, and let x_n be a sequence of independent random variables in G. Fix $r\in(0,1)$. Then $\hat{x}_n=x_1x_2\cdots x_n$ converges almost surely to a random matrix in G if and only if the following three conditions hold.*

(M1) $\sum_{n=1}^{\infty}P(\|x_n-I\|>r)<\infty$;
(M2) $b_1b_2\cdots b_n$ *converges in G as $n\to\infty$, where* $b_n=I+E[(x_n-I)1_{[\|x_n-I\|\leq r]}]$; *and*
(M3) $\sum_{n=1}^{\infty}E(\|x_n-b_n\|^2 1_{[\|x_n-I\|\leq r]})<\infty$.

Proof. For $x\in G$, let $U=\{x\in G;\ \|x-I\|\leq r\}$ and $\phi(x)=x-I\in\mathcal{M}_k$. If $\|y\|<1$, then $I+y$ is invertible with $(I+y)^{-1}=I+\sum_{p=1}^{\infty}(-1)^p y^p$. It follows that ϕ maps U diffeomorphically onto the ball of radius r centered at 0 in $\mathcal{M}_k\equiv\mathbb{R}^d$, and hence ϕ and U satisfy the required properties. Theorem 1 may be applied with b_n in (M2) being the U-truncated mean of x_n. (G1) and (G2) are just (M1) and (M2), and (G3) is $\sum_n E[\|(x_n-I)1_{H_n}-(b_n-I)\|^2]<\infty$, where $H_n=[\|x_n-I\|\leq r]$. Because $E[\|(x_n-I)1_{H_n}-(b_n-I)\|^2]=E[\|x_n-b_n\|^2 1_{H_n}]+\|b_n-I\|^2P(H_n^c)$, by (M1), (G3) is equivalent to (M3). □

Example 1. *Let y_n be a sequence of independent random variables in $\mathcal{M}_k\equiv\mathbb{R}^d$, $d=k^2$. Assume $x_n=I+y_n$ is almost surely invertible. Note that this holds if y_n has a continuous distribution. Also assume that for some $r\in(0,1)$, $E(y_n 1_{[\|y_n\|\leq r]})=0$ for all n. Then $\hat{x}_n=x_1x_2\cdots x_n$ converges to an invertible random matrix x_∞ almost surely if*

$$\sum_{n=1}^{\infty}E\bigl(\|y_n\|^2\bigr)<\infty. \tag{2}$$

To prove this claim, note that b_n in (M2) *is I and* (M2) *holds trivially. Now* (M1) *is $\sum_{n=1}^{\infty}P(\|y_n\|>r)<\infty$ and* (M3) *is $\sum_{n=1}^{\infty}E[\|y_n\|^2 1_{[\|y_n\|\leq r]}]<\infty$. Because $P(\|y_n\|>r)\leq E(\|y_n\|^2)/r^2$, so* (M1) *and* (M3) *are implied by* (2). *By Theorem 2, $\hat{x}_n$ converges almost surely in the matrix group G.*

Example 2. *Let y_n be independent random variables in $\mathcal{M}_k \equiv \mathbb{R}^d$, $d = k^2$. Assume y_n is normal of mean 0. Then $\hat{x}_n = (I + y_1)\cdots(I + y_n)$ converges almost surely in the matrix group G if and only if (2) holds. To prove this, note that by the symmetry of a normal distribution, $E(y_n 1_{[\|y_n\| \le r]}) = 0$ for all $r > 0$. By Example 1, (2) is a sufficient condition for the almost sure convergence of $\hat{x}_n$ in G. To see it is also necessary, it suffices to show that (2) is implied by $\sum_n E[\|y_n\|^2 1_{[\|y_n\| \le r]}] < \infty$ and $\sum_n P(\|y_n\| > r) < \infty$. This can be done by an elementary computation of the normal distribution.*

Example 3. *Let y_n be a sequence of independent random variables in $\mathcal{M}_k \equiv \mathbb{R}^d$, $d = k^2$. Assume there is $r > 0$, which may be chosen arbitrarily small, such that $E(y_n 1_{[\|y_n\| \le r]}) = 0$ for all n. Then $\exp(y_1)\exp(y_2)\cdots\exp(y_n)$ converges in the matrix group G almost surely if (2) holds. To prove this, apply Theorem 1 to $x_n = \exp(y_n)$ with $\phi = \exp^{-1}$ on U, where U is the diffeomorphic image of a small ball in $\mathcal{M}_k \equiv \mathbb{R}^d$ under $\exp$. The conditions may be verified as in Example 1.*

2. Sufficiency

For any sequence of independent random variables x_n in G, by the Borel–Cantelli lemma, if (G1) holds, then almost surely, $x_n \in U$ except for finitely many n. On the other hand, if $\hat{x}_n = x_1 x_2 \cdots x_n$ converges almost surely, then because $x_n = \hat{x}_{n-1}^{-1}\hat{x}_n \to e$, (G1) follows from the Borel–Cantelli lemma. Set $x'_n = x_n$ on $[x_n \in U]$ and $x'_n = e$ on $[x_n \in U^c]$. Then the almost sure convergence of $x_1 x_2 \cdots x_n$ is equivalent to that of $x'_1 x'_2 \cdots x'_n$ and (G1). Note that $\phi(x_n)1_{[x_n \in U]} = \phi(x'_n) = \phi(x'_n)1_{[x'_n \in U]}$, and all quantities in (G2) and (G3) (including b_n) only depend on the restriction of x_n on U. Therefore, (G2) and (G3) hold for x_n if and only if they hold for x'_n. Thus, as noted in [6], to prove Theorem 1, we may, and will, assume all $x_n \in U$, and prove that $\hat{x}_n$ converges almost surely in G if and only if (G2) and (G3) hold.

We will prove the sufficiency part of Theorem 1 in this section, and so assume (G2) and (G3). Let μ_n be the distribution of x_n. Because $x_n \in U$, the U-truncated mean b_n of x_n is defined by $\phi(b_n) = \mu_n(\phi)$, where $\mu_n(\phi) = \int \phi \, d\mu_n = E[\phi(x_n)]$. Set $\hat{x}_0 = \hat{b}_0 = e$. For $n \ge 1$, let $z_n = \hat{b}_{n-1} x_n b_n^{-1} \hat{b}_{n-1}^{-1}$ and $\hat{z}_n = z_1 z_2 \cdots z_n$, and set $\hat{z}_0 = e$. It is easy to show by a simple induction on n that for all $n \ge 0$,

$$\hat{x}_n = \hat{z}_n \hat{b}_n. \tag{3}$$

By (G2), it suffices to show that $\hat{z}_n$ converges in G almost surely.

Note that for $G = \mathbb{R}^d$, z_n is just the centered term $x_n - b_n$, and $\hat{z}_n = \hat{x}_n - \hat{b}_n$ is the sum of the centered terms. To have $\hat{x}_n = \hat{z}_n \hat{b}_n$ on a noncommutative multiplicative Lie group G, z_n has to be defined in the above rather complicated form.

By the lemma below, the almost sure convergence of $\hat{z}_n$ is equivalent to $z_m z_{m+1} \cdots z_n \to e$ almost surely as $m \to \infty$ with $m < n$.

Lemma 3. *Let u_n be independent random variables in G. Then $u_1 u_2 \cdots u_n$ converges almost surely as $n \to \infty$ if and only if $u_m u_{m+1} \cdots u_n \to e$ almost surely as $m \to \infty$ with $m < n$.*

Proof. This is an easy consequence of the existence of a complete metric on G that is invariant under left translations and is compatible with the topology on G. The metric can be any left invariant Riemannian metric on G. □

For any $f \in C_c^\infty(G)$, the space of smooth functions on G with compact supports, let $M_0 f = f(e)$ and for $n \ge 1$, let

$$M_n f = f(\hat{z}_n) - \sum_{p=1}^{n} \int \left[f\left(\hat{z}_{p-1}\hat{b}_{p-1} x b_p^{-1} \hat{b}_{p-1}^{-1}\right) - f(\hat{z}_{p-1}) \right] \mu_p(dx). \tag{4}$$

Lemma 4. *Let $\mathcal{F}_n$ be the σ-algebra generated by $x_1, x_2, \ldots, x_n$. Then $E[M_n f \mid \mathcal{F}_m] = M_m f$ for $m < n$, that is, $M_n f$ is a martingale under the filtration $\{\mathcal{F}_n\}$.*

Proof. Because x_n are independent, for $m < p$,

$$
\begin{aligned}
&E\left[\int f\left(\hat{z}_{p-1}\hat{b}_{p-1}xb_p^{-1}\hat{b}_{p-1}^{-1}\right)\mu_p(\mathrm{d}x)\,\middle|\,\mathcal{F}_m\right] \\
&\quad = E\left[\int f\left(\hat{z}_m z_{m+1}\cdots z_{p-1}\hat{b}_{p-1}xb_p^{-1}\hat{b}_{p-1}^{-1}\right)\mu_p(\mathrm{d}x)\,\middle|\,\mathcal{F}_m\right] \\
&\quad = E\left[f(\hat{z}z_{m+1}\cdots z_{p-1}z_p)\right]\Big|_{\hat{z}=\hat{z}_m} = E\left[f(\hat{z}_p)\mid\mathcal{F}_m\right].
\end{aligned}
$$

Then $E[M_n f \mid \mathcal{F}_m] = M_m f$. □

Fix an integer $m > 0$ and a neighborhood V of e. Let $f \in C_c^\infty(G)$ be such that $0 \le f \le 1$, $f(e) = 1$ and $f(x) = 0$ for $x \in V^c$. For $g \in G$, let l_g be the left translation $x \mapsto gx$ on G, and let $f_m = f \circ l_{\hat{z}_m^{-1}}$. Let $\Lambda(m, V)$ be the event that there is $n > m$ such that $z_{m+1}z_{m+2}\cdots z_n \in V^c$. To estimate $P[\Lambda(m, V)]$, let τ be the first time $n > m$ such that $z_{m+1}z_{m+2}\cdots z_n \in V^c$ and set $\tau = \infty$ if $z_{m+1}z_{m+2}\cdots z_n \in V$ for all $n > m$. Then

$$
P\left[\Lambda(m, V)\right] = E\left\{\left[f_m(\hat{z}_m) - f_m(\hat{z}_\tau)\right]1_{\Lambda(m,V)}\right\} = \lim_{n\to\infty} E\left\{\left[f_m(\hat{z}_m) - f_m(\hat{z}_{\tau\wedge n})\right]1_{\Lambda(m,V)}\right\}, \tag{5}
$$

where $\tau \wedge n = \min(\tau, n)$. Because $E\{[f_m(\hat{z}_m) - f_m(\hat{z}_{\tau\wedge n})]1_{\Lambda(m,V)}\} \le E[1 - f_m(\hat{z}_{\tau\wedge n})] = E[f_m(\hat{z}_m) - f_m(\hat{z}_{\tau\wedge n})]$ and $E[M_{\tau\wedge n} f_m] = E\{E[M_{\tau\wedge n} f_m \mid \mathcal{F}_m]\} = E[M_m f_m]$,

$$
\begin{aligned}
&E\left\{\left[f_m(\hat{z}_m) - f_m(\hat{z}_{\tau\wedge n})\right]1_{\Lambda(m,V)}\right\} \\
&\quad \le -E\left\{\sum_{p=m+1}^{\tau\wedge n}\int\left[f\left(\hat{z}_{p-1}\hat{b}_{p-1}xb_p^{-1}\hat{b}_{p-1}^{-1}\right) - f(\hat{z}_{p-1})\right]\mu_p(\mathrm{d}x)\right\} \\
&\quad \le \sum_{p=m}^{\infty} E\left\{\left|\int\left[f\left(\hat{z}_{p-1}\hat{b}_{p-1}xb_p^{-1}\hat{b}_{p-1}^{-1}\right) - f(\hat{z}_{p-1})\right]\mu_p(\mathrm{d}x)\right|\right\}.
\end{aligned} \tag{6}
$$

We will write $\hat{z}, \hat{b}, b, \mu$ for $\hat{z}_{p-1}, \hat{b}_{p-1}, b_p, \mu_p$ for simplicity. For $x \in U$, by the Taylor expansion of $f(\hat{z}\hat{b}xb^{-1} \times \hat{b}^{-1}) = f(\hat{z}\hat{b}\phi^{-1}(\phi(x))b^{-1}\hat{b}^{-1})$ at $x = b$, noting $\mu(U^c) = 0$,

$$
\int\left[f\left(\hat{z}\hat{b}xb^{-1}\hat{b}^{-1}\right) - f(\hat{z})\right]\mu(\mathrm{d}x) = \int\left\{\sum_i f_i(\hat{z}, \hat{b}, b)\left[\phi_i(x) - \phi_i(b)\right]\right\}\mu(\mathrm{d}x) + r, \tag{7}
$$

where

$$
f_i(\hat{z}, \hat{b}, b) = \frac{\partial}{\partial\phi_i} f\left(\hat{z}\hat{b}\phi^{-1}\left(\phi(x)\right)b^{-1}\hat{b}^{-1}\right)\Big|_{x=b} \tag{8}
$$

and the remainder r satisfies $|r| \le c\mu(\|\phi - \phi(b)\|^2)$ for some constant $c > 0$. Because $\mu(\phi_i) = \phi_i(b)$, $\int[\phi_i(x) - \phi_i(b)]\mu(\mathrm{d}x) = 0$, and then by (7),

$$
\left|\int\left[f\left(\hat{z}\hat{b}xb^{-1}\hat{b}^{-1}\right) - f(\hat{z})\right]\mu(\mathrm{d}x)\right| = |r| \le c\mu\left(\left\|\phi - \phi(b)\right\|^2\right). \tag{9}
$$

It now follows from (5) and (6) that $P[\Lambda(m, V)] \le c\sum_{n=m}^\infty \mu_n(\|\phi - \phi(b_n)\|^2)$. Let $\varepsilon \in (0, 1)$ and let V_k be a sequence of neighborhoods of e with $V_k \downarrow \{e\}$ as $k \uparrow \infty$. By (G3), for each $k \ge 1$, there is an integer m_k such that $P[\Lambda(m_k, V_k)] < \varepsilon^k$. Then $\sum_{k=1}^\infty P[\Lambda(m_k, V_k)] \le \sum_{k=1}^\infty \varepsilon^k = \varepsilon/(1 - \varepsilon)$. By Lemma 3, $P(\hat{z}_n \text{ converges}) \ge P[\bigcap_{k=1}^\infty \Lambda(m_k, V_k)^c] \ge 1 - \sum_{k=1}^\infty P[\Lambda(m_k, V_k)] \ge 1 - \varepsilon/(1 - \varepsilon) \to 1$ as $\varepsilon \to 0$. This proves $\hat{z}_n$ converges almost surely.

3. Necessity, part 1

We will now prove (G2) and (G3) under the assumption that $\hat{x}_n$ converges almost surely and all $x_n \in U$. This proof is more complicated and will require another section.

Because $x_n = \hat{x}_{n-1}^{-1}\hat{x}_n \to e$ almost surely, by the Borel–Cantelli lemma,

$$\forall \text{ neighborhood } V \text{ of } e, \quad \sum_{n=1}^{\infty} P\big(x_n \in V^c\big) < \infty. \tag{10}$$

We also have

$$b_n \to e \quad \text{and} \quad \mu_n\big(\|\phi - \phi(b_n)\|^2\big) \to 0 \quad \text{as } n \to \infty. \tag{11}$$

For $m < n$, let $\hat{x}_{m,n} = x_{m+1}x_{m+2}\cdots x_n$ and $\hat{b}_{m,n} = b_{m+1}b_{m+2}\cdots b_n$. If either (G2) or (G3) does not hold, then there are a neighborhood V of e, $\varepsilon > 0$ and two sequences of integers m_k and n_k with $VV \subset U$, $m_k < n_k$ and $m_k \uparrow \infty$ as $k \uparrow \infty$ such that for each $k \geq 1$,

$$\text{either} \quad \sum_{p=m_k+1}^{n_k} \mu_p\big(\|\phi - \phi(b_p)\|^2\big) \geq \varepsilon \quad \text{or} \quad \hat{b}_{m_k,n_k} \in V^c.$$

Because of (11), by choosing m_1 large enough, we have $b_n \in V$ and $\mu_n(\|\phi - \phi(b_n)\|^2) \leq \varepsilon$ for $n > m_1$. Thus, by suitably reducing n_k, we obtain that for each $k \geq 1$, either

(i) $\varepsilon \leq \sum_{p=m_k+1}^{n_k} \mu_p(\|\phi - \phi(b_p)\|^2) \leq 2\varepsilon$, and $\hat{b}_{m_k,p} \in U$ for $m_k < p \leq n_k$; or
(ii) $\sum_{p=m_k+1}^{n_k} \mu_p(\|\phi - \phi(b_p)\|^2) \leq 2\varepsilon$, $\hat{b}_{m_k,n_k} \in V^c$, and $\hat{b}_{m_k,p} \in U$ for $m_k < p \leq n_k$.

We will derive a contradiction from either (i) or (ii) above. We will embed the partial products $x_{m_k,p}$ and $b_{m_k,p}$, for $m_k < p \leq n_k$, into a process $\tilde{x}_t^k$ and a function $\tilde{b}_t^k$ on [0, 1] respectively. The main idea is to obtain a martingale property for the process $\tilde{z}_t^k$, defined by $\tilde{x}_t^k = \tilde{z}_t^k\tilde{b}_t^k$, similar to the martingale property for $\hat{z}_n$ in the last section, to show the limit $\tilde{z}_t$ of $\tilde{z}_t^k$ satisfies an integral equation, and then to derive a contradiction. This is similar to the approaches in [3,5] for processes in Lie groups with independent increments.

Let γ_k be a strictly increasing function from $\{m_k, m_k+1, \ldots, n_k\}$ into [0, 1] with $\gamma_k(m_k) = 0$ and $\gamma_k(n_k) = 1$. Let $t_{k,p} = \gamma_k(p)$ for $m_k \leq p \leq n_k$. Then $t_{k,m_k} = 0$ and $t_{k,n_k} = 1$. Let $\tilde{x}_t^k = \tilde{b}_t^k = e$ for $0 \leq t < t_{k,m_k+1}$. For $m_k < p < n_k$ and $t_{k,p} \leq t < t_{k,p+1}$, let

$$\tilde{x}_t^k = \hat{x}_{m_k,p} \quad \text{and} \quad \tilde{b}_t^k = \hat{b}_{m_k,p}. \tag{12}$$

Set $\tilde{x}_t^k = \hat{x}_{m_k,n_k}$ and $\tilde{b}_t^k = \hat{b}_{m_k,n_k}$ for $t \geq 1$. Then $\tilde{x}_t^k$ and $\tilde{b}_t^k$ are respectively a step process and a step function, which are right continuous with jumps x_p and b_p at $t = t_{k,p}$.

Note that by Lemma 3, almost surely, $\tilde{x}_t^k \to e$ as $k \to \infty$ uniformly in t.

A continuous function $A(t) = \{A_{ij}(t)\}$ from $\mathbb{R}_+ = [0, \infty)$ to the space of $d \times d$ symmetric real matrices is called a covariance matrix function if $A(0) = 0$ and for $s < t$, $A(t) - A(s) \geq 0$ (nonnegative definite). Let

$$A_{ij}^k(t) = \sum_{0 < t_{k,p} \leq t} \int_G \big[\phi_i(x) - \phi_i(b_p)\big]\big[\phi_j(x) - \phi_j(b_p)\big]\mu_p(\mathrm{d}x). \tag{13}$$

Then $A^k(t) = \{A_{ij}^k(t)\}$ is almost a covariance matrix function except that it is not continuous, but $A^k(t) = A^k(1)$ for $t \geq 1$. Let $Q^k(t)$ be the trace of $A^k(t)$. Then

$$Q^k(t) = \sum_{0 < t_{k,p} \leq t} \mu_p\big(\|\phi - \phi(b_p)\|^2\big), \tag{14}$$

and for $s < t$,

$$\left|A^k_{ij}(t) - A^k_{ij}(s)\right| \le Q^k(t) - Q^k(s). \tag{15}$$

Note that $Q^k(t)$ is a nondecreasing step function in t with a jump $\mu_p(\|\phi - \phi(b_p)\|^2)$ at $t = t_{k,p}$, $Q^k(t) = 0$ for $0 \le t < t_{m_k,m_k+1}$ and $Q^k(t) = Q^k(1) = \sum_{m_k<p\le n_k} \mu_p(\|\phi - \phi(b_p)\|^2)$ for $t \ge 1$. By either (i) or (ii), $Q^k(t) \le 2\varepsilon$, and by (11), the jumps of $Q^k(t)$ converge to 0 uniformly in t as $k \to \infty$. It follows that the function γ_k may be chosen properly such that

$$Q^k(t) - Q^k(s) \le 2\varepsilon(t-s) + \varepsilon_k, \quad 0 \le s < t \le 1, \tag{16}$$

where $\varepsilon_k \to 0$ as $k \to \infty$. Roughly speaking, this means the functions $Q^k(t)$ are equi-continuous for large k. Because of (11), by either (i) or (ii), $n_k - m_k \to \infty$ as $k \to \infty$, and hence γ_k may be chosen to satisfy, besides (16),

$$\max_{p>m_k+1} (t_{k,p} - t_{k,p-1}) \to 0 \quad \text{as } k \to \infty. \tag{17}$$

Lemma 5. *There is a covariance matrix function $A(t)$ with $A(t) = A(1)$ for $t \ge 1$ such that along a subsequence of $k \to \infty$, $A^k(t) \to A(t)$ for any $t \ge 0$.*

Proof. Let Λ be a countable dense subset of $[0,1]$. Under either (i) or (ii), $Q^k(t)$ is bounded. By (15), along a subsequence of $k \to \infty$, $A^k(t)$ converges for any $t \in \Lambda$. By (16), the convergence holds for all $t \ge 0$, and $A(t)$ is continuous in t. □

Let Y be a smooth manifold equipped with a compatible metric ρ and let $y : [0,1] \to Y$ be a continuous function. For each k, let $y^k : [0,1] \to Y$ be a step function that is constant on $[t_{k,p-1}, t_{k,p})$ for each $p = m_k + 1, \ldots, n_k$. Assume for any $t > 0$, $\rho(y^k(t_{k,p}), y(t_{k,p})) \to 0$ as $k \to \infty$ uniformly for $t_{k,p} \le t$. Let $F(y,g) = \{F_{ij}(y,g)\}$ be a bounded continuous matrix-valued function on $Y \times G$.

Lemma 6. *Assume the above and let $A(t)$ be the covariance matrix function in Lemma 5. Then for any $t > 0$, along the subsequence of $k \to \infty$ in Lemma 5,*

$$\begin{aligned} &\sum_{0<t_{k,p}\le t} \sum_{i,j=1}^d \int_G F_{ij}\big(y^k(t_{k,p-1}), b_p\big)\big[\phi_i(x) - \phi_i(b_p)\big]\big[\phi_j(x) - \phi_j(b_p)\big]\mu_p(\mathrm{d}x) \\ &\quad \to \sum_{i,j=1}^d \int_0^t F_{ij}\big(y(s), e\big)\,\mathrm{d}A_{ij}(s). \end{aligned} \tag{18}$$

Proof. By the uniform convergence $\rho(y^k(t_{k,p}), y(t_{k,p})) \to 0$, $F(y^k(t_{k,p}), b) - F(y(t_{k,p}), b) \to 0$ as $k \to \infty$ uniformly for $t_{k,p} \le t$ and for b in a compact set. Because when $k \to \infty$, $b_p \to e$ uniformly for $p > m_k$, we may replace y^k and b_p by y and e in the proof.

Let $r > 0$ be an integer. For any two expressions A and B depending on (k, r), we will write $A \approx B$ if $|A - B| \to 0$ as $r \to \infty$ uniformly in k. Then

$$\begin{aligned} &\sum_{0<t_{k,p}\le t} \sum_{i,j=1}^d \int_G F_{ij}\big(y(t_{k,p-1}), e\big)\big[\phi_i(x) - \phi_i(b_p)\big]\big[\phi_j(x) - \phi_j(b_p)\big]\mu_p(\mathrm{d}x) \\ &\quad \approx \sum_{i,j=1}^d \sum_{q=0}^{r-1} \sum_{qt/r<t_{k,p}\le(q+1)t/r} \int_G F_{ij}\left(y\left(\frac{qt}{r}\right), e\right)\big[\phi_i(x) - \phi_i(b_p)\big]\big[\phi_j(x) - \phi_j(b_p)\big]\mu_p(\mathrm{d}x) \\ &\quad \left(\text{where } \sum_{qt/r<t_{k,p}\le(q+1)t/r} (\cdots) = 0 \text{ if } \left(\frac{qt}{r}, \frac{(q+1)t}{r}\right] \text{ contains no } t_{k,p}\right) \end{aligned}$$

$$\to \sum_{i,j=1}^{d}\sum_{q=0}^{r-1} F_{ij}\left(y\left(\frac{qt}{r}\right), e\right)\left[A_{ij}\left(\frac{(q+1)t}{r}\right) - A_{ij}\left(\frac{qt}{r}\right)\right] \quad (\text{as } k\to\infty, \text{ by Lemma 5})$$

$$\approx \sum_{i,j=1}^{d}\int_0^t F_{ij}\big(y(s), e\big)\,\mathrm{d}A_{ij}(s).$$

□

We now define a new process $\tilde{z}_t^k$, similar in the way as the sequence z_n is defined from x_n and b_n in Section 2, by setting $\tilde{z}_t^k = e$ for $0 \le t < t_{k,m_k+1}$, and inductively

$$\tilde{z}_t^k = \tilde{z}_{t_{k,p-1}}^k \tilde{b}_{t_{k,p-1}}^k x_p b_p^{-1}\big(\tilde{b}_{t_{k,p-1}}^k\big)^{-1} \tag{19}$$

for $t_{k,p} \le t < t_{k,p+1}$, $p = m_k+1,\dots,n_k$, setting $t_{k,n_k+1} = \infty$ here. Then $\tilde{z}_t = \tilde{z}_1$ for $t > 1$, and a simple induction on p shows that $\tilde{x}_t^k = \tilde{z}_t^k \tilde{b}_t^k$ for all $t \ge 0$.

For $f \in C_c^\infty(G)$, let $\tilde{M}_t^k f = f(\tilde{z}_t^k) = f(e)$ for $0 \le t < t_{k,m_k+1}$, and let

$$\tilde{M}_t^k f = f\big(\tilde{z}_t^k\big) - \sum_{0<t_{k,p}\le t}\int_G \big[f\big(\tilde{z}_{t_{k,p-1}}^k \tilde{b}_{t_{k,p-1}}^k x b_p^{-1}\big(\tilde{b}_{t_{k,p-1}}^k\big)^{-1}\big) - f(\tilde{z}_{t_{k,p-1}})\big]\mu_p(\mathrm{d}x), \tag{20}$$

for $t \ge t_{k,m_k+1}$.

Lemma 7. *$\tilde{M}_t^k f$ is a martingale under the natural filtration of process $\tilde{z}_t^k$.*

Proof. This is proved in the same way as in Lemma 4 for $M_n f$ to be a martingale. □

Because $\tilde{x}_t^k = \tilde{z}_t^k \tilde{b}_t^k$ and $\tilde{x}_t^k \to e$ uniformly in t as $k\to\infty$ almost surely, if $\tilde{b}_t^k$ converges to some continuous path $\tilde{b}_t$ in G uniformly in t as $k\to\infty$, then $\tilde{z}_t^k \to \tilde{z}_t = \tilde{b}_t^{-1}$ uniformly in t almost surely. This will be assumed in the rest of this section.

By a computation using Taylor expansion similar to the one in the last section, but up to the second order, noting the integrals of the first order terms vanish as before,

$$\tilde{M}_t^k f = f\big(\tilde{z}_t^k\big) - \sum_{0<t_{k,p}\le t}\sum_{i,j}\int_G f_{ij}\big(\tilde{z}_{t_{k,p-1}}^k, \tilde{b}_{t_{k,p-1}}^k, b_p\big)\big[\phi_i(x) - \phi_i(b_p)\big]\big[\phi_j(x) - \phi_j(b_p)\big]\mu_p(\mathrm{d}x) + r_k,$$

where

$$f_{ij}(\tilde{z}, \tilde{b}, b) = \frac{\partial^2}{\partial\phi_i\,\partial\phi_j} f\big(\tilde{z}\tilde{b}\phi^{-1}\big(\phi(x)\big)b^{-1}\tilde{b}^{-1}\big)\big|_{x=b},$$

and the reminder r_k may be divided into an integral over a small neighborhood V of e and an integral over V^c. The former is controlled by $c_V Q^k(t) \le c_V(2\varepsilon)$, where the constant $c_V \to 0$ as $V \downarrow \{e\}$, and the latter is controlled by $\sum_{m_k<p\le n_k}\mu_p(V^c)$ which converges to 0 as $k\to\infty$ by (10). Therefore, $r_k \to 0$ as $k\to\infty$. By Lemma 6 with $Y = G\times G$ and $y^k(t) = (\tilde{z}_t^k, \tilde{b}_t^k) \to y(t) = (\tilde{z}_t, \tilde{b}_t)$, it follows that $\tilde{M}_t^k f$ converges to the martingale

$$\tilde{M}_t f = f(\tilde{z}_t) - \sum_{i,j}\int_0^t f_{ij}(\tilde{z}_s, \tilde{b}_s, e)\,\mathrm{d}A_{ij}(s)$$

as $k\to\infty$. Because $\tilde{z}_t = \tilde{b}_t^{-1}$ is nonrandom, the martingale $\tilde{M}_t f$ must be $f(e)$, and then for any $f \in C_c^\infty(G)$ with $f(e) = 0$,

$$f(\tilde{z}_t) = \sum_{i,j}\int_0^t \left[\frac{\partial^2}{\partial\phi_i\,\partial\phi_j} f\big(\phi^{-1}\big(\phi(x)\big)\tilde{z}_s\big)\big|_{x=e}\right]\mathrm{d}A_{ij}(s). \tag{21}$$

Let t_0 be the largest nonnegative real number ≤ 1 such that $\tilde{z}_s = e$ and $A(s) = 0$ for $s \leq t_0$. We will show $t_0 = 1$. Suppose $t_0 < 1$. Then (21) holds for $t \geq t_0$ with $\int_0^t$ replaced by $\int_{t_0}^t$. Without loss of generality, we will assume $t_0 = 0$. Substitute $f = \phi_\beta^2$ in (21), then the integrand is $2\delta_{i\beta}\delta_{j\beta} + \varepsilon_s$, where ε_s denotes any function satisfying $\varepsilon_s \to 0$ as $s \to 0$. It follows that $\phi_\beta(\tilde{z}_t)^2 = 2A_{\beta\beta}(t) + \varepsilon_t T_t$, where $T_t = \mathrm{Tr}[A(t)]$. Then $\|\phi(\tilde{z}_t)\|^2 = 2T_t + \varepsilon_t T_t$. Now let $f = \phi_\beta$ and then (21) yields $\phi_\beta(\tilde{z}_t) = \varepsilon_t T_t$. This implies $|\phi_\beta(\tilde{z}_t)| \leq c\|\phi(\tilde{z}_t)\|^2$ for some constant $c > 0$, which is clearly impossible. This shows that $t_0 = 1$, and hence $\tilde{z}_t = e$ and $A(t) = 0$ for all $t \geq 0$.

If (i) holds, then $\mathrm{Tr}[A(1)] = \lim_k Q^k(1) = \lim_k \sum_{p=m_k+1}^{n_k} \mu_p(\|\phi - \phi(b_p)\|^2) \geq \varepsilon$, which contradicts to $A(t) = 0$. Thus (i) cannot hold. If (ii) holds, then $\tilde{b}_1 = \lim_k \tilde{b}_1^k = \lim_k \hat{b}_{m_k,n_k}$ belongs to the closure of V^c, which contradicts to $\tilde{b}_t = \tilde{z}_t^{-1} = e$. We have proved that neither (i) nor (ii) holds, and hence (G2) and (G3) must hold, under the assumption that $\tilde{b}_t^k \to \tilde{b}_t$ as $k \to \infty$ uniformly in t for some continuous path $\tilde{b}_t$ in G.

4. Necessity, part 2

It remains to show that $\tilde{b}_t^k \to \tilde{b}_t$ as $k \to \infty$ uniformly in t for some continuous path $\tilde{b}_t$ in G. A rcll path is a right continuous path with left limits, and a process with rcll paths will be called a rcll process. Let $D(G)$ be the space of rcll paths in G. Equipped with the Skorohod metric, $D(G)$ is a complete separable metric space (see [2], Chapter 3). A sequence of rcll processes y_t^k in G are said to converge weakly to a rcll process y_t if $y_\cdot^k \to y_\cdot$ in distribution as $D(G)$-valued random variables. The sequence y_t^k are called relatively weak compact in $D(G)$ if any subsequence has a further subsequence that converge weakly.

We will show that $\tilde{z}_t^k$ are relatively weak compact. Let V be a neighborhood of e. The amount of time it takes for a rcll process y_t to make V^c-displacement from a stopping time σ (under the natural filtration of process y_t) is denoted as τ_V^σ, that is,

$$\tau_V^\sigma = \inf\{t > 0;\ y_\sigma^{-1} y_{\sigma+t} \in V^c\} \quad \text{(inf of an empty set is } \infty). \tag{22}$$

For a sequence of processes y_t^k in G, let $\tau_V^{\sigma,k}$ be the V^c-displacement time for y_t^k from σ.

The following lemma is Lemma 16 in [5] and provides a criterion for the relative compactness. It is a slightly improved version of a lemma in [3].

Lemma 8. *A sequence of rcll processes y_t^k in G are relatively weak compact in $D(G)$ if for any constant $T > 0$ and any neighborhood V of e,*

$$\overline{\lim_{k\to\infty}} \sup_{\sigma \leq T} P\big(\tau_V^{\sigma,k} < \delta\big) \to 0 \quad \text{as } \delta \to 0, \tag{23}$$

and

$$\overline{\lim_{k\to\infty}} \sup_{\sigma \leq T} P\big[\big(y_{\sigma-}^k\big)^{-1} y_\sigma^k \in K^c\big] \to 0 \quad \text{as compact } K \uparrow G, \tag{24}$$

where $\sup_{\sigma \leq T}$ is taken over all stopping times $\sigma \leq T$.

We will apply Lemma 8 to $y_t^k = \tilde{z}_t^k$. Because $\tilde{z}_t^k = \tilde{z}_1^k$ for $t > 1$, we may take $T = 1$ in Lemma 8. Let $f \in C_c^\infty(G)$ be such that $0 \leq f \leq 1$ on G, $f(e) = 1$ and $f = 0$ on V^c. For any stopping time $\sigma \leq 1$, write τ for $\tau_V^{\sigma,k}$ and let $f_\sigma = f \circ l_z$ with $z = (\tilde{z}_\sigma^k)^{-1}$. Then

$$P(\tau < \delta) = E\big[f_\sigma\big(\tilde{z}_\sigma^k\big) - f_\sigma\big(\tilde{z}_{\sigma+\tau}^k\big); \tau < \delta\big] \leq E\big[f_\sigma\big(\tilde{z}_\sigma^k\big) - f_\sigma\big(\tilde{z}_{\sigma+\tau\wedge\delta}^k\big)\big], \tag{25}$$

noting $f_\sigma(z_\sigma^k) = 1$, $f_\sigma(z_{\sigma+\tau}^k) = 0$ and $\tau = \tau \wedge \delta$ on $[\tau < \delta]$. Because $\tilde{M}_t^k f$ given by (20) is a martingale for any $f \in C_c^\infty(G)$, and σ and $\sigma + \tau \wedge \delta$ are stopping times,

$$E\big[\tilde{M}_\sigma^k f_\sigma - \tilde{M}_{\sigma+\tau\wedge\delta}^k f_\sigma\big] = E\big\{E\big[\tilde{M}_\sigma^k f_\sigma - \tilde{M}_{\sigma+\tau\wedge\delta}^k f_\sigma \mid \mathcal{F}_\sigma\big]\big\} = 0.$$

Writing $\tilde{z}, \tilde{b}, b, \mu$ for $\tilde{z}^k_{t_{k,p-1}}, \tilde{b}^k_{t_{k,p-1}}, b_p, \mu_p$, by (20) and (25), we obtain

$$\begin{aligned} P(\tau<\delta) &\le -E\left\{\sum_{\sigma<t_{k,p}\le\sigma+\tau\wedge\delta}\int_G\left[f_\sigma\left(\tilde{z}\tilde{b}xb^{-1}\tilde{b}^{-1}\right)-f_\sigma(\tilde{z})\right]\mu(\mathrm{d}x)\right\} \\ &\le E\left\{\sum_{\sigma<t_{k,p}\le\sigma+\delta}\left|\int_G\left[f_\sigma\left(\tilde{z}\tilde{b}xb^{-1}\tilde{b}^{-1}\right)-f_\sigma(\tilde{z})\right]\mu(\mathrm{d}x)\right|\right\}. \end{aligned} \tag{26}$$

Performing the same computation leading to (9) shows that for some constant $c>0$,

$$P(\tau<\delta)\le cE\left[Q^k(\sigma+\delta)-Q^k(\sigma)\right].$$

By (16), $E[Q^k(\sigma+\delta)-Q^k(\sigma)]\le 2\varepsilon\delta+\varepsilon_k$. It follows that $\overline{\lim}_{k\to\infty}\sup_{\sigma\le 1}P(\tau<\delta)\le 2c\varepsilon\delta$. This shows that the condition (23) is satisfied for $y^k_t=\tilde{z}^k_t$.

To verify (24), note that because $\tilde{x}^k_t=\tilde{z}^k_t\tilde{b}^k_t$,

$$P\left[\left(\tilde{z}^k_{\sigma-}\right)^{-1}\tilde{z}^k_\sigma\in K^c\right]=P\left[\left(\tilde{x}^k_{\sigma-}\right)^{-1}\tilde{x}^k_\sigma\in\left(\tilde{b}^k_{\sigma-}\right)^{-1}K^c\tilde{b}^k_\sigma\right].$$

By either (i) or (ii), $\tilde{b}^k_t$ are bounded in k, when K is large, $(\tilde{b}^k_{\sigma-})^{-1}K\tilde{b}^k_\sigma$ contains a fixed neighborhood H of e. Because $(\tilde{b}^k_{\sigma-})^{-1}K^c\tilde{b}^k_\sigma=((\tilde{b}^k_{\sigma-})^{-1}K\tilde{b}^k_\sigma)^c$, it follows that

$$P\left[\left(\tilde{z}^k_{\sigma-}\right)^{-1}\tilde{z}^k_\sigma\in K^c\right]\le P\left[\left(\tilde{x}^k_{\sigma-}\right)^{-1}\tilde{x}^k_\sigma\in H^c\right]\le\sum_{p>m_k}\mu_p\left(H^c\right)\to 0$$

as $k\to\infty$. This verifies (24) even before taking $K\uparrow G$.

By Lemma 8, $\tilde{z}^k_t$ are relatively weak compact, and hence along a subsequence of $k\to\infty$, $\tilde{z}^k_t$ converge weakly to a rcll process $\tilde{z}_t$ in G. As $D(G)$-valued random variables, $\tilde{z}^k_\cdot$ converge in distribution to $\tilde{z}_\cdot$. It is well known (see for example Theorem 1.8 in [2], Chapter 3) that there are $D(G)$-valued random variables $\tilde{z}'^k_\cdot$ and $\tilde{z}'_\cdot$, possibly on a different probability space, such that $\tilde{z}'_\cdot$ is equal to $\tilde{z}_\cdot$ in distribution, $\tilde{z}'^k_\cdot$ is equal to $\tilde{z}^k_\cdot$ in distribution for each k, and $\tilde{z}'^k_\cdot\to\tilde{z}'_\cdot$ almost surely. Because $\tilde{x}^k_\cdot=\tilde{z}^k_\cdot\tilde{b}^k_\cdot\to e$ almost surely, where e is regarded as a constant path in G, $\tilde{x}'^k_\cdot=\tilde{z}'^k_\cdot\tilde{b}^k_\cdot\to e$ in distribution. As the limit e is nonrandom, $\tilde{x}'^k_\cdot\to e$ in probability. Then along a further subsequence of $k\to\infty$, $\tilde{x}'^k_\cdot\to e$ almost surely, and hence $\tilde{b}^k_\cdot=(\tilde{z}'^k_\cdot)^{-1}\tilde{x}'^k_\cdot\to(\tilde{z}'_\cdot)^{-1}$.

The convergence $\tilde{b}^k_t\to\tilde{b}_t=(\tilde{z}'_t)^{-1}$ under the Skorohod metric means (see Proposition 5.3(c) in [2, Chapter 3]) that there are continuous strictly increasing functions $\lambda_k:\mathbb{R}_+\to\mathbb{R}_+$ such that as $k\to\infty$, $\lambda_k(t)-t\to 0$ and $r(\tilde{b}^k_t,\tilde{b}_{\lambda_k(t)})\to 0$ uniformly for $0\le t\le 1$, where r is a compatible metric on G. If $\tilde{b}_t$ has a jump of size $r(\tilde{b}_{s-},\tilde{b}_s)>0$ at time s, then $\tilde{b}^k_t$ would have a jump of size close to $r(\tilde{b}^\gamma_{s-},\tilde{b}_s)$ at time $t=\lambda_k^{-1}(s)$, which is impossible because the jumps of $\tilde{b}^k_t$ are uniformly small when k is large. It follows that $\tilde{b}_t$ is continuous in t and hence $\tilde{b}^k_t\to\tilde{b}_t$ uniformly in t as $k\to\infty$.

Acknowledgements

The author wishes to thank David Applebaum and an anonymous referee for some very helpful comments.

References

[1] P. Billingsley. *Probability and Measure*, 2nd edition. Wiley, New York, 1986. MR0830424
[2] S. N. Ethier and T. G. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley, New York, 1986. MR0838085
[3] P. Feinsilver. Processes with independent increments on a Lie group. *Trans. Amer. Math. Soc.* **242** (1978) 73–121. MR0483034
[4] H. Heyer. *Probability Measures on Locally Compact Groups*. Springer, Berlin, 1977. MR0501241
[5] M. Liao. Inhomogeneous Lévy processes in Lie groups and homogeneous spaces. *J. Theoret. Probab.* **27** (2014) 315–357. MR3195816
[6] V. M. Maksimov. The principle of convergence "almost everywhere" in Lie groups. *Mat. Sb. (N.S.)* **91** (133) (1973) 523–536, 629. MR0372936