Lyapunov exponents of orthogonal-plus-normal cocycles

Sam Bednarski and Anthony Quas

September 16, 2023

Abstract

We consider products of matrices of the form $A_n = O_n + \epsilon N_n$ where O_n is a sequence of $d \times d$ orthogonal matrices and N_n has independent standard normal entries and the (N_n) are mutually independent. We study the Lyapunov exponents of the cocycle as a function of ϵ, giving an exact expression for the jth Lyapunov exponent in terms of the Gram-Schmidt orthogonalization of $I + \epsilon N$. Further, we study the asymptotics of these exponents, showing that $\lambda_j = (d - 2j)\epsilon^2/2 + O(\epsilon^4 |\log \epsilon|^4)$.

1 Introduction and Statement of Results

Lyapunov exponents play a highly important role in dynamical systems, allowing quantification of chaos, the development of a theory of hyperbolic and non-uniformly hyperbolic dynamical systems and much more. We work in the framework of multiplicative ergodic theory, where one has a base dynamical system $\sigma\colon \Omega \to \Omega$ preserving an ergodic measure $\mathbb{P}$ and a measurable map $A\colon \Omega \to M_{d\times d}(\mathbb{R})$. One then takes the cocycle of partial products $A_\omega^{(n)}$ of the sequence of $d \times d$ matrices and one studies the limiting growth rate of the jth singular value of the products.

In the case $d = 1$ this is often straightforward to calculate: the Lyapunov exponent is just $\int \log |A(\omega)|\, d\mathbb{P}(\omega)$. In higher dimensions, Lyapunov exponents tend to be much harder to calculate, and it is rare to be able to give exact expressions.

In this paper, we are able to establish exact expressions for Lyapunov exponents for cocycles of a particular form, namely where the matrices A_ω are of the form $O_\omega + \epsilon N_\omega$, where the O_ω are orthogonal matrices and the N_ω are mutually independent Gaussian matrices with independent standard normal entries. We further assume that the $(N_{\sigma^n\omega})$ are independent of the cocycle O_ω. We then interpret the cocycle as an additive noise perturbation of a cocycle of orthogonal matrices.

Our main results are the following:

Theorem 1. *Let $\sigma\colon \Omega \to \Omega$ be an ergodic transformation preserving an invariant measure $\mathbb{P}$ and let $O\colon \Omega \to O(d,\mathbb{R})$ be a measurable map into the $d \times d$ orthogonal matrices. Suppose that $N\colon \Omega \to M_{d\times d}(\mathbb{R})$ is measurable, and that that conditioned on $(O_{\sigma^n\omega})_{n\in\mathbb{Z}}$ and $(N_{\sigma^n\omega})_{n\neq 0}$, N_ω has independent standard normal entries. Then for all $\epsilon \in \mathbb{R}$, the Lyapunov exponents of the cocycle $A_\omega = O_\omega + \epsilon N_\omega$ are given by*

$$\lambda_j = \mathbb{E} \log \|c_j^\perp(I + \epsilon N)\|,$$

where $c_j^\perp(A)$ is the jth column of the Gram-Schmidt orthogonalization of A.

The following theorem describes the asymptotic behaviour of the exponents as ϵ tends to 0.

Theorem 2. *Let the matrix cocycle be as above. Then the Lyapunov exponents satisfy*

$$\lambda_j(\epsilon) = (d-2j)\tfrac{\epsilon^2}{2} + O(\epsilon^4 |\log \epsilon|^4)$$

as $\epsilon \to 0$.

We make the following conjecture. Let σ be an ergodic measure-preserving transformation of a space $(\Omega, \mathbb{P})$. If $B\colon \Omega \to M_{d\times d}(\mathbb{R})$ is the generator of a matrix cocycle with the property that $\|B_\omega\| \le 1$ almost surely, and N_ω is Gaussian with the independence properties above, then setting $\lambda_j'(\epsilon)$ to be the sequence of Lyapunov exponents of the cocycle $A_\omega^\epsilon = B_\omega + \epsilon N_\omega$, one has

$$\lambda_j'(\epsilon) - \lambda_{j+1}'(\epsilon) \ge \lambda_j(\epsilon) - \lambda_{j+1}(\epsilon) \text{ for all } \epsilon > 0,$$

where $\lambda_j(\epsilon)$ are the Lyapunov exponents for the cocycle described in Theorem 1.

That is, we conjecture that there are universal lower bounds on the gaps between consecutive Lyapunov exponents of Gaussian perturbed cocycles of matrices where the matrices in the unperturbed cocycle have norm at most 1; and that these lower bounds are obtained in the case where all of the matrices are the identity matrix.

The results in this paper are closely related to results in Newman [7], where he gave a result similar to Theorem 1 for some i.i.d. cocycles involving Gaussian matrices and SDE flows on the space of non-singular matrices.

Newman also re-derives an important result of Dynkin [3] that also has intermediate proofs due to LeJan [6]; Baxendale and Harris [1]; and Norris, Rogers and Williams [8]. Dynkin's result concerns the Lyapunov exponents of a simple stochastic flow on $GL_d(\mathbb{R})$, the group of invertible $d \times d$ matrices, and identifies explicit exact Lyapunov exponents for the flow. Although this cocycle is not the same as ours, it is in the same spirit. The Lyapunov exponents in that paper have a similar form to ours and are given by $\lambda_k = (d - 2k + 1)\sigma^2/2$

2 Definitions and Preliminary lemmas

If N is a $d \times d$ matrix valued random variable whose entries are independent standard normal random variables, we will say that N is the *standard Gaussian matrix random variable.* We will need the following property of the normal distribution:

Lemma 3. *Let U be an orthogonal matrix and let N be a standard Gaussian matrix random variable of the same dimensions. Then the matrices N, UN and NU are equal in distribution.*

This follows from a more general fact about the multivariate normal distribution.

Proposition 4. *Let $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a d-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Suppose V is a $d \times n$ matrix of rank d. Then $VX \sim N(V\boldsymbol{\mu}, V\boldsymbol{\Sigma}V^T)$.*

Proof. Recall that $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $X \sim AZ + \boldsymbol{\mu}$ where $AA^T = \boldsymbol{\Sigma}$ and $Z \sim N(\mathbf{0}, I_d)$. If $X = AZ + \boldsymbol{\mu}$ this implies that

$$\begin{aligned} VX &= VAZ + V\boldsymbol{\mu} \\ &\sim N(V\boldsymbol{\mu}, VA(VA)^T) \text{ by the fact above} \\ &\sim N(V\boldsymbol{\mu}, VAA^TV^T) \\ &\sim N(V\boldsymbol{\mu}, V\boldsymbol{\Sigma}V^T) \end{aligned}$$

□

Lemma 5. *Let N be a standard normal random variable. Then for any a and for $b \neq 0$, $\mathbb{E}\log^- |a + bN| \leq \mathbb{E}\log^- |bN| < \infty$.*

Proof. We have

$$\begin{aligned}\mathbb{E}\log^-|a+bN| &= \int_0^\infty \mathbb{P}(\log^-|a+bN|<t)\,dt \\ &= \int_0^\infty \mathbb{P}(N\in[-a-e^{-t}/|b|,-a+e^{-t}/|b|]\,dt \\ &\le \int_0^\infty \mathbb{P}(N\in[-e^{-t}/|b|,e^{-t}/|b|]\,dt \\ &= \mathbb{E}\log^-|bN|.\end{aligned}$$

Since $\log^-|bN|\le\log^-|b|+\log^-|N|$, and $\mathbb{E}\log^-|N|=\frac{2}{\sqrt{2\pi}}\int_0^1|\log x|e^{-x^2/2}\,dx$, it is easy to see $\mathbb{E}\log^-|bN|<\infty$. □

For a matrix B, let $c_j(B)$ denote the jth column of B and let $\theta_j(B)=\operatorname{dist}\big(c_j(B),\operatorname{lin}(\{c_i(B)\colon i\neq j\})\big)$ and let $\Theta(B)=\min\theta_j(B)$.

Lemma 6. *Let A be an arbitrary matrix and let $\epsilon>0$. Let Z denote a standard $d\times d$ Gaussian matrix random variable and N denote a standard normal random variable. Then $\mathbb{E}\log^-\Theta(A+\epsilon Z)\le\mathbb{E}\log^-(\epsilon N)<\infty$.*

Further, if S is any set, $\mathbb{E}\big(\log^-\Theta(A+\epsilon Z)\,\mathbf{1}_S\big)\le d\,\mathbb{P}(S)(1+\log^-(\epsilon\mathbb{P}(S)))$.

Proof. Let $\mathcal{F}$ denote the σ-algebra generated by the columns of N except for the jth. Then $\mathbb{E}\log^-\theta_j(A+\epsilon Z)=\mathbb{E}\big(\mathbb{E}(\log^-\theta_j(A+\epsilon Z)|\mathcal{F})\big)$. Let $\mathbf{n}$ be an $\mathcal{F}$-measurable unit normal to the subspace spanned by $(c_i(A+\epsilon Z))_{i\neq j}$ (this is almost surely unique up to a change of sign). Then $\theta_j(A+\epsilon Z)=|\langle\mathbf{n},c_j(A+\epsilon Z)\rangle|=|\langle\mathbf{n},c_j(A))\rangle+\epsilon\langle\mathbf{n},c_j(Z)\rangle|$. Let $a=\langle\mathbf{n},c_j(A))\rangle$ (an $\mathcal{F}$-measurable random variable) and note that since $c_j(Z)$ is independent of the unit vector $\mathbf{n}$, conditioned on $\mathcal{F}$, by Proposition 4, $\langle\mathbf{n},c_j(Z)\rangle$ is distributed as a standard normal random variable. Hence we have $\mathbb{E}\log^-\theta_j(A+\epsilon Z)=\mathbb{E}\big(\mathbb{E}(\log^-\theta_j(A+\epsilon Z)\big|\mathcal{F})\big)=\mathbb{E}\big(\mathbb{E}(\log^-|a+\epsilon N|\big|\mathcal{F})\big)\le\mathbb{E}\big(\mathbb{E}(\log^-|\epsilon N|\big|\mathcal{F})\big)=\mathbb{E}\log^-|\epsilon N|$ which is finite by the lemma above.

By definition, $\Theta(A+\epsilon Z)=\min_j\theta_j(A+\epsilon Z)$ so that $\log^-\Theta(A+\epsilon Z)=\max_j\log^-\theta_j(A+\epsilon Z)\le\sum_j\log^-\theta_j(A+\epsilon Z)$. By Lemma 5, we see $\mathbb{E}\log^-\Theta(A+\epsilon Z)<\infty$ as required.

Now if S is any set, we have

$$\begin{aligned}\mathbb{E}\big(\log^- \theta_j(A+\epsilon Z)\,\mathbf{1}_S\big) &= \int_0^\infty \mathbb{P}(S \cap \{\log^- \theta_j(A+\epsilon Z) > t\})\,dt \\ &\le \int_0^\infty \min\big(\mathbb{P}(S), \mathbb{P}(\theta_j(A+\epsilon Z) < e^{-t})\big)\,dt \\ &\le \int_0^\infty \min\big(\mathbb{P}(S), \mathbb{P}(a+\epsilon N \in [-e^{-t}, e^{-t}])\big)\,dt \\ &\le \int_0^\infty \min\big(\mathbb{P}(S), e^{-t}/\epsilon\big)\,dt,\end{aligned}$$

where in the third line, as above, a is a random variable that is independent of N. For the fourth line, we used the fact that the density of a standard normal is bounded above by $(2\pi)^{-1/2} < \frac{1}{2}$. Separating the integration region into $[0, \log^-(\epsilon\mathbb{P}(S))]$ and $[\log^-(\epsilon\mathbb{P}(S)), \infty)$, we obtain $\mathbb{E}\big(\log^- \theta_j(A+\epsilon Z)\,\mathbf{1}_S\big) \le \mathbb{P}(S)\log^-(\epsilon\mathbb{P}(S)) + \mathbb{P}(S)$. Since $\log^- \Theta(B) \le \sum_{j=1}^d \log^- \theta_j(B)$, the result follows. □

For any vector y in $\mathbb{R}^d$, y has at least one coefficient of magnitude $\|y\|/\sqrt{d}$, say the jth, so $\|By\| \ge \|y_j c_j(B) + \sum_{i\neq j} y_i c_i(B)\| \ge |y_j|\theta_j(B) \ge (1/\sqrt{d})\Theta(B)\|y\|$. If B is invertible then $\Theta(B)$ is non-zero and substituting $y = B^{-1}x$ gives $\|B^{-1}\| \le \sqrt{d}/\Theta(B)$.

Corollary 7. *Let (A_n) denote an i.i.d. sequence of $d \times d$ random matrices where $A_n = I + \epsilon N_n$, and where N_n is a $d \times d$ standard Gaussian matrix random variables. Then (A_n) satisfies the following:*

1. *$A_n \in GL_d(\mathbb{R})$ a.s.;*
2. *the distribution of A_n is fully supported in $GL_d(\mathbb{R})$: for any non-empty open set $U \subset GL_d(\mathbb{R})$, $\mathbb{P}(A_n \in U) > 0$.*
3. *$\log \|A_n\| \in L^1(\Omega)$.*

This corollary establishes that the sequence (A_n) satisfies the conditions of the Gol'dsheid-Margulis theorem [4, Theorem 5.4] which ensures that the Lyapunov exponents of the cocycle $A^{(n)} = (I + \epsilon N_n)\cdots(I + \epsilon N_1)$ are all distinct.

Proof. The distribution of the matrices A_i is mutually absolutely continuous with respect to Lebesgue measure. Since the zero locus of the polynomial equation $\det(A) = 0$ is a measure zero set, the first and second conclusions are established. To show $\log \|A_i\|$ is integrable, we separately show that $\log^+ \|A_i\|$ and $\log^- \|A_i\|$ are integrable. First,

$$\mathbb{E} \log^+ \|A_1\| \leq \mathbb{E} \|A_1\| \leq \mathbb{E} \sum_{1 \leq i,j \leq d} |A_{ij}|$$

where each A_{ij} is an integrable normal random variable. The fact that $\mathbb{E} \log^+ \left\|A_1^{-1}\right\| < \infty$ follows from Lemma 6 and the observation that $\|A_1^{-1}\| \leq \sqrt{d}/\Theta(A_1)$ made above. □

We make extensive use of the singular value decomposition in what follows. More information on this topic may be found in Horn and Johnson [5] and Bhatia [2]. For a $d \times d$ matrix A, a singular value decomposition is a triple (L, D, R) where L and R are orthogonal matrices and D is a diagonal matrix with non-negative entries such that $A = LDR$. We impose without loss of generality the requirement that the diagonal entries of D are decreasing. The matrix D is uniquely determined by A while there is some freedom in the choice of L and R. The *singular values* of A are denoted $s_i(A)$, where $s_i(A)$ is the ith entry of the diagonal of A. It is known (see for example Ragunathan [10, Lemma 1]) that there exist measurable functions L, D and R mapping $M_d(\mathbb{R})$ to $O(d, \mathbb{R})$, $M_{\text{diag}}(d, \mathbb{R})$ and $O(d, \mathbb{R})$ respectively such that $A = L(A)D(A)R(A)$.

It is well known that $|s_i(A) - s_i(B)| \leq \|A - B\|$ where $\|\cdot\|$ is the standard operator norm on matrices. We also make use of the products $S_i^j(A) = s_i(A) \cdots s_j(A)$. These have an interpretation in terms of exterior algebra. We write $\bigwedge^k \mathbb{R}^d$ for the kth exterior power of $\mathbb{R}^d$ and equip it with the standard inner product coming from the Cauchy-Binet formula and the corresponding norm. In particular if $v_1, \ldots, v_d$ is an orthonormal basis for $\mathbb{R}^d$, then $\{v_{i_1} \wedge \cdots \wedge v_{i_k} : i_1 < i_2 < \ldots < i_k\}$ is an orthonormal basis for $\bigwedge^k \mathbb{R}^d$. With respect to the corresponding operator norm, it is well known that $\left\|A^{\wedge k}\right\| = S_1^k(A)$.

If (A_n) is an independent identically distributed sequence of random variables taking values in $GL_d(\mathbb{R})$ such that $\mathbb{E} \log \|A_1\|^{\pm 1} < \infty$, it was shown by Oseledets [9] and Raghunathan [10] that the limits

$$\lim_{n \to \infty} \frac{1}{n} \log s_j(A_n \ldots A_1)$$

exist and are almost surely constant for almost every realization of (A_n). The almost sure limit is denoted λ_j and the (λ_j) are the *Lyapunov exponents* of the cocycle.

3 Exact expressions for Lyapunov exponents

For $1 \le k \le d$, we define e_k to be the kth coordinate vector, so that, as previously defined, $c_k(A) := Ae_k$ is the kth column of A. Let $c_k^\perp(A)$ denote the component of the kth column of A which is orthogonal to the first $k-1$ columns. That is, suppressing the matrix A for brevity, $c_1^\perp = c_1$, and

$$c_k^\perp := c_k - \sum_{1\le j<k} \frac{\langle c_j^\perp, c_k \rangle}{\langle c_j^\perp, c_j^\perp \rangle} c_j^\perp.$$

We now prove Theorem 1 which we restate here for convenience.

Theorem. *Let $(U_n)_{n\in\mathbb{Z}}$ be a sequence of $d \times d$ orthogonal matrices and let $(N_n)_{n\in\mathbb{Z}}$ be a sequence of independent $d\times d$ matrices, each with independent standard normal coefficients. Let $\epsilon > 0$ and let $A_j^\epsilon = U_j + \epsilon Z_j$. Then for $1 \le k \le d$, the kth Lyapunov exponent of the cocycle $(A^\epsilon_{\sigma^{n-1}\omega} \cdots A^\epsilon_\omega)$ is given by*

$$\lambda_k = \mathbb{E}(\log \left\| c_k^\perp(I + \epsilon N)\right\|).$$

Fix $\epsilon > 0$ and set $A_i := U_i + \epsilon N_i$ for each i. To find the Lyapunov exponents of this sequence we work with the products $A^{(n)} = A_n \cdots A_1$.

We now define $\Sigma_n = D(A^{(n)})$ and study the evolution of Σ_n. More precisely we are interested in the stochastic process $(\Sigma_n)_{n\ge0}$. To write Σ_{n+1} in terms of Σ_n, we have $\Sigma_{n+1} = D\big(A_{n+1}L(A^{(n)})\Sigma_n R(A^{(n)})\big)$. The following lemma shows that this process (Σ_n) is Markov and that the process has the same distribution as the simpler process $\Sigma'_{n+1} = D((1 + \epsilon N_{n+1})\Sigma'_n)$.

Lemma 8. *((Σ_n) is a Markov process) Let the sequence of matrices (A_i) be given by $U_i + \epsilon N_i$ as above and let $\Sigma_n = D(A^{(n)})$. Then (Σ_n) is a Markov process: For any measurable set F of diagonal matrices,*

$$\begin{aligned}\mathbb{P}(\Sigma_{n+1} \in F|\Sigma_n, \ldots, \Sigma_1) &= \mathbb{P}(\Sigma_{n+1} \in F|\Sigma_n)\\ &= \mathbb{P}(D((I + \epsilon N)\Sigma_n) \in F|\Sigma_n).\end{aligned}$$

That is, the Markov process (Σ_n) has the same distribution as the Markov process (Σ'_n) where $\Sigma'_0 = I$ and $\Sigma'_{n+1} = D(A'_{n+1}\Sigma_n)$, where (A'_n) is an independent sequence of matrices, each distributed as $I + \epsilon N$.

Proof. Let $\mathcal{F}_n$ denote the smallest σ-algebra with respect to which $N_1, \ldots, N_n$ are measurable. Let $\mathcal{G}_n$ be the smallest σ-algebra with respect to which $\Sigma_1, \ldots, \Sigma_n$ are measurable (so that $\mathcal{G}_n$ is a sub-σ-algebra of $\mathcal{F}_n$).

As usual, we write $A^{(n)}$ for the product $A_n \dots A_1$. Let $L_n = L(A^{(n)})$, $\Sigma_n = D(A^{(n)})$, $R_n = R(A^{(n)})$. Let F be a measurable subset of the range of D. We compute

$$\begin{aligned}
\mathbb{P}(\Sigma_{n+1} \in F|\mathcal{F}_n) &= \mathbb{P}\Big(D(A_{n+1}L_n\Sigma_n R_n) \in F|\mathcal{F}_n\Big) \\
&= \mathbb{P}\Big(D(A_{n+1}L_n\Sigma_n) \in F|\mathcal{F}_n\Big) \\
&= \mathbb{P}\Big(D\big((U_{n+1} + \epsilon N_{n+1})L_n\Sigma_n\big) \in F|\mathcal{F}_n\Big) \\
&= \mathbb{P}\Big(D\big(U_{n+1}L_n(I + \epsilon L_n^{-1}U_{n+1}^{-1}N_{n+1}L_n)\Sigma_n\big) \in F|\mathcal{F}_n\Big) \\
&= \mathbb{P}\Big(D\big((I + \epsilon L_n^{-1}U_{n+1}^{-1}N_{n+1}L_n)\Sigma_n\big) \in F|\mathcal{F}_n\Big) \\
&= \mathbb{P}\Big(D\big((I + \epsilon N_{n+1})\Sigma_n\big) \in F|\mathcal{F}_n\Big),
\end{aligned}$$

where the second and fifth lines follow from that facts that $D(A) = D(AU) = D(UA)$ for any matrix A and any orthogonal matrix U. The sixth line uses the fact that N_{n+1} is independent of $\mathcal{F}_n$ and Lemma 3 so that conditioned on $\mathcal{F}_n$, $L_n^{-1}U_{n+1}^{-1}N_{n+1}L_n$ has the same distribution as N_{n+1}. Since N_{n+1} is independent of $\mathcal{F}_n$, this is equal to $\mathbb{P}\Big(D\big((I+\epsilon N_{n+1})\Sigma_n\big) \in F|\Sigma_n\Big)$. We have established that

$$\mathbb{P}(\Sigma_{n+1} \in F|\mathcal{F}_n) = \mathbb{P}\Big(D\big((I + \epsilon N_{n+1})\Sigma_n\big) \in F|\Sigma_n\Big).$$

Taking conditional expectations of both sides with respect to $\mathcal{G}_n$, we deduce

$$\mathbb{P}(\Sigma_{n+1} \in F|\Sigma_n, \dots, \Sigma_1) = \mathbb{P}(\Sigma_{n+1} \in F|\Sigma_n).$$

□

Proof of Theorem 1. Fix $1 \leq k \leq d$. We use the stochastic process described in Lemma 8: let $A_n = I + \epsilon N_n$, $\Sigma_0 = I$, $\Sigma_n = D(A_n\Sigma_{n-1}) = \mathrm{diag}(s_1(A_n\Sigma_{n-1}), \dots, s_d(A_n\Sigma_{n-1}))$. As before, we write $A^{(n)} = A_n \dots A_1$. We note that Σ_n is *not* equal to $D(A^{(n)})$, but using Lemma 8 the two processes $(\Sigma_n)_{n\geq 0}$ and $\big(D(A^{(n)})\big)_{n\geq 0}$ have the same distribution.

Let $B_n = A_n\Sigma_{n-1}\left(e_1 \ \dots \ e_k \ 0 \ \dots \ 0\right)$. Then for all $1 \leq j \leq k$,

$$\begin{aligned}
\big|s_j(\Sigma_n) - s_j(B_n)\big| &= \big|s_j(A_n\Sigma_{n-1}) - s_j(B_n)\big| \\
&\leq \big\|A_n\Sigma_{n-1} - B_n\big\| \\
&= \big\|A_n\Sigma_{n-1}\left(0 \ \dots 0 \ e_{k+1} \ \dots \ e_d\right)\big\| \\
&= \|A_n \,\mathrm{diag}(0, \dots 0, s_{k+1}(\Sigma_{n-1}), \dots, s_d(\Sigma_{n-1}))\| \\
&\leq s_{k+1}(\Sigma_{n-1})\,\|A_n\|
\end{aligned}$$

Then we have

$$\left|\frac{s_j(B_n)}{s_j(\Sigma_n)} - 1\right| = \left|\frac{s_j(B_n) - s_j(\Sigma_n)}{s_j(\Sigma_n)}\right|$$
$$\leq \frac{s_{k+1}(\Sigma_{n-1})}{s_j(\Sigma_n)} \|A_n\|$$

By Gol'dsheid and Margulis, [4, Theorem 5.4], $\frac{1}{n}\log s_j(A^{(n)}) \to \lambda_j$ and $\frac{1}{n}\log s_{k+1}(A^{(n)}) \to \lambda_{k+1}$ almost surely for some $\lambda_j > \lambda_{k+1}$. Since the processes $\big(D(A^{(n)})\big)$ and (Σ_n) have a common distribution, it follows that $\frac{1}{n}\log s_j(\Sigma_n) \to \lambda_j$ and $\frac{1}{n}\log s_{k+1}(\Sigma_n) \to \lambda_{k+1}$ almost surely. So

$$\frac{1}{n}\log\left(\frac{s_{k+1}(\Sigma_{n-1})}{s_j(\Sigma_n)}\right) \to \lambda_{k+1} - \lambda_j < 0$$

almost surely as $n \to \infty$. If this occurs, there is some $N \in \mathbb{N}$ such that $s_{k+1}(\Sigma_{n-1})/s_j(\Sigma_n) < e^{-n(\lambda_{k+1}-\lambda_j)/2}$ for all $n \geq N$. A well-known consequence of the Strong Law of Large Numbers ensures that $C(\omega) := \sup_n \|A_n\|/n$ is finite a.s., so that $\|A_n\|/n \leq C(\omega)$ for all n. For $n \geq N$ we then have

$$\left|\frac{s_{k+1}(\Sigma_{n-1})}{s_j(\Sigma_n)}\right| \|A_n\| \leq C(\omega) n e^{-n(\lambda_{k+1}-\lambda_j)/2} \to 0$$

as $n \to \infty$. Hence

$$\frac{s_j(B_n)}{s_j(\Sigma_n)} \to 1 \text{ as } n \to \infty. \tag{1}$$

For a matrix A, let $s_1^k(A) = s_1(A)\cdots s_k(A)$. Since B_n has k non-zero columns, ${B_n}^{\wedge k}$ has rank one and we have

$$\begin{aligned}
s_1^k(B_n) &= \|B_n e_1 \wedge B_n e_2 \wedge \cdots \wedge B_n e_k\| \\
&= \|(A_n\Sigma_{n-1})e_1 \wedge (A_n\Sigma_{n-1})e_2 \wedge \cdots \wedge (A_n\Sigma_{n-1})e_k\| \\
&= \|(A_n e_1)s_1(\Sigma_{n-1}) \wedge (A_n e_2)s_2(\Sigma_{n-1}) \wedge \cdots \wedge (A_n e_k)s_k(\Sigma_{n-1})\| \\
&= s_1^k(\Sigma_{n-1}) \|c_1(A_n) \wedge c_2(A_n) \wedge \cdots \wedge c_k(A_n)\| \\
&= s_1^k(\Sigma_{n-1}) \big\|c_1^\perp(A_n)\big\| \big\|c_2^\perp(A_n)\big\| \cdots \big\|c_k^\perp(A_n)\big\|,
\end{aligned}$$

where $\wedge$ denotes the wedge product.

For $n \in \mathbb{N}$ and $1 \leq k \leq d$, let $X_n^k := \big\|c_1^\perp(A_n)\big\|\big\|c_2^\perp(A_n)\big\|\cdots\big\|c_k^\perp(A_n)\big\|$. Then $X_1^k, X_2^k, \ldots$ is a sequence of i.i.d. random variables. Since $\Theta(A) \leq$

$\|c_i^\perp(A)\| \leq \|A\|$, we see, using Lemma 6 and Corollary 7 that $\log \|c_i^\perp(A)\|$ is integrable. We have

$$\begin{aligned} s_1^k(\Sigma_n) &= \frac{s_1^k(\Sigma_n)}{s_1^k(B_n)} s_1^k(B_n) \\ &= \frac{s_1^k(\Sigma_n)}{s_1^k(B_n)} X_n^k s_1^k(\Sigma_{n-1}). \end{aligned}$$

Using induction, we obtain

$$s_1^k(\Sigma_n) = \left(\prod_{j=1}^n \frac{s_1^k(\Sigma_j)}{s_1^k(B_j)} \right) X_1^k \ldots X_n^k.$$

Hence

$$\tfrac{1}{n} \log s_1^k(\Sigma_n) = \frac{1}{n} \sum_{j=1}^n \log \frac{s_1^k(\Sigma_j)}{s_1^k(B_j)} + \frac{1}{n} \sum_{j=1}^n \log X_j^k.$$

By (1), the first term on the right side converges almost surely to 0 and by the Strong Law of Large Numbers the second term converges almost surely to $\mathbb{E} \log X_1^k$. Hence we obtain

$$\lambda_1 + \ldots + \lambda_k = \mathbb{E}\big(\log \|c_1^\perp(I + \epsilon N)\| + \ldots + \log \|c_k^\perp(I + \epsilon N)\|\big).$$

Subtracting the $(k-1)$-fold partial sum from the k-fold partial sum, we obtain

$$\lambda_k = \mathbb{E} \log \|c_k^\perp(I + \epsilon N)\|,$$

as required. □

This gives us an explicit description of λ_k. However it is difficult to compute for large matrices. In the next section we find an approximation for λ_k which is easier to compute.

4 An approximation for λ_j

In this section we focus on the case where $A \sim I_d + \epsilon N$ and introduce the computationally simpler vectors $c_j'(A)$ approximating $c_j^\perp(A)$, defined by $c_1'(A) = c_1(A)$ and

$$c_k'(A) = c_k(A) - \sum_{1 \leq j < k} \langle c_j(A), c_k(A) \rangle \, c_j(A)$$

With the same setup as in the previous section, when $|\epsilon \log \epsilon| < (100d)^{-1}$ we have

Theorem 9. *For any $d \in \mathbb{N}$, if $A_1 \sim I_d + \epsilon N$ and $1 \leq k \leq d$ then* $\mathbb{E} \log \left\| c_k^{\perp} \right\| = \mathbb{E} \log \left\| c_k' \right\| + O(\epsilon^4 |\log \epsilon|^4)$.

We will say that $A = I + \epsilon N$ is *bad* if $|N_{ij}| > |\log \epsilon|$ for some i, j. Let bad denote the event that A is bad. We first control the contribution to $\mathbb{E} \log \left\| c_k^{\perp} \right\| - \mathbb{E} \log \| c_k' \|$ coming from the bad set.

Lemma 10. *Let $\epsilon > 0$. Then*

$$\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \big| \log \| c_j^{\perp}(I + \epsilon N) \| \big|\big) = O(|\log \epsilon| e^{-(\log \epsilon)^2/2}); \text{ and}$$
$$\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \big| \log \| c_j'(I + \epsilon N) \| \big|\big) = O(|\log \epsilon| e^{-(\log \epsilon)^2/2}).$$

Proof. We write $c_j^{\perp}$ and c_j' for $c_j^{\perp}(I + \epsilon N)$ and $c_j'(I + \epsilon N)$ respectively. We control the positive parts $\log^+ \| c_j' \|$ and $\log^+ \| c_j^{\perp} \|$, and the negative parts $\log^- \| c_j' \|$ and $\log^- \| c_j^{\perp} \|$. For the positive parts, notice that $\| c_j^{\perp} \| \leq \| c_j \| \leq \sum_{i,j} |a_{ij}|$ and $\| c_j' \| \leq \Big(1 + \sum_{i,j} |a_{ij}|\Big)^3$. The set bad is a union of d^2 parts of the form $\mathsf{bad}_{ij} = \{N \colon |N_{ij}| > |\log \epsilon|\}$. Using the bound $\log^+(x) \leq x$, this gives

$$\begin{aligned}
\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \log^+ \| c_j^{\perp} \|\big) &\leq \sum_{i,j} \int_{\mathsf{bad}_{i,j}} \Big(d + \epsilon \sum_{k,l} |x_{kl}|\Big) f_X\big((x_{kl})\big) d(x_{kl}) \\
&\leq d^2 \int_{\mathsf{bad}_{1,1}} \big(d + \epsilon d^2 |x_{11}|\big) f_N(x_{11}) dx_{11} \\
&= O(\exp(-(\log \epsilon)^2/2)).
\end{aligned}$$

A similar argument holds for $\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \log^+ \| c_j' \|\big)$.

To control $\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \log^- \| c_j^{\perp} \|\big)$ and $\mathbb{E}\big(\mathbb{1}_{\mathsf{bad}} \log^- \| c_j' \|\big)$, recall $\| c_j^{\perp} \|$ and $\| c_j' \|$ are bounded below by $\Theta(A)$. By standard estimates on the tail of the normal distribution, $\mathbb{P}(\mathsf{bad}) = O(e^{-(\log \epsilon)^2/2} / |\log \epsilon|)$. We see from Lemma 6, $\mathbb{E} \log^- \Theta(I + \epsilon N) \mathbf{1}_{\mathsf{bad}} = O(|\log \epsilon| e^{-(\log \epsilon)^2/2})$, which gives the required estimates.

□

We now give pointwise estimates for $\big| \log \| c_k^{\perp} \| - \log \| c_k' \| \big|$ when A is not bad. That is, when $A = I + \epsilon N_{ij}$ where $|N_{ij}| \leq |\log \epsilon|$ for all i, j.

Lemma 11. *There exist $\epsilon_0 > 0$ and $C > 0$ depending only on d such that for all matrices A of the form $A = I + \epsilon X$ where $|X_{ij}| \le |\log \epsilon|$ for each i, j, then for each k,*

$$\left| \log \left\| c_k^{\perp}(A) \right\| - \log \left\| c_k'(A) \right\| \right| \le C(\epsilon |\log \epsilon|)^4 \text{ for all } \epsilon < \epsilon_0.$$

As usual, we write c_j, $c_j^{\perp}$ and c_j' in place of $c_j(A)$, $c_j^{\perp}(A)$ and $c_j'(A)$ for brevity. We define $\alpha_i^j := \frac{\langle c_i^{\perp}, c_j \rangle}{\|c_i^{\perp}\|^2}$ so that $c_j^{\perp} = c_j - \sum_{i<j} \alpha_i^j c_i^{\perp}$. Throughout the proof, we let $\eta = |\log \epsilon|$. We let ϵ_0 be sufficiently small that $\epsilon\eta < 1/(100d)$ for $\epsilon < \epsilon_0$. The proof makes use of a number of claims.

Claim 1. *Let $A = I + \epsilon X$ where $|X_{ij}| \le \eta$ for all i, j. For all $1 \le n \le d$, the following hold:*

(i) $| \|c_n\|^2 - 1| \le 2\epsilon\eta + d\eta^2\epsilon^2 \le 3\epsilon\eta$;

(ii) $| \left\|c_n^{\perp}\right\|^2 - 1| \le 3\epsilon\eta$;

(iii) $|\alpha_i^k| \le 6\epsilon\eta$ *for all* $i \le n$ *and* $k > i$;

(iv) $| \langle c_n^{\perp}, c_k \rangle | \le 3\epsilon\eta$ *for all* $k > n$.

Proof. Since $|X_{ij}| \le \eta$ for all i, j, for any $1 \le n \le d$ and $i < j$ we have

$$\begin{aligned} | \|c_n\|^2 - 1| &\le 2\epsilon\eta + d\epsilon^2\eta^2 \quad \text{and} \\ | \langle c_i, c_j \rangle | &\le 2\epsilon\eta + d\epsilon^2\eta^2. \end{aligned}$$

This shows (i) for all n, as well as (ii), (iii) and (iv) for $n = 1$.

Now suppose for some $2 \le j \le d$, (ii)–(iv) each hold for all $n \le j - 1$. Then for all $k > j$ we have

$$\langle c_j^{\perp}, c_k \rangle = \left\langle c_j - \sum_{i<j} \alpha_i^j c_i^{\perp}, c_k \right\rangle = \langle c_j, c_k \rangle - \sum_{i<j} \alpha_i^j \langle c_i^{\perp}, c_k \rangle$$

This implies

$$\begin{aligned} |\langle c_j^{\perp}, c_k \rangle| &\le |\langle c_j, c_k \rangle| + \sum_{i<j} |\alpha_i^j| |\langle c_i^{\perp}, c_k \rangle| \\ &\le (2\epsilon\eta + d\epsilon^2\eta^2) + d \cdot (6\epsilon\eta)(3\epsilon\eta) \\ &\le 3\epsilon\eta, \end{aligned}$$

where we used (i) and the induction hypotheses in the second line and the condition on ϵ_0 in the third line. This establishes (iv) for $n = j$.

Since $c_1^\perp, \ldots, c_j^\perp$ are mutually perpendicular, it follows that

$$\|c_j\|^2 = \|c_j^\perp\|^2 + \sum_{i<j} (\alpha_i^j)^2 \|c_i^\perp\|^2$$

Thus we have

$$\begin{aligned} \left| \|c_j^\perp\|^2 - 1 \right| &= \left| \|c_j\|^2 - 1 + \sum_{i<j} (\alpha_i^j)^2 \|c_i^\perp\|^2 \right| \\ &\leq \left| \|c_j\|^2 - 1 \right| + \sum_{i<j} (\alpha_i^j)^2 \|c_i^\perp\|^2 \\ &\leq (2\epsilon\eta + d\epsilon^2\eta^2) + d(6\epsilon\eta)^2(1 + 3\epsilon\eta) \\ &\leq 3\epsilon\eta, \end{aligned}$$

establishing (ii) for $n = j$.

We show that (iii) holds for $n = j$. Since by the induction hypothesis, $|\alpha_i^k| \leq 6\epsilon\eta$ for all $i < j$ and $k > i$, if suffices to show that $|\alpha_j^k| \leq 6\epsilon\eta$ for all $k > j$. For any $k > j$, using (iv), we have

$$|\alpha_j^k| = \frac{|\langle c_j^\perp, c_k \rangle|}{\|c_j^\perp\|^2} \leq \frac{3\epsilon\eta}{1/2} = 6\epsilon\eta$$

which shows that (iii) holds for $n = j$. □

Claim 2. *For each* $1 \leq n \leq d$, $c_n^\perp = c_n + \sum_{j<n} \beta_j^n c_j$ *where* $|\beta_j^n| < 7\epsilon\eta$.

Proof. We use induction on j. The base case is $c_1^\perp = c_1$. Suppose the claim holds for all $n < j \leq d$. Then

$$\begin{aligned} c_j^\perp &= c_j - \sum_{i<j} \alpha_i^j c_i^\perp \\ &= c_j - \sum_{i<j} \alpha_i^j \Big(c_i + \sum_{\ell<i} \beta_\ell^i c_\ell \Big) \\ &= c_j - \sum_{\ell<j} \alpha_\ell^j c_\ell - \sum_{i<j} \alpha_i^j \sum_{\ell<i} \beta_\ell^i c_\ell \\ &= c_j - \sum_{\ell<j} \alpha_\ell^j c_\ell - \sum_{\ell<j-1} \Big(\sum_{i=\ell+1}^{j-1} \alpha_i^j \beta_\ell^i \Big) c_\ell \end{aligned}$$

For any $\ell < j$, the coefficient of c_ℓ in the above expression is bounded by

$$|\alpha_\ell^j| + \sum_{i=\ell+1}^{j-1} |\alpha_i^j \beta_\ell^i| \leq 6\epsilon\eta + d(6\epsilon\eta)(7\epsilon\eta) \leq 7\epsilon\eta$$

□

Claim 3. *For all* $1 \leq j \leq d$, $c_j' = c_j^\perp + \sum_{n<j} \gamma_n c_n$ *where* $\gamma_n = O(\epsilon^2\eta^2)$, *where the implicit constant depends only on* d

Proof. For any such j we have

$$c_j' - c_j^\perp = \sum_{i<j} \left(\frac{\langle c_i^\perp, c_j \rangle}{\langle c_i^\perp, c_i^\perp \rangle} c_i^\perp - \langle c_i, c_j \rangle c_i \right).$$

We identify the coefficient of c_ℓ when $c_j' - c_j^\perp$ is expanded in the basis (c_k). That coefficient may be seen to be

$$\begin{aligned} &\frac{\langle c_\ell^\perp, c_j \rangle}{\langle c_\ell^\perp, c_\ell^\perp \rangle} - \langle c_\ell, c_j \rangle + \sum_{\ell<i<j} \frac{\langle c_i^\perp, c_j \rangle}{\langle c_i^\perp, c_i^\perp \rangle} \beta_\ell^i \\ =& \frac{\langle c_\ell^\perp, c_j \rangle - \langle c_\ell, c_j \rangle}{\langle c_\ell^\perp, c_\ell^\perp \rangle} + \frac{\langle c_\ell, c_j \rangle \left(1 - \langle c_\ell^\perp, c_\ell^\perp \rangle\right)}{\langle c_\ell^\perp, c_\ell^\perp \rangle} + O(\epsilon^2\eta^2), \end{aligned}$$

where we added and subtracted $\langle c_\ell, c_j \rangle / \langle c_\ell^\perp, c_\ell^\perp \rangle$; and the estimate for the third term follows from Claims 1 and 2.

Since $\langle c_\ell^\perp, c_j \rangle - \langle c_\ell, c_j \rangle = -\left\langle \sum_{i<\ell} \beta_i^\ell c_i, c_j \right\rangle$, the estimates in Claims 1 and 2 show the first term is $O(\epsilon^2\eta^2)$. Finally since $\langle c_\ell, c_j \rangle = O(\epsilon\eta)$ and $1 - \langle c_\ell^\perp, c_\ell^\perp \rangle$ is $O(\epsilon\eta)$ by Claim 1, the middle term is also $O(\epsilon^2\eta^2)$.

□

Proof of Lemma 11. By orthogonality,

$$\left\|c_j'\right\|^2 = \left\| c_j^\perp + \sum_{n<j} \gamma_n c_n \right\|^2 = \|c_j^\perp\|^2 + \left\| \sum_{n<j} \gamma_n c_n \right\|^2,$$

where γ_n is as in Claim 3. Since $\gamma_n = O(\epsilon^2\eta^2)$, we obtain $\left\|c_j'\right\|^2 = \left\|c_j^\perp\right\|^2 + O(\epsilon^4\eta^4)$. Since $\left\|c_j^\perp\right\|^2$ is in the range $(\frac{1}{2}, \frac{3}{2})$, it follows that $\left|\log\left\|c_j'\right\| - \log\left\|c_j^\perp\right\|\right| = O(\epsilon^4\eta^4)$ as required.

□

Proof of Theorem 9. Lemma 10 shows that

$$\begin{aligned}
&\big|\mathbb{E}\big(\log\|c_k^\perp\| - \log\|c_k'\|)\mathbf{1}_{\mathsf{bad}}\big)\big| \\
&\le \mathbb{E}\big(\log\|c_k^\perp\|\mathbf{1}_{\mathsf{bad}}\big) + \mathbb{E}\big(\log\|c_k'\|\mathbf{1}_{\mathsf{bad}}\big) \\
&= O(|\log\epsilon|e^{-(\log\epsilon)^2}).
\end{aligned}$$

and Lemma 11 shows that $\big|\log\big\|c_k^\perp\big\| - \log\big\|c_k'\big\|\big|\mathbf{1}_{\mathsf{bad}^c} = O(\epsilon^4|\log\epsilon|^4)$. Taking the expectation of this and combining the estimates gives the theorem. □

5 Computing $\mathbb{E}\log\|c_k'\|$

Finally, we find the dominant term in the asymptotic expansion for $\mathbb{E}\log\|c_j'\|$ in the same setup as the previous section. This is Theorem 2 which we restate here for convenience.

Theorem. *Consider an orthogonal-plus-Gaussian cocycle as in Theorem 1. Then the Lyapunov exponents satisfy*

$$\lambda_k(\epsilon) = (d-2k)\tfrac{\epsilon^2}{2} + O(\epsilon^4|\log\epsilon|^4) \textit{ as } \epsilon \to 0.$$

As in the previous sections, let $A = I_d + \epsilon N$ where N is a standard Gaussian matrix random variable.

Proof. Let $\eta = |\log\epsilon|$ and let bad be defined as above. We assume ϵ is sufficiently small that $\|c_j'(I+\epsilon N)\|^2 \in (\frac{1}{2}, \frac{3}{2})$ for all $N \in \mathsf{bad}^c$. Expanding, we have that

$$\begin{aligned}
\|c_j'\|^2 &= \Big\langle c_j - \sum_{i<j}\langle c_i, c_j\rangle c_i\,,\, c_j - \sum_{k<j}\langle c_k, c_j\rangle c_k\Big\rangle \\
&= \|c_j\|^2 - 2\sum_{i<j}\langle c_i, c_j\rangle^2 + \sum_{i,k<j}\langle c_i, c_j\rangle\langle c_k, c_j\rangle\langle c_i, c_k\rangle \\
&= \|c_j\|^2 - 2\sum_{i<j}\langle c_i, c_j\rangle^2 + \sum_{i<j}\langle c_i, c_j\rangle^2\|c_i\|^2 + 2\sum_{i<k<j}\langle c_i, c_j\rangle\langle c_k, c_j\rangle\langle c_i, c_k\rangle \\
&= \|c_j\|^2 - \sum_{i<j}\langle c_i, c_j\rangle^2(2-\|c_i\|^2) + 2\sum_{i<k<j}\langle c_i, c_j\rangle\langle c_k, c_j\rangle\langle c_i, c_k\rangle,
\end{aligned}$$

where to obtain the third line from the second, we separated the case $i = k$ from the case $i \ne k$.

We take a finite Taylor expansion, valid for $t \in (-1,1)$: $\log(1+t) = t - \frac{t^2}{2} + \frac{t^3}{3} - R(t)$ where $R(t) = \frac{1}{4}(1+\xi)^{-4}t^4$ for some ξ with $|\xi| \le |t|$. Let

X_j be the random variable $\|c'_j(I+\epsilon N)\|^2 - 1$. Notice from the above that X_j is a polynomial of degree 6 (whose coefficients don't depend on ϵ) in the entries of ϵN. If $N = 0$, then $c'_j(I+\epsilon N) = e_j$ so that the constant term in the polynomial X_j is 0. Notice also that by Claim 1, on bad^c, all terms other than the first term in the expression for $\|c'_j\|^2$ are $O(\epsilon^2|\log\epsilon|^2)$, while a calculation shows that $\|c_j\|^2 = 1+O(\epsilon|\log\epsilon|)$. Hence $X_j\mathbf{1}_{\mathsf{bad}^c} = O(\epsilon|\log\epsilon|)$. Let $Y_j = X_j - \frac{1}{2}X_j^2 + \frac{1}{3}X_j^3$, so that Y_j is another polynomial in the entries of ϵN with no constant term. Combining the above, on bad^c

$$\log(\|c'_j(I+\epsilon N)\|^2) = \log(1+X_j) = Y_j + O(\epsilon^4|\log\epsilon|^4).$$

Then we have

$$\begin{aligned}
&\mathbb{E}\log(\|c'_j(I+\epsilon N)\|^2) \\
&= \mathbb{E}\log(\|c'_j(I+\epsilon N)\|^2\mathbf{1}_{\mathsf{bad}^c}) + \mathbb{E}\log(\|c'_j(I+\epsilon N)\|^2\mathbf{1}_{\mathsf{bad}}) \\
&= \mathbb{E}(Y_j\mathbf{1}_{\mathsf{bad}^c}) + O(\epsilon^4|\log\epsilon|^4) + \mathbb{E}\log(\|c'_j(I+\epsilon N)\|^2\mathbf{1}_{\mathsf{bad}}) \\
&= \mathbb{E}Y_j - \mathbb{E}(Y_j\mathbf{1}_{\mathsf{bad}}) + \mathbb{E}\log(\|c'_j(I+\epsilon N)\|^2\mathbf{1}_{\mathsf{bad}}) + O(\epsilon^4|\log\epsilon|^4).
\end{aligned} \tag{2}$$

Since Y_j is a fixed polynomial function of the entries of ϵN, and all monomials that are products of entries N have finite expectation, we see that $\mathbb{E}Y_j$ agrees up to order ϵ^4 with the expectation of its terms of degree 3 or lower. Also, since the entries of N are independent and each has a symmetric distribution, the constant term of Y_j being 0, the only terms that give a non-zero contribution to $\mathbb{E}Y_j$ are the terms of the forms N_{ab}^2. Since the lowest order terms in Y_j are polynomials of degree 1, and $Y_j = X_j - \frac{1}{2}X_j^2 + \frac{1}{3}X_j^3$, the terms of the form N_{ab}^2 in Y_j are those appearing in X_j and $\frac{1}{2}X_j^2$.

We established above

$$X_j = \|c_j\|^2 - 1 - \sum_{i<j}\langle c_i, c_j\rangle^2(2-\|c_i\|^2) + 2\sum_{i<k<j}\langle c_i,c_j\rangle\langle c_k,c_j\rangle\langle c_i,c_k\rangle.$$

We see that $\|c_j\|^2 - 1 = 2\epsilon N_{jj} + \epsilon^2\sum_i N_{ij}^2$ and $\langle c_i, c_j\rangle = \epsilon(N_{ij}+N_{ji}) + \epsilon^2\sum_k N_{ki}N_{kj}$.

Substituting these in the expression for X_j, we see

$$\begin{aligned}
\mathbb{E}X_j &= d\epsilon^2 - \epsilon^2\sum_{i<j}\mathbb{E}(N_{ij}+N_{ji})^2 + O(\epsilon^4) \\
&= (d-2j+2)\epsilon^2 + O(\epsilon^4).
\end{aligned}$$

We also see $\mathbb{E}X_j^2 = 4\epsilon^2\mathbb{E}N_{jj}^2 + O(\epsilon^4)$. Combining these gives

$$\mathbb{E}Y_j = \mathbb{E}(X_j - \tfrac{1}{2}X_j^2) + O(\epsilon^4) = (d-2j)\epsilon^2 + O(\epsilon^4).$$

Therefore by (2), to finish the argument, it suffices to show $\mathbb{E}(Y_j \mathbf{1}_{\mathsf{bad}}) = O(\epsilon^4 |\log \epsilon|^4)$ and $\mathbb{E} \log(\|c'_j(I + \epsilon N)\|^2 \mathbf{1}_{\mathsf{bad}}) = O(\epsilon^4 |\log \epsilon|^4)$.

Since $\|c'_j(A)\| \geq \Theta(A)$, Lemma 6 shows

$$\mathbb{E}(\log^- \|c'_j(I + \epsilon N)\| \mathbf{1}_{\mathsf{bad}}) = O(|\log \epsilon|^2 e^{-(\log \epsilon)^2/2}).$$

Since $\|c'_j(A)\| \leq 2(\sum_{k,l} |A_{kl}|)^3$, we see $\mathbb{E}(\log^+ \|c'_j(I+\epsilon N)\|^2 \mathbf{1}_{\mathsf{bad}}) = O(\mathbb{P}(\mathsf{bad})) = O(e^{-|\log \epsilon|^2/2}/|\log \epsilon|)$.

Finally, for any of the (finitely many) monomial terms M appearing in Y_j, we can check $\mathbb{E} M \mathbf{1}_{\mathsf{bad}} = O(\mathbb{P}(\mathsf{bad})) = O(e^{-|\log \epsilon|^2/2}/|\log \epsilon|)$. This completes the proof.

□

References

[1] P. Baxendale and T. E. Harris. Isotropic stochastic flows. *Ann. Probab.*, 14:1155–1179, 1986.

[2] R. Bhatia. *Matrix analysis.* Springer-Verlag, New York, 1997.

[3] E. Dynkin. Non-negative eigenfunctions of the Laplace-Beltrami operator and Brownian motion in certain symmetric spaces. *Dokl. Akad. Nauk SSSR*, pages 288–291, 1961.

[4] I. Y. Gol'dsheid and G. A. Margulis. Lyapunov exponents of a product of random matrices. *Russian Math. Surveys*, 44:11–71, 1989.

[5] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[6] Y. Le Jan. On isotropic Brownian motions. *Z. Wahrsch. Verw. Gebiete*, 70:609–620, 1985.

[7] C. M. Newman. The distribution of Lyapunov exponents: exact results for random matrices. *Comm. Math. Phys.*, 103:121–126, 1986.

[8] J. R. Norris, L. C. G. Rogers, and D. Williams. Brownian motions of ellipsoids. *Trans. Amer. Math. Soc.*, 294:757–765, 1986.

[9] V. I. Oseledec. A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. *Trudy Moskov. Mat. Obšč.*, pages 179–210, 1968.

[10] M. S. Raghunathan. A proof of Oseledec's multiplicative ergodic theorem. *Israel J. Math.*, 32:356–362, 1979.